

ARIMA-SVR 组合模型在基于标准化 降水指数干旱预测中的应用

许德合², 张 棋¹, 黄会平²

(1. 华北水利水电大学地球科学与工程学院, 河南 郑州 450000; 2. 华北水利水电大学测绘与地理信息学院, 河南 郑州 450000)

摘 要:开展干旱预测是有效应对干旱风险的前提基础,本研究利用 1951—2017 年河南省郑州气象站点逐日降水量数据计算多尺度标准化降水指数(*SPI*),并建立了 *SPI* 序列自回归移动平均模型(*ARIMA*)和自回归移动平均与支持向量机回归组合模型(*ARIMA-SVR*),对模型参数进行率定和验证后,利用所建立的模型对河南省郑州气象站点多尺度 *SPI* 值进行预测。借助均方根误差(*RMSE*)、平均绝对百分比误差(*MAPE*)对回归预测模型的有效性进行判定。结果表明:*ARIMA-SVR* 组合模型在 *SPI*1(1 个月)和 *SPI*12(12 个月)的 *RMSE* 值分别为 80.05 和 0.74,均低于 *ARIMA* 模型的 92.25 和 1.24,说明 *ARIMA-SVR* 组合模型与单一的 *ARIMA* 模型对 *SPI* 的预测精度都与该指数的时间尺度长短有关,都随时间尺度的增加而逐渐提高;*SPI*12 的两种模型预测精度均高于 *SPI*1、*SPI*3(3 个月)和 *SPI*6(6 个月)的预测精度。用实测数据与模型的预测数据相比较说明 *ARIMA-SVR* 组合模型相比于单一 *ARIMA* 模型预测精度更高,且能够很好拟合不同时间尺度的标准化降水指数。

关键词:干旱预测;标准化降水指数;*ARIMA-SVR* 组合模型;*ARIMA* 模型;*SVR* 模型

中图分类号:S423 **文献标志码:**A

Application of the combined ARIMA-SVR model in drought prediction based on the Standardized Precipitation Index

XU Dehe², ZHANG Qi¹, HUANG Huiping²

- (1. College of Geosciences and Engineering, North China University of Water Resources and
Electric Power, Zhengzhou, Henan 450000, China;
2. College of Surveying and Geo-informatics, North China University of Water Resources and
Electric Power, Zhengzhou, Henan 450000, China)

Abstract: Carrying out drought prediction is the premise basis for effectively coping with drought risk. The multi-scale Standardized Precipitation Index (*SPI*) was calculated by using the daily precipitation data of Zhengzhou meteorological station in Henan Province from 1951 to 2017, and the *SPI* sequence autoregressive moving average model (*ARIMA*) and autoregressive moving average and support vector machine regression combined model (*ARIMA-SVR*) were established. After the model parameters were determined and verified, the multi-scale *SPI* value of Zhengzhou meteorological station in Henan Province was predicted by using the established model. The validity of the regression prediction model was determined by means of the root mean square error(*RMSE*) and the mean absolute percentage error (*MAPE*). The results showed that *RMSE* values of *ARIMA-SVR* combined model in *SPI*1 (1 month) and *SPI*12 (12 months) were 80.05 and 0.74, respectively, which were lower than 92.25 and 1.24 of *ARIMA* model, indicating that both the prediction accuracies of *SPI* of the *ARIMA-SVR* combined model and the single *ARIMA* model were related to the time scale of the index, and they gradually increased with the increase of time scale. The prediction accuracy of the two models of *SPI*12 was higher than that of *SPI*1, *SPI*3 (3

months), and SPI6 (6 months). Comparing the measured data with the predicted data of the model showed that the ARIMA-SVR combined model had higher prediction accuracy than the single ARIMA model, and can well fit the standardized precipitation index at different time scales.

Keywords: drought prediction; Standardized Precipitation Index (*SPI*); ARIMA-SVR combined model; ARIMA; SVR

旱灾被认为是世界上最严重的自然灾害类型之一^[1],极大程度地影响了人们的日常生活以及农业产量^[2]。干旱是指水分收支或供求不平衡形成的水分短缺现象^[3],全球气候变暖、碳排放量超标等问题将加剧未来农业干旱情况,严重威胁全球粮食生产,因此准确评估、监测、分析干旱情况一直是国内外学者的热门话题^[4]。对于干旱进行量化研究有助于研究干旱时空变化特征,提升干旱监测能力,开展干旱预报工作,寻求干旱治理及应对策略,对未来我国农业生产以及防旱抗旱等方面具有重要意义^[5]。我们通常选用便于计算的干旱指标来监测评估干旱发生的强度、持续时间和受灾范围^[6]。由于干旱指标种类多、运用范围广,且不同专业和学科对于干旱理解不同,因此出现了多种干旱指标^[7]。标准化降水蒸散指数(Standard Precipitation Evaporation Index, *SPEI*)、帕默尔干旱指数(Plamer Drought Severity Index, *PDSI*)、降雨 *Z* 指数(*Z* Index)、标准化降水指数(Standard Precipitation Index, *SPI*)、综合干旱指数(Colligation Drought Index, *CI*)等在气象干旱、农业干旱、水文干旱等领域已得到广泛应用^[8-12]。其中,*SPI* 是用于表征某时段降水量出现概率多少的指标,计算结果对于干旱分级精度相对较高,所需源数据少(仅利用日或月降水量数据就可进行计算),适用范围广,并且不同时间尺度的 *SPI* 值可以适用于不同类型的干旱。由于该指标使用灵活,易于计算,已成为实际应用最广泛且适用于所有气候状况的干旱指标^[13-16]。

