

基于改进的灰色NGM(1,1)预测模型研究及应用

——以新发传染病为例

李胡蓉, 陈燕婷*

上海理工大学管理学院, 上海

收稿日期: 2023年6月1日; 录用日期: 2023年7月12日; 发布日期: 2023年7月18日

摘 要

2020年突发传染病严重威胁人们的生命健康, 精准的预测对于传染病的管理有重要意义。基于传染病发展的不确定性, 建立传统灰色NGM(1,1)模型, 而NGM(1,1)模型初始值的选取、背景值的构造以及参数估计方法是导致该模型精度不稳定的重要原因。首先, 本文以三参数的紧邻均值背景值来优化传统NGM(1,1)模型背景值为基础, 以平均相对误差最小为目标, 利用粒子群优化算法求最优初始值的取值, 然后将三参数的紧邻均值背景值改为加权的三参数背景值, 利用粒子群优化算法寻求最优的加权权数, 最后将参数估计方法最小二乘法改为加权最小二乘法求模型的灰参数, 构成NGM(1,1)模型初始值的选取、背景值的构造和参数估计方法的组合优化。通过算例来验证优化后的灰色NGM(1,1)模型性能, 并将改进的模型应用于传染病的预测。

关键词

NGM(1,1)预测模型, 初始值, 背景值, 参数估计, 传染病

Research and Application of Improved Gray NGM(1,1) Based Prediction Model

—Emerging Infectious Diseases as an Example

Hurong Li, Yanting Chen*

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Jun. 1st, 2023; accepted: Jul. 12th, 2023; published: Jul. 18th, 2023

*通讯作者。

Abstract

The outbreak of infectious diseases in 2020 is a serious threat to people's life and health, and accurate prediction is important for the management of infectious diseases. Based on the uncertainty of infectious disease development, the traditional gray NGM(1,1) model is established, and the selection of initial values of NGM(1,1) model, the construction of background values and the parameter estimation methods are important reasons for the unstable accuracy of this model. Firstly, this paper optimizes the background value of the traditional NGM(1,1) model based on the three-parameter tight-neighborhood mean background value, and uses the particle swarm optimization algorithm to find the optimal initial value with the objective of minimizing the average relative error, then changes the three-parameter tight-neighborhood mean background value to a weighted three-parameter background value, and uses the particle swarm optimization algorithm to seek the optimal weighting weights, as well as finally changes the parameter estimation method least squares. Finally, the least squares method of parameter estimation is changed to the weighted least squares method to find the gray parameters of the model, which constitutes the optimization of the combination of the initial value selection, background value construction and parameter estimation method of the NGM(1,1) model. The performance of the optimized gray NGM(1,1) model is verified by arithmetic examples, and the improved model is applied to the prediction of infectious diseases.

Keywords

NGM(1,1) Prediction Model, Initial, Background Values, Parameter Estimation, Infectious Diseases

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

灰色预测模型是解决“少数据，贫信息”的小数据的重要方法，灰色预测模型以不需要大量的数据样本，短期预测效果好，运算过程简单的特点，且在已知数据少的情况下仍然可以较好的拟合与预测数据的发展趋势，所以灰色预测广泛应用于各个领域[1] [2] [3]。朱晓宵等[4]以武汉 2020 年新冠肺炎情况为例，首先构建滚动灰色 GM(1,1)模型，用来预测疫情的变化趋势，并与传染病动力学模型预测精度比较，得出滚动 GM(1,1)模型 > SEIHIH2RD 传染病动力学模型，这说明对于具有不确定性的传染病的传播趋势，灰色预测模型有很好的预测效果。GM(1,1)模型是目前应用最为广泛的灰色预测模型，但是据学者研究，经典 GM(1,1)有时也会出现较大的误差，导致经典的 GM(1,1)模型的模拟精度不理想，而模拟精度主要与模型参数、模型结构和原始数据序列等有关，针对这一问题，目前学者们主要从模型参数优化和模型结构优化这两个方面研究了 GM(1,1)模型的优化问题。如何提高灰色预测模型来预测事物的发展情况也是本文的研究目标。

在模型结构优化方面，考虑到经典的 GM(1,1)预测模型参数估计来源于 GM(1,1)模型的差分形式，时间响应函数来源于微分方程，造成二者的“非同源性”，为了实现模型参数估计和时间响应函数来源相同，谢乃明等提出了离散的 GM(1,1)模型，即 DGM(1,1)模型，并且得到了很好的应用[5] [6]。不管是 GM(1,1)模型还是 DGM(1,1)模型都只在近似齐次指数数据序列上可以得到较好的模拟效果，对于近似非齐次指数数据的模拟效果不佳，所以有了 GM(1,1)模型和 DGM(1,1)模型的很多拓展。张和平等[7]将可以

描述趋势为 S 型过程的 Verhulst 模型应用于网络舆情的预测, 得出模型的适用性; 邱慧[8]等将 Verhulst 模型应用于人口老龄化的预测; 石季雨等[9]等将对于近似非齐次指数数据模拟效果好的 NGM(1,1)模型应用于中国能源消费情况的预测。

