计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200200059

基于多种模型的云南省农作物主产区域部分气候指标分析与预测

陈 沛 郑万波 刘文奇 肖 敏 张凌霄
1 昆明理工大学理学院 昆明 650500
2 昆明理工大学数据科学研究中心 昆明 650500
(1262955357@qq.com)

摘 要 针对目前云南省缺乏农作物种植与气候指标预测模型与建模方法的问题,首先,概述了降水量、温度、空气湿度等主要气候因素的数据分析预测模型的研究现状,分析温度、降雨量、湿度与农业气候资源的综合关系,进行数据清洗,并筛选出主要分析指标;其次,使用1981年到2010年共30年的数据分析云南省的降水量、温度、空气湿度模型;再次,采用Matlab Curve Fitting Tool拟合函数进行气候预测,得到了所选取地区的气候指标的预测模型并计算预测误差,进行数值拟合误差分析;最后,利用SPSS软件建立ARIMA模型,以此作为前述模型的补充修正。通过实验验证,成功地将模型90%的预测值误差控制在10%以内。本研究建立了针对云南省主产区域部分气候指标分析与预测模型,对云南农作物种植区域规划有一定指导作用。

关键词:数据挖掘与分析;环境监测;气候模型拟合;区域气候预测;MATLAB应用;ARIMA模型
中图法分类号 TP391.9

Analysis and Forecast of Some Climate Indexes in Main Producing Areas of Yunnan Province Based on Multiple Models

CHEN Pei,ZHENG Wan-bo,LIU Wen-qi,XIAO Min and ZHANG Ling-xiao
1 School of Science,Kunming University of Science and Technology,Kunming 650500,China
2 Data Science Research Center,Kunming University of Science and Technology,Kunming 650500,China

Abstract In view of the lack of prediction models and modeling methods of crop planting and climate index in Yunnan Province, firstly,the research status of data analysis and prediction models of main climatic factors such as precipitation,temperature and air humidity are summarized. The comprehensive relationship between temperature,rainfall,humidity and agroclimatic resources are analyzed,the data are cleaned,and the main analysis indexes are selected. Secondly,the precipitation,temperature and air humidity model of Yunnan Province are analyzed by using the data of 30 years from 1981 to 2010. Thirdly,the Matlab Curve Fitting Tool fitting function is used to predict the climate,and the prediction model of the climate index in the selected area is obtained and the prediction error is calculated,and the numerical fitting error analysis is carried out. Finally,the ARIMA model is established by SPSS software,which is used as a supplement to the above-mentioned model. Through experimental verification,the prediction error of model 90% is successfully controlled within 10%. Through this study,a model for analyzing and predicting some climate indexes in the main producing areas of Yunnan Province is established,which plays a guiding role in the regional planning of crop planting in Yunnan Province.

Keywords Data mining and analysis,Environmental monitoring,Climate model fitting,Regional climate prediction,MATLAB application,ARIMA model

文献[1]说明,截至2016年,云南省第一产业占全省经济总量的14.8%,相对全国总体第一产业占比8.6%,东部沿海发达省份的第一产业在本省经济总量占比普遍低于5%,云南省的产业结构发展较为落后,同时云南省的第一产业在全省的经济发展中仍发挥着较为重要的作用。本研究选取降水量、温度、空气湿度3项指标进行建模,其中,降水是重要的区域水资源补充方式,水是农作物生长的必备条件;温度是农作物生长发育的必要气候因子,热量的多少与农作物的产量直接相关;相对空气湿度(下文简称为空气湿度)与农作物生产没有直接关系,但是与人的体感有关。因为农业资源环境适宜性是一个复杂的模糊的量化指标,借鉴文献[3]的研究方法,我们将降水量、温度作为农作物的环境适宜性指标,因为空气湿度对农作物的影响较小,针对空气湿度只建立预测模型,不与降水量模型、温度模型一起作为环境适宜性的分析指

基金项目:国家自然科学基金项目(61472051);重庆市社会事业与民生保障科技创新专项(cstc2016shmszx90002);昆明理工大学人培项目(KKZ3201907003)
This work was supported by the National Natural Science Foundation of China(61472051),Chongqing Social Undertakings and Livelihood Security Science and Technology Innovation Project Special Program(cstc2016shmszx90002) and Kunming University of Science and Technology Talent Training Project Fund(KKZ3201907003).
通信作者:郑万波(zwanbo2001@163.com)

标。在基于数据模型预测领域，文献[4]利用人工合成新的序列对平稳随机项进行模拟和预测，并将其应用于降水量预测，取得了较好的效果。文献[5]利用序列集分析模型的降水量预测，获得了较高的精度。文献[6]基于极限学习机 ELM 模型、BP 神经网络模型和广义回归神经网络 GRNN 模型建立了降雨预测模型，达到了作为预报模型的精度。文献[7]利用逐步回归将温度预测模型的误差控制在 5% 以内。文献[8]利用改进的 BP 神经网络对土壤温度进行了预测。空气湿度的预测主要基于神经网络并应用于室内环境[9-10]。对降水、温度和空气湿度预测的成果较多，也较为成熟，但是降水预测主要侧重于中长期预测，温度预测和空气湿度预测主要使用神经网络技术或是某一种数学模型。就云南省地区的相关研究情况来看，研究成果有文献[11]使用 ARIMA 模型进行云南省气象干旱预测，文献[12]使用 WRF 模式进行云南省干旱预测，文献[13]使用灰色灾变理论对云南省干旱气候进行预测等，但是目前针对云南省主要地区利用多种模型的精确到天的气象指标预测尚无先例。

1　数据收集与清洗

本研究所使用的原始数据来自中国气象数据网，数据的时间区段为 1981—2010 年，根据区站号对数据所属区域进行划分。

利用 pandas 库读取数据，并且进行数据的特征值提取，之后利用 python 的 pandas 库中的 pandas.read_csv()函数设置数据清洗参数进行数据清洗，避免因数据缺陷而导致结果错误。

