



中国研究生创新实践系列大赛
中国光谷·“华为杯”第十九届中国研究生
数学建模竞赛

学 校 武汉大学

参赛队号 22104860014

1.陈苗苗

队员姓名 2.孙守强

3.南思宇

中国研究生创新实践系列大赛

中国光谷·“华为杯”第十九届中国研究生 数学建模竞赛

题 目

草原放牧策略研究

摘要：

自 2003 年起，“退牧还草政策”逐渐进入人们视野，合理的草原放牧策略管理成为维护生物多样性、调节水土流失等现象的重要手段。在此背景下，本文研究一系列放牧强度对草原土壤物理性质、化学性质、植被生物量影响的诸多问题，广泛收集全球土壤数据、文献相关数据并进行数据处理和合并，综合运用微积分方程、多项式回归模型、OLS 普通最小二乘法算法、时间序列预测模型、决策树模型、RandomForest 回归模型、单变量回归诊断方法、粒子群算法、SHapley Additive exPlanation 特征影响分析、相关性分析、因子分析、变异系数法、熵值法、多目标进化算法等方法建立相关问题的数学模型，具体做法如下：

针对问题一（放牧强度对土壤物理性质和植被生物量影响的数学模型），首先，从机理分析的角度厘清放牧策略对土壤湿度和植被生物量的影响关系和主要指标，采取逐步推理方式确定求解数学模型。然后，利用附件 3、4、5、6、9、10 等数据，研究降水量、蒸发量、叶面积指数、绿植覆盖率等主要指标与时间的关系。接着，采用 ARIMA 时序分析模型分别对降水量、蒸发量进行时序建模，并利用自相关/偏自相关系数、BIC、分位图进行最优模型选择和时序稳定性验证；同时，通过观察指标分布，采用多项式回归分析模型对叶面积指数、绿植覆盖率进行回归建模，利用 OLS 最小二乘法进行最优模型参数求解。最后，通过定性分析揭示不同放牧策略对土壤物理性质和植被生物量的影响关系。

针对问题二（对 2022 年、2023 年不同深度土壤湿度进行预测），首先基于问题一从机理分析视角所建立的土壤湿度数学模型，明确土壤湿度的影响指标；然后，采用线性/非线性回归分析分别对关键影响指标(降水量、土壤蒸发量)按指定月份进行建模并预测；之后，基于降水量、土壤蒸发量对不同深度的土壤湿度的多重回归模型，挖掘三者间的相互作用关系；最后对 2022 年、2023 年不同深度土壤湿度进行预测，并对预测结果进行多维度分析，揭示了土壤湿度随降水量、土壤湿度的变化规律及其在不同月份和季度的演化特征和原因。

针对问题三（放牧强度对土壤化学性质影响的数学模型及预测），首先，为了建立放牧强度对土壤化学性质影响的机理模型，根据已有方程模型，将具有实际意义的土壤化学性质（SOC、SIC 和全 N）与植物生物量的数据进行关联对应和数据均值处理并构建土壤化学性质同 Woodward 放牧与植物生长模型的等式。其次，根据问题一的方程求解方法，使用多项式回归求解与土壤化学性质随时间变化相关的系数并构建出数学模型。随后，构建 60 个基于 OLS 回归的不同放牧强度下不同小区土壤化学性质随时间变化的模型，使用

CCPR 图、拟合图、偏回归图、模型残差图检验 OLS 回归模型建立的有效性并进行回归预测。最后，进行扩展研究，建立了决策树回归、线性回归、随机森林回归、**AdaBoost** 决策树回归、**GradientBoost** 回归、**XGBoost** 回归、**CatBoost** 回归、神经网络线性回归、**LGBMBoost** 回归累计 9 种回归模型预测不同放牧强度对土壤化学性质的影响，考虑影响每种土壤化学性质的最优模型并进行参数调优，确定影响每种土壤化学性质的最佳模型然后输入至可解释机器学习 **SHapley Additive exPlanation** 模型中，建立了**不同放牧强度对土壤化学性质的影响机制**。研究结果表明使用 OLS 回归能够很好地拟合不同放牧强度下各小区土壤化学性质的变化趋势，基于 SHAP 模型的影响机制建立，完美阐释了不同放牧强度对土壤化学性质影响的结果，验证了机器学习模型在相关研究中的有效性。

针对问题四（沙漠化程度指数预测模型和土壤板结化），通过观察发现，现有的研究数据不能支撑该问题的解决，因此，本文广泛收集内蒙古水资源数据和全球土壤信息数据。针对沙漠化程度指数预测模型，根据现有文献，重新建立基于植物生长多样性因素、气象因素、地表与水资源因素、人文因素的沙漠化程度预警指标，整合附件 2、附件 6、附件 8 和附件 15 及来自内蒙古政务网站的水资源数据，使用随机森林缺失值填充法和迭代插值法填补缺失数据，并充分考虑指标正负性影响，构建了基于因子分析法和变异系数法构建沙漠化程度指数公式；针对土壤板结化程度模型，随机收集来自 WoSIS 的 50 个地区全球土壤信息数据，基于熵值法构建全球土壤板结化程度公式，确定了土壤板结化的定量定义，并使用 **flourish** 可视化方法展现该 50 个地区土壤板结化程度；最后，基于 **PSO** 粒子群算法就沙漠化程度预测模型和土壤板结化程度模型进行最优求解，确定最优放牧策略为：当月降水量约为 45mm、平均气温约为 19 度时，采用轻牧方式。

针对问题五（在特定降雨量条件下，预测草原可持续发展原则下的放牧标准羊数量最大值），通过对附件 8 中气候情况进行皮尔逊相关系数分析，显示降雨量与其他气候条件相关性较弱，可以作为独立变量。草原可持续发展原则是在保证生态环境良性健康发展中寻求经济利益的最大化，为此，本文定义了沙漠化指数、土壤板结化程度以及生物多样性指标来评估草原生态的良性发展，利用**标准羊数量评估经济利益**，并将该问题视为多目标优化问题。其中生物多样性指标使用 α 多样性中的 ACE 指标计算，然后利用多项式回归建立了植物多样性同降水量、羊群数量的回归模型；经济利益通过分析附件 13，构建了每公顷经济收益同羊群数量的回归模型。最后，针对该多目标优化问题，使用融合回归模型的基于分解的多目标进化算法，在求解目标函数最优值，且满足约束条件的情况下，求解放牧标准羊数量的最大阈值。

针对问题六（求解不同放牧策略下土地情况的变化），基于前五个问题研究的基础上，本文将问题六分解成四个子问题，分别建立不同放牧强度同土壤化学性质的回归模型，土壤化学性质、降雨量同土壤湿度的回归模型，植被指数同绿植覆盖率的回归模型，土壤湿度同绿植覆盖率的该回归模型，从而间接求解出不同放牧策略下土地肥力、土壤湿度、绿植覆盖率的变化情况。

关键词：多项式回归；普通最小二乘法；机器学习；相关性分析；粒子群 PSO 算法；多目标进化；SHAP 影响分析；综合评价法

目录

一、问题背景与重述	9
1.1 问题背景	9
1.2 问题重述	9
二、总体技术路线图	11
三、基本假设与符号说明	12
3.1 基本假设	12
3.2 符号说明	12
四、问题一的求解：时间序列分析与多项式回归分析	13
4.1 问题分析和技术路线图	13
4.2 数学模型推理建立	13
4.3 ARIMA 模型构建	15
4.3.1 平稳时间序列的定义	16
4.3.2 时间序列预处理	16
4.3.3 ARIMA 模型定阶	16
4.3.4 ARIMA 模型求解	17
4.3.5 锡林郭勒草原土壤蒸发量时序分析	17
4.3.6 锡林郭勒草原降水量时序分析	20
4.4 多项式回归分析	24
4.4.1 多项式回归原理	24
4.4.2 锡林郭勒草原叶面积分指数 LAI 多项式回归分析	26
4.4.3 锡林郭勒草原绿植覆盖率多项式回归模型求解	28
4.4.4 放牧强度对土壤湿度和植被生物量影响的定性分析	30
4.5 总结	31
五、问题二的求解：多元回归模型预测土壤湿度	33
5.1 问题分析和技术路线图	33
5.2 模型建立	33
5.3 回归模型求解	33
5.3.1 锡林郭勒草原土壤蒸发量回归分析	33
5.3.2 锡林郭勒草原降水量回归分析	36
5.4 多元回归分析和预测土壤湿度	39
5.4.1 10 厘米深度的土壤湿度多元回归分析	39
5.4.2 40 厘米深度的土壤湿度多元回归分析	40
5.4.3 100 厘米深度的土壤湿度多元回归分析	40
5.4.4 200 厘米深度的土壤湿度多元回归分析	41
5.4.5 土壤湿度预测结果与分析	42
5.5 总结	43
六、问题三的求解：OLS 回归与机器学习模型	44
6.1 问题分析与技术路线图	44
6.2 数据分析与数据预处理	45
6.2.1 土壤化学性质处理与数据概况	45
6.2.2 土壤植物干重数据处理与对应关系建立	46
6.3 基于多项式回归的放牧强度对土壤化学性质影响的机理模型	47

6.3.1 放牧强度对土壤化学性质影响的数学模型构建	47
6.3.2 基于多项式回归的放牧强度对土壤化学性质影响的数学模型求解	48
6.4 基于普通最小二乘法(OLS)回归的土壤化学性质预测	51
6.4.1 不同放牧强度土壤化学性质变化的线性总览	51
6.4.2 单变量回归诊断方法	52
6.4.3 基于 OLS 的土壤化学性质预测	53
6.5 扩展研究：基于机器学习的放牧强度对土壤化学性质的影响机制研究	56
6.5.1 回归方法调研与阐述	56
6.5.2 模型预测的最优回归方法选择	58
6.5.3 模型调优与放牧强度对土壤化学性质的影响机制分析	59
6.6 总结	64
七、问题四的求解：综合评价法与粒子群优化算法	65
7.1 问题分析和技术路线图	65
7.2 基于变异系数法和因子分析的沙漠化程度预测模型	66
7.2.1 变异系数法和因子分析	66
7.2.2 沙漠化程度预警指标筛选与相关数据收集和预处理	67
7.2.3 基于因子分析的沙漠化程度预警指标确定	70
7.2.4 基于变异系数法的沙漠化程度指数	72
7.3 基于熵值法的土壤板结化预测模型	73
7.3.1 基于熵值法的综合评价法	73
7.3.2 土壤板结化相关指标数据收集	74
7.3.3 基于熵值法的土壤板结化程度预测	76
7.4 基于粒子群算法（PSO）的放牧策略模型求解	78
7.4.1 粒子群算法	78
7.4.2 算法求解过程	79
7.4.3 放牧策略求解结果	81
7.4.4 结果分析	81
7.5 总结	81
八、问题五的求解：融合回归模型的基于分解的多目标进化算法	82
8.1 问题分析与技术路线图	82
8.2 数据预处理	82
8.3 模型介绍	83
8.3.1 降水量、放牧强度与植物多样性回归模型	83
8.3.2 经济收入与羊群关系回归模型	84
8.3.3 基于分解的多目标进化算法（MOEAD）	85
8.3.4 融合回归模型的多目标优化算法	86
8.3.5 求解结果	89
8.3.6 结果分析	89
8.5 总结	90
九、问题六的求解：模型综合与动态演示	91
9.1 问题分析与技术路线图	91
9.2 不同放牧策略描述	91
9.2.1 不同牧户放牧策略	91
9.2.2 示范区牧户平均放牧策略	91

9.2.3 问题四放牧策略	91
9.3 不同放牧强度下土地状态回归模型	92
9.3.1 不同放牧强度下土壤化学性质回归模型	92
9.3.2 土壤化学性质、降水量与土壤湿度回归模型	98
9.3.3 土壤湿度与植物覆盖率回归模型	98
9.4 结果分析	99
9.5 总结	100
十、模型评价与改进	101
10.1 模型优点	101
10.2 模型缺点	101
10.3 模型改进与推广	101
核心代码附录	102
参考文献	118

图录

图 1 总体技术路线图.....	11
图 2 问题一技术路线图	13
图 3 ARIMA 模型参数选择过程.....	17
图 4 ARIMA 模型拟合	17
图 5 2012-2022 年锡林郭勒草原土壤蒸发量时序和一阶、二阶差分图	18
图 6 自相关图-偏自相关图	18
图 7 BIC 模型评估结果	19
图 8 Q-Q 分位图	20
图 9 预测结果图	20
图 10 2012-2022 年锡林郭勒草原降水量时间序列、一阶差分和二阶差分图	22
图 11 自相关图-偏自相关图	22
图 12 BIC 模型评估	23
图 13 降水量时间序列 Q-Q 分位图	23
图 14 ARIMA 模型预测下一个月份降水量	24
图 15 1-12 月份叶面积指数 LAI	27
图 16 叶面指数多项式回归拟合效果	28
图 17 2021 年锡林郭勒草原的绿植覆盖率分布情况	29
图 18 绿植覆盖率多项式回归拟合效果	30
图 19 放牧强度对土壤湿度、植被生物量的影响关系	31
图 20 问题二技术路线图	33
图 21 2012-2022 年各月份锡林郭勒草原土壤蒸发量分布情况	34
图 22 1-12 月降水量	35
图 23 2012-2022 年各月份锡林郭勒草原降水量分布	37
图 24 锡林郭勒 2012-2022 年 1-12 月蒸发量	38
图 25 2012-2022 年锡林郭勒降水量、土壤蒸发量和土壤湿度	39
图 26 10 厘米深度下的 OLS 回归分析结果	40
图 27 40 厘米深度下的 OLS 回归分析结果	40
图 28 100 厘米深度下的 OLS 回归分析结果	41
图 29 200 厘米深度下的 OLS 回归分析结果	41
图 30 2022-2023 年土壤湿度预测结果	43
图 31 问题三技术路线图	45
图 32 数据处理后的样地小提琴图和抖动散点图	45
图 33 不同放牧强度下干重数据分布	46
图 34 不同年份下不同放牧强度干重数据分布	47
图 35 SOC 土壤有机碳数据拟合图形	49
图 36 SIC 土壤无机碳数据拟合图形	49
图 37 全氮 N 数据拟合图形	50
图 38 α 、 β 、 γ 系数求解结果	50
图 39 不同放牧强度下土壤化学性质的线性变化	52
图 40 G17 放牧区全氮 N 和土壤 C/N 比回归模型诊断	54
图 41 G17 放牧区 SOC、SIC、STC 回归模型诊断	56
图 42 二维平面决策树回归概念化模型	57

图 43	Adaboost 算法概念结构化模型	57
图 44	随机森林的构造过程	58
图 45	放牧强度对土壤化学性质影响的回归模型 MAE 和 MSE 图	59
图 46	SOC 情况下的参数调优图与特征影响图	60
图 47	SIC 情况下的参数调优图与特征影响图	61
图 48	STC 情况下的参数调优图与特征影响图	62
图 49	全 N 情况下的参数调优图与特征影响图	63
图 50	土壤 C/N 比情况下的参数调优图与特征影响图	64
图 51	问题四技术路线图	66
图 52	附件 15 真实数据和缺失数据展示	68
图 53	沙漠化预警指标体系	70
图 54	基于特征值的因子个数确定	71
图 55	沙漠化程度预警指标对因子影响的结果	72
图 56	世界土壤有机碳密度分布	74
图 57	世界土壤容重分布	75
图 58	基于熵值法的土壤板结化程度世界地图分布	78
图 59	粒子群算法流程图	79
图 60	问题五技术路线图	82
图 61	植物多样性与放牧强度、降水量关系折线图	83
图 62	锡林郭勒 2012-2021 年降水量趋势图	84
图 63	放牧强度与单位公顷经济收入关系图	84
图 64	MOEAD 算法流程图	86
图 65	气候指标相关性分析	87
图 66	问题六技术路线图	91
图 67	LGI 下 SOC 土壤有机碳拟合曲线	92
图 68	LGI 下 SIC 土壤无机碳拟合曲线	93
图 69	LGI 下全 N 拟合曲线	94
图 70	MGI 下 SOC 土壤有机碳拟合曲线	94
图 71	MGI 下 SIC 土壤无机碳拟合曲线	95
图 72	MGI 下全 N 拟合曲线	96
图 73	HGI 下 SOC 土壤有机碳拟合曲线	96
图 74	HGI 下 SIC 土壤无机碳拟合曲线	97
图 75	HGI 下全 N 拟合曲线	98
图 76	植被指数与绿植覆盖率关系图	99
图 77	不同放牧强度下土壤情况变化	100

表录

表 1 符号说明	12
表 2 问题一联合方程式相关变量说明	14
表 3 ARIMA 模型求解结果	20
表 4 锡林郭勒草原 2012-2022 年降水量情况	21
表 5 降水量 ARIMA 模型求解	24
表 6 2012-2017 年锡林郭勒草原叶面积指数 LAI 分布情况	26
表 7 LAI 多项式求解结果	27
表 8 2021 年锡林郭勒草原植被覆盖率数据预处理	28
表 9 绿植覆盖率多项式回归模型参数求解	29
表 10 2022-2023 年锡林郭勒草原降水量预测结果	35
表 11 降水量和蒸发量预测结果	38
表 12 2022-2023 年土壤湿度预测结果	42
表 13 放牧强度对土壤化学性质影响的数学模型符号解释	48
表 14 G17 放牧区土壤化学性质回归结果	53
表 15 锡林郭勒草原监测样地在不同放牧强度下 2022 年土壤化学性质预测值	56
表 16 不同放牧强度对土壤化学性质的影响机制	64
表 17 关于锡林郭勒盟水资源的数据收集概况	68
表 18 人文因素数据整合统计	69
表 19 沙漠化程度因子权重表	72
表 20 沙漠化程度及沙漠化程度指数划分标准	73
表 21 不同放牧强度下监测点沙漠化程度	73
表 22 土壤板结化相关数据收集	75
表 23 土壤板结化预警指标权重	76
表 24 各地区土壤板结化程度值（部分）	77
表 25 粒子群算法参数设定	79
表 26 粒子群算法子问题描述	80
表 27 约束条件定义	80
表 28 放牧策略求解结果	81
表 29 植物群落功能及物种数表示	82
表 30 MOEAD 算法参数列表	85
表 31 模型初始化参数列表	87
表 32 模型子问题及描述	88
表 33 模型子问题及描述	88
表 34 Water = 300 子问题解	89
表 35 Water = 600 子问题解	89
表 36 Water = 900 子问题解	89
表 37 Water = 1200 子问题解	89
表 38 问题六结果	99
表 39 土壤肥力及标准化后数据	100

一、问题背景与重述

1.1 问题背景

草原是我国生态系统重要组成部分，具有“钱库”的重要功能，是我国绿水青山和金山银山结合一起的重要有机体。草原的组成主要有生态、生产和生活。其中，生态意味着草原具有陆地自然生态系统的生态服务价值，如防沙固沙、调节气候、净化空气等^[1]，生产意味着草原可以作为人类、动物等生命体所需物质产出的来源，例如，牲畜可以采食饲草等，而草原上的重要药草如甘草、黄芪等是人类重要的药物来源；生活体现在多数少数民族生活在草原上，也是国家当前脱贫攻坚的重点地区。因此，草原的重要性不言而喻。

但近年来，全球沙漠化成为研究者极为关注的问题，尤其是草原土壤的沙漠化，土壤沙漠化^[2]将会严重破坏全球生态系统，危害草原人民生活和生产、当地生物多样性以及社会经济损失等。影响土壤沙漠化的一种原因可能与放牧强度相关^[3]，草原有多种放牧强度方式，如轻度放牧强度、中度放牧强度、重度放牧强度，当前研究主要通过对不同放牧强度的实验，探究放牧强度对土壤的物理性质（如土壤湿度、土壤容重）、化学性质（如土壤有机碳、土壤无机碳、全 N）、植被生物量等的影响，过度放牧可能导致草原植被结构破坏，土壤裸露面积增大，促进了土壤表面的蒸发，土体内水分相对运动受到不利影响，进而形成土壤沙漠化，尽管更多的放牧对于草原人民而言可能意味着更多的经济收益，但是在“现代草地资源的经营的可持续发展观念”下，调整放牧强度和放牧方式极其重要，能够保证生态平衡的同时，获得最大的收益。

本题目的命题宗旨是利用附件中的数据和自己所收集的数据，从机理分析角度建立不同放牧策略对土壤物理性质、化学性质和植被生物量影响的微积分数学模型，利用机器学习、统计学习等相关方法预测不同放牧强度下土壤湿度和土壤化学性质。此外还需要构建沙漠化程度指数预测模型和土壤板结化模型以及可持续发展观下求解放牧羊的数量，从而从科学的角度确定放牧的影响。

1.2 问题重述

根据上述背景，该题目提供了 15 个附件数据，以及相关参考文献、公式等，拟解决的关键问题如下：

(1) 建立放牧策略对土壤湿度和植被生物量影响的数学模型：从机理分析的角度，建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤物理性质（主要是土壤湿度）和植被生物量影响的数学模型。

(2) 建立模型进行不同深度土壤湿度预测：根据附件 3 土壤湿度数据、附件 4 土壤蒸发数据以及附件 8 中降水等数据，建立模型对保持目前放牧策略不变情况下对 2022 年、2023 年不同深度土壤湿度进行预测，并填写预测结果到表格。

(3) 建立放牧策略对土壤化学性质影响的数学模型并预测：从机理分析的角度，建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤化学性质影响的数学模型；结合附件 14 中数据预测锡林郭勒草原监测样地(12 个放牧小区)在不同放牧强度下 2022 年土壤同期有机碳、无机碳、全 N、土壤 C/N 比等值，并完成下表。

(4) 计算沙漠化程度指数、建立土壤板结化模型并选择最优策略：利用沙漠化程度指数预测模型和附件提供数据（包括自己收集的数据）确定不同放牧强度下监测点的沙漠化程度指数值。并请尝试给出定量的土壤板结化定义，在建立合理的土壤板结化模型基础上结合问题 3，给出放牧策略模型，使得沙漠化程度指数与板结化程度最小。

(5) 保持草原可持续发展的放羊数据量并找到最大阈值：锡林郭勒草原近 10 的年降水量（包含降雪）通常在 300 mm~1200 mm 之间，请在给定的降水量（300mm, 600mm、900 mm 和 1200mm）情形下，在保持草原可持续发展情况下对实验草场内（附件 14、15）

放牧羊的数量进行求解，找到最大阈值。

(6) 用图示或者动态演示方式分别预测示范区土地状态：在保持附件 13 的示范牧户放牧策略不变和问题 4 中得到的放牧方案两种情况下，用图示或者动态演示方式分别预测示范区 2023 年 9 月土地状态（比如土壤肥力变化、土壤湿度、植被覆盖等）。

二、总体技术路线图

本文的研究技术路线图如图 1 所示：

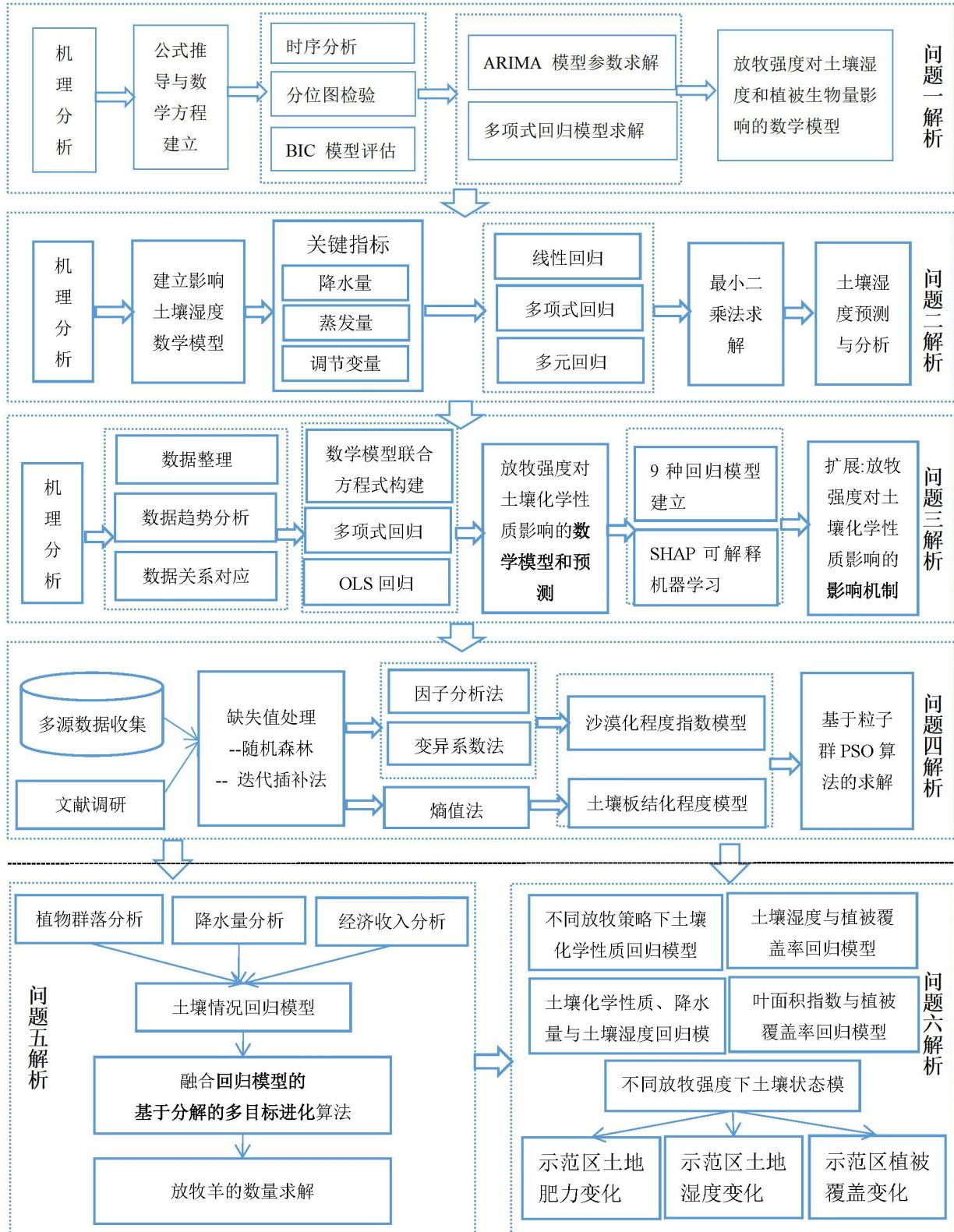


图 1 总体技术路线图

三、基本假设与符号说明

3.1 基本假设

根据内蒙古锡林郭勒盟草原监测、调查数据情况和本题所给出的条件，本文作出如下假设：

- (1) 所给数据除给定限制条件外，没有其他外部因素的影响，且数据人工统计误差可以忽略。
- (2) 除了所给的附件数据之外，其他外部因素对土壤物理性质、化学性质等的影响可以忽略。
- (3) 对于需要按年或者月份统计的数据可以通过平均的方式计算。
- (4) 对于缺失数据可以通过平均插值法进行填充。
- (5) 在放牧策略确定的情况下，人类活动对土壤等各项指标产生影响可以忽略。
- (6) 所有参考模型、公式和文献都真实无误。

3.2 符号说明

表 1 符号说明

序号	符号	说明
1	w	植被生物量
2	S	单位面积的载畜率
3	P	降水量
4	E	蒸发量
5	IC_{store}	植被截留率
6	R_{in}	入径流量
7	R_{out}	出径流量
8	G_u	地下水毛管上升量
9	G_d	土壤水渗透量
10	c_p	植被覆盖率
11	IC_{max}	特定植被的最大截流量
12	k	植被密度校正因子
13	R_{cum}	累积降雨量
14	LAI	叶面积指数
15	$w(t)$	随着时间变化的植被生物量含量变化
16	$SOC(t)$	随着时间变化的土壤有机碳含量的变化
17	$SIC(t)$	随着时间变化的土壤无机碳含量的变化
18	$N(t)$	随着时间变化的土壤全氮含量的变化
19	S	单位面积的载畜率
20	α	$SOC(t)$ 的系数
21	N	粒子数量
22	D	粒子维度
23	R	粒子范围
24	V_{max}	最大速度
25	c_1, c_2	社会认知
26	ω	惯性权重
27	I	迭代次数

四、问题一的求解：时间序列分析与多项式回归分析

4.1 问题分析和技术路线图

从机理分析的角度，建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤物理性质（主要是土壤湿度）和植被生物量影响的数学模型。

目前，不同放牧策略会对植被生物量产生影响，而生物量的变化关系到植被覆盖率和植被最大截流量。通过土壤-植被-大气系统的水平衡基本方程可知，植被截流率会影响到土壤土壤贮水变化量。因此，放牧强度会间接作用到土壤湿度变化。解析土壤湿度模型，需要获取影响土壤贮水变化量的降水量、蒸发量、植被截流率和植被覆盖率等因素。综上，确定问题一研究路线，如图 2 所示。

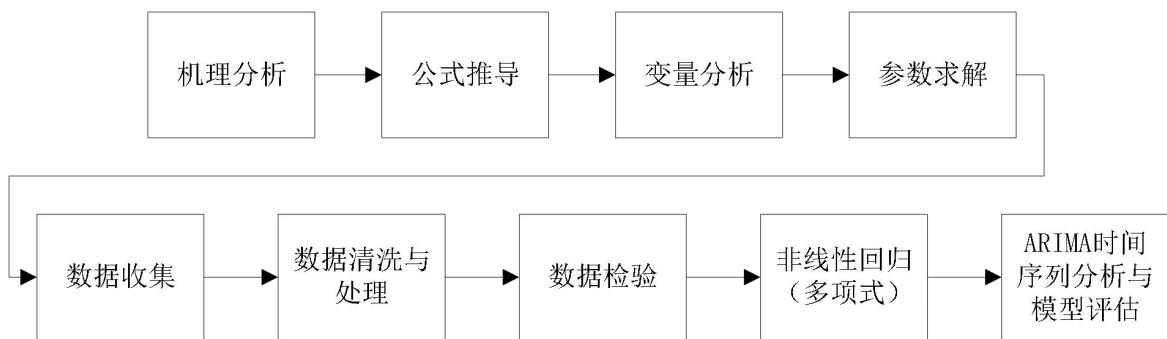


图 2 问题一技术路线图

4.2 数学模型推理建立

草原水分通过降水、入渗、蒸发等形式周而复始的循环，在无人干扰的情况下，土壤-植被-大气系统的水平衡基本方程为，

$$\Delta W = W_{t+1} - W_t = P + G_u + R_{in} - (E_{t_a} + G_d + R_{out} + IC_{store}) \quad (4-1)$$

由于锡林郭勒草原地势比较平坦，降水量和降水强度较少，水分循环以垂直方向的水量交换为主，绝大部分降水被蓄积在土壤中，尽管在遇到较大降水时会产生局地径流，但仍在整个草原区域内，其出入径流可视为相等。

所以，可建立公式 (4-2)：

$$\frac{dW(t)}{dt} = P(t) - E(t) - IC_{store}(t) - G_d + G_u \quad (4-2)$$

土壤含水量-降水量-地表蒸发模型可表示为公式 (4-3)。

$$\frac{d\beta}{dt} = P - E(a) \quad (4-3)$$

式 (4-3) 中， P 为该牧区供水率（主要为降水）； E 为地表蒸散发率； β 为土壤含水量； α 为土壤植被覆盖率可表达为 $\alpha^*G(w)$ ， w 为成草数量， $G(w) = (1 - e^{-\varepsilon_g w/w^*})$ 为草原的盖度，内蒙古草原盖度在 0.25~0.8 之间， α^* 为最大增长率，依赖于牧区草地除成草量外的环境条件（如光照、气温、土壤养分等）； $D = \beta * (e^{\varepsilon_g w/w^*} - 1)$ 为枯萎率，其中 w^* 为该草原群落的特征值只是一个大概的标度，引入它是使相应的无量纲量例如 w/w^* 中的量级为 1，便于计算， α^* 为有量纲系数，其余系数则是无量纲量。

由于植被截流量与降水量、植被覆盖度、叶面积指数 (LAI) 相关。植被覆盖率是植

物群落覆盖地表状况的一个综合量化指标。降水量小、植被覆盖度高、LAI 大时植被截流量大，其表达式为：

$$IC_{store} = c_p \cdot IC_{max} \cdot \left[1 - \exp\left(-k \cdot R_{cum} / IC_{max} \right) \right] \quad (4-4)$$

上式中， IC_{store} 为植被截流量 (mm)； c_p 为植被覆盖率； IC_{max} 特定植被的最大截流量 (mm)； k 为植被密度校正因子，与 LAI 有关； R_{cum} 为累积降雨量 (mm)。 IC_{max} 可以通过 LAI 来估算，

$$IC_{max} = 0.935 + 0.498 \cdot LAI - 0.00575 \cdot LAI^2 \quad (4-5)$$

式中，LAI 为一个分布式的时变参数。

所以，可进一步推理出公式 (4-6) 和 (4-7)。

$$IC_{store}(t) = c_p(t) \cdot IC_{max}(t) \cdot [1 - \exp(-k \cdot P(t)/IC_{max}(t))] \quad (4-6)$$

$$IC_{max}(t) = 0.935 + 0.498 \cdot LAI(t) - 0.00575 \cdot LAI(t)^2 \quad (4-7)$$

对于放牧与植物生长之间的关系，Woodward 等建立了如下模型[7]：

$$\frac{dw}{dt} = 0.049w(1 - \frac{w}{4000}) - 0.0047Sw \quad (4-8)$$

式中， w 为植被生物量， S 为单位面积的载畜率（与放牧强度有关/放牧方式）。

因此，公式 (4-8) 可写成公式 (4-9)：

$$\frac{dw(t)}{dt} = 0.049w(t)(1 - \frac{w(t)}{4000}) - 0.0047Sw(t) \quad (4-9)$$

综上所述，可以建立联合方程式如 (4-10) 所示。

$$\begin{cases} \frac{dw(t)}{dt} = 0.049w(t)(1 - \frac{w(t)}{4000}) - 0.0047Sw(t) \\ \frac{dW(t)}{dt} = P(t) - E(t) - IC_{store}(t) + R_{in} - R_{out} - G_d + G_u \\ c_p(t) = [a^* - W(t)(e^{\varepsilon_g w(t)/w^*} - 1)](1 - e^{-\varepsilon_g w(t)/w^*}) \\ IC_{store}(t) = c_p(t) \cdot IC_{max}(t) \cdot [1 - \exp(-k \cdot P(t)/IC_{max}(t))] \\ IC_{max}(t) = 0.935 + 0.498 \cdot LAI(t) - 0.00575 \cdot LAI(t)^2 \end{cases} \quad (4-10)$$

各符号说明如表 2 所示。

表 2 问题一联合方程式相关变量说明

序号	符号	说明
01	w	植被生物量
02	S	单位面积的载畜率（由放牧强度决定）
03	P	降水量
04	E	蒸发量

05	IC_{store}	植被截留率
06	R_{in}	入径流量
07	R_{out}	出径流量
08	G_u	地下水毛管上升量
09	G_d	土壤水渗透量
10	c_p	植被覆盖率
11	IC_{max}	特定植被的最大截流量
12	k	植被密度校正因子
13	R_{cum}	累积降雨量 (mm) (这里使用 $P(t)$ 表示)
14	LAI	叶面积指数

基于上述分析和公式 (4-10)，与时间无关的部分作为标量，与时间有关的部分作为求解项。

首先，放牧强度 S 与 $w(t)$ 植被生物量存在关系，即公式 (4-11)

$$\frac{dw}{dt} = 0.049w(1 - \frac{w}{4000}) - 0.0047Sw \quad (4-11)$$

然后，土壤贮水变化量表示土壤湿度，与降水量、地下水毛管上升量、土壤水渗透量、实际蒸发量、入和出径流量，以及植被截流量相关。单位时间内的土壤贮水变化量表示为公式 (4-12) 和公式 (4-13)。

$$\Delta W = W_{t+1} - W_t = P + G_u + R_{in} - (E_{ta} + G_d + R_{out} + IC_{store}) \quad (4-12)$$

$$\frac{dW(t)}{dt} = P(t) - E(t) - IC_{store}(t) + R - G_d + G_u \quad (4-13)$$

此外，植被截留率与 $IC_{store}(t)$ 与 $c_p(t)$ 植被覆盖率、 $IC_{max}(t)$ 特定植被的最大截流量及 $P(t)$ 降水量有关，如公式 (4-14)。

$$IC_{store}(t) = c_p(t) \cdot IC_{max}(t) \cdot [1 - exp(-k \cdot P(t)/IC_{max}(t))] \quad (4-14)$$

其中， $IC_{max}(t)$ 特定植被的最大截流量，由叶面积指数 (公式 (4-15)) 计算得出。这里叶面积指数可根据附件 10 得到，

$$IC_{max}(t) = 0.935 + 0.498 \cdot LAI(t) - 0.00575 \cdot LAI(t)^2 \quad (4-15)$$

然而，降水量由 $P(t)$ 通过附件 8 利用 ARIMA 时间序列模型分析得到。因此，需要求出 $c_p(t)$ 植被覆盖率。这里植被覆盖率可以根据附件 5 分析得出。由于植被覆盖率 $c_p(t)$ 与植被生物量存在关系，因而受到放牧强度影响。

$$c_p(t) = [a^* - W(e^{\varepsilon_g w(t)/w^*} - 1)](1 - e^{-\varepsilon_g w(t)/w^*}) \quad (4-16)$$

综上，本题求解放牧强度对土壤湿度和生物量的影响的数学模型，需要求解 $P(t)$ 降水量、 $LAI(t)$ 叶面积指数、 $E(t)$ 蒸发量和 $c_p(t)$ 植被覆盖率与时间的关系。

4.3 ARIMA 模型构建

时间序列处理常用的方法包括平滑法、趋势拟合法、组合模型、AR 模型、MA 模型和

ARMA 模型以及 ARIMA 模型^[4]。

4.3.1 平稳时间序列的定义

如果时间序列 x_t 在某一常数附件波动且范围有限，即有常数均值方差，且延迟 k 期的序列变量的子自协方差和自相关系数是相等的，则 x_t 是平稳序列。

4.3.2 时间序列预处理

(1) 时间序列判定

对于纯随机序列，它的各项之间是没有任何关系的，因此白噪声序列是没有信息可提取；

对于平稳的非白噪声序列，它的均值和方差是常数，可使用平稳序列拟合模型；

对于非平稳序列，即均值方差不稳定。这时可通过差分处理，判断差分后的时间序列是否平稳，如果具有平稳性，则该序列为差分平稳序列，进而通过模型进行拟合。

(2) 平稳性检验

①时序图检验

根据平稳时间序列的均值和方差都为常数的性质，平稳时间序列的时序图显示该序列始终在一个常数值附近随机波动，而且波动有界。

②自相关检验

平稳时间序列具有短期相关性，这表明平稳时间序列通常只有近期的序列值对现在的值影响较明显。随着延迟期数 k 的增加，平稳序列的自相关系数会衰减趋于零，并在零附近随机波动，而非平稳序列的自相关系数衰减速度较慢。

③单位根检验

判断时序是否存在单位根，如果存在即为非平稳时间序列。

4.3.3 ARIMA 模型定阶

ARIMA 模型参数定阶是指依据系统的输入输出数据建立模型结构的过程，包括模型的定阶与参数估计。由于实际的锡林郭勒草原数据收集缺失、数据处理过程等问题导致其分析过程存在不确定性。因此需要按照科学的流程进行建模和参数识别，其过程如图 3。本题中模型识别过程的具体步骤如下：

① 平稳性检测。首先画出内蒙古锡林郭勒地区统计指标的分布情况，观察变化是否稳定。

② 计算自相关系数。若自相关系数长期不等于 0，说明序列具有很强的长期相关性。再检验序列是否为非平稳序列。这里采用图检验方法，判断 p 值是否大于 0.05，可判断序列是否为非平稳序列。然后进行白噪声检验，(白噪声检验统计量对应的) p 值远小于 0.05，说明该序列属于非白噪声序列。

③ 如果不是非平稳序列，就对序列进行差分运算，再根据第 I 步进行平稳性验证，不满足就再进行差分运算，直至得到平稳序列。

④ 针对平稳非白噪声序列，利用 ARMA 模型进行分析。确定 p 、 q 值。

⑤ 相对最优模型识别。通过 BIC 模型进行评估，确定最佳的 p 和 q ，即找到 BIC 矩阵中最小值的坐标(a, b)，此时的 $p=a, q=b$ 。

⑥ 根据以上开展的 n 阶差分，确定的最佳 ARIMA 模型为 ARIMA(p, n, q)。

⑦ 针对建立好的 ARIMA 模型检验残差。

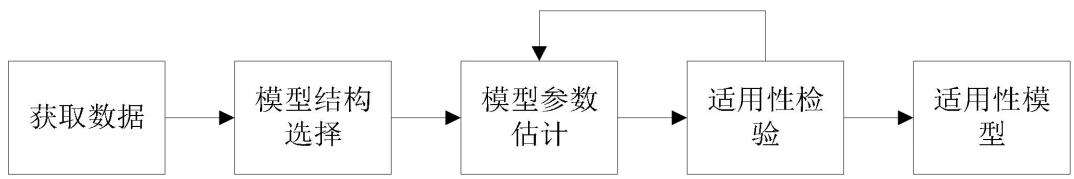


图 3 ARIMA 模型参数选择过程

4.3.4 ARIMA 模型求解

结合上述可知, $P(t)$ 降水量、 $E(t)$ 蒸发量、 $c_p(t)$ 植被覆盖率为平稳非白噪声序列, 符合 ARIMA 模型建模要求。因此我们可以采用 ARIMA 模型解题, 具体分析步骤如下:

- (1) 计算样本自相关系数 (ACF) 和样本偏自相关系数 (PACF) 的值;
- (2) 根据样本 ACF 系数和 PACF 系数, 选择表现突出的 ARIMA (p,q) 模型;
- (3) 检验模型的有效性。如果模型不通过检验, 转向上一步, 重新选择模型再拟合。
- (4) 模型优化。如果拟合模型通过检验, 还要充分考虑各种情况, 以建立多个拟合模型, 最后选择出最优模型。

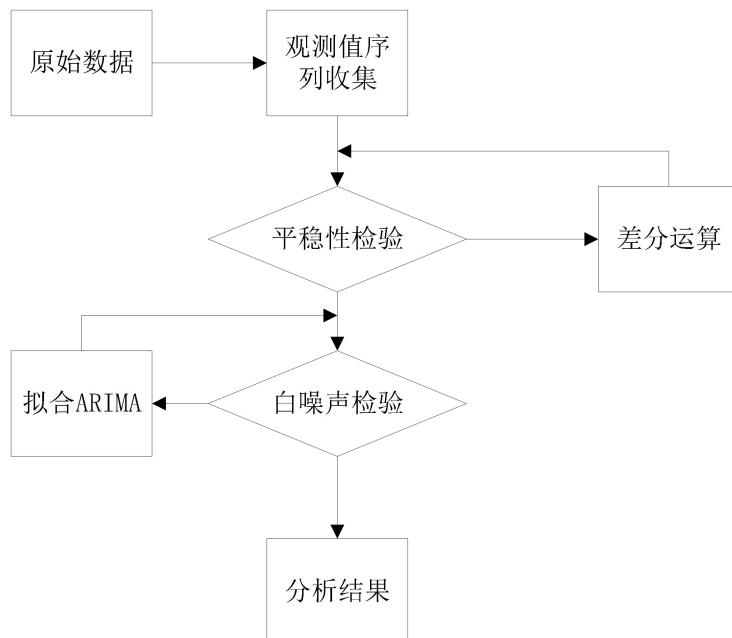


图 4 ARIMA 模型拟合

4.3.5 锡林郭勒草原土壤蒸发量时序分析

以内蒙古锡林郭勒草原土壤蒸发量为例, 图 5 描述了 2012-2022 年锡林郭勒草原土壤蒸发量时序和一阶、二阶差分效果。从图上可以发现, 通过二阶差分运算后, 锡林郭勒草原土壤蒸发量的时间序列为平稳时间序列。因此, 可以使用 ARIMA 进行时间序列建模与分析。

