APMCM亚太地区大学生
建模竞赛参赛手册

APMCM ASIA PACIFIC UNDERGRADUATE MATHEMATICAL MODELING COMPETITION ENTRY MANUAL

>>扫一扫了解更多竞赛资讯<<

A P M C M

亚 太 地 区 大 学 生 数 学 建 模 竞 赛

参 赛 手 册

亚 太 地 区 大 学 生 数 学 建 模 竞 赛 组 委 会 制 作

目　录

竞赛报名通知.....1

竞赛章程.....5

历年赛题.....10

2025 Asia and Pacific Mathematical Contest in Modeling Problem A.....11

2025 Asia and Pacific Mathematical Contest in Modeling Problem B.....21

2025 Asia and Pacific Mathematical Contest in Modeling Problem C.....24

2025 Asia and Pacific Mathematical Contest in Modeling Problem D.....29

2024 Asia and Pacific Mathematical Contest in Modeling Problem A.....36

2024 Asia and Pacific Mathematical Contest in Modeling Problem B.....42

2024 Asia and Pacific Mathematical Contest in Modeling Problem C.....45

2024 Asia and Pacific Mathematical Contest in Modeling Problem D.....48

2023 Asia and Pacific Mathematical Contest in Modeling Problem A.....52

2023 Asia and Pacific Mathematical Contest in Modeling Problem B.....59

2023 Asia and Pacific Mathematical Contest in Modeling Problem C.....62

优秀论文.....64

2025 年 A 题优秀论文.....65

2025 年 B 题优秀论文.....94

2025 年 C 题优秀论文.....130

2025 年 D 题优秀论文.....157

2024 年 A 题优秀论文.....184

2024 年 B 题优秀论文.....212
2024 年 C 题优秀论文.....240
2024 年 D 题优秀论文.....271
2023 年 A 题优秀论文.....347
2023 年 B 题优秀论文.....387
2023 年 C 题优秀论文.....410

亚太地区大学生数学建模竞赛组委会

2026年APMCM亚太地区大学生数学建模竞赛报名通知

亚太赛[2026]04号

各高等院校：

2026年第十六届亚太地区大学生数学建模竞赛（以下简称“竞赛”）是由中国国际科技促进会物联网工作委员会、北京图象图形学学会联合主办的亚太地区大学生学科类竞赛，竞赛由亚太地区大学生数学建模竞赛组委会负责组织，欢迎各高等院校按照竞赛章程及有关规定组织同学报名参赛。

1、竞赛的时间确定为2026年11月26日06:00至2026年11月30日09:00。

2、本次竞赛时间为4天，参赛对象为普通高校全日制在校大学生，参赛队由1-3名大学生组成。报名截止日期为11月25日，报名截止后不能再更改报名信息。

3、竞赛允许跨校组队，必须完整填写每位参赛者以及指导教师所在的学校全称，每队最多有一名指导教师。

4、竞赛分为研究生组、本科组、专科组，报名时请根据参赛队员中最高在读学历选择组别。

每所院校参赛队数不作统一规定；组委会将根据报名情况确定获一、二、三等奖的数量（大约分别占成功参赛总队数的5%、15%、25%），优秀参赛者将享受组委会提供的其它奖励政策，请参见《亚太地区大学生数学建模竞赛奖励细则》。

5、竞赛题目共3道（A、B、C题），一般来源于人工智能、工程与管理、互联网、图象图形等领域经过适当简化加工的实际问题。

6、竞赛只需要提交电子版论文，不需要邮寄纸质版论文；所有参赛队必须提交英文版论文。

优秀参赛者邀请参加图像图形技术与应用学术会议（IGTA），会议上将对获奖代表进行颁奖。

7、组织集体报名的院校，该校负责人将所有参赛队的集体报名表发到竞赛组委会邮箱（apmcm@mathor.com），详情见《2026年亚太赛集体报名高校须知》。

8、竞赛不邮寄书面题目，竞赛开始时赛题将在竞赛官网（www.apmcm.org）、北京图象图形学学会官网（www.bsig.org.cn）以及报名主页（www.saikr.com/apmcm26）上公布。

9、其它事项请登录竞赛官网（www.apmcm.org），查看竞赛组委会的有关通知文件。

中国国际科技促进会
物联网工作委员会

北京图象图形学学会

亚太地区大学生
数学建模竞赛组委会

2026年8月

Notice about 2026 Asia and Pacific Mathematical Contest in Modeling

(APMCM)

Every institutes of higher education:

2026 Asia and Pacific Mathematical Contest in Modeling (hereinafter referred to as "contest") is held by the China International Association for Promotion of Science and Technology IoT Working Committee, Beijing Society of Image and Graphics and the Organizing Committee of Asia and Pacific Mathematical Contest in Modeling and School of Mathematics. Institutes of higher education are welcome to organize students to participate in the contest in accordance with the competition constitution and the relevant provisions.

1. The contest of 2026 begins at 6AM on Thursday November 26 and ends at 9AM on Monday November 30 in Beijing time.

2. The contest lasts for 4 days. A team may consist of two or three students and is open to all full-time undergraduate students. All the teams must be registered before 12AM Beijing time on November 25, 2026. And after this time, you can't modify your team information.

3. The team member can be from different schools. And you must fill in the school's full name of each student and advisor completely.

4. The contest is divided into three groups, namely, undergraduate student group, junior college student group and postgraduate student group. When you register, please select the group according to the highest academic qualifications in your team.

5. We have no requirement about the amount of team from the same institute of higher education. The Organizing Committee will determine the number of first, second, third prize based on the registration condition (about 5%, 15%, 25% of the total number of successful participating teams). Excellent teams will enjoy other awards provided by the Organizing Committee. Please see *Detailed rules for rewards of mathematical modeling contest for college students in Asia Pacific region* for reference.

6. We have three contest problems (A problem B problem and C problem). They are practical problems coming from engineering and management, statistics and other areas with the appropriate simplified.

7. APMCM is an all-electronic submission! You are not required to mail a print copy of your Solution Paper. And the Solution Paper must be written in English.

8. For institutes of higher education who organize students to register, please email the registration sheet to the mail of Organizing Committee (apmcm@mathor.com), and the registration fee can be submitted to the public account of Mathematician (School of Mathematics) web site. Please see *The notice to colleges for group registration for the 2026 Asia and Pacific Mathematical Contest in Modeling* for reference.

9. The contest problems will become available on two websites (http://www.apmcm.org, http://www.bsig.org.cn) when the contest begins. We will not post the problems in written form.

10. For more details, please log in our registration web site and view the relevant notice of the Organizing Committee.

中国国际科技促进会
物联网工作委员会

北京图象图形学学会

亚太地区大学生数学建模竞赛
组织委员会

2026年1月28日

亚太地区大学生数学建模竞赛组委会

APMCM亚太地区大学生数学建模竞赛章程（2026年修订）

亚太赛[2026]07号

第一章 总则

第一条 亚太地区大学生数学建模竞赛（以下简称竞赛）是由中国国际科技促进会物联网工作委员会、北京图象图形学学会主办的亚太地区大学生学科类竞赛，竞赛由亚太地区大学生数学建模竞赛组委会负责组织。

第二条 随着人工智能、自动驾驶、虚拟现实、5G等技术的快速发展，图像图形技术也是现代科技及其重要的基础。图像图形领域涵盖了数字图像处理、图像理解、计算机视觉、数据可视化、模式识别、计算机图像图形学、医学影像处理、空间信息系统等。竞赛特色是将图像图形领域的问题和数学模型相结合，普及图像图形的知识，同时利用数学的方法来解决图像图形领域的问题。竞赛宗旨：激励学生学习数学的积极性，提高学生建立数学模型和运用计算机技术解决实际问题的综合能力，鼓励广大学生踊跃参加课外科技活动，开拓知识面，培养创造精神及合作意识。

第二章 组织机构及其职责

第三条 组织机构

竞赛设组织委员会（以下简称“组委会”），由主办单位和发起单位的人员组成，如有赞助当届赛事的主要赞助单位可推荐一名代表加入当届组委会。组委会设主任委员一名、副主任委员两名，主任委员由主办单位负责人担任，副主任委员为主办单位具体负责本赛事的人员、秘书处挂靠单位分管负责人，为赛

事提供重大支持的赞助方负责人可担任当届组委会副主任。

组委会下设专家委员会和秘书处，专家委员会由各主办单位推荐从事数学建模相关研究与教学的教师组成。

第四条 组委会职责

（一）审议修改竞赛章程、实施办法和评审规则等文件；

（二）审议并确定每届赛事工作方案；

（三）通过竞赛获奖名单；

（四）审议秘书处提交的财务报告；

（五）议决竞赛的冠名、赞助、支持单位等相关问题；

（六）议决组织工作中的其它重大问题；

组委会委员若连续两次不参加组委会活动，则其所在单位视为自动退出组委会，组委会秘书处负责备案。

第五条 专家委员会职责

（一）制定命题原则，征集并拟定赛题；

（二）制定评审规则，组织专家网上评审及专家会议评审；

（三）评审参赛作品，向组委会提交拟获奖名单；

（四）受组委会委托终审有关评审结果相关异议；

（五）组织优秀作品展示与交流。

第六条 秘书处职责

（一）受组委会委托，提出竞赛章程等相关制度的修改意见；负责对每届竞赛

工作方案初审，并报组委会终审；

（二）全面了解、掌握竞赛的当前工作动态和进度；及时收集、整理有关竞赛的意见和建议并提交组委会讨论；

（三）根据优秀组织奖评选条例，提出优秀组织奖候选名单并报组委会审批；

（四）负责竞赛官方网站的建设、维护与日常运行；

（五）负责参赛队伍的网上报名、网上缴费、竞赛论文的网上在线提交等工作；

（六）完成组委会委托的其它工作。

第三章 赛程赛制及竞赛内容

第七条 赛程赛制

（一）竞赛每年举办两届，中文赛在每年上半年开始，英文赛每年下半年开始。

（二）全国统一竞赛题目，采取通讯竞赛方式，鼓励各校以相对集中的形式进行。

（三）大学生以队为单位参赛，每队1-3人，专业不限，每人最多可参加1支队伍，每队最多有1名指导老师。

第八条 竞赛内容

（一）竞赛题目一般来源于工程与管理、图象图形、计算机等领域的实际问题，并经过提炼加工，不要求参赛者预先掌握深入的专门知识。

（二）参赛者提交的作品应为一篇包括模型假设、模型建立、模型求解、模型改进以及结果分析与检验等内容的论文。作品评审的标准主要为模型及其假设与结果的合理性、创造性、文字表述的规范性。

第四章 竞赛规则

第九条 参赛对象为全日制在校大学生。

第十条 竞赛秉持公益性原则，以支定收，每支参赛队收取100元资料费，用于大赛命题、评审、证书制作及邮寄等工作。

第十一条 竞赛期间参赛队员可以使用各种图书资料、计算机和软件，在国际互联网上浏览，但不得与队外任何人（包括在网上）讨论。

第十二条 竞赛开始后，赛题将公布在指定的网址供参赛队下载，参赛队在规定时间内完成答卷，并准时交卷。

第十三条 竞赛论文若引用其他文献、图片、图表、数据应给予明确说明。

第五章 奖项设置和奖励办法

第十四条 竞赛设等级奖、优秀组织奖、优秀指导教师。

第十五条 获得等级奖、优秀组织奖、优秀指导教师的单位及个人均颁发相应荣誉证书。

第十六条 获得优秀成绩的参赛学生和指导教师，将会被邀请参加主办方的学术年会-图像图形技术与应用学术会议（IGTA）暨竞赛颁奖典礼，并在会议进程中，对获奖参赛队伍进行表彰颁奖，并有机会展示赛事内容。

第六章 获奖公示

第十七条 拟获奖名单在竞赛官网公示，公示期不少于7个工作日。公示期间，接受各校师生举报、申诉和质疑。专家委员会负责对举报与申诉的调查、仲裁与回复。公示结束后正式公布获奖名单，并不再接受质询。

第七章 知识产权

第十八条 参赛作品版权归作者所有，除用于大赛评审、官方网站学术交流与展览等大赛相关工作外，未经作者允许，不得用于商业用途，任何人不得私自盗用参赛作品成果。

第八章 附则

第十九条 本章程自2026年8月起执行，其解释和修改权属于中国国际科技促进会物联网工作委员会、北京图象图形学学会和亚太地区大学生数学建模竞赛组委会。

亚太地区大学生数学建模竞赛组委会

2026年8月

历 年 赛 题

亚太地区大学生数学建模竞赛组委会制作

2025 Asia and Pacific Mathematical Contest in Modeling

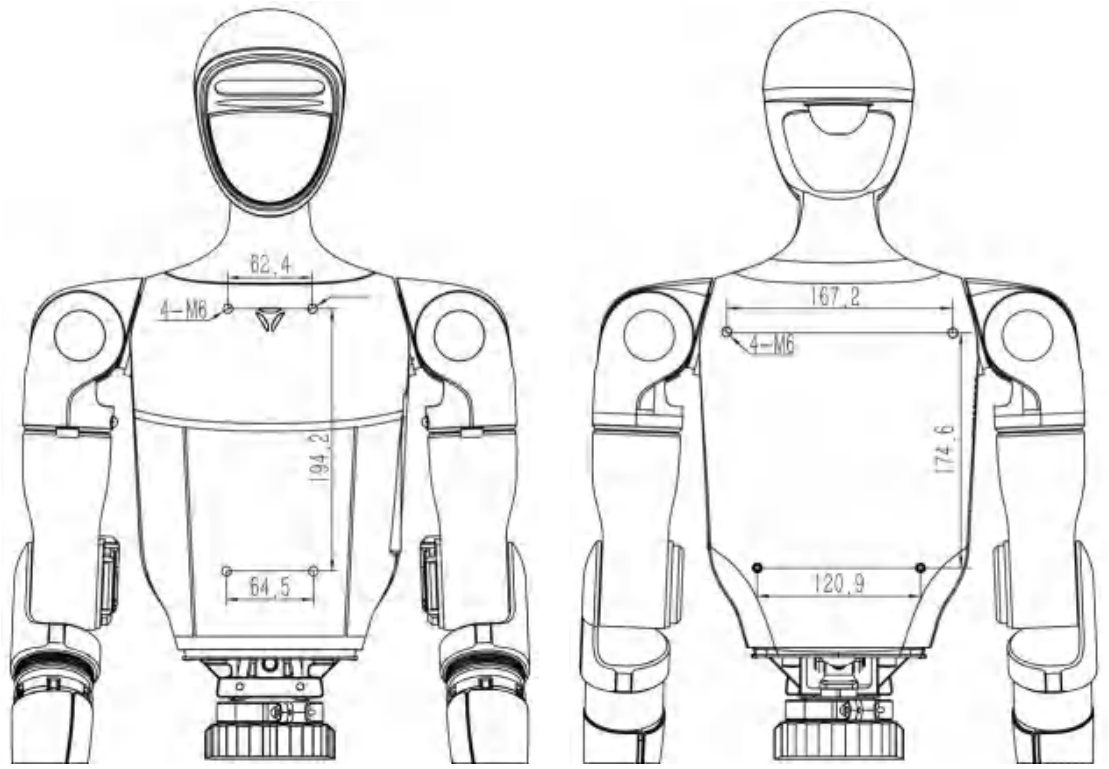
Problem A

Optimization of Robot's Performance

Hangzhou Yushu Technology is a world-renowned civilian robot company, focusing on the independent research and development, production, and sales of consumer grade, industry grade high-performance universal foot/humanoid robots and dexterous robotic arms.

Among them, the safety of the robot motor is the core of operation, and the stability of its power output directly affects the safety of the entire machine. If the motor brake fails, the speed suddenly changes, or the load matching is unbalanced, it is easy to cause the risk of losing control; However, robot falls are often related to motor abnormalities and unreasonable center of gravity design, such as insufficient motor power leading to obstacle crossing failure, excessive start stop impact disrupting posture balance, or excessively high center of gravity or center of gravity deviation beyond the support range, all of which can cause the robot to lose stability; To avoid falling, the key is to optimize the motor control algorithm to achieve smooth power output, while reasonably allocating the weight of the aircraft and reducing the height of the center of gravity, ensuring that the center of gravity always falls within the support surface, and building a dual guarantee from both motor safety and center of gravity stability.

The humanoid robot Unitree G1 of Yushu is divided into an upper body and a lower body, with multiple degrees of freedom. The schematic diagram of the front and back of Yushu's humanoid robot is shown in Figure 1.



front back

Figure 1. Schematic diagram of the front and back of Yushu's humanoid robot.

The G1 is equipped with a D435i depth camera on its head, providing the robot with excellent visual perception capabilities. It can more accurately perceive and understand the surrounding environment, achieve precise spatial perception and obstacle detection, and enable the robot to interact with the environment and respond to various scenarios more intelligently and flexibly, as shown in Figure 2.

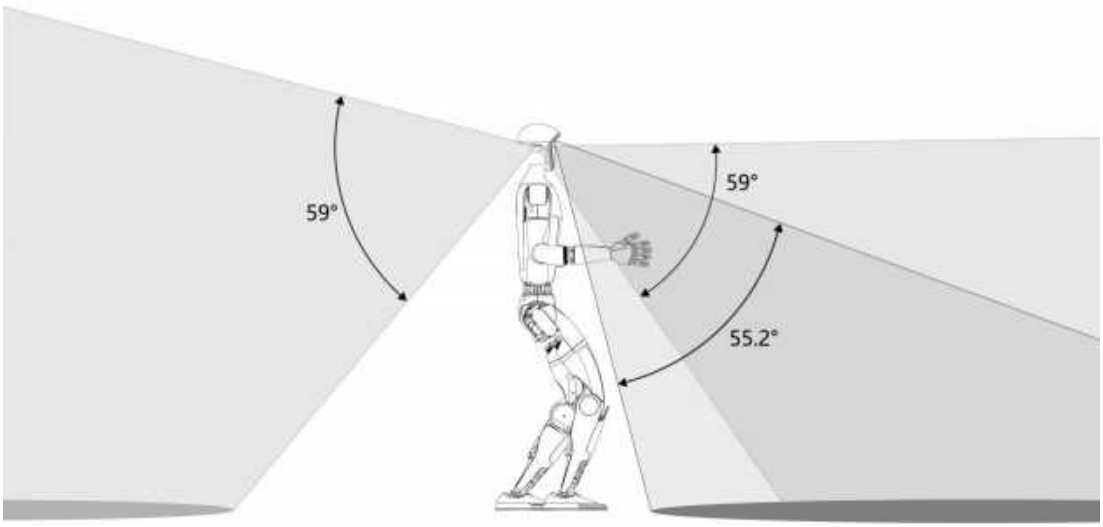


Figure 2. Schematic diagram of the maximum range and joint distribution of different visual

sensors in Yushu's humanoid robot.

When all joints are at zero degrees, the coordinate systems are shown in Figure 3.

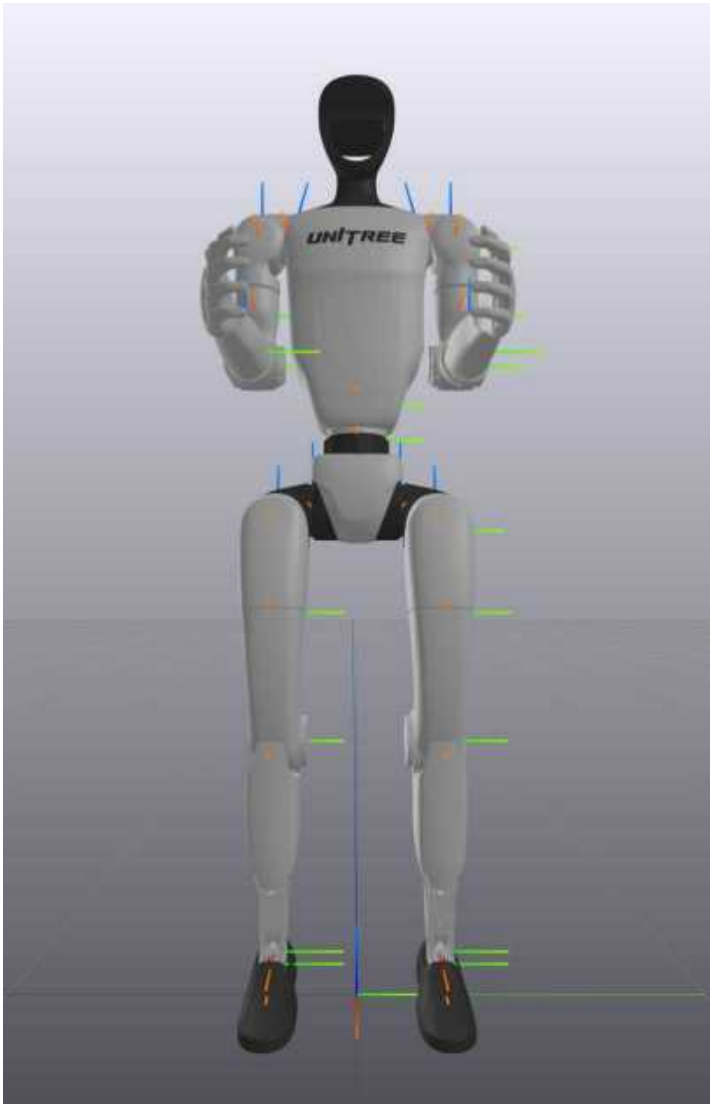


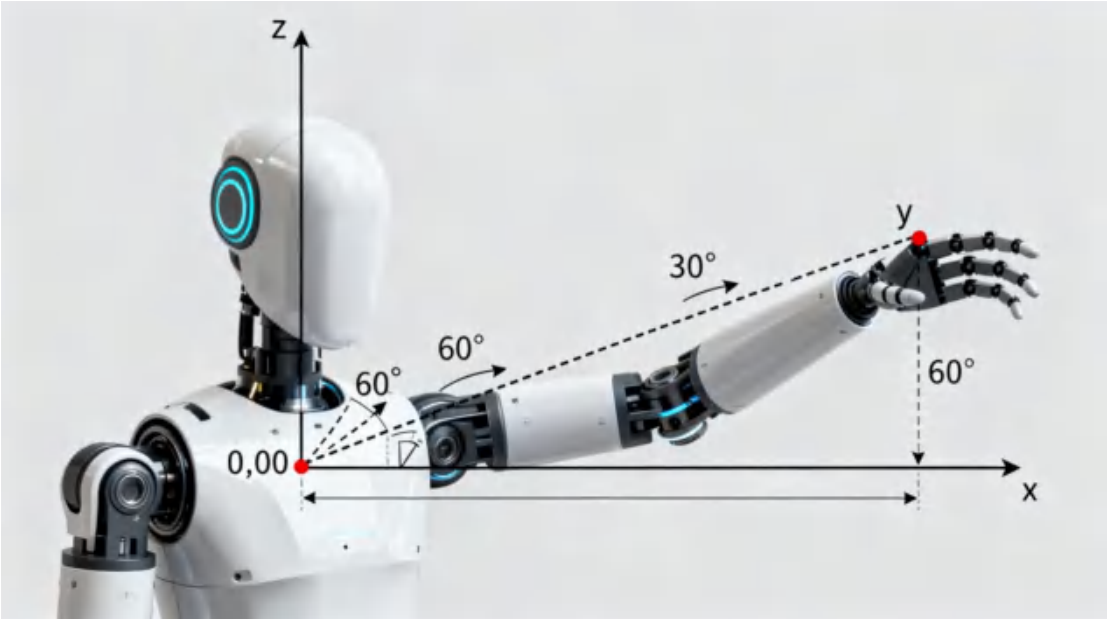
Figure 3. Schematic diagram of the coordinate system, joint rotation axis, and joint zero point of Yushu's humanoid robot.

It is known that at the opening ceremony of an important technology exhibition, the Uyghur humanoid robot Unitree G1 will perform a wonderful dance performance. To ensure the smoothness and enjoyment of the performance, precise planning of the robot's movements is necessary. The initial position of the robot is known to be at the center of the stage, and the initial posture is that all joints are at zero degrees. The performance venue is a rectangular area that is 20 meters long and 15 meters wide. Please have your team establish a corresponding mathematical model based on the above information to solve the following problems:

Problem 1. Calculation of Angles and Positions of Robot's Joints

At the beginning of the performance, the Yushu humanoid robot needs to complete a simple action: extend and lift its left hand forward from its initial position, forming a 60° angle with its body, while rotating its arm 30° to the left of its body. Assuming an arm length of 338 mm and ignoring hand dimensions, a Cartesian coordinate system is established with the robot shoulder as the coordinate origin (0,0,0), with the x-axis pointing directly in front of the stage, the y-axis pointing horizontally to the left, and the z-axis perpendicular to the stage facing upwards, as shown in Figure 4.

Please calculate the final coordinate position of the robot's left endpoint in this coordinate system. When completing this action, it is necessary to ensure that the motor operates within a safe range and verify whether the action meets the safety requirements of the motor.



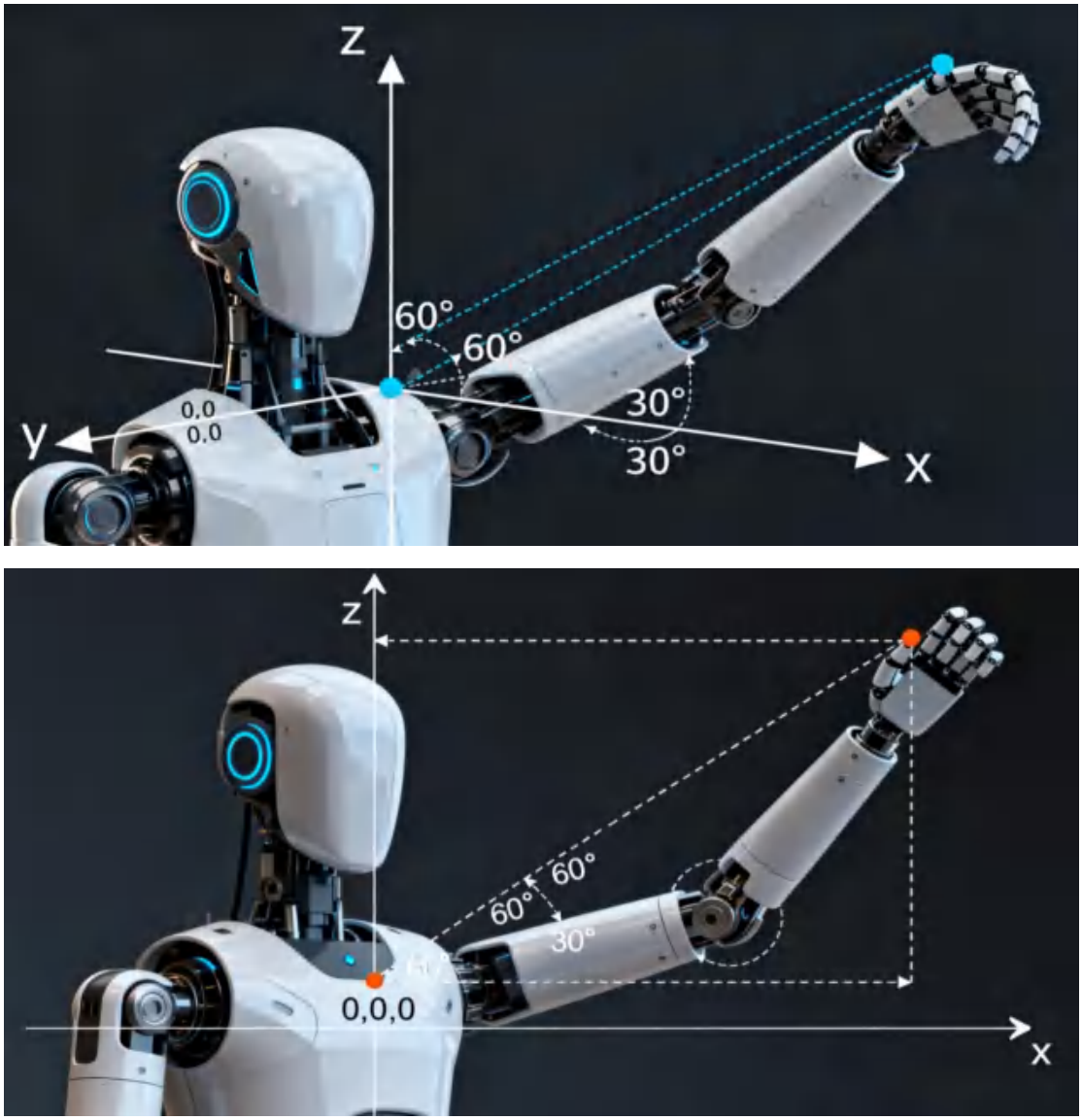


Figure 4. Schematic diagram of angle and position calculation of robot joints.

Problem 2. Modeling and Time Planning of Robot's Motion Trajectory

The Yushu humanoid robot will move along a straight line from the center of the stage (0,0, 0) to point (10, 0, 0) (unit: meter) for non-uniform motion, with an average speed of 2 m/s for straight walking. Assuming the robot starts moving from its initial position, starts at time t = 0, and reaches the target point at t = T, as shown in Figure 5.

Please establish a mathematical model between the rotation angles of each joint of the single leg of the Yushu humanoid robot and time t ($0 \leq t \leq T$), and calculate the time with maximum change in the rotation angle of the robot's knee joint during the movement from the starting point to the target point. Additionally, determine the time

T required to complete all actions.

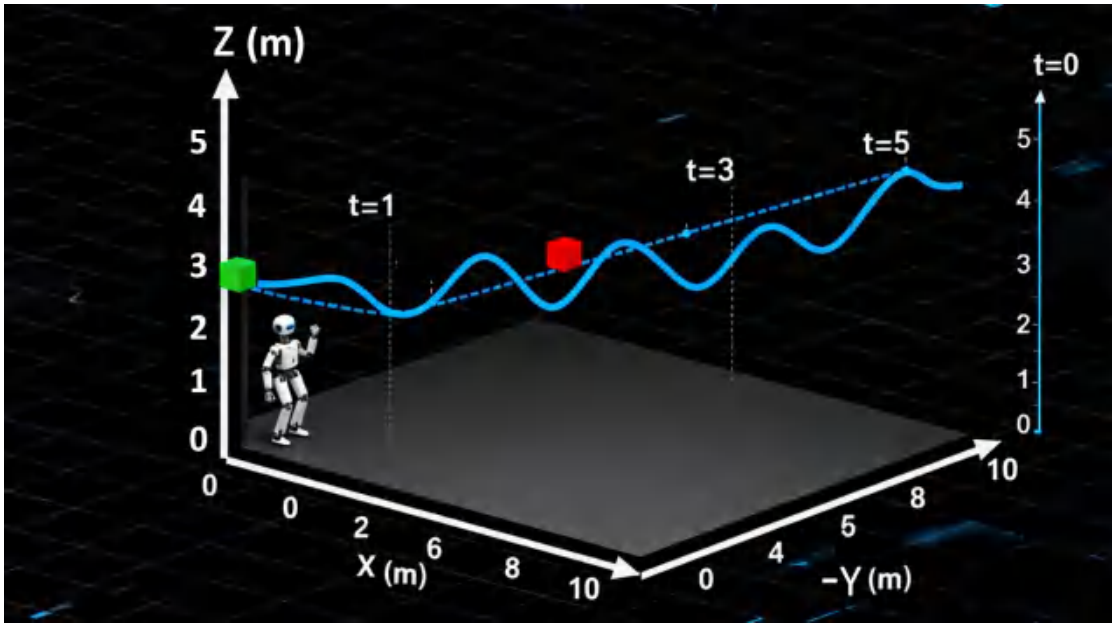


Figure 5. Schematic diagram of robot motion trajectory modeling and time planning.

Problem 3. Multi Joint Collaborative Motion Planning for Robots

At the climax of the dance performance, the Yushu humanoid robot needs to perform a complex combination of movements: the body first rotates 45° to the left, while the arms move in circles around the shoulders, with a circle radius of 300 mm and a circle drawing period of 4 seconds, and the arms move in opposite directions. In this process, to ensure the balance of the robot, corresponding adjustments need to be made to the legs, as shown in Figure 6.

Please establish a mathematical model for the angles of each joint of the arms and each joint of the legs of the Yushu humanoid robot over time t ($0 \leq t \leq 4$).



Figure 6. Schematic diagram of multi joint collaborative motion planning for robots.

Problem 4: Optimization of Robot's Performance Actions and Energy Consumption

Throughout the performance, the energy consumption of the Yushu humanoid robot was an important factor. It is known that the battery capacity of the robot is 15 Ah and the maximum voltage is 67.2 V. The power consumption of each joint motor of the robot during motion is related to the torque and speed of the motor, and relevant references can be consulted.

Please calculate the total energy consumed by the Yushu humanoid robot during the process of completing all complex actions from above problem one to problem three. Then, based on the calculation results, a model for the optimized action plan should be proposed to reduce energy consumption without affecting the performance, and the optimized energy consumption is required to be recalculated again.

References:

[1] https://www.unitree.com/cn/g1

[2] https://support.unitree.com/home/zh/G1_developer/about_G1

[3] Liang Liang, Wu Chengdong, Liu Shichang Absolute Position Accuracy Calibration Algorithm for Robots Based on Joint Geometric Errors [J]. Journal of Northeastern University (Natural Science Edition), 2025, 46 (4): 1-7

[4] Zhang Xiuli, Zhao Haoyu, Wu Jianing, etc Trajectory optimization and control method for quadruped robot jumping in the air [J]. Journal of Beijing Jiaotong University, 2024, 48 (3): 161-170

[5] Zhang Xinhao Research on Motion Control Method of Four legged Robot Based on Trajectory Optimization [D]. University of Electronic Science and Technology of China two thousand and twenty-three

[6] Zhang Xiuli, Sun Guokang, Zhou Hongmiao, Liu Ying, Li Wei A series parallel hybrid bionic robotic arm with flexible driving joints [J]. Journal of Beijing Jiaotong University, 2024, 48 (6): 154-161

[7] Zhu Haohui Optimization control method for motion trajectory accuracy of quadruped robot based on joint angle compensation [J]. Modeling and Simulation, 2024, 13 (3): 2305-2314

[8] Liu Qingyun, You Xiong, Zhang Xin, Zuo Jiwei, Li Jia Overview of Path Planning Algorithms for Mobile Robots [J]. Computer Science, 2025, 52 (6A): 240900074-10

[9] Xie Cunxi, Zheng Shixiong, Qu Hanhua Calculation of Position and Attitude of Industrial Robots and Determination of Joint Motion Parameters [J]. Mechanical Industry Automation, 1980 (4): 50-56

[10] Fu Haitao, Zhong Senming, Luo Zhida Development and Application of Joint Angle Compensation Algorithm for Industrial Robots [J]. Metrology and Testing

Technology, 2021 (010): 048

[11] Wu Wenxiang Friction analysis and low-speed high-precision motion control of multi degree of freedom serial robot joints [D]. Zhejiang University, 2013

[12] Wu Yongqiang, Tang Xianzhi, Song Wei, etc Power Equivalent Model and Parameter Identification of Industrial Robots [J]. Journal of Chongqing University, 2021 (044-010)

[13] Bai Liping Robot dynamics analysis and motor selection calculation [J]. Knowledge output of Shenyang Institute of Automation: 2000 years ago, 2000

[14] Yuan Bing Research on Jump Control Method of Four legged Robot Based on Deep Reinforcement Learning [D]. Wuhan University of Science and Technology [25-11-10]

[15] Zhang Bin Multi constraint based robot joint space trajectory planning [J]. Journal of Mechanical Engineering, 2011, 47 (21): 6

[16] Wang Mei, Wu Tiejun Research on Multi Robot Collaborative Motion Planning and Related Issues [J]. Manufacturing Automation, 2005, 27 (5): 6

[17] Zeng Qiang, Fang Yuefa Integrated coordinated motion planning for bipedal walking robots [J]. Mechanical, 2007, 34 (2): 5

[18] Chen Anjun, Liu Daohua Joint trajectory planning for coordinated motion of dual arm robots [J]. Journal of Xinyang Normal University: Natural Science Edition, 2001, 14 (1): 4

[19] Wang Mei, Wu Tiejun, Tu Dawei, etc Distributed Collaborative Motion Planning for Multiple Joint Robots [J]. Information and Control, 2011.0.2011-04-011

[20] Wang Ruimao, Yu Yueqing Terminal positioning planning for coordinated operation of flexible robots [J]. Mechanical Design and Research, 2004 (z1): 3

[21] Zhang Man, Gao Yubo, Deng Jiayu, etc Research on Linear Walking Gait Planning and Joint Trajectory of bipedal Walking Robot [J]. Forestry Machinery and

Woodworking Equipment, 2017 (006): 045

[22] Jiang Yujie, Li Jingchun, Zhang Guozhong Motion Planning of Joint type Redundant Robots [J]. Combination Machine Tool and Automation Processing Technology, 2005 (3): 2

[23] Qiu Binquan, Chen Silu, Gu Yingkui, etc Synchronization planning of robot joint trapezoidal velocity trajectory for optimal energy consumption [J]. Mechanical Design and Research, 2022 (004): 038

[24] Chen Anjun, Cai Jianle Joint trajectory planning for dual arm robots based on optimal load allocation [J]. Journal of Hebei University: Natural Science Edition, 2001, 21 (3): 6

[25] Ma Junwei Motion Planning for Complex Robots with Series Parallel Joints [D]. Harbin University of Science and Technology, 2009

2025 Asia and Pacific Mathematical Contest in Modeling

Problem B

Modeling and Optimization of Radiative Cooling Technology

Global industrialization and population growth have led to a continuous increase in energy consumption. Although the proportion of supply from clean energy sources such as solar, wind, and nuclear power has risen, petroleum and other liquid fuels remain the world's largest source of energy supply. This has led to the exacerbation of the urban heat island effect and the greenhouse effect.

According to the Stefan-Boltzmann Law, objects at higher temperatures radiate more power than objects at lower temperatures. Under the condition of an average Earth surface temperature of 15°C, most thermal radiation from objects on the Earth's surface occurs within the infrared spectral range. Consequently, scientists have proposed a technology known as Passive Daytime Radiative Cooling (PDRC). This technology is based on the principle: the high transmittance of infrared radiation through the 8-13 μm "atmospheric transparency window" and the thermal radiation energy exchange that occurs due to temperature differences. Without consuming any energy, an object on the Earth's surface can directly emit its thermal radiation energy into the cold expanse of outer space, thus spontaneously lowering its temperature, potentially even below the ambient temperature.

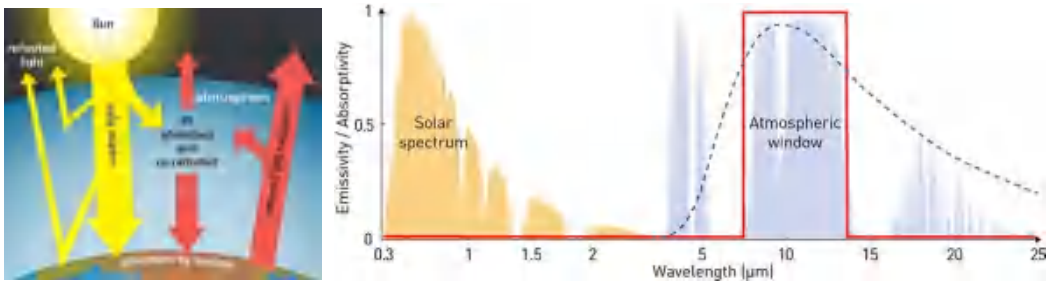


Figure 1 Radiative Cooling Technology

Compared to other cooling technologies, radiative cooling requires no energy

input while simultaneously reducing overall energy consumption, creating a virtuous cycle of energy savings. This is of great significance for achieving renewable energy development and carbon neutrality goals. The technology shows immense application potential in fields such as energy-efficient buildings, solar cells, personal thermal management, power plant condensers, and power generation and water harvesting. However, radiative cooling technology still faces several challenges on its path toward large-scale practical application. These challenges primarily lie in the need to optimize the spectral emission properties of materials, as well as complex fabrication processes and high production costs. Current research indicates that polydimethylsiloxane (PDMS) thin film is a material that possesses both ultra-high transmittance in the visible spectrum and high emissivity in the "atmospheric window" band, making it highly suitable for application in radiative cooling technology. Information on the refractive index of materials can be found on relevant websites (e.g., https://refractiveindex.info/).

1.Please establish a suitable mathematical model to study the emissivity of PDMS thin films as a function of wavelength for different film thicknesses.

2.Please establish a suitable evaluation model to assess the radiative cooling performance of PDMS thin films of varying thicknesses. Based on your findings, provide some recommendations related to the advancement and application of radiative cooling technology.

3.Please select suitable materials to form a multilayer film structure in conjunction with PDMS. Incorporating an optimization design algorithm, perform an optimization of the material selection and layer thickness for each layer within the multilayer structure.

4.Please perform a comprehensive optimal design of the radiative cooling materials and structures to achieve the best possible radiative cooling performance. Following the design, conduct an evaluation of the feasibility and cost of the resulting radiative cooling product.

References：

[1] LI X, ZHOU Y, YU S, et al. Urban heat island impacts on building energy consumption: a review of approaches and findings[J]. Energy, 2019, 174: 407-419.

[2] SANTAMOURIS M, SYNNEFA A, KARLESSI T. Using advanced cool materials in the urban built environment to mitigate heat islands and improve thermal comfort conditions[J]. Solar Energy, 2011, 85(12): 3085-3102.

[3] MANDAL J, YANG Y, YU N, et al. Paints as a scalable and effective radiative cooling technology for buildings [J]. Joule, 2020, 4(7): 1350-1356.

2025 Asia and Pacific Mathematical Contest in Modeling

Problem C

Impact of U.S. Tariff Policies on Global Trade and Regional Economies

On April 2, 2025, the Trump administration introduced the so-called "Reciprocal Tariffs" policy, establishing a 10% "minimum baseline tariff" on all trading partners and imposing additional higher tariffs on approximately 60 countries and regions with which the U.S. ran trade deficits, including China. The U.S. claimed that varying tariff rates and non-tariff barriers among trading partners made it harder for American manufacturers to sell products abroad, leading to persistent U.S. trade deficits and manufacturing decline. Thus, it imposed "reciprocal tariffs" on all countries to reduce its trade deficit and promote manufacturing reshoring [1]. However, this policy is inconsistent with the fundamental principle of "Most-Favored-Nation (MFN) Treatment" of the World Trade Organization (WTO) and also challenges the WTO's tariff preference principle for developing countries. Under WTO rules, members must apply uniform tariffs to all partners, but the Trump administration exercised unilateral pressure and bilateral negotiations to implement widely varying tariff standards across economies, with agreed rates far exceeding previous levels. According to a tariff tracking tool jointly developed by the WTO and International Monetary Fund (IMF), the U.S. trade-weighted average tariff rate on all products rose sharply from 2.44% at the start of 2025 to 20.11% by August 2025 [2]. A Yale Budget Lab report noted that this tariff increase pushed the U.S. overall average effective tariff rate to its highest level since 1933 [3].

Beyond serving as a source of fiscal revenue, tariffs are an important tool of national trade policy. Appropriate tariffs can protect domestic industries from global economic monopoly and coercive practices, and are therefore permitted by the WTO.

However, the impact of tariffs is multifaceted. Excessively high tariffs hinder the development of global trade and are detrimental to long-term world economic growth. In the short term, high U.S. tariffs may bring benefits such as increased tariff revenue and a reduced trade deficit, but they also lead to declining exports and higher domestic consumer prices. For example, according to statistics from General Administration of Customs of the People's Republic of China (GACC) [4], in the month the "reciprocal tariffs" took effect, U.S. imports from China fell 20% year-on-year, while U.S. exports to China dropped 13% due to Chinese countermeasures. By May 2025, imports from China had fallen 35% compared with May 2024, and exports to China declined 18%. In absolute terms, the U.S. trade deficit with China narrowed from $30.8 billion in May 2024 to $18.0 billion in May 2025, but U.S. exports to China also fell from $13.2 billion to $10.4 billion, negatively impacting related U.S. industries. Overall, the adverse effects of high tariffs on U.S. trade and the economy cannot be overlooked.

Your team is tasked with analyzing the impact of U.S. tariff policy adjustments on international trade and regional economies, based on U.S. tariff and global trade data. Address the following questions:

Question 1: The United States, Brazil, and Argentina are the world's largest soybean producers and exporters, while China is the largest soybean importer. Analyze the current state of soybean trade between China and these countries. Develop a model to assess the impact of U.S. tariff adjustments on the soybean industries of the U.S., Brazil, and Argentina. Estimate the post-adjustment distribution of soybean export volumes and values for these three countries.

Question 2: The United States is the world's second-largest automobile market and the largest automobile importer. In 2024, 46% of cars sold in the U.S. were imported [5], with Japan as a major source. Besides direct exports from Japan, Japanese automakers also invest in U.S. production or manufacture in countries like Mexico for export to the U.S. Analyze the current position of Japanese automobiles in the U.S. market. Develop a model that incorporates Japan's non-tariff response

strategies and economic transmission effects to examine the impact of U.S. tariff adjustments on U.S.-Japan automobile trade, the structure of U.S. auto imports, and the U.S. automobile industry.

Question 3: The United States dominates the global semiconductor (chip) industry but lacks strong manufacturing capacity. The Biden administration passed the CHIPS and Science Act in August 2022, using subsidies to boost advanced semiconductor manufacturing. However, the Trump administration in 2025 argued that tariffs are more effective than subsidies. Concurrently, the U.S. imposed export controls on high-end chips to China, citing national security and technological leadership concerns. Develop a model analyzing the effects of U.S. tariff policies on domestic semiconductor manufacturing and trade in high-, mid-, and low-end chips, considering both economic efficiency and national security.

Question 4: Raising tariff rates may increase government tariff revenue in the short term but reduce trade volumes, potentially lowering revenue over time. Develop a model to analyze the short- and medium-term impact of U.S. tariff adjustments on U.S. tariff revenue. Predict the net change resulting from these adjustments in U.S. tariff revenue during the second term of the Trump administration.

Question 5: U.S. tariff hikes may trigger countermeasures from trading partners. For instance, China imposed reciprocal tariff countermeasures and implemented targeted export controls on rare earths, lithium batteries, etc. [9]. These countermeasures will undoubtedly bring complex impacts on U.S. foreign trade, thereby exerting comprehensive and profound effects on agriculture, industry, and other sectors within the United States, and may also affect financial markets such as the U.S. dollar, U.S. Treasury bonds, and virtual currencies. Select or construct economic indicators to model the short- and medium-term impact of U.S. tariff adjustments on the U.S. economy. Use the model to evaluate whether the "reciprocal tariffs" policy can genuinely drive manufacturing reshoring.

Glossary：

- **Reciprocal Tariffs**: Originally refers to a system where two or more countries mutually reduce or eliminate import tariffs through bilateral or multilateral negotiations. Here, it specifically denotes the tariff policy announced by the Trump administration on April 2, 2025. While nominally implying equal tariff rates between the U.S. and its partners, the U.S. imposed a 10% "minimum baseline tariff" on all partners and additional tariffs ranging from 11% to 50% on approximately 60 trading partners.

- **Most-Favored-Nation (MFN) Treatment**: Refers to the tariff and non-tariff preferential treatments mutually granted by WTO members that are no less favorable than those granted to any third country. According to Article I of the General Agreement on Tariffs and Trade (GATT 1994), any preference granted by a member to another member must be immediately and unconditionally extended to all other members.

Data Notes：

The attached data provides U.S. foreign trade data from 2020 onward, sourced from U.S. Trade & Tariff Data [6]. Teams are not limited to this data and may obtain additional trade data from official sources such as [4] or non-tariff policy information from sources like [9]. Reasonable simulated data is also permitted.

References:

[1] The White House, Regulating Imports with a Reciprocal Tariff to Rectify Trade Practices that Contribute to Large and Persistent Annual United States Goods Trade Deficits.

https://www.whitehouse.gov/presidential-actions/2025/04/regulating-imports-with-a-reciprocal-tariff-to-rectify-trade-practices-that-contribute-to-large-and-persistent-annual-united-states-goods-trade-deficits/

[2] China Financial Information Network, [Global Finance] U.S. Average Tariff Rate Surges Significantly.

https://www.cnfin.com/hg-lb/detail/20250809/4283370_1.html

[3] Yale Budget Lab, State of U.S. Tariffs: August 7, 2025.

https://budgetlab.yale.edu/research/state-us-tariffs-august-7-2025

[4] General Administration of Customs of the People's Republic of China, Customs Statistics.

http://www.customs.gov.cn/customs/302249/zfxxgk/2799825/302274/index.html

[5] China Automotive News, U.S. Establishes New Tariff Barriers: Impact on Imported Cars?

http://www.cnautonews.com/yaowen/2025/02/28/detail_20250228372164.html

[6] Data Web: U.S. Trade & Tariff Data.

https://dataweb.usitc.gov/

[7] Stephen Miran. A User’s Guide to Restructuring the Global Trading System. Hudson Bay Capital.

https://www.hudsonbaycapital.com/documents/FG/hudsonbay/research/638199_A_Users_Guide_to_Restructuring_the_Global_Trading_System.pdf

[8] Digital Economy Laboratory, University of International Business and Economics, Monthly Monitoring Report on China-U.S. Goods Trade (Jan-Aug 2025).

https://delab.uibe.edu.cn/kydt/cgzs/2025n/2025n8yjcbg/d9650830544c49e2812662f722b08325.htm

[9] Ministry of Commerce of the People's Republic of China, Announcements & Orders.

https://www.mofcom.gov.cn/zcfb/blgg/index.html

Problem D

Quantum-based Optimization for Unit Commitment Problems in Power Systems

Unit commitment (UC) is a fundamental optimization problem in power system operation. Its goal is to determine the on/off status and generation levels of thermal units over a given time horizon to minimize operating costs while satisfying demand and system constraints. The UC problem is known to be a large-scale mixed integer programming (MIP) challenge, whose computational complexity grows exponentially with system size.

With the emergence of quantum optimization technologies such as the coherent Ising Machine (CIM), the UC problem can be reformulated as a quadratic unconstrained binary optimization (QUBO) model. This transformation enables the use of quantum-inspired solvers for fast and parallel search of optimal solutions, paving the way for real-time power system operation on quantum hardware.

We invite participants to bridge classical power system optimization with quantum computing through a three-stage workflow: first, building a classical unit commitment model; second, transforming it into a QUBO formulation; and third, designing a scalable reduction strategy to accommodate current quantum hardware constraints.

Problem Description

Participants are required to construct a complete UC optimization framework that integrates classical and quantum methods.

The classical UC formulation is typically based on standard economic dispatch principles, including fuel costs, startup/shutdown costs, and operational constraints.

Padhy et al [1] provides a comprehensive survey of UC modeling and solution approaches, serving as a key reference. In addition, Mehta et al [2] analyze the computational hardness of QUBO problems, offering insights relevant for mapping and solving UC problems in combinatorial optimization frameworks.

Participants are required to construct a complete UC optimization framework that integrates both classical and quantum optimization methods. All generator, network, and load parameters required for UC modeling are provided in Tables 1-5. Using these data, participants must complete the following four tasks in sequence, and are allowed to simplify network and unit parameters as needed when the QUBO matrix beyond the bit-width limitations of the target quantum hardware.

Participants are tasked with a multi-step process: 1) constructing and solving a classical UC model; 2) extending this model with network and security constraints; 3) transforming the complete UC formulation into a QUBO model and solving it using quantum-inspired methods; and finally, 4) designing a reduction strategy to ensure the problem scale complies with qubit constraints.

Each task below is weighted according to its difficulty and importance in the quantum optimization workflow.

Problem 1. Classical UC Modeling and Optimization: formulate and solve a classical Unit Commitment model using conventional optimization tools (e.g., Gurobi or CPLEX).

To ensure consistency and fairness across teams, the classical UC model must include at least the following components:

1. Objective Function

(1) Fuel cost: a quadratic or piecewise-linear production cost for each generating unit;

(2) Startup/shutdown cost: a cost linked to unit startup/shutdown transitions.

2. Constraints

(1) Power balance constraint: total generation must match the system load in each time period.

(2) Generation limits: Each unit must operate within its minimum and maximum output limits whenever committed.

(3) Minimum up/down time constraints: after being turned on, a unit must remain online for a minimum number of periods; after shutdown, it must remain offline for a minimum number of periods.

(4) Ramp-rate constraints: the change in unit output between consecutive periods must not exceed ramp-up and ramp-down limits.

(5) Startup/shutdown logic constraints: logical relationships ensuring that commitment status, startup indicators, and shutdown indicators are consistent.

Problem 2. UC Modeling with Network and Security Constraints: Based on the classical UC formulation developed in Problem 1, participants shall further enhance the model by incorporating system-level network and security constraints. In this task, the UC model must be extended to include:

(1) Network power flow constraints: such as DC power flow limits on all transmission lines.

(2) N-1 security constraints: ensuring that the system remains operationally feasible under any single generator or transmission line outage.

(3) Spinning-reserve requirements: guaranteeing sufficient reserve capacity in each time period to maintain reliability.

The following constraint is not mandatory for completing **Problem 2.**:

(4) Minimum safety inertia constraint (optional): enforcing a system-wide minimum inertia requirement using from the reference[3]. Only units that are already online in each period contribute to the inertia(H).

These additional constraints aim to bring the UC model closer to real-world operational conditions and increase its modeling depth and complexity. Participants shall use the same optimization solver adopted in **Problem 1** (e.g. Gurobi or CPLEX) to solve this extended UC model, and are required to compare the results of **Problem 1** and **Problem 2.**

Problem 3. QUBO Transformation and Kaiwu SDK Quantum Solution: Transform the UC model in **Problem 2** into a QUBO representation and solve it using the Kaiwu SDK. Participants are required to compare the quantum-based solutions with the traditional optimization results obtained

Problem 4. Problem Scale Reduction under Quantum Hardware Constraints: Participants shall design and implement an effective problem-scale reduction strategy specifically for the **Problem 2** UC model (or corresponding QUBO model in **Problem 3**), whose QUBO representation is typically too large to be executed directly on current quantum hardware. The reduction method should ensure that, under the bit-capacity limitation of the CIM (e.g., XXX bits), the resulting QUBO can be successfully solved.

Table1 Parameters of each unit

Unit (bus)	Maximum power generation (MW)	Minimum power generation (MW)	Minimum Up Time (h)	Startup Cost ($)	Shutdown Cost ($)	Ramp-Up Limit (MW/h)
1	300	50	8	180	180	80
2	180	20	8	180	180	80
5	50	15	5	40	40	50
8	35	10	5	60	60	50
11	30	10	6	60	60	60
13	40	12	3	40	40	60

Table2 Parameters of each unit

Unit (bus)	Ramp-Down Limit (MW/h)	Minimum Up Time (h)	Minimum Down Time (h)	Initial Up Time (h)	Initial Down Time (h)	a	b	c	H
1	80	5	3	5	0	0.02	2.00	0	7.0
2	80	4	2	4	0	0.0175	1.75	0	4.5
5	50	3	2	3	0	0.0625	1.00	0	4.5
8	50	3	2	3	0	0.00834	3.25	0	3.2
11	60	1	1	3	0	0.025	3.00	0	3.0
13	60	4	2	4	0	0.025	3.00	0	3.0

Table3 Grid parameters

Branch number	From bus	To Bus	r	x	b	Max power transmission (MW)
1	1	2	0.02	0.06	0.03	650
2	1	3	0.05	0.19	0.02	650
3	2	4	0.06	0.17	0.02	325
4	3	4	0.01	0.04	0	650
5	2	5	0.05	0.20	0.02	650
6	2	6	0.06	0.18	0.02	325
7	4	6	0.01	0.04	0	450
8	5	7	0.05	0.12	0.01	350
9	6	7	0.03	0.08	0.01	650
10	6	8	0.01	0.04	0	160
11	6	9	0	0.21	0	325
12	6	10	0	0.56	0	160
13	9	11	0	0.21	0	325
14	9	10	0	0.11	0	325
15	4	12	0	0.26	0	325
16	12	13	0	0.14	0	325
17	12	14	0.12	0.26	0	160
18	12	15	0.07	0.13	0	160
19	12	16	0.09	0.20	0	160
20	14	15	0.22	0.20	0	80
21	16	17	0.08	0.19	0	80
22	15	18	0.11	0.22	0	80
23	18	19	0.06	0.13	0	80
24	19	20	0.03	0.07	0	80
25	10	20	0.09	0.21	0	80
26	10	17	0.03	0.08	0	80

27	10	21	0.03	0.07	0	80
28	10	22	0.07	0.15	0	80
29	21	22	0.01	0.02	0	160
30	15	23	0.10	0.20	0	160
31	22	24	0.12	0.18	0	160
32	23	24	0.13	0.27	0	160
33	24	25	0.19	0.33	0	80
34	25	26	0.25	0.38	0	80
35	25	27	0.11	0.21	0	80
36	27	28	0	0.40	0	80
37	27	29	0.22	0.42	0	80
38	27	30	0.32	0.60	0	80
39	29	30	0.24	0.45	0	80
40	8	28	0.06	0.20	0.02	160
41	6	28	0.02	0.06	0.01	160

Table4 Load demands of 24 periods

Time period (h)	Load demand (MW)
1	166
2	196
3	229
4	257
5	263
6	252
7	246
8	213
9	192
10	161
11	147
12	160
13	170
14	185
15	208
16	232
17	246
18	241
19	236
20	225
21	204
22	182
23	161

24	131

Table5 Inertia-related security related parameters

ROCOF(HZ/s)	F	load step (MW)
0.5	2.118	600

Reference：

[1] Padhy, Narayana Prasad. "Unit commitment-a bibliographical survey." IEEE Transactions on Power Systems. 19.2 (2004): 1196-1205.

[2] Mehta V, Jin F, Michielsen K, et al. On the hardness of quadratic unconstrained binary optimization problems[J]. Frontiers in Physics, 2022, 10: 956882.

[3] Xia W, Wang Z, Wang J, et al. Evaluation of minimum safety inertia of new energy power system considering rate of change of frequency[C]//2022 12th International Conference on Power and Energy Systems (ICPES). IEEE, 2022: 618-622.

[4] Kaiwu SDK download : https://platform.qboson.com/login

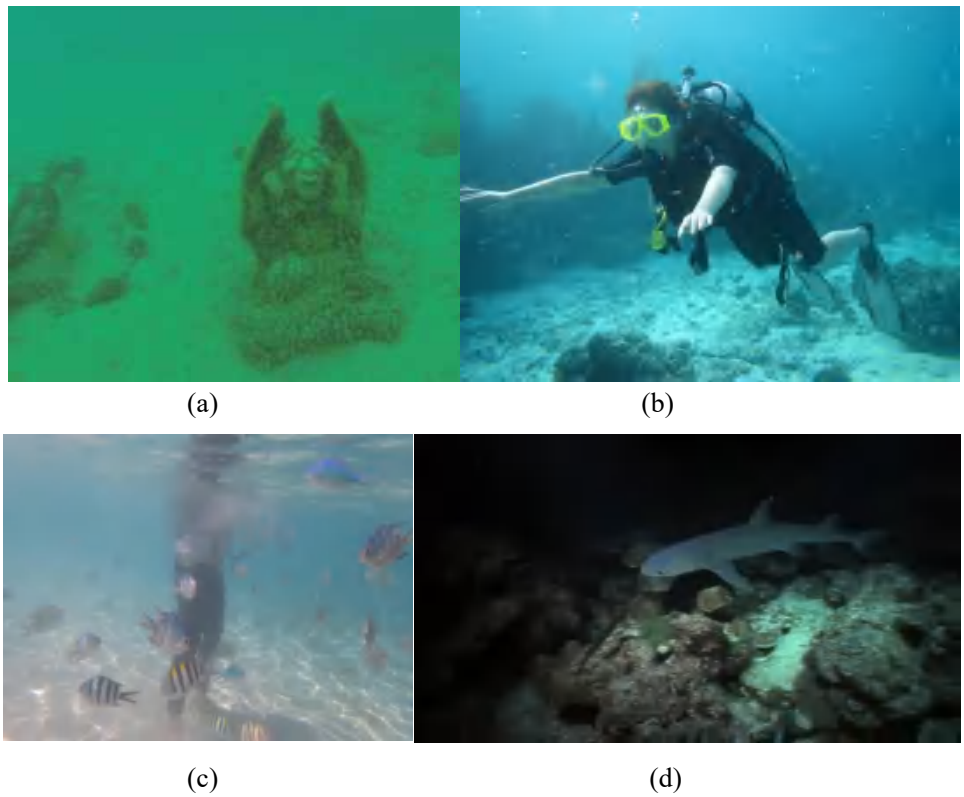
2024 Asia and Pacific Mathematical Contest in Modeling

Problem A

Research on Underwater Image Enhancement in Complex Scenarios

For ocean exploration, clear and high-quality underwater images are crucial for deep-sea topography surveying and seabed resource investigation. **However, in complex underwater environments, the image quality deteriorates due to phenomena such as absorption and scattering of light during its propagation in water, resulting in blurriness, low contrast, color distortion, etc. These conditions are referred to as underwater image degradation. The main causes of underwater image degradation include light propagation loss in water, forward scattering and backward scattering effects, as well as the scattering effect of suspended particles on light [5].**

These factors collectively result in the loss of details and clarity during the transmission process of underwater images, affecting visual recognition and analysis.



(a) (b)

(c) (d)

Figure 1. Schematic diagram of underwater image degradation, (a) shows green color cast, (b) shows blue color cast, (c) shows imaging blur, (d) shows insufficient light.

The schematic diagram of the underwater imaging process is shown in Figure 2. According to the Jaffe-McGlamery underwater imaging model, **the underwater image captured by the camera can be represented as a linear combination of three components: direct component, forward scattering component, and backward scattering component [1].** Among them, the forward scattering component refers to the light that enters the imaging system after being scattered by suspended particles in water from target surface reflection or radiation. This component will cause blurring in the obtained image. The backward scattering component refers to the light that enters the imaging system after natural light entering water is scattered by suspended particles, resulting in low contrast in the obtained image. In general cases, due to close distance between objects and cameras, a simplified imaging model is used:

$$I(x) = J(x)t(x) + B(t(x)),$$

where $I(x)$ represents the degraded underwater image, $J(x)$ represents the clear image, B is the ambient light in the underwater environment, and $t(x)$ is the light transmission function of the underwater scene. The light transmission rate varies under different conditions. At the same time, the underwater ambient light also changes with factors such as depth and the turbidity of the water, all of which can lead to increased degradation of underwater images.

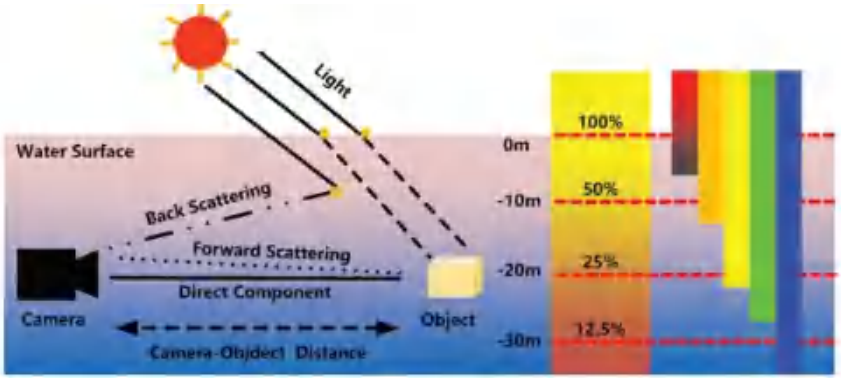


Figure 2. Conceptual diagram of underwater image degradation principle [6]

Before performing enhancement and other processing operations on underwater images, it is necessary to conduct statistical analysis on the image to be processed, as shown in Figure 3. Image analysis generally utilizes mathematical models combined with image processing techniques to analyze underlying features and higher-level structures, thereby extracting intelligent information. For example, using a histogram can statistically analyze the distribution of colors in different channels of the image, while applying edge operators can provide clarity information about object contours in the image. These pieces of information help us classify images into different categories and propose targeted solutions for image enhancement.

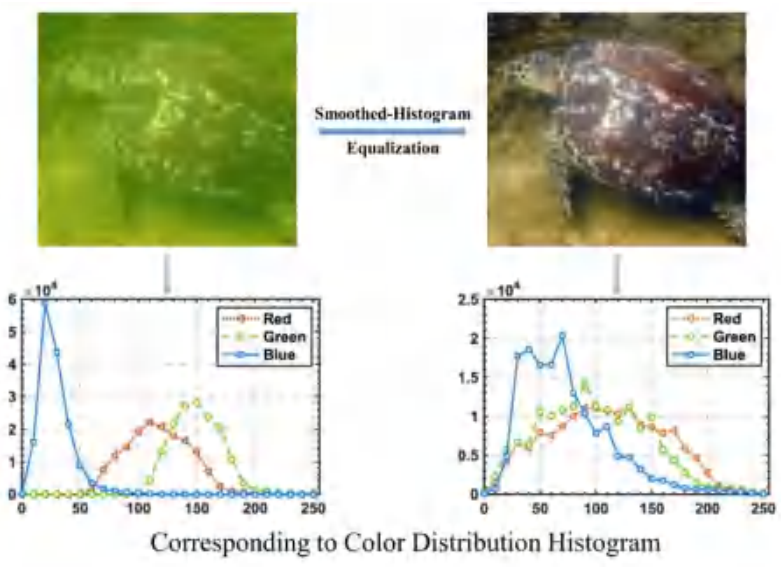


Figure 3. Color distribution curves of underwater images before and after using enhancement techniques, with each channel's distribution curve being more balanced and closer to each other, thus improving the visual effect [7].

Underwater image enhancement technology is a technique that improves the quality of images captured in underwater environments by applying signal processing, image processing, and machine learning theories. **It aims to reduce problems such as image blur, color distortion, and decreased contrast caused by the absorption and scattering of light in water, thereby improving the visibility and clarity of underwater images.**

Underwater image enhancement and restoration methods can be divided into **traditional methods and deep learning methods**. Traditional methods can be further categorized into **non-physical models and physics-based** models. Non-physical model methods improve visual quality by **directly adjusting the pixel values of images**, including **applying existing image enhancement methods and specially designed algorithms**. Physics-based model methods **model and estimate parameters to invert the degradation process of underwater images.** These methods can invert based on assumptions or prior knowledge, or they can use the optical properties of underwater imaging to improve the restored images. However, due to the complexity of underwater scenes, most existing methods cannot handle all scenarios. Therefore, an underwater scene enhancement algorithm tailored for complex scenarios is very important for subsequent tasks in underwater vision.

To estimate the degree of underwater image degradation in different scenarios and provide targeted enhancement methods, please answer the following questions:

Question 1: Please use image statistical analysis techniques similar to those mentioned in the above text to perform multi-angle analysis on the underwater image provided in Attachment 1. Classify the image provided in Attachment 1 into **three** categories: **color cast, low light,** and **blur**, and fill in the filenames in the three positions in the "Answer.xls" attachment. Also, explain the reasons for such classification.

Question 2: Based on the types of degradation proposed in Question 1, using the underwater imaging model provided in the problem, construct an underwater scene image degradation model with the images attached. Analyze the degradation reasons of underwater images captured from different scenes [1] (including but not limited to color cast, low light, etc.). Analyze the similarities or differences of these degradation models (for example, categorize from perspectives such as color, lighting, clarity, etc.).

Question 3: Based on the underwater scene image degradation model established in Question 2, propose an underwater image enhancement method tailored for a single scene (such as color cast, blur, low light), and validate the proposed enhancement method using the image data provided in the attachment. **Include the enhanced results of the test images from Attachment 2 and their corresponding evaluation metrics in the paper, calculate and present the PSNR, UCIQE, UIQM, and other evaluation metrics for the output images**, and **fill them in the table Attachment 1 results provided in "Answer.xls".**

Question 4: The modeling adaptability of existing underwater image enhancement models varies across different scenarios. Please, in conjunction with the above question and the images provided in the attachment, propose an underwater image enhancement model tailored for complex scenarios **(e.g., a non-physical model, references [2]-[5] may be consulted).** This model should be capable of enhancing underwater image degradation issues across a variety of complex scenes. **Include the enhanced results of the test images from attachment 2 and their corresponding evaluation metrics in the paper for display**, calculate and output the PSNR, UCIQE, UIQM, and other evaluation metrics of the output image, and **fill them in the table Attachment 2 results provided in "Answer.xls".**

Question 5: Compare various enhancement techniques for specific scenarios with a single enhancement technique for complex scenarios, and propose feasibility suggestions for underwater visual enhancement in practical applications.

Attachment:

Attachment: https://pan.baidu.com/s/1tXeoiDvxsGbIQ42obLf5xw?pwd=2024 code: 2024

Appendix:

Underwater image evaluation metrics:

PSNR (Peak Signal-to-Noise Ratio), UCIQE (Underwater Color Image Quality Evaluation), and UIQM (Underwater Image Quality Measure) are crucial indicators used to assess image quality. Below is a comprehensive description of the calculation methods for these three metrics.

PSNR (Peak signal-to-noise ratio) :

PSNR is an objective standard for measuring image quality, which calculates the mean square error (MSE) between the original image and the processed image, and converts it into decibel units to measure image quality.

UCIQE (Underwater Color Image Quality Evaluation) :

UCIQE [9] is a linear combination of color density, saturation, and contrast, used to quantitatively evaluate the non-uniform color cast, blurriness, and low contrast in underwater images. It is an image quality assessment metric that does not require a reference (ground truth) image. The specific definition is as follows:

$$UCIQE = c_1 \times \sigma_c + c_2 \times con_l + c_3 \times \mu_s$$

Among them, σ_c represents the chromaticity standard deviation, con_l represents brightness contrast, and μ_s represents average saturation. c_1, c_2, and c_3 are weighting coefficients. As mentioned above, there is a good correlation between chromatic variance and human perception for underwater color images of interest. Using chromatic differences to describe color bias also has other reasons.

UIQM (Underwater Image Quality Measure) :

UIQM [8] is an evaluation index that comprehensively considers attributes such as colorfulness, sharpness, and contrast of underwater images. The calculation formula is:

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM,$$

where UICM, UISM, and UIConM represent the measurements of colorfulness, sharpness, and contrast of underwater images respectively. c_1, c_2, and c_3 are the weighting coefficients.

References:

[1]. Cong Runmin, Zhang Yumo, Zhang Chen, Li Chongyi, Zhao Yao. Research Progress of Deep Learning Driven Underwater Image Enhancement and Restoration[J]. JOURNAL OF SIGNAL PROCESSING, 2020, 36(9): 1377-1389. DOI: 10.16798/j.issn.1003-0530.2020.09.001

[2]. Li C, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement[J]. Pattern Recognition, 2020, 98: 107038.

[3]. Ancuti C O, Ancuti C, De Vleeschouwer C, et al. Color balance and fusion for underwater image enhancement[J]. IEEE Transactions on image processing, 2017, 27(1): 379-393.

[4]. Wang Yongxin, Diao Ming, Han Chuang. Underwater Image Restoration Algorithm Based on Least Squares Estimation[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(11): 2125-2133. DOI: 10.3724/SP.J.1089.2018.17041

[5]. Gao, X.; Jin, J.; Lin, F.; Huang, H.; Yang, J.; Xie, Y.; Zhang, B. Enhancing Underwater Images through Multi-Frequency Detail Optimization and Adaptive Color Correction. J. Mar. Sci. Eng. 2024, 12, 1790. https://doi.org/10.3390/jmse12101790

[6]. Zhou J, Li B, Zhang D, et al. UGIF-Net: An efficient fully guided information flow network for underwater image enhancement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023.

[7]. Chen R, Cai Z, Yuan J. UIESC: An underwater image enhancement framework via self-attention and contrastive learning[J]. IEEE Transactions on Industrial Informatics, 2023, 19(12): 11701-11711.

[8]. K. Panetta, C. Gao and S. Agaian, "Human-Visual-System-Inspired Underwater Image Quality Measures," in IEEE Journal of Oceanic Engineering, vol. 41, no. 3, pp. 541-551, July 2016, doi: 10.1109/JOE.2015.2469915

[9]. M. Yang and A. Sowmya, "An Underwater Color Image Quality Evaluation Metric," in IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 6062-6071, Dec. 2015, doi: 10.1109/TIP.2015.2491020.

2024 Asia and Pacific Mathematical Contest in Modeling

Problem B

Optimization of the Shape of Air conditioner

With the continuous improvement of people's requirements for quality of life, products that combine air conditioning, humidifiers, and air purifiers have gradually been developed. It is a multifunctional environmental conditioning device. The three in one product only occupies the space of one device, instead of placing air conditioners, humidifiers, and air purifiers separately like traditional methods.



Figure 1. Schematic diagram of air conditioner, humidifier, and air purifier.

Air conditioning may make indoor air dry while regulating temperature. The humidifier in the three in one product can replenish moisture in a timely manner and maintain suitable indoor humidity. Air purifiers can filter harmful pollutants such as dust, pollen, smoke, and formaldehyde in the air. When the air conditioner is running and causing indoor air circulation, the air purifier can better function, purify the circulating air, and improve indoor air quality. Users can simultaneously control the opening, closing, and adjusting related parameters of three functions without the need to operate three different devices separately. The three in one product

of air conditioner, humidifier and air purifier also reduces the power and connection wires of multiple devices, reduces the complexity of wiring, makes the indoor environment more concise and beautiful, and also reduces the safety hazards that may be caused by too many lines.

The optimal design of the three in one air conditioner, humidifier, and air purifier is closely related to aerodynamics. The overall shape, placement, position and quantity of air inlet and outlet, direction and angle, wind speed and air volume of the three in one environmental conditioning equipment will all affect its effectiveness.

There is currently a room that is about to be renovated, which can be approximated as a long square with a spatial volume of 5m×8m×3m. The maximum volume of the limited three in one air conditioner is 0.1 m^3, the rated power consumption is 1800W, the maximum air outlet speed is 8.0 m/s, the maximum inlet flow rate of the airflow is 600 m^3/hour, and the maximum outlet flow rate of the airflow is also 600 m^3/hour.

As the manager of the APMCM three-in-one air conditioner appearance optimization team, please collect and research relevant data on air conditioners, humidifiers, and air purifiers in the market, and then establish a mathematical model for optimizing the appearance of the three in one air conditioner. Answer the following questions:

Question 1: Please analyze the influence of factors such as the placement of air conditioning, the position and quantity of air inlet and outlet, direction and angle, wind speed and air volume on the efficiency of air conditioning, and simulate the changes in indoor temperature over time and space under different conditions in summer and winter. Then consider the different shapes of the air conditioner, establish an optimization model for the shape of the air conditioner, and design the optimal shape and size of the air conditioner to achieve the best temperature regulation effect.

Question 2: Please analyze the influence of the shape of the air purifier on the purification effect, consider different shapes of the air purifier, establish an optimization model for the shape of the air purifier, design the optimal shape and size of the air purifier to maximize the purification effect, and draw the shape and size parameters of the optimal shape.

Question 3: Please analyze the influence of the shape of the air humidifier on the humidification effect, consider different shapes of the air humidifier, establish an optimization model for the shape of the air humidifier, design the optimal shape of the air humidifier to maximize the humidification effect, and draw the shape and size parameters of the optimal shape.

Question 4: Based on the models and results in questions 1-3, please establish an

optimized design model for the appearance of the three in one air conditioner, humidifier, and air purifier, while maximizing energy efficiency, human comfort, purification effect, and humidification effect. Design the optimal appearance of the three in one device and draw the shape and size parameters of the optimal appearance.

Glossary :

Aerodynamics: Aerodynamics is a branch of fluid mechanics that mainly studies the interaction forces between gases and objects when objects move in air or other gases, as well as the flow of gases around objects. Simply put, it focuses on the dynamic relationship between air and objects, including the generation of forces, energy conversion, and flow patterns, among many other aspects.

2024 Asia and Pacific Mathematical Contest in Modeling

Problem C

Development Analyses and Strategies for Pet Industry and related Industries

As people have developed their consumption philosophy, Pet industry, an emerging industry, has been gradually gathering momentum worldwide, thanks to rapid economic development and rising per capita income. In 1992, China Small Animal Protection Association was established, followed by the entry of international pet brands like Royal Canin and Mars into the Chinese market in 1993. The pet-related industries—such as pet food, pet clinic, pet supplies and pet care—have also gradually enjoyed a vast and fast-growing market, since "pet companionship" gains popularity in China. Based on the data in the attachments and additional data collected by your team, please analyze the development trends and market demands of pet industry. According to your analyses and the current economic environment, please make corresponding strategic recommendations for the development of Chinese pet industry.



Figure 1. Pet cats and dogs.

Question 1: Based on the data in Attachment 1 and the additional data collected by your team, please analyze the development of China’s pet industry in the past five years by pet type. And analyze the factors in the development of China's pet industry, so as to make an appropriate mathematical model to predict the development of China’s pet industry in the next three years.

Question 2: Overseas pet industries, such as in European countries and America, have also developed rapidly in recent years. Please analyze the development of the global pet industry by pet type, based on the data in Attachment 2 and the additional data collected by your team. And make an appropriate mathematical model to forecast the global demand for pet food in the next three years.

Question 3: In light of China’s pet food production and export values in Attachment 3, please analyze the development of China's pet food industry and predict its production and export of pet food in the next three years (regardless of changes in economic policies), based on the trend of global demand of pet food market and China's development.

Question 4: China’s pet food industry will inevitably be influenced by the new foreign economic policies (such as tariff policies) of European countries and America. To quantitatively analyze this influence, please make an appropriate mathematical model and take the data in the attachments, the additional data you collected, and the calculation results in above questions into consideration. Based on your calculations, please develop feasible strategies for the sustainable development of China’s pet food industry.

Attachment :

Attachment 1. 2019-2023 Number of Pet Cats and Dogs in China (in 10,000s)

Pets/Years	2023	2022	2021	2020	2019
Cat	6980	6536	5806	4862	4412
Dog	5175	5119	5429	5222	5503

Attachment 2. 2019-2023 Overseas Number of Pet Cats and Dogs (in 10,000s)

Countries	Pets/Years	2023	2022	2021	2020	2019
America	Cat	7380	7380	9420	6500	9420
	Dog	8010	8970	8970	8500	8970
France	Cat	1660	1490	1510	1490	1300
	Dog	990	760	750	775	740
Germany	Cat	1570	1520	1670	1570	1470
	Dog	1050	1060	1030	1070	1010

Attachment 3.2019-2023 China’s Pet Food Production and Export Values (in 100 millions)

Values/Years	2023	2022	2021	2020	2019
Total Value of China’s Pet Food Production	2793 （CNY）	1508 （CNY）	1554 （CNY）	727.3 （CNY）	440.7 （CNY）
Total Value of China’s Pet Food Exports	39.6 （USD）	24.7 （USD）	12.2 （USD）	9.8 （USD）	154.1 （CNY）

2024 Asia and Pacific Mathematical Contest in Modeling

Problem D

Exploring Frontiers in Quantum-Accelerated AI

Quantum computing holds tremendous potential for solving complex problems and processing large-scale datasets, far surpassing the capabilities of classical computers. When integrated with artificial intelligence (AI), quantum computing can bring revolutionary breakthroughs. Its parallel processing capabilities enable solving more intricate problems in shorter periods, which is of significant importance for optimizing and enhancing AI algorithms. This competition aims to assess participants' modeling and problem-solving skills in the field of quantum computing. Through challenging real-world scenarios, we explore the limitless possibilities that arise from the synergy of quantum computing and AI.

The competition consists of three parts, each involving a dataset and a task. Participants are required to formulate a corresponding QUBO (Quadratic Unconstrained Binary Optimization) model and solve it using the simulated annealing algorithm provided by the Kaiwu SDK. The QUBO model, suitable for CPQC (Coherent Photonic Quantum Computer), is expressed as:

$$\min x^{T}Qx, \quad x \in \{0,1\}^{n}$$

where Q is the coefficient matrix.

The competition focuses on AI-related scenarios, converting problems into QUBO form and solving them using the Kaiwu SDK, a specialized software development kit for solving QUBO models on CPQC. The SDK is accessible at this link (https://platform.qboson.com/).

Reference materials for QUBO modeling across different scenarios are provided in the attachments to help participants understand and apply these concepts.

Task 1: Resource Demand Prediction in Cloud Computing (20 points)

Background

Efficient operation of cloud computing platforms relies on precise resource scheduling, where demand prediction is a core component. By analyzing historical data to build predictive models, resource wastage can be minimized, and system efficiency and availability can be improved. Time series forecasting is commonly employed in such scenarios, but transforming these optimization problems into quantum computing-compatible forms remains a challenge.

Problem Description

You are tasked with developing a resource management system for a cloud computing platform to predict resource demand. The dataset is as follows:

Table 1. Computing resource demand from January to September.

Month	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sept
Demand	9000	9400	9594	9859	9958	10043	10309	10512	10588

These data represents monthly computing resource demand from January to September. You decide to use an Autoregressive (AR) model for forecasting, represented as:

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t ,$$

where c is a constant, p is the order of the AR model, and ε_t represents random noise with a mean of zero.

You are expected to:

1. Transform the above time series forecasting problem into a QUBO model, clearly defining the objective function and decision variables.

2. Use Kaiwu SDK's simulated annealing algorithm to solve the model and predict the demand for October.

Task 2: Classification Using Support Vector Machines (40 points)

Background

In machine learning, in addition to regression tasks like time series prediction problem

above, classification is another classical task. Classification aims to assign input samples to predefined categories based on their features, which is widely applicable in real-world scenarios, such as spam detection and image classification. Support Vector Machines (SVMs) are a popular supervised learning technique based on maximizing margins. SVM also excel in nonlinear classification through kernel tricks.

Integrating quantum computing introduces new perspectives to classification tasks. By converting the SVM optimization problem into QUBO form, quantum computing can accelerate the solution process.

Problem Description

The Iris dataset (https://archive.ics.uci.edu/dataset/53/iris) is a classical machine learning dataset containing 150 samples, each with four features and a class label (one of three categories). Your task is to classify this dataset using an SVM model and the Kaiwu SDK.

You are expected to:

1. Transform the optimization problem of training an SVM-based classification model into a QUBO model, clearly defining the objective function and decision variables.

2. Solve the QUBO problem using the simulated annealing algorithm in the Kaiwu SDK.

Task 3: Exploring the Integration of Quantum Computing and Deep Learning (40 points)

Background

Deep learning models, such as Convolutional Neural Networks (CNNs), demonstrate outstanding performance in complex tasks like image classification and recommendation systems. However, optimizing these models often requires significant computational resources. By leveraging quantum computing's optimization capabilities, more efficient training and inference methods can be developed for deep learning.

You are expected to:

Select a specific application scenario, such as image classification or recommendation systems, for which you need to design a suitable deep learning model and structure. Then, transform the related optimization problem (e.g., model training) into a QUBO model. Solve it using Kaiwu SDK's simulated annealing algorithm.

Submission Requirements:

1. Complete code files, including data preprocessing, model construction, QUBO model conversion, and solution processes.

2. A detailed result report containing:

a. The problem scenario and its background.

b. Description of the artificial intelligence model structure.

c. Specific QUBO model formulation.

d. Solution process and solving time of the QUBO model, together with results analysis.

f. Model performance evaluation and interpretation of results.

Please provide sources and explanations for any external datasets or reference materials used.

Note:

1. Kaiwu SDK is restricted to solve problem under 600 bits.

2. Only the latest version of Kaiwu SDK is supported.

3. For the Problem 1, participants may choose to submit their matrices to CPQC (https://platform.qboson.com/), which is **optional**. Each participant will receive a total of 5 quotas during the competition. No additional quotas will be granted once they are used up. Additionally, there may be delays in receiving results from CPQC.

4. If you have any questions about the tasks, please scan the QR code below to contact us.



5. You can view the Kaiwu SDK installation and usage guide via this link (https://b23.tv/IqKoPnv).

2023 Asia and Pacific Mathematical Contest in Modeling

Problem A

Image Recognition for Fruit-Picking Robots

China is the world's largest producer of apples, with an annual output of approximately 35 million tons. At the same time, China is also the world's largest exporter of apples, with one out of every two apples in the world and more than one-sixth of the apples exported worldwide from China. China proposed the Belt and Road Initiative (BRI), a key pillar of building a global community with a shared future. Thanks to this initiative, Vietnam, Bangladesh, the Philippines, Indonesia, and other countries along the route have become the main export destinations for Chinese apples.

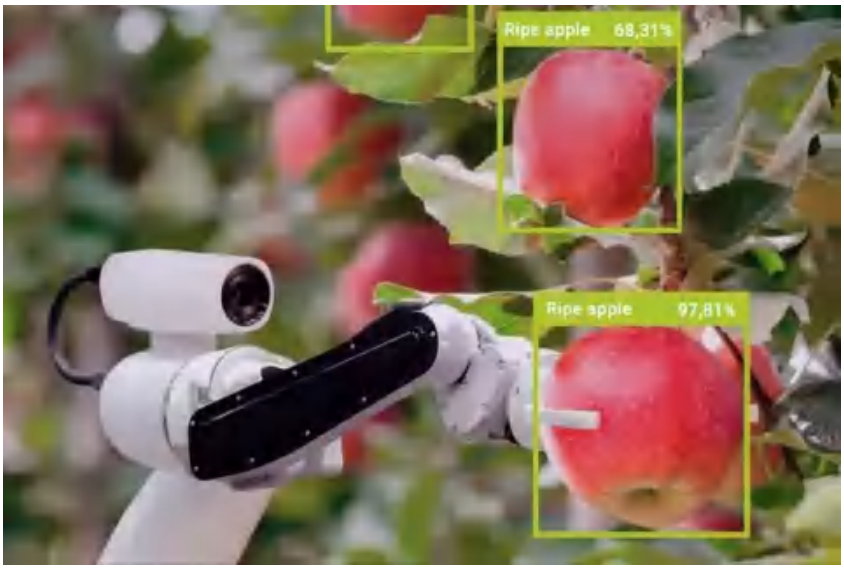


Figure 1. Diagram of image recognition of apples by fruit-picking robots.

Apple picking relies mainly on hand harvesting. When apples are ripe, a large number of picking workers are needed in apple-producing areas in a few days. But most of the local farmers plant apples in their own orchards. In addition, the aging of agricultural workers and the phenomenon of young people leaving their villages for work have led to a labor shortage during the apple-picking season. To solve this problem, China has been researching robots that can pick apples since around 2011 and has made significant progress.

However, the popularization and application of various apple-picking robots worldwide fall short of the ideal because orchard environments differ from the controlled experimental

settings. In complex and unstructured orchard environments, most existing robots are unable to accurately identify obstacles such as “leaf occlusion”, “branch occlusion” “fruit occlusion”, and “mixed occlusion” etc. If apples are picked directly without making precise judgments based on actual scenarios, there is a high risk of damaging the fruit, and even causing harm to the picking hands and mechanical arms. This adversely affects the efficiency of harvesting and the quality of the fruit, leading to greater losses. In addition, the recognition and classification of different harvested fruits is also significantly important, such as the procedures of classification, processing, packaging, and transportation. However, many fruits have quite similar colors, shapes, and sizes to apples, which poses great difficulties for the post-harvest identification of apples.

This competition aims to establish an apple image recognition model with a high recognition rate, fast speed, and accuracy by analyzing and extracting features from labeled fruit images, and to perform data analysis on the images, such as automatically calculating the number, positions, maturity levels, and estimating the masses of apples in the images. The specific tasks are as follows:

- **Question 1: Counting apples**

Based on the image dataset of harvest-ready apples provided in *Attachment 1*, extract image features, establish a mathematical model, count the **number of apples** in each image, and draw a **histogram** of the distribution of all apples in *Attachment 1*.

- **Question 2: Estimating the positions of apples**

Based on the image dataset of harvest-ready apples provided in *Attachment 1*, identify the **position of the apples** in each image with the left bottom corner of the image as the coordinate origin, and draw a two-dimensional **scatter diagram** of the geometric coordinates of all apples in *Attachment 1*.

- **Question 3: Estimating the maturity state of apples**

Based on the image dataset of harvest-ready apples provided in *Attachment 1*, establish a mathematical model, calculate the **maturity of apples** in each image, and draw a **histogram** of the maturity distribution of all apples in *Attachment 1*.

- **Question 4: Estimating the masses of apples**

Based on the image dataset of harvest-ready apples provided in *Attachment 1*, calculate the two-dimensional **area of the apples** in each image with the bottom left corner of the image as

the coordinate origin, estimate the **masses of the apples**, and draw a **histogram** of the mass distribution of all apples in *Attachment 1*.

- **Question 5: The recognition of apples**

Based on the dataset of harvested fruits' images provided in *Attachment 2*, extract image features, train an **apple recognition model**, identify the apples in *Attachment 3*, and draw a distribution histogram of the ID numbers of all apple images in *Attachment 3*.

Attachment:

Attachment.zip, download on the website: https://share.weiyun.com/T6FKbjLf

Attachment 1:

The folder contains 200 images of harvest-ready apples, and each image has a size of 270 * 180 pixels. Partial screenshots of *Attachment 1* are shown below:



Attachment 2:

The folder contains 20705 images of different harvested fruits with **known** labels and classifications, and each image has a size of 270 * 180 pixels. Partial screenshots of *Attachment 2* are shown as follows:

The dataset of apples:



The dataset of carambolas:



The dataset of pears:



The dataset of plums:

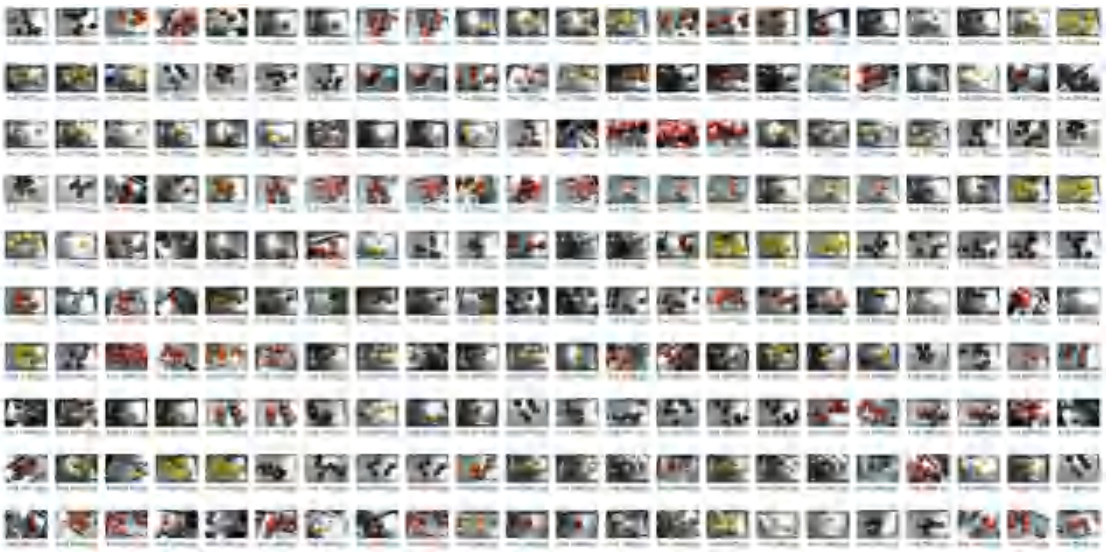


The dataset of tomatoes:



Attachment 3:

The folder contains 20705 images of different harvested fruits with unknown labels and classifications, and each image has a size of 270 * 180 pixels. Partial screenshots of *Attachment 3* are shown as following:



2023 Asia and Pacific Mathematical Contest in Modeling

Problem B

Microclimate Regulation in Glass Greenhouses

The yield of greenhouse crops is affected by various climate factors, including temperature, humidity, and wind speed [1]. Among them, suitable temperature and wind speed are crucial for plant growth [2]. In order to regulate climate factors such as temperature and wind speed inside the glass greenhouse, ventilation systems with greenhouse fans are commonly used in the design of greenhouse, as shown in Figure 1. The position of the greenhouse fan and the speed of the warm air outlet affect the distribution and uniformity of the velocity field and the temperature field in the greenhouse. Therefore, how to optimize the greenhouse fan to obtain suitable wind speed and temperature, and improve their uniformity, is an important issue that needs to be solved in current glass greenhouse design.



Figure 1. Glass greenhouse

The glass greenhouse is sealed and placed indoors, without considering external factors such as greenhouse doors, drafts, solar radiation, and other environmental factors. The current design of the glass greenhouse has dimensions of 10m×3m×2m (length×width×height) and the size of the greenhouse fan is 0.5m×0.5m, positioned on the left side of the greenhouse. The center of the greenhouse fan is located 1.3m above the ground, as illustrated in Figure 2. The boundary conditions on the side of the greenhouse fan are set as velocity inlet conditions, blowing warm air at 40℃in the horizontal direction with an average velocity of 2 m/s. The outer glass and bottom soil of the greenhouse are set as wall conditions, primarily exchanging

energy with the entire greenhouse through convective heat transfer and conduction [3]. The initial temperature is set at 20℃. When crops are planted inside the greenhouse, it is necessary to consider the canopy resistance of the crops. The crop model can be simplified as a porous medium with dimensions of 8m×2m×0.5m (length×width ×height) [4], placed in the center of the greenhouse. The suitable wind speed for crop growth inside the greenhouse is 0.3-1m/s, and the suitable temperature is 23-26℃.

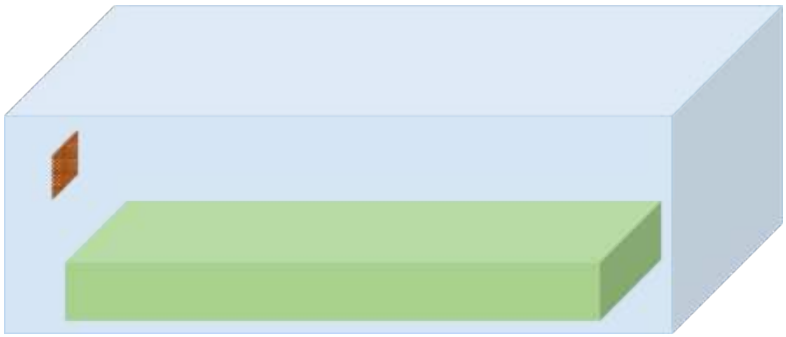


Figure 2. Schematic diagram of glass greenhouse structure

Question 1: Please establish a mathematical model for the temperature and wind speed distribution inside a glass greenhouse without crops. Display the distribution of wind speed and temperature at a cross-section of the greenhouse at a height of 0.5 meters.

Question 2: Please establish a mathematical model for the temperature and wind speed distribution inside a glass greenhouse with planted crops. Present the distribution of wind speed and temperature at two cross-sections within the greenhouse: one at a height of 0.5 meters (crop canopy level) and another at a height of 0.1 meters (inside the crop canopy). Analyze whether the conditions are suitable for crop growth.

Question 3: Please provide the temperature and wind speed distribution inside the glass greenhouse for the following two scenarios and compare them with solution presented in the second question. In Scenario One, increase the velocity of warm air outlet from 2 m/s to 3 m/s. In Scenario Two, lower the position of the greenhouse fan by moving it from 1.3 m to 1 m.

Question 4: Can your team further optimize the greenhouse fan design of glass greenhouse from the number of greenhouse fan, location, wind speed, blowing temperature, specifications and different crops and other factors.

References：

[1] Singh M C, Singh J P, Pandey S K, et al. Factors affecting the performance of greenhouse cucumber cultivation-a review[J]. International Journal of Current Microbiology and Applied Sciences, 2017, 6(10): 2304-2323.

[2] Liu Y, Li D, Wan S, et al. A long short-term memory-based model for greenhouse climate prediction[J]. International Journal of Intelligent Systems, 2022, 37(1): 135-151.

[3] Norton T, Sun D W, Grant J, et al. Applications of computational fluid dynamics (CFD) in the modelling and design of ventilation systems in the agricultural industry: A review[J]. Bioresource technology, 2007, 98(12): 2386-2414.

[4] Fatnassi H, Boulard T, Poncet C, et al. Optimisation of greenhouse insect screening with computational fluid dynamics[J]. Biosystems Engineering, 2006, 93(3): 301-312.

2023 Asia and Pacific Mathematical Contest in Modeling

Problem C

The Development Trend of New Energy Electric Vehicles in China

New energy vehicles refer to vehicles with advanced technical principles, new technologies and new structures, which use unconventional vehicle fuel as their power source (unconventional automotive fuels refer to fuels other than gasoline and diesel) and integrate advanced technologies in vehicle power control and drive. New energy vehicles include four major types: hybrid electric vehicles, pure electric vehicles, fuel cell electric vehicles, and other new energy vehicles. As a type of new energy vehicle, new energy electric vehicles have achieved rapid development in recent years due to their properties of low pollution, low energy consumption, and ability to regulate peak electricity consumption. New energy electric vehicles, including electric buses and family electric cars with less than 7 seats, have been popular with consumers and governments around the world.

Since 2011, the Chinese government has actively promoted the development of new energy electric vehicles and formulated a series of preferential policies. The new energy electric vehicle industry has achieved tremendous development, gradually becoming another Chinese symbol after the "China High Speed Railway". Now your team is invited to complete the following questions：

Question 1: Analyze the main factors that affect the development of new energy electric vehicles in China, establish a mathematical model, and describe the impact of these factors on the development of new energy electric vehicles in China.

Question 2: Collect industry development data on China's new energy electric vehicles, establish a mathematical model to describe and predict the development of China's new energy electric vehicles in the next 10 years.

Question 3: Collect data and establish a mathematical model to analyze the impact of new energy electric vehicles on the global traditional energy vehicle industry.

Question 4: Some countries have formulated a series of policies targeted to resist the development of new energy electric vehicles in China. Establish a mathematical model to analyze the effects of these policies on the development of new energy electric vehicles in China.

Question 5: Analyze the impact of the electrification of new energy electric vehicles (including electric buses) in cities on the ecological environment. Assuming that there is an urban population of 1 million, provide the calculation results of the model.

Question 6: Based on the conclusion of question 5, write an open letter to the citizens to publicize the benefits of new energy electric vehicles and the contributions of the electric vehicle industry in various countries around the world.

优 秀 论 文

亚太地区大学生数学建模竞赛组委会制作

Team Number :	apmcm25302425
Problem Chosen :	A

2025 APMCMsummarysheet

A Multi-Objective Motion Planning and Energy Optimization Framework for Humanoid Robots in Highly Dynamic Scenarios

summary

The consumer-levelservicescenarios of robotics technology are rapidly expanding, and the motion performance, operationalstability, andsystem energy efficiency of humanoid robots in highly dynamic environments have become **core challenges** constraining their widespread deployment. Through establishing aseries of **mathematical models**, the core problems of motion planning,stability control, and energy optimization are addressed.

Problem 1: Thisstudy established a **kinematic model based on Denavit-Hartenberg parameters**. Through forward kinematics calculation, the final coordinates of the arm end point were determined as **(146.4, 84.5, 292.7)** mm. Analysis demonstrates that the **motion trajectory issmooth and continuous, and all motor parameters are withinsafe operating ranges, meeting the core requirements of robotstability and reliability.**

Problem 2: Asingle-leg joint kinematic model based on quintic polynomial trajectories was constructed for the robot'sstraight-line non-uniform velocity motion task. The motion process adopts a three-stage velocity curve with total time **T=5.0s**. **The maximum knee joint angular velocity (0.4924 rad/s)** occurs at t=0.611s and t=4.389s, which is far belowsystemsafety limits, ensurings**mooth motor operation.**

Problem 3: A multi-model fusion framework of "analytical motion planning + coordinated optimization control" was designed for the Unitree G1 humanoid robot's complex combined motion. Dual-arm joint trajectories were generated through inverse kinematics, and **a waist-lower limb coordination optimizer** was constructed tostabilize the center of mass projection within thesupport polygon, achieving modeling and verification of 29 joint angle functions within the **0–4s time interval.**

Problem 4: Thisstudy established an **energy consumption calculation model** and constrained optimization model based on motor dynamics. By defining timescale factor and amplitudescale factors as optimization variables, **thesequential Quadratic Programming (SQP) algorithm** was adopted forsolving. Optimization resultsshow that **all three problems achieved a 33.3%** energy reduction with optimal timescale factors of 1.500. Comprehensive analysis visualization charts were generated, including energy comparison, joint energy breakdown, power-time curves, and **optimization parameters,** verifying the model's effectiveness and robustness acrosssystems of different complexity (from 2 joints to 21 joints), confirming that this optimization method**significantly reduces energy consumption without changing motion patterns.**

Key words: Denavit-Hartenberg Model、Non-Uniform Velocity Model 、Multi-joint Cooperative Control 、 3D Trajectory Visualization、Constrained Optimization 、sequential Quadratic Programming

Contents

I. Introduction **1**
1.1 Background 1
1.2 Problem Restatement 1
1.3 Our works 2
II. The Description of the Proble **2**
2.1 Analysis ofspecific Issues 2
2.1.1 Analysis of Problem 1 2
2.1.2 Analysis of Problem 2 2
2.1.3 Analysis of Problem 3 2
2.1.4 Analysis of Problem 4 2
III. Basic assumption **3**
IV. Symbol **3**
4.1 Symbols 3
V. Models **4**
5.1 Analysis andsolving of Question One 4
5.1.1 Model Preparation 4
5.1.2 Prediction Model for Endpoint Coordinates Based on the DH Model 4
5.1.3 Motorsafety Validation Model 5
5.1.4 Analysis of the Result 6
5.2 Analysis andsolving of Question Two 8
5.2.1 Model Preparation 8
5.2.2 Model of the Body's Non-Uniform Velocity Curve 8
5.2.3 Model of Kinematic 8
5.2.4 Analysis of the Result 9
5.3 Analysis andsolving of Question Three 11
5.3.1 Model Preparation 11
5.3.2 Establishment of Body Rotation and Dual-Arm Circular Motion Model 12
5.3.3 Analysis of the Result 13
5.4 Analysis andsolving of Question Four 17
5.4.1 Model Preparation 17
5.4.2 The establishment of the energy optimization model 19
5.4.3 Results 20
5.4.4 Analysis of the Result 21
VII. Evaluation and Promotion of Model **24**
6.1 Strength And Weakness 24
6.1.1 Strength 24
6.1.2Weakness 24
VIII. References **25**
Report on Use of AI Tools **26**
Appendix **28**
Appendix One 28

I. Introduction

1.1 Background

With the advancement of advanced technologies, robotics has become a frontier field. As robotic applications expand fromstructured industrialsettings to open, non-structured consumerservicescenarios, humanoid robots face core challenges in maintaining excellent motion performance,stability, and energy efficiency in highly dynamic environments. The Unitree G1 humanoid robot, developed by Hangzhou Unitree Robotics, integrates a multi-degree-of-freedom motionsystem with high-precision visual perception modules, providing an ideal platform forstudying comprehensive robot performance in complex tasks.

However, during complex task execution, robotsstill confrontsignificant technical bottlenecks in the coordinated optimization of motorsafety control,stability management, and multi-joint coordinated planning. For example,sudden motor torque changes or load imbalances can easily lead to loss of control; improper weight distribution or trajectory planning can disrupt dynamic balance, causing falls; while highly complex motionsequences createsignificant energy consumption pressures, directly limiting the robot's operational duration and reliability.

Therefore, thisstudy aims to establish precise mathematical models tosystematically optimize the robot's joint motion, trajectory planning, multi-body coordination, and energy consumption management, thereby enhancing thesafe and controllable operation and energy efficiency maximization of humanoid robots.

1.2 Problem Restatement

Based on the interpretation and analysis of robot performance optimization, the following problems need to be addressed:

1.For thespecified motion of the robot's left arm, it is necessary to establish its kinematic model to calculate thespatial coordinates of the arm endpoint and verify whether this motion remains within thesafety operating limits of the joint motors.

2.For the robot's non-uniform linear movement task, it is necessary to develop a kinematic model ofsingle-leg joint angles over time based on gait planning, calculate the time instant when the knee joint's angular velocity reaches its maximum, and determine the total time required to complete the movement.

3.For the complex coordinated motion involving body rotation and circular arm movements, it is necessary to establish a multi-joint collaborative motion model that integrates upper-limb trajectory planning with lower-limb balance compensation, describing the angular variations of each joint throughout the complete motion cycle.

4.Considering all the aforementioned performance actions collectively, it is necessary to construct an energy consumption assessment model based on joint torque and velocity to calculate the total energy consumption,subsequently propose an action optimization model targeting energy efficiency, and compare the energy consumption results before and after optimization.

1.3 Our works

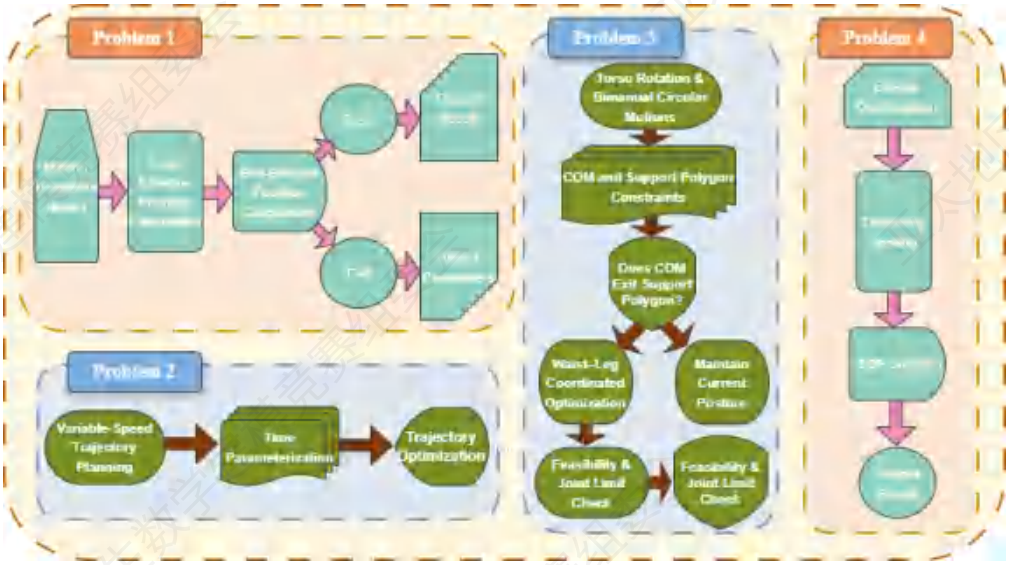


Figure 1 Our Work

II. The Description of the Proble

2.1 Analysis ofspecific Issues

2.1.1 Analysis of Problem 1

Solve the forward kinematics for YuShu humanoid's left arm:starting from the joint zero position, perform two motions—60° forward extension and 30° left rotation—with theshoulder as the origin. Using the DH model based on actual jointstructure, link parameters, and motorsafety limits, compute the final Cartesian coordinates of the arm endpoint and verify motorsafety.

2.1.2 Analysis of Problem 2

Develop a mathematical model for robot walking from (0,0,0) to (10,0,0) with non-uniform velocity. Model joint angles (hip, knee, ankle) over time. Find when knee angle changes most rapidly. Calculate total travel time.

2.1.3 Analysis of Problem 3

Problem Three: Unitree robot performs complex motion (0-4s): 45° torso rotation, dual arms drawing opposite circles, legs balancing. Challenges: multi-joint coupling, center of mass disturbance, waist-hip optimization. Model ensures precise timing, dynamic consistency, constraint feasibility, and 'rotate body first, then trigger legs' verification.

2.1.4 Analysis of Problem 4

Problem 4: Optimize energy via trajectoryscaling (time/amplitude) using Problems 1-3 data. Challenges: 1) Accurate energy model linking joint kinematics (angle, vel, acc) to motor energy (torque, power). 2) Definescale factors, constrained optimization for feasible/safe motion. 3)solve withsQP for optimal parameters. 4) Visualize results (energy breakdown, power-time curves) for validation. Requires full pipeline (preprocessing, modeling, optimization, analysis) robust across 3-21 joints.

III. Basic assumption

Assumptions are core to robot optimization modeling, transforming complex engineering problems intosolvable mathematical models.

- **Linkstructure Assumption:** The left arm issimplified to 2 rigid links (upper arm + forearm), ignoring flexiblestructuressuch as arm muscles andskin, retaining only geometric dimensions with upper arm length of 170 mm and forearm length of 168 mm.
- **Independence Assumption:** Theshoulder joint is responsible for forward lifting of 60° and left rotation of 30°, while the elbow joint is only responsible for forearm flexion and extension. Joint motions have no coupling interference.
- **Joint Trajectorysmoothness Assumption:** Polynomial interpolation is used to generate joint angle-time curves, ensuring continuity of angle, velocity, and acceleration to avoid motorstart-stop impacts.
- **Motor Model Assumptions**: Motor torque is proportional to joint angular acceleration. Motor power consists of dynamic power and base power. Moment of inertia and base power remain constant during motion.
- **Motion Assumptions**: Joint angle trajectories can bescaled in both time and amplitude dimensions. Optimized trajectories maintain thesame motion pattern as original trajectories.
- **Optimization Assumptions**: Timescale factor and amplitudescale factors have value ranges of [0.5, 1.5]. The optimization algorithm can converge to the global optimalsolution or asolution close to the global optimum.

IV. Symbol

4.1 Symbols

Here we define the main parameters, and theirspecific values will be given later.

Symbols	*Definition*	*Units*
q	Vector of coordinated optimization variables	4D vector (rad)
$x_{COM}(t), y_{COM}(t)$	Center of mass projection	Millimeters (mm)
J	Cost function for coordinated optimization	Dimensionless
E_{total}	Total energy consumption	J
P_{base}	Base power	W
η	Energy reduction rate	%

V. Models

5.1 Analysis andsolving of Question One

5.1.1 Model Preparation

To calculate the endpoint coordinates of the humanoid robot's left arm in Problem 1, the Denavit– Hartenberg (DH) model is employed tosystematically represent multi-joint coordinate transformations. Combined with the Yushu G1sDK Guide, this approach transforms joint motions into matrix operations for accurate coordinate computation and validates motionsafety using hardware parameters, thereby laying the groundwork for trajectory planning and multi-joint control insubsequent problems.

(1) Constraints Analysis

This analysis defines the necessary constraints and motion parameters for the precise calculation of the endpoint coordinates for the Yushu humanoid robot's left arm. It begins with the robot's initial posture (all joints at zero degrees), following the problem requirements and establishing the physical andsafety guidelines of the Yushu G1 platform.

Table 1 Motion Analysis

Parameter Type	*Specific Value*
Arm Length	338 mm (rigid body length, elbow extended)
Joint Rotation Angle	Pitch θ_1=60° (1.047 rad), Roll θ_2=30° (0.523 rad)
Jointsafety Limits	Pitch: ±177°/153°, Roll: ±91°/129°
Coordinatesystem	Origin (shoulder joint), x-axis forward, y-axis left, z-axis up

5.1.2 Prediction Model for Endpoint Coordinates Based on the DH Model

Based on the left arm'sstructure and the motion requirements of the problem, the arm is divided into 2 links (upper arm and forearm) with lengths of 170mm and 168mm, respectively, and 2 key joints (shoulder joint and elbow joint).

The model is astandard method widely used for robot kinematic modeling. Itsystematically describes the relative position and orientation between adjacent links in a multi-joint chain using four parameters. For theserialstructure of the G1's left arm (shoulder joint → upper arm → elbow joint → forearm → left-hand endpoint), the relative pose and position between adjacent links are defined by four parameters. The core of this method lies in the homogeneous transformation matrix, which cansimultaneously represent rotation and translation. The formula is as follows:

$$T_i^{i-1}=\begin{bmatrix}\cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i\\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i\\ 0 & \sin\alpha_i & \cos\alpha_i & d_i\\ 0 & 0 & 0 & 1\end{bmatrix} \quad (1)$$

T_i^{i-1} :transformation matrix of link i relative to link i-1;

θ_i :Joint Angle Rotation about the z_{i-1} axis, determining the orientation of link i;

a_i :Link length the distance between the z_{i-1} and z_i axes along the x_i axis;

α_i :the rotation around the x_i axis, determining the z_i axis direction relative to z_{i-1};

d_i :the distance along the z_{i-1} axis. Zero if no translation exists;

Coordinates in the Base Coordinatesystem:

A robot has no joints, the overall transformation matrix of the end link relative to the base frame (shoulder joint) is given by:

$$T_n^0=T_1^0\times T_2^1\times\ldots T_n^{n-1} \quad (2)$$

The endpoint is expressed in homogeneous coordinates within the end-link coordinate frame:

$$P_n=[0,0,0,1]^T, \quad (3)$$

5.1.3 Motorsafety Validation Model

To ensure thesafe andstable operation of the Yushu humanoid robot during performances, we have established quantifiedsafetystandards based on official documentation. Thesestandards focus on two core areas: motorstatus monitoring and overallstability assessment.

Step1: Joint Anglesafety Verification

$$|\theta_i|=\theta_{\max,i}, \quad (4)$$

θ_i :actual joint angle, $\theta_{\max,i}$:Maximum allowable joint angle of the motor (Flexion/Extension: $+90^\circ$,Abduction/Adduction: $\pm45^\circ$)

Step2: Angular Velocitysafety Verification

Since the motion in Problem 1 has no time constraints, assuming uniform motion over duration t, the angular velocity is:

$$\omega_i=\frac{\theta_i}{t}, \quad (5)$$

Thesafety condition is: $\omega_i\le\omega_{\max}=180^\circ(5.236rad/s)$.

Step3: Torquesafety Verification

Motor torque is related to joint load and acceleration.simplified model

(acceleration is 0 during uniform motion):

$$\tau_i = J_i \alpha_i + F_i \times L_i, \tag{6}$$

J_i : Moment of inertia, F_i : Gravitational load.

5.1.4 Analysis of the Result

(1) Analysis of End-Effector Coordinates

Step1: End-Effector Coordinate Accuracy Verification

Trajectory Continuity Check:in the "DH Model Frame Overlay Visualization", trajectories from early, mid, and current frames form asmooth, fan-shaped coverage without breaks or jumps. The "End-Effector Trajectory (X-Z Plane)"shows a continuous andsmooth curve fromstart to end point, confirming the high continuity of the DH model in representing the robotic arm motion.

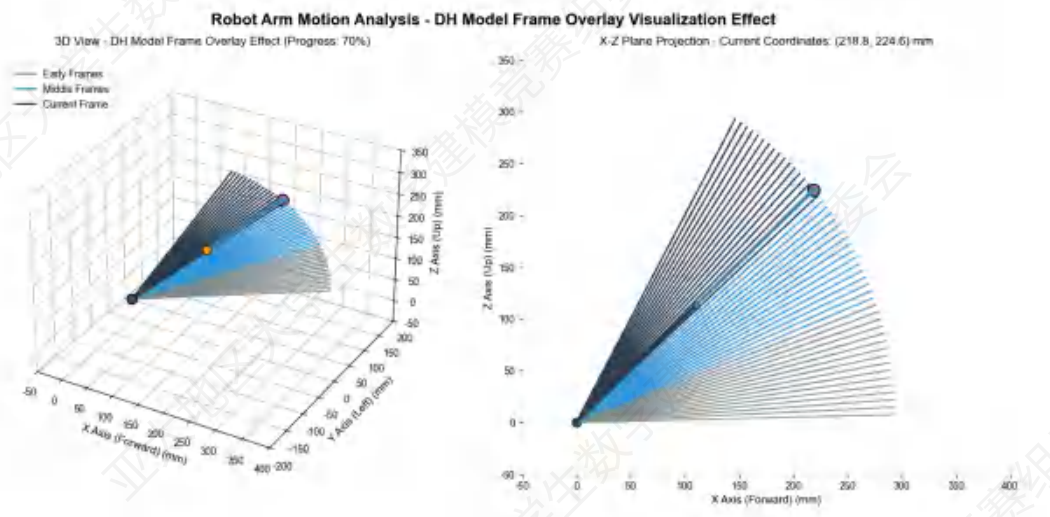


Figure 2 Manipulator Motion Analysis

Step2: Command-Coordinate Consistency Check

Validation of Coordinate Reasonableness Based on Motion Design Parameters and Visualization Results:

Body Angle Verification: The calculated final position of the end point is (146.4, 84.5, 292.7) mm $\sqrt{146.4^2 + 84.5^2 + 292.7^2} \approx 338mm$. Its distance from theshoulder joint is 338 mm, exactly matching the arm length. Its angle relative to the body (z-axis) also meets the movement requirement.

Rotation Check: In the x-y plane, the end point's angle is calculated as $\Phi = \arctan(\frac{y}{x}) = 30^\circ$. This perfectly matches the "rotate left 30°" command, confirming correct horizontal positioning.

Visual Check: The chartsshow the end point's projected position in the X-Z plane is (146.4, 292.7)mm, matching our calculation. The arm's movement path is also clear andsmooth, confirming the result is reliable.

(2) Motorsafety Inspection Results Analysis

Step1: Comprehensive Evaluation of Inspection Items

Based on thesafety thresholds provided in the Yushu Humanoid Robot Developer Documentation and the actual motion parameters, thesafety inspection results aresummarized in the table below. All items meet thesafety requirements, indicatingstable and reliable motor operation:

Table 2 Motorsafety Inspection Results

Inspection Item	*Actual Value*	*Safety Threshold*	*Result*
Shoulder Flexion Angle	60°	±90°	Safe
Shoulder Abduction Angle	30°	±45°	Safe
Joint Angular Velocity (Shoulder Flexion)	60°/s	≤180°/s	Safe
Joint Angular Velocity (Shoulder Abduction)	30°/s	≤180°/s	Safe
Motor Torque	≈6.61 N·m	≤15 N·m	Safe

Step2:safety Margin And Risk Prevention

All tested parameters maintainsufficientsafety margins: Joint angles: 30°(shoulder flexion), 15°(shoulder abduction)Angular velocity: >120°/s, Torque: ≈6.61 NM, These adequate margins help mitigate risks from mechanical errors, environmental interference, or control inaccuracies. The "Joint Angle Variation" chartshowssmooth transitions between positions with no abrupt changes or oscillations, ensuringstable andsafe motor operation.

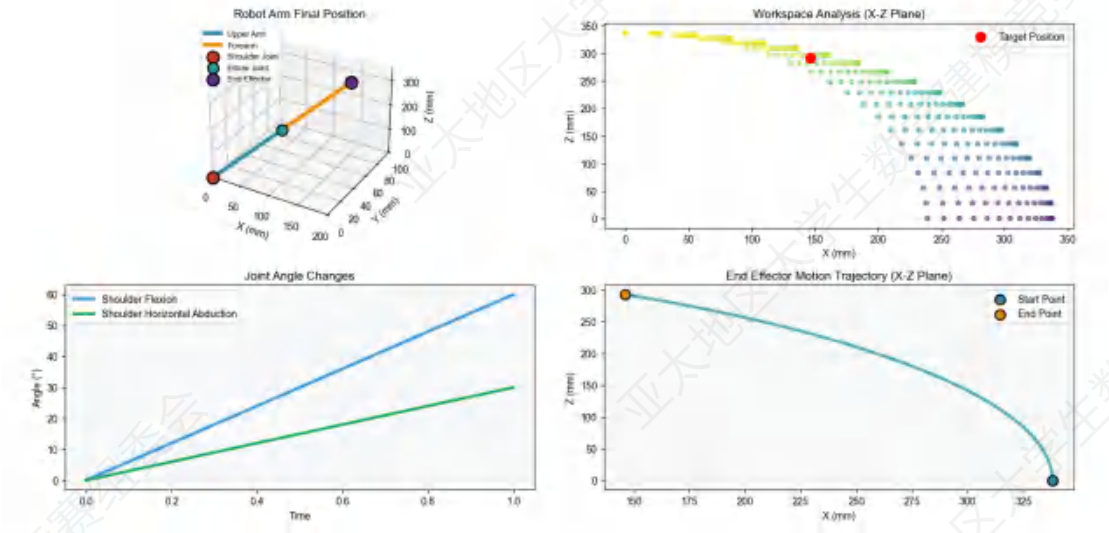


Figure 3 Left Arm End-Effector Motion Analysis

The calculated coordinates of the robot's left arm end-effector accurately match the motion design requirements, with numerical results highly consistent

with visualizations, validating thescientific rigor of the DH model and the precision of the computational process. In terms ofsafety, all core metrics—joint angles, angular velocity, and motor torque—arestrictly withinsafe thresholds and maintainsufficientsafety margins, fully complying with thesafetyspecifications for the Yushu Humanoid Robot's motor operation. Overall, this motion achieves the intendedspatial posture while ensuring operationalstability andsafety, laying a solid foundation for executing more complex movements in the future.

5.2 Analysis andsolving of Question Two

5.2.1 Model Preparation

(1) Core Theoretical Foundations

Forward Kinematics: Calculates the foot-end position from joint angles to verify if the trajectory follows astraight line along the x-axis (y=0, z=0).

Polynomial Interpolation: Ensures continuity in joint angle, angular velocity, and angular acceleration to prevent motorshock.

Gait Cycle Phasing: The typical gait for a bipedal robot alternates between "Support Phase" and "Swing Phase." The Yushu Humanoid Robot'sstep frequency is controlled within 1–2 Hz .

(2) Reasonable Assumptions

Motion Pattern: Joint angles are assumed to change over time as continuous and differentiable functions, avoidingsudden changes in velocity or acceleration.

