

基于时间序列分解方法的太湖未来特征水位预测

尹义星^{1,2,3}, 许有鹏², 张小娜¹, 焦士兴⁴, 闫桂霞¹

(1. 南京信息工程大学 应用水文气象研究院, 江苏 南京 210044; 2. 南京大学 地理与海洋科学学院, 江苏 南京 210093;
3. 河海大学 水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098; 4. 安阳师范学院 资源环境与旅游学院, 河南 安阳 455002)

摘要: 基于太湖沿湖 5 个水文台站 1956—2006 年的逐日平均水位数据, 提取得到了逐年特征(平均、最高和最低)水位, 运用时间序列分解方法模拟和预测了太湖未来 15 a 的特征水位, 并探讨了太湖未来洪灾情势。将水位序列分解为趋势成分、周期成分和随机成分, 得到时间序列分解模型, 并进行了水位序列模拟和预测。分析结果表明, 该模型的模拟精度比较理想, 可以实现未来较长期的特征水位预测。预测得到未来 15 a 后的最高水位可能达到 4.05 m, 显著超过太湖警戒水位, 因此需引起相关部门的重视。

关键词: 太湖; 特征水位; 时间序列分解; 预测

文献标识码: B

文章编号: 1000-288X(2012)04-0172-04

中图分类号: P343

Forecasting Characteristic Water Levels of Taihu Lake Based on Time Series Decomposing Method

YIN Yi-xing^{1,2,3}, XU You-peng², ZHANG Xiao-na¹, JIAO Shi-xing⁴, YAN Gui-xia¹

(1. *Applied Hydrometeorological Research Institute, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China*; 2. *School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing, Jiangsu 210093, China*; 3. *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, Jiangsu 210098, China*; 4. *Department of Resource & Environment and Tourism, Anyang Normal University, Anyang, He'nan 455002, China*)

Abstract: Based on the daily average water levels of the five gauges around the Taihu Lake observed from 1956 to 2006, the annual characteristic water levels(average, highest and lowest) were obtained. Time series analysis method was adopted, and the decomposition model of trend, period and random components was applied to simulate and forecast the annual characteristic water levels of Taihu Lake. The annual characteristic water levels were forecasted for the next 15 years and the future flood scenarios of Taihu Lake basin were also addressed. The results indicate that the highest water level of Taihu Lake may reach 4.05 m in 2021, which is significantly higher than the warning water level of Taihu Lake and thus worthy of attention for relevant department.

Keywords: Taihu Lake; characteristic water levels; time series decomposition; forecasting

太湖作为中国 5 大淡水湖泊之一, 是太湖流域最大的吞吐型平原湖泊, 也是太湖流域工农业供水和人民生活用水的主要水源; 另一方面, 汛期的太湖又是环湖地区洪水策源地, 给当地工农业生产和人民生命财产造成重大损失。太湖为四周高中间低的碟形地貌特征, 暴雨后水流向中心汇集, 容易造成短时间内水位快速上升, 由于其地势特征, 洪水不易于外泄, 加上人类活动影响等原因, 太湖水位时常居高不下, 对

周围和地势低洼平坦的下游地区构成严重的洪水威胁。作为太湖流域的洪水调节中枢, 太湖的水位高低反映流域洪涝灾害的严重程度; 20 世纪 90 年代以后太湖持续高水位, 而且经常出现“小水大灾”的现象, 引起了研究者的关注。

水位预测通常采用数理统计^[1-3]或水量平衡^[4]等方法。时间序列分析方法是应用水文要素的观测记录, 寻找其自身的演变规律来进行预报, 常用的有平

收稿日期: 2011-06-06

修回日期: 2011-09-26

资助项目: 水利部公益性行业科研专项经费项目“改善长江三角洲地区水系结构与河湖连通研究”(201201072); “长江三角洲地区城市化对洪涝孕灾环境的影响研究”(200701024); 国家自然科学基金重点项目(40730635); 水文水资源与水利工程科学国家重点实验室开放研究基金项目(2011491111); 南京信息工程大学科研基金项目(20100406); 江苏省高校自然科学基金基础研究资助项目(11KJB170008)

作者简介: 尹义星(1974—), 男(汉族), 安徽省和县人, 博士, 讲师, 主要从事变化环境下的水文水资源、流域洪涝灾害、水文气象方面的研究。
E-mail: yinyx@nuist.edu.cn.

