

不同放牧策略对草原土壤性质的影响研究

——基于机器学习

刘海东, 刘培莹, 常 钰

天津商业大学理学院, 天津

收稿日期: 2023年12月9日; 录用日期: 2024年1月11日; 发布日期: 2024年1月18日

摘 要

草原作为分布最广的重要陆地植被类型之一, 其放牧优化问题的研究可以为政府制定放牧政策提供科学依据。本文以内蒙古锡林郭勒草原为例, 基于不同放牧策略下的监测点数据, 首先运用岭回归模型、BP神经网络和XGBoost等机器学习算法研究不同放牧策略对该区域土壤性质的影响, 其次建立不同深度土壤湿度值的预测模型, 然后构建反映放牧强度对土壤化学性质影响的数学模型, 最后根据历史数据对2022年土壤性质的相关指标展开预测。结果表明随着放牧强度的加大, 植被生物量呈现先增后降的趋势, 同时无牧和重牧也不利于植被生长; 模型预测了锡林郭勒草原监测样地在不同放牧强度下2022年土壤同期有机碳、无机碳、全N、土壤C/N比的值, 经验证拟合效果都较好。为提高模型的实用性, 后续研究还应考虑到部分特征间的相关性。

关键词

放牧策略, 土壤性质, 植被生物量, 机器学习

Study on the Effects of Different Grazing Strategies on Soil Properties of Grassland

—Based on Machine Learning

Haidong Liu, Peiyong Liu, Yu Chang

School of Science, Tianjin University of Commerce, Tianjin

Received: Dec. 9th, 2023; accepted: Jan. 11th, 2024; published: Jan. 18th, 2024

Abstract

As one of the most widely distributed major terrestrial vegetation types, the optimization of grazing in grasslands could provide scientific evidence for governments to formulate grazing policies.

Taking the Xilin Gol Grassland in Inner Mongolia as an example and based on the monitoring data across sites under various grazing strategies, this study first leveraged machine learning algorithms including ridge regression, BP neural networks and XGBoost to investigate the impacts of different grazing strategies on soil properties in the region. Prediction models of soil moisture content across depths were then established. Afterwards, a mathematical model reflecting grazing intensity's influence on soil chemical properties was constructed. Finally, predictions were made with historical data on relevant indicators of soil properties in 2022. Results demonstrated that with the increase of grazing intensity, vegetation biomass first increased and then decreased, while no grazing and overgrazing were also detrimental to vegetation growth. The model predicted the 2022 values of soil organic carbon, inorganic carbon, total nitrogen and C/N ratio across monitoring sites in Xilin Gol Grassland under varied grazing intensities. Verification suggested relatively decent goodness of fit. To improve model applicability, future studies should consider correlations between certain features.

Keywords

Grazing Strategies, Soil Properties, Vegetation Biomass, Machine Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

草原生态系统是陆地生态系统的重要组成部分之一，其植被类型分布广泛，占陆地生态系统约三分之一的净初级生产力。内蒙古锡林郭勒草原位于内蒙古高原锡林河流域，气候类型属于大陆性气候中的温带半干旱草原气候，它不仅是我国重要的畜牧业生产基地，同时也是重要的绿色生态屏障。放牧是草地生态系统主要的利用方式，是影响草地演替和生态系统服务及功能的重要因素。适度放牧能改善草原土壤质量、增加草原生物量和生物多样性，刺激植物群落超补偿生长。但过度放牧会导致草原植被结构破坏，土壤盐碱化程度加重，甚至造成草场退化、土地沙漠化。因此，合理的放牧强度和方式对维持生态系统稳定和可持续性利用具有重要意义[1]。

目前，学术界对不同放牧策略下草原植物群落生物量和土壤性能的变化开展了较多研究。许宏斌等以羊草草甸草原为例研究不同放牧强度对植物群落地上、地下生物量分布的影响，得出轻度或中度放牧利用该区域草地比较合理的结论[2]。张倩等的研究表明不同放牧模式(选择划区轮牧、禁牧、轻度放牧和生长季休牧)显著影响了高寒草甸土壤养分和物种多样性，同时不同放牧模式下土壤全钾含量、含水量、容重和 pH 值是影响群落物种多样性的重要因素[3]。本文以内蒙古锡林郭勒草原为研究对象，研究不同放牧方式和不同放牧强度对该区域的植被群落生物量、土壤理化性质等草地生态系统的影响，探讨温性草原植物群落应对不同放牧强度的响应机制，从草原生产力和多样性维持角度探寻适宜的草地管理模式，以便为锡林郭勒温带草原的放牧管理提供科学参考。

2. 不同放牧策略对土壤湿度和植被生物量的影响

草原放牧通常考虑放牧方式和放牧强度两个因素。其中，放牧方式分为五种：全年连续放牧、禁牧、选择划区轮牧、轻度放牧、生长季休牧，放牧强度分为对照、轻度放牧强度、中度放牧强度、重度放牧强度。首先针对放牧强度对土壤湿度和植被生物量的影响展开分析，然后将放牧方式对土壤湿度和植被生物量的影响反映到比例系数上。针对资料的分析可知，放牧方式中的禁牧与轻度放牧与放牧强度中的

两项含义重复, 因此可不单独考虑; 此外, 放牧策略对草原的影响强调的是时间上的影响, 空间上的影响体现在草原的固有性质中, 所以作为自变量的放牧方式实际上只考虑全年连续放牧以及选择划区轮牧为简化模型, 放牧方式通过模型中的比例系数来体现[4]。

土壤湿度又称为土壤含水量, 影响因素有气象因素、土壤特征、植被状况、人为活动等, 而根据 Woodward 等建立的关于放牧与植物生长的模型, 反映出载畜率对植被生物量的作用[5]; 基于土壤含水量 - 降水量 - 地表蒸发模型可知, 降水与地表蒸散发率会对土壤湿度产生影响; 针对土壤 - 植被 - 大气系统的水平衡基本方程可以看出, 草地的植被直接决定放牧的强度, 而植被的截流量能最好反映植被的生长能力[6] [7]。另外本文选取了土壤鲜重、土壤蒸发量、植被指数(NDVI)、放牧强度和径流量五个因素为自变量, 对土壤湿度展开分析。植被生物量即为植物地上生物量, 其大小可以从植物的干重上间接体现出来, 因此本文将基于科研工作者采集的样品中的干重数量对植被生物量展开研究, 得出不同放牧强度对植被生物量的影响。

2.1. 数据预处理

根据上述思路, 利用 python 软件将数据表进行合并, 得到包含 10 cm 土壤湿度、40 cm 土壤湿度、100 cm 土壤湿度等变量的综合数据集, 并对数据进行相关预处理。

2.1.1. 缺失值处理

本文“鲜重(g)”和“株/丛数”等指标存在少量缺失值, 基于回归插值法, 根据同一指标的现有数据, 对缺失值进行估算。具体思路为构建自变量 $X_i (i=1, 2, \dots, m)$ 与目标变量 Y 的关系, 则第 k 个缺失值的插补估算值可由公式(1)计算, 其中 ξ 指随机因素。

$$y_k = \alpha_0 + \sum_{i=1}^m \alpha_i X_{ik} + \xi_k \quad (1)$$

2.1.2. 异常值、重复值处理

异常数据可能来自数据本身、存储或转换过程, 可删除记录或视为缺失值。本文采取 3σ 原则、绝对中位差法和分位数法对各指标进行处理, 其中 3σ 原则为设各指标在不同时间的取值为 X_{ij} , 算出各指标平均值 \bar{X}_i 及取值的误差 $V_{ij} = X_{ij} - \bar{X}_i$ 和指标标准差 σ_i 。

