

# ARIMA 模型预测土壤墒情研究<sup>\*</sup>

杨绍辉, 王一鸣, 郭正琴, 董亚峰

(中国农业大学信息与电气工程学院, 北京 100083)

**摘要:** 土壤水分的动态模拟是土壤墒情(旱情)监测及预测预报的重要内容。采用 ARIMA 模型进行土壤水分时间序列的拟合与趋势预测。在建立模型后,用实测数据与模型的预测数据相比较说明 ARIMA 时间序列模型能较好地拟合与预测土壤墒情的变化趋势。

**关键词:** 土壤墒情; 时间序列; ARIMA; 模型; 预测预报

**中图分类号:** S 152.7<sup>+</sup>3 **文献标识码:** A **文章编号:** 1000-7601(2006)02-0114-05

旱灾是我国主要自然灾害之一,与其他自然灾害相比,具有发生频繁、涉及面广、历时长等特点,对农业影响最为严重。

在影响旱情分析的诸多因素中,土壤墒情变化趋势的研究具有重要的意义。通过分析和预测土壤含水率的变化,可以预知作物的受旱情况,科学指导农业灌溉和抗旱救灾,为农业估产和旱灾损失评估提供依据。

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。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<sup>[1]</sup>。在土壤水分动态模拟方面,在北方气候条件下康绍忠<sup>[2]</sup>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。在三峡库区,刘洪斌、王伟等人<sup>[3]</sup>采用时间序列的自回归建模方法对紫色土丘陵旱坡地土壤水分动态进行了模拟和预测。在本文中为了能够预测土壤墒情的变化趋势,采用了时间序列的 ARIMA 模型进行了研究,并用实测数据与模型的预测数据相比较,结果显示 ARIMA 时间序列模型能良好地拟合与预测土壤墒情的变化趋势,有实际的应用价值。

## 1 求和自回归-移动平均模型 (ARIMA)

时间序列模型分为三类:自回归模型 AR( $p$ ), 移动平均模型 MA( $q$ ) 和自回归移动平均模型

ARMA( $p, q$ )。自回归模型和移动平均模型是自回归移动平均模型的特例。一般  $p$  阶自回归、 $q$  阶移动平均模型记作 ARMA( $p, q$ )。而运用 ARMA 模型的前提条件是,用作预测的时间序列已经是由一个均值为零的平稳随机过程产生的。反映在图形上就是所有的样本都围绕某一水平直线上下随机波动。所以对于某些不平稳的序列分析必须经过差分变换后才能进行。经过差分变换后序列应用 ARMA 模型,在习惯上就称之为求和自回归-移动平均模型 ARIMA。一般如果经过了  $d$  阶差分的 ARMA( $p, q$ ) 模型就记作 ARIMA( $p, d, q$ ) 模型<sup>[4]</sup>。如下式所示:

$$\begin{aligned} X_t - \Phi_1 X_{t-1} - \dots - \Phi_p X_{t-p} \\ = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \\ t = 0, \pm 1, \pm 2, \dots \quad (1) \end{aligned}$$

式(1)中,  $p, q$  为模型的阶数;常数系数  $\Phi_1, \Phi_2, \dots, \Phi_p$  和  $\theta_1, \theta_2, \dots, \theta_q$  为模型的参数;  $X_1, X_{t-1}, \dots, X_{t-p}$  为时间序列在不同时刻的取值;  $a_t, a_{t-1}, \dots, a_{t-q}$  为残差序列。

## 2 数据的采集与预分析

### 2.1 数据的采集

水利部 2003 年底立项,分三期建设全国范围的旱情信息管理系统,一些省份也在陆续开展墒情信息系统的建设。北京市现已经建成 38 个墒情监测站和 120 处墒情巡测点<sup>[5]</sup>。本文测墒的试验地点位于北京市大兴区芦城的北京水利水电中心基地,使用的是 SWR 2 型土壤水分传感器,传感器的埋设位置是地表下 10、20、30、40、50、60、80、100 cm 八个深

\* 收稿日期: 2005-08-02

基金项目: 国家高技术研究发展计划(863 计划) 专项课题“作物水分信息采集与精量控制用水技术与开发”(2002AA2Z4071)

