

参数由递推辨识的新型GM(1, 2)预测模型

雷鸣雳[†], 冯祖仁

(西安交通大学 系统工程研究所, 陕西 西安 710049; 机械制造系统工程国家重点实验室, 陕西 西安 710049)

摘要: 为改善模型预测性能, 提出一种GM(1, 2)预测新模型. 根据模型定义式直接推导获得模型预测值递推表达式, 应用粒子群算法对递推表达式参数进行辨识. 典型算例表明, 新模型收敛速度快, 较普通及文献中改进GM(1, 2)模型具有更高的预测精度.

关键词: 灰色预测; GM(1, 2)模型; 参数辨识; 粒子群算法

中图分类号: N94 **文献标识码:** A

A novel GM(1, 2) forecasting model with parameters identified recursively

LEI Ming-li[†], FENG Zu-ren

(Systems Engineering Institute, Xi'an Jiaotong University, Xi'an Shaanxi 710049, China;
State Key Laboratory for Manufacturing Systems Engineering, Xi'an Shaanxi 710049, China)

Abstract: To improve the prediction performance, we propose a novel GM(1,2) model for prediction. The recursive prediction equations are derived directly from the definition of the model. The parameters of prediction equations are identified by using the particle swarm optimization algorithm (PSO). Typical numerical examples are given to demonstrate that the novel GM(1,2) model provides faster convergence rate and higher prediction precision than conventional GM(1,2) models and other improved GM(1,2) models mentioned in references.

Key words: grey prediction; GM(1,2) model; parameter identification; particle swarm optimization (PSO)

1 引言(Introduction)

普通GM(1, 2)模型在输入端引入参考序列, 充分反映外界扰动变化情况, 增强了模型的预测能力, 适用波动变化序列建模预测. 通常根据模型定义式由最小二乘法进行参数辨识, 将辨识结果代入由模型影子方程解得的预测值白化响应表达式中进行预测值计算, 模型参数辨识及预测值计算分别由定义式及影子方程两个不同的模型式完成. 然而, 影子方程并不是由定义式直接推导获得, 只是一种借用和白化默认^[1], 如此建立的模型其预测性能必然降低. 已有研究者提出GM(1, 2)模型改进方法并取得较好效果, 改进方法主要分为两类: 一是引入其他优化算法基于定义式和白化响应式进行参数辨识及预测值计算^[2-4], 依然将参数辨识及预测值计算分割开来; 二是应用其他优化算法基于白化响应式进行参数辨识及预测值计算^[5], 虽然已将参数辨识与预测值计算统一于白化响应式, 但白化响应式的解只是近似解, 并非模型定义式的真实解.

本文由GM(1, 2)模型定义式直接推导建立模型

预测值递推表达式, 应用粒子群算法对递推式参数进行辨识, 建立了一种GM(1, 2)预测新模型, 并与普通及其他改进GM(1, 2)模型进行比较.

2 GM(1, 2)模型的建模思路(Modeling idea of GM(1, 2))

普通GM(1, 2)模型灰色微分方程即定义式为

$$y^{(0)}(k) + az^{(1)}(k) = br^{(1)}(k). \quad (1)$$

与式(1)对应的白化微分方程即影子方程为

$$\frac{dy^{(1)}}{dt} + ay^{(1)} = br^{(1)}, \quad (2)$$

式中: $y^{(0)}$ 为原始主序列; $y^{(1)}$ 为主序列的一次累加序列; $r^{(1)}$ 为原始参考序列 $r^{(0)}$ 的一次累加序列; a 称为发展系数; b 称为灰作用量; $z^{(1)}(k)$ 称为模型背景值.

由式(2)解得模型预测值白化响应式为

$$\hat{y}^{(1)}(k+1) = (y^{(0)}(1) - \frac{b}{a}r^{(1)}(k+1))e^{-ak} + \frac{b}{a}r^{(1)}(k+1). \quad (3)$$

通常,根据模型定义式(1),由最小二乘法对参数 a, b 进行估算,将估算结果代入白化响应式(3),就可获得原始主序列的拟合值及预测值.

