SCIPEDIA

Numerical Simulation and Optimization of Grey Relational Analysis Models for Panel Data

Xiaomei Liu[1], Fanghong Jian[1,*] and Xiaozhong Tang[2]

[1] College of Science, Jiujiang University, Jiujiang, 332005, China

[2] College of Economics and Management, Huangshan University, Huangshan, 245021, China

INFORMATION

Keywords:
Grey incidence degree
grey relational analysis model
panel data
mean value theory

DOI: 10.23967/j.rimni.2025.10.62052

SCIPEDIA Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

RIMNI

Numerical Simulation and Optimization of Grey Relational Analysis Models for Panel Data

Xiaomei Liu[1], **Fanghong Jian**[1,*] **and Xiaozhong Tang**[2]

[1]College of Science, Jiujiang University, Jiujiang, 332005, China

[2]College of Economics and Management, Huangshan University, Huangshan, 245021, China

ABSTRACT

The indicators-coupled grey relational analysis (ICGRA) models are important in clustering panel data with cross-sectional dependence. However, there is still little research on performance validation for the various ICGRA models. In this paper, we investigate the performance of the existing ICGRA models accounting for the reordering of indicators. Firstly, the robot execution failures (REF) dataset of the University of California Irvine (UCI) machine learning database is adopted to validate the robustness of four traditional ICGRA models. Then, we compared the grey relational orders for all arrangements of indicators in panel data. Simulation experiments showed that the four ICGRA models are not all robust against the grey relational order. To resolve this problem, we adopted the mean value theory and deep modeling to optimize the four models and compared them with the tetrahedral grey relational analysis (GRA) model that considers the coupling effect between indicators on the grey relational order, as well as with the k-nearest neighbor (KNN) algorithm. Results show that the classification accuracy of the averaged absolute GRA model was 97.73%, the other optimized ICGRA models and the k-nearest neighbor (KNN) method all achieved 100% accuracy, while the tetrahedral GRA model has an accuracy of 83.33%. Therefore, the average grey incidence degree for all arrangements of indicators and deep modeling significantly improves the stability of models and enhances the clustering accuracy in different cases.

OPEN ACCESS

Received: 09/12/2024

Accepted: 04/03/2025

Published: 30/06/2025

DOI
10.23967/j.rimni.2025.10.62052

Keywords:
Grey incidence degree
grey relational analysis model
panel data
mean value theory

1 Introduction

Panel data has been a focus of research, and grey relational analysis is one of the methods available for its clustering analysis [1]. Unlike the factor analysis of statistics inference, grey relational analysis is easy to implement and does not need a large number of samples. Therefore, grey relational analysis has been widely applied, such as ecological environment assessment [2], agricultural production assessment [3], and medical and health assessment [4,5].

*Correspondence: Fanghong Jian (jfhrecoba@126.com).

Grey relational analysis is an important branch of grey system theory, first proposed by Professor Deng [6]. It measures the likeness or proximity of subjects by computing the degree of incidence between them. To adapt to different panel data, various grey relational analysis (GRA) models have been presented. These GRA models for panel data can be classified into two categories: one is the indicators-independent models, which ignore the correlation between cross-sectional data, such as three-dimensional Deng's GRA model [1], modified Deng's GRA model for panel data [7], three-dimensional norm GRA model [8], three-dimensional B-type GRA model [9] and generalized matrix GRA model [10]; another is the indicators-coupled models, which contains the coupling effect between cross-sectional data, such as three-dimensional convex GRA model [11], grid GRA model [12], modified three-dimensional absolute GRA model [13], curvature GRA model [14], tetrahedral GRA model [15] and GRA model considering factor coupling relationship [16].

The cross-sectional data of the panel data may be independent among some indicators [17], but they may also be dependent due to some special characteristics of the indicators [18,19]. For the panel data with cross-sectional independence, the indicators-independent GRA models are applicable. However, for these panel data with cross-sectional dependence, the indicators-coupled GRA (ICGRA) models may be more reasonable. Generally, it cannot be simply assumed that all cross-sectional data are completely independent, and the potential correlations should be considered in the GRA modeling process. Therefore, the development and improvement of ICGRA models for panel data is particularly important.

However, the traditional ICGRA models for panel data mentioned above ignored two details. One is the inconsistency of the grey relational order that resulted from the reordering of indicators in the panel data, and the other is the inadequate validation of the models. Some researchers have discussed the inconsistency of the grey relational order in ICGRA models for panel data, such as permutating the indicators of panel data, and have found that the grey relational order has changed [10,20]. In addition, recent studies found that, even if the reordering of indicators in panel data is considered to modeling as a factor, the validation of models is performed only based on simple and low-order sample data [9].

