

## 第九届湖南省研究生数学建模竞赛承诺书

我们仔细阅读了湖南省高校研究生数学建模竞赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们完全清楚，在竞赛中必须合法合规地使用文献资料和软件工具，不能有任何侵犯知识产权的行为。否则我们将失去评奖资格，并可能受到严肃处理。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们授权湖南省研究生数学建模竞赛组委会，可将我们的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）。

所属学校和学院（请填写完整的全名）：国防科技大学系统工程学院

参赛队员（打印后签名）：  
1. 王彩琳  
2. 王子辰  
3. 曾得琪

指导教师或指导教师组负责人（打印后签名）：

程昆

日期：2024年 7 月 1 日

---

（请勿改动此页内容和格式。以上内容请仔细核对，如填写错误，论文可能被取消评奖资格。）

# 第九届湖南省研究生数学建模竞赛

## 题目：多视角下国家粮食安全风险建模与分析

本研究针对中国粮食安全面临的气候变化、耕地减少和种子短缺等关键问题,通过收集和分析多种数据,系统探讨了各因素对粮食生产及供应安全的影响及其解决方法。首先,基于历史气温变化和粮食产量数据,采用数学建模分析了水稻、小麦和玉米单产随气温变化的关系,预测了未来 10 年中国粮食净进口量的走势,揭示了气候变化对粮食安全的微观机理。其次,结合全国土地调查数据,采用定量分析方法描述了近 20 年来中国耕地面积和质量的变化情况,分析了政策在遏制耕地“非农化”“非粮化”方面的效果。再者,通过查询相关文献和数据,分析了近 20 年来中国粮食种子自给率的变化趋势及其背后的市场结构调整原因与理论逻辑。最后,综合分析了气候变化、耕地减少和种子短缺对中国粮食安全的综合影响及其演变趋势,为制定粮食安全政策提供了深入洞察和科学依据。

针对问题一,选取温度,降水量,日照等因素,探究其对稻谷,小麦,玉米等粮食作物产量的影响。首先对各因子和三种作物的产量的相关性进行分析,简单探究因子与输出之间的影响机制。在模型准备阶段,对各数据进行趋势检验,判断数据趋势。接着使用线性回归模型进行建模,该模型可解释性强,但精度相对较低。因此在分析三种粮食作物产量相关性的基础上,使用考虑产量相关性的多输出高斯过程模型,并引入 Shapley 方法对高斯过程插值模型进行解释。

针对问题二,以维持耕地面积增长的粮食大省黑龙江省为例,描述耕地面积变化情况。提取包括人口因素,经济因素,农村居民因素,政策因素以及其他因素共 18 个影响因子,建立影响因子与耕地面积变化的模型。为避免因子之间相关性对建模结果的影响,在主成分分析的基础上,使用主成分建立线性回归模型和非线性回归模型,选择精度较高的非线性回归模型,根据因子系数判断因子对耕地面积的影响趋势。

针对问题三,从种子数量和质量的视角,以农作物的播种面积和产量的作为度量指标,对 2000-2020 年我国粮食种子的自给率进行测算,收集调研了自 2000 年以来我国的粮食种子相关制度变革,并交叉对比了国内法律法规对于粮食种业的相关影响,以及针对种业即时状况法律法规的调整。同时分析了进出口量与国际产业链因素如贸易、供给关系、需求等对于国际市场的影响,详细对比了近 20 年中国与国际粮食价格的变化趋势,简要分析了国际的种子行业市场形式,对影响“种子按斤称到论粒卖”的内在机理进行相关分析。

针对问题四,分别从气候变化、耕地减少及种子短缺三方面对我国粮食安全影响进行剖析。根据气候-粮食产量效应模型分析气候变化对粮食安全影响;根据基于主成分分析的回归模型分析耕地减少对粮食安全影响;根据种子相关数据及国际种业环境分析种子短缺对粮食安全影响。结合前三问收集数据及所构建模型进行因果关系分析,进而根据我国粮食安全政策分析我国粮食安全形势的演变趋势。

最后,我们对提出的模型进行全面的评价:本文的模型贴合实际,能合理解决提出的问题,具有实用性强,算法效率高等特点。

**关键词: 粮食安全、趋势检验、多输出高斯过程、主成分分析、非线性回归、AHP**

# 目录

多视角下国家粮食安全风险建模与分析.....	I
摘要.....	I
1 问题综述.....	1
1.1 问题背景.....	1
1.2 问题提出.....	1
1.3 资料条件.....	2
2 模型假设与符号说明.....	3
2.1 模型基本假设.....	3
2.2 符号说明.....	3
3 问题一分析与模型建立.....	3
3.1 问题分析.....	3
3.2 数据分析.....	4
3.3 模型准备.....	11
3.4 基于 OLS 分析的气候-粮食产量效应模型.....	13
3.5 考虑产量相关性的气候-粮食产量效应模型.....	14
3.6 粮食净进口量影响模型构建及预测.....	18
4 问题二分析与模型建立.....	23
4.1 问题分析及数据来源.....	23
4.2 耕地面积变化定量描述及驱动机制研究.....	27
4.3 耕地质量变化定量描述.....	37
5 问题三的分析与模型构建.....	39
5.1 问题分析.....	39
5.2 模型构建.....	39
6 我国粮食安全形势的演变趋势分析.....	47
6.1 气候变化对粮食安全影响因果关系分析.....	47
6.2 耕地减少对粮食安全影响因果关系分析.....	48
6.3 种子短缺对粮食安全影响因果分析.....	49
7 模型评价与推广.....	49
7.1 模型的优点.....	49
7.2 模型的不足.....	49
7.3 模型的推广.....	49
参考文献.....	50
附    录.....	52
附录 I: 主要程序/关键代码.....	52

# 1 问题综述

## 1.1 问题背景

“五谷者,万民之命,国之重宝。”我国是一个人口大国,保障国家粮食安全是一个永恒的课题,也是国计民生的头等大事。中国人的饭碗任何时候都要牢牢端在自己手中,饭碗主要装中国粮。因此,分析粮食安全的影响因素,在粮食生产的多个环节,把握粮食安全,是作为农业大国的中国需要重点把握的问题。

粮食的生产首先在耕地,耕地面积和耕地质量是粮食生产的基础,只有维持耕地面积,保持耕地质量,才能保证粮食生产量。其次是对粮食种子的把握,针对国外种业巨头研发成本投入多,种子质量好的现状,国内如何制定相关政策,发展相关科技,在错综复杂的国际种业市场中良好地维护国家来给你是安全。此外,在粮食生长过程中,温度,湿度,日照等因素会对粮食的生长带来较大影响,针对不同粮食作物的特性,以及不同地域的区域特征,这种影响也会略有不同,影响机制需要分别进行分析。

## 1.2 问题提出

**气候变化对粮食产量的影响分析:** 我们将收集过去几十年的气温数据,并运用数学建模技术,研究水稻、小麦、玉米等主要粮食作物的产量如何随着气温变化而变化。具体而言,我们将分析每升高  $0.1^{\circ}\text{C}$  气温对这些作物产量的影响,以揭示气候变化背景下的微观机理。进一步,我们将预测未来 10 年中国粮食净进口量的趋势,以评估国家粮食供给的稳定性。

**耕地变化及政策效应分析:** 基于全国土地调查数据和相关文献资料,我们将选择特定国家、省份或地区为研究对象,定量描述近 20 年来耕地面积和耕地质量的变化趋势。我们将深入分析耕地“非农化”“非粮化”问题的背景原因,并评估国家政策在遏制这些趋势方面的效果。

**种子供给与质量变化分析:** 我们将调查并分析近 20 年来我国粮食种子的自给率及其质量变化情况。具体包括种子的退化和绝育问题,以及全球产业链布局对我国种子供给的影响。我们将从数量和质量的角,探讨“种子按斤称到论粒卖”销售模式变化的内在原因和理论逻辑。

**粮食安全形势的因果关系分析:** 最终,我们将结合上述研究成果,揭示气候变化、耕地减少和种子短缺对我国粮食安全的影响因果关系。通过深入分析这些因素的相互作用,我们将评估并预测未来我国粮食安全形势的演变趋势。

- (1) 问题 1: 首先通过相关性分析,分别研究温度,湿度,日照等因素与稻谷,小麦,玉米产量之间的关系。由于各粮食产量之间存在相关性,使用考虑输出相关性的高斯过程模型,同时使用 Shapley 分析方法,探究多个因子对产量的影响趋势。
- (2) 问题 2: 以粮食大省黑龙江省为例,描述耕地面积变化情况,提取包括人口因素,经济因素,农村居民因素,政策因素以及其他因素,使用基于主成分分析的非线性回归分析方法进行建模,通过主成分系数,描述多个因子对耕地面积变化的影响。
- (3) 问题 3: 查询最新文献资料和必要数据,分析中国近 20 年来粮食种子的自给率趋势。从数量和质量角度探讨种子退化、绝育问题及全球供应链布局的影响。此外,解析从按斤称售到按粒论卖销售方式转变背后的原因和理论逻辑。
- (4) 问题 4: 基于以上研究,揭示气候变化、耕地减少和种子短缺对中国粮食安全的因果关系。分析这些因素对中国粮食安全形势演变的趋势。

### 1.3 资料条件

本文针对我国粮食安全风险问题,从气候变化、耕地减少及种子短缺介三个方面开展研究。相应地,我们分别从这三个方面收集相关历史数据,包含非结构化数据和半结构化数据。总体收集数据包括近三十年中国水稻、小麦、玉米产量、播种面积、全国各省份年平均气温、年平均降雨量、平均日照时数、人口、GDP 增长率、黑龙江省耕地面积变化数据、黑龙江省人口及农业生产数据、黑龙江省土壤有机质、含氮量、PH 值等数据、海伦市耕地质量等级数据、黑龙江省耕地政策数据、全国粮食进出口量数据、2000-2023 国内国际粮价、2000-2022 相关法律法规、全球种业区域分布及市场份额数据等。具体数据集及来源见附件,详细说明见表 1。

表1 数据集介绍

数据集名称	来源
1990—2022 年中国 1km 分辨率平均降水量、气温、日照时数	国家地球科学数据中心 ( <a href="https://www.geodata.cn">https://www.geodata.cn</a> ) 青藏高原科学数据中心 ( <a href="https://data.tpdc.ac.cn">https://data.tpdc.ac.cn</a> )
“全国各地级市逐年平均降水数据”数据集	立方数据学院( <a href="https://www.lifangdata.com">https://www.lifangdata.com</a> ) 资源环境科学数据平台 ( <a href="http://www.resdc.cn">www.resdc.cn</a> ) 国家生态数据中心资源共享服务平台 ( <a href="http://nesdc.org.cn">nesdc.org.cn</a> )
“1981-2023 逐年平均气温”数据集	
“1960-2022 年中国各省份日照时数年度数据”数据集	
水稻、小麦、玉米播种面积及产量	《中国统计年鉴》 数字地球开放平台 ( <a href="https://open.geovisearth.com">https://open.geovisearth.com</a> )
社会经济及人口数据	《中国统计年鉴》
2003 年-2022 年黑龙江人口及农业生产数据(人口数、人口自然增长率、各产业生产总值、农业机械总动力、耕地面积等)	《黑龙江统计年鉴》,《中国农村统计年鉴》 科学数据银行 ( <a href="https://www.scidb.cn">https://www.scidb.cn</a> )
黑龙江省 250 米分辨率土壤粉粒含量、黏粒含量、砂粒含量、全氮含量、容重、酸碱度、有机碳含量数据集	国家地球科学数据中心 ( <a href="https://www.geodata.cn">https://www.geodata.cn</a> ) 国家土壤信息平台 ( <a href="http://www.soilinfo.cn">http://www.soilinfo.cn</a> ) 东北亚资源环境大数据中 ( <a href="http://www.igadc.cn/">http://www.igadc.cn/</a> )
海伦市耕地质量等级数据	文献[27-29]数据 科学数据银行 ( <a href="https://www.scidb.cn">https://www.scidb.cn</a> )
耕地政策信息	人民政府官网 ( <a href="https://www.gov.cn">https://www.gov.cn</a> ) 黑龙江省人民政府官网 ( <a href="https://www.hlj.gov.cn">https://www.hlj.gov.cn</a> )
全国粮食进出口量数据	文献[24], 中国农业生产指数 ( <a href="https://www.ceicdata.com.cn">https://www.ceicdata.com.cn</a> )
2000-2023 国内国际粮价	世界食品价格指数 ( <a href="https://zh.tradingeconomics.com">https://zh.tradingeconomics.com</a> ) 联合国粮食及农业组织 ( <a href="https://www.fao.org">https://www.fao.org</a> )
全球种业区域分布及市场份额数据	《2024 年中国种子行业深度研究报告》
2000-2022 种业相关法律法规	文献[13,18],新华网 ( <a href="http://xinhuane.com">http://xinhuane.com</a> ) 中国日报网 ( <a href="http://Chinadaily.com.cn">Chinadaily.com.cn</a> )

## 2 模型假设与符号说明

### 2.1 模型基本假设

(1) 假设我国未来不会出现颠覆性人口政策、经济政策和土地政策,不会对人口、耕地及社会经济造成颠覆性影响,国家经济平稳有序发展。

(2) 假设未来气候变化正常平稳,不会出现极端天气和重大极端自然灾害。

(3) Lesile 模型假设我国是一个完全封闭独立的人口系统,不考虑国际移民的数量,我国内部的人口变迁不会改变人口总量,不会出现人口骤降的情况。

(4) 不同地区耕地质量的变化对粮食作物产量的影响存在显著差异。

(5) 政府实施的耕地保护政策能有效减少耕地转非农化和非粮化的趋势。

(6) 所查阅的相关数据真实有效,且能准确反映指标所代表的含义。

### 2.2 符号说明

本文定义了如下 8 个使用次数较多的符号,其余符号在使用时注明。

表2 符号说明

符号	含义	单位
$Y_{\{it\}}$	粮食作物的产量(it 为玉米、稻谷、小麦)	万吨
$X_{Co_s}, X_{Ri_s}, X_{Wh_s}$	玉米、稻谷、小麦的播种面积	千公顷
$X_{Ra}$	降水量	米
$X_{Su}$	日照	小时
$X_{Te}$	温度	摄氏度
RMSE	均方根误差	无
RMAE	最大绝对误差	无
MAPE	平均绝对百分之误差	无

## 3 问题一分析与模型建立

### 3.1 问题分析

有关研究表明气温等气候因素的变化会对粮食造成减产影响,关系我国粮食安全。问题一以此为背景要求我们收集相关粮食产量及气候数据,研究气候对粮食产量的微观作用机理并预测未来 10 年我国的粮食净进口量走势。

可将问题一分为两部分进行分析解决。子问题 1: 收集有关数据,建模分析气候-粮食产量效应机理模型;子问题 2: 结合其他相关数据,建立粮食净进口量预测模型并预测我国未来 10 年粮食净进口量走势。

针对子问题 1,首先需要分析全国范围内各区域各省市气候及三大主粮产量相关数据统计特征,挖掘数据关联特征并分析相关性。针对多种作物粮食产量的多输出问题,采用单输出模型和多输出模型多种建模方法构建气候-粮食产量效应模型,并采用 Shapley 可解释性分析方法分析气候变化对中国粮食安全带来影响的微观机理。

针对子问题 2,考虑到粮食的进口量与国内的粮食供需关系存在着密不可分的联系,从粮食供需关系的角度出发,将进口量拆分为粮食供给与粮食需求两部分进行建模预测。基于子问题 1 提出的粮食产量模型,实现对全国粮食供给的预测建模;综合考虑人口、经

济、社会等因素,收集相关数据,共同构建粮食需求预测模型。进而组合两个预测模型,构建粮食净进口量预测模型,结合所收集数据对未来 10 年净进口量进行预测。

### 3.2 数据分析

问题一采用的数据主要包括：气候、土地利用、粮食产量、社会经济数据等。

气候数据包括近三十年中国各行政区划逐年平均气温、降雨量、日照时数等。气温数据来源于“1981-2023 逐年平均气温”数据集,降雨量数据来源于“全国各地级市逐年平均降水数据”数据集,日照时数数据来源于“1960-2022 年中国各省份日照时数年度数据”数据集。土地利用数据为近三十年中国各行政区划粮食播种面积数据,来源于《中国统计年鉴》。在问题一中主要研究水稻、小麦、玉米三种粮食作物播种面积。粮食产量数据为近三十年中国各行政区划粮食逐年产量数据,来源于《中国统计年鉴》。在问题一中主要研究水稻、小麦、玉米三种粮食作物的逐年产量数据。社会经济数据包括人口和 GDP 数据,均来源于《中国统计年鉴》。

#### 3.2.1 描述性统计

气候变化是影响粮食产量的主要影响因素之一,通常来说,温度、降水、日照等因素是直接影响粮食作物生长和产出的关键因子。图 1 为 1990—2022 年中国 1km 分辨率平均降水量分布图；图 2 为 1990—2022 年中国 1km 分辨率平均气温分布图。全国降水量与温度分布总体呈现东南降水量大温度高、西北降水量低温度低的分布趋势,不同降水量与温度决定全国各地区不同的粮食作物种植类型以及产量。随着时间变化,不同年份的降水量和气温分布略有不同,因此,需要对全国各行政区划逐年降水量、气温等气候数据进行统计分析。

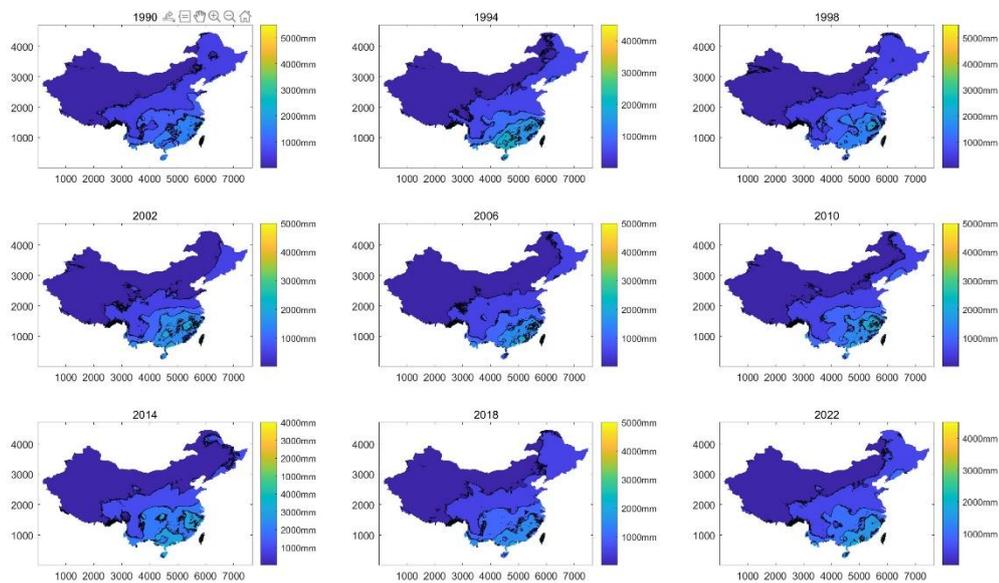


图1 1990—2022 年中国 1km 分辨率平均降水量分布图

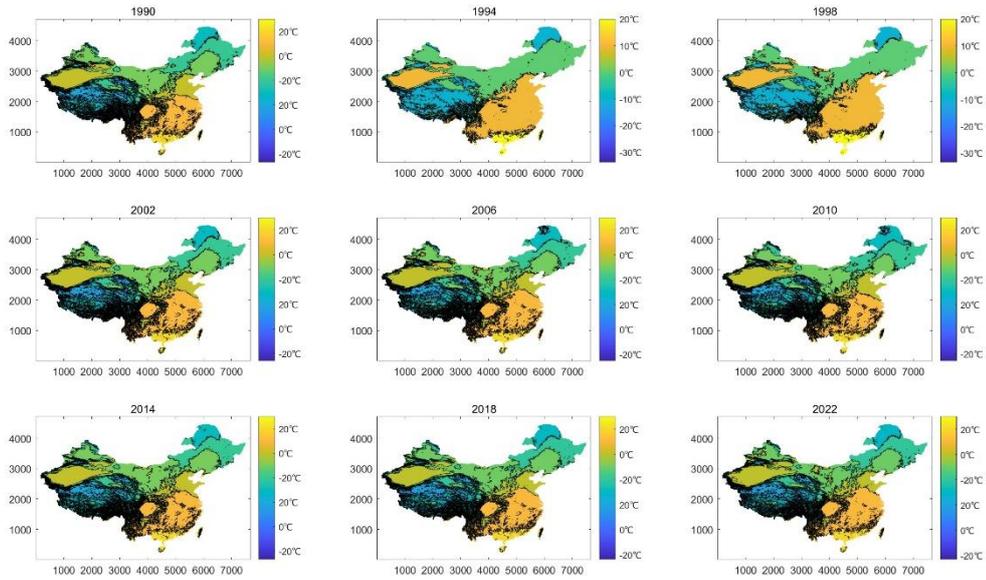
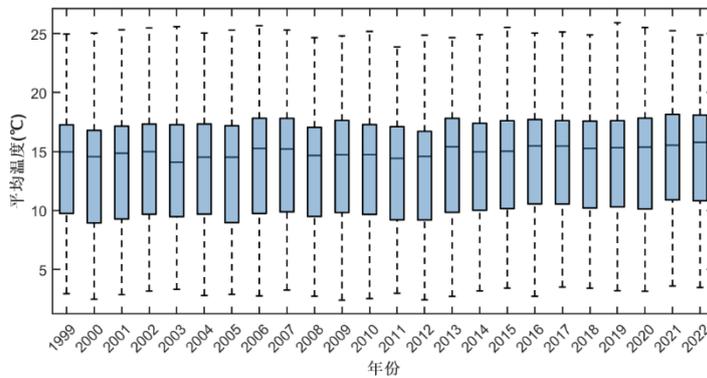
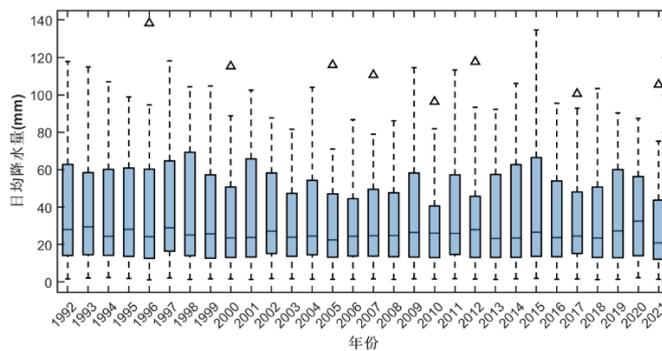


图2 1990—2022 年中国 1km 分辨率平均气温分布图

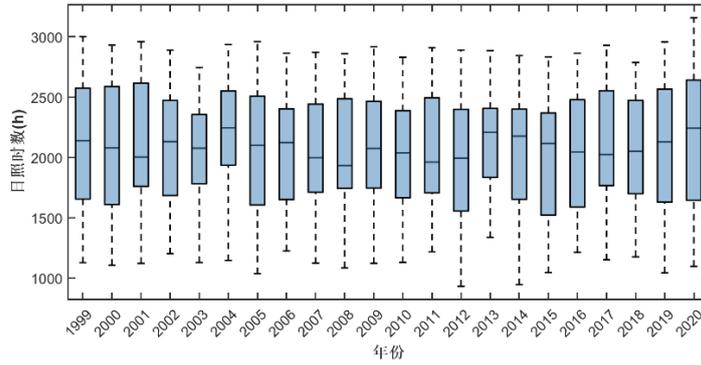
图 3 分别为各行政区划平均温度、平均降水量、平均日照时数随年份变化箱线图,可以发现,随着年份推移,全国各行政区划平均温度整体水平呈现一定的上升趋势,即全国气温在按照每年 0.1-0.2℃趋势逐年上升; 降水量总体趋势随时间波动,存在部分离群省份具有较高降水量; 日照时数随年份波动,未呈现明显变化趋势。



