



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 上海海事大学

参赛队号 22102540237

1. 米泽欣

队员姓名 2. 韦静

3. 王蕾

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 锡林郭勒草原放牧策略研究

摘 要：

本文主要针对锡林郭勒草原数据进行量化分析，通过机理分析，建立放牧方式与放牧强度对草原土壤物理性质及化学性质的影响模型，进而用锡林郭勒草原监测数据对未来时间段内的土壤湿度及有机化合物含量预测，探索放牧策略与沙漠化程度之间的相互关系，构建相应数学模型（问题 4、5、6），为保证生态环境可持续发展，提出了放牧方式与放牧强度的建议。文章综合采用了微分方程、ARIMA、目标规划、熵权法、遗传算法、线性回归、三次指数平滑等经典方法来研究锡林郭勒草原放牧方案的量化分析及预测相关数据的问题。

针对问题一，通过机理分析放牧强度与放牧方式对土壤物理性质及植被生物量的相互影响关系，将整个放牧过程作为一个连续且循环的过程，利用微分思想，结合文献信息推导出放牧方式与土壤物理性质的数学模型及放牧方式与植被生物量的模型，运用所给数据对模型进行验证，结果显示，放牧方式的改变与土壤物理性质及植被生物量变化趋势一致。

针对问题二，首先通过时序图分析了不同深度(10cm、40cm、100cm、200cm)土壤湿度 2012 年-2022 年 3 月的周期趋势，发现 10cm 土壤湿度存在明显的周期性，而 40cm、100cm、200cm 的周期波动不明显，但呈现出一定的上升或下降趋势。因此，本文决定对 10cm 土壤湿度采用需要协变量的统计模型预测其 2022 年-2023 年的数据，而对 40cm、100cm、200cm 的土壤湿度本文采用不需要协变量的三次指数平滑进行预测。对于 10cm 土壤湿度所需协变量的选择问题而言，本文通过线性拟合的变量筛选方法选择了 8 个高贡献度的土壤增发、降水等变量。由于所有协变量只提供了 2012 年-2022 年 3 月的相关数据，因此，通过对这 8 个协变量进行检验与处理后，采用适当的 ARIMA 模型预测其 2022 年 3 月-2023 年的数据，然后再通过预测得到的协变量数据进一步通过 6 种机器学习模型的交叉分析对比，选择了线性回归方法来预测 10cm 深度的土壤湿度在 2022 年 3 月-2023 年的数据。最终预测得到了不同深度(10cm、40cm、100cm、200cm)土壤湿度 2022 年-2023 年数据。

针对问题三，为取得不同放牧策略对锡林郭勒草原土壤化学性质影响的数学模型，本文在问题一的基础上分析了土壤化学成分的循环过程，通过推导得出放牧策略对土壤化学性质的影响模型，并在此基础上结合数据集和问题二的预测方法进行 2022 年草原土壤各个化学性质的分类预测，得出题目要求项的结果。

针对问题四，首先，分析影响沙漠化程度的因素，通过对收集的相关数据进行数据处理，进而计算相关因子的强度，然后分析因子之间的相关性，用层次分析法计算影响沙漠化程度的因素权重，再根据土壤板结化的影响因素的数据，对缺失的数据进行拉格朗格插值处理，用熵权法计算湿度、容重、有机含物因素的权重，结合参考资料，得出沙漠化程度指数预测模型及土壤板结化模型，以沙漠化程度指数最小以及土壤板结化最小为目标函数，问题三中放牧策略为约束条件，建立目标函数，采用遗传算法寻求最优解，结果显示，在划区轮牧且中度放牧强度情况下，土壤沙漠化程度指数和土壤板结化程度最小。

针对问题五，为在给定的降水量（300mm，600mm、900 mm 和 1200mm）和保持草原可持续发展情况下对实验草场内放牧羊的数量进行求解并找到最大阈值，将表 14 种的化学因素和表 15 的质量因素通过数据挖掘方法对数据处理，便于后续分析。对附件 14 中的数据处理建立线性回归模型量化放牧强度与土壤全氮之间的关系；对附件 15 中数据处理建立线性回归模型量化放牧强度与植被量之间的关系；对附件 5、7 的数据使用线性回归预测出 2016 年-2020 年的植被覆盖率。首先，以羊的总价值为经济指标为目标函数；其次，约束条件为土壤含氮量不能低于临界值，植物生长量不能低于植被需求量，并通过附件将降水量与植被量之间建立联系，并根据参考文献与附件数据得到放牧强度与植被截流量之间的关系；最后，建立非线性规划模型并使用 python 求解。

针对问题六，采用与问题二、三相同的预测流程，基础的初始数据不同，从附件中提取 2012 年到 2020 年间所有土壤有机物含量，土壤湿度的数据。将数据输入问题二中运用的 LSTM-ARIMA 模型中，得出 2012 后的一系列预测值，将预测值与提供的实际观察值通过计算方均误差进行模型性能评价，通过性能评价后，进而预测出 2022 年 9 月示范区土地状态，最后以图示的形式展现出 2012 年到 2022 年间土地状态的变化。

关键字：锡林郭勒草原；ARIMA；三次指数平滑；目标规划；熵权法；遗传算法

目录

1	问题重述.....	5
1.1	问题背景.....	5
1.2	本文拟解决的问题.....	5
2	模型假设.....	6
3	符号说明.....	6
4	问题一：放牧策略土壤物理性质和植被生物量影响的数学模型.....	7
4.1	问题分析.....	7
4.2	模型建立.....	8
4.3	模型检验.....	12
4.4	模型评价.....	13
5	问题二：不同深度土壤湿度预测.....	14
5.1	问题分析.....	14
5.2	模型建立和求解.....	14
5.3	模型构建与预测.....	15
5.4	结果分析.....	26
6	问题三：不同放牧策略土壤化学性质影响的数学模型.....	28
6.1	问题分析.....	28
6.2	模型建立与求解.....	29
6.2	土壤化学性质分析.....	29
6.3	数据预测.....	31
7	问题四：沙漠化程度指数与板结化程度.....	32
7.1	问题分析.....	32
7.2	模型建立与求解.....	33
7.3	结果分析.....	39
8	问题五：不同深度土壤湿度预测.....	40
8.1	问题分析.....	40
8.2	模型建立和求解.....	41
8.3	结果分析.....	47
9	问题六： 2023 年土地状况分析与预测.....	47

9.1	问题描述与分析.....	47
9.2	模型建立与求解.....	48
9.3	模型结果.....	48
10	模型的评价.....	50
5.1	模型的优点.....	50
5.2	模型的缺点.....	50
	参考文献.....	51
	附录.....	51
	附录 1	51
	附录 2	52
	附录 3	54
	附录 4	56
	附录 5	58

1 问题重述

1.1 问题背景

草原作为世界上分布最广的重要的陆地植被类型之一，具有维护生物多样性、涵养水土、净化空气、固碳、调节水土流失和沙尘暴等方面具有重要的生态功能。内蒙古锡林郭勒草原不仅是国家重要的畜牧业生产基地，同时也是重要的绿色生态屏障，在减少沙尘暴和恶劣天气的发生方面发挥着作用。

家畜放牧是天然草地最主要的利用方式，研究表明，放牧对草地生态有两方面影响，一方面，适度放牧对于草地生态系统具有一定程度的积极影响[1]，其原因是牲畜在放牧过程中的采食、践踏和排泄行为能够促进群落演替、维持群落结构、保护物种多样性和土壤养分状况。另一方面为负面效应，在荒漠草原地区，过度放牧对造成植被退化，土壤裸露并且肥力下降，从而加剧水土流失产生[2]，进而引发草地荒漠化和沙尘暴等问题，严重影响草原的生态环境及畜牧业等相关草原产业的可持续健康发展。

草地生物之间的复杂联系可能会通过直接或间接的各种途径对土壤产生不同的调控，土壤又会反作用于地上生物。因此，理解植物和动物相互作用及其与土壤因子之间的关系对于认识和探究在不同放牧方式下草地植被及土壤的响应和变化规律具有十分重要的意义，为实现草地合理利用和优化管理提供科学依据。

1.2 本文拟解决的问题

基于上述研究背景，题目提供了草原放牧相关的数据集，其中基本数据包括锡林郭勒盟相关环境信息，监测点数据包括锡林郭勒牧户生态畜牧业模式群落样方调查数据集、牲畜数量调查数据集、不同放牧强度土壤碳氮监测数据集、轮牧放牧样地群落结构监测数据集，并结合相关资料及数据要求建立模型描述，解决如下问题：

问题一：从机理分析的角度，探索放牧策略与锡林郭勒草原土壤物理性质相互影响的本质，建立放牧方式对土壤湿度和植被生物量影响的数学模型，以及放牧强度对土壤湿度和植被生物量影响的数学模型，验证模型的准确性及稳定性。

问题二：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，并完成相应的表格。

问题三：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。并结合附件 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，并完成表格。

问题四：根据参考文献中沙漠化程度指数预测模型，首先，从附件提供及网络查找的方式分析土壤沙漠化影响的因子的数据，确定不同放牧强度下监测点的沙漠化程度指数值；然后，从土壤板结化的定性描述，尝试给出定量的土壤板结化定义；最后，结合问题三中的放牧策略，建立合理的土壤板结化模型，使得沙漠化程度指数及板结化程度最小。

问题五：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。

问题六：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）。

2 模型假设

本文模型基于以下合理假设：

1. 假设本文的数据真实有效，不存在错误和虚假数据，符合数据统计分析的基本要求。
2. 假设模型建立过程中，可以仅考虑问题中的核心因素，不考虑次要因素的影响。
3. 假设每改变一次主要操作变量（放牧策略）的值，相应的系统因变量都会相应及时的变化。
4. 假设附件中的年均数据可以作为建模时使用的初始数据使用。
5. 假设放牧策略对系统各个因子的影响程度不会发生变化，同时该特征也不会因为时间的变动而变动。

3 符号说明

符号	说明
p_1	一只羊的单价
p_2	植被生长系数
IC_{store}	植被截流量
LAI	叶面积指数
c_p	植被覆盖率
IC_{max}	定植被的最大截流量（mm）
k	植被密度校正因子
R_{cum}	累积降雨量（mm）
$M_{\text{临界值}}$	可持续发展下最低土壤全碳量
x_{it}	有机碳的吸收量
x_{ot}	有机碳的消耗量
Δt	时间增量
h	吸收系数
D_{tn}	植被的枯萎率
u	代谢比例
M	放牧方式
S	放牧强度
β	土壤湿度
w	植被生物量

$\frac{dw}{dt}$	植物的生长速度
$\frac{d\beta}{dt}$	土壤含水量变化速度
P	降水量
E	地表蒸散发率
Q_i	沙漠化程度指数中的强度因子
W_{ci}	沙漠化程度指数中的权重因子系数
SM	沙漠化程度指数
η	沙漠化程度指数的调节系数

4 问题一：放牧策略土壤物理性质和植被生物量影响的数学模型

4.1. 问题分析

针对问题一，需要建立放牧方式对土壤湿度和植被生物量影响的数学模型，以及放牧强度对土壤湿度和植被生物量影响的数学模型。从材料和附录中可知，放牧方式分为以下五种，分别为：全年连续放牧、禁牧、选择划区轮牧、轻度放牧、生长季休牧。放牧强度可以分为四种，分别为：对照、轻度放牧强度、中度放牧强度、重度放牧强度，其中禁牧和对照实际属于同一种情况。土壤湿度是影响陆气耦合的重要地表变量，它可以影响地表蒸发、调控地表热通量，从而对气候产生局地及非局地影响[3]。植被生物量是单位面积内生物的量，一般以湿重或者干重计。

首先，将放牧方式看作是时序信息，放牧方式的不同实际是放牧时间间隔天数的不同，例如全年连续放牧则是无间隔放牧，轻度放牧可以是 10 天放牧一次，生长季休牧视为 60 天放牧一次等，因此，可以将整个畜牧过程作为一个连续且循环的过程，利用微分思想，把整个过程划分为无数个阶段，当阶段间隔时间短时，这一阶段的属性值（土壤湿度、PH 值、植被数量的初始值等）可以直接作用于下一阶段，则整片草原生态系统的植被生物量的循环变化过程如下图 4-1 所示：

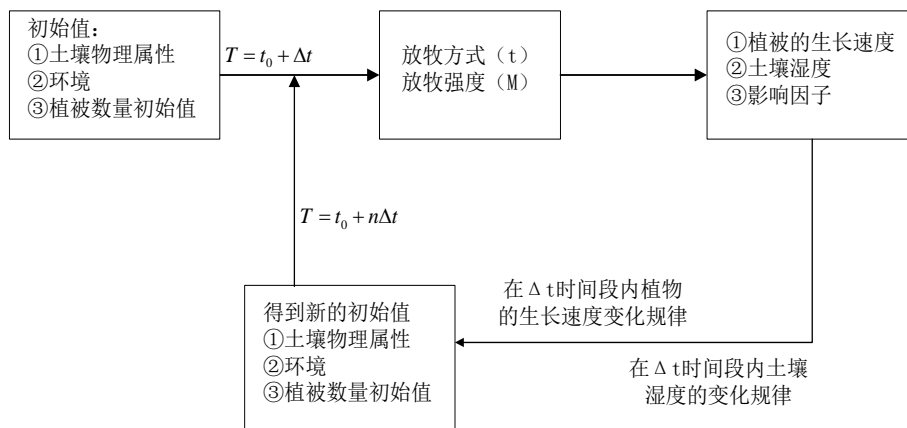


图 4-1 草原生态系统植被生物量的循环变化过程

然后，通过机理分析决定自变量和因变量，设放牧方式的决策变量为 M ，放牧强度的决策变量为 S ，土壤湿度的决策变量为 β ，植被生物量的决策变量为 w 。利用文献中相关资料，构建植被生物量 w 关于放牧方式 M 和放牧强度 S 的微分方程，土壤湿度 β 关于放牧方式 M 和放牧强度 S 的微分方程，并进行合理推导得出模型。

最后，根据相关文献以及附件中的数据，对模型进行准确性及有效性验证。解题思路如 4-2 所示：



图 4-2 问题一解题思路

4.2 模型建立

4.2.1 放牧策略-土壤湿度模型

植物的生长满足自身的生长规律，同时受到周围环境的影响。例如，降水、温度、土壤湿度、土壤 PH、营养等都决定植物的生长情况。当放牧对植物进行采食时，一方面植物的地上生物量减少；另一方面，放牧对植物有刺激作用，改变了植物原有的生长速率，适当的放牧会刺激植物的超补偿生长，同样不合理的放牧也会降低植物的生长速率。降水、温度对与本题来说不在控制范围内，营养受到人为影响因素较大不做考虑。简化为，本题中植物的生长速度 $\frac{dw}{dt}$ 受到土壤湿度 β 、土壤PH、放牧方式 M 和放牧强度 S 这四个因素的制约。同时，放牧活动能够降低表层土壤湿度、PH。

通过文献可知，对于放牧与植物生长之间的关系，Woodward 等建立简单模型：

$$\frac{dw}{dt} = 0.049w_0\left(1 - \frac{w}{4000}\right) - 0.0047Sw \quad (4-1)$$

式中， w 为植被生物量， S 为单位面积的载畜率。

通过问题分析放牧方式是一个连续循环的过程，利用微分方程的性质，基于植被生物量的循环变化过程递推出植被生物量 w 关于放牧方式 M 和放牧强度 S 的模型。

令放牧周期 $\Delta t = t_n - t_{n-1}$ ，设放牧方式 M 的表现为变量 Δt ，放牧强度 S 表现为单位面积的载畜率 S 。（当放牧方式为轻度放牧时， $\Delta t = 10$ ；当放牧方式为划区轮牧时， $\Delta t = 40$ ；当放牧方式为生长季休牧时， $\Delta t = 60$ ；当放牧方式为全年连续放牧时， $\Delta t = 360$ ）

当 $t=t_0$ 时，

$$\frac{dw_0}{dt} = 0.049w_0(1 - \frac{w_0}{4000}) - 0.0047Sw_0$$

其中， w_0 为 $t=t_0$ 时植被生物量， $\frac{dw_0}{dt}$ 是 $t=t_0$ 时植被生长速度。

可以求得 t_1 时的植被生物量

$$w_1 = \int_{t_0}^{t_1} \frac{dw_0}{dt} dt + w_0$$

此时，初始的环境条件发生了改变，受到这些改变的影响，需要对植被生物量的变化速度进行调整。可以通过设置影响因子来调整变化速度。

设置影响因子

$$a_{t_n} = f_1(w_{n-1}, M, S, PH_{n-1}, \beta_{n-1})$$

进而得到：

$$\frac{dw_2}{dt} = a_{t_1} \cdot \frac{dw_1}{dt} = a_{t_1}(0.049w_0(1 - \frac{w_0}{4000}) - 0.0047Sw)$$

同时，

$$w_2 = \int_{t_1}^{t_2} \frac{dw_2}{dt} dt + w_1 = \int_{t_1}^{t_2} \frac{dw_2}{dt} dt + \int_{t_0}^{t_1} \frac{dw_1}{dt} dt + w_0$$

通过数学的递推法类比可得：

$$\begin{aligned} \frac{dw_i}{dt} &= a_{t_i} \cdot \frac{dw_{i-1}}{dt} \\ w_n &= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{dw_i}{dt} dt + w_0 = \sum_{i=1}^n a_{t_i} \frac{dw_{i-1}}{dt} \Delta t + w_0 = \sum_{i=1}^n (a_{t_i} * \dots * a_{t_1}) \cdot \frac{dw_0}{dt} \cdot \Delta t + w_0 \end{aligned}$$

这里 $\sum_{i=1}^n (a_{t_i} * \dots * a_{t_1})$ 是影响因子，受到放牧方式 M 和放牧强度 S 影响； $\frac{dw_0}{dt}$ 由放牧强度 S 控制，当强度 S 上升， $\frac{dw_0}{dt}$ 变化量下降； Δt 由放牧方式 M 控制；式子中 w_0 作为初始量，是一个常量。

因此，得到植被生物量 w_n 关于放牧方式 M 和放牧强度 S 的模型如下：

$$w_n = \sum_{i=1}^n (a_{t_i} * \dots * a_{t_1}) \cdot \frac{dw_0}{dt} \cdot \Delta t + w_0 \quad (4-2)$$

4.2.2 放牧策略-植被生物量模型

由材料知土壤湿度的影响因素：气象因素（主要是降水）、土壤特征（孔隙度、容重、渗透性能等）、植被状况、人为活动都可能影响土壤含水量变化。植被状况可以由放牧方

式 M 和放牧强度 S 决定。所以题目中的土壤湿度 β 可以理解为只受到放牧方式 M 和放牧强度 S 的约束。

在参考文献中，有土壤含水量-降水量-地表蒸发模型：

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

式中， P 为该牧区供水率（主要为降水），是常数； 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/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。

将模型中的表达式代入，得到

$$\frac{d\beta}{dt} = P - E[\alpha^* \cdot G(w)] = P - E[\alpha^* \cdot (1 - e^{-\varepsilon_g w/w^*})]$$

式中除了 w 外的量都可以理解为常量，成草数量 w 几乎与（一）中所求植被生物量呈正相关关系，为了方便计算，这里直接将两者设为相等。

推理过程与（一）类似，得到：

$$\begin{aligned} \frac{d\beta_1}{dt} &= P - E[\alpha^* \cdot (1 - e^{-\varepsilon_g w_0/w^*})] \\ \frac{d\beta_1}{dt} &= P - E[\alpha^* \cdot (1 - e^{-\varepsilon_g w_0/w^*})] \\ \beta_1 &= \int_{t_0}^{t_1} [P - E[\alpha^* \cdot (1 - e^{-\varepsilon_g w_{t_1}/w^*})]] dt + \beta_0 \end{aligned}$$

因此，得到土壤湿度 β_n 关于放牧方式 M 和放牧强度 S 的模型如下：

$$\beta_n = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [P - E[\alpha^* \cdot (1 - e^{-\varepsilon_g w_{t_i}/w^*})]] dt + \beta_0 \quad (4-4)$$

至此，得到了植被生物量 w_n 与放牧方式 M 和放牧强度 S 的模型以及土壤湿度 β 与放牧方式 M 和放牧强度 S 的关系模型。其中，土壤湿度 β 的模型中除了 w_{t_i} 其他的量都可以看做已知量，而成草数量 w_{t_i} 与植被生物量 w_n 完全正相关。植被生物量 w_n 的模型中的影响因子受到放 z 。在不同的放牧方式 M 和放牧强度 S 的情况下，我们可以由参考文献大致确定影响因子 a_{t_i} 。

通过控制变量，在固定放牧方式的情况下，结合文献知识以及上述模型的大概关系画出植被生物量 w_n 、土壤湿度 β 在不同放牧强度影响下随时间变化曲线。通过与附件中提供的数据验证曲线是否符合实际情况，也就是验证了模型的正确性。

