

基于ANFIS模型的草原土壤湿度预测

李 帅, 陈 成, 马明明, 王自强

上海理工大学, 机械工程学院, 上海

收稿日期: 2023年2月17日; 录用日期: 2023年3月20日; 发布日期: 2023年3月27日

摘 要

中国是一个资源大国, 草地资源十分丰富, 面积分布也在世界前列。草原生态系统不仅是维护我国生态系统稳定的重要屏障, 同时还为我国的经济发展提供保障。近年来, 快速发展的畜牧业使得草地退化严重, 有的甚至出现了沙化现象。提供科学的草地管理方式迫在眉睫, 故对土壤中的湿度预测对于草原的保护和开发具有重要的意义。本文对往年的统计数据进行分析, 然后通过数学模型对2022、2023年的土壤湿度进行预测。首先对数据进行共线性分析, 对于共线性强的数据采用Lasso回归的方法进行降维。之后用ARIMA时间序列方法对以往年月数据进行预测。最后建立输入(Lasso回归所筛选变量和土壤蒸发量变量)和输出(不同深度土壤湿度) ANFIS模型, 对往年所测数据进行训练且整体数据集训练拟合度均在85%以上, 该模型的准确度较高。最后通过训练好的ANFIS模型预测2022年、2023年不同深度土壤湿度。

关键词

Lasso回归, ARIMA时间序列方法, ANFIS模型, 土壤湿度

Grassland Soil Moisture Prediction Based on ANFIS Model

Shuai Li, Cheng Chen, Mingming Ma, Ziqiang Wang

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 17th, 2023; accepted: Mar. 20th, 2023; published: Mar. 27th, 2023

Abstract

China is a big resource country. Grassland resources are very rich and the area distribution is also among the top in the world. Grassland ecosystem is not only an important barrier for maintaining the stability of Chinese ecosystem, but also provides a guarantee for the economic development of

our country. In recent years, the rapid development of animal husbandry has caused serious degradation of grassland, and some even appear desertification phenomenon. It is urgent to provide scientific grassland management methods, so soil moisture prediction is of great significance for grassland protection and development. This paper analyzes the statistical data of previous years, and then predicts the soil moisture in 2022, 2023 through mathematical models. Firstly, collinearity analysis was carried out on the data, and dimension reduction was carried out on the data with strong collinearity by Lasso regression method. Then the ARIMA time series method is used to predict the data of previous months. Finally, ANFIS models for input (variables screened by Lasso regression and soil evaporation variables) and output (soil moisture at different depths) were established. The data measured in previous years were trained and the overall data set training fit degree was above 85%, indicating a high accuracy of this model. Finally, the trained ANFIS model was used to predict the soil moisture at different depths in 2022 and 2023.

Keywords

Lasso Return, ARIMA Time Series Method, ANFIS Model, Soil Moisture

Copyright © 2023 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

草原生态系统是许多牧民生活的基础，同时也是我国生态安全和食品安全的重要支柱。近年来，随着我国的生活水平的提高，我国居民在食物方面的消费结构发生了很大的变化，对于肉类的需求开始大量增长[1]。其中就以内蒙古为例，草牧业快速发展，牲畜数量极具增加，造成草地退化严重。草地退化严重损害草地的生态功能，大大限制了牧区中的畜牧业生产能力，同时随着全球变暖、水资源短缺等情况下，我国的草地资源和保护面临严峻的挑战。草原中的土壤湿度是影响草地生态过程、生态承载能力以及退化后的重建和恢复能力的关键因素之一[2]。同时还对植物的生长期发育和产量有着直接作用[3]。故对于草原土壤中湿度预测是十分必要的，对于之后放牧策略调整和草原资源保护都有着十分重要的意义。因此本文根据草原历史湿度数据、草原土壤历史蒸发量和草原历史降水等数据，建立模型对保持当前的放牧策略不变的情况下对 2022 年、2023 年不同土壤深度的湿度进行预测(数据来源于锡林郭勒统计年鉴 2012~2021) [4]。