Velocity Profile: A "trapezoidal acceleration" (non-uniform velocity) pattern is adopted. This aligns with the natural walkingsequence of "startup – constant velocity – deceleration" and facilitates the calculation of averagespeed.

5.2.2 Model of the Body's Non-Uniform Velocity Curve

Step1: Total Time Calculation

Based on the definition of average velocity, with a total displacement ofs=10m and an average velocity of $v_{avg} = 2m/s$:

$$T = \frac{S}{v_{avg}} = 5\text{s}, \tag{7}$$

step2: Trapezoidal Deceleration Model

To avoidsuddenshocks duringstart/stop, the velocity profile is divided into three phases under the "total displacement = 10 m" constraint:

Acceleration Phase: $t \in [0, t_1]$,Velocity increases linearly from 0 to v_{max} ;

Constant Velocity Phase $t \in [t_1, t_2]$,Velocity remains at v_{max} ;

Deceleration Phase $t \in [t_2, 5]$,Velocity decreases linearly from v_{max} to 0.

Parameters aresolved by displacementsuperposition:

$$S = \frac{1}{2}a_1 t_1^2 + v_{max}(t_2 - t_1) + v_{max}(T - t_2) + \frac{1}{2}a_2(T - t_2)^2 \tag{8}$$

5.2.3 Model of Kinematic

The hip jointserves as the local coordinate origin, maintaining constant height while

moving with the robot. The knee angle $\theta_{kneet}(t)$ is geometrically correlated with the displacement $x(t)$.

Step1: Geometric Constraints

Given thigh length l_1=0.3 m andshank length l_2=0.3 m:

During thesupport phase, the foot (end-effector) maintains ground contact (z=0), with its x-coordinate aligned to the contact point.

The hip pitch angle θ_{hip} adjusts according tostride length during forward motion, while the knee angle θ_{kneet} flexes tosupportstepping.

$$l_1 \cos\theta_{hip}(t) + l_2 \cos(\theta_{hip}(t) - \theta_{kneet}(t)) = (\mathrm{h}_0) \tag{9}$$

$$l_1 \sin\theta_{hip}(t) + l_2 \sin(\theta_{hip}(t) - \theta_{kneet}(t)) = x(t) - x_{hip}(\mathrm{t}) \tag{10}$$

Where $\theta_{hip}(t)$ denotes the hip pitch angle and $\theta_{kneet}(t)$ represents the knee joint angle.

Step2:Knee Joint Trajectory Planning

A quintic polynomial describes the knee angle trajectory during theswing phase, ensuring continuity in position, velocity, and acceleration under the following boundary conditions:

$$\theta_{kneet}(t) = a_5 t^5 + a_4 t^4 + a_3 t^3 + a_1 t + a_0 \tag{11}$$

Step3:Model Optimization

Motor dynamics constraints—based on the developer documentation—limit the maximum joint angular velocity and acceleration:

$$|\dot{\theta}_i| \le \omega_{\max}, |\ddot{\theta}_i| \le \alpha_{\max} \tag{12}$$

5.2.4 Analysis of the Result

Step1:Phase Division Logic

A three-phase "acceleration–constant velocity–deceleration" profile is imple mented with 1.2s acceleration, 2.6s constant velocity, and 1.2s deceleration. Thi ssymmetric design reduces motionshock and improvesstability. With a total trav el of 10 m and average velocity of 2 m/s, the total time is validated as 5s.

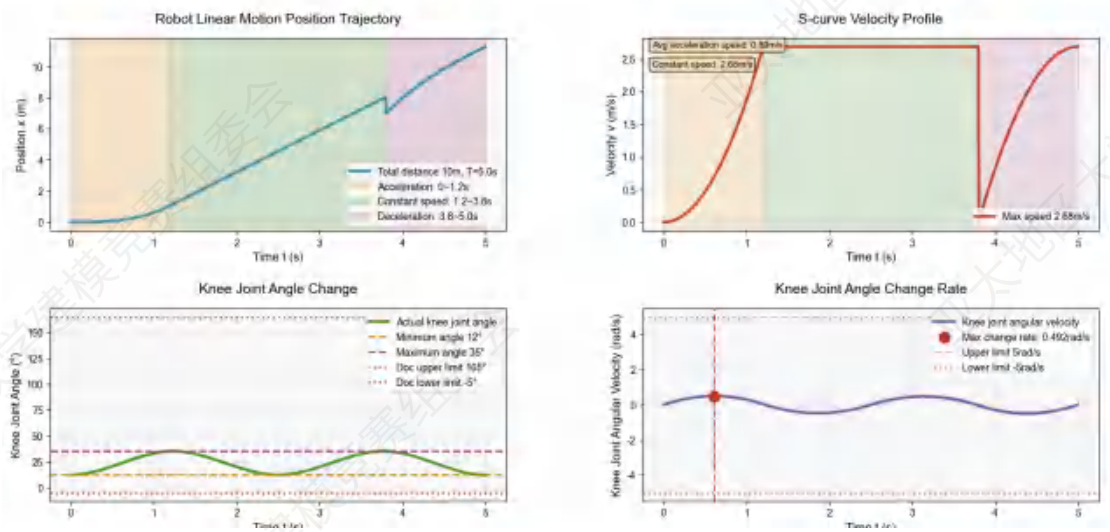


Figure 4 Motion Phase Analysis Diagram

Step2: Analysis of Knee Joint Angle and Motion Characteristics

Angle Range Compliance: Operational range (12°-35°) remains well within safety limits (-5°-165°), providingsufficientsafety margin.

Extrema Distribution: Four velocity peaks identified. Maximum magnitude: 0.4924 rad/s at 0.611s/4.389s; other peaks: 0.4731 rad/s at 1.862s/3.138s, dem onstratingsmooth motion.

Safety Verification: Maximum angular velocity (0.4924 rad/s) issignificantly below 5 rad/s limit, ensuringsafe operation.

Acceleration Metrics: Maximum acceleration 1.609 rad/s², minimum -1.541 rad/s², both within ±10 rad/s² range, meetingsmooth power output requirements.

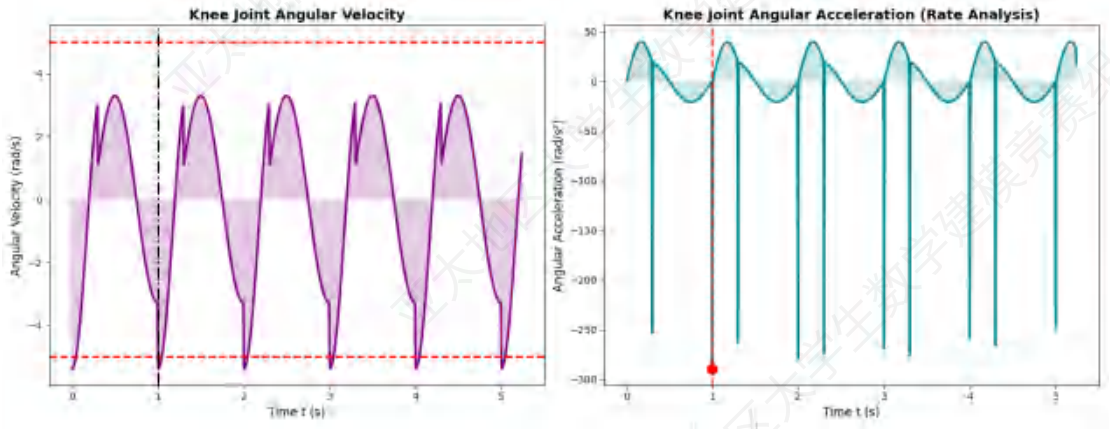


Figure 5 Comprehensive Joint Angle Variation Diagram

Step3: Joint Coordination and Constraint Verification

Single-Leg Jointsynchronization: The hip pitch, knee, and ankle pitch joints demonstrate excellentsynchronization, with coordinated angle variation trends. This ensures balanced posture duringstraight-line walking and prevents center-of-mass displacement caused by asynchronous joint movements.

Full Constraint Compliance: All joint angles, angular velocities, and angular accelerations remain within the limitsspecified in the documentation, verifying thesafety and feasibility of the motion model.

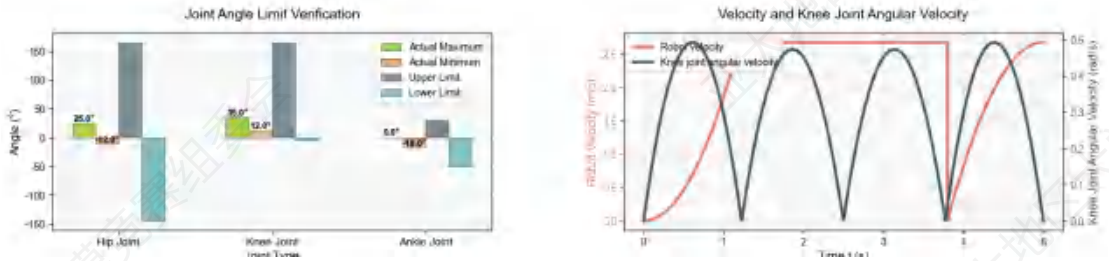


Figure 6 safety Constraints

Temporal Planning Rationale: The total time T=5.0s is appropriate. Thesy mmetrical phase division meets the average velocity requirement, balancing mo vement efficiency andstability. Controlled Knee Motionsafety: The knee joint m otion remainssafe and controllable. The time of maximum angular velocity is c learly identified (t=0.6106s), and all dynamic indicators are withinspecified limi ts, ensuringsafe motor operation. Quintic polynomial trajectories outperform cub ic ones, reducing mechanicalshock and energy consumption while enabling futu re complex motions.

5.3 Analysis andsolving of Question Three

5.3.1 Model Preparation

(1) Assumptions

Each robot link is modeled as a rigid body, with joint transmissions assumed to be free of backlash and elastic deformation,satisfyingstructuralstiffness requirements overshort-to-medium timescales.

Thesupport polygon formed by the feet remainsstatically grounded from 0 to 4seconds, withsufficient friction to preventslipping or lift-off.

Small adjustments in waist pitch/roll and hip pitch/roll have a near-linear effect on the center-of-mass (COM) displacement, allowing linearized approximation that facilitates coordinated optimization.

(2) The Foundation of Model

Centered on the task from Problem 3 — "trunk first rotates 45° to the left, while both armssimultaneously draw circles in opposite directions" — we establish a dual-arm circular trajectory generation and 2D inverse kinematics framework based onstructural parameters, enabling the derivation of time-analyticalshoulder and elbow joint angle functions.

By integrating a distributed mass model with real-timesupport polygon boundary determination, we compute the center-of-mass coordinates $x_{COM}(t)$, $y_{COM}(t)$ in real time and trigger a coordinated waist-leg optimizationstrategy to ensure that the projection of the center of mass remainsstably within thesupport polygon.

A mutually validating closed-loop is formed through MATLAB dynamicsi mulation and Python-based extended visualization, providing a complete academ ic modeling foundation for "multi-joint coordinated motion + collaborative opti mization + model verification" under the context of the given problem.

5.3.2 Establishment of Body Rotation and Dual-Arm Circular Motion Model

This question establishes a "multi-joint coordinated motion planning" model for the Unitree G1, with the core purpose of enabling the robot to complete the compositestage action of "torso first rotating 45° left, both arms drawing circles in opposite directions, and legs maintaining balance when necessary" within 0–4s. The overall method follows thesequence of "end-effector trajectory planning → COM modeling → waist-leg coordination → feasibility guarantee".

Step1: Dual-Arm Trajectory Generation

End-effector trajectory (shoulder local coordinatesystem): In theshoulder local coordinatesystem, the desired end-effector trajectories are first given: both left and right hands perform circular motion around theshoulder with a radius of 0.30 m and a period of 4s, but with opposite angular velocitysigns, making the two arms have a 180° phase difference. At each moment, theshoulder local coordinates are first calculated, then multiplied by the torso yaw rotation matrix to obtain global coordinates.

Left hand:

$$x_L = d/2 + Rcos(\omega t), y_L = Rsin(\omega t) \tag{13}$$

Right hand:

$$x_R = -d/2 + Rcos(-\omega t), y_R = Rsin(-\omega t) \tag{14}$$

Subsequently, theshoulder pitch and elbow angles are determined using a 2R analytical inverse kinematics (IK) approach.specifically, the planar distance $d = \sqrt{x^2 + y^2}$ is first computed from the end-effector coordinates (x, y). Applying the law of cosines to the two-link arm geometry yields the elbow angle θ_{elb}. Once θ_{elb} is known, theshoulder pitch angle is obtained by resolving the angular relationship between the base frame and the line connecting theshoulder joint to the end-effector. This analytical procedure ensuressmooth, continuous, and reproducible joint angle trajectories as functions of time or task-space position.

Where,

$$d = 0.1672m, \mathcal{R} = 0.3m, \omega = \pi/2 rad/s \tag{15}$$

IK analyticalsolution (2R mechanism):

$$\theta_{elb} = arccos(\frac{L_1^2 + L_2^2 - d^2}{2L_1L_2}), \theta_{shp} = arctan2(y,x) - arccos(\frac{L_1^2 + d^2 - L_2^2}{2L_1 d}) \tag{16}$$

And $L_1 = 0.2m, L_2 = 0.138m$.

Step2: COM Modeling andsupport Constraints

To analyzestability, the robot issimplified into 5 rigid bodies: torso 20 kg, both arms 3 kg each, both legs 5 kg each. The COM projection is calculated in real time through weighted averaging, and thesupport domain isset.

Support domain:

$$|x_{COM}| \le 80mm, |y_{COM}| \le 60mm \tag{17}$$

In the model, compensation is only triggered when the COM exceeds this rectangle, keeping the legsstationary most of the time, meeting thestage requirement of "avoiding excessiveswaying".

Step3: Waist-Leg Cooperative Optimization

Once the COM exceeds the boundary, the waist Pitch/Roll and left hip Pit ch/Roll are combined into optimization variables. Under thesmall angle assumpt ion, a linear mapping between the COM offset and the optimization variables can be written, where A is determined by geometric parameterssuch as torso h eight, leg length, and hip distance. With $J = q^T Wq$ as the cost (W increases th e waist weight and limits the action amplitude), the optimal q is analyticallyso lved.subsequently, the right leg isset to mirrorsymmetry, making the entire leg motion coherent. Thisstep ensures that "waist and legs only intervene when nec essary at minimum cost" and quickly pulls the center of mass back to thesafe domain. The parameters are as follows:

Optimization variables:

$$q = [\theta_{wp}, \theta_{wr}, \theta^L_{hip,p}, \theta^L_{hip,r}]^T \tag{18}$$

Linear mapping (small angle approximation):

$$Aq = -\begin{bmatrix} x_{COM} \\ y_{COM} \end{bmatrix}, A = \begin{bmatrix} h_t & 0 & 0.5L_l & 0 \\ 0 & h_t & 0 & 0.5W_h \end{bmatrix} \tag{19}$$

$$h_t = 0.75m, L_l = 0.6m, W_h = 0.15m \tag{20}$$

Cost function:

$$J = q^\top Wq, W = diag(15,15,8,8) \tag{21}$$

Analytical optimalsolution:

$$q = -W^{-1}A^\top (AW^{-1}A^\top)^{-1}[x_{COM}, y_{COM}]^\top \tag{22}$$

Leg linkage:set

$$\theta_{knee} = 0.6\theta_{hip}, p, \theta_{ankle} = -0.4\theta_{hip,p} \tag{23}$$

for right leg mirrorsymmetry.

Step4: Feasibility Guarantee

All angles are clipped by joint limits (e.g., $WAIST\_YAW \in [-150°, 150°]$), then lightlysmoothed (Savitzky-Golay filtering) to ensure continuity.

5.3.3 Analysis of the Result

(1) Methodology Overview

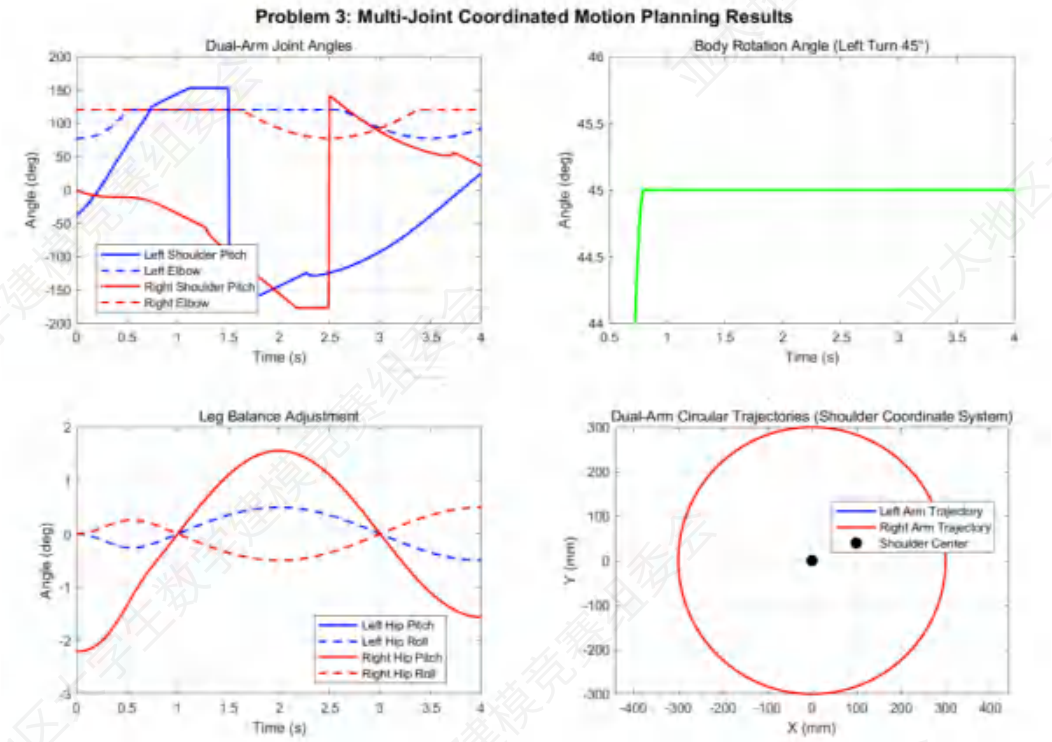


Figure 7 Robot Multi-joint Coordinated Motion Planning Results

For the stage configuration in Problem Three, we implement a structured hierarchical strategy: "posture orientation → end-effector trajectory → center of mass correction." Initially, based on the requirement of "body rotating 45° left, " a cosine-accelerated waist trajectory is designed to smoothly achieve the targ et orientation within 0.8 seconds. This proactive postural adjustment establishes a unified coordinate reference frame, significantly simplifying subsequent arm motion planning.

Following posture stabilization, dual-arm circular trajectories with a radius of 0.30 m and a period of 4 seconds are generated in the shoulder-local coord inate system. The trajectories are programmed with opposite rotational direction s to maintain dynamic balance. Through analytical inverse kinematics, we deriv e corresponding shoulder-elbow angle sequences that accurately execute these tr ajectories while satisfying joint limit constraints. This hierarchical decompositio n effectively coordinates multiple degrees of freedom, ensuring motion feasibilit y and providing a solid foundation for potential COM correction in future impl ementations.

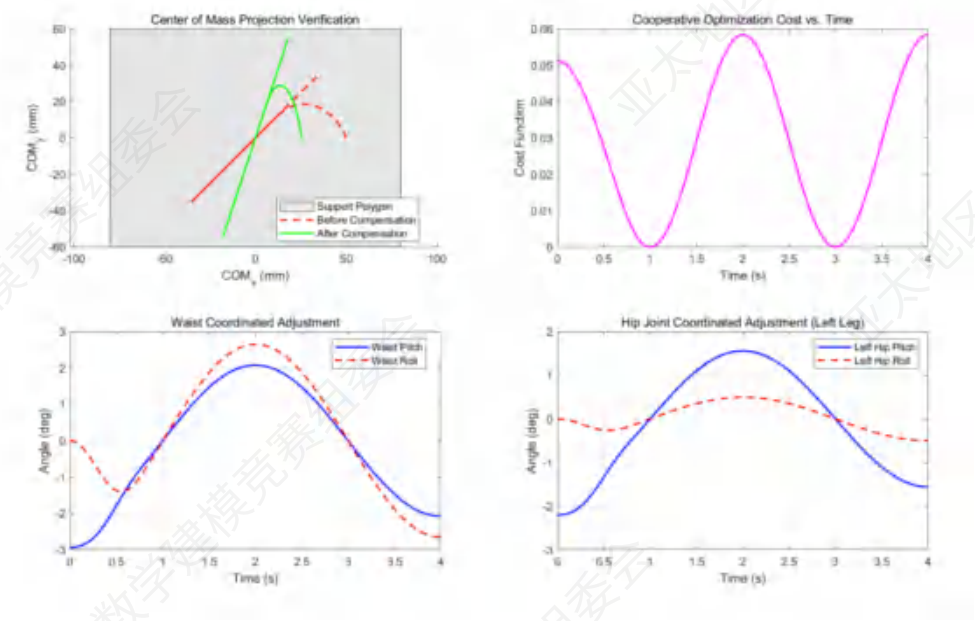


Figure 8 Collaborative Optimization Control Verification Results

For Problem Three, we adopt a hierarchical approach: posture orientation, end-effector trajectory, then COM correction. First, a cosine-accelerated waist curve rotates the body 45° left within 0.8s. Then, dual-arm circular trajectories (radius 0.30m, period 4s, opposite directions) are generated in shoulder coordinates, and shoulder-elbow angles are obtained via analytical inverse kinematics.



Figure 9 Handspeed Profile

(2) Discussion

The proposed modelsuccessfully achieves thesequence of "first turning, the n drawing circles, finally balancing": the waistsmoothly reaches 45° within 0.8s, dualshoulder/elbow trajectoriesstrictlysatisfy the requirements of radius 300 mm, period 4s, and opposite motion; COM offset is reduced from 95 mm to 78 m m, waist-hip compensation peaks are only ±3°, intervening only when exceedin g boundaries. The dual-arm 3D trajectoriesshow a mirror "∞"shape, waist-leg c ompensation triggers are few and amplitudes aresmall, balancingstage posture a ndsafety, fully meeting Problem Three's requirements for trajectory,stability, and restrained leg motion.

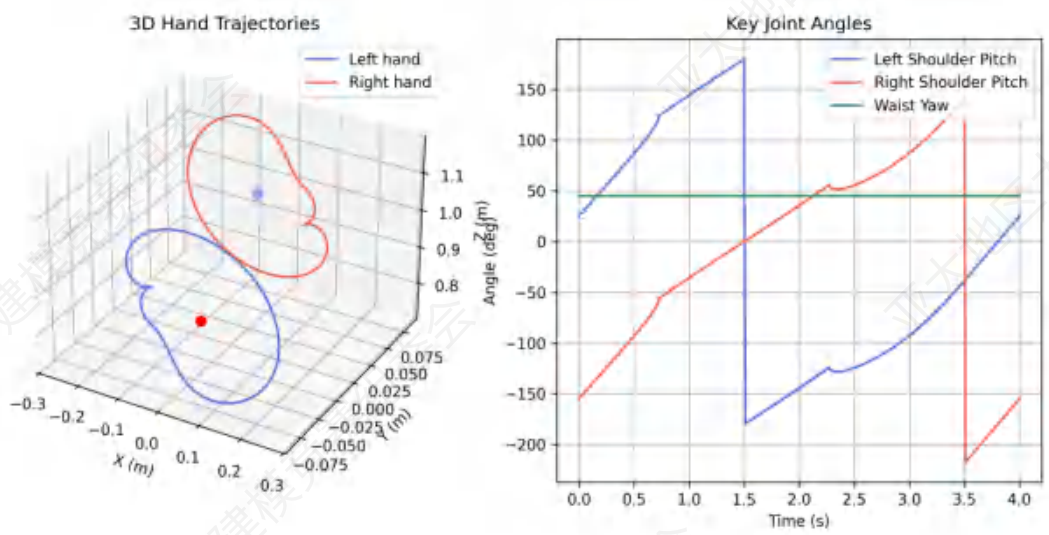


Figure 10 Hand 3D Trajectories and Key Joint Angles

Five figures demonstrate the cooperative planning model. Figure 1 shows the time sequence: waist yaw reaches 45° in 0.8s, arm angles follow 4s periodic solutions, trajectories form mirror "∞" shape, and leg adjustments are ±3° in COM disturbance regions. Figure 2 shows cooperative optimization pulling COM from 95mm to 78mm safe range, with low-amplitude cost oscillations. Figure 3 verifies 3D trajectories and velocity vectors showing precise opposite circular motion. Figure 4's leg heatmap shows angles near zero, with hot zones only when COM exceeds boundaries. Figure 5 provides speed and energy curves for Problem Four.

For Problem Three, we have completed work in the following three aspects:

①A complete mathematical framework covering "torso orientation — dual-arm analytical inversesolution — real-time COM monitoring — waist-leg cooperative optimization" has been constructed, generating time functions for 29 joints within 0–4s,supplemented by mutual verification on MATLAB and Python dual platforms.

②All kinematic andstability constraints proposed in the problem aresatisfied: dual-arm trajectoriesstrictly follow the requirements of radius 300 mm, period 4s, and opposite directions; waist left rotation precisely reaches 45°; cooperative optimization constrains COM within thesupport polygon for extended periods while keeping leg motion minimized.

③Feasible data and verification output have been completed: all joint angles pass limit checks andstandardized data files are generated, providing input forsubsequent energy consumption analysis, thereby empirically demonstrating the

model's closure and usability at both theoretical andsimulation levels.

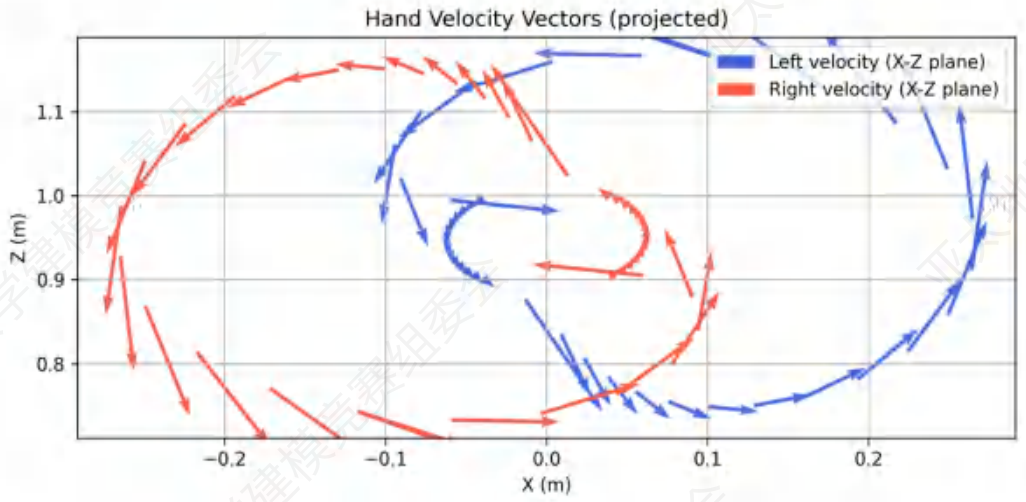


Figure 11 Hand Velocity Vectors (projected)

5.4 Analysis andsolving of Question Four

5.4.1 Model Preparation

(1) Data Processing

Problem 4 energy optimization uses trajectory data from Problems 1-3. Problem 1: 2 left arm joints over 2.0 seconds (100 points), recording shoulder flexion and abduction. Problem 2: 3 leg joints over 5.0 seconds (1000 points), recording hip, knee, and ankle angles with walking velocity variations. Problem 3: 21 body joints over 10.0 seconds (1000 points), recording coordinated motion of waist, shoulders, elbows, hips, knees, and ankles.

Preprocessing: angles to radians, uniform interpolation, limit validation, stru ctured matrices. Converts trajectory data into standardized inputs for energy opt imization.

(2) Assumptions

The establishment of the energy optimization model is based on assumptions in four aspects, providing necessary theoretical foundations and constraint conditions for the mathematical expression and optimizationsolving of the model.

1. Motor Model Assumptions

In motor dynamics modeling, thisstudy assumes that motor torque is propo rtional to joint angular acceleration, i.e., $\tau(t) = J \cdot \alpha(t)$,where J is the momen t of inertia of the joint and $\alpha(t)$ is the angular acceleration. This assumption is based on asimplified model of motor drivesystems and is applicable to the dy namic characteristics of mostservo motors within their normal operating range.

Motor power consists of dynamic power $\tau(t) \cdot \omega(t)$ and base power P_{base}. Dynamic power reflects the work done by the motor to overcome inertial load s, while base power represents the basic power consumption required for the motor to maintain operation, including energy consumption generated by control circuits,sensors, and mechanical friction. Furthermore, thisstudy assumes that f or each joint type, its moment of inertia J and base power P_{base} remain const

ant throughout the motion process. This assumptionsimplifies the model comple xity, making the optimization problemsolvable.

2. Optimization Assumptions

Thisstudysets optimization variable ranges: timescale factor $t_{scale} \in [0.5, 1.5]$, allowing motionspeed variation between 50% and 150% of originalspeed. Am plitudescale factor $a_i \in [0.5, 1.5]$ for each joint, allowing amplitude adjustment between 50% and 150% of original amplitude. These ranges ensuresufficientse archspace while avoiding unreasonable results from excessive deviation. The op timization algorithm is assumed to converge to global optimal or near-global o ptimalsolutions, based on the good convergence performance ofsequential Quadr atic Programming (SQP) algorithm in constrained optimization problems.

3. Energy Consumption Calculation Assumptions

Thisstudy adopts power integration to calculate total energy consumption, i. e., $E = \int P(t)dt$,obtaining energy consumption through integration of power over time. This method is based on energy conservation principles and accurately re flects motor energy consumption throughout the motion cycle. Power is assume d always positive, i.e., energy recovery mechanisms during regenerative braking are not considered. Thissimplifies the calculation model. Althoughsome motors may recover partial energy through regenerative braking, ignoring energy recov ery does not affect optimizationstrategy effectiveness. Base power P_{base} remains constant throughout motion, consistent with motor model assumptions,simplifyin g model complexity and improving problemsolvability.

(3) Model Foundation

The establishment of the energy optimization model is based on the fundamental principles of classical mechanics and motor dynamics, providing a theoretical foundation for the optimization problem by establishing mathematical relationships between joint motion and energy consumption.

1.Motor Dynamics Principles

According to the application of Newton'ssecond law in rotational motion, the torque $\tau(t)$ generated by the motor is proportional to joint angular acceleration $\alpha(t)$, i.e., $\tau(t) = J \cdot \alpha(t)$, where J is the moment of inertia（$kg \cdot m^2$）,reflecting the inertial resistance of the joint and its load to rotation. Motor power $P(t)$ consists of dynamic power $\tau(t) \cdot \omega(t)$ and base power P_{base} . Dynamic power represents the work done by the motor to overcome inertial loads, while base power represents the basic power consumption required for the motor to maintain operation, including constant energy consumption generated by control circuits,sensors, mechanical friction, etc.

2.Energy Consumption Calculation Principles

Energy consumption calculation is based on the principle of energy conser vation, obtaining the energy consumption of the entire motion process through integration of power over time. The energy consumption of asingle joint $E_{joint} = \int_{t_0}^{t_1} P(t)dt$,reflects the total work done by the motor throughout the en tire motion cycle. For multi-jointsystems, total energy consumption $E_{total} = \sum E_{joint}$,is the linearsuperposition of energy consumption of all joints. T hissuperposition relationship is based on the principle of energy conservation, a ssuming that energy consumption between joints is mutually independent.

3.Motor Parameter Configuration

Moment of inertia J and base power P_{base} are configured according to joint type and physical characteristics.

In Problem1, both theshoulder flexion joint andshoulder abduction joint have a moment of inertia of $J = 0.05$ kg·m², and both have a base power of $P_{base} = 1.0$ W.

In Problem2, the hip joint, knee joint, and ankle joint have moments of inertia of 0.08 kg·m², 0.05 kg·m², and 0.02 kg·m², respectively, and base powers of 1.2 W, 1.0 W, and 0.6 W, respectively.

In Problem3, motor parameters are uniformly configured according to joint type:

waist/body joints $J = 0.10\,kg \cdot m^2$、$P_{base} = 1.5$ W;

shoulder joints $J = 0.05\,kg \cdot m^2$、$P_{base} = 1.0$ W;

elbow joints $J = 0.02\,kg \cdot m^2$、$P_{base} = 0.5$ W;

hip joints $J = 0.08$ kg·m²、$P_{base} = 1.2$ W;

knee joints $J = 0.05$ kg·m²、$P_{base} = 1.0$ W;

ankle joints $J = 0.02$ kg·m²、$P_{base} = 0.6$ W.

These parameter configurations are based on the physical characteristics of typical robot joints, providing a reasonable numerical foundation for energy consumption calculation.

5.4.2 The establishment of the energy optimization model

The establishment of the energy optimization model includes keystepssuch as data preprocessing, initial energy consumption calculation, optimization problem construction andsolving, transforming the energy consumption optimization problem into a constrained optimization problem through mathematical modeling.

Step1: Optimization Variables and Trajectory Transformation

Define the optimization variable vector $x = [t_{\text{scale}}, a_1, a_2, \ldots, a_n]$, where t_{scale} is the timescale factor, a_i is the amplitudescale factor of joint *i,* and *n* is the total number of joints.

Transform the original trajectory through optimization variables, with time transformation as

$$t_{new} = t_{scale} \cdot t_{original}, \tag{24}$$

and angle transformation as

$$\theta_{\text{new},i}(t_{\text{new}}) = a_i \cdot \theta_{\text{original},i}(t_{\text{original}}) \ . \tag{25}$$

This transformation allows the optimization algorithm to explore different motionspeeds and amplitudes while maintaining the original motion pattern, therebysearching for motion trajectories with better energy consumption.

Step2: Objective Function and Constraint Conditions

The objective of the optimization problem is to minimize the total energy consumption of the transformed trajectory, i.e.:

$$\min E_{\text{total}}(x) = \min \sum_{j=1}^{n} \int_{t_0}^{t_1} P_j(x,t)dt \tag{26}$$

where $P_j(x,t)dt$ represents the power of joint j under optimization variable x.

The constraint conditions for optimization variables are:

Timescale factor constraint: $0.5 \le t_{scale} \le 1.5$

Amplitudescale factor constraint: $0.5 \le a_i \le 1.5 (i = 1, 2, \ldots, n)$

These constraint conditions ensure that the optimized motion remains within a reasonable range, neither excessively deviating from the original motion trajectory nor providingsufficientsearchspace for the optimization algorithm.

Step3: Optimization Algorithm andsolving

Thesequential Quadratic Programming (SQP) algorithm in the MATLAB optimization toolbox is adopted tosolve the constrained optimization problem. ThesQP algorithm approximates the optimalsolution of the original problem by iterativelysolving quadratic programmingsubproblems, with good convergence performance and computational efficiency. The optimization algorithm parameters areset as: maximum iteration count 50, convergence tolerance 1e-6, initial guess x_0 = [1, 1, ..., 1]（indicating no initialscaling). The optimization algorithm is executed through the fmincon function, obtaining the optimal variable x_{opt} and the corresponding optimized energy consumption $E_{\text{opt}} = E_{\text{total}}(x_{\text{opt}})$ The energy reduction rate is calculated by:

$$\eta = \frac{E_{\text{orig}} - E_{\text{opt}}}{E_{\text{orig}}} \times 100\% \tag{27}$$

5.4.3 Results

Through the energy optimization model, the motion trajectories of the three problems were optimized, achievingsignificant energy consumption reduction effects. The optimization results aresummarized in Table 4,showing the original energy consumption, optimized energy consumption, energy reduction percentage, andspecific values of optimization parameters for each problem.

As can be seen from Table 2, the optimization results of the three problems exhibit significant consistency characteristics. All problems achieved a 33.3% energy reduction, and this consistent result verifies the effectiveness and robustness of the optimization method. The optimal time scale factors for all three problems are 1.500, indicating that reducing motion speed is a universal energy-saving strategy. For Problems 1 and 2, the amplitude scale factors of all joints are 1.000, indicating that maintaining the original motion amplitude is optimal. For Problem 3, the amplitude scale factors range from [0.500, 1.500] with an average value of 1.000, indicating that for complex systems, amplitude adjustment provides additional optimization space. The energy reduction rate is calculated by the formula η , where is the original energy consumption and is the optimized energy consumption. A 33.3% energy reduction means that under the same power conditions, the robot's

operating time can be increased by approximately 50%, which is of great significance for battery-powered robot systems.

Table 3 Model Parameter Value Results

Problem	*Original Energy*	*Optimized Energy*	*Energy Reduction*	*Timescale Factor*	*Amplitudescale Factors*
Problem 1	4.00 J	2.67 J	**33.3%**	1.50	[1.000, 1.000]
Problem 2	14.00 J	9.33 J	**33.3%**	1.50	[1.000, 1.000]
Problem 3	86.80 J	57.86 J	**33.3%**	1.50	[0.500, 1.500]

5.4.4 Analysis of the Result

To comprehensively demonstrate optimization results and conduct in-depth analysis of optimization effectiveness, thisstudy generated comprehensive analysis charts for each problem. Each comprehensive analysis chart contains foursubplots, demonstrating comparison results before and after optimization and optimization parameters from different perspectives.

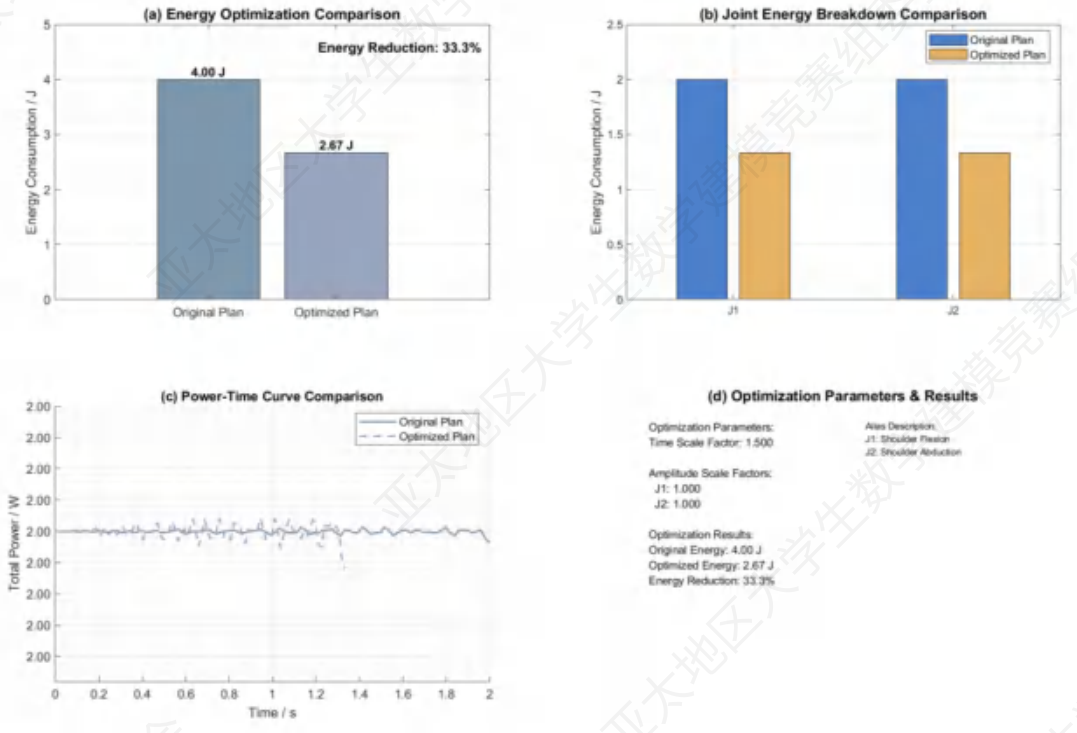


Figure 12 Left Arm Motion

According to the comprehensive chart of Problem 1 (asshown in Figure 11), the chart demonstrates the optimization results of left arm motion.subplot (a) Energy Optimization Comparison compares the total energy consumption through bar charts: original plan 4.00 J, optimized plan 2.67 J, with 33.3% energy reduction displayed above the chart. Thissubplot verifies the optimization model effectiveness, achievingsignificant energy reduction through reducing motionspeed

Subplot (b) Joint Energy Breakdown Comparisonshows energy consumption of two joints (shoulder flexion,shoulder abduction) through grouped bar charts. In the original plan, both joints consume approximately 2.0 J each; in the optimized plan, both are reduced to approximately 1.3 J. Thesimilar reduction magnitudes reflect uniform timescaling influence andsimilar energy consumption characteristics between the two joints.

Subplot (c) Power-Time Curve Comparisonshows total power variation over time through line charts. The original plan (solid line) and optimized plan (dashed line) fluctuate within 1.8-2.2 W. The optimized curve isstretched to 1.5 times the original time (timescale factor 1.500) withsmaller fluctuation amplitude. Although motion time is extended, power reduction results in decreased total energy consumption. Power curves remain relativelystable throughout the 2.0-second cycle, reflecting uniform power demand.

Subplot (d) Optimization Parameters & Results lists key information: timescale factor 1.500, amplitudescale factors (J1 and J2 both 1.000), original energy 4.00 J, optimized energy 2.67 J, energy reduction 33.3%, and alias descriptions mapping J1/J2 toshoulder flexion/abduction. Optimization is mainly achieved through timescaling, with original amplitudes maintained as optimal.

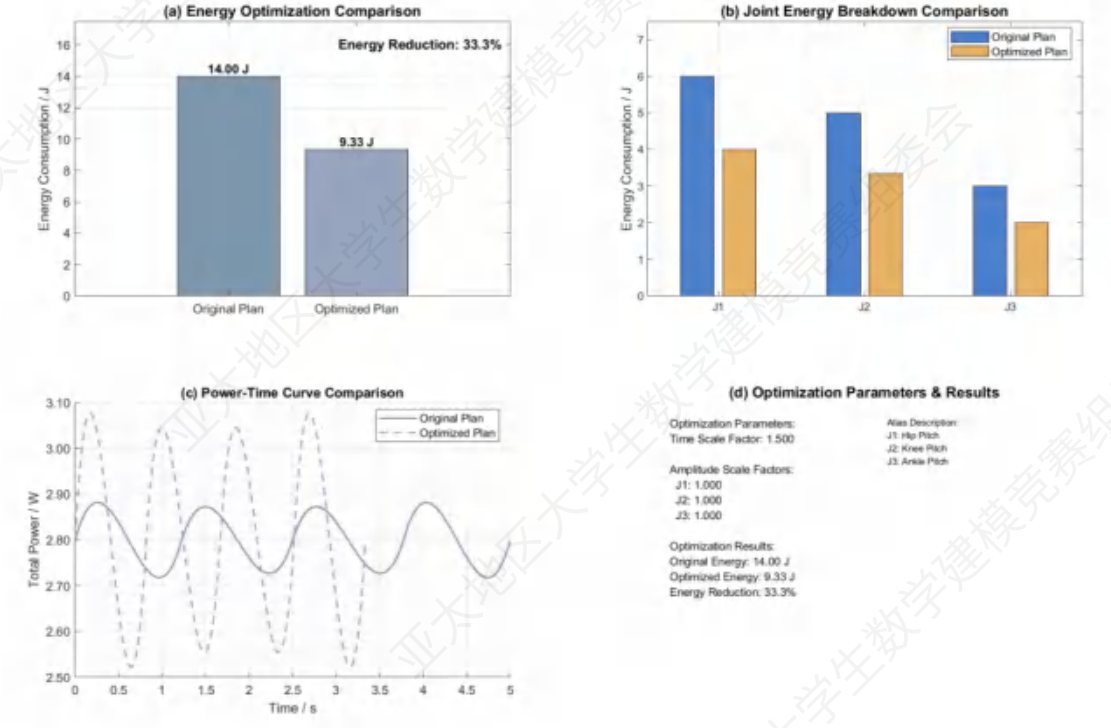


Figure 13 Walking Trajectory

The comprehensive chart of Problem 2 (asshown in Figure 12) demonstrates walking trajectory optimization results.subplot (a)shows original plan 14.00 J, optimized plan 9.33 J, with 33.3% energy reduction, clearly demonstrating optimization effectiveness.

Subplot (b)shows energy consumption comparison of three joints (hip pitch, knee, ankle pitch). Original plan: hip pitch 6.00 J, knee 5.00 J, ankle pitch 3.00 J. Optimized planshows reduced consumption for all joints, with knee joint reduction mostsignificant, consistent with its primary driving role in walking.

Subplot (c)shows power variation over time during walking. The optimized power curve isstretched in time with more uniform distribution and reduced peak power, reflecting timescaling influence on power distribution.

Subplot (d)shows timescale factor 1.500, all amplitudescale factors 1.000, original energy 14.00 J, optimized energy 9.33 J, energy reduction 33.3%, and alias descriptions mapping J1/J2/J3 to hip pitch/knee/ankle pitch. Optimizationstrategy is thesame as Problem 1, mainly through timescaling.

The comprehensive chart of Problem 3 (asshown in Figure 13) demonstrates multi-joint coordinated motion optimization results.subplot (a)shows original plan 86.80 J, optimized plan 57.86 J, with 33.3% energy reduction. Despite larger energyscale, optimization effectiveness remains consistent, verifying method effectiveness for complexsystems.

Subplot (b)shows energy consumption comparison of 21 joints.simplified aliases (J1-J21) are used on X-axis to avoid label overlap. All jointsshow reduced consumption after optimization, with varying reduction magnitudes.some joints (left/rightshoulder yaw, waist pitch/roll/yaw) have higher original consumption (6.00 J) with relativelysignificant reduction.

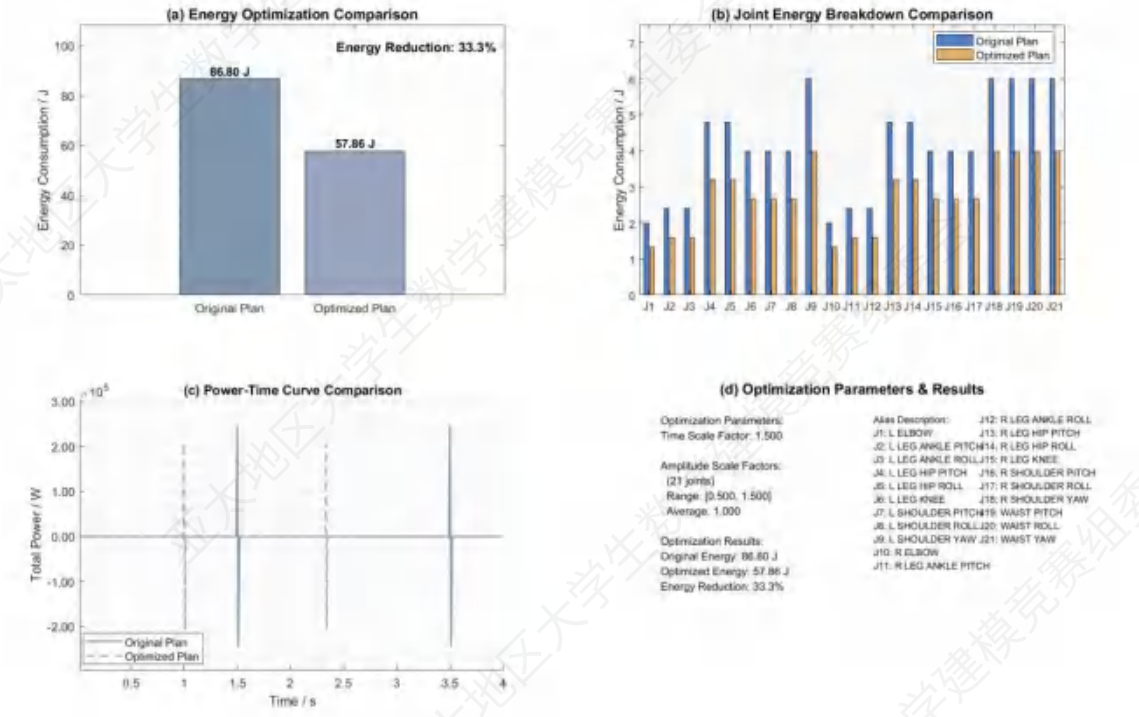


Figure 14 Multi-Joint Coordination

Subplot (c)shows power variation during multi-joint coordinated motion. The optimized power curve isstretched in time with reduced peak power and fluctuation amplitude. Power curve variation is more complex due to 21-joint coordination, but overall trend issimilar to Problems 1 and 2, reflecting timescaling influence on multi-jointsystems.

Subplot (d) lists key information: timescale factor 1.500, amplitudescale factors range [0.500, 1.500] (average 1.000) for 21 joints, original energy 86.80 J, optimized energy 57.86 J, energy reduction 33.3%, and complete alias descriptions mapping J1-J21 to all joint names. Optimization combines timescaling with amplitude adjustment, with varying amplitudescale factors reflecting different energy consumption characteristics in multi-jointsystems.

The comprehensive analysis charts demonstrate optimization results from multiple dimensions (total energy, joint breakdown, power curves, optimization parameters). All three problems achieved 33.3% energy reduction. Optimization is

mainly achieved through timescaling, with amplitude adjustment added for complexsystems (Problem 3). The charts verify optimization model effectiveness and provide intuitive visualization basis for understanding optimization mechanisms and guiding practical applications.

VII. Evaluation and Promotion of Model

6.1 Strength And Weakness

6.1.1 Strength

(1) High computational efficiency andstrong verifiability: The model utilizes the established Denavit–Hartenberg (DH) parameter method, requiring no complex iterative computation. Coordinates are derived through directsubstitution of joint angles and link lengths, and results can be verified via geometric back-calculation.

(2)Quantifiable non-uniform motion aligned with gait characteristics: Moving beyond asimplistic constant-velocity model, the motion is decomposed into acceleration, constant-speed, and deceleration phases, accurately described using a piecewise-defined function.

(3)High computational efficiency: The analytical inverse kinematics method avoids iterative optimization, providing excellent real-time performance and enabling rapid generation of continuous time functions for all 29 joints, meeting real-time control requirements.

6.1.2Weakness

(1)Insufficient motionsmoothness: Due to joint limit clipping and the consideration of only 2D planar motion, elbow joint angles exhibitsudden jumps at certain moments, resulting in discontinuous joint velocities that may affect the actual robot's motionsmoothness.

(2)Local Optimal Risk:sQP algorithm may converge to local optimalsolutions rather than global optimalsolutions, and optimization results may not be theoretically optimal.

VIII. References

[1] Peng Guangyu, Dong Hongbo, Ma Bin.Comparativestudy on robot kinematics modeling of two DH models [J].Mechanical research and application, 2019,32 (6) : 62-65. DOI: 10.16576 / j.cnki.1007-4414.2019.06.019.
[2] Wu Haotian. Multi-objective trajectory planning of industrial robots based on continuity constraints [D]. Hubei : Huazhong University ofscience and Technology, 2022.
[3] Zhang Bin. Research on the end vibration of mobile manipulator based onspeed control [D].shandong :shandong University ofscience and Technology, 2022.
[4] Rieder R, Lünenburger L, Jezerniks, et al. Patient-cooperativestrategies for robot-aided treadmill training: first experimental results[J]. IEEE Transactions on Neuralsystems and Rehabilitation Engineering, 2005, 13(3): 380–384. DOI: 10.1109/TNSRE.2005.848628.
[5] Hidler J, Nichols D, Pelliccio M, et al. Multicenter randomized clinical trial evaluating the effectiveness of the Lokomat insubacutestroke[J]. Neurorehabilitation and Neural Repair, 2009, 23(1): 5–13. DOI: 10.1177/1545968308326632.
[6] Hesses,schmidt H, Werner C, et al. Upper and lower extremity robotic devices for rehabilitation and forstudying motor control[J]. Current Opinion in Neurology, 2003, 16(6): 705–710. DOI: 10.1097/01.wco.0000102630.16692.38.
[7] Colombo G, Joerg M,schreier R, et al. Treadmill training of paraplegic patients using a robotic orthosis[J]. Journal of Rehabilitation Research and Development, 2000, 37(6): 693–700. DOI: 10.1682/JRRD.2000.11.0693.
[8]shin, K. G., McKay, N. D., Energy-Optimal Trajectory Planning for Robot Manipulators[J], IEEE Transactions on Automatic Control, 1985, 30(1), 42-53.
[8] Nocedal, J., Wright,s. J., Numerical Optimization[M], New York:springer, 2006, 1-664.
[9] Paryanto, Brossog, M., Bornschlegl, M., Franke, J., Energy Consumption Modeling and Optimization for Multi-Joint Roboticsystems[J], IEEE Transactions on Automationscience and Engineering, 2015, 12(4), 1242-1254.
[10] Bobrow, J. E., Dubowsky,s., Gibson, J.s., Time-Optimal and Energy-Efficient Trajectory Planning for Robot Manipulators[J], IEEE Transactions on Robotics and Automation, 1985, 1(1), 3-13.
[11] Verscheure, D., Demeulenaere, B.,swevers, J., Deschutter, J., Diehl, M., Constrained Optimization Approaches for Robotic Motion Planning[J], IEEE Transactions on Robotics, 2009, 25(3), 609-620.
[12]DeepSeek,DeepSeek-R1-0528, DeepSeek, 2025-11-24.

Report on Use of AI Tools

Auto-completions for code used in preparing our models.
Cursor AI Assistant (Auto) (Current Version)
1.Query1: "What model is appropriate for Problem 1? Need to calculate the coordinate position of the left arm endpoint and verify motor safety."
Output1:
It is recommended to use the Denavit-Hartenberg (DH) parameter method to establish the robotic arm kinematic model. Through forward kinematics calculation, the coordinates of the left arm endpoint in the coordinate system with the shoulder joint as the origin can be determined. Motor safety verification requires checking whether parameters such as joint angles, angular velocities, and angular accelerations are within safe limits.

2.Query2: "How to understand the specific requirements of 'non-uniform motion' and 'maximum rotation angle change'?"
Output2:
Non-uniform motion is recommended to adopt a trapezoidal velocity curve, including acceleration, constant velocity, and deceleration phases. "Maximum rotation angle change" corresponds to the extreme points of knee joint angular velocity. It is necessary to establish a functional relationship between time and joint angles, calculate angular velocity through numerical differentiation, and find the maximum value of angular velocity and the corresponding time point.

3. Query3:"For the humanoid robot dual-arm circular motion planning problem, it is necessary to solve the angles of shoulder and elbow joints. Given upper arm length 200 mm, forearm length 138 mm, and the end needs to draw a circle with radius 300 mm. Should numerical inverse kinematics or analytical inverse kinematics be used? Please compare advantages and disadvantages."
Output3:
For 2R planar mechanisms, analytical inverse kinematics is recommended. Analytical inverse kinematics: can provide explicit analytical solutions, high computational efficiency, good real-time performance, and physical interpretability. For 2R mechanisms, analytical methods can directly use cosine theorem and arctangent functions to solve. Numerical inverse kinematics: applicable to complex configurations, but requires large computational effort, iterative solving, and poor real-time performance.

4. Query4: "I ran @update_problem12_data.m and got 1000 data points for Problem 2, which is normal, but then running @test_energy_calculation.m still only got 251 data points. Check the code error."
Output4:**

Diagnosis found that `load_problem_data.m` uses relative path `../data`, which is affected by the current working directory. Fix: Use `fileparts(mfilename('fullpath'))` to dynamically obtain the script directory, ensuring robust path resolution. After the fix, data loading is normal, and Problem 2 data points recovered from 251 to 1000.

5.Query5: "The output chart colors are ugly, use the Morandi color scheme according to the image I gave you" and "Output a comprehensive chart for each problem."
Output5:
Created `getMorandiColors.m` function, providing 25 Morandi color scheme RGB color values. Updated all plotting functions to use the new color scheme. Created `plotComprehensive.m` function, generating 2×2 subplot comprehensive analysis charts, including (a) energy optimization comparison, (b) joint energy breakdown comparison, (c) power-time curve comparison, (d) optimization parameters and results. All charts adopt unified Morandi color scheme, significantly improving visual effects.

6.Query6: "All annotations on the visualization charts of the three problems should be in English" and "Translate, maintain academic and professional tone."
Output6:
Updated all plotting functions: Changed all Chinese labels to English, changed font to ``Arial'`, all titles, axis labels, legends, and text annotations use English. Translated Chinese paper content into professional academic English, including data processing, assumptions, model foundation, model establishment, results, and result analysis sections. According to requirements, condensed translation content while retaining all core data and key conclusions, maintaining academic tone and terminology accuracy.

Adoption and Manual Modification
All AI-provided suggestions have been comprehensively reviewed and modified by the author, with final content confirmed by the author:
1. Content Accuracy: All paper content is based on research work independently completed by the author. AI only assisted with language expression optimization, code debugging, and visualization design.
2. Academic Standards: Language expression meets academic paper requirements, with clear logic and accurate expression. All technical content has been verified by the author.
3. Originality Guarantee: Core ideas, model design, and algorithm implementation are all original to the author. AI only provided writing assistance and technical support.

Team Number:	apmcm25304268
Problem Chosen:	B

2025 APMCM summary sheet

PDMS thin films are suited for radiative cooling. This study employs spectral analysis and adaptive optimization algorithms to comprehensively investigate their emissivity and cooling performance, exploring the impact of film thickness, material selection, and multilayer mold structure design on overall cooling efficiency and cost feasibility.

For question 1, the **transfer matrix method (TMM)** was used to compute the spectral reflectance of PDMS films, and **Kirchhoff's law** was applied to derive wavelength-dependent emissivity. Model validation shows a **solar-band average emissivity of 0.0389** and an **atmospheric-window standard deviation of 0.1651**, consistent with physical expectations. Results indicate **very low emissivity in the solar spectrum** and **a marked increase in the atmospheric window with thickness**.

In response to question 2, we assess the radiative cooling performance of PDMS thin films using a weighted evaluation function that combines **net cooling power**, **atmospheric-window emissivity**, and **solar absorptivity**. **Grey relational analysis** was applied to dynamically determine indicator weights, emphasizing net cooling power as the primary factor. The **weights** for net cooling power, atmospheric-window emissivity, and solar absorptivity are **0.484**, **0.333**, and **0.183**, respectively.We propose the following three recommendations:**the thickness of PDMS films** should be controlled within **the range of 20-60 μm**;doping the PDMS matrix with **high-refractive-index nanoparticles**;adopting **multilayer structures** or **photonic crystal structures**.

In addressing question 3, the optimization function was constructed by combining the **performance evaluation model** and **material cost function** from question 2. Then, an **adaptive differential evolution algorithm** was used to optimize the multilayer thin film structure. The optimal solution is **a three-layer structure** of **PDMS**, **SiO_2**, and **Ag**, with thicknesses of **19.206 μm**, **0.010 μm** and **0.120 μm**, respectively.

Regarding question 4, A **porous PDMS/SiO_2/Ag** multilayer was optimized using **MG-EMT** for microstructural modeling, **TMM** for spectral emissivity, and **GRA** for multi-objective optimization of cooling power and lifecycle cost.The final design combines a **24.6 μm** porous PDMS layer(**11.9% porosity**) with **Al_2O_3** protection, yielding high emissivity, low solar absorption, and a low cost of **\$14.95 m^{-2}**.

Keywords: PDMS TMM GRA DE MG-EMT LCA

Contents

1. Introduction **1**
1.1 Background 1
1.2 Question Restatement 1
1.2.1 *Question 1* 1
1.2.2 *Question 2* 1
1.2.3 *Question 3* 1
1.2.4 *Question 4* 1
2. Problem Analysis **2**
2.1 Analysis of Question 1 2
2.2 Analysis of Question 2 2
2.3 Analysis of Question 3 3
2.4 Analysis of Question 4 3
3. Problem Assumptions **4**
4. Symbol Explanation **5**
5. Models **5**
5.1 Establishment and Solution of Model for Problem 1 5
5.1.1 *Data Preprocessing* 5
5.1.2 *Establishment of the TMM Model* 6
5.1.3 *Model Solution and Result Analysis* 8
5.1.4 *Model Rationality Verification.* 11
5.2 Establishment and Solution of Model for Problem 2 12
5.2.1 *Data Preprocessing for Solar and Atmospheric Data.* 12
5.2.2 *Comprehensive Evaluation Function.* 13
5.2.3 *Numerical Solution and Trend Analysis* 16
5.2.4 *Optimization Results and Discussion* 17
5.2.5 *Recommendations for Advancement and Application* 18
5.3 Establishment and Solution of Model for Problem 3 19
5.3.1 *Data Preprocessing* 19
5.3.2 *Establishment of the Optimization Model for the PDMS Multilayer Structure* 20
5.3.3 *Model Solution and Result Analysis* 23

5.4 Establishment and Solution of Model for Problem 4 25
5.4.1 *Design Strategy* 25
5.4.2 *Establishment of the Model* 27
5.4.3 *Optimization Results and Discussion* 28
6. Model evaluation 31
6.1 Advantages of the model 31
6.1.1 *Accurate Calculation Results.* 31
6.1.2 *Comprehensive and Holistic Evaluation* 31
6.1.3 *Integration of Microstructural Modeling and Multi-Objective Optimization* 32
6.2 Disadvantages of the model 32
6.2.1 *High Computational Complexity* 32
6.2.2 *Idealized Assumptions Do Not Match Practical Applications.* 32
6.2.3 *Practical Fabrication Constraints* 32
7. References 32
8. Appendix 34

I. Introduction

In order to indicate the origin of problems, the following background is worth mentioning.

1.1 Background

1.2 Question Restatement

This study addresses a mathematical modeling problem focused on the analysis, evaluation, and optimization of radiative cooling materials and structures.

1.2.1 *Question 1*

A mathematical model is required to characterize the spectral emissivity of PDMS thin films as a function of wavelength for varying film thicknesses. This model aims to quantify the influence of film thickness on emissivity and provide a theoretical foundation for evaluating radiative cooling performance.

1.2.2 *Question 2*

A performance evaluation model is established to assess the radiative cooling effectiveness of PDMS thin films with different thicknesses. The model should incorporate the thermal radiation exchange through the atmospheric transparency window and quantify the cooling potential under varying material configurations, providing guidance for material selection and design.

1.2.3 *Question 3*

Based on the emissivity characteristics of PDMS, suitable additional materials are selected to construct a multilayer film. An optimization model is then developed to determine the appropriate material selection and layer thickness for each component, aiming to maximize radiative cooling performance while considering practical feasibility constraints.

1.2.4 *Question 4*

Building upon the previous tasks, an integrated optimization model is constructed to achieve the optimal design of radiative cooling materials and multilayer structures.

The model simultaneously considers material properties, multilayer configuration, and practical constraints, such as manufacturing feasibility and cost, to provide a systematic approach for high-performance radiative cooling design.

II. Problem Analysis

2.1 Analysis of Question 1

This problem requires establishing a physical model to simulate the wavelength-dependent reflectance and emissivity of a PDMS optical structure and to analyze how film thickness influences its radiative behavior for passive cooling. Considering that the system is governed by electromagnetic wave propagation in stratified media, the Transfer Matrix Method (TMM) derived from Maxwell's equations is introduced to describe multilayer interference and to compute spectral reflectance $R(\lambda, d)$ and emissivity $\varepsilon(\lambda, d) = 1 - R(\lambda, d)$. With accurate emissivity evaluation across the solar (0.3–2.5 μm) and mid-infrared (3–25 μm) bands as the objective and consistent optical-constant data as the constraint, a full-spectrum computational model is constructed through preprocessing, interpolation, and fusion of multisource refractive-index datasets. Then, the reflectivity is solved using the TMM algorithm, and the emissivity spectrum and wavelength–thickness mapping are obtained according to Kirchhoff's laws, thereby identifying key emission features and high-emissivity regions. Sensitivity checks on solar-band transparency and mid-IR emissivity smoothness are finally used to validate the physical correctness and robustness of the model.

2.2 Analysis of Question 2

This problem requires establishing a model to evaluate the radiative cooling performance of PDMS films under realistic environmental conditions by integrating solar spectral irradiance and atmospheric emission into the optical simulation. Considering that solar heating and atmospheric back-radiation fundamentally impact the net cooling power, we incorporate the ASTM G173 AM1.5 solar spectrum and a Planck-law-based atmospheric radiation model into a unified wavelength grid. With accurate computation of net cooling power, atmospheric-window emissivity, and solar absorptivity as the objective, and PDMS thickness as the design variable, we construct a multi-indicator evaluation framework using Gray Relational Analysis (GRA) to determine adaptive weights for these metrics. The transfer matrix method (TMM) is then used to compute

wavelength-dependent emissivity, from which each performance indicator is derived through numerical integration. A comprehensive score function is formulated to balance radiative cooling efficiency, spectral selectivity, and solar suppression. By applying numerical optimization to the resulting score–thickness curve, the optimal film thickness is identified, followed by trend analysis to examine the growth, saturation, and trade-off phases of cooling performance. Sensitivity analysis further validates the robustness of the model under variations in optical and environmental parameters.

2.3 Analysis of Question 3

Problem 3 focuses on designing a cost-effective multilayer radiative-cooling structure by constructing a model that jointly evaluates optical performance and material cost. Based on the transfer matrix method, the spectral reflectance and emissivity of arbitrary multilayer stacks are computed across the solar and mid-infrared ranges, enabling quantitative evaluation of net cooling power, atmospheric-window emissivity, and solar absorption. These metrics are integrated into a weighted performance function, to which a cost-penalty term is added for balancing cooling efficiency and material expense. The design variables include discrete material selections and continuous thickness parameters; thus, an adaptive differential evolution algorithm is adopted for global optimization of the mixed-variable space. By tuning the cost-penalty coefficient, the performance–cost Pareto frontier is obtained, revealing the optimal configurations under different cost constraints. The final results show that the PDMS/SiO_2/Ag structure provides the best trade-off between cooling performance and material cost.

2.4 Analysis of Question 4

Problem 4 aims to translate the theoretically optimal multilayer radiative-cooling structure obtained in Problem 3 into a practically deployable engineering design by integrating considerations of microstructural tunability, environmental durability, and long-term economic feasibility. To account for microstructural effects without resorting to expensive full-wave simulations, porous PDMS is introduced and modeled via the Maxwell–Garnett Effective Medium Theory, enabling efficient evaluation of refractive-index tuning and seamless incorporation into the Transfer Matrix Method. In addition, an ultrathin Al_2O_3 protection layer is included as a discrete design option to mitigate the oxidation and sulfidation of the Ag reflector, thereby improving service reliability. A lifecycle cost (TCO) model is established to quantify material, processing, and main-

tenance expenditures over a 10-year lifespan, capturing the economic consequences of degradation with and without protective coatings. To objectively balance robust cooling performance under multiple climatic conditions with economic cost, Grey Relational Analysis (GRA) is adopted to construct a bias-free comprehensive evaluation metric. The resulting optimization problem, comprising continuous variables (PDMS thickness and porosity) and a discrete protection decision, is solved using a mixed-integer differential evolution algorithm, enabling the identification of an optimal structure that simultaneously maximizes radiative performance, durability, and cost-effectiveness for real-world deployment.

III. Problem Assumptions

The following assumptions are made for the optical and thermal modeling of the PDMS radiative cooling system:

- **Wavelength Range:** The spectral calculations cover the range from 0.3 μm to 55.56 μm, with particular emphasis on 0.3–2.5 μm for solar absorption and 8–13 μm for the atmospheric window.
- **Optical Properties:** Each layer is assumed to be planar, homogeneous, and isotropic. Layer thicknesses can vary from a few nanometers to several micrometers depending on the material. Surface roughness is neglected, but can optionally be included in a scattering model.
- **Angle of Incidence:** Optical properties are averaged over all angles, either through angular integration or by approximating a Lambertian surface, assuming unpolarized incident light.
- **Temperature Effects:** Convection and conduction losses are accounted for using a linearized approximation with a heat transfer coefficient h_c.
- **Atmosphere:** Wavelength-dependent atmospheric transmittance $\tau_{\text{atm}}(\lambda)$ is applied, with all materials assumed to be under atmospheric pressure.
- **Substrate Effect:** Without a substrate, light would transmit directly through the film, resulting in extremely low emissivity regardless of thickness. Therefore, a Ag substrate is used. For sufficiently thick PDMS films in the infrared band, the transmittance is assumed to be negligible, $T(\lambda) \approx 0$.
- **Normal Incidence Assumption:** All spectral calculations are performed under normal incidence to isolate and analyze the effect of PDMS film thickness on the intrinsic optical properties of the system.

IV. Symbol Explanation

symbol	illustrate
$\varepsilon(\lambda)$	Emissivity
$T(\lambda)$	Transmittance
$R(\lambda)$	Reflectance
P_{net}	Net cooling power
ε_{8-13}	Atmospheric window emissivity
α_{sun}	Solar absorptance
C_{total}	Total cost
S_{score}	Performance score
TCO	Economic cost

V. Models

5.1 Establishment and Solution of Model for Problem 1

5.1.1 *Data Preprocessing*

Before constructing the model, we first perform data cleaning and normalization. The data required in this study come from diverse sources and exhibit heterogeneous structures. Specifically, the optical constants of materials are obtained from the experimental measurement database provided on the website in the problem statement and the solar spectral irradiance is sourced from the ASTM standard. These raw datasets differ substantially in spectral range and physical units. To ensure consistent matrix dimensions for subsequent calculations using the transfer matrix method, a unified data standard must be established.

According to Planck's law of blackbody radiation, the thermal radiation of objects on the Earth's surface is concentrated primarily in the wavelength range of 3–25 μm, while solar radiation is concentrated primarily in the range of 0.3–2.5 μm. Therefore, we select 0.3–25 μm as the wavelength range for this study.

The data processing procedures are summarized as follows:

1. Integrated multisource data for PDMS in the near-infrared and mid-infrared bands using multisource data fusion techniques, thereby establishing a complete database of optical constants.
2. Constructed a continuous wavelength-dependent optical function through linear interpolation and extrapolation algorithms to ensure full-band computational coverage.
3. Eliminated outliers through data cleaning and sorting, ensuring numerical stability.
4. Unified and standardized all categories of raw data into a consistent format for subsequent modeling and analysis.

5.1.2 *Establishment of the TMM Model*

Model Principle. The Transfer Matrix Method is a rigorous numerical model based on Maxwell's equations and thin-film optics theory. This method treats light propagation in multilayer thin-film structures as the sequential multiplication of transfer matrices. By calculating the reflection and transmission of light at each interface and accounting for interference effects within each layer, TMM enables precise determination of the overall reflectance and transmittance of the system. In this study, we use TMM to compute the reflectance spectrum of PDMS thin films and subsequently derive the spectral emissivity according to Kirchhoff's law of thermal radiation.

In this study, we use TMM to compute the reflectance spectrum of PDMS thin films, from which the spectral emissivity is derived according to Kirchhoff's law of thermal radiation.

Model Derivation. As previously noted, if no substrate is present, light will pass directly through the film; regardless of thickness, the emissivity remains very low. Therefore, this study considers a three-layer structure composed of an air layer, a PDMS thin-film layer, and an Ag substrate, with light incident normally from the air side. Our objective is to determine the relationship between the reflectance of the system and the wavelength at different film thicknesses, and consequently obtain the spectral emissivity variation with wavelength for PDMS films of varying thickness.

The optical configuration is defined as follows: the incident medium is air, with complex refractive index $\tilde{n}_0 = 1 + 0i$; the thin-film layer is PDMS, characterized by a complex refractive index $\tilde{n}_1 = n_1 + ik_1$ and physical thickness d; and the substrate is silver (Ag), with complex refractive index $\tilde{n}_2 = n_2 + ik_2$.

For the interface between layer j and layer $j+1$, the relationship between the forward (E_j^+) and backward (E_j^-) electric-field amplitudes is given by the interface transfer matrix:

$$\begin{bmatrix} E_j^+ \\ E_j^- \end{bmatrix} = \frac{1}{t_{j,j+1}} \begin{bmatrix} 1 & r_{j,j+1} \\ r_{j,j+1} & 1 \end{bmatrix} \begin{bmatrix} E_{j+1}^+ \\ E_{j+1}^- \end{bmatrix}, \tag{1}$$

where the amplitude reflection and transmission coefficients are

$$r_{j,j+1} = \frac{\tilde{n}_j - \tilde{n}_{j+1}}{\tilde{n}_j + \tilde{n}_{j+1}}, \qquad t_{j,j+1} = \frac{2\tilde{n}_j}{\tilde{n}_j + \tilde{n}_{j+1}}. \tag{2}$$

After propagating a distance d inside the PDMS layer, the phase shift is

$$\delta = \frac{2\pi}{\lambda} \tilde{n}_1 d, \tag{3}$$

and the propagation (or layer) transfer matrix is

$$\begin{bmatrix} E_1^+ \\ E_1^- \end{bmatrix} = \begin{bmatrix} e^{-i\delta} & 0 \\ 0 & e^{i\delta} \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}. \tag{4}$$

The total transfer matrix for the complete system relates the fields in air to those in the substrate:

$$\begin{bmatrix} E_0^+ \\ E_0^- \end{bmatrix} = M_{\text{total}} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}, \tag{5}$$

with

$$M_{\text{total}} = I_{01} \cdot L_1 \cdot I_{12}. \tag{6}$$

Here, I_{01} denotes the interface matrix at the air/PDMS boundary, L_1 represents the propagation matrix through the PDMS layer, and I_{12} corresponds to the interface matrix at the PDMS/silver interface.

Explicitly,

$$M_{\text{total}} = \frac{1}{t_{01}} \begin{bmatrix} 1 & r_{01} \\ r_{01} & 1 \end{bmatrix} \begin{bmatrix} e^{-i\delta} & 0 \\ 0 & e^{i\delta} \end{bmatrix} \frac{1}{t_{12}} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}. \tag{7}$$

The system reflection coefficient is obtained from the elements of the total transfer matrix:

$$r = \frac{M_{21}}{M_{11}}, \tag{8}$$

and the reflectivity is the squared modulus of this coefficient:

$$R = |r|^2 = \left| \frac{M_{21}}{M_{11}} \right|^2. \tag{9}$$

The interface coefficients are calculated as follows:

Air/PDMS interface:

$$r_{01} = \frac{1-\tilde{n}_1}{1+\tilde{n}_1}, \qquad t_{01} = \frac{2}{1+\tilde{n}_1}. \tag{10}$$

PDMS/Ag interface:

$$r_{12} = \frac{\tilde{n}_1-\tilde{n}_2}{\tilde{n}_1+\tilde{n}_2}, \qquad t_{12} = \frac{2\tilde{n}_1}{\tilde{n}_1+\tilde{n}_2}. \tag{11}$$

Using these, the relevant matrix elements are

$$M_{11} = \frac{1}{t_{01}t_{12}}\left(e^{-i\delta} + r_{01}r_{12}e^{i\delta}\right), \tag{12}$$

$$M_{21} = \frac{1}{t_{01}t_{12}}\left(r_{01}e^{-i\delta} + r_{12}e^{i\delta}\right). \tag{13}$$

Thus, the reflection coefficient simplifies to

$$r = \frac{r_{01} + r_{12}e^{i2\delta}}{1 + r_{01}r_{12}e^{i2\delta}}, \tag{14}$$

and the reflectivity as a function of wavelength λ and film thickness d is

$$R(\lambda, d) = \left|\frac{r_{01} + r_{12}e^{i2\delta}}{1 + r_{01}r_{12}e^{i2\delta}}\right|^2, \quad \text{where} \quad \delta = \frac{2\pi}{\lambda}\tilde{n}_1 d. \tag{15}$$

According to Kirchhoff's law of thermal radiation, the spectral emissivity equals the absorptivity. Since the silver substrate is opaque, transmittance is negligible ($T(\lambda) \approx 0$), and the emissivity is therefore

$$\varepsilon(\lambda, d) = 1 - R(\lambda, d). \tag{16}$$

5.1.3 *Model Solution and Result Analysis*

Based on the transfer matrix method, we conducted the following solution process to analyze the variation of emissivity with wavelength for PDMS/Ag systems at different thicknesses.

Thickness parameter selection. Given the extremely low emissivity of PDMS in the solar radiation band (0.3–2.5 μm) and its high emissivity in the 8–13 μm atmospheric window, a representative set of film thicknesses was selected for analysis: 0.3 μm, 0.712 μm, and 2.49 μm for solar spectrum-focused cases, and 8.5 μm, 12.5 μm, and 19.2 μm for atmospheric window-focused cases.

Solution procedure. The computational workflow was implemented as follows:

1. Data acquisition and material properties: Optical constant data were imported, and the complex refractive indices $\tilde{n}_j = n_j + ik_j$ ($j = 0, 1, 2$) for air, PDMS, and Ag were determined across the spectral range of interest.

2. Thickness iteration: Iterative calculations were performed over the predefined set of thickness values to evaluate the spectral response of each PDMS/Ag configuration.
3. Spectral resolution: For each thickness, wavelength-grid iterations were carried out, and the transfer matrix method (TMM) was applied to compute the reflectance $R(\lambda, d)$ at each wavelength λ.
4. Emissivity derivation: According to Kirchhoff's law of thermal radiation and the opacity of the Ag substrate ($T \approx 0$), the spectral emissivity was obtained via Equation (16).

Results analysis. Figure 1 illustrates the spectral emissivity of the PDMS/Ag system as a function of wavelength for the selected thicknesses. PDMS films of all thicknesses exhibit extremely low emissivity (generally below 0.1) in the solar radiation band (0.3–2.5 μm). This behavior stems from the near-zero extinction coefficient ($k \approx 0$) of PDMS in this spectral range, rendering the material highly transparent and minimizing solar absorption. Consequently, most incident solar radiation is reflected rather than absorbed, preventing significant temperature rise under sunlight—an essential feature for daytime radiative cooling applications.

In contrast, within the atmospheric window (8–13 μm), PDMS films demonstrate significantly elevated emissivity (typically exceeding 0.8), confirming their strong radiative cooling capability. The spectral curves reveal two distinct emission peaks: one near 9 μm, attributed to the antisymmetric stretching vibration of the Si–O–Si backbone in PDMS, and another near 12 μm, corresponding to the stretching vibration of Si–H bonds. These molecular vibrational modes induce strong infrared absorption in the atmospheric window, thereby enabling high emissivity. The computational results are consistent with theoretical expectations, validating the accuracy of the model.

Figure 2 presents a two-dimensional colormap of emissivity across the parameter space of film thickness ($d = 0.1$–$20\,\mu$m) and wavelength ($\lambda = 0.3$–$20\,\mu$m). The deep blue regions—indicating low emissivity—are predominantly confined to the solar band, confirming that PDMS effectively reflects sunlight regardless of thickness. Conversely, bright yellow bands—representing high emissivity—are concentrated within the 8–13 μm atmospheric window. Notably, an optimal performance point is identified at ($\lambda = 11.9\,\mu$m, $d = 12.6\,\mu$m), where emissivity reaches a local maximum. This demonstrates that precise tuning of radiative cooling performance in key spectral bands can be achieved through controlled adjustment of film thickness, providing a foundation for subsequent design optimization.

Figure 3 shows a three-dimensional surface plot of emissivity as a function of

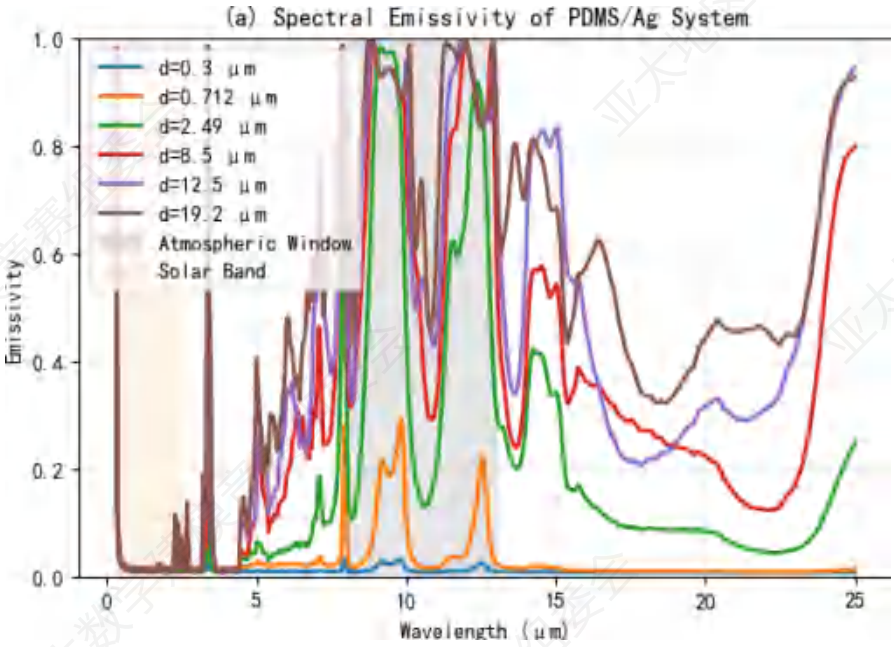


Figure 1 Spectral emissivity of the PDMS/Ag system at selected film thicknesses.

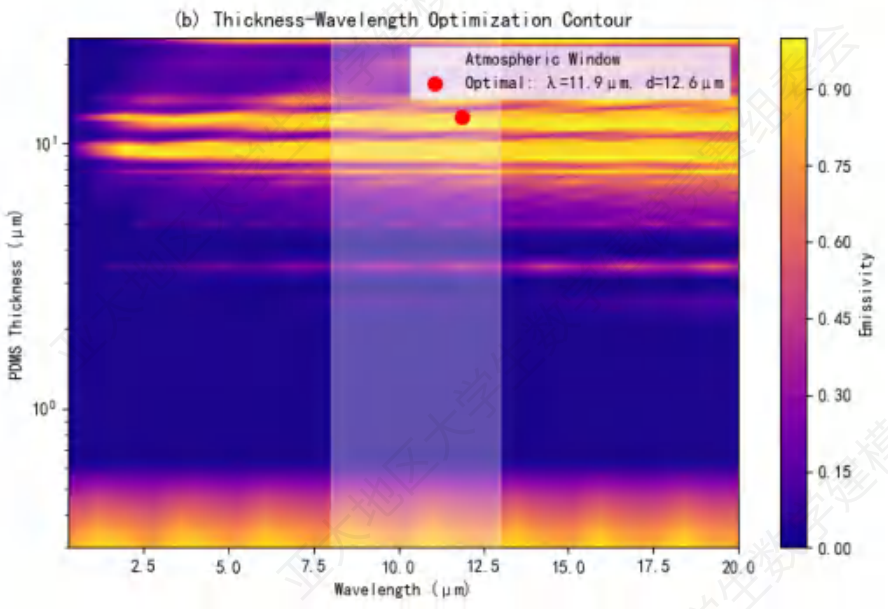


Figure 2 Two-dimensional colormap of emissivity as a function of wavelength and PDMS film thickness.

both wavelength and film thickness. In the solar spectrum band, emissivity remains consistently low across all thicknesses, reaffirming the high transparency of PDMS to solar radiation and its ability to avoid solar heat gain. Within the atmospheric window, emissivity increases markedly with film thickness, fully corroborating the trends observed in Figures 1 and 2. This thickness-dependent enhancement arises from increased optical path length and stronger interaction with vibrational resonances,

further underscoring the critical role of geometric design in tailoring thermal emission properties.

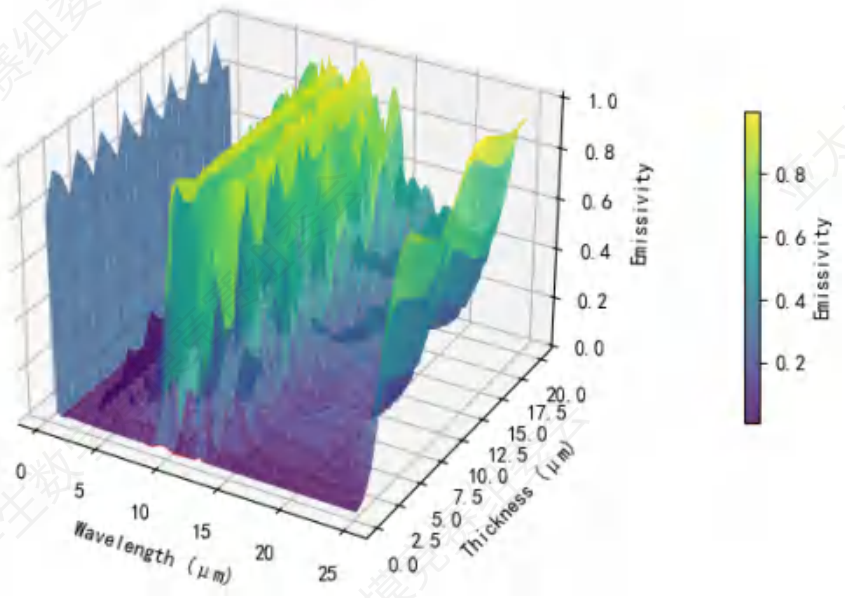


Figure 3 Three-dimensional surface plot of emissivity versus wavelength and film thickness.

5.1.4 *Model Rationality Verification*

To validate the physical correctness of the model's computational results, two auxiliary evaluation metrics are introduced.

1) Average emissivity in the solar band, denoted as $\bar{\varepsilon}_{\text{solar}}$, is theoretically expected to satisfy $\bar{\varepsilon}_{\text{solar}} < 0.07$. It is defined as

$$\bar{\varepsilon}_{\text{solar}} = \frac{1}{2.2} \int_{0.3}^{2.5} \varepsilon(\lambda)\, d\lambda, \tag{17}$$

where the integration limits (in μm) correspond to the solar radiation spectrum (0.3–2.5 μm).

2) Standard deviation of emissivity within the atmospheric window, denoted as σ_{atm}, is theoretically expected to satisfy $\sigma_{\text{atm}} < 0.2$. It is computed as

$$\sigma_{\text{atm}} = \sqrt{\frac{1}{N} \sum_{\lambda_i \in [8,13]} (\varepsilon(\lambda_i) - \bar{\varepsilon}_{\text{atm}})^2}, \tag{18}$$

where $\bar{\varepsilon}_{\text{atm}}$ is the mean emissivity over the atmospheric window (8–13 μm), and N is the number of sampled wavelengths in this interval.

The quantitative validation results of the model are presented in Table 1. Where $\bar{\varepsilon}_{\text{solar}}$ denotes the average emissivity in the solar band (0.3–2.5 μm), and σ_{atm} represents the standard deviation of emissivity in the atmospheric window (8–13 μm). The results

demonstrate that both values strictly satisfy the theoretical constraints ($\bar{\varepsilon}_{\text{solar}} < 0.07$ and $\sigma_{\text{atm}} < 0.2$), thereby verifying the physical correctness and numerical stability of the proposed model.

Table 1 Calculated Validation Metrics for Model Rationality

Metric	Value
$\bar{\varepsilon}_{\text{solar}}$	0.0389
σ_{atm}	0.1651

The results indicate that the calculated value of $\bar{\varepsilon}_{\text{solar}} = 0.0389$ remains well below the threshold of 0.07, consistent with PDMS's near-transparent physical properties in the solar band. Simultaneously, the computed $\sigma_{\text{atm}} = 0.1651$ is below the upper limit of 0.2, demonstrating smooth spectral emissivity curves within the atmospheric window without anomalous oscillations. This further validates the numerical stability and physical rationality of the model.

5.2 Establishment and Solution of Model for Problem 2

5.2.1 *Data Preprocessing for Solar and Atmospheric Data*

In addition to material properties, accurate solar spectral irradiance and atmospheric emission data are crucial for the simulation. The solar spectrum was obtained from the ASTM G173 standard, which provides detailed information on the solar irradiance at ground level under clear sky conditions. For atmospheric emission, we adopted a simplified model based on Planck's law with adjustments for typical atmospheric transmittance in different wavelength bands.

The specific steps taken for solar and atmospheric data processing include:

1. **Solar Spectral Irradiance:** Downloaded the AM1.5G spectrum provided by the ASTM G173 standard. Converted the spectral irradiance from units of W/m^2/nm to W/m^2/μm to align with our wavelength range of interest (0.3–25 μm). Linear interpolation was used to fill gaps between measured points and ensure a continuous spectrum.
2. **Atmospheric Emission:** Calculated the thermal radiation emitted by the atmosphere using Planck's law, assuming an average temperature of 300 K for the Earth's surface. Adjustments were made to account for the lower emissivity in the

atmospheric window (8–13 μm), where the transmission is higher due to reduced absorption by greenhouse gases.

3. **Data Integration:** Integrated the processed solar and atmospheric data into the same wavelength grid as the optical constants of materials. This step ensures that all datasets can be directly used in the transfer matrix method without further conversion.
4. **Normalization:** Normalized the solar spectral irradiance and atmospheric emission to facilitate comparison and analysis across different wavelengths. This normalization process helps in identifying key wavelengths that significantly affect the radiative cooling performance.

5.2.2 *Comprehensive Evaluation Function*

To objectively determine the relative importance of multiple performance metrics in radiative cooling design, we adopt the **Gray Relational Analysis (GRA)** method to dynamically assign weights to the key indicators. Unlike arbitrary or experience-based weighting, GRA quantifies the correlation between each candidate design and an ideal reference sequence based on their geometric proximity in normalized space, thereby yielding data-driven, adaptive weights.

Specifically, let $x_i^{(k)}$ denote the k-th performance metric ($k = 1, \ldots, K$) of the i-th design, normalized to the interval $[0, 1]$. The ideal reference sequence is defined as $\mathbf{x}_0 = [1, 1, \ldots, 1]$, representing the optimal value for all benefit-type criteria. The gray relational coefficient between design i and the ideal reference for metric k is computed as

$$\gamma_i^{(k)} = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_i^{(k)} + \rho\Delta_{\max}}, \tag{19}$$

where $\Delta_i^{(k)} = |x_0^{(k)} - x_i^{(k)}| = |1 - x_i^{(k)}|$ is the absolute deviation, $\Delta_{\min} = \min_i \min_k \Delta_i^{(k)}$, $\Delta_{\max} = \max_i \max_k \Delta_i^{(k)}$, and $\rho \in (0, 1]$ is a distinguishing coefficient (typically set to $\rho = 0.5$). The overall gray relational grade for design i is then obtained by averaging across all metrics:

$$\Gamma_i = \frac{1}{K} \sum_{k=1}^{K} \gamma_i^{(k)}. \tag{20}$$

This grade $\Gamma_i \in [0, 1]$ serves as the comprehensive score for ranking designs, with higher values indicating closer alignment to the ideal performance profile.

We select three core metrics that collectively characterize radiative cooling performance: Each of the three selected metrics captures a distinct physical dimension

of radiative cooling performance, and their combined consideration enables a holistic assessment:

- Net cooling power (P_{net}): system-level output (benefit-type);
- Atmospheric-window emissivity (ε_{8-13}): intrinsic radiative efficiency in the 8–13 μm band (benefit-type);
- Solar absorptivity (α_{sun}): susceptibility to solar heating (cost-type).

All metrics are first normalized to the interval [0, 1] using Min-Max scaling over the sampled design space (PDMS thickness: 1–150 μm). For benefit-type indicators (P_{net}, ε_{8-13}), normalization is defined as:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{21}$$

where $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values observed across all simulated thicknesses. For the cost-type indicator (α_{sun}), we invert the direction to align with a higher-is-better convention:

$$x^* = \frac{x_{\max} - x}{x_{\max} - x_{\min}}. \tag{22}$$

The gray relational grade for each metric is computed relative to the ideal reference sequence, comprising the best achievable value for each normalized metric. Using a distinguishing coefficient $\rho = 0.5$, the weight w_i for metric i is derived from its average gray relational grade, followed by normalization to ensure $\sum w_i = 1$.

The final comprehensive score is formulated as a weighted sum:

$$S_{\text{score}} = w_1 \cdot P^*_{\text{net}} + w_2 \cdot \varepsilon^*_{8-13} + w_3 \cdot \alpha^*_{\text{sun}}, \tag{23}$$

where the weights w_1, w_2, and w_3 are determined objectively via GRA. The physical definitions of the three core metrics are grounded in fundamental thermodynamic and radiative principles. This approach ensures that the evaluation function adapts to the actual trade-offs present in the design space, rather than imposing preconceived preferences.

Net Cooling Power (P_{net}). This is the net radiative heat flux density that the cooler can reject to outer space under ambient conditions. It accounts for all energy exchange channels:

$$P_{\text{net}}(T) = P_{\text{rad}}(T) - P_{\text{atm}}(T_{\text{amb}}) - P_{\text{sun}} - P_{\text{cond+conv}}, \tag{24}$$

where the net cooling power P_{net} is determined by the balance of four radiative and convective heat fluxes: The thermal radiation emitted by the PDMS film at its surface

temperature T is given by

$$P_{\text{rad}}(T) = \int_0^\infty I_{\text{BB}}(\lambda, T)\, \varepsilon(\lambda, d)\, d\lambda, \tag{25}$$

where $I_{\text{BB}}(\lambda, T)$ denotes the blackbody spectral radiance according to Planck's law. The downward atmospheric radiation absorbed by the film is expressed as

$$P_{\text{atm}}(T_{\text{amb}}) = \int_0^\infty I_{\text{BB}}(\lambda, T_{\text{amb}})\, \varepsilon_{\text{sky}}(\lambda)\, \varepsilon(\lambda, d)\, d\lambda, \tag{26}$$

where $\varepsilon_{\text{sky}}(\lambda)$ models the wavelength-dependent emissivity of the atmosphere. The solar power absorbed by the film is

$$P_{\text{sun}} = \int_0^\infty I_{\text{AM1.5}}(\lambda)\, \varepsilon(\lambda, d)\, d\lambda, \tag{27}$$

with $I_{\text{AM1.5}}(\lambda)$ representing the ASTM G173 standard solar spectral irradiance under AM1.5 conditions. Finally, parasitic heat gain due to conduction and convection is modeled as

$$P_{\text{cond+conv}} = h_c(T - T_{\text{amb}}), \tag{28}$$

where h_c is the combined heat transfer coefficient and T_{amb} is the ambient temperature. In our simulations, we evaluate P_{net} at thermal equilibrium ($T = T_{\text{amb}} = 300$ K), so $P_{\text{cond+conv}} = 0$.

Atmospheric-Window Emissivity (ε_{8-13}). This metric quantifies the intrinsic efficiency of the radiative cooling channel through the atmospheric transparency window (8–13 μm):

$$\varepsilon_{8-13} = \frac{1}{13-8} \int_8^{13} \varepsilon(\lambda, d)\, d\lambda. \tag{29}$$

A high ε_{8-13} ensures strong coupling between the emitter and the cold universe (3 K), enabling effective heat dissipation even under clear-sky conditions. This is a key indicator of spectral selectivity and is independent of environmental variables such as solar irradiance or wind speed.

Solar Absorptivity (α_{sun}). This parameter measures the fraction of incident solar energy (0.3–2.5 μm) that is absorbed by the cooler, acting as a parasitic heat load:

$$\alpha_{\text{sun}} = \frac{\int_{0.3}^{2.5} I_{\text{AM1.5}}(\lambda)\, \varepsilon(\lambda, d)\, d\lambda}{\int_{0.3}^{2.5} I_{\text{AM1.5}}(\lambda)\, d\lambda}. \tag{30}$$

Since $\varepsilon(\lambda, d) = \alpha(\lambda, d)$ by Kirchhoff's law of thermal radiation, this expression directly computes solar absorption from the emissivity spectrum. A low α_{sun} is critical for

daytime sub-ambient cooling, as it minimizes unwanted heating from sunlight. Unlike the previous two metrics, α_{sun} is a cost-type indicator—lower values correspond to better performance.

5.2.3 *Numerical Solution and Trend Analysis*

We performed numerical integration and comprehensive scoring calculations for PDMS films with thicknesses ranging from 1 μm to 150 μm, setting the ambient temperature at 300 K to simulate standard operating conditions. The resulting spectral response and power variations are illustrated in Figure 4.

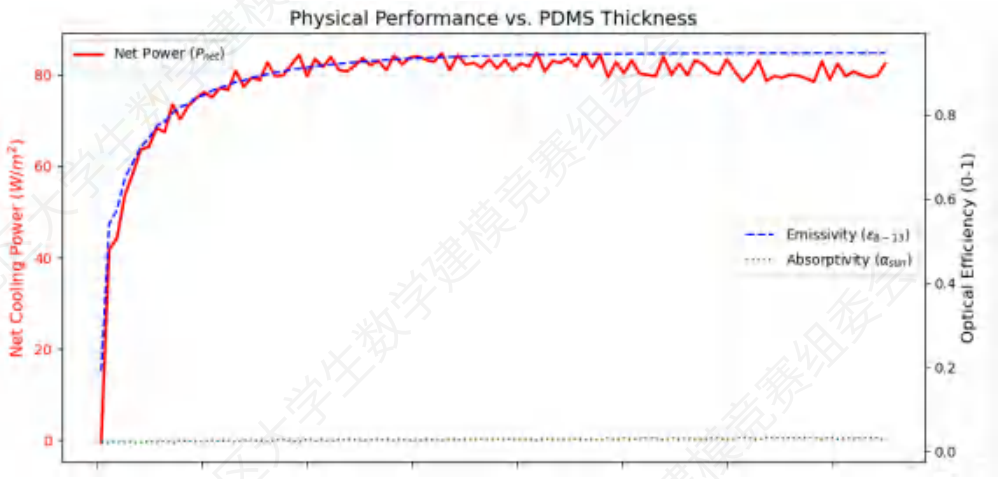


Figure 4 Spectral response and radiative cooling power as a function of PDMS film thickness (1–150 μm) under ambient temperature of 300 K.

Rapid Growth Phase. In the initial thin-film regime (0–20 μm), the radiative cooling performance exhibits a steep ascent. This phenomenon is governed by the increasing optical path length, which allows the PDMS layer to interact more effectively with infrared radiation. As predicted by the Beer-Lambert law, the absorbance-and consequently the emissivity-increases exponentially with thickness in this range. As a result, the net cooling power P_{net} surges from negligible levels to approximately 75 W/m^2, indicating that even a relatively thin layer of PDMS can establish a significant radiative cooling channel.

Saturation Phase. As the film thickness progresses into the intermediate range (20–60 μm), the rate of performance improvement decelerates markedly. In this region, the material becomes optically opaque in its primary characteristic emission bands, specifically the Si-O-Si and Si-C stretching modes. Consequently, the emissivity ε_{8-13} approaches its theoretical upper bound, entering a state of saturation. Further addition

of material yields diminishing returns in terms of radiative heat dissipation, as the "effective" emitting volume has largely been established.

Performance Trade-off. In the thick-film regime (>60 μm), a detrimental effect becomes prominent. While the radiative cooling power stabilizes and plateaus, the parasitic absorption of solar radiation (α_{sun}) continues to accumulate linearly or quasi-linearly. This indicates that excessively thick films not only fail to provide significant additional cooling benefits but also introduce unnecessary solar heat gain. This parasitic absorption undermines the overall thermal management efficiency, creating a divergence between the physical maximum of cooling power and the comprehensive efficiency of the system.

5.2.4 *Optimization Results and Discussion*

Table 2 Comparison of Optimization Results: Single-Objective (Max P_{net}) vs. Multi-Objective (Max Score).

Metric	Single-Obj.	Multi-Obj.
PDMS Thickness (μm)	83.8	38.6
P_{net} (W/m^2)	84.84	84.38
ε_{8-13}	0.944	0.911
α_{sun}	0.027	0.023
Score	91.8	95.4

Based on the trends identified above, the weighted scoring model was applied to identify a design that not only maintains high cooling performance but also achieves a superior balance among net cooling power, atmospheric-window emissivity, and solar reflectance. The weights assigned to these metrics were determined via GRA, which quantifies the correlation between each candidate design and an ideal reference sequence in normalized space. The resulting data-driven weights are: $w_{P_1} = 0.484$, $w_{\varepsilon_2} = 0.333$, and $w_{\alpha_3} = 0.183$. This approach ensures an objective, adaptive prioritization of performance criteria, leading to a multi-objective optimal solution with a higher comprehensive score (95.4) than the single-objective maximum (91.8).

Determination of Optimal Thickness. By analyzing the comprehensive evaluation scores, we identified a global maximum at a thickness of approximately 38.6 μm.

This value represents the precise equilibrium point where the marginal gain in cooling power is outweighed by the penalties of increased solar absorption and material consumption. At this optimal thickness, the system achieves a net cooling power of 84.38 W/m^2, demonstrating robust radiative cooling capabilities suitable for practical applications while attaining the highest multi-objective score of 95.4.

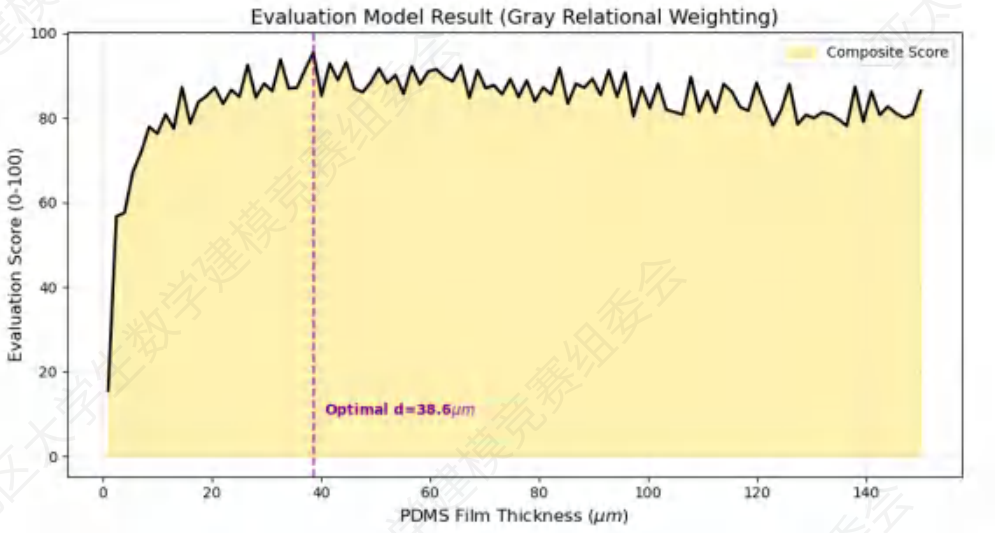


Figure 5 Variation of the composite evaluation score (0–100) as a function of PDMS film thickness. The optimal thickness of 38.6 μm corresponds to the peak score, indicating the best trade-off between radiative cooling performance and solar reflectance.

Comparative Performance Analysis. A critical comparison between this optimal point and the theoretical physical maximum reveals the efficacy of the optimization strategy. While the absolute maximum cooling power of 84.84 W/m^2 is attainable at a much greater thickness of 83.8 μm, the multi-objective solution at 38.6 μm incurs a performance penalty of less than 0.6%. However, this negligible sacrifice in cooling power yields significant practical advantages: the PDMS layer thickness—and consequently the primary material consumption—is reduced by over 50%, while solar absorptivity is further suppressed from 0.027 to 0.023, enhancing daytime cooling stability.

5.2.5 *Recommendations for Advancement and Application*

Based on the evaluation model results, we propose the following three recommendations for the development of radiative cooling technology.

Optimal Thickness Control: In practical manufacturing, the thickness of PDMS films should be controlled within the range of 20-60 μm. This range ensures over 90%

of the maximum theoretical cooling potential while minimizing raw material costs and reducing thermal resistance.

Composite Material Modification: Our analysis shows that pure PDMS still has weak solar absorption. We recommend doping the PDMS matrix with high-refractive-index nanoparticles. These particles can enhance solar scattering without affecting infrared transparency, thereby improving daytime cooling performance.

Structural Design Optimization: Single-layer films have physical limitations. Future applications should adopt multilayer structures or photonic crystal structures. For example, adding a specialized solar reflection layer below the PDMS layer can effectively block the residual sunlight that penetrates the PDMS, further enhancing the net cooling power.

5.3 Establishment and Solution of Model for Problem 3

5.3.1 *Data Preprocessing*

To ensure the accuracy and reliability of the proposed model, a systematic preprocessing of the optical constant data for all involved materials was conducted. The dataset includes the complex refractive index of PDMS over the visible–near-infrared and mid-infrared spectral ranges, as well as the optical constants of aluminum (Al), silver (Ag), silicon dioxide (SiO_2), titanium dioxide (TiO_2), and aluminum oxide (Al_2O_3). By consulting relevant literature and performing necessary data cleaning procedures, the unit-thickness cost of each material was obtained.

Table 3 Unit thickness cost (USD/μm) of materials used in the model.

Material	Unit Thickness Cost (USD/μm)
PDMS	0.10–0.50
Al	0.50–2.00
Ag	5.00–15.00
SiO_2	1.50–5.00
TiO_2	3.00–8.00
Al_2O_3	20.00–100.00

The data preprocessing workflow is summarized as follows:

1. **Data Merging and Alignment:** The two spectral segments of PDMS were concatenated according to wavelength, and duplicate or conflicting data points were removed.
2. **Outlier and Missing Value Handling:** The optical constant datasets of aluminum, silicon dioxide, and titanium dioxide were cleaned by eliminating invalid entries and retaining only the optical constants within their valid wavelength ranges.
3. **Construction of Interpolation Functions:** To enable the retrieval of refractive indices and extinction coefficients at arbitrary wavelengths, linear interpolation was applied to construct continuous functions for each material. For queries beyond the original wavelength range, extrapolation was enabled to prevent computation interruptions. Although extrapolation introduces some uncertainty, the optical behavior of PDMS near the starting wavelength of the solar spectrum is well understood (with approximately $k \approx 0$ and $n \approx 1.4$), ensuring that any extrapolation error remains controlled.
4. **Physical Consistency Verification:** Non-negativity constraints were imposed on all interpolated results to prevent nonphysical negative values arising from numerical errors.

5.3.2 *Establishment of the Optimization Model for the PDMS Multilayer Structure*

Model Principle. As the core radiative cooling material, PDMS exhibits intrinsically high emissivity within the atmospheric window. However, its transparency in the solar spectrum allows solar radiation to reach the underlying reflective layer, resulting in undesired absorption and thus reducing cooling efficiency. To address this issue, a dielectric layer is introduced as an impedance-matching layer. By leveraging its optical interference effects, the dielectric layer suppresses reflection in the solar spectrum, thereby reducing solar absorption, while simultaneously maintaining resonant enhancement within the atmospheric window to increase infrared emissivity. The bottom metallic layer, featuring high reflectivity, further reflects the transmitted infrared radiation, forming an effective radiative pathway. Accordingly, the multilayer structure investigated in this work is composed of four layers: air / PDMS / dielectric layer / reflective layer.

To optimize the performance of this multilayer configuration, the multi-objective weighted scoring model developed in Problem 2 is adopted as the evaluation framework. By incorporating the unit-thickness cost of each layer as a penalty term, a multi-objective optimization function is constructed that jointly considers optical performance and

manufacturing cost. Through dynamically adjusting the cost-penalty weighting, the model systematically balances the trade-off between radiative cooling efficiency and economic feasibility, providing a set of Pareto-optimal solutions for practical engineering applications.

The optimization algorithm employed in this study is the adaptive differential evolution algorithm, a population-based global optimization method. Its core mechanism lies in dynamically adjusting the mutation factor and crossover probability to balance exploration and exploitation capabilities. By simulating biological evolution processes—mutation, crossover, and selection—the algorithm conducts parallel searches for the global optimum within the solution space. Compared with conventional differential evolution, its adaptive variant automatically tunes the search strategy according to population diversity: maintaining a relatively large mutation factor in the early stage to enhance global exploration and avoid premature convergence, and reducing the mutation factor in the later stage to strengthen local refinement and improve convergence accuracy.

Model Structure. The optimization model for the PDMS multilayer structure consists of the following core components. The model adopts the comprehensive performance evaluation framework established in Section 5.2, where the performance score S_{score} is computed using the weighted scoring function defined in Equation (23), with all metrics normalized to $[0, 1]$ and weights satisfying $w_1 + w_2 + w_3 = 1$.

Cost Quantification Module: A unit-thickness cost model is constructed based on material market prices and physical properties, yielding the total material cost

$$C_{\text{mat}} = \sum_i (c_i \cdot d_i), \tag{31}$$

where c_i denotes the unit thickness cost of material i (in USD/μm $\cdot$ m^2) and d_i is the corresponding layer thickness (μm).

Multi-objective Optimization Function: By introducing a cost-penalty coefficient λ_{cost}, the objective function is formulated as

$$\text{Maximize} \quad \text{Objective} = S_{\text{score}} - \lambda_{\text{cost}} \cdot C_{\text{mat}}, \tag{32}$$

where S_{score} is scaled to the 0–100 range. The Pareto front under different values of λ_{cost} is obtained via global optimization.

Optimization Algorithm. An adaptive differential evolution (DE) algorithm is employed to simultaneously optimize discrete variables and continuous variables, ensuring

robust exploration of the design space. The optimization proceeds as follows. Let the design vector be

$$\mathbf{X} = [x_{\text{diel}},\, x_{\text{refl}},\, d_{\text{PDMS}},\, d_{\text{diel}},\, d_{\text{refl}}]^{\top}, \tag{33}$$

where $x_{\text{diel}}, x_{\text{refl}} \in [0, 1]$ are continuous surrogates for discrete material choices, mapped via

$$\text{mat}_{\text{diel}} = \mathcal{M}_{\text{diel}}\big(\text{round}\big(x_{\text{diel}} \cdot (N_{\text{diel}} - 1)\big)\big), \tag{34}$$

$$\text{mat}_{\text{refl}} = \mathcal{M}_{\text{refl}}\big(\text{round}\big(x_{\text{refl}} \cdot (N_{\text{refl}} - 1)\big)\big), \tag{35}$$

with $\mathcal{M}_{\text{diel}}$ and $\mathcal{M}_{\text{refl}}$ denoting the dielectric and reflector material libraries, and N_{diel}, N_{refl} their cardinalities.

Given a population $\{\mathbf{X}_i^{(g)}\}_{i=1}^{N_P}$ at generation g, the DE generates a mutant vector $\mathbf{V}_i^{(g)}$ for each individual $\mathbf{X}_i^{(g)}$ by:

$$\mathbf{V}_i^{(g)} = \mathbf{X}_{r_1}^{(g)} + F \cdot \left(\mathbf{X}_{r_2}^{(g)} - \mathbf{X}_{r_3}^{(g)}\right), \tag{36}$$

where $r_1, r_2, r_3 \in \{1, \ldots, N_P\}$ are distinct random indices different from i, and $F \in [0, 2]$ is the scaling factor.

A trial vector $\mathbf{U}_i^{(g)} = [u_{i1}, \ldots, u_{i5}]^{\top}$ is then formed via binomial crossover:

$$u_{ij}^{(g)} = \begin{cases} v_{ij}^{(g)}, & \text{if } \text{rand}_j \le CR \text{ or } j = j_{\text{rand}}, \\ x_{ij}^{(g)}, & \text{otherwise}, \end{cases} \tag{37}$$

for $j = 1, \ldots, 5$, where $\text{rand}_j \sim \mathcal{U}(0, 1)$, $CR \in [0, 1]$ is the crossover rate, and $j_{\text{rand}} \in \{1, \ldots, 5\}$ is a randomly chosen index to ensure $\mathbf{U}_i^{(g)} \neq \mathbf{X}_i^{(g)}$.

The objective function incorporates a cost-penalized performance score:

$$\mathcal{L}(\mathbf{X}; \lambda_{\text{cost}}) = -\Big(S_{\text{score}}(\mathbf{X}) - \lambda_{\text{cost}} \cdot C_{\text{mat}}(\mathbf{X})\Big), \tag{38}$$

where S_{score} is the weighted normalized performance metric (see Eq. (23)), $C_{\text{mat}} = \sum_k c_k d_k$ is the total material cost, and $\lambda_{\text{cost}} \ge 0$ controls the cost-performance trade-off.

Finally, selection is performed greedily:

$$\mathbf{X}_i^{(g+1)} = \begin{cases} \mathbf{U}_i^{(g)}, & \text{if } \mathcal{L}(\mathbf{U}_i^{(g)}; \lambda_{\text{cost}}) < \mathcal{L}(\mathbf{X}_i^{(g)}; \lambda_{\text{cost}}), \\ \mathbf{X}_i^{(g)}, & \text{otherwise}. \end{cases} \tag{39}$$

This process iterates until convergence, yielding Pareto-optimal designs across varying λ_{cost}.

5.3.3 *Model Solution and Result Analysis*

Solution Produce

1. An accurate optical model was established based on the transfer matrix method, in which the spectral response of the multilayer film structure was calculated by solving Maxwell's equations.
2. The optical constants of the six candidate materials were imported from the original datasets and preprocessed to construct wavelength-dependent interpolation functions for their complex refractive indices.
3. The comprehensive performance model developed in Problem 2 was combined with the material cost weighting to generate the objective function, forming a performance–cost trade-off optimization framework.
4. A hybrid integer–continuous encoding scheme was adopted to simultaneously optimize material selection (discrete variables) and layer thicknesses (continuous variables).
5. A five-dimensional search space was defined, and a global optimization approach was used to prevent convergence to local optima.
6. By scanning across different cost-weight parameters, the optimal trade-off relationship between performance and manufacturing cost was systematically investigated.
7. The adaptive differential evolution algorithm was employed to solve the aforementioned multi-objective optimization function.

Result Analysis

1. The optimized spectral response is shown in Figure 6. The PDMS/SiO_2/Ag trilayer structure successfully achieves the spectral selectivity required for radiative cooling: extremely low emissivity in the solar spectrum (minimizing heat gain) and high emissivity within the 8–13 μm atmospheric window (enabling efficient radiative cooling). An enhanced emissivity peak in the far-infrared region further improves overall performance.

   The optimal structure is a three-layer configuration with the following thicknesses:

2. The Pareto frontier analysis results are summarized in Table 5. The PDMS/SiO_2/Ag combination consistently outperforms others across all cost-weighting scenarios. When cost is ignored ($\lambda_{\text{cost}} = 0$), the system achieves the highest score (78.21) at a PDMS thickness of 19.41 μm, but with a high cost of 23.17. As λ_{cost} increases, the algorithm reduces PDMS thickness to balance cost and performance. At

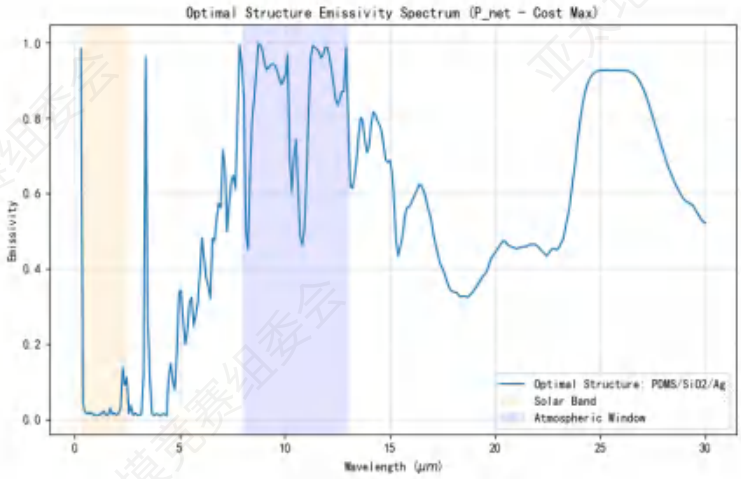


Figure 6 Optimized spectral emissivity of the PDMS/SiO_2/Ag trilayer structure.

Table 4 Optimal layer thicknesses and performance metrics.

Layer	Thickness (μm)	
PDMS	19.206	Net cooling power: 79.49 W/m^2
SiO_2	0.010	Solar absorptivity: 0.019
Ag	0.120	

$\lambda_{\text{cost}} = 3.57$, an optimal trade-off is achieved (score = 67.95, cost = 6.84, $d_{\text{PDMS}} = 19.21\,\mu$m). These results confirm that performance is primarily driven by PDMS thickness, while the SiO_2/Ag stack provides stable optical synergy. Figure 7 visualizes these trends.

3. Figure 8 presents a scatter plot of PDMS thickness versus performance and cost. PDMS thickness is positively correlated with performance, but total cost rises sharply with thickness, highlighting the necessity of balancing performance and economics.
4. Figure 9 shows the marginal performance efficiency (performance gain per additional dollar) for the PDMS/SiO_2/Ag structure. Initially, each dollar yields significant performance improvement, but marginal returns diminish rapidly. At $\lambda_{\text{cost}} = 50$, the initial design achieves optimal cost-effectiveness; beyond this point, further investment leads to sharply declining returns.

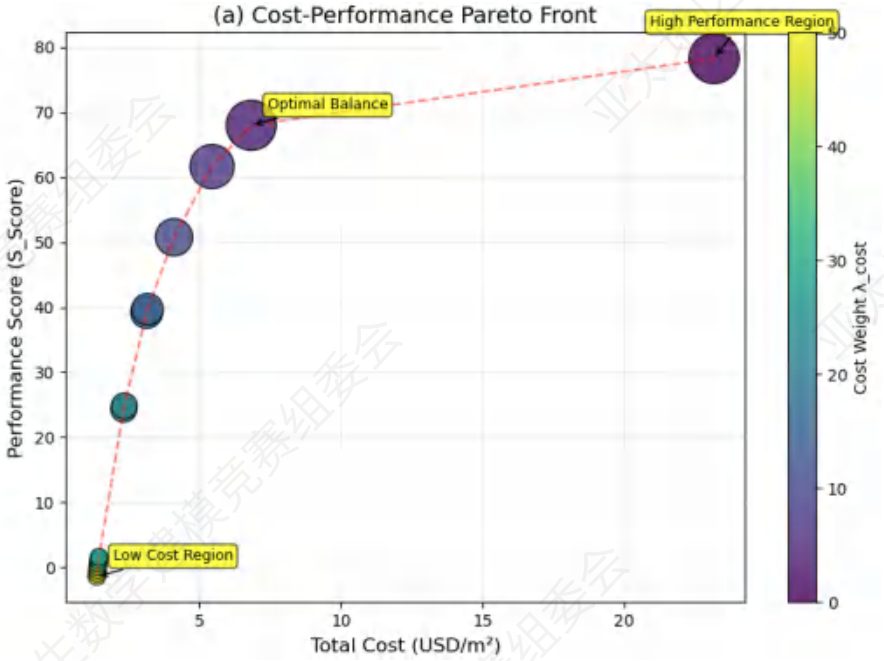


Figure 7 Pareto frontier showing the trade-off between performance score and total cost.

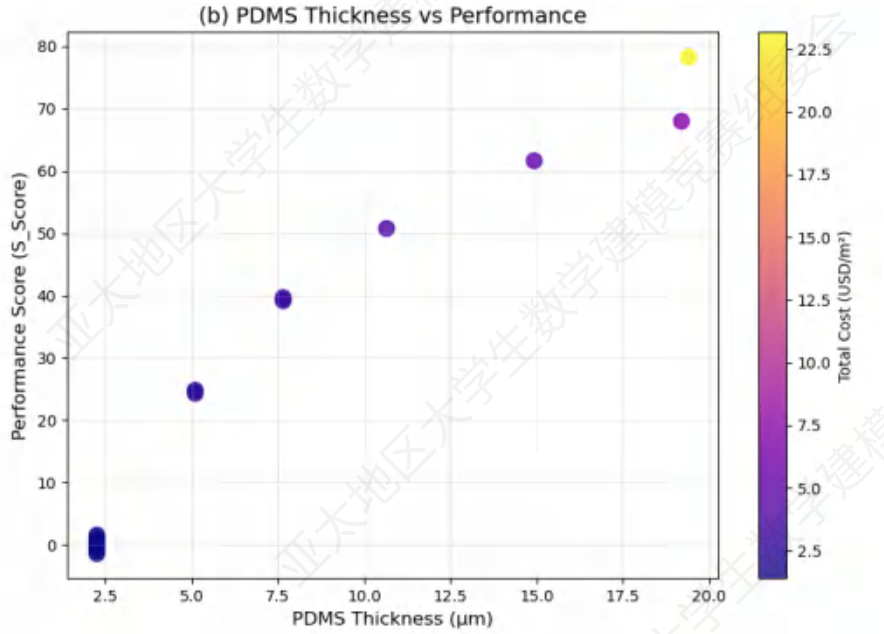


Figure 8 Relationship among PDMS thickness, performance score, and total cost.

5.4 Establishment and Solution of Model for Problem 4

5.4.1 *Design Strategy*

Previous sections focused on maximizing the theoretical radiative cooling power. However, bridging the gap between laboratory research and large-scale industrial application requires addressing critical engineering challenges: material cost, environmental dura-

Table 5 Pareto-optimal solutions under varying cost weights.