(1) 全国各行政区划平均气温统计



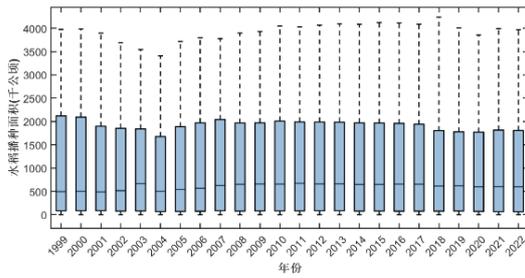
(2) 全国各行政区划平均降水量统计



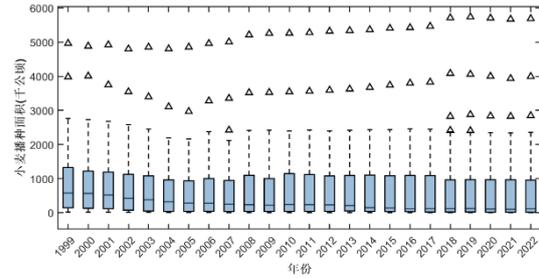
(3) 全国各行政区划日照时数统计

图3 全国各行政区划气候数据箱线图

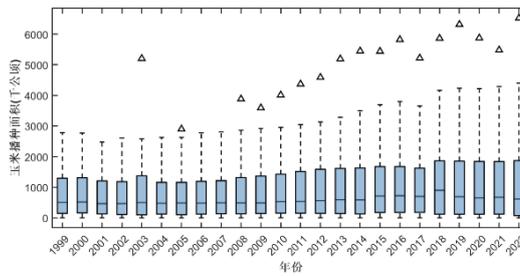
此外,对收集到的粮食产量及播种面积等相关数据进行统计分析。如图 3 和图 4 所示。可以发现不同行政区划的水稻播种面积总体分布较几种,但小麦和玉米播种面积不同省份差异较大,存在较多离群值。小麦和玉米年产量数据同样存在这样的特征。这说明不同行政区划间粮食作物种植及产量存在一定差异,这与不同地区的地域气候特征有关。另外,根据统计特征发现我国粮食产量随年份变化存在一定的增长趋势。



水稻播种面积统计

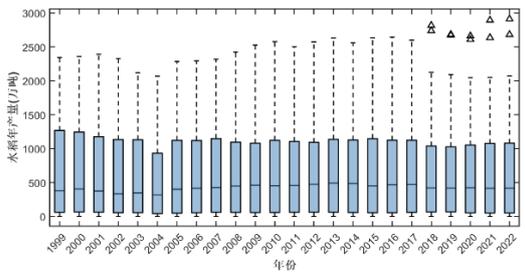


小麦播种面积统计

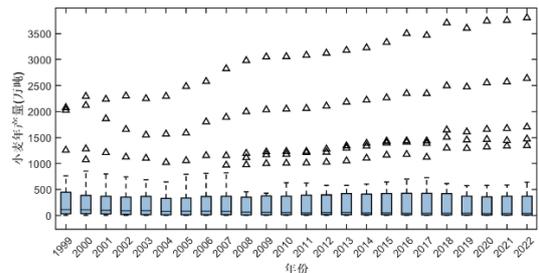


玉米播种面积统计

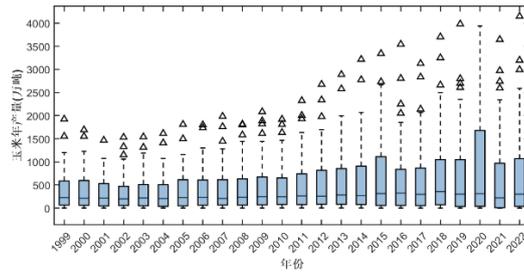
主粮播种面积统计数据



水稻年产量



小麦年产量



玉米年产量

图4 三大主粮年产量统计数据

综合上述统计分析结果可以得出结论：全国水稻、小麦、玉米等产量和播种面积存在一定的变化趋势,且不同行政区划的粮食产量分布和气候变化数据存在明显差异,但在相近区域中粮食产量与气候分布也分别存在一定相关性。在下一节中,将对农作物产量间相关性及农作物与气候数据相关性等进行进一步分析。

### 3.2.2 相关性分析方法

经典的相关性分析方法包含 Pearson 相关性分析,Spearman 相关性分析以及 Kendall 相关性分析。由于本文所解决问题涉及到的数据均为定量历史数据,故使用 Pearson 相关系数也叫简单相关系数或者线性相关系数,用来检测两个连续型变量之间的线性相关程度。

总体的 Pearson 相关系数用  $\rho$  表示,计算公式为

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x\sigma_y} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x\sigma_y} \quad (1)$$

样本的 Pearson 相关系数用字母  $r$  表示,用来度量两个变量间的线性关系,计算公式为:

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} \quad (2)$$

即用两个变量的协方差除以两个变量标准差的乘积。其中,Cov(X,Y): 表示的是协方差。Var[X]: 表示的是方差,开根号之后就变成了标准差。也就是说,两个变量之间的 Pearson 相关系数定义为两个变量之间的协方差和标准差的商。

### 3.2.3 农作物产量相关性分析

首先对水稻,小麦,玉米产量之间的相关性进行分析。可以看到,三种粮食作物的产量存在几种明显的趋势特征。

1、以东三省为代表的地区,该类地区稻谷和小麦的产量存在较强负相关,而稻谷和玉米的产量存在较强正相关,湖北,湖南,江西等地与该地区相同,可能与稻谷和玉米的耕地使用重复有关。

2、以福建,广东为代表的区域,该类地区稻谷和小麦存在较强正相关,而玉米和小麦无明显关系。

3、山东,河南两地,以及云南,西藏,重庆,宁夏,四川等西部地区,小麦和玉米的产量存在较强正相关。

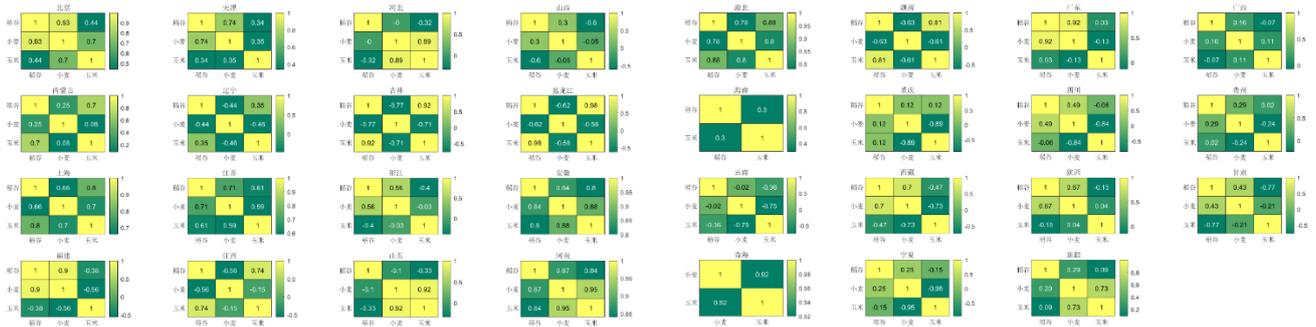


图5 农作物产量相关性分析

### 3.2.4 农作物产量与气候相关性分析

分别对农作物产量输出与影响因子之间的相关性进行分析。相关系数可以基本表明哪些因素会显著影响农作物的产量。对农作物产量和气候因素以及年份的相关性进行分析。相关性系数的大小代表正相关和负相关。图 6 至图 7 分别表示稻谷总产量,小麦总产量,水稻总产量与气候因素（降水,日照以及气温）的关系。

图 6 中可以看到,对某些地区例如北京,黑龙江等而言,稻谷产量与稻谷种植面积成较强正相关,也就是说,稻谷平均产量基本保持不变,受其他因素影响较小。且各省产量与年份均存在较强相关性,正相关表明产量逐年上升而负相关表明产量逐年下降。降水量,日照以及平均气温三个气候影响因子中,降水量的影响相对而言较小,而日照和平均气温对部分省份的影响较大。例如,日照时长对内蒙古,天津,黑龙江,湖北湖南等地有负面影响,对云南,西藏,新疆,宁夏等西部地区等地则存在正向影响。平均气温对内蒙古,吉林,湖北等地存在负面影响,对新疆,宁夏等地区则是正面影响。这些趋势表明,日照和气温对稻谷产量的影响,可能与地域分布有关,对于不同的区域,气候条件对产量的影响略有不同。

水稻为高温,多湿植物。在山西,黑龙江,海南,宁夏等地,不同地域的水稻产量均随着降水量的升高而升高。相关性系数分别为 0.52,0.37,0.38。

水稻为短日照植物,但相关研究发现水稻在长日照情况下仍能正常生长,日照对水稻产量的影响也存在地域差异。在内蒙古,黑龙江,湖北湖南等地,相关性系数分别为-0.54,-0.48,-0.46,-0.39。但在云南,西藏,陕西,宁夏,新疆等地区,水稻产量与日照时长的相关性系数分别为 0.49,0.43,0.38,0.52,0.41,存在较强正相关。这一现象在论文[1]中也有较为全面的解释。温度与日照对应,对水稻的影响也存在地域差异,且趋势与水稻相同。

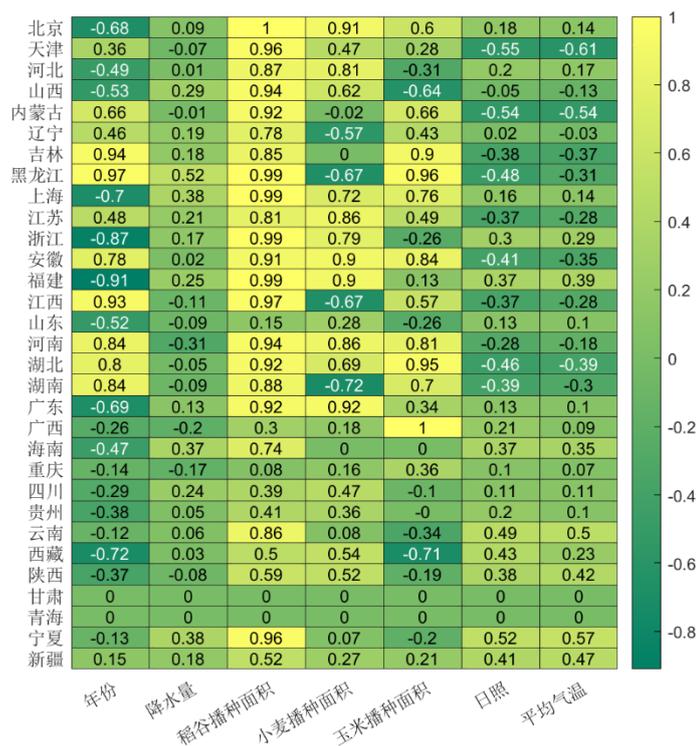


图6 各省因子与稻谷产量相关性分析

图 7 为各因素对小麦总产量的影响,其中,大多数地区小麦产量与耕地面积存在较大正相关,例如北京,福建,广东等地区。这说明在 1999 年至 2020 年之间,这些地区的平均产量维持平衡,未有较大增长。而山西,青海,陕西等地区相关性则较低。值得注意的是,河北的总产量与面积成负相关,从图 8 中的详细数据可以看出,虽然耕地面积略有下降,但其小麦总产量却保持了上升趋势。此外,从图 8 中可以看出,相关性较低的山西等地,耕地面积也呈下降趋势,但小麦总产量基本保持稳定。

小麦喜欢温凉,湿润的条件,因此降水量对于小麦的影响大多数是正向的。例如在江西,湖南,贵州,云南等地,降水量与小麦产量的相关性可以达到 0.5,具有较强相关性。

日照时长对于山东,河南,江苏,安徽等地区的小麦产量也是存在负向的影响,相关系数分别为-0.45,-0.46,-0.5,-0.48。对于重庆,四川,西藏,宁夏等地存在正向的影响,相关系数分别为 0.41,0.47,0.52,0.57。从分布可以看出,日照时长对小麦产量的结果也存在地域差异。这一差异与日照对水稻的影响相同。

对于气温条件,在江苏,浙江,安徽等江浙一带,小麦产量随着平均气温的升高而降低,同样在山东河南地区,小麦产量也随着平均气温的升高而降低。但对于宁夏,四川,重庆等西部地区而言,小麦产量随着平均气温的升高而升高,同样表现与明显的地域差异。

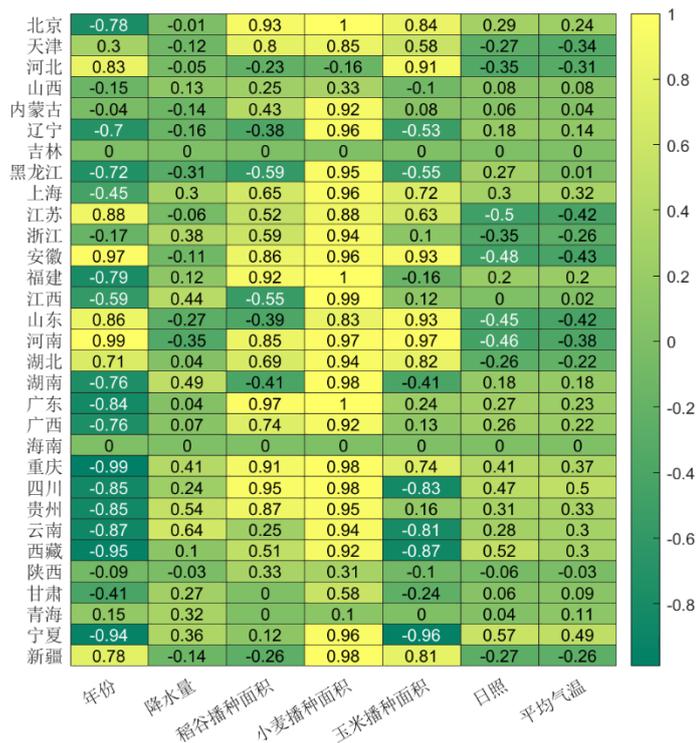
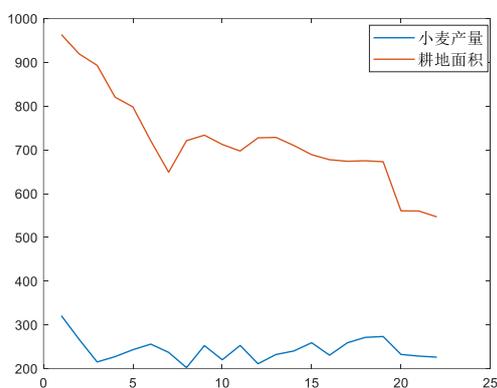
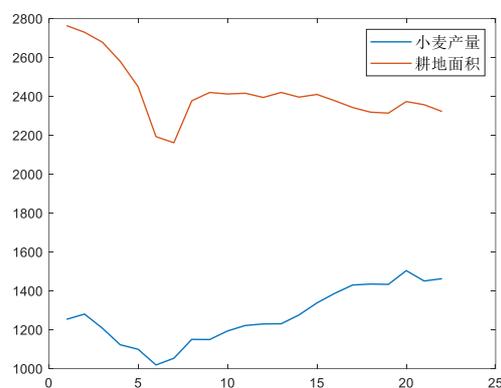


图7 各省因子与小麦产量相关性分析



山西小麦产量与耕地面积的关系



河北小麦产量与耕地面积的关系

图8 各省因子与小麦产量相关性分析

图9为玉米产量与气候条件以及年份的关系。其中大部分地区产量与面积具有强相关性,也就是说,单位面积的玉米产量保持稳定。且与水稻小麦不同,大多数地区的玉米总产量随年份增长趋势明显。而对于气候因素,降水量主要对黑龙江,新疆等地有影响,影响分别为正向影响和负向影响。日照对大多数地区的玉米产量都是负向影响,其中对上海为正向影响。与日照类似,气温对玉米的产量也大多为负向影响,因此可以看出,气温和日照对玉米产量的影响均为负向。

玉米为短日照植物,在短日照时间下可以开花结果,因此随着日照时长的增加,玉米产量反而呈下降趋势。且与水稻小麦不同,这一趋势没有表现出较强的地域差异。

对于温度而言,大多数地区的玉米产量均随着温度的升高而降低,呈现出显示的负相关。

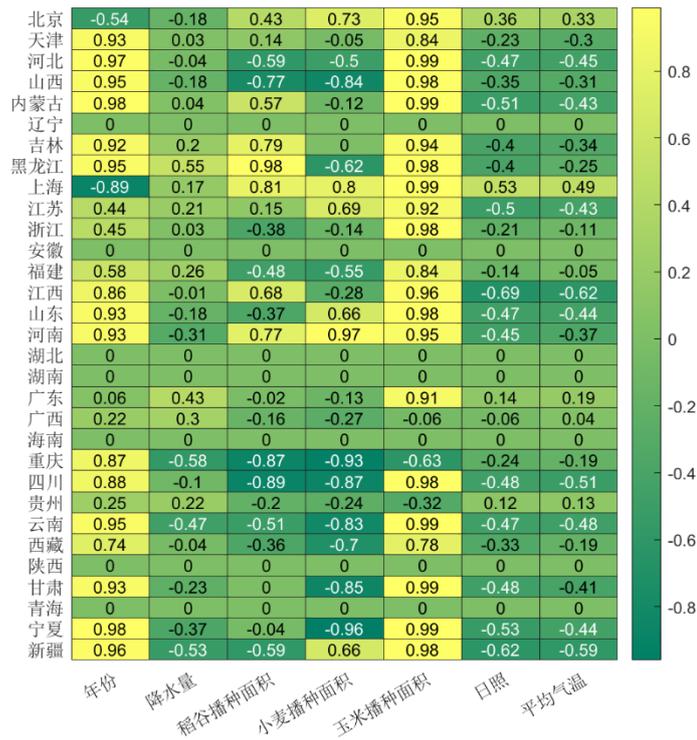


图9 各省玉米产量与因子相关性分析

### 3.3 模型准备

#### 3.3.1 Mann-Kendall 趋势检验

Mann-Kendall 是一种用于检测时序数据大致趋势的方法,趋势分为无明显趋势,趋势上升以及趋势下降。MK 检验不要求数据是正态分布,也不要求趋势是线性的,但需要时间序列数据要具有代表性。该方法被广泛用于降水,温度,干旱水质等的分析。本文将应用于 23 个省,5 个自治区,4 个直辖市,2 个特别行政区的粮食作物产量,进出口数量以及气候的时序数据中。

MK 检验是一种假设检验方法。在 MK 检验中,假设原数据为时间序列数据,算法步骤如下:

1. 零假设为该数据没有单调趋势
2. 数据按照时间序列依次取出,假设该时间序列数据为  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$
3. 确定所有  $n(n-1)/2$  个  $x_i - x_j$  的差值函数  $f(x_i - x_j), 1 \leq j < i \leq n$

$$f(x_i, x_j) = \begin{cases} -1, & x_i - x_j < 0 \\ 0, & x_i - x_j = 0 \\ 1, & x_i - x_j > 0 \end{cases} \quad (3)$$

4. 计算 S 的值,用差值为正的数量减去差值为负的数量

$$S = \sum_{j=1}^{n-1} \sum_{i=j+1}^n f(x_i - x_j) \quad (4)$$

如果  $S$  是正数,则表明后一部分观测值相对于前一部分观测值会趋向于变大,若  $S$  是负数,则表明后一部分观测值相比于之前的观测值趋向于变小。

### 5. 计算 $S$ 的方差

若每个  $x$  都是唯一的,则:

$$VAR(S) = \frac{1}{18}n(n-1)(2n+5) \quad (5)$$

若数据中存在  $x$  不唯一,则:

$$VAR(S) = \frac{1}{18}[n(n-1)(2n+5) - \sum_{p=1}^g t_p(t_p-1)(2t_p+5)] \quad (6)$$

其中,  $p$  为重复  $x$  的数量,  $g$  为唯一  $x$  的数量,  $t_p$  为重复的次数,

### 6. 计算 MK 的 Z 值

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{VAR(S)}}, S > 0 \\ 0, S = 0 \\ \frac{S+1}{\sqrt{VAR(S)}}, S < 0 \end{cases} \quad (7)$$

7. 采用双边趋势检验。在给定显著性水平为 0.05 的情况下,当统计量  $Z_{MK}$  的绝对值大于 1.64 时,通过置信度为 95% 的显著性检验。

本文分别对稻谷,玉米,小麦产量以及气温,湿度,日照等数据进行时间序列趋势检验。将相同坐标不同年份的值当做一个时间序列,逐个计算。我们对每个省份的三种粮食产量、播种面积,气候降水量、日照强度、温度等指标进行了如下步骤的分析:

#### (1) Mann-Kendall 检验:

- 计算每个时间序列的秩。
- 计算 Mann-Kendall 统计量和对应的 p 值。
- 判断是否拒绝原假设 (即数据没有趋势)。

#### (2) Sen's Slope 估计:

- 计算每个时间序列的斜率 (趋势)。
- 利用置信区间评估斜率的显著性。

粮食产量趋势分析以福建省对应粮食产量的 Mann-Kendall 检验结果为例:

表3 福建省对应粮食产量的 Mann-Kendall 检验结果

	Mann-Kendall Z	Mann-Kendall p	Sen's Slope
稻谷产量	-5.3085	1.11E-07	-10.1729
小麦产量	-5.86312	4.54E-09	-0.22

玉米产量	3.908748	9.28E-05	0.488889
降水量	-0.73949	0.459608	-0.00015
稻谷播种面积	-5.96876	2.39E-09	-27.4231
小麦播种面积	-6.04799	1.47E-09	-0.6
玉米播种面积	2.033605	0.041991	0.527273
日照	0	1	-0.12943
温度	1.901553	0.05723	0.028959

其中稻谷、小麦、玉米的产量与播种面积检验  $p$  值均小于 0.05, 可以认为对应的时序数据存在趋势性, 而降水量等气象因素变化较为复杂, 随时间趋势不够明显。而对应趋势, 稻谷、小麦产量与播种面积随时间呈显著下降, 玉米产量与播种面积显著上升。

气象变化趋势分析以黑龙江省对应粮食产量的 Mann-Kendall 检验结果为例:

表4 黑龙江省对应粮食产量的 Mann-Kendall 检验结果

	Mann-Kendall Z	Mann-Kendall p	Sen's Slope
稻谷产量	5.229270364	1.70E-07	97.725
小麦产量	-3.644642981	0.000267763	-4.5
玉米产量	5.017986713	5.22E-07	180.225
降水量	2.4825829	0.013043371	0.000272
稻谷播种面积	5.070807626	3.96E-07	127.8
小麦播种面积	-4.120031196	3.79E-05	-17.94
玉米播种面积	4.806703062	1.53E-06	238.48
日照	-1.267701906	0.204904424	-9.24752
温度	1.109239168	0.26732702	0.03204