2　数据分析与显示

2.1　温度、降雨量、湿度与农业气候资源的综合关系分析

文献[14]表明，温度、降水量、光照时数与自然灾害情况与农业气候资源有着紧密联系，能够直接影响农作物产量，因为光照时数据的缺失，只能选取温度、降水量来判断农业气候资源的质量优劣。

空气湿度不直接影响农作物生长和产量，但是作为影响人体感觉的重要气候指标一并进行分析，但不作为农业气候资源指标参与本研究。

2.2　数据可视化

根据以上读取数据，对云南省各个有数据记录的地区的空气湿度、温度、降水量集中绘制散点图，如图 1—图 4 所示，分析 3 项气候指标所遵循的函数规律。

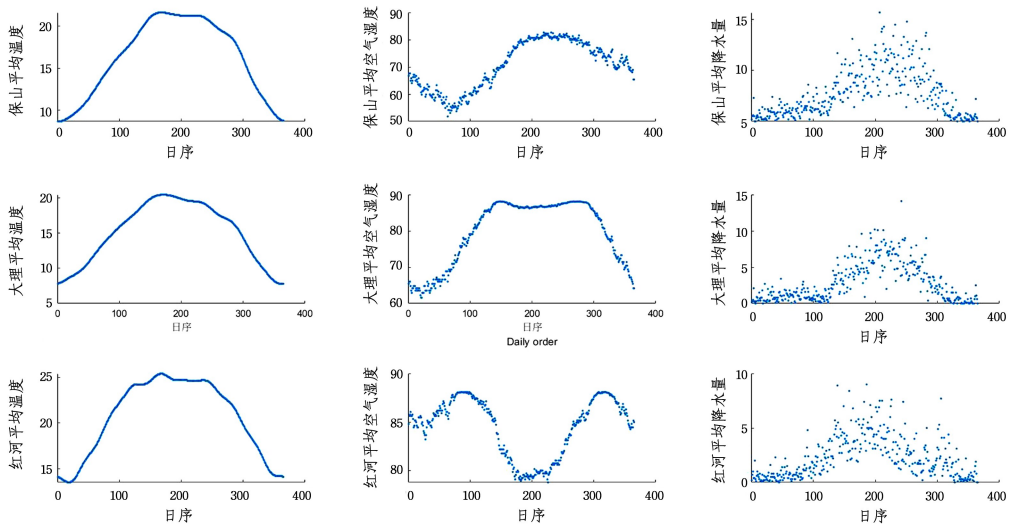

图 1　云南省保山、大理、红河地区日序平均温度、平均空气湿度、平均降水量散点图

Fig. 1　Daily sequence average temperature, average air humidity and average precipitation scatter chart in Baoshan, Dali and Honghe area of Yunnan province

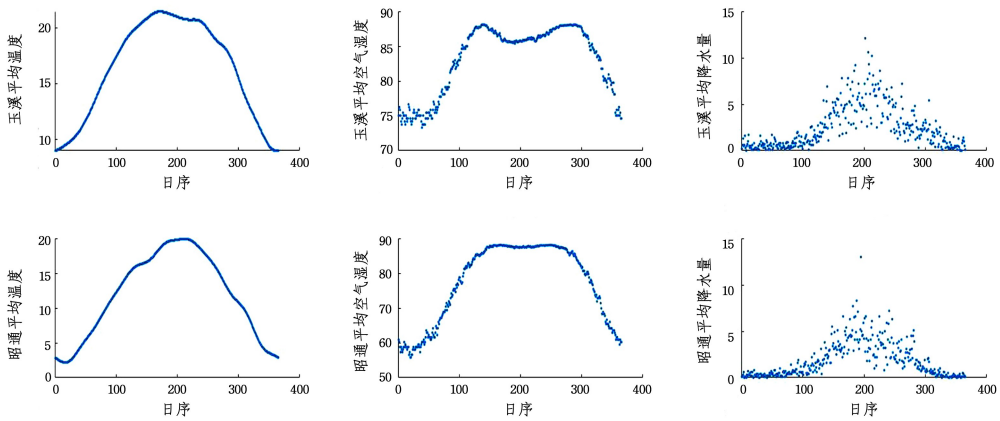

图 2　云南省玉溪、昭通地区日序平均温度、平均空气湿度、平均降水量散点图

Fig. 2　Daily sequence average temperature, average air humidity and average precipitation scatter chart in Yuxi and Zhaotong area of Yunnan province

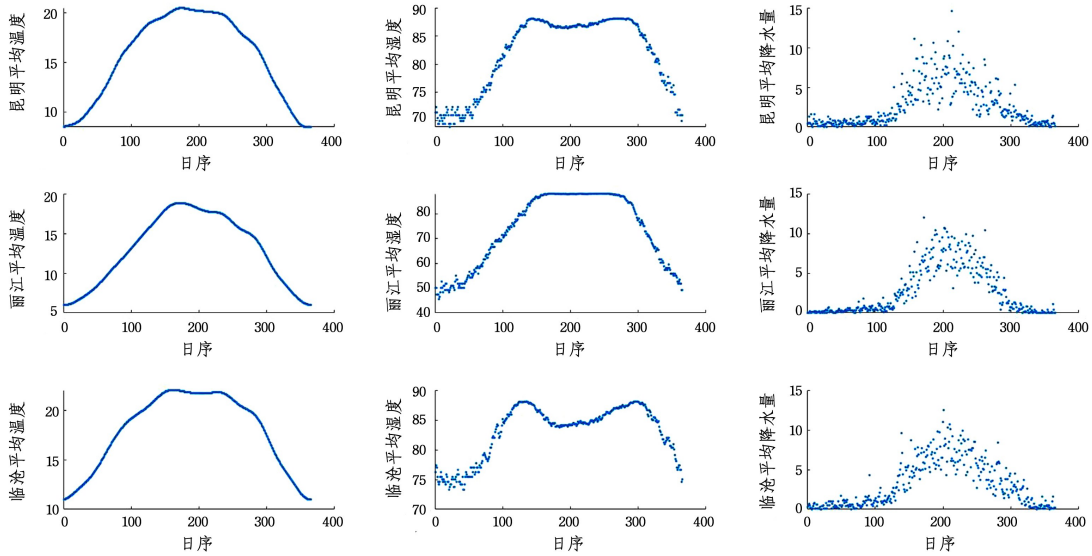

图 3 云南省昆明、丽江、临沧地区日序平均温度、平均空气湿度、平均降水量散点图

Fig. 3 Daily sequence average temperature, average air humidity and average precipitation scatter chart in Kunming, Lijiang and Lincang of Yunnan province

