Application of non-equidistant gray model based on optimization of background value in settlement prediction

Jiameng Wei 1,a, Huagen Jiang2ab and Jianpeng Diao3,c

1,2,3Surveying Engineering, Southwest Forestry University, kunming, Yunnan Province, 650000, China
*aEmail: wei14056939*47@126.com

Abstract. In order to solve the influence of different original data series on model adaptability and fitting accuracy in non-equidistant grey model. Because the background value in the gray model is a key factor affecting the accuracy and adaptability of the model. Faced with this situation, this article uses the Newton-Cotes quadrature formula to optimize the background value, which improves the adaptability and fitting accuracy of the non-equidistant gray model. By constructing the model, the Newton-Cotes quadrature formula is applied to the application of the non-equidistant GM (1,1) modeling method in the settlement and deformation of buildings. The analysis shows that the new model method prediction is more suitable for building settlement monitoring and analysis, and the degree of fit is improved by about 30%. The new model is better than the traditional model.

1. Introduction
The settlement phenomenon is common during the construction and operation of the building. As the progress of the project develops, the building will inevitably settle as the load increases. To ensure the smooth progress of the work, the deformation of the building must be monitored. Dynamic buildings carry out static and periodic measurements, and through data analysis, safety monitoring and forecasting of buildings [1].

The application of gray model in the settlement prediction of buildings has achieved great results, but the research on improving the accuracy of model fitting has never stopped. Ji Peirong, Hu Xiangyong, through the study of the error characteristics of the gray model, shows that the establishment of the gray model is essentially an exponential function [5]. Zhang Dahai, Jiang Shifang and others pointed out the theoretical flaws in the model and changed the choice of initial values. [3]. Xiao Wen and Fan Zhiping gave the method and grade index of gray model prediction accuracy evaluation [4]. Some institutions and scholars through the optimization research of the model, found that the background value optimization can improve the practicality of the model, expand the applicability of the model, and achieve good results [5][6].

For some sequences with high exponential growth, the model often has a lag, and the background value in the gray model is a key factor affecting the accuracy and adaptability of the model. Therefore, by reconstructing the background value, it is necessary to find a sequence that can adapt to both high growth and low growth [7]. Li Kai et al. proposed a combined interpolation method based on Simpson's formula and Newton's interpolation formula to construct a new background value, and applied the first data to the modeling process through the coordination coefficient [8]. Common background value optimization methods include integration method, interval area and background value optimal generation coefficient method. By reconstructing the background value method and using software to
simulate and predict the building, this paper verifies whether the background value optimization can improve the fitting accuracy based on the new method of non-equidistant model prediction, which is more suitable for data analysis in settlement monitoring and prediction.

2. research method

2.1. Grey theory model

Let the original data sequence be \( x^0 = \{x^0(t_1), x^0(t_2), \ldots, x^0(t_i)\} \), among them \( \Delta t_i = t_i - t_{i-1} \neq \text{const} \).

Calculate the time interval between each observation period and the first observation:

\[ t_i = T_i - T_1, i = 1, 2, 3, \ldots, n. \]  

Average time interval:

\[ \Delta t_0 = \frac{T_n - T_1}{n - 1}, z^{(i)}(i) = \frac{1}{2} A^i(t_i) + A^i(t_{i+1}) \]  

The coefficient of the unit time difference between the time interval \( T_i \) and the average time interval \( \Delta t \) in each period:

\[ u(t_i) = \frac{t_i - (i - 1) \Delta t_0}{\Delta t_0} \]  

Then the total difference in each period is:

\[ \Delta x^{(i)}(t_i) = u(t_i)\left\{ \Delta x^{(i)}(t_i) - x^{(i)}(t_{i-1}) \right\} \]  

Calculate the gray value at equal intervals:

\[ A^0(t_i) = x^0(t_i) - \Delta x^{(0)}(t_i) \]  

Based on this, an evenly spaced data sequence can be established:

\[ A^0 = \{A^0(t_1), A^0(t_2), \ldots, A^0(t_n)\} \]  

The above formula is cumulatively generated:

\[ A^1 = \{A^1(t_1), A^1(t_2), \ldots, A^1(t_n)\} \]  

Fit \( A^0 \) to a first-order linear differential equation:

\[ \frac{dA^0}{dt} + A^1 = u \]  

The immediate mean of the sequence is:

\[ z^{(i)}(i) = \frac{1}{2} A^1(t_i) + A^1(t_{i+1}) \]  

Use least squares to solve the undetermined parameters and get the equally spaced time response function:

\[ A(k + 1) = [A^0(1) - \frac{u}{a}]e^{-ak} + \frac{u}{a}, k = 1, 2, 3, \ldots, n. \]  

Revert the above to a function related to time \( t \) in a non-equal interval sequence:
\[ x^{(1)}(t) = \left[ x^{(0)}(1) - \frac{u}{a} \right] e^{-\frac{ut}{a\Delta t_0}} + \frac{u}{a} \]  
\[ x^{(0)}(t) = x^{(1)}(t) - x^{(1)}(t - \Delta t_0) \]  

(11)  
(12)

Where \( t \) is the time interval from the first cycle.