加强干旱预测方面的研究对相关部门预防干旱灾害、减少干旱损失具有重要意义^[17]。常用来预测干旱的模型有很多,如人工神经网络(Artificial Neural Network, ANN)、支持向量机(Support Vector Machine, SVM)和差分自回归移动平均模型(Autoregressive Integrated Moving Average, ARIMA)等。其中 ARIMA 模型是时间序列中常用的模型,通常用来预测线性数据;SVM 模型是一种二分类模型,通常用来处理非线性数据;支持向量回归机(Support Vector Regression, SVR)是 SVM 的一种拓展,多用来进行非线性数据的回归预测,而 ARIMA 与 SVR 组合模型分别对线性模型及非线性模型处

理具有优势,所以它们之间存在优势互补^[18]。

河南省位于中国中部和黄河中、下游,是中国重要的粮食作物产区和农业大省,对于保障国家粮食安全发挥着至关重要的作用^[19]。本研究以河南省国家级气象观测站郑州站为例,计算不同时间尺度的 *SPI* 值,利用 ARIMA 模型及 ARIMA-SVR 组合模型对其进行预测,并采用均方根误差(Root Mean Squared Error, *RMSE*)和平均绝对百分比误差(Mean Absolute Percentage Error, *MAPE*)对 2 种模型预测能力进行分析。

1 研究方法

1.1 *SPI* 的计算

降水量通常是一种偏态分布,常采用 Γ 分布概率描述降水量的变化,再将偏态概率分布进行正态标准化处理,最后用标准化降水累计频率分布划分干旱等级^[20-22]。*SPI* 可以量化研究多时间尺度的降水量不足。

SPI 指数公式为^[23-24]:

$$SPI = G \frac{t - (u_2 t + u_1) t + u_0}{[(l_3 t + l_2) t + l_1] t + 1} \quad (1)$$

$$t = \sqrt{\ln \frac{1}{Y(x)^2}} \quad (2)$$

式中, $Y(x)^2$ 为与 Γ 函数相关的降水量分布概率; x 为样本值(即降水量); G 为正负系数; u_0 、 u_1 、 u_2 和 l_1 、 l_2 、 l_3 为常数:

$$u_0 = 2.515517, u_1 = 0.802853, u_2 = 0.010328$$

$$l_1 = 1.432788, l_2 = 0.189269, l_3 = 0.001308$$

当 $Y(x) > 0.5$ 时, $G = 1$, 当 $Y(x) \leq 0.5$ 时, $G = -1$ 。

$Y(x)$ 由 Γ 函数求得,其中 Γ 为概率密度积分公式:

$$Y(x) = \frac{2}{\beta \gamma \Gamma(\gamma_0)} \int_0^x x^{\gamma-1} e^{-\frac{x}{\beta}} dx, \quad x > 0 \quad (3)$$

式中, γ 、 β 为 Γ 分布函数的形状和尺度参数。

1.2 ARIMA 模型

ARIMA 分为自回归模型(Autoregressive model, AR)、滑动平均模型(Moving average model, MA)以及自回归移动平均模型(Autoregressive Moving Average Model, ARMA),是传统的时间序列预测模

型。其建模流程是首先判断模型平稳程度,其次利用差分法对非平稳时间序列进行平稳化处理,然后选取 $AR(p)$ 、 $MA(q)$ 对模型进行定阶,差分次数记为 d , $ARIMA(p,d,q)$ 模型就是经过了 d 阶差分后的 $ARMA(p,q)$ 模型。如下式所示^[25]:

$$\phi(L)x_t = \theta(L)u_t \tag{4}$$

式中, $\phi(L)$ 和 $\theta(L)$ 分别为:

$$\begin{aligned} \phi(L)x_t &= (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)x_t \\ \theta(L)x_t &= (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)x_t \end{aligned} \tag{5}$$

$ARIMA(p,d,q)$ 模型的一般式:

$$\begin{aligned} w_t &= \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} \\ &+ \phi + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q} \\ \phi(L)\Delta^d x_t &= \theta(L)u_t \end{aligned} \tag{6}$$

上式中, d 为差分次数, $\Delta = 1 - L$, $(\phi_1, \phi_2, \cdots, \phi_p)$ 为自回归系数, p 为自回归阶数, $(\theta_1, \theta_2, \cdots, \theta_p)$ 为移动平均系数; u_t 为白噪声序列(服从 0 均值、正态分布且相互独立的白噪声序列)。

其中 p 、 q 阶数采用赤池信息准则(Akaike Information Criterion, AIC)和贝叶斯信息准则(Bayesian Information Criterion, BIC)来确定,当样本数 N 固定时,选择 AIC 和 BIC 最小值来确定 p 、 q 。公式如下:

$$\begin{aligned} AIC(p,q) &= N\ln\sigma^2(p,q) + 2(p+q+1) \\ BIC(p,q) &= N\ln\sigma^2(p,q) + (p+q+1)\ln N \end{aligned} \tag{7}$$