差分图

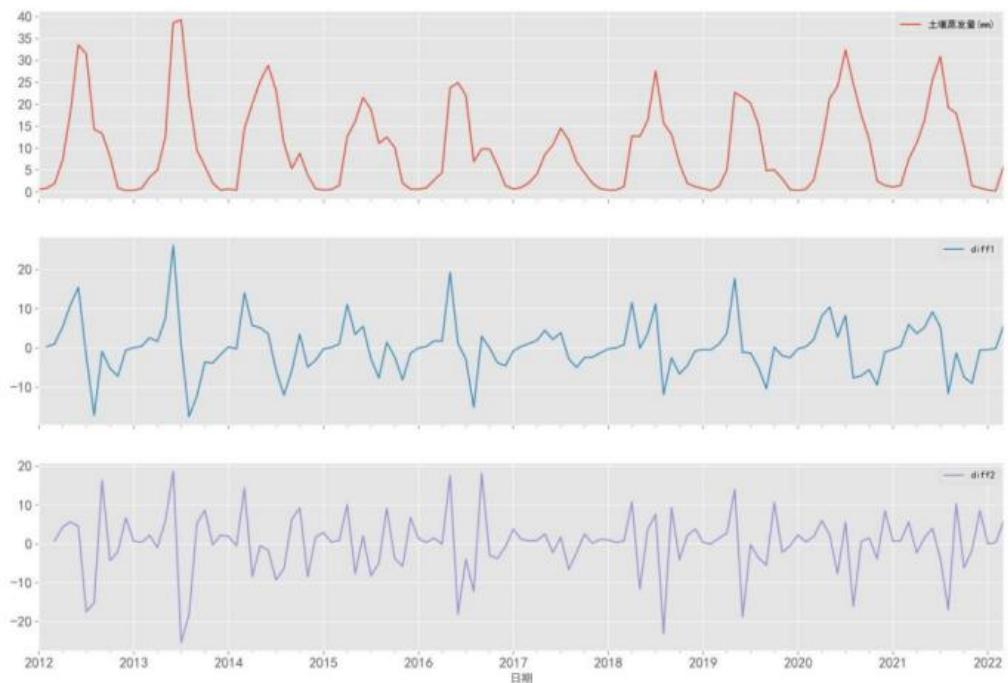


图 5 2012-2022 年锡林郭勒草原土壤蒸发量时序和一阶、二阶差分图

内蒙古锡林郭勒草原的土壤蒸发量时间序列的自相关函数和偏自相关函数分析见图(a)和图 6(b)。经过自相关图和偏自相关图分析，可见序列的自相关函数拖尾，所以通过 ARIMA(2, 0, 0) 模型进行拟合。最后使用贝叶斯信息准则 BIC 对 ARIMA 模型模型进行评估，评估结果见图 7。

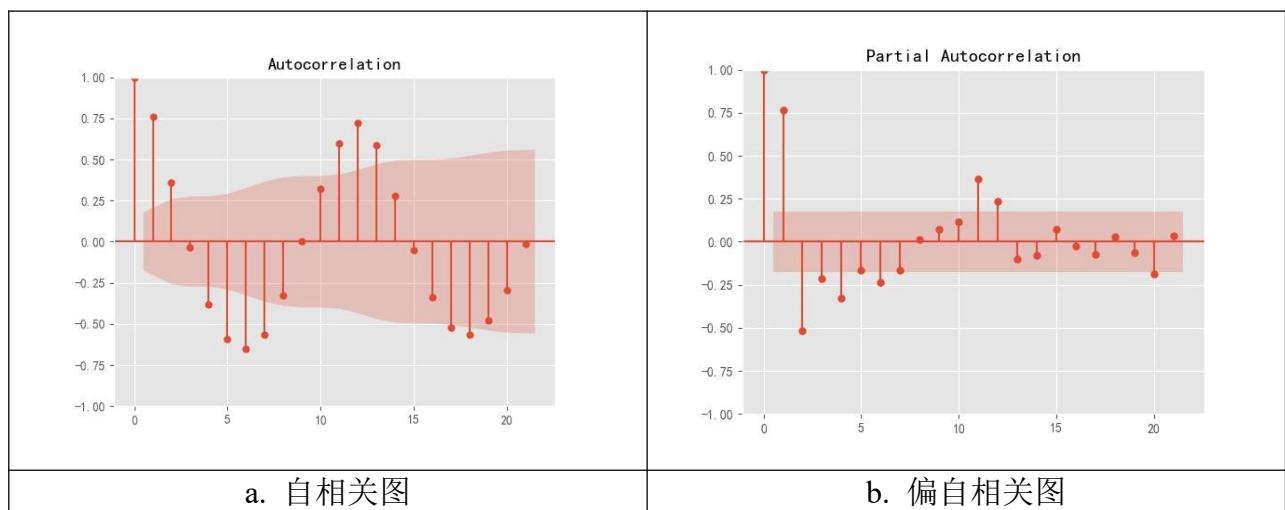


图 6 自相关图-偏自相关图

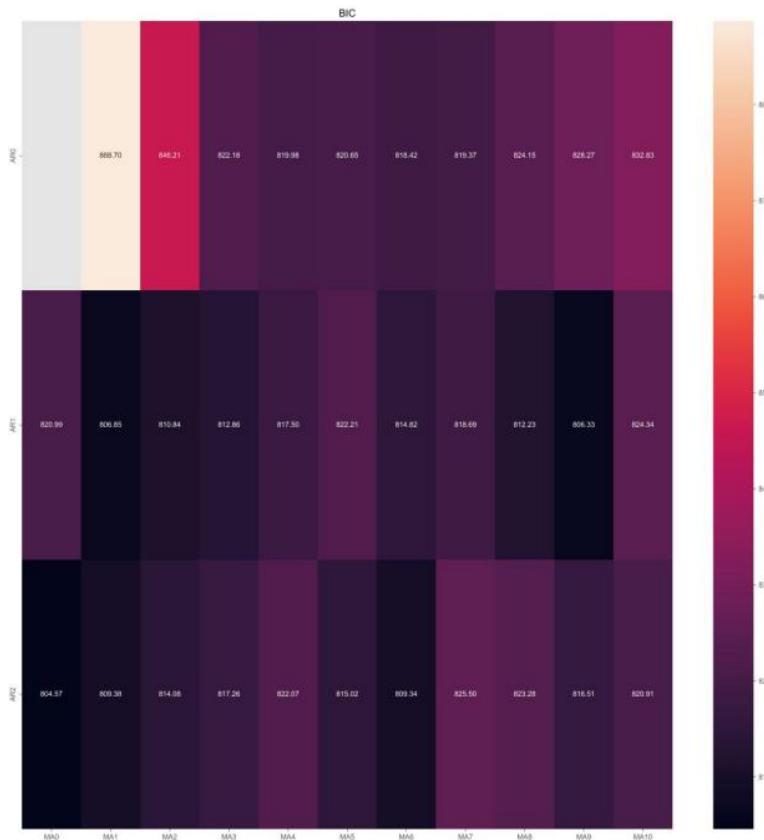


图 7 BIC 模型评估结果

Quantile-quantile plot, 也就是分位图(Q-Q 图), 在统计学上是一种通过比较两个概率分布的分位数从而实现对两个概率分布进行比较的概率图方法。如果两个概率分布相同, 那么它们的分位数也应该相同或者重叠在同一条直线上。也就是说重合度越高, 说明拟合的效果越好。通过图 8 揭示了锡林郭勒草原土壤蒸发量时间序列处理的 Q-Q 分位图, 表明利用 ARIMA 构建模型能够很好地对该时间序列进行拟合。

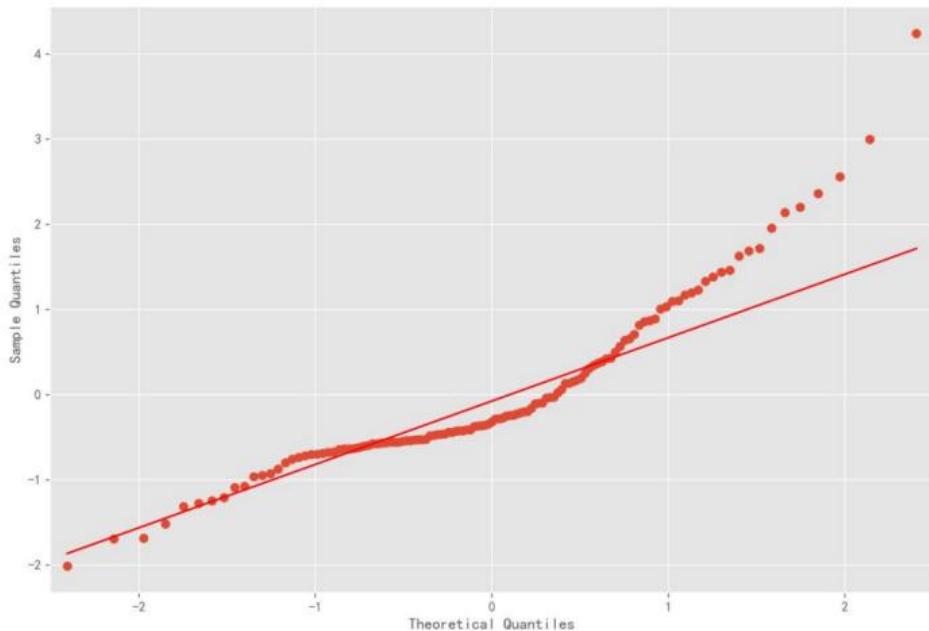


图 8 Q-Q 分位图

ARIMA 模型求解结果，如下表所示。

表 3 ARIMA 模型求解结果

Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
const	9.8924	1.935	5.112	0.000	6.100	13.685
ar.L1	1.1364	0.083	13.697	0.000	0.974	1.299
ar.L2	-0.4989	0.081	-6.171	0.000	-0.657	-0.340
sigma2	29.7312	3.754	7.920	0.000	22.374	37.089
Ljung-Box (L1) (Q):		1.22	Jarque-Bera (JB):		64.17	
Prob(Q):		0.27	Prob(JB):		0.00	
Heteroskedasticity (H):		0.64	Skew:		1.32	
Prob(H) (two-sided):		0.16	Kurtosis:		5.37	

最后，对锡林郭勒草原土壤蒸发量时间序列进行预测处理，预测下一个月的土壤蒸发效果如图 9 所示。

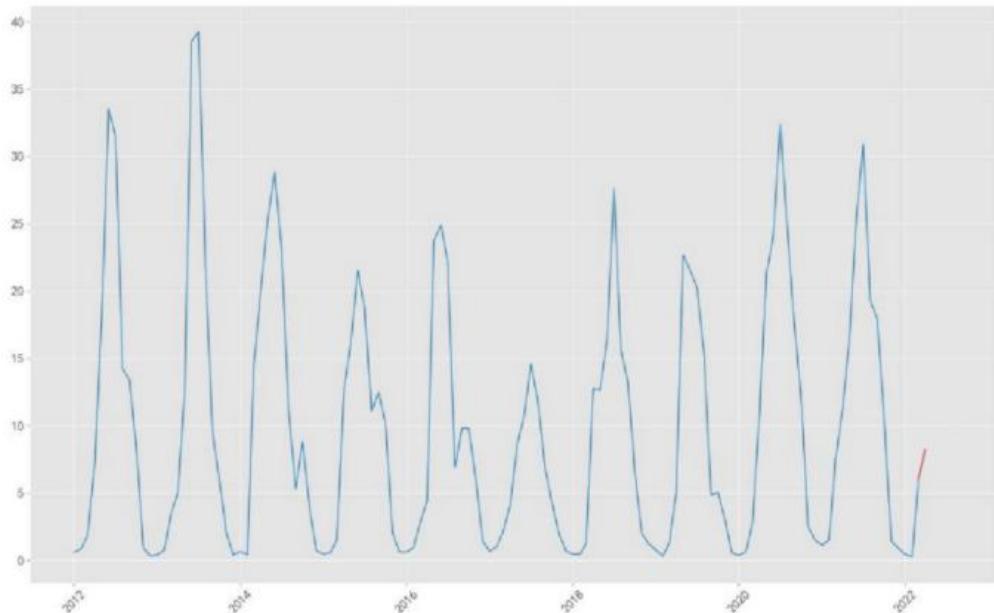


图 9 预测结果图

4.3.6 锡林郭勒草原降水量时序分析

锡林郭勒草原为高平原，海拔在 800-1200 米之间，属于中温带亚干旱大区。其气候类型属于大陆性气候中的温带半干旱草原气候。该地区年日照时数为 2603.8h，年平均温度 0.96°C、年均蒸发量 1664.6mm/年，平均降雨量 340mm 左右，降水月际间变异较大，主要集中在 6-9 月份，约占全年降水量 74.4%。表 4 是收集整理的锡林郭勒草原 2012-2022 年降水量情况。

表 4 锡林郭勒草原 2012-2022 年降水量情况

日期	2012-1	2012-2	2012-3	2012-4	2012-5	2012-6	2012-7	2012-8	2012-9	2012-10	2012-11	2012-12
	2.29	2.29	21.84	6.1	50.8	98.81	183.39	62.23	62.74	16.51	30.48	5.59
日期	2013-1	2013-2	2013-3	2013-4	2013-5	2013-6	2013-7	2013-8	2013-9	2013-10	2013-11	2013-12
	3.3	2.79	9.14	19.56	11.18	65.53	101.35	68.07	7.37	5.59	3.56	4.57
日期	2014-1	2014-2	2014-3	2014-4	2014-5	2014-6	2014-7	2014-8	2014-9	2014-10	2014-11	2014-12
	30.73	2.54	2.54	5.84	62.48	66.04	63.5	27.94	23.37	34.29	1.02	5.33
日期	2015-1	2015-2	2015-3	2015-4	2015-5	2015-6	2015-7	2015-8	2015-9	2015-10	2015-11	2015-12
	153.16	34.8	0.51	36.58	45.97	143.0	76.71	46.48	48.77	16.0	156.97	270.0
日期	2016-1	2016-2	2016-3	2016-4	2016-5	2016-6	2016-7	2016-8	2016-9	2016-10	2016-11	2016-12
	4.32	9.65	93.47	6.1	16.76	77.72	51.31	54.36	30.48	57.15	44.45	481.33
日期	2017-1	2017-2	2017-3	2017-4	2017-5	2017-6	2017-7	2017-8	2017-9	2017-10	2017-11	2017-12
	891.79	202.44	8.38	6.6	6.6	39.88	36.58	56.64	12.19	10.92	207.52	205.49
日期	2018-1	2018-2	2018-3	2018-4	2018-5	2018-6	2018-7	2018-8	2018-9	2018-10	2018-11	2018-12
	414.02	298.7	29.72	19.56	20.83	19.56	95.76	50.8	61.47	24.64	236.73	0.0
日期	2019-1	2019-2	2019-3	2019-4	2019-5	2019-6	2019-7	2019-8	2019-9	2019-10	2019-11	2019-12
	0.51	184.4	2.79	5.84	30.23	64.77	59.18	113.54	16.51	28.45	5.33	20.57
日期	2020-1	2020-2	2020-3	2020-4	2020-5	2020-6	2020-7	2020-8	2020-9	2020-10	2020-11	2020-12
	4.06	8.64	4.32	10.67	39.12	35.81	171.96	83.31	25.65	5.84	5.08	2.29
日期	2021-1	2021-2	2021-3	2021-4	2021-5	2021-6	2021-7	2021-8	2021-9	2021-10	2021-11	2021-12
	4.57	240.03	118.62	10.67	17.78	52.07	134.37	37.34	65.53	4.83	530.86	193.8
日期	2022-1	2022-2	2022-3	2022-4	2022-5	2022-6	2022-7					
降水量	607.31	796.04	115.57	3.3	13.21	27.43	29.21					

首先对锡林郭勒草原降水量进行描述和差分处理，如图 10 所示。根据单位根计算，在 $p=0.000 < 0.005$ 的情况下，如果 P 值小于 0.05 则拒绝原假设，不存在单位根，即序列平稳。

差分图

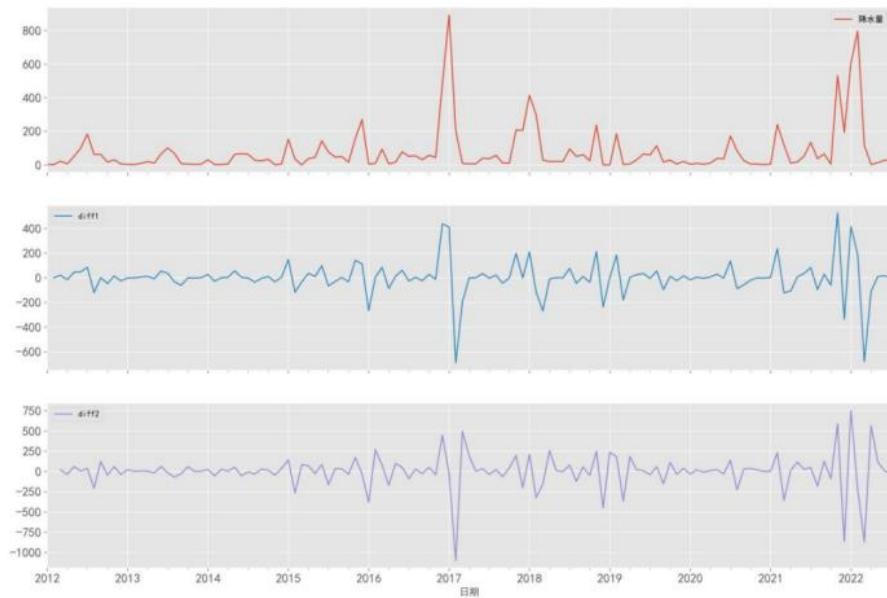


图 10 2012-2022 年锡林郭勒草原降水量时间序列、一阶差分和二阶差分图

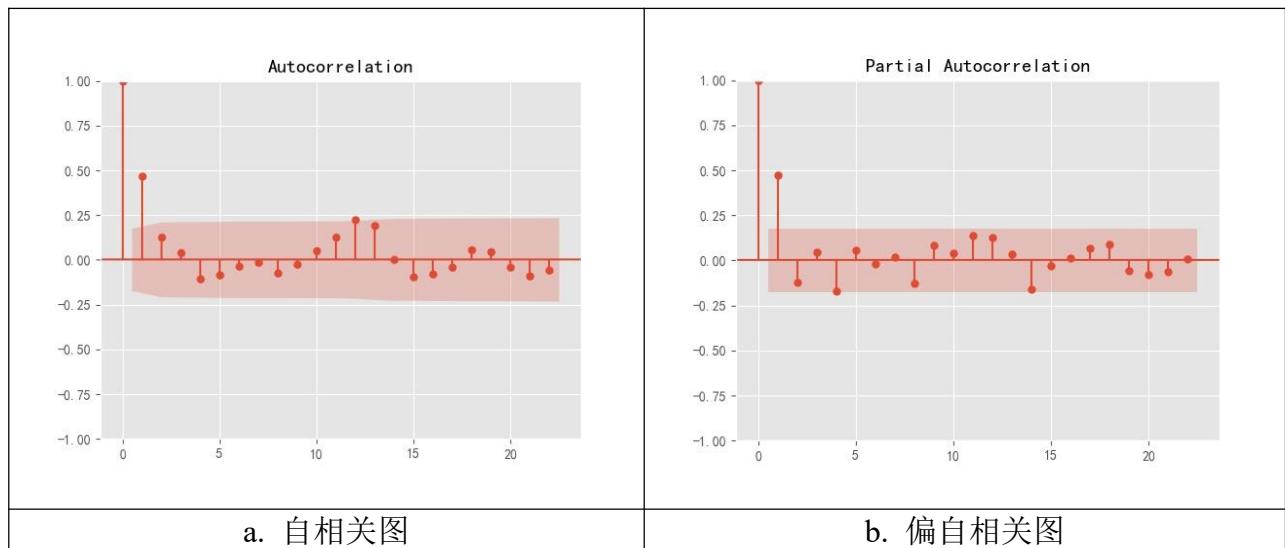


图 11 自相关图-偏自相关图

根据 BIC 模型评估结果，发现 ARIMA 模型在 AR1 和 MR0 的情况下表现最佳，故 ARIMA(p, n, q)中， p=1, n=2, q=0。

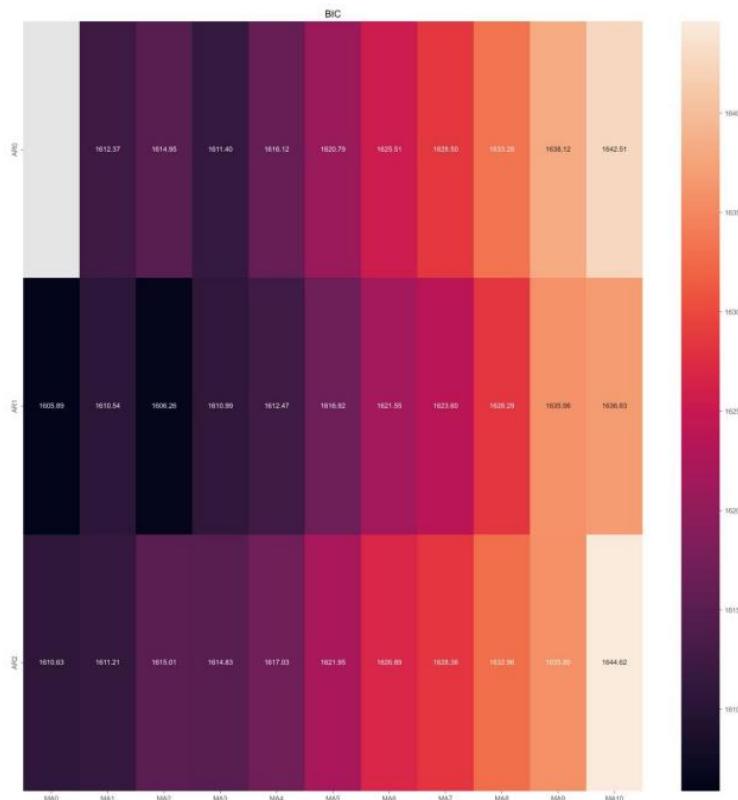


图 12 BIC 模型评估

此外，根据 Q-Q 分位图发现，通过锡林郭勒草原降水量时间序列具有较好地相关性。

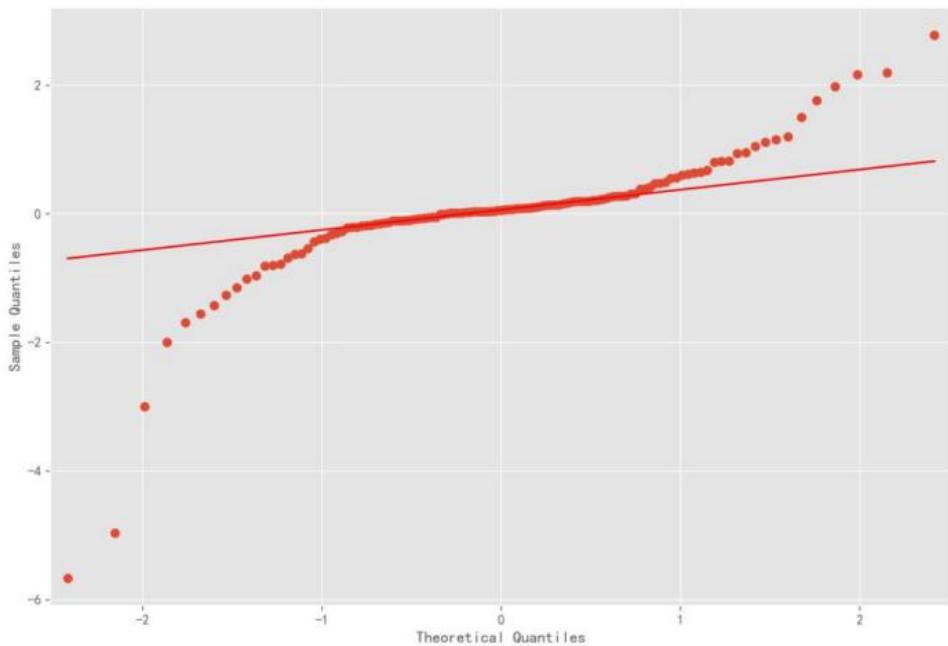


图 13 降水量时间序列 Q-Q 分位图

最后，本文利用 python 中 statsmodels 库中的 ARIMA 模型进行计算求解，其中对锡林郭勒草原降水量的 ARIMA 模型求解如表 5 所示。

表 5 降水量 ARIMA 模型求解

Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-1.4033	0.391	-3.587	0.000	-2.170	-0.637
ma.L2	0.4043	0.138	2.927	0.003	0.134	0.675
sigma2	1.981e+04	8636.259	2.294	0.022	2883.501	3.67e+04

Ljung-Box (L1) (Q):	1.40	Jarque-Bera (JB):
177.69		
Prob(Q):	0.24	Prob(JB):
Heteroskedasticity (H):	12.32	Skew:
Prob(H) (two-sided):	0.00	Kurtosis:

根据 ARIMA 模型进行下一个月份降水量的预测效果，如下图 14 所示。该图主要根据 2022 年 6 月和 7 月的数据预测 8 月降水情况。

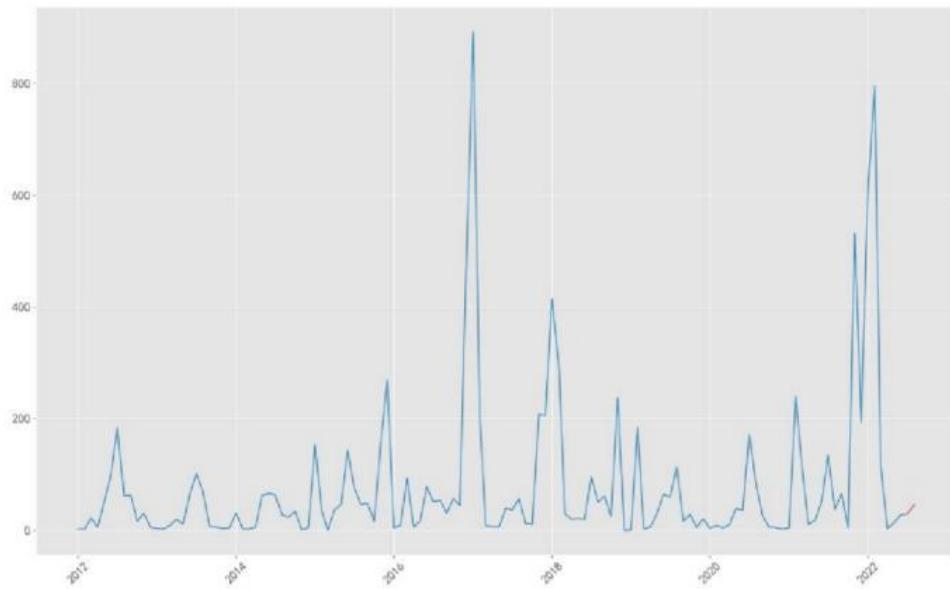


图 14 ARIMA 模型预测下一个月份降水量

4.4 多项式回归分析

4.4.1 多项式回归原理

多项式回归是统计学中回归分析的一种，其中自变量 x 和因变量 y 之间的关系被建模为关于 x 的 n 次多项式。多项式回归拟合 x 的值与 y 的相应条件均值之间的非线性关系，表示为 $E(y|x)$ ，并且已被用于描述非线性现象。

多项式回归使用的情况是当因变量 y 与自变量 x 为非线性关系时，线性关系不满足拟合状态，找不到合适的曲线拟合，可采用一元多项式回归 $f(x) = \theta_0 + \theta_1x + \theta_2x^2 + \theta_3x^3$ 通过增加 x 的高次项对实测点进行逼近，直到拟合结果到达满意条件。

多项式回归模型通常使用最小二乘法来拟合。在高斯-马尔可夫定理的条件下，最小二乘法最小化系数的无偏估计的方差^[5]。多项式回归的原理具体如下：

对于 n 阶多项式函数，可以表示为： $f(x) = \sum_{i=0}^n a_i * x^i$

对于数据集中有 n 个点 (x, y) ，都有：

$$\begin{cases} \hat{y}_1(x, w) = w_0 + w_1 * x_1 + w_2 * x_1^2 \dots + w_{m-1} * x_1^{m-1} + w_m * x_1^m \\ \hat{y}_2(x, w) = w_0 + w_1 * x_2 + w_2 * x_2^2 \dots + w_{m-1} * x_2^{m-1} + w_m * x_2^m \\ \dots \\ \hat{y}_n(x, w) = w_0 + w_1 * x_n + w_2 * x_n^2 \dots + w_{m-1} * x_n^{m-1} + w_m * x_n^m \end{cases} \quad (4-17)$$

将其写为矩阵形式：

$$\hat{Y} = XW \quad (4-18)$$

其中， $W = (w_0, w_1, w_2, \dots, w_{m-1}, w_m)^T$ ，为 $(m+1) * 1$ 的矩阵， X 为 $n * (m+1)$ 的矩阵

$$X = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ x_1^m & \dots & x_n^m \end{bmatrix}^T \quad (4-19)$$

根据上述给定的多项式模型，通常使用利用最小二乘法来求解 W 。具体做法如下：

对于矩阵 $Ax = b$ ，其中 $b \in R^{m*1}$ ， $A \in R^{m*n}$ ，且 $m > n$ ，假定 b 存在加行观测误差或噪声，即 $b = b_0 + e$ ，其中 b_0 为无误差数据向量， e 为误差向量。

为了抵消误差对矩阵方程求解的影响，使用 Δb ，使得 $b + \Delta b = b_0 + e + \Delta b \rightarrow b_0$ 从而实现公式 (4-20) 的转换

$$Ax = b + \Delta b \rightarrow Ax = b_0 \quad (4-20)$$

即意味着，选择校正向量 $\Delta b = Ax - b$ ，使校正向量尽可能小，可以实现无误差的矩阵方程 $Ax = b_0$ 的求解。其可以描述为，

$$\min_x \|\Delta b\|^2 = \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) \quad (4-21)$$

该方法被称为普通最小二乘法（Ordinary least squares, OLS）法，也称为最小二乘法。于是，矩阵方程的 $Ax = b$ 的最小二乘解为，

$$\hat{x}_{LS} = argmin_x \|Ax - b\|_2^2 \quad (4-22)$$

展开 $\emptyset = (Ax - b)^T(Ax - b)$ 得，

$$\emptyset = x^T A^T A x - x^T A^T b - b^T A^T x + b^T b \quad (4-23)$$

求 \emptyset 相对于 x 的导数，并让结果等于零，则有，

$$\frac{d\emptyset}{dx} = 2A^T A x - 2A^T b = 0 \quad (4-24)$$

即， x 满足公式 (4-25)

$$A^T A x = A^T b \quad (4-25)$$

当矩阵 $A_{m \times n}$ 具有不同的秩时，上述方程解不同：

- (1) 对于正定方程 ($m > n$) 列满秩时，即 $\text{rank}(A)=n$ ，
- (2) 由于 $A^T A$ 非奇异，所以方程有唯一解，

$$x_{LS} = (A^T A)^{-1} A^T b \quad (4-26)$$

- (3) 对于不满秩 ($\text{rank}(A) < n$) 的正定方程，则最小二乘的解为，

$$x_{LS} = (A^T A)^{\dagger} A^T b \quad (4-27)$$

其中， \dagger 表示该矩阵的 Moore-Penrose 逆矩阵，即伪逆。

4.4.2 锡林郭勒草原叶面积分指数 LAI 多项式回归分析

通过提取附件 10 中锡林郭勒草原 2012-2017 年的叶面积指数，数据整理发现每年的叶面指数只在当年的月份间发生变化，不同年份直接基本没有差异。因此，叶面积指数 $LAI(t)$ 与时间的关系主要在年份内有关，并具有明显的分布特征。

表 6 2012-2017 年锡林郭勒草原叶面积指数 LAI 分布情况

日期	2012-1	2012-2	2012-3	2012-4	2012-5	2012-6	2012-7	2012-8	2012-9	2012-10	2012-11	2012-12
LAI	0.636	0.652	0.717	0.727	0.739	0.827	0.964	0.983	0.847	0.738	0.71	0.648
日期	2013-1	2013-2	2013-3	2013-4	2013-5	2013-6	2013-7	2013-8	2013-9	2013-10	2013-11	2013-12
LAI	0.636	0.651	0.717	0.727	0.739	0.827	0.964	0.983	0.847	0.738	0.71	0.648
日期	2014-1	2014-2	2014-3	2014-4	2014-5	2014-6	2014-7	2014-8	2014-9	2014-10	2014-11	2014-12
LAI	0.636	0.651	0.717	0.727	0.739	0.827	0.964	0.983	0.847	0.738	0.71	0.648
日期	2015-1	2015-2	2015-3	2015-4	2015-5	2015-6	2015-7	2015-8	2015-9	2015-10	2015-11	2015-12
LAI	0.636	0.651	0.717	0.727	0.739	0.827	0.964	0.983	0.847	0.738	0.71	0.648
日期	2016-1	2016-2	2016-3	2016-4	2016-5	2016-6	2016-7	2016-8	2016-9	2016-10	2016-11	2016-12
LAI	0.636	0.652	0.717	0.727	0.739	0.827	0.964	0.983	0.847	0.738	0.71	0.648
日期	2017-1	2017-2	2017-3	2017-4	2017-5	2017-6	2017-7	2017-8	2017-9	2017-10	2017-11	2017-12
LAI	0.636	0.651	0.717	0.727	0.739	0.827	0.964	0.983	0.847	0.738	0.71	0.648

通过对 2017 年锡林郭勒草原每个月的叶面积指数统计分析，如图 15，发现其分布情况呈现非线性，故选择多项式回归分析。

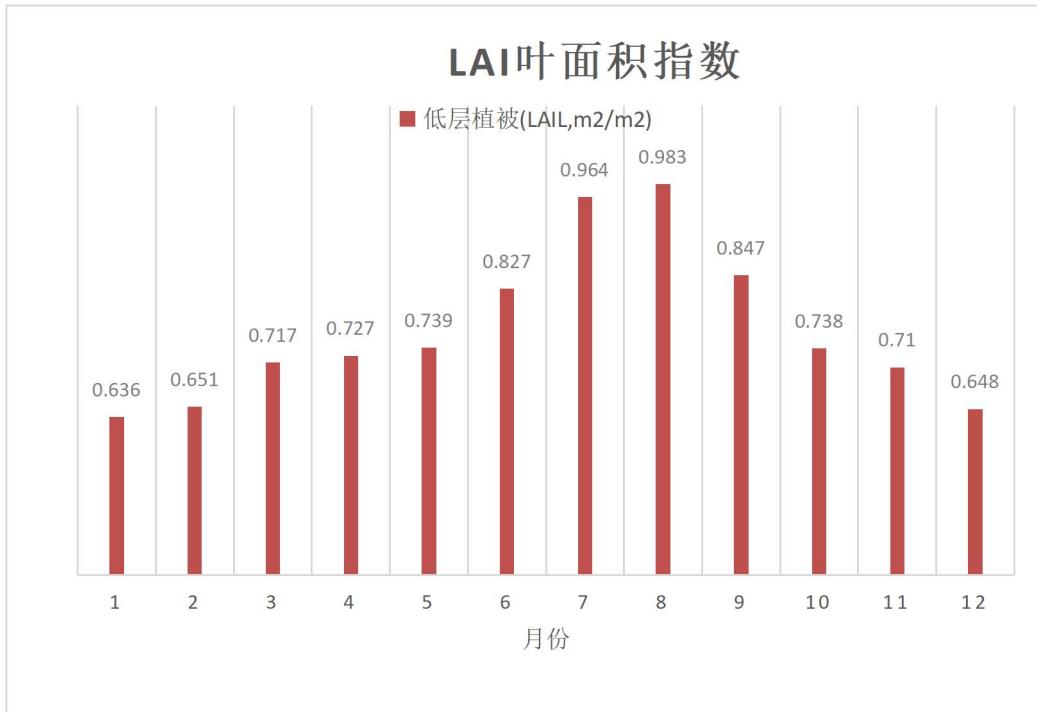


图 15 1-12 月份叶面积指数 LAI

本文主要利用 python 工具中 Polynomial Features 函数进行多项式回归模型求解。

表 7 LAI 多项式求解结果

参数	值
回归截距:	w0=[0.61751515]
回归系数:	w1=[[0. -0.00443645 0.01395321 -0.0011418]]
R2 确定系数:	0.7801
均方误差:	0.0027
平均绝对值误差:	0.0423
中位绝对值误差:	0.0279

根据表 7, 可知叶面指数 LAI(t) 为公式 (4-28),

$$LAI(t) = 0.6175 - 0.0044 * t + 0.014 * t^2 - 0.0011 * t^3 \quad (4-28)$$

通过非线性多项式回归模型拟合得到 1-12 月份的叶面指数分布情况, 如图 16 所示。

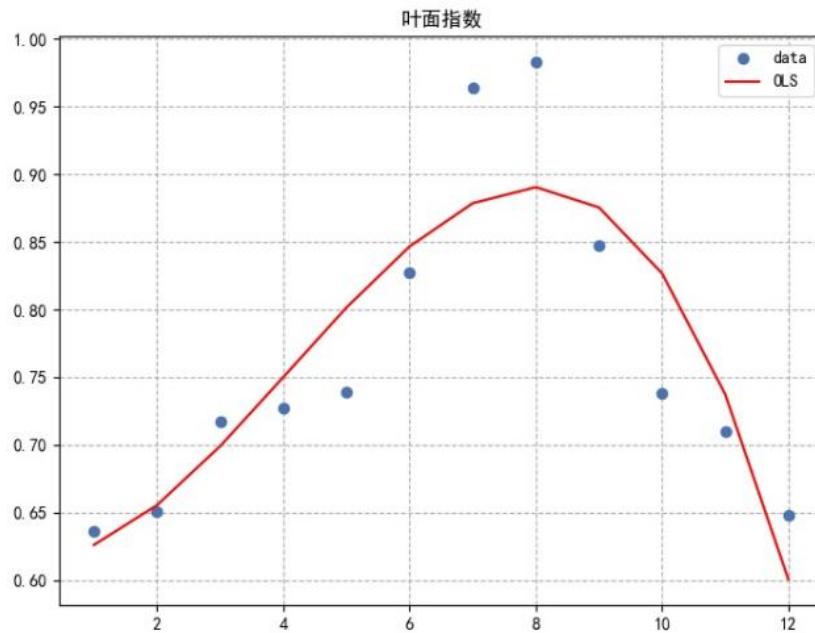


图 16 叶面指数多项式回归拟合效果

4.4.3 锡林郭勒草原绿植覆盖率多项式回归模型求解

由于每年的 11 月份到来年的 3 月份为大雪覆盖天气，这里没有植被覆盖，即将空值设为 0。同时，每月采集数据 3 次，因此将数据以月份为单位求取平均值。如此，数据预处理后的植被覆盖情况如下表 8 所示，数据预处理后的统计分布情况如图 17。选择的数据为采集数据完整的 2021 年的绿植覆盖率数据作为多项式模型数据输入，拟合求解。

表 8 2021 年锡林郭勒草原植被覆盖率数据预处理

时间	绿植覆盖率	处理后绿植覆盖率
2021-01	NAN	0.0000
2021-02	NAN	0.0000
2021-03	0.0147	0.0147
2021-04	0.0387	0.0387
2021-05	0.0813	0.0813
2021-06	0.1573	0.1573
2021-07	0.3227	0.3227
2021-08	0.3627	0.3627
2021-09	0.2533	0.2533
2021-10	0.0880	0.0880
2021-11	NAN	0.0000
2021-12	NAN	0.0000

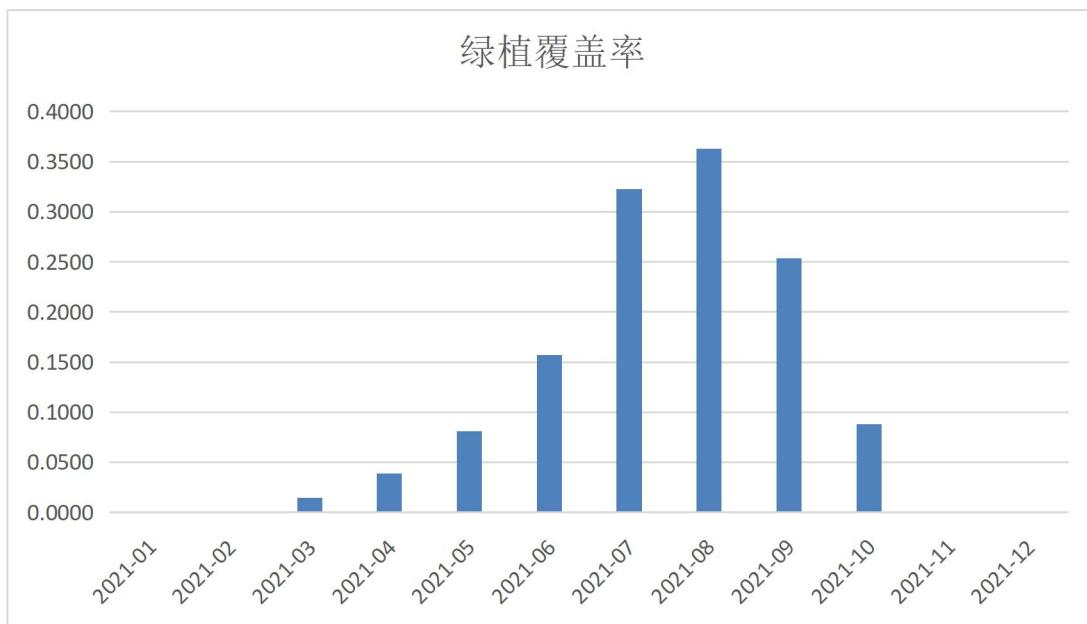


图 17 2021 年锡林郭勒草原的绿植覆盖率分布情况

利用 python 处理绿植覆盖率数据，对数据分布情况作多项式回归分析，构建多项式回归模型，并求解模型参数。通过数据处理，计算多项式回归模型参数，如表 9 所示。

表 9 绿植覆盖率多项式回归模型参数求解

参数	值
回归截距:	w0=[0.02627609]
回归系数:	w1=[[0. -0.06199073 0.02523625 -0.00173651]]
R2 确定系数:	0.7460
均方误差:	0.0041
平均绝对值误差:	0.0516
中位绝对值误差:	0.0406