通过对得到的模型固定变量来绘制因变量随时间变化的图，得到图 4-3 到图 4-6：

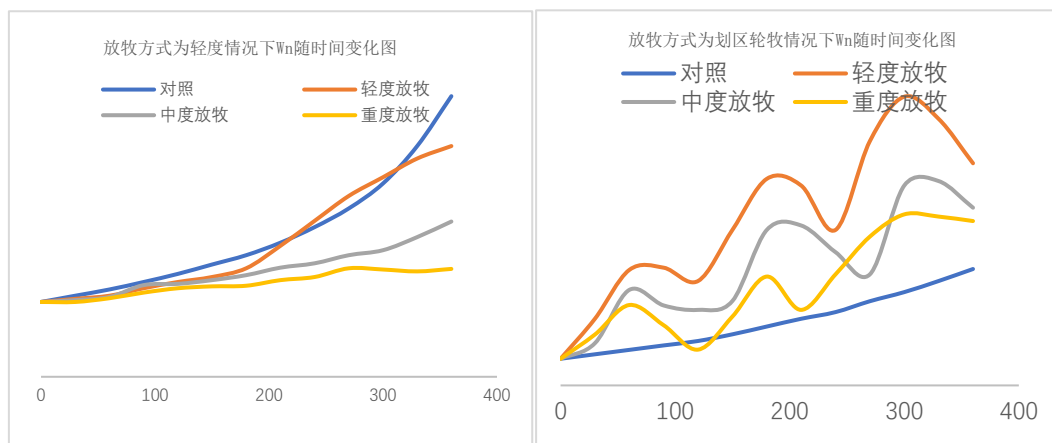


图 4-3 轻度与划区轮牧情况下 W_n 随时间变化图

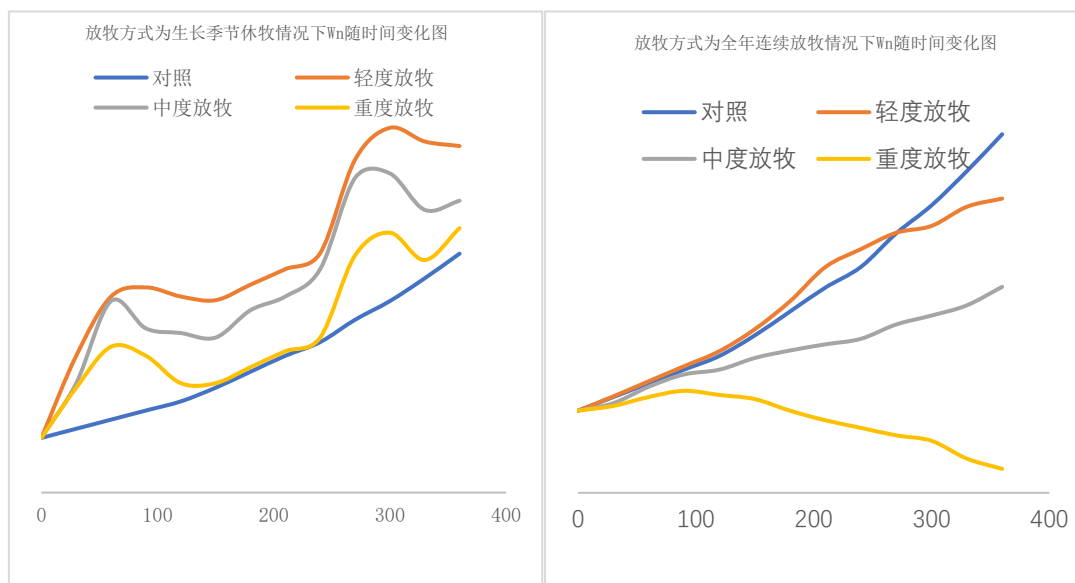


图 4-4 生长季节休牧与连续放牧情况下 W_n 随时间变化图

当放牧方式固定时，随着放牧强度的增大， w_n 整体来说在不断减小。在此基础上，改变放牧方式（或者说延长周期 Δt ），首先 w_n 一个变化周期前后的变化量越来越小，即延长放牧周期从整体上在减少 w_n 的量。

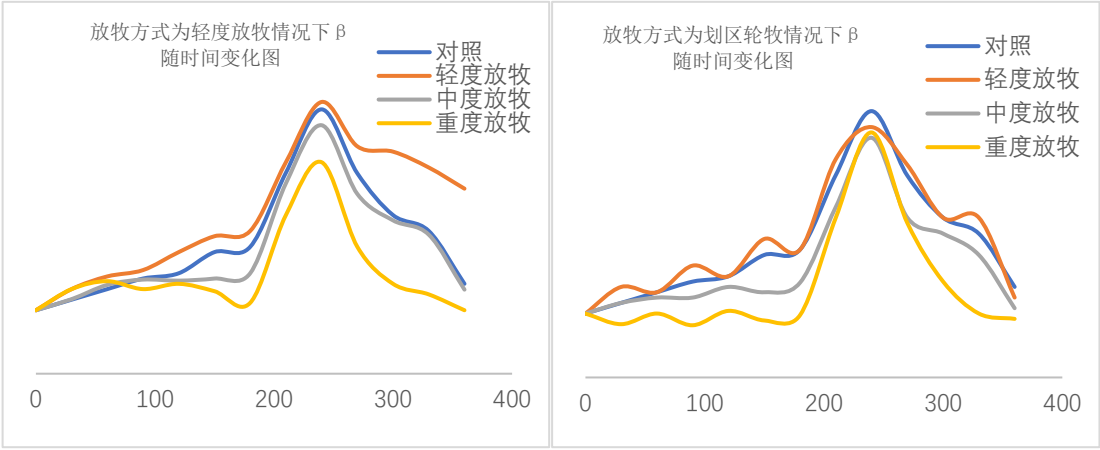


图 4-5 轻度放牧与划区轮牧情况下 β 随时间变化图

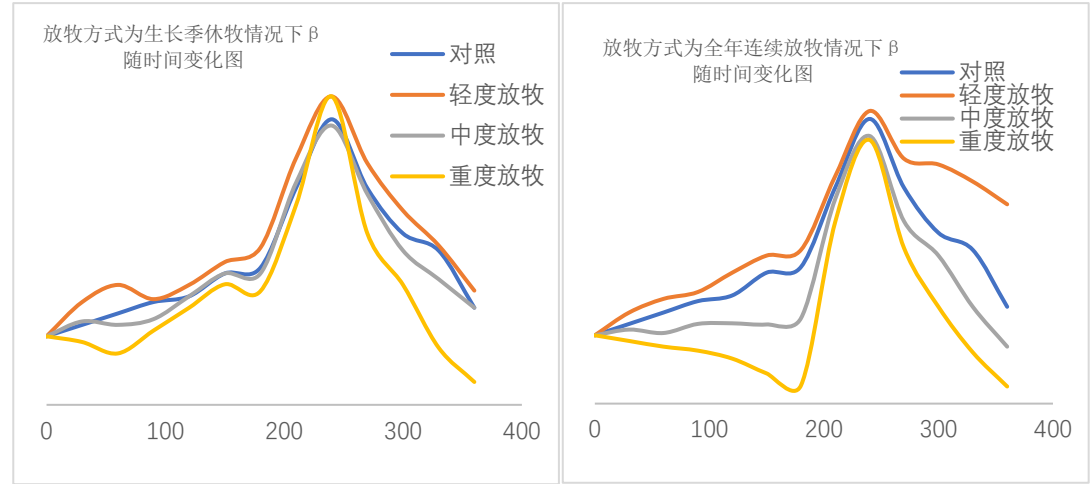


图 4-6 生长季休牧与全年连续放牧情况下 β 随时间变化

当放牧方式固定时，随着放牧强度的增大，土壤含水量 β 整体一开始有略微增加，但随着放牧强度继续增大， β 开始大幅下降；在此基础上，改变放牧方式（延长周期），加快了含水量 β 的变化速度，同时也使得 β 整体有一定下降。

4.3 模型检验

将附件 15 中的数据（全都固定放牧方式为划区轮牧）进行筛选，发现对于某一实验区域实行的放牧强度一直都是相同的。首先选取一个区的数据，其中每一年的数据不具有连续性。只能单独提出每一年的五月到九月的牧前和四轮放牧后的数据进行检验，由附件 15 的说明可知总生物量：植物干重之和（g），将时间与对应植被生物量进行画图观察趋势，同时与模型中放牧方式为划区轮牧的对应强度曲线进行对比。例如利用 G6 区域的数据进行正确性验证，G6 区域采取轻牧的放牧强度（每一轮放牧持续时间约为一个月），计算 2019 年每一个阶段的生物总量得到表 4-1：

表 4-1 2019 年每一个阶段的生物总量

阶段	生物总量 (g)
放牧前	114.49
一轮牧后	378.57
二轮牧后	328.96
三轮牧后	434.05
四轮牧后	377.63

绘图并对数据进行拟合后得出这段时期的变化趋势：

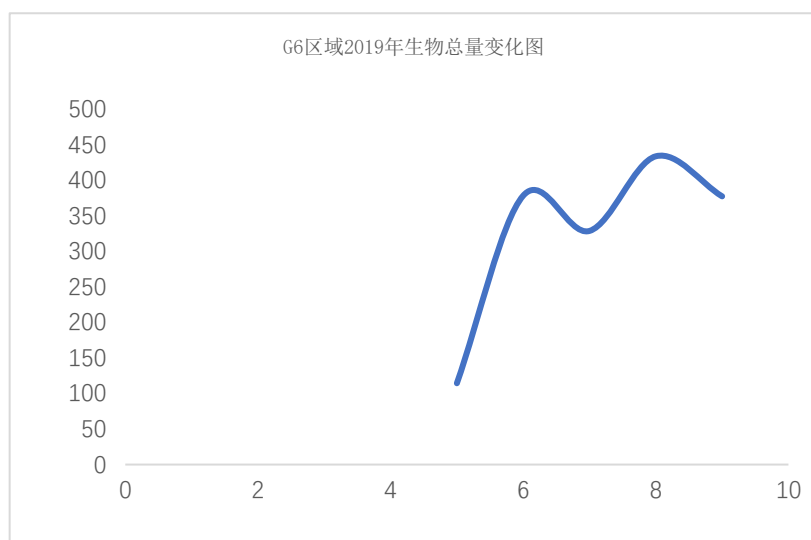


图 4-7 G6 区域 2019 年生物总量变化图

与放牧方式为划区轮牧情况下 W_n 随时间变化图进行对比，发现大致符合其中的变化趋势，证明该模型在这个例子是可行的。

在对附件 15 中 12 个不同放牧强度下的实验区域进行绘图观察后，验证了该模型正确且稳定。

4.4 模型评价

本题将整个放牧方式刻画为循环且连续的过程，通过机理分析，放牧策略与土壤湿度之间的相互影响关系，利用微分思想刻画函数关系，突出了放牧方式对土壤及生物种类的影响，同时也是根据客观事务发展的连续规律性。

运用相关数据，对模型的检验，对规律的变化大致符合，说明该模型可以很好的反应放牧方式与土壤湿度及土壤生物种类的关系。

5 问题二：不同深度土壤湿度预测

5.1 问题分析

问题二中，需要根据 2012 年-2022 年 3 月的土壤湿度，土壤蒸发数据，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。本问题分析思路详见下图 5-1。

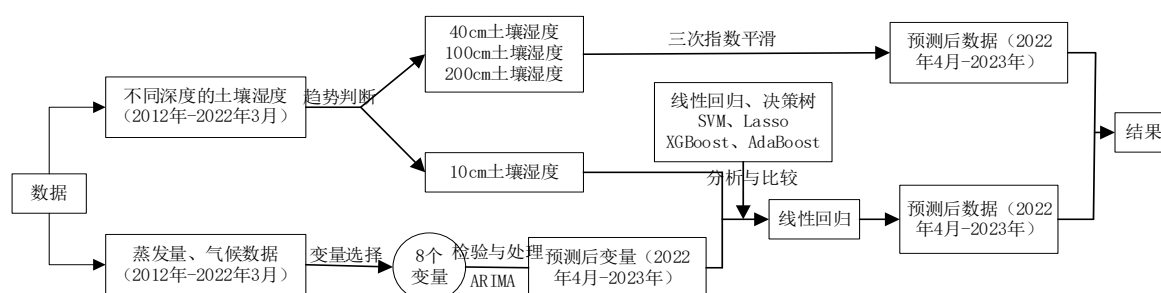


图 5-1 问题二解析思路图

首先通过时序图分析了不同深度(10cm、40cm、100cm、200cm)土壤湿度 2012 年-2022 年 3 月的周期趋势，发现 10cm 土壤湿度存在明显的周期性，而 40cm、100cm、200cm 的周期波动不明显，但呈现出一定的上升或下降趋势。因此，本文决定对 10cm 土壤湿度采用需要协变量的统计模型预测其 2022 年-2023 年的数据，而对 40cm、100cm、200cm 的土壤湿度本文采用不需要协变量的三次指数平滑进行预测。对于 10cm 土壤湿度所需协变量的选择问题而言，本文通过线性拟合的变量筛选方法选择了 8 个高贡献度的土壤增发、降水等变量。由于所有协变量只提供了 2012 年-2022 年 3 月的相关数据，因此，通过对这 8 个协变量进行检验与处理后，采用适当的 ARIMA 模型预测其 2022 年 3 月-2023 年的数据，然后再通过预测得到的协变量数据进一步通过 6 种机器学习模型的交叉分析对比，选择了线性回归方法来预测 10cm 深度的土壤湿度在 2022 年 3 月-2023 年的数据。最终预测得到了不同深度(10cm、40cm、100cm、200cm)土壤湿度 2022 年-2023 年数据。

以下为解题详细过程及重要结果。

5.2 模型建立和求解

5.2.1 描述性统计分析

通过对 4 个不同深度的土壤湿度数据的分析，得出个湿度的相关统计特征，包括均值、标准差、最小值、最大值、2 阶中心距、波峰、趋势斜率等特征，具体数据如下表所示：

表 5-1 不同深度的土壤湿度值的统计特征

指标	10cm 湿度(kg/m ²)	40cm 湿度(kg/m ²)	100cm 湿度(kg/m ²)	200cm 湿度(kg/m ²)
均值	14.56	42.38	55.351	167.21

标准差	2.849	7.476	14.536	0.9767
最小值	9.64	29.71	42.36	164.48
25 分位数	12.46	35.22	44.69	166.78
50 分位数	14.24	42.71	48.39	167.37
75 分位数	16.64	46.76	60.86	167.85
最大值	22.1	61.7	93.45	168.56
趋势斜率	0.01	0.08	0.23	-0.02
2 阶中心距	8.05	55.44	209.58	0.94
波谷	22	18	14	0
波峰	22	18	14	0

此可以分析出，随着深度越深，土壤湿度平均值越大，且 10cm 湿度趋势斜率不明显（0.01），而 40cm、100cm、200cm、趋势斜率较明显。此外可以发现 10cm、40cm、100cm 深度都存在波峰、波谷，但 200cm 深度的波峰和波谷数为 0。为了进一步分析数据特征，我们绘制了 4 个不同深度的土壤湿度数据的时序图（2012 年-2022 年 3 月），详见下图。

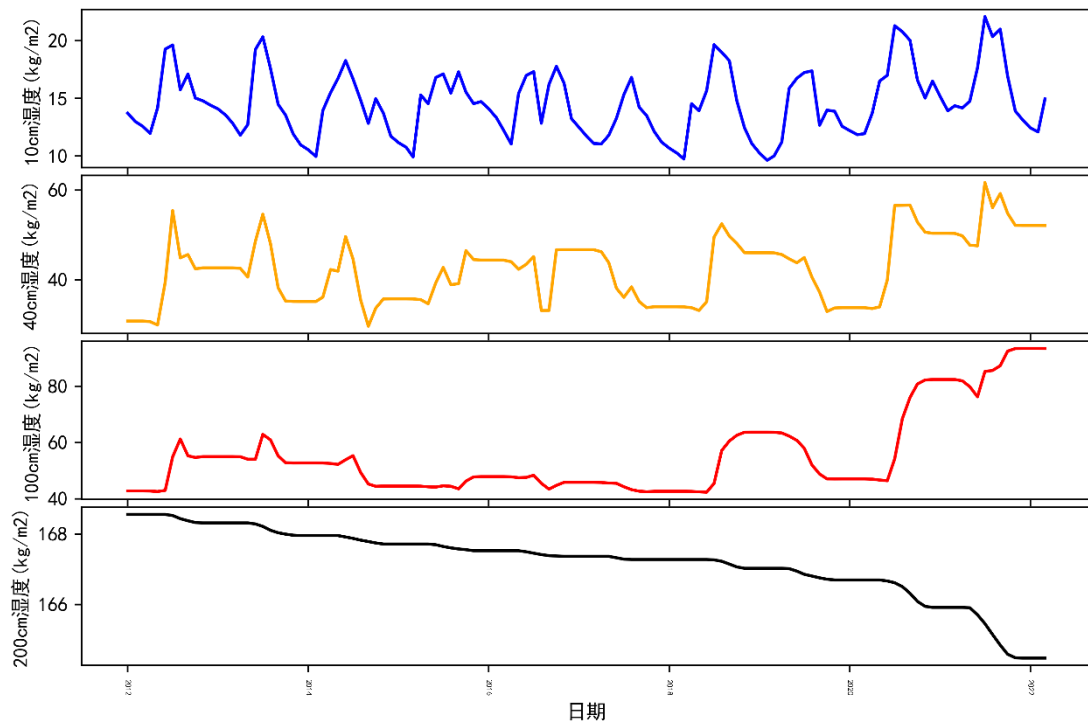


图 5-2 4 个不同深度的土壤湿度的时序图

从图中可以发现，10cm 深度的土壤湿度在 9.64~22.1 区间波动，且存在明显的周期模型，而 20cm 深度的土壤湿度周期波动规律不明显，但任能较明显的识别出其中的波峰和波谷。而对于 100cm 和 200cm 深度而言，其周期波动更加不明显，但存在明显的上升或下降趋势。因此，综合上面图表分析，对于 10cm 深度的土壤湿度和其他深度的土壤湿度进行分别预测处理。对于 10cm 深度的土壤湿度，本文采用线性回归的方法进行预测，而对于其他深度的土壤湿度本文采用 3 次指数平滑的方法进行预测。

5.3 模型构建与预测

5.3.1 10cm 深度土壤湿度预测

对于 10cm 土壤湿度的预测，首先通过线性拟合的变量筛选方法选择了 8 个高贡献度的土壤增发、降水等变量。然后对这 8 个协变量进行检验与处理后，分别采用差分移动平均自回归（Autoregressive Integrated Moving Average model, ARIMA）模型预测其 2022 年 3 月-2023 年的数据，然后再通过预测得到的协变量数据进一步通过线性回归、决策树、支持向量机回归、Lasso 回归、XGBoost 和 AdaBoost 共 6 种机器学习模型进行交叉分析对比，最终选择了表现最佳的线性回归方法来预测 10cm 深度的土壤湿度在 2022 年 3 月-2023 年的数据。

1) 变量选择

通过合并附件 4，附件 8 中 2012-2022 年土壤蒸发数据和气候数据，共得到 123 条观测样本和 27 个变量。对于本研究中，得到的数据变量大多是无相关或低相关变量，对于涉及到的第相关数据特征首先应当对数据降维，以便计算和可视化。常用的数据降维方法有主成分分析、树模型、相关系数等，问题二中我们采用线性拟合从中选取重要性变量，以便更为有效地综合提取有用数据，摒弃无效信息。根据线性拟合，对得到的各个变量的重要性（线性拟合系数）进行条形图绘制，如下图所示，图中系数绝对值越大说明变量对目标变量的贡献程度约高。

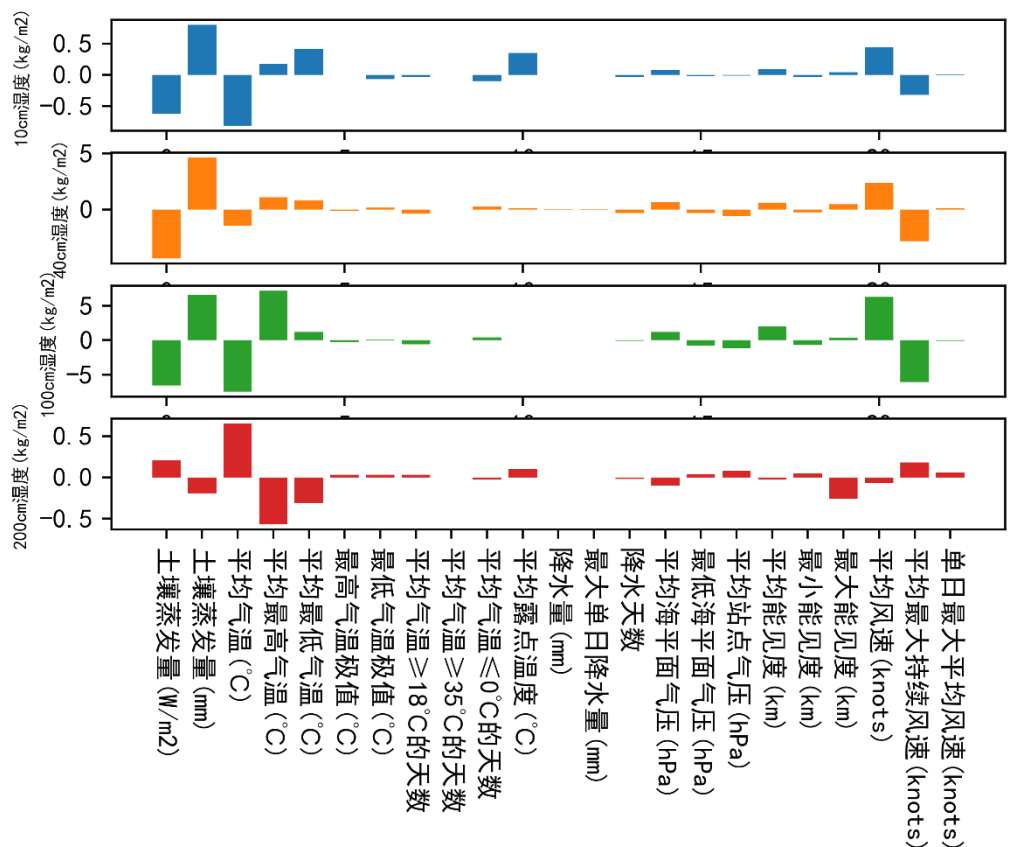


图 5-3 各个变量贡献度直方图

在计算出所有变量的线性拟合系数后本文进一步计算了各个变量对 4 种不同深度的土壤湿度的平均贡献系数，见下表。选择绝对值排序为前 8 个变量作为本文用于预测的主要自变量，分别为：“土壤蒸发量(W/m2)”，“土壤蒸发量(mm)”，“平均气温(℃)”，“平均最高气温(℃)”，“平均最低气温(℃)”，“平均能见度(km)”，“平均风速(knots)”和“平均最大持续风速(knots)”。

表 5-2 各个变量线性拟合系数值

变量	10cm 湿度	40cm 湿度	100cm 湿度	200cm 湿度
土壤蒸发量(W/m2)	-0.62	-4.37	-6.58	0.21
土壤蒸发量(mm)	0.8	4.6	6.56	-0.19
平均气温(℃)	-0.81	-1.46	-7.46	0.65
平均最高气温(℃)	0.17	1.08	7.17	-0.57
平均最低气温(℃)	0.42	0.81	1.18	-0.31
最高气温极值(℃)	<0.001	-0.13	-0.27	0.03
最低气温极值(℃)	-0.07	0.17	0.07	0.03
平均气温≥18℃的天数	-0.04	-0.38	-0.61	0.03
平均气温≥35℃的天数	<0.001	<0.001	<0.001	<0.001
平均气温≤0℃的天数	-0.1	0.26	0.41	-0.03
平均露点温度(℃)	0.35	0.09	0.01	0.1
降水量(mm)	<0.001	0.02	0.02	<0.001
最大单日降水量	<0.001	0.03	0.01	<0.001
降水天数	-0.03	-0.31	-0.1	-0.02
平均海平面气压	0.08	0.68	1.16	-0.1
最低海平面气压	-0.02	-0.33	-0.76	0.04
平均站点气压	-0.01	-0.59	-1.17	0.08
平均能见度(km)	0.09	0.61	2	-0.03
最小能见度(km)	-0.03	-0.26	-0.74	0.05
最大能见度(km)	0.04	0.48	0.34	-0.26
平均风速(knots)	0.44	2.39	6.25	-0.07
平均最大持续风速	-0.32	-2.83	-6.08	0.18
单日最大平均风速	0.01	0.13	-0.08	0.06