2. 模型建立与求解

本文用到的符号及其含义如表 1 所示：

Table 1. Symbol description

表 1. 符号说明

符号	含义
VIF	方差膨胀系数
ARIMA	移动平均自回归模型
ANFIS	自适应模糊神经推理系统

2.1. Lasso 回归

2.1.1. 共线性分析

首先观察锡林郭勒统计年鉴中的降水数据,发现给定的样本中存在大量为 NAN 值的变量,这些变量无法进行相应分析预测,故全部删去。之后进行多重共线性分析,多重共线性是指在线性回归模型当中两个或者两个以上的变量具有精确的相关关系或者是有着高度相关关系,这样就会导致模型估计不准确更有甚者会导致模型失真。故要对模型的变量进行共线性分析。对处理后的数据进行初步回归并检查其方差膨胀因子 VIF,其方差膨胀因子值远大于 10,根据经验规则:若 $VIF > 10$,则认为回归存在严重的多重共线性。

对于具有严重多重共线性的高维数据进行降维可以采用 Lasso 回归的方法,相比传统的 OLS (ordinary least squares)和岭回归,Lasso (Least absolute shrinkage and selection operator)回归可以将与因变量相关性极小的变量的回归系数压缩至零,从而提供一个稀疏解,用此方法筛选出的变量相互独立且具有代表性。并使用 K 折交叉验证来最小化均方误差。

共线性分析结果见表 2 所示。

Table 2. Collinear analysis results (part)
表 2. 共线性分析结果(部分)

变量	VIF
平均气温	1840.17
平均最低气温	1026.53
平均最高气温	481.13
...	...
单日最大平均气压	2.29
最大能见度	2.14
Mean VIF	189.25

2.1.2. 回归结果

在 Lasso 回归前,由于数据中各变量的量纲不同,需要对数据进行标准化,通过 Matlab 进行标准化操作,然后用 Stata 将处理完的数据不同深度土壤湿度进行 Lasso 回归。

当调整 Lasso 回归中的参数于不同的值时,回归系数的结果也会产生很大的不同。这里可以使用 K 折交叉验证的方法来选择最佳的调整参数,即将样本随机分为 K 份,先将第一个子样本作为验证集,使用剩下的子样本来估计模型并预测第一个子样本,(计算与其他的 MSPE 值。再将第二个子样本作为验证集,类推下去,通过调整参数,使所有子样本的 MSPE 值之和达到最小,从而得到最好的预测能力。

表 3 列出了 Lasso 回归的部分结果,共计筛选出 6 个变量,表中第二列为筛选出变量的 Lasso 回归的系数,其余变量对不同深度土壤湿度的回归系数均为 0。表中第三列数据为仅用筛选出的 6 个变量针对不同深度土壤湿度进行普通最小二乘法回归的系数。

2.2. ARIMA 模型

通过对统计数据进行分析,得出平均最高气温,平均最低气温,最高气温极值,平均气温 $\leq 0^{\circ}\text{C}$ 的天数,平均海平面气压,平均最大持续风速的六个变量对不同深度土壤湿度的回归系数不为 0。

Table 3. Lasso regression results (part)
表 3. Lasso 回归结果(部分)

变量名	Lasso	OLS
平均最高气温	0.4408493	0.4479843
平均最低气温	0.4706232	0.5260698
最高气温极值	0.0467226	0.0254644
...
平均气温 < 0 的天数	-0.0079671	0.0009340
平均海平面气压	-0.0365896	0.0027272

故土壤湿度与平均最高气温, 平均最低气温, 最高气温极值, 平均气温 $\leq 0^{\circ}\text{C}$ 的天数, 平均海平面气压, 平均最大持续风速和土壤蒸发量有关, 建立定性的数学模型如下:

$$Q = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7) \quad (1)$$

其中 X_1 指平均最高气温; X_2 指平均最低气温; X_3 指最高气温极值; X_4 指平均气温 $\leq 0^{\circ}\text{C}$ 的天数; X_5 指平均海平面气压; X_6 指平均最大持续风速; X_7 指土壤蒸发量。

在预测 2022 年、2023 年不同深度土壤湿度时, 需要对 Lasso 回归所筛选变量和土壤蒸发量变量的相应月份数据进行确认, 由于材料中并未给定 2022 年, 2023 年的相关数据, 需要针对以往年月数据进行预测。

