

基于 GM (1,1) 和 ARIMA 模型的山西焦炭产量预测研究

盖奇峰 (山西省沁水县自然资源局, 山西 沁水 048200)

摘要: 焦炭作为我国重要的工业原材料, 合理预测焦炭产量, 对于我国钢铁行业的发展具有重要的战略价值和经济意义。本文基于山西焦炭产量实际数据, 分别建立灰色预测 GM (1,1) 模型和 ARIMA 模型对未来山西焦炭年产量进行预测, 运用最小二乘法、AIC 和 BIC 准则确定模型参数并进行后验差比值检验, 检验结果表明 GM (1,1) 模型预测焦炭年产量准确, ARIMA 模型预测焦炭月产量有效。这些预测模型的建立为山西省未来焦炭产业发展和规划具有重要的决策意义。

关键词: 焦炭产量; GM (1,1) 模型; 求和自回归移动平均模型; 预测

焦炭作为我国重要的工业原材料, 其生产过程先后经历了快速增长、产能过剩、企稳回升三个阶段。山西省作为中国焦煤、焦炭资源的主要产区, 寻找合理的模型对山西焦炭产量进行预测对于我国钢铁行业结构优化升级具有重要的意义。目前, 诸多学者对未来我国焦炭产量进行了分析预测。杨峰华^[1]运用一元线性回归法和趋势预测法对贵州煤炭产量进行预测; 赵殿玉^[2]、李和平^[3]、张晓晔^[4]、王火雷^[5]等人建立灰色系统动态预测模型预测煤炭产量; 彭新^[6]等人建立 GM (1,1) 模型、GM (1,1) 残差模型等 GM (1,1) 模型对我国煤炭产量进行拟合与预测, 通过比较得出拟合精度较高的模型。王芳^[7]、周晓明^[8]等提出一种预测精度更高的优化 GM (1,1) 模型方法。夏劲彪^[9]基于非线性协整的 BP 神经网络, 结合多元线性回归技术建立综合模型预测煤炭产量。左佼^[10]等人采用灰色神经网络对我国煤炭产量进行预测分析, 该模型提高了传统预测模型的精度和准确性。刘寿兰^[11]等人建立改进 GM (1,1) 模型和灰色马尔柯夫模型预测煤炭产量。朱美峰^[12]等人利用卡尔曼滤波算法对中国焦炭产量进行准确的跟踪和估测焦炭产量。本文基于 1995-2019 年山西焦炭产量实际数据, 建立焦炭产量预测模型。首先, 根据山西焦炭年产量历史数据建立灰色预测 GM (1,1) 模型, 得到响应函数, 经过精度检验分析, 表明 GM (1,1) 模型预测山西焦炭年产量可靠性较高。其次, 基于山西焦炭月产量建立 ARIMA (0, 1, 3) 模型, 结合 AIC 和 BIC 准则确定模型参数, 预测结果表明该模型能够对我国山西焦炭月产量进行有效预测。

1 GM (1,1) 模型预测焦炭年产量

1.1 GM (1,1) 模型

GM (1,1) 模型是灰色预测模型, 基于原始的离散非负数据列, 通过累加生成一组新的离散数据列, 然后通过建立数列满足的一阶线性微分方程模型, 求解得到离散点处的解以及原始数据的预测值。

首先对非负原始序列 $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), X^{(0)}(3), \dots, X^{(0)}(n)\}$ 进行累加得到新的序列 $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), X^{(1)}(3), \dots, X^{(1)}(n)\}$ 满足下列方程:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (1)$$

其中 a 称为发展灰数, u 称为生控制灰数。

令 $Y_n = \{x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\}^T$, $\hat{a}$ 为待估参数向量。

于是得到模型为 $Y_n = B\hat{a}$, 通过最小二乘法求解得到微分方程的解为 $\hat{a} = (B^TB)^{-1}B^TY_n$, 以及灰色预测模型的响应函数:

$$\hat{X}^{(1)}(t+1) = (X^{(0)}(1) - \frac{u}{a})e^{-at} + \frac{u}{a}, t = 1, 2, \dots, n-1, \quad (2)$$

其中:

$$B = \begin{bmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(X^{(1)}(n-1) + X^{(1)}(n)) & 1 \end{bmatrix} \quad (3)$$