作者简介: 杨绍辉(1979-), 男, 湖北恩施人, 博士生, 从事土壤墒情监测和预测预报的研究开发工作。E-mail: yangshaohui1979@163

度。试验由 2003 年 11 月 16 日开始至 2005 年 2 月 24 日,采集与发送装置每隔 4h 采集一次土壤水分体积含水量并通过 GSM 通信网络远距离实时发送数据到试验室上位机。通过现代的传感器测量技术和计算机、通信技术相结合,使墒情数据的采集工作变得非常可靠、准确、方便。

## 2.2 数据的预分析

数据采集到上位机后,首先对原始数据进行数理统计分析。本文用 2003 年 11 月 16 日至 2005 年 2 月 24 日共获得的 450 组数据进行了初步的统计分析。统计发现土壤水分变异系数的变化有很明显的规律。10cm、20cm 深处土壤水分的变异系数接近 30%,属于比较剧烈的变异,30cm、40cm、50cm、60cm、80cm 深处的土壤变异系数在 10%~20%之间,属于中等程度的变异,而在 100cm 处,变异系数只有 4.9%,属于弱变异。这说明越浅层的土壤水分值变化越快,不稳定,而深层土壤水分值却变化缓慢,比较稳定。这样在做时间序列的分析时将考虑用不同的时间序列模型来拟合不同埋设深度的土壤水分变化。

## 3 时间序列模型分析

根据不同埋设深度土壤水分值的相关性分析,一般 10cm、40cm、60cm、100cm 处的土壤水分值具有代表性的意义<sup>[9]</sup>。在本文中将以 40cm 处的土壤水分数据为例进行分析,建立时间序列模型。模型的分析分为数据的整理、时间序列模型的选定、模型的评价以及模型的预测等四步进行。

### 3.1 数据的整理

为了满足后面分析的需要,同时提高数据运算精度,首先应对原始数据进行标准化处理,标准化公式为:

$$Y_t = \frac{X_t - u}{v} \quad (3)$$

式(3)中, $Y_t$  为标准化后的时间序列, $X_t$  为原始时间序列, $u$  是时间序列  $\{X_t\}$  的均值, $v$  是时间序列  $\{X_t\}$  的标准差,标准化后的数据见图 1 图中横坐标为测量的天数,纵坐标为标准化后 40cm 埋深处的土壤水分值在一天中的平均含水量。

时间序列模型的分析都是建立在序列平稳的条件上的。一个平稳的随机过程有以下要求:均值不随时间变化;自相关系数只与时间间隔有关,而与所处的时间无关。实际上,大多数时间序列都是不平稳的。由图 1 可知,标准化后 40cm 深处土壤水分值的

序列显然也不满足平稳的要求,需要对序列进行差分处理,差分处理的公式如下:

$$DY_t = Y_t - Y_{t-1} \quad (3)$$

式(3)中, $DY_t$  表示一阶差分。 $Y_{t-1}$  差分后的序列通常表现出没有明显的时间趋势以及剧烈波动等特征。

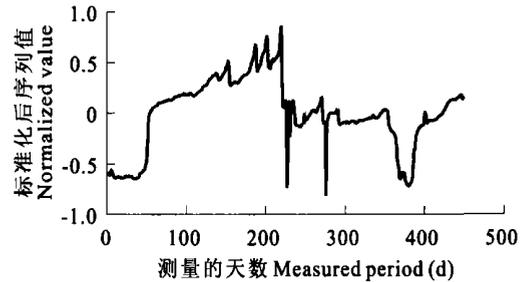


图 1 标准化后 40cm 埋深处土壤水分值

Fig. 1 The normalized soil moisture at 40cm depth

图 2 所示即为一阶差分处理后的结果。

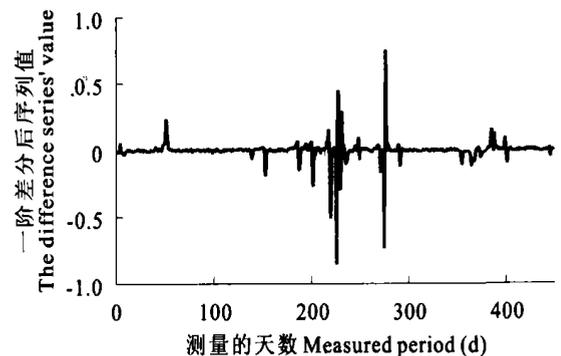


图 2 差分处理后的序列

Fig. 2 The difference series of soil moisture

由图 2 可知,一阶差分后序列的均值十分接近于零,序列没有明显的时间变化趋势。可以认为一阶差分后的土壤水分序列为均值是零,平稳的数据序列,符合时间序列的建模要求。