1.3 SVR 模型

SVR 模型是 SVM 的推广,SVR 的本质属性不再是原来的二分类方法,而是回归方法。由于 SVR 模型在对非线性数据预测方面具有优势,因此采用 ARIMA 模型对线性数据 SPI 进行预测,将所得残差(非线性数据)传入 SVR 模型,再利用 SVR 模型对残差进行预测。通过引入径向基核函数(RBF)来把训练样本映射到高维空间下来进行回归预测,再将回归问题转为优化问题^[26-28]:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*), \\ \text{s.t.} \quad & \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i, \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{8}$$

式中, w 为权重系数, ξ_i, ξ_i^* 为松弛变量,规定了模型的误差要求; C 为惩罚参数, C 越大则支持向量的决策边界越大; b 为偏置项。RBF 公式中 σ 和 γ 关系如下:

$$\begin{aligned} K(x_i, y_i) &= \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) = \exp(-\gamma \cdot d \\ &(x_i, x_j)^2), i = 1, 2, \cdots, n \end{aligned} \tag{9}$$

式中, $K(x_i, y_i)$ 为 RBF 核函数, σ 为核密度。

$\gamma = \frac{1}{2 \cdot \sigma^2}$,因此 γ 越大, σ 越小,使得 RBF 宽幅越窄。

1.4 ARIMA-SVR 组合模型

由于 ARIMA 模型和 SVR 模型在线性和非线性预测中各有优点,因此本文分别采用 ARIMA 与 SVR 模型的优点建立组合模型 ARIMA-SVR,假设时间序列 Y_t 可视为线性自相关部分 L_t 与非线性残差 N_t 两部分的组合,利用 ARIMA 模型对 SPI 值进行预测,将结果与实际值相减得到残差,将残差记为非线性部分,带入 SVR 模型进行预测,最后把两者预测结果相加得到组合结果,即:

$$Y_t = L_t + N_t \tag{10}$$

1.5 评价验证指标

在回归模型当中,平均绝对误差(Mean Absolute Error, MAE)、均方误差(Mean Square Error, MSE)、RMSE 和 MAPE 是常见的回归预测评估指标,其中 MAE 和 MSE 是最基础的评估方法, RMSE 和 MAPE 是回归任务最常用的性能度量,是前两种指标的扩展,不但在多场景下可以使用,而且相对前两种更加准确,因此本文采用 RMSE 和 MAEP 作为模型评价的指标。

RMSE 均方根误差

$$RMSE = \sqrt{(x_i - y_i)^2} \tag{11}$$

MAPE 平均绝对百分误差

$$MAPE = \sum_{i=1}^N \left| \frac{x_i - y_i}{x_i} \right| \times \frac{100}{N} \tag{12}$$

式中, x_i 是观测值, y_i 是预测值, N 是样本数。RMSE 和 MAPE 越接近 0,表示预测值与观测值越接近。

2 实例应用

2.1 数据资料及 SPI 计算

本文选用 1951—2017 年河南省国家级气象站郑州站逐日降水量数据来进行计算,原始数据来源于国家气象信息中心提供的中国地面气候资料日值数据集。运用 Matlab 数学建模软件编写 SPI 计算程序,分别计算了 1951—2017 年的 1、3、6、12 个月共 4 个尺度的 SPI 值,记为 SPI1, SPI3, SPI6, SPI12,并通过国家标准气象干旱等级(GB/T20481-2006)规定的干旱分级标准(表 1)来表征干旱情况^[29]。

2.2 时间序列生成及建模流程

本文使用 Python 3.6 平台对 ARIMA 进行建模,并利用 Python 中 matplotlib 可视化库对多尺度 SPI

计算结果进行可视化展示,如图 1 所示。

2.2.1 平稳化处理及 ARIMA 模型定阶 由于 ARIMA 是经过 d 次差分的平稳时间序列 ARMA 模型,且通常针对平稳时间序列进行建模,因此在建模前首先应对时间序列的平稳性进行判断,本文采用观察时间序列的时序图和单位根检验 (Augmented Dickey-Fuller Test, ADF) 来判断平稳性。

表 1 标准化降水指数干旱分级

Table 1 Drought classification based on SPI		
等级 Level	类型 Type	SPI 范围 Range of SPI
1	无旱 No drought	$SPI \geq -0.5$
2	轻旱 Mild drought	$-1.0 \leq SPI < -0.5$
3	中旱 Moderate drought	$-1.5 \leq SPI < -1.0$
4	重旱 Severe drought	$-2.0 \leq SPI < -1.5$
5	特旱 Extreme drought	$SPI \leq -2.0$

