文章编号:1007- 4929(2007)05- 0054-03

城市工业用水量的灰色马尔可夫预测模型

宋巧娜[1],唐德善[2]

(1.河海大学商学院,南京 210098;2.河海大学水利水电学院,南京 210098)

摘　要:灰色 GM(1,1)是预测城市工业用水量的模型,这种模型不适合长期的、随机和波动性较大的数据序列预测,但是马尔可夫模型适合描述随机波动性较大的预测问题。可以将这两种模型结合,构建灰色马尔可夫预测模型。按特定的状态划分方法,先用灰色 GM(1,1)预测模型进行预测,再用马尔可夫模型预测结果进行优化,使预测精度大大提高。最后以抚顺市为例,预测结果证明了该模型的优势。

关键词:城市工业用水量;预测模型;灰色 GM(1,1)模型;马尔可夫模型

中图分类号:TV213.4　　文献标识码:A

Grey Markov Prediction Model for City Industrial Water Quantity

SONG Qiao-na[1], TANG De-shan[2]

(1. Business School of Hohai University, Nanjing 210098, China;

2. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

Abstract: GM(1,1) model is used for forecasting city industrial water quantity. This model is not suitable for forecasting data sequence in long term with randomness and great fluctuation. While Markov model is suitable to treat with problems with these characteristics. Two models are combined together to form a new model, namely grey Markov model, in this paper. According to the special state-division ways proposed in this paper, grey GM(1,1) model is used to forecast the data sequence firstly, then, Markov model is used to optimize the results from the grey GM(1,1) model. So, the prediction result is much more precise. Finally, taking the Fushun city as an example, the advantage of the grey Markov model is proved by the prediction result.

Key words: city industrial water quantity; prediction model; grey GM(1,1) model; Markov model

0　引　言

城市用水量预测是进行城市建设规划、输配水系统优化调度的一项十分重要的前提工作。而城市中工业用水量占城市用水量的 2/3,其创造的效益更是其他行业无法比拟的,所以预测城市工业用水量,无论在经济效益还是在宏观调控上都有重要意义。因为影响城市工业用水量的因素很多,如工业企业的数目、企业生产规模和工艺路线等,工业用水量有很大的波动性、随机性。现在的预测方法主要有回归分析法时间序列法、指数平滑法[1]和灰色模型法[2]等。每种方法,既有其优点,同时又不可避免其缺点。本文尝试将灰色马尔可夫模型运用到城市工业用水量预测领域,即在 GM(1,1)模型的基础上,进一步运用马尔可夫模型对其结果进行优化。

灰色 GM(1,1)模型要求数据序列成指数规律变化,且在建模前通过累加处理使数据的随机性弱化,因此,该模型不适合长期的、随机性和波动性较大的数据序列预测,通常GM(1,

收稿日期:2006-12-28

作者简介:宋巧娜(1978-),女,博士研究生,研究方向为技术经济及管理、水资源规划与管理。

1)模型用来揭示数据的发展趋势。而马尔可夫随机性过程理论指出:系统将来所处的状态只与现在系统状态有关,而与系统过去状态无关(即无后效性)。它可以描绘一个随机变化的动态系统,根据状态之间的转移概率来推测一个系统未来的发展变化,适合描述随机波动性较大的预测问题,通常该模型用来确定状态的转移规律。因而,将两模型结合使用,建立灰色马尔可夫预测模型,一方面利用 GM(1,1)模型使数据序列满足了马尔可夫的前提条件,即无后效性和平稳过程等均值特点;另一方面,马尔可夫模型又解决了对随机波动性较大序列的预测问题。因此,预测精度大大提高。

1　灰色预测模型 GM(1,1)