在模型参数优化方面, 主要包括初始值优化、背景值优化和参数估计方法优化三个方面, 这也是目前对于灰色预测模型优化最多的三个方面。经典的灰色预测模型是选择一阶累加生成序列的第一个数据作为模型的初始值, 而这个选择并不能突出新信息的重要性。所以考虑到新信息的重要性, Zhao 等[10]引入滚动机制; 也有学者引入新陈代谢模型[11][12], 用每天更新的新数据来替换最初始数据, 给予新数据更高的优先级; 丁松等[13]考虑新信息优先原理, 引入权重对生成序列加权求和来优化 GM(1,1)模型初始值, 然后利用平均相对误差最小来确定权重值; 郑坚等[14]同样考虑了生成序列的综合影响, 按生成数据的大小对数据赋权值, 并且引入了时间权重函数来反应各个误差平方和的重要性。经典的 GM(1,1)模型的背景值是采用紧邻均值构造法, 构造出的梯形面积就是经典的灰色模型的背景值, 该值与实际的曲边梯形的面积之差就是经典 GM(1,1)模型背景值构造的误差值。为减小经典背景值选取的误差, 卢捷等[15]将背景值设为变量, 然后用平均相对误差最小来确定背景值参数, 证明了优化后的模型精度提高, 同时也对初始值重新取值对模型进行优化; 杨孝良等[16]考虑到原始序列中极端值的影响, 提出了三参数紧邻均值来构造背景值, 弱化了原始序列中极端值的影响; 刘震等[17]对 NGM(1,1)的微分方程进行积分, 通过分析误差产生原因来重新推导背景值表达式。对于灰色预测模型的参数估计方法, 经典 GM(1,1)模型的参数估计方式是最小二乘法, 最小二乘法要求数据总体为正态分布, 而经典 GM(1,1)模型的数据总体特征不明确, 且最小二乘法存在稳健性差的问题。许多研究者对此进行了改进, 比如何霞[18]利用加权最小二乘法来提高模型的精度和稳健性; 周德强[19]利用 LS-SVM 来估计灰色 Verhulst 模型的参数, 符合灰色 Verhulst 模型小样本的建模, 提高了灰色 Verhulst 预测模型的精度。

以上研究都是从灰色预测误差主要来源的一个方面或者两个方面对灰色预测模型进行优化, 在一定程度上提高了模型的精度和减少了预测模型误差, 但是没有总体解决误差来源问题, 所以基于以上研究, 本研究首先构建 NGM(1,1)模型, 从背景值、初始值和参数估计方法三个方面对 NGM(1,1)灰色预测模型进行组合优化。用优化后的灰色 NGM(1,1)预测模型进行算例分析, 计算出模型的相对平均误差、方差比(C)和小误差概率(p)三个指标值, 通过与其他模型对比三个指标值的大小, 比较模型的性能。最后将本文构建的组合优化的模型应用于武汉市新型冠状病毒肺炎新增感染人数模拟, 分别采用不同模型建模并与本文模型进行对比, 结果表明所提出的优化模型能够获得较高精度。

2. NGM(1,1)模型优化

NGM(1,1)模型是基于经典 GM(1,1)模型构建的一种能够同时模拟近似齐次指数数据和近似非齐次指数数据的灰色预测模型, 而 GM(1,1)模型是针对近似齐次指数数据有较为理想的预测效果, 由于现实生活中大多数数据都是近似非齐次指数的, 近似齐次指数数据只是一种理想状态, 所以基于现实数据, NGM(1,1)模型具有更好的预测效果。本节基于文献[16]依次从初始值取值、背景的构造和参数估计方法三个方面来优化 NGM(1,1)模型。

2.1. 三参数背景值的 NGM(1,1)模型

NGM(1,1)模型[20]原始数列为: $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 累加生成 $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, 其中 $k=1, 2, \dots, n$, 用一阶微分方程拟合 $X^{(1)}$, 得到 NGM(1,1)的白化微分方程形式:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = bt + c \quad (1)$$

生成紧邻均值生成序列 $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$, 其中 $z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1))$, $k = 2, 3, \dots, n$, 作为模型的背景值, 已知数据是离散的而不是连续的, 将 $\frac{dx^{(1)}}{dt} \rightarrow \frac{\Delta x^{(1)}}{\Delta t} = \frac{x^{(1)}(k) - x^{(1)}(k-1)}{t - (t-1)} = \sum_{i=1}^k x^{(0)}(i) - \sum_{i=1}^{k-1} x^{(0)}(i) = x^{(0)}(k)$, $k = 2, 3, \dots, n$, 和背景值带入式(1)得到 NGM(1,1)的基本形式:

$$x^{(0)}(k) + az^{(1)}(k) = bk + c \quad (2)$$

式(2) a 、 b 和 c 为待估灰参数。根据杨孝良等[16]为了提高 NGM(1,1)模型背景值的平滑效果, 减少模型中极端值对模型精度的影响, 提出三参数紧邻均值作为 NGM(1,1)模型的背景值, 即:

$$z^{(1)}(k) = \frac{1}{3}(x^{(1)}(k) + x^{(1)}(k-1) + x^{(1)}(k-2)) \quad (3)$$

其中 $k = 3, 4, \dots, n$, 代替传统 NGM(1,1)背景值 $z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1))$, 其中 $k = 2, 3, \dots, n$, 将三参数背景值带入式(2), 利用最小二乘法求解参数 a 、 b 和 c 。将 $x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$ 和式(3)带入式(2)得:

$$\hat{x}^{(1)}(k) = \frac{1 - \frac{a}{3}}{1 + \frac{a}{3}} x^{(1)}(k-1) - \frac{\frac{a}{3}}{1 + \frac{a}{3}} x^{(1)}(k-2) + \frac{b}{1 + \frac{a}{3}} k + \frac{c}{1 + \frac{a}{3}} \quad (4)$$

$$\text{令 } a_1 = \frac{1 - \frac{a}{3}}{1 + \frac{a}{3}}, \quad a_2 = -\frac{\frac{a}{3}}{1 + \frac{a}{3}}, \quad a_3 = \frac{b}{1 + \frac{a}{3}}, \quad a_4 = \frac{c}{1 + \frac{a}{3}}, \text{ 得到离散的 NGM(1,1)模型:}$$

$$\hat{x}^{(1)}(k) = a_1 x^{(1)}(k-1) + a_2 x^{(1)}(k-2) + a_3 k + a_4 \quad (5)$$