除其中稻谷、小麦、玉米的产量与播种面积检验  $p$  值均小于 0.05, 可以认为对应的时序数据存在趋势性以外, 黑龙江省的降水量随时间的变化也存在一定上升趋势。

### 3.3.2 一阶差分法

由于粮食作物产量不仅受气候因素的影响, 还受到环境, 农作物品种, 土壤管理, 化肥, 农药等的影响。为评估气候变化对粮食作物产量的影响, 需要消除非气候因子的效应。这里对数据进行一阶差分变换, 尽量减少农作物品种, 肥料和其他非气候因素对产量的影响 [1]。

## 3.4 基于 OLS 分析的气候-粮食产量效应模型

使用 OLS 模型对气候与粮食产量间影响关系建模, 分析并描述气候变化对粮食产量及我国粮食安全的微观机理。对每个因变量 (粮食产量、降水量、日照强度和温度) 进行 PanelOLS 分析, 以下是主要步骤:

(1) 模型设定:

- 控制变量选择: 确定需要控制的个体和时间固定效应。
- 模型设定: 建立面板数据模型

$$Y_{\{it\}} = \beta_0 + \beta_1 X_{\{it\}} + \gamma_i + \lambda_t + \epsilon_{\{it\}} \quad (8)$$

其中, $Y_{\{it\}}$ 是因变量, $X_{\{it\}}$ 是解释变量, $\gamma_i$ 和 $\lambda_t$ 是个体和时间固定效应, $\epsilon_{\{it\}}$ 是误差项。

(2) 估计和推断:

- 估计模型: 使用 PanelOLS 方法估计模型系数。
- 推断和显著性检验: 评估系数的显著性和模型的拟合度。

将历史数据带入模型,针对每个粮食作物的产量,我们得到了以下主要结果:

$$Y_{Co} = 18.51 - 0.57X_{Co_s} + 7.37X_{Ra} - 7.05X_{Su} + 23.14X_{Te} \quad (9)$$

$$Y_{Ri} = 1.25 - 0.67X_{Ri_s} + 21.76X_{Ra} - 3.46X_{Su} + 2.93X_{Te} \quad (10)$$

$$Y_{Wh} = 74.20 + 36.34X_{Wh_s} - 35.28X_{Ra} - 61.84X_{Su} - 0.55X_{Te} \quad (11)$$

其中, $Y_{Co}$ , $Y_{Ri}$ , $Y_{Wh}$ 为玉米、稻谷、小麦的产量, $X_{Co_s}$ , $X_{Ri_s}$ , $X_{Wh_s}$ 为玉米、稻谷、小麦的播种面积, $X_{Ra}$ , $X_{Su}$ , $X_{Te}$ 为降水量,日照,温度,通过对模型参数的分析,玉米与稻谷的产量与降水量、温度正相关,与日照负相关,小麦的产量与降水量、温度、日照均为负相关。

这与作物的相关特性十分相关,如玉米为喜温和对温度反应敏感的作物,并且对光照条件要求较高,全生育期都需要较为强烈的光照,另外对于水分要求虽然较多,但需要适当控制水量。类似的,稻谷起源于热带、亚热带,也是喜温作物,需要较高的生长温度。相反小麦为喜凉作物,生长对于水分的要求较少,模型的参数与作物自然习性也相对吻合。

分析气候变化的微观机理影响,随着气温和降雨量的升高玉米和水稻产量可能会随之增加,而小麦产量可能出现下降趋势,因此在分析气候对我国粮食安全影响时,应当注意气温的升高和降雨量的陡增对小麦等作物带来的负面影响,同时关注温度骤降和干旱可能带来的水稻和玉米的减产问题。

### 3.5 考虑产量相关性的气候-粮食产量效应模型

#### 3.5.1 多输出高斯过程

粮食产量由水稻、小麦、玉米三种主要作物产量共同组成,前文 OLS 模型仅对三种作物的产量分别进行建模分析,尽管这种模型可以直观解释气候因素对粮食产量影响的微观机理,但对单一输出建模的方法往往忽略了多个输出间的相关性,一定程度造成数据信息的损失。因此,为了提高模型精度和预测准确性,介绍一种多输出建模方法,将多输出相关性信息纳入建模,构建基于多输出高斯过程的气候-粮食产量效应模型。

对于多个输出的问题,在进行高斯过程模型核函数的计算时,其输入支持多维  $x$ ,但是输出只有一维,这是因为由输入变量之间的“距离”产生的协方差对于不同响应造成的影响是一致的。且各响应之间时互不相关的。因此,单输出高斯过程虽然可以较好地构建单响应模型,但无法模拟响应存在相关性的情况。两者的模型构造和训练过程相似,主要区别在于协方差矩阵的构造。在多输出高斯过程中,协方差矩阵的构造需要增加考虑响应之间的相关关系,因此其协方差矩阵的构造方法尤为重要。

Intrinsic coregionalization model (ICM) 是一种用来描述不同响应之间关系的高斯过程模型。ICM 包含基础高斯过程和相关性矩阵。其中基础高斯过程用于描述因子与响应之间的关系,相关性矩阵则描述不同响应之间的关系。

对于多输出高斯过程,即是存在多个高斯分布,以两个高斯分布为例:

$$f_1(x) = GP\left(0, k_1(x, x')\right) \quad (12)$$

$$f_2(x) = GP\left(0, k_2(x, x')\right) \quad (13)$$

对应样本数据集分别为:

$$S_1 = \{(x_{i,1}, f_1(x_{i,1}, 1)), i = 1, 2, 3, \dots, N_1\} \quad (14)$$

$$S_2 = \{(x_{i,2}, f_2(x_{i,2}, 1)), i = 1, 2, 3, \dots, N_2\} \quad (15)$$

$$f_1 \sim \mathcal{N}(0, K_1) \quad (16)$$

$$f_2 \sim \mathcal{N}(0, K_2) \quad (17)$$

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}\right) \quad (18)$$

响应建模结果可以表示为:

$$y_1 \square \mathcal{N}(0, K_1 + \sigma_1^2 I) \quad y_2 \square \mathcal{N}(0, K_2 + \sigma_2^2 I)$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \square \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 I & 0 \\ 0 & \sigma_2^2 I \end{bmatrix}\right)$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 I & 0 \\ 0 & \sigma_2^2 I \end{bmatrix} K_{f,f} = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}$$

此时的 $K_{f,f}$ 未考虑响应之间的相关关系,即斜对角线上的元素为0,那么可以考虑设置一种协方差形式,使得斜对角元素不为0,可以描述多个响应之间的关系,因此需要考虑如何设置协方差矩阵描述这种关系:

假设从一个高斯过程 GP 中采样得到 $u^1(x)$ 和 $u^2(x)$ ,通过将 $u^1(x)$ 和 $u^2(x)$ 进行线性转化得到 $f_1(x)$ 和 $f_2(x)$

$$f_1(x) = a_1^1 u^1(x) + a_1^2 u^2(x)$$

$$f_2(x) = a_2^1 u^1(x) + a_2^2 u^2(x)$$

其中, $u^1(x)$ 和 $u^2(x)$ 虽然共享一个协方差函数,但是相互独立, $f_1(x)$ 和 $f_2(x)$ 可以写成:

$$f(x) = a^1 u^1(x) + a^2 u^2(x)$$

其中, $a^1 = [a_1^1 \ a_2^1]$ ,  $a^2 = [a_1^2 \ a_2^2]$ ,

$f(x)$  的方差可以计算得到:

$$\begin{aligned} \text{cov}(f(x), f(x')) &= a^1 (a^1)^T \text{cov}(u^1(x), u^1(x')) + a^2 (a^2)^T \text{cov}(u^2(x), u^2(x')) \\ &= a^1 (a^1)^T k(x, x') + a^2 (a^2)^T k(x, x') \\ &= [a^1 (a^1)^T + a^2 (a^2)^T] k(x, x') \end{aligned}$$

定义 $B = [a^1 (a^1)^T + a^2 (a^2)^T]$ ,可以得到

$$\text{cov}(f(x), f(x')) = B k(x, x') = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} k(x, x')$$

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} f_1(x_1) \\ \vdots \\ f_1(x_N) \\ f_2(x_1) \\ \vdots \\ f_2(x_N) \end{bmatrix} \sim \mathbf{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} b_{11}K & b_{12}K \\ b_{21}K & b_{22}K \end{bmatrix} \right)$$

预测分布为:

$$p(y_* | y, X, \theta) = \mathbf{N}(y_* | \mu_*, K_{y_*, y_*})$$

均值和方差分别可以表示为:

$$\begin{aligned} \mu_* &= K_{f_*, f} (K_{f, f} + \Sigma)^{-1} y \\ K_{y_*, y_*} &= K_{f_*, f_*} - (K_{f, f} + \Sigma)^{-1} K_{f_*, f}^T + \Sigma_* \end{aligned}$$

### 3.5.2 模型可解释性分析

由于多输出高斯过程模型的复杂性,尽管可以带来良好的预测精度,但却对模型的解释性和微观机理的研究带来了困难。模型可解释性 Shapley 分析是一种解释机器学习模型的有效方法,该方法源自博弈论中解决多个局中人在合作过程中因利益分配而产生矛盾的问题,广泛应用于可解释性机器学习领域。

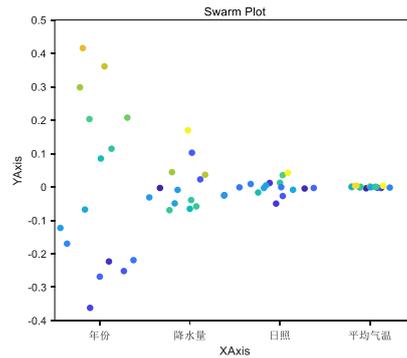
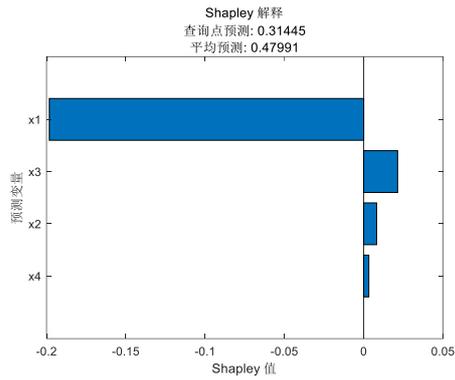
本文使用 Shapley 分析对多输出高斯过程模型做可解释性分析,分析气候因素等对粮食产量影响的作用机理。称温度、降水量、日照时数等气候因素为因子,粮食产量为响应。

令  $val(p) \in R$  是因子子集  $p \subseteq \{1, 2, \dots, n\} \equiv [n]$  的贡献,假设  $val(\emptyset) = 0$ 。对任意个体  $j \in \{1, 2, \dots, n\}$ , 其 Shapley 值都表示为  $\phi_j$ 。Shapley 值是解释为成员参与到每一个可能的集体中带来的边际贡献的权重之和。公式如下:

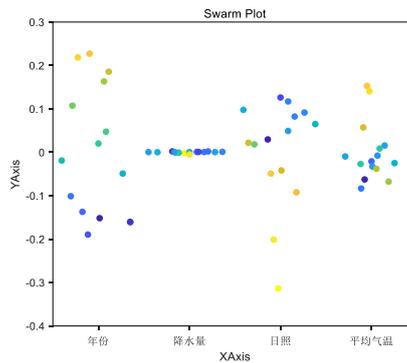
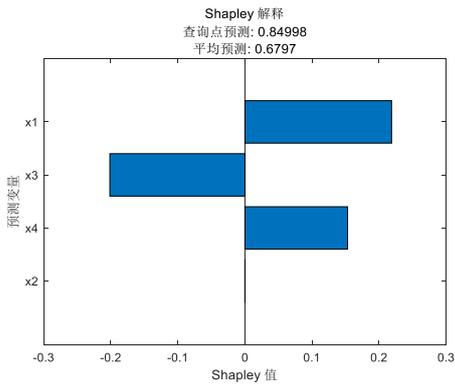
$$\begin{aligned} \phi_j &= \frac{1}{n} \sum_{p \subseteq [n] - \{j\}} C_{n-1}^{|p|} (val(p \cup \{j\}) - val(p)) \\ &= \sum_{p \subseteq [n] - \{j\}} \frac{(n - |p| - 1)! |p|!}{n!} (val(p \cup \{j\}) - val(p)) \end{aligned} \quad (19)$$

其中  $val(p \cup \{j\}) - val(p)$  为因子  $j$  对于因子子集  $p$  的边际贡献,  $|p|$  代表因子子集的因子数量,  $\frac{(n - |p| - 1)! |p|!}{n!}$  代表因子子集  $p$  的权重,该权重与顺序相关。  $val(p)$  为因子子集  $p$  的预测值,定义为:  $val(p) = \int_{i \notin p} f(X) dx_i - E(f(X))$ 。

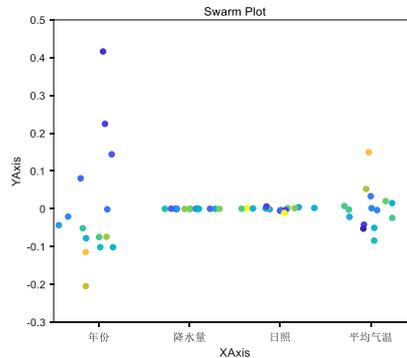
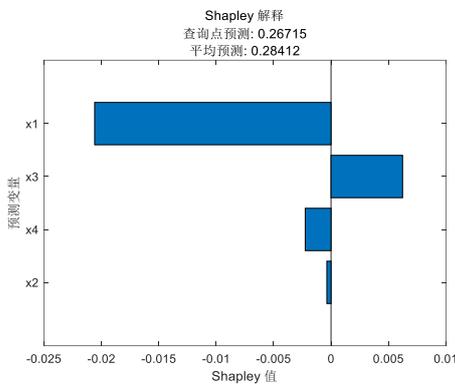
分别以湖北、河南、广东三省的粮食产量数据和气候数据为例进行分析,计算各因子 Shapley 值。Shapley 值越大则代表因子对响应存在较大影响。图 10 为三省的稻谷产量 Shapley 分析结果,另外两种作物分析结果见附录。



湖北稻谷产量 Shapley 分析



河南稻谷产量 Shapley 分析



广东稻谷产量 Shapley 分析

图10 三省稻谷产量 Shapley 分析

从上述分析中发现一些明显的地区特征。“降水量”在不同地区和作物中影响也不同。“黑龙江水稻”、“湖北玉米”、“湖北小麦”、“河南小麦”随着降水量增加,对其产量增加均有正向作用。而“黑龙江小麦”随着降水量增加,对其产量增加均有负向作用。降水量对“黑龙江玉米”、“湖北稻谷”、“河南水稻”、“河南玉米”、“广东各作物”的影响不明显。“日照”对广东地区的作物基本不影响,对“河南地区作物”、“黑龙江小麦”、“黑龙江玉米”影响较大,并且随着“日照”水平的增大,对作物产量增加均有负向作用。“日照”对湖北地区作物有一定影响,但趋势并不明显。“平均气温”对“黑龙江小麦”、“河南各作物”、“广东各作物”又较大影响。对“湖北各作物”、“黑龙江玉米”、“黑龙江水稻”的影响较小。“平均气温”对“黑龙江小麦”的影响趋势并不明确。而随着“平均气温”的增大,对“河南各作物”与“广东各作物”产量的增大均有正向作用。

### 3.6 粮食净进口量影响模型构建及预测

研究粮食净进口量变化情况及影响因素首先需要研究分析全国粮食供需关系及其影响因素。粮食作为人类赖以生存的基本要素,体现出需求层次理论汇总生理需求的要求,是不可或缺的。影响粮食需求的因素主要包括:人口、商品价位、消费政策等。结合中国实际,从以往文献中梳理出粮食需求的影响因子主要包括人口、经济、城市化水平、价格、气候等因素。

根据需求理论,当粮食总需求量=粮食总供给量才能保证国家的粮食安全总体平稳。相关文献研究发现,粮食的供给来源主要包括国内粮食生产与国外粮食进口。而我国所生产粮食主要流向三个方向,即国内自给、战略储备库存及出口,其中战略储备库存可视为固定需求,因此可认为我国粮食总需求 $N_S$ 为粮食产量 $N_P$ 与粮食净进口 $N_B$ 的总和,即

$$N_S = N_P + N_B \quad (20)$$

因此,为了对我国未来粮食净进口量进行预测研究,需要对每年粮食需求总量和每年粮食生产总量进行建模和预测分析。结合上文气候因素对以水稻、小麦、玉米为主体的粮食产量的影响关系建模,本文对问题进行简化,认为粮食总产量为水稻、小麦、玉米三大主粮产量的总和。根据粮食总产量假设条件,则粮食总产量预测值可根据前文所构建气候-粮食产量效应模型计算得到。接下来只需要研究粮食需求影响因素并建立粮食需求预测模型,即可对我国粮食净进口量进行建模和预测。

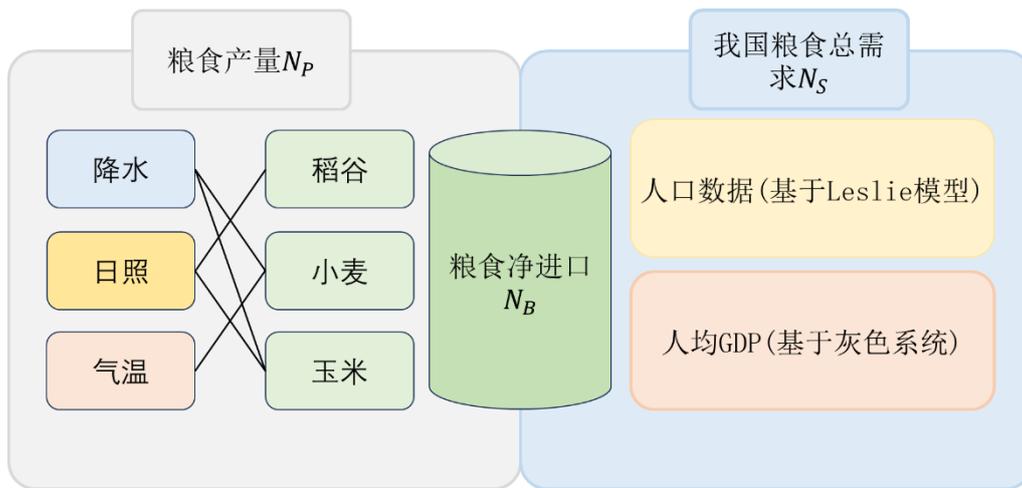


图11 粮食净进口量模型框架图

#### 3.6.2 粮食需求预测建模

要分析和预测粮食净进口量变动趋势需要建立粮食需求的长期模型。本文首先确定人口和 GDP 两个影响因素,分别从人口和经济两方面综合对粮食需求进行建模。人口数据来源于 1999-2022 年《中国统计年鉴数据》,GDP 数据来源于世界银行官网 1999-2022 年数据。参照 Hayami 和 Godo 的研究,给出粮食长期需求预测模型:

$$\frac{\Delta D}{D} \cong \frac{\Delta N}{N} + \beta \frac{\Delta \left( \frac{Y}{N} \right)}{\frac{Y}{N}} \quad (21)$$

其中 $D$ 代表粮食需求量, $N$ 代表人口总量, $Y$ 代表 GDP, $\Delta$ 代表增量, $\beta$ 代表粮食需求收入弹性。模型第一部分代表人口总量对粮食需求的影响,第二部分反映经济增长对粮食需求的拉动作用,同时受到粮食需求收入弹性的限制。进一步,对人口数据和 GDP 数据预测建模,带入上述模型,即可求解粮食需求预测值。

### 3.6.3 基于 Leslie 模型的人口数据预测建模

Leslie 模型处理人口预测问题,对人口性别进行分组,根据女性分年龄别人口数据、年龄别死亡率和年龄别生育率构造矩阵并进行计算,对女性人口数量作出合理预测,同时根据男女比例测算男性人口。相比于灰色模型、Logistic 模型等,Leslie 模型处理人口预测问题时,可将人口年龄、性别、层次等变化趋势纳入建模,有助于对人口数量变化趋势进行合理预测。

该模型假设我国是一个完全封闭独立的人口系统,不考虑国际移民的数量,我国内部的人口变迁不会改变人口总量。预测时间内,默认不发生战争以及重大自然灾害等,即不会出现人口骤降的情况。假定各年龄别生育率、死亡率和性别比例不随人口区域间流动变化。年龄别生育率只和年龄相关。预测时间内,我国人口生育政策不会变化。

设  $x_i(t)$  表示第  $t$  年第  $i$  组的女性人口数,其中  $i = 1, 2, \dots, 90$ ;  $b_i(t)$  表示第  $t$  年,第  $i$  组的女性生育率,根据女性生理条件,15 周岁到 49 周岁为正常的生育年龄区间,其他年龄段的生育率均为零;  $d_i(t)$  表示第  $t$  年第  $i$  组女性的死亡率;  $s_i(t)$  表示第  $t$  年第  $i$  组女性的存活率,则有  $s_i(t) = 1 - d_i(t)$ 。则有

$$x_i(t+1) = \sum_{i=1}^m b_i x_i(t) \quad (22)$$

$$x_{i+1}(t+1) = s_i(t)x_i(t), i = 1, 2, \dots, m-1; t = 0, 1, 2 \quad (23)$$

根据 Leslie 模型构造方法,可得以下矩阵:

$$L = \begin{pmatrix} b_1 & b_2 & \dots & b_{m_1} & b_m \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & 0 & \dots & \dots \\ \dots & 0 & \dots & 0 & \dots \\ 0 & \dots & 0 & s_{m_1} & 0 \end{pmatrix} \quad (24)$$

由以上可得  $x(t+1) = Lx(t)$ ,进而得到预测公式:  $x(t) = L^t x(0)$ 。通过搭建  $L$  矩阵,将所需数据带入上式,可求出第  $t$  年各年龄段女性人口分布数量,通过男女性别比例测算预测未来我国长期总人口量。

如图 12 所示为 2020-2035 年我国人口数量变化趋势预测情况,红线所在位置为 2024 年,红线右侧区域为人口数量预测值。从趋势上看,我国人口在未来将呈现由增长到缓慢下降的趋势,人口数量基本保持在 14.8 亿左右,在 2029 年左右将进入人口由增长到下降的转折点。

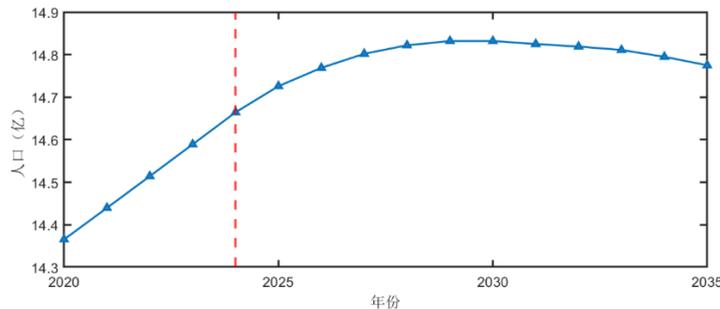


图12 人口预测结果

### 3.6.4 基于灰色系统理论的人均 GDP 预测建模

由于人均 GDP 的发展受社会、经济、政策等多种因素影响,这些因素影响作用复杂,很难通过以往经验加以判断。灰色系统 GM(1,1)模型预测精度高且所需信息较少,广泛应用与各领域数据分析。因此本文使用灰色系统理论对我国人均 GDP 数据进行预测分析。

