



# 基于多元线性回归分析的水位短期预测方法

范先友, 李朝阳, 杨传华, 曹晓勤

(长江武汉航道局, 湖北 武汉 430014)

**摘要:** 为了给长江电子航道图终端用户提供更为准确的水深预测信息, 以统计学为基础, 通过回归分析的方法, 以预测长江监利站水位为例, 探讨基于多元线性回归分析的水位短期预测方法实施步骤和实际应用, 取得了很好的效果。

**关键词:** 回归分析; 水位; 短期预测

中图分类号: U 69

文献标志码: A

文章编号: 1002-4972(2014)01-0159-04

## Short-term water level prediction method based on multiple linear regression analysis

FAN Xian-you, LI Zhao-yang, YANG Chuan-hua, CAO Xiao-qin

(Changjiang Wuhan Waterway Bureau, Wuhan 430014, China)

**Abstract:** To provide more accurate water depth prediction data for the terminal users of the Yangtze electronic navigational chart, this paper discusses the implementation procedure and practice of water level short-term forecasting method based on the statistics and using the regression analysis method by taking the Yangtze Jianli gauging station for example, and has achieved satisfactory results.

**Key words:** regression analysis; water level; short-term prediction

长江电子航道图系统是基于最新的河床地形和实时水位数据向航行船舶提供可通航水域的智能化动态航行参考系统, 能够为船舶提供准确的自身位置信息、实际水深数据、航标位置、周围航行船舶信息、最新航行通告、航道地物地貌信息、水深短期预测信息等<sup>[1]</sup>。通过长江电子航道图系统给营运船舶终端用户提供及时准确的航道水深短期预测信息, 能够为船舶合理配载、优化营运提供重要参考依据, 具有一定的经济价值。

回归分析是研究两个变量或多个变量之间因果关系的统计方法, 是一种“由因果果”的定量分析、预测技术。其基本思想是, 在相关分析基础上, 对具有相关关系的两个或多个变量之间数量变化的一般关系进行测定, 确立合适的数学模型, 从而通过自变量的已知或设定值来估计和预测因变量均值。本文以统计学为基础, 应用回归

分析方法, 以长江中游监利水位站为例, 利用长江中游宜昌至城陵矶各站点历史水位资料, 建立回归分析预测模型, 并运用此模型预测监利站短期内的水位变化情况。

### 1 基于多元线性回归分析的水位短期预测方法实施步骤

某河段水位的涨落变化, 主要受其上游来水、支流入汇和降雨等综合因素影响。一方面, 上游来水量的变化, 会体现在上游水位的涨落变化上, 而上游来水量的变化需经过一段流程时间, 才能体现在其下游某河段的水位涨落上; 另一方面, 某河段下游河段由于受支流入汇、降雨等多种原因产生的壅水作用, 也会对其上游河段的水位变化有较大影响。基于此思路, 我们把水位涨落波动的差值作为回归分析的变量, 再利用

收稿日期: 2013-04-02

作者简介: 范先友(1955—), 男, 高级工程师, 主要从事航道管理工作。

统计软件 (SPSS软件) 进行多元线性回归分析。

### 1.1 根据预测目标确定自变量和因变量

根据需要预测监利水位的目标, 把预测开始时间作为当日, 确定以预测未来1天 (预测日) 的监利水位与其前1日 (当日) 的水位差值作为因变量 (y)。

通过前面分析和对长江水位统计资料的初步观察, 找到可能影响监利水位涨落变化的因素, 并经过初步筛选, 选出主要的影响因素作为自变量。确定以下因素作为自变量: 宜昌当日水位与其前1日水位差值 (x<sub>1</sub>)、宜昌前1日水位与前2日水位差值 (x<sub>2</sub>)、沙市当日水位与其前1日水位差值 (x<sub>3</sub>)、沙市前1日水位与前2日水位差值 (x<sub>4</sub>)、监利当日水位与其前1日水位差值 (x<sub>5</sub>)、城陵矶当日水位与其前1日水位差值 (x<sub>6</sub>)。

