

灰色马尔可夫预测模型在工业 SO₂ 排放量中的应用*

王 艳 玲

(重庆师范大学 数学与计算机科学学院, 重庆 400047)

摘 要 :灰色马尔可夫预测模型是将灰色系统理论和马尔可夫链理论相结合建立的预测模型 ,它不仅充分发挥了灰色预测模型和马尔可夫预测模型的优点 ,而且因为马尔可夫链理论的引入 ,有效地解决了灰色预测模型对于随机波动性较大的数列预测精度低问题。首先建立 GM(1 , 1)灰色动态拟合模型 ,并以此作为工业 SO₂ 排放量发展变化的动态基准线模型 ,在此基础上应用马尔可夫链确定系统状态转移概率矩阵 ,通过系统状态的划分、样本值与模型拟合值之间的残差等指标的分析计算 ,最终以概率形式分析和预测工业 SO₂ 排放量的发展变化区间。理论分析和实践都表明 ,该法不但预测结果更可靠 ,而且能够对工业 SO₂ 排放量的发展趋势进行宏观的把握 ,有利于决策者的决策行为。

关键词 :灰色系统 ;灰色马尔可夫模型 ;SO₂

中图分类号 : F409. 6 ; O211. 62

文献标识码 : A

文章编号 : 1672-6693(2008)02-0074- 04

我国学者邓聚龙^[1]提出灰色理论以来 ,灰色预测已广泛地应用于许多科学领域 ,如农业、经济、运输、气象等。灰色预测^[2-7]的优势在于短期预测 ,缺点在于对长期预测和波动性较大数据列的拟合较差。马尔可夫链预测^[8-10]适用于随机波动性大的动态过程 ,在这一点上恰好可以弥补灰色预测的局限。但马尔可夫链预测对象要求具有马氏性和平稳性。而客观世界中的预测问题大多是随时间变化或呈某种变化趋势的非平稳过程 ,因此把 GM(1 , 1) (灰色预测)与马尔可夫链预测两者结合起来形成灰色马尔可夫链预测模型 ,两者的结合可以避免考虑其他多种影响因素 ,同时又可以预测未来一段时间的变化 ,具有较大的科学性和实用性。

我国 SO₂ 的排放量逐年提高 ,SO₂ 对人体、动植物、金属和生态环境都有很大的危害。它的大量排放使城市的空气污染不断加重 ,致使我国出现大面积的酸雨 ,并不断发展。随着环保意识的加强 ,控制和治理 SO₂ 污染成为我国当前和今后一段时期内紧迫的任务。而 SO₂ 的排放量受能源结构、环保政策和环保技术等多方面原因影响 ,是一个随机波动性很大的动态过程 ,正是基于以上分析 ,笔者认为工业 SO₂ 排放量是一个部分信息已知 ,部分信息未知的系统 ,即工业 SO₂ 排放量作为系统行为的主要特征

量是一个灰色量 ,把灰色预测和马尔可夫链预测两种方法结合为一 ,取长补短 ,用 GM(1 , 1)模型来揭示其长期发展变化的某种总趋势 ,用马尔可夫模型来确定现象状态之间的转移 ,建立灰色马尔可夫预测模型 ,对工业 SO₂ 排放量进行实证研究 ,希望能对研究控制和治理 SO₂ 问题提供一定的参考。

1 灰色马尔可夫链预测过程

1.1 建立 GM(1 , 1)模型^[5-6, 11]

将 1994—2004 年工业 SO₂ 排放量视为原始灰色序列 $X^{(0)}$,则

$$X^{(0)} = (x^{(0)}(1) , x^{(0)}(2) , \dots , x^{(0)}(11)) = \\ (1825 , 1405 , 1364 , 1363 , 1594 , 1460 , 1612 , \\ 1566 , 1562 , 1792 , 1891)$$

弱化 $X^{(0)}$ 的随机性作一次累加生成得到

$$X^{(1)} = (x^{(1)}(1) , x^{(1)}(2) , \dots , x^{(1)}(11))$$

作 $X^{(1)}$ 的紧邻均值生成序列

$$Z^{(1)} = (z^{(1)}(2) , z^{(1)}(3) , \dots , z^{(1)}(10)) ,$$

则一次累加生成序列 $X^{(1)}$ 的规律可用最小二乘法求解 GM(1 , 1)模型

$$x^{(0)}(k) + a z^{(1)}(k) = b \quad (1)$$

得 $\hat{u}=(a\ b)=(0.033\ 1241.76)$
其中 a 称为发展系数 b 称为灰色作用量。则模型
(1)的时间响应函数为

$$\hat{x}^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a}=39\ 330.259\ 8e^{0.033k}-37\ 162.259\ 8\quad (2)$$