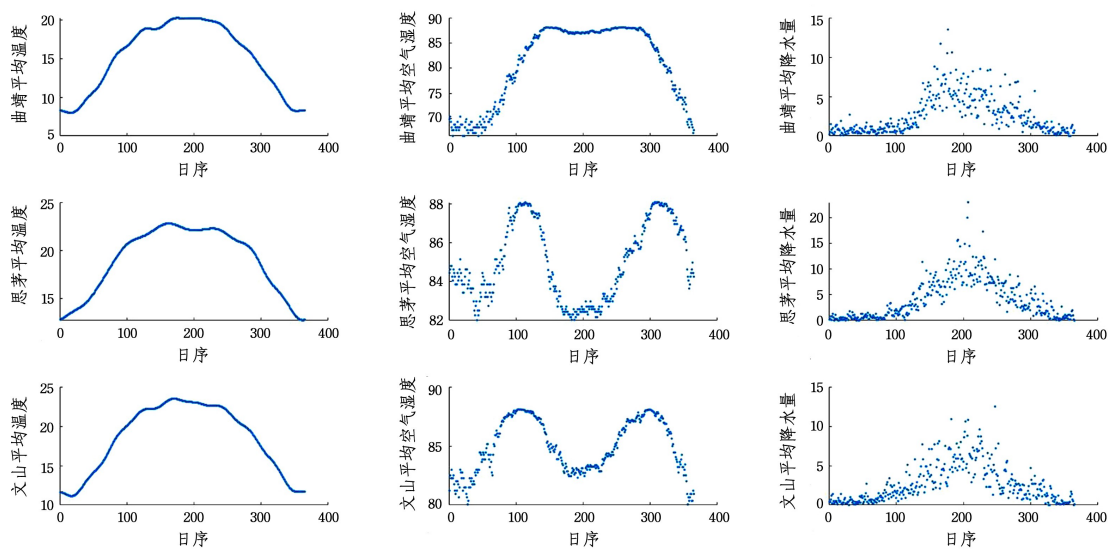

图 4 云南省曲靖、思茅、文山地区日序平均温度、平均空气湿度、平均降水量散点图

Fig. 4 Daily sequence average temperature, average air humidity and average precipitation scatter chart in Qujing, Simao and Wenshan area of Yunnan province

3 数据拟合建模与预测分析

通过 Matlab 的 Curve Fitting Tool，根据数据的特点和适用性选择数据拟合的方法。本研究的数据可以使用的拟合方法主要有：fourier 模型、gaussian 模型、interpolant 插值模型、polynomial 多项式模型、power 功率函数、rational 有理方程模型、spline 立方样条和平滑样条模型。因为 interpolant 插值模型、power 功率函数明显不能在本次研究中得到确定的函数关系，因此经过筛选在本次研究中可以较好地对数据拟合的拟合模型有 polynomial 多项式模型、rational 有理方程模型、fourier 傅里叶级数模型、gaussian 模型。上述 4 种模型的阶数均可根据 matlab Curve Fitting Tool 统计值的情况进行调整。4 种模型的简单介绍如下：

polynomial 多项式模型，例如 $f(x)=p_1x^9+p_2x^8+p_3x^7+p_4x^6+p_5x^5+p_6x^4+p_7x^3+p_8x^2+p_9x+p_{10}$ 是一个九阶多项式，其中 p_i 为常数，x 为自变量。

rational 有理方程模型，例如 $f(x)=\frac{p_1x^5+p_2x^4+p_3x^3+p_4x^2+p_5x+p_6}{x^5+q_1x^4+q_2x^3+q_3x^2+q_4x+q_5}$ 为最高阶 5 度/5 度，即 5 次多项式 a/5 次多项式 b，p_i 为常数，x 为自变量。

fourier 傅里叶级数模型，例如 $f(x)=a_0+a_1\cos(\omega x)+b_1\sin(\omega x)+a_2\cos(2\omega x)+b_2\sin(2\omega x)+a_3\cos(3\omega x)+b_3\sin(3\omega x)+a_4\cos(4\omega x)+b_4\cos(4\omega x)$ 是一个四阶傅里叶级数模型，a_i，b_i，ω 为常数，x 为自变量。

gaussian 模型即高斯模型，例如 $f(x)=a_1e^{(-\frac{x-b_1}{c_1})^2}+a_2e^{(-\frac{x-b_2}{c_2})^2}+a_3e^{(-\frac{x-b_3}{c_3})^2}$ 是一个三阶高斯模型 a_i，b_i，c_i 为常数。

通过观察分析发现，很难直接观察出较为准确的函数类型，因此采用上述 4 种可用的拟合函数模型进行比较得到抽

取的 5 个地区的 3 项气候指标的拟合函数，结合 Matlab Curve Fitting Tool 统计值，取最优拟合结果。统计值类型及效果优劣判断标准如下。

SSE 是拟合误差的平方和，其值越接近 0，说明拟合结果推断的准确性越高。SSE 的完整公式为：

$$SSR=\sum_{i=1}^{n}\omega_i(y_i-\hat{y}_i)^2 \tag{1}$$

其中，ω_i 为权值，y_i 为测量值，$\hat{y}_i$ 为估计值。

R-Square 是表达值与推测值之间相关系数的平方值。它越接近 1，说明模型越能更好地解释变量间的比例关系。R-Square 的公式为：

$$R\text{-}square=\frac{SSR}{SST} \tag{2}$$

其中，$SSR=\sum_{i=1}^{n}\omega_i(\hat{y}_i-\bar{y}_i)^2$，$SST=\sum_{i=1}^{n}\omega_i(y_i-\bar{y}_2)^2$。