若某测量值 X_b 的剩余误差 $V_b (1 \leq b \leq n)$ 满足 $|V_b| = |X_b - \bar{X}_i| > 3$, 则认为 X_b 应予修正。由于数据量较大, 合并难免会产生相同的行数据, 以每一行数据为基础, 剔除完全相同的行, 便于后续的数据分析。

2.1.3. 数据归一化

本文为消除不同量纲的影响, 采取归一化方式对数据标准化, 使得原始数据结果映射到 0~1 之间, 即公式(2):

$$x_{ij}(t) = \frac{x_{ij}(t)^* - \min x_{ij}(t)^*}{\max x_{ij}(t)^* - \min x_{ij}(t)^*} \quad (2)$$

此为多指标面板数据, $x_{ij}^*(t)$ 表示第 i 个指标在 t 时刻的第 j 个指标的原始观测值。 $\{x_{ij}^*(t)\}_{n \times p \times d}$ 为标准化结果, $\max x_{ij}(t)^*$ 和 $\min x_{ij}(t)^*$ 分别为指标的最大值和最小值。

2.2. 模型建立与求解

2.2.1. 相关性分析

本文选取土壤鲜重、土壤蒸发量、植被指数(NDVI)、放牧强度和径流量作为自变量, 因变量是四种不同深度的土壤湿度。针对上述变量进行相关性分析, 绘制热力图如图 1 所示。

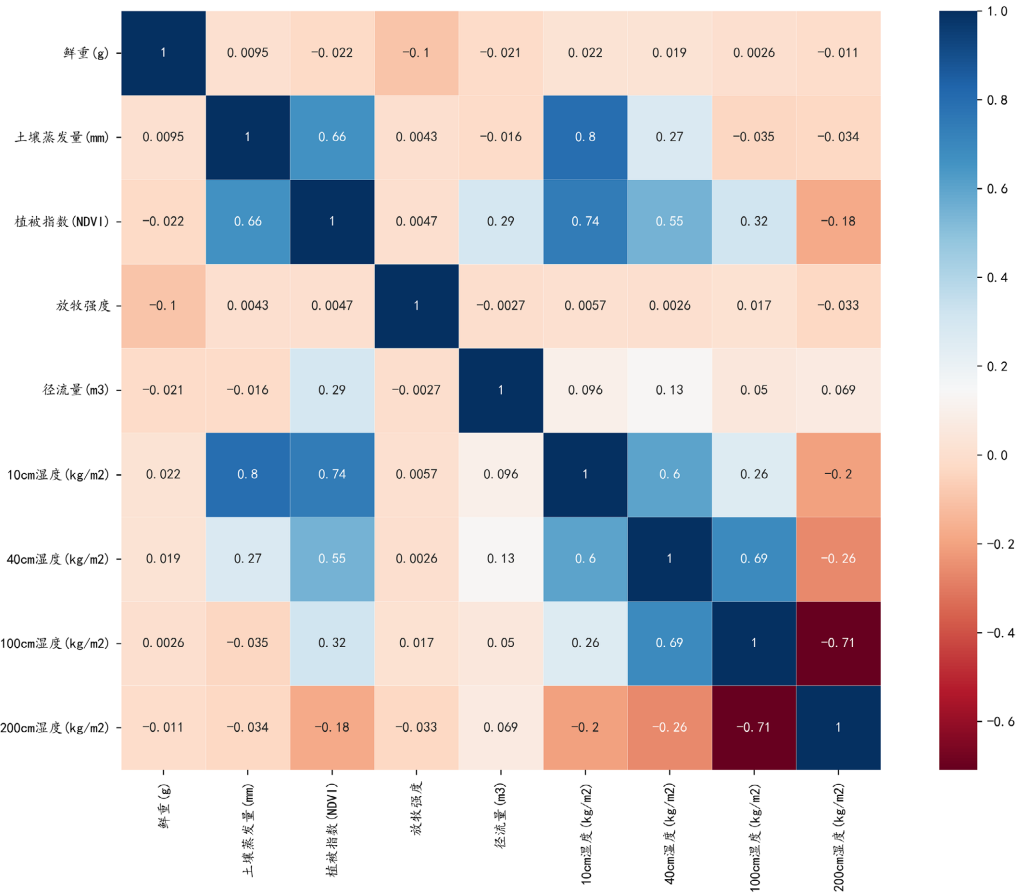


Figure 1. Heat map of correlation coefficients between variables
图 1. 变量的相关系数热力图

由热力图可知，不同深度土壤的湿度存在较强的相关性，所以在进行分析相关因素对土壤湿度的影响时，可以逐步添加变量，比如在分析 40 cm 土壤湿度受各因素的影响时可以将 10 cm 土壤湿度纳入模型，在分析 100 cm 土壤湿度受各因素的影响后可以将 10 cm 和 40 cm 的土壤湿度纳入模型，在分析 200 cm 土壤湿度受各因素的影响后可以将前三种深度土壤湿度纳入模型[8]。此外，由图 1 可知鲜重、土壤蒸发量、植被指数等 5 个变量间的相关性并不高，且根据资料表明各变量对土壤湿度存在影响，因此将这 5 个变量同时纳入模型是切实可行的。

2.2.2. 岭回归模型简介

1962 年，Hoerl 首次提出岭回归，并在 1970 年和 Kennard 进一步对岭回归模型做出了详细讨论，证明自变量间存在多重共线性，岭回归是改良后的普通最小二乘估计，通过对最小二乘估计进行了改进，以达到消除共线性影响的效果[9]。它是一种改良的最小二乘法，实际上是在线性回归的损失函数后加一个 L_2 正则化项，以达到消除共线性影响的效果，能根据回归结果分析各因素对因变量的影响程度和方向。

$$\begin{cases} \min \|y - Xw\|_2^2 \\ \text{s.t. } \|w\|_2^2 \leq C \end{cases} \quad (3)$$

公式(3)中， X 是输入的特征矩阵； y 是输出矩阵； w 是模型的参数向量； C 是大于零的常数。在公式(3)中加入拉格朗日乘子法，将有约束的优化问题转换为公式(4)的无约束的惩罚函数优化问题。

$$\begin{cases} \min_w \|y - Xw\|_2^2 + \lambda \|w\|_2^2 \\ \text{s.t. } \|w\|_2^2 = \sqrt{\sum_i x_i^2} \\ \lambda > 0 \end{cases} \quad (4)$$

岭回归的解为:

$$w = \arg \min_w (\|y - Xw\|_2^2 + \lambda \|w\|_2^2) = (X^T X + \lambda I)^{-1} X^T y$$

2.2.3. 岭回归模型建立与求解

根据是否带变量交叉项, 针对 10 cm 土壤湿度和放牧强度等 5 个因素进行两次岭回归分析, 得出两次回归的判定系数如表 1 所示, 同理对其他三种土壤湿度作相同操作。对比可知, 本次岭回归应包含交叉项, 进一步对回归的各系数大小进行可视化展示如图 2 所示。

