文章编号: 1008-1542(2006)02-0129-04

GM(1, 1)模型在长江水质预测中的应用

刘玉敬, 阎晨光, 索秀云, 王 群

(河北科技大学理学院, 河北石家庄 050054)

摘 要: 介绍了 GM(1, 1)的原息模型、新息模型和等维新息模型及残差修正模型的建模方法, 分别建立了长江水质预测的等维新息模型和残差修正模型。通过对模型的预测结果进行分析, 并对比原数据的趋势, 发现两模型在数据突变点附近有明显差别, 其数据说明残差修正模型在细节方面更接近实际, 其结果体现了数据的内在规律。

关键词: GM(1, 1)模型; 等维新息模型; 残差修正模型; 水质预测

中图分类号: N941.5; O212 文献标识码: A

Application of GM(1, 1) model in prediction of the quality of the Yangtse river

LIU Yu-jing, YAN Chen-guang, SUO Xiu-yun, WANG Qun

(College of Sciences, Hebei University of Science and Technology, Shijiazhuang Hebei 050054, China)

Abstract The modelling methods of old information model, new information model, equal dimension new information model and modified model of residual based on GM(1, 1) are introduced; the equal dimension new information model and the modified model of residual are set up separately. According to the result and the trend of the data, we think that the result by the modified model of residual could represent the inherence of the data. The two models are obviously different at the aberrant point, but the data show the modified model of residual is more actual in details.

Key words GM(1, 1) model; equal dimension new information model; modified model of residual; predicting of water quality

长江是中国第一、世界第三大河流, 长江水质的污染程度日趋严重, 已引起了政府相关部门和专家们的高度重视。2004年 10月, 由全国政协与中国发展研究院联合组成“保护长江万里行”考察团, 从长江上游宜宾到下游上海, 对沿线 21个重点城市作了实地考察, 揭示了一幅长江污染的真实画面, 其污染程度让人触目惊心。目前面临的问题是如何根据过去 10年的主要统测数据, 对长江未来水质污染的发展趋势作出准确的分析、预测。笔者以 2005年全国大学生数学建模竞赛 A题及其数据为例, 分析在建立预测模型中出现问题的原因, 通过建立 GM(1, 1)的改进模型, 提高了模型的预测能力和精确度。

1 问题分析

经分析、计算发现长江污水排放量除存在缓慢增长的大趋势外, 还有一个潜在周期震荡现象, 据以作出近 10年长江流域污水和可饮用水所占比例的数据图 (见图 1)。

收稿日期: 2005-11-10; 修回日期: 2005-12-13; 责任编辑: 冯 民
基金项目: 国家“八六三”计划课题(2002AA639300)
作者简介: 刘玉敬(1976-), 女, 河北唐山人, 助教, 主要从事计算机应用数学方面的研究。

由于该数据的特殊性, 笔者选择 GM(1, 1)模型。GM(1, 1)模型是灰色系统理论的预测模型[1], 在许多领域有实际应用, 用其进行预测既有很多成功的实例, 同时也存在一些预测偏差过大的情况。因此, 对 GM(1, 1)模型进行深入研究, 找出影响模型精度及其适应性的关键因素, 提高其预报精确度, 具有非常重要的理论价值和实践意义。许多研究者已经对此课题进行了大量研究[1~3], 并已有实际应用[4~6]。但理论研究一般只限于对模型本身进行修改, 而实际预测就必须针对数据的不断反馈进行后验检验。笔者分析了各种 GM(1, 1)模型预测原理, 并对其试验结果进行对比, 结合实际数据变化趋势, 得到了影响模型预测精度的大体原因, 认为 GM(1, 1)残差修正模型针对某些不平稳随机数列可以提高精度, 并在数据细节方面更接近实际。