往年数据所给定的是以月份为周期, 时间序列分析便是根据过去的变化趋势利用统计学方式预测未来, 通常符合事物发展的规律; 在考虑发展趋势的同时, 更注重周期性变化对具体时间点的影响, 更加准确; 承认随机变量可能对最终结果造成的影响。ARIMA 模型是时间序列预测分析方法之一, 模型将预测对象随时间推移而形成的数据序列视为一个随机序列, 用一定的数学模型来近似描述这个序列, 这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值; 其模型简单, 仅需要内生变量而不需要借助其他外生变量。

2.2.1. ARIMA 建模

通过 ARIMA(p, q, d) 预测筛选变量和土壤蒸发量变量数据[5]。其中, p、d 和 q 分别为自回归项数、时间序列成为平稳序列时所做的差分次数和移动平均数。构建 ARIMA 模型, 基于基期(2012~2022 年)数据预测(2022~2023 年)各个变量相应月份的数据, 从经过 Lasso 回归后(2012~2022 年)的数据(平均最高气温, 平均最低气温, 最高气温极值, 平均气温 $\leq 0^{\circ}\text{C}$ 的天数, 平均海平面气压, 平均最大持续风速, 土壤蒸发量)作为 ARIMA 模型的初始条件, 具体逻辑步骤如下: 首先, 为消除异方差, 统计(2012~2022 年)各月份变量值并取对数, 然后对该时间序列进行平稳性检验, 如果检验不平稳, 可对原序列进行差分, 直至序列平稳。其次, 在以上步骤的基础上建立平稳时间序列的基础模型, 再根据赤池信息准则选择 AIC 值最小的基础模型构建相应的 ARIMA(p, q, d) 模型, 用于拟合(2012~2022 年)各月份变量值。最后, 应用赤池信息准则选择 AIC 值最小的模型 ARIMA(p, q, d), 对(2022~2023 年)各月份变量值进行预测分析。操作流程如图 1。

2.2.2. ARIMA 预测结果

针对 Lasso 回归所筛选变量和土壤蒸发量变量分别进行建模, 对于每个模型, 预测从所请求估算期范围内的最后一个非缺失值之后开始, 并结束于最后一个所有预测变量都有可用的非缺失值的周期, 或者在所请求预测期的结束日期结束, 以较早者为准。