根据表 9 多项式回归模型求解中，回归截距 $w0=[0.02627609]$ ，回归系数 $w1=[[0. -0.06199073 0.02523625 -0.00173651]]$ ；根据 R2 确定系数 0.7460，均方误差 0.0041，和平均绝对值误差 0.0516 和中位绝对值误差 0.0406，判断多项式回归分析效果显著。因此，可知叶面指数 $c_p(t)$ 为公式 (4-29)。

$$c_p(t) = 0.0263 - 0.0620 * t + 0.0252 * t^2 - 0.0017 * t^3 \quad (4-29)$$

根据多项式回归模型求解参数结果，绘制回归分析效果图，如下图 18 所示。结果表明，回归分析模型能够较好地将锡林郭勒草原每月份的绿植覆盖率预测出来。

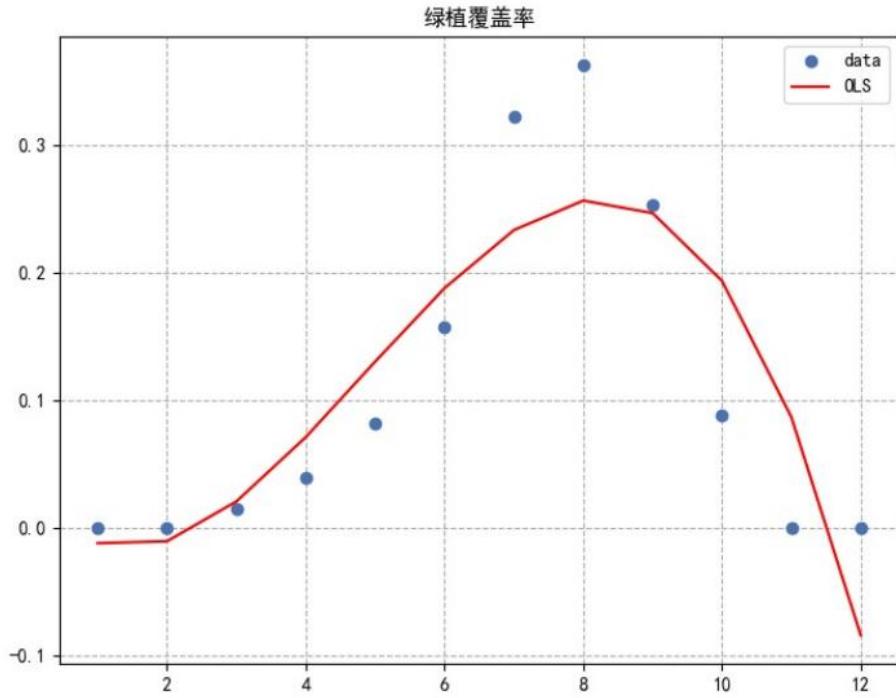


图 18 绿植覆盖率多项式回归拟合效果

4.4.4 放牧强度对土壤湿度和植被生物量影响的定性分析

经过机理分析，查找相关文献，总结出内蒙古锡林郭勒草原上不同放牧策略对会被生物量和土壤湿度的影响关系，如图 10。首先，根据公式（4-30），获取生物量与放牧强度的关系。其中，放牧强度分为：对照（NG，0 羊/天/公顷）、轻度放牧强度（LGI，2 羊/天/公顷）、中度放牧强度（MGI，4 羊/天/公顷）和重度放牧强度（HGI，8 羊/天/公顷）。实际中，也可以做如下的划分：对照（NG，0 羊/天/公顷）、轻度放牧强度（LGI，1-2 羊/天/公顷）、中度放牧强度（MGI，3-4 羊/天/公顷）和重度放牧强度（HGI，5-8 羊天/公顷）。

$$\frac{dw(t)}{dt} = 0.049w(t)(1 - \frac{w(t)}{4000}) - 0.0047Sw(t) \quad (4-30)$$

然后，根据公式土壤-植被-大气系统的水平衡基本方程^[6-7]，确定土壤湿度与降水量、蒸发量、地下水毛管上升量、土壤水渗透量、土壤蒸发量、入和出径流量，以及植被截流量有关。草原水分通过降水、入渗、蒸发等形式周而复始的循环，在无人干扰的情况下，土壤-植被-大气系统的水平衡基本方程为公式（4-31）。

$$\Delta W = W_{t+1} - W_t = P + G_u + R_{in} - (E_{ta} + G_d + R_{out} + IC_{store}) \quad (4-31)$$

此外，根据公式（4-32）能够判断绿植覆盖率与土壤湿度、植被生物量存在关联关系，

$$c_p(t) = [a^* - W(t)(e^{\varepsilon_g w(t)/w^*} - 1)](1 - e^{-\varepsilon_g w(t)/w^*}) \quad (4-32)$$

根据公式（4-33）和公式（4-34）可知，植被截流率与特定绿植覆盖率、降水量、最大植被截流量直接相关，其中叶面积指数通过影响最大植被截流量间接关系到植被截流率。

$$IC_{store}(t) = c_p(t) \cdot IC_{max}(t) \cdot [1 - \exp(-k \cdot P(t)/IC_{max}(t))] \quad (4-33)$$

$$IC_{max}(t) = 0.935 + 0.498 \cdot LAI(t) - 0.00575 \cdot LAI(t)^2 \quad (4-34)$$

因此，土壤湿度受到植被截流率进而受到植被覆盖率的影响，通过植被覆盖率关系到植被生物量。因为，植被生物量与放牧强度有关，所以土壤湿度受到放牧强度的影响。最终影响关系如图 19 所示。

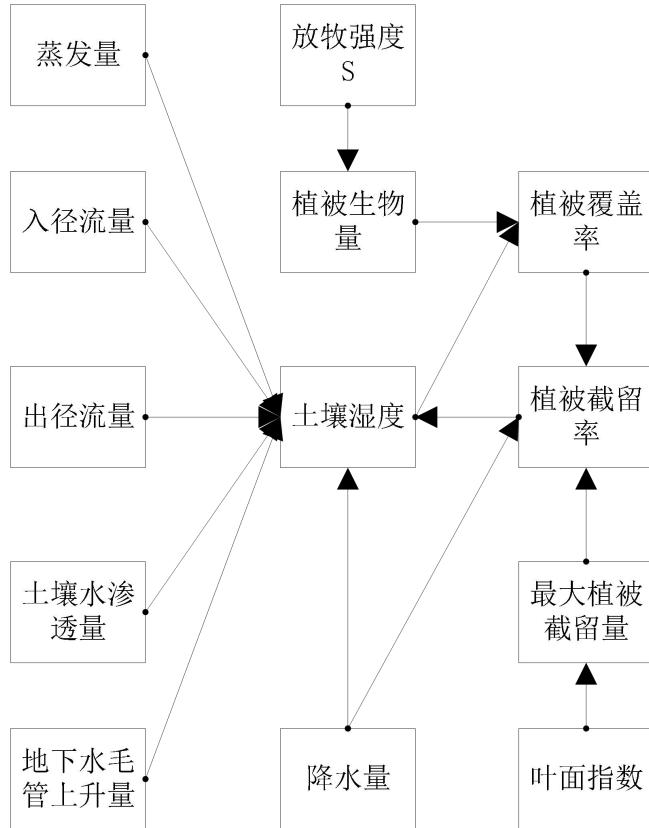


图 19 放牧强度对土壤湿度、植被生物量的影响关系

4.5 总结

本章节建立了不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤物理性质（主要是土壤湿度）和植被生物量影响的数学模型。根据上文对未知变量的预测和回归分析，可以获取放牧强度对土壤湿度和植被生物量的影响，其影响关系表达如公式（4-35）。

$$\left\{ \begin{array}{l} \frac{dw(t)}{dt} = 0.049w(t)(1 - \frac{w(t)}{4000}) - 0.0047Sw(t) \\ \frac{dW(t)}{dt} = P(t) - E(t) - IC_{store}(t) + R_{in} - R_{out} - G_d + G_u \\ c_p(t) = [a^* - W(t)(e^{\varepsilon_g w(t)/w^*} - 1)](1 - e^{-\varepsilon_g w(t)/w^*}) \\ IC_{store}(t) = c_p(t) \cdot IC_{max}(t) \cdot [1 - \exp(-k \cdot P(t)/IC_{max}(t))] \\ IC_{max}(t) = 0.935 + 0.498 \cdot LAI(t) - 0.00575 \cdot LAI(t)^2 \end{array} \right. \quad (4-35)$$

此外，本章节还做了以下工作：

- (1) 通过对锡林郭勒草原土壤蒸发量进行时间序列分析，发现土壤蒸发量序列为平稳序列，并利用 ARIMA 模型进行时序分析和预测。
- (2) 锡林郭勒地区降水量时间序列平稳，其中降水月际间变异较大，主要集中在 6-9 月份；并且和土壤蒸发量一样可以通过 ARIMA 模型进行时间序列分析和预测。
- (3) 叶面积指数 LAI 只在年内的月份间发生变化，不同年份之间基本没有差异。通过统计发现，可以通过多项式非线性回归分析。

$$LAI(t) = 0.6175 - 0.0044 * t + 0.014 * t^2 - 0.0011 * t^3 \quad (4-36)$$

(4) 对锡林郭勒草原绿植覆盖率进行统计分析,发现一年之中6-9月份的植被是非常茂盛的,每年的1-3和11-12月份处于冬季,植物凋零,植被覆盖率基本为0。此外,通过统计分布情况发现,其呈非线性分布;利用多项式回归分析,能够拟合绿植覆盖率随着时间变化的趋势图。

$$c_p(t) = 0.0263 - 0.0620 * t + 0.0252 * t^2 - 0.0017 * t^3 \quad (4-37)$$

五、问题二的求解：多元回归模型预测土壤湿度

5.1 问题分析和技术路线图

问题二需要建立一些模型对保持目前放牧策略不变情况下对 2022 年、2023 年不同深度土壤湿度进行预测。根据问题一中土壤湿度与降水量、土壤蒸发量、植被截流量、水毛管上升量、土壤水渗透量和径流量等有关，因此在保持放牧策略不变的情况下，植被生物量确定（根据问题一放牧方式与植被生物量关系公式）。由于植被的截流量与植被生物量有关其中，假设内蒙古土壤水毛管上升量和土壤水渗透量为特定量，同时在草原中径流量的出入相等。因此，因此影响土壤湿度的关键指标为土壤蒸发量和降水量。

预测 2022 年、2023 年不同深度土壤湿度，首先预测土壤蒸发量和降水量，进而根据拟合的土壤蒸发量和降水量对土壤湿度的关系，获取指定时间的土壤湿度数据。针对以上分析和处理过程，研究路线图如图 20。

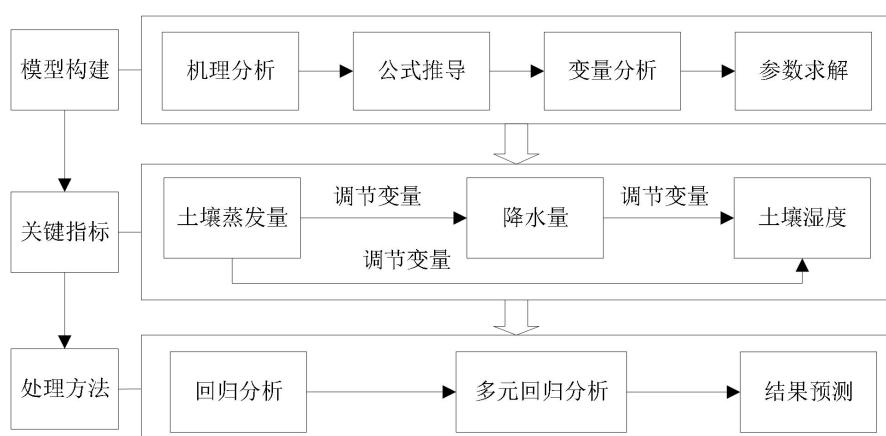


图 20 问题二技术路线图

5.2 模型建立

针对该问题，需要确定土壤湿度与其他影响因素的关系。根据水平衡公式（5-1），在放牧策略不变的情况下，土壤湿度关系到降水量、蒸发量两个主要指标。

$$\frac{dW}{dt} = P(t) - E(t) - IC_{store}(t) + R_{in} - R_{out} - G_d \quad (5-1)$$

5.3 回归模型求解

根据问题一种使用的回归分析方法，对 $P(t)$ 和 $E(t)$ 降水量和蒸发量按指定月份拟合，以预测 2022 年和 2023 年对应月份的降水量和蒸发量。在预测到指定月份的降水量和蒸发量后，建立并求解降水量和蒸发量这两个关键指标对土壤湿度的影响，在放牧策略不变的情况下。通过 OLS 最小二乘法求解模型参数，最终预测和分析不同深度土壤湿度情况。

5.3.1 锡林郭勒草原土壤蒸发量回归分析

利用附件 4 中的土壤蒸发量数据，统计不同月份在各年的分布情况，数据处理如下图 21 所示。根据图 21 发现，5-9 月份的土壤蒸发量是比较高的，这可能与锡林郭勒地区日照时间长，夏季温度高的原因导致的。



图 21 2012-2022 年各月份锡林郭勒草原土壤蒸发量分布情况

这里本文使用回归的方式拟合出土壤蒸发量的预测模型，进而对 2022 年和 2023 年指定月份进预测。拟合结果如下图 22 所示。

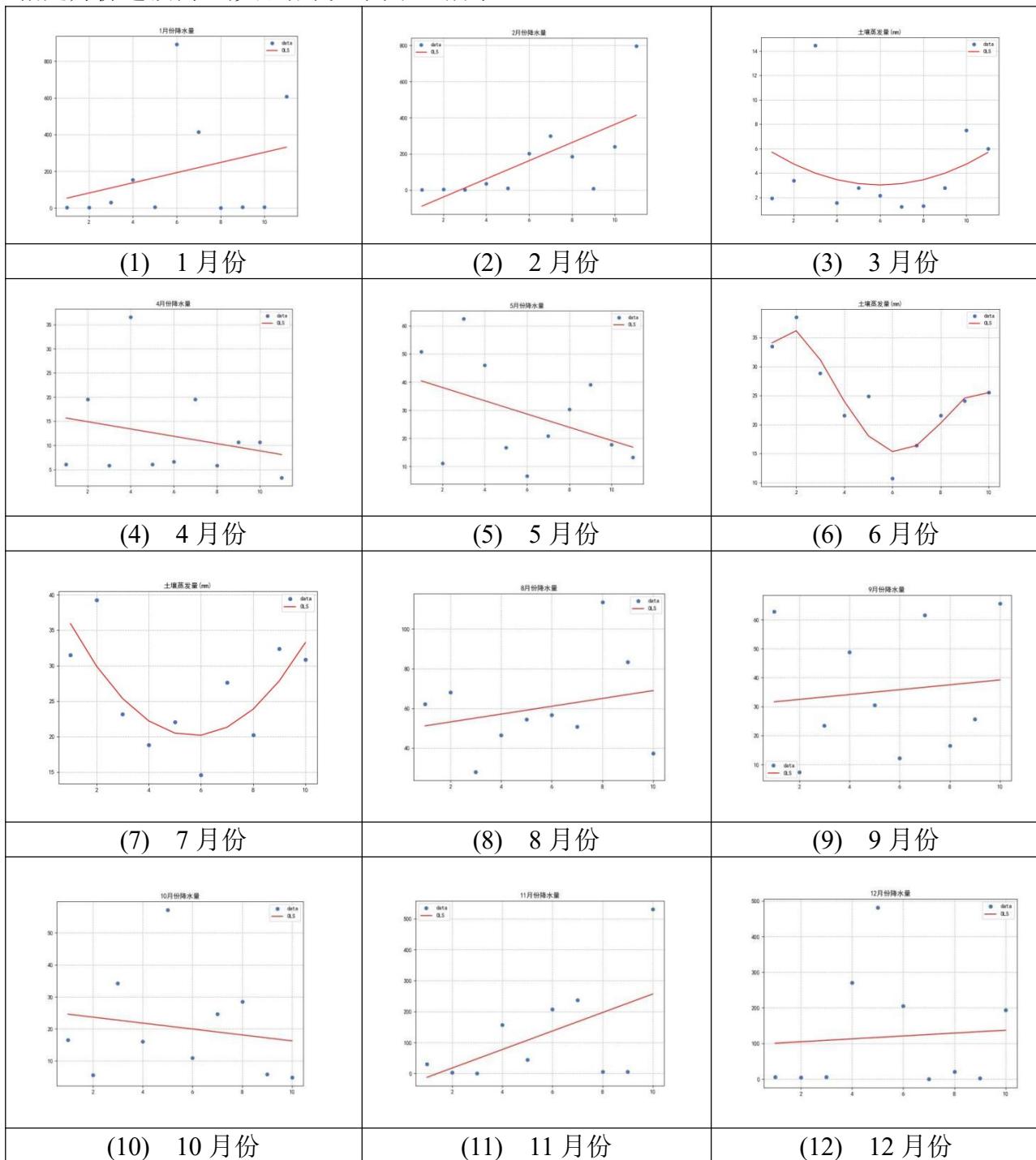


图 22 1-12 月降水量

根据锡林郭勒各月份对应各年的降水量回归模型，对 2022-2023 年指定月份的降水量进行预测。预测结果如下表 10 所示。

表 10 2022-2023 年锡林郭勒草原降水量预测结果

年份	月份	蒸发量 (mm)
2022	4	9.37
	5	17.99
	6	16.6

	7	24.4
	8	18.68
	9	14.24
	10	9.08
	11	2.58
	12	1.29
<hr/>		
2023	1	0.69
	2	0.69
	3	4.08
	4	9.38
	5	18.04
	6	15.15
	7	24.11
	8	19.32
	9	14.81
	10	9.25
	11	2.57
	12	1.37

5.3.2 锡林郭勒草原降水量回归分析

选择附件 8 中锡林郭勒 2012-2022 年降水数据，整理出每个月份的各年降水情况，如下图 23 所示。锡林郭勒地区的降水量主要集中在 5-10 月份，冬季的降水量普遍比较少。



图 23 2012-2022 年各月份锡林郭勒草原降水量分布

接着，对每个月份的各年降水情况进行回归分析，找出降水量与时间的关系。回归分析结果如下图 24 所示。

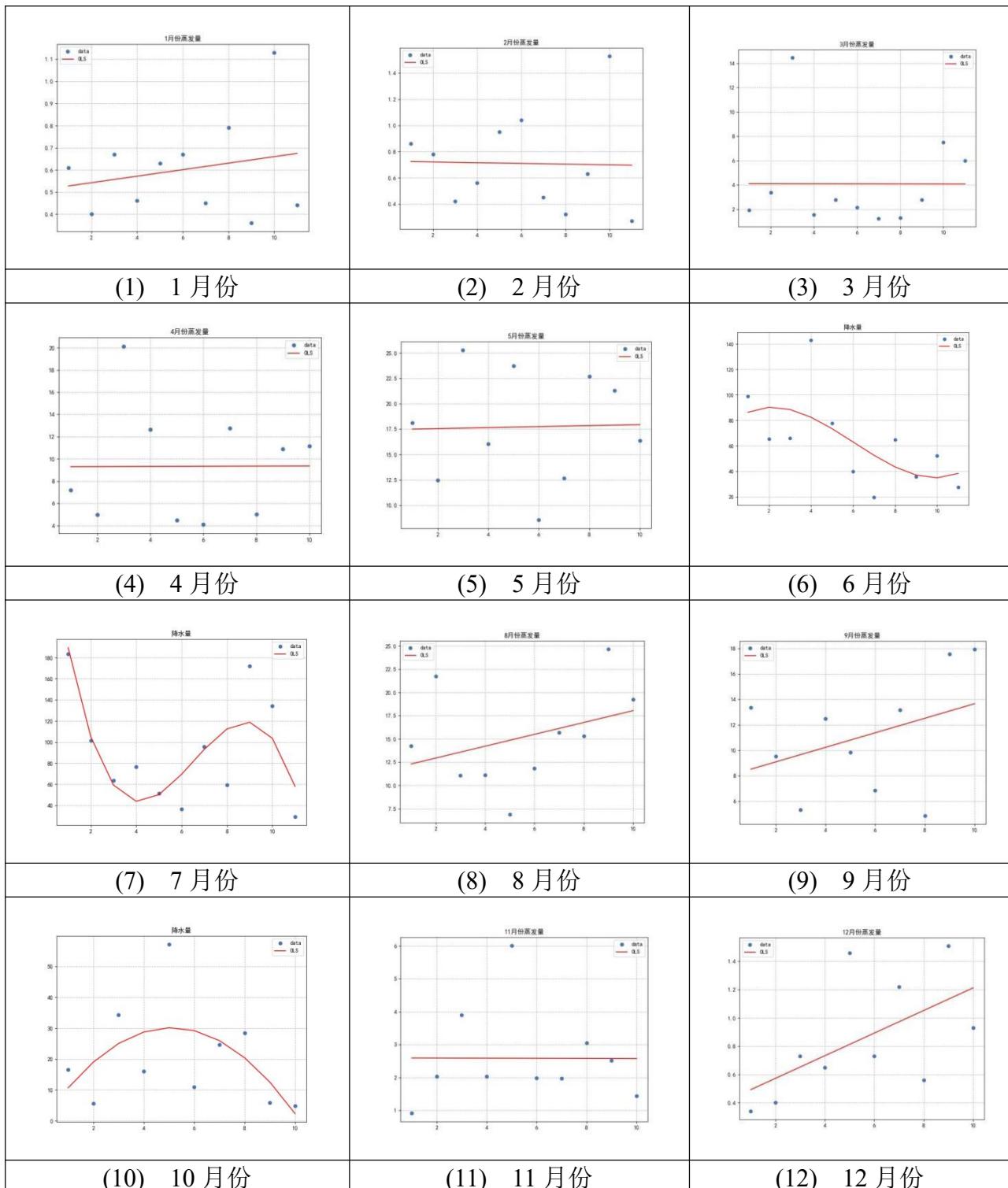


图 24 锡林郭勒 2012-2022 年 1-12 月蒸发量

在建立各月份在各年的降水量回归模型的基础上，对 2022-2023 年指定月份的降水量进行预测，预测结果如下表 11 所示。

表 11 降水量和蒸发量预测结果

年份	月份	降水量	蒸发量 (mm)
2022	4	3.3	9.37
	5	13.21	17.99
	6	27.43	16.6

	7	29.21	24.4
	8	70.93	18.68
	9	40.03	14.24
	10	22.26	9.08
	11	286.83	2.58
	12	140.62	1.29
2023	1	359.0	0.69
	2	463.34	0.69
	3	83.5	4.08
	4	7.36	9.38
	5	14.5	18.04
	6	23.72	15.15
	7	74.68	24.11
	8	72.91	19.32
	9	40.87	14.81
	10	14.4	9.25
	11	316.72	2.57
	12	144.62	1.37

5.4 多元回归分析和预测土壤湿度

根据附件3土壤湿度数据、附件4土壤蒸发数据以及附件8中降水等数据，建立对应各时间的降水量、蒸发量和土壤湿度数据，如下图25所示。

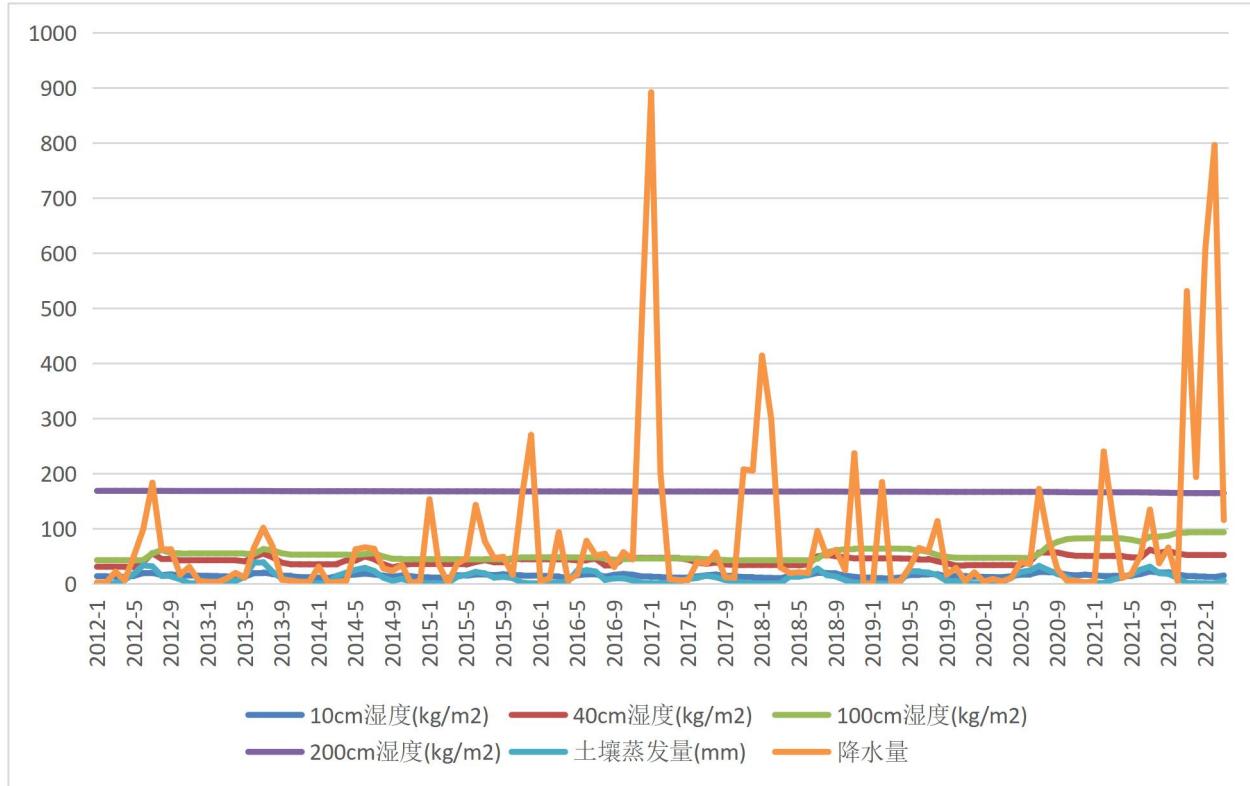


图 25 2012-2022 年锡林郭勒降水量、土壤蒸发量和土壤湿度

5.4.1 10 厘米深度的土壤湿度多元回归分析

对 10 厘米深度的土壤湿度与土壤蒸发量和降水量进行多元回归分析，通过最小二乘法求解参数结果如图 26 所示。结果显示 R^2 为 0.664 时，土壤蒸发量的比例系数为 0.2390，降水量的比例系数为 $0.5e^{-2}$ 。因此，10cm 土壤与土壤蒸发量和降水量的关系式为，

$$W(t) = 12.1671 + 0.2390 * P(t) + 0.5e^{-2} * E(t) \quad (5-2)$$

OLS Regression Results						
Dep. Variable:	10cm湿度(kg/m ²)	R-squared:	0.664			
Model:	OLS	Adj. R-squared:	0.658			
Method:	Least Squares	F-statistic:	118.6			
Date:	Fri, 07 Oct 2022	Prob (F-statistic):	3.81e-29			
Time:	10:20:45	Log-Likelihood:	-235.75			
No. Observations:	123	AIC:	477.5			
Df Residuals:	120	BIC:	485.9			
Df Model:	2					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
const	12.1671	0.239	50.890	0.000	11.694	12.641
土壤蒸发量(mm)	0.2390	0.016	15.320	0.000	0.208	0.270
降水量	0.0005	0.001	0.474	0.637	-0.002	0.003
Omnibus:	5.748	Durbin-Watson:	0.720			
Prob(Omnibus):	0.056	Jarque-Bera (JB):	5.675			
Skew:	0.483	Prob(JB):	0.0586			
Kurtosis:	2.581	Cond. No.	261.			

图 26 10 厘米深度下的 OLS 回归分析结果

5.4.2 40 厘米深度的土壤湿度多元回归分析

对 40 厘米深度的土壤湿度与降水量、蒸发量进行多元回归分析。通过最小二乘法求解回归模型参数结果如图 27 所示。因此，40cm 土壤与土壤蒸发量和降水量的关系式为，

$$W(t) = 38.2686 + 0.2861 * P(t) + 0.0160 * E(t) \quad (5-3)$$

OLS Regression Results						
Dep. Variable:	40cm湿度(kg/m ²)	R-squared:	0.202			
Model:	OLS	Adj. R-squared:	0.189			
Method:	Least Squares	F-statistic:	15.22			
Date:	Fri, 07 Oct 2022	Prob (F-statistic):	1.29e-06			
Time:	10:23:13	Log-Likelihood:	-407.58			
No. Observations:	123	AIC:	821.2			
Df Residuals:	120	BIC:	829.6			
Df Model:	2					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
const	38.2686	0.967	39.590	0.000	36.355	40.182
土壤蒸发量(mm)	0.2861	0.063	4.534	0.000	0.161	0.411
降水量	0.0160	0.004	3.714	0.000	0.007	0.024
Omnibus:	14.200	Durbin-Watson:	0.532			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	5.321			
Skew:	0.215	Prob(JB):	0.0699			
Kurtosis:	2.076	Cond. No.	261.			

图 27 40 厘米深度下的 OLS 回归分析结果

5.4.3 100 厘米深度的土壤湿度多元回归分析

对 100 厘米深度的土壤湿度与降水量、蒸发量进行多元回归分析，以求去在放牧

策略不变的情况下降水量和蒸发量对土壤湿度的影响。通过 OLS 求解回归分析结果，如下图 28 所示。因此，100cm 土壤与土壤蒸发量和降水量的关系式为，

$$W(t) = 52.2547 + 0.1085 * P(t) + 0.0251 * E(t) \quad (5-4)$$

OLS Regression Results						
Dep. Variable:	100cm湿度(kg/m ²)	R-squared:	0.061			
Model:	OLS	Adj. R-squared:	0.046			
Method:	Least Squares	F-statistic:	3.923			
Date:	Fri, 07 Oct 2022	Prob (F-statistic):	0.0224			
Time:	10:30:58	Log-Likelihood:	-499.36			
No. Observations:	123	AIC:	1005.			
Df Residuals:	120	BIC:	1013.			
Df Model:	2					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
const	52.2547	2.039	25.633	0.000	48.218	56.291
土壤蒸发量(mm)	0.1085	0.133	0.816	0.416	-0.155	0.372
降水量	0.0251	0.009	2.764	0.007	0.007	0.043
Omnibus:	21.629	Durbin-Watson:	0.288			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26.980			
Skew:	1.122	Prob(JB):	1.39e-06			
Kurtosis:	3.479	Cond. No.	261.			

图 28 100 厘米深度下的 OLS 回归分析结果

5.4.4 200 厘米深度的土壤湿度多元回归分析

最后，对 200 厘米深度土壤湿度与降水量和蒸发量进行多元回归分析。通过 OLS 求解回归分析结果，如下图 29 所示。因此，200cm 土壤与土壤蒸发量和降水量的关系式为，

$$W(t) = 167.3766 + 0.0022 * P(t) - 0.0022 * E(t) \quad (5-5)$$

OLS Regression Results						
Dep. Variable:	200cm湿度(kg/m ²)	R-squared:	0.109			
Model:	OLS	Adj. R-squared:	0.094			
Method:	Least Squares	F-statistic:	7.342			
Date:	Fri, 07 Oct 2022	Prob (F-statistic):	0.000981			
Time:	10:29:37	Log-Likelihood:	-164.04			
No. Observations:	123	AIC:	334.1			
Df Residuals:	120	BIC:	342.5			
Df Model:	2					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
const	167.3766	0.133	1254.132	0.000	167.112	167.641
土壤蒸发量(mm)	0.0022	0.009	0.257	0.797	-0.015	0.01
降水量	-0.0022	0.001	-3.756	0.000	-0.003	-0.001
Omnibus:	15.750	Durbin-Watson:	0.247			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17.518			
Skew:	-0.868	Prob(JB):	0.000157			
Kurtosis:	3.638	Cond. No.	261.			

图 29 200 厘米深度下的 OLS 回归分析结果

5.4.5 土壤湿度预测结果与分析

基于上述对不同深度土壤湿度的多元回归分析，预测 2022-2023 年各月份土壤湿度，预测结果如表 12 所示。根据预测结果，土壤深度为 10 厘米时 5-8 月份的土壤湿度最高，这个时节处于夏季，也是雨季来临的时候，因此土壤湿度比较高。进入 11 月份后土壤湿度降低，因为冬季雪层覆盖，雨水较少。相比更深层的土壤湿度来说，土壤湿度会随着深度增加而升高，受到外界条件影响的波动越小，如图 30。

表 12 2022-2023 年土壤湿度预测结果

年份	月份	10cm 湿度 (kg/m ²)	40cm 湿度 (kg/m ²)	100cm 湿度 (kg/m ²)	200cm 湿度 (kg/m ²)
2022	04	14.42	41.00	53.35	167.39
	05	16.53	43.63	54.54	167.39
	06	16.27	43.46	54.74	167.35
	07	18.14	45.72	55.64	167.37
	08	16.99	44.75	56.06	167.26
	09	15.77	42.98	54.80	167.32
	10	14.45	41.22	53.80	167.35
	11	14.22	43.60	59.73	166.75
	12	13.18	40.89	55.92	167.07
	01	14.13	44.21	61.34	166.59
	02	14.65	45.88	63.96	166.36
	03	13.56	40.77	54.79	167.20
2023	04	14.45	41.07	53.46	167.38
	05	16.55	43.66	54.58	167.38
	06	15.91	42.98	54.49	167.36
	07	18.30	46.36	56.75	167.27
	08	17.15	44.96	56.18	167.26
	09	15.91	43.16	54.89	167.32
	10	14.45	41.15	53.62	167.37
	11	14.36	44.07	60.48	166.69
	12	13.22	40.97	56.03	167.06

根据图 30 可知，土壤深度相同的情况下，40 厘米深度的水分波动最大，200 厘米深度的土壤湿度波动波动最小。同时在冬季土壤湿度整体都比较低，在夏季时土壤湿度都比较高。

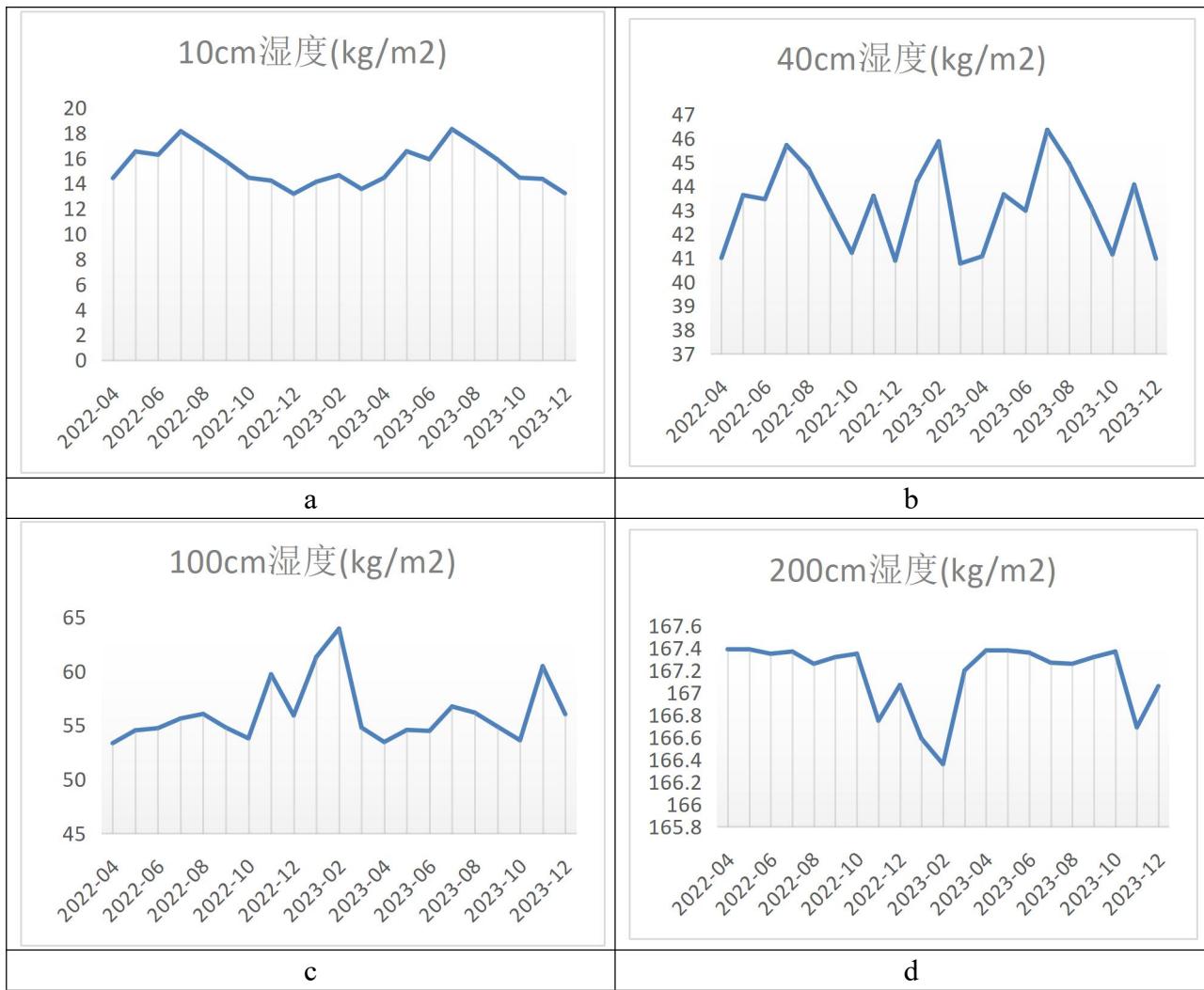


图 30 2022-2023 年土壤湿度预测结果

5.5 总结

本章主要问题是在问题一的基础上，利用附件 3 土壤湿度数据、附件 4 土壤蒸发数据以及附件 8 中降水等数据，在放牧策略不变的情况下，建立多元回归模型、线性回归模型和非线性回归模型对土壤湿度、土壤蒸发量、降水量进行预测。

(1) 通过建立线性、非线性回归模型对 2012-2022 年锡林郭勒地区的土壤蒸发量、降水量等进行建模，利用附件 3-4 进行数据处理，通过 OLS 求解回归模型。

(2) 土壤蒸发量主要集中在 5-9 月份，这与锡林郭勒地区所处位置的日照时间和气温有关。降水量主要集中在 6-9 月份，冬季降水量普遍较少。

(3) 土壤湿度与降水量和蒸发量存在线性关系，通过多元回归分析，能够有效进行土壤湿度预测。

(4) 土壤深度相同的情况下，40 厘米深度的水分波动最大，200 厘米深度的土壤湿度波动波动最小；10 厘米深度波动比较小的原因是接近地面，与地面动植物活动有关，同时更易于受到大气的影响；土壤湿度在夏季雨水充沛的情况下整体湿度比较高，冬季的土壤湿度比较低，且深度越深土壤湿度越大但受外界环境的影响比较少。

六、问题三的求解: OLS 回归与机器学习模型

6.1 问题分析与技术路线图

问题三分为两个子问题，第一个子问题同问题一一样需要从机理分析角度建立不同放牧策略对锡林郭勒草原土壤化学性质影响，为更进一步探究放牧策略对土壤化学性质的影响方向，我们建立放牧策略对土壤化学性质影响模型；第二子问题需要完成对不同放牧策略下 2022 年各类土壤化学性质的预测。针对以上两个问题，确定了该题的总体研究思路如图 31 所示：

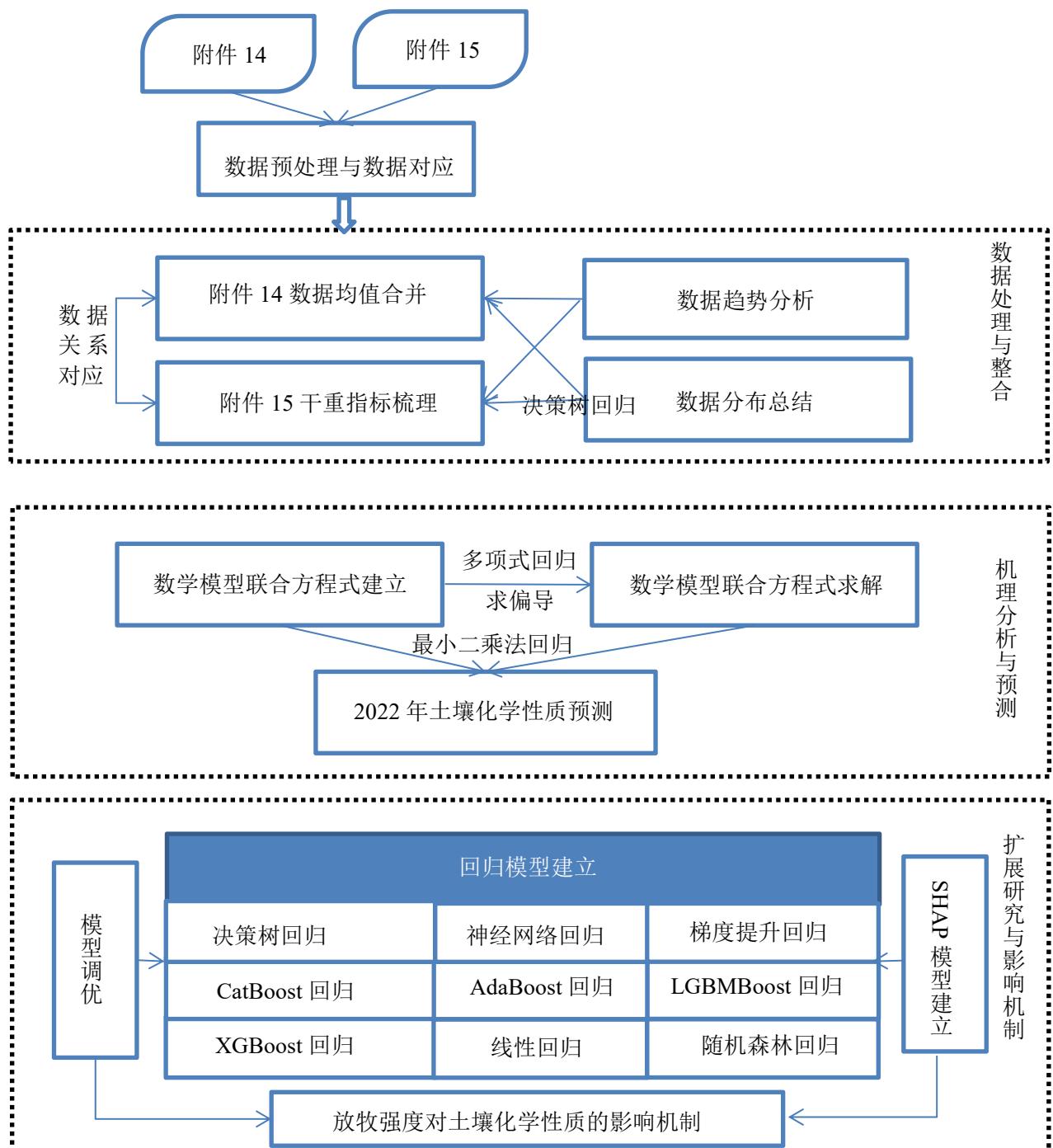


图 31 问题三技术路线图

针对问题三的子问题一，本文拟基于问题一放牧强度对土壤物理性质等的影响，建立关于放牧强度对土壤化学性质的影响，为求出不同方程式的系数，主要使用最小二乘法求系数；其次，针对子问题二，基于普通最小二乘法进行回归拟合预测；此外，本文建立了9种回归模型探索放牧强度对不同土壤化学性质的影响机制。

6.2 数据分析与数据预处理

6.2.1 土壤化学性质处理与数据概况

在附件 14 和附件 15 中，全年采取轮牧的方式进行放牧，因此本文只考虑放牧强度而不考虑放牧方式。不同放牧强度对土壤化学性质的影响主要在附件 14 中，对附件 14 的数据分析可知，2012 年、2014 年不同地区放牧强度设置为 2 个重复，2016 年、2018 年不同地区放牧强度设置为 3 个重复，2020 年为不同样本地区的数据均值，为统一数据年份和放牧强度的对应关系，同 2020 年数据一样，本文对其他年份不同地区的土壤化学性质取均值，最终得到 59 行数据，对这 59 行数据不同年份、不同地区及不同放牧强度下的数据情况绘制小提琴图和抖动散点图如图 32 所示。