首先设定原始数据序列:

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$$

对原始序列做一阶累加生成:

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$$

其中

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$$

接着构造 $z^{(1)}$ 序列,令 $x^{(1)}(k) = \frac{1}{2}[x^{(1)}(k) + x^{(1)}(k-1)]$

可得

$$z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$$

建立白化方程

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b$$

求取参数 $a$ 、 $b$ 的值,若 $\hat{a} = [a, b]^T$ 为参数序列,且

$$B = \begin{bmatrix} -\frac{1}{2}z^{(1)}(2) & 1 \\ -\frac{1}{2}z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}z^{(1)}(n) & 1 \end{bmatrix}, Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

运用最小二乘法进行求解,进一步处理白化方程,保证其离散化,尤其可得 GM(1,1)微分方程

$$b = x^{(0)}(k) + az^{(1)}(k)$$

求解白化微分方程,得到模型的时间响应方程如下:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}$$

还原值:

$$\hat{x}^{(0)}(k+1) = x^{(1)}(k) - x^{(1)}(k-1)$$

其中,发展系数 $-a \in [-2, 2]$ , $b$ 为灰色作用量。

由于政策、国际环境等不确定性因素对 GDP 的显著影响,根据所收集人均 GDP 增长率数据,选择近 20 年数据代入灰色系统 GM(1,1)模型进行计算并预测未来十年变化趋势,如图 13 所示。

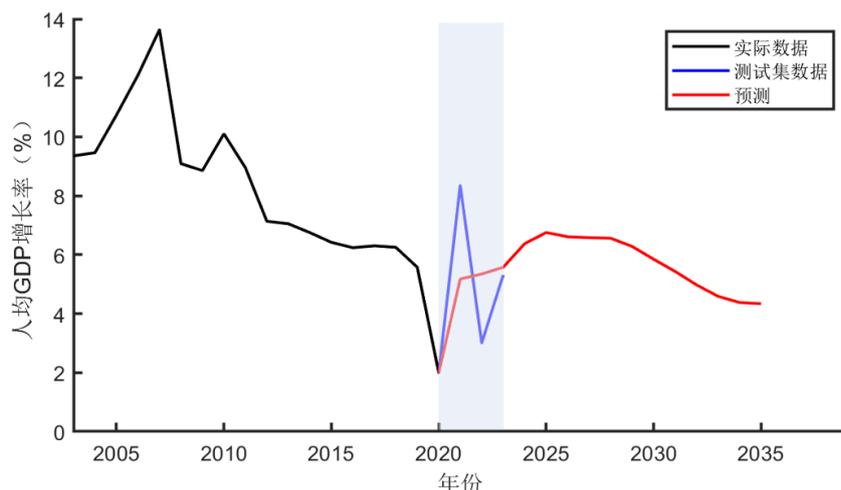


图13 人均 GDP 增长率预测结果

图 13 中蓝色区域为测试集区域,由于 2022 年新冠疫情这一黑天鹅事件的不可预测性导致人均 GDP 增长率预测值未出现下降情况,但总体预测趋势较符合。从未来时间预测情况看,人均 GDP 增长率呈现平稳下降趋势,基本维持在 4%-5%之间波动。

### 3.6.5 未来十年粮食净进口量建模与预测

根据粮食总体供需关系式,推导出第*i*年粮食净进口量预测模型为

$$\begin{cases} N_B(i) = N_S(i) - N_P(i) \\ N_P(i) = N_P(i-1) \left( \frac{\Delta N}{N} + \beta \frac{\Delta \left( \frac{Y}{N} \right)}{\frac{Y}{N}} \right) \\ N_S(i) = Y_{Ri}^i + Y_{Wh}^i + Y_{Co}^i \\ N_P(0) = D_0 \end{cases} \quad (25)$$

其中, $D_0$ 为已知初始年份粮食总需求量,  $Y_{Ri}^i$ 、 $Y_{Wh}^i$ 、 $Y_{Co}^i$ 分别表示第*i*年水稻、小麦、玉米产量。

根据未来十年预测人口数据及人均 GDP 增长率,计算未来十年中国粮食需求平均增长率,并根据 2023 年粮食总需求量 86700 万吨为初始值计算每年粮食需求总量,其中未来十年粮食需求收入弹性 $\beta$ 根据文献确定,结果如表 5 所示。

表5 未来十年粮食总需求量及增长率预测结果

年份	人口 (亿)	人均 GDP 增长率 (%)	粮食需求收入弹性	经济增长带来的粮食需求增长率 (%)	人口增长带来的粮食需求增长率 (%)	粮食需求增长率 (%)	粮食需求预测 (万吨)
2024	14.6645	6.3753	0.038	0.242261	0.516	0.758261	87357.41265
2025	14.726	6.753541	0.038	0.256635	0.419384	0.676019	87947.96523
2026	14.769	6.608818	0.028	0.185047	0.292001	0.477047	88367.51876
2027	14.802	6.579609	0.028	0.184229	0.223441	0.40767	88727.76666
2028	14.822	6.559711	0.028	0.183672	0.135117	0.318789	89010.62082
2029	14.8321	6.281049	0.028	0.175869	0.068142	0.244011	89227.81682
2030	14.832	5.848049	0.028	0.163745	-0.00067	0.163071	89373.32164

2031	14.825	5.434203	0.019	0.10325	-0.0472	0.056055	89423.41951
2032	14.819	4.976399	0.019	0.094552	-0.04047	0.054079	89471.77915
2033	14.811	4.589187	0.019	0.087195	-0.05398	0.03321	89501.49255
2034	14.795	4.37786	0.019	0.083179	-0.10803	-0.02485	89479.25279
2035	14.775	4.337826	0.019	0.082419	-0.13518	-0.05276	89432.04165

对于粮食总产量的预测计算,可通过对水稻、小麦、玉米三种主粮产量预测直接加和得到。根据气候-粮食产量效应模型,合理推算未来十年的气候情况即可预测未来十年粮食产量。由于气温、降雨、日照等气候因素随时间变化不确定性强,系统机理复杂,同样地,使用灰色系统 GM(1,1)模型对气温、降雨、日照等气候因素进行预测。进而得到未来十年三大主粮产量及粮食总产量预测值,如表 6 所示。

表6 未来十年三大主粮产量及粮食总产量预测值

年份	稻谷产量(万吨)	小麦产量(万吨)	玉米产量(万吨)	粮食总产量(万吨)
2024	21388.75	15244.52	27019.05	63652.32
2025	21441.07	16364.02	26766.55	64571.64
2026	21493.39	16650.44	27556.63	65700.47
2027	21545.72	16651.66	27786.64	65984.01
2028	21598.04	16993.09	28317.51	66908.64
2029	21650.36	16482.62	28686.76	66819.74
2030	21702.69	17134.57	29142.84	67980.1
2031	21755.01	17438.27	29552.27	68745.55
2032	21807.33	18200.26	29986.76	69994.35
2033	21859.66	19457.33	30407.79	71724.78
2034	21911.98	19816.27	30836.05	72564.3
2035	21964.3	20876.67	31260.43	74101.4

根据前文粮食总需求与总产量的预测值,由公式中国未来粮食净进口=中国未来粮食总需求-中国未来粮食总产量,可得未来十年中国粮食净进口预测结果,将净进口量、总需求量、总产量共同绘制预测趋势,如图 14 所示。

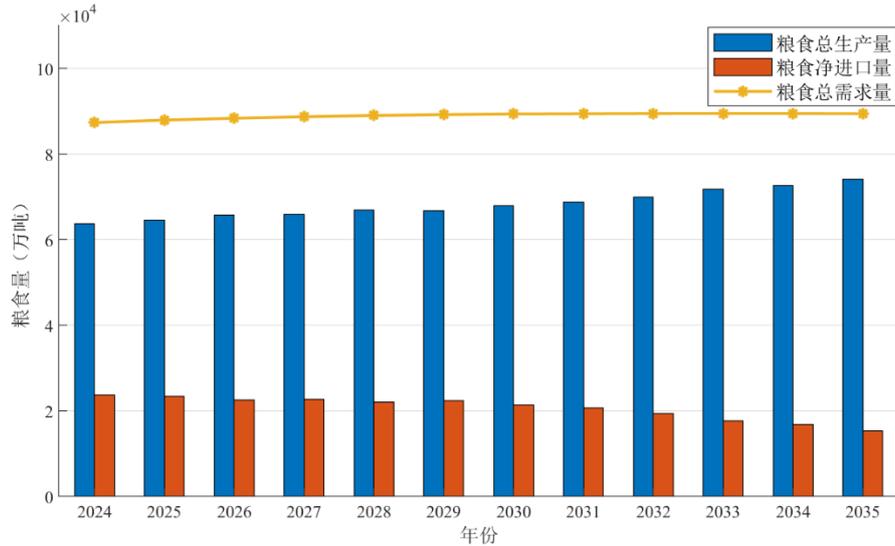


图14 未来十年净进口量走势图

可以看出,未来十年,粮食总生产量呈现上升趋势,而由于人口等因素的变化,粮食总需求量上升较缓慢,粮食自给率提高,因此预测未来十年我国粮食需求稳定,粮食净进口量呈现逐年下降走势。

## 4 问题二分析与模型建立

### 4.1 问题分析及数据来源

黑龙江省位于我国东北部,东部和北部以乌苏里江,黑龙江为界河与俄罗斯相邻,西接内蒙古自治区,南连吉林省,是中国位置最北,最东,纬度最高的省份。总面积 47.3 万平方公里,横跨 10 个纬度,14 个经度。黑龙江省下辖 12 个地级市,包括哈尔滨,齐齐哈尔,佳木斯,牡丹江,大庆,黑河,绥化,伊春,鸡西,鹤岗,七台河,双鸭山以及一个大兴安岭地区。黑龙江省内平原较多,其中西部为松嫩平原,东北部为三江平原,平原总面积占全省总面积的 37%。黑龙江省是我国耕地面积最大的省份,位于我国东北黑土区,物产丰盛,是我国最大的粮食产区,黑龙江省的耕地变化对于粮食产业有着举足轻重的作用。因此本文以黑龙江省为例,研究 2003 年-2022 年 20 年来,探究黑龙江 20 年来耕地面积和耕地质量的变化情况以及驱动因素。选取数据包括耕地面积数据以及其他相关社会经济数据。

其中,耕地面积数据来源于黑龙江省统计局中各年统计年鉴的农业部分,驱动因子数据页来源于 2003 年至 2023 年黑龙江省统计年鉴。

耕地质量作为描述耕地产量优劣的重要指标,通常由耕地养分、耕地酸碱性、耕地土壤侵蚀等多种因素表征。为此,建立综合评价指标体系描述耕地质量。海伦市地处松嫩平原与小兴安岭余脉交接过渡带的黑土核心地区,是黑龙江省重要粮食产地。综合考虑数据的获取难易程度,选择海伦市为代表,研究黑龙江黑土地核心区土壤质量变化情况,并结合我国相关遏制耕地“非农化”“非粮化”的政策,分析前后土壤质量变化趋势。

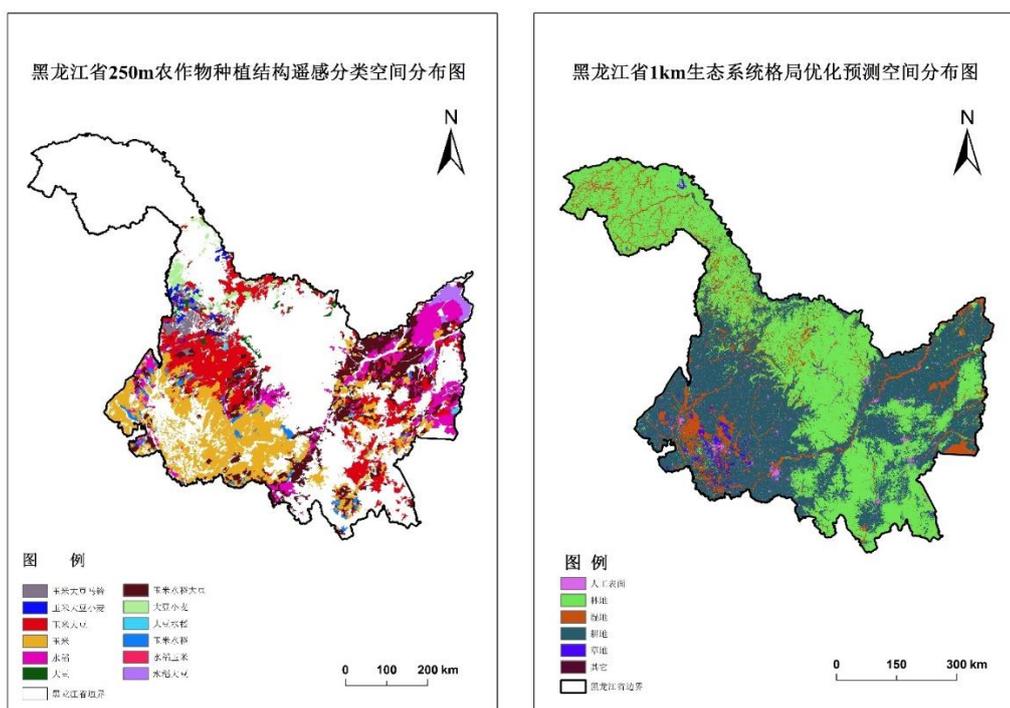


图15 黑龙江省空间分布图

#### 4.1.2 数据准备及数据补充

根据黑龙江省耕地面积变化情况已经实际发展情况,依据黑龙江省统计年鉴及相关数据资料,研究黑龙江省社会经济因素对耕地面积的影响。选取表 7 来表征相关影响因素。主要驱动因子包括人口因素,经济因素,政策因素,农村居民因素以及其他因素共 18 个驱动因子。

表7 驱动因子信息

人口因素	经济因素	农村居民因素	政策因素	其他因素
总人口	GDP	农村居民人均可支配收入	涉农支出金额	农业机械总动力
人口自然增长率	人均GDP	农村居民家庭恩格尔系数	涉农支出占财政支出比例	化肥施用量(氮肥,磷肥,钾肥,复合肥)
农村人口数	公共财政收入		农业生产总值	农村用电量
	公共财政支出		农林渔业生产总产值	
	第二产业生产总值			
	第三产业生产总值			

表8 各因子描述性统计结果

	N	最小值	最大值	平均值	标准差
化肥使用量	20	125700.00	2553071.00	1860070.39	911539.39

GDP（万元）	20	3609.73	15900.98	9913.53	3779.89
生产总值指数	20	100.86	111.70	107.48	3.19
农村居民人均可支配收入	20	2509.00	18577.00	9299.33	5200.56
人口自然增长率	20	(5.75)	2.67	0.14	2.55
农村居民家庭基本情况恩格尔系数	20	26.30	40.90	32.93	4.51
农林渔业生产总值	20	903.30	6718.24	3812.07	2047.07
农业生产总值	20	502.90	4320.48	2349.95	1363.53
第二产业生产总值	20	1956.41	5202.65	3780.10	894.17
第三产业生产总值	20	1160.30	7642.23	4080.57	2194.85
年末人口数	20	3099.00	3833.00	3586.54	268.06
农村人口数	20	1047.00	1808.70	1496.53	272.57
农业机械总动力	20	1807.74	7090.81	4457.90	1698.28
农村用电量	20	307402.00	967408.01	636786.83	217483.51
公共财政收入	20	248.86	1301.31	912.25	402.99
公共财政支出	20	564.91	5452.00	3061.55	1721.97
涉农支出	20	20.66	937.59	468.49	346.23
财政支出占比	20	3.66	18.97	12.65	5.36
有效个案数（成列）	20				

由于因子指标重多,且相关领域可能存在重叠,因此首先对因子之间相关性进行分析。但其中,与2022年省统计年鉴不同,黑龙江省2023年鉴数据中,不包括有效灌溉面积以及农村用电量。其中2023年统计年鉴数据中农业机械总动力,仅包括农业机械数量数据,未进行总动力计算。因此考虑使用趋势外推的方法,计算2021年,2022年有效灌溉面积以及2022年农村用电量。

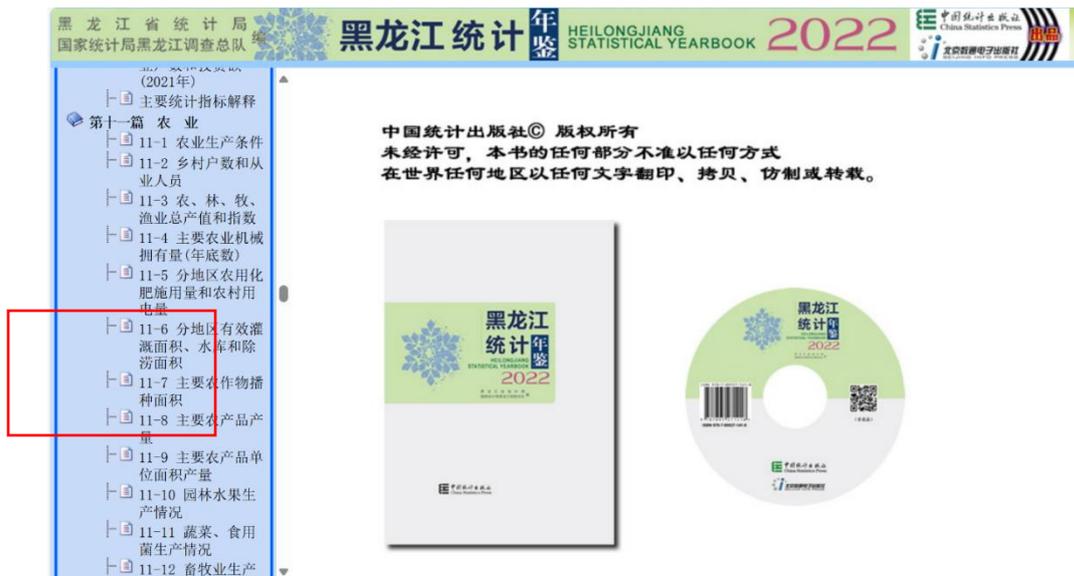


图16 2022 年统计年鉴

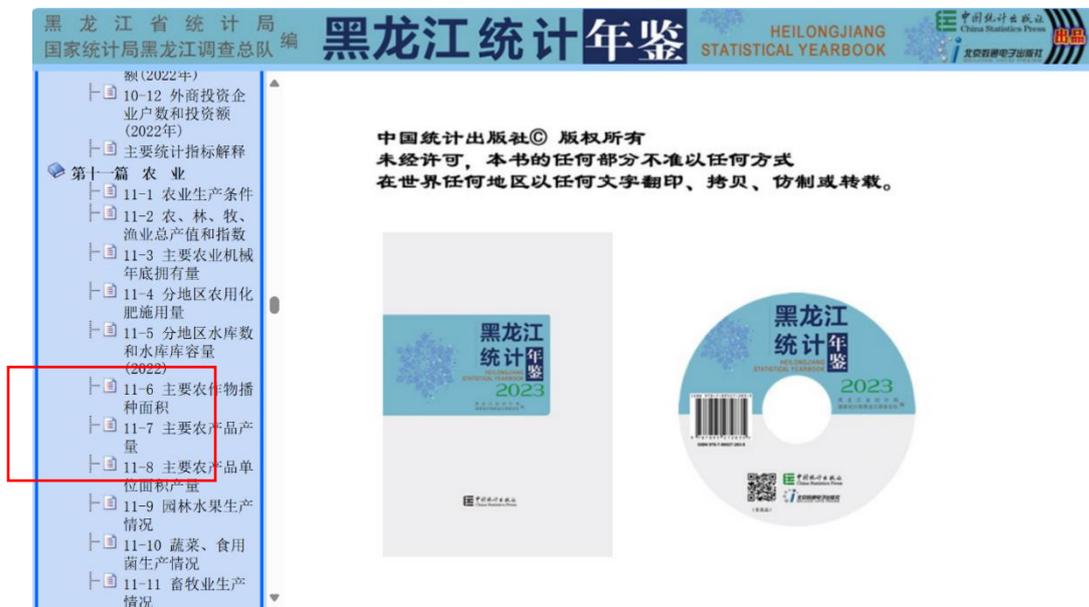


图17 2023 年统计年鉴

使用二次指数平滑方法预测 2021,2022 年灌溉数据以及 2022 年农村用电量。当时间序列数据的变动出现直线趋势时,用一次指数平滑方法进行预测,会存在明显的滞后偏差,因此,需要使用二次指数平滑的方法进行预测,利用滞后偏差的规律建立趋势模型。二次指数平滑的计算公式为:

$$S_t^{(1)} = \alpha y_t + (1 - \alpha)S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha)S_{t-1}^{(2)}$$

其中,  $S_t^{(1)}$  为一次指数的平滑值,  $S_t^{(2)}$  为二次指数的平滑值,预测公式为:

$$\hat{y}_{t+T} = a_t + b_t T, T = 1, 2, \dots$$

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \frac{\alpha}{1 - \alpha} (S_t^{(1)} - S_t^{(2)}) \end{cases}$$

使用二次指数平滑得到 2021,2022 的有效灌溉面积如图 18 所示,其中红色点标记的为预测值。

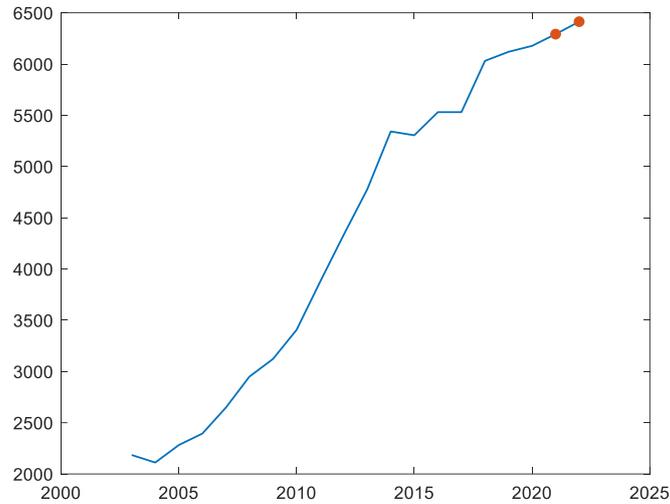


图18 有效灌溉面积趋势外推结果

同样对农村用电量进行指数平滑,  $\alpha$  取 0.3, 得到 2022 年预测值如图 19 所示:

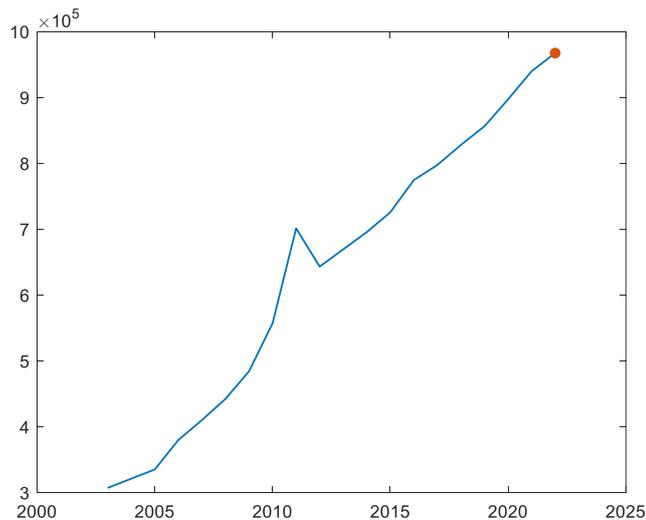


图19 农村用电量趋势外推结果

## 4.2 耕地面积变化定量描述及驱动机制研究

### 4.2.1 耕地面积定量描述

本小节选取数据为粮食作物实际播种面积,数据来源为黑龙江省 2023 年统计年鉴。为分析国家“非农化”,“非粮化”政策对于耕地面积的影响选择当前用于种植粮食作物的实际耕地面积作为研究对象。如图 20 为 2003 年至 2022 年耕地面积变化情况,随着城市用地,工业用地的不断增加,在其他省耕地逐年下降的同时,黑龙江省作为我国最大的粮食种植省,其粮食作物面积在稳定中保持上升趋势。从粮食作物耕地面积的变化可以看出,耕地面积仅在 2016 年,2017 年出现短暂的下降,而后保持稳定。