稳时间序列中的自回归模型、周期均值迭加、马尔可夫链等。齐文等^[5]研究表明时间序列分析方法是预测存在趋势性、周期性和季节性波动的湖泊水位复杂变化的较好途径。自回归滑动平均模型(auto-regression moving average, ARMA)是通用的时间序列预测方法,也是一种精确度相当高的短期预测方法。湖泊水位是各种物理因子综合作用的结果,由于这些因子表现出趋势变化、周期变化和随机变化等特征,因而水位序列有趋势、周期和随机变化等成分。本研究采用时间序列分解法来分别模拟水位序列的趋势、周期和随机成分,从而实现太湖水位的中长期预报,进一步对流域未来的洪涝灾害进行预评估。

1 数据与方法

1.1 数据来源

太湖水位是衡量太湖水情的最重要指标。1997年起,太湖流域管理局经过分析,采用沿湖5站(大浦口、望亭(太)、西山、小梅口、夹浦)水位平均值代表太湖水位,沿用至今^[6]。这里采用的数据是这5个台站1956—2009年逐日水位平均值,并在此基础上提取得到年最高和最低水位,最终得到1956—2009年逐年的年平均水位,最高水位和最低水位,用于后文的模拟和预测。

1.2 模拟和预测方法

从分析水文要素自身演变规律这一角度出发,选择时间序列分解法来建立水文序列中长期预报模型。将年水位时间序列分为周期(谐波)成分、趋势成分和随机成分,即

$$X_t = T_t + P_t + R_t \quad (t=1, \dots, N) \quad (1)$$

式中: X_t ——年水位时间序列; T_t ——趋势项; P_t ——周期项; R_t ——随机项; N ——序列长度。

趋势项反映水文现象因水文或气象因素而引起的多年变化趋势;周期项反映水文现象的周期性变化;随机项是水位序列中去除了趋势和周期之后的成分。其中的趋势成分采用多项式进行拟合,周期成分采用谐波分析进行拟合。趋势和周期成分均属于确定性成分,剔除确定性成分后,剩余序列为随机成分,其随机成分由ARMA模型来进行模拟和预测。ARMA模型阶数确定根据自相关图,并结合最小信息准则AIC^[7]和贝叶斯准则BIC^[8]等确定。

这里实际谐波成分数目的确定,主要通过方差贡献率来进行分析,其各谐波分量所解释的方差计算公式为:

$$C_j^2 = \frac{A_j^2 + B_j^2}{2s^2} \quad (2)$$

式中: C_j^2 ——第 j 个谐波的方差贡献率; A_j, B_j ——傅里叶系数; s^2 ——序列 X_t' [见公式(3)] 的总方差。这里的分解模型为:

$$X_t = T_t + X_t', \quad X_t' = P_t + X_t'', \quad X_t'' = R_t \quad (3)$$

因此得到模型的一般形式为:

$$X_t = (a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p) + (A_0 + \sum_{j=1}^m [A_j \cos(\frac{2\pi j}{N} t) + B_j \sin(\frac{2\pi j}{N} t)]) + \text{ARMA} \quad (4)$$

预测模型为:

$$T_{N+z} = a_0 + a_1(N+z) + a_2(N+z)^2 + \dots + a_p(N+z)^p \quad (5)$$

$$P_{N+z} = A_0 + \sum_{j=1}^m \{A_j [\cos(\frac{2\pi}{T_j}(N+z))] + B_j \sin[\frac{2\pi}{T_j}(N+z)]\} \quad (6)$$

