

# 基于组合预测模型的煤炭能源消费总量预测

戴剑勇, 郭睿<sup>1</sup>

(南华大学 资源环境与安全工程学院, 湖南 衡阳 421000)

**[摘要]** 文章以2000—2016年的煤炭能源消费总量为原始数据,分析改进灰色预测模型,同时引入逻辑斯蒂预测模型,以组合模型与原始数据的残差平方和最小为目标函数确定组合模型权重系数,建立组合预测模型,通过验证3者的预测结果,得出组合模型比单一预测模型预测效果更好,适用性更强,对于煤炭能源消费总量的预测具有较好的指导意义与可行性。

**[关键词]** 煤炭能源消费总量; 逻辑斯蒂模型; 灰色模型; 组合预测

**[中图分类号]** F405 **[文献标识码]** A **[文章编号]** 1673-0755(2019)03-0059-05

煤炭是中国的重要能源,中国经济的较快发展以及工业化、城镇化的推进将拉动煤炭需求的上升,而煤炭消费的增长也将对中国的能源环境产生一定的压力。因此,一方面,我们需要对煤炭的未来消费总量、结构、趋势进行科学建模与预测,对政府能源战略的制定、环境质量的预测、环境政策的实施提供参考与预警;另一方面,鉴于需求量决定生产量,供不应求和供过于求的市场对相关企业的损失都是不可忽视的,因此科学的煤炭消费预测方法对评估煤炭需求的中长期变化、稳定煤炭市场平衡、提高煤炭行业效率、保障煤炭企业经营成果均有重要的价值。

预测理论是一种对已知数据进行分析处理,挖掘其潜在规律,进而达到预测未来数据的方法。目前在这方面的研究已有不少,如徐荣辉等<sup>[1-4]</sup>做了逻辑斯蒂回归方程方面的预测研究;张鑫等<sup>[5-7]</sup>通过对灰色预测结果进行马尔可夫修正得到了改进灰色预测模型;王安等<sup>[8-9]</sup>根据灰色模型原理,构建了灰色增量模型和灰色组合预测模型。对于国内的能源预测方面,孙涵等<sup>[10]</sup>运用生活方式分析法(CLA)对城镇居民生活能源消费量进行了预测分析;向思桐等<sup>[11]</sup>通过Holt-Winter模型分析了铁路煤炭运输的预测;王鉴雪等<sup>[12]</sup>采用灰色预测分析了煤炭消费的影响因素;时合生等<sup>[13]</sup>通过马尔可夫残差修正的灰色理论对煤炭需求进行了预测。而在国外,Ujjwal kumar等<sup>[14]</sup>采用时间序列模型对印度的能源消费进行了预测分析;Rout U K等<sup>[15]</sup>通过时间的线

性编程模型分析并预测了中国的能源发展和排放。

考虑到当前煤炭消费量的发展形势,单靠灰色预测显得有些不足,本文将引入逻辑斯蒂预测,利用其S曲线的特性,将二者进行组合构建组合预测模型,得到一种更加符合实际情况的预测模型。

## 一 改进灰色预测模型

### (一)传统灰色GM(1,1)预测模型

灰色系统是指系统内部分信息已知,而另一部分信息未知,同时各因素间有不确定关系的系统。灰色预测法是一种预测灰色系统的预测方法。灰色预测通过鉴别系统因素之间发展趋势的相异程度,即进行关联分析,并对原始数据进行生成处理来寻找系统变动的规律,生成有较强规律性的数据序列,然后建立相应的微分方程模型,从而预测事物未来发展趋势的状况。其建模步骤如下:

#### 1. 设原始数据列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (1)$$

#### 2. 计算数列级比

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 1, 2, \dots, n \quad (2)$$

如果所有级比都落在可容覆盖区间  $X = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$  内,则数列  $x^{(0)}$  可建立GM(1,1)进行预测,否则对数据做平移变换:

$$y^{(0)}(k) = x^{(0)}(k) + c, k = 1, 2, \dots, n \quad (3)$$

取合适的  $c$  使新数据列的级比都落在可容覆盖区间内。

**[收稿日期]** 2019-03-07

**[作者简介]** 戴剑勇(1969-),男,湖南新化人,南华大学资源环境与安全工程学院教授。

<sup>1</sup> 南华大学资源环境与安全工程学院硕士研究生。

## 3. 累加生成(AGO)

令:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n \quad (4)$$

得到1次累加的数据列

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \quad (5)$$

4. 确定向量矩阵  $B$  和  $Y$ 

$$B = \left\{ \begin{array}{cc} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{array} \right\}$$

$$Y = \left\{ \begin{array}{c} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{array} \right\} \quad (6)$$

通过最小二乘法计算参数  $a, b$ , 计算式为:

$$[a, b]^T = (B^T B)^{-1} B^T Y \quad (7)$$

## 5. 建立预测模型白化方程

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \quad (8)$$

解得

$$x^{(1)}(t) = \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-a(t-1)} + \frac{b}{a} \quad (9)$$

因此, 预测模型为:

$$x^{(1)}(k+1) = \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad (10)$$

$$k = 1, 2, \dots, n$$

## 6. 原始数列的预测模型为

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (11)$$

$$k = 1, 2, \dots, n$$

## (二) 滑动无偏 GM(1,1) 预测模型

滑动无偏 GM(1,1) 预测模型建立步骤如下:

## 1. 设原始序列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (12)$$

## 2. 对原始序列进行一次加权滑动处理

$$x_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(n)) \quad (13)$$

其中:

$$x_1^{(0)}(1) = \frac{1}{4} [3x^{(0)}(1) + x^{(0)}(2)]$$

$$x_1^{(0)}(n) = \frac{1}{4} [x^{(0)}(n-1) + 3x^{(0)}(n)]$$

$$x_1^{(0)}(i) = \frac{1}{4} [x^{(0)}(i-1) + 2x^{(0)}(i) + x^{(0)}(i+1)]$$

3. 将数列按传统 GM(1,1) 建模步骤求出参数  $a, b$ , 根据文献[7], 无偏 GM(1,1) 的模型参数  $u, A$  为

$$u = \ln \frac{2-a}{2+a}, A = \frac{2b}{2+a} \quad (14)$$

## 4. 原始数据预测模型为

$$\hat{x}^{(0)}(1) = x^{(0)}(1)$$

$$\hat{x}^{(0)}(k+1) = Ae^{uk} \quad (15)$$

此模型相较传统 GM(1,1) 消除了模型的固有误差。

## (三) 马尔可夫模型修正

考虑到灰色预测模型在长期预测中的不足, 而马尔可夫模型正好可以弥补其不足, 因此通过马尔可夫模型对组合模型的预测结果进行修正, 可以获得更高精度的结果。其步骤如下:

1. 对组合模型预测值和实际值的残差曲线  $Y(k)$  上下两边分别做  $n, m$  条与  $Y(k)$  平行的曲线, 因此,  $m+n+1$  条平行曲线共组成  $m+n$  个状态区间  $Q_1, Q_2, \dots, Q_{m+n}$ 。

## 2. 建立1步转移概率矩阵:

$$p_{ij} = \frac{n_{ij}}{N_i} \quad (16)$$

其中  $p_{ij}$  表示数据由状态  $Q_i$  经过1步转移到  $Q_j$  的概率,  $n_{ij}$  表示由状态  $Q_i$  经过1步转移到  $Q_j$  的点的数目,  $N_i$  表示处于状态  $Q_i$  的所有点的数目。因此1步转移概率矩阵为

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1j} \\ \vdots & & \vdots \\ p_{j1} & \cdots & p_{jj} \end{bmatrix} \quad (17)$$

## 3. 预测值求解:

$$w(t) = \hat{e}(t) + 0.5(\theta_1 + \theta_2) \quad (18)$$

其中  $w(t)$  为修正后的预测值,  $\hat{e}(t)$  为滑动无偏 GM(1,1) 模型所求的预测值,  $\theta_1$  和  $\theta_2$  为残差所在状态区间的上下限。