### 1.2 建立回归预测模型<sup>[2-3]</sup>

建立多元线性回归预测模型时, 为了保证回归模型具有较好的解释能力和预测效果, 在选择自变量时应遵循以下一些准则: 1) 自变量对因变量必须有显著的影响, 并呈密切的线性相关; 2) 自变量与因变量之间的线性相关必须是真实的, 而不是形式上的; 3) 自变量之间应具有一定的互斥性, 即自变量之间的相关程度不应高于自变量与因变量之间的相关程度; 4) 自变量应具有完整的统计数据, 其预测值容易确定。

遵照自变量的选择准则, 对前文确定的自变量和因变量进行适应性分析:

1) 长江监利水位站位于宜昌、沙市水位站下游, 位于城陵矶水位站上游。上游来水情况和下游壅水作用对监利水位有直接的、显著的影响, 而且上下游的水位涨落关系有较密切的线性相关性。

2) 以长江相邻各水位站点的水位涨落差值作为自变量和因变量, 它们之间的依存关系和线性相关性是真实存在的。

3) 由于长江沿程水位涨落相互影响、制约, 所以确定的各自变量之间有一定的相关性, 但自变量之间的相关程度低于自变量与因变量之间的相关程度。

4) 长江中游重点港埠历史水位资料齐全、完

整, 所有水位资料均逐日记录, 其预测值也比较容易确定, 因此经计算处理可以作为自变量的统计数据。

如上所述, 可以依据前文确定的自变量和因变量, 建立经验回归方程, 即多元线性回归分析预测模型为:

$$y=b_0+b_1x_1+b_2x_2+b_3x_3+b_4x_4+b_5x_5+b_6x_6+e \quad (1)$$

式中: b<sub>0</sub>为常数项, e为随机误差, b<sub>1</sub>, b<sub>2</sub>, ..., b<sub>6</sub>为回归系数。b<sub>1</sub>为x<sub>2</sub>, x<sub>3</sub>, ..., x<sub>6</sub>固定时, x<sub>1</sub>每增加一个单位对y的效应, 即x<sub>1</sub>对y的偏回归系数; 同理b<sub>2</sub>为x<sub>1</sub>, x<sub>3</sub>, ..., x<sub>6</sub>固定时, x<sub>2</sub>每增加一个单位对y的效应, 即x<sub>2</sub>对y的偏回归系数, 等等。

### 1.3 水位数据整理与计算

所采用的样本水位数据为2008年9月1日—2012年6月30日的长江宜昌、沙市、监利和城陵矶逐日水位数据资料。按照确定的自变量、因变量和建立的回归模型, 根据回归分析对水位数据的统计需要, 对样本水位数据资料进行整理和计算, 包括对水位资料的补缺和推算、对不可靠水位数据进行核实、对确定有误的水位值加以修正、对逐日水位涨落幅度进行差值计算、规范化处理计算数据等。

### 1.4 利用SPSS软件进行线性回归分析<sup>[4]</sup>

将经过规范化处理后的数据导入到SPSS软件中, 选用进入法进行线性回归分析, 并设置运行参数为:

- 1) 按前文所述设置因变量和自变量;
- 2) 选别变量的准则为系统默认值, 即使用回归方程的显著性统计量F所对应的概率P (Sig) 值;
- 3) 选入变量的显著性水平为0.05, 即当P≤0.05时, 就将该变量选入回归方程;
- 4) 剔除变量的显著性水平为0.10, 即当P≥0.10时, 就该变量剔除回归方程。

通过设置相应参数, 利用SPSS运行分析后, 得到水位预测模型为:  $y = (-4.121 \times 10^{-5}) + 0.113x_1 - 0.033x_2 + 0.320x_3 + 0.043x_4 - 0.094x_5 + 0.469x_6$ , 具体输出结果见表1~3。

表1 模型汇总

模型	R	R <sup>2</sup>	调整R <sup>2</sup>	标准估计误差
1	0.907	0.822	0.822	0.075

表1中,复相关系数 $R$ 反映了模型自变量 $(x_1, \dots, x_6)$ 与因变量 $y$ 之间线性回归关系的密切程度。建立的模型越接近于实际,通过样本数据计算得出的构造统计量 $R^2$ 越接近于1。由表中可知,模型调整 $R^2$ 系数为0.822,标准估计误差为0.075,说明因变量与各自变量的相关度较高,回归模型比较接近实际情况。

