

文章编号:1672-3961(2008)02-0096-05

地下水位预报中的组合时间序列分析法

廖伙木¹, 董增川¹, 束龙仓², 汝安³

(1. 河海大学水文水资源及水利工程科学国家重点实验室, 江苏 南京 210098;

2. 河海大学水资源环境学院, 江苏 南京 210098;

3. 山东大学 土建与水利学院, 山东 济南 250061)

摘要:将系统分析方法中的传统GM(1,1)模型与时间序列分析方法相结合建立地下水位的预报模型. 为更好地反映地下水位的统计规律随时间而缓慢变化的现象,采用衰减记忆最小二乘法率定GM(1,1)模型参数. 用方差分析法进行地下水位时间序列的周期分析时,对初相进行优选;在选定的置信度水平下,当有多个周期都通过检验,此时该如何优先选择哪个周期,文中提出按拟选周期的F分布检验统计量与检验区间界限值的比值最大原则来选择周期成分. 最后,采用本文所述方法进行福建省龙岩市的年最高地下水位的预报.

关键词:地下水位预报;GM(1,1)模型;方差分析;初相优选;时间序列分析

中图分类号:TV213 文献标志码:A

Combinative time series analysis method for the prediction of the groundwater level

LIAO Huo-mu¹, DONG Zeng-chuan¹, SHU Long-cang², YUN Ru-an³

(1. State Key Laboratory of Hydrology on Water Resource and Hydraulic Engineering, Hohai University, Nanjing 210098, China;

2. The College of Water Resources and Environment, Hohai University, Nanjing 210098, China;

3. School of Civil Engineering, Shandong University, Jinan 250061, China)

Abstract: Traditional GM(1,1) and time series analysis method were integrated together to establish the groundwater level forecast model. The fading memory least square methods was applied in the process of confirming the parameters of GM(1,1) in order to more factually reflect the phenomena of the groundwater level's statistic rule varying with the time. The periods components were analyzed with the variance analysis method and the starting phases were optimized. Under the condition that more than one F-statistic values of periods were over the critical values of the selected confidence degree, what rules can be adopted to decide which period should be first selected. The rule was put forward that the ratio of F-statistic value of selected period's being divided by the critical value of the examination interval should be the greatest. Finally, the yearly highest groundwater levels in Longyan Basin in the west of Fujian Province were forecast with this method.

Key words: groundwater level forecast; GM(1,1); variance analysis; starting phase optimization; time series analysis

0 引言

地下水资源系统比较复杂,地下水位的变化受

到系统补给、径流、排泄作用的影响. 地下水位预报的方法有很多,如解析法、均衡法、数值模拟法,人工神经网络分析、小波、系统分析方法等^[1-8]. 这些方法中,解析法适用于理想条件下的含水层,均衡法需

收稿日期:2007-06-25

作者简介:廖伙木(1968-),男,福建宁化人,博士研究生,从事水文学及水资源研究.

E-mail: fjiaohm@163.com

要获知系统的补给量、排泄量,数值模拟法则不需获知补给量、排泄量,还必须进行水文地质参数的

分区,知晓各分区的水文地质参数等,一般来说,由于无法准确获得水文地质参数与将来的补给量、排泄量等,使得上述方法所得结果仍具有某种程度的不确定性.与前述水位预报方法相比,系统分析方法只是通过对一定长度的观测水位序列的分析来获取地下水位的变化规律,从而实现对地下水位的预报,所需资料相对较少,适合资料缺乏地区应用. GM(1,1)模型与时间序列分析法都属系统分析方法,本文拟将 GM(1,1)模型与时间序列分析法耦合,构建地下水位的预报模型.

一般而言,对于时间系列,其预报模型如后.

$$H_t = F_t + P_t + X_t, \quad (1)$$

式中: F_t —趋势项, P_t —周期项, X_t —平稳随机成分.

同样,地下水时间序列分析模型通常也包含上述3个部分.

