

基于灰色马尔科夫模型的煤炭产量预测

宋晓震¹, 施式亮^{1,2*}, 曹建¹

(1. 湖南科技大学 资源环境与安全工程学院, 湖南 湘潭 411201;
2. 湖南科技大学 煤矿安全开采技术湖南省重点实验室, 湖南 湘潭 411201)

摘要: 煤炭资源是我国经济建设中的重要能源, 准确预测我国煤炭产量有助于国家能源政策的制定. 为此, 在单一的灰色 GM(1,1) 预测模型的基础上, 与马尔科夫预测模型组合形成灰色 GM(1,1)-马尔科夫组合预测模型, 以相对误差、均方差比值和小概率误差 3 个指标对模型的精度做检验. 结果表明, 组合预测模型优于单一预测模型, 组合预测模型不仅反映了煤炭产量短中期呈上升趋势, 又突显了该模型能较优地处理波动性序列的优点, 能够有效地预测 2019 年和 2020 年煤炭产量.

关键词: 煤炭产量; 预测; 精度; 灰色马尔科夫模型

中图分类号: TD821 **文献标志码:** A **文章编号:** 1672-9102(2019)02-0029-06

Study on Coal Production Prediction Based on Gray Markov Model

Song Xiaozhen¹, Shi Shiliang^{1,2}, Cao Jian¹

(1. School of Resources, Environment and Safety Engineering, Hunan University of Science and Technology, Xiangtan 411201, China;
2. Hunan Provincial Key Lab on Prevention and Control of Gas and Roof Disasters for Southern Coal Mines,
Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: Coal resource is an important energy source in China's economic construction. Accurate prediction of China's coal output is beneficial to the formulation of national energy policy. Based on the single grey GM(1,1) prediction model, it is combined with Markov Prediction Model to form Grey GM(1,1)-Markov combination prediction model. The relative error mean square ratio and small probability error are used to test the accuracy of the three index models. Results show that the combined forecasting model is superior to the single forecasting model. The combination prediction model not only reflects the rising trend of coal production in the short and medium term, but also highlights the advantage of the model in dealing with the fluctuation series. The predicted coal production for the period 2019~2020 is 38.97 and 39.16 billion tons.

Keywords: coal production; prediction; accuracy; Grey Markov Model

我国是煤炭生产大国, 煤炭产量关乎我国国家发展战略^[1]. 21 世纪以来, GDP 的快速增长其中很大一部分要依赖于煤炭产量的增长^[2]. 煤作为第一能源供给的现状, 较短时间内, 难以改变.

在煤炭开采行业中, 准确地预测我国煤炭产量有利于国家合理制定能源政策, 提高生产效率, 有益于我国经济发展和社会稳定^[3]. 煤炭产量预测方法通常采用弹性系数预测、概率统计、灰色预测等模型^[4-7]. 但是, 弹性系数预测需要掌握难以获得的变量之间的变动比率关系, 概率统计方法需要大量数据, 在实际中往往难以获得足够的支撑, 而灰色系统理论虽然能够利用少量已得到的数据快速建模, 但该模型对于波动性较大的数据无法进行较为精确的预测. 基于此构建灰色马尔科夫模型, 该模型兼具灰色预测

收稿日期: 2019-04-25

基金项目: 国家自然科学基金资助项目(51774135); 湖南省 2017 年安全生产专项资金资助项目(湘财企指[2017]20 号)

* 通信作者, E-mail: 99850399@qq.com

模型所需样本量小、精度高可进行研究对象的短中期预测和马尔科夫模型能较精确拟合波动性强的随机序列的双重优势,适用于预测具有明显灰度特征的煤炭产量^[8,9].灰色马尔科夫组合预测模型自1992年提出以来,广泛应用于农业、交通等领域^[10-12],而运用该模型进行煤炭产量预测的研究相对较少.因此笔者以我国2010~2018年度的煤炭产量为研究对象,利用灰色马尔科夫组合预测模型对我国2019~2020年的煤炭生产总量进行预测及研究,以应对近年来我国煤炭产量随国家的政策变动已有较大变化的情况,并为企业提升生产效率与产能、国家发展战略的制定提供借鉴与参考.