1.1　原始数据处理

城市工业用水量的原始数据具有随机性,为了找出其中的内部规律,弱化原始数据的随机性,增强其规律性,可以借助累加(累减)生成的方法对其作必要的处理。

记 $X^{(0)}$ 为原始序列:

$$X^{(0)}=(X_{(1)}^{(0)},X_{(2)}^{(0)},\cdots,X_{(n)}^{(0)})$$

记 $X^{(1)}$ 为生成序列:

$$X^{(1)}=(X_{(1)}^{(1)},X_{(2)}^{(1)},\cdots,X_{(n)}^{(1)})$$

且 $X_{(k)}^{(1)}=\sum_{i=1}^{k}X_{(i)}^{(1)}$,$X^{(1)}$ 的上标(1) 表示一次累加生成,记作 1-AGO。

因为灰色 GM(1,1)模型实质为指数方程,要求用于预测的样本数据也要符合指数规律,因此,要进行序列的规律性检验。

(1)光滑性检验,若序列 X 满足:

$$\frac{\rho(k+1)}{\rho(k)}<1,k=2,3,\cdots,n-1$$

$$\rho(k)\in[0,\varepsilon];\quad k=3,4,\cdots,n$$

$$\varepsilon<0.5$$

则称 X 为准光滑序列。

式中:$\rho(k)=\frac{X_{(k)}^{(0)}}{X_{(k-1)}^{(0)}}$;$k=2,3,\cdots,n$,为序列 X 的光滑比。

(2) 指数规律性检验,若序列 X 满足:

$$\sigma_{(k)}^{(1)}\in[1,1+\delta];k=3,4,\cdots,n$$

$$\delta=0.5$$

则称 X 具有准指数规律。

式中:$\sigma_{(k)}^{(1)}=\frac{X_{(k)}^{(1)}}{X_{(k-1)}^{(1)}}$,$k=2,3,\cdots,n$,为序列 X 的级比。

一般情况下,对于非负的准光滑序列通过(一次)累加呈现出(准)指数规律,即可建立指数模型。

1.2　模型的建立

GM(1,1)模型是指一阶、一个变量的微分方程预测模型,是一阶单序列的线性动态模型,用于时间序列预测的是离散形式的微分方程模型[3]。其具体形式是:

$$\frac{\mathrm{d}x}{\mathrm{d}t}+\alpha x=u \tag{1}$$

式中:α,u 为待识别的灰色参数。式(1) 表示一个单变量 x 对时间 t 的一阶微分方程是连续的。它的离散形式为:

$$\hat{x}_{(t+1)}^{(1)}(k)=\left(x_{(1)}^{(0)}-\frac{u}{\alpha}\right)\mathrm{e}^{-\alpha t}+\frac{u}{\alpha}\quad(t=1,2,\cdots,n)$$

式中:$\hat{\alpha}=(\alpha,u)^T=(B^TB)^{-1}B^TY_N$,其中:

$$B=\begin{bmatrix}-z^{(1)}(2) & 1\\ -z^{(1)}(3) & 1\\ \vdots & \vdots\\ -z^{(1)}(n) & 1\end{bmatrix}\qquad Y_N=\begin{bmatrix}x_{(2)}^{(0)}\\ x_{(3)}^{(0)}\\ \vdots\\ x_{(n)}^{(0)}\end{bmatrix}$$

$z^{(1)}(t)=\frac{x_{(t)}^{(1)}+x_{(t+1)}^{(1)}}{2}\quad(t=2,3,\cdots,n)$ 为均值生成序列。