Adjusted R-square 是按照误差自由度调整后的 R-square。它越接近 1，说明拟合结果越好。

RMSE 是均方差，又称标准误差（注意不是标准差）。它越接近 0 说明拟合结果的推断越有用。

选取的云南省气候特征比较具有代表性的 5 个地区的拟合函数（以一年当中的天序数为自变量）如下：

（1）保山

1）降水量

$$\begin{aligned}f(x)=&-4.832\,e^{-(\frac{x-204.6}{1.484})^2}+2.31\,e^{-(\frac{x-249.1}{10.93})^2}+\\&3.114\,e^{-(\frac{x-230.4}{2.85})^2}+2.63\,e^{-(\frac{x-280.6}{24.5})^2}-4.697\,e^{-(\frac{x-189.2}{13.99})^2}+\\&3.585e^{-(\frac{x-191.6}{108.6})^2}+2.973\,e^{-(\frac{x-260.3}{1.647})^2}+\\&5.354\,e^{-(\frac{x-196.9}{26.51})^2}\end{aligned} \tag{3}$$

2）温度

$$\begin{aligned}f(x)=&16.26-6.112\cos(0.01722x)-1.461\sin(0.01722x)-1.052\cos(0.03444x)-0.253\sin(0.03444x)-0.2861\cos(0.05166x)+0.3199\sin(0.05166x)+0.00964\cos(0.06888x)-0.08519\sin(0.06888x)+0.1354\cos(0.0861x)+0.0927\sin(0.0861x)-0.09188\cos(0.10322x)-0.1037\sin(0.10322x)-0.08612\cos(0.12054x)-0.01252\sin(0.12054x)-0.03436\cos(0.13776x)+0.06267\sin(0.13776x)\end{aligned} \tag{4}$$

3）空气湿度

$$f(x)=-8.037\times10^{-18}x^8+1.242\times10^{-14}x^7-9.045\times10^{-12}x^6+3.998\times10^{-9}x^5-1.076\times10^{-6}x^4+0.0001546x^3-0.008366x^2-0.02975x+66.47 \tag{5}$$

（2）昆明

1）降水量

$$\begin{aligned}f(x)=&2.69-2.797\cos(0.01728x)-1.402\sin(0.01728x)+0.7501\cos(0.03456x)+0.6537\sin(0.03456x)-0.2863\cos(0.05184x)+0.03652\sin(0.05184x)-0.02273\cos(0.06912x)-0.01766\sin(0.06912x)+0.3013\cos(0.0864x)+0.188\sin(0.0864x)-0.2002\cos(0.10368x)-0.2008\sin(0.10368x)+0.1048\cos(0.12096x)+0.1918\sin(0.12096x)+0.08882\cos(0.13824x)-0.08944\sin(0.13824x)\end{aligned} \tag{6}$$

2）温度

$$\begin{aligned}f(x)=&15.76-5.566\cos(0.01779x)-1.187\sin(0.01779x)-1.169\cos(0.03598x)-0.4513\sin(0.03598x)-0.2009\cos(0.05377x)-0.0537\sin(0.05377x)-0.05068\cos(0.07116x)+0.03679\sin(0.07116x)-0.0597\cos(0.08895x)+0.09311\sin(0.08895x)+0.033\cos(0.10674x)-0.03591\sin(0.10674x)-0.1042\cos(0.12453x)-0.05709\sin(0.12453x)\end{aligned} \tag{7}$$

3）空气湿度

$$\begin{aligned}f(x)=&18.1\,e^{-(\frac{x-295.3}{63.6})^2}+0.8036\,e^{-(\frac{x-139.8}{5.477})^2}+\\&4.283\,e^{-(\frac{x-149}{26.65})^2}+0.9919\,e^{-(\frac{x-124.9}{7.242})^2}+\\&8.341\,e^{-(\frac{x-107.2}{39.65})^2}+17.22\,e^{-(\frac{x-193.4}{77.66})^2}+\\&82.68\,e^{-(\frac{x-1439}{3634})^2}+1.935\,e^{-(\frac{x-348.6}{5.368})^2}\end{aligned} \tag{8}$$

（3）红河

1）降水量

$$\begin{aligned}f(x)=&2.464\,e^{-(\frac{x-185.8}{9.062})^2}+2.875\,e^{-(\frac{x-138.1}{2.921})^2}+\\&1.672\,e^{-(\frac{x-152}{24.26})^2}+2.727\,e^{-(\frac{x-304.2}{5.054})^2}+\\&5.107\,e^{-(\frac{x-225.2}{1.283})^2}+2.352\,e^{-(\frac{x-207.1}{11.97})^2}+\\&1.861\,e^{-(\frac{x-246.1}{2.702})^2}+2.807\,e^{-(\frac{x-201.2}{129})^2}\end{aligned} \tag{9}$$

2）温度

$$\begin{aligned}f(x)=&2.479\,e^{-(\frac{x-160.1}{21.89})^2}+24.83\,e^{-(\frac{x-214.2}{184.7})^2}+\\&2.9992\,e^{-(\frac{x-125.2}{19.46})^2}+0.2577\,e^{-(\frac{x-106.6}{8.598})^2}-\\&3.101\,e^{-(\frac{x-65.63}{22.78})^2}-6.237\,e^{-(\frac{x-83.77}{54.31})^2}+\\&7.411\,e^{-(\frac{x+9.137}{27.45})^2}+13.94\,e^{-(\frac{x-73.45}{50.21})^2}\end{aligned} \tag{10}$$

3）空气湿度

$$f(x)=-2.687\times10^{-17}x^8+4.129\times10^{-14}x^7-2.492\times10^{-11}x^6+7.287\times10^{-9}x^5-1.0041\times10^{-6}x^4+4.442\times10^{-5}x^3+0.001757x^2-0.1147x+86.01 \tag{11}$$