由于式(5)是三阶差分方程求解很困难, 所以直接利用递推算法进行求解 $\hat{x}^{(1)}(k)$, 然后利用 $x^{(0)}(k)$ 与 $x^{(1)}(k)$ 和 $x^{(1)}(k-1)$ 关系求 $x^{(0)}(k)$ 的预测值 $\hat{x}^{(0)}(k)$ 。

2.2. 优化三参数背景值 NGM(1,1)模型的初始值

杨孝良等[16]中是将具有三参数背景值的 NGM(1,1)模型的初始值 $\hat{x}^{(1)}(1)$ 和 $\hat{x}^{(1)}(2)$ 视为已知数据, 即令 $\hat{x}^{(1)}(1) = x^{(1)}(1)$, $\hat{x}^{(1)}(2) = x^{(1)}(2)$, 直接给预测值赋原值, 虽然序列前两个值的没有预测误差, 但是初始值不是参与模型运算得出可能会导致后面的预测值出现较大的误差, 所以这里将初始值 $\hat{x}^{(1)}(1)$ 和 $\hat{x}^{(1)}(2)$ 设为未知, 令:

$$\hat{x}^{(1)}(1) = \xi_1 \quad (6)$$

$$\hat{x}^{(1)}(2) = \xi_2 \quad (7)$$

利用式(5)和 $x^{(0)}(k)$ 与 $x^{(1)}(k)$ 和 $x^{(1)}(k-1)$ 关系求预测值 $\hat{x}^{(0)}(k)$, 然后以平均相对误差式(8)最小为目标, 利用粒子群优化算法确定 $\hat{x}^{(1)}(1)$ 和 $\hat{x}^{(1)}(2)$ 的最优值 ξ_1 和 ξ_2 。

$$fitness = \frac{1}{n-1} \sum_{k=1}^n \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \quad (8)$$

2.3. 优化三参数背景值 NGM(1,1)模型的背景值

杨孝良等[16]提出的三参数背景值与传统的 NGM(1,1)模型的背景值构造一样是取的紧邻生成序列的均值, 为了提高模型预测精度, 将紧邻均值生成序列的均值改为未知的加权数, 即将式(3)改为:

$$z^{(1)}(k) = \alpha x^{(1)}(k) + \beta x^{(1)}(k-1) + \gamma x^{(1)}(k-2) \quad (9)$$

并满足 $\alpha + \beta + \gamma = 1$ 。将式(9)和 $x^{(0)}(k)$ 与 $x^{(1)}(k)$ 和 $x^{(1)}(k-1)$ 的关系式代入式(2)得:

$$x^{(1)}(k) = \frac{1-\alpha\beta}{1+a\alpha} x^{(1)}(k-1) - \frac{\gamma}{1+a\alpha} x^{(1)}(k-2) + \frac{b}{1+a\alpha} k + \frac{c}{1+a\alpha} \quad (10)$$

将寻求的最优初始值 ξ_1 和 ξ_2 带入模型中, 再次以平均相对误差最小为目标(式(8)), 通过粒子群优化算法寻求最优的 α 、 β 和 γ 值, 当 $\alpha = \beta = \gamma = \frac{1}{3}$, 背景值变为杨孝良等[16]提出的紧邻三参数均值背景值构造。

2.4. 优化三参数背景值 NGM(1,1)模型的参数估计方法

传统的 NGM(1,1)模型及其他的灰色预测模型都是利用最小二乘法求解模型的灰参数, 即利用残差平方和最小为目标来求灰参数, 即:

$$s = \sum_{k=3}^n (x^{(0)}(k) - \hat{x}^{(0)}(k))^2 \quad (11)$$

式(11)没有考虑到新旧信息数据对于模型的贡献不同, 对于序列长度为 n 的原始数据, 根据数据值发生的时间的不同, 给予不同重要程度, 当数据值发生时间距离当前时刻越远, 则其在建模时所占重要程度就越小, 应该占有更小的比重; 当数据值发生时间离当前时刻越近的数据, 则在建模时的重要程度就越高, 应该占有更大的比重。本文参考何霞等[18]用赋予残差平方和(式(11))权值 w 来反应建模序列值的重要程度, 令:

$$s = \sum_{k=3}^n w(k) (x^{(0)}(k) - \hat{x}^{(0)}(k))^2 \quad (12)$$

其中 $w(k) = \frac{1}{(z^{(1)}(k))^2}$ 。

将式(2)改为

$$x^{(0)}(k) = -az^{(1)}(k) + bk + c \quad (13)$$

代入式(11)得:

$$s = \sum_{k=3}^n w(k) (x^{(0)}(k) + az^{(1)}(k) - bk - c)^2 \quad (14)$$

$$\text{令 } y = \begin{bmatrix} w(3) \cdot x^{(0)}(3) \\ w(4) \cdot x^{(0)}(4) \\ \vdots \\ w(n) \cdot x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} w(3) \cdot z^{(1)}(3) & 3w(3) & w(3) \\ w(4) \cdot z^{(1)}(4) & 4w(4) & w(4) \\ \vdots & \vdots & \vdots \\ w(n) \cdot z^{(1)}(n) & nw(n) & w(n) \end{bmatrix}, \quad d = [a, b, c]^T$$