虽然改进灰色预测模型在处理不确定性因素数据组时具有很大优势, 但在预测精度方面还存在不足, 而且随着数据量的增加其劣势也会愈加明显。

## 二 逻辑斯蒂预测

逻辑斯蒂曲线又称逻辑斯蒂函数, 是一种常见的 S 形函数, 它是数学生物学家皮埃尔·弗朗索瓦·韦吕勒于 1844 年在研究它与人口增长的关系时命名的。逻辑斯蒂曲线常被用来描述在某些限定条件下的种群增长模型。开始阶段, 曲线大致成指数增长; 后期开始变得饱和, 增长变慢; 最后, 达到完全饱和时增长停止。其函数图像如图 1 所示。

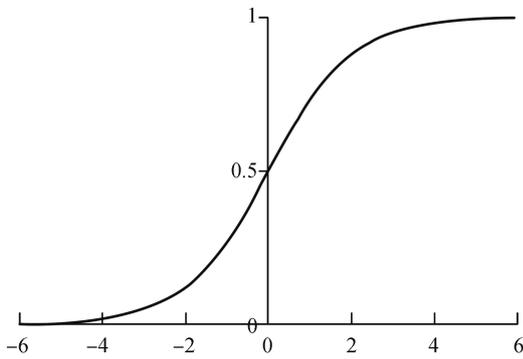


图 1 逻辑斯蒂函数图

逻辑斯蒂方程:

$$M = \frac{A}{1 + Be^{-rt}} \quad (19)$$

其中  $M$  为种群数量,  $A$  为环境容量,  $r$  为增长率,  $t$  为时间变量,  $B$  为常数。

逻辑斯蒂预测的优点在于对波动较小数据组的拟合具有更高的精度。

### 三 组合模型

为了得到一种既能处理不确定因素数据组同时又能保证精度要求, 这里将建立灰色预测和逻辑斯蒂预测的组合预测模型。

#### (一) 建立组合模型

设以时间为指标的煤炭消费量序列  $\{A(t)\}$ ,  $t = 1, 2, \dots, n$ , 采用滑动无偏 GM(1, 1) 预测的序列为  $\{A_1(t)\}$ , 逻辑斯蒂方程预测的序列为  $\{A_2(t)\}$ , 则组合预测模型为

$$\hat{e}(t) = \omega_1 A_1(t) + \omega_2 A_2(t), \quad t = 1, 2, \dots, n \quad (20)$$

其中  $A_1(t)$  表示滑动无偏 GM(1, 1) 的预测值,  $A_2(t)$  表示逻辑斯蒂方程预测值,  $\omega_1, \omega_2$  为各预测值所占权重。

#### (二) 权重值计算

这里以组合预测模型的残差平方和  $S$  最小为目标函数, 则建立非线性规划模型:

$$S = \sum_{t=1}^n (A(t) - (\omega_1 A_1(t) + \omega_2 A_2(t)))^2$$

$$\text{s. t. } \begin{cases} \omega_1 + \omega_2 = 1 \\ \omega_1, \omega_2 \geq 0 \end{cases} \quad (21)$$

其中  $A(t)$  表示实际值,  $A_1(t)$  表示第 1 种预测方法的预测值,  $A_2(t)$  表示第 2 中预测方法的预测值,  $\omega_1, \omega_2$  为各预测值所占权重。

### 四 实例分析

以 2000 年至 2016 年的国家煤炭能源消费总量

为例, 其统计数据如表 1 所示。

表 1 2000 年至 2016 年煤炭消费总量统计表

年份	煤炭消费总量 / 千万吨	年份	煤炭消费总量 / 千万吨
2000	100.670	2009	240.666
2001	105.772	2010	249.568
2002	116.160	2011	271.704
2003	138.352	2012	275.464
2004	161.657	2013	280.999
2005	189.231	2014	279.329
2006	207.402	2015	273.849
2007	225.795	2016	270.320
2008	229.237		