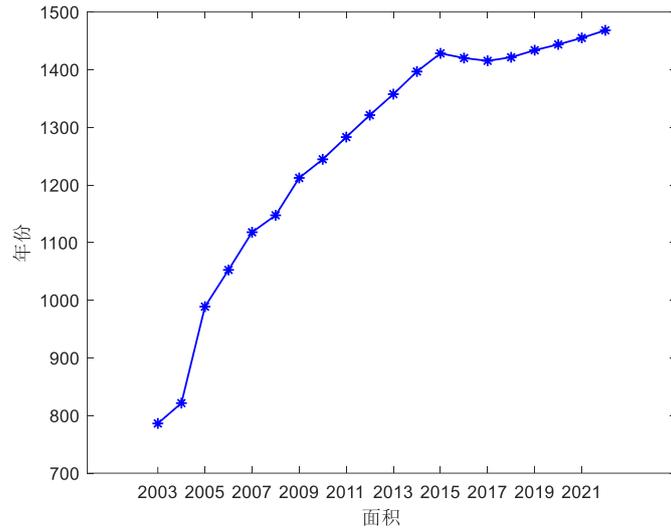


图20 黑龙江省粮食用地面积变化情况

图 21 为近 20 年来黑龙江土地利用情况,可以看到,2000-2020 年,耕地面积和建设用 地面积逐年上升,而林地,水域和草地面积逐年下降。其中,林地和草地在 2000-2015 年降 速较为缓慢,在 2015-2020 年急剧下降。水域面积在 2000-2015 年维持稳定,2015-2020 年 快速下降。未利用土地面积在 2000-2015 年逐年下降,在 2015-2020 则快速上升。与上述 分析对应,建设用地呈现上升趋势,对应工业产值的增长和 GDP 的增长,抢夺了可以转化 为耕地的其他用地。

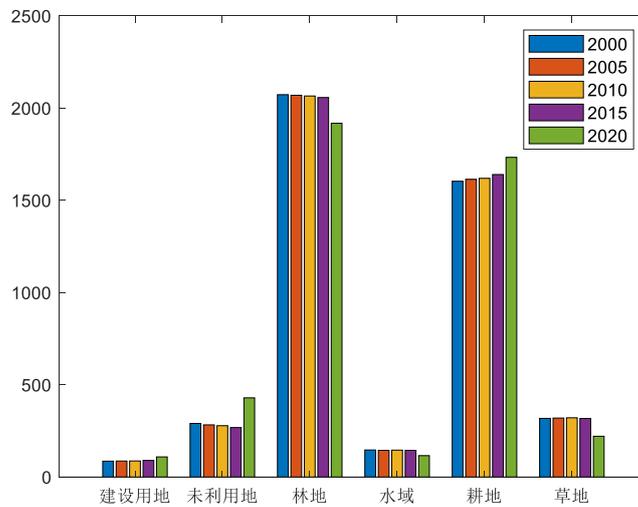


图21 2000-2020 年黑龙江土地利用情况

#### 4.2.2 耕地面积变化驱动机制主成分分析

如下表为 2003-2022 年黑龙江省上述信息的数据合集。

表9 黑龙江省数据 1

年份	耕地面积	化肥使用量	GDP	生产总值指数	人口自然增长率	农村居民人均可支配收入
2003	786.3	129700.0	3609.7	109.8	2.03	2509
2004	821.6	125700.0	4134.7	110.4	1.82	3005
2005	988.9	143810.0	4756.4	110.6	2.67	3221

2006	1052.6	150920.0	5329.8	110.4	2.39	3552
2007	1118.0	1751982	6126.3	109.8	2.49	4132
2008	1147.4	1807259	7134.2	110.3	2.23	4856
2009	1212.2	1988724	7218.9	110.1	2.06	5207
2010	1244.5	2148852	8308.3	111.7	1.52	6211
2011	1283.1	2284366	9935.0	110.9	1.07	7591
2012	1321.2	2402818	11015.8	108.7	1.27	8604
2013	1357.6	2449560	11849.1	107.6	0.78	9634
2014	1396.8	2519295	12170.8	105.3	0.91	10453
2015	1428.3	2553071	11690.0	105.4	-0.60	11095
2016	1420.2	2527469	11895.0	104.4	-0.49	11832
2017	1415.4	2511953	12313.0	106.0	-0.41	12665
2018	1421.5	2456410	12846.5	104.5	-0.69	13804
2019	1433.8	2232662	13544.4	104.0	-1.01	14982
2020	1443.8	2242160	13633.4	100.9	-4.48	16168
2021	1455.1	2389656	14858.2	106.1	-5.11	17889
2022	1468.3	2385041	15901.0	102.7	-5.75	18577

表10 黑龙江省数据 2

年份	农村居民家庭恩格尔系数	农林渔业生产总产值	农业生产总值	第二产业生产总值	第三产业生产总值	年末人口数
2003	40.7	903.3	502.9	1956.4	1160.3	3815.0
2004	40.9	1136.6	620.2	2270.3	1283.5	3816.8
2005	36.3	1294.4	718.6	2656.4	1425.4	3820.0
2006	35.3	1391.1	817.5	2998.0	1599.9	3823.0
2007	34.6	1591.7	873.4	3383.0	1850.9	3824.0
2008	33.0	2004.1	1051.7	3935.0	2125.4	3825.0
2009	31.4	2136.8	1136.5	3668.1	2409.0	3826.0
2010	33.8	2422.2	1320.1	4146.1	2870.4	3833.0
2011	35.1	3103.0	1778.5	4916.3	3323.2	3782.0
2012	37.9	3842.0	2339.8	5099.8	3796.5	3724.0
2013	35.2	4560.2	2954.7	5202.7	4106.9	3666.0
2014	28.2	4865.8	3193.6	4872.4	4607.4	3608.0
2015	27.5	5030.1	3156.9	3926.9	5050.9	3529.0
2016	27.7	5202.9	3189.7	3689.7	5454.0	3463.0
2017	26.5	5586.6	3471.3	3519.5	5828.2	3399.0
2018	26.3	5624.3	3635.0	3536.0	6309.3	3327.0
2019	26.8	5930.0	3774.5	3640.1	6721.1	3255.0
2020	34.3	6438.1	4044.1	3449.9	6738.5	3171.0
2021	33.6	6460.0	4099.5	4086.5	7308.3	3125.0
2022	33.5	6718.2	4320.5	4648.9	7642.2	3099.0

表11 黑龙江省数据 3

年份	乡村人口数	农业机械总动力	用电量	公共财政收入	公共财政支出	涉农支出	财政支出占比
2003	1808.7	1807.7	307402	248.9	564.9	20.6589	3.657
2004	1802.3	1952.2	321382	289.4	697.6	26.6089	3.8146
2005	1791.6	2234.0	335362	318.2	787.8	29.5465	3.7506
2006	1777.7	2570.6	380124	386.6	968.5	36.6294	3.782
2007	1762.9	2785.3	410335	440.2	1187.3	106.0181	8.9296

2008	1706.0	3018.4	442525	578.4	1542.3	148.1514	9.6059
2009	1702.6	3401.3	484379	641.6	1877.7	192.4199	10.2474
2010	1703.4	3736.3	557278	755.6	2253.3	338.0638	15.0033
2011	1645.5	4097.8	701381	997.5	2794.1	355.98	12.7405
2012	1605.8	4549.3	643269	1163.2	3171.5	430.3941	13.5706
2013	1538.3	4848.7	669533	1277.4	3369.2	461.6956	13.7035
2014	1471.3	5155.5	695625	1301.3	3434.2	487.667	14.2002
2015	1395.0	5442.7	725812	1165.9	4020.7	681.48	16.9495
2016	1347.5	5634.3	774675	1148.4	4227.3	801.77	18.9663
2017	1295.0	5813.8	797667	1243.3	4641.1	815.1611	17.564
2018	1215.7	6082.4	828042	1282.5	4675.7	834.4713	17.843
2019	1151.6	5273.5	856368	1262.8	5011.6	881.9893	17.5991
2020	1090.5	6775.1	897225	1152.5	5449.4	914.53	16.7822
2021	1072.2	6888.4	939945	1300.5	5104.8	868.9781	17.0227
2022	1047.0	7090.8	967408	1290.7	5452.0	938	17.197252

如图 22 为这 21 个因子之间的相关性矩阵,可以看到,多个因子之间存在较强的相关性,为避免这种相关性对结果的影响,首先对因子进行主成分分析。主成分分析是一种多元统计分析方法,可以消除影响因素彼此的相关影响,减少评价指标选择的工作量。进而回归分析可以准确地计量各个影响因素之间的相关程度与回归拟合程度的高低,提高预测方程式的效果。

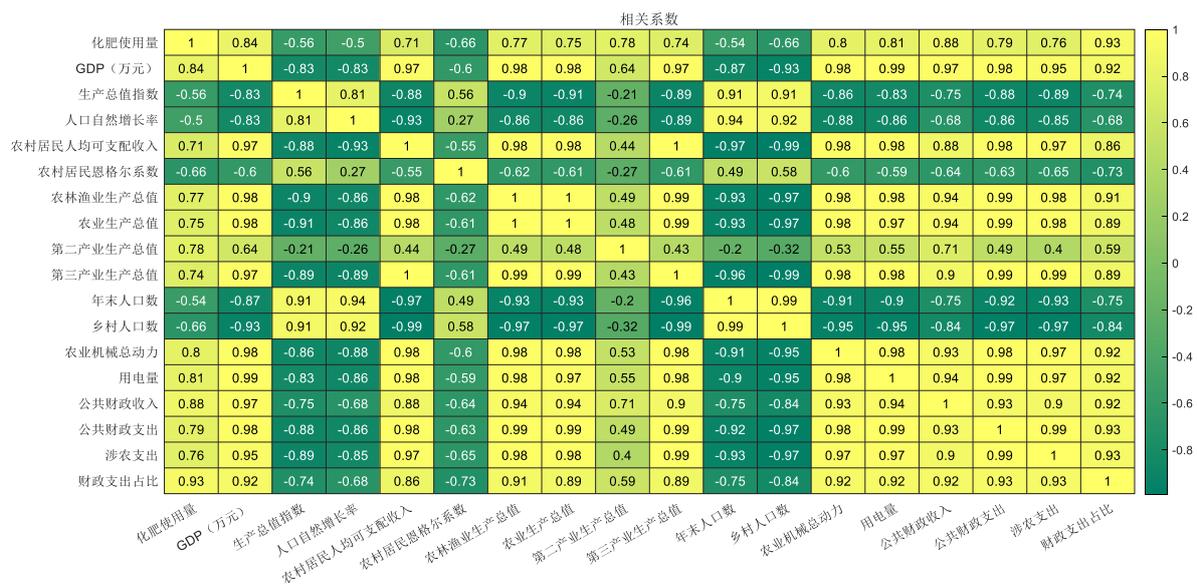


图22 因子相关性分析结果

因此,本文考虑采用主成分分析法确定影响因素主成分,并进行回归分析。主成分分析基本原理如下: 根据 min-max 标准化公式,将原始样本矩阵标准化变为:

$$X = (X_1, X_2, \dots, X_p)$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p1} & \cdots & r_{pp} \end{bmatrix}$$

其中,  $p$  表示数据集的维度,  $R$  为半正定矩阵,

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) = \frac{1}{n-1} \sum_{k=1}^n X_{ki} X_{kj}, \quad i=1,2,\dots,p, \quad j=1,2,\dots,p。$$

计算  $R$  的特征值  $\lambda_m \geq 0, m=1,2,\dots,p$ , 以及特征向量  $a_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{pn} \end{bmatrix}, n=1,2,\dots,p。$

从原始数据中提取的主成分可以表示为指标的线性组合:

$$F_q = a_{1q}X_1 + a_{2q}X_2 + \dots + a_{pq}X_p, q=1,2,\dots,m$$

对于某个主成分而言,指标前面的系数越大,说明该指标对于主成分的影响越大。根据  $p$  个特征值计算各主成分的贡献率以及累计贡献率分别为:

$$\text{贡献率} = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k}, i=1,2,\dots,p$$

$$\text{累计贡献率} = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k}, i=1,2,\dots,p$$

取累计贡献率超过 90% 的特征值对应的第一、第二、...、第  $m(m \leq p)$  个主成分。使用主成分分析的前提条件是原始数据各个变量之间应有较强的线性相关关系。如果原始变量之间的线性相关程度很小,它们之间不存在简化的数据结构,这时进行主成分分析实际是没有意义的。所以,应用主成分分析时,首先要对其适用性进行统计检验。检验方法有巴特莱特球形检验,KMO 检验。下面对数据相关性进行检验。从相关系数矩阵可以看出,多个变量之间的相关性较强。接下来对数据进行 KMO 检验和巴特利特球形检验。

KMO(Kaiser-Meyer-Olkin-Measure of Sampling Adequacy)检验是从比较原始变量之间的简单相关系数和偏相关系数的相对大小出发来进行的检验。一般来说,KMO 值在 0.8 以上适合做主成分分析,在实证分析中能达到 0.6 以上,就认为可以进行主成分分析。巴特利特球形检验法的统计量是根据相关系数矩阵的行列式得到的。如果该值较大,且其对应的相伴概率值小于指定的显著水平时,拒绝零假设,表明相关系数矩阵不是单位阵,原有变量之间存在相关性,适合进行主成分分析。数据进行检验的结果如下表:

表12 KMO 和巴特利特检验结果

KMO 取样適切性量数。		.681
巴特利特球形度检验	近似卡方	1112.324
	自由度	153
	显著性	.000

根据检验结果,其 KMO 检验值为 0.681,大于 0.5,巴特利特球形度检验小于 0.001,满足主成分分析条件,可以对数据进行主成分分析。接下来,根据主成分分析步骤,计算特征值以及特征向量并提取主成分,将特征值进行排序并获取累积贡献度,合并入特征向量得到:

	eig_value	eig_cum	0	1	2	3	4	5	6	7	8	9	10	11
0	13.723486	0.857718	-0.265198	0.241889	0.237249	-0.267472	0.167175	-0.269098	-0.268193	-0.130506	-0.268812	0.253603	0.264196	-0.266802
1	1.195217	0.932419	0.157197	0.241022	0.286972	-0.074128	-0.190293	0.017641	0.008803	0.751049	-0.053885	0.295142	0.173142	0.048870
2	0.759829	0.979908	-0.074890	-0.121303	0.352703	-0.089072	-0.853640	0.008601	0.006538	-0.318680	0.002786	0.052907	-0.005031	-0.047779
3	0.160209	0.989921	-0.013174	-0.811388	0.283588	-0.138806	0.248047	0.120744	0.191119	0.058738	-0.125276	0.117851	0.110121	-0.053661
4	0.079944	0.994918	-0.016221	0.336592	0.451083	-0.049759	0.342855	0.123507	0.125400	-0.399743	0.076986	0.053485	0.050398	-0.106079
5	0.037700	0.997274	0.139056	0.138450	0.224989	0.231920	-0.019437	0.068048	0.243093	0.086304	0.145690	-0.453834	-0.247524	-0.460203
6	0.023832	0.998764	-0.057664	0.208352	-0.246537	-0.052495	-0.065881	0.215175	0.339915	-0.215787	-0.126981	0.070200	0.163032	0.512763
7	0.007098	0.999207	0.071465	0.078618	0.426926	0.067539	0.114347	-0.015673	-0.103193	0.131248	0.101599	-0.219419	-0.133132	0.497393
8	0.006446	0.999610	-0.107016	-0.073023	0.338821	0.004864	0.062568	-0.191457	-0.129988	-0.083135	-0.065256	-0.191720	-0.153783	0.417491
9	0.003201	0.999810	0.100643	-0.111887	0.049178	0.128991	0.019812	-0.323653	-0.054099	-0.024856	0.342167	-0.109185	0.338443	0.108494
10	0.001886	0.999928	0.012857	-0.086916	-0.139102	0.102249	0.000561	-0.377424	-0.451328	-0.067706	0.221562	0.222385	-0.375470	0.022251
11	0.000692	0.999971	-0.181951	0.021430	0.001891	-0.207371	0.018947	0.204600	-0.112495	0.104654	-0.419913	-0.040494	-0.579957	0.017186
14	0.000228	0.999986	0.121564	-0.027938	0.126152	0.576115	0.034989	0.533122	-0.339561	-0.120548	0.017255	0.391004	0.017494	0.006259
15	0.000218	0.999999	-0.042236	0.032160	0.082952	0.254422	0.042532	-0.369034	0.581110	-0.024171	0.036415	0.516357	-0.344640	0.026588
13	0.000011	1.000000	-0.488206	0.009456	0.003896	-0.328861	0.030391	0.306812	0.018929	0.120758	0.693992	0.139261	-0.123382	0.004008
12	0.000003	1.000000	0.749257	0.022588	0.003703	-0.507221	0.040577	0.077040	-0.029846	-0.173462	0.159794	0.176960	-0.173587	0.016672

图23 特征值计算结果

根据特征值可以看出,两个主成分的累积贡献度可以到达 93.24%,因此选择两个主成分,得到主成分因子载荷系数如表 13:

表13 因子载荷值

变量名称	主成分 1	主成分 2	变量名称	主成分 1	主成分 2
化肥使用量	-0.20745	-0.43248	第三产业生产总值	-0.25431	0.088817
GDP (万元)	-0.25329	-0.08558	年末人口数	0.236064	-0.30389
生产总值指数	0.226131	-0.24235	乡村人口数	0.248238	-0.19147
人口自然增长率	0.220379	-0.30693	农业机械总动力	-0.25396	-0.0019
农村居民人均可支配收入	-0.25234	0.114689	用电量	-0.2539	-0.01919
农村居民恩格尔系数	0.164034	0.206599	公共财政收入	-0.24095	-0.22717
农林渔业生产总值	-0.25532	0.030036	公共财政支出	-0.25544	0.021355
农业生产总值	-0.254	0.042556	涉农支出	-0.25297	0.066437
第二产业生产总值	-0.1325	-0.59802	财政支出占比	-0.23871	-0.20682

得到两个主成分的表达式为:

$$comp1 = -0.207x_1 - 0.253x_2 + 0.226x_3 + 0.220x_4 - 0.252x_5 + 0.164x_6 - 0.255x_7 - 0.254x_8 - 0.132x_9 - 0.254x_{10} + 0.236x_{11} - 0.248x_{12} - 0.253x_{13} - 0.253x_{14} - 0.240x_{15} - 0.255x_{16} - 0.253x_{17} - 0.239x_{18}$$

$$comp2 = -0.432x_1 - 0.085x_2 - 0.242x_3 - 0.307x_4 + 0.114x_5 + 0.206x_6 + 0.03x_7 + 0.04x_8 - 0.598x_9 + 0.089x_{10} - 0.304x_{11} - 0.191x_{12} - 0.001x_{13} - 0.019x_{14} - 0.227x_{15} + 0.021x_{16} + 0.066x_{17} - 0.207x_{18}$$

可以看出,在第一主成分中,各因素对第一主成分影响较大,第二产业生产总值对第二主成分影响较大。

#### 4.2.3 基于主成分分析的线性回归和非线性回归

回归分析包含线性回归与非线性回归,本节根据回归模型预测粮食作物耕地面积,利用统计方法建立因变量和自变量的关系,根据提供的自变量,研究因变量的变化,可以用于预计分析。本文在上面研究了影响因素的主成分,以主成分 Comp1、Comp2 为自变量,粮食作物耕地面积 y 量为因变量,进行多元线性回归分析后结果如下图所示:

```

Call:
lm(formula = y ~ comp1 + comp2, data = my_data)

Residuals:
    Min       1Q   Median       3Q      Max
-78.335 -35.063   5.322  40.355  63.265

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1260.831    10.708  117.752 < 2e-16 ***
comp1       -52.289     2.965  -17.633 2.32e-12 ***
comp2        65.934    10.049   6.562 4.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47.89 on 17 degrees of freedom
Multiple R-squared:  0.9542,    Adjusted R-squared:  0.9488
F-statistic: 177 on 2 and 17 DF,  p-value: 4.164e-12

```

图24 线性回归结果

得到线性回归的表达式为

$$y = -52.289Comp1 + 65.934Comp2 + 1260.831$$

此时,R-squared 的值为 0.9542,表示线性关系显著。且常数项和非常数项的 p 值均较小,因此可以使用最终表达式。

进一步地,为选择精度更高的模型,以主成分 Comp1、Comp2 为自变量,耕地面积 y 为因变量,进行多元非线性回归分析后结果如下图,自变量包括常数项、comp1、comp2,comp1<sup>2</sup>、comp2<sup>2</sup>、comp1\*comp2 可以看出解释变量均显著,可得非线性回归系数为:

```

Call:
lm(formula = y ~ comp1 * comp2 + comp1 + comp2 + I(comp1^2) +
    I(comp2^2), data = my_data)

Residuals:
    Min       1Q   Median       3Q      Max
-35.195 -12.712   2.092  14.898  24.843

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1334.2074    13.2354  100.806 < 2e-16 ***
comp1       -44.7767     1.7008  -26.327 2.52e-13 ***
comp2        28.4854     8.8258   3.228 0.006079 **
I(comp1^2)   -4.0842     0.9256  -4.413 0.000590 ***
I(comp2^2)  -17.7282     3.8733  -4.577 0.000431 ***
comp1:comp2   9.3057     1.6767   5.550 7.16e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.88 on 14 degrees of freedom
Multiple R-squared:  0.9935,    Adjusted R-squared:  0.9912
F-statistic: 427.8 on 5 and 14 DF,  p-value: 8.683e-15

```

图25 非线性回归结果

可以看到,每一项均显著,且 R-squared 的值大于 0.99,回归方程表达式为:

$$y = 1334.20 - 44.77Comp1 + 28.48Comp2 - 4.08Comp1^2 - 17.72Comp2^2$$

此时 RMSE 的值为 39.42。线性回归方程的 RMSE 的值为 44.14,相比之下,非线性回归方法效果更好。对两模型进行比较,可以得到:

Analysis of Variance Table							
Model 1: y ~ comp1 + comp2							
Model 2: y ~ comp1 * comp2 + comp1 + comp2 + I(comp1^2) + I(comp2^2)							
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)		
1	17	38982					
2	14	5532	3	33450	28.219	3.407e-06	***
---							
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

图26 线性回归模型和非线性回归模型对比

两者的残差平方和在非线性模型中为 38982,在线性模型中为 5532,。两者进行 F 检验,p 小于 0.05,即两者残差平方和具有统计学差异。综上,更好的模型为非线性模型。