使 s 最小的灰参数 a 、 b 和 c 应该满足:

$$\begin{cases} \frac{\partial S}{\partial a} = 2 \sum_{k=3}^n w(k) (x^{(0)}(k) + az^{(1)}(k) - bk - c) \cdot w(k) \cdot z^{(1)}(k) = 0 \\ \frac{\partial S}{\partial b} = 2 \sum_{k=3}^n w(k) (x^{(0)}(k) + az^{(1)}(k) - bk - c) \cdot (-k) \cdot w(k) = 0 \\ \frac{\partial S}{\partial c} = 2 \sum_{k=3}^n w(k) (x^{(0)}(k) + az^{(1)}(k) - bk - c) \cdot (-1) \cdot w(k) = 0 \end{cases} \quad (15)$$

根据式(15)解得 $d = [a, b, c]^T =$

$$\begin{pmatrix} \frac{(CE - ZF)H + \left(ZE - \frac{CE^2}{F}\right)I}{2ZEC - \frac{C^2E^2}{F} - Z^2F} \\ \frac{(CE - ZF)G + (AF - C^2)H + (ZC - AE)I}{2ZEC - \frac{C^2E^2}{F} - Z^2F} \\ \frac{\left(ZE - \frac{CE^2}{F}\right)G + (ZC - AE)H + \left(\frac{AE^2}{F} - Z^2\right)I}{2ZEC - \frac{C^2E^2}{F} - Z^2F} \end{pmatrix}$$

其中:

$$A = \sum_{k=3}^n w^2(k) (z^{(1)}(k))^2, \quad Z = \sum_{k=3}^n w^2(k) z^{(1)}(k)k, \quad C = \sum_{k=3}^n w^2(k) z^{(1)}(k), \quad E = \sum_{k=3}^n w^2(k)k, \quad F = \sum_{k=3}^n w^2(k),$$

$$G = \sum_{k=3}^n w^2(k) z^{(1)}(k) x^{(0)}(k), \quad H = \sum_{k=3}^n w^2(k) x^{(0)}(k), \quad I = \sum_{k=3}^n w^2(k) x^{(0)}(k)$$

$$\begin{aligned} B^T B &= \begin{pmatrix} w(3)z^{(1)}(3) & 3w(3) & w(3) \\ w(4)z^{(1)}(4) & 4w(4) & w(4) \\ \vdots & \vdots & \vdots \\ w(n)z^{(1)}(n) & nw(n) & w(n) \end{pmatrix}^T \begin{pmatrix} w(3)z^{(1)}(3) & 3w(3) & w(3) \\ w(4)z^{(1)}(4) & 4w(4) & w(4) \\ \vdots & \vdots & \vdots \\ w(n)z^{(1)}(n) & nw(n) & w(n) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=3}^n w^2(k) (z^{(1)}(k))^2 & \sum_{k=3}^n w^2(k) z^{(1)}(k)k & \sum_{k=3}^n w^2(k) z^{(1)}(k) \\ \sum_{k=3}^n w^2(k) z^{(1)}(k)k & \sum_{k=3}^n w^2(k)k^2 \sum_{k=3}^n w^2(k)k & \sum_{k=3}^n w^2(k)k \\ \vdots & \vdots & \vdots \\ \sum_{k=3}^n w^2(k) z^{(1)}(k) & \sum_{k=3}^n w^2(k)k & \sum_{k=3}^n w^2(k) \end{pmatrix} = \begin{pmatrix} A & Z & C \\ Z & \frac{E^2}{F} & E \\ C & E & F \end{pmatrix} \\ B^T Y &= \begin{pmatrix} w(3)z^{(1)}(3) & 3w(3) & w(3) \\ w(4)z^{(1)}(4) & 4w(4) & w(4) \\ \vdots & \vdots & \vdots \\ w(n)z^{(1)}(n) & nw(n) & w(n) \end{pmatrix}^T \begin{pmatrix} w(3)x^{(0)}(3) \\ w(4)x^{(0)}(4) \\ \vdots \\ w(n)x^{(0)}(n) \end{pmatrix} = \begin{pmatrix} \sum_{k=3}^n w^2(k) z^{(1)}(k) x^{(0)}(k) \\ \sum_{k=3}^n w^2(k) x^{(0)}(k)k \\ \sum_{k=3}^n w^2(k) x^{(0)}(k) \end{pmatrix} = \begin{pmatrix} G \\ H \\ I \end{pmatrix} \end{aligned}$$

$$|B^T B| = 2ZEC - \frac{C^2 E^2}{F} - Z^2 F = \theta$$

由克拉默法则求得:

$$(B^T B)^{-1} = \begin{pmatrix} 0 & \frac{CE-ZF}{\theta} & \frac{ZE-\frac{CE^2}{F}}{\theta} \\ \frac{CE-ZF}{\theta} & \frac{AF-C^2}{\theta} & \frac{ZC-AE}{\theta} \\ \frac{ZE-\frac{CE^2}{F}}{\theta} & \frac{ZC-AE}{\theta} & \frac{\frac{AE^2}{F}-Z^2}{\theta} \end{pmatrix}$$