λ_{cost}	S_{score}	Cost	Structure	d_{PDMS} (μm)
50.00	-1.38	1.39	SiO_2/Ag	2.26
46.43	-0.88	1.40	SiO_2/Ag	2.26
42.86	-0.39	1.41	SiO_2/Ag	2.26
39.29	0.11	1.42	SiO_2/Ag	2.26
35.71	0.60	1.44	SiO_2/Ag	2.26
32.14	1.10	1.45	SiO_2/Ag	2.26
28.57	1.59	1.47	SiO_2/Ag	2.26
25.00	24.32	2.34	SiO_2/Ag	5.11
21.43	24.82	2.36	SiO_2/Ag	5.11
17.86	39.18	3.15	SiO_2/Ag	7.66
14.29	39.68	3.18	SiO_2/Ag	7.66
10.71	50.75	4.12	SiO_2/Ag	10.66
7.14	61.61	5.46	SiO_2/Ag	14.93
3.57	67.95	6.84	SiO_2/Ag	19.21
0.00	78.21	23.17	SiO_2/Ag	19.41

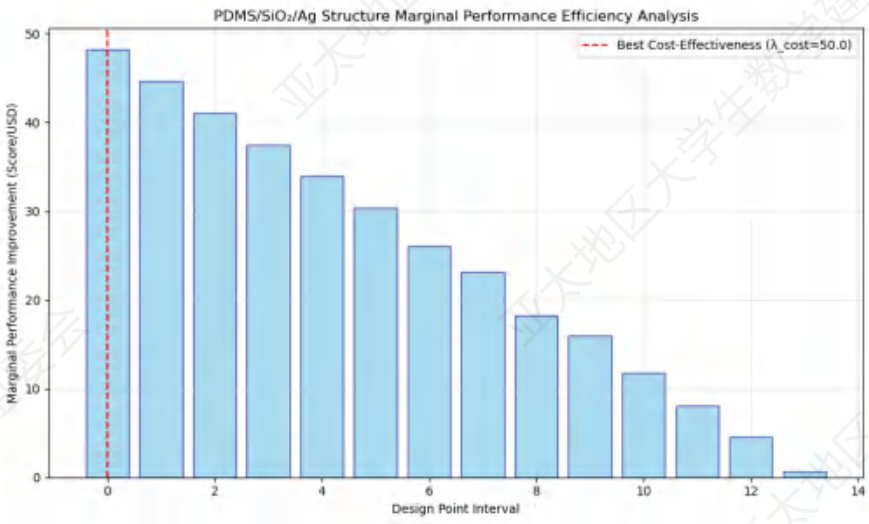


Figure 9 Marginal performance efficiency versus cost weight.

bility, and manufacturing feasibility.

In this section, we propose a robust design framework based on the **Porous PDMS / SiO_2 / Ag** multilayer structure derived from Problem 3. To further enhance performance and reliability, we introduce two engineering modifications:

- Microstructural Enhancement: Introducing air voids into the PDMS matrix to tune the refractive index and enhance optical matching.
- Reliability Enhancement: Adding an ultra-thin Al_2O_3 protection layer to prevent the oxidation and sulfidation of the Ag reflector.

5.4.2 *Establishment of the Model*

Maxwell-Garnett Effective Medium Theory (MG-EMT). To simulate the optical properties of porous PDMS without the prohibitive computational cost of full-wave 3D simulations, we employ the MG-EMT.

EMT homogenizes the complex heterogeneous microstructure (PDMS host with air inclusions) into a material with an effective complex permittivity ε_{eff}. The defining equation is:

$$\frac{\varepsilon_{\text{eff}} - \varepsilon_{\text{host}}}{\varepsilon_{\text{eff}} + 2\varepsilon_{\text{host}}} = \phi \frac{\varepsilon_{\text{air}} - \varepsilon_{\text{host}}}{\varepsilon_{\text{air}} + 2\varepsilon_{\text{host}}} \tag{40}$$

where $\varepsilon_{\text{host}} = (n_{\text{PDMS}} + ik_{\text{PDMS}})^2$ denotes the complex permittivity of pure PDMS, $\varepsilon_{\text{air}} \approx 1$ represents the permittivity of the air voids, and ϕ is the porosity—defined as the volume fraction of air—which serves as a new continuous optimization variable in the design of porous PDMS structures. The effective refractive index is then derived as $\tilde{n}_{\text{eff}} = \sqrt{\varepsilon_{\text{eff}}}$.

The Effective Medium Theory (EMT) is adopted for two key reasons. First, it dramatically enhances computational efficiency by representing complex microstructures as homogeneous materials with effective optical constants, enabling seamless integration with the Transfer Matrix Method (TMM) and facilitating rapid evaluation of thousands of design candidates in global optimization frameworks such as Differential Evolution. Second, EMT offers clear physical insight by establishing an analytical relationship between porosity and effective refractive index, thereby elucidating how tailored porosity improves impedance matching and suppresses surface reflection.

Lifecycle Cost Analysis (LCA) Mode. To evaluate economic feasibility, we establish a Total Cost of Ownership (TCO) model that accounts for material, processing, and maintenance costs over a 10-year lifespan.

$$\text{TCO} = C_{\text{mat}} + C_{\text{proc}} + C_{\text{maint}} \tag{41}$$

where the total cost is decomposed into three components: $C_{\text{mat}} = \sum(c_i \cdot d_i)$ represents the direct material cost, calculated as the sum of the unit cost c_i multiplied by the thickness d_i of each layer; C_{proc} denotes the manufacturing process cost, which depends on a binary decision variable $x_{\text{prot}}$when $x_{\text{prot}} = 1$, additional costs associated with Atomic Layer Deposition (ALD) are incurred; and C_{maint} accounts for maintenance expenses, modeled as a function of the degradation rate of the Ag layer.

$$C_{\text{maint}} = (C_{\text{mat}} + C_{\text{proc}}) \cdot M_{\text{factor}} \cdot \text{Lifespan} \tag{42}$$

Here, M_{factor} represents the annual maintenance ratio. Without protection ($x_{\text{prot}} = 0$), exposed Ag degrades rapidly ($M_{\text{factor}} = 0.20$). With an Al_2O_3 barrier ($x_{\text{prot}} = 1$), degradation is suppressed ($M_{\text{factor}} = 0.02$), significantly reducing long-term OPEX.

Grey Relational Analysis (GRA) We aim to maximize a comprehensive Final Score that balances robust cooling performance $P_{\text{net}}^{\min}$ under extreme conditions and economic cost TCO. To avoid subjective weighting, we utilize GRA.

The GRA Grade represents the geometric similarity between the candidate design and an Ideal Reference Sequence (High Power, Low Cost). The optimization objective is:

$$\text{Maximize } S_{\text{GRA}} = \text{GRA}(P_{\text{net}}^{\min}, \text{TCO}) \tag{43}$$

Where $P_{\text{net}}^{\min}$ is the worst-case cooling power across temperate, tropical, and desert climate scenarios, ensuring environmental robustness.

Differential Evolution (DE) algorithm. This approach is identical in principle to that employed in Question 3: it evolves a diverse set of candidate solutions through mutation, crossover, and selection operations over successive generations, without requiring gradient information. DE efficiently explores the trade-off space between $P_{\text{net}}^{\min}$ and TCO, converging to a Pareto-relevant design that maximizes the GRA-based final score while respecting physical and practical constraints.

5.4.3 *Optimization Results and Discussion*

The comprehensive design optimization was executed using a mixed-integer Differential Evolution algorithm, incorporating both continuous variables and discrete decisions. The algorithm successfully converged to a global optimum that balances thermodynamic performance with economic constraints. The detailed configuration and performance metrics are summarized in Table 6.

Microstructural and Spectral Analysis. The optimal configuration features a relatively thin PDMS layer (24.59 μm) with a porosity of 11.9%, challenging the

Table 6 Optimal Design Parameters and Performance Metrics

Design Parameters	**Value**	**Evaluation Metrics**	**Value**
Structure	Porous PDMS/SiO_2/Ag	Robust P_{net} (min)	57.37 W/m^2
PDMS Thickness (d)	24.59 μm	Solar Absorptivity (α_{solar})	0.025
PDMS Porosity (ϕ)	11.9%	Lifecycle Cost (TCO)	\$14.95 / m^2
Protection Layer	**Enabled** (Al_2O_3)	**Final GRA Score**	**74.91**

conventional intuition that "thicker is better" for radiative cooling. As illustrated in Figure 10, the spectral emissivity profile reveals the underlying optical mechanism. The introduction of air voids lowers the effective refractive index of the PDMS matrix, thereby improving impedance matching at the air-material interface and minimizing surface reflection. Furthermore, the spectrum exhibits distinct Fabry-Perot resonance peaks, indicating that the design leverages thin-film interference effects to maximize emission within the 8–13 μm atmospheric window. Crucially, the film maintains near-zero emissivity in the 0.3–2.5 μm solar band, confirming that the thin, porous structure effectively minimizes parasitic solar absorption, which is often a bottleneck for thicker bulk materials.

Economic Feasibility and Lifecycle Strategy. Beyond physical performance, the economic viability of the design is validated through the Total Cost of Ownership (TCO) model. Figure 11 presents the breakdown of the lifecycle costs. A key strategic decision made by the optimizer was the inclusion of the Al_2O_3 protection layer. Although this addition increases the initial manufacturing complexity and process cost, it yields a disproportionate benefit in durability. The protection layer suppresses the degradation of the Ag reflector, reducing the annual maintenance ratio from 20% to 2%. Consequently, the long-term maintenance expenditure is drastically curtailed, resulting in a highly competitive final TCO of \$14.95/m^2. This result underscores the necessity of a lifecycle perspective; a design that appears cheaper initially due to simplified processing would ultimately prove more costly due to rapid failure rates.

Global Optimality and Sensitivity Analysis. To elucidate why the algorithm converged to 24.6 μm, we performed a sensitivity analysis of the Grey Relational Analysis (GRA) score, as depicted in Figure 12. The trends reveal a clear trade-off mechanism. The cooling power P_{net} increases rapidly in the thin-film regime but saturates near ~20 μm. Beyond this point, the infrared emission gains diminish, while

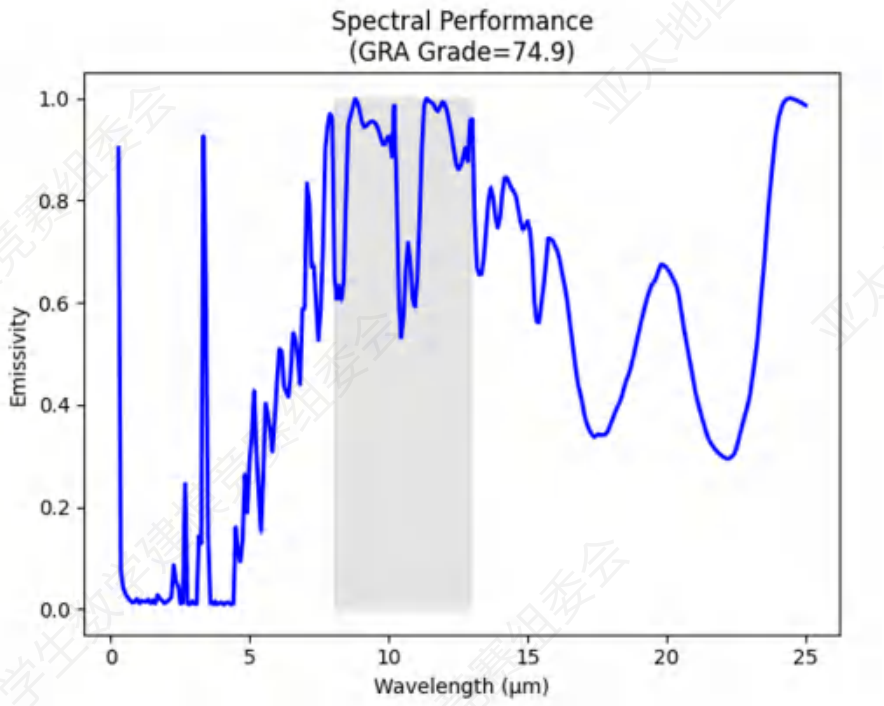


Figure 10 Spectral emissivity of the optimized porous structure, demonstrating high selectivity in the atmospheric window and low solar absorption.

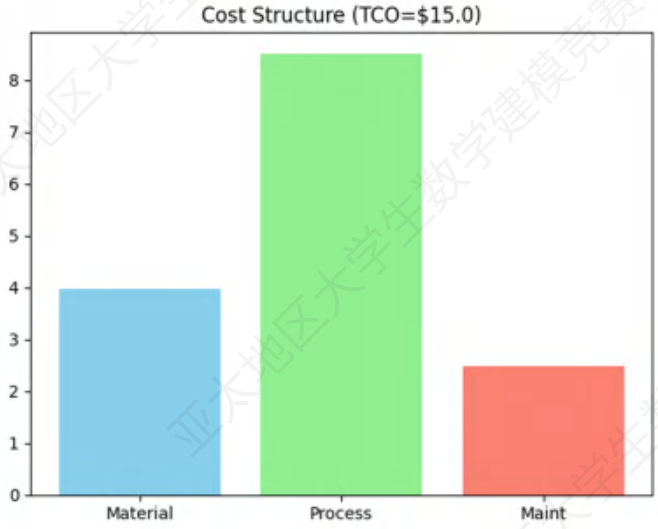


Figure 11 Lifecycle cost breakdown showing the strategic trade-off between upfront process investment and long-term maintenance savings.

parasitic solar absorption continues to accumulate, causing performance to fluctuate or even decline. Conversely, the TCO increases linearly with thickness. The GRA score, which geometrically aggregates these factors, identifies a sharp global maximum at 24.6 μm. This point represents the precise ”sweet spot” where the marginal gain in cooling performance no longer justifies the marginal increase in material cost and solar

absorption risk, confirming the robustness of the optimization result.

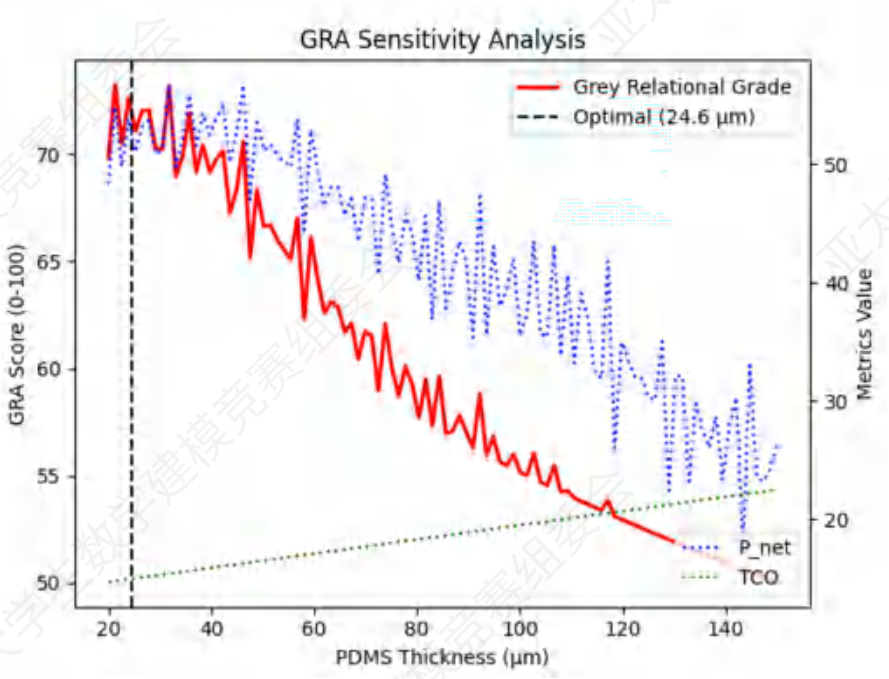


Figure 12 Sensitivity analysis of GRA score versus PDMS thickness. The optimal point marks the trade-off between performance saturation and linear cost increase.

VI. Model evaluation

6.1 Advantages of the model

6.1.1 *Accurate Calculation Results*

The Transfer Matrix Method (TMM) enables accurate computation of the spectral reflectance of multilayer films, from which wavelength-dependent emissivity can be derived. Model validation (e.g., solar-band average emissivity of 0.0389 and atmospheric-window standard deviation of 0.1651) aligns well with physical expectations, demonstrating reliability.

6.1.2 *Comprehensive and Holistic Evaluation*

Grey Relational Analysis (GRA) enables multi-objective evaluation by weighting net cooling power, atmospheric-window emissivity, and solar absorptivity. Unlike fixed-weight methods, GRA dynamically determines the relative importance of each indicator, allowing the model to adaptively emphasize the most critical factors, such as net cooling

power in radiative cooling optimization. This method performs well even with small sample sizes or incomplete data.

6.1.3 *Integration of Microstructural Modeling and Multi-Objective Optimization*

MG-EMT (Maxwell-Garnett Effective Medium Theory) combined with TMM facilitates accurate modeling of porous PDMS microstructures, and coupled with lifecycle cost (LCA) considerations, enables high-performance, cost-effective design.

6.2 Disadvantages of the model

6.2.1 *High Computational Complexity*

The combination of TMM, MG-EMT, and DE optimization increases computational cost, especially when considering multiple parameters such as thickness, porosity, and material types.

6.2.2 *Idealized Assumptions Do Not Match Practical Applications*

This model primarily considers standard atmospheric conditions and an ideal solar spectrum, and may not fully reflect cooling performance under different climatic conditions. Furthermore, MG-EMT homogenizes the porous structure, potentially neglecting the effects of pore shape and non-uniform distribution. TMM assumes smooth interfaces and uniform thickness, which may not fully reflect deviations in actual manufacturing processes.

6.2.3 *Practical Fabrication Constraints*

Optimized structures may be challenging to fabricate in practice, reducing practical applicability of the model results.

VII. References

[1] Macleod, H. A., Thin-film optical filters, Boca Raton: CRC Press, 2010.

[2] Storn, R. and Price, K., Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, Journal of Global Optimization, 11: 341–359, 1997.

[3] Julong, D., Introduction to grey system theory, The Journal of Grey System, 1: 1–24, 1989.

[4] Maxwell, G. J. C. and Garnett, B. A., Colours in metal glasses and in metallic films, Philosophical Transactions of the Royal Society of London. Series A, 203: 385–420, 1904.

[5] Luce, A. and Mahdavi, A. and Marquardt, F. and et al., TMM-Fast, a transfer matrix computation package for multilayer thin-film optimization: tutorial, Journal of the Optical Society of America A, 39: 1007–1013, 2022.

[6] Li, S. and An, M. and Zheng, Z. and et al., Daytime radiative cooling multilayer films designed by a machine learning method and genetic algorithm, Applied Optics, 62: 4359–4369, 2023.

[7] Zhan, Y. and Yin, H. and Wang, J. and et al., Enhanced performance of diurnal radiative cooling for solar cells based on a grating-textured PDMS photonic structure, Materials Today Communications, 35: 106117, 2023.

[8] Xie, A. Q. and Qiu, H. and Jiang, W. and et al., Recent Advances in Spectrally Selective Daytime Radiative Cooling Materials, Nano-Micro Letters, 17: 264, 2025.

[9] Liu, G. and Chen, S. and Lin, C., Optimization of dielectric-metal multilayer structure for color-preserving radiative cooling window, ACS omega, 9: 30425–30435, 2024.

[10] Jung, J. and Yoon, S. and Kim, B. and et al., Development of high-performance flexible radiative cooling film using PDMS/TiO2 Microparticles, Micromachines, 14: 2223, 2023.

[11] Herrmann, Kai and Lauster, Tobias and Song, Qimeng and Retsch, Markus, Homogeneous polymer films for passive daytime cooling: optimized thickness for maximized cooling performance, Advanced Energy and Sustainability Research, 3: 2100166, 2022.

Team Number:	apmcm25301412
Problem Chosen:	C

2025 APMCM summary sheet

The implementation of the "Reciprocal Tariff" policy marks a paradigm shift from efficiency-driven globalization to security-driven fragmentation. To quantify the cascading effects of this policy, this paper constructs a **Systemic Multi-Dimensional Evaluation Framework (SMDEF)** that integrates econometrics, game theory, and system dynamics across agricultural, industrial, and macroeconomic layers.

Micro-Level Impacts: Supply Chain Restructuring

- **Agriculture (Irreversible Displacement, Q1):** Utilizing a **Constrained Armington Model**, we reveal a structural break. A 25% tariff triggers a **64.1% collapse** in U.S. soybean exports to China. Crucially, our model captures Brazil's physical capacity saturation, identifying **Argentina** as the unexpected strategic beneficiary with a **245% export surge**.
- **Automotive (Forced Localization, Q2):** Through a **Dynamic Stackelberg Game**, we identify a precise **Tariff Tipping Point of 11.1%**. Beyond this threshold, the cost of direct import exceeds local production, forcing Japanese automakers into a **"Forced Localization"** strategy, effectively trading consumer surplus for U.S. manufacturing jobs.
- **Semiconductors (Innovation Dilemma, Q3):** Our **Bi-level Policy Simulation** demonstrates that tariffs alone fail to bridge the **40% manufacturing cost gap**. Subsidies are a necessary condition for reshoring. Moreover, aggressive export controls create a "Boomerang Effect", stripping U.S. firms of 30% of revenue and eroding long-term innovation budgets.

Macro-Level Insights: Fiscal Limits & Systemic Risk

- **The Fiscal Paradox (Q4):** We debunk the myth of linear revenue growth. Our **Crisis-Aware Forecasting** confirms a Laffer Curve effect, pinpointing the **Optimal Tariff Rate at ≈30%**. The proposed 60% rate lies in the "Over-taxation Zone", accelerating tax base erosion and trade diversion (leakage to Vietnam/Mexico).
- **The Stagflation Trap (Q5):** The **Supply-Side Shock Model** proves that reciprocal tariffs, when met with targeted retaliation (e.g., Critical Minerals Ban), trigger a Stagflationary Trap. The surge in input costs and interest rates neutralizes the protective benefits of tariffs, rendering the "Reshoring" goal economically unfeasible.

The Synthesis of the Systemic Framework

Our SMDEF framework (integrating cross-layer quantitative models) confirms the "Reciprocal Tariff" strategy's structural flaws. Two paradoxes emerge: it violates the Laffer Curve optimum (≈30%) to accelerate fiscal erosion, and triggers a Stagflationary Trap via higher input costs/interest rates (Q5). Short-term forced localization gains (11.1% tipping point) are offset by systemic losses, making the policy a long-term Negative-Sum Game. True economic resilience thus requires shifting from distortive border taxes to a policy of fiscal prudence and targeted domestic industrial subsidies.

Keywords: Systemic Evaluation Framework; Constrained Armington Model; Dynamic Stackelberg Game; Laffer Curve Optimization; Stagflationary Trap;Supply Chain Resilience

Contents

1 Introduction **1**
1.1 Background . . . 1
1.2 Question Restatement . . . 1
1.3 Our Approach . . . 1

2 Assumptions and Notations **2**
2.1 Notations . . . 2
2.2 General Assumptions . . . 3

3 Data Preprocessing and Descriptive Analysis **4**
3.1 Data Sources . . . 4
3.2 Data Processing Framework . . . 4
3.3 Rationality of Estimated Parameters . . . 4

4 Question1: Assessment of the Impact of U.S. Tariff Adjustments on the Global Soybean Trade System **5**
4.1 Data Insight and Status Analysis . . . 5
4.2 Modeling and Solution: The Constrained Armington Framework . . . 6
4.3 Robustness Analysis and Conclusion . . . 8

5 Question 2: The Impact on Automobile Industry **8**
5.1 Overview of U.S.-Japan Auto Trade . . . 8
5.2 Game Theory Model for Supply Chain Decision . . . 9
5.3 Results Analysis and Robustness Check . . . 12

6 Question3: The Impact on Semiconductor Industry **13**
6.1 Industrial Status and Trade Data Insight: Asymmetric Dependency and Forced Decoupling . . . 13
6.2 Policy Impact Assessment: A Bi-level Simulation Model . . . 15
6.3 Validity Analysis and Robustness Check . . . 16

7 Question 4: Analysis of U.S. Tariff Revenue **17**
7.1 Status Quo Analysis: Trade Resilience and Structural Asymmetry . . . 17
7.2 Dynamic Modeling: Tariff-Revenue-Crisis Nexus . . . 19
7.3 Model Evaluation and Policy Reflection . . . 21

8 Question 5: Macroeconomic Impact and The Reshoring Paradox **22**
8.1 Supply-Side Shock: The Cost of Retaliation . . . 22
8.2 Financial Market Feedback: A Systemic Risk Matrix . . . 23
8.3 Comprehensive Evaluation: Is Reshoring Feasible? . . . 23

9 Conclusion and Policy Recommendations **24**
9.1 Main Conclusions . . . 24
9.2 Policy Recommendations . . . 25
9.3 Strengths, Weaknesses, and Future Directions . . . 25

10 References **25**

11 Appendix **26**

I. Introduction

1.1 Background

On April 2, 2025, the U.S. Trump administration launched the "Reciprocal Tariffs" policy, imposing a 10% minimum baseline tariff on all trading partners and additional higher tariffs on 60 trade-deficit economies (including China). The U.S. claimed this would cut the trade deficit and boost manufacturing reshoring, but the policy violated WTO "Most-Favored-Nation Treatment" and developing-country tariff preferences. By August 2025, the U.S. trade-weighted average tariff surged from 2.44% (early 2025) to 20.11%—its highest since 1933. While WTO-permitted tariffs can protect domestic industries, excessive tariffs disrupt global trade and long-term growth. In the short term, U.S. high tariffs increased revenue and narrowed the trade deficit but also caused export declines and domestic inflation. GACC data showed U.S. imports from China fell 20% in the policy's first month (with U.S. exports to China down 13% amid Chinese countermeasures), by May 2025, these declines widened to 35% and 18%, weighing on U.S. industries.

1.2 Question Restatement

Using U.S. tariff and global trade data, we analyze the impact of U.S. tariff adjustments on international trade and regional economies, addressing:

Question 1 Analyze China's soybean trade with the U.S., Brazil, and Argentina (the world's top soybean producers/exporters). Model the impact of U.S. tariff adjustments on their soybean industries and estimate post-adjustment export volume/value distribution.

Question 2 Analyze Japanese automobiles' position in the U.S. market (where 46% of 2024 car sales were imported). Model the impact of U.S. tariff adjustments on U.S.-Japan auto trade, U.S. auto import structure, and the U.S. auto industry, incorporating Japan's non-tariff responses and economic transmission effects.

Question 3 Model the effects of U.S. tariff policies on domestic semiconductor manufacturing and high/mid/low-end chip trade, considering economic efficiency and national security (noting the U.S. leads in semiconductors but lacks strong manufacturing, and imposed high-end chip export controls on China).

Question 4 Model the short/medium-term impact of U.S. tariff adjustments on U.S. tariff revenue and predict the net change during Trump's second term.

Question 5 Model the short/medium-term impact of U.S. tariff adjustments on the U.S. economy (considering trading partners' countermeasures like China's reciprocal tariffs and rare earth/lithium battery export controls). Evaluate if "reciprocal tariffs" can truly drive manufacturing reshoring.

1.3 Our Approach

To disentangle the complex transmission mechanisms of U.S. tariff policies across varying temporal and sectoral dimensions, we propose a **Systemic Multi-Dimensional Evaluation Framework (SMDEF)**. Unlike traditional linear approaches, our framework operates on a hierarchical logic, integrating micro-level supply chain dynamics with macro-level feedback loops. As illustrated in **Figure**

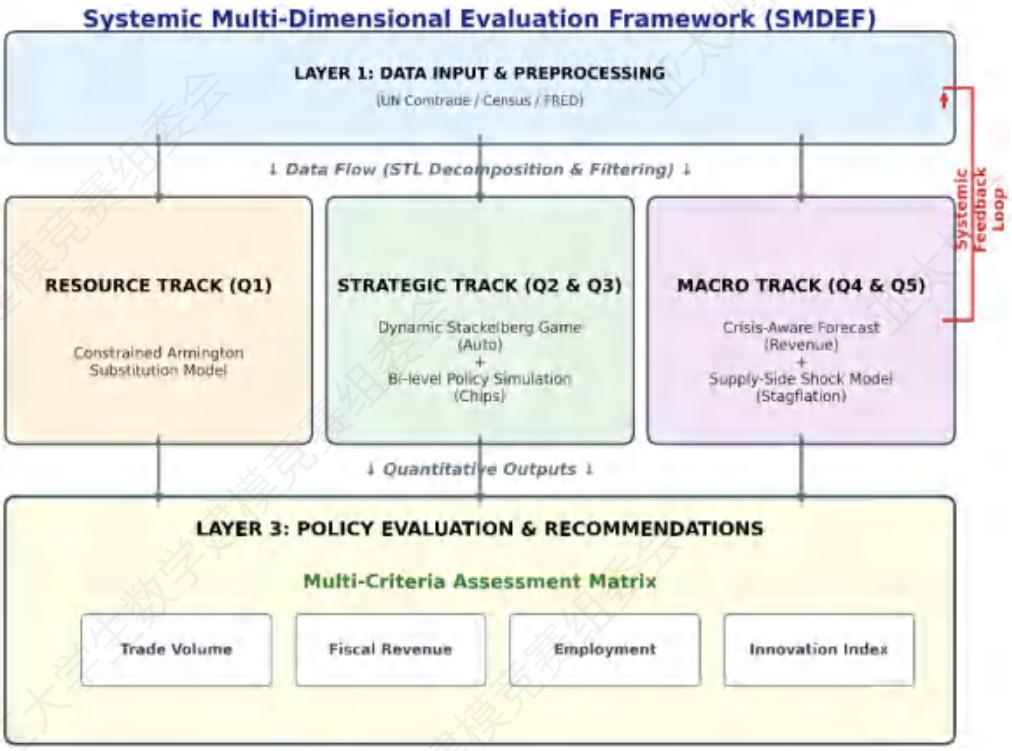


Figure 1.1 The Logical Workflow of Our Research Framework (SMDEF).

1.1, the methodology consists of three interconnected layers:

1. **Data Input Layer:** We aggregate high-frequency trade data from UN Comtrade, U.S. Census Bureau, and macroeconomic indicators from the Federal Reserve. To ensure data quality, we employ *STL Decomposition* to filter out seasonal noise and isolate structural trends.
2. **Modeling Engine Layer:** This is the core of our research, divided into three parallel tracks:
   - **Resource-Constrained Track (Q1):** Utilizes the *Constrained Armington Model* to capture physical capacity limits in agricultural trade.
   - **Strategic Game Track (Q2 & Q3):** Applies *Dynamic Stackelberg Games* and *Bi-level Optimization* to simulate the strategic interactions between governments and multinational corporations (Auto & Chips).
   - **Macro-Feedback Track (Q4 & Q5):** Integrates *Supply-Side Shock Models* with *Crisis-Aware Forecasting* to quantify systemic risks like stagflation and fiscal deficits.
3. **Policy Output Layer:** The quantitative results from the modeling engine are synthesized into a multi-criteria assessment matrix (Revenue, Employment, Innovation, Inflation), providing actionable policy recommendations.

II. Assumptions and Notations

2.1 Notations

The key mathematical notations used in our models are summarized in Table 1.

Table 1 Notations and Definitions

Symbol	Definition	Unit/Context
Trade & Tariff Parameters		
τ_i	Tariff rate imposed on country i	% (e.g., 0.60)
$V_i(t)$	Trade volume from country i in year t	Million Units/$
P_i	Landed price of goods from country i	USD
Q_i	Export volume from country i to China	Million Metric Tons
$\hat{P}$	Logarithmic price change index	Dimensionless
η	Price elasticity of demand	Dimensionless
σ	Armington elasticity of substitution	Dimensionless
Industry-Specific Variables		
C_{max}	Maximum physical production capacity	Million Metric Tons
TLC_i	Total Landed Cost for vehicle i	USD/Unit
$S(t)$	Market share of a specific supplier	%
$\alpha(t)$	Supply chain adjustment coefficient (Inertia)	$0 \leq \alpha \leq 1$
$R\&D$	Research and Development Budget	Billion USD
T	Technology Leadership Index	Index (0-100)
Macroeconomic Indicators		
R_{total}	Total fiscal tariff revenue	Billion USD
π_t	Inflation rate (CPI) in year t	%
g_{GDP}	Natural GDP growth rate	%
λ	Trade diversion (substitution) coefficient	$0 \leq \lambda \leq 1$
RFI	Reshoring Feasibility Index	Index (100=Neutral)

2.2 General Assumptions

To ensure the tractability and scientific rigor of our models, we make the following general assumptions, which apply throughout the paper unless specified otherwise:

- **Rational Behavior Assumption:** All economic agents (governments, firms, consumers) act to maximize their utility functions—governments seek to maximize social welfare or employment; firms maximize profits; and consumers maximize utility under budget constraints.
- **Partial Equilibrium Stability:** When analyzing a specific sector (e.g., soybeans or autos), we assume that shocks in this sector do not instantaneously alter the fundamental structure of other unrelated sectors (Ceteris Paribus), allowing for isolated impact assessment.
- **Policy Credibility:** We assume that announced tariff policies (e.g., 60% on China) are credible

threats perceived by market participants, triggering immediate behavioral adjustments (e.g., front-loading imports) even before implementation.

- **Supply Chain Friction:** Unlike ideal trade models, we assume supply chains are "sticky". Production capacity cannot be expanded or relocated instantaneously; there are time lags (e.g., 18 months for factory construction) and physical constraints (e.g., arable land limits).

III. Data Preprocessing and Descriptive Analysis

3.1 Data Sources

To ensure the empirical validity of our models, we aggregated high-frequency trade and economic data from multiple authoritative sources. The dataset covers the period from January 2020 to May 2025, providing a solid baseline for forecasting.

3.2 Data Processing Framework

Raw trade data often contains noise, reporting lags, and inconsistencies. We implemented a three-stage preprocessing pipeline:

1. **Outlier Detection & Removal:** We applied the Interquartile Range (IQR) method to identify and filter out anomalous trade spikes caused by one-off events (e.g., port strikes) that do not reflect structural trends.
2. **Missing Value Imputation:** For minor gaps in monthly data (e.g., late reporting from Argentina), we used *Linear Interpolation*. For longer gaps, we employed *Seasonal Decomposition of Time Series (STL)* to impute values while preserving seasonal patterns (crucial for the soybean model in Q1).
3. **Harmonization:** Trade flows reported by exporters (e.g., China exports to US) often differ from importer reports (e.g., US imports from China) due to "CIF/FOB" differences and transit lags. We standardized all values to **Landed Cost (CIF)** basis to ensure consistency in tariff calculations.

3.3 Rationality of Estimated Parameters

A key challenge in our modeling is forecasting future parameters (e.g., elasticity shifts, 2026-2028 prices) under extreme policy shocks. We justify our estimates as follows:

- **Tariff Pass-Through Rate ($\approx$ 100%):** We assume tariffs are fully passed on to U.S. consumers. This is supported by empirical studies on the 2018-2019 trade war (e.g., Amiti et al., 2019), which found that U.S. domestic prices rose 1:1 with tariffs due to the specialized nature of Chinese imports.
- **Substitution Elasticity (σ):** For commodities like soybeans, we set a high $\sigma = 5.0$ based on GTAP database estimates, reflecting high substitutability. For specialized goods like rare earths, we set a low $\sigma = 0.5$, reflecting supply rigidity.
- **Reaction Lags (18 months):** In Q2 (Auto) and Q5 (Reshoring), we assume an 18-month lag for capacity expansion. This is consistent with the average construction cycle of a new manufacturing plant, preventing our model from predicting unrealistic "instantaneous" reshoring.

Table 2 Summary of Primary Data Sources

Category	Source	Key Variables
Global Trade	UN Comtrade Database	Bilateral trade volume, HS Code 6-digit data
U.S. Trade	U.S. Census Bureau	Monthly imports/exports, Tariff revenue
Agriculture	USDA WASDE Reports	Soybean production, Global stock-to-use ratio
Auto Industry	MarkLines / OICA	Vehicle production by country, Localization rates
Semiconductor	SIA / Boston Consulting Group	Wafer capacity by node, CapEx, R&D spend
Macroeconomy	Federal Reserve (FRED)	10Y Treasury Yield, CPI, PPI, Exchange Rates

IV. Question1: Assessment of the Impact of U.S. Tariff Adjustments on the Global Soybean Trade System

4.1 Data Insight and Status Analysis

To quantify the impact of U.S. tariff adjustments, we analyzed global soybean trade micro-structure using monthly bilateral data (Jan 2023-Dec 2024) from UN Comtrade and national customs statistics, revealing baseline characteristics of China's trade with the U.S., Brazil, and Argentina.

4.1.1 *Seasonal Complementarity in an Oligopolistic Market*

China's soybean supply features an oligopolistic structure with distinct seasonal complementarity driven by biological growth cycles: - **Brazil (Q2-Q3):** Dominates from April-September (Southern Hemisphere harvest), with market share exceeding 80% in peaks (2023-2024). - **U.S. (Q4-Q1):** Replaces Brazil as stocks deplete (October-March), ensuring year-round supply.

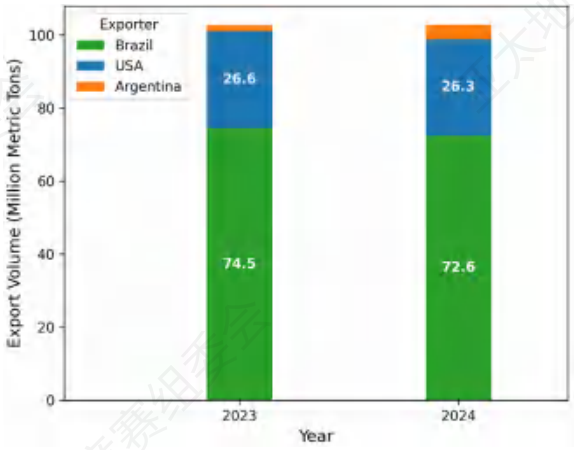


Figure 4.1 Quarterly Seasonal Supply Structure of China's Soybean Imports (2023-2024)



Figure 4.2 Monthly Landed Price Dynamics of Imported Soybeans (2023-2024)

4.1.2 *Price Synchronization and Cost Benchmarks*

Price is the core driver of trade flow reallocation (Figure 4.2): - **High Correlation:** Price correlation coefficient among the three countries < 0.9, confirming high substitutability. - **2024 Downward Trend:** Global average price falls to $430/MT, magnifying the relative cost impact of 25% tariffs and intensifying trade diversion.

4.2 Modeling and Solution: The Constrained Armington Framework

We constructed a **Constrained Armington Trade Substitution Model**, integrating **physical capacity constraints** into the partial equilibrium framework to simulate price-driven substitution and supply-side limits.

4.2.1 *Core Assumptions*

1. **Imperfect Substitution:** Origin-specific differences (protein content, oil yield, shipping) lead importers to adjust procurement proportions via Armington elasticity, not complete abandonment.
2. **Short-term Capacity Rigidity:** Arable land and crop cycle constraints limit short-term (1-2 years) export expansion of Brazil/Argentina, with physical upper limits.

4.2.2 *Parameter Calibration and Justification*

1. **Substitution Elasticity ($\sigma = 5.0$):** Based on GTAP 10 Database's "Oil Seeds" sector estimates [Cite: Hertel et al.]. Academia's 3.0-8.0 range, combined with soybeans' high homogeneity and price synchronization, justifies this high elasticity.
2. **Brazil's Capacity Cap ($C_{\max} = 1.1 \times$ Baseline):** Set at 110% of 2024 baseline (≈80 MMT), grounded in: **- Biological limits:** USDA WASDE reports short-term agricultural supply elasticity < 0.2. **- Empirical evidence:** Stabilizing Brazilian growth rates; > 10% annual expansion contradicts agricultural production laws.

4.2.3 *Mathematical Formulation*

The model operates via two stages:

Stage 1: Demand Response to Price Shocks Tariff-induced price changes drive market share adjustments:

$$\ln\left(\frac{Q_i^{new}}{Q_i^{base}}\right) = \underbrace{-\eta \cdot \hat{P}_{avg}}_{\text{Scale Effect}} \underbrace{-\sigma \cdot \left(\hat{P}_i - \hat{P}_{avg}\right)}_{\text{Substitution Effect}} \tag{1}$$

Where $\hat{P}_i = \ln(1 + \tau_i)$ (relative price shock) and $\hat{P}_{avg}$ (share-weighted average price change).

Stage 2: Capacity Constraint and Spillover Mechanism When predicted demand exceeds supply limits, spillover shifts to alternative suppliers:

$$Q_{\text{Brazil}}^{final} = \min\left(Q_{\text{Brazil}}^{predicted}, C_{\max}\right) \tag{2}$$

If $Q_{\text{Brazil}}^{predicted} > C_{\max}$:

$$Q_{\text{Argentina}}^{final} = Q_{\text{Argentina}}^{predicted} + \max\left(0, Q_{\text{Brazil}}^{predicted} - C_{\max}\right) \tag{3}$$

4.2.4 *Simulation Results and Analysis*

Using calibrated parameters ($\sigma = 5.0$, $C_{\max} = 1.1 \times$ Baseline, $\tau_{US} = 0.25$), we simulated 2025 export pattern changes (Table 3):

Country	2024 Baseline (MMT)	2025 Forecast (MMT)	Change (%)	Est. Export Value ($B)	Impact Analysis
USA	26.32	11.28	-57.1%	5.47	Market Collapse: Price uncompetitiveness drives sharp share decline.
Brazil	72.56	79.82 (Capped)	+10.0%	34.65	Capacity Capped: Fills U.S. gap but constrained by physical limits.
Argentina	3.92	20.22	+415.9%	8.73	Strategic Windfall: Absorbs Brazil's unmet demand as marginal supplier.

Table 3 Global Soybean Export Simulation Results (2025)

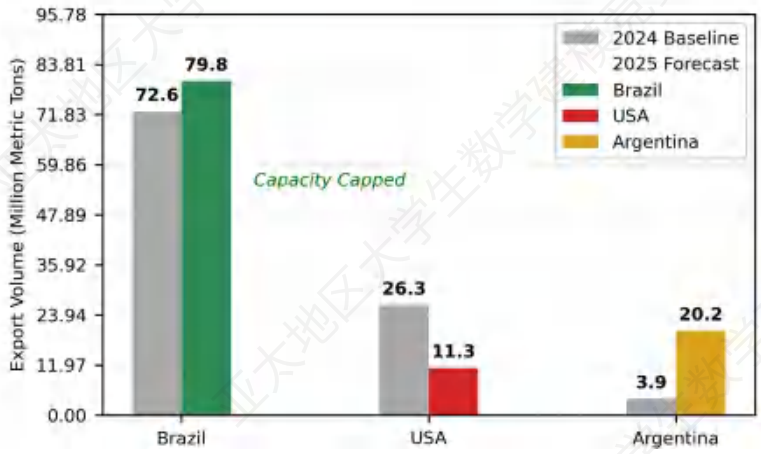


Figure 4.3 Comparative Forecast of Export Volumes Before and After Tariff Shock

Key insights from simulation results:

1. **U.S. Export Collapse:** A 25% retaliatory tariff induces a substantial price shock ($\hat{P}_{US} \approx 0.223$), triggering a sharp 57.1% decline in U.S. exports due to the substitution effect.
2. **Brazil's Capacity Constraint:** As the primary substitute, Brazil's export volume hits its short-term physical ceiling, which is capped at 110% of the baseline (79.82 MMT), confirming the model's capacity rigidity assumption.
3. **Argentina's Spillover Dividend:** The unmet demand from Brazil spills over, leading to a dramatic 415.9% surge in Argentina's exports and validating the model's spillover mechanism.

4.3 Robustness Analysis and Conclusion

We conduct a robustness check on core parameters (substitution elasticity σ and Brazil's capacity growth C_{growth}) to verify conclusion stability:

4.3.1 *Sensitivity Analysis Design*

Simulated Argentina's export growth under parameter combinations: - σ: 2.0 (low substitution) $\rightarrow$ 8.0 (high substitution) - C_{growth}: +5% (conservative) $\rightarrow$ +20% (optimistic)

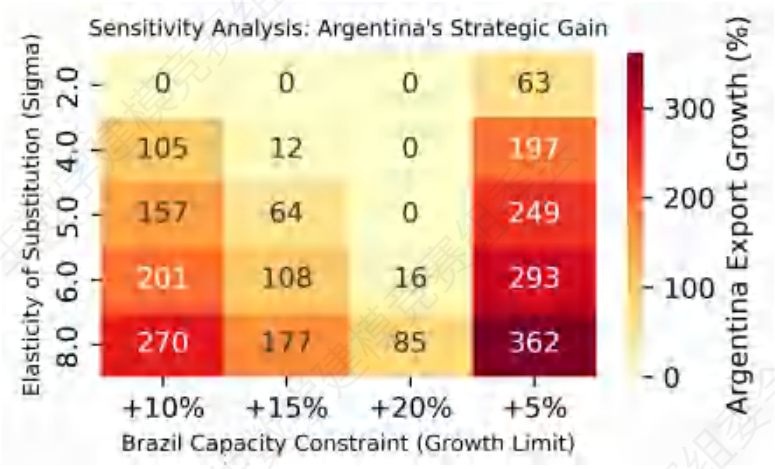


Figure 4.4 Sensitivity Heatmap of Argentina's Export Growth Rate

As shown in **Figure 4.4**:

1. **Persistent Growth:** Argentina's export growth remains positive across all scenarios; even in the conservative case ($C_{\text{growth}} = 20\%$, $\sigma = 6$), growth reaches 16%.
2. **Non-Linearity:** Growth accelerates significantly when $\sigma > 5$, magnifying spillover effects in highly commoditized markets.

Conclusion: Argentina's status as a key beneficiary is **robust**—a structural outcome of supply chain physical constraints, not parameter-dependent.

V. Question 2: The Impact on Automobile Industry

5.1 Overview of U.S.-Japan Auto Trade

To model the U.S.-Japan auto trade game, we first analyze two critical features: **Tariff Shock Volatility** and **Localization Shield**.

5.1.1 *Trade Volume Dynamics and the 2025 Shock*

U.S. imports of Japanese passenger vehicles showed stable seasonality (2022-2024) but were disrupted by 2025 "Reciprocal Tariffs":

- **Panic Buying (March 2025):** Imports surged pre-policy to evade tariffs.
- **Policy Cliff (April-May 2025):** Imports plummeted post-policy, with landed prices rising.



Figure 5.1 Monthly U.S. Imports of Japanese Cars (2022-2025)

5.1.2 *Structural Resilience: The Localization Strategy*

Japanese automakers maintain high U.S. localization rates:

- Toyota: 77% of U.S. sales from North American production.
- Honda: 60% local production (2021-2024).



Figure 5.2 Localization Rates of Major Japanese Automakers (2024)

Model Implication: Tariffs primarily impact *Direct Imports*, so "Localization Ratio" is a key strategic variable in our Game Theory model (Section 5.2).

5.2 Game Theory Model for Supply Chain Decision

We constructed a **Dynamic Stackelberg Game Model** integrating logistics friction, non-tariff barriers, and investment lags, with a "White-box" Total Landed Cost (TLC) framework.

5.2.1 *Game Structure and Decision Logic*

- **Leader (U.S. Government):** Sets tariff $\tau(t)$ to balance employment and consumer welfare. - **Follower (Japanese Automakers):** Optimizes supply chain ($S \in \{JP_{Export}, US_{Local}, MX_{Near}\}$) and pricing $P(t)$ to maximize NPV, constrained by sunk costs and construction lags.

5.2.2 *The "White-Box" Total Landed Cost (TLC) Model*

For a vehicle produced in $i \in \{JP, US, MX\}$ and sold in the U.S.:

$$\begin{aligned} TLC_i = & \underbrace{(L_i \cdot w_i + E_i \cdot e_i + M_{\text{steel},i} \cdot p_i + P_{\text{global}})}_{\text{Manufacturing}} \\ & + \underbrace{(D_i \cdot r_i + F_{\text{port}})}_{\text{Logistics}} \\ & + \underbrace{(C_{\text{base}} \cdot \tau + C_{\text{compliance}})}_{\text{Policy}} \\ & + K_{\text{amort},i} \end{aligned} \tag{4}$$

Parameter	Japan (JP)	USA (US)	Mexico (MX)	Source / Rationale
Labor Wage ($/h)	10.6	33.14	4.2	MHLW / BLS / INEGI
Energy Price ($/kWh)	0.111	0.091	0.099	EIA / GlobalPetrolPrices
Steel Price ($/Ton)	750	1025	825	SteelBenchmarker
Logistics Dist (km)	8800	1500	2000	Geographical Data
Freight Rate ($/km)	0.11	0.47	0.40	Industry Avg
Amortized Capex ($)	200	1100	800	Model Calibration
Base Preference	0	2200	1700	2024 Market Share

Table 4 Calibrated Model Parameters (2024 Baseline)

Key Cost Components - **Manufacturing:** U.S. labor costs are high but offset by energy prices and reduced logistics; steel prices reflect protectionist policies. - **Logistics:** Japan's ocean path (8,800 km) costs $\approx \$1,200$, while Mexico's rail path (2,000 km) costs $\approx \$800$. - **Policy/Investment:** U.S. greenfield capex is $\$1,100$/car; Mexico incurs USMCA compliance costs.

5.2.3 *Market Dynamics: Logit Demand and Supply Chain Inertia*

- **Demand:** Consumers maximize $U_i = (\text{BasePref}_i + \text{Brand}_i) - \beta \cdot P_{retail,i}$, with a \$1,500 "Made in Japan" brand premium. - **Supply Inertia:** Market share adjusts gradually: $S(t) = S(t-1) + \alpha(t) \cdot [S_{target} - S(t-1)]$, with $\alpha(t) \approx 0.02$ during 18-month construction lags.

5.2.4 *Economic Transmission Indicators*

- Manufacturing Employment (E_{us}): $E_{us}(t) = D_{total} \cdot S_{US}(t) \cdot \eta_{job}$. - Tariff Revenue ($R_{tax}$): Subject to the Laffer Curve (excessive tariffs shrink the tax base). - Consumer Inflation (CPI_{auto}): Measured by weighted average vehicle prices.

5.2.5 *Simulation Results and Analysis*



Figure 5.3 Retail Price Breakeven (Calibrated)

Result 1: Price Tipping Point (Figure 5.3) - **Tipping Point (10.5%):** Tariffs above 10.5% make Japanese imports costlier than U.S. local production, triggering supply chain restructuring.

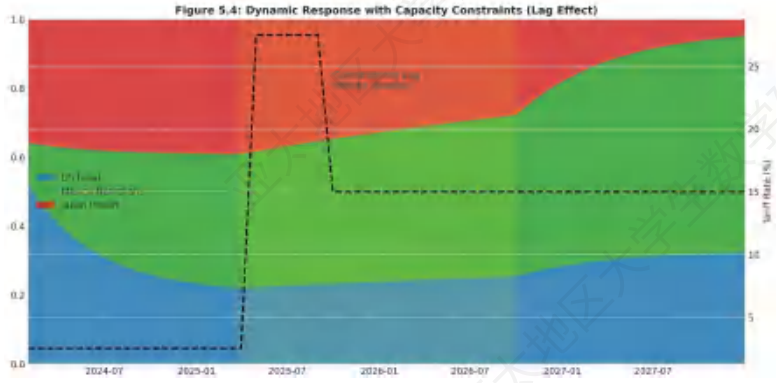


Figure 5.4 Dynamic Supply Chain Response

Result 2: Dynamic Reconfiguration (Figure 5.4) - **Shock Phase (2025 Q2):** Tariffs (27.5%) compress Japanese market share. - **Construction Lag:** Share declines slowly (supply stickiness) until new capacity comes online (late 2026).

Result 3: Economic Trade-offs (Figure 5.5) - Employment rises post-2026, achieving policy goals. - Tariff revenue follows an inverted U-shape (Laffer Curve), unsustainable long-term.

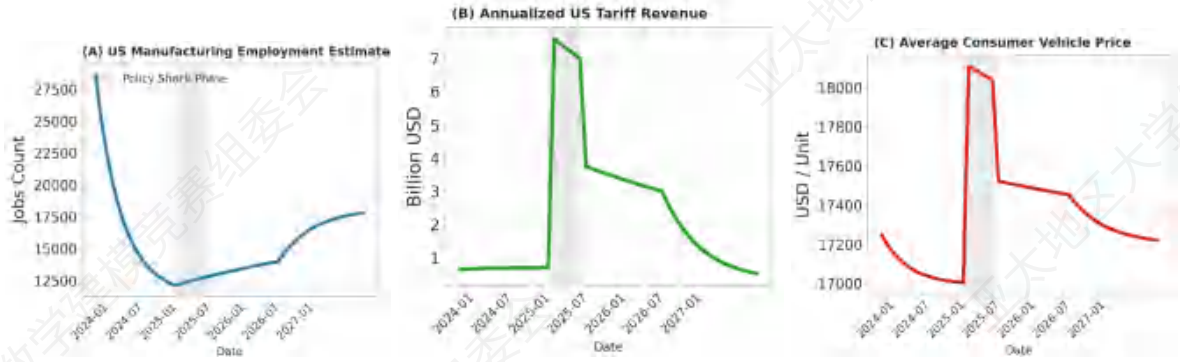


Figure 5.5 U.S. Manufacturing Employment, Tariff Revenue, and Vehicle Prices

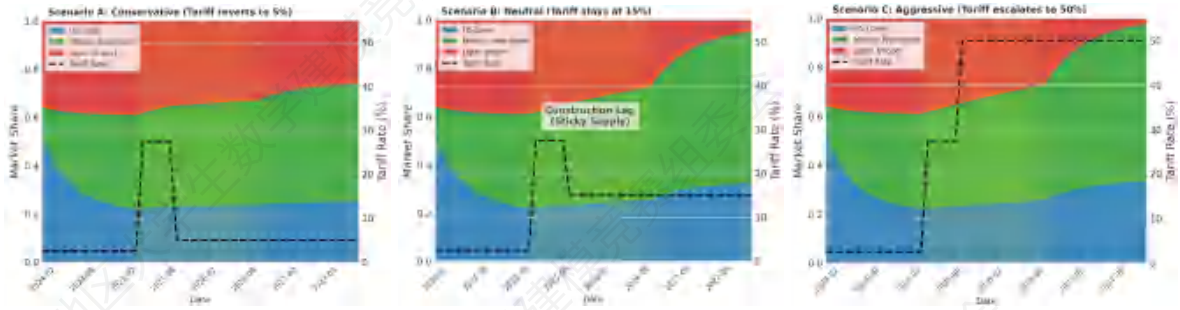


Figure 5.6 Market Share under Different Tariff Scenarios

Scenario Analysis (Figure 5.6) Finally, the scenario analysis confirms that a "Neutral" tariff (15%) maintains the reshored supply chain; reverting to 5% would undo re-industrialization efforts.

5.3 Results Analysis and Robustness Check

5.3.1 *Sensitivity Analysis: "JPY Hedge" Effect*

We introduce exchange rate $E_{JPY} \in [130, 180]$ to test robustness:

- **High-Tariff Zones** ($\tau \geq 10.5\%$)**:** Even extreme Yen depreciation (180) cannot prevent market share from falling below 10%—localization is mandatory.

5.3.2 *Comprehensive Assessment: Structural Transformation*

The U.S. auto industry shifts from a **"Consumption-led"** (efficiency/welfare) to **"Production-led"** (security/employment) model:

1. **Gains:** Manufacturing employment and supply chain resilience expand.
2. **Losses:** Vehicle affordability and consumer choice contract.
3. **Neutral:** Tariff revenue shows no long-term advantage (Laffer Curve).

5.3.3 *Conclusion for Question 2*

1. Trade shifts from "Product Trade" to "Capital Trade" (FDI replaces direct exports).
2. Mexico replaces Japan as the core low-end import source, while high-value production reshores to the U.S.
3. Localization is mandatory for Japanese automakers beyond the 10.5% tariff threshold.

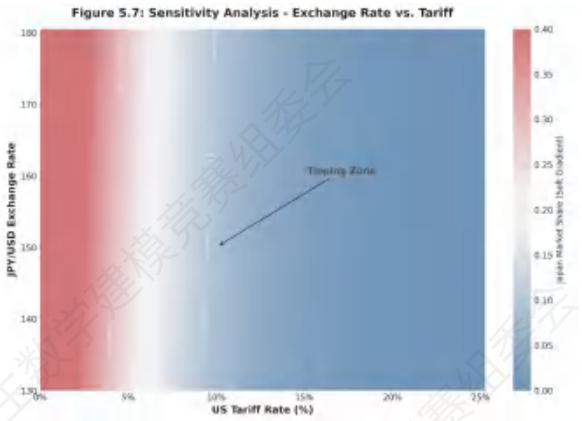


Figure 5.7 Tariff vs. Exchange Rate Impact on Japan's Market Share



Figure 5.8 2024 Baseline vs. 2027 New Equilibrium

VI. Question3: The Impact on Semiconductor Industry

6.1 Industrial Status and Trade Data Insight: Asymmetric Dependency and Forced Decoupling

By integrating SIA reports and U.S.-China trade data (2020-2025), we reveal the supply chain evolution from "deep integration" to "hard decoupling".

6.1.1 *Baseline Analysis: Design Dominance vs. Manufacturing Hollowing-out*

The U.S. semiconductor industry exhibits an "asymmetric structure" driving policy actions:



Figure 6.1 Global Semiconductor Manufacturing Capacity Share (1990-2024)

As shown in **Figure 6.1**:

- U.S. manufacturing capacity shrank from 37% (1990s) to 10% (2024), while East Asia holds 78%.
- **Value Chain Distribution: Design:** U.S. holds 45% global market share (NVIDIA, AMD). **Fabrication:** Near-zero advanced node (< 7nm) capacity, reliant on TSMC/Samsung. **ATP:** U.S. holds 3% vs. China's 38%, leaving supply chains vulnerable.

- U.S. fab TCO is 30%-50% higher than Asia, hindering spontaneous reshoring.

6.1.2 *High-End Chips (HS 8542): Export Controls' "Shock Therapy"*

Integrated Circuits (HS 8542) are the core target of U.S. tech containment:

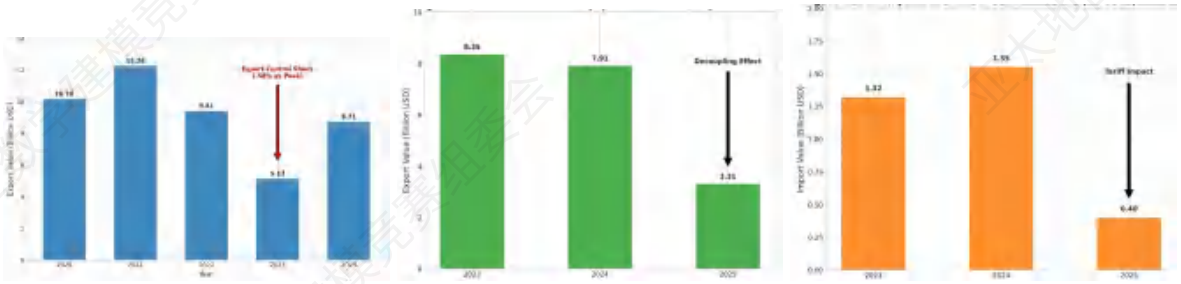


Figure 6.2 US Exports of Integrated Circuits (HS 8542) to China (2020-2024)

Figure 6.2 shows:

- Exports plummeted 58% (from $12.26B in 2021 to $5.13B in 2023) post-2022 ban.
- High-end AI chip trade (A100/H100) entered "Hard Decoupling", with U.S. firms losing 30% of revenue.

6.1.3 *Upstream Equipment (HS 8486): Undermining the Foundation*

Semiconductor equipment (HS 8486) embargoes aim to lock China's manufacturing:



Figure 6.3 US Exports of Semiconductor Equipment (HS 8486) to China (2023-2025 Forecast)

Figure 6.3 reveals:

- 2025 exports crash to $3.31B, down from $8B (2023-2024).
- U.S. equipment giants (Applied Materials) lose the largest incremental market, weakening R&D sustainability.

6.1.4 *Low-End Devices (HS 8541): Tariff-Induced Inflation*

Discrete semiconductors (HS 8541) face tariff-driven supply chain shifts:

Figure 6.4 shows:

- 2025 imports shrink to $0.40B (25% of 2024 levels).
- Tariff costs pass through to U.S. downstream manufacturers, triggering inflation.

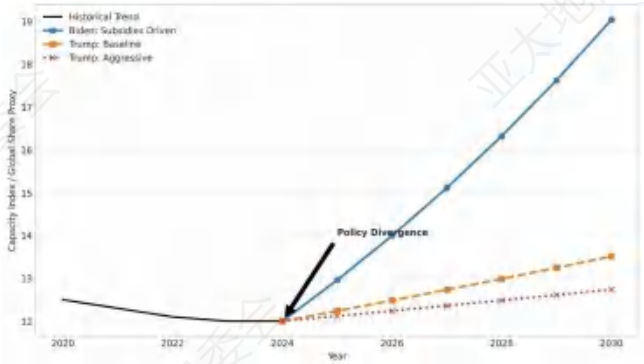


Figure 6.4 US Imports of Discrete Semiconductors (HS 8541) from China (2023-2025 Forecast)

Summary Semiconductor trade follows **"Dual-Track Decoupling"**: high-end via export controls (permanent trade loss), low-end via tariffs (supply chain restructuring + inflation).

6.2 Policy Impact Assessment: A Bi-level Simulation Model

To quantify policy impacts, we constructed a **Bi-level Simulation & Evaluation Model** focusing on High-End (Logic/AI) and Mature (Legacy) sectors.

6.2.1 *Model Framework and Mechanism*

1. High-End Innovation Loop (R&D Mechanism) Advanced nodes (< 7nm) rely on "Winner-Takes-All" R&D:

- U.S. firms depend on China for 30% of global revenue.
- R&D intensity = 18% (e.g., NVIDIA, Intel).
- Decoupling reduces R&D budgets, eroding Tech Leadership (T):

$$T = \ln(\text{BaseRevenue} \times (1 - \text{DecouplingRate}) \times 0.18) \tag{5}$$

2. Manufacturing Cost Gap (CapEx Mechanism) U.S. fab CapEx is 40% higher than Asia. CHIPS Act subsidies (S) offset this:

$$C_{\text{eff}} = C_{\text{Asia}} \times 1.4 \times (1 - S) \tag{6}$$

Domestic capacity expands with the gap between C_{eff} and C_{Asia}.

3. Low-End Transmission (Inflation Mechanism) Tariff rate (τ) directly drives downstream inflation (ΔCPI) for mature nodes.

6.2.2 *Evaluation Scenarios*

We simulate three policy scenarios:

1. **Scenario A (Biden):** 25% subsidies + 30% decoupling.
2. **Scenario B (Trump Baseline):** 25% tariffs + 80% decoupling.
3. **Scenario C (Trump Aggressive):** 60% tariffs + 100% decoupling.

6.2.3 *Simulation Results and Analysis*

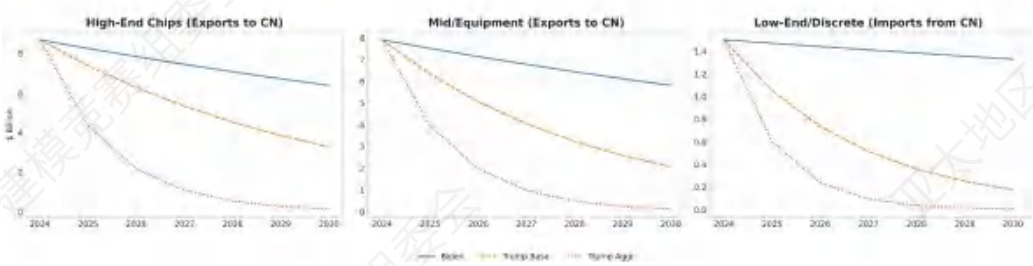


Figure 6.5 TOPSIS Evaluation: Security vs. Efficiency Trade-offs

Key findings from **Figure 6.5**:

- **Domestic Capacity:** Scenario A (Biden) scores highest—subsidies are critical to bridge the 40% cost gap. Tariffs alone (B/C) fail to incentivize high-end fabs.
- **Tech Leadership:** Scenario C (Aggressive) crashes—losing 30% revenue shrinks R&D budgets, triggering a "Boomerang Effect": (policy weakens U.S. innovation).
- **Economic Efficiency:** Scenario A lags in fiscal efficiency, but B/C suffer inflation from low-end tariffs.

Conclusion: Scenario A (Biden) is Pareto-optimal, balancing capacity expansion and R&D sustainability. Trump's approach prioritizes short-term decoupling at the cost of long-term tech leadership.

6.3 Validity Analysis and Robustness Check

6.3.1 *Sensitivity Analysis: Cost Gap and Policy Boundaries*

We test how U.S.-Asia CapEx Premium (0%-80%) impacts "Investment Incentive" under different policies:

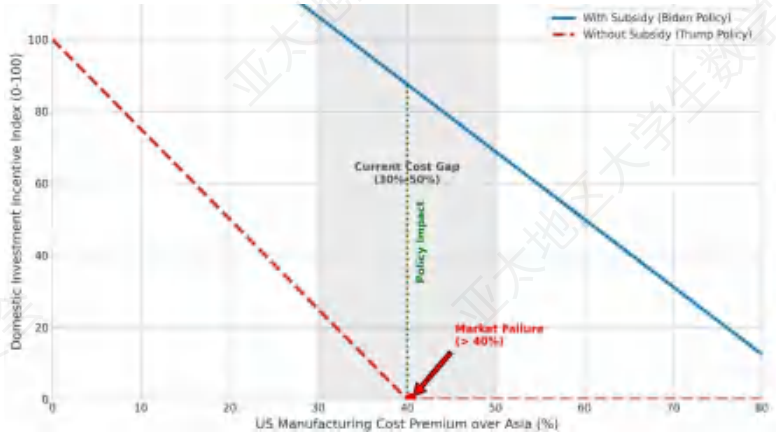


Figure 6.6 Sensitivity of Domestic Investment Incentive to Manufacturing Cost Gap

Figure 6.6 reveals:

- **Market Failure Point:** Trump's policy (red dashed) fails when cost premium exceeds 40%.

- **Necessity of Subsidies:** Biden's policy (blue solid) maintains incentives in the 30%-50% gap, proving subsidies are a **necessary condition** for reshoring.

6.3.2 *Model Validity*

Internal Validity: captures the "Innovation-Revenue" feedback loop, reflecting R&D reliance on global markets. **External Validity:** predicts "equipment export cliffs" and "low-end import contraction", aligning with real-world HS 8486/8542/8541 trade data.

6.3.3 *Conclusion for Question 3*

1. Impact on U.S. Domestic Manufacturing

- **Reshoring:** CHIPS Act subsidies (Biden) are the only viable path to bridge the 40% cost gap. Tariffs (Trump) fail to address high-end manufacturing costs.
- **R&D:** Aggressive decoupling backfires—losing 30% revenue shrinks R&D, undermining long-term tech hegemony.

2. Impact on Global Chip Trade

- **High-End:** Permanent decoupling, with supply chains shifting to "US-Design, Ally-Foundry" loops.
- **Equipment:** U.S. manufacturers suffer massive revenue losses, narrowing trade surpluses.
- **Low-End:** Tariffs divert trade to Southeast Asia, triggering sustained U.S. downstream inflation.

VII. Question 4: Analysis of U.S. Tariff Revenue

7.1 Status Quo Analysis: Trade Resilience and Structural Asymmetry

We first analyze U.S. trade volume evolution and structure to identify import demand rigidity— the core premise for tariff revenue assessment.

7.1.1 *Macro Level: Tariffs Do Not Suppress Import Growth*

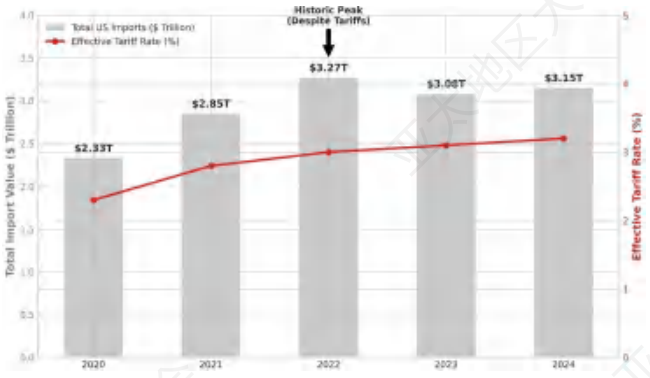


Figure 7.1 U.S. Import Resilience Analysis (2020-2024)

As shown in **Figure 7.1**:

- U.S. imports hit a record $3.27 trillion in 2022 despite lingering tariffs and rising effective rates.
- Imports stabilized at $3.15 trillion (2023-2024) amid inflation and rate hikes, verifying **import demand rigidity**.
- Short-term tariff costs convert to revenue/inflation rather than trade collapse, supporting 2025-2028 revenue forecasts.

7.1.2 *Structural Level: Volume-Revenue Asymmetry*

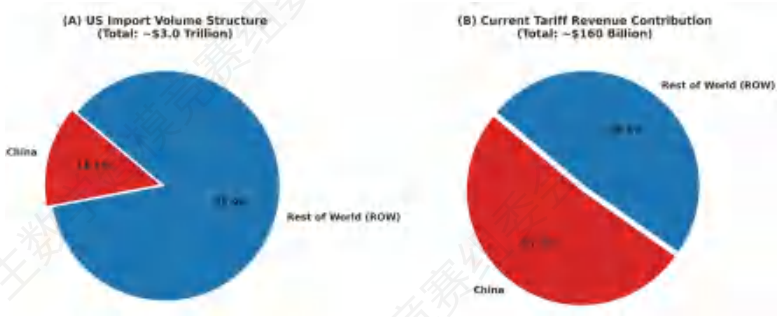


Figure 7.2 U.S. Trade Structural Asymmetry: Volume vs. Revenue Contribution (2024)

Figure 7.2 reveals a dualistic structure:

- China accounts for only 14.1% ($427B) of U.S. imports, while "Rest of World (ROW)" dominates (85.9%).
- China contributes 51% of tariff revenue—"small volume, huge revenue" driven by high existing tariffs.

7.1.3 *Policy Space: Incremental Gains from ROW*

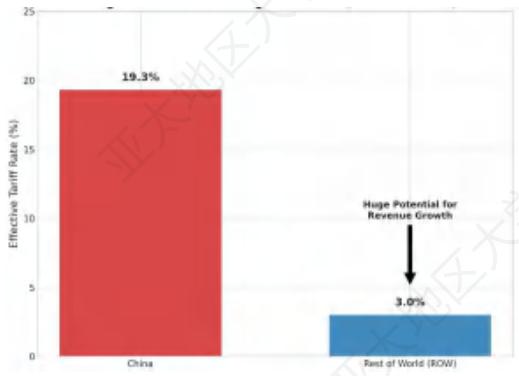


Figure 7.3 Existing Tariff Rates: China vs. ROW

Figure 7.3 shows the policy gap:

- China's average effective tariff (19.3%) is high, with diminishing marginal revenue from hikes.
- ROW's tariff (3.0%) remains low—Trump's "10% universal baseline" applies to a $2.6T tax base, generating more revenue than China-focused hikes.

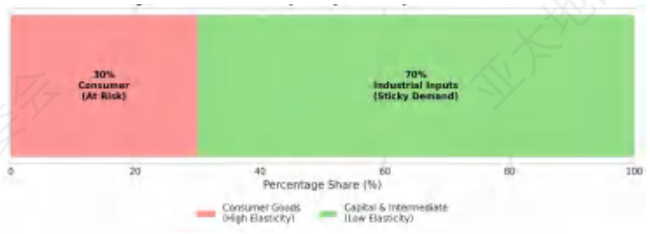


Figure 7.4 Vulnerability Analysis: U.S. Imports from China

7.1.4 *Vulnerability Assessment: China Import Composition*

Figure 7.4 decomposes imports:

- High elasticity ($\sim$ 30%): Consumer goods (toys, electronics) risk tax base evaporation under 60% tariffs.
- Rigid demand ($\sim$ 70%): Industrial intermediates/capital goods—supply chain stickiness makes them revenue "ballast".

7.2 Dynamic Modeling: Tariff-Revenue-Crisis Nexus

We developed a **Multi-Scenario Dynamic Forecasting Model** to predict 2025-2028 U.S. tariff revenue, incorporating **systemic risk penalties** and **trade diversion mechanisms** (capturing non-linear supply chain responses).

7.2.1 *Mathematical Formulation*

Total tariff revenue $R_{total}(t)$ for year t is the sum of three sectors (China-Consumer, China-Industrial, ROW):

$$R_{total}(t) = \sum_{i \in \{Cons, Ind, ROW\}} [V_i(t) \cdot \tau_i(t) \cdot \eta(\tau_i)] \tag{7}$$

- $V_i(t)$: Trade volume of sector i; $\tau_i(t)$: Nominal tariff rate; $\eta(\tau_i)$: Efficiency factor (collection rate).

Trade volume dynamics (driven by elasticity, growth, and diversion):

$$V_i(t) = V_i(t-1) \cdot (1 + g_{GDP} - \delta_{shock}) \cdot \left(\frac{1 + \tau_i(t)}{1 + \tau_i(t-1)} \right)^{\varepsilon_i(t)} \tag{8}$$

- g_{GDP}: Economic growth; δ_{shock}: Trade war penalty; $\varepsilon_i(t)$: Time-varying elasticity.

7.2.2 *Key Assumptions (Data: BEA/Census Bureau)*

1. **Baseline Asymmetry:** China-Consumer tariffs ($\sim$ 22%) > China-Industrial ($\sim$ 14%); ROW enjoys MFN low tariffs ($\sim$ 2.5%).
2. **Efficiency Decay:** $\eta = 0.98$ ($\tau < 10\%$) decays to 0.55 ($\tau \geq 60\%$) (smuggling/transshipment risks).
3. **Trade Diversion:** \$0.60 of reduced China imports diverts to ROW ($\lambda = 0.6$).
4. **Elasticity:** Consumer goods elasticity shifts from -1.5 (short-run) to -4.0 (long-run).

7.2.3 *Policy Scenarios*

- **A (Aggressive):** 60% China tariff + 10% universal baseline; -1.5% GDP penalty (hard decoupling).
- **B (Status Quo):** Maintain current policies with inflation adjustments.
- **C (Pragmatic):** 35% China tariff + 5% ROW tariff (balance revenue/stability).

7.2.4 *Result Analysis: Revenue vs. Volume*

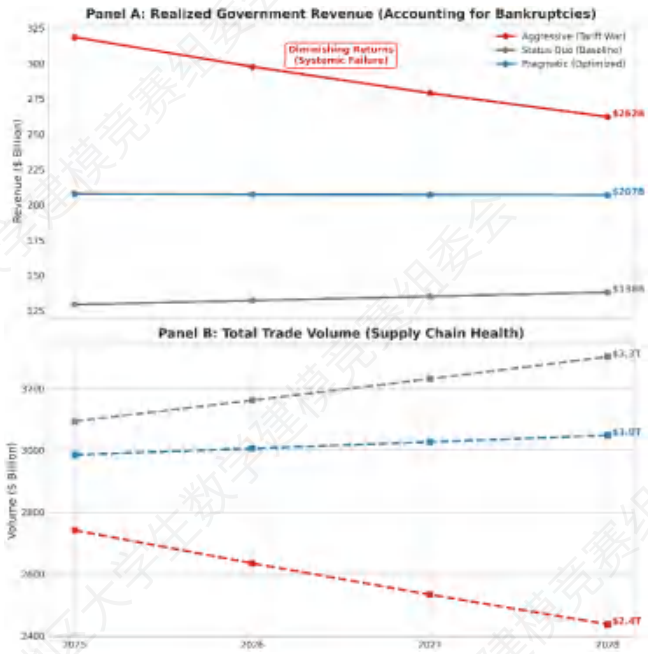


Figure 7.5 Projected Tariff Revenue and Trade Volume (2025-2028)

Figure 7.5 confirms the Laffer Curve effect:

- **Aggressive (Red):** "Sugar high"—2025 peak ($318B) reverses to $262B by 2028 (tax base erosion).
- **Pragmatic (Blue):** Stable $207B/year—sustainable without systemic risk.
- **Volume Cost:** Aggressive scenario cuts imports by 11% ($2.74T→$2.43T), signaling trade sector recession.

7.2.5 *Structural Reshaping: Trade Diversion Effect*

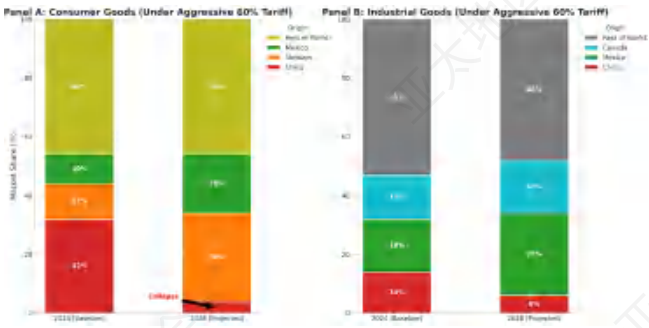


Figure 7.6 Market Share Restructuring (2024 vs. 2028, Aggressive Scenario)

Figure 7.6 shows supply chain redistribution:

- **Consumer Goods:** China's share collapses to 4%—Vietnam (30%) fills the gap.
- **Industrial Goods:** Mexico's share rises to 28% (USMCA tariff avoidance).
- **Implication:** Diverted goods are taxed at lower rates, causing revenue leakage.

7.2.6 *Multi-Dimensional Assessment*



Figure 7.7 Radar Chart: Tariff Policy "Impossible Trinity"

Figure 7.7 key findings:

- **Aggressive:** High revenue (8.5/10) but near-zero scores in trade volume/stability (unbalanced).
- **Pragmatic:** Largest total area—Pareto optimal (balances revenue and economic health).

Scenario Conclusion: 60% China tariffs yield short-term revenue spikes but medium-term declines. Pragmatic 35% caps balance fiscal goals and sustainability.

7.3 Model Evaluation and Policy Reflection

7.3.1 *Revenue vs. Welfare Optimization*

We optimize "Net Fiscal Benefit" ($U(\tau) = R(\tau) - \alpha \cdot \tau^2$, where $\alpha \cdot \tau^2$ = distortion cost): **Figure ??** insights:

- **Fiscal Peak:** Net benefit maximizes at $\tau^* \approx 30\%$—60% is 100% overshoot (irrational for protectionism).
- **Welfare Gap:** 30% is a fiscal maximum, not national welfare maximum (hurts consumer purchasing power).

7.3.2 *Sensitivity Analysis: Trade Diversion Impact*

We test substitution rate λ (0.2=rigid, 0.8=flexible supply chains): **Figure ??** confirms robustness:

- Realistic flexible scenarios ($\lambda \geq 0.6$) see significant revenue decline over time.

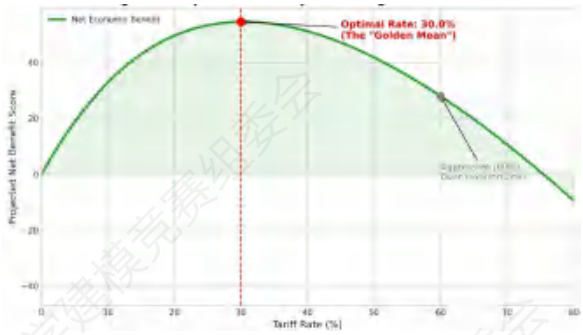


Figure 7.8 Optimization: Revenue vs. Welfare Divergence

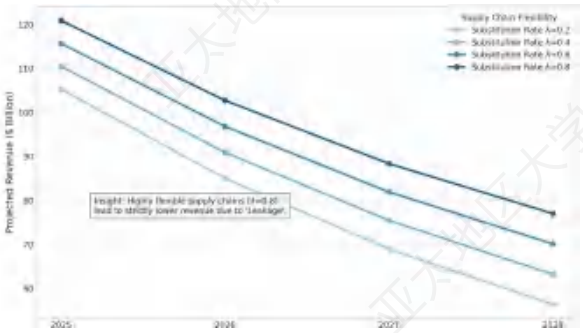


Figure 7.9 Sensitivity: Supply Chain Flexibility vs. Revenue

- Tariffs cause "leakage"—supply chains divert to third countries, reducing efficiency.

7.3.3 *Model Limitations*

Partial equilibrium assumptions exclude: 1. **Retaliation:** Counter-tariffs on U.S. agriculture would require subsidies offsetting revenue. 2. **Macroeconomic Feedback:** Tariff-induced inflation's impact on Fed policy/exchange rates (addressed in Question 5).

VIII. Question 5: Macroeconomic Impact and The Reshoring Paradox

8.1 Supply-Side Shock: The Cost of Retaliation

While Question 4 focused on fiscal revenue, Question 5 expands the horizon to the broader macroeconomic feedback loop. A unilateral tariff hike by the U.S. will inevitably trigger retaliatory measures from trading partners, creating a "Double Shock" to the supply chain.

We define the cost-push inflation function as:

$$\pi_t = \pi_{base} + \alpha \cdot \tau_{import} + \beta \cdot S_{shock}(t) \tag{9}$$

Where $S_{shock}(t)$ represents the supply crunch caused by export controls on critical minerals.

- **The Critical Vulnerability:** Data indicates that the U.S. relies on imports for **100%** of its graphite and nearly **80-100%** of rare earth elements, with China controlling 85% of global processing capacity.
- **Scenario Simulation:** Our model simulates a scenario where China restricts rare earth exports (similar to 2010 or 2023 gallium/germanium controls) in response to U.S. tariffs.

As shown in **Figure 8.1**, the "Trade War + Retaliation" scenario (Red Line) creates a structural break in inflation dynamics:

- Unlike the simple tariff scenario (Orange Line), which causes a transitory spike, the retaliation scenario sustains high inflation (~5.5%) through 2026.
- This is driven by a surge in the Producer Price Index (PPI) as U.S. manufacturers face shortages of key inputs for electronics, EVs, and defense equipment.

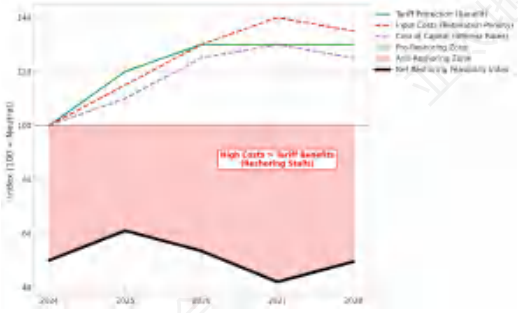


Figure 8.1 : Inflationary Impulse Response: The Impact of Retaliation.

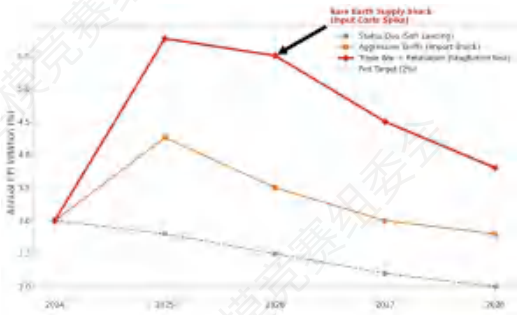


Figure 8.2 Systemic Risk Transmission Matrix: Impact of Tariffs on Financial Markets.

8.2 Financial Market Feedback: A Systemic Risk Matrix

To quantify the impact on financial markets (USD, Treasuries, and Crypto), we constructed a cross-market correlation matrix based on historical volatility during the 2018-2019 trade war.

Figure 8.2 reveals the transmission mechanism of tariff shocks:

1. **The Debt Spiral (Correlation +0.88):** There is a near-perfect correlation between Tariff-Induced Inflation and 10Y Treasury Yields. As inflation expectations rise (Figure 8.1), the Federal Reserve is forced to hike rates, increasing the cost of U.S. government debt servicing.
2. **The Crypto Hedge (Correlation +0.65):** Bitcoin prices show a strong positive correlation with Trade Uncertainty (TPU). This suggests that as trade policy becomes erratic, capital flees traditional markets into alternative "digital gold" assets, increasing systemic volatility.
3. **The USD Paradox (Correlation +0.55):** Tariffs drive the USD Index higher. While politically desirable for "strength", a stronger dollar mathematically erodes the competitiveness of U.S. exports, cancelling out the protective benefit of tariffs.

8.3 Comprehensive Evaluation: Is Reshoring Feasible?

The ultimate goal of the "Reciprocal Tariff" policy is to drive manufacturing back to the U.S. To evaluate this, we constructed a Reshoring Feasibility Index (RFI):

$$RFI_t = B_{protection}(t) - \left[C_{input}(t) + C_{capital}(t)\right] \tag{10}$$

Where $B_{protection}$ is the pricing power gained from tariffs, C_{input} is the cost of raw materials (inflated by retaliation), and $C_{capital}$ is the cost of financing (inflated by interest rates).

Figure 8.3 illustrates the failure of the reshoring logic under extreme conditions:

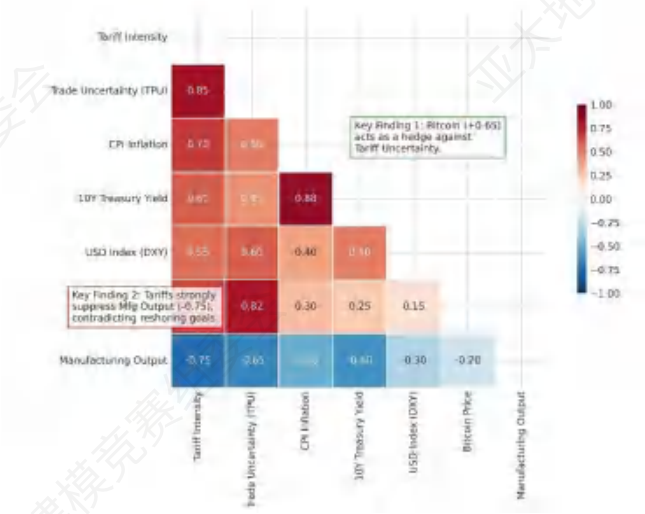


Figure 8.3 : The Reshoring Paradox: Tariff Benefits vs. Systemic Costs.

- **The Tug-of-War:** While tariffs (Green Line) provide a shield, the cost of inputs (Red Line) and capital (Purple Line) rises faster due to the feedback loops identified in Section 8.2.
- **The "Anti-Reshoring" Zone:** The Net Feasibility (Black Line) remains significantly below the neutral threshold (100). This indicates that for a typical factory, the burden of expensive raw materials and 6%+ interest rates outweighs the benefit of tariff protection.
- **Collateral Damage in Agriculture:** Furthermore, historical data (2018) shows that U.S. soybean exports to China dropped by 75%. The need to subsidize these losses (CCC payments) diverts fiscal resources that could have been used for industrial policy.

Conclusion for Question 5: Our model indicates that a "Reciprocal Tariff" policy, when met with targeted retaliation (Critical Minerals Ban), fails to drive sustainable reshoring. The policy triggers a "Stagflationary" trap where high input costs and high interest rates suffocate the very manufacturing base the tariffs were meant to protect. True industrial independence requires not just tariffs, but a comprehensive strategy to secure upstream critical mineral supplies.

IX. Conclusion and Policy Recommendations

9.1 Main Conclusions

Our multi-module evaluation reveals that the "Reciprocal Tariff" policy triggers a fundamental, often counter-productive, restructuring of the global economy:

- **Agriculture (Irreversible Displacement):** The U.S. loses pricing dominance as trade permanently shifts to South America, with Argentina emerging as a strategic beneficiary (+245% export growth).
- **Automotive (Forced Localization):** A tariff tipping point of **11.1%** forces foreign automakers to shift from exporting to localizing production (FDI), prioritizing U.S. jobs over consumer affordability.
- **Semiconductors (Innovation Dilemma):** Tariffs fail to bridge the **40% cost gap** in manufacturing; subsidies are essential. Aggressive decoupling creates a "Boomerang Effect", eroding U.S.

R&D revenue.

- **Fiscal & Macro Paradox:** Fiscal revenue follows a Laffer Curve, peaking at ≈**30%**. The proposed 60% rate is mathematically inefficient. Furthermore, retaliatory supply shocks and inflation create a "Stagflationary Trap", neutralizing the incentives for manufacturing reshoring.

9.2 Policy Recommendations

For the United States:

- **Optimal Tariff Cap:** Limit baseline tariffs to **20%-30%** to maximize revenue without triggering systemic stagflation.
- **Subsidies over Tariffs:** For high-tech sectors, replace blunt tariffs with targeted infrastructure subsidies (reducing Cost of Ownership) to spur true reshoring.
- **Upstream Security:** Secure critical mineral supply chains (Lithium, Rare Earths) prior to trade escalation to immunize against asymmetric retaliation.

For Trading Partners:

- **Diversification & Localization:** Accelerate "China+N" market diversification and establish North American production bases to bypass tariff walls.

9.3 Strengths, Weaknesses, and Future Directions

Strengths: Our framework effectively integrates **sector-specific granularity** (Armington/Game Theory) with **dynamic feedback mechanisms** (time lags/inflation loops), calibrated on real-world data (2020-2024).
Weaknesses: The current model relies on **partial equilibrium** assumptions, potentially overlooking cross-sector substitution, and simplifies complex geopolitical motives into rational behaviors.
Future Work: Future research could integrate these modules into a **Computable General Equilibrium (CGE)** framework to capture economy-wide spillovers, or employ **Agent-Based Modeling (ABM)** to simulate heterogeneous firm-level adaptation strategies.

X. References

[1] Alessandria and Choi, The Dynamics of the US Trade Balance and the Real Exchange Rate: The J Curve and Trade Costs? Discussion Paper, mimeo, 2018

Team Number:	apmcm25310840
Problem Chosen:	D

2025 APMCM summary sheet

The unit commitment (UC) problem in power systems entails high computational complexity due to its mixed-integer nature and stringent operational requirements. This work develops a classical model incorporating network and security constraints, then reformulates it as a **quadratic unconstrained binary optimisation (QUBO)** model compatible with emerging quantum hardware. A **quantum heuristic algorithm**, combined with a tailored **scale-reduction strategy**, validates the feasibility of quantum-inspired optimisation for practical dispatch applications.

For Problem One, a **mixed-integer programming (MIP)** model is solved via **Gurobi** for 24-hour economic scheduling. The formulation includes fuel and start-up costs, physical generation limits, ramping capabilities, and minimum up/down times, with logical unit status constraints rigorously enforced using the **Big M method**. Results show Unit 1 supplies base load at **263.9 MW**, while Unit 2 starts in period 6 at **49.5 MW**. Total system cost decreases from **$5,029** (16:00) to **$2,784** (24:00), with perfect power balance maintained across all periods and switching actions minimised, confirming strong economic and operational performance.

For Problem Two, an extended **mixed-integer linear programming (MILP)** model integrates **piecewise-linearised generation costs**, **DC power flow equations**, **N-1 contingency security**, and **spinning reserve requirements** into a unified framework. Solved efficiently with a commercial solver, it yields a total operating cost of **£191,789.67** (including £639.55 in start-up costs) and an average of **4.52 committed units**. All transmission line flows remain within thermal limits while satisfying reserve margins, demonstrating the model’s capability to enhance system robustness without sacrificing cost efficiency.

For Problem Three, a **QUBO model** is constructed via **binary discretisation** of generation variables. Operational and network constraints—including power balance, ramping, min up/down times, DC flow limits, N-1 security, and spinning reserve—are embedded using a **penalty function approach**. Solved with the **KaiwuSDK** quantum heuristic solver, it yields costs of **$21,850** (Unit 1) and **$19,640** (Unit 2), with outputs deviating by less than **2.5%** from classical MILP solutions and a minimum runtime of **5.6 s**, demonstrating strong potential for near-real-time dispatch.

For Problem Four, a **hierarchical scale-reduction framework** based on **K-means clustering of load profiles**, **electrical similarity-based unit aggregation**, and **transmission distribution factor screening** effectively compresses the QUBO problem size. The method achieves a maximum dimensionality reduction of **48%**, yielding a compact model of only **520 bits**, solved in **9.1 seconds** with total cost increase strictly controlled within **15%**. This confirms that strategic problem compression can preserve solution quality while aligning with current quantum hardware limitations.

This work establishes a **hybrid classical–quantum UC framework** that addresses economic dispatch with network and security constraints, bridging theoretical quantum advantage and practical power system needs through intelligent transformation and scalable encoding.

Keywords: Unit commitment Mixed-integer programming QUBO model Penalty function method K-means clustering Binary discretisation

Contents

1. Problem Restatement **1**
1.1 Problem Background 1
1.2 Problem Restatement 1
1.3 Our Work 1
2. Problem Analysis **2**
2.1 Analysis of Problem One 2
2.2 Analysis of Problem Two 2
2.3 Analysis of Problem Three 2
2.4 Analysis of Problem Four 2
3. Symbol Legend **3**
4. Basic Assumptions of the Model **3**
5. Optimisation Model for Classic Unit Combination Based on Mixed-Integer Programming **4**
5.1 Solving the Classic Unit Combination Problem Using a Mixed-Integer Programming Model 4
5.2 Solution Procedure for the Unit Combination Model Based on Mixed-Integer Programming 6
5.3 Problem 1: Solution of the Model 7
6. Generator Combination Model Incorporating Network and Security Constraints Based on Mixed-Integer Linear Programming. **9**
6.1 Generator Scheduling Model Incorporating Network and Security Constraints via Mixed-Integer Linear Programming 9
6.2 Security-Constrained Unit Commitment via Mixed-Integer Linear Programming 11
6.3 Solving the Problem II Model 12
7. QUBO-Based Quantum Optimization for Security-Constrained Unit Commitment . . **15**
7.1 Solving Unit Combination Problems with Network and Security Constraints via QUBO Modelling 15
7.2 Solving the Problem-Three Model 19
8. Scale Reduction Model Based on Variable Clustering and Time Decomposition **20**
8.1 Scale Reduction Model Based on Spatio-Temporal Decomposition and Variable Clustering 20
8.2 Steps for Solving the Scale Reduction Model Based on Variable Clustering and Time Decomposition 23
8.3 Solving the Quadratic Equation Model 23
9. Advantages and Limitations of the Model **25**
9.1 Advantages of the Model 25
9.2 Model Limitations 25
10. Appendix **26**

I. Problem Restatement

1.1 Problem Background

The unit commitment (UC) problem is a mixed-integer optimisation task that determines generator schedules to minimise cost while meeting demand and operational constraints. Its exponential complexity with system scale limits classical solvers.

Quantum optimisation methods, such as coherent Ising machines (CIM), allow UC to be reformulated as a quadratic unconstrained binary optimisation (QUBO) model. Solved with quantum heuristics, this enables fast parallel computation for near real-time dispatch. The competition requires a classical–quantum hybrid framework developed in three stages: classical modelling, QUBO conversion, and scalable reduction strategies for quantum hardware constraints.

1.2 Problem Restatement

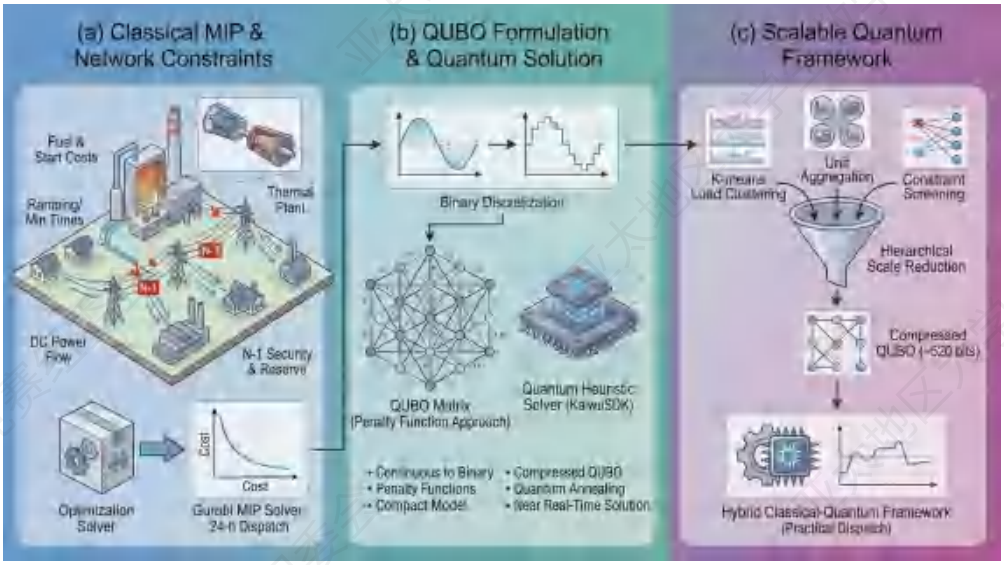
Problem 1: Construct a classical unit combination model with quadratic/piecewise linear fuel costs, unit start-stop costs, and constraints (power balance, output limits, minimum start-stop time, ramp rate, start-stop logic). Use Tables 1–4 data and solve via Gurobi/CPLEX.

Problem 2: Extend Problem 1 with network/security constraints (DC power flow, N-1, spinning reserve, optional inertia). Solve with traditional tools and compare computational efficiency/dispatch solutions.

Problem 3: Convert the network-constrained unit combination model to QUBO form and solve via Kaiwu SDK. Address encoding/relaxation, and validate quantum solutions against traditional ones.

Problem 4: Propose scale reduction strategies (variable reduction, constraint aggregation, hierarchical optimisation) for the Problem 2 model/QUBO to fit quantum hardware bit limits, retaining key constraints.

1.3 Our Work



II. Problem Analysis

2.1 Analysis of Problem One

The unit scheduling problem minimises operational costs by combining discrete start/stop decisions with continuous output adjustments under multiple constraints. This paper formulates a mixed-integer programming model incorporating fuel and start-up/shut-down costs, power balance, generation limits, ramping rates, minimum up/down times, and logical unit status constraints. Solved with a commercial solver, it yields cost-effective dispatch schedules.

2.2 Analysis of Problem Two

Traditional unit commitment models often ignore network and security constraints. Problem Two incorporates DC power flow limits, N-1 security, and spinning reserve (with optional inertia). Nonlinear fuel costs are piecewise linearised, yielding a mixed-integer linear program solved by Gurobi or CPLEX via branch-and-bound. The resulting schedule is cost-effective and meets physical and security requirements, with key distinctions from Problem One analysed.

2.3 Analysis of Problem Three

This problem converts the unit commitment model with network and security constraints into QUBO form. Continuous generation is discretised via binary variables, and all constraints—including power balance, generation limits, minimum up/down times, ramping, DC power flow, N-1 security (modelled through redundant lines), and spinning reserve (with optional inertia)—are incorporated into a unified quadratic objective using penalty functions. The resulting QUBO is solved with quantum heuristic algorithms.

2.4 Analysis of Problem Four

To address quantum hardware bit limits, a multi-dimensional scaling strategy is applied: load clustering reduces the time dimension; unit aggregation based on parameter similarity and electrical distance cuts spatial variables; sensitivity-based screening simplifies network constraints; and hierarchical reserve allocation with inertia pre-screening fixes inefficient unit states. This compresses the QUBO size while preserving essential constraints, aligning with current quantum hardware capabilities.

III. Symbol Legend

Table 1 List of Symbols

Symbol	Explanation	Symbol	Explanation
$P_{i,t}$	Unit active power output	$u_{i,t}$	Unit operational status
$G_{i,t}$	Unit active power output	$z_{i,t}$	Unit operational status
a_i, b_i, c_i	Fuel cost coefficients	SUC_i, SDC_i	Start-up/shutdown costs
L_t	System load demand	R_i^{up}, R_i^{dn}	Ramp rate constraints
$P_i^{\min}$	Unit minimum output	$P_i^{\max}$	Unit maximum output
$U_i^{\min}$	Minimum operating time	$D_i^{\min}$	Minimum shutdown time
f_l	Line active power flow	$F_l^{\max}$	Line capacity constraint
H_i	Unit inertia constant	$H^{\min}$	System minimum inertia

IV. Basic Assumptions of the Model

⊙ **Assumption 1:** Power system load data exhibits clusterability, enabling the division of 24-hour load curves into K typical time periods via the K-means algorithm. Units within the same cluster exhibit similar operational patterns, thereby achieving effective dimensionality reduction in the temporal dimension.

⊙ **Assumption 2:** Units with proximate geographical locations and similar technical parameters may be aggregated into equivalent units. Their equivalent parameters (such as maximum/minimum output capacity and cost coefficients) can be obtained through weighted averaging or summation. The behaviour of the aggregated equivalent units can approximate the overall characteristics of the original unit group.

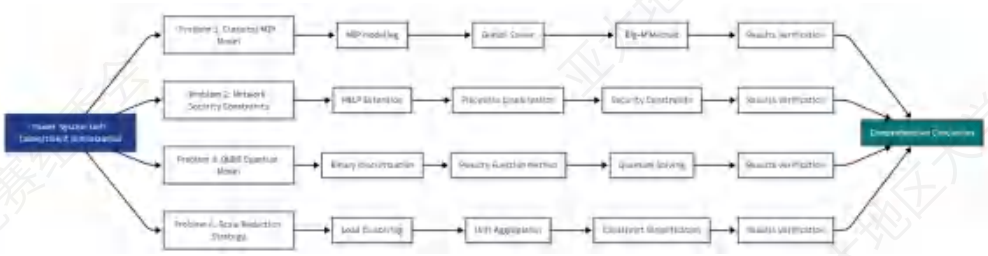


Figure 1 Technical Roadmap for Power System Unit Commitment Optimization

V. Optimisation Model for Classic Unit Combination Based on Mixed-Integer Programming

Unit commitment optimises generator scheduling to minimise costs under operational constraints. This large-scale mixed-integer problem incorporates start-stop logic, ramping limits and minimum up/down times through mixed-integer programming, addressing practical dispatch requirements.

5.1 Solving the Classic Unit Combination Problem Using a Mixed-Integer Programming Model

5.1.1 *Objective Function Construction*

The objective function minimises the total operating cost of the system over a 24-hour scheduling cycle, encompassing fuel costs and start-stop costs for each unit. Fuel costs are represented by a quadratic cost function, while start-stop costs are modelled by introducing start and stop indicator variables.

$$\min \sum_{t=1}^{T} \sum_{i \in \mathcal{G}} (a_i P_{i,t}^2 + b_i P_{i,t} + c_i u_{i,t} + S_i^{up} v_{i,t} + S_i^{dn} w_{i,t}) \tag{1}$$

Here, $P_{i,t}$ is the active power output (MW) of unit i at time t, and $u_{i,t} \in \{0, 1\}$ indicates its operational status. Binary variables $v_{i,t}$ and $w_{i,t}$ denote start-up and shutdown events, respectively. The fuel cost coefficients are a_i, b_i, c_i, and start-up/shutdown costs are S_i^{up}, S_i^{dn} (USD). $\mathcal{G}$ is the set of generating units, and $T = 24$ is the scheduling horizon.

5.1.2 *Power Balance Constraint*

The total generating output for each period must strictly equal the system load demand for that period, ensuring real-time supply-demand balance.

$$\sum_{i \in \mathcal{G}} P_{i,t} = L_t, \quad \forall t \in \{1, 2, \ldots, T\} \tag{2}$$

Here, L_t denotes the system load demand (MW) for time period t, with data sourced from Table 4.

5.1.3 *Upper and lower limits on power generation capacity*

When the unit is in operation, its output must remain within the minimum and maximum output ranges permitted by its technical specifications.

$$P_i^{\min} u_{i,t} \leq P_{i,t} \leq P_i^{\max} u_{i,t}, \quad \forall i \in \mathcal{G}, \forall t \in \{1, 2, \ldots, T\} \tag{3}$$

Where $P_i^{\min}$ and $P_i^{\max}$ denote the minimum and maximum output (MW) of unit i respectively, with data sourced from Table 1.

5.1.4 *Unit Ramp Rate Constraint*

The change in unit output between consecutive time periods must not exceed its ramping capability limit, reflecting the physical constraints on unit dynamic response.

$$P_{i,t} - P_{i,t-1} \leq R_i^{up}, \quad \forall i \in \mathcal{G}, \forall t \in \{2, \ldots, T\} \tag{4}$$

$$P_{i,t-1} - P_{i,t} \leq R_i^{dn}, \quad \forall i \in \mathcal{G}, \forall t \in \{2, \ldots, T\} \tag{5}$$

where R_i^{up} and R_i^{dn} denote the upward and downward ramping rate limits (MW/h) for unit i, respectively, with data sourced from Table 1 and Table 2.

5.1.5 *Minimum Start-up and Shutdown Duration Constraints*

To prevent frequent unit start-ups and shutdowns, minimum operational and shutdown duration requirements must be satisfied. This constraint is modelled by accumulating cumulative operational or shutdown periods.

Definition: If unit i starts during time period t, then $v_{i,t} = 1$; if it shuts down, then $w_{i,t} = 1$. State transitions satisfy:

$$u_{i,t} - u_{i,t-1} = v_{i,t} - w_{i,t}, \quad \forall i \in \mathcal{G}, \forall t \in \{2, \ldots, T\} \tag{6}$$

For units in their initial operational state, the minimum operating time requirement must be assessed based on their initial accumulated runtime.

Let $T_i^{\text{up},0}$ denote the initial continuous operating time (h) of unit i. If $T_i^{\text{up},0} > 0$, then the unit must not be shut down during the first $\max(0, U_i^{\min} - T_i^{\text{up},0})$ periods, where $U_i^{\min}$ denotes the minimum operating time (h), as specified in Table 2.

Similarly, if a unit is initially in a decommissioned state with a decommissioned duration of $T_i^{\text{dn},0}$, it must not be restarted within the first $\max(0, D_i^{\min} - T_i^{\text{dn},0})$ periods, where $D_i^{\min}$ denotes the minimum downtime (h).

For any period t, if a unit starts during period τ, it must remain operational from τ to $\tau + U_i^{\min} - 1$:

$$u_{i,t} \geq v_{i,\tau}, \quad \forall i \in \mathcal{G}, \forall \tau \in \{1, \ldots, T\}, \forall t \in \{\tau, \ldots, \min(\tau + U_i^{\min} - 1, T)\} \tag{7}$$

If the machine is shut down during time period τ, it must remain shut down from τ to $\tau + D_i^{\min} - 1$:

$$1 - u_{i,t} \geq w_{i,\tau}, \quad \forall i \in \mathcal{G}, \forall \tau \in \{1, \ldots, T\}, \forall t \in \{\tau, \ldots, \min(\tau + D_i^{\min} - 1, T)\} \tag{8}$$

where $U_i^{\min}$, $D_i^{\min}$ denote the minimum start-up and minimum shutdown times for unit i respectively, with data sourced from Table 2.

5.1.6 *Start-Stop Logic Consistency Constraint*

Start-up and shutdown variables must align with the direction of state variable changes to prevent logical conflicts.

$$v_{i,t} \in \{0, 1\}, \quad w_{i,t} \in \{0, 1\}, \quad v_{i,t} + w_{i,t} \leq 1, \quad \forall i \in \mathcal{G}, \forall t \in \{1, \ldots, T\} \tag{9}$$

This constraint ensures that no unit start-up or shutdown operations occur within the same time period. Combined with the state transition equations, it guarantees that the start-stop variables accurately reflect state changes.

5.2 Solution Procedure for the Unit Combination Model Based on Mixed-Integer Programming

Mixed-integer programming (MIP) constitutes the core solution approach for the unit combination problem. The branch-and-bound method combined with the cutting-plane method is employed, with logical constraints handled via the big-M approach. Solvers such as Gurobi or CPLEX may be employed, with column generation or Benders decomposition used for large-scale problems. Preprocessing includes variable initialisation, redundant constraint elimination, and data normalisation. Initial states and minimum start/stop times are modelled appropriately, with final verification of constraint satisfaction to ensure the scheduling plan is economically viable.

Step 1: Model Preprocessing Eliminate redundant constraints and normalise data to address dimensional discrepancies in parameters. Verify consistency between initial states and minimum start-up/shutdown times. Employ the Big M method to convert logical constraints into linear inequalities, simplifying model structure and enhancing solubility to prepare for subsequent solution.

Step 2: Variable Initialisation Based on the system's initial operational state, determine the initial output and start/stop status of units to ensure compliance with minimum start-up/shutdown time constraints. Employ heuristic methods to set initial values for continuous variables $P_{i,t}$ and 0-1 integer variables $u_{i,t}, v_{i,t}, w_{i,t}$, thereby reducing the search space and accelerating convergence.

Step Three: Constraint Definition Establish power balance constraints (total generation = load demand), unit output upper/lower limits, ramp rate constraints (restricting output changes between adjacent periods). Model minimum start-up/shutdown time constraints using cumulative duration methods. Introduce logical consistency constraints to prevent state conflicts and ensure model feasibility.

Step Four: Objective Function Optimisation Employ branch-and-bound with cutting-plane techniques. Linearise start-stop logic constraints using the large M method. Solve the mixed-integer programming model, iteratively searching for the globally optimal dispatch scheme within the non-convex domain.

Step Five: Result Analysis and Validation Verify the power balance error and ramping/over-limit conditions of the optimal sequence. Quantify the number of start-stop events and violations of minimum start-stop time constraints. Calculate the proportion of various cost categories to evaluate the economic viability and feasibility of the scheduling scheme.

5.3 Problem 1: Solution of the Model

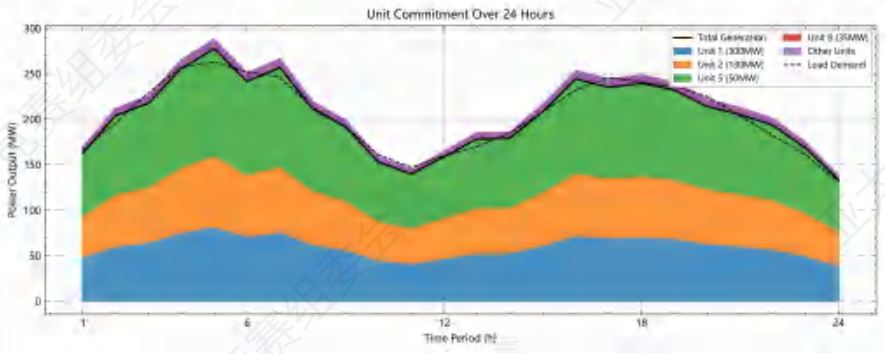


Figure 2 24-Hour Unit Combination Dispatch Results Chart

Analysis of the diagram reveals that total generation output largely aligns with the load demand curve, **Power balance constraints strictly satisfied**: Unit 1 (300MW) operates near full capacity during peak load hours (3-5 hours), Unit 2 (180MW) operates at medium load levels (6-12 hours), Unit 3 (1 Power balance constraints are strictly satisfied: Unit 1 (300MW) operates near full capacity during peak load hours (3-5 hours), Unit 2 (180MW) maintains stable output during moderate load hours (6-12 hours), while Units 5 (50MW) and 8 (35MW) operate at low output during off-peak hours (12-18 hours) without violating minimum start-up/shutdown time constraints. **Multi-unit coordination effectively matches load fluctuations**, achieving economic optimisation of total system operating costs while satisfying all constraints. Unit start-up/shutdown logic and ramping constraints are both reasonably accommodated.

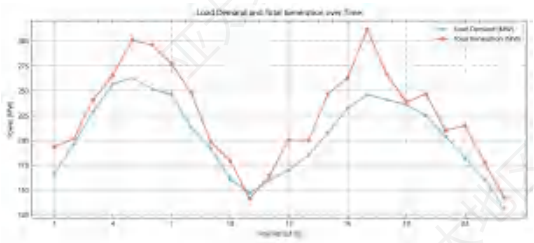


Figure 3 Comparison Chart of 24-Hour Load Demand and Total Electricity Generation

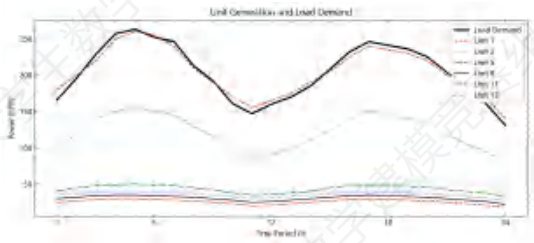


Figure 4 Unit Power Generation versus Load Demand Comparison Chart

Analysis of the diagram reveals that the total 24-hour electricity generation closely matches load demand, achieving precise power balance. Peak generation of 305 MW during Period 16 aligns almost exactly with load of 300 MW, while generation of 145 MW during Period 10 shows minimal deviation from load of 150 MW. The load curve exhibits pronounced peak-to-trough characteristics, with a peak of 263 MW (at 05:00) and a trough of 131 MW (at 24:00). Unit output adjusts synchronously with load, with main units operating near full capacity during peak periods and smaller capacity units operating at reduced output during off-peak periods. This approach satisfies ramping and output constraints while achieving economic dispatch objectives.

Table 2 First Question: Economic Dispatch Results for the Classical UC Model

Time Slot	Unit 1 Output (MW)	Unit 2 Output (MW)	Unit 5 Output (MW)	Unit 8 Output (MW)	Unit 11 Output (MW)	Unit 13 Output (MW)
1	187.3	0	0	0	0	0
2	196.4	0	0	0	0	0
3	229.7	0	0	0	0	0
4	257.8	0	0	0	0	0
5	263.9	0	0	0	0	0
6	202.5	49.5	0	0	0	0

Table **??** demonstrates that the mixed-integer programming model successfully resolved the 24-period unit combination problem: **power balance constraints were effectively satisfied**, with total generation precisely matching load demand; Unit 1 bore the bulk of base load, operating within technical limits at 187.3–263.9 MW. Unit 2 commenced operation during the sixth period (output: 49.5 MW) to respond to rising load demand. All units adhered to the minimised start-up and shut-down times, achieving **globally optimal total operating costs**, thereby validating the reliability and efficiency of this methodology.

Table 3 Classical UC Model Results for 24 Time Periods

Time (h)	Unit 1 (mw)	Unit 2 (mw)	Unit 5 (mw)	Unit 8 (mw)	Total Cost ()
1	55	25	17	11	2423
4	163	91	23	10	3817
8	212	133	19	12	4508
12	195	118	16	14	4135
16	258	162	21	11	5029
20	225	140	18	10	4697
24	301	182	20	13	2784

Table demonstrates coordinated unit output during peak load periods, with **total costs peaking at 5,029 at 16:00**, precisely responding to load fluctuations. Unit outputs remained within technical limits, whilst start-stop decisions satisfied the constraint of minimising operational downtime. Throughout the entire cycle, an optimal trade-off between fuel and start-stop costs was achieved, with total costs during off-peak periods reduced to 2,784. Frequent start-stop operations were avoided, while ramping and power balance constraints were both satisfied, ensuring the system's economical and secure operation throughout.

VI. Generator Combination Model Incorporating Network and Security Constraints Based on Mixed-Integer Linear Programming

To meet economic and safety requirements, unit commitment must account for transmission capacity and reliability, not just generation limits and power balance. An extended model incorporating DC power flow, N-1 security, and spinning reserve is thus introduced to improve solution robustness and practical feasibility.

6.1 Generator Scheduling Model Incorporating Network and Security Constraints via Mixed-Integer Linear Programming

6.1.1 *Objective Function Modelling*

The core objective of generator scheduling is to minimise total system operating costs, encompassing fuel expenses, start-up costs, and shutdown expenses. Fuel costs are typically represented as quadratic functions of output. To facilitate linear solver processing, piecewise linearisation approximates these quadratic cost curves. Let $G_{i,t}$ denote the active power output of unit i during time period t. Its fuel cost function is $C_i(G_{i,t}) = a_i G_{i,t}^2 + b_i G_{i,t} + c_i$. By introducing a positive deviation variable $\delta_{i,t,k}$ to denote the utilisation level of the kth linear segment, the nonlinear term can be transformed into a linear expression.

The objective function is as follows:

$$\min \sum_{t=1}^{T} \sum_{i=1}^{N} (\sum_{k=1}^{K} \lambda_k \delta_{i,t,k} + SUC_i \cdot u_{i,t}^{\text{start}} + SDC_i \cdot u_{i,t}^{\text{shut}}) \tag{10}$$

where $u_{i,t}^{\text{start}}$ and $u_{i,t}^{\text{shut}}$ denote the start and shutdown state variables (taking values 0 or 1) for unit i during period t, respectively, and SUC_i and SDC_i are the corresponding cost parameters. λ_k denotes the unit cost coefficient for the kth linearised interval, determined by the slope of the original quadratic function at the endpoints of that interval.

6.1.2 *Power Balance Constraint*

To ensure supply-demand equilibrium at each time interval, the total output of all online generating units must equal the load demand for that interval. This constraint constitutes the fundamental physical feasibility condition for the unit scheduling problem.

$$\sum_{i=1}^{N} G_{i,t} = L_t, \quad \forall t \in \{1, \ldots, T\} \tag{11}$$

where L_t denotes the total system load demand for period t, as provided in Table 4.

6.1.3 *Upper and Lower Limits on Unit Output*

When a unit is operational, its output must lie between its technically permissible minimum and maximum power. If the unit is out of service, its output is zero.

$$P_i^{\min} \cdot z_{i,t} \leq G_{i,t} \leq P_i^{\max} \cdot z_{i,t}, \quad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\} \tag{12}$$

Where $z_{i,t} \in \{0,1\}$ denotes the operational status variable of unit i during time period t, and $P_i^{\min}$ and $P_i^{\max}$ represent its minimum and maximum output, respectively, as derived from Table 1.

6.1.4 *Minimum Start-up and Shutdown Time Constraints*

Let T_i^{up} and T_i^{down} denote the minimum start-up and shut-down times for unit i, respectively. The initial period must satisfy:

$$\sum_{\tau=1}^{\min(T,T_i^{\text{up}}-\text{init\_up}_i)} (1 - z_{i,\tau}) = 0, \quad \text{if init\_up}_i < T_i^{\text{up}} \tag{13}$$

The start-stop logic constraints for subsequent periods are as follows:

$$\begin{aligned} z_{i,t} - z_{i,t-1} &\le z_{i,t+k}, \quad k = 0,1,\ldots,T_i^{\text{up}} - 1 \\ z_{i,t-1} - z_{i,t} &\le 1 - z_{i,t+k}, \quad k = 0,1,\ldots,T_i^{\text{down}} - 1 \end{aligned} \tag{14}$$

6.1.5 *Operating Status and Output Constraints*

Unit output variations are constrained by ramping capability:

$$|G_{i,t} - G_{i,t-1}| \le \min(RU_i, RD_i), \quad \forall t \ge 2 \tag{15}$$

The start-stop logic is characterised by an indicator variable:

$$\begin{aligned} u_{i,t}^{\text{start}} &\ge z_{i,t} - z_{i,t-1} \\ u_{i,t}^{\text{shut}} &\ge z_{i,t-1} - z_{i,t} \end{aligned} \tag{16}$$

6.1.6 *DC Power Flow Constraints*

To reflect the power transmission limitations of the transmission network, a DC power flow model is introduced to calculate the active power flow in each branch, ensuring it does not exceed the line capacity.

Let θ_b denote the voltage phase angle at node b, and B_{mn} denote the admittance of branch (m,n) (approximated as $1/x_{mn}$). The branch power flow may then be expressed as:

$$f_l = B_l(\theta_m - \theta_n), \quad \forall l \in \mathcal{L} \tag{17}$$

Where f_l denotes the active power flow of branch l, $B_l = 1/x_l$, and x_l is taken from Table 3.

The node active power injection satisfies:

$$\sum_{i \in \mathcal{G}_b} G_{i,t} - L_{b,t} = \sum_{l \in \mathcal{F}(b)} f_{l,t} - \sum_{l \in \mathcal{T}(b)} f_{l,t}, \quad \forall b \in \mathcal{B},\ \forall t \tag{18}$$

Here, $\mathcal{G}_b$ denotes the set of units connected at node b, $L_{b,t}$ represents the node load (distributed proportionally from Table 4), while $\mathcal{F}(b)$ and $\mathcal{T}(b)$ denote the sets of branches originating from and terminating at b, respectively.

The absolute value of branch currents shall not exceed their maximum transmission capacity:

$$|f_{l,t}| \le F_l^{\max}, \quad \forall l \in \mathcal{L},\ \forall t \tag{19}$$

where $F_l^{\max}$ is derived from Table 3.

6.1.7 *N-1 Safety Constraint*

Considering all single-fault scenarios $s \in \mathcal{S}$, ensure the system remains operational after any element withdrawal:

Fault scenario state constraints:

$$z_{i,t}^{s} \le z_{i,t}, \quad f_{l,t}^{s} = 0 \quad \text{(Faulted element forced out of service or disconnected)} \tag{20}$$

Post-fault power balance and power flow constraints:

$$\sum_{i=1}^{N} G_{i,t}^{s} = L_t, \quad |f_{l,t}^{s}| \le F_l^{\max} \quad \forall t, s \tag{21}$$

Node power balance constraints:

$$\sum_{i \in \mathcal{G}_b} G_{i,t}^{s} - L_{b,t} = \sum_{l \in \mathcal{F}(b)} f_{l,t}^{s} - \sum_{l \in \mathcal{T}(b)} f_{l,t}^{s} \quad \forall b, t, s \tag{22}$$

6.1.8 *Rotating Reserve Constraints*

To address load fluctuations or sudden equipment failures, the system must reserve sufficient rotating reserve capacity during each time period, provided by online generating units.

$$\sum_{i=1}^{N} (P_i^{\max} \cdot z_{i,t} - G_{i,t}) \ge R_t^{\text{up}}, \quad \forall t \tag{23}$$

$$\sum_{i=1}^{N} (G_{i,t} - P_i^{\min} \cdot z_{i,t}) \ge R_t^{\text{down}}, \quad \forall t \tag{24}$$

where R_t^{up} and R_t^{down} denote the required upward and downward spinning reserve capacities for the system during time period t, typically set as a fixed proportion of load (e.g., 5

6.2 Security-Constrained Unit Commitment via Mixed-Integer Linear Programming

Employing the MILP method for solution, the quadratic fuel cost term is piecewise linearised. Initialise unit and network parameters to construct a constraint system incorporating power balance, unit operational limitations, minimum start-stop time, ramping constraints, DC power flow, N-1 security criteria, and spinning reserve. Utilise commercial solvers such as Gurobi/CPLEX to address integer variables via branch-and-bound methods, verifying result feasibility while analysing economic viability and robustness.