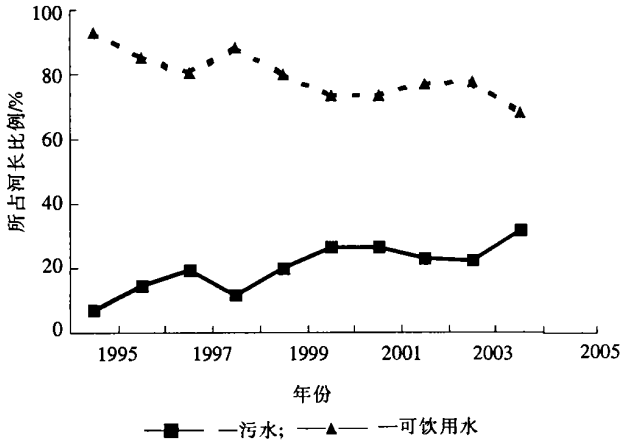

图 1 近 10年数据散点图

Fig 1 Scatter diagram of data of last ten years

2 GM(1, 1)预测模型

2.1 GM(1, 1)原息模型

设原始数据列为 $X^{(0)}(k)=\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$, 累加得到生成数据列

$X^{(1)}(k)=\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\}$, 其中: $x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i)$, $k=0, 1, \cdots, n$。

该累加生成数据列的微分方程描述为

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=u, \tag{1}$$

其时间响应函数为

$$\hat{x}^{(1)}(k+1)=\left(x^{(0)}(1)-\frac{u}{a}\right)\mathrm{e}^{-ak}+\frac{u}{a},$$

则经过累减后的还原值为

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)=(1-\mathrm{e}^{a})\left(x^{(0)}(1)-\frac{u}{a}\right)\mathrm{e}^{-ak}, \tag{2}$$

式(2)中参数 a, u 可用最小二乘法拟合, $\begin{pmatrix}a\\u\end{pmatrix}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_n$,

其中: $\boldsymbol{B}=\begin{pmatrix}-Z^{(1)}(2) & 1\\ -Z^{(1)}(3) & 1\\ \vdots & \vdots\\ -Z^{(1)}(n) & 1\end{pmatrix}$, $\boldsymbol{Y}_n=\begin{pmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{pmatrix}$。

一般背景值 $Z^{(1)}(k+1)=\frac{x^{(1)}(k+1)+x^{(1)}(k)}{2}$, 在文献[2]和文献[3]中背景值改进为

$$Z^{(1)}(k+1)=\frac{x^{(1)}(k+1)-x^{(1)}(k)}{\ln[x^{(1)}(k+1)]-\ln[x^{(1)}(k)]}。$$

2.2 GM(1, 1)新息模型

一般而言, 随着时间的推移, 影响预测对象的因素在不断变化, 若直接用 GM(1, 1)原息模型进行长期预测, 一方面预测精度不断降低, 另一方面模型未能反映出预测对象的变化趋势, 其预测的可信度很小。因此, 必须充分引入已知信息来反映预测对象的变化和状态, 或在无已知信息的情况下, 用灰色信息来淡化灰平面的灰度。具体建模方法: 建立 GM(1, 1)原息模型后可得 $n+1$时刻的预测值 $\hat{x}^{(0)}(n+1)$, 加入灰数 $\hat{x}^{(0)}(n+1)$, 得到新息数据列 $X^{(0)}=$

$\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n), \hat{x}^{(0)}(n+1)\}$, 可在此数据列基础上建立 GM(1, 1)新息模型[7]。

2.3　GM(1, 1)等维新息模型

其思想是在原息模型的基础上, 预测得 $n+1$时刻的预测值 $\hat{x}^{(0)}(n+1)$, 加入灰数 $\hat{x}^{(0)}(n+1)$, 去掉$x^{(0)}(1)$, 重新构成等维新息序列: $X^{(0)} = \{x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n), \hat{x}^{(0)}(n+1)\}$, 建立新的 GM(1, 1)模型, 预测$n+2$时刻的值 $\hat{x}^{(0)}(n+2)$, 加入 $\hat{x}^{(0)}(n+2)$, 去掉 $x^{(0)}(2)$, 构成新的等维新息序列, 如此类推, 建立新模型[7]。