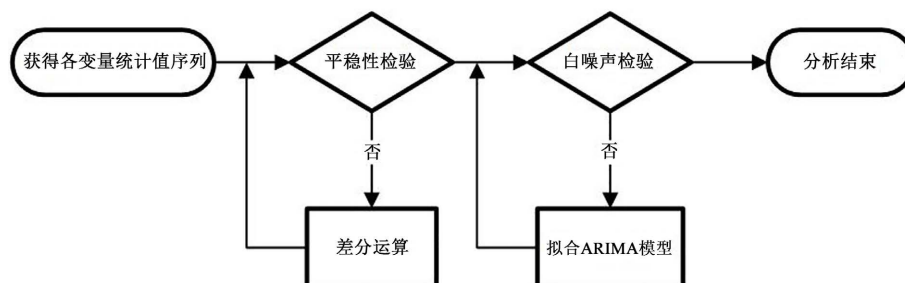
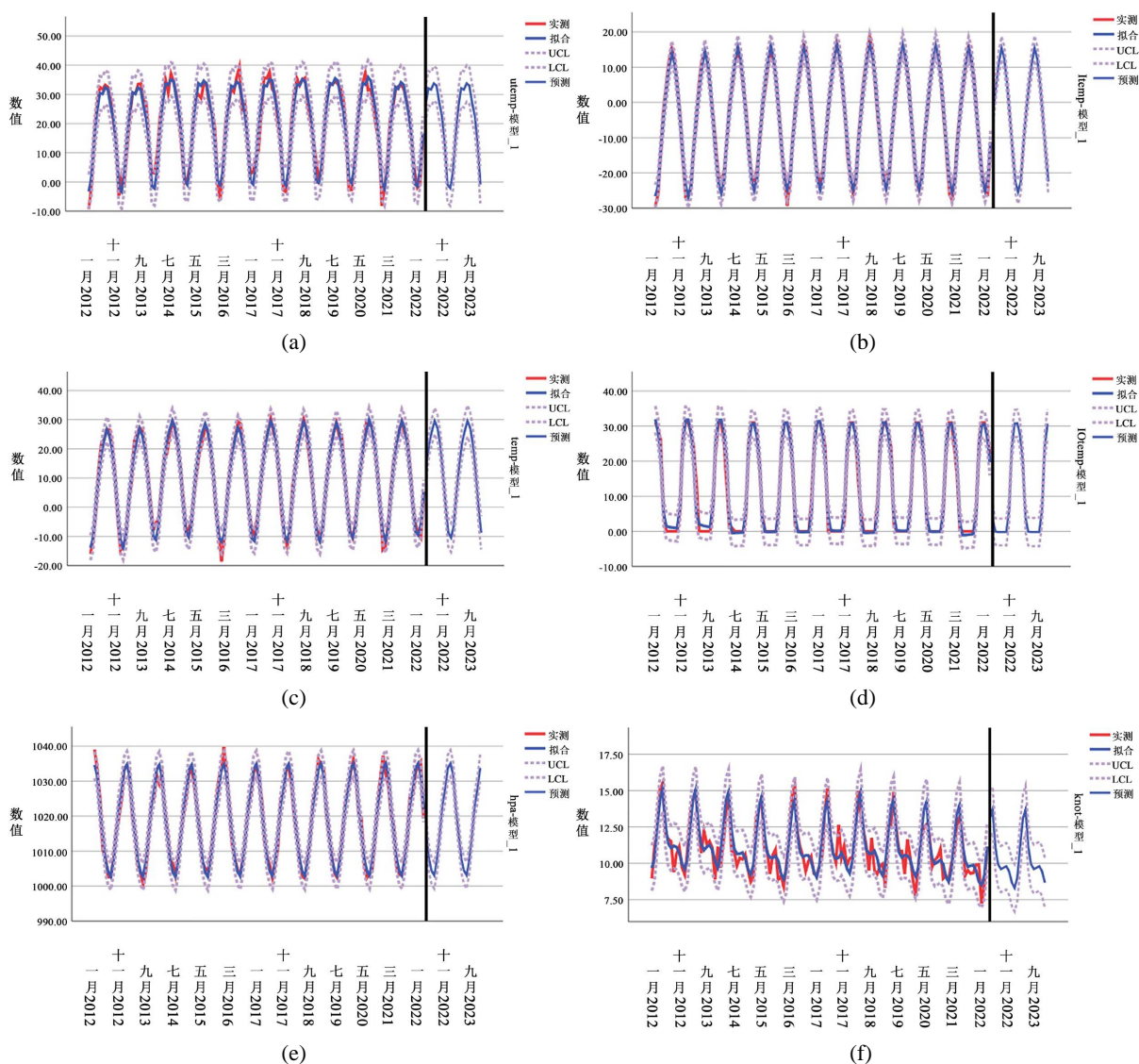


Figure 1. ARIMA modeling process

图 1. ARIMA 建模流程

图 2 列出了 ARIMA 预测的结果，共计 7 个变量预测模型，图中黑直线后表示后续预测数据，从图中可以看出预测线与实测线高度重合，说明 ARIMA 模型的预测结果十分贴合实际结果。预测结果具有较高可信度。2022 年、2023 年具体预测结果见表 4。



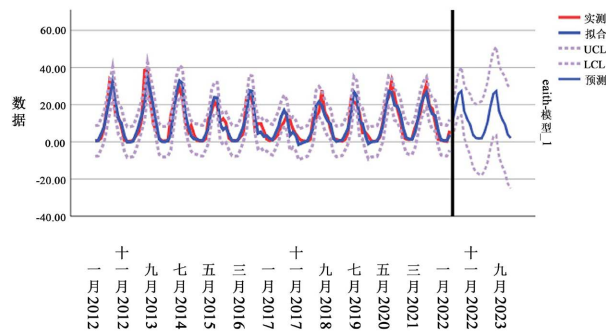


Figure 2. ARIMA prediction results graph. (a) Mean maximum temperature; (b) Mean minimum temperature; (c) Maximum temperature extreme; (d) The number of days when the mean temperature is < 0; (e) Mean sea level pressure; (f) Mean maximum sustained wind; (g) Soil evaporativeness

图 2. ARIMA 预测结果图。(a) 平均最高气温; (b) 平均最低气温; (c) 最高气温极值; (d) 平均气温 < 0 的天数; (e) 平均海平面气压; (f) 平均最大持续风; (g) 土壤蒸发量