（4）曲靖

1）降水量

$$\begin{aligned}f(x)=&2.708-2.262\cos(0.01828x)-1.389\sin(0.01828x)+0.6203\cos(0.03656x)+0.2541\sin(0.03656x)-0.6384\cos(0.05484x)+0.1579\sin(0.05484x)+0.2397\cos(0.07312x)-0.04743\sin(0.07312x)-0.124\cos(0.0914x)+0.244\sin(0.0914x)+0.1918\cos(0.10968x)-0.1067\sin(0.10968x)-0.1284\cos(0.12796x)-0.1939\sin(0.12796x)-0.09574\cos(0.14624x)+0.2\sin(0.14624x)\end{aligned} \tag{12}$$

2）温度

$$f(x)=14.64-6.261\cos(0.01582x)+1.083\sin$$

$(0.01582x)-0.5369\cos(0.03164x)+0.2446\sin(0.03164x)+0.2635\cos(0.04746x)-0.3553\sin(0.04746x)+0.341\cos(0.06328x)-0.2022\sin(0.06328x)+0.1592\cos(0.0791x)-0.185\sin(0.0791x)-0.04113\cos(0.09492x)-0.2117\sin(0.09492x)-0.01214\cos(0.11074x)-0.05202\sin(0.11074x)-0.1286\cos(0.12656x)-0.1276\sin(0.12656x)$ (13)

3)空气湿度

$f(x)=1\ 979+511.6\cos(0.00876x)-3\ 444\sin(0.00876x)-2566\cos(0.01752x)-794.4\sin(0.01752x)-757.6\cos(0.02628x)+1523\sin(0.02628x)+692.7\cos(0.03504x)+521.4\sin(0.03504x)+265.7\cos(0.0438x)-218.1\sin(0.0438x)-31.98\cos(0.05256x)-93.03\sin(0.05256x)-19.45\cos(0.06132x)-6.12\sin(0.06132x)-3.449\cos(0.07008x)+1.548\sin(0.07008x)$ (14)

(5)昭通

1)降水量

$f(x)=2.13-1.26\cos(0.02034x)-1.79\sin(0.02034x)+0.04845\cos(0.04068x)+0.3009\sin(0.04068x)-0.1194\cos(0.06102x)-0.3731\sin(0.06102x)-0.1339\cos(0.08136x)+0.08936\sin(0.08136x)+0.07579\cos(0.1017x)+0.007145\sin(0.1017x)-0.1717\cos(0.12204x)-0.2102\sin(0.12204x)-0.1763\cos(0.14238x)+0.301\sin(0.1438x)+0.1854\cos(0.16272x)+0.05852\sin(0.16272x)$ (15)

2)温度

$f(x)=11.86-8.213\cos(0.01722x)-1.92\sin(0.01722x)-0.9405\cos(0.03444x)+0.03234\sin(0.03444x)-0.09826\cos(0.05166x)-0.3677\sin(0.05166x)-0.1274\cos(0.06888x)+0.1405\sin(0.06888x)-0.04722\cos(0.0861x)-0.002401\sin(0.0861x)+0.2123\cos(0.10332x)+0.05377\sin(0.10332x)+0.09129\cos(0.12054x)-0.004323\sin(0.12054x)+0.1512\cos(0.13776x)-0.1377\sin(0.13776x)$ (16)

3)空气湿度

$$f(x)=28.99\,e^{-(\frac{x-254.4}{76.97})^2}+0.6414\,e^{-(\frac{x-1481}{6.628})^2}+1.15\,e^{-(\frac{x-121.6}{14.84})^2}+24.86\,e^{-(\frac{x-147.7}{63.96})^2}+5.307\,e^{-(\frac{x-93.41}{32.67})^2}+5.615\,e^{-(\frac{x-307.3}{36.87})^2}+58.16\,e^{-(\frac{x+121.8}{2341})^2} \quad (17)$$

4 数据校验与偏差计算

4.1 预测和偏差计算

对一年十二个月中随机抽样出的月份使用与求得的拟合函数计算与实际月序数据比对，使用偏差计算公式 $z=\frac{|y-x|}{y}\cdot 100\%$（其中 y 为实际值，x 为预测值，z 为拟合偏差）得到了拟合偏差。云南省部分地区降水量、温度和湿度值拟合结果如表 1－表 3 所列。

表 1 云南省部分地区降水量值拟合结果

Table 1 Fitting results of precipitation in some areas of Yunnan province

Area	Month	Actual value	Predicted value	Fitting deviation/%
保山	1	16.90	8.5283	49.54
	4	48.30	58.29	11.97
	7	153.90	159.32	3.52
	11	38.10	35.10	7.87
昆明	1	15.80	16.23	2.72
	4	25.20	26.74	6.11
	7	200.20	198.05	1.07
	11	36.70	33.67	8.26
昭通	1	8.40	11.33	34.88
	4	30.10	30.19	0.30
	7	151.90	149.97	1.27
	11	15.90	12.32	22.51

表 2 云南省部分地区温度值拟合结果

Table 2 Fitting results of temperature values in some areas of Yunnan province

Area	Month	Actual value	Predicted value	Fitting deviation/%
保山	1	9.10	9.40	3.29
	4	16.80	16.95	0.89
	7	21.30	21.27	0.14
	11	13.20	12.91	2.20
昆明	1	8.90	9.16	2.92
	4	17.30	17.45	0.87
	7	20.30	20.21	0.89
	11	12.10	11.47	5.21
昭通	1	2.40	2.55	6.25
	4	12.80	13.08	2.19
	7	19.70	20.42	3.65
	11	8.10	7.77	4.07

表 3 云南省部分地区空气湿度量值拟合结果

Table 3 Fitting results of air humidity in some areas of Yunnan province

Area	Month	Actual value	Predicted value	Fitting deviation/%
保山	1	67	66.35	0.97
	4	65	60.25	7.30
	7	81	79.58	1.75
	11	75	70.31	6.25
昆明	1	66	72.88	10.42
	4	56	82.5617	47.43
	7	81	86.61	6.92
	11	75	81.92	9.23
昭通	1	74	60.28	18.54
	4	67	79.56	18.75
	7	78	87.50	12.18
	11	78	74.83	4.06