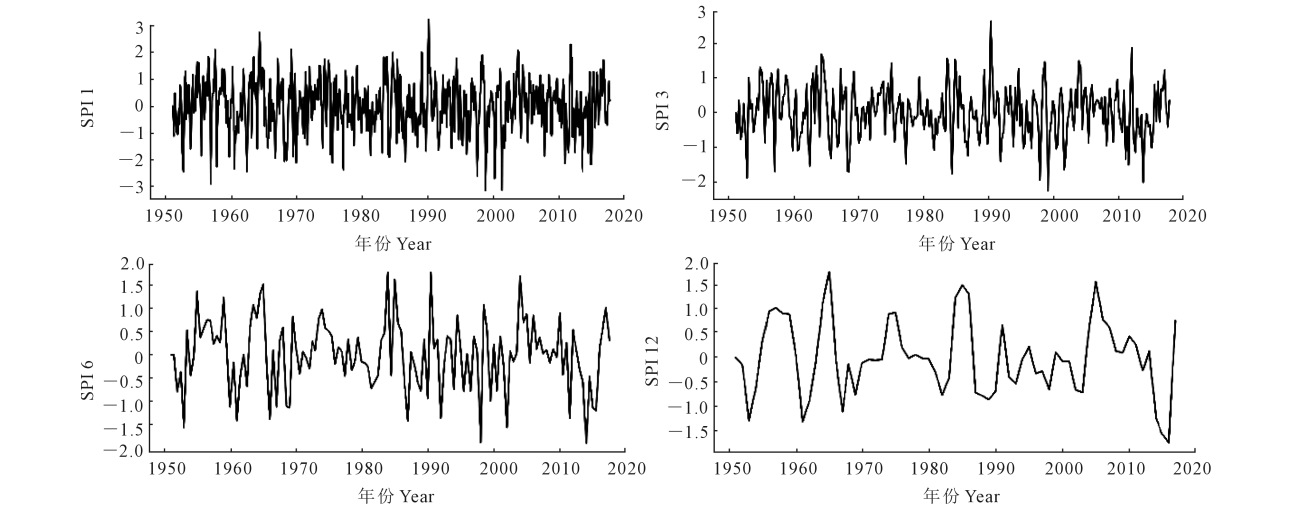


图 1 1951—2017 年郑州站多时间尺度 SPI 变化趋势

Fig.1 Trend charts of multi-time scale SPI of Zhengzhou Station from 1951 to 2017

表 2 原始序列 SPI12 单位根检验

Table 2 Unit root test of original sequence SPI12				
单位根检验 Augmented dickey- fuller test	临界值 C			P
	1%	5%	10%	
-4.02418	-3.56024	-2.91785	-2.59679	0.36753

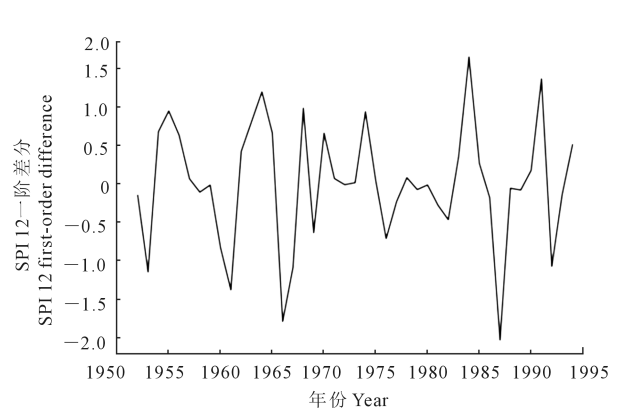


图 2 一阶差分 (SPI12)

Fig.2 First-order difference of SPI12

由图 1 可得, SPI1、SPI3、SPI6 序列无明显的上升和下降趋势, SPI12 序列略有上升趋势, 进一步对 SPI1、SPI3、SPI6 和 SPI12 进行 ADF 检验, 在 ADF 检验中, 原假设为非平稳时间序列且存在单位根, 给定显著水平 $\alpha=0.05$, 如果检验统计量对应的概率值 $P<0.05$, 则拒绝原假设。其中 SPI1、SPI3、SPI6 的 ADF 检验 P 值均小于 0.05, SPI12 的 P 值为 0.36753 且显著大于 0.05, 检验结果见表 2, 因此判断 SPI12 序列为非平稳时间序列 (非平稳时间序列一定不是白噪声序列), SPI1、SPI3、SPI6 为平稳时间序列。利用差分法对 SPI12 非平稳时间序列进行平稳化处理, 经一阶差分后时间序列趋于平稳 (图 2), 再对一阶差分后的时间序列进行单位根检验 (结果见表 3)。

表 3 一阶差分后的 SPI12 单位根检验

Table 3 Unit root test after first order difference of SPI12				
单位根检验 Augmented dickey-fuller test	临界值 C			P
	1%	5%	10%	
-6.28749	-3.59663	-2.93329	-2.60499	0.00128