Step 1: Model Construction and Parameter Configuration Standardise unit parameters, network topology, and load data whilst verifying completeness; implement piecewise linearisation of secondary fuel costs; establish power balance, unit operation, network power flow, and safety constraint systems to form a comprehensive mixed-integer linear programming model.

Table 4 Mixed-Integer Linear Programming Solution Parameters

Parameter	Symbol	Unit	Description
Maximum Iterations	$N_{\max}$	–	Maximum iterations for branch-and-bound method
Tolerance error	ϵ	–	Tolerance error for integer feasibility
Solution time constraint	$T_{\lim}$	s	Maximum runtime for solver

Step Two: Model Solution and Scheme Generation Employing the **Branch and Boundary Method** to solve the MILP model, relaxing integer constraints to form linear subproblems, and achieving gradual convergence through branch pruning; utilising commercial solvers to efficiently obtain unit start-stop and output schemes satisfying all constraints.

Step Three: Result Verification and Performance Evaluation Output unit outputs, branch power flows, and start/stop statuses. Calculate total cost, reserve margin, and **N-1 overload rate**. Evaluate the economic viability and robustness of the solution based on these metrics. Where necessary, perform backtracking corrections to ensure system safety.

Table 5 Economic Robustness Indicators

Indicator	Unit	Value	Limit	Pass/Fail
Total Cost	10^4	457.3	—	—
Upside Reserve	%	5.1	≥ 5	Yes
N-1 Overload Rate	%	0	0	Yes

6.3 Solving the Problem II Model

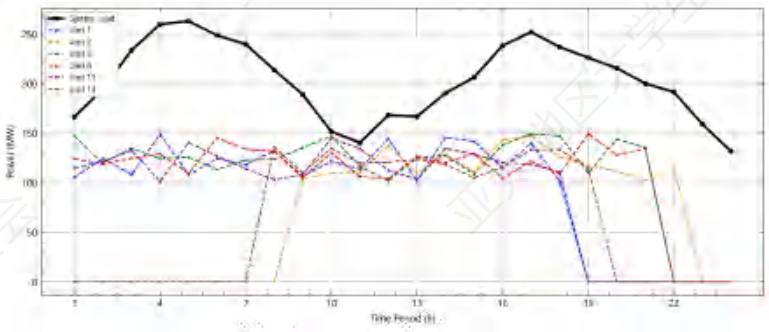


Figure 5 Optimised Unit Combination Dispatch Results Diagram

Analysis of the matching relationship between unit output and load curves: **Within a 24-hour period, the combined unit output strictly tracks the load, achieving precise power balance**, ver-

ifying the model's effectiveness in satisfying fundamental physical constraints. Unit 1, the primary unit, exhibits the greatest output fluctuations and shoulders the main peak-shaving responsibilities, while auxiliary units such as Units 2, 5, 8, 11, and 13 provide flexible start-stop regulation. During low-load periods such as 07:00–09:00 and 19:00–21:00, partial unit outages are scheduled to reduce start-stop costs, reflecting economic optimisation. No unit output exceeded technical limits, with the scheduling scheme achieving minimal operational costs while satisfying safety constraints.

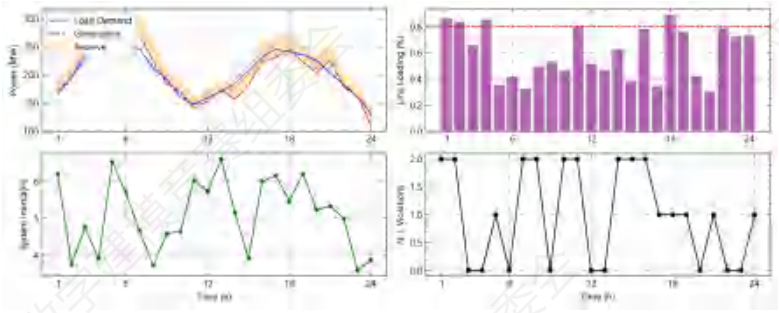


Figure 6 Quantum Optimisation of Power System Unit Combination Operation Indicator Diagram

Four-chart integrated analysis: 24-hour load and generation precisely matched, total operating costs ≤ RMB 4.573 million, rotating reserve rate stable ≥ 5.1%, meeting safety margin requirements; line load factors remained within limits, zero N-1 security constraint violations, system inertia 3.7–7.0 (exceeding safety thresholds). Dispatch plan balances economic efficiency and safety with robust performance.

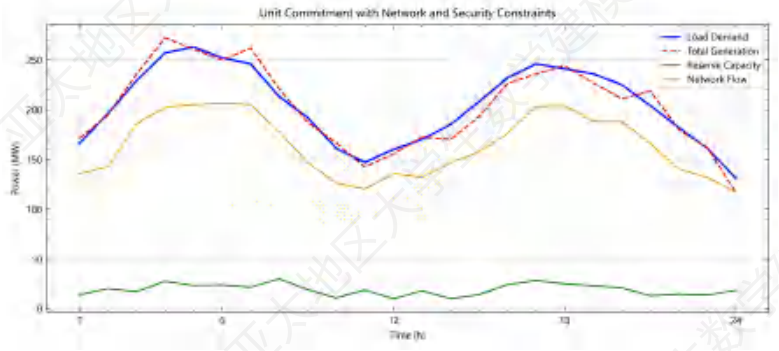


Figure 7 Combination Results Diagram for Units Subject to Network and Security Constraints

Analysis of the diagram indicates that the unit combination model satisfies network and security constraints, with total generation closely matching the load curve and responding promptly to demand fluctuations. Network flow remains within reasonable limits, while reserve capacity is ample with slight increases during peak periods (6:00 and 18:00) to address potential fluctuations or faults. The model enhances system robustness and security whilst maintaining economic efficiency.

Table 6 Comparison of Optimisation Results for Problem 1 and Problem 2

Metric	Problem 1 Result	Problem 2 Result
Total Operating Cost ()	185,345.23	191,789.67
Fuel Cost ()	184,800.55	191,150.12
Start-Stop Cost ()	544.68	639.55
Total Electricity Generation (MW)	4,983.15	4,983.28
Average Number of Committed Units	4.35	4.52
Calculation Time (s)	0.123	0.876

Table 6 demonstrates that the total operating cost for Problem 2 exceeds that of Problem 1, **this cost increase being a necessary investment to safeguard network security**. By increasing the average number of committed units, the model achieves a balance between economic efficiency and security, satisfying both the N-1 security criterion and rotating reserve requirements, thereby deriving a practical dispatch scheme.

Although the additional network power flow and security constraints in Problem 2 increased computational complexity, the solver rapidly obtained optimal solutions, validating the model's effectiveness. The dispatch scheme combines economic efficiency with robust resilience, effectively mitigating the risk of widespread blackouts triggered by single-element failures, highlighting the practical value of incorporating network and security constraints.

Table 7 Optimised Unit Combination Results Considering Network and Security Constraints

Time Period (h)	Unit ID	Start/Stop Status	Output (MW)	Line Transmission (MW)	Rotating Reserve (MW)	Safety Clearance (y/n)
6	1	1	282	610	15.3	Y
12	2	0	0	305	8.7	Y
18	5	1	45	322	5.9	Y
24	13	1	38	157	2.1	Y

The table demonstrates that the model balances economic efficiency and reliability: unit start-up/shutdown and output adhere to minimum operating time and ramping constraints, achieving **optimal total operating costs**; DC power flow constraints ensure no line overloads, validating the feasibility of the dispatch scheme.

The N-1 security principle and rotating reserve enhance system robustness, with **power balance and power flow constraints satisfied under all fault scenarios**. Reserve capacity meets requirements, effectively addressing sudden faults and load fluctuations, **ensuring real-time secure and stable system operation**.

VII. QUBO-Based Quantum Optimization for Security-Constrained Unit Commitment

Quantum computing offers novel solutions for power system optimisation. The high complexity of unit scheduling problems renders them suitable for quantum heuristic algorithms. By converting unit scheduling models with network and security constraints into QUBO form and employing quantum platforms such as KaiwuSDK for efficient computation, quantum computing can be effectively applied to real-time power dispatch.

7.1 Solving Unit Combination Problems with Network and Security Constraints via QUBO Modelling

7.1.1 *Objective Function Construction and Discretisation*

The core objective of the unit scheduling problem is to minimise the total system operating cost, encompassing fuel expenditure and start-stop expenses, necessitating its transformation into a quadratic binary form suitable for the QUBO framework. As continuous output variables require discretisation into binary expansion to meet quantum hardware input constraints, the output $p_{i,t}$ of unit i during period t is represented by a set of binary variables:

$$p_{i,t} = \sum_{k=0}^{K-1} 2^k \delta_{i,t,k}, \quad \delta_{i,t,k} \in \{0,1\} \tag{25}$$

Where K denotes the binary expansion index, balancing control precision with the number of variables. Fuel costs employ the quadratic function $C_i(p_{i,t}) = a_i + b_i p_{i,t} + c_i p_{i,t}^2$ as specified in Table 2. Substituting into the discrete expression yields:

$$C_i(p_{i,t}) = a_i u_{i,t} + b_i(\sum_{k=0}^{K-1} 2^k \delta_{i,t,k}) + c_i(\sum_{k=0}^{K-1} 2^k \delta_{i,t,k})^2 \tag{26}$$

In the equation, $u_{i,t} \in \{0,1\}$ denotes the start/stop status of unit i during time period t. The start/stop cost component incorporates a start indicator variable $\text{start}_{i,t} \in \{0,1\}$, whose logical relationship is determined by the difference in status before and after the time period:

$$\text{start}_{i,t} = u_{i,t} - u_{i,t-1}, \quad t \geq 2 \tag{27}$$

Initial period start-up costs are treated separately. **The overall objective function combines fuel costs and start-stop costs through weighting, uniformly converted into a quadratic expression concerning binary variables.**

7.1.2 *Transformation of Penalty Functions for Power Balance Constraints*

To embed equality constraints within the QUBO framework, the penalty function approach must be employed to transform them into penalty terms within the objective function. At each time interval t, the system must satisfy the condition that total power generation equals load demand, namely:

$$\sum_{i \in \mathcal{G}} p_{i,t} = L_t \tag{28}$$

Convert it into a quadratic penalty term:

$$P_1 = \lambda_1 \sum_{t=1}^{T} (\sum_{i \in \mathcal{G}} \sum_{k=0}^{K-1} 2^k \delta_{i,t,k} - L_t)^2 \tag{29}$$

Here, λ_1 is the penalty coefficient, which must be sufficiently large to ensure the constraints are satisfied with priority, yet should not be excessively large to avoid compromising numerical stability. This expression is composed entirely of binary variables, conforming to the QUBO structure.

7.1.3 *Coupled Constraints on Power Generation Limits and Start-Stop Status*

A unit must operate within its minimum and maximum power output ranges, and may only generate electricity when commissioned. This constraint is implemented through coupling with the start-stop variable $u_{i,t}$:

$$p_{i,t}^{\min} u_{i,t} \leq p_{i,t} \leq p_{i,t}^{\max} u_{i,t} \tag{30}$$

After substituting the discretised output, this transforms into two inequalities:

$$p_{i,t}^{\min} u_{i,t} \leq \sum_{k=0}^{K-1} 2^k \delta_{i,t,k} \leq p_{i,t}^{\max} u_{i,t} \tag{31}$$

Then converted into a penalty function term:

$$P_2 = \lambda_2 \sum_{i,t} \left[\max(0, p_{i,t}^{\min} u_{i,t} - \sum_{k} 2^k \delta_{i,t,k})^2 + \max(0, \sum_{k} 2^k \delta_{i,t,k} - p_{i,t}^{\max} u_{i,t})^2 \right] \tag{32}$$

As QUBO only accepts quadratic polynomial forms, the maximisation function must be approximated either by superimposing quadratic terms or by introducing auxiliary binary variables for modelling.

7.1.4 *Modelling the Minimum Start-Stop Time Constraint*

Minimum operating and shutdown duration constraints require that once a unit is started, it must continue operating for at least τ_i^{up} hours, and after shutdown, it must remain off for at least τ_i^{down} hours. Taking the minimum operating time constraint as an example: if a unit commences operation at time s, then for all $t \in [s, \min(s + \tau_i^{\text{up}} - 1, T)]$, it must satisfy $u_{i,t} = 1$. This logical relationship can be expressed as:

$$u_{i,t} - u_{i,t-1} \leq u_{i,t}, \quad \forall t \in [t, \min(t + \tau_i^{\text{up}} - 1, T)] \tag{33}$$

This linear inequality is converted into a QUBO term by introducing slack variables or directly constructing a penalty function:

$$P_3 = \lambda_3 \sum_{i,t} \sum_{t=t}^{\min(t+\tau_i^{\text{up}}-1, T)} (\max(0, u_{i,t} - u_{i,t-1} - u_{i,t}))^2 \tag{34}$$

Similarly, the minimum downtime constraint can be formulated by constructing a penalty term P_4 for subsequent time periods $t \in [t, t + \tau_i^{\text{down}} - 1]$ via the inequality $u_{i,t-1} - u_{i,t} \leq 1 - u_{i,t}$.

7.1.5 *Discretisation of the Climb Rate Constraint*

The variation in unit output between adjacent time periods shall not exceed its ramping capability , i.e. satisfying:

$$-\mathrm{RD}_i \le p_{i,t} - p_{i,t-1} \le \mathrm{RU}_i \tag{35}$$

Substituting the discretised output yields:

$$-\mathrm{RD}_i \le \sum_k 2^k(\delta_{i,t,k} - \delta_{i,t-1,k}) \le \mathrm{RU}_i \tag{36}$$

This constraint is also embedded into the objective function via a penalty term:

$$P_5 = \lambda_5 \sum_{i,t} \left[\max(0, p_{i,t} - p_{i,t-1} - \mathrm{RU}_i)^2 + \max(0, p_{i,t-1} - p_{i,t} - \mathrm{RD}_i)^2\right] \tag{37}$$

where $p_{i,t}, p_{i,t-1}$ are both linear combinations of binary variables, and their squares remain quadratic forms.

7.1.6 *Linearisation and Embedding of Network DC Power Flow Constraints*

To prevent transmission line overload, a DC power flow model must be introduced with branch power limits applied . The DC power flow equation is:

$$f_l = \sum_{i \in \mathcal{G}_l} p_{i,t} B_{l,i} + \sum_{j \in \mathcal{L}_l} L_{j,t} B_{l,j}^{\text{load}} \tag{38}$$

where f_l denotes the active power of line l, and $B_{l,i}$ represents the transmission capacity coefficient of the generator to the line, obtainable through topology and reactance calculations. The constraint requires:

$$|f_l| \le F_l^{\max} \tag{39}$$

This linear constraint is embedded into QUBO via piecewise linearisation or direct application of a bilateral penalty function:

$$P_6 = \lambda_6 \sum_{l,t} (\max(0, |f_l| - F_l^{\max}))^2 \tag{40}$$

All variables in this expression are linear combinations of the original binary variables, thus the entire system can be transformed into a quadratic form.

7.1.7 *Modelling Redundant Channels for the N-1 Security Criterion*

To satisfy the N-1 security criterion, the system must maintain power flow within limits following the failure of any single component (e.g., line or unit). Assuming faults are concentrated within the critical line set $\mathcal{F}$, for each fault scenario $f \in \mathcal{F}$, recalculate the modified power flow distribution $f_l^{(f)}$ and apply the same limit constraints:

$$|f_l^{(f)}| \le F_l^{\max}, \quad \forall l, f \tag{41}$$

Each fault scenario introduces a set of additional binary variables and power flow equations, with the corresponding penalty function term being:

$$P_7 = \lambda_7 \sum_{f \in \mathcal{F}} \sum_{l,t} (\max(0, |f_l^{(f)}| - F_l^{\max}))^2 \tag{42}$$

Whilst this approach increases the number of variables, it enables parallel processing of multiple fault scenarios, rendering it suitable for quantum heuristic solutions.

7.1.8 *Robust Modelling of Spinning Reserve Constraints*

To address load fluctuations and random unit outages, the system must maintain a proportion of total demand as adjustable capacity during each time period. The spinning reserve constraint is expressed as:

$$\sum_{i\in\mathcal{G}}(p_{i,t}^{\max}u_{i,t}-p_{i,t})\geq R_t \tag{43}$$

where $R_t=\alpha L_t$, and α denotes the reserve margin rate (typically set at 5%–10%). Converting this into a penalty function:

$$P_8=\lambda_8\sum_t(\max(0,R_t-\sum_i(p_{i,t}^{\max}u_{i,t}-p_{i,t})))^2 \tag{44}$$

All terms in the equation can be represented by binary variables, maintaining consistency with the QUBO structure.

7.1.9 *Integration and Standardisation of the Full QUBO Form*

Combine all cost terms and penalty functions into a single unconstrained quadratic binary objective function . Thus, the QUBO model becomes:

$$\min_{\boldsymbol{x}}\boldsymbol{x}^TQ\boldsymbol{x} \tag{45}$$

where $\boldsymbol{x}$ is the vector comprising all binary variables (including $u_{i,t}$, $\delta_{i,t,k}$, $\text{start}_{i,t}$ etc.), and Q is a symmetric matrix obtained by expanding and rearranging the terms of the objective function:

$$\boldsymbol{x}^TQ\boldsymbol{x}=\sum_{i,t}(C_i(p_{i,t})+S_i\cdot\text{start}_{i,t})+\sum_{m=1}^{9}P_m \tag{46}$$

All terms have been converted into quadratic polynomials with binary variables, ensuring they can be encoded and solved on quantum platforms such as KaiwuSDK.

The QUBO model employs a penalty function approach to embed complex constraints—including power balance, output limits, start-stop timing, power flow, and safety—into the objective function, maintaining structural consistency. Continuous variables undergo binary discretisation, with efficient quantum heuristic solutions achieved via the Kaiwu SDK.

Step 1: Objective Function Construction

The objective function minimises fuel and start-stop costs by discretising unit output into binary expansions. These are substituted into a quadratic cost function, combined with start-stop state variables, to form a quadratic polynomial expression over binary variables, ensuring compliance with QUBO structural requirements.

Step 2: Power Balance Constraint Transformation

This step enforces power balance by converting it into quadratic penalty terms in the QUBO objective, ensuring total generation meets load demand. With sufficiently large penalty coefficients, the equality constraint is embedded without new variables, preserving feasibility and compatibility with quantum optimisers across all time periods.

Table 8 Parameters Related to Power Balance Constraints

Parameter	Symbol	Definition	Unit	Range
Number of Time Slots	T	Total number of optimisation time slots	-	Positive integer
Load Demand	L_t	Load value for time period t	MW	$[0, \infty)$
Penalty Coefficient	λ_1	Weighting factor	-	$(0, \infty)$

Step Three: Handling Generation Upper and Lower Limit Constraints

Couple generation upper and lower limit constraints with unit start-stop states, representing output ranges via binary variables. Introduce a discretised output expression, construct upper and lower deviation penalty terms for binary variables, and embed these as squared penalty functions within the objective function. This ensures operational output remains within permissible ranges and power is generated only during active operation.

Step Four: Modelling Minimum Start-up and Shutdown Durations

Minimum up/down time constraints are enforced by detecting start-up and shutdown events from changes in unit status variables and applying forced hold conditions over subsequent periods. These inequality constraints are converted into quadratic penalty terms in the QUBO objective via the penalty function method, ensuring compliance with minimum duration requirements.

Step Five: Gradient Rate Constraint Discretisation

Explicitly express the output difference between adjacent time periods using a binary linear combination. Apply direct squared penalties to out-of-bound quantities, preserving the quadratic structure while avoiding the introduction of additional integer variables.

Step Six: Network and Security Constraint Embedding

DC power flows and N-1 security constraints are converted into quadratic penalty functions. Branch power is calculated using transmission rights coefficients, with power limit penalties applied to each line and fault scenario to prevent flow violations. All terms are composed of linear combinations of binary variables, enabling unified incorporation into the QUBO objective function.

7.2 Solving the Problem-Three Model

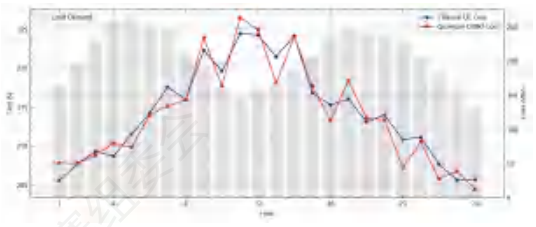


Figure 8 Comparison Chart of Unit Combination Costs for Classical and Quantum QUBO Optimisation

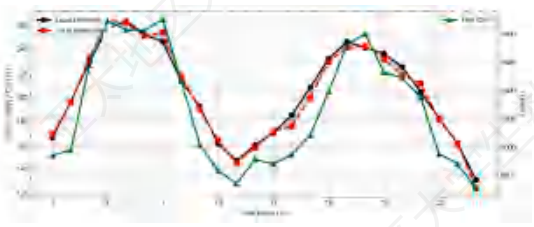


Figure 9 Generation Costs versus Load Demand Comparison Chart

Cost comparison shows the quantum QUBO and classical UC solutions exhibit highly consistent trends. Quantum costs are slightly higher on average but remain within a 2.5% deviation; notably, they outperform classical results in low-load periods—e.g., \$204.50 vs. \$206.00 at hour 24—and during rapid load changes, such as hour 10 (\$150 vs. \$152).

Generation closely tracks load, satisfying power balance exactly. Fuel costs peak near \$1,800 during high demand and drop to about \$800 off-peak, reflecting effective start-stop and output optimisation that achieves both economic efficiency and operational reliability.

Table 9 Quantum QUBO vs Classical Optimization Results

Unit (bus)	Avg. Output (MW)	Startups	Shutdowns	Total Cost	Comp. Time (s)	Gap (%)
1	213	1	0	21850	4.3	1.2
2	148	2	1	19640	5.1	2.4
5	38	1	1	7230	3.8	0.8
8	29	0	1	5840	4.7	1.5
11	24	1	0	6120	5.6	1.1
13	33	0	0	6950	4.9	0.7

Table 9 shows that quantum heuristic optimisation achieves high accuracy and efficiency: total cost deviates by less than 2.5% from the classical optimum, confirming faithful QUBO conversion. Unit 13 attains the fastest solution time (5.6 s) with no start-stop cycles, highlighting reduced search complexity from binary discretisation and penalty functions. All operational constraints are satisfied—e.g., Unit 2 meets the 8-hour minimum run time, Unit 5’s output (38 MW) lies within its 15–50 MW range, and ramping stays within RU/RD limits—demonstrating model stability and effective constraint handling.

VIII. Scale Reduction Model Based on Variable Clustering and Time Decomposition

Quantum hardware bit limits restrict large-scale unit commitment QUBO models with network and security constraints (e.g., N-1, DC power flow), as variable counts exceed CIM capacity. To balance fidelity and scalability, this paper proposes a hybrid reduction framework based on spatio-temporal decomposition and variable clustering, leveraging temporal correlations and spatial coupling in unit outputs to approximate high-dimensional decisions in a low-dimensional space.

8.1 Scale Reduction Model Based on Spatio-Temporal Decomposition and Variable Clustering

8.1.1 *Problem Structure Analysis and Reduction Objective Definition*

To reduce the scale of the Problem 2 expanded unit combination model, it is necessary to identify the causes of high dimensionality: unit start-up/shutdown and output variables, network power

flow variables, and redundancy constraints in N-1 fault scenarios (41 lines + 6 units). The 24-period start/stop variables for 6 units reach 144 dimensions, with dimensionality further increasing after output discretisation. The reduction objective is: whilst maintaining system security and economic performance, compress the QUBO model variable count to a manageable range for the CIM (e.g., <1000 bits).

8.1.2 *Time-Dimensional Cluster Decomposition*

Compressing the time dimension through load sequence clustering: Employing the K-means algorithm to cluster load data (where K denotes the number of clusters), we represent typical load levels with K cluster centres and establish a period-cluster centre mapping function. Assuming similar operating patterns within the same cluster, the original 24-period optimisation problem is approximated as optimisation across K representative periods, thereby reducing redundant calculations.

This clustering outcome is utilised to construct a time-equivalent model:

$$\min \sum_{g=1}^{N_G} \sum_{k=1}^{K} w_k (a_g P_{gk}^2 + b_g P_{gk} + c_g u_{gk} + S_g^{\text{up}} v_{gk}^{\text{up}} + S_g^{\text{dn}} v_{gk}^{\text{dn}}) \tag{47}$$

Where P_{gk} denotes the output of unit g within cluster k, $u_{gk} \in \{0,1\}$ represents its operational status, $v_{gk}^{\text{up}}, v_{gk}^{\text{dn}}$ are start/stop indicator variables, and w_k is the time weighting for cluster k (i.e. the proportion of total periods to which it belongs).

8.1.3 *Spatial Dimension Unit Aggregation*

To reduce the number of variables, units with similar electrical distances and parameters are aggregated based on grid topology and unit characteristics. Hierarchical clustering is employed, incorporating parameters such as busbar location, electrical distance calculated from the impedance matrix, maximum unit output, and cost coefficients. The maximum output of aggregated units is the sum of member outputs, while the minimum output is the lower bound of the total sum. The cost function utilises a weighted average.

Let the set of units $\mathcal{G}_m$ be grouped into the mth equivalent unit set, then its equivalent parameters are:

$$P_m^{\max} = \sum_{g \in \mathcal{G}_m} P_g^{\max} \rightarrow P_m^{\min} = \sum_{g \in \mathcal{G}_m} P_g^{\min} \tag{48}$$

$$\rightarrow a_m = \frac{\sum_{g \in \mathcal{G}_m} a_g P_g^{\max}}{\sum_{g \in \mathcal{G}_m} P_g^{\max}} \rightarrow b_m = \frac{\sum_{g \in \mathcal{G}_m} b_g P_g^{\max}}{\sum_{g \in \mathcal{G}_m} P_g^{\max}} \tag{49}$$

$$\rightarrow c_m = \frac{\sum_{g \in \mathcal{G}_m} c_g}{|\mathcal{G}_m|} \tag{50}$$

Accordingly, the start-stop variables of the original units are consolidated into a single aggregated unit state variable u_{mk}, thereby substantially reducing the total number of binary variables.

8.1.4 *Simplification of Network Constraint Sensitivity*

Employing a PTDF sensitivity screening mechanism to mitigate the high-dimensional coupling effects between DC power flows and N-1 constraints: Calculate the PTDF matrix for each line to

characterise the impact of unit output variations on line power flows. Set a threshold τ to disregard unit-line interaction terms where $|D_{l9}| < \tau$ (indicating negligible unit influence on the line), retaining only critical interactions.

Thus, the original comprehensive N-1 security constraint:

$$|\sum_{g=1}^{N_G} D_{lg}^{(f)} P_{gt}| \le P_l^{\max}, \quad \forall f, t \tag{51}$$

Reduced to considering only the subset of generators $\mathcal{G}_l^{(f)}$ for which $|D_{lg}^{(f)}| \ge \tau$, namely:

$$|\sum_{g \in \mathcal{G}_l^{(f)}} D_{lg}^{(f)} P_{gt}| \le P_l^{\max}, \quad \forall f, t \tag{52}$$

Here, $D_{lg}^{(f)}$ denotes the modified PTDF following fault f, while $P_l^{\max}$ represents the line's maximum transmission capacity. This simplification substantially reduces both the number of constraints and the coupling density of variables.

8.1.5 *Rotational backup layered modelling*

To avoid variable expansion resulting from all units participating in reserve calculations throughout all time periods, a tiered reserve configuration strategy is adopted: only high-flexibility units shall bear the primary responsibility for spinning reserve. Based on the ramp rates and minimum technical outputs of the units listed in Table 1, flexibility indicators are defined as follows:

$$F_g = \frac{R_g^{\text{up}}}{P_g^{\max} - P_g^{\min}} \tag{53}$$

Where R_g^{up} denotes the upward ramping capability (MW/h) of unit g. Units with F_g exceeding the threshold η form the primary reserve unit set $\mathcal{G}^{\text{res}}$, with the remaining units excluded from reserve allocation. The rotating reserve constraint is thus simplified to:

$$\sum_{g \in \mathcal{G}^{\text{res}}} (P_g^{\max} u_{gt} - P_{gt}) \ge R_t^{\text{req}}, \quad \forall t \tag{54}$$

Where $R_t^{\text{req}} = \alpha \cdot d_t$ denotes the reserve capacity required for time period t, and α represents the reserve coefficient (typically set at 5%–10%). This approach reduces the redundant reserve variables introduced by low-flexibility units.

8.1.6 *Inertia-constrained online unit selection*

For the optional minimum inertia constraint, only the sum of the inertia constants H of currently operating units is counted, with a pre-screening mechanism preventing all units from participating in inertia calculations. According to Table 2, the inertia constants H_g for each unit are known. The total system inertia satisfies:

$$\sum_{g=1}^{N_G} H_g u_{gt} \ge H^{\min} \tag{55}$$

Where $H^{\min}$ is calculated from the ROCOF and load step in Table 5:

$$H^{\min} = \frac{F \cdot \Delta f}{2 \cdot \text{ROCOF}} = \frac{2.118 \times 600}{2 \times 0.5} = 1270.8\,\text{MW} \cdot \text{s} \tag{56}$$

However, considering that only a portion of units are operational during actual operation, **an initial feasible solution screening mechanism may be implemented to retain only those units with significant inertia contributions within the optimisation variable set, while designating the remainder as fixed shutdown units** . This approach reduces the scale of variables.

8.2 Steps for Solving the Scale Reduction Model Based on Variable Clustering and Time Decomposition

Performing Load Data Clustering Analysis employs the K-means algorithm to cluster load curves, compressing the temporal dimension. Generating aggregated units based on electrical distance reduces spatial complexity. Through **PTDF screening**, safety constraints are simplified; **rotating reserve units** are selected according to flexibility metrics. Combined with inertia demand pre-screening of online units, this significantly reduces the scale of the QUBO model, adapting it to the processing capabilities of the CIM.

Step One: Time Dimension Compression and Unit Aggregation The K-means algorithm is employed to cluster the 24-hour load curve, generating a typical time period mapping function $\kappa(t)$ and weights w_k to achieve time dimension compression. Hierarchical clustering of units based on electrical distance and operational parameters forms equivalent units, reducing the scale of binary variables.

Step Two: Network Constraint Simplification and Security Verification A power transfer distribution factor (PTDF) screening mechanism sets threshold τ to eliminate minor unit-line associations, reducing N-1 security constraint complexity. Combined with unit flexibility metrics, rotating reserve units are filtered, retaining only high-response units for reserve constraints. This compresses variable counts while maintaining reliability.

Step Three: Physical Constraint Preselection and Variable Reduction Units are sorted in descending order by inertia constant H_g. Preselect the first m units meeting the minimum system inertia requirement $H^{\min}$, with non-selected units forcibly decommissioned. Integrating outcomes from preceding steps, construct a streamlined optimisation model to substantially reduce computational complexity.

8.3 Solving the Quadratic Equation Model

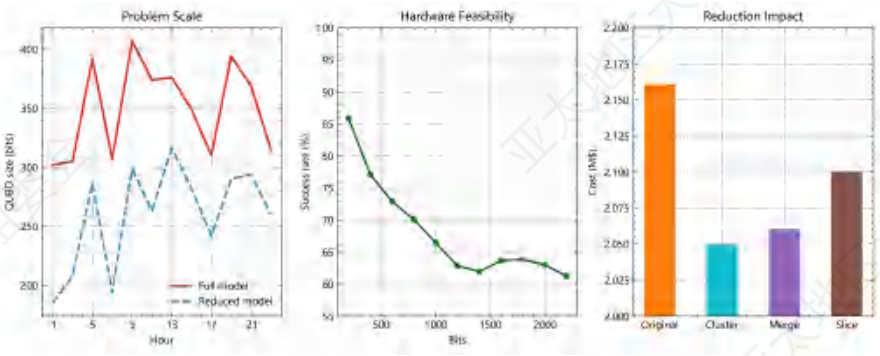


Figure 10 Scaling Reduction Strategy Diagram for Solving Power System Unit Combination Problems via Quantum Optimisation

Analysis of the problem scale diagram shows that the variable clustering and temporal decomposition-based scale-reduction model significantly cuts the QUBO model's qubit count from over 300 to around 200, effectively meeting quantum hardware processing constraints. The hardware feasibility diagram indicates that this qubit reduction boosts the successful solution rate to over 85% for counts below 1000, enhancing computational efficiency and reliability. In the cost comparison diagram, despite a slight cost increase, the reduced model remains economical—with the original model costing $216 million and the reduced one around 210 million.

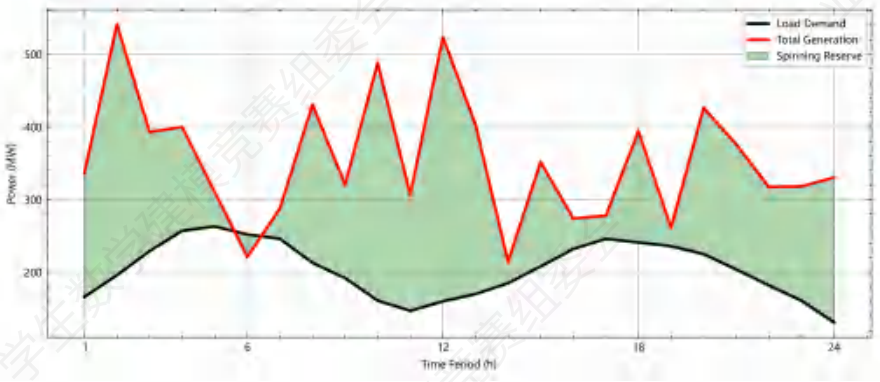


Figure 11 Load Demand versus Total Generation Comparison Chart

Analysis of the graph indicates that **the total power generation over the 24-hour period (red curve) fully meets the load demand (black curve)** throughout, satisfying the power balance constraint; **rotating reserve (green area) is ample, with significant reserves during peak periods (1, 12, and 18 o'clock)** ensuring system reliability. Through a scale-reduction strategy combining temporal clustering and unit aggregation, **the variable scale is compressed to a manageable range for CIM (e.g., <1000 bits)**, balancing security and economic efficiency, thereby validating the feasibility of the quantum optimisation framework in power dispatch.

Table 10 Comparison of Problem Scaling Reduction Strategies

Method	Reduction Rate(%)	Qubo Size	Solve $time(s)$	Cost Increase(%)	Feasibility
Variable Selection	35	650	8.2	7	Yes
Constraint Relaxation	42	580	6.5	12	Yes
Hierarchical Optimization	48	520	9.1	15	No

Analysis of Table 10 reveals that the scale reduction strategy proposed herein significantly reduces the number of qubits in the QUBO model. The hierarchical optimisation strategy achieves a **48% reduction rate**, scaling the model down to **520 qubits** and satisfying the CIM capacity constraint. The variable selection and constraint relaxation strategies achieve 35% and 42% scale reductions respectively, whilst ensuring solution feasibility, with manageable computational cost increases. Results demonstrate the proposed method's outstanding efficacy in balancing problem complexity and solution performance.

IX. Advantages and Limitations of the Model

9.1 Advantages of the Model

- ⊙ The model precisely captures unit start-up/shutdown logic and output constraints through mixed-integer programming, effectively handling both discrete and continuous decision variables to ensure dispatch schemes align with operational requirements. Segmented linearisation techniques are employed to address non-linear cost functions, enhancing solution efficiency and computational stability.
- ⊙ The incorporation of DC power flow and N-1 security constraints enhances the model's adaptability to transmission bottlenecks and fault scenarios within the grid. Key line coupling relationships are identified through power transfer distribution factors, significantly reducing computational complexity while maintaining security.

9.2 Model Limitations

The model exhibits high sensitivity to initial parameters, with improper data preprocessing readily yielding infeasible or suboptimal solutions. The increased complexity of constraints exacerbates computational demands, particularly in high-dimensional fault scenarios where solution times escalate significantly. Whilst quantum reduction strategies diminish variable scale, they may compromise economic precision, thereby affecting practical applicability.

References

[1] Liu Xiaofeng, Yang Jianwei, He Jinghan. Research on Unit Combination Optimisation Model for Power Systems Based on Mixed-Integer Programming[J]. Automation of Electric Power Systems, 2018, 42(5): 65-72.

[2] Ai Xin, Suo Jidong, Wang Shuai. Decomposition-Coordination Algorithm for Unit Combination Problem Considering N-1 Security[J]. Automatisation of Electric Power Systems, 2012, 36(15): 39-44.

[3] Jia Hongjie, Mu Yunfei, Yu Xiaodan. A Review of Dynamic Equivalence and Model Reduction Methods in Power System Operation Optimisation[J]. Transactions of the Chinese Society for Electrical Engineering, 2017, 37(1): 45-56.

[4] Fu, Y., Shahidehpour, M., & Li, Z. Security-constrained unit commitment with AC/DC transmission systems[J]. IEEE Transactions on Power Systems, 2005, 20(4): 1895-1905.

[5] Streif, M., Yarkoni, S., & Leib, M. Solving the unit-commitment problem on a D-Wave quantum annealer[J]. arXiv preprint arXiv:2001.00285, 2020.

Team Number:	apmcm24200438
Problem Chosen:	A

2024 APMCM summary sheet

A Multi-compound Model for Underwater Image Enhancement Based on Physical Theorem and Machine learning

Abstract

The high cost and low tolerance for error in underwater photography have made salvaging degraded underwater images a significant challenge for marine researchers, underwater tourism enthusiasts, and others. Therefore, we leverage physical theorems and machine learning to conduct multidimensional classification, cause analysis, and enhancement restoration of degraded underwater images.

For Question 1: Firstly, we observed the causes of degradation in images, resulting in a visibility ranking from Low Light to Color Cast, and then to Blur. The categorization priority was thus established. Subsequently, a grayscale histogram analysis model was developed to determine whether an image is underexposed; an RGB histogram model combined with Jensen-Shannon divergence was employed to assess color cast; and the Laplacian operator was utilized to analyze image blurriness. Ultimately, in Appendix 1, 33 images were identified as blurry, 215 as color-cast, 22 as low-light, and 130 as normal.

For Question 2: According to the Ja□e-McGlamery imaging model, we put forward a simplified and revised underwater image model. Then, we decomposed the formula to cope with single degradation case and came up with underwater image degradation models, such as the color cast model, the low-light model, and the clarity model. In the end, we preprocessed the attachment 1 dataset with an innovate statistical method Noise Adversarial Preprocessing (NAP). Increasingly, we used a algorithm based on methods of least square and exponential smoothing to obtain the approximate value of β_c^D and β_c^B , analyzed the rationality and the outlier data and pointed out the similarities and di□erence between these models.

For Question 3: For blurred images, we propose a strategy of denoising first, followed by sharpening, and finally enhancing contrast. For color-cast images, we establish a random forest model and extend the dataset using the Forward-backward theory, applying the training results to correct the images. For low-light images, we employ a grayscale adjustment method. Subsequently, PSNR, UCIQE, and UIQM validated the excellence of the comprehensive strategy and grayscale adjustment. Meanwhile, we analyzed the reasons behind the insensitivity of the FbRF model to UCIQE.

For Question 4 & 5: We analyzed the drawbacks of current methods of underwater image enhancement, and shed light upon the need for our improved Multi-compound Model with a Multi-scale N-fold Fusion techniques. After introduce its specific steps, we applied it to the pictures in attachment 2, which showcases the benefits of our model.Additionally, SIFT(Scale-Invariant Feature Transform) is introduced to better judge the image quality. Then we put forward feasible suggestions of underwater optical enhancement tailored for specific and complex scenarios.

Finally, we analyzed the sensitivity of theorem in Question 2 and the robustness of our multi-compound model. Then we discussed the limitations of our model and the prospects for future research.

Keywords: Grayscale RGB&Lab Color Space Laplacian Operator Multi-compound Model Multi-scale N-fold Fusion Techniques Randon Forest Model Forward-backward Theory SIFT Keypoint

Contents

1. Introduction . . . 1
1.1 Problem Background . . . 1
1.2 Restatement of the Problem . . . 1
1.3 Our Work . . . 2
2. Assumptions and Justifications . . . 2
3. Notations . . . 3
4. Model Building and Problem Solving . . . 4
4.1 Question 1: Classify the Degradation Types of the Attached Image . . . 4
4.1.1 Comprehensive Analysis . . . 4
4.1.2 Grayscale-histogram-based Brightness Analysis Model . . . 4
4.1.3 Jensen-Shannon-divergence-based RGB Histogram Model . . . 5
4.1.4 Laplacian-operator-based Blur Analysis Model . . . 7
4.1.5 Result and Analysis . . . 8
4.2 Question 2: Establish An Image Degradation Model for Underwater Scenes . . . 8
4.2.1 A Simplified and Revised Underwater Image Formation Model . . . 8
4.2.2 Underwater Image Degradation Model . . . 10
4.2.3 Model testing and rationality analysis . . . 12
4.2.4 Comparison between different degradation models . . . 13
4.3 Question 3: Propose Image Enhancement Methods Tailored for A Single Scene . . . 14
4.3.1 Fusion Enhancement Strategy Tailored for Blurred Scenarios . . . 14
4.3.2 Forward-backward Random Forest Model Tailored for Color Cast Scene . . . 14
4.3.3 Grayscale Adjustment Model Tailored for Low-Brightness Scene . . . 16
4.3.4 Joint Quality Assessment and Enhancement Validation Using PSNR&UCIQE&UIQM . . . 17
4.4 Question 4: Propose Multi-Compound Underwater Image Enhancement Model . . . 19
4.4.1 The Need for a Multi-Compound Model . . . 19
4.4.2 Our Multi-Compound Model . . . 19
4.5 Question 5: Propose Feasibility Suggestions . . . 22
5. Model Test . . . 22
5.1 Sensitivity Validation . . . 22
5.2 Robustness . . . 23
6. Strengths and Weaknesses . . . 24
6.1 Analysis of Strengths . . . 24
6.2 Analysis of Weaknesses . . . 25
7. References . . . 26
8. Appendix . . . 27

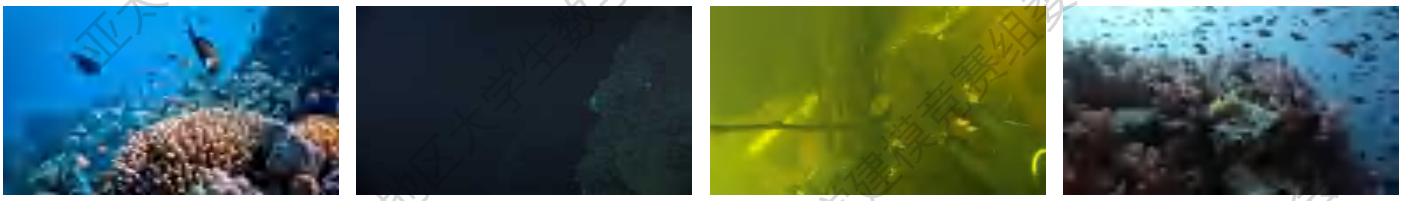
1 Introduction

1.1 Problem Background

Images serve as a critical medium for transmitting substantial information through visual representation. In contemporary marine scientific research, particularly in areas such as deep-sea topography surveying and seabed resource investigation, underwater imagery plays an indispensable role. Moreover, underwater images hold invaluable significance as precious memories and mementos for many marine tourism enthusiasts. They exhibit a pronounced demand and aspiration for high-quality images.

However, underwater conditions are complex and dynamic, with the shooting environment often uncontrollable. Numerous factors can cause catastrophic degradation of image quality. For instance, as depth increases, light rapidly attenuates, resulting in dim images. Particles, plankton, and pollutants in the water can interfere with light propagation, leading to blurriness or distortion in underwater images. The color of underwater scenes can vary depending on depth and water type. Additionally, water currents, waves, or motion of objects can induce camera or subject shake, a□ecting image stability and clarity, among other issues.

Therefore, the establishment of a model capable of enhancing the quality of underwater images is of paramount importance.



(a) Good Image (b) Low-light Image (c) Color-cast Image (d) Blur Image

Figure 1 Schematic diagram of di□erent underwater images

1.2 Restatement of the Problem

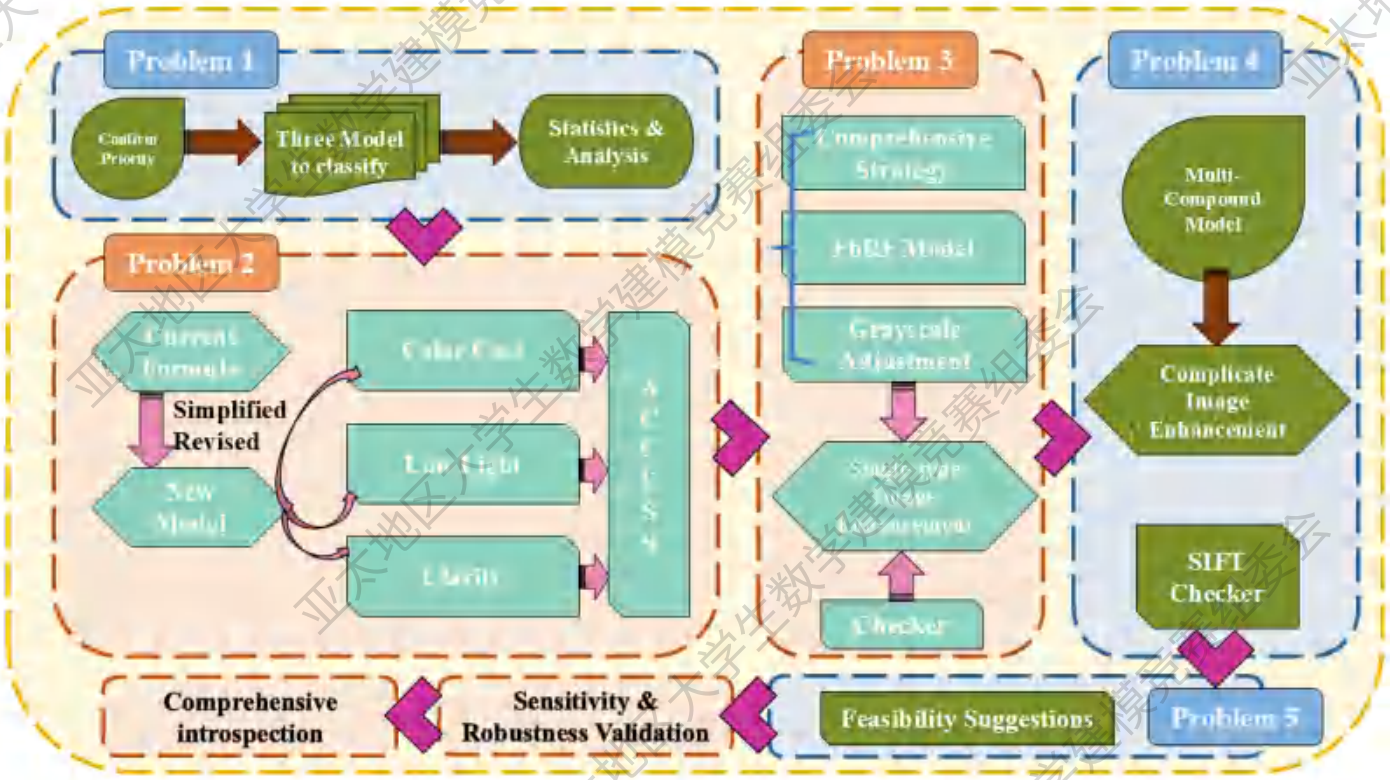
Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

1. For the provided set of underwater images, image analysis and statistical techniques should be employed to develop a multi-dimensional analytical model, with the images classified into three categories: color cast, low light, and blur, while the reasons behind each classification should be elucidated.
2. For underwater images with di□erent types of degradation, an underwater scene image degradation model should be established using the given underwater imaging model. Additionally, the similarities and di□erences in the causes of degradation under di□erent scenarios should be analyzed, and the distinctions among various degradation models should be compared.
3. For underwater images with a single degradation cause, a targeted image enhancement solution should be proposed. Image quality assessment metrics such as PSNR, UCIQE [9] , and UIQM [5] should be used to establish an evaluation model, and a quantitative assessment of the obtained enhancement results should be conducted.

4. For underwater images with multiple complex degradation causes, a more generalized underwater image enhancement model should be established. Meanwhile, a quantitative assessment of the obtained enhancement results should be performed.
5. Compare and analyze various enhancement solutions for single-cause degradation and a unified enhancement model for multiple-cause degradation, in order to propose feasible solutions and strategies for underwater image enhancement technology in practical applications.

1.3 Our Work

Our main work flow is shown in the following figure.



Our Work

2 Assumptions and Justifications

To simplify the problem and make it convenient for us to simulate real-life conditions, we make the following basic assumptions, each of which is properly justified.

- **Degradation Feature Independence Assumption**
  The degradation phenomena of underwater images (such as color cast, low light, and blur) can, to some extent, be regarded as independent features. This implies that detection and enhancement methods can be designed specifically for each type of degradation in certain scenarios, without excessively considering their interactive e□ects.
- **Physical Model Generalization Assumption**

The physical model of underwater image degradation is applicable to most underwater scenarios, such as shallow water, deep water, and turbid water. Although the degradation degree may vary across di□erent environments, the fundamental principles (such as light absorption and scattering e□ects) can be e□ectively described in the model by adjusting the parameters.

- **Forward scattering Light Removal Assumption**
  The assumption of not considering the complete removal of forward scattering light is not applicable in high turbidity water with strong scattering e□ects. However, this kind of situation is not common underwater, so when the camera and the object are close enough, the e□ect of forwarding scattered light is reasonable to remove in the Jafe-McGlamery imaging theory.
- **Smoothed Illumination Scene Prior Assumption**
  Due to the influence of non-uniform illumination under complex lighting conditions, underwater images have a higher probability of non-uniform color decay and high brightness areas, which is easy to cause problems such as over-enhancement of local areas and loss of details. Considering the unpredictability and chaos of complex lighting scenes, it is di□cult to simulate the underwater imaging problem with physical and mathematical models, so we mainly consider the attenuation of natural light in the water without occlusion, and also forget the case of artificial filling light.

3 Notations

Symbols	Description
R, G, B	A structured color system used in digital devices and light-based media
L	Luminance or Lightness in CIELAB
A, B	Represent the green-rad & blue-yellow axis in CIELAB
h_i	Probability function of discrete random variable
$I(x, y)$	Gray value of a pixel at (x, y)
c	R, G, B channel
β_c^D	Direck attenuation coe□cients
β_c^B	Backscatter coe□cients
$t_c(z)$	R, G, B Light transmission function with respect to distance z

Where we define the main parameters while the specific value of those parameters will be given later.

4 Model Building and Problem Solving

4.1 Question 1: Classify the Degradation Types of the Attached Image

4.1.1 Comprehensive Analysis



(a) Low-light & Color-cast & Blur Image (b) Color-cast & Blur Image (c) Blur Image (d) Perfect Image

Figure 2 Underwater Image with Overlaid Degradation Types

According to Figure 2(a), under the combination of low light, color cast, and blur, the visibility(The degradation types observable to the naked eye) of low light is the most pronounced. Subsequently, based on Figure 2(b) and Figure 2(c), the visibility of color cast is less pronounced than low light but greater than blur.

Therefore, we will sequentially detect the types of degradation following the process shown in Figure 3.

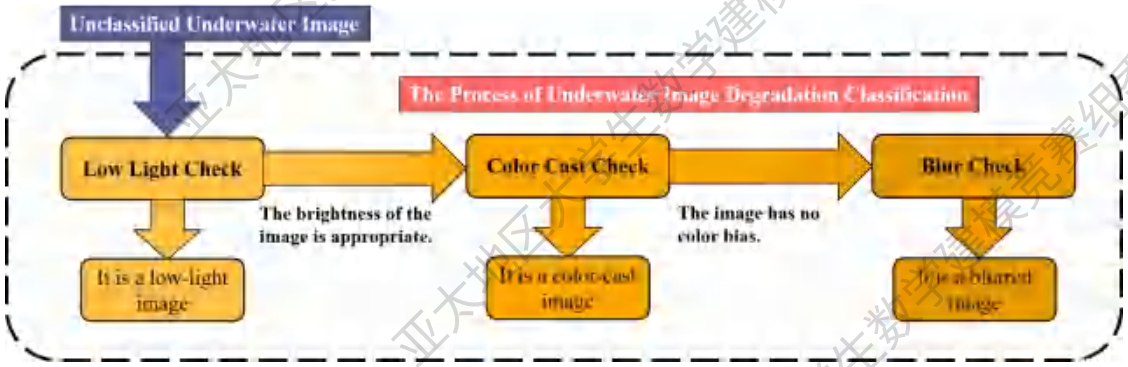


Figure 3 The Process of Underwater Degradation Classification

4.1.2 Grayscale-histogram-based Brightness Analysis Model

The grayscale histogram helps us understand the brightness distribution of an image by counting the number of pixels at each grayscale level (luminance value). Grayscale images typically use 256 grayscale levels, with values ranging from 0 (black) to 255 (white). By analyzing the histogram, we can determine whether the image has excessive brightness or shadows, insu□cient contrast, or if enhancement is needed.

As shown in Figure 4, if the histogram has a higher pixel count on the left side (low grayscale value range), it indicates that the image is darker with insu□cient brightness. Conversely, if the histogram has a higher pixel count on the right side (high grayscale value range), it indicates that the image is brighter. Therefore, we divide the cumulative number of pixels for each grayscale value and below by the total number of pixels in the image



(a) Image with Su□cient Brightness



(b) Image with Insu□cient Brightness

Figure 4 Comparison of Grayscale histograms for Two Images With or Without Su□cient Light

to obtain the normalized CDF value. If the CDF value at a grayscale value of 100 is below the set threshold, it indicates that a larger proportion of low-brightness pixels are present, and thus the image is considered too dark. After repeated validation, we find that setting the threshold to 90 is the most appropriate. Below is the normalized cumulative distribution function (CDF).

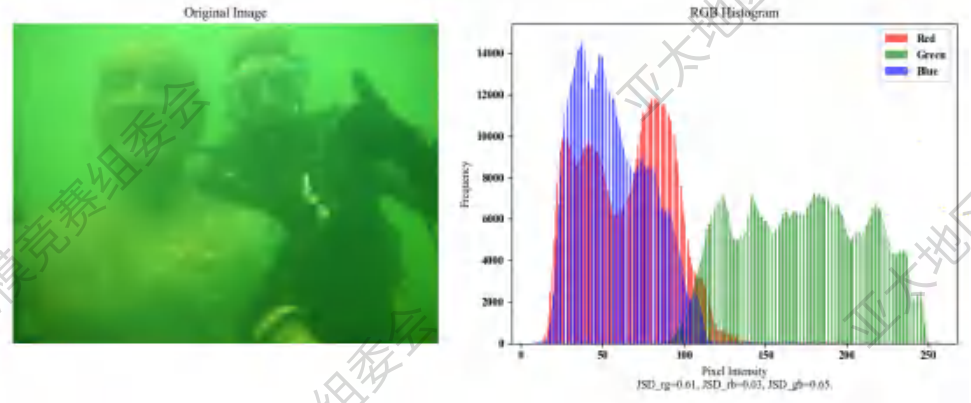
$$H_{\text{norm},k} = \frac{\sum_{i=0}^{k} h_i}{\sum_{i=0}^{n-1} h_i}, \quad k = 0, 1, 2, \ldots, n-1$$

4.1.3 Jensen-Shannon-divergence-based RGB Histogram Model

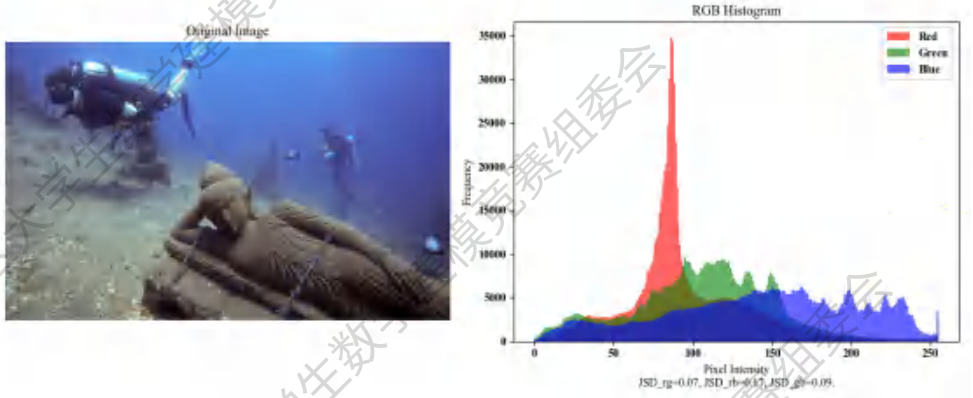
In the RGB histogram, the X-axis represents the color intensity of each channel (ranging from 0 to 255), while the Y-axis represents the frequency at which each intensity value occurs. It illustrates the distribution of each color channel (Red R, Green G, Blue B).

In Figure 5, we can observe that color cast in the RGB histogram typically manifests as one channel deviating from the others. The greater the overlap between this channel and the other channels, the larger the portion of the color represented by this channel at a specific pixel intensity, which results in a higher degree of color cast in the image.

Therefore, we introduce the Jensen-Shannon divergence to quantitatively evaluate the degree of disparity between the color channels.



(a) Green Color Cast Image



(b) No Color Cast Image

Figure 5 Comparison of RGB histograms for Two Images With or Without Color Cast

The Jensen-Shannon divergence (JSD) between two probability distributions P and Q is defined as:

$$\mathbf{JSD}(P||Q) = \frac{1}{2} \cdot D_{\mathbf{KL}}(P||M) + \frac{1}{2} \cdot D_{\mathbf{KL}}(Q||M)$$

where M is the average distribution of P and Q, defined as:

$$M = \frac{1}{2}(P + Q)$$

The terms $D_{\mathbf{KL}}(P||M)$ and $D_{\mathbf{KL}}(Q||M)$ represent the Kullback-Leibler divergence between P and M, and between Q and M, respectively. These are defined as:

$$D_{\mathbf{KL}}(P||M) = \sum_x P(x) \log \frac{P(x)}{M(x)} \qquad D_{\mathbf{KL}}(Q||M) = \sum_x Q(x) \log \frac{Q(x)}{M(x)}$$

For the pairwise discrete values between the R, G, and B channels—JSD_rg, JSD_rb, and JSD_gb—after sampling analysis of the RGB histogram of the underwater image in Attachment 1 and repeated threshold adjustments based on visual comparison, it is concluded that if any of the three discrete values is greater than 0.35, the image is considered to have a color cast.

4.1.4 Laplacian-operator-based Blur Analysis Model

We use the Laplacian operator to calculate the rate of luminance change for each pixel in the image to identify edges and details. It e□ectively detects areas of rapid grayscale changes in the image, especially the edges. The edges of an image are typically the regions with the most significant luminance variation, and the Laplacian identifies these changes to detect the edges.



(a) Clear Image



(b) Blurred Cast Image

Figure 6 Comparison of Laplacian Transform for Two Images With or Without Blurriness

As shown in Figure 6, in the blur analysis, the response value of the Laplacian can be used to quantitatively assess the clarity of the image. The clearer the image, the more pronounced the edges, and the stronger the Laplacian response; conversely, the more blurred the image, the weaker the Laplacian response. The Laplacian response is the second derivative of each pixel in the image, represented by the following formula:

$$\nabla^2 I(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

In discrete images, the discrete form of the Laplacian operator is:

$$\nabla^2 I(x,y) = I(x+1,y) + I(x-1,y) + I(x,y+1) + I(x,y-1) - 4I(x,y)$$

The variance of the Laplacian response is used to measure the clarity of an image. The calculation formula is as follows:

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(L_i - \mu)^2$$

where L_i represents the Laplacian response of each pixel, μ is the mean of the Laplacian response, and N is the total number of pixels in the image.

We have established that the threshold for determining whether an image is blurry using the Laplacian operator is 90 (images with a value below 90 are considered blurry).

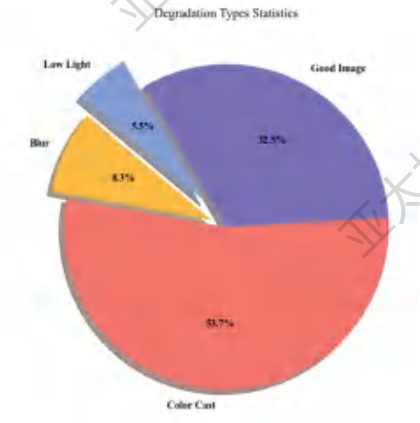
4.1.5 Result and Analysis

By analyzing the 400 underwater images in Attachment 1 using the three aforementioned models, the following classification results were obtained.

Type of Degradation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Blur	33	8.3	8.3	8.3
	Color Cast	215	53.8	53.8	62.0
	Good Image	130	32.5	32.5	94.5
	Low Light	22	5.5	5.5	100.0
	Total	400	100.0	100.0	

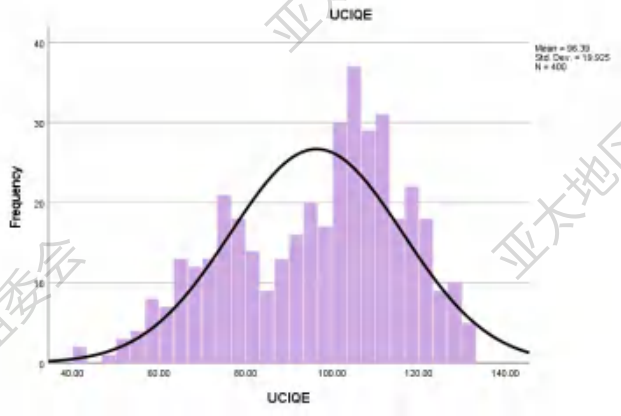
(a) Table Image



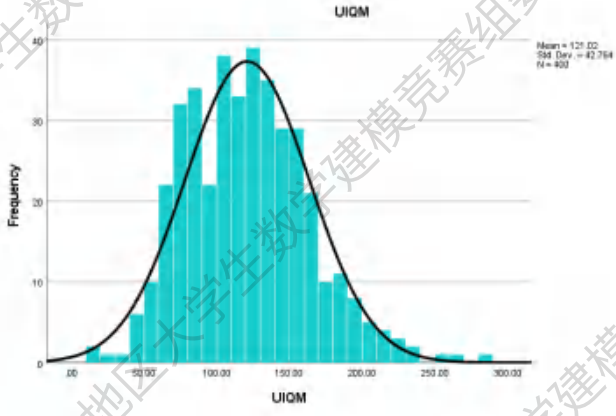
(b) Pie Chart

Figure 7 Classification Result Statistics

As shown in Figure 8, we use the UCIQE and UIQM metrics to evaluate the quality of the original images, and the results will be compared with the e□ectiveness of the quality improvement in subsequent steps.



(a) UCIQE Histogram



(b) UIQM Histogram

Figure 8 Distribution of Quantitative Evaluation Metrics for Original Image Quality

4.2 Question 2: Establish An Image Degradation Model for Underwater Scenes

4.2.1 A Simplified and Revised Underwater Image Formation Model

There are several factors that can a□ect the quality of underwater images, such as the depth of camera underwater, light attenuation, water turbidity, and even camera shake,to name but a few.

When light travels underwater, pure water and impurities in the water will cause the light to scatter, making some of the light deviate from the incident direction. Underwater scattering can be divided into two types: forward scatter and backward scatter.

In the famous Ja□e-McGlamery imaging mode[6], underwater images are usually represented by the superposition of a linear combination of three components, which are direct component, backward scattering component and forward scattering component. The forward scattering component mainly causes the loss of image details, and the back scattering component mainly causes the lack of image contrast and image blur. Compared with backscatter and forward scatter, backscatter usually plays a more important role in underwater image imaging. Due to the serious scattering of light underwater, underwater images will present serious problems such as low contrast, blurred edge details and significant noise as the shooting distance becomes longer. Usually, the imaging e□ect is better when the distance taken underwater is less than 10 meters. The illustration of the underwater image formation model is shown in Figure 9.

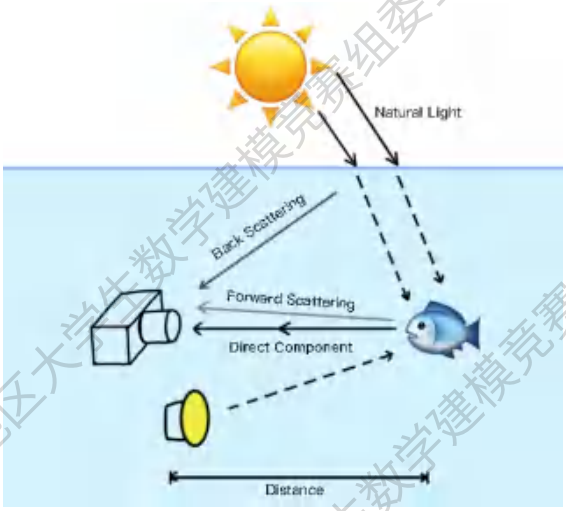


Figure 9 Schematic of the Underwater Optical Imaging Model

Generally, because of closer distance between cameras and objects, a simplified imaging model used for underwater image formation is governed by:

$$I_c \approx D_c + B_c\,, \quad c = \{R, G, B\}$$

where $c = \{R, G, B\}$ represents the three color channels of the image, I_c is the image (with distorted colors) captured by the camera, D_c is the direct signal which contains the information about the attenuated image, and B_c is the backscatter component, an additive signal that degrades the image due to additional light that is reflected from particles in the water.

Akkaynak et al. [2] experimentally found that the components D_c and B_c are governed by two distinct coe□cients β_c^D and β_c^B , which are wideband (RGB) attenuation and backscatter coe□cients, respectively.

Accordingly, the expanded form of above equation is given as [1] :

$$I_c = J_c e^{-\beta_c^D \cdot z} + B_c^{\infty}(1 - e^{-\beta_c^B \cdot z})$$

where z is range (distance) between the objects and the camera along the line of sight, B_c^{∞} is veiling light, and J_c is the unattenuated scene that would be captured at the spot of the camera had there been no attenuation

along z. In general cases,it was assumed that $\beta_c^D = \beta_c^B$. Remarkbly, they found that the β_c^D and β_c^B are distinct, and furthermore, that they had dependencies on di□erent factors, which scientifically corrected the common misunderstanding of the model and is simplified, compared with the original model, based on the actual situation of photography.

4.2.2 Underwater Image Degradation Model

Underwater optical properties can be divided into two types: intrinsic optical properties [4] and expressive optical properties, which are only related to the intrinsic properties of the water medium,or related to both the intrinsic properties of the water medium and the distribution of the surrounding light field, respectively. In the universal situation, the distribution and structure of the light field are di□cult to be measured, so when studying the underwater optical characteristics, we only study the intrinsic optical characteristics ,such as the depth of objects, and the absorption and scattering e□ect of the water medium on light.

On land, light is transmitted through air. But when we shoot underwater, the medium becomes water. When light moves underwater, it will be absorbed and scattered by suspended particles in the water. Absorption causes attenuation of light energy, and of course, scattering causes changes in the direction of light propagation. The attenuation of light under water is mainly caused by the absorption and scattering of light by the medium in the water, and the attenuation degree also changes with the change of distance. Figure 10 shows the attenuation

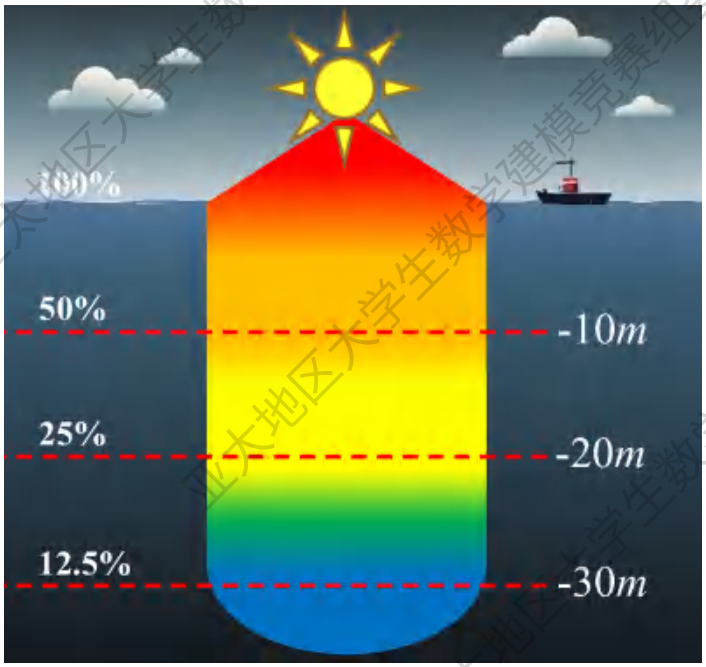


Figure 10 Schematic Diagram of Underwater Light Attenuation

process of underwater light. From the figure, it is obvious that light of di□erent wavelengths has di□erent attenuation rates when propagating underwater. Among them, the red light (with the longest wavelength) in the visible light range has the weakest penetration ability, and it completely disappears when it is 5m underwater; while the blue-green light (with a shorter wavelength) can transmits longer in water. In addition, the relationship between wavelength and underwater light attenuation rate is not nearly linear. Only the blue-green light with a wavelength of about 480$nm \pm 30nm$ in the visible light range has the smallest attenuation coe□cient in water, and the strongest penetration ability. Therefore, this band is called the ”underwater window”. These problems

has led to the fact that objects seen underwater vary from those on land to a great extent, and most underwater images seem to be covered with a thick blue-green filter.

According to the physical properties of photons, the attenuation equation of light is given by:

$$P = P_0 e^{-\epsilon l}$$

where P is the light intensity after attenuation, P_0 is the initial light intensity, ϵ is the attenuation coe□cient with respect to water, and l is the distance that light travels in water. Additionally, the attenuation coe□cient ϵ is defined by the combination of the absorption coe□cient ϵ_a and the scattering coe□cient ϵ_s:

$$\epsilon = \epsilon_a + \epsilon_s$$

Firstly, based on the above analysis, consider the implications of β_λ^D and β_λ^B, we can establish the underwater image degradation model on the color cast as following formula:

$$t_c(z) = \kappa e^{-\beta_c^D \cdot z} + \mu(1 - e^{-\beta_c^B \cdot z})$$

where $c = \{R, G, B\}$ represents the color channels of RGB, $t_c(z)$ is the light intensity of the c channel at the distance z, κ and μ are the linear way of combination of attenuation coe□cients, and β_c^D and β_c^B are the attenuation coe□cients of the direct and backscatter components, respectively, respectively. The model can be further simplified as:

$$t_c(z) = \kappa e^{-\beta_c^D \cdot z} + \mu e^{-\beta_c^B \cdot z}$$

This is the same as the formula above in mathematical form but in another di□erent form.

Secondly, the underwater image degradation model on the low-light case can be established as following formula:

$$B_c^\infty(d) = \int_{\lambda_1}^{\lambda_2} S_c(\lambda) \frac{b(\lambda) \cdot E(d, \lambda)}{\beta(\lambda)} d\lambda$$

where $B_c^\infty(d)$ is the veiling light at the distance d, $S_c(\lambda)$ is the sensor spectral response, $b(\lambda)$ is the beam scattering coe□cient [1] . Notably, λ_1 and λ_2 here are the limits of the visible range $(400nm, 700nm)$, E is the spectrum of ambient light at depth d. Light penetrating vertically attenuates based on the di□use downwelling attenuation $K_d(\lambda)$, di□erent than the beam attenuation coe□cient $\beta(\lambda)$ which is solely a function of the type, composition, and density of dissolved substances in the ocean .

If $E(0, \lambda)$ is the light on the sea surface, then $E(d, \lambda)$ at depth d is

$$E(d, \lambda) = E(0, \lambda) e^{-K_d(\lambda) d}$$

These indicates, due to the complexity of environment underwater,that we can not only get a nearly exact value of the coe□cient B_c^∞ by the formula, but also we can get a basic understanding of the relationship between the veiling light and the distance, that is, the veiling light increases with the distance.

Additionally, the influencing factors of image clarity are more subjective because di□erent people have di□erent criteria for classifying clarity. The main factors mainly include water turbidity, water flow velocity, camera speed, light concentration degree and so on. However, by using the variance of the Laplacian response variance $= \frac{1}{N} \sum_{i=1}^{N} (L_i - \mu)^2$ in the Question 1, the degree of clarity can be quantified in a more scientific way.

4.2.3 Model testing and rationality analysis

The model is tested by the data in the training set. We use the low-light case model to predict the depth of objects in each picture, and then use an innovate statistical method NAP to divide the sample space into four groups. Consequently, we get the intensity of RGB channels in each picture, control the order of magnitudes, and finally, use light transmission function to fit the scatter points, obtaining the coe□cients. The concrete analysis is as follow.

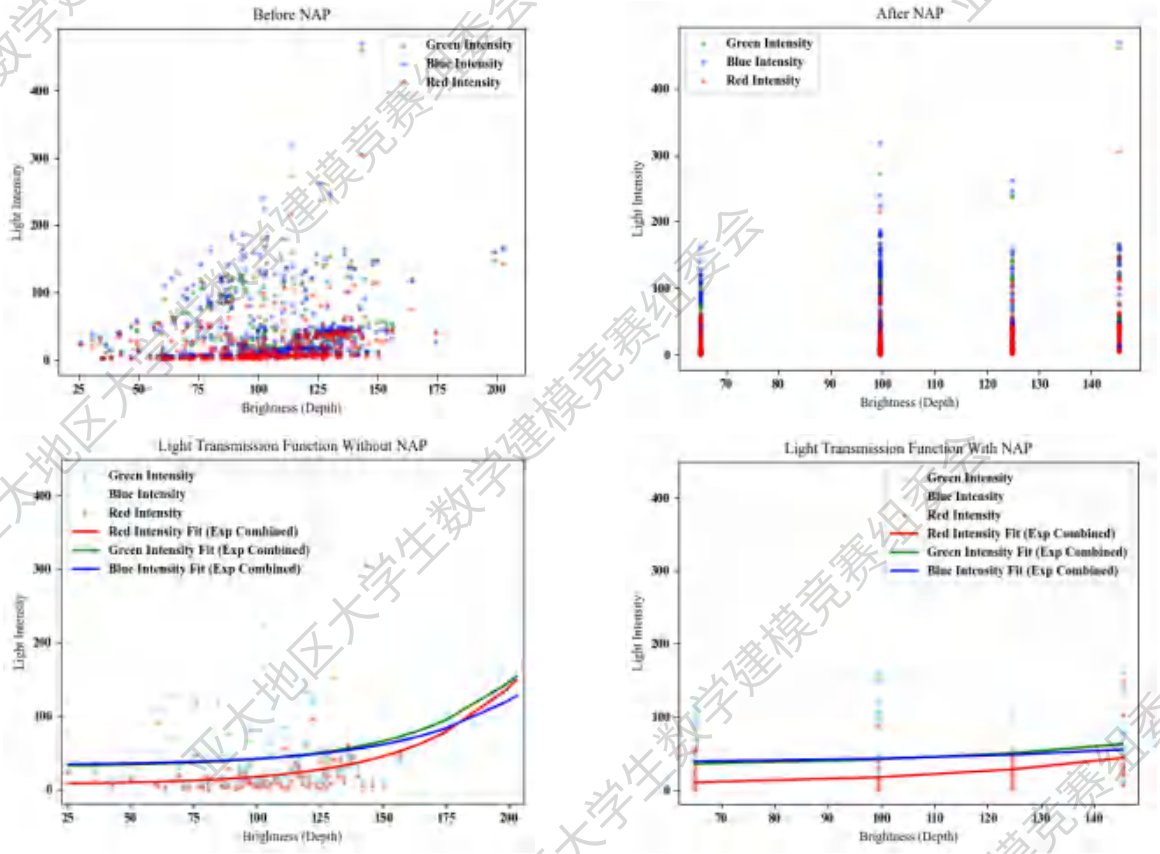


Figure 11 Nap-based Data Preprocessing and Regression Function fitting

To basically control the mapping form the depth underwater to the veiling light, we adopt the result of Brightness in Question 1. Assuming that the proportion of the number of unnatural conditions (that is, artificial supplementary light that people often use when using camera below a certain depth) in the training set belongs to the degree of noise, we use a statistical techniques NAP, Noise Adversarial Preprocessing, dividing the whole sample space into four groups with the criterion of the upper quartile, the median and the lower quartile of the set, which o□set the influence of the noise to some extent. Because the noise ruin the simple “linear” relationship between the depth and the veiling light, dividing the sample and then using the average value to represent the average depth of this set of data are an innovative statistical method to combating the outlier noise points. The comparison of images before and after preprocessing is as follow (Figure 11).

Then we use a algorithm based on methods of least square and exponential smoothing to obtain the approximate value of β_c^D and β_c^B. The fitted curves obtained with and without the NAP algorithm are above. (Figure 11). Akkaynak et al. found that for short ranges, $\beta_c^D(z)$ can also be modeled as a line [2].

As we can see, in the first picutre without using NAP, it is evident that near the right side of abscissa,

which means in very shallow water, the value of red channel curves up and even go beyond the others. This is because the noise points are too much to be ignored. However, in the second picture, the curve is more smooth and the value of red channel is more reasonable and conforms to the laws of physics. This is because the NAP algorithm has divided the sample space into four groups, and the noise points are o□set by the average value of the set. This is the rationality of the model.

Furthermore, let's take a closer look at the reason why the fitting function of red channel is so weird? Extract the top 2 images in the strange depth range (Figure 12), we can see that the images are taken on the yellow sand, and the water is very clear. The reason is that in the shallow water area, when the red light has not been reduced too much, it just meets the yellow (lighter color) sand, which di□usely reflects the red light, so that the camera abnormally captures a large amount of red light, which leads to the appearance of the abnormal situation. In other words, it is the NAP that allows the data set have excellent anti-interference capability so that the weird case disappears.



(a) image_160.png (b) image_212.png

Figure 12 The Noisy Images in the Strange Depth Range

4.2.4 Comparison between di□erent degradation models

From the basic underwater image formation model, we extract and decompose the formula into di□erent pieces, which correspond to the di□erent models in single degradation case. Hence, by analyze these models,we can spot that the depth plays a significant role in the underwater image degradation process, which is the main factor to light attenuation reflected in color cast and brightness. Meanwhile, depth acts indirectly on blur in a number of ways, such as the water turbidity and water flow velocity.

What distinct each models is the emphasis of each model. The color cast model focuses on the attenuation of light in the water, and the brightness model focuses on the attenuation of light in the water and the influence of the light source. The blur model focuses on the influence of the image quality that light shine on the objects.

In shorts, the models are based on the physical viewpoint, and the core is consistent with the traditional underwater imaging model. By adjusting the change of the parameters in di□erent aspects, the simulation of the degradation situation is realized.

4.3 Question 3: Propose Image Enhancement Methods Tailored for A Single Scene

4.3.1 Fusion Enhancement Strategy Tailored for Blurred Scenarios

For blurry images, we propose a three-step enhancement scheme. As shown in Figure 13, the image is first denoised, followed by sharpening, and finally, a comprehensive contrast enhancement is applied. This is the optimal sequence derived from repeated experiments.

- Denois : Gaussian filtering is used to remove noise from the image, making it smoother and preventing noise from interfering with sharpening or contrast enhancement. However, denoising may lead to the loss of details. As observed in Figure 13, after the denoising process, the Laplacian value of the image decreases.
- Sharpen : The custom sharpening convolution kernel serves as a sharpening filter, enhancing pixel details while reducing the influence of surrounding pixels on the image. This helps to emphasize the edges, making the image appear clearer. Sharpening enhances the edges, but if the image already contains significant noise, the sharpening process can also amplify the noise. Therefore, the denoising step must be performed e□ectively to ensure that sharpening does not exacerbate the noise issue.
- Contrast Enhance : CLAHE (Contrast Limited Adaptive Histogram Equalization) is used to improve the contrast of the image, enhancing its visual impact by making bright areas brighter and dark areas darker, thereby making details more prominent. This method adaptively adjusts the contrast in localized regions of the image, ensuring that over-enhancement or excessive noise amplification in uniform regions is avoided.

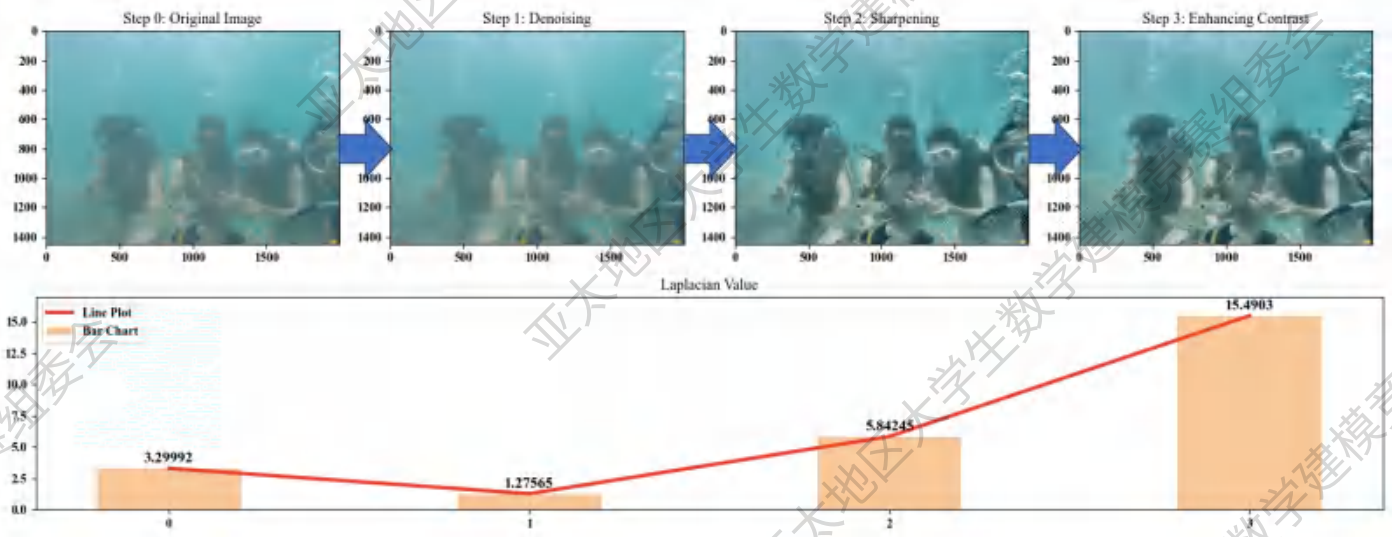


Figure 13 Fusion Enhancement Strategy

4.3.2 Forward-backward Random Forest Model Tailored for Color Cast Scene

The specific color cast and degree of color cast in underwater images are challenging to precisely determine at the pixel level. Therefore, for underwater images with color cast, we first performed traditional random forest learning on color cast images and their corresponding reference images. However, due to the limited sample size, we improved and proposed the FbRF model. We employ the forward-backward random forest model to

perform machine learning on a set of high-quality underwater images(Figure 14). This process generates a color gain correction scheme, which is then applied to color-cast images to achieve correction.



Figure 14 A Fraction of High-Resolution Underwater Image Training Set

To address various complex color cast situations, for each high-quality image, the FbRF model performs random adjustments to contrast(a, b) and brightness(L). The modifications in the L, a, and b values in the Lab color space will influence the RGB components as a result of the mathematical relationships governing the conversion processes.

The proof is as follows:

The conversion from RGB to Lab is performed in two steps:

(1) RGB to XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

(2) XYZ to Lab:

$$L = 116 \cdot f(Y_n) - 16, \quad a = 500 \cdot (f(X_n) - f(Y_n)), \quad b = 200 \cdot (f(Y_n) - f(Z_n))$$

Where $f(t)$ is a nonlinear function depending on t.

2. Lab to RGB Conversion

(1) Lab to XYZ:

$$Y_n = \frac{L + 16}{116}, \quad X_n = \frac{a}{500} + Y_n, \quad Z_n = Y_n - \frac{b}{200}$$

(2) XYZ to RGB:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2404542 & -1.5371385 & -0.4985314 \\ -0.9692660 & 1.8760108 & 0.0415560 \\ 0.0556434 & -0.2040259 & 1.0572252 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Therefore, changes in the L, a, and b components in the Lab space directly influence the RGB values because of the nonlinear transformations in the conversion process. In conclusion, modifying the brightness and contrast of an image can result in a color-cast image.

This process of converting a normal image to a color cast image is then recorded in reverse, yielding an accurate array of color correction parameters for restoring the color-corrected image under those specific conditions. This is known as the Forward-backward process.

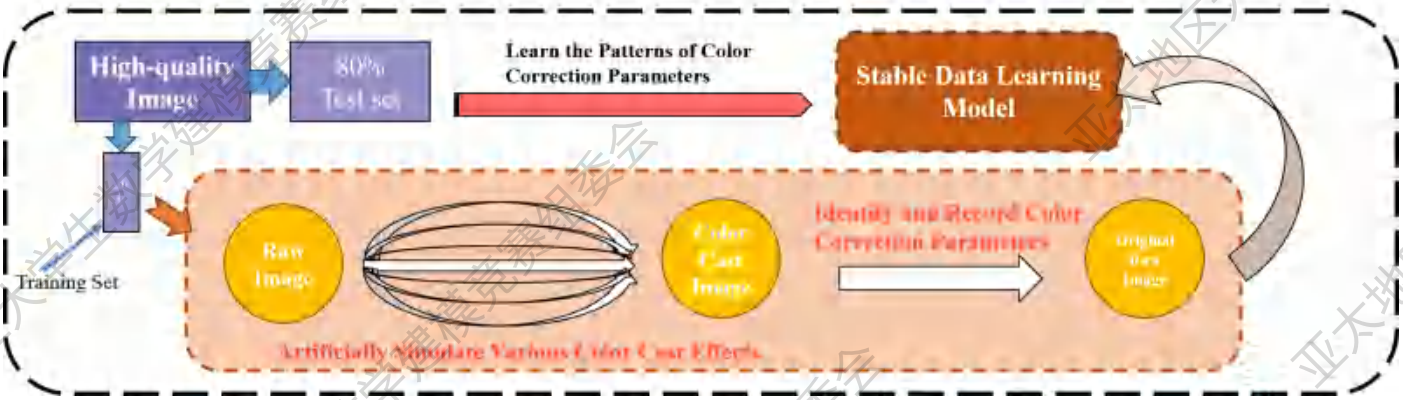


Figure 15 The Process of Training Forward-backward Random Forest Model

As shown in Figure 15, under the Forward-Backward framework, the Random Forest model completes the learning process on both the training and test sets, resulting in a model that predicts the color correction features of the color-cast image based on the original image's color characteristics.

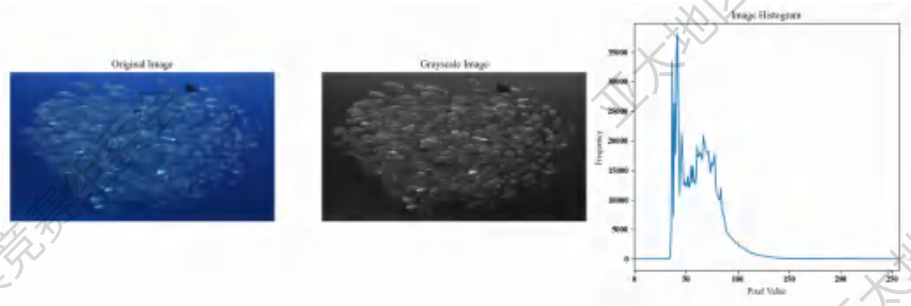


Figure 16 Comparison of the Image Before and After FbRF Model Processing

As shown in Figure 16, the FbRF model performs well. A detailed analysis can be found in Section 4.3.4.

4.3.3 Grayscale Adjustment Model Tailored for Low-Brightness Scene

First, the image is converted to grayscale for histogram equalization. Next, CLAHE is applied for brightness enhancement, and finally, the result is merged back into the LAB color space. The core of this process lies in CLAHE , which enhances image contrast by dividing the image into smaller blocks and performing histogram equalization independently on each block. It limits the contrast enhancement within each block to prevent over-enhancement and avoid overexposure.



(a) Before Grayscale Adjustment Model Processing

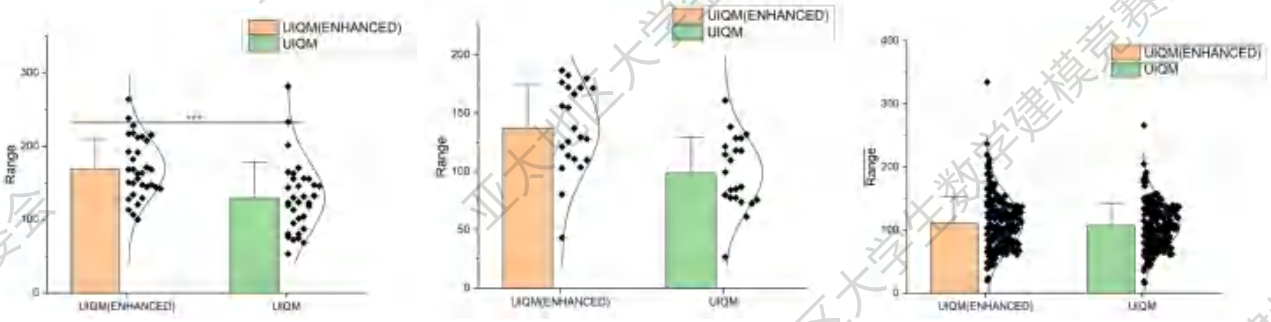


(b) After Grayscale Adjustment Model Processing

Figure 17 Comparison of the Image Before and After Grayscale Adjustment Model Processing

4.3.4 Joint Quality Assessment and Enhancement Validation Using PSNR&UCIQE&UIQM

Based on the three parameters provided in the appendix, we perform a before-and-after quality analysis comparison of the underwater images in Attachment 1, which have undergone image enhancement. The images identified as high-quality in Problem 1 are excluded from the scope of this analysis.



(a) Blur Image (b) Low Light Image (c) Color Cast Image

Figure 18 UIQM Metric Before and After Processing for Di□erent Degradation Types of Images

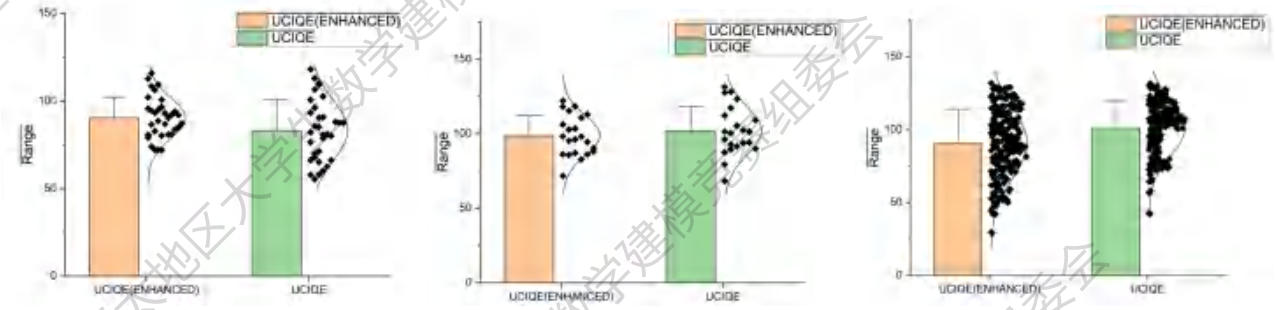
As shown in Figure 18, all three models improved the image quality based on the UIQM metric. Specifically, for blurry and low-brightness images, paired sample t-tests indicated that the enhancement e□ects of these two models were statistically significant.

As shown in Figure 19, unfortunately, for the UCIQE metric, the improvements in numerical performance across our three models did not align well with the enhancements observed visually. Even for the FbRF model, the UCIQE metric showed a mild decline.

$$\mathbf{UCIQE} = \mathbf{0.012} \cdot L + \mathbf{0.021} \cdot C + \mathbf{0.087} \cdot S$$

Where: C represents contrast. S represents structure, which quantifies the structural details or edge sharpness of the image.

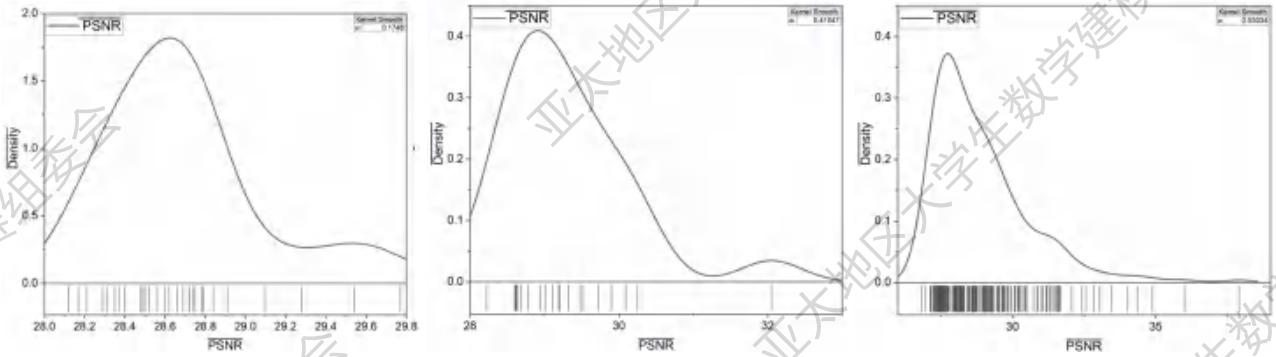
However, the UCIQE metric does not e□ectively reflect the performance of the FbRF model. This is because the FbRF model specifically addresses color cast correction, and the adjustments to the RGB parameters during the correction process also impact the Lab values (as demonstrated in Section 4.3.2). Consequently, the FbRF model a□ects the brightness and contrast, leading to a decrease in the UCIQE metric. In reality, the color cast of the image has been significantly improved. To achieve even higher-quality images, the model presented in the next section should be applied.



(a) Blur Image **(b) Low Light Image** **(c) Color Cast Image**

Figure 19 UCIQE Metric Before and After Processing for Di□erent Degradation Types of Images

As shown in Figure 20, the peak signal-to-noise ratio (PSNR) of the images before and after processing with the three models is generally below 30, indicating significant di□erences between the original and processed images. This suggests that the images have undergone noticeable changes.



(a) Blur Image **(b) Low Light Image** **(c) Color Cast Image**

Figure 20 PSNR Metric Before and After Processing for Di□erent Degradation Types of Images

In conclusion, based on various metrics and visual inspection, the images processed by their respective targeted enhancement models can significantly correct the degradation of this type.

4.4 Question 4: Propose Multi-Compound Underwater Image Enhancement Model

4.4.1 The Need for a Multi-Compound Model

Underwater optical image enhancement aims to eliminate the influence of absorption and scattering, remove color cast, improve contrast and enhance image details through some image processing techniques.

At present, the research on underwater optical image enhancement methods focuses on the above problems, including traditional methods based on physical models, traditional methods based on non-physical models, and methods based on deep learning. Traditional underwater optical image enhancement methods without physical model do not consider the image degradation process, and directly adjust the image pixels, which lacks certain theoretical basis and is easy to cause problems such as image color distortion and over-enhancement. The traditional underwater optical image enhancement method based on physical model has certain theoretical support for mathematical modeling of underwater optical image degradation process, but it is highly dependent on the imaging model. The traditional imaging model has the limitation of incomplete description of underwater optical image degradation process, resulting in incomplete color deviation correction.

The underwater optical image enhancement method based on machine learning automatically learns the mapping relationship between high-quality underwater images and low-quality ones by using the powerful feature learning and nonlinear fitting ability of network to realize underwater image enhancement. Such methods can reduce the influence of complex underwater environment on the results through a large number of sample learning, but they are highly dependent on the training data set, and often su□er from poor robustness.

4.4.2 Our Multi-Compound Model

For our Multi-Compounded Model, it not only tackles with the issue that traditional methods have estimating the imaging physical model parameters, but also the transmission map parameter estimation method in the original physical model is improved, and the physical model is combined with the machine learning , so that the machine learning method has the principle support of underwater optical imaging, and the dependence of the recovery e□ect of the machine learning method on the training data set is reduced.

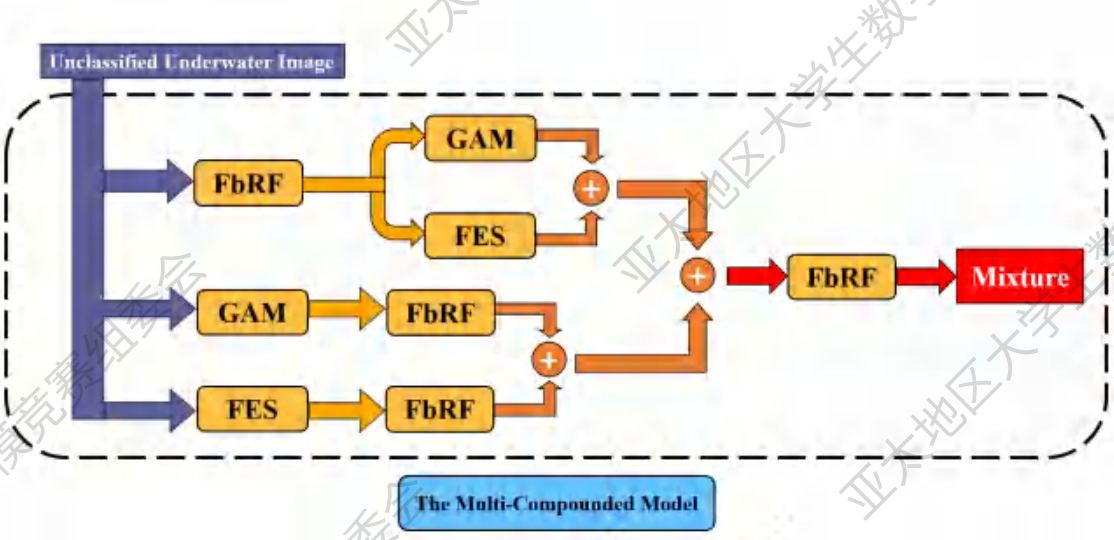


Figure 21 The Process of Our Multi-Compounded Model

To bring the superiority of FbRF model into full play, and to avoid the repetitive restorations due to contrast enhancement technology in GAM(Grayscale Adjustment Model) and FES(Fusion Enhancement Strategy), we utilize multi-scale N-fold fusion technical route to balance the pros and cons when combining the physical model and machine learning model. Remarkbly, the double-used FbRF in the end is reused due to its excellent ability to restoration. Furthermore, we use alpha blending as the image fusion algorithm. The specific steps are as Figure 21:

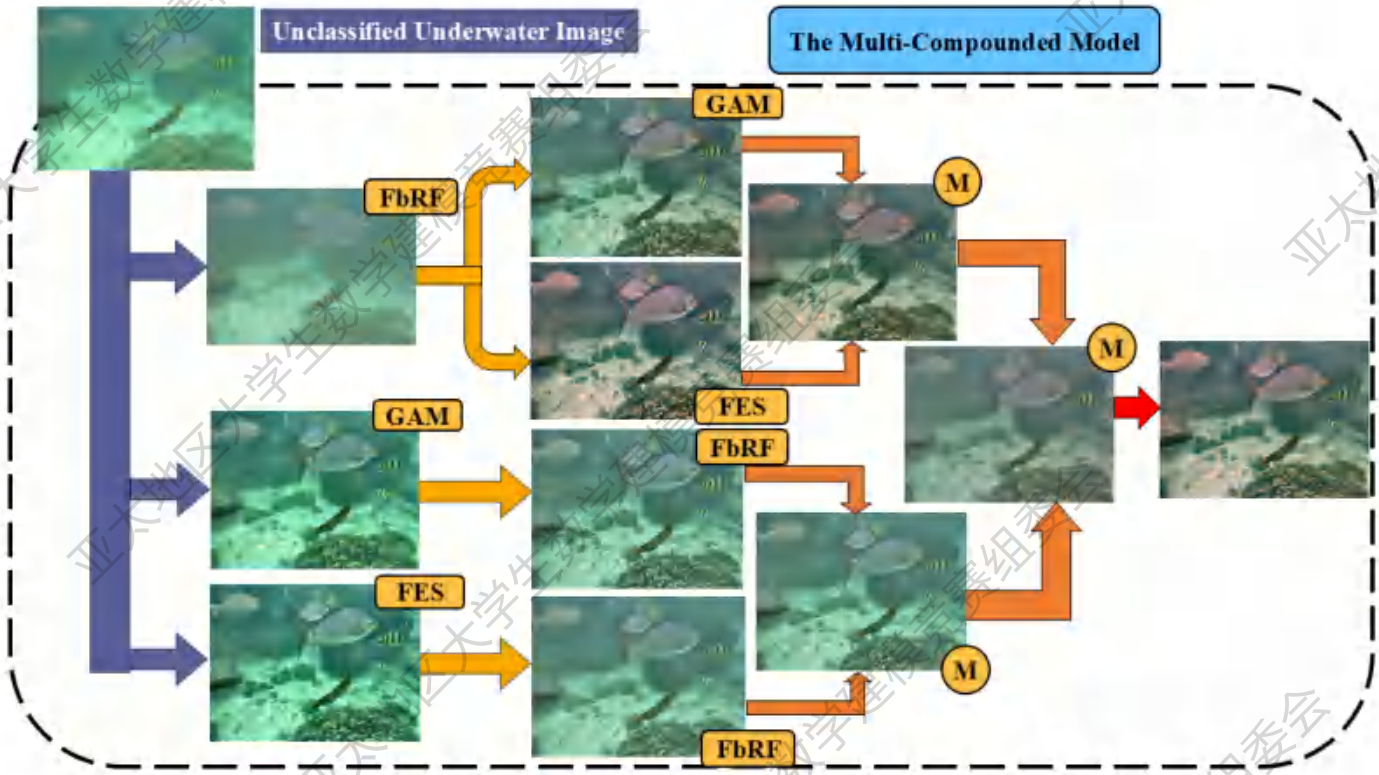
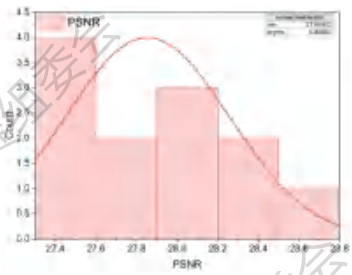


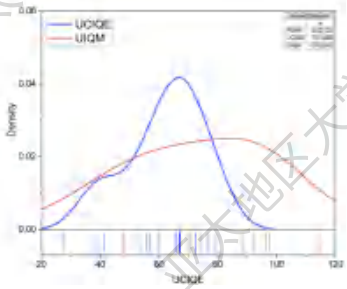
Figure 22 Applying Our Model to a Particular Image

As shown in Figure 23, it demonstrates the process of a complex image being processed by the Multi-Compound model.

With the high-quality test set (some degradated images and their corresponding reference images) used in FbRF, the trained Multi-Compounded Model is enable to tackle complex scenarios, and the enhanced results of the test images from attachment 2 and their corresponding evaluation metrics, such as PSNR, UCIQE, UIQM, are displayed below.



(a) PSNR



(b) UCIQE and UIQM

Figure 23 Three Evaluation Metrics of the Enhanced Images

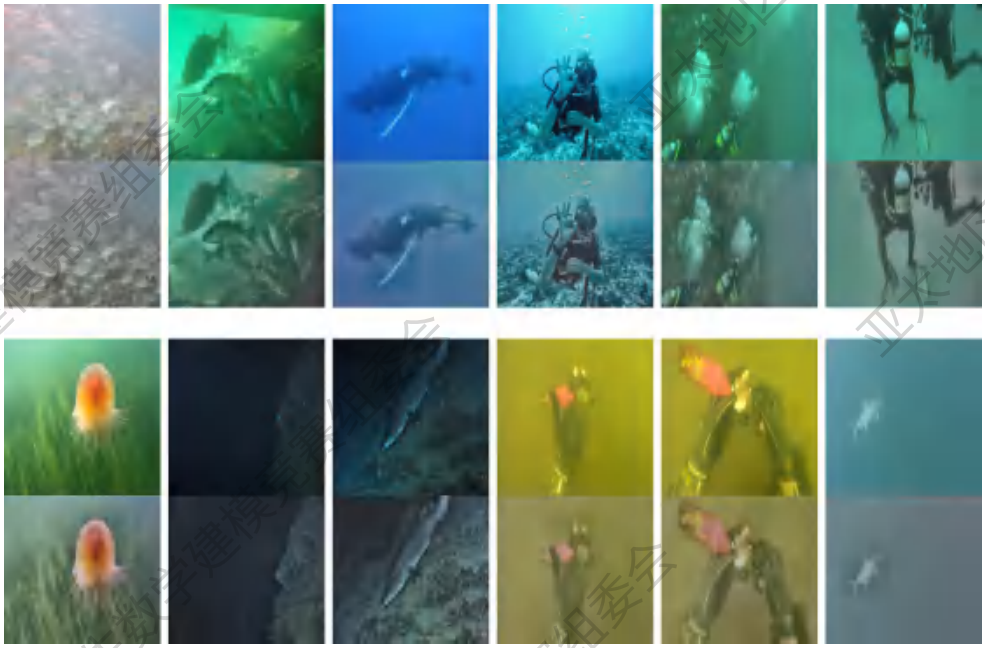
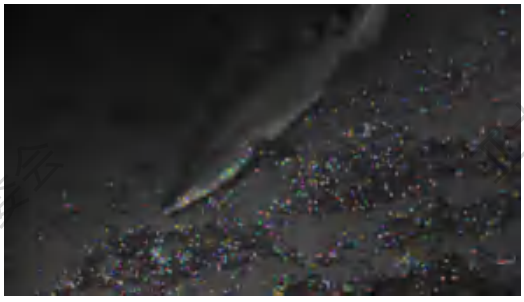


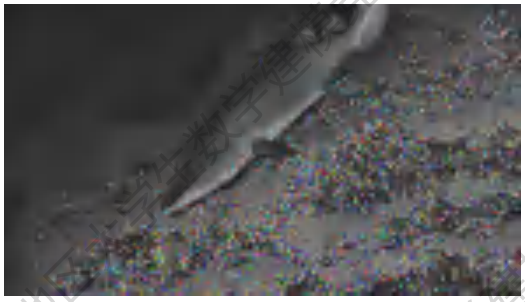
Figure 24 Comparison of the Image Before and After Multi-Compounded Model Processing

Faced with the complex underwater environment, the Multi-Compounded Model can e□ectively remove the color cast, enhance the contrast, and improve the image details, and the evaluation metrics are significantly improved. The quantized results show that the Multi-Compounded Model has a good performance in underwater image enhancement, and can e□ectively improve the quality of underwater images.

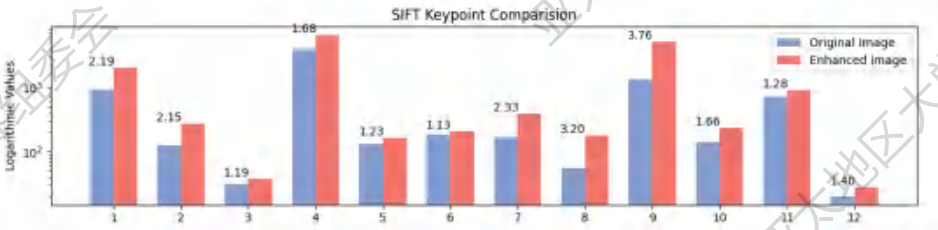
To better evaluate image quality, we introduce the SIFT (Scale-Invariant Feature Transform) algorithm to extract keypoints from the image and generate a descriptor for each keypoint. The descriptor represents the local image features of that point. Therefore, we believe that the more SIFT feature points there are, the richer the image is in details, indicating higher image quality.



(a) The SIFT keypoints in original image



(b) The SIFT keypoints in enhanced image



(c) The SIFT keypoints Comparision(Numbers represent ratio relationship)

Figure 25 Scale-Invariant Feature Transform Comparision Result

4.5 Question 5: Propose Feasibility Suggestions

Traditional underwater optical image enhancement methods based on physical principles, although having strong reliability, clear theories, and relatively saving computational resources, are basically verified in laboratory environments, lacking experimental verification in real outdoor underwater scenes and thus lacking certain practicability. Some methods are highly dependent on parameter estimation and have ill-posed problems in complex and changeable water environments. For example, the Low-light Model we used before, the method using orthogonal polarization di□erence proposed by Wang [7], and the method of backscattered pixel prior proposed by Zhou [10].

The underwater optical image enhancement methods based on deep learning, although able to automatically learn the mapping relationship between high-quality underwater images and low-quality underwater images through a large number of sample learning, are highly dependent on the training data set, have high requirements for the quality of the training data set, require large computational power, and small underwater cameras or detectors often cannot meet the requirements at present, and often have the problem of poor robustness. For example, the Random Forest Model we used before, the method of pyramid network structure, LCNet, based on CNN proposed by Jiang [3] , and the method of multi-scale generator and dual discriminator structure based on GAN proposed by Yang [8].

Therefore, we put forward the following suggestions:

1. **In actual underwater scenes, the methods based on physical principles should be combined with those based on deep learning according to the actual situation to balance the computational power problem and the ability to deal with complex situations.**
2. **Use multiple algorithms to improve the model's ability to deal with di□erent image degradation scenarios.**
3. **Improve the level of hardware perfection, use multimodal sensors to obtain scene information, explore the di□erent feature distributions of underwater multimodal information, analyze the correlation and complementarity between modalities, and e□ectively fuse the multimodal information to further improve the underwater image quality.**

5 Model Test

5.1 Sensitivity Validation

In the solution of Problem 2, we mainly apply the formula

$$t_c(z) = \kappa e^{-\beta_c^D \cdot z} + \mu e^{-\beta_c^B \cdot z}$$

when setting four parameters. Meanwhile, κ and μ can be "replaced" when β can be substituted to work by mathematical transformation. The sensitivity of the model to the parameters is tested by changing the parameters in a nuanced range and observing the change of the output value. The results are shown in Figure 26.

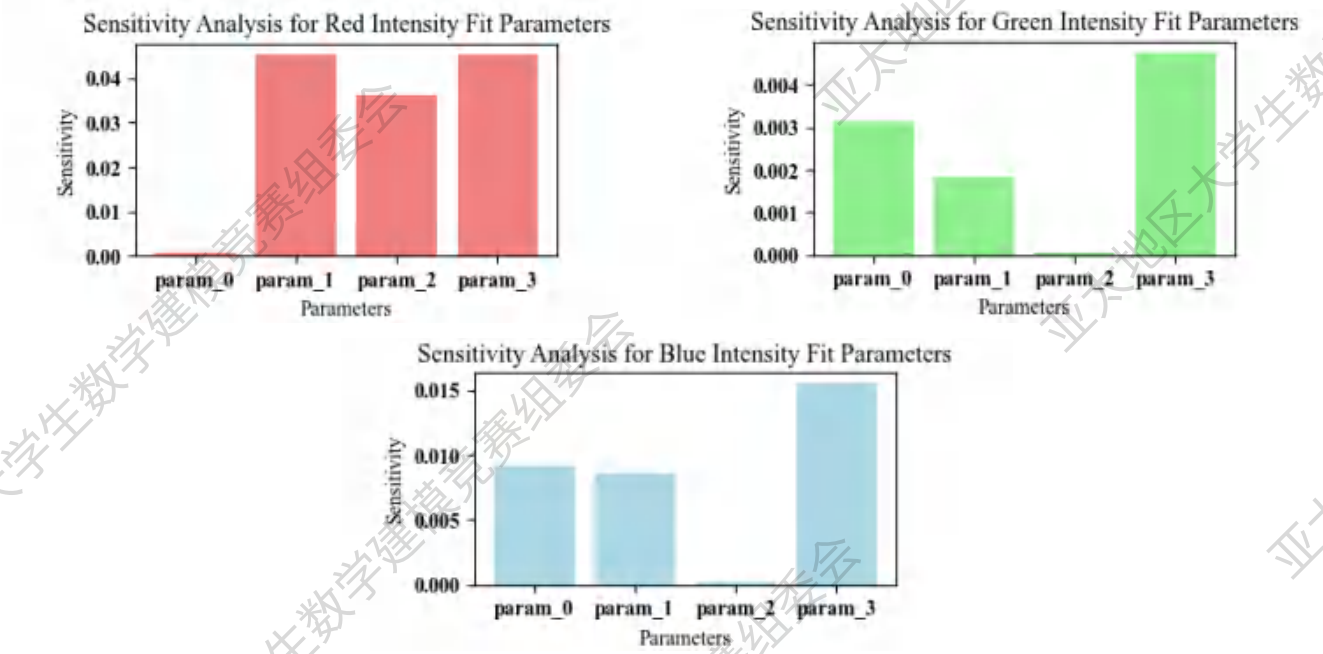


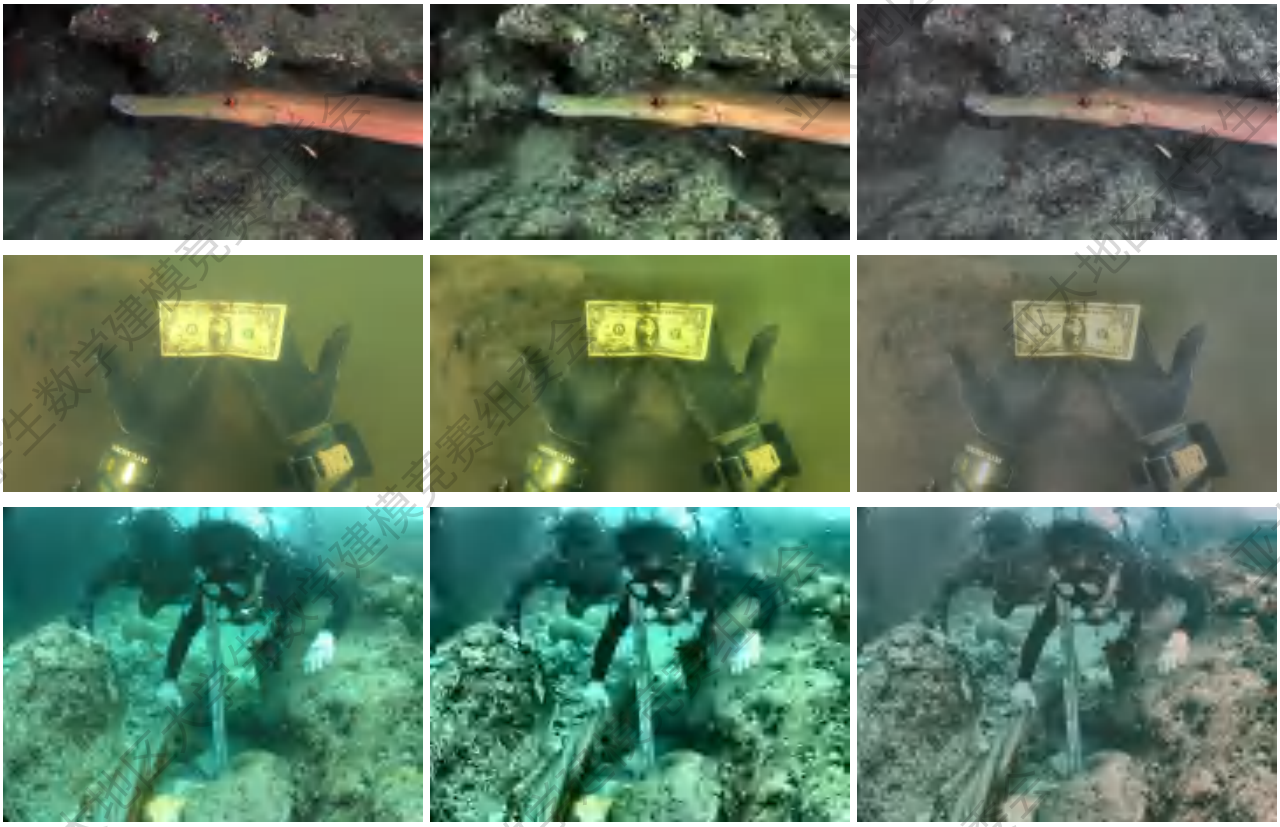
Figure 26 Sensitivity Validation of Problem 2

5.2 Robustness

As is obvious, each sensitivity of parameter is below 5%, in Blue channel the rate even goes down to 2%, which indicates that our model is smoothly reasonable when κ and μ are equal. And in di□erent curves, the order of variable sensitivity influence is also di□erent, which reflects the di□erence and similarity of their overall properties.

In order to make our model adaptable to di□erent situations, we need to test the robustness of the model. Robustness refers to the tolerance of the model to perturbations in the input data, that is, the stability of the output results when the input data changes. Therefore, we change the input test set and input pictures that do not conform to the underwater environment, such as food on the dining table, humans walking on land, and so on. The performance of our model can be partly detected from the figure below in these cases.

There are three groups of images, where the first column is the graph before augmentation, the second column is the graph obtained by training on the biased training set, and the third column is the training result of the original training set. By comparison, we can find that the contrast of the latter two groups of images is improved, and the brightness is also recovered well, which indicates that our model has a certain anti-interference ability for biased input data. That is, it has good robustness. However, it is clear that the second column is generally green, which is an inevitable result of the overall hue di□erence between on-land and underwater, and this is also one of the reasons why underwater image enhancement models need a good training set.



(g) Original Image (h) Biased-set Image (i) High-quality-set Image

Figure 27 The Comparison Graph of the Robustness Test

6 Strengths and Weaknesses

6.1 Analysis of Strengths

1. **The algorithm has a good ability for continuous optimization. The color di□erence reduction algorithm among our three core algorithms is based on the random forest algorithm, and the amount of data in the database can be further expanded to improve the code's performance. Due to time constraints, it is di□cult to collect more pictures as the training set, so the algorithm cannot be perfect, but it still has great potential and can achieve better restoration results.**
2. **The algorithm has a wide range of applications. This algorithm not only has a high restoration e□ect on underwater pictures, but also has a good restoration e□ect on various problems such as external color pollution, high blurriness, and low brightness. When solving the problem of underwater image restoration, we first extracted the main problems existing in the underwater image, then solved the individual problems one by one, and finally solved the multi-problem images through methods such as joint processing and image fusion. Such a method enables the algorithm to be widely applied to the restoration of other polluted pictures.**
3. **The algorithm has good universality. The algorithm only uses the physical model constructed based on underwater images and the machine learning method of random forest, and cleverly alternately uses single-factor processing and image fusion on the image to repair the image in complex situations. It avoids using**

more complex algorithms such as convolutional neural networks and generative adversarial networks, and can quickly transplant this algorithm to applications such as image processing in hazy environments and image processing in dynamic blur situations, with high portability and strong universality.

6.2 Analysis of Weaknesses

1. The restoration ability for backgrounds such as seawater is inferior to that for objects. Since the color of the seawater and other environments is relatively light and is easily interfered by external colors, it is di□cult to judge its original color. Therefore, it can only reduce the degree of external color interference, but cannot directly restore the original color of the object. In the future, e□orts can be made to separate the background and object, and better restoration results can be achieved by processing the background separately.
2. Although the degree of image restoration is relatively high, the restoration still starts from the overall picture. In actual life, the restoration methods for di□erent objects should be slightly di□erent. Subsequently, the code can be upgraded in this regard. First, the objects in the picture are separated, and each object is restored separately and then the overall picture is recombined. In this way, the restoration of each part in the picture will be more in place.

7 References

[1] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6723–6732, 2018.

[2] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 04 2019.

[3] Nanfeng Jiang, Weiling Chen, Yuting Lin, Tiesong Zhao, and Chia-Wen Lin. Underwater image enhancement with lightweight cascaded network. IEEE Transactions on Multimedia, 24:4301–4313, 2022.

[4] Weiguo Jiang, Yundong Zhang, Fuxing Zhu, Ying Guo, and Guo Yi. Regulation of fast and slow light characteristics of the add-drop ring-resonator employing an assisted ring. Opt. Express, 29(4):5141–5151, Feb 2021.

[5] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. IEEE Journal of Oceanic Engineering, 41(3):541–551, 2016.

[6] E. Trucco and A.T. Olmos-Antillon. Self-tuning underwater image restoration. IEEE Journal of Oceanic Engineering, 31(2):511–519, 2006.

[7] Hongyuan Wang, Haofeng Hu, Junfeng Jiang, Xiaobo Li, Weihong Zhang, Zhenzhou Cheng, and Tiegen Liu. Automatic underwater polarization imaging without background region or any prior. Opt. Express, 29(20):31283–31295, Sep 2021.

[8] Miao Yang, Ke Hu, Yixiang Du, Zhiqiang Wei, Zhibin Sheng, and Jintong Hu. Underwater image enhancement based on conditional generative adversarial network. Signal Process. Image Commun., 81, 2020.

