Advances in Applied Mathematics 应用数学进展, 2022, 11(6), 4003-4009
Published Online June 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.116428

Hans汉斯

基于ARIMA和GM模型的日最高温预测

刘　芳，张　兵，刘　岩，郭　旗，葛瑞婷

淄博市气象局，山东　淄博

收稿日期：2022年5月27日；录用日期：2022年6月19日；发布日期：2022年6月29日

摘　要

基于差分整合移动平均自回归模型(Autoregressive Integrated Moving Average Model, ARIMA)、灰色预测模型(Gray Model, GM)、ARIMA和GM组合模型(ARIMA-GM)来预测夏季日最高温度。本文收集了某市2019~2021年6~8月日最高温数据作为实验数据。实验结果表明，GM、ARIMA-GM组合模型具有更好的预测结果。与GM模型相比，ARIMA-GM组合模型的预测结果更加稳定。

关键词

最高温度，预测，ARIMA模型，GM模型

The Prediction of Daily Maximum Temperature Based on ARIMA and GM Models

Fang Liu, Bing Zhang, Yan Liu, Qi Guo, Ruiting Ge

Zibo Meteorological Bureau, Zibo Shandong

Received: May 27th, 2022; accepted: Jun. 19th, 2022; published: Jun. 29th, 2022

Abstract

Based on the Autoregressive Integrated Moving Average model (ARIMA), Gray Model (GM), ARIMA and GM combined models (ARIMA-GM), the daily maximum temperature in summer is predicted. This paper collects the maximum temperature data in a city from June to August in 2019~2021, and used it as experimental data. The experimental results show that the GM model and ARIMA-GM model have better prediction results. Compared with the GM model, the prediction results of the ARIMA-GM model are more stable.

文章引用: 刘芳, 张兵, 刘岩, 郭旗, 葛瑞婷. 基于 ARIMA 和 GM 模型的日最高温预测[J]. 应用数学进展, 2022, 11(6): 4003-4009. DOI: 10.12677/aam.2022.116428

Keywords

Maximum Temperature, Prediction, ARIMA Model, GM Model

1. 引言

天气与人们的日常生活息息相关，随着社会的进步、科技、经济的发展，其对生活中多个方面都产生了巨大的影响，特别是高温、低温、大风、暴雨、冰雹和暴雪等极端天气对农业、交通、航空、建筑等行业带来严重的经济损失，甚至对人们的生命也会产生威胁，这就对天气预报准确性有更高的要求。此外，天气预报的重要特点之一就是时效性，人们需要提前预知天气情况，并根据不同的天气做出合理的应对措施。因此，预测未来天气是气象领域的关键任务之一。

由于传输气象数据设备的不断完善、升级，其收集的数据越来越多[1]，这给预测天气技术带来了新的机遇和挑战。原始的预测方法不仅不能满足公众对天气要素预测准确率的需求，而且无法充分利用目前已收集到的气象数据。因此，面对这些挑战，许多研究者提出了新的预测方法、技术，如物理预测法[2] [3]、基于统计学的方法[4] [5] [6] [7] [8]、综合预测方法[9]。其中，基于统计学的方法是采用数据统计分析方法，统计某一预测对象在某个时间内出现的频率，从而推算出未来时间段内，在相似环境条件下该对象出现的概率。