1 趋势项

设有原始地下水观测序列 $H_t (t=1, 2, \dots, n)$, 在以往的时间序列分析中,大多用多项式来拟合序列的趋势成分,但率定多项式的形式和项数的工作较为繁琐. GM(1,1)模型通过对微分方程的近似所得的灰微分方程 $dH/dt + aH = b$ 的求解,能够较好地拟合系列的变化趋势,于上世纪80年代就被单独作为地下水的预报模型.但在率定模型参数 a, b 时,对序列中的前后期数据是等权重的,不能反映统计规律随时间的变迁而缓慢变化的情况.为此,本文在率定参数的过程中,引入衰减因子以不断消弱序列前期观测数据对当前估计结果的影响,使模型参数能跟踪所研究过程特性的变化.具体步骤为:

1.1 均值序列的生成

对原始观测水位序列进行一次累加生成(AGO),得到累加序列 $H_t^1 (t=2, 3, \dots, n)$,再对 H_t^1 进行一次均值化处理,得到均值序列 $Z_t^1 (t=2, 3, \dots, n)$. 其中,

$$H_t^1 = \sum_{i=1}^t H_i, \quad Z_t^1 = 0.5(H_{t-1}^1 + H_t^1). \quad (2)$$

1.2 参数 a, b 的率定

对微分方程 $dH/dt + aH = b$ 进行离散近似可得到灰微分方程 $H_t + aZ_t^1 = b$,引入衰减因子 α ,采用衰减记忆最小二乘法率定模型参数 a, b .

令 $\Phi = [a, b]^T, H = [H_2, H_3, \dots, H_n]^T$,

$$Z = \begin{bmatrix} -Z_2^1 & 1 \\ -Z_3^1 & 1 \\ \vdots & \vdots \\ -Z_n^1 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} \alpha^{n-2} & 0 & \cdots & 0 \\ 0 & \alpha^{n-3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \alpha^0 \end{bmatrix},$$

$$\tilde{Z} = AZ, \quad \tilde{H} = AH, \quad \text{则 } \Phi = [\tilde{Z}^T \tilde{Z}]^{-1} \tilde{Z}^T \tilde{H}. \quad (3)$$

1.3 趋势项的提取

通过对灰微分方程的求解可得到原始观测水位序列趋势项为

$$\left. \begin{aligned} H_t^1 &= (H_1 - b/a)\exp(a(t-1)) + b/a \\ F_t &= H_t^1 - H_{t-1}^1 (t=2, 3, \dots, n, n+1, \dots) \\ F_1 &= F_2 \exp(a) \end{aligned} \right\}. \quad (4)$$

2 周期项

将原始观测水位序列扣除趋势项后便可得到新序列 $H_t^p (t=1, 2, \dots, n)$,对新序列是否存在周期成分及周期成分组成的分析一般采用频谱分析法、方差分析法等.文献【9】在采用方差分析法提取周期成分时,排序过程中固定地将拟提取周期序列的第一项 H_1^p 排在第一组的第一项,即 $w_{1,1}^{T,m} = H_1^p$.这实际上相当于对于潜在周期项的初相没有进行优选而加以固定,在某些情况下有可能使提取的周期成分失真或导致原本有周期的序列提不到周期项;还有文献【10】中对于在一次分析中当有多个周期都通过检验时,该如何优先选择哪个周期,没有加以说明.为更好地进行周期成分的分析,本文对拟提取周期序列 $H_t^p (t=1, 2, \dots, n)$ 的分组排序加以改进,并提出当有多个周期成分同时通过检验时,按拟选择周期的 F 分布检验统计量与检验区间界限值的比值最大的原则来确定被优先选取的周期成分,具体步骤如下:

2.1 数据序列的分组

根据序列长度 n ,确定可能的周期为 $T=2, 3, \dots, \text{int}(n/2)$ ($\text{int}(x)$ 表示不大于 x 的最大整数).分别令 $T=2, 3, \dots, \text{int}(n/2)$,固定 T 值,再分别令 $m=1, 2, \dots, T$ (m 表示将序列中的第1个数据 H_1^p 排在第 m 位,本文简称为初相),对固定的 T, m 值,将序列按以下公式分成 T 组.