1 预测模型的构建

1.1 灰色GM(1,1)模型的构建

由预测对象提供的信息生成原始序列: $X^{(0)}(k) = \{x(1), x(2), \dots, x(n)\}$; 其中 $X^{(0)}(k) \geq 0, k = 1, 2, \dots, n$; $X^{(1)}$ 为 $X^{(0)}$ 的一次累加生成序列: $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$; 其中 $X^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$; 紧邻均值生成序列 $Z(k) = \mu X^{(1)}(k) + (1 - \mu)X^{(1)}(k - 1), k = 2, 3, \dots, n, \mu$ 取0.5. 由一次累加生成序列 $X^{(1)}(k)$, 构建线性微分方程: $(\frac{dX^{(1)}(k)}{dt}) + aX^{(1)}(k) = b$, 求其特殊解为

$$X^{(1)}(k) = [X^{(0)}(1) - \frac{b}{a}]e^{-a(k-1)} + \frac{b}{a}, k = 2, 3, \dots, n. \quad (1)$$

式中: a 和 b 为待定系数,本研究采用最小二乘法进行参数向量 $\hat{a} = [a, b]^T$ 估测,计算得到 a 和 b 的待定系数表达式如下:

$$D = (n - 1) \sum_{k=2}^n [Z(k)]^2 - [\sum_{k=2}^n Z(k)]^2; \quad (2)$$

$$a = \frac{\{(1 - n) \sum_{k=2}^n x^{(0)}(k)Z(k) + \sum_{k=2}^n x^{(0)}(k) \sum_{k=2}^n Z(k)\}}{D}; \quad (3)$$

$$b = \frac{\{\sum_{k=2}^n x^{(0)}(k) \sum_{k=2}^n [Z(k)]^2 - \sum_{k=2}^n x^{(0)}(k)Z(k) \sum_{k=2}^n Z(k)\}}{D}. \quad (4)$$

1.2 灰色马尔科夫组合预测模型

由预测对象提供的信息组成原始序列 $X^{(0)}(k)$, 设 t_0 时的状态已知,随机过程 $t > t_0$ 时的状态,仅与 t_0 时的状态有关,与 t_0 之前的状态无关,称该随机过程为马尔科夫过程,即马尔科夫的无后效性.通常情况下,将时间和状态均离散的马尔科夫过程称为马尔科夫链.以其能够有效处理波动特征的随机序列为基础,与灰色GM(1,1)预测模型组合形成灰色GM(1,1)-马尔科夫组合预测模型.

1.2.1 划分状态

由原始系列中的实际值与单一的灰色GM(1,1)预测模型所得预测值相比得其比值: $M = \frac{x^{(0)}(k)}{\hat{x}^{(0)}(k)}$; 由

步长 $S = \frac{M_{\max} - M_{\min}}{i}$, 合理划分状态区间: $E_i = [M_{i1}, M_{i2}]$, $i = 1, 2, \dots, n$, 其中 M_{\max}, M_{\min}, i 分别表示相对最大值、相对最小值、状态区间个数; M_{i1}, M_{i2} 分别为比值的上限和下限.

1.2.2 状态转移概率矩阵的计算

上述划分的状态区间,与灰色GM(1,1)预测模型计算所得的相对值一一对应,通过公式 $P_{ij}^{(k)} = \frac{n_{ij}(k)}{n_i}$, 其中 $n_{ij}(k)$ 表示研究对象由 E_i 状态通过 k 步转移到 E_j 状态时出现的次数; n_i 表示研究对象处于 E_i 状态的总数.由此得到状态转移概率矩阵如下:

$$P^{(k)} = \begin{bmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \cdots \\ P_{21}^{(k)} & P_{22}^{(k)} & \cdots \\ \cdots & \cdots & \cdots \\ P_{m1}^{(k)} & P_{m2}^{(k)} & \cdots \end{bmatrix}.$$

1.2.3 灰色马尔科夫预测值的修正

以马尔科夫转移概率矩阵预测灰色 GM(1,1) 模型的相对值所在的区间 $[M_{i1}, M_{i2}]$, 以 $[M_{i1}, M_{i2}]$ 中值作为修正值, 计算组合模型的预测值 $\hat{X}^{(0)}(k)$, 计算式:

$$\hat{X}^{(0)}(k) = 0.5(M_{i1} + M_{i2})X^{(0)}(k). \quad (5)$$

1.3 精度检验

$$1) \text{ 平均相对误差 } \beta: \beta = \left(\frac{1}{n}\right) \sum_{k=1}^n \left[\frac{|\varepsilon(k)|}{x^{(0)}(k)}\right] \times 100\%;$$

$$2) \text{ 均方差比值 } B: B = \frac{S_2}{S_1};$$

$$3) \text{ 小概率误差 } R: R = \{|\varepsilon(k) - \bar{\varepsilon}(k)|\} < 0.6745S_1;$$

式中: $\varepsilon(k)$ 代表残差序列; $\bar{\varepsilon}(k)$ 代表残差序列平均值; S_1 代表原始序列的标准差; S_2 代表残差序列的标准差.