则:

$$\begin{aligned} (B^T B)^{-1} B^T Y &= \begin{pmatrix} 0 & \frac{CE-ZF}{\theta} & \frac{ZE-\frac{CE^2}{F}}{\theta} \\ \frac{CE-ZF}{\theta} & \frac{AF-C^2}{\theta} & \frac{ZC-AE}{\theta} \\ \frac{ZE-\frac{CE^2}{F}}{\theta} & \frac{ZC-AE}{\theta} & \frac{\frac{AE^2}{F}-Z^2}{\theta} \end{pmatrix} \begin{pmatrix} G \\ H \\ I \end{pmatrix} \\ &= \begin{pmatrix} \frac{(CE-ZF)H + \left(ZE - \frac{CE^2}{F}\right)I}{\theta} \\ \frac{(CE-ZF)G + (AF-C^2)H + (ZC-AE)I}{\theta} \\ \frac{\left(ZE - \frac{CE^2}{F}\right)G + (ZC-AE)H + \left(\frac{AE^2}{F} - Z^2\right)I}{\theta} \end{pmatrix} = (a, b, c)^T \end{aligned}$$

所以利用加权最小二乘法求解模型参数依然满足:

$$d = [a, b, c]^T = (B^T B)^{-1} B^T Y \quad (16)$$

3. NGM(1,1)模型应用

首先, 本节利用的数据实例来验证本文构建的组合优化模型的可行性, 然后, 通过与其他模型进行对比得出本文构建的组合优化模型的优化性, 最后将本文构建的组合优化模型应用于武汉市 2020 年 2 月 6 日~2 月 15 日的新增感染人数的拟合与预测, 来检验本文构建模型的适用性。

3.1. 算例分析

通过文献[16]中的算例分析数据, 分别建立 GM(1,1)模型(模型 1)、DGM(1,1)模型(模型 2)、文献[14]改进初始值模型(模型 3)及本文建立背景值、初始值和参数估计方式组合优化模型(模型 4), 并且计算模型的相对误差(δ)、平均相误差($avg_ \delta$)、方差比(C)和小误差概率(P)来比较模型的拟合性能和精度。

Table 1. Raw data
表 1. 原始数据

年份	2002	2003	2004	2005	2006	2007
数据	21.62	22.83	23.74	23.89	24.50	26.88
年份	2008	2009	2010	2011	2012	
数据	27.66	26.31	32.44	37.67	39.50	

根据表 1 中的原始数据, 计算灰色预测模型模拟值、相对误差和平均相对误差如表 2 所示。其中相对误差(δ)和平均相误差($avg_ \delta$)的计算公式分别为:

$$\delta(k) = \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\% \quad (17)$$

$$avg_ \delta = \frac{\sum_{k=1}^n \delta(k)}{n-1} \quad (18)$$

通过寻优得到最优初始值为 $\xi_1 = 21.5204$, $\xi_1 = 44.3504$, 代入式(10), 通过寻优得到的最优的三参数背景值加权权数为 $\alpha = 0.2365$, $\beta = 0.7635$, $\gamma = 0.1744 \times 10^{-10}$, 最后通过加权最小二乘法(式(16))计算得到灰参数 $a = -0.3904$, $b = -8.8525$, $c = 30.7150$, 从而通过计算得参数 $a_1 = 1.4302$, $a_2 = 7.0521 \times 10^{-12}$, $a_3 = -9.7532$, $a_4 = 33.8402$, 根据式(10)可知, 表 1 数据序列的 NGM(1,1)模型为:

$$x^{(1)}(k) = 1.4302x^{(1)}(k-1) + 7.0521 \times 10^{-12}x^{(1)}(k-2) - 9.7532k + 33.8402 \quad (19)$$

其中 $k = 3, 4, \dots, n$ 。同时计算 GM(1,1)模型(模型 1)、DGM(1,1)模型(模型 2)、文献[14]改进初始值模型(模型 3)及本文建立背景值、初始值和参数估计方式组合优化模型(模型 4)得模拟值、相对误差和平均相对误差来比较模型的模拟性能, 如表 2 所示。

Table 2. Average error and average relative error of simulated values
表 2. 模拟值的平均误差和平均相对误差

原始数据	模型 1		模型 2		模型 3		模型 4	
	模拟值	相对误差(%)	模拟值	相对误差(%)	模拟值	相对误差(%)	模拟值	相对误差(%)
21.62	21.62	0.00	21.62	0.00	21.62	0.00	21.62	0.00
22.83	20.73	9.21	20.76	9.09	20.84	8.72	22.83	0.00
23.74	22.15	6.70	22.18	6.57	22.27	6.19	23.74	0.02
23.89	23.67	0.91	23.70	0.77	23.80	0.38	23.98	0.37
24.5	25.30	3.26	25.33	3.40	25.44	3.82	24.73	0.95
26.88	27.04	0.58	27.07	0.72	27.18	1.12	25.64	4.60
27.66	28.89	4.46	28.93	4.60	29.05	5.02	27.00	2.38
26.31	30.88	17.36	30.92	17.52	31.04	18.00	28.90	9.86
32.44	33.00	1.72	33.04	1.86	33.18	2.27	31.62	2.53
37.67	35.27	6.38	35.31	6.25	35.46	5.88	35.47	5.83
39.5	37.69	4.59	37.74	4.45	37.89	4.07	40.95	3.67
平均相对误差(%)	5.517		5.525		5.547		3.021	