2.4　GM(1, 1)残差修正模型

为提高模型预测精度, 可对 GM(1, 1)残差序列进行周期分析及修正。文献[6]中将残差序列分别按 2, 3, …, $\frac{n}{2}$年的周期进行排列, 计算出各周期的拟合误差值, 经比较选出拟合误差最小的周期作为残差序列拟合的主要周期, 然后用残差序列减去主要周期之差建立新的时间序列, 找出校正周期。笔者按此方法计算出残差序列主要周期 $T=6$, 校正周期 $T=5$或者 $T=6$。

2.5　模型的对比分析

从理论上讲, 上述 4个模型并不属于同种类型。

GM(1, 1)新息模型和 GM(1, 1)等维新息模型均从模型的构造入手, 意图从根本上提高模型的精确度。但都有各自缺陷: 随着数据量的加大和时间推移, GM(1, 1)新息模型的原有数据可能会干扰和影响预测结果, 模型会逐渐失去其实际应用的意义; 而 GM(1, 1)等维新息模型逐渐将原有数据舍去, 把未经任何修正的预测结果作为已知数据建立新模型, 也存在放大误差的可能。

GM(1, 1)的残差修正模型属于后验模型。根据原数据对预测结果进行残差分析后, 结合其周期性, 得到了误差最小的复合残差, 并用该复合残差去修正预测结果, 很好地再现了数据的趋势。

3　模型求解和分析

用等维新息模型与残差修正模型预测长江可饮用水所占河长比结果见表 1。

表 1　等维新息模型与残差修正模型预测值比较

Tab. 1　Comparison for the predicting between the equal dimension new information model and the modified of residual

%

年份	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
实测值	93.1	85.3	80.7	88.4	80.2	74	73.7	76.7	77.5	68
等维新息模型		85.630 0	83.683 3	81.780 5	79.920 9	78.103 7	76.327 8	74.592 4	72.896 3	71.238 8
精确度		0.996	0.963	0.925	0.997	0.945	0.964	0.973	0.941	0.952
残差修正模型		85.252 4	83.206 1	87.832 2	77.665 6	75.025 9	73.747 5	74.193 8	78.067 8	70.534 3
精确度		0.999	0.969	0.994	0.968	0.986	0.999	0.967	0.993	0.963
年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
等维新息模型	69.619 0	68.010 4	65.432 4	65.034 2	64.118 5	62.126 5	59.888	58.414 5	58.102 2	55.700 5
残差修正模型	69.619 0	69.241 4	67.558 9	72.540 8	62.722	60.422 1	59.475 7	60.246 4	64.437 6	57.214 1

从表 1可以看出, 虽然两模型对现有数据的预测精确度都很高, 基本达到精度要求, 但在预测未来数据上的表现有所不同, 相差较大。为比较模型的优劣, 作出预测值的比较图, 见图 2。

从图 2可清楚地看到: 等维新息模型未能很好地体现原始数据的周期震荡性质, 只是在一定程度上反映了原始数据的缓慢降低的趋势; 相对而言, 残差修正模型不仅较好地反映了原始数据递减的大趋势, 更难能可贵的是, 该模型还几乎完美地再现了数据周期震荡的趋势, 较好地表现了数据的本质趋势。从这一点上可以说, 残差修正模型要优于等维新息模型。

另外, 1998年洪灾导致当年长江全流域水量猛增, 可饮用水比例突然升高, 这属于数据上的突变点。两模型预报时表现也不同。等维新息模型在 1998年的准确度降到了 92.5%, 2000年的准确度也只有 94.5%。而残差修正模型在预测时表现较好, 1998年准确度为 99.4%, 2000年准确度也高达 98.6%。笔者利用残差修正模型预测其在 2008年的数据为 72.540 8%。两模型孰优孰劣, 有待证实。