式中: z ——预测步数。

建立模型的过程即是从已知序列中提取各成分的过程,提取的顺序为,先趋势成分,其次周期成分,最后为随机成分。一旦建立了各成分的数学模型后,再将其线性叠加,就得到了公式(1)形式的分解模型。采用上述时间序列分解模型来对趋势、周期和随机成分分别进行模拟和预测,综合它们的结果,得到最终的预测成果。

1.3 结果检验

对模拟效果的检验采用效率系数(CoE)。对预测误差的评价采用二个指标:均方根误差(RMSE)和平均绝对误差(MAE)。具体为:

$$\text{CoE} = 1 - \frac{\sum_{i=1}^N (L_i - \hat{L}_i)^2}{\sum_{i=1}^N (L_i - \bar{L}_i)^2} \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (L_i - \hat{L}_i)^2}{N}} \quad (8)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |L_i - \hat{L}_i| \quad (9)$$

式中: L_i ——实测水位; \hat{L}_i ——模拟所得水位; \bar{L}_i ——实测水位的均值。其余符号意义与前面相同。

2 结果与分析

2.1 序列分解

为了对太湖特征水位序列进行成分分解,首先对平均、最高和最低水位序列进行趋势检验,从而进一步分解趋势项。趋势检验方法采用文献[9]中的改进MK方法。分析中由于最低水位序列1阶自相关系数比较显著(0.498),且其Sen坡度仅为0.003,小于0.01,因此进行预置白处理。平均水位和最高水位的

1 阶自相关系数均不显著,因此未进行预置白处理。MK 趋势检验表明,年平均和最低水位均存在显著上升,MK 检验 Z 值分别为 2.91 和 2.94,均达到 0.01 显著性水平,而最高水位 Z 值为 1.48,没有达到显著性水平。因此对平均水位和最低水位序列进行多项式拟合,得到消除趋势后的序列,结果如图 1 所示。

进一步进行谐波分析和周期成分拟合。表 1 为平均水位序列谐波分析的主要参数,由表 1 可见其第 1 至第 14 个谐波成分累计占原序列方差贡献率的 90.7%,因此这里选择前 14 个谐波进行周期成分的拟合。最高和最低水位序列与其比较类似,因篇幅所限,此处不再赘述。