时间序列预测方法也是统计学方法中的一种。时间序列是根据时间排序的一组随机数据，时间序列预测是根据历史数据来对未来预测[10] [11]。差分整合移动平均自回归模型(Autoregressive Integrated Moving Average model, ARIMA)是一个经典的时间序列预测模型[12]，该模型较为简单、灵活，在预测股票走势[13] [14] [15] [16]、人口增长数量、区域经济 GDP 增量[17]、交通流量和气象要素等各个方面具有重要的意义。例如、Kwong 等人[5] [6]利用人工神经网络(ANN)方法对风速进行了预测评估。郭[18]等使用自回归模型来预测季节的降雨量。其次，灰色预测模型(GM) [19]也是常用的预测模型之一，它可对即含有已知信息又含有不确定信息的系统进行预测。其优点就是所需历史数据少，预测越近期的数据越有效，解决了大量的生产、生活和科学研究等方面的预测问题[20] [21]。此外，为了获得更加准确的预测结果，许多学者利用 ARIMA 和 GM 的组合模型来进行预测。Wang 等[22]利用 ARIMA-GM 组合模型来预测美国页岩油产量。文献[23]、[24]利用 ARIMA-GM 组合模型分别对湖北省电力需求、PM2.5 浓度做了预测分析。ARIMA-GM 组合相对于单一的 ARIMA、GM 预测模型的最大优势，把两个单一模型组合起来来降低单个模型的敏感度，从而提高预测准确率[24]。

基于以上的分析和对日常气象服务内容的总结，本文主要利用 ARIMA 模型、GM 模型及 ARIMA-GM 组合模型来预测夏季日最高温。我们收集了某市 2019~2021 年 6~8 月的每日最高温数据来生成时间序列数据集，并分别分析了 ARIMA、GM 及 ARIMA-GM 模型在该数据集上的预测效果。实验发现，GM、ARIMA-GM 具有较好的预测结果，但 ARIMA-GM 的预测结果更加稳定。

2. ARIMA 模型

2.1. ARIMA 模型组成

ARIMA模型主要包含四部分：自回归模型(AR)、移动平均模型(MA)、自回归移动平均模型

(ARMA)和差分模型。其中ARMA模型是AR和MA模型的组合。

AR模型表示为：

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t \tag{1}$$

其中，p是自回归阶数，表示当前值与前p个历史值相关，ε_t为误差，γ_i为自相关系数。

MA模型表示为：

$$y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2}$$

其中，q为移动平均阶数。

因此，ARMA 模型可表示为：

$$y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3}$$

μ是常数项。

2.2. 差分模型

AR模型需要历史时间序列数据具有平稳性，而现实生活中获得的数据分布是复杂且多样的，但预测未来数据需要时间序列数据具有惯性，这就要求数据具有平稳性。平稳的时间序列数据集满足其均值、协方差和方差不随时间变化而发生明显变化。为了让一个非平稳的时间序列数据呈现出平稳性，本文利用差分模型实现。给定时间序列数据$Y^0 = \left\{Y_t^0, t = 1,2,\cdots,n\right\}$，差分后的时间序列$F_t$满足：$F_t = Y_t^0 - Y_{t-1}^0$，且$Y_0^0 = 0$，该方法称为差分法。

若做一次差分，数据集仍不具备平稳性，可做多次差分，直至数据平稳。做一次差分称为一阶差分，d次差分称作d阶差分。因此ARIMA模型可表示为ARIMA(p,d,q)，p,d,q分别为AR模型、差分模型、MA模型中的参数。

2.3. ARIMA 模型过程

本小节详细描述了 ARIMA 预测方法的过程，如算法 1 所示。

算法1：ARIMA方法

Input：Historical dataset Y^0，forecast days n；

Output：predicted value for n days

1) ini(Y^0)←preprocessing dataset Y^0

2) d-difference method for Y^0

3) p、q←calculateACF、PACF

4) Build ARIMA(p,d,q) model

5) Set_p←ARIMA(p,d,q)

6) return Set_p

首先，剔除时间序列数据集 Y^0 中的空值、异常值，然后对 Y^0 做差分操作，使其具有平稳性(如 1)、2)所示)。然后，通过采用自相关函数(ACF)和偏自相关函数(PACF)来获得 AR、MA 模型中参数 p、q 值(如3)所示)。最后，构建 ARIMA(p,d,q)模型，并输出预测值集合 Set_p (如 4)~6)所示)。