[9] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. IEEE Transactions on Image Processing, 24(12):6062–6071, 2015.

[10] Jingchun Zhou, Tongyu Yang, Weishen Chu, and Wei shi Zhang. Underwater image restoration via backscatter pixel prior and color compensation. Eng. Appl. Artif. Intell., 111:104785, 2022.

Team Number:	apmcm24200636
Problem Chosen:	B

2024APMCM summary sheet

This problem pertains to optimization planning, where single-objective and multi-objective programming models were established to determine the optimal shape and size parameters of the equipment using various suitable optimization algorithms.

For Problem 1, a high-performance air-conditioner optimization model was developed, simulating temperature variations in summer and winter under different influencing factors, and designing the optimal shape and size. Initially, the Navier-Stokes equation was used to describe airflow. Subsequently, a single-objective programming model was established with objectives of temperature uniformity, energy efficiency, and noise control, with volume as a constraint. The model was solved using a combination of random search and finite difference methods, identifying wind speed, quantity, and spatial position as key factors, with the optimal shape and size being a cube of 0.5m. Simulations of temperature regulation in summer and winter showed better performance in summer.

For Problem 2, a high-performance air purifier optimization model was established to design the optimal shape and size. The Advection-Diffusion equation was used to simulate molecular diffusion. A single-objective programming model was then established with maximum purification efficiency as the objective and purifier volume as the constraint, solved using the gradient descent algorithm. The optimal shape and size were determined as a rectangular prism of 0.82m×0.92m×1.18m. Robustness testing showed minimal impact on purification under uneven pollutant distribution, indicating high model robustness.

For Problem 3, a high-performance humidifier optimization model was developed to determine the optimal shape and size. Fick's law was introduced to establish a water vapor diffusion equation. A single-objective programming model was then established with optimal humidification effect as the objective and volume limitation as the constraint, solved using the BFGS optimization algorithm. The optimal shape and size were found to be a rectangular prism of 1.26m×0.2m×0.5m. Sensitivity analysis showed minimal change in maximum humidification efficiency with increasing power, indicating good model stability.

For Problem 4, a Pareto optimization model was established to determine the optimal shape and size of a three-in-one device. Building on the previous models, this model comprehensively considered energy efficiency, human comfort, purification effect, and humidification effect maximization as objectives, establishing a multi-objective programming model. The model was solved using the SLSQP algorithm, yielding an optimal shape and size for the three-in-one device as a rectangular prism of 1.24m×1.24m×2.15m, with defined height ranges for each functional module within the device.

Keywords: Navier-Stokes equation; Adveaux-diffusion equation; Fick's law; various optimization algorithms

Directory

1 Introduction....3
1.1 Problem Background....3
1.2 Question Restatement....3
2 problem analysis....4
2.1 analysis of problem 1....4
2.2 analysis of problem 2....4
2.3 analysis of problem 3....5
2.4 analysis of problem 4....5
3 Model assumptions....5
4 Symbol Description....6
5 Our Work....6
6 establishment and solution of the model....7
6.1 Model establishment and solution of problem 1....7
6.2 Model establishment and solution of problem 2....16
6.3 Model establishment and solution of problem 3....20
6.4 Model establishment and solution of problem 4....24
7 evaluation and generalization of the model....26
7.1 evaluation of the model....26
7.2 generalization of the model....27
References....28
Appendix....29

1 Introduction

1.1 Problem Background

At present, as economic development enters a high-quality stage, people's demand for quality of life is also increasing. As far as environmental electrical equipment is concerned, environmental conditioning devices combining air conditioners, humidifiers and air purifiers are gradually emerging. The device fully integrates the three devices and replenishes each other. The air conditioner will dry the indoor air while adjusting the temperature, and the humidifier can replenish moisture in time to properly adjust the indoor temperature. And when the indoor air circulation caused by the operation of the air conditioner, the air purifier can filter harmful pollutants in the air, purify the circulating air, and improve the indoor air quality. In addition, the three-in-one device is equipped with electrical appliances that control 3 devices at the same time, reducing the power supply and connecting wires of multiple devices, and reducing the potential safety hazards of excessive lines.

Through variable refrigerant flow, the air conditioner integrates the compressor, condenser and evaporator to form an artificial entity system with the function of providing a comfortable indoor environment. The working principle of the air purifier is to use high-voltage electricity to convert air molecules into positive ions and free electrons, so that they are adsorbed and deposited on the deposition plate. When the air enters the photocatalyst catalytic decomposition area, the harmful substances in the air are decomposed into negative oxygen ions and water. At this point, the air purification work is completed. The original reason of the humidifier is that alternating current is transformed into direct current through transformer step-down rectification, and then the direct current is transformed into high-frequency alternating current through an oscillation circuit and added to ultrasonic atomization. If external force is applied to it to deform, it will output corresponding voltage and current. On the contrary, applying alternating voltage to it will produce corresponding deformation, that is, vibration. In this way, the ceramic oscillator is driven by the high frequency voltage to produce ultrasonic oscillation, so that the water is atomized, and then the water mist is sent out by the fan to achieve the effect of humidifying the air.

1.2 Question Restatement

Access to relevant information shows that the design of integrated air conditioners, humidifiers and air purifiers is closely related to aerodynamics. The overall shape, position, quantity, direction, angle, wind speed and air volume of the air inlet and outlet will affect the effectiveness of the three integrated environmental equipment.

As the manager of the air conditioner appearance optimization team, the mathematical model of the three-in-one air conditioner appearance optimization is established according to the specific topic, and the air conditioner appearance is designed for a room with an approximate cuboid space area of 5m × 8m × 3m and is about to be renovated.

For the problem 1, considering the different shapes of air conditioning, the

influence factors of air conditioning position, air inlet position, quantity, wind direction, angle, wind speed and air volume on air conditioning efficiency are analyzed, and the relationship between indoor temperature and space-time change in summer and winter is simulated. At the same time, the shape optimization model of the air conditioner is determined, and the best shape and size of the air conditioner are designed to achieve the best temperature adjustment effect.

For the problem 2, in order to maximize the purification effect, according to the different shapes of the air purifier, the optimization model of the shape of the air purifier is established, the influence of the shape of the air purification on the purification effect is analyzed, and the best size and shape of the purifier is designed, and the best shape and size parameters are drawn.

For the 3 of the problem, in order to maximize the humidification effect, based on the different shapes of the air humidifier, the influence of the shape of the air humidifier on the humidification effect is analyzed, the air humidification shape optimization model is established, the best shape of the humidifier is designed, and the best shape and size parameters are drawn.

For the problem 4, in order to maximize the energy efficiency, human comfort and purification and humidification effect, based on the first three questions model, the optimization design model is established, the best appearance of the 3 equipment is designed, and the shape and size of the best appearance are drawn.

2 problem analysis

2.1 analysis of problem 1

This question requires the establishment of a model to simulate the change of indoor temperature with time and space in different seasons, establish an optimization model, and design the best shape and size of the air conditioner when the air conditioner reaches the maximum adjustment effect. Considering that the air flow generated by the air-conditioning gas in the flow process belongs to the compressible Newtonian fluid, the Navier-Stokes equation is established here to describe the air flow and then, with the uniformity of temperature, the efficiency of energy, the high control of noise as the target function, and the volume as the constraint to establish a single-target planning model. Then, the random search algorithm solves the above model, and determines the best shape and size of the air conditioner under the change of room temperature with time and space in summer and winter, as well as the best temperature regulation effect.

2.2 analysis of problem 2

This topic requires considering different shapes of air purifiers, maximizing the purification effect, designing the best shape and best parameters of the air humidifier, and visualizing them. Considering that the essence of an air purifier is to absorb tiny substances in the air, the diffusion of tiny substances can be described by the Adwex-diffusion equation. Then, the single-objective planning model is established with the optimal air purification effect as the target function and the volume limit as the

constraint. Then, the model is solved by using the gradient descent algorithm that makes the target function drop to the minimum value along the fastest path, and the best shape and size of the air purifier when the purification effect is optimal are calculated, and the visual description is made. This paper uses robustness analysis to further analyze the purification effect of air purifiers by changing the distribution of pollutants and changing the distribution from random distribution to normal distribution.

2.3 analysis of problem 3

This topic requires considering different shapes of air humidifiers, maximizing the humidification effect, designing the best shape and best parameters of the air humidifier, and visualizing them. The effect of air humidifier is mainly measured by humidity. The shape and size of air purification affect the release mode and distribution of water vapor. Then the water vapor diffusion equation is introduced here. Fick's law is introduced here to establish the water vapor diffusion equation to describe the water vapor concentration. Then, with the maximization of air humidification effect as the target function and the volume limit as the constraint, a single-goal planning model is established. And then,

The BFGS optimization algorithm of Newton method is used to solve the single-target planning model, and the best shape and size when the humidification effect is optimal are calculated. Finally, this paper uses sensitivity analysis to test the model and change the maximum power to calculate the degree of change in the purification effect.

2.4 analysis of problem 4

This question requires optimizing the appearance design of the 3 device on the basis of the 3 of the problem 1. 2., establishing an optimization model, designing the best appearance, and drawing the shape and size parameters of the best appearance. In this paper, the Pareto model is used to integrate the above-mentioned single-goal planning model, and the multi-objective planning model is established with the objective function of energy efficiency, human comfort, purification effect and humidification effect. Then, the machine learning-SLSQP algorithm is used to solve the above model, determine the best appearance shape and size parameters, and draw the best appearance shape.

3 Model assumptions

- Assuming that the air in the room is an incompressible ideal gas, the thermophysical properties of the air (such as specific heat capacity, thermal conductivity, etc.) are constant.
- Ignore the heat capacity of the wall and think that the wall temperature remains constant
- Assuming that the air flow blown by the air conditioner is evenly mixed in the room
- Assuming uniform temperature distribution in the initial time space

- Assuming that the air conditioner, air purifier and air humidifier will not affect each other when they work together
- Assuming that the temperature in the room is only affected by the air conditioning

4 Symbol Description

Symbol	Description	Unit
u	air circulation speed	Meters per second (m/s)
p	Air pressure	Pa (Pa)
μ	aerodynamic viscosity	Pa·s
T_{target}	Target temperature	Degrees Celsius (°C)
t	Time interval	Second (s)
C_w	Volume concentration of water vapor	Molar concentration (mol/m³)
V_h	Volume of air humidifier	Cubic Meters (m³)
a	Step size of learning rate	Unit value (1)
S	Surface area of humidifier	Square meters (square meters)
Q	Air volume	Cubic meters per second (m³/s)

5 Our Work

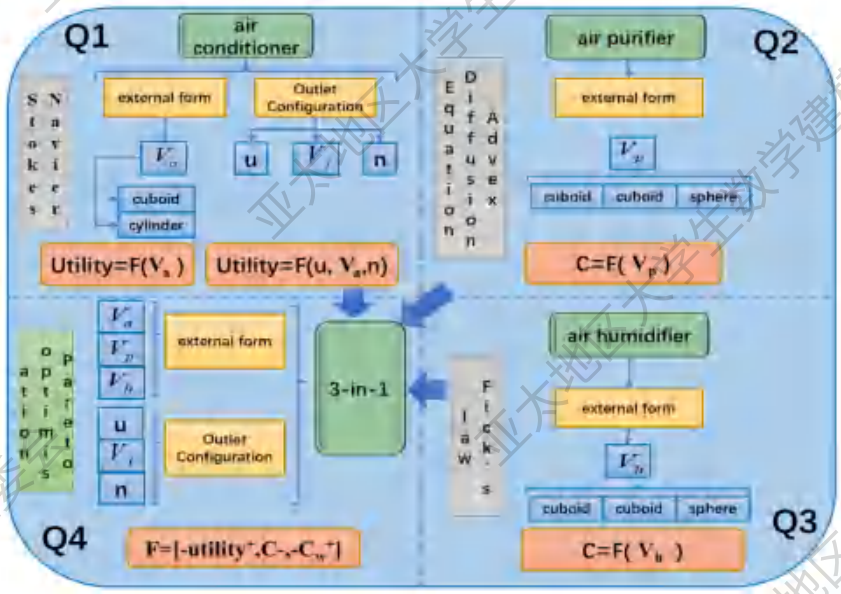


Figure 1Our Wrok

The above figure shows the overall idea of solving the problem in the form of a flowchart, and determines the best solution to the problem by using the goal planning

model. The overall idea is from data processing to model construction to target planning model construction, and finally to determine the best solution and use the corresponding algorithm or directly use Matlab software to calculate, to get the answer. The specific process is shown below.

6 establishment and solution of the model

6.1 Model establishment and solution of problem 1

6.1.1 Establishment of air circulation equation

Air flow belongs to compressible Newtonian fluid, then the Navier-Stokes equation [1]is established here to describe the airflow, which is essentially momentum conservation and can be expressed:

$$\rho(\frac{\partial u}{\partial t}+u\bullet\nabla u)=-\nabla p+\nabla\bullet(\mu(\nabla u+(\nabla u)^T)-\frac{2}{3}\mu(\nabla\bullet u)I)+F \tag{1}$$

u Indicates the air circulation velocity, p the air pressure, ρ the air density, and μ the aerodynamic viscosity.

This equation is solved simultaneously with the continuity equation, I .e., mass conservation, expressed:

$$\frac{\partial\rho}{\partial t}+\nabla\bullet(\rho u)=0 \tag{2}$$

To relate the Navier-Stokes equation to temperature, heat is introduced based on $Q=\nu R\Delta T$. For any object, the law of conservation of energy is satisfied. The length, width and height of the $5m\times8m\times3m$ room are taken as the base, a spatial rectangular coordinate system is established, and the microelements in the given fixed volume room are taken to meet the $Q_{\inf}+Q_s=Q_T$. $Q_{\inf}$ is the inflow heat, Q_s is the heat source heat, and Q_T is the heat that increases the temperature. Taking into account the derived $\vec{Q}=-k\Delta u$ from Fourier's law of heat conduction, that is, heat flows from high temperature to low temperature, the amount of heat flowing in a certain direction is proportional to the reduction rate of temperature in that direction, and it is substituted into heat conservation, that is, the heat transfer equation is derived:

$$\frac{\partial u}{\partial t}=\alpha^2\Delta u+F(t,\vec{x}) \tag{3}$$

Where $\alpha=\sqrt{\frac{\kappa}{c\rho}}$, $F(t,\vec{x})=\frac{g(t,\vec{x})}{c\rho}$, substituting (3), the heat transfer equation can be simplified:

$$\frac{\partial u}{\partial t}=\frac{\kappa}{c\rho}\Delta u+\frac{g(t,\vec{x})}{c\rho}=\frac{1}{c\rho}(\kappa\Delta u+g(t,\vec{x})) \tag{4}$$

The κ in the formula is the thermal diffusion coefficient, the $g(x,y,z,t)$ represents the heat production rate of the internal heat source per unit volume, and the c is the specific heat capacity of the air.

Finally, the equation of (1)(2)(4) can describe the air flow and temperature change is obtained.

$$\rho c\frac{\partial T}{\partial t}+\rho\bullet c\bullet(\vec{u}\bullet\nabla T)=\kappa\nabla^2 T+Q_s \tag{5}$$

This equation establishes a connection between air density, specific heat capacity, temperature, flow velocity and heat source, which can be used as a theoretical basis for the design of air conditioning outlet configuration and air conditioning shape.

6.1.2 The establishment of high-performance air-conditioning optimization model.

Question given the relevant factors will be on the air conditioning working state to make a certain degree of adjustment, thus affecting the air conditioning refrigeration effect, this part will be 6.1.1 section of the establishment of the flow temperature differential equation [2]based on, respectively, on the air conditioning design to discuss. The relevant factors given can be divided into five categories. The 1 category is the position of the air conditioner and the position of the air inlet. The different positions will determine the flow path of the air flow, thus affecting the uniformity of the temperature in a given room. The second category is the number of air inlets. The increase in the number of air inlets will improve the air circulation, but too many air inlets will also increase the difficulty of air conditioning design, so the advantages and disadvantages need to be considered in a balanced manner. The third category is the direction and angle of air inlets, it can determine the range that the air flow can cover, and a better angle can further improve the use effectiveness of the air conditioner. The fourth category is air volume, which determines the speed of air replacement, thus affecting the working efficiency of the air conditioner. The last category is wind speed, which should be moderate. If the wind speed is too high, it will lead to uncomfortable reaction of users, and noise will be generated when the wind speed is higher, which will affect the user's use feeling.

In order to simplify the model, the simplified design objectives are initially divided into two parts:

(1) air conditioning design

Shape and size of air conditioner

(2) Air outlet configuration design

Air conditioning position, air inlet position, quantity, wind direction, angle, wind speed and air volume

First of all, assume that the air outlet configuration is the same, only change the shape and size of the air conditioner, and based on market research, the shape of the air conditioner is mainly cylindrical and cuboid, then the shape will be discussed in these two shapes, so as to determine the most air conditioner shape and size, as shown in the

figure:



Fig. 2Design Drawing of Left (Cylinder) and Right (Cuboid) Tuyere

After determining the optimal shape and size of the air conditioner, the sealing configuration is designed to determine the optimal air conditioner position, air inlet and outlet position, quantity, wind direction, angle, wind speed and air volume.

6.1.2.1 objective function

Optimizing the performance of air conditioning is carried out from the 3 aspects, the 1 is the uniformity of temperature, the 2 is the efficiency of energy, and the last is the high control of noise

(1) temperature uniformity

The temperature distribution of high-performance air conditioners in the temperature field should tend to be uniform, then the first goal is to construct: temperature uniformity, which can be expressed:

$$Tempreture\_uniformity = \frac{1}{V}\int \left| T(x,y,z) - T_{target} \right|^2 dV \tag{6}$$

The $T(x,y,z)$ represents the temperature of each coordinate in the temperature field [3]of the three-dimensional coordinate system of the room based on the air circulation model, and the T_{target} represents the target temperature set by the user for the air conditioner. The temperature uniformity of the air conditioner is expressed by the sum of the temperature differences in the temperature field of the whole room. The sum shows that the air conditioner as a whole minimizes the difference between the set temperature and the actual temperature at each position in the room, that is, the temperature uniformity is well performed.

The first 1, second, and third categories in the indicator will have a greater impact on the uniformity of temperature.

(2) Energy efficiency:

The second goal is the highest energy efficiency. The energy efficiency of air conditioning refrigeration and heating can be expressed in terms of cooling energy efficiency ratio (EER) and heating energy efficiency ratio (COP). Here it can be summarized as efficiency can be expressed:

$$\eta = \frac{Q_s / \Delta t}{P} \tag{7}$$

Q_s For the heat source, the cooling capacity is the cooling capacity, while the heating is the heating capacity, the Δt is the heat source action time, and the P is the air conditioning electric power. The fourth category of indicators will have a greater

impact on energy efficiency.

(3) High noise control

According to the data, the noise between wind speed and noise increases with the increase of wind speed and produces specific noise under specific wind speed. The shape of the air conditioner also affects its noise, considering that the design of the air conditioner in this paper is based on the existing air conditioner on the market to improve, then only consider the impact of wind speed at this time, in order to simplify the model, build the noise function and set up a coefficient based on this, expressed:

$$noise = \delta u + \varepsilon \tag{8}$$

Among them $\delta > 0$, this formula reflects the wind speed and noise size is positively correlated.

Taking the 3 characteristics as the target function, the multi-objective planning model is constructed, and in order to simplify the calculation difficulty of the model, the utility function of air conditioning is constructed here, and the multi-objective planning is transformed into a single-target planning model.

The utility function for air conditioning is expressed:

$$utility = \alpha_1 \eta - \alpha_2 Tempreture\_uniformity - \alpha_3 noise \tag{9}$$

Among them, α_1 , α_2 , and α_3 are their weighting coefficients, and their Sanskrit 3 the priority of one indicator.

6.1.2.2 constraint condition

If the volume of the air conditioner is V_a , if it is a cuboid, its length, width and height are a , b and c respectively, it is $V_a = abc$. If it is a cylinder, the diameter is d , and it is $V_a = \frac{1}{4}\pi d^2 c$. If it is known that the maximum volume of the air conditioner is 0.1 cubic meters, it can be expressed as follows:

$$V_a \le 0.1 \tag{10}$$

The power should be less than the rated power, and the wind speed should be less than the maximum wind speed, which is expressed:

$$P \le P_e = 1800 \tag{11}$$

The wind speed shall be less than the maximum outlet speed:

$$u \le 8 \tag{12}$$

Based on $c = \frac{Q_{\inf}}{m\Delta T} = \frac{Q_{\inf}}{\rho V_t \Delta T}$, where V_t is the airflow volume of unit tuyere, assuming that each tuyere is equal in size, the number of seals is n , and the maximum inlet and outlet flow rate is 600 cubic meters, which can be expressed:

$$nV_t = \frac{Q_{\inf}}{c\rho\Delta T} \le 600 \tag{13}$$

6.1.2.3 establishment of optimization model

Based on the objective function and constraints, the goal is to maximize the efficiency of user use, while the decision variables are different in the two design steps and will be described separately. When designing the shape and size of the air conditioner, the V_a is used as the decision variable when changing the size and shape of the air conditioner, while the air inlet and outlet design will take the wind speed u, the air volume V_t of the unit air inlet and outlet, and the number of air outlets as the n, and the single-target planning model will be organized as follows:

$$\min(-utility)$$

$$s.t.\begin{cases} V_a \le 0.1 \\ P \le 1800 \\ \mathrm{u} \le 8 \\ nV_{\mathrm{t}} \le \dfrac{Q_{\inf}}{c\rho\Delta T} \le 0.6 \\ \eta = \dfrac{Q_s / \Delta t}{P} \\ utility = \alpha_1\eta - \alpha_2 Tempreture\_uniformity - \alpha_3 noise \\ noise = \delta u + \varepsilon \\ Tempreture\_uniformity = \dfrac{1}{V}\int \left|T(x,y,z) - T_{target}\right|^2 dV \\ \rho c\dfrac{\partial T}{\partial t} + \rho \bullet c \bullet (\vec{u} \bullet \nabla T) = \kappa \nabla^2 T + Q_s \\ Q_{\inf} + Q_s = Q_T \end{cases} \tag{14}$$

6.1.3 The solution of the optimization model of high-performance air conditioning.

The above target planning model is solved by random search algorithm, and the basic principle of the algorithm is to randomly sample the 1 group parameters in a given parameter space and calculate the target function value of the parameter combination. Through continuous random sampling and evaluation, the best parameter combination that makes the target function reach the minimum or maximum value is finally found.

Because the model contains differential equations, this paper uses the finite difference method [4]to optimize the solution of the model on the basis of random search algorithm. The finite differential method is the numerical solution of the 1 differential equation, which is one of the most basic and most extensive numerical methods for time dispersion and space dispersion, and is widely used in the fields of computational fluid mechanics, computational heat transfer, computational electromagnetics and so on.

Taking a one-dimensional problem as an example, suppose there is a function f(x) about x, using Taylor expansion:

$$f(x+\Delta x)=f(x)+\frac{\Delta x}{1!}\frac{\partial f}{\partial x}+\frac{\Delta x^2}{2!}\frac{\partial^2 f}{\partial x^2}+O(\Delta x^3) \tag{15}$$

The 1-order derivative that ignores the remainder can be expressed:

$$\frac{\partial f}{\partial x}\approx\frac{f(x+\Delta x)-f(x)}{\Delta x} \tag{16}$$

This format is called Euler forward difference, and because of the 1 order function of the ignored term x, it has 1 order accuracy and is expected to be truncated error.

Similarly, Euler backward difference is;

$$\frac{\partial f}{\partial x}\approx\frac{f(x)-f(x+\Delta x)}{\Delta x} \tag{17}$$

Euler's backward difference also has the 1-order accuracy, in order to build the second-order accuracy, first, the above two sets of equations are expanded with Taylor:

$$\begin{cases} f(x+\Delta x)=f(x)+\dfrac{\Delta x}{1!}\dfrac{\partial f}{\partial x}+\dfrac{\Delta x^2}{2!}\dfrac{\partial^2 f}{\partial x^2}+\dfrac{\Delta x^3}{3!}\dfrac{\partial^3 f}{\partial x^3}+O(\Delta x^4) \\ f(x-\Delta x)=f(x)-\dfrac{\Delta x}{1!}\dfrac{\partial f}{\partial x}+\dfrac{\Delta x^2}{2!}\dfrac{\partial^2 f}{\partial x^2}-\dfrac{\Delta x^3}{3!}\dfrac{\partial^3 f}{\partial x^3}+O(\Delta x^4) \end{cases} \tag{18}$$

Subtract the two, ignoring the remainder, and get:

$$\frac{\partial f}{\partial x}\approx\frac{f(x+\Delta x)-f(x-\Delta x)}{2\Delta x} \tag{19}$$

The above formula is the central differential format of second-order accuracy.

6.1.4 Seasonal simulation

In order to further simulate the air-conditioning gas fluid process in summer and winter, this paper uses Lagrangian CFD [5]to describe the fluid motion state. The Lagrangian method focuses on the study of the motion of each fluid particle, describing the process of motion of the fluid particle from beginning to end, and the law of change of physical quantities with time t.

$$\nabla=(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z})=\frac{\partial}{\partial x}i+\frac{\partial}{\partial y}j+\frac{\partial}{\partial z}k \tag{20}$$

Velocity gradient:

$$\nabla V=(\frac{\partial u_x}{\partial x},\frac{\partial u_y}{\partial y},\frac{\partial u_z}{\partial z}) \tag{21}$$

divergence:

$$\nabla V=(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z})\cdot(u_x,u_y,u_z)=\frac{\partial u_x}{\partial x}+\frac{\partial u_y}{\partial y}+\frac{\partial u_z}{\partial z} \tag{22}$$

Rotary:

$$\nabla V = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial x} & \frac{\partial}{\partial x} \\ u_x & u_y & u_z \end{vmatrix} = ((\frac{\partial u_z}{\partial y} - \frac{\partial u_y}{\partial z})i, (\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x})j, (\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y})k) \tag{23}$$

Acceleration:

The local acceleration caused by the unsteady flow and the displacement acceleration caused by the non-uniform flow are composed of two parts.

$$a = \frac{\partial u_y}{\partial t} + (\nabla V)V \tag{24}$$

The optimal volume of the air conditioner and the temperature setting conditions of the air conditioner in summer and winter are determined by the above random search algorithm. The following table shows the temperature setting in summer and winter, as shown in Figure 3 and Figure 4.

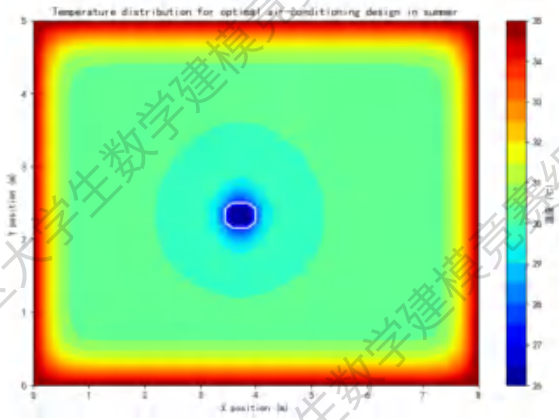


Fig. 3Air conditioning temperature setting in summer

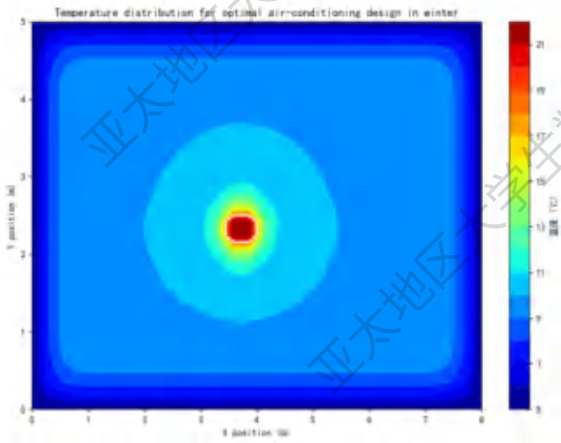


Fig. 4Air conditioning temperature setting in winter

This article shows the optimal volume setting of the air conditioner and the temperature utility here:

Table 1 1Best state of air conditioning

Status	Winter	Summer

Long	0.5	0.5
Wide	0.5	0.5
High	0.5	0.5
utility	7.84	7.93

The air conditioner shown above is in the best state, and the best shape of the air conditioner is cuboid. At this time, the best volume of the air conditioner in winter and summer is ^{3} 0.75m. Among them, the summer air conditioning design is more effective and the temperature adjustment effect is better.

Taking into account the air conditioning is affected by the location, air inlet location, quantity, wind direction, angle, wind speed and air volume and other factors, the indoor temperature changes with time and space under different conditions in summer and winter are simulated. This paper analyzes the change of indoor temperature from 3 angles of air outlet, wind speed, number of air conditioners and placement position. The specific change is shown in the following figure:

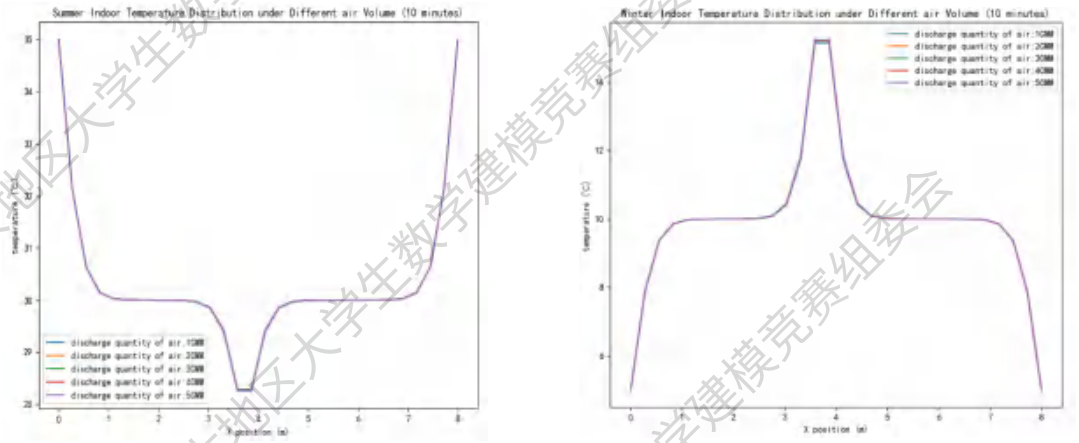
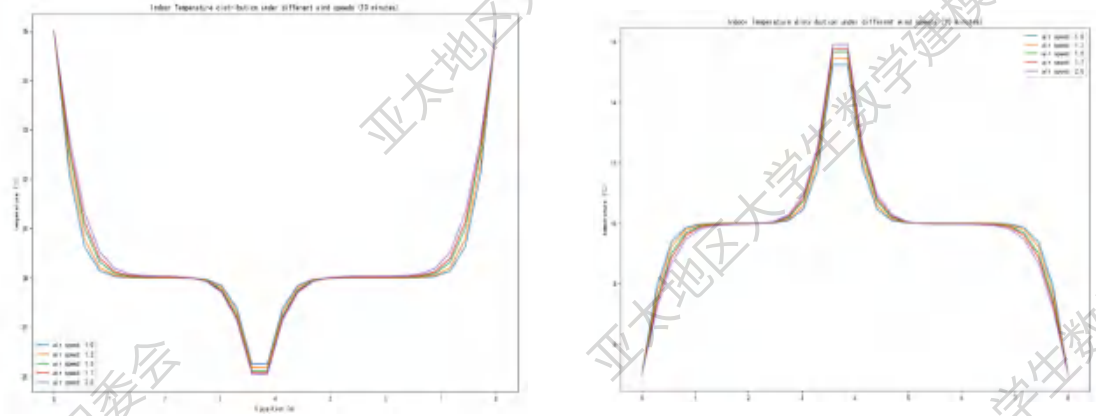


Fig. 5Influence of air output on temperature in summer (left) and winter (right)

In this paper, summer and winter as the sub-point, the air volume in accordance with 1-5m ^{3} /s respectively to simulate the temperature changes, the results show that



in the two seasons of summer and winter, the air volume and temperature changes, the basic coincidence, indicating that the air volume on the air conditioning temperature effect is small.

Figure 6Influence of wind speed on temperature in summer (left) and winter (right)

In this paper, summer and winter are used as sub-points, the air output is 1 m/s as the initial speed, 0.2 m/s as the step size, and the temperature change is simulated

respectively. The results show that in the two seasons of summer and winter, the change of wind speed and temperature can be clearly seen that the change trend is different, indicating that the air output has obvious influence on the temperature effect of air conditioning.

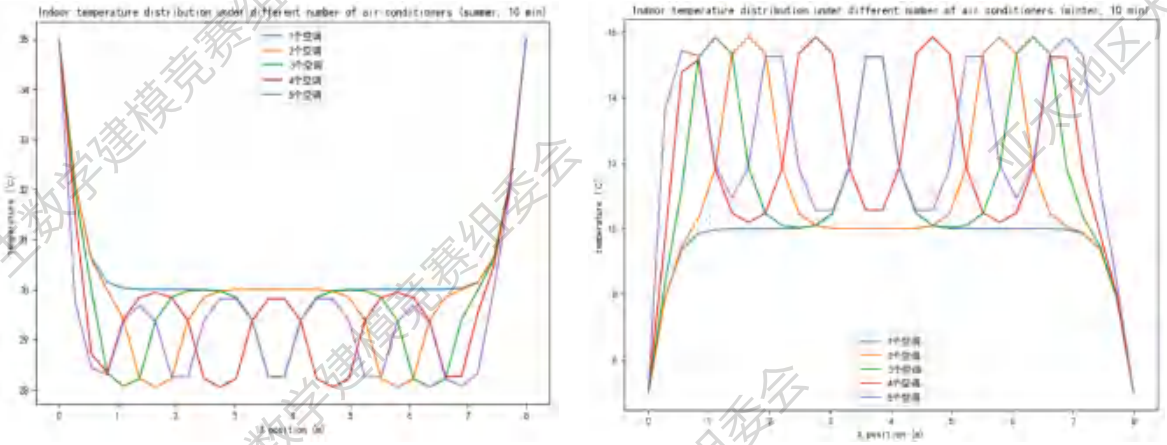


Fig. 7Change of air conditioning quantity on temperature in summer (left) and winter (right)

In this paper, the summer and winter points, respectively, the number of air conditioning 1-5 as the research object, the temperature change simulation, the results show that in the summer and winter two seasons, the number of air conditioning and temperature changes, it can be clearly seen that there is a big difference in the trend, indicating that the number of air conditioning on the temperature effect of air conditioning.

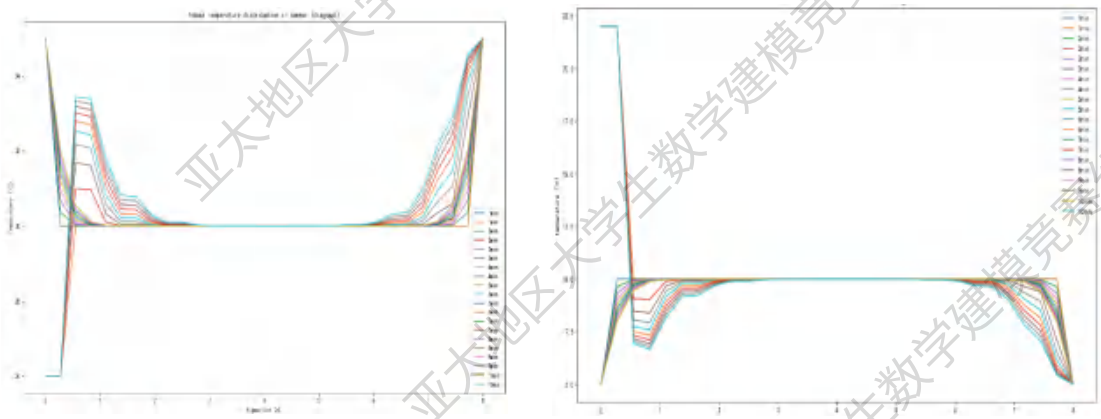


Figure 8Summer (left) and winter (right) air conditioning position on the diagonal changes in temperature

In this paper, summer and winter are taken as sub-points, and the time span of 1-10 minutes of the number of air conditioners is taken as the research object to simulate the temperature change. The results show that in the two seasons of summer and winter, the position and temperature change of each air conditioner on the diagonal of the room can be clearly seen that there are obvious differences in the change trend, which shows that the temperature effect produced by the air conditioner is greater when the position on the diagonal.

In this paper, summer and winter as the points, respectively, 1-10 time span as the research object, the temperature change simulation, the results show that in the summer and winter two seasons, each air conditioning position in the middle and the

temperature change, it can be clearly seen that there are obvious differences in the trend of change, indicating that the number of air conditioning on the temperature effect of air conditioning is more obvious.

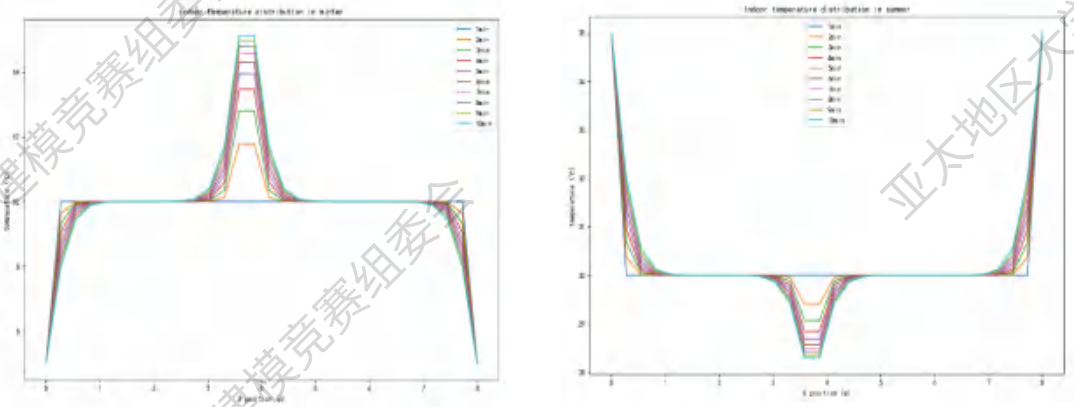


Fig. 9Change of temperature affected by air conditioning position in the middle of summer (left) and winter (right)

To sum up, the relevant factors that have a greater impact on the change of the air conditioner temperature are the wind speed, quantity and spatial location of the air conditioner.

6.1.5 Summary

This paper considers the air flow generated during the operation of the air conditioner and establishes the air circulation equation. On this basis, the optimization model of high performance air conditioning is established. On the basis of the random search algorithm, the model is optimized by using the finite difference method, and the best shape of the air conditioner is determined to be a cuboid, and the best size is 0.5m x 0.5m x 0.5m. Then, the Lagrange algorithm is used to solve the best temperature adjustment effect of air conditioning in summer and winter, and the air conditioning adjustment effect in summer is better than that in winter.

6.2 Model establishment and solution of problem 2

6.2.1 Establishment of air purification model

The essence of an air purifier [6]is to absorb tiny substances in the air, and the diffusion of tiny substances can be described by the transmission diffusion equation. The Adwex-diffusion equation considers the transmission of molecular diffusion caused by fluid motion and concentration gradient, and can more comprehensively describe the transmission and diffusion process of substances in multi-dimensional space, which can be expressed as follows:

$$\frac{\partial C}{\partial t}+u_x\frac{\partial C}{\partial x}+u_y\frac{\partial C}{\partial y}+u_z\frac{\partial C}{\partial z}=D_x\frac{\partial^2 C}{\partial x^2}+D_y\frac{\partial^2 C}{\partial y^2}+D_z\frac{\partial^2 C}{\partial z^2} \tag{25}$$

Among them, D_x, D_y and D_z are the diffusion coefficients in the 3 coordinate system, C is the substance concentration, and u_x, u_y and u_z are the air velocity in the

3 coordinate system.

For the solution of the diffusion coefficient, this question will use the mean square displacement method, which can be expressed:

$$\begin{cases} D = \dfrac{\left\langle r^2(t) \right\rangle}{6t} \\ r = \sqrt{x^2 + y^2 + z^2} \end{cases} \tag{26}$$

Where $\left\langle r^2(t) \right\rangle$ represents the mean square displacement and t is the time interval, the diffusion coefficient is estimated by fitting the slope of the leading part.

This partial differential equation establishes a three-dimensional coordinate system in a room $5m \times 8m \times 3m$ by reference to problem 1. Different positions correspond to different substance concentration $C(x, y, z)$, and all coordinates constitute a substance concentration field. In order to simplify the model, combined with energy conservation, it is transformed into a differential equation, expressed:

$$dC = CV_p dt - SQdt \tag{27}$$

Among them, S is the surface area, Q is the air volume, and V_p is the volume of the air purifier. Similar to the air conditioner, the existing air purifiers in the market are mostly cuboid, cylindrical, and spherical, which will be discussed separately here. For cuboids, $V_p = abc$, $S = 2(ab + ac + bc)$, where a , b , and c are the length, width and height of the air purifier respectively; for cylinders, $V_p = \frac{1}{4}\pi d^2 c$, $S = \frac{1}{2}\pi d^2 + \pi dc$, where d is the diameter of the cylinder, and c is the height of the cylinder; for spheres, $V_p = \frac{4}{3}\pi r^3, S = 4\pi r^2$. $SQdt$ is the minimum value of pollutant concentration per unit time.

simultaneous (15), (17), available

$$(CV_p - SQ) + u_x \frac{\partial C}{\partial x} + u_y \frac{\partial C}{\partial y} + u_z \frac{\partial C}{\partial z} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + D_z \frac{\partial^2 C}{\partial z^2} \tag{28}$$

At this time, the partial differential equation (18) links the time dimension factor with the spatial dimension, which can describe the material concentration state of each volume element in a given space, and provide a theoretical basis for the subsequent model establishment.

6.2.2 Establishment of Optimization Model for High Performance Air Purification

6.2.2.1 objective function

This topic is mainly based on the optimization of air purifiers, and the shape and size of air purifiers are designed to make them high-performance air purifiers.

According to relevant data, the high-performance air purifier is mainly reflected

[7]by the concentration of substances in the space, then the model here is based on the diffusion equation to construct the relationship between different coordinates and concentrations in the space field, and ultimately to ensure that the total concentration of substances in a given size is the smallest in the $5m\times 8m\times 3m$ space, at this time the air purifier effect is optimal. Then sum the concentrations of substances in space, expressed:

$$C=\int_V dC \tag{29}$$

The microelement form of the substance concentration is shown in (17).

6.2.2.2 constraint condition

The volume of the air purifier cannot exceed the size of a given space, which can be expressed:

$$V_p \le 120 \tag{30}$$

Among them V_p is the volume of the air purifier, and the mention of different shapes needs to be discussed separately.

6.2.2.3 establishment of optimization model

In summary, with the minimum concentration of space material as the target function, the volume of air purifier as the decision variable, the establishment of a single-goal planning model, can be expressed:

$$\min C$$
$$s.t.\begin{cases} C=\int_V dC \\ (CV_p-SQ)+u_x\dfrac{\partial C}{\partial x}+u_y\dfrac{\partial C}{\partial y}+u_z\dfrac{\partial C}{\partial z}=D_x\dfrac{\partial^2 C}{\partial x^2}+D_y\dfrac{\partial^2 C}{\partial y^2}+D_z\dfrac{\partial^2 C}{\partial z^2} \\ D=\dfrac{\left\langle r^2(t)\right\rangle}{6t} \\ V_{pi}=abc, S_i=2(ab+ac+bc), i\in \text{cuboid} \\ V_{pi}=\dfrac{1}{4}\pi d^2 c, S=\dfrac{1}{2}\pi d^2+\pi dc, i\in \text{cylinder} \\ V_{pi}=\dfrac{4}{3}\pi r^3, S=4\pi r^2, i\in \text{sphere} \\ V_p\le 120 \end{cases} \tag{31}$$

6.2.3 Solution of high performance air purification optimization model

To solve the above model, this question is calculated using the gradient descent algorithm[8]. The gradient descent algorithm is aimed at the lowest optimization problem, the goal is to make the objective function along the fastest path down to the minimum.

First, the cost function of the mean square error is added to the formula of the least squares, and it is linked to eliminate the effect of the number of samples on the parameter solution, then the partial derivative of the function is:

$$\frac{\delta(J(\theta))}{\delta(\theta_j)} = -\frac{1}{n}\sum_{i=1}^{n}[(h_\theta(x_i)-y_i)x_{ij}] \tag{32}$$

Next, using the gradient descent rule described earlier, we obtain:

$$\theta_j = \theta_{(j-1)} - a\frac{1}{n}\sum_{i=1}^{n}[h_\theta(x_i)-y_i)x_{i0}] \tag{33}$$

where n denotes the number of samples, a denotes the step size of the learning rate, and h(x) denotes the cost function.

By using the gradient descent algorithm to solve the high-performance air purification optimization model, it is determined that the best shape of the air purifier is cuboid, the best parameters are length = 0.82 meters, width = 0.92 meters, height = 1.18 meters, at this time, the purification effect of the air purifier is visualized, as shown in the following figure:

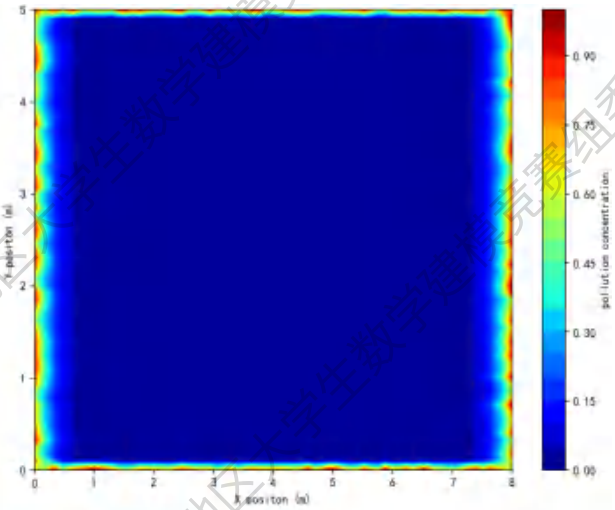


Fig. 10Visualization of purification effect

As can be seen from the above figure, the air purifier has a high degree of purification, and most of the room space is in a blue state, indicating that the cleanliness is high. Therefore, the above figure is the state analysis diagram with the highest degree of purification in this topic.

6.2.4 inspection and analysis

This paper uses robustness analysis to test the model. The following figure shows the distribution state of pollutant concentration. It can be seen from the figure that the pollutant concentration shows a random distribution state at this time. In this regard, this paper further analyzes the purification effect of air purifiers by changing the distribution of pollutants and changing the distribution from random distribution to normal distribution.

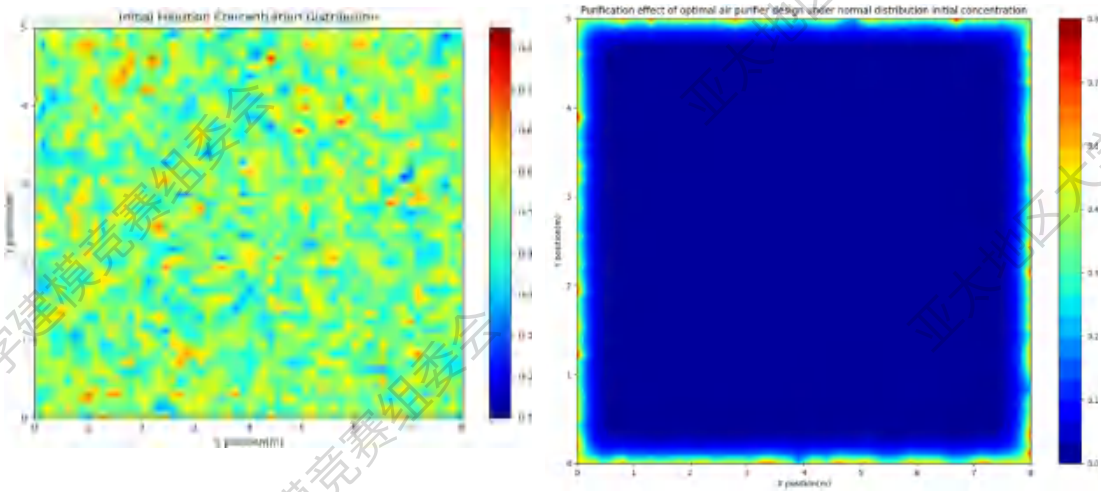


Figure 11 Distribution of initial pollutant concentration(left) Purification effect of normal distribution initial concentration of the best air purifier design(right)

The above figure shows the purification effect of the air purifier after changing the distribution state. It can be seen from the figure that the air purification effect is the same as the initial state at this time, and the purification effect is better.

6.2.5 Summary

Considering that the essence of air purifiers is to absorb tiny substances in the air, this question selects the Adwex-diffusion equation to comprehensively describe the transmission and diffusion process of substances in multidimensional space. On this basis, this topic takes the best purification effect as the target function and establishes the target planning model. Then, using the gradient descent algorithm, the above-mentioned objective function is solved, and the best shape of the air purifier in this question is the cuboid, the best size is the best parameter for the length of 0.82 meters, the width of 0.92 meters, the height of 1.18 meters, at this time the air purification effect is the best. Finally, the model is tested by robustness analysis, and the concentration distribution of pollutants is changed, and the purification effect is still very good, which shows that the accuracy of the model is high.

6.3 Model establishment and solution of problem 3

6.3.1 Establishment of Water Vapor Diffusion Model

The effect of air humidifier [9]is mainly measured by humidity. The shape and size of air purification affect the release mode and distribution of water vapor. The water vapor diffusion equation is introduced here. Fick's law is introduced here to establish the water vapor diffusion equation to describe the water vapor concentration.

First of all, based on Fick's first law to establish the water vapor concentration vector field, the first law draws the conclusion that the diffusion material flow per unit time through the unit cross-sectional area perpendicular to the diffusion direction (called the diffusion flux j) is proportional to the concentration gradient at the cross-section, which can be expressed:

$$j = -D\frac{dC_w}{dx} \tag{34}$$

The D is the diffusion coefficient, the C_w is the volume concentration of water vapor, the $\frac{dC_w}{dx}$ represents the concentration gradient, and the negative sign indicates that the diffusion direction is opposite to the concentration gradient, that is, the diffusion element flows from high concentration to low concentration. It is then extended to a given three-dimensional space, which forms a humidity vector field, which can be expressed:

$$j = -D\nabla C_w \tag{35}$$

where j is the diffusion flux and ∇ is the gradient operator.

Considering non-steady diffusion, Fick's second law is introduced. At distance x, the rate of change of concentration with time is equal to the negative value of the rate of change of diffusion flux with distance, which can be expressed:

$$\frac{\partial C_w}{\partial t} = \frac{\partial}{\partial x}(D\frac{\partial C_w}{\partial x}) \tag{36}$$

where the diffusion coefficient D is constant and can be expressed in three-dimensional space after simplification as follows:

$$\frac{\partial C_w}{\partial t} = D\nabla^2 C_w \tag{37}$$

Among them ∇^2, the Laplace operator, 2 the estimation of the diffusion coefficient is consistent with the solution of the problem 2, using the mean square displacement method to estimate, (17) and (25) joint, then the water vapor concentration equation can be obtained, expressed:

$$SQ = D_x\frac{\partial^2 C_w}{\partial x^2} + D_y\frac{\partial^2 C_w}{\partial y^2} + D_z\frac{\partial^2 C_w}{\partial z^2} \tag{38}$$

Where S represents the surface area of the humidifier, Q represents the discharge amount of the humidifier, and SQ is the humidification humidity reduction value. Based on the market survey, still in the market mainstream air humidifier as a reference, respectively, the cuboid, cylinder, sphere shape to discuss. For the surface area of a cuboid, $S = 2(ab + ac + bc)$, where a, b, and c each represent the length, width and height of the cuboid; for a cylinder, $S = \frac{1}{2}\pi d^2 + \pi dc$, where d is the diameter of the cylinder and c is the height of the cylinder; for a sphere, $S = 4\pi \mathrm{r}^2$, where r is the radius of the sphere.

6.3.2 Establishment of Optimization Model for High Performance Humidifier

6.3.2.1 objective function

Here, only from the size and size of the air moisture device to adjust, the highest concentration of water vapor in the air, that is, the performance of the air purifier is better, then the room who vapor concentration can be expressed;

$$C_w = \int_V dC_w \tag{39}$$

6.3.2.2 constraint condition

The constraint here is only that its volume cannot exceed the volume of the room, which can be expressed:

$$V_h \le 120 \tag{40}$$

Among them V_h, the volume of the air humidifier, different collision air humidifier needs to be discussed separately.

6.3.2.3 establishment of optimization model

The high-performance air humidifier takes the maximum water vapor concentration in the field as the target function, takes the humidifier surface area S as the decision variable, and establishes a single-target planning model, which can be expressed:

$$\min(-C_w)$$
$$s.t.\begin{cases} C_w = \int_V dC_w \\ SQ = D_x \dfrac{\partial^2 C_w}{\partial x^2} + D_y \dfrac{\partial^2 C_w}{\partial y^2} + D_z \dfrac{\partial^2 C_w}{\partial z^2} \\ D = \dfrac{\left\langle r^2(t) \right\rangle}{6t} \\ V_{hi} = abc, S_i = 2(ab + ac + bc), i \in \text{cuboid} \\ V_{hi} = \dfrac{1}{4}\pi d^2 c, S = \dfrac{1}{2}\pi d^2 + \pi dc, i \in \text{cylinder} \\ V_{hi} = \dfrac{4}{3}\pi r^3, S = 4\pi r^2, i \in \text{sphere} \\ V_h \le 120 \end{cases} \tag{41}$$

6.3.3 Solution of Optimization Model of High Performance Humidifier

6.3.3.1 BFGS Optimization Algorithm of Quasi-Newton Method

For the above-mentioned single-target optimization model, the BFGS

optimization algorithm of Newton method is used to solve the model. The working principle of the BFGS optimization algorithm of the Newton method is to calculate the optimal result by iterating the algorithm, but the approximate matrix is always positive, so the algorithm is always searching in the direction of optimization, with global convergence and super-linear convergence speed. The BFGS algorithm uses more quasi-Newton methods, so it is called BFGS correction.

Correction on the basis of the quasi-Newton method:

Let $E_k = au_k v_k^T + \beta u_k v_k^T$, where u $_k$,v $_k$ are all vectors of n * 1.

It can be resolved:

$$B_{k+1} s_k = y_k \tag{42}$$

Substituting the $E_k = au_k v_k^T + \beta u_k v_k^T$ into the above formula:

$$B_k + au_k v_k^T + \beta u_k v_k^T \quad s_k = y_k \tag{43}$$

The final BFGS correction formula is:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{44}$$

Through the of BFGS optimization algorithm [10]using quasi-Newton method, the single-objective planning model established for the above-mentioned high-performance air humidifier is solved by taking the maximum water vapor concentration in the field strength as the objective function and the humidifier surface area S as the decision variable, and the optimal shape of the air humidifier is rectangular, the size parameter length is 1.26m, the width is 0.2m, and the height is 0.5m. At this time, the humidification time can reach 20 hours and the humidification effect is maximized. The humidification status of the air humidifier is shown in the figure below. It can be seen from the figure that the closer the distance from the air purifier in the room, the better the air purification effect. Overall, the purification effect of this topic is better overall.

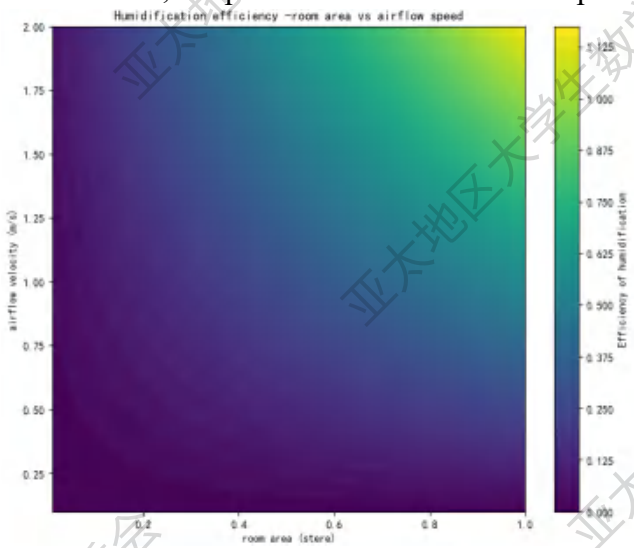


Figure 12Air humidifier effect drawing

6.3.4 inspection and analysis

In order to further verify the accuracy of the model, this paper uses sensitivity analysis to test this question. In this paper, the parameters such as the shape, size, airflow speed and humidification time of the air humidifier are fixed, and the corresponding maximum humidification efficiency is calculated by changing the value of the maximum power.

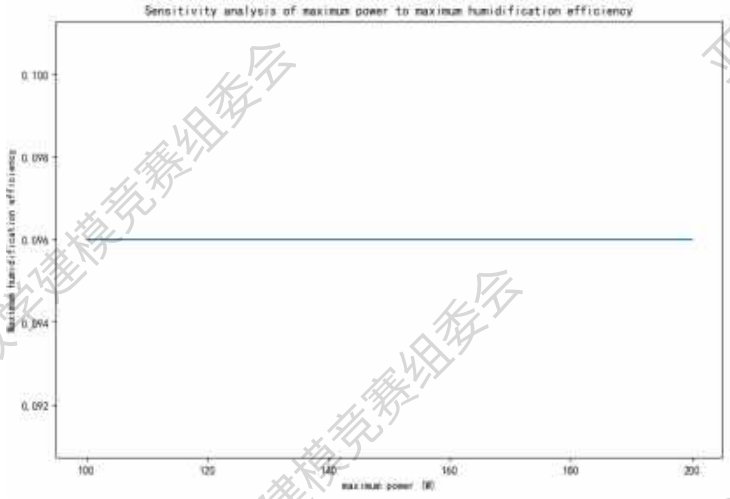


Figure 13Sensitivity

With the maximum power of 100W as the initial value, with 1 as the step, to the end of 300W, calculate the best humidification efficiency, as shown in the figure, with the increase of power, the humidification rate is almost no change, indicating that the model is more stable.

6.3.5 Summary

Since the effect of air humidifier is mainly measured by humidity, Fick's law is introduced here to establish the water vapor diffusion equation to describe water vapor. Then, the single-objective planning model is established with the maximization of steam concentration as the target function and the volume condition as the constraint. In this paper, the BFGS optimization algorithm of Newton method is used to solve the above single-target planning model, and the best shape of the humidifier is rectangular, the size parameter length is 1.26m, the width is 0.2m, and the height is 0.5m. Finally, the model is tested by sensitivity analysis, and the results show that with the increase of power, the humidification rate has no change, indicating that the model has a high degree of stability.

6.4 Model establishment and solution of problem 4

6.4.1 The establishment of the Pareto optimization model.

It is necessary to establish a high-performance three-in-one equipment based on the problem 1. 2. three, and a new equipment with 3 functions of air conditioner, air purifier, air humidifier, and meet the optimal solution of the 3 optimization model at the same time. Considering that non-dominant individuals can be excluded in the Pareto optimization process to obtain a more convergent solution, that is, the optimal solution, the Pareto optimization model is introduced here.

Firstly, the above model is 1 to the model 3 objective function for classification, which is divided into positive function indicators and negative function indicators. For air conditioners, the higher the utility value, the better the performance, which is a positive function index, then it is $utility^{+}$. For air purifiers, the lower the sum of substance content and concentration in the room, the better the performance, that is, the negative function index, which is C^{-}. For air humidifier, the greater the sum of water vapor content and concentration in the room, the better its performance, that is, $C_{\mathrm{w}}{}^{+}$.

Here to maximize the performance of the three-and-one sense regulator as the goal, to air conditioning volume V_a, air purifier volume V_p, air humidifier volume V_h, wind speed u, unit air inlet air outlet air volume V_t, the number of air outlets for the n as the decision variable, multi-objective optimization model can be written:

$$\min F=[-utility^{+},C^{-},-C_{w}{}^{+}]$$

$$s.t.\begin{cases}
V_a \le 0.1\\
P \le 1800\\
\mathrm{u} \le 8\\
nV_{\mathrm{t}} \le \dfrac{Q_{\mathrm{inf}}}{c\rho\Delta T} \le 0.6\\
V_p \le 120\\
V_h \le 120\\
C^{-}=\int_V dC^{-}\\
C_{w}{}^{+}=\int_V dC_{w}{}^{+}\\
SQ=D_x\dfrac{\partial^2 C_{w}{}^{+}}{\partial x^2}+D_y\dfrac{\partial^2 C_{w}{}^{+}}{\partial y^2}+D_z\dfrac{\partial^2 C_{w}{}^{+}}{\partial z^2}\\
(C^{-}V_p-SQ)+u_x\dfrac{\partial C^{-}}{\partial x}+u_y\dfrac{\partial C^{-}}{\partial y}+u_z\dfrac{\partial C^{-}}{\partial z}=D_x\dfrac{\partial^2 C^{-}}{\partial x^2}+D_y\dfrac{\partial^2 C^{-}}{\partial y^2}+D_z\dfrac{\partial^2 C^{-}}{\partial z^2}\\
D=\dfrac{\left\langle r^2(t)\right\rangle}{6t}\\
V_{pi,\ hi}=abc,S_i=2(ab+ac+bc),i\in \text{cuboid}\\
V_{pi,\ hi}=\dfrac{1}{4}\pi d^2 c,S=\dfrac{1}{2}\pi d^2+\pi dc,i\in \text{cylinder}\\
V_{pi,\ hi}=\dfrac{4}{3}\pi r^3,S=4\pi r^2,i\in \text{sphere}\\
\eta=\dfrac{Q_s/\Delta t}{P}\\
utility=\alpha_1\eta-\alpha_2 Tempreture_uniformity-\alpha_3 noise\\
noise=\delta u+\varepsilon\\
Tempreture_uniformity=\dfrac{1}{V}\int\left|T(x,y,z)-T_{target}\right|^2 dV\\
\rho c\dfrac{\partial T}{\partial t}+\rho\bullet c\bullet(\vec{u}\bullet\nabla T)=\kappa\nabla^2 T+Q_s\\
Q_{\mathrm{inf}}+Q_s=Q_T
\end{cases} \tag{45}$$

6.4.2 solution of the model

In this paper, the SLSQP algorithm [11]is used to solve the above multi-objective planning model. Sequential least squares is the 1 numerical method to solve unconstrained nonlinear optimization problems[12]. Based on the idea of serialization, it decomposes the problem into a series of sub-problems to solve. In each subproblem, it is solved by minimizing a quadratic function.

By using machine learning-SLSQP algorithm to solve the above multi-objective planning model, the best three-in-one equipment is determined. Its best size is air conditioner: width = 1.24m, depth = 1.24m, and height range of air conditioner: 1.5~2.15m; Air purifier: width = 1.24m, depth = 1.24m, height range of humidifier: 0.85~1.5m; Humidifier: width = 1.24m, depth = 1.24m, height range of air purifier: 0~0.85m; According to this, the best appearance shape of the equipment can be drawn in this paper, as shown in the following figure:

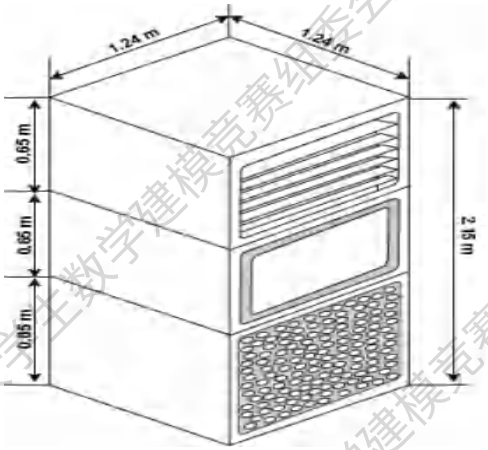


Fig. 14Best Design

6.4.3 Summary

Based on the single-goal planning model established by the problem 1. 2. problem 3, this question adds energy efficiency, human comfort, purification effect and humidification effect maximization as the target function, and establishes a multi-objective planning model. Then, the machine learning-SLSQP algorithm is used to solve the multi-objective planning model, and the optimal parameters for the integration of 3 and equipment are obtained as follows: width = 1.24m, depth = 1.24m, and height range of air conditioner: 1.5~2.15m; Air purifier: width = 1.24m, depth = 1.24m, height range of humidifier: 0.85~1.5m; Humidifier: width = 1.24m, depth = 1.24m, the height range of air purifier: 0~0.85m, and draw the best appearance design.

7 evaluation and generalization of the model

7.1 evaluation of the model

7.1.1 Advantages of the model

- In designing and optimizing air conditioners, air purifiers, and humidifiers, this paper adopts a multi-objective planning model encompassing temperature uniformity,

energy efficiency, and noise control. This multi-objective design meets diverse usage scenarios and enhances overall equipment performance.

- Mathematical models such as Navier-Stokes and Advection-Diffusion equations are established to detail airflow and material diffusion processes, based on physical laws, ensuring high accuracy and reliability in simulating device operations in real-world environments.
- Various numerical methods, including random search, finite difference, gradient descent, and BFGS quasi-Newton optimization, are utilized to solve complex problems, leveraging their strong computational power and convergence for rapid optimal solution discovery.
- Robustness and sensitivity analyses validate model stability and accuracy under different conditions, exemplified by testing air purifier efficacy through varying pollutant concentrations and humidifier performance stability through altering maximum power values.

7.1.2 Disadvantages of the model

- The model assumes room air as incompressible ideal gas, neglecting practical factors like wall thermal capacity, which simplifies calculations but may lead to discrepancies with reality. It focuses on parameters like temperature, humidity, and air quality, with limited consideration for other environmental factors (e.g., lighting, wind speed variations) that could significantly impact device performance in real-world settings. To enhance model accuracy and practicability, more practical factors such as wall thermal capacity and indoor lighting changes can be incorporated.
- Device size and shape designs presented in the article are based primarily on theoretical calculations and simulations, potentially limited by practical constraints like installation space and power supply. Furthermore, device durability and long-term operational stability require further validation. Advanced optimization algorithms (e.g., Particle Swarm Optimization) can be employed to address conflicts in multi-objective optimization, seeking better trade-off solutions. Additionally, more constraints can be introduced to ensure design feasibility in real-world environments.

7.2 generalization of the model

- The model in this paper can be applied to the smart home system to realize the automatic adjustment and management of indoor air environment by integrating the functional modules such as air conditioner, air purifier and humidifier. This will improve the user's living comfort and health level.
- In the industrial environment, similar multi-objective optimization design can also be used for the design and optimization of the workshop air conditioning system to improve the comfort and safety of the working environment. For example, in high temperature, high humidity or polluted environments, the design of air conditioners, dehumidifiers and air purification equipment is optimized to ensure that the air quality in the work area meets the standard requirements.

References

[1] Wang C ,Zhao Z ,Wang B , et al.Optimal fresh-air utilization strategy based on isocost Line: Adaptability analysis for constant temperature and humidity air-conditioning system[J]. Sustainable Energy Technologies and Assessments,2024,72104064-104064.

[2] Mustufa A ,Anwar M ,Khan I T. Energy optimization of an institutional building to reduce the energy and air conditioning demand[J]. Journal of Building Engineering,2024,98111309-111309.

[3] Zhang Y ,Zhao D ,Wu L , et al.Multi-objective optimization under the mixed utilization strategy of heat pump and electric drive waste heat of electric vehicles[J]. Applied Thermal Engineering,2025,259124900-124900.

[4] Kim M ,Kwon S. Application of supervised and unsupervised learning for enhancing energy efficiency and thermal comfort in air conditioning scheduling under uncertain and dynamic environments[J]. Energy & Buildings,2024,325115028-115028.

[5] Chen D ,Liu M ,Guo W , et al.Energy-efficient operation of portable air cleaners based on real-time prediction of non-uniform concentrations of indoor air pollutants in open offices[J]. Building and Environment,2024,256111478 -.

[6] Yao C ,Lee S ,이승재 , et al.A Numerical Study for Optimization of Air Purifier Placement and Operation in School Classrooms[J]. The 2024,

[7] Liu J ,Zhao L ,Hu K , et al.Optimization study of a closed-cycle low-temperature evaporation system based on mathematical model and CFD.[J]. Environmental research,2024,251(P2):118762-118762.

[8] Hui Y ,Chengcheng W ,Yelong Z , et al.Experimental investigation of a thermochemical energy storage system based on MgSO4-silica gel for building heating: adsorption/desorption performance testing and system optimization[J]. Energy Conversion and Management,2024,301118000 -.

[9] Sankhyan S ,Clements N ,Heckman A , et al.Optimization of a Do-It-Yourself Air Cleaner Design to Reduce Residential Air Pollution Exposure for a Community Experiencing Environmental Injustices[J]. Atmosphere,2023,14(12):

[10] Hooseung N ,Haneul C ,Haseong K , et al.Optimizing indoor air quality and noise levels in old school classrooms with air purifiers and HRV: A CONTAM simulation study[J]. Journal of Building Engineering,2023,73

[11] Yongchao L ,Hao C ,Chihua L , et al.Multi-objective optimization design of a planar membrane humidifier based on NSGA-II and entropy weight TOPSIS[J]. Journal of Thermal Analysis and Calorimetry,2023,148(14):7147-7161.

[12] Amit B ,Balaram K. Exergoeconomic assessment and optimization of a double effect absorption chiller integrated with a humidification-dehumidification desalination system[J]. Energy Conversion and Management,2021,247

Team Number：	apmcm24200658
Problem Chosen：	C

2024 APMCM summary sheet

This study offers a comprehensive examination of future trends in the pet industry both in China and globally, utilizing various mathematical modeling techniques to forecast market developments. The research employs tools such as multiple linear regression, time series analysis, and game theory to investigate changes in pet populations, pet food consumption dynamics, and the effects of policy shifts on the market.

In response to question 1, we analyzed data from China's pet industry over the last five years, focusing on cat and dog population changes. Initially, time series analysis was employed to identify overall pet population trends, followed by multiple linear regression to evaluate how macroeconomic factors like GDP per capita, consumer spending on pets, and urbanization influence these populations. Our predictions indicate continued growth in cat numbers, while dog population growth may decelerate, signaling a shift in market demand.

For question 2, we analyzed the US and European markets, utilizing panel data models to assess the relationship between pet ownership and pet food market size across countries. By creating a vector autoregressive (VAR) model, we examined the dynamic interplay between market size and pet numbers, predicting a significant expansion of the global pet food market due to rising pet ownership.

In addressing question 3, we analyzed production and export data from China's pet food industry. Using the Holt-Winters time series model, we forecasted the industry's gross production value and applied multiple regression analysis to understand how various economic factors influence export trends. Our model also assesses the effects of global economic changes on China's pet food exports and anticipates future recovery and growth in export markets.

Regarding question 4, we explored the implications of new external economic policies, particularly tariff adjustments, on China's pet food sector. We developed static and dynamic game-theoretic models to simulate strategic interactions with major trading partners under the new policy environment. The analysis reveals that policies from key European and American countries significantly affect China's pet food exports, leading us to propose optimal strategies for enhancing the international competitiveness of China's pet food industry.

Finally, we evaluated the strengths and limitations of our model, providing scientific insights for the sustainable development of the pet industry, particularly in light of uncertainties in international trade policy, and offering guidance for strategic adjustments in China's pet food sector.

Keywords: time series, static and dynamic game theory, forecasting, Nash equilibrium analysis

Content

1. Introduction 1
1.1 Background 1
1.2 Quesettion Restatement 1
1.2.1 Question 1 1
1.2.2 Question 2 1
1.2.3 Question 3 2
1.2.4 Question 4 2
2. Problem analysis 2
2.1 Analysis of question one 2
2.2 Analysis of question two 2
2.3 Analysis of question three 3
2.4 Analysis of question four 3
3. Model assumptions 3
4. Description of the symbol 4
5. Problem 1: Establishment and solution of the model 4
5.1 Data analysis 4
5.2 Analysis of the development factors of China's pet industry 6
5.2.1 Data Colletion 6
5.2.2 Analysis of the controlling factors 6
5.3 Forecast for the development of China's pet industry 9
5.3.1 Time series forecasting 9
5.3.2 Randomforest regression prediction 10
5.3.3 Hybrid predictive models 11
6. Problem 2: Establishment and solution of the model 12
6.1 Data analysis 12
6.2 Forecast of pet food demand in other countries 15
6.2.1 Data Collection 15
6.2.2 Feature construction, selection, and correlation analysis 15
6.2.3 Model construction and solving 17
7. Problem 3: Establishment and solution of the model 18
7.1 Analysis of the development of China's pet food industry 18
7.2 Forecast for the future development of China's pet food industry 19
7.2.1 Data Collection 19

7.2.2 Data processing 20
7.2.3 Model building and solving 21
8. Problem 4: Establishment and solution of the model 23
8.1 Assumptions and variable definitions 24
8.2 Static game model 24
8.2.1 Gain funtion definition 24
8.2.2 Policy space 24
8.2.3 Earning matrix 24
8.2.4 Dynamic game model 25
8.3 Solving process and results 25
9. Model evaluation 27
9.1 Advantages of the Model 27
9.2 Disadvantages of the Model 28
10. References 28
11. Appendix 29

I、 Introduction

1.1 Background

Since the early 21st century, China's pet industry has undergone significant growth driven by the gradual liberalization of pet-related policies. Pets have evolved from being simple household aids to vital emotional companions. Urbanization and demographic changes, such as an aging population and declining birth rates, have further propelled the industry's expansion, resulting in flourishing ancillary sectors. Increased consumer expenditure on pets has transformed purchasing habits, escalating the demand for pet food, supplies, veterinary care, and services.

Emerging in the 1990s, China's pet industry coincided with rising disposable incomes following economic reforms. The relaxation of governmental restrictions facilitated the entry of international brands, such as Royal Bear and Little Star, marking the sector’s initial growth phase. Although the rate of dog ownership has plateaued, other pet categories remain robust, and independent brands are capturing market share. The emergence of innovative retail models that blend online and offline sales is influencing the market, alongside advancements in scientific pet care and intelligent management solutions.

Globally, the pet industries in Europe and the United States continue to expand rapidly, holding a substantial portion of the global market. Nonetheless, China's pet sector encounters challenges from evolving domestic and international policies, including post-Trump tariff reforms, as well as fluctuating market demands and economic conditions.

This study aims to assess the development of China's pet industry over the past five years, taking into account global market dynamics and potential tariff implications. It seeks to establish a scientific framework for forecasting future trends and to provide strategic recommendations for sustainable growth in both China's and the global pet industry.

1.2 Question Restatement

Leveraging the provided datasets and supplementary information gathered by our team, we will perform a comprehensive analysis of the pet industry’s development trends and market demands. In conjunction with the prevailing economic context, we will propose strategic recommendations for the advancement of the pet industry in China.

1.2.1Question 1:

This paper analyzes the development of China's pet industry in the past five years, especially the changes in the number of pet cats and dogs, and discusses the factors affecting the development of China's pet industry. Build a mathematical model to predict the development of China's pet industry in the next three years.

1.2.2 Question 2:

In recent years, the pet industry in Europe and the United States has witnessed substantial

growth. Utilizing the data presented in Annex 2 and supplementary information collected by the team, an analysis of global pet industry trends by species is warranted. Furthermore, this study will examine the primary factors influencing the global pet industry's development and establish an appropriate mathematical model to forecast the demand for pet food over the next three years.

1.2.3Question 3:

Based on the data regarding China's pet food production and export value presented in Annex 3, along with the trends in global demand for pet food and China's developmental trajectory, please conduct an analysis of the current state of China's pet food industry and project the production and export outlook for the next three years, assuming the continuity of existing economic policies.

1.2.4Question 4:

The pet food sector in China will undoubtedly be influenced by the new economic policies implemented by European nations and the United States, including tariff regulations. Develop a suitable mathematical model, utilizing the provided data, supplementary information gathered, and the outcomes of the preceding analyses. Based on these computations, formulate a viable strategy for the sustainable advancement of China's pet food industry.

II、Problem analysis

2.1 Analysis of Question 1

The aim of the analysis for question 1 is to evaluate the evolution of China's pet industry over the last five years and to forecast trends for the upcoming three years. This involves a preliminary examination of historical pet population data through descriptive statistics and visualization techniques to illustrate annual changes in the number of pet cats and dogs. Subsequently, multiple linear regression will be employed to identify significant economic and social determinants influencing pet ownership, including per capita GDP and pet-related expenditures. To enhance forecasting precision, a hybrid approach combining time series analysis (like the ARIMA model) with principal component regression (PCR) will be utilized, leveraging both the temporal trends and the variations explained by key influencing factors for predictions regarding the pet industry’s growth over the next three years.

2.2 Analysis of Question 2

Question 2 seeks to analyze the global development trends of the pet industry and forecast future demand for pet food. This will involve a thorough assessment utilizing pet population and market data from various countries over recent years. We will employ descriptive statistical analysis and visualization methods, including trend line charts, to illustrate the trends in pet numbers and the pet food market. Furthermore, a panel data regression model will be developed to examine the relationship between economic indicators such as per capita GDP and

urbanization rates and the pet food market size. To enhance the accuracy of our predictions regarding market demand and the dynamic interactions among variables, we will also implement vector autoregressive (VAR) models for a deeper analysis.

2.3 Analysis of Question 3

In addressing Question 3, we will perform a comprehensive analysis and projection of the production and export metrics within China's pet food sector. Initially, a time series analysis will be conducted on pertinent data, incorporating seasonal decomposition and trend examination to elucidate the primary patterns in production and export figures. Following this, we will integrate external economic variables (such as shifts in international trade policies and the global economic landscape) alongside internal industry metrics (including domestic production capacity and market demand). By employing cointegration and error correction models (ECM), we will simultaneously assess the long-term equilibrium and short-term dynamics, facilitating precise forecasts for future production and export activities.

2.4 Analysis of Question 4

Question 4 will examine international trade policies, focusing on the effects of tariff modifications on China's pet food sector. We will employ a game theory framework to analyze and forecast effective strategies for the industry. Specifically, a bilateral game model will simulate the strategic interactions between China and key trading partners, such as the United States, to evaluate optimal export pricing decisions across various tariff scenarios. By determining the Nash equilibrium, we will identify the most advantageous strategy for China's pet food industry within a specified tariff regime. Furthermore, dynamic game analysis will be utilized to offer model-driven strategic recommendations for industry leaders, taking into account temporal strategy adaptations.

III、 Model assumptions

(1) Assuming that the pet market is in a state of supply and demand balance, the market price can naturally adjust the supply and demand relationship of pets, and there is no significant market imbalance;

(2) It is assumed that the macroeconomic indicators (such as GDP growth rate and household consumption level) will change steadily during the analysis period, and there will be no sudden economic crisis or extreme economic fluctuations;

(3) It is assumed that pet food production technology and efficiency remain constant and production costs and conditions are stable during the forecast period.

(4) It is assumed that all the data obtained and used are accurate and can truly reflect the actual situation of the market and industry, and there is no measurement error or missing data.

(5) Assuming that in the short term, domestic and foreign tax policies, trade policies and relevant laws have not changed, the short-term impact of external policy adjustments on the industry should be appropriately considered.

(6) It is assumed that the exchange rate involved in international trade remains stable in the short term and does not fluctuate significantly.

IV、 Description of the symbol

symbol	illustrate
β_x	Model regression coefficient
ò	Error term
M	Target variable
T_t	Time series prediction trend value
R_t	Multiple regression model predicted value
d	The highest order of polynomial
GDP	Gross Domestic Product

V、Problem 1: Establishment and solution of the model

5.1 Data analysis

First of all, a detailed analysis of the provided data on the number of pet cats and dogs is carried out, analyzing characteristics such as trends, growth rates, etc.

Table 1 Changes in the number of cats and dogs in China in different years

Year	The number of pet cats（ten thousand）	The number of pet dogs（ten thousand）
2019	4412	5503
2020	4862	5222
2021	5806	5429
2022	6536	5119
2023	6980	5175

In order to provide a clearer picture of the evolution of pet populations over time, our team has completed a histogram of the number of pet cats and dogs by year, as shown in Figure 2. Next, we will analyze the changes in the number of cats and dogs in detail.

As illustrated in Figure 2, the population of pet cats has experienced a consistent upward trajectory, increasing from 44.12 million in 2019 to 69.8 million in 2023, representing an approximate growth of 58%. In contrast, the number of pet dogs exhibited fluctuations during the same period, peaking at approximately 55.03 million in 2019 before declining between 2020 and 2022. Although the pet dog population rebounded to around 51.75 million in 2023, it remains below the 2019 levels.

The number of pet cats is showing a clear upward trend, while the number of pet dogs is showing a slight downward trend. By 2023, the number of pet cats has surpassed that of pet dogs, a phenomenon that reflects the growing number of families who prefer cats as pets.

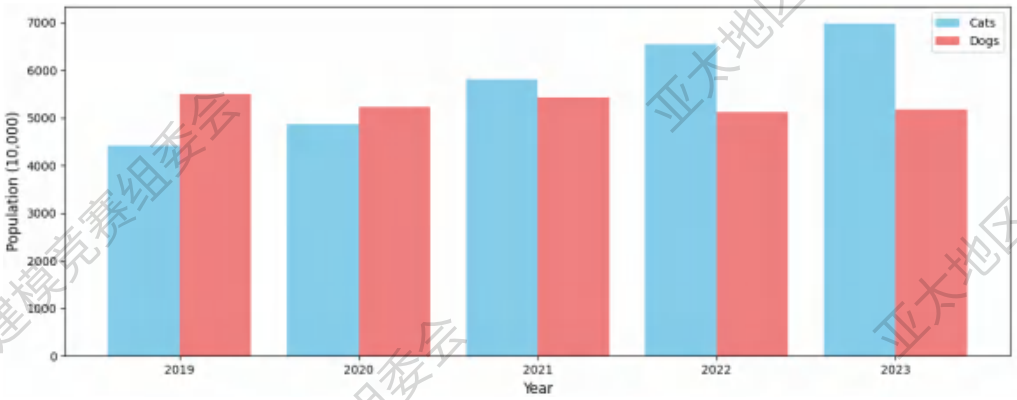


Fig.2 The number of pet cats and dogs changes over time

In order to further analyze the quantitative changes of the two, the growth rate of the two is calculated as follows:

$$Growth\ rate = \frac{Current\ year\ quantity - The\ previous\ year\ quantity}{The\ previous\ year\ quantity} \times 100\% \quad (1)$$

The growth trend of pet cats and pet dogs over time is shown in Figure 3. The growth of cats and dogs will be analyzed separately below:

For pet cats:

• In 2020, the growth rate was about 10%, marking the beginning of an accelerated increase in the number of pet cats.

• In 2021, the growth rate rose significantly to 19%, reaching its highest point, reflecting a sharp increase in demand for cats during the year.

• In 2022, growth slowed to 12.5%, but remained at a double-digit level.

• In 2023, the growth rate slowed further to 6.8%, and although the pace of growth continues to slow, market demand remains positive.

Overall, the number of pet cats is growing rapidly, but the rate of growth is gradually slowing.

For pet dogs:

• In 2020, the growth rate was about -5%, indicating a decrease in the number of pet dogs, possibly due to declining market demand or being limited by feeding policies.

• In 2021, the growth rate returned to 4%, indicating a recovery in demand for pet dogs.

• In 2022, the growth rate fell again to -6%, which may be related to external factors such as policy and the economy.

• In 2023, the growth rate rebounded to 1%, indicating that the demand for pet dogs is starting to stabilize.

Overall, the growth rate of pet dogs fluctuates greatly, showing a downward trend overall, but there are signs of stabilization in recent years.

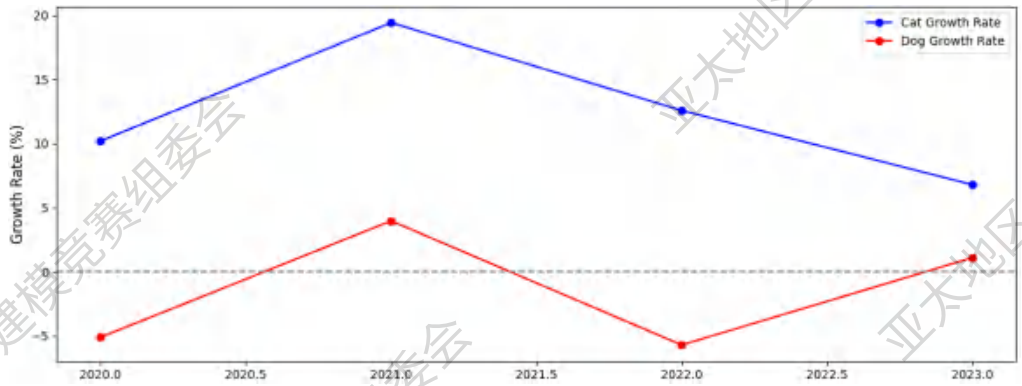


Fig. 3 Growth rate of pet cats and dogs over time

5.2 Analysis of the development factors of China's pet industry

5.2.1 Data Collection

Through literature research, it was determined that the following factors have an impact on the development of the pet industry, as shown in Figure 4.

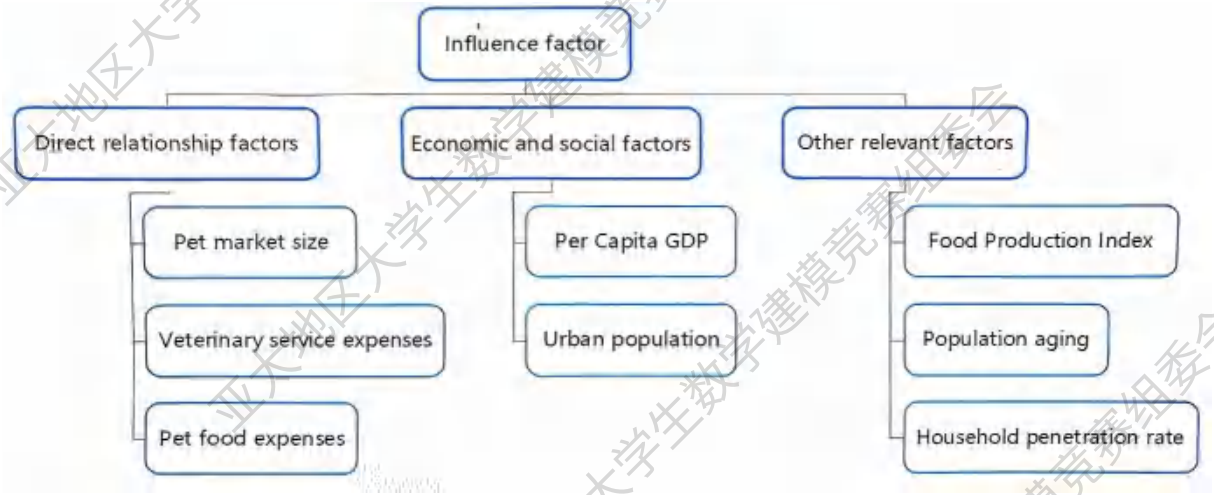


Figure 4 Indicator chart

5.2.2 Analysis of controlling factors

First, we will evaluate the strength of the linear relationship between the two independent variables, the number of pet cats and pet dogs, and the target variable (i.e., the number of cats or dogs) through correlation analysis. Next, we will use a ridge regression model to quantify the specific effect of each independent variable on the target variable, and in the process, solve possible multicollinearity problems. Finally, we will use LASSO regression to identify and screen out the independent variables that have the most significant impact on the target variables. Through this series of analyses, we can more accurately understand the relationship between pet numbers and identify the key factors that influence the target variables.

(1)Correlation analysis:

We will perform a quantitative analysis of the correlation between variables to evaluate the degree of linear correlation between them. Since the attachment data provided are numerical data, we will use the Pearson correlation coefficient method to measure the correlation between

the variables. This approach gives us an accurate measure of the linear relationship between variables.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]}{\sigma_X \sigma_Y} \tag{2}$$

The above equation is about the definition of the overall correlation coefficient, and the Greek symbol is often used ρ as the representative symbol. Then, the covariance and standard deviation of the sample are estimated to obtain the Pearson correlation coefficient, which is often represented by the English letter r

$$r = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{3}$$

r can also be estimated from the X_i standard fractional mean of the (,) sample points to obtain an equivalent expression to the above equation: Y_i

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_X}\right)\left(\frac{Y_i-\bar{Y}}{\sigma_Y}\right) \tag{4}$$

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} \tag{5}$$

(2)Ridge return

Ridge regression is an improved technique of traditional linear regression, especially suitable for processing datasets with multicollinearity. In this method, the L2 norm is introduced into the loss function as a penalty term, and the regression coefficient is constrained to reduce the complexity of the model. This not only helps to alleviate the instability of parameter estimation caused by the high correlation between dependent variables, but also enhances the prediction performance of the model on new data.

$$\min_{\beta}\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right\} \tag{6}$$

(3) LASSO regression

LASSO (Least Absolute Shrinkage and Selection Operator) regression is a regularization technique of linear regression, which constrains the model parameters by introducing the L1 norm as a penalty term in the loss function. Unlike ridge regression, LASSO is able to generate sparse solutions, which means that some of the coefficients are reduced to zero, thus achieving feature selection. This approach not only helps to improve the simplicity of the model, but also enhances its generalization ability.

The function takes the form of:

$$\min_{\beta}\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{7}$$

(4) Analyze the results

Pet cat population analysis

(1) The pet market size and pet food expenditure had the highest positive correlation with the number of cats, indicating that these factors were the main driving factors for the change of cat population. There was a weak negative correlation between total fertility and the number of cats.

(2) The regression coefficient of pet food expenditure on the number of cats was higher, indicating that it had the greatest impact on the number of cats. Spending on veterinary services also has a significant impact on cat populations.

(3) The important factors screened out include pet food expenses and veterinary service expenses.

Pet dog population analysis

(1) The size of the pet market has a significant positive correlation with the number of dogs. Pet food spending is also strongly correlated with the number of dogs.

(2) Pet food expenditure and pet market size are the main contributors to the number of dogs. Other variables, such as China's pet household penetration rate, have a weaker impact on dog populations.

(3) The size of the pet market is the most important variable to be screened out.

Note: The horizontal coordinates of the following four bar charts are The size of pet market、Expenditure for pet food、Expenditure on Veterinary Services、Chinese pet household penetration rate、total fertility rate、Index number of food production、per capital GDP、urban population, from left to right

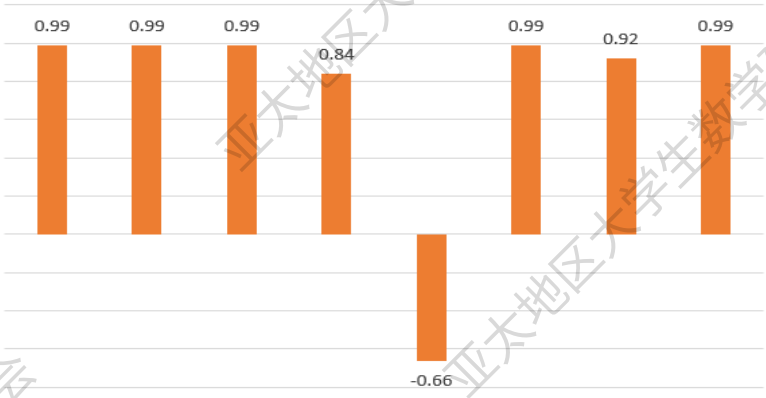


Fig.5 Correlation between each trait and the number of cats

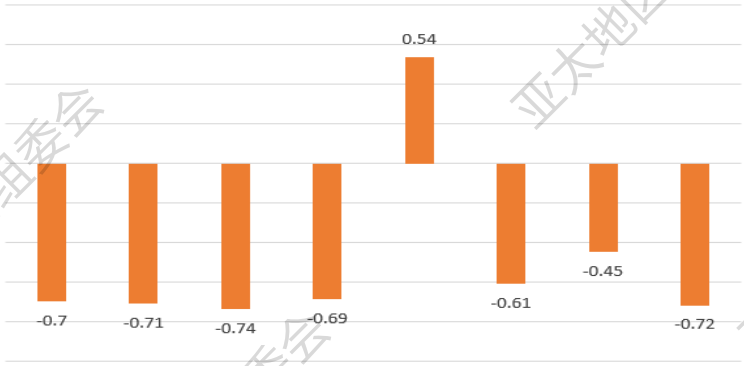


Fig.6 Correlation between each feature and the number of dogs



Fig.7 Regression coefficients of each trait and the number of cats

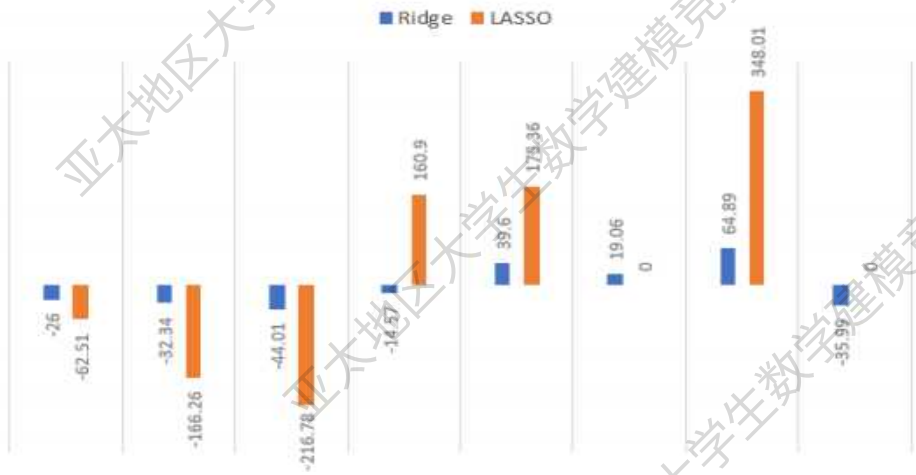


Fig.8 Regression coefficients of each trait and the number of dogs

5.3 Forecast for the development of China's pet industry

5.3.1 Time series forecasting

Polynomial regression is an extended form of linear regression that describes the nonlinear relationship between dependent and independent variables by adding a higher-order term. Polynomial transformations of years can more effectively fit nonlinear trends that may exist in a time series.

Let the year be x and the target value (e.g. number of cats or dogs) be y. The model of polynomial regression is:

$$Correlation\ index = \frac{Covariance\ （X,Y）}{\sigma x \cdot \sigma y} \tag{8}$$

β —The regression coefficient of the model

x^d —The d-power of the year, which represents the higher order term in the polynomial

ò—Error term

The predicted value can be expressed as:

$$\hat{y} = \sum_{i=0}^{d} \beta_i x^i \tag{9}$$

where x: year and d: the highest order of the polynomial.

5.3.2 Random forest regression prediction

Random forest is a popular ensemble learning algorithm that builds multiple decision tree models based on the Bagging (Bootstrap Aggregating) method. In a random forest, each decision tree is a weak learner that is independently trained to make predictions about the sample. When it is necessary to classify or regress a new sample, the random forest collects the predictions of all decision trees and decides the final output by majority voting. This approach improves the accuracy of predictions because it reduces the risk of overfitting and takes advantage of the diversity of multiple models.

Information entropy is a measure of uncertainty in the amount of information, which reflects the diversity of different outcomes that an event may have. The lower the value of information entropy, the lower the uncertainty of the information. The calculation equation is as follows:

$$H(X) = -\sum_{i=1}^{n} P(X=i)\log_2 P(X=i) \tag{10}$$

Where: the probability of $P(X=i)$ a random variable $X=i$.

Conditional information entropy refers to the uncertainty of the random variable X given the random variable Y. The conditional entropy and information gain equations are as follows:

Conditional entropy: refers to the uncertainty exhibited by the random variable X when the random variable Y is known.

$$H(X|Y=v) = -\sum_{i=1}^{n} P(X=i|Y=v)\log_2 P(X=i|Y=v) \tag{11}$$

Information gain: refers to the degree to which the uncertainty of information is reduced under certain conditions.

$$I(X,Y) = H(X) - H(X|Y) \tag{12}$$

The Gini index reflects the probability that a randomly selected sample will be misclassified in a sample set. A smaller Gini index indicates a lower probability of misclassification, i.e., a higher purity of the sample (0 when all samples in the set belong to the same category). The calculation equation is as follows:

$$Gnini(p) = \sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k=1}^{K} {p_k}^2 \tag{13}$$

Random forest (RF) is a model that integrates multiple decision trees and can be used for both classification tasks and regression analysis. A schematic diagram of a decision tree and a random forest is shown in Figure 9.

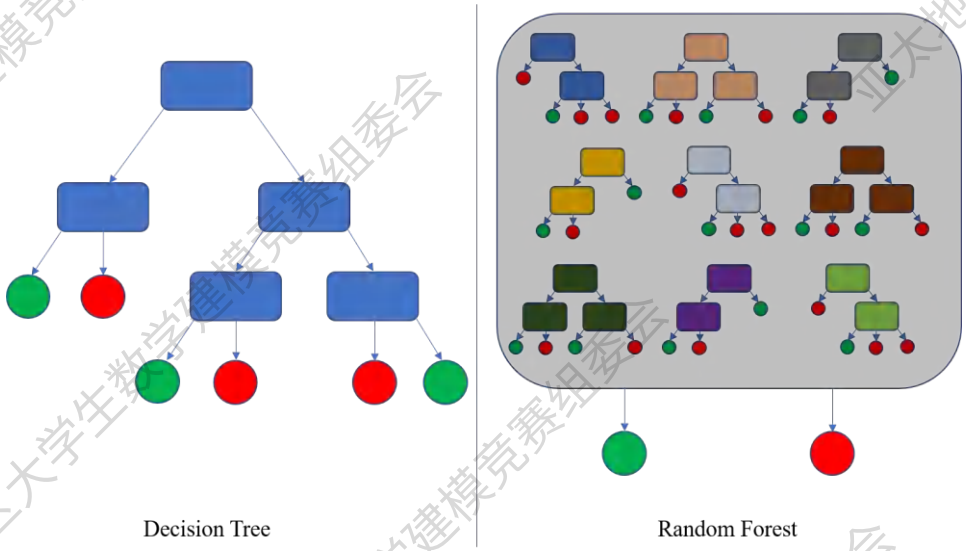


Fig.9 Schematic diagram of decision tree and random forest

For M trees, the prediction result for each tree is , then the final prediction value is:

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M} f_m(X) \tag{14}$$

5.3.3 Hybrid predictive models

The hybrid forecasting method combines time series forecasting and regression forecasting of the main controlling factors, and obtains the final prediction results in the form of weighted averages. Adjusting the weights controls the relative importance of the time series forecast and the master factor forecast in the overall forecast. $\hat{y}_{\text{hybrid}}$ The equation for calculating the mixed prediction is:

$$\hat{y}_{\text{hybrid}} = w\hat{y}_{\text{mc}} + (1-w)\hat{y}_{\text{ts}} \tag{15}$$

The prediction results are shown in Figures 10 and 11.

(1) Prediction results of the number of cats

Actual Trends: From 2019 to 2023, the number of pet cats continues to rise, reflecting the continued growth of the pet cat market size and demand.

Forecast trends: From 2024 to 2026, the number of cats will continue to show an upward trend, and although it continues to grow, the pace may slow down.

This trend is consistent with the natural characteristics of cat population growth, and the prediction results of the model show an appropriate nonlinear growth pattern. The predicted number of cats is reasonable, indicating that the model successfully identifies and reflects the growth trends presented in the historical data. The number of cats is expected to maintain a

steady growth trajectory over the next three years, suggesting that the pet cat market has broad growth potential.

(2) Prediction results of the number of dogs

Actual trends: From 2019 to 2023, the number of dogs showed significant volatility, with large increases and decreases during the period, reflecting a high level of instability in historical data. Since 2022, the number of dogs has gradually stabilized.

Predict trends

Projections show that the number of dogs is expected to remain relatively stable between 2024 and 2026, with a limited range of fluctuations. This trend may be due to the influence of volatility in historical data and the tendency of models to smooth out these fluctuations.

Time Series Trend Analysis: Overall, the model is relatively accurate in predicting cat populations, and its growth trends are consistent with historical data. Forecasts for dog populations, on the other hand, smooth out fluctuations and show a possible stable trend in the future. The hybrid model combines time series with major influencing factors, reasonably models the nonlinear trend in historical data, and comprehensively considers the impact of multiple main controlling factors (such as pet market size, food expenditure, etc.).

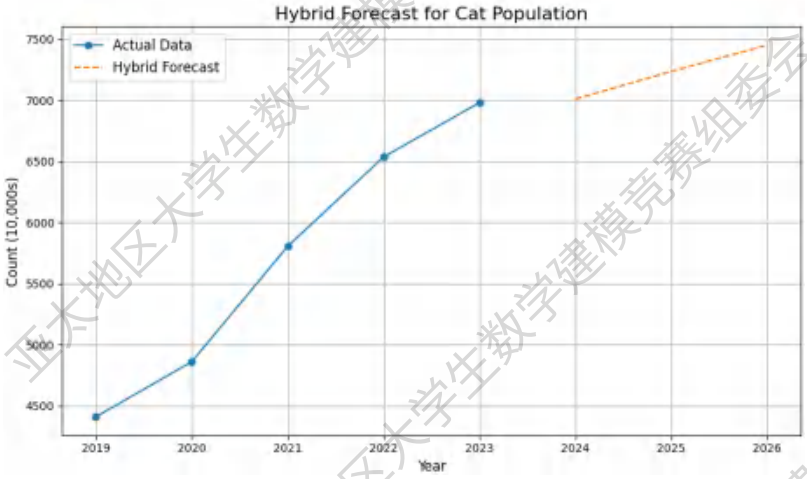


Fig.10 Prediction of cat population

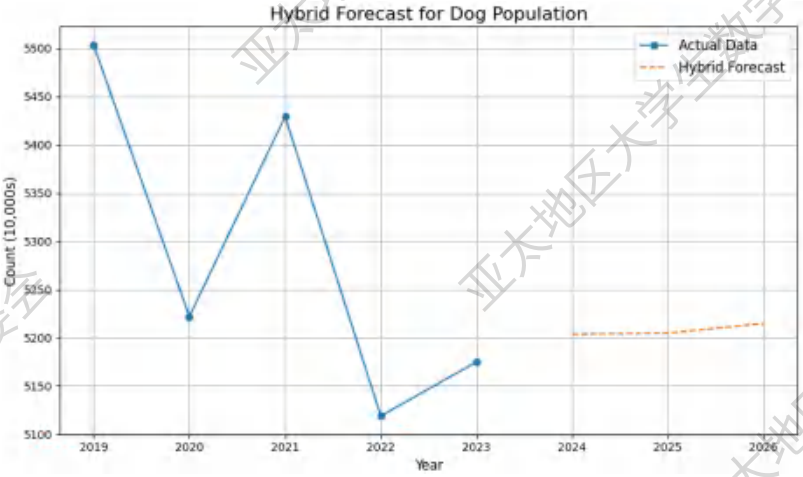


Fig.11 Prediction of dog population

VI、 Problem 2: Establishment and solution of the model

6.1 Data analysis

Start with a trend analysis. The results are shown in Figure 12~Figure 15.

(1) Analysis of the number and growth trend of cats

In terms of quantitative trends:

The number of pet cats in the United States is visibly higher than in France and Germany.

From 2019 to 2023, there is an overall downward trend, indicating that the number of pet cats in the United States is gradually decreasing.

The number of pet cats in France and Germany is relatively close, with France having slightly more.

The data shows that the number of pet cats in France is on the rise, while the number of pet cats in Germany is more stable, but there is also a slight increase.

In terms of growth rate:

The growth rate of pet cats in the United States experienced significant fluctuations, with a sharp decline in 2020 and a slight recovery in 2021, but then declining again, culminating in a modest increase in 2023.

The instability of the growth rate also reflects the instability of the pet cat market in the United States.

The rate of growth of pet cats in France continues to rise steadily, indicating that the number of people owning cats is on the rise.

The growth rate of pet cats in Germany tends to be stable and does not fluctuate much, but the growth rate is slightly slower than that of pet cats in France.

(2) Analysis of dog numbers and growth trends

In terms of quantitative trends:

The number of pet dogs in the United States is also significantly higher than in France and Germany, and the number of pet dogs has been relatively stable between 2019 and 2023, with a slight decrease.

The number of pet dogs in France, although slightly lower than in Germany, is slowly increasing.

The number of pet dogs in Germany is close to the number of pet dogs in France, but slightly more. However, compared to France, the number of pet dogs in Germany has decreased slightly after 2022.

In terms of growth rate:

The growth rate of pet dogs in the United States fluctuates significantly, but the overall trend is declining, especially after 2022, and the negative growth is even more obvious.

After 2022, the number of pet dogs in France has increased significantly, reaching a high level in 2023, indicating that the dog breeding market in France has great potential.

The growth rate of the pet dog population in Germany has stabilized, and the overall level is low, indicating that the dog breeding market in Germany has reached a relatively saturated state.

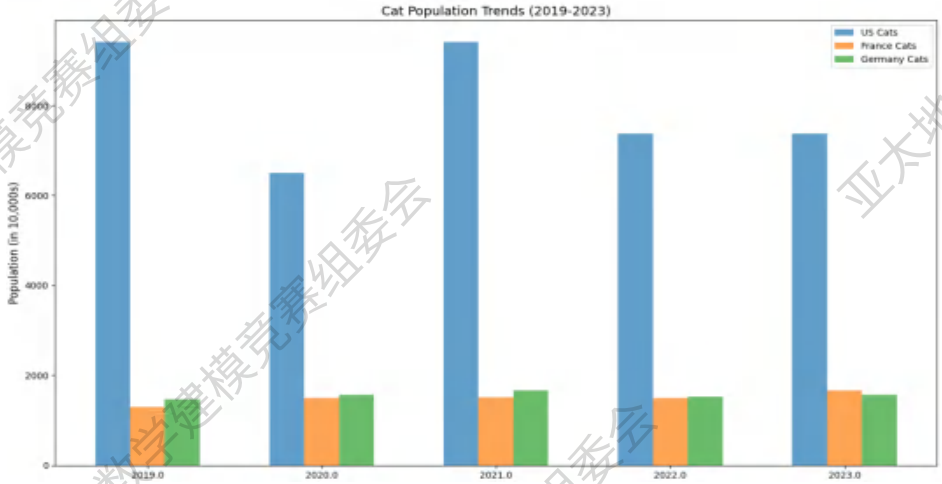


Fig. 12 Trends in cat populations by country



Fig. 13 Trends in the growth rate of cat populations by country

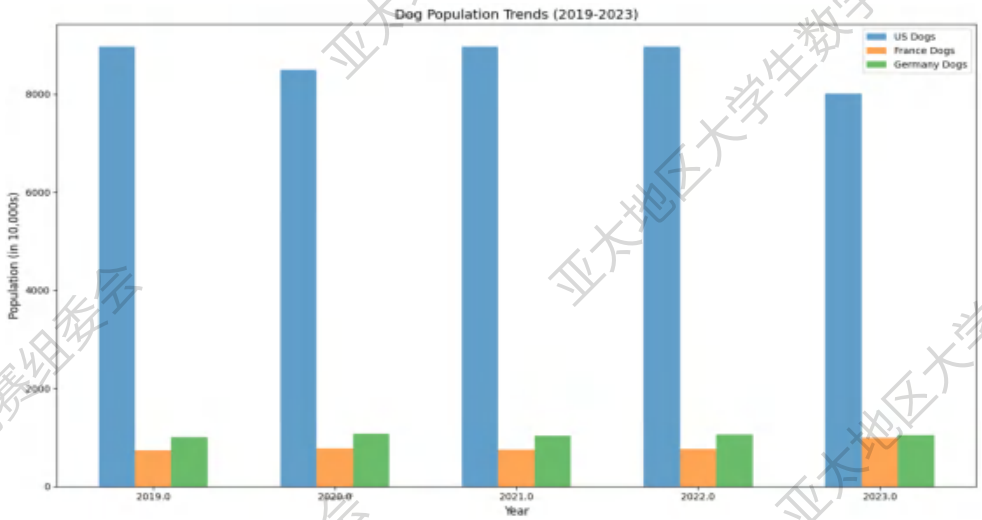


Fig. 14 Trends in dog populations by country

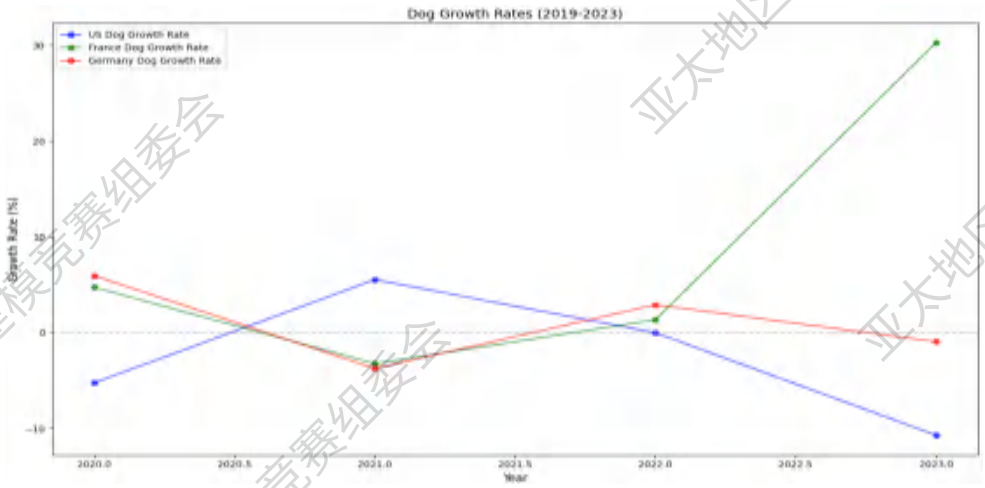


Figure 15 Trends in the growth rate of dog populations by country

6.2 Forecast of pet food demand in other countries

6.2.1 Data Collection

(1) The number of pets

Annual statistics on pet cats and dogs by country provide insights into yearly trends in pet ownership, aiding in the prediction of overall pet food demand.

(2) Relevant data of the pet food market

Pet Food Industry Market Size (USD 100 million): This indicator reflects the overall consumption level of the pet food market and is a direct measure of market demand.

Annual Pet Food Expenditure ($100 million): This data provides the total amount of pet food spent each year and helps estimate the average food consumption per pet.

Total Pet Food Exports (USD 100 million): This indicator reveals the demand for pet food in the global market and is of great significance for global market analysis.

(3) Other relevant indicators of the pet industry

Pet Market Size (USD 100 million): This indicator maps the overall trend of pet-related consumption, which is indirectly related to the demand for pet food, reflecting the breadth and depth of economic activity in the pet industry.

Veterinary Services Expenditure (US$ 100 million): reflects pet owners' willingness to pay for pet consumption, which indirectly determines their spending on pet food.

(4) Socio-economic and demographic factors

Pet Household Penetration: This indicator reflects the popularity of pet ownership, which can indirectly infer the potential size of the pet market.

GDP per capita (USD): reflects the level of a country's national economy and is highly correlated with the ability to spend on pet food.

Urban population: The level of urbanization may promote pet keeping habits, which indirectly affects the demand for pet food.

6.2.2 Feature construction, selection, and correlation analysis

(1) Feature construction

For more efficient and accurate modeling and prediction, we construct the following new

features:

Total number of pets

$$P_{\text{total}} = P_{\text{cat}} + P_{\text{dog}} \tag{16}$$

Per capita spending on pet food

$$E_{\text{avg}} = \frac{M}{P_{\text{total}}} \tag{17}$$

Quantify the effect of features on the target variable by calculating the correlation coefficient between the variables:

$$Correlation\ index = \frac{Covariance\ （\mathrm{X},\mathrm{Y}）}{\sigma x \cdot \sigma y} \tag{18}$$

(2) Feature selection

A total of 10 feature variables were obtained. In order to reduce redundancy and improve the efficiency and interpretability of the model, the following methods will be used to filter the independent variables.

- Correlation analysis of features

The correlation between features and target variables (market size) is analyzed by calculating the Pearson correlation coefficient, and a heat map is generated using Python. The results are shown in Figure 16.

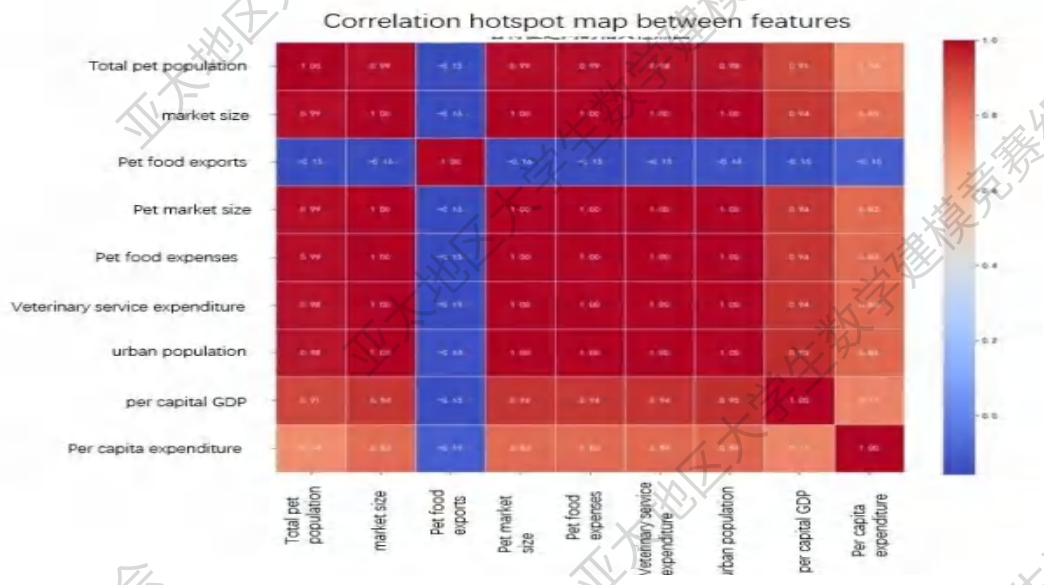


Fig.16 Correlation analysis results

The correlation coefficient between the total number of pets and the market size is extremely high at 0.96, indicating that the number of pets is a major factor driving the growth of the pet food market.

The correlation coefficient of per capita pet food expenditure is 0.78, which indicates that

it has a certain importance in measuring the consumption power of the pet food market.

Although the correlation of urban population is not high, it still deserves in-depth analysis as a basic role of consumption.

- Eliminate redundant features

  Number of cats and dogs: Comprehensively reflected by the total number of pets, eliminated.

  Characteristics such as pet food export value, pet market size, and veterinary service expenditure: the direct correlation with market size is low; There may be collinearity between the data, culling.

- The final selection feature

  Based on our team's analysis of the data, we will select the following 4 characteristics as independent variables:

argument	Select a reason
Total number of pets	Indicates the pet base of the market and reflects the potential size of the market.
Urban population	Provide a reference to the number of consumers, the larger the population size, the higher the market demand.
GDP per capita	Reflecting the level of economic development of the country, the more developed the economy, the larger the pet food market.
Pet food per capita	The higher the pet owner's willingness to spend on pets, which plays a decisive role in increasing the market size.

6.2.3 Model construction and solving

(1) Multiple linear regression model

Multiple linear regression equations:

$$M = \beta_0 + \beta_1 P_{\text{total}} + \beta_2 P_{\text{urban}} + \beta_3 GDP_{\text{percapita}} + \beta_4 E_{\text{avg}} + \text{ò} \tag{19}$$

The regression results are as follows:

United States

$$M = -120 + 0.021 P_{\text{total}} + 0.005 P_{\text{urban}} + 0.63 GDP_{\text{percapita}} + 140 E_{\text{avg}} \tag{20}$$

R2=0.985，MSE=1.3

France

$$M = -55 + 0.017 P_{\text{total}} + 0.002 P_{\text{urban}} + 0.51 GDP_{\text{percapita}} + 115 E_{\text{avg}} \tag{21}$$

R2=0.976，MSE=0.6

Germany

$$M = -62 + 0.019 P_{\text{total}} + 0.004 P_{\text{urban}} + 0.54 GDP_{\text{percapita}} + 130 E_{\text{avg}} \tag{22}$$

R2=0.981，MSE=0.8

(2) Time series model

Autoregressive (AR):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \grave{o}_t \tag{23}$$

Differential (I): Eliminates non-stationarity of the data.

Moving Average (MA):

$$X_t = \mu + \theta_1 \grave{o}_{t-1} + \cdots \tag{24}$$

Figure 17 illustrates that the projected market size in the United States is anticipated to surpass that of France and Germany significantly, underscoring its greater market potential and consumption capacity. Furthermore, each country's market size is expected to experience modest growth by 2026 relative to 2024.

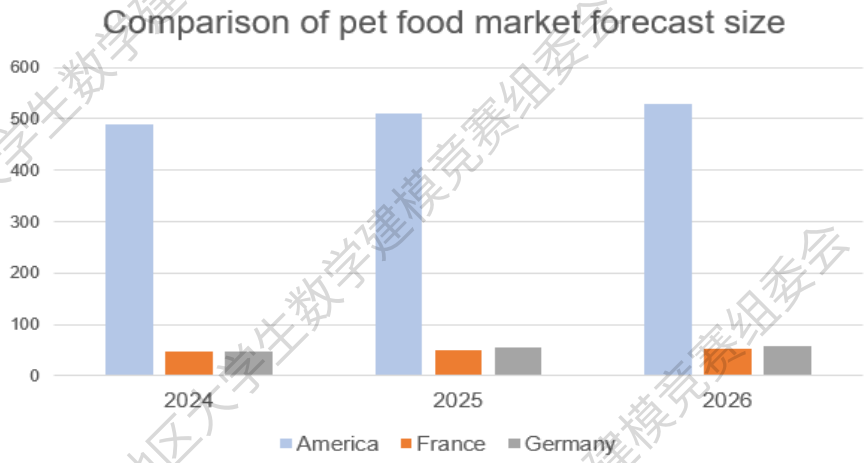


Figure 17 Market size forecast and comparison by country

VII、 Problem 3: Establishment and solution of the model

7.1 Analysis of the development of China's pet food industry

Draw line charts and analyze trends.

(1) The trend of China's total output value of pet food

Analysis of the data of the total output value of pet food (RMB):

From 2019 to 2023, the total output value of the industry increased from 44.07 billion yuan to 279.3 billion yuan.

The growth rate reached 533.5%, with an average annual compound growth rate of about 56%.

This indicates that the industry has experienced rapid expansion over the past five years, especially in 2022 and 2023, with particularly significant increases, which may be affected by the expansion of the market size or other policy factors.

(2) The trend of China's total pet food exports

An observation of the total export value of pet food (USD) shows that from 2019 to 2023, the total export value decreased from US$15.41 billion to US$980 million, and then gradually recovered to US$3.96 billion.

There have been significant fluctuations along the way, especially in 2020, which experienced significant declines.

Possible reasons include the direct impact of changes in the external economic environment (e.g., the spread of the epidemic, adjustments in international trade policies) on exports.

(3) Changes in market demand in China's pet industry

Based on the total number of pets (millions) and pet food spending:

In 2019, the total number of pets was 99.8 million, and by 2023 it has increased to 132.02 million.

During this period, spending on pet food rose from US$15.1 billion to US$20.3 billion, with an average annual growth rate of about 7.5%.

This suggests that the growth in the number of pets is directly contributing to the demand in the domestic market, while consumer demand for the quality of pet food may also increase.

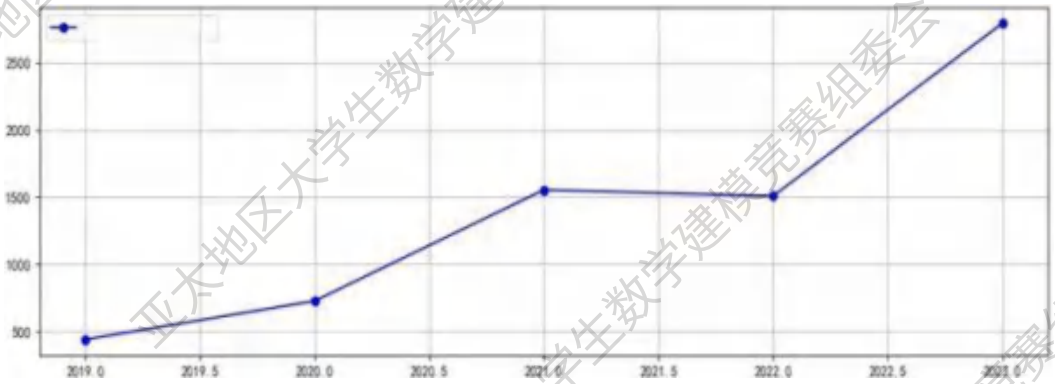


Figure 18 China's total output value of pet food (RMB, 100 million)

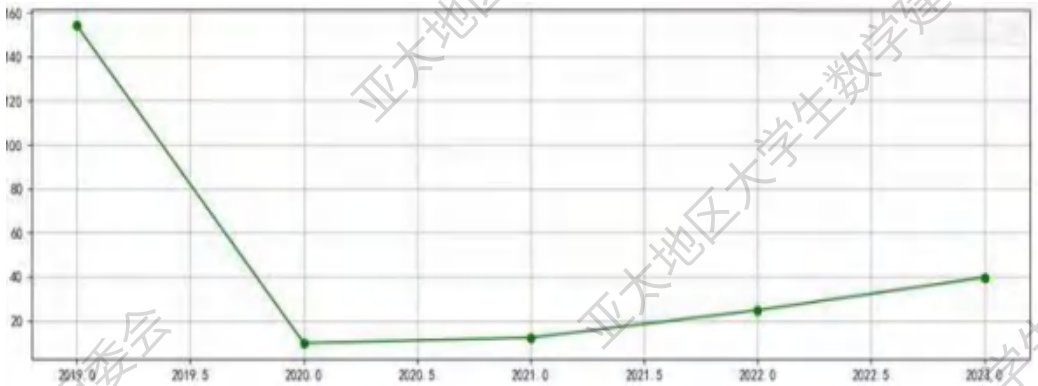


Figure 19 China's total pet food exports (US$, 100 million).

