

AR 模型在土壤水分动态模拟中的应用

刘洪斌¹, 武 伟², 魏朝富¹, 谢德体¹

(1. 西南农业大学资源与环境学院, 重庆 400716; 2. 西南农业大学信息学院, 重庆 400716)

摘 要: 土壤水分运动是一个复杂的时间序列系统, 其变化与区域气候条件和生态环境密切相关, 具有明显的随机性波动。建立土壤水分动态变化模型可以使田间土壤水分的适时适量调节方便可行, 对于农田水利工程的规划和管理也有利。利用时间序列分析方法进行丘陵地区紫色土土壤水分含量变化的预测建模研究, 应用 AR 时间序列模型进行土壤水分含量的预测, 结果表明 AR 模型能够较好地拟和及预测土壤水分时间序列。

关键词: AR 模型; 土壤水分; 时间序列; 预测

中图分类号: S157; TP 18

文献标识码: A

土壤水分运移建模及预测预报是近代农业水土工程生产管理的重要依据。土壤水分含量受气候条件的影响, 具有较大的随机性。时间序列分析主要是采用参数模型对所观测到的有序的随机数据进行分析与处理的一种方法, 它所建立的模型是随机模型。目前, 时间序列分析的方法主要应用在医学、金融及各种工程应用领域^[1]。在土壤水分动态模拟方面, 在北方气候条件下康绍忠^[2]对较厚土层用提取周期趋势的方法作了土壤水分动态的随机模拟研究, 取得了一定的效果。

三峡库区生态环境脆弱, 伏旱严重, 常常是夏旱连伏旱。长江河谷是伏旱重灾区, 伏旱发生频率为 80%~ 90%, 持续时间 30~ 50 d, 对大春作物的播种和生长产生严重影响。加上土壤蓄水性能差, 丘陵地区土层薄弱, 抗旱能力弱, 土壤耕层的含水量不能满足作物的需要, 严重影响农业生产, 导致农牧业经济效益低。因此进行该地区的土壤水分动态预报研究极为重要。本文采用时间序列自回归建模方法对紫色土丘陵旱坡地土壤水分动态进行模拟和预测, 从数理统计的角度探讨紫色土水分运动的规律, 试图在紫色土水分随机建模、水分预报及探索土壤干旱原因等方面作一些尝试。

1 自回归模型

时间序列是指按时间顺序排列的一组数据, 或一组有序的随机数据。时间序列分析是采用参数模型对所观测到的有序的随机数据进行分析与处理的一种数据处理方法, 时间序列方法将所观测到的时序作为系统的一维或多维输出, 将模型所描述的等价系统视为在与输出同维白噪声驱动下产生这一输出的系统^[3~ 5]。

1.1 自回归模型定义

设由等时段(或等间隔)的 n 个样本组成的平稳、正态、零均值时间序列 $x_t (t = 1, 2, \dots, n)$ 。当 $k > p$ 时, 若其偏相关函数 $\varphi_{k, k} \approx 0$, 则序列 x_t 符合自回归模型(Autoregression)

$$\phi(B)x_t = a_t \quad (1)$$

或

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p} = a_t \quad (2)$$

式中 $\phi_1, \phi_2, \dots, \phi_p$ 为自回归系数, a_t 为残差。

AR 模型描述了变量 x 在 t 时刻的值与在 $t-1, t-2, \dots, t-p$ 时刻的值之间的统计关系, 自回归系数直观地反映了变量 x 在 t 时刻值与在 $t-1, t-$

收稿日期(Received date): 2003- 03- 14; 改回日期(Accepted): 2003- 06- 05。

基金项目(Foundation item): 重庆市科委攻关项目(6217)资助。[Supported by the Key Project of Chongqing Science and Technology Commission, (No. 6217)].

作者简介(Biography): 刘洪斌(1966-), 男, 重庆人, 博士, 副研究员, 主要从事土壤学、遥感、农业信息化研究等方面的工作。联系电话: (023)68251069(0)。[Liu Hongbin(1966-), Male, born in Chongqing, Associate Researcher in Southwest Agriculture University, Ph. D, his present research interest focuses on soil science, remote sense and agriculture information etc.]