#### 4.2.4 模型检验

接着对对回归结果进行分析,首先对残差进行分析,进行残差正态性检验。

##### (1) 残差正态性检验

在理想模型中有五大假设其中之一便是残差应该是一个正态分布,与估计值无关。如果残差还和估计值有关,那就说明模型仍然有值得去改进的地方,当然,这样的模型准确程度也就大打折扣。下图分别为线性模型和非线性模型中残差与估计值之间的关系图,Residuals 即为残差的意思 (估计值与真实值之差)。可以看到残差与估计值无关。

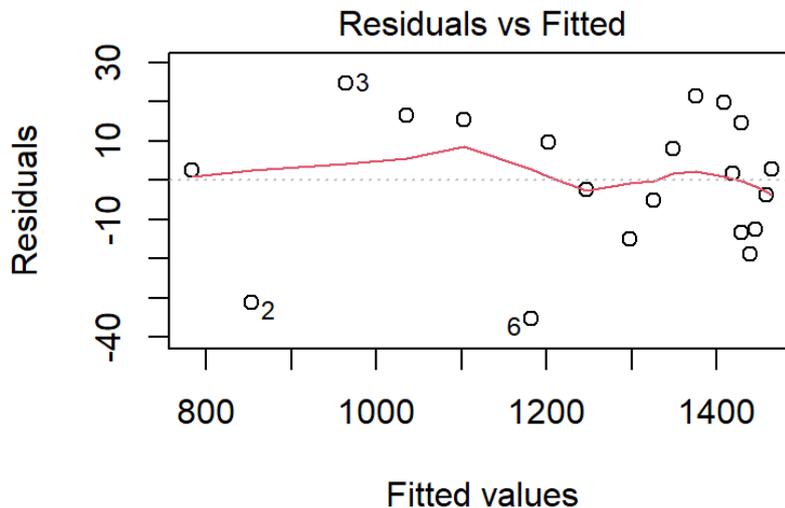


图27 残差正态性检验

进一步使用 QQ 图对模型估计残差进行正态性检验,下图分别为线性模型与非线性模型的残差 QQ 图。

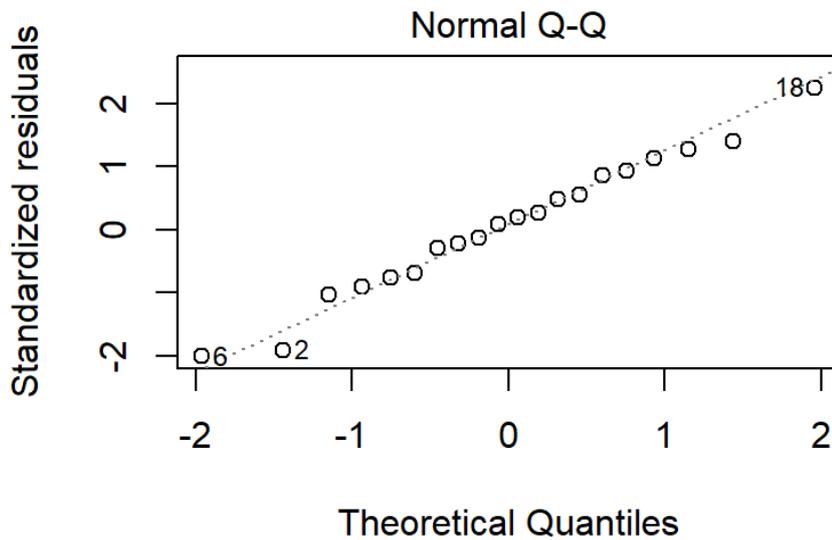


图28 QQ 图

可以看到,非线性模型残差均大致服从正态分布。

#### (2)方差齐性检验

在进行残差正态性检验后,对模型估计值进行方差齐性检验,假设预测的模型里方差是一个定值。如果方差不是一个定值那么这个模型的可靠性也是大打折扣的。下图为非线性模型的方差图:

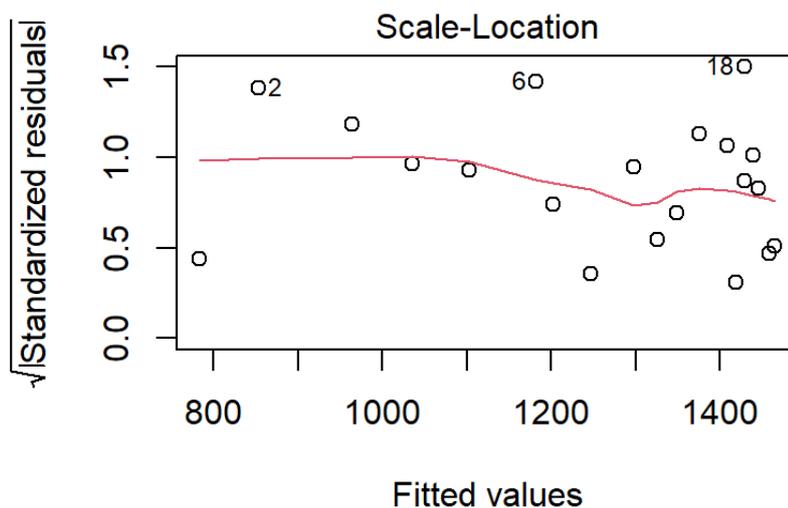


图29 方差齐性检验

从图中可以看出,其残差大致恒定,进一步的,可以对残差进行BP检验(Breusch-Pagan Test),对非线性模型进行检验,得到检验结果为:

```
studentized Breusch-Pagan test
data: model2
BP = 3.6153, df = 5, p-value = 0.606
```

图30 方差齐性检验计算结果

检验统计量分别为 3.6153,对应的 p 值为 0.606。由于 p 值不小于 0.05,因此无法否定原假设。我们没有足够的证据表明回归模型中存在异方差性,检验通过。

#### 4.2.5 耕地面积变化原因及相关政策分析

基于主成分分析结果和非线性模型结果,可以得到最终具有负向影响极正向影响的因子,其中具有负向影响的因子包括:

1、人口自然增长率,人口数等人口因素也就是说随着人口的增长,可能带来的工业用地,建设用地等其他用地的扩张,耕地面积会随之下降。

##### 2、第二产业生产总值

第二产业生产总值为工业生产总值,工业生产总值的增长势必造成工业用地的增长,这也是造成耕地面积减少的因素之一。

具有较大正向影响的因子包括:

##### 1、GDP,农业生产总值,第三产业生产总值

值得注意的是,生产总值中第三产业生产总值的增长与全部生产总值的增长对应,第三产业为服务业,与工业不同,服务业对于耕地的需求明显弱于工业生产,因此服务业的增长,刺激了经济的发展,使得对于工业的发展需求减弱,工业用地减少,也会带来耕地的增长。

##### 2、农业机械总动力,农村用电量以及相应的财政支出

农村用电量,农村机械总动力以及相应财政支出,均为耕地面积增长的促进因素,这表明,耕地面积的维持需要相关的投入,相关投入的增加使得有更多的财力精力对耕地进行维护。

##### 3、农村居民人均可支配收入

农村居民人均可支配收入的增长,使得农民不需要放弃耕地寻找其他工作,因此对于农民收入的支持,也会带来耕地面积的增长。

现代经济的发展往往以牺牲耕地为代价,事实上,大多数省份的耕地面积逐年下降,有文献指出,1978 年以来,我国总体耕地面积下降明显,尤其东南地区肥沃的土地急剧减少。其中 1986 至 2000 年间耕地转为草地和林地现象占耕地减少的 47%。有学者利用江苏省昆山市的数据分析得出,工业化、城市化、人口增长和国家经济改革措施是土地利用变化的四个主要驱动力,后又有研究进一步证实了城市化对耕地“非农化”的影响。也正因如此,2020 年,国务院办公厅连续发布《关于坚决制止耕地“非农化”行为的通知》《关于防止耕地“非粮化”稳定粮食生产的意见》,明确提出,要采取有力举措坚决遏制耕地“非农化”、防止“非粮化”,切实稳定粮食生产,牢牢守住国家粮食安全的生命线。黑龙江省成功实现 GDP 与耕地面积同步增长,证明经济的发展并不一定要以牺牲耕地为代价,可以尝试其他发展方向尤其是服务业的发展。

2021 年 12 月 29 日,中共黑龙江省委举行新闻发布会,解读省委十二届十一次全会暨省委经济工作会议精神。发布会上,省农业农村厅副厅长孙化库介绍,2021 年,全省粮食生产夺得“十八连丰”,粮食播种面积、粮食总产量均创历史新高。落实会议精神,要重点采取

六项措施。其中一个重要措施就是稳面积。严格遏制耕地“非农化”、防止“非粮化”,明确耕地利用优先序,严禁耕地撂荒,种满种严,确保全省粮食作物播种面积稳定在 2.18 亿亩以上。

高效的科技使用,也是保障粮食生产的途径之一,“推广大马力高端智能农业机械,集成组装技术模式,不断提高单产,增加总产”也是黑龙江实施的政策之一。除此之外,保障农民的生产生活,提升农民生活水平,使得农民没有后顾之忧,也是保证粮食生产的重要途径。

### 4.3 耕地质量变化定量描述

对耕地质量进行评价时,需要考虑多种多方面综合因素并选择合适的评价指标。选择合适的指标需要确保指标的代表性和主导性,使其可以充分代表各土壤质量各方面特征;另外还需要考虑指标的多样性,保证涵盖耕地质量各方面特征。因此,结合所调研数据,确定评价指标,构建土壤质量综合评价指标体系。

#### 4.3.1 基于 PSR 模型的耕地质量评价体系构建

本文选择 PSR 模型对耕地质量综合评价指标体系进行量化评价。PSR 模型用于耕地质量评价体系,P (Production Press Index,PPI) 代表生产压力指数,即不同耕地措施对耕地产生压力影响,S (Land Status Index, LSI) 代表耕地现状条件,即耕地状态对耕地质量的影响程度,R (Social Response Index, SRI) 代表社会经济响应,即耕地所带来的经济效益。

《中国土壤分类》(GB/T 17296-2009) 中耕地质量是指土壤的物理性质、化学性质和生物性质,以及土地的土地利用和管理状况对农业生产和环境保护的影响程度。结合国土部《农用地质量分等规程》、农业部《耕地质量等级》、《耕地地力调查与质量评价技术规程》、以及《耕地质量监测技术规程》等相关资料,问题二认为耕地质量可从耕地压力、耕地理化条件、社会响应状态等方面指标进行综合评价。

考虑耕地质量评价指标的多样性、代表性、科学性等原则,构建耕地质量综合评价指标体系如图 31 所示。分为耕地压力 (P)、耕地理化条件 (S)、社会响应状态 (R) 三个方面。耕地压力包括土壤侵蚀速率、土地退化指数等;耕地理化条件包括土壤有机质、氮等含量、PH 值、土壤质地(砂粒、黏粒、粉粒)等;社会响应状态为土壤灌溉面积、土壤利用综合指数等。其中,土壤有机质、总氮量、土壤砂粒、黏粒、粉粒容量、土壤 PH 值等数据主要来源于于国家地球系统科学数据中心 (<https://www.geodata.cn>); 土壤灌溉面积数据来源于《黑龙江统计年鉴》;土壤侵蚀速率、土地退化指数及土壤利用综合指数等数据来源已发表文章。

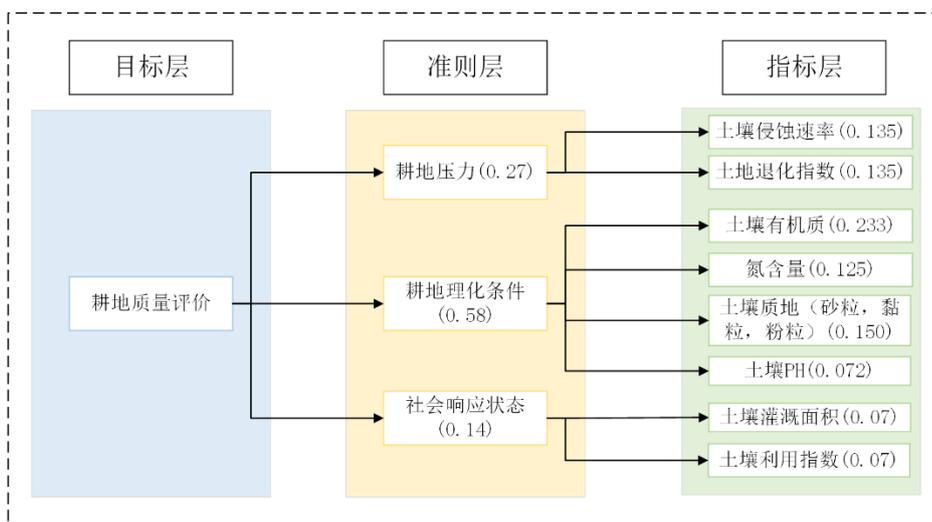


图31 耕地质量综合评价指标体系

使用 AHP 法对所构建指标体系进行权重计算,以耕地质量为目标层,耕地压力、耕地理化条件、社会响应状态三类为准则层,各类细化指标为指标层。确定各因素间相对重要性并构建判断矩阵,计算各层次权重值,综合评价权重结果已在图 31 中展示。

### 4.3.2 耕地质量评价分析结果

研究文献发现,描述耕地质量的土壤有机质、含氮量、PH 值、土壤质地等数据往往需要实地监测,获取黑龙江全省范围内土壤理化性质数据较为困难,且缺失常年累月的历史数据。因此,结合所收集有关数据,本节以黑龙江省海伦市为代表,研究耕地质量评价结果及时空变化。

海伦市位于黑龙江省中部,如图 32,地处松嫩平原与小兴安岭余脉交接过渡带的黑土核心地区,是黑龙江省重要粮食产地。全市占地面积 4667km<sup>2</sup>,其中耕地面积 3100km<sup>2</sup>,占土地总面积的 63.42%,约 2727km<sup>2</sup>的耕地主要土壤类型是黑土。

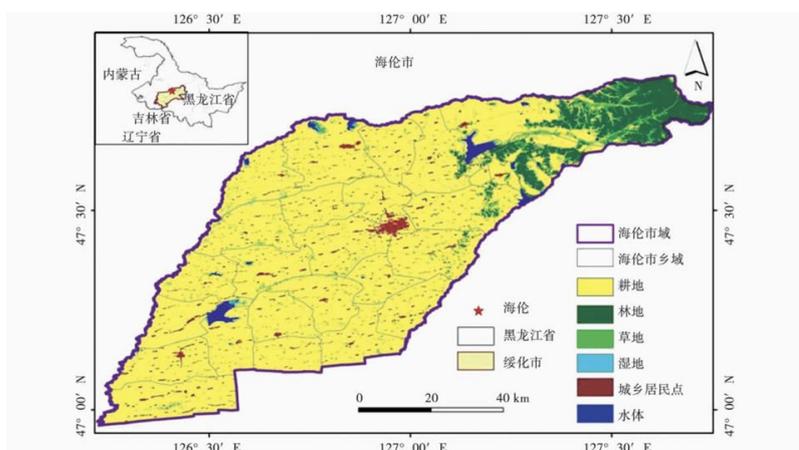


图32 海伦市地理位置说明

为了便于描述土壤质量评价指标,根据农业农村部出台的《耕地质量等级》国家标准 (GB/T33469-2016),通过对地块大规模野外实测调查历史数据整理,结合制定的指标体系评估耕地质量等级。绘制 1985-2020 年海伦市耕地质量等级占比时空变化图如图 33。

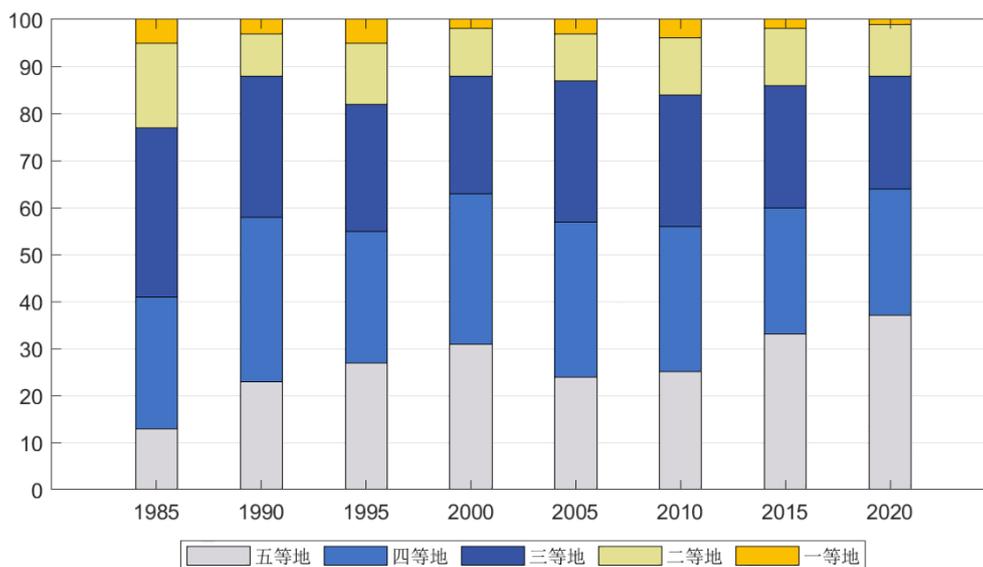


图33 耕地质量等级占比变化

从图中看,随年份变化海伦市耕地质量等级占比发生明显变化,五级耕地面积明显升高,一、二、三级耕地占比逐年下降。1985-2010 间,一级和二级耕地呈现波动趋势,2010-2020 年呈现明显下降趋势;三级耕地从 2005 年以后呈逐年下降趋势;四级耕地整体变化较为缓慢。这说明海伦市整体耕地质量呈下降趋势。

结合历史分析原因可知,我国自上世纪 50 年代以来不断开发利用黑土资源,由于早期环境保护意识与可持续发展意识较弱,对黑土地的过度利用造成了如今耕地质量不断下降的结果。

### 4.3.3 耕地质量变化原因及相关政策分析

随着耕地利用强度加大,农户管理意识与养分均衡观念不足,可持续利用意识差,引起土壤板结、土壤营养结构失衡等问题,导致耕地质量下降,土地退化,五级耕地占比增加。

2020 年国务院办公厅连续发布《关于坚决制止耕地“非农化”行为的通知》《关于防止耕地“非粮化”稳定粮食生产的意见》,提出,要采取有力举措坚决遏制耕地“非农化”、防止“非粮化”,坚持科学合理利用耕地资源。随后,黑龙江省人民政府印发黑龙江省防止耕地“非粮化”稳定粮食生产工作方案通知,明确指出加强耕地保护修复,稳步提升耕地质量,大力开展测土配方施肥和有机肥替代化肥试点示范,改善土壤理化性质,优化养分结构,提升耕地质量。

## 5 问题三的分析与模型构建

### 5.1 问题分析

近年来,全球粮食种子产业面临着种子“退化”、“绝育”等挑战,而中国作为全球重要的粮食生产国,其粮食种子的自给率受到广泛关注。农作物的播种面积是种子市场需求的直接来源,而我国种业品种分布集中,主要由玉米、水稻、小麦、马铃薯、大豆等粮食作物种子组成,这些作物种子的市场需求量大,是种子行业的主要收入来源。本章旨在从数量和质量的角度的角度,利用粮食播种面积与产量分析近 20 年来我国粮食种子的自给率变化,并结合粮食种业的现状、政策法规与国际环境,探讨“种子按斤称到论粒卖”销售模式背后的原因与理论逻辑。

### 5.2 模型构建

#### 5.2.1 粮食自给率测度

粮食自给率是国内粮食产量占国内粮食消费量的百分比。粮食净进口率,又称为粮食进口依存率,是指粮食净进口量占国内粮食消费量的百分比。粮食自给率受粮食产量定义和粮食净进口计算方法的影响。

$$\text{粮食自给率} = \frac{\text{国内粮食产量}}{\text{国内粮食产量} + \text{进口量} - \text{出口量}} \times 100\%$$

为计算对应粮食自给率,我们收集了从 2000 年至 2020 年来中国粮食主要作物(如小麦、水稻、玉米等)的种子生产和进出口数据进行计算。

表14 2000-2020 年中国粮食进出口量情况(单位:万吨)

年份	出口量			进口量			净进口量		
	小麦	玉米	稻谷	小麦	玉米	稻谷	小麦	玉米	稻谷
2000	18.8	1047.9	296.2	91.9	0.3	24.9	73.1	-1047.6	-271.3

年份	出口量			进口量			净进口量		
	小麦	玉米	稻谷	小麦	玉米	稻谷	小麦	玉米	稻谷
2001	71.3	600	187	73.9	3.9	29.3	2.6	-596.1	-157.7
2002	97.7	1167.5	199	63.2	0.8	23.8	-34.5	-1166.7	-175.2
2003	251.4	1639.1	261.7	44.7	0.1	25.9	-206.7	-1639	-235.8
2004	108.9	232.4	90.9	725.8	0.2	76.6	616.9	-232.2	-14.3
2005	60.5	864.2	68.6	353.9	0.4	52.2	293.4	-863.8	-16.4
2006	151	309.9	125.3	61.3	6.5	73	-89.7	-303.4	-52.3
2007	307.3	491.8	134.3	10.1	3.5	48.7	-297.2	-488.3	-85.6
2008	31	27.3	97.2	4.3	5	33	-26.7	-22.3	-64.2
2009	24.5	13	78.6	90.4	8.4	35.7	65.9	-4.6	-42.9
2010	27.7	12.7	62.2	123.1	157.3	38.8	95.4	144.6	-23.4
2011	32.8	13.6	51.6	125.8	175.4	59.8	93	161.8	8.2
2012	28.6	25.7	27.9	370.1	520.8	236.9	341.5	495.1	209
2013	27.8	7.8	47.8	553.5	326.6	227.1	525.7	318.8	179.3
2014	19	2	41.9	300.4	259.9	257.9	281.4	257.9	216
2015	12.2	1.1	28.7	300.7	473	337.7	288.5	471.9	309
2016	11.3	0.4	39.5	341.2	316.8	356.2	329.9	316.4	316.7
2017	18.3	8.6	119.7	442.2	282.7	402.6	423.9	274.1	282.9
2018	28.6	1.2	209.1	309.9	352.4	307.7	281.3	351.2	98.6
2019	31.3	2.6	274.8	348.8	479.3	254.6	317.5	476.7	-20.2
2020	18.1	0.3	230.5	837.6	1129.6	294.3	819.5	1129.3	63.8

同时结合前面汇总的三种粮食作物的产量信息,对从 2000-2020 年的粮食自给率进行计算:

表15 2000-2020 年中国粮食产量情况 (单位: 万吨)

年份	稻谷	小麦	玉米	年份	稻谷	小麦	玉米
2000	18790.8	9963.6	10600	2011	20288.3	11857	21131.6
2001	17758	9387.3	11408.8	2012	20653.2	12247.5	22955.9
2002	17453.9	9029	12130.8	2013	20628.6	12363.9	24845.3
2003	16065.6	8648.8	11583	2014	20960.9	12823.5	24976.4
2004	17908.8	9195.2	13028.7	2015	21214.2	13255.5	26499.2
2005	18058.8	9744.5	13936.5	2016	21109.4	13318.8	26361.3
2006	18171.8	10846.6	15160.3	2017	21367.6	13424.1	25907.1
2007	18638.1	10949.2	15512.3	2018	21212.9	13144	25717.4
2008	19261.2	11290.1	17212	2019	20961.4	13359.6	26077.9
2009	19619.7	11579.6	17325.9	2020	21186	134425.4	26066.5
2010	19722.6	11609.3	19075.2				