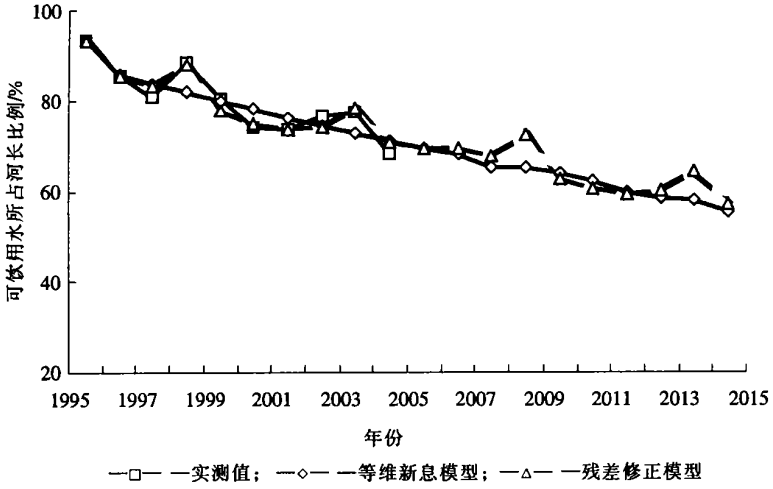

图 2 模型预测值比较

Fig. 2 Comparison of the predicting of the models

4 结 论

1)从理论上讲,由于引进了新灰数,所以新息模型要优于原息模型,更适用于动态预测。等维新息模型还具有所需数据量较少、计算方便等优点,从预测结果看,其精度较高。

2) GM(1, 1)等维新息模型虽然预测现有数据精度很好,但不能很好反映原始数据的周期震荡性,必须结合残差序列的周期分析来修正模型,从而将模型的预报精确度大大提高。

3)通过分析可知,两模型的不同表现原因在于等维新息模型只注重了数据降低的长期趋势,而残差修正模型还模拟了数据的周期震荡性,从而在数据的突变点预报方面更胜一筹。但这种对周期震荡性的模拟只是出于对原始数据的分析,并没有考虑到影响长江水质的经济、天气、地质等因素,可能不尽完善。

参考文献:

[1] 邓聚龙.灰色系统基本方法[M]. 武汉: 华中工学院出版社, 1987.
[2] 罗 党,刘思峰,党耀国.灰色模型 GM (1, 1)优化[J].中国工程科学, 2003, 5(8): 50-53.
[3] 王钟羡,吴春笃. GM (1, 1)改进模型及其应用[J].数学实践与认识, 2003, 33(9): 20-25.
[4] 郝永红,黄登宇,张文忠,等. GM (1, 1)的周期修正模型及其在泉水流量预测中的应用[J].数学实践与认识, 2003, 33(9): 35-37.
[5] 叶明全,胡学钢. GM(1, 1)残差修正的季节性神经网络预测模型及其应用[J].计算机工程与应用, 2005, (4): 194-196.
[6] 陈观浩.灰色系统理论在水稻纹枯病发生面积预报中的应用[J]. 广东农业科学, 2000, (1): 37-38.
[7] 张立亭,祝国瑞,周世健,等. 时间序列的灰色建模试验[J].武汉大学学报(工学版), 2005, (2): 117-120.

(上接第 118页)

参考文献:

[1] DARIS J M, ELOE P W, HENDERSON J. Triple positive solutions and dependence on higher order derivatives[J]. J Math Appl, 1999, 237: 710-720.
[2] DARIS J M, HENDERSON J. Triple positive symmetric solctions for a Lidstone boundary value problem[J]. Differential Equations Dynam Systems, 1999, 7: 321-330.
[3] DARIS J M, HENDERSON J, WONG P J Y. General Lidstone problems: multiplicity and symmetry of solutions[J]. J Math Anal Appl, 2000, 251: 527-548.
[4] 郭彦平,葛渭高.高阶微分方程边值问题的多个正解存在性[J].系统科学与数学, 2003, 23(4): 529-535.
[5] AERY R I A generalization of the Leggett-Williams fixed point theorem[J]. Math Sci Res Hot-Line, 1999, 3: 9-14.