(数据来源: 国家统计局能源消费总量数据)

#### (一) 改进灰色预测

按照上文提到的滑动无偏 GM(1, 1) 预测模型建立步骤, 求得  $a = -0.0498$ ,  $b = 139.630$ , 因此  $u = 0.0498$ ,  $A = 143.196$ , 则预测模型为

$$\hat{x}^{(0)}(k+1) = 143.196e^{0.0498k}, k = 1, 2, \dots, n$$

再由马尔可夫修正得其预测结果如表 2 所示。

表 2 改进灰色预测值对照表

年份	实际值	预测值	残差	相对误差
2000	100.670	100.670	0	0
2001	105.772	125.508	-19.736	0.1866
2002	116.160	133.193	-17.033	0.1466
2003	138.352	141.270	-2.918	0.0211
2004	161.657	149.760	11.897	0.0736
2005	189.231	158.684	30.547	0.1614
2006	207.402	168.063	39.339	0.1897
2007	225.795	217.921	7.874	0.0349
2008	229.237	228.282	0.955	0.0042
2009	240.666	239.172	1.494	0.0062
2010	249.568	250.619	-1.051	0.0042
2011	271.704	262.650	9.054	0.0333
2012	275.464	275.295	0.169	0.0006
2013	280.999	288.586	-7.587	0.0270
2014	279.329	262.555	16.774	0.0601
2015	273.849	277.238	-3.389	0.0124
2016	270.320	292.670	-22.350	0.0827

通过表 2 可以看出改进灰色预测结果的部分数据具有较大偏差, 其平均误差为 0.0614, 并不理想。这主要由于灰色预测本身模型的不足, 另外在马尔可夫修正的时候, 状态空间的划分也会对预测结果产生影响。观察煤炭消费量在后面几年的数据也可

以看出改进灰色预测的结果与实际值的偏差越来越大,显然单靠灰色预测已经不能满足实际需求了。

(二)逻辑斯蒂方程预测

由 SPSS 软件曲线拟合可得

$$A = 290, B = 2.61, k = -0.267$$

逻辑斯蒂方程为

$$M = \frac{290}{1 + 2.61e^{-0.267t}}, t = 1, \dots, n$$

拟合曲线如图 2 所示。

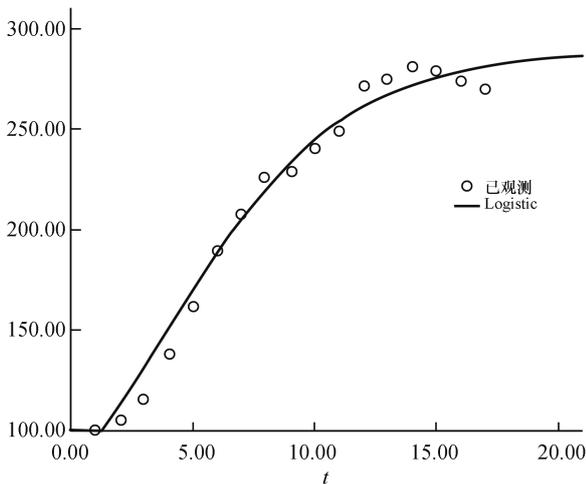


图 2 逻辑斯蒂拟合曲线图

其预测结果如表 3 所示。

表 3 逻辑斯蒂预测值对照表

年份	实际值	预测值	残差	相对误差
2000	100.670	96.718	3.952	0.0393
2001	105.772	114.619	-8.847	0.0836
2002	116.160	133.544	-17.384	0.1497
2003	138.352	152.869	-14.517	0.1049
2004	161.657	171.919	-10.262	0.0635
2005	189.231	190.052	-0.821	0.0043
2006	207.402	206.749	0.653	0.0031
2007	225.795	221.660	4.135	0.0183
2008	229.237	234.616	-5.379	0.0235
2009	240.666	245.607	-4.941	0.0205
2010	249.568	254.745	-5.177	0.0207
2011	271.704	262.215	9.489	0.0349
2012	275.464	268.237	7.227	0.0262
2013	280.999	273.038	7.961	0.0283
2014	279.329	276.832	2.497	0.0089
2015	273.849	279.810	-5.961	0.0218
2016	270.320	282.133	-11.813	0.0437