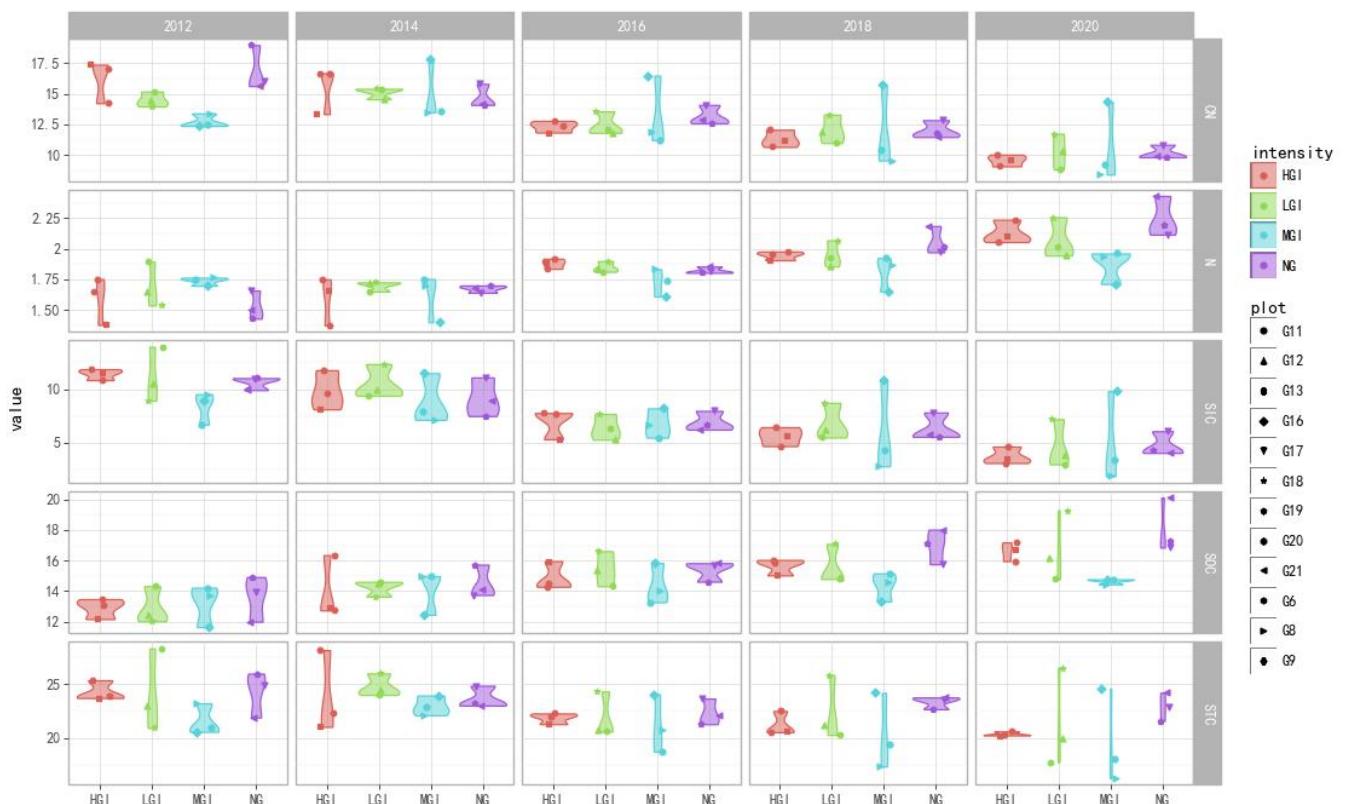


图 32 数据处理后的样地小提琴图和抖动散点图

从图 32 可知，不同放牧强度下各地区土壤 C/N 比、SIC 土壤无机碳逐年呈现下降趋势；全氮 N、SOC 土壤有机碳呈现逐年上升趋势；SIC 土壤无机碳呈现下降趋势但是下降幅度不是特别明显；从各个地区来看，中度放牧强度（即 MGI，4 羊/天/公顷）各地区土壤化学性质差别最大，在大多数情况下，不同放牧强度对不同地区的影响是相似的，尤其是无放牧的情况下，各地区土壤化学性质是接近的。

6.2.2 土壤植物干重数据处理与对应关系建立

由于本问题的子问题一需要从机理分析角度建立不同放牧策略对锡林郭勒草原土壤化学性质影响，考虑到 Woodward 等学者^[8]建立了放牧强度同植物生长的关系，为了将植物生长和土壤化学性质关系进行对应，本文进一步处理附件 15 放牧样地群落结构监测数据集，对附件 15 的时间、轮次及放牧强度进行统计，计算干重的均值，得到如图 33 和图 34 所示的数据概况图，从图中可以看出不同放牧强度下的干重（即植物地上生物量）从全部年份来看无明显差异，但是具体到每一年份，差异较为明显，如 2016 年，重牧（12 天）干重均值较为集中，集中在 15g 左右，而轻牧（3 天）数据相对分散。可见，不同的放牧强度对植物生物量具有较大的影响。

在轻牧、中牧、重牧的情况下，植物生物量的生长可能会受到牛羊等生物产出粪便的微生物的影响，为仅考虑植物单位时间内土壤有机碳和无机碳的变化与植物生长量的直接关系，本文有且仅考虑无牧情况下干重（植物地上生物量）。考虑到附件 14、附件 15 在放牧小区上是对应的，我们选取附件 15 的无牧（0 天）的干重数据和附件 14 的对照组的土壤化学性质进行放牧小区和年份上的对应，并进行后续的机理分析。

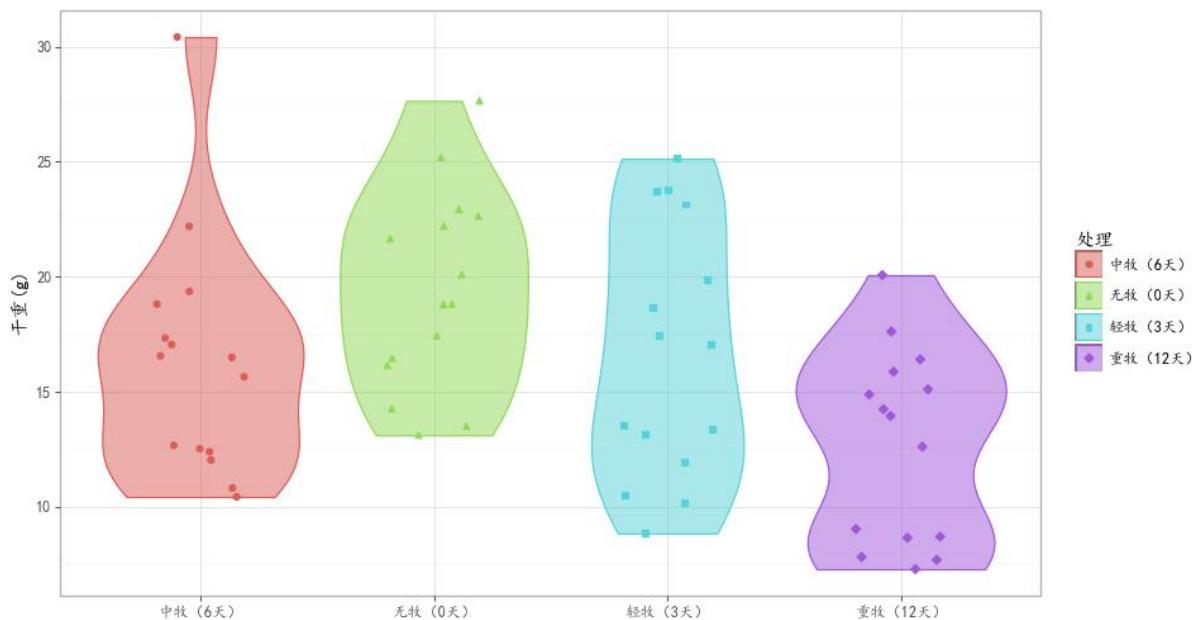


图 33 不同放牧强度下干重数据分布

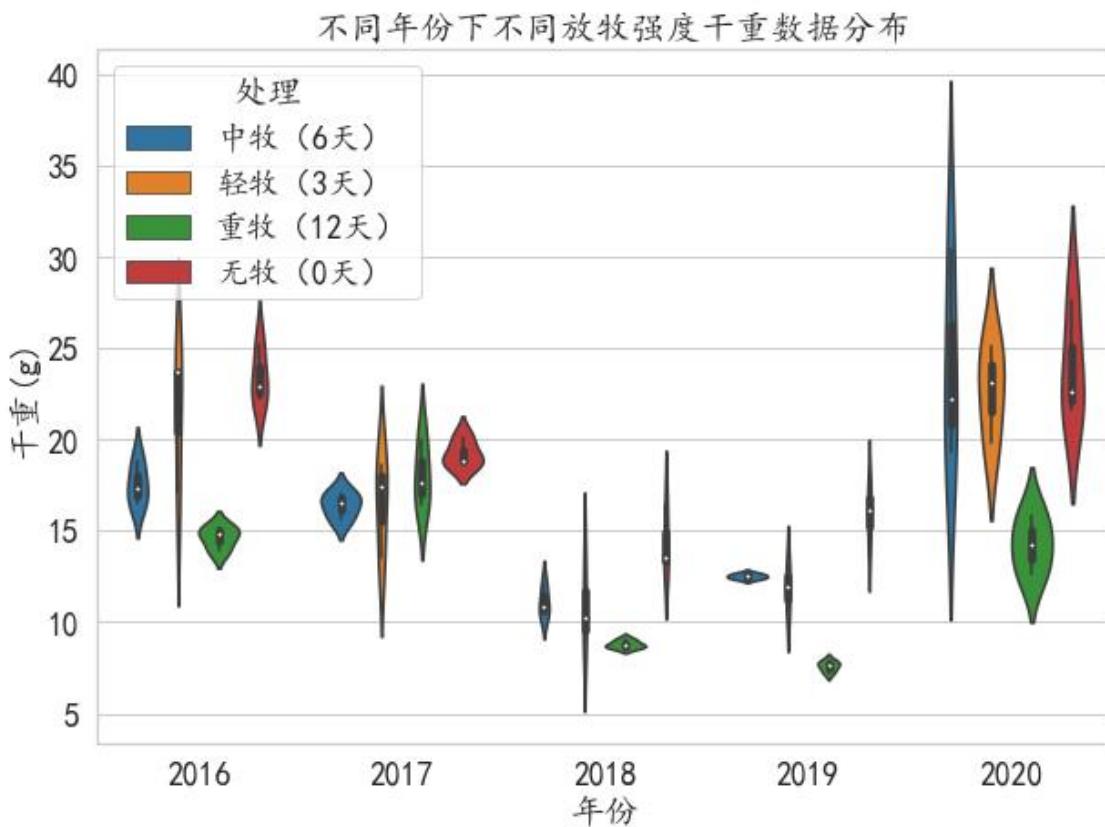


图 34 不同年份下不同放牧强度干重数据分布

6.3 基于多项式回归的放牧强度对土壤化学性质影响的机理模型

如前所述，当数据样本够大的时候，时间序列模型的拟合效果较好，但是当数据样本较小时，简单线性回归和多项式回归比时间序列更为合适。考虑到附件 14 和附件 15 中的样本年份数量较少，仅有 2012 年、2014 年、2016 年、2018 年和 2020 年的数据，且无具体的月份数据，本文延续问题一的机理模型构建方法，考虑基于普通最小二乘法的线性回归拟合进行机理模型构建。

6.3.1 放牧强度对土壤化学性质影响的数学模型构建

在构建放牧强度对土壤化学性质影响的数学模型，需要考虑附件 14 的土壤化学性质的有机合并，土壤的化学性质包括土壤有机碳、土壤无机碳、土壤全碳、全氮和土壤 C/N 比，在所有化学性质中， $STC = SOC + SIC$ ，土壤 C/N 比 $= STC / N$ 。

为了使得建立的数学模型清晰、简洁且可解释，在建立放牧强度对土壤化学性质的影响时，仅考虑土壤有机碳 (SOC)、土壤无机碳 (SIC) 和全氮 (N)。

如前所述，在构建土壤化学性质影响的机理模型时，需要考虑不放牧的情况，即单位时间内土壤有机碳和无机碳的变化与植被生物量（植物生长状况）直接相关，该条件相当于建立在假设一只羊每天所摄入得入的植物量生物量不变的基础上。

因此，根据已有的条件我们可以建立如下数学模型联合方程式：

$$\left. \begin{aligned}
 w(t) &= \alpha * SOC(t) + \beta * SIC(t) + \gamma * N(t) + \theta \\
 \frac{dw(t)}{dt} &= 0.049w(t)\left(1 - \frac{w(t)}{4000}\right) - 0.0047Sw(t) \\
 \frac{dw(t)}{dt} &= \alpha * \frac{dSOC(t)}{dt} + \beta * \frac{dSIC(t)}{dt} + \gamma * \frac{N(t)}{dt} \\
 \alpha * \frac{dSOC(t)}{dt} + \beta * \frac{dSIC(t)}{dt} + \gamma * \frac{N(t)}{dt} &= 0.049w(t)\left(1 - \frac{w(t)}{4000}\right) - 0.0047Sw(t)
 \end{aligned} \right\} \quad (6-1)$$

该数学模型联合方程式各个变量的解释如表 13 所示。

表 13 放牧强度对土壤化学性质影响的数学模型符号解释

序号	符号	说明
1	$w(t)$	随着时间变化的植被生物量含量变化
2	$SOC(t)$	随着时间变化的土壤有机碳含量的变化
3	$SIC(t)$	随着时间变化的土壤无机碳含量的变化
4	$N(t)$	随着时间变化的土壤全氮含量的变化
5	S	单位面积的载畜率
6	α	$SOC(t)$ 的系数
7	β	$SIC(t)$ 的系数
8	γ	$N(t)$ 的系数
9	θ	常数项

该数学模型的联合方程式具体解释如下，首先建立植物生物量同土壤有机碳、无机碳、全氮含量的数学关系并对该函数进行求导，然后根据 Woodward 等学者所建立的放牧同植被生物量的关系可以建立联合方程等式。因此模型重要的是进一步求出 α 、 β 、 γ 的值。模型的系数值将在下一节给出。

6.3.2 基于多项式回归的放牧强度对土壤化学性质影响的数学模型求解

同问题一一样，本文使用基于多项式回归的方法求 α 、 β 、 γ 的数值。基于 6.2 章节对附件 14 和附件 15 的数据分析与预处理的情况下，我们建立了 3 个多项式回归模型。具体如下：

(1) SOC 土壤有机碳的多项式回归模型

根据附件 14 无牧情况下 G17、G19、G21 的 SOC 土壤有机碳随年份的变化，我们拟合出如图 35 所示的回归模型。其中回归截距为 $w0=[85581.26081947]$ 、回归系数为 $w1=[[0.0000000e+00 -8.54552771e+01 2.13353017e-02]]$ ，该模型拟合效果非常好， R^2 高达 0.9943，均方误差为 0.0149，平均绝对误差值为 0.0968，中位绝对值误差为 0.0695。根据模型的拟合结果可以构建 SOC 随时间变化的公式如下所示。 e 为科学计数法，代表 10。

$$SOC(t) = 8.5581e^4 - 0.8546 * t + 0.02134 * t^2 \quad (6-2)$$

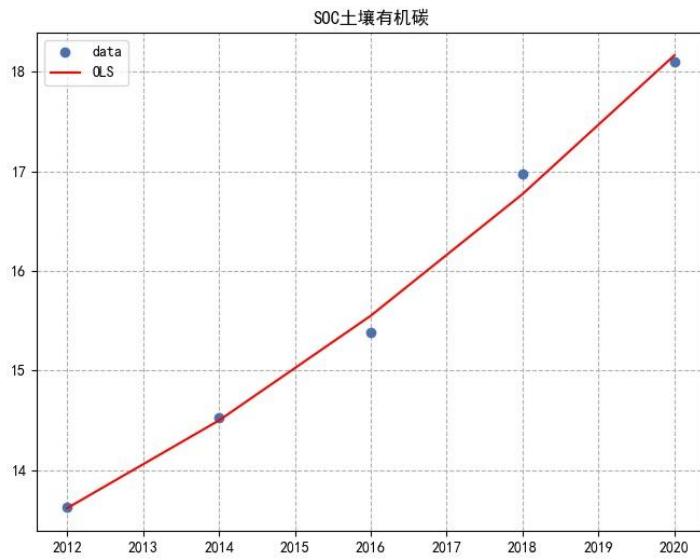


图 35 SOC 土壤有机碳数据拟合图形

(2) SIC 土壤无机碳的多项式回归模型

根据附件 14 无牧情况下 G17、G19、G21 的 SIC 土壤无机碳随年份的变化，我们拟合出如图 36 所示的回归模型。其中回归截距为 $w0=[106013.96919199]$ 、回归系数为 $w1=[[0.00000000e+00 -1.04443697e+02 2.57247354e-02]]$ ，该模型拟合效果也很好， R^2 高达 0.9834，均方误差为 0.0708，平均绝对值误差为 0.2257，中位绝对值误差为 0.2334。根据模型的拟合结果可以构建 SIC 随时间变化的公式如下所示。 e 为科学计数法，代表 10。

$$SIC(t) = 10.6014e^4 - 0.0104 * t + 0.0257 * t^2 \quad (6-3)$$

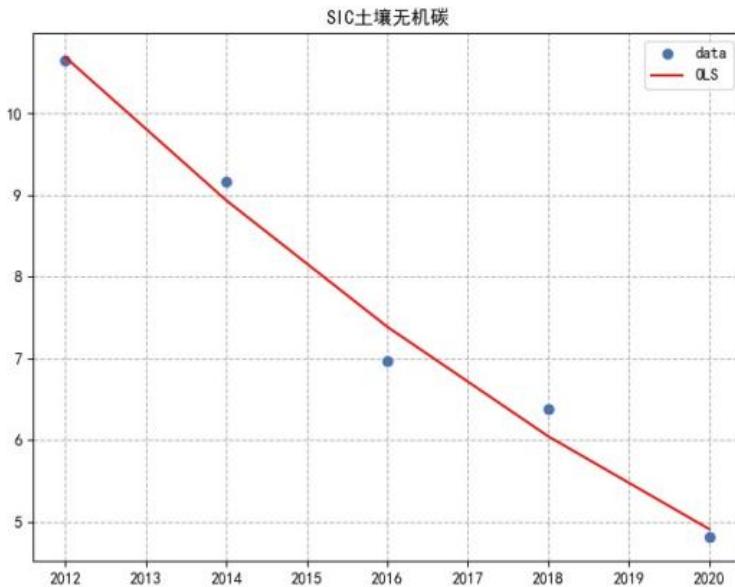


图 36 SIC 土壤无机碳数据拟合图形

(3) 全氮 N 的多项式回归模型

根据附件 14 无牧情况下 G17、G19、G21 的全氮 N 随年份的变化，我们拟合出如图

37 所示的回归模型。其中回归截距 $w_0=[12420.53700646]$ 、回归系数 $w_1=[[0.00000000e+00 -1.24112309e+01 3.10077577e-03]]$ ，该模型拟合效果也很好， R^2 高达 0.9974，均方误差为 0.0002，平均绝对值误差为 0.0103，中位绝对值误差为 0.0078。根据模型的拟合结果可以构建全氮 N 随时间变化的公式如下所示。 e 为科学计数法，代表 10。

$$N(t) = 1.2421e^4 - 0.1241 * t + 0.0031 * t^2 \quad (6-4)$$

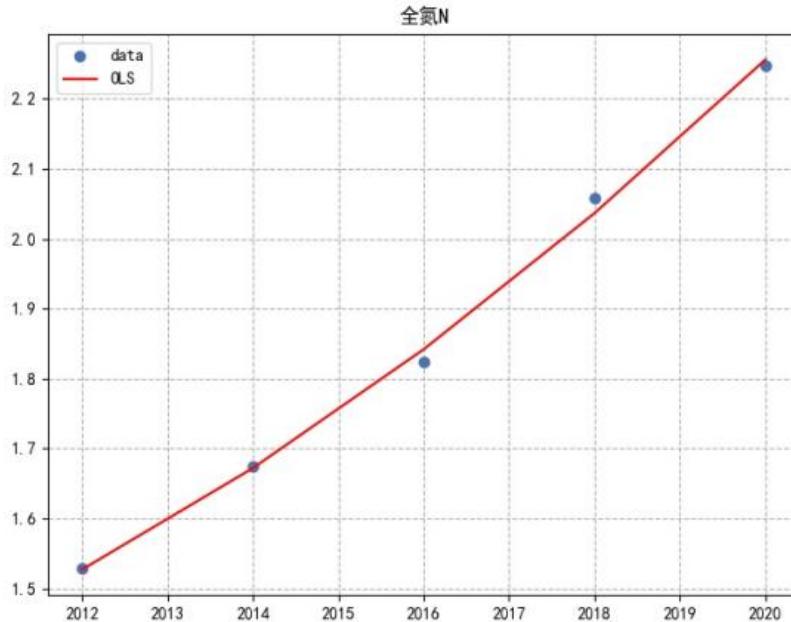


图 37 全氮 N 数据拟合图形

(4) 数学模型求解

根据 $SOC(t)$ 、 $SIC(t)$ 、 $N(t)$ 的数学方程及对应的普通最小二乘法 OLS 对 α 、 β 、 γ 及 θ 进行求解。求解结果如图 38 所示。从图中可得 $\alpha = -1.8051$ 、 $\beta = -0.3676$ 、 $\gamma = 8.5545$ 、 $\theta = 35.6882$ 。

OLS Regression Results						
Dep. Variable:		植物量	R-squared:		0.048	
Model:		OLS	Adj. R-squared:		-0.523	
Method:		Least Squares	F-statistic:		0.08414	
Date:		Sat, 08 Oct 2022	Prob (F-statistic):		0.966	
Time:		10:25:52	Log-Likelihood:		-26.666	
No. Observations:	9	AIC:		61.33		
Df Residuals:	5	BIC:		62.12		
Df Model:	3					
Covariance Type:		nonrobust				
	coef	std err	t	P> t	[0.025	0.975]
const	35.6882	52.045	0.686	0.523	-98.098	169.475
SOC土壤有机碳	-1.8051	4.362	-0.414	0.696	-13.019	9.408
SIC土壤无机碳	-0.3676	2.768	-0.133	0.900	-7.484	6.749
全氮N	8.5545	36.009	0.238	0.822	-84.009	101.118
Omnibus:		1.567	Durbin-Watson:		1.457	
Prob(Omnibus):		0.457	Jarque-Bera (JB):		0.758	
Skew:		-0.204	Prob(JB):		0.685	
Kurtosis:		1.638	Cond. No.		465.	

图 38 α 、 β 、 γ 系数求解结果

因此，不同放牧强度 S 对土壤化学性质影响的数据模型为：

$$\begin{aligned}
& -1.8051 * \frac{dSOC(t)}{dt} - 0.3676 * \frac{dSIC(t)}{dt} + 8.5545 * \frac{N(t)}{dt} \\
& = 0.049w(t)(1 - \frac{w(t)}{4000}) - 0.0047Sw(t)
\end{aligned} \tag{6-5}$$

6.4 基于普通最小二乘法(OLS)回归的土壤化学性质预测

6.4.1 不同放牧强度土壤化学性质变化的线性总览

如前所述，考虑到每年不同放牧强度下数据量较少，在方法预测上，主要选择最小二乘法进行预测。为此，本文首先对数据进行线性拟合，观察数据是否符合线性回归的要求，绘制图形如图 39 所示。从图中可以看出，不同的土壤化学性质与时间呈现不同的相关性，其中土壤有机碳 SOC、全氮 N 在不同放牧强度下与时间基本呈现线性关系，且随时间逐渐增长，土壤无机碳 SIC、土壤全碳 STC、土壤 C/N 比的线性关系在不同的放牧强度下皆呈现下降趋势，但线性关系强度在不同强度下有所区别，如土壤全碳 STC 的在 MGI 中度放牧强度下一元线性拟合不够好，土壤全碳 STC 在 LGI 轻度放牧强度下一元线性拟合欠佳，土壤 C/N 比在 MGI 中度放牧强度下的一元线性拟合有待加强等，对于这些数据的拟合，可考虑使用多项式回归。

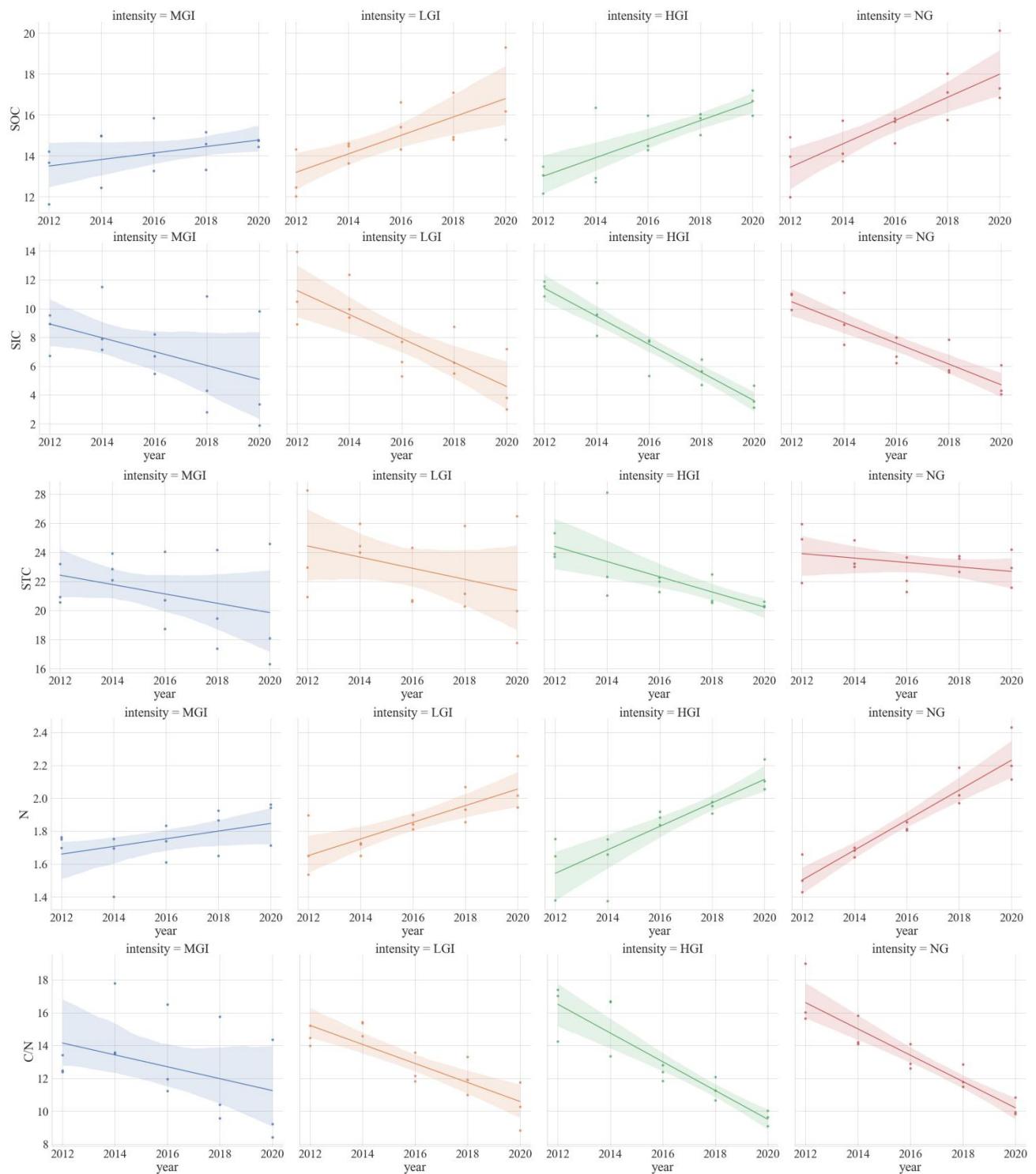


图 39 不同放牧强度下土壤化学性质的线性变化

6.4.2 单变量回归诊断方法

为了能够确定本章节所使用的数据能够很好地阐释 OLS 模型的系数，本文对数据进行单变量回归检验，绘制了如下四个图形来观察数据拟合效果。

(1) 组成部分残差 (CCPR) 图检验

CCPR 图提供一种新的方法来判断数据是否能够用于回归，即自变量与因变量是否呈

现线性回归，它考虑其他自变量的影响来判断一个回归变量对响应变量的影响。部分残差图可以定义为 *Residuals* (残差) + $B_i X_i$ 与 X_i 的关系，即组成成分添加 $B_i X_i$ 和 X_i 来展现拟合情况。如果变量与其他自变量高度相关，则需要高度小心，因为这将会低估真实的方差情况。

(2) 拟合图

拟合图使用置信区间来判别拟合效果，在拟合图中可以很好地看到真实数据和拟合数据在数据值上的接近性，如果真实数据与拟合数据高度重合，说明回归拟合效果较好。

(3) 偏回归图

在进行回归的时候，简单查看单个双变量图来判别变量之间的关系是不合理的，自变量与因变量的回归可以使用其他变量作为条件，偏回归图就提供了这样一种条件，自变量 X_i 的偏回归图是因变量 Y 与其他自变量回归残差同自变量 X_i 与其他自变量回归残差的散点图，简单而言，图形展现了两种残差，一种是变量 X 的残差，另一种是因变量 Y 的残差，前者指自变量 X 对模型中其他所有预测变量上进行回归并计算的残差，后者指因变量 Y 对自变量 X 以外的所有预测变量上进行回归获取的残差。偏回归图展现了这两者之间的关系，需要注意的是，在偏回归图中，拟合线通常有斜率 β_k ，并且截距为 0。

(4) 模型残差图

在回归分析的时候还需要检验模型的残差图，模型残差图是以指定变量为横坐标、残差为纵坐标的图形，回归分析后需要使用残差图检验模型，因为使用最小二乘法进行回归分析涉及回归模型的随机性和不可预测性。一个有效的回归模型通常可以用以下公式进行表示。

$$\text{Response} = (\text{Constant} + \text{Predictors}) + \text{Error} \quad (6-6)$$

在该公式中，涉及了确定性部分和随机性部分。其中确定性部分是由预测自变量函数组成，即 Constant + Predictors，而随机性部分是由 Error 组成，在回归模型中，预测值和预测值之间的差异应当是随机的且不可预测的，如果 Error 可预测，则说明预测模型缺少一些可预测的因素。残差图可以帮助检验观测到或者预测道德误差 Error 与随机误差是否一致，进而帮助解释模型的科学性。

6.4.3 基于 OLS 的土壤化学性质预测

根据 6.2 章节数据分析与数据预处理结果，以及 6.4.1 中数据线性拟合情况，本文对不同放牧强度下不同小区及不同土壤性质建立普通最小二乘法回归（OLS）模型进行预测，累积建立 60 个回归模型，为了检验回归模型的合理性，使用 6.4.2 章节中单变量回归诊断方法建立了如图 40 和图 41 所示的 G17 放牧区 SOC、SIC、STC 回归模型诊断和 G17 放牧区全氮 N 和土壤 C/N 比回归模型诊断，由于回归模型过多，本文仅展示 G17 放牧区的回归模型诊断结果，从图 40 和图 41 及表 14 回归结果可知，各个回归模型建立优良合理。根据各回归拟合预测模型可得到锡林郭勒草原监测样地(12 个放牧小区)在不同放牧强度下 2022 年土壤同期有机碳、无机碳、全 N、土壤 C/N 比等值如表 15 所示。

表 14 G17 放牧区土壤化学性质回归结果

预测变量	R-squared	Prob (F-statisti c)	Log- Likelih ood	AIC	BIC	Interce pt- Coef	Interce pt- Std err	Intercep t- t
------	-----------	---------------------------	------------------------	-----	-----	------------------------	---------------------------	---------------------

SOC 土壤有机碳	0.877	0.019	-2.6649	9.33	8.55	-769.59	169.67	-4.536
SIC 土壤无机碳	0.894	0.0151	-4.7951	13.5 9	12.8 1	1318.50	259.81	5.075
STC 土壤全碳	0.916	0.0106	0.41234	3.17 5	2.39 4	548.91	91.692	5.986
全 N	0.930	0.00803	8.0662	-12. 13	-12.9 1	-123.41	19.83	-6.220
土壤 C/N 比	0.950	0.00478	-2.8778	9.75	8.97	1354.12	177.06	7.648

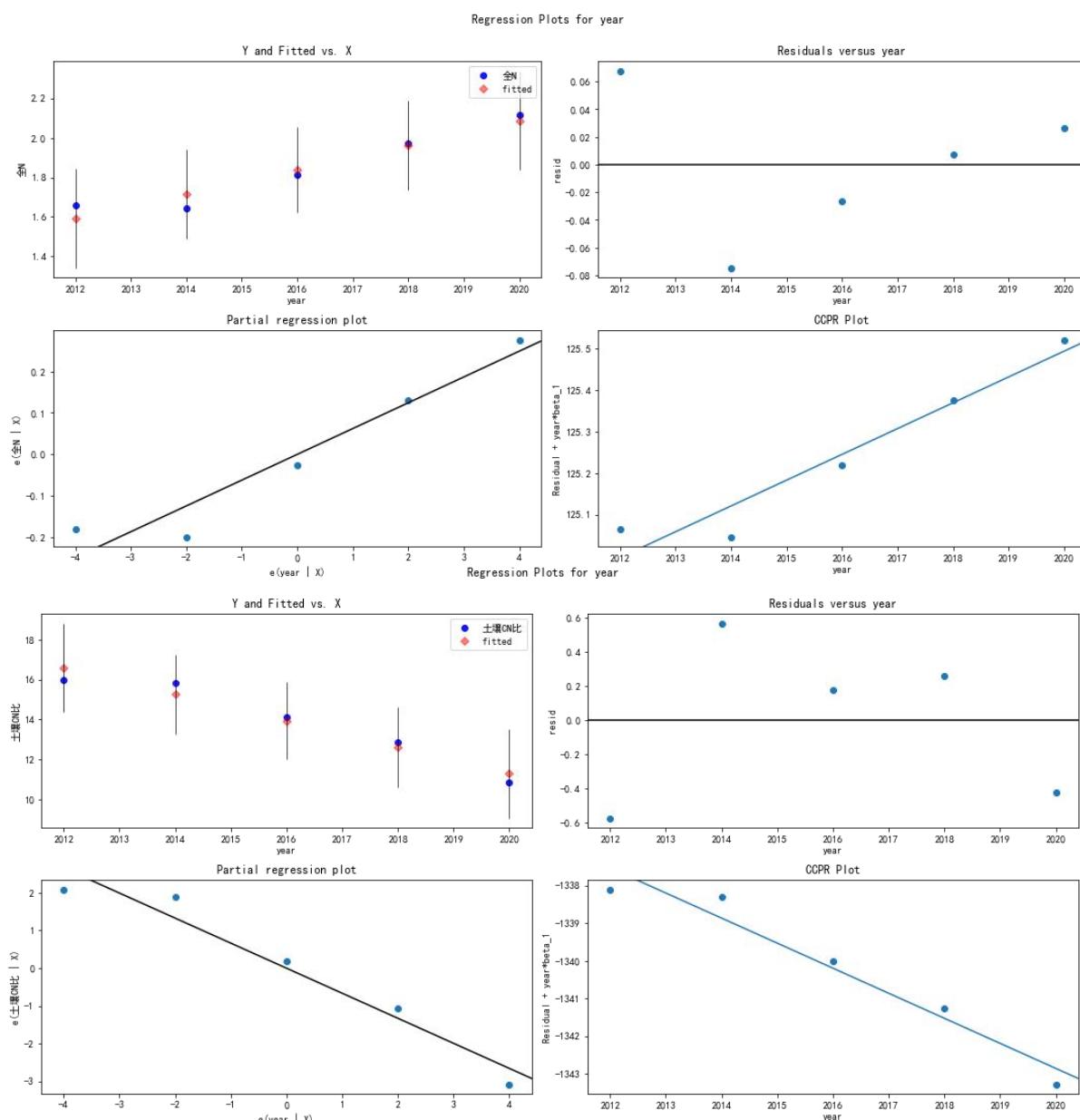


图 40 G17 放牧区全氮 N 和土壤 C/N 比回归模型诊断

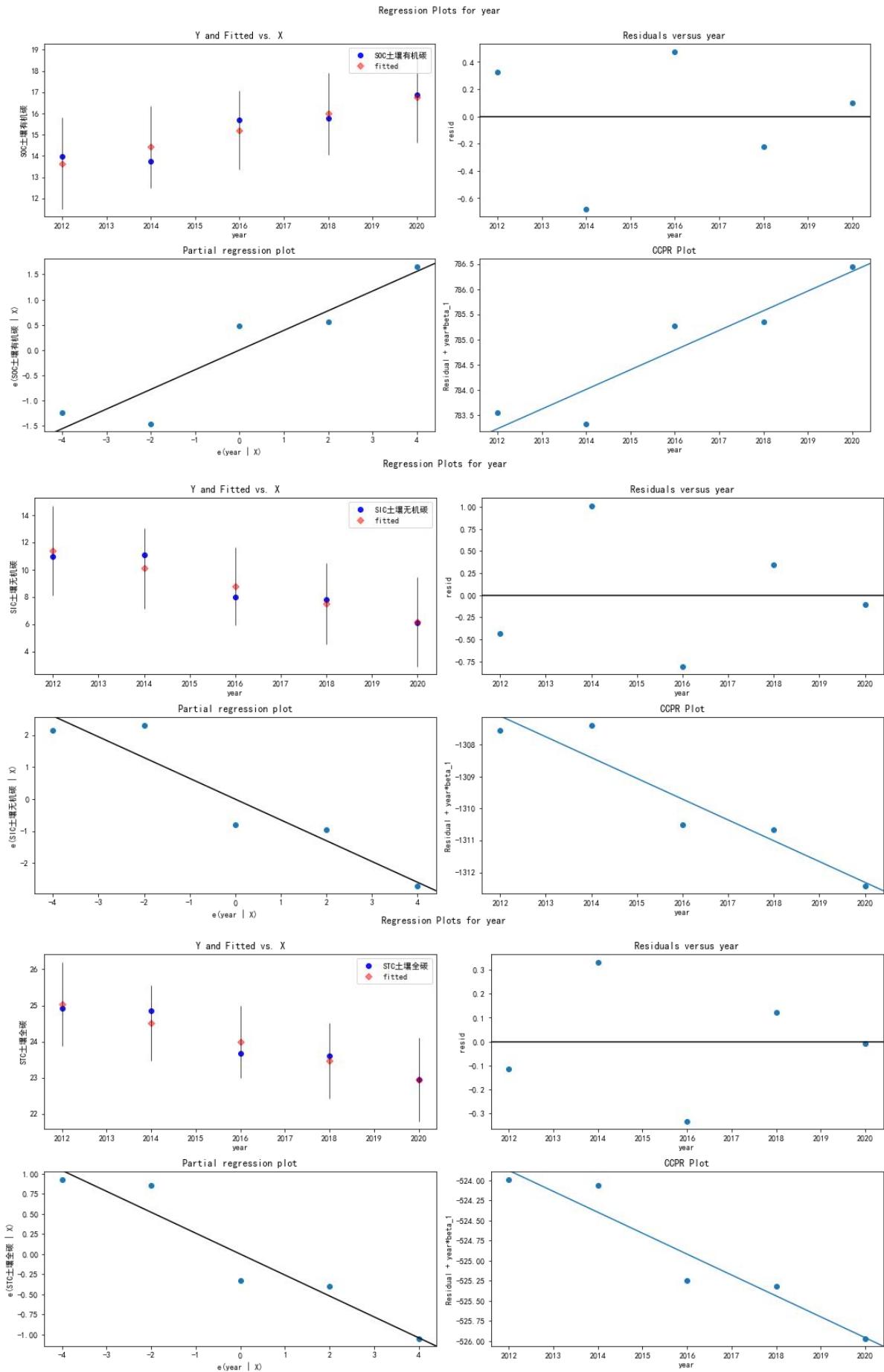


图 41 G17 放牧区 SOC、SIC、STC 回归模型诊断

表 15 锡林郭勒草原监测样地在不同放牧强度下 2022 年土壤化学性质预测值

放牧强度	Plot 放牧小区	SOC 土壤有机碳	SIC 土壤无机碳	STC 土壤全碳	全 N	土壤 C/N 比
NG	G17	16.923	5.145	22.821	2.059	8.246
	G19	17.395	3.254	21.437	2.38	8.782
	G21	22.041	3.212	24.45	2.663	8.058
LGI	G6	14.569	1.39	15.274	2.139	6.32
	G12	15.943	1.421	20.512	1.955	10.306
	G18	21.644	7.982	27.178	2.497	12.427
MGI	G8	14.571	1.014	14.955	1.982	6.646
	G11	14.023	2.57	16.284	2.056	8.388
	G16	13.847	10.205	24.843	1.78	13.291
HGI	G9	18.572	1.966	20.169	2.465	7.267
	G13	16.22	2.844	19.728	2.302	8.975
	G20	15.145	3.129	20.96	2.118	9.482

6.5 扩展研究：基于机器学习的放牧强度对土壤化学性质的影响机制研究

6.5.1 回归方法调研与阐述

为了发现不同放牧强度对土壤化学性质的具体影响方向，我们使用基于机器学习的方法进行探究。在现有的回归方法中，主要分为四个方向：（1）回归算法，主要包括线性回归和逻辑回归等方法，这类方法使用简单适用于简单数据集的拟合；（2）正则化方法，主要包括 LASSO 回归、岭回归等，与回归算法相比，复杂度更高并且能够提升模型的泛化能力；（3）决策树学习回归方法，主要采取 CART 等核心算法建立分类及回归树，相较于前两种方法复杂度更高，模型也更容易解释；（4）集成算法，集成算法是将多个算法进行集合并预测，精准率相对而言更高。

本文选择了以下 9 种流行回归模型以用于后续回归模型的比较选优。包括决策树回归、线性回归、随机森林回归、AdaBoost 决策树回归、GradientBoost 回归、XGBoost 回归、CatBoost 回归、神经网络线性回归、LGBMBoost 回归。在这些回归方法种，绝大多数为集成回归和决策树回归，并且包括了简单的线性回归。现就在后续实验中表现较好的回归模型进行原理的进一步阐释。由于线性回归已在前面进行过阐释，这里不再赘述。

（1）决策树回归

决策树回归模型应用广泛，在健康素养领域^[9]、烟考化学指标识别^[10]、遥感数据领域^[11]等都有所应用，在决策回归树模型中，CART 算法是其核心内容，其决策树回归的概念化模型如图 42 所示，即对于某个节点而言需要对每个结点进行“是”与“否”的判断，并根据节点是否的判断可以将其对应到右边的二维平面中。