接下来本研究采用 ARIMA 模型对选择的这 8 个变量预测其 2022 年 4 月至 2023 年 12 月的数据。

2) 平稳性检验

数据的平稳性检验是建立 ARIMA 条件的前提，如果不满足，那么需要对时间序列数据的平稳性进行修正和调整。对序列的平稳性有两种检验方法，一种是根据时序图和自相关图显示的特征做出判断的图检验法；一种是构造检验统计量进行假设检验的方法。据平稳时间序列均值、方差为常数的性质，平稳序列的时序图应该显示出该序列始终在一个常数附近随机波动。以年作为单位长度，对选取的关于土壤湿度、土壤蒸发量、降水量、降水天数的时间序列数据绘制初始的时序走势图，如下图所示。

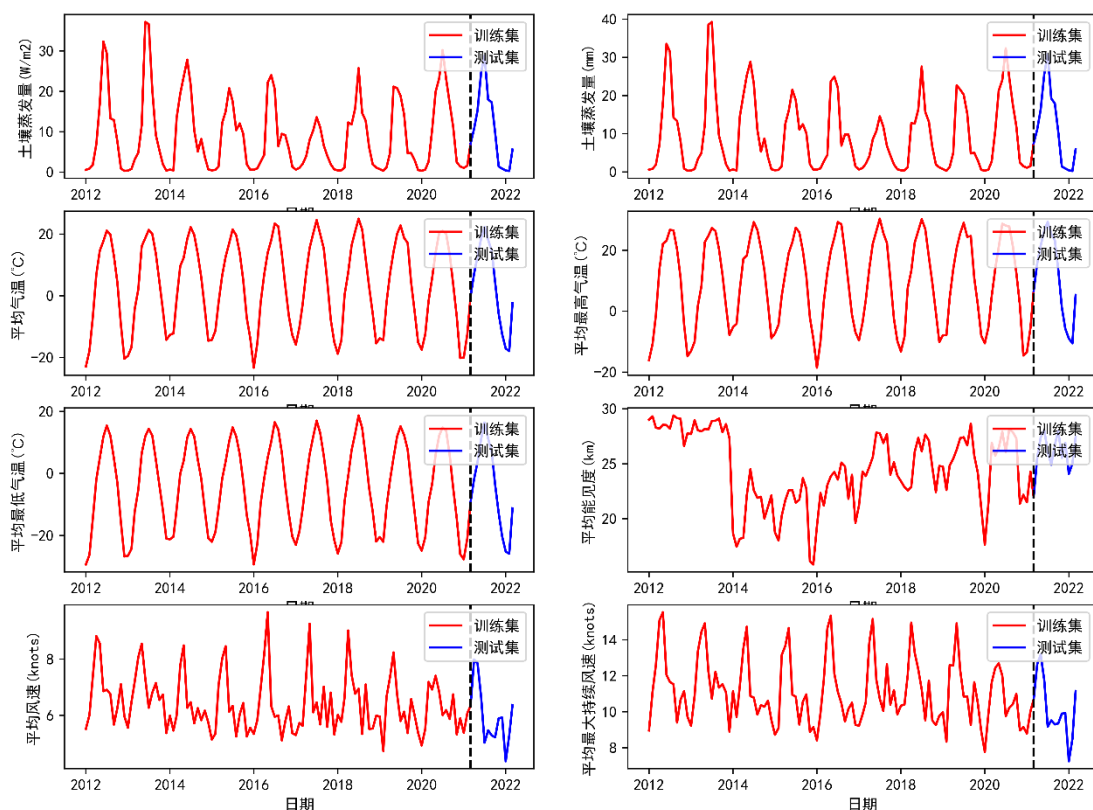


图 5-4 各协变量的时间序列图

观察时序图可以发现，8 个变量的时序图呈现出剧烈的上下波动，周期性非常明显，所以初步判断该序列存在某种趋势，且都为非平稳序列。但真实结论还需要通过假设检验进行验证，即如下的单位根检验。

3) 单位根检验

单位根检验是对序列是否存在单位根进行判断：如果序列平稳，就不存在单位根；否则就会存在单位根。对原序列检验结果如下所示：

表 5-3 各变量的单位根检验结果

变量	统计量	P	阶数	1%	5%	10%	是否平稳
土壤蒸发量	-1.271	0.642	11	-3.498	-2.891	-2.583	否
土壤蒸发量	-1.257	0.649	11	-3.498	-2.891	-2.583	否
平均气温	-2.677	0.078	11	-3.498	-2.891	-2.583	否
平均最高气温	-2.404	0.141	13	-3.5	-2.892	-2.583	否
平均最低气温	-2.811	0.057	11	-3.498	-2.891	-2.583	否
平均能见度	-3.757	0.003	1	-3.492	-2.888	-2.581	是
平均风速	-1.755	0.403	12	-3.499	-2.892	-2.583	否
平均最大持续风速	-0.94	0.775	12	-3.499	-2.892	-2.583	否

原假设为存在单位根，根据上表可以得到，除“平均能见度”外，其余 7 各变量在检验水平为 0.01 的条件下，接受原假设，原序列存在单位根，序列不平稳。故需要对原序列进行平稳化调整。

4) 差分调整

差分是调整平稳序列的重要方式，本文对 8 个协变量的原序列均采用了一阶差分。差分之后数据的时序图如下所示：

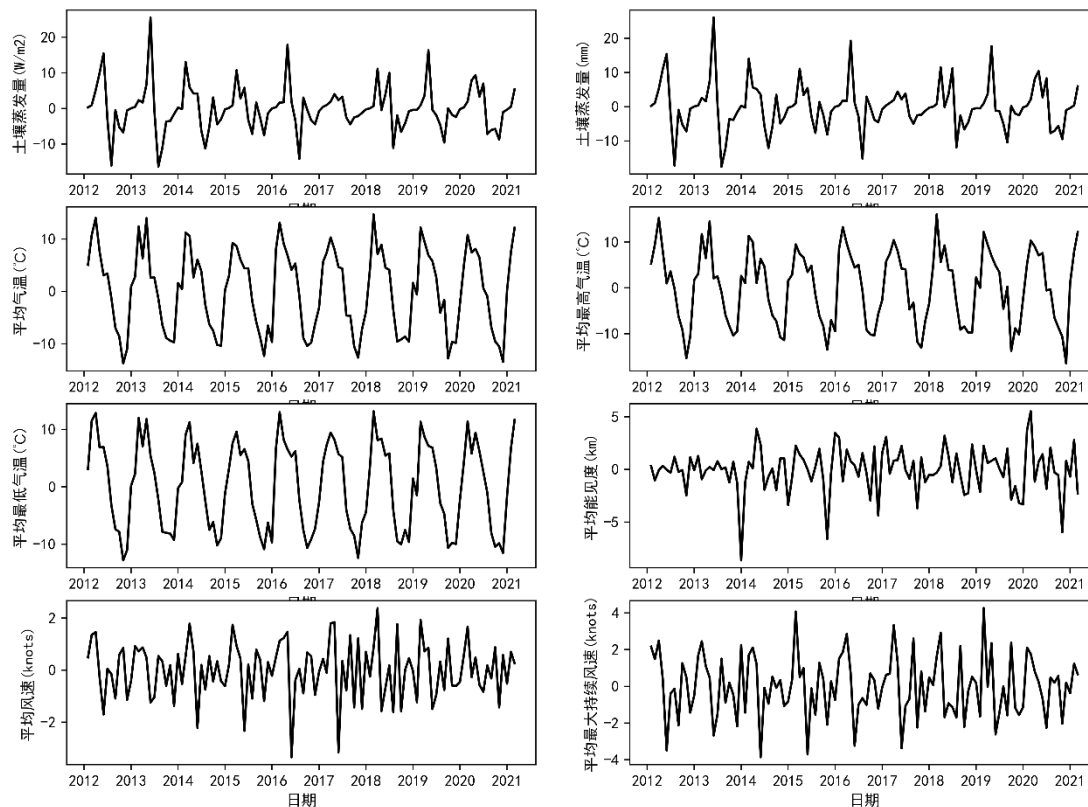


图 5-5 土壤蒸发量一阶差分后时间序列图

从时序图中初步可以发现，8 个协变量的序列经过差分调整之后序列变得平稳有界。进一步通过单位根检验，结果如下：

表 5-4 各变量的单位根检验结果

变量	统计量	P	阶数	1%	5%	10%	是否平稳
土壤蒸发量	-9.2	0	10	-3.498	-2.891	-2.583	是
土壤蒸发量	-9.273	0	10	-3.498	-2.891	-2.583	是
平均气温	-9.981	0	10	-3.498	-2.891	-2.583	是
平均最高气温	-9.329	0	10	-3.498	-2.891	-2.583	是
平均最低气温	-10.644	0	10	-3.498	-2.891	-2.583	是
平均能见度	-6.741	0	7	-3.496	-2.89	-2.582	是
平均风速	-4.242	0.001	13	-3.5	-2.892	-2.583	是
平均最大持续风速	-4.134	0.001	13	-3.5	-2.892	-2.583	是

从表中可以看出，经过一阶差分之后，所有序列的单位根检验结果表明拒绝原假设，差分后序列平稳。可以进行下一步的时间序列分析。

5) 纯随机检验

纯随机检验也称为白噪声检验，是专门用来检验序列是否为纯随机序列的一种方法。本报告中计算了差分后序列的 Q 统计量和 Ljung-Box 统计量，结果如下（此处只展示部分结果，其余详见附录 1）：

表 6-5 部分协变量的纯随机检验结果

滞后阶数	土壤蒸发量(W/m ²)		土壤蒸发量(mm)		平均气温(°C)		平均最高气温(°C)	
	Q	Prob(>Q)	Q	Prob(>Q)	Q	Prob(>Q)	Q	Prob(>Q)
1	11.466	0.001	11.128	0.001	59.193	<0.001	53.45	<0.001
2	11.557	0.003	11.165	0.004	77.964	<0.001	70.208	<0.001
3	13.023	0.005	12.708	0.005	78.141	<0.001	70.562	<0.001
4	20.164	<0.001	20.078	<0.001	98.97	<0.001	90.243	<0.001
5	27.799	<0.001	27.606	<0.001	152.923	<0.001	138.083	<0.001
6	38.196	<0.001	38.212	<0.001	225.457	<0.001	202.649	<0.001
7	51.194	<0.001	50.969	<0.001	283.766	<0.001	255.526	<0.001
8	54.202	<0.001	54.01	<0.001	302.903	<0.001	274.156	<0.001
9	54.202	<0.001	54.012	<0.001	303.219	<0.001	274.532	<0.001
10	54.661	<0.001	54.583	<0.001	321.208	<0.001	290.553	<0.001
11	65.564	<0.001	65.232	<0.001	377.166	<0.001	342.45	<0.001
12	97.833	<0.001	97.469	<0.001	457.116	<0.001	416.384	<0.001
13	113.602	<0.001	113.008	<0.001	513.426	<0.001	469.83	<0.001
14	113.615	<0.001	113.041	<0.001	529.681	<0.001	484.456	<0.001
15	114.578	<0.001	113.988	<0.001	529.823	<0.001	484.734	<0.001
16	118.048	<0.001	117.662	<0.001	547.147	<0.001	500.903	<0.001
17	125.932	<0.001	125.448	<0.001	595.783	<0.001	544.964	<0.001
18	133.126	<0.001	132.719	<0.001	656.03	<0.001	596.451	<0.001
19	139.355	<0.001	138.936	<0.001	705.904	<0.001	642.236	<0.001
20	146.124	<0.001	145.728	<0.001	723.273	<0.001	658.986	<0.001

由于 8 个差分后序列在各阶延迟下 Q 统计量的 P 值都非常小（<0.1），所以差分后序列显著拒绝纯随机的原假设。因而可以认为 8 个协变量序列的变动不属于纯随机波动，也即序列属于非白噪声序列。

6) ARIMA 模型定阶

通过绘制各协变量的一阶差分自相关和偏自相关图进行初步定阶，此处只展示其中一个协变量（平均风速）的一阶差分自相关和偏自相关图，其余结果见附件 2。

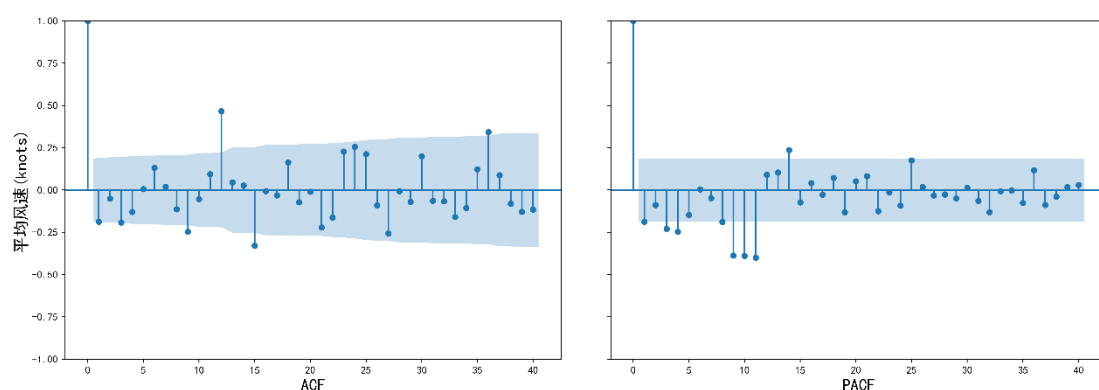


图 5-7 平均风速差分序列的自相关图和偏自相关图

注意在滞后阶数较高时也存在较高相关性，可能的原因是因为序列中存在季节周期性。则对于土壤蒸发量(W/m²)有以下模型可以供选择：ARMA(4, 6)模型：即自相关图在滞后 4 阶之后缩小为 0，且偏自相关滞后 6 阶缩小至 0；ARMA(6, 6)模型：即自相关图在滞后 6 阶之后缩小为 0，且偏自相关滞后 6 阶缩小至 0；ARMA(2, 4)模型：即自相关图在滞后 2 阶之后缩小为 0，且偏自相关滞后 4 阶缩小至 0；对于平均气温(℃)有以下模型可以供选择：ARMA(11, 0)模型：即自相关图在滞后 11 阶之后缩小为 0，且偏自相关缩小至 0，则是一个阶数 p=1 的自回归模型；其他协变量的差分序列同理。

由于自相关图和偏自相关图的定阶方法存在一定的主观性，故可以采用客观的模型选择准则来确定阶数。本研究主要采用三种准则：AIC、BIC、HQIC。AIC=-2 ln(L) + 2 k：赤池信息量；BIC=-2 ln(L) + ln(n)*k：贝叶斯信息量；HQIC=-2 ln(L) + ln(ln(n))*k。构造这些统计量所遵循的统计思想是一致的，就是在考虑拟合残差的同时，依自变量个数施加“惩罚”。但要注意的是，这些准则不能说明某一个模型的精确度，也即是说，对于三个模型 A，B，C，我们能够判断出 C 模型是最好的，但不能保证 C 模型能够很好地刻画数据，因为有可能三个模型都是糟糕的。计算最大 p、q 为 3 以内的所有组合情况的模型。设置最大阶数为 13 阶，得到 8 各协变量在各准则信息量的最小信息量的阶数如下表所示：

表 5-6 自动定阶各准则结果

变量	AIC	BIC	HQIC
土壤蒸发量(W/m ²)	(4, 6)	(2, 4)	(2, 4)
土壤蒸发量(mm)	(4, 6)	(2, 4)	(2, 4)
平均气温(℃)	(11, 0)	(11, 0)	(11, 0)
平均最高气温(℃)	(6, 7)	(2, 5)	(6, 7)
平均最低气温(℃)	(11, 0)	(2, 4)	(11, 0)
平均能见度(km)	(2, 4)	(0, 0)	(2, 4)
平均风速(knots)	(9, 8)	(11, 0)	(9, 7)
平均最大持续风速(knots)	(6, 9)	(6, 7)	(6, 7)

针对土壤蒸发量(W/m²)：由 AIC 准则可以得到，最优阶数为 ARMA(4, 6)；由 BIC 准则确定的最优模型为 ARMA(2, 4)；由 HQIC 信息量确定的最有模型为 ARMA(2, 4)。针对土壤蒸发量(mm)：由 AIC 准则可以得到，最优阶数为 ARMA(4, 6)；由 BIC 准则确定的最优模型为 ARMA(2, 4)；由 HQIC 信息量确定的最有模型为 ARMA(2, 4)。针对平均气温(℃)：

由 AIC 准则可以得到, 最优阶数为 ARMA(11, 0); 由 BIC 准则确定的最优模型为 ARMA (11, 0); 由 HQIC 信息量确定的最有模型为 ARMA (11, 0)。

AIC、BIC 是两种不同的模型选择方法, AIC 选择出来的模型往往预测精度更好, 而 BIC 准则不具备该优良性质。BIC 主要的优点是选择相合性。结合自相关和偏自相关图, 本报告对各协变量分别采用了如下模型。

土壤蒸发量(W/m2): ARIMA (2, 1, 4) 模型、土壤蒸发量(mm): ARIMA (2, 1, 4) 模型、平均气温(℃): ARIMA (11, 1, 0) 模型、平均最高气温(℃): ARIMA (6, 1, 7) 模型、平均最低气温(℃): ARIMA (11, 1, 0) 模型、平均能见度(km): ARIMA (2, 1, 4) 模型、平均风速(knots): ARIMA (9, 1, 7) 模型、平均最大持续风速(knots): ARIMA (6, 1, 7) 模型。

7) ARIMA 模型结果

对 8 个协变量的 2012 年-2022 年 3 月的序列数据按照日期顺序, 选择前 90%为训练集, 后 10%的数据为测试集。通过 python 软件得到各模型的拟合结果如下表所示 (此处只展示了部分结果, 全部结果见附录三):

表 5-7 8 个协变量的各 ARIMA 模型结果

变量	模型	coef	std err	z	P> z	[0.025	0.975]
土壤蒸发量 (W/m2)							
	ar. L1	1.7316	0.004	478.830	0.000	1.725	1.739
	ar. L2	-0.9997	0.001	-750.207	0.000	-1.002	-0.997
	ma. L1	-2.1321	58.183	-0.037	0.971	-116.168	111.904
	ma. L2	1.0825	65.829	0.016	0.987	-127.941	130.106
	ma. L3	0.6348	2.890	0.220	0.826	-5.029	6.299
	ma. L4	-0.5853	34.048	-0.017	0.986	-67.319	66.148
	sigma2	16.2850	947.373	0.017	0.986	-1840.531	1873.101
.....							
平均最大持续风速							
	ar. L1	0.7801	0.105	7.441	0.000	0.575	0.986
	ar. L2	0.6550	0.183	3.573	0.000	0.296	1.014
	ar. L3	-1.9265	0.111	-17.286	0.000	-2.145	-1.708
	ar. L4	0.7212	0.108	6.698	0.000	0.510	0.932
	ar. L5	0.7087	0.179	3.959	0.000	0.358	1.060
	ar. L6	-0.9704	0.102	-9.484	0.000	-1.171	-0.770
	ma. L1	-1.6942	0.176	-9.639	0.000	-2.039	-1.350
	ma. L2	0.0554	0.395	0.140	0.889	-0.719	0.829
	ma. L3	2.4433	0.373	6.549	0.000	1.712	3.175
	ma. L4	-2.3679	0.263	-9.002	0.000	-2.883	-1.852
	ma. L5	-0.0813	0.364	-0.223	0.823	-0.795	0.632
	ma. L6	1.5218	0.341	4.459	0.000	0.853	2.191
	ma. L7	-0.8690	0.167	-5.194	0.000	-1.197	-0.541
	sigma2	0.8149	0.169	4.831	0.000	0.484	1.145

8) ARIMA 模型诊断

残差正态性检验：首先对各模型的残差进行正态性检验，观察 ARIMA 模型的残差是否是平均值为 0 且方差为常数的正态分布（服从零均值、方差不变的正态分布）。

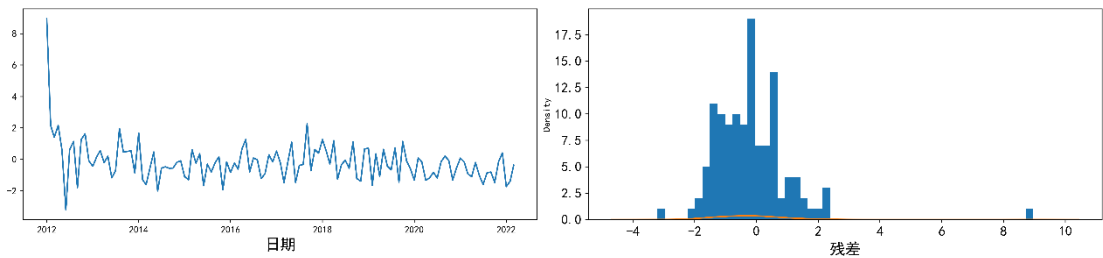


图 5-8 平均风速模型残差的序列图和核密度估计图

绘制模型残差的序列图，可以发现残差序列在 0 附近波动，且波动无异常，说明残差序列服从 0 均值同方差的分布，可以发现残差序列的直方图近似服从正态分布。

残差纯随机检验：对于 ARIMA 模型，其残差被假定为高斯白噪声序列，所以当用 ARIMA 模型去拟合数据时，拟合后我们要对残差的估计序列进行 Ljung-Box 检验，判断其是否是高斯白噪声，如果不是那么就说明 ARIMA 模型也许并不是一个适合样本的模型。对本报告得到的模型残差进行纯随机检验，如果模型建立可行，则残差应为一个白噪声序列。进行的纯随机检验结果如下所示（全部结果见附录 5）：