累减还原:式(1)是数据序列预测值的累加,进行累减运算将其还原成原始数据序列的预测值。

$$\hat{X}_{(1)}^{(0)}=X_{(1)}^{(0)}$$

$$\hat{X}_{(t+1)}^{(0)}=\hat{X}_{(t+1)}^{(1)}-\hat{X}_{(t)}^{(1)}$$

$$\hat{y}_{(t)}=\hat{X}_{(t+1)}^{(0)}$$

上式中:$\hat{y}(1)$ 对应的是原始序列第二个数据的预测值 $\hat{X}_{(2)}^{(0)}$。

2　马尔可夫预测模型

马尔可夫预测法,是一种预测事件发生的概率的方法,它是基于马尔可夫链,根据事件的目前状况预测其将来各个时刻变动状况的一种预测方法。

2.1　状态划分

了解状态的转移规律是利用马尔可夫链对系统未来发展变化做出预测的关键。因此,为构造状态转移概率矩阵,首先作状态的划分,即将数据序列分成若干状态。

状态划分与预测正确与否有很大关系。根据数据序列的不同特点采用不同方法进行划分,主要有相对值法、平行曲线法、残差的标准化离差法及相对误差法。对于抚顺地区工业用水量的预测,采用相对值法。

将预测值与实际值的偏差放大 100 倍后得 δ,再加以分析。

$$\delta=100\times[x_{(k+1)}^{(0)}-\hat{x}_{(k+1)}^{(0)}]/x_{(k+1)}^{(0)}\quad k=1,2,\cdots,n-1$$

确定以 δ 所处的不同上下阈值作为状态划分的标准。

2.2　状态转移概率(矩阵)的计算状态转移概率

$$P_{ij}(k)=\frac{M_{ij}(k)}{M_i}\quad i=1,2,\cdots,n$$

式中:$M_{ij}(k)$ 表示由状态 E_i 经过 k 步转移到状态 E_j 的原始数据的个数;M_i 为处于状态 E 的原始数据的个数;$P_{ij}(k)$ 表示由状态 E_i 经 k 步转移到状态 E_j 的概率。状态转移概率矩阵[4]:

$$P(k)=\begin{bmatrix}p_{11}(k) & p_{12}(k) & \cdots & p_{1n}(k)\\ p_{21}(k) & p_{22}(k) & \cdots & p_{2n}(k)\\ \vdots & \vdots & \cdots & \vdots\\ p_{n1}(k) & p_{n2}(k) & \cdots & p_{nn}(k)\end{bmatrix}$$

由于数据序列最后的状态转向不确定,故在构建 k 步转移概率矩阵时要去掉数据序列中最末的 k 个数据。

2.3 预测值的计算

确定了系统的未来转移状态 E_i，也就确定了预测值的变动区间$[E_{1i},E_{2i}]$。由于状态的划分有多种方法，预测值的计算方法也不相同。本文采用如下公式计算：

$$\hat{y}(k)=x_{(k+1)}^{(0)}(1-\bar{\delta}\%)$$

式中：$\bar{\delta}=\dfrac{(\delta_1+\delta_2)}{2}$，$\delta_1$，$\delta_2$ 分别为某一特定状态的上下限。

2.4 模型的检验

模型建立是否合理可信，要通过一定的检验，常用方法有相对误差检验和后验差检验。这里选用相对误差检验法。

绝对误差为：

$$\varepsilon_{(t)}=x_{(t)}^{(0)}-\hat{x}_{(t)}^{(0)}$$

相对误差为：

$$\Delta_{(t)}=\left|\frac{\varepsilon_{(t)}}{x_{(t)}^{(0)}}\right|$$

预测精度为：

$$\theta_{(t)}=(1-\Delta_{(t)})\times100\%$$

3 案例分析

用 1994～2003 年抚顺地区工业用水量数据来预测 2004 年抚顺地区的工业用水量。

3.1 建立 GM(1,1)模型

经检验样本能够通过序列的规律性检验，经计算得 $\alpha=-0.0098$，$u=5.1614$，白化形式微分方程的解为：

$$\hat{x}_{(t+1)}^{(1)}(k)=\left(x^{(0)}(1)-\frac{5.1614}{-0.0098}\right)e^{0.0098t}-\frac{5.1614}{0.0098}$$