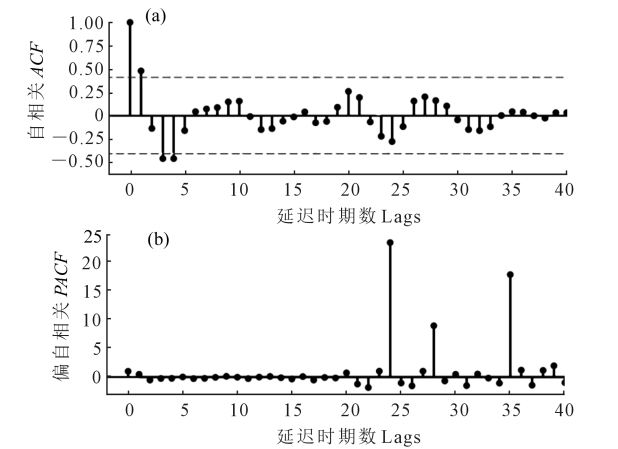
表 3 中 P 值为 0.00128, 小于 0.05, 因此差分之后为平稳时间序列。采用纯随机性检验 (Ljung-Box Test) 进行白噪声检验, 得 P 值为 0.006025, 远小于 0.05, 因此经一阶差分后该序列为平稳非白噪声序列。然后用自相关函数 (Autocorrelation Function, ACF) 及偏自相关函数 (Partial Autocorrelation Function, PACF) 来为 ARMA 模型定阶 (图 3)。这里采用观察法为模型 ARIMA(p, d, q) 定阶, 由于图 3 中 ACF 在二阶之后均落在置信区间上, 所以直接定 $q=2$ 或 1 或 0; 而 PACF 图置信区间较小, 看到在二阶后落在置信区间上, 但最后几阶则在置信区间外, 因此需要通过时间序列相关性来判断, 判断结果如

图 4 所示。由图 4 可见,曲线在横轴 0 处相关性最强,最接近 0.5,在 2 处小于-0.25,即 p 可取的值有 0 和 2。由于 d 取 1,且 q 可取 3 个值, p 可取 2 个值,所以 $ARIMA(p,d,q)$ 模型有 6 个可取值。

2.2.2 ARIMA 模型参数估计及适用性检验 对 $ARIMA(p,d,q)$ 的 6 个可取值分别进行计算,将相关检验结果汇总(表 4)。本文采用 AIC、BIC 准则和标准误差来选取最优模型,由表 4 可得, $ARIMA(2,1,0)$ 的 AIC、BIC 和标准误差值最小,因此判定为最优模型,即一阶差分后的 $AR(2)$ 模型为最优模型。对 $ARIMA(2,1,0)$ 进行参数估计(结果见表 5),可得 $ARIMA(2,1,0)$ 模型的具体形式:

$$(1+0.281B)(1-B)X_t=U_t$$

对一阶差分后的序列拟合 $AR(2)$ 模型进行残差检验,以评价所建模型的稳定程度,在这里选择 QQ 残差图及正态分布图来检验模型残差是否是平均值为 0 且方差为常数的正态分布,QQ 残差图中散



注:图(a)中虚线所示范围为置信区间。下同。
Note: The dotted line in the figure shows the confidence interval. The same below.

图 3 ACF 和 PACF (SPI12)
Fig.3 Autocorrelation function and partial autocorrelation function of SPI12

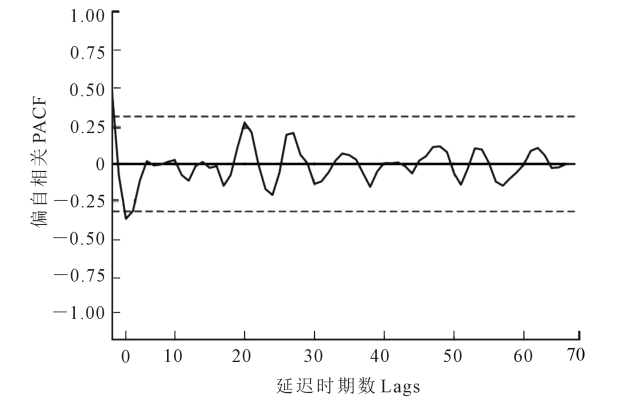


图 4 PACF 折线图 (SPI12)
Fig.4 Line chart of PACF (SPI12)

点均落在拟合直线附近(见图 5),且正态图残差曲线也满足正态分布(见图 6),再通过 Ljung-Box 检验,得到 p 值为 0.88566,远大于 0.1,综合图 5 和图 6,明显符合白噪声序列特征,说明该模型适用于拟合与预测 SPI12 的变化趋势。选取 1951—1995 年数据作为训练集,1996—2017 年数据作为测试集,应用 $ARIMA(2,1,0)$ 来对 SPI12 进行 22 a 预测(1996—2017),对于 SPI1、SPI3、SPI6 的 $ARIMA$ 建模流程同 SPI12,选定模型分别为:SPI1(0,0,2)、SPI3(2,0,0)和 SPI6(1,0,0),预测结果见图 7。

表 4 模型参数检验结果

项目 Item	ARIMA(p,d,q)					
	(0,1,0)	(0,1,1)	(0,1,2)	(2,1,0)	(2,1,1)	(2,1,2)
AIC	159.4	161.1	142.9	137.1	158.7	158.6
BIC	163.8	167.6	151.6	148.1	167.4	162.3
标准误差 Standard error	0.791	0.789	0.677	0.639	0.763	0.753

表 5 ARIMA(2,1,0) 模型参数

Table 5 Parameters of ARIMA (2, 1, 0)				
模型 Model	系数 Coefficient	标准差 Std. error	Z	P
AR(2)	-0.281	0.129	-2.180	0.013