其中 $k=0\ 1\ \dots\ 10$ 。由(2)式算得的数据为累加生成序列 $X^{(1)}$ 的拟合值,当 $k>10$ 时算得的是预测值。然后再用累减生成还原得到原始数据序列 $X^{(0)}$ 的一次拟合值

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)=(1-e^a)(x^{(0)}(1)-\frac{b}{a})e^{-ak}\quad k=0\ 1\ 2\ \dots\ 10\quad (3)$$

1.2 马尔可夫链预测模型

1.2.1 状态划分 对于一个符合 n 阶马尔可夫非平稳随机序列 $y(t)\ (t=1\ 2\ \dots\ n)$,其任一状态 E_i 可表述为 $E_i\in[\otimes_{1j},\otimes_{2j}]$ 其中 $\otimes_{1j}=y(t)+A_i,\otimes_{2j}=y(t)+B_i\ y(t)=\hat{x}^{(0)}(k+1)\ A_i=a_i\bar{y}\ B_i=b_i\bar{y}\ j=1\ 2\ \dots\ m\ k=0\ 1\ 2\ \dots\ n\ t=1\ 2\ \dots\ n\ \bar{y}$ 为原始数据的均值。

根据工业 SO_2 排放量的数据分布状况和一次拟合值的残差,将残差序列可划分为 3 个、4 个或 5 个状态,经代入计算 3 个和 5 个状态时二次拟合值误差较大,故本文选取划分为 4 个状态,即

$$\begin{aligned}\otimes_1:\otimes_{11}&=y(t)-0.08\bar{y} & \otimes_{12}&=y(t)-0.04\bar{y} \\ \otimes_2:\otimes_{21}&=y(t)-0.04\bar{y} & \otimes_{22}&=y(t) \\ \otimes_3:\otimes_{31}&=y(t) & \otimes_{32}&=y(t)+0.04\bar{y} \\ \otimes_4:\otimes_{41}&=y(t)+0.04\bar{y} & \otimes_{42}&=y(t)+0.08\bar{y}\end{aligned}$$

由于 $y(t)$ 是一个时间函数,因而灰元 $\otimes_{1j},\otimes_{2j}$ 也随时序发生变化,即状态 $\otimes_i$ 具有动态性。

1.2.2 确定转移概率矩阵 马尔可夫过程^[9]是一种“无后效性”的随机过程,即一个时间序列,它在未来是什么状态,将取什么数值只与它现在的状态、现在取什么数值有关,而与以前状态和数值无关。所以,状态划分好后,就可以利用 m 步状态转移概率的计算公式

$$p_{ij}^{(m)}=\frac{M_{ij}^{(m)}}{M_i}\quad (4)$$

其中 $p_{ij}^{(m)},M_{ij}^{(m)}$ 为工业 SO_2 排放量从状态 E_i 经过 m 步转移到状态 E_j 的概率和次数, M_i 为状态 E_i 出现的年数。针对本文的状态划分,可得各年份所处的状态(表 1)。

由于 2004 年的下一步状态未知,在求状态转移概率矩阵时不考虑该年的状态转移。依(4)式 $P_{11}^{(1)}$

是由状态 E_1 经一步转移到 E_1 的概率,表 1 中处于状态 E_1 的有 4 年,经一步转移后状态仍为 E_1 的有 1 年。

表 1 各年份所属状态

年份	1994	1995	1996	1997	1998	1999
状态	E_3	E_4	E_2	E_1	E_4	E_1
年份	2000	2001	2002	2003	2004	
状态	E_3	E_1	E_1	E_3	E_4	

即 $P_{11}^{(1)}=\frac{1}{4}$,同理可求 $P_{ij}^{(1)}\ i,j=1\ 2\ 3\ 4$,得到一步转移概率矩阵

$$P^{(1)}=\begin{bmatrix}1/4 & 0 & 1/2 & 1/4 \\ 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 2/3 \\ 1/2 & 1/2 & 0 & 0\end{bmatrix}$$