### 3.2 模型的选定

3.2.1 模型类别的确定 时间序列模型一般可以分为自回归模型  $AR(p)$ , 滑动平均模型  $MA(q)$  和自回归滑动平均模型  $ARMA(p, q)$  三个大类,而 ARIMA 模型属于 ARMA 模型。序列需用哪种模型进行拟合,一般是通过自回归系数图(ACF)和偏自回归系数图(PACF)进行判断。表 1 为判断的规则,表中 Lag 表示滞后数。

为了判断 40cm 深度土壤水分值序列的模型类

别,现作出该序列的 ACF 图和 PACF 图。

表 1 时间序列模型的判断规则

Table 1 The rule to judge time series models

模型 Model	自回归系数 Autoregressive coefficient (ACF)	偏自回归系数 Partial autore gressive coefficient (PACF)
AR(p)	拖尾 Tailing	Lag = p 后骤减 Sudden decrease after Lag = p
MA(q)	Lag = q 后骤减 Sudden decrease after Lag = q	拖尾 Tailing
ARMA(p, q) 和 ARIMA(p, d, q)	拖尾 Tailing	拖尾 Tailing

由图 3和图 4看出,当滞后数增大时,自回归系数与偏自回归系数均拖尾。同时,在整理序列时曾经一阶差分处理过,所以应该选取 ARIMA 模型进行建模分析。

3.2.2 模型阶数的确定 本文使用残差的白噪声

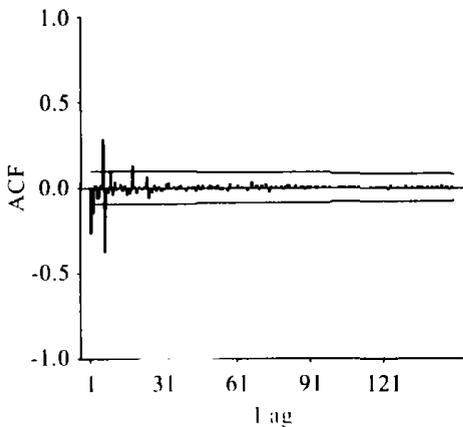


图 3 ACF 图  
Fig 3 Chart of ACF

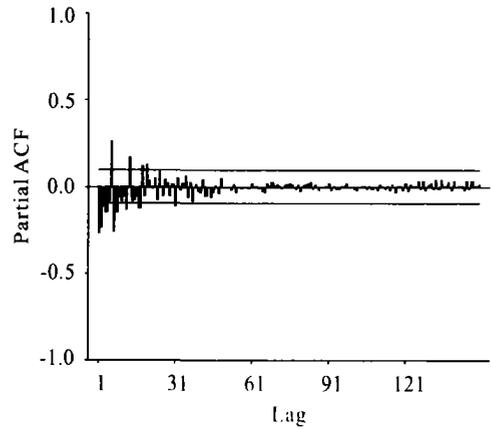


图 4 PACF 图  
Fig 4 Chart of PACF

表 2 不同阶次的模型对应的 AIC 值

Table 2 The AIC value of different models

阶次 p Order	2	4	5	6	7	8	9	10	28
AIC	935.0	932.7	934.7	930.6	929.8	924.2	926.0	928.6	938.6

表 3 ARIMA(8,1,7) 模型参数表

Table 3 The parameters of ARIMA(8,1,7) model

$\Phi_1$	$\Phi_2$	$\Phi_3$	$\Phi_4$	$\Phi_5$	$\Phi_6$	$\Phi_7$	$\Phi_8$
0.2372	-0.7859	0.4427	-0.1334	0.2049	0.2826	0.3881	0.1244
$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\theta_7$	$\theta_8$
0.4154	-0.6984	0.5319	-0.2376	0.2425	0.0782	0.5303	