通过图 2 和表 3 可以看出,煤炭消费总量的增长趋势更符合逻辑斯蒂曲线,且判定系数  $R^2$  为 0.926,也表明拟合效果较好,其平均相对误差为 0.0409,比改进灰色预测的精度要高。

(三)组合模型预测

通过 matlab 计算上述第 3 部分中的非线性规划模型可得  $\omega_1 = 0.1198, \omega_2 = 0.8802$ ,组合预测模型为

$$\hat{e}(t) = 0.1198A_1(t) + 0.8802A_2(t), t = 1, 2, \dots, n$$

其预测结果如表 4 所示。

表 4 组合模型预测值对照表

年份	实际值	预测值	残差	相对误差
2000	100.670	97.191	3.479	0.0346
2001	105.772	115.924	-10.152	0.0960
2002	116.160	133.502	-17.342	0.1493
2003	138.352	151.479	-13.127	0.0949
2004	161.657	169.264	-7.607	0.0471
2005	189.231	186.294	2.937	0.0155
2006	207.402	202.114	5.288	0.0255
2007	225.795	221.212	4.583	0.0203
2008	229.237	233.857	-4.620	0.0202
2009	240.666	244.836	-4.170	0.0173
2010	249.568	254.251	-4.683	0.0188
2011	271.704	262.267	9.437	0.0347
2012	275.464	269.083	6.382	0.0232
2013	280.999	274.901	6.098	0.0217
2014	279.329	275.122	4.207	0.0151
2015	273.849	279.502	-5.653	0.0206
2016	270.320	283.395	-13.075	0.0484

通过表 4 可以看出组合预测模型的平均误差为 0.0413 在预测精度上比改进灰色预测高 30%,而与逻辑斯蒂预测相差不到 1%,但对比残差平方和可以知道,组合预测模型的残差平方和为  $1.1519 \times 10^3$ ,逻辑斯蒂预测的残差平方和为  $1.1975 \times 10^3$ ,因此综合来看组合预测模型更好,3 种预测的相对误差如图 3 所示。另外,随着样本数据量的增加组合模型的预测效果会更加明显。

通过计算得出 2017 年至 2019 年的煤炭消费量(千万吨),结果如表 5 所示。

表 5 三种预测结果对照表

年份	改进灰色预测	逻辑斯蒂	组合预测
2017	308.891	283.938	286.927
2018	325.940	285.335	290.200
2019	343.859	286.415	293.297

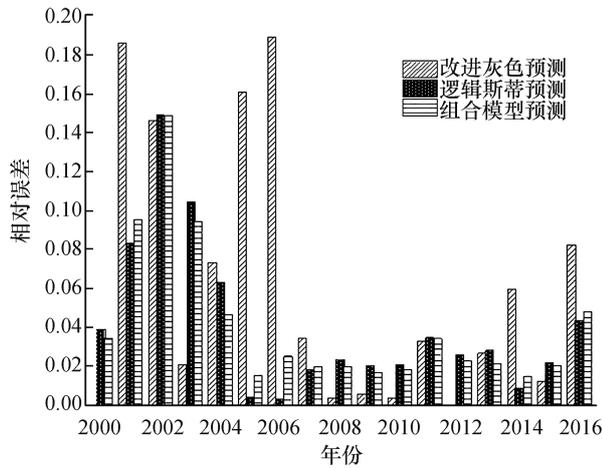


图3 三种预测相对误差对比图

## 五 结 论

通过对比改进灰色预测、逻辑斯蒂预测及组合预测得到的结果,可以知道,改进灰色预测仍然存在着固有的局限性,而且在处理非线性较明显的数据时存在很大偏差,而逻辑斯蒂曲线则相对较好,但当数据量小且波动较大的时候,曲线拟合的效果就大打折扣,而组合模型以残差平方和最小为目标函数确定组合权重,在保留前两者优点的同时减弱异常值的影响,提高了预测精度。总体来说,组合模型对于煤炭消费总量的预测具有较好的准确性和适用性。