3 GM(1,2)模型的缺陷(Defects of GM(1,2))

GM(1,2)模型参数由定义式通过最小二乘法辨识确定,预测值根据影子方程求得的白化响应式计算.可见,模型参数辨识和预测值计算分别由两个不同的模型式获得.然而,定义式为差分方程,影子方程为微分方程,两个方程之间不存在严格的等同关系,只是一种近似的替代,因此,将由定义式辨识获得的参数 a, b 估算值代入白化响应式进行预测值计算,结果只能是近似解.模型定义式及影子方程的具体关系推导如下.

推导:对模型影子方程(2)两端同时在区间 $[k-1, k]$ 上进行积分,可得

$$\int_{k-1}^k \dot{y}^{(1)}(\tau)dt + \int_{k-1}^k ay^{(1)}(\tau)dt = \int_{k-1}^k br^{(1)}(\tau)dt. \quad (4)$$

化简可得

$$y^{(1)}(k) - y^{(1)}(k-1) + a \int_{k-1}^k y^{(1)}(\tau)dt = b \int_{k-1}^k r^{(1)}(\tau)dt. \quad (5)$$

由一次累加变换定义可知

$$y^{(0)}(k) = y^{(1)}(k) - y^{(1)}(k-1). \quad (6)$$

将式(6)代入式(5)可得

$$y^{(0)}(k) + a \int_{k-1}^k y^{(1)}(\tau)dt = b \int_{k-1}^k r^{(1)}(\tau)dt. \quad (7)$$

上式中, $\int_{k-1}^k y^{(1)}(\tau)dt$ 和 $\int_{k-1}^k r^{(1)}(\tau)dt$ 分别代表图1(a)及图1(c)中所示曲顶梯形的面积.用图1(b)中梯形近似替代图1(a)中曲顶梯形,用图1(d)中矩形近似替代图1(c)中曲顶梯形,进行计算可得

$$\int_{k-1}^k y^{(1)}(\tau)dt \approx gy^{(1)}(k-1) + (1-g)y^{(1)}(k), \quad (8)$$

$$\int_{k-1}^k r^{(1)}(\tau)dt \approx r^{(1)}(k). \quad (9)$$

令

$$z^{(1)}(k) = gy^{(1)}(k-1) + (1-g)y^{(1)}(k), \quad (10)$$

式中 $g \in [0, 1]$ 称为模型背景值权重系数.

将式(9)–(10)代入式(7)可得模型的定义式

$$y^{(0)}(k) + az^{(1)}(k) = br^{(1)}(k). \quad (11)$$

以上推导说明,模型定义式与影子方程不存在严格的等同关系,影子方程只是定义式的近似替代,因而,使得定义式解最优的参数 a, b 辨识值未必使得影子方程取得最优解.

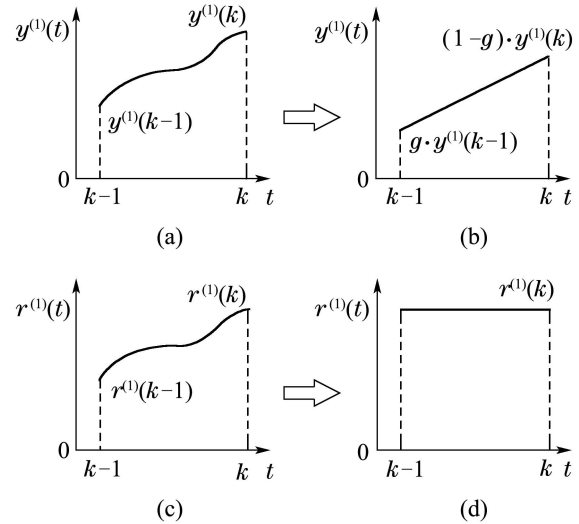


图1 GM(1,2)模型背景值示意图

Fig. 1 Background value diagram of GM(1,2)

4 GM(1,2)预测新模型(Novel GM(1,2) forecasting model)

应用粒子群算法对模型预测值递推表达式参数进行辨识,将辨识结果代入递推表达式进行拟合值及预测值计算,建立GM(1,2)预测新模型.