Table 1. Model evaluation of ridge regression based on soil moisture at different depths

表 1. 基于不同深度土壤湿度岭回归的模型得分表

	10 cm 土壤湿度	40 cm 土壤湿度	100 cm 土壤湿度	200 cm 土壤湿度
无交叉项	0.717	0.522	0.547	0.663
有交叉项	0.826	0.592	0.314	0.117

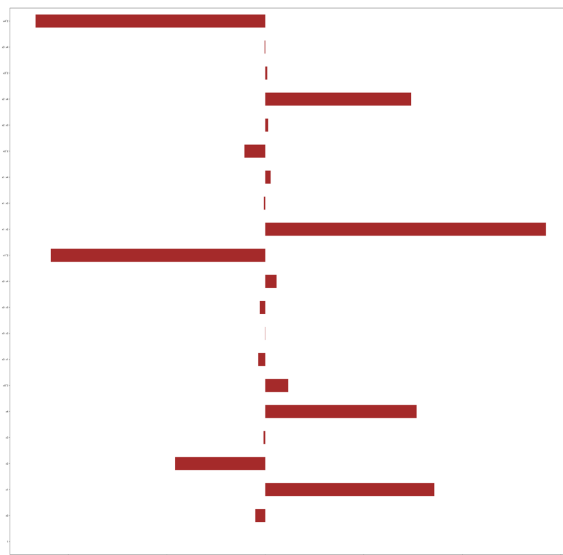


Figure 2. Regression coefficient plot of 10 cm soil moisture ridge regression

图 2. 10 cm 土壤湿度岭回归的回归系数图

根据岭回归的结果可知, 5 种因素对 10 cm 湿度的回归方程为:

$$\begin{aligned} Y = & 1.17X_1^2 - 10.88X_2^2 - 1.06X_3^2 + 0.1X_4^2 - 11.66X_5^2 - 0.51X_1 + 8.57X_2 - 4.8X_3 - 0.09X_4 \\ & + 7.68X_5 + 0.36X_1X_2 - 0.02X_1X_3 - 0.28X_1X_4 + 0.57X_1X_5 + 14.23X_2X_3 - 0.08X_2X_4 \\ & + 0.26X_2X_5 + 0.13X_3X_4 + 7.4X_3X_5 - 0.04X_4X_5 + \varepsilon_t \end{aligned}$$

其中, X_1 表示鲜重(g), X_2 表示土壤蒸发量(mm), X_3 表示植被指数(NDVI), X_4 表示放牧强度, X_5 表示径流量(m^3), ε_t 为滞后项, 用于拟合前期数据对当期数据的影响。

2.3. 结果分析

2.3.1. 基于不同深度土壤湿度的分析

根据岭回归方程各参数可知，植物鲜重与 10 cm 土壤湿度(本段下同，均针对 10 cm 的土壤分析)呈现正相关，植物鲜重越大土壤湿度越大；土壤蒸发量的二次项系数为负，说明土壤蒸发量越大土壤湿度越小；在一定范围内，植被指数越大土壤湿度会相应加大，在范围之外土壤湿度会相应减小；放牧强度和径流量在草原可承受的区间范围里，对土壤湿度的影响也与植被指数所产生的相应规律类似。

同理，针对 40 cm 土壤湿度、100 cm 土壤湿度和 200 cm 土壤湿度展开岭回归分析，通过对比是否考虑交叉项所得的两种模型得分确定最终模型，相应的回归系数图如图 3、图 4 和图 5 所示，相应的回归方程可参考 10 cm 土壤湿度的岭回归结果给出，据此也可得出类似的规律。

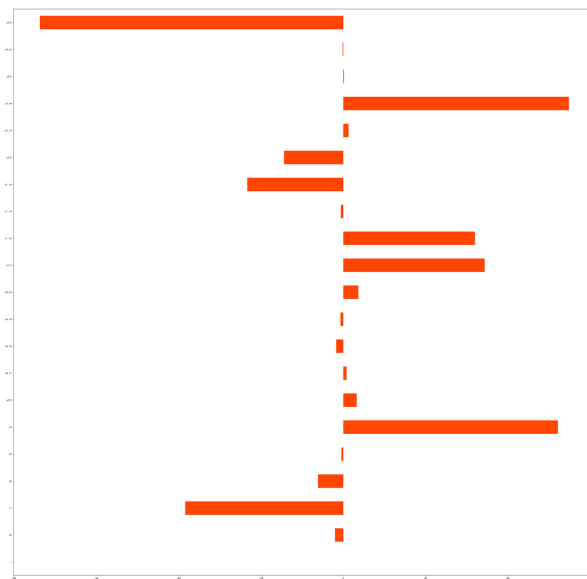


Figure 3. Regression coefficient plot of 40 cm soil moisture ridge regression
图 3. 40 cm 土壤湿度岭回归的回归系数图

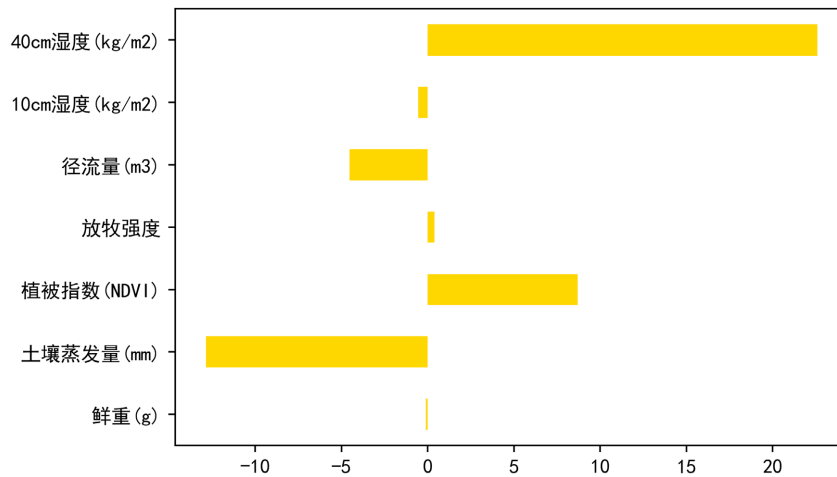


Figure 4. Regression coefficient plot of 100 cm soil moisture ridge regression
图 4. 100 cm 土壤湿度岭回归的回归系数图

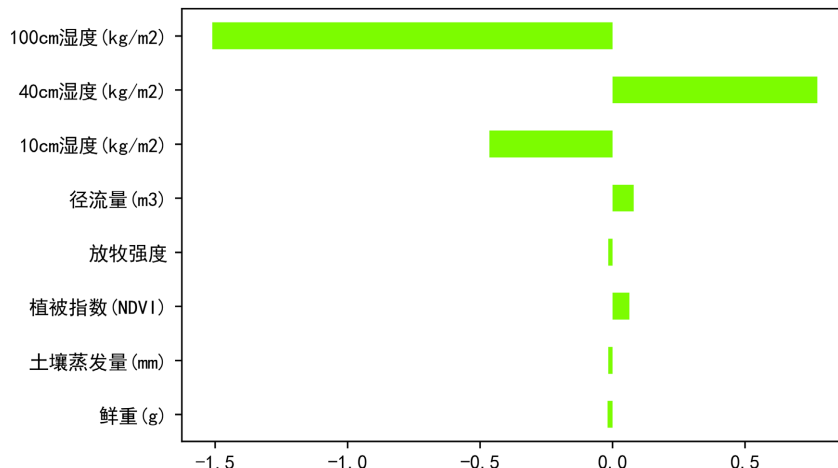


Figure 5. Regression coefficient plot of 200 cm soil moisture ridge regression
图 5. 200 cm 土壤湿度岭回归的回归系数图