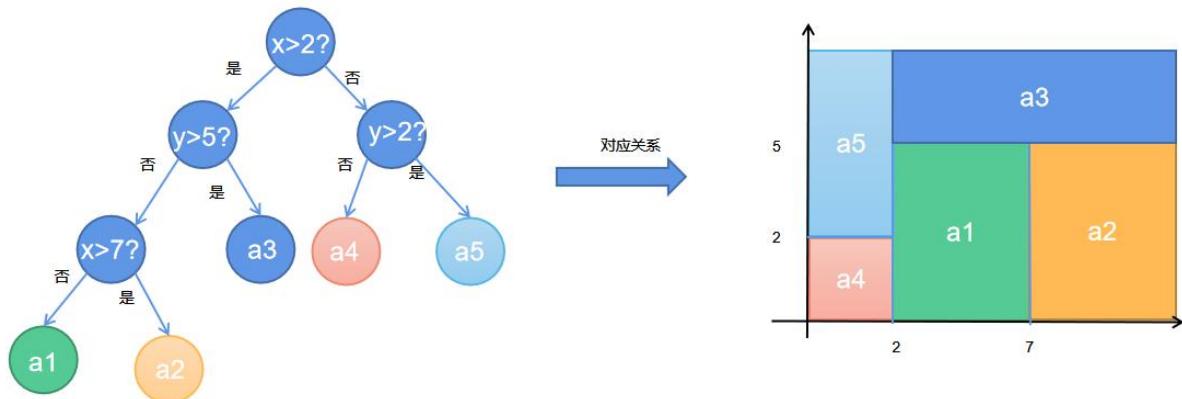


图 42 二维平面决策树回归概念化模型

回归树建立的重点是确定切分点和输出值，其中切分点的选择可以使用最小二乘法，输出值的确定则可以使用均值。例如 X 为输入变量，Y 为输出变量（连续变量），那么对于给定的数据集合 $Dataset = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$ ，可以采用启发式搜索，逐一划分确定该数据集合中特征取值，然后基于平方误差最小化准则划分最优切分点，即求解以下公式。其中 $R1(j, s)$ 和 $R2(j, s)$ 两个区域，该函数将主要求出使得这两个区域平方误差和最小的 j 和 s。

$$\min_{j,s} [\min_c \sum_{x_i \in R1(j,s)} (y_i - c_1)^2 + \sum_{x_i \in R2(j,s)} (y_i - c_2)^2] \quad (6-7)$$

(2) Boost 算法综述

Boost 算法包括 AdaBoost、GradientBoost、CatBoost、XGBoost 等，这些模型都是基于决策树进行进一步的改进。其中，XGBoost 以梯度提升(Gradient Boost)为框架的机器学习集成算法，其特点在于采用了稀疏感知算法、近似树学习、核外计算优化方法等^[12]；Catboost 算法最强大的特征是可以混合使用独热编码和平均编码处理类别特征并且提出了一种全新的梯度提升机制，丰富了特征维度^[13]。Adaboost^[14]算法通过使用错误率不断更新分布权值向量构建多分类器，然后基于加权方法得到决策结果，其结构图可如图 43 所示。

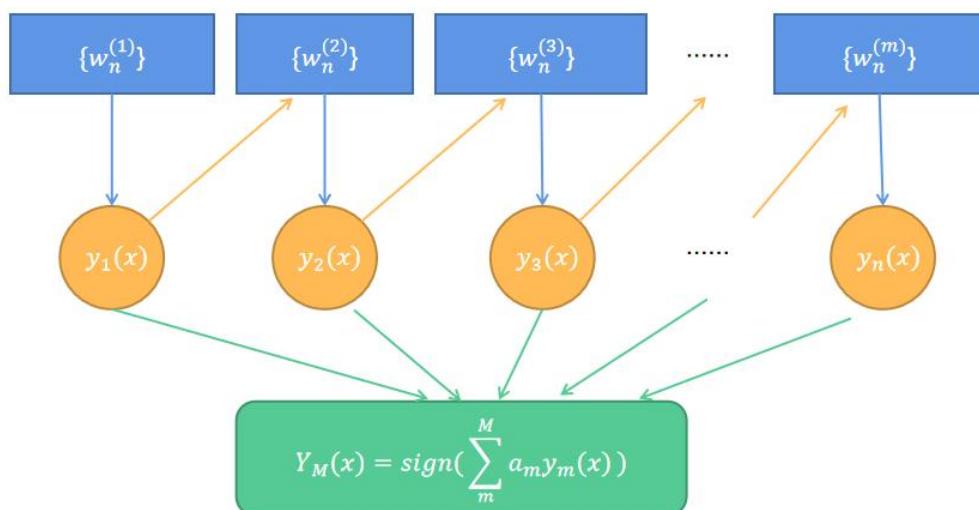


图 43 Adaboost 算法概念结构化模型

(3) 随机森林回归

随机森林的回归构造过程如图 44 所示^[15]。在随机森林中需要有放回抽取样本，并依据如信息增益、信息增益比、基尼指数等选择最佳分裂变量，直至所有节点都已经分裂，则停止分裂并建立深林。随机森林可以很好地解决过拟合问题，虽然算法和计算成本更高，但是结果更加稳定。

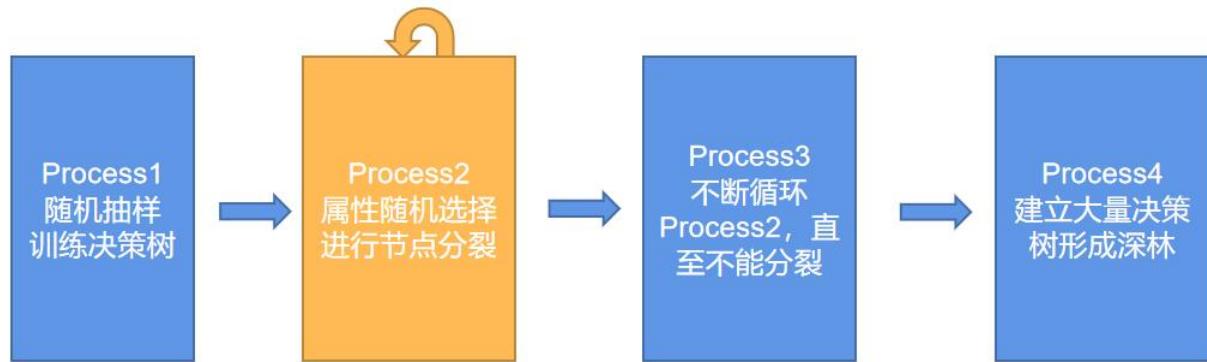


图 44 随机森林的构造过程

6.5.2 模型预测的最优回归方法选择

通常而言，对于回归模型好坏的评价可以使用平均绝对误差 MAE 和均方误差 MSE(Mean Squared Error)。通常值越小，说明模型拟合效果越好。

基于以上 9 种模型，我们将放牧强度作为自变量，并使用 one-hot 编码对变量进行虚拟变量分类，因变量分别为 SOC 土壤有机碳、SIC 土壤无机碳、STC 土壤全碳、全 N、土壤 C/N 比。使用 MAE 和 MSE 进行模型评估，各模型的 MAE 和 MSE 的效果如图 45 所示。从图中可以看出这些模型神经网络 MLP 算法对于土壤化学性质的预测效果最差，尤其是对 SOC 土壤有机碳、STC 土壤全碳和土壤 C/N 比。其他各类模型在 MAE 和 MSE 的差别上并不大，但综合比较而言，对于预测各个因变量的回归模型我们可以选择以下最优模型：

- (1) 对于使用放牧强度预测 SOC 土壤有机碳的最优回归模型选择决策树模型；
- (2) 对于使用放牧强度预测 SIC 土壤无机碳的最优回归模型选择决策树模型；
- (3) 对于使用放牧强度预测 STC 土壤全碳的最优回归模型选择决策树模型；
- (4) 对于使用放牧强度预测全 N 的最优回归模型选择随机森林回归模型；
- (5) 对于使用放牧强度预测土壤 C/N 比的最优回归模型选择随机森林回归模型。

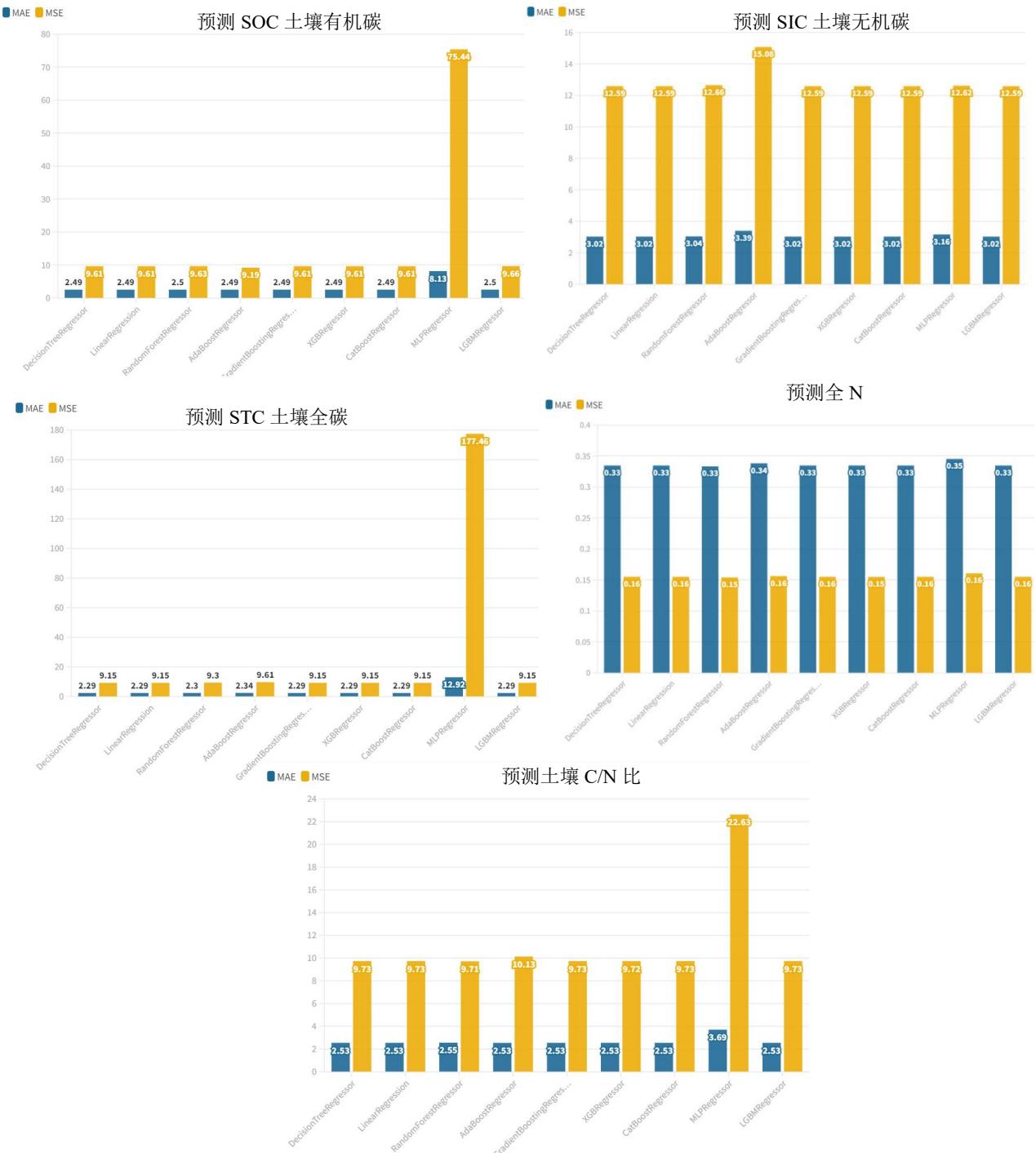


图 45 放牧强度对土壤化学性质影响的回归模型 MAE 和 MSE 图

6.5.3 模型调优与放牧强度对土壤化学性质的影响机制分析

考虑到尽管选择了最优模型，但模型可能还不是最佳模型，因此，本文进一步对最优模型进行参数调整，以获取最佳模型。然后使用 SHAP 方法对模型进行解释，SHAP 方法是可解释机器学习模型^[16]，能够对“黑盒模型”的输入进行有效的解释。

(1) 模型调优与放牧强度对 SOC 土壤有机碳的影响

如图图 46 所示为放牧强度对 SOC 土壤有机碳影响的模型调优和特征影响图。从图中

可以看出，使用决策树模型的最优回归模型的 `min_samples_split` 和 `min_samples_leaf` 值分别为 8 和 3，能够使得 `neg_mean_absolute_error` 最小。使用该最佳模型建立 SHAP 模型，探索放牧强度对 SOC 土壤有机碳。从图中可以看出，在无牧（NG，0 羊/天/公顷）和轻牧（LGI，2 羊/天/公顷）的情况下，可以有效促进 SOC 土壤有机碳的含量，但是在中牧（MGI，4 羊/天/公顷）的情况下，将会降低 SOC 土壤有机碳的含量，而在重牧（HGI，8 羊/天/公顷）的情况下，该影响不明显。

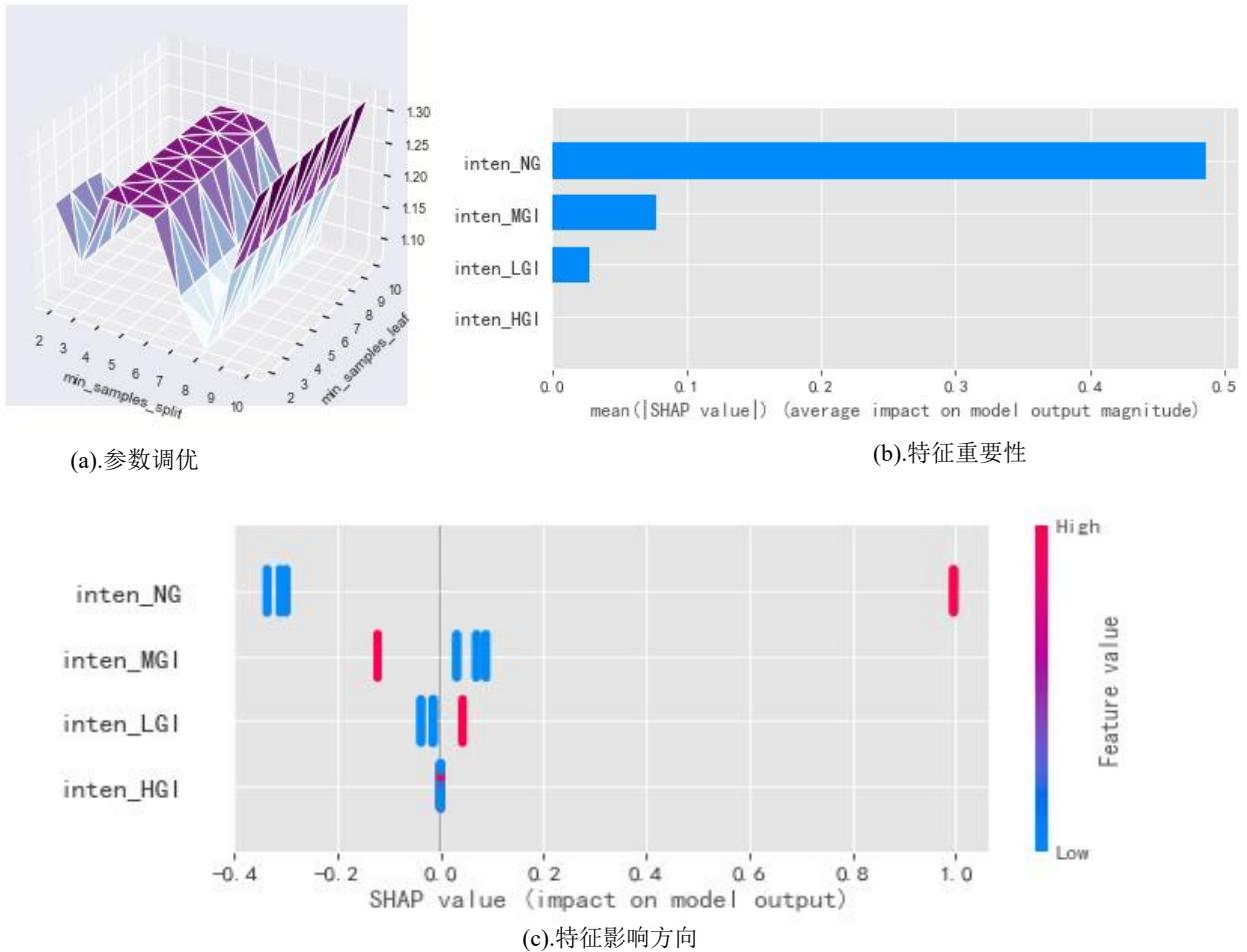


图 46 SOC 情况下的参数调优图与特征影响图

（2）模型调优与放牧强度对 SIC 土壤无机碳的影响

如图图 47 所示为放牧强度对 SIC 土壤无机碳影响的模型调优和特征影响图。从图中可以看出，决策树模型的最佳回归模型的 `min_samples_split` 和 `min_samples_leaf` 值分别为 2 和 4，这时，能够使得 `neg_mean_absolute_error` 最小。使用该最佳模型建立 SHAP 模型，探索放牧强度对 SIC 土壤无机碳的影响。从图中可以看出，在无牧（NG，0 羊/天/公顷）和中牧（MGI，4 羊/天/公顷）的情况下，都会降低 SIC 土壤无机碳的含量，而在重牧（HGI，8 羊/天/公顷）的情况下，反而能够增加 SIC 土壤无机碳的含量，轻牧（LGI，2 羊/天/公顷）对 SIC 土壤无机碳的含量影响并不明显，总体而言，放牧强度对 SIC 土壤有机碳的含量的影响呈现先下降后增加的趋势。

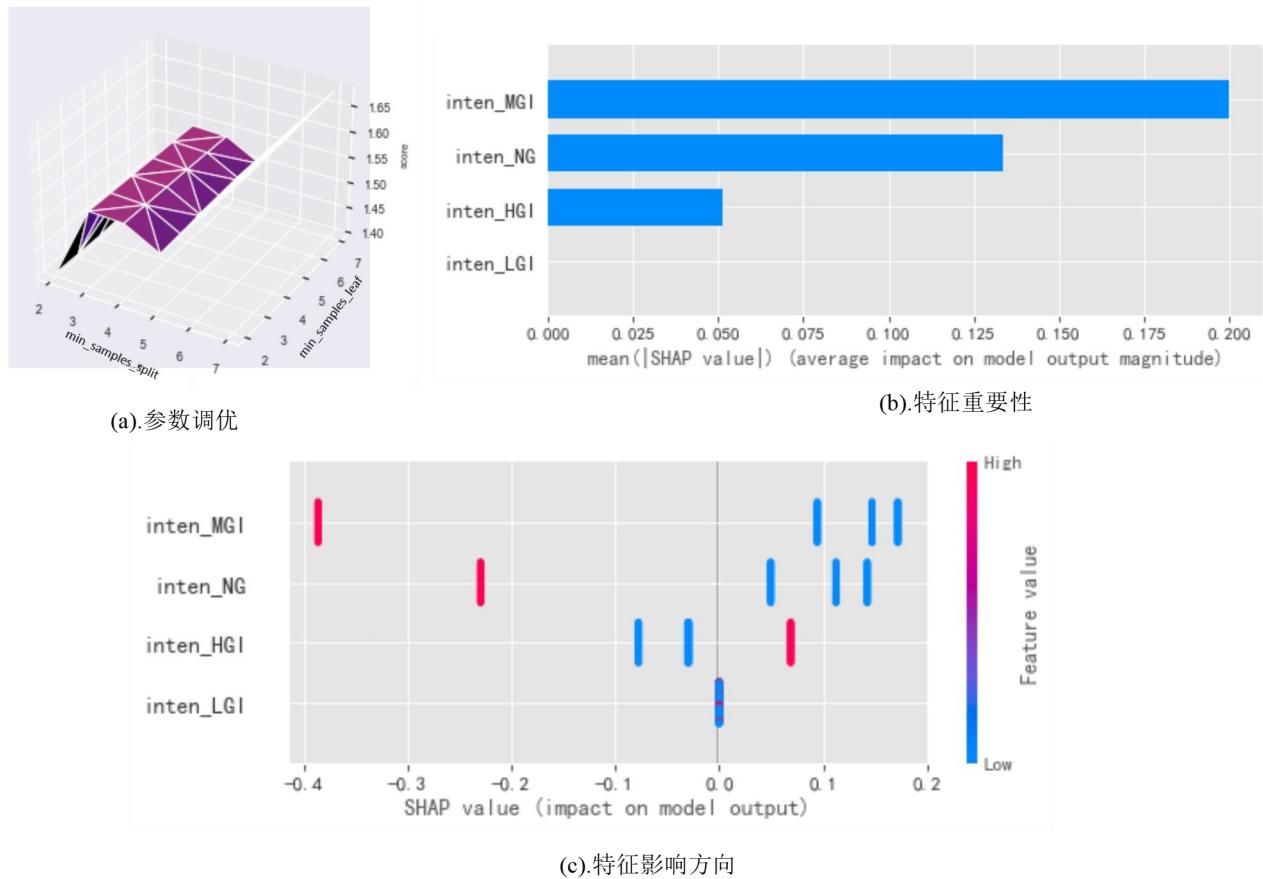


图 47 SIC 情况下的参数调优图与特征影响图

(3) 模型调优与放牧强度对 STC 土壤全碳的影响

如图图 48 所示为放牧强度对 STC 土壤全碳影响的模型调优和特征影响图。从图中可以看出，决策树模型的最佳回归模型的 `min_samples_split` 和 `min_samples_leaf` 值分别为 4 和 5，这时，能够使得 `neg_mean_absolute_error` 最小。使用该最佳模型建立 SHAP 模型，探索放牧强度对 STC 土壤全碳的影响。从图中可以看出，在轻牧（LGI，2 羊/天/公顷）和重牧（HGI，8 羊/天/公顷）的情况下，可以促进 STC 土壤全碳的含量，在中牧（MGI，4 羊/天/公顷）的情况下，会降低 STC 土壤全碳的含量，轻牧（LGI，2 羊/天/公顷）对 STC 土壤全碳的含量影响不明显。

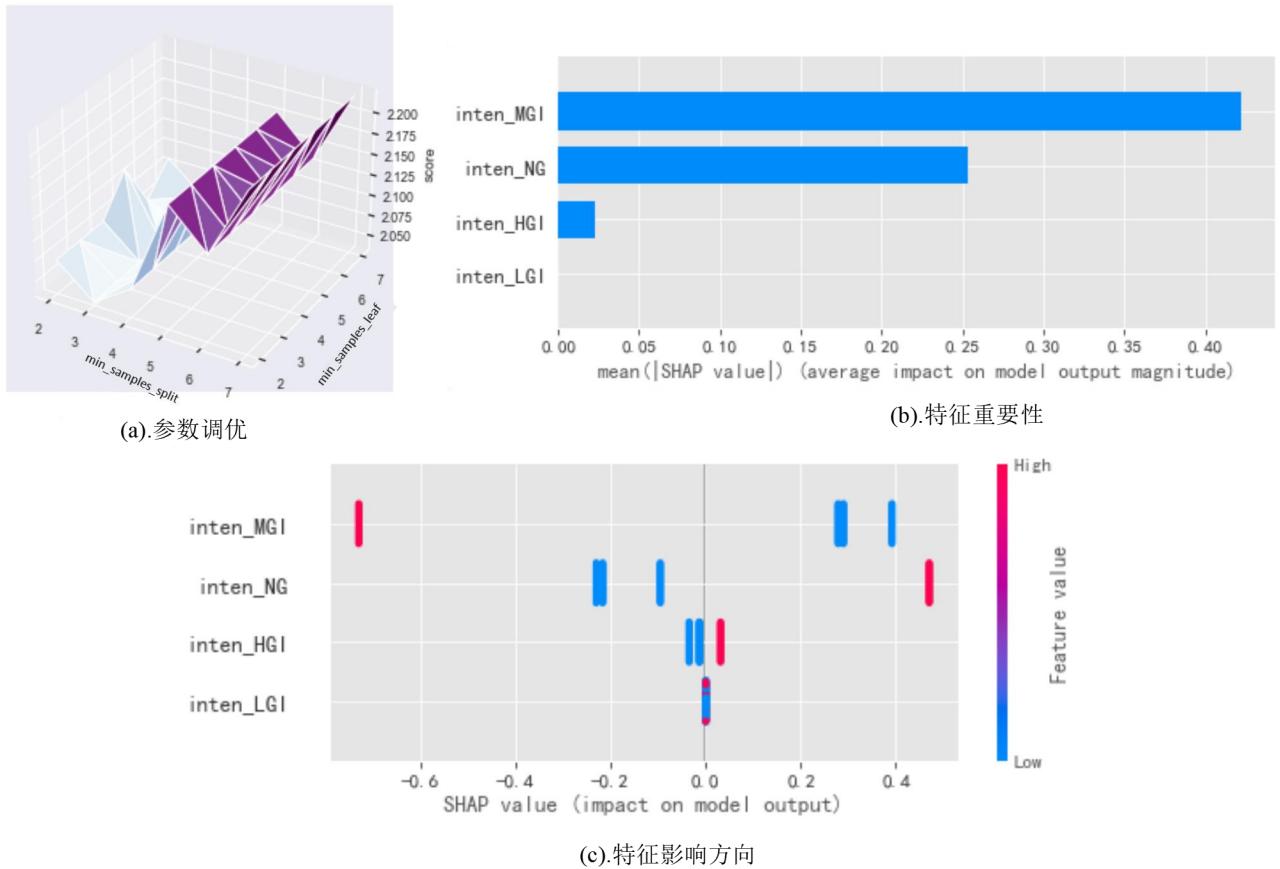


图 48 STC 情况下的参数调优图与特征影响图

(4) 模型调优与放牧强度对全 N 的影响

如图图 49 所示为放牧强度对全 N 影响的模型调优和特征影响图。从图中可以看出，随机森林模型(RandomForest)的最佳回归模型的 n_estimators 和 max_depth 值分别为 10 和 4，这时，误差最小。使用该最佳模型建立 SHAP 模型，探索放牧强度对全 N 的影响。从图中可以看出，在无牧(NG, 0 羊/天/公顷)的情况下，会增加全 N 的含量，而在轻牧(LGI, 2 羊/天/公顷)、中牧(MGI, 4 羊/天/公顷)、重牧(HGI, 8 羊/天/公顷)的情况下均会降低全 N 的含量且放牧强度越强，对于全 N 的含量降低的影响越大。

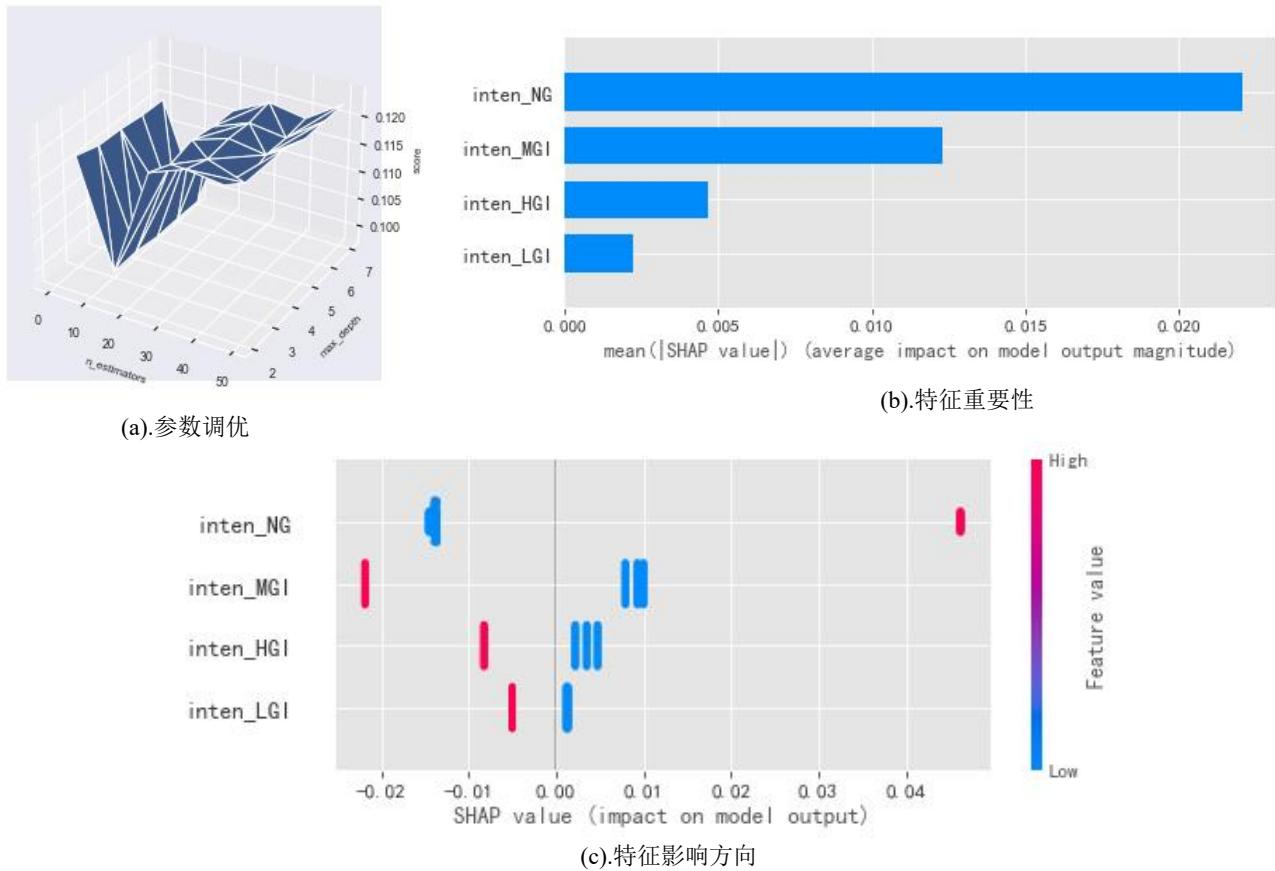


图 49 全 N 情况下的参数调优图与特征影响图

(5) 模型调优与放牧强度对土壤 C/N 比的影响

如图图 50 所示为放牧强度对土壤 C/N 比影响的模型调优和特征影响图。从图中可以看出，随机森林模型(RandomForest)的最佳回归模型的 n_estimators 和 max_depth 值分别为 30 和 5，这时，误差最小。使用该最佳模型建立 SHAP 模型，探索放牧强度对土壤 C/N 比的影响。从图中可以看出，无牧 (NG, 0 羊/天/公顷)、重牧 (HGI, 8 羊/天/公顷) 都能促进土壤 C/N 比值的增加，而轻牧 (LGI, 2 羊/天/公顷) 和中牧 (MGI, 4 羊/天/公顷) 都会降低土壤 C/N 比的值。

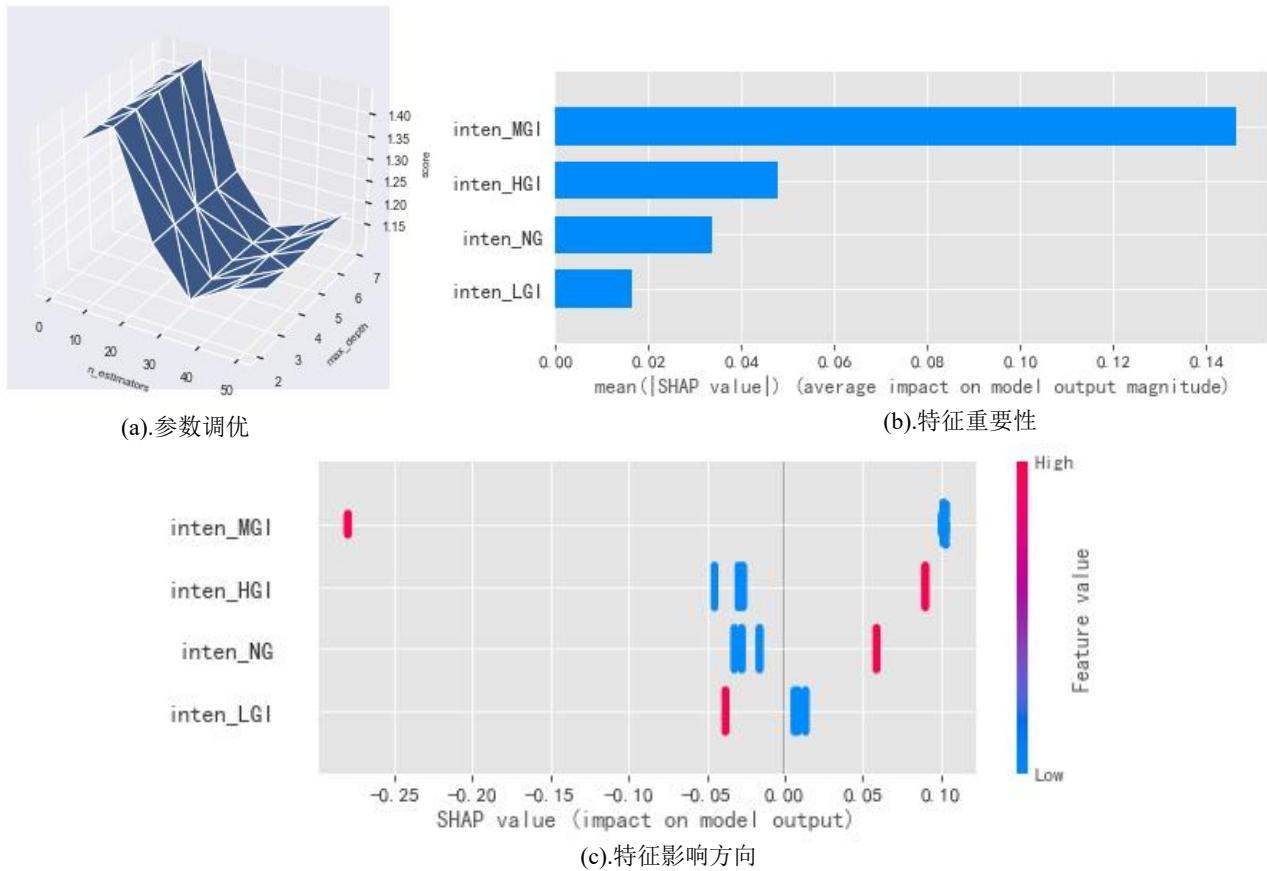


图 50 土壤 C/N 比情况下的参数调优图与特征影响图

6.6 总结

本章节基于多项式回归从机理分析的角度构建并求解了放牧强度对土壤化学性质影响的数学模型，并利用普通最小二乘法 OLS 回归对 2022 年各放牧强度下各小区的土壤化学性质进行预测，其中，主要使用 CCPR 图检验、拟合图、偏回归图、模型残差图对模型的拟合效果进行检验。然后本文使用决策树回归、线性回归、随机森林回归、AdaBoost 决策树回归、GradientBoost 回归、XGBoost 回归、CatBoost 回归、神经网络线性回归、LGBMBoost 回归等 9 种回归模型进一步建立了放牧强度对土壤化学性质影响的回归模型，利用 SHapley Additive exPlanation(SHAP) 模型分别对影响不同土壤化学性质的最佳回归模型进行建模，利用机器学习解释放牧强度对不同土壤化学性质的影响，其影响机制可总结如表 16 所示，从表中可以看出不同放牧强度对土壤化学性质的促进作用是不一样的，本文所得出的结论与前人研究基本保持一致^[17-18]，如全 N 含量会随着放牧强度的增加而下降，但由于数据样本不一样，不同的论文在研究结论上有所差异，研究中结论一致的内容说明本文构造方法的有效性，即本文利用机器学习的方法，对预测方法进行了优化。

表 16 不同放牧强度对土壤化学性质的影响机制

放牧强度	SOC 土壤有机碳	SIC 土壤无机碳	STC 土壤全碳	全 N	土壤 C/N 比
NG	上升	下降	----	上升	上升
LGI	上升	----	上升	下降	下降
MGI	下降	下降	下降	下降	下降
HGI	----	上升	上升	下降	上升

七、问题四的求解：综合评价法与粒子群优化算法

7.1 问题分析和技术路线图

问题四包括三个子问题，各个子问题分析如下：

子问题一：要求利用沙漠化程度指数预测模型确定不同放牧强度下监测点的沙漠化程度。针对该问题，首先需要整理确定影响沙漠化程度指数的变量，其次需要确定沙漠化指数的各变量的系数。根据题目提供的沙漠化程度指数指标，我们发现题目所提供的数据集有限，且难以收集到更多其他相关数据，因此，在该问题中，我们根据已有文献重新定义影响沙漠化程度的相关指标，然后使用变异系数法确定各指标的系数，进而来预测不同放牧强度下监测点的沙漠化程度。在该部分涉及到附件 15、附件 8、附件 6 和附件 2 等相关数据。

子问题二：子问题二需要给出定量的土壤板结化定义和建立合理的土壤板结化模型。由于题目中没有更多相关的数据，因此，首先我们从其他网站上收集关于土壤有机物、土壤湿度和土壤的容重的数据，然后使用熵值法确定各指标的权重系数，进而建立土壤板结化模型。

子问题三：给出放牧策略模型，使得沙漠化程度指数与土壤板结化程度最小。考虑到根据子问题和子问题二可以求解出关于沙漠化程度指数和土壤板结化的公式，基于这两个公式，使用粒子群算法进行求解。随机森林缺失值填充法和 IterativeImputer 多变量缺失值法。问题四的求解技术路线图如图 51 所示。

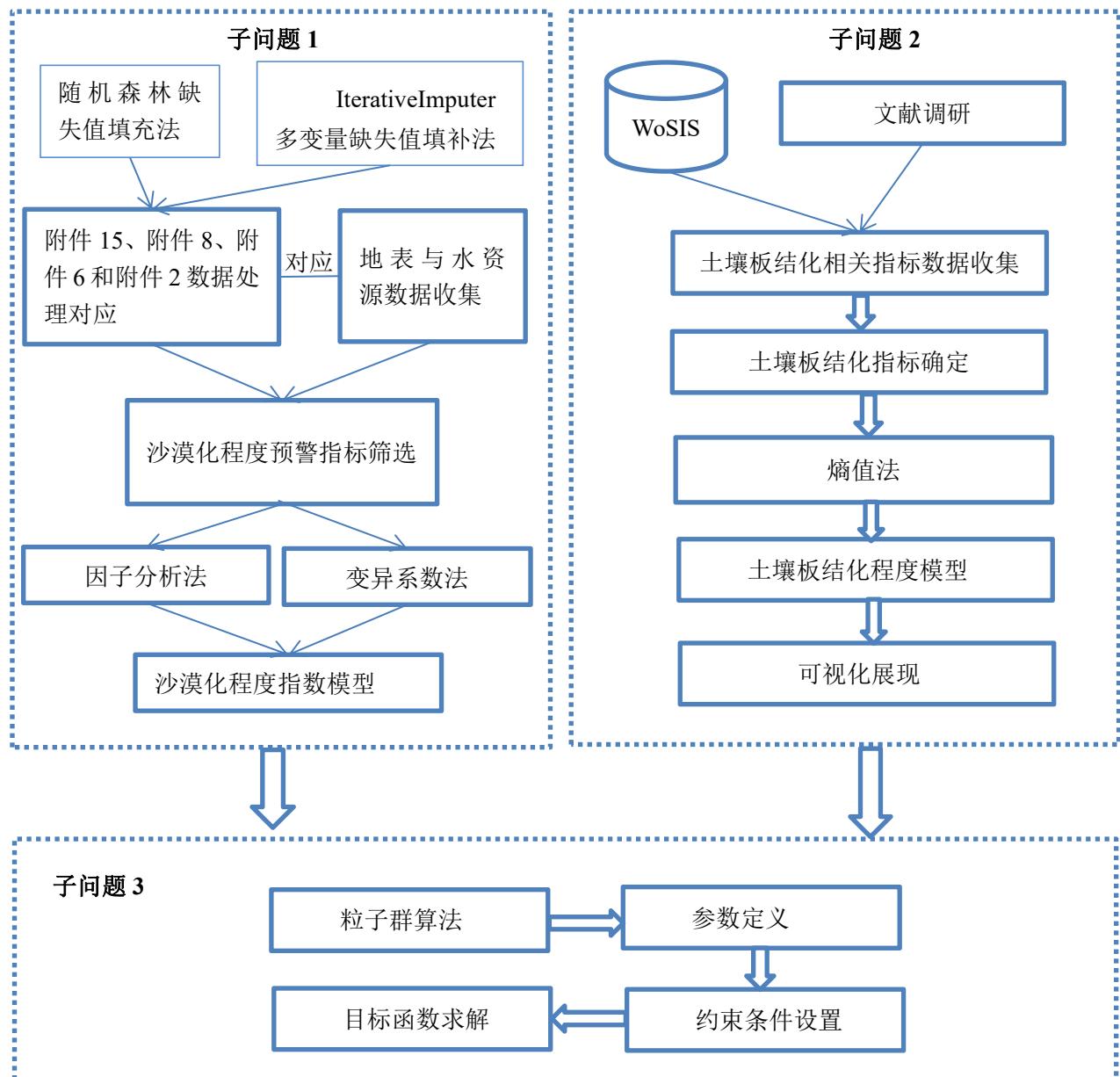


图 51 问题四技术路线图

7.2 基于变异系数法和因子分析的沙漠化程度预测模型

7.2.1 变异系数法和因子分析

变异系数法是一种客观权重赋值法^[19]，变异系数法的主要原理在于其能够直接利用指标中含有的具体信息计算指标内部和横向的规律，使用变异系数法能够消除人为因素的影响。当指标当前值和目标值的变异程度较大时，说明该指标更难实现目标，因此应当赋予更大的权重，反之，应对该指标赋予更小的权重。变异系数法的主要步骤如下。

(1) 数据收集与整理，首先需要进行数据收集，确定数据的 m 个评价指标以及 n 个数据样本，其可概念化为公式 (7-1) 所示，其中 x_{ij} 表示第 i 个样本第 j 个评价指标的数值。

$$X = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{pmatrix} \quad (7-1)$$

(2) 计算均值和标准差, 如公式 (7-2) 所示, 需要计算每个评价指标的均值和标准差。

$$\begin{cases} \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \\ S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \end{cases} \quad (7-2)$$

(3) 计算变异程度, v_j 表示第 j 个指标的变异系数, \bar{x}_j 和 S_j 为上一步得到的第 j 个指标的均值和标准差。

$$v_j = \frac{S_j}{\bar{x}_j} \quad (7-3)$$

(4) 计算各个指标的权重, 如公式 (7-4) 所示

$$W_j = \frac{v_j}{\sum_{j=1}^m v_j} \quad (7-4)$$

(5) 综合评分公式, 如公式 (7-5) 所示

$$M_i = \sum_{j=1}^m W_j * x_{ij} \quad (7-5)$$

此外, 需要对数据进行标准化, 本文中对数据标准化的处理都是使用最大最小标准化法, 其计算公式如 (7-5) 所示。其中, \max 为第 j 个指标中的最大值, \min 为第 j 个指标的最小值。

$$x'_i = \frac{x_i - \min}{\max - \min} \quad (7-6)$$

因子分析是从变量群中提取共性因子的统计分析技术^[20], 其与主成分思想类似。考虑到后续所使用的变量过多, 本文使用因子分析法提取其中若干个因子以进行建模。使用特征值对因子数量进行筛选。因子分析法的重点在于根据初等载荷矩阵, 构造因子模型, 如公式 (7-7) 所示。

$$\begin{cases} \widetilde{x_1} = b_{11}F_1 + b_{12}F_2 + \dots + b_{1t}F_t \\ \widetilde{x_p} = b_{p1}F_1 + b_{pm}F_2 + \dots + b_{pt}F_t \end{cases} \quad (7-7)$$

