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

APMCM 亚太地区大学生数学建模竞赛

参 赛 手 册

亚 太 赛 组 委 会 制 作

目　录

亚太地区大学生数学建模竞赛组委会

2024年APMCM亚太地区大学生数学建模竞赛报名通知

亚太赛[2024]01号

各高等院校：

2024年第十四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2024年11月21日06:00至2024年11月25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20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《2024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/apmcm/2024）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

2024年8月

Notice about 2024 Asia and Pacific Mathematical Contest in Modeling (APMCM)

Every institutes of higher education:

2024 Asia and Pacific Mathematical Contest in Modeling (hereinafter referred to as “contest”) is held by the China Association for the Promotion of International 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 2024 begins at 6AM on Thursday November 21 and ends at 9AM on Monday November 25 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 20, 2024. 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 2024 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
数学建模竞赛组委会

2024年8月

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（2024年修订）

亚太赛[2024]04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。

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、优秀指导教师奖。

第十五条 获得等级奖、优秀组织奖、优秀指导教师奖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2024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

2024年8月

亚太地区大学生数学建模竞赛组委会

2024年APMCM亚太地区大学生数学建模竞赛参赛规则

亚太赛[2024]03号

一、竞赛组织

主办单位：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

二、竞赛日期和时间

注册截止日期：北京时间2024年11月20日

竞赛开始时间：北京时间2024年11月21日（星期四）上午6点

竞赛结束时间：北京时间2024年11月25日（星期一）上午9点

提交论文截止日期：北京时间2024年11月25日上午9点

承诺书及附件提交截止日期：北京时间2024年11月25日上午9点

竞赛结果：竞赛结果预计于2025年1月30日前发布

三、参赛队员承诺

（1）论文中所有图像，图形，图片，表格和附图如果不是队伍原创，就要在引用的位置标注。

（2）所有的直接引语都放在了引号中进行标注，或者进行了特殊的引用注明。

（3）授权APMCM组委会，可将参赛论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

四、竞赛评阅原则

（1）对赛题给出了满意的解读方法，并对模糊概念给予了必要的解释和说明；

（2）明确列出了建模用到的前提条件及假设，并对其合理性给出解释或论证；

（3）通过对赛题的分析给出建模的思路或论证建模的合理性；

（4）设计出了能有效地解决赛题的数学模型；

（5）对模型给出了稳定性的检验；

（6）讨论了模型的优缺点，并给出了清晰的结论；

（7）给出圆满准确的摘要。

五、参赛具体流程和规则

（一）竞赛前

1.报名

（1）所有队伍都必须在北京时间2024年11月20日12点前完成注册。

建议所有的队伍都提前完成注册，因为一旦超过截止时间，报名系统将关闭，不接受任何新报名队伍。

在线报名官网为：https://www.saikr.com/vse/apmcm/2024

a.进入报名官网，点击位于屏幕右侧的报名入口，选择学校所在的省份，再选择学校，如果无法找到自己的学校，可以返回选择其他高校进行报名；填写所有需要的信息包括队伍的电子邮件地址和联系方式。若有多名队员参赛时，需要每个队员都注册，然后由队长邀请其他队友，邀请发出后，其他队友需同意，完成自己的队伍成员，然后由队长在线缴纳注册费。

注意：报名截止后将不能再更改报名信息，确保使用的是当前有效的电子邮件地址以便我们在竞赛前、中、后的必要的时候可以联系到你。

b.如果你是学校集体报名的，请学校集体报名负责人，与竞赛工作人员联系。

c.每所院校参赛队数不作统一规定，也不对指导教师的指导队伍数做限制。

（2）注册费

参赛高校需为每个参赛队伍缴纳100元参赛费，个人报名通过报名官网缴费，学校集体报名，请联系竞赛工作人员，填写集体报名表统一缴费。

（3）个人通过报名网站报名完成后即可得到参赛编号，集体报名的队伍，由工作人员审核后获得参赛编号。

2.团队成员

（1）由队长先注册创建队伍，创建后再邀请队员加入团队。

（2）每支队伍人数上限是3人。

（3）比赛期间，参赛队成员必须是全日制在校大学生。

（二）竞赛开始后

1.从竞赛官网下载赛题

竞赛题目将会在北京时间2024年11月21日（星期四）上午6点在竞赛官网（www.apmcm.org）、北京图象图形学学会官网（www.bsig.org.cn）以及报名主页上公布。

2.选题

每个参赛队可以从三个题目中任选一个题目作答：参赛队可以选择赛题A、B或C。

3.参赛队准备做题

（1）参赛队可以利用任何非生命提供的数据和资料——包括计算机，软件，参考书目，网站，书籍等，但是所有引用的资料必须注明出处，如有参赛队未注明引用的内容的出处，将被取消参赛资格。

（2）参赛队成员不允许向指导教师或者除了本团队成员以外的其他寻求帮助或讨论问题。与除本团队成员以外的人的任何形式的接触都是严格禁止的。这包括通过Email、电话联系、私人交谈、通过网络聊天联络或是其他的任何问答系统，或者其他任何的交流方式。

（3）不完整的论文也是接受的。大赛不存在通过或是不通过的分数分界点，也不会有一个数字形式的分数。APMCM的评判主要是依据参赛队的解决方法和步骤。

（4）摘要页：摘要是论文中的重要部分，（摘要页要作为论文的第一页）。在评卷过程中，摘要占据了相当大的比重，以至于有的时候获奖论文之所以能在众多论文中脱颖而出是因为其高质量的摘要。

好的摘要可以使读者通过摘要就能判断自己是否要通读论文的正文部分。如此一来，摘要就必须清楚的描述解决问题的方法，显著的表达论文中最重要的结论。摘要应该能够激发出读者阅读论文详细内容的兴趣。那些简单重复比赛题目和复制粘贴引言中的样板文件的摘要一般将被认为是没有竞争力的。

除了摘要页以外每篇论文还需要包括以下的一些部分：问题重述、模型假设、符号说明、模型评价、模型分析、参考文献、附录等。

（5）论文需按规范的格式用英文书写，用可读的字体，字号一般为11或12。

（6）整个解决方案必须由正文主体部分、必要的图形、图表或其他类型的材料组成且只能以论文形式递交。不是文档形式的材料如计算机文件或磁盘等将不被接受。

（7）解决方案的每一页的顶部都需要有参赛队的参赛编号以及页码。我们建议在每页上使用页眉，例如：

Team#apmcm2400001Page10of20

（8）参赛队的成员名单、指导教师名单以及学校名称均不能出现在解决方案的任何一页上。整个赛题解决方案不能包含有除了参赛队编号以外的任何身份识别信息。

（9）任何不遵守以上规则的参赛队将被取消参赛资格。

（三）比赛结束前

1.提交论文

要求参赛队（由学生或者指导教师）通过报名官网提交一份PDF格式的论文（不包含附件和承诺书）。论文必须在2024年11月25日上午9点之前提交，否则取消参评资格，论文的命名格式为：题号+参赛编号。

例：Aapmcm2400001

2.打印承诺书并签名：

承诺书和竞赛题目一起公布，公布后，队伍应打印承诺书，打印之后，需每个队员手写签名，然后通过扫描或者拍照的形式把签名的承诺书生成PDF文件

或JPG、PNG图片，命名为：题号+参赛编号+cns，如Aapmcm2400001cns。然后上传至报名官网的“提交承诺书”一栏，格式为PDF、JPG、PNG。

注意：可以用手机或者相机拍照，但要保证图片清晰，保证文字内容均在图片之内。如果不方便打印可以采用电子签名的方式签字。

3.提交支撑材料

将与竞赛相关的其他所有文件（包括程序、数据（赛题中的原始数据除外）和结果等）压缩打包上传。“支撑材料”由参赛队员在审慎考虑的基础上，选择性地添加，如竞赛题目中没有明确的要求，则不作为必须提交的材料。（如果有自己编写的程序，虽然按照论文格式要求，程序必须放入正文附录，但源程序仍然要以支撑材料的形式提供。）命名格式：题号+参赛编号+fj，例Aapmcm2400001fj。然后上传至报名官网的“提交支撑材料”一栏，格式为rar、zip。

4.保留论文附件

为防止差错及评审时验证的需要自己保留附件至获奖最终名单公布之后，参赛队需保存论文的pdf和word版本以及程序、计算结果等在内的附件。

（四）比赛结束之后

1.比赛初稿结果公布

（1）评阅：评阅工作一般在一个月结束，结果将在2025年1月30日之前公布，所有的参赛论文将被评定为SuccessfulParticipant(成功参赛奖)、ThirdPrize（三等奖）、SecondPrize(二等奖)、FirstPrize(一等奖)或InnovationPrize（创新奖）等几个等级。

（2）一旦所有的结果被确定，我们会立刻在网页上公布，请定时登陆报名官网查询比赛结果。

2.竞赛异议期

（1）获奖名单初稿公布之日起的7天内，任何个人或单位可以提出异议，由竞赛组委会负责受理。

（2）欢迎各位监督举报，包括竞赛期间教师参与、队员与他人讨论，抄袭他人论文、不公正的评阅等。对于要求将答卷复评以提高获奖等级的申诉，原则上不予受理，特殊情况可经组委会讨论后，决定是否给予复评。

（3）针对初稿获奖名单，出现学校名称、姓名等错误信息的，请及时上报竞赛组委会。原则上只能改正错误的字，不能更换指导教师姓名、队员姓名，也不能添加队员等。

3. 奖金或证书

此次竞赛将设立等级奖和优秀指导教师奖、优秀组织奖，比例及奖品如表1所示。

由于题目难度不可能完全相同，评审中将向难度较大的题目倾斜，每题的特等奖均为2支队伍，一等奖队数差别不大。

表1获奖等级以及比例

获奖等级	证书	奖金	比例
亚太杯”创新奖	亚太杯”创新奖证书	1000元/队	3支（每题2支）
一等奖	一等奖证书	无	5%
二等奖	二等奖证书	无	15%
三等奖	三等奖证书	无	25%
成功参赛奖	成功参赛奖证书	无	

亚太地区大学生数学建模竞赛组委会

2024年8月

历 年 赛 题

亚太赛组委会制作

2021 Asia and Pacific Mathematical Contest in Modeling

Problem A

Image Edge Analysis and Application

With the development of science and technology, the demand for measurement accuracy of various workpieces and parts is getting higher and higher, and the requirements for measurement instruments are also getting more and more demanding. Various image measuring equipment such as digital image size measuring instrument are now gradually replacing the traditional manual caliper measurement application. Generally, after the camera is calibrated, based on the the dot matrix or checkerboard feature information of the calibrated image, the image can be corrected for distortion and the mapping relationship between the image coordinate space and the world coordinate space can be calculated.

The edge of the target object is very useful for image recognition and computer analysis. Image edge is the reflection of discontinuity of the local characteristics of an image. The edge can outline the target object and make it clear to the observer at a glance. The edge contains rich intrinsic information (such as orientation, step property and shape, etc.), which is an important attribute for extracting image features in image recognition. Image edge contour extraction is a very important processing in boundary segmentation and also a classical problem in image processing. The purpose of both contour extraction and contour tracking is to obtain the external contour features of an image. Applying certain methods where necessary to express the features of the contours to prepare for image shape analysis has a significant impact on performing high-level processing such as feature description, recognition and understanding.

The contour can be described as a set of ordered points, and the common expression of the contour is a polygon. Contours can be either closed or open. The closed contours on an image are all connected start to end, and the open contours generally intersect with the image boundary. In Figure 1, there are five closed contour curves. Although edge detection algorithms such as sobel and canny can detect the image edge pixels boundary based on the difference of image gray value, it does not take the contour as a whole. On an image, a contour corresponds to a series of pixel points. The contour describes a continuous sequence of points, and the edge pixel points can be assembled into a contour curve to describe the edge information of the image.

A sub-pixel is a virtual pixel defined between two physical pixels of an image acquisition sensor. To improve resolution or image quality, sub-pixel calculation is very useful. Image sub-pixel edge extraction is a more accurate method than traditional pixel edge extraction. Sub-pixel

means that the coordinate value of each pixel point on the image is no longer integer positioning, but floating-point number positioning. If the accuracy is increased to 0.1 pixel using subpixel technique, it is equivalent to 10 times higher resolution of image system analysis.

For the following three schematic diagrams, in Figure 1, the object edge contour lines of the image have been extracted and the image edge contour has been segmented into basic graphics such as straight line segments, circular arc segments, and circles. In Figure 2, The edge contour of a rounded rectangle is divided into several geometric shapes. In Figure 3, an elliptical sub-pixel contour curve is shown drawn on the background of a grayscale pixel image grid.

Figure 1. Image Edge Detection

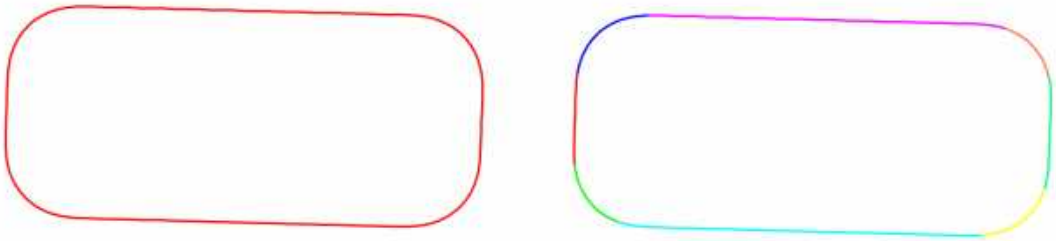


Figure 2. Segmentation Image Edge Contour

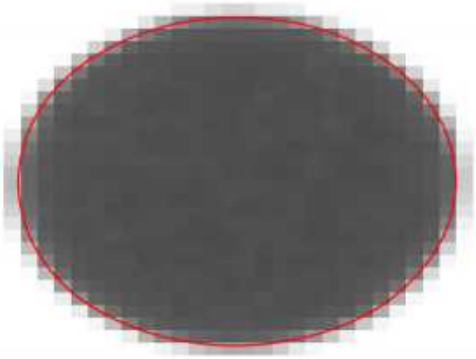


Figure 3. Sub-pixel Edge Contour of Image

Question 1: Build a mathematical model, analyze the method and process of extracting sub-pixel edge with 1/10 pixel accuracy and above, extract sub-pixel edge contour boundaries of the main edge parts of the objects on the three images (Pic1_1, Pic1_2, Pic1_3) in Annex 1, and convert the edge sub-pixel point data into ordered edge contour curve data, with the need to considering how to eliminate the interference effects of edge burrs and shadow parts of the edges. Note that the Pic1_3 image was taken under relatively complex lighting conditions, with more interference information.

a) Please draw the extracted edge contours in different colors on the image, output it as a color edge contour image and save it as png image format for submission. The file names are pic1_1.png, pic1_2.png, pic1_3.png.

b) Output the edge contour data in the format of EdgeContoursOutput.xls file in Annex 1, and output the data of the Pic1_1 and Pic1_2 images to the corresponding Sheet1 and Sheet2 of the worksheet respectively. The output data contains the total edge contours count, the total edge contours length in the image coordinate space, point count and length of each contour curve, and the X and Y coordinate data of each contour point.

c) The total contour curves count on each image and the point count and length data on each curve should be given in the paper. See Table 1, Table 2 and Table 3.

Table 1. Pic1_1 Edge Contour Data Output Format

Total Edge Contours Count		
Total Edge Contours Length		
Edge Contour 1	Length	
	PointCount	
Edge Contour 2	Length	
	PointCount	
…	…	…

Table 2. Pic1_2 Edge Contour Data Output Format

Total Edge Contours Count		
Total Edge Contours Length		
Edge Contour 1	Length	
	PointCount	
Edge Contour 2	Length	
	PointCount	
…	…	…

Table 3. Pic1_3 Edge Contour Data Output Format

Total Edge Contours Count	
Total Edge Contours Length	

Question 2: While the measured image is taken, there is a dot matrix calibration plate placed at the same horizontal height of the target object. The diameter of the dots on the calibration plate is 1 mm, and the center distance between two dots is 2 mm. Annex 2 contains three calibration plate images taken at different angles and one product image (Pic2_4.bmp). Please build a mathematical model, use the calibration plate image information to conduct image rectification analysis of the product image and consider how to calculate, as accurately as possible, the actual physical sizes of the edge segmentation fitting curve segments on the product image. Please calculate the length (mm) of each edge contour, and finally calculate the total edge contours length (mm). According to the contour data labeling shown in Figure 4, output the data results of the table format files such as EdgeContoursLengthOutput.xls in Annex 2.

Table 4. Edge Contour Length Output Format (mm)

Contour ID	Length(mm)
Total Edge Contours	
Edge Contour 1	
Edge Contour 2	
Edge Contour 3	
Edge Contour 4	
Edge Contour 5	
Edge Contour 6	

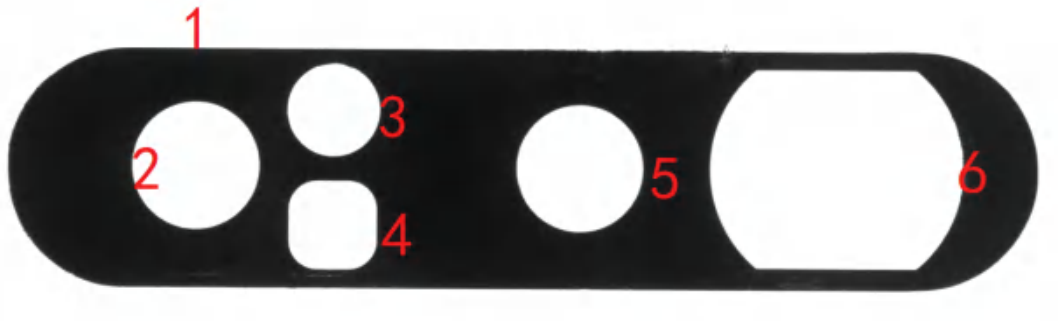


Figure 4. Image Contour Data Labeling

Question 3: Two sub-pixel contour edge data (EdgeContour1.xls and EdgeContour2.xls) are provided in Annex 3, and the shape are shown in Figure 5. Please build a mathematical model, analyze the automated segmentation and fitting of edge contour curve data into straight line segments, circular arc segments (including circles), or elliptical arc segments (including ellipses), and discuss the model method or strategy for automated segmentation and fitting of edge contours. The blue curve starts from the blue digit 1 label and outputs the model calculation result data along the arrow direction. The green curve starts from the green digit 1 label and outputs the model calculation result data along the arrow direction. Please fill in the parameters of the segmented curve segments into the table in the table format. Submit Table 7 and Table 8 (regarding contour 1 and contour 2 segmentation data) in the paper. Note that the type of the lines in this table is populated according to the actual type.

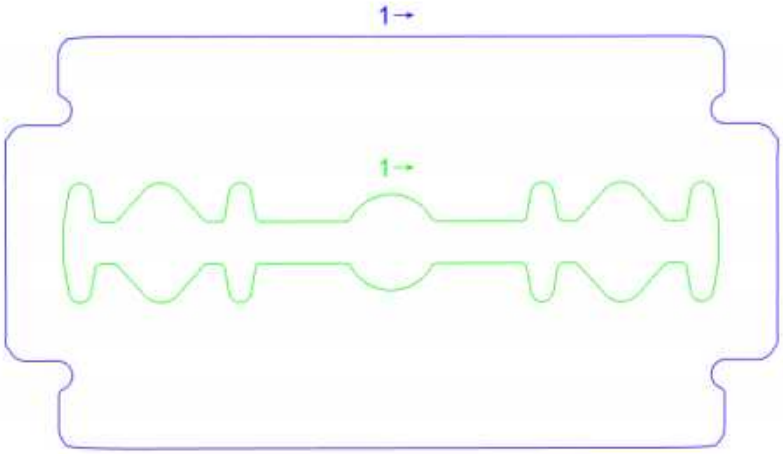


Figure 5. Edge Contour Curve Data

Table 5. Segmentation Data Geometric Shapes Parameter Format

TYPE	PARAM					
Line	StartPt	EndPt	Length (pixel)			
Circle	Center	Radius	Length (pixel)			
CircularArc	Center	StartPt	EndPt	SweepAngle		
Ellipse	Center	Size	RotationAngle			
EllipticArc	Center	Size	StartPt	EndPt	SweepAngle	RotationAngle

Table 6.Demo Coutour Segmentation Data

Edge Contour Demo							
NO	TYPE	PARAM					
S1	Line	(20,100)	(80,100)	60 pixel			
S2	CircularArc	(80,90)	(80,100)	(90,90)	-90°		
S3	Line	(90,90)	(90,60)	30 pixel			
S4	EllipticArc	(90,40)	[20,10]	(90,60)	(90,20)	+180°	-90°

Table 7. Contour 1 Segmentation Data Output Format

Edge Contour 1							
NO	TYPE	PARAM					
S1							
S2							
S3							
S4							
S5							
...							

Table 8. Contour 2 Segmentation Data Output Format

Edge Contour 2							
NO	TYPE	PARAM					
S1							
S2							
S3							
S4							
S5							
...							

Remark:

1. SweepAngle indicates the sweep angle from the start point to the end point, angular system;

2. Size indicates the radius value of specified ellipse or elliptic arc in the X and Y directions;

3. RotationAngle indicates the rotation angle value of specified ellipse or elliptic arc, angular system;

4. For the direction of rotation angle, the rotation direction from positive direction of x-axis to positive direction of y-axis is positive direction, and vice versa is negative direction.

5. All image coordinate points are expressed under the image coordinate system, that is, the upper left corner is the (0,0) origin, the positive direction of the X-axis is to the right, and the positive direction of the Y-axis is downward.

2021 Asia and Pacific Mathematical Contest in Modeling

Problem B

Optimal Design of Thermal Emitter in Thermophotovoltaic Technology

In recent years, the world's major powers have turned their eyes one after another to "the sea of stars", formulating various plans of space exploration. In 2020, China's "Tianwen-1" was launched and navigated through space to Mars; then in 2021, the rover "Zhurong" completed its projected mission, and staying on Mars, it is still working for more findings of the vast universe. To ensure that the various instruments and devices carried by the rover could operate well without sun exposure, and to provide necessary technical support for its long-time work, scientists explored and developed thermophotovoltaic technology. This picture below shows a test prototype of a thermophotovoltaic device.

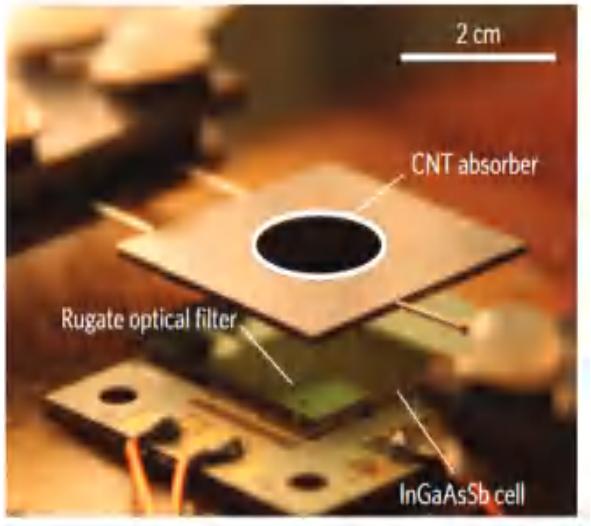


Figure 1: The test prototype of the thermophotovoltaic device

The thermophotovoltaic technology is a technology that uses various heat sources to heat a thermal emitter (absorber), and then converts the infrared radiation of the thermal emitter into electricity through photovoltaic cells. There are many types of heat sources, including chemical energy, solar energy, nuclear energy, etc. The thermal emitter in the system mainly utilizes different material structures to regulate the emission of the heat absorbed, so that most of the emitted photons are below the band-gap wavelength of the photovoltaic cell. The photovoltaic cell mainly converts the high-energy photons that are below a particular band-gap wavelength. It possesses a certain band-gap energy, and therefore a corresponding band-gap wavelength. For example, a silicon solar cell with a band-gap wavelength of 1100 nanometers can only absorb the high-energy photons below the wavelength aforementioned and convert them into

electrical energy, while low-energy photons above the wavelength absorbed by the cell cannot be converted into electrical energy through the photoelectric effect. Instead, they can only be transformed into thermal energy, and the photoelectric conversion efficiency of the cell is thus reduced. Therefore, to improve the thermoelectric conversion efficiency of the thermophotovoltaic system, it is essential to regulate the emission spectrum of the thermal emitter. Calculation methods of the emission spectrum mainly include the Transfer Matrix Method (TMM) [1-2], the Finite Difference Time Domain Method (FDTD), and the Rigorous Coupled Wave Analysis Method (RCWA). And chief factors affecting the emission spectrum of the thermal emitter are the optical properties (refractive index or dielectric constant) and structural properties (thickness) of the material. Wang et al. [3] have developed a sub-micron-thick multilayer selective solar absorber, which is composed of tungsten, silicon dioxide and silicon nitride, and has an absorption rate up to 0.95 in the solar waveband. In 2014, the Evelyn Wang team of the Massachusetts Institute of Technology designed a photon-controlled solar thermophotovoltaic device that operated well in experiments. In their work, the thermal emitter is a multilayer film structure composed of silicon and silicon dioxide. The thickness of each layer is optimized so that its emission spectrum corresponds to the band gap of indium gallium arsenide antimonide (InGaAsSb) batteries. The refractive index or dielectric constant of different materials can be found by searching documents or referring to the optical property database of materials [5], which provides the refractive index of common materials. Please solve the following problems according to the above introduction.

(1) Please express the relationship between the emission spectrum of the single-layer structure and the material properties (refractive index, thickness), and calculate the emission spectrum of a 50-nanometer-thick tungsten (as shown in the picture below) within 0.3-5 microns.

Figure 2: The schematic of 50-nanometer-thick tungsten

(2) Please express the relationship between the emission spectrum of the multilayer structure and the material properties (refractive index, thickness), and calculate the emission spectrum of the composite structure formed by tungsten (50 nm) and silica (50 nm) (as shown in the picture below) within 0.3-5 microns.

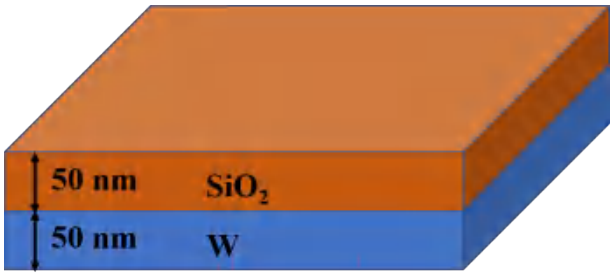


Figure 3: The schematic of multilayer structure formed by tungsten and silica

(3) In order to improve the spectral-control ability of the radiator, sometimes the thermal emitter is designed to emit in a narrow-band form, that is, the emission is concentrated within a very small band, thereby improving the thermoelectric conversion efficiency of the thermophotovoltaic device, such as the multilayer narrow-band emitter designed by Sakurai et al. [6] with silicon, silicon dioxide and germanium. Please select reasonable materials, design a multiplayer thermal emitter to make its emission as narrow-band and high as possible, and give the design parameters of the multilayer structure (including the number of layers, the material and thickness of each layer), as well as its emission spectrum. Note that the idea thermal emitter in this question has the sharp and high thermal emission at 1.5 microns, and the calculated wavelength range is 0.3-5 microns.

(4) The gallium antimonide (GaSb) battery is currently more advanced among others. Assume that its band-gap wavelength is 1.71 microns. The emission spectrum of its idealized thermal emitter is roughly shown by the red-dotted line in the picture below. The blue-dotted line represents the External Quantum Efficiency (EQE), and its influence can be considered properly. Please select reasonable materials, design a multilayer thermal emitter for the GaSb battery to achieve the highest possible thermoelectric conversion efficiency, and give the design parameters of the multilayer structure (including the number of layers, the material and thickness of each layer), as well as its emission spectrum.

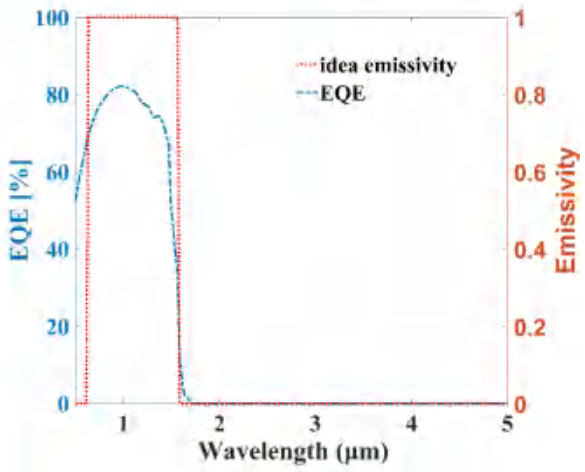


Figure 4: The EQE and ideal emission spectrum of GaSb for thermophotovoltaic

Note: The emission spectra here are all of vertical emission.

Explanation of professional terms:

Thermophotovoltaic: a device that includes heat sources, thermal emitters, photovoltaic cells, and heat-dissipation systems etc. for converting thermal energy into electrical energy.

Thermal emitter: a component that emits thermal radiation when heated.

References:

[1] Mohammed Z H. The fresnel coefficient of thin film multilayer using transfer matrix method tmm[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019, 518(3): 032026.

[2] Katsidis C C, Siapkas D I. General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference[J]. Applied optics, 2002, 41(19): 3978-3987.

[3] Wang H, Alshehri H, Su H, et al. Design, fabrication and optical characterizations of large-area lithography-free ultrathin multilayer selective solar coatings with excellent thermal stability in air[J]. Solar Energy Materials and Solar Cells, 2018, 174: 445-452.

[4] Lenert A, Bierman D M, Nam Y, et al. A nanophotonic solar thermophotovoltaic device[J]. Nature nanotechnology, 2014, 9(2): 126-130.

[5] https://refractiveindex.info/ (Refractive index database)

[6] Sakurai A, Yada K, Simomura T, et al. Ultranarrow-band wavelength-selective thermal emission with aperiodic multilayered metamaterials designed by Bayesian optimization[J]. ACS central science, 2019, 5(2): 319-326.

2021 Asia and Pacific Mathematical Contest in Modeling

Problem C

Construction of Ecological Conservation and Assessment of Its Impact on Environment

Adhere to the concept that lucid waters and lush mountains are invaluable assets, China insists on respecting for, being harmony with and protecting nature, giving high priority to conserving resources, protecting the environment and letting nature restore itself, implementing sustainable development strategies, improving the overall coordination mechanism in the field of ecological civilization, building an ecological civilization system, promoting the transformation of economic and social development toward comprehensive green growth and building a beautiful country. With the help of the Chinese government, China's Saihanba Tree Farm has recovered from the desert and has now become an eco-friendly and green farm with the stable sand prevention function.

Since 1962, 369 young people with an average age under 24 have come to this wasteland filled with yellow sand. From then on, they dedicated their lives here, advanced wave upon wave, to planting seeds in the sand and planting green in the crevices of the stones, like nails fastening millions of acres of forest on the wasteland. Planting trees to fix the sand and conserve water sources, they build a green barrier to block wind and sand. Today, the forest coverage in Saihanba area has reached 80%. It supplies Beijing and Tianjin with 137 million cubic meters of clean water each year, sequesters 747,000 tons of carbon, and releases 545,000 tons of oxygen.

With over a half century's struggle, the world's largest artificial forest was built on the earth of Saihanba. Expanding afforestation of 1.12 million mu with more than 400 million trees, the builders created a green sea on the plateau wasteland which is 400 kilometers north of Beijing.

On the one hand, there is the historical mission that "civilization will develop followed with zoology thriving". On the other hand, there are new issues encountered on the road of green development. Therefore, Saihanba people now have a higher goal, which is to restore the ecology. Since the 18th National Congress of the Communist Party of China, they have successively launched three major projects, namely, afforestation, natural improvement of artificial forests, and near-naturalization cultivation of natural forests. They have tried to make artificial forests closer to natural ones.

Please build the mathematical models by your team and answer the following questions:

1. Saihanba plays an important role in resisting wind and sand, protecting environment, maintaining ecological balance and stability, etc. Please select appropriate indicators, collect relevant data, and build the evaluation model for the impact of Saihanba on the ecological environment, in order to quantitatively evaluate the impact on the environment after the restoration of Saihanba; that is, to comparatively analyze the environmental conditions before and after the Saihanba restoration.

2. The restoration of Saihanba Tree Farm has played an important role in resisting sandstorms for Beijing. Please select appropriate indicators and collect relevant data to build a mathematical model for evaluating the Saihanba's impact on Beijing's ability on sandstorm resistance, and quantitatively evaluate Saihanba's role in resisting sandstorms in Beijing.

3. Suppose we plan to extend the ecological protection model of Saihanba to the whole country, please build a mathematical model and collect relevant data, to determine which geographical locations in China need to build ecological areas (ie. Ecological Reservation), and fix the number or scale of ecological areas to be built; moreover, to evaluate its impact on achieving China's carbon neutral target.

4. China's Saihanba ecological protection model has made an example for the Asia-Pacific region. Please choose another country from the Asia-Pacific region to establish a mathematical model and collect relevant data, and then discuss which geographical locations in this country need to build an ecological area (ie. Ecological Reservation), as well as determine the number or scale of ecological areas to be built; moreover, to evaluate its impact on absorbing greenhouse gases and mitigating carbon emissions.

5. Please write a non-technical report to the Asia-Pacific Mathematical Contest in Modeling Organizing Committee (APMCM), describing your models, and proposing feasible plans and suggestions for building ecological reservation.

Tips: In the process of building the models, you can consider condition of the existing ecological forests in China and other Asia-Pacific regions; the requirement of growth environment for different trees (that is, the specific area for certain tree); how to balance the layout of ecological forest land, economic development land and industrial land; and whether the aimed geographical area has enough land available to develop the ecological reservation.

Reference: http://lycy.hebei.gov.cn/shb/show_article.php?id=5876

2022 Asia and Pacific Mathematical Contest in Modeling

Problem A

Feature Extraction of Sequence Images and Modeling Analysis of Mold Flux Melting and Crystallization

Mold fluxes in continuous casting process thermally insulate the molten steel meniscus, prevent reoxidation of liquid steel during continuous casting of liquid steels, control heat transfer, provide lubrication of strand, and absorb nonmetallic inclusions. Metallurgical functions of mold flux is mainly determined by its melting rate and crystallization rate under the temperature control curve. It is therefore important to study the phase distribution of mold fluxes in the gap between mold wall and strand shell.

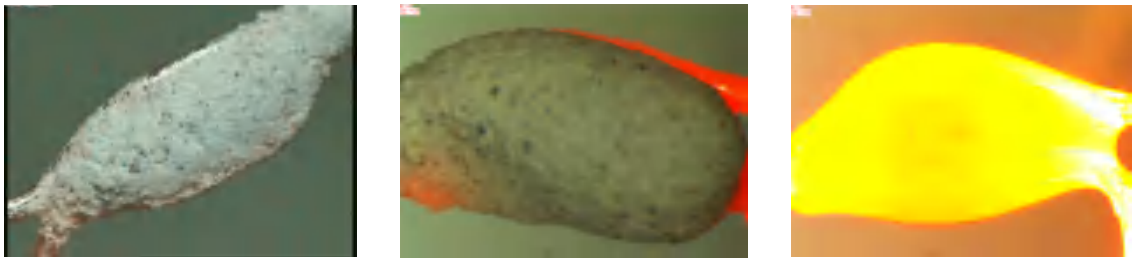
Continuous casting mold fluxes are added to the top of liquid steel in the mold. These solid slags, accumulating on the surface of liquid steel as a powder layer, can prevent liquid steel level crusting due to excessive temperature drop of liquid steel. The temperature of mold fluxes then gradually rise to the melting point, and mold fluxes are melted to form a sintered layer. Raw materials of mold fluxes form low-melting-point substances and then liquid slag through chemical reactions, and the composition of mold fluxes will change to a certain extent. **It is melting process**.

As mold fluxes are completely fused, a liquid slag layer will be formed and covering on the surface of liquid steel. The slag film will be formed when the liquid slag infiltrates from the slag pool at the steel liquid surface into the gap between the shell and the copper mold wall. The slag against the strand still maintains liquid phase, because of the high temperature of the strand surface. However, as the temperature of liquid slag decreases with that of the strand surface in the longitudinal direction of the mold, the slag film, against the copper mold wall, is quenched and solidified to form a glassy solid slag film (solidification behavior of slag film), with mold's forced cooling, while slag film will crystallize at certain areas and form a crystalline layer (crystallization behavior of slag film) under suitable conditions, finally creating a typical three-layer slag film structure: glassy layer, crystalline layer and liquid slag layer. **This process is crystallization.**

Because of the high temperature, transient fluid flow, complex phase transitions and chemical reactions as well as the opacity of mold wall, it is difficult to observe the phase changes of mold fluxes directly. The SHTT II tester of melting and crystallization temperature

is now widely applied to observe the crystallization behaviors of mold fluxes. After the experiment is finished, experimenters demonstrate the images one by one, record the information in the upper left corner of the images, and identify the key node images with naked eyes and experience (see Figure 1), so as to guide the design of mold fluxes to meet the solidification requirements of steel grades. This process wastes manpower and hinders the development of experimental process information. It is urgent to develop automatic feature extraction and mathematical modeling technology of sequence images.

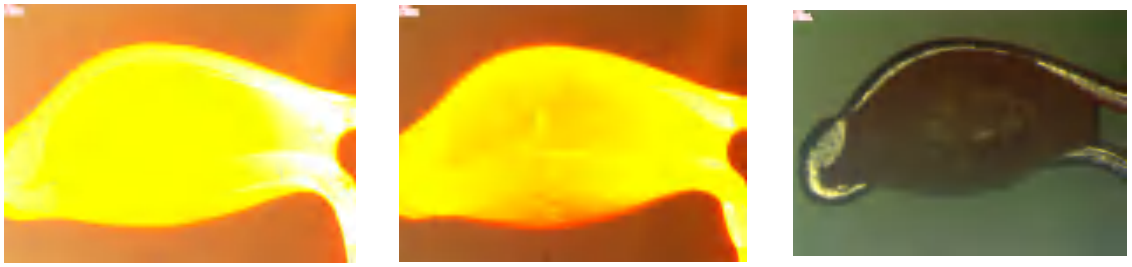
Attachment 1 has 562 sequence images of mold fluxes' melting and crystallization. These sequence images are collected from the 110th to 671st seconds when the experiment starts. The file serial numbers follow the collection time sequence, and images are collected every 1s. The information is presented by digital images in Attachment 1 (see Figure 1). The upper left corner of each image is marked with the corresponding time of the image and the temperature values of No.1 thermocouple and No.2 thermocouple.



(a) The sample

(b) The sample begins to melt at 110s

(c) The sample is completely melted at 141s



(d) The sample sees steels melting at 149s

(e) Crystallization begins at 150s

(f) Crystallization ends at 671s

Fig. 1 Melting and crystallization of mold fluxes

To realize automatic feature extraction and mathematical modeling of sequence images of mold fluxes melting and crystallization, please answer the following three questions.

Question 1: With image segmentation and recognition or other technologies, please automatically extract temperatures of No.1 thermocouple and No.2 thermocouple in the upper left corner of each image and import them automatically into the corresponding table in

Attachment 2 (please write a step-by-step technical operation document), and please make a temperature-time curve diagram (1#wire temperature-2#wire temperature-time diagram; 1#wire average temperature-2#wire average temperature-time diagram). In addition, the test result of 1#wire or 2 #wire is inaccurate. Please point it out and explain.

Question 2: According to the six node images in Fig.1, please study and quantify the dynamic differences between adjacent sequence images in the process of mold fluxes melting and crystallization by applying digital image processing technology. On this basis, please make a time series modeling of the quantified different characteristics, and discuss the melting and crystallization process curve of mold fluxes based on the simulation results of mathematical model.

Question 3: Given the changes of temperature and time, as well as the research results of Question 2, please make a mathematical model to discuss the functional relationship between the changes of temperature and time as well as the melting and crystallization process of mold fluxes, and discuss the kinetics of melting and crystallization of mold fluxes (the relationship between temperature, melting rate and crystallizing rate) based on numerical simulation results.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.

Note: The APMCM Contest has a 25-page limit. All aspects of your submission count toward the 25-page limit (Summary Sheet, Table of Contents, your complete solution). However, the pages of Reference List and Appendices are not limited.

Attachment:

Attachment 1.zip, download on the website: https://share.weiyun.com/ubtXPGz0

Attachment 2.xlsx

2022 Asia and Pacific Mathematical Contest in Modeling

Problem B

Optimal Design of High-speed Train

On April 12, 2022, China High Speed Railway Fuxing CR450 multiple units successfully achieved a single train speed of 435 km/h and a relative speed of 870 km/h on the open line, creating a world record for the speed of high-speed rail multiple units trains crossing open lines and tunnels. The new generation standard EMU "Fuxing" is a new generation of high-speed train independently researched and developed by China with full intellectual property rights. It integrates a large number of modern domestic high-tech, and achieves important breakthroughs in key technologies such as traction, braking, network, bogie, and axle. It is another major achievement of China's scientific and technological innovation. Figure 1 is a simplified model of the high-speed rail geometric structure.

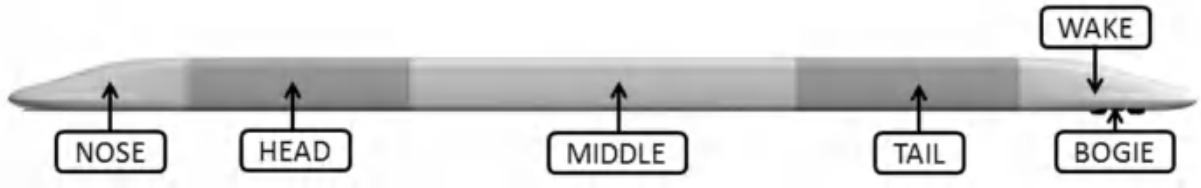


Figure 1. Simplified model of high-speed rail geometry.[1]

The head structure of China's high-speed railway is bullet head, and the duckbill structure is adopted by Japan's high-speed railway. Figure 2 shows the simplified models of four typical high-speed railway head structures including the types of TP1, TP2, TP3 and TP4. Among them, the design of high-speed rail head structure mainly considers air resistance and noise level.

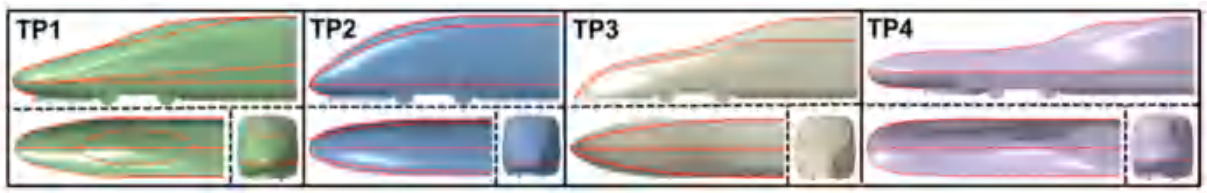


Figure 2. Simplified models of four typical high-speed rail head structures.[2]

The design process of bullet head of high-speed railway should not only take aerodynamics as the basic principle, but also conduct simulation and experiment repeatedly. To achieve the optimal scheme between the air flow, aerodynamic force, other relevant parameters around the car head and the car body can be optimized after thousands of calculations and experiments.

Figure 3 shows the different areas of the streamlined high-speed rail head structure.

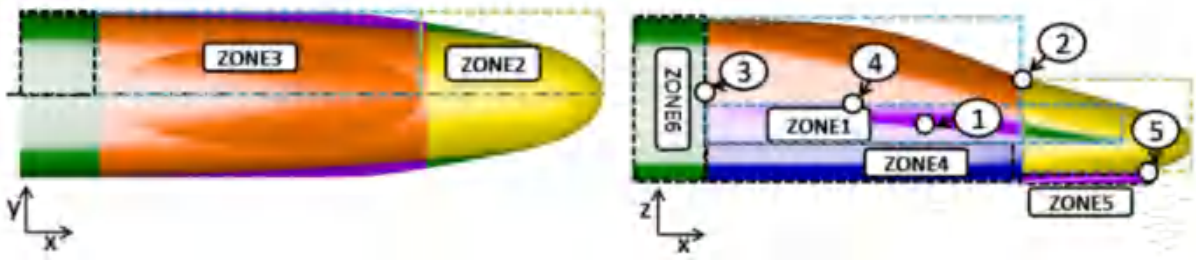


Figure 3. Different areas of streamlined high-speed rail head structure.[3]

Please collect relevant data, establish several mathematical models, and answer the following questions:

Question 1: Please establish a mathematical model for the air resistance of the high-speed railway, consider the relationship between the high-speed railway geometry and the force under the general condition and extreme weather (such as rain, snow and wind), simulate the distribution of air resistance for the conical type and the four typical high-speed railways as shown in Figure 2, and select the best high-speed railway shape with the smallest air resistance.

Question 2: Please analyze the influence of the radian of the high-speed rail head's curve on the air resistance, establish an optimization model of the high-speed rail's shape, design the best high-speed rail's shape, making the high-speed rail suffer the least air resistance, and draw a sketch of the optimized high-speed rail's shape.

Question 3: Please establish a mathematical model for the noise generated by high-speed rail, analyze the intensity of the noise generated by the conical type and the four typical high-speed rail as shown in Figure 2, simulate the distribution of their respective noise, and select the best high-speed rail shape which generates the least noise.

Question 4: Please combine the results of the previous three questions to establish a comprehensive optimization model of the high-speed rail shape and design the best high-speed rail shape, while improving the speed of high-speed trains and reducing noise. Draw a sketch of the high-speed rail's shape and give the corresponding structural parameters.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.

Note: The APMCM Contest has a 25-page limit. All aspects of your submission count toward the 25-page limit (Summary Sheet, Table of Contents, your complete solution). However, the pages of Reference List and Appendices are not limited.

Attachment:

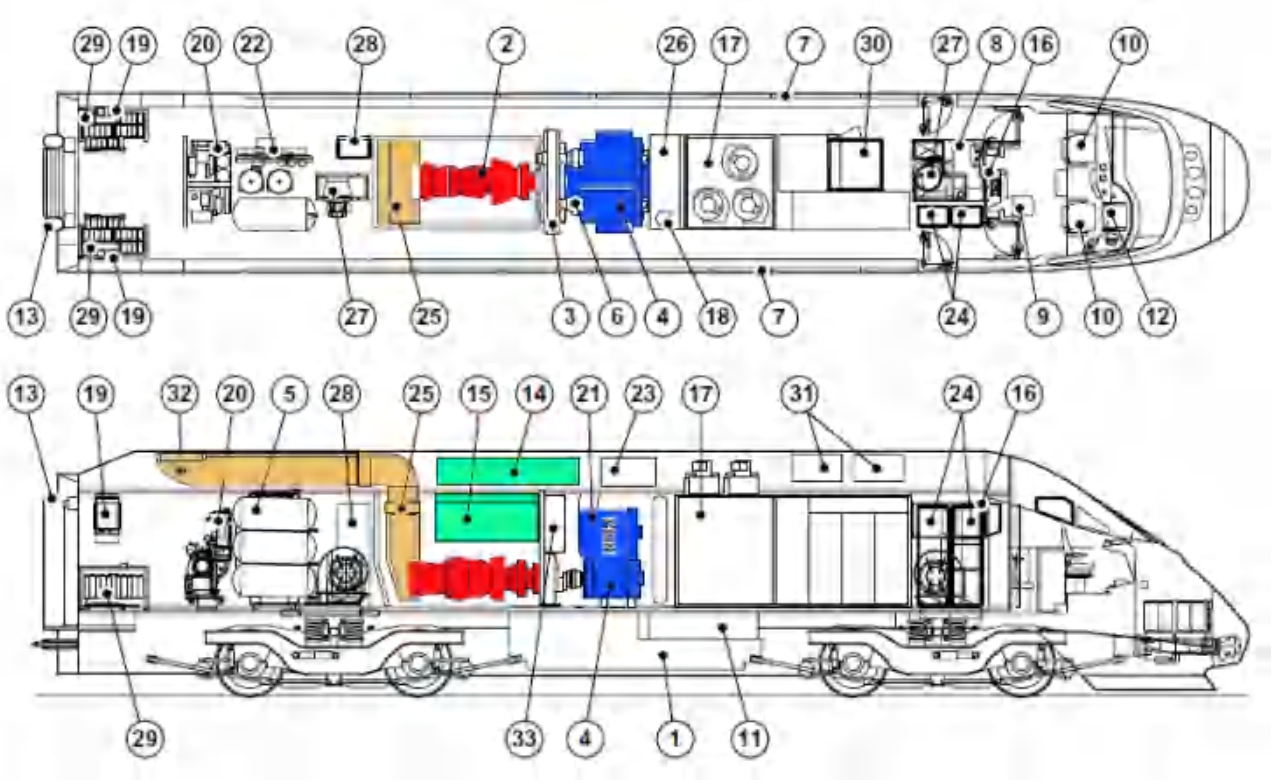


Figure 4. Schematic diagram of different regional structures of general high-speed railway.

References:

[1] Zhenxu Sun, Shuanbao Yao & Guowei Yang (2020) Research on aerodynamic optimization of high-speed train's slipstream, Engineering Applications of Computational Fluid Mechanics, 14:1, 1106-1127, DOI: 10.1080/19942060.2020.1810128.

[2] Sun, Zhenxu, Shuanbao Yao, Lianyi Wei, Yongfang Yao, and Guowei Yang. 2021. "Numerical Investigation on the Influence of the Streamlined Structures of the High-Speed Train's Nose on Aerodynamic Performances" Applied Sciences 11, no. 2: 784.

https://doi.org/10.3390/app11020784

[3] Sun Zhenxu, Yao Yongfang, Guo Dilong, Yang Guowei, Yao Shuanbao, Zhang Ye, Chen Dawei, Li Guibo, Shang Keming, Jia Ling. RESEARCH PROGRESS IN AERODYNAMIC OPTIMIZATION OF HIGH-SPEED TRAINS 1). Chinese Journal of Theoretical and Applied Mechanics[J], 2021, 531: 51-74 DOI:10.6052/0459-1879-20-205.

[4] Wang, M.Y., Hashmi, S.A., Sun, Z.X. et al. Effect of surface roughness on the aerodynamics of a high-speed train subjected to crosswinds. Acta Mech. Sin. 37, 1090–1103 (2021). https://doi.org/10.1007/s10409-021-01099-7

[5] Sun, Z.X., Wang, M.Y., Wei, L.Y. et al. Aerodynamic shape optimization of an urban maglev train. Acta Mech. Sin. 37, 954–969 (2021). https://doi.org/10.1007/s10409-021-01094-y

[6] Yao, Y., Sun, Z., Li, G. et al. Aerodynamic optimization using passive control devices near the bogie cabin of high-speed trains. Acta Mech. Sin. 38, 321363 (2022). https://doi.org/10.1007/s10409-022-21363-x

[7] Wang, M, Sun, Z, Ju, S, & Yang, G. "Influence of Rectangular Strips' Size on Aerodynamic Performance of a High-Speed Train Subjected to Crosswind." Proceedings of the ASME 2021 Fluids Engineering Division Summer Meeting. Volume 2: Fluid Applications and Systems; Fluid Measurement and Instrumentation. Virtual, Online. August 10–12, 2021. V002T03A003. ASME. https://doi.org/10.1115/FEDSM2021-65692

[8] Wang, Y, & Sun, Z. "Influence of the Topological Structures of the Nose of High-Speed Maglev Train on Aerodynamic Performances." Proceedings of the ASME 2021 Fluids Engineering Division Summer Meeting. Volume 2: Fluid Applications and Systems; Fluid Measurement and Instrumentation. Virtual, Online. August 10–12, 2021. V002T03A004. ASME. https://doi.org/10.1115/FEDSM2021-65711

[9] Sun, Z, Yao, Y, Kong, F, & Yang, G. "Numerical Study on Unsteady Wake Characteristics of an Urban Maglev Train." Proceedings of the ASME-JSME-KSME 2019 8th Joint Fluids Engineering Conference. Volume 3A: Fluid Applications and Systems. San Francisco, California, USA. July 28–August 1, 2019. V03AT03A003. ASME. https://doi.org/10.1115/AJKFluids2019-5041

2022 Asia and Pacific Mathematical Contest in Modeling

Problem C

Global Warming OR Not?

Canada's 49.6 °C has set a new temperature record for regions above 50 °N of the earth, and hundreds of people died of heat within a week; Death Valley, California, USA, is 54.4 °C, which is the highest temperature ever recorded on the earth; Kuwait 53.5 °C, even more than 70 °C under direct sunlight, and more than 50 °C in many countries in the Middle East.

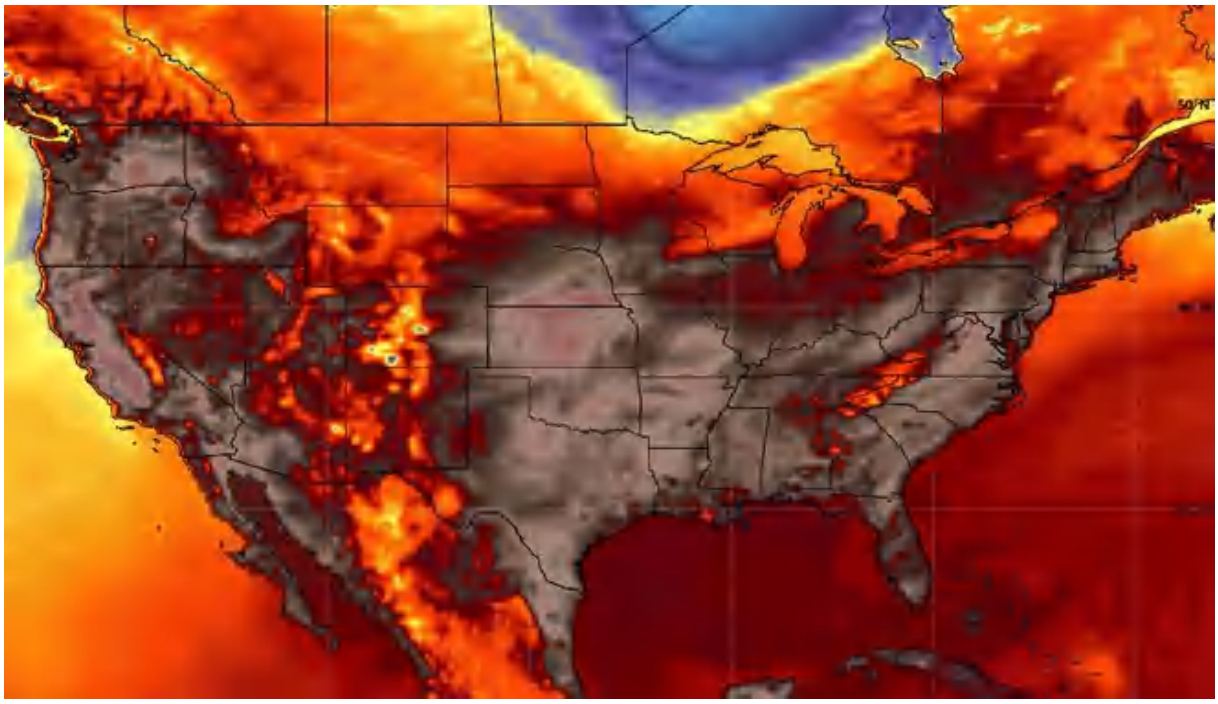


Figure 1. The high temperature in North American.

Since this year, we have seen a large amount of amazing temperature reports. The reality that the earth is burning is beyond doubt. Following the terrible high temperature in these regions from the end of June to the beginning of July, Italy once again set a European temperature record, reaching an astonishing 48.8 °C, and many countries declared a state of emergency.

Global climate warming is a phenomenon related to nature. It is due to the continuous accumulation of greenhouse effect, which leads to the imbalance of energy absorbed and emitted by the earth atmosphere system, and the continuous accumulation of energy in the earth atmosphere system, leading to temperature rise and global climate warming.

Prior to the Industrial Revolution, carbon dioxide (CO_2) in the atmosphere was consistently around 280 parts per million (ppm). The concentration of CO_2 in the atmosphere reached 377.7 ppm in March of 2004, resulting in the largest 10-year average increase up to that time.[1] According to scientists from National Oceanographic and Atmospheric Administration (NOAA) and Scripps Institution of Oceanography (SIO) the monthly mean CO_2 concentration level peaked at 421 ppm in May 2022.[2] An Organisation for Economic Co-Operations and Development (OECD) report predicts a CO_2 level of 685 ppm by 2050.[3]

The APMCM organizing committee has asked your team to address these claims of the current reported and future predictions of global temperature levels. They provided data set 2022_APMCM_C_Data.csv which contains 239177 records to assist in your research.

Requirements

1. Do you agree with the claims of global temperature? Use 2022_APMCM_C_Data.csv in the attachment and other datasets your team collects to analyze global temperature changes.

a) Do you agree that the increase of global temperature in March 2022 resulted in a larger increase than observed over any previous 10-year period? Why or why not?

b) Based on the historical data, please build two or more mathematical models to describe the past and predict the future global temperature level.

c) Use each of your models in 1(b) to predict global temperatures in 2050 and 2100, respectively. Do any of your models agree with the prediction that the average global temperature of observation points in 2050 or 2100 will reach 20.00 °C? If not in 2050 or 2100, when will the average temperature of observation points in your prediction models reach 20.00 °C?

d) Which model you built in 1(b) do you consider most accurate? Why?

2. What are the reasons that affect the temperature change?

a) Use the results of question 1 and the data in the attachment 2022_ APMCM_C_ Data.csv and other datasets collected by your team, build a mathematical model to analyze the relationship (if any) between global temperature, time and location, and explain the relationship or prove that there is no relationship between them.

b) Please collect relevant data and analyze the factors of natural disasters (such as volcanic eruptions, forest fires and the COVID-19). Is there any impact on global temperature?

c) What do you think is the main reason that affects the global temperature change?

d) Do you think there are some measures to curb or slow down global warming?

3. Prepare a non-technical article (1 page maximum) Please write a non-technical article (1 page at most) to the APMCM organizing committee, explaining your team's findings and suggestions for the future.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.
- One-page non-technical Article.

Note: The APMCM Contest has a 25-page limit. All aspects of your submission count toward the 25-page limit (Summary Sheet, Table of Contents, Article). However, the pages of Reference List and Appendices are not limited.

Attachment:

2022_APMCM_C_Data.csv

Data Source: Berkeley Earth data page http://berkeleyearth.org/data/

References:

[1] National Oceanographic and Atmospheric Administration. NOAA Earth System Research Laboratory. (2022, October). Trends in atmospheric carbon dioxide [Internet]. https://gml.noaa.gov/ccgg/trends/data.html

[2] National Oceanographic and Atmospheric Administration. NOAA Research News & Features. (2022, June 3). Carbon dioxide now mare than 50% higher than pre-industrial levels [Internet].https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels.

[3] Organisation for Economic Co-Operations and Development. (2012). The OECD environmental outlook to 2050 [Internet]. https://www.oecd.org/env/cc/Outlook%20to%202050_Climate%20Change%20Chapter_HIGLIGHTS-FINA-8pager-UPDATED%20NOV2012.pdf.

2022 Asia and Pacific Mathematical Contest in Modeling

Problem D

Structural Optimization of Heat Transfer Fins in the Energy Storage System

Efficient energy storage technology is the core technology to solve the volatility and intermittency of renewable energy and waste heat resources. Phase change heat storage is widely used because of its high energy storage density and heat storage and release at a nearly constant temperature. Solid-liquid phase change materials (PCMs) feature high latent heat and small volume change before and after phase change, which makes them comparatively easy to store and encapsulate. However, their generally low thermal conductivity factor leads to a prolonged heat storage and release process, which has become a key factor hindering the broad application of PCMs. To improve the rapid heat transfer capability of a heat storage system, we need to research the important topic to optimize the structural design and parameters of the system. At present, adding fin(s) has been widely used as a simple, economical, and effective means to enhance the solid-liquid phase change heat transfer process.

Fig. 1 Heat storage tank adopting phase change heat storage technology

A company needs to design the structure of heat transfer fins in the tank of the phase change heat storage system to further improve the heat transfer performance of heat storage products. The core component of the heat storage system is a shell-and-tube heat storage tank, as shown in Figure 1. The cross-section is shown in Figure 2. The circular ring needs to be filled with heat storage PCMs and fin structures. When the PCMs absorb heat, the high-temperature working fluid circulates in the inner tube, storing and utilizing its waste heat. When the PCMs release heat, the low-temperature working fluid circulates in the inner tube, absorbing and reusing the heat stored in the PCMs.

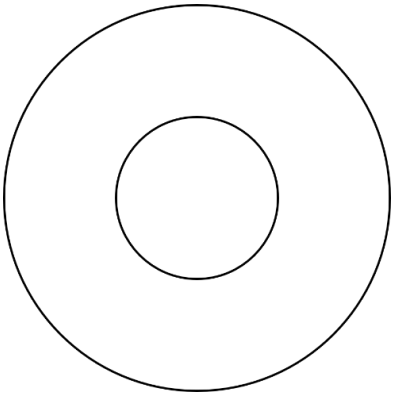


Fig. 2 Cross-section of phase change heat storage tank

The company plans to develop a small phase change heat storage tank, in which the inner radius of the tank is 0.02 m, the outer radius of the tank is 0.05 m, the thermal conductivity of the fins is 214 W/(m K), the density of PCMs is 780 kg/m3, the thermal conductivity of the phase change material is 0.15 W/(m K), and the phase change temperature is 333 K. The exterior of the tube is adiabatic, and the interior is filled with working fluid with a temperature of 373 K. Now your team is needed to answer the following questions to improve the heat transfer rate of the heat storage system.

Question 1: Assume that the cross-section of the heat storage tank with evenly distributed rectangular fins as shown in Fig. 3 is adopted. The length of rectangular fins is 0.018 m, the width is 0.006 m, and the interval angle between the fins is θ. PCMs in a solid state are heated by absorbing the heat energy of the working fluid in the tube. Please model the heat transfer process in the heat storage tank, optimize the interval angle θ, and state the time it takes for the average temperature of PCMs to rise from room temperature (293 K) to the phase change temperature in this case.

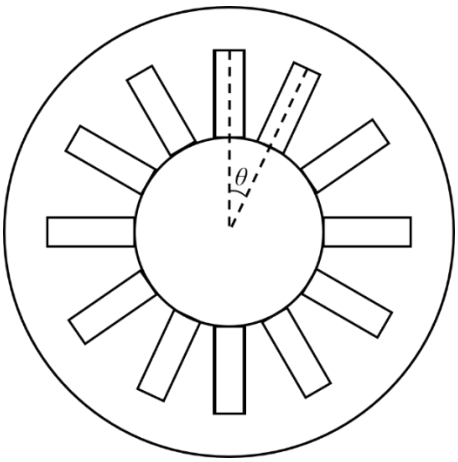


Fig. 3 Cross-section of phase change heat storage tank (rectangular fin structure)

Question 2: The shape and geometric size of the fins can have great influence on the heat transfer rate of the heat storage tank. Please take Fig. 4 as an example to optimize the size and distribution of triangular fins, study the influence of their size on the heating rate of PCMs, and compare and analyze their heat transfer effect with the rectangular fin structure in Question 1.

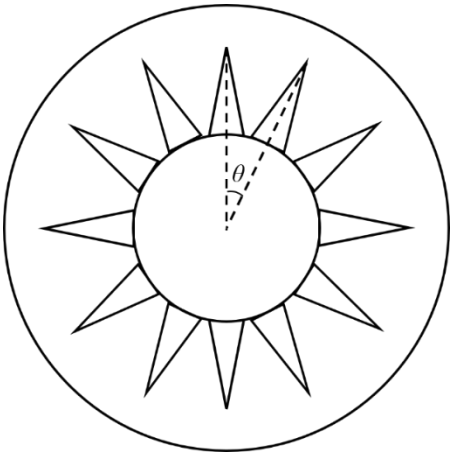


Fig. 4 Cross-section of phase change heat storage tank (triangular fin structure)

Question 3: Further optimize the design of the shape, parameters, and spatial distribution of fins to achieve the optimal heat transfer capacity of PCMs.

Question 4: Please write a letter to the company to suggest the fin design of the heat storage tank.

2022 Asia and Pacific Mathematical Contest in Modeling

Problem E

How Many Nuclear Bombs can Destroy the Earth?

On August 6, 1945, the Second World War was drawing to a close. In order to end the war as soon as possible, the United States dropped the next atomic bomb called "Little Boy" in Hiroshima, Japan. Such an atomic bomb killed 200000 people in Hiroshima, and all buildings in Hiroshima collapsed. This is the first time in human history that the atomic bomb has been put into practice, and it also makes many people realize the terrible power of the atomic bomb for the first time.

Figure 1. Simulation diagram of atomic bomb explosion.

Nuclear weapons refer to huge lethal weapons related to nuclear reaction, including hydrogen bombs, atomic bombs, neutron bombs, etc. Nuclear weapons are one of the most powerful weapons ever developed by human beings, and they often remind people of the scene of destroying heaven and earth. The instantaneous explosion temperature of an atomic bomb can reach tens of millions of degrees. The explosive yield of the atomic bomb is about tens of thousands to hundreds of thousands of tons of TNT equivalent. The explosion of an atomic bomb and the area of its radiation can destroy a city.

Because of the terrible power of the atomic bomb, many countries hope to use it to deter other countries and protect their countries from foreign invasion. After World War II, all

countries in the world began to research and manufacture atomic bombs crazily, and even created destructive weapons such as "Big Ivan".

The "Big Ivan" is the most powerful nuclear bomb known in the world at present, that is, the "czar bomb" built in the Soviet period. It is not only the most powerful nuclear bomb in the world, but also the largest nuclear bomb. How big is the "Tsar Bomb"? The data shows that its length is 8 meters, diameter is 2.1 meters, weight is up to 27 tons, and design TNT equivalent is 50 megatons. The Soviet Union originally planned to design an equivalent of 100 megatons, but because the destructive force was too great to find a suitable test site, the power was reduced by half. Even so, the Czar Bomb is still the most powerful nuclear weapon in the world. The explosive power of the "little boy" is about 14000 tons of TNT equivalent, making Hiroshima a ruin, while the power of the "tsar bomb" is comparable to dozens or even hundreds of "little boys". After the test explosion of the "Tsar Bomb" on the Soviet Union's Xindi Island, even in some states as far away as the United States, an earthquake of about magnitude 5 was detected, resulting in a 9 mm southward movement of Eurasia.

Many people believe that these nuclear weapons can destroy the earth several times. Is this really the case? As far as the power of nuclear weapons developed by mankind is concerned, if they destroy the earth, they do not mean that they can blow the earth into pieces, but that the living environment of human beings and creatures on the earth has been destroyed.

The APMCM Organizing Committee requires your team to address the current report and future nuclear weapons projections. They provided data set 2022_APMCM_E_Data.csv to help you with your research. Please collect the corresponding data, establish a mathematical model and answer the following questions.

Requirements

1. Basic data analysis

a) Which countries have ever possessed nuclear weapons?

b) Which country has the largest reduction or increase in its nuclear weapons stockpiles in the last 20 years?

c) During which five years did nuclear weapon tests occur the most?

d) Which country has been the most active in nuclear weapons research in the last 10 years?

e) Which country has made the fastest transition from "not considering nuclear weapons" to "possessing nuclear weapons"?

2. Predict the number of nuclear weapons

a) According to the attached data or the data you collected, establish a mathematical model to predict the number of nuclear weapons, and predict the countries with nuclear weapons in the next 100 years;

b) Predict the change trend of the number of nuclear weapons in the next 100 years, the total number of nuclear weapons in 2123, and the number of nuclear weapons in each country.

3. Protect our planet

a) Establish an mathematical model for the detonation position of nuclear weapons, and calculate how many nuclear bombs are required at least to destroy the earth?

b) According to the mathematical model, what is the maximum destructive power of the nuclear bomb currently possessed? Is it enough to destroy the earth?

c) In order to protect the earth and the environment on which we live, what should the total number of nuclear bombs in the world be limited to, and what should the countries that already have nuclear weapons be limited to theoretically?

4. Prepare a non-technical article (1 page maximum). Please write a non-technical article (1 page at most) to the United Nations (U.N.), explaining your team's findings and putting forward several suggestions for all countries.

Attachment

2022_APMCM_E_Data.xlsx

Position sheet status: 0-Does not consider, 1-Considers, 2-Pursues, 3-Possesses

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.

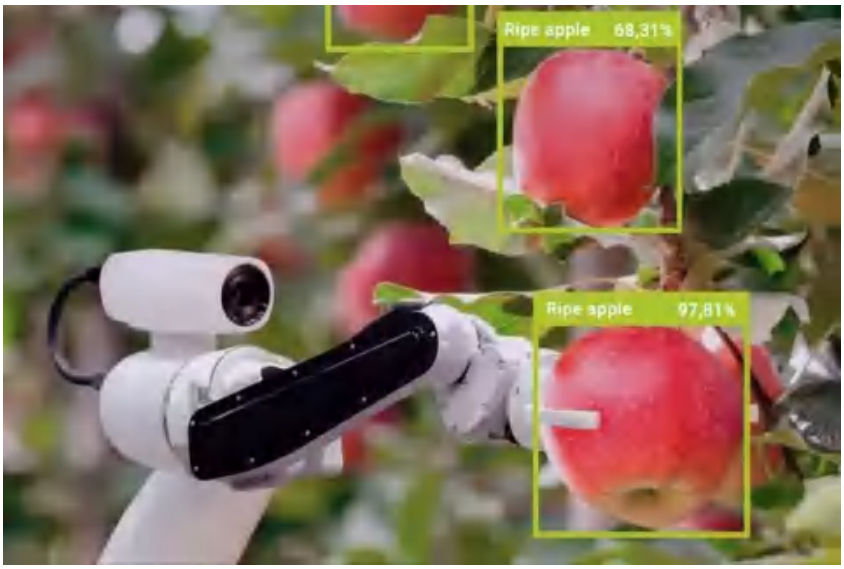


Figure 1. Diagram of image recognition of apples by fruit-picking robots.

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:

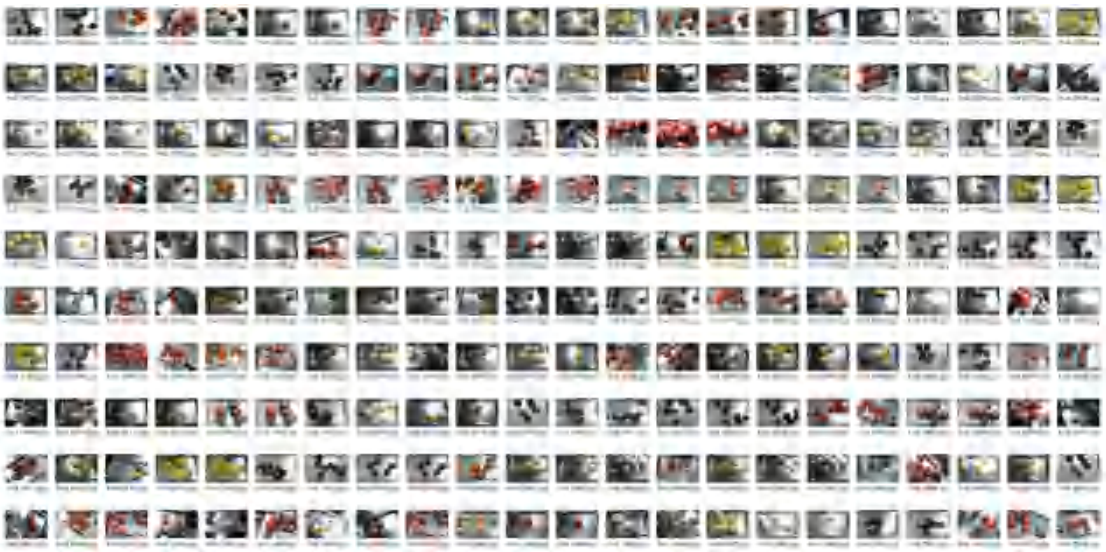
The dataset of pears:

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℃.

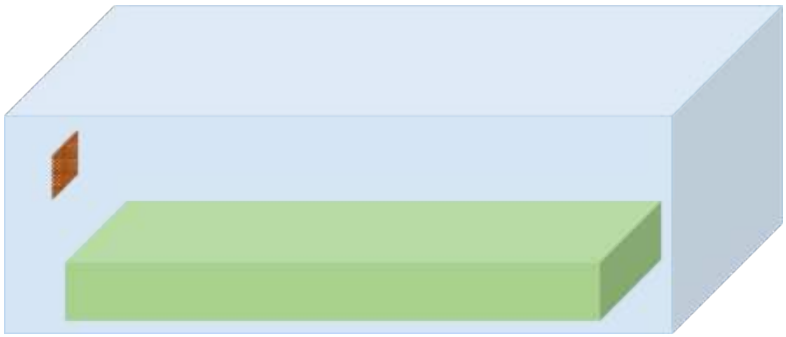


Figure 2. Schematic diagram of glass greenhouse structure

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

亚太赛组委会制作

Team Number :	apmcm2103847
Problem Chosen :	A

Research On The Application Of Sub-pixel Edge Detection Technology In The Precision Measurement Of Workpiece Dimensions

Summary

As technology improves by leaps and bounds, the requirements for measurement instruments are increasingly higher. This article introduces edge image segmentation and Canny algorithm and sub-pixel images. The actual length of the part is measured based on the relationship of the calibration plate pixels to the real distance. Finally, the Ramer algorithm is used to fit the image, which then accurately extract the actual data of the image edge profile.

For Question 1, by drawing the subpixel edge contour, it is first necessary to transform the edge subpixel data into the orderly edge contour curve data, considering how to eliminate the interference effect of the edge hairy margin and shadow parts. Two images, Pic1_1 and Pic1_2, are processed in the same way, sharpening the images using the Laplace operator, the image contour smoothed, followed by thresholding of OTSU binarization and Gaussian filtering for denoising, and finally coarse segmentation with the Canny algorithm. Processing Pic1_3 images were taken under conditions that were environmentally complex. The PSO-OTSU algorithm is used to reduce the highlight of the image, and finally the part defect detection and contour repair to obtain the ideal binarized image for contour inspection. Then the subpixel edge detection is to determine whether the image has reached 0.1 pixels, and finally the image edge profile is segmented by the Ramer algorithm.

For Question 2, when measuring the true length of the given parts pictures, measure it according to the calibration version of the given level. The true length of the provided product image is measured according to the calibration board images of three different angles provided.

For image pre-processing, OTSU binarization segmentation, Gaussian filtering denoising, image correction, and Canny algorithm segmentation are required for image preprocessing. The image sharpening of the Laplace operator of the part gives the final segmented part segmentation image. Later, the subpixel-level image segmentation of the calibration board and the parts needs to extract the contour edge of the mark point, improve the precision of the center distance measurement, and the parts need to measure the precision to obtain the pixel circumference of each area of the part. Finally, the true length of each part is measured.

For Question 3, model the two subpixel contour edge data provided in the attachment, analyze the edge contour curve data to fit the straight segment, arc segment or ellipse arc segment, find the segmentation of Edges through Halcon, get the outline

of the line, circle and ellipse according to the read subpixel coordinate data, and finally the center, start position and termination position of the contour everywhere. First, perform contour segmentation on the paper, then perform the Ramer algorithm used in contour fitting, and perform edge contour data extraction.
Key words: Subpixel Detection; Image Segmentation; Ramer Algorithm; Part Measurement; Contour Fitting

Content

1 Introduction

1.1 Question Background

As technology and science grows by leaps and bounds, the demand for intensive reading of work piece parts measurement continues to improve, and the intensive reading requirements of measuring instruments are becoming more and more demanding. In this paper, the edge analysis of the digital image size calculates the image edge profile curve under the subpixel processing.

1.2 Question Restatement

Question 1: For the subpixel edge extraction method and process analysis of 0.1 pixel accuracy or above, the subpixel edge contour boundary of the three images of Annex 1 is extracted and the subpixel point data is converted into the orderly edge contour curve data. The interference effect of the edge hair margin and the shadow part should need to be taken into account. Of the three images, Pic1_3 images taken in complex environments have more interference information. The problem requires extracting the three image edge profiles and drawing different colors, calculating the number of total edge profiles, the total edge contour length of the image coordinate space, the number of points and length of the contour curve, and determining the coordinates of each contour point. Finally, the number of points and points and length data on each curve are calculated.

Question 2: Measured images need to be standard with a point matrix calibration version. On the calibration version, the point diameter is 1mm and the distance between the center of the two points is 2mm.The attachment provided attachment contains three calibration map images taken at different angles and a product image. The length of the total edge profile was calculated by building the model with actual measurements of the edge of the product image profile.

Question 3: Two subpixel contour edge data are provided, and the mathematical model is analyzed and discussed. It is also necessary for the length of each sub-pixel contour edge of the part image segmentation to be calculated. The blue curve starts with the blue number 1 label and outputs the model calculated result data along the direction of the arrow. The green curve starts with the green number 1 label and outputs the model calculated result data along the direction of the arrow. The parameters of the segment segment are written to the table as shown in Annex 3.

2 Model Hypothesis

·Suppose that when extracting the image location point, the point where the most line intersects is the unique center point of the image;
·Suppose that the calibration board height is consistent with the part shooting height.

3 Symbol Description

Table 1 symbol description

Symbol	Explain
ω_k	The proportion of pixels points to the entire image
μ_0	Average gray scale
g	Interclass variance

σ	Standard deviation of the Gaussian distribution
I_i	Pixel distance
Z_i	actual distance

4 Problem Analysis

4.1 For Question 1

To draw the sub-pixel edge contour, first extract the edge contour, and transform the edge subpixel data into the curve data of the orderly edge profile, taking into account how the interference effect of the edge margin and shadow parts can be eliminated. Two images, Pic1_1 and Pic1_2, were processed in the same way, sharpening the images using the Laplace operator, the image contour smoothed, followed by thresholding for OTSU binarization and Gaussian filtering for denoising, and finally coarse segmentation with the Canny algorithm. Processing Pic1_3 images were taken under conditions that are environmentally complex. The PSO-OTSU algorithm was used to reduce the highlight of the image, and finally to detect the part defects and contour repair, to obtain the ideal binarized image for contour inspection..1Then the subpixel edge detection is performed to determine whether the image has reached 0 pixels, and finally the image edge profile is segmented with the Ramer algorithm. Observe Figure 1 for the overall flow chart.

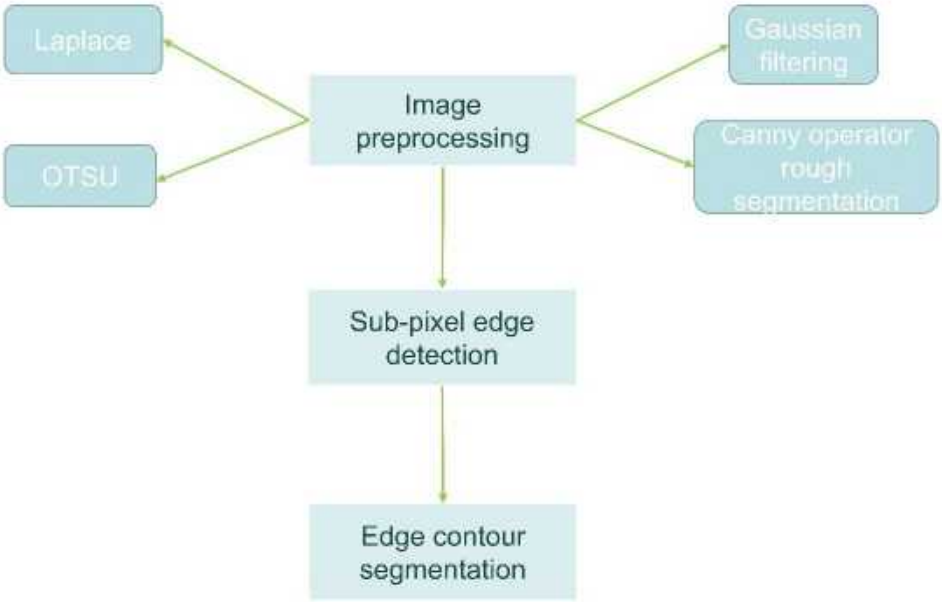


Figure 1 Question 1 flowchart

4.2 For Question 2

When measuring the true length of the given component picture, measurements need to be made according to the calibration version of the given level. The true length

of the provided product image is measured according to the calibration board images of three different angles provided. Image processing requires image preprocessing with OTSU binarization segmentation, Gaussian filtering denoising, image correction, and Canny algorithm segmentation. The image sharpening of the Laplace operator of the part gives the final segmented part segmentation image. Later, the subpixel-level image segmentation of the calibration board and the parts needs to extract the contour edge of the mark point, improve the precision of the center distance measurement, and the parts need to measure the precision to obtain the pixel circumference of each area of the part. Finally, the true length of each part is measured. Observe Figure 2 for the modeling flow chart.

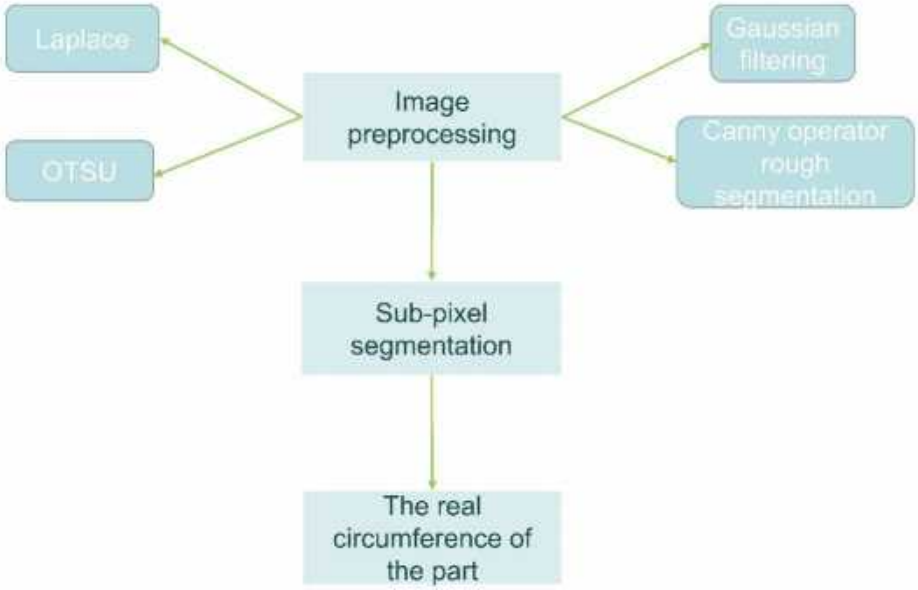


Figure 2 Question 2 flowchart

4.3 For Question 3

Through the two subpixel contour edge data provided in the attachment, model the edge profile segmentation, analyze the edge profile curve data to fit the straight segment, arc segment or ellipse arc segment, find the Edges through Halcon, get the line, circle and ellipse contour according to the read subpixel coordinate data, and finally the center, start position and termination position of the contour everywhere. First, the Ramer algorithm used for contour fitting and data extraction of edge profile. Observe Figure 3 for the flow chart.

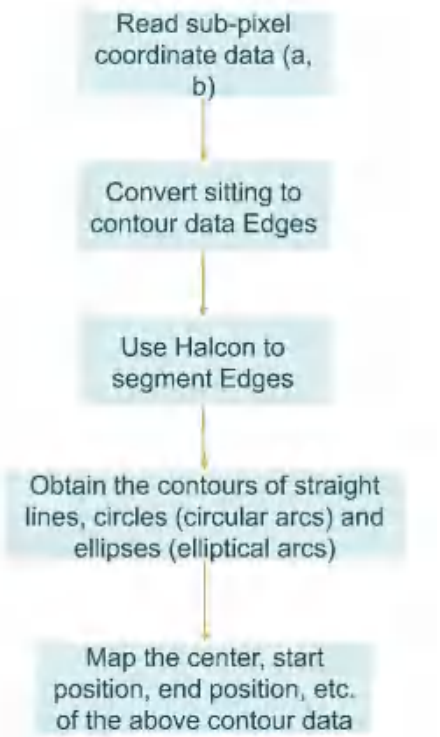


Figure 3 Question 3 flowchart

5. Model Establishment

5.1 Pic1_1 And Pic1_2 Pre-processing

During extracting three image edge profiles, images have noisy image edge profiles, which usually requires edge sharpening and shadow filtering denoising of the image. In this paper, the second derivative of the Laplace operator is used for image sharpening and curvature Gaussian filter denoising. Observe Figure 4 and 5 for the comparison and original images.

Figure 4 Original picture

Figure 5 The Laplace sharpening image

The Laplacian operator is able to enhance the edges and details of the image and improve the contrast. According to the degree of change of the image pixels, when the gray degree of the center pixel in the center pixel of the center pixel, the gray degree of

the center pixel shall be further improved to realize the image sharpening process. When sharpening, it is necessary to derive the relationship between second order differential and pixel, first to measure the first order partial differential method, first order partial differential observe equation (1).

$$\nabla f = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} = 2f(x,y) - f(x-1,y) - f(x,y-1) \tag{1}$$

Second-order differential methods can determine the position of the edges, with expressions shown in (2).

$$\nabla^2 f = 4f(x,y) - f(x-1,y) - f(x,y+1) - f\ （x+1,y）\ - f(x,y-1) \tag{2}$$

This makes the white dots brighter based on the template matrix. Areas where the grayscale jump can be solved, thus making the image clearer.[1]

The contour smoothing operator is the smooth_contours_xld() of the machine vision library. It is often used to calculate distances between profiles. Observe Figure 6 and 7 for contour smoothing operator and contour extraction effects.

Figure 6 Profile smoothing operator processing

Figure 7 Profile extraction rendering diagram

In image processing, the OTSU method is commonly used for image segmentation-based clustering. The theoretical basis of the algorithm is that the image contains two classes of pixels (foreground and background pixels), the histogram is a bimodal histogram, and then calculates the optimal threshold (intra-class variance) that the two classes can be separated, or the equivalent inter-class variance.

For each coordinate point of the image, the segmentation threshold of the foreground and background is T recorded, the proportion of pixels to the whole image is ω_0, the average gray degree is μ_0; the proportion of background pixels to the whole image is ω_1, the average gray degree is μ_1; the average gray scale of the whole image isμ, and the variance is$g$. Assuming that the image size is M*N, the number of pixels in the pixels of the image and the number of pixels less than the threshold is T, the histogram is N_0 obtained from a low valuation between the two peaks. OTSU results for image comparison are shown in Figure 8.

Figure 8 OTSU binarization thresholding segmentation

Denoising was performed using Gaussian filtering. Gaussian filtering is a linear smoothing filter with a good inhibition of noise obeying a normal distribution. Assuming that the noise for this figure is Gaussian white noise, in the pre-processing part of practical application, Gaussian filtering is usually used.

The image was denoised with Gaussian filtering only after the sharpened image and profile were smoothed with the OTSU binarized segmentation process. Gaussian filtering is convolutional using a mask and image, and the template coefficient of the Gaussian filter decreases with the increasing distance from the center of the template. A Gaussian filter maintains the overall details of the image, with the distribution expression shown in (3).

$$f\left(x,y\right)=\frac{1}{\left(\sqrt{2\pi}\,\sigma\right)^{2}}e^{-\left((x-ux)^{2}+(y-uy)^{2}\right)/2\sigma^{2}} \tag{3}$$

σCreating a Gaussian filter template where x, y is the coordinate of the image, is the standard deviation of the Gaussian distribution. Its size represents the central coefficient of the generated template. Sampling was performed at the center of the template as the coordinate origin. The coordinates of each position are also brought into the Gaussian function to obtain the coefficient of the filter. Assuming that the size of the resulting window template is (2k + 1) (2k-1), then obtain the calculation formula for each element of the template, observe formula (4).

$$H_{i,j}=\frac{1}{2\pi\sigma^{2}}e^{-\frac{(i-k-1)^{2}+(j-k-1)^{2}}{2\sigma^{2}}} \tag{4}$$

The calculated results will be in two forms: decimal and integer. Where the form of the decimal number does not need to be processed, the integer form needs to be normalized. After Gaussian filtering processing, the two plots are compared.[2]

Figure 9 Gaussian filter effect map comparison

Figure 10 Gaussian filter renderings

Comparing Figure 9 and 10, the profile of the image is clearer after Gaussian filtering processing.