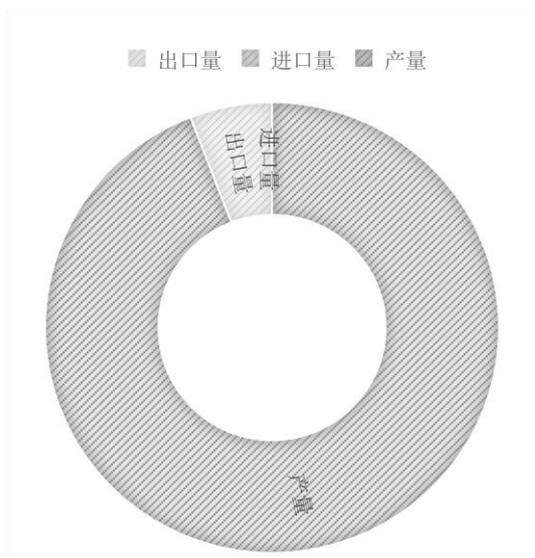
计算粮食的自给率,即国内的粮食产量与总需求量的比例。

表16 2000-2020 年中国粮食自给率

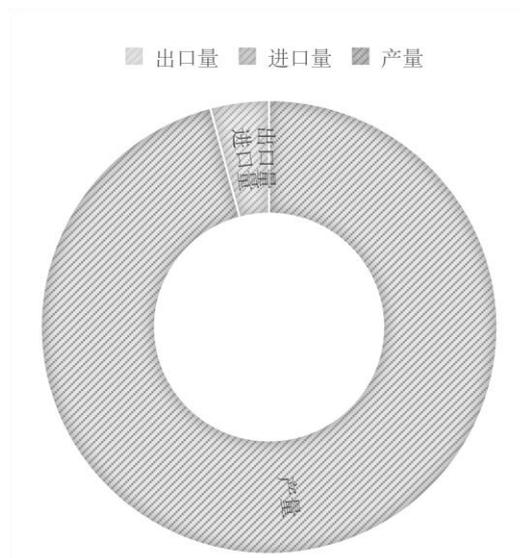
年份	粮食自给率 (百分比)			
	小麦	玉米	稻谷	全部谷物
2000	99.27	110.96	101.46	102.69
2001	99.97	105.51	100.89	101.36
2002	100.38	110.64	101.01	103.10

2003	102.44	116.48	101.48	105.62
2004	93.71	101.81	100.07	98.80
2005	97.07	106.60	100.09	100.81
2006	100.83	102.04	100.28	100.50
2007	102.79	103.25	100.46	101.68
2008	100.23	100.12	100.33	100.05
2009	99.43	100.02	100.21	99.67
2010	99.18	99.24	100.11	99.20
2011	99.22	99.24	99.95	99.28
2012	97.28	97.88	98.99	97.92
2013	95.92	98.73	99.13	97.89
2014	97.85	98.97	98.98	97.15
2015	97.86	98.25	98.56	95.35
2016	97.58	98.81	98.52	96.86
2017	96.93	98.95	98.69	96.50
2018	97.90	98.65	99.53	97.34
2019	97.67	98.20	100.09	97.83
2020	99.39	95.84	99.69	95.27

由于粮食种子主要用于粮食的生长,因此粮食的产量可以侧面反映粮食种子的数量与质量,因此我们选用粮食的自给率代替粮食种子的自给率,基于以上数据可以得到对应粮食自给率的变化特点。



a. 2005年玉米进出口与产量



b. 2020年玉米进出口与产量

图34 粮食种子的自给率变化对比

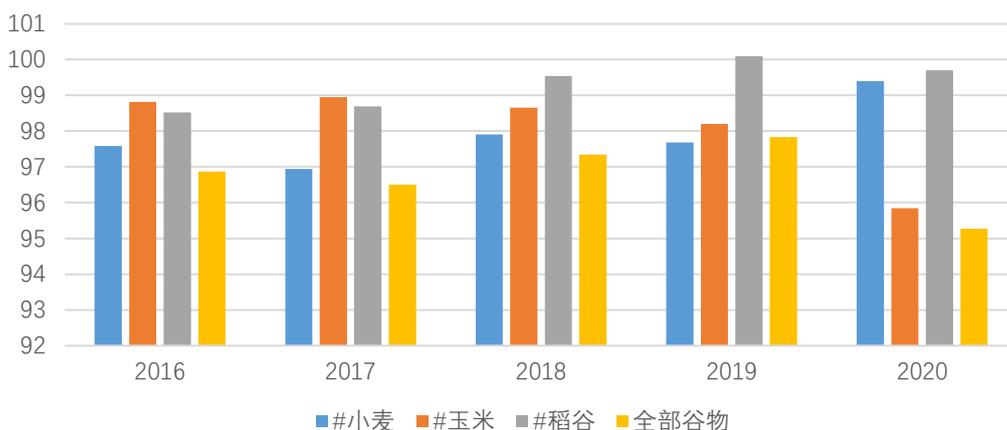


图35 2016-2020 年粮食自给率柱状图

### 5.2.2 粮食进出口与国家政策

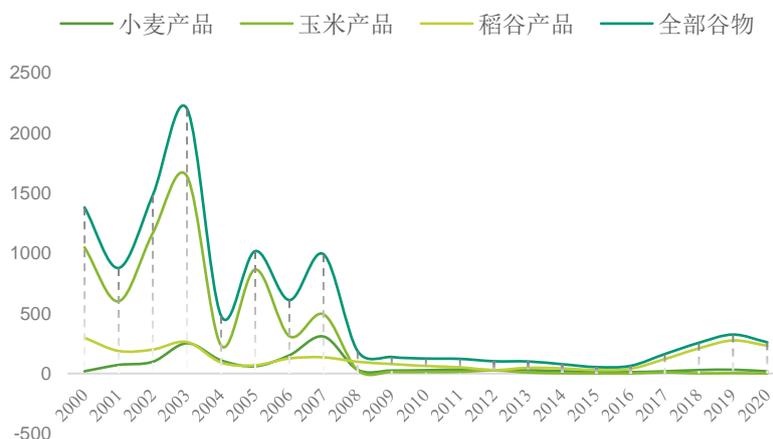


图36 中国 2000-2020 年粮食出口量变化趋势



图37 中国 2000-2020 年粮食进口量变化趋势

可以发现我国的粮食出口量总体呈逐年下降,而进口量逐年上升的趋势,而粮食自给率方面小麦、稻谷逐渐稳步提高,玉米自给率有所下降。

另外,国家政策对于粮食种子的质量也具有一定影响,因此还需要对国家种业相关政策进行调研整理,根据题目要求,我们对国家种子相关政策进行调研,结合粮食进出口数量的变化时间拐点,汇总可能影响数量、质量的法规如下:

表17 主要种子政策法规发展

时间	颁布机构	事件及政策法规	主要内容及意义
2004.12	中共中央、国务院	《中共中央国务院关于进一步加强农村工作提高农业综合生产能力若干政策的意见》	实施良种补贴,强调提高农民种植积极性。
2015.05	农业部	《关于加快推进种业“事企脱钩”工作的通知》	明确公益性的科研院所和高等院校所办的种子企业“事企脱钩”的标准,即“产权清晰”“股权多元”“权责明确”。
2015.07	国务院	《关于加快转变农业发展方式的意见》	深化种业体制改革,在完善种业科研成果权益分配,健全种业科技资源、人才向企业流动机制,做大做强育繁推一体化种子企业。
2015.12	中共中央、国务院	《关于落实发展新理念加快农业现代化实现全面小康目标的若干意见》	指出要大力推进“育繁推一体化”,加快现代种业发展,提升种业自主创新能力,保障国家种业安全;加快推进育种制种基地和良种繁育基地建设;强化企业育种创新主体地位,加快培训具有国际竞争力的现代种业企业。
2019.01	中共中央、国务院	《中共中央国务院关于坚持农业农村优先发展做好“三农”工作的若干意见》	要求加快农业关键核心技术的攻关,推动生物种业、智慧农业领域的自主创新。继续组织主要农作物良种选育联合攻关,推进新品种新技术示范推广。
2019.06	国务院	《关于促进乡村产业振兴的指导意见》	支持种业育繁推一体化,培育一批竞争力强的大型种业企业集团。
2019.12	国务院	《关于加强农业种质资源保护与利用的意见》	到 2035 年种质资源保护建设目标:建成系统完整、科学高校的农业种质资源保护与利用体系,提升重要种质资源的收集、保护、评价与利用水平,资源保存总量位居世界前列、资源创新利用达到国际先进水平。

结合数据进行分析,2004 年粮食出口量大幅度降低,进口量大幅度提升,自给率显著下降,因此在年末出台相关提升农民种植积极性的相关法案,加强农村工作提高农业综合生产能力。像斯蒂,2015,2019 年分别为进口量的拐点,都是当年颁布一系列措施促进种业创新、核心技术攻关,促进了种业发展与粮食产量双丰收。同时可以发现,政策法规的制定具有一定的滞后性,作用发挥也无法实时展现,这对于决策者法规的制定具有一定挑战性。

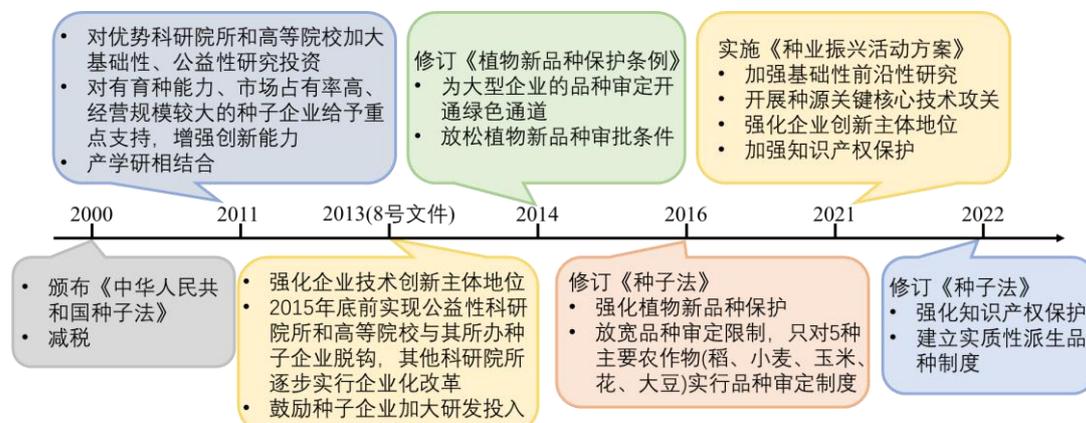


图38 2000-2022 年中国种业相关政策措施

另外,2020 年的新冠肺炎疫情席卷全国,并正值农民春播农忙之际,由于疫情对于市场、交通、人员等等的巨大影响,对种子的加工、运输、购种都有较大影响。而电商的逐渐兴起,也给种业带来了新的路径,国内人口数量逐渐增加与耕地面积有限的矛盾使得农产品的需求仍为刚性,种业的发展仍是充满动力。

### 5.2.3 国际种业环境影响

国际种业环境也是对国内影响的重要因素之一,其中一个重要反应指标就是国际粮价的时序变化,因此将对应的粮食价格进行汇总对比,如下图:

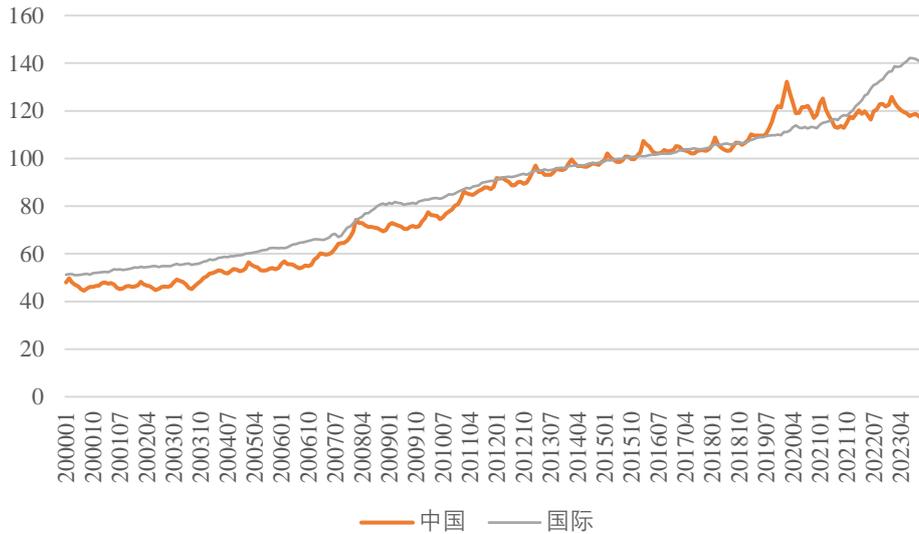


图39 2000-2023 年中国与国际粮价变化曲线

从数据中可以发现,中国的粮食价格大多时间低于国际的粮价,而在 2020 年 1 月至 2021 年 4 月高于国际粮价,正是疫情期间,可见疫情对于国内的粮食产业影响较大。而对比国内相对高昂的粮价以及硬性需求,对于粮食的进口需求也会产生增加。自新冠疫情爆发以来,各国纷纷采取措施以应对其消极影响,包括调整利率、货币供给等金融政策,以及农产品进出口贸易政策。这些国家在政策选择上的趋同导致了当前粮价波动的原因之一。2020 年 3 月,国际市场粮价增长速率明显加快,2022 年,俄乌冲突的爆发,进一步影响国际粮食市场,粮食供需关系发生较大变化,粮食价格持续上涨,继续给国内的粮食市场和粮食安全带来冲击。

表18 产业链因素对国际市场的影响

驱动因素	分类	国际市场表现
产业链因素	贸易 (运输、运费)	(1)2020 年 3、4 月份至 2022 年 4、5 月,价格上涨; (2)2022 年 4、5 月份之后,价格下降
	供给 (天气、战争)	
	需求 (经济增长)	

由于新冠疫情影响,全球货运物流受到阻碍。尽管粮食总量相对充足,但全球粮食供需不平衡问题凸显,导致全球粮食危机。俄乌冲突的影响首先体现在贸易环节,其次才是生产环节。因此,在当前国际粮价波动中,贸易因素相比于生产和需求端更为主导。

## 5.2.4 “种子按斤称到论粒卖”内在机理

“种子按斤称到论粒卖”这句话揭示了种子产业从生产到销售的价值链变化和种子本身价值的提升。以前的种子都是用麻袋装,按斤称着算钱,农民买回去种到地里,发芽率难以保证,更难保证产量。而随着国家对种子技术的逐渐重视与不断发展,目前不少种子的售卖方式已经逐渐转变为“论粒卖”,种子的质量提升了,长出的作物再抗药性、口感上也越发优良,基于前面的分析,我们主要从国内、国际因素,对种子质量的提升进行分析。

### 1. 国内因素

自新中国成立初期起,中国政府就高度重视种子产业的发展。1949 年的第一次全国农业工作会议便提出了加强良种繁育和推广的方针,为我国种子事业的起步奠定了基础。经过 70 多年的发展,尤其是近年来政府更加重视种子产业,并相继推出了多项相关政策,加强了种业建设和监管工作。这些措施形成了一套法律规章体系,以《中华人民共和国种子法》为核心,辅以行政法规、地方法规和部门规章的支持。

同时在国家政策的支持下,中国种子研发和科技创新取得了显著成就。我国拥有世界上规模最大的公共研发团队,特别是在水稻、小麦等主要农作物领域,许多研究处于国际领先地位,尤其是杂交水稻和转基因水稻技术更是全球领先。资本实力雄厚的企业如中化、中信等已进入国内种业市场,并通过收购中国种业集团、隆平高科等龙头企业来加强其在种业领域的影响力,国内种业也形成了良好的合作竞争氛围。近年来,中国化工收购了先正达,中粮集团完成了对荷兰尼德拉种子公司的并购,中信农业也扩展了其在南美洲玉米种子业务的市场份额。

另外,对粮食自给率的分析表明,中国在粮食和其他主要农产品的种子自给率居于世界前列。主要粮食作物如水稻、小麦、玉米和大豆的种子需求量约为 900 万吨,几乎全部由国内生产。特别是杂交水稻、杂交玉米和小麦品种的种子产量远超过需求量,保障了种业安全。种子技术也在农业科技进步中扮演着关键角色,对于提升国家粮食安全和主要农产品生产力发挥了重要作用。自 20 世纪 80 年代初以来,新育成的水稻、小麦和玉米等作物品种持续提升产量,目前这些现代品种的广泛应用率已超过 96%。这一成就得益于中国强大的种子研发体系,同时也受益于国际公共研发机构和外国种业引进的优良遗传资源。

### 2. 国际因素

国际种业市场的集中度确实显示出逐年增加的趋势,特别是几大种业巨头在市场份额上的扩张。此外,全球种业市场的集中度也表现为越来越多的市场份额被前四大种业公司所掌握,约占全球种业市场的一半左右。

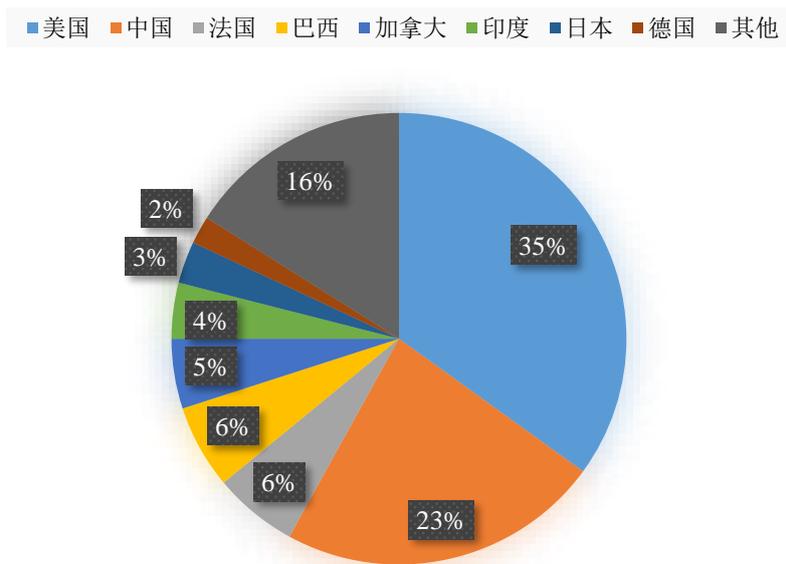


图40 全球种业区域分布

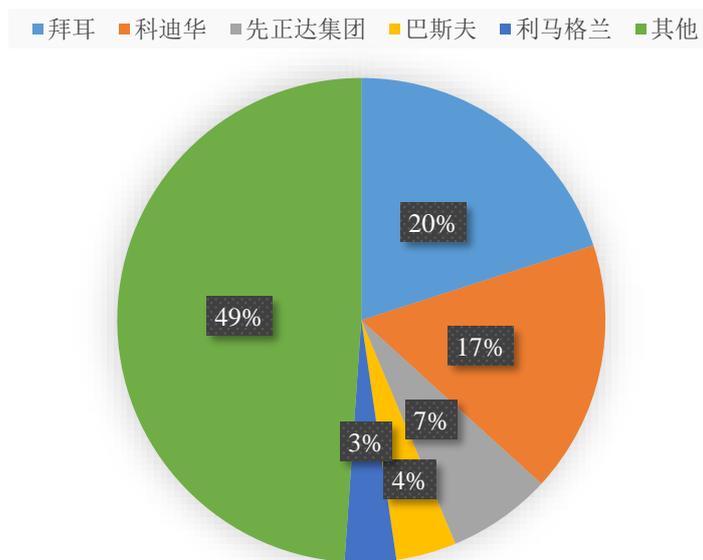


图41 2020 年全球种子行业市场份额

这种集中度增加的背景可以归因于几个主要因素：

(1) 技术与创新：大型种业公司通常投入大量资金用于种子研发和生物技术创新,这使得它们能够提供更具竞争力的种子产品,并在市场上占据领先地位。

(2) 市场合并与收购：种业市场近年来出现了多起合并和收购案例,这导致市场份额逐渐集中于少数几家大型企业。这些企业通过合并与收购扩大了其市场影响力和供应链能力。

(3) 知识产权与专利：大型种业公司拥有丰富的知识产权和专利,这些能够保护其研发成果和市场竞争优势,进一步增强了市场份额和集中度。

(4) 全球化和市场透明度：全球化使得市场更加透明和互联互通,大型种业公司能够更有效地在全球范围内扩展业务和市场份额。

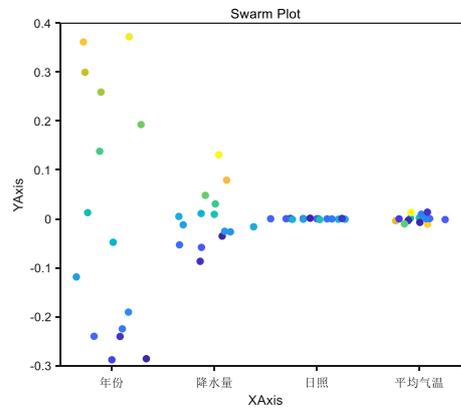
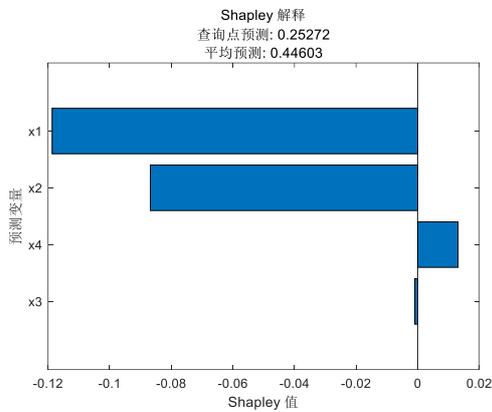
这种市场集中度增加的趋势,虽然可能会带来经济效益和技术进步,但也引发了一些关于市场竞争性和市场透明度的讨论。另外,国际市场对国内的不断冲击,对国内种业的发展既是机遇,也是挑战,如何跟上全球的种业步伐,维护国内粮食安全,任重道远。

## 6 我国粮食安全形势的演变趋势分析

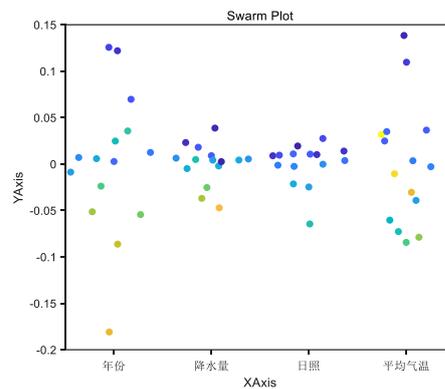
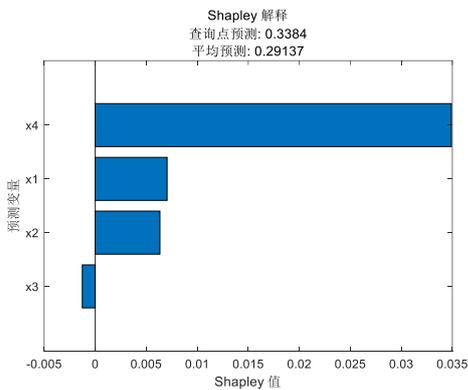
### 6.1 气候变化对粮食安全影响因果关系分析