2, ... t-p 时刻值之间的依赖权重。

1.2 AR 模型的参数估计

对于参数估计的前提是所获取的时间序列是平稳、正态分布的序列。对于不平稳的、带有某些趋势的时间序列不能够直接采用 AR 建模方法, 必须经过适当的变换去除其中的趋势变化, 获得平稳时间序列, 然后再进行建模。

时间序列 AR 模型的参数估计, 就是选择合适的参数使得模型的残差 a_t 为白噪声。常用的有矩估计、极大似然估计、最小平方和估计及最小二乘估计等方法, 本研究采用最小二乘方法估计模型参数。具体计算过程见文献[3]和[5]。

1.3 AR 模型的检验

本文使用残差的白噪声检验准则及 AIC 准则 (An Information Criterion) 进行 AR 模型的检验。AIC 准则又称为信息准则, 其计算公式为

$$AIC(p) = N \ln \sigma_a^2 + 2p \quad (3)$$

式中 p 为模型阶次, N 是时序 $\{x_t\}$ 的长度, 当 $\{x_t\}$ 是平稳、正态分布的时间序列时

$$\sigma_a^2 = \frac{1}{N} \sum_{t=1}^N a_t^2 \quad (4)$$

式中 N 为数据个数, a_t 为残差。参数估计是在给定阶次的情况下进行的, 由于事先无法判断模型的阶次, 因此在建模过程中先给定模型的阶次, 然后按照最小二乘法估计出 AR 模型的参数, 得到各阶模型, 最后取 AIC 值最小的阶次作为模型的最佳阶次, 同时也确定了 AR 模型。

2 研究区简介及数据收集

2.1 研究区简介

本研究在西南农业大学试验场进行。该场地处 $106^{\circ}26' E, 30^{\circ}26' N$, 海拔 230 m。属亚热带湿润气候, 年平均温度 $18.3^{\circ} C$, 年际间平均气温变化不大。雨量在季节分配上极不均匀, 多集中在 5~9 月, 占年降雨量的 70%。实验区为丘陵旱坡地, 面积 0.1 ha, 土壤为灰棕紫泥土。土壤基本性质见表 1。

表 1 土壤基本性质

Table 1 The basic properties of soil

有机质 (g/kg)	全 N (g/kg)	全 P (g/kg)	全 K (g/kg)	有效 N (mg/kg)	有效 P (mg/kg)	有效 K (mg/kg)	pH	物理性粘 粒 (g/kg)	土壤容重 (mg/m ³)
14.300	0.627	0.810	32.180	3.550	6.860	76.800	6.3	317.400	1.475

2.2 数据收集

从 1998-11~2001-04, 每隔 5 天测定土壤含水量。土壤含水量的测定方法为: 用土钻在相应的土层取土, 恒温箱烘干法测定, 重复 3 次。土壤含水量计算公式如式(5)

$$w \% = (g_1 - g_2) / (g_2 - g) \times 100\% \quad (5)$$

式中 w 表示土壤含水量, 单位为%; g 表示铝盒重, 单位为克; g_1 表示铝盒加湿土重, 单位为克; g_2 表示铝盒加干土重, 单位为克。

3 土壤水分预测 AR 模型

本研究使用 SPSS 9.0 for Windows 建立 AR 模型。建模过程包括: 数据整理、建立 AR 模型、模型评价、模型预测。

3.1 数据整理

本文采用土层 0~30 cm 的月平均观测数据进行 AR 建模研究。为了提高数据运算精度, 减少运算次数, 应对原始数据进行标准化处理, 标准化公式见式(6)

$$y_t = \frac{x_t - \mu}{v} \quad (6)$$

式中 x_t 为原始序列, μ 是时序 $\{x_t\}$ 的均值, v 是时序 $\{x_t\}$ 方差。

首先建立 $\{y_t\}$ 的 AR 模型, 根据模型求出预测值, 然后再利用式(6)反变换求出序列 $\{x_t'\}$, 将 $\{x_t'\}$ 与观测值 $\{x_t\}$ 进行比较, 检验模型预测效果。

原始数据见图 1(a), 标准化的数据见图 1(b)。本研究采用前 28 个土壤含水量数据构建 AR 模型, 获取模型参数, 其后的 8 个土壤含水量数据进行模型的预测评价。