Table 4. ARIMA forecasts results in 2022 and 2023

表 4. ARIMA 预测 2022、2023 年结果

年份	月份	平均最高气温(°C)	平均最低气温(°C)	最高气温极值(°C)	平均气温 ≤ 0°C 的天数	平均海平面气压(hPa)	平均最大持续风速(knots)	土壤蒸发量(mm)	
2022	04	14.39	-2.38	25.99	3.72	1015.09	13.04	10.56	
	05	21.84	4.97	32.13	-0.08	1007.78	13.7	18.95	
	06	26.11	11.28	31.66	-0.18	1004.38	11.1	25.79	
	07	29.53	15.44	33.77	-0.18	1003.05	9.94	27.27	
	08	27.26	12.42	33	-0.18	1007.42	9.61	16.41	
	09	21.83	5.46	27.13	-0.18	1014.62	9.7	12.32	
	10	12.29	-3.7	22.51	5.92	1023.11	9.8	9.35	
	11	0.75	-13.67	13.21	24.22	1027.59	9.46	3.81	
	12	-8.76	-22.42	-0.93	30.72	1033.71	8.67	2.08	
	2023	01	-10.44	-25.38	-2.11	30.82	1035.06	8.32	1.83
		02	-6.52	-22.39	3.4	27.55	1031.8	9.05	1.94
		03	5.24	-11.25	16.52	19.91	1022.06	11.16	5.31
04		14.39	-2.38	25.99	3.72	1015.14	13.04	10.56	
05		21.84	4.97	32.13	-0.08	1007.84	13.7	18.95	
06		26.11	11.28	31.66	-0.18	1004.43	11.1	25.79	
07		29.53	15.44	33.77	-0.18	1003.1	9.94	27.27	
08		27.26	12.42	33	-0.18	1007.47	9.61	16.41	
09		21.83	5.46	27.13	-0.18	1014.67	9.7	12.32	
10		12.29	-3.7	22.51	5.92	1023.16	9.8	9.35	
11		0.75	-13.67	13.21	24.22	1027.64	9.46	3.81	
12		-8.76	-22.42	-0.93	30.72	1033.77	8.67	2.08	

2.3. 自适应神经模糊推理系统

根据上述分析所得变量及预测数据结果，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。首先需要建立 Lasso 回归所筛选变量和土壤蒸发量变量与不同深度土壤湿度的模型，进而通过模型去预测(2022~2023)不同深度土壤湿度数据。

神经网络对于处理非线性的问题有很好的效果同时还具有自学习和自适应的能力，但是它不能很好的表现人脑的推理功能，其中的原因之一就是缺少透明度。仅仅模糊系统自身是没有自适应能力的。自适应神经模糊推理系统(ANFIS)融合了神经网络和模糊系统的优点，使得不仅具有自学习，自适应的能力，同时还具有了语言推理的能力，弥补了单一系统的不足。自适应神经模糊推理系统(ANFIS)属于神经模糊系统的一种。同其他神经模糊系统相比，ANFIS 具有便捷高效的特点，能处理复杂预测的问题。

2.3.1. ANFIS 建模

为了实现 T-S 模糊模型的学习过程，一般将其转化为一个自适应网络，即 ANFIS 模型[6]，模型结构如图 3。

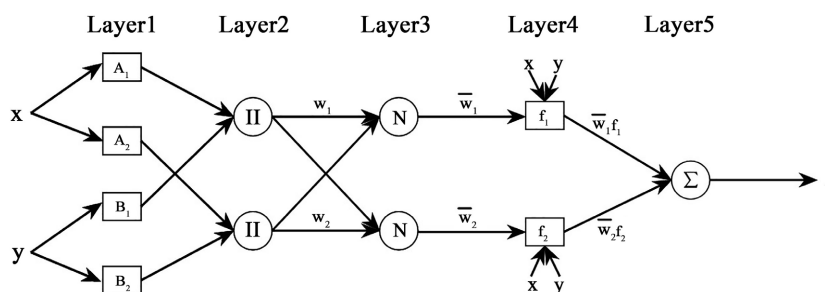


Figure 3. Adaptive network-based fuzzy inference system (ANFIS)