$$W_{i,j}^{T,m} = H_{(i-1)T+j-m+1}^p (RB_j^{T,m} \leq i \leq RE_j^{T,m}, 1 \leq j \leq T) \quad (5)$$

$$RB_j^{T,m} = \begin{cases} 2 & 1 \leq j < m \text{ 且 } m > 1 \\ 1 & m \leq j \leq T \end{cases}.$$

$$RE_j^{T,m} = \begin{cases} \text{int}((n+m-1)/T) + 1, & 1 \leq j \leq \text{mod}(m+n-1, T) \text{ 且 } \text{mod}(m+n-1, T) \neq 0; \\ \text{int}((n+m-1)/T), & \text{mod}(m+n-1, T) < j \leq T. \end{cases}$$

式中: $\text{mod}(x, y)$ —— x 被 y 整除后的余数, $W_{i,j}^{T,m}$ ——周期为 T 、初相为 m 时分组后第 j 组第 i 行的元素.

2.2 计算组间离差平方和 S_1 , 组内离差平方和 S_2

$$S_1^{T,m} = \sum_{j=1}^T \frac{(G_j^{T,m})^2}{RE_j^{T,m} - RB_j^{T,m}} - \frac{(G^{T,m})^2}{n}, \quad (6)$$

$$S_2^{T,m} = \sum_{j=1}^T \sum_{i=1}^{RE_j^{T,m}} (W_{i,j}^{T,m})^2 - \sum_{j=1}^T \frac{(G_j^{T,m})^2}{RE_j^{T,m} - RB_j^{T,m}}. \quad (7)$$

式中, $G_j^{T,m} = \sum_{i=RB_j^{T,m}}^{RE_j^{T,m}} W_{i,j}^{T,m}$, $G^{T,m} = \sum_{j=1}^T \sum_{i=RB_j^{T,m}}^{RE_j^{T,m}} W_{i,j}^{T,m}$.

2.3 计算检验统计量 F , 并进行检验

首先计算固定 T, m 值时序列的检验统计量 $F^{T,m}$.

$$F^{T,m} = (s_1^{T,m}/f_1)/(s_2^{T,m}/f_2). \quad (8)$$

式中, $f_1 = T - 1$ ——组间离差平方和 $S_1^{T,m}$ 的自由度, $f_2 = n - T$ ——组内离差平方和 $S_2^{T,m}$ 的自由度.

再求取固定 T 值的最大的检验统计量 F^{T,m_i} .

$$F^{T,m_i} = \max_{0 \leq m \leq T-1} F^{T,m}. \quad (9)$$

选定置信度 β , 根据自由度 f_1, f_2 查出 F_β^T . 如果 $F^{T,m_i} > F_\beta^T$, 说明在选定置信度 β 下周期 T 可能存在, 计算 $FB^{T,m_i} = F^{T,m_i}/F_\beta^T$ 的值; 如果 $F^{T,m_i} \leq F_\beta^T$, 表明在选定置信度 β 下周期 T 不存在. 转入下一个可能周期的分析.

所有的 T 值都分析后, 如果通过检验的可能周期不止一个而有 l 个, 周期分别为 $T = T_1, T_2, \dots, T_l$, 则以序列中最大的 FB 所对应的 T_i 作为序列的周期.

$$FB^{T_i} = \max_{T=T_1 \dots T_l} FB^{T,m_i}. \quad (10)$$

2.4 确定周期成分的各位相的振幅

如经以上步骤确定序列的周期为 T_i , 所对应的初相为 m_i , 则需求取该周期各位相共 T_i 个振幅, 各振幅采用序列按周期 T_i 、初相 m_i 分成 T_i 组后各组的平均值.

$$ZF_k^{T_i, m_i} = \frac{G_k^{T_i, m_i}}{RE_k^{T_i, m_i} - RB_k^{T_i, m_i}} \quad (k = 1, 2, \dots, T_i). \quad (11)$$

式中, $ZF_k^{T_i, m_i}$ ——第 k 个位相的振幅.