表 5-8 模型对部分协变量残差序列的纯随机检验结果

滞后阶数	土壤蒸发量(W/m2)			土壤蒸发量(mm)			平均气温(℃)		
	AC	Q	P 值	AC	Q	P 值	AC	Q	P 值
1	0.07	0.612	0.434	0.081	0.826	0.364	-0.061	0.468	0.494
2	-0.002	0.613	0.736	0.013	0.846	0.655	-0.144	3.089	0.213
3	-0.096	1.805	0.614	-0.097	2.061	0.56	-0.073	3.775	0.287
4	-0.04	2.014	0.733	-0.039	2.256	0.689	0.12	5.646	0.227
5	0.051	2.354	0.798	0.057	2.685	0.748	0.071	6.306	0.278
6	0.114	4.066	0.668	0.112	4.334	0.632	-0.077	7.094	0.312
7	0.016	4.099	0.768	0.019	4.38	0.735	-0.033	7.236	0.405
8	0.034	4.25	0.834	0.032	4.517	0.808	0.006	7.242	0.511
9	0.057	4.689	0.861	0.044	4.782	0.853	-0.03	7.362	0.599
10	-0.114	6.444	0.777	-0.106	6.322	0.788	0.052	7.728	0.655
11	-0.001	6.444	0.842	-0.006	6.326	0.851	0.016	7.762	0.734
12	0.209	12.523	0.405	0.217	12.868	0.379	-0.153	11.007	0.528
13	0.148	15.589	0.272	0.143	15.719	0.265	0.014	11.034	0.608
14	-0.053	15.991	0.314	-0.042	15.964	0.316	-0.013	11.059	0.681
15	-0.06	16.504	0.349	-0.059	16.466	0.352	0.049	11.402	0.724
16	-0.043	16.776	0.4	-0.044	16.739	0.403	-0.077	12.258	0.726
17	0.017	16.82	0.467	0.023	16.814	0.467	0.076	13.091	0.73
18	0.063	17.405	0.495	0.066	17.446	0.493	0.035	13.267	0.775
19	0.046	17.712	0.542	0.044	17.731	0.54	-0.004	13.27	0.824
20	-0.077	18.607	0.548	-0.073	18.527	0.553	0.031	13.414	0.859

由于在各阶延迟下 Q 统计量的 P 值都非常大 (>0.1)，所以模型残差序列显著不拒绝纯随机的原假设。因而可以认为残差序列的波动属于纯随机波动，也即建立的 ARIMA 模型可行。

9) ARIMA 模型预测

至此土壤蒸发量、降水量、降水天数的 ARIMA 模型已经建立完成，为了进一步评估模型的效果，需要对原序列进行预测评价。首先在原序列的基础上进行回代预测，然后对 2022 年 4 月份之后的数据进行预测。进一步绘制预测结果的折线图如下，黑色的线代表实际的数据，红色的线代表对建模所用数据进行回代预测，蓝色的线代表对 2022 年 3 月-2023 年数据进行预测。

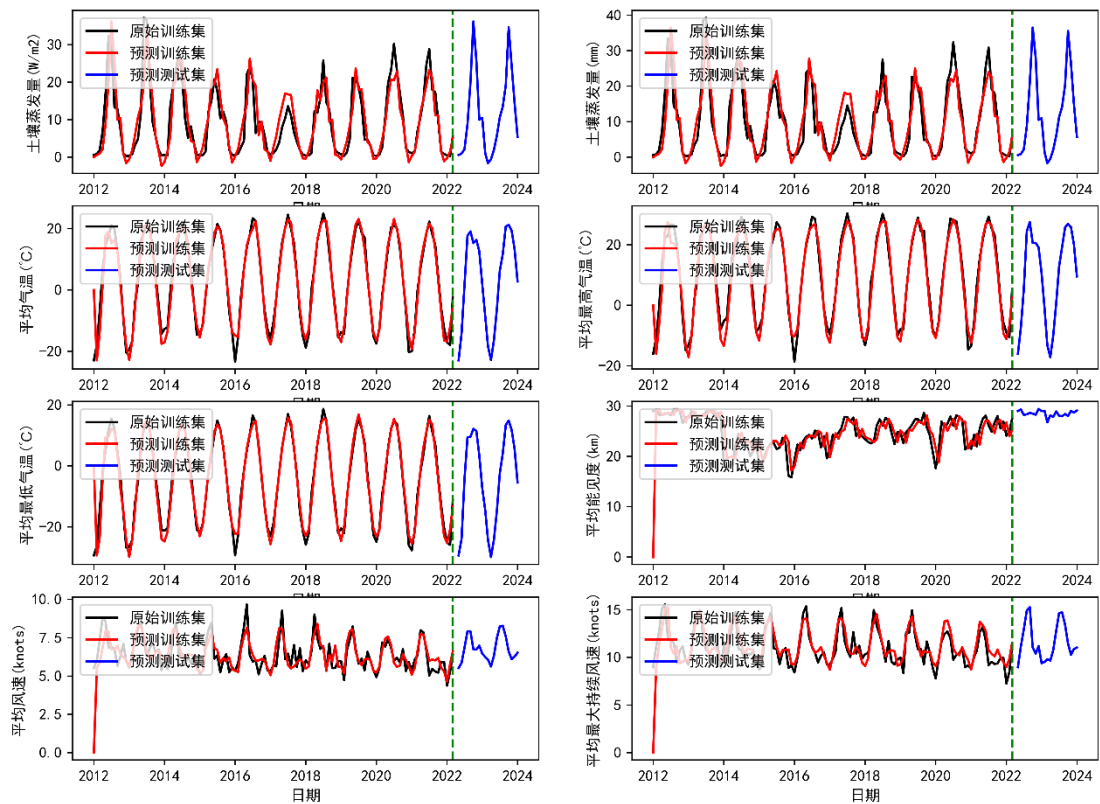


图 5-9 8 个协变量时间序列预测效果图

从图形来看，在测试集数据上预测效果拟合较好。另外通过计算预测效果的均方误差得：

表 5-9 各 ARIMA 模型预测效果的均方误差	
变量	均方误差
土壤蒸发量(W/m2)	4.469
土壤蒸发量(mm)	4.7309
平均气温(℃)	3.2147
平均最高气温(℃)	3.171
平均最低气温(℃)	3.4931

平均能见度(km)	3.3976
平均风速(knots)	0.8116
平均最大持续风速(knots)	1.3136

随着对未来预测的期数越长，总体来说预测效果越差。但是在对建模数据的回代拟合上来看，效果较好。即所有的协变量预测的均方误差都小于 5，说明模型的预测效果较好，预测结果具有一定的可信度。

10) 预测模型分析与比较

接下来本文通过 ARIMA 模型预测得到的 8 个协变量 2022 年 4 月-2023 年 12 月的数据作为自变量，10cm 深度土壤湿度作为因变量。为了构建最佳的统计模型，我们使用多种机器学习方法：线性回归（Linear Regression）、套索回归（Lasso Regression）、随机森林（Random Forest）、支持向量机回归、XGBoost、AdaBoost。对 2012 年-2022 年 3 月的数据按照日期顺序，取前 70%作为训练集，后 30%作为测试集。基于此数据使用这六种算法进行 10 折交叉验证。各算法得到的结果如下图所示：

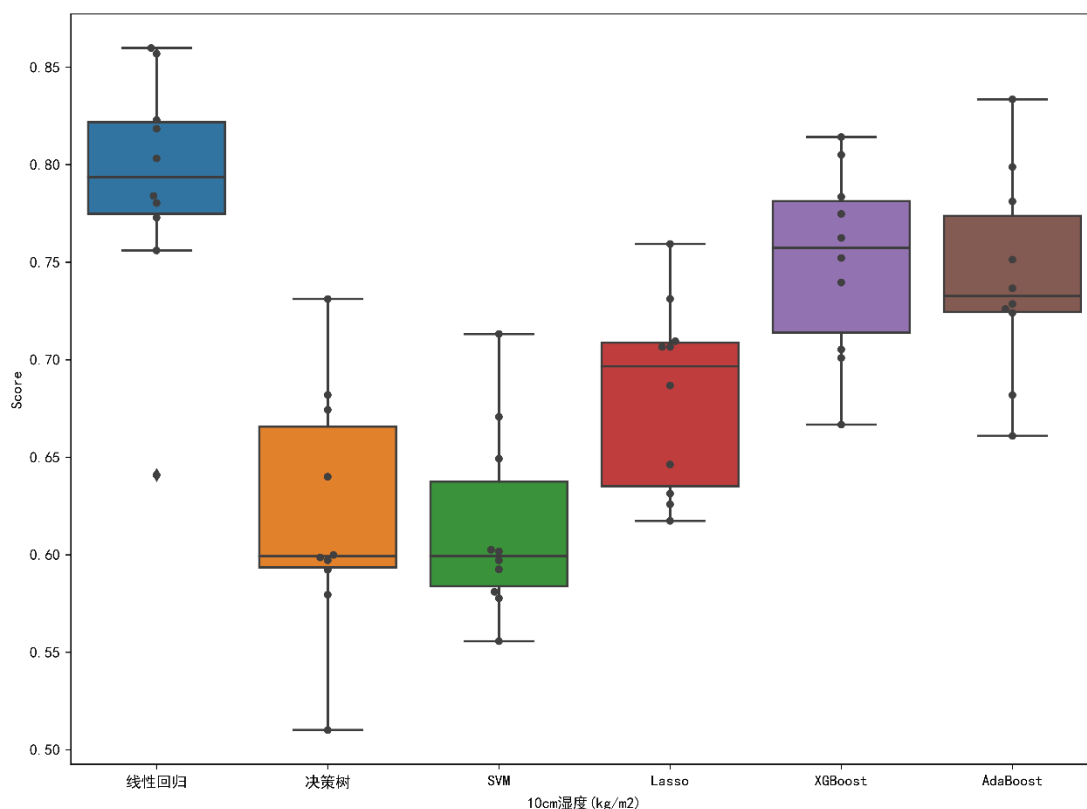


图 5-10 各算法对 10cm 土壤湿度 10 折交叉验证结果的箱型图

通过分析发现，线性回归模型所得出的效果最好，10 折交叉验证的平均决定系数接近 0.8。因此我们选择线性回归方法来构建 10cm 深度土壤湿度的数学模型。

通过 python 软件，拟合上述训练集和测试集，得到如下的线性回归方程：

$$Y = 0.031X_1 + 0.16X_2 - 0.193X_3 - 0.192X_4 + 0.427X_5 + 0.049X_6 + 0.762X_7 - 0.735X_8$$

模型在训练集中的均方误差 MSE 为 1.40，在测试集的均方误差 MSE 为：2.99，因此认为模型可行度较高。进一步，通过将前文预测得到的 2022 年 4 月-2023 年的 8 个协变量数

据代入其中，即可预测得到 2022 年 4 月-2023 年的 10cm 深度土壤湿度数据。具体预测结果见本章后文统一展示。

5.3.2 其他深度土壤湿度预测

由于本章前文所分析，对 40cm、100cm、200cm 的土壤湿度本文将采用不需要协变量的三次指数平滑进行预测。

1) 三次指数平滑预测

虽然一次指数平滑在产生新的数列的时候考虑了所有的历史数据，只是考虑其静态值，但没有考虑时间序列当前的变化趋势。所以二次指数平滑一方面考虑了所有的历史数据，另一方面也兼顾了时间序列的变化趋势。若时间序列的变动呈现出二次曲线趋势，则需要采用三次指数平滑法进行预测。实际上是在二次指数平滑的基础上再进行一次指数平滑。它与二次指数平滑的区别就是三次平滑还考虑到了季节效应。因此本文能够考虑到季节效应的三次指数平滑模型。具体预测模型如下：

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \quad (5-1)$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2} \left[(6 - 5\alpha)S_t^{(1)} - 2(5 - 4\alpha)S_t^{(2)} + (4 - 3\alpha)S_t^{(3)} \right] \quad (5-2)$$

$$c_t = \left(\frac{\alpha^2}{2(1-\alpha)^2} \left[S_t^1 - 2S_t^{(2)} + S_t^3 \right] \right) \quad (5-3)$$

$$y'_{t+T} = a_t + b_t T + c_t T^2 \quad (5-4)$$

其中 y_t 是 t 期的实际值； $S_t^{(1)}$ 是第 t 周期的一次指数平滑值； $S_t^{(2)}$ 是第 t 周期的二次指数平滑值； $S_{t-1}^{(2)}$ 是第 $t-1$ 周期的二次指数平滑值； α 是平滑系数；

对于三次指数平滑模型而言，平滑系数的选择至关重要，平滑系数值越大，实际的数据所占的比例就越大，原预测值所占比重就越小；反之，平滑系数取小值，权数变化较迅速，平滑作用越强，但对实际的数据的变化反应迟缓。此外，还需要根据时间序列的波动大小，选择相应的平滑系数值；当时间序列波动较缓时，选择较小的平滑系数值，反之，选择较大的平滑系数值。

2) 模型构建与预测结果

综合考虑，本文对于 40cm 深度的土壤湿度设置平滑系数为 0.3，对于 100cm 深度的土壤湿度设置平滑系数为 0.5，对于 200cm 深度的土壤湿度设置平滑系数为 0.7。而对于周期季节设置，从前文的时序图以及序列的季节性分解可以看出，设置为 12 将是一个最好的选择。

进一步，通过求解前文所述三次指数平滑预测模型，即可预测得到 2022 年 4 月-2023 年的 40cm、100cm 和 200cm 深度土壤湿度数据。通过 python 软件，拟合 2012 年-2022 年 3 月各深度土壤湿度数据，具体预测结果见本章后文统一展示。其中 40cm 深度土壤湿度原序列预测均方误差 MSE 为：12.56；100cm 深度土壤湿度原序列预测均方误差 MSE 为：7.11；200cm 深度土壤湿度原序列预测均方误差 MSE 为：0.01；因此认为模型可行性较高。

5.4 结果分析

通过前文介绍，2022 年 4 月-2023 年预测结果如下表所示：

表 5-10 模型预测最终结果

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	12.93546	52.4924	92.89019	164.48
	05	10.55426	51.35786	91.79438	164.4781
	06	13.16471	54.58107	90.83684	164.4321
	07	12.85707	61.2063	95.31026	164.3652
	08	14.14679	56.53798	96.85321	164.2802
	09	22.51796	54.45215	97.57871	164.1952
	10	20.79075	54.77956	97.90615	164.1181
	11	14.73236	54.20812	96.59067	164.0771
	12	14.85879	54.17276	96.43181	164.072
2023	01	13.89564	54.17289	96.70347	164.072
	02	12.9203	54.17297	96.8811	164.072
	03	12.22033	54.24851	96.7931	164.072
	04	12.22432	54.61091	96.28599	164.072
	05	12.45043	53.47637	95.19018	164.07
	06	13.13323	56.69958	94.23264	164.0241
	07	14.18362	63.32481	98.70606	163.9571
	08	16.79566	58.65649	100.249	163.8721
	09	20.68703	56.57066	100.9745	163.7871
	10	18.7052	56.89807	101.302	163.7101
	11	15.82144	56.32663	99.98647	163.669
	12	13.49121	56.29127	99.82762	163.664

为了便于观测数据预测效果，进一步绘制了原序列和预测序列，见下图。其中 10cm 深度是由包括 8 个协变量的线性回归预测所得，40cm、100cm 和 200cm 深度是由三次指数平滑预测而得。可以明显的看出，在测试集中红色线条几乎符合黑色线条走势，因此 2022 年 3 月-2023 年的红色线条预测数据具有一定的准确性。

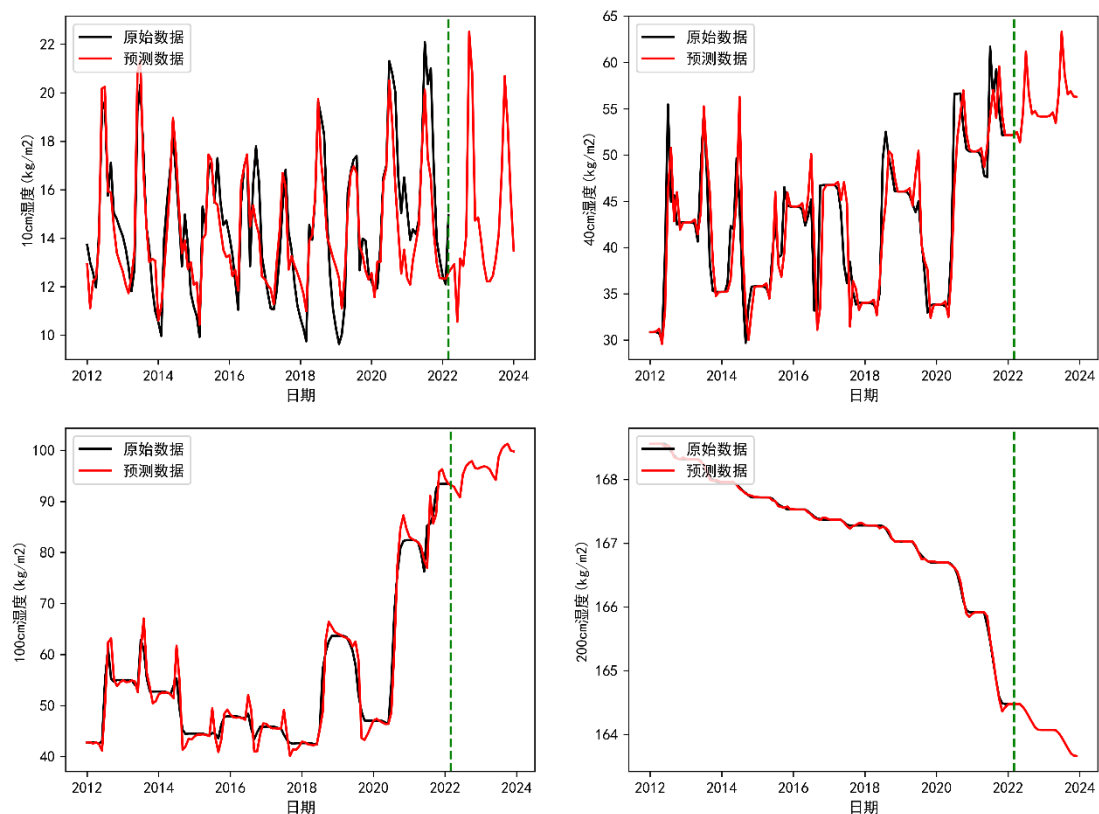


图 5-11 模型预测结果时序图

6 问题三：不同放牧策略土壤化学性质影响的数学模型

6.1 问题分析

需要建立不同放牧策略(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，接着对有机碳、无机碳、全 N、土壤 C/N 比等值，这是对前面模型扩展，基于前面模型基础上进行优化整合，在预测时需要结合放牧策略对化学性质的影响，需要时间序列模型加上一个影响因子进行修正。现在对第一问的方案进行优化，然后进行分析，本题解题思路如下 6-1 所示；

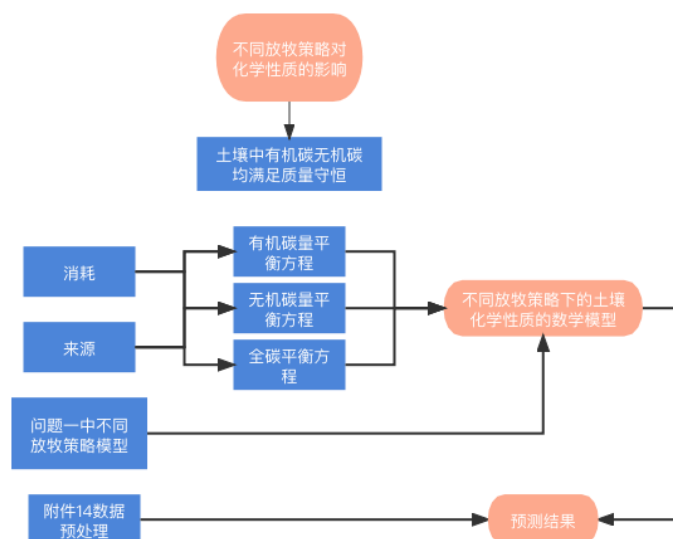


图 6-1 问题三解题思路

6.2 模型建立与求解

根据题目分析可知土壤中的化学物质满足以下动态图 6-2:

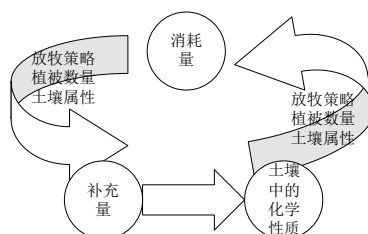


图 6-2 土壤中的化学物质满足以下动态图

从题目有关信息中可以了解到,土壤中的有机碳、无机碳、全 N 都是动态循环的过程,均满足质量守恒,即初始化学物质加吸收的化学物质减去消耗的化学物质会稳定在一定水平。故此对土壤中的各个化学元素进行逐一分析。通过上图可知在一段时间内土壤中的各化学物质满足质量守恒。

6.2 土壤化学性质分析

6.2.1 SOC 土壤有机碳含量分析

令土壤有机碳的含量为 x ，则 t 时刻土壤有机碳含量计为 x_t ，则存在，

$$x_t = x_{t-1} + x_{it} - x_{ot} \quad (t = 1, 2 \dots n) \quad (3-1)$$

其中， x_{it} 表示从 $t-1$ 时刻到 t 时刻土壤中有机碳的吸收量， x_{ot} 表示从 $t-1$ 时刻到 t 时刻土壤中有机碳的消耗量，当 $t = 1$ 时 x_0 表示土壤中有机碳的初始值。

对于 x_{ot} （土壤中有机碳的消耗量）而言，根据资料可知，土壤中的有机碳的主要作用是用于植被生长，而部分有机碳会随着水土流失而流失，本文在研究过程中 Δt 的取值较小，而实际生活中因水土流失带来的土壤中的有机碳消耗几乎可忽略不计[1]，因此土壤中的有机碳的消耗量与植被数量相关，令植被的吸收有机碳的系数为 h_1 则，

$$x_{ot} = hw_{tn} = h_1 \sum_{i=2}^{i=n} \int_{t_{i-1}}^{t_i} v_i dt + w_0 \quad (3-2)$$

对于 x_{it} 而言，根据资料可知土壤中的无机碳来源主要由两部分组成，第一部分是植被自然枯萎转化为土壤中的无机碳 x_1 ，第二部分是放牧时生出排泄物转化为土壤中的无机碳 x_2 ，逐一分析两部分来源构成。

令枯萎的植被吸收有机碳的系数为 h_2 ，则

$$x_1 = w_{tn} \cdot D_{tn} \cdot h_2 = w_{tn} \cdot \beta_{tn} \left(e^{\frac{\varepsilon \varphi w_{tn}}{w}} \cdot -1 \right) h_2 \quad (3-3)$$

根据扩展资料， $D = \beta^* (e^{\varepsilon_g w/w^*} - 1)$ 为枯萎率， D_{tn} 表示植被的枯萎率，推导方式与问题一相同。