为了进一步比较模型的模拟精度, 计算出四个模型的方差比(C)和小误差概率(P), 首先计算出四个模型得出的模拟值与原始序列(表 1)的残差值, 残差计算公式:

$$\varepsilon_i(k) = x^{(0)}(k) - \hat{x}^{(0)}(k) \quad (20)$$

计算得到残差序列如表 3 所示:

Table 3. Residual

表 3. 残差

模型 1 残差 ε_1	模型 2 残差 ε_2	模型 3 残差 ε_3	模型 4 残差 ε_4
0.00	0.00	0.00	0.00
2.10	1.99	1.99	0.00
1.59	1.47	1.47	0.00
0.22	0.09	0.09	-0.09
-0.80	-0.94	-0.94	-0.23
-0.16	-0.30	-0.30	1.24
-1.23	-1.39	-1.39	0.66
-4.57	-4.73	-4.73	-2.59
-0.56	-0.74	0.75	0.82
2.40	2.21	2.36	2.20

通过表 3 计算得到的残差序列值, 进一步计算原始序列的均值($\bar{x}^{(0)}$)和标准差(s_1)以及三个模型与原始序列的残差序列均值($\bar{\varepsilon}_i$)与标准差(s_{1i})、方差比(C)和小误差概率(P), 计算公式如下:

$$\bar{x}^{(0)} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k) \quad (21)$$

$$s_1 = \sqrt{\frac{\sum_{k=1}^n (x^{(0)}(k) - \bar{x}^{(0)})^2}{n}} \quad (22)$$

$$\bar{\varepsilon}_i = \frac{1}{n} \sum_{i=1}^4 \varepsilon_i \quad (23)$$

$$s_{1i} = \sqrt{\frac{\sum_{i=1}^4 \varepsilon_{1i}^2}{n}} \quad (24)$$

$$C_i = \frac{s_1}{s_{1i}} \quad (25)$$

$$P_i = p(|\varepsilon_i(t) - \bar{\varepsilon}_i| < 0.6745s_{1i}) \quad (26)$$

式(20)~(21)计算表 1 原始数据的均值为 27.91, 标准差为 5.76, 通过式(22)~(26)计算四个模型的平均相对残差(q)、方差比(C)和小误差概率(P)如表 4 所示。小误差概率和方差比对于模型预测精度的评价准则如下:

Table 4. Accuracy evaluation criteria
表 4. 精度评价准则

C 和 P 的取值	精度等级
$C \leq 0.35, P \geq 0.95$	一级
$0.35 < C \leq 0.50, 0.80 \leq P < 0.9$	二级
$0.50 < C \leq 0.65, 0.6 \leq P < 0.8$	三级
$C > 0.65, P < 0.6$	四级

Table 5. Mean relative error, varianceratio, and small error probability test
表 5. 平均相对残差、方差比和小误差概率检验

指标	模型 1	模型 2	模型 3	模型 4
残差均值	-0.10	0.04	-0.23	0.20
标准差	1.90 (s_{11})	1.91 (s_{12})	1.91 (s_{13})	1.18 (s_{14})
C	0.550	0.329	0.329	0.212
P	0.818	1	0.909	1
精度等级	2	1	1	1

通过表 2、表 5 和表 4 的评价准则比较 4 个模型相关指标得出每个模型的性能, 从表 2 可以看出本文所构建初始值、背景值和参数估计方法组合优化的模型(模型 4)的平均相对误差为 3.021%, 都小于 GM(1,1)模型(模型 1)、将作为初始值的改进 GM(1,1)模型(模型 2)和文献[14]改进初始值模型(模型 3), 通过表 4 的预测精度评价准则以及表 5 计算得出的指标值比较 4 个模型的方差比(C)和小误差概率(P), 得出本文所构建模型(模型 4)的模拟精度等级、将作为初始值的改进 GM(1,1)模型(模型 2)和文献[14]改进初始值模型(模型 3)的模拟精度等级最高, 且本文所构建模型(模型 4)的模拟精度最高, GM(1,1)模型(模型 1)的模拟精度等级为 2 最差。通过该算例得出本文将初始值通过文献[16]将初始值设为未知变量 ξ_1 和 ξ_2 , 以平均相对误差最小为目标利用粒子群优化算法求最优初始值的取值, 将三参数的紧邻均值背景值改为加权的三参数背景值, 利用粒子群优化算法寻求最优的加权权数, 最后将参数估计方法最小二乘法改为加权最小二乘法求模型的灰参数, 构成 NGM(1,1)模型初始值的选取、背景值的构造和参数估计方法的组合优化一定程度上使得模型精度提高, 因此本文所构建组合优化模型优于其他三种模型。