通常情况灰色预测模型精度等级设置为 4 级, 指标 C 值越小越好, 如表 1 所示。

表 1 GM (1,1) 模型后验差误差精度检验表

误差精度	C 值	P
1 级	< 0.35	好
2 级	< 0.5	合格
3 级	< 0.65	勉强合格
4 级	< 0.7	不合格

1.2 GM (1,1) 模型检验

以下选取 1995-2019 年山西省焦炭产量年度数据建立灰色预测 GM (1,1) 模型, 对山西 2020 年至 2025 年的焦炭年产量进行预测。根据灰色预测理论建立 GM (1,1) 预测模型:

$$\frac{dX^{(1)}}{dt} + 0.01219X^{(1)} = 7865.163 \quad (4)$$

$$\hat{X}_1^{(1)} = 162.4188e^{0.0921(t-1)} - 151.5288 \quad (5)$$

通过 R 软件计算原始数据序列残差的方差值, 运用

后验差比值对模型精度进行检验,其中C值为 $0.29 < 0.35$,GM(1,1)预测精度等级为“好”,检验结果表明可以用GM(1,1)模型对未来山西焦炭总产量进行预测。首先利用预测模型(4)预测2015至2019年山西省焦炭年产量以及残差和相对误差分析,如表2、图1、图2所示。其次利用山西焦炭年产量的灰色预测模型预测山西省2020-2025年焦炭年产量,预测结果见表3。

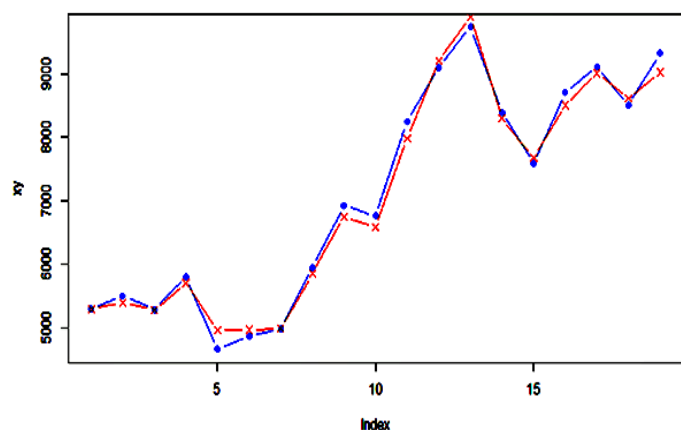


图1 1995-2014年真实值与预测值拟合曲线

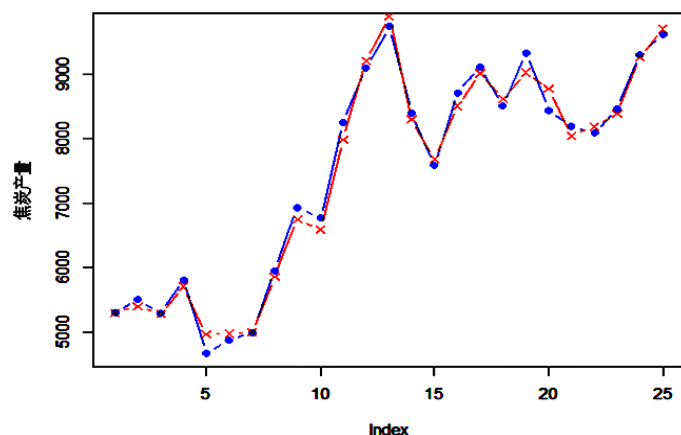


图2 1995年-2019年山西焦炭产量预测

表2 2015年-2019年山西焦炭总产量预测结果

序号	年份	实际值	预测值	残差值	相对误差
1	2015年	8040	8185.99	-45.99	-0.005720149
2	2016年	8185.99	8183.14	2.85	0.000348156
3	2017年	8383.14	8556.16	-73.02	-0.00871034
4	2018年	9256.16	9199.53	-43.37	-0.004685528
5	2019年	9699.53	9510	89.53	0.009230344

表3 2020-2025年山西焦炭年产量预测结果

年份	2020	2021	2022	2023	2024	2025
预测值 (万 t)	10371.34	10433.91	11012.22	11520.14	11928.24	12049.22

2 ARIMA 模型预测焦炭月产量

2.1 ARIMA 模型原理

求和自回归移动平均模型,简称ARIMA(p, d, q)模型,通常用于时间序列预测,主要将非平稳时间序列转化为平稳时间序列的过程中,将因变量的滞后值以及随机误差项的现值进行回归建立模型:

$$\phi(B)\nabla^d x_t = \theta(B)\varepsilon_t \quad (6)$$

$$E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_s^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t$$

$$E(x_s \varepsilon_t) = 0, \forall_s < t$$

其中 $\nabla^d = (1-B)^d$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_q B^q$ 为平稳可逆ARIMA(p, q)模型的自回归系数多项式, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为平稳可逆ARIMA(p, q)模型的移动平滑系数多项式。