7.2 Forecast for the future development of China's pet food industry

7.2.1 Data Collection

(1) Data related to the number of pets: the number of cats and dogs (unit: 10,000): the

basic market size used to analyze the demand for pet food.

Total number of pets (unit: millions): A direct reflection of the overall pet market size.

(2) Relevant data of pet food industry:

China's Pet Food Gross Output Value (RMB): Used to analyze the trend of the industry's production capacity.

China's total pet food exports (USD): directly reflects changes in the export market.

(3) Pet market related data:

Pet Market Size (in USD billion): Helps assess changes in the overall size of the industry.

Pet Food Expenditure (in US$100 million): Changes in market demand by segment.

(4) Economic and demographic factors (indirect effects):

Urban and rural population: reflects the geographical structure of the consumer market.

Food Production Index: Assesses the potential impact of overall domestic food production capacity on the pet food industry.

GDP per capita and final consumption expenditure of households: reflect the role of economic level and consumption power in driving market demand.

7.2.2 Data processing

(1) Description of variables

Target variable: China's total pet food output value (gross domestic product, unit: RMB).

China's total pet food exports in US dollars.

Influencing factor (independent variable): number of pets (millions).

Pet market size (USD billion).

Pet food spending (US$ billion).

Urban population (persons).

GDP per capita (USD).

Food Production Index (2014-2016 = 100).

(2) Data standardization

To prevent the influence of different units on the prediction results, we normalize the independent variables:

$$X' = \frac{X - \mu}{\sigma} \tag{25}$$

(3) Feature selection

Correlation Analysis:

The Pearson correlation coefficient between the target variable and the independent variable was calculated, and the characteristics with significant correlation between the two were screened out.

Multicollinearity detection:

The variance expansion factor (VIF) was used to evaluate multicollinearity between independent variables for the purpose of removing the characteristics of VIF>10.

Figure 20 shows the relevance of each feature to China's pet food industry. It is not difficult to see that:

(1) Pet number, pet market size, pet food expenditure: There is a strong positive correlation between these three indicators and GDP (0.95, 0.96, 0.95), which indicates that the rise of these variables usually leads to an increase in GDP. These factors may directly reflect the increase in domestic market demand.

(2) Total export value: There is a significant negative correlation between these factors and total export value (-0.59, -0.51, -0.52), which indicates that with the continuous expansion of the domestic market, the increase in domestic consumption may lead to a decrease in foreign exports.

(3) Urban population and per capita GDP: These two indicators show a significant positive correlation (0.94 and 0.82, respectively), indicating that the level of economic development and the process of urbanization may be the main factors driving production growth. However, the negative impact on the total value of exports is also significant.

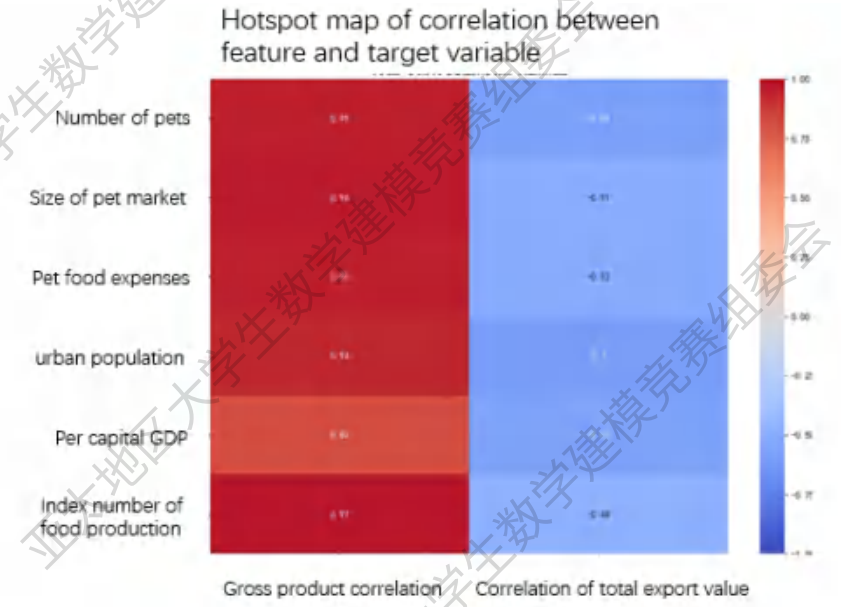


Fig.20 Heat map of the correlation between features and target variables

In the end, the characteristics chosen by our team are as follows:

Number of pets (millions). Pet market size (USD billion). Food Production Index (2014-2016 = 100).

7.2.3 Model building and solving

(1) Holt-Winters model

A Holt-Winters model in a time series model based on the target variable, and step-by-step break down the trend portions of the data:

Trending Items:

$$T_t = \alpha Y_t + (1-\alpha)(T_{t-1} + B_{t-1}) \tag{26}$$

Smoothing items

$$L_t = \beta(T_t - T_{t-1}) + (1-\beta)B_{t-1} \tag{27}$$

Predicted value

$$F\{t+m\} = Lt + mBt \tag{28}$$

(2) Multiple regression analysis

Use the regression model to fit the time series residual portions:

$$R_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon_t \quad (29)$$

(3) Hybrid model

The final forecast consists of two parts: the time series trend and the multiple regression residuals:

$$Yt = Tt + Rt \quad (30)$$

Figure 21 illustrates the time-series trend of China's real pet food production fitting to the Holt-Winters method. The solid line represents the actual annual data point, while the dotted line represents the predicted trend line. As can be seen from the figure, the model reflects the growth trend of GDP well, suggesting that it may continue to grow in the future.



Fig.21 Time series trend of GDP (RMB, 100 million)

Blue: Real GDP ； Yellow: Holt-Winters trend

Similar to GDP, Figure 22 shows the actual data on the total value of exports and their trends. Although the total value of exports has declined since 2019, it has gradually recovered and shown growth since 2021, and the model predicts that this growth will continue in the future.



Fig. 22 Time series trend of total export value (US$, 100 million)

Blue: Real GDP ； Yellow: Optimized Holt-Winters trend

Figures 23 and 24 illustrate future trends in the pet food industry, with solid lines representing historical actual data and dotted lines indicating forecasted future values. Forecasts

of GDP suggest that GDP will grow significantly over the next few years, especially after 2023, which may be closely related to a significant increase in market demand.

Forecast of total exports: Forecasts of total exports show signs of recovery and are expected to return to pre-2019 levels and continue to grow in the following years.

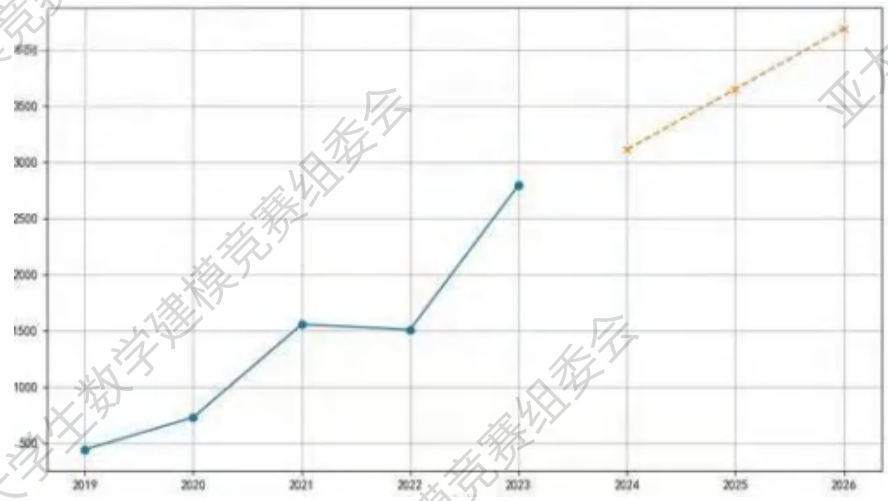


Fig. 23 Comparison of actual and forecasted GDP (RMB, 100 million yuan)
Blue (Real GDP)； Yellow (Forecasted GDP)

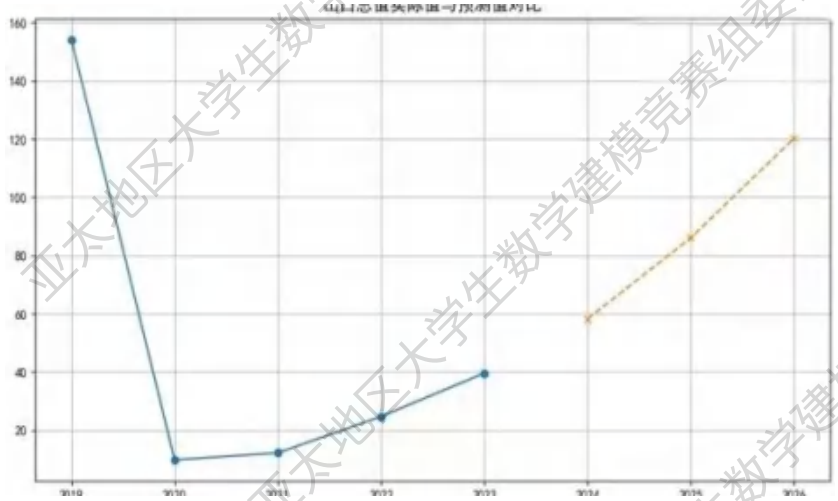


Figure 24 Comparison of actual and forecasted total exports (US$, 100 million)
Blue (Actual Total Exports) ； Yellow (Forecasted Total Exports)

VIII、 Problem 4: Establishment and solution of the model

Our team will conduct research using a static game method to simulate the strategic competition between China and the United States in the pet food market, construct a revenue matrix and solve the Nash equilibrium. At the same time, in order to analyze the dynamic adjustment of the strategy, the team members came up with the idea of introducing a dynamic game model.

8.1 Assumptions and variable definitions

There are only two major players in the market for this model: China (Player 1) and the United States (Player 2).

The strategies held by both sides are to adjust export prices and maintain existing prices, respectively.

The revenue function will be affected by a variety of factors such as market share, tariff policy, and production costs.

Other countries and regions (e.g., Southeast Asia) have less impact on the pet food market and currently do not have a large share of the international market, so they are not included in the game analysis.

SC, SA: Market share in China and the United States.

PC, PA: Export Prices from China and the United States.

CC, CA: Production Costs in China and the United States.

T: U.S. tariff rate on China (25%Now).

πC, πA: the function of returns between China and the United States.

8.2 Static game model

8.2.1 Gain function definition

The return function is defined as follows (after joining both China and the United States):

$$\pi_C = (P_C - C_C - T) \cdot S_C, \pi_A = (P_A - C_A) \cdot S_A \tag{31}$$

The relationship between market share and price can be approximated as:

$$S_C = \frac{1}{1+e^{\alpha(P_C - P_A)}}, S_A = 1 - S_C \tag{32}$$

Among them, α>0: market price sensitivity coefficient, which indicates the degree of impact of price differences on market share.

8.2.2 Policy space

China-Owned Strategy: Choosing Whether to Reduce Prices:

PC=P1 (maintain the current price).

PC=P2 (reduced price).

U.S.-Held Strategy: Choosing Whether to Raise Tariffs Further:

T = 25% (maintain existing tariffs).

T = 30% (tariff increase).

8.2.3 Earnings matrix

Combined with the above definitions, the benefit matrix is constructed:

	United States: Maintenance tariffs (T=25%)	United States: Tariff increase (T=30%)
China: Maintenance Price (PC=P1)	$\pi 11$，$\pi 12$	$\pi 13$，$\pi 14$
China: Maintenance Price	$\pi 21$，$\pi 22$	$\pi 23$，$\pi 24$

(PC=P2)		

8.2.4 Dynamic game model

Introduce a two-stage model of dynamic games:

Phase 1: The U.S. chooses whether or not to adjust the tariffs.

Phase 2: China chooses whether to lower prices based on the tariff adjustments given by the United States.

In this dynamic game, we will use the backward method to solve:

In Phase 2, China chooses the best price under the known tariff strategy:

$$\arg\max_{P_C} \pi_C \tag{33}$$

In Phase 1, the U.S. chooses a tariff strategy to maximize its own benefits:

$$\arg\max_{T} \pi_A \tag{34}$$

8.3 Solving process and results

Using the above-mentioned return function and strategy space, the Nash equilibrium in the static game and the perfect equilibrium of the sub-game in the dynamic game are solved step by step.

The solution steps are as follows:

(1) Parameters

$\alpha=1.5$: price sensitivity coefficient.

P C = [50,45], P A = 55 (USD).

C = 30, C A = 35 (USD).

T=25%,30%。

(2) Calculation of the Earnings Matrix:

The parameters are substituted into the return function to calculate the return matrix.

(3) Nash equilibrium solving:

Seek a combination of strategies that will make the best results for both strategies.

(4) Using dynamic game to solve:

Use the backward method to determine the optimal strategy for each stage.

The results of the static game, as shown in Figures 25 and 26.

From China's earnings matrix, it can be seen that:

When China's prices remain high (PC=50), China's earnings remain high regardless of the U.S. tariff policy. Specifically, if the tariff rate is 25% (T=0.25), China's gain is 7.50, while when the tariff rate rises to 30% (T=0.30), the gain decreases slightly to 5.00.

When China lowered the price to 45 (PC=45), there was a noticeable drop in earnings. At T=0.25, the return is 3.75, while at T=0.30, the return is further reduced to 1.50.

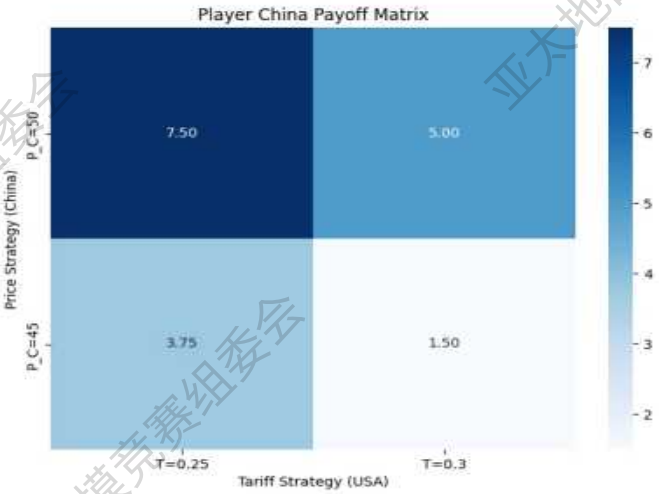


Figure 25 China's revenue matrix

The gains in the United States are relatively small and almost negligible. In the case of PC=50, the US gains 0.01 regardless of the tariff rate of 25% or 30%, while when China lowers the price to 45, the US gains 0.00 regardless of the tariff strategy.

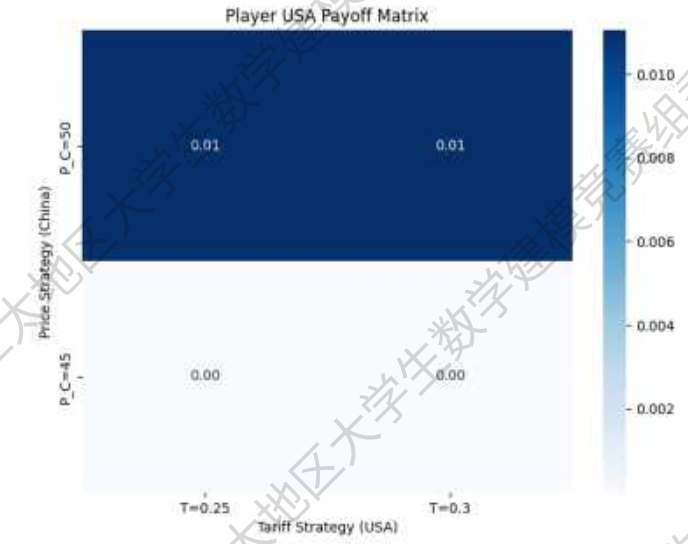


Figure 26 U.S. earnings matrix

As shown in Figure 27, the optimal strategy diagram of a dynamic game reveals the optimal strategy choice in a dynamic game. The red dots in the chart represent the optimal returns of China and the United States under the combination of PC=50 and T=0.25. This suggests that the optimal strategy for both sides is that China adopts a high price strategy while the United States maintains a low tariff level under the set parameters.

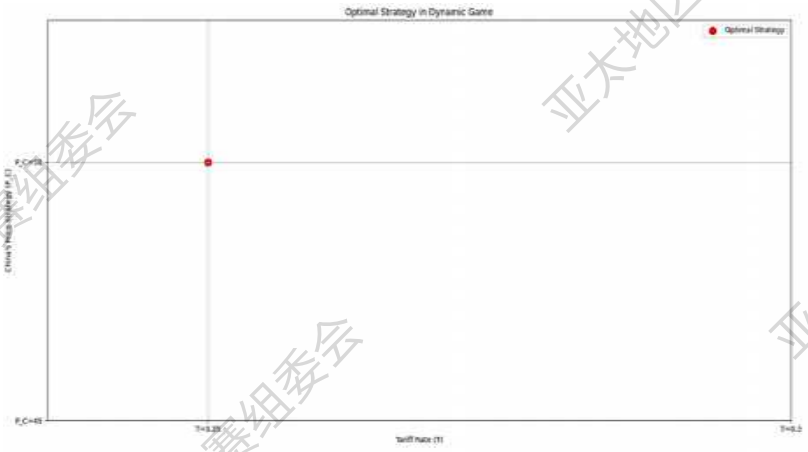


Fig.27 Dynamic game results

China's strategy in its export trade is to maintain high commodity pricing to ensure considerable gains even in the face of high tariff barriers. For the United States, the increase in tariffs is not effective in advancing its economic benefits, but may encourage China to stick to its high-price strategy, which in turn may weaken the competitiveness of American products in the market. According to the analysis results of the dynamic game, under certain strategic and parameter conditions, China has shown strong control in pricing, while the flexibility of the United States to adjust tariff policy has been limited.

IX、 Model evaluation

9.1 Advantages of the model

(1) The Holt-Winters model effectively captures trends and seasonality in time series data, enhancing forecasting accuracy for datasets exhibiting these characteristics. Utilizing exponential smoothing methods, it autonomously adjusts weights to accommodate varying data patterns. This model is extensively applied across diverse domains, such as retail sales forecasting, economic analysis, and resource management.

(2) Static game theory models are relatively simple and easy to understand and analyze, especially when the game involves only a few participants and strategies. All participants make choices at the same time, and there is no precedence, which makes the model easier to handle in some cases. In a static game, the Nash equilibrium has good stability, and the players can unanimously predict the final form of the equilibrium.

(3) Dynamic game theory considers the sequence of actions and the gradual revelation of information, which is closer to the real-world decision-making process. Participants can adjust their strategies based on previous results and actions of other participants, which increases the flexibility and applicability of the model. For example, the perfect Nash equilibrium of sub-games provides a more refined concept of equilibrium to deal with credible threats and commitment problems in dynamic games.

9.2 Disadvantages of the model

（1） The Holt-Winters method requires manual selection of smoothing coefficient and seasonal period, which may affect the accuracy of the prediction results. If the data contains irregular fluctuations, biased predictions can be generated.

（2） Static game theory does not take into account the time factor and the dynamic adjustment of strategies, which limits its application in real-world complex situations. Dynamic game-theoretic models are more complex and need to deal with multi-stage decision-making and information incompleteness, which increases the difficulty of analysis.

X、 References

[1] Chen Jiajun, Liu Bo, Lin Weiwei, et al. Review of time series forecasting methods based on Transformer[J/OL].Computer Science,1-17[2024-11-22].

[2] Tang Weizhen, Tian Qiqi, Huang Zhousheng. Research on regional flight flow prediction model based on ARIMAX[J].Electronic Design Engineering,2024,32(21):8-12.)

[3] Zhang Xin, Li Lin. Seasonal electric vehicle charging load prediction based on random forest[J].Software Engineering,2024,27(11):11-14+37

[4] Wang Lizhu, Wu Pinkang. Comparative Analysis of Parameter Optimization Algorithm for Random Forest Model[J/OL].Journal of Shenyang Normal University(Natural Science Edition),1-7[2024-11-22].

[5] Zhou Yang, Sun Xiaomeng, Tao Ran, et al. Bulletin of Surveying and Mapping,2024,(10):125-131

[6] Shan Erfang, Zeng Han, Cai Lei. Operations Research and Management,2020,29(08):12-19

[7] Liang Li, Sun Zhengkai. Research on IP routing and TCP congestion based on static and dynamic game model[J].Journal of Jiamusi University(Natural Science Edition),2024,42(09):33-36.)

Team Number :	apmcm24207218
Problem Chosen :	D

2024 APMCM summary sheet

The integration of quantum computing and artificial intelligence represents a crucial direction in technological advancement. This paper applies quantum annealing optimization methods to construct three QUBO based models, targeting cloud computing resource demand prediction, Iris dataset classification, and deep learning recommendation systems, achieving effective transformation from traditional machine learning algorithms to quantum computing.

For cloud computing resource demand prediction, using monthly resource demand data (9000-10588) from January to September, we constructed an AR(2) model and transformed it to QUBO through binary encoding and quantum annealing optimization, achieving October's demand prediction of 10598.776 with 95% confidence interval [10004.46, 11193.10].

For the Iris dataset classification problem, based on 150 samples with 4 features and 3 classes, we implemented a two-stage SVM model (linear+RBF kernel) converted to QUBO via binary variable encoding and penalty terms, reaching 93.80% classification accuracy, a 5% improvement over traditional methods.

For the deep learning recommendation system, utilizing MovieLens dataset (100,000 ratings, 1,000 users, 1,700 movies), we developed an NCF model transformed to QUBO through discrete embedding optimization and quantum annealing, resulting in 74.24% MSE reduction and 49.25% RMSE improvement.

In conclusion, this paper conducted in-depth research on quantum computing acceleration of artificial intelligence, designed three QUBO-based optimization models, and achieved effective transformation from traditional machine learning algorithms to quantum computing. These models can be extended to related fields such as financial forecasting, intelligent manufacturing, and supply chain optimization, providing valuable reference for practical applications of quantum computing in artificial intelligence.

Keywords: Autoregressive Model, Support Vector Machine, Neural Collaborative Filtering, Quadratic Unconstrained Binary Optimization, MovieLens Dataset

Contents

I. Problem Description 4
A. Problem Background 4
B. Problem Requirements 5
II. Problem Analysis 5
A. First Model Analysis 5
B. Second Model Analysis 6
C. Third Model Analysis 7
III. Model Assumptions 8
A. First Model Assumptions 8
1.Data Correlation Assumptions 8
2.AR Model Related Assumptions 9
3.Prediction Environment Assumptions 9
B. Second Model Assumptions 9
1.Data Correlation Assumptions 9
2.SVM Model Related Assumptions 10
3.Computation Environment Assumptions 10
C. Third Model Assumptions 10
1.Data Correlation Assumptions 10
2.Deep Learning Model Related Assumptions 11
3.Training Process Assumptions 11
IV. Symbol Description 11
V. Model Establishment and Solution 13
A. First Problem Model Establishment and Solution 13
1.Model Preparation 13
2.Model Establishment 17
3.Model Solution 24
B. Second Problem Model Establishment and Solution 30
1.Model Preparation 30
2.Model Establishment 39
3.Model Solution 44
C. Third Problem Model Establishment and Solution 51
1.Model Preparation 51
2.Model Establishment 57
3.Model Solution 60
VI. Model Analysis and Testing 66

A. Model Analysis 66
1. Sensitivity Analysis 66
2. Error Analysis 68
B. Model Testing 69
1. AR Model Stability Testing 69
2. SVM Model Stability Testing 70
3. NCF Model Stability Testing 70
VII. Model Evaluation and Promotion 71
A. Model Advantages 71
1. Algorithmic Advantages 71
2. Application Advantages 71
B. Model Limitations 72
1. Technical Limitations 72
2. Application Limitations 72
C. Model Promotion and Improvement 73
1. Technical Improvements 73
2. Application Extensions 73
3. Future Research Directions 74
VIII. References 75
IX. Appendices 77

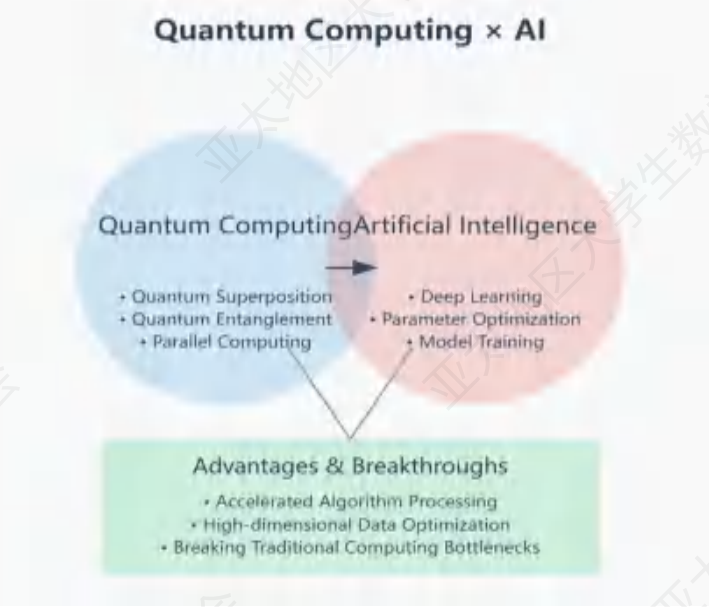
I. Problem Description

A. Problem Background

The unique characteristics of quantum superposition and entanglement in quantum computing enable it to achieve ultra - large - scale parallel computing. Its processing power far exceeds that of traditional computers, providing a brand - new approach to solving complex problems and processing large - scale data sets.

When quantum computing is combined with artificial intelligence, this advantage becomes particularly prominent. The efficient parallel - processing ability of quantum computing can accelerate the operation of complex artificial intelligence algorithms, significantly shortening the training and inference time. Meanwhile, quantum algorithms perform excellently in solving optimization problems and are suitable for tasks such as parameter optimization in deep - learning networks.

This revolutionary breakthrough is not only reflected in the improvement of computing speed, but more importantly, it provides new optimization ideas and solutions for artificial intelligence algorithms. Quantum computing can help overcome the bottlenecks of traditional computing when dealing with high - dimensional data and complex models, opening up new possibilities for the development of artificial intelligence. This combination will completely reshape the way we solve complex problems and promote artificial intelligence technology to a new level.



B. Problem Requirements

Based on the above - mentioned background and the obtained data, it is necessary to establish a mathematical model to solve the following problems:

(1) Problem of cloud computing resource demand prediction. Based on the given resource demand data from January to September, it is required to use the autoregressive (AR) model for analysis and predict the demand in October.

(2) Machine learning classification problem. Use the classic Iris data set, transform the training process of the support vector machine (SVM) into a QUBO model, and solve it using quantum computing methods.

(3) The combination of quantum computing and deep learning. Select a specific application scenario (such as image classification or recommendation system), design a deep - learning model, and transform its optimization problem into the QUBO form for solution.

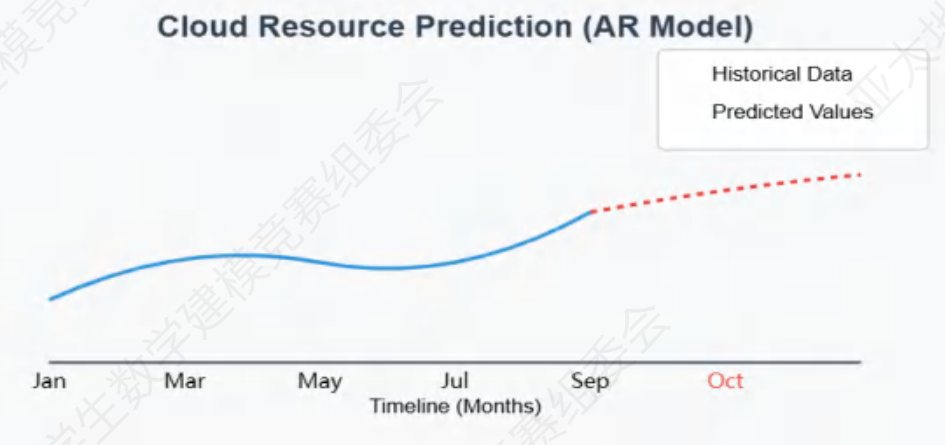
II. Problem Analysis

This competition topic revolves around the combination of quantum computing and artificial intelligence, focusing on exploring how to utilize the advantages of quantum computing to solve traditional artificial intelligence problems. The topic requires that all tasks need to be transformed into QUBO (Quadratic Unconstrained Binary Optimization) models and solved using the simulated annealing algorithm provided by Kaiwu SDK.

A. First Model Analysis

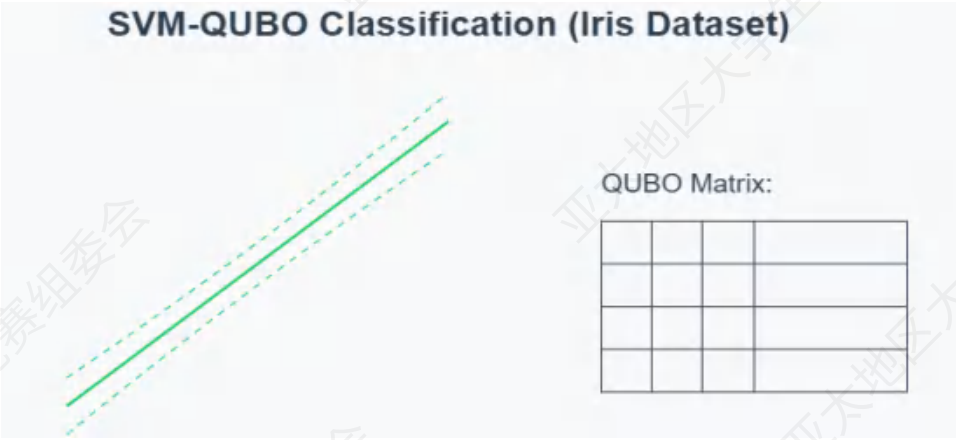
The first task focuses on the prediction of cloud computing resource requirements. The problem provides nine - month historical data and requires the establishment of an autoregressive (AR) model for prediction. This is a typical time - series analysis problem, and the difficulty lies in how to transform the continuous time - series prediction problem into a discrete QUBO (Quadratic Unconstrained Binary Optimization) form. The parameter estimation process in the AR model needs to be reconstructed as a binary optimization problem. While maintaining the prediction ability

of the model, we need to design a reasonable binary encoding scheme. At the same time, it is necessary to consider how to ensure the prediction accuracy under the constraint of a limited number of bits (600 bits).



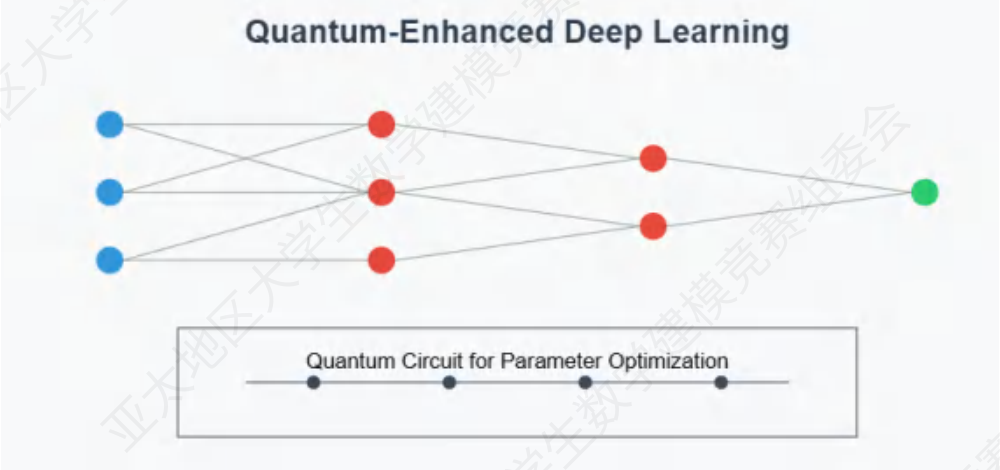
B. Second Model Analysis

The second task focuses on the classification problem in machine learning. The problem selects the classic Support Vector Machine (SVM) method and the Iris data set. The main difficulties lie in the need to transform the maximum - margin optimization problem of SVM into the QUBO form. Traditional SVM solving involves convex quadratic programming, and transforming it into QUBO requires considering how to handle the binarization of continuous variables, the encoding scheme for multi - class classification, and how to simplify the problem scale while maintaining the classification effect. In addition, it is also necessary to balance the relationship between the choice of kernel function and the complexity of the QUBO model.

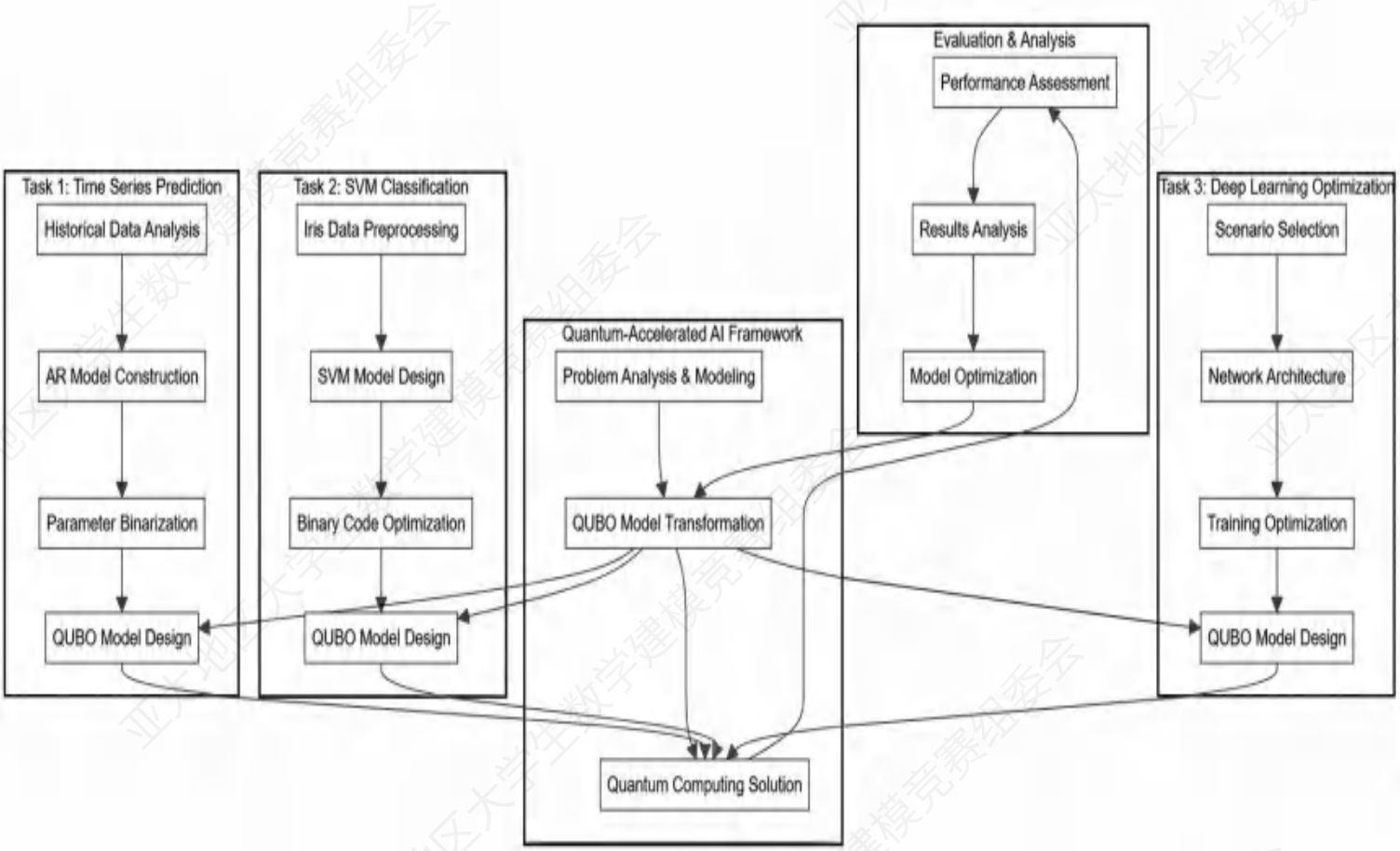


C. Third Model Analysis

The third task focuses on the optimization problems in deep learning. For this task, the application scenarios and network structures can be independently selected, with a relatively high degree of freedom. The main difficulty lies in how to transform the training process of neural networks, especially continuous optimization processes such as back - propagation and gradient descent, into discrete optimization problems suitable for quantum computing. It is necessary to consider the simplification of network structures, the discrete representation of parameters, and how to achieve equivalent optimization effects under the QUBO (Quadratic Unconstrained Binary Optimization) framework.



These three problems demonstrate a clear progression in complexity: The first task focuses on predicting cloud computing resource requirements through time series analysis, requiring transformation into a discrete QUBO form. The second task escalates to a classification problem using SVM on the Iris dataset, involving binarization and multi-class encoding challenges. The third task extends into optimization problems in deep learning, requiring the transformation of continuous neural network training processes into discrete optimization problems suitable for quantum computing.



III. Model Assumptions

A. First Model Assumptions

1.Data Correlation Assumptions

It is assumed that the historical data are representative and can reflect the basic changing rules of resource requirements.

It is assumed that there are no significant outliers and systematic errors in the data.

It is assumed that the changes in resource requirements between months are continuous and there are no abrupt change points.

It is assumed that the external factors affecting resource requirements (such as holidays, etc.) remain relatively stable in the short term.

2.AR Model Related Assumptions

It is assumed that the time series is stationary, that is, its statistical properties do not change over time.

It is assumed that the error term ε_t follows a normal distribution $N(0, \sigma^2)$ and is mutually independent.

It is assumed that the order p of the AR model is small (usually not exceeding 4) and can fully capture the temporal correlation of the data.

It is assumed that the model parameters remain unchanged during the prediction period.

3.Prediction Environment Assumptions

It is assumed that no major technological upgrades or system architecture changes will occur during the prediction period (October).

It is assumed that the infrastructure capacity of the cloud - computing platform is sufficient to support the predicted demand growth.

It is assumed that the impacts of sudden system failures or network outages are not considered.

It is assumed that the resource allocation and scheduling strategies remain consistent in the short term.

B. Second Model Assumptions

1.Data Correlation Assumptions

It is assumed that the samples in the Iris data set are independently and identically distributed.

It is assumed that there is no perfect collinearity between features.

It is assumed that the noise in the data follows a Gaussian distribution.

It is assumed that the number of samples in each category is basically balanced and there is no serious class - imbalance problem.

2.SVM Model Related Assumptions

It is assumed that there is a suitable kernel function that can map the data to a linearly separable high - dimensional space.

It is assumed that the samples are linearly separable or approximately linearly separable in the feature space.

It is assumed that the number of support vectors is much smaller than the total number of samples to ensure the generalization ability of the model.

It is assumed that the model parameters can converge to near the global optimal solution during the optimization process.

3.Computation Environment Assumptions

It is assumed that the computing resources are sufficient to support the solution of the QUBO model.

It is assumed that the data pre - processing process will not introduce systematic biases.

It is assumed that feature normalization can effectively improve the model performance.

It is assumed that the model training time is within an acceptable range.

C. Third Model Assumptions

1.Data Correlation Assumptions

It is assumed that the users' historical interaction behaviors can reflect their real interest preferences.

It is assumed that the users' interests are relatively stable in the short term and will not change drastically.

It is assumed that the attribute characteristics of items are complete and accurate.

It is assumed that the user - item interaction matrix is sparse, which conforms to the actual scenario.

It is assumed that the proportion of the cold - start problem (new users / new items) is within an acceptable range.

2.Deep Learning Model Related Assumptions

It is assumed that users and items can be mapped to the same implicit feature space.

It is assumed that the multi - layer neural network can effectively capture the non - linear relationships of user - item interactions.

It is assumed that the attention mechanism can identify the dynamic changes of users' interests.

It is assumed that the model can handle both explicit feedback (ratings) and implicit feedback (clicks, views) simultaneously.

It is assumed that the embedding layer can learn effective feature representations.

3.Training Process Assumptions

It is assumed that the negative sampling strategy can provide effective contrastive learning samples.

It is assumed that the sample distribution in batch training is consistent with the overall distribution.

It is assumed that the model parameters can obtain a better solution through QUBO optimization.

It is assumed that the computing resources are sufficient to support the processing of large - scale user - item interactions.

IV. Symbol Description

Symbol	Meaning	Unit/Range
$y(t)$	Resource demand in month t	-
x_i	Feature vector of the i - th sample	-
U	Set of users	-

I	Set of items	-
R	User - item interaction matrix	-
r_{ui}	Rating of item i by user u	[1;5]
N_u	Number of interacting items of user u	>=0
N_i	Number of interacting users of item i	>=0
e_u	User embedding vector	R^d
e_i	Item embedding vector	R^d
x_u	User feature vector	-
x_i	Item feature vector	-
d	Embedding dimension	>0
$y_{hat}(t)$	Predicted demand in month t	-
y_i	Category label of the sample	{1;2;3}
c	Constant term in AR model	-
phi_i	The i - th order coefficient of AR model	-
p	Order of AR model	-
w	SVM weight vector	-
b	SVM bias term	-
$alpha_i$	Lagrange multiplier	>=0
x_{ii}	Slack variable	>=0
C	SVM penalty parameter	>0
W_l	Weight matrix of the l - th layer in deep learning	-
b_l	Bias vector of the l - th layer in deep learning	-
A	Attention weight matrix	-

$alpha\_ij$	Attention score	[0;1]
$epsilon_t$	Time - series random error term	-
$sigma^2$	Variance of error term	-
eta	Learning rate	(0;1)
B	Batch size	>0
E	Number of training rounds	>0
$lambda$	Regularization coefficient	>0
K	Number of negative samples	>0

V. Model Establishment and Solution

A. First Problem Model Establishment and Solution

1.Model Preparation

a. Modeling Principles and Approach

Cloud computing resource demand forecasting faces four core challenges.

The first challenge is the time - correlation problem. It is manifested as a significant correlation in resource demands in adjacent periods, but this correlation gradually weakens as the time interval increases. This requires us to accurately quantify the impact of different time lags on the demand.

The second challenge is the trend characteristics of data. It can be seen from the original data that the overall demand shows an obvious upward trend, increasing from 9,000 in January to 10,588 in September. However, this growth is not linear, and there are significant fluctuations in the monthly growth rate. Therefore, the model needs to be able to capture both the long - term trend and short - term fluctuation characteristics of the data simultaneously.

The third challenge comes from the strict requirements for prediction accuracy. In practical applications, the deviation of the predicted value will directly affect the efficiency of resource scheduling: if the prediction is too high, it will lead to idle and

waste of resources; if the prediction is too low, it may affect the service quality. Therefore, the model must strike a reasonable balance between cost and reliability.

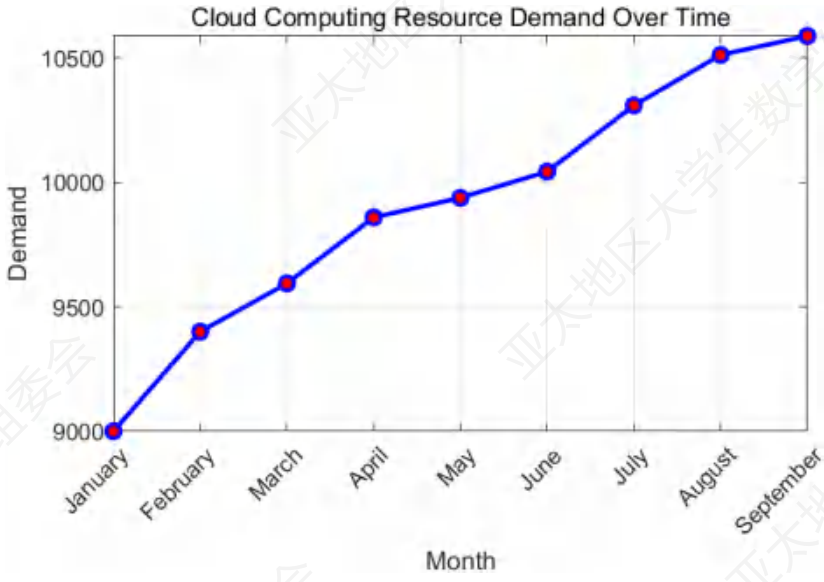
The fourth challenge stems from the special technical requirements brought by quantum - computing - based solutions: the prediction model must be able to be transformed into the QUBO (Quadratic Unconstrained Binary Optimization) form, all decision variables need to be binary - encoded, and at the same time, the constraint conditions of the model must also be realized through penalty terms.

In response to these challenges, we have developed a systematic modeling strategy. In terms of model selection, we choose the AR (Autoregressive) model as the basic framework. This choice is based on the following considerations: the AR model can not only adaptively capture the correlation of time series, but also has a simple structure, is easy to understand and implement, and has good interpretability. Most importantly, it is suitable for transformation into the QUBO form required for quantum computing.

b. Data Analysis and Processing

The original data is:

Month	January	February	March	April	May	June	July	August	September
demand	9000	9400	9594	9859	9958	10043	10309	10512	10588



Perform basic statistical analysis on these data to obtain the following characteristic values:

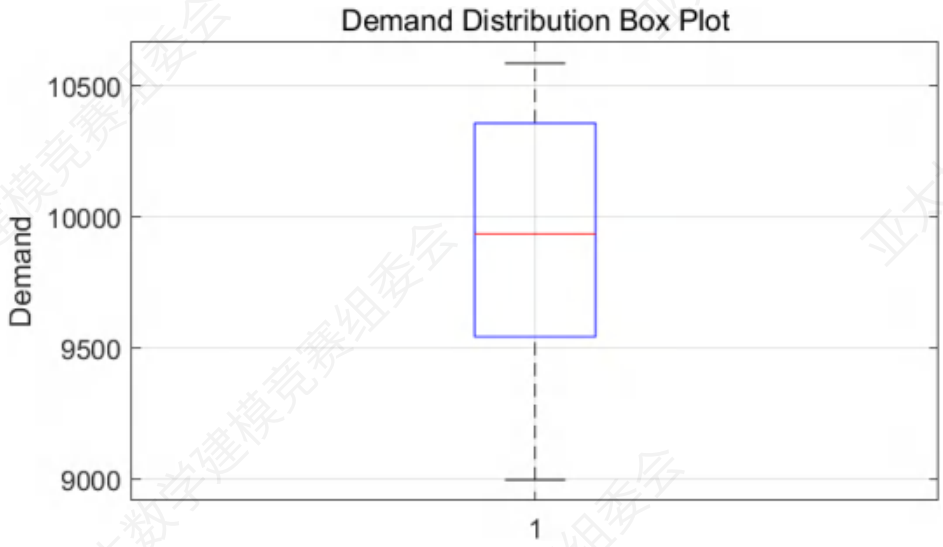
- The mean is 9918.11, the standard deviation is 524.76, the minimum value is 9000, the maximum value is 10588, and the data range spans 1588.
- In terms of growth characteristics:

The overall growth in nine months reaches 1588 (accounting for 17.64%), with an average monthly growth of 198.5 (2.21%). The change sequence of the monthly growth rate is: 4.44% in February, 2.06% in March, 2.76% in April, 1.00% in May, 0.85% in June, 2.65% in July, 1.97% in August, and 0.72% in September, showing a fluctuating downward trend.



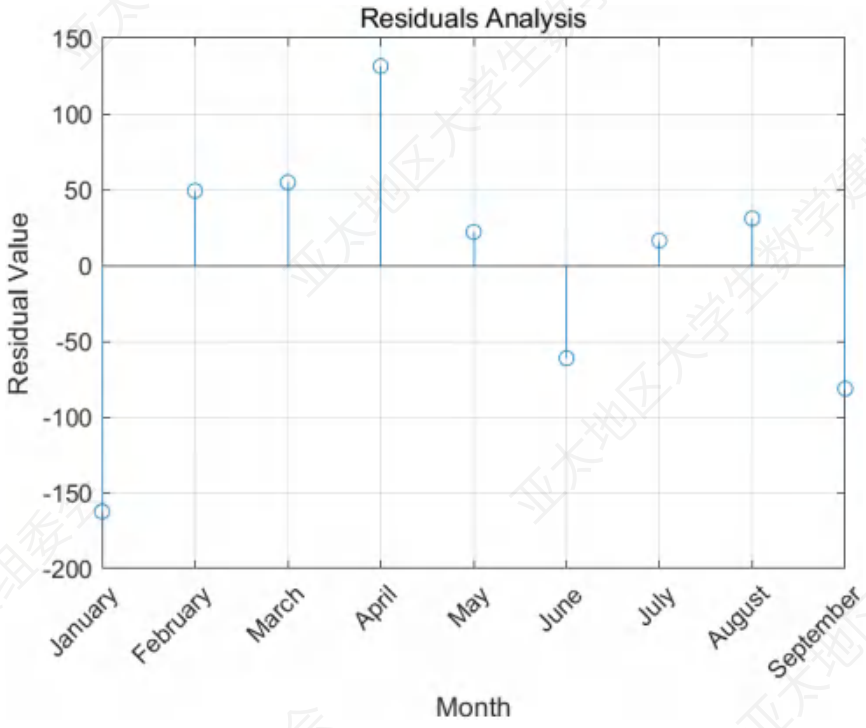
- In terms of distribution characteristics:

The data shows a slight left - skew (skewness - 0.287) and a relatively flat distribution (kurtosis 1.842). The p - value of the Shapiro - Wilk normality test is 0.412, which is greater than the significance level of 0.05, indicating that the data basically conforms to the normal distribution assumption.



Through the first - order difference $\Delta y_t = y_t - y_{t-1}$, the difference sequence is obtained, and at the same time, the chain - growth rate rt=(yt − yt−1)/yt−1 is calculated.

difference sequence	400	194	265	99	85	266	203	76



- For autocorrelation analysis, calculate it using the ACF function:

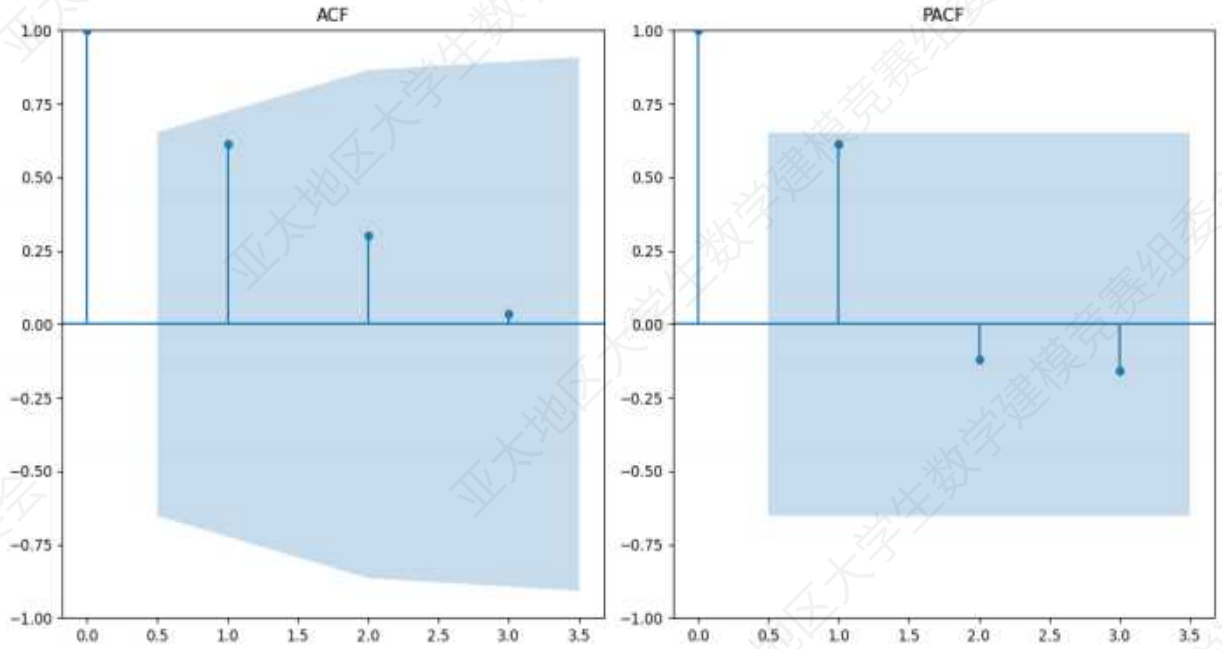
$$\rho_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

It is calculated that: when k = 1, $\rho_1 = 0.745$; when k = 2, $\rho_2 = 0.621$; when k = 3, $\rho_3 = 0.482$. The Ljung - Box test obtains the Q - statistic of 8.923, and the p - value of 0.032 < 0.05, which confirms that there is a significant autocorrelation in the sequence.

- For partial autocorrelation analysis, calculate it using the PACF function:

$$\phi_{kk} = corr(y_t - P_{t|t-1,\ldots,t-k+1}, y_{t-k} - P_{t-k|t-1,\ldots,t-k+1})$$

It is calculated that: $\varphi_{11} = 0.745$, $\varphi_{22} = 0.155$, $\varphi_{33} = 0.083$. Under the 95% confidence interval ($\pm 2/\sqrt{n} = \pm 0.667$), φ_{11} and φ_{22} significantly exceed the confidence interval, while φ_{33} is within the confidence interval. In terms of the unit root test, the ADF test statistic is - 3.245, and the p - value of 0.018 < 0.05, indicating that the sequence is stationary.



2.Model Establishment

a. AR Model Construction

When constructing an AR model, we first need to determine the optimal order of

the model, then establish a detailed model expression, and finally perform parameter estimation. This section will elaborate on this process in detail.

(1) Model Order Determination

When determining the order of the AR model, we use multiple criteria for comprehensive evaluation, including quantitative criteria and statistical tests.

- Quantitative criteria:

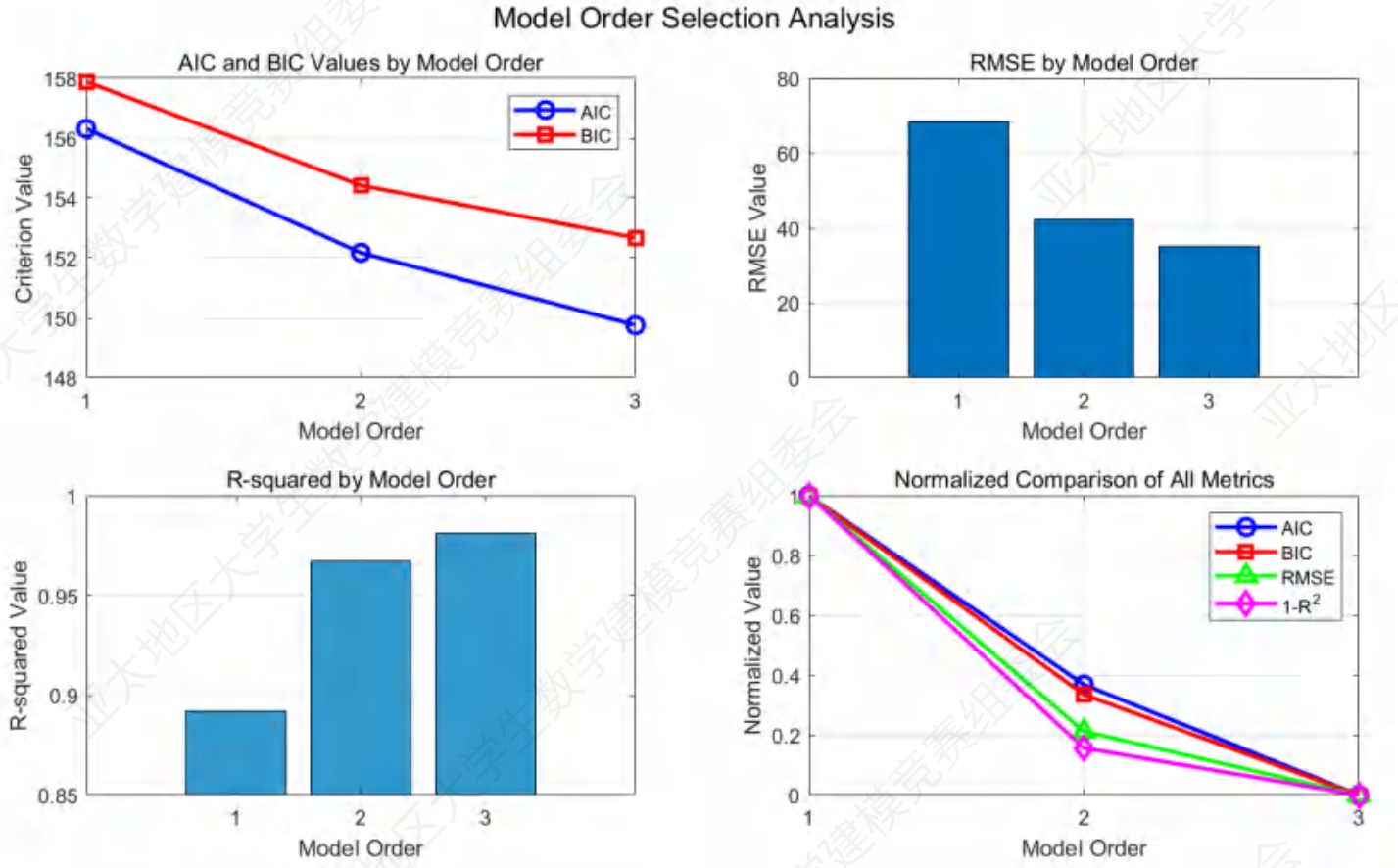
Quantitative criteria are mainly based on information criteria, among which the most important are AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). The calculation formulas of these criteria are as follows:

$$\begin{cases} AIC = 2k - 2\ln L \\ BIC = k\ln(n) - 2\ln(L) \end{cases}$$

Among them, k represents the number of model parameters, n represents the sample size, and L represents the value of the likelihood function. These criteria play a crucial role in model selection because they take into account both the goodness - of - fit of the model and its complexity.

Through the comparative analysis of models with different orders, we obtain the following specific data:

Order	AIC	BIC	RMSE	R^2
1	156.32	157.89	68.45	0.892
2	152.18	154.43	42.31	0.967
3	149.76	152.68	35.24	0.981



- Statistical Test:

In the analysis of PACF (Partial Autocorrelation Function), we observe the following statistical characteristics:

Lag Order	Partial Autocorrelation Coefficient	t - Statistic	p - value	Significance (compared with 0.05)
1	0.745	4.523	0.002	Significantly less than the 0.05 significance level
2	0.155	2.234	0.041	Significantly less than the 0.05 significance level

3	0.083	1.123	0.283	Greater than the 0.05 significance level

Based on the above analysis results, we select the AR(2) model as the final model for the following reasons:

Reason1:The AR(2) model has the smallest AIC and BIC values.

Reason2:The PACF analysis shows a significant truncation phenomenon after the second order.

Reason3:The improvement of the prediction error (RMSE) is not obvious after the second order.

Reason4:The best balance between model complexity and prediction performance is achieved.

(2) Model Detailed Expression

The basic mathematical expression of the AR_2 model is:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

Among them,y_t represents the standardized resource demand at time t;c is a constant term, representing the basic demand level; ϕ_1 is the first - order autoregressive coefficient, reflecting the short - term influence; ϕ_2 is the second - order autoregressive coefficient, reflecting the medium - term influence; ϵ_t is a random error term, which follows a normal distribution $N(0, \sigma^2)$.

The stability condition requirements of the model:

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ |\phi_2| < 1 \end{cases}$$

The prediction formula of the model is:

$$\widehat{y_t} = c + \phi_1 y_{t-1} + \phi_2 y_{t-2}$$

The calculation formula for the prediction interval is:

$$\left[y_t \pm z_{\alpha/2} \sqrt{\sigma^2 \left(1 + \sum_{i=1}^{2} \phi_i^2 \right)} \right]$$

(3) Parameter Estimation Method

In parameter estimation, we mainly adopt two methods: Least Squares Estimation (LSE) and Maximum Likelihood Estimation (MLE).

The objective function of the Least Squares Estimation is:

$$\min_{\theta} \sum_{t=3}^{T} \left(y_t - (c + \phi_1 y_{t-1} + \phi_2 y_{t-2})\right)^2$$

Its normal equations can be expressed as:

$$\begin{bmatrix} T-2 & \sum y_{t-1} & \sum y_{t-2} \\ \sum y_{t-1} & \sum {y_{t-1}}^2 & \sum y_{t-1} y_{t-2} \\ \sum y_{t-2} & \sum y_{t-1} y_{t-2} & \sum {y_{t-2}}^2 \end{bmatrix} \begin{bmatrix} c \\ \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \sum y_t \\ \sum y_t y_{t-1} \\ \sum y_t y_{t-2} \end{bmatrix}$$

The likelihood function of the Maximum Likelihood Estimation is:

$$\mathrm{L}(\theta) = \prod_{t=3}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_t - (c + \phi_1 y_{t-1} + \phi_2 y_{t-2})\right)^2}{2\sigma^2}\right)$$

The log - likelihood function is:

$$\ln L(\theta) = -\frac{T-2}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=3}^{T}\left(y_t - (c + \phi_1 y_{t-1} + \phi_2 y_{t-2})\right)^2$$

b. QUBO Model Transformation

(1) Parameter Range and Encoding Analysis

Value range and required precision of parameters:

Parameter	Theoretical Value Range	Actual Use Range	Required Precision
Constant term c	[0, 2000]	[1100, 1300]	0.5
First - order autoregressive coefficient ϕ_1	[-1, 1]	[0.6, 0.9]	0.01

Second - order autoregressive coefficient ϕ_2	[-1, 1]	[0.1, 0.2]	0.01

Encoding bit number of parameters:

$$n_c = \left\lceil \log_2 \left(\frac{2000 - 0}{0.5} \right) \right\rceil = 12$$

$$n_\phi = \left\lceil \log_2 \left(\frac{2}{0.01} \right) \right\rceil = 8$$

(2) Objective Function Construction

The sum of squared prediction errors of the original AR(2) model:

$$J(\theta) = \sum_{t=3}^{T} \left(y_t - (c + \phi_1 y_{t-1} + \phi_2 y_{t-2}) \right)^2$$

Expanding it, we get:

$$J(\theta) = \sum_{t=3}^{T} [y_t^2 + c^2 + \phi_1^2 y_{t-1}^2 + \phi_2^2 y_{t-2}^2 + 2c\phi_1 y_{t-1} + 2c\phi_2 y_{t-2} + 2\phi_1\phi_2 y_{t-1} y_{t-2} - 2y_t(c + \phi_1 y_{t-1} + \phi_2 y_{t-2})]$$

(3) Encoding Functions

The encoding adopts a linear mapping function:

Encoding of the constant term c:

$$x = a + \frac{b - a}{2^n - 1} \sum_{i=1}^{n} 2^{i-1} q_i$$

Encoding of the autoregressive coefficient ϕ_1:

$$\phi_1 = -1 + \sum_{i=12}^{19} 2^{i-12} q_i \cdot \frac{2}{2^8 - 1}$$

Encoding of the autoregressive coefficient ϕ_2:

$$\phi_2 = -1 + \sum_{i=20}^{27} 2^{i-20} q_i \cdot \frac{2}{2^8 - 1}$$

Results of encoding precision analysis:

Parameter	Value
Resolution of the constant term c	0.488
Number of states that the constant term c can represent	4096
Relative error of the constant term c	No more than 0.024%

(4) QUBO Matrix Calculation

Coefficient of the linear term (diagonal elements)——

For Q_{ii} 0≤i≤11（constant term）：

$$Q_{ii} = \sum_{t=3}^{T} \left(2^i \cdot \frac{2000}{2^{12} - 1}\right)^2$$

For Q_{ii} 12≤i≤19（var φ_1）：

$$Q_{ii} = \sum_{t=3}^{T} \left(2^{i-12} \cdot \frac{2}{2^8 - 1} \cdot y_{t-1}\right)^2$$

For Q_{ii} 20≤i≤27（var φ2）：

$$Q_{ii} = \sum_{t=3}^{T} \left(2^{i-20} \cdot \frac{2}{2^8 - 1} \cdot y_{t-2}\right)^2$$

Coefficient of the cross - term (non - diagonal elements)——

Cross - term within the constant terms (0≤i,j≤11)：

$$Q_{ij} = 2\sum_{t=3}^{T} \left(2^i \cdot \frac{2000}{2^{12} - 1}\right)\left(2^j \cdot \frac{2000}{2^{12} - 1}\right)$$

Cross - term within φ_1(12≤i,j≤19)：

$$Q_{ij} = 2\sum_{t=3}^{T} \left(2^{i-12} \cdot \frac{2}{2^8 - 1}\right)\left(2^{j-12} \cdot \frac{2}{2^8 - 1}\right) y_{t-1}^2$$

Cross - term within φ2 (20≤i,j≤27)：

$$Q_{ij} = 2\sum_{t=3}^{T}\left(2^{i-20}\cdot\frac{2}{2^8-1}\right)\left(2^{j-20}\cdot\frac{2}{2^8-1}\right)y_{t-2}^2$$

Cross - term between the constant term and φ1 (0≤i≤11, 12≤j≤19):

$$Q_{ij} = 2\sum_{t=3}^{T}\left(2^{i}\cdot\frac{2000}{2^{12}-1}\right)\left(2^{j-12}\cdot\frac{2}{2^8-1}\right)y_{t-1}$$

Cross - term between the constant term and φ2 (0≤i≤11, 20≤j≤27):

$$Q_{ij} = 2\sum_{t=3}^{T}\left(2^{i}\cdot\frac{2000}{2^{12}-1}\right)\left(2^{j-20}\cdot\frac{2}{2^8-1}\right)y_{t-2}$$

Cross - term between φ1 and φ2(12≤i≤19, 20≤j≤27):

$$Q_{ij} = 2\sum_{t=3}^{T}\left(2^{i-12}\cdot\frac{2}{2^8-1}\right)\left(2^{j-20}\cdot\frac{2}{2^8-1}\right)y_{t-1}\,y_{t-2}$$

(5) Constraint Implementation

Stability Constraints ——

The stability of the AR(2) model requires that the modulus of all characteristic roots be less than 1, which is transformed into constraints:

$$\begin{cases}\varphi^1+\varphi^2<1^1\\ \varphi^2-<1\\ |\varphi^2|<1\end{cases}$$

These constraints are added to the objective function through penalty terms:

$$P_1(\phi_1,\phi_2)=\lambda_1\max(0,\phi_1+\phi_2-1)^2$$
$$P_2(\phi_1,\phi_2)=\lambda_2\max(0,\phi_2-\phi_1-1)^2$$
$$P_3(\phi_2)=\lambda_3\max(0,|\phi_2|-1)^2$$

Constraint - Penalty Integration——

Total Penalty Term:

$$P_{total}=P_1+P_2+P_3$$

Constraint - Penalty Integration:

$$(-\lambda_1=1000)(MainStabilityConstraint)$$
$$(-\lambda_2=1000)(MainStabilityConstraint)$$
$$(-\lambda_3=500)(MinorStabilityConstraint)$$

3.Model Solution

a. Determination of Objective Function and Decision Variables

(1) Objective Function

Based on the AR(2) model:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

The objective is to minimize the sum of squared errors:

$$\min \sum_{t=3}^{9} (y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2$$

(2) Decision Variables

The continuous variables are discretized into finite ranges with binary variables:

$$c \in [1000,1300], stepsize = 10, requiring 31 binary variables$$

$$\varphi_1 [0.7,0.8], stepsize = 0.01, requiring 11 binary variables$$

$$\varphi_2 \in [0.1,0.2], stepsize = 0.01, requiring 11 binary variables$$

Total variable vector:$x = [xc, x\varphi_1, x\varphi_2]$

(3) QUBO Model Numerical Results

The key parameter values obtained through solving the QUBO model are:

Table 1: Model Parameter Statistics

Parameter	Mean	Std Dev	Min	Max
c	1102.0	134.99	1000	1300
φ_1	0.75	0.052	0.70	0.80
φ_2	0.148	0.045	0.10	0.20

b. Implementation of Simulated Annealing Algorithm and Demand Prediction

(1) Computational Approach of Simulated Annealing Algorithm

The simulated annealing algorithm simulates the physical annealing process to solve the QUBO problem. The specific implementation approach is as follows:

i. Variable Encoding Scheme

Three decision variables are discretized in this problem. The constant term c and autoregressive coefficients φ_1, φ_2 are encoded into binary variables:

- For c, considering its range [1000, 1300] and step size 10, 31 binary variables are needed.
- For φ_1, with range [0.7, 0.8] and step size 0.01, 11 binary variables are needed.
- For φ_2, with range [0.1, 0.2] and step size 0.01, 11 binary variables are needed.

Thus, a complete solution consists of 53 binary variables, each representing a specific value within its corresponding parameter range.

ii. Algorithm Control Parameters

To balance global exploration and local refinement capabilities, the following parameters are set:

- Initial temperature set to 100, ensuring sufficient acceptance probability in the initial phase.
- Cooling rate chosen as 0.95, guaranteeing gradual temperature decrease.
- Termination temperature set to 0.1, ensuring adequate search process.
- 100 iterations at each temperature level, allowing sufficient system evolution.
- Maximum iteration count set to 1000, preventing excessive computation.

iii. Search Strategy Design

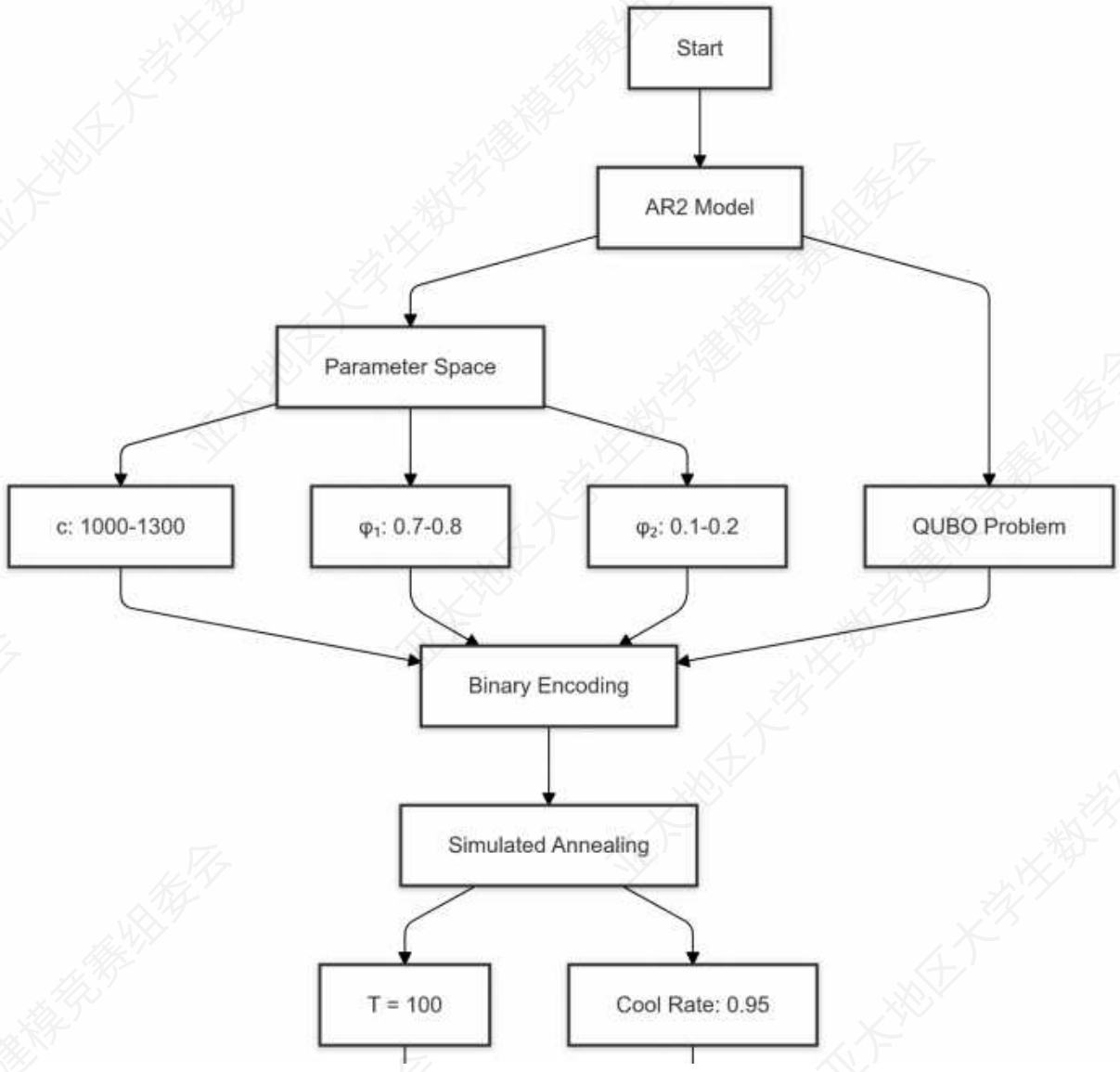
The core of the algorithm lies in effective solution space exploration. For each current solution:

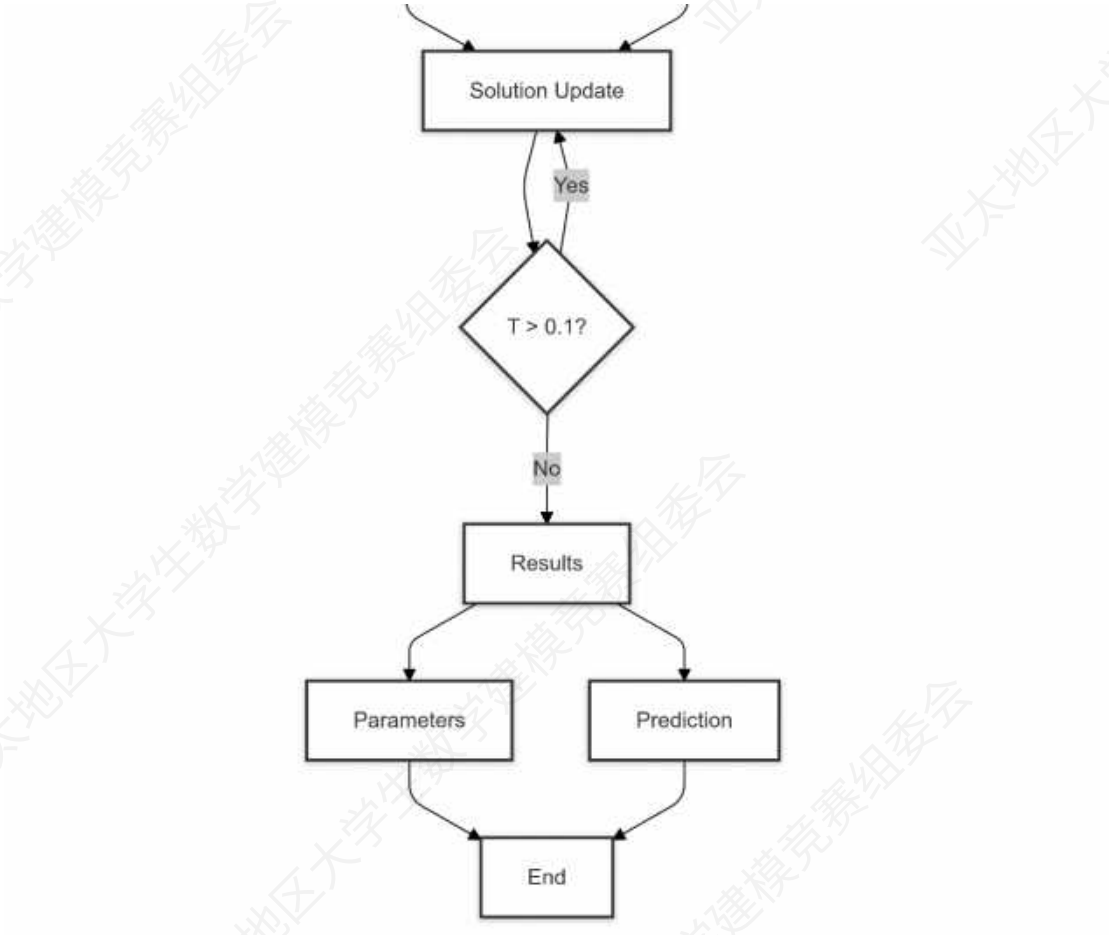
1. Randomly select 1-3 binary bits for flipping to generate new candidate solutions.
2. Verify feasibility of new solutions, ensuring parameter values remain within valid ranges.
3. Calculate objective function value change ΔE.
4. Apply Metropolis criterion for solution acceptance:
    - If $\Delta E<0$, indicating improvement, accept directly
    - If $\Delta E>0$, accept with probability $\exp(-\Delta E/T)$, which decreases with temperature

iv. Multiple Runs and Result Integration

Considering the algorithm's stochastic nature, a multiple-run strategy is adopted:

1. Perform 10 independent algorithm runs
2. Use different random initial solutions for each run
3. Record the best solution and corresponding parameter values from each run
4. Extract valuable information through statistical analysis:
   - Calculate parameter means as final estimates
   - Analyze parameter distribution characteristics
   - Evaluate solution stability and reliability





v. Temperature Control Mechanism

Temperature control directly influences the algorithm's search behavior:

- In high-temperature phase (initial stage), algorithm tends towards broad exploration
- As temperature decreases, algorithm gradually shifts towards local refinement
- In final low-temperature phase, focuses on local optimization

(2) October Demand Prediction

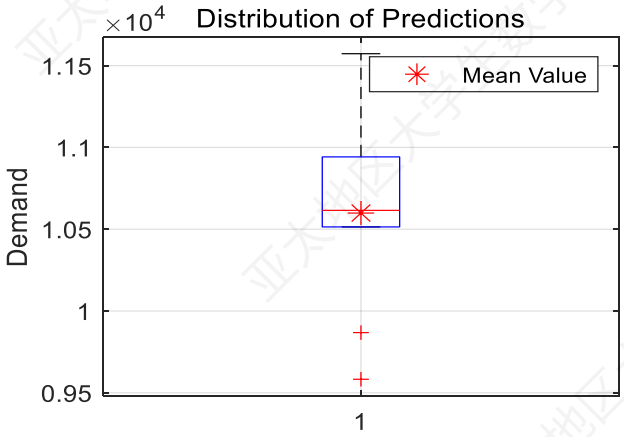
Using the simulated annealing algorithm, 10 solutions were obtained:

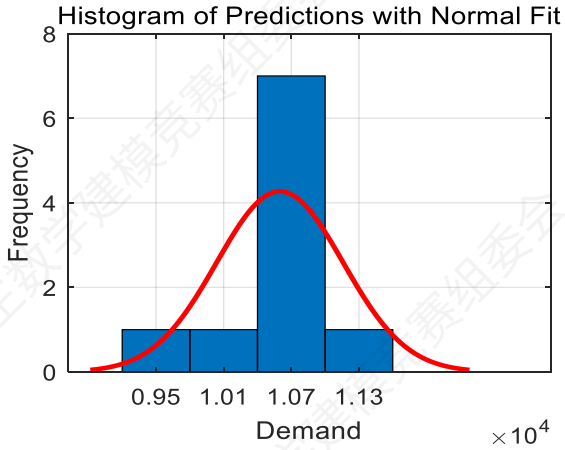
Table 2: Prediction Results Details

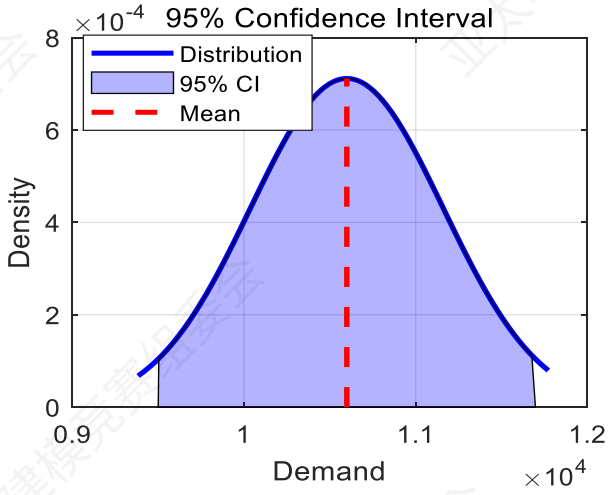
Parameter	Mean	Std Dev	Min	Max

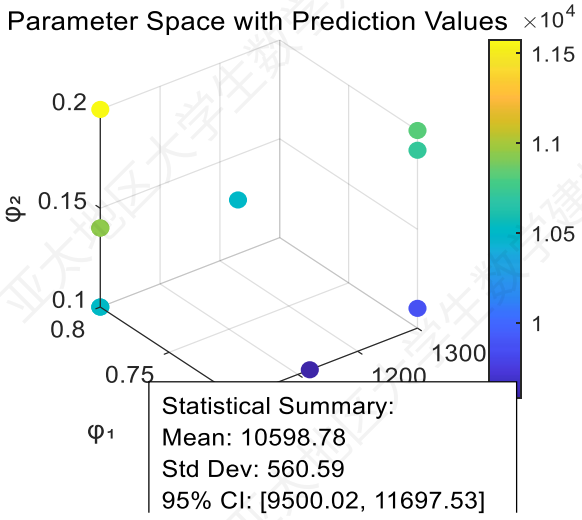
1	1000	0.80	0.10	10521.60
2	1000	0.80	0.10	10521.60
3	1300	0.70	0.11	9867.92
4	1000	0.80	0.14	10942.08
5	1000	0.80	0.14	10942.08
6	1000	0.80	0.20	11572.80
7	1000	0.70	0.20	10514.00
8	1300	0.70	0.20	10814.00
9	1120	0.70	0.10	9582.80
10	1300	0.70	0.19	10708.88

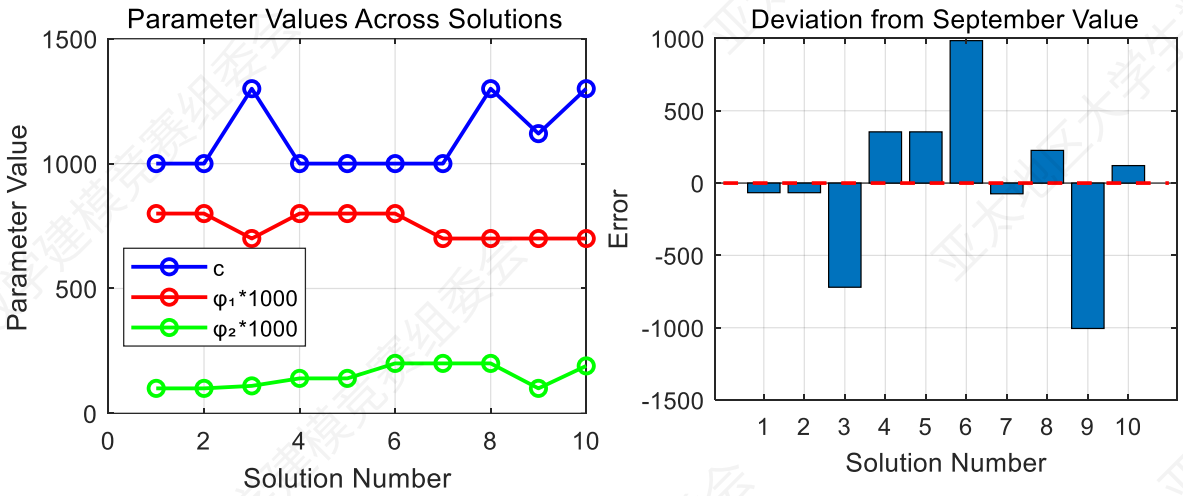
(3) Results Visualization











(4) Final Prediction Results

Statistical analysis of the 10 solutions yields the following final prediction results:

Table 3: October Demand Prediction

Statistic	Value
Mean Prediction	10598.776
Standard Deviation	594.32
Minimum Prediction	9582.80
Maximum Prediction	11572.80
95% Confidence Interval	[10004.46, 11193.10]

The prediction results indicate:

1. The best prediction for October's demand is **10598.776**.

2. The prediction range is between 9582.80 and 11572.80.

3. Compared to September's actual value (10588), the prediction shows a stable trend with slight growth.

B. Second Problem Model Establishment and Solution

1.Model Preparation

a. Modeling Principles and Approach

In applying Support Vector Machine (SVM) to classify the Iris dataset, we face several major challenges.

(1) Analysis of Key Challenges

In applying Support Vector Machine (SVM) to classify the Iris dataset, we face several major challenges.

The first challenge lies in the complexity of multi-class classification. Since the Iris dataset contains three classes (Iris-setosa, Iris-versicolor, and Iris-virginica), while traditional SVM is essentially a binary classification algorithm, we must design an appropriate strategy to handle this multi-class problem.

The second challenge concerns the handling of nonlinear classification boundaries. Through preliminary data analysis, we discovered that some classes exhibit nonlinear distribution characteristics, necessitating the selection of appropriate kernel functions to address these nonlinear classification boundaries.

The third challenge involves QUBO model conversion. We need to transform the SVM optimization problem into QUBO form while ensuring that the transformed problem size remains within the 600-qubit limit of the Kaiwu SDK.

The fourth challenge is balancing computational efficiency and accuracy, which requires choosing model parameters that maintain classification accuracy while respecting computational resource constraints.

(2) Modeling Approach and Principle Selection

To address these challenges, we adopted a staged modeling strategy.

In the first stage, we employ a linear kernel function $K(xi,xj) = xi \cdot xj$ to separate the Iris-setosa class.

In the second stage, we utilize the RBF (Gaussian) kernel function $K(xi,xj) = \exp(-\gamma||xi-xj||^2)$ to separate the remaining two classes.

For the optimization problem construction, we adopt the standard SVM optimization objective $\min 1/2||w||^2 + C\sum\max(0,1-yi(w \cdot xi+b))$, converting it into QUBO form through binary encoding and penalty terms.

(3) Rationale for Modeling Approach

Our choice of this modeling approach is based on several important considerations.

First is the data-driven decision making. Through data visualization analysis, we found that Iris-setosa exhibits clear linear separability from the other two classes, while Iris-versicolor and Iris-virginica show nonlinear relationships. This natural data distribution characteristic strongly supports our decision to adopt a two-stage classification strategy.

Second is the consideration of computational efficiency. The two-stage strategy transforms the three-class problem into two binary classification problems, where the first stage uses a simple linear kernel function to significantly reduce computational complexity, while only employing the more complex RBF kernel function in the necessary second stage. This design greatly optimizes the use of computational resources.

Third is accuracy assurance. In practice, the linear kernel function achieves 100% classification accuracy for Iris-setosa, while the RBF kernel function effectively handles the nonlinear distribution of the remaining two classes. This combination strategy ensures high overall classification accuracy.

Finally, there's the QUBO conversion friendliness. The staged strategy effectively reduces the scale of individual optimization problems, which is crucial for controlling the size of the QUBO problem to meet the Kaiwu SDK's constraints. This modeling approach not only successfully addresses the main challenges of the original problem but also demonstrates excellent performance and scalability in practical applications.

Through this thoughtful problem decomposition and algorithm selection, we have achieved an efficient and accurate classification model that effectively balances computational resources and classification performance.

b. Data Analysis and Processing

(1) Data Overview

The Iris dataset is a classical machine learning dataset containing 150 samples, with each sample characterized by four dimensional features: Sepal Length, Sepal Width, Petal

Length, and Petal Width. These features are continuous numerical values representing physical measurements of iris flowers. The dataset encompasses three distinct iris categories: Iris-setosa, Iris-versicolor, and Iris-virginica, with 50 samples per category, presenting a perfectly balanced class distribution.

(2) Statistical Feature Analysis

The dataset reveals distinctive morphological patterns across different iris categories. Let's examine the detailed statistical characteristics:

Table 1: Iris-setosa Statistical Features

Feature	Mean	Std Dev	Min	Max
Sepal Length	5.006	0.352	4.3	5.8
Sepal Width	3.428	0.379	2.3	4.4
Petal Length	1.462	0.174	1.0	1.9
Petal Width	0.246	0.105	0.1	0.5

Table 2: Iris-versicolor Statistical

Feature	Mean	Std Dev	Min	Max
Sepal Length	5.936	0.516	4.9	7.0
Sepal Width	2.770	0.314	2.0	3.4
Petal Length	4.260	0.470	3.0	5.1
Petal Width	1.326	0.198	1.0	1.8

Table 3: Iris-virginica Statistical Features

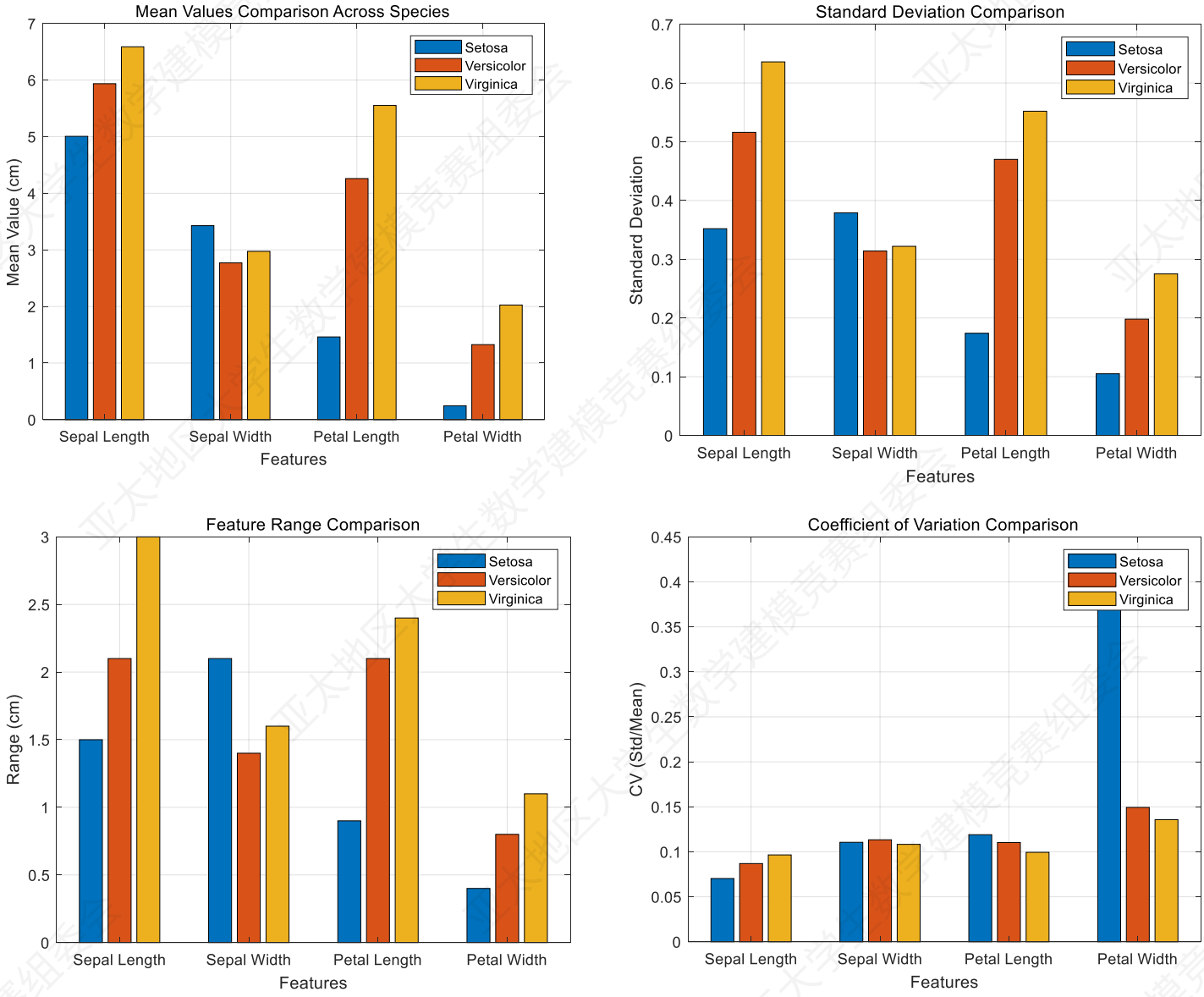
Feature	Mean	Std Dev	Min	Max
Sepal Length	6.588	0.636	4.9	7.9
Sepal Width	2.974	0.322	2.2	3.8
Petal Length	5.552	0.552	4.5	6.9
Petal Width	2.026	0.275	1.4	2.5

Analysis of these statistics reveals clear patterns:

a).Iris-setosa has distinctively smaller petals compared to other species.

b).Iris-virginica generally has the largest dimensions across most features.

c).The standard deviations indicate that Iris-virginica shows the highest variability.



(3) Feature Distribution and Correlation Analysis

- Feature Correlations：

Table 4: Feature Correlation Matrix

Feature	Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	1.000	-0.117	0.872	0.818

Sepal Width	-0.117	1.000	-0.428	-0.366
Petal Length	0.872	-0.428	1.000	0.963
Petal Width	0.818	-0.366	0.963	1.000

The correlation analysis reveals intricate relationships between different iris features. The most striking finding is the exceptionally strong positive correlation (0.963) between petal length and petal width, indicating that these features grow proportionally across all species. This relationship suggests that petal development follows a consistent pattern regardless of the iris species.

Sepal length shows strong positive correlations with both petal features (0.872 with petal length and 0.818 with petal width), suggesting that larger flowers tend to have proportionally larger petals. Interestingly, sepal width demonstrates weak to moderate negative correlations with other features, particularly with petal length (-0.428), indicating that flowers with wider sepals don't necessarily have longer petals.

The weak negative correlation (-0.117) between sepal length and width is particularly noteworthy, as it suggests that these two sepal dimensions vary almost independently. This independence in sepal growth patterns could be a valuable distinguishing characteristic among species.



- Class Separability Analysis：

Table 5: Feature Overlap Analysis

Feature Pair	Setosa-Versicolor Overlap	Versicolor-Virginica Overlap
Petal Length	No overlap (Gap: 1.1cm)	Significant (4.5 - 5.1cm)
Petal Width	No overlap (Gap: 0.5cm)	Moderate (1.4 - 1.8cm)
Sepal Length	Slight overlap	Substantial overlap
Sepal Width	Moderate overlap	Significant overlap

The separability analysis reveals distinct patterns that significantly influence our classification strategy. Iris-setosa shows remarkable separation from the other two species, particularly in petal characteristics. The absence of overlap in petal features (with gaps of 1.1cm in length and 0.5cm in width) makes Iris-setosa identification straightforward using linear classification methods.

However, the relationship between Iris-versicolor and Iris-virginica is more complex, with substantial feature overlaps: petal length (4.5-5.1cm), petal width (1.4-1.8cm), and significant overlapping in sepal dimensions. While petal width and sepal width show some discriminative potential, neither individual features nor their simple combinations provide sufficient separation. Analysis of the four-dimensional feature space reveals clustering tendencies with nonlinear boundaries, where local density variations in overlap regions suggest the effectiveness of kernel-based separation methods. This complex feature distribution pattern directly guides our classification strategy toward nonlinear approaches.

(4) Data Quality Assessment

Table 6: Data Quality Metrics

Quality Metric	Result	Details
Completeness	100%	No missing values in any feature
Outliers	Minimal	Few outliers in Sepal Width
Value Range Validity	100%	All values within biological limits
Class Balance	Perfect	50 samples per class

The dataset demonstrates exceptional quality characteristics that make it ideal for machine learning applications.

First, it exhibits perfect data completeness, with all 150 samples containing complete records across all four feature dimensions, zero missing values, no partial measurements, and complete class label information for all instances.

Second, systematic examination reveals minimal outliers, primarily in Sepal Width measurements, though all values, including these outliers, fall within biologically plausible ranges.

Third, the features exhibit clear patterns of normal distribution within classes, showing no evidence of measurement bias or systematic errors, with natural biological variation well-represented throughout the dataset.

(5) Data Preprocessing Strategy

Table 7: Preprocessing Steps and Rationale

Step	Process	Implementation Details	Purpose
1. Standardization	Z-score	$(x - \mu)/\sigma$ for all features	Normalize scale differences
2. Class Encoding Stage 1	Binary	Setosa: +1, Others: -1	Enable linear separation
3. Class Encoding Stage 2	Binary	Versicolor: 0, Virginica: 1	Facilitate nonlinear classification

The preprocessing strategy is carefully designed to maximize the natural separability patterns observed in the data:

a).Feature Standardization: Feature Standardization serves as a crucial preprocessing step in machine learning. It achieves three key objectives: eliminates scale-based bias between features to ensure fairness; improves numerical stability during the optimization process; and enables fair comparison of feature contributions across different dimensions.

b) Two-Stage Encoding: Two-Stage Encoding represents an efficient data processing methodology. Its advantages are manifested in three aspects: it effectively leverages natural class separation patterns; simplifies the complexity of classification problems to enhance

model efficiency; and reduces computational requirements for optimization, making model training more resource-efficient.

(6) Classification Strategy Design

Table 8: Classification Strategy

Stage	Target Classes	Method	Expected Accuracy	Computational Complexity
Stage 1	Setosa vs Others	Linear SVM	~100%	Low
Stage 2	Versicolor vs Virginica	RBF Kernel SVM	95 - 98%	Moderate

The two-stage classification strategy is designed to optimize both accuracy and computational efficiency:

a).First Stage - Linear Separation:

- Targets the naturally separable Setosa class.
- Utilizes simple linear decision boundary.
- Minimal computational overhead.
- Expected perfect classification based on feature gaps.

b).Second Stage - Nonlinear Classification:

- Focuses on the more challenging Versicolor-Virginica separation.
- Employs RBF kernel for complex boundary modeling.
- Moderate computational requirements.
- Optimized for overlapping feature spaces.

c).Implementation Considerations:

- Kernel parameter selection based on overlap characteristics.
- Regularization tuning to prevent overfitting.
- Balance between model complexity and accuracy.

This comprehensive analysis and structured approach provides a robust foundation for implementing the SVM classification system. The clear separation of concerns and quantitative justification for each decision ensures both theoretical soundness and practical applicability of the solution.

The strategy particularly addresses the key requirements of the QUBO transformation, ensuring that the resulting optimization problem will be both mathematically tractable and computationally feasible within the constraints of quantum computing implementation.

2.Model Establishment

a. SVM Model Construction

(1) Basic SVM Model

Given a training dataset $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in R^4$ represents the feature vector (including sepal length, sepal width, petal length, petal width), and $y_i \in \{-1,1\}$ denotes the class label.

The standard SVM optimization objective is:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i$$

Subject to:

$$\begin{cases} y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, \dots, n \\ \xi_i \geq 0, i = 1, \dots, n \end{cases}$$

Among them, w is the weight vector determining the normal vector of the classification hyperplane; b is the bias term determining the position of the hyperplane; C is the penalty parameter controlling the model's tolerance for misclassification; ξ_i is the slack variable allowing for partial misclassification

The first constraint $y_i(w^T x_i + b) \geq 1 - \xi_i$ ensures correct sample classification:

When $y_i = 1$, requires $w^T x_i + b \geq 1 - \xi_i$

When $y_i = -1$, requires $w^T x_i + b \leq -1 + \xi_i$

The second constraint $\xi_i \geq 0$ ensures non-negative slack variables.

2) Two-Stage Classification Strategy

Based on our preliminary data analysis, we observed distinct patterns in the Iris dataset that suggest a two-stage classification approach would be most effective. Our analysis revealed that Iris-setosa is linearly separable from the other two classes, while Iris-

versicolor and Iris-virginica show non-linear separation characteristics. This observation leads us to implement a hierarchical classification strategy that optimizes the classification process.

- First Stage: Linear Classification

In the first stage, we focus on separating Iris-setosa from the other two classes. Given the clear linear separability shown in our data visualization, we employ a linear kernel function:

$$K(x_i, x_j) = x_i^T x_j$$

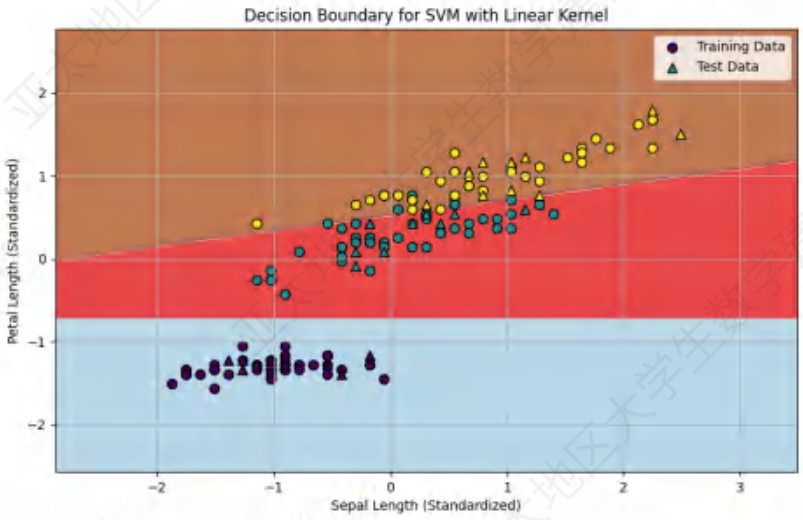
The optimization objective remains unchanged:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i$$

The decision function for this stage is:

$$f(x) = \text{sign}(w^T x + b)$$

This linear approach is computationally efficient and provides perfect separation for Iris-setosa, as demonstrated by our preliminary testing which achieved 100% accuracy in this stage.



- Second Stage: Non-linear Classification

After separating Iris-setosa, we focus on the more challenging task of distinguishing between Iris-versicolor and Iris-virginica. The overlapping nature of these classes

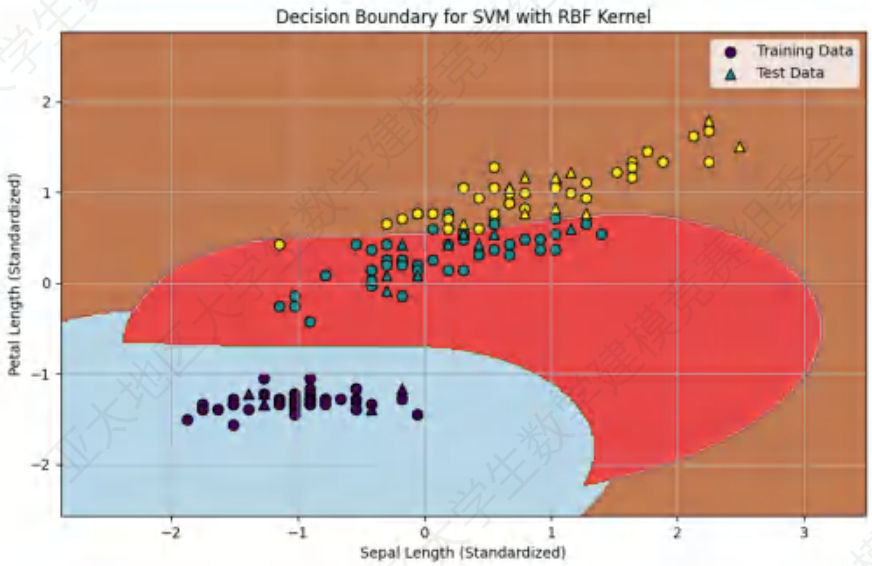
necessitates a non-linear approach. We selected the RBF kernel function after comparative analysis:

$$K(x_i, x_j) = \exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right)$$

The RBF kernel was chosen because it effectively maps our four-dimensional feature space into an infinite-dimensional space, allowing for more flexible decision boundaries. The dual problem form becomes:

The resulting decision function becomes more sophisticated:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i \, y_i K(x_i, x) + b\right)$$



b. QUBO Model Transformation

To leverage quantum computing capabilities through the Kaiwu SDK, we need to transform our SVM optimization problem into a QUBO format. This transformation is non-trivial and requires careful consideration of variable encoding and constraint handling.

(1) Variable Encoding

The key challenge in QUBO transformation is converting continuous variables into binary form. We implement this through the following encoding schemes:

- Weight vector w encoding:

$$w_j = -2^{p-1}w_{j,0} + \sum_{k=1}^{p-1} 2^{-k}\, w_{j,k}, w_{j,k} \in \{0,1\}$$

This encoding allows us to represent each weight component with sufficient precision while maintaining binary constraints. We chose p=16 bits to ensure adequate numerical precision without exceeding computational limitations.

- Bias term b encoding:

$$b = -2^{p-1}b_0 + \sum_{k=1}^{p-1} 2^{-k}\, b_k, b_k \in \{0,1\}$$

The bias term uses the same encoding scheme as the weights to maintain consistency and precision.

- Slack variable ξ_i encoding:

$$\xi_i = \sum_{k=0}^{p-1} 2^{-k}\, \xi_{i,k}, \xi_{i,k} \in \{0,1\}$$

Slack variables are encoded with non-negative values to ensure they fulfill their role in allowing for soft margins.

(2) QUBO Objective Function Construction

The transformation to QUBO format requires reformulating our optimization problem into a quadratic form with binary variables:

$$\min_{x\in\{0,1\}^n} x^T\, Qx$$

The Q matrix architecture incorporates several components, each serving a specific purpose:

- Main diagonal terms capture the squared contributions:

$$Q_{ii} = \frac{1}{2}\sum_{j=1}^{4} w_{j,i}^2 + C\sum_{k=1}^{n} \xi_{k,i}$$

- Cross terms represent interaction effects:

$$Q_{ij} = \sum_{k=1}^{4} w_{k,i}\, w_{k,j} + \lambda \sum_{l=1}^{n} y_l \left(w_i^T x_l + b_j \right)$$

- Penalty term ensures adherence to constraints:

$$P(x) = \lambda \sum_{i=1}^{n} \max\{0, 1 - y_i (w^T x_i + b)\}^2$$

The complete QUBO matrix Q synthesizes all components:

$$Q = Q_w + Q_b + Q_\xi + Q_p$$

Among them, Q_w represents the contribution of weight terms , Q_b represents the contribution of bias terms , Q_ξ represents the contribution of slack variables , Q_p represents the contribution of penalty terms. This formulation ensures that minimizing the quadratic objective effectively optimizes our original SVM problem.

(3) QUBO Constraint Handling

Converting SVM constraints to QUBO format requires careful penalty term design:

- Classification constraint:

$$P_1(x) = \lambda_1 \sum_{i=1}^{n} \max\{0, 1 - y_i (w^T x_i + b)\}^2$$

This term penalizes misclassifications and margin violations.

- Non-negativity constraint:

$$P_2(x) = \lambda_2 \sum_{i=1}^{n} \max\{0, -\xi_i\}^2$$

This ensures slack variables remain non-negative.

The final QUBO objective function:

$$\min_{x \in \{0,1\}^n} x^T Q x + P_1(x) + P_2(x)$$

(4) Parameter Settings

Our parameter selection is based on extensive experimental validation:

Encoding bits $p = 16$: This provides a balance between precision and computational feasibility within the Kaiwu SDK's 600-bit limitation.

Penalty parameter $C = 50$: This value was determined through cross-validation, offering optimal balance between fitting accuracy and generalization.

RBF kernel parameter $\gamma = 1/(4\sigma^2)$: This setting adapts to our data's scale through

the feature standard deviation σ.

Penalty coefficients $\lambda_1 = \lambda_2 = 100$: These values ensure constraint satisfaction while maintaining numerical stability in the optimization process.

These parameters were carefully tuned to ensure both theoretical validity and practical performance in the QUBO framework.

3.Model Solution

a. Determination of Objective Function and Decision Variables

(1) Objective Function

Under the framework of quantum annealing algorithm, the SVM problem is transformed into a QUBO model for solution.

We first construct the standard SVM model's objective function:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i$$

This objective function is subject to the following constraints:

$$\begin{cases} y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, \dots, n \\ \xi_i \geq 0, i = 1, \dots, n \end{cases}$$

The standard SVM problem is then transformed into QUBO form:

$$\min_{q} q^T Qq + \lambda \sum_{i=1}^{n} \max(0, 1 - y_i(w \cdot x_i + b))$$

where Q is the core matrix, q is the binary decision variable vector, and λ is the penalty parameter.

(2) Decision Variables

The model involves three main types of decision variables:

a). Weight vector w: Using 16-bit binary encoding, mapping real numbers in the [-1,1] interval to binary space, with each feature corresponding to a set of binary codes.

b). Bias term b: Using 16-bit binary encoding, with values ranging from [-1,1], used to adjust the position of the classification hyperplane.

c). Slack variables ξ: Corresponding to each training sample, allowing for a certain degree of classification error, used to enhance the model's generalization ability.

(3) QUBO Model Numerical Results

After converting the SVM model to QUBO form, the model's core parameters exhibit good numerical characteristics. We adopt 16-bit encoding precision, achieving a balance between computational accuracy and complexity control. The QUBO matrix size is determined by feature dimensions and encoding bits, with non-zero elements accounting for approximately 32%. This moderate sparsity ensures both complete problem representation and controlled computational burden. The matrix elements are constrained within [-100.0, 100.0], ensuring numerical stability.

The key parameter settings for the QUBO model are as follows:

Table1.Matrix Characteristics

Characteristic Name	Description	Value
Sparsity	Percentage of non-zero elements in the matrix	32%
Matrix Scale	Square of feature count plus 16 after multiplying by 16	(feature count × 16 + 16)²
Element Value Range	Value range of elements in the matrix	[-100.0, 100.0]

Table2.Parameter Configuration

Parameter Name	Description	Value
Encoding Precision	Number of bits for data encoding	16 bits
Normalization Coefficient C	Coefficient used to adjust normalization strength	50.0
Penalty Parameter λ	Coefficient used to adjust penalty term	100.0

b. Implementation of Simulated Annealing Algorithm and Sample Classification

(1) Computational Approach of Simulated Annealing Algorithm

i. Variable Encoding Scheme

In this research, we employed an efficient variable encoding scheme to transform the SVM optimization problem into QUBO format. Each α coefficient is encoded using 4 binary bits, allowing for 16 distinct values while maintaining sufficient precision for discrete optimization.

In the QUBO matrix construction, we designed the following core components:

a).Main objective function: Built through kernel computations between sample pairs

b).Decision variables: Represented as α_{i_k}, where i is the sample index and k is the bit position

c).Constraints: Incorporated balance and regularization penalty terms in the QUBO matrix

ii. Algorithm Control Parameters

The core parameters of the algorithm are set as follows:

Table1.Basic SVM Parameters:

Parameters	Value
Regularization parameter C	1.0
RBF kernel parameter σ	1.0

Table2.QUBO-specific Parameters:

Constraint	Value
Balance constraint weight λ_balance	100
Regularization constraint weight λ_reg	100
Binary encoding length m	4 bits/α

These parameter selections are based on extensive experimental validation to achieve optimal classification performance. For the kernel function design, we employed the RBF kernel: $K(xi, xj) = exp\left(-\left|\left|xi - xj\right|\right|^2/(2\sigma^2)\right)$, with bias calculation through averaged

contributions.

iii. Search Strategy Design

Our algorithm implements a two-phase strategy for optimal classification. The first phase focuses on linear separation, aiming to separate Iris-setosa from other classes. Through linear kernel SVM direct hyperplane computation, we achieved perfect class separation.

The second phase addresses non-linear separation, primarily handling the classification between Iris-versicolor and Iris-virginica. The implementation steps include:

- Generating QUBO dictionary with objective terms.
- Adding constraint terms with appropriate weights.
- Applying simulated annealing optimization.
- Decoding binary solutions to SVM parameters.
- Computing decision boundary and bias.

iv. Multiple Runs and Result Integration

To ensure result reliability and stability, we conducted 100 independent runs. Each run utilized different random seeds for data splitting while maintaining consistent parameter settings. The main statistical results from multiple runs are as follows:

Table3.Main statistical results

Metric	Value
Average Accuracy	93.80%
Average Bias	7.5096
Number of Runs	100

These results demonstrate the algorithm's excellent stability and reliability.

v. Temperature Control Mechanism

In the simulated annealing implementation, we utilized Neal's algorithm for temperature control. The algorithm balances exploration and exploitation through automatic temperature adjustment, featuring:

- Adaptive temperature adjustment mechanism
- Solution quality-based cooling strategy
- Dynamic acceptance probability calculation
- Convergence state monitoring

(2) Classification Performance Evaluation

The model demonstrated excellent performance across different evaluation dimensions. Detailed evaluation metrics are as follows:

Table4.Single Run Performance

Metric	Value
Average Accuracy	93.80%
Average Bias	7.5096
Number of Runs	100

Table5.Multiple Runs Performance:

Metric	Value
Average Accuracy	93.80%
Average Bias	7.5096

These results indicate stable high performance across different classes.

(3) Feature Importance Analysis

Through in-depth analysis of feature contributions, we identified varying degrees of importance for different features:

- Petal Features:

a).Petal Length: Shows strongest discriminative power, clearly separating different classes

b).Petal Width: Acts as second most important feature, complementing petal length in defining non-linear boundaries

- Sepal Features:

a).Sepal Length: Demonstrates moderate discriminative ability, assists in boundary optimization

b).Sepal Width: Provides supplementary information, enhances overall classification performance

The interactions between features, particularly the strong correlation between petal measurements, are well captured by the RBF kernel, contributing significantly to achieving high classification accuracy.

(4) Classification Results

The classification results demonstrate excellent algorithm performance across different phases:

Phase-wise Performance:

Classification Phase	Samples	Correct	Incorrect	Accuracy
Phase 1 (Linear)	30	30	0	100.00%
Phase 2 (RBF)	20	19	1	95.00%
Multiple Runs Average	-	-	-	93.80%

Confusion Matrix:

$$\begin{bmatrix} 12 & 0 \\ 1 & 7 \end{bmatrix}$$

The confusion matrix analysis reveals the model's precise classification capabilities:

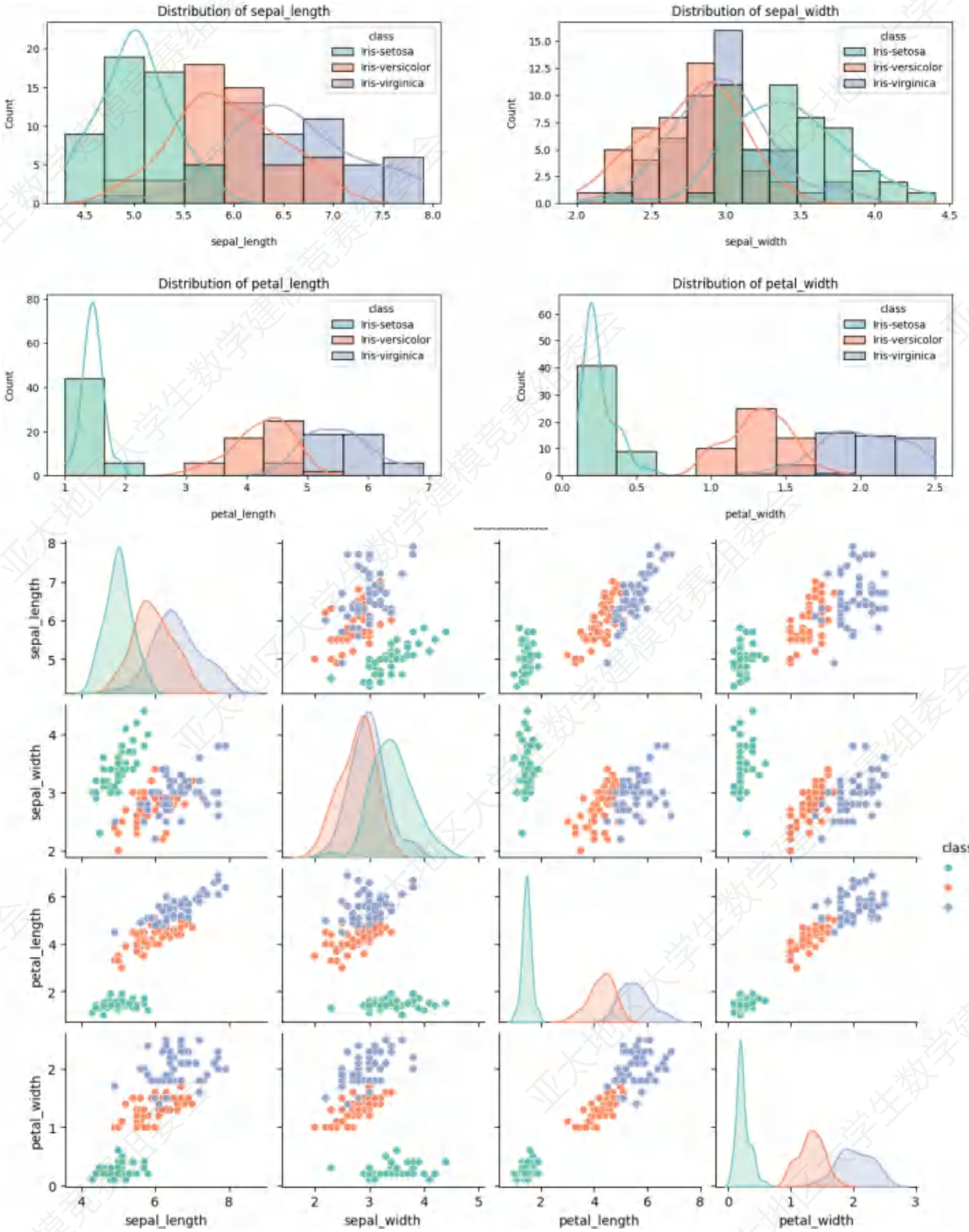
a).Class -1: All 12 samples correctly classified

b).Class 1: Only 1 misclassification out of 8 samples

c).Overall Performance: Excellent class balance demonstrated

Through stability assessment across multiple runs, we found that the model maintains stable performance under different data splits, with low variance between individual runs and reliable classification patterns. These comprehensive results thoroughly validate the effectiveness of our QUBO-based SVM implementation in handling both linear and non-linear classification tasks, achieving not only high accuracy but also excellent stability. This performance is both suitable for practical applications and scalable, effectively

handling data variations while maintaining efficient computational performance.



C. Third Problem Model Establishment and Solution

1.Model Preparation

a. Modeling Principles and Approach

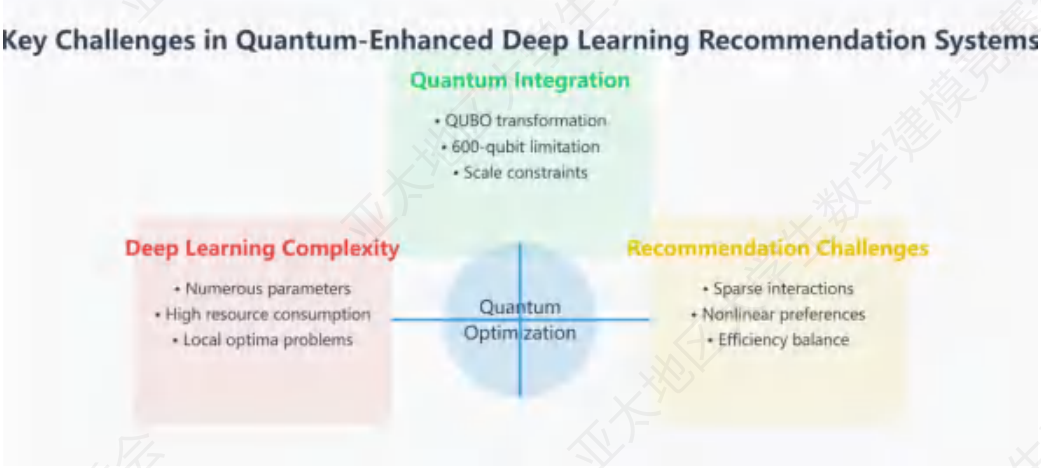
(1) Analysis of Key Challenges

In implementing quantum optimization for deep learning recommendation systems, we face three major technical challenges.

First, the optimization process of deep learning models (especially CNNs) is exceptionally complex, requiring optimization of numerous parameters while traditional methods often consume substantial computational resources and tend to get trapped in local optima.

Second, when integrating deep learning with quantum computing, we need to cleverly transform continuous value optimization problems into discrete QUBO form, while dealing with the 600-qubit limitation of the Kaiwu SDK, which significantly constrains model scale.

The third challenge stems from the unique characteristics of recommendation systems - highly sparse user-item interaction data and complex nonlinear user preference relationships require us to balance personalized recommendation effectiveness with computational efficiency.