3. GM 模型

一阶灰色模型GM(1,1)是一阶的，且包含一个变量的灰色模型。建模步骤如下：

1) 将历史时间序列数据$Y^0 = \left\{Y_t^0, t = 1,2,\cdots,n\right\}$做1次累加求和处理，获得$Y^1 = \left\{Y_t^1, t = 1,2,\cdots,n\right\}$，其中

$$Y_t^1 = \sum_{i=1}^{t} Y_i^0, \quad t = 1,2,\cdots,n \tag{4}$$

2) 对获得的Y^1建立微分方程：

$$\frac{\mathrm{d}Y^1}{\mathrm{d}t} + aY^1 = u \tag{5}$$

式中a、u为未确定参数。

3) 确定参数a、u，令：

$$B = \begin{bmatrix} -\frac{1}{2}\left[Y_1^1 + Y_2^1\right] & 1 \\ -\frac{1}{2}\left[Y_2^1 + Y_3^1\right] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[Y_{t-1}^1 + Y_t^1\right] & 1 \end{bmatrix}, \quad X = \begin{bmatrix} Y_2^0 \\ Y_3^0 \\ \vdots \\ Y_t^0 \end{bmatrix}$$

利用最小二乘法估计a、u值：

$$\begin{bmatrix} a & u \end{bmatrix}^{\mathrm{T}} = \left(B^{\mathrm{T}}B\right)^{-1} B^{\mathrm{T}}X \tag{6}$$

根据公式(6)获得参数a、u值，a是控制模型发展态势，称为发展系数；u的数值大小反应数据变化的关系，称作灰色作用量。

4) 通过解微分方程(5)，可得GM(1,1)预测模型为：

$$\widehat{Y_t^1} = \left(Y_1^0 - \frac{u}{a}\right)\mathrm{e}^{-a(t-1)} + \frac{u}{a} \tag{7}$$

$$\widehat{Y_t^0} = \widehat{Y_t^1} - \widehat{Y_{t-1}^1} \tag{8}$$

式(8)中，$t = 1,2,\cdots,n$，且令$Y_0^1 = 0$。根据式(8)获得最终的预测值$\widehat{Y^0} = \left\{\widehat{Y_t^0}, t = 1,2,\cdots,n\right\}$。

接下来，给出了GM预测方法的伪代码，如算法2所示。

算法2：GM方法

Input：Historical dataset Y^0，forecast days n；

Output：predicted value for n days

1) ini(Y^0)←preprocessing dataset Y^0

2) Y^1←accumulate(Y^0)

3) creatDiffEquation(Y^1)

4) a、u←obtain by the ordinary least squares

5) Set_{preGM}←CalDiffEquation(Y^0,Y^1)

6) return Set_{preGM}

首先预处理数据集Y^0，并对处理后的Y^0中每个数据对象按照公式(4)做累加处理(1)、2)所示)。其次，创建微分方程并利用最小二乘法估计获得微分方程中的参数值a、u(3)、4)所示)。最后，把a、u值代入微分方程，据公式(7) (8)获得预测值(5)、6)所示)。

4. ARIMA-GM 模型

ARIMA和GM组合模型是按照不同权重，对两者进行组合，表示为：

$$x_t = w_1 m_1 + w_2 m_2$$

其中，x_t 为预测值，w_1、w_2 分别为ARIMA、GM模型权重，m_1、m_2 分别为ARIMA、GM的预测值。

本文通过利用差分倒数法来确定权重 w_1、w_2 值，即 $w_i = (1/s_i)\sum_{i=1}^{k}(1/s_i)$，$s_i$ 为第i个模型的误差平方和，k为模型个数。

本文因为关注于预测夏季(6、7、8月)的日最高温度，且每个月份的日最高温数据差异比较大，因此本文在预测过程中，针对每个月份分别求出一组权重值，来确保预测的数据结果更加准确。下面给出ARIMA-GM模型的详细过程：

