

文章编号:0253-9993(2008)04-0415-04

GM (1, 1) 模型与线性回归组合方法在矿井 瓦斯涌出量预测中的应用

施式亮, 伍爱友

(湖南科技大学 能源与安全工程学院, 湖南 湘潭 411201)

摘要: 以矿井瓦斯涌出量的预测为主要研究目的, 讨论了 GM (1, 1) 方法适用于单一指数增长模型、对预测序列数据异常情况难以准确预测的局限性, 依据灰色灾变预测原理, 利用线性回归适用短期预测的特点, 提出了基于 GM (1, 1) 与线性回归组合预测矿井瓦斯涌出量的新方法. 应用结果表明: 该方法能很好地解决历史数据的跳变问题, 使预测结果更为可靠、精确.

关键词: GM (1, 1) 模型; 线性回归; 瓦斯涌出量; 预测

中图分类号: TD713.2 **文献标识码:** A

Application of GM (1, 1) and line regression for predicting amount of mine gas emission in coal mine

SHI Shi-liang, WU Ai-you

(School of Energy and Safety Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: Regarding forecasting the amount of mine gas emission in coal mine as the main research purpose, discussed the limitation of GM (1, 1) model in forecasting accidents with simplex index increase model which had some difficult in deal with the abnormity circs of list data. Depending on the grey disaster forecasting theory and using the character of linear regression adapting to the short term forecast, a new combined method based on GM (1, 1) and line regression was come up with. Using the new method, a conclusion was obtained, this method can resolve the problem of the original data change suddenly and make the predicting result more reliability and accuracy.

Key words: GM (1, 1) model; linear regression; amount of mine gas emission; prediction

统计资料表明, 我国煤矿瓦斯事故已占到煤矿生产过程所发生事故的 80% 以上, 造成的伤亡占特大事故伤亡人数的 90%^[1]. 矿井瓦斯涌出量的预测方法可以分为矿山统计法、分源预测法、瓦斯梯度法、煤层瓦斯含量法和瓦斯地质数学模型等^[2-4]. 其中, 矿山统计法与瓦斯含量法以原始瓦斯含量的数理统计为基础, 依据瓦斯涌出量随开采深度变化的统计规律, 外推到预测新区; 分源预测法通过对影响瓦斯涌出量预测精度的各种系数的研究, 找出相应条件下的适合系数值; 瓦斯梯度法用于瓦斯带内, 外推深度不宜超过 100~200 m; 瓦斯地质数学模型以瓦斯地质条件为预测指标, 以数量化理论为基础建立瓦斯涌出量数学模型. 上述的每一种方法都有自己的适用条件, 其预测过程都是静态的, 没有考虑瓦斯涌出量是一个

收稿日期: 2007-06-17 责任编辑: 毕永华

基金项目: 国家自然科学基金资助项目 (50774033, 50274060); 教育部科学技术研究重点资助项目 (206100); 国家安全生产监督管理总局安全生产科技发展计划资助项目 (06-394, 03-103); 湖南省自然科学基金资助项目 (06JJ2069); 湖南省教育厅资助科研项目 (05A013)

作者简介: 施式亮 (1962—), 男, 浙江天台人, 教授, 博士, 博士生导师. Tel: 0732-8290179, E-mail: xtpussl@vip.sina.com

动态的非线性复杂系统^[5], 所以对矿井瓦斯涌出量预测存在着较大的困难.

1 灰色 GM (1, 1) 与线性回归组合模型

1.1 模型的基本特点

灰色预测是根据过去及现在已知或不确定的信息建立一个从过去引申到未来发展趋势的灰色模型 (GM 模型). 它突破了概率统计、回归分析的局限性, 解决了连续微分方程的建模问题, 得到了很好的应用. 然而, 该方法是用指数函数来模拟生成数据的, 要求原始数据服从一定的分布, 只适用于变形呈指数趋势变化情况, 对在趋势线上发生跳变的序列数据出现异常 (如指数偏离过大甚至减小) 情况难以考虑. 线性回归预测法通过分析预测对象的统计资料, 确定一个与预测对象相关的、呈线性关系的因素, 建立两者的一元线性回归方程. 该预测法的缺陷就是它对一些不呈线性变化规律的不能进行分析^[6].