确定了预测对象未来的状态转移后,即确定了预测值变动的区间 $[\otimes_{j1},\otimes_{j2}]$,得到序列的模拟值

$$\begin{aligned}\hat{y}(k) &= \hat{y}(k)=0.5\times(\otimes_{j1}+\otimes_{j2})= \\ &= \hat{x}^{(0)}(k)+0.5\times(A_i+B_i) \\ &(k=1\ 2\ \dots\ 11;j=1\ 2\ 3\ 4)\end{aligned}\quad (5)$$

由(5)得二次拟合值,见表 2。

由表 2 可知,马尔可夫链模型的相对误差要比 GM(1,1)模型小得多。

1.3 确定预测值

为了预测 2005 年的工业 SO_2 排放量,需考虑 2004 年的排放量 $p_{4j}^{(1)},j=1\ 2\ 3\ 4$,该年处于状态 E_4 ,若 $\max_k p_{4k}^{(1)}=p_{4j}^{(1)}\ (k,j=1\ 2\ 3\ 4)$ 。则认为序列的下一状态将处于 E_j ,因第 4 行中有两个概率相近,需要考虑多步转移概率 $p_{4j}^{(2)}=(1/2\ 0\ 1/2\ 0),p_{4j}^{(3)}=(1/2\ 0\ 0\ 1/2),p_{4j}^{(4)}=(1\ 0\ 0\ 0)^{[10]}$ 。此时有 $\max_k p_{4k}^{(4)}=p_{41}^{(4)}$,则 2005 年的工业 SO_2 排放量将处于状态 $E_1\in[\otimes_{11},\otimes_{12}]$,得 2005 年的工业 SO_2 排放量为 1795.9 万 t。

2 模型精度检验^[11]

只有通过检验的模型才能用来预测。通常,一个模型要经过多种检验才能判定其是否合理。这里用残差和均方差比值来检验模型。

2.1 残差检验

根据表 2 中灰色马尔可夫预测模型计算出的模拟序列 $\hat{X}^{(0)}$ 的残差序列和相对误差序列,得到平均

$$\begin{aligned}\text{相对误差}\ \bar{\Delta} &= \frac{1}{n}\sum_{k=1}^n\Delta_k=\frac{1}{11}(|\Delta_1|+|\Delta_2|+\dots \\ &+|\Delta_{11}|)=0.012\approx0.01\end{aligned}$$

表 2 我国工业 SO₂ 排放量的拟合值、预测值比较

项目		GM(1, 1) 模型			灰色马尔可夫模型		
年份	实际值 / 万 t	一次拟合值 / 万 t	残差	相对误差	二次拟合值 / 万 t	残差	相对误差
1994	1825	1825	0	0	1825	0	0
1995	1405	1336.4	68.6	4.9	1431.5	- 26.5	- 1.9
1996	1364	1381.8	- 17.8	- 1.3	1350.1	13.9	1.0
1997	1363	1428.8	- 65.8	- 4.8	1333.7	29.3	2.1
1998	1594	1477.3	116.7	7.3	1572.4	21.6	1.4
1999	1460	1527.5	- 67.5	- 4.6	1432.4	27.6	1.9
2000	1612	1579.4	32.6	2.0	1611.1	0.9	0.05
2001	1566	1633.1	- 67.1	- 4.3	1538.0	28	1.8
2002	1562	1686.6	- 124.6	- 8.0	1591.5	- 29.5	- 1.9
2003	1792	1746.0	46.0	2.6	1777.7	14.3	0.8
2004	1891	1805.6	85.4	4.5	1900.7	- 9.7	- 0.5

注 :表中实际值摘自《中国统计年鉴》1995—2005 年

给定 α ,当 $\bar{\Delta} < \alpha$ 且 $\Delta_n < \alpha$ 时 ,称模型为残差合格模型。

2.2 均方差比值检验

设 $\bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k) , S_1^2 = \frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \bar{x}]^2$ 分别为 $X^{(0)}$ 的均值和方差 $\bar{\varepsilon} = \frac{1}{n} \sum_{k=1}^n \varepsilon(k) , S_2^2 = \frac{1}{n} \sum_{k=1}^n [\varepsilon(k) - \bar{\varepsilon}]^2$ 分别为残差序列 $\varepsilon(0)$ 的均值和方差 ; $C = S_2/S_1$ 为均方差比值 ,代入表 2 中数据得 $C = 0.000\ 04$ 。