Later, the pixel-level profile of the image is extracted by the Canny operator coarse segmentation for subsequent subpixel edge detection. Effective suppression noise is required to satisfying the coarse segmentation of the Canny operator, smoothing the image using a Gaussian operator and with precise determination of the edge position. After image smoothing and precise edge position respectively using Gaussian filter and OTSU, the time of image edge detection conditions are met, after a three-step processing, first use Gaussian smoothing function to eliminate noise, then use first-order differential convolution template for edge enhancement, and finally use non-maximum inhibition to retain the maximum in the gradient direction. The coarse segmented images are shown in Figure 11.

Figure 11 The Canny algorithm for coarse segmentation

Edge enhancement simultaneously obtains both the magnitude of the gradient amplitude and the direction angle, showing differences in the X and Y direction, performing first-order differential convolution operations to obtain the image where the features are enhanced at the edge position. When making a global gradient, it is not enough to determine the edge, to determine the point with the largest local gradient.

5.2 Pic1_3 Image Pre-processing

- Highlight processing of metal artifacts based on the PSO-OTSU algorithm

The area of high reflectivity metal working surface due to the uneven light

distribution, which then affects the visual measurement accuracy. Therefore, we choose the PSO-OTSU algorithm to treat metal artifacts with highlights. First, the highlight regions were extracted using the PSO-OTSU algorithm, secondly using the histogram specification method, and finally fused the processed highlight regions with the original or non-highlight regions to form the de-highlight image.

• The highlight regions were extracted by the PSO-OTSU algorithm

Figure 12 High photosensitive processing process

The PSO-OTSU algorithm is calculated by selecting the fitness function in the P SO algorithm as the maximum inter-class variance criterion. In the process of finding the optimal solution, the particles change the current speed and threshold are based on their own inter-class variance maximum and the global inter-class variance maximum. Until the particles reach the optimal threshold. Observe the iterative algorithm in formula (5).

$$\begin{aligned} V_i^{k+1} &= wV_i^k + c_1 r_1 (pbest_i^k - x_i^k) + c_2 r_2 (gbest_i^k - x_i^k) \\ X_i^{k+1} &= X_i^k + V_i^{k+1} \end{aligned} \tag{5}$$

Where, w for the inertial weight,c_1, c_2 for the acceleration constant, r_1 and $r_2$is the random number.

In order to be able to balance the local c_1、c_2 and w global search capabilities, we need to be linearly dynamically adjusted, observe formula for (6), formula for (7).

$$c_1 = 1 + \frac{iter_{now} + 1}{iter_{\max}} \tag{6}$$

$$\begin{aligned} c_2 &= 2 - \frac{iter_{now} + 1}{iter_{\max}(w_{\max} - w_{\min})} \\ w(t) &= w_{\max} - \frac{iter_{now}}{iter_{\max}}(w_{\max} - w_{\min}) \end{aligned} \tag{7}$$

Among them, $iter_{now}$ for the current number of iterations and the maximum number of iterations, the initial value of our parameter is $iter_{\max}$ set according to the standard particle group algorithm, respectively: $w_{\min} = 0.4, w_{\max} = 0.9$, $c_1 = c_2 = 2, iter_{\max} = 20$ learning factor. The resulting de-highlighted images are shown

in Figure 13.

Figure 13 Original picture

Figure 14 After removing the highlights

Figure 15 after binarization using OTSU Oxford method.

Figure 15 Binarization of the OTSU method

• High of area areas based on histogram specification

The implementation of the histogram specification method requires the balanced processing of the original and desired images, and the formula (8),(9),(10).

$$s = \int_0^r P_r(r)\,dr \tag{8}$$

$$v = \int_0^z P_z(z)\,dz \tag{9}$$

$$z = G^{-1}(v) \tag{10}$$

Among them, $P_r(r)$ is the gray degree probability density function of the original image, $P_z(z)$ the gray degree probability density function of the desired image, s the gray level of the original image, v the gray level of the fixed image, and z the gray level of the desired image.

With the above processing, the transformed original image gray scale level can be used to replace the gray scale level of s a given image v .The above inverse transformation analytics is suitable for continuous images. For the discrete images, the prescribed expression is type (11),(12),(13).

$$P_Z(z_k) = \frac{n_k}{n}, 0 \leqslant z_k \leqslant 1 \tag{11}$$

$$v_k = G(z_k) = \sum_{k=0}^{m-1} P_z(z_k) \tag{12}$$

$$z_k = G^{-1}(s_k) \tag{13}$$

Among them, n is the total number of pixels in the image, n_k is the number of k gray scale appearance, m is the number of gray scale level, $P_z(z_k)$ is the probability of k gray scale level.

In conclusion, the histogram regulation method is to reduce the gray degree value of the highlight area by adjusting the gray degree histogram of the highlight area to a predetermined shape.

- Fusion algorithm

A fusion algorithm based on non-highlight region images, which is used as follows:

(1) Find the pixel points of the non-highlight area image is all 0 from the predetermined images and the histograms, and obtain the coordinates of the pixel points; (x_i, y_j)

(2) Find the corresponding pixel point according to the coordinate value, and replace it with the pixel point in the non-highlight area;

(3) The completed replacement image is the final highlight-removed image.

There are multiple scratch noises in the Pic1_3 picture, and the color is similar to the background color, thus leaving blank areas in the components after the binary operation, causing interference with the profile detection. After defect detection and repair processing, an ideal binary image can be obtained for subsequent contour detection. The image were detected by machine vision. The detection of surface defects using deep learning is an end-to-end scheme featuring the convolutional neural network. Its feature extraction ability allows the network to automatically learn the image features of the defect, to understand the defect target from local to global, distributed in different layers of the network, eventually forming the overall perception of the defect target. The advantage is that it more accurately expresses and understands various types of defects than artificial feature extraction, and detects higher precision reading. Its main key lies in the detection algorithm. The temporal image segmentation is used to determine the object type to which each pixel position of the image belongs. Observe Figure 16 for the effect and original drawings.

Figure 16 A metal scratch detection image

Blue is short scratch, red is long scratch. After the scratch position of the image is detected, the marked blue-red mark is replaced with black for convenient subsequent step processing to facilitate the extraction of the subsequent profile. The repaired images are shown in Figure 17.

Figure 17 Metal scratch repair image

5.2.1 Product Measurement And Preprocessing

Length measurement of the product requires an image segmentation, separating the part area from the background area. Image segmentation is performed by the threshold segmentation method, and the pixel gray scale amplitude of the product image is divided to divide different grades, and parts regions are extracted by setting the gray scale threshold. Observe (14).

$$g\ （i,j）\ = \begin{cases} 1 & f(i,j) \geqslant T \\ 0 & f(i,j) < T \end{cases} \tag{14}$$

$g\ （x）\ T$Set as the binary image after the threshold operation; the threshold determined by the gray scale histogram of the image; 1 is the value of the target image; 0 is the value of the background image.

Appropriate thresholds were determined using gray-scale histogram tools in H ALCON. The post-filtered binarized image parts were segmented by the threshold algorithm. According to HALCON, the gray value of the part image is between 0 and 114, and the background image is greater than 114.The operator connection is then divided into different connected domains, and part regions are extracted using area features.[3]

After part binarization processing, the image is denoised. Call open_framegrabber

() with the grab_image_start () operator in the HALCON software. The rgb_to_gray () operator was called to grayscale the image. During image acquisition and transmission, it is not conducive to image processing and analysis, using the median filter operator median_image () and Gaussian filter operators to eliminate pretzel and Gaussian noise by gauss_ image ().

5.2.2 Preprocessing Of The Image Determination

The image determination first requires the calibration board, whose main process is shown in Figure 18.

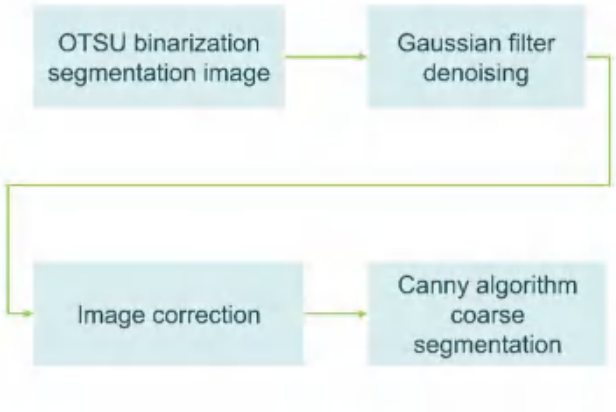


Figure 18 Main process for the calibration board

First considering the OTSU binarization treatment, the basic principle is the same as the first question, except here is to split the circular marker point and the background, used to calculate the pixel length of the center distance of the extracted dot. Comparison of the resulting OTSU binarized images with the original image is shown in Pic2_1 Figure 19 and Figure 20.[4]

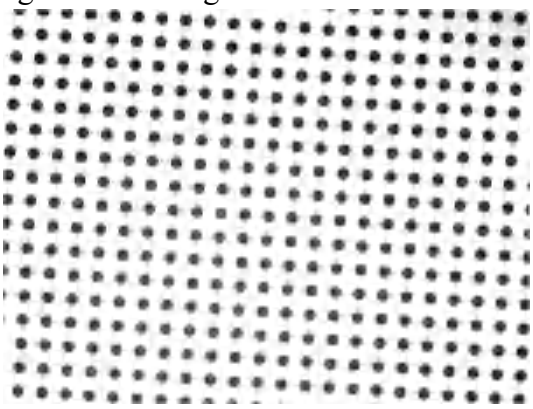


Figure 19 Original picture

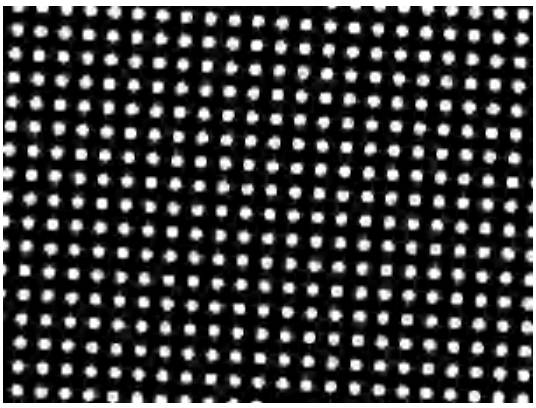


Figure 20 The OTSU binarization processing

The remaining two plots were also treated in the same form.

Afterwards, Gaussian filtering was used to denoise it. Denoising works the same as above, using a Gaussian filter template that then removes the noise on the image. Figure 19 , taken at a non-horizontal angle. Before the image segmentation, the image correction is needed to ensure the precision of the measurement and the collected three scalar images to the horizontal position.

The method of image correction employed is the image area correction. In order to perform the parts work quickly and efficiently and ensure the measurement accuracy, the collected part image needs to be affine and transformed to the horizontal position. The expression for the affine changes is (15).

$$\begin{bmatrix} u \\ v \end{bmatrix} = A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_2 & a_1 & a_0 \\ b_2 & b_1 & b_0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{15}$$

Among them, A for the affine transformation matrix, the linear parts and the translation parts composition.

Affine changes mainly include translation transformation and rotation transformation. The translation transformation is to move all the pixels of an image horizontally and vertically at the required offset; while the rotation transformation rotates an image at an angle around a certain point. The expressions for the two are, respectively.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{16}$$

$$\begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{17}$$

Where, respectively, x_0, y_0 is the coordinate translation quantity; θ is the rotation angle.

In HALCON, affine transformation: orientation_region, vector_angle_to_rigid, affine_trans_image. The orientation_region operator is first called to calculate the image deflection angle; then use the vector_angle_to_rigid operator; finally, convert the part image to a horizontal position using the affine_trans_image operator, and the positive Pic2_1 image is shown in Figure 21.

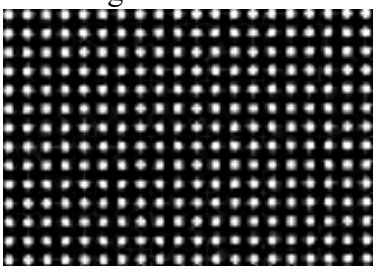


Figure 21 Post-corrected renderings

The corrected image is the same as that taken at a horizontal angle, which can be used as a basis for measuring the part pixels. Finally, the Canny algorithm is coarse divided with the same principle as in the first question. The resulting Pic2_1 images of the Canny algorithm are coarse-segmented in Figure22.

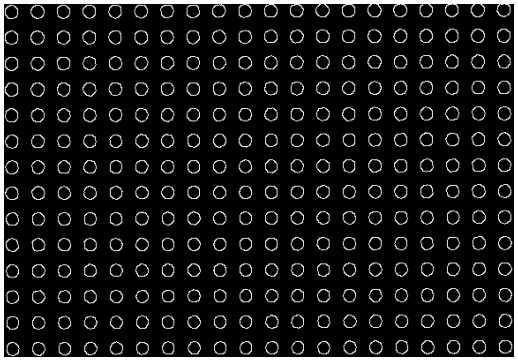


Figure 22 Canny algorithm

6 Problem Solving

6.1 For Question 1

For Pic1_1 and Pic1_2 processing, unlike Pic1_3, Pic1_3 had more noise causing image profile extraction, so the first two plots were processed together. The outline of the first two plots can detect the edge well after preprocessing, and Pic1 _1 and Pic1_2 detect the edge through the Canny subpixels in only four steps:

(1) Smooth the input image with a Gaussian filter;

(2) Calculate the gradient amplitude image and the direction angle image;

(3) Apply non-maximum suppression to the gradient amplitude images;

(4) Detection and connect edges with double threshold processing and connection analysis.

First, to effectively smooth the image and reduce the identification of pseudo-edges, use Gaussian filtering to suppress noise, in the pre-processing has been conducted, a suitable Gaussian filtering standard deviation is important for the effect of edge detection. The parameters of Canny edge detection can be calculated at only the width and height of the edge to the bottom. The Sobel algorithm was selected as the gradient algorithm for the edge detection and extraction process. The finite difference of the first-order bias guide is used to calculate the value of the amplitude of the gradient.

The gradient amplitude is (18).

$$|gradg\ （x,y）| = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2} \tag{18}$$

Do the Sobel algorithm on the image, and each pixel in the image represents the gradient of this point. With the help of Sobel, image preprocessing and segmentation by Canny, and the image are shown in Figure 23 and Figure 24.

Figure 23 Pic1_1 Contour

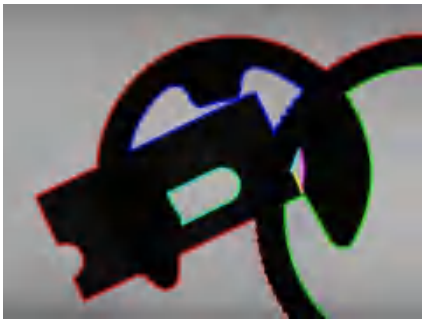
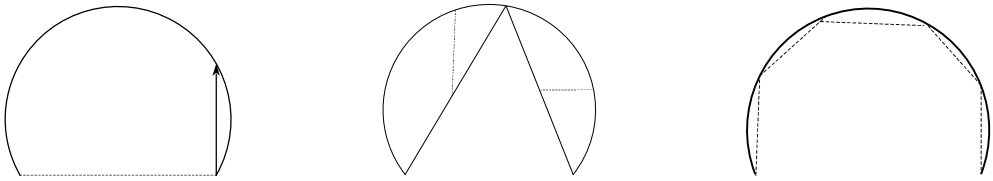


Figure 24 Pic1_2 Countour

Figure 23 and 24 can clearly observe the contour shape of the two images, after which the edge profile is cut.

Edge profile cutting requires the Ramer algorithm. In fact, recursive segmentation is analyzed in Table 2.

Table 2 for a recursive subdivision



Near polygons are represented by thick lines and then the edges are correctly split apart with the Ramer algorithm for Ramer: the algorithm sometimes places the polygon control point to deviate slightly from the true corner. For this question, these deviations do not pose a problem, due to the maximum accuracy that requires fitting the lines to these contour segments with robust algorithms. Observe Figure 25, Figure 26.

Figure 25 Profile edge segmentation renderings

Figure 26 Profile edge segmentation renderings

For Pic1_3, Pic1_3 image were taken under relatively complex light conditions and with more interference information. Therefore, in the first step, we need to preprocess it, use the Laplace operators and contour smooth operators to complete the image sharpening, remove the edge burr and smooth the contour, to avoid the interference effects caused by the edge hairy edges and shaded parts. Then the PSO-OTSU algorithm was used to remove highlight. Since the three pictures were taken under relatively complex light conditions on the parts, the color of the reflective area is similar to the background color, which will affect the subsequent OTSU binary image

segmentation, so the PSO-OTSU algorithm was used to remove highlight and reduce the effect. Then the OTSU binarization threshold segmentation is used to separate the image foreground from the background, filtering out the background image to easily extract the outline. Gaussian filter is used to remove the noise. For the scratches on the parts, the surface defect detection algorithm and marker feature contour repair are used to obtain the ideal binary image for subsequent contour detection. Observe the resulting contour edge distribution images in Figure 27 and 28.[5]

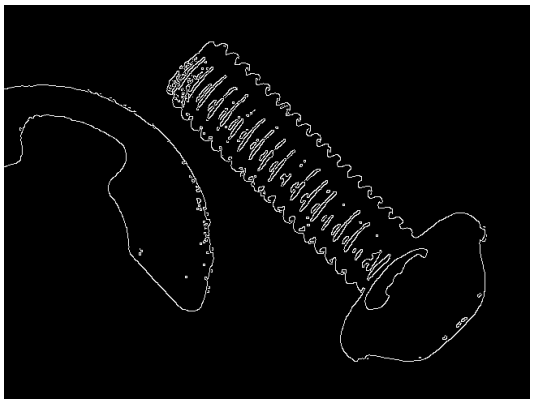


Figure 27 The Canny rough segmentation effect

Figure 28 Edge detection effect

The pixel-level profile of the image was extracted by the Canny operator edge detection algorithm for subsequent subpixel edge detection. The second step regarding subpixel edge detection, we detect edge information using a Gaussian fitting-based subpixel edge detection algorithm. Finally, we use the variable weight least squares based image edge feature extraction algorithm for profile segmentation, observe Figure 29.

Figure 29 Profile cut effect

Finally, the total number of contour curves for Pic1_1 and the number of points and length data on each curve are shown in Table 3.

Table 3 Total number of contour curves for Pic1_2 and the number of points and length on each curve

Total Edge Contours Count:		3610
Total Edge Contours Length:		3183.21
Edge Contour 1	Length	PointCount
	1617.99	1854
Edge Contour 2	Length	PointCount
	55.86	64
Edge Contour 3	Length	PointCount
	154.47	148
Edge Contour 4	Length	PointCount
	136.56	135
Edge Contour 5	Length	PointCount
	60.16	77
Edge Contour 6	Length	PointCount
	268.77	338
Edge Contour 7	Length	PointCount
	61.83	71
Edge Contour 8	Length	PointCount
	109.34	115
Edge Contour 9	Length	PointCount
	51.56	61
Edge Contour 10	Length	PointCount
	259.7	328
Edge Contour 11	Length	PointCount
	69.5	81
Edge Contour 12	Length	PointCount
	138.31	146
Edge Contour 13	Length	PointCount
	151.36	149
Edge Contour 14	Length	PointCount
	47.79	56

The number of edge total contour curves for Pic1_2 and the number of points and length data on each curve are shown in Table 4.

Table4 Total number of contour curves for Pic1_2 and the number of points and length on each curve

Total edge profile count	7970	
Total edge contour length	6763.46	
Edge Contour 1	Length 346.82	PointCount 403
Edge Contour 2	Length 955.34	PointCount 1099
Edge Contour 3	Length 205.74	PointCount 266
Edge Contour 4	Length 163.64	PointCount 200
Edge Contour 5	Length 173.07	PointCount 205
Edge Contour 6	Length 79.56	PointCount 100
Edge Contour 7	Length 188.74	PointCount 247
Edge Contour 8	Length 81.59	PointCount 114
Edge Contour 9	Length 69.64	PointCount 82
Edge Contour 10	Length 76.38	PointCount 78
Edge Contour 11	Length 249.16	PointCount 317
Edge Contour 12	Length 251.56	PointCount 280
Edge Contour 13	Length 497.51	PointCount 530
Edge Contour 14	Length 62.3	PointCount 72
Edge Contour 15	Length 170.78	PointCount 226
Edge Contour 16	Length 67.97	PointCount 83
Edge Contour 17	Length 385.82	PointCount 441
Edge Contour 18	Length 142.11	PointCount 177

Edge Contour 19	Length	PointCount
	422.99	534
Edge Contour 20	Length	PointCount
	273.61	333
Edge Contour 21	Length	PointCount
	64.29	88
Edge Contour 22	Length	PointCount
	139.93	144
Edge Contour 23	Length	PointCount
	77.39	83
Edge Contour 24	Length	PointCount
	208.56	252
Edge Contour 25	Length	PointCount
	169.82	200
Edge Contour 26	Length	PointCount
	78.7	77
Edge Contour 27	Length	PointCount
	137.29	136
Edge Contour 28	Length	PointCount
	73.74	85
Edge Contour 29	Length	PointCount
	349.37	399
Edge Contour 30	Length	PointCount
	107.77	132
Edge Contour 31	Length	PointCount
	172.65	215
Edge Contour 32	Length	PointCount
	136.63	159
Edge Contour 33	Length	PointCount
	182.99	238

The representations of the total outline technique and length are shown in Table 5.

Table5 Total edge profile counts and length of Pic1_3

Total edge profile count	194
Total edge contour length	6870.37

6.2 For Problem 2

Subpixel segmentation of the image is required before making the image measurements. Subpixel segmentation is mainly to extract contour edges of marker points and improve the precision of center distance measurements. The algorithm used is for the calibration board to improve the measurement accuracy of the image of each area of each part. The process of extracting the edge and line characteristics of the large spatial direction of the gray value is edge detection. In this paper, a gradient is used to

reflect the edge position of the image, and the gradient of $f(x,y)$ a continuous image function is expressed as (19).[6]

$$\nabla f(x,y) = [G_x \quad G_y]^T = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}\right]^T \tag{19}$$

Part image edges were extracted by using the edge detection of the Canny operator. The basic principle is to first select a Gaussian filter to smoothly filter the image, and then use the non-extreme suppression technology to obtain the image edge. The steps are as follows.

First: smooth the image by using $f_s(x,y)$ a circular 2D Gaussian filter to smooth $G(x,y)\, f(x,y)$.

$$G(x,y) = e^{\frac{x^2+y^2}{2\sigma^2}} \tag{20}$$

$$f_s(x,y) = G(x,y) \cdot f(x,y) \tag{21}$$

Second: calculate the gradient amplitude $M(x,y)$ and direction of $\alpha(x,y)$ each pixel of the smooth image, and the edge emphasis and direction of each point are calculated using the gradient amplitude and direction.

$$M(x,y) = \sqrt{(g_x)^2 + (g_y)^2} \tag{22}$$

$$\alpha(x,y) = \arctan\left(\frac{g_x}{g_y}\right) \tag{23}$$

Third: perform the non-maximum suppression of the gradient amplitude to determine the point with the largest local gradient and the inhibition non-maximum, the non-local maximum point is set to 0 to refine the edge.

Fourth: set the low and high threshold to detect the edges T_1 and T_2 the connected edges by comparing the edge pixels of the image with the set high and low threshold. The images obtained from segmentation according to the Canny algorithm are shown in Figure 30.

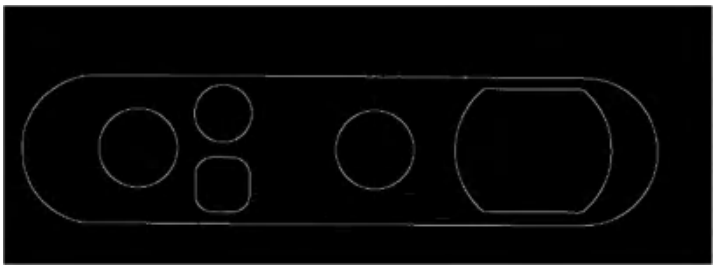


Figure 30 The Canny algorithm for coarse segmentation

Following the above principles, we extracted the subpixel edges of the part images.

The resulting image edges are shown in Figure 31.

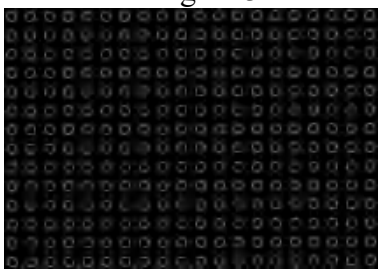


Figure 31 Subpixel edge detection rendering

After subpixel edge detection, image edges were performed. Image edge extraction included linear fitting and circle fitting. Where the line fitting is performed by calling the select_shape_xld()operator in HALCON, using the characteristic histogram of XLD to remove the aperture edge using the width, retaining only the outer outline of the part. The segment_contours_xld() operator is then called for segmentation, and the segmented profile is determined by the global profile attribute cont_approx.

After the part profile segmentation is completed, calling the get_contour_global_attrib_xld() operator obtains the properties of each profile. When Attrib = -1 belongs to the line segment, the fit_line_contour_xld () can fit the boundary value of the image to obtain the actual size of the part according to the conversion relationship.

Sometimes the aperture of the shot part can not obtain the complete circular outline during the edge extraction, which may be due to shooting instability or field environment, so the arc, which is not conducive to the fitting of the circle, so we combined the arc and call the union_cocircular_contours_xld() operator to combine the arc.

In this paper, the circle parameter is fitted by least squares, so fitting the circle does not affect even the edge localization and detection. The expressions of the circle equation is (24).

$$(x-x_0)^2+(y-y_0)^2=r^2 \tag{24}$$

The minimum sum of square of error of the objective function is (25).

$$S=\sum_{i=1}^{n}\left[\sqrt{(x_i-x_0)^2+(y_i-y_0)^2}-r\right]^2 \tag{25}$$

These are the coordinates of $(x_i,y_i),i=1,2,3$ the feature points on the arc n in the image coordinate system and the number of feature points involved in the fitting.

To obtain a direct solution to a minimization problem, while avoiding the square root, the objective function can be overwritten as (26).

$$E=\sum_{i=1}^{n}\left[(x_i-x_0)^2+(y_i-y_0)^2\right]^2 \tag{26}$$

To sort out the top formula (27).

$$E=\sum_{i=1}^{n}\left(x_i^2-2x_0x_i+x_0^2+y_i-2y_0y_i+y_0^2-r^2\right)^2 \tag{27}$$

a surname $B=-2y_0,\ A=-2x_0,\ C=x_0^2+y_0^2-r^2$

The upper formula can be expressed as (28).

$$E=\sum_{i=0}^{n}\left(x_i^2+y_i^2+Ax_i+By_i+C\right)^2 \tag{28}$$

According to the principle of E least squares, the conditions to be met to A,B,C solve the minimum are (29),(30),(31).

$$\frac{\partial E}{\partial A}=2\sum_{i=0}^{n}\left(x_i^2+y_i^2+Ax_i+By_i+C\right)x_i=0 \tag{29}$$

$$\frac{\partial E}{\partial B}=2\sum_{i=0}^{n}\left(x_i^2+y_i^2+Ax_i+By_i+C\right)y_i=0 \tag{30}$$

$$\frac{\partial E}{\partial C}=2\sum_{i=0}^{n}\left(x_i^2+y_i^2+Ax_i+By_i+C\right)=0 \tag{31}$$

Solving the above three formulas, the fitting value of the center coordinate (x_0,y_0) and r radius is (32).

$$x_0=-\frac{A}{2},y_0=-\frac{B}{2},r=\frac{1}{2}\sqrt{A^2+B^2-4C} \tag{32}$$

For curve fitting and circle fitting processing, the components can be detected by subpixel edges. The resulting component effect 31 is compared with the original Figure 32.

Figure 32 Original picture

Figure 33 Subpixel edge detection effect diagram of the components

This is then the need to measure the length of the part. First, we can find the pixel distance. Then according to the (33).[9]

$$\frac{I_i}{I_k} = \frac{Z_i}{Z_k} \tag{33}$$

The (33) I_i represents the pixel value corresponding to the two centers of Z_i the calibration plate, I_k the real distance of the calibration plate, $Z_k$and the real distance of the calibration plate.

The real length represented is Z_i known to be 2mm, given the pixel values and the mean values of the three images.

Table 6 Calibration plot center distance

order number	Pic2_1	Pic2_2	Pic2_3	average
The center distance	58.8742	58.2367	59.3419	58.8176

The same algorithm is adopted to measure the contour pixels of parts and the data are shown in Table 7.

Table 7 Contour pixel length

order number	1	2	3	4	5	6
length	4532	436.349	321.472	329.124	437.874	834.231

Enter Type (33) for the true length value of the outline, observe Table 8.

Table 8 Edge contour length output format (mm)

Contour ID	Length(mm)
Total Edge Contours	234.3193
Edge Contour 1	154.1035
Edge Contour 2	14.8374
Edge Contour 3	10.9311
Edge Contour 4	11.1913
Edge Contour 5	14.8892
Edge Contour 6	28.3667

The desired length is the true length of the required profile.

6.3 For Question 3

6.3.1 Allocation Method Of Power At The Inflection Point

Using the general weight assignment method, some non-near points will have a large weight at the inflection point. Considering the geometric characteristics of the closed discrete point, this paper uses the curve fitting method to calculate the weights at the inflection point.

x_{i+1} and x $x_1 = x$ definition between two adjacent x_i discrete points (x_1, y_1), $w(s(x_i))$ takes a point on two lines, so that the weights of each point in the support domain are determined by the distance between the point to the point, and type(34).

$$s(x_i)=\frac{\sqrt{(x_1-x_i)^2+(y_1-y_i)^2}}{k\max\left(\sqrt{(x_1-x_j)^2+(y_1-y_j)^2}\right)} \tag{34}$$

Among$i,j=1,2,\cdots,N,k>1$.

The distance from discrete to point (x_1,y_1) within the support domain determines the radius of the support domain, and reasonable selection values can determine that the weights of $w(s(x_i))$ each point in the support domain are in the decay region of the weight function, and control the degree of local approximation of the curve. The assignment of such weights is x related to the geometric characteristics of the discrete point at the corner, supporting the gradual decay along the two directions of the domain, and ensuring the approximation nature of the fitting curve at the corner.

In these two cases, the calculation formulas are (35),(36).

$$y_1=\frac{x_1-x_i}{x_{i+1}-x_i}(y_{i+1}-y_i)+y_i \tag{35}$$

$$y_1=\frac{x_1-x_{i+1}}{x_i-x_{i+1}}(y_i-y_{i+1})+y_{i+1} \tag{36}$$

7 Test Of The Model

7.1 Model Checking Of The First Question

This article compares the five sub-pixel edge detection algorithms of Gaussian fitting, spatial matrix, Zernike moment method, OFMM method, and PZM method to detect the positioning accuracy and the running time of the detection, observe the Table 9.

Table 9 The detection results of several sub-pixel edge detection algorithms for step-shaped edges

	Gaussian fitting	Space matrix method	Zernike matrix	OFMM	PZM method
Positioning accuracy/pixel	0.084	0.141	0.115	0.102	0.187
Detection time/ms	103	142	1.5	130	109

It can be seen from the table that the positioning accuracy of the sub-pixel edge detection algorithm based on Gaussian fitting is higher than that of the sub-pixel edge detection algorithm based on spatial matrix, Zernike moment method, OFMM method, and PZM method. And the detection based on the sub-pixel edge detection algorithm takes the shortest running time, so its efficiency is the highest.[10]

7.2 Model Checking Of The Second Question

The starting and ending point coordinates of the part boundary are obtained by straight line fitting, the image_points_to_world_plane() operator is adjusted in HALCON to convert the image coordinates of the starting and ending points of the straight line into world coordinates, and the distance_pp() operator for calculating the distance between two points is called to obtain the actual side length of the part . Let x represent the actual size and p represent the distance of 1 pixel in the image. Through the obtained pixel coordinates and world coordinates of the circular mark of the HALCON calibration board, combined with the result of the camera calibration, the actual size corresponding to each pixel is 0.00011 mm. In order to obtain the actual size of the aperture r. Use high-precision measuring equipment to measure the actual size of the part, compare it with the visual measurement result and calculate the error.

Table 10 Operation time of 3 methods

	L_1	L_2	L_3	L_4	radius
Actual size /mm	30.012	74.987	29.989	75.011	5.000
measure size /mm	30.023	75.004	30.005	74.994	4.991
error /mm	-0.011	-0.017	-0.016	0.017	0.009

It can be seen from the table that the size of the parts measured by vision is very small from the actual size, and the error of each size is within ± 0.02 mm, which satisfies the acceptable error range, and the accuracy and real-time performance meet the inspection needs of industrial production.[11]

8 Extension And Evaluation Of The Model

8.1 Extension Of The Model

For the shortcomings of Sobel algorithm, this algorithm makes two improvements

based on the classical Sobel algorithm to compensate for the shortcomings of Sobel algorithm.

8.1.1 Direction Templates Are Set To 8

The edge of the image has many directions. In addition to the horizontal and vertical directions, there are other edge directions, such as 600, 1200, etc.

8.1.2 Use The Edge Tracking To Discharge The Noise Points

After the calculation of the 8-point orientation template and the determination of the threshold value, we detected the pixels with the new gray-scale values greater than or equal to the threshold value. However, it is not possible to determine that these pixels are the edge points, because the noise also causes a jump in the pixel gray degree, resulting in new gray values greater than or equal to the threshold value. Therefore, the corresponding method must be further taken to determine whether the pixel is an edge point or a noise point, and the edge tracking method is adopted to determine below.

8.1.3 Basic Ideas Of Edge Tracking

The edge has two features of direction and amplitude, perpendicular to the edge direction, and the pixel gray value changes dramatically, while the pixel gray value changes gently along the edge direction[8]. The edges of the object are continuously smooth, and along either edge point, another edge point with a small difference in gray value and direction can always be found. However, the noise is different. Because the noise is random, it is difficult to find another noise point with a small difference in the gray degree value and direction along the edge direction. Using this basic idea, the actual edge point can be distinguished from the noise point. In order to more accurately determine whether a pixel is an edge point, the pixels are generally tracked continuously along the edge direction.[12]

8.2 Evaluation Of The Model

8.2.1 Merit

1.Sobel usually has a larger absolute value for the elements of the output image (array);

2. The detection method has a good effect on the image processing with more gray scale gradient and more noise;

3. Can identify as many actual edges in the image as possible;

4. The identified edges should be as close as possible to the actual edges in the actual image;

5. Provide more accurate edge direction information.

8.2.2 Shortcoming

Disadvantages of Sobel algorithm Because the edge is a sign of position, they are not sensitive to the change of gray scale.

9 Reference

[1] Gu Tianqi, Zhang Lei, Ji Shijun, Tan Xiaodan, Hu Ming.Curve fitting method for closed discrete points [J].Journal of Jilin University (Engineering Edition), 2015,45(02):437-441.DOI:10.13229/j.cnki.jdxbgxb 201502015.

[2] Lin Haibo, Pan Wanui.Research and application of the image edge feature extraction algorithm [J].Combined machine tool and automatic processing technology, 2015(06):118-120.DOI:10.13462/j.cnki.mmtamt.2015.06.032.

[3] Han Dong, Li Yuqi, Wu Yanhui.Sub-pixel edge detection algorithm based on a Gaussian fitting [J].Computer applications and Software, 2018,35 (06): 210-213 + 229.

[4] Gao Maoyuan, Wang Haochen, Cong Zhiwen, Wang Zezheng, Li Jiapeng.Fast and precision measurement of mechanical part dimensions based on HALCON [J].Electromechanical Engineering technology, 2020,49 (03): 71-74 + 127.

[5] Tang Song, Yang Qihua, Liu Gang.Sub-pixel level of core size measurement based on Halcon [J].Computer Engineering and Application, 2017,53 (03): 237-241.

[6] Chen Yifeng. Application of Halcon on artifact 2 D size detection [J].Electromechanical technology, 2011,34 (04): 12-13 + 17.

[7] Cai Yan, Ye Lianxiang, Sun Dawei, Tian Hua, Zhu Junjie, Hu Zhiyun. Based on an improved Zernike moment subpixel edge extraction algorithm for external threaded non-contact measurements [J].Journal of Shanghai Jiao Tong University, 2014,48(10):1468-1472+1478.DOI:10.16183/j.cnki.jsjtu.2014.10.021.

[8] Lu Zhengqi, Han Yongxiang, Liu Yubao. Improving the Ramer scheme for diagnosis of freezing rain in China[J].Atmospheric Research,2021,254:

[9] Elworthy K.D..Infinite-Dimensional Degree Theory and Ramer'S Finite Co-Dimensional Differential Forms[J].The Quarterly Journal of Mathematics,2021,72(1-2):

[10] Chen Fangsheng, He Chunyu, Chen Yafei. Improving application of Canny algorithm in thickness edge detection [J].Steel rolling, 2021,38(05):81-85.DOI:10.13228/j.boyuan.issn 1003-9996.20210514.

[11] Yuan Weiqi, Zhao Peiyao. Adaptive heat exchanger plate edge detection method based on Canny [J].Computer Technology and Development, 2021,31 (11): 81-85 + 94.

[12] Fei Jiyu, Wang Zheng, Li Hua, Liu Zhibo, Jiang Wei, Liu Xiaodong, Zhang Junfei. Metal artifact highlight processing based on the PSO-OTSU algorithm [J].Journal of Dalian Jiaotong University, 2020,41(03):28-33.DOI:10.13291/j.cnki.djdxac.2020.03.006.

Team Number:	apmcm2106398
Problem Chosen:	B

2021 APMCM summary sheet

Optimal Design of Thermal Emitter in Thermophotovoltaic Technology

Abstract

Recent years witness the development of thermophotovoltaic technology. However, the thermal energy produced during the conversion process is not efficiently utilized. Therefore, it is required to regulate the emission spectrum of the thermal emitter to improve the thermoelectric conversion efficiency.

For Problem 1, this paper employs the Transfer Matrix Method to obtain the relationship between the emissivity, the wavelength, and the material properties. Through adjusting the thickness and refractive index, this paper explores the impact of the above factors. Further, the calculation is conducted on the $50nm$ single-layer tungsten, which reaches the peak emissivity 0.53 at the wavelength of $0.33\mu m$.

For Problem 2, this paper extends the model proposed in Problem 1 to adapt to the multilayer structure. Apart from exploring the impact of material properties on the emission spectrum, this paper also calculates the spectrum of the specific two-layer structure. Results show that the composite of $50nmSiO_2$ and $50nmW$ can reach the peak of 0.796 at the wavelength of $0.34\mu m$.

For Problem 3, it is regarded as the efficiency optimization task. This paper designs a multi-objective optimization model to make the peak higher and sharper within the ideal range. A genetic algorithm-based method is proposed to calculate the results. The optimal results have 6 layers(from the top layer to the bottom layer):$Ge(51.23nm)$, $W(66.33nm)$, $InAs(41.62nm)$, $Ta(70.30nm)$, $MgF_2(43.11nm)$, $GaAs(31.64nm)$.

For Problem 4, this paper takes the battery efficiency into account and adjusts the optimization function proposed in Problem 3. The optimal design is:$SiC(43.92nm)$, $GaSb(15.84nm)$, $W(32.29nm)$, $GaSb(23.50nm)$, $Pt(53.10nm)$, $Cr(28.51nm)$. Further, sensitivity analysis is conducted to prove the effectiveness of our model.

Keywords: Thermal Emitter Genetic Algorithm Multi-objective Optimization

Contents

I. Introduction

1.1 Problem Background

Thermophotovoltaic technology (TPV) is a technology that directly converts the energy of a heated high-temperature thermal emitter into electrical energy through a semiconductor PN junction cell. Its basic principle is the same as that of a solar cell, except that the conversion spectrum is different. Due to its fuel versatility, portability, silent operation, and independent work from the sun, TPV can be applied to sailing power systems, lunar rovers, and other equipment in the future [1], providing technical guarantee for their normal operation under conditions of no sunlight. Therefore, the research of thermal photovoltaic technology has always been a hot issue.

The actual thermal photovoltaic system mainly includes thermal emitters, optical filters, photovoltaic cell components, heat recovery devices, and auxiliary components. Among them, the thermal emitter mainly utilizes different material structures to adjust the emission of absorbed heat, so that most of the emitted photons are lower than the bandgap wavelength of the photovoltaic cell. The photovoltaic cells can convert high-energy photons below this specific bandgap wavelength to collect thermal radiation energy and convert it into electrical energy. Therefore, photovoltaic cells operate with certain bandgap energy and a corresponding bandgap wavelength, and low-energy photons above the wavelength cannot be converted into electrical energy. Thus, the thermoelectric conversion efficiency of the thermal photovoltaic system is limited by the emission spectrum of the thermal emitter.

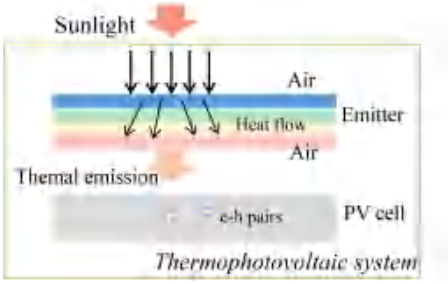


Figure 1 The diagram for the working process.

The emission spectrum of the thermal emitter varies with the optical and structural properties of the material. By combining materials with a specific refractive index or dielectric constant and optimizing the thickness of each of them, the thermoelectric conversion efficiency of the thermophotovoltaic system can be improved.

1.2 Restatement of the Problem

Problem 1: For the single-layer structure, establish the relationship between the emission spectrum and the material properties (refractive index, thickness), and then calculate the emission spectrum of 50-nm-thick tungsten in the range of 0.3-5 μm.

Problem 2: For the multilayer structure, establish the relationship between its emission spectrum and the material properties (refractive index, thickness), and calculate the emission spectrum of the composite structure formed by tungsten (50 nm) and silicon dioxide (50 nm) in the range of 0.3-5 μm.

Problem 3: Designing the thermal emitter as a narrow-band emission can effectively improve the spectral control capability of the radiator. Please design a multi-layer thermal emitter based on Problem 2, and by selecting suitable materials, optimize the structural parameters and emission spectrum to make it emit narrow-band and highest.

Problem 4: Select reasonable materials, optimize the structural parameters and emission spectrum according to the emission spectrum of its idealized thermal emitter, and design a multilayer thermal emitter for the GaSb battery to obtain the highest thermoelectric conversion efficiency.

1.3 Our Work

To tackle the problems above, this paper conducts the following work:

(1) For Problem 1, this paper will construct a single layer model based on the calculation of the emission spectrum to explore the impacts of material properties on the spectrum. Further, we will fix the thickness of the metal and give the output of the spectrum.

(2) For Problem 2, this paper will extend the model proposed for Problem 1 to the multilayer structure. Similarly, we will conduct the analysis of the impact of material properties and the calculation on the specific two-layer condition.

(3) For Problem 3, this paper will analyze from the perspective of efficiency and design a multi-objective optimization problem to make the emission spectrum as high and sharp as possible at the ideal wavelength. The genetic algorithm will be employed to tackle the optimization.

(4) For Problem 4, further take the battery efficiency into consideration. Based on the optimization model in Problem 3, this paper will adjust the optimization objective function to adapt to the condition of battery efficiency.

II. Symbol Decriptions

Some of the main symbols and their meanings are included in the table.

Table 1 Symbol descriptions.

Symbols	Meaning	Unit
n	The real part of the refractive index	-
k	The imaginary part of the refractive index	-
λ	Wavelength	μm
R	Reflection	-
T	Transmittance	-
A	Absorption	-
$e(\lambda)$	The function of the emission spectrum	-
N	The maximum number of layers to be selected	-
d_i	The thickness of i^{th} layer	nm
$Q(\lambda)$	The function of EQE curve	-

III. Assumptions

- We assume that there exists no energy loss during the emission period. Because our model is based on the ideal thermal equilibrium law, it ignores all extra loss.
- Because the given data is limited to the optical and structural properties of the materials, we assume that the material selection is not related to the cost and does not consider other factors(e.g. pollution effects).
- We assume that the total thickness of the multilayer structure has no strict restrictions, which means we can search for the optimal within a wider range.

IV. Single Layer Thermal Emitter for Problem 1

For Problem 1, this paper conducts an analysis of the single-layer structure and illustrates the relationship of the emission spectrum and the material properties. In addition, this paper shows the emission spectrum of tungsten with a fixed thickness of 50 nm.

4.1 The Model for Single Layer Structure

Considering that the tungsten performs as the only medium in the single structure, the incident wave will pass through the interface of the air and the tungsten twice, from air to metal as well as from metal to air. The process diagram is shown in Figure 2.

Figure 2 The process diagram for single layer.

In view of the transmission of the electromagnetic wave in the simple structure, this paper employs the classical Maxwell functions as follows.

$$\begin{cases} \nabla \times \overrightarrow{H} = \overrightarrow{J} + \frac{\partial \overrightarrow{D}}{\partial t} \\ \nabla \times \overrightarrow{E} = -\frac{\partial \overrightarrow{B}}{\partial t} \\ \nabla \cdot \overrightarrow{B} = 0 \\ \nabla \cdot \overrightarrow{D} = \rho \end{cases}, \tag{1}$$

where, H, J, D, E, and B denotes the magnetic field intensity, conduction current density, electric displacement vector, electric field intensity, and magnetic induction intensity respectively.

Note that this paper only considers the emission in the vertical direction, all the angles are set to zero, that is $\theta = 0$. Thus, we can derive the reflection coefficient r and the transmission coefficient t. Without losing generality, this paper shows the relationship between Layer i and Layer $i + 1$,

$$r_{i,i+1} = \frac{n_i - n_{i+1}}{n_i + n_{i+1}}, \quad t_{i,i+1} = \frac{2n_i}{n_i + n_{i+1}}, \tag{2}$$

where $i \in \{0, 1\}$ for Problem 1 since there exists three layers of medium. n_i denotes the refractive index for i^{th} layer. Note that in the calculation, this paper further considers

the propagation loss in the layer. Thus, the refractive index is in the form of the complex number, that is $n_i + jk_i$.

According to Transfer Matrix Method (TMM) [2, 3], we can first give the transmission matrix between i^{th} layer and $(i+1)^{th}$ layer,

$$D_{i,i+1} = \frac{1}{t_{i,i+1}} \begin{pmatrix} 1 & r_{i,i+1} \\ r_{i,i+1} & 1 \end{pmatrix}, \tag{3}$$

Meanwhile, the propagation matrix P_i of i^{th} layer can be represented as follows,

$$P_i = \begin{pmatrix} e^{-j\frac{2\pi}{\lambda}n_i d_i} & 0 \\ 0 & e^{-j\frac{2\pi}{\lambda}n_i d_i} \end{pmatrix}, \tag{4}$$

where λ is the incident wavelength, d_i denotes the thickness of i^{th} layer. In Problem 1, $d_1 = 50nm$ and we can obtain the matrices of D_{01}, D_{12}, and P_1. Therefore, the following equation is constructed based on [4],

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = D_{01}P_1D_{12}. \tag{5}$$

The reflection R and transmittance T of the tungsten layer can be derived,

$$R = \left|\frac{M_{21}}{M_{11}}\right|^2, \quad T = \frac{n_2}{n_0}\left|\frac{1}{M_{11}}\right|^2. \tag{6}$$

Since the boundary layer is air, we set $n_0 = n_2 = 1$ for simplicity. Meanwhile, the absorption A of the single layer structure satisfies the following equation,

$$A = 1 - R - T. \tag{7}$$

Assuming that the system reaches the thermal equilibrium state, the emission of the system is equal to the absorption [5].

$$|\Delta E_{emit}| = |\Delta E_{absorb}|. \tag{8}$$

For simplicity, this paper utilizes the absorption A to substitute the emissivity E. In conclusion, the emissivity of the single layer tungsten can be represented as,

$$E(\lambda) = 1 - \left|\frac{M_{21}}{M_{11}}\right|^2 - \left|\frac{1}{M_{11}}\right|^2. \tag{9}$$

4.2 Relationship of Emission Spectrum and Material Properties

Given different wavelengths λ, we can obtain various emissivities from Equation (9). These data pairs form the emission spectrum of the single layer material. Because refractive index $(n + jk)$ and thickness (d) are related to the calculation of emissivity, this paper will demonstrate the relationship between emission spectrum and material properties for single layer below. Take the metal tungsten as an instance.

Firstly, consider the optical characteristic of tungsten. According to the reference data materials, the refractive index has a close relation to the wavelength. We separate the analysis of the real part and imaginary part, and draw the curves of refractive index and wavelength in Figure 3.

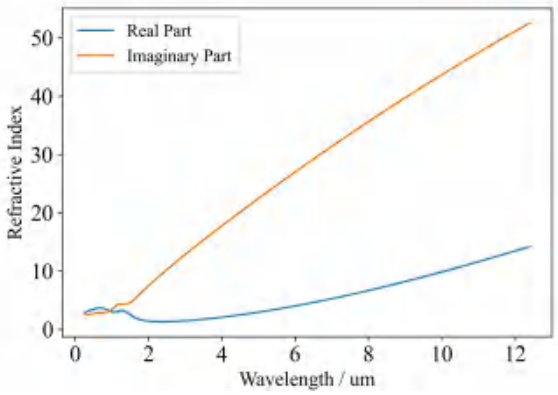
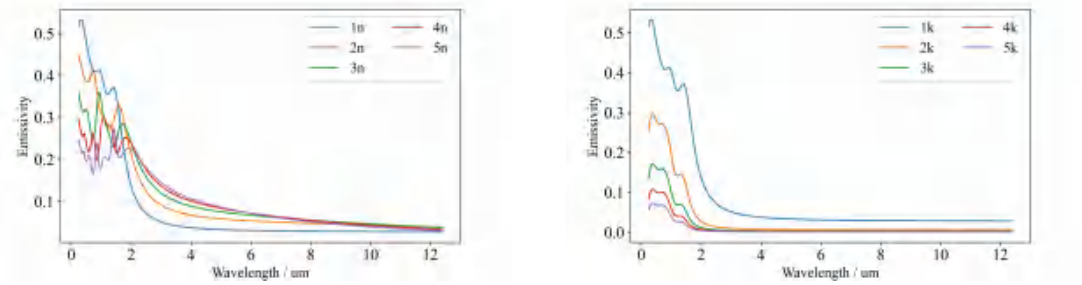


Figure 3 The relationship between wavelength and refractive index.

In general, with wavelength increasing, both parts of the refractive index increase. For the real part, the curve fluctuates in a small range when the wavelength is within $1.5\mu m$. And it reaches the minimum point when the wavelength is around $2\mu m$. For the imaginary part, the curve increases slowly within $1.5\mu m$ while it witnesses a sharp increase after it. Further, the real part of the refractive index is connected to the propagation characteristic and the imaginary part represents the extinction coefficient which can be regarded as a loss. That is to say, when the wavelength is over $2\mu m$, the loss is significant.

To further explore the function of refractive index on the emission spectrum, this paper increase the real part and the imaginary part of the refractive index k by multiple respectively. The trends of the spectrum changes are shown in Figure 4.

In Figure 4(a), we fix the thickness of the metal tungsten to $50nm$, remain the imaginary part of its refractive index to be original and change the real part by multiple. Within the range of 2 μm, the smaller value n achieves higher emissivity. With the



(a) Emission Spectrum with different n (b) Emission Spectrum with different k

Figure 4 Emission Spectrum with different refractive index.

wavelength increasing, higher value n shows its superior performance on emissivity. In Figure 4(b), we make the real part of the refractive index unchanged and increase the imaginary part by multiple. The results illustrate that a higher k value leads to lower emissivity. It proves the extinction effect of value k.

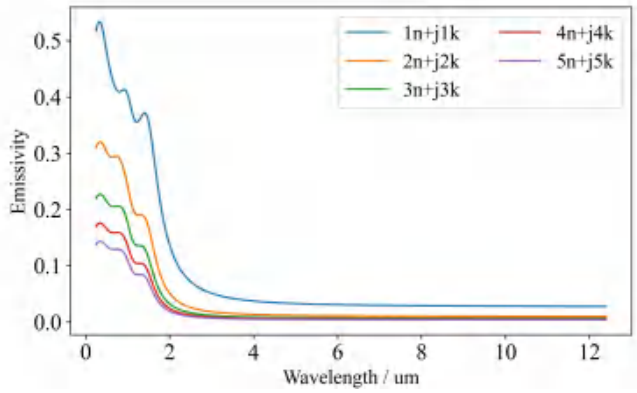


Figure 5 Emission Spectrum with different refractive index.

Also, we explore the impact of simultaneously increasing n and k. The results shown in Figure 5 are similar to the situation of singly changing k. A higher refractive index brings more extinction effects.

Secondly, consider the structure characteristic of the metal tungsten. This paper selects ten thickness numbers from $10nm$ to $100nm$ with an interval of 10. The relationship between the emission spectrum and thickness is shown in Figure 6.

Within the wavelength of $1\mu m$, the tungsten with a thickness above $60nm$ ($60nm$-$100nm$) has the highest and sharpest emissivity. With the wavelength increasing, the thinner one witnesses the better emissivity effects. In general, the moderate thickness can take all the wavelengths into account and obtain the best performance.

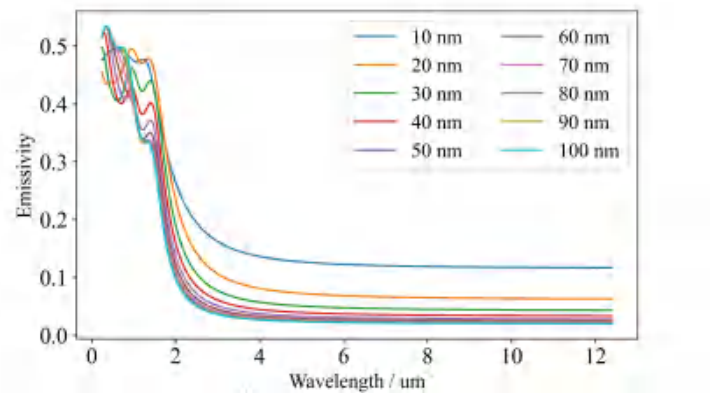


Figure 6 Emission spectrum with different thickness.

4.3 Calculation of the Emission Spectrum

In this part, this paper fixes the thickness of tungsten to $50nm$ and calculates the emission spectrum within the wavelength of $0.3-5\mu m$. According to the derived equation above, we can obtain Figure 7.

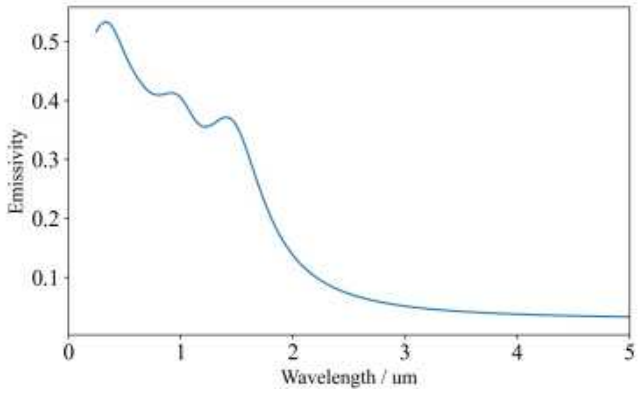


Figure 7 Emission spectrum with fixed thickness of 50nm.

The emission spectrum reaches the peak emissivity at the wavelength of $0.33\mu m$ and the peak emissivity value is around 0.53. The emissivity maintains a relatively high level under $1.5\mu m$, and it drops sharply afterward.

V. Multilayer Thermal Emitter for Problem 2

For Problem 2, this paper extends the single layer structure to the multilayer. Also, the relationship between the emission spectrum and material properties is illustrated. Further, this paper presents the emission spectrum of the composite structure with specific settings.

5.1 The Model for Multilayer Structure

Multilayer structure is shown in Figure 8. Assume that the number of the composite materials is N. Therefore, the structure consists of $N + 2$ layers, including the two boundary layers of air. According to TMM, the formula can be extended to:

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = D_{0,1}P_1D_{1,2}P_2 \cdots D_{N-1,N}P_N D_{N,N+1}. \tag{10}$$

The forms of the transmission matrix D and the propagation matrix P remain the same with the single layer situation.

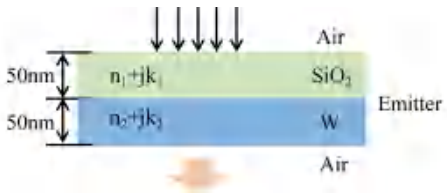


Figure 8 The process diagram for multilayer structure.

Therefore, the emissivity of the multilayer structure can also be concluded as:

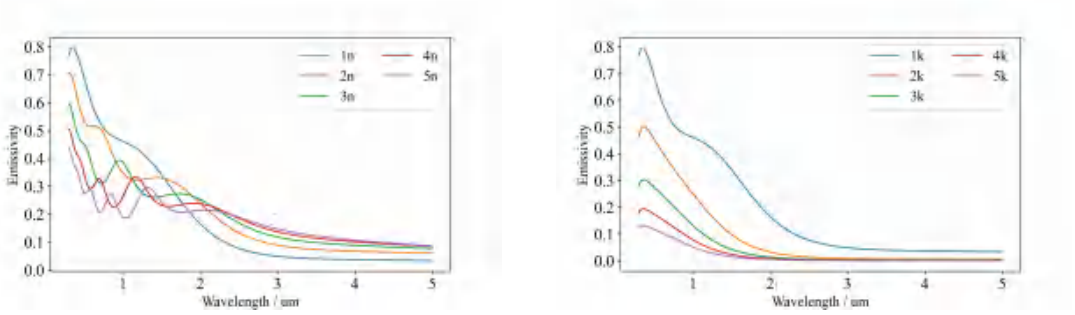
$$e(\lambda) = 1 - \left|\frac{M_{21}}{M_{11}}\right|^2 - \left|\frac{1}{M_{11}}\right|^2. \tag{11}$$

The formula form is the same as the single layer situation. Their differences lie in the calculation of M_{11}, M_{12}, M_{21}, and M_{22}, where the influences of multilayer are stacked in the multiplication.

5.2 Relationship of Emission Spectrum and Material Properties

Similar to Section IV, this paper considers the different impacts of material properties on the emission spectrum. For simplicity, we only consider the two-layer structures with metal tungsten and silica.

First, consider the impact of the refractive index. Similar to the single layer analysis, this paper increases the refractive index to explore its relationship with the emission spectrum. Fix the thickness of both layers to 50 *nm*, remain the refractive index of silica unchanged, and singly increase the value n of tungsten by multiple, we obtain Figure 9(a). Similarly, increase the value k and obtain Figure 9(b).



(a) Emission Spectrum with different n of tungsten (b) Emission Spectrum with different k of tungsten

Figure 9 Emission Spectrum with different refractive index of W.

From the curves, we can draw the conclusion that changing one of the coefficients will have an obvious influence on the spectrum of the composite structure. Also, it shows that the two-layer structure significantly improves the emissivity. Compared with single layer tungsten, the tungsten-silica structure reaches the peak value of 0.796, improving by 33.4%. Also, we simultaneously increase n and k of tungsten and obtain the curves in Figure 10.

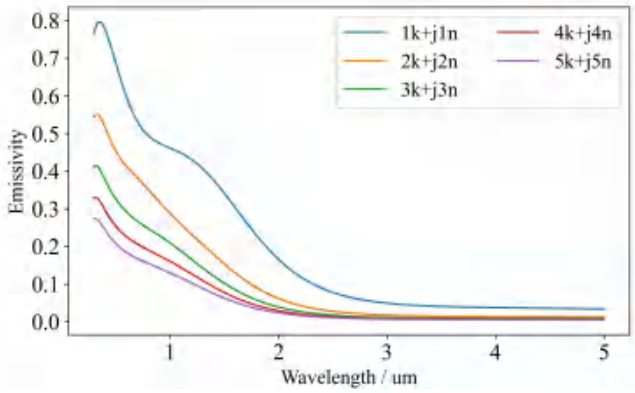


Figure 10 Emission spectrum with different refractive index of W.

Because the refractive index data of silica does not have the value of k, this paper merely adjusts the value of n to explore its relationship with the emission spectrum. The curves are shown below in Figure 11. When n increases by three times or more, it

appears more than two peaks in the curve, which will have a bad effect on the conversion efficiency.

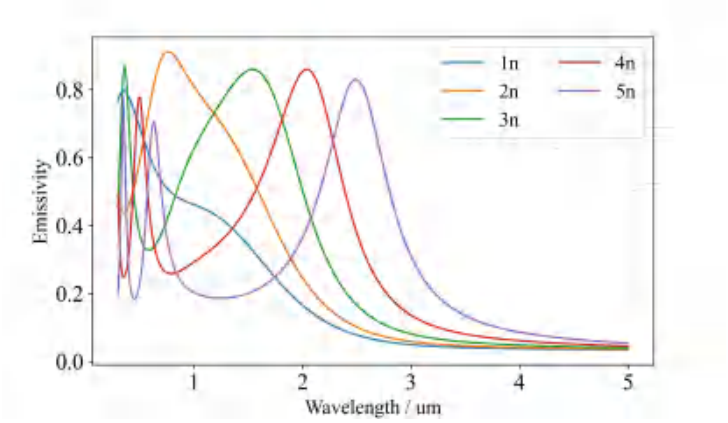
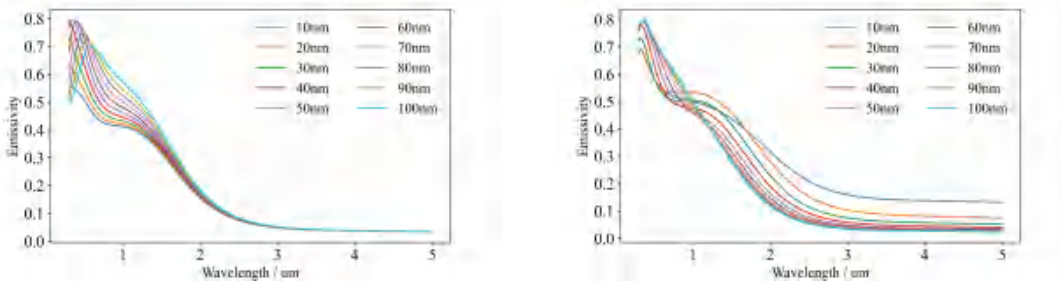


Figure 11 Emission spectrum with different n for SiO2.

Second, consider the impact of thickness on the emission spectrum. We first maintain the tungsten and adjust the thickness of SiO_2 in the range of $[10nm, 100nm]$ with an interval of $10nm$. The curves are shown in Figure 12(a).



(a) Different thickness of SiO2 (b) Different thickness of W

Figure 12 Emission Spectrum with different refractive index of W.

Within the range of $10nm - 40nm$, the emissivity witnesses an obvious increase, but still has not formed the peak. When the thickness reaches $50nm$, the peak of the spectrum appears at the wavelength of $0.34\mu m$. With the thickness increasing, the peak starts to shift to the right and the peak value drops.

Then, we fix the thickness of SiO_2 to 50 nm and change that of W from $10nm$ to $100nm$. Similar to the situation in single layer structure, the thicker ones have a higher peak and sharper curve. It is meaningful for the thickness selection in the following problems.

5.3 Calculation of the Emission Spectrum

We fix the thickness of both materials to $50nm$ and obtain the spectrum in Figure 13.

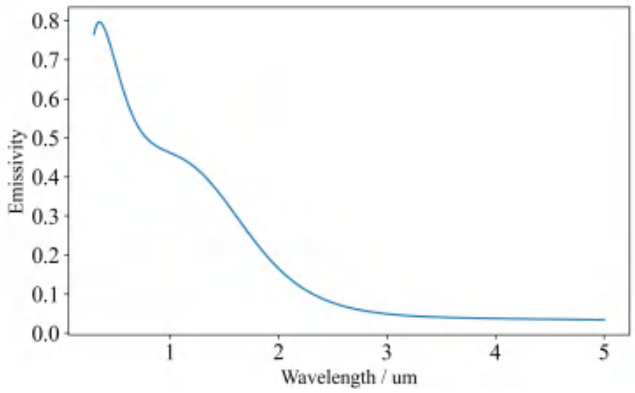


Figure 13 Calculation of emission spectrum with fixed thickness.

The emission spectrum reaches the peak emissivity at the wavelength of $0.34\mu m$ and the peak emissivity value is around 0.796. Similarly, the emissivity maintains a relatively high level under $1.5\mu m$, and it drops sharply afterward.

VI. Model Establishment and Solution for Problem 3

6.1 Model Establishment

Considering that our proposal is to maximize the output electric energy with the fixed input thermal energy, optimizing thermoelectric conversion efficiency is of great importance. This paper divides the total conversion efficiency η into two parts, thermal optical conversion efficiency η_{to} and photoelectric conversion efficiency η_{pe}. In Problem 3, we ignore the impacts of η_{pe}. Thus, thermal optical conversion efficiency η_{to} is regarded as the potential optimization objective.

To make a connection between η_{to} and the emission spectrum, we make the assumption that the thermal energy is equally divided among all excited wavelengths. That is to say, the fewer the excited wavelengths, the higher average energy they get. Therefore, the ideal emission spectrum can gather the wavelengths among the range with the highest emissivity $e(\lambda)$. From the spectrum, the thermal emission is required to be as sharp and high as possible.

Note that the ideal highest thermal emission is obtained at the wavelength of $1.5\mu m$, thus one of the optimization objects is to maximize the value of $e(1.5)$. Meanwhile, in order to make the spectrum as sharp as possible around $1.5\mu m$, we propose to consider the wavelength at $e(1.5)/2$. Since $\lambda = 1.5\mu m$ is the maximum point, both sides of it are approximate monotone functions. So there exist two wavelengths at $e(1.5)/2$. The difference between the two values $(\lambda_1 - \lambda_2)$ can be employed to measure the sharpness of the thermal emission. Therefore, this multi-objective optimization task can be expressed as follows:

$$\begin{aligned} &\max w_1 \cdot e\,(1.5) + w_2 \cdot \tfrac{1}{\lambda_2 - \lambda_1} \\ &s.t. \begin{cases} \lambda_1 = \min\{e^{-1}(e(1.5)/2)\} \\ \lambda_2 = \max\{e^{-1}(e(1.5)/2)\} \\ 0.3\mu m \le \lambda \le 5\mu m \\ 1 \le N \le 7 \\ 10nm \le d_i \le 100nm, i \in [0, N) \end{cases} \end{aligned} \tag{12}$$

where w_1, w_2 are the weights of the multi-objective, which are hyperparameters. $e^{-1}()$ denotes the inverse function of $e()$. N is the layer number of the designed structure (excluding air layers). d_i denotes the thickness of i^{th} layer.

6.2 Solution and Analysis

6.2.1 *Genetic Algorithm for Optimization Task*

For the established multi-objective dynamic programming model to optimize the parameter design of multi-layer structure thermal emitter, the targets to be solved are the number of material layers, the type and thickness of each layer of material. At the same time, the range of selectable values of the solution target is not fixed, and the search space is large, which results in the complexity of solving the optimization problem of parameter design of multi-layer structure thermal emitter, and it is difficult to find the absolute optimal solution. However, genetic algorithm adopts stochastic optimization technology, which requires no continuity and differentiability. Therefore, a genetic algorithm is considered in this paper to solve the optimization results of the parameter design of multilayer structure thermal emitter.

First, take the requested variable coding, this paper takes required variables for a thermal emitter material layer number, species, and thickness, set each multilayer structure design of thermal emitter parameters of material layer for n, types of materials for k, the thickness of the material for d, At the same time, each variable satisfies the following constraint conditions:

$$\begin{aligned} &n, k, d \in [\min, \max], \\ &k = \{k_1, k_2, ..., k_n\}, \\ &d = \{d_1, d_2, ..., d_n\}, \end{aligned} \tag{13}$$

Among them, since the material type is not a specific value, the material type is labeled with a digital label, representing different kinds of materials with discretized data. That is, the materials in the database are numbered, the minimum value of k is 1, and the maximum value is the total number of materials in the database. Each corresponding parameter of the multi-layer thermal emitter parameter design scheme is coded according to the range of available solutions, and then the parameter coding results of all the design schemes are spliced together to form an individual genotype, whose phenotype is as follows.

$$Y = \{n, k_1, k_2, ..., k_n, d_1, d_2, ..., d_n\}. \tag{14}$$

The population size POP is initialized, and each individual in the population corresponds to a feasible solution, and the population with several feasible solutions is generated by random method.

$$POP = rand\{Y_1, Y_2, ..., Y_n\}, \tag{15}$$

where $Y_i = \{rand(n), rand(k_1), ..., rand(k_n), rand(d_1), ..., rand(d_n)\}$.

Fitness function design is the most important part of the genetic algorithm solution model constructed in this paper. According to the established multi-objective dynamic programming model, an evaluation mechanism that can effectively reflect individual fitness is constructed in this paper.

$$fitness = D = w_1 e\,(1.5) + w_2 \frac{1}{\lambda_2 - \lambda_1} \tag{16}$$

In order to increase the speed of evolution and optimization, an elitist strategy was introduced into the solving model of genetic algorithm constructed in this paper, that is, by comparing individual fitness of each generation population, several of the highest genotypes were selected and the genotypes were reserved for the next generation. Then select, cross and mutate other individuals in the population.

Selection: According to the fitness of each individual in the population, the selection probability is calculated. Each individual corresponds to a feasible solution of the parameter design scheme of a multilayer thermal emitter. The specific probability calculation is as follows.

$$p_l = \frac{D_l}{\sum_{k=1}^{C} D_k} = \frac{\left(w_1 e(1.5) + w_2 \frac{1}{\lambda_2 - \lambda_1}\right)_l}{\sum_{k=1}^{C} \left(w_1 e(1.5) + w_2 \frac{1}{\lambda_2 - \lambda_1}\right)_k}. \tag{17}$$

According to the probability, 0-1 region is divided, where the region size is positively correlated with the probability. The feasible solution of a multilayer thermal emitter parameters corresponding to each individual of the selected population is determined by generating random numbers between 0-1.

Cross: The genotype of each individual in the population consists of feasible solutions to the parameters of the multi-layer thermal emitter design scheme. According to the set cross probability cross, cross-exchange of a gene fragment in different genotypes between individuals (cross-exchange of feasible solutions of parameters representing a multi-layer thermal emitter design scheme).

$$G_l\{n, k_1, ..., k_n, d_1, ..., d_n\} \leftrightarrow G_p\{n, k_1, ..., k_n, d_1, ..., d_n\}. \tag{18}$$

Variation: in order to jump out of the local optimal solution trap, better in global optimization, by setting mutation probability var, in every time evolution of the one on

the individual genes (i.e., the parameters of a multi-layer thermal emitter design feasible solution) on variation, variation value within the design parameters of the feasible solution space randomly generated.

$$G_l\{n, k_1, ..., k_n, d_1, ..., d_n\} = rand\{n, k_1, ..., k_n, d_1, ..., d_n\}. \tag{19}$$

Through continuous selection, crossover and mutation, the population is constantly evolving and iterating, so as to seek the optimal individual, that is, the optimal feasible solution of multi-layer thermal emitter parameter design scheme. The overall design framework is shown in Figure 14.

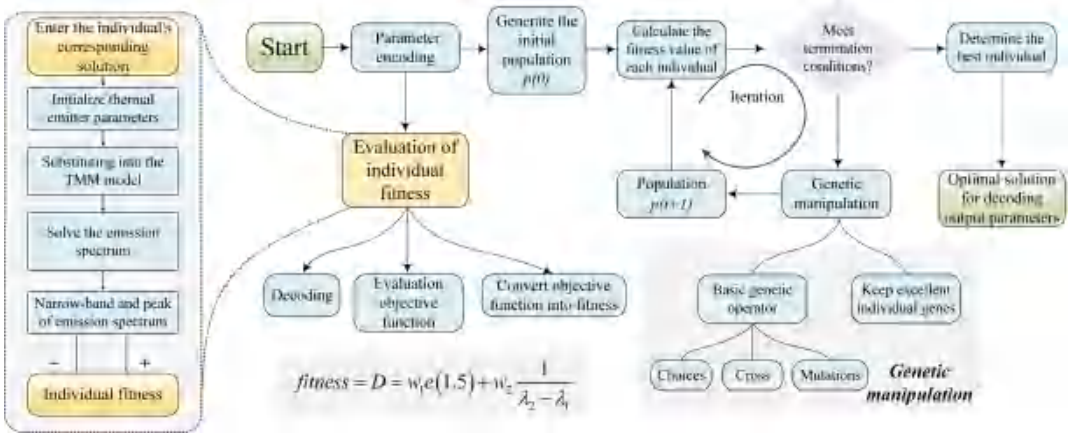


Figure 14 The process of Genetic Algorithm.

6.2.2 *Model Solving and Main Results*

Since the problem can be regarded as a multi-object optimization task, we propose to build an optimization solution model based on the genetic algorithm. The objectives to be solved and their search scopes are listed in Table 2.

Table 2 The objectives to be solved and their search scopes.

Objectives	Material for each layer	Layer Number	Thickness for each layer
Search Scope	26 in total	1-7	$10nm - 100nm$

Under the above conditions, the main result of Problem 3 is obtained by the genetic algorithm. The optimal layer number is 6 (excluding the two air layers). This paper lists the material selected and its thickness for each layer in Table 3.

Table 3 The optimal results for the designed multilayer structure.

Layer	Material	Thickness
I	Ge	$51.23nm$
II	W	$66.33nm$
III	$InAs$	$41.62nm$
IV	Ta	$70.30nm$
V	MgF_2	$43.11nm$
VI	$GaAs$	$31.64nm$

According to the optimal structure, this paper draws the emission spectrum to analyze the results. The schematic of the optimal multilayer structure is shown in Figure 15(a) and the emission spectrum is shown in Figure 15(b).

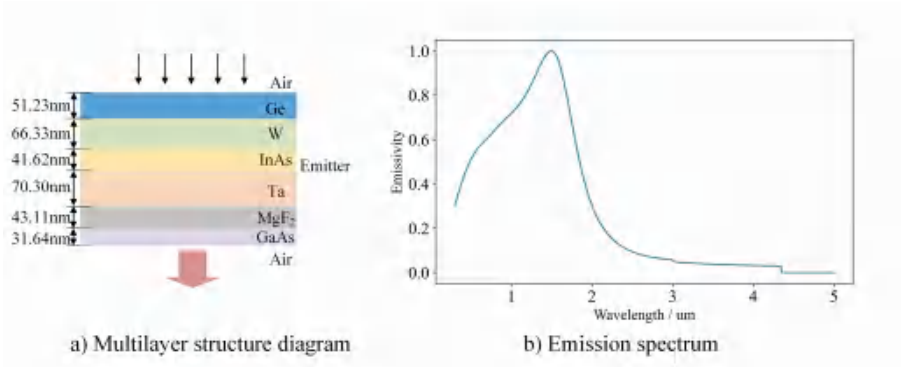


Figure 15 The optimal schematic and the emission spectrum.

From the curve, the peak value is obtained at the wavelength of 1.5 μm and the peak value is over 0.99. It meets the first optimal objective, which requires the peak to be as high as possible. Meanwhile, the curve is sharp around the wavelength of $1.5\mu m$, which makes the energy to be gathered among the most efficient wavelength.

Further, this paper reduces the maximum layer number condition to explore the optimal results under each layer number.

When the maximum layer number is 1, the problem degenerates to the optimization of single layer structure. The optimal material is tungsten with the thickness of $16.66nm$. Under this condition, the peak value of emissivity is nearly 0.5 and the curve around the peak is not sharp. This also confirms that the single layer structure is far from efficient for the thermal emiiter.

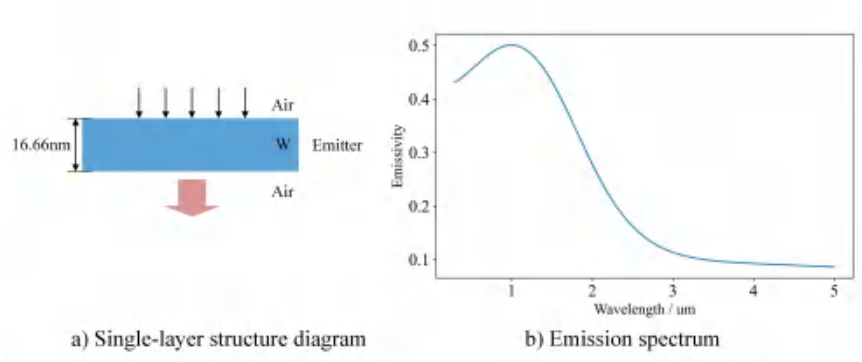


Figure 16 Maximum Layer Number is 1.

When it comes to the two-layer structure, the results show that 69.20nm GaSb and 53.76nm Pt are the best combination for the emitter. However, from the curve, the peak value is obtained when the wavelength is over 1.5μm, which is not ideal.

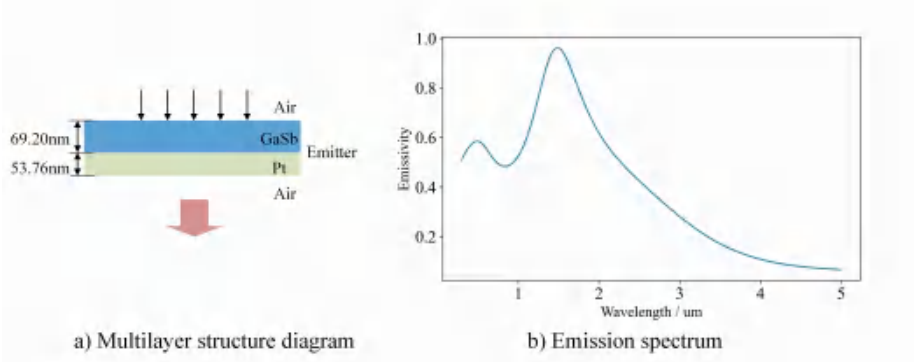


Figure 17 Maximum Layer Number is 2.

When the maximum layer number is set to 3, 4, and 5 respectively, the curve is optimized step by step. The total trend of the optimization process is making the peak higher and sharper among the wavelength of 1.5 μm. This paper only shows the condition when maximum layer number is 5 in Figure 18.

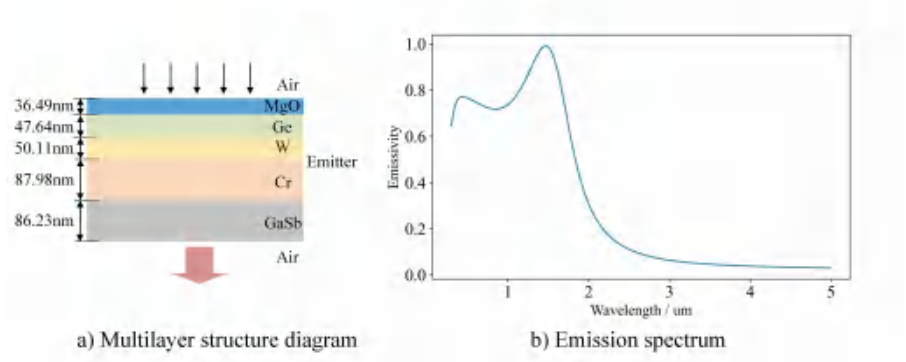


Figure 18 Maximum Layer Number is 5.

6.2.3 *Sensitivity Analysis*

Table 4 The optimal results for the ideal peak wavelength of 1μm.

Layer	Material	Thickness
I	Ge	75.75nm
II	Ti	26.53nm
III	Pt	35.01nm
IV	Cu	42.89nm
V	MgF_2	46.76nm
VI	Au	50.28nm

Table 5 The optimal results for the ideal peak wavelength of 2μm.

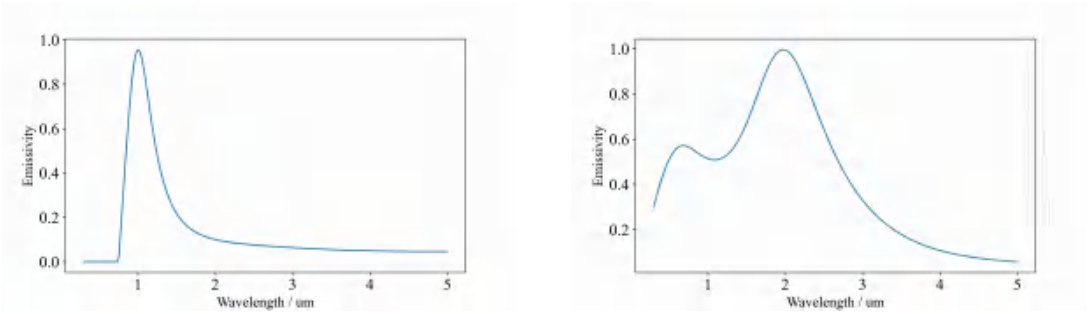
Layer	Material	Thickness
I	MgF_2	15.78nm
II	Zn	20.61nm
III	SiC	23.98nm
IV	MgF_P2	92.04nm
V	Al_2O_3	33.44nm
VI	Cr	72.75nm

In order to further verify the sensitivity of our proposed model, we conduct the following experiments. The main idea is to change the ideal peak wavelength and observe the capability of the model to respond to changes in constraints.

We fix the maximum layer number to 6. Firstly, change the ideal peak wavelength to 1 μm. The optimal layers are obtained in Table 4.

Meanwhile, the optimal emission spectrum is given in Figure 19(a). As can be seen from the curve, the ideal peak wavelength of 1μm is well satisfied.

Secondly, change the ideal peak wavelength to 2 μm. The optimal layers are obtained in Table 5. The optimal emission spectrum is given in Figure 19(b). Similar to the situation above, the ideal peak wavelength is obtained.



(a) Ideal wavelength of $1\mu m$ (b) Ideal wavelength of $2\mu m$

Figure 19 Emission spectrum with different ideal peak wavelengths.

In conclusion, the sensitivity analysis further proves the effectiveness of our model.

VII. Model Establishment and Solution for Problem 4

7.1 Model Establishment

In Problem 3, we only consider the thermal optical conversion efficiency η_{to} and ignore the photoelectric conversion efficiency η_{pe}. In the real situation, the battery has a characteristic of wavelength selective. Within the band-gap of the battery material, it has the External Quantum Efficiency (EQE), which is an embodiment of photoelectric conversion efficiency. It restricts the emission spectrum to a certain range. Therefore, our proposed optimization model for Problem 4 will cover both parts of η_{to} and η_{pe}.

On one hand, in order to improve the thermal optical conversion efficiency η_{to}, the emission spectrum is required to be gathered within the band-gap wavelength. Therefore, we propose the following optimization expression:

$$\max \frac{\int_{0.3}^{1.7} e(\lambda)d\lambda}{\int_{0.3}^{5} e(\lambda)d\lambda}. \tag{20}$$

Note that the calculated wavelength range is 0.3-5 μm, we set the starting point of the integral interval to 0.3.

On the other hand, in order to improve the photoelectric conversion efficiency η_{pe}, we will make the product of EQE and emissivity within the band-gap as large as possible. In Problem 4, The gallium antimonide battery is considered and its band-gap wavelength is 1.71 μm. Therefore, the optimization expression is:

$$\max \int_{0.3}^{1.71} e(\lambda) \cdot Q(\lambda)d\lambda, \tag{21}$$

where Q_λ is the expression of external quantum efficiency curve.

To make it easy for software to do the optimization task, we transform the continuous integral form into a discrete summation form. Also, we integrate two optimization expressions together with weights,

$$\begin{aligned} &\max w_1 \sum_{\lambda\in[0.3,1.71]} e(\lambda)Q(\lambda) + w_2 \frac{\sum_{\lambda\in[0.3,1.71]} e(\lambda)}{\sum_{\lambda\in[0.3,5]} e(\lambda)} \\ &s.t. \begin{cases} 1 \le N \le 7 \\ 10nm \le d_i \le 100nm, i \in [0, N) \end{cases} \end{aligned}. \tag{22}$$

Thus, the optimization model is constructed. The solution and analysis of Problem 4 will be followed in the next subsection.

7.2 Solution and Analysis

7.2.1 *Model Solving and Main Results*

Similar to Problem 3, this paper continues to solve the optimization task based on genetic algorithm. The optimal layers with thickness parameters are as follows(from the top layer to the bottom layer): $SiC(43.92nm)$, $GaSb(15.84nm)$, $W(32.29nm)$, $GaSb(23.50nm)$, $Pt(53.10nm)$, $Cr(28.51nm)$. The optimal structure diagram and the emission spectrum are shown in Figure 20(a) and 20(b) respectively.

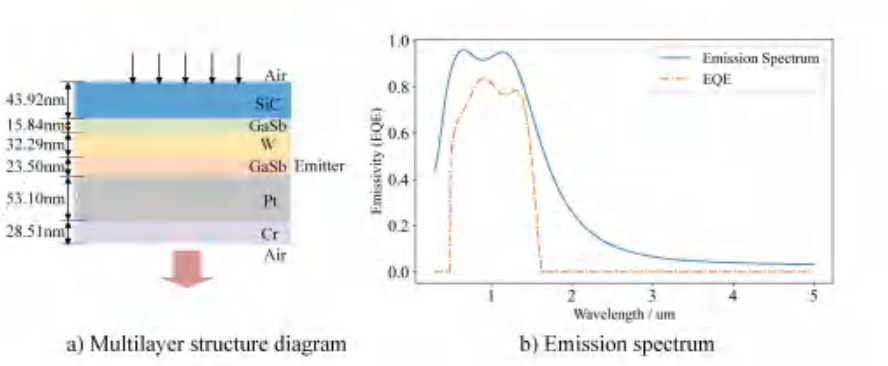


Figure 20 The optimal results for Problem 4.

We draw the EQE curve and the emission spectrum in the same figure. From the emissivity itself, the curve is sharp within the band-gap of $1.71\mu m$ and the peak value is over 0.95. Meanwhile, the obtained emission spectrum is highly matched with the EQE curve. That is to say, the emitted wavelengths are gathered among the points with higher EQE values.

7.2.2 *Sensitivity Analysis*

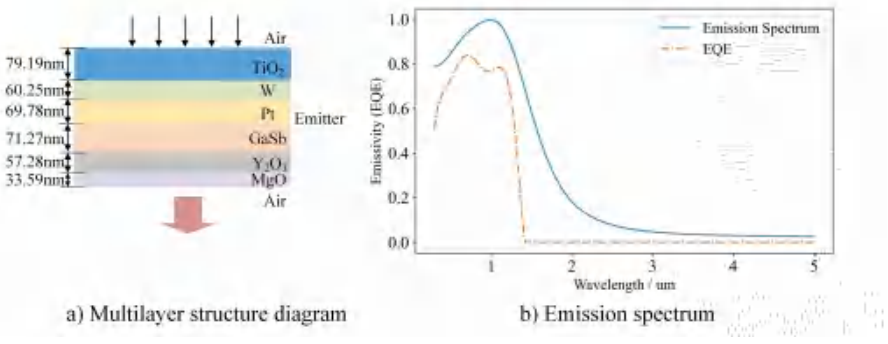


Figure 21 The results of the sensitivity analysis.

In order to verify the sensitivity of our model, this paper conducts an additional analysis experiment. In detail, we offset the EQE curve to the left side with 0.2 μm while remaining the relative values unchanged. The optimal results are shown in Figure 21 below. The optimal layers with thickness parameters are as follows(from the top layer to the bottom layer): $TiO_2(79.19nm)$, $W(60.25nm)$, $Pt(69.78nm)$, $GaSb(71.27nm)$, $Y_2O_3(57.28nm)$, $MgO(33.59nm)$.

The curves and results support our claims and prove the sensitivity of our model, as they adapt to the changes of the EQE curve.

VIII. Conclusion

8.1 Strength and Weakness

Strength

(1) Our model divides the process into two parts, that is thermal optical conversion and photoelectric conversion. And we design the weighted optimization function, considering both of the two parts. The core idea of the model is simple and clear, with excellent interpretability.

(2) Our model has good generalization performance. According to the sensitivity analysis, our model can properly respond to changes in the requirements (e.g. different EQE curves), which is of great importance in real situations.

Weakness

(1) Limited by the data collected, we do not consider some factors, such as energy loss and material cost in the design process. Meanwhile, we do not give the simulation results to further verify our model.