4.1 模型预测值递推表达式(Model recurrence expression)

将背景值表达式(1)代入模型定义式(2)可得

$$y^{(0)}(k) + a(gy^{(1)}(k-1) + (1-g)y^{(1)}(k)) = br^{(1)}(k). \quad (12)$$

由一次累加变换定义可得

$$y^{(1)}(k) = y^{(1)}(k-1) + y^{(0)}(k). \quad (13)$$

将式(13)代入式(12)并进行整理可得

$$(1 + a - ag)y^{(0)}(k) + ay^{(1)}(k-1) = br^{(1)}(k), \quad (14)$$

整理可得

$$y^{(0)}(k) = \frac{br^{(1)}(k) - ay^{(1)}(k-1)}{1 + a - ag}. \quad (15)$$

由式(15)可得

$$y^{(0)}(k-1) = \frac{br^{(1)}(k-1) - ay^{(1)}(k-2)}{1 + a - ag}. \quad (16)$$

将式(15)及式(16)等式左右两边相减并整理可得

$$y^{(0)}(k) - y^{(0)}(k-1) = \frac{br^{(0)}(k) - ay^{(0)}(k-1)}{1 + a - ag}. \quad (17)$$

进一步化简得

$$y^{(0)}(k) = \frac{br^{(0)}(k)}{1 + a - ag} + \frac{(1 - ag)y^{(0)}(k-1)}{1 + a - ag}. \quad (18)$$

令

$$\lambda_1 = y^{(0)}(1), \lambda_2 = \frac{b}{1 + a - ag}, \lambda_3 = \frac{1 - ag}{1 + a - ag}, \quad (19)$$

可得模型预测值递推表达式为

$$\begin{cases} \hat{y}^{(0)}(1) = \lambda_1, \\ \hat{y}^{(0)}(k) = \lambda_2 r^{(0)}(k) + \lambda_3 \hat{y}^{(0)}(k-1), \end{cases} \quad (20)$$

式中: $\hat{y}^{(0)}(k)$ 代表原始主序列 k 时刻预测值, $\lambda_1, \lambda_2, \lambda_3$ 为待辨识参数.

4.2 递推表达式参数辨识(Recurrence expression parameters identification)

采用粒子群算法^[6]对模型预测值递推表达式参数 $\lambda_1, \lambda_2, \lambda_3$ 进行辨识, 具体过程描述如下:

Step 1 初始化粒子群.

在 $D = 3$ 维空间中随机产生 $m = 50$ 个粒子, 初始化粒子位置和速度. 第 k 次迭代中第 i 个粒子位置向量 $\vec{x}_i^k = \{x_{i1}^k, x_{i2}^k, x_{i3}^k\}$, 速度向量 $\vec{v}_i^k = \{v_{i1}^k, v_{i2}^k, v_{i3}^k\}$, 粒子空间坐标分别对应待辨识参数的某组潜在解.

Step 2 计算粒子适应度.

第 k 次迭代中第 i 个粒子的适应度定义为

$$f_i^k(a, b) = \frac{1}{1 + e_{MAPE}}, \quad (21)$$

式中: e_{MAPE} 代表预测平均绝对百分比误差:

$$e_{MAPE} = \frac{1}{N} \sum_{j=1}^N e_{APE}(j), \quad (22)$$

$$e_{APE}(j) = \frac{|y^{(0)}(j) - \hat{y}^{(0)}(j)|}{y^{(0)}(j)} \times 100\%, \quad (23)$$

式中 $y^{(0)}(j), \hat{y}^{(0)}(j)$ 和 $e_{APE}(j)$ 分别代表原始序列 j 时刻真实值、预测值及预测绝对百分比误差.

Step 3 更新粒子速度及位置.

$$\begin{cases} v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_i - x_{id}^k) + c_2 r_2 (p_g - x_{id}^k), \\ x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1}, \end{cases} \quad (24)$$

式中: v_{id}^k 和 x_{id}^k 分别代表当前粒子速度及位置; v_{id}^{k+1} 和 x_{id}^{k+1} 分别代表更新后粒子速度及位置; c_1 和 c_2 设为2; r_1 和 r_2 为 $[0, 1]$ 内变化的随机数, ω 为惯性权重系数; p_i 和 p_g 分别为粒子个体及种群最优位置.