2.5 其余周期成分的分析

将序列中的各观测值减去所提取周期的各相应振幅, 得到新序列, 返回 2.2 步对新序列开展新的周期分析.

在 2.3 步中通过检验周期为 T_i , 初相为 m_i 的

情况下, 将序列按以下公式减去周期组份, 得到新序列 H_t^p .

$$H_t^p = H_t^p - ZF_{\text{mod}(T_i+1+t-m_i, T_i)}^{T_i, m_i}. \quad (12)$$

原则上, 所提取周期成分一般不超过 6 个. 如原始观测序列提取的周期共有 e 个, 其周期、初相分别为 $T = T_1, T_2, \dots, T_e, m = m_1, m_2, \dots, m_e$ 则原始观测序列的周期成分如下.

$$p_t = \sum_{i=1}^e ZF_{\text{mod}(T_i+1+t-m_i, T_i)}. \quad (13)$$

3 随机项

原始观测水位序列减去趋势项与周期项便得到随机平稳序列 $X_t (t = 1, 2, \dots, n)$. 对于随机序列大多采用 ARMA(p, q) 进行拟合, 而其中较为常用的为 ARMA(1,1) 模型.

$$X_t = u + \varphi(X_{t-1} - u) + \varepsilon_t - \theta\varepsilon_{t-1}. \quad (14)$$

式中: $u = \sum_{i=1}^n X_i / n$ ——序列 X_t 的均值, $\varphi = \rho_1/\rho_2$ ——模型的参数.

$$\rho_i = \frac{c_i}{c_0}, \quad c_i = \frac{\sum_{t=1}^{n-i} (X_t - u)(X_{t+i} - u)}{n - i}, \quad i = 1, 2.$$

$\theta =$

$$\frac{2\varphi c_1 - \varphi^2 c_0 - c_0 \pm \sqrt{(c_0 + \varphi^2 c_0 - 2\varphi c_1)^2 - 4(c_1 - \varphi c_0)^2}}{2(c_1 - \varphi c_0)}$$

(“+ , -”符号的选择应保证 $|\theta| < 1$).

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ ——服从均值为 0, 方差为 σ_ε^2 ($\sigma_\varepsilon^2 = (\varphi c_0 - c_1)/\theta$) 的正态分布随机变量.

4 实例分析

龙岩盆地地处福建西部, 是龙岩市城区所在地. 盆地内岩溶地下水发育, 是当地主要的供水水源, 2006 年岩溶地下水年开采量为 $6.070 \times 10^7 \text{ m}^3$, 是福建省地下水开采量最大的城市. 为掌握地下水位的动态变化规律, 自 1984 年始, 在盆地内布设有 4 个国家级地下水位监测孔 (孔号分别为 3508020029, 3508020077, 3508020094, 3508020102) 对盆地内的地下水位开展监测. 为更好地为有关部门的水资源规划提供依据, 每年都对下一年的地下水最高水位进行预报. 目前已积累有 23 年的水位观测资料, 本文利用 18 年的水位资料建立模型, 其余 5 年数据用于预报检验. 为节省篇幅, 本文以 3508020029 孔的资

料为例,介绍建模与预报过程.

4.1 模型的建立

对监测孔的1984年至2001年共18年的原始观测年最高水位序列,引入衰减因子,的取值一般在0.9~1之间,本文取0.98.利用公式(2)~(3)求得趋势成分的参数 $a = 0.0006211$, $b = 329.63$.

有了GM(1,1)模型参数 a, b 后,由公式(4)计算序列的趋势成分.将原始观测序列减去趋势成分得到新序列,利用公式(5)~(12)对新序列进行周期分析,本文对序列提取3个周期.各周期长、初相、统计量自由度、检验区间见表1.