8.2 Future work

Firstly, we are required to design more complex models, which takes some extra factors into account. For example, we can design an additional optimization objective to minimize the cost during the selection process. The economic factors and energy conversion efficiency can be combined together to design better solutions.

Secondly, from the aspects of the models, we can utilize the BP network to do the precise material data fitting. Because the data lengths of different materials differ, it is required to preprocess the data to match the total number. The BP network can output fine-grained data points.

IX. References

[1] Coutts T J. An overview of thermophotovoltaic generation of electricity[J]. Solar energy materials and solar cells, 2001, 66(1-4): 443-452.

[2] Mohammed Z H. The fresnel coefficient of thin film multilayer using transfer matrix method tmm[C]. IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019, 518(3): 032026.

[3] Katsidis C C, Siapkas D I. General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference[J]. Applied optics, 2002, 41(19): 3978-3987.

[4] Wang H, Alshehri H, Su H, et al. Design, fabrication and optical characterizations of large-area lithography-free ultrathin multilayer selective solar coatings with excellent thermal stability in air[J]. Solar Energy Materials and Solar Cells, 2018, 174: 445-452.

[5] Manabe S, Wetherald R T. Thermal equilibrium of the atmosphere with a given distribution of relative humidity[J]. 1967.

Team Number :	apmcm2112772
Problem Chosen :	C

2021 APMCM summary sheet

The transformation of China's Saihanba Forest Farm from a desert to an oasis is an excellent example of the establishment of an ecological reserve. This article mainly uses the TOPSIS evaluation model based on the entropy weight method, the multiple regression model and the evaluation model based on BP Neural network to study the impact of the establishment of the Saihanba Ecological Reserve on the local ecological environment and the ability to resist wind and sand, and how to scientifically Planning to establish ecological protection zones.

For question 1, we need to establish a TOPSIS evaluation model of Saihanba's impact on the environment. First, select the indicators and data that have an impact on the ecological environment in Saihanba area. For data that cannot be queried, we use cubic spline interpolation to fill in. In order to reduce the complexity of the model, we use the entropy method to reduce the dimensionality to obtain the index set. Finally, we can get the environmental score of Saihanba over the years and find that the ecological environment has improved significantly.

In response to question 2, we first use the Spearman correlation coefficient to test the ecological environment of Saihanba and Beijing's ability to resist sand and dust, and find that the two have strong convergence, indicating that the indicator set in question 1 can be used to represent Saihanba. Next, establish a multiple regression model of the relationship between Saihanba's ecological indicators and Beijing's ability to resist sand and dust. It is found that the restoration of Saihanba Forest Farm has a significant effect on Beijing's ability to resist sand and dust.

Aiming at problem three, we first establish a BP neural network model between ecological evaluation indicators and three ecological capabilities based on the data of Saihanba. Then, the comprehensive ecological index is obtained through the weighted average of these three ecological capacity factors, and the scale of the ecological area that needs to be established in the place is judged through the negative index equation. Then, the carbon absorption capacity of the area was recalculated and compared with the carbon absorption capacity before construction. It was found that the carbon absorption capacity of each area had a significant upward trend.

In response to question 4, we chose Australia, which accounts for a large amount of desert area, to analyze and divide it into 7 regions. Based on the model established in question 3, we calculated the actual scale of ecological regions that can be built in each region of Australia and the carbon absorption after construction. The conclusion of the change trend of ability is similar to the third question.

Finally, we also wrote a non-technical report to describe the built model and put forward feasible plans and suggestions for the establishment of ecological reserves.

Keywords: Entropy method; TOPSIS; Spearman correlation coefficient; Multiple regression model; BP Neural network

Contents

1 Introduction

1.1 Background

We all know that to build a beautiful China, we must adhere to the concept that green waters and green mountains are golden mountains and silver mountains, and sustainable development. China's Saihanba Forest Farm is a good example. With the help of the Chinese government, the Saihanba Forest Farm has been transformed from a desert before its establishment into an oasis today. Compared with 1962, the forest coverage of Saihanba has increased from 11.4% to 80% in more than half a century. Above, Saihanba can absorb 747,000 tons of carbon for Beijing and Tianjin every year. It can also effectively resist sandstorms and conserve water. And it can be seen that the afforestation history of Saihanba has far-reaching developmental significance. Therefore, we can build models based on the historical data of Saihanba and reflect the mission of prospering ecology in more places.

1.2 Problem requirements

(1) Select appropriate indicators and collect relevant data, establish an evaluation model of Saihanba's impact on the ecological environment, and analyze the environmental conditions before and after Saihanba′s restoration.

(2) Select appropriate indicators and collect relevant data to establish an evaluation model of Saihanba's impact on Beijing's Anti-dust ability, and evaluate Saihanba's influence on Beijing's Anti-dust ability.

(3) Collect data to establish a mathematical model, extend Saihanba's ecological model to the whole of China, determine where the ecological zone needs to be established, and determine the scale of the proposed ecological zone.

(4) Choose another country in the Asia-Pacific region to build a mathematical model and collect data, analyze which locations in this country need to establish an ecological zone, determine the scale of the proposed construction, and evaluate its impact on the absorption of greenhouse gases and the reduction of carbon emissions.

2 Problem analysis

2.1 Analysis of Problem One

In order to analyze the environmental conditions before and after the restoration of Saihanba, we can objectively assign an evaluation score to the two for comparison. After collecting relevant indicators and data, we can use interpolation to fill in missing data, and use entropy to find several important indicators. Then, we can use the TOPSIS method to evaluate

the ecological environment of Saihanba through these weighted indicators, and finally get the annual ecological environment score of Saihanba from 1962 to 2020, and conduct a comparative analysis.

2.2 Analysis of Problem Two

Problem Two requires us to study the relationship between Saihanba's recovery and Beijing's Anti-sand dust ability. We need to find a new set of indicators to evaluate this pair of relationships. But before that, we can also try to test the correlation between Saihanba's ecological environment and Beijing's Anti-sand dust ability. If the correlation between the two is good, we can use five indicators for evaluating Saihanba's ecological environment to evaluate Beijing's Anti-sand dust ability. Next, we can roughly determine a multiple regression equation based on the line chart of Beijing's Anti-sand dust ability over time, and finally establish a mathematical model of the relationship between the recovery of Saihanba and Beijing's Anti-sand dust ability.

2.3 Analysis of Problem Three

In order to determine the scale of the establishment of the ecological zone, we can set a comprehensive ecological index for evaluation. The comprehensive ecological index can be determined by three ecological ability factors, such as the Anti-sand dust ability, Carbon absorption capacity and Water purification capacity. We can establish a model of the relationship between evaluation indicators and ecological capabilities through the BP Neural network, and then we can use the weighted average of the three ecological ability factors to determine the comprehensive ecological indicators. After obtaining the comprehensive ecological indicators, we can use the corresponding equations to judge whether the area needs to be established and the scale of the ecological area. Assuming that the ecological area is successfully completed in accordance with the construction plan, we can recalculate the Forest coverage of the area and bring it into the Neural network model to find the Carbon absorption capacity after the ecological area is constructed, and then combine it with the ecological area. The previous Carbon absorption capacity is compared, and a conclusion is drawn on the impact of the ecological zone on carbon neutrality.

2.4 Analysis of Problem Four

Problem Four is similar to Problem Three, we only need to find the relevant indicator data of a country in the Asia-Pacific region and bring it into the model in Problem Three to draw conclusions.

3 Model assumptions

(1) Assuming that the change in Beijing's Anti-sand dust ability is only related to the ecological changes in Saihanba.

(2) The change trend of Beijing's Anti-sand dust ability is logarithmically related to the evaluation index.

(3) It is assumed that the establishment of Saihanba Ecological Zone will have the same impact on the environment as other regions in the world.

(4) Assuming that when ecological zones are established in various administrative regions in China, the weights of ecological capabilities as evaluation factors are consistent with those of Saihanba.

(5) Suppose that in the initial stage of the establishment of the ecological zone, a larger scale of ecological zone needs to be built to significantly change the ecological environment; as the scale of the ecological zone gradually increases, the rate of improvement of the ecological environment also increases accordingly.

(6) It is assumed that all the ecologically-developed land planned by the government can be used to build an ecological zone.

4 Symbol description

Symbol	Description	Unit
V_{ij}	Normalized standard matrix	/
P_{ij}	Under the index i, the characteristic proportion of the j-th evaluation object	%
e_j	Entropy of the j-th indicator	/
d_j	The coefficient of variance for the index j	%
w_j	Entropy weight of the index j	%
Nh	Number of hidden layer neurons	PCS
N_i	Number of neurons in the input layer	PCS
No	Number of neurons in the output layer	PCS

5 Model building and solution of question 1

In order to analyze the environmental conditions before and after the restoration of Saihanba, we can objectively assign an evaluation score to the two for comparison. Among them, the TOPSIS method is a good objective evaluation method. TOPSIS is a systematic evaluation

method suitable for multi-index analysis. We can collect a series of indicators representing the ecological environment. By calculating the weighted Euclidean distance between the relevant indicators and the positive and negative ideal solutions for each year, we can obtain the indicators and positive ideals for each year. The closeness of the solution is used as the basis for evaluation.

5.1 Data preprocessing

First, we need to select the representative factors that affect the ecological environment in the Saihanba area as the index set. By consulting relevant information, we selected 10 indicators of Forest coverage, Air quality index, Water quality index, Soil pollution index, Biological abundance index, Fixed carbon dioxide, Oxygen released, Forest tree species structure, Climate index, and Ecological resilience[1].

We tried to collect data on these indicators in the Saihanba area since 1962, but we found a problem. It is difficult for us to accurately find the historical data of the Saihanba area over a long period of time. In order to establish a more reasonable model, we consider Use the collected data to interpolate the unknown data.

We can use cubic spline interpolation algorithm to smoothly connect missing data. Let $S(x)$ be a cubic spline interpolation function, and its construction method has the following form:

$$S(x)=\begin{cases} S_0(x), & x\in[x_0,x_1], \\ S_1(x), & x\in[x_1,x_2], \\ \quad\vdots & \\ S_{n-1}(x), & x\in[x_{n-1},x_n]; \end{cases} \quad S_i(x)\in C^3([x_i,x_{i+1}]) \tag{1}$$

At the same time meet the conditions:

$$\begin{cases} S(x_i)=f_i, \\ S_{i-1}(x_i)=S_i(x_i), \\ S'_{i-1}(x_i)=S'_i(x_i), \\ S''_{i-1}(x_i)=S''_i(x_i); \end{cases} \tag{2}$$

In summary, we can get complete and more accurate data. The complete data can be found in the appendix. Among them, the results of the interpolation algorithm for forest coverage are shown in the figure below:

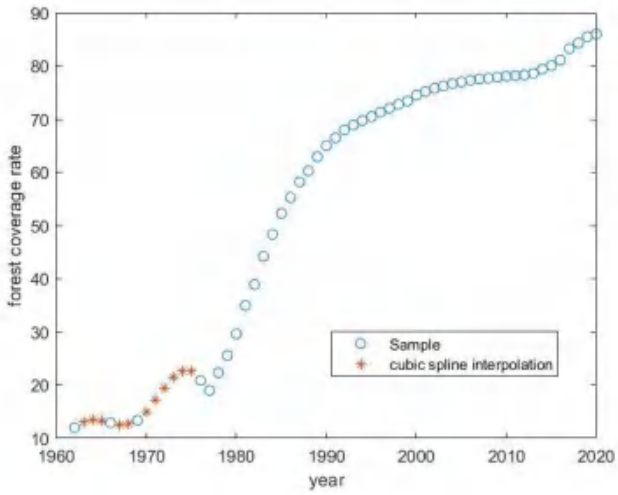


Figure 1:Forest coverage rate

5.2 Dimensionality reduction of indexes by entropy method

After obtaining the data of 10 indicators, we found that if all these indicators are used to evaluate the ecological environment of Saihanba, the complexity of the model will be slightly larger. In order to simplify the model, we can extract several of the most representative characteristic indicators from these 10 indicators as evaluation criteria.

Among them, the entropy method is a method that can objectively assign weights, which can reflect the effective information amount of the indicator. The larger the entropy value, the smaller the effective information amount of the indicator, and the smaller the weight of the indicator. Compared with the subjective weighting method, the entropy weight method is more explanatory for indicators. Therefore, we can use the entropy method to more accurately extract the most representative feature indicators among the above 10 indicators.

5.2.1 Establish and standardize the decision matrix

Firstly, We will select the years from 1962 to 2020 as the object set M_i, Multiple indicators are used as indicator set D_j. The specific data of different indicators in 59 years were selected horizontally to form the decision matrix.

Secondly, we need to divide the indicators into two categories according to their nature: benefit indicators and cost indicators. The so-called benefit index actually indicates that the larger the index is, the better; The cost index indicates that the smaller the index is, the better. Among these 10 indicators, we can divide them into the following categories:

- **Benefit indicators:**
  - **Forest coverage rate:** Forests play an important role in conserving water sources and conserving water and soil. It is easy to know that the higher the forest coverage rate, the more beneficial the impact on the terrestrial ecosystem with forests as the main body.
  - **Water quality index:** This index can reflect the quality of water with a dimensionless

value by integrating the test results of multiple water quality parameters, including ph, dissolved oxygen and other indicators, which is a reliable indicator to measure water quality.

❖ **Biological abundance index:** It reflects the abundance and poverty of species in the evaluated area. The higher the species abundance, the stronger the anti-disturbance ability of the ecosystem.

❖ **The amount of released oxygen:** This indicator can reflect the amount of oxygen released by the forest or vegetation in the evaluation area in a certain period of time. The higher the amount of oxygen released, the higher the ecological and environmental benefits.

❖ **Vegetation Index:** This index is a simple, effective and empirical measure of the vegetation status on the ground, reflecting the difference between the reflection of vegetation in the visible light and near-infrared bands and the soil background[2].

❖ **Climate index:** refers to a quantity that is composed of two or more climatic elements and represents a certain climatic characteristic. It includes aridity index, humidity index, monsoon index, and degree of land. It is mainly used for climate classification and zoning.

❖ **Ecological resilience:** also known as resilience, refers to the ability of the ecosystem to maintain the structure and pattern, that is, the ability of the system to restore its original function after being disturbed. The greater the ecological resilience, the faster the restoration of the ecosystem.

➢ **Cost indicators:**

❖ **Soil pollution index:** an index used to quantitatively describe the degree of soil pollution, the higher the degree of soil pollution, the greater the index.

❖ **Air Quality Index:** Reflects the concentration of pollutants (smoke, NO_2, SO_2, etc.) in the air. The greater the degree of air pollution, the higher the index.

❖ **Soil erosion rate:** refers to the ratio of land surface erosion and water and soil loss to the area of the evaluation area, and is a macro management indicator that reflects the overall situation of soil and water conservation.

For the benefit index, we use the following formula to normalize:

$$v_{ij} = \frac{\max\{x_{1j}, x_{2j}, \cdots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, x_{2j}, \cdots, x_{nj}\} - \min\{x_{1j}, x_{2j}, \cdots, x_{nj}\}} \tag{3}$$

For cost indicators, we use the following formula to normalize:

$$v_{ij} = \frac{x_{ij} - \min\{x_{1j}, x_{2j}, \cdots, x_{nj}\}}{\max\{x_{1j}, x_{2j}, \cdots, x_{nj}\} - \min\{x_{1j}, x_{2j}, \cdots, x_{nj}\}} \tag{4}$$

According to the above formula, we can finally get the normalized standard matrix.

5.2.2 Calculate feature weights and establish entropy values

❖ **Calculate feature weights:**

In the standard matrix, for a certain index j, the larger the value of V_{ij} is, the more useful information the index can provide to the evaluated object. We calculate the feature weight P_{ij} by the following formula:

$$P_{ij} = \frac{v_{ij}}{\sum_{i=1}^{m} v_{ij}} (0 \leqslant v_{ij} \leqslant 1, 0 \leqslant p_{ij} \leqslant 1) \tag{5}$$

❖ **Calculate the entropy value:**

The entropy value e_j of the indicator is calculated by the following formula, the closer the entropy value is to 1, the more stable the column is.

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^{m} p_{ij} \ln(p_{ij}) \tag{6}$$

❖ **Calculate the coefficient of difference:**

The greater the difference coefficient d_j is, the greater the amount of information provided by the index, and a higher index weight should be given.

$$d_j = 1 - e_j \tag{7}$$

❖ **Determine the index entropy weight:**

$$w_j = \frac{d_j}{\sum_{k=1}^{n} d_k}, j = 1, 2, \cdots, n \tag{8}$$

Through the above formula, the entropy weight can be obtained by dividing the difference coefficient calculated in the previous step by the sum of the difference coefficient of all indicators.

5.2.3 Solution of Entropy Method

We import the data into ***matlab*** to solve, sort the results of the entropy weight of many indicators, and extract the top 5 indicators with the most influence as our final indicator set according to the fault condition of the entropy value. The weights of the indicators are normalized to obtain the following weight matrix:

Table 1: Index weight matrix

Ranking	Indicators	Entropy weight
1	Forest coverage rate	0.3564
2	Air Quality Index	0.2529
3	Water quality index	0.1776
4	Soil pollution index	0.1399
5	Biological abundance index	0.0732

It can be seen that the Forest coverage rate is the most important for evaluating the quality of the environment, and its entropy weight reaches 0.35, followed by Air quality, Water quality, and Soil quality. Finally, it can be seen that the diversity of species has an impact on the evaluation of the environment. The quality of it also has a slight impact.

5.3 TOPSIS method to evaluate Saihanba's environmental impact

After obtaining the relevant indicators and data, we can use the TOPSIS method to calculate the annual environmental score of Saihanba since 1962. By comparing the environmental scores over the years, we can evaluate the impact of Saihanba on the environment.

5.3.1 Data set processing

We set the five indicators obtained in the previous step of dimensionality reduction as indicator set $D=(D_1,D_2,\cdots,D_n)$, and the situation of Saihanba from 1962 to 2020 as object set $M=(M_1,M_2,\cdots,M_m)$. The relevant data of each indicator from 1962 to 2020 is recorded as $x_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$.

Similarly, in order to eliminate the influence of dimensions, we can also normalize the data through benefit indicators and cost indicators. Through the two normalization formulas (3) (4), the standardized matrix is finally obtained.

Since the weight of each indicator is different, we can use the result obtained by the entropy weight method in 5.2.3 as the weight matrix W_j of indicator set $D=(D_1,D_2,\cdots,D_n)$. We multiply the weight matrix W with the standard matrix obtained in the previous step to obtain the weighted decision matrix:

$$\boldsymbol{R=(r_{ij})_{m\times n}\ :\quad r_{ij}=w_j\cdot v_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)} \tag{9}$$

5.3.2 Establishment of TOPSIS model

(1) Calculate positive ideal solution and negative ideal solution

First, we need to use the weighted decision matrix obtained in the previous step to first calculate the positive and negative ideal solutions. The so-called positive ideal solution is a set consisting of the maximum value of the benefit index and the minimum value of the cost index. It can also be called the best solution. The negative ideal solution is the opposite:

$$\boldsymbol{S_j^+=\begin{cases}\max\limits_{1\le i\le m}\{r_{ij}\},j=1,2,\cdots,n;\ D_j\ is\ the\ benefit\ indicator\\ \min\limits_{1\le i\le m}\{r_{ij}\},j=1,2,\cdots,n;\ D_j\ is\ the\ cost\ indicator\end{cases}} \tag{10}$$

The formula above is the positive ideal solution, and the formula below is the negative ideal solution.

$$S_j^- = \begin{cases} \min\limits_{1\le i\le m}\{r_{ij}\}, j=1,2,\cdots,n;\ D_j \text{ is the benefit indicator} \\ \max\limits_{1\le i\le m}\{r_{ij}\}, j=1,2,\cdots,n;\ D_j \text{ is the cost indicator} \end{cases} \tag{11}$$

(2) Calculate Euclidean distance

Next, we calculate the Euclidean distance between each object and the positive and negative ideal solutions by the following formula:

$$\begin{cases} Sd_i^+ = \sqrt{\sum_{j=1}^{n}(S_j^+ - r_{ij})^2}, i=1,2,\cdots,m \\ Sd_i^- = \sqrt{\sum_{j=1}^{n}(S_j^- - r_{ij})^2}, i=1,2,\cdots,m \end{cases} \tag{12}$$

(3) Calculate the relative closeness between object and the positive ideal solution

In order to explore the impact of changes in Saihanba on the ecological environment, we need to calculate the relative closeness η_i between the indicators of each year and the positive ideal solution, and evaluate the pros and cons of the ecological environment according to η_i excellent.

The expression for η_i is as follows:

$$\eta_i = \frac{Sd_i^-}{Sd_i^+ + Sd_i^-}, i=1,2,\cdots,m \tag{13}$$

5.3.3 Solution of TOPSIS model

Next, we import the sorted data, use $matlab$ programming to solve the established model, we can get the environmental score of Saihanba over the years, as shown in the following figure:

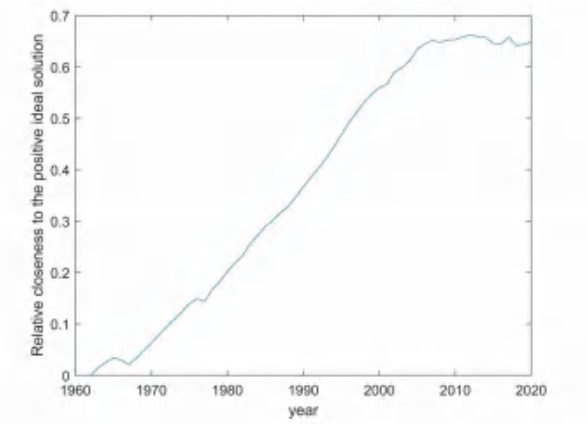


Figure 2：The relative closeness between the set of indicators and the positive ideal solution from 1962 to 2020

6 Establishment and solution of Problem Two Model

6.1 Model preprocessing

6.1.1 Handling of problems

Promblem 2 requires us to study the relationship between the restoration of Saihanba and the ability of Beijing to resist sand and dust, and establish a mathematical model that can quantitatively evaluate the impact of Saihanba on the ability of Beijing to resist sand and dust. At this time, we need to find a new set of indicators to evaluate this pair of relationships, but this is inefficient, and cumbersome data collection will increase the difficulty of building the model. In view of the fact that we have found five indicators of Saihanba's impact on the ecological environment in Promblem 1, we can try to test the correlation between Saihanba's ecological environment and Beijing's ability to resist sand and dust, if the correlation between the two is good, We can use the five indicators to evaluate the ecological environment of Saihanba to evaluate Beijing's ability to resist sand and dust.

Among them, the Spearman correlation coefficient is a method to study the convergence between two groups of variables. It does not require high sample size, and does not need to assume the normality of the population, and good results can be obtained. Therefore, here we use the Spearman rank correlation coefficient to test.

The definition of Spearman correlation coefficient is as follows:

$$\begin{cases} r_{SP} = 1 - \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \\ t = r_{SP}\sqrt{\dfrac{n-2}{1-r_{SP}^2}} \sim t(n-2) \end{cases} \tag{14}$$

Where n is the sample size and d_i is the difference between the levels of the two sets of data. When $|t| \geqslant t_{1-\alpha/2}$, the null hypothesis is rejected: there is no correlation between the two groups of variables studied, that is, there is a significant convergent relationship between the two variables.

The test results are as follows:

Table 2: The test results

Spearman rank correlation ρ	p-value
0.9680	6.4030e-36

We can see that the value of ρ is as high as 0.9680, and p-value is almost 0, indicating that the ability to resist sand and dust storms has a significant convergent relationship with the impact

of the first question on the ecological environment. We can use the five indicators in question one for subsequent modeling.

6.1.2 Data processing

In order to facilitate the establishment of the model, we need to quantify the two abstract concepts of Saihanba's annual restoration process and the ability to resist sand and dust: we use the five indicators obtained in Problem 1 (Forest coverage, Air quality index, Water quality index, Soil Pollution index and Biological abundance index) during the period of 1962-2020 to represent the annual recovery process of Saihanba; for the Anti-dust ability, we also based on the frequency and severity of sandstorms that occurred in Beijing each year from 1962 to 2020. Quantify the abstract index of Anti-sand dust ability into concrete data, and finally organize all the data into an excel table.

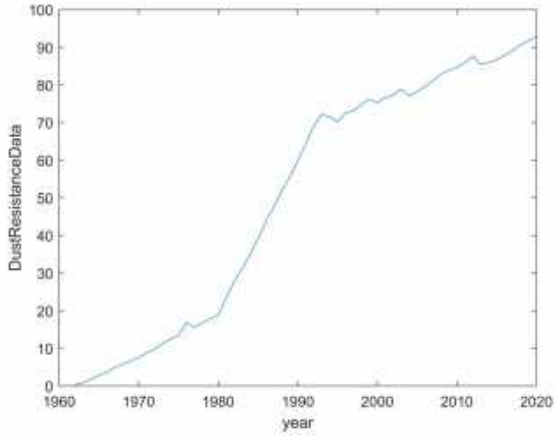


Figure 3:Variation of dust resistance in Beijing from 1962 to 2020

By observing the above figure, we found that the change trend of Beijing's Anti-dust ability index is similar to the logarithmic growth form, so we tried to use the logarithmic form of the function to establish a function that can reflect the recovery and Anti-sand dust of Saihanba Mathematical model of the relationship between abilities.

6.2 Establishment and solution of multiple regression model

We define the ability to resist sand and dust as y^D and the index set as $G(x^f, x^w, x^b, x^s, x^a)$. The variables x^f, x^w, x^b, x^s, x^a represent Forest coverage, Water quality index, Biological abundance index, Soil pollution index and Air quality index.

Assuming that the change in Beijing's ability to resist sand and dust is only related to the ecological changes in Saihanba, according to the obtained data, we take the five indicators of Forest coverage, Air quality index, Water quality index, Soil pollution index, and Biological abundance index as self Variables, take Beijing's ability to resist sand and dust as the dependent variable, and establish a multiple regression model that can reflect the relationship between

them.

As mentioned above, the change trend of the ability to resist sand and dust is very similar to the logarithmic function. We try to build a model based on this, and the form is as follows:

$$y^D = \alpha_0 + \alpha_1 lnx^f + \alpha_2 lnx^a + \alpha_3 lnx^w + \alpha_4 lnx^s + \alpha_5 lnx^b \tag{15}$$

We can set:

$$\begin{aligned} Q &= \sum_{i=1}^{5} \left(y^D - \overline{y^D}\right)^2 \\ &= \sum_{i=1}^{5} \left(y^D - (\alpha_0 + \alpha_1 lnx^f + \alpha_2 lnx^a + \alpha_3 lnx^w + \alpha_4 lnx^s + \alpha_5 lnx^b)\right) \end{aligned} \tag{16}$$

Let Q be the smallest to get an equation with the best fit, and then let:

$$\frac{\partial Q}{\partial \alpha_i} = 0 (i = 0, 1, 2, 3, 4, 5) \tag{17}$$

We can use matlab programming calculation to find the value of each coefficient:

Table 3: The value of each coefficient

parameter	α_0	α_1	α_2	α_3	α_4	α_5
estimated value	81.8011	22.027	-15.455	25.376	0.912	-30.199

Then bring in the value of each parameter to get the final mathematical model:

$$y^D = 81.8011 + 22.02761 lnx^f - 15.45556 lnx^a + 25.3765 lnx^w + 0.9127 lnx^b - 30.1999 lnx^s \tag{18}$$

Through the above model, it can be found that the coefficients before the variables lnx^f and lnx^w are as high as 22.0276 and 25.3765. Sand and dust capabilities have also been significantly improved. The coefficients of the variables lnx^a and lnx^s are -15.45556 and -30.1999, indicating that as the degree of Air pollution and Soil pollution in Saihanba decreases, Beijing's ability to resist sand and dust has increased significantly.

According to the nature of the logarithmic function, we can know that when the index set $G(x^f, x^w, x^b, x^s, x^a)$ just started to change, that is, when the Saihanba project was initially constructed, Beijing's Anti-dust ability was still relatively slow. However, in the middle of construction, the ability to resist sand and dust began to increase significantly, and the frequency and severity of sandstorms in Beijing were significantly reduced. Now, the Forest coverage rate is close to 86%, gradually reaching a state of saturation, and Beijing's Anti-dust ability has also begun to slow down and gradually maintained at a relatively good value.

6.3 Model checking

We use hypothesis testing methods for the obtained model, use the coefficient of

determination test, F test and T test to test the model, and draw a fitting effect graph to observe the fitting effect of the model.

(1) Coefficient of determination test and F test

Table 4: Coefficient of F test

Source of Variance	Degree of freedom	sum of square	Mean square	F-value	p-value
Return	5.000000	62044.03	12408.81	654.22	0.000000
Residual	53.00000	1005.27	18.97		
Total	58.000000	63049.31			

We can see that the p-value of the F test is 0, rejecting the null hypothesis: the coefficients of all independent variables are 0, and the F test is passed. Moreover, the value of the determination coefficient is as high as 0.9826, indicating that the fitting effect of the model is still relatively good.

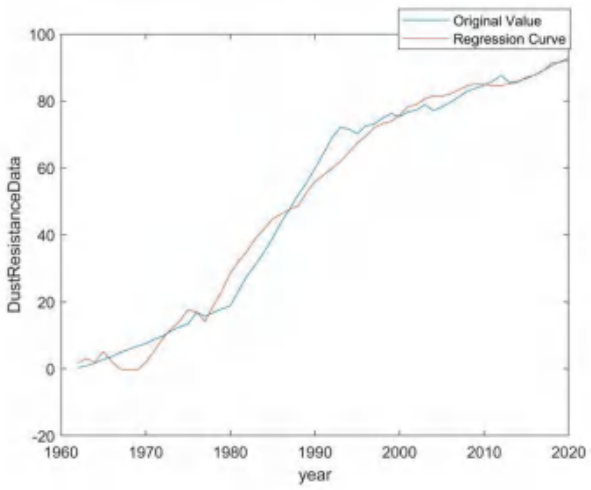
(2) T test

Table 5: Coefficient of T test

Variable	estimated value	Standard error	t-value	p-value
Constant term	81.8011	63.9919	1.2783	0.0267
x^f	22.0276	2.9625	7.4356	0.0000
x^a	-15.4556	14.3875	-1.0742	0.0876
x^w	25.3765	3.2583	7.7883	0.0000
x^b	0.9127	2.6677	0.3421	0.0336
x^s	-30.1999	3.9228	-7.6986	0.0000

From the above table, we can see that 3 variables of the model: x^f, x^w, x^s ,the p-value of the t-test is all equal to 0, and the values of the remaining variables and constant terms are also less than 0.10. The more excellent ones have passed the t-test, indicating The Linear Relationship between the independent variable and the dependent variable of the model is significant, and the model fitting effect is better.

(3) Fitting effect diagram



Figuer 4:Model fitting effect diagram of "impact of Saihan dam on anti dust capacity of Beijing from 1962 to 2020"

By observing the regression curve and the original value, we can see that around 1970, the fitted value deviates from the original value trend. As mentioned above, our original value Anti-dust ability is based on the number and severity of sandstorms that occurred in Beijing that year. We suspect that the climate of Beijing during this time period reduced the frequency of sandstorms, but the actual Anti-dust ability is not as high as the original value.

Looking at the information, we found that Beijing experienced 114 consecutive days of drought-free rain from 1970 to 1971[3]. The surface was loose and the surface was loose. When strong winds passed, it was very easy to produce sandstorms. Fortunately, there were no strong winds in Beijing during this period, which reduced the occurrence of sandstorms. The frequency of resulting in the false high of the original value. Instead, the regression curve we fitted is more able to reflect Beijing's sudden drop in Anti-dust ability at that time.

From 1981 to 1996, the regression curve was slightly higher than the original value before 1988, and slightly lower after 1988, but the fitting effect is still good, and the unconsidered climate may be one of the reasons for the error[4]. After 1996, there was almost no deviation between the regression curve and the original value.

On the whole, the fitting effect of this model is still relatively good.

7 Establishment and solution of Problem Three Model

7.1 Model building ideas

In Problem 1, we have found the impact of Saihanba on the local ecological environment through relevant indicators. In order to extend Saihanba’s ecoloehensive Ecological indicators to judge these geographic locations. The comprehensive ecological index can be determined by three ecological capacity factors, such as the ability to resist sand and dust storms, as well as the

carbon absorption capacity and the water purification capacity described in Problem 2.

The specific modeling steps are as follows:

- ✓ **Step1:** We can build a model of gical protection model to the whole country and determine which geographical locations need to build ecological zones, we can use a comprthe relationship between evaluation indicators and ecological capabilities based on the data of Saihanba over the years, and then we can use the weighted average of the three ecological ability factors to determine the comprehensive ecological indicators.
- ✓ **Step2:** Divide the geographical location into 34 administrative regions in China, collect the relevant evaluation indicators of these administrative regions, and obtain the comprehensive ecological indicators of each administrative region. Through comprehensive ecological indicators, we can judge whether it is necessary to build an ecological zone in this area, and if it needs to be built, go to the next step.
- ✓ **Step3:** Establish a relationship between the comprehensive ecological indicators and the scale of the ecological zone through an equation to determine the proportion of the ecological zone area that needs to be increased in the administrative area. Then, we compare it with the maximum ecological area that can be planned in the administrative area. If the area that can be planned is greater than the area that needs to be increased, the actual area that can be increased is the area that needs to be increased. Otherwise, the ecological area will be based on the largest area. The area can be planned for construction.
- ✓ **Step4:** Assuming that the ecological zone is successfully completed in accordance with the construction plan, we can recalculate the forest coverage of the area and bring it into the previously established model to solve the carbon absorption capacity of the area after the ecological zone is constructed, and compare it with The carbon absorption capacity of the ecological zone was compared before the construction, and the conclusion of the ecological zone's impact on carbon neutrality was reached.

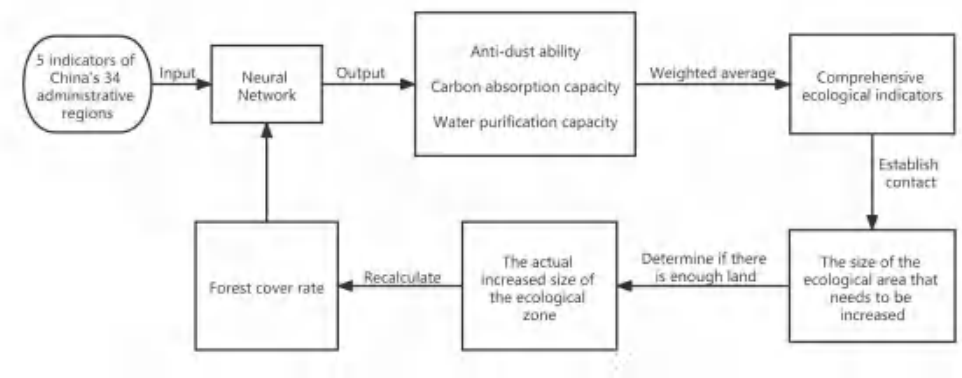


Figure 5:Step flow chart

7.2 The establishment of BP Neural Network

In Step 1, we need to establish a model of the relationship between evaluation indicators and ecological capabilities. At this time, the evaluation indicators can be the five factors in question 1: Forest coverage rate, Air quality index, Water quality index, Soil pollution index and Biological abundance Index; and the Ecological ability can be the ability to resist sand and dust storms, the ability to absorb carbon, and the ability to purify water. Therefore, we need to find out the mapping relationship between these two pairs of indicators, namely:

$$f: (x_1, x_2, x_3, x_4, x_5) \rightarrow (y_1, y_2, y_3) \tag{19}$$

This is a mapping relationship from 5 dimensions to 3 dimensions. We cannot simply use multiple regression to establish a suitable mathematical model. In the field of artificial intelligence, deep learning solves this problem well. Therefore, we decided to use BP Neural Network to achieve this mapping relationship.

7.2.1 Neural network design

Since the data we collect is non-linear, and in neural networks, if the data needs non-linear separation, there must be a hidden layer. Therefore, we can divide the architecture of the neural network into an input layer, a hidden layer, and an output layer. The input layer is mainly used to obtain input information; the hidden layer is mainly used for feature extraction, so that the neurons in the hidden layer react to a certain pattern; the output layer is used to receive the hidden layer and output the model results, which can be adjusted by adjusting the weights. Neuron stimulation forms a more superior response, in which the excitability of the output is the result.

In the neural network model we built, the input layer is forest coverage, air quality index, water quality index, soil pollution index and biological abundance index, and the output layer is the three indicators of sandstorm resistance, carbon absorption and water purification. For the hidden layer, the deeper the number of network layers, the stronger the ability to fit the function theoretically, and the better the effect, but in fact, the deeper the number of layers may also cause over-fitting problems and increase the difficulty of training, which makes the model difficult to converge[5]. Therefore, when determining the number of network layers, we only give priority to the case of one hidden layer. If the result of the model training is not ideal, we will appropriately add the number of hidden layers.

The number of neurons in the hidden layer is also a very particular variable. If too few neurons are used in the hidden layer, the model will underfit, and too many neurons may also lead to overfitting. Therefore, choosing an appropriate number of hidden layer neurons is crucial. However, we do not have an accurate way to determine this value. We can only rely on empirical formulas to determine the initial value of the number of neurons, and finally dynamically adjust

the number of neurons according to the fitting results. In this model, the empirical model we propose is:

$$Nh = \frac{2}{3} N_i + No + \alpha \tag{20}$$

Among them, nh represents the number of neurons in the hidden layer, ni represents the number of neurons in the input layer, no represents the number of neurons in the output layer, and α can be any integer, generally starting from 0 and going to both ends. Since in this model, there are 5 neurons in the input layer and 3 neurons in the output layer, we can initially set the number of neurons in the hidden layer to 7.

In addition, when solving nonlinear problems, we need to select an appropriate activation function to introduce nonlinear features into the network. If we don't use the activation function, then the output signal will be just a simple linear function, and its ability to learn more complex function mapping from the data is even smaller. In addition, since the number of network layers in this model is not very deep, we can use the traditional tansig activation function.

Finally, the Neural network model we get is as follows:

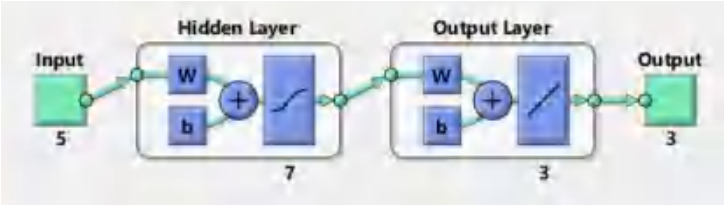


Figure 6:Neural network model

7.2.2 Neural network solution

We import the collected data into $matlab$, use 75% of the index data for training, 15% of the index data for verification, and 15% of the index data for testing. As the number of iterations increases, the model gradually tends to converge. When the Neural network is iterated to 52 times, the optimal value of the Neural network parameters can be found and output as the result.

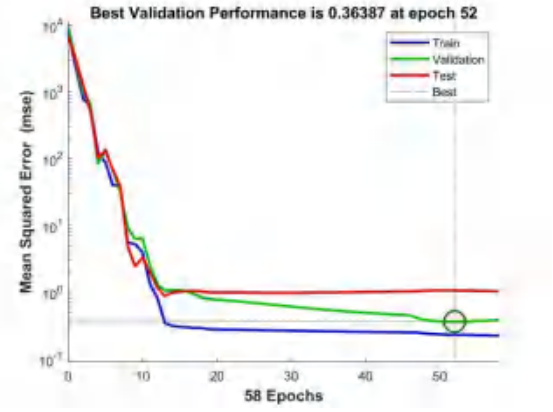


Figure 7:Index data training results

Next, we test the results of the model training, we can get the fitting effect of all the data sets, and find that the R value of the model is very close to 1, and we can think that the fitting effect of the neural network model is perfect.

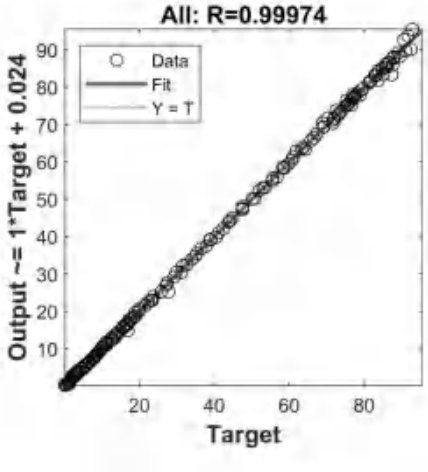


Figure 8:Fitting effect of all data sets

Therefore, we can bring the index data of 34 administrative regions into the neural network model for prediction, and we can get the anti-sand storm ability, carbon absorption ability and water purification ability of these administrative regions. Part of the data is shown in the following figure:

7.3 The location where the ecoregion was established

In order to explore whether these 34 administrative regions need to build ecological zones, we need to have an exact judgment basis, so we can set a comprehensive ecological index to make judgments.

In the previous section, we have calculated the three ecological abilities of 34 administrative regions, and the comprehensive Ecological index can be determined by the three ecological abilities. The easiest way is to take the average of the three. . However, since these three indicators have slightly different assessments of the local ecological environment, we must distinguish the importance of these three indicators.

We can assign a specific weight to various ecological abilities according to the establishment of existing ecological regions in China. By collecting relevant data, we found that the most important capacity for the establishment of the ecological zone is the capacity to absorb carbon, followed by Anti-dust ability and Water purification capacity. We can assign values to these indicators:

Table 6:Indicator assignment

	Anti-dust ability	Carbon absorption capacity	Water purification capacity

Weight	0.3	0.45	0.25

Therefore, the relationship between comprehensive ecological indicators and ecological capabilities is:

$$\beta = 0.3Y_1 + 0.45Y_2 + 0.25Y_3 \tag{21}$$

Through this relationship, we can find the comprehensive ecological indicators of 34 administrative regions, as shown in the following figure and table:

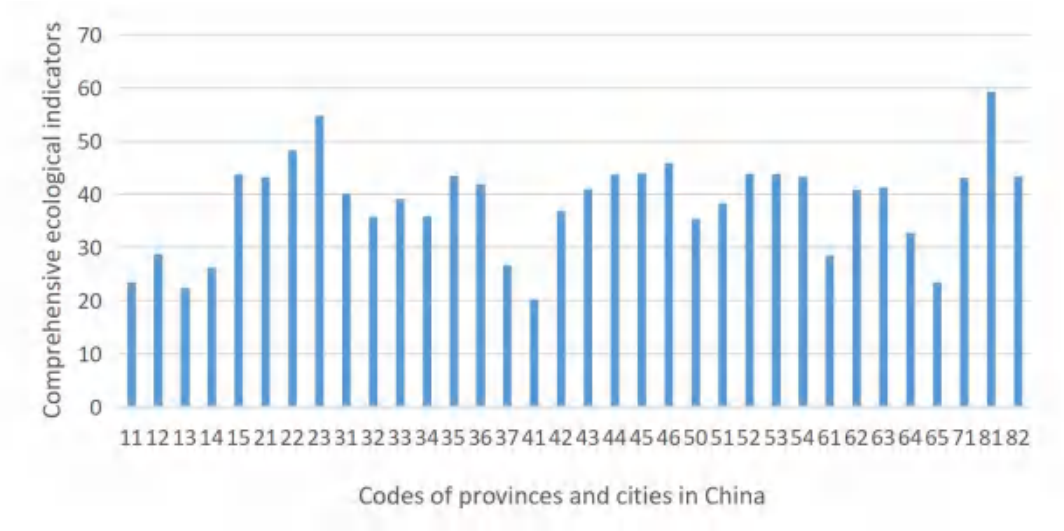


Figure 9:The comprehensive ecological indicators of 34 administrative regions

When the comprehensive ecological index is greater than 50, we can think that there is no need to build an ecological zone in the area, and when the comprehensive ecological index is less than 50, we think that the area needs to build a new ecological zone. We set the variable w as whether a new ecological area needs to be built in the area. A value of 1 indicates that it needs to be built, and a value of 0 indicates that it does not need to be built. Then the mathematical model can be expressed as:

$$w_i = \begin{cases} 1, & \beta_i \leqslant 50 \\ 0, & \beta_i > 50 \end{cases} \tag{22}$$

Therefore, it can be found from the above formula that there are 32 administrative regions in China that need to build ecological zones.

7.4 The scale of the establishment of the ecological zone

Next, we need to determine the scale of the ecological zone they need to build through the comprehensive ecological indicators of each administrative region. However, due to the inconsistency of the size of the various administrative regions, in order to facilitate the identification of the connection between the two, we can use the proportion of the ecological area that the administrative region needs to increase as a measurement standard.

Since in the initial stage of establishing an ecological zone from scratch, it is necessary to build a larger scale of ecological zone to have an impact on comprehensive ecological indicators. Let S_i be the proportion of ecological area that needs to be increased, and the function of S_i with respect to β_i is in the defined domain [0,50] Is continuous and derivable, that is, the constraint condition can be expressed as:

$$\frac{dS}{d\beta_0} > \frac{dS}{d\beta_1}(\beta_0 < \beta_1) \tag{23}$$

We can think that the equation of S_i with respect to β_i is a concave function in the domain of definition, and try to establish the connection between the two using an exponential equation. The preliminary model is as follows, where a and b are undetermined coefficients:

$$S_i = b(a^{-\beta_i}) \tag{24}$$

In order to determine the two undetermined coefficients a and b, we can perform the following analysis: When the comprehensive ecological index is 0, it means that the ecological environment of the area is extremely bad. By consulting relevant information, we can see that we will need at least 40% of the area to build an ecological zone in order to significantly improve the local ecological environment[6]. Only by building an ecological zone can the local ecological environment be significantly improved. At the same time, we believe that once the ecological zone is determined to be established, it is more cost-effective to build at least 5% of the area. Therefore, when the comprehensive ecological index is 80, we also believe that at least 5% of the area is needed to build an ecological zone.

That is, when $\beta_i = 0, S_i = 40$ and $\beta_i = 50, S_i = 5$. Substituting into (x) formula can be solved to get $a = 1.026, b = 40$, namely:

$$S_i = 40(1.042^{-\beta_i}) \tag{25}$$

In addition, we can also extend the domain of the model to [0,100], which is:

$$S_i = 40w_i(1.042^{-\beta_i}) \tag{26}$$

At this point, we can calculate the scale of the ecological zone that needs to be built in each administrative region, as shown in the following figure:

Table 7: The scale of the ecological zone that needs to be built in each administrative region

	Beijing	Tianjin	Hebei	Shanxi	Liaoning	Jilin	Heilongjiang
Anti-dust ability	45.39	45.85	35.68	45.34	50.65	60.95	66.54
Carbon absorption capacity	13.10	20.22	15.86	15.61	38.86	41.33	48.14

Water purification capacity	15.96	23.75	18.10	22.43	42.07	45.38	52.72

However, each administrative area is not completely consistent with regard to the free land planning. For example, areas such as Beijing, Shanghai, and Hong Kong have higher economic levels and denser populations. The remaining free land in these areas is less, and the proportion of land used for economic development and people's livelihood Larger, resulting in less land used to build ecological zones in these areas. However, through calculations, it is known that the scale of the ecological zone that needs to be built in these areas is relatively large, and the construction plan of the area cannot meet the optimization of the ecological environment of the area. At this time, we can only appropriately reduce the construction scale of the ecological zone.

We set the actual increaseable ecological area ratio of this place as Q_i and the planned ecological area ratio of this place as A_i, then the model can be updated as follows:

$$Q_i = \begin{cases} 40w_i(1.026^{-\beta_i}), & S_i < A_i \\ A_i, & S_i \geq A_i \end{cases} \tag{27}$$

In the end, we can get the actual scale of ecological zones that can be built in each administrative region, which can be represented by the following figure:

Figure 10:The percentage of increase in the actual ecological area that can be built in each administrative region

7.5 Assess changes in carbon neutrality

Assuming that the ecological zone is successfully completed in accordance with the

construction plan, we can recalculate the forest coverage of the area and bring it back into the neural network model to calculate the carbon absorption capacity of the area after the construction of the ecological zone. We compare it with the carbon absorption capacity of the ecological zone before construction, as shown in the figure below:

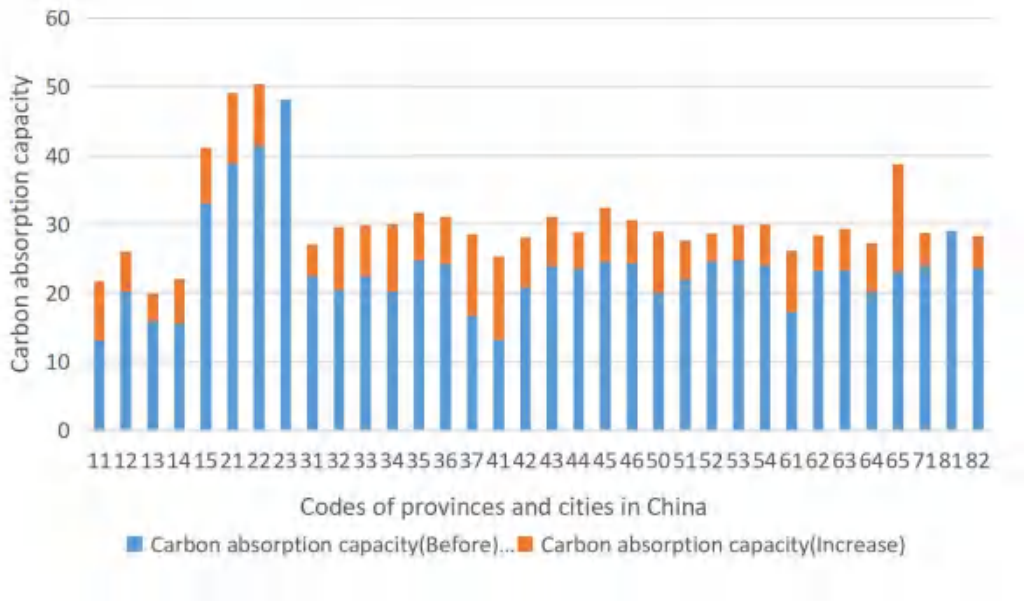


Figure 11:Comparison of carbon absorption capacity

It can be found that the carbon absorption capacity of each administrative region has a clear upward trend. We averaged the difference before and after the establishment of the ecological zone. As a result, the overall carbon absorption capacity increased by 7.17, which shows that the construction of the ecological zone will help China achieve carbon The neutralization goal has a better promoting effect.

8 Establishment and solution of Problem Four Model

Problem 4 requires us to analyze a country in the Asia-Pacific region, discuss which geographical locations of the country need to build an ecological zone and the scale of the ecological zone, and at the same time conduct an assessment of the carbon absorption capacity of the ecological zone. This is actually an extension of the model in question three, and we can use the model established in question three to solve it.

For countries in the Asia-Pacific region, we find that most parts of Australia are perennially dry, and a large area of land is desert. Therefore, the Australian government needs to build more ecological zones to enhance the territory's ability to resist sand and dust, so this can become Our ecoregion model is a more suitable research object.

We divide Australia into 7 geographical locations, namely Queensland, New South Wales, Victoria, Tasmania, Northern Territory, South Australia and Western Australia. By collecting the

relevant index data of these seven geographic locations, and then bringing them into the neural network model for prediction, the anti-sand storm capacity, carbon absorption capacity and water purification capacity of these administrative regions can be obtained respectively.

In Australia, the important basis for establishing an ecological zone is its ability to resist sand and dust, we can slightly adjust the relationship (21) in the model to increase the weight of its ability to resist sand and dust storms, and obtain the following formula:

$$\beta = 0.45Y_1 + 0.35Y_2 + 0.2Y_3 \tag{28}$$

Then, we can calculate the comprehensive ecological indicators of each geographic location through the relational formula, as shown in the following table:

Table 8:Comprehensive ecological indicators for each geographic location

	Anti-dust ability	Carbon absorption capacity	Water purification capacity	Comprehensive ecological indicators
Queensland	67.97	24.10	33.30	45.68
New South Wales	72.38	25.19	35.14	48.41
Victoria	70.45	24.45	34.17	47.09
Tasmania	82.48	25.70	36.97	53.51
Northern Territory	53.45	25.62	31.79	39.38
South Australia	30.58	14.80	17.10	22.36
Western Australia	47.50	19.51	23.59	32.92

We can calculate the actual scale of the ecological area that can be built in each geographic location according to the size of the ecological area that can be planned for each geographic location in Australia and the formula (27), as shown in the following figure:

Figure 12:The percentage of increase in the actual ecological area that can be built in each administrative region

Assuming that the ecological zone is successfully completed in accordance with the construction plan, we can recalculate the forest coverage of the area and bring it back into the Neural Network model to calculate the carbon absorption capacity of the area after the

construction of the ecological zone. We compare it with the carbon absorption capacity of the ecological zone before construction, as shown in the figure below:

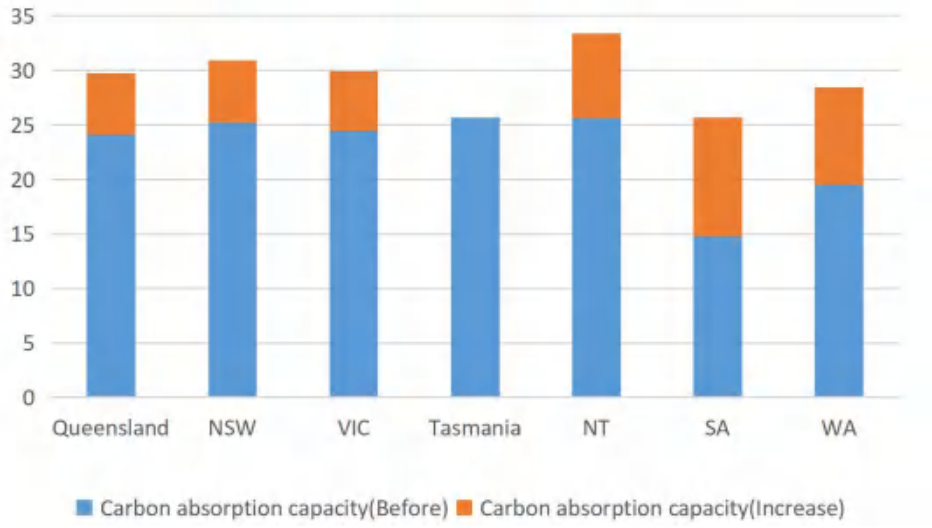


Figure 13:Comparison of carbon absorption capacity

9 Evaluation, Improvement and promotion of the Model

9.1 Evaluation of the Model

- **Advantage:**

(1) We have found the relationship between Saihanba's ecological environment and Beijing's ability to resist sand and dust through correlation research. There is no need to continue to collect complex data for research, which simplifies the difficulty of modeling.

(2) We use artificial intelligence algorithm neural network to analyze ecological capabilities. Compared with traditional regression models, the results we get will be more convincing.

(3) In the fourth question, Australia, a country with serious desertification, was selected as the research object, which has good representative significance and analytical value

- **Shortcoming:**

(1) In reality, there are many indicators for correctly evaluating an ecological region, and our model only uses a few more important indicators, which may not fully reflect the characteristics of the ecological region.

(2) When evaluating the ability to resist sand and dust, we judge by studying the frequency and severity of the occurrence of sand and dust storms. It is inevitable that there will be a bit of subjective evaluation, and the evaluation that may be obtained is not completely correct.

9.2 Improvement and promotion of the Model

(1) We can collect more indicators and data, and build a more complete and more complex mathematical model for more accurate evaluation of ecological indicators in various regions.

(2) We can apply the Saihanba ecological model to the whole country or even the whole world through the establishment of the Saihanba ecological model, and draw out the scale of the ecological zone that needs to be built according to the environment of each region.

10 References

[1] Study on the value of forest ecosystem service function in Saihanba Nature Reserve[D].Zhang Yunling.Hebei Normal University

[2] Design and implementation of forest resources data visualization system in Saihanba Forest Farm[D].Zhu Qianyu.Beijing Forestry University

[3] Dust weather in Beijing in recent 50 years and Its Control Countermeasures.Chen Guangting[J].Institute of environment and Engineering in cold and arid regions, Chinese Academy of Sciences

[4] History of severe dust storms and changes of surrounding ecological environment in Beijing[J].Chen Guangting.Institute of environment and Engineering in cold and arid regions, Chinese Academy of Sciences

[5] Liu lianglin, Wang Quanfeng, Lin Huangbin, parameter setting and application of BP neural network[J], infrastructure optimization, Vol. 28, No. 2, April 2007

[6] Quantitative discussion on the best protection benefit and forest coverage[D].Zhang Jian, Gong Yuanbo, Chen Linwu.College of forestry, Sichuan Agricultural University

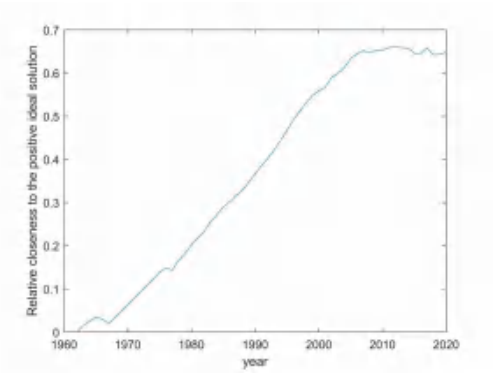
Report:

The construction of ecological civilization plays a very important role in the sustainable development of the world. China has always placed the construction of ecological civilization in a prominent position in the overall work. The transformation of Saihanba forest farm from desert to oasis in China is an excellent case of improving the local ecological environment by establishing an ecological reserve.

Figure:Comparison of Saihan dam in the past and present

Before **Now**

In order to study the effect of the establishment of Saihanba ecological area, we extract the indexes that can significantly represent the ecological environment of Saihanba by Entropy method: forest coverage rate, air quality index, water quality index, soil pollution index and biological abundance index. We can establish TOPSIS evaluation model of Saihan dam's impact on ecological environment through relevant data, and get the score of Saihan dam's ecological environment from 1962 to 2020. The higher the value, the better the local ecological environment.



We can see that when Saihanba forest farm has not been established, the score of the area is very low, indicating that the environment is very bad. However, in the following 60 years, Saihanba people jumped the ecological score to 0.65 through their hard-working hands, indicating that the local environment has improved greatly.

Then, in order to study the impact of Saihan dam on the anti dust ability of Beijing, we established a multiple regression model:

$$y^D = \alpha_0 + \alpha_1 lnx^f + \alpha_2 lnx^a + \alpha_3 lnx^w + \alpha_4 lnx^s + \alpha_5 lnx^b$$

According to the model, with the increase of Saihanba forest coverage and water quality index and the decrease of air pollution and soil pollution, the anti dust ability of Beijing shows an obvious upward trend. Therefore, we can think that the establishment of Saihanba ecological area has made a significant contribution to Beijing's anti dust ability.

Then, in order to extend the ecological protection model of Saihan dam to China, we calculate the three ecological indicators of anti sandstorm capacity, carbon absorption capacity and water purification capacity through BP neural network, and introduce comprehensive ecological indicators to judge whether it is necessary to establish an ecological area. Then, the proportion of ecological area to be increased in an administrative region is calculated according to the size of comprehensive ecological indicators, and the actual proportion of ecological area to be increased is obtained according to the idle land of each province.

At the same time, we can also extend the model to other countries in the Asia Pacific region, and can also calculate which geographical locations of these countries need to increase the ecological area. We take Australia, which is dry all the year round and has a wide desert area, as an example for analysis, as shown in the figure below:

The darker the color in the figure, the larger the proportion of ecological area actually needs to be increased in this area. If all regions are constructed according to this plan, we recalculate the forest coverage after the implementation of the scheme and bring it into the neural network model. It is found that the carbon absorption capacity of all regions has been significantly improved, which also shows that the construction of ecological areas is of great significance to the realization of carbon neutrality and, to a certain extent, also proves the popularization of our model.

To sum up, we suggest that each region can reasonably expand the ecological area according to these results without sacrificing our valuable ecological environment for rapid economic development, which is our bounden responsibility. The establishment of ecological area is to provide a more comfortable environment for our human habitat and add more beauty to our home.

Team Number :	apmcm2208815
Problem Chosen :	A

2022 APMCM summary sheet

Feature Extraction of Sequence Images and Modeling Analysis of Mold Flux Melting and Crystallization

By establishing a pixel discrimination model, the problem of identifying temperature data in images is solved. By establishing a feature-based equation of state model, the problem of quantifying image features and establishing the process curve of mold melting and crystallization is solved. By establishing a polynomial regression model, the problems of temperature and time change and the functional relationship between mold flux melting and crystallization process are solved, and the reliability of the results is verified by combining the JMA formula in crystallization kinetics.

For problem one, we establishes a pixel discrimination model, extracts the temperature in the upper left corner of each picture, and makes temperature time curves of 1# and 2# respectively. Firstly, the picture is grayscale processed and the global threshold is selected to binarize the picture, and the image is represented by a 0-1 pixel matrix; Then, each preprocessed picture is cut into 8 parts according to the characteristics of the digital font and position of the picture, and then the recognition rate is improved by filtering and noise reduction processing. Then, 0-9 numbers and °C are selected as templates, and by calculating the correlation coefficient between the image and the template, the template with the largest value is selected as the recognition result, and automatically stored in the table; Finally, the convolutional neural network is used to test the results, and the results are reliable.

For problem two, we establish a feature-based equation of state model, study the dynamic differences in the melting and crystallization processes, and discuss the process curves of melting and crystallization for different characteristics. First, the image is preprocessed to reduce the pixel size to 300*400 size; Then, the "Canny" operator is used to detect the edge, the edge contour image is obtained, and the find function is used to obtain the index vector for the contour edge image to separate the mold from the background. Then, the Tamura texture feature algorithm is used to extract the roughness, contrast, directionality and linearity, observe the change law of the mold in the process of melting and crystallization, and find its recursive formula about time. Finally, the average gray value and roughness are visualized, and the polynomial fitting is realized, and the melting and crystallization process curves after fitting are obtained. The results are tested and analyzed by the five features of roughness, contrast, directionality, linearity and average gray value, which proves that the results are consistent with the experimental phenomena and verifies the reliability of the feature-based equation of state model.

For problem three, a polynomial regression model is developed in this paper to solve the problem of establishing the temperature as a function of time and the

melting and crystallisation of the mould melt. Firstly, a polynomial regression is fitted to the temperature at different moments from Problem 1 to establish a segmental function expression. Then, based on the process curves for mould melting and crystallisation found in Problem 2, polynomial regression expressions were created for the amount of melting versus temperature and the amount of crystallisation versus temperature at the same moments. The crystallisation rate is then related to time, and the melting rate is related to time. Finally, in conjunction with the crystallisation kinetics, it was found that the graph of crystallisation rate versus temperature found in this paper was in high agreement with the image of the JMA equation, proving the accuracy and reliability of the results obtained.

Keywords: Convolutional neural networks; Equation of state; Tamura algorithm; Polynomial regression model

Contents

1 Introduction

1.1 Problem Background

The phase distribution of the mould flux in the gap between the mould wall and the strand shell is of great importance for continuous casting. However, high temperatures, transient flows, the complexity of phase changes and chemical reactions and the opacity of the mould wall make it difficult to observe the phase changes of the mould flux directly in this paper. Furthermore the data obtained by the SHTT II tester at melting and crystallisation temperatures needs to be processed by a large number of experimenters before it can guide the design of the mould co-solvent. For this reason, there is an urgent need to develop automatic feature extraction and mathematical modelling techniques for column sequential images.

Melting process: A flux for continuous casting moulds is added to the top of the molten steel in the mould. The solid slag that accumulates on the surface of the molten steel prevents horizontal crusting of the molten steel due to the temperature of the molten steel dropping too quickly. The mould flux is then melted when the temperature of the mould flux reaches its melting point and a sintered layer is formed. The raw material for the flux starts as a low melting point substance and is melted to form a liquid slag, the composition of which changes during the process.

Crystallisation process: During the melting process, a liquid slag layer is produced and covers the surface of the molten steel. The liquid slag may penetrate from the steel surface into the gap between the shell and the copper mould wall, thus forming a slag film. Since the temperature of the liquid slag is positively correlated with the temperature of the steel strand surface, if the mould is forced to cool, the slag film against the copper mould wall will quench and solidify, forming a glassy solid slag film (solidification behaviour of the slag film), while the slag film will, under certain conditions, form a crystalline layer in certain areas (crystallisation behaviour of the slag film), thus constituting a three-layer slag film structure: glassy layer, crystalline layer and liquid slag layer.1.2 Restatement of the Problem

1.2 Problem requirements

(1) Question 1: Using image segmentation and recognition or other techniques, extract the corresponding data for each image and make a temperature time profile. one set of test results in line 1# or 2# is inaccurate and needs to be pointed out and explained.

(2) Question 2: Based on the images in Figure 1, use digital image processing techniques to investigate and quantify the dynamic differences between adjacent sequences of images during mould flow melting and crystallisation. Using this as a basis for time series modelling of the quantified different features, discuss the process profiles of mould melting and crystallisation based on the simulation results obtained from the mathematical model.

(3) Question 3: Based on the results of the studies in questions 1 and 2, develop a mathematical model to discuss the variation of temperature and time as well as the melting and crystallisation processes of the mould melt as a function of temperature. The relationship between the temperature, melting rate and crystallisation rate of the mould melt is then discussed in the light of these results.

2 Problem analysis

Three problems are proposed for the complex phase change and chemical reaction of the mold flux under high temperature heating in the continuous casting process, which makes it difficult to directly observe the phase of the mold flux. First, in order to reduce manpower, it is necessary to develop digital recognition of sequence images and record temperature data; Then, the sequence image is extracted by features, and the model is established through the features to study the process curve of mold melting and crystallization. Finally, the extracted temperature data and the process curve of mold melting and crystallization are deduced to derive the functional relationship expression, and the functional relationship expression of temperature, melting rate and crystallization rate is determined. The three questions, from shallow to deep, are intertwined.

2.1 Analysis of Question One

In the analysis of the images, the numbers to be identified in each of the 562 images are located in the top left corner of the image and the spacing between the data font and the numbers is fixed. The cut will result in a smaller image size and fewer pixel points. It is envisaged that the image will be enlarged by a factor of 10 in preparation for the subsequent operation. Before cutting, this paper envisages grey-scale processing and binarisation of the image, simplifying the image into black and white bicolour i.e. two values of 0,1, splitting out the numbers in the image and then cutting the numbers into 158*93 pixels in size, and setting the size of the numbers to 28*28 pixels by scaling uniformly; this paper notes that the edges of the numbers may not be smooth, and in order to make the edges of the numbers in the image smoother In this paper, it is noted that the edges of the figures may not be smooth, so in order to make the edges of the figures in the image smoother, the image is filtered and noise reduced; this paper envisages cutting out 8 figures per image, but upon observation the cut images are not all figures, this paper envisages identifying the non-numbered °C as the number 10, and then carrying out uniform processing. The correlation coefficient is calculated between the cut image data and the designed standard data. According to the prediction, the larger the correlation coefficient, the higher the similarity between the image data and the standard data. Based on this principle, this paper envisages the establishment of a pixel discrimination model, which will be used to perform a circular traversal to identify the temperature of each image in comparison with the standard sample, exported and input into a table, and

thus make a temperature time profile. The paper envisages a comparative analysis of this model and the use of convolutional neural networks to build a digital image recognition model. If the conclusions reached are not significantly different from those of the pixel discrimination model, then the pixel discrimination model is reasonable and the test results are accurate.

2.2 Analysis of Question Two

Problem 2 is a feature extraction type of problem. In this paper, Tamura texture features are used to extract four features: roughness, contrast, directionality and linearity. To make the computation more sensitive, the image is pre-processed with a greyscale process that turns the image into only black and white, and then scales the image to a uniform size. This paper envisages the use of the "Canny" operator to detect edges, but the edge operator has the limitation that it only responds to edges and finds all edges, while non-edges should be discarded. Therefore, in order to make the edge operator results more accurate, this paper envisages a series of image processing: first a Gaussian smoothing filter and convolution of the image to make the image edges smooth; then a non-maximum suppression of the amplitude of the gradient to make the blurred boundary clear; and then a double thresholding technique to further eliminate the noise. The edges are tracked in the image obtained after processing to obtain an edge profile image and to separate the mould from the background. This paper envisages the use of Tamura texture feature extraction followed by observing the changes in the four features during the crystallisation and melting process and finding out their change patterns, listing the feature equations and plotting the results obtained after the calculation into a graph to make the results more intuitive.

2.3 Analysis of Question Three

Problem 3 belongs to the functional relationship derivation class problem. Based on the temperature change data in question 1 and the process curve of mold melting and crystallization in question 2, the functional relationship between them is derived. For the accuracy of the function expression, we chose to use polynomial regression fitting to make the discrete data more relevant to the function image. The function relationship between mold flux melting and crystallization and temperature can be derived by establishing the function relationship between temperature and time and mold flux melting and crystallization and time. The functional relationship is visualized and combined with the crystallization kinetic formula to test the accuracy of the establishment of the functional relationship. If the images are similar, the results are reliable.

3 Model assumptions

(1) Assume that the text in the upper left corner of the image is the top frame.

(2) Assume that the moment of image recording is accurate, and the error is not considered.
(3) Assuming that the mold flux is normal, there is no problem with the use process.

4 Symbol description

Symbol	Description
A_i	A matrix of pixels of 28*28 size for the i number
r_{ij}	The 0-1 pixel matrix of the i digital image has a correlation coefficient with the matrix of the j template
$\max r_{ij}$	The 0-1 pixel matrix of the i digital image is most similar to the matrix of the j template
T_{1t}	Thermocouple No. 1 temperature
T_{2t}	Thermocouple No. 2 temperature
$C_k(x,y)$	The average gray value of the cell
$g(i,j)$	The grayscale value of the pixel at (i,j)
$\Delta G(x,y)$	The gradient vector at each pixel
$N_\theta(i)$	The total number of edge pixels on the corresponding direction angle θ
n_p	The number of histogram peaks
ω_p	Peak quantization range
P_a	The distance point of the $m \times m$ locally oriented symbiotic matrix
b_t	The melting amount of the mold at time t
c_t	The roughness of the mold at time t
I	The number of grains that can be formed per unit volume per unit of time
V^s	The volume of liquid crystallized
V^t	The remaining volume of uncrystallized liquid
a_i	The amount of mold crystallization at the i moment
b_i	The amount of mold melt at the i moment
m_i	Crystallization rate at moment i

n_i	Melting rate at moment i

5 Model building and solutions

5.1 Model buiding and solving Problem One

5.1.1 Image pre-processing

In this paper, the images are first greyed out as well as binarised, and then segmented based on the fact that the data is in the same position.

Step1: The image in Annex 1 was greyed out to give 1316*1792 pixel points, each with a value of 0-255 for radians, 255 for white and 0 for black. Next, the image is binarised. The image is made up of a matrix, each point in the matrix has a different RGB value, which results in a different colour, and the final overall presentation to this paper is a coloured image. The binarisation process changes the range of colours from 256 to 2, thus increasing the speed of computation. A global threshold is then chosen to divide the image into a black or white binary image, with a grey value greater than or equal to the threshold assigned to 1 and vice versa, so that each pixel is represented by 0 or 1, with 0 being black and 1 being white.

Step2: Based on the size of each image, the position of the top left corner data and the fixed font and spacing of the data, this paper sets a fixed cut line to cut the top left corner data of each image. 65*110 pixel points are cut out from the top left corner and the image is enlarged by a factor of 10 in preparation for the subsequent operation.

5.1.2 Pixel discriminant modeling

The model building process in this paper is represented in a flow chart, which is shown below.

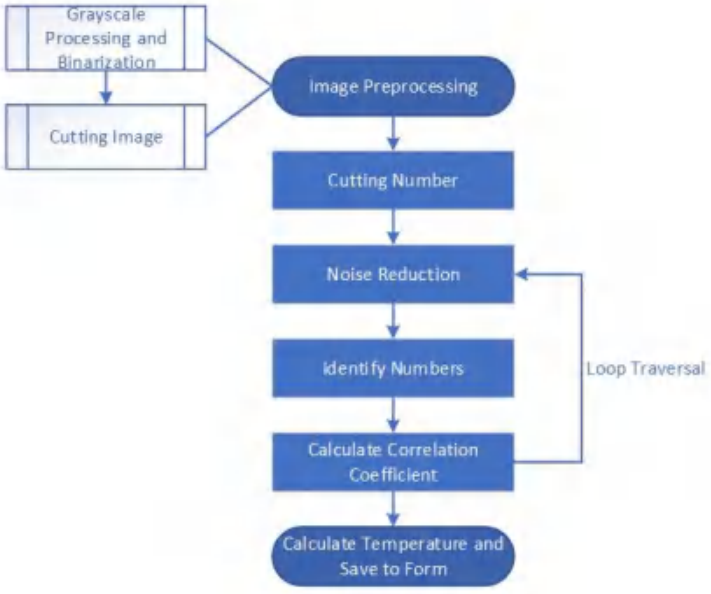


Figure1 Model building flow chart

This article uses a pixel point as a unit and sets the coordinates of the first pixel in the upper left corner to $(1,1)$ to position the number in the image. Since the temperature values of thermocouple 1 and thermocouple 2 are either 3 or 4 digits, the paper is cut at 4 digit size spacing and in pixels, the diagonal coordinates of the size spacing of thermocouple 1 temperature are $(501,280)$ to $(873,366)$ and similarly the diagonal coordinates of thermocouple 2 temperature are $(501,393)$ to $(873,551)$. The difference between the horizontal coordinates of the four figures is 372 pixel points, i.e. each figure is 92 pixel points wide. The vertical coordinate difference is 158 pixel points, i.e. each digit is 158 pixel points high. By this criterion, each digit is cut into an image of 158*93 pixel size, and by scaling, the size of each digit is set to 28*28 pixel points.

Next, the image is filtered for noise reduction, which corrodes the image and then expands it to remove small deviating pixel points. In this paper, the borders of larger objects are smoothed without changing their area, giving the figures in the image smoother edges.

Then, among these digital images, a total of 11 images from 0-9 as well as °C were identified as templates. The correlation coefficient was calculated for each digital image in turn with the 11 templates through a 0-1 matrix of size 28*28, and the correlation coefficient was calculated as follows.

$$r_{ij}=\frac{\sum_m\sum_n\left(A_i-\overline{A}_i\right)\left(B_j-\overline{B}_j\right)}{\sqrt{\left(\sum_m\sum_n\left(A_i-\overline{A}_i\right)\right)^2\left(\sum_m\sum_n\left(B_j-\overline{B}_j\right)\right)^2}} \tag{1}$$

$$\left(i=1,2,\cdots\cdots,8;\ j=1,2,\cdots\cdots,11\right)$$

Where denotes the 0-1 pixel matrix of size 28*28 for the i th number.When $i=1,2,3,4$， the 4 digits of thermocouple No.1 from left to right are depicted， when $i=5,6,7,8$， Respectively, the No.2 thermocouple is represented by 4 digits from left to right. B_j represents a matrix of 0-1 pixels of 28*28 size of the j template。When $j=1,2,\cdots\cdots,10$， A template representing the number "$j-1$"， For example, B_1 represents a template with the number "0". r_{ij} represents the correlation coefficient between the 0-1 pixel matrix of the i digital image and the matrix of the j template，Then $\max r_{ij}$ represents the 0-1 pixel matrix of the i digital image and the matrix of the j template.The i digit in the image at this moment is recognized as $j-1$, so $N_{t,i}=j-1$ means that the i digit at time t is $j-1$.

Finally the data from each image recognition is processed. Where the individual data are of different orders of magnitude, different orders of magnitude are processed for different digits.

$$
\begin{aligned}
T_{1t} &= \begin{cases} N_{t,1}\times 1000 + N_{t,2}\times 100 + N_{t,3}\times 10 + N_{t,4} \left(N_{t,4} \neq 10\right) \\ N_{t,1}\times 100 + N_{t,2}\times 10 + N_{t,3} \left(N_{t,4} = 10\right) \end{cases} \\
T_{2t} &= \begin{cases} N_{t,5}\times 1000 + N_{t,6}\times 100 + N_{t,7}\times 10 + N_{t,8} \left(N_{t,8} \neq 10\right) \\ N_{t,5}\times 100 + N_{t,6}\times 10 + N_{t,7} \left(N_{t,8} = 10\right) \end{cases} \\
&\left\{ N_{t,i} \middle| 110 \le t \le 671, 1 \le i \le 8, t, i \in N \right\}
\end{aligned} \tag{2}
$$

Where T_{1t} is the temperature of thermocouple number 1 and T_{2t} is the temperature of thermocouple number 2. The data output is obtained after this processing.

5.1.3 Pixel discriminant model solution and results

In this paper, the 4,496 sets of processed image data are traversed cyclically using the model, and the output is corrected for changes in order of magnitude and output again to obtain the following table (see appendix and supporting material for details)

Table.1 Temperature table for lines 1# and 2# at various times

NO	Time	1#Temperature	2#Temperature
1	110	900	1142
2	111	904	1146
3	112	910	1146
4	113	914	1144
⋮	⋮	⋮	⋮
560	669	815	1211
561	670	814	1146
562	671	812	1144

The output was visualised as required by the question and the 1# line temperature - 2# temperature time graph is shown in the following figure.

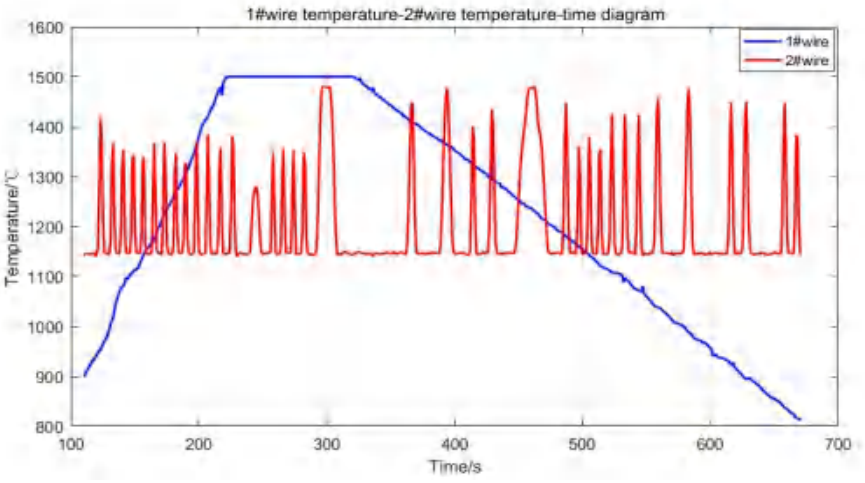


Figure.1 1# line temperature - 2# line temperature time graph

It was observed that the temperature of the 1# line rose smoothly and part of the time at an approximate rate of 5°C per second and 10°C per second, rising to a maximum temperature of 1500°C at 222s. After reaching the maximum temperature the experiment was held for up to 100s then cooled down at a rate of 2°C per second. Observation of the 2# line, the temperature of the 2# line is generally stable and stable for a long time at about 1145°C. At the same time, the temperature of the 2# line fluctuates up and down and some of the temperature changes are large, and there is no obvious pattern.

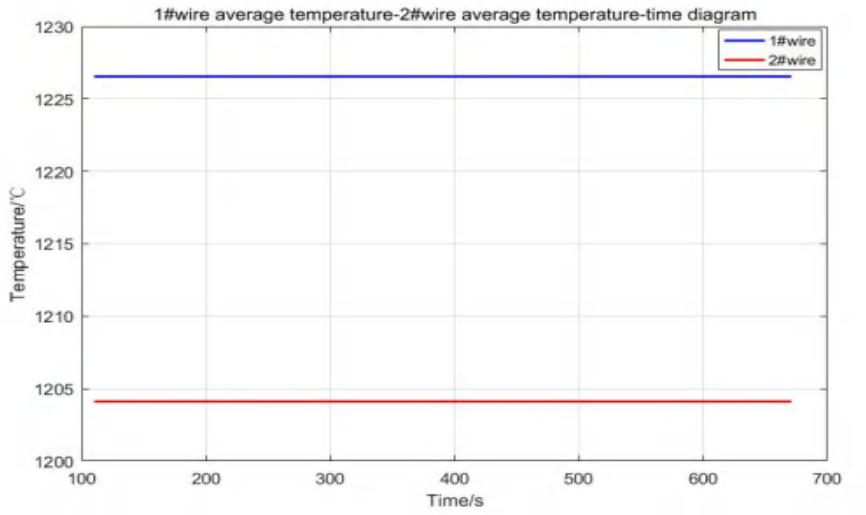


Figure.2 Time diagram of the average temperature of line 1# - average temperature of line 2

From the above graph, it can be seen that the temperature of line 1# increases with time, the temperature first increases, then stays the same and finally decreases; the temperature of line 2# is not very correlated with time and is jagged up and down. In the average temperature image it can be seen that the average temperature of line 1# is higher than the average temperature of line 2#.

The analysis of the 1# line temperature after a linear temperature rise and constant temperature for 100s, followed by a uniform decline, this heating process is consistent with the hot wire method of continuous experiments of temperature control mode. For the 2# line, the temperature always showed irregular jumping lack of reliability. In summary, this paper considers that the temperature data of the 2# line is a problem, the first point may be the 2# line temperature measurement contact poor contact caused by data fluctuations; the second point may be due to the lack of compensation wire and the use of copper wire, then even if there is a temperature gradient in the part will not produce a thermoelectric potential, resulting in errors in the measurement of temperature results.

5.1.4 Test analysis

A convolutional neural network is used to build a digital image recognition model to recognise the image information and compare the output with that of the pixel discrimination model.

5.1.4.1 Introduction to Convolutional Neural Network Principles

Convolutional neural network is a deep feed-forward neural network with features such as local connectivity and shared weights. It is good at processing image, especially image recognition and other related machine learning problems, such as image classification, target detection, image segmentation and other various vision tasks with significant enhancement effects. It is able to classify input information in a translation-invariant manner according to its hierarchical structure, and can perform supervised and unsupervised learning. The shared parameters of the convolutional kernel within its hidden layers and the sparsity of the inter-layer connections enable convolutional neural networks to lattice dotted features with a small amount of computation.

5.1.4.2 How Convolutional Neural Networks Work

A convolutional neural network consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer and an output layer. There are multiple convolutional layers in the network, the purpose of which is to extract image features from the input data, and the sharing of convolutional parameters greatly reduces the number of operations. The presence of multiple convolutional kernels allows for the extraction of image features from multiple perspectives. The convolutional layer is immediately followed by the pooling layer, which acts as a compression of the input features, making the feature map smaller to simplify the complexity of the neural network computation or to re-extract features. [1]

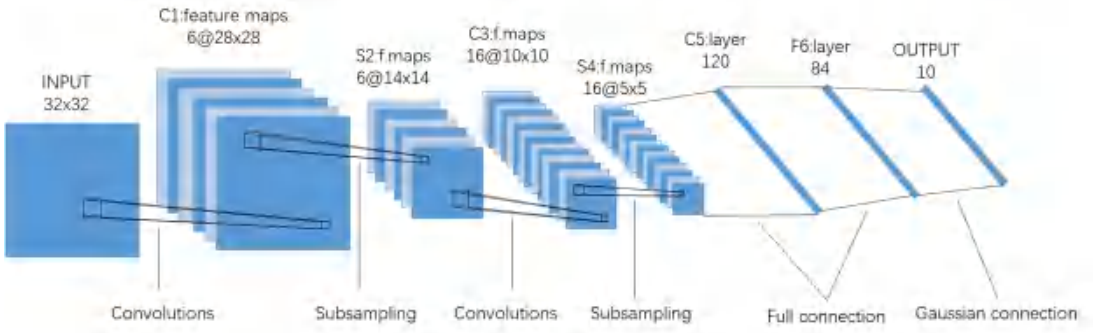


Figure.3 Convolutional neural network working principle diagram

5.1.4.3 Training of Convolutional Neural Networks

Step1: Preparation of the image data to be trained. After the convolutional neural network for handwritten digit recognition has been built, the image data to be trained is input. In this paper, we input eleven sets of image data to be trained, each set contains 23 different handwritten data, a total of 253 handwritten data samples, of which 17 data samples are taken from each set as training data.

Step2: The input layer, convolutional layer and final layer layer parameters are defined. The output layer sets the pixel size template to 28*28 for the sample image data; the convolution layer creates a 2-D convolution layer containing 6 filters of size [5,5]; for the pooling layer this paper uses the maximum pooling function to perform downsampling by dividing the input into rectangular pool regions and then calculating the maximum value for each region. A 2-D convolution layer is then created

containing 16 filters of size [5,5]; a pooling kernel of size 2*2 is then created in the pooling layer, and a 2-D convolution layer containing 120 filters of size [5,5] is created for the final feature extraction.

Step3: The convolutional neural network was trained. The prepared training data samples were imported and 50 iterations of training were performed with the following results

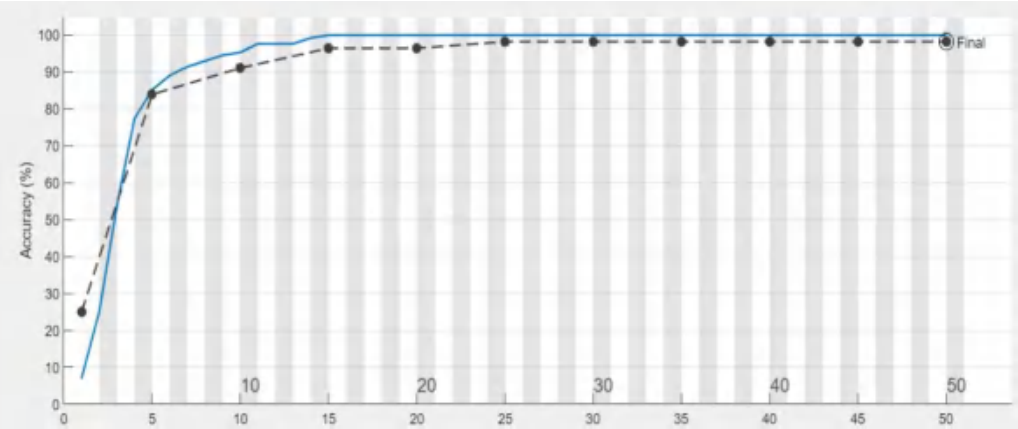


Figure.4 CNN Training Accuracy Image

The actual post-training linearity in the accuracy image is highly consistent with the post-training linearity, with an extremely high training accuracy of 98.21% and valid training results.

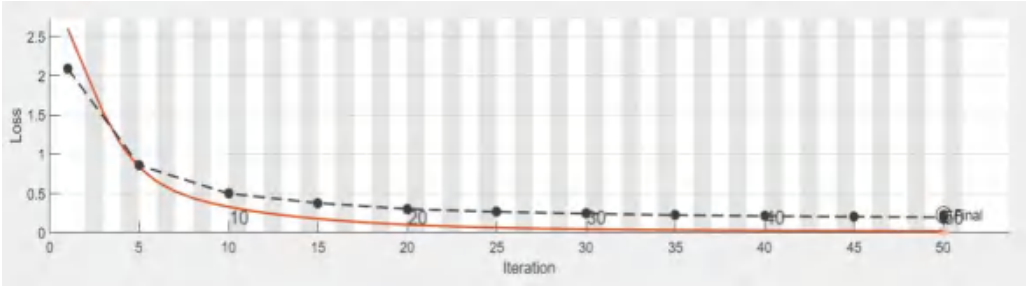


Figure.5 CNN Training Loss Image

Most of the loss curve lines are below 0.5, with less loss, meeting the training requirements. After several training sessions, the training results are all above 90%, and after dozens of training sessions, the training accuracy is up to 98.21%, which basically meets the training requirements. [2]

Step4: The test set was imported for validation. This paper selected eleven experimental groups, respectively 0 to 9 and ℃ (degrees Celsius) in random order for identification, 0 to 9 identification out of normal for the output, for ℃ identification, this paper special treatment output a, and finally the output results of a unified processing, identification results are shown in the following figure.