根据精度不同, 灰色马尔科夫组合预测模型一般分为 4 个等级, 详见表 1.

表 1 检验模型精度等级标准

| 模型精度等级 | 平均相对误差 β | 均方差比值 B | 小误差概率 R |
|---------|----------------|----------------------|-----------|
| 一级(优) | 0.01 | $B \leq 0.35$ | 0.95 |
| 二级(良) | 0.05 | $0.35 < B \leq 0.50$ | 0.80 |
| 三级(合格) | 0.10 | $0.50 < B \leq 0.65$ | 0.70 |
| 四级(不合格) | 0.20 | $0.65 < B \leq 0.80$ | 0.60 |

2 实证计算与分析

2.1 数据处理

由全国煤炭生产量提供的原始信息, 组成 2010~2018 年的原始序列如下: $X(t) = \{x(2010), x(2011), x(2012), x(2013), x(2014), x(2015), x(2016), x(2017), x(2018)\}$, 时间序列对应的实际值如表 2 所示.

表 2 2010~2018 年灰色 GM(1,1) 预测模型预测煤炭产量结果

| 年份 | 实际值/亿 t | 预测值/亿 t | 相对值 M |
|------|---------|---------|---------|
| 2010 | 32.40 | 32.40 | 1.000 0 |
| 2011 | 35.20 | 36.37 | 0.967 8 |
| 2012 | 36.60 | 36.47 | 1.003 6 |
| 2013 | 36.80 | 36.58 | 1.006 0 |
| 2014 | 38.70 | 36.69 | 1.054 8 |
| 2015 | 37.50 | 36.80 | 1.019 0 |
| 2016 | 33.60 | 36.91 | 0.910 3 |
| 2017 | 39.75 | 37.02 | 1.073 7 |
| 2018 | 35.81 | 37.13 | 0.964 4 |
| 2019 | | 37.24 | |
| 2020 | | 37.35 | |

2.2 灰色 GM(1,1) 模型建立

由上述灰色 GM(1,1) 预测模型的建模步骤及计算方法,计算煤矿产量的预测值,同时实际值与预测值的比值 M 如表 2 所示.

2.3 组合预测模型的计算

2.3.1 划分状态区间

由马尔科夫理论可知,为防止数据波动性对预测产生影响,应根据相对值范围集中原则对数据进行状态区间划分.由相对值 M ,用 0.910 3, 0.964 8, 1.019 3, 1.073 8 这 4 个数据为划分单位,得出 3 个状态区间 E1, E2, E3, 如表 3 所示.

表 3 状态区间的划分

| 状态编号 | 相对值范围 | 含义 |
|------|-----------------|-----|
| E1 | 0.910 3~0.964 8 | 高估 |
| E2 | 0.964 8~1.019 3 | 较准确 |
| E3 | 1.019 3~1.073 8 | 低估 |

2.3.2 计算煤炭产量状态转移矩阵

$$P^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}; \quad P^{(2)} = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}; \quad P^{(3)} = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

2.3.3 煤炭产量预测

根据所求煤炭状态转移矩阵,选取与 2019 年相隔时间最近的 2018, 2017, 2016 年的实际值,分别对应 1, 2, 3 步转移,在各个转移步数所对应的状态转移概率矩阵中,同时,加入 2016~2018 年的初始状态,组成新的状态转移矩阵,计算各列向量之和,计算所得数值最大者即为 2019 年对应的状态,如表 4 所示.

通过表 4 可知,2019 年煤炭产量状态处于“E3”低估状态,所对应状态区间为 1.019 3~1.073 8,基于灰色马尔科夫预测模型,得到 2019 年煤炭产量预测值为 $\hat{X}^{(0)}(k) = 38.97$.

表 4 煤炭产量的状态预测

| 起始年份 | 初始状态 | 转移步数 | E1 | E2 | E3 |
|-------|------|------|-----|-----|----|
| 2018 | E1 | 1 | 0 | 0 | 1 |
| 2017 | E3 | 2 | 1/2 | 1/2 | 0 |
| 2016 | E1 | 3 | 0 | 0 | 1 |
| 列向量之和 | | | 1/2 | 1/2 | 2 |

由此,基于灰色马尔科夫模型的 2010~2019 年煤炭产量预测值、残差、相对误差等如表 5 所示.