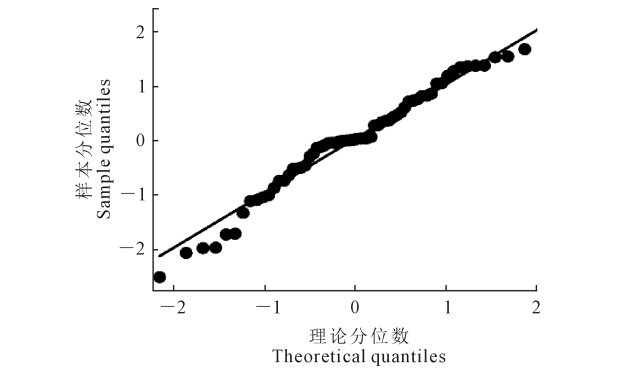


图 5 QQ 残差图
Fig.5 Residual chart of QQ model

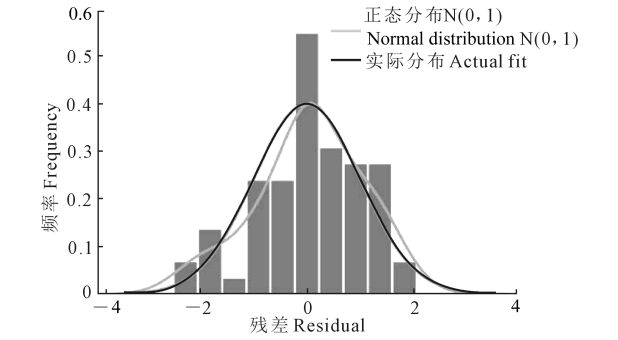


图 6 残差正态分布
Fig.6 Normal distribution map of residual

2.2.3 SVR 模型残差定阶及参数寻优 SVR 模型对 ARIMA 模型残差预测之前,首先要知道过去几个时期的残差会对下一个时期的残差产生的影响,即 SVR 模型残差定阶。从起始时间开始每次取 k 个按时间顺序排列的残差数据,并依次排列下来,将排列好的矩阵作为 SVR 模型的输入, $k+1$ 个数据作为模型的输出并保留误差。利用交叉验证的方法,随机取 80% 的数据作为模型的训练集,其余作为测试集,利用 SVR 模型训练后对测试集的结果进行预测。在选定阶数为 k 的情况下,若 $RMSE$ 第 k 阶小于第 $k+1$ 阶的时候,停止循环并输出 k ;反之,则继续增加阶数。对于 SVR 模型中的惩罚参数 C 和径向基函数(Radial basis function, RBF)中的参数 γ 采用网格寻优(Grid Search, GS)算法进行率定,并完成 SVR 模型的建模。以 1951—1995 年作为训练集,1996—2017 年作为测试集,利用 SVR 模型对 ARIMA 模型所预测的 SPI12 值的残差进行预测,结果见图 7。

2.2.4 组合模型的预测与检验 在 ARIMA 模型的基础上加入 SVR 模型进行组合预测时通常有两种方式,一种是并联型,一种是串联型,并联型是通过两种模型预测结果分别加上权重后重组得到;串联型则是用 ARIMA 模型进行预测后所得到的残差值输入 SVR 模型再进行预测,将 SVR 输出的残差修正值和 ARIMA 预测值组合得到最终组合模型。两种模型在线性和非线性预测中各有优势,并且 ARIMA 模型随着时间增加,预测结果越趋于平稳,在用并列组合方式给 ARIMA 模型权重赋值时无法随着时间长度的增加而改变。因此本文采用串联型进行组合预测,选取 1951—1995 年 4 个时间尺度 SPI 值作为训练集,1996—2017 年作为测试集,将 ARIMA 模型与 ARIMA+SVR 组合模型预测结果与实际 4 个时间尺度 SPI 值进行比对,结果见图 8。

利用 $RMSE$ 和 $MAPE$ 两种评价指标对 4 个时间尺度的两种模型预测结果进行评价,结果见表 6。

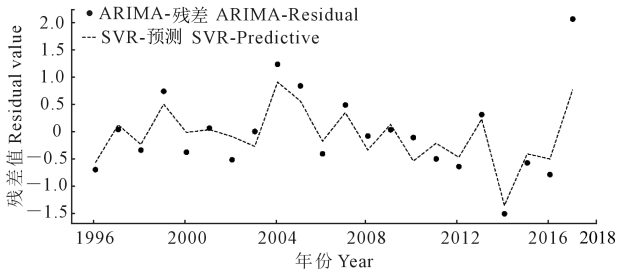


图 7 SVR 模型残差预测 (SPI12)

Fig.7 Residual prediction of SVR model (SPI12)