在子问题一的解决中, 综合使用因子分析法和变异系数法对沙漠化指数进行建模。

7.2.2 沙漠化程度预警指标筛选与相关数据收集和预处理

根据刘敦利学者所构建的沙漠化程度指数预测模型表达式^[21], 可知, 与沙漠化指数相关的数据包括三个气象因素: 风速、气象、降水; 三个地表因素: 植被盖度、地表水资源、地下水位; 三个人文因素: 人口数量、牲畜数量、社会经济水平等。

我们对现有的数据集进行分析, 发现数据存在缺失, 因此单纯使用刘敦利学者所构建的沙漠化程度指数预测难以保证效果。为了使建立的模型能够具有一定的专业性, 我们收集现有的文献, 发现沙漠化同植物生长和多样性也具有密切的相关关系, 因此, 本文构建沙漠化相关的指标流程如下:

(1) 植物生长多样性因素。考虑附件 15 的植物群落功能群、营养苗、株/丛数、鲜重(g)、干重(g)等与植物生长多样性相关的指标。从图 52 可知, 附件 15 存在一定的数据缺失, 因此需要对数据进行数据缺失值填补, 使用随机森林缺失值填充法和 IterativeImputer 多变

量缺失值填补法对缺失数据进行填补。

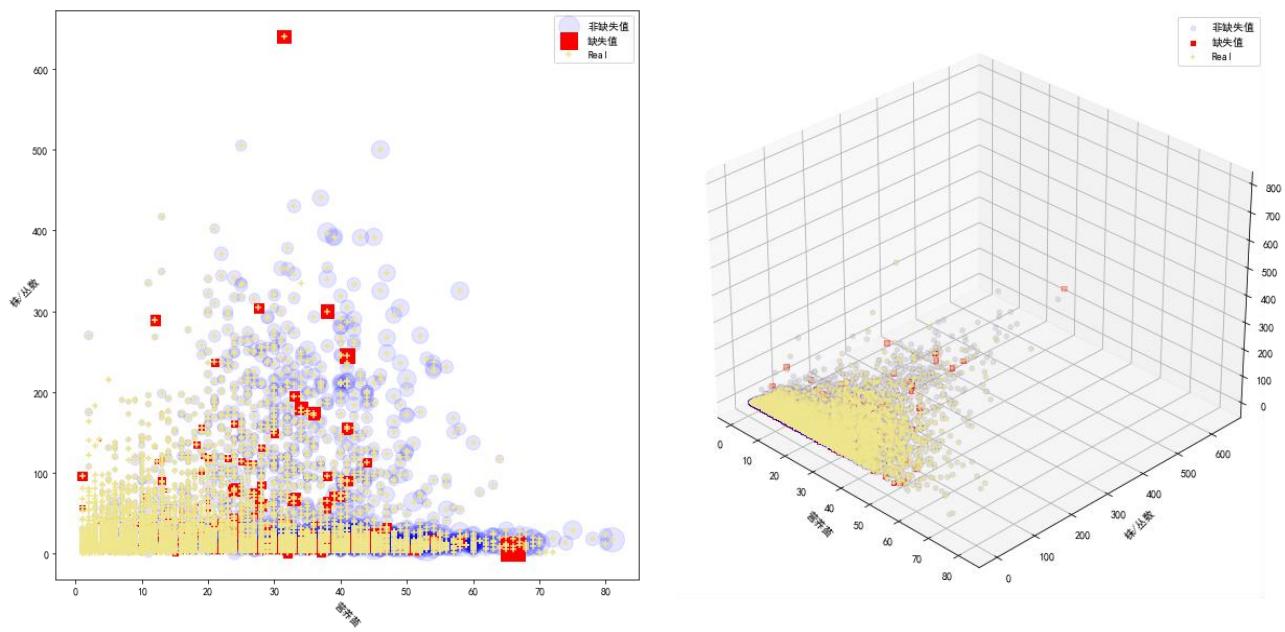


图 52 附件 15 真实数据和缺失数据展示

(2) 气象因素。对附件 8 锡林郭勒盟 2012-2022 年数据进行处理，考虑到气象因素与风速、气象、降水有关，我们选择与气象因素相关的指标为平均气温($^{\circ}\text{C}$)、降水量(mm)、降水天数和平均风速(knots)。当气温越高的时候、降水量越少、降水天数越小、平均风速越大时，沙漠化程度越高。值得注意的是，在刘敦利学者所构建的预警指标体系中，尚未考虑指标的正负性，如降水量，事实上，当降水越多的时候，沙漠化指标应当是越小的^[22]。根据附件 8 气象因素数据，我们将其匹配到附件 15 中所筛选的植物生长多样性因素中。

(3) 地表与水资源因素。地表因素主要有植被盖度、地表水资源、地下水位等。其中植被盖度我们使用“附件 6、植被指数-NDVI2012-2022 年”中的植被指数(NDVI)作为替换。考虑到锡林郭勒草原地表因素缺失，本文通过大量的数据收集，最终确定锡林郭勒地表数据来源于内蒙古自治区水利厅网站中关于内蒙古自治区水资源公报的文档数据中^[23]，具体网址如下:<http://slt.nmg.gov.cn/xxgk/bmxxgk/gbxx/szygb/>。最终本文整合 2012-2022 年内蒙古自治区水资源公报的水资源因素如表 17 所示。最终地表与水资源因素中设计的指标包括植被指数(NDVI)、地表水资源、地下水资源、地表水与地下水资源重复量、总水资源和产水系数。考虑总水资源的目的和地下水的目的是考虑到水资源的调用会缓慢沙漠化进程，因此一个地方的水资源对于沙漠化的影响也是重要的。

表 17 关于锡林郭勒盟水资源的数据收集概况

年份	地表水资源	地下水资源	地表水与地下水资源重复量	总水资源	产水系数
2012	8	30.43	3.94	34.49	0.06
2013	9.47	28.31	3.97	33.81	0.06
2014	6.84	26.3	3.69	29.45	0.06
2015	3.75	26.42	2.29	27.87	0.05
2016	4.9	33.53	2.91	35.52	0.07
2017	4.67	19.37	3.19	20.85	0.06

2018	5.09	35.35	4.06	36.38	0.05
2019	4.05	25.64	2.63	27.06	0.05
2020	3.78	28.85	2.27	30.36	0.05

(4) 人文因素。人文因素包括人口数量、牲畜数量、社会经济水平，我们从“附件 2、锡林郭勒统计年鉴（2016-2021）”中收集关于人文因素的数据，具体而言，包括常驻人口、牲畜数量和社会经济水平，具体数据情况如表 18 所示。其中牲畜数量以放牧强度表示

表 18 人文因素数据整合统计

年份	常驻人口	牲畜数量[放牧数量]	社会经济水平
2012	104.9	1178.52	114.5
2013	105.25	1288.19	109.9
2014	105.83	1476.89	108
2015	106.53	1555.67	107.6
2016	107.39	1630.17	107
2017	108.23	1460.16	104.2
2018	109.08	1340.29	105.3
2019	109.85	1297.71	103.4
2020	110.88	1311.99	103.4

(5) 最终关于沙漠化程度指标的数据如图 53 所示。在图中标记了各个指标对沙漠化程度影响的正负性，这一点在现有文献中大多未考虑到。其中，正负性意味着，指标对于结果大小的影响，当指标为正时，意味着指标越大，结果越大；当指标为负时，意味着指标越大，结果越小。

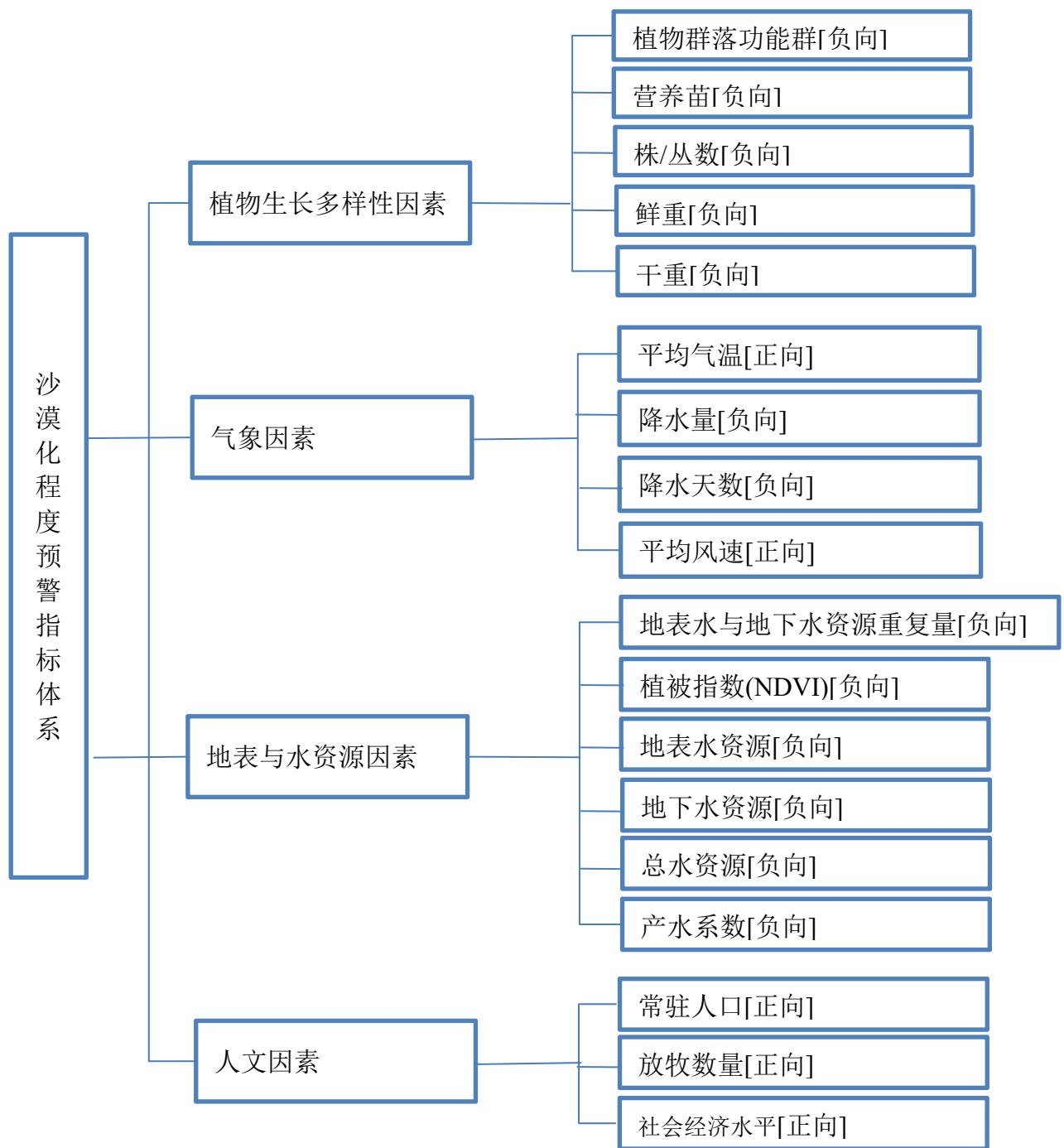


图 53 沙漠化预警指标体系

7.2.3 基于因子分析的沙漠化程度预警指标确定

考虑到现有的指标的维度过多，基于 7.2.1 章节的对因子分析的描述，本文基于因子分析法重新构建新的因子。图 54 为基于特征值的因子个数筛选的方法，当因子个数为 10 的时候，特征值逐渐趋向于稳定。因此，选择因子个数为 10。各个沙漠化程度预警指标对因子的贡献如图 55 所示。从图中可以看出，各个变量对因子分析贡献不同。

对于因子 0，主要的沙漠化程度预警指标贡献是产水系数、常驻人口、社会经济水平、地表水资源，因此因子 0 是负向指标。

对于因子 1，主要的沙漠化程度预警指标贡献是鲜重、干重，因此因子 1 是负向指标。

对于因子 2，主要的沙漠化程度预警指标贡献是地下水资源、总水资源、社会经济水平，因此因子 2 是负向指标。

对于因子 3，主要的沙漠化程度预警指标贡献是地表水资源、地表水与地下水资源重复量，因此因子 3 是负向指标。

对于因子 4，主要的沙漠化程度预警指标贡献是降水量、降水天数，因此因子 4 是负向指标。

对于因子 5，主要的沙漠化程度预警指标贡献是平均气温，因此因子 5 是正向指标。

对于因子 6，主要的沙漠化程度预警指标贡献是放牧数量，因此因子 6 是正向指标。

对于因子 7，主要的沙漠化程度预警指标贡献是植物群落功能群，因此因子 7 是负向指标。

对于因子 8，主要的沙漠化程度预警指标贡献是平均风速，因此因子 8 是正向指标。

对于因子 9，主要的沙漠化程度预警指标贡献是营养苗，因此因子 9 是负向指标。

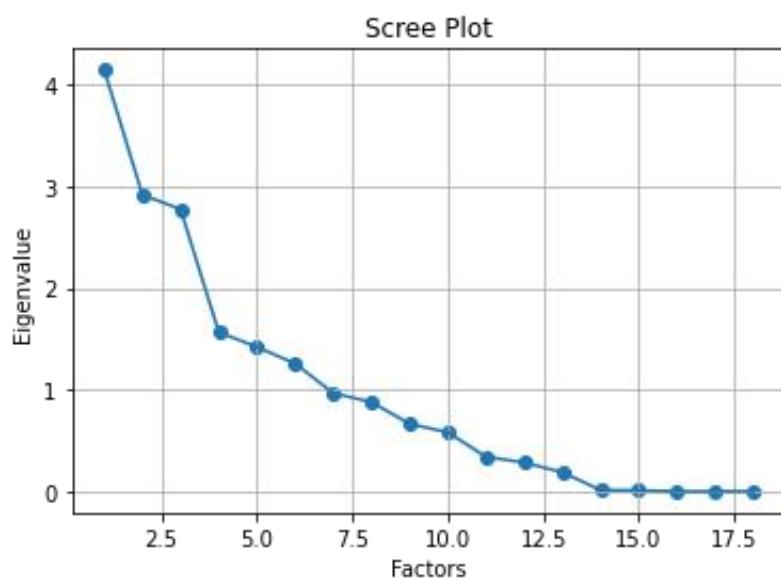


图 54 基于特征值的因子个数确定



图 55 沙漠化程度预警指标对因子影响的结果

7.2.4 基于变异系数法的沙漠化程度指数

基于 7.2.1 章节的对客观权重赋值法变异系数法的描述以及 7.2.3 章节筛选出的因子，得到各个因子的权重如表 19 所示

表 19 沙漠化程度因子权重表

因子序号	权重系数	因子正负性
因子 0	0.123496	负向
因子 1	0.203641	负向
因子 2	0.093844	负向
因子 3	0.127602	负向
因子 4	0.137632	负向
因子 5	0.071563	正向
因子 6	0.109575	正向
因子 7	0.066930	负向
因子 8	0.042229	正向
因子 9	0.023489	负向

对于各个因子，可以建立沙漠化程度指标公式如 (7-8) 所示。

$$SM = -0.123496X_0 - 0.203641X_1 - 0.093844X_2 - 0.127602X_3 - 0.137632X_4 + 0.071563X_5 + 0.109575X_6 - 0.066930X_7 + 0.042229X_8 - 0.023489X_9 \quad (7-8)$$

然后将 SM 标准化为[0-100]之间的数值，最终，得到不同放牧强度下监测点的沙漠化程度指数值如表 21 所示。根据沙漠化程度预警体系，即表 20 所示，可以得到每个地区的沙漠化程度。

表 20 沙漠化程度及沙漠化程度指数划分标准

划分内容	划分类型				
沙漠化程度	非沙漠化	轻度沙漠化	中度沙漠化	重度沙漠化	极重度沙漠化
沙漠化程度指数	[0,0.20]	(0.20,0.40]	(0.40,0.60]	(0.60,0.80]	(0.80,1.00]

表 21 不同放牧强度下监测点沙漠化程度

监测点 Block	放牧强度	沙漠化程度	沙漠化预警程度
G11	MGI	70.30785742	重度
G12	LGI	57.6438557	中度
G13	HGI	85.4772628	极重度
G16	MGI	70.64090929	重度
G17	NG	39.59474692	轻度
G18	LGI	61.59454453	重度
G19	NG	44.44388848	中度
G20	HGI	76.45256228	重度
G21	NG	33.93166652	轻度
G6	LGI	47.48898541	中度
G8	MGI	69.60135851	重度
G9	HGI	76.25881605	重度

7.3 基于熵值法的土壤板结化预测模型

7.3.1 基于熵值法的综合评价法

熵值法同变异系数法一样是一种客观权重赋值法^[24]，熵值法主要用于判断某个指标的离散程度，当该指标的离散程度越大的时候，说明该指标提供的信息量也就越多，这时，指标对于结果的影响也就越大。熵值法的重点在于计算信息熵和信息效用值。对于公式(7-1)所示的数据原始矩阵 $X=((x_{ij})_{n*m})$ ，则第 j 个指标的信息熵计算公式如公式 (7-9) 所示。

其中 e_j 表示第 j 个指标的信息熵， p_{ij} 为第 j 个指标下的第 i 个数值占该列全部数值和的比例

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (7-9)$$

然后计算第 j 个指标的信息效用值如公式 7-10 所示。

$$d_j = 1 - e_j \quad (7-10)$$

之后，计算第 j 个指标的权重如公式 (7-11) 所示。

$$W_j = \frac{d_j}{\sum_{j=1}^m d_j} \quad (7-11)$$

最后，基于权重和指标数据，计算土壤板结化的综合评价值如公式（7-12）所示。

$$F = \sum_{j=1}^m W_j * X_{ij} \quad (7-12)$$

7.3.2 土壤板结化相关指标数据收集

根据题目提示，土壤板结化与土壤有机物、土壤湿度和土壤的容重有关，其数学模型可定性描述为如下：

$$B = f(W, C, O) \quad (7-13)$$

土壤湿度 w 越少，容重 c 越大，有机物含量 o 越低，土壤板结化程度 B 越严重。

在题目中，由于题目中的数据样本各指标数据不全，关于土壤板结化相关的指标数据考虑另外收集。在阅读大量文献的基础上，本文从 Batjes^[25]、Hengl^[26]等学者发布的论文中找到关于世界土壤信息可视化数据，这些数据经过质量评估和标准化，考虑了土壤的化学性质：有机碳、总碳、总碳酸盐当量、总氮、磷（可提取磷、总磷和磷保留）、土壤 pH 值、阳离子交换容量和电导率以及土壤的物理特性：土壤质地（沙子、淤泥和粘土）、容重、粗碎块和保水性。根据学者所提供的 WoSIS 服务，如图 56 和图 57 所示，展示了部分世界土壤有机碳密度分布和世界土壤容重分布情况，从土壤有机碳密度分布来看，中国丽江、昆明等地土壤有机碳含量较高，从土壤容重来看，中国土壤容重分布较为良好，而印度、巴基斯坦等国家土壤容重值较高；此外，中国的北京、济南等地区土壤容重相较于中部地区如武汉、贵阳等地而言，其容重值更大。

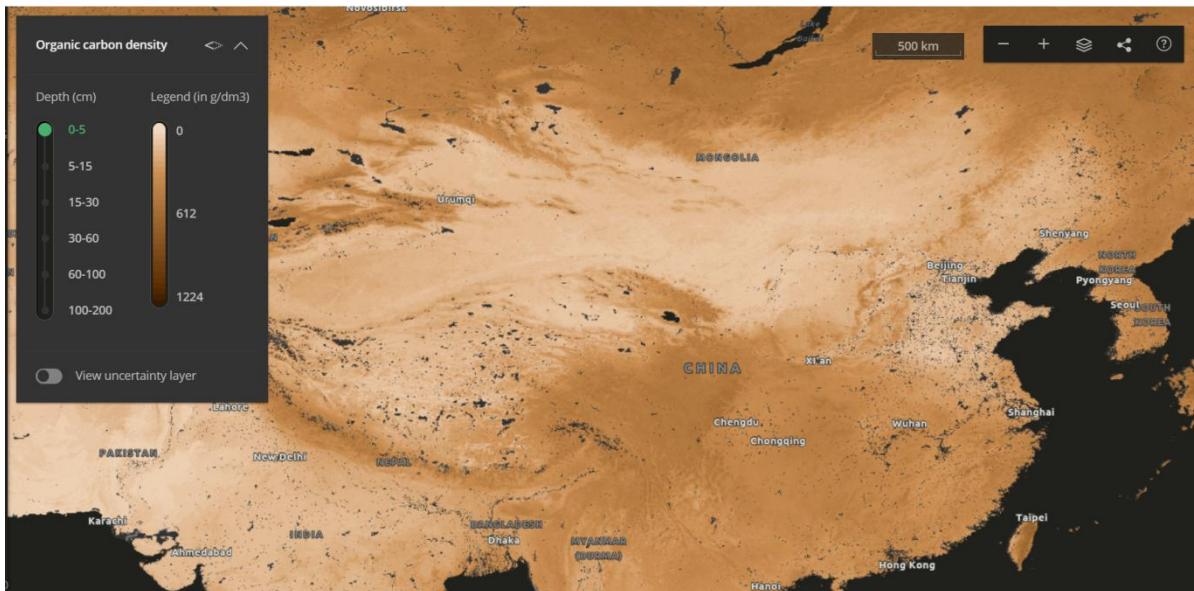


图 56 世界土壤有机碳密度分布



图 57 世界土壤容重分布

根据 WoSIS 服务提供的可视化数据，本文从地图中随机选择地图点，根据每个地图点所展示的数据一个个收集信息，最终随机选择确定了如表 22 所示的有关土壤物理性质和化学性质的 50 个数据点。

表 22 土壤板结化相关数据收集

序号	维度	经度	土壤容重 (cg/cm³)	有机碳 密度 (g/dm³)	土壤有机 碳存储 (t/ha)	粘土含 量 (g/kg)	沙砾 (g/kg)	氮(cg/kg)	土壤有 机碳 (cg/kg)		水的 pH (pH*10)
1	29.4117	106.0091	127	426	45	255	333	352	338	62	
2	39.3345	104.3835	140	155	25	262	359	140	73	84	
3	31.2004	98.3182	107	540	87	177	406	529	843	64	
4	32.48	95.3748	110	505	77	155	464	467	655	74	
5	35.5731	106.9695	123	319	37	252	239	320	339	80	
6	28.7601	98.4002	103	542	102	180	443	641	1018	58	
7	32.5639	97.848	101	457	69	151	422	544	736	65	
8	40.862	94.0356	132	144	19	191	322	113	60	81	
9	27.6271	114.8685	116	384	42	249	312	287	350	55	
10	33.8536	87.5349	111	162	30	261	382	170	123	81	
11	25.2061	114.8755	124	340	48	289	375	184	213	55	
12	41.0126	74.1108	112	464	56	212	385	417	704	65	
13	13.8455	105.3015	123	301	48	125	306	235	284	57	
14	34.8935	89.1414	109	183	42	179	614	231	187	79	
15	-16.2925	132.2743	143	209	32	251	607	78	111	65	
16	50.8909	96.8215	90	518	73	229	534	388	709	62	
17	-2.9724	25.947	130	324	45	291	470	215	347	54	
18	63.3896	95.3217	58	645	85	163	510	1064	1855	52	
19	-23.1712	-57.2695	119	267	33	180	601	223	294	58	
20	69.7723	105.8353	52	648	70	198	379	794	1440	58	
21	63.7754	-69.1587	67	608	69	169	488	725	1089	57	

22	52.8717	82.3885	97	498	63	251	544	458	921	68
23	67.0458	-51.8276	69	634	91	231	421	739	1306	54
24	37.1601	58.0252	137	214	29	244	289	166	197	77
25	53.7832	-93.0167	96	499	66	278	383	459	582	57
26	53.561	13.6955	133	419	41	145	618	701	513	61
27	47.6125	-97.9725	128	532	143	286	368	479	664	70
28	31.6934	8.7857	142	141	18	266	414	56	51	81
29	81.0283	-94.0518	84	659	99	220	298	1153	2217	66
30	50.4207	-72.4989	96	592	90	163	475	676	742	51
31	6.0862	23.2812	124	301	44	320	430	258	387	57
32	66.1967	-141.9901	55	517	43	176	255	979	2227	56
33	-22.2843	27.911	156	162	25	169	703	88	73	67
34	-41.0744	-65.9799	111	227	19	131	662	65	74	78
35	-23.0231	-48.551	109	466	63	309	505	359	479	53
36	1.1148	-56.7624	115	304	43	296	412	223	282	44
37	54.9563	-69.7463	65	618	84	135	611	829	335	53
38	70.3219	-93.8972	96	443	52	290	348	442	732	66
39	49.0199	-99.5451	119	541	71	292	307	465	676	68
40	73.2959	-94.4271	102	528	60	306	264	597	1034	70
41	32.4242	43.1667	148	152	16	354	264	49	52	80
42	42.5119	100.0548	140	165	21	325	333	138	90	80
43	64.7197	-70.7736	66	618	75	216	427	735	1241	58
44	61.1925	-98.2898	64	654	59	261	464	501	790	59
45	24.4833	82.3482	161	192	34	325	292	130	128	72
46	52.9622	-112.9522	115	508	75	293	330	489	611	63
47	63.8206	-135.1078	73	500	63	185	337	624	811	55
48	56.0053	-117.1256	109	577	78	286	326	497	467	63
49	42.0345	-100.9416	133	380	41	121	723	517	298	65
50	40.0905	113.107	138	157	22	124	529	119	75	84

7.3.3 基于熵值法的土壤板结化程度预测

土壤板结化与土壤有机物、土壤湿度和土壤的容重有关，根据收集到的数据，其中土壤有机物包括有机碳密度、土壤有机碳存储、氮和土壤有机碳，由于世界土壤湿度难以获取，本文在调研文献的基础上，发现相当一部分研究探索了土壤湿度同水的 pH 值的关系，Ghazali 等人^[27]建立多个二元回归模型发现土壤湿度触发土壤 pH 值的变化高达 100%，在一些周期性洪水中，土壤水分含量显然增加了土壤 pH 值^[28-29]，因此，本文使用水的 pH 值代表土壤湿度，土壤容重则由数据给出。

根据 7.3.1 的熵值法计算方法，得到土壤板结化的各个指标权重如下表所示。

表 23 土壤板结化预警指标权重

	权重系数	影响因素	权重系数	指标因子	权重系数
土壤板结化预	1	土壤容重 C 土壤有机物 O	0.179107 0.597139	土壤容重C ₁ 有机碳密度O ₁	0.179107 0.325018

警指标体系	土壤有机碳存储O ₂	0.070349
	氮O ₃	0.108816
	土壤有机碳O ₄	0.092956
土壤湿度 W	0.223754	土壤水分 pH 值

根据表 7-7 土壤板结化预警指标权重，构建土壤板结化程度预测公式如公式（7-14）和（7-15）所示，其中，公式（7-15）为土壤板结化程度预测的简写模式。

$$B = 0.179107C_1 - 0.325018O_1 - 0.070349O_2 - 0.108816O_3 - 0.092956O_4 - 0.223754W_1 \quad (7-14)$$

$$B = f(W, C, O) = 0.179107C - 0.597139O - 0.223754W \quad (7-15)$$

根据公式（7-14）和公式（7-15）可以得到表 7-8 所示的土壤板结化程度值，根据前文对土壤沙漠化程度指数的划分，土壤板结化也可以划分为无板结化[0,20]、轻度板结化(20,40]、中度板结化(40,60]、重度板结化(60,80]和极重度板结化(80,100]。基于该标准，展示收集到的数据土壤板结化程度世界地图分布如图 58 所示。

表 24 各地区土壤板结化程度值（部分）

序号	维度	经度	土壤板结化程度值	土壤板结化程度
1	29.4117	106.0091	60.64303846	重度
2	39.3345	104.3835	71.80598983	重度
3	31.2004	98.3182	42.85660002	中度
4	32.48	95.3748	41.91990648	中度
5	35.5731	106.9695	57.38478089	中度
6	28.7601	98.4002	42.74775118	中度
7	32.5639	97.848	47.82542828	中度
8	40.862	94.0356	73.51380942	重度
9	27.6271	114.8685	66.14210484	重度
10	33.8536	87.5349	67.49344694	重度
11	25.2061	114.8755	71.48553167	重度
12	41.0126	74.1108	51.30230259	中度
13	13.8455	105.3015	71.84349898	重度
14	34.8935	89.1414	65.42658138	重度
15	-16.2925	132.2743	79.6000868	重度
16	50.8909	96.8215	45.30004295	中度
17	-2.9724	25.947	73.32293778	重度
18	63.3896	95.3217	25.443839	轻度
19	-23.1712	-57.2695	73.66661098	重度
20	69.7723	105.8353	26.17836156	轻度

■ 重度板结 ■ 中度板结 ■ 轻度板结 ■ 无板结 ■ 极重度板结

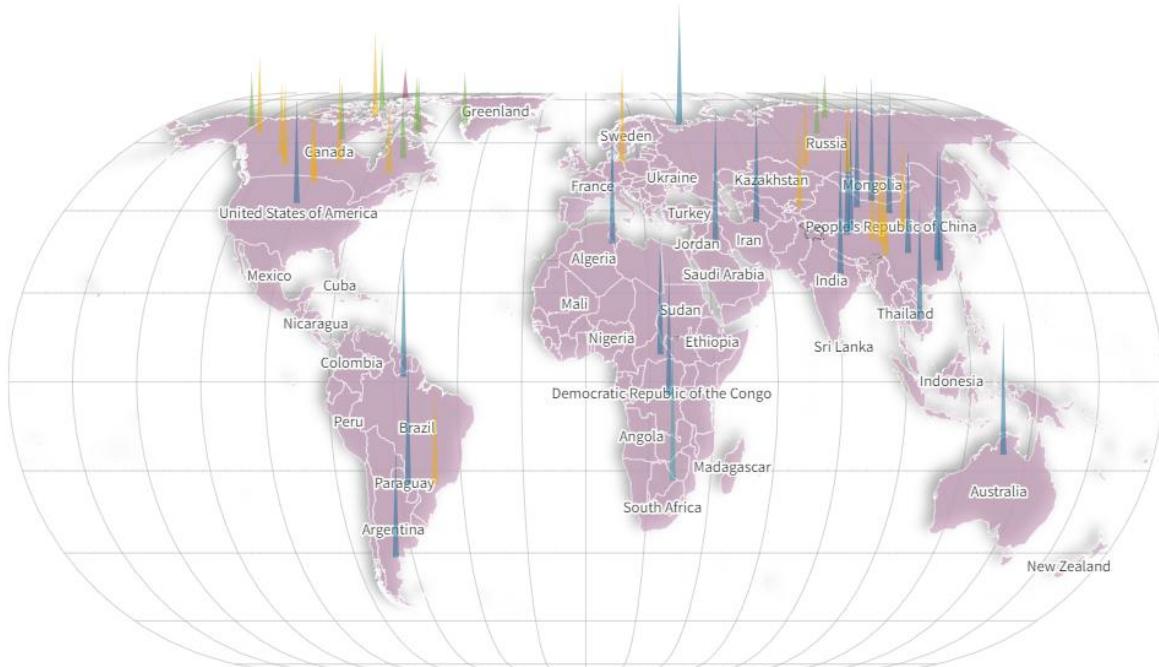


图 58 基于熵值法的土壤板结化程度世界地图分布

7.4 基于粒子群算法（PSO）的放牧策略模型求解

该子问题需给出放牧策略模型，使得沙漠化程度指数与土壤板结化程度最小，我们将放牧策略模型定义为离散的放牧强度，即 NG, LGI, MGI, HGI，在求解荒漠化程度指数与土壤板结化程度最小时时，将其作为单目标优化问题，在本问题中，依据上文给出的土壤板结化定义和沙漠化指数定义，采用粒子群算法来选取放牧强度及其他与荒漠化指数、土壤板结化程度相关的变量，获取最优解，从而得出放牧策略。

7.4.1 粒子群算法

粒子群算法是一种精度高、易控制、收敛快，源于鸟群捕食行为研究的进化算法，通过设计粒子来模拟鸟类^[30]，代表优化问题中可行性的解，每个粒子拥有三个属性——速度、位置和适应度值。每个粒子在搜索空间中单独搜索个体最优解 $Pbest$ ，并将其记为当前个体极值与整个粒子群的粒子共享，找到的最优的那个个体极值将作为整个粒子群的当前全局最优解 $Gbest$ ，所有粒子再通过当前全局最优解和个体极值来调整自己的位置，直到找到满足条件的全局最优解。

我们定义粒子维度为 D ，粒子的数量即种群大小为 N ，假定种群在第 t 次迭代产生的第 $i \in \{1, 2, \dots, N\}$ 个粒子的位置为：

$$present_i(t) = (x_{i,1}(t), x_{i,2}(t), \dots, x_{i,D}(t)) \quad (7-16)$$

那么在第 t 次迭代产生的第 i 个粒子的速度可以表示为：

$$v_i(t) = (v_{i,1}(t), v_{i,2}(t), \dots, v_{i,D}(t)) \quad (7-17)$$

对于种群在 $t + 1$ 次迭代时粒子的速度以及位置依据分别进行更新：

$$\begin{aligned} v_i(t+1) &= \omega * v_i(t) + c_1 * rand * (Pbest_i - present) \\ &\quad + c_2 * rand * (Gbest_i - present) \end{aligned} \quad (7-18)$$

$$present_i(t+1) = present_i(t) + v_i(t+1) \quad (7-19)$$

其中 $Pbest_i$ 表示第*i*个粒子历史飞行中的最优位置, $Gbest_i$ 表示全部粒子历史飞行中的最优位置。 ω 称为惯性因子, 其值为非负, 值越大, 全局寻优能力越强, 局部寻优能力越弱, 这个值的引入可以让粒子群算法性能大大提高, 针对不同的搜索问题可以调整全局和局部搜索能力。 c_1 和 c_2 表示控制粒子探索时的非负数学习因子, 取值为[0,1], 增加搜索随机性, $rand$ 表示随机值。在更新粒子速度时分别为记忆项、自身认知项和群体认知项相加, 记忆项指的是上一次迭代的速度大小和方向, 自身认知项是指当前点指向粒子自身最好点的一个矢量, 表示粒子的动作来源于自己经验部分, 群体认知是指一个从当前点指向种群目前最好点的矢量, 反映了粒子间的协同合作和知识同享。粒子通过自己和其他粒子的经验来确定下一步运动。

7.4.2 算法求解过程

在利用 PSO 算法求解问题时, 流程图如图 59 所示。首先进行初始化, 将*N*个粒子的位置和速度进行随机初始化, 并计算对应的目标函数的取值, 得到目标函数的最小值, 寻找到个体和群体历史最优位置。再基于设定的参数*R*和 V_{max} , 运用粒子群算法的迭代步骤, 依据当前最优位置和自身位置的评估, 得出粒子的速度更新, 进而计算位置更新。之后评估粒子更新后的适应性函数值, 再次选取个体和全局最优解, 如果达到最大迭代次数, 则输出结果, 否则再次进入迭代步骤。

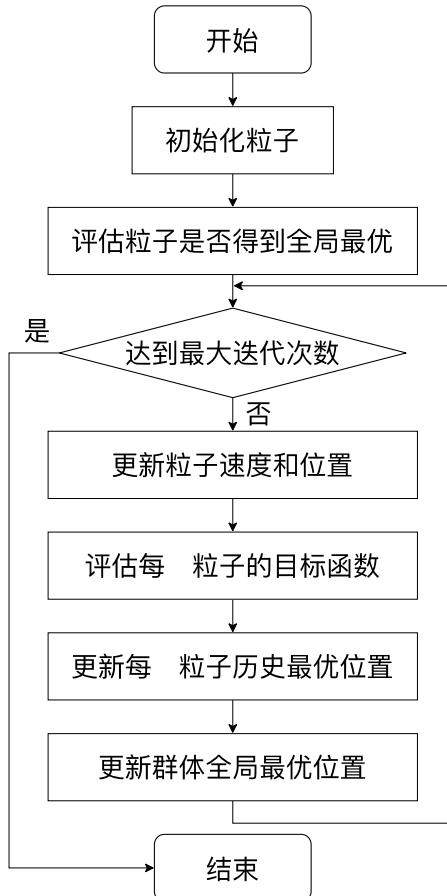


图 59 粒子群算法流程图

(5) 参数设定

在本文中上述参数设定如表 25 所示:

表 25 粒子群算法参数设定

参数	描述	取值
N	粒子数量	100
D	粒子维度	196
R	粒子范围	$[\min(x_i), \max(x_i)]$
V_{max}	最大速度	$R * 0.15$
c_1, c_2	社会认知	0.8
ω	惯性权重	1
I	迭代次数	1000

(6) 子问题描述

待求解的子问题及其描述如表 26 所示:

表 26 粒子群算法子问题描述

子问题	描述
X_1	产水系数
X_2	常驻人口
X_3	地下水资源
X_4	总水资源
X_5	地表水与地下水资源重复量
X_6	降水量
X_7	平均气温
X_8	放牧强度

(7) 目标函数

该目标函数我们将沙漠化指数和土壤板结化指数视为同等重要的指标，所以我们对沙漠化指数和土壤板结化程度求解调和平均数作为最终目标函数，以平衡两个指标的互相影响。其形式化定义如下：

$$F = \frac{2*(SM(X_1, X_2, \dots, X_8))*B(SM(X_6, X_8))}{(SM(X_1, X_2, \dots, X_8))+B(SM(X_6, X_8))} \quad (7-20)$$

其中 SM 和 B 为上文定义的沙漠化指数和土壤板结化程度。

(8) 约束条件

本文定义的约束条件如表 27 所示:

表 27 约束条件定义

约束条件	描述
$0.5 \leq X_1 \leq 0.7$	
$107 \leq X_2 \leq 111$	
$19 \leq X_3 \leq 36$	该约束条件均为分析锡林郭勒草原进五年涉及的子问题波动范围定义。
$20 \leq X_4 \leq 37$	
$2.2 \leq X_5 \leq 4.1$	
$0 \leq X_6 \leq 180$	
$10 \leq X_7 \leq 25$	
$X_8 \in \{0,1,2,3\}$	放牧强度为离散变量，分别代表

	NG, LGI, MGI 和 HGI
--	--------------------

7.4.3 放牧策略求解结果

根据本文设定的参数、子问题、目标函数及约束条件，求得子问题的解如表 28 所示：

表 28 放牧策略求解结果

子问题	描述	结果
X_1	产水系数	0.5762
X_2	常住人口	107.5883
X_3	地下水资源	26.3739
X_4	总水资源	26.4570
X_5	地表水与地下水资源重复量	3.4347
X_6	降水量	44.6670
X_7	平均气温	19.7374
X_8	放牧强度	1

7.4.4 结果分析

根据粒子群算法的求解结果，可以得到放牧策略为，当月降水量约为 45mm 时时，平均气温约为 19 度时，采用轻牧的放牧强度，将会使沙漠化指数和土壤板结化程度最低。

7.5 总结

本章节将问题四分为三个子问题，对于第一个子问题，由于数据缺失，本章节重新定义了沙漠化程度预警指标体系，在结合附件 2、附件 6、附件 8 和附件 15 以及人工收集相关的文本数据的基础上，使用基于变异系数法和因子分析法建立沙漠化程度指数模型，考虑到现有指标未确定指标的正负性，本文在对建立模型时考虑了不同因子指标对结果影响的方向，最终确定了沙漠化程度指数模型公式和同放牧强度下监测点沙漠化程度。

对于第二个子问题，关于土壤板结化预测模型的建立，由于现有的数据不能满足模型建立的标准，本文随机收集了来自 WoSIS 服务的 50 个地区的土壤化学性质和物理性质，对于其中的土壤湿度，考虑到土壤湿度和水的 pH 值高度相关，使用水的 pH 值作为土壤湿度的评价指标，最后基于熵值法确定 50 个地区的土壤板结化程度，并可视化这 50 个地区土壤板结化程度，其中当土壤板结化程度值在 80 以上为极重度土壤板结化、在 60 到 80 之间为重度土壤板结化，在 40 到 60 间为中度土壤板结化，在 20 到 40 间为轻度板结化，小于 20 为无土壤板结化。

对于第三个子问题，需要给出放牧策略模型，使得沙漠化程度指数与板结化程度最小，根据公式 (7-8) 和公式 (7-16)，使用基于粒子群算法求解使得沙漠化程度指数与板结化程度最小的值，最终得到月降水量约为 45mm 时时，平均气温约为 19 度时，采用轻牧的放牧强度可以使得沙漠化程度指数与板结化程度最小。

八、问题五的求解：融合回归模型的基于分解的多目标进化算法

8.1 问题分析与技术路线图

问题五需在给定降水量的情形下，在保持草原可持续发展情况下找到实验草场内放牧羊数量的最大阈值。现代草地资源经营的可持续发展原则是在保证生态环境良好健康发展中寻求经济利益的最大化，依据问题四，我们将生态环境良好定义为草场土壤板结化程度和沙漠化程度小，且生物多样性丰富。一味得追求经济效益即扩大草场放牧羊的数量，将对生态环境造成很大影响，为实现草原资源的可持续发展，需在土壤板结化、沙漠化、生物多样性以及经济效益中进行权衡，因此我们将该问题看作为一个多目标优化问题，并主要利用融合回归模型的多目标优化算法来获取在保持草原可持续发展情况下的放牧羊数量的最大阈值。求解问题五的技术路线图如图 60 所示。

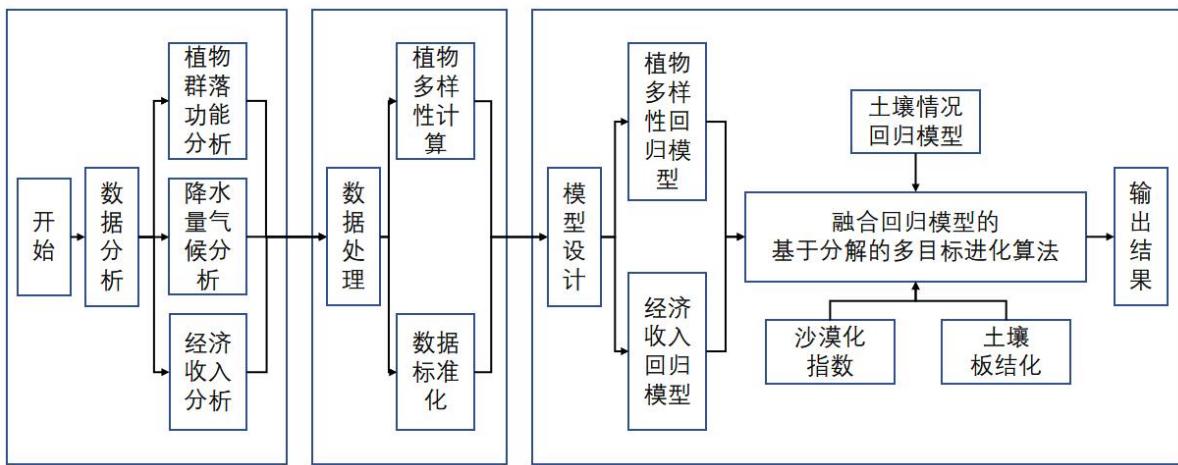


图 60 问题五技术路线图

8.2 数据预处理

对于问题五的四个目标，土壤板结化和沙漠化模型在问题四中已经给出，经济效益和放牧羊数量存在正相关关系，在数据处理部分，我们主要进行草原植物多样性的计算。

在附件 15 中，给出了实验草地 12 个地块在放牧前、四次不同强度放牧后的群落结构监测数据，其中，我们使用了植物群落功能 CF ，统计对于具备群落功能的 CF_i 物种丰度 F ，其中 $i \in \{1, 2, 3, 4\}$ ， CF_1 表示植物群落功能为多年生丛草型禾草（PB）， CF_2 表示植物群落功能为多年生根茎型禾草（PR）， CF_3 表示植物群落功能为一二年生草本（AB）， CF_4 表示植物群落功能为多年生杂草（PF），如表 29 所示。