4.2 误差分析

预测结果与实际数据存在偏差是正常现象，从预测偏差值来看，90%的误差值在 10%以内，15%的误差值在 1%以内，但是也有占拟合结果总数量的 10%的预测值对应拟合偏差达到了 10%以上甚至 40%以上，明显存在错误。本研究认为对降水量的预测偏差主要原因有：1)部分时期出现极端天气；2)气候数据时间跨度只有 30 年，导致数据量不够大；3)各

年的降水量变化幅度较大。对空气湿度进行预测检验，是根据日相对水汽压计算空气湿度并与已有数据的月平均空气湿度进行比较进行的，产生误差的主要原因可能是：1）数据源存在错误的可能；2）空气湿度在一天内变化速度和范围均比较大；3）使用的数学模型对于比较分散的数据普遍不能达到使统计值表现足够好的效果。

4.3　模型修正

经过计算拟合偏差的结果说明，当前使用的模型不足以在任何时间段内较好地预测。为了提高数据预测的精准度，借鉴文献[15-18]的方法，运用 SPSS 软件建立 ARIMA 模型，以此作为前述模型的补充修正部分。

ARIMA 模型全称为自回归积分滑动平均模型，是一种最常见的对时间序列进行预测的模型，本研究所有涉及的气候指标均为时间序列，所以该模型对本次研究具有重要作用。

ARIMA 模型：$ARIMA(p,d,q)$。其有 3 个参数，p 代表预测模型中采用的时序数据本身的滞后数，d 代表时序数据需要进行几阶差分化才达到稳定状态，q 代表预测模型中采用的预测误差的滞后数。保山降水量 ARIMA 模型为 ARIMA(0,1,2)；昭通降水量 ARIMA 模型为 ARIMA(0,1,1)；昆明空气湿度 ARIMA 模型为 ARIMA(0,1,17)；昭通空气湿度 ARIMA 模型为 ARIMA(0,1,8)。4 个 ARIMA 模型的拟合结果如表 4－表 7 所列。

表 4　保山地区降水量 ARIMA 模型拟合结果

Table 4　Fitting results of ARIMA model of precipitation in Baoshan area

Fitting statistics	Mean value	Minimum value	Maximum value	Percentile						
				5	10	25	50	75	90	95
平稳的 R^2	0.351	0.351	0.351	0.351	0.351	0.351	0.351	0.351	0.351	0.351
R^2	0.652	0.652	0.652	0.652	0.652	0.652	0.652	0.652	0.652	0.652
RMSE	1.380	1.380	1.380	1.380	1.380	1.380	1.380	1.380	1.380	1.380
MAPE	69.18	69.18	69.18	69.18	69.18	69.18	69.18	69.18	69.18	69.18
MaxAPE	1081.230	1081.230	1081.230	1081.230	1081.230	1081.230	1081.230	1081.230	1081.230	1081.230
MAE	0.989	0.989	0.989	0.989	0.989	0.989	0.989	0.989	0.989	0.989
MaxAE	6.098	6.098	6.098	6.098	6.098	6.098	6.098	6.098	6.098	6.098
正态化的 *BIC*	0.677	0.677	0.677	0.677	0.677	0.677	0.677	0.677	0.677	0.677

表 5　昭通地区降水量 ARIMA 模型拟合结果

Table 5　Fitting results of ARIMA model of precipitation in Zhaotong area

Fitting statistics	Mean value	Minimum value	Maximum value	Percentile						
				5	10	25	50	75	90	95
平稳的 R^2	0.368	0.368	0.368	0.368	0.368	0.368	0.368	0.368	0.368	0.368
R^2	0.685	0.685	0.685	0.685	0.685	0.685	0.685	0.685	0.685	0.685
RMSE	1.068	1.068	1.068	1.068	1.068	1.068	1.068	1.068	1.068	1.068
MAPE	58.165	58.165	58.165	58.165	58.165	58.165	58.165	58.165	58.165	58.165
MaxAPE	603.520	603.520	603.520	603.520	603.520	603.520	603.520	603.520	603.520	603.520
MAE	0.677	0.677	0.677	0.677	0.677	0.677	0.677	0.677	0.677	0.677
MaxAE	8.130	8.130	8.130	8.130	8.130	8.130	8.130	8.130	8.130	8.130
正态化的 *BIC*	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148

表 6　昆明地区空气湿度 ARIMA 模型拟合结果

Table 6　Fitting results of ARIMA model of air humidity in Kunming area

Fitting statistics	Mean value	Minimum value	Maximum value	Percentile						
				5	10	25	50	75	90	95
平稳的 R^2	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
R^2	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
RMSE	0.669	0.669	0.669	0.669	0.669	0.669	0.669	0.669	0.669	0.669
MAPE	0.647	0.647	0.647	0.647	0.647	0.647	0.647	0.647	0.647	0.647
MaxAPE	3.881	3.881	3.881	3.881	3.881	3.881	3.881	3.881	3.881	3.881
MAE	0.447	0.447	0.447	0.447	0.447	0.447	0.447	0.447	0.447	0.447
MaxAE	2.196	2.196	2.196	2.196	2.196	2.196	2.196	2.196	2.196	2.196
正态化的 *BIC*	−0.771	−0.771	−0.771	−0.771	−0.771	−0.771	−0.771	−0.771	−0.771	−0.771

表 7　昭通地区空气湿度 ARIMA 模型拟合结果

Table 7　Fitting results of ARIMA model of air humidity in Zhaotong area

Fitting statistics	Mean value	Minimum value	Maximum value	Percentile						
				5	10	25	50	75	90	95
平稳的 R^2	0.123	0.123	0.123	0.123	0.123	0.123	0.123	0.123	0.123	0.123
R^2	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993
RMSE	0.544	0.544	0.544	0.544	0.544	0.544	0.544	0.544	0.544	0.544
MAPE	0.452	0.452	0.452	0.452	0.452	0.452	0.452	0.452	0.452	0.452
MaxAPE	3.734	3.734	3.734	3.734	3.734	3.734	3.734	3.734	3.734	3.734
MAE	0.347	0.347	0.347	0.347	0.347	0.347	0.347	0.347	0.347	0.347
MaxAE	2.667	2.667	2.667	2.667	2.667	2.667	2.667	2.667	2.667	2.667
正态化的 *BIC*	−1.169	−1.169	−1.169	−1.169	−1.169	−1.169	−1.169	−1.169	−1.169	−1.169