表1 3508020029孔年最高地下水水位周期成分表
Table 1 The period components table of the yearly highest groundwater levels in 3508020029 bore

序号	周期长	初相	检验统计量 F 值	自由度 (f_1, f_2)	检验区间
1	4	1	3.52	(3, 14)	(2.52, +∞)
2	9	2	2.15	(8, 9)	(2.47, +∞)
3	5	1	5.45	(4, 13)	(2.43, +∞)

注:表中选定的置信度水平为0.10

从表中可以看出,周期长为4年、5年的检验统计量值较大,在对应的自由度下,能通过 $\alpha = 0.10$ 水平的检验,周期长为9年的检验统计量值相对于相

表2 3508020029孔年最高水位实测值与预报值对比表

Table 2 The forecast and observation values comparison of the yearly highest groundwater levels in 3508020029 bore

年份	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
实测值	330.35	329.58	329.64	327.69	329.21	328.23	328.19	325.93	330.86	328.71	328.23	327.23
预报值	330.68	329.83	329.37	327.60	329.41	328.33	328.42	325.53	330.59	328.71	328.03	326.97

年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
实测值	326.26	326.89	326.61	325.36	326.55	326.67	326.32	326.1	323.78	327.06	328.95
预报值	326.53	327.44	326.45	325.18	326.40	327.13	326.97	326.31	324.02	326.43	327.77

4.3 年最高水位的预报

所建模型经检验后便可用于对将来地下水位进行预报,本文选定地下水位的预报期为一年.由预报模型所得的2007年3508020029孔的年最高水位为327.21 m.

5 结论

本文引入衰减因子采用衰减记忆最小二乘法率定GM(1,1)模型的参数,使所得GM(1,1)模型能更好地拟合序列趋势成分的近期变化规律.由公式(9)可知,初相优选可以得到拟选周期 T 所对应的检验统计量的最大值,可避免周期的漏选,更有效地提取序列的周期成分.公式(10)则表明,根据拟选周期所对应的 F 检验统计量值与检验区间的界限

应的自由度接近 $\alpha = 0.01$ 水平所要求的检验值.

得到各周期的周期长与初相后,由公式(11), (13)计算序列的周期成分.将原始观测序列减去趋势成分与周期成分便得到随机序列.由公式(14)得到拟合的ARMA(1,1)模型如下.

$$x_t = 0.037 + 0.34x_{t-1} + \epsilon_t + 0.11\epsilon_{t-1}.$$

式中, $\epsilon_t \sim N(0, 0.06)$ 即 ϵ_t 服从均值为0,方差为0.06的正态分布.

4.2 模型的检验

以上所建立模型是否有效与可靠还得经过验证.有效性是通过用于建立模型序列数据的实测值与模型预报值的拟合程度对比所得的确定性系数来评判的,可靠性是通过未用于建立模型序列数据的实测值与模型预报值的对比所得的合格点数的百分率来判断.

计算表明模型拟合的确定性系数达到98%.试预报的绝对误差除2006年为1.18 m外,其余均小于0.65 m,相对误差均小于1%.若选定合格点的相对误差为1%、绝对误差为1.0 m,则合格率达80%.由《水文情报预报规范》可知,所建模型的有效性与可靠性均为甲等,可用于实际应用.

值之间的比值(如图1)最大原则确定的周期,可以使更显著的周期被优先提取,从而在一定程度上提高了所选周期成分的有效性.将传统的GM(1,1)预报模型耦合到时间序列分析模型中所建立的组合时间序列分析预报模型,具有比传统GM(1,1)模型精度高,比通常时间序列分析模型更为简练的特点,具有一定的适用性.