表 29 植物群落功能及物种数表示

群落功能 小区	CF_1 (PB)	CF_2 (PR)	CF_3 (AB)	CF_4 (PF)
$Block_n$	j_1	j_2	j_3	j_4

表中 $Block_n$ 代表某个放牧小区， j_i 代表属于 CF_i 植物群落功能的物种数。

接下来利用 α 多样性中的 ACE 指数来计算实验地块植物丰富度指数 R ，公式如下：

$$\left\{ \begin{array}{l} R = F_{abund} + \frac{F_{rare}}{C_{ACE}} + \frac{F_1}{C_{ACE}} \gamma_{ACE}^2 \\ \gamma_{ACE}^2 = \max \left[\frac{F_{rare} \sum_{j=1}^{10} j(j-1) F_j}{C_{ACE} N_{rare} (N_{rare}-1)} - 1, 0 \right] \\ C_{ACE} = 1 - \frac{F_1}{N_{rare}} \\ N_{rare} = \sum_{j=1}^{10} j * F_j \end{array} \right. \quad (8-1)$$

其中, F_{abund} 指丰度大于 10 的群落功能数, F_{rare} 指丰度大于 0 且小于等于 10 的群落功能数, F_j 指丰度为 j 的群落功能数。

对于每个地块的重复实验, 我们取重复实验的平均值, 作为该地块的最终植物丰富度指数。

8.3 模型介绍

8.3.1 降水量、放牧强度与植物多样性回归模型

通过对每年每实验地块 (G6、G8、G9、G11、G12、G13、G16、G17、G18、G19、G20、G21) 的植物多样性取平均值, 得到了图 61 所示的植物多样性与放牧强度、降水量关系折线图, 其中 G17、G19、G21 代表无放牧强度 (NG), G6、G12、G18 代表低强度放牧 (LGI), G8、G11、G16 代表中强度放牧 (MGI), G9、G13、G20 代表重度放牧 (HGI)。

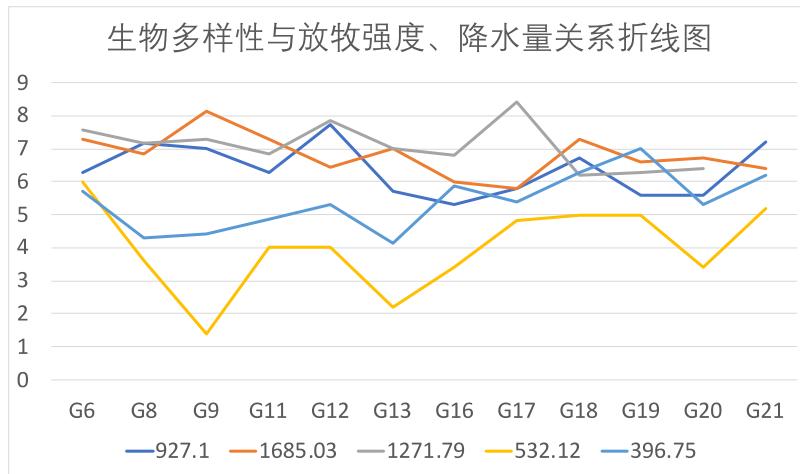


图 61 植物多样性与放牧强度、降水量关系折线图

我们使用了多维多项式线性回归来建模实验草场中, 植物多样性和降水量、放牧强度的关系模型。

$$\tilde{R} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 (x_1)^2 + a_5 (x_2)^2 + \dots \quad (8-2)$$

其中 \tilde{R} 是植物多样性的预测值, x_1 是降水量, x_2 是放牧强度。根据 8.2 中生物多样性指标计算公式, 可以得到生物多样性指标的取值范围为 0.0-10.0。在该问题中, 我们将放牧强度定义为 0 代表 NG, 1 代表 LGI, 2 代表 MGI, 4 代表 HGI。

同时分析了气候附件中近十年锡林郭勒草原的年降水量情况。2012-2021 年年降水量如图 62 所示, 最大值为 1685.03mm, 最小值为 302.01mm, 平均值为 842.29mm。

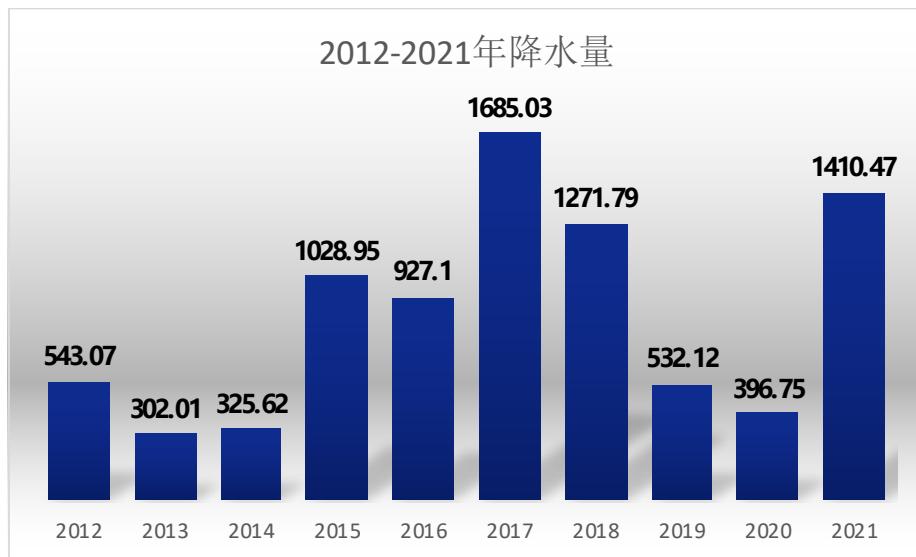


图 62 锡林郭勒 2012-2021 年降水量趋势图

年降水量值相对于生物多样性指标及放牧强度变量，数值较大，为此，我们采用最小最大 (min-max) 数据标准化方法，将回归模型涉及的变量进行标准化。

$$M(V) = \frac{V - \text{Min}}{\text{Max} - \text{Min}} \quad (8-3)$$

8.3.2 经济收入与羊群关系回归模型

依据附件 13，我们分析了放牧强度与每公顷经济收益的关系如图 63 所示，其中横坐标表示每天每公顷放牧羊的数量。从图中可以看到，随着放牧强度强度的增加，单位面积的经济收入会出现波动，当每公顷每天放牧羊数量在 3 到 4 之间时，单位面积经济收入较低。

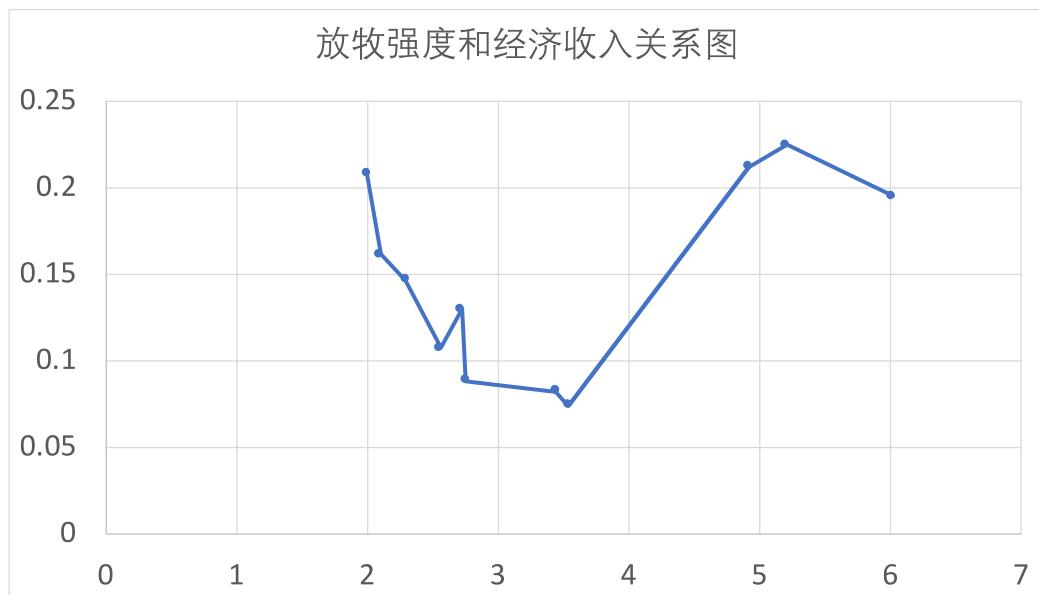


图 63 放牧强度与单位公顷经济收入关系图

我们使用多项式回归来建模放牧强度与经济收益的关系模型。

$$\tilde{Y} = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3 \quad (8-4)$$

其中 \tilde{Y} 表示每公顷面积的经济收益， x_1 表示放牧强度。

8.3.3 基于分解的多目标进化算法 (MOEAD)

基于分解的多目标进化算法^[31]将一个多目标优化问题分解为一组单目标优化问题，并对它们同时优化，通过利用与每一个子问题相邻的子问题的优化信息来优化它本身计算复杂度耕地，正态化的 MOEAD 算法还可以解决规模范围相异的多目标问题。

一个多目标优化问题可以表示为：

$$\text{Max or Min } F(x) = [f_1(x), f_2(x), \dots, f_m(x)]^T, x \in \Omega \quad (8-5)$$

其中， m 表示目标函数的数量， $\Omega = \{x \in R^n | h_j(x) \leq 0, j = 1, 2, \dots, n\}$ 表示决策空间， $h_j(x)$ 是连续的函数。

如果目标函数互斥，同时对所有目标求最优解往往是没有意义的，传统的多目标优化算法通常获取一个维持目标函数间平衡的解，即 Pareto 最优解，作为最终的解。而 MOEAD 算法利用分解的思想，将多目标优化问题分解为 N 个标量的子问题，通过进化出一个解的种群来同时解决所有子问题。对于每一代种群，种群是从所有代中选出的每一个子问题的最优解的集合。相邻的两个子问题的关联程度是由聚合系数向量间的距离所决定的，对于两个相邻子问题来说，最优解是非常相似的，对于每一个子问题来说，MOEAD 使用相邻的子问题的信息来优化。

(1) 参数列表

MOEAD 算法所需要的参数如表 30 所示。

表 30 MOEAD 算法参数列表

参数	符号	意义
迭代次数	I	停止条件
子问题	X	$X = \{x^1, x^2, \dots, x^N\}$
子问题个数	N	
目标函数	f	$f = \{f_1, f_2, \dots, f_m\}, m$ 为目标函数的数量
目标函数值	F	$F = \{F_1, F_2, \dots, F_N\}, F_i$ 是问题的目标函数值向量
最优解集合	z	$z = \{z_1, z_2, \dots, z_m\}$
权重向量	$\lambda^1, \lambda^2, \dots, \lambda^N$	均匀分布的 N 个权重向量
邻居个数	T	每个权重的邻居个数
邻居	B	$B(i) = \{i_1, i_2, \dots, i_T\}, i \in N$
初始化种群	P	
外部种群	EP	存储搜索过程中找到的非支配解

(2) 求解过程

MOEAD 算法的求解过程如图 64 所示，首先应进行初始化，设定一个空集 EP ，计算任意两个权重向量之间的欧式距离，然后求解出每个权重向量的 T 的邻居，用 $B(i) = \{i_1, i_2, \dots, i_T\}, i \in N$ 表示。之后随机初始化种群 P ，计算 $F_i = F(x^i)$ 。再初始化 $\bar{z} = (z_1, z_2, \dots, z_m)^T$ ，令 $z_i = \min \text{ or } \max \{f_i(x^1), f_i(x^2), \dots, f_i(x^N)\}$ 。之后对 $i = 1, 2, \dots, N$ 开始进行迭代，从 $B(i)$ 中选取该子问题邻居问题的当前最优解的索引 k, l ，融合基因重组思想，使用遗传算子由 x^k, x^l 生成新解，对改节运用切比雪夫法进行改善，以满足约束条件，生成 y' ，如果 $z_j < f_j(y')$ ，更新最优解结合，另 $z_j = f_j(y')$ ，之后在更新邻居解，

令 $F^j = F(y')$, $j \in B(i)$, 并通过聚合函数值 F 更新父代种群, 从从 EP 中移除 $F(y')$ 的支配向量;如果EP中不存在支配 $F(y')$ 的向量就将 $F(y')$ 加入EP。如此一直迭代, 直到满足停止条件, 即达到最大迭代数, 停止迭代, 输出结果。

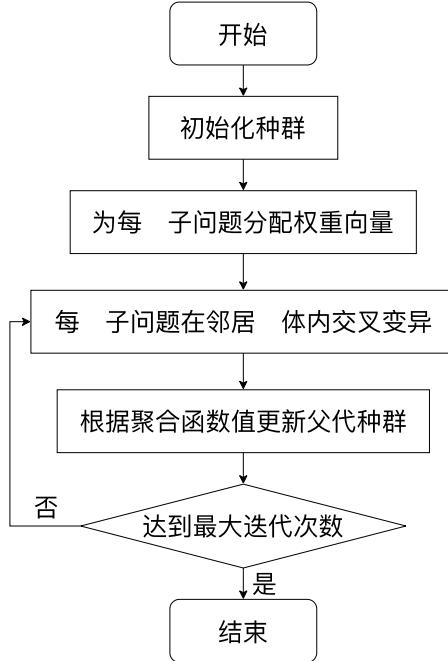


图 64 MOEAD 算法流程图

8.3.4 融合回归模型的多目标优化算法

经上述分析, 我们已经拟合了放牧强度、降水量和生物多样性的模型, 放牧强度和经济发展的模型, 在问题三中, 我们已经得到了土壤板结化和沙漠化指数模型。问题五当中, 需分析在给定降水量的情况下, 求得实验操场内羊群在保持草原可持续发展的情况下最大阈值。为了保持草场可持续发展, 我们需在草原植物多样性、放牧经济、土壤板结化及沙漠化指数中取得平衡。

根据问题三得到的土壤板结化和沙漠化指数模型, 其与降水量、放牧强度、平均气温等各项指标均有关联。为此, 我们分析了气候相关条件, 并利用皮尔逊相关系数对气候的各项指标进行相关性分析, 其公式如下:

$$\begin{cases} \rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \\ r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \end{cases} \quad (8-6)$$

其中 X , Y 分别为气候的指标变量, \bar{X} , σ_X 分别是对 X_i 样本的样本平均值和样本标准差。

得到的气候相关指标, 如月份、平均气温、平均最高气温、平均最高气温极值, 最低气温极值, 平均露点气温、降水量、降水天数、积雪深度、能见度、风速等相关性系数如图 65 所示, 其中颜色越偏近绿色, 即系数值偏近 0, 代表两个指标之间没有线性相关关系。

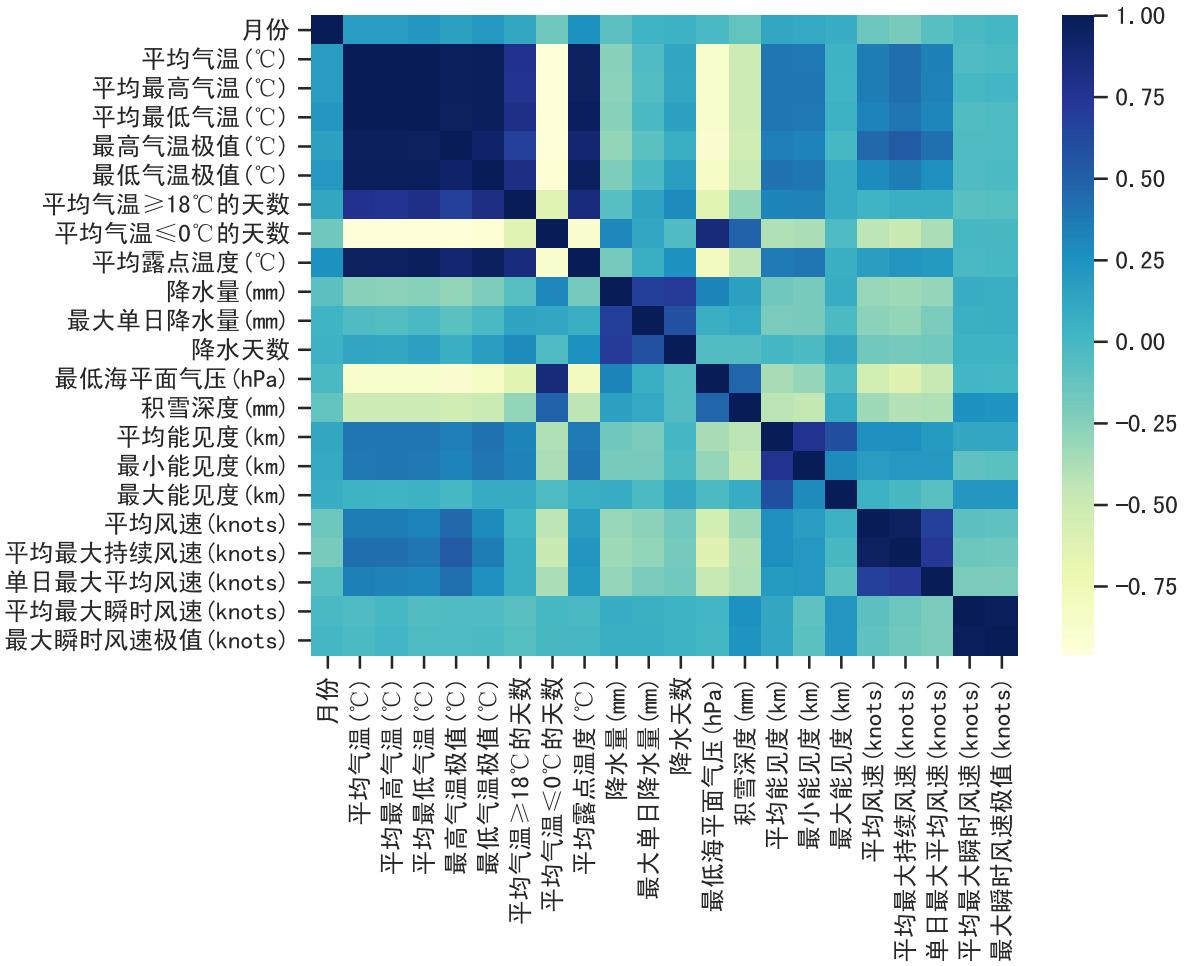


图 65 气候指标相关性分析

从图中可以发现，降水量和其他指标的相关性均较弱，在沙漠化指数和土壤板结化指数中涉及的相关指标与其他指标的相关性均较弱，为此，我们融合回归模型对基于分解的多目标优化算法设计如下的初始化参数、子问题、目标函数及约束条件。

(1) 初始化参数列表

我们设置的融合回归模型的 MOEAD 算法初始化参数如下表所示。

表 31 模型初始化参数列表

初始化参数	描述
I	迭代次数，设置为 1000
N	种群数量，设置为 100
$Water$	降水量, $Water = \{300, 600, 900, 1200\}$
$DReg$	植物多样性指数回归模型
$MReg$	经济水平回归模型
$TReg$	土壤情况回归模型
SM	沙漠化指数
B	土壤板结化程度

(2) 子问题描述

我们设置的融合回归模型的 MOEAD 算法子问题如下表所示。

表 32 模型子问题及描述

子问题	描述
X_1	放牧强度
X_2	平均气温
X_3	总水资源
X_4	常住人口

(3) 目标函数

根据问题五需在草原可持续发展的原则下，求得放牧标准羊最大阈值的要求，我们定义了如下目标函数。

目标函数 1：

该目标函数是评估放牧带来的经济收益，该目标函数仅与放牧强度相关，其根据经济收入和放牧强度的回归模型得到，该目标函数应该越大越好。其形式化定义如下：

$$F1 = M(MReg(X_1)) \quad (8-7)$$

目标函数 2：

该目标函数来评估在一定放牧强度和降水量情况下，生物多样性的情况，在满足草原可持续发展的条件下，生物多样性应该越大越好。其形式化定义如下：

$$F2 = M(DReg(X_1, Water)) \quad (8-8)$$

目标函数 3：

该目标函数来评估在求解子问题的过程中，对于沙漠化指数的影响。草场荒漠化和草原植物多样性是矛盾的目标，且荒漠化会对草原的可持续发展造成不可逆的影响，因此，该目标函数应该越小越好。其形式化定义如下：

$$F3 = M(SM(X_1, X_2, X_3, X_4)) \quad (8-9)$$

目标函数 4：

该目标函数来评价在不同放牧情况下，对于土壤板结化的影响。由于放牧强度不同，会对土壤有机物和土壤容重有影响，而降水会对土壤湿度造成影响，土壤板结化表示土壤表层的有机质早到了严重破坏，不符合草原可持续发展原则，因此该目标应该越小越好。其形式化定义如下：

$$F4 = M(B(TReg_1(X_1, W), TReg_2(X_1), TReg_3(X_1))) \quad (8-10)$$

其中 $M()$ 为标准化函数，我们认为这四个目标函数在草原可持续发展中是同等重要的指标，因此我们均将其标准化为 0-1 之间的结果，使得融合回归模型的基于分解的多目标进化算法可以在四个目标函数间取得较好的平衡。

(4) 约束条件

我们设置的融合回归模型的 MOEAD 算法子问题如下表所示。

表 33 模型子问题及描述

约束条件	描述
$0 \leq X_1 \leq 4$	0 代表无牧强度，4 代表重度放牧 锡林郭勒草原放牧集中于每年 6-9 月， 近十年这四个月中，平均气温分布在
$10 \leq X_2 \leq 25$	15 度到 25 度之间

$20 \leq X_3 \leq 37$	经外部收集数据，该变量约束条件值设为 20-37
$107 \leq X_4 \leq 111$	经分析锡林郭勒统计年鉴，常住人口变量约为 107-111
$DReg(X_1, W) \geq 4.0$	该约束条件对草场的植物多样性进行约束，其在求解过程中需满足该约束，以保证草场的可持续发展

8.3.5 求解结果

根据题目要求，当降水量为 300 时，子问题解如表 34 所示。

表 34 Water = 300 子问题解

子问题	取值
X_1	3.2775
X_2	15.2897
X_3	24.3778
X_4	110.0371

根据题目要求，当降水量为 600 时，子问题解如表 35 所示。

表 35 Water = 600 子问题解

子问题	取值
X_1	2.6298
X_2	21.6047
X_3	34.1128
X_4	108.9525

根据题目要求，当降水量为 900 时，子问题解如表 36 所示。

表 36 Water = 900 子问题解

子问题	取值
X_1	3.1841
X_2	15.8201
X_3	30.9022
X_4	108.6267

根据题目要求，当降水量为 1200 时，子问题解如表 37 所示。

表 37 Water = 1200 子问题解

子问题	取值
X_1	4.1406
X_2	15.86075
X_3	29.6090
X_4	110.7115

8.3.6 结果分析

通过利用融合回归模型的基于分解的多目标进化算法，我们已经得到了不同降水量条件下的子问题求解结果，下面我们利用下述公式来求解不同降水量情况下的放牧标准羊的最大阈值。

$$N = X_{1,i} * 2 * PA \quad (8-11)$$

其中 $X_{1,i}$ 表示第*i*种降水量情况下的放牧强度取值， PA 表示实验草场的面积（公顷）。

通过利用该公式计算，可以求得，在满足草原可持续发展的原则下，当降水量为300时，实验草场放牧羊最大阈值为151.03；当降水量为600时，实验草场放牧羊最大阈值为121.18；当降水量为900时，实验草场放牧羊最大阈值为146.72；当降水量为1200时，实验草场放牧羊最大阈值为190.90。

8.5 总结

首先对和问题五相关的附件及外部分件进行分析并清理，得到植物多样性与降水量、放牧强度回归模型以及经济收入与放牧强度回归模型，并利用问题三、问题四得到的土壤情况回归模型、沙漠化指数及土壤板结化模型，之后设计和草原可持续发展原则相关的函数，并应用到融合回归模型的多目标进化算法当中，求解出放牧标准羊最大阈值的中间结果，即放牧强度，最后利用实验草场面积和放牧强度间的相关关系求出不同降水量条件下放牧羊的最大阈值。

九、问题六的求解：模型综合与动态演示

9.1 问题分析与技术路线图

该问题需要预测在示范区牧户放牧策略不变和问题四得到的放牧方案两种情况下，示范区 2023 年 9 月的土壤肥力、土壤湿度、植被覆盖等土地状态的情况。为求解该问题，我们利用了问题一、问题三得到的放牧强度同植被覆盖率模型和放牧强度同土壤情况模型来进行计算。其技术路线图如图 66 所示。

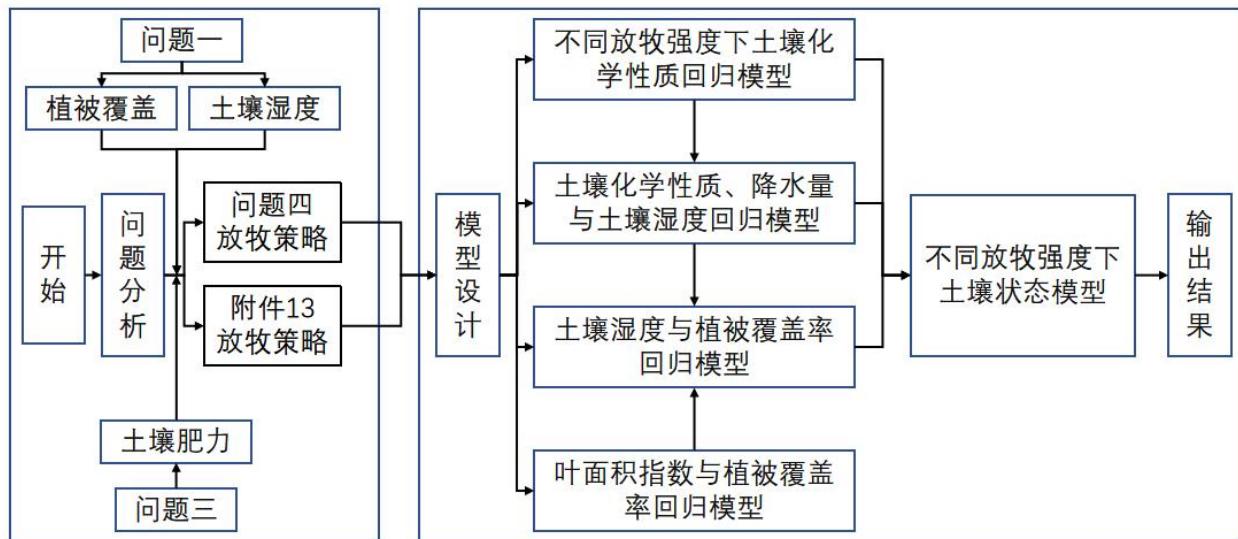


图 66 问题六技术路线图

9.2 不同放牧策略描述

9.2.1 不同牧户放牧策略

我们分析了附件 13 示范区的不同牧户放牧策略，根据牧户的标准羊数量、放牧面积，利用下属公式进行求解，得到每个牧户的放牧强度。

$$S_i = \frac{\sum_{j=1}^y N_{ij}}{M_i} \quad (9-1)$$

其中， N_{ij} 表示第 i 个牧户第 j 年的放牧标准羊数量， M_i 表示第 i 个牧户的放牧面积。

根据公式计算，得到四个牧户的放牧强度分别为 2.68、2.13、5.38、3.48（羊/天/公顷），对应题目描述，可以得到牧户 1 为中度放牧强度，牧户 2 为轻度放牧强度，牧户 3 为重度放牧强度，牧户 4 为中度放牧强度。

9.2.2 示范区牧户平均放牧策略

我们还根据牧户的标准羊数量、放牧面积，利用下属公式进行求解，得到牧户的平均放牧强度。

$$\bar{S} = \frac{\sum_{i=1}^n \bar{N}_i}{\sum_{i=1}^n M_i} \quad (9-2)$$

其中， n 表示牧户数量， \bar{N}_i 表示第 i 个牧户三年平均放牧标准羊数量。最终求解出 $\bar{S} = 3.54$ ，根据题目描述，我们将示范区内牧户的平均放牧策略定义为中等放牧强度。

9.2.3 问题四放牧策略

根据问题四求解结果，我们得到了在使沙漠化指数和土壤板结化程度最低时，应采用轻度放牧强度。

综合以上不同放牧策略，在下文中，主要分析轻度放牧强度策略（牧户 2、问题四）、中度放牧强度策略（牧户 1、牧户 4、示范区平均放牧强度）以及重度放牧强度（牧户 3）对示范区内 2023 年 9 月土地状态的影响。

9.3 不同放牧强度下土地状态回归模型

本问题需要求解在不同放牧策略下，2023 年 9 月土壤肥力、土壤湿度和植被覆盖率的情况。根据题目描述，土壤肥力的重要影响因素有土壤的化学性质（土壤有机碳 SOC、土壤无机碳 SIC、土壤全碳 STC、全 N、土壤 C/N 比等）和土壤物理性质（土壤湿度、土壤容重等），而土壤湿度与土壤化学性质和降水量相关，植被覆盖率与土壤湿度有关。所以我们将问题六分解为如下子问题：（1）子问题一：依据问题三，求解不同放牧强度下，土壤化学性质与放牧强度的关系模型；（2）子问题二：依据子问题一得到的土壤化学性质与放牧强度关系模型，建立土壤湿度与降水量、土壤化学性质的关系模型，在问题二中已经预测了 2022 和 2023 年锡林郭勒草原的降水情况，从而间接求解到土壤湿度与放牧强度的关系模型；（3）子问题三：依据子问题二得到的土壤湿度与放牧强度的关系模型，建立植被覆盖率与土壤湿度的关系模型，从而也可以间接求解植被覆盖率与放牧强度的关系模型；（4）子问题四：子问题一得到的土壤化学性质和放牧强度关系模型以及子问题而得到的土壤湿度和放牧强度关系模型，求解土壤肥力与放牧强度的关系模型。为此，我们通过多项式回归建立了不同放牧强度下土壤化学性质回归模型，土壤化学性质、降水量与土壤湿度回归模型，植被覆盖率与土壤湿度回归模型。

9.3.1 不同放牧强度下土壤化学性质回归模型

在不放牧情况下，单位时间内土壤有机碳和无机碳的变化与植被生物量（植物生长状况）直接相关（假设一只羊摄每天入植物量不变的情况下）。植被生物量与放牧情况存在关系。因此，依据问题三的求解过程，我们得到了在 9.2 提及的不同放牧强度策略情况下，土壤有机碳 SOC、土壤无机碳 SIC 以及全 N 同放牧强度以及时间的拟合模型。

（1）轻度放牧（LGI）下土壤化学性质拟合模型

A. SOC 土壤有机碳

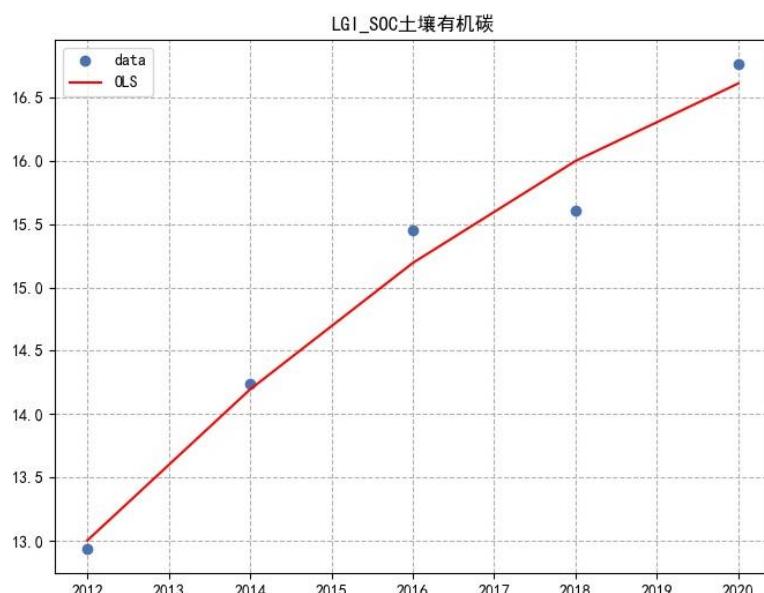


图 67 LGI 下 SOC 土壤有机碳拟合曲线

拟合曲线的参数如下:

回归截距: w0=[-99007.95866885]

回归系数: w1=[[0.00000000e+00 9.77870634e+01 -2.41410896e-02]]

R2 确定系数: 0.9703

均方误差: 0.0505

平均绝对值误差: 0.1838

中位绝对值误差: 0.1531

$$SOC(t) = -99007.95866885 + 97.7870634 * t - 0.0241410896 * t^2 \quad (9-3)$$

B. SIC 土壤无机碳

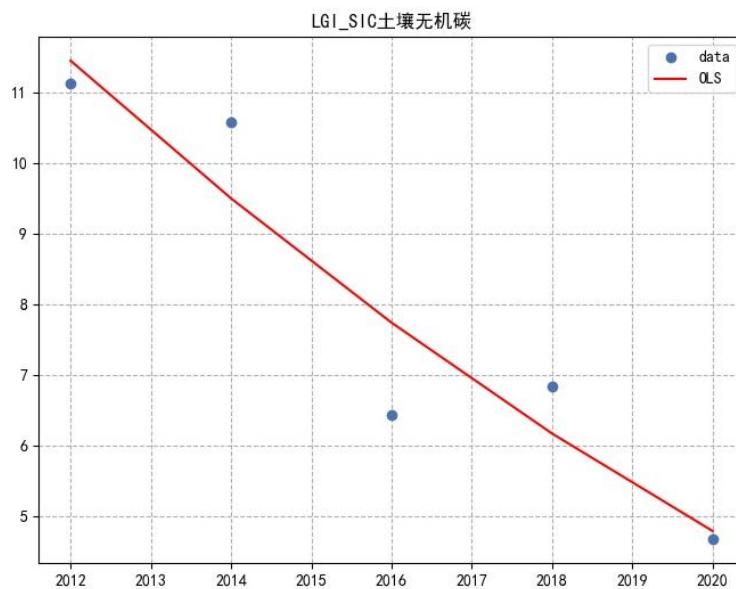


图 68 LGI 下 SIC 土壤无机碳拟合曲线

拟合曲线的参数如下:

回归截距: w0=[98311.13254523]

回归系数: w1=[[0.00000000e+00 -9.66905392e+01 2.37742724e-02]]

R2 确定系数: 0.8907

均方误差: 0.6837

平均绝对值误差: 0.6960

中位绝对值误差: 0.6632

$$SIC(t) = 98311.13254523 - 96.6905392 * t + 0.0237742724 * t^2 \quad (9-4)$$

C. 全 N

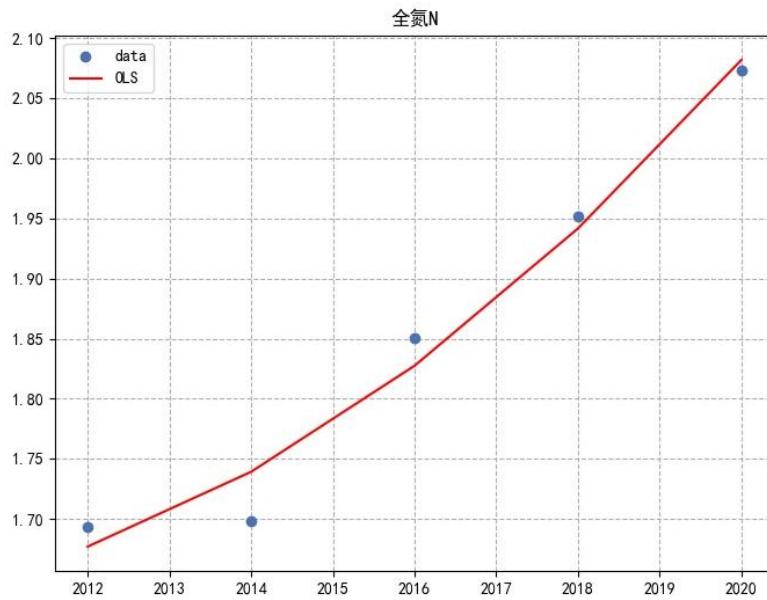


图 69 LGI 下全 N 拟合曲线

拟合曲线的参数如下：

回归截距: $w_0 = [13096.04036917]$

回归系数: $w_1 = [[0.0000000e+00 -1.30408505e+01 3.24687756e-03]]$

R2 确定系数: 0.9753

均方误差: 0.0005

平均绝对值误差: 0.0199

中位绝对值误差: 0.0166

$$N(t) = 13096.04036917 - 13.0408505 * t + 0.00324687756 * t^2 \quad (9-5)$$

(2) 中度放牧 (MGI) 与土壤化学性质拟合模型

A. SOC 土壤有机碳

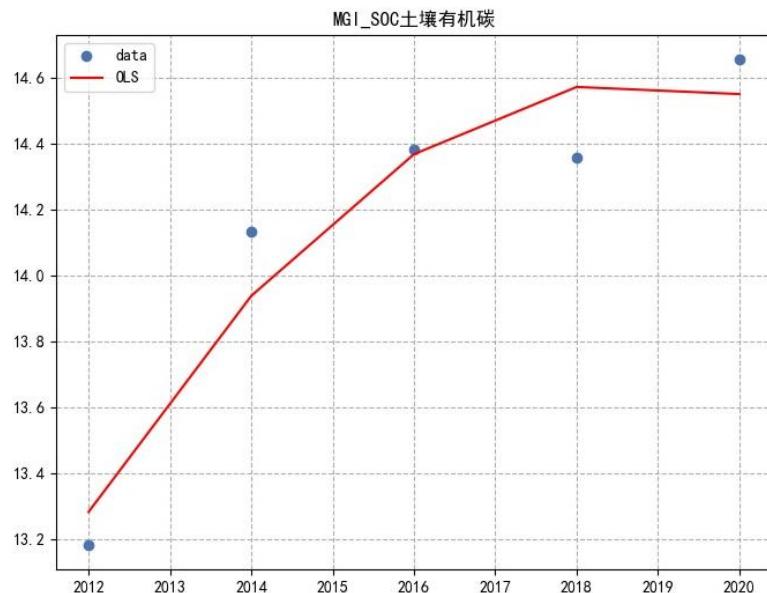


图 70 MGI 下 SOC 土壤有机碳拟合曲线

拟合曲线的参数如下：

回归截距: $w_0 = [-114950.35336289]$
 回归系数: $w_1 = [[0.00000000e+00 \quad 1.13893863e+02 \quad -2.82081906e-02]]$
 R2 确定系数: 0.9172
 均方误差: 0.0213
 平均绝对值误差: 0.1266
 中位绝对值误差: 0.1055
 $SOC(t) = -114950.35336289 + 113.893863 * t - 0.0282081906 * t^2 \quad (9-6)$

B. SIC 土壤无机碳

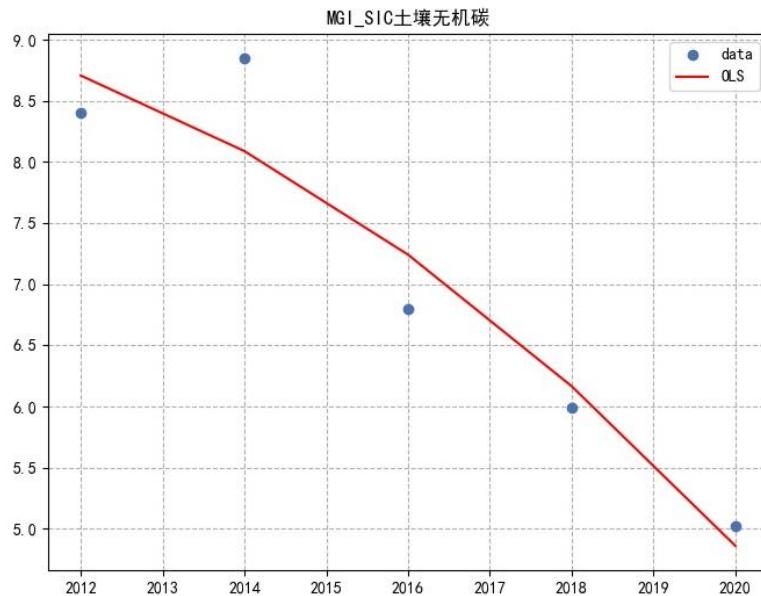


图 71 MGI 下 SIC 土壤无机碳拟合曲线

拟合曲线的参数如下:

回归截距: $w_0 = [-115389.60919689]$
 回归系数: $w_1 = [[0.00000000e+00 \quad 1.14961968e+02 \quad -2.86316799e-02]]$
 R2 确定系数: 0.9110
 均方误差: 0.1844
 平均绝对值误差: 0.3679
 中位绝对值误差: 0.3066
 $SIC(t) = -115389.60919689 + 114.961968 * t - 0.0286316799 * t^2 \quad (9-7)$