表2 方差分析

模型	平方和	df	均方	F	Sig
回归	36.460	6	6.077	1 070.318	0.000
残差	7.880	1 388	0.006		
总计	44.340	1 394			

表2列出了回归拟合的方差分析结果,其中包括回归平方和、残差平方和及自由度等。通过方差分析构造统计量 $F$ 可以检验回归方程的显著性,统计量 $F$ 同样是通过样本数据计算得出。从表2中可知,当预测因子为 $x_1, \dots, x_6$ 时,检验值 $F=1 070.318$ ,显著性概率值 $Sig=0.000<0.05$ ,说明 $x_1, \dots, x_6$ 的回归系数不为零,模型具有统计学意义。

表3 模型系数

模型 因变量: y-监差0	非标准化系数		标准系数		t	Sig
	B	标准误差	试用版			
(常量)	$-4.121 \times 10^{-5}$	0.002			-0.020	0.984
$x_1$ -宜差1	0.113	0.007	0.303		17.121	0.000
$x_2$ -宜差2	-0.033	0.008	-0.088		-4.114	0.000
$x_3$ -沙差1	0.320	0.018	0.518		18.262	0.000
$x_4$ -沙差2	0.043	0.013	0.069		3.382	0.001
$x_5$ -监差1	-0.094	0.033	-0.094		-2.852	0.004
$x_6$ -城差1	0.469	0.020	0.485		23.436	0.000

从标准差分析,非标准化系数均具有较高可靠性;从显著性分析,非标准化常量系数( $b_0$ )的显著性概率值 $Sig=0.984>0.10$ ,不具有统计学意义,剔除该常数项。由此可以得到,非标准化方程为 $y=0.113x_1-0.033x_2+0.320x_3+0.043x_4-0.094x_5+0.469x_6$ ;标准化方程为 $y=0.303x_1-0.088x_2+0.518x_3+0.069x_4-0.094x_5+0.485x_6$ 。标准系数表示为在方程其它变量不变的情况下, $x_1$ 改变一个单位, $y$ 改变0.303个单位; $x_2$ 改变一个单位, $y$ 改变-0.088个单位;其它依此类推。可以看

出,自变量 $x_1, x_3, x_6$ 与因变量 $y$ 的关系最密切,即预测日监利水位的涨落幅度主要受宜昌当日水位涨落幅度、沙市当日水位涨落幅度和和城陵矶当日水位涨落幅度的影响。

### 1.5 计算并确定水位预测值

利用前述分析得到的非标准化回归方程,通过已知或设定自变量的值,计算得出水位涨落幅度的预测值,并对预测值进行综合分析,考虑局部流域降雨、水位大幅涨落等因素影响,估计随机误差,确定最终的预测值。将最终确定的水位涨落幅度预测值,加上当日水位,得到未来一日的水位预测值。

### 1.6 模型实际预测检验

回归预测模型能否用于实际预测,还必须对模型的预测能力加以检验。不难理解,回归方程对于样本期间来说是正确的,但是对于实际预测是否合适,还必须检验模型参数值的稳定性和相对于样本容量变化时的灵敏度,也必须研究确定模型是否可以用于样本数据以外的范围,具体做法是:

1) 采用把增大样本容量以后模型分析的结果与原来的模型分析结果进行比较,并检验其差异的显著性。

2) 把预测模型用于样本以外某一段时间的实际预测,并将预测值与实际观测值作一比较,然后检验其差异的显著性。

表4为增大样本容量(增加2012年7月1日—9月30日水位数据作为样本)的模型分析系数,通过增大样本容量进行回归分析,得到的模型系数与原样本分析结果对比,差异不大,说明模型参数有很好的稳定性。