从图 1(a) 可见, 观测数据呈近似于平稳的线性变化, 观测数据及标准化数据的统计学特征见表 2。

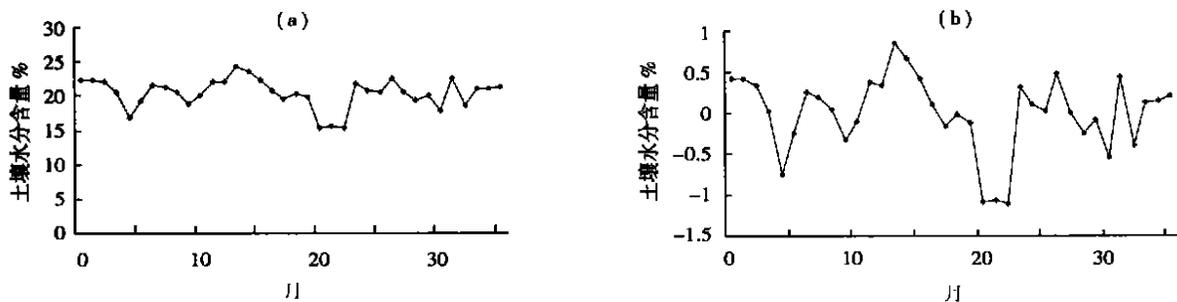


图 1 观测数据(a)及标准化后数据(b)
Fig. 1 The raw (a) and the normalized data (b)

表 2 观测数据及标准化后的数据的统计学特征

Table 2 Statistics of the raw and the normalized data

数据	最大值	最小值	均值	标准差	方差
观测数据	24.4	15.37	20.44	2.14	4.59
标准化数据	0.86	-0.94	-9.67×10^{-6}	0.47	0.22

首先检验标准化的数据是否符合 AR 建模要求。计算其自相关及偏相关系数, 结果见图 2。
从表 2 及图 2 可知, 标准化的数据均值接近于

0, 自相关及偏相关系数随着滞后阶数的递增逐渐趋近于零。因此, 可以认为该数据是零均值、平稳、正态分布的数据序列, 符合 AR 建模的要求。

3.2 建立 AR 模型

用最小二乘法估计 p 取不同值时的 AR 模型参数, 并计算出相应的 AIC 值及标准差, 结果见表 3。

从表 3 中可知, 阶次为 10 的 AR 模型 AIC 值最小, 同时其标准差也最小, 所以考虑使用 AR(10) 进行土壤水分含量的预测建模。模型参数见表 4。

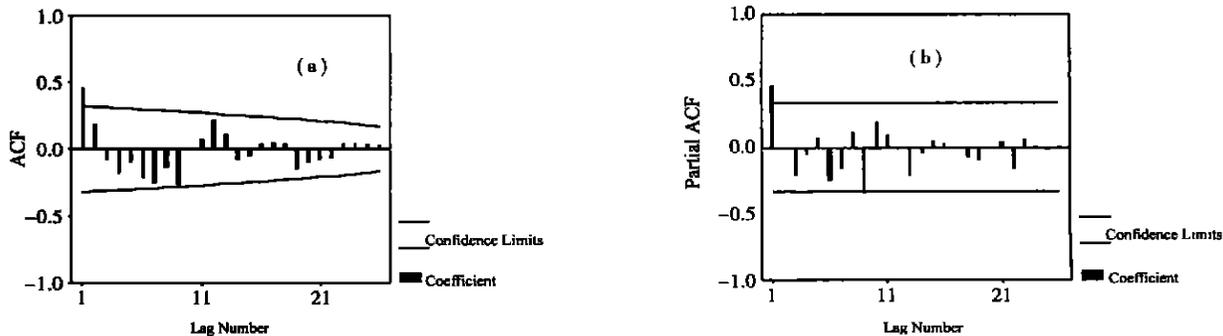


图 2 标准化的数据自相关(a)与偏相关函数图(b)

Fig. 2 The autocorrelation function (a) and partial autocorrelation function (b) of the normalized data

表 3 模型阶次及其对应的 AIC 值及标准差

Table 3 AIC and standard error of the models

阶次	1	2	3	4	5	6	7	8	9	10	11	12
AIC	41.39	43.45	43.79	45.66	47.63	46.25	47.96	49.38	44.6	40.8	43.06	45.42
标准差	0.42	0.42	0.42	0.42	0.43	0.41	0.42	0.42	0.38	0.34	0.35	0.36

表 4 AR(10) 模型参数表

Table 4 The parameters of AR(10) model

常量	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
-0.02	0.82	-0.28	-0.24	0.26	0.12	-0.43	0.13	0.28	-0.78	0.48

3.3 模型评价

对 AR(10) 的模型残差进行统计验证, 残差序列及自相关函数图见图 3。从图中可知, 模型残差围绕零值上下波动, 残差自相关函数图表明残差之间没有呈现出显著自相关。并且第一个残差自相关系数, 符合残差的白噪声检验准则。说明该模型适用。