C. 全 N

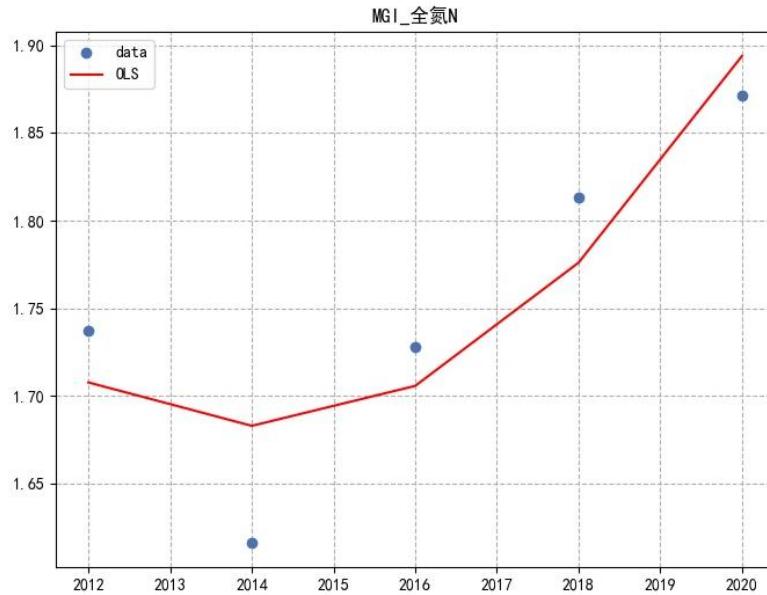


图 72 MGI 下全 N 拟合曲线

拟合曲线的参数如下：

回归截距: $w_0 = [24116.10911112]$

回归系数: $w_1 = [[0.0000000e+00 -2.39463078e+01 5.94484035e-03]]$

R2 确定系数: 0.7933

均方误差: 0.0015

平均绝对值误差: 0.0356

中位绝对值误差: 0.0297

$$N(t) = 24116.10911112 - 23.9463078 * t + 0.00594484035 * t^2 \quad (9-8)$$

(3) 重度放牧 (HGI) 与土壤化学性质拟合模型

A. SOC 土壤有机碳

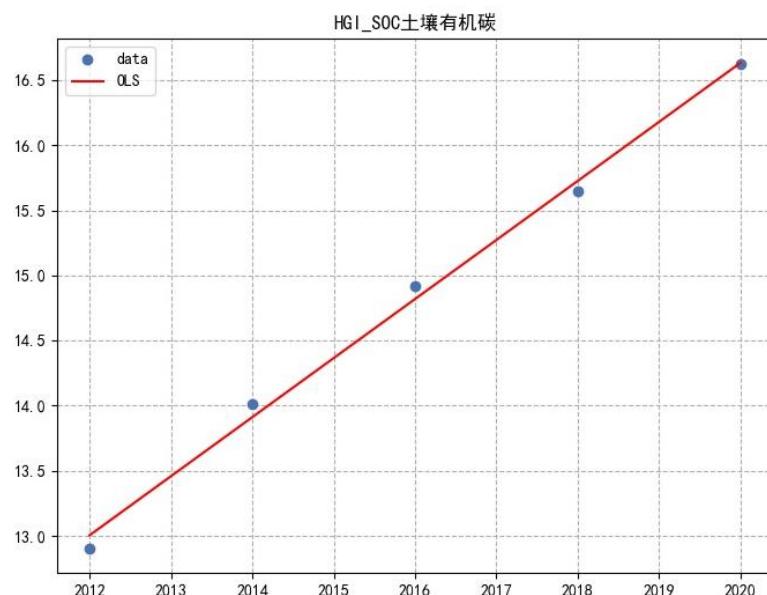


图 73 HGI 下 SOC 土壤有机碳拟合曲线

拟合曲线的参数如下：

回归截距: $w_0 = [-899.8219526]$
 回归系数: $w_1 = [[0. \quad 0.45369083]]$
 R2 确定系数: 0.9956
 均方误差: 0.0073
 平均绝对值误差: 0.0787
 中位绝对值误差: 0.0972

$$SOC(t) = -899.8219526 + 0.45369083 * t \quad (9-9)$$

B. SIC 土壤无机碳

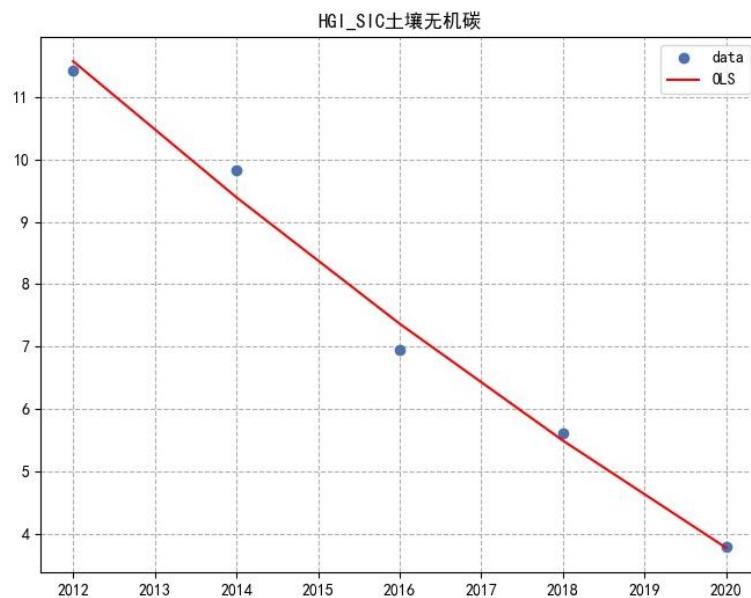


图 74 HGI 下 SIC 土壤无机碳拟合曲线

拟合曲线的参数如下:

回归截距: $w_0 = [80152.80174554]$
 回归系数: $w_1 = [[0.00000000e+00 \quad -7.85336143e+01 \quad 1.92355820e-02]]$
 R2 确定系数: 0.9895
 均方误差: 0.0807
 平均绝对值误差: 0.2269
 中位绝对值误差: 0.1506

$$SIC(t) = 80152.80174554 - 78.5336143 * t + 0.0192355820 * t^2 \quad (9-10)$$

C. 全 N

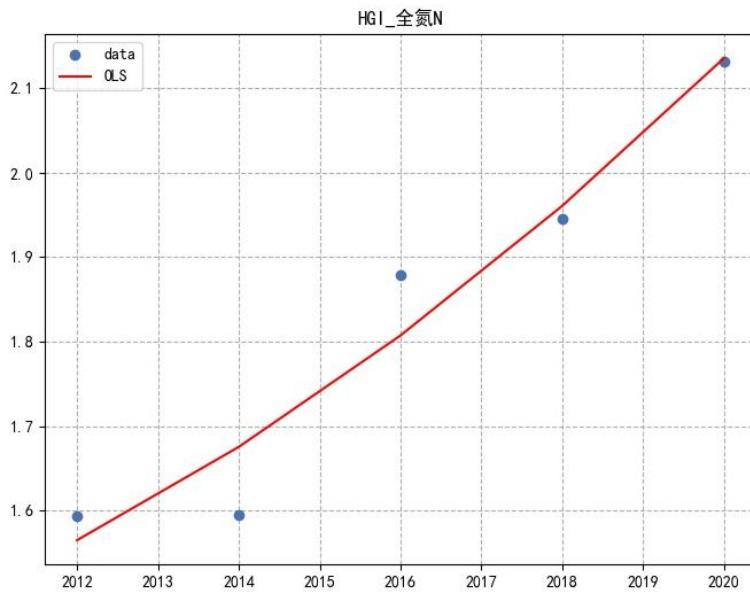


图 75 HGI 下全 N 拟合曲线

拟合曲线的参数如下：

回归截距: $w0=[10781.75890748]$

回归系数: $w1=[[0.0000000e+00 -1.07658143e+01 2.68780563e-03]]$

R2 确定系数: 0.9412

均方误差: 0.0026

平均绝对值误差: 0.0402

中位绝对值误差: 0.0286

$$N(t) = 10781.75890748 - 10.7658143 * t + 0.00268780563 * t^2 \quad (9-11)$$

对于这三种放牧强度情况下，对于土壤全碳和土壤 C/N 比化学性质均有如下关系：

$$STC(t) = SOC(t) + SIC(t) \quad (9-12)$$

$$C/N(t) = \frac{STC(t)}{N(t)} \quad (9-13)$$

9.3.2 土壤化学性质、降水量与土壤湿度回归模型

为求解该模型，我们利用多输出多维多项式回归模型分别建立了土壤化学性质、降水量与 10cm 湿度、40cm 湿度、100cm 湿度及 200cm 湿度的回归模型。

$$\check{Y} = \{\check{Y}_1, \check{Y}_2, \check{Y}_3, \check{Y}_4\} = Polynomial(X_1, X_2, X_3, X_4, X_5, X_6) \quad (9-14)$$

其中， $\check{Y}_1, \check{Y}_2, \check{Y}_3, \check{Y}_4$ 分别代表 10cm 湿度、40cm 湿度、100cm 湿度及 200cm 湿度的预测值， $X_1, X_2, X_3, X_4, X_5, X_6$ 分别代表土壤有机碳、土壤无机碳、土壤全碳、土壤全 N、土壤 C/N 比以及降水量。

9.3.3 土壤湿度与植物覆盖率回归模型

(1) 绿植覆盖率与植被指数拟合模型

在求解该回归模型时，我们首先分析了附件 5 中的 2020-2022 年绿植覆盖率，绿植覆盖率可以录到值的时间聚集在每年 6-11 月，且其与植被指数存在相关关系。我们将每个月的绿植覆盖率取平均值后，对应植被指数的情况如图 X 所示。

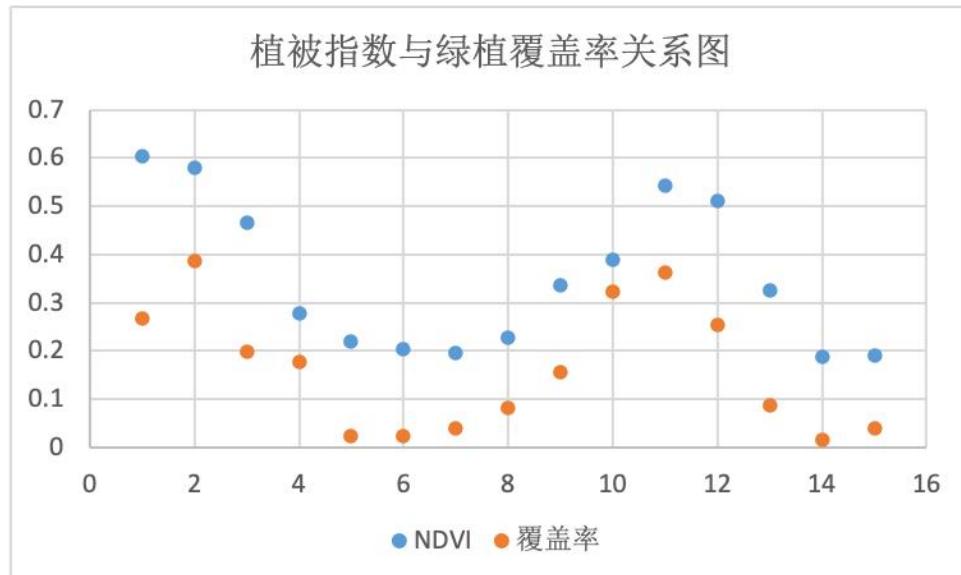


图 76 植被指数与绿植覆盖率关系图

因此，我们利用多项式回归拟合植被指数与绿植覆盖率，通过附件 6 中的植被指数，预测绿植覆盖率，该拟合模型公式如下：

$$\tilde{Y} = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3 \quad (9-15)$$

其中， \tilde{Y} 为绿植覆盖率的预测值， x_1 为植被指数，在该拟合函数中， $a_0 = -0.20512732$,

$$a_1 = 1.23489766, a_2 = 0.0400446, a_3 = -1.05044181。$$

(2) 土壤湿度与绿植覆盖率拟合模型

根据上文得到的绿植覆盖率与植被指数拟合模型，我们将每年 6-11 月的绿植覆盖率平均值作为该年的绿植覆盖率标准值，与 10cm 土壤湿度、40cm 土壤湿度、100cm 土壤湿度和 200cm 土壤湿度利用多维多项式回归建立拟合模型，定义如下：

$$\tilde{Y} = \text{Polynomial}(X_1, X_2, X_3, X_4) \quad (9-16)$$

其中， \tilde{Y} 为绿植覆盖率的预测值， X_1, X_2, X_3, X_4 分别为 10cm 土壤湿度、40cm 土壤湿度、100cm 土壤湿度和 200cm 土壤湿度。

9.4 结果分析

根据 9.3 得到的多个拟合模型，以及问题二中得到的 2023 年降水量预测情况，对每种放牧策略，可以得到如表 X 所示的结果，结果已保留两位小数，其中由于回归模型的精度问题，会导致结果出现负数，负数结果在我们的问题当中没有意义，所以我们将负数值设定为 0。

表 38 问题六结果

放牧强度	SOC	SIC	STC	N	C/N	10cm 湿度	40cm 湿度	100cm 湿度	200cm 湿度	植物 覆盖 率
LGI	17.16	3.07	20.23	2.34	8.64	18.91	65.11	116.06	157.32	0.24
MGI	14.10	2.47	16.57	2.16	7.67	0	9.34	20.71	166.08	0
HGI	17.99	1.48	19.47	2.44	7.98	15.84	73.12	112.44	153.08	0.54

对于土壤的肥力，我们利用化学性质和物理性质的调和平均数控制，其中的化学性质与物理性质也分别通过调和平均数计算，其公式定义如下：

$$Fer = \frac{2*C*P}{C+P} \quad (9-17)$$

$$C = \frac{n \prod_i^n C_i}{\sum_{i=1}^n C_i}, \text{ if } C_i > 0 \quad (9-18)$$

$$P = \frac{m \prod_i^m P_i}{\sum_{i=1}^m P_i}, \text{ if } P_i > 0 \quad (9-19)$$

其中 C 表示化学性质， P 表示物理性质， C_i 表示 SOC、SIC、N 中第 i 个化学性质， P_i 表示 10cm 湿度、40cm 湿度、100cm 湿度、200cm 湿度中的第 i 个物理性质。根据该公式求出的土壤肥力情况及标准化后数据如表 39 所示。

表 39 土壤肥力及标准化后数据

放牧强度	土壤肥力
LGI	32.75
MGI	23.53
HGI	17.76

我们将不同放牧强度下的土壤肥力、植被覆盖率、土壤化学性质、土壤湿度数据标准化后，在图 77 中进行展示。

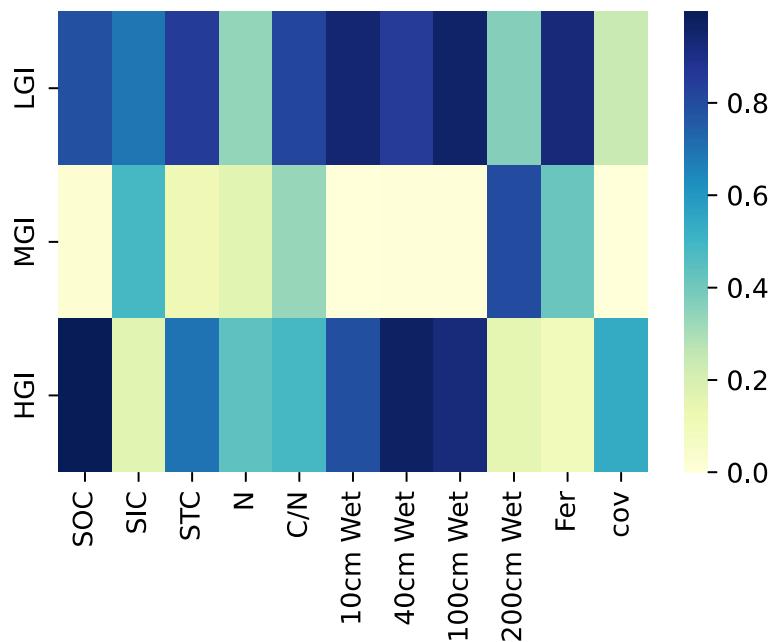


图 77 不同放牧强度下土壤情况变化

9.5 总结

在问题六的求解过程中，我们将问题六分解成为多个子问题，并结合问题 1-4 形成的相关模型进行分析、求解，最终拟合了不同放牧强度下土壤化学性质回归模型、土壤化学性质及降水量与土壤湿度回归模型、植被指数与绿植覆盖率回归模型、土壤湿度与绿植覆盖率回归模型，进而直接或者间接求解到针对牧户 1-4 放牧强度、示范区平均放牧强度以及问题 4 的放牧策略的分类土壤情况变化图。

十、模型评价与改进

10.1 模型优点

(1) 对于问题一，利用时间序列分析与多项式回归分析，建立了土壤蒸发量、降水量、叶面积指数、绿植覆盖率与时间变量的回归模型，并建立数学模型揭示锡林郭勒草原土壤物理性质、植被生物量同放牧策略的机理，使得针对问题一的结论具有可解释性。

(2) 对于问题二，利用多元回归模型与 OLS 最小二乘法进行求解，由于本问题中使用的附件中数据量较少，使用这两种模型可以取得较好的效果。

(3) 对于问题三，对于使用的模型进行了重组的扩展研究，将模型与多种树模型、机器学习方法进行对比，验证了问题求解模型的有效性。

(4) 对于问题四，广泛收集了内蒙古水资源数据和全球土壤信息数据，并整合题目附件中的数据，构建了沙漠化指数组合公式与土壤板结化定量定义，在大量数据来源下，两个公式的定义具有实际意义。

(5) 对于问题五，使用融合回归模型的基于分解的多目标进化算法，在草原生态健康发展指标和经济利益指标中寻求平衡，拥有较低的计算复杂度。

(6) 对于问题六，对所求解进行关系分析，分解为多个子问题，并利用多个间接模型进行求解，降低了建模复杂度。

10.2 模型缺点

(1) 问题一主要依据附件中给出的数据以及理想状态下的基本方程进行计算，附件中的数据量较少，且实际状态同理想方程会存在偏差，仅依靠数学模型推断可能会对结果造成影响。

(2) 对于问题二和问题三使用的多元回归模型和 OSL 最小二乘法，其对因子和因子表达式是基于推测，对某些因子具有不可测性，在某些情况下，会受到限制。

(3) 问题四中利用粒子群算法寻找使沙漠化指数和土壤板结化程度最低的放牧策略，该算法虽然搜索速度快、参数较少，但是容易陷入局部最优，且不能有效解决离散和组合优化问题。

(4) 利用基于分解的多目标进化算法，可以寻找到质量较高的解集，但是其计算成本非常高，呈线性增长。

(5) 问题六中通过多个间接模型来求解最终结果，虽然问题分解可以将复杂问题简单化，但也会造成错误累计。

10.3 模型改进与推广

(1) 对于求解问题使用的多个回归模型，我们好需要尝试从各种渠道获取高质量的数据集，来验证实际状况下的模型有效性。

(2) 对于进行机理分析时，建立的数学模型，应从多种情况考虑，对理想状态下的数学模型优化，以满足实际情况。

(3) 针对于沙漠化指数因子、气候情况因子，我们利用因子分析、相关系数分析评估了因子之间的相关关系，但并未进一步深究，我们可以利用特征降维的方式，进一步分析。

(4) 针对问题五的求解，与问题四定义的沙漠化指数和土壤板结化指数息息相关，由于部分因子我们未获取到相关关系，且维度太高会增大模型优化难度，在未来我们可以尝试设计新的目标函数、优化算法来进行提升。

(5) 在本文中，融合多个高质量的外部数据集，给出了沙漠化指数、土壤板结化程度的定义及公式，这对于草原的可持续发展具有一定的指导意义。

核心代码附录

程序 1	9种回归模型
<pre>import pandas as pd caoyuan=pd.read_excel('D:\\DataSet\\19届研究生数模竞赛\\2022年E题数据\\数据集\\监测点数据\\附件14：内蒙古自治区锡林郭勒盟典型草原不同放牧强度土壤碳氮监测数据集（2012年8月15日-2020年8月15日）\\内蒙古自治区锡林郭勒盟典型草原不同放牧强度土壤碳氮监测数据集（2012年8月15日-2020年8月15日）.xlsx') caoyuan=caoyuan[['intensity','SOC土壤有机碳','SIC土壤无机碳','STC土壤全碳','全氮N','土壤C/N比']] caoyuan.head() from sklearn.preprocessing import OneHotEncoder caoyuanH=pd.get_dummies(caoyuan, prefix=['inten','year','plot'],columns=['intensity','year','plot']) import sklearn #机器学习算法模型 from sklearn.linear_model import LogisticRegression from sklearn.tree import DecisionTreeRegressor #准确率，精确率，召回率，f1 from sklearn.metrics import accuracy_score,precision_score,recall_score,f1_score,classification_report import xgboost as xgb import joblib from xgboost import XGBRegressor from sklearn.model_selection import train_test_split X = caoyuanH.drop(['SOC','SIC','STC','N','CN'],axis=1) X # X=caoyuanH[['SOC土壤有机碳','SIC土壤无机碳','STC土壤全碳','全氮N','土壤C/N比']] Y=caoyuanH['SOC'] test_size = 0.2 X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=test_size, random_state=123) from sklearn.metrics import mean_squared_error,mean_absolute_error,r2_score modelDecision= DecisionTreeRegressor() modelDecision.fit(X_train, y_train) # 对测试集做预测 y_pred = modelDecision.predict(X_test)</pre>	

```

predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelMLP= MLPRegressor()
modelMLP.fit(X_train, y_train)
# 对测试集做预测
y_pred =modelMLP.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

from sklearn import linear_model
modelLinear=linear_model.LinearRegression()
modelLinear.fit(X_train, y_train)
# 对测试集做预测
y_pred = modelLinear.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelRandom= RandomForestRegressor()
modelRandom.fit(X_train, y_train)
# 对测试集做预测
y_pred = modelRandom.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelADA= AdaBoostRegressor()
modelADA.fit(X_train, y_train)
# 对测试集做预测
y_pred =modelADA.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)

```

```
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelGradient= GradientBoostingRegressor()
modelGradient.fit(X_train, y_train)
# 对测试集做预测
y_pred =modelGradient.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelXGB= XGBRegressor()
modelXGB.fit(X_train, y_train)
# 对测试集做预测
y_pred =modelXGB.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelCat= CatBoostRegressor()
modelCat.fit(X_train, y_train)
# 对测试集做预测
y_pred =modelCat.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)

modelLGB= lgb.LGBMRegressor()
modelLGB.fit(X_train, y_train)
# 对测试集做预测
y_pred =modelLGB.predict(X_test)
predictions = [round(value) for value in y_pred]
# 评估预测结果
MAE=mean_absolute_error(y_pred, y_test)
MSE=mean_squared_error(y_pred, y_test)
print(MAE,MSE)
```

程序 2

多项式回归代码

```
# -*- coding: utf-8 -*-

import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error,
median_absolute_error
from sklearn.preprocessing import PolynomialFeatures

plt.rcParams['font.family'] = ['sans-serif']
plt.rcParams['font.sans-serif'] = ['SimHei']

import pandas as pd

df = pd.read_excel(r'Data/f8_weather_3.xls', sheet_name='10月')

# 生成测试数据:
nSample = 100
x = df['月份'].values      # = np.linspace(0, 10, nSample) # 起点为 0, 终点为 10, 均分为
nSample个点
e = np.random.normal(size=len(x)) # 正态分布随机数
y = df['降水量'].values    #= 2.36 + 1.58 * x + e # y = b0 + b1*x1

# 按照模型要求进行数据转换: 输入是 array类型的 n*m 矩阵, 输出是 array类型的 n*1
数组
x = x.reshape(-1, 1) # 输入转换为 n行 1列 (多元回归则为多列) 的二维数组
y = y.reshape(-1, 1) # 输出转换为 n行1列的二维数组
# print(x.shape,y.shape)

# 定义多项式回归, degree的值可以调节多项式的特征
poly_reg = PolynomialFeatures(degree=2)
# 特征处理
x_poly = poly_reg.fit_transform(x)

# 定义回归模型
reg = LinearRegression()
reg.fit(x_poly, y)

#plt.plot(x, y, 'b.')
#plt.plot(x, reg.predict(x_poly), 'r')
```

```

# plt.show()

plt.style.use("seaborn-deep")
# 绘图: 原始数据点, 拟合曲线
fig, ax = plt.subplots(figsize=(8, 6))
ax.plot(x, y, 'o', label="data") # 原始数据
ax.plot(x, reg.predict(x_poly), 'r-', label="OLS") # 拟合数据

ax.legend(loc='best') # 显示图例
ax.legend(loc='best') # 显示图例
plt.title('降水量')
plt.grid(ls='--')
plt.savefig('降水量_duo10.jpg')
plt.show()

# 输出回归结果 XUPT
print('回归截距: w0={}'.format(reg.intercept_)) # w0: 截距
print('回归系数: w1={}'.format(reg.coef_)) # w1,..wm: 回归系数

# 回归模型的评价指标 YouCans
print('R2 确定系数: {:.4f}'.format(reg.score(x_poly, y))) # R2 判定系数
print('均方误差: {:.4f}'.format(mean_squared_error(y, reg.predict(x_poly)))) # MSE 均方误差
print('平均绝对值误差: {:.4f}'.format(mean_absolute_error(y, reg.predict(x_poly)))) # MAE 平均绝对误差
print('中位绝对值误差: {:.4f}'.format(median_absolute_error(y, reg.predict(x_poly)))) # 中值绝对误差

```

程序 3

ARIMA代码

```
import sys
import os
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller
from statsmodels.stats.diagnostic import acorr_ljungbox
from statsmodels.graphics.api import qqplot
import matplotlib.pyplot as plt
from matplotlib.pyplot import style
style.use('ggplot')
from arch.unitroot import ADF
import warnings
warnings.filterwarnings("ignore")
pd.set_option('display.float_format', lambda x: '%.5f' % x)
np.set_printoptions(precision=5, suppress=True)
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_rows', 100)

plt.rcParams['font.family'] = ['sans-serif']
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus']=False    # 解决负号问题

plt.rcParams['font.family'] = ['sans-serif']
plt.rcParams['font.sans-serif'] = ['SimHei']

data = pd.read_excel("Data/f8_weather.xls", sheet_name='Sheet2', index_col="日期",
parse_dates=True)
print(data.head())

data["diff1"] = data["降水量"].diff(1).dropna()
data["diff2"] = data["diff1"].diff(1).dropna()
data1 = data.loc[:,["降水量","diff1","diff2"]]

data1.plot(subplots=True, figsize=(18, 12), title="差分图", fontsize=15)
plt.savefig("降水量_arrm1.jpg", dpi=300)
plt.show()

print("一阶差分单位根检验:\n")
```

```

print(ADF(data.diff1.dropna()))

print("二阶差分单位根检验:\n")
print(ADF(data.diff2.dropna()))

plot_acf(data["降水量"])
plt.savefig('降水量.jpg')
plt.show() #自相关图

print(u'差分序列的白噪声检验结果为: \n')
print(acorr_ljungbox(data["降水量"], lags=1)) #返回统计量和p值
plot_pacf(data["降水量"])
plt.savefig('降水量_偏自相关.jpg')
plt.show()

from statsmodels.stats.diagnostic import acorr_ljungbox
acorr_ljungbox(data.diff1.dropna(), lags = [i for i in range(1,12)],boxpierce=True)

arma_mod20 = sm.tsa.ARIMA(data["降水量"], order=(1,1,0)).fit()
arma_mod30 = sm.tsa.ARIMA(data["降水量"], order=(0,1,1)).fit()
arma_mod40 = sm.tsa.ARIMA(data["降水量"], order=(1,1,1)).fit()
values = [[arma_mod20.aic,arma_mod20.bic,arma_mod20.hqic],[arma_mod30.aic,arma_mod30.bic,arma_mod30.hqic],[arma_mod40.aic,arma_mod40.bic,arma_mod40.hqic]]
df2 = pd.DataFrame(values,index=["AR(1,1,0)","MA(0,1,1)","ARMA(1,1,1)"],columns=["AIC","BIC","hqic"])
print(df2)

import statsmodels.api as sm
model = sm.tsa.ARIMA(data["降水量"], order=(0, 2, 2))
result = model.fit()
print(result.summary())
print('222222', result.params)

resid = result.resid # 残差
fig = plt.figure(figsize=(12,8))
ax = fig.add_subplot(111)
fig = qqplot(resid, line='q', ax=ax, fit=True) # qq 分位图
plt.savefig("降水量_arrm2.jpg", dpi=300)
plt.show()

```

```

pred = result.predict(start='2022-07-01', end='2022-08-01', dynamic=True, typ='levels')
print(pred)

plt.figure(figsize=(20, 12))
plt.xticks(rotation=45)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.plot(pred)
print("11", data)
plt.plot(data.loc[:,["降水量"]])
plt.savefig("降水量_arrm3.jpg", dpi=300)
plt.show()

# BIC评估图
import itertools
#当多组值都不符合时，遍历多组值，得出最好的值
p_min = 0
d_min = 0
q_min = 0
p_max = 2
d_max = 1
q_max = 10

# Initialize a DataFrame to store the results
results_bic = pd.DataFrame(index=['AR {}'.format(i) for i in range(p_min,p_max+1)],
                           columns=['MA {}'.format(i) for i in range(q_min,q_max+1)])

for p,d,q in itertools.product(range(p_min, p_max+1),
                                range(d_min, d_max+1),
                                range(q_min, q_max+1)):
    if p==0 and d==0 and q==0:
        results_bic.loc['AR {}'.format(p), 'MA {}'.format(q)] = np.nan
        continue

    try:
        model = sm.tsa.SARIMAX(data["降水量"], order=(p, d, q),
                               enforce_stationarity=False,
                               enforce_invertibility=False,
                               )
        results = model.fit()
        results_bic.loc['AR {}'.format(p), 'MA {}'.format(q)] = results.bic
    except:

```

```
    continue
results_bic = results_bic[results_bic.columns].astype(float)
fig, ax = plt.subplots(figsize=(20, 20))
import seaborn as sns
ax = sns.heatmap(results_bic,
                  mask=results_bic.isnull(),
                  ax=ax,
                  annot=True,
                  fmt='%.2f',
                  )
ax.set_title('BIC', fontsize=14)
sns.set(context='notebook', font_scale=2)
hist_fig = ax.get_figure()
hist_fig.savefig('降水量_BIC.jpg', dpi = 400)
plt.show()
```

程序 4

气候条件相关性分析

```
import pandas as pd

import os
#待搜索的目录路径
path = "./Q5/data/qihou"
#定义保存结果的数组

def findfiles(path):
    result = []
    file_list = os.listdir(path)
    for file in file_list:
        cur_path = os.path.join(path, file)
        if os.path.isdir(cur_path):
            findfiles(cur_path)
        else:
            result.append(cur_path)
    return result

result = findfiles(path)
print(result)
df = pd.DataFrame()
for i in range(0,len(result)):
    if i==0:
        df = pd.read_excel(result[i])
        #df.drop(columns=["站点号","海拔高度(m)","经度","纬度","年份","月份"],inplace=True)
    else:
        df_t = pd.read_excel(result[i])
        #df_t.drop(columns=["站点号","海拔高度(m)","经度","纬度","年份","月份"],inplace=True)
        df = df.append(df_t,ignore_index=True)
print(df[df.index.duplicated()])
import math
for i in range(len(df)):
    clums_list = df.columns.to_list()
    for j in range(len(clums_list)):
        df_i_j = df.iloc[i][clums_list[j]]
        if math.isnan(df_i_j):
            df.loc[i,clums_list[j]] = 0
df.to_csv('./Q5/data/qihou.csv')
```

```
df.drop(columns=["站点号","海拔高度(m)","经度","纬度","年份","平均气温≥35℃的天数  
","平均海平面气压(hPa)","平均海平面气压(hPa)","平均站点气压(hPa)"],inplace = True)  
correlations = df.corr()  
print(df)  
print(correlations)  
  
import seaborn as sns  
import matplotlib.pyplot as plt  
import matplotlib  
print(matplotlib.matplotlib_fname())  
plt.rcParams['font.family']='SimHei'  
plt.rcParams['axes.unicode_minus']= False  
rc = {'font.sans-serif': 'SimHei',  
       'axes.unicode_minus': False}  
sns.set(font="SimHei")  
sns.set(context='notebook', style='ticks', rc=rc)  
ax4 = sns.heatmap(correlations, cmap="YlGnBu")  
plt.gcf().subplots_adjust(left=0.3,bottom=0.38)  
plt.show()  
#print(matplotlib.get_cachedir())
```

程序 5

植物多样性回归模型

```
import pandas as pd
import numpy as np
def get_duoyang(yangqun,water):
    df = pd.read_excel('./Q5/data/duoyangxing.xlsx',sheet_name='Sheet2')
    print(df)
    x=np.arange(2*len(df),dtype=float).reshape(len(df),2)
    for i in range(len(df)):
        df.loc[i,'qiangdu']=df.loc[i,'qiangdu']/8.0
        df.loc[i,'jiangshui']=(df.loc[i,'jiangshui']-300.0)/1500.0
        df.loc[i,'duoyang']=df.loc[i,'duoyang']/10.0
        x[i][0]=df.loc[i,'qiangdu']
        x[i][1]=df.loc[i,'jiangshui']
    print(df)
    print(x)

from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
reg = PolynomialFeatures(degree=3)
x=reg.fit_transform(x)
print(x)

f = LinearRegression()
f.fit(X=x,y=df['duoyang'])
y = f.predict(x)
print(y)

x=np.array([
[0.0,0.0],[0.0,0.2],[0.0,0.4],[0.0,0.6],
[0.25,0.0],[0.25,0.2],[0.25,0.4],[0.25,0.6],
[0.5,0.0],[0.5,0.2],[0.5,0.4],[0.5,0.6],
[1.0,0.0],[1.0,0.2],[1.0,0.4],[1.0,0.6]
])
x=reg.fit_transform(x)
print(f)
y = f.predict(x)
print(y)
xx=np.array([[yangqun/4.0,(water-300.0)/1500.0],[yangqun/4.0,(water-300.0)/1500.0]])
xx = reg.fit_transform(xx)
xy = f.predict(xx)
print(1)
print(xy)
return xy[0]
```

程序 6

融合回归模型的基于分解的多目标进化算法

```
water = 1200
from operator import le
import geatpy as ea
import numpy as np
import pandas as pd
def get_turang(x,water):
    y = 0.179107*10-0.223754*water/300+0.597139*(x-2.3)**2
    yy = np.zeros(shape=(len(y),1),dtype=float)
    for i in range(len(y)):
        yy[i][0] = y[i][0]
    return yy
def get_shamo(X1,X2,X3,X4):
    y=-(-0.1235*X4-0.093844*X3+0.071563*X2+0.109575*X1*2+0.137632*water/60)
    yy = np.zeros(shape=(len(y),1),dtype=float)
    for i in range(len(y)):
        yy[i][0] = y[i][0]
    return yy

def get_jingji(x1):
    y = -0.02161358*(x1**3)+0.27854107*(x1**2) -1.10540993*(x1**1)+1.474103
    yy = np.zeros(shape=(len(y),1),dtype=float)
    for i in range(len(y)):
        yy[i][0] = y[i][0]**4
    return yy
def get_duoyang(yangqun,water):
    df = pd.read_excel('./Q5/data/duoyangxing.xlsx',sheet_name='Sheet2')
    x=np.arange(2*len(df),dtype=float).reshape(len(df),2)
    for i in range(len(df)):
        df.loc[i,'qiangdu']=df.loc[i,'qiangdu']/8.0
        df.loc[i,'jiangshui']=(df.loc[i,'jiangshui']-300.0)/1500.0
        df.loc[i,'duoyang']=df.loc[i,'duoyang']/10.0
        x[i][0]=df.loc[i,'qiangdu']
        x[i][1]=df.loc[i,'jiangshui']
    from sklearn.preprocessing import PolynomialFeatures
    from sklearn.linear_model import LinearRegression
    reg = PolynomialFeatures(degree=3)
    x=reg.fit_transform(x)
    f = LinearRegression()
    f.fit(X=x,y=df['duoyang'])

    # x=np.array([
```

```

# [0.0,0.0],[0.0,0.2],[0.0,0.4],[0.0,0.6],
# [0.25,0.0],[0.25,0.2],[0.25,0.4],[0.25,0.6],
# [0.5,0.0],[0.5,0.2],[0.5,0.4],[0.5,0.6],
# [1.0,0.0],[1.0,0.2],[1.0,0.4],[1.0,0.6]
# ])
# x=reg.fit_transform(x)
xx = np.zeros(shape=(len(yangqun),2))
for i in range(len(yangqun)):
    xx[i][0] = yangqun[i][0]/4.0
    xx[i][1] = (water-300.0)/1500.0

##xx=np.array([[yangqun/4.0,(water-300.0)/1500.0],[yangqun/4.0,(water-300.0)/1500.0]],dtype=object)

xx = reg.fit_transform(xx)
xy = f.predict(xx)
yy = np.zeros(shape=(len(xy),1),dtype=float)
for i in range(len(xy)):
    yy[i][0] = xy[i]
return yy

class MyProblem(ea.Problem):
    def __init__(self, M=4):
        name = 'MyProblem'
        Dim = 4
        maxormins = [1,0,0,1]
        varTypes = [0,0,0,0]
        lb = [0,10,20,107]
        ub = [4,25,37,111]
        lbin = [1]*Dim
        ubin = [1]*Dim

        ea.Problem.__init__(self,
                            name,
                            M,
                            maxormins,
                            Dim,
                            varTypes,
                            lb,
                            ub,
                            lbin,
                            ubin)

    def evalVars(self, Vars):

```

```

x1 = Vars[:,[0]]
x2 = Vars[:,[1]]
x3 = Vars[:,[2]]
x4 = Vars[:,[3]]
f1 = get_duoyang(x1,water)*10/5
f2 = get_shamo(x1,x2,x3,x4)/3/5
f3 = get_turang(x1,water)*2/5
f4 = get_jingji(x1)*10/5

CV = np.hstack([
    0.3-get_duoyang(x1,water),
])

objV = np.hstack([f1,f2,f3,f4])
return objV,CV

problem = MyProblem()

algorithm = ea.moea_MOEAD_templet(problem,
    ea.Population(Encoding='RI', NIND=100),
    MAXGEN=1000, # 最大进化代数。
    logTras=0)
res = ea.optimize(algorithm, verbose=False, drawing=1, outputMsg=True, drawLog=False,
saveFlag=False, dirName='result')

print(res)

```

```
import pandas as pd
df = pd.read_excel('./Q5/data/q6.xlsx',sheet_name='r1')
for i in range(len(df)):
    df.loc[i,'SOC']=(df.loc[i,'SOC']-14)/4
    df.loc[i,'SIC']=(df.loc[i,'SIC']-1)/3
    df.loc[i,'STC']=(df.loc[i,'STC']-16)/5
    df.loc[i,'N']=(df.loc[i,'N']-2)/1
    df.loc[i,'C/N']=(df.loc[i,'C/N']-7)/2
    df.loc[i,'10cm Wet']=(df.loc[i,'10cm Wet']-0)/20
    df.loc[i,'40cm Wet']=(df.loc[i,'40cm Wet']-9)/66
    df.loc[i,'100cm Wet']=(df.loc[i,'100cm Wet']-20)/100
    df.loc[i,'200cm Wet']=(df.loc[i,'200cm Wet']-150)/20
    df.loc[i,'Fer']=(df.loc[i,'Fer']-16)/18
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib
# print(matplotlib.matplotlib_fname())
# plt.rcParams['font.family']='SimHei'
# plt.rcParams['axes.unicode_minus']=False
# rc = {'font.sans-serif': 'SimHei',
#        'axes.unicode_minus': False}
# sns.set(font="SimHei")
# sns.set(context='notebook', style='ticks', rc=rc)
ax4 = sns.heatmap(df, cmap="YlGnBu")
plt.gcf().subplots_adjust(left=0.3,bottom=0.38)
import numpy as np
plt.yticks(np.arange(0.5, 3.5, step=1),list(['LGI','MGI','HGI']))
plt.show()
print(matplotlib.get_cachedir())
```

参考文献

- [1] 阎庆文, 刘寿东, 杨霞. 内蒙古典型草原生态系统服务功能价值评估研究[J]. 草地学报, 2004, 12(3): 165-169.
- [2] 王涛, 朱震达. 中国沙漠化研究[J]. 中国生态农业学报(2):7-12.
- [3] 李金花, 李镇清, 任继周. 放牧对草原植物的影响[J]. 草业学报, 2002, 11(1): 4-11.
- [4] Siami-Namini S, Tavakoli N, Namin A S. A comparison of ARIMA and LSTM in forecasting time series[C]//2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2018: 1394-1401.
- [5] 多项式曲线拟合和最小二乘法.
[https://sikasjc.github.io/2018/10/24/curvefitting/.\(2022.10.09\)](https://sikasjc.github.io/2018/10/24/curvefitting/.(2022.10.09))
- [6] Simon J. R. Woodward, Graeme C. Wake, et al, A Simple Model for Optimizing Rotational Grazing, Agricultural Systems ,41 (1993) 123-155.
- [7] 王悦骅. 模拟降水对不同载畜率放牧荒漠草原植物多样性的影响[D]. 内蒙古农业大学, 2019.
- [8] Woodward S J R. Wake G C. McCall D G. Optimal grazing of a multi—paddock system using a discrete time model[j]. Agri—cultural Systems. 1995, 48: 119—139.
- [9] 牵现文, 李春玉, Miyong Kim, 等. 决策树与 Logistic 回归在高血压患者健康素养预测中的应用[J]. 护士进修杂志, 2012, 27(13):3.
- [10] 张毅, 李强, 王政, 等. 基于分类-回归决策树烤烟化学指标识别模型的建立与验证[J]. 中国烟草学报, 2014, 20(6):28-33.
- [11] Xu M, Watanachaturaporn P, Varshney P K, et al. Decision tree regression for soft classification of remote sensing data[J]. Remote Sensing of Environment, 2005, 97(3): 322-336.
- [12] Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting[J]. R package version 0.4-2, 2015, 1(4): 1-4.
- [13] Hancock J T, Khoshgoftaar T M. CatBoost for big data: an interdisciplinary review[J]. Journal of big data, 2020, 7(1): 1-45.
- [14] Schapire R E. Explaining adaboost[M]//Empirical inference. Springer, Berlin, Heidelberg, 2013: 37-52.
- [15] Biau G, Scornet E. A random forest guided tour[J]. Test, 2016, 25(2): 197-227.
- [16] Bi Y, Xiang D, Ge Z, et al. An interpretable prediction model for identifying N7-methylguanosine sites based on XGBoost and SHAP[J]. Molecular Therapy-Nucleic Acids, 2020, 22: 362-372.
- [17] 赵熠. 不同放牧强度对短花针茅荒漠草原土壤线虫群落组成及多样性的影响[D]. 内蒙古农业大学, 2021. DOI:10.27229/d.cnki.gnmnu.2021.001065.
- [18] 徐新影, 张卫青, 李金霞, 赵利萌, 乌兰图雅, 万志强. 克氏针茅(*Stipa krylovii*)草原土壤与优势植物碳氮含量对放牧强度的响应[J]. 草地学报, 2022, 30(09):2255-2263.
- [19] 严军, 王婷, 秦珏. 基于变异系数法的马鞍山江心洲生态敏感性定量研究[J]. 生态科学, 2020, 39(02):124-132.
- [20] 李祎嘉, 吕玉花. 基于因子分析法的河南省农产品物流发展水平研究[J]. 中国商论, 2022(15):4-6.
- [21] 刘敦利. 基于栅格尺度的土地沙漠化预警模式研究[D]. 乌鲁木齐: 新疆大学, 2010.
- [22] 胡光印, 董治宝, 逯军峰, 杨林海, 南维鸽, 肖锋军. 黄河流域沙漠化空间格局与成因[J]. 中国沙漠, 2021, 41(04):213-224.
- [23] 内蒙古自治区水利厅. 内蒙古自治区水资源公报.
[http://slt.nmg.gov.cn/xxgk/bmxxgk/gbxx/szygb/.\(2022.10.08\)](http://slt.nmg.gov.cn/xxgk/bmxxgk/gbxx/szygb/.(2022.10.08))

- [24]程建华,肖军琪,刘菲菲.基于熵值法的地方政府优化营商环境途径探索——以安徽省为例[J].河北北方学院学报(社会科学版),2022,38(04):53-58+68.
- [25] Batjes N H, Ribeiro E, Van Oostrum A. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019)[J]. Earth System Science Data, 2020, 12(1): 299-320.
- [26] Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., ... & Gonzalez, M. R. (2014). SoilGrids1km—global soil information based on automated mapping. PloS one, 9(8), e105992.
- [27] Ghazali M F, Wikantika K, Harto A B, et al. Generating soil salinity, soil moisture, soil pH from satellite imagery and its analysis[J]. Information Processing in Agriculture, 2020, 7(2): 294-306.
- [28] Guo Z, Chai M, Zhan Z, et al. Spatiotemporal variation of soil PH in the past 30 years of Guangdong Province, China[C]//2011 19th International Conference on Geoinformatics. IEEE, 2011: 1-5.
- [29] Misra A, Tyler G. Influence of soil moisture on soil solution chemistry and concentrations of minerals in the calcicoles Phleum phleoides and Veronica spicata grown on a limestone soil[J]. Annals of Botany, 1999, 84(3): 401-410.
- [30] Alkeshuosh A H, Moghadam M Z, Al Mansoori I, et al. Using PSO algorithm for producing best rules in diagnosis of heart disease[C]//2017 international conference on computer and applications (ICCA). IEEE, 2017: 306-311.
- [31] 侯薇, 董红斌, 印桂生. 一种改进的基于分解的多目标进化算法[J]. 计算机科学, 2014, 41(2): 114-118.