检验准则及 AIC 准则(An Information Criterion) 进行 ARIMA 模型的检验。AIC 准则又称为信息准则,其计算公式为

$$AIC(p) = N \ln \alpha^2 + 2p \quad (4)$$

式中, p 为模型阶次, N 是时间序列  $\{x_t\}$  的长度。

$$\alpha^2 = \frac{1}{N} \sum_{t=1}^N \alpha_t^2 \quad (5)$$

式中,  $\alpha_t$  为残差序列。由于事先无法判断模型的阶次,因此在建模过程中先根据 PACF 图给定模型的阶次,然后按照最小二乘法估计出模型的参数,得到各阶模型,最后根据 AIC 值的大小,取定模型的最佳阶次。AIC 值越小的,模型越好<sup>[9]</sup>。由图 4 可知,当 Lag = 2, 6, 8, 28 时,偏自回归系数超出了可信区间。说明序列在这些阶次相关性是比较大的。应该以这些阶次作为主要参考对象。

由表 2 可知,当 p = 8 时 AIC 达到最小值,所以模型应取为 8 阶。根据 Pandit 建模策略<sup>[9]</sup>,可得 q 取为 7。因此,模型可确定为 ARIMA(8,1,7)。

利用 SPSS 11.5 软件分析,可得模型的各个常数参数。如表 3 所示。

3.3 模型的评价

对该模型残差进行统计验证,残差序列及自相关函数图见图 5 和图 6。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。并且由图 6 可知,残差自相关系数符合残差的白噪声检验准则。说明该模型适用于拟合与预测土壤水分值的变化趋势。

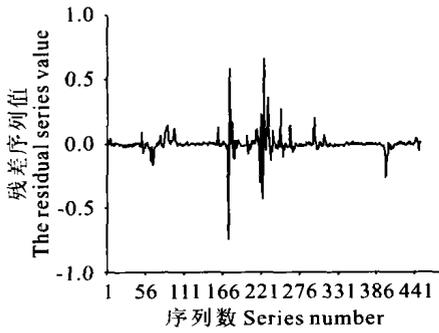


图 5 残差序列  
Fig 5 The residual series

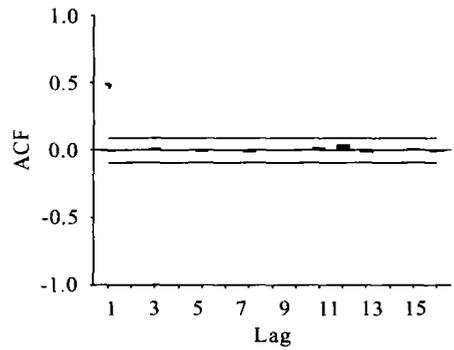


图 6 残差 ACF 图  
Fig 6 The residual ACF function

### 3.4 模型预测

本文用 2003 年 11 月 16 日至 2005 年 2 月 24 日所测得的 450d 的数据作为模型拟合点,估算出序列的估计值。用 2005 年 2 月 24 日到 3 月 10 日所测的 15d 数据作为预测的样本,检验预测的效果。经过对式(1)的反变换后可得计算土壤水分的公式。

$$X_t = Y_t \times v + u \quad (6)$$

通过公式(6)可以计算出每天的土壤水分值,最

后可生成图 7,为了便于观察,图中只绘出了 30d 之内的拟合值和 15d 之内的预测值。由图 7 可知,模型拟合土壤墒情的变化趋势较好,而且预测值也十分接近于观测值,最大绝对误差为 0.8%,最大相对误差为 7.2%。这说明使用 ARIMA(8,1,7) 时间序列模型可以较好地模拟与预测 40cm 埋深处土壤墒情的变化趋势。