该部分将对所涉及的云南省各地区的降水量、温度、空气湿度进行建模，已有模型对温度以及部分地区部分时段的降水量和空气湿度能够达到较好的预测效果，所以，只对已有模型拟合误差超过 10%的部分利用 ARIMA 模型进行修正，以得到一个新模型。根据软件计算得到的 R^2 值来判断新的模型性能，作为上述数学模型的有效补充和修正。

4 个 ARIMA 模型的拟合曲线如图 2—图 5 所示。

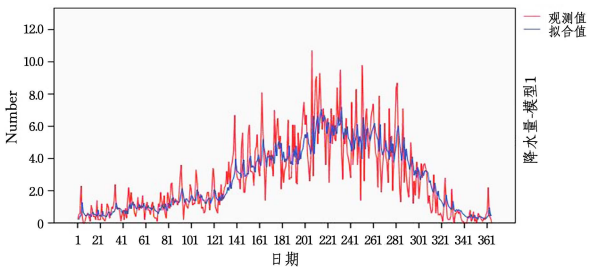

图 2 保山降水量 ARIMA 模型拟合曲线

Fig. 2 Fitting curve of Baoshan precipitation ARIMA model

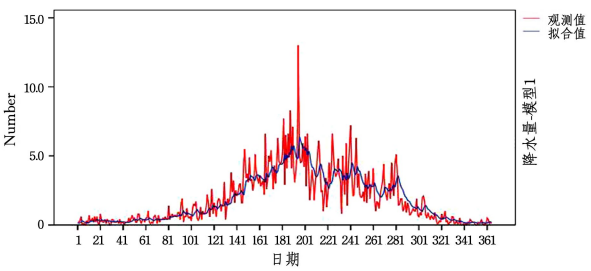

图 3 昭通降水量 ARIMA 模型拟合曲线

Fig. 3 Fitting curve of Zhaotong precipitation ARIMA model

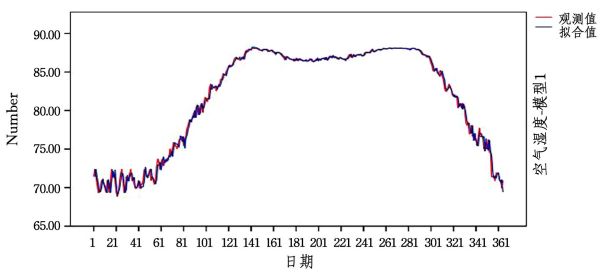

图 4 昆明空气湿度 ARIMA 模型拟合曲线

Fig. 4 Fitting curve of Kunming air humidity ARIMA model

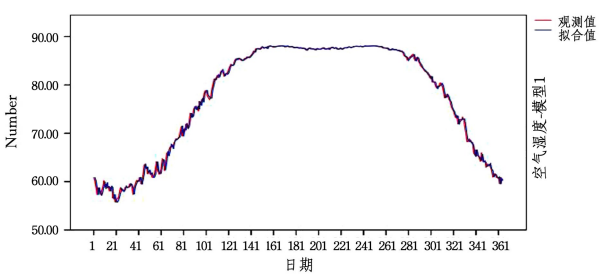

图 5 昭通空气湿度 ARIMA 模型拟合曲线

Fig. 5 Fitting curve of Zhaotong air humidity ARIMA model

由此可见，尤其在空气湿度部分，对使用已有模型误差很大的昆明地区和昭通地区使用 ARIMA 模型进行空气湿度建模，可以得到新的拟合模型，R^2 值全部在 0.99 以上，说明新的模型性能超过了前文所得的 Matlab 拟合模型，可以作为已有模型的有效补充。

结束语 本文研究结论如下：

(1)相对文献[19]使用的地理信息技术，文献[20]的农业气候资源数据特征分析和文献[21]介绍的农作物气候适宜性的动态模拟方法，本研究使用的数据科学方法利用气象观测数据进行数据分析，绘制云南省主要地区的温度、空气湿度、降水量图，建立用于预测本文的 3 项气候指标的数学模型，并分析其偏差和数据可信度。

(2)本研究首次对云南省的主要地区的 3 种气候指标进行基于 Matlab Curve Fitting Tool 和 ARIMA 模型的数据建模和数据预测。

(3)利用本研究的方法，可以对云南省各地区的降水量、温度、空气湿度进行建模，得到较为准确的预测模型，已有的模型 90%的预测误差在 10%以内，其中空气湿度模型 R^2 全部在 0.99 以上，可以为农业生产提供数据支持和气候指标保障。

相对文献[22]的跟进测验和文献[23]的全面准确数据，本研究还有不足，尤其是 ARIMA 模型的使用上，没能找到完善的模型使对降水量的预测误差全部控制在 10%以内，未来需要在模型互补上继续研究，使模型在实现快速计算的同时具有较高的精度。

参 考 文 献

[1] LIU D. An Empirical Study on the Effect of Industrial Structure Optimization on Economic Growth in Yunnan Province [D]. Kunming: Yunnan University of Finance and Economics, 2018.

[2] LIN YU, LI X L, JIANG H Q, et al. Discussion on Comparative Advantage and Distribution of Main Crop Production in Yunnan Province [J]. Exploration of Economic problems, 2005(5): 142-144.

[3] ZHANG C X. Study on Climate Suitability of Main Cropping Systems in South China under the Background of Climate Change [D]. Nanchang: Jiangxi Agricultural University, 2016.