表 5 基于灰色 GM(1,1)-马尔科夫组合模型的煤炭产量预测结果

| 年份 | 实际值/亿 t | 预测值/亿 t | 残差 | 相对误差 |
|------|---------|----------|----------|---------|
| 2010 | 32.40 | 34.142 4 | -1.742 4 | 0.053 8 |
| 2011 | 35.20 | 36.080 9 | -0.880 9 | 0.025 0 |
| 2012 | 36.60 | 36.180 1 | 0.419 9 | 0.011 5 |
| 2013 | 36.80 | 36.289 2 | 0.510 8 | 0.013 9 |
| 2014 | 38.70 | 38.398 0 | 0.302 0 | 0.007 8 |
| 2015 | 37.50 | 36.507 4 | 0.992 6 | 0.026 5 |
| 2016 | 33.60 | 34.605 0 | -1.005 0 | 0.029 9 |
| 2017 | 39.75 | 38.858 4 | 0.891 6 | 0.022 4 |
| 2018 | 35.81 | 34.914 3 | 0.895 7 | 0.025 0 |
| 2019 | | 38.973 5 | | |
| 2020 | | 39.160 5 | | |

2.4 预测精度对比

由表5可得组合预测模型的平均相对误差、均方差比值和小误差概率分别为:0.024 0, 0.427 2, 0.800 0.比较灰色GM(1,1)模型与灰色马尔科夫组合预测模型的煤矿产量相关数据,虽然预测方法的平均相对误差,均方差比值及小误差概率等级相同,但组合预测模型的相关数据预测精确度优于灰色GM(1,1)模型.运用2种方法预测精度的相关数据如表6所示.

表6 预测模型精度比较

| 模型名称 | 平均相对误差 | 均方差比值 | 小误差概率 |
|-------------|-------------|-------------|---------|
| 灰色GM(1,1)模型 | 0.039 6(二级) | 0.775 9(四级) | 0.6(四级) |
| 组合预测模型 | 0.024 0(二级) | 0.427 2(二级) | 0.8(二级) |

2.5 数据分析

图1为2种预测模型与原始数据序列的对比,可得出结论:

1)灰色GM(1,1)模型与原始序列拟合程度较小,预测情况与实际值相差较大,除去2010年,2011~2018年的预测值呈现线性趋势,反映出灰色GM(1,1)预测模型能够预测时间序列短中期发展趋势的优点.而灰色马尔科夫组合预测,与原始序列拟合度较高,更好地反映出马尔科夫能够较好地处理波动性随机变化过程.灰色马尔科夫组合预测模型具有更高的拟合度,不仅能够预测时间序列短中期发展趋势,且能较好地处理具有波动性特征的系统,使预测结果更加准确.

2)灰色马尔科夫组合预测模型,反映出我国煤炭生产量呈现逐渐上升的趋势.因国内外经济宏观环境影响煤炭价格与煤炭产量、煤炭消费需求发生改变等原因的影响,促使我国煤炭生产量呈现波动性,组合预测模型较优的将各方面因素考虑在内.

综上所述,为准确把握我国煤炭产量未来发展趋势,为相关单位制定煤炭产量使用方案提供指导与借鉴,引入新陈代谢原理,结合组合预测模型预测2019~2020年煤炭产量.首先在2010~2018年煤炭产量数据中加入2019年煤炭产量组合预测值38.97亿t,然后去掉2010年数据,由此组成2011~2019年新的原始序列,以此建模计算,即可预测出2020年煤炭产量,拟合精度如表7所示.