对于放牧过程中，牲畜的日食量为 1.8，代谢比例为 u ，排泄物中无机碳的吸收系数为 h_3 ，则，

$$x_2 = \sum_{i=2}^{i=n} \int_{t_{i-1}}^{t_i} s 1.8 u h_3 dt \quad (3-4)$$

由(3-3), (3-4)得，

$$x_t = x_1 + x_2 \quad (3-5)$$

综上，由（3-1）（3-2）（3-5）得到，

$$x_t = x_{t-1} + w_{tn} \cdot \beta_{tn} \left(e^{\frac{\varepsilon \varphi w_{tn}}{w}} \cdot -1 \right) h_2 + \sum_{i=2}^{i=n} \int_{t_{i-1}}^{t_i} 1.8 u h_3 dt - h_1 \sum_{i=2}^{i=n} \int_{t_{i-1}}^{t_i} v_i dt + w_0 \quad (3-6)$$

6.2.2 SIC 土壤无机碳含量分析

通过资料可知土壤无机碳主要受土壤表面植被数量的影响，当植被数量增大时有助于防止土壤钙化，无机盐的成分降低，反之则增高，则可构建一下函数量化其数量

$$y_t = h_4 y_0 e^{(w_0 - w_{tn})} \quad (3-6)$$

其中， y_t 表示无机碳含量， y_0 表示 w_{tn} 取 w_0 初始状态无机碳含量， h_4 表示参数。

1) 全 N

根据资料显示土壤中的含量取决于土壤中的有机质[2]，令土壤中的有机质含量为 z_t ，则。

$$z_t = h_5 \cdot x_t \quad (3-6)$$

h_5 取 1.724。令土壤中的全 N 含量为 F_t ，则，

$$F_t = a z_t + b \quad (3-7)$$

a、b 为拟合参数值。

2) 全碳 D

由附件 14,

$$D = x_t + y_t \quad (3-8)$$

以上为对锡林郭勒草原土壤化学性质影响的数学模型。

6.3 数据预测

对附件 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) 中同年同放牧校区进行平均值处理。

利用问题二中 ARIMA 模型结合锡林郭勒草原土壤化学性质影响的数学模型进行预测, 三种化学物质的预测效果图如下:

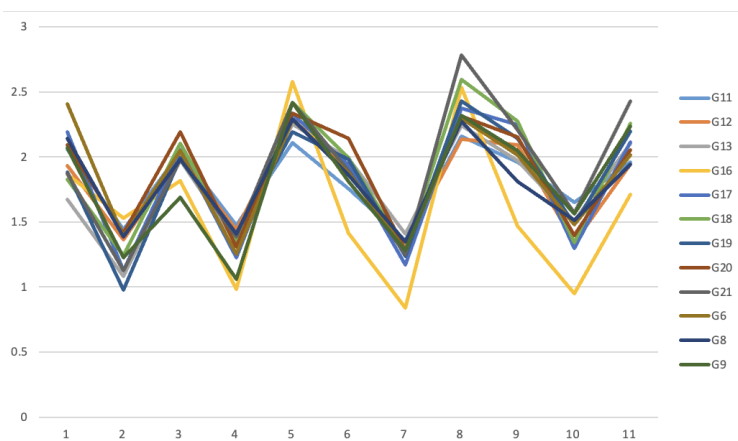


图 6-3 SOC 预测图

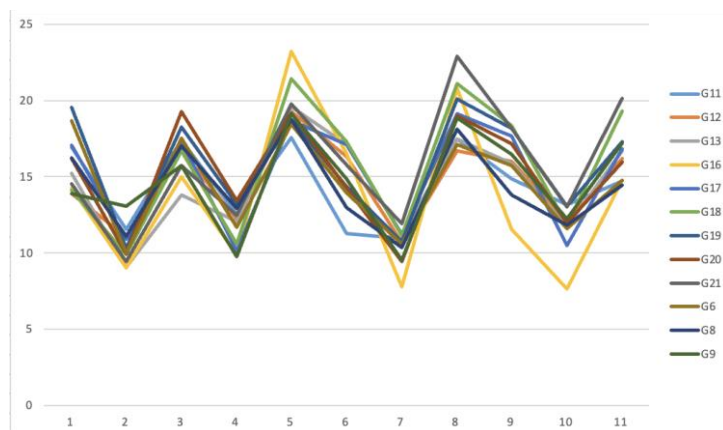


图 6-4 SIC 预测图

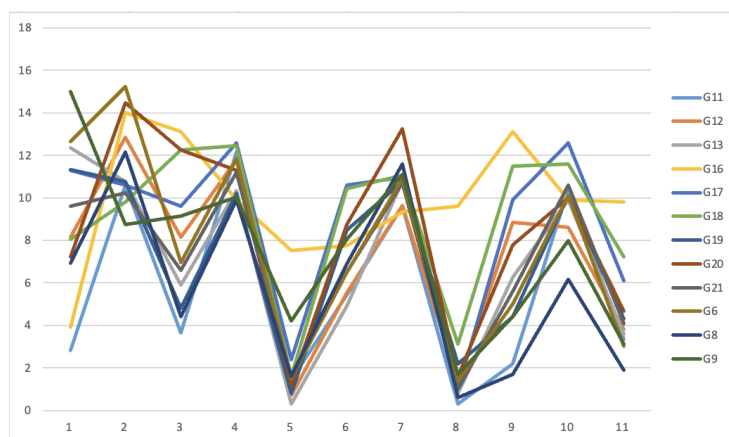


图 6-5 全 N 预测图

使用上述模型对 2022 年数据进行预测：

表 6-1 2022 年数据预测结果

放牧强度	Plot 放牧小区	SOC 土壤有机碳	STC 土壤全碳	全 N	土壤 C/N 比
NG	G17	16.38	24.42	1.53	15.97
	G19	15.88	22.50	1.91	11.81
	G21	20.21	27.55	2.17	12.69
LGI	G6	15.88	22.50	1.91	11.81
	G12	13.77	19.97	1.90	10.52
	G18	15.46	23.99	2.17	11.07
MGI	G8	14.03	16.64	1.89	8.81
	G11	14.27	18.92	1.96	9.65
	G16	11.66	21.52	1.56	13.80
HGI	G9	16.14	21.10	2.09	10.09
	G13	16.04	21.14	2.03	10.40
	G20	15.69	22.36	1.86	12.04

7 问题四：沙漠化程度指数与板结化程度

7.1 问题分析

查阅相关文献及材料，可知沙漠化相关关系的指标因子可以考虑风速、降水、气温（三个气象因素）；植被盖度、地表水资源、地下水位（三个地表因素）；人口数量、社会经济水平（三个人文因素）等，土壤板结化定义与土壤湿度、土壤容重、土壤有机物含量相关。

首先，由文献可知，沙漠化程度指数预测模型与相关因子强度及因子权重相关，通过查找相关数据，对 9 个影响沙漠化相关因子原始数据进行处理，连续时间内缺失的数据用

拉格朗日插值法插值，对检测区的数据取均值，个别缺失因素数据舍去，通过计算得出因子强度，再运用参考文献中的方法，确定因子权重，最后将因子强度与因子权重带入沙漠化程度指数预测模型可得出不同放牧强度下检测点的沙漠化程度值[6]。

然后，土壤板结化定义中，土壤湿度越少，容重越大，有机物含量越低，土壤板结化程度越严重，需要通过构建综合评价体系，确定 3 个因素之间的相关系数，同上，先对 3 个数据进行收集，再通过数据预处理，再选择合适的权重确定方法，确定指标权重。

最后，根据上述得出的沙漠化程度指数预测模型及土壤板结化模型构建目标函数，结合问题三中的放牧策略为约束条件，构建多目标函数，采用优化算法求解模型，得出沙漠化程度指数及板结化程度最小时的放牧策略。

问题四的解题思路如下 6-1 所示：

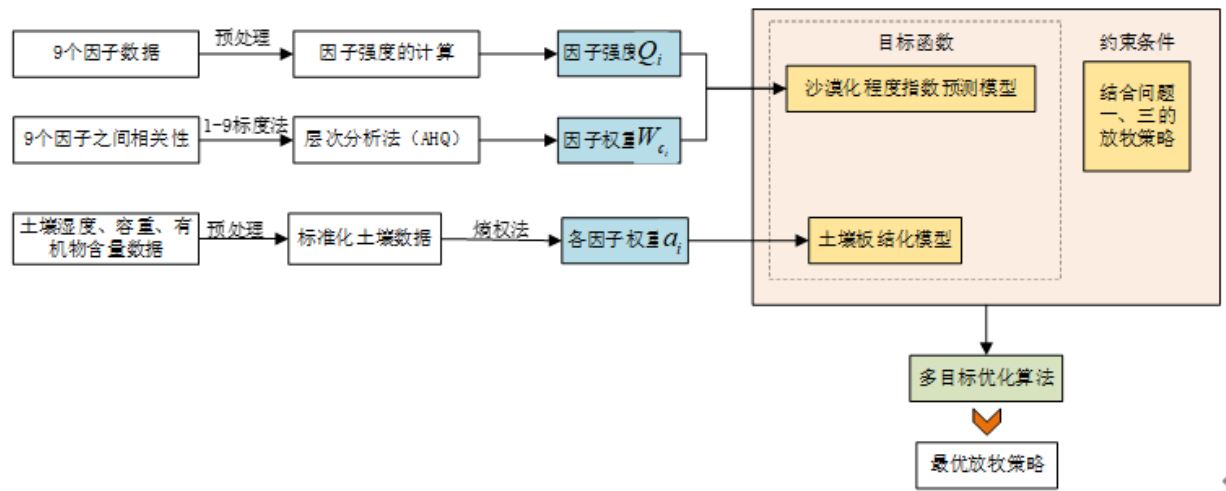


图 7-1 问题四解题思路

7.2 模型建立与求解

7.2.1 沙漠化程度预测指数

1) 因子强度计算

通过查找相关资料，在《锡林郭勒统计年鉴 2020》中找到关于锡林浩特地区每个月的气象情况[7]，通过计算得到监测区域暖季平均风速、暖季降水量、暖季平均气温，具体结果如下表 6-1：

表 7-1 监测区域暖季平均值

指标	暖季平均风速 c_1 (m/s)	暖季降水量 c_2 (mm)	暖季平均气温 c_3 (°C)
数值	3.35	264.3	12.78

在附件 5 的 2020 绿植覆盖率中，只有 7 月到 8 月的绿植覆盖率数据，通过参考往年的数据变化趋势，用拉格朗日插值法插值六月的数据（0.113、0.128、0.152），从而计算夏季 6 月到 8 月中绿植覆盖率为 0.262。由网络查询得知，锡林郭勒盟水资源总量为 34.93 亿立方米/年，其中地表水资源量 9.08 亿立方米/年，地下水资源量 30.23 亿立方米/年。地表水资源很大程度受到降雨量的影响，通过全年降水量趋势，估计该区域夏季地表水资

源占全年的百分之六十二左右，得出夏季地表水资源量大约为 5.62 亿立方米，除以锡林郭勒盟的面积为 20300 平方公里，得到夏季水资源量为 $2.77 \times 10^4 \text{m}^3/\text{km}^2$ 。人文因素中的影响因子全都通过附件 14 的检测点数据计算，得到人口密度 1.75 人/平方公里，牲畜密度 708 头/平方公里，人均家庭经营纯收入 199605 元。由于搜集不到具体的关于夏季地下水位的数据，也无法从本题的模型中推得数据，故舍弃该因素。

按照参考文献中的计算方法，计算得出因子强度为 6-2:

表 7-2 因子强度计算值

因子	数值	对应的 Q_i
暖季平均风速	3.35	1
暖季降水量	264.3	0
暖季平均气温	12.78	0.611
夏季植被盖度	26.2	0.595
夏季地表水资源	2.77	0.923
人口密度	1.75	0
牲畜密度	708	1
人均家庭经营收入	199605	1

2) 因子权重系数计算

目前常用的方法有层次分析法和主成分分析法，鉴于地下水埋深和植被盖度因子缺少长期连续的观测数据，不能很好的支持 PCA 方法，本文采用 AHP 方法确定各指标因子的权重系数，首先通过 1~9 标度方法构建研究区域的判断矩阵，然后计算权重对判断矩阵计算最大特征值 λ_{max} 和对应的特征向量 W，W 经过标准化后即为一层次中相应元素对上一层次因素的相对重要性的排序权值，即为权重，相对应的结果如下表 6-3 到表 6-5。

表 7-3 A-B 层判断矩阵、权重及一致性检验

A	B_1	B_2	B_3	权重 w_{c_i}	一致性检验
B_1	1	1	1	0.3275	$\lambda_{max} = 3.05$
B_2	1	1	2	0.4126	CI=0.025
B_3	1	1/2	1	0.2599	CR=0.0431 \leq 0.1

表 7-4 B_1 -C 层判断矩阵、权重及一致性检验

B_1	C_1	C_2	C_3	权重 w_{c_i}	一致性检验
C_1	1	2	3	0.5503	$\lambda_{max} = 3.02$
C_2	1/2	1	1	0.2404	CI=0.01
C_3	1/3	1	1	0.2093	CR=0.0172 \leq 0.1

表 7-5 B_2 -C 层判断矩阵、权重及一致性检验

B_2	C_4	C_5	权重 w_{c_i}
C_4	1	2	0.66
C_5	1/2	1	0.33

表 7-6 B ₃ -C 层判断矩阵、权重及一致性检验					
B ₃	C ₇	C ₈	C ₉	权重w _{C_i}	一致性检验
C ₇	1	1/2	1/2	0.1958	λ _{max} = 3.05
C ₈	2	1	2	0.4934	CI=0.025
C ₉	2	1/2	1	0.3108	CR=0.0432≤0.1

所谓一致性指标，是指用来衡量判断矩阵不一致程度的数量指标，是保证判断矩阵可接受的一个必要的方法和过程。本文采用 CR 修正 CI 后得到的 CR 代替 CI 作为一致性检验的指标，当 CR≤0.1 时认为判断矩阵具有可接受的一致性；否则需要对判断矩阵进行修正。定义：

$$CI = (\lambda_{\max}-n)/(n - 1) \quad (6-1)$$

$$CR = CI/RI \quad (6-2)$$

式中，λ_{max} 为判断矩阵的最大特征根，n 为判断矩阵的维数 RI 为随机一致性指标，取值根据矩阵维数而定，本文 n=3，RI=0.58。计算得到 CI 和 CR 的值由通用表可知均通过一致性检验，相应的 W 均可作为其权向量。

对以上层次单排序后，还需进行因子权重合成，即计算最底层各元素相对于最高层的权重值，结果见表 6-7

表 7-7 最底层各元素相对于最高层的权重值					
SM	1	B ₁	0.3275	C ₁	0.1802
				C ₂	0.0787
				C ₃	0.0685
		B ₂	0.4126	C ₄	0.275
				C ₅	0.1375
				C ₇	0.0509
		B ₃	0.2599	C ₈	0.1282
				C ₉	0.0808

3) 沙漠化程度预测指数

将因子强度以及因子权重系数代入 SM 模型，通过调节系数，调节 SM 模型的界限，使之适合于模型测算位置，由于缺少历年荒漠化程度划分的临界值进行界定，修正因子考虑使用初始值 1，将因子权值及因子强度带入到模型SM = ∑_{i=1}ⁿ(Q_i · W_{C_i})=0.7207，对比文献中的表 3-2，得出结论锡林郭勒地区的监测点的沙漠化程度属于重度沙漠化。

表 7-8 沙漠化程度及沙漠化程度指数划分标准					
沙漠化程度	非沙漠化	轻度沙漠化	中度沙漠化	中度沙漠化	极重度沙漠化
沙漠化指标	[0, 0.20]	(0.20, 0.40]	(0.40, 0.60]	(0.60, 0.80]	(0.80, 1.00]

7.2.2 土壤板结化定义

1) 板结化程度定义

板结化程度和土壤有机物、土壤湿度、土壤的容重因素之间的关系，满足土壤湿度 W 越少，容重 C 越大，有机物含量O越低，土壤板结化程度B越严重的规律。可以采用与上一问类似的模型结构，即

$$B = a_1 \cdot W + a_2 \cdot C + a_3 \cdot O \quad (6-2)$$

其中 a_i 是每个因素的权重，确定板结化程度模型也就变成了确定权重 a_i ，最终得到的土壤板结化程度 B 模型，

2) 熵权法原理

根据信息论基本原理的解释，信息是系统有序度的度量，熵是系统扰动程度的度量。根据信息熵的定义，给定指标，熵值可以用来评价给定指标的方差程度。如果所有指标值都相同，则该指标对整体评价没有影响。因此，可以利用信息熵工具计算各个指标的权重，为综合评价多个指标提供依据，本题中因子数量仅为 3，不好通过主观判断来计算因子的权重系数，采用熵权法确定各指标因子的权重系数。

3) 权重系数计算

从附件中，得到土壤湿度、土壤有机碳含量数据，以及部分参考容重数据，其中土壤有机碳含量以及土壤湿度的数据仅为最近偶数年份的数据，通过拉格朗日插值法将土壤湿度的数据对应含碳量的数据情况进行插值补全奇数年份的数据，资料显示土壤中有有机物含量大约是有机碳含量的 1.724 倍。此外，通过网络资料补充获取部分锡林郭勒地区土壤容重后，进行插值补全，得到如下数据表 6-8：

表 7-9 插值补全结果

年份	土壤湿度 g/cm^2	容重	有机物含量
2015	14.18	1.31	23.97
2016	15.82	1.40	19.38
2017	17.61	1.29	17.17
2018	16.61	1.32	18.67
2019	15.47	1.33	18.22
2020	14.22	1.23	16.74

x_1, x_2, \dots, x_k 其中 $x_i = \{x_1, x_2, \dots, x_n\}$ ，进行数据标准化。假设给定了 k 个数据。假设对各指标数据标准化后的值为 Y_1, Y_2, \dots, Y_K ，那么 $Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)}$ 。得到标准化数据表：

表 7-10 标准化数据表

年份	土壤湿度 g/cm^2	容重	有机物含量
2015	0	0.47	1
2016	0.47	1	0.36
2017	1	0.35	0.05
2018	0.52	0.52	0.26
2019	0.37	0.58	0.2
2020	0.01	0	0

求取各个指标的信息熵，根据信息论中信息熵的定义，一组数据的信息熵

$$E_j = -\frac{1}{\ln n} \sum_{i=1}^n p_{ij} \ln p_{ij} \quad (6-3)$$

其中， $p_{ij} = \frac{Y_{ij}}{\sum_{i=1}^n Y_n}$ ，如果， $p_{ij} = 0$ ，则定义， $\lim_{p_{ij} \rightarrow 0} p_{ij} \ln p_{ij} = 0$ 。首先计算 p_{ij} 。

表 7-11 计算 p_{ij} 后结果

年份	土壤湿度 g/cm^2	容重	有机物含量
2015	0	0.1	0.35
2016	0.13	0.22	0.12
2017	0.28	0.07	0.01
2018	0.14	0.11	0.09
2019	0.1	0.13	0.07
2020	0	0	0

通过 p_{ij} 计算信息熵 E_j 。

表 7-12 计算信息熵 E_j 结果

指标	土壤湿度 g/cm^2	容重	有机物含量
E_j	0.62	0.69	0.58

最后通过信息熵计算各指标的权重： $W_i = \frac{1-E_i}{k-\sum E_i}$ ($i = 1, 2, \dots, k$)。结果如下。

表 7-13 最终计算结果

指标	土壤湿度 g/cm^2	容重	有机物含量
a_i	0.34	0.28	0.38

4) 土壤板结化定义

最后结合题目中“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”相关关系。土壤板结化程度 B 与容重 C 呈正相关，与土壤湿度 W 、有机物含量 O 呈负相关。最终土壤板结化定义：

$$B = -0.34 \cdot W + 0.28 \cdot C - 0.38 \cdot O \quad (6-5)$$

7.2.3 模型建立

1) 目标规划

由于问题三中得到的模型建立在放牧方式为划区轮牧的基础上，所以这里的放牧策略中放牧方式也为划区轮牧，剩下的自变量只剩放牧强度。题目简化为在划区轮牧的情况下，求最佳放牧强度，使得沙漠化程度指数和板结化程度最小。

这是一个多目标优化问题，优化目标函数有两个，分别是沙漠化程度指数 $SM = \sum_{i=1}^n (Q_i \cdot W_{Ci})$ 、土壤板结化 $B = -0.34 \cdot W + 0.28 \cdot C - 0.38 \cdot O$ 。约束条件为放牧强度 S 影响 SM 和 B 的模型关系，即放牧强度通过影响单位面积内牲畜的数量、人均家庭纯收入、植被盖度、地表水资源量来决定沙漠化程度，同时也可以通过影响土壤湿度、土壤容重、土壤中有机物含量来决定土壤板结化程度。如下图所示。

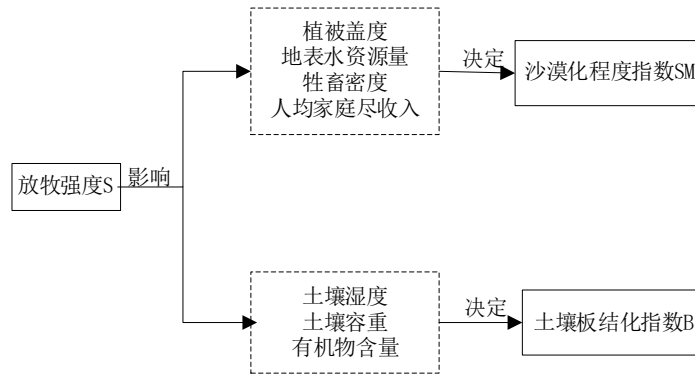


图 7-2 土壤因子影响关系

多目标优化问题与单目标优化问题有很大差异。当只有一个目标函数时，人们寻找最好的解，这个解优于其他所有解，通常是全局最大或最小，即全局最优解。而当存在多个目标时，由于目标之间存在冲突无法比较，所以很难找到一个解使得所有的目标函数同时最优，也就是说，一个解可能对于某个目标函数是最好的，但对于其他的目标函数却不是最好的，甚至是最差的。因此，对于多目标优化问题，通常存在一个解集，这些解之间就全体目标函数而言是无法比较优劣的，其特点是：无法在改进任何目标函数的同时不削弱至少一个其他目标函数。这种解称作 Pareto 最优解，针对这种情况，我们采用非支配排序遗传算法优化求解。