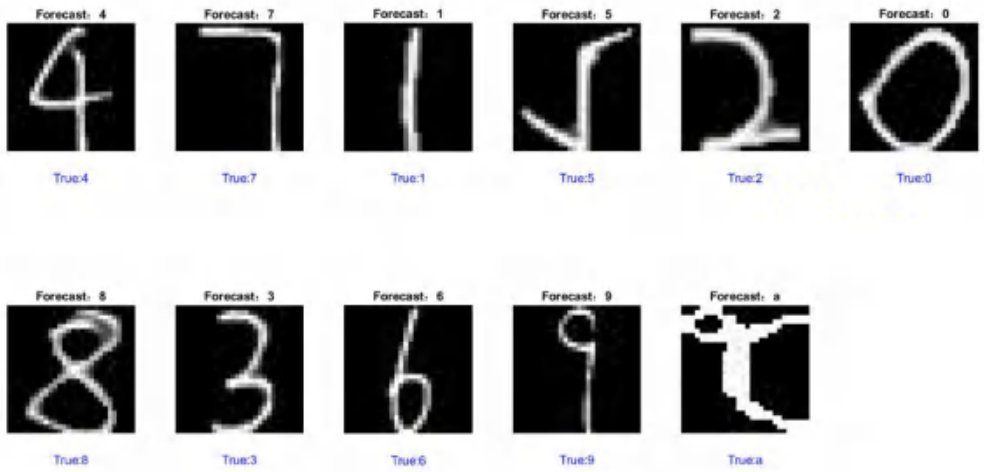


Figure.6 Output of test set validation results

After the image has been analysed, the image is correctly identified for regular digital recognition and output.

5.1.4.4 Use of Convolutional Neural Networks

In this paper, each image will be positioned for cutting as in the preparation above, each image can be separated into eight images data to be analysed after cutting, after all segmentation a total of 4496 images. The images are identified and the results identified are compared with the output of the previous method, the output is shown in the appendix and supporting material, the output is shown in the following figure.

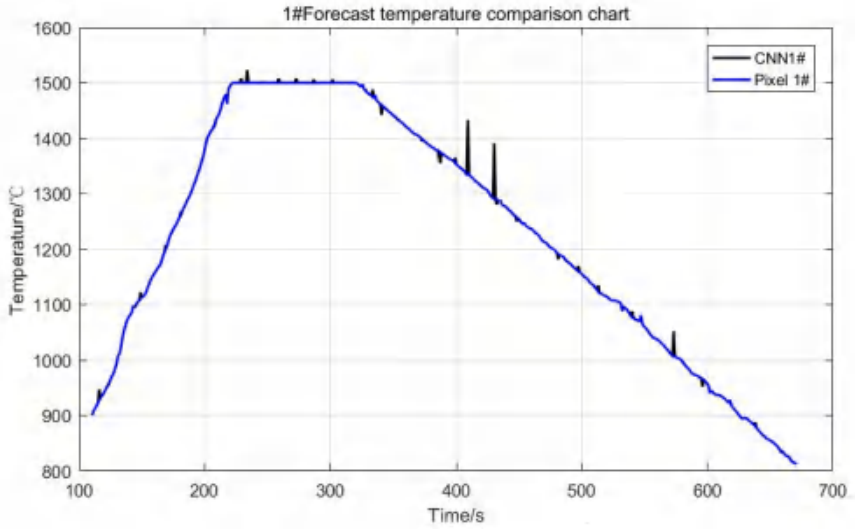


Figure.7 Comparison of CNN and pixel discrimination for 1# line temperature time

After analysis of the image, the black line graph is the result of the convolutional neural network recognition output, the blue line graph is the data obtained from the pixel discrimination method, after comparing the convolutional neural network output

results on the image there are a lot of protrusions, part of the data there is a recognition error makes the image jitter, the convolutional neural network used in this paper in the recognition of the same direction and the same picture pixels, the results are not as expected, except for the jitter part of the results and the pixel discrimination model to reach a highly consistent conclusion.

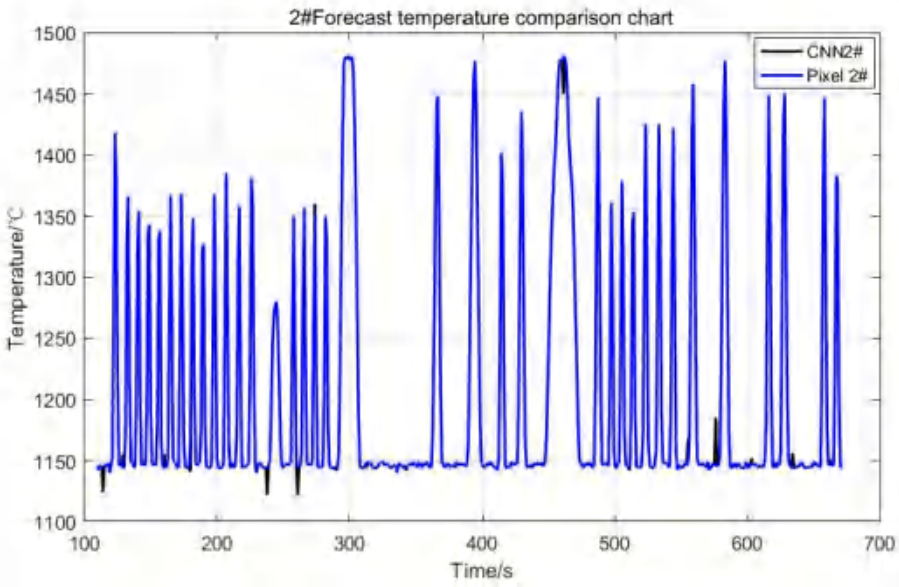


Figure.8 CNN vs Pixel discrimination for 2# line temperature time comparison

Similarly, in the temperature comparison of thermocouple No. 2, the convolutional neural network identified data that was highly consistent with the pixel recognition model except for the dithering part, and thus the pixel discrimination model could pass the test of the convolutional neural network set up in this paper, and thus the pixel discrimination method was chosen appropriately for this paper, so the pixel discrimination model was established.

5.1.5 Short summary

Problem 1 requires the development of a digital image recognition model to automatically extract temperature-related data from images and output them, visualising the output as 1# and 2# lines of time temperature profiles. For this problem, a pixel discrimination model was developed, firstly, the image was pre-processed, the image was greyed out and binarised, the image was turned into a black or white binary image, then based on the location and size of the digits on each image, the data was cut out and enlarged for subsequent preparation; then, each digit on the image was cut into a 158*93 pixel size image, and the size of the digits was uniformly Next, the image is filtered and noise-reduced to make the edges of the numbers smoother; then, the correlation coefficient of each image data to be recognized is calculated with 11 samples, and the two images with the largest correlation coefficients are selected and output numerically, and finally the data is visualized and processed to draw the temperature image; finally, this paper uses the established convolutional neural network-based digital Finally, this paper uses a digital image recognition model based on convolutional neural network to test and

analyse the data, which is basically consistent with the conclusion reached by the pixel discrimination model except for the dithering part, so it shows that the model recognition is accurate.

5.2 Problem 2 Model Building and Solving

5.2.1 Preparation for model building

In this paper, 562 adjacent sequential curves of mould melting and crystallisation are greyscaled so that each pixel point has a value corresponding to the depth of the colour used to represent the melting and crystallisation process. The grey scale range is 0-255, where white is 255 and black is 0. The final image is scaled down to a uniform 300*400 pixels for uniform processing later. [3]

5.2.2 Feature-based equation of state modelling

5.2.2.1 Edge detection using the "Canny" operator

The Canny edge detection algorithm locates objects and boundaries in the image and determines the set of sub-regions or contour lines on the image.

Step1： First, an edge-enhanced image is obtained by smoothing a Gaussian filter and convolving the image. Gaussian filtering uses a two-dimensional Gaussian kernel to convolve with the image. This is because the digital image data is in the form of a discrete matrix and the Gaussian kernel is a discrete approximation to a continuous Gaussian function, derived by discrete sampling and normalisation of the Gaussian surface. The edge images obtained by Gaussian convolution also have some points with high gradient values, non-edges, which are a distraction to the true edges and should be removed. In this paper, the magnitude and direction of their gradients are calculated for each image using equation (3).

$$\begin{cases} M(x,y)=\sqrt{G_x(x,y)^2+G_y(x,y)^2} \\ \theta(x,y)=\arctan\left(\dfrac{G_x(x,y)}{G_y(x,y)}\right) \end{cases} \tag{3}$$

Step2： In order to make the blurred boundaries clear, this paper retains the extreme values of the gradient intensity at each pixel point and performs non-extreme suppression of the magnitude of the gradient, suppressing all gradient values except the local maxima to obtain the ideal gradient image. Then, a double-thresholding means is used to further eliminate the noise. In other words, an upper threshold and a lower threshold are set. Pixel points in the image that are larger than the upper threshold are considered to be boundaries, while those that are smaller than the lower threshold are considered to be not boundaries, and those in between are considered to be candidates and need further processing. [4]

Step3： The edges are traced in the processed image and the data contour in the top left corner is replaced with a 0 matrix to obtain the edge contour image (shown in Figure 8).

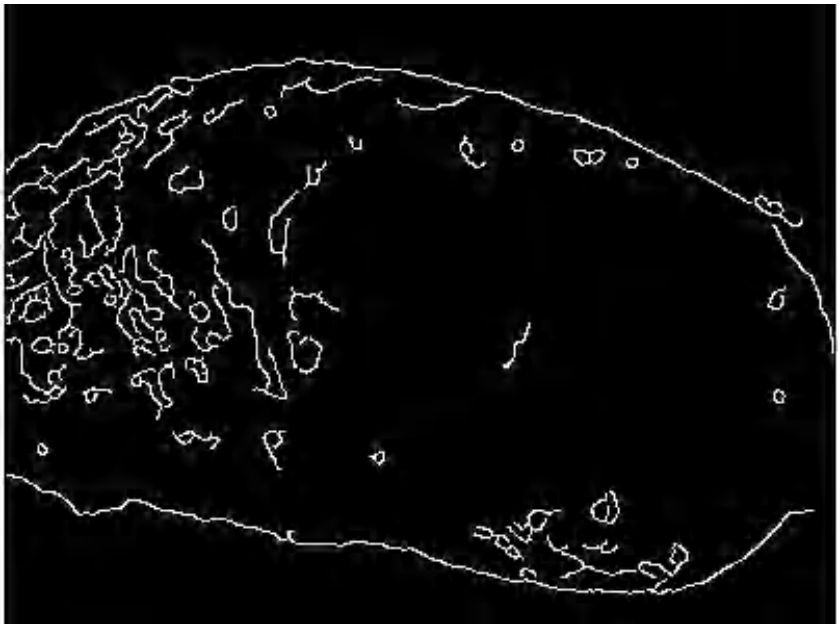


Figure.9 Edge profile at 110s

Step4： For edge contour images, the find function is used to obtain the index vector which separates the mould from the background in the grey-scale map. [5]

5.2.2.2 Feature extraction using Tamura textures

Tamura texture features are widely used to extract information from images, including roughness, contrast, directionality, linearity, regularity and coarseness.

❖ **Extraction of roughness features**

Roughness reflects the amount of granularity in the image texture, with a larger image roughness indicating a more granular image texture pattern and vice versa. The roughness feature extraction method is as follows.

Step1： Find the average gray value of each image element of the image, i.e.

$$C_k(x,y)=\frac{1}{2^{2i}}\sum_{i=x+2^{l-1}}^{x+2^{l-1}}\sum_{j=y+2^{l-1}}^{y+2^{l-1}} g(i,j) \tag{4}$$

where $C_k(x,y)$ is the average gray value of the cell, $g(i,j)$ is the gray value of the pixel located at (i,j), and $l=0,1,\cdots,5$, the active window size is $2^l\times 2^l$ pixels.

Step2： Calculate the average intensity difference between the unoverlapped windows in the horizontal and vertical directions for each pixel point separately, i.e.

$$E_{l,h}(x,y)=\left|C_l\left(x+2^{l-1},y\right)-C_l\left(x-2^{l-1},y\right)\right| \tag{5}$$

$$E_{l,v}(x,y)=\left|C_l\left(x,y+2^{l-1}\right)-C_l\left(x,y-2^{l-1}\right)\right| \tag{6}$$

Where,1 is the grey scale value of the pixel point located at that coordinate.

Step3: Calculation of roughness

$$T_{coa}=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}S_{best}\left(i,j\right) \tag{7}$$

where m and n are the length and width of the image, and the value of l is $E\left(x,y\right)=\max$, the optimal size of the setting window is $S_{best}\left(i,j\right)=2^{l}$.

❖ **Extraction of contrast features**

Contrast is the degree of polarisation of light and darkness in an image and the dynamic range of the grey level, which reflects the depth of the texture grooves and the sharpness of the image. The higher the contrast ratio, the deeper the grooves and the clearer the image; the lower the contrast ratio, the shallower the grooves and the blurrier the image.

Contrast is calculated as

$$T_{con}=\sigma a_4^{-\frac{1}{4}} \tag{8}$$

Where , $a_4=\frac{\mu_4}{\sigma_4}$, μ_4 is the fourth moment and σ is the variance.

❖ **Extraction of directionality features**

Directivity can reflect how concentrated or divergent an image texture is in a certain direction, and is related to the shape of texture primitives and rules for arrangement. A large degree of directivity indicates that the shape of the image is irregular and the distribution range of light and dark is large.

The directivity feature extraction method is as follows:

Step1：Convolution of the image and the operator of Equation (9) to obtain the gradient vector change ΔH in the horizontal direction and the change in the vertical direction ΔV , which can be expressed as:

$$\Delta H=\begin{bmatrix}-1 & 0 & 1\\ -1 & 0 & 1\\ -1 & 0 & 1\end{bmatrix},\Delta V=\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix} \tag{9}$$

Step2：Calculate the gradient vector for each pixel, ie:

$$\theta=\tan^{-1}\left(\frac{\Delta V}{\Delta H}\right)+\frac{\pi}{2} \tag{10}$$

$$\left|\Delta G\left(x,y\right)\right|=\frac{1}{2}\left(\left|\Delta H\right|+\left|\Delta V\right|\right) \tag{11}$$

where $\Delta G\left(x,y\right)$ is the gradient vector at each pixel.

Step3: Calculate the histogram $H_D\left(l\right)$ of the direction of the construction vector, i.e.:

$$H_D(l)=\frac{N_\theta(l)}{\sum_{i=0}^{n-1}N_\theta(i)} \tag{12}$$

where n is the quantization level of the directional angle, and $N_\theta(i)$ is the total number of edge pixels on the corresponding directional angle θ ,when $|\Delta G|\le t$, $\frac{(2l-1)\pi}{2n}\le\theta\le\frac{(2l+1)\pi}{2n}$, the number of pixels is $N_\theta(l)$ .where ΔG is the gradient vector for each pixe， t is the threshold of the gradient vector modulus.For images with an unobvious direction, the value of $H_D(l)$ is relatively flat, and the image $H_D(l)$ with a more obvious opponent has a more obvious peak.

Step4： Calculate the directionality, i.e.:

$$T_{dir}=\sum_{p=1}^{n_p}\sum_{\varphi\in\omega_p}(\varphi-\varphi_p)^2H_D(\varphi) \tag{13}$$

Where $\varphi=1,2,\cdots,n-1$, n_p is the number of histogram peaks, $p(p>0)$ is the peak of histogram $H_D(l)$, is the ω_p peak quantization range, and φ_p is the quantization value in the maximum histogram in ω_p .

❖ **Extraction of linearity features**

Linearity refers to the degree of deviation in the distance between pixels when the co-occurrence matrix is calculated. The larger the interval between pixels in the image, the greater the linearity and vice versa. The formula for calculating linearity is

$$T_{\lim}=\frac{\sum_{i=1}^{m}\sum_{j=1}^{m}P(i,j)\cos\left[(i-j)\frac{2\pi}{n}\right]}{\sum_{i=1}^{m}\sum_{j=1}^{m}P_a(i,j)} \tag{14}$$

where P_a is the distance point of the $m\times m$ local directional symbiotic matrix. [6]

5.2.3 Solution and results of feature-based equation of state model

5.2.3.1 Crystallization process

Observing the image, it was found that the brightness of the image increased as the crystals grew. Therefore, it is speculated that the change of average gray value is affected by crystal growth, crystal growth, and gray scale decreases, so it can be concluded that gray scale is negatively correlated with crystal area. Assuming a linear relationship between the two and a scale factor of k, the crystallization growth at time t is

$$a_t=-k\Delta h_t+a_{t-1} \tag{15}$$

where $\Delta h_t=h_t-h_{t-1}$, h_t refers to the gray value of the image at time t .

At the same time, the influence of crystal growth on image roughness is also

considered. Observing the curved image of roughness over time obtained by using the Tamura texture feature, it is found that the roughness is at a minimum value at 220s, 409s and 537s, and then gradually increases to increase the duration by about 60s. Therefore, it is speculated that the crystallization growth rate is slower at this time. If the coefficient of crystal growth and average gray value change is 0.5k, the relationship between crystal growth and gray value change can be obtained:

$$a_t = -0.5k\Delta h_t + a_{t-1} \tag{16}$$

From this, the recursive formula of the crystallization growth amount at different moments is obtained. Line plots are drawn by images of mean gray difference, and fitted lines are obtained by polynomial fitting, as shown in Figure 9.

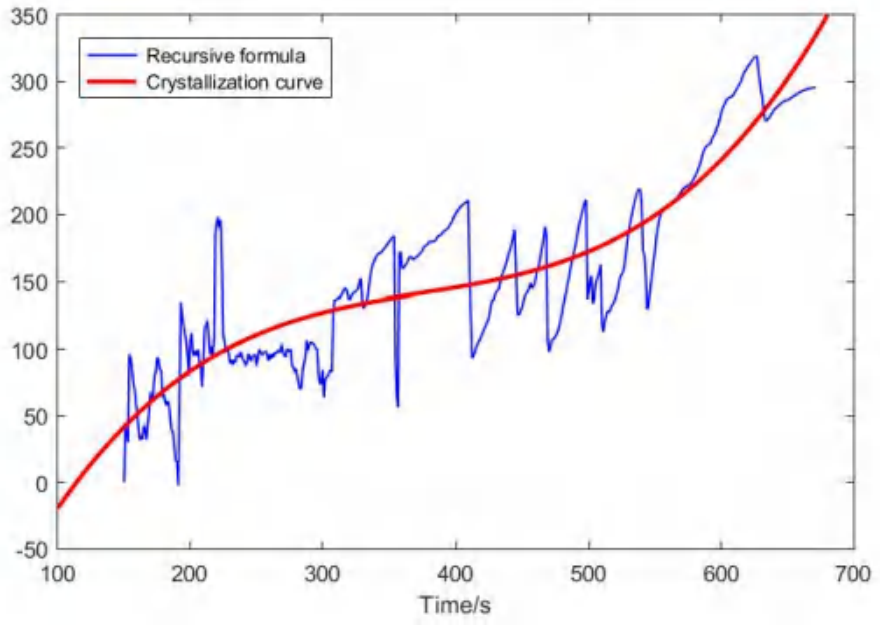


Figure.10 Crystal growth status vs. time plot

As shown in the figure, the blue line is the average gray value, and the red curve is the relationship between crystallization and time after polynomial fitting. It can be seen from the figure that the crystallization rate first increases and then slows down and then increases again with time, and because there is no crystallization result from 110s to 149s in the polynomial fitting image, the image is only the fitting result and has no practical physical significance.

5.2.3.2 Melting process

Looking at the Annex 1 image, as the mold melts, the brightness of the image changes from gray to red, and then from red to gold, and the average gray value also increases. Therefore, it is speculated that the average gray value is positively correlated with the melting amount, and the change of the average gray value increases with the increase of melting amount. At the same time, as the melting process progresses, the surface of the mold becomes smoother and smoother, that is, the roughness decreases. Therefore, it can be deduced that roughness and melting amount are negatively correlated, so assume that the relationship between melting

amount and average gray value and roughness is:

$$b_t = \frac{h_t}{c_t} \tag{17}$$

where b_t is the melting amount of the mold at time t, h_t is the average gray value of the image at time t, and c_t is the roughness of the mold at time t. From this, the melting amount at different moments can be obtained, and then compared with the fitting curve, as shown in the figure below.

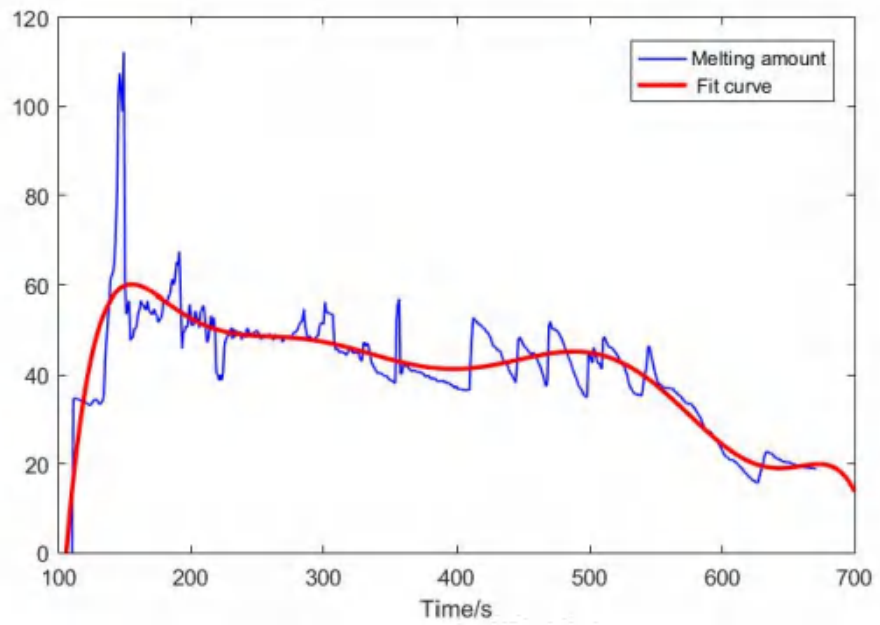


Figure.11 Melting quantity vs. time plot

As shown in the figure, the blue line is the melting amount, that is, the ratio of the average gray value to the roughness, and the relationship diagram with time. In this article, the graph is polynomically fitted, and the fitted line is the red line in the plot. It can be seen from the figure that the melting rate increases with the increase of temperature in the time period from 110s to 149s. [7]

5.2.3.3 Melt crystallization process curve

In this paper, the melting curve and crystallization curve are normalized, and the curve of their relationship with time is combined into one image for better observation. The image is as follows.

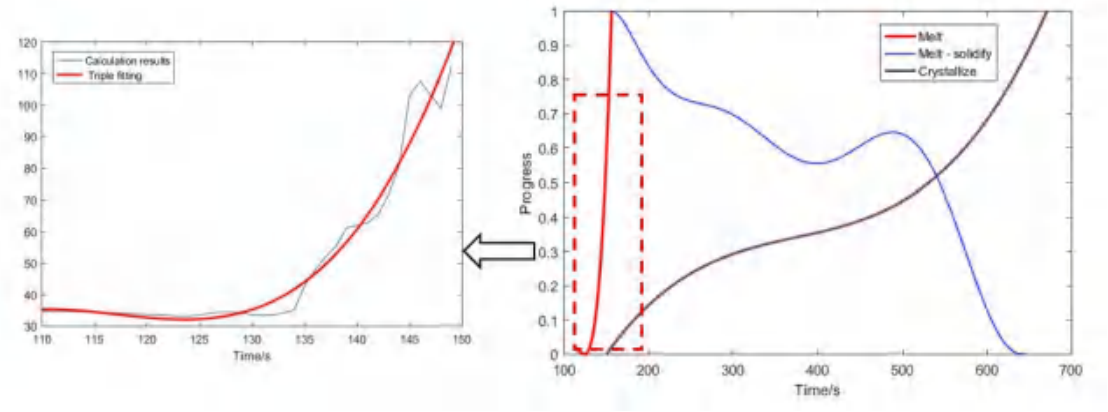


Figure.12 Melt crystallization process curve

As shown in the figure, the red line is the melting curve from 110s to 149s, the blue line is the polynomial fitting result, and the purple line is the result after crystal fitting. Due to the nature of the image conditions and the occlusion of the crystallization process, it is impossible to perform a specific and accurate analysis of the melting process after 150s, and the results of polynomial fitting are used in this paper, and the images before 150s are enlarged and presented for easy reading and grasping the melting characteristics. For the melting process, it is completely melted at 141s, at which point the melting progress reaches its maximum. The melting process has existed since then, but has been reduced relative to the melting progress at 141s. At the same time, the progress of crystallization continued to rise, and at the end of 671s, the maximum progress of crystallization was reached.

5.2.4 Test analysis

In this paper, it is found that roughness, contrast, directionality and linearity reflect the progress characteristics of melting and crystallization to varying degrees. First, this article analyzes the roughness, as shown in the following figure:

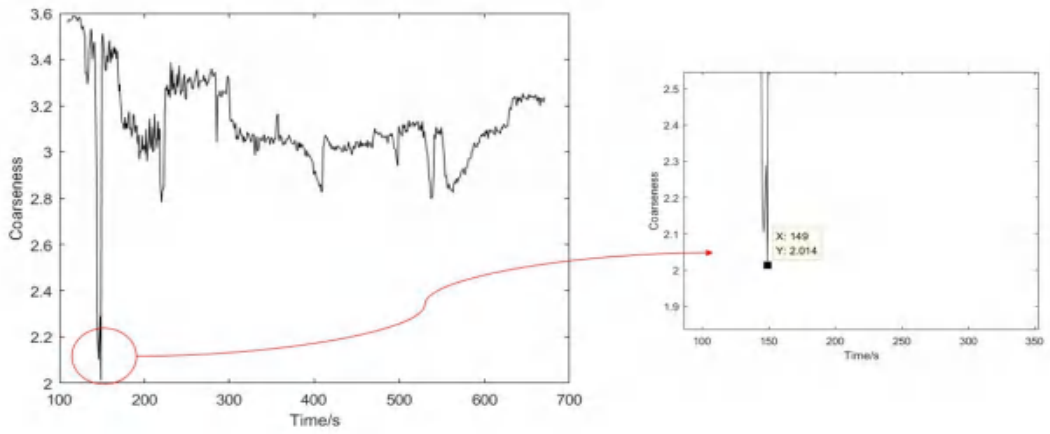


Figure.13 Roughness characteristic analysis diagram

For a detailed analysis of the lowest point of the drastic change in roughness image, it can be seen from the figure that the roughness lowest point time is 149s, when the image is the smoothest, that is, when the color image is the brightest. At this time, the corresponding sample has completely melted, and the steel has also melted, and the degree of melting of the sample is very high.

Then, the melting crystallization process is discussed in this paper with four characteristics: average gray value, linearity, contrast and directionality.

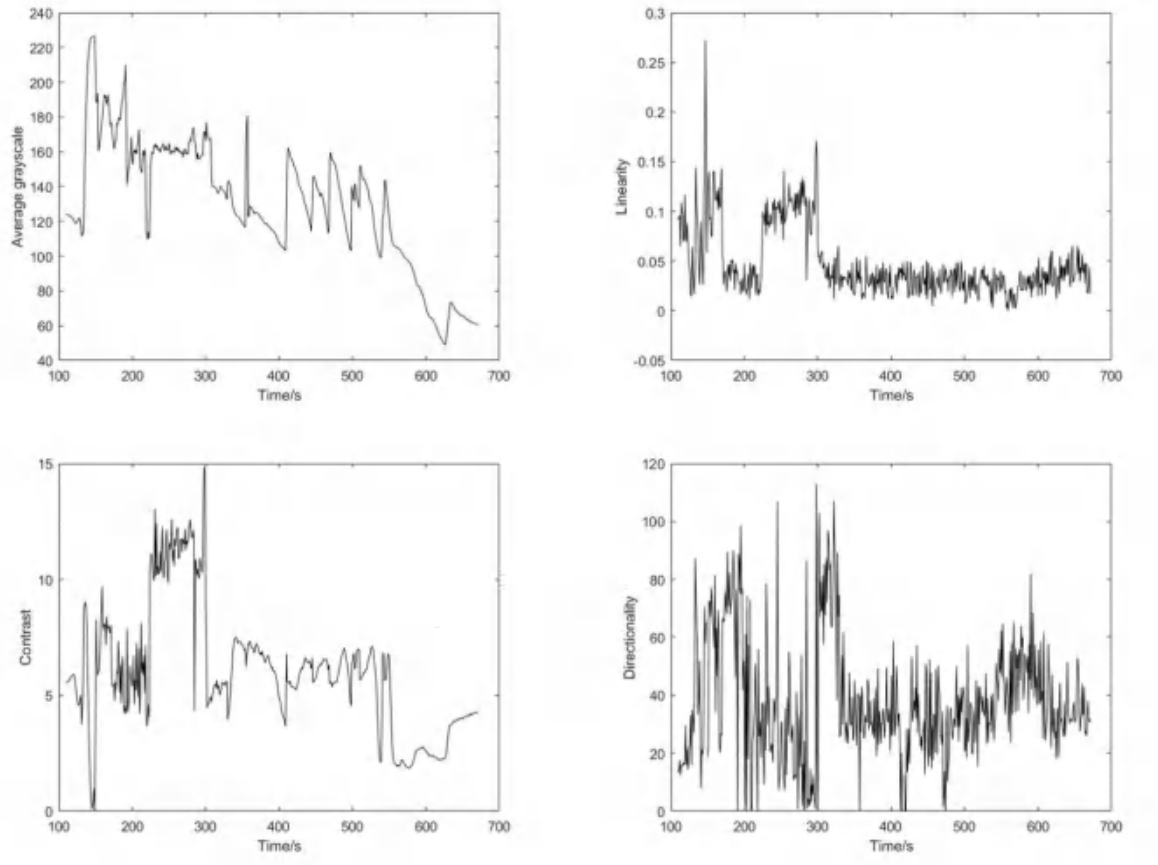


Figure.14 Four feature group diagrams

In the image above, the first row is the average gray value image and the linearity image, and the second row is the contrast image and the direction image, respectively. In this paper, when analyzing the four characteristics, the melting and crystallization are linked, and the image is used to test. In the process of 110s to 149s, due to the higher and higher temperature, the brightness of the picture is increasing, and the average grayscale map as a whole is on the rise. In the crystallization process, that is, after 150s, the average gray value of the image changes when there are crystal nuclei, and then the gray value decreases during the crystal growth process; As the temperature drops, more nuclei appear and the average gray value rises and falls repeatedly. For linearity images, contrast images and directionality images, this paper finds that the images change drastically at the junction of complete melting and crystallization, which shows the process of melting and crystallization, which proves that the conclusions obtained by the feature-based equation of state model are consistent with the experimental phenomenon, and verifies the reliability of the equation of state model of features.

5.2.5 Short summary

The second problem requires studying and quantifying the dynamic differences of images during mold melting and crystallization, establishing models of different image features according to time development, and discussing the change curves of mold melting and crystallization processes. For this question, a feature-based equation

of state model is established. First, the picture is preprocessed, and after the grayscale processing of the picture, the image is uniformly reduced to 300*400 pixels according to the proportion. Then, a Gaussian smoothing filter is used to convolve the image, and then the image ladder amplitude is suppressed by a non-maximum value to make the image blurred boundary clearer. Then trace the boundary in the processed image to obtain the outline image, and use the find function on the outline image to separate the mold from the background. Then, the Tamura texture features are used to extract the four image features of roughness, contrast, directionality and linearity, and the change law of the four texture features and the recursive formula about time are found in the process of crystallization and melting, and the calculated data is visualized to obtain a line chart, and finally polynomial fitting is performed to obtain the process curve of melting and crystallization. In this paper, the five characteristics of roughness, contrast, directionality, linearity and average gray value are combined with images and experimental phenomena for examination and analysis, and it is found that the average gray value can obviously reflect the emergence of crystals in the crystallization process. The other four features changed drastically between 141s and 150s, which showed the process of melting and crystallization, which proved that the conclusions obtained by the feature-based equation of state model were consistent with the experimental phenomenon, and verified the reliability of the equation of state model of features.

5.3 Model establishment and solution of problem three

5.3.1 Establishment of polynomial regression model

Observing Figure 1, it can be seen that within the first 222 s, the temperature and time have a linear relationship; Within 222s-321s, the temperature is basically maintained at 1500°C unchanged; After 321s, the temperature has a linear relationship with time and shows a downward trend.

In the first 222 s, it can be seen from observation that the temperature increases linearly with time, so let

$$T = at + b \tag{18}$$

For a given data point $(t_i, T_i) i = 110, 111, \cdots, 222$, the primary formula $T = at + b$ is calculated so that the total error is minimized

$$Q = \sum_{i=110}^{222} (T - at + b)^2 \tag{19}$$

Find $a = 5.3408, b = 307.3435$. So from 110s to 222s, the function of time and temperature is $T = 5.3408t + 307.343$. In the same way, when $t \in (321, 671)$, the equation of time and temperature is $T = -1.9663t + 2135.7$.

In summary, the function of time and temperature is

$$T = \begin{cases} 5.3408t + 307.3435 \ (110 \le t \le 222) \\ 1500 \ (222 \le t \le 321) \\ -1.9663t + 2135.7 \ (321 \le t \le 671) \end{cases} \tag{20}$$

According to the process curve of mold melting and crystallization obtained in question 2, the polynomial regression expression is established in the relationship between melting amount and temperature, and the relationship between crystallization quantity and temperature at the same time

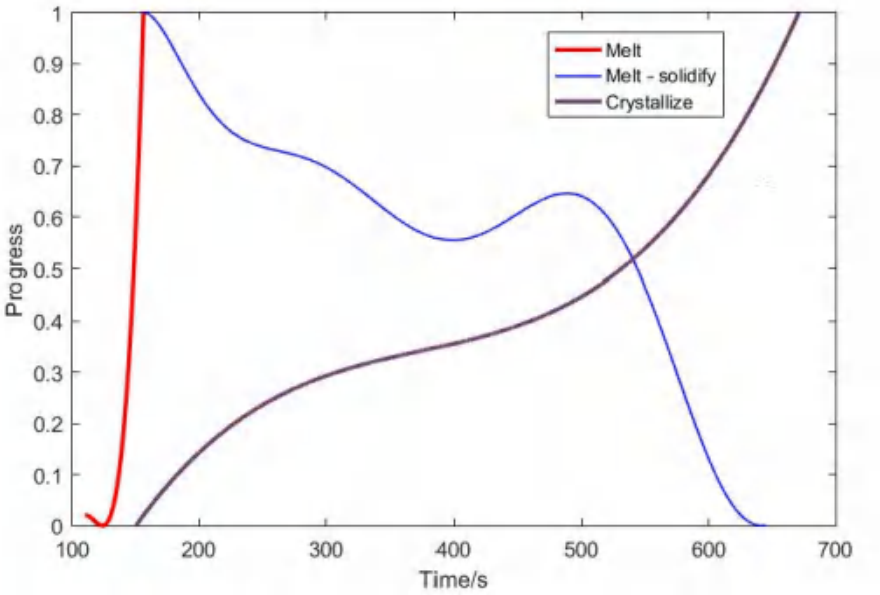


Figure.15 Diagram of melting and crystallization processes

Considering that the fitted results are inaccurate because the data points from 110s to 150s are fewer than the population, the data from 110s to 150s are fitted separately, and the fitting results are obtained as follows:

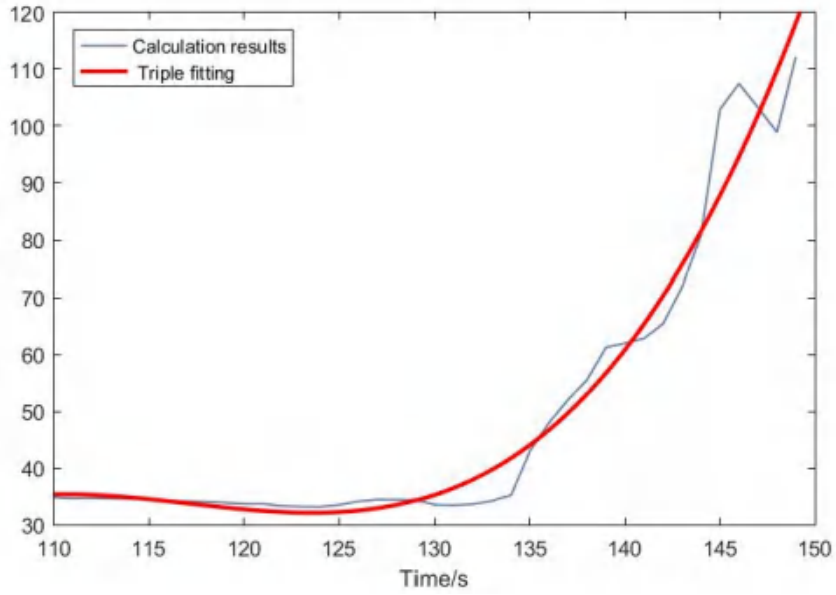


Figure.16 110s-150s melting status fitting diagram

It can be seen from the above figure that during the 110s-150s, the change trend of the melting process is basically the same as the fitting curve, and the difference is not much.

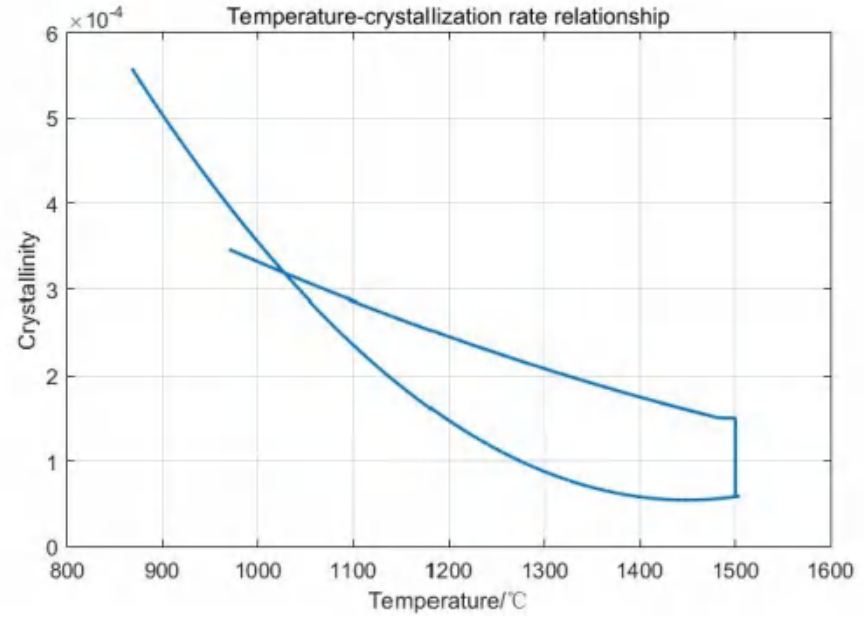


Figure.17 Temperature vs. crystallization plot

From the above figure, it can be found that the crystallization process can be divided into heating and cooling processes, and the two processes are fitted separately

$$\begin{aligned}&\text{Temperature rises：}\ m_i = 2.026\times10^{-10}T_i^2 - 8.7885\times10^{-7}T_i + 0.00100\,82\\&\text{Temperature drops：}\ m_i = 1.4897\times10^{-10}T_i^2 - 4.3176\times10^{-7}T_i + 0.0031\,828\end{aligned} \tag{22}$$

where T_i is the temperature at i-time and m_i is the crystallization rate at i-time.

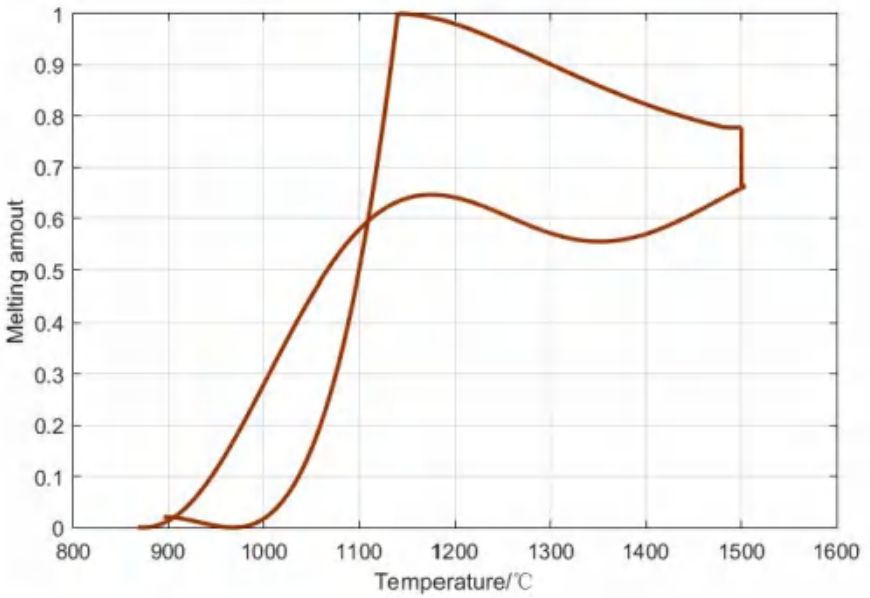


Figure.18 Melt-temperature diagram

As can be seen from the figure above, the relationship between melting amount and temperature reaches a maximum at 1140 °C, and then changes dramatically. Therefore, the two processes before and after 1140 °C are fitted as a function

relationship, and the function relationship of the melting rate before 1140 °C is obtained：

$$n_i = 1.9273\times10^{-7}T_i^2 - 3.594\times10^{-4}T_i + 0.16751 \tag{23}$$

Before 1140 °C, the melting rate as a function of ：

$$n_i = -3.9813\times10^{-11}T_i^3 + 1.6791\times10^{-7}T_i^2 - 2.3449\times10^{-4}T_i + 0.10857 \tag{24}$$

Draw the function as an image：

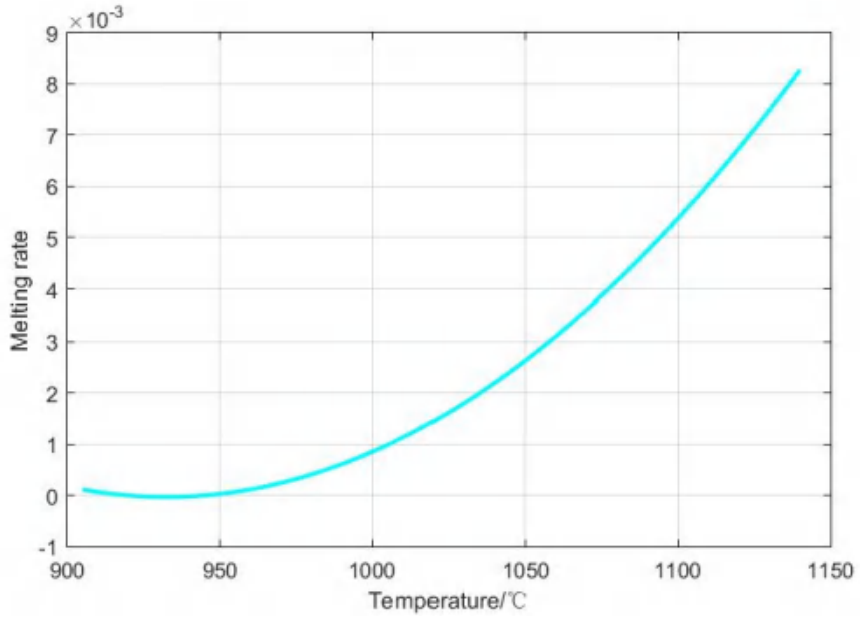


Figure.19 Melting rate-temperature plot below 1140°C

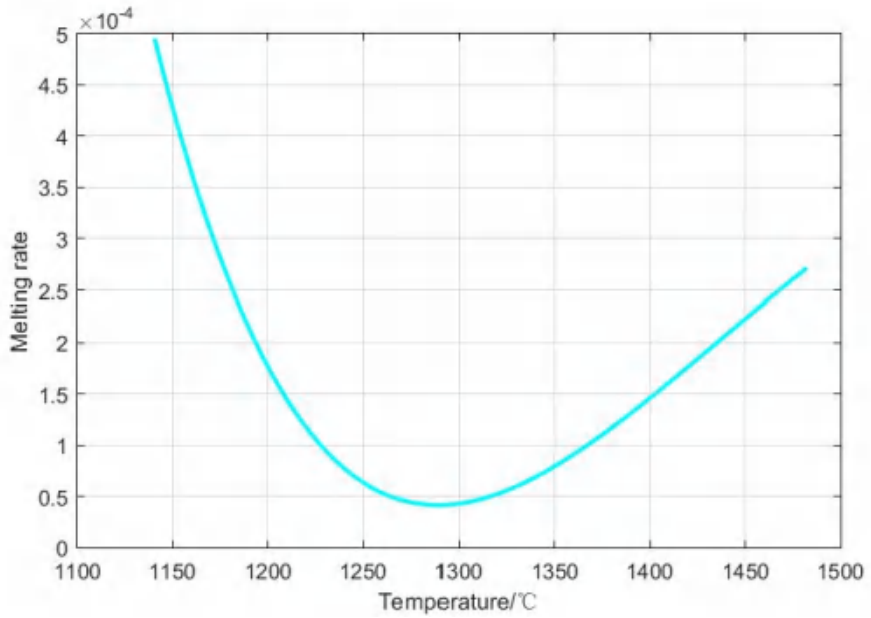


Figure.20 Melting rate-temperature plot above 1140°C

5.3.2 Test analysis

The crystallization rate can be multiplied by the proportion of the crystallization volume that has been crystallized and the total volume of the original liquid by the total crystallization time, assuming that the proportion of the crystallization volume that has been crystallized and the total volume of the original liquid is x, the total crystallization time is t, the crystallized liquid volume is V^s, the remaining uncrystallized liquid volume is V^t, the original total liquid volume is V, and the liquid easily reaches a new temperature, and the holding time at this temperature is τ. The number of crystallized particles in $\mathrm{d}\tau$ time is:

$$N_\tau = IV^t \,\mathrm{d}\tau \tag{25}$$

where I is the amount of crystallization produced per unit volume per unit of time. Let U be the rate at which a single grain interface produces crystals, and assume that the crystals grow at the same rate in all directions and that the grains are spherical. After the τ-time, the volume of the crystal produced in the total time t is

$$V_\tau^s = \frac{4\pi}{3}U^3\left(t-\tau\right)^3 \tag{26}$$

At the beginning of crystallization, $V^\tau \approx V$ is assumed, so at the t moment, the volume of liquid that has been crystallized is the crystal volume produced by τ in $\tau+\mathrm{d}t$ and the period, that is

$$\mathrm{d}V^s = N_\tau V_\tau^s \approx \frac{4\pi}{3}U^3\left(t-\tau\right)^3 \tag{27}$$

From this, it can be deduced that the proportion x of the crystallized crystal volume to the total volume of the original liquid is

$$x = \frac{V^s}{V} = \frac{4\pi}{3}\int_0^t IU^3\left(t-\tau\right)^3 \tag{28}$$

$$\mathrm{d}x = \frac{4\pi}{3}IU^3\left(t-\tau\right)^3 \mathrm{d}t \tag{29}$$

Given the interparticle collision and the correction factor $1-x$ of the reduction of the mother liquor, so

$$\mathrm{d}x = \left(1-x\right)\frac{4\pi}{3}IU^3\left(t-\tau\right)^3 \mathrm{d}t \tag{30}$$

Points are obtained

$$x = \exp\left(-\frac{\pi}{3}IU^3t^4\right) \tag{31}$$

Given the nucleation rate and growth rate over time, equation (24) can be collated：

$$x = 1 - \exp\left(-Kt^n\right) \tag{32}$$

where K is the crystallization rate constant and n is the Avrami index. Equation (28) is that the JMA formula[8] substituts the data from Annex II to obtain $K = 0.00038$, $t = t_i - 420, n = 3$, i.e：

$$x = 1 - \exp\left(-0.00038\left(t_i - 420\right)^3\right) \tag{33}$$

From the above equation, the relationship between temperature and crystallization rate can be plotted:

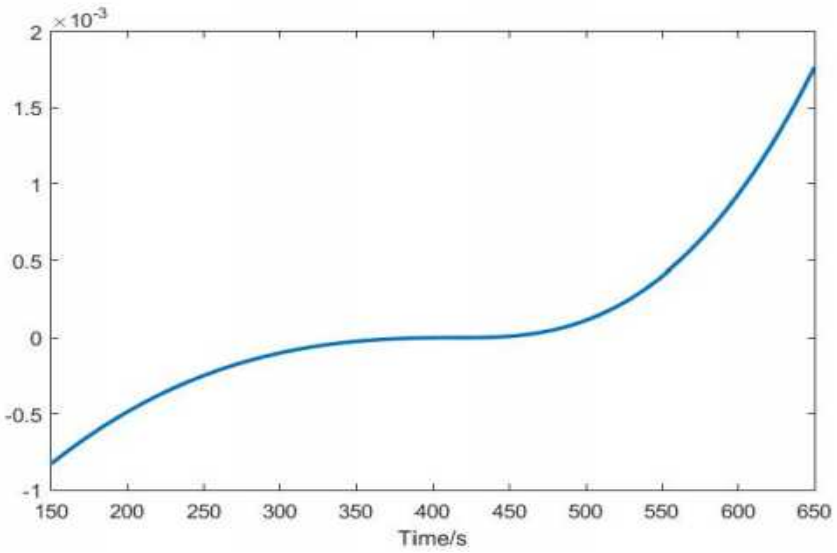


Figure.21 JMA formula crystallinity

Looking at the image, it can be found that Figure 14 is highly consistent with the image obtained in the second question, so it can be proved that the image obtained in the third question is accurate and reliable.

5.3.3 Short summary

Question 3 requires exploring temperature and time changes and the functional relationship between melting and crystallization, and discussing the melting and crystallization kinetics of melt fluxes. For this question, a polynomial regression model is established. Firstly, based on the temperature at different moments found in problem 1, polynomial fitting regression is carried out, so as to establish the time-temperature piecewise function. Then, according to the process curve of mold melting and crystallization obtained in question 2, the relationship between melting amount and temperature and the relationship between crystallization quantity and temperature at the same time are integrated, so as to establish a polynomial regression expression and draw an image. Through the change curve of mold crystallization with respect to time, the crystallization amount is derived to time, so as to determine the relationship between crystallinity and time, and the relationship between melting rate and time can be found in the same way. Finally, combined with the crystallization kinetics, it is found that the relationship between crystallinity and temperature obtained in this paper is highly consistent with the image of the JMA formula, which

proves the accuracy and reliability of the obtained results.

6 Evaluation and promotion of models

6.1 Advantages of the model

(1)For image digital recognition, a pixel discrimination model is established, which segments the numbers in the image according to a fixed size, and uses the numbers in the image as the discriminant template to make the results of the model solution more accurate. By calculating the correlation coefficient of the pixel matrix, the template with the largest correlation coefficient is selected as the recognition result, which ensures the reliability of the model.

(2)For the extraction of image features, the Tamura algorithm is used to extract texture features as the basis for model establishment, which ensures the reliability of the model. By analyzing the relationship between the characteristics and mold melting and crystallization, the equation of state model is established to ensure the rationality of the model.

(3)In order to establish a functional relationship between temperature and time discrete data, we choose to use a polynomial regression model to establish it, which ensures the reliability of the functional relationship. Through the visual processing of data, the rationality of the function relationship is guaranteed.

6.2 Disadvantages of the model

(1)Although the pixel discrimination model has high accuracy, it is less flexible and can only recognize fixed images.

(2)Polynomial regression models establish complex functional relationships.

6.3 Promotion of the model

The pixel discrimination model has high accuracy and good recognition effect, and is suitable for identifying images with consistent formats, such as certificate recognition, icon recognition, and so on.

7 References

[1]陈群贤.TensorFlow 下基于 CNN 卷积神经网络的手写数字识别研究[J].信息记录材料,2022,23(09):

[2] Li Ziyang. Single-well lithofacies identification based on logging response and convolutional neural network[J]. Journal of Applied Geophysics,2022,207:68-75.

[3] 刘峰.基于图像边缘检测的绘画机器人轨迹规划研究[D].辽宁：大连理工大学,2021.10.

[4] 赵娅岐.基于 Canny 算子的边缘检测算法研究改进与电路实现[D].湖北：华中

科技大学，2021.
[5] Jadwaa Sana’a Khudayer.X-Ray Lung Image Classification Using a Canny Edge Detector[J].Journal of Electrical and Computer Engineering,2022.
[6] Taguchi Toshio. The post-war rebirth of Yokohama: the planner Akira Tamura’s contributions to municipal reform[J]. Planning Perspectives,2022,37(05):1073-1095.
[7] Xiao Zikang. Quantitative Evaluation of Reservoir Heterogeneity in the Ordos Basin Based on Tamura Texture Features[J]. Frontiers in Earth Science.2022.
[8] Liu Hui. Erratum to “Light nuclei production in Au+Au collisions at [formula omitted] = 5–200 GeV from JAM model”[J].Physics Letters,2022,829.

Team Number:	apmcm2205838
Problem Chosen:	B

High-Speed Train Aerodynamic Shape Optimization Research

Abstract

With the rapid development of the computing field, the pursuit of high speed and low noise pollution of the aerodynamic shape is particularly important. In this paper, a corresponding mathematical model has been established to study the speed as well as noise domain of high speed railways and simulations have been carried out.

For problem 1, based on Standard $k-\varepsilon$ Models, four preliminary high-speed train models are established respectively, and the force analysis of the universal train head is carried out in two dimensions, as well as the simulation of the four models by RWIND wind tunnel software, and the Δ factors used for description are defined, and the Δ factors of each model are compared, and it is concluded that **TP1 is the best high-speed train shape with the lowest air resistance**.

For problem 2, we divided TP1 into 5 parts, optimized and fine-tuned them by Pareto search method, and built the optimized high speed rail model.

In response to problem 3, and the fact that many scholars have studied the noise pollution of high-speed trains, many countries have also introduced corresponding speed limit regulations. We model and analyze a high-speed train traveling at 85m/s based on the physical properties of turbulence, and conclude that **TP4 is the optimal high-speed train shape that generates the least noise**.

Finally, we combine the characteristics of TP1 and TP4 and use a multi-objective particle swarm algorithm to design a new train which has a more balanced force distribution, i.e., good upper speed limit, and low noise to the environment in wind tunnel experiments, and the Δ of the new train is excellent when compared by the Δ factor. $\mathbf{\Delta_{TP_1} = 0.0105}$, $\mathbf{\Delta_{TP_4} = 0.0031}$, $\mathbf{\Delta_{TP_{best}} = 0.0029}$. Comparing the models before and after optimization, it can be found that the nose cone height decreases, the nose cone length increases, and the nose cone becomes smoother; the driver's chamber height is adjusted downward, and the front half of the streamline shape becomes narrower and the width of the rear half increases.

Keywords: High-Speed Train Standard $k-\varepsilon$ Models Wind Tunnel Simulation MOPSO Blender Modeling Software

Contents

I. Introduction

1.1 Problem Background

In order to indicate the origin of problems, the following background is worth mentioning. High-speed high-speed rails have become the way of travel for most people, and the continuous research on high-speed high-speed rails makes the development of high-speed high-speed rails play a role in promoting, and with the accelerated pace of people's lives and improve the transport capacity and transport efficiency of high-speed railroads, railroad speed has become the current development trend, which needs to be continuously optimized, and this paper analyzes the air resistance to its high-speed railroad head, generates noise and other problems, and establishes an optimization model.

1.2 Restatement of the Problem

Question 1

Establish the mathematical model of air resistance of high speed railroad, analyze the force relationship of high speed railroad head in general weather and extreme weather, and simulate accordingly out the force relationship of cone type and four kinds of high-speed railroad heads in the figure. The best high-speed railroad shape with the lowest air resistance is obtained.

Question 2

Analyze the effect of the curvature of the head of the high speed rail on the air resistance, and in order to minimize the air resistance, build an optimization model, design the best shape of the high speed rail, and draw a sketch of the optimized high speed rail.

Question 3

The noise intensity generated by the conical shape as well as the four classical high speed high-speed rails shown in Figure 2 is analyzed, and the noise distribution is simulated to select the best shape of high speed high-speed train with the least noise.

Question 4

The results of the above three problems were combined to create an optimized model of the shape of the high speed rail, to design the best shape of the high speed rail to increase its speed and reduce its noise, and to draw a sketch of its shape and its structural parameters.

1.3 Our Work

Question 1

Through our analysis of the continuity principle of the fluid, in essence, the fluid in the flow of mass conservation, after Bernoulli's equation can be obtained for the ideal fluid it is the conservation of mechanical energy, through the Poiseuille theorem can be obtained for the actual fluid is the existence of viscous friction, this viscous friction will cause a kind of resistance to the relative motion of fluid and solid, according to this establishment of air resistance model to study high-speed railroad Air resistance. When the fluid speed is very fast, turbulence will be generated. In this regard, the standard $k-\varepsilon$ model is applied to solve the turbulent kinetic energy and its dissipation rate equation. For the four high-speed railroad shapes, in extreme weather (such as rain and snow), we use the force analysis method to represent the force relationship, for the different shapes of high-speed railway we make four models of high-speed railway front end by Blender software, and the four models are carried out wind tunnel experiments, which can intuitively feel the air resistance to each part, and by comparing the air resistance distribution to select the By comparing the air resistance distribution, we choose the shape of high-speed railroad with the least air resistance. **The final TP1 is the best high speed rail model with the least air resistance.**

Question 2

On the basis of TP1, an optimization model is established based on the Pareto search process to design the optimal high speed rail model with the minimum air resistance.

Question 3

We firstly collect the relevant noise data to have a certain understanding of the noise of high-speed train (train) in each country, and secondly get that when the speed of the car is very fast, the wave surface accelerates the buildup, which makes the air friction increase and the noise also increases. In this regard, the external aerodynamic noise model of high-speed train is established, and the turbulent kinetic energy turbulent dissipation rate at each node is extracted by the air resistance model in turbulence, so as to determine the sound power at each node. Then the *Lighthill-Curle acoustic analogy theory* is compared to the high-speed railroad pressure distribution to realize the aerodynamic noise simulation. **TP4 is finally determined to be the optimal shape of the high-speed railway that generates the least noise.**

Question 4

Based on the Pareto search in the second question, we used the multi-objective particle swarm algorithm (MOPSO) to combine the characteristics of TP1 and TP4 to find the approximate model, and the model was simulated in a wind tunnel to obtain the relevant data.

Combined with the simulated values of the theoretical conic train in the third question, the optimized model $\mathbf{\Delta}_{\mathrm{TP}_{\mathrm{best}}} = \mathbf{0.0029}$ is the smallest among the three of TP1 and TP4 in terms of Δ coefficients (the less the better), reducing **72.38**% compared to TP1 and reducing **6.45**%, which verifies the feasibility of the model.

II. Model Assumptions

1. The material is the same for different high-speed railroads.
2. High-speed rail is straight.
3. The wind speed is the same.
4. Body length is the same.

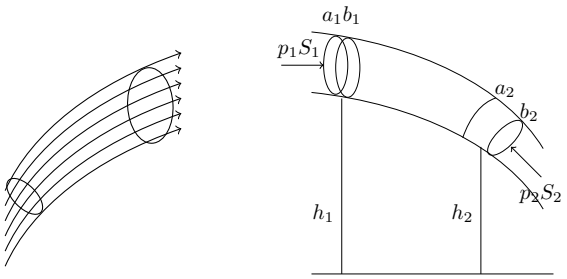
III. Symbol Description

Symbols	Meaning	Unit
u	Turbulence velocity	m/s
c_0	Air speed	m/s
S	Cross-sectional area	m^2
p_1	Pressure	Pa
R	Radius	m
P_{ij}	Pulsating pressure of air on the outer surface of the body	Pa
P	Sound pressure	Pa
P_0	Reference sound pressure	Pa

IV. Models

4.1 Physical basis of the model

4.1.1 *Principle of Continuity*



(a) Tube of flow (b) Bernoulli's principle

Figure 1 The basic theory of ideal fluids.

In a constant flow field, take any section of the flow tube and set the vertical cross sections at its two ends as $\mathrm{d}S_1$ and $\mathrm{d}S_2$. In a constant flow, the tube of flow is stationary and the density ρ at each point in the fluid should not change with time, so the mass of the fluid in this section is constant, and thus the mass flow rate $\mathrm{d}Q_{m1}$ from one section is always equal to the mass flow rate $\mathrm{d}Q_{m2}$ from the other end. So

$$\rho_1 v_1 \mathrm{d}S_1 = \rho_2 v_2 \mathrm{d}S_2 \tag{1}$$

Or, along the arbitrary tube of flow

$$\rho v \mathrm{d}S = Constant \tag{2}$$

If we further assume that the fluid is incompressible, then its density does not change and we have $\rho_1 = \rho_2$, thus

$$v_1 \mathrm{d}S_1 = v_2 \mathrm{d}S_2 \tag{3}$$

Or, along the arbitrary tube of flow

$$v \mathrm{d}S = Constant \tag{4}$$

The above equations are called the continuity principle of fluids, which in physical essence embodies the conservation of mass in the flow of fluids.

4.1.2 *Daniel Bernoulli's Equation*

As shown in *Figure 1*, a section of fluid is sectioned off by cross sections S_1 and S_2 for any of the flow tubes in an ideal fluid with constant flow. During the time interval Δt, S_1 at the left end moves from position a_1 to b_1 and S_2 at the right end moves from position a_2 to b_2, such that $\overline{a_1 b_1} = \Delta l_1$ and in $\overline{a_2 b_2} = l_2$, then $\Delta V_1 = S_1 \Delta l_1$ and $\Delta V_2 = S_2 \Delta l_2$ are the volumes of fluid flowing in and out during the same time interval, respectively, for the incompressible fluid constant flow, $\Delta V_1 \equiv \Delta V_2 \equiv \Delta V$. Because there is no viscosity, i.e., no dissipation, we can apply the law of

conservation of mechanical energy to the fluid in this section of the flow tube. Although the fluid changes in the section from b_1 to a_2, the flow is constant and the state of motion remains unchanged. Thus, the kinetic and potential energies remain unchanged. Therefore, when examining the change in energy, we only need to calculate the difference in capacity between the two end elements ΔV_2 and ΔV_1. First, consider the change in kinetic energy:

$$\Delta E_k = \frac{1}{2}\rho v_2^2 \Delta V - \frac{1}{2}\rho v_1^2 \Delta V$$

Consider again the change in gravitational potential energy:

$$\Delta E_p = \rho g \left(h_2 - h_1\right) \Delta V$$

Now look at the work done by the external force on the fluid in this section of the flow tube. Let the pressure at the left end be p_1 and the force $F_1 = p_1 S_1$ acting on S_1. The work done by the external force is $A_1 = F_1 \Delta l_1 = p_1 S_1 \Delta l_1 = p_1 \Delta V$; the pressure at the right end is p_2, the force acting on S_2 The force on $F_2 = p_2 S_2$, the external work is $A_2 = -F_2 \Delta l_2 = -p_2 S_2 \Delta l_2 = -p_2 \Delta V$. Therefore

$$A_{out} = A_1 + A_2 = \left(p_1 - p_2\right) \Delta V$$

From the conservation of mechanical energy $A_{out} = \Delta E_k + \Delta E_p$ we get

$$\left(p_1 - p_2\right) \Delta V = \frac{1}{2}\rho \left(v_2^2 - v_1^2\right) \Delta V + \rho g \left(h_2 - h_1\right) \Delta V$$

So

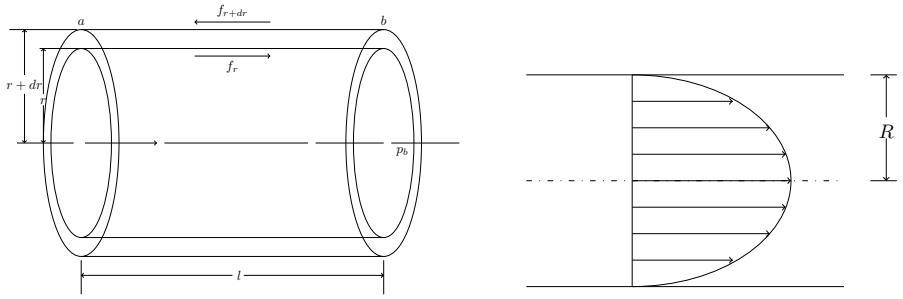
$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2 \tag{5}$$

Since 1 and 2 are any two points in the same stream, Equation 5 can also be expressed as the same *streamline*.

$$p + \frac{1}{2}\rho v^2 + \rho g h = Constant \tag{6}$$

Equation 5 and 6 are the *Daniel Bernoulli's Equation*.

4.1.3 *J. L. M. Poiseuille's Equation*



(a) Derivation of J. L. M. Poiseuille's Equation (b) Poiseuille flow velocity distribution

Figure 2 Derivation of J. L. M. Poiseuille's Equation and Flow Rate Distribution.

Due to the existence of viscosity, attached to the wall of the solid immersed in the fluid and the relative speed of the solid surface is always 0. To keep the fluid for constant flow, there must be

pressure; to keep the solid for uniform motion, there must be a support force. This means that the relative motion of the fluid and the solid is subject to a resistance. This resistance comes from the viscous nature of the fluid[2].

Consider a horizontal tube ab of radius R and length l, in which the fluid flows along the axis and the velocity of the fluid attached to the wall of the tube is 0 due to viscosity. For a given pressure difference, the velocity of the fluid has a distribution along the radial direction, with the maximum velocity v in the central tube axis ($r = 0$), decreasing to 0 around $r \to R$. In the following we first determine the radial distribution function $v(r)$ of the velocity, for which it is envisaged to isolate a cylindrical thin flow layer within the fluid with inner and outer radii r and $r + \mathrm{d}r$ respectively, with side areas of $2\pi rl$ and $2\pi(r + \mathrm{d}r)l$ respectively, subject to viscous forces [1] respectively.

$$\begin{cases} f_r = -\eta \left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r 2\pi rl \\ f_{r+\mathrm{d}r} = \eta \left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r} 2\pi(r+\mathrm{d}r)l \end{cases}$$

Here the velocity gradient $\dfrac{\mathrm{d}v}{\mathrm{d}r} < 0$, the positive and negative signs in Eq. are determined by specifically analyzing the direction of the viscous forces on both sides of this thin layer. The combined force of the viscous forces on the flow layer is:

$$f = f_{r+\mathrm{d}r} + f_r = 2\pi\eta l \left[(r+\mathrm{d}r)\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r} - r\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r\right] = 2\pi\eta l \frac{\mathrm{d}}{\mathrm{d}r}\left[r\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)\right]\mathrm{d}r$$

In the case of constant flow this force should be balanced by the pressure difference on the end face $2\pi r\mathrm{d}r$.

$$(p_b - p_a)\,2\pi r\mathrm{d}r = 2\pi\eta l \frac{\mathrm{d}}{\mathrm{d}r}\left(f\frac{\mathrm{d}v}{\mathrm{d}r}\right)\mathrm{d}r \Longrightarrow \mathrm{d}\left(r\frac{\mathrm{d}v}{\mathrm{d}r}\right) = \frac{p_b - p_a}{\eta l} r\mathrm{d}r$$

The two sides are integrated from $r = 0$ to r, later then integrate from r to R, we get:

$$r\frac{\mathrm{d}v}{\mathrm{d}r} = \frac{p_b - p_a}{2\eta l} r^2, \text{ or } \mathrm{d}v = \frac{p_b - p_a}{2\eta l} r\mathrm{d}r$$

$$v(R) - v(r) = \left[\frac{p_b - p_a}{4\eta l} r^2\right]_r^R = \frac{p_b - p_a}{4\eta l}\left(R^2 - r^2\right)$$

Because $v(R) = 0$ on the wall of the tube, the final radial distribution of the flow velocity in the tube is obtained as:

$$v(r) = \frac{p_a - p_b}{4\eta l}\left(R^2 - r^2\right) \tag{7}$$

It has the form of a rotating paraboloid. Now calculate the flow rate, the flow rate through the circular area $2\pi r\mathrm{d}r$ is $\mathrm{d}Q_V = 2\pi v r\mathrm{d}r$, so the total flow rate in the tube is:

$$Q_v = 2\pi \int_0^R v(r) r\mathrm{d}r = \frac{\pi(p_a - p_b)}{2\eta l} \int_0^R \left(R^2 - r^2\right) r\mathrm{d}r = \frac{\pi}{8}\frac{p_a - p_b}{\eta l} R^4 \tag{8}$$

The average velocity $\bar{v}$ in the tube can be defined as the volume flow Q_V divided by the cross-sectional area $S = \pi R^2 : \bar{v} = \dfrac{Q_V}{\pi R^2}$. In addition, the pressure difference $p_a - p_b$ divided by the tube length l is the pressure gradient along the tube, and multiplied by the cross-sectional area of the tube S is the pressure gradient, expressed by $\dfrac{\Delta F}{\Delta l}$, so that by equation 8 we have:

$$\frac{\Delta F}{\Delta l} = \frac{p_a - p_b}{l}\pi R^2 = \frac{8\eta Q_V}{R^2} = 8\pi\eta\bar{v} \tag{9}$$

[1] 1 P = 1 dyn · s/cm^2 = 0.1 N · · · /m^2 = 0.1 Pa · s

4.2 Air Resistance model and Force Analysis of the high-speed train

4.2.1 *Air Resistance model*

Turbulent Flow and Reynolds Count

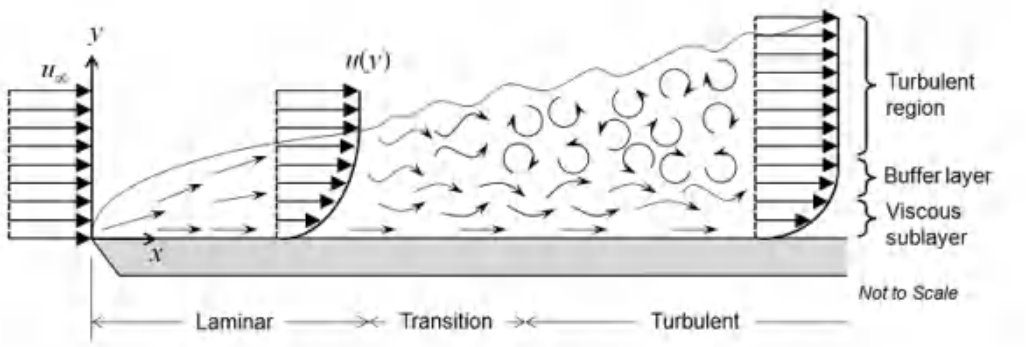


Figure 3 Turbulent flow schematic.

Turbulence we can think of as the motion of a superposition of various vortices. There are many models of turbulent flow, but they can be broadly grouped into the following three categories:

1. The first type is the *turbulent transport coefficient model*, proposed by Boussinesq in 1877 for two-dimensional flows, which expresses the second-order correlation of the velocity pulsation as the product of the average velocity gradient and the turbulent viscosity coefficient.
2. The second type is to discard the concept of turbulent transport coefficients and directly establish *transport equations* for turbulent stresses and other second-order correlated quantities.
3. The third category is *large turbulent simulations*. The first two types are based on the statistical structure of turbulence and statistical averaging of all vortices. The large turbulent simulations divides turbulence into large scale turbulence and small scale turbulence and obtains the characteristics of the motion of large eddies by solving the three-dimensional modified *Navier-Stokes Equation*, while the above approach is still adopted for the motion of small eddies.

Figure 3 is a diagram of the vortex flow[2], In practice, the model chosen for the solution is determined by the characteristics of the specific problem. The general principles of selection are high accuracy, simplicity of application, saving time of calculation, and generality at the same time.

In this paper, the viscosity of fluid is considered and the *standard* $k-\varepsilon$ *model*, which is the most widely used, stable and relatively accurate in engineering applications, is used to simulate the high speed rail and avoid the limitations of wall calculation.

Standard $k-\varepsilon$ Models

The the standard $k-\varepsilon$ model requires the solution of the turbulent kinetic energy and its dissipation rate equations. The turbulent kinetic energy transport equations are obtained by exact equation derivation, but the dissipation rate equations are obtained by physical derivation and mathematically simulating similar proto-equations. The model assumes that the flow is fully turbulent and the effect of molecular viscosity can be neglected. Therefore, the standard $k-\varepsilon$ model is only suitable for the

[2]Image from https://zhuanlan.zhihu.com/p/21338549

simulation of fully turbulent flow processes. The turbulent kinetic energy k and dissipation rate ε equations of the standard $k-\varepsilon$ model have the following form.

$$\frac{\partial}{\partial t}(\rho k)+\frac{\partial}{\partial x_i}(\rho k u_i)=\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_k}\frac{\partial k}{\partial x_j}\right)\right]+G_k+G_b-\rho\varepsilon-Y_M+S_k \tag{10}$$

$$\frac{\partial}{\partial t}(\rho\varepsilon)+\frac{\partial}{\partial x_i}(\rho\varepsilon u_i)=\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial\varepsilon}{\partial x_j}\right]+C_{1\varepsilon}\frac{\varepsilon}{k}(G_k+C_{3\varepsilon}G_b)-C_{2\varepsilon}\rho\frac{\varepsilon^2}{k}+S_\varepsilon \tag{11}$$

where G_k denotes the turbulent kinetic energy generation due to the mean velocity gradient; G_b denotes the turbulent kinetic energy generated by buoyancy; Y_M denotes the contribution of fluctuating expansions in compressible turbulence to the total dissipation rate; S_k and S_ε are user-defined source terms; $C_{i\varepsilon}$ $(i=1,2,3)$ denotes the constants; σ_k and σ_ε denote the turbulent Prandtl number of turbulent kinetic energy k and dissipation rate ε.

4.2.2 *Force Analysis*

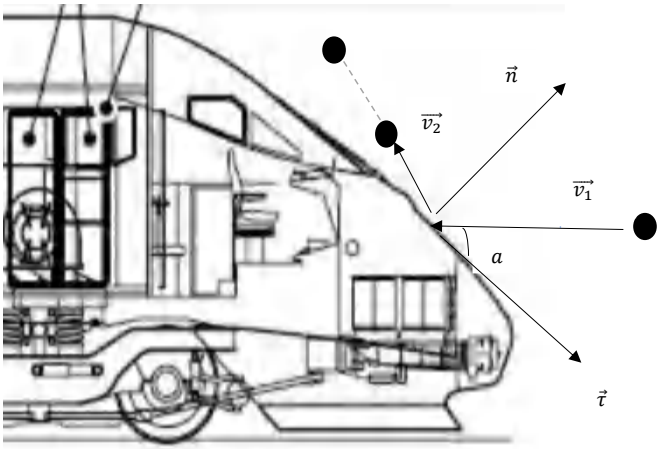


Figure 4 Particle-based force analysis.

Considering the extreme weather, such as rain, snow and wind conditions, because of the fast speed of the train relative to the rain and snow, so the rain and snow will also impact the head of the train at a very fast speed. *Figure 4* shows the trajectory of the rain and snow when it touches the surface of the head of the train, when the particle touches the wall, the tangential force component by $\vec{\tau}$ has a small value, and the surface normal direction $\vec{n}$ on the z axis provides the particle motion. When the collision nature is fully elastic, $|v_1|=|v_2|$ motion rate is the rate in front of the contact wall, while in the actual impact is inelastic, that is, $|v_1|>|v_2|$, so if you want to reduce the impact of rain and snow on the head of the train, you can choose the elastic material coating.

4.2.3 *Simulation of Conical and Four Typical models*

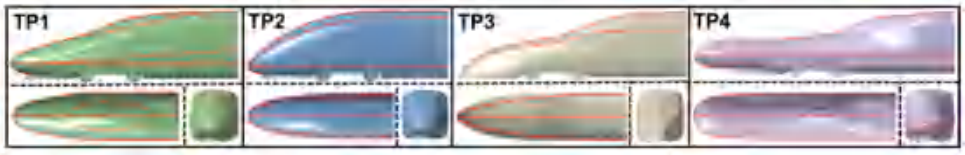


Figure 5 Simplified models of four typical high-speed train head structures[5].

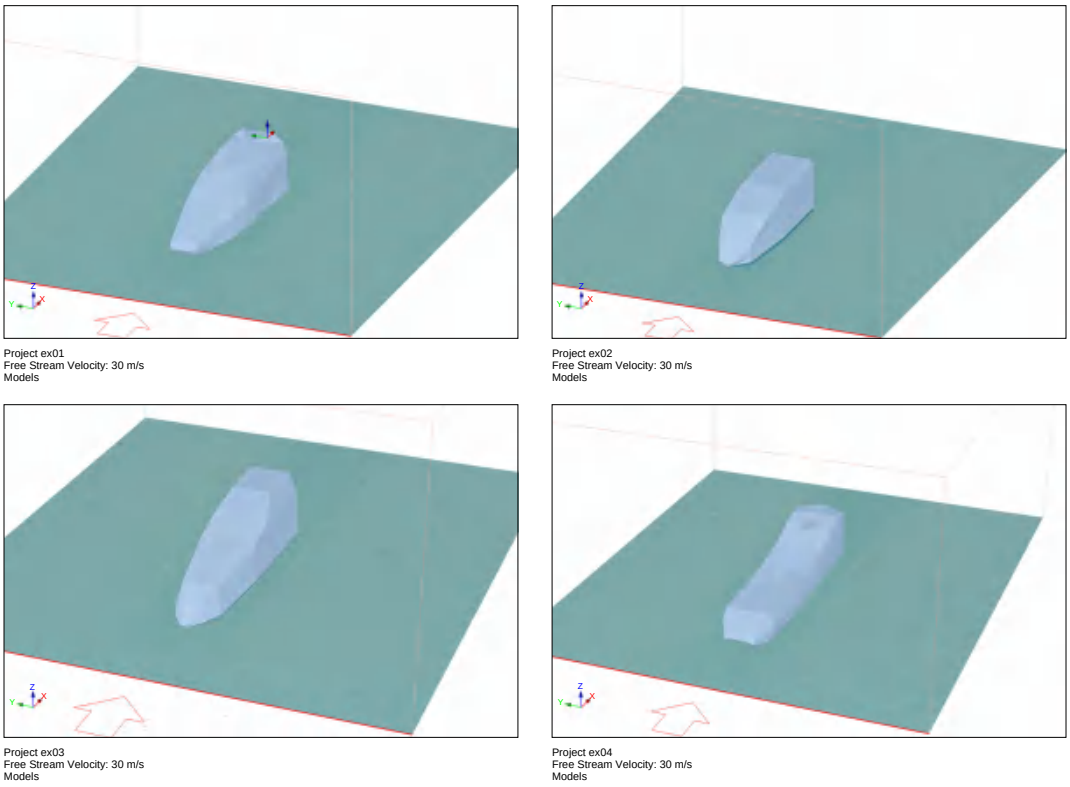
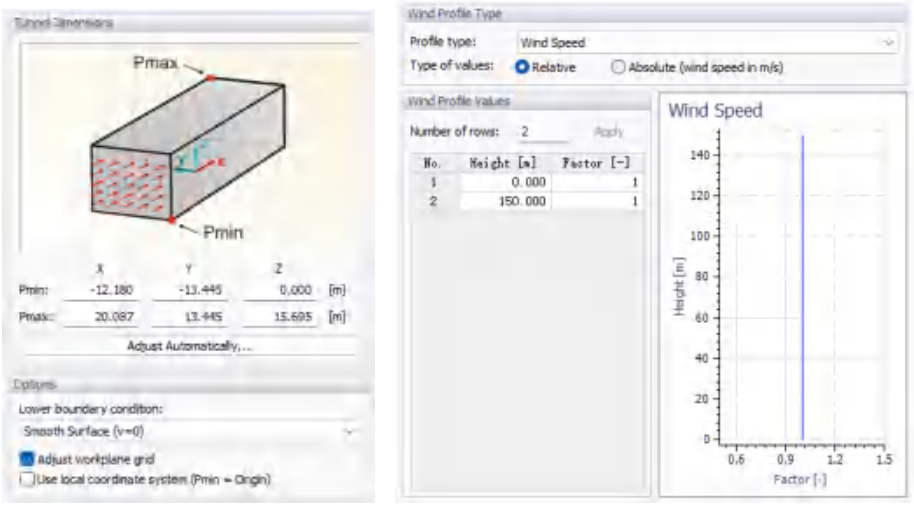


Figure 6 Use Blender software to create the models and then put them into RWIND.

We first create what we think the head of the high speed train will look like by using *Blender* software in conjunction with *Figure 5* provided in the topic, as shown in *Figure 6*. We first consider the general case that comes with the software to verify the feasibility of the four head models. After setting the corresponding parameters, we conduct wind tunnel experiments on each of the four models, as shown in *Figure 8*, it can be found that the vortex and the *Phenomenon of the Poiseuille* leaf appear at the rear of the front end, we continue to simulate the surface pressure of the front end and get as shown in *Figure 9*.



(a) Setting of simulation parameters (b) Setting of simulation parameters

Figure 7 Setup of RWIND.

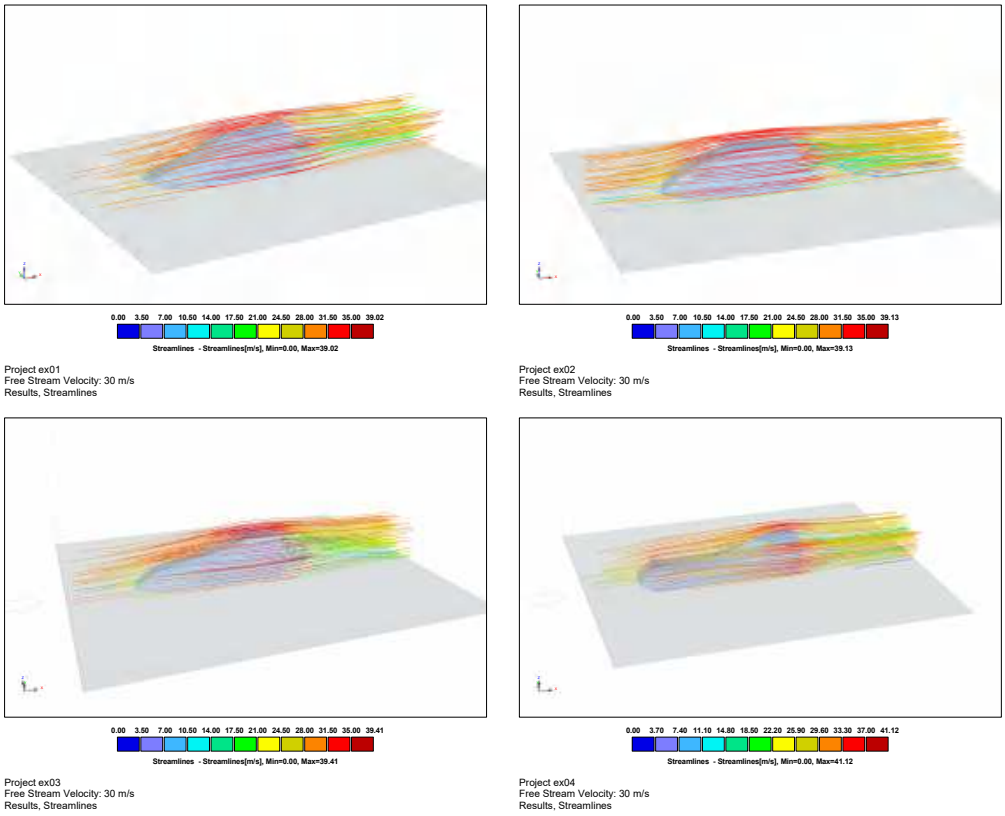


Figure 8 Simulate the streamline of the front surface with RWIND software.

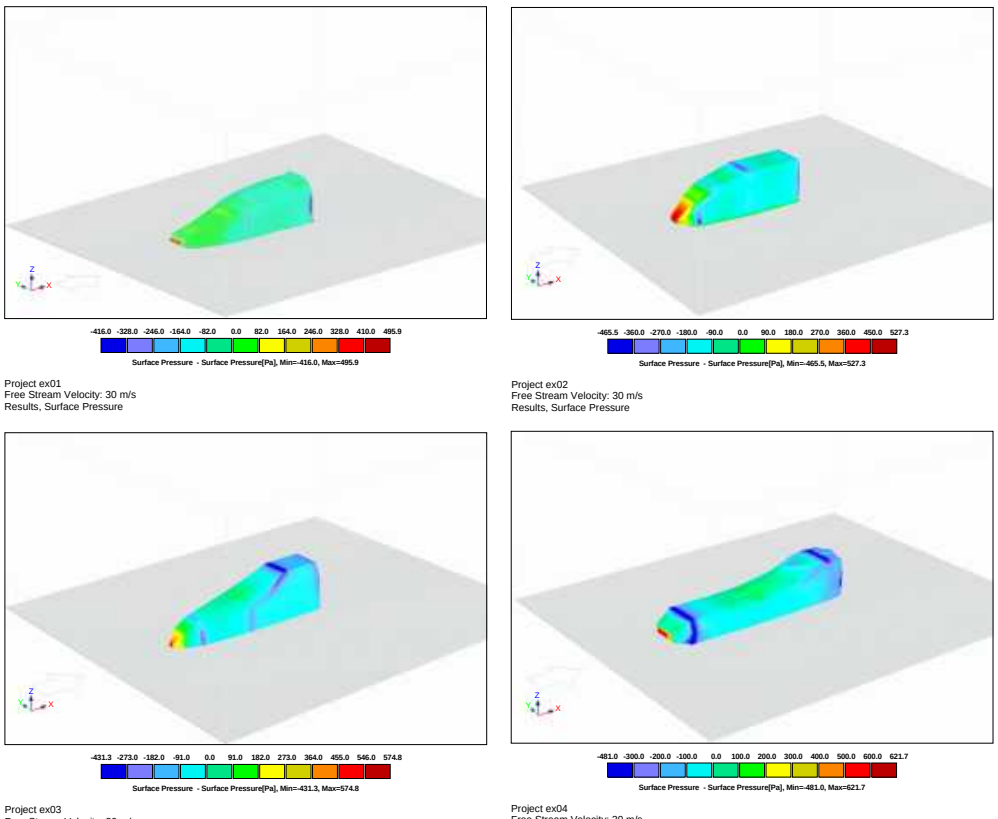
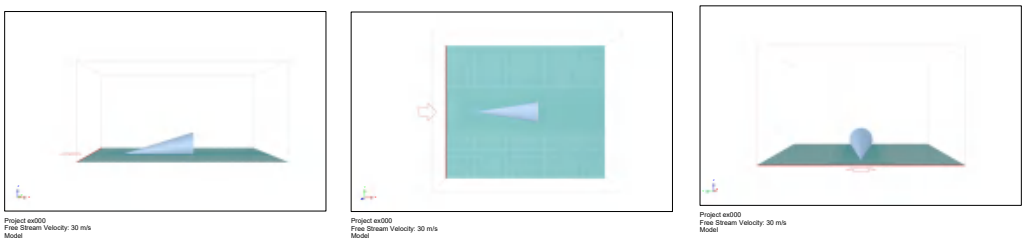


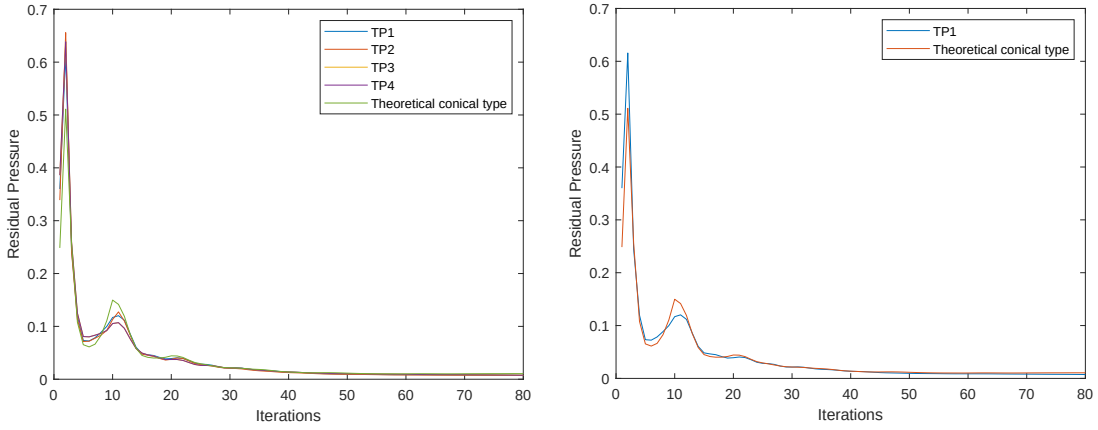
Figure 9 Pressure simulation of the head of a high speed train.