2.2. Parameter solving based on Newton-Cotes background value optimization

It can be seen from the conventional background value formula, the \( z^{(1)}(i) \) expression is essentially calculated by the trapezoidal formula. However, when the slope of the curve increases, replacing the trapezoidal formula as the background value will increase the error [9]. As shown in the error source diagram (1), the shaded part in the figure is the error part. The Newton-Cotes integration formula is a quadrature formula constructed by taking the equal points of the integration interval as the integration node. Studies have shown that the cotes coefficient of Newton-Cotes is independent of the interval, only depends on \( n \) and \( k \), and has nothing to do with the interval \( f(x) \) and \([a, b]\) [10]. That is, the construction of the integrand is completed without being affected by the original sequence. Suppose the integration interval \([a, b]\) is divided into \( n \) equal parts, and each node is \( x_k = a + kh \), \( k = 1, 2, 3, \ldots, n \). The integral formula is:

\[ I_n(f) = \int_a^b f(x)dx \approx (b - a)\sum_{k=0}^{n} C_k^{(n)} f(x_k) \]  
\[ C_k^{(n)} = \frac{1}{n} \frac{n!}{k!(n-k)!} \int_0^1 \prod_{i\neq k}^n (t-i)dt \]  

among them \( C_k^{(n)} \) is Cotes coefficient. According to the equidistant gray model, the gray differential equation is obtained:

\[ x^{(0)} + ax^{(1)}(k) = u \]  

(13)

(14)

After deforming the whitening equation, we can get:

\[ dx^{(1)}(t) + ax^{(1)}(t)dt = udt \]  

(15)

Integrate the above formula in the interval \([k-2, k+2]\) to get:

\[ x^{(1)}(k + 2) - x^{(1)}(k - 2) + a\int_{k-2}^{k+2} x^{(1)}(t)dt = 4u \]  

(16)

Use the Newton-Cotes fourth-order quadrature formula to solve the integral term on the interval \([k-2, k+2]:\)

\[ \int_{k-2}^{k+2} x^{(1)}(t)dt = \frac{4}{90} \left( 7x^{(1)}(k - 2) + 32x^{(1)}(k - 1) + 12x^{(1)}(k) + 32x^{(1)}(k + 1) + 7x^{(1)}(k + 2) \right) \]  

(17)

Finished up:

\[ x^{(1)}(k + 2) - x^{(1)}(k - 2) + a\frac{4}{90} \left( 7x^{(1)}(k - 2) + 32x^{(1)}(k - 1) + 12x^{(1)}(k) + 32x^{(1)}(k + 1) + 7x^{(1)}(k + 2) \right) = 4u \]  

(18)

Arrange the above according to the grey differential equation to get:

\[ \frac{(x^{(0)}(k + 2) + x^{(0)}(k + 1) + x^{(0)}(k) + x^{(0)}(k - 1))}{4} + a\frac{1}{90} \left( 7x^{(0)}(k - 2) + 32x^{(0)}(k - 1) + 12x^{(0)}(k) + 32x^{(0)}(k + 1) + 7x^{(0)}(k + 2) \right) = u \]  

(19)

The final gray model background value is:
\[ z^{(i)}(k) = \frac{1}{90}(7x^{(i)}(k-2) + 32x^{(i)}(k-1) + 12x^{(i)}(k) + 32x^{(i)}(k+1) + 7x^{(i)}(k+2)) \]  
\[(20)\]

Parameter solving:
\[ (a,u)^T = (B^TB)^{-1}B^TY \]
\[(21)\]

among them

\[
B = \begin{bmatrix}
-\frac{1}{90}(7x^{(i)}(2) + 32x^{(i)}(3) + 12x^{(i)}(4) + 32x^{(i)}(5) + 7x^{(i)}(6)) & 1 \\
-\frac{1}{90}(7x^{(i)}(2) + 32x^{(i)}(3) + 12x^{(i)}(4) + 32x^{(i)}(5) + 7x^{(i)}(6)) & 1 \\
\vdots & \vdots \\
-\frac{1}{90}(7x^{(i)}(n-4) + 32x^{(i)}(n-3) + 12x^{(i)}(n-2) + 32x^{(i)}(n-1) + 7x^{(i)}(n)) & 1 \\
\end{bmatrix}
\]