### [参考文献]

- [1] 徐荣辉. 逻辑斯蒂方程及其应用[J]. 山西财经大学学报, 2010, 32(Z2): 311-312.
- [2] 代涛, 黄显峰, 徐学军. 离散 Logistic 人口增长预测模型研究[J]. 三峡大学学报(自然科学版), 2010, 32(5): 102-105.
- [3] 张鹏丽, 李育. 逻辑斯蒂模型在河谷型城市洪水事件研究中的验证[J]. 西北师范大学学报(自然科学版), 2017, 53(1): 128-134.
- [4] 梁慧玲, 王文辉, 郭福涛, 林芳芳, 林玉蕊. 比较逻辑斯蒂与地理加权逻辑斯蒂回归模型在福建林火发生的适用性[J]. 生态学报, 2017, 37(12): 4128-4141.
- [5] 张鑫, 任永泰, 王福林, 付强. 基于改进灰色马尔科夫模型的年降水量预测[J]. 数学的实践与认识, 2011, 41(11): 51-57.
- [6] 李义华, 杜康, 周洁. 基于改进灰色马尔科夫模型的木材需求量预测[J]. 中南林业科技大学学报(自然科学版), 2017, 37(12): 133-138.
- [7] 侯翔龙, 阳辉. 基于滑动无偏灰色马尔科夫模型的水库年降雨量预测[J]. 水土保持通报, 2014, 34(3): 181-184.
- [8] 王安, 杨雨. 基于最小二乘法的灰色组合预测模型及其应用[J]. 数学的实践与认识, 2018, 48(22): 291-299.
- [9] 石端银, 李文宇, 蔡吉花, 董志超. 改进的灰色增量模型及其在哈市人口预测中的应用[J]. 数学的实践与认识, 2013, 43(10): 71-76.
- [10] 孙涵, 周丽娜, 郭海湘. 中国城镇居民生活完全能源消费的测算及趋势分析[J]. 中国能源, 2014, 36(4): 40-44.
- [11] 向思桐, 祝继常, 宫薇薇. 基于季节性特征分析的铁路煤炭运输预测方法[J]. 统计与决策, 2019(5): 85-88.
- [12] 王鉴雪, 宁云才. 基于灰色理论的煤炭消费影响因素分析[J]. 煤炭技术, 2009, 28(11): 193-195.
- [13] 时合生, 樊爱宛, 王巍. 用马尔可夫残差修正灰色理论的煤炭需求预测[J]. 计算机仿真, 2011, 28(10): 187-190.
- [14] KUMAR U, JAIN V K. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India[J]. Energy, 2010, 35(4): 1709-1716.
- [15] ROUT U K, Alfred VoB, SINGH A, et al. Energy and emissions forecast of China over a long-time horizon[J]. Energy, 2011, 36(1): 1-11.

## Forecast of Total Coal Energy Consumption Based on Combination Forecasting Model

DAI Jian-yong, GUO Rui

(University of South China, Hengyang 421000, China)

**Abstract:** This paper takes the total coal energy consumption from 2000 to 2016 as the original data, analyzes the improved grey prediction model, and introduces the logistic prediction model. The weight coefficient of the combined model is determined by the minimum sum of the residuals of the combined model and the original data. The combination forecasting model is established. By verifying the prediction results of the three, the combined model is better than the single forecasting model and has better applicability. It has good guiding significance and feasibility for the prediction of total coal energy consumption.

**Key words:** total consumption of coal; Logistics model; Grey Model; combination forecasting