2.2 ARIMA 模型构建

依据山西省2002年4月至2019年12月的焦炭月产量数据,经过缺失值插值、数据试验和模型拟合,确定山西省焦炭月产量预测的自回归移动平均模型,并利用AIC和BIC准则确定模型参数,最后对山西省2021年焦炭的月产量进行预测。

基于山西焦炭总产量数据存在1、2月份的缺失值,故采用样条插值法将数据填充完整,以我国山西焦炭月产量为变量,以单变量时间序列模型为分析工具,绘制山西焦炭总产量时序图,如图3所示。

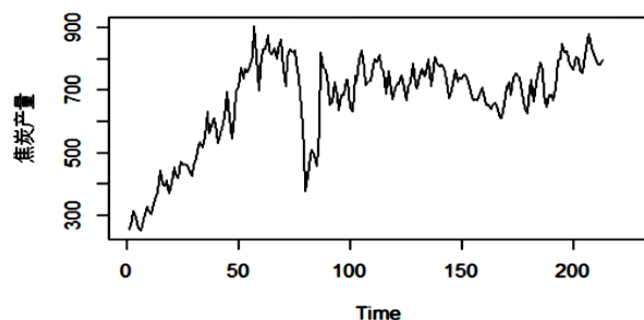


图3 2002年-2019年焦炭产量时序图

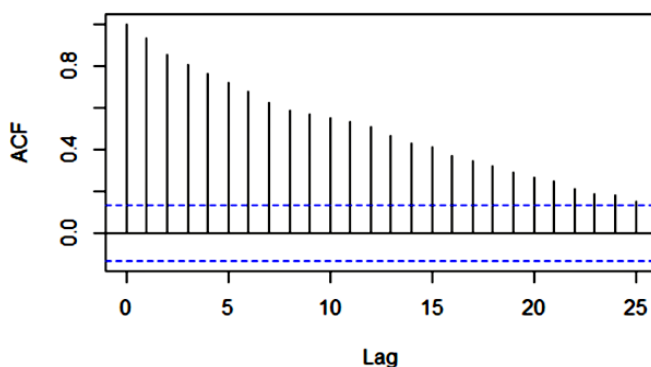


图4 2002年-2019年焦炭产量自相关图

由图4分析得出,随着延迟期数的增加,平稳序列自相关系数一直为正,且衰减向零的速度较为缓慢,可以判断焦炭产量数据为非平稳的时间序列,故将其进行一阶差分处理,消除增长趋势,得到一阶差分图,如图5所示。根据白噪声检验 $P=2.2e-16$,远小于显著性水平

0.05, 可表明其为非白噪声序列, 具备数据挖掘的研究意义, 于是结合 AIC 和 BIC 准则对模型进行定阶, 得到 ARIMA (0, 1, 3) 模型, 利用最小二乘法对 ARIMA (0, 1, 3) 模型进行拟合, 通过白噪声检验得到 $P=0.812$, 拟合模型效果显著, 如表 4。最后, 利用线性最小方差法预测 2020 年 1 月至 2020 年 12 月焦炭产量, 预测结果如表 5, 模型拟合为:

$$\nabla x_t = \varepsilon_t + 0.1085\varepsilon_{t-1} - 0.304\varepsilon_{t-2} - 0.1072\varepsilon_{t-3} \quad (7)$$