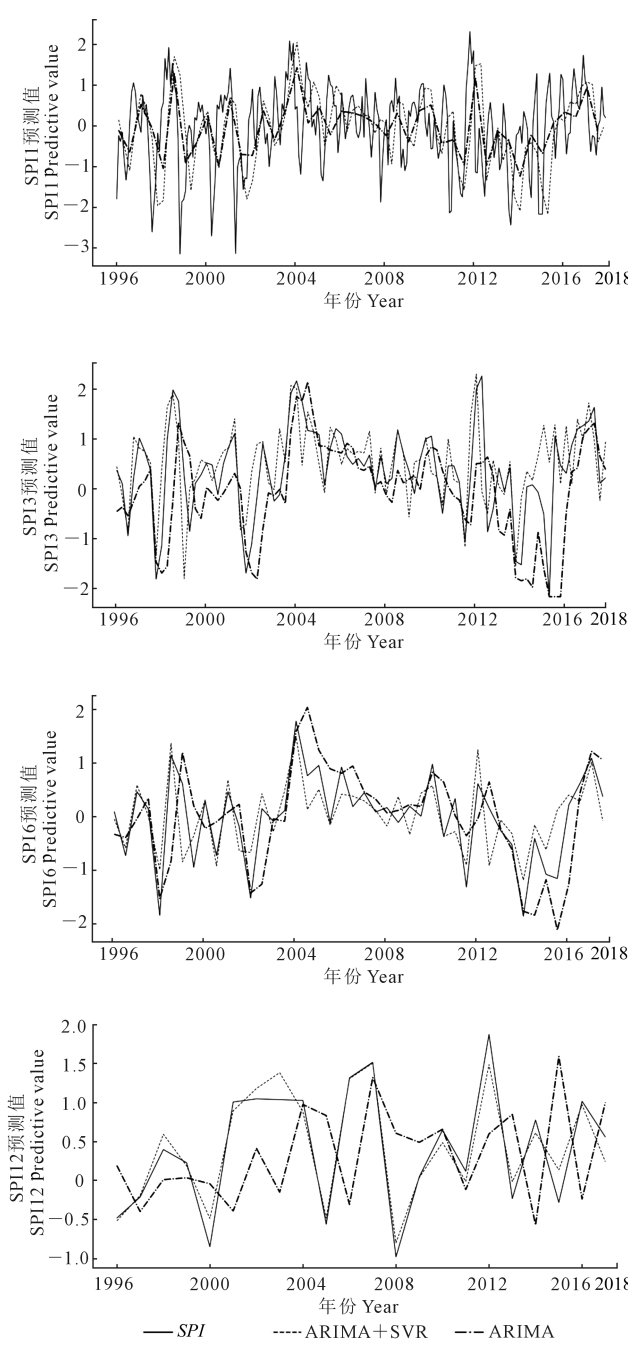


图 8 ARIMA 模型与 ARIMA-SVR 组合模型的多时间尺度 SPI 值的预测 (1996—2017)

Fig.8 Prediction of multi time-scale SPI based on ARIMA and ARIMA-SVR models

表 6 ARIMA 模型与 ARIMA-SVR 组合模型的均方根误差和平均绝对百分误差

Table 6 $RMSE$ and $MAPE$ values for ARIMA and ARIMA-SVR model

时间尺度 Time-scale	ARIMA		ARIMA+SVR	
	$RMSE$	$MAPE$	$RMSE$	$MAPE$
SPI1	92.25	68.56	80.05	13.68
SPI3	55.45	25.98	43.87	6.88
SPI6	12.35	21.06	18.68	0.85
SPI12	1.24	48.68	0.74	1.97

表中数据可以发现两种模型在 SPI1 时间尺度的预测效果最差,ARIMA 模型在 SPI12 尺度预测效果最好,在 SPI3 和 SPI6 时间尺度预测效果仅次于 SPI12。组合模型在各时间尺度预测效果都比单独 ARIMA 模型预测效果好,在 SPI12 预测效果最好,随着时间尺度减小,预测精度降低。

3 讨论与结论

本文引入了反映干旱强度和持续时间的 SPI 指数,以郑州市气象站点为例,利用 ARIMA 模型和 ARIMA-SVR 组合模型对不同时间尺度的 SPI 序列进行建模预测,并利用 RMSE 和 MAPE 两种回归模型预测指标进行评价,得到如下结论:

1)从 ARIMA 模型的预测结果来看,该模型对较长时间尺度的预测精度较高,对较短时间尺度的预测精度较低,随着时间尺度增加,拟合精度提高。对于 SPI12 预测效果最好,对 SPI1 预测效果最差,主要由以下原因导致:ARIMA 模型本质上是一种整体线性自回归模型,该模型预测趋势会随着测试集时间增长而逐渐趋于平稳。因为 SPI1 相对于其他 3 个时间尺度数据量最多,整体趋于严平稳(严平稳表示的分布不随时间的改变而改变),所以预测精度最低,同理,SPI3、SPI6 和 SPI12 时间尺度逐渐增加,数据量逐渐减少,越来越趋于弱平稳(期望与相关系数不变,未来时刻值依赖于过去时刻的值),因此拟合精度整体逐渐提高。

2)从组合模型的预测结果来看,对多尺度 SPI 值利用 ARIMA 模型预测线性部分,用 SVR 模型预测非线性部分,叠加在一起的组合模型在各个时间尺度的预测精度均比单一 ARIMA 模型的预测精度高。从多时间尺度来看,该模型预测精度随着时间尺度的增加而提高,对 SPI12 预测效果最佳,主要是由于 ARIMA 模型对 SPI12 预测精度最高,残差最小,所以传入 SVR 模型中的误差就小,最后组合预测结果也更加准确。同理,随着时间尺度的减小,SPI6、SPI3 和 SPI1 的 ARIMA 模型预测精度也降低,则组合模型的预测精度也随之降低。

参 考 文 献:

[1] 夏露,宋孝玉,马细霞.新乡市近 60 年降水序列变化规律及干旱预测[J].干旱地区农业研究,2013,31(5):14-18.