(a) Main View (b) Top view (c) Side view

Figure 10 Theoretical conical type.

As shown in *Figure 10*, we designed a theoretical head model according to the requirements of the topic, and used it to compare with other four types of typical high-speed trains. In combination with the *Figure 9*, we set up a set of TP1 and theoretical conical type for comparison.



(a) Comparison of four typical trains with theoretical conical (b) Comparison of TP1 high speed trains with theoretical conical

Figure 11 Comparison of high speed rail with theoretical conical.

The data of relevant simulations are exported from *RWIND* software, and we use the following formula to calculate the difference between four typical high speed rail $x_l^{(\mathrm{TP_i})}$ and the theoretical values x^*, and the results are presented in Table 1, where N equals 300.

$$\Delta = \sum_{j=1}^{N} \left(x_j^{(\mathrm{TP_i})} - x^* \right)^2 \tag{12}$$

Table 1 The difference between four typical high speed rail and theoretical values.

	TP1	TP2	TP3	TP4
Δ	0.0695	0.1031	0.0885	0.1120

From the difference Δ: Among these four typical HSRs, **TP1 is the best HSR shape with the lowest air resistance**. Detailed values can be found in *Table 6*.

4.3 Optimization model of high-speed train head

4.3.1 *Introduction to Pareto Optimality*

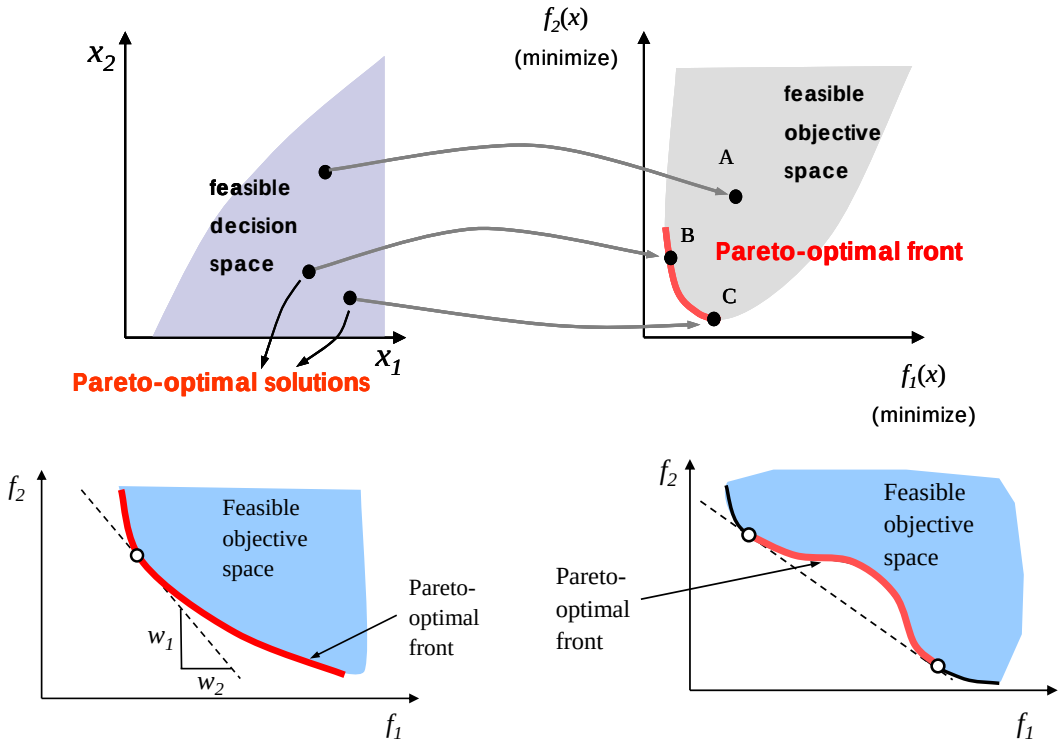


Figure 12 Pareto optimality schematic.

The deformation of each region is done sequentially, and after obtaining the coordinate increments of all the deformed regions, the deformed surface is obtained by adding the coordinate increments with the original coordinate values. In the optimization analysis, a total of 5 optimization variables were designed, which are *Figure 13* Zone1-5 (Because Zone6 belongs to the carriage part, this paper does not consider for the time being).

There are usually several methods for solving multi-objective planning as follows.

1. *multi-objective into single-objective*, synthesizing multiple objectives into a single objective function weighted by their importance, or using the ideal point method, i.e., first finding each single-objective optimal function value and then expressing the objective function as the Euclidean distance to these single-objective optima.
2. *stratified sequence method*, a sequence is given by the importance of the objectives, and each time the optimal solution of the next objective is found within the optimal solution set of the previous objective until a common optimal solution is found.
3. *Pareto optimality search*, as shown in *Figure 12* that is, the set of *Pareto solutions* is found [3], for Pareto solutions, the improvement of one goal is at the expense of the deterioration of the other goal.

[3]Image from https://hpzhao.github.io

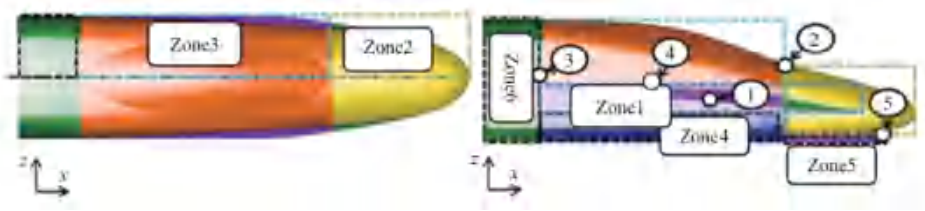


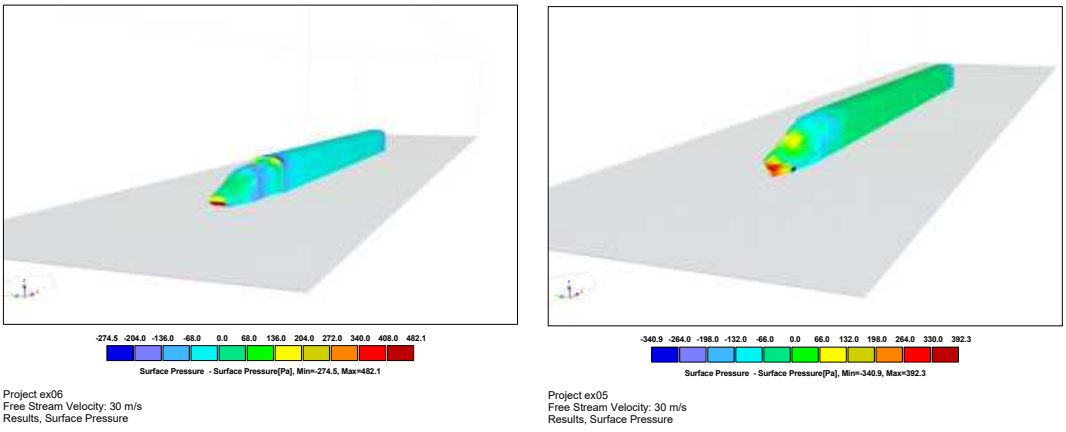
Figure 13 Different areas of streamlined high-speed train head structure[6].

4.3.2 *Optimal numerical solution*

The introduction of weighting factors to weight multiple objectives is transformed into a single-objective planning problem to be solved in a model with parameters, which is actually a series of models that vary with reference. Solving a series of models is achieved using the *for loop* and a series of useful results are saved. The following is a *Pareto weighting calculation*.

$$\min F(x) = \sum_{m=1}^{M} w_m f_m(x)$$

$$s.t. \begin{cases} g_j(x) \geq 0 & j = 1, 2, \cdots, J \\ h_k(x) = 0 & k = 1, 2, \cdots, K \\ x_i^{(L)} \leq x_i \leq x_i^{(U)} & i = 1, 2, \cdots, n \end{cases}$$



(a) Original model (b) Optimized model

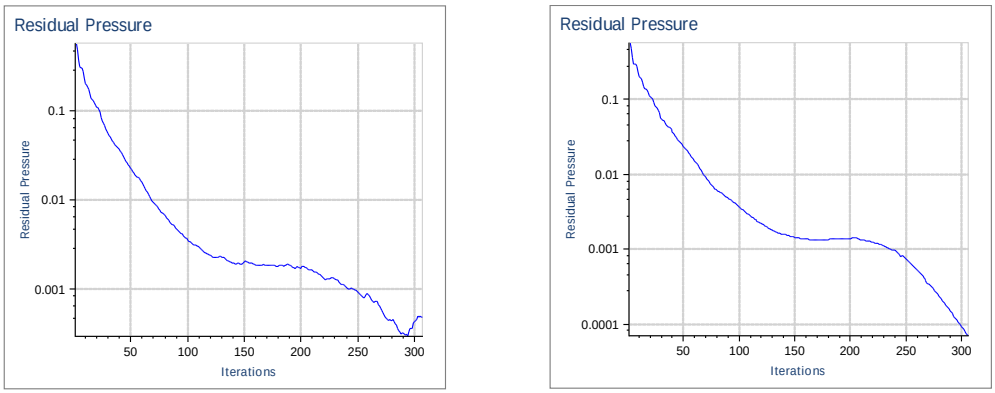


Figure 14 High-speed rail optimization before and after comparison chart.

4.4 Noise model of high-speed train

4.4.1 *Status of foreign high-speed train noise research*

In recent years,environmental noise pollution caused by traffic noise continues to increase and has become a serious problem in many countries, the following section will introduce respectively The current status of foreign research on high-speed train (railroad) noise.

Researchers in Germany believe that HSR operation noise has a greater impact on the environment along the line than the vibration caused by HSR, when the maximum speed reaches 300km/h or more , noise will increase significantly[4]. See *Table 2, 3, 4* for details.

Table 2 The relationship between high-speed train speed and noise level.

Speed/km · h^{-1}	Noise level/dB(A)
200	82
250	85
300	89
400	102

Table 3 Japan Railway Environmental Noise Standards.

Region	Shinkansen Railway (high-speed railroad) (environmental standards) L_{Amax}/dB(6:00-24:00)	New construction of general railroad (guidance) Daytime	Nighttime
Residential	70	60	55
Commercial and industrial areas	75	60	55

Table 4 France and Italy railroad noise standard limit value L_{pAeq}/dB.

Country	Daytime	Night
France (high speed)	60	55
France (new general speed)	63	58
France (highway conversion)	65	60
France (conversion of general speed)	68	63
Italy (new highway)	65	55
Italy (existing, reconstruction, new general speed)(0~100m)	70	60
Italy (existing, reconstruction, new general speed)(100~250m)	65	55

[4]High-speed rail model is ICE-V. The measurement point is 25m from the centerline of the track.

4.4.2 *Bow wave and Mach cone*

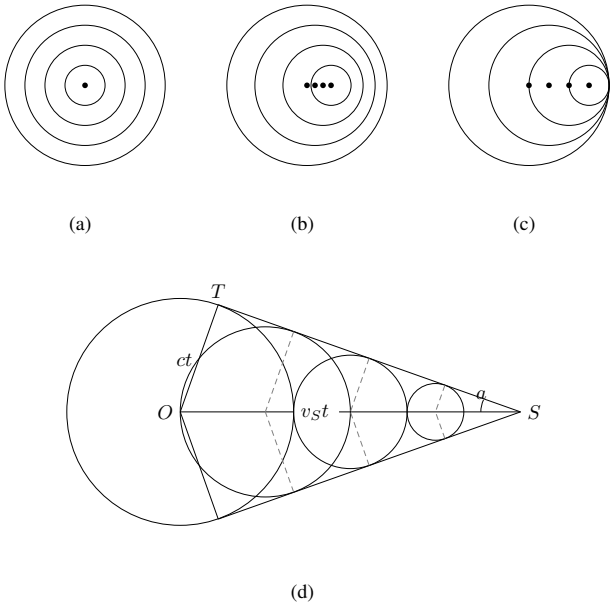


Figure 15 Wavefront of the moving wave source.

The *Figure 15a* the wave source is stationary, wavefront is concentric; In the *figure 15b* the wave source is in motion, but its speed is less than the wave speed, the center of the wavefront is misaligned, creating a *Doppler effect*; In the *Figure 15c* the velocity of the wave source tends to the wave speed and all wavefronts are tangent at one point, frequency $\nu' \to \infty$; In the *Figure 15d* the speed of the wave source exceeds the wave speed, the envelope of the wavefront is conical in shape and is called *Mach cone*. Since in this case the propagation of the wave does not exceed the moving object itself, mach cone surface is the leading edge of the wave, no disturbance waves outside of it. This form of fluctuation is called *Bow wave*. Such that the half-top angle of the *Mach cone* is α, As can be seen from the figure:

$$\sin\alpha = \frac{c}{v_S}$$

The dimensionless parameter $\frac{v_S}{c}$ is called the *Mach number*,it is a very useful parameter in aerodynamics. Air under high velocity impact can be considered as being under an adiabatic process, so the pressure change will trigger a change in air temperature. [5] In the humid air, In humid air, the temperature of the thinnest part of the shockwave air (the wall portion) drops below the dew point, causing the moisture in the air to condense quickly and present a visible cone of fog [6]. And present a visible cone of fog.

Although the speed of high-speed train operation is much less than the speed of sound, the wave surface intensifies the buildup in the interval of the forward direction of the high-speed train (Similar to *figure 15b*), it makes the air friction increase, and the noise will increase.

[5]The speed of sound in air is about $340 m/s$ at 1 standard atmosphere and $15°C$.

[6]In the temperature range where the adiabatic index and the gas constant change negligibly, the speed of sound is proportional to the temperature open root sign.

4.4.3 *High-speed rail external Aerodynamic Noise model*

The turbulent kinetic energy turbulent dissipation rate at each node in the steady-state flow field is extracted from the air resistance model in turbulence above, and the acoustic power at each node is determined by the expression[3]:

$$W_A = \alpha\rho\left(\frac{u^3}{l}\right)\frac{u^5}{c_0^5} \tag{13}$$

where α is a constant; l is the turbulence scale; u is the turbulence velocity; and c_0 is the speed of sound.

Then, the large eddy model (see 4.2.1) is used to simulate the transient flow field in combination with the steady-state flow field. The pressure distribution on the surface of the high speed rail is obtained, and finally, the prediction of the aerodynamic noise of the external flow field of the high speed rail is done by using *Lighthill-Curle acoustic analogy theory*, and the expression[4] is:

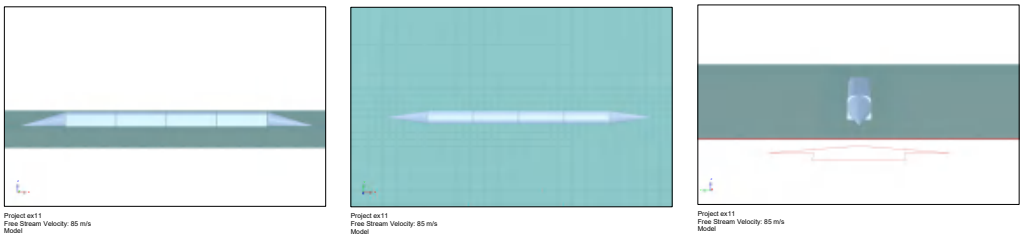
$$\rho'(\boldsymbol{x},t) = -\frac{1}{4\pi c_0^2}\frac{\partial}{\partial x_i'}\int_S \frac{n_j P_{ij}\left(\boldsymbol{y}, t-\frac{\boldsymbol{R}}{c_0}\right)}{\boldsymbol{R}}\mathrm{d}\boldsymbol{y} \tag{14}$$

$$P(\boldsymbol{x},t) - P_0 = c_0^2\rho'(\boldsymbol{x},t) \tag{15}$$

where $\boldsymbol{x}$ is the sound measurement point vector; t is the vector; $\boldsymbol{y}$ is the sound source point vector; n_j is the directional cosine of the vertical high iron outer wall outward (pointing to the outflow field); P_{ij} is the pulsating pressure of air on the outer surface of the body; $\boldsymbol{R} = \boldsymbol{x} - \boldsymbol{y}_j$; P and P_0 are the sound pressure and the reference sound pressure ($P_0 = 2\times10^{-5}$Pa).

4.4.4 *Aerodynamic noise Simulation*

To better reflect the relationship between high-speed train model and external noise, we reset the wind tunnel speed to 85m/s (306km/s), which is also processed to correlate with the data in *Table 2* for the train model. And we follow the shape of *Figure 10* to produce a theoretical conical type of high speed rail model as shown in *Figure 16* (for the calculations below).



(a) Main View (b) Top view (c) Side view

Figure 16 High-speed rail model based on theoretical conical type.

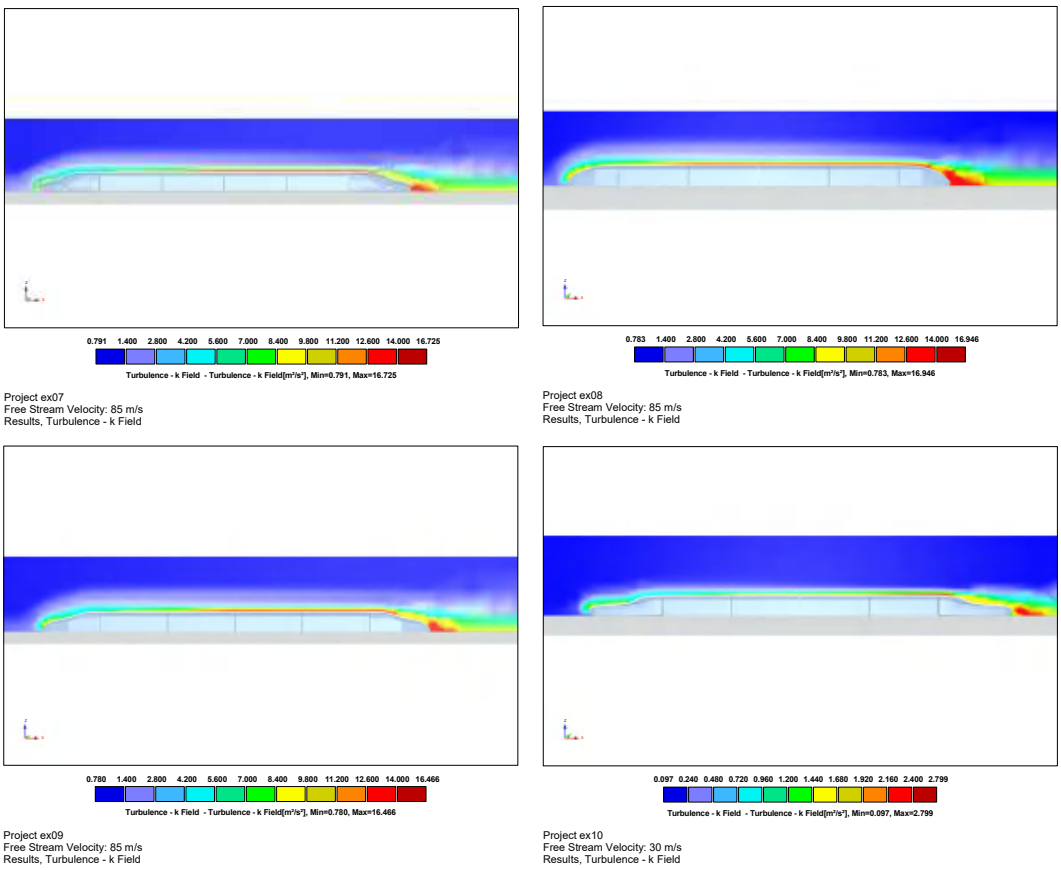


Figure 17 Turbulence - k Field - Turbulence - k Field (Images)[m^2/s^2]**.**

Table 5 Turbulence - k Field - Turbulence - k Field (Values)[m^2/s^2]**.**

	TP1	TP2	TP3	TP4
min	0.791	0.783	0.780	0.097
max	16.725	16.946	16.466	2.799

Compare *Figure 17* and *Table 5*. It can be found that the strong value region of TP4 high speed rail is the smallest, and not only the minimum value is the best, but also the maximum value is much more desirable than the other three models. Therefore, **TP4 is the best shape of high speed rail that generates the least noise**. TP4 is designed to comply with the relevant regulations in Japan to a certain extent.

In summary, it is necessary to analyze the TP1 and TP4 trains more thoroughly and combine their advantages to design a new type of high-speed train that takes into account both speed and noise reduction.

4.5 Optimized model for both Resistance and Noise

4.5.1 *Introduction to Multi-Objective Particle Swarm Algorithm*

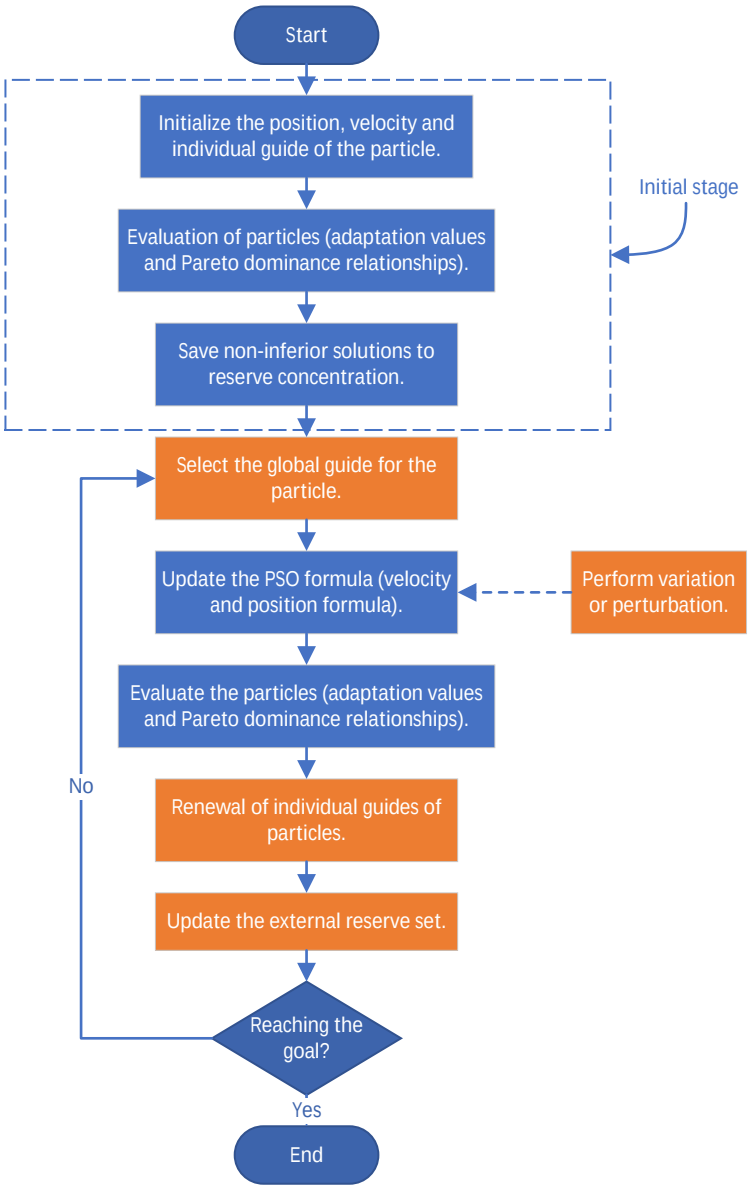


Figure 18 The basic workflow of the MOPSO algorithm.

The expected return and variance (or standard deviation) of all Pareto solutions are expressed on a two-dimensional plane, giving a curve called the Pareto Frontier, where the points above the Pareto Frontier are unavailable and the points below the Pareto Frontier, corresponding to the portfolio, are always available.

In order to obtain a non-inferior solution set that covers the entire search space, is uniformly distributed and close to the true Pareto front with as few computational resources as possible is the goal of the MOPSO algorithm. In this paper, this algorithm will be adopted to adjust the train head model that takes into account both speed and noise.

4.5.2 *Numerical solution of the model*

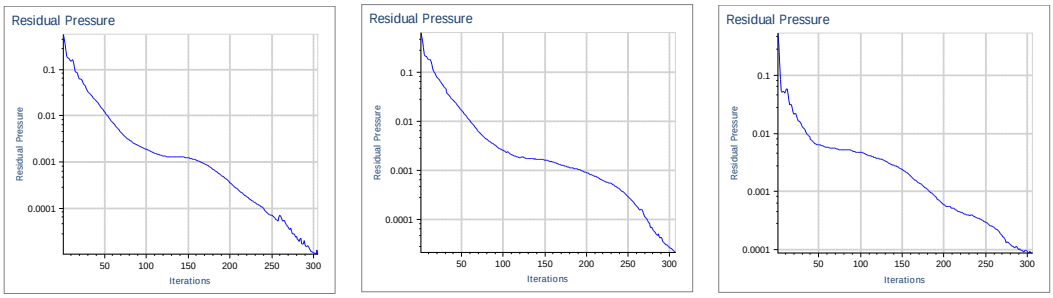


Figure 19 Numerical simulation of TP1 (left) and TP2 (middle) high speed rail with carriages. And numerical simulation of an optimal high-speed train model (right) that balances speed and noise

From the above analysis, TP1 has less air resistance property and TP4 has less noise property, so we made *Figure 17 1 and 4* the relationship between Residual Pressure and iterations as shown in *Figure 19 1 and 2*. We find that the first and last parts of the TP1 curve converge faster, and the middle part of the TP4 curve is flatter, and the designed high speed rail model inherits these two properties. The following are the relevant wind tunnel simulation values for the model, with specific size reference *Figure 23*. Combine *Figure 16* get $\Delta_{TP_1} = 0.0105$, $\Delta_{TP_4} = 0.0031$, $\Delta_{TP_{best}} = 0.0029$.

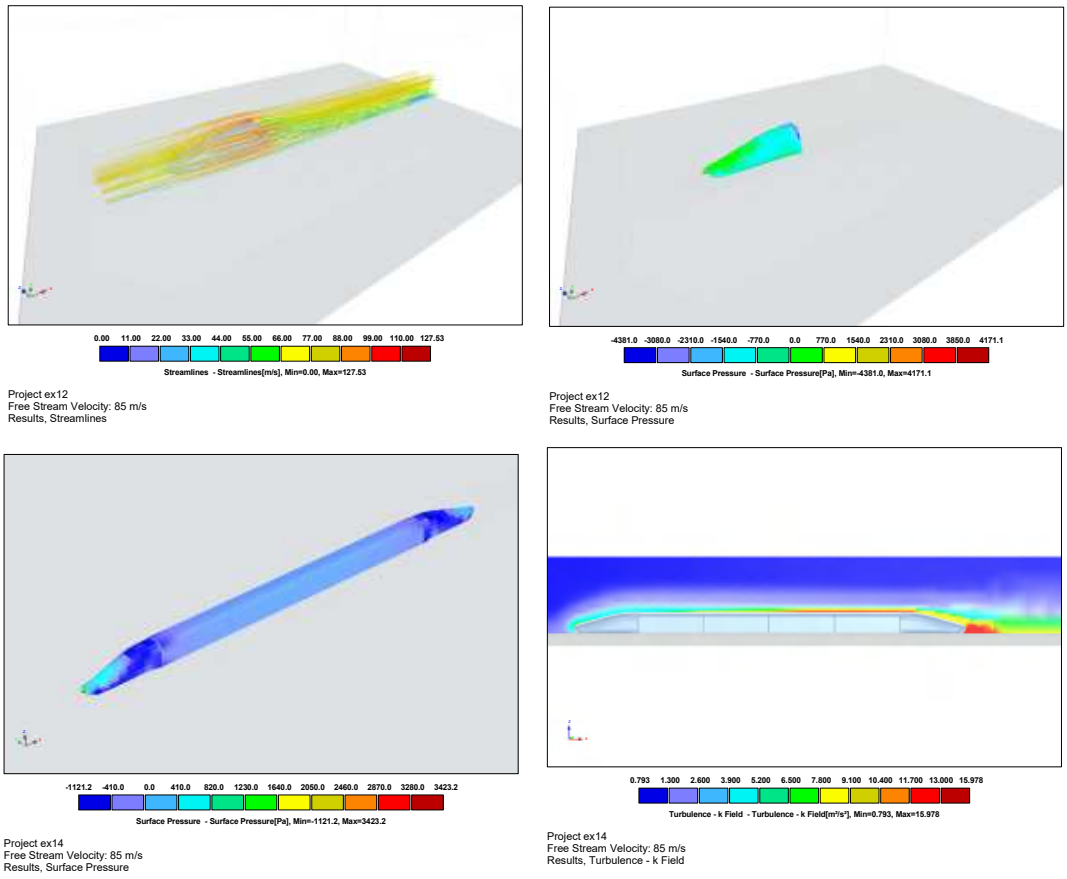


Figure 20 Relevant test data of the integrated and optimized high speed rail model in the wind tunnel.

4.5.3 *Internal adjustment of the integrated optimization train model*

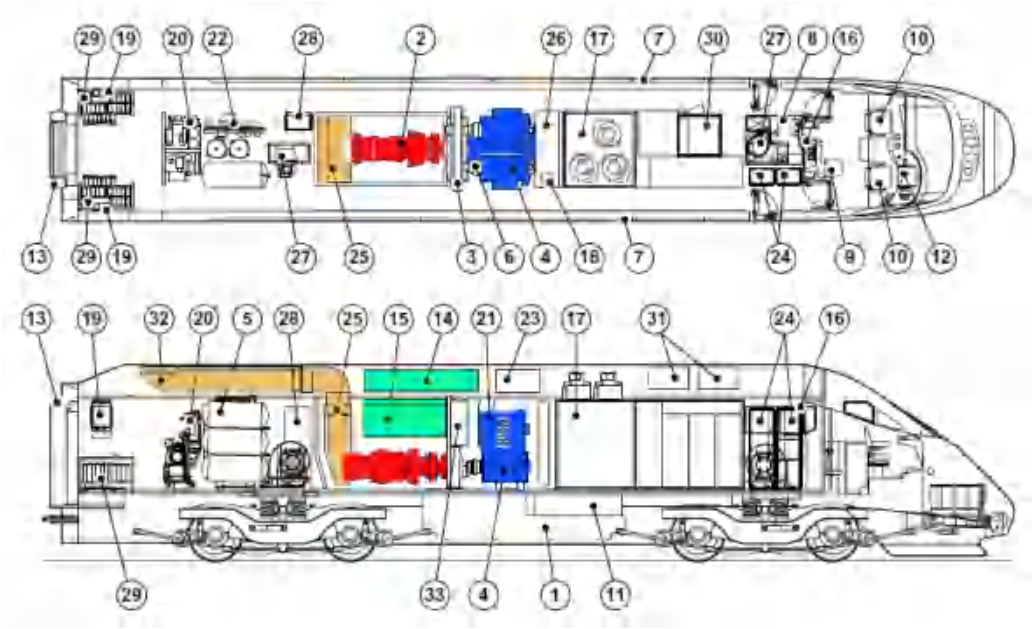


Figure 21 Schematic diagram of the general train interior structure.

Figure 21 is a schematic diagram of the general train interior equipment. For the comprehensive optimization model, we should consider not only its external aerodynamic design, but also the adjustment of the corresponding equipment inside the train.

Comparing the models before and after optimization, it can be found that the nose cone height decreases, the nose cone length increases, and the nose cone becomes smoother; the driver's chamber height is adjusted downward, and the front half of the streamline shape becomes narrower and the width of the rear half increases. So we have to adjust the device numbers accordingly: 8, 9, 10, 12, 16, 24, 27 are rearranged; other device numbers of the corresponding new models are designed.

V. Evaluation of the models

5.1 Advantages of the model

1. The established model is closely related to the actual, general and generalization is strong;
2. The consideration of the model is very comprehensive, the simulation results are reasonable and the results are accurate;
3. The use of wind tunnel experiments, can accurately control the experimental conditions, high precision experiments;

5.2 Disadvantages of the model

1. There is a certain amount of error;
2. For the model can not do exactly the same;
3. There is still a lot of room for improvement.

VI. Optimization of the model

Model for Problem 1

The force analysis can be considered as a three-dimensional case, no longer limited to two-dimensional simple mechanical analysis, and can spend more time in the modeling.

Model for Problem 2

The process of Pareto search can be mentioned over the algorithmic implementation, as well as a more detailed theoretical portrayal of the Pareto frontier can be attempted.

Model for Problem 3

More parameters can be compared, making the results more convincing than mentioning min and max and simple wind tunnel simulations to make the case.

Model for Problem 4

The algorithm of multi-objective particle swarm needs to set the initial number of particle swarms according to the width and breadth of the optimization search, and the particle swarm algorithm is easy to fall into the local optimal solution.

VII. Future work

Future research should more carefully consider the potential impact of high-speed train on environmental noise, such as when it passes through caves or forests. And how to design a new type of environmentally friendly and fast high-speed train needs to be studied in more depth and detail.

For the optimization methods described in the paper, the reader can also use excellent modern optimization algorithms such as Simulated Annealing Algorithm or Heuristic Algorithm.

In this paper, we analyze the optimization process of high-speed trains and hope that it will be useful for scholars who are researching in this field. We will continue to pay attention to the research results in this field in the future, and we hope that more and more scholars will participate in it and contribute to the cause they love.

VIII. References

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein. Introduction to Algorithms Third Edition [M], China Machine Press, 2013:170-190.

[2] 赵凯华, 罗蔚茵. 新概念物理教程·力学 (第二版) [M], 北京：高等教育出版社, 2004: 234-235.

[3] Lighthill M J. On sound generated aerodynamically.part II :turbulence as a source of sound[J]. Proceedings of the Royal Society of London,1954 ,222:1-32.

[4] Sarkar S,Erlebacher G, Hussaini M Y,et al. The analysis and modelling of dilatational terms in compressible turbulence[J].Joural of Fluid Mechanics,2006,227(12):473-493.

[5] Sun, Zhenxu, Shuanbao Yao, Lianyi Wei, Yongfang Yao, and Guowei Yang. 2021. ”Numerical Investigation on the Influence of the Streamlined Structures of the High-Speed Train’s Nose on Aerodynamic Performances” Applied Sciences 11, no. 2: 784.

[6] Sun Zhenxu, Yao Yongfang, Guo Dilong, Yang Guowei, Yao Shuanbao, Zhang Ye, Chen Dawei, Li Guibo, Shang Keming, Jia Ling. RESEARCH PROGRESS IN AERODYNAMIC OPTIMIZATION OF HIGH-SPEED TRAINS 1. Chinese Journal of Theoretical and Applied Mechanics[J], 2021, 5311: 51-74 DOI:10.6052/0459-1879-20-205.

```
  \node[above]at(0.67,1.89){$T$};
  \node[left]at(0.34,0.92){$ct$};
  \draw[black] (5.3,0) arc (180:155 :0.5) node[above, midway] {$a$};
  \node[left]at(0,0){$O$};
\end{tikzpicture}
```

Table 6 Simulation values of four typical high speed trains.

No.	TP1	TP2	TP3	TP4
1	0.36000	0.33897	0.28682	0.38613
2	0.61573	0.65582	0.61214	0.63859
3	0.24439	0.24061	0.25190	0.26011
4	0.11960	0.11287	0.10850	0.12508
5	0.07326	0.07139	0.06738	0.08075
6	0.07238	0.07194	0.07130	0.08020
7	0.07843	0.07752	0.08235	0.08330
8	0.08850	0.08329	0.09854	0.08680
9	0.09990	0.09240	0.11668	0.92158
10	0.11683	0.11307	0.14264	0.10541
...	...	...	...	...
39	0.01414	0.01301	0.01213	0.01407
40	0.01346	0.01254	0.01189	0.01368
41	0.01318	0.01220	0.01181	0.01334
...	...	...	...	...
71	0.00802	0.00763	0.00873	0.00845
72	0.00794	0.00757	0.00866	0.00836
73	0.00784	0.00752	0.00861	0.00827
74	0.00776	0.00746	0.00869	0.00819
75	0.00770	0.00741	0.00872	0.00809
76	0.00769	0.00733	0.00880	0.00799
77	0.00760	0.00726	0.00891	0.00788
78	0.00755	0.00722	0.00902	0.00778
79	0.00745	0.00722	0.00911	0.00765
80	0.00735	0.00725	0.00918	0.00753

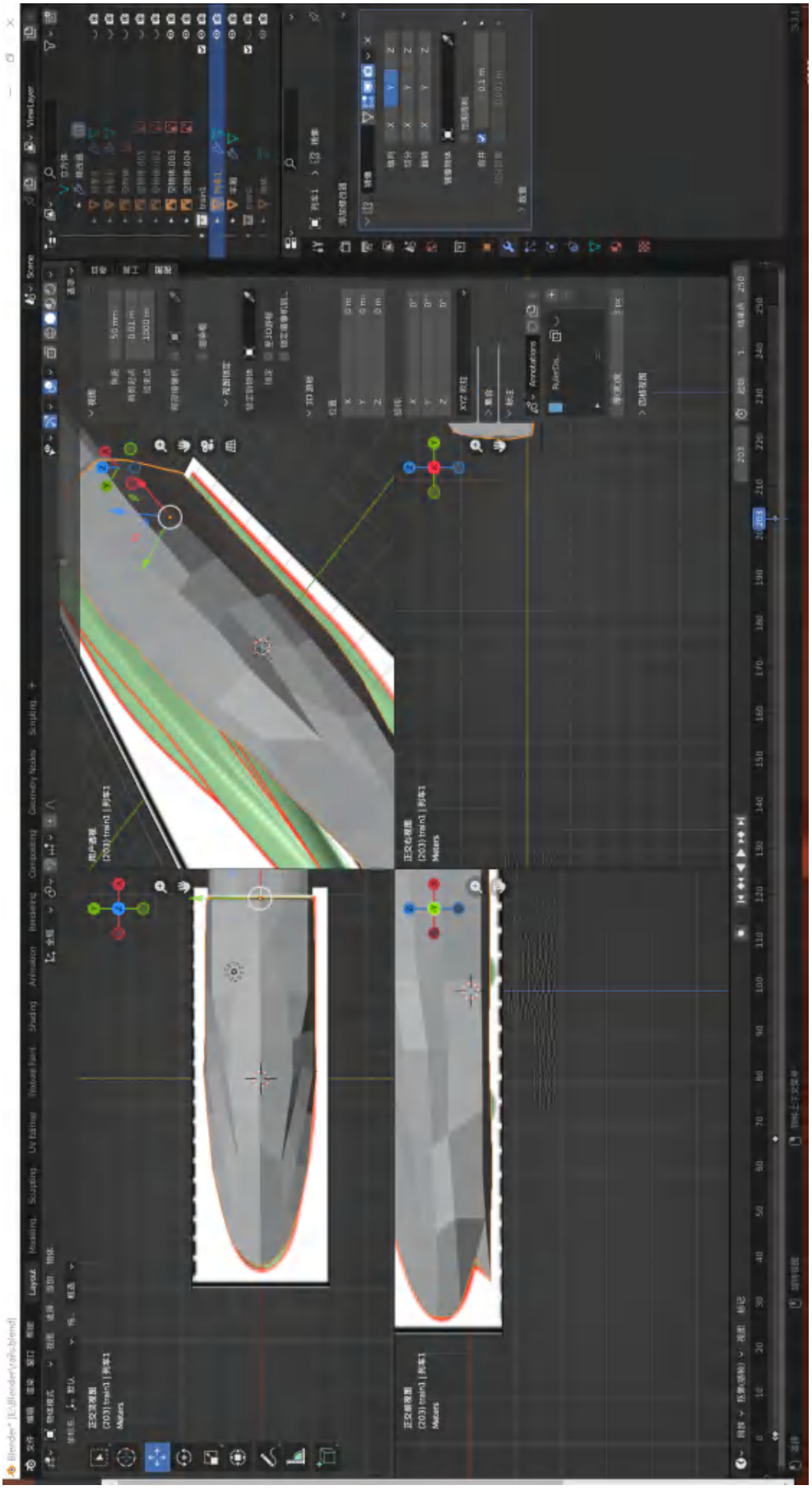


Figure 22 Production process.

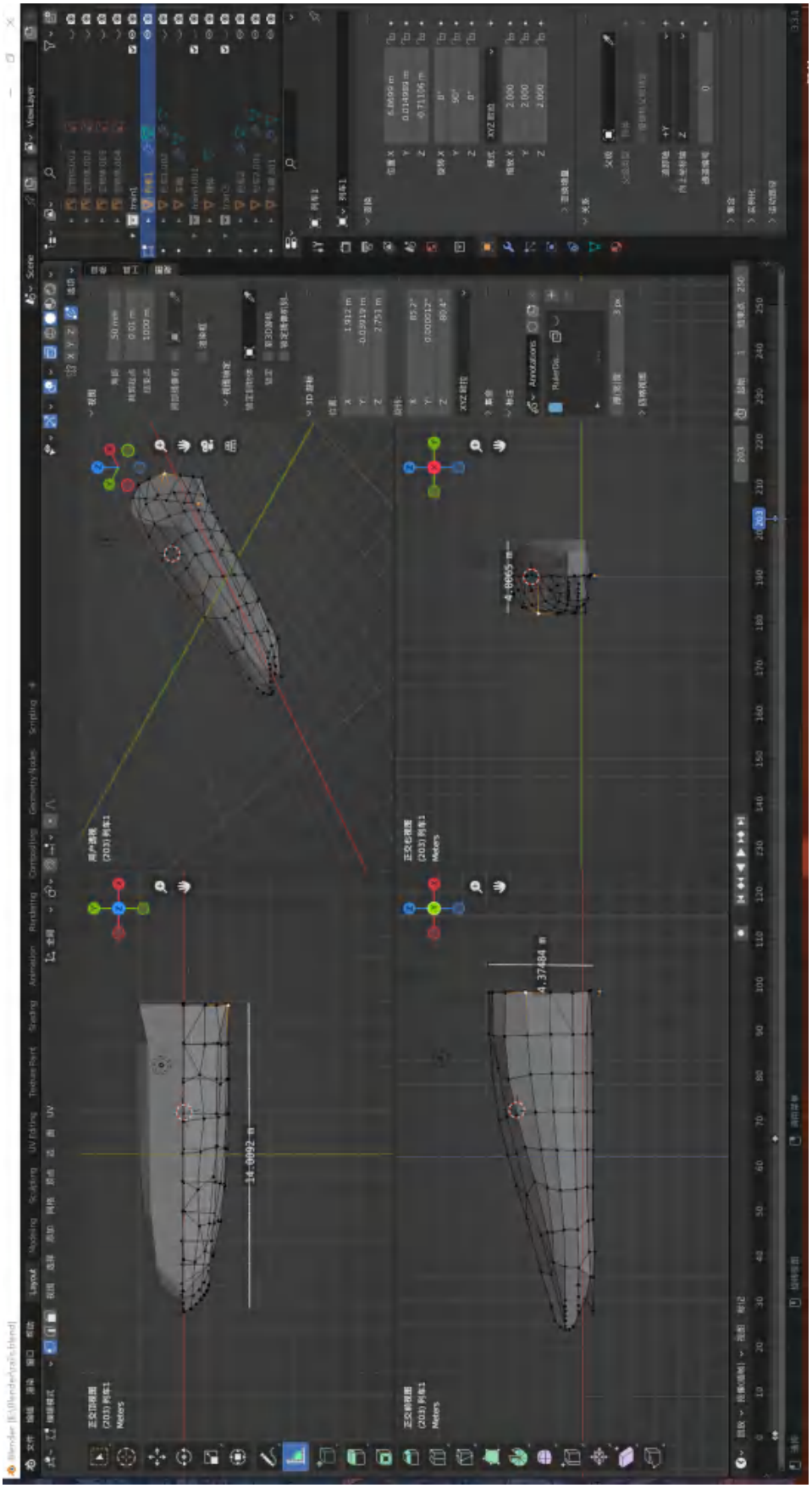


Figure 23 The sketch of the shape of the high speed rail is the corresponding structural parameters.

Team Number :	apmcm2203180
Problem Chosen :	C

2022 APMCM summary sheet

Abstract:

This paper combines the knowledge of relevant statistics, based on various methods such as pivot analysis, correlation analysis, t-test, multiple regression model, ARIMA time series model, random forest algorithm, GA-BP neural network algorithm and other software such as MATLAB, SPSS and EXCEL, to solve the problems extended by "Global Warming OR Not?"

In response to question 1, we first investigated the validity of the global warming argument using a normal distribution test and a one-sample t-test and conducted a data perspective analysis to determine whether the increase of global temperature in March 2022 resulted in a larger increase than observed over any previous 10-year period. And then, on the basis of the data collection and processing of global temperature impact factors and correlation analysis, we established a multiple regression prediction model, an ARIMA time series prediction model, and a random forest-based nonlinear prediction model, working together to complete a forecast of future global temperatures. After the accuracy assessment and sensitivity analysis with confidence intervals ranging from 95% to 105%, the non-linear prediction model based on the Random Forest algorithm was selected to further study the future global temperature: the global average temperature will reach 20 degrees in 2055. By 2050, the global average temperature is predicted to be 19.96 degrees and by 2100, the global average temperature is predicted to be 20.07 degrees.

In response to question 2, the historical temperatures around the world were analysed and considering the multiplicity of variables, which do not apply to traditional mathematical models, the GA-BP neural network algorithm was introduced to prove the Spatio-temporal relationship of the global average temperature. Four representative points were selected as criteria for the analysis, and the climatic tendencies of the different geographical locations were highlighted through data insights and trend lines.The analysis is then based on the uncertainty of the temperature, the time series graphs and the differential relationships to investigate the smoothness and periodicity of the temperature changes, and then the volcanic eruption as a typical representative of natural disasters, to explore its relationship with the global temperature. Based on the above study, the global temperature impact factors were further analysed and discussed, and six major impact factors were identified. Finally, based on the above conclusions, we propose specific initiatives to curb global warming mitigation at the global level.

In response to question 3, based on the specific analysis and discussion of the surface phenomena and intrinsic incentives of global warming in questions 1 and 2, the above conclusions are further elaborated and a series of feasible suggestions for the future are proposed, taking into account the real-life conditions and applying the conclusions to real-life practice.

In this paper, we study a series of global warming problems from a multidimensional perspective, and use relevant data indicators to synthesize various models, optimize and improve them from multiple perspectives, and complete the evaluation of their accuracy, rationality and sensitivity. At the end of the paper, the applicability of the model is extended and its reference value in real life is improved.

Key words: T-test, Multiple regression, ARIMA time series, Random Forest, GA-BP Neural Network

Content

1. Introduction

1.1 Problem background

In recent years, many places have broken historical records for high temperatures, and the instability of the global climate system has increased, with frequent, widespread, intense and concurrent extreme weather and climate events.

Global warming has moved from the 'behind the scenes' to the 'in front of the stage'. Global warming is a phenomenon associated with nature. It is caused by the continued accumulation of the greenhouse effect [1] , which leads to an imbalance in the amount of energy absorbed and emitted by the Earth-atmosphere system, and the continued accumulation of energy in the Earth-atmosphere system, which leads to an increase in temperature and global warming.

Global warming is a major environmental problem facing mankind, with a wide range of influence, long duration, complex constraints and serious consequences, which is related to the social and economic life of mankind and has attracted the high attention of scholars, the public and governments worldwide. To investigate the impact of global warming, predict global temperature changes, formulate reasonable and efficient development strategies based on scientific evidence, and reduce and control greenhouse gas emissions is an urgent issue to achieve sustainable human development.

1.2 Restatement of the Problem

Problem 1

Investigate the validity of global warming claims, determine whether the increase of global temperature in March 2022 resulted in a larger increase than observed over any previous 10-year period, develop multiple models to complete descriptions of past global temperature levels and projections of future global temperatures, conduct point-specific analyses, and complete assessments of model accuracy and precision

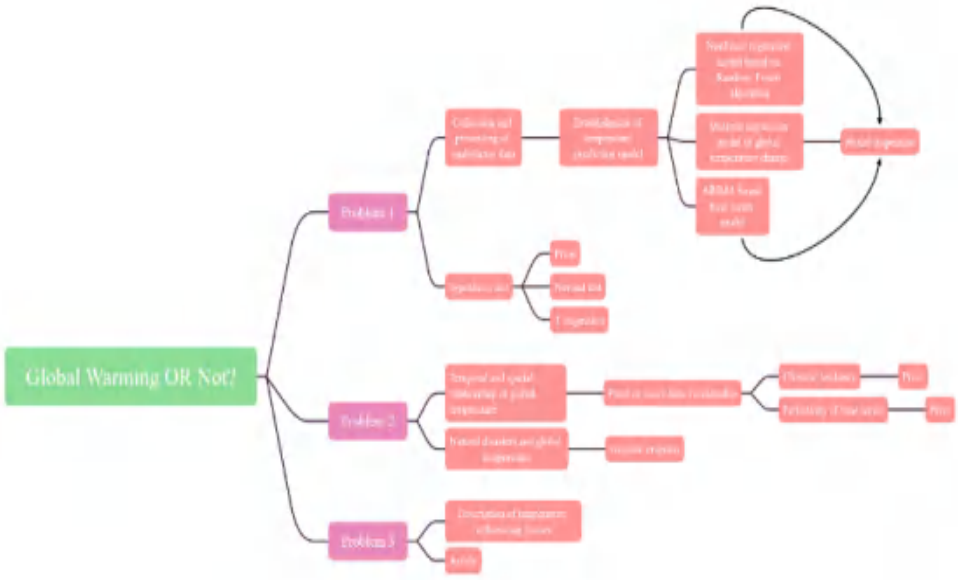
Problem 2

investigate the causes of global temperature change and develop a mathematical model to analyze and explain in detail the relationship between global temperature, time and location, and then, through the collection and analysis of relevant data, determine whether natural disaster factors (such as volcanic eruptions, forest fires and Corona Virus Disease 2019) have an impact on global temperature and, based on these findings, propose effective measures to curb global warming.

Problem 3

Based on the findings of questions 1 and 2, and taking into account their relevance, write a non-technical paper to the APMCM organising committee explaining the findings of the survey and making recommendations for the future.

1.3 Our work



Picture 1 Mind map

1.3.1 Problem 1

Based on the requirements of question 1, firstly, by collecting relevant data and hypothesis testing methods, compare and explore the range of global temperature rise in the past ten years, and verify the statement of "global warming", and then select Relevant impact factors are selected according to the statistical correlation, and the establishment of three prediction models is completed on the basis of scientific statistical analysis. Finally, through the analysis of sensitivity and error, the models are evaluated and selected and conclusions are drawn.

1.3.2 Problem 2

Based on the requirements of question 2, in order to analyze the temporal and spatial variation of historical temperatures around the world, first of all, considering the vast land and many observation points, select a few representative points as the analysis standard, and analyze the climate of different geographical locations. Then, the stability and periodicity of temperature changes will be discussed, and the corresponding time series will be constructed to further analyze the relationship between global temperature, time and location. Then, volcanic eruptions will be selected as typical representatives of natural disasters. Based on the above research, further analysis and discussion were carried out on the factors affecting global temperature, the influencing factors were determined, and specific measures to curb and slow down global warming were proposed on a global scale.

1.3.3 Problem 3

Based on the specific analysis and discussion of the surface phenomenon and internal incentives of global warming based on Problem 1 and Problem 2, combined with realistic conditions, the conclusions are applied to life practice, further elaborating on the above conclusions, and putting forward some constructive suggestions for the future.

2. Assumptions and Justifications

Assumption 1： It is assumed that the temperature data in each region is real and valid, and the sampling error is small. Only when the model is built on real data can reliable prediction results be obtained.

Assumption 2： It is assumed that the influence of other factors such as terrain on the temperature is not considered. The model selects some factors that have a greater impact on temperature as

independent variables. Only by ignoring the influence of other factors can a reasonable prediction equation be obtained.

Assumption 3： It is assumed that the human factors affecting the climate in the next 87 years starting from 2013 will not experience excessive variation. If there is too much variation due to human factors, climate change will not obey this rule and lead to distortion.

Assumption 4： It is assumed that the non-human factors affecting the climate will not have excessive variation in the next 87 years starting in 2013. If the non-human factors have excessive variation, the climate change will not obey this law and cause distortion. If we want to base on the current climate data To predict future data, it is necessary to keep non-human factors from large variations.

Assumption 5： Assume that the data obtained from the search has a certain degree of credibility and rationality. The data used comes from the data provided by the organizing committee and the OurWorldInData.com website, and the data is relatively complete.

3. Notations

Table 1 Symbol Description

Symbol	Explanation
Temperature	Global mean land and ocean surface temperatures
Population	global population
CO2	Carbon dioxide emissions
Urbanization	urbanization rate
Land_use_area	land use area
Forest	Forest cover rate
SO2	Sulfur dioxide emissions
leaves	The number of leaves used in the model
trees	The number of trees used in the model
pk	The sample data selected by the random forest model
k	class probability
UCL	upper bound on the temperature forecast
LCL	The lower bound for temperature prediction
Xi	predictor variable
Yi	response
Wi	connection weight
y	objective function
x	dimension
wt	weight

4. Data Desceiption

4.1 Data Collection

In order to analyze the relationship between various factors and global climate change, relevant data sets need to be collected. Due to the different statistical departments of different elements, the following table 4-1 shows the sources of data sets used in this chapter and the meanings of variables.

Table 2 Dataset source

Global Average Temperature (Celsius)	Berkeley Earth
Degree of urbanization	Ourworldindata website
Ozone hole area	Ourworldindata website

Sulfur dioxide	Ourworldindata website
Forest cover rate	Ourworldindata website
Built-up area	Ourworldindata website
Grazing area	Ourworldindata website
Cultivated area	Ourworldindata website
Annual carbon dioxide emissions	Ourworldindata website

4.2 Data Cleaning

4.2.1 Data missing value processing

Due to the different departments and cycles of data recording, the amount of data in different data sets is not the same. In order to make the data sets of each variable available, the missing values in the data sets should be dealt with accordingly. In this paper, the regression interpolation method based on the serial trend is used to deal with the missing values of the data.

4.2.2 Data standardization processing

Since the units and orders of magnitude of each variable are different, in order to facilitate the construction of a multivariate regression model, the data of each variable can be standardized. In this paper, z-score standardization is used to process the data. It can be represented by the following formula:

$$z = (x - \mu) / \sigma \qquad (1)$$

x is a specific index of the variable, μ is the mean of the data sequence, and σ is the standard deviation of the data sequence.

4.2.3 Factor correlation analysis chart

	Temperature	Population	CO2	Urbanization	Land area	Forest	SO2
Temperature	1	0.785	0.761	0.847	0.847	-0.897	0.724
Population	0.785	1	0.994	0.992	0.787	-0.954	-0.244
CO2	0.761	0.994	1	0.984	0.731	-0.932	-0.331
Urbanization	0.847	0.992	0.984	1	0.836	-0.98	-0.184
Land area	0.847	0.787	0.731	0.836	1	-0.914	0.367
Forest	-0.897	-0.954	-0.932	-0.98	-0.914	1	0.032
SO2	0.724	-0.244	-0.331	-0.184	0.367	0.032	1

Picture 2 Correlation coefficient heat map

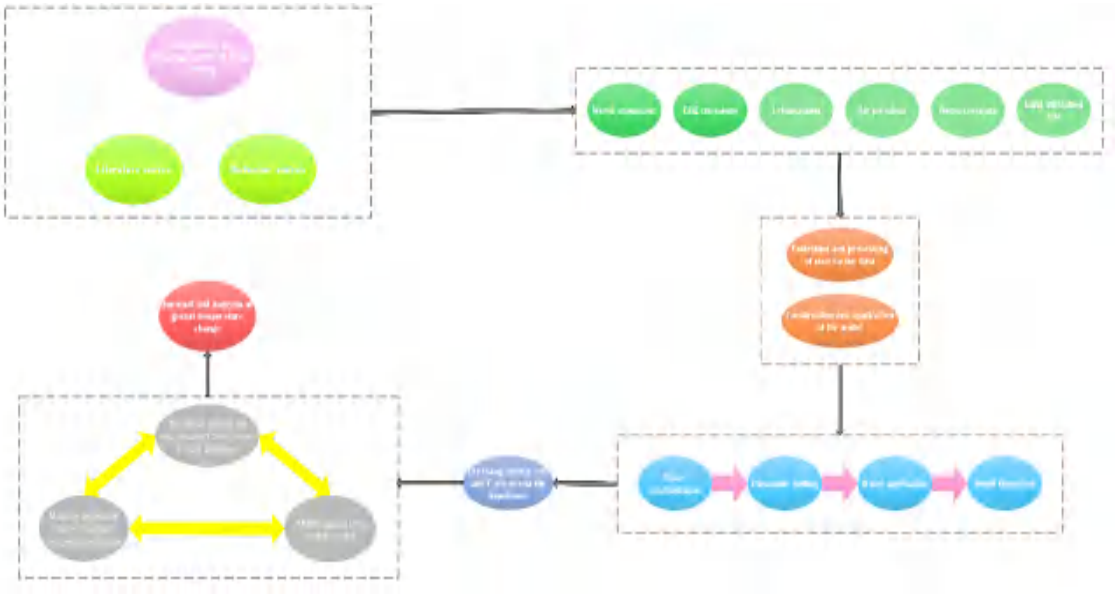
The degree of correlation between each factor and temperature is visualized, and the correlation coefficient is drawn as the above heat map. It can be seen from the figure that the above six influencing factors have high correlation and can be used as predictive temperature factors.

4.2.4 Non-numeric variable processing

Digitize the non-numeric features, split the date into year and month, split the latitude into positive and negative, with 0 as the boundary, the north latitude is a positive number, and the south latitude is a negative number. Split the longitude into positive and negative, with 0 as the boundary, the east longitude is a positive number, and the west longitude is a negative number.

5. Models and Solutions

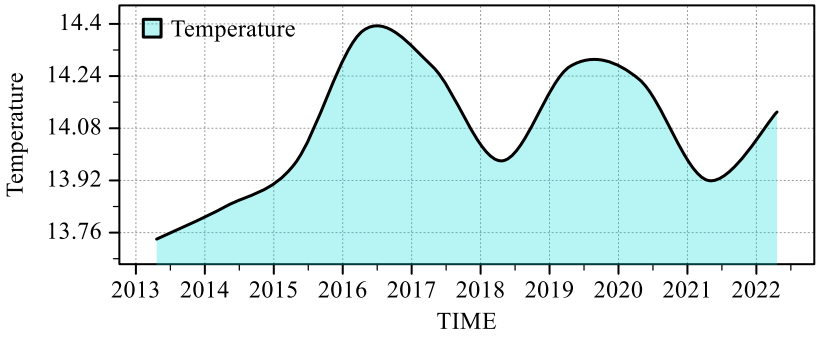
5.1 Models and Solutions for Problem1



Picture 3 Problem 1 structure diagram

5.1.1 Verification and judgment

A. The sequence diagram of the global average temperature from March 2013 to March 2022 by collecting data is as follows:

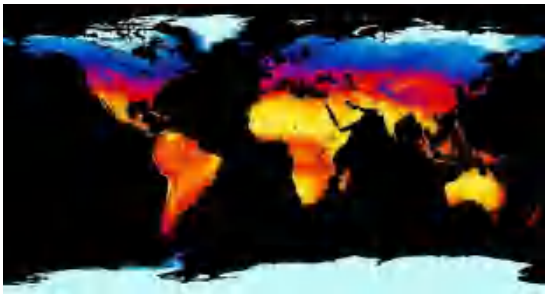


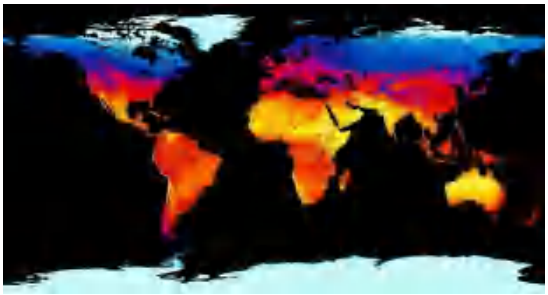
Picture 4 Series map of global average temperature from March 2013 to March 2022

It can be seen from the figure that the global average temperature in March every year has fluctuated and increased in the past ten years.

The trendline is:

$$y = 0.0288x - 44.015 \quad \textbf{(2)}$$

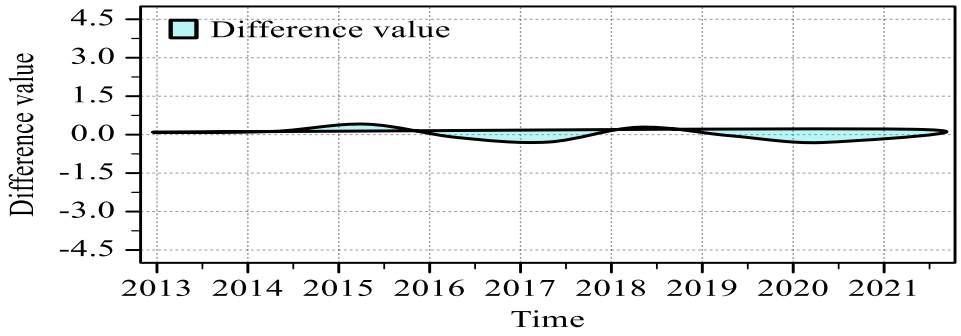




Picture 5 Comparison of global temperature heat maps between March 2013 and March 2022

B. As can be seen from the above figure, the global temperature will rise to a certain extent from March 2013 to March 2022.

C. Make a first-order difference of the global average temperature in March every year in the past ten years, and the results are as follows:



Picture 6 The first-order difference map of the global average temperature in March each year in the past ten years

It can be seen that the global average temperature has fluctuated and increased in the past ten years, and the growth rate fluctuates around 0.

D. Perform a hypothesis test on the first-order difference of the global average temperature in the past ten years, and take the original hypothesis H0: the temperature difference between 2021 and 2022 is less than or equal to the temperature difference in the past ten years. Considering the small amount of data, the normal distribution test The Shapiro-Wilk test is suitable. Carry out the Shapiro-Wilk test on the temperature difference in the past ten years to verify whether the data obey the normal distribution. After calculating the Wilk statistics, the corresponding p value can be obtained. Compare the p value with 0.05, if If it is less than 0.05, the null hypothesis can be rejected, that is, the data does not obey the normal distribution, otherwise the null hypothesis cannot be rejected, that is, the data obeys the normal distribution.

The results of the Shapiro-Wilk test are as follows:

Table 3 Normality Test

	Kolmogolov-Sminov a			Shapiro-Wilk		
	statistics	degrees of freedom	significant	statistics	degrees of freedom	significant
difference	0.136	8	0.200*	0.954	8	0.756

***. True lower limit of significance.**

a. Riley's significance correction

It can be seen from the table that the p-value (that is, significance) of this group of data is 0.756, which is greater than 0.05, indicating that this group of data obeys a normal distribution. If the data satisfies a normal distribution, a one-sample t-test can be performed on the sample, and the p-value of the t-test can be used to determine whether the null hypothesis H0 can be rejected.

E. The results of the one-sample t-test are as follows:

Table 4 One-Sample Tests

	test value = 0.21					
	t	degrees of freedom	significant (double tail)	average value difference	95% confidence interval for the difference	
					lower limit	upper limit

difference	2.049	7	0.080	0.18750	-0.0289	0.4039

It can be seen from the table that the p value of this group of data is 0.08, which is less than 0.1, indicating that the null hypothesis H0 can be rejected at the 90% confidence level, and the alternative hypothesis H1 can be accepted, that is, the temperature difference between 2021 and 2022 is greater than the temperature in the past ten years difference.

Our team agrees that the increase in global temperatures in March 2022 resulted in a larger increase than observed over the past 10 years.

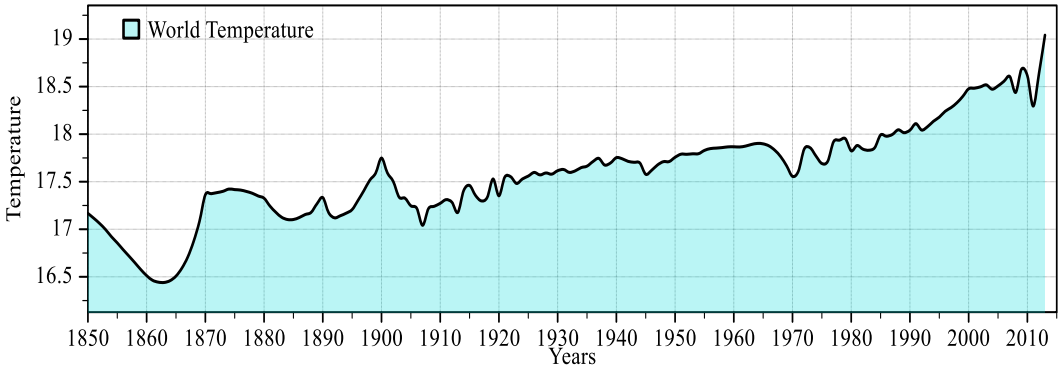
5.1.2 Establishment and solution of prediction model

(1) Multiple regression prediction

Step 1 Model establishment

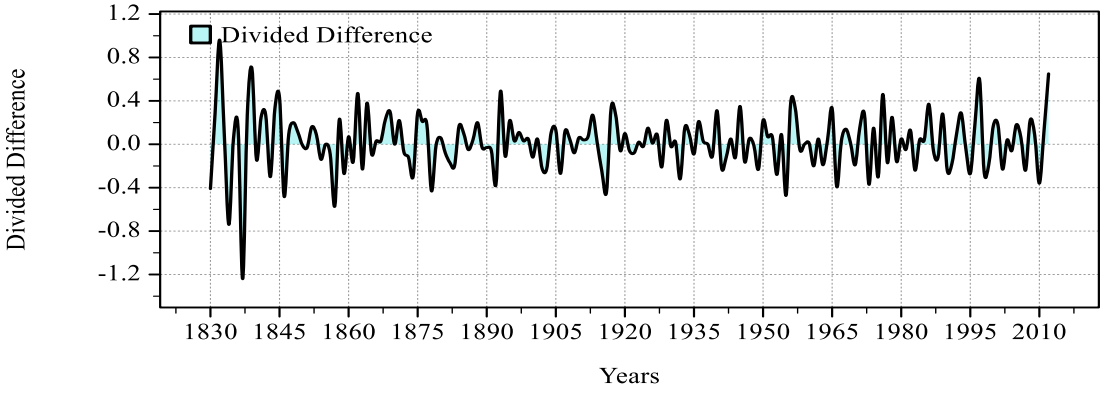
- Stationary analysis of global climate change

Plotting the year-by-year data of global temperature into the following sequence diagram, it can be seen that the sequence has a gradual upward trend, which shows that the trend of global warming does exist.



Picture 7 Year-by-year data series of global temperature

After the yearly data of global temperature are processed by first-order difference, the sequence obtained is relatively stable.



Picture 8 Sequence diagram of yearly data of global temperature after first-order difference processing

- Model construction

Combined with practical considerations and according to relevant research [2] , the main factors affecting global climate change include world population, atmospheric carbon dioxide emissions, forest coverage, urbanization rate, land use area, sulfur dioxide emissions, etc.

Establish a multiple linear regression model between each factor and global temperature, and use the training set and test set to evaluate the accuracy of the model, and finally use the constructed model to predict the global temperature in 2050 and 2100.

- Build multiple linear models

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon \quad \textbf{(3)}$$

- Multiple linear regression model establishment

Using SPSS software to build a multiple linear regression model to obtain model coefficients, the regression model can be expressed as:

$$y = 17.439 + 0.400x_1 - 0.461x_2 + 0.180x_3 - 0.900x_4 - 1.357x_5 - 0.004x_6 \quad \textbf{(4)}$$

Table 5 Model Summary

Model	R	R square	Adjusted R square	Standard Estimate Error
	0.925[a]	0.855	0.850	0.25703

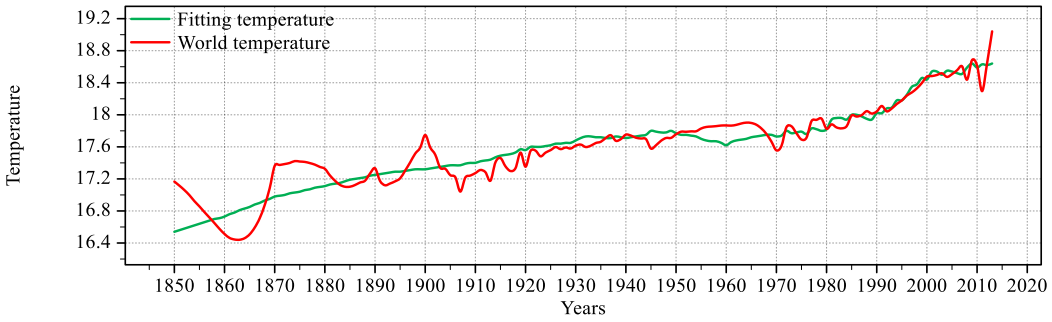
The model summary shows that the R-square of the model is 0.855, the adjusted R-square is 0.850, and the overall fitting effect is good.

Hypothesis is raised in F-test: all coefficients are 0. From the ANOVA table (see appendix), it can be seen that the p value of the F test is <0.001, and the null hypothesis is rejected at the significance level of 0.05, and the test is passed.

Step 2 Forecast and Analysis

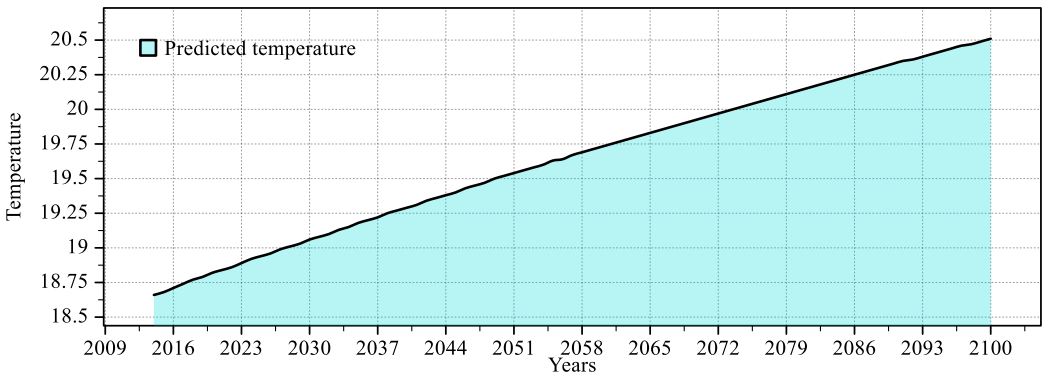
- Model prediction results and testing

The multiple linear regression model is used to fit and predict the global temperature. The fitting results of the model are shown in the figure below. It can be seen that the model fits the historical data well, and the change trend is the same as the real trend.



Picture 9 Temperature Forecast Time Series

The forecast results of the global average temperature for the future to 2100 are as follows:



Picture 10 Average temperature forecast results

Conclusion: From the temperature curve obtained by the multiple linear regression model, it can be concluded that the trend of global climate change in the future is that the temperature will continue to rise and the world will continue to warm.

(2) ARIMA time series forecasting

The ARIMA model is Autoregressive Integrated Moving Average Model, which is a model established for non-stationary time series.

Step 1 Model building

From the establishment process of the establishment of the multiple regression model, it can be seen that the stability of the first-order difference sequence diagram has been significantly improved, which is conducive to the establishment of the ARIMA model. After the analysis of stationarity, periodicity, autocorrelation and partial autocorrelation, it is found that the effect of the model is the best when the ARIMA (0, 1, 5) model is used for fitting and forecasting.

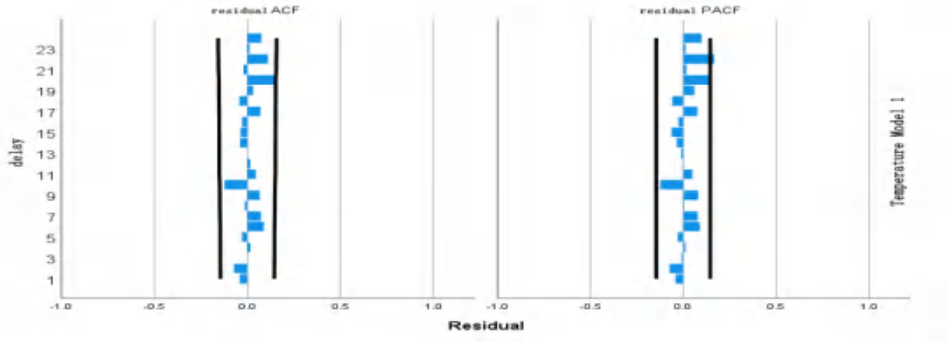
Table 6 Model Statistics

Model	predictor variable	Model Fit Statistics		Young Box Q(18)			Outliers
		Stationary R square	Rsquare	statistics	DF	significance	
Air temperature - model	0	0.218	0.881	10.496	15	0.787	0

The R-square of the model is 0.881, indicating that the results of the model have high credibility.

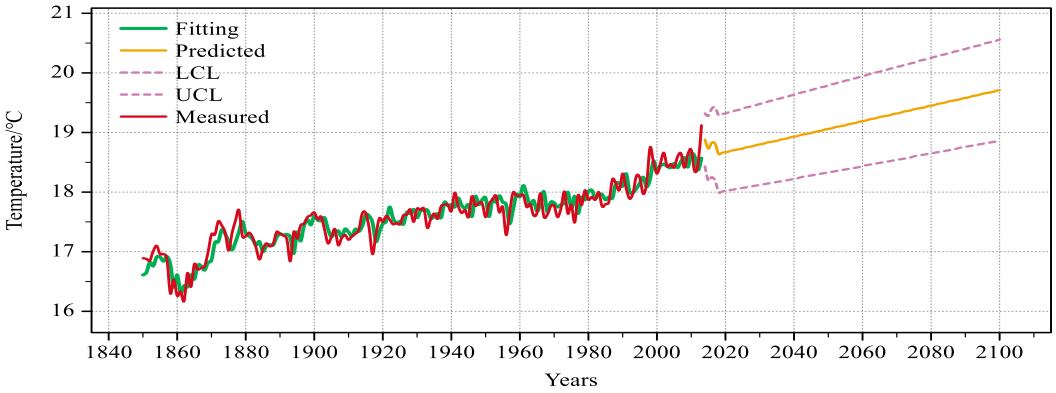
Step 2 Forecast and Analysis

The residual ACF (autocorrelation coefficient) and PACF (partial autocorrelation coefficient) plots of the model are as follows:



Picture 11 Residuals plot of time series model based on global temperature

The model's fitting and forecasting graph of global temperature is as follows:



Picture 12 Fitting and forecasting plot of time series model based on global temperature

It can be seen from the figure that the model fits the historical data better, and the change trend is consistent with the real trend.

(3) Nonlinear prediction based on random forest algorithm

Step 1 Model building

Random forest regression is a model that integrates multiple CART trees for regression prediction through the idea of ensemble learning. In this paper, the CART regression tree is used to build the model. Based on the principle of minimum mean square error, each factor leaf node (division point) s is divided for each climate influencing factor. For any division factor A, the corresponding arbitrary division point s , the parent node splits into two child nodes (hypothesis), the data sets of the two nodes are D_1 and D_2, find the corresponding factors that minimize the mean square error of the respective sets of D_1 and D_2, and minimize the sum of the mean square errors of D_1 and D_2 and factor value as the dividing point.

The expression is:

$$\min_{A,s}\left[\min_{c_1}\sum_{x_i\in D_1(A,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in D_2(A,s)}(y_i-c_2)^2\right] \quad (5)$$

In the formula, c_1 represents the mean value of the first node, and c_2 represents the mean value of the second node. The above formula is used as the principle to determine the factor division points, and finally establish a climate prediction regression model based on each factor.

Step 2 Random Forest Model Parameter Tuning

From the optimal parameter training diagram (see appendix), it can be seen that when the Number of Leaves is 5, the parameters are optimal, and the mean square error is 0.01.

Step 3 Evaluation of factor contribution rate

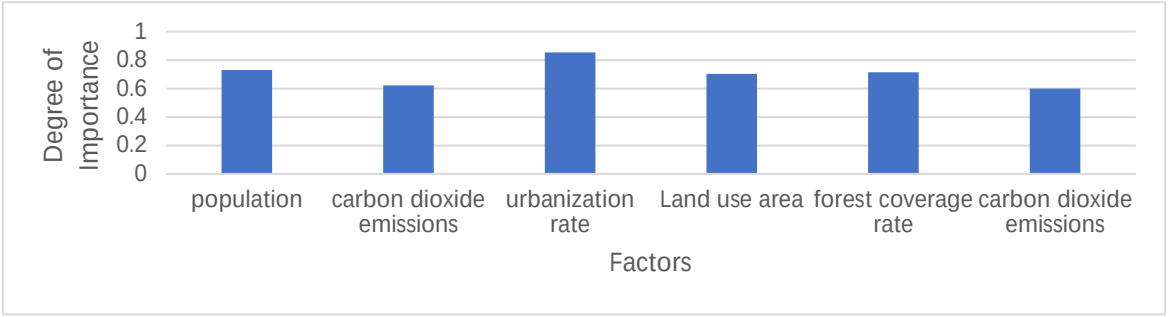
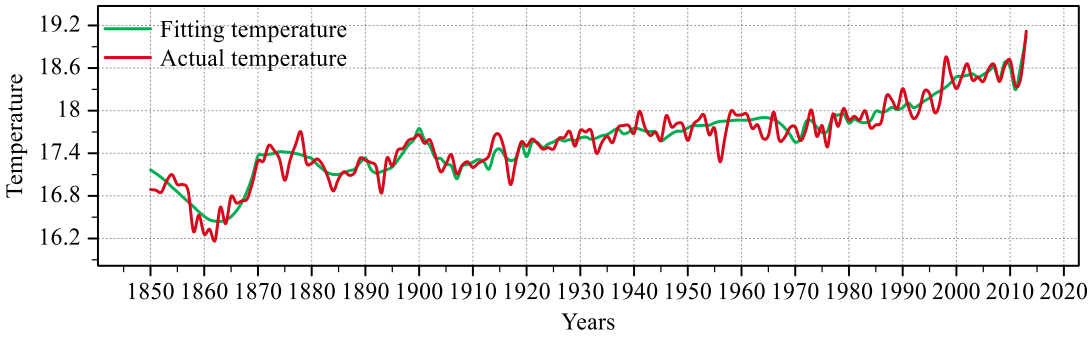


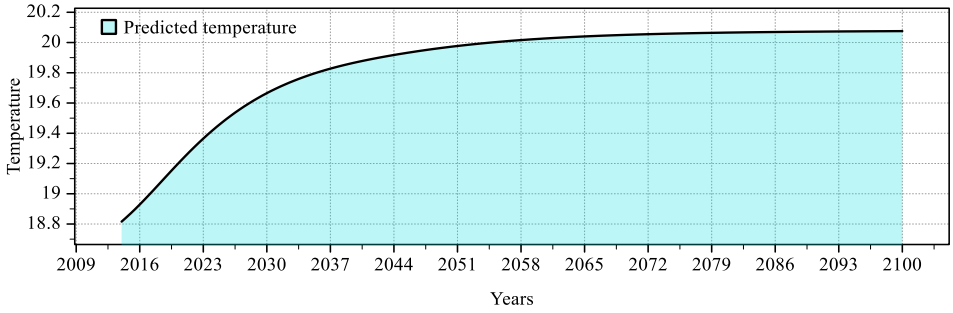
Table 7 Importance of climate-influencing factors

Step 4 random forest prediction effect evaluation

The prediction effect of the random forest was tested on the test set, and the results showed that the adjusted R^2 value of the prediction reached 0.9586, and the prediction effect was good.



Picture 13 Random forest fitting data



Picture 14 Global Average Temperature Forecast

5.1.3 Conclusions of the model

(1) Multiple regression prediction model

By 2050, the global average temperature is predicted to be 19.51 degrees, and by 2100, the global average temperature is predicted to be 20.51 degrees, that is, the model does not believe that the global average temperature will reach 20 degrees in 2050 or 2100, and the average temperature will reach 20 degrees The time forecast is 2074.

(2) ARIMA time series model

By 2050, the global average temperature is predicted to be 19.06 degrees, and by 2100, the global average temperature is predicted to be 19.71 degrees, that is, the model does not believe that the global average temperature will reach 20 degrees in 2050 or 2100, and the average temperature will reach 20 degrees The time forecast is 2122.

(3) Nonlinear prediction model based on random forest algorithm

By 2050, the global average temperature is predicted to be 19.96 degrees, and by 2100, the global average temperature is predicted to be 20.07 degrees, that is, the model believes that the global average temperature will reach 20 degrees in 2050 or 2100, and the average temperature will reach 20 degrees The time forecast is 2055.

5.1.4 Model Precision and Accuracy

It can be seen from Table 9 that when the six influencing factors are used for modeling, the absolute error and relative error of the five-year independent sample forecast by the multiple regression method are 0.148°C and 0.80% respectively; Under these conditions, the absolute error and relative error of the ARIMA time series forecast for the six-year independent sample forecast are 0.016°C and 0.87%, respectively, and the absolute error and relative error of the six-year independent sample forecast for the nonlinear forecast based on random forest are 0.081579°C and 0.44%

Therefore, it can be concluded that in this question, the nonlinear forecasting ability based on random forest is stronger than the multiple regression forecasting model and the ARIMA time series forecasting model.

Table 8 Forecasts and errors of the three models

years	reality	Multiple Regression Forecasting			ARIMA time series forecasting			Nonlinear Forecasting Based on Random Forest		
		forecast	absolute error	relative error	forecast	absolute error	relative error	forecast	absolute error	relative error
2008	18.41	18.59	0.18	0.98%	18.58	0.17	0.92%	18.43831	0.02831	0.15%
2009	18.63	18.64	0.01	0.05%	18.42	0.21	1.13%	18.68301	0.05301	0.28%
2010	18.71	18.59	0.12	0.64%	18.64	0.07	0.37%	18.61449	0.09551	0.51%
2011	18.35	18.63	0.28	1.53%	18.6	0.25	1.36%	18.29478	0.055222	0.30%
2012	18.47	18.62	0.15	0.81%	18.37	0.1	0.54%	18.64674	0.17674	0.96%
	Average		0.148	0.80%		0.16	0.87%		0.081579	0.44%

5.2 Models and Solutions for Problem2: Explore the causes of temperature changes

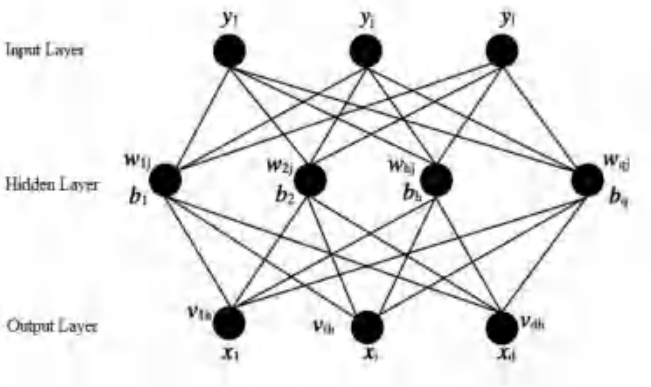
Picture 15 Structure diagram of Problem 2

5.2.1 Spatiotemporal relationship of temperature change

(1) Proof of relationship

Considering the diversity of its variables, it is not suitable for traditional mathematical models. Consider introducing neural network algorithms to prove the relationship and use GA genetic algorithm to process the weights and thresholds of neural networks, that is, to mutate and combine weights and thresholds so that Avoid falling into a local optimal solution, and minimize the possibility that the model training results are affected by the initial weight and threshold pair.

$$BP\ \ neural\ network\ given\ training\ set\ D\ =\ \{(x_1,y_1).(x_2,y_2)\ldots(x_n,y_n)\},x_n\in R_d.y_n\in R_l.$$

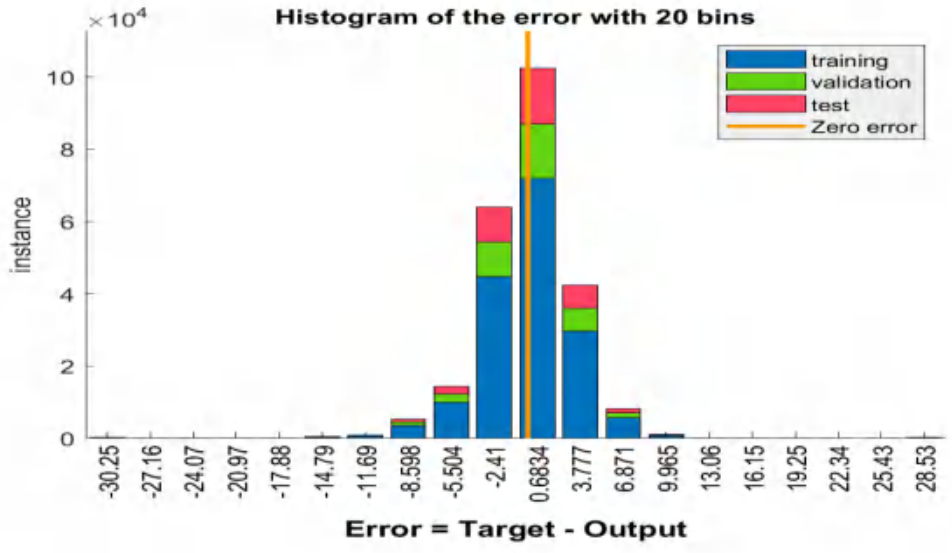


Picture 16 Neural Network Structure Diagram

For the i-th neuron, x1, x2... are the input variables of the neuron, w1, w2... are the connection weights

The net output of neurons can be obtained as $Neti_{in}=\sum_{i=1}^{n}w_i\times x_i$, output is: $y_i=f\left(Net_{in}-\theta_j\right)$, Express the results using the least squares method: $E_k=\frac{1}{2}\sum_{j=1}^{l}\left(y_j^{k'}-y_j^k\right)^2$, adjust the parameters according to the error, and then continuously reduce the error.

Genetic algorithm is a search algorithm inspired by the theory of natural evolution, which is helpful to overcome some obstacles encountered by traditional search and optimization algorithms. In this question, it can be used to select, cross or recombine the weights and thresholds of BP neural network and input them into BP neural network in a loop get a better solution. The model is trained with year, month, longitude, and latitude as predictors and temperature as the response.



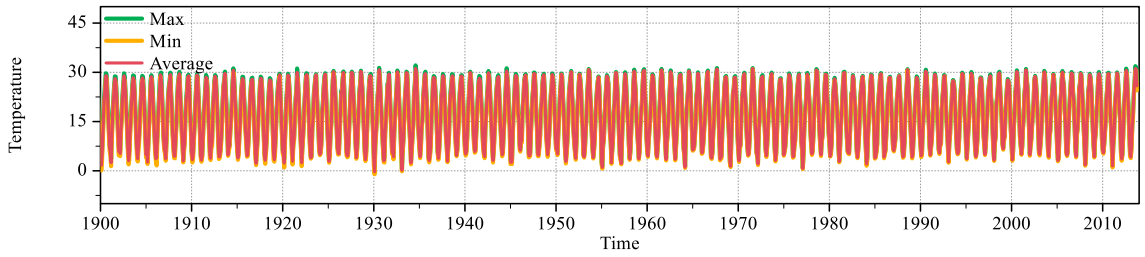
Picture 17 Histogram of the error with 20 bins

See the appendix for the iteration diagram. From the error histogram, it can be seen that the MSE of the test results is 9.1270, and the R is 0.9528. The degree of fitting is good, which can prove that there is a relationship.