[4] HUANG X F, SHAO D G, YANG S M. A Decomposition Prediction Model for Rainfall Time Series and Its Application [J]. China Rural Water Conservancy and Hydropower, 2007(9): 6-8.

[5] HUANG X F, SHAO D G, YANG S M. Application of Ordinal Set Pair Analysis in Annual Rainfall Prediction of Liao River Basin [J]. China Rural Water Conservancy and Hydropower, 2007(9): 6-8.

[6] ZHAO G C, ZHAO P F, WANG G Y. Study on Temporal and Spatial Distribution Characteristics and Prediction Model of Rainfall in Tianjin City[J]. Haihe Water Conservancy, 2019(4): 41-43.

[7] CHEN C, XU F X, PANG Y M, et al. Prediction of threshold temperature start date for rice at critical development stages in Southwest China [J]. Chinese Journal of Ecological Agriculture (in English and Chinese), 2019, 27(8): 1172.

[8] HE Q Q, GUO X H, LEI T, et al. Prediction of WinterSoil Temperature of Apple Orchard Under Water Storage Pit Irrigation Based on improved BP Neural Networks [J]. Water-saving Irrigation, 2019(7): 16-20.

[9] XU T Y, WANG T K, ZHANG X B, et al. Application of RBF

Neural Network in Humidity Simulation and Prediction of Northern Sunlight Greenhouse[J]. Journal of Shenyang Agricultural University，2014，45(6)：726-730.

[10] YAN M L. Characteristic Analysis and Prediction of Farmland Microclimate Time Series Data [D]. Tai'an：Shandong Agricultural University，2018.

[11] BAI Z W，ZHANG L，WANG J，et al. Study on meteorological drought prediction in Yunnan Province based on ARIMA[J]. People's Yangtze River，2015，46 (15)：6.

[12] DENG X Z，ZHAO C H，YUAN Y W. Analysis of the drought risk in Yunnan Province Based on the WRF model[J]. Population，Resources and Environment of China，2013，23(10)：95-101.

[13] LU X J，XIE H P，MU S L. Study on drought Climate Prediction of Yunnan Province based on Grey catastrophe Theory[J]. Mathematics In Practice And Theory. 2015，45(1)：164-168.

[14] JI W J，ZHANG M S，HU X Q. Characteristics of agroclimatic resources and their effects on agricultural production in Yunnan in 2016 [J]. Yunnan Agricultural Science and Technology，2017(5)：8-11.

[15] LI Z C，LIU S. Prediction comparison based on ARIMA model，grey model and regression model [J]. Statistics and decision-making，2019，35(23)：38-41.

[16] ZHANG T，PAN E S. Degradation Modeling of Capacitors Based on Time Series Analysis [J]. Journal of Shanghai Jiaotong University，2019，53(11)：1316-1325.

[17] ZHANG X W，GUO Z Y，CHEN L H. Brake temperature prediction method based on ARIMA model [J/OL]. Journal of Jilin University (Engineering Edition). [2020-01-19]. https：//doi.org/10.13229/j.cnki.jdxbgxb20190656.

[18] WANG D G，HUANG L，CHANG J，et al. Load forecasting model based on ARIMA and CART [J]. Journal of Shenzhen University (Science and Technology Edition)，2019，36(3)：245-251.

[19] WANG D. GIS-Based Rrgional Climatic Suitability for Potato Planting in Yunnan Province [D]. Kunming：Yunnan Agricultural University，2017.

[20] ZHU Y，LI S Y，ZHANG S B. Climate Resource and Adaptability of Rice and Winter-Sown Maize in Southern Subtropical Zone of Yunnan[J]. Journal of Nanjing Institute of Meteorology，2000，23(2)：299404.

[21] ZHAO F，QIAN H S. The research advances on the crop climate suitability influenced by global warming [J]. Chinese Journal of Ecological Agriculture，2004，12(12)：134-137.

[22] XU S T，HE X，YANG P W，et al. Effect of Climatic Factors on the Banana Growth and ItsFruit Quality at Different Altitudes [J/OL]. Journal of Yunnan Agricultural University (Natural Science). [2020-01-17]. http：//kns.cnki.net/kcms/detail/53.1044.S.20200115.0855.001.html.

[23] ZHAO P，LIN K H. Effect of Potassium Fertilizer on Quality of Crops [J]. Journal of Yunnan Agricultural University，2001，16(1)：56-59.

CHEN Pei，born in 2000，postgraduate. His main research interests include application of data science and big data technology in intelligent industry.

ZHENG Wan-bo，born in 1981，Ph.D，associate researcher. His main research interests include big data intersects with cloud computing，data analysis and mining，mine tunnel emergency management，as well as research on mine tunnel emergency rescue and disposal information technology and equipment，engineering geophysical exploration，measurement and control technology and instruments.

(上接第 481 页)

[7] TSAI C F，LIN W C，KE S W. Big data mining with parallel computing：A comparison of distributed and MapReduce methodologies[J]. The Journal of Systems & Software，2016，122(1)：83-92.

[8] LIU BAOLONG，SU JIN. Improved Canopy-Kmeans Algorithm based on Double-MapReduce[J]. Journal of Xi'an Technological University，2016，36(9)：730-737.

[9] CHEN J Y，HE H H. Research on Density-based Clustering Algorithm for Mixed Data with Determine Cluster Centers Automatically[J]. Acta Automatica Sinica，2015，41(10)：1798-1813.

[10] CAO H，JIA L，SI G，et al. A clustering-analysis-based membership functions formation method for fuzzy controller of ball mill pulverizing system [J]. Journal of Process Control，2013，23(1)：34-43.

[11] NING K，SUN T J，XU J J. Improved Nearest Neighbor Absorption First Clustering Algorithm for Massive Data[J]. Computer Engineering，2018，44(4)：35-40.

[12] HU X J，WEI C H. Order batching optimization based on Canopy and k-means algorithm[J]. Journal of Hefei University of Technology (Natural Science Edition)，2017，40(3)：414-419.

SHAO Xin-xin，born in 1980，postgraduate，assistant professor. Her main research interests include computer software and theory，and big data.