灰色 GM (1, 1) 与线性回归组合模型根据灰色灾变模型与线性回归模型各自的特点, 将两者有机地组合起来, 从历史数据的映射图 ($F \rightarrow Q$ 图) 中找出灾变日期点, 将数据中偏离度比较高的跳变值构成数据序列, 采用 GM (1, 1) 模型分别计算出日期点和相应的函数, 并且预测出下一个或几个跳变日期点和跳变值. 显然, 该方法应用到矿井瓦斯涌出量预测中, 很好地克服了单一的 GM (1, 1) 模型和线性回归的缺陷, 不仅具有高度的概括性, 而且使预测精度高, 具有明显的确定性.

1.2 基本分析步骤

预测时, 如果预测日期点是根据改进灰色模型预测出的跳变日期点, 则它的函数值是通过跳变预测函数计算出的跳变值; 如果预测日期点不是预测出的跳变日期点, 则预测值是通过线性回归曲线计算出的函数值. 其基本的分析流程如图 1 所示, 分析步骤^[7,8]如下:

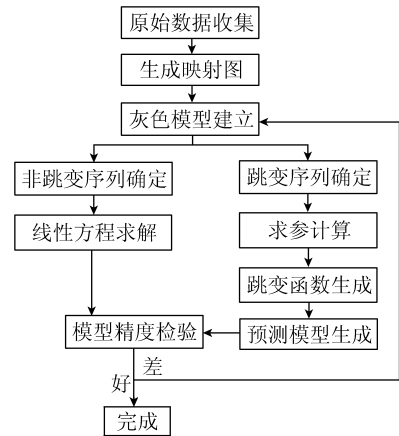


图 1 灰色模型与线性组合方法分析程序

Fig. 1 Analysis process composed gray and linear regression

(1) 作出历史数据映射图 ($F \rightarrow Q$ 图). 设原始数据列 $F_0(X)$ 及原始日期列 $Q_0(X)$ 分别为 $F_0(X) = \{f_0(x_1), f_0(x_2), \dots, f_0(x_n)\}$, $Q_0(X) = \{q_0(x_1), q_0(x_2), \dots, q_0(x_n)\}$. 其中 n 为步长; $f_0(x_n)$, $q_0(x_n)$ 为第 n 个观测点的对应值; 作 $F \rightarrow Q$ 图.

(2) 跳变日期及跳变序列的确定. 以步骤 (1) 为基础, 进行跳变日期及跳变序列的选择, 其选择原则: ① 该原始数据点相对于前后两个连续点显示出非线性特点, 有明显的不在同一直线上的特点; ② 由于跳变点要根据前两个连续点而定, 折线图的第 1 个点不作为跳变点处理. 跳变日期原始序列 $Q_\lambda^0(X)$ 及跳变序列 $F_\lambda^0(X)$ 函数表达式: $Q_\lambda^0(X) = \{q_\lambda^0(x_1), q_\lambda^0(x_2), \dots, q_\lambda^0(x_k)\} \subset Q^0(X)$, $F_\lambda^0(X) = \{f_\lambda^0(x_1), f_\lambda^0(x_2), \dots, f_\lambda^0(x_k)\} \subset F^0(X)$. 其中, $q_\lambda^0(x_k)$, $f_\lambda^0(x_k)$ 为第 k 个跳变序列点.

(3) 建立时序线性回归分析方程. 以步骤 (1), (2) 为基础, 对于两变量间具有线性依存关系的序列 (即非跳变序列: 原始数据点相对于前后两点有明显在同一直线上的特征) 进行线性回归拟合. 其任务是以统计图为基础, 找出描述用变量 x 来推算变量 $f(x)$ 的直线方程, 其表达式: $f(x) = ax + b$, 其中 a 为斜率, 也称回归系数; b 为截距.