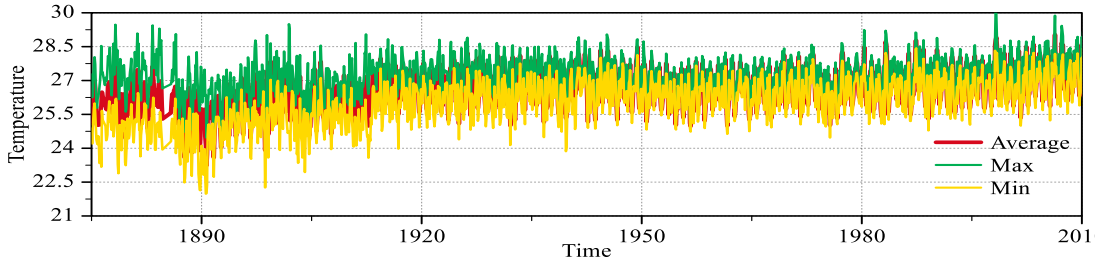
(2) Relationship Analysis

Due to the vastness of the world, it is difficult to analyze the annual average temperature changes at various latitudes and longitudes. Therefore, a typical point is selected in Asia, Africa, South America, and Australia to represent the climate change in the region.

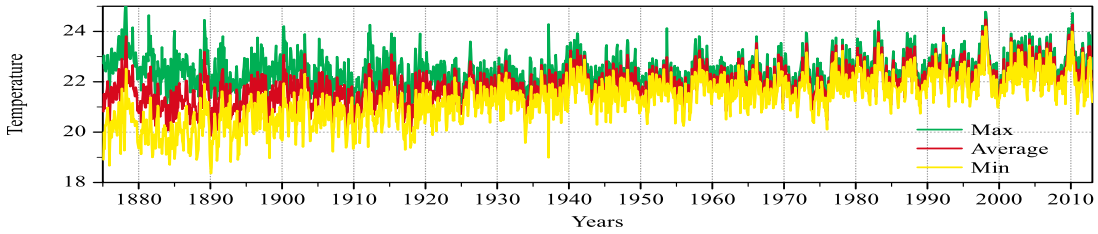
1) Climate tendency rate



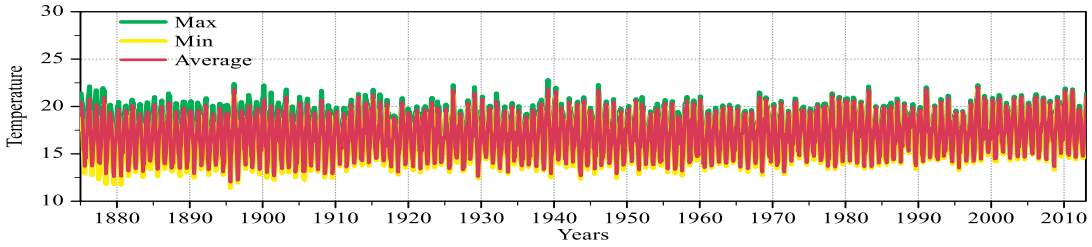
Picture 18 Temperature time series in Asia



Picture 19 Temperature time series in Africa



Picture 20 Time series of temperature in South America



Picture 21 Temperature time series in Australia

It can be seen from the correlation calculation analysis that

The temperature trend in Asia is as follows:

$$y = 0.0008x + 16.488 \quad \textbf{(6)}$$

The temperature trend in Africa is as follows:

$$y = 0.0071x + 12.819 \quad \textbf{(7)}$$

The temperature trend in South America is as follows:

$$y = 0.0095x + 3.4455 \quad \textbf{(8)}$$

The temperature trend in Australia is as follows:

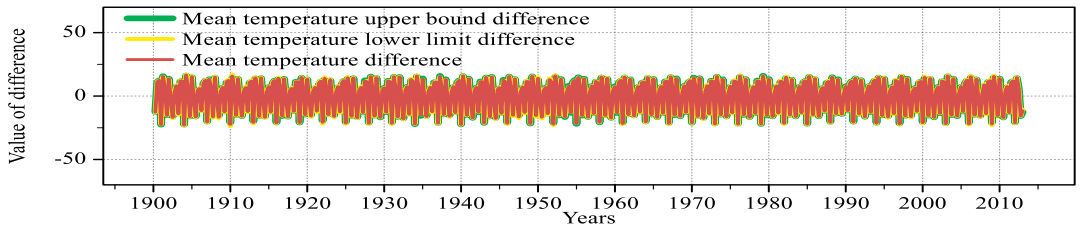
$$y = 0.0005x + 16.671 \quad \textbf{(9)}$$

Through comparative analysis, it can be seen that

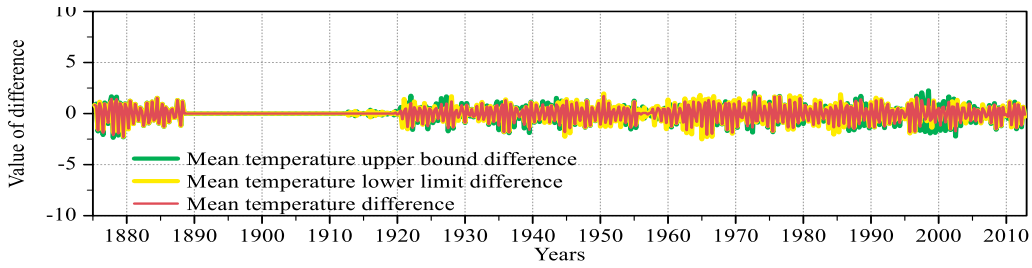
A. The average temperature of Asia, Africa, South America, and Australia all have an upward trend with the increase of the year. Among them, the temperature increase trend of South America is more obvious, and its climate tendency rate is 0.0950 years, that is, the temperature in the interior of Canada will increase by 0.095° every ten years.

B. It can be seen from the time series diagram that there is a clear periodicity in the temperature change.

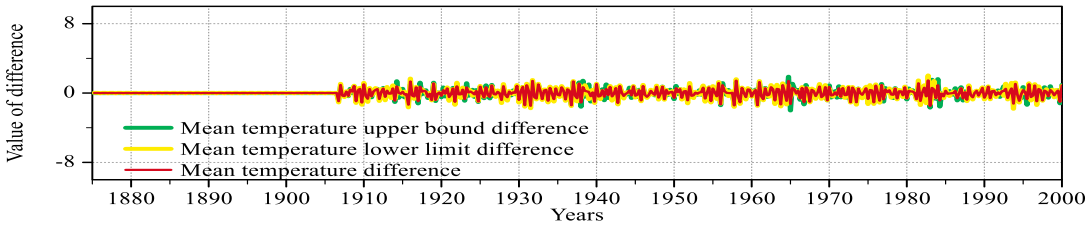
2) Stationarity of time series



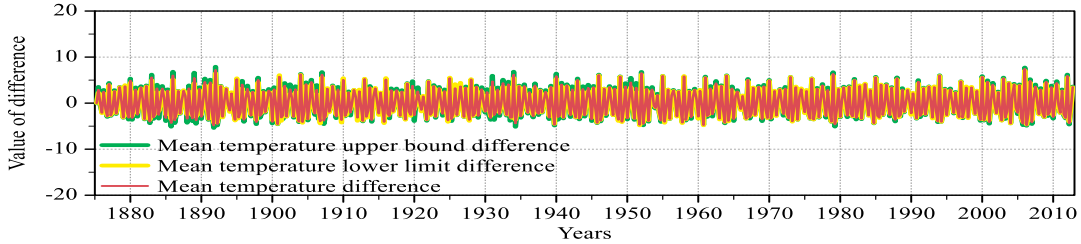
Picture 22 First-order difference of Asian temperature time series



Picture 23 First-order difference of temperature time series in Africa



Picture 24 First-order difference of temperature time series in South America



Picture 25 First-order difference of Australian temperature time series

It can be seen from the above figure that the temperature time series is relatively stable.

5.2.2 Relationship between natural disasters and global temperature

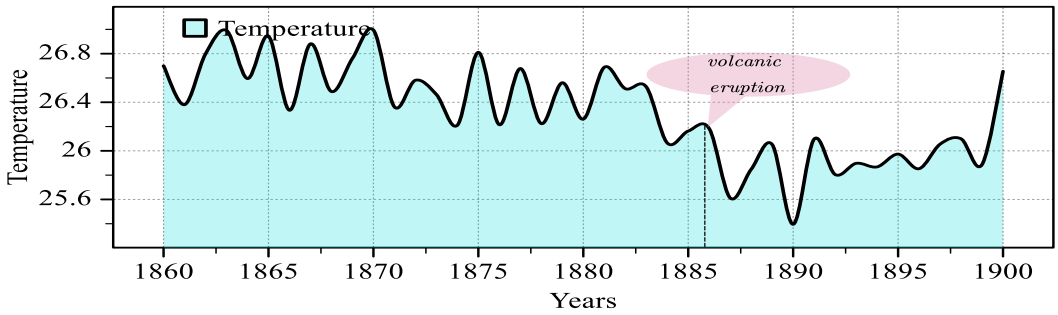
Considering the wide range of natural disasters, we selected volcanic eruptions as an outstanding representative to measure the relationship between natural disasters and global temperature. Combining the factors of major natural disasters in the world in the past 200 years and the data given by the organizing committee, our team selected the Krakatau volcano in 1883 Outbreak events and combined with the given data for analysis.

(1) Background introduction

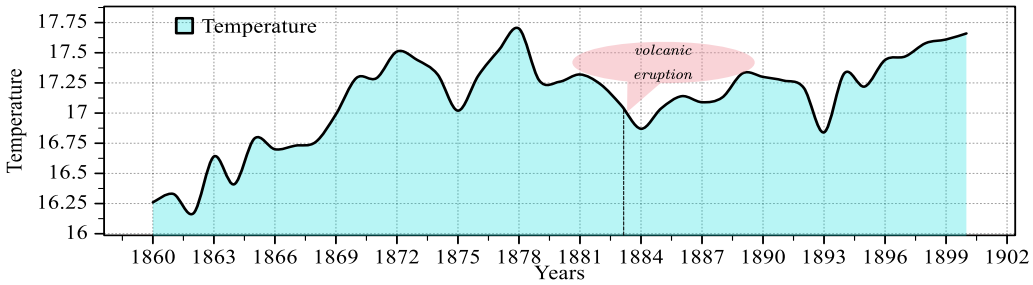
Krakatoa volcano is located in the Sunda Strait of Indonesia, near Lakata Island. The VEI-6 eruption in 1883 released 21 cubic kilometers of volcanic ash, which was one of the largest volcanic eruptions in human history. According to relevant data, the eruption and the resulting tsunami destroyed hundreds of villages and cities. Two-thirds of the original Krakatoa volcano disappeared in the eruption, and the entire planet cooled by an average of 0.6 degrees Celsius.

(2) Analysis of results

Our team intercepted the August temperature in Singapore from 1860 to 1900 for comparison. It was found that the temperature dropped significantly after 1883, with a range of about 0.6 degrees. It proves that the volcanic eruption will cause the local temperature to rise briefly and then drop, and the global temperature will drop briefly.



Picture 26 Impact of volcanic eruption on temperature in Singapore



Picture 27 Impact of volcanic eruptions on global temperature

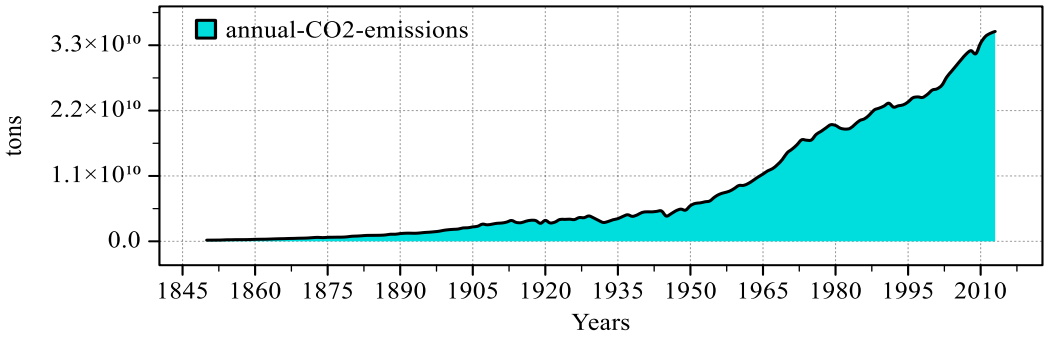
5.2.3 Main causes of global temperature change

Considering that global climate change is a change on a macroscopic space and time scale, and as a systemic variable, it will be affected by many factors. After the above analysis, the influencing factors mainly include: the global population, greenhouse gases (in the form of CO2 is a typical representative), land use rate, ozone concentration, air pollution, urbanization, forest coverage.

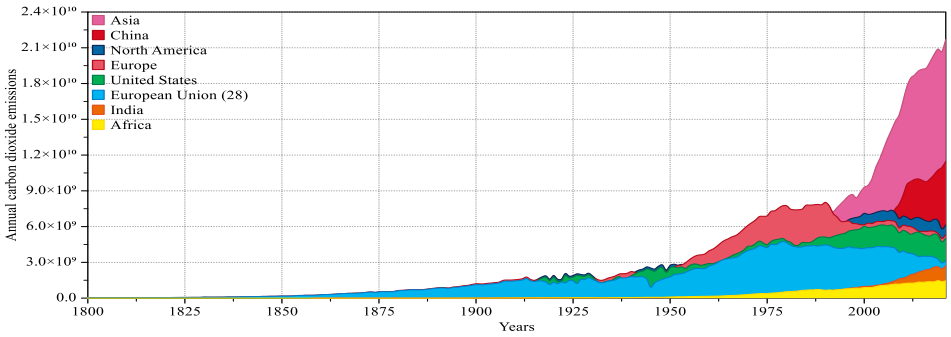
(1) Greenhouse gases

Greenhouse gases refer to some gases in the atmosphere that can absorb long-wave radiation reflected by the ground and re-emit radiation, such as water vapor (H_2O), carbon dioxide (CO_2), nitrous oxide (N_2O), Freon, methane

(CH_4), etc., which make the earth The surface becomes warmer, similar to how a greenhouse traps solar radiation and heats the air inside it.



Picture 28 CO2 emissions

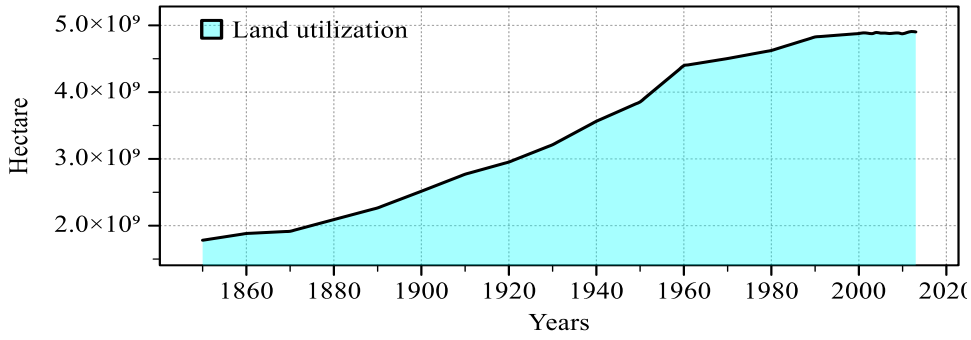


Picture 29 Carbon Dioxide Emissions by Regions

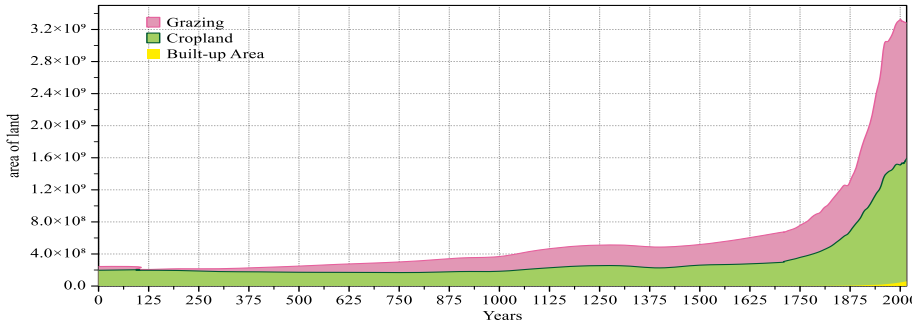
It can be seen from the figure above that since the industrial revolution, carbon dioxide emissions have shown a blowout upward trend, the concentration of atmospheric CO2 has continued to increase, and the ecological function of balancing CO2 and O2 has continued to decline, thus significantly increasing the "greenhouse effect" of CO2.

(2) Land use

Land use mainly refers to the impact of human activities on the natural environment, and is the most important driving factor of land cover change. Land use emissions account for more than one-third of global greenhouse gas emissions. Changes in land cover can cause changes in greenhouse gas emission absorption, ground albedo, and evaporation, thereby causing climate change on a global scale. The following figure shows the use of land by humans over the past two hundred years:



Picture 30 Land use

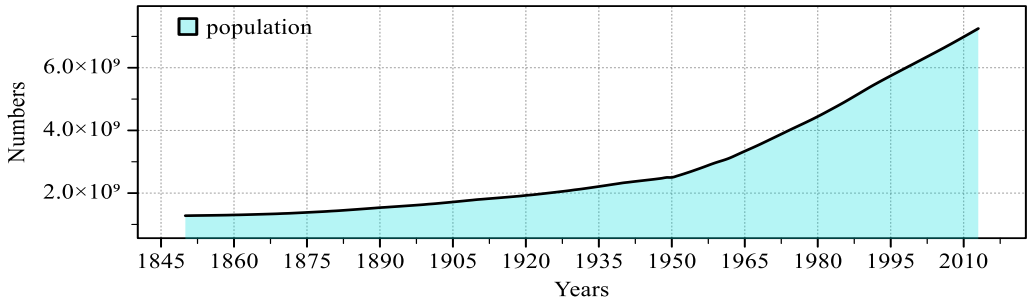


Picture 31 Area of various types of land use

(3) World population

[3] The appearance of human beings on the earth has a history of 3 million years until today. The general trend is that the population growth rate is getting faster and faster. According to the statistics of the relevant departments of the United Nations, the world population was 1.6 billion at the beginning of the 20th century, and it increased to 41 in 1975. billion, predicted to be 10 billion by 2030, and close to 30 billion by 2100. According to literature data [4] Population and environment are closely related. The rapid increase in population, the sharp decline in forests, and the sharp decline in the ecological function of balancing CO2 and O2 in the atmosphere have greatly enhanced the "greenhouse effect" of CO2.

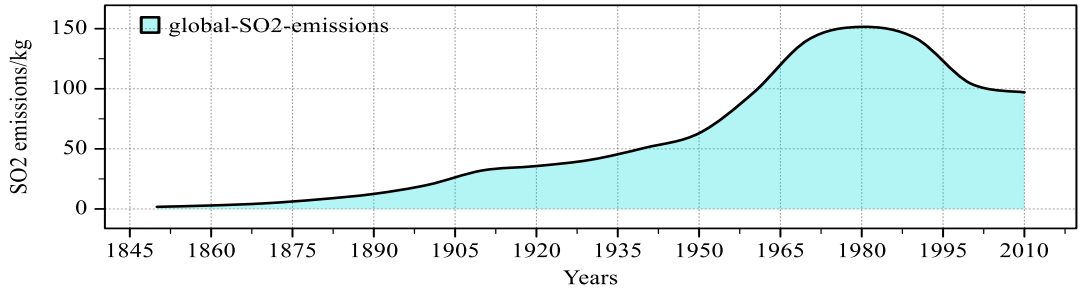
The figure below shows the growth of the world population over the past two hundred years:



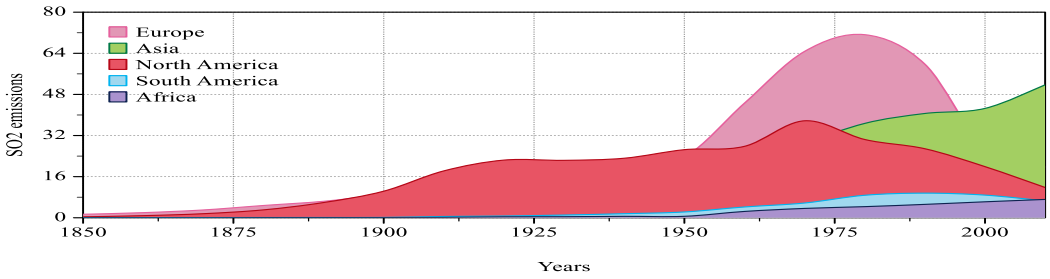
Picture 32 World population

(4) Air pollution

When tracing the historical sources of man-made air pollution, we usually use SO2 as an indicator for data processing.



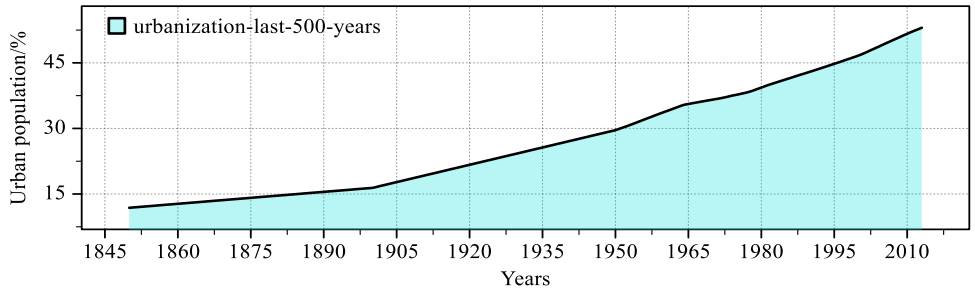
Picture 33 SO_2 emissions



Picture 34 SO_2 emissions by region

(5) Urbanization

One of the most notable characteristics of urban climate is the heat island effect, which means that when a city develops to a certain scale, due to changes in the nature of the underlying surface of the city and air pollution, the temperature in the city is significantly higher than that in the suburbs, forming a phenomenon similar to a high-temperature island [5].

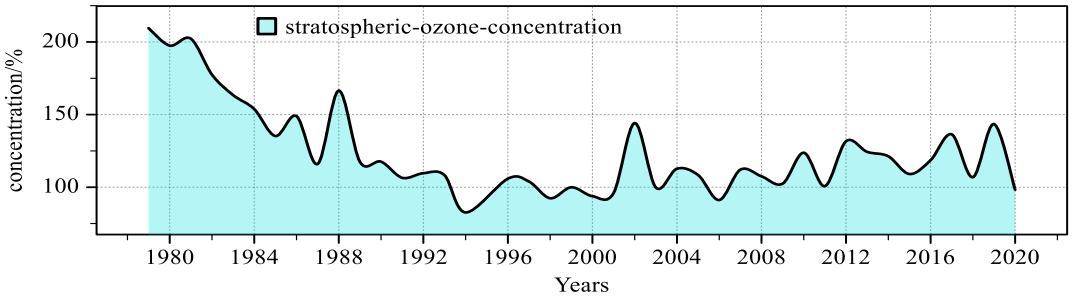


Picture 35 Degree of Urbanization

As can be seen from the above figure, the world is developing towards a high degree of urbanization.

(6) Ozone concentration in the stratosphere

Ozone in the stratosphere also has a certain effect on climate regulation, and the increase in the amount of ultraviolet radiation radiated to the ground will warm the earth.

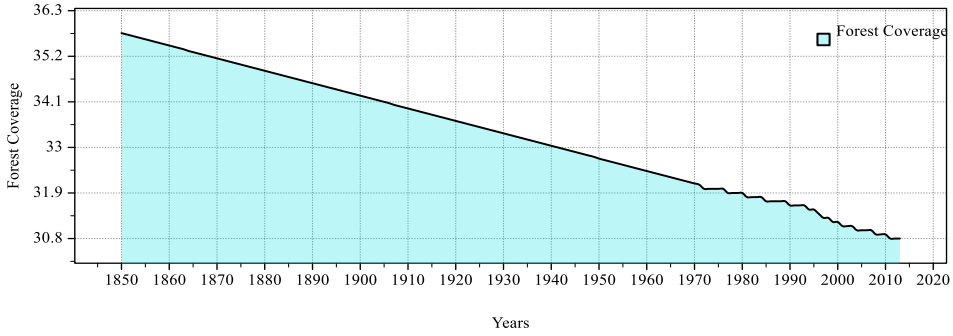


Picture 36 Stratospheric Ozone Concentrations

(7) Forest coverage

Forest cover refers to the ratio of forest area to total land area, and trees absorb a lot of greenhouse gases that cause global warming. Fewer forests mean more greenhouse gases enter the air, which intensifies the rate and extent of global warming.

As can be seen from the figure below, the forest coverage rate will drop sharply, and the ecological function of the forest will be greatly weakened. As a result, the balance ratio of CO2 and O2 in the atmosphere will be seriously out of balance, resulting in a greatly enhanced "greenhouse effect", thereby accelerating the process of global warming.



Picture 37 Forest Coverage

5.2.4 Initiatives to curb and mitigate global warming

Considering that climate change has many impacts on production and life, some of them are irreversible or catastrophic. To sum up, increasing research efforts and in-depth research on the application of countermeasures against global climate change can reduce a series of adverse effects caused by climate change to a certain extent. Globally, the following measures are proposed:

1) Considering that climate change is a prominent global issue, the international community must strengthen cooperation among countries, learn from each other's strengths, and share advanced experience.

2) Therefore, countries around the world should adopt active population policies, implement family planning, and effectively control population growth.

3) Regarding carbon dioxide emissions:

A. Improve energy utilization technology and energy utilization efficiency, thereby reducing the use of natural gas, coal, and oil, and reducing carbon dioxide emissions;

B. Develop clean energy and adopt new energy, such as: solar energy, wind energy, geothermal energy, hydrogen energy, ocean energy, nuclear energy, etc.

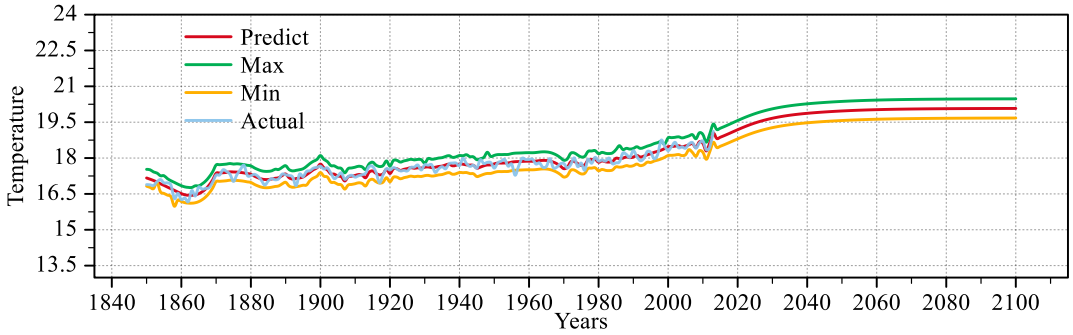
4) For global underlying surface changes (forest coverage, desertification, urbanization, physical landscape, etc.):

A. Increase the intensity of afforestation, increase the protection and management of nature reserves, and at the same time formulate relevant laws and regulations, improve systems and policies, and strengthen the prevention of forest fires and the prevention and control of pests and diseases;

B. Adapt measures to local conditions and make timely adjustments according to climate change trends. Use grass to determine livestock, avoid overgrazing that leads to grassland degradation, improve the disaster resistance of animal husbandry, and effectively curb the trend of desertification.

6. Sensitivity Analysis

When conducting a nonlinear prediction model based on the random forest algorithm, in order to explore the sensitivity of the model, a 95%-105% confidence interval for each input variable of the six influencing factors is given to study the impact of the input variable data on the solution when it fluctuates within the interval.



Picture 38 Sensitivity analysis

It can be seen from the conclusion of the model that the fluctuation range of the solution is less than or equal to 2% (see the appendix for specific data). The above figure shows that the sensitivity of the model is low and the robustness of the model is good.

7. Model Evalustion and Further Discussion

7.1 Multi-global climate change multivariate linear regression model

The multivariate linear regression model establishes the correlation between the multivariate and the dependent variable, so that the factors and the degree of influence on the dependent variable can be directly judged and verified. The mechanism is relatively simple, and the global Factors influencing climate change and the extent of their impact.

However, when making future predictions of global temperature changes, there is a certain lag, which has a lot to do with the selection of the prediction data set, which will cause a large loss of the characteristics of the data set and may cause deviations in the prediction results.

7.2 ARIMA-based time series model

ARIMA adopts auto-regression method, which can predict the future value of a single variable based on historical data. When the amount of data is large, the result is reasonable, but it is difficult for the ARIMA model to give the correlation between multiple variables or factors and dependent variables. Therefore, it can be considered to establish a multivariate ARIMA model, and establish a relationship between the historical data of multiple variables and the dependent variable to complete the optimization of the model.

7.3 Multivariate nonlinear regression model based on random forest algorithm

The model is based on machine learning algorithms and has a high degree of flexibility. Certain data can be selected from a multivariate historical data set as a training sample. After parameter tuning, the fitting degree is extremely high. However, it is difficult to intuitively give Or the correlation between the factor and the dependent variable. When it is popularized and applied, the requirements for the size, quality and training samples of the data set are relatively high.

7.4 GA-BP neural network model

The BP neural network algorithm is a multi-layer feed-forward neural network trained according to the error backpropagation algorithm. It is one of the most widely used neural network models and can make the solution of complex functions simple and feasible. Therefore, this paper uses the genetic algorithm [6] to optimize the connection weight of the BP network and its network structure to optimize the BP neural network model. In view of its strong nonlinear mapping ability and parallel information processing ability, it can be extended to actual production and life.

8. Conclusions

8.1 For Problem one

In part a of problem 1, in the judgment of global temperature rise in March 2022, the methods of normal distribution test and one-sample t test are used, and the data is pivoted, and the conclusion is drawn: the phenomenon of global warming exists, and The increase in global temperatures in March 2022 resulted in a larger increase than observed in the past decade.

In part b of problem 1, in order to describe and predict global temperature changes, multiple regression prediction models, ARIMA time series prediction models and random forest prediction models were respectively established. The linear forecasting model further studies the future global temperature: in 2050, the global average temperature is predicted to be 19.96 degrees, and by 2100, the global average temperature is predicted to be 20.07 degrees, and the time when the average temperature reaches 20 degrees is predicted to be 2055.

8.2 For Problem two

In part a of problem 2, use the BP neural network algorithm optimized by the GA genetic algorithm to prove the relationship and pivot the data, and draw conclusions: ①There is a relationship between temperature change and time and space. ②The average temperature of all continents has an upward trend with the increase of the year, and the temperature increase trend in South America is more obvious. ③ There is obvious periodicity and stationarity in the temperature change in the time series graph.

In part b of problem 2, explores the relationship between global temperature changes and natural disasters, chooses volcanic eruptions as a representative of natural disasters, analyzes the impact on local and global temperatures before and after volcanic eruptions, and draws a conclusion: volcanic eruptions will cause local temperatures to rise temporarily After that, it dropped and caused a brief drop in global temperature.

In part c of problem 2, explores the main causes of global temperature change. By referring to relevant literature and discussions, it is concluded that the main influencing factors of global temperature change include: global population, greenhouse gas emissions, land use, ozone concentration, air Pollution, urbanization rate, forest cover.

9. Article

The high concentration of carbon dioxide in the atmosphere in 2022[7] is unprecedented in the past 2 million years; 1970-2020 is the fastest 50-year increase in global temperature in the past 2000 years. These listed key indicators of climate change fully illustrate The warming process that the earth is experiencing is beyond the laws of nature. Global warming is a common problem faced by human beings, and it is having a profound impact on the natural environment on the surface and the human and social environment. Therefore, the team aimed at a series of questions extended from "Global Warming OR Not?" and obtained related conclusions.

The global climate will generally maintain a "warming" trend in the future, and the main influencing factors are human activities: **global population, greenhouse gases (typically represented by CO2), land use rate, ozone concentration, air pollution, urbanization, forest cover**,etc.Therefore, three prediction models for the global average temperature were established: multiple regression prediction model, ARIMA time series prediction model, and **nonlinear prediction model based on random forest algorithm**. After accuracy evaluation, the nonlinear prediction model based on random forest algorithm was selected. The model predicts the global temperature in the future: in 2050, the global average temperature is predicted to be 19.96 degrees, **the global average temperature will reach 20 degrees in 2055**, and by 2100, the global average temperature is predicted to be 20.07 degrees, and then the global temperature The analysis of the temporal and spatial relationship of changes shows that: from **the climate tendency rate**, the average temperature of all continents has an upward trend with the increase of the year, among which the temperature rise trend in South America is more obvious, and the temperature change has obvious **periodicity and stability**. Finally, choose As a representative of natural disasters, **volcanic eruptions** clarify the relationship between natural disasters and global temperature: volcanic eruptions will cause the local temperature to rise briefly and then drop, and the global temperature will drop briefly.

Therefore, our team makes the following recommendations:

A. Introduce corresponding policies, speed up the improvement of the institutional system, and actively explore new models of low-carbon and development coexistence. Constantly adjust the industrial structure and energy structure, such as improving the fuel consumption of vehicles, vigorously promoting energy conservation and emission reduction, and afforestation and other specific measures.

B. Increase funding and policy support for climate change research, strengthen the organization and coordination of related research, form a complete research system, and improve the level of global climate research.

C. Strengthen the construction of comprehensive coping capacity for extreme climate disasters, and establish a long-term strategic mechanism for mitigating and defending against extreme climate disasters.

D. Formulate and improve relevant laws and regulations, such as the complete ban on chlorofluorocarbons, and based on specific national conditions and regional climate phenomena, adapt measures to local conditions and establish a management system that is different from each other.

E. Protection of biodiversity and ecosystems and sustainable urban planning and infrastructure design, to a certain extent reduce the urban heat island effect and reduce air pollution.

F. Reform the energy system and land use, change consumption patterns; change the energy structure, improve energy efficiency, and reduce greenhouse gas emissions.

10. References and Appendices

[1] Xu Shixiao, Zhao Xinquan, Sun Ping, Zhao Tongbiao, Zhao Wei. Greenhouse effect and global warming [J]. Journal of Qinghai Normal University (Natural Science Edition),2001(04):43-47+52.DOI:10.16229/j.cnki.issn1001-7542.2001.04.014.

[2] Wang Mengchi, Du Yelin, Liu Xintong. Discussion on multi-factors of global climate change [J]. Science and Technology Innovation Herald,2013 (36) : 193-197+199.DOI: 10.16660/j.cnki.1674-098x.2013.36.006.

[3] Li Guochen. Analysis of causes of global warming [J]. Journal of Natural Disasters,2005(05):42-46.

[4] Ke Yan. Analysis of the Impact of Population Development on Global Climate Change [D]. Jilin University,2010.

[5] Peng Shaolin, Zhou Kai, Ye Youhua, Su Juan. Research Progress on Urban Heat Island Effect[J]. Ecological Environment,2005(04):574-579.DOI:10.16258/j.cnki.1674-5906.2005.04.024.

[6] Jin Long, Wu Jiansheng, Lin Kaiping, Chen Binglian. Neural network short-term climate prediction model based on genetic algorithm [J]. Plateau Meteorology,2005(06):981-987.

[7] Xu Qimin. Survival in the climate crisis, the key to choosing the future lies in these few years [N]. Wen Wei Po,2022-08-02(007).DOI:10.28814/n.cnki.nwehu.2022.002355.

Team Number :	apmcm2201058
Problem Chosen :	D

Topology and fractal optimization of heat transfer fins

Abstract

Incorporating fins into heat storage tank is an effective approach to overcome the low thermal conductivity of phase change materials (PCMs) and enhance the heat transfer efficiency. The optimal configuration and distribution of fins are numerically obtained by topology and fractal optimization.

For question 1, a **two-dimensional finite element model** (FEM) of **computational fluid dynamics** (CFD) is established to investigate the heat transfer process. **The enthalpy-porosity method** is used to describe the phase change of PCM. The phase change starts around the rectangular fins and subsequently extends to the outer tank wall, causing the PCM to vary from the solid to liquid phase. Considering the PCM temperature rising from 293K to 333K, the initial phase change time (t_1) at the θ of 72°, 60°, 45°, 36°, 30° and 24° are **47.6min, 39.2min, 31.4min, 25.2min, 20.7min, 17.2min,** and the complete phase change time (t_2) are **158.5min, 142.9min, 123.4min, 106.2min, 92.4min, and 81.9min**, respectively. The t_1, liquid fraction (f_m), and enhancement ratio (E_R) are adopted as indicators to evaluate thermal behaviors. **The optimal $\boldsymbol{\theta}$ is 24°** with minimum t_1 of 17.2min and maximum E_R of 12.3%.

For question 2, the effect of θ, fin length (L), and width (W) of triangular fins are systematically investigated. The results show that θ=24° promotes **63.83%, 56.34%, 46.73%, 46.73%, and 31.71%** heating rate compared with θ=72°, 60°, 45°, 36°, 30°. L=0.024m enhances **42.17%, 70.04%, and 79.52%** heating rate than L=0.006m, 0.012m, and 0.018m. W=0.01m increases by **7.60%, 22.55%, and 35.91%** heating rate than W=0.008m, 0.006m, and 0.004m. **The priority order of fin parameters is $\boldsymbol{L > \theta > W}$**. The heating efficiency of rectangular fins at θ=72°, 45°, 30°, and 24° can increase by **11.95%, 14.99%, 15.43%, and 15.62%** compared with that of triangular fans.

For question 3, two models are proposed to achieve the optimal fin distribution: topology optimization and fractal optimization. For **topology optimization**, the **variable density method** is used and **image reconstruction** is performed to achieve maximum average temperature. For **fractal optimization**, optimal parent-children fractal trees are generated by **Murray's law** and growth ratio. The results show that topology optimization can promote **18.84% and 28.17%** heating rate compared with rectangular fins and triangular fins, respectively, and fractal optimization can promote **14.01% and 24.25%**.

For question 4, a recommendation letter is written to fin company with suggestions on design optimization.

Keywords: heat transfer fins, CFD, topology optimization, fractal geometry.

Contents

1. Introduction

1.1 Background

A double-wall tank with a phase change material (PCM) is commonly used for energy storage. Heated transfer liquid flows into the tank and loses energy to the PCM. The PCM absorbs heat by changing phase from solid to liquid, and releases heat from liquid to solid. Fins are embedded in the PCM between the tank walls to accelerate the heat transfer process. Using fins decreases the melting time and increases the rate of energy storage significantly. The optimal fin distribution is required for the promotion of heat transfer process.

1.2 Question restatement

- ✧ **Problem 1:** A cross-section with rectangular fins evenly distributed is given. Propose a model to simulate the heat transfer process in the tank. Study the effect of fin interval angle on heat transfer and calculate the transfer time when the temperature of PCM changes from 293K to 333K.
- ✧ **Problem 2:** Based on the new cross-section with triangular fins, investigate the effect of fin size on heat transfer. Compare and analyze the results with that of the rectangular fins.
- ✧ **Problem 3:** Propose a mathematical model to obtain the optimal fin distribution.
- ✧ **Problem 4:** Write a recommendation letter to then fin company for the optimal fin design.

2. Model hypothesis

1) The physical properties of the PCM are homogenous, isotropic, and temperature independent (except the density in the liquid state).

2) The phase change process of PCM is regarded as laminar, unsteady, and incompressible.

3) Natural convection of liquid PCM conforms to the Boussinesq assumption.

4) The liquid fraction is assumed to change linearly with temperature.

5) The effect of viscous dissipation is ignored during the simulation.

6) Heat loss to the environment is neglected

3. Symbol description

Symbol	Meaning	Unit
L	Fin length	m
W	Fin width	m
θ	Interval angle	°

T	Temperature	K
t_1	Initial phase change time	min
t_2	Complete phase change time	min
R_1	Radius of the inner wall	m
R_2	Radius of the outer wall	m
ρ	Density	kg/m^3
f_m	Liquid fraction	/
λ	thermal conductivity	W/(m·K)
E_R	Enhancement ratio	/
η	Penalty factor	/
α	Fractal length ratio	/

4. Modeling and solution of Problem 1

4.1 Phase change model

A two-dimensional finite element model (FEM) was established to investigate the thermal and transient responses of the heat transfer process. The phase change of the PCM was described by the enthalpy-porosity method. ANSYS Fluent Modular based on Computational Fluid Dynamics (CFD) was used to simulate the heat transfer process.

4.1.1 Governing equations

The heat storage tank and cross section are depicted in Fig. 1. The interior domain Ω_1 is regarded as the heat source with a temperature of 373K, and the exterior domain Ω_3 is adiabatic. The middle domain Ω_2 is filled with the PCM.

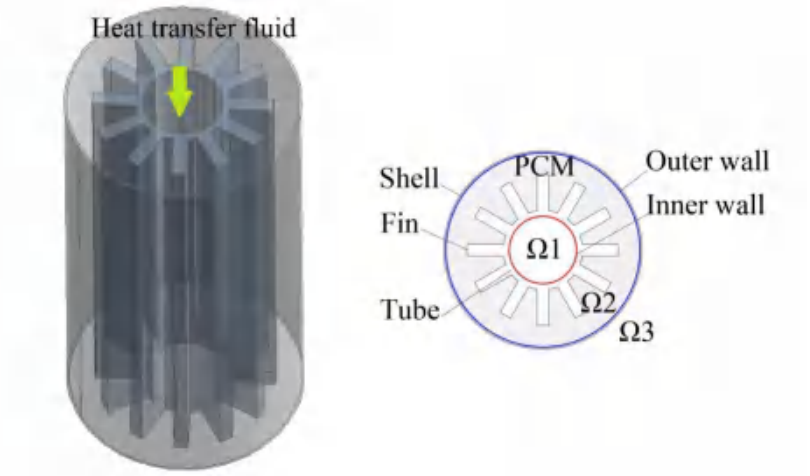


Fig. 1. Diagram of heat storage tank and cross-section

The enthalpy-porosity method is applied to describe the phase change process in the heat storage tank[1]. Governing equations of continuity, momentum, and energy equations predict the flow, heat transfer, and PCM melting, respectively[2].

1) Continuity equation

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0 \tag{1}$$

2) Momentum equations

$$\frac{\partial(\rho_{\mathrm{PCM}}u)}{\partial t}+u\frac{\partial(\rho_{\mathrm{PCM}}u)}{\partial x}+v\frac{\partial(\rho_{\mathrm{PCM}}u)}{\partial y}=\frac{\partial}{\partial x}\left(\mu_{\mathrm{PCM}}\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(\mu_{\mathrm{PCM}}\frac{\partial u}{\partial y}\right)-\frac{\partial p}{\partial x}+Su \tag{2}$$

$$\frac{\partial(\rho_{\mathrm{PCM}}v)}{\partial t}+u\frac{\partial(\rho_{\mathrm{PCM}}v)}{\partial x}+v\frac{\partial(\rho_{\mathrm{PCM}}v)}{\partial y}=\frac{\partial}{\partial x}\left(\mu_{\mathrm{PCM}}\frac{\partial v}{\partial x}\right)+\frac{\partial}{\partial y}\left(\mu_{\mathrm{PCM}}\frac{\partial v}{\partial y}\right)-\frac{\partial p}{\partial y}+Sv \tag{3}$$

3) Energy equations

For PCM:

$$\rho_{\mathrm{PCM}}c_{\mathrm{PCM}}\frac{\partial T}{\partial t}+\rho_{\mathrm{PCM}}c_{\mathrm{PCM}}\left(u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right)=\frac{\partial}{\partial x}\left(\lambda_{\mathrm{PCM}}\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(\lambda_{\mathrm{PCM}}\frac{\partial T}{\partial y}\right)+Se \tag{4}$$

For fin:

$$\rho_{\text{fin}} c_{\text{fin}} \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\lambda_{\text{fin}} \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(\lambda_{\text{fin}} \frac{\partial T}{\partial y}\right) \tag{5}$$

where T is the temperature, p is the pressure, u and v are the velocities in the x and y direction, c is the specific heat, λ is the thermal conductivity, ρ is the density, Su, Sv, and Se are the source terms of the enthalpy method can be expressed by:

$$Su = \frac{(1-f_m)^2}{f_m^3 + \xi} u C_{mush} \tag{6}$$

$$Sv = \frac{(1-f_m)^2}{f_m^3 + \xi} v C_{mush} + \rho_{\text{PCM}} g \delta (T_{\text{PCM}} - T_{m2}) \tag{7}$$

$$Se = \rho_{\text{PCM}} L_{\text{PCM}} \frac{\partial f_m}{\partial t} \tag{8}$$

where δ is the thermal expansion coefficient, μ is the dynamic viscosity of the PCM, f_m is the liquid fraction, g is the acceleration of gravity, C_{mush} is the mushy zone constant C_{mush}=10^7, and ξ us the numerical coefficient ξ=10^{-3} [3].

4) Enthalpy-porosity method

The enthalpy-porosity method uses liquid fraction to describe the state of the PCM, and assumes that the liquid phase ratio equals the porosity. The mathematical equations can be expressed as[4]:

$$f_m = \begin{cases} 0 & T \le T_{m1} \\ \dfrac{T - T_{m1}}{T_{m2} - T_{m1}} & T_{m1} < T < T_{m2} \\ 1 & T \ge T_{m2} \end{cases} \tag{9}$$

where T_{m1} and T_{m2} are solidus temperature and liquidus temperature, respectively.

The energy equation of this method is:

$$\frac{\partial}{\partial t}(\rho_{\text{PCM}} H) + \frac{\partial}{\partial x}(\rho_{\text{PCM}} H v_x) + \frac{\partial}{\partial y}(\rho_{\text{PCM}} H v_x) = \frac{\partial}{\partial x}(\lambda \Delta T_{\text{PCM}}) + \frac{\partial}{\partial y}(\lambda \Delta T_{\text{PCM}}) \tag{10}$$

$$\frac{\partial}{\partial t}(\rho_{\text{PCM}} H) + \frac{\partial}{\partial x}(\rho_{\text{PCM}} H v_y) + \frac{\partial}{\partial y}(\rho_{\text{PCM}} H v_y) = \frac{\partial}{\partial x}(\lambda \Delta T_{\text{PCM}}) + \frac{\partial}{\partial y}(\lambda \Delta T_{\text{PCM}}) \tag{11}$$

where v_x and v_y are the velocities in the x and y direction, and H is the enthalpy which is calculated by the sum of the sensible enthalpy h and latent enthalpy ΔH:

$$H = h + \Delta H \tag{12}$$

$$h = h_{ref} + \int_{T_{ref}}^{T} c \mathrm{d}T \tag{13}$$

$$\Delta H = f_m L \tag{14}$$

where h_{ref} is the reference enthalpy, T_{ref} is the reference temperature, and L is the fin length.

4.1.2 Initial and boundary conditions

1) Initial condition

$$T_{\mathrm{PCM}}(x,y,0) = T_{\mathrm{fin}}(x,y,0) = T_0, u = v = 0, R_1^2 \le x^2 + y^2 \le R_2^2 \quad (15)$$

2) Boundary conditions

$$T_{\mathrm{PCM}}(x,y,t) = T_{\mathrm{fin}}(x,y,t) = T_w, u = v = 0, x^2 + y^2 = R_1^2 \quad (16)$$

$$\frac{\partial T(x,y,t)}{\partial n} = 0, u = v = 0, x^2 + y^2 = R_2^2 \quad (17)$$

where T_0 is the initial temperature, T_w is the temperature of heat source, R_1 is the radius of the inner wall, and R_2 is the radius of the outer wall.

4.1.3 Numerical procedure

1) Geometry and material

Eight fin structures with corresponding interval angles were systematically studied by ANSYS Fluent. Fig. 2 depicts the cross-section under various interval angles θ. The length of rectangular fins (L) was 0.018m, and the width (W) was 0.006m. All rectangular fins are evenly distributed in the cross section. The interval angle θ changes from 180° to 30 ° accompanying fin number from 2 to 12.

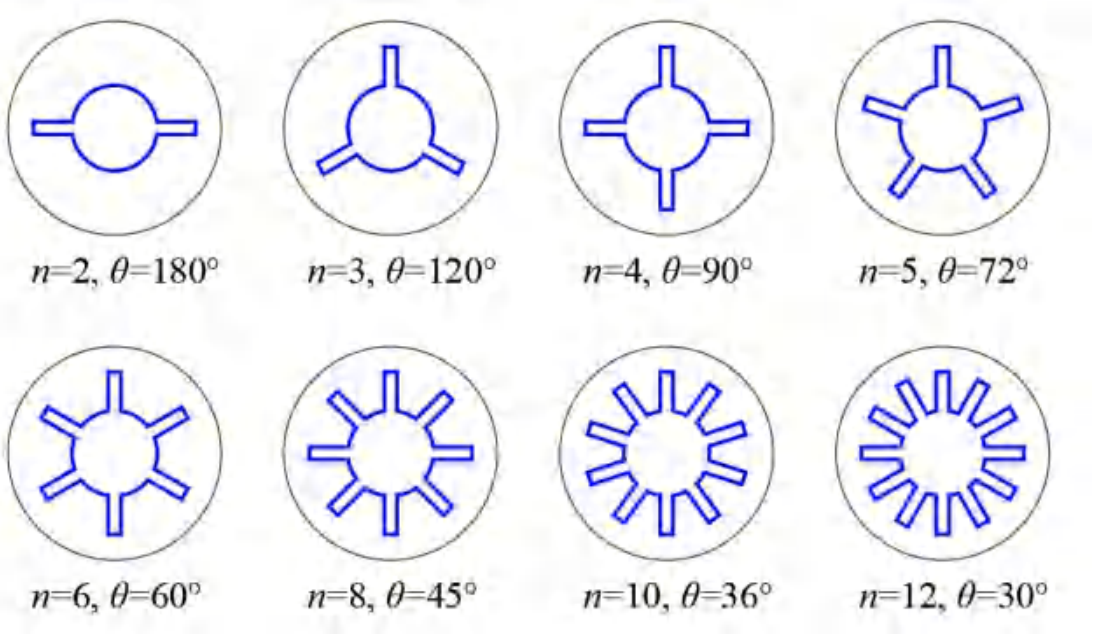


Fig. 2. Cross-section under various interval angles θ

PCM can absorb or release heat energy in the process of changing from solid to liquid or from liquid to solid. The shell-and-fin tank structure (marked as Fin) and PCM domain (marked as PCM) were established in the simulation. Based on the physical properties of PCM and fin and common constants, the material parameters used in the numerical simulation are shown in Table 1.

Table 1 Material property parameters

Symbol	Property	Unit	PCM	Fin (Quartz)
ρ	Density	kg/m^3	780	2650
λ	Thermal conductivity	W/(m·K)	0.15	214
c	Specific heat	J/(Kg·K)	2000	-
T_{m1}	Solidus temperature	K	273.15	-
T_{m2}	Liquidus temperature	K	333	-
δ	Thermal expansion coefficient	K^{-1}	0.0002	-
μ	Dynamic viscosity	Kg/(m·s)	/	-

2) Mesh size and time step

To validate the finite element model, the independence of the grid and time step was examined. The temperature evolution was compared to obtain the accurate and efficient grid, and the phase change time was compared for the optimal time step.

✧ Mesh size

The mesh size was adjusted from 4mm (Grid1) to 1mm (Grid4) to obtain the accurate and efficient grid. Fig. 3 Shows the meshing conditions under different mesh sizes. Fig. 4(a) presents the temperature of the PCM for four grid conditions. With grid size from Grid3 to Grid4, the temperature evolutions were almost identical. Considering the computational efficiency, the mesh size in this study is 2mm.

✧ Time step

Fig. 4(b) shows that the result of phase change time is stable at the time step of 1s or 2s. Therefore, the time step used in the simulation was selected as 2s.

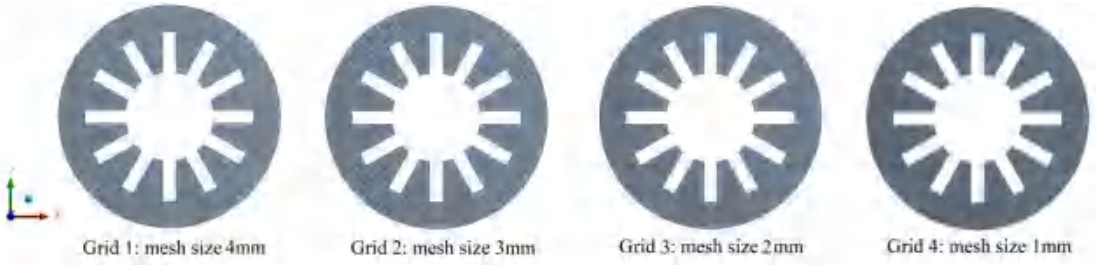


Fig. 3. Meshing conditions under different mesh sizes

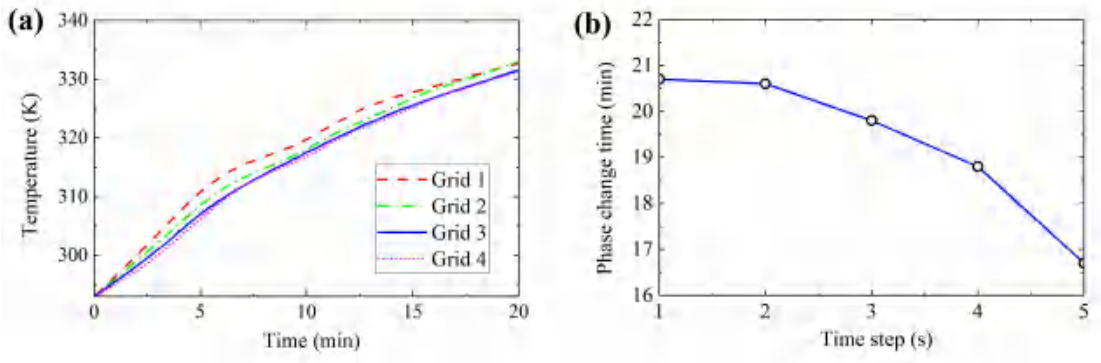


Fig. 4. Validation of the numerical model: (a) grid test, (b) time step test

4.2 Heat transfer process

4.2.1 Temperature distribution

Fig. 5 shows the temperature distribution considering the influence of the interval angle of the fin structure. It can be seen that the phase change first starts in the area around the fins, because of the heat transfer from the fins to the PCM. As time grows, the phase change extends to the outer shell through fins, causing the PCM within the tank completely melts to liquid phase.

The addition of the fin significantly accelerates the phase change process of PCM in the system. This is because the gap between fins becomes smaller and the internal PCM can quickly absorb the heat to melt under the action of the heat conduction mechanism.

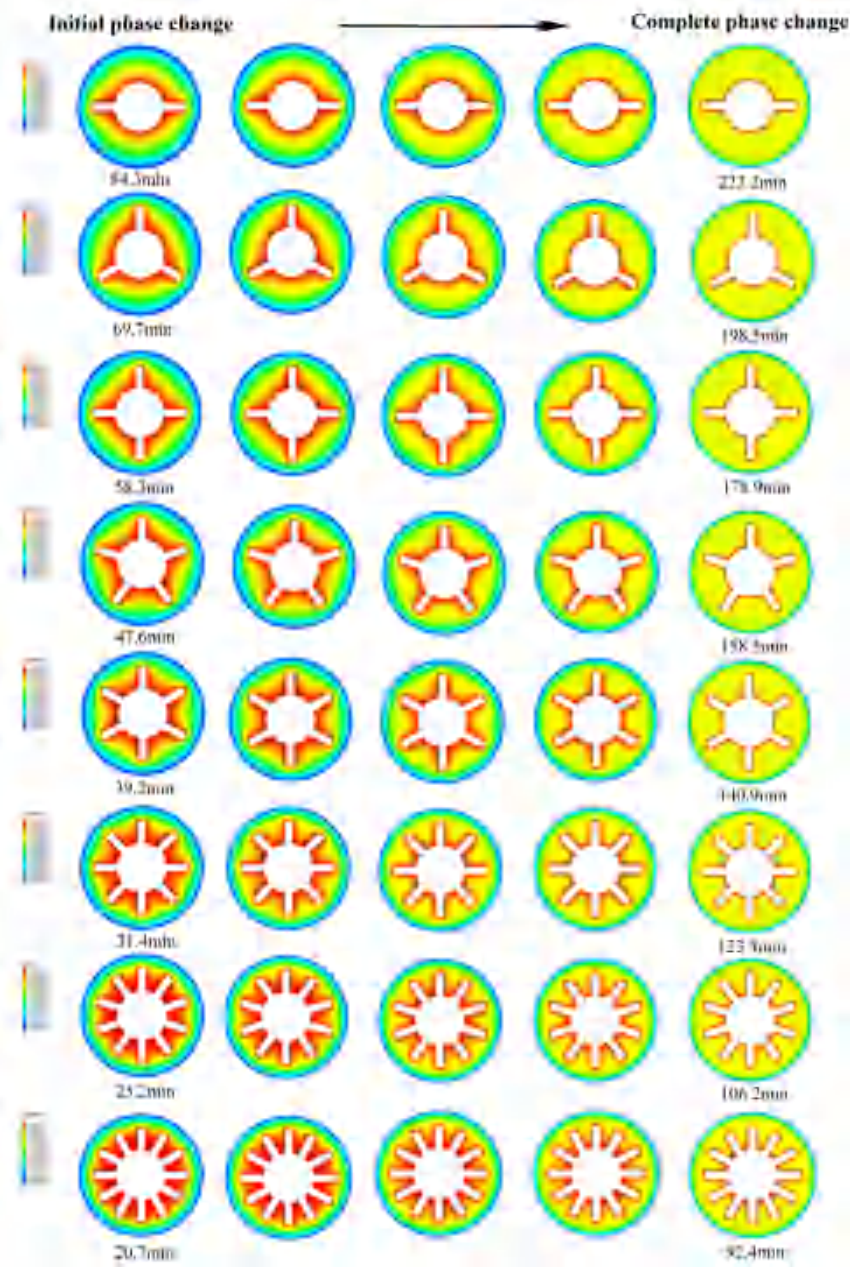


Fig. 5. Temperature distribution under different θ

4.2.2 Phase change time

The time of initial phase change and complete phase change for PCM melting were both studied.

- ✧ **Time of initial phase change** t_1: it means the start time of the PCM rising from room temperature (293 K) to the phase change temperature (333K).
- ✧ **Time of complete phase change** t_2: it refers to the final time when the PCM completely melts to the liquid phase.

Fig. 6 shows the variation of phase change time under different interval angles. The initial and complete phase change time decreases with the decrease of the interval angle (θ) or the increase of the fin number (n). When θ decreases from 60°, 45°, 36°, to 30°, the initial phase change time is 39.2min, 31.4min, 25.2min, and 20.7min, respectively.

However, as the decrease of θ, fin efficiency weakens slightly. This is due to the shrinkage of the spacing in the middle of the tank, resulting in a slight reduction of the volume of the PCM. Therefore, the optimal interval angle should be investigated to get the most efficient heat transfer configuration.

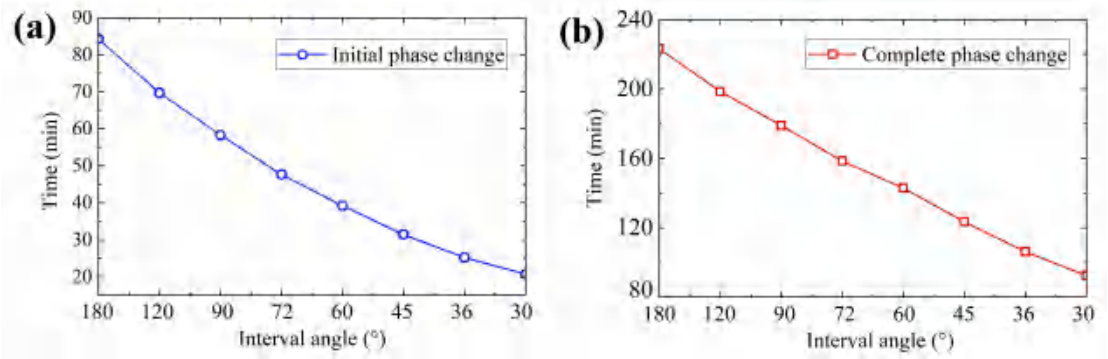


Fig. 6. Phase change time in different cases: (a) initial phase change time, (b) complete phase change time

Table 2. Phase change time

Fin number	Interval angle (°)	Initial phase change time (min)	Complete phase change (min)
2	180	84.3	92.4
3	120	69.7	106.2
4	90	58.3	123.4
5	72	47.6	142.9
6	60	39.2	158.5
8	45	31.4	178.9
10	36	25.2	198.5
12	40	**20.7**	**223.2**

4.2.3 Liquid fraction

Fig. 7 exhibits the variation of the liquid fraction over time under different fin numbers. It can be seen that the melting rate of the PCM increases with the increase

of n. However, with the increase of n, or in other words, the decrease of θ, the growth rate of melting starts to slow down to a certain critical value. The reason for this phenomenon will be explained in '4.4.2 Optimal interval angle'. Therefore, the θ of the fin should match the overall arrangement of the fins to reach optimal thermal efficiency.

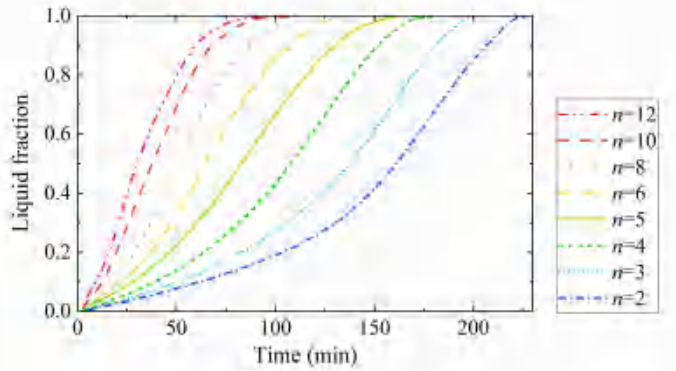


Fig. 7. Liquid fraction in different cases

4.3 Optimization of the interval angle

1) Enhancement ratio

To explore the optimal interval angle, the numerical simulation was expanded to n=20 to further the effect of n on the heat transfer. The enhancement ratio E_R was proposed to evaluate the influence of n on heat transfer rate, which can be defined as:

$$E_R = \frac{f_m(n) - f_m(\text{no fins})}{f_m(\text{no fins})} \times 100\% \tag{18}$$

where $f_m(n)$ is the liquid fraction of the PCM under the n fin number, and $l(\text{no fins})$ is the liquid fraction without fins added.

Fig. 8(a) shows the enhancement effects of different n on the PCM melting rate. With the growth of n, the E_R increases significantly, which indicates that larger n shows a greater advantage in the immediate stage after heating. **The E_R of 15 fins (θ=24°) reaches the maximum value of 12.3%**, and E_R of 14 fins and 16 fins are relatively low at 12.1% and 11.8, respectively.

However, with the further increase of n, especially at n is greater than 15, the E_R gradually decreases and excessive fins limit the heat transfer, causing the PCM at the wall surface to barely melt. This is because smaller spacing is left between adjacent fins and the flow resistance is larger, which weakens the amplitude of fluid movement[5].

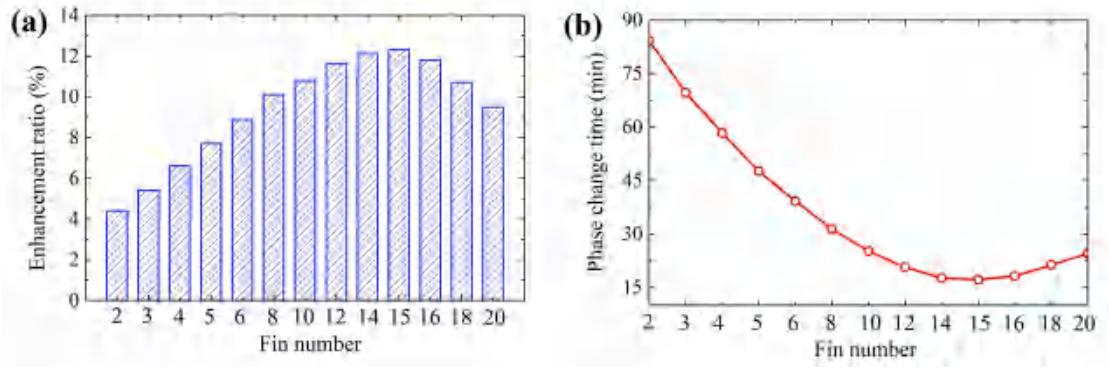


Fig. 8. Comparison of different θ: (a) enhancement ratio, (b) initial phase change time

3) Phase change time

The initial phase change times under different n were calculated correspondingly. Fig. 8(b) shows the initial phase change time of different n. It can still be seen that the **15-fin structure (θ=24°) has the best capacity for heating with a minimum melting time of 17.2 min**.

5. Modeling and solution of Problem 2

5.1 Effect of the interval angle

Same as rectangular fins, eight interval angles of triangular fins were systematically studied in this section. Fig. 9 depicts the cross-section under various θ. The L was 0.018m and the W was 0.006m. The θ changes from 90° to 24 ° accompanying the fin number from 4 to 15.

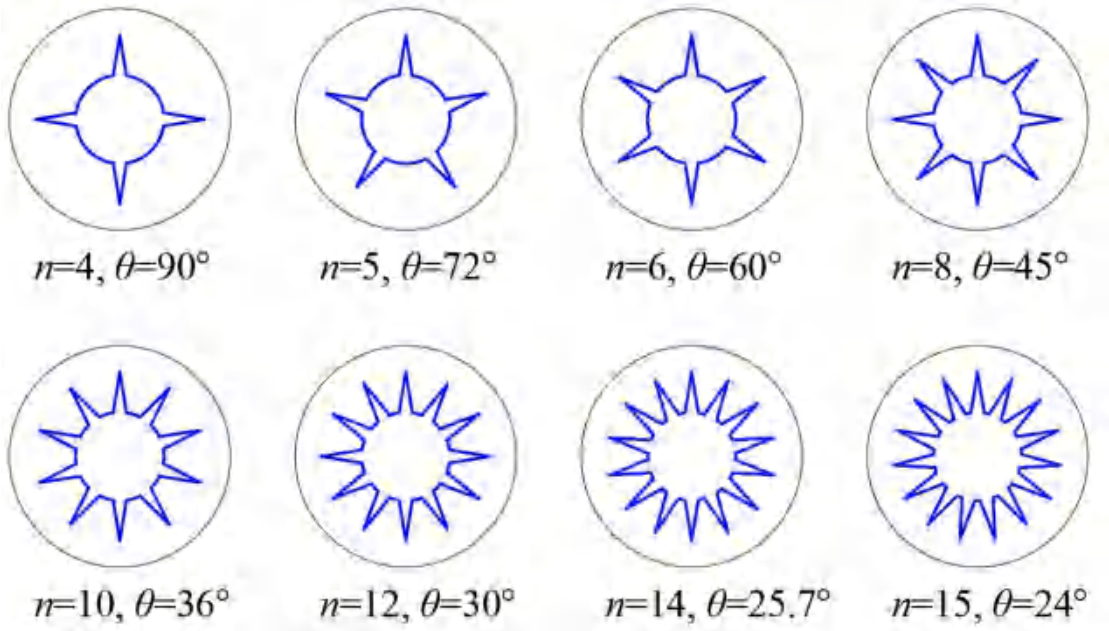


Fig. 9. Cross-section under various θ

1) Influence on heat transfer

The performance of the heat storage was analyzed by varying the θ while keeping other parameters constant. Fig. 10 shows the temperature distribution under various θ of triangular fins. The melting time decreases by about 44.1% when the θ changes from 90° (n=4) to 45° (n=8), and significantly decreases by about 70.2% when θ equals 24° (n=15). This is due to the enhancement of natural convection and reduction of the thermal resistance, which are caused by the larger heat transfer surface area for more fins.

The more fins set, the faster the heating process in the system. However, the change rate of phase change time slightly decreases when the θ larger than 36° (n=10), especially for 26° (n=14) and 24° (n=15). This is due to the fact that the presence of fins interferes with natural convection. Therefore, there is a limit to the interval angle θ and fin number n.

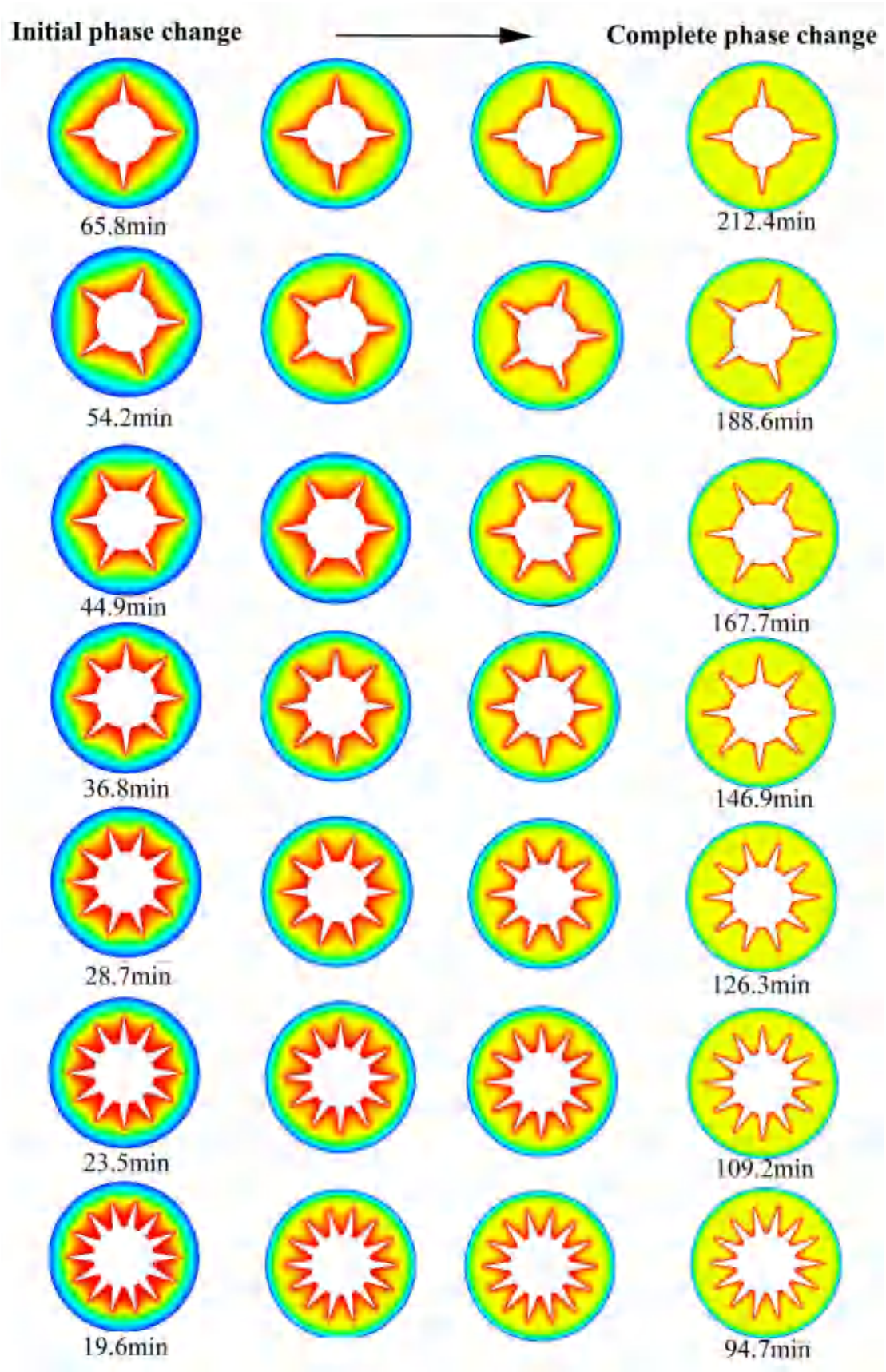


Fig. 10. Temperature distribution under different θ

2) Comparison with rectangular fins

Fig. 11(a) shows the comparison of phase change time between rectangular fins and triangular fins for various θ. The heating efficiency **gradually improves** as the fin design evolves from triangular to rectangular. θ=24° promote **63.83%, 56.34%, 46.73%, 46.73%, and 31.71%** heating rate compared with θ=72°, 60°, 45°, 36°, 30°. It should be pointed out that the time difference between both fins becomes smaller as the θ decreases.

Fig. 11(b) shows the liquid fraction curves for different θ during melting. The rectangular fins are more efficient than triangular fins by reducing the total melting time. Taking the θ of 72°, 45°, 30°, and 24° for examples, the heating efficiency of rectangular fins can improve by **11.95%, 14.99%, 15.43, and 15.62%** compared with that of triangular fins. The improved heating efficiency here means a relative difference in melting time between both cases at the same θ.

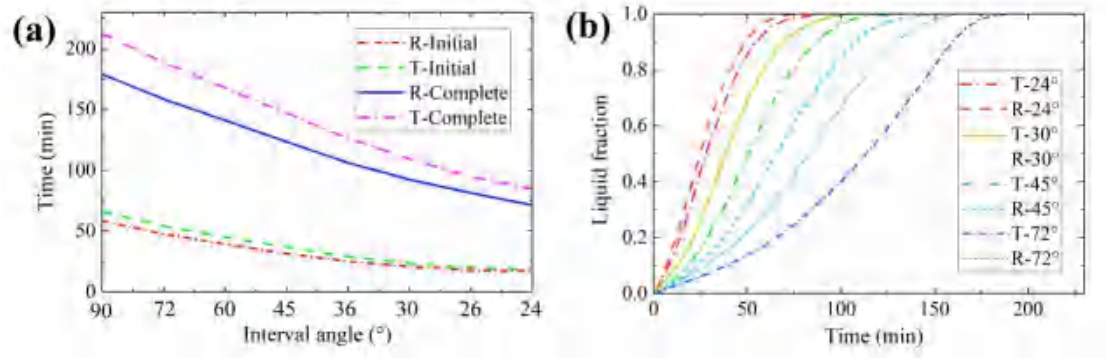


Fig. 11. Comparison between rectangular fins and triangular fins for various θ: (a) phase change time, (b) liquid fraction over time

(Note: T means triangular fins, and R means rectangular fins)

5.2 Effect of the fin length

Four fin lengths (L) changing from 24mm to 6mm were studied. Fig. 12 depicts the cross-section under various L. The θ was 30° and and the W was 6mm.

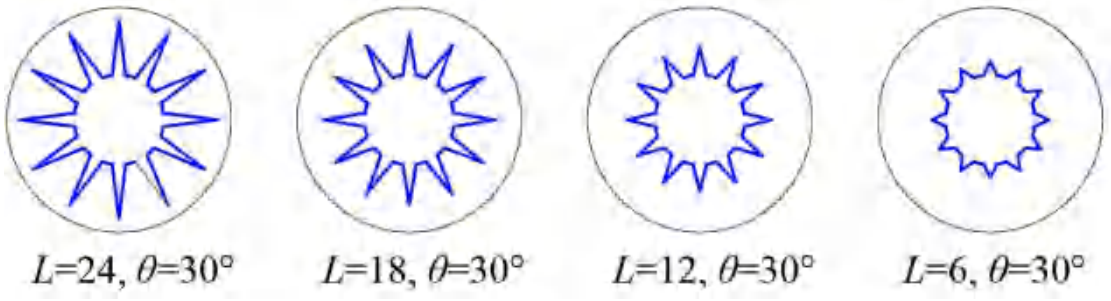


Fig. 12. Cross-section under various L

1) Influence on heat transfer

Fig. 13 shows the temperature distribution under various L of triangular fins. The melting time decreases by about 70.4% when the L changes from 12mm to 24mm. This is because, at the early stage of melting, the fin tip of the longer fin first contacts

the PCM far away. This leads to the promotion of natural convection in the periphery of the tank. Therefore, longer L can improve the heat transfer process obviously.

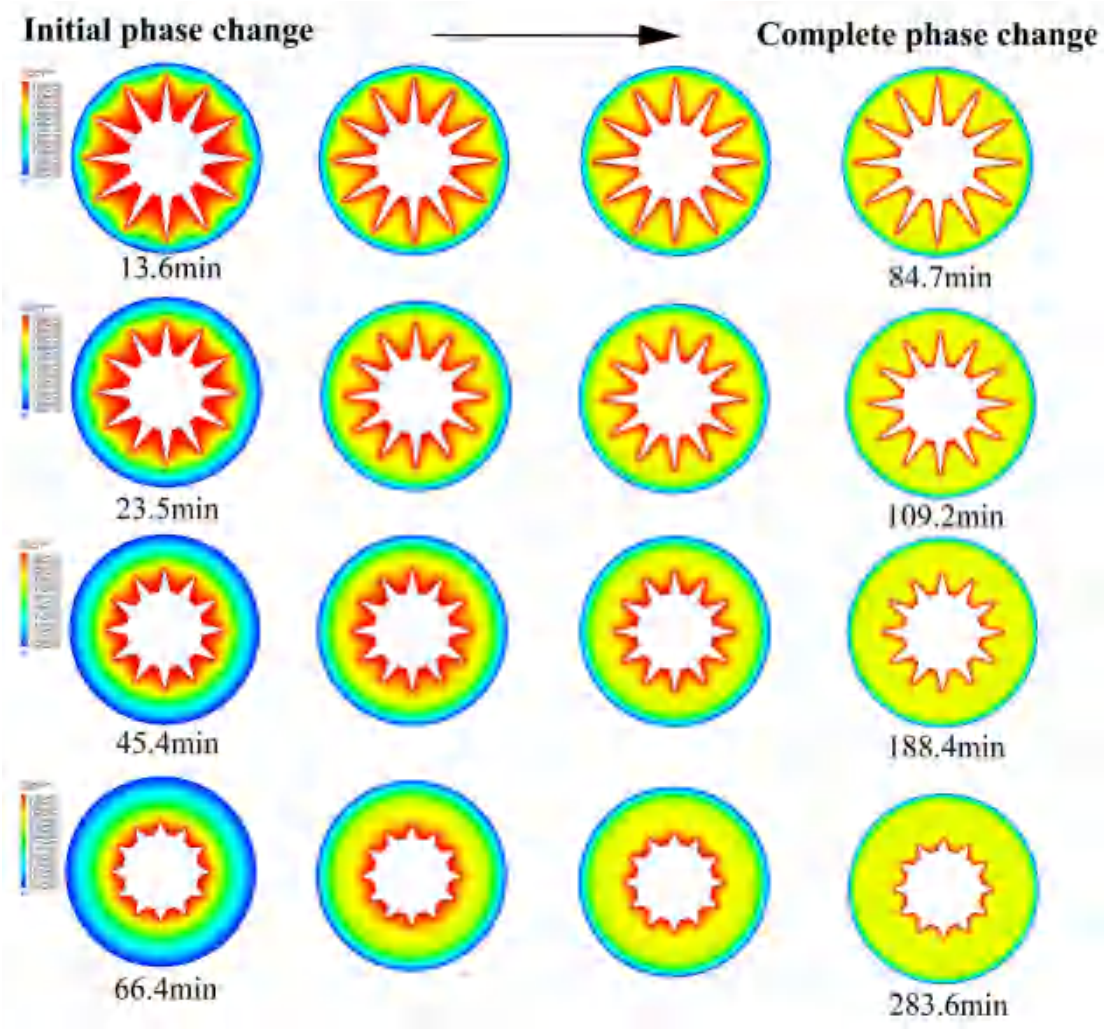


Fig. 13. Temperature distribution under different L

2) Comparison and analysis

Fig. 14(a) shows the comparison of phase change time between different L. The heating efficiency **significantly improves** as the increase of L. **For initial phase change**, L =24mm can promote **79.52%, 70.04%, and 42.17%** heating rate compared with L =6mm, 12mm, and 18mm, respectively. **For complete phase change**, the melting time of L =24mm can save **70.13%, 55.06%, and 22.43%** compared with L=6mm, 12mm, and 18mm, respectively.

Fig. 14(b) shows the liquid fraction curves for different L during melting. It still can be seen that long fins are more efficient than short fins.

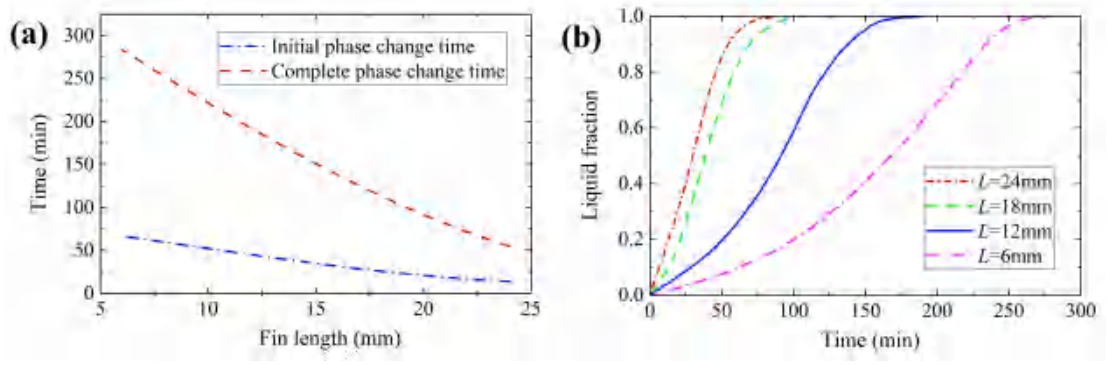


Fig. 14. Comparison between various L: (a) phase change time, (b) liquid fraction over time

5.3 Effect of the fin width

Four fin widths (W) changing from 10mm to 4mm were studied. Fig. 15 depicts the cross-section under various W. The θ was 30°and the L was 18mm.

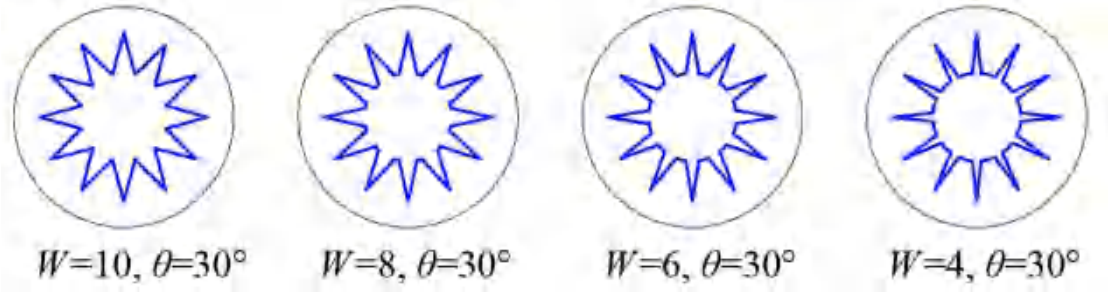


Fig. 15. Cross-section under various W

1) Influence on heat transfer

Fig. 16 shows the temperature distribution under various W of triangular fins. The melting time decreases by about 35.9% when the W changes from 4mm to 10mm. Therefore, larger W can slightly improve the heat transfer process.

2) Comparison and analysis

Fig. 17(a) shows the comparison of phase change time between different W. The heating efficiency **slightly improves** with the increase of W. **For initial phase change**, W=10mm can promote **7.6%, 22.55%, and 35.91%** heating rate compared with W=8mm, 6mm, and 4mm, respectively. **For complete phase change**, the melting time of W=10mm can save **11.13%, 21.79%, and 30.52%** compared with W=8mm, 6mm, and 4mm, respectively.

Fig. 17(b) shows the liquid fraction curves for different W during melting. It still can be seen that wide fins are more efficient than narrow fins.

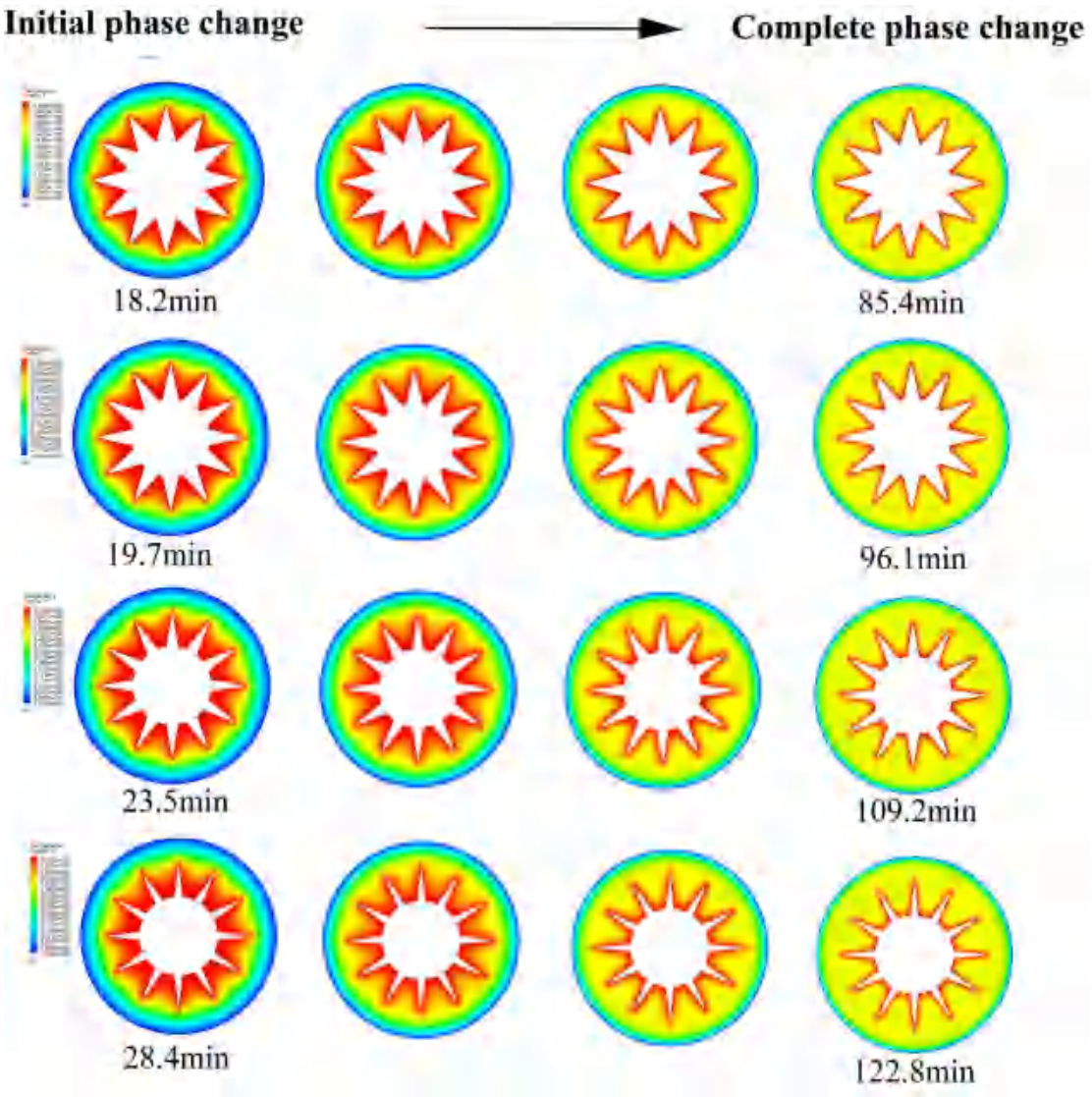


Fig. 16. Temperature distribution under different W

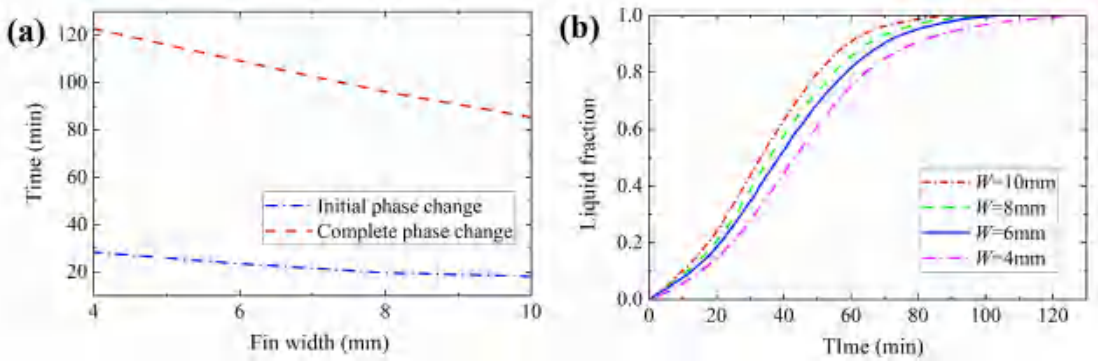


Fig. 17. Comparison between various W: (a) phase change time, (b) liquid fraction over time

6. Modeling and solution of Problem 3

6.1 Problem analysis

The optimal spatial distribution of fins is required to achieve the best heat transfer capacity. The optimization aim is to maximize the heat transfer efficiency in the heat storage tank.