算法3：ARIMA-GM预测模型

Input：Historical dataset Y^0，forecast days n；

Output：predicted value for n days

1) ini(Y^0)←preprocessing dataset Y^0

2) Set_{preA}←predictby ARIMA(p,d,q) model

3) Set_{preGM}←predictby GM(1,1) model

4) w_1、w_2←obtain by thereciprocal variance method

5) $Set_{pre} = w_1 * Set_{preA} + w_2 * Set_{preGM}$

6) return Set_{pre}

首先，预处理数据集Y^0，并分别利用ARIMA(p,d,q)、GM(1,1)进行预测(如1)~3)所示)。然后，利用差分倒数法获得ARIMA(p,d,q)、GM(1,1)的预测权重，根据所得权重分别对ARIMA、GM预测结果进行加权计算，获得最终的预测结果(4)~6)所示)。

5. 实验评估

本文主要采集了某市 2019~2021 年 6、7、8 月的日最高气温作为实验数据集。本文利用每个月份 70%的数据作为训练数据，30%的数据作为验证数据来评估所用模型的有效性。

预测有效性评估

本小节主要利用平均绝对误差(Mean Absolute Error, MAE)来评估 ARIMA 模型、GM 模型和 ARIMA-GM 模型的预测有效性。

平均绝对误差(Mean Absolute Error, MAE)表示为：

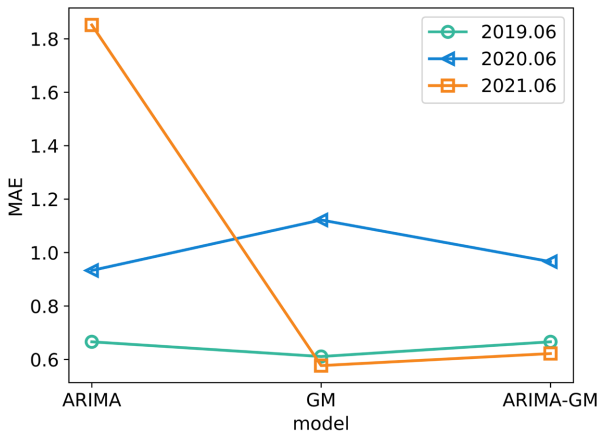
$$\mathrm{MAE} = \frac{1}{n} * \sum_{i=1}^{n} \left| y_i - \widehat{y_i} \right| \tag{9}$$

其中，y_i 是真实值，$\widehat{y_i}$ 为预测值，n 为预测天数。MAE 值越小，表示预测值与真实值越相近，误差越小。相反，MAE 值越大，表示预测值与真实值相差越大，误差越大。

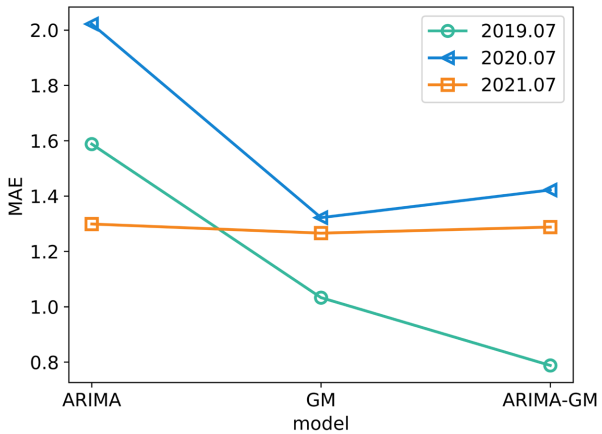
Table 1. The mean of the mean absolute errors of the different forecasting models
表 1. 不同预测模型的平均绝对误差均值