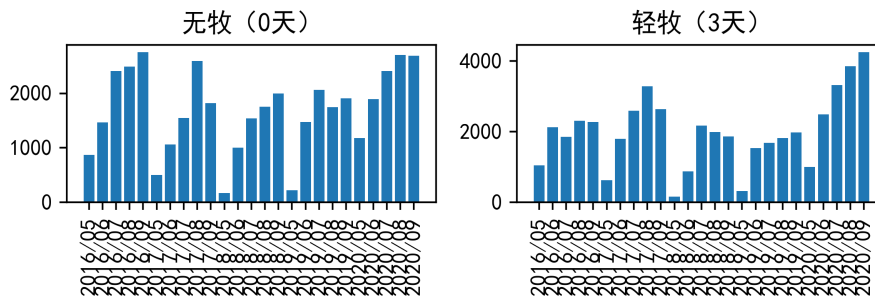
2.3.2. 基于植被生物量的分析

根据预处理过后的监测数据集，将各年份 6 月 1 日的数据代表 5 月的监测数据。首先根据放牧强度将监测结果进行分组，然后以月为频率对数据进行重采样，最后对植物的干重进行求和汇总得到不同月份在不同放牧强度下的植被生物量，对所有月份的植被生物量计算均值，结果如表 2 所示。

Table 2. Average vegetation biomass under different grazing intensities
表 2. 不同放牧强度下植被生物量均值情况

放牧强度	植被生物量(g)均值
无牧(0 天)	1691.21
轻牧(3 天)	1994.43
中牧(6 天)	1820.54
重牧(12 天)	1445.66

由表可知，随着放牧强度的加大，植被生物量呈现先增后降的趋势，即轻牧条件下对植被生长的促进能力最强，重牧则会对植被造成破坏。然后不同放牧强度下，根据各月份的植被生物量绘制柱形图如图 6 所示，可以看出每年的 7、8、9 月份的植被生物量达到峰值，这可能与植物的自身生长规律有关，另外也能清晰发现无牧和重牧条件下的植被生物量相比轻牧和中牧条件下更少，也能间接反映出轻牧条件下对植被生长的促进能力最强，重牧则会对植被造成破坏。



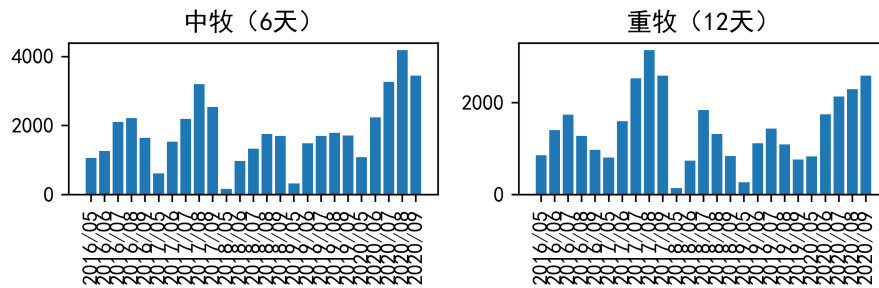


Figure 6. Vegetation biomass under different grazing intensities
图 6. 不同放牧强度下的植被生物量

2.3.3. 基于放牧方式的分析

根据前文的分析可知，在已有与放牧强度有关的分析的基础上，可以通过加入一个比例系数展开讨论，如草原的放牧方式为全年连续放牧时，相当于放牧强度按照一定程度加大；只选择生长季放牧时，等价于放牧强度按照一定程度减小，其对土壤湿度和植被生物量的影响会相应减小。

3. 不同深度土壤湿度的预测模型

通过 2012~2022 年土壤湿度、土壤蒸发量及锡林郭勒盟气候的相关数据，建立数学模型，用来预测 2022 年、2023 年 4 种不同深度下的土壤湿度值，将土壤湿度在不同深度的数据作为预测输出值，土壤蒸发量和草原气象等数据作为输入值。气候变化通过影响区域降水量、蒸发量而引起水资源供需平衡的变化，从而影响区域水资源可持续利用与农牧业可持续发展。

为了对不同深度的土壤湿度值进行预测，首先绘制不同深度土壤湿度值的时间序列图，如图 7 所示。其次，分别构建 BP 神经网络、随机森林和 XGBoost 模型，使用 2012~2021 年数据训练模型，并选择最优模型作为最终预测模型。然后，采用 ARIMA 模型根据以往数据对 2022~2023 年的各自变量值进行预测。考虑用时间序列模型进行自变量未来值的预测，之后输入到预测模型中对最终结果进行回归预测。最后，将 ARIMA 模型预测出来的自变量值输入到训练好的机器学习模型中，对不同土壤深度的湿度值进行回归预测，并得到最终的预测结果表。具体实现路线如图 8 所示。

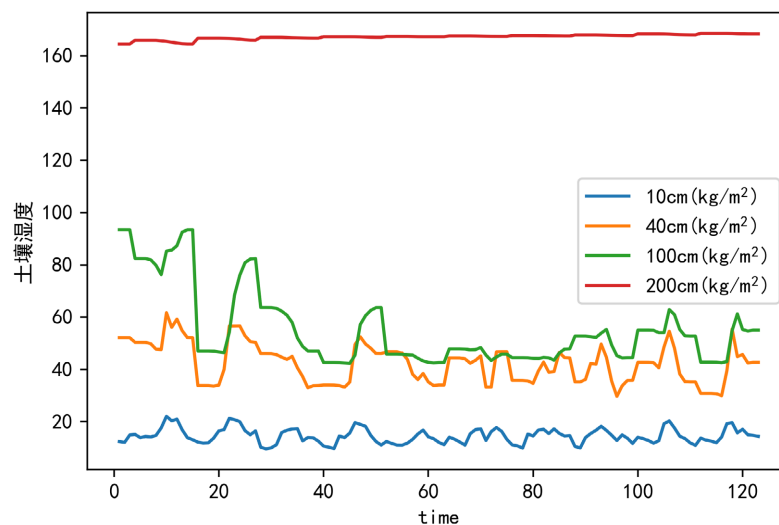


Figure 7. Time series plot of soil moisture at different depths
图 7. 不同深度土壤湿度值时间序列图