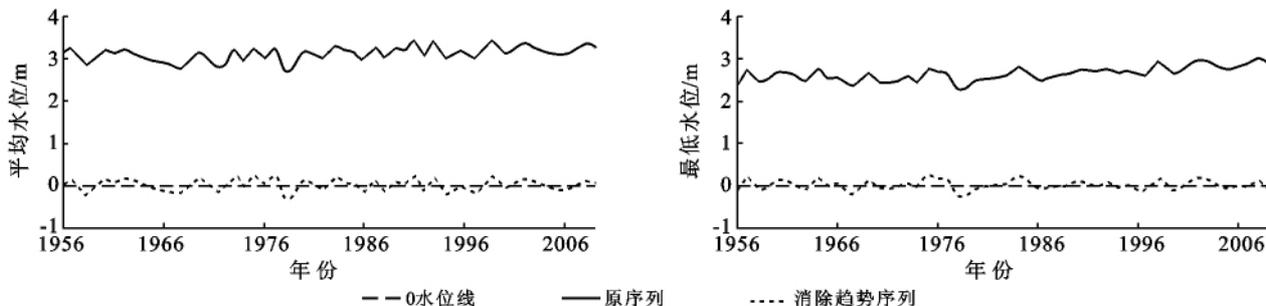


图 1 太湖平均和最低水位及消除趋势序列

表 1 平均水位序列谐波分析的参数

序号	周期/a	傅立叶系数 A	傅立叶系数 B	累积方差贡献率/%	序号	周期/a	傅立叶系数 A	傅立叶系数 B	累积方差贡献率/%
1	2.00	-0.083	0.000	15.4	8	3.60	0.001	-0.051	63.6
2	9.00	0.052	-0.034	24.1	9	3.00	-0.007	0.049	69.2
3	6.75	0.053	-0.031	32.6	10	3.18	-0.046	0.016	74.7
4	2.57	-0.045	0.033	39.6	11	2.25	-0.046	-0.015	79.9
5	13.50	-0.029	0.046	46.2	12	3.38	0.034	0.030	84.6
6	2.08	0.051	-0.009	52.1	13	3.86	0.002	0.043	88.7
7	7.71	-0.028	-0.042	57.9	14	4.15	0.017	0.024	90.7

在消除趋势和周期序列后,即得到随机序列。一般认为,在原始序列滤去趋势项和周期项后的剩余序列为平稳随机序列,但这里仍然对剩余序列进行平稳性检验,即对平均、最高和最低水位各自消除趋势和周期成分后的随机序列进行单位根检验。由于已经对原序列消除了趋势(最高水位序列趋势不显著,除外),故采用不包含趋势项的模型进行单位根检验。由麦金农法计算的 1%,5%,10% 临界 τ 统计量分别为 -3.565, -2.920 和 -2.598,而平均、最高和最低水位的计算 τ 值分别为 -7.318, -8.590 和 -10.272。证明均不存在单位根,所得的序列在 99% 的置信水平上是平稳的,符合平稳序列的要求。

2.2 随机模型识别与定阶

ARMA 模型应用的关键是模型识别与定阶,不同阶次的模型预测结果差别较大。模型的识别与定阶可以通过样本的自相关与偏自相关函数的观察获得。需要强调的是,虽然通常采用 AIC 值选择合适的 ARMA 模型,但 AIC 值最小化并不是得出最优 ARMA 模型的充分条件。本研究采用的方法是先通

过最小 AIC 和 BIC 准则来建立模型,并对估计结果进行参数显著性检验和残差随机性检验。如果通过检验,则此模型可以看作为最优模型;如果不能通过,则选取次小的 AIC 和 BIC 值并进行相关的统计检验,依此类推,直至选到合适的模型为止。

随机序列的自相关和偏自相关分析可得,对于平均水位序列,其自相关系数在 1 阶后截尾,偏自相关系数在滞后 6 阶仍然显著。依据残差序列是否为随机序列、AIC 与 BIC 最小等原则,最后确定采用 AR-MA(1,1)模型;最高水位序列自相关系数在 1 阶后截尾,偏自相关系数 2 阶截尾;最低水位序列自相关系数在 1 阶后截尾,偏自相关系数也是 1 阶后截尾。同样依据残差序列是否为随机序列、AIC 与 BIC 最小等原则,最终确定最高水位采用 MA(1)模型;最低水位采用 AR(1)模型。

2.3 随机模型检验

一个好的模型应得到统计上独立的残差,该残差序列只包含噪声而没有规律性变化。为验证残差序列的相互独立性,对残差自相关函数进行分析。自相

关图分析结果表明其没有明显的相关性(图略),因此可以看作白噪声序列。同时,由各序列随机模拟残差的QQ图可见(图略),残差序列符合正态分布,因此选用的模型是合适的。