Based on the conclusions of questions 1 and 2, we have known that the optimal geometrical shape is closely related to the length L, width W, and number n of fins. For branching structure design, topology optimization and fractal design can be used to obtain the optimal spatial distribution.

1) Design by topology optimization

Topology optimization is a shape optimization method that uses algorithms to optimize material layout within a limited space. It maximizes the performance and efficiency of structures by removing redundant material and reaching the optimal design goal. The principle of topology optimization is shown in Fig. 18(a).

2) Design by fractal optimization

Fractal geometry is a geometric shape containing detailed structure at arbitrary scales, usually having a fractal dimension strictly exceeding the topological dimension. Fractal can generate the optimal fractal shape under the constraints of the objective function. Fig. 18(b) shows the principle of fractal optimization.

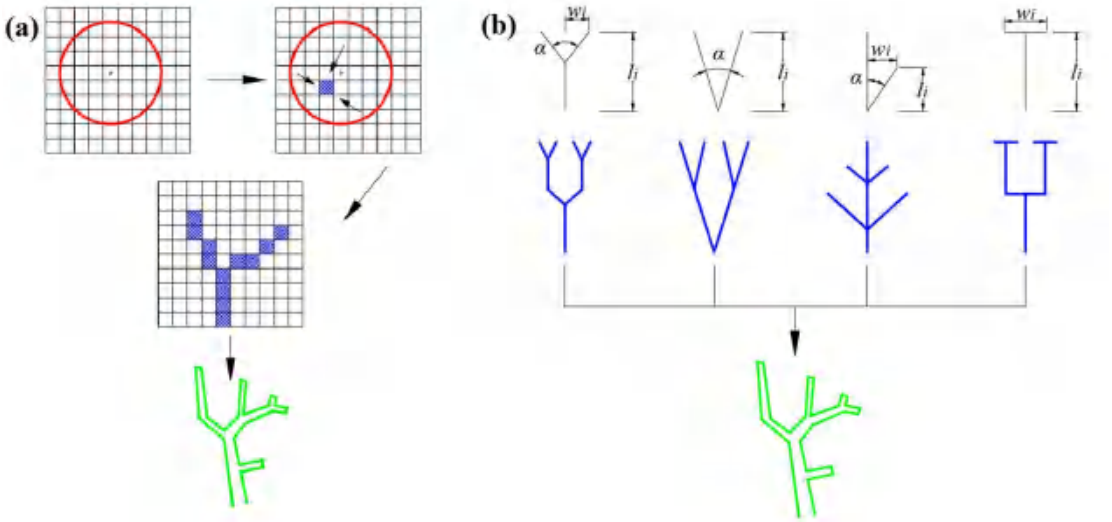


Fig. 18. Principle diagram of fin optimization: (a) topology optimization, (b) fractal design

6.2 Topology optimization

6.2.1 Variable density method

To find the optimal fin shape, the variable density method (VDM)[6] is used to optimize the spatial distribution of fins. The VDM introduces material with variable imaginary density and corresponding physical properties, to transform structure

optimization problems into material layout problems.

The interpolation relationships of each variable between original and imaginary are redefined by solid isotropic material with penalization (SIMP) [6], which can be written as:

$$k(x) = (k_{HCM} - k_{PCM})(\rho_e)^{\eta} + k_{PCM} \tag{19}$$

$$\rho(x) = (\rho_{HCM} - \rho_{PCM})\rho_e + \rho_{PCM} \tag{20}$$

$$\lambda(x) = (\lambda_{HCM} - \lambda_{PCM})\rho_e + \lambda_{PCM} \tag{21}$$

$$L(x) = (1 - \rho_e)L \tag{22}$$

$$\theta(x) = (\theta_{HCM} - \theta_{PCM})\rho_e + \theta_{PCM} \tag{23}$$

where $k(x)$ is the element stiffness, $\rho(x)$, $\lambda(x)$ $L(x)$, and $\theta(x)$ are pseudothermal density, conductivity, fin length, and interval angle respectively, ρ_e is the pseudo density, $\rho_e \in [0,1]$, η is penalty factor, $\eta \in [0,1]$.

The solution steps of topology optimization based on VDM are expressed as:

$$\begin{cases} \text{find } \boldsymbol{x} = (x_1, x_2, \cdots, x_n) \\ \text{minimum } \mathbf{T} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbf{T_i} \\ \mathbf{KT} = \mathbf{Q} \\ \text{constraint} \quad \dfrac{1}{V}\displaystyle\sum_{i=1}^{n} V_i \rho_i \le \varphi \\ \rho_{min} \le \rho_i \le 1 \end{cases} \tag{24}$$

where **T** is the temperature array, **K** is the stiffness matrix, **Q** is the heat load matrix, n is the number of elements, V_i is the volume of element i, φ is the maximum volume fraction, and $\rho_{\min}$ is the minimum pseudo density which equals 10^{-9} to avoid the occurrence of singular matric.

The problem from Eq. (19) to Eq. (24) is an approximate mathematical programming problem of continuous variables, whose Lagrangian function is[7]:

$$L = Q + \beta_1(V - f_v V_0) + \beta_2^T(KT - Q) + \sum_{i=1}^{n} \beta_3^i (\underline{x}_i - x_i) + \sum_{i=1}^{n} \beta_4^i (x_i - \overline{x}_i) \tag{25}$$

According to the Kuhn-Tucker rule[7], at the optimum point there is:

$$\frac{\rho(x)^{\eta} T_i k_0 T_i}{V_i} = \beta_1 \tag{26}$$

6.2.2 Optimization result

The optimization goal was the maximum average temperature in the PCM region. The finite element software ANSYS was still used in the optimization process for the steady-state solution. The method of moving asymptotes (MMA)[8] was used for calculation. Optimized tolerance was 10^{-3}. The initial value of the element pseudo density ρ_i was 0.5, and the maximum volume fraction φ was set to 0.05. To get

explicit topology results, the penalty factor η is usually set to 3.

The maximum average temperature was 362.46K after 500 iterations in MATLAB. The fin distribution obtained by topology optimization is shown in Fig. 19(a).

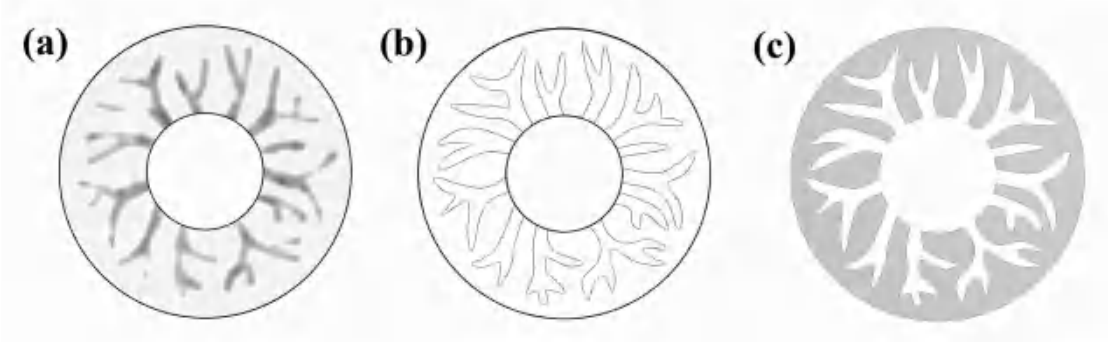


Fig. 19. Topological shape of fin distribution: (a) topology optimization, (b) image reconstruction, (c) final fin distribution used in ANSYS

6.2.3 Image reconstruction

Although SIMP can effectively reduce the intermediate value, the junctions between fins and PCMs exist due to the continuity of design variables. Therefore, the fuzzy boundary needs to be filtered out to make the topological graph sharper. The reconstruction function is chosen as[9]:

$$x_p = \frac{\tanh(\beta\mu) + \tanh(\beta(x_f - \mu))}{\tanh(\beta\mu) + \tanh(\beta(1-\mu))} \tag{27}$$

where x_p is the mapped variable, x_f is the initial variable, μ is the mapping threshold which is taken as 0.5, β is the control parameter of mapping steepness.

The shape image of optimal fins can be obtained by drawing the node coordinates. For the convenience of numerical simulation in ANSYS, the optimized shape is converted into an 'Auto CAD' image for further processing. The reconstruction result of the fin distribution is shown in Fig. 19(b). The fin distribution used in ANSYS is depicted in Fig. 19(c).

6.2.4 Heat transfer process

Fig. 20 shows the temperature distribution of the topological optimization. Fig. 21 shows the phase change time and liquid fraction, respectively. Compared with rectangular fins and triangular fins, the heating efficiency of topology optimization fins significantly improves. **For initial phase change**, topology optimization fins can promote **18.84% and 28.17%** heating rate compared with rectangular fins and triangular fins. **For complete phase change**, the melting time of topology optimization fins can save **41.81% and 50.82%** compared with rectangular fins and triangular fins, respectively.

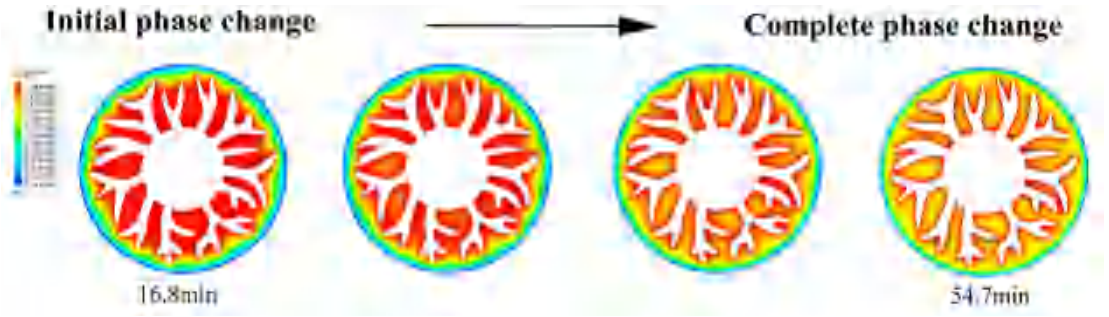


Fig. 20. Temperature distribution of topological optimization shape

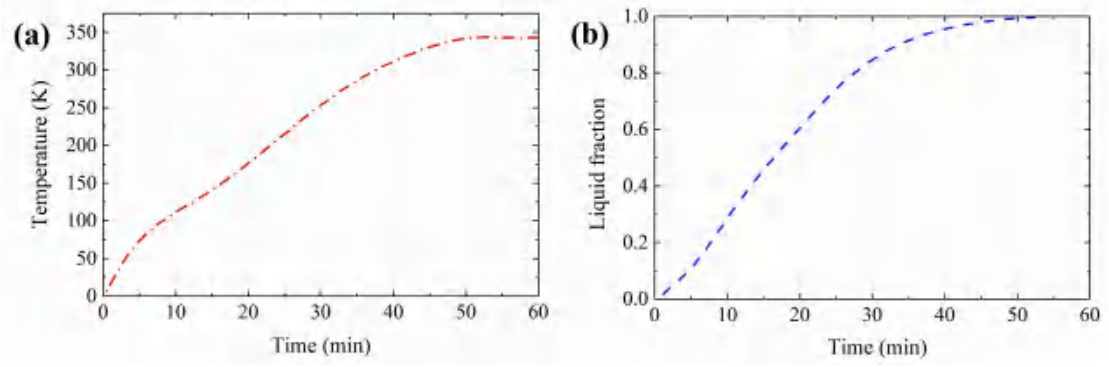


Fig. 21. Temperature variation and liquid fraction

6.3 Fractal optimization

As can be seen that the optimized fin structure obtained by topology optimization is irregular, which may cause difficulties in the manufacture and application of fin components. Therefore, A regular fin distribution is proposed optimally by fractal optimization in this section.

6.3.1 Murray's law and branch generation

The fractal fins have statistically self-similar and multi-scale properties, which can be considered as a mimicking of a fern leaf. The optimal transportation of leaf is derived by a relationship known as Murray's law[10]:

$$r^3 = r_1^3 + r_2^3 + r_3^3 + \cdots + r_n^3 \tag{28}$$

where is the radius of the parent tube, r_1 to r_n are the radii of the branching tube. Based on the coupling of Murray's law, the generalized Murray's law can be obtained as[11]:

$$r^\gamma = \frac{1}{1-\zeta}\sum_i^n r_i^\gamma \tag{29}$$

where ζ is the heat variation ratio, γ is determined by the mode of transport, and n is the number of child tubes.

The fern-fractal fins were designed by the generalized Murray's law. The exponent γ was taken 2 to achieve better thermal performance for the heat conduction [12]. The construction method of the fractal fin was introduced as follows:

Step 1: the parent branch grows two children branches. The length and width of

the parent branch were L_0 and W_0, respectively. The length and width of the first level of children's branches were L_1 and W_1, and the second level were L_2 and W_2. The terminals of the children branches were close to the outer wall of the tank.

Step 2: The lengths of any two successive branches (i.e., the parent branch and the middle child branch) meet the following relationship, which is written as

$$L_i = \alpha L_{i-1} \tag{30}$$

where α is the length ratio that reflects the length growth law.

Step 3: The correlation between the widths of the parent and children branches can be calculated by

$$W_i^2 = \frac{1}{3} W_{i-1}^2 \tag{31}$$

Step 4: Repeat the above steps 1 to 3 until the desired branching level. The number of branches per level was optimized by maximum average temperature as well. Then the fern-fractal fins are constructed.

6.3.2 Optimization result

The fin distribution obtained by Fractal design is shown in Fig. 22(a). The fin distribution used in ANSYS is depicted in Fig. 22(b).

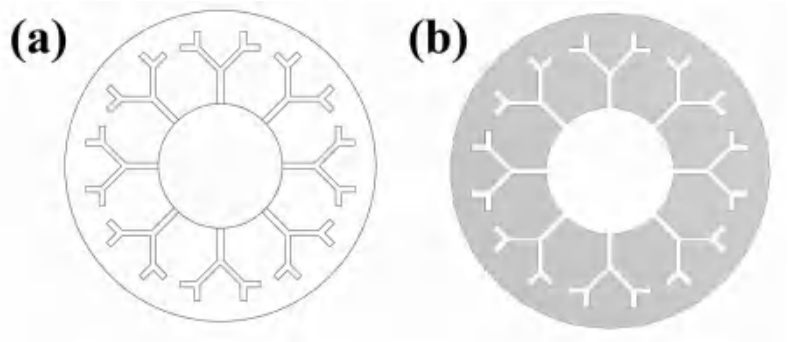


Fig. 22. Fractal shape of fin distribution: (a) fractal optimization, (b) finial fin distribution used in ANSYS

6.3.3 Heat transfer process

Fig. 23 shows the temperature distribution of the fractal optimization. Fig. 24 shows the phase change time and liquid fraction, respectively. Compared with rectangular fins and triangular fins, the heating efficiency of fractal optimization fins significantly improves. **For initial phase change**, fractal optimization fins can promote **14.01% and 24.25%** heating rate compared with rectangular fins and triangular fins, respectively. **For complete phase change**, the melting time of topology optimization fins can save **32.57% and 42.94%** compared with rectangular fins and triangular fins, respectively.

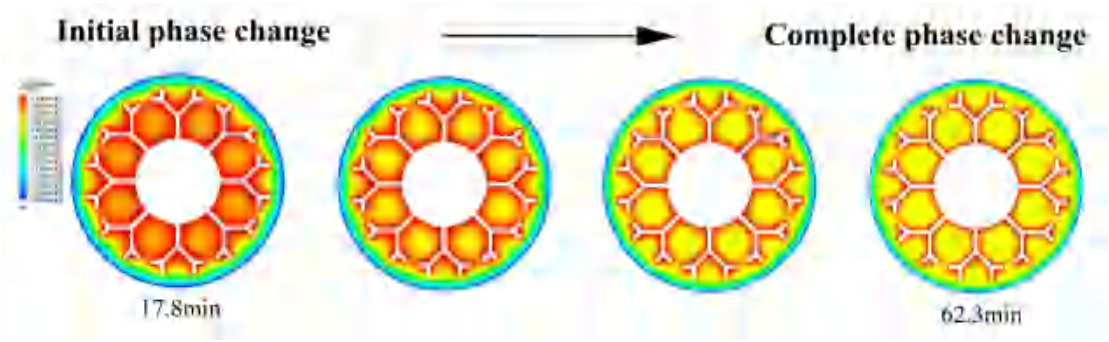


Fig. 23. Temperature distribution of fractal optimization shape

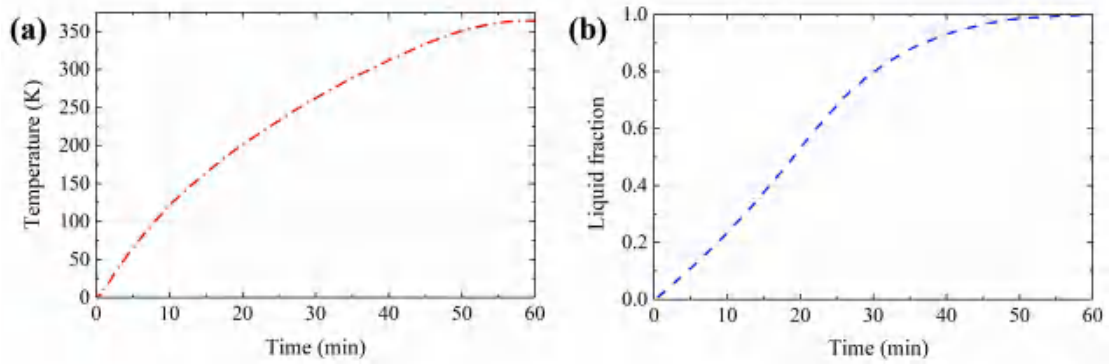


Fig. 24. Temperature variation and liquid fraction

6.4 Comparison of two models

According to the phase change time and liquid fraction, topology optimization is more efficient than fractal optimization. For initial and complete phase change time, **topology optimization increases by 5.61% and 13.8% heating rate compared with fractal optimization.**

7. Recommendation letter

Dear company leaders,

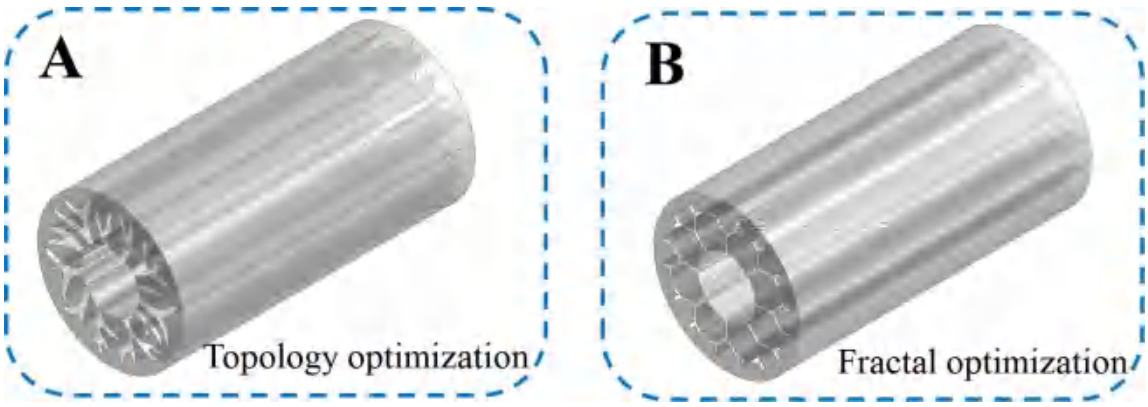
We are writing this letter to provide you with some suggestions to optimize the heat transfer fins in the heat storage tank. To achieve the optimal fin configurations, We have devised two optimization models to obtain the optimal fin configurations. The design suggestions are summarized based on the model simulation and structural analysis.

First, parametric studies of fins were conducted to preliminarily investigate the effects of shape, geometry, and spatial distribution on heat transfer efficiency. Generally speaking, fins configurations with smaller interval angles and larger lengths and widths are more efficient. The priority order of fin parameters is **fin length > interval angle > fin width**. But the thermal efficiency will weaken when the fin configurations exceed a certain limited value. Based on our research, **the recommended ranges of design parameters** are as follows:

Parameter	Interval angle (°)	Length (mm)	Width (mm)
Recommended range	24~45	18~24	6~10

In addition, topology optimization was conducted to generate an arborescent fin distribution as shown in Fig. A. However, the optimized fin structure obtained by topology optimization is **irregular** most of the time, which may cause difficulties in the manufacture and application of fin components. Therefore, I suggest that this topology optimization structure can be adopted in **PCMs with very low thermal conductivity**.

Finally, fractal optimization was performed to obtain a **regular** fin distribution as shown in Fig. B. This method provides more possibilities for engineering applications. But it should be pointed out that the hearing efficiency of fractal optimization is less efficient than that of topology optimization. Therefore, this fractal optimization can be used in **general heat storage tanks** and **PCMs with low thermal conductivity**.



We hope that the proposed models, results, and suggestions could provide you with valuable information.

Yours sincerely,

Team Dapmcm2201058

8. Model evaluation and promotion

7.1 Evaluation

The sensitivity of exponent γ in the fractal model was tested by changing the fractal dimension and comparing the difference between different configurations. The exponent γ was taken 2, 3, and 4, respectively, to investigate the influence of fractal dimension (i.e. bifurcation number) on the temperature variation.

Fig. 25 shows the comparisons of temperature variation under different γ. It can be seen that the temperature rise is slightly promoted with the increase of γ. This is because no significant change in the spacing between adjacent fins due to the control of fractal tree. Therefore, the fractal design model has good stability.

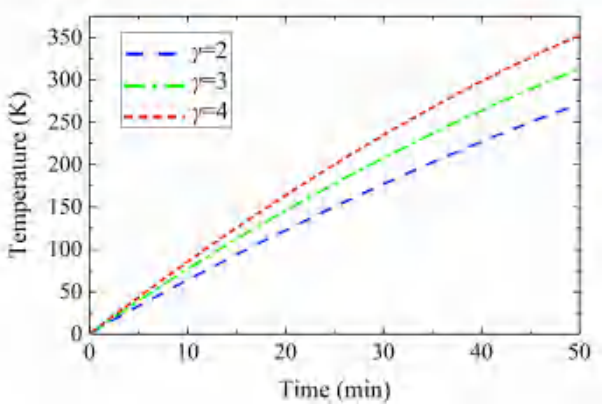


Fig. 25. Comparisons of temperature variation under different γ

7.2 Promotion

1) In order to obtain the real features of heat transfer, the viscous dissipation and thermal expansion should be considered in Eq. (7) and (9), because large velocity gradients may occur during the heat flow in the PCM domain.

2) Considering the actual structure, three-dimensional topology optimization should be conducted to obtain a tridimensional structure. The fins can be designed along the longitudinal direction or layered layout.

9. Strength and weakness

7.1 Strength

1) When predicting the optimal fin distribution, the fractal optimization model has **better model stability** compared with the topology optimization model. However, it also should be pointed out that the fin distribution obtained by topology optimization us more efficient for the heat transfer.

2) The topology optimization model can find the optimal fin distribution **considering various balance factors** in optimization objectives.

3) The proposed topology and fractal optimization models can **automatically**

search the optimal fin distribution under the constraint of objective functions.

7.2 Weakness

1) Most of the time, the optimized shape obtained by topology optimization is irregular, which causes difficulties in the manufacture and application of fin components.

2) Objective function plays a critical role in the final optimization results. When selecting different optimization objectives, the final optimization results may be different. Therefore, objective function could be diverse in solving the optimal fin distribution.

Reference

[1] Voller VR, Brent AD, Prakash C. Modelling the mushy region in a binary alloy [J]. Applied Mathematical Modelling, 1990, 14(6):320-326.

[2] Cheng Y, Xuan ZB, Xi CB, et al. Melting performance enhancement of a latent heat storage unit using gradient fins [J]. International Journal of Heat and Mass Transfer, 2022, 150:11930.

[3] Al-Abidi AA, Mat S, Sopian K, et al. Internal and external fin heat transfer enhancement technique for latent heat thermal energy storage in triplex tube heat exchangers [J]. Applied Thermal Engineering, 2013, 53(1):147-156.

[4] Seddegh S, Joybari MM, Wang X, et al. Experimental and numerical characterization of natural convection in a vertical shell-and-tube latent thermal energy storage system [J]. Sustainable Cities and Society, 2017, 35:13-24.

[5] Mao Q, Li Y, Li G, et al. Study on the influence of tank structure and fin configuration on heat transfer performance of phase change thermal storage system [J]. Energy, 2021, 235:121382.

[6] Bendse MP, Sigmund O. Material interpolation schemes in topology optimization [J]. Archive of Applied Mechanics, 1999, 36: 635-654.

[7] Geng DC. Basis of engineering structure optimization design [M]. Dalian University of Technology Press, 2012.

[8] Krister D. The method of moving asymptotes - a new method for structural optimization [J]. International Journal of Numerical Methods Engineering, 1987, 24:359-373.

[9] Zhao CY, Wu ZG. Heat transfer enhancement of high temperature thermal energy storage using metal foams and expanded graphite [J]. Solar Energy Materials and Solar Cells, 2011, 95(2): 636-43.

[10] Mcculloh KA, Sperry JS, Adler FR. Water transport in plants obeys Murray's law [J]. Nature, 2003, 421:939-942.

[11] Zhou Y, Kassab GS Molloi S. On the design of the coronary arterial tree: a generalization of Murray's law [J]. Physics in medicine and biology, 1999, 44(12):2929-2945.

[12] Xu P, Sasmito AP, Yu B, et al. Transport phenomena and properties in treelike networks [J]. Applied Mechanics Reviews, 2016, 68:1-9.

Team Number :	apmcm2213503
Problem Chosen :	E

2022 APMCM summary sheet

Analysis and Prediction of Global Nuclear Weapon

Abstract

Nuclear weapons are the most powerful weapons ever developed in the history of mankind, and it is because of the power of nuclear weapons that countries want to protect themselves through the deterrent effect of nuclear weapons. Although nuclear weapons are powerful, when they explode, they release enormous energy and cause serious damage to the ecological environment.

In response to question one, this paper makes a **basic data analysis** based on the relevant data on nuclear weapons of various countries given in the question. First, visualize the status of nuclear weapons owned by all countries, the number of nuclear weapons stored and the number of nuclear tests conducted. Use python to traverse relevant data and complete **descriptive statistical analysis**. Build a **critical weight scoring model**, select seven indicators to score the attitude of different countries towards nuclear weapons, and judge that **North Korea** is the most active country towards nuclear weapons.

For the second question, in order to predict the number and trend of nuclear weapons in the future, this paper constructs a prediction model based on **ARIMA time series analysis**. First, the data is preprocessed, and the **ADF stability test and white noise test** are performed. Then we evaluate and test the model to find the most appropriate **super parameter** to bring into the **prediction model**. Next, the data of possession status and storage of nuclear weapons of each country are brought in, and the prediction and numerical analysis are carried out. Finally, we get that the total number of nuclear weapons in **2123 will reach 2152** and the number of nuclear weapons in each country.

For the third question, this paper gives two definitions of destroying the earth. One is to blow the earth to pieces with nuclear weapons. Through the calculation of the **gravitational binding energy** of the earth and the **explosive capacity of nuclear weapons**, it is found that the current number of nuclear weapons is **far from enough**. The second definition is to destroy the majority of living things on the earth with **nuclear weapons and its radiation**. We convert the world map into a grayscale image, build the **DBSCAN clustering model**, and set the explosion radius and propagation radius of nuclear weapons to obtain the total number of nuclear bombs that destroyed the earth finally.

In response to question four， we submitted a **non-technical article** to the United Nations, explained our team's views on nuclear weapons and related findings with the previous model analysis results, and made several suggestions for all countries. We believe that the ultimate goal of developing nuclear weapons is to **eliminate nuclear weapons**.

Key words： Critical weight method; ADF stability test; ARIMA time series analysis; DBSCAN clustering model

Contents

1 Introduction

1.1 Background

Nuclear weapons, including neutron bomb, atomic bomb and hydrogen bomb, are related to nuclear reaction and have great destruction. Among all the weapons invented in human history, nuclear weapons are one of the most powerful weapons. It is because of the power of nuclear weapons that many countries hope to protect their homeland from invasion through the deterrent of nuclear weapons.

A single atomic bomb has the power to destroy a city. Considering the specific explosion parameters, the temperature of the explosion moment of the atomic bomb can reach tens of millions of degrees, and the explosive yield is similar to tens of thousands to hundreds of thousands of tons of TNT. But although nuclear weapons are very powerful, but the destruction of the ecological environment is extremely serious. The detonation of a nuclear weapon can make it difficult for the ecological environment of the explosion area to recover for many years.

1.2 Problem Restatement

Question one：Basic data analysis

According to the requirements of the title, data analysis should be conducted on the acquired data. Firstly, it needs to determine which countries the nuclear weapons have existed in the past. The countries that have reduced their stockpiles the most in the last 20 years and those that have increased their stockpiles the most; Taking five years as a span, the time period of the most nuclear weapon tests from 1945 to 2019 is calculated, and the most active countries in the research of nuclear weapons in the recent 10 years are selected. Finally, the countries that have changed from "not considering nuclear weapons" to "possessing nuclear weapons" are analyzed.

Question two：Forecast nuclear weapons

Mathematical models are built from the data to predict the number of nuclear weapons and predict which countries will have nuclear weapons within the next 100 years. A mathematical model is used to predict the number of nuclear weapons in the next 100 years, the total number of nuclear weapons in the world in 2123, and the total number of nuclear weapons possessed by each country.

Question three：Protect our planet

Build a mathematical model of where a nuclear weapon will detonate, and calculate how many nuclear bombs it will take to destroy the Earth. From the perspective of the individual nuclear weapon, the mathematical model is established to determine the maximum destructive power of the current nuclear bomb and analyze whether it can destroy the Earth. The above study is about the destructive power of nuclear weapons, while the protection of the ecological environment is a topic of constant human research. In order to protect the ecological environment, the limit of the total number of nuclear bombs in the world and the theoretical limit of nuclear weapons of the states possessing nuclear weapons should be analyzed.

Question four：Complete a non-technical essay

Combining the team's findings and writing to the UN in a non-technical article, it makes several recommendations for all countries.

2 Analysis of Question

2.1 Analysis of Question One

The primary purpose of Question 1 is to conduct a basic analysis of the data, which contains five small questions:

Question a: status=3 in the data (country-position-nuclear-weapons.csv) indicates that the country possessed nuclear weapons in that year. Therefore, by traversing the data, we can find that the country with status=3 is the country that once possessed nuclear weapons.

Question b: In order to analyze the problem more accurately, we will first visualize the data of nuclear weapons stockpile (nuclear-chirf-stock.csv), and then analyze which country's nuclear weapons stockpile has decreased or increased the most. Since there are different measures, two methods have been developed in order to analyze the problem more comprehensively. The first is the difference between the number of nuclear weapons in each country's last year and the number in each country's earliest year. The second approach is to look at the very different number of nuclear weapons in each country. Then the results of the two methods are compared. If they are consistent, the accuracy of the results can be better explained.

Question c: Firstly, the number of nuclear weapon tests in each year is obtained by processing the number-of-nuclear-weapons-weapons tests. Then, the number of nuclear weapon tests occurring in each "5-year" is calculated, and the five years with the highest number of nuclear weapon tests can be obtained after comparison.

Question d: By establishing CRITIC model and combining with nuclear weapon test data, we scored each country's nuclear weapon research activity, so as to get the most active country in nuclear weapon research.

Question e: We will analyze the selected countries that once possessed nuclear weapons, subtract the earliest year when status=3 and the latest year when status=0 to get the time interval for these countries to change from "not considering nuclear weapons" to "possessing nuclear weapons", and then compare them to get the result.

2.2 Analysis of Question Two

The main objective of question two is the projection of nuclear weapons, including the projection of the future nuclear-weapon States and the projection of the future number of nuclear weapons. Based on the data analysis, we choose to establish the nuclear weapon number prediction model based on ARIMA, and the establishment flow chart is as follows

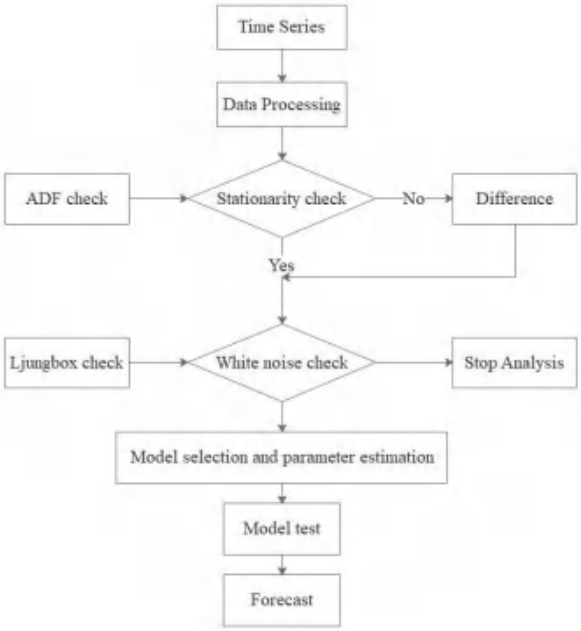


Figure 1 ARIMA model building flowchart

As can be seen from the figure, we need to preprocess the time series data first. Due to the prerequisite requirements of ARIMA model establishment, we need to conduct stationarity test and white noise test. After passing the stationarity test and white noise test, the model estimation and test are carried out, and finally the ARIMA model can be successfully established.

The number of nuclear weapons forecast model based on ARIMA is established to predict the countries with nuclear weapons in the next 100 years and the change trend of the number of nuclear weapons, and to predict the total number of nuclear weapons in the world and the number of nuclear weapons in each country in 2123.

2.3 Analysis of Question Three

There are different definitions of destroying the earth. In order to analyze the problem more comprehensively, we set up two definitions of destroying the earth. The first definition is to tear the earth apart, and on the basis of the first definition a proper analysis is carried out. The second definition is to cause serious damage to the ecological environment on the earth surface. Based on this definition, we will establish the detonation location model of nuclear weapons through DBSCAN clustering algorithm based on the population density as the standard, and then analyze and solve the problem, judge how many nuclear weapons are needed to destroy the earth, and calculate the limit of nuclear weapons in the world.

2.4 Analysis of Question Four

Based on the results of the above solution and our findings, we will write a non-technical article to the United Nations with our recommendations for all countries.

3 General Assumptions and Notations

To simplify our modeling, we make the following assumptions:

1) It is assumed that the collected data ignores the influence of information disclosure degree.
2) Suppose that the standard for the earth to be destroyed is not the earth to shatter, but the ecological environment on the earth surface to be seriously damaged.
3) The analysis of the number of nuclear weapons in this paper is limited to the countries in the collected data, and other countries are not analyzed.

Furthermore, we have a list of the main notations we define during our modeling, as shown in Table 1.

Table 1 notations

Symbol	Description
x_{ij}	The number of nuclear bombs that the i country has in the j year
δ_i	The number of nuclear bombs that the i country has possessed in the last 20 years is extremely low
$score$	CRITIC review score
U	Gravitational binding energy
R	The damage range of a nuclear weapon explosion
Eps	The minimum distance between two points
minPoints	The minimum number of points that form a dense area

4 Model and Solution of Question One

4.1 Data Collection

Question one requires a quantitative analysis of the situation related to nuclear weapons, so we study the question by collecting sufficient data. The data include four aspects: (1) data on the proliferation of nuclear weapons. This part contains two data, one is the annual proliferation of nuclear weapons in the world (nuclear-weapons.csv), and the other is the proliferation of nuclear weapons in each country (country-position-nuclear-weapons.csv). In the latter, "proliferation" is a description of the status of countries' possession of nuclear weapons. In the data, the number of states of countries' nuclear weapons has four different values, each of which has its own significance. status=0 means not considering nuclear weapons, status=1 means considering nuclear weapons, status=2 means pursuing nuclear weapons, and status=3 means possessing nuclear weapons. (2) Storage data of nuclear weapons (nuclear-warhead- stock.csv). To some extent, the stockpile of nuclear weapons can fully reflect the

distribution of nuclear weapons. (3) number of nuclear weapon tests (number-of-nuclear-weapons-tests.csv). To some extent, the number of nuclear weapons tests can reflect a country's activity in nuclear weapons research and the degree of demand for nuclear weapons.

4.2 Problem Solving

4.2.1 The Solution of Problem a

In order to determine which country has ever possessed nuclear weapons, we look for data (country position-nuclear-weapons.csv) on the number of states of possession of nuclear weapons by different countries from 1938 to 2022, and traverse all the years with python. A state of 3 indicates that the country had nuclear weapons in that year. As is shown below, a total of ten countries have possessed nuclear weapons.

Table 2 Countries that once possessed nuclear weapons

Country	China	France	India	Israel	North Korea
	Pakistan	Russia	South Africa	United Kingdom	United States

South Africa, in particular, only had nuclear weapons between 1979 and 1991. Nine other countries still have nuclear weapons by 2022.

4.2.2 The Solution of Problem b

4.2.2.1 Data Visualization

Firstly, we searched the data of nuclear weapons storage (nuclear-kit-stock.csv), and visualized the data to show the changing trend from 2003 to 2022 with the help of python drawing, as shown in the figure below.

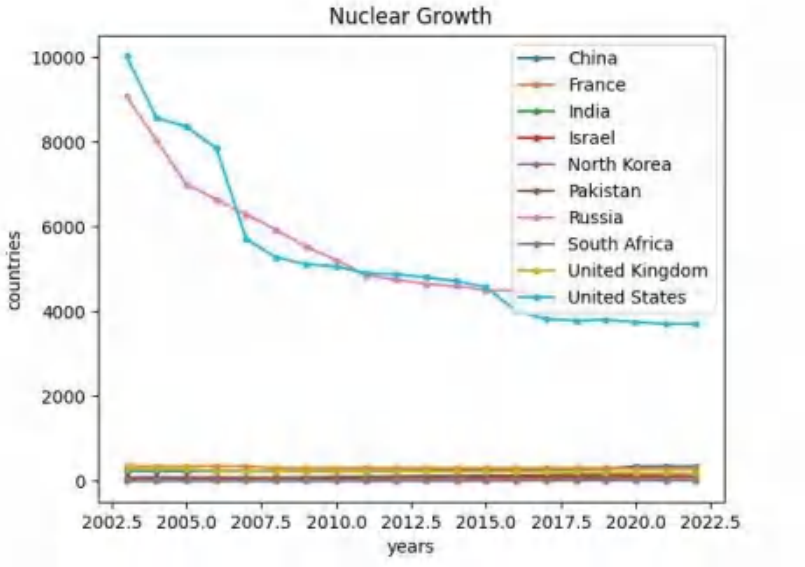


Figure 2 Nuclear growth

The image shows that the number of nuclear weapons in the United States and Russia far exceeded that of other countries in the past 20 years, and showed a rapidly decreasing trend. The number of nuclear weapons in other countries during this period was basically no more than 500.

Since the number of nuclear weapons of the United States and Russia is far more than that of other countries, in order to make the trend of the change of the number of nuclear weapons of other countries clearer, we have made the chart of the change of the number of nuclear weapons of other countries except the United States and Russia, as shown in the figure below.

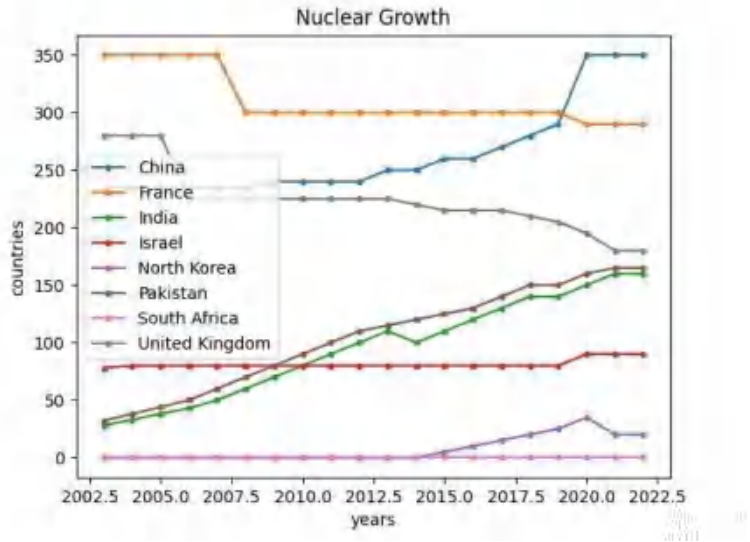


Figure 3 Nuclear growth(others)

As can be seen from the chart, the general trend of changes in the number of nuclear weapons varies from country to country. There are countries that are on an overall downward trend, and there are countries that are on an overall upward trend.

4.2.2.2 Problem Solving

To find out which country's nuclear weapons stockpile has decreased or increased the most over the past 20 years, we objectively divide it into two parts.

Method one： Will ever have nuclear weapons by the first letter sorting, respectively $x_1, x_2 \cdots x_{10}$, might as well t for years, is the first i country in the first j the number of years the bomb for x_{ij}, 20 years of this variable for $change$, among them $j = t - 2003$, $change = x_{i20} - x_{i1}$, Calculate the final result by plugging in the following formula. If the change is greater than 0, then

$$in = \max(x_{i20} - x_{i1}), i = 1, 2 \cdots 10 \tag{1}$$

If the change is less than 0, then

$$de = \min(x_{i20} - x_{i1}), i = 1, 2 \cdots 10 \tag{2}$$

in Represents the maximum increase, de represents the maximum decrease.

Method two： Note that the difference between the most and least number of nuclear

bombs possessed by i countries in the past 20 years is δ_i, which is calculated as

$$\delta_i = \max\{x_{ij}\} - \min\{x_{ij}\}, i = 1,2\cdots 10,\ j = 1,2\cdots 20 \tag{3}$$

And compare the results with Method one, as shown in the following table.

Table 3 Variation of nuclear weapons

Country	Variation（Method one）	Range（Method two）
United States	-6319	6319
China	115	115
France	-60	60
India	132	132
Israel	12	12
North Korea	20	35
Pakistan	133	133
Russia	-4599	4776
South Africa	0	0
United Kingdom	-100	100

It is easy to see that according to Method one, the stockpile of nuclear weapons in the United States has been reduced the most, by 6319, or 60.43%. Pakistan's nuclear weapons stockpile increased by the most, by 516%, with 133 weapons. According to Method two, America has the largest range in the past 20 years. Combined with the visual chart, it is not difficult to see that America has the largest reduction during this period. The second is Pakistan, which has the largest increase in the past 20 years.

By comparing the two methods, it can be seen that the results are consistent: the United States is the country with the largest reduction in nuclear weapons inventory, while Pakistan is the country with the largest increase in nuclear weapons inventory, which can better illustrate the accuracy of the results.

4.2.3 The Solution of Problem c

By adding the test times of each country in the same year in the data table of nuclear weapon test times (number-of-nuclear-weapons-tests) through python, we get the number of nuclear weapon test times of each year in the world. Then, taking each year as the basis, we extend back for five years, respectively. Each "five years" is regarded as a group, and the number of nuclear weapon test occurrence of each group is calculated. To better visualize the change in the data, the graph below shows the change in the number of nuclear weapons tests.

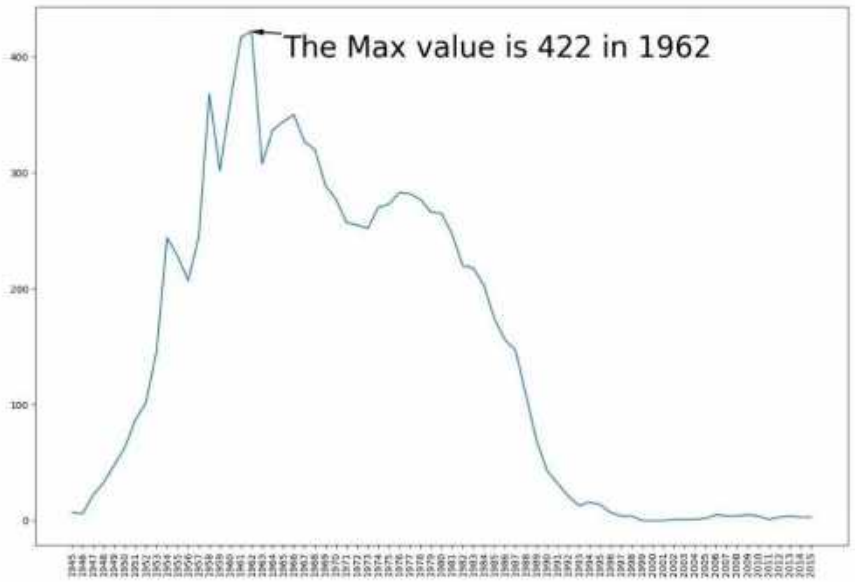


Figure 4 Changes in the number of nuclear weapon tests

Due to limited space, only some groups are shown in the abscissa annotations in the figure. The figure can intuitively observe the changes in the number of nuclear weapon tests, and it can be known that over time, the number of nuclear weapon tests in the world is gradually declining.

A comparison of the total number of tests in each group gives the five years with the highest number of nuclear weapons tests, as shown in the table below.

Table 4 The five years with the highest number of nuclear weapons tests

Year	Number of nuclear weapons tests
1962	178
1963	50
1964	60
1965	58
1966	76
Total	422

It can be seen from the table that the five years with the largest number of nuclear weapon tests were 1962, 1963, 1964, 1965 and 1966, with a total of 422 times.

4.2.4 The Solution of Problem d

4.2.4.1 Data Processing

The valid data is obtained by secondary processing of the data found on the previous web address. Firstly, the following indicators are selected as the criteria to judge whether a country

is active in nuclear weapons, namely "nuclear weapons status", "nuclear weapons stockpile" and "nuclear weapons tests". "the difference over ten years", "deployed strategic", "centrally deployed nonstrategic", "centrally deployed non deployed", "retired warhead". There are 8 indicators in total, which are respectively denoted as $s_1, s_2 \cdots s_n, n = 8$.

Where s_1 is the state of nuclear weapons, respectively 0,1,2,3 according to the meaning of the question to represent different degrees of state. Since s_2 and s_3 were given data in different years, we added the average value of the data into the table. In order to describe the activity degree of a country more objectively, From the initial data itself, we construct a new list of indicators called s_4, representing the difference before and after the decade. A table with 10 rows and 8 columns can be obtained by adding the following 4 indicators.

Since the order of magnitude of each index is different, we need to compare them in the same range, that is to say, all indexes need to be normalized. For convenience, we will directly standardize them here. There are *m* objects to be evaluated and *n* evaluation indicators, which can constitute the data matrix $X = \left(x_{ij}\right)_{m \times n}$. Let the elements in the data matrix be $x_{ij}^{'}$ after the positive and standardized processing of indicators. All the selected elements in this paper are positive indicators, so there is no need for positive processing.

4.2.4.2 Establishment and Solution of the Model

Here, the CRITIC method is adopted to evaluate and score these countries, and the country with the highest score is selected as the most active country. This method considers the comparison intensity of evaluation indicators and the conflict between indicators to comprehensively measure the weight of indicators, while taking into account the correlation between indicators. In essence, it means that the larger the number is, the greater the weight is not assigned. Instead, it makes full use of the attributes of data to evaluate projects, which is an objective and scientific weighting method.

The first point to be considered here is comparison intensity, which refers to the value gap between the same index and other evaluation schemes, that is, to consider the variability between different indicators, which is expressed by standard deviation. In other words, the higher the standard deviation, the greater the fluctuation, the higher the weight should be. The specific calculation is as follows:

$$\begin{cases} \overline{x}_j = \dfrac{1}{n}\sum_{i=1}^{n} x_{ij} \\ S_j = \sqrt{\dfrac{\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2}{n-1}} \end{cases} \tag{4}$$

Next, we consider the conflicts among indicators, which can be expressed by the correlation coefficient R_j. The calculation method is as follows:

$$R_j = \sum_{i=1}^{p}\left(1 - r_{ij}\right) \tag{5}$$

When there is a strong positive correlation between two indicators, the smaller the conflict is, the lower the weight will be. For the CRITIC method, when the standard deviation is constant,

when the degree of positive correlation between the two indicators is greater, that is, when the correlation coefficient is close to 1, the conflict will reach the minimum, which indicates that the information reflected by the two indicators in the evaluation scheme has great similarity.

Finally, calculate the amount of information C_j, which takes the product of the conflict and variability of the previously calculated indicators and actually represents the objective proportion of the j evaluation indicator subchild in the whole evaluation system. The larger the value, the more weight should be assigned, so it is positively correlated with the final weight.

$$C_j = S_j \sum_{i=1}^{p}\left(1-r_{ij}\right) = S_j \times R_j \tag{6}$$

Finally, the objective weight is calculated as follows:

$$W_j = \frac{C_j}{\sum_{j=1}^{p} C_j} \tag{7}$$

Add the weights and indices here to get the final score

$$score = \sum_{j=1} w_j * x_{ij} \tag{8}$$

Here we use MATLAB calculation to write corresponding programs to solve this problem, the final score is as follows

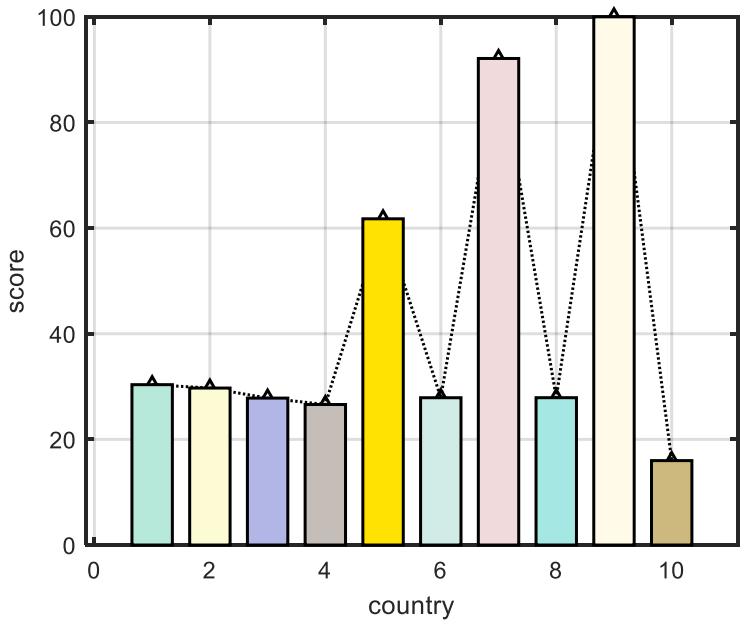


Figure 5 CRITIC score bar chart

The figure above intuitively shows the rating of countries by CRITIC, among which three countries have significant scores and others have low scores. The table below shows the specific scores of these countries.

Table 5 CRITIC rating result

Country	Evaluation score	Country	Evaluation score
North Korea	100.0000	Indian	27.8934
China	92.1177	Israel	27.8865
France	61.7613	Russia	27.8156
Britain	30.3602	United States	26.6016
Pakistan	29.7192	South Africa	15.9810

According to the table, North Korea received the highest score of 100.0. Although North Korea has a small stockpile of nuclear weapons, it has increased its nuclear weapons significantly in recent years, has a large stockpile of nuclear weapons, has tested more nuclear weapons, and scored the highest score of 100.0. So overall, North Korea has been the most active in nuclear weapons research over the past decade.

4.2.5 The Solution of Problem e

We searched the data of different countries' possession states of nuclear weapons from 1938 to 2022 (country-position-nuclear-weapons.csv), screened out 10 countries from 3.2.1 for analysis, and iterated all the years through python. When the status quantity is 0, it means that the country does not consider researching nuclear weapons in that year (status=0); when the status quantity is 3, it means that the country has possessed nuclear weapons in that year (status=3). In these countries, we subtract the earliest year when status=3 appears from the latest year when status=0 to obtain the time interval for these countries to change from "no consideration of nuclear weapons" to "possessing nuclear weapons", and the results are shown in the table below

Table 6 The time interval of transformation

Country	Year	Experience time（year）
China	1954~1964	10
France	1953~1960	7
India	1963~1987	24
Israel	1954~1967	13
North Korea	1979~2006	27
Pakistan	1971~1987	16
Russia	1942~1949	7
South Africa	1973~1979	6
United Kingdom	1940~1952	12
United States	1941~1945	4

As can be seen from the table, the United States experienced the fastest transition from "not considering nuclear weapons" to "possessing nuclear weapons" in only four years, from 1941 to 1945.

5 Model and Solution of Question Two

5.1 Prediction Model of Nuclear Weapon Number Based on ARIMA Time Series Analysis

5.1.1 Prediction Model Based on ARIMA Time Series

A time series is a sequence of numbers in chronological order that shows the values of different periods of time. ARIMA model is an analysis of random time series, its core is the probability analysis of the sequence itself, that is, the random attribute. ARIMA(p,d,q) model is used to establish a stationary sequence, where d is the order of difference, AR is autoregressive, and p is the corresponding autoregressive term. MA is the moving average and q is the number of moving average terms.

The ARIMA model will capture a series of different standard time structures in the time series, and the ARIMA model can separate noise and signal, and deduce them, so as to obtain accurate prediction results, especially suitable for the fitting of non-stationary data. The mathematical expression of ARIMA model is as follows:

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)(1-L)^d X_t = \alpha_0 + \left(1+\sum_{i=1}^{q}\beta_i L^i\right)\varepsilon_t \tag{9}$$

$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)$ is $\mathrm{AR}(p)$, $(1-L)^d X_t$, $(1-L)^d X_t$ is d Degree Differential, $\left(1+\sum_{i=1}^{q}\beta_i L^i\right)\varepsilon_t$ is $\mathrm{MA}(q)$. The ARIMA model is $X_t \sim \mathrm{ARIMA}(p,d,q)$.

5.1.2 Stationarity Test

Before constructing the ARIMA model, the stationarity test of the sequence should be carried out, and the following three conditions should be met to determine that $\{x_t\}$ is covariance stationarity:

1) $E(x_t) = E(x_{t-s}) = u$ (The mean is a constant)
2) $Var(x_t) = Var(x_{t-s}) = \sigma^2$ (The variance is constant)
3) $\mathrm{cov}(x_t, x_{t-s}) = \gamma_s$ (The covariance is only somewhat related to the interval s)

If any $t_1, t_2, \cdots, t_k$ (k is an arbitrary value) and h are satisfied, the multidimensional random variables $(x_{t_1}, x_{t_2}, \cdots, x_{t_k})$ and $(x_{t_1+h}, x_{t_2+h}, \cdots, x_{t_k+h})$ have the same joint distribution, then $\{x_t\}$ can be considered to be in a strictly stationary condition.

Logarithmic sequence of the number of nuclear weapons of each country is taken for ADF test, and the results are shown in the table below.

Table 7 ADF test results

Variable name	*t* statistic	1%	5%	10%	*p* value	Inspection result
Num	-2.2912	-3.8724	-3.3958	-3.2132	0.1379	unstable
Δ*Num*	-12.3285	-2.6123	-1.8921	-1.6037	0.0000	steady

As can be seen from the table, the test statistics after the difference of the sequence show significance at the level of 1%. Therefore, the sequence is stable, and the ARIMA model can be used to predict the number of nuclear weapons in each country.

5.1.3 White Noise Test

In this part, we carry out white noise test on the sequence to further explore the rationality of the model. If it is white noise, it means that the sequence is extremely random, and the historical situation has nothing to do with the future development, so there is no need for more research. If the sequence is not white noise, the model can be used for quantitative prediction. In order to prevent the data from being full of randomness, this paper must detect the white noise of the residual sequence. Its essence is to judge whether the data has autocorrelation, and use autocorrelation graph to further test whether the sequence data has validity.

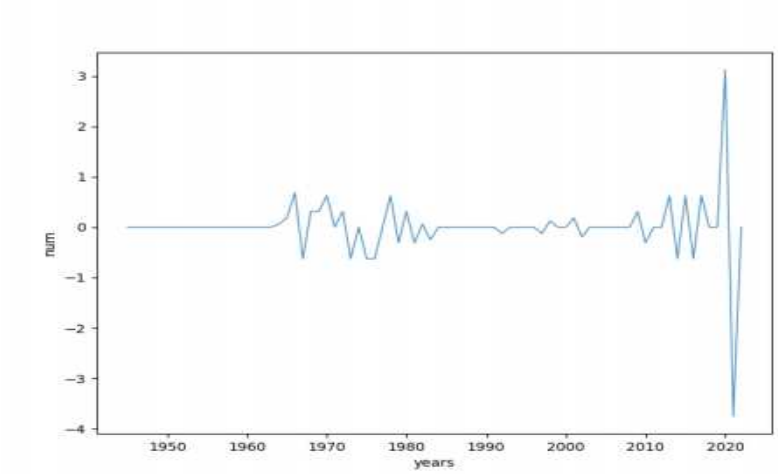


Figure 6 White noise

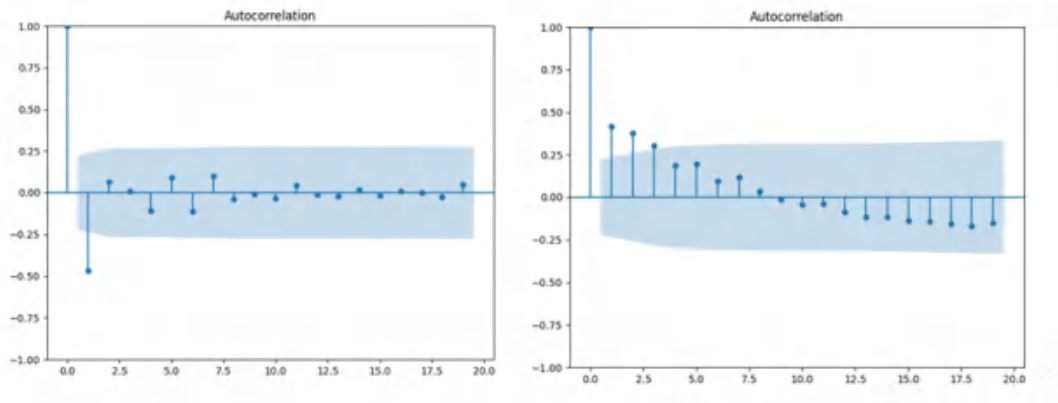


Figure 7 Autocorrelation(left) and Partial Autocorrelation(right)

Autocorrelates and partial autocorrelates are bounded by light blue bars in the graph, and the length of the sequence may be denoted as L. Since the data follows the normal distribution, it can be considered that the data is autocorrelated. Meanwhile, it is found that the characteristics of the data and white noise are inconsistent in many places. Therefore, the data successfully passes the white noise test, so the ARIMA model can be established.

5.1.4 Model Estimation and Test

We substituted the data set into the model for repeated training, so as to obtain the results meeting the error requirements, as shown in the figure below

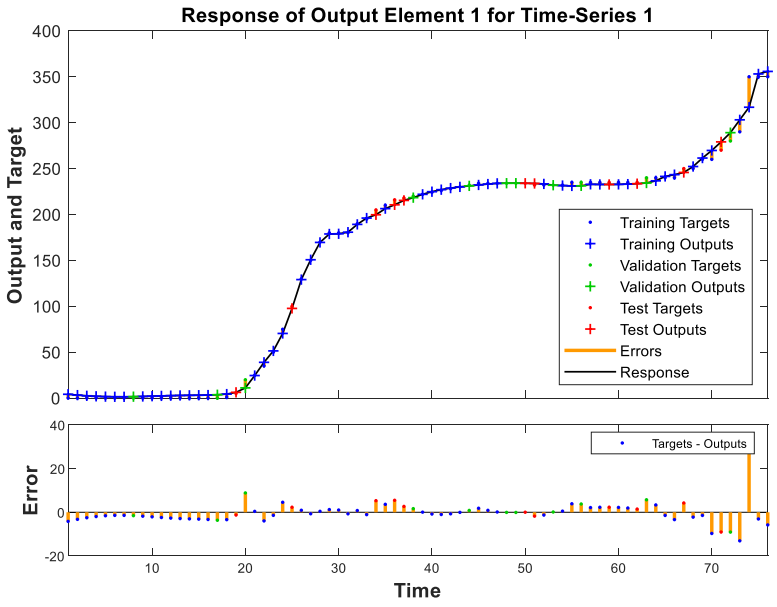


Figure 8 Error analysis diagram

As can be seen from the figure, the errors are generally within a reasonable range and are always distributed around 0, which indicates that the model has high accuracy and can obtain effective analysis results. After repeated experiments, the optimal model is determined. In the ARIMA(2,1,2) model, AIC and BIC were the smallest, and the best model ARIMA(2,1,2) with better fitting effect was obtained after experimental comparison. The specific model is as follows:

$$\Delta Num_t = 1.3124\Delta Num_{t-1} - 0.6524\Delta Num_{t-2} - 1.0058\mu_{t-1} + 0.4632\mu_{t-2} \quad (10)$$

Where, Num_t represents the sequence of the number of nuclear weapons after logarithm. μ_t indicates the random error term.

Because of

$$\Delta Num = dNum = Num - Num(-1)$$
$$\Delta Num = (1-L)Num_t \quad (11)$$

namely

$$(1-L)Num_t = 1.3124(1-L)Num_{t-1} - 0.6524(1-L)Num_{t-2} - 1.0058\mu_{t-1} + 0.4632\mu_{t-2}$$

$$Num_t = 2.3121Num_{t-1} - 1.8947Num_{t-2} + 0.6524Num_{t-3} - 1.0079\mu_{t-1} + 0.4632\mu_{t-2}$$

In order to ensure the correctness and applicability of the model, the correlation test of residual error is carried out in this paper. By running python, it is known that the values of AC and PAC both tend to be close to 0, indicating that the residual sequence of ARIMA(2,1,2) mode has no autocorrelation, which means that the residual is white noise.

5.2 The Number of Nuclear Weapons and National Projections

After the time series successfully passed the stationarity test and white noise test, this paper established the nuclear weapon number prediction model based on ARIMA. By building a good model to predict the countries possessing nuclear weapons in the next 100 years, the specific results are shown in the figure below:

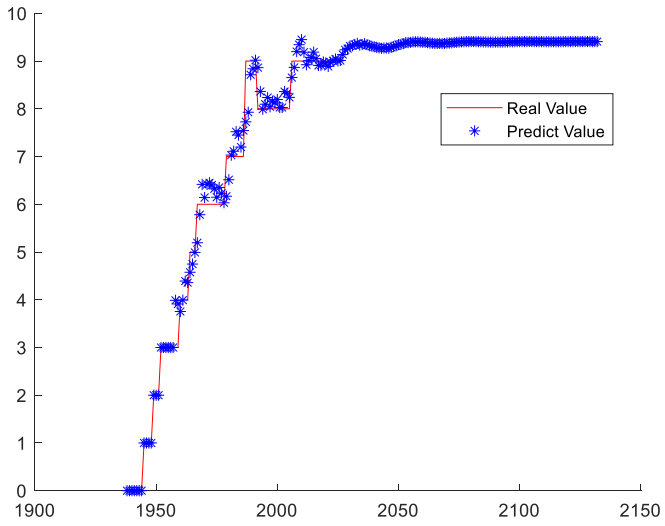


Figure 9 The predictions of nuclear-weapon states

As can be seen from the figure, with the passage of years, the number of states possessing nuclear weapons gradually rose to 9.62 and then leveled off, no longer increasing significantly, which is also related to the current environment of nuclear non-proliferation.It is easy to know that under the Nuclear Ban treaty, nine of the future nuclear weapon states should be the same as the current nine nuclear weapon states: China, France, India, Israel, North Korea, Pakistan, Russia, South Africa, United Kingdom, United States. Because of 9.62>9.5, hence the possibility of a tenth nuclear-weapon state in the future. Based on data analysis, it is presumed to be one of three countries: Brazil, Iran and South Africa. Brazil and Iran are pursuing nuclear weapons, so the possibility is high. There is also a strong possibility that South Africa has a certain research base because it once possessed nuclear weapons.

We built a good model to predict the number of nuclear weapons in the next 100 years, the total number of nuclear weapons in 2121. In order to more intuitively show the change of the number of nuclear weapons in the next 100 years, python is used to draw a graph to show the change trend of the number of nuclear weapons in the next 100 years, as shown in the figure below.

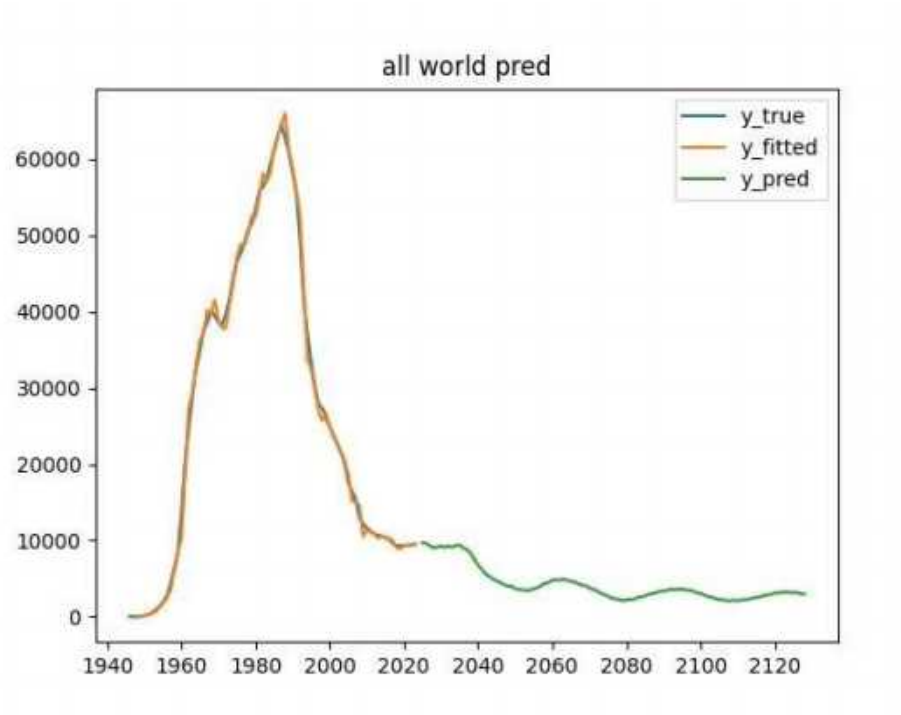


Figure 10 Trends in the number of nuclear weapons

The model predicts a total of 2,152 nuclear weapons in 2123. As can be seen from the figure, the real value is close to the predicted value, which indicates that the prediction effect of the model is good. As you can see from the graph, the number of nuclear weapons over the next 100 years will go up and down, but overall it will go down. The ban on nuclear weapons testing promoted in the Comprehensive Nuclear-Test-Ban Treaty adopted by the United Nations General Assembly has limited the growth in numbers to some extent, while the total number of nuclear weapons will continue to decline as nuclear weapons are decommissioned and dismantled, which is consistent with our model.

Since these countries have the same forecasting methods, we will use China as an example to analyze the number of nuclear weapons in China in the next 100 years.

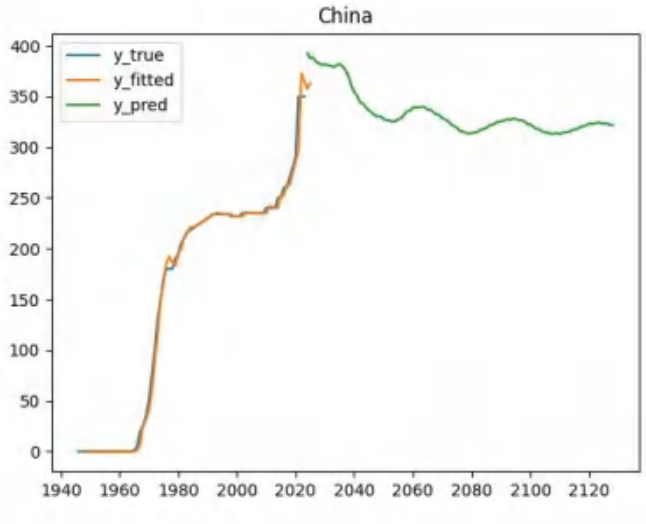


Figure 11 China's nuclear weapons numbers forecast

As can be seen from the figure, the number of nuclear weapons in China will generally show a downward trend in the next 100 years and gradually stabilize to 323.

6 Model and Solution of Question Three

6.1 Measurement of Nuclear Weapon Energy

（1）A brief introduction to nuclear Weapons

Nuclear bombs, also known as nuclear weapons, essentially use nuclear radiation, thermal radiation, shock waves and pulses to create enormous damage and radiation to the surrounding environment.

Nuclear weapons can be divided into tactical and strategic nuclear weapons. The use of tactical nuclear weapons is to detonate nuclear weapons in the strategic vanguard area of the target enemy, while the strategic nuclear weapons are to detonate intercontinental missiles in the important places far away from the enemy. The modes of occurrence are generally divided into sea, land and air. In contrast, launching in the sea is the most effective, directly reflected in the sea can reduce the likelihood of enemy detection.

（2）Nuclear weapon power measure

Equivalent is an important unit that reflects the power of nuclear weapons. It means the energy of nuclear weapons explosion is equivalent to the energy of TNT explosion. The most powerful nuclear weapon ever developed was the Soviet Czar Bomb, which had a yield of 50 megatons of TNT. So far, the total yield of nuclear weapons possessed by all countries adds up to about 15 billion tons. That's still a far cry from the 250 trillion tons produced when the planet hit Earth.

6.2 Definition of Earth Destruction

We currently define the destruction of the planet in two ways:

① Rocks were gravitationally bound together to form the Earth, so destroying the Earth would mean breaking it up into pieces of rock that would not be gravitationally bound back together.

② From the point of view of biological survival, the destruction of the earth means that the ecological environment and living conditions on the surface of the earth have been seriously destroyed, even can be said to be razed to the ground, resulting in the destruction of the earth is impossible to have biological survival.

6.3 Model Establishment and Solution

6.3.1 Model Establishment and Solution Based on Definition ①

Matter in nature is held together by its attraction to each other. If a body contains more than one part, the parts form a whole because of the gravitational interaction between them. If you want to separate the parts, you need to do external work. And the work needed to overcome

gravity increases as the force of gravity increases. Binding energy means the total energy of the different parts of an object bound together.

In the nucleus as a whole, protons and neutrons are bound together by the strong nuclear force, and the binding energy of these microscopic particles together is the nucleus binding energy. Similar to that is the gravitational binding energy, which holds the Earth together from matter in a stable state. So to blow up the Earth, the energy from the explosion has to be greater than the binding energy of gravity. The gravitational binding energy of the earth is calculated by the following formula.

$$U = \frac{\frac{3}{5}GM^2}{r} \tag{12}$$

G stands for gravity constant, $6.67\times10^{-11}\mathrm{N}\cdot\mathrm{m}^2/\mathrm{kg}^2$; M is the mass of the earth, is $5.97\times10^{24}\mathrm{kg}$; r is the radius of the Earth, which is 6371 km. Through calculation, the gravitational binding energy of the earth can be known $2.24\times10^{32}\mathrm{J}$.

The total number of nuclear weapons currently in existence is about 9,440, of which the atomic bomb is the main. According to the literature, the yield of the atomic bomb is relatively small, only a few tons, while the yield of the hydrogen bomb is measured in megatons. If all the nuclear bombs in the world exploded at the same time, it would generate the energy of $3.14\times10^{19}\mathrm{J}$, and it would take at least 7 trillion times more energy to destroy a planet on the basis of definition ①.

According to the descriptive statistical analysis of the data above, it can be known that there are 9,440 nuclear weapons in the world at present. The most powerful nuclear weapon known to have ever been developed was the Soviet Czar Bomb, which had an explosive yield of about 5 kilotons. Since the emphasis in the title is on the minimum number of nuclear weapons needed, we can assume that all existing nuclear weapons are the strength of Czar bombs. Based on this assumption, the formula for calculating the sum of the current explosive energy of all nuclear weapons is as follows

$$W = Num \cdot m_{\mathrm{TNT}} \cdot W_{\mathrm{TNT}} \tag{13}$$

In the formula, Num represents the total number of existing nuclear weapons; m_{TNT} represents the yield of the assumed nuclear weapons; W_{TNT} represents the energy released when one ton of TNT is exploded.

Since the current total number of nuclear weapons is 9440, the yield of Czar bomb is 5000 tons, and the energy released by one ton of TNT explosive is about 4.18×10^9 J, the total energy of nuclear weapons in the world can be obtained by substituting it into the formula: $W = 1.97296\times10^{17}$ J. The earth's gravitational binding energy 2.24×10^{32} J, by comparison, we can see that the total energy of global nuclear weapons and the earth's gravitational binding energy is far from each other, so it is not possible to destroy the earth on the basis of the definition of ①. From another point of view, if the Earth's gravitational binding energy is converted into the energy of the Czar bombs, a total of 316 trillion Czar bombs would be needed, which is far from enough given the current level of human technology. Therefore, on the basis of definition 1, it is impossible to blow the earth apart, and the ARIMA model predicts that

nuclear weapons will not blow the Earth up as the number of nuclear weapons is decreasing.

6.3.2 Model Establishment and Solution Based on Definition ②

6.3.2.1 Main Components of Nuclear Weapon Damage Radius

Based on the center of the explosion, the closest distance to the center is the radius of the fireball, that is, the central circle of the damage, followed by the detonation radius, radiation radius, and the first region of the shock wave. The power of a nuclear weapon decreases as its radius increases.

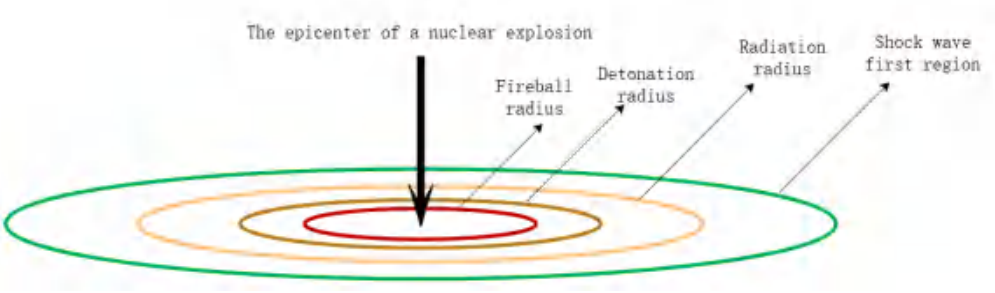


Figure 12 Nuclear explosion radius

1) The radius of the fireball is the central part of the explosion. Within this radius, the nuclear bomb is hot enough to evaporate everything. In other words, the nuclear weapon is energetic enough to burn everything within this radius.
2) The detonation radius is the second range circle outward from the central explosion point. The energy released by a nuclear weapon can severely destroy all buildings within the detonation radius. Since the radius of the fireball and the radius of the detonation are very small from the center of the explosion, the death rate of the creatures in it would be a horrible 100% if they were not hidden deep underground.
3) The radiation radius is the third circle of range outward from the center of the explosion. It's called the radius because the main thing that causes damage to living things in this area is radiation. Most living things will be seriously affected if they are exposed to severe radiation.
4) The first region of the shock wave is the fourth circle of range outward from the center of the explosion. An important component of the shock wave is atmospheric pressure, which is caused by the rapid expansion of the fireball and expands outward at a staggering speed of 200m/s, releasing the energy of a nuclear weapon explosion. After the shock wave, the expansion creates a vacuum, and the atmospheric pressure causes the atmosphere to backfill rapidly, which can cause serious damage.

The specific damage range of a nuclear weapon explosion is calculated by the following formula:

$$R = C \cdot \sqrt[3]{m_{\mathrm{TNT}}} \tag{14}$$

Where, C is the fixed value of 1.493885, and m_{TNT} is the explosive equivalent.

6.3.2.2 Nuclear Weapon Detonation Location Model Based on DBSCAN Clustering Algorithm

(1)Establishment of model

Based on the definition ② of destroying the Earth, we construct the DBSCAN clustering model to further judge the number of nuclear weapons destroying the earth. Since there are a large number of human beings living in areas with ecological environment and living conditions, we will perform DBSCAN clustering based on population density to obtain the total number of clustering categories, and the number of clustering is the number of nuclear weapons that destroy the earth. In order to visually display the situation of population density, we make the following figure to show.

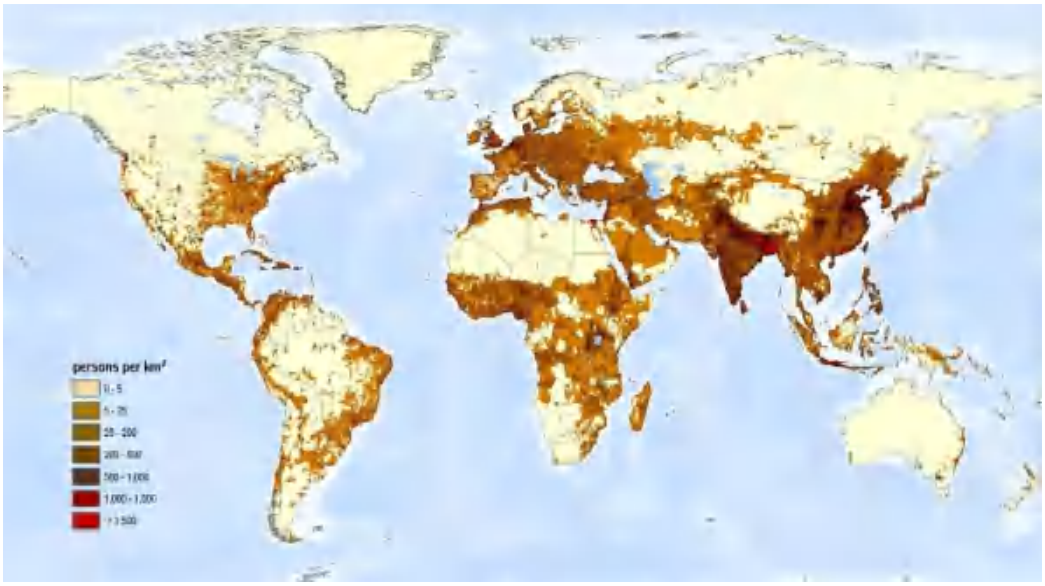


Figure 13 Global population density

The above figure is the distribution of global population density. We converted the above figure into a gray scale of 60000*10000 by using python as follows

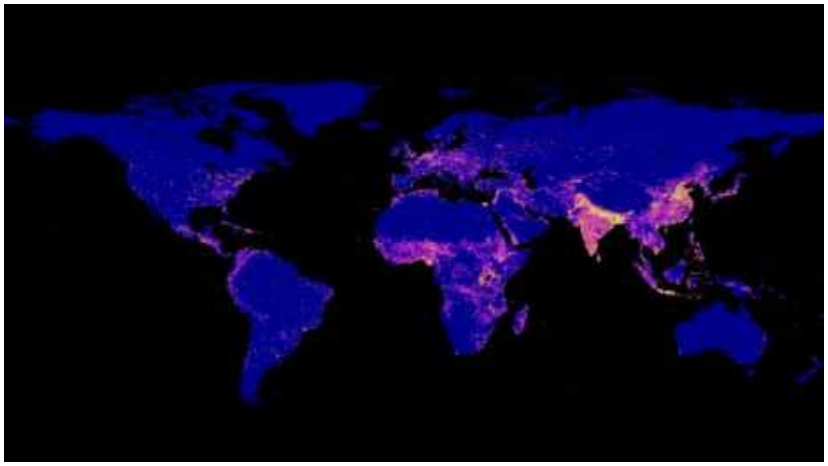


Figure 14 Grayscale map

The above figure is the gray level after the population density distribution map is converted, from which the general situation of population density distribution can be seen. Because we

consider irreversible damage to the human and much of the ecological environment of the Earth as a way to destroy the planet, we need to calculate the range and number of nuclear explosions needed to cover the entire planet.

DBSCAB is a density-based spatial clustering method. It does not need to determine the number of clusters in advance, and the number of clusters generated is random. The algorithm adopts density-based clustering method, and the number of targets must be contained in the clustering space should not be lower than the preset threshold. Even if there are noise points in the spatial database, the algorithm can still find clusters of arbitrary shapes. Large density areas close to each other can be connected by this algorithm, which can improve the effectiveness of data processing.

DBSCAN clustering algorithm has three types of data points:

Core points: With the core point as the center of a circle and Eps as the radius, the number of points in Eps is not less than minPoints

Boundary points: With boundary points as the center of the circle and Eps as the radius, the number of points in Eps is less than minPoints, but they are in the Eps neighborhood of the first type of data points

Noise points: With noise points as the center of the circle and Eps as the radius, the number of points in Eps is less than minPoints, and they are not in the Eps neighborhood of the first type of data points

Eps represents the minimum distance between two points. When the distance between two points is $d \le Eps$, they are considered to be adjacent. Eps selection is based on data sets. minPoints indicates the minimum number of points to form a dense area.

The flowchart of DBSCAN clustering algorithm is as follows

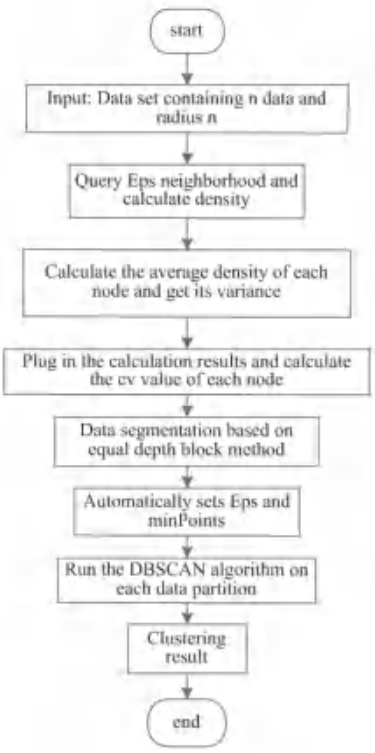


Figure 15 DBSCAN clustering algorithm flowchart

Based on DBSCAN clustering algorithm, this paper combined with the idea of density detection for analysis.

Density:

Step1：Density formula of node i:

$$\rho_i = \frac{|Pts(i)|}{\pi \cdot Eps^2} \tag{15}$$

Where Eps is the radius, $Pts(i)$ is the point set with center i and radius Eps, and $|Pts(i)|$ is the number of elements.

k is other points in the neighborhood of node i, whose density is defined as

$$\rho_k = get\_\rho\left(Pts_k\right), Pts_k \in Pts(i) \tag{16}$$

In the circle neighborhood, k is the center of the circle and Eps is the radius

Step2：Mean density

$$\overline{\rho}_i = \frac{\sum_{k=1}^{|Pts(i)|} \rho_k}{|Pts(i)|} \tag{17}$$

Step3： Density variance reflects the degree of deviation between each point in the Eps neighborhood and the mean value

$$s^2 = \frac{1}{n-1}\left[\sum \rho_i^2 - n \cdot \left(\overline{\rho}_i\right)^2\right], \quad n = |Pts(i)| \tag{18}$$

Step4：Coefficient of density variation, labeled in the neighborhood. The value of node density decreases with increasing number of points.

$$cv_i = \frac{s}{\overline{\rho}_i}, s = \sqrt{s^2} \tag{19}$$

Automatic Eps parameter setting for each partition:

The Eps of each region is defined as the average distance between each point and the previous point within its neighborhood

① Determine the value of minPoints

② $|Pts'(i)| = \rho_i \cdot \pi \cdot Eps^2$ Calculate the number of nodes in the Eps neighborhood of each point

③ $Eps(i) = \frac{|Pts'(i)|}{MinPts} \cdot Eps$

④ $Eps(i)$ is averaged after summation， $E(Eps) = \frac{\sum_{i=1}^{n} Eps(i)}{n}$

(2) Solution of model

The nuclear weapon detonation location model based on DBSCAN clustering algorithm was established to solve the problem, and the results as shown in the following figure were obtained

Figure 16 DBSCAN Cluster result

The graph above shows the population density of one of these countries. The same color represents the range of nuclear weapons and the range of nuclear radiation. According to the program, 5752 categories of clustering are obtained. Therefore, it is equivalent to 5752 Czar bombs or 14300 ordinary nuclear bombs are needed to destroy the earth on the basis of definition ②, which is not enough to completely radiate and contaminate the whole earth with the current scientific and technological level. With the development of science and technology, the power of nuclear weapons will continue to increase in the future, and the total number of nuclear weapons destroyed by the earth will decrease. Therefore, according to the results of the model, we believe that the total number of nuclear weapons in the world should be limited to 5,000 in theory, which is conducive to protecting the ecological environment. Russia and the United States have far more nuclear weapons than other countries, and the number of nuclear weapons in other countries is relatively small. Therefore, Russia and the United States are important factors in destroying the earth. Therefore, the first thing we should do is strictly limit Russia and the United States to 2000, and the rest of the countries cannot grow any more. Only on this basis can human beings further promote sustainable development and effectively protect the ecological environment.

7 A Non-Technical Article

Nuclear Weapons, Restriction or Development

Our findings:

In the 1960s, many countries conducted simultaneous high-volume nuclear tests to consolidate their dominance. Currently, there are about 10 countries with nuclear weapons in the world, and the number of nuclear bombs is negligible except for the P5, of which the United States and Russia have the most. To this day, many countries are still pursuing vigorous development of nuclear weapons.

From the prediction results of our model, even if all the harsh criteria are met, the energy generated by the explosion of all the existing nuclear bombs in the world will not be able to completely destroy the whole earth. However, if countries are allowed to develop nuclear weapons, once a large-scale war breaks out, it will cause enormous damage to the ecology of the entire planet. Therefore, it is imperative that we limit the number of nuclear weapons in the countries that possess them and change the direction of their development and competition.

Our suggestions:

From a military perspective, a nuclear threat policy is usually a policy that uses the power of nuclear weapons to threaten other countries. In the last century, countries have developed nuclear weapons to protect their territories. However, in the future, if countries still have conflicting interests, nuclear weapons will continue to exist due to their deterrent effect. Nuclear weapons will become the core military power of countries and promote the formation of mutual checks and balances among them. Therefore, in order to reduce the threat of nuclear weapons to human beings and the ecological environment, it is necessary to conclude agreements between countries to reduce the number of nuclear weapons in each country within a certain limit and to limit the use of nuclear weapons.

From the point of view of technological development, nuclear weapons are a great technological breakthrough in the history of mankind. In the 21st century, we should change the direction of its development and devote ourselves to the use of this highly sophisticated technology in agriculture, industry and medicine for the benefit of mankind. In order to achieve a sustainable planet and better use of resources in the future, we should also continue to study the theory and application of nuclear technology with the primary responsibility of protecting the ecological environment. We have always believed that the ultimate goal of developing nuclear weapons is to eliminate them.

8 Model Evaluation and Further Discussion

8.1 Strengths

1 Time series is a series of numbers obtained by arranging the total changes of a variable in time based on the sequence of time. All the information of the historical behavior that produced the sequence is included in the sequence, and the historical information can be used to predict the development trend through the time series model, so the prediction of the number of nuclear weapons in this paper is more accurate.

2 The ARIMA model, which requires only endogenous variables and does not require exogenous variables, is adopted in this paper, so it has a good effect on the long-term prediction of the number of nuclear weapons.

8.2 Weaknesses

1 As the degree of information disclosure has certain limitations on model building, the accuracy of prediction will be affected to some extent.

8.3 Further Discussion

To limit the number of nuclear weapons, we can analyze from the living standards of residents, especially ecological and environmental carrying capacity and other factors. These factors are important constraints on the number of nuclear weapons. We can do further analysis using the Vensim cause-and-effect flow chart to get a limit on the number of nuclear weapons.

Reference

[1] Hargreaves, Ben. Nuclear explosion.[J]. Professional Engineering, 2010.

[2] Thurber C , Trabant C , Haslinger F , et al. Nuclear explosion locations at the Balapan, Kazakhstan, nuclear test site: the effects of high-precision arrival times and three-dimensional structure[J]. Physics of the Earth & Planetary Interiors, 2001, 123(2):283-301.

[3] C, H, Thurber, et al. Accurate locations of nuclear explosions in Balapan, Kazakhstan, 1987 to 1989[J]. GEOPHYSICAL RESEARCH LETTERS, 1993, 20(5):399-402.