Step 4 更新个体最优位置及种群最优位置.

$$\begin{cases} \text{if } f_i^k(a, b) > p_i, \text{ then } p_i = f_i^k(a, b), \\ \text{if } f_i^k(a, b) > p_g, \text{ then } p_g = f_i^k(a, b). \end{cases} \quad (25)$$

Step 5 判定是否停止算法. 若迭代次数达到预设的最大代数Gmax(文中取Gmax = 200), 转向Step 6; 否则, 转向Step 2.

Step 6 输出最优辨识参数, 算法结束.

5 实例分析(Example analysis)

例 1 降低数据序列^[3]

苧麻纤维投影宽度(μm)数据序列 $r = \{39.12, 36.17, 34.85, 33.57, 25.67, 27.76\}$, Tex数据序列 $y =$

$\{0.813, 0.690, 0.645, 0.617, 0.566, 0.518\}$.

以投影宽度数据为参考序列, Tex数据为主序列, 建立GM(1, 2)苧麻纤维Tex数据预测模型. 序列中前4个数据用于模型拟合训练, 对后两个数据进行预测, 结果如表1所示, 新模型粒子适应度进化曲线如图2所示.

表 1 例1预测及拟合结果

Table 1 Results for the example 1

| 原始数据 | 普通GM(1, 2)模型 | | 本文模型 | |
|-----------------|--------------|----------------|-------|----------------|
| | 拟合值 | $e_{APE} / \%$ | 拟合值 | $e_{APE} / \%$ |
| 0.813 | 0.813 | 0.00 | 0.813 | 0.00 |
| 0.690 | 0.554 | 19.71 | 0.733 | 6.23 |
| 0.645 | 0.678 | 5.16 | 0.687 | 6.45 |
| 0.617 | 0.628 | 1.85 | 0.654 | 5.96 |
| $e_{MAPE} / \%$ | 6.68 | | 4.66 | |

| 原始数据 | 预测值 | $e_{APE} / \%$ | 预测值 | $e_{APE} / \%$ |
|-----------------|-------|----------------|-------|----------------|
| 0.566 | 0.477 | 15.65 | 0.550 | 2.85 |
| 0.518 | 0.517 | 0.28 | 0.534 | 3.02 |
| $e_{MAPE} / \%$ | 13.36 | | 2.93 | |

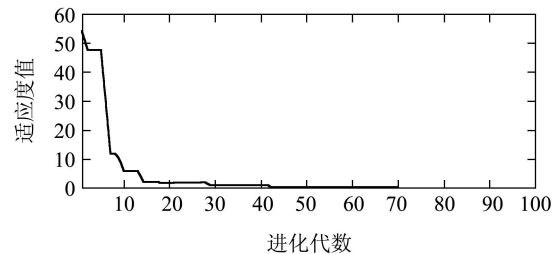


图 2 例1粒子适应度进化曲线

Fig. 2 The particle fitness evolution curve of example 1

例 2 低增长数据序列^[7]

江苏省1996年-2003年内河货运量(万吨)数据序列为 $y = \{35258, 34256, 33800, 33256, 36695, 35012, 37306, 39600\}$, 同期GDP(亿元)数据序列为

$$r = \{30, 35, 55, 60, 70, 90, 110, 140, 260, 320\},$$

以 y 为主序列, r 为参考序列, 建立GM(1, 2)内河货运量预测模型. 1996年-1999年数据用于模型训练, 对2000年-2003年内河货运量进行预测. 结果如表2所示, 粒子适应度进化曲线如图3所示.