(4) 灰色 GM (1, 1) 预测方程的生成. 对原始数据序列与原始时间序列进行累加生成处理, 构成生成序列并建立灰色生成模型, 在此基础上得出原始序列 (包括原始测量数据序列及原始时间序列) 的跳变预测函数模型, 即

$$\hat{f}_\lambda^1(k+1) = (f^0(0) - \lambda_f/\omega_f) e^{-\omega k} + \lambda_f/\omega_f, \quad \hat{q}_\lambda^1(k+1) = (q^0(0) - \lambda_q/\omega_q) e^{-\omega k} + \lambda_q/\omega_q, \quad (1)$$

其中, $f^0(0), q^0(0)$ 分别为原始测量数据序列及原始时间序列; $\omega_f, \omega_q, \lambda_f, \lambda_q$ 为待辨识参数. 通过对待辨识参数的分析计算, 运用跳变预测函数预测下面若干个跳变日期及跳变值.

$$f_A^0(k+1) = f_A^1(k+1) - f_A^1(k), \hat{q}_A^0(k+1) = \hat{q}_A^1(k+1) - \hat{q}_A^1(k). \tag{2}$$

(5) 预测. ① 利用式(2)预测下一次跳变时间 $\hat{q}_A^0(k+1)$, 以此反推下一次预测数据的跳变值 $f_A^0(k+1)$; ② 预测日期不是跳变日期点, 运用线性回归函数($f(x) = ax + b$)进行预测.

(6) 预测精度分析. 模型的拟合残差中往往还有一部分动态有效信息, 可以通过建立残差模型对原模型进行修正^[9]. ① 计算原始序列(s_1)与残差序列(s_2)标准差, 即

$$s_1 = \sqrt{\frac{1}{k-1} \sum_{i=1}^k [f^{(0)}(k) - \bar{f}]^2}, s_2 = \sqrt{\frac{1}{k-1} \sum_{i=1}^k [\beta^{(0)}(k) - \bar{\beta}]^2}, \tag{3}$$

其中, $\bar{f} = \frac{1}{k} \sum_{i=1}^k f^{(0)}(k); \beta^{(0)}(k) = f^{(0)}(k) - \hat{f}_k^{(0)}(k);$
 $\bar{\beta} = \frac{1}{k} \sum_{i=1}^k \beta^{(0)}(k).$ ② 精度检验. 后验差比值 $C =$

s_2/s_1 ; 小误差概率 P 检验 $P = P(|\beta^{(0)}(k) - \bar{\beta}| < 0.674 5s_1)$. 模型预测精度分为4个等级, 见表1. 两个指标必须同时满足, 若有一指标在高等级区间, 另一指标在低等级区间, 其预测精度为低等级, 否则模型需要通过残差修正以提高精度.

表1 模型精度分级标准

Table 1 Grade standard of the model's accuracy

精度	C	P	等级
良好	$C \leq 0.35$	$P \geq 0.95$	1
合格	$0.35 < C \leq 0.50$	$0.95 > P \geq 0.80$	2
勉强	$0.50 < C \leq 0.65$	$0.80 > P \geq 0.70$	3
不合格	$C > 0.65$	$P < 0.70$	4

2 预测模型在瓦斯涌出预测实例中的应用

湖南某矿属于低瓦斯矿井, 瓦斯隐患或瓦斯涌出异常现象常有出现, 特别是掘进工作面进入某些区域的底板砂岩石、褶曲的脊背部位时, 瓦斯涌出量会突然加大, 并有时出现瓦斯嘶叫等动力现象, 因此, 对该矿的瓦斯预测与防治不可忽视. 该矿1994—2006(年份编码分别为1~13)相对瓦斯涌出量分别为1.74, 1.88, 3.79, 1.65, 1.93, 1.84, 3.95, 4.14, 2.05, 4.52, 1.79, 4.63, 1.86 m^3/t .