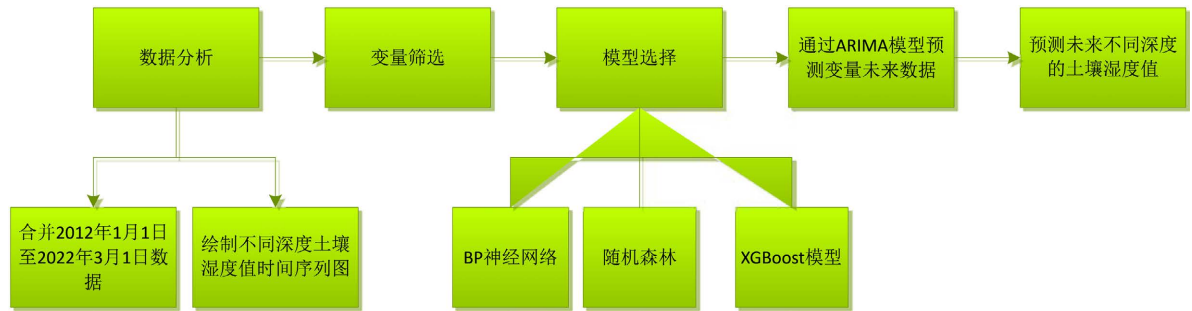


Figure 8. Prediction process of soil moisture at different depths
图 8. 不同深度土壤湿度预测过程

3.1. 模型建立与求解

3.1.1. 变量选择

2012~2022 年土地蒸发量数据中所包含两个指标：土壤蒸发量(W/m^2)和土壤蒸发量(mm)，由于两个指标之间可以相互转换，故选取其一土壤蒸发量(mm)作为自变量。分析锡林格勒气象条件数据，发现平均最大瞬时风速(knots)、最大瞬时风速极值(knots)、积雪深度(mm)和平均温度 $\geq 35^\circ C$ 的天数变量存在缺失值过多或取值全为 0，因此选择平均气温($^\circ C$)、平均最高气温($^\circ C$)、平均最低气温($^\circ C$)、最高气温极值($^\circ C$)、最低气温极值($^\circ C$)、平均气温 $\geq 18^\circ C$ 的天数、平均气温 $\leq 0^\circ C$ 的天数等 20 个草原气候变量为自变量。

考虑土壤深度为 10 cm、40 cm、100 cm 和 200 cm 时，土壤湿度不仅与气候条件、土壤蒸发量有关，不同土壤深度之间也存在相关性。在不同的土壤深度下，水分会有着不同的沉降、上升等情况，某一湿度值会受到其他土壤深度和湿度的影响。所以选择某一土壤湿度值作为预测模型的因变量时，其余三个不同土壤深度的湿度值也看作自变量[10]。综上所述，选取区别于自变量的另一深度土壤湿度值作为因变量，自变量和因变量均为连续性变量。具体选取变量如图 9 所示。



Figure 9. Variable selection plot
图 9. 变量选择图

3.1.2. 算法简介

1) 随机森林模型

随机森林(Random Forests, RF)，又名随机决策森林，其基本思想是通过随机形式建立由很多棵彼此

之间没有任何联系的决策树组成的一个森林。创建好了森林，每当新样本进入随机森林，会使得每颗决策树对其进行判定，最终得出新样本的类别[11]。

随机森林模型的优点如下：(a) 随机森林可以同时处理数值特征和分类，随机森林可以用在解决回归和分类问题方面。(b) 具有较强的抗过拟合能力，利用平均决策树的作用，从而降低过拟合的风险性。(c) 随机森林模型较稳定，因为它只会影响到一颗决策树，所以不容易对整个随机森林产生影响。(d) 随机森林模型鲁棒性强，对数据有很好的包容性。

2) XGBoost 模型

XGBoost 算法是基于梯度提升树的集成学习算法[12]。其基本的组成部分是 CART 树，每一棵新树的建立都是为了尽可能多地减小损失函数值，以此对目标函数进行优化和估计。在合理的参数设置下，需要生成一定数量 CART 树才能达到令人满意的预测准确率。将含 m 个特征、容量为 N 的数据集 $D = \{(x_i, y_i) : x_i \in R^m, y_i \in R, |D| = N\}$ ，所有 CART 树的集合记为 $F = \{f(x) = W_{q(x)}, q : R^m \rightarrow T, w \in R^T\}$ 。其中， q 代表样本映射到相应的叶子结点的决策规则， T 代表一棵树的叶子结点数量， w 代表叶子结点的得分。 f 代表 CART 树，包括树的结构 q 和叶子结点的得分 w 。基于 XGBoost 算法的 y_i 的预测值可以表示为：

$$\hat{y}_i = \varphi(x_i) = \sum_{k=1}^K f_k(x_i) \quad (5)$$

其中， $f_k \in F$ ， K 为 CART 树的数量。XGBoost 算法在每一次模型训练时保留前面 $t-1$ 轮的预测不变，加入新函数 f_t 到模型中， $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$ 为第 i 个样本在第 t 次模型训练时的预测结果。假设基学习器的误差相互独立，XGBoost 算法的学习目标是找到 f_t ，最小化以下目标函数式：

$$L^{(t)} = \sum_{i=1}^T l[y_i, \hat{y}_i^{(t-1)} + f_t(x_i)] + \Omega(f_t) + \text{constant} \quad (6)$$

3) BP 神经网络模型

BP 神经网络是一种按照误差逆传播算法训练的多层前馈网络，由众多的神经元可调的连接权值连接而成，具有大规模并行处理、分布式信息存储、良好的自组织自学习能力等特点，在信息处理、模式识别、智能控制及系统建模等领域有越来越广泛的应用。BP 网络可以学习和贮存大量的输入-输出模式映射关系，而无需事前揭示描述这种映射关系的数学方程。它的学习规则是使用最速下降法，通过反向传播来不断调整网络的权值和阈值，使网络的误差平方和最小[13]。

4) ARIMA 模型

时间序列是按时间顺序的一组数字序列。时间序列的预测分为几个步骤，首先需要对历史数据进行预处理(异常值、缺失值等的处理)，接下来需要对处理后的数据进行分析，根据其所展现的趋势和方向。时间序列预测模型主要分为两大部分，线性预测模型和非线性预测模型。ARIMA 模型比较擅长对线性时间序列进行预测。

ARIMA 模型包含 3 个部分，即自回归(AR)、差分(I)和移动平均模型(MA)。ARIMA 模型的构建步骤可以分为：时间序列平稳性检测、非平稳序列处理、模式识别、模型优化和模型检验几个步骤：(a) 平稳性检测；(b) 非平稳序列处理；(c) 纯随机性检验；(d) 模式识别；(e) 模型优化；(f) 模型检验[14]。

3.1.3. 回归预测模型建立与求解

本文针对不同深度土壤湿度值选用随机森林模型、XGBoost 回归模型和 BP 神经网络模型建立回归预测模型，并结合交叉验证的方法选择效果最优的模型进行最终预测。评价指标均方误差 MSE、 R^2 去评估模型效果，MSE 是预测值与实际值之差平方的期望值，其取值越小，模型准确率越高。 R^2 越接近 1，模型的效果越好。针对四种不同深度的土壤湿度值的最优回归预测模型如表 3 所示。

Table 3. Performance table of optimal regression prediction models for soil moisture at different depths
表 3. 不同深度土壤湿度的最优回归预测模型效果表