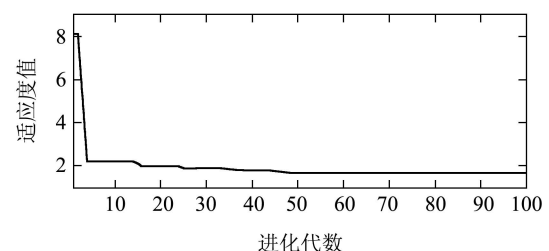


图 3 例2粒子适应度进化曲线

Fig. 3 The particle fitness evolution curve of example 2

表2 例2拟合及预测结果

Table 2 Results for the example 2

| 年份 | 原始数据 | 普通GM(1,2)模型 | | 本文模型 | |
|---------------|-------|-------------|--------------|-------|--------------|
| | | 拟合值 | $e_{APE}/\%$ | 拟合值 | $e_{APE}/\%$ |
| 1996 | 32528 | 32528 | 0.00 | 32528 | 0.00 |
| 1997 | 34256 | 23544 | 31.27 | 32769 | 4.34 |
| 1998 | 33800 | 32740 | 3.14 | 33011 | 2.33 |
| 1999 | 33256 | 35976 | 8.18 | 33256 | 0.00 |
| $e_{MAPE}/\%$ | | 10.65 | | 1.67 | |

| 年份 | 原始数据 | 预测值 | $e_{APE}/\%$ | 预测值 | $e_{APE}/\%$ |
|---------------|-------|-------|--------------|-------|--------------|
| 2000 | 32528 | 40682 | 10.87 | 33502 | 8.70 |
| 2001 | 34256 | 45593 | 30.22 | 33750 | 3.61 |
| 2002 | 33800 | 51778 | 38.79 | 33999 | 8.86 |
| 2003 | 33256 | 60807 | 53.55 | 34251 | 13.51 |
| $e_{MAPE}/\%$ | | 33.36 | | 8.67 | |

例3 高增长数据序列

原始主数据序列及参考数据序列分别为

$$y = \{20, 22, 40, 45, 60, 80, 100, 120, 200, 240\},$$

$$r = \{30, 35, 55, 60, 70, 90, 110, 140, 260, 320\}.$$

在此, 建立GM(1,2)模型对主序列数据进行预测. 数据序列中前6个数据用于模型拟合训练, 对主序列后4个数据进行预测. 新模型粒子适应度进化曲线如图4所示, 结果如表3所示.

表3 例3拟合及预测结果

Table 3 Results for the example 3

| 原始数据 | 普通GM(1,2)模型 | | 本文模型 | | |
|---------------|-------------|--------------|-------|--------------|--|
| | 拟合值 | $e_{APE}/\%$ | 拟合值 | $e_{APE}/\%$ | |
| 20 | 20.00 | 0.00 | 20.00 | 0.00 | |
| 22 | 20.79 | 5.51 | 27.06 | 23.02 | |
| 40 | 51.74 | 29.35 | 39.94 | 0.16 | |
| 45 | 69.49 | 54.43 | 49.76 | 10.57 | |
| 60 | 85.03 | 41.72 | 59.93 | 0.12 | |
| 80 | 108.29 | 35.36 | 74.65 | 6.69 | |
| $e_{MAPE}/\%$ | | 27.73 | | 6.76 | |

| 原始数据 | 预测值 | $e_{APE}/\%$ | 预测值 | $e_{APE}/\%$ | |
|---------------|--------|--------------|--------|--------------|--|
| 100 | 131.24 | 31.24 | 92.07 | 7.93 | |
| 120 | 164.79 | 37.33 | 115.43 | 3.80 | |
| 200 | 299.08 | 49.54 | 181.36 | 9.32 | |
| 240 | 368.18 | 53.41 | 246.56 | 2.73 | |
| $e_{MAPE}/\%$ | | 42.88 | | 5.95 | |

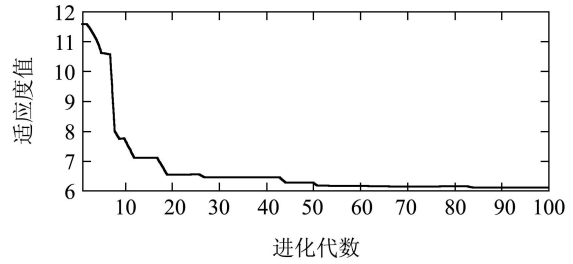


图4 例3粒子适应度进化曲线

Fig. 4 The particle fitness evolution curve of example 3

例4 波动变化数据序列^[8]