[2] 韩海涛,胡文超,陈学君,等. 三种气象干旱指标的应用比较研究[J]. 干旱地区农业研究,2009,27(1):237-247.

[3] 李凤霞,伏洋,张国胜,等. 青海省干旱预警服务系统设计与建立[J].干旱地区农业研究,2004,22(1):1-5.

[4] 袁文平,周广胜. 干旱指标的理论分析与研究展望[J]. 地球科学进展,2004,19(6):982-991.

[5] 史本林,朱新玉,胡云川,等. 基于 SPEI 指数的近 53 年河南省干旱时空变化特征[J]. 地理研究,2015,34(8):1547-1558.

[6] Vasiliades L, Loukas A, Liberis N. A water balance derived drought index for Pinios River Basin, Greece [J]. Water Resource Manage, 2011, 25:1087-1101.

[7] 容锦盟,周丹,罗静,等. 4 种干旱指标在华北地区气象干旱监测中的适用性分析[J].干旱地区农业研究,2019,37(1):295-276.

[8] 刘庚山,郭安红,安顺清,等. 帕默尔干旱指标及其应用研究进展[J]. 自然灾害学报,2004,13(4):21-27.

[9] 林盛吉,许月萍,田烨,等. 基于 Z 指数和 SPI 指数的钱塘江流域干旱时空分析[J].水力发电学报,2012,31(2):20-26.

[10] 邹旭恺,任国玉,张强. 基于综合气象干旱指数的中国干旱变化趋势研究[J]. 气候与环境研究,2010,15(4):371-378.

[11] 李树岩,刘荣花,马志红. 基于降水距平的黄淮平原夏玉米干旱评估指标研究[J]. 干旱地区农业研究,2012,30(3):252-256.

[12] 李星敏,杨文峰,高蓓,等. 气象与农业业务化干旱指标的研究与应用现状[J]. 西北农林科技大学学报(自然科学版),2007,35(7):111-116.

[13] 郭伟,李莹,杜莉莉,等.基于 SPI 的山西省 1972—2012 年春夏干旱特征及对玉米产量的影响分析 [J]. 干旱地区农业研究,2018,36(1):230-236.

[14] McKee T B, Doesken N J, Kleist J. The relationship of drought frequency and duration to time scales [R]. California; Eighth Conference on Applied Climatology, American Meteorological Society, 1993.

[15] 姚瑶,张鑫,马全,等. 青海省东部农业区作物生长期不同气象干旱指标应用研究[J]. 自然灾害学报,2014,23(4):177-184.

[16] 许小峰.中国气象灾害年鉴[M]. 北京:气象出版社,2014.

[17] 张玉虎,向柳,孙庆,等.贝叶斯框架的 Copula 季节水文干旱预报模型构建及应用 [J].地理科学,2016,36(9):1437-1444.

[18] 王佳敏,张红燕.基于 ARIMA-SVR 组合模型的移动通信用户数预测[J].计算机时代,2014,(1):12-17.

[19] 刘晓璐,周延刚,温莉,等.基于 VSWI 和 SPI 的 2000—2016 年河南省干旱特征研究[J].干旱区地理,2018,41(50):984-991.

[20] 杨晓静,左德鹏,徐宗学.基于标准化降水指数的云南省近 55 年旱涝演变特征[J].资源科学,2014,36(3):473-480.

[21] 翟禄新,冯起. 基于 SPI 的西北地区气候干湿变化[J]. 自然资源学报,2011,26(5): 847-857.

[22] 陈莹,陈兴伟. 福建省近 50 年旱涝时空特征演变—基于标准化降水指数分析[J].自然灾害学报,2011,20(3):57-63.

[23] 袁云,李栋梁,安迪.基于标准化降水指数的中国冬季干旱分区及气候特征[J].中国沙漠,2010,30(4):917-925.

[24] 黄晚华,杨晓光,李茂松,等. 基于标准化降水指数的中国南方季节性干旱近 58a 演变特征[J].农业工程学报,2010,26(7):50-59.

[25] 魏丽华,任志远.陕西省人均生态足迹 ARIMA 模型测算[J].干旱地区农业研究,2009,27(4):245-248.

[26] Vapnik V N, Chervonenkis A Y. On the uniform convergence of relative frequencies of events to their probabilities [J]. Theory of Probability & Its Applications,1971, 16(2):264-280.

[27] 王霞,万占岐,金贵,等.基于核函数支持向量回归机的耕地面积预测[J]. 农业工程学报,2014,30(4):204-211.

[28] 梁昌勇,马银超,陈荣,等.基于 SVR-ARIMA 组合模型的日旅游需求预测[J].管理工程学报,2015,29(1):122-127.

[29] 中华人民共和国国家质量监督检验检疫总局. GB/T 20481—2006 中华人民共和国国家标准:气象干旱等级 [S]. 北京:中国标准出版社,2006.