	ARIMA	GM	ARIMA_GM
2019~2021.06	1.150	0.770	0.751
2019~2021.07	1.636	1.207	1.166
2020~2022.08	0.869	0.680	0.655

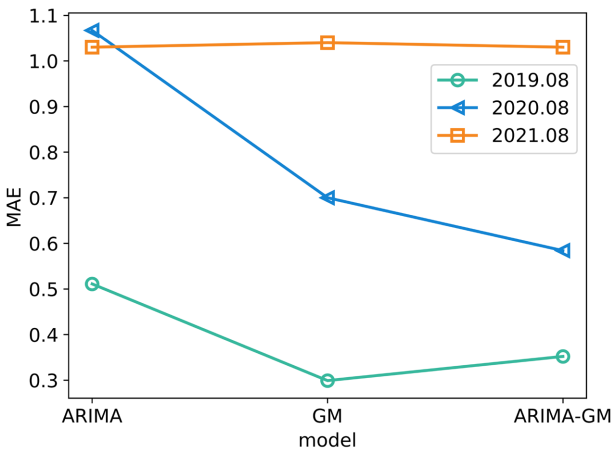
图 1 展示了 ARIMA、GM、ARIMA-GM 模型预测 2019~2021 年 6、7、8 月日最高温的绝对误差值。表 1 展示了 ARIMA、GM、ARIMA-GM 模型预测这三年(2019~2021 年) 6、7、8 月，每月的绝对误差值和的均值。通过图 1 和表 1 可知，ARIMA 的绝对误差最大，预测效果最不佳，最低平均误差为 0.869。而 GM、ARIMA-GM 误差值相近，但是从表 1 可知，ARIMA-GM 预测 2019~2021 年 6 月的平均绝对误差和的均值小于 GM 预测 2019~2021 年 6 月的平均绝对误差和的均值。同样，ARIMA-GM 预测 2019~2021 年 7、8 月的平均绝对误差和的均值也小于 GM 预测 2019~2021 年 7、8 月的平均绝对误差和的均值。因此，ARIMA-GM 的预测效果比 GM 的预测效果更加稳定。



(a) 2019~2021 年 06 月



(b) 2019~2021 年 07 月



(c) 2019~2021 年 08 月

Figure 1. Mean absolute errors of ARIMA, GM, ARIMA-GM models
图 1. ARIMA、GM、ARIMA-GM 模型平均绝对误差值

6. 总结

本文利用 ARIMA、GM、ARIMA-GM 模型对某市 2019~2021 年 6~8 月日最高温进行了预测、分析。实验表明，ARIMA 模型预测的平均绝对误差比 GM、ARIMA-GM 模型的大，而 GM 和 ARIMA-GM 模型的误差值比较接近，且相对较小，但 ARIMA-GM 组合模型具有更加稳定的预测效果。因此，在实际生活中，可以采用 GM 或 ARIMA-GM 组合模型来对夏季每日的最高温提前预测，从而减少甚至避免因为高温对公众、各部门造成的损失。

基金项目

淄博市气象局气象科研项目(2022zbqx07)。

参考文献

[1] Overpeck, J.T., Meehl, G.A., Bony, S., *et al.* (2011) Climate Data Challenges in the 21st Century. *Science*, **331**, 700-702. https://doi.org/10.1126/science.1197869

[2] 曾庆存. 大气运动的特征参数和动力学方程[J]. 气象学报, 1963, 33(4): 64-75.

[3] 曾庆存. 大气中的适应过程和发展过程(一)——物理分析和线性理论[J]. 气象学报, 1963, 33(2): 35-46.

[4] 段文广, 周晓军, 石永炜. 数据挖掘技术在精细化温度预报中的应用[J]. 干旱气象, 2012, 30(1): 130-135.