图 3. 自适应神经模糊推理系统(ANFIS)

2.3.2. ANFIS 预测结果

建立输入(平均最高气温, 平均最低气温, 最高气温极值, 平均气温 $\leq 0^{\circ}\text{C}$ 的天数, 平均海平面气压, 平均最大持续风速和土壤蒸发量变量)和输出(不同深度土壤湿度)模型, 针对 ANFIS 模型, 其建模数据为(2012~2021 年)每个月的检测结果, 其训练集、测试集和验证集比例: 7:2:1。本次使用模型结构为七个输入一个输出, 网络结构见图 4 所示, 针对不同深度土壤湿度(10 cm, 40 cm, 100 cm, 200 cm)分别建立四个模型, 以下以 10 cm 深度土壤湿度建模为例。其模型训练结果如图 5 所示, 模型训练平均误差如表 5 所示, 整体数据集训练拟合度均在 85%以上, 该模型较为准确。

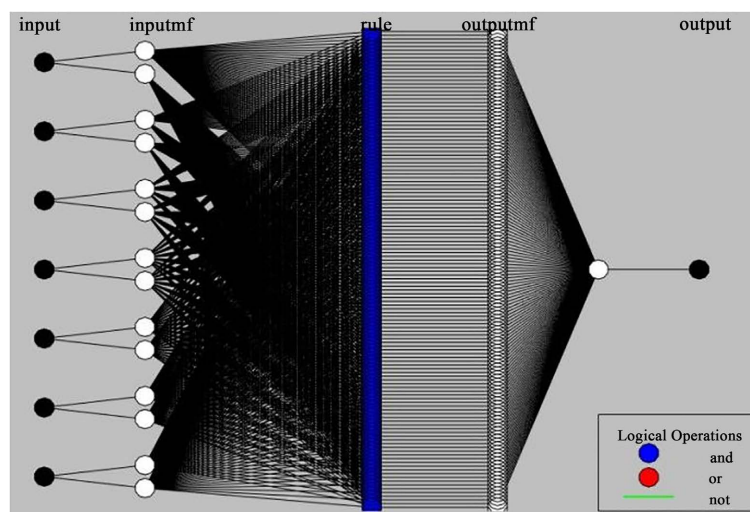


Figure 4. Seven input and one output network structure diagram

图 4. 七输入一输出网络结构图