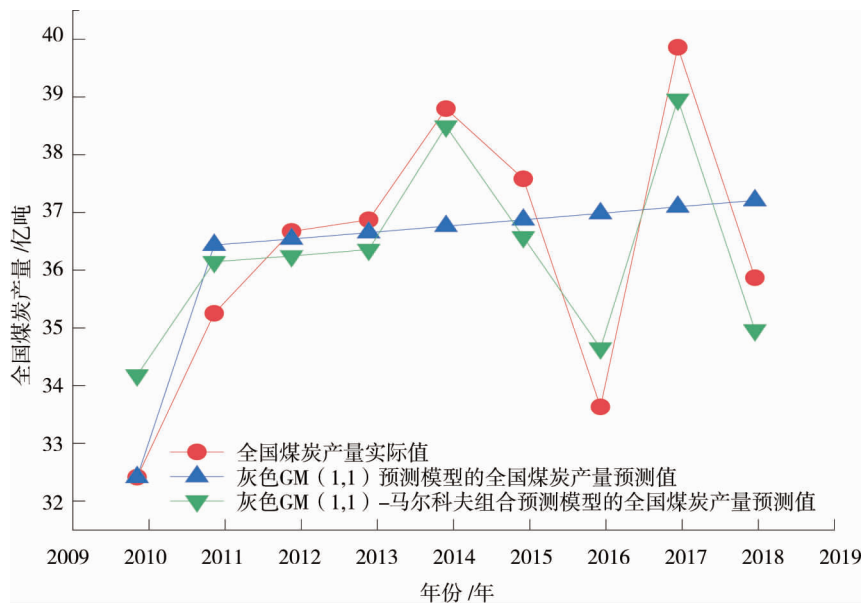


图1 2种预测模型与原始序列对比

表7 灰色GM(1,1)-马尔科夫模型及新陈代谢模型拟合精度

| 年份 | 平均相对误差 | 均方差比值 | 小误差概率 | 精度等级 |
|------------|-------------|-------------|---------|------|
| 2010~2018年 | 0.024 0(二级) | 0.427 2(二级) | 0.8(二级) | 二级 |
| 2011~2019年 | 0.018 5(二级) | 0.384 1(二级) | 1.0(一级) | 二级 |

通过表7可知,灰色GM(1,1)-马尔科夫组合预测模型的精度比单一灰色GM(1,1)模型的精度有所提高,因此采用该预测方法对2019~2020年煤炭产量进行预测是有效、可靠的.2019年和2020年煤炭产量预测值如表8所示.

表8 2019~2020年组合模型预测值

| 煤矿产量预测 | 2019年 | 2020年 |
|--------|-----------|-----------|
| | 38.97 亿 t | 39.16 亿 t |

3 结论

1)对比单一灰色预测模型与灰色马尔科夫组合预测模型预测结果可发现,灰色马尔科夫组合预测模型预测精度高,降低了数据的随机性,结合了灰色预测模型所需样本数量少和马尔科夫模型能较好处理波动性较大数据的优势,能够有效地预测2019年和2020年煤炭产量.

2)由2019年和2020年煤炭产量预测数据可分析,煤炭产量今后2年会呈现增长趋势,国家应该就此制定相关政策,合理处理煤炭消费需求 and 煤炭产量之间的关系,减少煤炭行业相关事故,合理控制产量.

参考文献:

- [1] 刘伟,刘晨君.改革开放四十年来煤炭行业安全发展之路[J].煤炭经济研究,2018,38(11):34-42.
- [2] 罗文柯,施式亮,李润求,等.灰色预测模型在能源消费需求预测中的应用[J].中国安全科学学报,2010,20(4):32-37.
- [3] 赵卫.电煤价格影响因素及预测分析[J].发电与空调,2017,38(2):1-6.
- [4] 焦光华.2020年煤炭消费量预测及情景分析[J].内蒙古煤炭经济,2015(2):32-33.
- [5] 刘延松,吴翠.煤炭产需量预测方法讨论与实证分析[J].西安科技大学学报,2008(3):503-506.
- [6] 王芳.基于改进GM(1,1)灰色模型的煤炭产量预测与分析[J].煤炭技术,2014,33(1):84-86.
- [7] 叶斌.我国原煤产量的灰色预测研究[J].科学技术与工程,2011,11(20):4947-4949.
- [8] 曹建,施式亮,曹华娟,等.基于GM(1,1)-Markov的危化品道路运输事故与交通事故预测及关系研究[J].中国安全生产科学技术,2019,15(1):26-31.
- [9] 刘寿兰,周新良,罗文柯,等.基于改进灰色马尔柯夫模型对我国煤炭生产总量的预测[J].矿业工程研究,2011,26(1):76-80.
- [10] 樊超,郭亚菲,曹培格,等.基于灰色-马尔科夫模型的粮食产量预测[J].江苏农业科学,2018,46(9):346-349.
- [11] 刘成勇,万伟强,陈蜀喆,等.基于灰色马尔科夫模型的船舶交通流预测[J].中国航海,2018,41(3):95-100.
- [12] 王星,刘小勇.基于灰色马尔科夫模型的交通事故预测研究[J].交通科技与经济,2017,19(4):9-13.