评价指标	10 cm 湿度 BP 神经网络回归		40 cm 湿度 XGBoost 回归		100 cm 湿度 XGBoost 回归		200 cm 湿度 XGBoost 回归	
	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²
训练集	0.658	0.921	0.001	1	0.001	1	0	1
测试集	1.248	0.816	7.729	0.9	5.874	0.961	0.101	0.896

3.1.4. ARIMA 模型建立与求解

以变量 100 cm 湿度(kg/m²)为例, 进行 ARIMA 模型的建立与预测。首先, 通过 ADF 单根检验法来检验时间序列是否平稳, 由 ADF 检验表 4 可知, 在差分为 1 阶和 2 阶时, 显著性 P 值为 0.003、0.000, 水平上呈现显著性, 拒绝原假设, 通过平稳性检验(1 阶差分后的时序图如图 10 所示)。

Table 4. ADF test table

表 4. ADF 检验表

变量	差分阶数	t	P	AIC	临界值		
					1%	5%	10%
100 cm 湿度(kg/m ²)	0	0.802	0.992	516.581	-3.492	-2.888	-2.581
	1	-3.792	0.003***	511.828	-3.492	-2.888	-2.581
	2	-6.31	0.000***	520.327	-3.491	-2.888	-2.581

注: ***、**、* 分别代表 1%、5%、10% 的显著性水平。

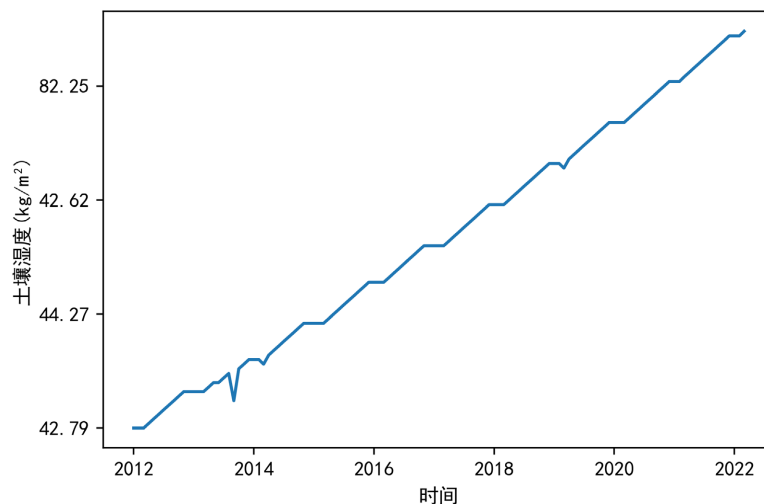


Figure 10. First order differenced time series plot of original data
图 10. 原始数据 1 阶差分后的时序图

然后基于 AIC 信息准则自动寻找最优参数, 确定 100 cm 湿度(kg/m²)预测模型为 ARIMA(0, 1, 1), 基于变量 100 cm 湿度(kg/m²), 从表 5 中 Q 统计量结果分析可知: Q6 在水平上不呈现显著性, 不能拒绝模型的残差为白噪声序列的假设, 同时模型的拟合优度 R² 为 0.966, 效果良好。

Table 5. Test table of ARIMA model (0, 1, 1)
表 5. ARIMA 模型(0, 1, 1)检验表

项	符号	值
样本数量	Df Residuals	120
	N	123
Q 统计量	Q6 (P 值)	0.028 (0.867)
	Q12 (P 值)	0.765 (0.993)
	Q18 (P 值)	9.229 (0.683)
	Q24 (P 值)	16.039 (0.590)
	Q30 (P 值)	21.27 (0.623)
信息准则	AIC	593.21
	BIC	601.622
拟合优度	R ²	0.966

注：***、**、*分别代表 1%、5%、10%的显著性水平。

最后，用 ARIMA(0, 1, 1)模型对 100 cm 湿度(kg/m²)变量值进行预测。同理，可根据 ARIMA 模型得出剩余变量的未来预测值。

3.2. 模型预测结果

通过 ARIMA 模型对 2022 年 4~2023 年 12 月的各自变量值进行预测，并将预测数据代入训练好的最佳模型中，得出在目前放牧策略不变情况下，2022、2023 年不同深度的土壤湿度值，具体预测结果详见表 6 所示：

Table 6. Prediction table of soil moisture at different depths
表 6. 不同深度的土壤湿度预测表

年份	月份	10 cm 湿度(kg/m ²)	40 cm 湿度(kg/m ²)	100 cm 湿度(kg/m ²)	200 cm 湿度(kg/m ²)
2022	04	13.0361	51.5338	47.8925	168.3608
	05	14.5551	51.6532	48.9803	168.4069
	06	16.6938	51.4858	48.3077	167.2532
	07	15.5350	51.4786	49.3574	167.7170
	08	14.9076	51.7093	48.2133	167.8939
	09	14.5804	51.7648	48.2528	167.8789
	10	15.8794	52.1969	48.2086	167.8696
	11	16.6891	52.1148	48.5181	167.8219
	12	16.8967	52.1325	48.5656	167.8225

续表

2023	01	16.2420	52.1301	48.7531	167.8218
	02	15.5761	52.1068	48.4083	167.8262
	03	14.2645	51.9323	48.4110	167.8220
	04	13.1394	51.8273	48.3990	167.8214
	05	12.1608	51.7145	48.3818	167.8810
	06	12.1622	51.7145	48.3272	167.4638
	07	12.5797	51.7135	48.7427	167.4507
	08	13.7917	51.7137	48.6813	167.5320
	09	15.0412	51.8270	48.6090	167.5705
	10	16.4960	52.1460	48.5769	167.5695
	11	17.2449	52.1105	48.6363	167.5113
	12	17.5584	52.1198	48.6363	167.5113

4. 不同放牧策略对土壤化学性质的影响

首先使用不同放牧强度下草原监测数据，建立不同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。选取相关因素作为自变量对土壤有机碳 SOC、土壤无机碳 SIC、土壤全碳 STC 建立岭回归模型，对比考虑交叉项和不考虑交叉项时的模型性能，以确定最终选取哪种方式建模。同时，结合模型各指标系数取值情况，分析不同放牧强度对土壤化学性质的影响程度，并通过比例系数反映放牧方式对土壤化学性质的影响。

其次给出 12 个小区监测点，利用不同放牧强度土壤碳氮监测数据，建立数学模型预测 2022 年 12 个放牧小区在四种不同放牧强度下的土壤化学性质数值。由于考虑不同放牧强度对各种化学性质的影响，相当于单变量回归，同时数据包含 2012~2020 年中偶数年草原监测数据，每个放牧强度设置多个重复，共 132 条数据，数据量较少，所以该问选择 XGBoost 模型对五种化合物在不同放牧情况下进行回归预测。

土壤化学性质数据包含土壤有机碳 SOC、土壤无机碳 SIC、土壤全碳 STC、全 N、土壤 C/N 比，而通过分析可知土壤全碳 STC 和土壤 C/N 比可通过其余 3 个指标计算得来，其计算公式如下所示。故选择土壤有机碳 SOC、土壤无机碳 SIC 和全 N 三项土壤化学性质指标进行分析求解。

$$\text{土壤全碳 STC} = \text{土壤有机碳 SOC} + \text{土壤无机碳 SIC}。$$

$$\text{土壤 C/N} = \text{土壤全碳 STC} / \text{全 N}。$$

4.1. 模型建立与求解

首先，考虑到数据的可获取性和相关文献，本文考虑上述 3 个指标作为因变量，自变量是放牧强度和放牧小区。因变量是连续性变量，自变量通过设置哑变量的方式进行变量转换。