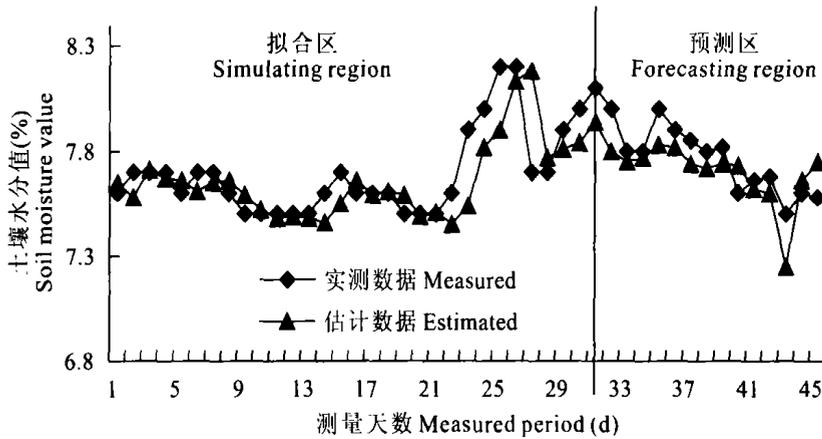


图 7 土壤水分拟合及预测结果  
Fig 7 The simulating and forecasting of soil moisture

### 3.5 其它土壤埋深的时间序列预测模型

根据前面的数据预测分析,不同的土壤埋深,土壤水分序列具有不同的变异系数。因此也应该用不

同的时间序列模型来拟合。用上述的方法,表 4 列出了在大兴基地不同埋深的土壤水分时间序列模型。

表 4 不同埋深的时间序列模型  
Table 4 The time series model at different depth

测量深度 Depth (cm)	10	20	30	40	50	60	80	100
模型 Model	ARIMA (6,1,3)	ARIMA (8,1,4)	ARIMA (8,1,3)	ARIMA (8,1,7)	ARIMA (8,1,7)	ARIMA (5,1,4)	ARIMA (4,1,4)	ARIMA (1,1,4)

由表 4 所示,不同深度的土壤水分序列具有不同的时间序列模型,这些模型模拟所得结果误差较小,而且预测值也十分接近于观测值。这也说明了 ARIMA 模型适用于土壤墒情的模拟与预测。同时,虽然不同深度的土壤水分序列具有不同的时间序列模型,但也有共同点,即所有模型中  $d = 1$ ,说明不同埋深的土壤水分序列都只需要进行一阶差分,就可使序列平稳。在埋深为 100 cm 时,只需要一阶模型 ( $p = 1$ ) 就可以拟合 100 cm 深度土壤水分值的变化,而其他埋深则需要较高阶的模型。这是由于在 100 cm 深度,土壤水分值变异不明显的结果。

## 4 结 论

本文的研究对象是某一固定土壤水分采样点在不同埋设深度的各个土壤水分采样序列,以 40 cm 深度的土壤水分序列为例给出了分析土壤水分时间序列的具体方法和步骤。建立了时间序列模型,对该模型进行了评价并用来进行预测,预测效果良好。最后根据对不同深度的土壤水分序列分析得出,不同

深度的土壤水分序列均需要对原序列进行一阶差分后才能成为稳定的时间序列。而且各个埋深的土壤水分序列都适合运用 ARIMA 模型进行时间序列的建模分析。

## 参 考 文 献:

- [1] 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] 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] 刘洪斌,武 伟,魏朝富,等. AR 模型在土壤水分动态模拟中的应用[J].山地学报,2004,22(1):121~125
- [4] 田 铮.时间序列的理论与方法[M].北京:高等教育出版社,2001.214~250
- [5] 孙 凯.墒 情(旱情)监测与预测预报方法研究[D].北京:中国农业大学,2004
- [6] 吴宪民.时间序列及系统分析与应用[M].北京:机械工业出版社,1988.13~66
- [7] 孙 凯,王一鸣,杨绍辉,等.区域内旱情监测站点的布设研究[J].水利学报,2005,36(3):208~212

## Research on soil moisture forecast based on ARIMA model

YANG Shao hui, WANG Yi ming, GUO Zheng qin, DONG Ya feng

(College of Information and Electrical Engineer, China Agricultural University, Beijing 10083, China)

**Abstract:** Soil moisture dynamics simulation is an important item of soil moisture monitoring and forecasting. In this paper, ARIMA model is used to forecast the change trend of soil moisture time series. By comparing the measured value with the estimated value, it is showed that the ARIMA model can preferably fit soil moisture time series.

**Key words:** soil moisture content; time series; ARIMA model; forecast