2.1 跳变序列确定

以该矿13a的相对瓦斯涌出量统计资料为基础, 建立 $f(X) \rightarrow X$ 映射图(图2), 其中 $f(X)$ 为相对瓦斯涌出量; 由于变量“X”是以1a为步长递增的时间序列, 为方便计算, 以1994年为基准点“1”计算. 3, 7, 8, 10, 12为明显的跳变点, 对应1996, 2000, 2001, 2003, 2005年, 则依此可建立原始日期序列 $Q_A^0(X)$ 及原始瓦斯涌出量的跳变序列 $F_A^0(X)$.

$$Q_A^0(X) = \{ Q_A^0(3), Q_A^0(7), Q_A^0(8), Q_A^0(10), Q_A^0(12) \} = \{ 3, 7, 8, 10, 12 \},$$

$$F_A^0(X) = \{ F_A^0(3), F_A^0(7), F_A^0(8), F_A^0(10), F_A^0(12) \} = \{ 3.79, 3.95, 4.14, 4.52, 4.63 \}.$$

2.2 跳变预测模型的求解

运用灰色GM(1, 1)模型计算 Q^0 和 F^0 的跳变预测函数 $\hat{q}^0(k+1)$ 和 $\hat{f}^0(i+1)$. 利用式(1), (2)进行预测模型的求解. 为了方便计算, 减少工作量, 采用Matlab7.0工具箱中矩阵的基本数值计算及特征参数计算法则对特征参数进行了求解, 最终得出该矿瓦斯涌出量的跳变预测模型为

$$\hat{q}_A^0(k+1) = 2.839 2e^{0.312 9k}, \hat{f}_A^1(k+1) = 3.456 4e^{0.024 2k}. \tag{4}$$

2.3 线性预测模型的求解

测点3, 7, 8, 10, 12为明显的跳变点, 而对应的测点1, 2, 4, 5, 6, 9, 11, 13为非跳变点. 对

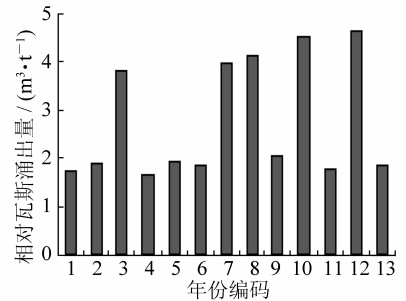


图2 矿井相对瓦斯涌出量

Fig. 2 Representation of gas relative emission

非跳变点建立一元线性回归模型, 通过拟合, 得出线性预测方程表达式为

$$y = 0.0089x + 1.7855. \quad (5)$$

2.4 矿井瓦斯涌出量预测

由以上分析, 可得出组合预测趋势 (图 3).

(1) 原始数据中共有 5 个跳变点, 考察下一次跳变 (即第 6 个跳变点) 情况. 令 $k=5$, 代入式 (4), 得: $\hat{q}_A^0(6) = 2.8392e^{0.3129 \times 5} \approx 14$. 即第 6 个跳变点出现在第 14 年, 对应 $1994 + (14 - 1) = 2007$ 年, 对应图 3 中跳变序列趋势预测 (第 14 年), 可得其相对瓦斯涌出量为 $4.85 \text{ m}^3/\text{t}$. 同理可知, 第 7 跳变点出现在第 19 年, 对应 2012 年的相对瓦斯涌出量为 $5.45 \text{ m}^3/\text{t}$.

(2) 第 15, 16 点不是跳变点, 对应 2008, 2009 年的相对瓦斯涌出量, 综合图 2, 采用线性回归公式 (5) 进行确定, 分别为 $1.93, 1.93 \text{ m}^3/\text{t}$.

2.5 预测结果分析

数据统计到 2006 年, 利用式 (5) 及表 1, 比较其实际值与预测值, 对原始数据与预测数据进行精度分析, 应用式 (3), 通过求标准差及预测精度, 可知 $C = 0.318 < 0.35$, $P = 1 > 0.95$, 预测精度良好 (表 2), 证明了该方法的合理性与精确性.