根据是否带变量交叉项，针对土壤有机碳 SOC 与放牧强度、放牧小区哑变量等 15 个因素进行两次岭回归分析，得出两次回归的判定系数如表 7 所示。由下表结果对比可知，在无交叉项时岭回归模型得分均较高，因此建模时不考虑交叉项。此外，进一步对回归的各系数进行可视化展示如图 11 所示。

Table 7. Model evaluation table of ridge regression based on soil moisture at different depths
表 7. 基于不同深度土壤湿度岭回归的模型得分表

	土壤有机碳 SOC	土壤无机碳 SIC	全 N
无交叉项	0.827	0.378	0.79
有交叉项	0.818	0.371	0.774

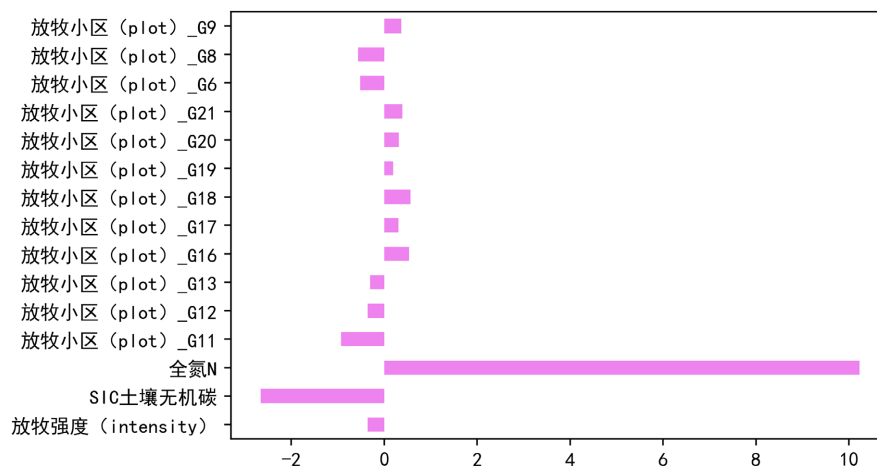


Figure 11. Regression coefficient plot of soil organic carbon (SOC) ridge regression
图 11. 土壤有机碳 SOC 岭回归的回归系数图

由岭回归的结果可知，15 种因素对土壤有机碳 SOC 的回归方程如下：

$$Y_5 = -0.36X_6 - 0.51X_7 - 0.56X_8 + 0.37X_9 - 0.92X_{10} - 0.36X_{11} - 0.3X_{12} + 0.53X_{13} + 0.31X_{14} + 0.56X_{15} + 0.19X_{16} + 0.32X_{17} + 0.39X_{18} + \varepsilon_i$$

其中， Y_5 表示土壤有机碳 SOC， X_6 表示放牧强度(intensity)， $X_7 \sim X_{18}$ 表示 12 个放牧小区(plot)哑变量， ε_i 为滞后项，用于拟合前期数据对当期数据的影响。

4.2. 结果分析

4.2.1. 基于放牧强度的分析

根据方程的系数可知，在其他条件保持不变的情况下，随着放牧强度或 SIC 土壤无机碳含量的增大土壤有机碳 SOC 的含量会降低，而全氮 N 的含量与土壤有机碳 SOC 的含量呈现正相关，且其含量对有机碳含量的影响显著；而对比放牧小区的情况可知，小区 6、小区 8、小区 9、小区 11、小区 12、小区 13 这 6 个小区相比其他小区而言，有机碳含量相对更低一些。

同理，我们对土壤无机碳 SIC、全 N 含量展开岭回归分析，相应的回归系数如图 12 和图 13 所示，相应的回归方程可参考土壤无机碳 SIC 的岭回归结果给出，据此也可得到类似的规律。

4.2.2. 基于放牧方式的分析

根据上述推理可知，在已有与放牧强度有关分析的基础上，可以通过加入比例系数展开放牧方式对土壤化学性质影响的研究得出结论：如果草原的放牧方式为全年连续放牧时，相当于放牧强度按照一定程度加大；只选择划区轮收时，等价于放牧强度按照一定程度减小，其对土壤化学性质的影响会相应减小，其他规律也可参考以上分析相应给出。

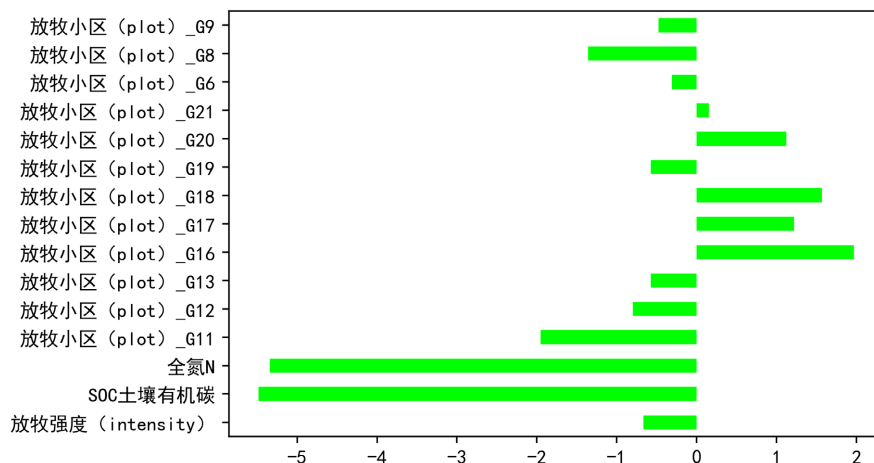


Figure 12. Regression coefficient plot of soil inorganic carbon (SIC) ridge regression
图 12. 土壤无机碳 SIC 岭回归的回归系数图

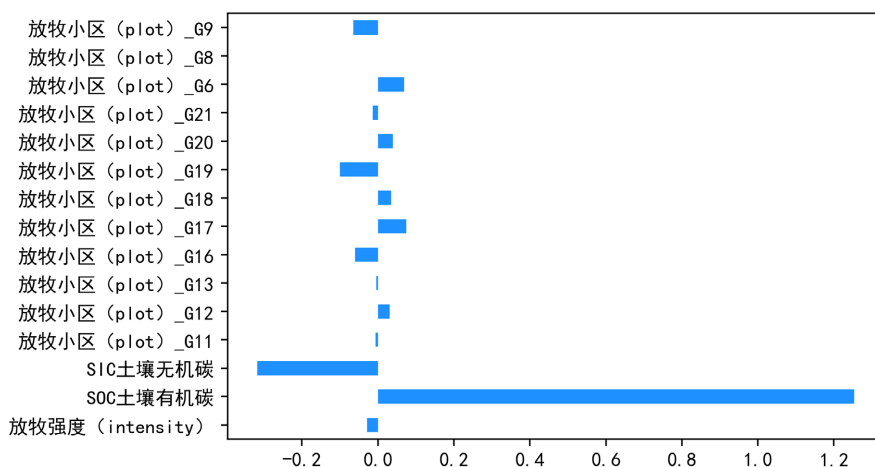


Figure 13. Regression coefficient plot of total N ridge regression
图 13. 全 N 岭回归的回归系数图