3.2. 模型应用

下面通过将模型应用于武汉市新型冠状病毒肺炎新增感染人数的预测来分析本文的组合优化模型的效果, 将武汉市 2020 年 2 月 6 日~2020 年 2 月 15 日(封城阶段, 假设不存在人口流动)作为原始数据, 比较传统的 GM(1,1)模型、文献[16]模型、文献[14]模型和本文组合优化模型平均相对误差来比较模型的预测精度, 如表 6 所示。通过寻优得到最优初始值为 $\xi_1 = 1686$, $\xi_2 = 3671$, 代入式(10), 通过寻优得到的最优的三参数背景值加权权数为 $\alpha = 0.4910$, $\beta = 6.1000 \times 10^{-14}$, $\gamma = 0.5090$, 最后通过加权最小二乘法(式(16))计算得到灰参数 $a = -0.8661$, $b = -1.3818 \times 10^3$, $c = 2816.5495$, 从而通过计算得参数 $a_1 = 1.7399$, $a_2 = 0.7670$, $a_3 = -2.4042 \times 10^3$, $a_4 = 4.9004 \times 10^3$, 根据式(10)可知, 表 1 数据序列的 NGM(1,1)模型为:

$$x^{(1)}(k) = 1.7399x^{(1)}(k-1) + 0.7670x^{(1)}(k-2) - 2.4042 \times 10^3 k + 4.9004 \times 10^3 \quad (27)$$

其中 $k = 3, 4, \dots, n$, 通过式(27)计算出本文组合优化模型的 $x^{(1)}(k)$ 的模拟值, 之后根据 $x^{(1)}(k)$ 、 $x^{(1)}(k-1)$

与 $x^{(0)}(k)$ 的关系, 计算出原始数据序列的模拟值, 根据模拟值与原始数据本文组合优化模型的相对误差和平均相对误差。同时计算 GM(1,1)模型、文献[16]模型、文献[14]模型的模拟值、相对误差和平均相对误差。

从表 6 可以看出本文组合优化模型的平均相对误差小于 GM(1,1)模型, 且远小于文献[16]和文献[14]模型的平均相对误差, 且本文基于的文献[16]的三参数紧邻均值背景值 NGM(1,1)模型平均相对误差最大, 这也说明本文在此基础上对初始值、背景值和参数估计的组合优化的有效性。为了更加直观的比较 GM(1,1)模型、文献[16]、文献[14]和本文组合优化模型对于武汉市 2020 年 2 月 6 日~2020 年 2 月 15 日的新增感染肺炎人数的模拟情况, 本文绘制 GM(1,1)模型、文献[16]、文献[14]和本文组合优化模型的相对误差折线图, 如图 1 所示:

Table 6. Average error and average relative error
表 6. 平均误差和平均相对误差

	GM(1,1)模型		文献[16]模型		文献[14]模型		本文组合优化模型	
	模拟值	相对误差	模拟值	相对误差	模拟值	相对误差	模拟值	相对误差
1501	1501.00	0.00	1501.00	0.00	1501.00	0.00	1686.00	0.12
1985	1935.67	0.02	1985.00	0.00	2526.53	0.27	1985.00	0.00
1379	1832.85	0.33	693.01	0.50	2669.51	0.94	1233.19	0.11
1921	1735.48	0.10	2363.53	0.23	2820.58	0.47	1727.56	0.10
1552	1643.29	0.06	3671.80	1.37	2980.19	0.92	1547.32	0.00
1104	1555.99	0.41	4222.18	2.82	3148.85	1.85	1612.88	0.46
13436	1473.33	0.89	4346.67	0.68	3327.04	0.75	1588.73	0.88
3910	1395.07	0.64	4329.85	0.11	3515.32	0.10	1596.98	0.59
1923	1320.96	0.31	4296.97	1.23	3714.26	0.93	1592.81	0.17
1548	1250.78	0.19	4278.69	1.76	3924.45	1.54	1591.88	0.03
平均相对误差	0.330		0.967		0.863		0.274	

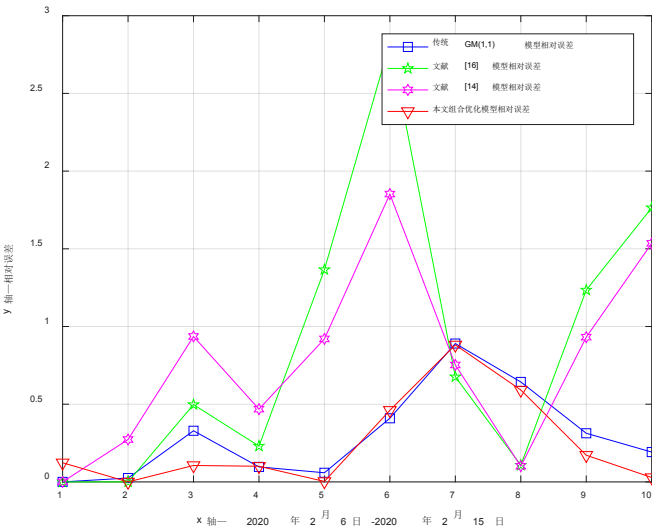


Figure 1. Relative error line chart
图 1. 相对误差折线图

从图 1 可以更加清楚地看出本文构建的组合优化模型优于 GM(1,1)模型, 且远优于文献[16]和文献[14]模型。这一结果表明本文构建的组合优化模型具有合理性和有效性。

4. 结论

本文构建对于近似非齐次指数数据模拟效果好的 NGM(1,1)预测模型, 考虑到灰色模型的背景值容易受到极端值的影响, 所以为了减少原始数据中极端值对于灰色预测模型模拟性能的影响, 达到提高灰色预测模型模拟性能稳定性的目的, 本文参考文献[16]构建 NGM(1,1)模型的三参数紧邻均值背景值, 基于此, 依次对三参数紧邻均值背景值的 NGM(1,1)模型的初始值、背景值和参数估计方式进行优化。首先, 以平均相对误差最小为目标利用粒子群优化算法求最优初始值的取值, 然后将三参数的紧邻均值背景值改为加权的三参数背景值, 利用粒子群优化算法寻求最优的加权权数, 最后将参数估计方法最小二乘法改为加权最小二乘法求模型的灰参数, 实现 NGM(1,1)模型初始值、背景值和参数估计方式的协同优化。通过算例分析, 验证了本文构建的初始值、背景值和参数估计方式组合优化模型模拟性能优于经典 GM(1,1)模型、DGM(1,1)模型和文献[14]的初始值优化模型。最后, 将本文的组合优化模型应用于武汉市 2020 年 2 月 6 日~2020 年 2 月 15 日新增感染人数的模拟, 模拟值的平均相对误差为 0.274 低于其余三个灰色预测模型的平均相对误差, 再次验证了本文构建的组合优化模型的有效性和合理性。