[5] Kwong, K.M., Wong, M.H.Y., Liu, J.N.K., *et al.* (2012) An Artificial Neural Network with Chaotic Oscillator for Wind Shear Alerting. *Journal of Atmospheric and Oceanic Technology*, **29**, 1518-1531. https://doi.org/10.1175/2011JTECHA1501.1

[6] Liu, J.N.K., Kwong, K.M. and Chan, P.W. (2012) Chaoticoscillatory-Based Neural Network for Wind Shear and Turbulence Forecast with LiDAR Data. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, *Part C* (*Applications and Reviews*), **42**, 1412-1423. https://doi.org/10.1109/TSMCC.2012.2188284

[7] Chen, S.M. and Hwang, J.R. (2000) Temperature Prediction Using Fuzzy Time Series. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, *Part B* (*Cybernetics*), **30**, 263-275. https://doi.org/10.1109/3477.836375

[8] Chen, S.M. and Tanuwijaya, K. (2011) Multivariate Fuzzy Forecasting Based on Fuzzy Time Series and Automatic Clustering Techniques. *Expert Systems with Applications*, **38**, 10594-10605. https://doi.org/10.1016/j.eswa.2011.02.098

[9] 路志英, 赵智超, 郝为, 等. 基于人工神经网络的多模型综合预报方法[J]. 计算机应用, 2004, 24(4): 50-51, 88.

[10] 何书元. 应用时间序列分析[M]. 北京: 北京大学出版社, 2003.

[11] 杨海民, 潘志松, 白玮. 时间序列预测方法综述[J]. 计算机科学, 2019, 46(1): 21-28.

[12] Box, G.E.P. and Pierce, D.A. (1970) Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American statistical Association*, **65**, 1509-1526. https://doi.org/10.1080/01621459.1970.10481180

[13] Ariyo, A.A., Adewumi, A.O. and Ayo, C.K. (2014) Stock Price Prediction Using the ARIMA Model. 2014 *UKSim-AMSS* 16*th International Conference on Computer Modelling and Simulation*, Cambridge, UK, 26-28 March 2014, 106-112. https://doi.org/10.1109/UKSim.2014.67

[14] Mondal, P., Shit, L. and Goswami, S. (2014) Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science*, *Engineering and Applications*, **4**, 13. https://doi.org/10.5121/ijcsea.2014.4202

[15] Siami-Namini, S. and Namin, A.S. (2018) Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. arXiv preprint arXiv:1803.06386.

[16] Wabomba, M.S., Mutwiri, M.P. and Fredrick, M. (2016) Modeling and Forecasting Kenyan GDP Using Autoregressive Integrated Moving Average (ARIMA) Models. *Science Journal of Applied Mathematics and Statistics*, **4**, 64-73. https://doi.org/10.11648/j.sjams.20160402.18

[17] 王鄂, 张霆. 时间序列在湖南省 GDP 预测中的应用——基于 ARIMA 模型[J]. 青岛大学学报: 自然科学版, 2019, 32(3): 136-140.

[18] 郭化文, 尹爱芹. 时间序列分析预测法在气象上的应用[J]. 泰安师专学报, 2002(6): 1-5.

[19] 邓聚龙. 灰色预测与决策[M]. 武汉: 华中理工大学出版社, 1988.

[20] 杨国华, 颜艳, 杨慧中. GM(1,1)灰色预测模型的改进与应用[J]. 南京理工大学学报, 2020, 44(5): 69-76.

[21] 许泽东, 柳福祥. 灰色 GM(1,1)模型优化研究进展综述[J]. 计算机科学, 2016, 43(S2): 6-10.

[22] Wang, Q., Song, X.X. and Li, R.R. (2018) A Novel Hybridization of Nonlinear Grey Model and Linear ARIMA Residual Correction for Forecasting US Shale Oil Production. *Energy*, **165**, 1320-1331. https://doi.org/10.1016/j.energy.2018.10.032

[23] 王莉琳, 张维, 赖敏, 等. 基于 ARIMA-GM 组合模型的湖北省电力需求预测研究[J]. 中国农村水利水电, 2013(4): 101-105.

[24] 徐辉军, 张林男. 基于 GM-ARMA 组合模型的 PM2.5 浓度预测——以扬州市为例[J]. 南通职业大学学报, 2018, 32(4): 67-71.