3.4 模型预测

用前 28 个数据做模型拟合点, 估算出序列的估计值, 后 8 个数据作为样本进行预测检验, 同时预测后期(4 个月) 走势。然后经过相应的反变换得到土壤水分趋势的拟合、预测结果及 95% 的置信区间(图 4), 误差分析结果见表 5。从图 4 可以看出, 模型所得结果与观测值的拟和效果较好。从表 5 可知, 除第 31 观测数据与模型预测数据相差较大(绝对误差为- 5.39, 相对误差为 30.1%), 其余的预测值均较接近于观测值。说明使用 AR 模型可以进行土壤水分动态的预测, 其预测效果良好。

地区紫色土土壤水分运动趋势模型, 其模拟和预测的结果表明, 该方法能够以一定的精度来模拟和预测土壤水分的动态变化过程, 对土壤水分的趋势预测有一定的适用意义, 为采取浇灌排水和蓄水保墒等调节农田土壤水分的技术措施提供了一定的理论依据。

表 5 土壤水分趋势预测结果误差分析

Table 5 The forecasting error analysis of the model

序号	实测值(%)	预测值(%)	绝对误差(%)	相对误差%
29	19.33	19.94	- 0.61	3.14
30	20.08	20.73	- 0.65	3.21
31	17.90	23.29	- 5.39	30.10
32	22.45	18.16	4.29	19.10
33	18.64	18.04	0.60	3.22
34	21.01	19.58	1.43	6.82
35	21.09	20.03	1.06	5.04
36	21.36	21.22	0.14	0.66
37		20.03		
38		20.73		
39		21.43		
40		18.67		

4 结论

本文利用时间序列自回归建模方法建立了丘陵

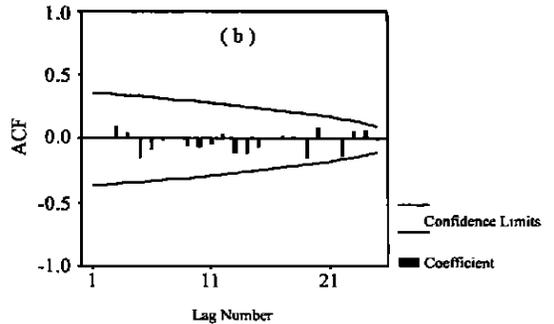
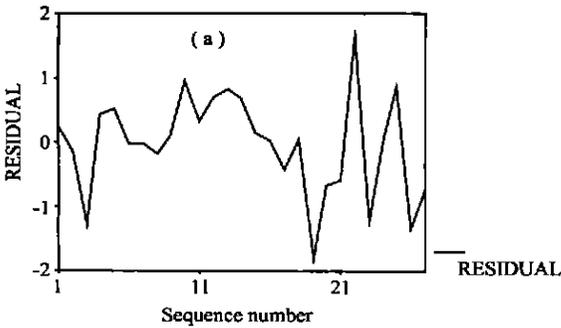


图 3 残差序列(a)及残差自相关图(b)

Fig. 3 The residual series (a) and residual ACF function (b)

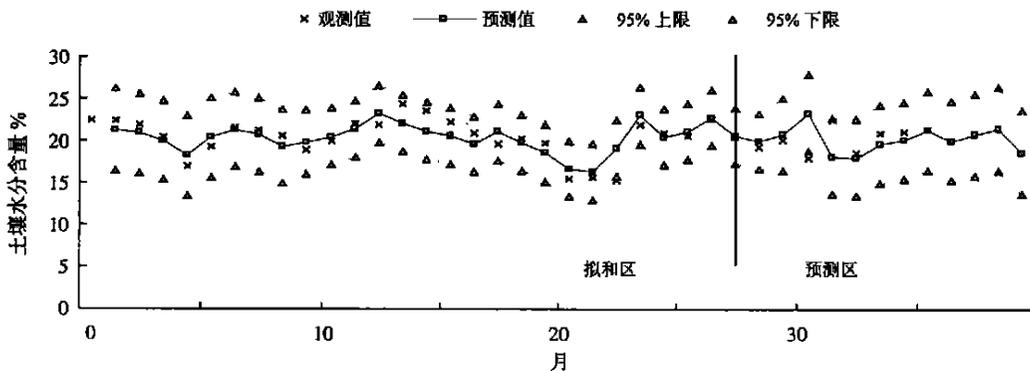


图 4 土壤水分趋势拟和及预测结果

Fig. 4 The fitting and forecasting of the model

在利用 AR 方法建立时间序列预测模型时, 最为重要的是建模数据必须符合平稳、正态分布的条件, 对于非平稳序列, 可以通过适当的变换(如标准化、一次或多次差分等)使之成为满足建模需要的平稳序列。然后才能建立 AR 模型。