\[
Y = \begin{bmatrix}
\frac{1}{4}(x^{(0)}(2) + x^{(0)}(3) + x^{(0)}(4) + x^{(0)}(5)) & 1 \\
\frac{1}{4}(x^{(0)}(3) + x^{(0)}(4) + x^{(0)}(5) + x^{(0)}(6)) & 1 \\
\vdots & \vdots \\
\frac{1}{4}(x^{(0)}(n-3) + x^{(0)}(n-2) + x^{(0)}(n-1) + x^{(0)}(n)) & 1 \\
\end{bmatrix}
\]

2.3. Model accuracy check

In order to ensure the prediction accuracy and scientificity of the gray model, the model must be tested to ensure the reliability of the model. Generally, three methods are used to test the accuracy and feasibility of the model: correlation test method, relative error test method, and posterior error test method. Here we use the posterior difference test method for evaluation, which is described by the posterior difference ratio \(c\) and the probability of small error \(P\) [11].

Model residual calculation:
\[ e(k) = x^{(0)}(k) - x^{(0)}(k) \]
\[(22)\]

The original sequence variance value:
\[ S_i^2 = \frac{1}{n} \sum_{k=1}^{n} (x^{(0)}(k) - \bar{x})^2 \] (23)

\[ \bar{x} = \frac{1}{n} \sum_{k=1}^{n} x^{(0)}(k) \]

among them \[ \bar{x} \].

Variance value of residual sequence:

\[ S_2^2 = \frac{1}{n} \sum_{k=1}^{n} (e(k) - \bar{e})^2 \] (24)

\[ \bar{e} = \frac{1}{n} \sum_{k=1}^{n} e(k) \]

among them \[ \bar{e} \].

The posterior difference ratio \( c \) and the small error probability \( P \):

\[ c = \frac{S_2}{S_1} \] (25)

\[ p = p \{ |e(k) - \bar{e}| < 0.6745S_1 \} \] (26)

Compare the posterior difference ratio and small probability error obtained above with the gray model accuracy evaluation standard Table 1 to determine whether it meets the requirements.

| Table 1 Accuracy assessment table |
|-----------------------------------|
| Forecast accuracy level | Posterior variance ratio | Small error probability value P |
| First level (good) | \( c < 0.35 \) | 0.95 \( \leq p \) |
| Level 2 (Qualified) | 0.35 \( \leq c < 0.5 \) | 0.8 \( \leq p < 0.95 \) |
| Level 3 (barely) | 0.5 \( \leq c < 0.65 \) | 0.7 \( \leq p < 0.8 \) |
| Level 4 (unqualified) | \( c \geq 0.65 \) | \( p < 0.7 \) |

3. Case Analysis

Take the observation results of an affordable housing in Foshan City as an example in Reference [12], as shown in Table 2 for the original data table. The data is from August 10, 2016 to November 21 of the same year, a total of eight periods of data. Among them, the first six periods are modeling and fitting, and the latter two periods are predicted as follows:

| Table 2 Original data |
|-----------------------|
| Observation date | 8-10 | 9-1 | 10-8 | 10-18 | 10-24 | 10-29 | 11-8 | 11-20 |
| Observation period t | 0 | 22 | 59 | 69 | 75 | 80 | 91 | 103 |
| Cumulative settlement \( (\text{mm}) \) | 1.0510 | 1.5497 | 2.1226 | 2.1575 | 2.2038 | 2.2319 | 2.2653 | 2.2653 |

The modeling method of this article is as follows:

The first step is to generate an equally spaced sequence:

\[ A^0 = [1.051, 1.3627, 1.1558, 2.1117, 2.172, 2.2319] ; \]

The second step is to generate an accumulation of the equally spaced sequence \( A^0 \):

\[ A^1 = [1.0510, 2.4137, 3.569, 5.6812, 7.8532, 10.085] ; \]

The third step is to generate \( z(1) = [4.047, 5.791] \) on the adjacent mean value of \( A^1 \);
The fourth step is to solve the parameters using the least square method:
a=0.12457, u=1.1964;
In the fifth step, the time response equation is:
\[ x^{(1)}(t) = 10.6551 \times e^{0.12457t} - 9.6041 \]
The sixth step, the prediction results are displayed, the simulated prediction diagram is shown in Figure 2; the residual analysis diagram is shown in Figure 3; the simulation prediction results table is shown in Table 3;

![Figure 2 Simulation prediction](image)