$$\hat{X}_{(t+1)}^{(0)}=\hat{X}_{(t+1)}^{(1)}-\hat{X}_{(t)}^{(1)}=(1-e^{\alpha})\left(x_{(1)}^{(0)}-\frac{u}{\alpha}\right)e^{-\alpha t}=$$

$$5.1781e^{0.0098t}$$

$$\hat{y}=\hat{X}_{(t+1)}^{(0)}$$

1994～2003 年实际工业用水量及预测值见表 1。

2004 年模型预测值为 $\hat{X}_{(11)}^{(0)}=\hat{y}(10)=5.7113$ 亿 m^3。

表 1 抚顺 1994～2003 年实际工业用水量及预测值 亿 m^3

年份	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
实际值	4.3017	5.1620	5.1944	5.3278	5.4283	5.2091	5.4204	5.3892	5.6394	5.9307
预测值	5.1781	5.2291	5.2806	5.3326	5.3851	5.4381	5.4917	5.5458	5.6004	5.6556
δ	1.59	1.28	1.63	0.09	−0.80	4.21	1.30	2.81	−0.70	−4.87
状态	3	2	3	2	1	4	2	4	1	1

3.2 状态划分

采用相对值法进行状态划分，划分标准如表 1，状态划分结果如表 2。

表 2 状态划分标准

状态	状态界限
E1	$\delta=(-5.0,0)$
E2	$\delta=(0,1.5)$
E3	$\delta=(1.5,2)$
E4	$\delta=(2.0,5.0)$

3.3 构造状态转移概率矩阵

这里，只考虑一步转移概率，将最后一个数据状态去掉，故：

$$M_1=2,M_2=3,M_3=2,M_4=2$$

$$P(1)=\begin{bmatrix}1/2 & 0 & 0 & 1/2\\1/3 & 0 & 1/3 & 1/3\\0 & 1 & 0 & 0\\1/2 & 1/2 & 0 & 0\end{bmatrix}$$

3.4 预测值的计算

2003 年的工业用水量处于状态 1，经过 1 年的转移，由 $\max P_{1j}=1/2$，可以认为 2004 年处于状态 1 的可能性最大，则预测值为：

$$\hat{y}(10)=x_{(k+1)}^{(0)}(1-\bar{\delta}\%)=5.5827\text{ 亿 m}^3$$

3.5 模型的检验

2004 年抚顺地区工业用水量实际值为 5.4982 亿 m^3，则相对误差 $\Delta_{(11)}=0.0154$，预测精度 $\theta_{(11)}=98.46\%$；单纯用 GM(1,1) 计算的结果 $\Delta'_{(11)}=0.0387$，预测精度 $\theta'_{(11)}=96.13\%$。将两种方法的结果进行比较，可见灰色马尔可夫模型比单纯的 GM(1,1)精度明显提高，充分体现出其优越性。

4 结 语

本文将传统的 GM(1,1)模型与马尔可夫模型有机结合起来，建立灰色马尔可夫模型，并用于抚顺地区工业用水量预测。新建模型兼具 GM(1,1)模型与马尔可夫模型优点，既能揭示数据的发展趋势，又适合于波动性较大的随机序列预测，预测精度明显高于单一的灰色模型。

新建模型预测效果良好，该方法对水资源管理具有重要的理论和实际意义。

参考文献：

[1] 张 雄. 城市用水量预测模型综合研究[J]. 水资源与水工程学报，2005，16(4)：24—28.

[2] 邓聚龙. 灰色预测与决策[M]. 武汉：华中理工学院出版社，1985：7—9.

[3] 傅 立. 灰色系统理论及其应用[M]. 北京：科学技术文献出版社，1991：54—56.

[4] 胡奇英，刘建庸. 马尔可夫决策过程引论[M]. 西安：西安电子科技大学出版社，2000：2—4.

[5] 周 刚，王弘宇，胡春雪，等. 应用灰色新陈代谢 GM(1,1)模型预测中长期城市需水量[J]. 中国农村水利水电，2005，(8)：16—18.