2.4 模型预测

采用时间序列分解模型对趋势、周期和随机项分别进行模拟,并加和得到综合的模拟和预测结果。预

测过程中,将2006—2009年作为检验时段,其预报值与实测值的相对误差绝对值均在5%以下,可见其拟合精度较高。水位实测值与模型计算值对比如图2所示。

总体来说,模拟值与实测值变化趋势一致,且误差较小,因此模型的模拟效果较好,可以用于未来水位预报。

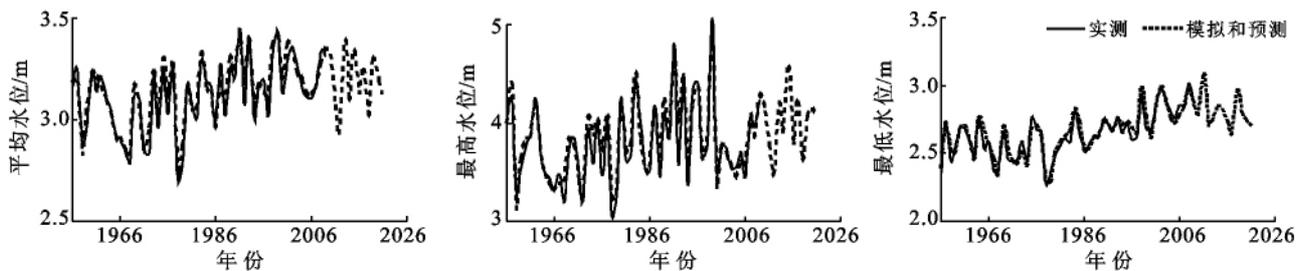


图2 太湖特征水位综合模拟和预测

模拟和预测结果的评定详见表2,平均、最高和最低水位模拟的效率系数均在90%以上,且其均方根误差RMSE和平均绝对误差MAE均小于0.1m,因此预测效果比较理想。预测得到未来15a后即2021年的平均、最高和最低水位分别为3.09、4.05和2.70m。太湖的警戒水位为3.50m,西山站的警戒水位为3.60m。一般而言,太湖水位超过4.00m时,防汛就开始紧张,并可能出现一些区域性灾害。因此预测得到的2021年最高水位值得重视。

表2 太湖未来特征水位模拟和预测结果

特征水位	效率系数/%	均方根误差/m	平均绝对误差/m	2021年水位/m
平均水位	92.04	0.05	0.04	3.09
最高水位	90.64	0.09	0.09	4.05
最低水位	90.84	0.05	0.04	2.70

3 结论

时间序列分析是对有序的数据进行数据处理、寻求规律、预测未来等的有力工具,特别是对规律性强的随机数据的预测,其结果是令人满意的。本研究运用时间序列分解方法建立太湖年特征水位预测模型,结果表明,其拟合精度和预测精度均较高。值得注意的是,在模型的实际应用过程中,有效谐波数和ARMA模型阶数的确定是建模的关键,采用方差贡献率分析

来进行谐波数的确定,在ARMA模型阶数的确定中,综合考虑了AIC和BIC准则以及残差随机性检验等,取得了比较理想的模拟效果。因此,基于趋势变化、周期变化和随机变化分析的时间序列分解模型,用于模拟和预测太湖特征水位的效果较理想。

[参考文献]

- [1] 门玉明,晏长根.季节性指数平滑法在地下水水位预报中的应用[J].工程勘察,2000(2):25-27.
- [2] 方中英,钟茂生,王明文,等.鄱阳湖水位动态预测模型[J].江西师范大学学报:自然科学版,2003,27(3):232-236.
- [3] 吴震.用滑动平均值分析下关年最高水位的变化趋势[J].气象科学,2000,20(1):106-110.
- [4] 曲耀光.青海湖水量平衡及水位变化预测[J].湖泊科学,1994,6(4):298-307.
- [5] 齐文,郑绵平.西藏扎布耶盐湖水位 Winters 和 ARIMA 模型分析[J].湖泊科学,2006,18(1):21-28.
- [6] 戈礼宾,潘彩英,张泉荣,等.太湖水位资料系列分析与水位特征值计算[J].水资源研究,2008,29(4):7-8.
- [7] Akaike H. A new look at the statistical model identification[J]. IEEE Transactions on Automatic Control, 1974, 19(6):716-723.
- [8] Schwarz G. Estimating the dimension of a model[J]. The Annals of Statistics, 1978, 6(2):461-464.
- [9] 尹义星,许有鹏,陈莹.1950—2003年太湖流域洪旱灾害变化与东亚夏季风的关系[J].冰川冻土,2010,31(2):381-388.