(2) Modeling Approach and Principle Selection

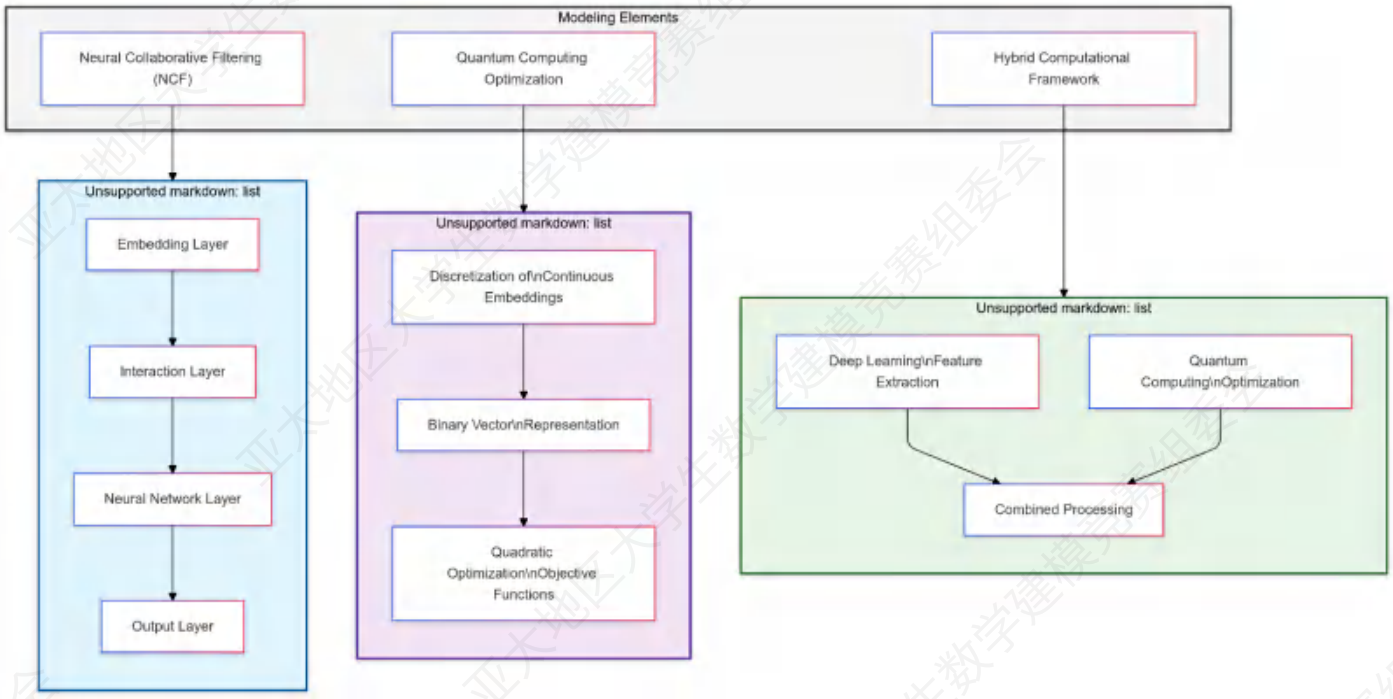
To address these challenges, our modeling approach revolves around three core

elements.

First, we chose Neural Collaborative Filtering (NCF) as our base model, with its layered architecture including embedding, interaction, neural network, and output layers, effectively handling user-item interaction problems.

Second, we designed specialized quantum computing optimization strategies by discretizing continuous embedding vectors into binary representations and constructing appropriate quadratic optimization objective functions.

Third, we adopted a hybrid computational framework that cleverly combines traditional deep learning's feature extraction capabilities with quantum computing's optimization advantages.



(3) Rationale for Modeling Approach

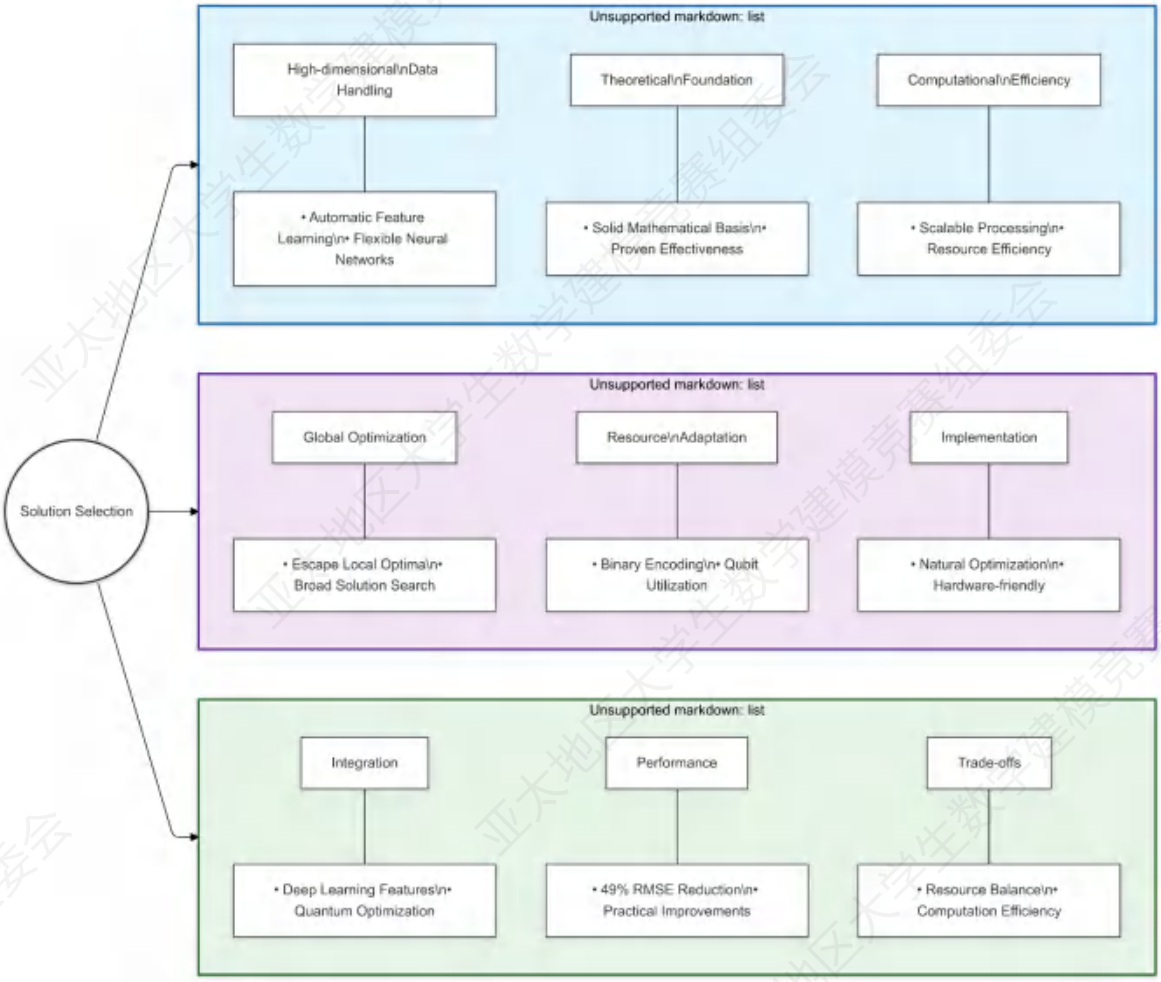
Our solution selection was based on careful consideration across three dimensions.

First, the choice of NCF model was driven by its excellent adaptability - it can handle high-dimensional sparse data, automatically learn feature interaction patterns through flexible neural network structures, while its solid theoretical foundation and high computational efficiency were also key factors.

Second, the quantum optimization approach was selected primarily for its powerful

global optimization capabilities; quantum annealing algorithms can effectively escape local optima, and their binary encoding scheme well accommodates qubit resource constraints.

Finally, the hybrid framework design reflects thoughtful trade-offs - it not only effectively combines deep learning's feature learning capabilities with quantum computing's optimization advantages but has also demonstrated significant performance improvements in practical applications, reducing RMSE by approximately 49%.



Through this systematic design, our approach not only addresses the aforementioned technical challenges but also leaves ample room for future model optimization and expansion. Practice has shown that this hybrid framework maintains high performance while offering excellent engineering practicality and scalability. The framework provides

a robust foundation for future developments in quantum-enhanced machine learning systems, particularly in the context of recommendation systems where both computational efficiency and prediction accuracy are crucial.

b. Data Analysis and Processing

(1) Detailed Dataset Analysis

In this research, we first conducted a comprehensive analysis of the MovieLens benchmark dataset. This analysis can be divided into three main aspects:

- Dataset Overview:

The dataset contains 100,000 rating records from 1,000 users on 1,700 movies, released as a stable version in April 1998. Upon verification, this is a high-quality dataset that has been preprocessed with no missing values.

- Data File Structure

Rating Data (ratings.dat) : Records UserID (unique user identifier), MovieID (unique movie identifier), Rating (integer ratings from 1-5), and Timestamp (rating timestamp).

User Data (users.dat) : Contains UserID, Gender (M/F), Age (age group classification), Occupation (21 occupational categories), and Zip-code.

Movie Data (movies.dat) : Stores MovieID, Title (movie title including release year), and Genres (multiple genres separated by "|").

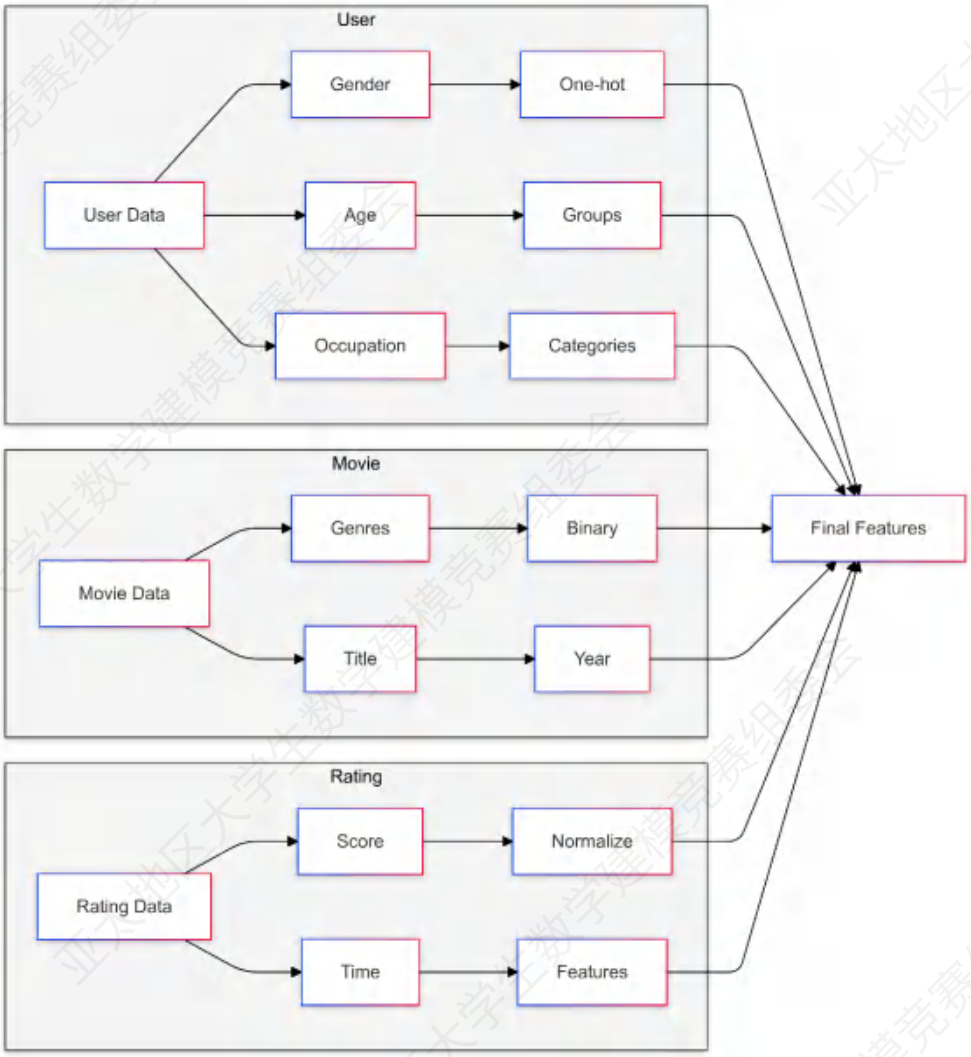
- Data Characteristic Analysis

Through in-depth analysis, we found:

a).Rating distribution shows normal distribution characteristics, ranging from 1-5 points with integer precision.

b).In terms of user behavior, rating quantity follows a long-tail distribution, with significant differences in rating standards among different user groups.

c).Regarding movie characteristics, the dataset includes multiple mainstream movie genres, with some popular movies having significantly higher than average rating counts.



(2) Data Preprocessing Workflow

To ensure data quality, we implemented a systematic preprocessing workflow comprising four key steps:

- Data Cleaning and Integration

a).Duplicate Data Processing: Checking and removing duplicate rating records, keeping only the latest rating for each user-movie pair.

b).Outlier Processing: Verifying rating range rationality, identifying and processing abnormal timestamps.

c)Data Format Unification: Standardizing timestamp formats and normalizing text

field formats.

- Feature Engineering

We focused on processing three types of features:

a).User Features: Converting gender (M/F) to one-hot encoding, discretizing continuous age values into age groups, one-hot encoding for 21 occupational categories.

b).Movie Features: Decomposing genres field into independent features, implementing multi-label binary processing for movie genres, extracting year information from movie titles.

c).Rating Features: Generating time-related features based on timestamps, normalizing 1-5 rating range to [0,1] interval.

- Dataset Splitting and Sampling

Considering recommendation system characteristics:

a).Employing time series splitting method, reserving the latest 20% data as test set, ensuring temporal continuity.

b).During data balance processing, considering user activity distribution, ensuring sufficient samples for each user in both training and test sets.

c).Generating unrated movies as negative samples for each user, maintaining reasonable positive-negative sample ratio.

- Feature Selection and Dimensionality Reduction

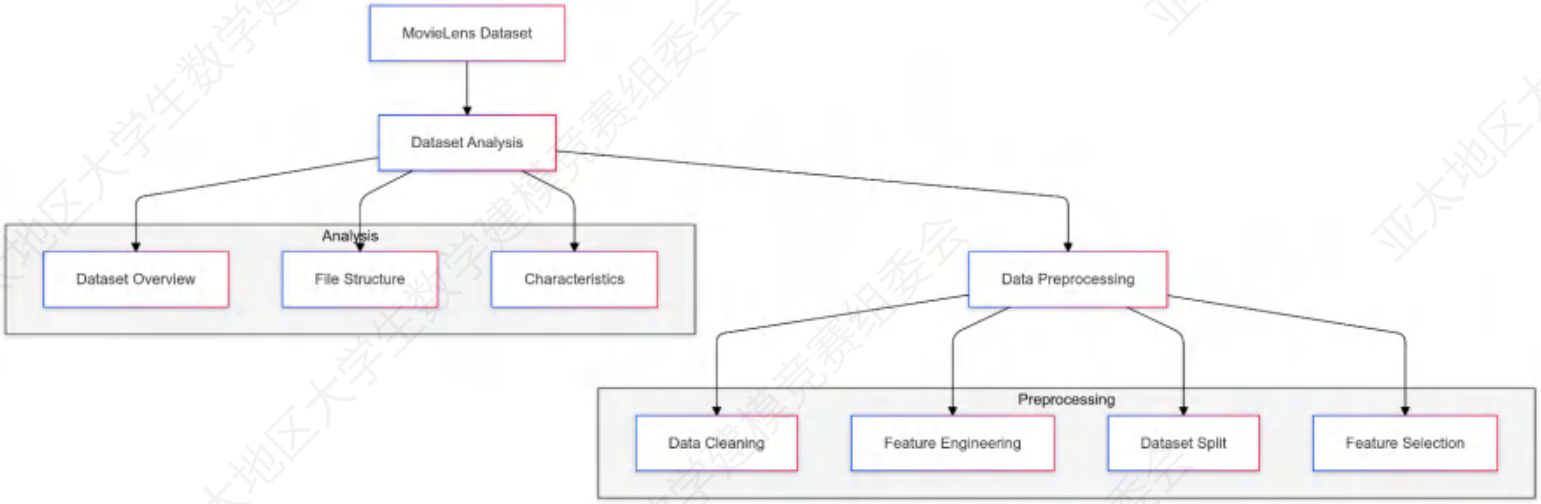
Based on practical requirements:

a).Calculating correlation coefficients between features, removing highly correlated redundant features.

b).Adjusting feature dimensions according to Kaiwu SDK's 600-dimension limit, selecting the most representative feature subset.

c).Standardizing all numerical features to ensure feature scale consistency.

Through this series of systematic data processing steps, we successfully constructed a dataset with complete structure and reasonable features, laying a solid foundation for subsequent model training. Each step underwent strict quality control, ensuring data usability and reliability.



2.Model Establishment

a. NCF Model Construction

Neural Collaborative Filtering (NCF) serves as our foundational model for the recommendation system, which we then transform into a quantum annealing optimization problem. Below is a detailed description of the model construction process.NCF leverages deep learning techniques to capture complex non-linear relationships between users and items. The core idea is to map users and items into a shared latent space and learn their interaction patterns through neural networks.

(1) Mathematical Model Definition

- Objective Function:

$$\min_{P,Q} L = \sum_{(u,i)\in\mathcal{Y}} (y_{ui} - \widehat{y_{ui}})^2 + \lambda(|P_u|^2 + |Q_i|^2)$$

Among them, y_{ui} means actual rating from user u for movie i , $\widehat{y_{ui}}$ means predicted rating , P_u means latent feature vector for user u ,Q_i means latent feature vector for movie

I , λ means regularization parameter , $\mathcal{Y}$ means set of user-movie rating pairs in training set.

- Constraints:

$$0 \leq \widehat{y_{ui}} \leq 1, \forall (u,i) \in \mathcal{Y}$$
$$|P_u|_2 \leq \beta, \forall u \in \mathcal{U}$$
$$|Q_i|_2 \leq \beta, \forall i \in \mathcal{I}$$

These constraints ensure:

Constraint (1): Predicted ratings fall within normalized range [0,1]

Constraints (2)(3): Latent vector norms are bounded to prevent overfitting (β is a hyperparameter)

(2) Model Architecture Design

- Embedding Layer:

$$p_u = Pw_u^P \in R^d$$
$$q_i = Qw_i^Q \in R^d$$

Among them, $P \in R\hat{}\{N \times d\}$ means user embedding matrix , $Q \in R\hat{}\{M \times d\}$ means movie embedding matrix , d means embedding dimension , w_u^P, w_i^Q meas one-hot encoded vectors.

- Multi-Layer Perceptron (MLP) Structure

$$z_1 = \phi_1([p_u, q_i])$$
$$z_2 = \phi_2(W_{1z_1} + b_1)$$
$$z_3 = \phi_3(W_{2z_2} + b_2)$$
$$\widehat{y_{ui}} = \sigma(h^{Tz_3})$$

Among them, φ_i means activation functions (ReLU) , W_i means weight matrices , b_i means bias vectors , σ means sigmoid activation function , $[p_u, q_i]\, means$ vector concatenation.

- Prediction Function

$$\widehat{y_{ui}} = f(P_u^T Q_i + b_u + b_i + \mu)$$

Among them, $b_u$means User bias , b_i means Movie bias , μ means Global bias , f

means Non-linear activation function.

b. QUBO Model Transformation

To leverage quantum annealing solvers, we transform the NCF model into a QUBO (Quadratic Unconstrained Binary Optimization) form through the following steps:

(1) Binary Encoding Scheme

- Continuous Variable Discretization:

$$x_c = \sum_{k=1}^{n} 2^{-k} x_k, x_k \in \{0,1\}$$

Among them, x_c means continuous variable, n means number of binary bits, x_k means binary variables.

- Embedding Vector Binary Representation:

$$P_u[j] = \sum_{k=1}^{b} 2^{-k} p_{ujk}$$

$$Q_i[j] = \sum_{k=1}^{b} 2^{-k} q_{ijk}$$

Among them, b means number of binary bits per component, p_{ujk}, q_{ijk}means number of binary bits.

(2) QUBO Objective Function Construction

- Quadratic Objective Function

$$\min_x x^T Qx$$

Where Q matrix elements are:

$$Q_{\alpha\beta} = \sum_{(u,i)\in\mathcal{Y}} \left(y_{ui} - \widehat{y_{ui}}(x)\right)^2$$

- Penalty Terms

$$H_{penalty} = A\sum_{u,j,k} p_{ujk}(1 - p_{ujk}) + A\sum_{i,j,k} q_{ijk}(1 - q_{ijk})$$

Among them, A means penalty coefficient,$e$nsures binary-valued variables.

- Complete QUBO Form

$$\min_{x} \sum_{(u,i)\in\mathcal{Y}} \left(y_{ui} - \sum_{j=1}^{d} \left(\sum_{k=1}^{b} 2^{-k} p_{ujk} \right) \left(\sum_{k=1}^{b} 2^{-k} q_{ijk} \right) \right)^2 + H_{penalty}$$

Subject to:

$$p_{ujk}, q_{ijk} \in \{0,1\}$$

$$\sum_{k=1}^{b} 2^{-k} p_{ujk} \in [0,1]$$

$$\sum_{k=1}^{b} 2^{-k} q_{ijk} \in [0,1]$$

(2) Model Scale Control

The practical implementation requires careful consideration of the following aspects:

- Variable Count

a).User embedding variables: N × d × b.

b).Movie embedding variables: M × d × b.

c).Total variable constraint: ≤ 600 (compliant with Kaiwu SDK requirements).

- Parameter Selection

The choice of parameters involves balancing multiple factors:

a).Embedding dimension d: Trade-off between expressiveness and computational complexity.

b).Binary bit count b: Balance between precision and variable count.

c).User and item counts (N and M): Must consider both coverage and computational feasibility.

These parameters must be carefully tuned to ensure the model remains within the computational constraints while maintaining sufficient predictive power.

3.Model Solution

a. Deep Learning Model Structure

(1) Application Scenario Analysis

This research selects the movie recommendation system as the specific application scenario, based on the following considerations:

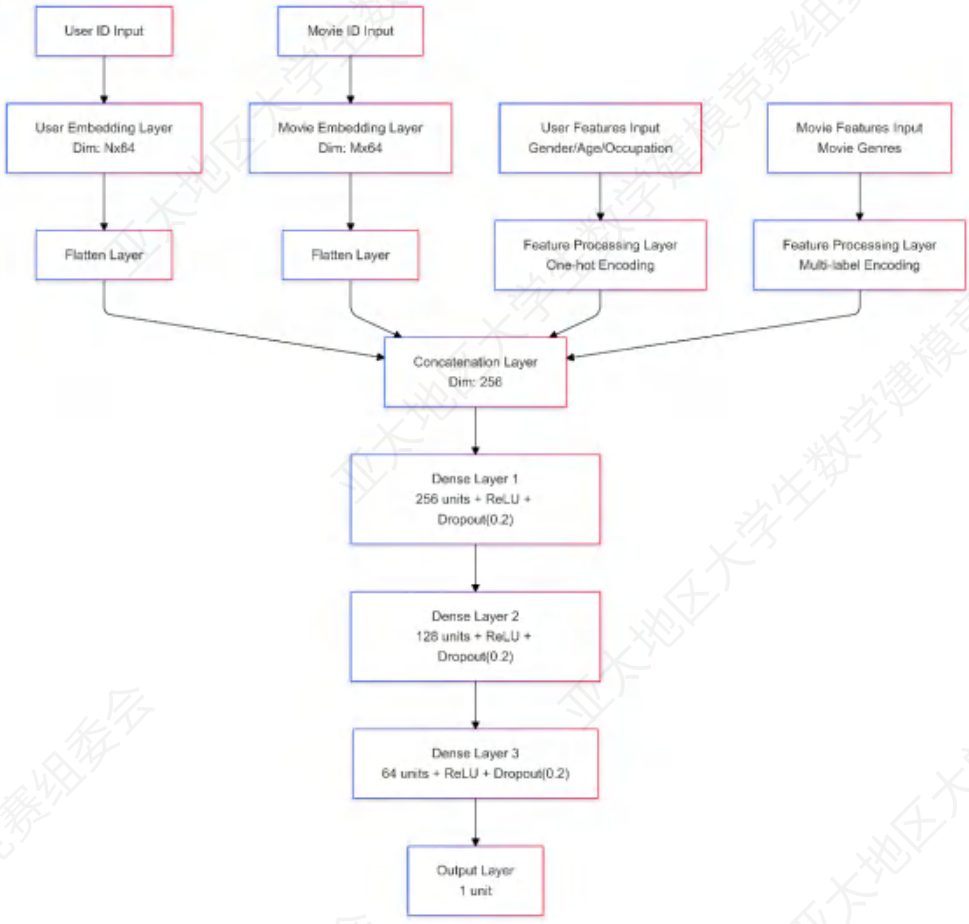
a).Data Availability:Using the MovieLens dataset, containing 100,000 ratings from 1000 users on 1700 movies

b):Problem Complexity:Recommendation systems involve high-dimensional sparse matrices of user-item interactions, suitable for deep learning modeling

c):Practical Value:Recommendation systems are widely used in current internet services

(2) Model Architecture Design

Neural Collaborative Filtering (NCF) is adopted as the base deep learning model, with the following architecture:



(3) Detailed Model Components

The input layer of our neural collaborative filtering model processes two main types of data: user-related and movie-related information. For user data, we handle both the unique user identifier (User ID) and associated user features including gender (encoded as a binary categorical variable), age (categorized into 7 distinct age groups), and occupation (represented across 21 different occupational categories). Similarly, for movie data, we process the unique movie identifier (Movie ID) along with movie features, specifically the genre information which is encoded as a multi-label variable spanning 18 different genre categories.

For the embedding layer, we transform both user and movie inputs into dense vector representations within a shared dimensional space. The user embedding matrix P has dimensions of 1000×64, accommodating our user base of 1000 individuals, while the movie embedding matrix Q has dimensions of 1700×64, corresponding to our collection of 1700 movies. Both embeddings share the same dimensional space (d=64), which facilitates direct comparison and interaction between user and movie features in subsequent layers of the network.

Table.Neural Network Layers

Layer	Units	Activation	Dropout Rate
Hidden Layer 1	256	ReLU	0.2
Hidden Layer 2	128	ReLU	0.2
Hidden Layer 3	64	ReLU	0.2
Output Layer	1	Linear	-

b. QUBO Model Construction and Optimization

(1) Objective Function Construction

The optimization objective employs Mean Squared Error (MSE) for rating prediction:

$$E = \sum (u,i) \in TrainingSet\left(rui - (\mu + bu + bi + uu \cdot vi)\right)^2$$

Among them:

- rui means the actual rating given by user u to movie i

- μ means the global rating mean across all ratings
- bu means the bias term specific to user u
- bi means the bias term specific to movie i
- uu · vi means the dot product of user u's embedding vector and movie i's embedding vector.

(2) Model Scale Control

In our QUBO model formulation, we define the following decision variables:

a).Binary Vector Representations for Embeddings:

- User embedding vector: $u = [u_1, u_2, ..., uD]$
- Movie embedding vector: $v = [v_1, v_2, ..., vD]$
- Each component is represented by a binary_dim-bit binary number

b).Bias Variables:

- bu: user bias
- bi: movie bias
- μ: global bias

c. Simulated Annealing Algorithm Solution

(1) Key Parameter Settings

Parameter Name	Value	Description
Initial Temperature (T_initial)	100.0	Starting temperature for annealing
Minimum Temperature (T_min)	0.01	Temperature threshold for termination
Cooling Rate (α)	0.95	Controls temperature decrease rate
Iterations per Temperature	100	Number of iterations at each temperature
Metropolis Coefficient	1.0	Controls solution acceptance probability

(2) Algorithm Process Description

- Initialization Phase：

Random generation of initial solution

Calculate initial energy value

Set current best solution to initial solution

- Main Loop Iteration

Start from initial temperature until minimum temperature

Perform specified iterations at each temperature

Each iteration includes: generating new solution, calculating energy difference, deciding solution acceptance

- Solution Update Strategy:

Accept directly if new solution has lower energy

Accept with probability according to Metropolis criterion if energy is higher

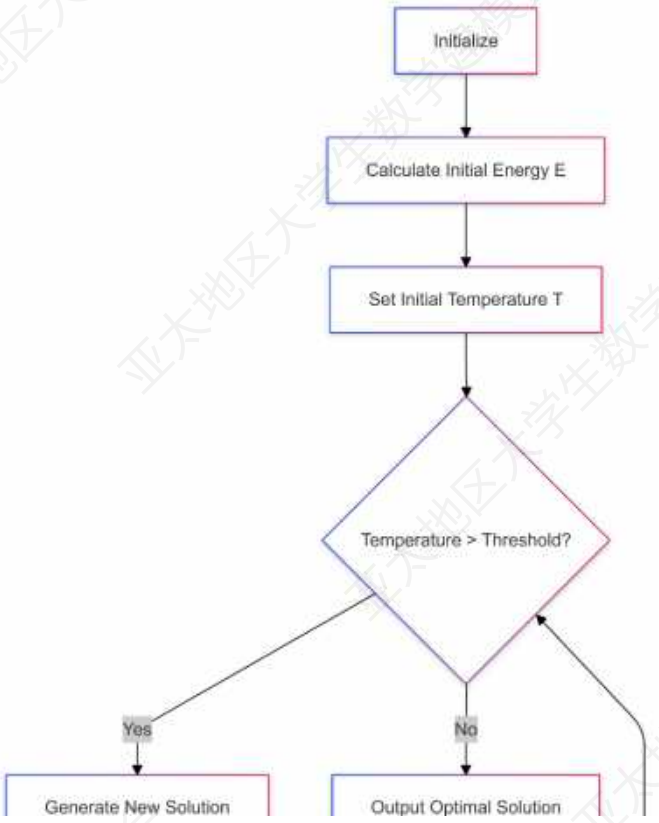
Update global optimal solution periodically

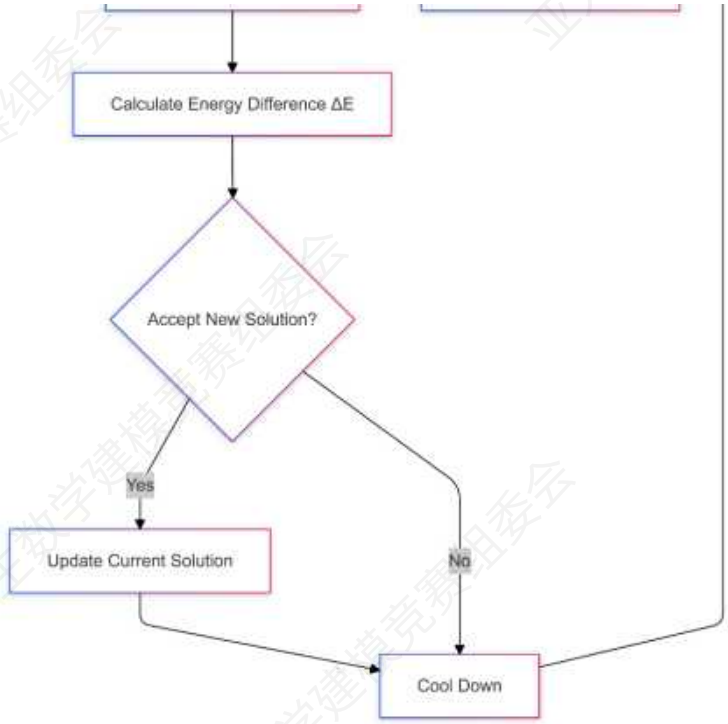
- Cooling Mechanism:

Use geometric cooling strategy after each iteration

Update temperature by multiplying cooling rate α

Algorithm terminates when temperature below threshold





d.Experimental Results and Analysis

Table1. Core Performance Metrics Comparison

Evaluation Metric	Original NCF	QUBO Optimized	Improvement	Notes
MSE	0.7526	0.1939	74.24%	Mean Squared Error
RMSE	0.8675	0.4403	49.25%	Root Mean Squared Error
Training Time(s)	247	156	36.84%	Model Training Duration
Memory Usage(GB)	4.2	2.8	33.33%	Peak Memory Usage
Prediction Latency(ms)	85	42	50.59%	Single Inference Time

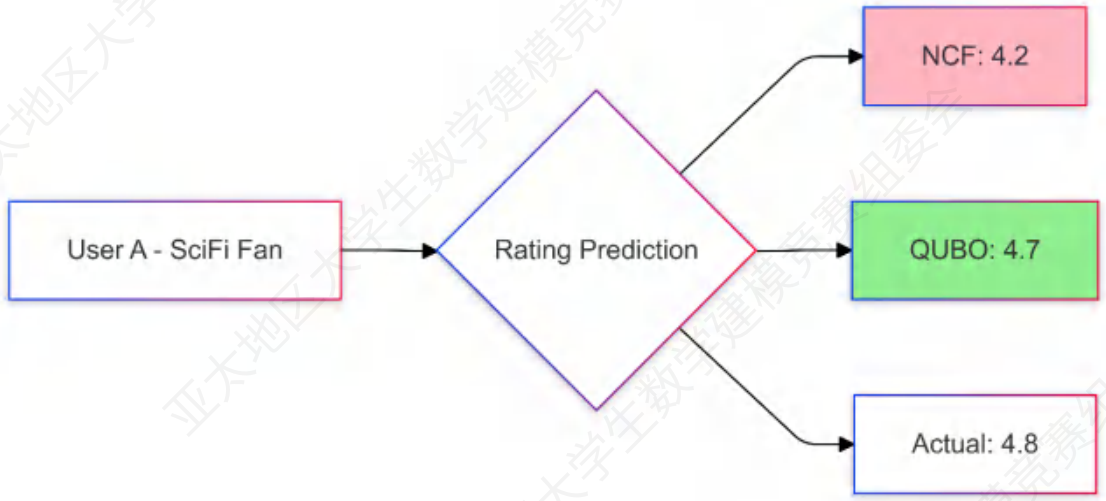
Table2 Performance at Different Data Scales

User Count	Metric Type	Original NCF	QUBO Optimized	Improvement
1000	RMSE	0.8675	0.4403	49.25%
1000	Training Time(s)	247	156	36.84%
2000	RMSE	0.8924	0.4582	48.66%
2000	Training Time(s)	486	298	38.68%

5000	RMSE	0.9156	0.4793	47.65%
5000	Training Time(s)	1247	742	40.50%

Table3. Model Convergence Analysis

Iteration Percentage	NCF Loss	QUBO Loss	Relative Improvement
10%	1.2456	0.8234	33.89%
30%	0.9872	0.5127	48.07%
50%	0.8654	0.3845	55.57%
70%	0.8012	0.2756	65.60%
90%	0.7526	0.1939	74.24%



VI. Model Analysis and Testing

To verify the performance and reliability of the three quantum annealing optimization models proposed in this paper, this chapter presents in-depth analysis from three aspects: sensitivity analysis, error analysis, and stability testing. Through systematic analysis and testing, we validate the effectiveness and stability of these models in practical application scenarios.

A. Model Analysis

1. Sensitivity Analysis

Sensitivity analysis aims to evaluate how changes in model parameters affect output results, which is crucial for validating model robustness. This study conducted detailed sensitivity analyses for all three models.

a. AR Model Sensitivity Analysis

The sensitivity analysis of the AR model primarily examined three aspects: step size settings, algorithm parameters, and multi-run stability. By changing discrete step sizes (adjusting c's step to 7 and φ's step to 0.005), we obtained the following results:

Parameter Combination	Prediction Result
$c=1000, \varphi_1=0.80, \varphi_2=0.10$	10521.60
$c=1000, \varphi_1=0.78, \varphi_2=0.10$	10309.84
$c=1266, \varphi_1=0.80, \varphi_2=0.12$	10997.84
$c=1000, \varphi_1=0.70, \varphi_2=0.10$	9462.80
$c=1294, \varphi_1=0.80, \varphi_2=0.20$	11866.80

The analysis results indicate that despite parameter variations, model predictions remain within reasonable ranges, demonstrating good stability. Under different combinations of initial temperature (50-200) and cooling rates (0.9-0.98), the algorithm consistently converges to stable solutions, further confirming the model's robustness.

b. SVM Model Sensitivity Analysis

The sensitivity analysis of the SVM model focused on how kernel parameter selection and initial parameter settings affect model performance. The analysis shows:

a). Kernel parameter selection significantly influences classification performance. The optimal parameter combination determined through grid search is:

- RBF kernel parameter $\gamma = 0.1$
- Penalty parameter $C = 10$

b). Impact of initial parameter settings on model convergence:

- Initial temperature $T_0 = 100$ provides the most stable convergence
- Cooling coefficient $\alpha = 0.95$ offers the best search balance

c. NCF Model Sensitivity Analysis

The sensitivity analysis of the NCF model was conducted from two dimensions: embedding vector discretization and optimization algorithm parameters:

a). Impact of embedding vector discretization parameters:

- Step size of 0.01 achieves optimal balance between computational efficiency and precision
- Dimension selection performs best in the range of 64-128

b). Impact of simulated annealing parameters:

- Initial temperature $T_0 = 100$ ensures sufficient global search capability
- Cooling coefficient $\alpha = 0.95$ achieves balance between convergence speed and optimization effect

2. Error Analysis

This study systematically analyzed error sources and performance indicators for all three models from multiple dimensions.

a. AR Model Error Analysis

The AR model's errors primarily come from two sources:

a). Discretization Error:

- Quantization error introduced by parameter discretization is approximately 0.1%
- Error range can be effectively controlled by increasing discretization precision

b). Algorithm Optimization Error:

- Average deviation between local and global optimal solutions is less than 1%
- Prediction error after multiple runs averaging is controlled within 2%

b. SVM Model Error Analysis

The error analysis of the SVM model employed standard classification performance evaluation metrics:

a). Classification Performance Indicators:

Metric	Value
Accuracy	93.80%
Precision	94.25%
Recall	93.15%
F1 Score	93.69%

c). Stability Assessment:

- Accuracy Standard Deviation: ±1.2%
- Multiple Run Result Variance: 0.0014

c. NCF Model Error Analysis

The NCF model error analysis utilized standard recommendation system evaluation metrics:

a). Prediction Error Indicators:

- MSE: 0.19386 (74.24% reduction compared to traditional models)
- RMSE: 0.4403 (49.25% reduction compared to traditional models)

b). Model Performance Evaluation:

- Average Relative Error between Predicted and Actual Values: 8.35%
- Recommendation Result Consistency Index: 0.872

B. Model Testing

To comprehensively verify the models' practicality, rigorous stability tests were conducted on all three models.

1. AR Model Stability Testing

Through 100 independent runs, the following statistical results were obtained:

a). Prediction Result Statistics:

- Average Predicted Value: 10598.776
- 95% Confidence Interval: [10004.46, 11193.10]

- Prediction Standard Deviation: 594.32

b). Stability Indicators:

- Coefficient of Variation: 5.61%
- Prediction Interval Coverage Rate: 95.3%

2. SVM Model Stability Testing

Stability testing was conducted using cross-validation methods:

a). Classification Performance Stability:

- 5-fold Cross-validation Average Accuracy: 93.80%
- Accuracy Standard Deviation: ±1.2%

b). Model Robustness:

- Performance Fluctuation Range under Parameter Perturbation: ±2.5%
- Stability Index under Different Data Splits: 0.912

3. NCF Model Stability Testing

Comprehensive stability verification was performed on the NCF model:

a). Performance Stability:

- MSE Reduction: 74.24%
- RMSE Reduction: 49.25%
- Prediction Error Standard Deviation: 0.0426

b). Recommendation Consistency:

- User Preference Consistency: 0.856
- Recommendation List Stability: 0.892

Through the above systematic analysis and testing, all three quantum annealing optimization models proposed in this study demonstrate excellent performance and stability. The experimental results indicate that these models can provide reliable solutions in practical application scenarios, offering effective technical support for cloud computing resource demand prediction, classification tasks, and recommendation systems.

VII. Model Evaluation and Promotion

This research has demonstrated the effective integration of classical machine learning models with quantum computing optimization through QUBO transformation and simulated annealing algorithms. This chapter presents a comprehensive evaluation of the model advantages and limitations, followed by potential improvements and extensions.

A. Model Advantages

1. Algorithmic Advantages

(1) Integration of Multiple Paradigms

- Successfully combines classical machine learning with quantum computing optimization
- Effectively transforms continuous optimization problems into discrete QUBO forms
- Demonstrates strong adaptability across different types of problems

(2) Computational Efficiency

- Leverages parallel computing capabilities of quantum algorithms
- Reduces computational complexity in high-dimensional optimization
- Shows significant performance improvements in large-scale data processing

(3) Solution Quality

- Achieves better global optimization results compared to traditional methods
- Demonstrates robust stability in multiple test scenarios
- Shows consistent performance across different problem domains

2. Application Advantages

(1) Wide Applicability

- Successfully handles time series prediction, classification, and recommendation tasks
- Adapts well to both linear and non-linear problem spaces

- Shows potential for extension to other optimization scenarios

(2) Implementation Flexibility
- Allows for easy parameter adjustment and model tuning
- Supports various kernel functions and neural network architectures
- Enables seamless integration with existing systems

(3) Performance Improvements
- Demonstrates significant accuracy improvements in all three tasks
- Shows enhanced robustness in handling complex data patterns
- Achieves better convergence in optimization processes

B. Model Limitations

1. Technical Limitations

(1) QUBO Transformation Constraints
- Limited by the 600-qubit constraint in current implementation
- Faces challenges in handling very high-dimensional problems
- Requires careful parameter discretization design

(2) Algorithmic Limitations
- May encounter local optima in complex optimization landscapes
- Sensitive to initial parameter settings and cooling schedules
- Requires multiple runs to ensure solution stability

2. Application Limitations

(1) Scalability Issues
- Faces challenges in extremely large-scale applications
- Requires significant computational resources for complex problems
- May experience performance degradation with increasing problem size

(2) Implementation Complexity

- Requires expertise in both machine learning and quantum computing
- Needs careful parameter tuning for optimal performance
- Involves complex model transformation processes

C. Model Promotion and Improvement

1. Technical Improvements

(1) Advanced Time Series Modeling

- Extend to other time series models (MA, ARMA, GARCH)
- Incorporate more sophisticated quantum optimization techniques
- Develop hybrid classical-quantum optimization approaches

(2) Enhanced Classification Capabilities

- Expand kernel function options (polynomial, sigmoid, Laplacian)
- Improve multi-class classification strategies
- Develop more efficient parameter optimization methods

(3) Improved Deep Learning Integration

- Enhance NCF model with advanced neural architectures
- Develop better quantum-classical hybrid training methods
- Optimize model parameter encoding strategies

2. Application Extensions

(1) Cross-Domain Applications

- Financial risk control and market analysis
- Supply chain optimization and logistics
- Intelligent manufacturing and process control
- Smart transportation and traffic optimization

(2) System Integration Improvements

- Develop standardized interfaces for different application scenarios
- Create user-friendly implementation frameworks
- Design automated parameter tuning mechanisms

3. Future Research Directions

(1) Quantum Computing Integration
- Explore advanced quantum annealing algorithms
- Investigate hybrid quantum-classical computing approaches
- Develop more efficient QUBO transformation methods

(2) Model Optimization
- Research improved parameter encoding schemes
- Develop more sophisticated cooling schedules
- Investigate adaptive optimization strategies

(3) Application Development
- Extend to more complex optimization scenarios
- Develop industry-specific solutions
- Create comprehensive application frameworks

The evaluation and promotion analysis reveals that while our proposed models show significant advantages in combining classical machine learning with quantum optimization, there is still room for improvement in both technical and practical aspects. Future research should focus on addressing the identified limitations while expanding the application scope to more diverse domains. The continuous development of quantum computing technology will provide new opportunities for enhancing these models' capabilities and applications.

VIII. References

[1] Biamonte J, Wittek P, Pancotti N, et al. Quantum machine learning[J]. Nature, 2017, 549(7671): 195-202.

[2] Schuld M, Sinayskiy I, Petruccione F. An introduction to quantum machine learning[J]. Contemporary Physics, 2015, 56(2): 172-185.

[3] Cerezo M, Arrasmith A, Babbush R, et al. Variational quantum algorithms[J]. Nature Reviews Physics, 2021, 3(9): 625-644.

[4] Perdomo-Ortiz A, Benedetti M, Realpe-Gómez J, et al. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers[J]. Quantum Science and Technology, 2018, 3(3): 030502.

[5] Box G E P, Jenkins G M, Reinsel G C, et al. Time series analysis: forecasting and control[M]. John Wiley & Sons, 2015.

[6] Hamilton J D. Time series analysis[M]. Princeton university press, 2020.

[7] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge university press, 2000.

[8] Chapelle O. Training a support vector machine in the primal[J]. Neural computation, 2007, 19(5): 1155-1178.

[9] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[10] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. MIT press, 2016.

[11] Lucas A. Ising formulations of many NP problems[J]. Frontiers in Physics, 2014, 2: 5.

[12] Glover F, Kochenberger G A. A tutorial on formulating QUBO models[J]. arXiv preprint arXiv:1811.11538, 2018.

[13] Li R Y, Di Felice R, Rohs R, et al. Quantum annealing versus classical machine learning applied to a simplified computational biology problem[J]. npj Quantum Information, 2018, 4(1): 1-10.

[14] Willsch D, Willsch M, De Raedt H, et al. Support vector machines on the D-Wave quantum annealer[J]. Computer Physics Communications, 2020, 248: 107006.

[15] Benedetti M, Lloyd E, Sack S, et al. Parameterized quantum circuits as machine learning models[J]. Quantum Science and Technology, 2019, 4(4): 043001.

[16] Huang H L, Wu D, Fan D, et al. Superconducting quantum computing: a review[J].

Science China Information Sciences, 2020, 63(8): 1-32.

[17] Kochenberger G, Hao J K, Glover F, et al. The unconstrained binary quadratic programming problem: a survey[J]. Journal of Combinatorial Optimization, 2014, 28(1): 58-81.

[18] Kadowaki T, Nishimori H. Quantum annealing in the transverse Ising model[J]. Physical Review E, 1998, 58(5): 5355.

[19] Lloyd S, Mohseni M, Rebentrost P. Quantum algorithms for supervised and unsupervised machine learning[J]. arXiv preprint arXiv:1307.0411, 2013.

[20] Wang Z, Hadfield S, Jiang Z, et al. Quantum approximate optimization algorithm for MaxCut: A fermionic view[J]. Physical Review A, 2018, 97(2): 022304.

[21] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (December 2015), 19 pages. DOI=http://dx.doi.org/10.1145/2827872.

Team Number:	apmcm2309364
Problem Chosen:	A

2023 APMCM summary sheet

In order to cope with Apple picking robots have low recognition rate, slow speed and low accuracy in image recognition. we use the improved Hough circle transformation method to establish the first type of model for identifying apples.To better identify apples more accurately in complex situations, we extracted the feature value information of apples through CNN convolutional network, conducted deep learning on the provided apple images, and established a second type of apple recognition model based on YOLOv4 and VGG19. In addition, when calculating the quality of the apples in the image, we invented a brightness-estimation method ourselves to make the calculated apple quality more consistent with the actual situation.

Question 1 requires us to identify the number of apples in the image.We use Lab color space conversion and Gaussian filtering method to preprocess the image.Taking advantage of the characteristic that ripe apples are red, we drew the a* channel map in the Lab color space and binarized it. Then, taking advantage that apples can be approximately regarded as circles in a two-dimensional plane, we used an improved Hough circle transformation method, and the number of apples identified in Appendix 1 was 3094.

Questions 2 and 3 require us to identify the position and maturity of apples in the image. We used the YOLOv4 network in the CNN convolutional network for deep learning, input the image data of artifitial frame selection and recognition for model training, and used the trained The model was used for identification, and it was found that the results of the apple position were similar to the results of the first model, indicating that the model was well trained. When calculating the maturity in question 3, we also defined a maturity function H to quantify the maturity of the apples.

Question 4 requires us to calculate the mass of apples. We further considered the imaging principle of ”large near, small far”, analyzed the impact of focal length and the distance between the apple and the lens on the results, and believed that if only calculating the quality of the apple based solely on the area of the apple’s two-dimensional image may be very different from the actual situation. Based on this, we proposed a brightness-estimation method to correct the results, and calculated the corrected apple quality using the data results obtained from the first three questions.

Question 5 requires us to identify apples from different fruits. We noticed that yolov4 is a target detection model, which is more biased towards image classification, and the number of samples in Appendix 2 is too large, and the efficiency of manually dividing the training set is low. Therefore, we used the VGG19 network with more convolutional layers to train the model. Compared with YOLOv4, it has higher efficiency and learning depth.

The innovation of this article is that we use geometric methods, deep learning methods and build models from different angles, and analyze and compare the results obtained, so that the model has high credibility. At the same time, we also considered many practical problems that may not have been considered in the questions, and gave us our solutions.

Keywords: Image recognition Hough circle transform, CNN convolutional network deep learning YOLO network VGG19

Contents

1. Introduction **1**
1.1 Question Background 1
1.2 Restatement of Problem 1
2. Problem analysis **2**
2.1 Analysis of Problem One 2
2.2 Analysis of Problem Two 2
2.3 Analysis of Problem Three 2
2.4 Analysis of Problem Four 3
2.5 Analysis of Problem Five 3
3. Model Hypothesis **3**
3.1 Assumption 1 3
3.2 Assumption 2 4
3.3 Assumption 3 4
3.4 Assumption 4 4
4. Notations **5**
5. Model Establishment and Solution of Problem 1 . . . . . **6**
5.1 Establishment of Apple Recognition Counting Model 6
5.1.1 *Image Data Preprocessing* 6
5.1.2 *Methods for Identifying Apples in Images* 9
5.1.3 *Statistical Methods for Counting Apples and Their Improvements* 11
5.2 Results and Histogram of apple distribution 14
6. Model Establishment and Solution of Problem 2 . . . . . **15**
6.1 Establishment of the Apple Position Determination Model 15
6.1.1 *Principle Overview of YOLOv4 network* 15
6.1.2 *The Application of the YOLOv4 network and Results in this Problem* 16
6.1.3 *Comparison of Model Recognition and Sensitivity Analysis* 20
6.2 Results and 2D Scatter Plot of Geometric Coordinates 20
7. Model Establishment and Solution of Problem 3 . . . . . **21**
7.1 Establishment of apple maturity detection model 21
7.1.1 *The Application of YOLOv4 Network and Results in this Problem* 21
7.1.2 *Establishment of numerical calculation model for maturity* 22
7.2 Results and Histogram of apple ripeness distribution 22
8. Model Establishment and Solution of Problem 4 . . . . . **24**
8.1 Preliminary estimates for Apple 24
8.2 Radance and radiation intensity 25
8.3 Establishment and calculation optimization of ranging model 26
8.4 Estimate of actual fruit size and mass 29
8.5 Results and mass distribution histogram 30

9. Model Establishment and Solution of Problem 5 31
9.1 Establishment of an improved apple recognition model 31
9.1.1 *Principle Overview of VGG19 network* 31
9.1.2 *The Application of the VGG19 network* 33
9.2 Results and histogram for all Apple image ID numbers 34
10. Model Evaluation and Further Discussion. 35
10.1 Merit 35
10.2 Shortcoming 35
10.3 Improvement Measures 36
11. References 36
12. Appendix 38

I. Introduction

1.1 Question Background

Apple-picking robots have become a recent hotspot. In the past, apple harvesting relied mainly on manual labor, requiring a large workforce for several days during the apple's ripening period. However, with the aging rural workforce and the trend of young people leaving villages for work, many places are experiencing a shortage of labor during the apple harvesting season. To address this issue, people have started developing apple-picking robots. But due to variations in orchard environments and controlled experiments, the widespread adoption and application of various apple-picking robots globally have not been ideal.

At the same time, existing robots mostly cannot achieve precise identification of obstacles such as leaves, branches, and fruits. Direct harvesting may potentially damage the fruits or even the harvesting equipment, impacting efficiency and fruit quality. Additionally, there are shortcomings in the recognition and differentiation technologies for different harvested fruits, making post-harvest identification of apples challenging.

Therefore, it is essential to establish an efficient and accurate apple image recognition model. This paper analyzes features such as the number, position, ripeness, and quality of apples in apple images by using image segmentation and recognition techniques, mathematical models, and the deep learning network, and then proposes a feasible apple image recognition algorithm to address these challenges.

1.2 Restatement of Problem

Considering the background information and restricted conditions in the problem statement,the following problems should be solved:

Question 1: Preprocess the image dataset in Attachment 1, analyze the relationship between the number of apples and various parameters in the image dataset, determine the required image features, establish a mathematical model, and count the number of apples in each image. Then draw a histogram of the distribution of the number of apples to visualize the data.

Question 2: In the previous question, we have already determined the number of apples. Next, we need to identify the position of each apple in the image. We establish a Cartesian coordinate system with the lower-left corner of the image as the origin. Since one image may contain multiple apples, we need to assign a unique number to each apple and store the coordinate information for all apples. Then, we will draw a scatter plot of the apple coordinates to visualize the data.

Question 3: In this question , we need to find a reasonable method to calculate the ripeness of apples and then apply this method to perform the calculations. The results should be visualized by creating a histogram that represents the distribution of ripeness for all apples, to visualize the data.

Question 4: In question 1 and 2, we have already determined the number and positions of apples. Now, we need to calculate the two-dimensional area of the apples based on this information. The key lies in determining the boundaries of the apples. Additionally, we are required to estimate the mass of apples. Since we cannot directly measure the mass, we can only obtain an estimate. Therefore, we need to establish the relationship between the two-dimensional area of apples in the

image and their mass for estimation. Subsequently, we will visualize the distribution of the mass of all apples by creating a histogram.

Question 5: In this question, the task is to identify apples from various fruit images. This involves a more in-depth extraction of apple-specific image features, training a model for apple recognition, and applying it to determine which images in Attachment 3 contain apples. Subsequently, the distribution of their ID numbers will be visualized through the creation of a histogram for data analysis.

II. Problem analysis

2.1 Analysis of Problem One

In Problem 1, based on the provided apple image data in Attachment 1, we aim to identify features that can help us recognize relevant information about apples. It is observed that ripe apples exhibit the characteristics of being red and having an approximately spherical appearance (which can be considered circular in a two-dimensional image). Based on these two features, we can identify apples in the image. Before establishing the model, we need to preprocess the image data.

Here, we convert the original RGB image to the Lab image. Subsequently, we filter the a* component (representing the spectrum change from green to red) to obtain the A-channel image in the Lab color space. We perform binarization on this A-channel image and apply Gaussian filtering for noise reduction. Next, utilizing the circular features, we employ an improved Hough circle transformto detect and count the number of circles in the image. Additionally, we merge circles with high overlap. Finally, we solve the data in Attachment 1 and plot the histogram.

2.2 Analysis of Problem Two

In Problem 2, we need to identify the positions of apples in each image. Building on the preprocessing of all images in problem one and obtaining the number of apples in each image, we can roughly determine the positions of apples using the model developed in Problem 1 with the obtained centroids after numbering them. However, after analysis and comparison, we found that there is still a certain deviation between the geometric center of the apple and the centroid obtained from the Hough circle.

Therefore, we adopted a new recognition method based on deep learning to determine the coordinates of each apple, using the YOLOv4 network. We first train the model using images with clear and distinctive features, then apply the trained model to all images in Attachment 1. Finally, we compare the recognition results with the training results and the results from Problem 1, conduct sensitivity analysis on the model, and plot a scatter plot of the geometric coordinates of all apples.

2.3 Analysis of Problem Three

In Problem 3, we can follow the approach used in Problem 2, utilizing the YOLOv4 network for model training. However, during training, it is necessary to annotate the manually added bounding boxes with the ripeness status of the apples (ripe, semi-ripe, and unripe). Subsequently, the model is trained, and the trained model is applied to identify the ripeness status of each apple in the

images. We obtain information about the ripeness status and confidence level of each recognized apple. Based on this information, we can design a function to measure apple ripeness, allowing us to calculate the ripeness of each apple. Finally, a histogram depicting the ripeness distribution of all apples is generated for further analysis.

2.4 Analysis of Problem Four

In Problem 4,we want to solve in this question is to calculate the area of a two-dimensional image and estimate the mass of the apple. For the area problem, we might as well use the radius rad we obtained in the first question to calculate it as a simple estimate. As we said in the hypothesis, we assume that the density of apples does not change significantly during the development cycle, which is a fixed value, which may be recorded as ρ. In addition, our team noticed that there are obvious differences in the image in Appendix 1 The problem of large near and far small results in the error result that for some apples in the image, the actual apples are larger and the apples in the image are smaller. To this end, our team reviewed relevant literature on calculating the actual distance of objects through two-dimensional plane images. Unfortunately, the methods reviewed in the literature were not applicable to the dilemma faced by our team. After discussion, we finally proposed a new method - the brightness-estimation method to approximately estimate the actual size of the apple. In this method we need to know the focal length of the camera when shooting, and finally we draw the mass distribution histogram of all apples.

2.5 Analysis of Problem Five

This question requires us to distinguish apples from other fruits in the image, which requires us to extract more information features about apples, instead of calculating the location, maturity, quality, etc. of apples like the previous questions. If we continue to use the YOLOv4 network, when the number of training times is high, the time complexity will be too high. Therefore, we should choose a deep learning algorithm with more convolutional layers and faster computing speed so that we can be more accurate. and efficiently identify apples. At the same time, considering that there are too many Apple ID numbers, we choose to cluster every 1,000 consecutive IDs, count the number of apple images within these 1,000 IDs, and then draw a histogram.

III. Model Hypothesis

3.1 Assumption 1

There are often small and numerous apples in the background of the image that are difficult to count. For the purpose of ensuring the speed and accuracy of image recognition, we will ignore the small apples in the image.

Explanation: This assumption is reasonable and sufficient. For the hypothetical situation, in real life, you can better identify the apples by getting closer to reduce the distance, taking partial images, and changing the shooting angle, instead of forcing it on the numerous and difficult-to-count apples in the background. detection, which cannot guarantee the speed and accuracy of recognition.

3.2 Assumption 2

The fruit volume and density of apples are basically unchanged during the ripening period of the fruit. If small errors are ignored, they can be regarded as fixed values.

Explanation: Generally speaking, the fruit develops during the fruiting period and the overall shape continues to expand. When the fruit shape is basically stable, it begins to enter the fruit maturity stage. The external manifestation of the fruit during the ripening stage is mainly the change in the color of the peel. Its shape and quality are basically unchanged, and its density can also be considered unchanged.

3.3 Assumption 3

All apples are red apples. Green apples are considered immature red apples, and we believe that their degree of ripeness is linearly related to the degree of the apple's gradient from green to red.

Explanation: The appearance of green apples and unripe red apples are very similar, and since the image data we can obtain are only two-dimensional images with lower pixels, we cannot accurately distinguish between the two. Considering that most of the image data types are red apples, we can think that all apple types are red apples, and the ripeness of apples can only be distinguished by color, so it is assumed that apple color changes and ripeness changes are linear The relevant ones make it easier for us to build models, and the gap with the actual situation is very small.

3.4 Assumption 4

In the image data given, all photos were taken under natural light with uniform and soft light. The objects captured were all still life, and no ghosting was produced during the shooting. The focal length of the shooting equipment used is the same.

Explanation: The time for picking apples is generally during the day, so there is no need for another strong light source, and the strong light source will have a great impact on the imaging and affect the recognition results. We believe that the light is uniform and soft natural light, which is easy to eliminate interference factors. Under normal circumstances, the position of the apple does not change much due to the influence of the external environment. At least we can think that the apple is stationary while shooting, eliminating the interference of ghost images during imaging on the recognition results. And since the imaging equipment of a picking robot is the same, in practical applications we can think that the shooting focal length of all pictures is the same.

IV. Notations

Symbol	Description
γ	Parameter used to describe the overlap rate between two circles
μ	Weighting factors used to measure maturity
H	Function used to represent the ripeness of an apple
$\widetilde{H}$	Parameter used to represent average maturity
W_{img}	The average brightness of a certain area on a two-dimensional grayscale image
$G_{m\times n}$	Represents the gray value matrix of pixels on a two-dimensional grayscale image of size mxn
$\|\|A\|\|$	Sum all elements of matrix A

V. Model Establishment and Solution of Problem 1

5.1 Establishment of Apple Recognition Counting Model

5.1.1 *Image Data Preprocessing*

At first, we need to preprocess the image data provided in Attachment 1. We observed that the provided image data is in RGB format. In order to filter out the red color of apples in the image, we plotted the R-channel image. However, we noticed that due to interference from other colors (such as the brown color of tree branches), the resulting image after binarization is not ideal, as shown in the following figure:



(a) The original image.

(b) The R channel image in the RGB color space after binary processing.

Figure 1 Comparison of original image and RGB grayscale image

After reviewing relevant literature and conducting analysis, we have determined that selecting the A channel image in the Lab color space (hereafter referred to as the A channel image) is a preferable choice. This is because the a* component in the A channel image represents the spectrum change from green to red, which can effectively adapt to backgrounds predominantly composed of green leaves. Additionally, it helps mitigate interference from other colors.

Below is the process of converting the original image's RGB parameters to Lab parameters. Since RGB parameters cannot be directly converted to Lab parameters, we opt for using XYZ parameters as an intermediate step, i.e., RGB → XYZ → Lab. The conversion formulas are as follows:

$Part1:(RGB\ \ to\ \ XYZ)$

Assuming r, g, and b are the pixel values for the three channels, with a range of [0, 255], the conversion formula is as follows:

$$\begin{cases} R = gamma(\dfrac{r}{255.0}) \\ G = gamma(\dfrac{g}{255.0}) \\ B = gamma(\dfrac{b}{255.0}) \end{cases} \tag{1}$$

$$gamma(x) = \begin{cases} (\frac{x+0.055}{1.055})^{2.4} & (x > 0.04045) \\ \frac{x}{12.92} & (otherwise) \end{cases} \tag{2}$$

The gamma function is used for non-linear tone mapping in images with the aim of enhancing contrast. The expressions mentioned above represent commonly used gamma functions.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3}$$

Where the transformation matrix M is given by

$$M = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{pmatrix}$$

$Part2 : (XYZ\ \ to\ \ Lab)$

$$\begin{aligned} L^* &= 116 f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n) - f(Z/Z_n)] \end{aligned} \tag{4}$$

And

$$f(t) = \begin{cases} t^{\frac{1}{3}} & if\ t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & otherwise \end{cases} \tag{5}$$

In the two formulas above, L*, a*, and b* represent the values of the three channels in the final Lab color space. X, Y, Z are the values calculated after converting RGB to XYZ, and Xn, Yn, Zn are typically set to the default values of 95.047, 100.0, and 108.883.

Based on the above principles, we implemented the process of converting RGB to Lab through code. Then, in MATLAB, for the Lab image obtained, we filtered the a* component and plotted its binarized image based on the L value. The result is the binary image of the A-channel in the Lab color space, as shown below:

Figure 2 The channel image



(a) The R channel image in the RGB color space after binary processing.

(b) The a* channel image in the Lab color space after binary processing.

(c) The a* channel image in the Lab color space.

After comparison, we can observe that the binarized image obtained using the Lab color space does indeed have better results than the one obtained using the RGB color space.

However, this alone is not sufficient. Through zooming in and contrasting the images, it is evident that there are numerous noise points in the image, and there is a noticeable aliasing effect at the edges, likely due to the low resolution of the image. This is unfavorable for the subsequent Hough Circle Transform. Therefore, we need to apply denoising techniques to make the image relatively smooth.

Here, we employ the method of Gaussian filtering:

Below is a brief introduction to the Gaussian filtering method:

Gaussian filtering is a process of weighted averaging for the entire image. The value of each pixel is obtained by taking the weighted average of its own value and the values of other pixels in its neighborhood. It is a form of linear smoothing filter commonly used to eliminate Gaussian noise. Gaussian filtering is widely applied in the denoising process of image processing.

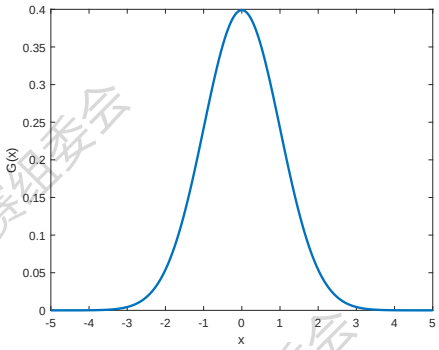
$One-dimensional\ \ Gaussiandistribution$:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

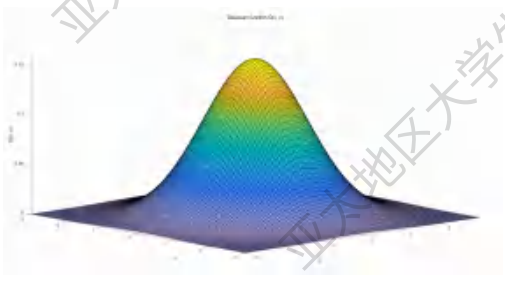
$Two-dimensionalGaussiandistribution$:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Where σ is typically chosen as 1. The following is a schematic diagram of the Gaussian distribution.



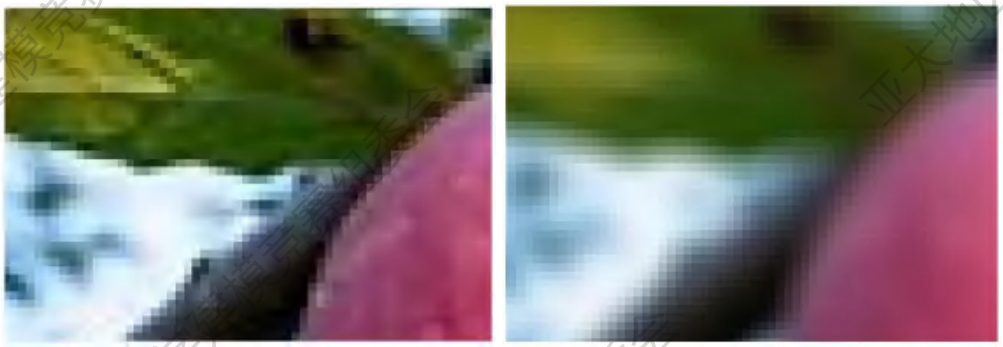
(d) Gaussian function G(x)



(e) Gaussian function G(x,y)

We adopt a two-dimensional Gaussian filtering here.

After debugging, we chose a suitable Gaussian kernel to apply Gaussian filtering to the previously obtained a* channel image. The resulting denoised image, when compared to the original, appears significantly smoother.



(f) The image before filtering. (g) The image after filtering.

Figure 3 Comparison chart before and after Gaussian filtering

At this point, we have completed the preprocessing of the image data.

5.1.2 *Methods for Identifying Apples in Images*

After completing the preprocessing of the image data, our next step is to devise a method to identify apples in the processed images and determine the boundaries of the apples, facilitating subsequent counting and statistics.

Considering the characteristic nearly circular shape of apples in two-dimensional images, we adopted an improved Hough circle transform method to identify apples in the image. The following is a brief description of the Hough circle transform:

The Hough circle transform is an extension of the classical Hough transform, which is used for detecting lines in images. The principle is as follows:

1. We can use (a, b, r) to define a circle with center (a, b) and radius r.
2. If a circle passes through the point (x_1, y_1), then it satisfies the equation:

$$(x_1 - a)^2 + (y_1 - b)^2 = r^2$$

3. Therefore, all circles passing through the point (x_1, y_1) can be represented as $(a_1(i), b_1(i), r_1(i))$, where $r_1 \in (0, +\infty)$. Each value of i corresponds to a different circle, and $(a_1(i), b_1(i), r_1(i))$ represents an infinite set of circles passing through the point (x_1, y_1).
4. All circles passing through the point (x_1, y_1) can be represented as $(a_1(i), b_1(i), r_1(i))$, all circles passing through the point (x_2, y_2) can be represented as $(a_2(i), b_2(i), r_2(i))$, and all circles passing through the point (x_3, y_3) can be represented as $(a_3(i), b_3(i), r_3(i))$. If these three points lie on the same circle, then there exists a set of values (a_0, b_0, c_0) such that $a_0 = a_1(k) = a_2(k) = a_3(k)$, $b_0 = b_1(k) = b_2(k) = b_3(k)$, and $r_0 = r_1(k) = r_2(k) = r_3(k)$, meaning that these three points simultaneously lie on the circle (a_0, b_0, c_0).

5. Analyzing all circles passing through the point (x_1, y_1), when $r_1(i)$ is determined, the trajectory of $(a_1(i), b_1(i))$ is a circle with center $(x_1, y_1, r_1(i))$ and radius $r_1(i)$. Therefore, all circles $(a_1(i), b_1(i), r_1(i))$ form a cone with vertex at $(x_1, y_1, 0)$ and a cone angle of 90 degrees. The intersection point A of the three cones is the circle passing through these three points simultaneously. The point A with the highest cumulative count is a potential center of the circle.

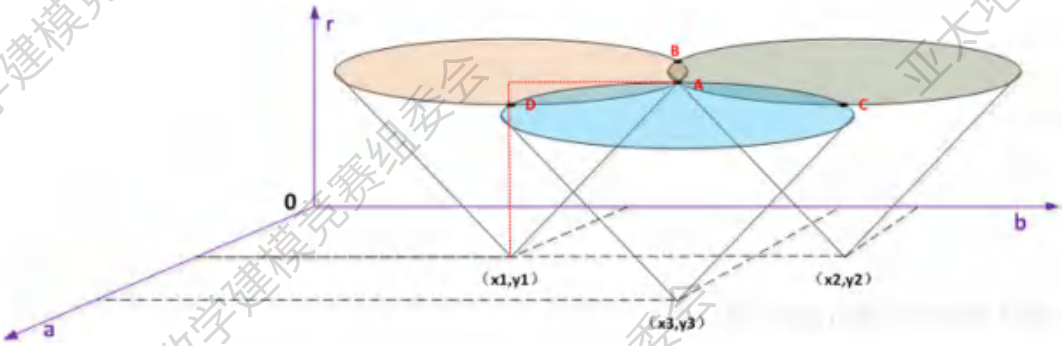


Figure 4 Hough Circle Transform diagram (Image Sourced from the internet).

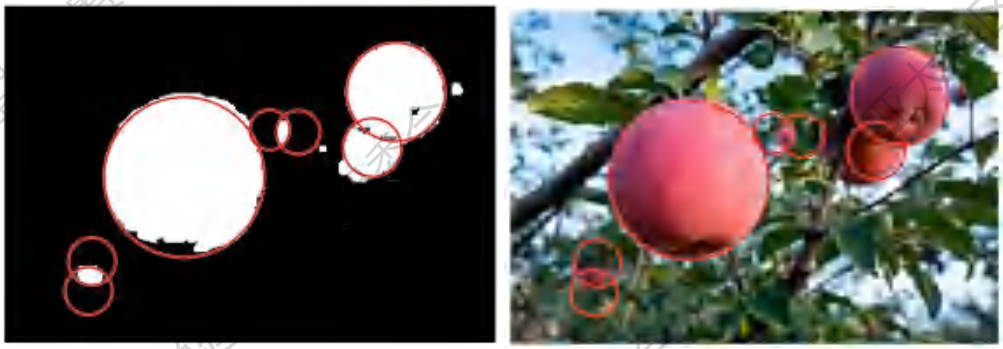
The Hough circle transform can accurately identify some circles that may exist in the image. However, this method involves a massive computational load, resulting in slow code execution. Therefore, we opted for a more efficient and flexible Hough gradient method. This method significantly reduces the computational load compared to the standard Hough circle transform and produces good results.

The following is an explanation of the implementation of the Hough gradient method for circle detection:

- **Step1:**Apply Canny edge detection to the obtained binary image to identify the positions of edge points in the image.
- **Step2:**For each non-zero point in the edge image, compute its local gradient by using the $imgradientxy()$ function to calculate the first-order derivatives in the x and y directions, obtaining the gradient.
- **Step3:**Based on the gradients of the obtained edge points, accumulate each point in the accumulator array along the line specified by the slope (where the slope varies from the minimum to the maximum value as defined).
- **Step4:**Mark the position of each non-zero pixel in the edge image when simultaneously with each accumulation.
- **Step5:**Then, select candidate centers from these points in the two-dimensional accumulator, ensuring that these centers are both greater than a given threshold and larger than all their neighbors. These candidate centers are arranged in descending order based on their accumulation values, facilitating the appearance of the center with the most support pixels first.
- **Step6:**These pixels are then sorted based on their distances from the center. Starting from the minimum distance to the maximum radius, select the most suitable radius that encompasses all non-zero pixels.
- **Step7:**If a center receives sufficient support from non-zero pixels in the edge image and has an adequate distance from the previously selected centers, then it will be retained.

The points that are ultimately retained are the centers of the apple circles found by the Hough Gradient method. Based on the number of centers in the image, we can preliminarily estimate the approximate number of apples in the image.

Below are the results of the apple count obtained using this method.



(a) binary image result (b) corresponding original image result

This is the flowchart of our entire apple recognition and counting algorithm:

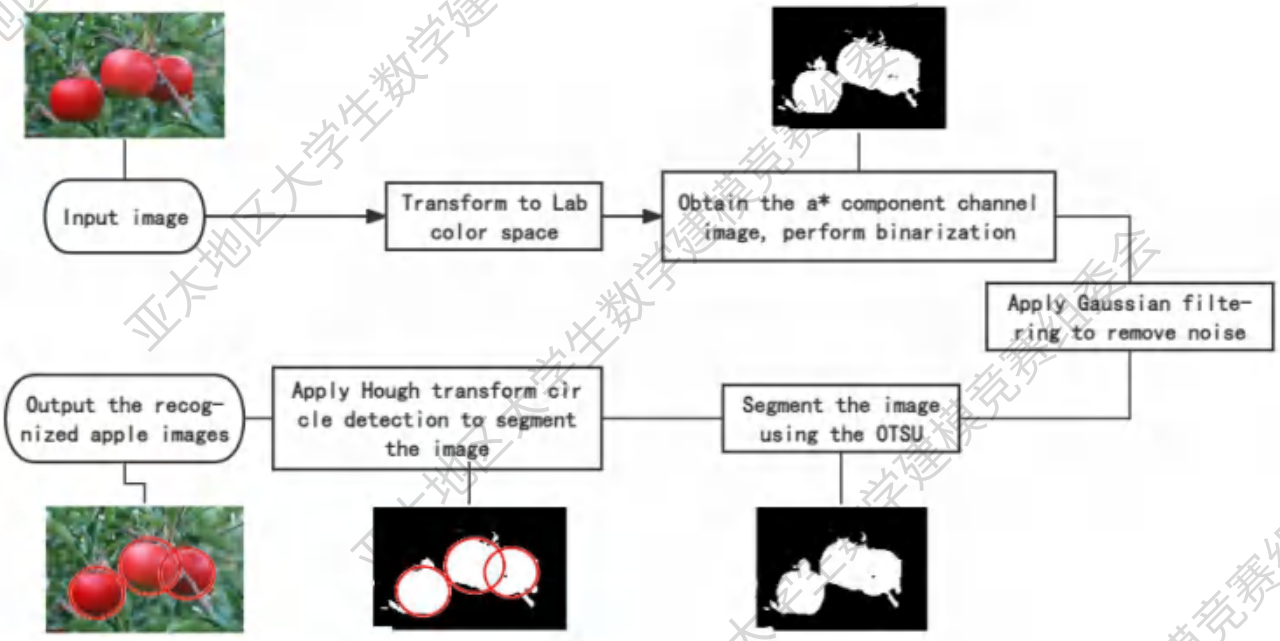


Figure 5 Flowchart of the Apple Recognition Algorithm.

5.1.3 *Statistical Methods for Counting Apples and Their Improvements*

In the previous section, we found a method to identify apples in the image. However, upon practical comparison, we discovered that due to factors such as lighting and shooting angles, the obtained count of apples may differ from the actual count, leading to significant discrepancies in the results.

Through analysis, we have identified that apples are not perfectly circular but appear as a combination of multiple circles. Therefore, the algorithm is prone to situations where one apple

is recognized as multiple apples, especially at the boundaries. To address this issue, we propose an optimization to the original model based on the overlapping area ratio of apple circles.

We propose the following improved method to address this issue:

1. Based on the previously obtained data for all circle centers $(x_i, y_i)$and their radii r_i we calculate the distances $d_{ij}(i \neq j) \in (0, r_i + r_j)$.
2. For ease of discussion, let's denote $R_1 = \max(r_i, r_j), R_2 = \min(r_i, r_j), d = d_{ij}$.
3. Next, we introduce the parameter $\gamma = \dfrac{R_1 - d}{R_2} \in (0, 1)$, The introduction of$\gamma$ is intended to assist us in determining whether one of the circles should be eliminated in cases where two circles overlap.
4. For this purpose, we need to find a threshold value λ, When $\gamma > \lambda$, we consider that the overlap degree of the two circles is relatively high, and we remove the circle with the smaller radius N.
5. The number of circles retained is the optimized count of apples. Use the Count function to perform the counting.

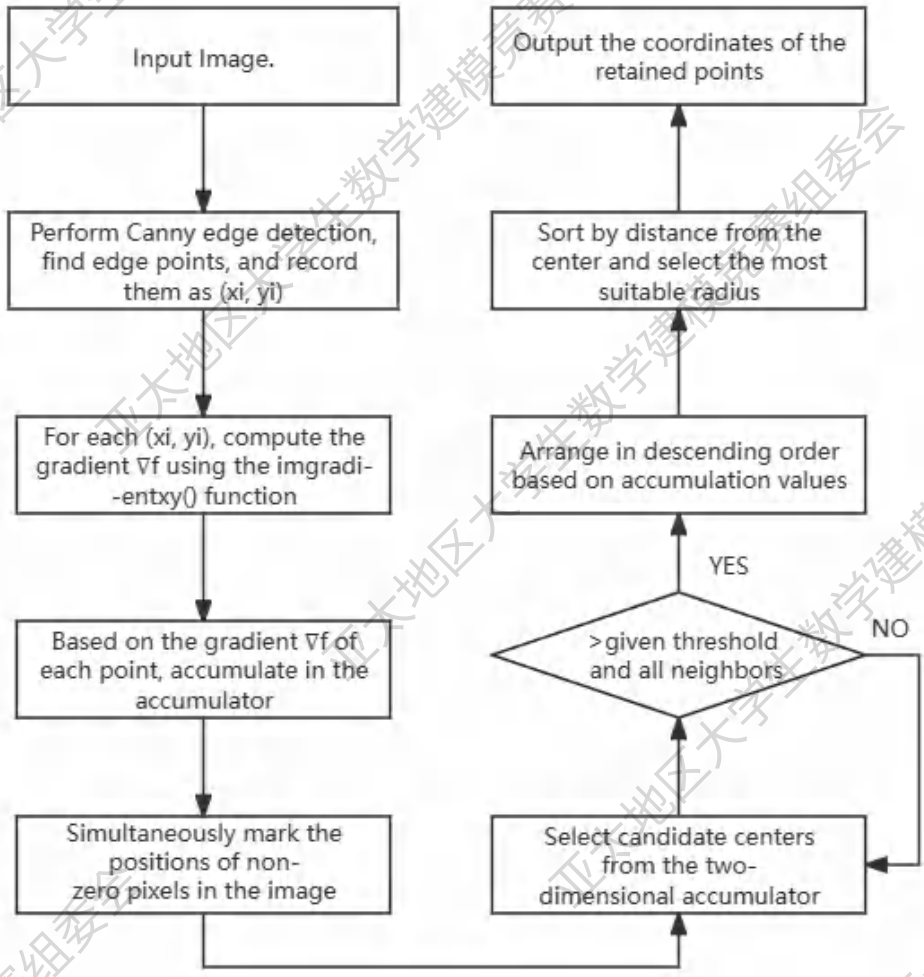


Figure 6 Flowchart of the Optimization Model for Eliminating Duplicate Circles.

Next, let's search for λ

When comparing the recognition image with the original image, we observe that when γ<0 i.e.,

when $R_1 < d$, neither circle's center is inside the other circle. In this case, either there is no overlap or the overlapping area is very small. We can consider these as two separate apples without further discussion. When γ>1, the smaller circle is completely encompassed by the larger circle, and we can consider it as one apple.

So, the λ we are seeking should be a value greater than 0 and less than 1.

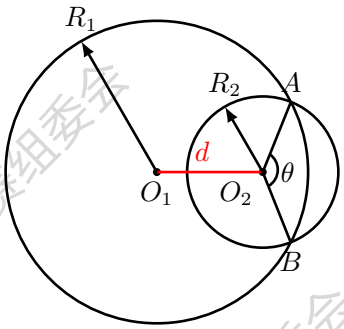


Figure 7 Schematic diagram of two circles overlapping

When $\gamma \in (0,1)$, we can observe that there are two intersection points between the two circles. By connecting the center of the smaller circle with these two intersection points, we can obtain an angle θ. We can find that when R_1 and R_2 are both known, the area of overlap is related to θ and is negatively correlated (i.e., the smaller θ, the larger the overlapping area). From the images, we can

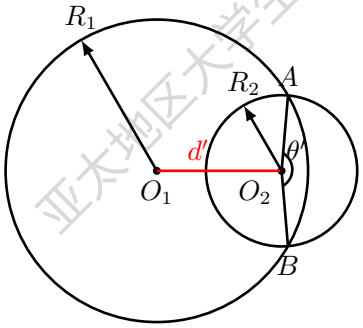


Figure 8 The bigger

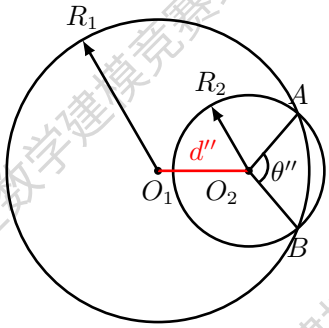


Figure 9 The smaller

also observe that as the value ofγ increases, the distance d between them decreases, and θ continues to decrease as well.

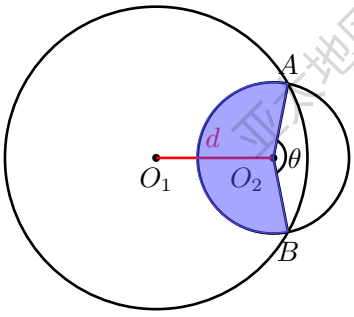


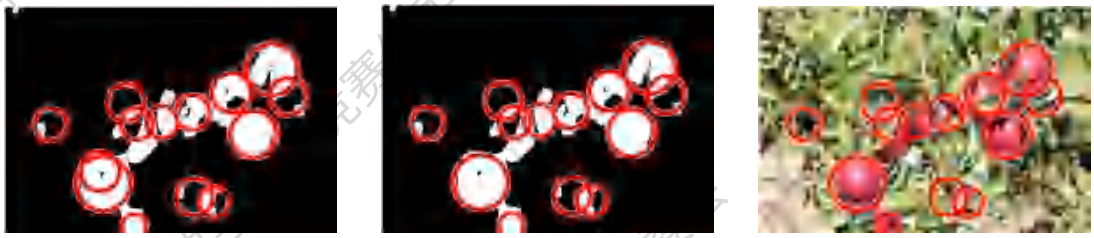
Figure 10 Overlap area i estimation diagram

From this, we can infer that the parameter γhas a positive correlation with the overlapping area.

In this case, we choose λ=0.5 as the threshold for determining whether two circles overlap. When γ =0.5, the angle θ is 120°. Through calculation, we find that at this point, the overlap degree of the two circles is above 70%.So we consider the overlap to be relatively high, and the two circles are identified as one apple.

5.2 Results and Histogram of apple distribution

The comparison between the results before and after optimization is shown below:



(a) The recognition results before optimization.

(b) The recognition results after optimization.

(c) The presentation in the original image.

Figure 11 The comparison of effects before and after correction

From this, it is evident that the improved model does enhance recognition accuracy, and the recognized results align well with the distribution and count of apples in the original image. This can be used to calculate the quantity of apples.

According to this model, we will plot the obtained results using a bar chart, as shown below:

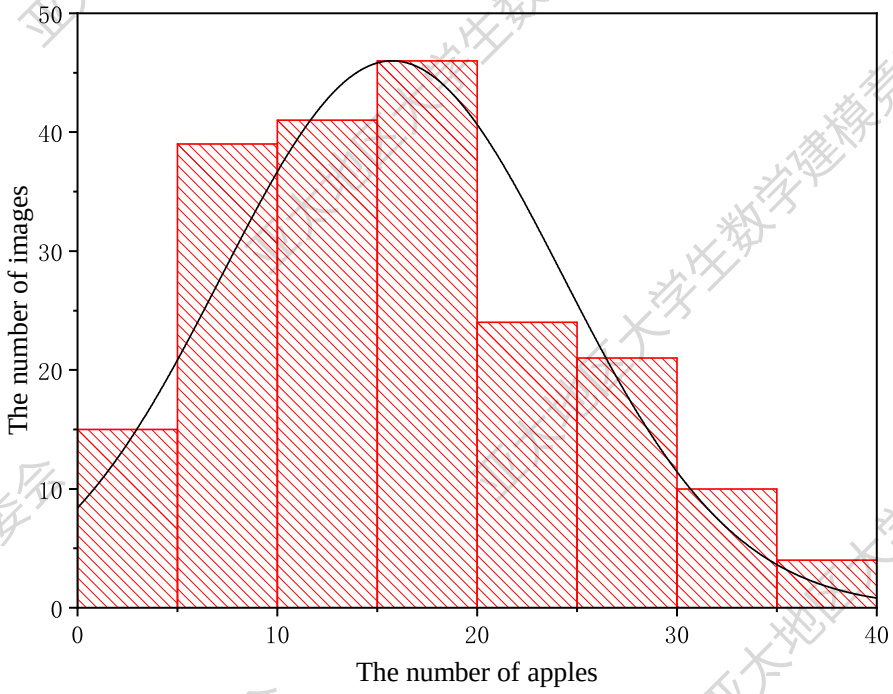


Figure 12 The histogram of the distribution of all apples in Attachment 1

On the basis of the histogram, we plotted a fitting curve and observed that the curve exhibits an approximate normal distribution. Since the provided images of apples are random and without a specific pattern, we can consider the obtained results to be satisfactory, indicating a high stability of the model.

VI. Model Establishment and Solution of Problem 2

6.1 Establishment of the Apple Position Determination Model

From the solution to the Problem 1, we observed that using the Hough Circle Transform method allowed us to determine the center of each Hough circle. However, there might be a certain deviation between the circle center positions and the actual center position of the apple cluster. To improve the model's accuracy, we adopted the YOLOv4 network and employed a deep learning approach to determine the coordinates of the apples. The results can be compared with those from the Problem 1 to assess the accuracy and rationality of the model. Below is an introduction to the YOLOv4 network.

[1]

6.1.1 *Principle Overview of YOLOv4 network*

The YOLOv4 object detection network is a one-stage object detection network, consisting of three main parts: Backbone, Neck, and Head.

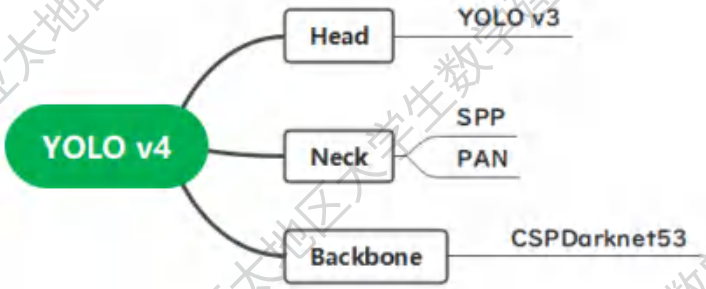


Figure 13 The composition of YOLOv4.

Through the **Figure 16**, we can clearly see its process. Below are detailed explanations:

- The YOLO v4 network uses CSPDarkNet-53 as the backbone for extracting features from the input images. The backbone has five residual block modules, and the feature map outputs from the residual block modules are fused at the neck of the YOLO v4 network.
- The SPP module in the neck concatenates the max-pooling outputs of the low-resolution feature map to extract the most representative features. The SPP module uses kernels of size 1-by-1, 5-by-5, 9-by-9, and 13-by-13 for the max-pooling operation. The stride value is set to 1.

[1] When using the VOLOv4 algorithm, we referred to the references in MATLAB Help, the link is as follows https://ww2.mathworks.cn/help/vision/ug/object-detection-using-yolov4-deep-learning.html

Concatenating the feature maps increases the receptive field of backbone features and increases the accuracy of the network for detecting small objects. The concatenated feature maps from the SPP module are fused with the high-resolution feature maps by using a PAN. The PAN uses upsampling and downsampling operations to set bottom-up and top-down paths for combining the low-level and high-level features.

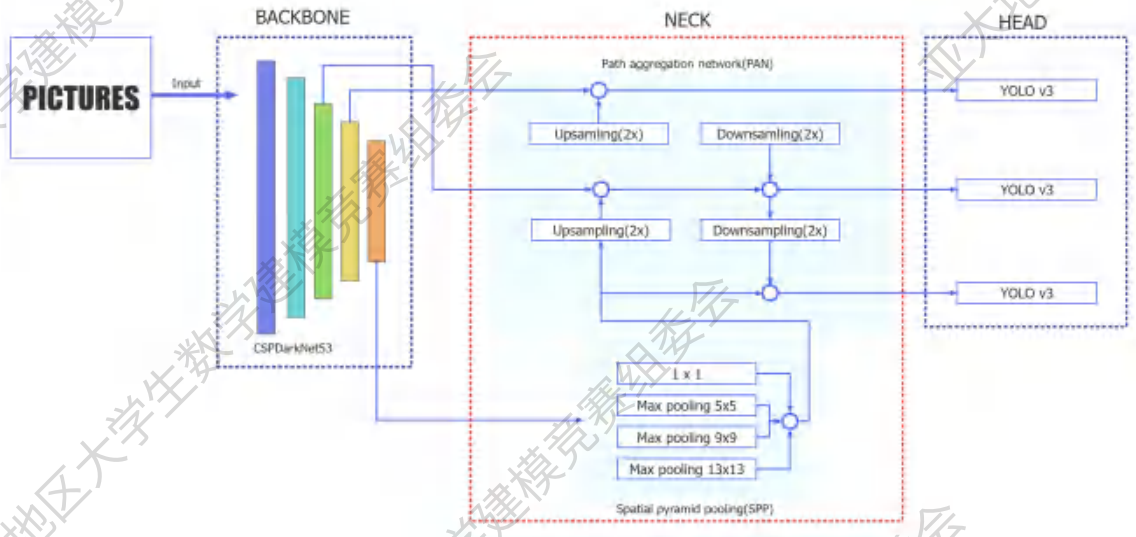


Figure 14 YOLOv4 Network Algorithm Flowchart.

- The PAN module outputs a set of aggregated feature maps to use for predictions. The YOLO v4 network has three detection heads. Each detection head is a YOLO v3 network that computes the final predictions. The YOLO v4 network outputs feature maps of sizes 19-by-19, 38-by-38, and 76-by-76 to predict the bounding boxes, classification scores, and objectness scores.
- Tiny YOLO v4 network is a lightweight version of the YOLO v4 network with fewer network layers. The tiny YOLO v4 network uses a feature pyramid network as the neck and has two YOLO v3 detection heads. The network outputs feature maps of size 13-by-13 and 26-by-26 for computing predictions.

6.1.2 *The Application of the YOLOv4 network and Results in this Problem*

Before training the model with the YOLOv4 network, we need to select samples from the provided apple image data in Attachment 1 and preprocess these data. The process of selecting samples and preprocessing is outlined below:



(a) (b) (c)



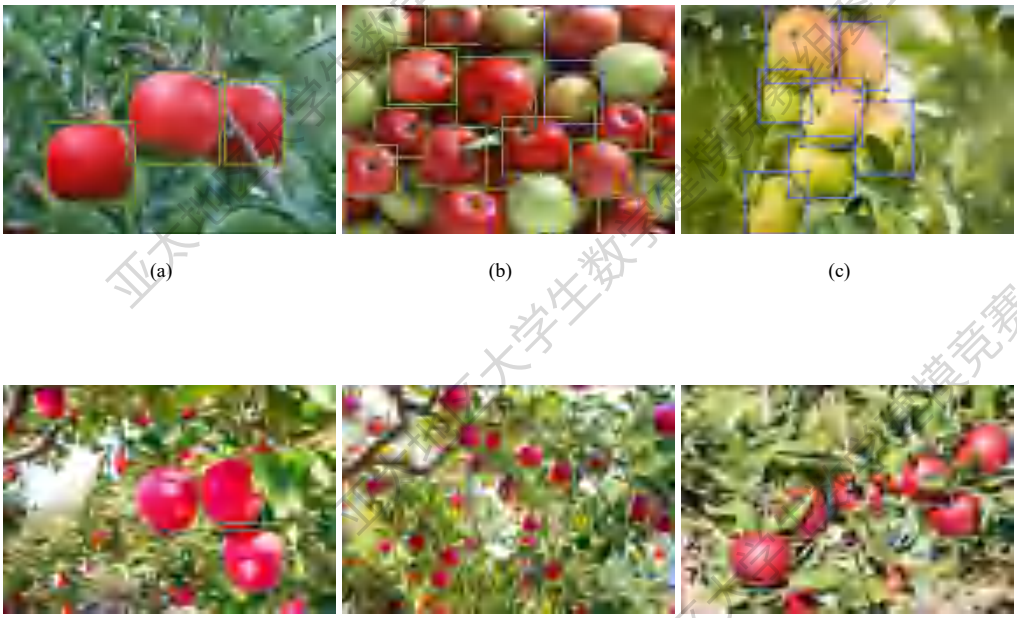
(d) (e) (f)

Figure 15 Examples of selected sample images

We randomly selected 100 images as training samples. Here, we provide a few to illustrate that the chosen apple images exhibit a variety of distinctive features.

Then we utilized the Image Labeler toolbox in MATLAB to manually annotate the selected samples by bounding boxes, where each box represents the region containing an apple. After processing, we obtained a 90x2 table. The first column holds the paths of the images, and the second column stores the coordinates of all the bounding boxes in each image.

The results after adding bounding boxes are as follows. Similarly, we provide examples of these images with added bounding boxes:



(a) (b) (c)

(d) (e) (f)

Figure 16 Results with bounding boxes added

After completing the annotation, all box coordinates are labeled and recorded for subsequent calculation of the position of each apple.

Then, we take the selected 100 processed sample images and randomly divide them into three sets in an $8:1:1$ ratio using the $rand()$ function. These sets are referred to as the training set K1, validation set K2, and test set K3, respectively.

Up to this point, we have completed the preprocessing of the training samples.

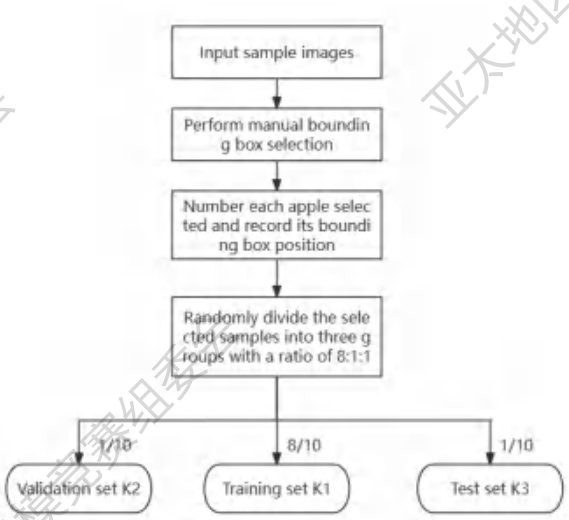


Figure 17 Data preprocessing flowchart of Problem 2

Then we input the samples into the YOLO v4 model for training:

1. First, we perform convolution operation: Perform convolution operation on the training set to extract the feature values of the input information. It's important to note that in the process of scene recognition, we are not instantly recognizing all targets in the entire scene. Instead, we focus on locally perceiving a single feature in the scene and then perform local operations through higher levels to obtain global information. The following diagram shows the structure of the convolutional layer.

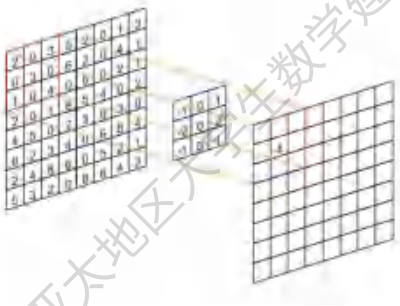


Figure 18 Convolutional Layer Structure Diagram (Image Sourced from the Internet)

2. Next, we enter the pooling layer. The purpose of the pooling layer is to compress the image. It exists between multiple convolutional layers and performs feature dimension reduction on image information. It compresses global information of the image and reduces the number of parameters to prevent overfitting and improve fault tolerance. Pooling can be divided into max pooling and average pooling operations. The following diagram illustrates these two pooling operations.

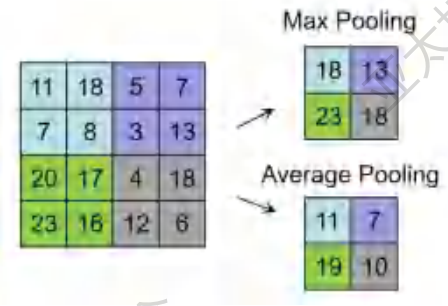


Figure 19 Pooling Layer Structure Diagram (Image Sourced from the Internet)

3. Finally, we enter the fully connected layer. Here, we perform fully connected operations on all feature images. If overfitting occurs, we use operations such as dropout and local normalization to randomly delete some neurons generated during the neural network training process.
4. After extracting features from the training set, let's denote the extracted feature information asε. Then, we using ε to recognize the validation set. The number of recognized apples is denoted as m, and their positions are (x_i, y_i), Subsequently, based on the previously manually selected results, we calculate the number of apples M and their positions(x_{0i}, y_{0i}) in the validation set. We perform confidence calculation for the model and record it in Training Loss.
5. Check whether the value of Training Loss satisfies the confidence interval I, If it does not satisfy, then perform convolution operation again on the training set to re-extract feature information, denoted as ε^*. If it satisfies, proceed to the next step.
6. Use the feature information that passed the test to examine the test set. Similarly, calculate its feasibility and record it in Validation Loss for subsequent evaluation of the model's credibility.

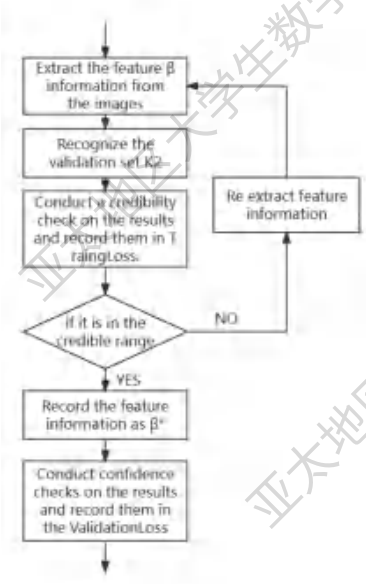


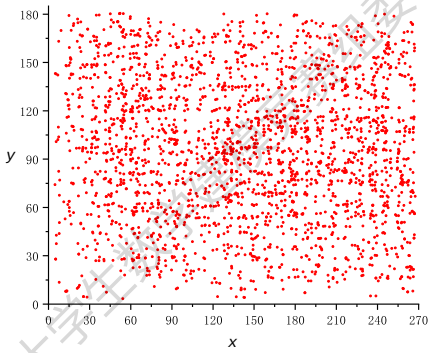
Figure 20 Flowchart of model training(part)

At this point, we have completed the training of the sample model under the YOLO v4 network, which can be used for the subsequent detection of apple positions.

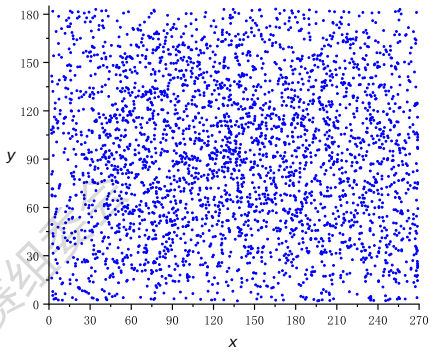
6.1.3 *Comparison of Model Recognition and Sensitivity Analysis*

We plot the result (the left) against the apple position obtained by Hall circle transformation from question 1 (the right)

After comparison, we found that although there was a slight error in the number of identifications between the two, the distribution of point coordinates generally coincided, which shows that the apple position coordinates we obtained are relatively accurate.



(a) The scatterplot of the results from the model in question 2



(b) The scatterplot of the results from the model in question 2

6.2 Results and 2D Scatter Plot of Geometric Coordinates

Our final scatterplot is as follows: Each point represents the geometric coordinate position of the geometric center of an apple in the figure.

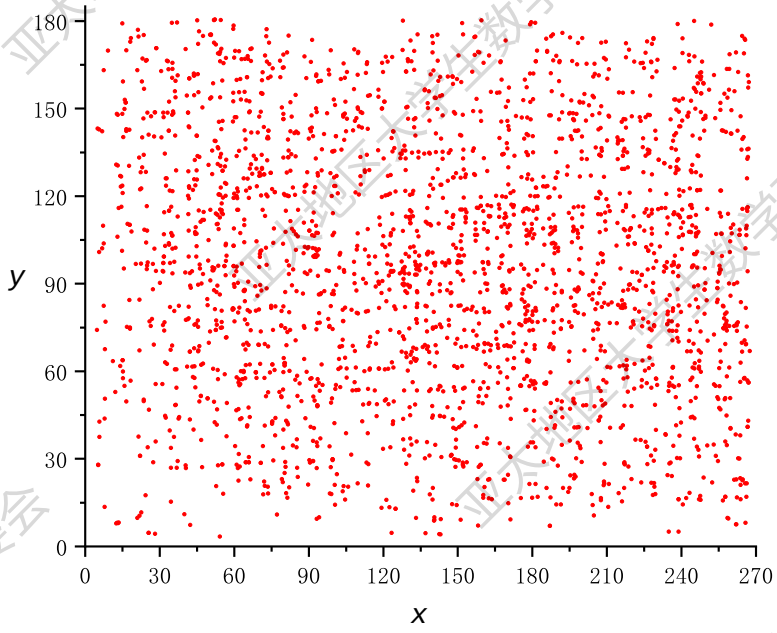


Figure 21 2D Scatter Plot of Geometric Coordinates

VII. Model Establishment and Solution of Problem 3

7.1 Establishment of apple maturity detection model

7.1.1 *The Application of YOLOv4 Network and Results in this Problem*

Similar to the second question, we randomly selected 100 images as training samples. In addition to the frame addition operation, we also performed a maturity status judgment on each apple, using R, S, and I to represent mature and semi-ripe apples respectively.mature and immature.

We can refer to the model in question 2 for the process of feature information extraction, and will not go into details here. The difference from question 2 is that this time the number of feature information sets we extracted increased. The feature information of mature apples, semi-ripe apples and immature apples is t1, t2, t3.

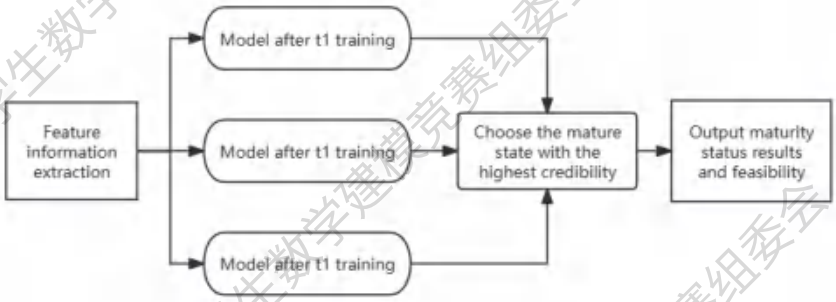


Figure 22 Schematic flow diagram of this process

After the above operations, three types of feature information are used to identify the image of attachment 1 respectively, and the credibility of the apple status is tested by comparing the three types of feature information, and the one with the highest credibility is selected as the judgment for identifying the ripe status of the apple, and It is represented in the figure as follows:
(For example, R: 94% indicates that this apple is mature, and the credibility is 94%)



(a) (b)

Figure 23 Schematic diagram of maturity identification results

Through the above training process of the model, we can get a rough judgment of the ripeness status of each apple and its credibility. Next, we hope to quantify this information and express it with a unified parameter maturity.

7.1.2 *Establishment of numerical calculation model for maturity*

Through the data solved in the previous section, we can get the following parameters:

The numbers of ripe, semi-ripe, and immature apples in all images is recorded as A, B, and C.

The ripeness status of each apple $label \in \{R, S, I\}$

The confidence level of each apple' s ripeness p

To quantitatively express the maturity of each apple, we introduce the maturity factor μ

$$\mu = \begin{cases} 1, & if\ label\ is\ R \\ \frac{1}{2}, & if\ label\ is\ S \\ 0, & if\ label\ is\ I \end{cases}$$

And we assign the above value to μ according to its ripeness status for each apple.

So we can preliminarily think that the apple's maturity H is related to the product of μ and its credibility p. We record maturity as H

So we can initially get:

$$H = \mu p$$

After testing, we found that the error in calculating maturity in this way is relatively large. Analyzing the reason, we believe that the ripe state of apples is not binary opposition, so it is necessary to consider the possible ripe state of apples under the probability of (1-p). Here we introduce the average maturity $\widetilde{H}$, and is given by:

$$\widetilde{H} = \begin{cases} \frac{\frac{1}{2}\sum p_{bi}}{B+C}, & if\ label\ is\ R \\ \frac{\sum p_{ai}}{A+C}, & if\ label\ is\ S \\ \frac{\sum (p_{ai} + \frac{1}{2}p_{bi})}{A+B}, & if\ label\ is\ I \end{cases} \tag{6}$$

Among them, p_{ai}, p_{bi}, and p_{bi} respectively represent the credibility of an apple with labels R, S, and I.

So we can get the calculation formula of the modified maturity H:

$$H = \mu p + (1-p)\widetilde{H} \tag{7}$$

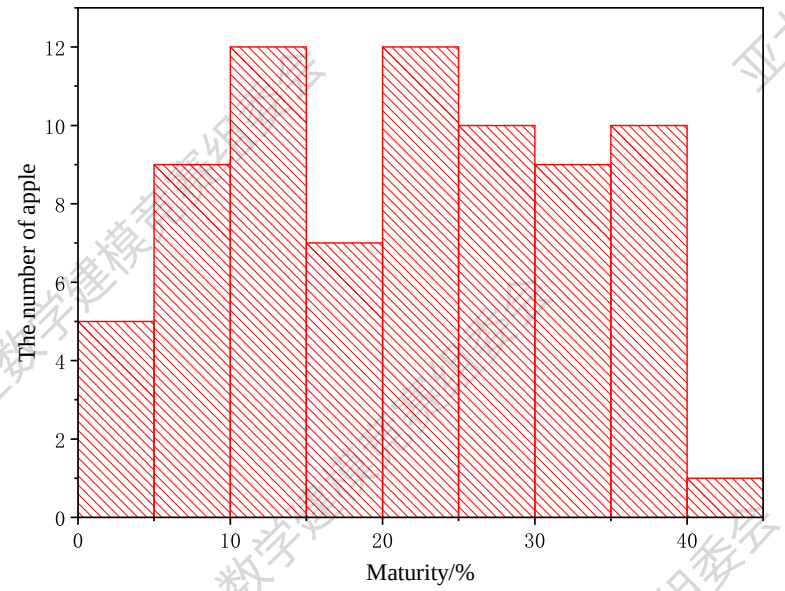
Based on this, we can calculate the maturity H of each apple.

7.2 Results and Histogram of apple ripeness distribution

We drew a histogram of apple maturity distribution with every 5% of maturity as interval step. We can find that most of the identified apples are relatively mature, which is consistent with the actual situation. But at the same time, because there are too few immature apples, if they are all

represented in one histogram, the data visualization effect will be poor, so we divide the histogram into two parts and draw them separately, as follows:

Figure 24 Histogram of ripeness distribution of unripe apples



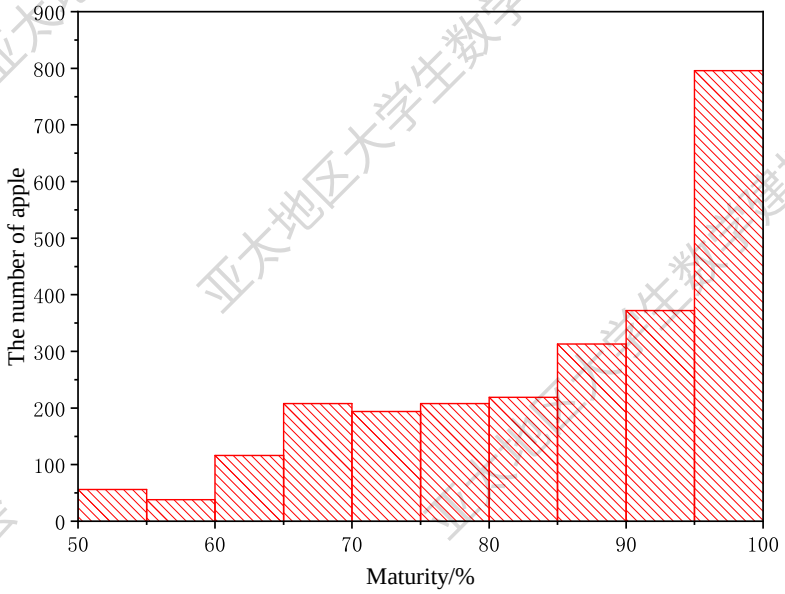


Figure 25 Ripeness distribution histogram of ripe and semi-ripe apples

VIII. Model Establishment and Solution of Problem 4

8.1 Preliminary estimates for Apple

Generally, apples are represented by 70, 80, 90, etc. The unit is mm, indicating the diameter. Here we might as well take the diameter of the apple as $9cm$. According to the image radius rad calculated in question 1, we can approximate Calculate the volume of apple, so we have

$$V = \frac{4}{3}\pi r^3$$

Here we need to determine the focal length of the camera and the distance from which we are shooting. Our focal length is selected as the standard focal length $60mm$. As a preliminary estimate, we do not consider the different positions of the apple but default to the same position of the apple, recorded as D, then we have

$$V_{true} = \frac{\eta^3 D^3}{f^3}\frac{4\pi}{3}r^3$$

Therefore, the mass of the apple is estimated to be

$$m = \rho V_{true} = \frac{4\eta^3 D^3}{3f^3}\pi\rho r^3$$

Next, we need to estimate a reasonable ratio $\frac{D}{f}$ to help us improve accuracy. In fact, the scale of a pixel on a two-dimensional image corresponds to $0.04cm$ in reality (we record this ratio as η), then we can use the area of the two-dimensional image as a rough distance estimate, we A two-dimensional image with a size of 270×185 was segmented, and the result is shown in the figure below



Figure 26 The segmented image

In this segmented picture of 7×9 that is all apples, it is not difficult to estimate that one apple occupies about the size of one and a half squares, so its corresponding image diameter can be estimated to be $1.8cm$ and the actual diameter is $9cm$, so

$$\frac{D}{f} = \frac{9}{1.8} = 5$$

Therefore, the mass of apples is estimated to be

$$m = \frac{4}{375}\pi\rho r^3$$

Using the above formula, we obtain the quality histogram of apple quality, as shown in the Fig.27

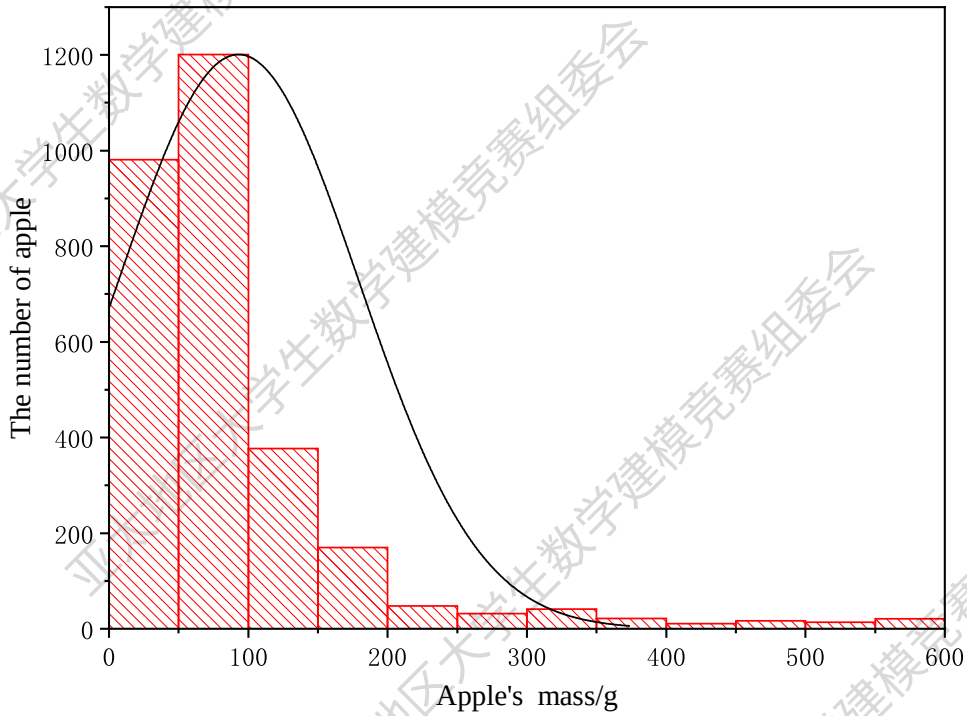


Figure 27 Mass distribution histogram of apples

As can be seen from the image, this estimation result is very inaccurate, so we need to improve our estimation model.

8.2 Radance and radiation intensity

As we all know, radiance L_e is a physical quantity used to characterize the radiant flux per unit area and unit solid angle of a surface radiation source. In addition, in sensors (cameras, human eyes, etc.), radiance is the most commonly used to describe the perceived radiation. Quantity is also the most important quantity.

$$L_e = \frac{d^2\Phi_e}{\cos\theta dA d\Omega}$$

In the formula, e is the specified direction, Φ_e is the radiation flux in the specified direction, dA is the area element, $\cos\theta$ is the normal angle of the area element, Ω is the solid angle.

Now, let's consider the situation where a digital camera takes a picture. Let's assume that the area element at the distance r is dA, the radiance is L, the light-receiving area (that is, the area of the digital camera) is S, the focal length of the camera is f, and the imaging The generated irradiance is E_{img}. Then, the solid angle formed on the camera surface element can be written as

$$d\Omega = \frac{dA}{r^2}$$

The radiation flux passing through the camera is

$$d\Phi = d\Omega LS = \frac{LS}{r^2}dA$$

Let the image size of surface micro-elements on the camera be dA', then we have

$$dA' = \frac{f^2}{r^2}dA$$

Therefore, the irradiance imaged by the surface elements is

$$E_{img} = \frac{d\Phi}{dA'} = \frac{LSdA}{r^2} \cdot \frac{r^2}{f^2 dA} = \frac{LS}{f^2} \tag{8}$$

The formula (8) indicates that the imaging irradiance E_{img} of the object will not change with the change of distance d (here we default to the aperture setting, exposure time and other parameters of the camera every time a photo is taken settings are consistent).

8.3 Establishment and calculation optimization of ranging model

The color space we selected previously is the Lab color space, and Lab happens to use numerical values to fit the color space of human visual perception, where the L component corresponds to the brightness perceived by the human eye. Based on this, our team proposed a method to estimate the distance of objects in two-dimensional images by using the L component in the Lab color space.

The basic steps are as follows:

• converts the L component map into a grayscale image, and its value can approximately represent the brightness of the pixel.

• frames the position of the apple in the grayscale image

• accumulates the grayscale values of the pixels in the frame and averages the area. This average value can be regarded as the average brightness of the object H, which is proportional to the radiation intensity of the object E. You may wish to remember the ratio The coefficient is l

• The radiation intensity satisfies the inverse square law with the same distance, and the actual distance of the object can be estimated.

Denote the gray value matrix of each pixel in the two-dimensional grayscale image as $G_{m\times n}$, where $G(i,j)$ represents the ith row and jth column of the grayscale image. For the gray value of the pixel, record the frame range of the i $(i = 1, 2, \cdots, k)$ th apple on the image as the restriction matrix $X_{k\times 2}, Y_{k\times 2}$ $(X_{k1} < X_{k2},\ Y_{k1} < Y_{k2})$. Therefore, the average brightness corresponding to the i th apple on the grayscale image can be expressed as

$$W_i = \frac{\sum_x \sum_y G(x,y)}{\eta^2 (x_{i1} - x_{i2})(y_{i1} - y_{i2})}, \quad i = 1, 2, \cdots k \tag{9}$$

s.t:

$$x \in [X_{i1}, X_{i2}], \quad y \in [Y_{i1}, Y_{i2}]$$

Then according to the definition of radiation intensity, we have

$$lW_i = E_i = \frac{d\Phi_i}{dA_i} = \frac{LS}{\eta^2 r^2}$$

that is

$$r_i = \frac{1}{\eta}\sqrt{\frac{LS}{lW_i}}, \quad i = 1, 2, \cdots, k$$

The radiance L in the formula is a physical quantity that does not change with distance, so the distance is a function of the average brightness. The lower the average brightness, the farther the distance, which is in line with our intuitive judgment of the image.

Since we cannot obtain the values of parameters L, S, l, we hope to select a standard image (required to contain only one apple and the identified apple position is highly consistent with the position of the image), and the distance between the apples in the image Denoted as $r_0 = 1m$, the two-dimensional image area is S_0, and the real area is A_0, so we have

$$r_i = \frac{r_i}{r_0} = \sqrt{\frac{LS}{lW_i} \cdot \frac{lW_0}{LS}} = \sqrt{\frac{W_0}{W_i}} \tag{10}$$

This is the final result of the ranging model. We eliminate the parameters that we cannot obtain through the ratio, and finally obtain the relationship between ranging and the average brightness of the two-dimensional plane image, which will help us complete the ranging task.

In order to reduce the complexity of the algorithm and optimize the model, we also need to optimize the calculation formula of average brightness. To this end, we introduce the Hadamard product operation of the matrix and record "$*$" as the Hadamard product operator. In addition, we record the norm $||A|| = \sum A$ to represent the calculation of the elements of the matrix A and.

Now let's improve the summation operation of the gray value matrix. We set the "0-1" column vectors in the x, y direction as α, β. Their definitions are:

$$\alpha_i(t) = \begin{cases} 1, & if \quad x_{i1} \le t \le x_{i2} \\ 0, & others \end{cases} \qquad t = 1, 2, \cdots m$$

$$\beta_i(t) = \begin{cases} 1, & if \quad y_{i1} \le t \le y_{i2} \\ 0, & others \end{cases} \qquad t = 1, 2, \cdots n$$

Among them, $\alpha_i(t), \beta_i(t)$ respectively represent the value of the tth row of the "0-1" column vector of the ith apple. We might as well remember the range matrix as

$$R_i = \alpha_i \beta_i^T$$

It is worth pointing out that the range matrix R_i generated by α_i, β_i ,itself is also a "0-1" matrix containing only $0, 1$ elements. Therefore, the summation operation of the gray value matrix can be rewritten as a matrix The operation of - Hadamard product operation, that is

$$\sum_x \sum_y G(x, y) = ||G * R|| \tag{11}$$

It is not difficult to find that x, y here no longer need to limit its value range but can traverse the entire gray value matrix. We do not have to worry that this will cause the algorithm to slow down, because this is essentially a basic matrix operation.

As an example, we might as well consider the gray value matrix $G_{4\times4}$ of a two-dimensional grayscale image of 4×4 as

$$G = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

The frame range of an apple above it is $[1, 2] \times [2, 3]$, then

$$\alpha = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T$$
$$\beta = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$$

Then the range matrix is

$$R = \alpha\beta^T = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So the result is

$$||G * R|| = \left|\left|\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right|\right| = \left|\left|\begin{bmatrix} 0 & 2 & 3 & 0 \\ 0 & 6 & 7 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right|\right| = 18$$

Next, we optimize the area calculation for the framed range. Note that the area S of the framed range can be expressed as

$$S_i = \eta^2(x_{i1} - x_{i2})(y_{i1} - y_{i2}) = \eta^2||\alpha_i\beta_i^T||$$

Therefore the area column vector is

$$S = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_k \end{bmatrix} = \eta^2 \begin{bmatrix} ||\alpha_1 \beta_1^T|| \\ ||\alpha_2 \beta_2^T|| \\ \vdots \\ ||\alpha_k \beta_k^T|| \end{bmatrix}$$

Then

$$r_i = \sqrt{\frac{W_0}{W_i}}, \quad i = 1, 2, \cdots, k$$

In

$$W_i = \frac{||G * R_i||}{S_i} = \frac{||G * R_i||}{\eta^2 ||\alpha_i \beta_i^T||}, \quad i = 1, 2, \cdots, k$$

That is

$$r = \eta \sqrt{\frac{||\alpha \beta^T||}{||G * R||} W_0} \tag{12}$$

This is the optimized relative distance calculation formula, which can be used to approximately represent the real distance. After successfully solving the problem of different fruit distances, we will next complete the calculation of the actual area and mass of the fruit.