本文进行了丘陵地区紫色土土壤水分含量月平均值的趋势预测, 下一步的工作将考虑进行多变量土壤水分运动趋势预测, 加入土层、作物、气候等变量对土壤水分含量进行预测, 探讨多变量之间的相互影响、相互作用关系。

参考文献(References):

[1] Yu Shuzi, Wu Ya. Engineering application of time series analysis [M]. Wuhan: Huazhong University of Science and Technology

Press, 1991. 1~ 395. [杨叔子, 吴雅. 时间序列分析的工程应用 [M]. 武汉: 华中理工大学出版社, 1991. 1~ 395.]

[2] Kang Shaozhong. Stochastic modeling of dynamic process of soil moisture[J]. *Acta Pedologica Sinica*, 1990, 27(1): 17~ 24. [康绍忠. 土壤水分动态的随机模拟研究[J]. 土壤学报, 1990, 27(1): 17~ 24.]

[3] Xiang Jingtian, Shi Jiuen, Zhou Qinfang, *et al.*. Dynamic data process time series analysis [M]. Beijing: China Meteorological Press. 1~ 433. [项静恬, 史久恩, 周琴芳等. 动态数据处理: 时间序列分析[M]. 北京: 气象出版社, 1991. 1~ 433.]

[4] Cheng Baoren, Hong Zaiji, *et al.*. Groudwater dynamic and forecast [M]. Beijing: Science Press. 1~ 382. [陈葆仁, 洪再吉等. 地下水动态及其预测[M]. 北京: 科学出版社, 1988. 1~ 382.]

[5] An Hongzhi. Time series analysis [M]. Beijing: Science Press. 1992. 1~ 344. [安鸿志. 时间序列分析[M]. 北京: 科学出版社, 1992. 1~ 344.]

Soil Water Dynamics Simulation by Autoregression Models

LIU Hongbin¹, WU Wei², WEI Chaofu¹, XIE Deti¹

(1. Resources and Environment College, Southwest Agriculture University, Chongqing 400716, China;

2. Information College, Southwest Agriculture University, Chongqing 400716, China)

Abstract: The environment is friable and the drought is a serious in summer in the Three Gorges Reservoir Area. Changjiang is a serious disaster area of drought, of which frequency is 80%~ 90% and duration are 30~ 50 days. Drought has a serious impact on planting and growth of spring crops. In addition, soil sluce is low, soil layers of upland is weak, ability of fighting a drought is feeble, the contained water of soil planting layer can not satisfy the needs of crops. All of these influence agriculture production seriously and make for low economy benefit of agriculture and graze.

Soil water movement is a complex time series system. Its variety has a close relationship with regional climate conditions and environment and has obvious random undulation. In using AR method of time series analysis to constitute forecast model, the most important is that datum must accord with balance and normal distribution. By proper transformation(for example standardization, one or many times difference etc), AR model can be constituted. The test of AR model uses residual and An Information Criterion.

From November 1998 to April 2001, by measuring soil water of soil surface layer(0~ 30 cm) every six days, we worked out mensal average soil water of purple soil in hilly region. Using 36 months soil water data, adopting time series analysis method(AR model), we studied soil water dynamic trend.

Firstly, we normalized raw data. After the standardization of raw datum, its autocorrelation and partial autocorrelation coefficients gradually trend to zero along with lags. It showed that managed data is a series accorded with zero average, balance, normal distribution and needs of AR model.

Secondly, using former 28 months data as model simulating point, we calculated parameters of AR model and using the subsequent 8 months data as detecting stylebook, we tested model and forecasted the anaphase(4 months)trend of soil water at the same time.

The results showed that the AIC of order 10 AR model is least and its standard deviation is also least. So we considered to use AR(10) to constitute soil water forecast model. By validating residual of AR(10) model statistically, we found that model residual fluctuated surround zero and the first residual autocorrelation coefficient ρ_1 equaled to -0.001, which showed that there was no obvious autocorrelation among residuals. The simulated showed that model could preferably fit the soil water time series.

Key words: autoregression model; soil water; time series; forecast; purple soil