基金项目

基金项目: 国家自然科学基金青年项目(71701066); 上海理工大学管理学院 2021 年度科研启动项目(KYQD202101)资助课题。

参考文献

- [1] 莫定源. 基于 GM(1,1)模型对广西人均地区生产总值的预测[J]. 价值工程, 2022, 41(23): 17-19.
- [2] 项莹莹, 耿献辉, 彭世广. 中国玉米产量与消费量预测研究——基于改进 GM(1,1)模型[J]. 数学的实践与认识, 2022, 52(4): 39-47.
- [3] 杨鑫波, 龙勇, 曾波. 季度性用电需求预测的移动平均 GM(1,1)模型[J]. 统计与决策, 2021, 37(9): 168-171. <https://doi.org/10.13546/j.cnki.tjyjc.2021.09.040>
- [4] 朱晓宵, 刘明, 曹杰. 基于滚动灰色 GM(1,1)修正的新冠肺炎疫情演化情景重建[J/OL]. 运筹与管理: 1-8. <http://kns.cnki.net/kcms/detail/34.1133.G3.20220905.1557.006.html>, 2023-04-12.
- [5] 吕燕梅, 王苏林. 基于离散灰色模型的铁路货运量预测仿真[J]. 计算机仿真, 2022, 39(5): 150-153+190.
- [6] 张可, 马成文, 丰景春, 薛松. 基于离散灰色模型的农村水环境政策减排效应及其空间分异性研究[J]. 中国管理科学, 2017, 25(5): 157-166. <https://doi.org/10.16381/j.cnki.issn1003-207x.2017.05.019>
- [7] 张和平, 陈齐海. 基于灰色马尔可夫模型的网络舆情预测研究[J]. 情报科学, 2018, 36(1): 75-79. <https://doi.org/10.13833/j.issn.1007-7634.2018.01.013>
- [8] 邱慧, 胡运红, 冯青梅. 基于新陈代谢和 Verhulst 灰色模型的山西省人口老龄化趋势预测[J]. 数学的实践与认识, 2015, 45(13): 192-197.
- [9] 石季雨, 余思瑶, 易叶青. 优化 NGM(1,1,k)模型在国内能源消费中的应用[J]. 吉首大学学报(自然科学版), 2022, 43(4): 27-33. <https://doi.org/10.13438/j.cnki.jdzk.2022.04.006>
- [10] Zhao, Y.-F., Shou, M.-H. and Wang, Z.-X. (2020) Prediction of the Number of Patients Infected with COVID-19 Based on Rolling Grey Verhulst Models. *International Journal of Environmental Research and Public Health*, 17, 4582.
- [11] 贾沛, 张健, 王方, 刘小杰. 基于新陈代谢灰色线性回归组合模型的我国技术市场发展趋势预测[J]. 数学的实践与认识, 2021, 51(22): 10-18.
- [12] 熊华平, 李鹏辉. 基于新陈代谢 GM(1,1)模型的湖北省城市生活垃圾清运量预测[J]. 数学的实践与认识, 2019, 49(1): 9-14.
- [13] 丁松, 李若瑾, 党耀国. 基于初始条件优化的 GM(1,1)幂模型及其应用[J]. 中国管理科学, 2020, 28(1): 153-161. <https://doi.org/10.16381/j.cnki.issn1003-207x.2020.01.013>

- [14] 郑坚, 陈斌. 基于时间权重序列的 GM(1,1)初始条件优化模型[J]. 控制与决策, 2018, 33(3): 529-534.
<https://doi.org/10.13195/j.kzyjc.2017.0033>
- [15] 卢捷, 李峰. 基于初始值和背景值改进的 GM(1,1)模型优化与应用[J]. 运筹与管理, 2020, 29(9): 27-33.
- [16] 杨孝良, 周猛, 曾波. 灰色预测模型背景值构造的新方法[J]. 统计与决策, 2018, 34(19): 14-18.
<https://doi.org/10.13546/j.cnki.tjyjc.2018.19.003>
- [17] 刘震, 党耀国, 魏龙. NGM(1,1,k)模型的背景值及时间响应函数优化[J]. 控制与决策, 2016, 31(12): 2225-2231.
<https://doi.org/10.13195/j.kzyjc.2015.1370>
- [18] 何霞. 灰色 GM(1,1)模型参数估计的加权最小二乘方法[J]. 运筹与管理, 2012, 21(6): 23-27.
- [19] 周德强. 估计灰色 Verhulst 模型参数的 LS-SVM 方法及应用[J]. 中国管理科学, 2022, 30(3): 280-286.
<https://doi.org/10.16381/j.cnki.issn1003-207x.2017.1233>
- [20] 刘思峰. 灰色系统理论及其应用[M]. 北京: 科学出版社, 2021.