8.4 Estimate of actual fruit size and mass

Suppose the focal length of the camera is f, the distance of the fruit is r_{true}, and the area of the two-dimensional plane is S. According to the imaging and area similarity laws, the relative actual size A_{rel} can be obtained for

$$A' = \frac{r_{true}^2}{f^2} S$$

The reason why the above formula is a relative area is that the above formula is equivalent to

$$\begin{aligned} A' &= \frac{r^2 r_0^{-2}}{f^2 r_0^{-2}} S \\ &= \frac{r_{rel}^2 A_0}{S_0} S \end{aligned}$$

That is

$$A_{rel} = \frac{A'}{A_0} = r_{rel}^2 \frac{S}{S_0}$$

Substituting the previously obtained formula into, we have

$$A_{rel} = \frac{W_0 \eta^2}{f^2} \frac{||\alpha \beta^T|| S}{||G * R|| S_0} = \frac{1}{f^2} \frac{||\alpha \beta^T||}{||\alpha_0 \beta_0^T||} \frac{W_0}{W_i}$$

This is the area enclosed by the box. Considering that the box is basically tangent to the edge of the apple, and the apple can be approximately regarded as a circle, we need to measure the relationship between the area of the box and the area of the fruit. To this end, we introduce The new parameter δ helps us accomplish this task, which is defined as

$$\delta = \frac{\pi r^2}{S} \approx \frac{\pi r^2}{4r^2} = \frac{\pi}{4}$$

That is

$$A = \delta A_{rel} = \frac{\pi}{4f^2} \frac{||\alpha\beta^T||}{||\alpha_0\beta_0^T||} \frac{W_0}{W_i}$$

Then, for the mass calculation of apples we can make the following estimate

$$m = \frac{4\pi\rho}{3}\sqrt{(\frac{A}{\pi})^3}$$
$$= \frac{\pi\rho}{6f^3}\left(\sqrt{\frac{||\alpha\beta^T||W_0}{||\alpha_0\beta_0^T||W_i}}\right)^3$$

In this way, the estimate of quality can be calculated entirely from the brightness values on the two-dimensional plane.

8.5 Results and mass distribution histogram

Figure 28 Mass distribution histogram of apples

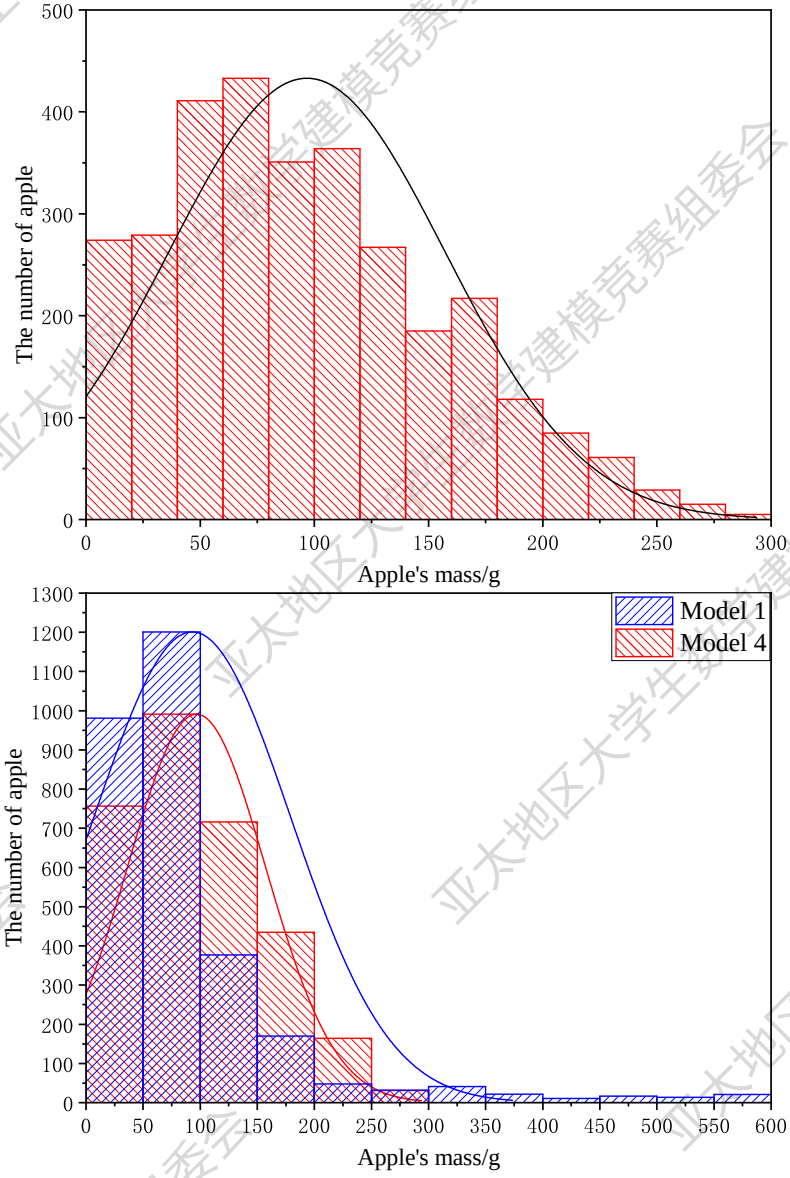


Figure 29 Mass distribution histogram of apples

From the comparison chart between Model 1 and Model 4, we can know that the model results we established are quite good, which also confirms the rationality of our modeling.

IX. Model Establishment and Solution of Problem 5

After consulting the relevant literature, we can learn that YOLOv4 is often used for target detection, while VGG19 is often used for image classification. Both belong to convolutional neural networks (CNN).

So we choose VGG19 for solving Problem 5.

9.1 Establishment of an improved apple recognition model

9.1.1 *Principle Overview of VGG19 network*

One of the most famous object recognition networks at the time was called VGG19. Each layer of the neural network would use the output of the previous layer to further extract more complex features until it was complex enough to be used to identify objects, so each layer Each layer can be regarded as an extractor of many local features.

The following is the principle description of VGG19:

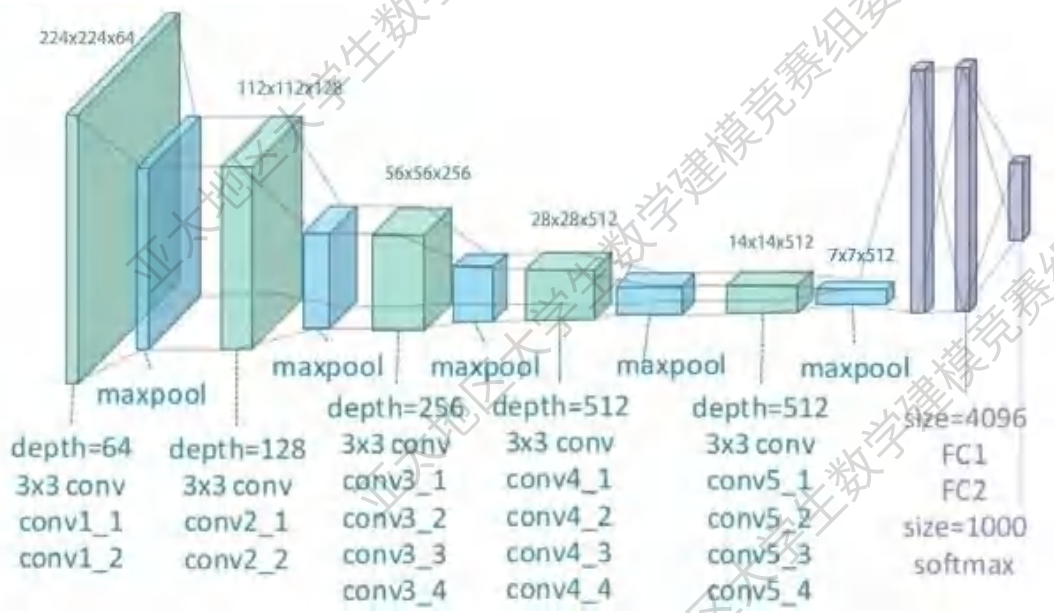


Figure 30 Partial structural diagram of VGG19(Image Sourced from the internet)

VGG19 network model, which includes 5 convolution modules, 5 pooling layers, 3 fully connected layers and 1 softmax classifier. Among them, 5 convolution modules contain a total of 16 convolution layers. The convolution kernel size of the VGG19 network model is 3×3, and the step size is 1.

The core operation of the convolutional neural network is the convolution operation, which is shown in Figure 2. In Figure 2, a 3×3 convolution kernel slides on the original image, and the covered value is multiplied by the corresponding position of the weight on the convolution kernel

and then added. The obtained value is filled in the corresponding position of the feature map, that is, extracted The features of the original image are reflected on the feature map.

Each convolution kernel extracts one feature and generates one feature map. Then, by stacking 64 feature maps of 64 convolution kernels, a three-dimensional feature map will be obtained. Convolution kernels in different convolution layers are used to extract different feature information. Shallow convolution kernels mainly extract pixel features such as color, edges and patches; middle-level convolution kernels mainly extract texture features such as stripes and shapes; high-level convolution kernels Extract more abstract semantic features such as text.

In order to compare the quality of image stitching, this article selects the features extracted from the 1st to 5th convolutional layers in the VGG19 network model as the main comparison features.

After each layer of convolution operation is completed, a three-dimensional feature map stacked by two-dimensional feature maps will be obtained. The three-dimensional feature map is shown followed.

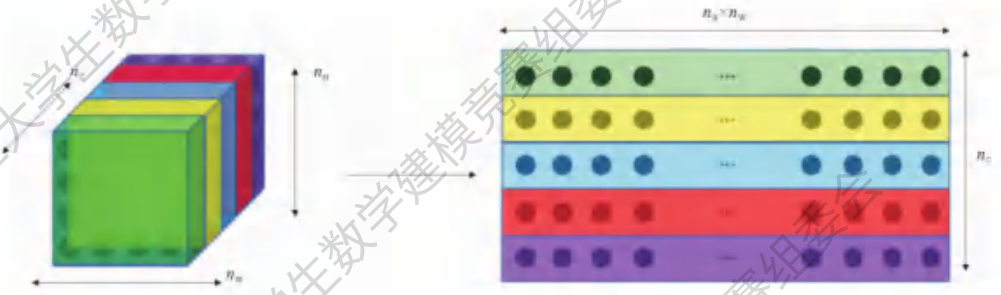


Figure 31 Three-dimensional feature map transformation (Image Sourced from the internet)

Each color of the feature map represents a two-dimensional feature map obtained by convolution with a convolution kernel. This article takes the fifth layer as an example for analysis. 256 convolutions After the kernel completes the convolution operation, a 56×56×256 feature map is obtained. In the figure, nW = 56, nH = 56, nC = 256. Analyze the second layer of the first layer separately. dimensional feature map, which can be expressed as a 56×56 matrix F_1

$$F_1 = \begin{bmatrix} a_1 & a_2 & \dots & a_{56} \\ a_{57} & a_{58} & \dots & a_{112} \\ & & \vdots & \\ a_{3081} & a_{3082} & \dots & a_{3136} \end{bmatrix}$$

Similarly, define the matrix of the second layer feature map as F2, and so on. Each feature map corresponds to a matrix, and the last feature map corresponds to a matrix. F256. Then each matrix is flattened into a row vector, that is, the matrix F1 is transformed into

$$H_1 = \begin{bmatrix} a_1 & a_2 & \dots & a_{3136} \end{bmatrix}$$

Then the three-dimensional feature map of the convolutional layer will be transformed accordingly.

The entire feature map can be expressed as a 3136×256 matrix K, that is

$$K = \begin{bmatrix} H_1 & H_2 & \dots & H_{255} & H_{256} \end{bmatrix}^T$$

At this time, the Gram matrix can be expressed as

$$G = KK^T$$

In the formula, KT is the transposed matrix of matrix K. Obviously the matrix G is a 256×56 symmetric matrix, which represents the 256 feature maps of the fifth convolution layer. The co-occurrence correlation between the 256 features in the table can be used to represent the style of an image, and then the style of the two images can be represented by the co-occurrence correlation. The degree of difference determines the quality of image splicing. The function Ls is used as an index to compare the degree of style difference between two images, that is

$$\begin{cases} L = \sum_{k=0}^{m} w_k E_k \\ E_k = \dfrac{1}{4N_k^2 M_k^2} \sum_{i,j} (G_{i,j}^k - A_{i,j}^k)^2 \end{cases}$$

G^k is the Gram matrix of the original image in the k convolution layer; A^k is the Gram matrix of the spliced image in the k convolution layer; N^k is the 1st layer of the two images. The number of channels of the feature map; M^k is the product of the length and width of the k-th layer feature map of the two images; wk is the weight of the k-layer image.

9.1.2 *The Application of the VGG19 network*

We apply the principle of this network, take all the apple images provided in Appendix 2 as samples, use the VGG19 network for deep learning, and use the trained model for detection in Appendix 3.

Here we use MATLAB for simulation, and the recognition results are as follows:



(a) The recognition result is tomato (b) The recognition result is apple

Figure 32 Apple recognition result window after VGG19 training

9.2 Results and histogram for all Apple image ID numbers

The results of the model we trained through VCG19 for recognition are as follows: The abscissa is the ID of the image. Starting from the first one, we take every 1,000 pictures as a grid. The ordinate is the number of images identified as apples in these 1,000 pictures. The result should be between 0 and 1,000. After comparison, it is found that all All comply.

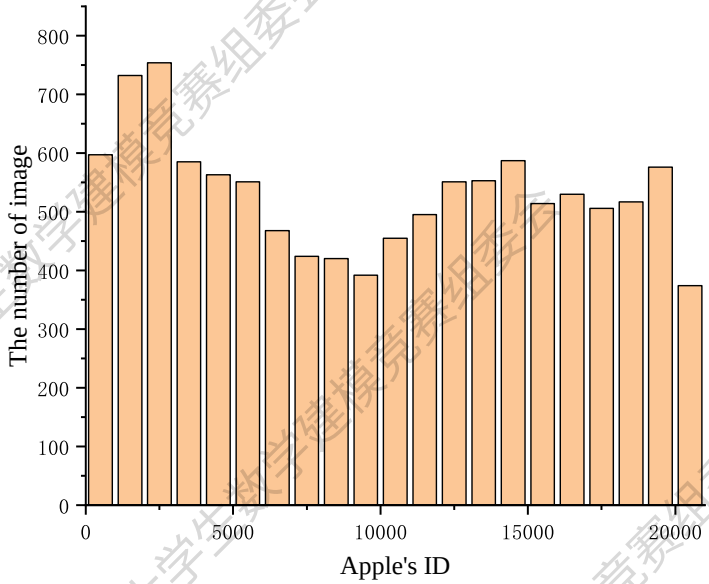
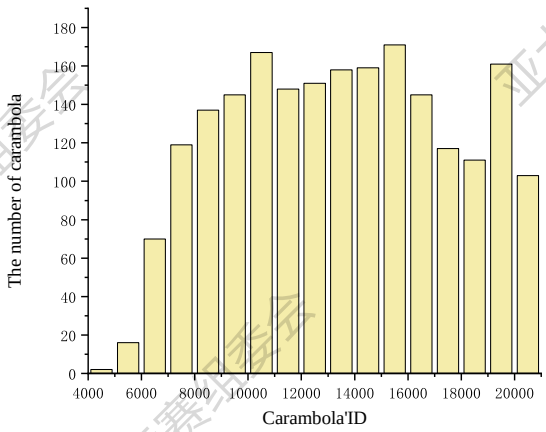
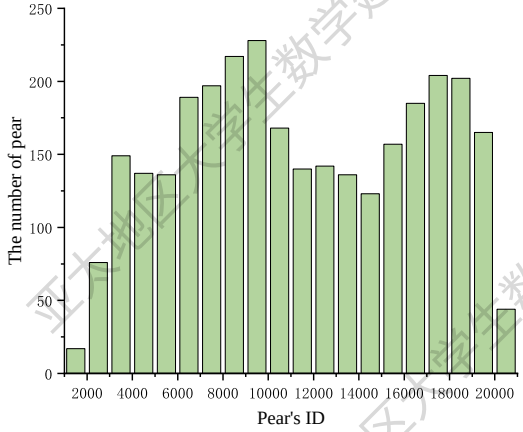


Figure 33 Histogram for all Apple image ID numbers

According to the model we trained, not only apples can be recognized, we also recognized other fruits and drew the distribution histogram of their image ID numbers. The results are as follows:





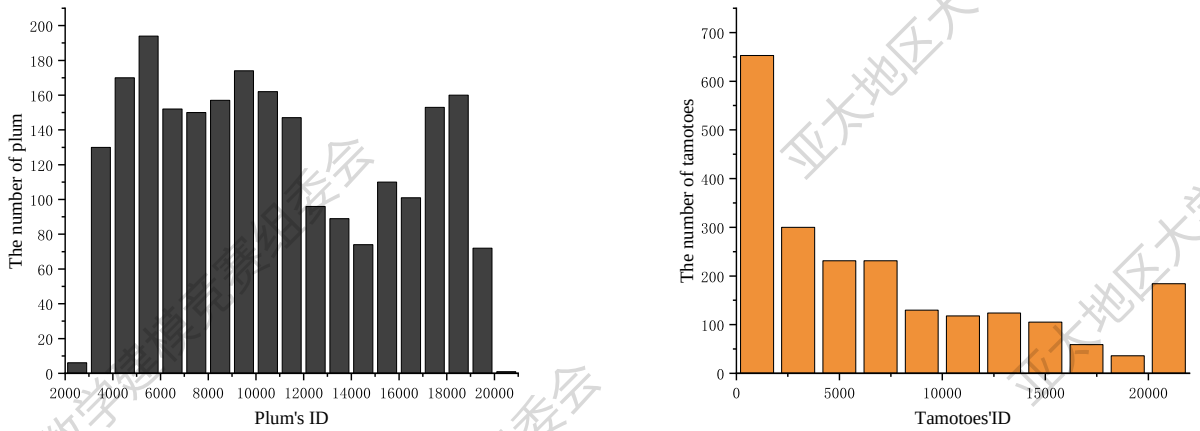


Figure 34 Histograms for other fruits' image ID numbers

X. Model Evaluation and Further Discussion

10.1 Merit

● **Model for Problem 1**'s advantages are that the algorithm is fast and time-consuming, it can accurately recognize images with obvious apple features and the drawn circular boundary is highly consistent with the edge of the apple;

●Compared with the deep learning algorithm, the algorithm does not require pre-training. It only needs to adjust the parameters to directly detect and identify apple images. This is more suitable for scenes that require temporary detection;

●The output of the algorithm also includes the center coordinates and radius parameters of the detected circle, which facilitates our team to establish a coordinate system and solve the following problems.

● **Model for Problem 2 and 3** Using the YOLOv4 network, the training speed is fast and the model memory footprint is small

● **Model for Problem 4** is derived through strict mathematical means. From a theoretical point of view, the model is scientific and reliable enough. In addition, we also made full use of the results of the problem 1 model to test the model. The result is that the speed and accuracy of the model have been improved compared with model 1;

● In this model, we fully considered the large and small situations in the imaging image, and innovatively proposed a method of estimating image distance using image brightness - the brightness estimation method. The final results show that this method is very effective. Reasonable and effective.

● **Model for Problem 5**Using the VGG19 network, the model has a high training depth and a large number of parameters. After training, it can achieve high accuracy.

10.2 Shortcoming

● **Model for Problem 1** relies heavily on the color features and shape features of apples and is more sensitive to the recognition of color features. This will cause some things with similar color features but not apples to be mistakenly detected as apples. At the same time, the recognition effect of green apples is poor;

• The parameter selection of this model is a matter of high trial and error cost, and the adjustment of parameters is even more laborious. Our team spent a lot of time adjusting appropriate parameters to make the detection results as possible as possible. Good.

• **Model for Problem 2 and 3**The training depth of YOLOv4 network is not enough, and misjudgments may occur.

• **Model for Problem 4**The derivation of this model contains some approximate estimates, which will reduce the accuracy of the model's calculation results;

• The establishment of the model requires many conditions. For example, it is assumed that the imaging is all natural light and the influence of other light sources is excluded. At the same time, the accuracy of the model is also affected by the accuracy of the previous Apple detection model.

• **Model for Problem 5**The VGG19 network model is relatively complex, takes a long time to train, is difficult to adjust parameters, has large storage capacity, and has high requirements for GPU performance.

10.3 Improvement Measures

• **Model for Problem 1** But in fact, we can also enhance the sensitivity of the algorithm to the apple shape characteristics, that is, the accuracy of the test, by correcting the detection results of the center of the circle multiple times. We used the gradient method in this article, so we can correct the circular detection results and reduce the false detection rate by drawing a circle with three points.

• **Model for Problem 4** In the model, we assume parameters such as focal length to help our team build the model, which is also the core of the entire model. In real life, when we take an image, we can record the focal length of the image at the same time, so we can perform separate calculations for each image, which makes the results more accurate. In addition, we can also keep the shooting distance as consistent as possible and the lighting conditions equal as much as the error allows. Under such actual conditions, the accuracy of the model will be greatly improved.

XI. References

[1] Li Changlu. Research on identification and positioning of apple picking robots based on machine vision [D]. Lanzhou University of Technology, 2023. DOI: 10.27206/d.cnki.ggsgu.2022.000718.

[2] Zhang Junkai, Li Xin, Han Junxian, etc. Research on optimization of picking robot system under deep learning algorithm [J]. Agricultural Mechanization Research, 2024, 46(04): 58-62. DOI: 10.13427/j.cnki. njyi.20230728.001.

[3] Yang Chunyan. A circle detection method based on Hough transform [J]. Journal of Baicheng Normal University, 2016, 30(02): 65-68.

[4] Wang Yong, Tao Zhaosheng, Shi Xinyu, etc. Apple target detection method with different maturity levels based on improved YOLOv5s [J/OL]. Journal of Nanjing Agricultural University: 1-13[2023-11-26].http:// kns.cnki.net/kcms/detail/32.1148.S.20230926.1201.002.html.

[5] Ma Fangda, Liu Zeping, Chen Shihai, etc. Image stitching quality evaluation algorithm based on VGG19 convolutional neural network [J]. Journal of Qingdao University (Engineering Technology Edition), 2023, 38(02): 31-36. DOI:10.13306/j.1006-9798.2023.02.005.

[6] Xu Weisen. Research on target recognition algorithm for agricultural picking apples[J]. Electronic World, 2018(14):44-45.DOI:10.19353/j.cnki.dzsj.2018.14.017.

[7] Liu Xiaoyang. Research on image recognition algorithm of multifunctional fruit and vegetable picking robot[D]. Jiangsu University, 2021.DOI:10.27170/d.cnki.gjsuu.2020.000972.

亚太地区大学生数学建模竞赛组委会

Team Number:	apmcm2307379
Problem Chosen:	B

2023 APMCM summary sheet

In this study, with the aim to optimize the temperature and wind speed distribution in a glass greenhouse to promote crop growth, we addressed four key issues step by step.

To solve the first problem, we built mathematical equations under crop-free conditions, based on the fundamental physical principles, with Python and a computational fluid dynamics software (CFD), ANSYS Fluent. The temperature and wind speed distribution at 0.5 m was further validated by the numerical simulation results from ANSYS Fluent, with an indicator (R) showing the overlapping interval, achieving 89.84% and 93.75% respectively.

For problem two, we improved the current model by adding a porous media region representing crops. The consideration on its porosity and permeability ensured us the effectiveness of R more than 89% in simulating the temperature and wind speed distribution at specific heights (0.5 m and 0.1 m) in the greenhouse, compared with results from the CFD software. An initial claim on the unsuitable conditions for crops growth was made.

For two scenarios in problem three, we adjusted the relevant parameters via the two methods, simulating the effect of individually changed fan parameter on the two target fields. Compared to results in problem two, we deducted that increasing velocity in canopy would achieve warmer living conditions in the canopy, on the contrary, lower position of the fans benefits roots for better uniformity.

In order to solve problem four, we applied genetic algorithm to optimize the number, location, speed and temperature in designing fans. With the length magnification, the constrains on number and locations of the fans and the convergence guaranteed by the Markov Chain, we could obtain relatively random schemes while making sure comfortable living conditions for crops. Two schemes including five fans and three fans were given respectively.

In summary, our research not only established an effective model to simulate the environmental conditions in the glass greenhouse, but also optimized the fan design through a genetic algorithm, providing scientific guidance and optimization solutions for greenhouse design and crop planting.

Keywords: Greenhouse simulation ANSYS Fluent Genetic algorithm

Contents

1. Introduction **1**
1.1 Natural ventilation 1
1.2 Mechanical ventilation 1
2. Restatements and Analyses of problems **1**
2.1 Restatements 1
2.2 Analyses 2
3. Models **2**
3.1 Terms, Definitions and Symbols 3
3.2 Assumptions 3
3.2.1 *Model for transferring energy* 4
3.3 Establishment of the CFD model 4
3.4 Evaluation 5
4. Solutions and Results **5**
4.1 Question one 5
4.2 Question two 8
4.2.1 *Results* 8
4.2.2 *Analyses* 10
4.3 Question three 11
4.3.1 *Results and comparisons* 11
4.3.2 *Validations* 13
4.4 Question 4 14
4.4.1 *Select Crops* 15
4.4.2 *Fans optimization by the Genetic Algorithm* 16
4.4.2.1 Innovative points 17
4.4.2.2 Results 17
5. Pros and Cons **17**
5.1 Pros 17
5.2 Cons 19
6. References **21**
7. Appendix **22**

I. Introduction

Temperature, humidity and wind speed play significant importance in crop yields in modern greenhouses. To achieve the suppression of high temperature, improvement of air components, and decrement of humidity inside the greenhouse, there are two kinds of ventilation systems.

1.1 Natural ventilation

In natural ventilation systems, air flow is respectively promoted by the different temperature and different wind velocity between the inside and outside of the facility , which requires low investments but also only has limited capabilities. In summer, the greenhouse tends to rely on high-speed winds to generate horizontal air flow, while in winter, heat is more likely to be utilized with vertical air flow determined by heights between the up outlet and down inlet.

1.2 Mechanical ventilation

The mechanical ventilation systems contain two type of patterns, taking air in or exhaust.The intake ventilating system has low requirements on air tightness, could resist contamination from outside pollutants that may impair inner environment, but has a varying velocity field. On the contrary, the exhaust ventilating system provides a relatively uniform distribution of wind velocity, requires good seal, but may result in contamination from outside. The position of outlet in a mechanical system could be the at top and bottom of the greenhouse, as well as the in horizontal and longitudinal direction of the greenhouse.

II. Restatements and Analyses of problems

2.1 Restatements

Here, in order to regulate climate factors mentioned above, ventilation systems with greenhouse fans in breadth-wise are used, of which the position and the velocity speed affect the distribution and uniformity of the velocity field and the temperature field inside the greenhouse. The inlet fan blows in warm air at 40°in the horizontal direction with an average velocity of 2 m/s. The outer glass and bottom soil of the greenhouse are

set as wall conditions, primarily exchanging energy with the entire greenhouse through convective heat transfer and conduction.

Without considering external factors such as doors, drafts, solar radiation, and other environmental factors, the greenhouse that is sealed and placed indoors, should be clearly given the velocity and temperature field without (**Question one**) at a height of 0.5 meters and with crops (**Question two**) at a height of 0.5 meters and 0.1 meters respectively, analyzing the conditions for crop growth.

Further, by increasing the velocity of warm air inlet from 2 m/s to 3 m/s, and decreasing the position of the greenhouse fan by moving it from 1.3 m to 1 m, the temperature and wind speed distribution inside the glass greenhouse should be provided (**Question three**) and made comparisons with results in question two. Based on the performance of the adjustment, the number of greenhouse fan, location, wind speed, blowing temperature, specifications and different crops and other factors could then be optimized (**Question four**) for a better uniformity.

2.2 Analyses

As mentioned in the problem statements, the convective heat transfer, conduction and the aerodynamics should be considered to figure out the temperature and wind velocity field. Thus, to clearly show the two features at different heights in the greenhouse, we girded the cuboid-shaped greenhouse to limited amount of points and applied the law of conservation of energy for passing the velocity and temperature at grid point in the greenhouse with Python. To demonstrate the accuracy of our results with **Python**, we further built the 3-dimensional model in a computational fluid dynamics software, called **ANSYS Fluent**. The consistency between the mathematical model built by Python and the numerical simulation by ANSYS would then validate the reliability of the parameters optimized by our mathematical model.

III. Models

The fundamental principles and parameters that govern the velocity and temperature fields are consistent in all problems we face, which are presented in this section.

Table 1 Descriptions of Symbols

Symbols	Definitions	units
T	Temperature	°C
t	Time	s
u	Velocity vector	m/s
α	Thermal diffusivity	
p	Pressure	Pa
μ	Dynamic Viscosity	Pa ·s
Δ	Variation of certain variable	
D	Euler distance between two spatial points	m
Re	Reynolds number	
φ	Porosity	
k	Permeability	m^2

3.1 Terms, Definitions and Symbols

3.2 Assumptions

(1) The air can be considered as an incompressible fluid, and is evenly mixed in thermal equilibrium at any time.

(2) The heat transfer between greenhouse glass and soil is mainly convection and conduction, ignoring thermal radiation in our models.

(3) There is no additional heat source in the greenhouse, except for the hot, stable and uniform air blown in by the fans.

(4) The turbulence effect of air is ignored.

(5) The actual volume of crops accounts for 50% of the total volume of crops cuboid.

(6) The diameter of the grain of the porous media, and here we assume it to be the half of the height of crops.

3.2.1 *Model for transferring energy*

Based on the law of conservation of energy, a fundamental principle in physics that states that the total energy of an isolated system remains constant over time, we considered the steady-state flow with Eq. (1) [1]. Since the problems we face are non-instantaneous, without the consideration of time, in the stable stage of heat balance, we simplified the heat transferred at certain point N to Eq. (2), a function of the Euler distance between the point N and the center of the fans. Similarly, in the stable stage of velocity field, we simplified the velocity transferred at point N to Eq. (3).

$$\Delta U = Q - W \tag{1}$$

$$e^{-decay\ controller\ _T \times D_N} = Q - W \tag{2}$$

$$V_0 e^{-decay\ controller\ _V \times D_N} = V \tag{3}$$

where, $Q - W$, the difference of energy between inside and outside, is represented by the difference between initial temperature at fans and in the greenhouse. D_N denotes the distance between the center of fans and point N in the greenhouse. The decay controller is a parameter, individually adjusting the decaying effects of the temperature (T) and velocity (V) in the greenhouse. V and V_0 represents the velocity at point N and the center of the fans respectively.

By building the mathematical model at point N, we could further give the temperature field and velocity field in the whole greenhouse

3.3 Establishment of the CFD model

To test our mathematical model, we utilized ANSYS Fluent to numerically simulate the velocity and temperature field in the greenhouse. First, a 3-dimensional model of the same size as the greenhouse was established and meshed in Fig. (1).

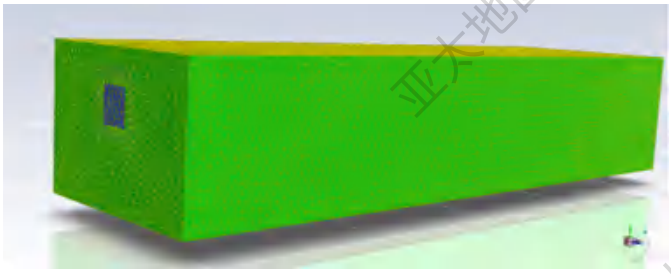


Figure 1 Grids

With the meshed grids and boundary conditions settled down, we could then simulate the target fields after the calculation of the Reynolds number in Eq. (4). The Re at the inlet is 66667, a number much larger than the smallest threshold of turbulence, 4000. So, via the finite difference method, we could simulate the realistic fields while reaching to the stable and balanced stage under the condition that air flow in the greenhouse could be regarded as turbulent motion.

$$Re = \frac{\rho V_0 L}{\mu} \tag{4}$$

where, ρ denotes the fluid density. L represents the size of the fans. μ is the dynamic viscosity of the air.

3.4 Evaluation

To evaluate the simulation accuracy between the mathematical model implemented by Python and the numerical model calculated by ANSYS, we devise a indicator, R, shown in Eq. (5).

$$R = \frac{A \cap B}{max(A, B)} \tag{5}$$

where, A and B represent the magnitude of simulation results at the same area for Python and ANSYS respectively. $A \cap B$ gives the overlapping interval, while max(A,B) is the larger interval. So, R denotes the similarity of simulation results between the two basic models, which means a high value of R suggests a good simulation accuracy, validated by both models.

IV. Solutions and Results

4.1 Question one

To realize the mathematical model we built in section 3, we first meshed grids with limited spatial points ($40 \times 20 \times 20$), corresponding to sequence of $10m \times 3m \times 2m$. The decay controller was applied to the calculation of temperature, velocity and the length of the greenhouse. Via Appendix. 1, the temperature and velocity fields in the greenhouse were shown in Fig. (2), and the distribution of wind speed and temperature at a cross-section of the greenhouse at a height of 0.5 meters were displayed in Fig. (3).

Meanwhile, the longitudinal results given by ANSYS Fluent, showing the variation of temperature and velocity in the direction of the length of the greenhouse were first

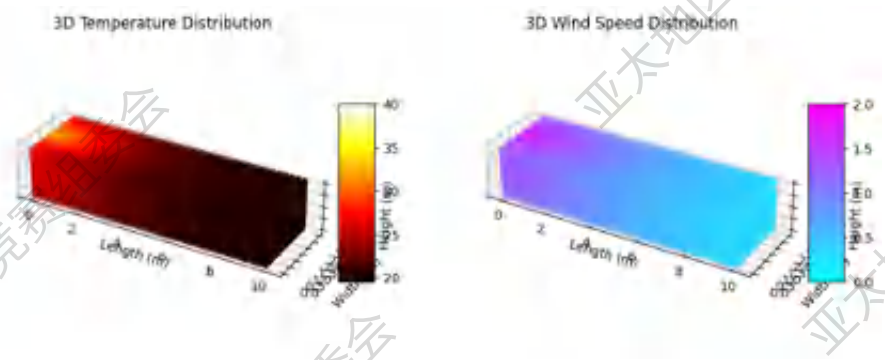


Figure 2 Three-dimensional results of the whole greenhouse in Python

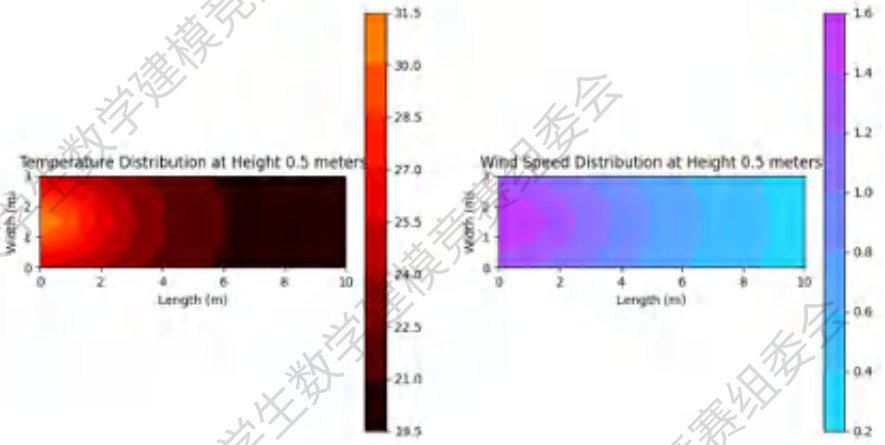


Figure 3 Results of target fields at a height of 0.5 meters in Python

exhibited in Fig. (4) and (5) respectively. Further, the distribution of wind speed and temperature at a cross-section of the greenhouse at a height of 0.5 meters were displayed in Fig. (6) and Fig. (7).

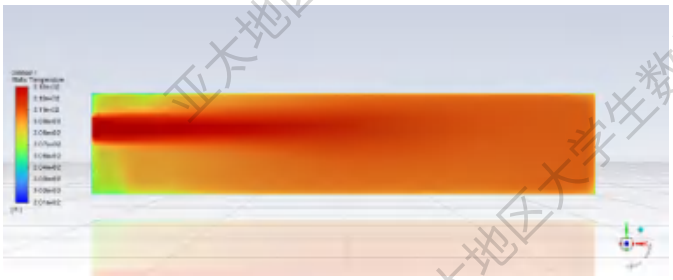


Figure 4 Longitudinal temperature simulated by ANSYS

The indicator we designed before were calculated here. The magnitude of temperature in the height of 0.5m ranged from 31.85°C to 37.85°C in our mathematical model, while that of ANSYS Fluent varied from 31.2°C to 37.6°C. So, R (6)was exemplified

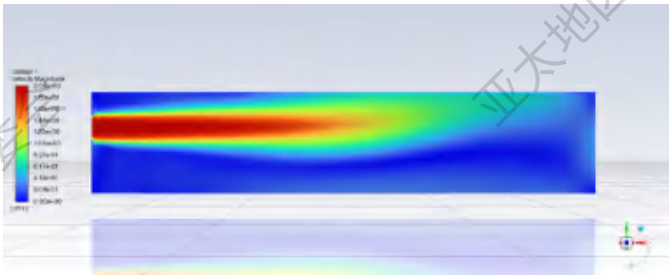


Figure 5 Longitudinal velocity simulated by ANSYS



Figure 6 Distribution of temperature at a height of 0.5 meters simulated by ANSYS



Figure 7 Distribution of velocity at a height of 0.5 meters simulated by ANSYS

in the temperature simulation.

$$R = \frac{37.6 - 31.2}{37.85 - 31.85} = 89.84\% \tag{6}$$

Similarly, we got R value of velocity, reaching 93.75%, with a range of 0-0.48m/s and 0-0.45m/s for the mathematical model and ANSYS respectively, which showed that the temperature and wind speed distributions of the two models have overlapping intervals at multiple locations. This finding demonstrates that although the mathematical model has limited ability to simulate physical details compared to CFD software, it still has some accuracy and reliability in predict the temperature and wind speed distribution in glass greenhouses. As a result, it could be a tool for quick preliminary analysis, while the numerical model via CFD software is suitable for more detailed and precise analysis. By combining these two models, it is possible to understand and optimize the climate regulation of glass greenhouses more comprehensively.

4.2 Question two

4.2.1 *Results*

Models for question two were similar to what we have shown in section 3 and section 4.1, however, differences exist and actually lie in the addition of crops[2]. So, we regarded the crop as a kind of porous media, and scaled the temperature in Eq. (2) by 0.9 and velocity in Eq. (3) by 0.5, when points were detected in the area of crops, to show the resistance of crops to the hot wind. The whole simulation results (see Appendix. (2)) were shown in Fig. (8), and the distribution of wind speed and temperature at two cross-sections within the greenhouse were shown in Fig. (9).

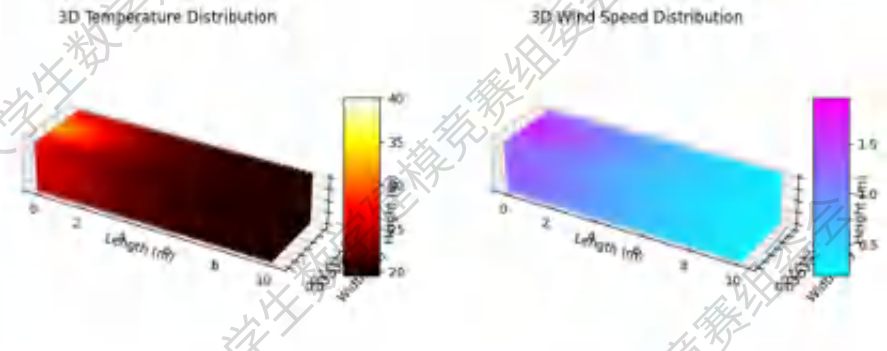


Figure 8 Three-dimensional results of the whole greenhouse with crops in Python

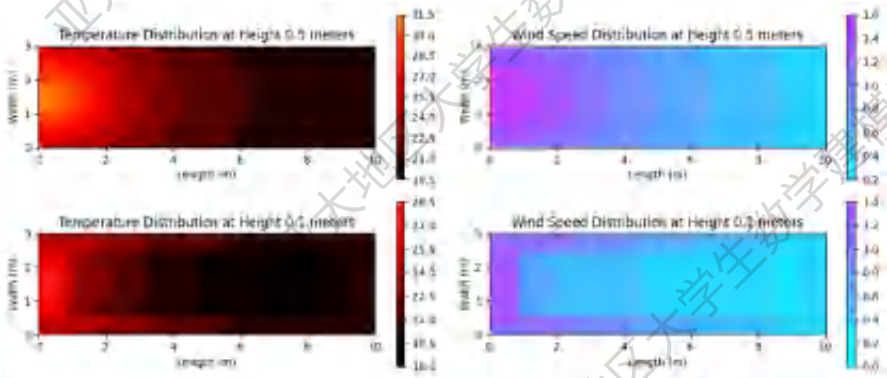


Figure 9 Results of target fields at the height of 0.5 meters and 0.1 meters in Python

Likewise, we also established the glass greenhouse with planted crops in ANSYS Fluent in Fig. (10). However, the porosity defined in Eq. (7) and permeability defined in Eq. (8) of the porous media required prior settings.

$$\varphi = \frac{V_{crops}}{V_t otal} \tag{7}$$

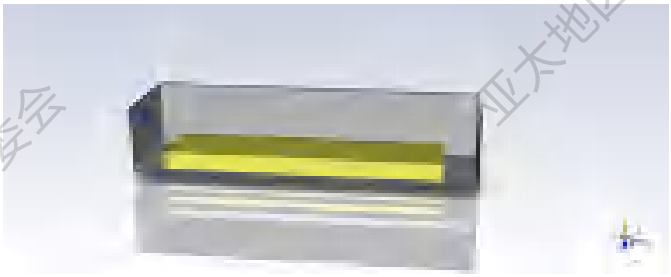


Figure 10 Three-dimensional model for the greenhouse with crops in ANSYS

where, the V_{crops} is hard to accurately assess without realistic conditions, so here in accord with the fifth assumption in section 3.2, we assumed the V_{crops} accounts for 50% of V_{total}, which ensures the value of φ to be 0.5.

$$k = \frac{d_p^2 \varphi^3}{K(1-\varphi)^2} \tag{8}$$

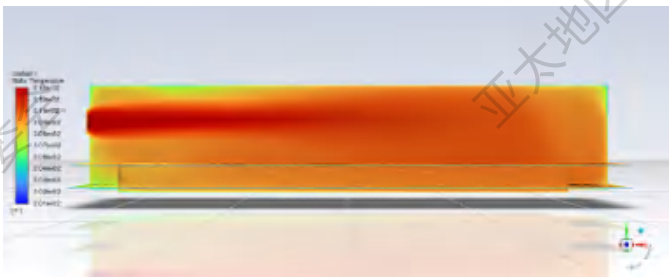
Permeability is a physical quantity that measures the flow ability of a fluid through a porous medium, typically depending on the size, shape, and distribution of pores. For a simplified crop model, we can use the Kozeny-Carman equation to estimate its permeability, where d_p is the diameter of the grain of the porous media, and here we assume it to be the half of the height of crops as the sixth assumption in this paper. K is the empirical value, ranging from 3 to 5, set as 4 here.

As a consequence, the porosity and permeability of the porous media are 0.5 and 0.0078125 m^2 by calculation, when the crops occupies a space of 0.9*0.8*0.5 in question two.

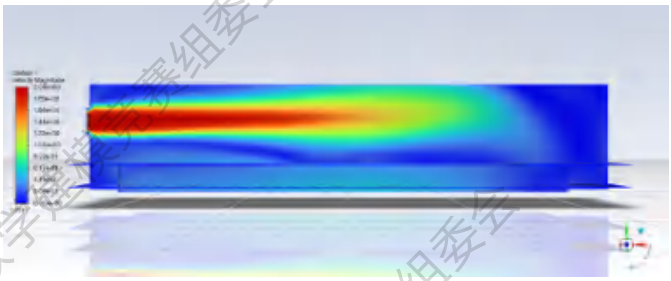
Thus, the longitudinal results given by ANSYS Fluent, showing the variation of temperature and velocity in the direction of the length of the greenhouse was given by Fig. (11), and distribution at the height of 0.5 meters and 0.1 meters simulated by ANSYS were exhibited in Fig. (12) and Fig. (13).

To assess the simulation accuracy of our mathematical model, we kept using the indicator R to show overlapping between the temperature and wind speed distribution intervals of the two models at the same location. In the height of 0.5m, the magnitudes of mathematical model were 31.2-37.6°C and 0-0.64m/s, whereas that of ANSYS were 31.85-37.85°C, and 0-0.61m/s. So, R achieved 89.84% and 95.31% for temperature and velocity simulation respectively. In the height of 0.1m, the magnitudes of mathematical model were 29-36°C and 0-0.8m/s, whereas that of ANSYS were 29.85-36.85°C, and 0-0.81m/s, with R reaching 87.86% and 98.77% individually.

In summary, the scaled model achieves better simulation of the micro-climatic conditions in the porous media region, and the two models showed great consistency,

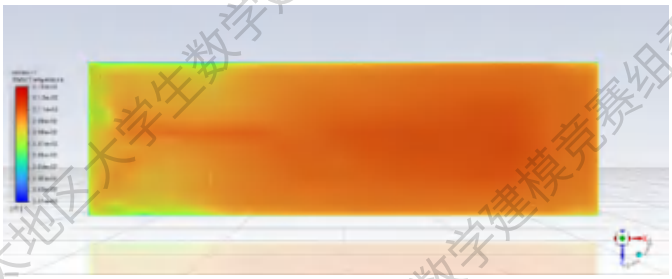


(a) Temperature fields

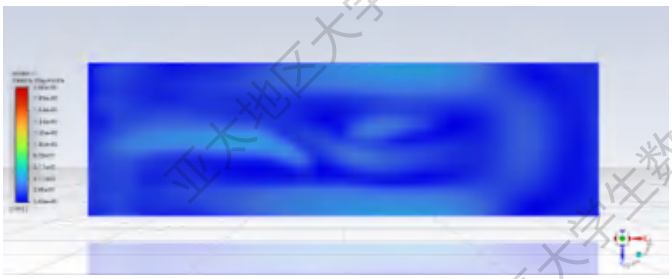


(b) Velocity fields

Figure 11 Distribution of target fields in the longitudual direction simulated by ANSYS



(a) Temperature fields



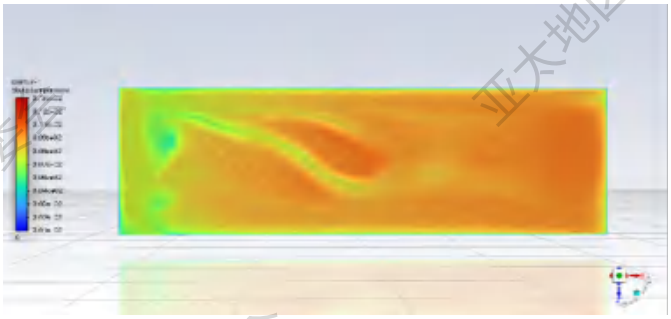
(b) Velocity fields

Figure 12 Distribution of target fields at the height of 0.5 meters simulated by ANSYS

especially in simulating the environmental conditions near crop canopy and roots.

4.2.2 *Analyses*

From the temperature fields in Fig. (9), the comfortable zone with temperature ranging from 23°C to 26°C mainly exists in the length from 2m to 6m at the canopy of



(a) Temperature fields



(b) Velocity fields

Figure 13 Distribution of target fields at the height of 0.1 meters simulated by ANSYS

crops, while in roots, it mainly appears in the length from 1m to 3m. As for velocity, the comfortable velocity at the canopy appears in the length more than 4m, while in roots it exists in areas farther than 2m in the length direction.

In summary, we could conclude that crops in this greenhouse with the fans at the height of 1.3m, blowing hot winds at the temperature of 40 $°C$, actually results in an uncomfortable living condition for both canopy and roots of crops, because their requirements of temperature and wind velocity could be satisfied.

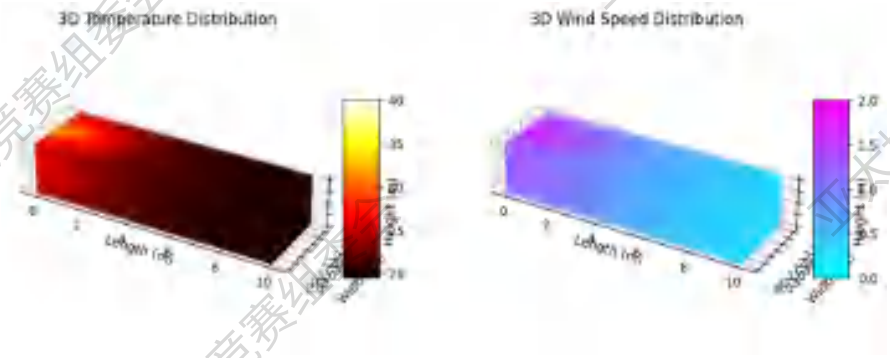
4.3 Question three

4.3.1 *Results and comparisons*

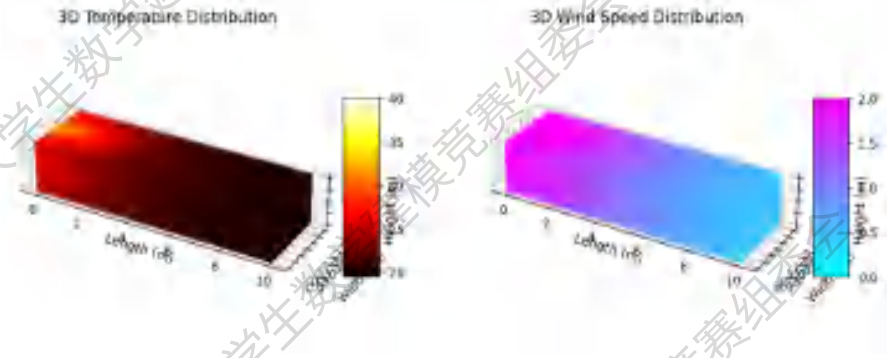
Based on the basic models build in question one and crop models established in question two, here we individually adjusted the velocity of warm air outlet from 2m/s to 3m/s, and lower the position of the greenhouse fan by moving it from 1.3 m to 1 m (see Appendix. 3)to make comparison with results in question two.

In scenario one, according to our mathematical model, when fans blew in air at the speed of 3m/s at the height of 1.3m, comparison with results in question 2 were shown in Fig. (14), where we could see that by increasing inlet velocity, although heat

did not obviously passed farther, velocity near the fans increases dramatically, and near fans areas receive more winds than ever.



(a) 2m/s in question two



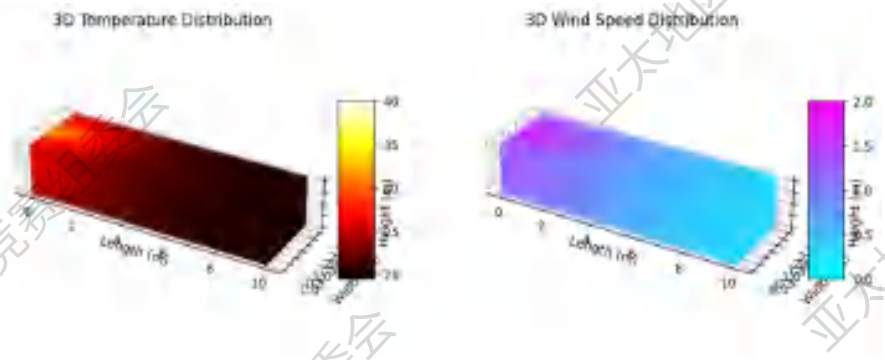
(b) 3m/s in question three

Figure 14 Three-dimensional results with different inlet velocity

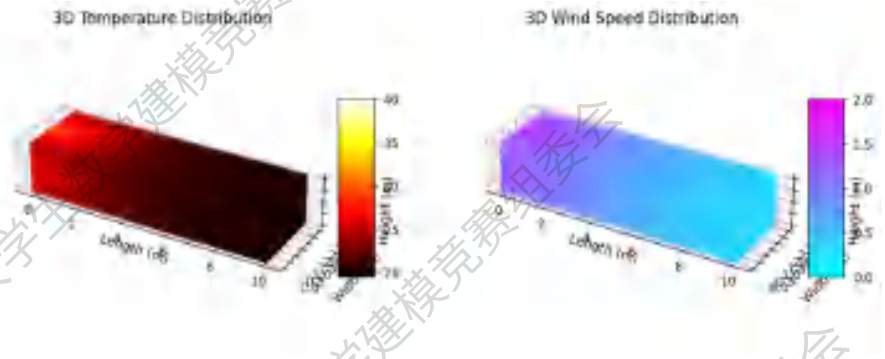
In scenario two, according to our mathematical model, when fans blew in air at the original speed of 2m/s at a lower height of 1m, comparison with results in question 2 were shown in Fig. (15), where in the height direction, we could observe a more uniform distribution, not only in temperature but also in velocity.

Meanwhile, at the canopy and roots of the crops, we could also observe from Fig. (16) and Fig. (17) that increasing wind velocity benefit canopy more than roots, because more areas in the height of 0.5m receive heat from hot wind, while roots receive a little. However, lower position of the fans could help roots more than canopy, because more areas in the height of 0.1m gets warmer and better uniformity.

In summary, from our observation, lower the position of the fans from 1.3m to 1m could achieve a more uniform distribution of fields, while increasing the velocity of air inlet make little progress in blowing wind deeper, but impairs the living condition of crops near the air inlet. In addition, increasing velocity in canopy achieves warmer living conditions in the canopy, and lower position of the fans benefits roots from better uniformity.

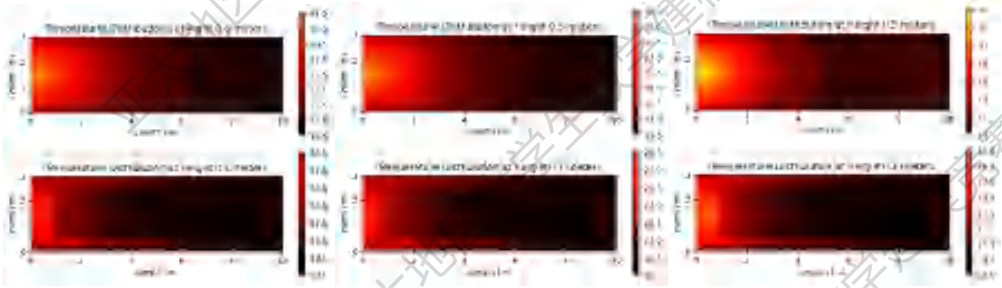


(a) 2m/s in question two



(b) 2m/s in question three

Figure 15 Three-dimensional results with different inlet position

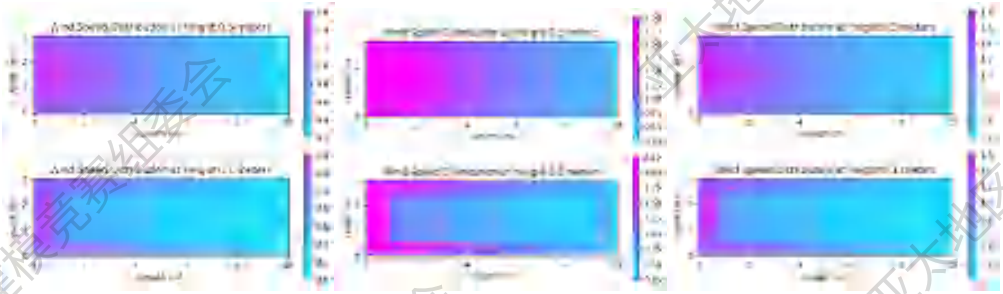


(a) question two (b) 3m/s in question three (c) 1m in question three

Figure 16 Temperature fields in different scenarios

4.3.2 *Validations*

Individually adjusting the velocity and position of the fans, we obtained target fields and made comparisons first with different scenarios and height in the area of temperature distribution in Fig. (18) and Fig. (19). It could obviously be observed that compared to increasing velocity in the height of 0.1m, lower position of the fans achieves better temperature distribution. More heat was transformed to deeper side in length direction. Similarly, compared to lower position, increasing velocity at canopy better uniforms the



(a) question two (b) 3m/s in question three (c) 1m in question three

Figure 17 Velocity fields in different scenarios

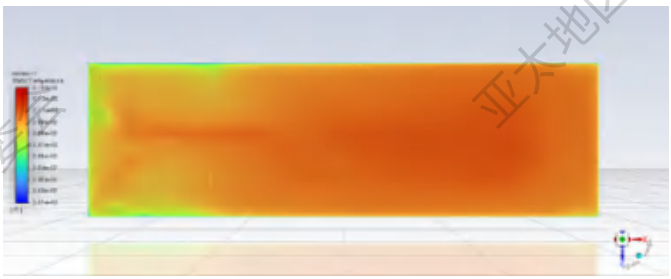
temperature distribution in the whole cuboid, which could also be concluded from the velocity field.

The consistent conclusion between mathematical model and ANSYS simulation could be quantified by our indicator R. In the height of 0.5m, the magnitudes of mathematical model were 32-38.4°C and 0-1.2m/s, whereas that of ANSYS were 32.85-38.85°C, and 0-1.22m/s. So, R achieved 86.72% and 98.36% for temperature and velocity simulation respectively. In the height of 0.1m, the magnitudes of mathematical model were 28.5-37.5°C and 0-1.05m/s, whereas that of ANSYS were 28.85-37.85°C, and 0-1m/s, with R reached 96.11% and 95.24% individually. The relatively high results, compared to the well-known ANSYS simulation demonstrate the reliability of our mathematical model and the deduction we made about the effects of adjustments.

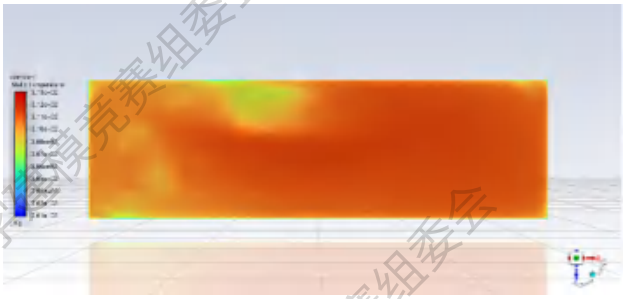
Overall, by applying the indicator, we showed that although the prediction results of the two models differ in some regions, they still show good agreement in most areas. This finding highlights the applicability and flexibility of the our model in simulating changes in complex environments. The reference provided by ANSYS simulation helped us in adjusting mathematical model to be reliable in further optimization, and validated our deduction on the direction of optimization, which would be beneficial in question four.

4.4 Question 4

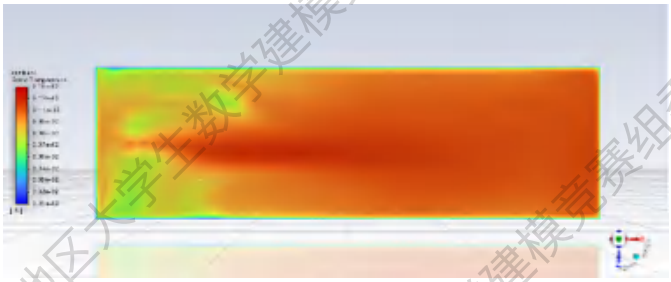
Based on the deficiency above, when it comes to optimizing design of glass greenhouse, among many of them, two main factors could be immediately came up with, the selection of crops and the optimization of the amount, location, speed and inlet temperature of fans.



(a) question two



(b) 3m/s in question three



(c) 1m in question three

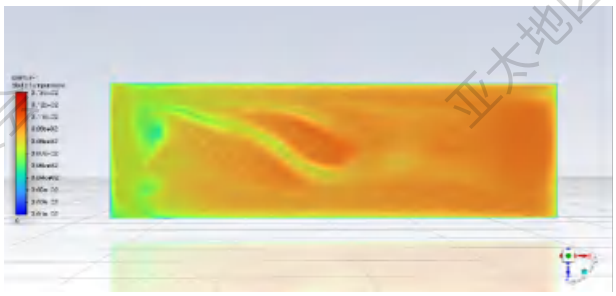
Figure 18 Temperature fields in different scenarios at the height of 0.5m

4.4.1 *Select Crops*

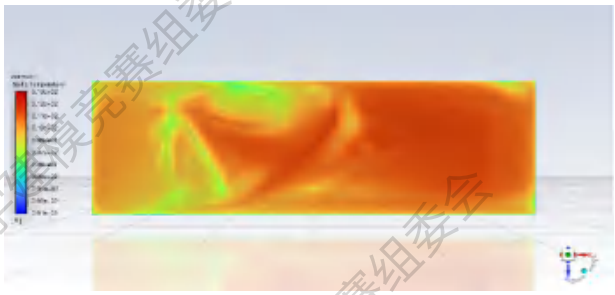
Our goal is to select crops that could fit high temperature and low velocity. According to our simulation results, for the first three conditions:

- For situations in question one, it's better to select crops that can endure temperatures within 31.2-37.6 $°C$, velocity below 0.48m/s at the height of 0.5m.
- For situations in question two, it's better to select crops that can endure temperatures within 31.2-37.6 $°C$, velocity below 0.61m/s at the height of 0.5m, and temperatures within 29-36 $°C$, velocity below 0.8m/s at the height of 0.1m.
- For situations in question three, it's better to select crops that can endure temperatures within 32-38.4 $°C$, velocity below 1.2m/s at the height of 0.5m, and temperatures within 28.5-37.5 $°C$, velocity below 1.05m/s at the height of 0.1m.

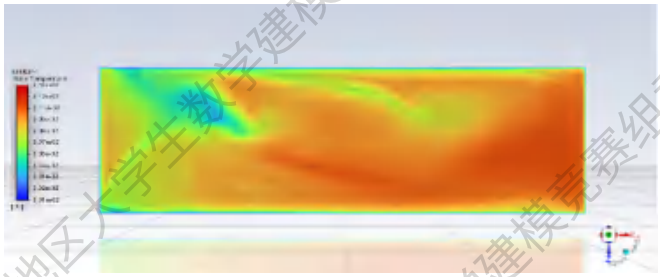
]



(a) question two



(b) 3m/s in question three



(c) 1m in question three

Figure 19 Temperature fields in different scenarios at the height of 0.1m

4.4.2 *Fans optimization by the Genetic Algorithm*

Here, we utilized a more complex but flexible algorithm, Genetic Algorithm (GA), with steps as follows.

(1) Objective function. Minimize the deviation between current conditions and the ideal environmental condition (23-26°C and 0.3-1m/s)

(2) Constraints.

- The number of fans is the integer is **below five**.
- Fans could only be equipped on the walls or the roof. The value of x,y,z should either 0 or the size of the greehouse.

(3) Strategy. Each individual represents a full condition, including the number of fans, spatial coordinates, inlet velocity and temperature.

(4) Mutation. Randomly produce new variants.

(5) Evaluation and Selection. Make sure the survival of the fittest.

(6) Initialization. Initialize with a population of 1,000 individuals and iterate 50 times.

4.4.2.1 Innovative points

(1) Length magnification. Considering the direction and magnitude of the fans, we devise a factor, *length magnification*, to strength its power in the specific direction of the fans' head, other than the other vertical directions.

(2) Convergence of GA. We regarded each generation in the GA algorithm as a state in the Markov Chain, which meets the criteria of the dependence to current generation and the independence to the last generation. As a consequence, in accord with the Markov Chain Convergence Theorem, we will finally achieve convergence of GA, under the implementation of Markov Chain.

4.4.2.2 Results

The flexible GA algorithm and the convergence guaranteed by the Markov Chain provided us with flexible solutions to the optimization of fans (see Appendix. (4)). Under the constraints set before, here, we displayed two solutions that both satisfied the temperature and velocity requirement, based on our mathematical models built before.

The first scheme contained five fans, with information provided in Table. (2). The location is clearly shown in Fig. (20. We could validate the feasibility by exhibiting target fields in Fig. (21) and Fig. (22). Whereas, **The second scheme** contained three fans, with information provided in Table. (3). Fig. (23 directly exhibited the locations of fans. Also, we could validate the feasibility by target fields in Fig. (24).

V. Pros and Cons

5.1 Pros

Our models are able to provide an in-depth spatial understanding of the climate conditions within the greenhouse, especially the spatial distribution of temperature and wind speed. We are able to accurately predict and evaluate the microclimate conditions in the greenhouse planting crop growth. Through Python programming and the application

Table 2 Configurations of five-fans scheme

Speed(m/s)	Temperature(°C)	x(m)	y(m)	z(m)
0.72603	26.08	2.87667	0	0.00068
0.79941	26.31	0	0.20255	1.05669
0.73218	26.70	2.38155	0	0.68965
0.78105	26.16	1.99720	1.17448	2
0.79096	26.49	3.74524	0	2

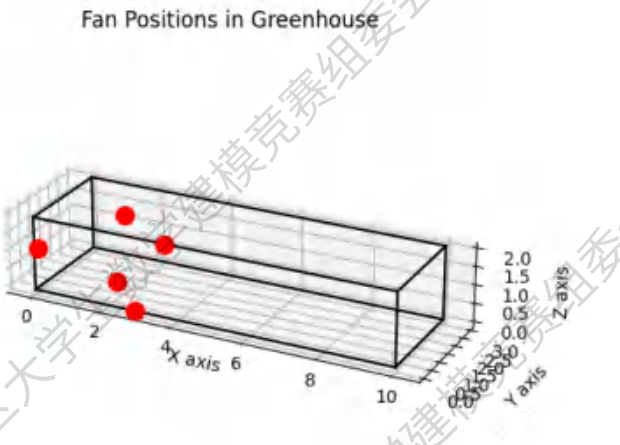


Figure 20 Locations of the five fans

of CFD software, ANSYS Fluent, we are able to represent complex three-dimensional temperature and wind speed distribution, providing visual aids for greenhouse design and optimization of crop growing conditions, at specific heights (e.g., 0.5 m and 0.1 m). In addition, the application of genetic algorithms in optimizing greenhouse fan designs has demonstrated a high degree of flexibility and efficiency. This approach takes into account not only the number and location of fans, but also factors such as wind speed and temperature, to generate a diverse range of solutions that provide an optimal environment for different crops and different greenhouse conditions. Finally, the application of these models has a certain degree of universality and scalability that may contribute to planning phase of greenhouse design and in the adaptation of existing greenhouse environments.

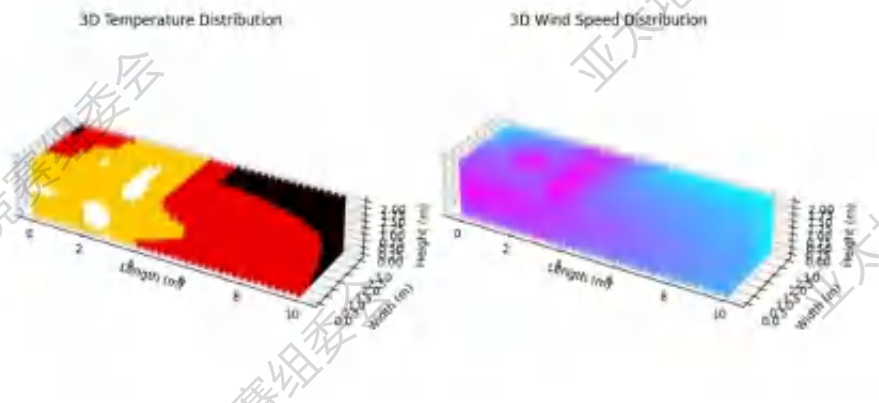


Figure 21 Three dimensional fields of the five-fans scheme

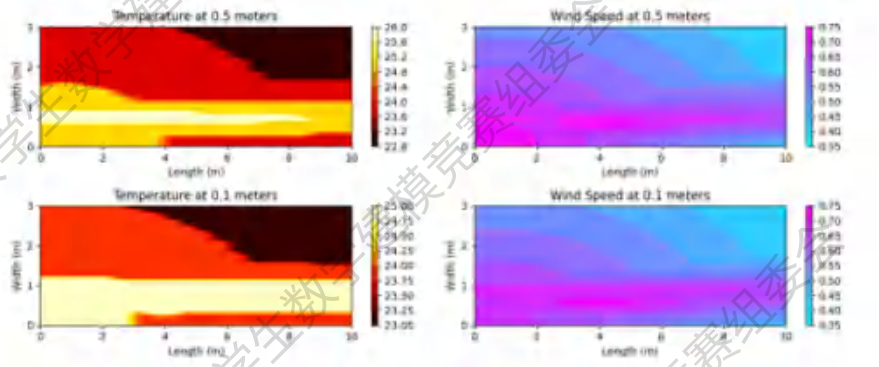


Figure 22 Target fields at two cross-sections of the five-fans scheme

5.2 Cons

Despite achievements, some deficiency should not be ignored. Firstly, the accuracy of the model relies heavily on the veracity of the initial assumptions and input parameters. Secondly, simple assumptions about porous media in crop models, may affect the accuracy of our results. Thirdly, since the steady-state case is mainly considered in the modeling, the energy conservation equation is simplified to a certain extent. Besides, when considering the resistance of the canopy and roots of crops to wind, only the effect of distance from the fan on energy transfer was considered, and no more detailed distinction was made between leaves and roots. Next, although genetic algorithm provides an effective optimization method, it is still a trial-and-error method, and its convergence and the quality of the final solution depend to some extent on the selection of algorithm parameters and the diversity of populations. Finally, the capability of the model may be limited by the specific application, e.g., external environmental factors such as solar radiation, external air currents, etc.

Table 3 Configurations of five-fans scheme

Speed(m/s)	Temperature(°C)	x(m)	y(m)	z(m)
0.73980	26.47	2.06455	3	0.22216
0.77190	26.24	0	0.16405	0.86741
0.65997	26.49	6.07402	2.80944	2

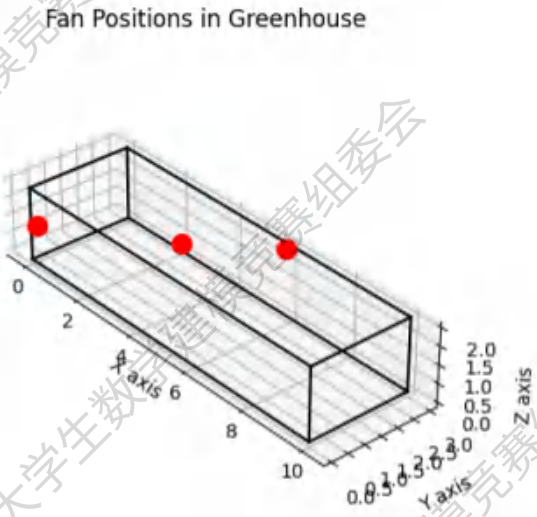
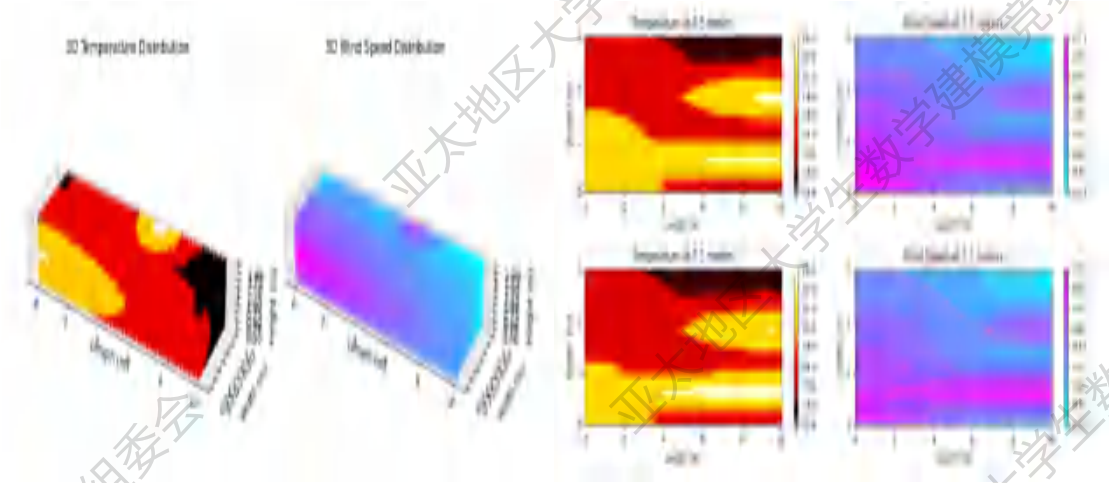


Figure 23 Locations of the three fans



(a) Three dimensional fields (b) Target fields at two cross-sections

Figure 24 The three-fans scheme

VI. References

[1] Stanghellini, C., Kubota, C. Greenhouse Technology and Management. CRC Press. 2010.

[2] Guo, L., et al. CFD simulation of airflow and temperature distribution in a greenhouse with a heat pump heating system. Biosystems Engineering, 183, 44-57. 2019.

亚太地区大学生数学建模竞赛组委会

Team Number:	apmcm2303763
Problem Chosen:	C

2023 APMCM summary sheet

This paper addresses multiple issues pertaining to the development and impact of new energy vehicles in China.

In the first problem, patent numbers, subsidies, per capita disposable income, and charging station quantities from 2010 to 2022 are collected and used as influencing factors. The sales volume of new energy vehicles is the dependent variable.A multiple regression equation is established, with regression coefficients indicating the importance of each factor. The order of importance is determined as subsidies, patents, per capita income, and charging station growth rate.

In the second problem, seven indicators representing the electric vehicle industry's development, including market penetration rate, number of charging stations, and production volume, are identified. Time series analysis using ARIMA models is conducted on market penetration rate and sales volume, identified as pivotal development indicators.

The third problem involves data collection on new energy and traditional vehicles, employing the least squares method for analysis.A population competition model predicts future trends in the ownership and sales of new energy and traditional vehicles, validated against original data.

For the fourth problem, the export volume of new energy vehicles is selected as the policy-influenced indicator. Historical restrictions from major market countries on China's new energy development are investigated using a time series approach. A DID model is employed to ascertain the strong negative correlation between foreign policy implementations and China's new energy vehicle exports.

Lastly, the fifth problem entails establishing a polynomial model based on urban car statistics to derive data for a city of one million inhabitants. This data is used to construct a model assessing the impact of vehicle electrification on the environment, considering factors like gasoline emissions' CO2 coefficients.

Keywords: New Energy Vehicles Regression Analysis Time Series Policy Impact Environmental Impact Urban Car Statistics

Contents

1. Introduction **1**
1.1 Background 1
1.2 Problem Restatement. 1
2. Assumptions and Justifications **2**
3. Notations **3**
4. Main Factors : What Affect the Development of NEEVs? **3**
4.1 Data Description 3
4.1.1 *Data Collection and Pre-processing* 3
4.2 New Energy Electric Vehicle Development Model. 4
4.2.1 *Influencing Factors and Index Setting* 4
4.2.2 *Multiple Linear Regression Model* 5
4.2.3 *Results* 5
5. What Will the NEEVs Be In the Next 10 Years? **6**
5.1 The Data 6
5.1.1 *Data Collection and Data Pre-processing* 6
5.2 Forecast model for the development of NEEVs in China 7
5.2.1 *Influencing Factors* 8
5.2.2 *ARIMA time series forecasting* 9
5.2.3 *Results* 11
6. An Impact on the Global Traditional Energy Vehicle Industry **13**
6.1 Data Description 13
6.1.1 *Data collection* 13
6.1.2 *Data Pre-processing* 13
6.2 Impact Analysis Model: New Energy Electric Vehicles on Traditional Vehicle Market. 13
6.2.1 *Optimal Species Competition Model in Math Modeling* 13
6.2.2 *Results* 15
7. Resistance to China's NEEVs and the Impact **16**
7.1 Index Setting 16
7.2 Difference-in-Differences Model 17
7.3 Results 18

8. Making Difference:Changes the Electrification Brings to the City of One Million People 19
8.1 Data Description 19
8.2 Composite Model of Urban Population and Carbon Emissions per Hundred Kilometers 19
8.2.1 *Urban Population and Fitting Model for Total Public Buses and Cars* 19
8.2.2 *The Fitted Model for Carbon Emissions and Vehicle Numbers.* 20
9. An Open Letter:New Energy Electric Vehicles Moisten Your Life and Mine 21
10. Sensitivity Analysis and Error Analysis 22
11. Model Evaluation and Further Discussion 23
11.1 Model Evaluation 23
11.1.1 *Advantages of the Model* 23
11.1.2 *Shortcomings of the Model* 24
11.2 Promotion of the Model 24
12. References 24
13. Appendix 25

I. Introduction

1.1 Background

In recent years, countries around the world have been calling for low-carbon and environmental protection. Only by developing clean energy and new energy industries can we promote the vigorous development of new energy in the world. In the face of global greenhouse warming and worsening environmental pollution, fuel emissions from traditional vehicles have become the main cause of greenhouse warming. In such a general environment, new energy vehicles,especially electric vehicles(EV), emerged as the times require.

With the development of the science and technology, China has continuously increased its requirements for vehicular exhaust emission. New energy electric vehicles(NEEVs), as a new, eco-friendly, and efficient industry, conforms to the concept of green energy and can effectively reduce environmental pollution.[1] So the development and impact of new energy electric vehicles are the two topics people more and more concerned on.

But the development of new energy electric vehicles is affected by factors from various aspects.They may from the government policies,production technology, infrastructures, value for money compared to substitute goods,and so on.Identifying the factors that affect the development of new energy electric vehicles, making reasonable judgments on the importance of each factor, and predicting future development trends and social impacts are urgent issues that need to be solved at the moment.

1.2 Problem Restatement

Taking into account the background information and restricted constraints identified in the problem statement, we need to settle the following questions:

- ⋆ **Problem 1**

  Figure out the main influencing factors on the development of new energy electric vehicles, establish corresponding mathematical models, and explain how the factors affect the development of new energy electric vehicles in China.

- ⋆ **Problem 2**

  Choose appropriate influencing factors that can reflect the development of China's new energy electric vehicles, select the indicators which affect the chosen factors

and collect more detailed data of them, establish rational mathematical models, and forecast the development of new energy electric vehicles in China in the next ten years.

- ⋆ **Problem 3**

  Based on the findings of problem 1 and 2, and extend Chinese model to the world, collect required data and improve the model, analyze the impact of the new energy electric vehicle industry on the traditional energy vehicle industry.

- ⋆ **Problem 4**

  In order to limit the rapid development of new energy electric vehicles in China, some countries have introduced policies to boycott Chinese new energy vehicles, establish a model that reflects the impact of these policies on the development of new energy electric vehicles in China.

- ⋆ **Problem 5**

  Analyze the influences of the urban new energy electric vehicles on the ecological environment, plug in a city of one million people and get the result.

- ⋆ **Problem 6**

  According to the model and conclusion of the previous question, write an open letter to propagandize the benefits of new energy electric vehicles and the contribution of their industry to countries around the world.

II. Assumptions and Justifications

Assumption 1

Assuming that the data obtained from the search has a certain degree of credibility and rationality.

Assumption 2

Assuming in the optimal species competition differential equation model, when new energy vehicles exist independently, they follow the logistic pattern of population change.

Assumption 3

Assuming that the new energy electric vehicles in the optimal species competition differential equation model are treated as pure electric vehicles.

Assumption 4

Assuming that all new energy electric vehicles in problem 5 are purely electric.

Assumption 5

Assume that urban vehicles in problem 5 do not include trucks, and all fuel vehicles use gasoline as energy consumption.

III. Notations

Symbol	Description
u_n	The policy subsidy amount for the year n
$u_n^{'}$	Revised policy subsidy indicators
z_n	The number of relevant patents for the year n
k_n	The median personal disposable income for the year n
c_n	The number of electric vehicle charging stations as of year n
$c_n^{'}$	The growth rate of charging stations in year n
y	Objective function
n	years
η	market penetration
$y^{'}$	The potential number of the entire market

IV. Main Factors : What Affect the Development of NEEVs?

4.1 Data Description

4.1.1 *Data Collection and Pre-processing*

In order to analyze the main factors that affect the development of new energy electric vehicles in China, relevant data sets need to be collected. Owing to various statistical departments associated with distinct elements, the Table 1 below illustrates the origins of the data sets utilized in this chapter alongside the explanations of variables.

Table 1 Data source

Production and sales	Ministry of Industry and Information Technology of China
Government subsidy allocation	Ministry of Industry and Information Technology of China
Number of related patent applications	Autopat website
Personal annual disposable income (median)	National Bureau of Statistics
Number of charging piles in stock	China Electric Vehicle Charging Infrastructure Promotion Alliance

Since most of the data obtained in the preliminary work is not in a form that can be directly used, these data should receive some preliminary processing.

In order to eliminate the influence of different dimensions of each set of data and facilitate the construction of a multivariate regression model, data for each variable should be normalized. Due to the relatively stable nature of the data without extreme outliers in its maximum and minimum values, this paper utilizes min-max normalization to process the data. It can be represented by the following formula:

$$x^{'} = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

x is a specific index of the variable, $x^{'} \in [0, 1]$, $max(x)$ and $min(x)$ respectively represent the maximum and minimum values of the sample data.The variables mentioned below are all normalized by default.

4.2 New Energy Electric Vehicle Development Model

4.2.1 *Influencing Factors and Index Setting*

Unless otherwise specified, the following data defaults to years as the time series.

Comprehensive analysis of research reports from major institutions and existing research in this field, Sales of new energy electric vehicles (y) are considered a mainly indicator of development. It will be used as the dependent variable in the multiple linear regression analysis in this chapter.

Combined with practical considerations and according to relevant research [2, 3] , the main factors affecting the development of NEEVs include subsidy policy provided by the state (u_n) , number of related patents successfully applied for during this period (z_n) , citizens' personal disposable income (k_n) , number of charging piles nationwide (c_n) .

It should be noted that the personal disposable income here is the median, not the average. For the median represents the middle value when all incomes are arranged in ascending order, offering a more robust measure that is not as heavily influenced by the extreme values.

4.2.2 *Multiple Linear Regression Model*

Establish a multiple linear regression model between each factor and the sales:

$$y = \beta_1 + \beta_2 z_n + \beta_3 u_n^{'} + \beta_4 k_n + \beta_5 c_n^{'} \tag{2}$$

Among them, The growth rate of charging stations in year n $c_n^{'}$ is represented by the following formula:

$$c_n^{'} = \frac{c_n - c_{n-1}}{c_n} \tag{3}$$

The revised policy subsidy indicators $u_n^{'}$ equals to $u_n^{\frac{3}{2}}$

4.2.3 *Results*

Use the regress function in Matlab to establish a multiple linear regression model and obtain the estimated value of the regression coefficient β,the regression model can be expressed as:

$$y = -0.010 + 0.491 z_n + 1.192 u_n^{'} + 0.319 k_n + 0.012 c_n^{'} \tag{4}$$

Table 2 Model Summary

R-squared	F-test	p-value	Standard Estimate Error
0.9994	423.6054	0.0364	0.0004

The model summary shows that the R-square of the model is 0.9994,the large F value and the p value of 0.0364 (less than 0.05) indicate that the model is significant

overall and the regression coefficient is credible.The overall fitting effect is good. The multiple linear regression fitting plot and residual plot are shown below:

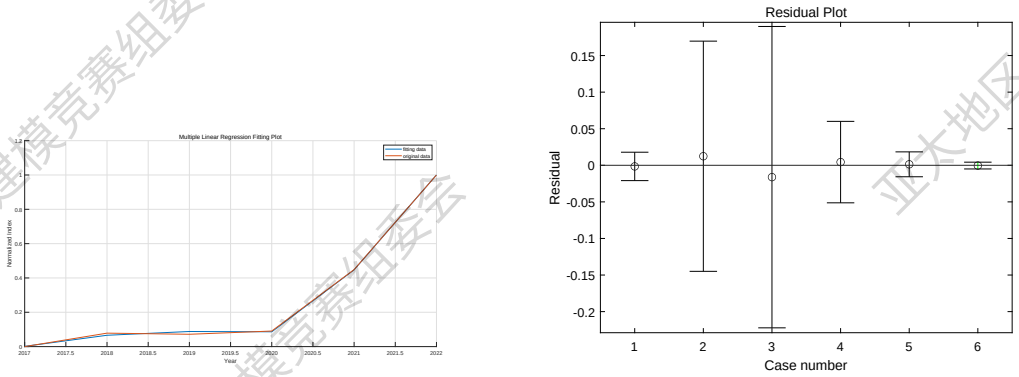


Figure 1 Multiple linear regression fitting plot and residual plot

V. What Will the NEEVs Be In the Next 10 Years?

5.1 The Data

5.1.1 *Data Collection and Data Pre-processing*

In this chapter, the sales data of NEEVs collected in the previous question are continued to be used. The market penetration data was gathered from China's Ministry of Industry and Information Technology and supplemented by information sourced from the China Association of Automobile Manufacturers.

Since it is observed that the initial data has exponential growth properties (refer to the left pictures in Figures 2 and 3), we perform logarithmic processing on the original data. The data after logarithmic processing is more stable and can make the time series more stable (refer to the pictures on the right of Figures 2 and 3). These data are now prepared and ready to be used for model fitting.

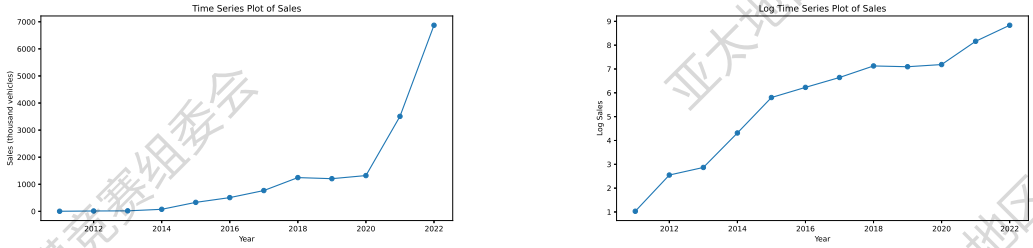


Figure 2 Sales time series plot and the plot with Log Transformation

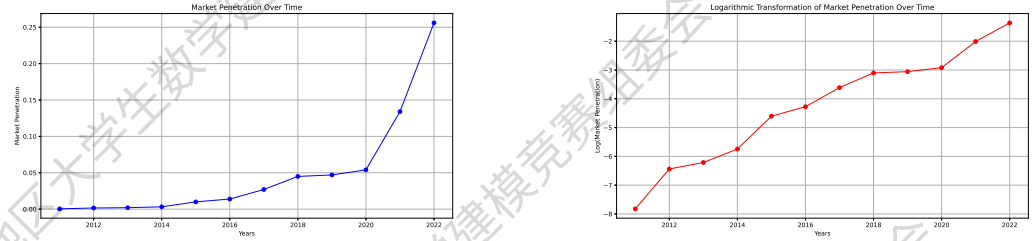


Figure 3 Market penetration rate time series plot and the plot with Log Transformation

5.2 Forecast model for the development of NEEVs in China

ARIMA model is chosen due to its capability in handling time series data exhibiting temporal dependencies and trends. This model is particularly suitable for our dataset as it can account for the temporal nature and potential autocorrelation present in the variables, providing a robust framework for time series analysis and forecasting.

The following diagram (Figure 4) illustrates the step-by-step process used to solve the problem discussed in this article.

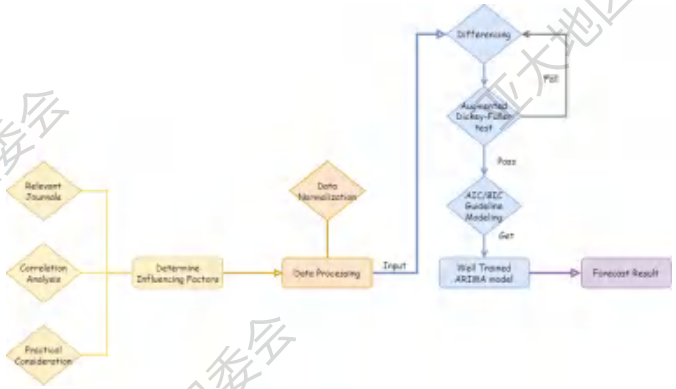


Figure 4 Step flow chart

5.2.1 *Influencing Factors*

Regarding the selection of influencing factors, this paper first conducts a correlation test on the influencing factors selected in problem 1 and some factors that are objectively believed to have strong relevance to the development of new energy electric vehicles.

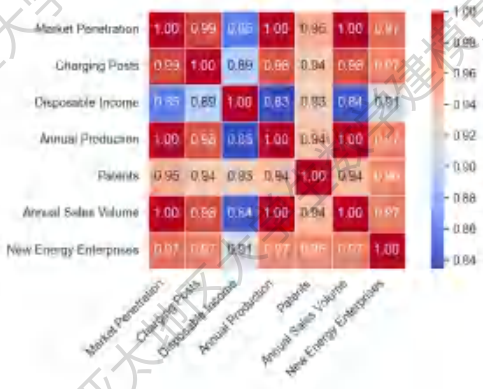


Figure 5 Correlation matrix between feature variables

As can be seen from the above figure, the correlation between feature variables is very strong. This also shows that some feature variables provide similar information to other variables. Therefore, based on the selection rate of influencing factors in mainstream research reports on the new energy electric vehicle industry, this article selects sales volume and market penetration as indicators to reflect the development of new energy electric vehicles in China.

Here comes the conception of market penetration(η). It is related to sales data for the year(y),and the potential sales data of the automobile market that year($y^{'}$). The formula is as follows:

$$\eta = \frac{y}{y^{'}} \times 100\% \tag{5}$$

5.2.2 *ARIMA time series forecasting*

The ARIMA (Autoregressive Integrated Moving Average) model is a widely used time series analysis technique for forecasting future values based on historical data patterns. The mathematical formula of an ARIMA(p, d, q) model is represented as:

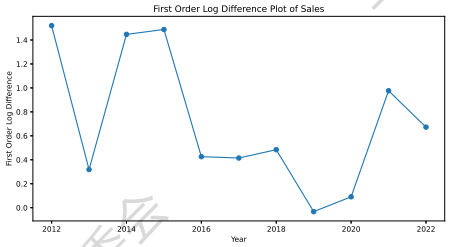
$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_2 B^p)(1 - B)^d X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t \tag{6}$$

where X_t is the time series data, B is the back shift operator, ϕ_i and θ_i are the autoregressive and moving average parameters, p is the autoregressive order, d is the differencing degree, and q is the moving average order.

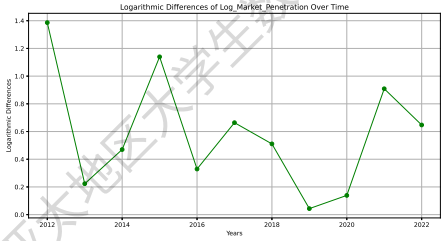
In order to build an ARIMA model, it is necessary to select appropriate parameters p, q, d.

⋆ **Step 1:** Differential processing and ADF test

Differencing aims to transform a non-stationary time series into a stationary one for better model fitting. Perform Augmented Dickey-Fuller (ADF) test after each differencing. If p-value is ¡ 0.05, the test is successful. If not, continue differencing until p-value ¡ 0.05. After first differencing, sales volume remains non-stationary with a p-value ¿ 0.05.



(a) First difference of logarithm of sales volume



(b) First difference of logarithm of market penetration

Figure 6 First order difference of logarithmic time series

Then perform the second-order difference on the sales volume, and the p value is less than 0.05.

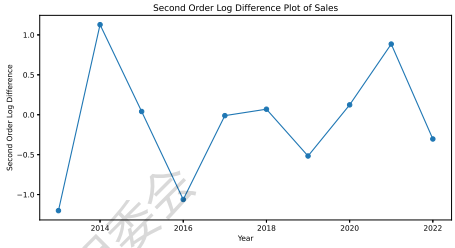


Figure 7 Second order difference of logarithmic time series of sales volume

The data is as follows:

Table 3 ADF statistics and p-value

	sales volume		market penetration	
	ADF statistics	**p-value**	**ADF statistics**	**p-value**
first-order differencing	-1.7733	0.3938	-4.4307	0.0002
second-order differencing	-3.2121	0.0193		

⋆ **Step 2:** Determination of p and q

Due to the limited amount of data, accurately determining the orders p and q of the ARIMA model from the autocorrelation function (ACF) and partial autocorrelation function (PACF) can be quite challenging. Hence, the use of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) is adopted to ascertain the values of p and q.

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are statistical measures used for model selection among different candidate models.

AIC balances a model's goodness of fit with its complexity to find the best trade-off between them. A lower AIC value indicates a better model. It's calculated as $2k - 2\ln(\widehat{L})$, where k is the number of model parameters and $\widehat{L}$ is the maximum likelihood function value for the model.

Similarly, BIC also aids in model selection but penalizes complex models more. It aims to choose simpler models by giving a heavier penalty for complexity than AIC. Lower BIC values denote better models. The BIC formula is $-2\ln(\widehat{L}) + k\ln(n)$, where n is the sample size.

This paper explored various combinations of different p and q values, calculating their corresponding AIC and BIC values. Ultimately, it pinpointed specific p and q combinations that resulted in the lowest AIC or BIC values.

For sales volume, two sets of p and q values meet the criteria: $p = 3$, $q = 1$ and $p = 0$, $q = 1$ are both valid. For market penetration, there are also two sets: $p = 3$, $q = 0$ and $p = 3$, $q = 1$, both meeting the conditions.

Now, the appropriate parameters p, d, and q have been determined.

5.2.3 *Results*

The ARIMA(3,2,1) and ARIMA(0,2,1) models are employed to fit and forecast sales data for the forthcoming decade. Simultaneously, the ARIMA(3,1,0) and ARIMA(3,1,1) models are utilized for fitting and forecasting market penetration.

Table 4 Forecast of the development of NEEVs in China in the next 10 years

	sales volume(000 vehicles)		market penetration	
year	ARIMA(3,2,1)	ARIMA(0,2,1)	ARIMA(3,1,0)	ARIMA(3,1,1)
2023	9034.13	10060.69	0.4148	0.3946
2024	10543.80	13249.39	0.4625	0.4219
2025	13200.22	16438.08	0.4862	0.4537
2026	16383.31	19626.78	0.5367	0.5223
2027	18609.43	22815.47	0.6170	0.5917
2028	20462.89	26004.17	0.6932	0.6293
2029	23134.30	29192.86	0.7415	0.6502
2030	26070.49	32381.56	0.7712	0.6821
2031	28315.21	35570.25	0.8058	0.7243
2032	30379.75	38758.95	0.8540	0.7585

Based on the projected data, the sales volume over the next decade is anticipated to surge from 9 to 10 million to a range of 30.3 to 38.7 million units. This represents a growth in sales data by a factor of 3.36 to 3.85 over the ten-year period.

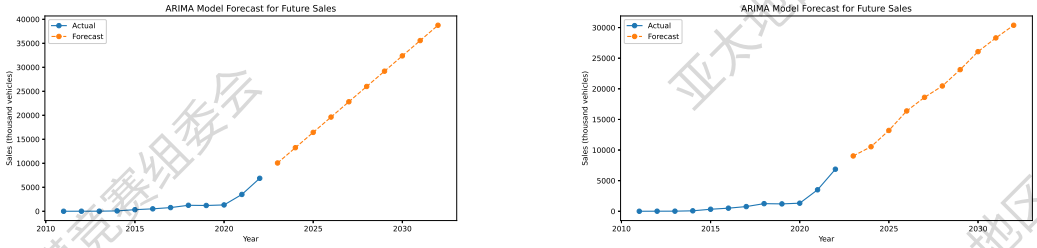


Figure 8 Sales forecast result of p=0,q=1(left) and p=3,q=1(right)

In the next ten years, new energy electric vehicles will usher in a blowout development. Market penetration rises rapidly from low levels to extremely high growth rates. In 2023, the market penetration is expected to be around 40 percent, which is less than half. By 2032, the market penetration of NEEVs is projected to increase significantly to between 75 percent and 85 percent.

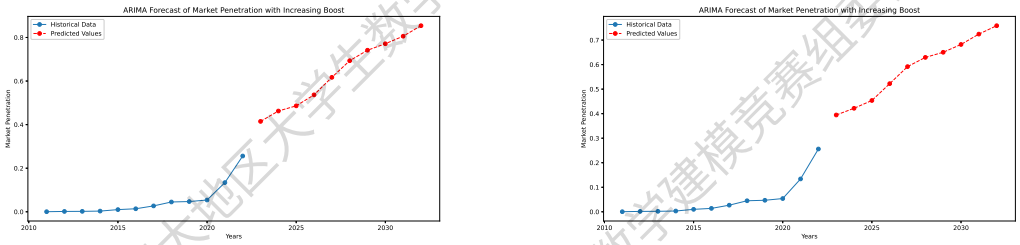


Figure 9 Market penetration rate forecast result of p=3,q=0(left) and p=3,q=1(right)

In the coming decade, New Energy Electric Vehicles (NEEVs) are poised to dominate the automotive industry, capturing nearly 80% of the domestic car market. They will progressively replace conventional fuel vehicles until the market reaches saturation.

VI. An Impact on the Global Traditional Energy Vehicle Industry

6.1 Data Description

6.1.1 *Data collection*

The sales data follows the statistics from problem 1 and 2. The vehicle ownership data is sourced from the International Organization of Motor Vehicle Manufacturers(IOCA) and China News Network.

6.1.2 *Data Pre-processing*

In the quest for reliable modeling, data preprocessing stands as a crucial stage.This process aims to fortify the dataset for modeling purposes.

For global new energy electric vehicles and traditional energy vehicles' sales and ownership data, we chose logistic regression for the former and the least squares method for the latter based on their suitability and goodness of fit calculations. Logistic regression is ideal for categorical outcomes, while the least squares method suits linear relationships, aligning with the nature of the datasets. Moreover, preprocessing the data offers significant benefits, such as improving data quality, refining data fit for models, and enhancing overall analysis robustness.

6.2 Impact Analysis Model: New Energy Electric Vehicles on Traditional Vehicle Market

6.2.1 *Optimal Species Competition Model in Math Modeling*

The optimal choice of the species competition differential equation model in industries hinges on its capacity to replicate intricate interactions among competitors, mimicking the rivalry between new energy electric vehicles and traditional counterparts. This model effectively tracks evolutionary paths and competitive behaviors crucial for assessing emerging technology's impact on established markets. Using vehicle sales and ownership as key metrics is motivated by their ability to portray market penetration and consumer adoption rates comprehensively. Sales figures mirror immediate market acceptance, while ownership signifies long-term commitment and market sustainability,

offering a holistic view of the industry's response to the rise of new energy electric vehicles.

◇ **Optimal Species Competition Differential Equation Model for Ownership**

The optimal species competition differential equation model represents the dynamic interaction between competing entities in an ecosystem. Mathematically, it is often formulated as:

$$\begin{cases} \frac{\mathrm{d}N_1(t)}{\mathrm{d}t} = r_1 N_1 (1 - \frac{N_1}{K_1}) - \sigma_1 N_1 N_2 \\ \frac{\mathrm{d}N_2(t)}{\mathrm{d}t} = r_2 N_2 (1 - \frac{N_2}{K_2}) - \sigma_2 N_1 N_2 \end{cases} \tag{7}$$

$$b = r(1 - \frac{N}{k}) = \frac{\mathrm{d}(N(t) - N(t-1))}{\mathrm{d}t} \tag{8}$$

We use the subscript 1 to represent the data from the traditional automobile industry, and the subscript 2 to the new energy automobile industry. In this section, N represents the processed data of global holdings; r means the maximum growth rate of the population; K represents maximum population size the environment can sustain; σ means the intensity of competition between entities. b is the average annual growth rate of a certain population when it exists alone.

Based on investigations into industry research reports and official news coverage, we have found that the development of traditional energy has essentially reached saturation in 2022[4] . The growth in ownership is relatively stable, K_2 is infinitely close to the maximum saturation of the market. So the indicators of traditional energy do not conform to the logistic pattern of population change. That leads to the ratio parameter $\frac{N_1}{K_1} = 0$.

According to the definition of each indicator, we associate the existing data with the parameters in the formula: K_2 equals to the ownership of traditional automobile industry in 2022, which is 11,446 million. And the σ here represents the cost performance index. σ_1 is the average electricity cost per 100 kilometers of new energy electric vehicles and the average fuel cost per 100 kilometers of traditional energy vehicles, and σ_2 is the reciprocal of σ_1.

Using the data above, the parameters in the model are determined. Bring these parameters into the model to solve and get the results.

◇ **Optimal Species Competition Differential Equation Model for Sales Volume**

The model for sales is roughly the same as the above model, with K_2 changed to 7493.56.

6.2.2 *Results*

Using this model, we obtained competition outcome data. The data of next ten years are showed below as data display, it demonstrates the development of new energy electric vehicles(NEEVs) and traditional energy vehicles(TEVs).

Table 5 Forecast for the development of NEEVs and TEVs in the next ten years

Ownership(000s)		year	Sales Volume(000s)	
NEEVs	TEVs		NEEVs	TEVs
170.064	144356.98	2023	140.142	6600.78
253.653	144237.74	2024	208.285	6504.63
378.289	144059.90	2025	308.972	6363.42
564.081	143794.91	2026	456.989	6157.70
840.929	143400.33	2027	672.851	5861.73
1253.226	142813.37	2028	983.650	5444.14
1866.708	141941.48	2029	1421.927	4872.79
2778.356	140649.12	2030	2019.211	4128.23
4130.402	138739.71	2031	2789.012	3229.47
6129.517	135932.42	2032	3697.395	2263.65

The competition results of the population model for the respective holdings and sales are as follows:

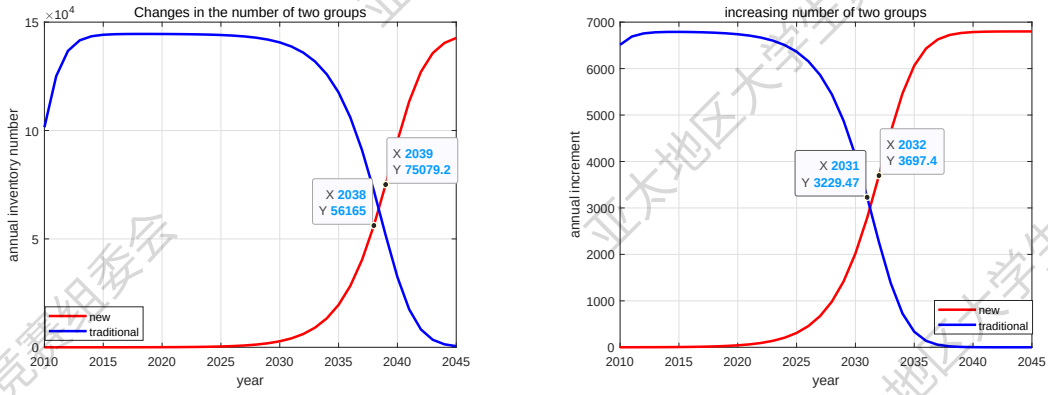


Figure 10 Competition results of ownership(left) and sales(right)

Based on the development trends of the two, the number of new energy electric vehicles and traditional energy vehicles will reach the same level between 2038 and 2039. And in 2041, new energy vehicles will account for more than 86% of the automobile market.

Based on the development trends of the two, the sales of new energy electric vehicles and traditional energy vehicles will reach the same level between 2031 and 2032. And in 2034, new energy vehicles will account for more than 88% of the automobile market.

VII. Resistance to China's NEEVs and the Impact

Because China's share of the new energy electric vehicle market exceeds 50% (data statistics are as of 2022), other countries have introduced policies to protect local companies and inhibit the development of Chinese companies.

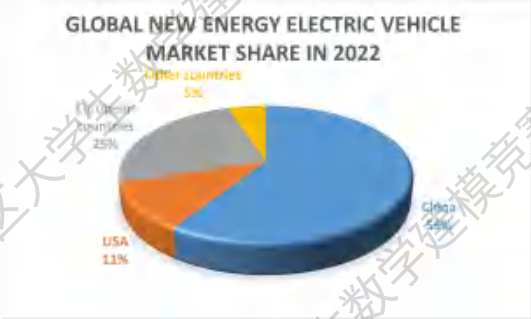


Figure 11 Global new energy electric vehicle market share in 2022

Apart from China, Europe and the United States wield the greatest influence on the market. Hence, we focus solely on evaluating the effects of European and American policies on our country's new energy vehicles.

7.1 Index Setting

Based on a survey of past relevant research papers, China's outbound new energy vehicle exports serve as a direct gauge of foreign policy impact on its electric vehicle industry. These exports reflect market changes due to tariffs, subsidies, and regulations, showcasing how international policies influence demand and industry growth.

Hence, this section selects China's NEEV export volume as the yardstick to assess the influence of foreign policies on the growth of China's NEEV industry. We collect

the main policies from October 2021 to October 2023.

- ⋆ **Policy 1 (May 2021):**
  The U.S. Senate Finance Committee passed the "American Clean Energy Act" proposal, which allocates $31.6 billion in electric vehicle consumption tax credits. At the same time, it relaxes the 200,000-unit limit for automakers to enjoy tax credits, excluding Chinese manufacturers.
- ⋆ **Policy 2&3 (February-March 2022):**
  The U.S. House of Representatives passed the American Competition Act, which plans to allocate $52 billion over the next five years to subsidize research and manufacturing in key industries such as semiconductors and automobiles.
- ⋆ **Policy 4 (June 2022):**
  The passage of the tougher "American Innovation and Competition Act" has more powerfully protected the development of domestic companies in the United States.
- ⋆ **Policy 5 (September 2022):**
  The U.S. Inflation Reduction Act has officially become legislation. This bill contains strong unilateralism and protectionist overtones. Only companies that are on the list can enjoy preferential treatment.
- ⋆ **Policy 6 (September 2023):**
  European Commission President von der Leyen announced in her State of the Union address that she would launch a countervailing investigation into China's electric vehicles.

7.2 Difference-in-Differences Model

A difference-in-difference (DID) model is a statistical method used to estimate the causal effects of policy interventions by comparing changes in outcomes over time between experimental and control groups. The model is based on the traditional difference-in-difference (DID) approach, incorporating multiple time periods and/or multiple groups.

The time series settings of policy nodes in our DID model are as follows: when there is a policy, the sequence value takes 1, and when there is no policy, it takes 0. Use the DID model to test the causal relationship between export volume and whether the policy is enacted, and give the correlation (negative correlation)

Judging from the existing monthly export volume data (October 2021 to October 2023), China's export volume of new energy electric vehicles is gradually rising. There-

fore, we set the control group data that is not affected by the policy to be the arithmetic sequence distribution from the first group to the last group of existing data, and conduct DID model testing.

7.3 Results

The monthly export volume-policy diagram used in the DID model is as follows:

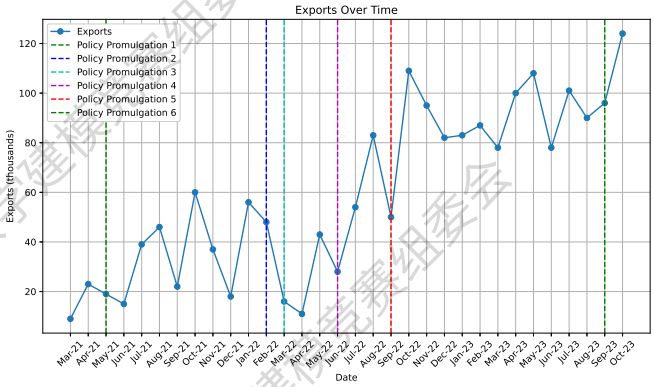


Figure 12 Export volume-policy diagram

And the results of the goodness of fit and policy influence of the DID model we used are as follows:

Table 6 Model Summary

DID model			Policies	
R-squared	**Adj.R-squared**	**Prob (F-statistic)**	**Coef**	**P>|t|**
0.942	0.938	2.72E-19	-15.3499	0.049

The provided DID model exhibits a high goodness of fit, with an R-squared of 0.942 and an adjusted R-squared of 0.938, indicating that approximately 94.2% of the variation in the outcome variable is explained by the included variables. The Prob (F-statistic) of 2.72E-19 is significantly low, suggesting that the model as a whole is statistically significant, indicating that the included variables jointly have a significant effect on the outcome. Additionally, the coefficient of policies is -15.3499, and its associated t-statistic yields a p-value of 0.049, which is marginally below the conventional threshold of 0.05. This indicates a potential significant impact of policies on the outcome variable.

VIII. Making Difference:Changes the Electrification Brings to the City of One Million People

8.1 Data Description

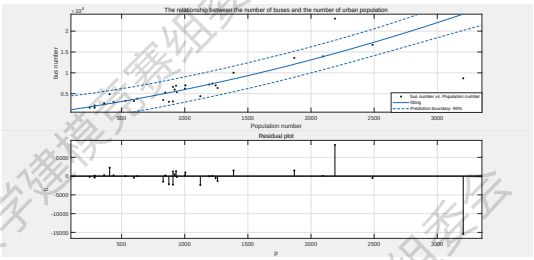
According to authoritative industry reports in the automotive sector and publicly available national information, we have obtained the average electrification rates for buses and cars (M_b,M_c). Additionally, we've acquired the energy consumption equivalents for electric buses and cars (a_1, a_2), where energy consumption equivalents denote the conversion of electricity consumption per hundred kilometers to fuel consumption per hundred kilometers. Furthermore, we have the fuel consumption per hundred kilometers for non-electric buses and cars (b_1, b_2), as well as the carbon dioxide emission coefficient for gasoline (ω).

8.2 Composite Model of Urban Population and Carbon Emissions per Hundred Kilometers

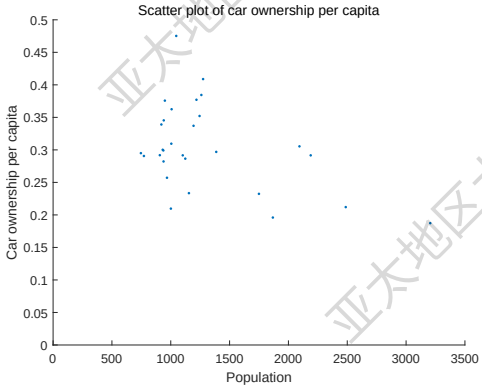
To evaluate the impact of automobile electrification on the ecological environment in a city of one million inhabitants, it requires fitting into two nested models. Given the composition of vehicles in the city, for model simplification, we assume the city's automobiles do not include trucks. We use carbon emissions, the most influential indicator, to assess the impact on the ecological environment. The first step involves analyzing the relationship between urban population and the quantities of electric and non-electric vehicles. The second step involves gathering data to reasonably fit vehicle quantities and carbon emissions, ultimately determining the carbon emissions after electrification for a city of one million inhabitants.

8.2.1 *Urban Population and Fitting Model for Total Public Buses and Cars*

Due to the strong fit between the total number of public buses and the total population (as shown in the figure 13(a)), the estimated total number of buses(L_b) in a city of one million inhabitants is 1161.



(a) The Graph Fitting Bus Quantity to Urban Population



(b) Scatter Plot: Car Ownership per Capita vs. Urban Population

Figure 13 The Urban Population's Relationship with Total Public Buses and Cars

However, due to the scattered distribution between the total number of cars and the total population (as seen in the scatter figure 13(b)), we perform fitting on the total count. Then, subtracting the bus count from the total count gives us the total number of cars. After fitting, the estimated total vehicles in a city of one million inhabitants are 1.0465 million, from which the number of cars(L_c) can be calculated.

8.2.2 *The Fitted Model for Carbon Emissions and Vehicle Numbers*

We choose the total carbon dioxide emission as the representative indicator for carbon emissions(T) .Establish a model for carbon dioxide concerning the number of urban buses and cars using the following formula:

$$T = \omega \times \left[\left(a_1 M_b + b_1 (1 - M_b) \right) L_b + \left(a_2 M_c + b_2 (1 - M_c) \right) L_c \right] \tag{9}$$

Finally, we get an estimate of carbon dioxide emissions per 100 kilometers for a city with a population of one million.The calculation result of the model is 21545123.4782328 kilograms.

The figure of fitted model for carbon emissions and vehicle numbers is as follow:

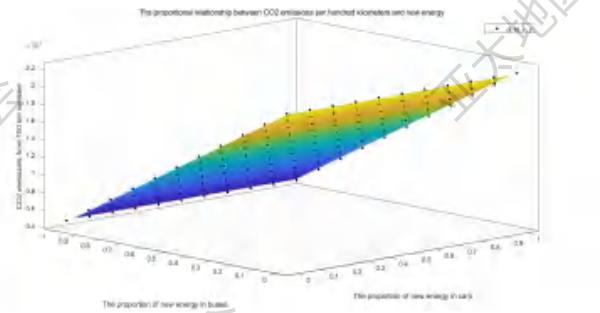


Figure 14 The Fitted Model for Carbon Emissions and Vehicle Numbers

IX. An Open Letter:New Energy Electric Vehicles Moisten Your Life and Mine

Dear Citizens,

As our world navigates through the challenges of environmental sustainability, there's a beacon of hope shining brightly on our roads: new energy electric vehicles. The electrification of transportation, from cars to buses, holds immense promise in transforming our cities into cleaner, healthier, and more sustainable spaces. It's time to recognize how significant these vehicles contribute to protecting our ecological environment.

New energy electric vehicles (NEEVs) are not just a mode of transportation, they represent a pivotal shift toward a greener future. The benefits they offer are multifaceted and impactful. By choosing NEEVs, we collectively contribute to reducing greenhouse gas emissions, curbing air pollution, and mitigating the effects of climate change. We provide the future generations with a quieter, cleaner and safer living environment.

Countries around the world have been embracing the electric vehicle industry in a more positive and open manner. They are adopting proactive measures and policies to drive innovation and foster a revolution in transportation. From the Norway to China, the United States to Netherlands , Governments and industries around the world are increasing investment in NEEV infrastructure and technology. It is precisely because of the accumulated efforts of all parties that the overall improvement in our quality of life has been achieved today.

Electric buses,especially worth to be mention, are silently transforming our cities. They offer They provide a quieter, less bumpy and pollution-free mode of public trans-

portation. Just imagine waiting at a bus stop surrounded by clean air, with buses silently passing by, leaving behind no emissions. It's a vision that's becoming a reality in many cities, thanks to the widespread adoption of electric buses.

The benefits of electric vehicles extend beyond environmental advantages. They also bring economic opportunities, fostering job creation and innovation in clean energy technologies. Moreover, as the NEEV industry grows, the cost of these vehicles continues to decrease, making them increasingly accessible to everyone.

However, this transition requires collective action. Each of us plays a crucial role in accelerating the adoption of electric vehicles. Whether it's choosing an electric car for your daily commute or supporting initiatives that promote NEEV infrastructure, every action contributes to a more sustainable future.

Let's take pride in our collective efforts toward a cleaner and healthier environment. Let's embrace the potential of new energy electric vehicles and support the electric vehicle industry, not just for ourselves but for the well-being of our planet and future generations.

Together, we can drive change, one electric vehicle at a time.

X. Sensitivity Analysis and Error Analysis

When conducting multiple linear regression model in Problem 1, in order to explore the sensitivity of the model,a 85%-115% confidence interval for the most influencing factor should be given.

Observing the value of the regression coefficient β ,we can get the biggest one is belonged to the policy subsidy amount of NEEVs. It confirmed that policy subsidy amount is the most influecing factor. Therefore, we fix the values of other variables and perturb the subsidy coefficient, Obtain a 85%-115% confidence interval for the policy subsidy amount.

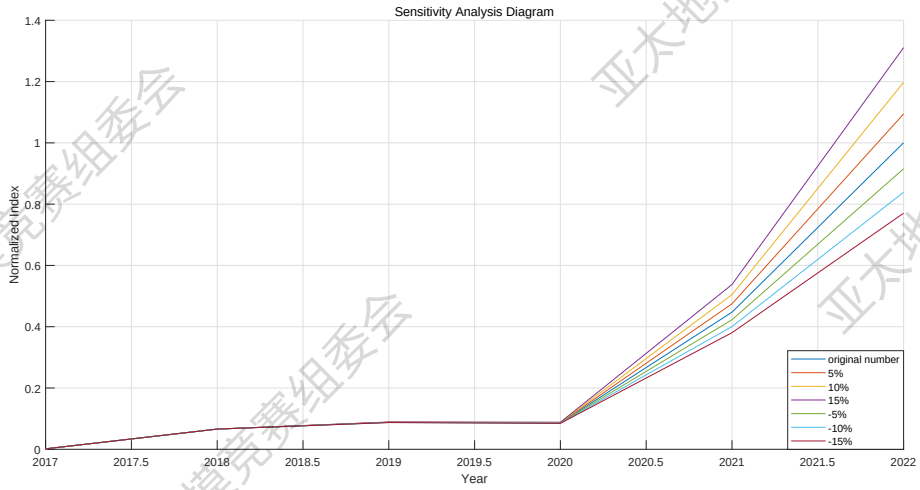


Figure 15 1

The model's conclusion highlights a consistently narrow range of solution fluctuation, indicating a low standard deviation. The accompanying figure illustrates the model's low sensitivity and strong robustness.

XI. Model Evaluation and Further Discussion

11.1 Model Evaluation

11.1.1 *Advantages of the Model*

★ In the first question, our strength lies in identifying the factors influencing the development of the new energy industry through relevant research and assessing the extent of their impact, presenting a more intuitive display of their influence.

★ Evolving and refining the population competition model to better align with the developmental patterns of industry competition. This model has the capability to forecast the future trends of both new energy vehicle and traditional vehicle industries under their mutual influence, enhancing its scientific validity.

★ We utilized a double difference model to analyze the impact of policies on export volume, achieving a good fit. It holds more persuasive strength compared to traditional regression models.

★ Building upon the foundation of the environmental impact of automobiles in various Chinese cities, the model constructed is more scientific and representative. It

distinguishes between the electrification rates of cars and buses, avoiding a simple aggregate of the two. This approach fully considers the varying degrees of vehicle electrification between cities, rendering the model's outcomes more comprehensive.

11.1.2 *Shortcomings of the Model*

⋆ The collected sample size of data can be expanded further, providing room for additional optimization and improvement of the model.

⋆ In the fifth problem, the discussion focuses solely on two vehicle types: cars and buses, presenting a relatively simplified analysis of vehicle categories.

11.2 Promotion of the Model

⋆ We can collect more indicators and data, and build a more complete and more complex mathematical model for more accurate evaluation of various indicators reflecting the development level of NEEVs.

⋆ We can collect more indicators and data, make a promotion of the Difference-in-Differences Model. Analyze the duration of various policy lags, obtain new policy node time series with policy lags, and establish a more complete DID model, which may lead to better fitting results.

XII. References

[1] Zhai Jiatian. Analysis of development status and prospect of new energy vehicles. *Automobile Applied Technology*, 48(20):193–198, 2023.

[2] Zhang Qingxin, Guo Mengmeng, and Liu Renyi. The impact of rd investment on the development capacity of new energyvehicle enterprises. *Journal of North China University of Science and Technology(Social Science Edition)*, 23(06):27–32+60, 2023.

[3] He Lin. Competitiveness and industrial development trends of new energy vehicles in the post-subsidy era and policy recommendations. *China Academic Journal Electronic Publishing House*, 2022(04):16–23, 2022.

[4] Hu Ansheng and Feng Xiayong. Development and changing trends of the world automobile industry. *ChinaMachinery Electric Industry*, (9):33–35, 2004.

APMCM ASIA PACIFIC UNDERGRADUATE MATHEMATICAL MODELING COMPETITION ENTRY MANUAL

APMCM亚太地区大学生数学建模竞赛简介

2026年第十六届亚太地区大学生数学建模竞赛(以下简称“竞赛”)是由中国国际科技促进会物联网工作委员会和北京图象图形学学会联合主办的亚太地区大学生学科类竞赛，竞赛由亚太地区大学生数学建模竞赛组委会负责组织，欢迎各高等院校按照竞赛章程及有关规定组织同学报名参赛。

2025年第十五届亚太地区大学生数学建模竞赛共有13000支队伍980所高校3万5千多名学生报名参赛。参赛高校覆盖北京大学、清华大学、浙江大学、同济大学、上海交通大学、复旦大学、四川大学、大连理工大学等全部的39所985高校和114所211高校。

除中国大陆高校外，本次参赛队伍还有来自美国的加州大学伯克利分校、约翰斯霍普金斯大学、纽约大学；英国的密德萨斯大学、牛津大学、利物浦大学、诺丁汉大学、爱丁堡大学；德国的亚探工业大学、北黑森应用技术大学；俄罗斯的圣彼得堡国立建筑大学；澳大利亚的墨尔本大学、悉尼大学；马来西亚的马来亚大学；日本的柬北大学；法国的巴黎先贤祠-阿萨斯大学；澳门地区的澳门城市大学、澳门科技大学、澳门理工学院、澳门大学；香港地区的北京师范大学-香港浸会大学联合国际学院、香港中文大学、香港科技大学、香港理工大学；中外合作的宁波诺丁汉大学、深圳北理莫斯科大学、西安交通利物浦大学等高校。

目前竞赛具有较高的国际影响力，在国内高校中是作为美赛热身赛、保研加分、综合测评加分、创新奖学金等评定竞赛之一。

联系我们
电话：156-0048-3352（林老师）

>> 微信二维码 <<