PJM电力市场2007年5月1日-12日5点及4点, 电价(\$/MWh)原始数据序列分别为

$$y = \{35.55, 25.81, 31.47, 26.48, 43.69, 31.24, 59.20, 82.42, 53.69, 64.32, 34.44, 38.38\},$$

$$r = \{35.57, 26.70, 31.43, 30.59, 53.69, 43.31, 67.41, 80.42, 50.11, 58.72, 31.41, 27.50\}.$$

以 y 为主序列, r 为参考序列建立GM(1,2)电价预测模型, 5月1日-8日电价数据用于模型拟合训练, 对5月9日-12日5点电价进行预测. 结果如表4所示, 新模型粒子适应度进化曲线如图5所示.

预测实例结果显示, 新建GM(1,2)模型拟合及预测性能均优于普通模型; 降低数据序列新模型拟合及预测 e_{MAPE} 较普通模型分别降低2个和10个百分点; 低增长数据序列分别降低9个和25个百分点; 高增长数据序列分别降低21个和37个百分点; 波动变化数据序列分别降低60个和148个百分点. 另外, 粒子适应度进化曲线说明新模型具有较快收敛速度, 在50代内粒子适应度收敛于常数.

仍以上述4个数据序列为例, 将新模型与文献[2, 5]中改进GM(1,2)模型进行比较. 文献[2]应用粒子群算法对模型背景值进行优化; 文献[5]采用粒子群算法对模型参数 a, b 进行辨识. 记文献[2]改进模型为模型1, 文献[5]改进模型为模型2, 比较结果如表5所示. 结果显示, 本文新建模型较文献中的两种改进模型拟合精度及预测精度均有所改善.

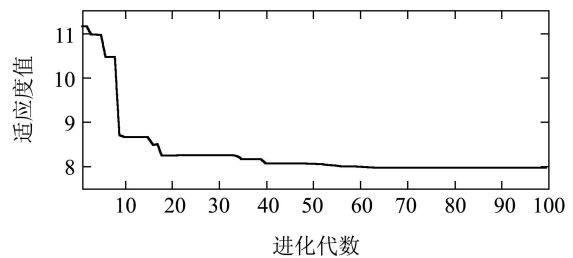


图5 例4粒子适应度进化曲线

Fig. 5 The particle fitness evolution curve of example 4

表 4 例4拟合及预测结果
Table 4 Results of example 4

| 原始数据 | 普通GM(1, 2)模型 | | 本文模型 | |
|---------------|--------------|--------------|-------|--------------|
| | 拟合值 | $e_{APE}/\%$ | 拟合值 | $e_{APE}/\%$ |
| 35.55 | 35.55 | 0.00 | 35.55 | 0.00 |
| 25.81 | 14.00 | 45.74 | 25.81 | 0.01 |
| 31.47 | 29.43 | 6.49 | 27.41 | 12.89 |
| 26.48 | 41.90 | 58.24 | 27.11 | 2.37 |
| 43.69 | 75.05 | 71.78 | 43.69 | 0.00 |
| 31.24 | 80.72 | 158.37 | 39.28 | 25.73 |
| 59.20 | 124.03 | 109.50 | 55.83 | 5.70 |
| 82.42 | 160.05 | 94.19 | 68.26 | 17.18 |
| $e_{MAPE}/\%$ | 68.04 | | 7.98 | |

| 原始数据 | 预测值 | $e_{APE}/\%$ | 预测值 | $e_{APE}/\%$ |
|---------------|--------|--------------|-------|--------------|
| 53.69 | 129.03 | 140.34 | 48.72 | 9.26 |
| 64.32 | 151.42 | 135.42 | 51.31 | 20.23 |
| 34.44 | 109.18 | 217.01 | 32.11 | 6.75 |
| 38.38 | 102.77 | 167.75 | 25.75 | 32.91 |
| $e_{MAPE}/\%$ | 165.13 | | 17.29 | |