对于给定的 $C_0 > 0$,当 $C < C_0$ 时模型为均方差比合格模型。

表 3 给出了精度检验的等级参照表。经验证 ,该模型的预测精度达到一级 ,可用来预测。

表 3 精度检验等级参照表^[11]

等级	一级	二级	三级	四级
平均相对误差 $\bar{\Delta}$	0.01	0.05	0.10	0.20
均方差比值 C_0	0.35	0.50	0.65	0.80

3 结语

把 GM(1, 1) 灰色预测)与马尔可夫链预测两者结合起来形成灰色马尔可夫链预测模型 ,用 GM(1, 1)揭示工业 SO₂ 排放量数据列的发展变化总趋势 ,用马尔可夫链预测来确定其状态的变化规律。两者结合可以避免考虑其他多种影响因素 ,能充分利用时间序列数据给予的信息进行分析预测 ,从而比较准确地预测未来几年的工业 SO₂ 排放量 ,

提高了预测精度。

参考文献 :

[1] 邓聚龙. 灰色系统基本方法[M]. 郑州 :河南大学出版社 ,1987 5-25.

[2] 王丰效. 基于初值修正的非等距灰色预测模型[J]. 重庆师范大学学报(自然科学版) 2006 23(3) 42-44.

[3] 李东 ,苏小红 ,马双玉. 基于新维灰色马尔科夫模型的股价预测算法[J]. 哈尔滨工业大学学报 2003 35(2) : 244-248.

[4] 潘晓琳. 一种改进的人力资源价值灰色评价方法[J]. 重庆师范大学学报(自然科学版) 2004 21(1) 59-61.

[5] 邓聚龙. 灰色系统理论教程[M]. 武汉 :华中理工大学出版社 ,1991 89-95.

[6] 李茂林 ,魏勇. 灰色 GM(1, 1)模型的一种优化组合方法[J]. 西华师范大学学报(自然科学版) 2007 28(1) : 40-44.

[7] 周介铭 ,彭文甫. 影响四川省粮食生产因素的灰色分析与粮食产量预测[J]. 四川师范大学学报(自然科学版) 2005 28(3) 350-353.

[8] 叶宗文 ,罗珊 ,曾波 ,等. 股票价格的马氏链预测法[J]. 重庆师范大学学报(自然科学版) 2006 23(1) 64-66.

[9] 孙荣恒. 随机过程及其应用[M]. 北京 :清华大学出版社 2004 38-53.

[10] 贺福利 ,胡勇 ,陈淳. 用灰色马尔柯夫链预测模型对我国粮食产量的预测[J]. 数学的实践与认识 2003 33 (12) 45-47.

[11] 刘思峰 ,党耀国 ,方志耕. 灰色系统理论及其应用[M]. 北京 :科学出版社 2005.

Applying Grey Markov Model to SO₂ Emissions from Industry

WANG Yan-ling

(College of Mathematics and Computer Science , Chongqing Normal University , Chongqing 400047 , China)

Abstract Grey Markov model is a forecast model based on the grey system theory and Markov chain theory. It not only makes good use of the advantages of grey forecast model and Markov model , but also solves efficiently the problem of low precision of number list forecast. What's more , according to the forecast about the sulfur dioxide emissions , the grey Markov forecast model can satisfy the requirement in forecast precision. In this article , a Grey Markov Model for forecasting sulfur dioxide emissions is presented by means of combining Grey system theory with dispersed Markov Chains theory. For the first time , grey model and Markov chain are integrated to forecast sulfur dioxide emissions. Firstly a GM(1 , 1) is built to get the dynamic base line for sulfur dioxide emissions. Secondly on the basis of the GM(1 , 1) , Markov chain is applied to achieve state transition probability matrix. Thirdly sulfur dioxide emissions interval is forecasted and analyzed in the form of probability by the system state classification , the calculation of the residue between true value and model fitting value and the standardization of deviation of the residue. It's proved in theory and practice that the forecast results not only are more reliable but also can help the decisionmaker with grasping the sulfur dioxide emissions development tendency in general and making proper decision. Results show that the Grey Markov Model has higher accuracy than that of GM(1 , 1) model.

Key words Grey system ; Grey Markov model ; sulfur dioxide

(责任编辑 游中胜)