2) 遗传算法

遗传算法，Genetic Algorithm，是一种解决问题的方法，它模拟大自然中种群在选择压力下的演化，从而得到一个问题的近似解。遗传算法遵循适者生存、优胜劣汰的原则，是一类借鉴生物界自然选择和自然遗传机制的随机化搜索算法。遗传算法模拟一个人工种群的进化过程，通过选择(Selection)、交叉(Crossover)以及变异(Mutation)等方法，在每次迭代中都保留一组候选个体，重复此过程，种群经过若干代进化后，理想情况下适应度达到近似最优的状态。遗传算法自从被提出来，得到了广泛的应用，特别是在函数优化、生产调度、模式识别、神经网络、自适应控制等领域，遗传算法发挥了很大的作用，提高了一些问题求解的效率。

本问题基于遗传算法的优化模型构建流程为：

1、首先采用二进制编码的方式对输入分子描述符数据 x_i 进行编码，随机产生基于二进

制编码的初始化种群，定义适应度函数和终止条件；

2、计算初始化种群中的每个个体的适应度函数值；

3、判断当前种群迭代次数下是否满足终止条件，若满足，则解码输出当前最优解，否则采用交叉、变异的方式对种群进行下一轮的迭代优化，采用轮盘赌法在当前种群的基础上选择进化后的种群，直到种群满足优化的终止条件，停止对分子描述符数据的优化过程，并解码输出近似最优解。

NSGA 与简单的遗传算法的主要区别在于：该算法在选择算子执行之前根据个体之间的支配关系进行了分层。其选择算子、交叉算子和变异算子与简单遗传算法没有区别。

在选择操作执行之前，种群根据个体之间的支配与非支配关系进行排序：

首先，找出该种群中的所有非支配个体，并赋予他们一个共享的虚拟适应度值。得到第一个非支配最优层（第一层的适应度值应该大于所有下面层数的适应度）；然后，忽略这组已分层的个体，对种群中的其它个体继续按照支配与非支配关系进行分层，并赋予它们一个新的虚拟适应度值，该值要小于上一层的值，对剩下的个体重复以上操作，直到种群中的所有个体都被分层。

这样做可以保持优良个体适应度的优势，以获得更多的复制机会，同时也维持了种群的多样性。

具体流程如图所示：

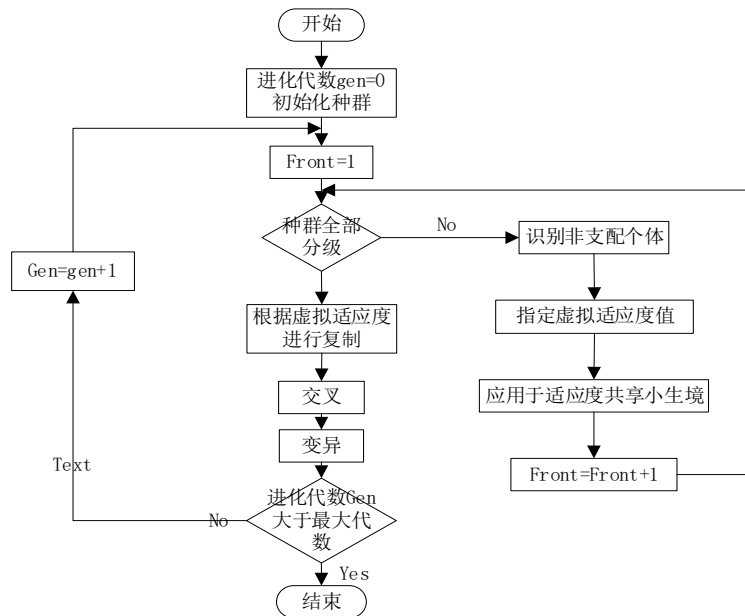


图 7-3 SGA 算法流程图

7.3 结果分析

通过算法得出结果如下图所示：

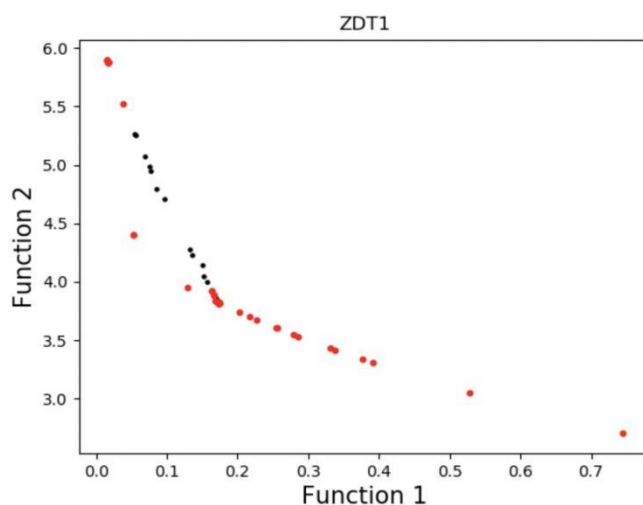


图 7-4 遗传算法结果迭代图

这些点集是经过迭代后得到 Paerot 最优解（无法在改进任何目标函数的同时不削弱至少一个其他目标函数）。按照定义这些解都满足题目要求，这里取沙漠化程度为 0.165, 板结化指数为 3.44 时的放牧强度大约为 3.4 羊/天/公顷。也就是说划区轮牧且中度放牧强度情况下，土壤沙漠化程度指数和土壤板结化程度最小。

8 问题五：不同深度土壤湿度预测

8.1 问题分析

问题五在给定的降水量（300mm、600mm、900 mm 和 1200mm）和保持草原可持续发展情况下对实验草场内放牧羊的数量进行求解，找到最大阈值。将表 14 种的化学因素和表 15 的质量因素通过数据挖掘对数据处理，便于后续分析。通过题目及相关信息可以知道，土壤合适的状态由土壤含氮含量决定，而放牧强度决定了土壤含氮含量；降水影响植被截流量和植物生长速度，从而影响植物量；放牧羊数量消耗植被的速度影响植被需求量。题目要求土壤含氮量不能低于临界值，植物生长量不能低于植被需求量，因此，本节以经济指标为目标函数构造一个目标规划对该问题求解。流程图如下所示。

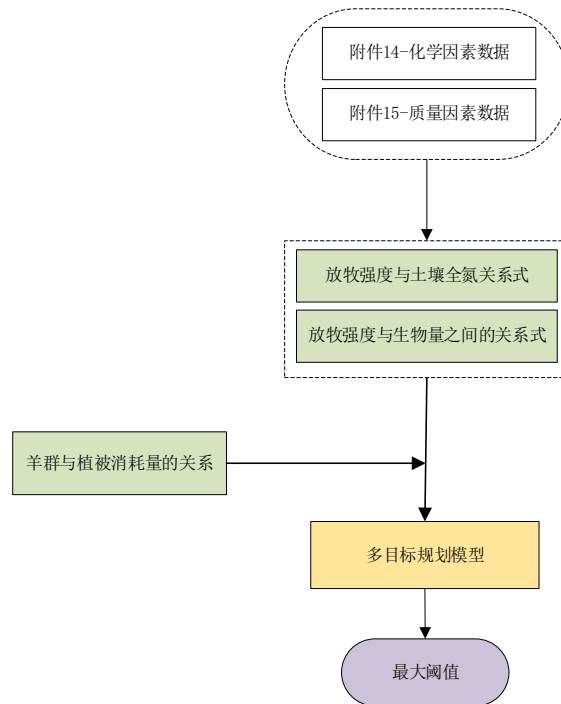


图 8-1 问题五思路流程图

首先，确定以经济指标为目标函数。其次，对附件 14 中的数据处理建立线性回归模型，以此来量化放牧强度与土壤全氮之间的关系；通过对附件 15 中数据处理同样建立线性回归模型，来量化放牧强度与植被量之间的关系，并根据参考文献与附件数据得到放牧强度与植被截流量之间的关系。再次，查找相关资料得到羊群与植被消耗量之间的关系；最后，建立非线性规划模型并使用 python 求解。

8.2 模型建立和求解

8.2.1 目标函数

为方便建模，令羊的单价为 p_1 ，数量为 i ，则通过放牧羊得到的总价值为 I ，

$$I = p_1 * i \quad (1)$$

其中， p_1 值取当前市场均价，取 1200 元。

8.2.2 约束条件

1) 放牧强度与土壤全氮之间的关系

由于土壤全氮随着放牧强度的增加而降低，因此假设放牧强度与全氮含量之间存在如下关系：

令放牧强度为 i ，全氮含量为 y_1 ，则

$$y_1 = f(i) \quad (2)$$

为了量化上述关系 $f(i)$ ，对表 14 中的数据作如下处理：

将相同年份相同强度下的全氮值求平均值，并同统计在表格中，统计结果如表 X 所示。

表 8-1 全氮含量在不同年份下的平均值		
年份	放牧强度	全氮含量平均值
2012	0	1.53
	2	1.69
	4	1.74
	8	1.59
2014	0	1.67
	2	1.70
	4	1.62
	8	1.59
2016	0	1.82
	2	1.85
	4	1.73
	8	1.88
2018	0	2.06
	2	1.95
	4	1.81
	8	1.95
2020	0	2.25
	2	2.07
	4	1.87
	8	2.13

为进一步描述土壤全氮含量与放牧强度之间的关系，以放牧强度为自变量，全氮含量为因变量，使用线性回归求解每年对应的函数表达式，如表 8-3 所示。

表 8-2 不同年份下放牧强度与全氮含量的关系表达式	
年份	放牧强度与全氮之间的关系 $f(i)$
2012	$y = -0.0115x^2 + 0.099x + 1.5322$
2014	$y = -0.0007x^2 - 0.0066x + 1.6825$
2016	$y = 0.0051x^2 - 0.0372x + 1.8429$
2018	$y = 0.0107x^2 - 0.102x + 2.0739$
2020	$y = 0.018x^2 - 0.1619x + 2.2696$

2) 植被生长速度与放牧强度、降水量之间的关系

由问题四可知，植被量满足质量守恒，则植被初始量+生长量-牧羊消耗量=植被剩余量
对上述数量关系量化，且植被剩余量动态过程满足以下关系。

在 t 时刻，
 $t - 1$ 时刻的植被剩余量+ $t - 1$ 时刻的生长增量- $t - 1$ 时刻的牧羊消耗量= t 时刻植被剩余量，

对上式进行变换，得到：
 $y_2 = t$ 时刻植被剩余量- $t - 1$ 时刻植被剩余量；

$$y_2 = f_2(i) \tag{3}$$

为了量化上述 $f_2(i)$ 关系式，这里对附件 15 的数据作如下处理：对相同年份相同放牧强度的总植被量求平均值，并统计在表格中，统计结果如表 8-4 所示。

表 8-3 不同月份下放牧强度与 y_2 关系式

年份	月份	放牧强度	y_2
2016	5	0	673.32
		2	751.5
		4	745.64
		8	621.36
	6	0	454.47
		2	633.23
		4	51.26
		8	248.81
	7	0	500.48
		2	-178.17
		4	502.26
		8	162.21
	8	0	-337.36
		2	15.88
		4	-108.87
		8	-249.83
	9	0	295.29
		2	108.33
		4	-294.1
		8	-250.41
2017	5	0	437.91
		2	463.04
		4	510.24
		8	539.21
	6	0	276.81
		2	608.74
		4	348.33
		8	326.92
	7	0	330.92
		2	526.76
		4	482.23
		8	724.67
	8	0	632.00
		2	485.06
		4	640.36
		8	380.90
	9	0	-723.89
		2	-760.82

		4	-709.27
		8	-700.72
2018	5	0	159.92
		2	114.97
		4	110.47
		8	89.42
	6	0	415.88
		2	436.33
		4	461.26
		8	354.07
	7	0	655.7
		2	909.45
		4	329.89
		8	813.96
	8	0	-12.38
		2	-232.65
		4	159.44
		8	-402.49
	9	0	-1116.58
		2	-783.34
		4	-438.43
		8	-496.72
2019	5	0	211.07
		2	233.35
		4	231.87
		8	194.21
	6	0	739.66
		2	2693.86
		4	794.48
		8	655.92
	7	0	372.74
		2	-3.07
		4	77.12
		8	128.47
	8	0	-29.03
		2	222.08
		4	201.17
		8	-74.05
	9	0	-147.97
		2	-9.95
		4	-155
		8	-276.51
2020	5	0	945.18

		2	748.3
		4	741.23
		8	577.8
	6	0	453.58
		2	846.37
		4	633.45
		8	528.53
	7	0	424.61
		2	622.48
		4	799.29
		8	391.81
	8	0	-81.37
		2	104.21
		4	251.15
		8	131.99
	9	0	-63.61
		2	29.37
		4	-477.82
		8	-186.69

以放牧强度为自变量， y_2 为因变量，使用线性回归求解每年对应的函数表达式，如表8-5所示。

表 8-4 不同月份下放牧强度与植被量的关系表达式

年份	月份	y_2 表达式
2016	5	$y = -6.5402x^2 + 45.192x + 677.13$
	6	$y = 8.4224x^2 - 109.65x + 553.84$
	7	$y = 5.733x^2 - 61.884x + 342.89$
	8	$y = -15.927x^2 + 131.35x - 295.29$
	9	$y = 15.537x^2 - 199.3x + 336.06$
2017	5	$y = -0.981x^2 + 21.109x + 434.32$
	6	$y = -9.3712x^2 + 70.886x + 338.89$
	7	$y = -0.1426x^2 + 45.575x + 359.63$
	8	$y = -4.174x^2 + 8.717x + 591.72$
	9	$y = 0.833x^2 - 2.1315x - 733.71$
2018	5	$y = 1.2665x^2 - 18.338x + 156.28$
	6	$y = -4.2835x^2 + 27.316x + 411.23$
	7	$y = 13.539x^2 - 107.36x + 768.69$
	8	$y = -13.887x^2 + 76.798x - 99.194$
	9	$y = -20.789x^2 + 247.35x - 1137.9$
2019	5	$y = -1.9308x^2 + 13.173x + 212.07$
	6	$y = -49.284x^2 + 314.62x + 1154.8$
	7	$y = 15.039x^2 - 144.13x + 332.46$
	8	$y = -17.45x^2 + 131.32x - 13.139$

	9	$y = -6.469x^2 + 31.058x - 120.21$
2020	5	$y = 3.3081x^2 - 69.12x + 925.58$
	6	$y = -14.974x^2 + 119.44x + 511.89$
	7	$y = -22.46x^2 + 178.75x + 405.58$
	8	$y = -13.271x^2 + 134.19x - 89.473$
	9	$y = 13.218x^2 - 135.25x + 21.1$

牧羊消耗量:

$$y_3 = 1.8 * i \quad (4)$$

生长增量:

$$y_4 = p_2 * IC_{store} \quad (5)$$

经查阅相关资料, p_2 取值为 0.6。另外, 由题中附件可知, 植被截流量 IC_{store} 表达式如下。

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

其中, c_p 为植被覆盖率, IC_{max} 为特定植被的最大截流量 (mm), k 为植被密度校正因子, 与叶面积指数 LAI 有关, 查阅文献【基于水分平衡原理的内蒙典型草原土壤水动态模型研究】本题中 k 取 0.5; 为一个分布式的时变参数, R_{cum} 为累积降雨量 (mm)。 IC_{max} 与 LAI 的关系式为:

$$IC_{max} = 0.935 + 0.498 \cdot LAI - 0.00575 \cdot LAI^2$$

另外, 植被覆盖率数据不完整, 经查阅未找到有效信息, 因此, 通过附件 5 与附件 7 数据, 采用线性回归法拟合出 2016 年-2019 年中 5 月-9 月的植被覆盖率数据, 拟合结果与 LAI 合并后如下表 8-5 所示。

表 8-5 2016 年-2020 年叶面积指数与植被覆盖率

年份	月份	LAI	植被覆盖率
2016	05	0.739	0.199
2016	06	0.827	0.270
2016	07	0.964	0.349
2016	08	0.983	0.357
2016	09	0.847	0.317
2017	05	0.739	0.180
2017	06	0.827	0.203
2017	07	0.964	0.302
2017	08	0.983	0.380
2017	09	0.847	0.298
2018	05	0.739	0.191
2018	06	0.827	0.207
2018	07	0.964	0.569
2018	08	0.983	0.577

2018	09	0.847	0.479
2019	05	0.739	0.207
2019	06	0.827	0.266
2019	07	0.964	0.329
2019	08	0.983	0.325
2019	09	0.847	0.282
2020	05	0.739	0.187
2020	06	0.827	0.258
2020	07	0.964	0.605
2020	08	0.983	0.581
2020	09	0.847	0.467

8.3 结果分析

将附件中数据按照年份和对应的月份代入上式，组成多个表达式。综上，当降雨量为 300、600、900、1200 时，求解目标函数最优值：

$$\begin{cases} \max V = p_1 * i \\ Y_4 - Y_3 = Y_2 \\ i \geq 0 \\ Y_1 \geq M_{\text{临界值}} \end{cases}$$

经查阅资料 $M_{\text{临界值}}$ 取 0.4。综上， $R_{cum}=300$ 、 $R_{cum}=600$ 、 $R_{cum}=900$ 、 $R_{cum}=1200$ 分别代入约束条件进行非线性规划，在 python 运行后得到最大阈值分为 2.7、3.6、5.4、6.6。

9 问题六： 2023 年土地状况分析与预测

9.1 问题描述与分析

首先确定附件 13 的示范牧户放牧策略，通过附件 13 得知示范牧户的放牧方式为划区轮牧，牧户 1、2、4 的放牧强度是轻度放牧，牧户 3 的放牧强度为重度放牧方式，考虑到作为示范牧户，不太可能会采取重度放牧的方式[10]。这里确定附件 13 中示范牧户的放牧策略为在划区轮牧情况下轻度放牧。而问题四的结论为在划区轮牧且中度放牧强度情况下，土壤沙漠化程度指数和土壤板结化程度最小。

由于题目中以及附件中并未规定示范区域具体为哪个区域，这里自己选取区域作为展示。

调用问题 2 中不同放牧策略对土壤湿度影响的预测模型，以及问题 3 中不同放牧策略对土壤化学性质影响的预测模型，针对 G18 区域的划区轮牧且中度放牧和 G16 区域划区轮牧切轻度放牧进行调整后，得出预测结果。

9.2 模型建立与求解

首先寻找 2012 到 2020 年所有的 G18 和 G16 区域土壤肥力（土壤中有机碳、无机碳、全碳、全氮、土壤 C/N 比）的数据以及土壤湿度的数据，将数据输入 LSTM 多变量模型，从 2012 年开始预测下一年的土壤肥力以及土壤湿度的数据，不断重复，并且与真实值对比，求取均方差以查看模型的准确性，进行模型迭代[9]。对 G18 区域进行前期预测迭代，得到图 9-1 和 9-2：

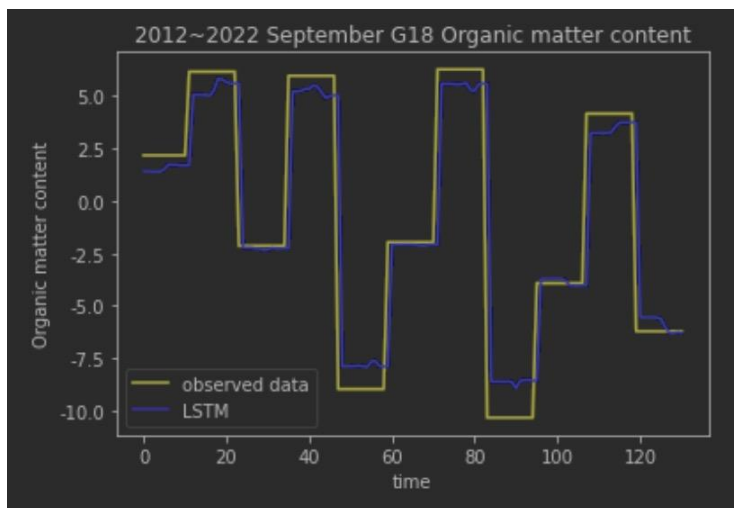


图 9-1 2012-2022 年 G18 区域前期预测迭代结果 1

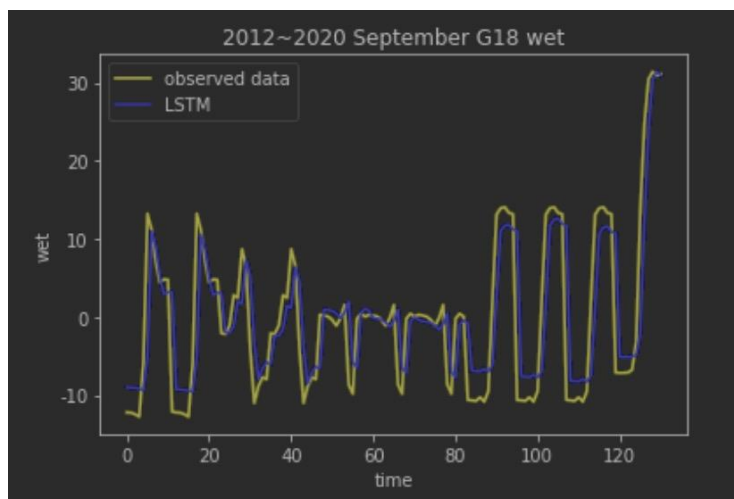


图 9-1 2012-2020 年 G18 区域前期预测迭代结果 2

湿度值预测的方均误差（RMSE）：5.979458 土壤有机物预测的方均误差（RMSE）：2.6822317。

9.3 模型结果

通过模型继续预测 2024 年九月土壤湿度值以及土壤有机物含量的值。为了体现历史年份间数据的不断变化，选择以 X 轴表示土壤湿度，Y 轴表示土壤有机物含量，得到 G18 示范区（划区轮牧+中度放牧）2023 年 9 月土地状态预测图 9-3。