表 5 新模型与其他改进模型预测结果比较
Table 5 Results of the novel model compared with other improved models

| 实例 | 拟合误差 $e_{MAPE}(\%)$ | | | 预测误差 $e_{MAPE}(\%)$ | | |
|----|---------------------|-------|------|---------------------|-------|-------|
| | 模型1 | 模型2 | 本文模型 | 模型1 | 模型2 | 本文模型 |
| 例1 | 4.13 | 4.12 | 4.66 | 7.92 | 8.28 | 2.93 |
| 例2 | 9.43 | 13.05 | 1.67 | 37.61 | 31.19 | 8.67 |
| 例3 | 23.48 | 9.37 | 6.76 | 38.98 | 11.39 | 5.95 |
| 例4 | 14.20 | 10.45 | 7.98 | 23.88 | 23.71 | 17.29 |

6 结论(Conclusions)

为改善普通GM(1, 2)模型预测性能, 本文提出一种GM(1, 2)预测新模型. 通过理论分析及实例验证, 形成结论如下:

1) GM(1, 2)模型影子方程与模型定义式不存在严格的等同关系, 影子方程的解只是模型定义式的近似解;

2) 直接根据模型定义式推导获得的预测值递推表达式较由影子方程获得的预测值白化响应表达式更适合于数据序列预测求解;

3) 基于预测值递推表达式参数优化的GM(1, 2)新模型具有更好的预测性能.

参考文献(References):

- [1] 邓聚龙. 灰预测与灰决策 [M]. 武汉: 华中科技大学出版社, 2002. (DENG Julong. *Grey Forecasting and Grey Decision* [M]. Wuhan: Huazhong University of Science and Technology press, 2002.)
- [2] LEI M L, FENG Z R. Modeling electricity price forecast with grey and correlation method in competitive markets [C] // *Asia-Pacific Power and Energy Engineering Conference*. Piscataway, NJ: IEEE, 2010: 10.1109/APPEEC.2010.5448701 .
- [3] 李晓峰, 罗佑新. 苎麻纤维细度测试的灰色优化GM(1, 2)模型与误差分析 [J]. 应用科学学报, 2003, 21(1): 25 - 29. (LI Xiaofeng, LUO Youxin. The grey optimum model for the test of the fineness of ramie fiber and its error analysis [J]. *Journal of Applied Sciences*, 2003, 21(1): 25 - 29.)
- [4] 王瑞庆, 李渝曾. 含误差校正的粒子群优化GM(1, 2)短期电价预测方法 [J]. 电力系统保护与控制, 2011, 39(13): 41 - 45. (WANG Ruiqing, LI Yuzeng. Particle swarm optimization based GM(1, 2) method for short-term electricity price forecasting with predicted error improvement [J]. *Power System Protection and Control*, 2011, 39(13): 41 - 45.)
- [5] LEI M L, FENG Z R. A novel grey model to short-term electricity price forecasting for NordPool power market [C] // *Proceedings of the International Conference on Systems, Man, and Cybernetics*. San Antonio: IEEE, 2009: 4347 - 4352.
- [6] KENNEDY J, EBERHART R C. Particle swarm optimization [C] // *Proceedings of the International Conference on Neural Networks*. Piscataway: IEEE, 1995: 1942 - 1948.
- [7] 缪岩. GM(1, 2)模型在内河货运量预测中的应用 [J]. 现代交通技术, 2008, 5(1): 84 - 86. (MIAO Yan. Application of GM(1, 2) in the forecast of transport volume of inland river [J]. *Modern Transportation Technology*, 2008, 5(1): 84 - 86.)
- [8] <http://www.PJM.com>.

作者简介:

雷鸣雳 (1971-), 男, 博士研究生, 主要从事系统优化及预测方面的研究工作, E-mail: mllei768@126.com;

冯祖仁 (1953-), 男, 教授, 博士生导师, 主要从事机器人控制及视觉导航方面的研究工作, E-mail: fzr9910@mail.xjtu.edu.cn.