4.3. 基于土壤化学性质的 XGBoost 预测模型

4.3.1. 模型建立

选择土壤有机碳、土壤无机碳和全氮建模，根据三种土壤化合物预测结果计算土壤全碳和土壤 C/N 比。此外，数据不存在异常情况，因此不需要进行预处理操作。采用 XGBoost 算法建立单变量回归预测模型，用来预测 2022 年 12 个小区的土壤化合物值。每一个小区包含 2012、2014、2016、2018 和 2020 这五个年份的放牧强度和土壤化合物值。由于每一个小区只对应一个放牧强度，在此不考虑放牧强度，根据 12 个小区进行分类，针对每个小区依次使用 XGBoost 算法构建三个土壤化合物对应的单变量回归预测模型，分别对应对 3 种土壤化合物的预测。将年份作为自变量，每一个化合物作为因变量单独构建回归预测模型。

数据集的特征与模型的选取决定了预测结果准确度的上限，而对模型参数的调整能够帮助无限接近准确度的上限值。为了进一步提高 XGBoost 回归预测的准确率，本文对建立的基于 XGBoost 算法的回归预测模型进行了参数调优，主要选取三个超参数：learning_rate (控制每个弱学习器的权重缩减系数)、max_depth (基学习器的树最大深度，该参数用于避免过拟合)和 n_estimators (梯度增强树 GB 的数量，也

等于最大迭代的次数作为优化目标)。XGBoost 取 $\text{max_depth} = 3$, $\text{n_estimators} = 100$, $\text{learning_rate} = 0.1$, 其余参数为默认参数。

4.3.2. 模型预测结果

根据本文建立的 XGBoost 回归模型, 对 2022 年锡林郭勒草原监测样地在四种放牧强度下的土壤化合物有机碳、无机碳、全 N 进行预测, 计算相应的土壤全碳和土壤 C/N 比, 将结果展示如表 8 所示。

Table 8. The prediction table of soil chemical properties in 2022
表 8. 2022 年土壤化学性质预测表

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	16.4575	6.1341	22.5916	2.0555	10.9909
	G19	16.9931	4.2831	21.2762	2.1493	9.8993
	G21	19.8665	4.0546	23.9211	2.3969	9.9800
LGI	G6	14.3854	3.0000	17.3854	1.9505	8.9135
	G12	15.9067	3.8341	19.7408	1.9155	10.3058
	G18	19.1562	7.2444	26.4006	2.2214	11.8844
MGI	G8	14.2493	1.9530	16.2023	1.9069	8.4966
	G11	14.5753	3.6537	18.2290	1.9328	9.4314
	G16	14.7594	9.8243	24.5837	1.7113	14.3652
HGI	G9	17.1049	3.1593	20.2642	2.1774	9.3065
	G13	16.3682	3.5297	19.8978	2.0691	9.6166
	G20	15.6223	4.7445	20.3667	2.0064	10.1509

5. 结语和未来工作

本文以锡林郭勒草原为例, 根据监测点数据, 采用岭回归分析、BP 神经网络和 XGBoost 回归模型等机器学习方法, 通过 Python、Matlab 以及 SPSS 等软件进行求解, 得出如下结论: 一方面, 随着放牧强度的加大, 植被生物量呈现先增后降的趋势, 同时无牧和重牧也不利于植被生长; 另一方面, 针对 2022 年 4 月~2023 年 12 月不同深度的土壤湿度值进行了预测, 也分析得出锡林郭勒草原监测样地在不同放牧强度下 2022 年土壤同期有机碳、无机碳、全 N、土壤 C/N 比的值, 对实施可持续发展的草原放牧策略具有战略意义。为提高模型的实用性, 后续研究会考虑到部分特征间的相关性。

参考文献

- [1] 凯撒·米吉提. 当前新疆草原畜牧及其可持续发展策略[J]. 中国畜禽种业, 2019, 15(12): 27.
- [2] 许宏斌, 辛晓平, 宝音陶格涛, 等. 放牧对呼伦贝尔羊草草甸草原生物量分布的影响[J]. 草地学报, 2020, 28(3): 768-774.
- [3] 张倩, 杨晶, 姚宝辉, 蔡志远, 王小燕, 苏军虎. 放牧模式对祁连山东缘高寒草甸土壤理化特性和物种多样性的影响[J]. 草原与草坪, 2021, 41(2): 105-112. <https://doi.org/10.13817/j.cnki.cvyyp.2021.02.015>

-
- [4] Wang, X.F., Ma, H.-B., Shen, Y, *et al.* (2019) Effects of Different Rotational Grazing Patterns on Plant Community Characteristics in Desert Steppe Grassland. *Acta Prataculturae Sinica*.
- [5] Woodward, S.J.R., Wake, G.C. and McCall, D.G. (1995) Optimal Grazing of a Multi-Paddock System Using a Discrete Time Model. *Agricultural Systems*, **48**, 119-139. [https://doi.org/10.1016/0308-521X\(94\)00013-H](https://doi.org/10.1016/0308-521X(94)00013-H)
- [6] 闫婷, 黄海广, 闫德仁, 松晓. 浑善达克沙地地下水动态变化及其影响[J]. 内蒙古林业科技, 2021, 47(2): 20-23.
- [7] 宋珊. 锡林郭勒盟地下水动态监测工作现状及对策[J]. 内蒙古水利, 2021(8): 21-22.
- [8] Zhou, Y., Fan, J., Zhang, W., *et al.* (2011) Factors Influencing Altitudinal Patterns of C3 Plant Foliar Carbon Isotope Composition of Grasslands on the Qinghai-Tibet Plateau, China. *Alpine Botany*, **121**, 79-90. <https://doi.org/10.1007/s00035-011-0093-5>
- [9] 朱海龙, 李萍萍. 基于岭回归和 LASSO 回归的安徽省财政收入影响因素分析[J]. 江西理工大学学报, 2022, 43(1): 59-65. <https://doi.org/10.13265/j.cnki.jxlgdxxb.2022.01.009>
- [10] Xu, L., Xu, X., Tang, X., *et al.* (2018) Managed Grassland Alters Soil N Dynamics and N₂O Emissions in Temperate Steppe. *Journal of Environmental Sciences*, **66**, 20-30. <https://doi.org/10.1016/j.jes.2017.04.003>
- [11] 袁红. 数据挖掘模型在股市预测中的应用综述[J]. 中国集体经济, 2017(33): 66-67.
- [12] 黄卿, 谢合亮. 机器学习方法在股指期货预测中的应用研究——基于 BP 神经网络、SVM 和 XGBoost 的比较分析[J]. 数学的实践与认识, 2018, 48(8): 297-307.
- [13] 冯玉芳, 卢厚清, 殷宏, 等. 基于 BP 神经网络的故障诊断模型研究[J]. 计算机工程与应用, 2019, 55(6): 24-30.
- [14] 陈艳玫, 刘子锋, 李贤德, 等. 2015-2050 年中国人口老龄化趋势与老年人口预测[J]. 中国社会医学杂志, 2018, 35(5): 480-483.