3 结 语

应用 GM(1, 1) 模型与线性回归耦合的方法, 建立了矿井瓦斯涌出量预测模型. 通过残差检验, 得到其灰色系统预测拟合精度好, 预测结果正确可靠, 能够反映矿井瓦斯涌出量的客观存在与发展态势, 具有实用价值. 用灰色系统进行矿井瓦斯涌出量预测的最大优点是: 数学模型简单易于建立, 要求的原始数据少, 预测精度高. 因此, GM(1, 1) 模型与线性回归组合方法用于矿井瓦斯涌出量预测是可行的, 并且能获得足够的预测精度, 为煤矿瓦斯事故预防和控制、保障煤矿安全生产具有较大的指导意义.

参考文献:

- [1] 施式亮, 宋 译, 何利文, 等. 矿井掘进工作面瓦斯涌出混沌特性判别研究 [J]. 煤炭学报, 2006, 31(6): 58~62.
- [2] 曾 勇, 吴财芳. 矿井瓦斯涌出量预测的模糊形神经网络研究 [J]. 煤炭科学技术, 2004, 32(2): 62~65.
- [3] 刘超儒, 马云东. 矿井延深瓦斯涌出量的灰色预测 [J]. 矿业安全与环保, 2005, 32(3): 1~4.
- [4] 郁 云, 陆金桂. 基于灰色理论和人工神经网络的瓦斯涌出量预测 [J]. 人工智能, 2006, 22(3): 269~272.
- [5] 吕贵春, 马云东. 矿井瓦斯涌出量预测的灰色建模法 [J]. 中国安全科学学报, 2004, 14(10): 22~25.
- [6] 李洪波, 帅 斌. 灰色-线性回归组合模型在预测中的应用 [J]. 陕西工学院学报, 2003, 19(4): 51~55.
- [7] 陈梅英. 回归分析方法在用电量预测中的应用 [J]. 福建水力发电, 2002, 12(1): 60~62.
- [8] 鲍一丹, 吴燕萍, 何 勇. 基于 GM(1, 1) 模型和线性回归的组合预测新方法 [J]. 系统工程理论与实践, 2004, 3(3): 95~98.
- [9] 伍爱友, 肖红飞, 王从陆, 等. 煤与瓦斯突出控制因素加权灰色关联模型的建立与应用 [J]. 煤炭学报, 2005, 30(1): 58~62.

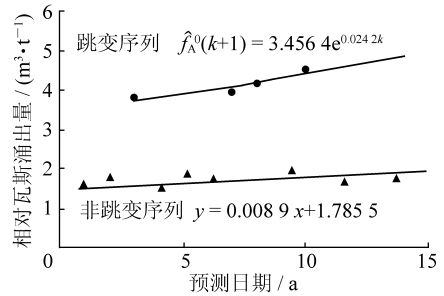


图 3 GM(1, 1) 与线性回归组合趋势预测

Fig. 3 Forecasting trends based on linear regression & GM(1, 1)

表 2 预测精度分析

Table 2 Analysis of the prediction precision

年份	实际值	预测值	$f^0(k) - \hat{f}^0(k)$
1994	1.74	1.79	-0.05
1995	1.88	1.80	0.08
1996	3.79	3.72	0.07
1997	1.65	1.82	-0.17
1998	1.93	1.83	0.10
1999	1.84	1.84	0.00
2000	3.95	4.09	-0.14
2001	4.14	4.19	-0.05
2002	2.05	1.87	0.22
2003	4.52	4.40	0.12
2004	1.79	1.88	0.11
2005	4.63	4.62	0.01
2006	1.86	1.90	-0.04
2007		4.85	

注: 由预测精度分析可知, $s_1 = 0.842$, $s_2 = 0.268$,

$$C = s_2/s_1 = 0.318, P = 1 > 0.95.$$