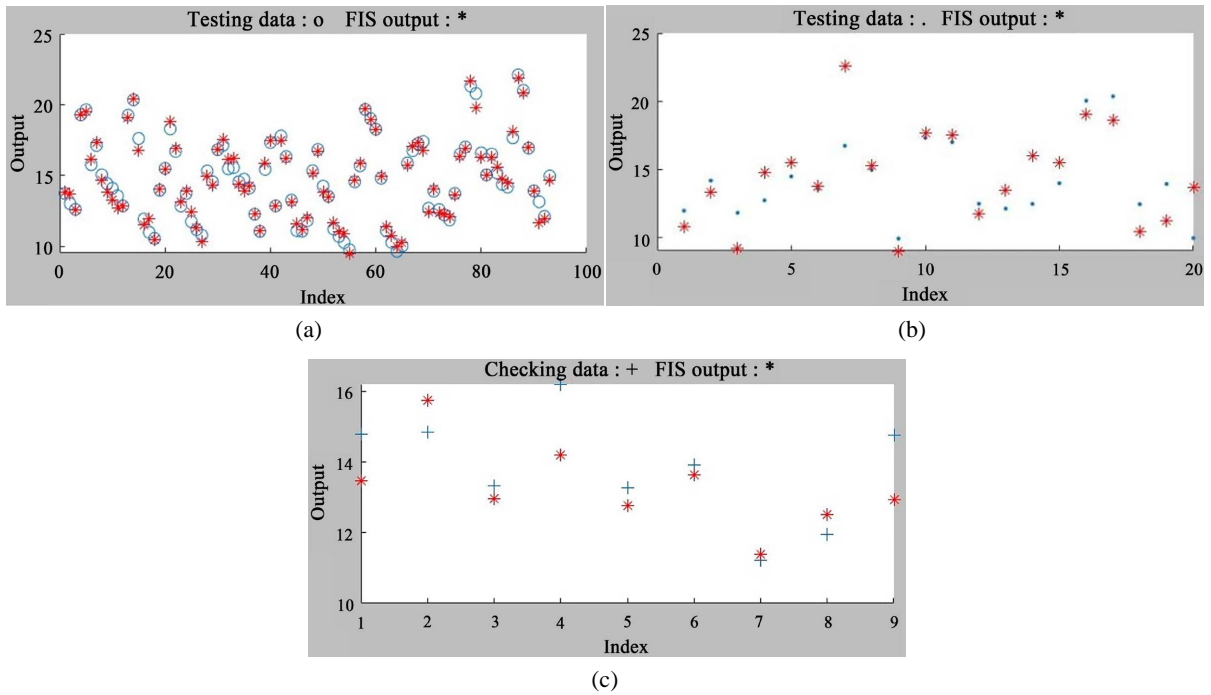
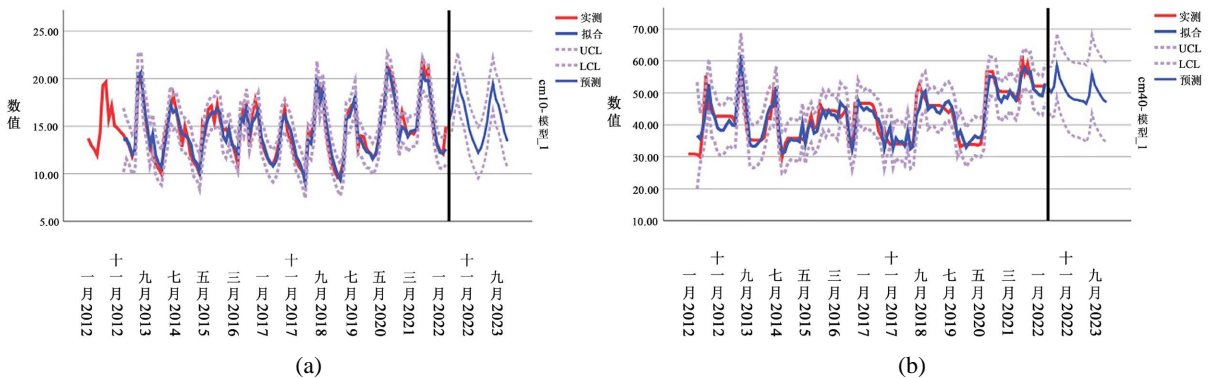


Figure 5. Model training results. (a) Training set; (b) Test set; (c) Verification set
图 5. 模型训练结果图。(a) 训练集；(b) 测试集；(c) 验证集

Table 5. Model training mean error
表 5. 模型训练平均误差

数据集合	平均误差	平均拟合度
训练集	0.394	0.973
测试集	2.187	0.854
验证集	1.903	0.873

针对 ANFIS 建立的模型进行预测, 其输入为 ARIMA 模型预测所获得的(2022~2023 年)数据, 输出为不同深度土壤湿度(10 cm, 40 cm, 100 cm, 200 cm), 具体预测结果见图 6, 具体数据见表 6 所示。2012~2021 年的预测值与实际值重合率很高, 故可以认为预测模型是比较准确的, 预测值具有较高的可信度。