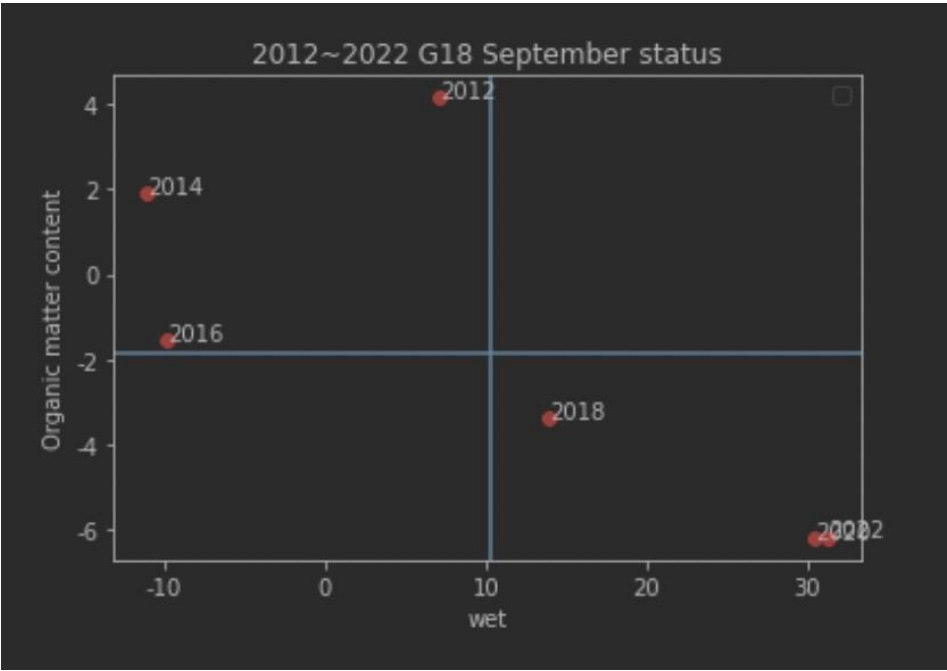


图 9-3 G18 示范区 2023 年 9 月土地状态预测图

从图中可以看出，在划区轮牧和中度放牧强度的情况下，土壤的湿度值在逐渐变大，而土壤中有机物含量也在轻微减少。

同理，对 G16 区域进行同样的处理，得到在划区轮牧和轻度放牧强度情况下的预测图 9-4：

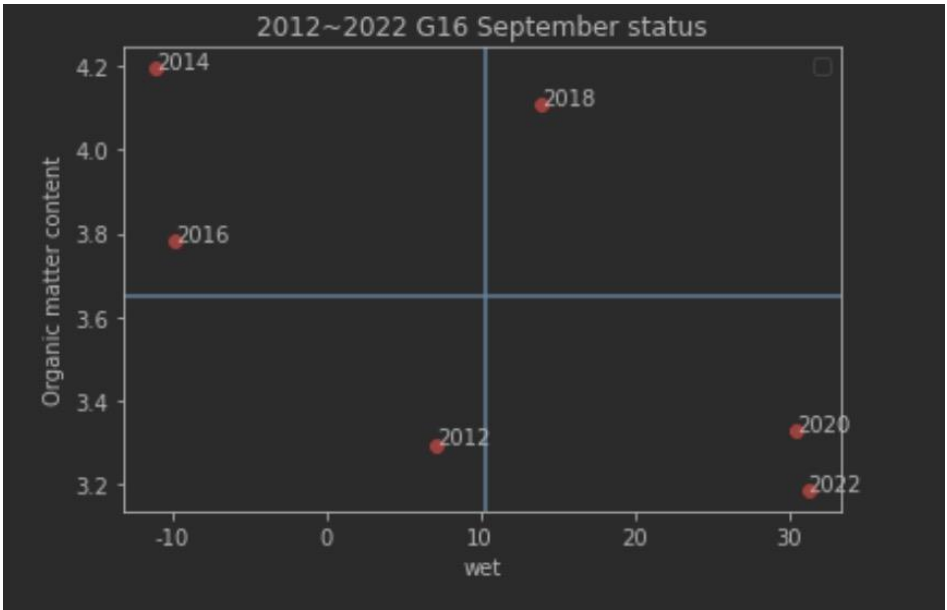


图 9-4 G16 示范区 2023 年 9 月土地状态预测图

从图中可以得出结论，在划区轮牧以及轻度放牧的放牧策略下，土地状况能够维持相对的稳定性。

10 模型的评价

5.1 模型的优点

问题二中通过时序图分析不同深度土壤湿度，采用统计模型预测其 2022 年-2023 年的数据。对于协变量的选择问题，本文通过线性拟合的变量筛选方法选择了 8 个高贡献度的土壤增发、降水等变量，并对缺失数据进行检验与处理，采用 ARIMA 模型，然后再通过预测得到的协变量数据进一步通过 6 种机器学习模型的交叉分析对比。所建立的 ARIMA 模型预测结果良好，验证了模型的有效性和稳定性。

对于问题三的建模及预测，本文对草原土壤的各化学成分进行机理分析，直接确定影响因素，各化学因子的变化规律均有参考依据，确保了数据处理方式的稳定性，进而保证了预测结果的准确性。

问题四设计了较为合理的方法对缺失数据进行预处理，使数据更加全面和准确，通过数学模型和验证相结合的方法，使得验证的结果说明模型符合现实放牧方式与物理性质和化学性质之间的变化规律。

问题五通过数据挖掘方法对数据处理，建立线性回归模型量化放牧强度与土壤全氮、植被量之间的关系，并通过线性回归拟合出植被覆盖率等缺失数据。确定目标函数，构造约束条件，将植被初始量、植被生长量、植被消耗量和植被剩余量通过题目信息建立了关系等式，将问题转化为非线性规划模型求解得到不同降水量下的牧羊阈值。

考虑研究放牧强度与植被生态之间的成本，本文得到的预测模型具有一定参考价值。

5.2 模型的缺点

1. 本文数据分析与预测模型建立采用多个方法对比，综合取优。
2. 得到的模型适应性强且稳定，对放牧强度与植被生态研究有一定的现实参考意义。
3. 为提高模型预测结果，可考虑深度学习进行预测分析。
4. 沙漠化指数及土壤板结化程度模型中，所用到的影响因素数据样本较少且极有可能数据在统计方式上存在误差，在后续的研究中可以根据新的样本进行分析，探讨各变量之间的内在联系。
5. 遗传算法的编程实现比较复杂，首先需要对问题进行编码，找到最优解之后还需要对问题进行解码，另外三个算子的实现也有许多参数，如交叉率和变异率并且这些参数的选择严重影响解的品质，而目前这些参数的选择大部分是依靠经验，求得的最优解可以改进。

参考文献

[1] 闫春霞,赵曼,李浩,张林,张晓严,祁智.放牧和施肥模式对轻度退化天然羊草草原羊草叶钙、铁、锌含量的影响[J].北方农业学报,2022,50(04):83-95.

[2] 张启新,李洁.土壤有机质与全氮相关关系分析[J].硅谷,2010(16):122+162.

[3] 尹晓雷,李先德,林少颖,阳祥,王维奇,张永勋.不同轮作模式下土壤细菌群落特征及其对土壤全碳、全氮与温室气体释放潜力影响[J].环境科学学报,2021,41(12):5161-5173

[4] 侯琼,王英舜,杨泽龙,师桂花.内蒙古典型草原作物系数的动态模拟与确定[J].植物生态学报,2010,34(12):1414-1423.

[5] 侯琼,王英舜,杨泽龙,等.基于水分平衡原理的内蒙古典型草原土壤水动态模型研究[J].干旱地区农业研究,2011,29(05):197-203.

[6] 李丹青,杨超宇.基于 ARIMA 和 LSTM 的瓦斯浓度预测研究[J].新乡学院学报,2022,39(09):45-48.

[7] 宋影.基于熵权法赋权的因子分析水质评价模型[J].水土保持应用技术,2022(05):45-46.

[8] 胡晓蕾,王艺臻.层次分析法(AHP)在示范河湖评价中的应用[J].水资源开发与管理,2020(09):77-81.DOI:10.16616/j.cnki.10-1326/TV.2020.09.15

[9] 于萍,胡卉芪,钱卫宁.基于遗传算法的多目标货物配载研究[J].华东师范大学学报(自然科学版),2021(05):185-198.

[10] 王大双,邹进,刘磊,赖敬明,李琳静.基于 NSGA-II算法的水库生态调度研究[J].中国水运(下半月),2022,22(01):92-93+96.

[11] Neda Mohseni,Yasin Sayyad Salar. Terrain indices control the quality of soil total carbon stock within water erosion-prone environments[J]. Ecohydrology & Hydrobiology,2020(prepublish).

附录

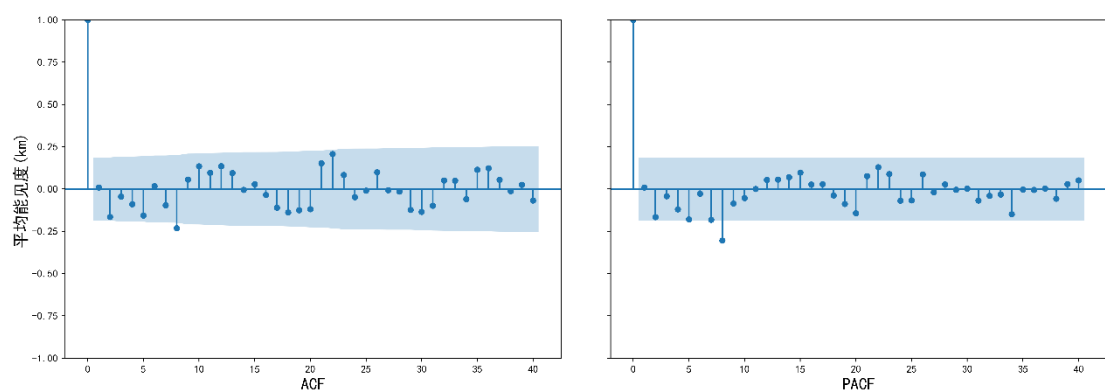
附录 1

附录 1 其余协变量的纯随机检验结果

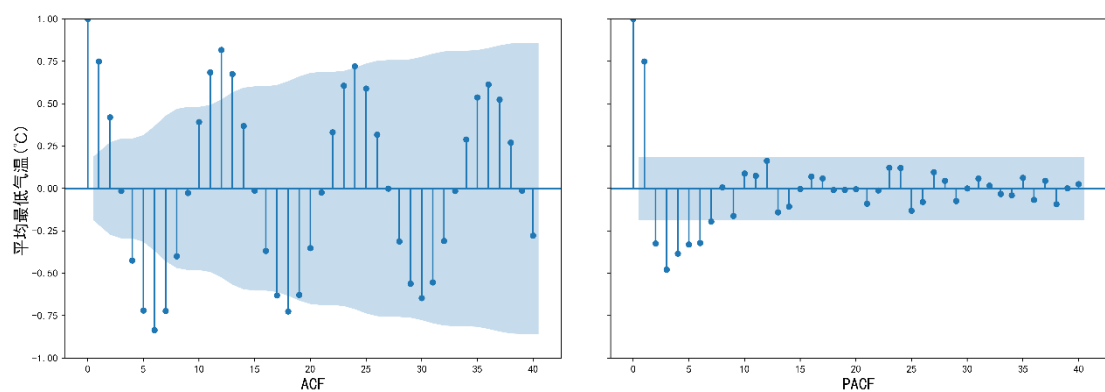
滞后阶数	土壤蒸发量(W/m2)		土壤蒸发量(mm)		平均气温(℃)		平均最高气温(℃)	
	Q	Prob(>Q)	Q	Prob(>Q)	Q	Prob(>Q)	Q	Prob(>Q)
1	63.451	<0.001	0.008	0.927	3.995	0.046	0.051	0.821
2	83.489	<0.001	3.129	0.209	4.289	0.117	0.284	0.868
3	83.515	<0.001	3.357	0.34	8.57	0.036	10.163	0.017
4	104.453	<0.001	4.299	0.367	10.537	0.032	13.697	0.008
5	165.132	<0.001	7.183	0.207	10.54	0.061	13.698	0.018
6	247.794	<0.001	7.215	0.301	12.575	0.05	15.818	0.015
7	310.105	<0.001	8.325	0.305	12.62	0.082	15.82	0.027
8	329.442	<0.001	14.787	0.063	14.188	0.077	21.352	0.006

9	329.526	<0.001	15.171	0.086	21.582	0.01	32.011	<0.001
10	348.51	<0.001	17.4	0.066	21.942	0.015	32.366	<0.001
11	406.759	<0.001	18.563	0.069	23.033	0.017	38.738	<0.001
12	490.612	<0.001	20.853	0.053	50.337	<0.001	66.841	<0.001
13	548.429	<0.001	21.981	0.056	50.591	<0.001	71.743	<0.001
14	565.814	<0.001	21.985	0.079	50.682	<0.001	71.809	<0.001
15	565.833	<0.001	22.092	0.105	64.743	<0.001	84.801	<0.001
16	583.611	<0.001	22.244	0.135	64.749	<0.001	86.447	<0.001
17	636.218	<0.001	23.855	0.123	64.881	<0.001	86.459	<0.001
18	706.581	<0.001	26.418	0.091	68.458	<0.001	88.65	<0.001
19	759.908	<0.001	28.538	0.074	69.161	<0.001	88.799	<0.001
20	776.79	<0.001	30.467	0.063	69.175	<0.001	91.577	<0.001

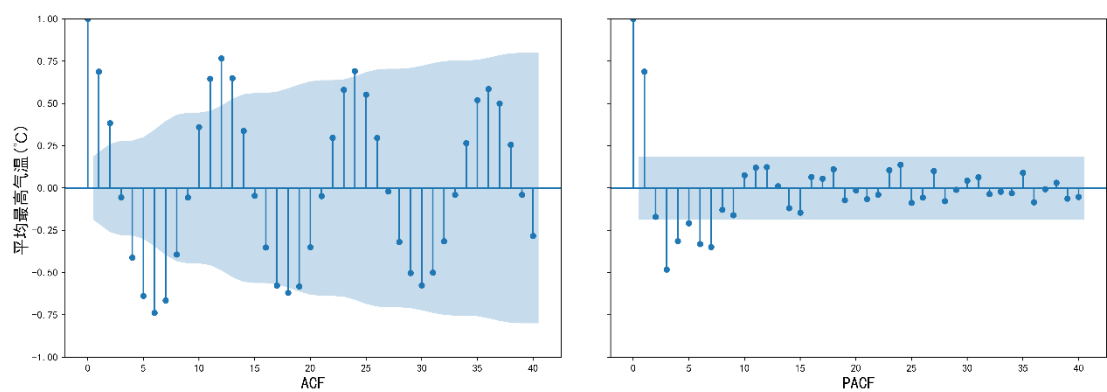
附录 2



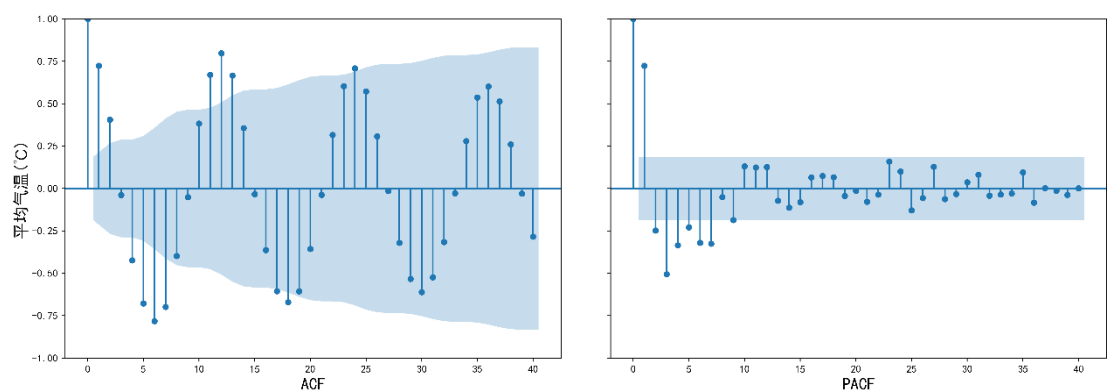
附录 2-1 平均能见度差分序列的自相关图和偏自相关图



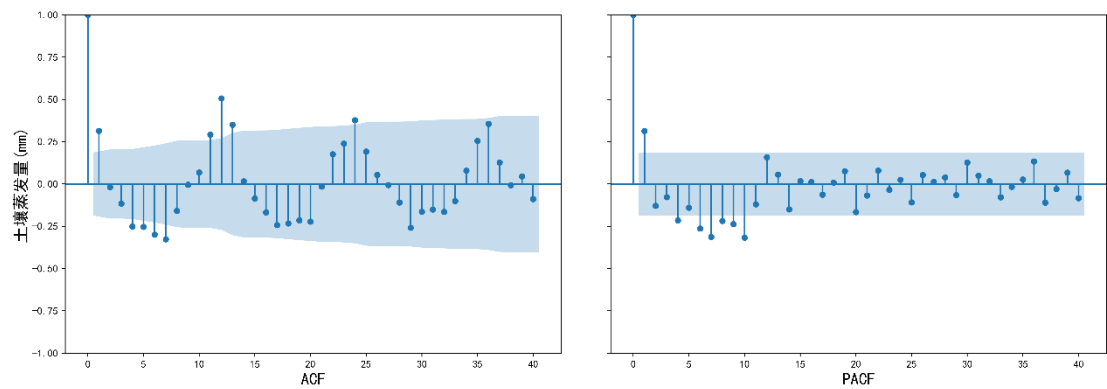
附录 2-2 平均最低气温差分序列的自相关图和偏自相关图



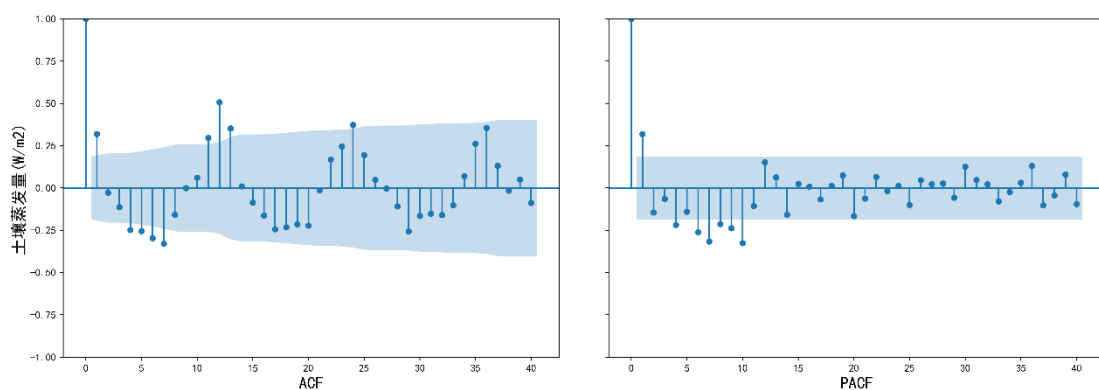
附录 2-3 平均最高气温差分序列的自相关图和偏自相关图



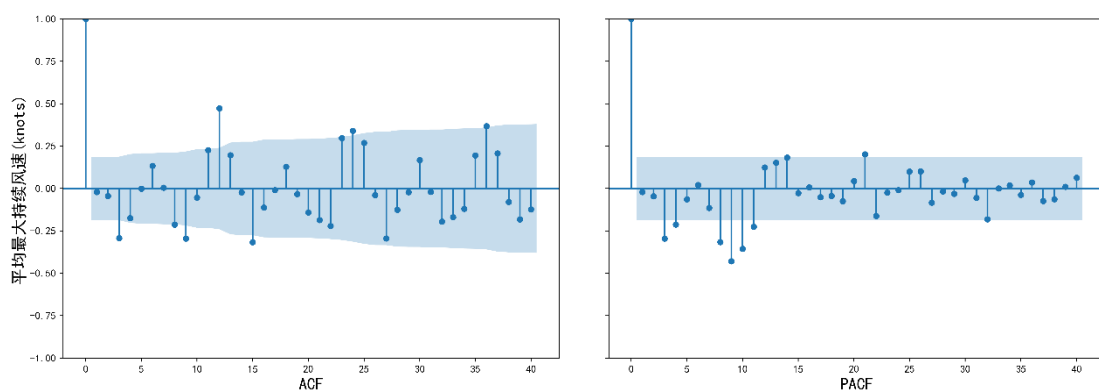
附录 2-4 平均气温差分序列的自相关图和偏自相关图



附录 2-5 土壤蒸发量差分序列的自相关图和偏自相关图



附录 2-6 土壤蒸发量 (W/m^2) 差分序列的自相关图和偏自相关图



附录 2-7 平均最大持续风速差分序列的自相关图和偏自相关图

附录 3

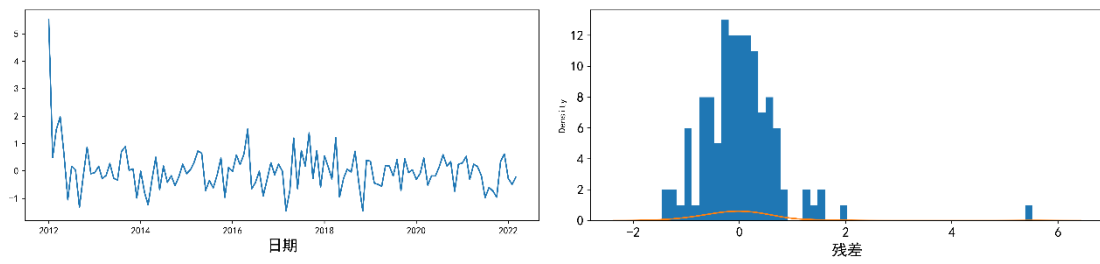
附录 3 8 个协变量的各 ARIMA 模型结果

变量	模型	coef	std err	z	P> z	[0.025	0.975]
土壤蒸发量 (W/m^2)							
	ar.L1	1.7316	0.004	478.830	0.000	1.725	1.739
	ar.L2	-0.9997	0.001	-750.207	0.000	-1.002	-0.997
	ma.L1	-2.1321	58.183	-0.037	0.971	-116.168	111.904
	ma.L2	1.0825	65.829	0.016	0.987	-127.941	130.106
	ma.L3	0.6348	2.890	0.220	0.826	-5.029	6.299
	ma.L4	-0.5853	34.048	-0.017	0.986	-67.319	66.148
	sigma2	16.2850	947.373	0.017	0.986	-1840.531	1873.101
土壤蒸发量 (mm)							
	ar.L1	1.7314	0.004	481.504	0.000	1.724	1.738
	ar.L2	-0.9997	0.001	-684.067	0.000	-1.003	-0.997
	ma.L1	-2.1639	150.064	-0.014	0.988	-296.285	291.957
	ma.L2	1.1496	174.492	0.007	0.995	-340.848	343.147
	ma.L3	0.5919	2.128	0.278	0.781	-3.578	4.762