[4] Latter R , Karzas W J . Electromagnetic Radiation from a Nuclear Explosion in Space[J]. Phys Rev, 1962, 126(6):40.

[5] Karzas W J , Latter R . Detection of the Electromagnetic Radiation from Nuclear Explosions in Space. 1964.

[6] KeylisBorok, V.I, Malinovskaya, et al. DETECTION OF NUCLEAR EXPLOSIONS[J]. Trudy Instituta Fiziki Zemli Akademiya Nauk Sssr, 1962, vol: no. 20.

[7] Lars, Nielsen, Hans, et al. Origin of upper-mantle seismic scattering – evidence from Russian peaceful nuclear explosion data[J]. Geophysical Journal International, 2003.

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

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

$$M = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{pmatrix}$$

$Part2 : (XYZ\ \ to\ \ Lab)$

$$\begin{aligned} L^* &= 116f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n) - f(Z/Z_n)] \end{aligned} \tag{4}$$

And

$$f(t) = \begin{cases} t^{\frac{1}{3}} & if\ t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & otherwise \end{cases} \tag{5}$$

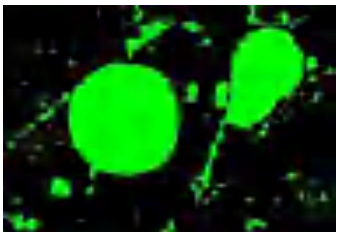
In the two formulas above, L*, a*, and b* represent the values of the three channels in the final Lab color space. X, Y, Z are the values calculated after converting RGB to XYZ, and Xn, Yn, Zn are typically set to the default values of 95.047, 100.0, and 108.883.

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.

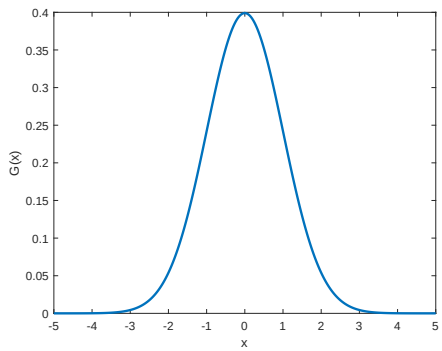
$One-dimensional\ \ Gaussiandistribution$:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

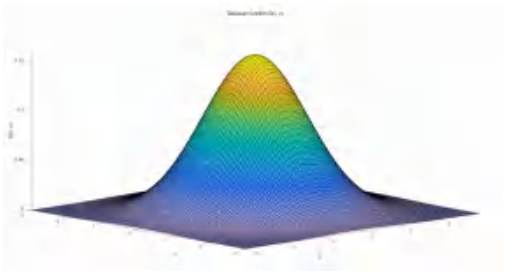
$Two-dimensionalGaussiandistribution$:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Where σ is typically chosen as 1. The following is a schematic diagram of the Gaussian distribution.



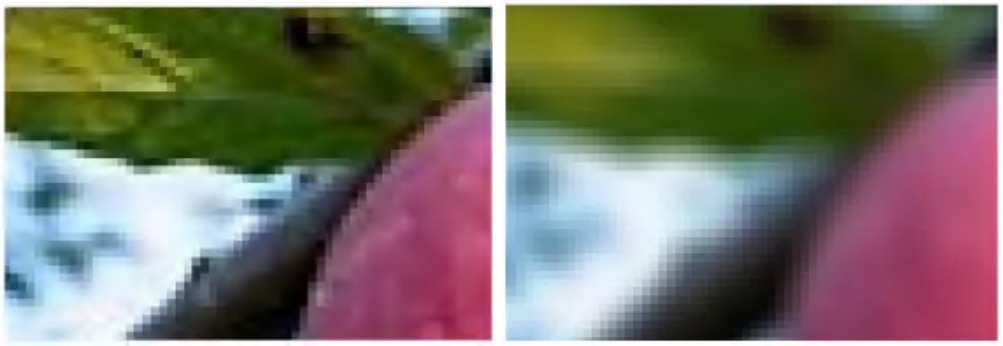
(d) Gaussian function G(x)



(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.

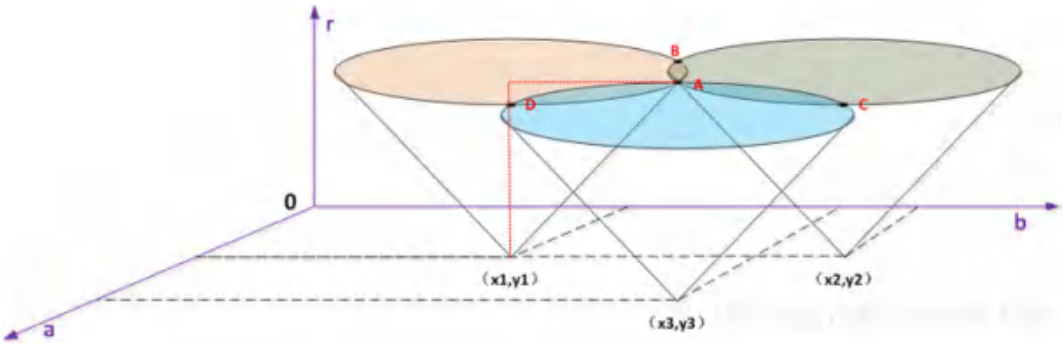


Figure 4 Hough Circle Transform diagram (Image Sourced from the internet).

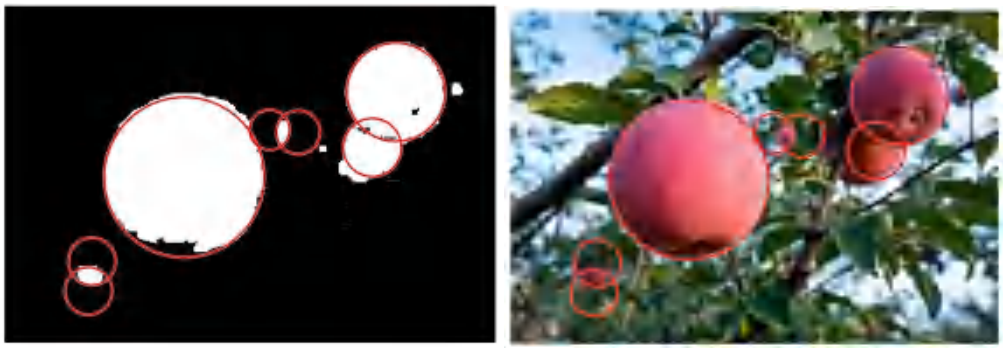
The Hough circle transform can accurately identify some circles that may exist in the image. However, this method involves a massive computational load, resulting in slow code execution. Therefore, we opted for a more efficient and flexible Hough gradient method. This method significantly reduces the 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:

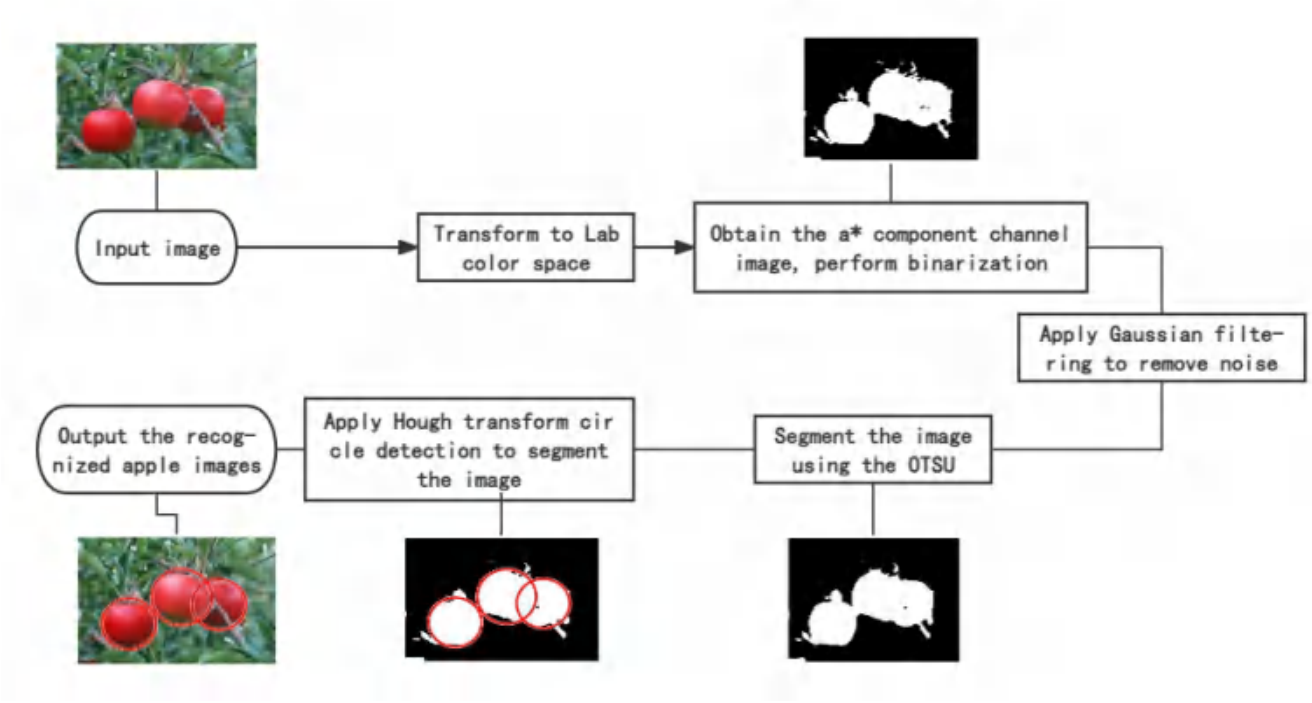


Figure 5 Flowchart of the Apple Recognition Algorithm.

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.

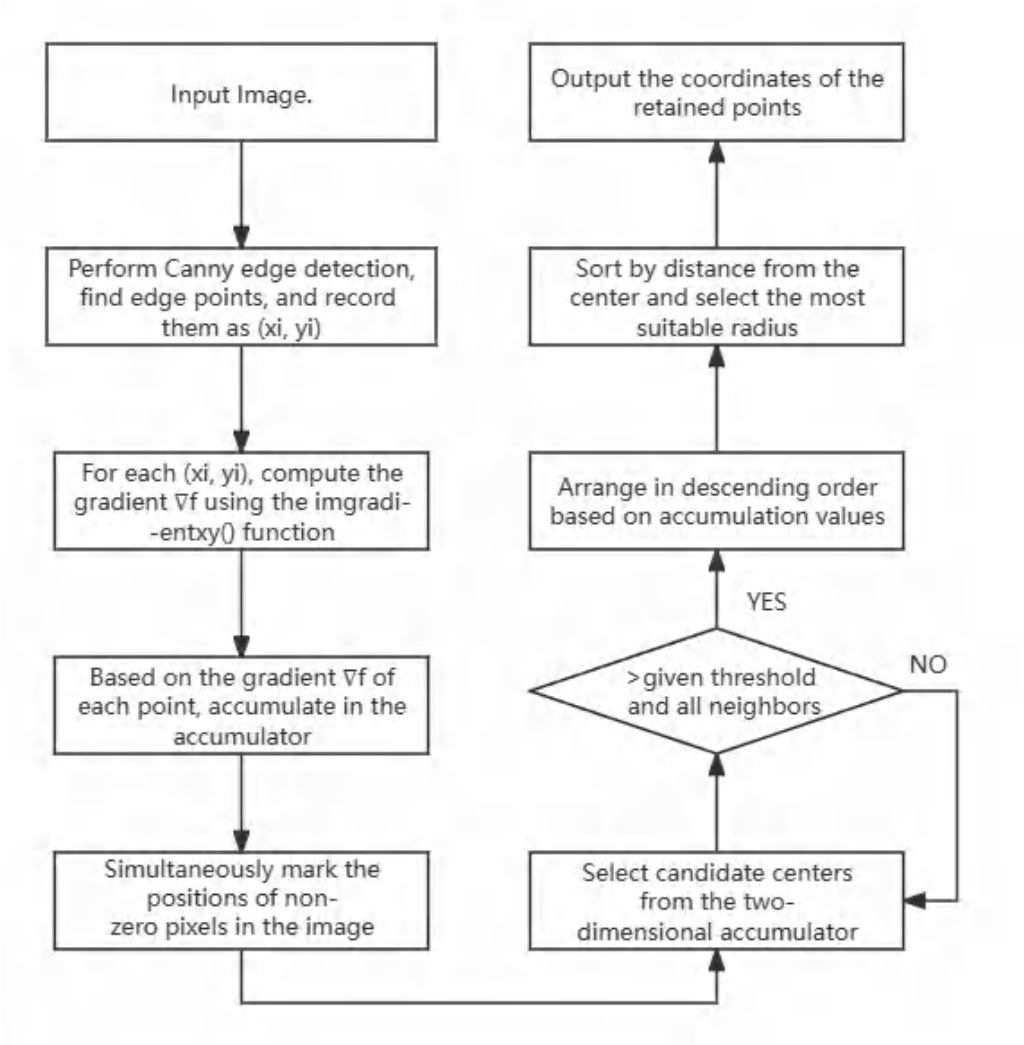


Figure 6 Flowchart of the Optimization Model for Eliminating Duplicate Circles.

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.

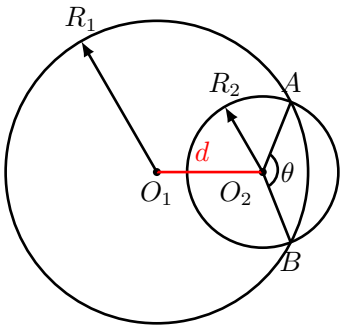


Figure 7 Schematic diagram of two circles overlapping

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

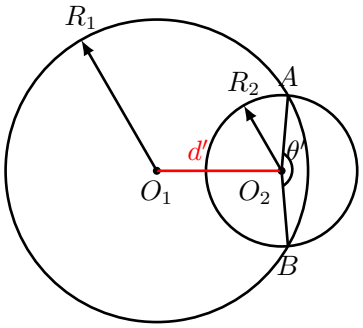


Figure 8 The bigger

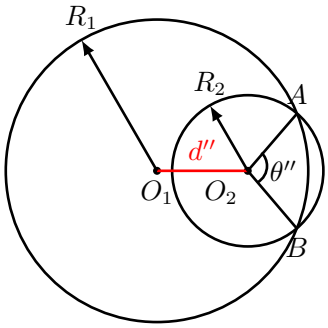


Figure 9 The smaller

also observe that as the value ofγ increases, the distance d between them decreases, and θ continues to decrease as well.

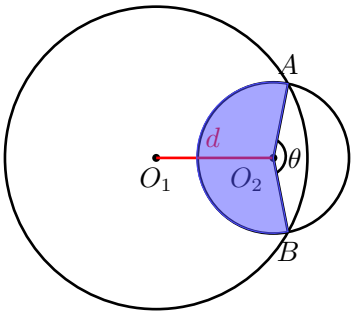


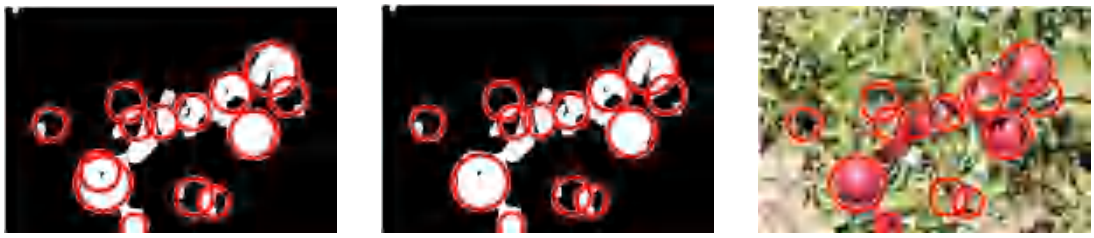
Figure 10 Overlap area i estimation diagram

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:

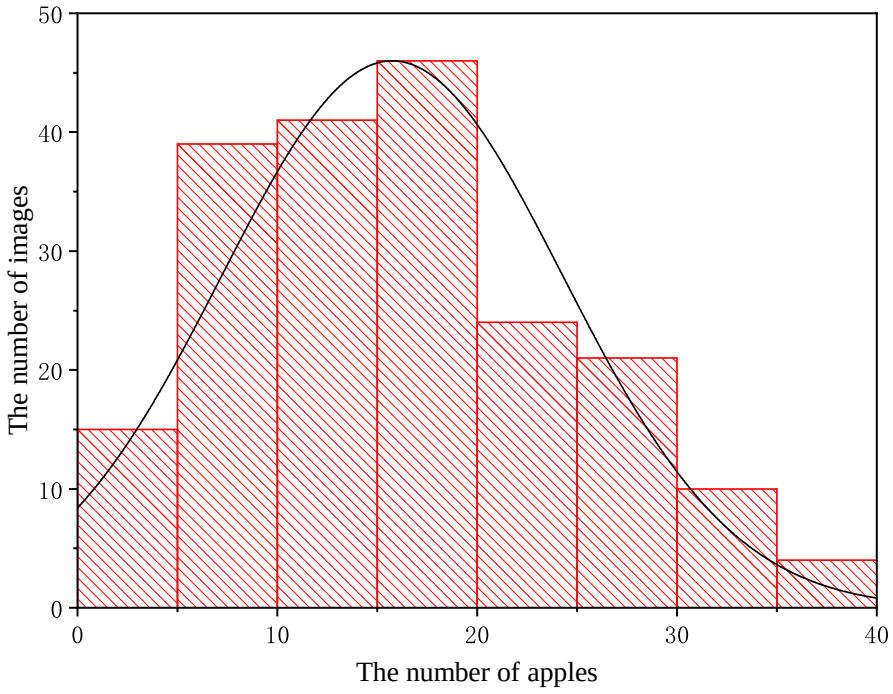


Figure 12 The histogram of the distribution of all apples in Attachment 1

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.

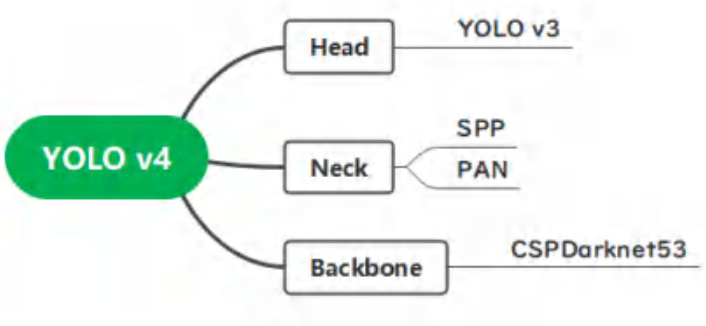


Figure 13 The composition of YOLOv4.

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.

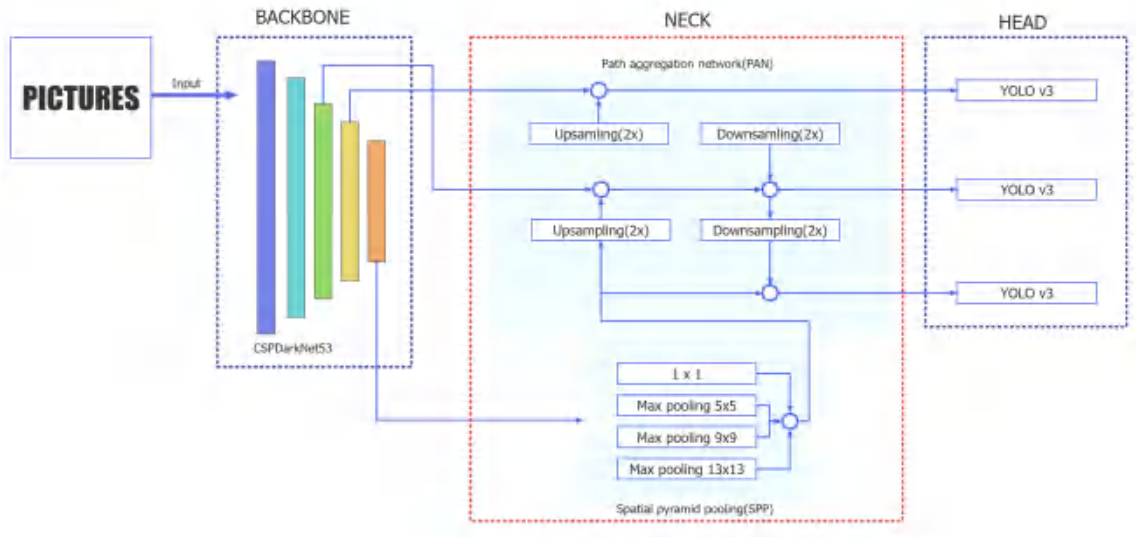


Figure 14 YOLOv4 Network Algorithm Flowchart.

- 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)

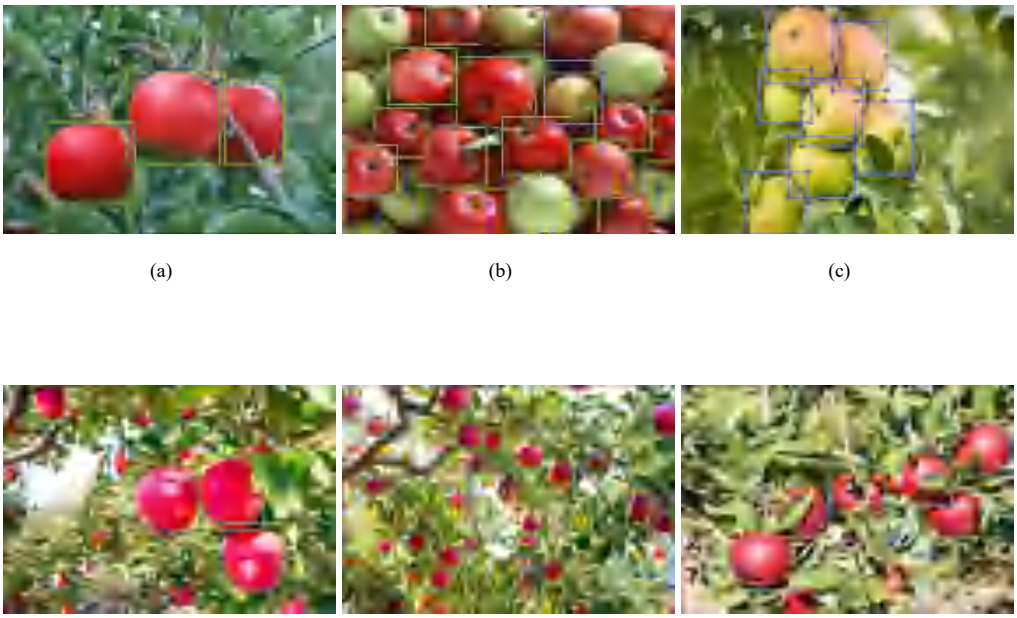
(d) (e) (f)

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.

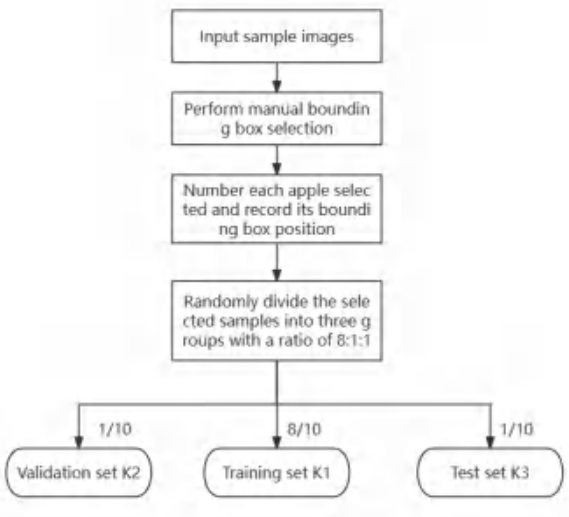


Figure 17 Data preprocessing flowchart of Problem 2

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.

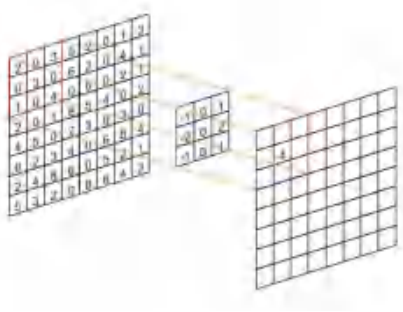


Figure 18 Convolutional Layer Structure Diagram (Image Sourced from the Internet)

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.

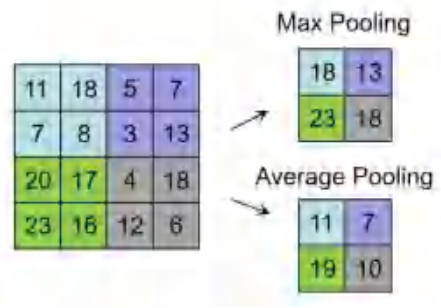


Figure 19 Pooling Layer Structure Diagram (Image Sourced from the Internet)

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.

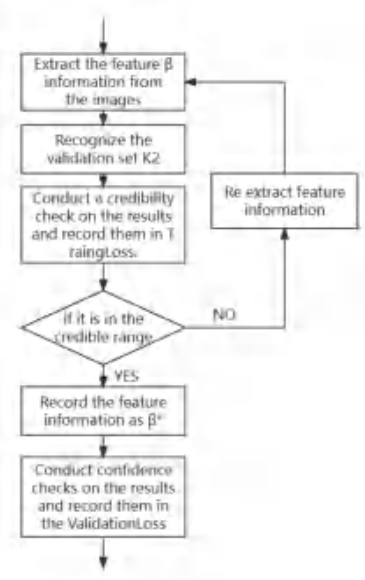


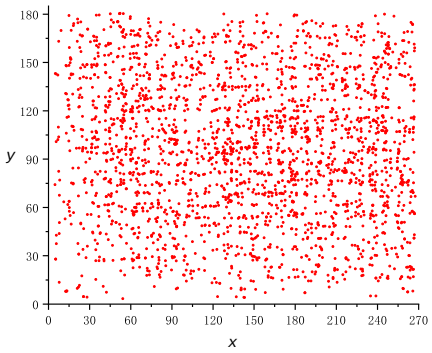
Figure 20 Flowchart of model training(part)

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.

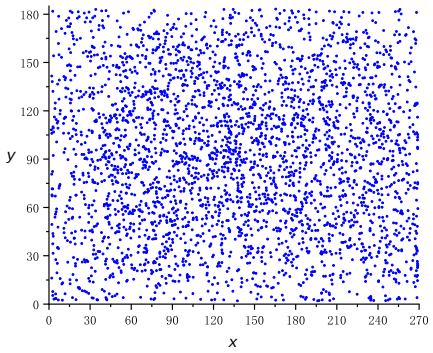
6.1.3 *Comparison of Model Recognition and Sensitivity Analysis*

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.

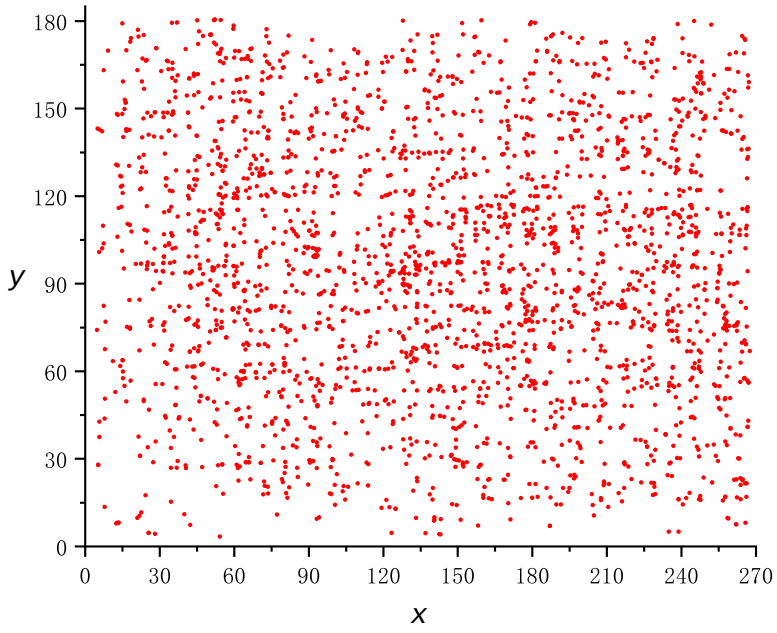


Figure 21 2D Scatter Plot of Geometric Coordinates

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.

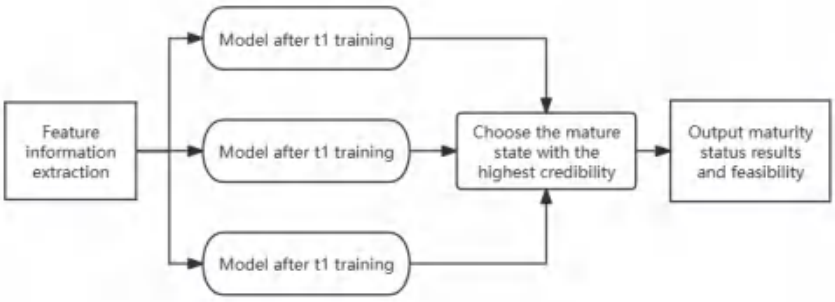


Figure 22 Schematic flow diagram of this process

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:

$$H = \mu p + (1-p)\widetilde{H} \tag{7}$$

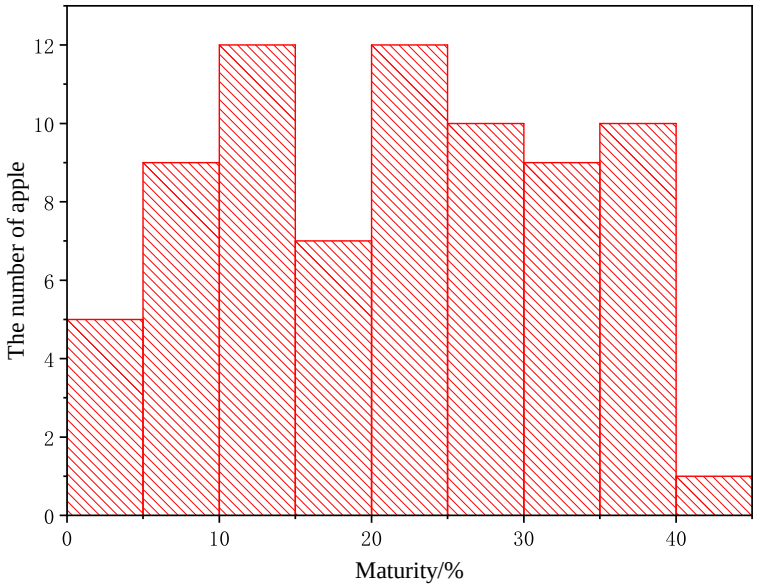
Based on this, we can calculate the maturity H of each apple.

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



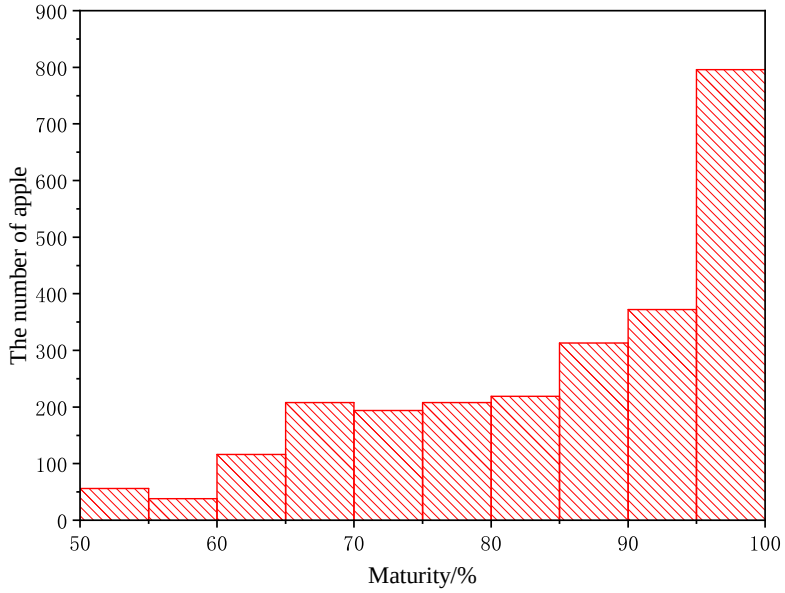


Figure 25 Ripeness distribution histogram of ripe and semi-ripe apples

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

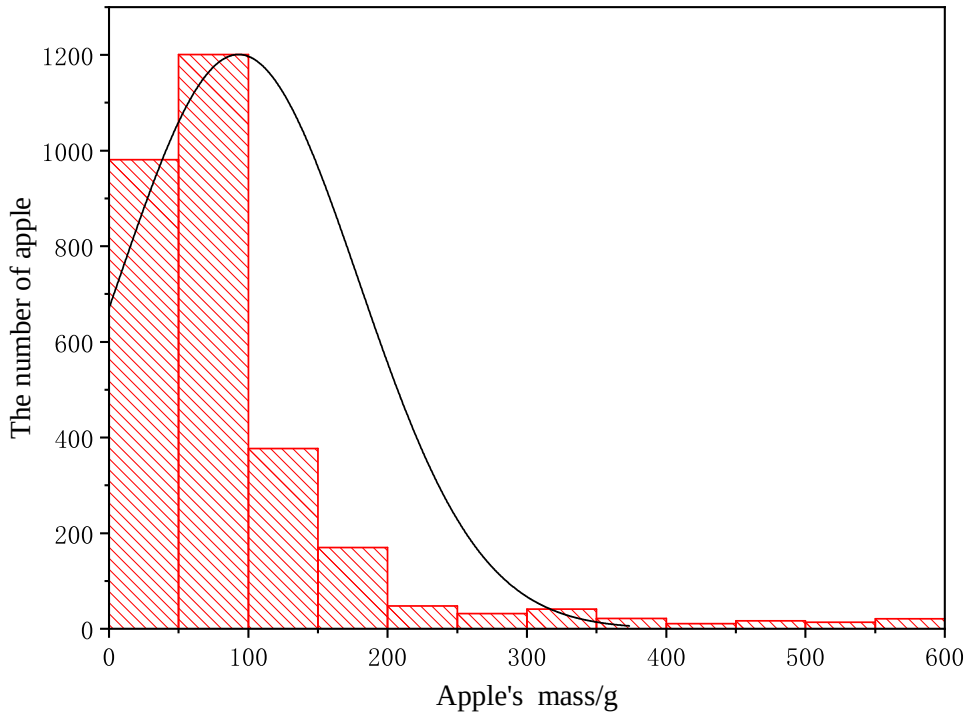


Figure 27 Mass distribution histogram of apples

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

$$\begin{aligned} m &= \frac{4\pi\rho}{3}\sqrt{(\frac{A}{\pi})^3} \\ &= \frac{\pi\rho}{6f^3}\left(\sqrt{\frac{||\alpha\beta^T||W_0}{||\alpha_0\beta_0^T||W_i}}\right)^3 \end{aligned}$$

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

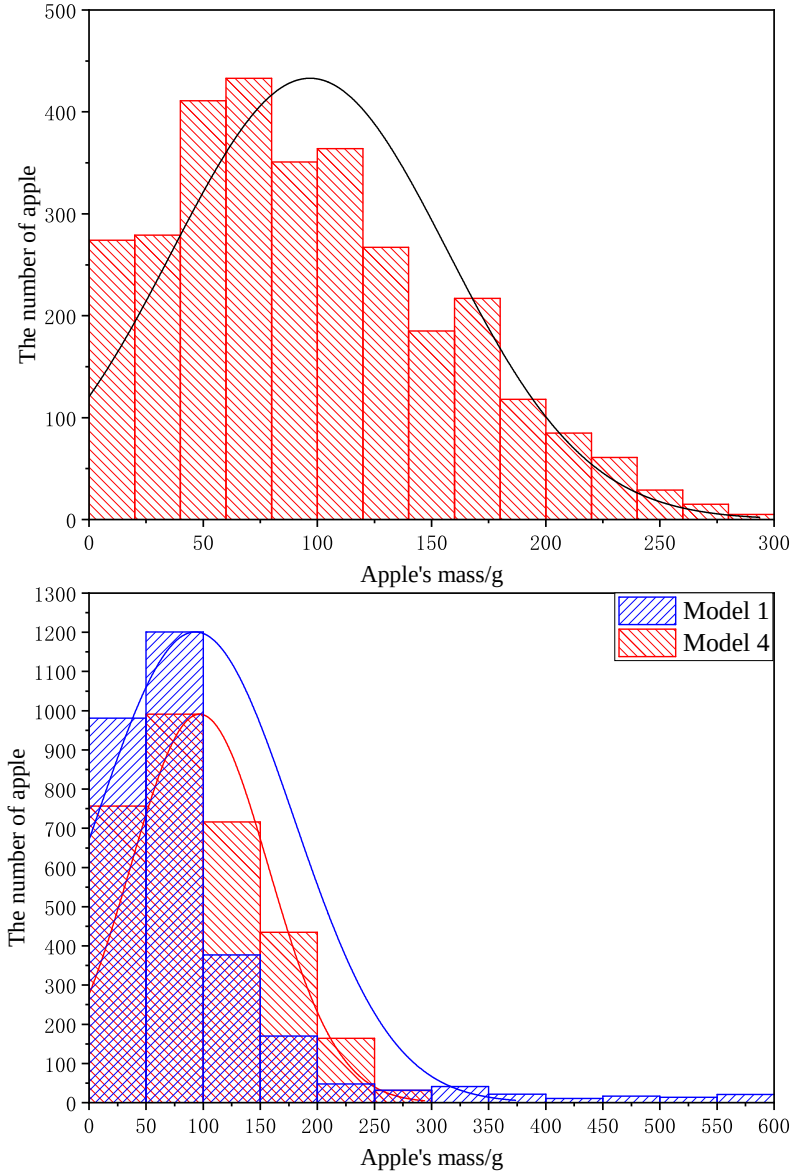


Figure 29 Mass distribution histogram of apples

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:

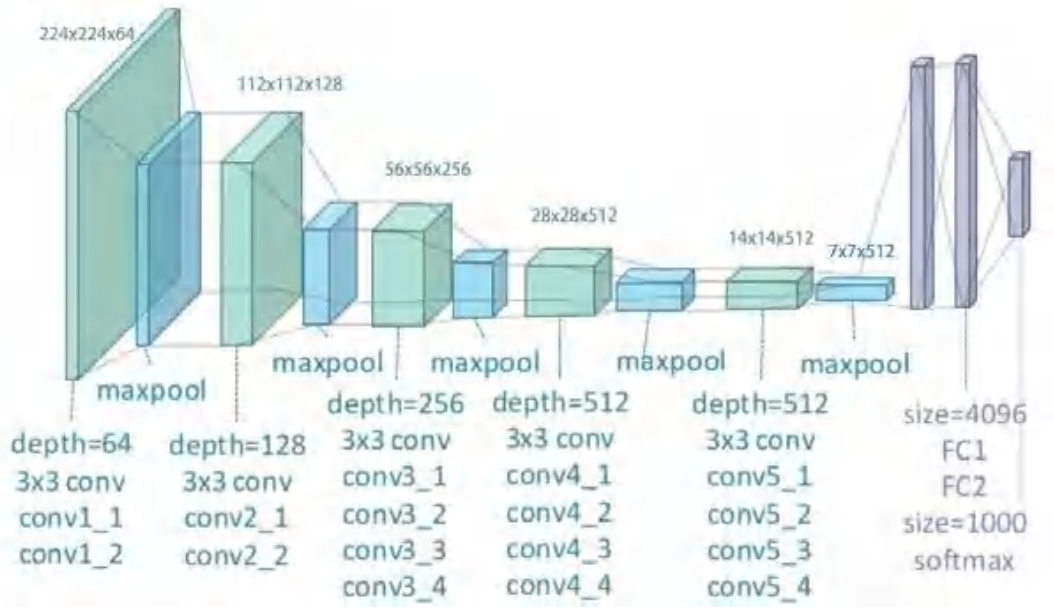


Figure 30 Partial structural diagram of VGG19(Image Sourced from the internet)

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.

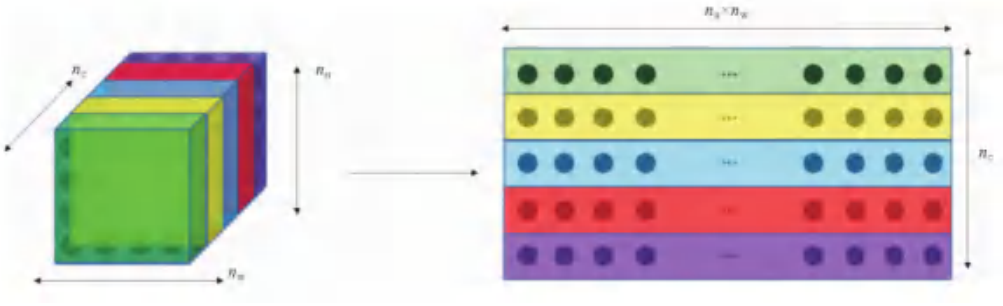


Figure 31 Three-dimensional feature map transformation (Image Sourced from the internet)

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.

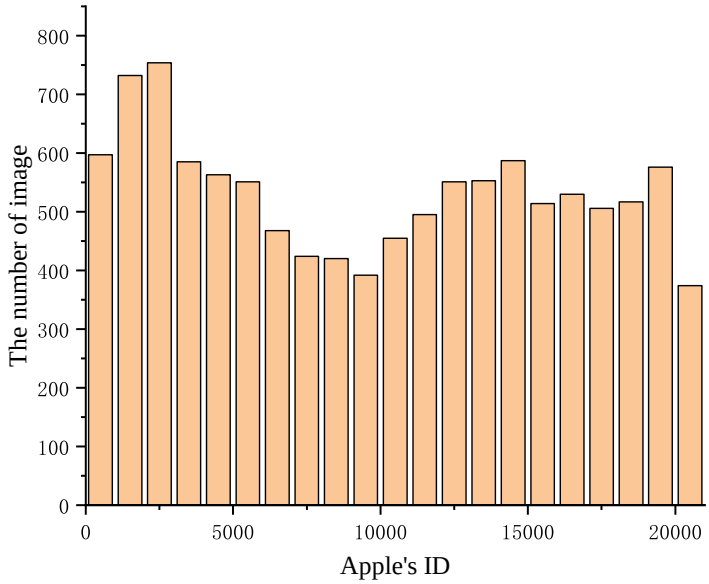
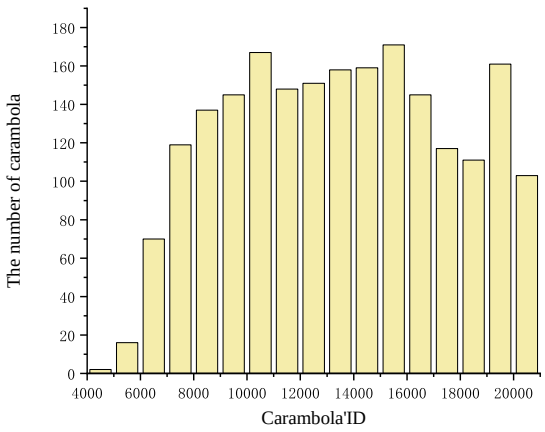
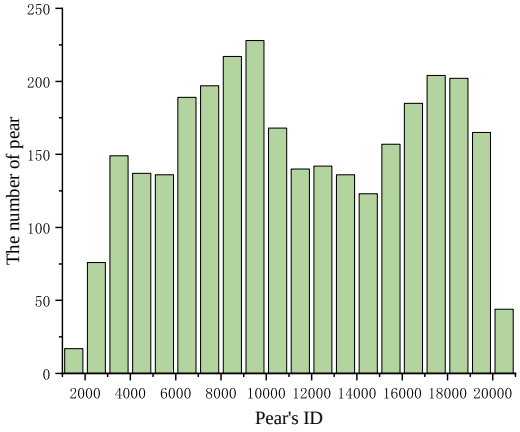


Figure 33 Histogram for all Apple image ID numbers

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:





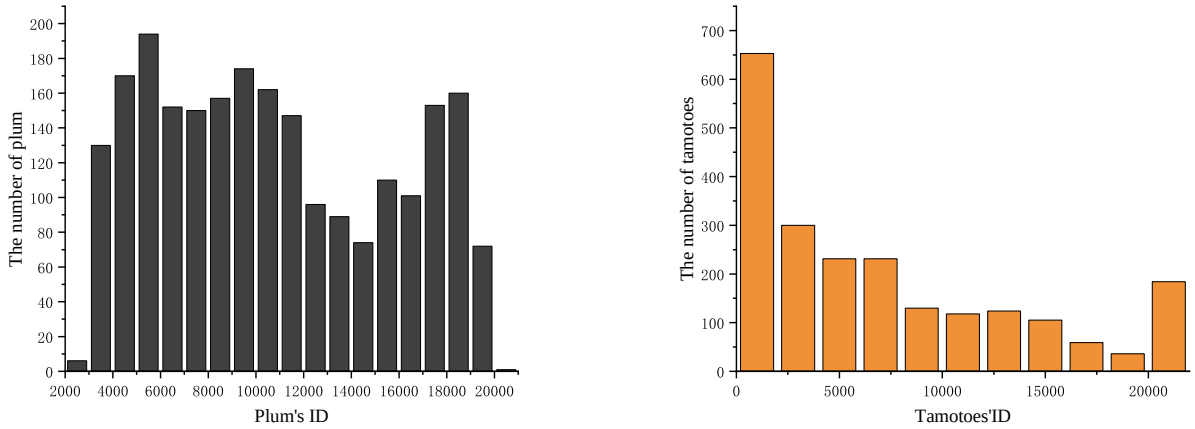


Figure 34 Histograms for other fruits' image ID numbers

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.

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

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).

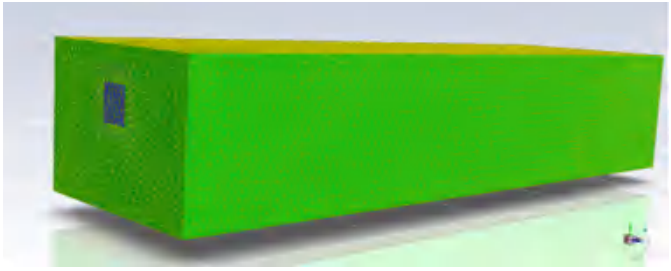


Figure 1 Grids

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

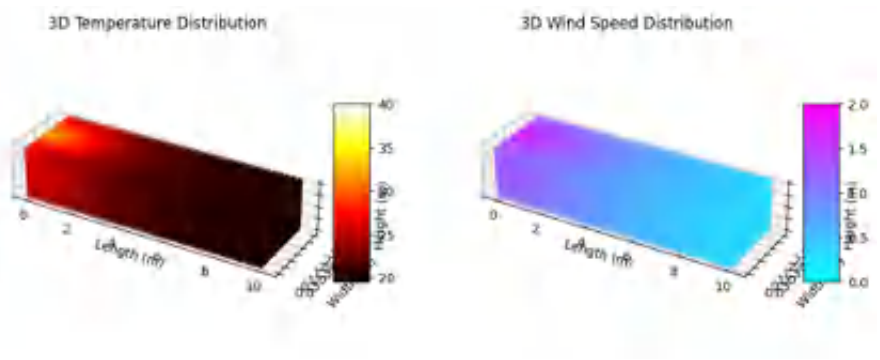


Figure 2 Three-dimensional results of the whole greenhouse in Python

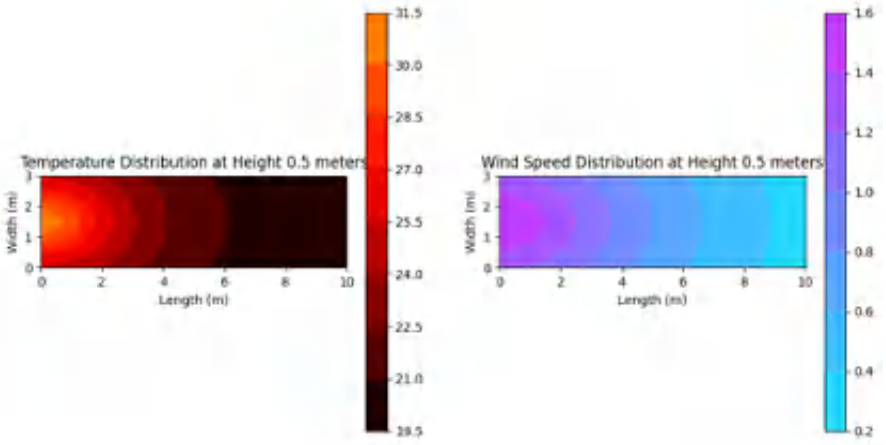


Figure 3 Results of target fields at a height of 0.5 meters in Python

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).

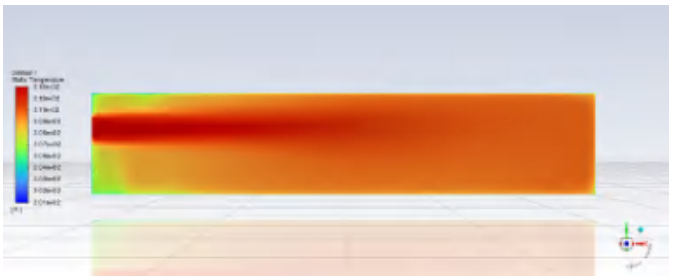


Figure 4 Longitudinal temperature simulated by ANSYS

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

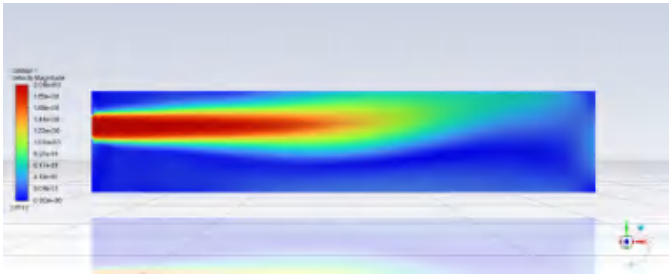


Figure 5 Longitudinal velocity simulated by ANSYS

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).

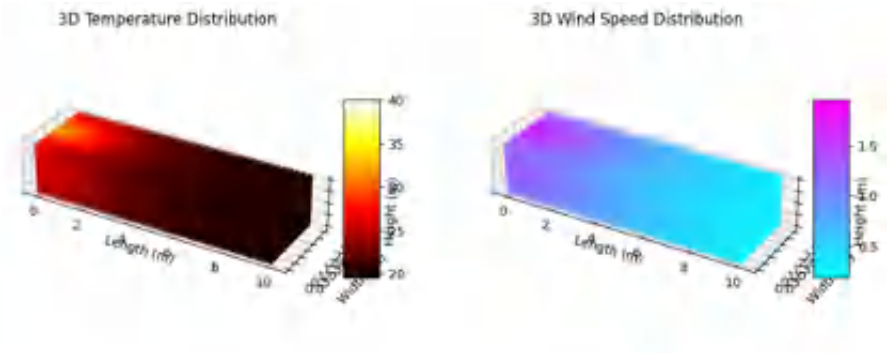


Figure 8 Three-dimensional results of the whole greenhouse with crops in Python

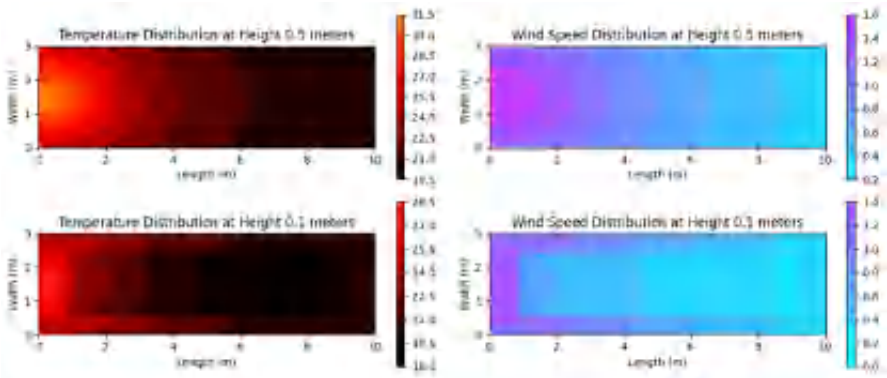


Figure 9 Results of target fields at the height of 0.5 meters and 0.1 meters in Python

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.

$$\varphi = \frac{V_{crops}}{V_t otal} \tag{7}$$

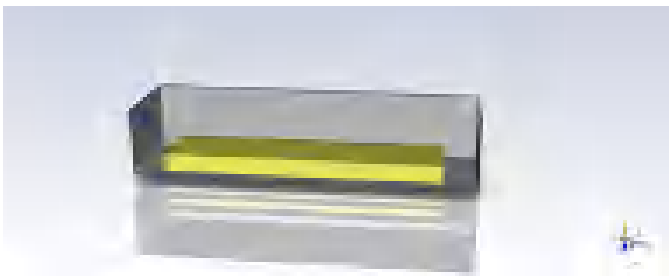


Figure 10 Three-dimensional model for the greenhouse with crops in ANSYS

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.

$$k = \frac{d_p^2 \varphi^3}{K(1-\varphi)^2} \tag{8}$$

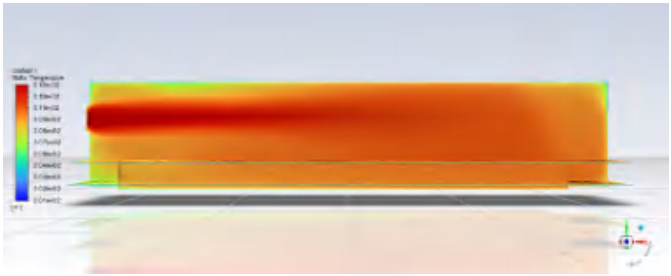
Permeability is a physical quantity that measures the flow ability of a fluid through a porous medium, typically depending on the size, shape, and distribution of pores. For a simplified crop model, we can use the Kozeny-Carman equation to estimate its permeability, where d_p is the diameter of the grain 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.

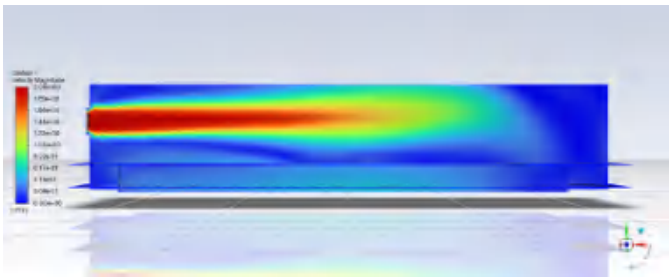
Thus, the longitudinal results given by ANSYS Fluent, showing the variation of temperature and velocity in the direction of the length of the greenhouse was given by Fig. (11), and distribution at the height of 0.5 meters and 0.1 meters simulated by ANSYS were exhibited in Fig. (12) and Fig. (13).

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,

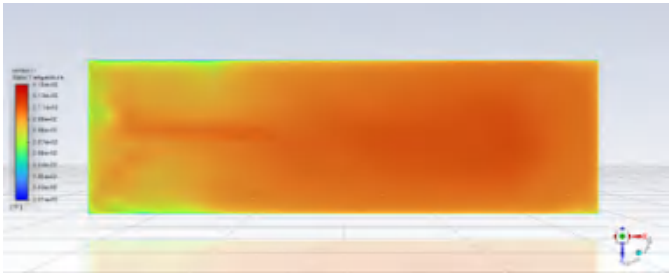


(a) Temperature fields



(b) Velocity fields

Figure 11 Distribution of target fields in the longitudual direction simulated by ANSYS



(a) Temperature fields

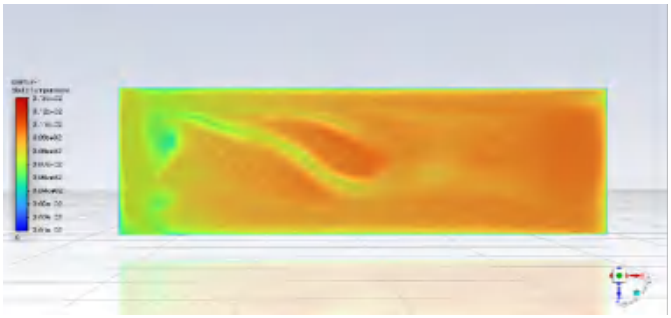
(b) Velocity fields

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.

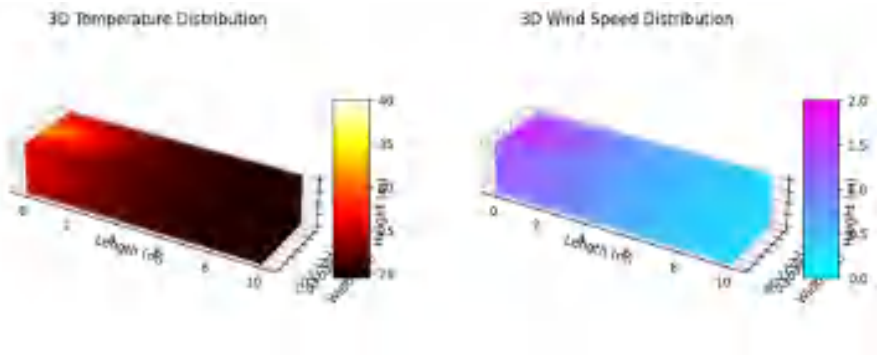
4.3 Question three

4.3.1 *Results and comparisons*

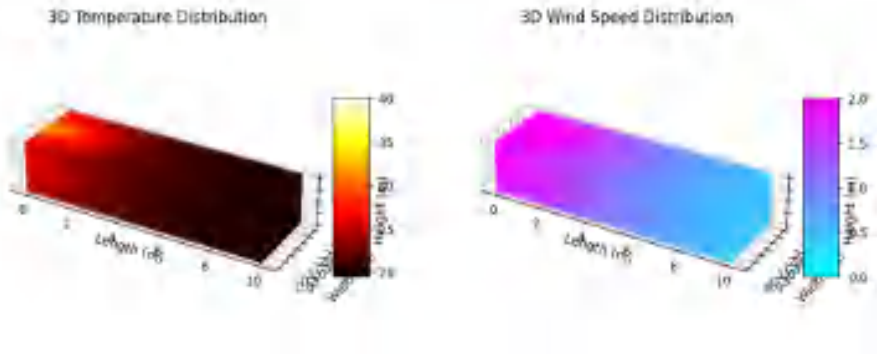
Based on the basic models build in question one and crop models established in question two, here we individually adjusted the velocity of warm air outlet from 2m/s to 3m/s, and lower the position of the greenhouse fan by moving it from 1.3 m to 1 m (see Appendix. 3)to make comparison with results in 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



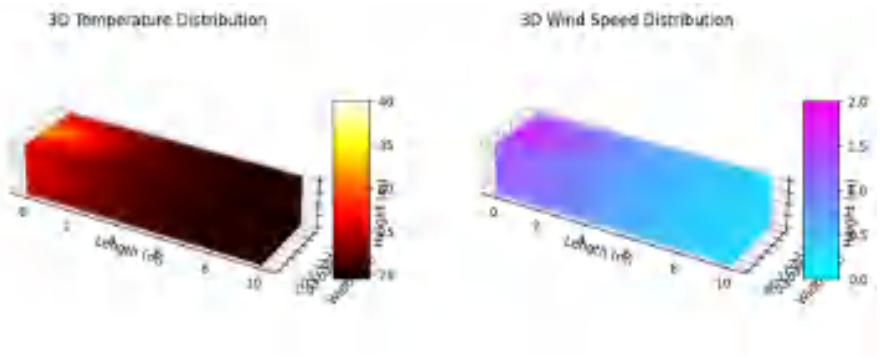
(b) 3m/s in question three

Figure 14 Three-dimensional results with different inlet velocity

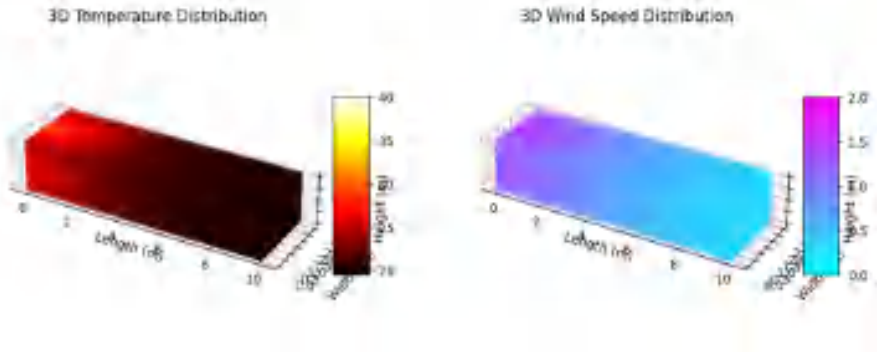
In scenario two, according to our mathematical model, when fans blew in air at the original speed of 2m/s at a lower height of 1m, comparison with results in question 2 were shown in Fig. (15), where in the height direction, we could observe a more uniform distribution, not only in temperature but also 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.

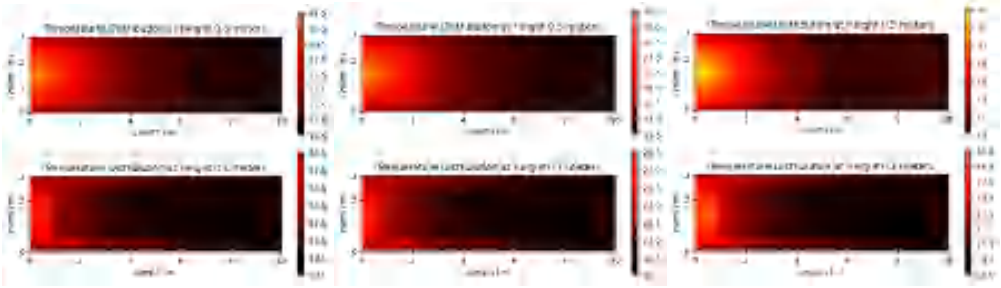


(a) 2m/s in question two



(b) 2m/s in question three

Figure 15 Three-dimensional results with different inlet position

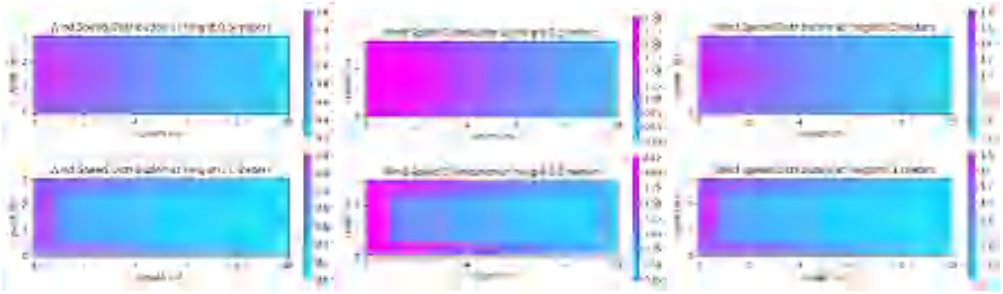


(a) question two (b) 3m/s in question three (c) 1m in question three

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

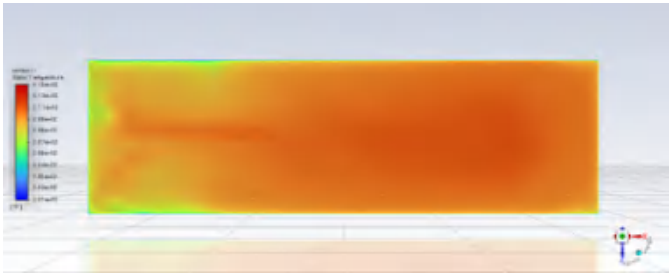
temperature distribution in the whole cuboid, which could also be concluded from the velocity field.

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.

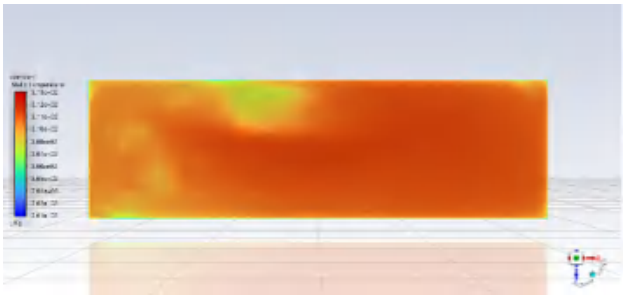
Overall, by applying the indicator, we showed that although the prediction results of the two models differ in some regions, they still show good agreement in most areas. This finding highlights the applicability and flexibility of the our model in simulating changes in complex environments. The reference 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

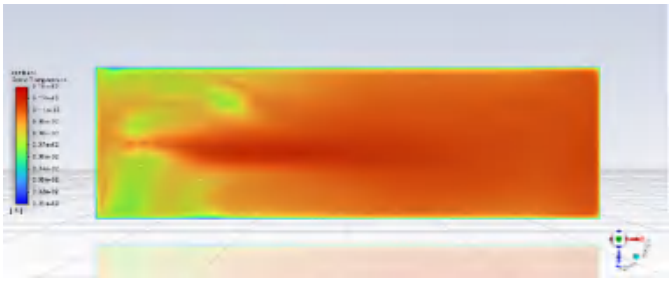
Based on the deficiency above, when it comes to optimizing design of glass greenhouse, among many of them, two main factors could be immediately came up with, the selection of crops and the optimization of the amount, location, speed and inlet temperature of fans.



(a) question two



(b) 3m/s in question three



(c) 1m in question three

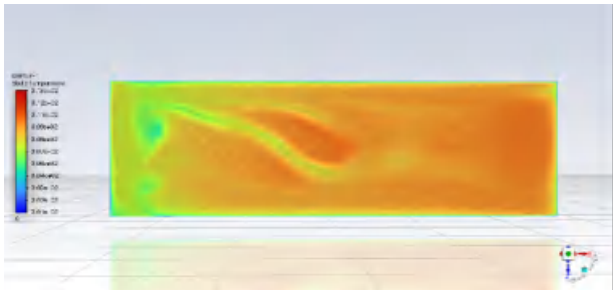
Figure 18 Temperature fields in different scenarios at the height of 0.5m

4.4.1 *Select Crops*

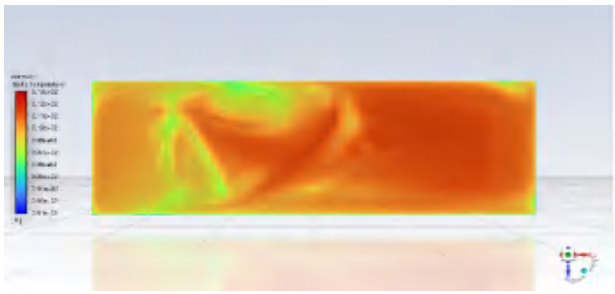
Our goal is to select crops that could fit high temperature and low velocity. According to our simulation results, for the first three conditions:

- 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.

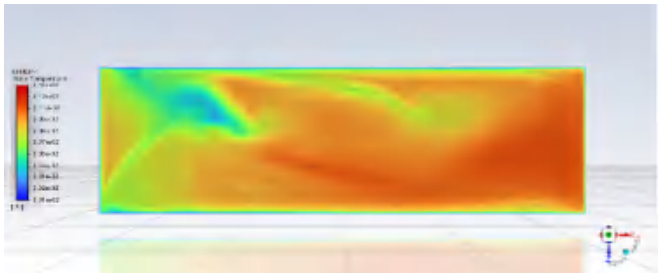
]



(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

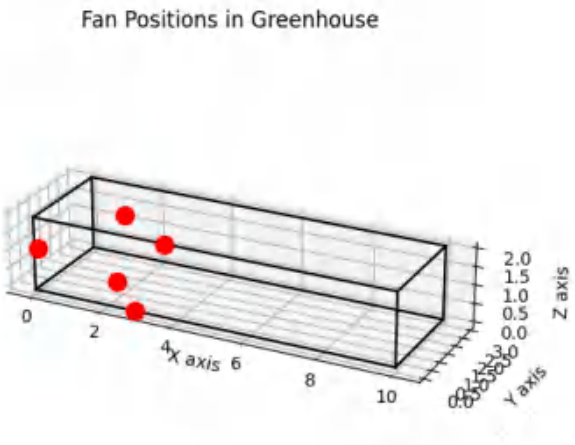


Figure 20 Locations of the five fans

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.

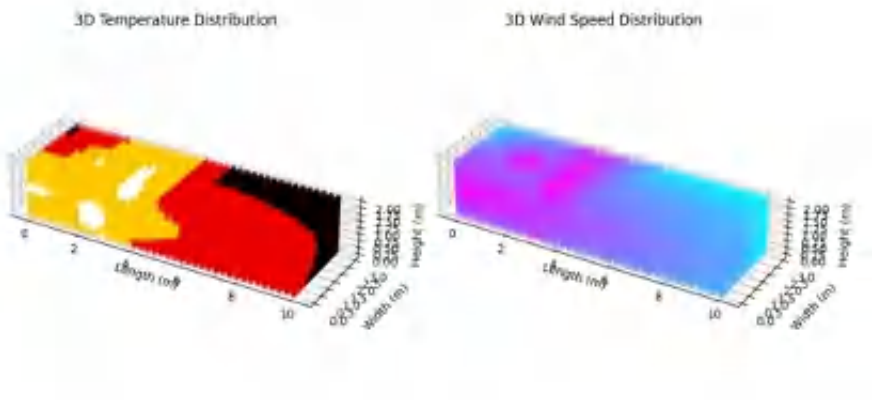


Figure 21 Three dimensional fields of the five-fans scheme

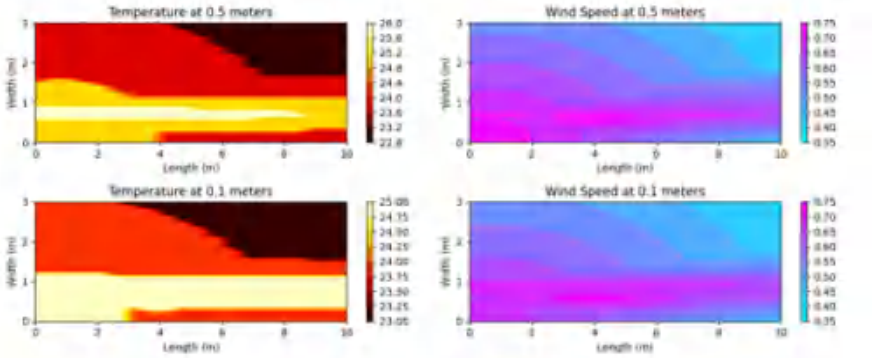


Figure 22 Target fields at two cross-sections of the five-fans scheme

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

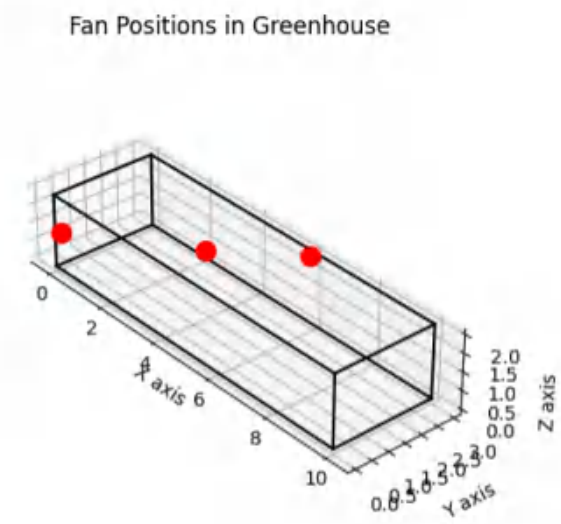
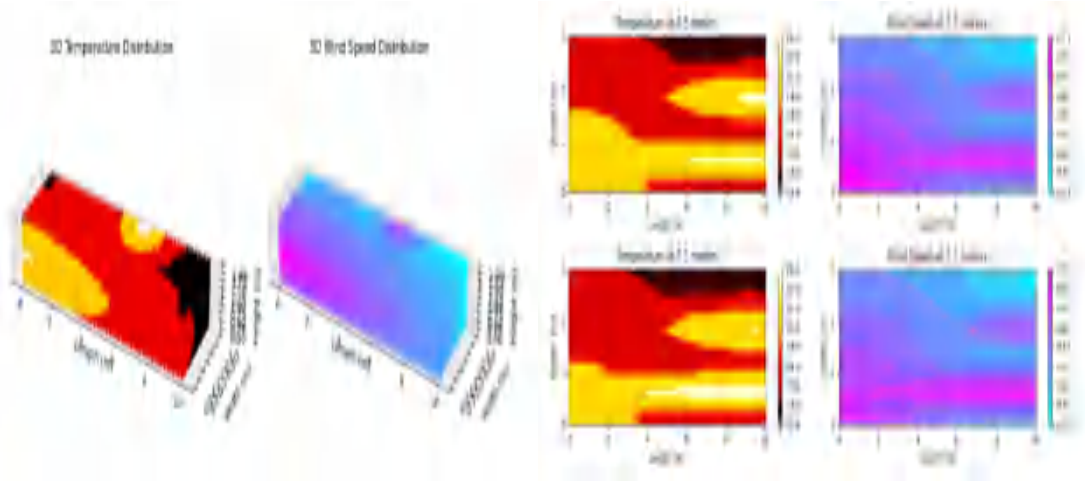


Figure 23 Locations of the three fans



(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.

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

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:

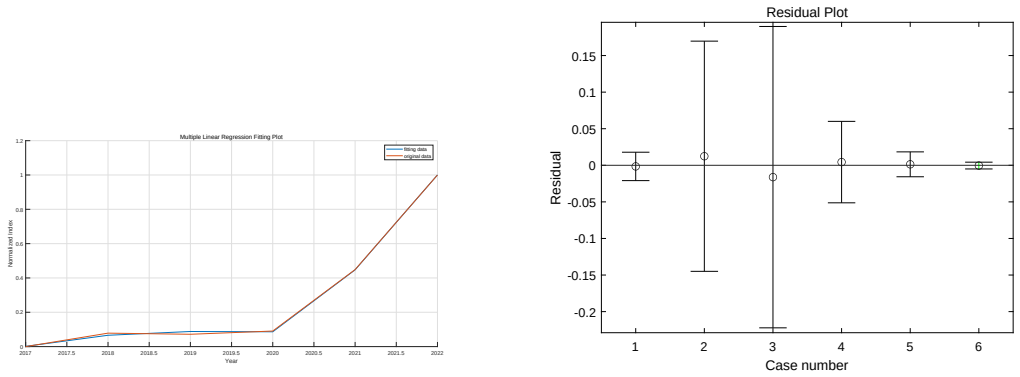


Figure 1 Multiple linear regression fitting plot and residual plot

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.

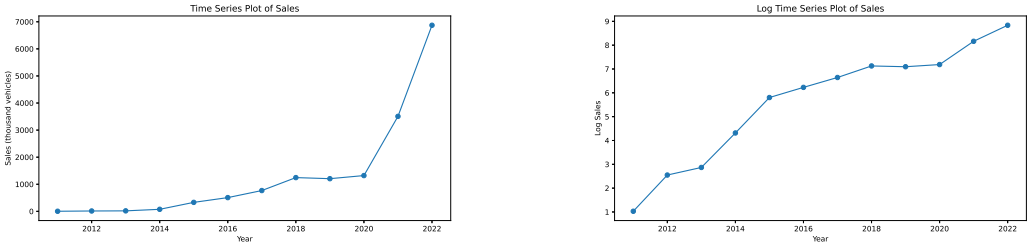


Figure 2 Sales time series plot and the plot with Log Transformation

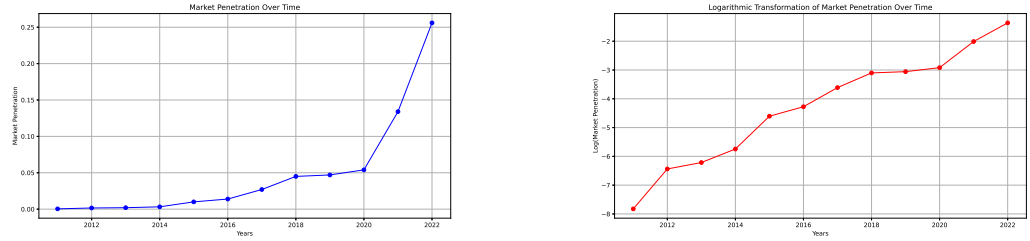


Figure 3 Market penetration rate time series plot and the plot with Log Transformation

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.

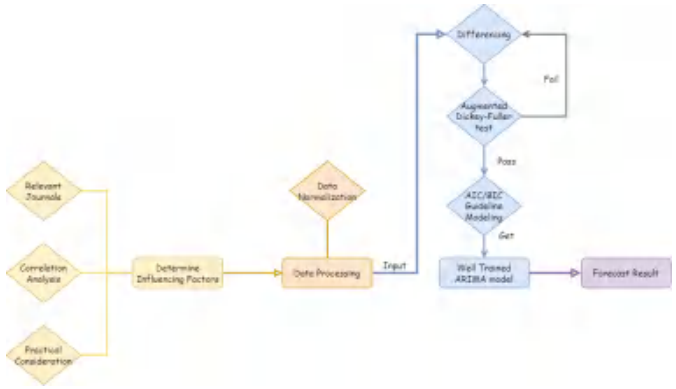


Figure 4 Step flow chart

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.

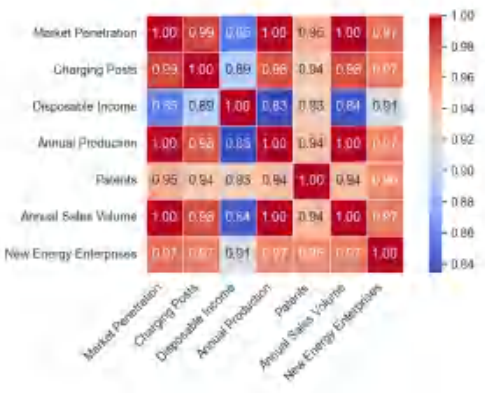


Figure 5 Correlation matrix between feature variables

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:

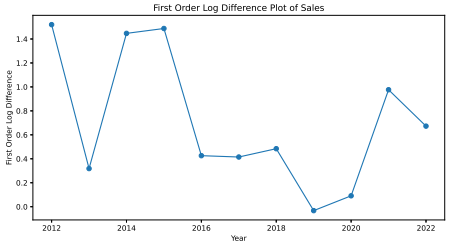
$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_2 B^p)(1 - B)^d X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t \tag{6}$$

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.

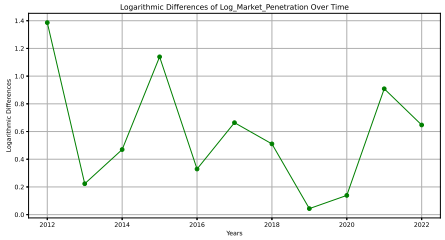
In order to build an ARIMA model, it is necessary to select appropriate parameters p, q, d.

⋆ **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.

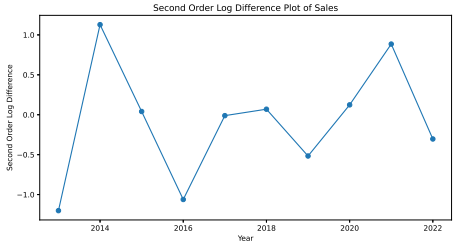


Figure 7 Second order difference of logarithmic time series of sales volume

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.

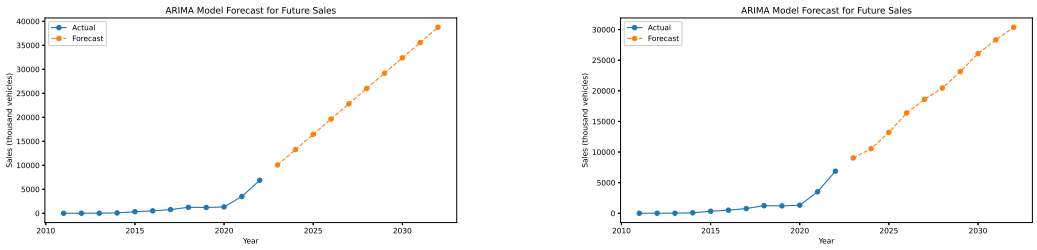


Figure 8 Sales forecast result of p=0,q=1(left) and p=3,q=1(right)

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.

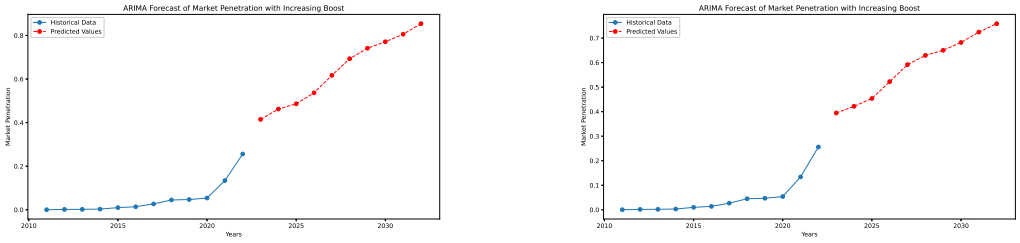


Figure 9 Market penetration rate forecast result of p=3,q=0(left) and p=3,q=1(right)

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:

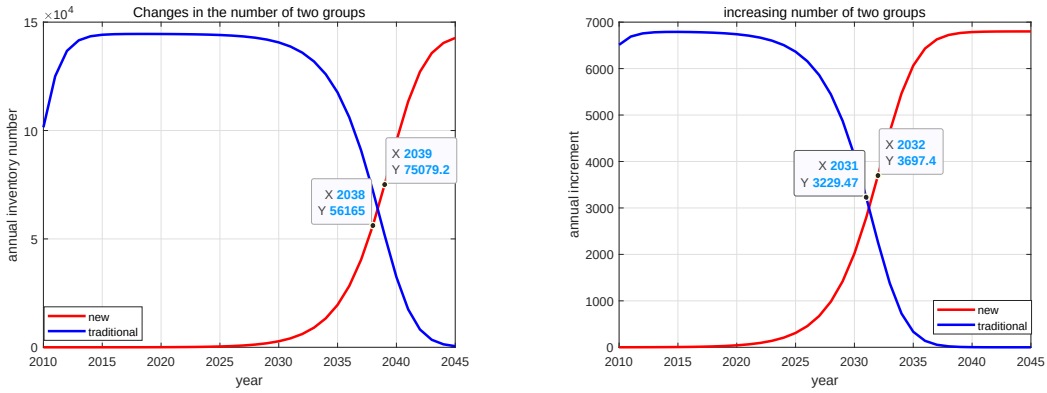


Figure 10 Competition results of ownership(left) and sales(right)

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.

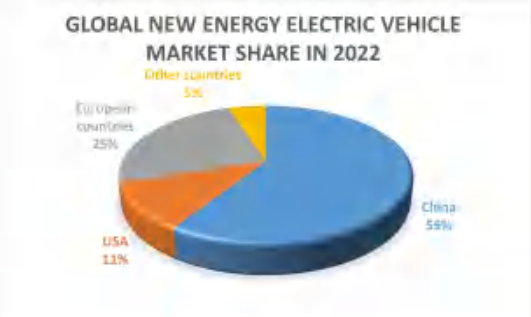


Figure 11 Global new energy electric vehicle market share in 2022

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:

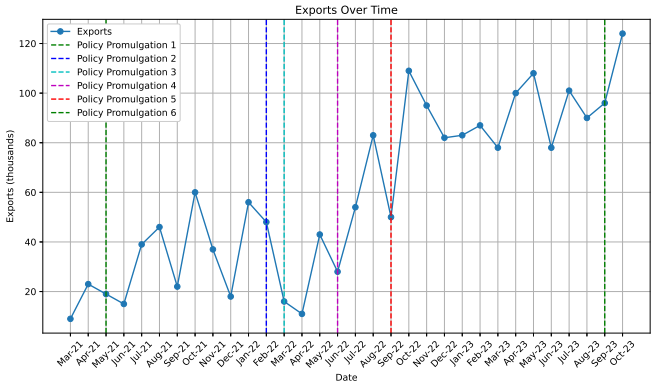


Figure 12 Export volume-policy diagram

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

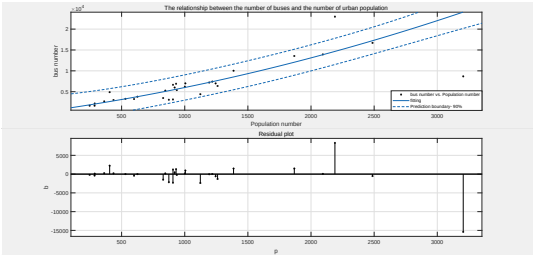
According to authoritative industry reports in the automotive sector and publicly available national information, we have obtained the average electrification rates for buses and cars (M_b,M_c). Additionally, we've acquired the energy consumption equivalents for electric buses and cars (a_1, a_2), where 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

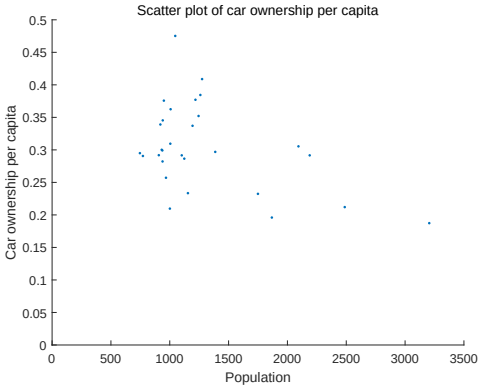
To evaluate the impact of automobile electrification on the ecological environment in a city of one million inhabitants, it requires fitting into two nested models. Given the composition of vehicles in the city, for model simplification, we assume the city's automobiles do not include trucks. We use 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:

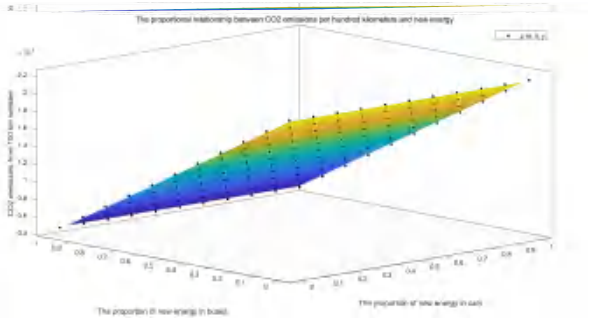


Figure 14 The Fitted Model for Carbon Emissions and Vehicle Numbers

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.

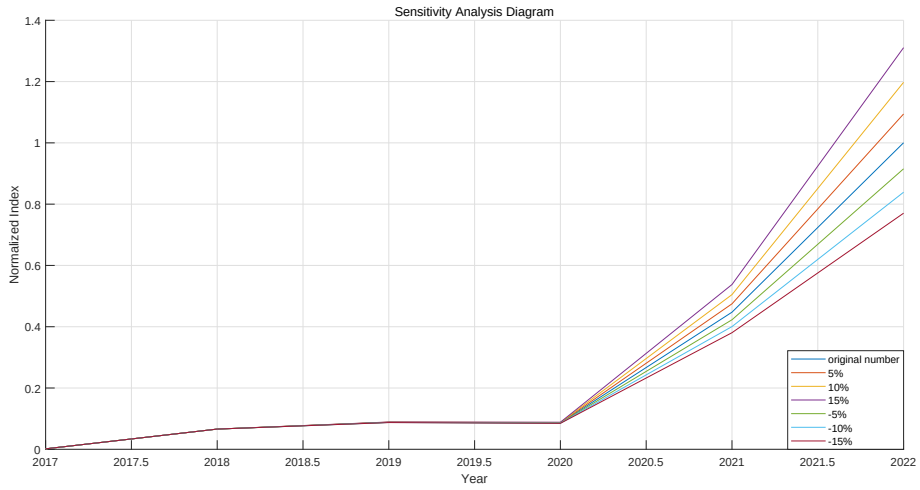


Figure 15 1

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

2024年第十四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。

2023年第十三届亚太地区大学生数学建模竞赛共有9700支队伍969所高校2万7千多名学生报名参赛。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；德国的亚琛工业大学、 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；日本的東北大学；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。

>> 微信二维码 <<

联系我们
电话：156-0048-3352（林老师）