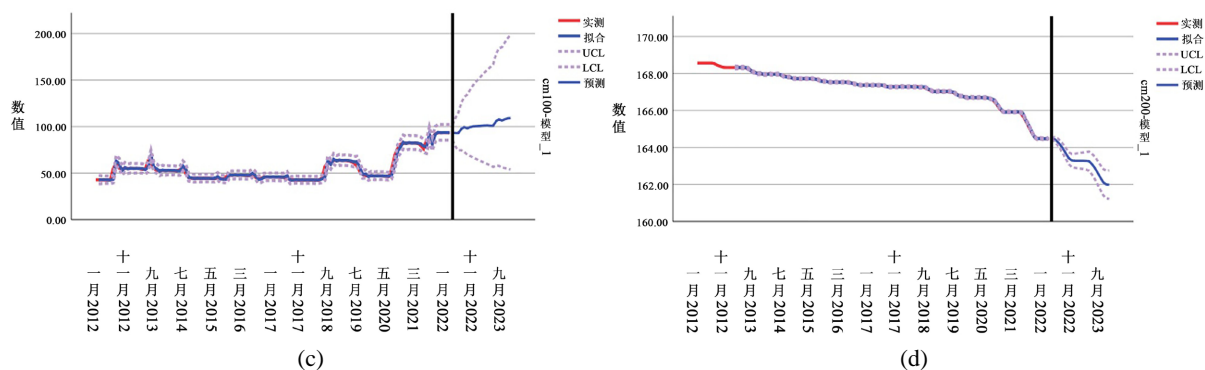


Figure 6. Prediction results of soil moisture at different depths. (a) Prediction results of soil moisture at a depth of 10 cm; (b) Prediction results of soil moisture at a depth of 40 cm; (c) Prediction results of soil moisture at a depth of 100 cm; (d) Prediction results of soil moisture at a depth of 200 cm

图 6. 不同深度土壤湿度预测结果。(a) 10 cm 深度土壤湿度预测结果；(b) 40 cm 深度土壤湿度预测结果；(c) 100 cm 深度土壤湿度预测结果；(d) 200 cm 深度土壤湿度预测结果

Table 6. Prediction results of soil moisture at different depths during 2022~2023

表 6. 2022~2023 年不同深度土壤湿度预测结果

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	15.44	51.87	95.11	164.47	
	05	16.63	50.20	96.68	164.45	
	06	18.84	51.88	99.45	164.32	
	07	20.26	58.61	101.80	164.15	
	08	18.47	54.54	102.49	163.92	
	09	17.65	52.55	99.28	163.67	
	10	16.15	50.65	98.23	163.42	
	11	14.6	49.09	97.52	163.30	
	12	13.67	48.39	96.05	163.28	
	2023	01	12.94	47.86	95.73	163.28
		02	12.22	47.83	95.86	163.28
		03	12.76	47.48	96.09	163.28
04		13.84	47.42	97.31	163.28	
05		15.36	46.61	99.12	163.26	
06		17.83	48.91	101.96	163.10	
07		19.45	55.97	104.37	162.90	
08		17.83	52.39	105.07	162.64	
09		17.14	50.74	101.79	162.37	
10		15.74	49.11	100.70	162.11	
11		14.28	47.69	99.98	162.00	
12		13.42	47.08	98.48	161.99	

3. 结论

本文发现提供的数据具有较强的共线性, 通过 Lasso 回归的方法对参数进行降维, 在预测 2022 年、2023 年不同深度土壤湿度时, 材料中并未给定相关数据, 需要 ARIMA 时间序列方法对以往年月数据进行预测。通过 ARIMA 模型预测所筛选的主变量和土壤蒸发量在 2022~2023 年的数据。建立输入(Lasso 回归所筛选变量和土壤蒸发量变量)和输出(不同深度土壤湿度) ANFIS 模型, 对往年所测数据进行训练且整体数据集训练拟合度均在 85% 以上, 该模型较为准确。通过 ARIMA 模型预测得到的相关参数数据去预测 2022 年、2023 年不同深度土壤湿度。本研究对于以后放牧策略的变化和草原生态系统的保护具有重要意义。

参考文献

- [1] 黄季焜. 农业供给侧结构性改革的关键问题: 政府职能和市场作用[J]. 中国农村经济, 2018(2): 2-14.
- [2] 于格, 鲁春霞, 谢高地, 郭广猛. 基于 RS 和 GIS 的青藏高原草地生态系统土壤水分保持功能及其经济价值评估——以生长季为例[J]. 山地学报, 2006, 24(4): 498-503.
- [3] 邓国伦, 邓振镛, 万信, 方德彪. 陇东黄土高原土壤贮水量对小麦产量影响的研究[J]. 气象科技, 1998, 26(1): 56-60.
- [4] 中国统计数据库——锡林郭勒盟统计年鉴 2012-2022 [EB/OL]. <http://www.shujuku.org/xilin-gol-league-statistical-yearbook.html>, 2022-12-26.
- [5] 耿芳艳, 蔡仕茂. 基于 ARIMA 模型的贵州省玉米产量预测研究[J]. 作物研究, 2022, 36(4): 379-383.
- [6] 刘开启. 基于自适应神经模糊推理系统的 MTDC 系统下垂控制策略[J]. 自动化应用, 2022(1): 17-20.