![Figure 3 Residual analysis chart](image)

| Observation time t | Observation on x | Analog value (mm) | Absolute residual value (mm) | Absolute relative error % | Analog value (mm) | Analog value (mm) | Absolute residual value % |
|--------------------|------------------|-------------------|-----------------------------|----------------------------|-------------------|-------------------|---------------------------|
| 1                  |                  |                   |                             |                            |                   |                   |                           |
| 2                  |                  |                   |                             |                            |                   |                   |                           |
| 3                  |                  |                   |                             |                            |                   |                   |                           |
| 4                  |                  |                   |                             |                            |                   |                   |                           |
| 5                  |                  |                   |                             |                            |                   |                   |                           |
| 6                  |                  |                   |                             |                            |                   |                   |                           |
| 7                  |                  |                   |                             |                            |                   |                   |                           |
| 8                  |                  |                   |                             |                            |                   |                   |                           |
After calculation: the model posterior difference ratio c is 0.085, the probability of small error is 1, and the model prediction level is level 1, which fully meets the accuracy requirements. It can be seen from Fig. 2 that the curve fitted by NC is closer to the curve formed by the true value than the reference, indicating that the fitting effect is better. This model has improved the prediction accuracy to a certain extent. From Figure 3, it can be seen that 83% of the residual error distribution of NC is within millimeters, and 50% of the residual error distribution of reference documents is within millimeters. By comparison, the prediction accuracy can be increased by about 35%. That is to say, the gray model background value optimization based on the Newton-Cotes formula can be applied to the settlement and prediction of buildings, and the accuracy can be improved on the basis of the conventional non-equal interval prediction model, and the feasibility is higher.

4. Conclusion
According to the background value and initial conditions of the two influencing factors that mainly affect the accuracy of GM(1,1) in the gray model prediction simulation, this article focuses on the background value problem, starting from the geometric meaning of the background value, according to the number between the background value and the development coefficient Relations, using the background value optimization based on Newton-Cotes to improve the simulation effect; and verify the reliability and practicability of the method in building settlement prediction.

References
[1] Wu Jianghuai. W. J. H. The analysis and processing of building static deformation monitoring data [D]. Tongji University, 2007.Podani, J. (1994) Multivariate Data Analysis in Ecology and Systematics. SPB Publishing, The Hague.
[2] Huang Weisong, H. W. S Ji Pei rong, J.P.R. Hu Xiangyong. H.X.Y. Experimental research on the error characteristics of gray GM (1,1) model[J]. Journal of Wuhan University of Hydraulic and Electric Power (Yichang), 2000, 22(01): 69-72.
[3] Zhang dahai, Z. D. H. Jiang shifang. J.S.F. Shi kaiquna. S.K.Q. Theoretical Defect of Grey Prediction Formula and Its Improvement[J]. Systems Engineering —Theory & Practice,2002,22(8):140-142.
[4] Xiaowen, X. W.  Fan zhi ping, F. Z. P. Cai renlan, C.R.L. et al. Application of Grey System GM (1,1) Model in Building’s Subsidence Monitoring[J]. Geospatial Information, 2010, 8(2):148-150.
[5] Luo dang, L. D. Liu si feng, L.S.F. Dangyaoguo. D. Y. G. The Optimization of Grey Model GM (1,1) [J]. Engineering Science, 2003, 5(8):50-53.
[6] Libo, L. B. Jiangyan, J. Y. Yangxiuwen. Y. X. W. MGM (1, n) Model of Optimized Background [J]. Journal of Southwest Agricultural University, 2013, 35(1):89-94.

[7] TanGuan-jun. T. G. J. The Structure Method and Application of Background Value in Grey System GM (1,1) Model (I) [J]. Systems Engineering—Theory & Practice, 2000, 20(4):98-103.

[8] Li kai, L. K. Zhang Tao. Z. T. Improvement of background value of GM (1,1) model based on combination interpolation [J]. Computer Application Research, 2018, 35(10):2994-2999.

[9] Huaiyin Institute of Technology. GM (1,1) model prediction method based on Newton-Cortez formula to construct background values: China, 107811180 A [P], 2018.04.03.

[10] Xu Xinxin, X.X.X. Su Huayou, S. H. Y. Zhang Chunping, Z. C. P. et al. Application of optimized non-equidistant GM (1,1) model in prediction of surface settlement around foundation pits [J]. Mathematics in Practice and Theory, 2014(01):93-98.

[11] Peng Guangliang. P. G. L. Research on data processing and analysis method of deformation monitoring [D]. Liaoning Technical University, 2007.

[12] Wei Hao, W. H. Wei Ke, W. K. Chen Shan. Application of non-equidistant GM (1,1) model in building settlement prediction [J]. Science and Technology Guide, 2017(11):75-76.