表4 增大样本容量的模型系数

模型 因变量: y-监差0	非标准化系数		标准系数		t	Sig
	B	标准误差	试用版			
(常量)	-0.001	0.002			-0.395	0.693
$x_1$ -宜差1	0.127	0.006	0.349		19.862	0.000
$x_2$ -宜差2	-0.023	0.008	-0.063		-2.998	0.003
$x_3$ -沙差1	0.282	0.017	0.462		16.398	0.000
$x_4$ -沙差2	0.050	0.012	0.082		4.201	0.000
$x_5$ -监差1	-0.116	0.032	-0.116		-3.615	0.000
$x_6$ -城差1	0.490	0.020	0.495		24.413	0.000

## 2 水位短期预测应用

长江电子航道图(2.0版)系统全线试运行后,根据电子航道图试运行工作需要,对长江中游水位站点开展了水位短期预测预报工作。通过一段时间的水位预测预报,并将预测结果与实际水位观测值进行比较(图1)。由图1可以看出,在水位涨落幅度不大的中枯水期和相对平稳的高洪水期,水位预测准确性相对较高;在水位经常有陡涨陡落的中洪水期,预测误差较大。总体来说,从水位实际预测应用情况看,短期内水位预测误差不大,说明模型有很好的适应性,能够用于对水位的短期预测。

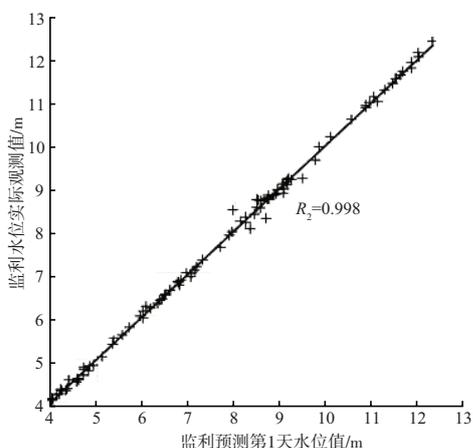


图1 监利预测第1天水位与实际水位对比

需要说明的是,一个回归模型只是针对某一个水位站点的水位短期预测,要预测各个不同站点水位,需根据每个水位站的具体情况,分析影响因素,建立相应回归模型,进行回归分析和参数估计,进而得到各自的回归预测方程。

## 3 结语

1) 预测江河水位短期变化,在资料较充分的条件下,多元线性回归分析是一个有力的研究工具。利用多元线性回归分析方法,对长江中游水位进行短期预测,具有一定的可操作性和可信度,其预报精度较高、误差较小。

2) 回归预测方程是一种经验公式,其准确性决定于因变量与自变量之间的依存关系和依赖程度,选择适当的相关因子,对预测精度有很大的影响。

3) 应用该方法进行水位预测,应综合考虑河道流域的自然特性和地理因素;要进一步提高预测精度,还应结合实践经验和其它预测方法进行综合分析和验证。

### 参考文献:

- [1] 何明宪. 长江电子航道图系统的研发与建设[EB/OL]. (2012-08-08) [2013-03-27] <http://www.cjhdj.com.cn/main.do?method=item&id=43319>.
- [2] 包凤达, 翁心真. 多元回归分析的软件求解和案例解读[J]. 数据和管理, 2000(5): 56-61.
- [3] 多元线性回归分析预测法[DB/OL]. (2011-07-14) [2013-03-27] <http://wenku.baidu.com/link?url=fMR5NVSywhNt4ksh9k416NCHUpzWDTmDyYNbp7nEKqvEkkNL3aF7zX01fjXgR5HC3c-sJQwjM3sXU1FEh-TiIF2dBw9mT-zveQqccbQ0PO3>.
- [4] 顾伟欣, 周红. 基于逐步回归分析的孔隙度预测方法[J]. 石油地质与工程, 2008(1): 37-39.

( 本文编辑 郭雪珍 )



## · 消 息 ·

### 中国交建2013年获国家授权专利512项再创新记录

中国交建2013年获国家授权专利总数再创新记录,全年共获国家授权专利512项,同比增长24%,其中发明专利70项。公司获得授权专利数量已连续7年实现大幅增长,创新科技成果不断涌现,充分体现了公司及各单位对知识产权保护工作的高度重视,凝聚了广大科技人员的不懈努力。

摘编自《中国交通建设网》