表 4 模型系数检验表

t1	0.04755707
t2	1.925816e-05
t3	0.04631261

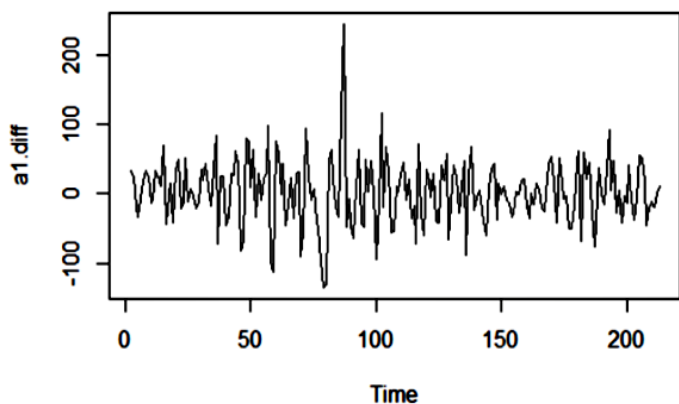


图 5 2002 年 -2019 年焦炭产量一阶差分图

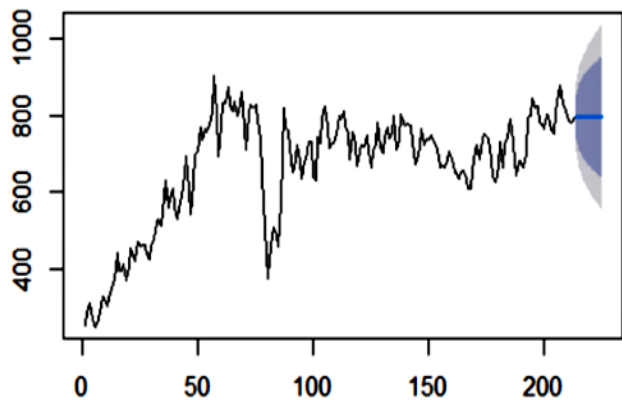


图 6 2020 年月度焦炭产量预测图

表 5 2020 年 1 月 -2020 年 12 月真实值与预测值

序号	真实值	预测值	序号	真实值	预测值
2020 年 1 月	798.71	7881.34	2020 年 7 月	963.2	958.98
2020 年 2 月	807.97	801.21	2020 年 8 月	976.4	974.3
2020 年 3 月	837.7	834.50	2020 年 9 月	960	961.93
2020 年 4 月	886.6	890.30	2020 年 10 月	920.5	915, 97
2020 年 5 月	910.3	907.12	2020 年 11 月	821.4	823.56

2020 年 6 月	939.7	934.45	2020 年 12 月	785.1	794.31
------------	-------	--------	-------------	-------	--------

2.3 ARIMA 模型的应用

利用山西省焦炭月产量 ARIMA (0, 1, 3) 模型预测山西省 2021 年 1 月 -9 月的焦炭月产量, 预测结果如表 6 所示。

表 6 2021 年 1 月 -9 月山西焦炭月产量预测结果

月份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月
预测值 (万吨)	819.2	825.1	849.6	828.7	860.7	859.6	846.2	867.3	821.9

3 结论

首先基于灰色系统理论处理后的数据, 让原始数据呈现出明显的特征, 据此建立山西省焦炭年产量的灰色 GM (1,1) 模型, 并对山西省焦炭产量进行预测。模型结果表明: 后验差比值检验 C 值为 0.29, 预测等级为“好”, 说明该模型对山西焦炭年产量的预测科学准确, 对山西省焦炭行业或能源市场的预测发展具有一定的参考价值。其次, 针对山西省焦炭月产量建立 ARIMA 预测模型, 模型检验表明: 参数检验均小于显著性水平, 说明该模型对山西省煤炭月产量的预测效果较好, 精度高。

参考文献:

- [1] 杨峰华. 贵州煤炭产量的预测研究 [J]. 中国煤炭, 2003 (01):23-25.
- [2] 赵殿玉, 邓泽民. 灰色动态模型与我国煤炭产量消费量的预测 [J]. 阜新矿业学院学报 (自然科学版), 1991 (01):36-43.
- [3] 李和平, 乔旺, 王瑞智. 基于 GM(1,1) 模型的中国煤炭总产量预测的研究 [J]. 煤炭技术, 2011, 30(03):1-3.
- [4] 张晓晔. 基于灰色联系数的煤炭产量预测方法 [J]. 安阳工学院学报, 2009(04):93-95.
- [5] 王火雷, 蒋超, 刘宇博. 基于灰色系统模型的中国煤炭生产总量预测 [J]. 中国科技信息, 2009(15):18-19.
- [6] 彭新, 李润求. 基于等维新息 GM(1,1) 模型群的煤炭产量预测研究 [J]. 煤矿现代化, 2016(05):98-101.
- [7] 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(01):84-86.
- [8] 周晓明, 罗文柯, 李润球. 改进 GM(1,1) 预测模型对我国煤炭消费需求的预测分析 [J]. 矿业工程研究, 2010, 25(02):65-68.
- [9] 夏劲彪. 基于 BP 神经网络的煤炭价格预测模型 [J]. 数码世界, 2020(08):192-193.
- [10] 左佼, 卢建军, 贾云飞, 卫晨. 基于灰色神经网络的我国原煤产量预测 [J]. 煤炭技术, 2011, 30(10):5-7.
- [11] 刘寿兰, 周新良, 罗文柯, 蒲宏桂. 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(01):76-80.
- [12] 朱美峰, 赵国浩. 卡尔曼滤波算法在焦炭产量预测中的应用 [J]. 未来与发展, 2013, 36(05):110-113.