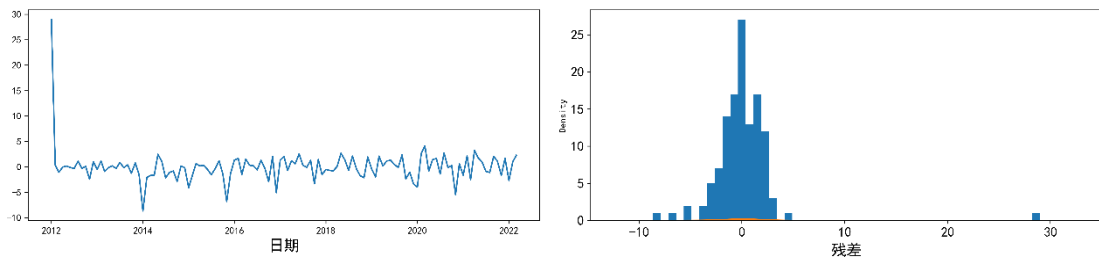
	ma. L4	-0.5775	86.625	-0.007	0.995	-170.359	169.204
	sigma2	18.1920	2728.838	0.007	0.995	-5330.233	5366.617
平均气温(℃)							
	ar. L1	-0.4954	0.104	-4.772	0.000	-0.699	-0.292
	ar. L2	-0.5130	0.097	-5.288	0.000	-0.703	-0.323
	ar. L3	-0.6196	0.090	-6.856	0.000	-0.797	-0.442
	ar. L4	-0.6923	0.101	-6.877	0.000	-0.890	-0.495
	ar. L5	-0.5439	0.085	-6.373	0.000	-0.711	-0.377
	ar. L6	-0.5596	0.079	-7.118	0.000	-0.714	-0.406
	ar. L7	-0.7313	0.089	-8.191	0.000	-0.906	-0.556
	ar. L8	-0.5347	0.095	-5.651	0.000	-0.720	-0.349
	ar. L9	-0.6946	0.080	-8.696	0.000	-0.851	-0.538
	ar. L10	-0.5386	0.091	-5.948	0.000	-0.716	-0.361
	ar. L11	-0.3791	0.088	-4.315	0.000	-0.551	-0.207
	sigma2	4.5709	0.545	8.386	0.000	3.503	5.639
平均最高气温(℃)							
	ar. L1	0.8771	0.119	7.391	0.000	0.645	1.110
	ar. L2	0.4100	0.210	1.948	0.051	-0.003	0.823
	ar. L3	-1.6395	0.150	-10.959	0.000	-1.933	-1.346
	ar. L4	0.5660	0.130	4.351	0.000	0.311	0.821
	ar. L5	0.7113	0.193	3.680	0.000	0.332	1.090
	ar. L6	-0.9340	0.111	-8.390	0.000	-1.152	-0.716
	ma. L1	-1.6109	0.200	-8.043	0.000	-2.003	-1.218
	ma. L2	0.1102	0.443	0.248	0.804	-0.759	0.979
	ma. L3	1.9780	0.413	4.793	0.000	1.169	2.787
	ma. L4	-1.8154	0.255	-7.129	0.000	-2.315	-1.316
	ma. L5	-0.3722	0.465	-0.801	0.423	-1.283	0.539
	ma. L6	1.5412	0.454	3.392	0.001	0.651	2.432
	ma. L7	-0.8281	0.213	-3.879	0.000	-1.246	-0.410
	sigma2	5.6760	1.213	4.679	0.000	3.298	8.054
平均最低气温(℃)							
	ar. L1	-0.5530	0.101	-5.462	0.000	-0.751	-0.355
	ar. L2	-0.6023	0.087	-6.899	0.000	-0.773	-0.431
	ar. L3	-0.6403	0.089	-7.185	0.000	-0.815	-0.466
	ar. L4	-0.6761	0.094	-7.202	0.000	-0.860	-0.492
	ar. L5	-0.5954	0.080	-7.469	0.000	-0.752	-0.439
	ar. L6	-0.6343	0.071	-8.967	0.000	-0.773	-0.496
	ar. L7	-0.7882	0.084	-9.343	0.000	-0.954	-0.623
	ar. L8	-0.5588	0.093	-6.039	0.000	-0.740	-0.377
	ar. L9	-0.7102	0.079	-9.021	0.000	-0.864	-0.556
	ar. L10	-0.5915	0.087	-6.820	0.000	-0.761	-0.421
	ar. L11	-0.4375	0.095	-4.625	0.000	-0.623	-0.252
	sigma2	3.4839	0.420	8.287	0.000	2.660	4.308

平均能见度 (km)							
ar. L1	-0.0292	0.498	-0.058	0.953	-1.006	0.948	
ar. L2	0.6861	0.362	1.898	0.058	-0.022	1.395	
ma. L1	-0.0683	0.770	-0.089	0.929	-1.577	1.440	
ma. L2	-0.9478	0.528	-1.794	0.073	-1.983	0.088	
ma. L3	-0.0922	0.152	-0.608	0.543	-0.389	0.205	
ma. L4	0.1108	0.133	0.831	0.406	-0.151	0.372	
sigma2	3.8836	1.699	2.286	0.022	0.554	7.213	
平均风速 (knots)							
ar. L1	0.1573	0.186	0.846	0.398	-0.207	0.522	
ar. L2	0.4605	0.135	3.404	0.001	0.195	0.726	
ar. L3	-1.2106	0.173	-7.007	0.000	-1.549	-0.872	
ar. L4	0.0006	0.200	0.003	0.998	-0.391	0.392	
ar. L5	0.2098	0.182	1.155	0.248	-0.146	0.566	
ar. L6	-0.5571	0.194	-2.879	0.004	-0.936	-0.178	
ar. L7	-0.1235	0.116	-1.064	0.287	-0.351	0.104	
ar. L8	-0.2831	0.138	-2.050	0.040	-0.554	-0.012	
ar. L9	-0.3070	0.141	-2.181	0.029	-0.583	-0.031	
ma. L1	-1.1597	0.181	-6.390	0.000	-1.515	-0.804	
ma. L2	-0.2012	0.243	-0.829	0.407	-0.677	0.274	
ma. L3	1.8607	0.238	7.828	0.000	1.395	2.327	
ma. L4	-1.4071	0.377	-3.735	0.000	-2.145	-0.669	
ma. L5	-0.3092	0.205	-1.512	0.131	-0.710	0.092	
ma. L6	1.1119	0.231	4.823	0.000	0.660	1.564	
ma. L7	-0.4460	0.176	-2.532	0.011	-0.791	-0.101	
sigma2	0.3234	0.057	5.685	0.000	0.212	0.435	
平均最大持续风速 (knots)							
ar. L1	0.7801	0.105	7.441	0.000	0.575	0.986	
ar. L2	0.6550	0.183	3.573	0.000	0.296	1.014	
ar. L3	-1.9265	0.111	-17.286	0.000	-2.145	-1.708	
ar. L4	0.7212	0.108	6.698	0.000	0.510	0.932	
ar. L5	0.7087	0.179	3.959	0.000	0.358	1.060	
ar. L6	-0.9704	0.102	-9.484	0.000	-1.171	-0.770	
ma. L1	-1.6942	0.176	-9.639	0.000	-2.039	-1.350	
ma. L2	0.0554	0.395	0.140	0.889	-0.719	0.829	
ma. L3	2.4433	0.373	6.549	0.000	1.712	3.175	
ma. L4	-2.3679	0.263	-9.002	0.000	-2.883	-1.852	
ma. L5	-0.0813	0.364	-0.223	0.823	-0.795	0.632	
ma. L6	1.5218	0.341	4.459	0.000	0.853	2.191	
ma. L7	-0.8690	0.167	-5.194	0.000	-1.197	-0.541	
sigma2	0.8149	0.169	4.831	0.000	0.484	1.145	

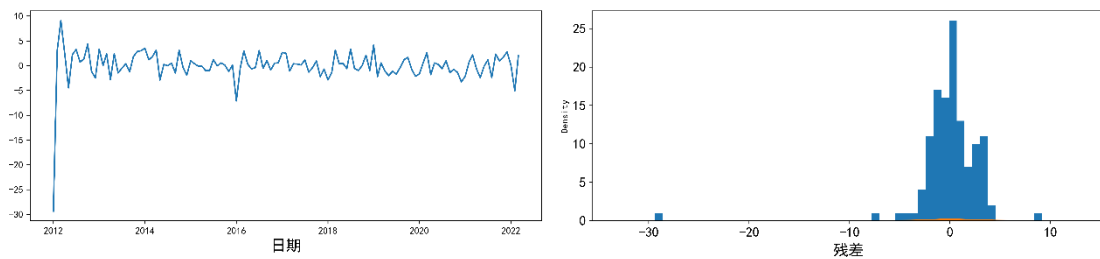
附录 4



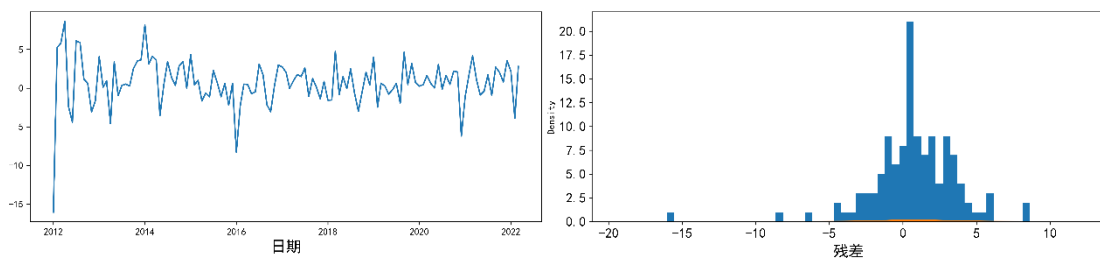
附录 4-1 平均能见度模型残差的序列图和核密度估计图



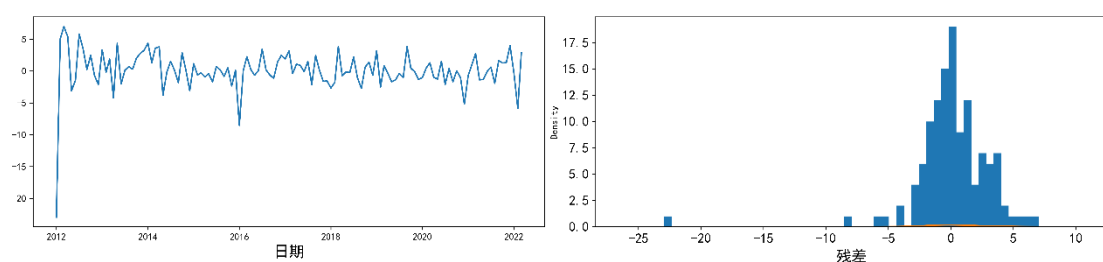
附录 4-2 平均最低气温模型残差的序列图和核密度估计图



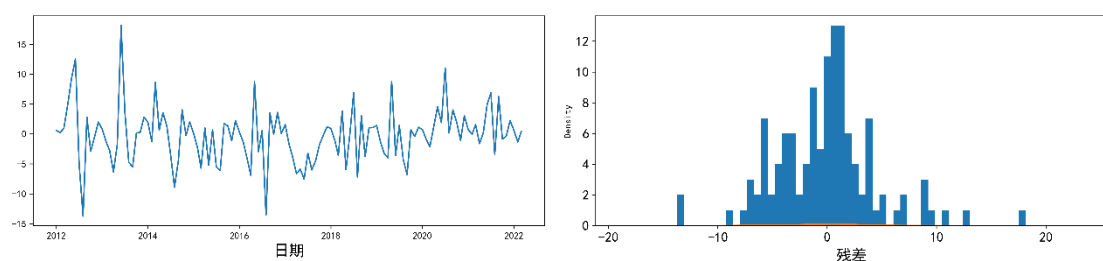
附录 4-3 平均最高气温模型残差的序列图和核密度估计图



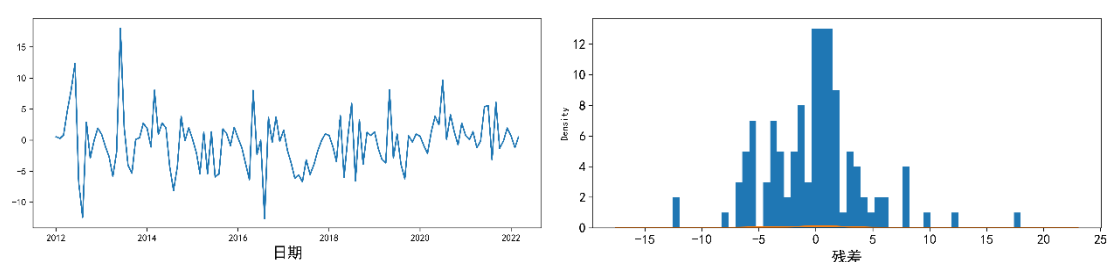
附录 4-4 平均气温模型残差的序列图和核密度估计图



附录 4-5 土壤蒸发量模型残差的序列图和核密度估计图



附录 4-6 土壤蒸发量 (W/m^2) 模型残差的序列图和核密度估计图



附录 4-7 平均最大持续风速模型残差的序列图和核密度估计图

附录 5

附录 5 模型对各协变量残差序列的纯随机检验结果

变量	lag	AC	Q	Prob(>Q)
土壤蒸发量 (W/m^2)	1	0.07	0.612	0.434
土壤蒸发量 (W/m^2)	2	-0.002	0.613	0.736
土壤蒸发量 (W/m^2)	3	-0.096	1.805	0.614
土壤蒸发量 (W/m^2)	4	-0.04	2.014	0.733
土壤蒸发量 (W/m^2)	5	0.051	2.354	0.798
土壤蒸发量 (W/m^2)	6	0.114	4.066	0.668
土壤蒸发量 (W/m^2)	7	0.016	4.099	0.768
土壤蒸发量 (W/m^2)	8	0.034	4.25	0.834
土壤蒸发量 (W/m^2)	9	0.057	4.689	0.861
土壤蒸发量 (W/m^2)	10	-0.114	6.444	0.777
土壤蒸发量 (W/m^2)	11	-0.001	6.444	0.842

土壤蒸发量(W/m2)	12	0.209	12.523	0.405
土壤蒸发量(W/m2)	13	0.148	15.589	0.272
土壤蒸发量(W/m2)	14	-0.053	15.991	0.314
土壤蒸发量(W/m2)	15	-0.06	16.504	0.349
土壤蒸发量(W/m2)	16	-0.043	16.776	0.4
土壤蒸发量(W/m2)	17	0.017	16.82	0.467
土壤蒸发量(W/m2)	18	0.063	17.405	0.495
土壤蒸发量(W/m2)	19	0.046	17.712	0.542
土壤蒸发量(W/m2)	20	-0.077	18.607	0.548
土壤蒸发量(mm)	1	0.081	0.826	0.364
土壤蒸发量(mm)	2	0.013	0.846	0.655
土壤蒸发量(mm)	3	-0.097	2.061	0.56
土壤蒸发量(mm)	4	-0.039	2.256	0.689
土壤蒸发量(mm)	5	0.057	2.685	0.748
土壤蒸发量(mm)	6	0.112	4.334	0.632
土壤蒸发量(mm)	7	0.019	4.38	0.735
土壤蒸发量(mm)	8	0.032	4.517	0.808
土壤蒸发量(mm)	9	0.044	4.782	0.853
土壤蒸发量(mm)	10	-0.106	6.322	0.788
土壤蒸发量(mm)	11	-0.006	6.326	0.851
土壤蒸发量(mm)	12	0.217	12.868	0.379
土壤蒸发量(mm)	13	0.143	15.719	0.265
土壤蒸发量(mm)	14	-0.042	15.964	0.316
土壤蒸发量(mm)	15	-0.059	16.466	0.352
土壤蒸发量(mm)	16	-0.044	16.739	0.403
土壤蒸发量(mm)	17	0.023	16.814	0.467
土壤蒸发量(mm)	18	0.066	17.446	0.493
土壤蒸发量(mm)	19	0.044	17.731	0.54
土壤蒸发量(mm)	20	-0.073	18.527	0.553
平均气温(℃)	1	-0.061	0.468	0.494
平均气温(℃)	2	-0.144	3.089	0.213
平均气温(℃)	3	-0.073	3.775	0.287
平均气温(℃)	4	0.12	5.646	0.227
平均气温(℃)	5	0.071	6.306	0.278
平均气温(℃)	6	-0.077	7.094	0.312
平均气温(℃)	7	-0.033	7.236	0.405
平均气温(℃)	8	0.006	7.242	0.511
平均气温(℃)	9	-0.03	7.362	0.599
平均气温(℃)	10	0.052	7.728	0.655
平均气温(℃)	11	0.016	7.762	0.734
平均气温(℃)	12	-0.153	11.007	0.528
平均气温(℃)	13	0.014	11.034	0.608
平均气温(℃)	14	-0.013	11.059	0.681

平均气温(℃)	15	0.049	11.402	0.724
平均气温(℃)	16	-0.077	12.258	0.726
平均气温(℃)	17	0.076	13.091	0.73
平均气温(℃)	18	0.035	13.267	0.775
平均气温(℃)	19	-0.004	13.27	0.824
平均气温(℃)	20	0.031	13.414	0.859
平均最高气温(℃)	1	0.024	0.07	0.791
平均最高气温(℃)	2	-0.097	1.267	0.531
平均最高气温(℃)	3	-0.148	4.079	0.253
平均最高气温(℃)	4	0.081	4.935	0.294
平均最高气温(℃)	5	0.189	9.581	0.088
平均最高气温(℃)	6	-0.04	9.786	0.134
平均最高气温(℃)	7	-0.144	12.541	0.084
平均最高气温(℃)	8	0.013	12.562	0.128
平均最高气温(℃)	9	0.022	12.629	0.18
平均最高气温(℃)	10	0.127	14.824	0.139
平均最高气温(℃)	11	0.021	14.886	0.188
平均最高气温(℃)	12	-0.149	17.982	0.116
平均最高气温(℃)	13	-0.015	18.014	0.157
平均最高气温(℃)	14	-0.058	18.481	0.186
平均最高气温(℃)	15	-0.007	18.487	0.238
平均最高气温(℃)	16	-0.06	19.012	0.268
平均最高气温(℃)	17	0.012	19.031	0.327
平均最高气温(℃)	18	0.071	19.775	0.346
平均最高气温(℃)	19	-0.014	19.803	0.407
平均最高气温(℃)	20	0.02	19.861	0.467
平均最低气温(℃)	1	-0.044	0.248	0.618
平均最低气温(℃)	2	-0.219	6.364	0.042
平均最低气温(℃)	3	-0.028	6.466	0.091
平均最低气温(℃)	4	0.126	8.507	0.075
平均最低气温(℃)	5	-0.05	8.83	0.116
平均最低气温(℃)	6	-0.056	9.237	0.161
平均最低气温(℃)	7	0.022	9.3	0.232
平均最低气温(℃)	8	-0.029	9.414	0.309
平均最低气温(℃)	9	-0.088	10.455	0.315
平均最低气温(℃)	10	0.04	10.668	0.384
平均最低气温(℃)	11	0.046	10.958	0.447
平均最低气温(℃)	12	-0.111	12.66	0.394
平均最低气温(℃)	13	0.006	12.664	0.474
平均最低气温(℃)	14	-0.037	12.853	0.538
平均最低气温(℃)	15	0.054	13.265	0.582
平均最低气温(℃)	16	-0.037	13.467	0.638
平均最低气温(℃)	17	0.068	14.144	0.657

平均最低气温(°C)	18	0.031	14.285	0.71
平均最低气温(°C)	19	0.016	14.324	0.764
平均最低气温(°C)	20	0.06	14.87	0.784
平均能见度(km)	1	0.007	0.006	0.936
平均能见度(km)	2	-0.02	0.055	0.973
平均能见度(km)	3	0.01	0.069	0.995
平均能见度(km)	4	-0.012	0.087	0.999
平均能见度(km)	5	-0.036	0.255	0.998
平均能见度(km)	6	-0.001	0.255	1
平均能见度(km)	7	0.006	0.26	1
平均能见度(km)	8	-0.048	0.569	1
平均能见度(km)	9	0.035	0.733	1
平均能见度(km)	10	0.01	0.747	1
平均能见度(km)	11	0.064	1.316	1
平均能见度(km)	12	0.06	1.807	1
平均能见度(km)	13	0.082	2.738	0.999
平均能见度(km)	14	0.017	2.777	0.999
平均能见度(km)	15	0.012	2.797	1
平均能见度(km)	16	-0.01	2.812	1
平均能见度(km)	17	-0.056	3.271	1
平均能见度(km)	18	-0.015	3.302	1
平均能见度(km)	19	-0.009	3.315	1
平均能见度(km)	20	-0.03	3.452	1
平均风速(knots)	1	0.084	0.89	0.345
平均风速(knots)	2	0.093	1.978	0.372
平均风速(knots)	3	0.154	4.999	0.172
平均风速(knots)	4	0.013	5.022	0.285
平均风速(knots)	5	-0.127	7.138	0.211
平均风速(knots)	6	-0.034	7.287	0.295
平均风速(knots)	7	-0.012	7.305	0.398
平均风速(knots)	8	-0.091	8.402	0.395
平均风速(knots)	9	0.023	8.473	0.487
平均风速(knots)	10	0.049	8.799	0.551
平均风速(knots)	11	-0.044	9.068	0.616
平均风速(knots)	12	-0.045	9.348	0.673
平均风速(knots)	13	-0.056	9.79	0.711
平均风速(knots)	14	-0.025	9.88	0.771
平均风速(knots)	15	0.007	9.887	0.827
平均风速(knots)	16	0.059	10.382	0.846
平均风速(knots)	17	0.013	10.406	0.886
平均风速(knots)	18	0.012	10.427	0.917
平均风速(knots)	19	0.07	11.148	0.919
平均风速(knots)	20	0.123	13.398	0.86

平均最大持续风速 (knots)	1	0.105	1.4	0.237
平均最大持续风速 (knots)	2	0.044	1.645	0.439
平均最大持续风速 (knots)	3	0.208	7.185	0.066
平均最大持续风速 (knots)	4	0.084	8.103	0.088
平均最大持续风速 (knots)	5	-0.104	9.501	0.091
平均最大持续风速 (knots)	6	0.046	9.776	0.134
平均最大持续风速 (knots)	7	0.116	11.573	0.116
平均最大持续风速 (knots)	8	-0.047	11.865	0.157
平均最大持续风速 (knots)	9	0.119	13.78	0.13
平均最大持续风速 (knots)	10	0.117	15.652	0.11
平均最大持续风速 (knots)	11	0.029	15.766	0.15
平均最大持续风速 (knots)	12	0.037	15.958	0.193
平均最大持续风速 (knots)	13	0.022	16.025	0.248
平均最大持续风速 (knots)	14	0.053	16.428	0.288
平均最大持续风速 (knots)	15	0.01	16.442	0.353
平均最大持续风速 (knots)	16	0.047	16.754	0.402
平均最大持续风速 (knots)	17	-0.052	17.146	0.445
平均最大持续风速 (knots)	18	0.031	17.287	0.503
平均最大持续风速 (knots)	19	0.099	18.722	0.475
平均最大持续风速 (knots)	20	0.105	20.353	0.436