根据气候-粮食产量效应模型,可以通过 Shapley 分析方法对多输出高斯过程模型中气候对三大主粮产量的因果关系展开分析。同样以黑龙江省粮食产量数据为例,计算各气候因素对水稻、小麦、玉米三种粮食产量 Shapley 值,如图 42。

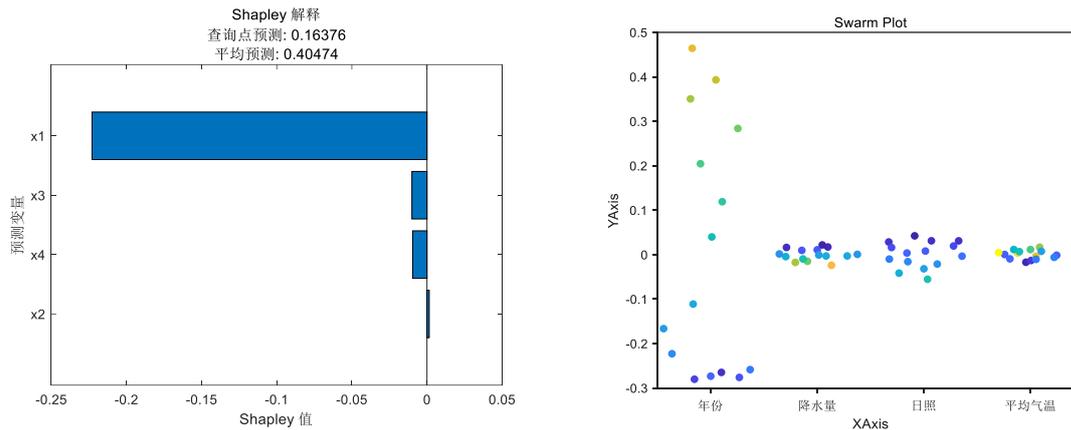
降水量在不同地区和作物中影响也不同。水稻产量随着降水量增加,对其产量增加均有正向作用。小麦产量随着降水量增加,对其产量增加有负向作用。日照对小麦和玉米影响较大,并且随着日照水平的增大,对作物产量增加具有负向作用。平均气温对小麦有较大影响。对玉米和水稻的影响较小。但平均气温对“小麦产量的影响趋势并不明确。



稻谷产量 Shapley 分析



小麦产量 Shapley 分析



玉米产量 Shapley 分析

图42 黑龙江三种主粮产量 Shapley 分析

随着全球气候变化加剧,高温对水稻、小麦和玉米等主要作物的生长有负面影响,可能导致生长期缩短、产量减少及品质下降;不规律的降水可能导致干旱或洪涝,对作物生长造成严重影响,尤其是影响灌溉依赖的地区;暴雨、冰雹等极端天气事件频繁发生,可能导致农作物受灾面积增加。未来气候变化对农业的负面影响可能进一步加剧。

同时随着国内经济发展,国内需求的增长和生产不确定性,中国可能会增加粮食进口依赖,通过多元化供应来保障国内粮食安全。对此,政府应继续加大对农业科技研发的支持,推动节水灌溉、抗旱、抗病虫害的新技术应用,提高农业生产适应气候变化的能力。

## 6.2 耕地减少对粮食安全影响因果关系分析

耕地面积的变化包含多方面因素,其中包括人口因素,经济因素以及农业投入等因素:

人口的自然增长和总人口的增加会导致对土地资源的需求增加,特别是工业和建设用地的扩张。这将直接导致耕地面积的减少。耕地面积的减少可能会导致粮食产量的下降,从而对粮食安全构成威胁。

第二产业,特别是工业生产的增加,也会占用大量的耕地。随着工业化进程的加快,工业用地的需求会进一步增加,从而导致耕地面积的减少。耕地面积的减少,同样会对粮食产量产生负面影响。

第三产业(服务业)的增长对耕地的需求较小。随着服务业的发展,经济结构逐渐从工业主导转向服务业主导,这可能会减缓工业用地的扩张速度,从而有助于保护耕地面积。此外,服务业的发展可以促进经济增长,减少对工业化的过度依赖,从而间接保护耕地。

农业机械总动力、农村用电量以及相应的财政支出等因素对耕地面积有正向影响。这些投入能够提高农业生产效率,促进耕地的维护和保护,从而增加粮食产量。

农村居民人均可支配收入的增加,能够提高农民的生活水平,使得农民不必放弃耕地寻找其他工作。这有助于保持农业劳动力的稳定,从而有利于耕地的维持和粮食生产的稳定。

考虑到我国发展现状,我国粮食安全形势的演变趋势为:随着人口增长和工业化进程的持续推进,耕面积面临一定程度的减少压力,但通过政策保护和农业投入的增加,可以在一定程度上维持粮食产量的稳定。但随着经济结构调整,服务业增长将缓解耕地减少的压力。同时,通过提高农业生产效率和优化粮食结构,可以进一步保障粮食安全。长期来看,我们需要持续关注人口增长和耕地保护的平衡,通过科技创新、政策支持和国际合作,确

保粮食安全的长期稳定。我国粮食安全形势的演变趋势是一个复杂的过程,受到多种因素的综合影响。为了确保粮食安全,需要在保护耕地、提高农业生产效率、优化经济结构和加强政策支持等方面综合施策,才能实现粮食安全的长期稳定。

## 6.3 种子短缺对粮食安全影响因果分析

种子供应链的中断可能由于多种因素引起,如天气灾害、自然灾害如洪水、干旱、飓风等可能严重影响种子生产和运输,导致供给链中断;市场垄断,少数种子生产公司垄断市场可能导致供给不足或价格上涨,使供给链不稳定。政策限制、贸易限制或法规变更可能影响种子的生产、进口或出口,从而打断供应链等。种子生产过程中的问题,如疾病、虫害或技术性失败,可能导致种子的产量减少或质量下降。全球性事件如疾病爆发(例如新冠疫情)、地缘政治紧张局势或国际贸易争端可能导致供给链中断或延迟。

种子供应不足会直接影响到农民的播种计划和作物的生产量。缺乏适当的种子品种也会降低作物的抗病虫害能力和适应环境的能力,种子价格上涨或供应不足也会增加农业生产的成本,影响农民的经济收益。高昂的种子成本可能迫使农民减少种植面积或者转向其他经济作物,从而间接影响到粮食供应的稳定性和安全性。

长期的种子供应问题可能影响到农业生产的持续性和可持续性。如果种子供应问题长期存在,可能会影响到农业技术进步、农民的投资意愿和政府的粮食安全政策的实施效果,从而长期影响到国家粮食安全的整体形势。

政府可以通过多样化种子供应来源,包括支持本地种子生产、鼓励农业科技创新和跨国合作,来减少种子供应的单一依赖和风险。建立健全的农业保护政策和法律框架,保障农民获取合理价格的高质量种子的权利。技术支持和信息分享:加强农业技术研发和推广,提供农民种植技术和种子选择的信息支持,以提高农业生产的效率和可持续性。

## 7 模型评价与推广

### 7.1 模型的优点

- (1) 论文所提出模型对现实问题进行简化,将复杂问题简约化,容易理解且便于使用,模型贴合实际,具有较高的应用价值;
- (2) 气候-粮食产量效应预测模型考虑多输出相关性信息,有效提升了数据预测准确性,为粮食产量预测及气候影响研究提供了一条有效途径;
- (3) 本文使用的耕地质量综合评价指标体系,在已有耕地质量评价方法研究基础上,采用采用 PSR 模型进行多方面评价,指标全面,代表性强。

### 7.2 模型的不足

由于部分数据来源于已有参考文献的相关研究,数据来源可信度难以求证,所建模型可能存在偏差;

本文提出的模型对于现有数据使用效果较好,由于时间问题没有对其他历史数据进行检验.对于更早期历史数据及长远外来预测,可能无法达到较好的效果。

### 7.3 模型的推广

本文在考虑收集数据信息时,主要考虑气候、土壤环境、粮食产量、社会经济状况、粮食贸易等数据集。关于粮食产量和耕地面积的预测,如果将地质条件、政策影响、粮食进出口数据等信息纳入考虑,可将气候-粮食产量效应预测模型推广为基于多维信息的粮食产量效应预测模型,进而推广至国家粮食安全战略层面、经济发展战略层面的预测及决策分析。

---

## 参考文献

- [1] 韩芳玉, 张俊飏, 程琳琳, 等. 气候变化对中国水稻产量及其区域差异性的影响[J]. 生态与农村环境学报, 2019, 35(3): 283-289. DOI: 10.19741/j.issn.1673-4831.2018.0224
- [2] LIN, LIN H, WANG T, et al. Impact of climate change on cotton growth and yields in Xinjiang, China[J]. Field Crops Research, 2020,247: 107590.
- [3] 王柳, 熊伟, 温小乐, 冯灵芝. 温度降水等气候因子变化对中国玉米产量的影响[J]. 农业工程学报, 2014, 30(21): 138-146. DOI: doi : 10.3969/j.issn.1002-6819.2014.21.017
- [4] 师晨迪, 许明祥, 邱宇洁, 张志霞, 张晓伟. 黄土丘陵区县域农田土壤近 30 年有机碳变化及影响因素研究: 以甘肃庄浪县为例[J]. 环境科学, 2014, 35(3): 1098-1104.
- [5] 李涛辉, 张文翔, 吕爱锋, 等. 1978—2019 年云南省气候生产潜力的时空演变及粮食产量响应[J]. 中国农学通报, 2023, 39(17): 55-63. DOI: 10.11924/j.issn.1000-6850.casb2022-0441.
- [6] 孟斌, 王劲峰, 张文忠, 等. 基于空间分析方法的中国区域差异研究[J]. 地理科学, 2005, 25(4): 393-400.
- [7] 段德宏, 孙林, 史印芳. 基于主成分分析法的青岛市耕地面积变化及驱动机制研究[J]. 山东科技大学学报(自然科学版), 2012, 31(4): 44-50.
- [8] 李丹, 周嘉, 战大庆. 黑龙江省耕地时空变化及驱动因素分析[J]. 地理科学, 2021, 41(7): 1266-1275.
- [9] 刘佳宁, 孙丽娜. 1980—2020 年黑龙江省土地利用碳排放时空分异研究[J]. 中国农学通报, 2023, 39(26): 86-97.
- [10] 李丹, 周嘉, 战大庆. 黑龙江省耕地时空变化及驱动因素分析[J]. 地理科学, 2021, 41(7): 10. DOI: 10.13249/j.cnki.sgs.2021.07.017.
- [11] 邓光联, 韩奎. 新冠肺炎疫情形势下对种业的一些思考[J]. 中国种业, 2020(6): 1-4.
- [12] 方怡晖. 长期疫情影响国际粮食产业链[J]. 小康(中旬刊), 2020(5): 32-34.
- [13] 黄季焜, 胡瑞法. 中国种子产业: 成就、挑战和发展思路[J]. 华南农业大学学报(社会科学版), 2023, 22(01): 1-8.
- [14] 黄毅, 柳思维. 国际种业垄断: 理论解释、实证测算及趋势[J]. 华南农业大学学报(社会科学版), 2015, 14(01): 79-91.
- [15] 孔令博, 林巧, 聂迎利, 等. 中国农作物种业发展现状及对策分析[J]. 中国农业科技导报, 2023, 25(4): 1-13.
- [16] 李继军. 疫情对种业的影响及中小企业战略调整的思考[J]. 种业导刊, 2020(1): 7-10.
- [17] 李丽菲. 新发展格局下保障粮食产业链供应链安全研究[J]. 现代食品, 2022, 28(14): 44-46.

- 
- [18] 凌婉阳, 陈士伟, 郑乾坤, 等. 浅析我国农作物种子政策法规的沿革与现状[J]. 甘蔗糖业, 2022,51(3): 1-13.
- [19] 刘石磊. 警惕全球粮食供应链新变局[J]. 黑龙江粮食, 2024(2): 11-12.
- [20] 孙梦爽. 评论 为种业装上更多“中国芯” [J]. 中国人大, 2024(01): 43.
- [21] 王柳, 熊伟, 温小乐, 等. 温度降水等气候因子变化对中国玉米产量的影响[J]. 农业工程学报, 2014(21): 138-146.
- [22] 王术坤, 韩磊. 中国种业发展形势与国际比较[J]. 农业现代化研究, 2022,43(5): 814-822.
- [23] 吴海彬, 朱志锋, 谭城, 等. 基于产业链视角的小麦品种品质需求调研分析 [J]. 中国种业, 2024(1): 1-8.
- [24] 熊启泉. 中国粮食的真实进口规模与自给率[J]. 华南农业大学学报 (社会科学版), 2022,21(3): 85-101.
- [25] 张秀青. 国际市场动荡不定 粮食安全首当其冲——2020年以来国际粮食市场及其对我国的影响研究[J]. 价格理论与实践, 2022(10): 37-41.
- [26] 赵军明, 金琰, 胡盛红, 等. 中国农作物种业贸易现状及发展趋势分析[J]. 热带农业科学, 2020,40(4): 93-98.
- [27] 郑可心,张锡煜,张清.1985-2020年海伦市耕地质量遥感监测与空间格局变化分析[J].黑龙江农业科学,2024,(01):29-36.
- [28] 王建华, 陶培峰, 袁月, 李志忠, 杨佳佳. PSR 框架下的黑龙江省海伦市耕地质量评价. 地质与资源, 2020, 29(6): 525-532.
- [29] 侯永华.典型白浆土区耕地质量评价研究[D].东北农业大学,2023

## 附录

### 附录 I: 主要程序/关键代码

#### 代码清单 1 问题一趋势检验方法(Python)

```
import pandas as pd
import numpy as np
from scipy.stats import norm
from sklearn.linear_model import LinearRegression
import os

# Define a function to perform the Mann-Kendall test. This function will take an
array of time series data as input and
# return the statistics and p-values of the Mann-Kendall test.
def trend_analysis(data):
    mk_z = []
    mk_p = []
    sen_slope_val = []
    for i in range(len(data)):
        x = data.iloc[i, 0:].values
        # Mann-Kendall test
        s = 0
        for j in range(1, len(x) - 1):
            for k in range(j + 1, len(x)):
                s += np.sign(float(x[k]) - float(x[j]))
        var_s = len(x) * (len(x) - 1) * (2 * len(x) + 5) / 18
        if s > 0:
            z = (s - 1) / np.sqrt(var_s)
        elif s < 0:
            z = (s + 1) / np.sqrt(var_s)
        else:
            z = 0
        p = 2 * (1 - norm.cdf(abs(z)))
        mk_z.append(z)
        mk_p.append(p)
        # Sen's slope method
        slopes = []
        for j in range(1, len(x) - 1):
            for k in range(j + 1, len(x)):
                slope = (x[k] - x[j]) / (k - j)
                slopes.append(slope)
        sen_slope_val.append(np.median(slopes))

    result = pd.DataFrame({'Mann-Kendall Z': mk_z, 'Mann-Kendall p': mk_p,
                          'Sen\'s Slope': sen_slope_val})

    return result
```

#### 代码清单 2 问题一回归方法(Python)

```
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np
from linearmodels import PanelOLS
from sklearn.preprocessing import StandardScaler
```

```

# Reading data
data = pd.read_excel("combined_data.xlsx")
# data.fillna('0', inplace=True)
# Select the desired columns
data = data[['PR', 'Pa', 'Wh', 'Co', 'Year', 'Ra', 'Pa_s', 'Wh_s', 'Co_s', 'Su',
'Te']]
# Standardized features
data = data.set_index(['PR', 'Year'])
scaler = StandardScaler()
data[['Ra', 'Su', 'Te']] = scaler.fit_transform(data[['Ra', 'Su', 'Te']])
# 拟 Fit the panelOLS model
# data['Pa_p'] = data['Pa']/data['Pa_s']
model = PanelOLS.from_formula('Co ~ 1 + Co_s + Te + Ra + Su', data=data).fit()
print(model)
# Model prediction
data['Pa_predicted'] = model.predict()
# # calculate RMSE
rmse = np.sqrt(mean_squared_error(data['Co'], data['Pa_predicted']))
# # calculate MAE
mae = mean_absolute_error(data['Co'], data['Pa_predicted'])
# print("RMSE:", rmse)
# print("MAE:", mae)

```

### 代码清单 3 问题一高斯过程建模及 shapley(MATLAB)

```

errors={};
xs=[4,5,9,10];
order=randperm(20);
test=order(1:5);
others=order(6:end);
% for i=1:31
%   data=datas{i};
%   y=data(:,1:3);
%   for j=2:size(y,1)
%     y(j,:)=y(j,:)-y(j-1,:);
%   end
%   data(:,1:3)=y;
% end
for i=1: size(province,1)
  data=datas{i};
  newdata=data;
  %   for k=1:3
  %     newdata(:,k)=data(:,k)./data(:,k+5);
  %   end
  y=newdata(:,1:3);
  new_y=y;
  %   for j=2:size(y,1)
  %     new_y(j,:)=y(j,:)-y(j-1,:);
  %   end
  y=new_y;
  x=data(:,xs);
  x(1,:)=[];
  y(1,:)=[];
  [x,settings1]=mapminmax(x',0.1,1);
  x=x';
  [y,settings2]=mapminmax(y',0.1,1);
  y=y';
  x_test=x(test,:);

```

```

y_test=y(test,:);
input=x(others,:);
output=y(others,:);
model = MRSM (input, output);
% X_ = [2.5, 3.5;2,4];
[Y_, S_] = predict_resp(model, x_test);
f = @(input) predict_resp1(model,input);
explainer = shapley(f,input);
Values=[];
figure;
for m =1:size(input,1)
    queryPoint = input(m,:);
    explainer = fit(explainer,queryPoint);
    Values=[Values;explainer.ShapleyValues];
    values=Values(:,2).Variables;
%每一行对应一个 x, 每一列对应一个 y
Values0=reshape(values,size(input,2),[]);
    plot(explainer,'NumImportantPredictors',4)
end
% Y_=mapminmax('reverse',Y_',settings2);
% Y_=Y_';
for j=1:size(Y_,2)
    error=in_cal_error(Y_(:,j),y_test(:,j));
    errors{j}(i,:)=error;
end
end

C=slanCL(1924);
C1 = C(1,:);
C2 = C(2,:);
C3 = C(3,:);
C4 = C(4,:);
% 图片尺寸设置 (单位: 厘米)
figureUnits = 'centimeters';
figureWidth = 15;
figureHeight = 12;
% 窗口设置
figureHandle = figure;
set(gcf, 'Units', figureUnits, 'Position', [0 0 figureWidth figureHeight]);
hold on
% 分簇散点图绘制
ms = 8;
ny=size(input,2);
for j=1:ny
% scatter(ones(size(input,1),1)*j,Values0(:,j),[],input(:,j),'filled');
    s1 = swarmchart(ones(size(input,1),1)*j,Values0(:,j),[],input(:,j),'filled');
% colorbar;
    hold on;
end
hTitle = title('Swarm Plot');
hXLabel = xlabel('XAxis');
hYLabel = ylabel('YAxis');
xjs = 0.6;
set(s1,'XJitterWidth',xjs)
% 坐标轴美化
set(gca, 'Box', 'off', ... % 边框
        'LineWidth',1,...
        'XGrid', 'off', 'YGrid', 'off', ... % 网格

```

```

    'TickDir', 'out', 'TickLength', [.01 .01], ...           % 刻度
    'XMinorTick', 'off', 'YMinorTick', 'off', ...           % 小刻度
    'XColor', [.1 .1 .1], 'YColor', [.1 .1 .1],...         % 坐标轴颜色
    'XTick', 1:8,...
    'XTickLabel',{'年份';'降水量';'日照';'平均气温';'x5';'x6';'x7';'x8'})
% 添加上右框线
XL = get(gca,'xlim'); XR = XL(2);
YL = get(gca,'ylim'); YT = YL(2);
xc = get(gca,'XColor');
yc = get(gca,'YColor');
plot(XL,YT*ones(size(XL)),'color', xc,'linewidth',1)
plot(XR*ones(size(YL)),YL,'color', yc,'linewidth',1)
% 字体和字号
% set(gca, 'FontName', 'Arial', 'FontSize', 10)
% set([hXLabel,hYLabel], 'FontName', 'Arial', 'FontSize', 10)
% set(hTitle, 'FontSize', 12, 'FontWeight', 'bold')
% 背景颜色
    set(gcf, 'Color', [1 1 1])

```

#### 代码清单 4 问题二数据补充：趋势外推(MATLAB)

```

n=length(yt);
alpha=0.3;
st1(1)=yt(1);
st2(1)=yt(1);
for i=2:n
    st1(i)=alpha*yt(i)+(1-alpha)*st1(i-1);
    st2(i)=alpha*st1(i)+(1-alpha)*st2(i-1);
end
% xlswrite('fadian.xls',[st1',st2']) ;
a=2*st1-st2;
b=alpha/(1-alpha)*(st1-st2);
y_new=a(19)+b(19);

```

#### 代码清单 5 问题二主成分分析(Python)

```

df=data
df std = (df - df.mean())/df.std()
df std
dataset corr= df std.corr()
dataset corr
import numpy as np
import math as math

def kmo(dataset corr):
    corr inv = np.linalg.inv(dataset corr)
    nrow inv corr, ncol inv corr = dataset corr.shape
    A = np.ones((nrow inv corr,ncol inv corr))
    for i in range(0,nrow inv corr,1):
        for j in range(i,ncol inv corr,1):
            A[i,j] = -(corr inv[i,j])/(math.sqrt(corr inv[i,i]*corr inv[j,j]))
            A[j,i] = A[i,j]
    dataset corr = np.asarray(dataset corr)
    kmo num = np.sum(np.square(dataset corr)) - np.sum(np.square(np.diagonal(A)))
    kmo denom = kmo num + np.sum(np.square(A)) - np.sum(np.square(np.diagonal(A)))
    kmo value = kmo num / kmo denom
    return kmo value

kmo(dataset corr)

```

```

eig value,eig vector=np.linalg.eig(dataset corr)
#特征值排序
eig=pd.DataFrame({"eig value":eig value})
eig=eig.sort values(by=["eig value"], ascending=False)
#获取累积贡献度
eig["eig cum"] = (eig["eig value"]/eig["eig value"].sum()).cumsum()
#合并入特征向量
eig=eig.merge(pd.DataFrame(eig vector).T, left index=True, right index=True)
eig
loading = eig.iloc[:2,2:].T
loading["vars"]=df std.columns
filepath = r'X2.xlsx'
loading.to excel(filepath, index=False)
loading
score = pd.DataFrame(np.dot(df std,loading.iloc[:,0:2]))
filepath = r'X1.xlsx'
score.to excel(filepath, index=False)
score

```

### 代码清单 6 问题二基于主成分分析的回归(R)

```

library(tidyverse)
library(caret)
library(readxl)
library(lavaan)
my data<-read excel("")
head(my data)
model1 <- lm(y ~ comp1 + comp2, data = my data)
summary(model1)
model2 <- lm(y ~ comp1* comp2 + comp1 + comp2 + I(comp1^2) + I(comp2^2), data = my data)
summary(model2)
predictions1 <- model2 %>% predict(my data)
RMSE(predictions1, my data$y)
predictions2 <- model1 %>% predict(my data)
RMSE(predictions2, my data$y)
anova(model1,model2)
res<-residuals(model2)
shapiro.test(res)
plot(model2)
library(lmtest)
bptest(model2)

```