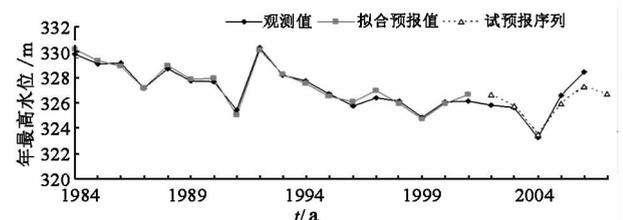


图1 3508020029孔年最高地下水水位实测值与预报值对比图
Fig. 1 The comparison of the forecast and observation values of the yearly highest groundwater levels in 3508020029 bore

参考文献:

- [1] 兰盈盈,赵康文,肖长来. 吉林省洮儿河扇形地地下水库数值模拟[J]. 人民长江,2006,37(8):15-17.
LAN Ying-Ying, ZHAO Kang-wen, XIAO Chang-lai. The numerical simulation analysis of the groundwater reservoir in Taore river's sector region of Jilin province [J]. Yangtze River, 2006, 37(8):15-17.
- [2] 陈南祥,曹连海,徐建新. 地下水位预报的相空间重构神经网络模型研究[J]. 西北农林科技大学学报:自然科学版,2005,33(12):139-142.
CHEN Nan-xiang, CAO Lian-hai, XU Jian-xin. The model of phase space reconstruction and neural network in the groundwater level[J]. Journal of Northwest A and F University: Natural Science Edition, 2005, 33(12):139-142.
- [3] 常亮,解建仓. 应用优化神经网络算法预报地下水位[J]. 水利水运工程学报,2005,5(1):66-70.
CHANG Liang, XIE Jian-cang. Prediction of groundwater level by optimized neural network algorithm[J]. Hydro-Science and Engineering, 2005, 5(1):66-70.
- [4] 王文圣,廖杰,丁晶. 浅层地下水位预测的小波网络模型[J]. 土木工程学报,2004,37(12):62-66.
WANG Wen-sheng, LIAO Jie, DING Jing. Wavelet network model for shallow water table prediction[J]. China Civil Engineering Journal, 2004, 37(12):62-66.
- [5] 李平,卢文喜,杨忠平. 频谱分析法在吉林西部地下水动态预报中的应用[J]. 水文地质工程地质,2005,6(2):70-73.
LI Ping, LU Wen-xi, YANG Zhong-ping. Application of spectrum analysis method to the prediction of groundwater regime in west Jilin province[J]. Hydrogeology and Engineering Geology, 2005, 6(2):70-73.
- [6] 杨金忠,蔡树英. 地下水动态预报的多层递阶组合模型[J]. 水科学进展,1995,6(2):101-106.
YANG Jin-zhong, CAI Shu-ying. A multi-layer hierarchical combinative model for prediction of regional groundwater level [J]. Advance in Water Science, 1995, 6(2):101-106.
- [7] 杨忠平,卢文喜,李平. 时间序列模型在吉林西部地下水动态变化预测中的应用[J]. 水利学报,2005,36(12):1475-1479.
YANG Zhong-ping, LU Wen-xi, LI Ping. Application of time-series model to predict groundwater regime[J]. Shuili Xuebao, 2005, 36(12):1475-1479.
- [8] 何宗花. 一种区域地下水位预报的时间序列分析组合模型[J]. 广东水利电力职业技术学院学报,2003,1(4):33-35.
HE Zong-hua. A combinative model of time series analysis for prediction of regional groundwater depth[J]. Journal of Guangdong Technical College of Water Resources and Electric Engineering, 2003, 1(4):33-35.
- [9] 范钟秀. 中长期水文预报[M]. 南京:河海大学出版社,1999:6-20.
FAN Zhong-xiu. Medium and long-term hydrology prediction [M]. Nanjing: Press of Hohai University, 1999: 6-20.
- [10] 邓聚龙. 灰理论基础[M]. 武汉:华中科技大学出版社,2002:225-238.
DENG Ju-long. Base of grey theory[M]. Wuhan: Press of Huazhong University of Science and Technology, 2002: 225-238.

(编辑:孙广增)