REF dataset is an important component of the UCI machine learning database and represents the small-scale multivariate sample data, which plays an important role in testing the performance of clustering models and greatly improves the generalization potential of models in complex dynamic environments [20–22]. Meanwhile, Liu et al. [13], and Liu et al. [23] have successfully used the first sub dataset LP1 to validate the performance of the modified three-dimensional absolute GRA model. However, they did not pay attention to the inconsistency of grey relational order caused by the reordering of indicators in panel data. On the other hand, Wu et al. [14] proposed a tetrahedral GRA model, which is an ICGRA model, and discussed the inconsistency of grey relational order caused by the reordering of indicators. Unfortunately, they do not provide adequate validation for the reliability and effectiveness of the model.

Given these, this paper will use small-scale multivariate sample data to simulate and analyze the performance of existing ICGRA models for panel data and attempt to optimize and improve these models. Specifically, the REF dataset is used for testing, to compare and discuss the changes in the grey relational order after reordering the indicators in panel data, and aims to obtain more reliable clustering accuracy. The main work of this paper lies in the following aspects:

(1) Perform a numerical simulation of existing ICGRA models for panel data using the UCI machine learning database.

(2) Optimize existing ICGRA models for panel data to improve clustering accuracy.

(3) Reveal some principles for integrating indicators of panel data in grey relational analysis modeling.

The framework of the paper is presented in the following sections. Section 2 reviews the existing ICGRA models for panel data. Section 3 performs a numerical simulation for the ICGRA models. Section 4 optimizes the existing ICGRA models based on the testing results. Section 5 compares the optimization form of existing ICGRA models with the tetrahedral GRA model and the KNN method. Section 6 makes conclusions.

2 Review of Indicators-Coupled Grey Relational Analysis Models for Panel Data

In this section, we review five traditional ICGRA models for panel data, including the convex GRA model, grid GRA model, modified absolute GRA model, curvature GRA model and the tetrahedral GRA model. These models have presented diverse approaches in the modeling mechanisms, and have been extensively utilized in practical applications.

Firstly, we suppose that $X_p = (a_{ij})_{m\times n}$ represents the system feature behaviors, $X_q = (b_{ij})_{m\times n}$ represents the factor interaction matrix, $1 \le i \le m$ is index dimension, $1 \le j \le n$ is time dimension.

2.1 Convex GRA Model

Convex GRA model [11] mainly utilizes the second-order difference of panel data and the positive semi-definite of Hessian matrix to calculate the three-dimensional grey convex incidence degree between samples.

The three-dimensional grey convex incidence degree of X_p and X_q is defined as

$$r_{pq} = \frac{1}{3m(n-2)}\sum_{i=1}^{m}\sum_{j=1}^{n-2}\frac{1}{1+\left|r_{pq}^{1}\right|} + \frac{1}{3(m-2)n}\sum_{i=1}^{m-2}\sum_{j=1}^{n}\frac{1}{1+\left|r_{pq}^{2}\right|} + \frac{1}{3(m-2)(n-2)}\sum_{i=1}^{m-2}\sum_{j=1}^{n-2}\frac{1}{1+\left|r_{pq}^{3}\right|} \quad (1)$$

where

$r_{pq}^{1} = a_{i,j+2} - 2a_{i,j+1} + a_{ij} - (b_{i,j+2} - 2b_{i,j+1} + b_{ij})$,

$r_{pq}^{2} = a_{i+2,j} - 2a_{i+1,j} + a_{ij} - (b_{i+2,j} - 2b_{i+1,j} + b_{ij})$,

$r_{pq}^{3} = (a_{i+2,j} - 2a_{i+1,j} + a_{ij})(a_{i,j+2} - 2a_{i,j+1} + a_{ij}) - (a_{i+1,j+1} - a_{i+1,j} - a_{i,j+1} + a_{ij})^2 - (b_{i+2,j} - 2b_{i+1,j} + b_{ij})(b_{i,j+2} - 2b_{i,j+1} + b_{ij}) + (b_{i+1,j+1} - b_{i+1,j} - b_{i,j+1} + b_{ij})^2$.

2.2 Grid GRA Model

Grid GRA model [12] uses the line slope and geometric characteristics of the panel data in three dimensions to calculate the incidence degree of the grey grid between samples.

The three-dimensional grey grid incidence degree of X_p and X_q is defined as

$$r_{pq} = \frac{1}{2(m-1)n}\sum_{i=1}^{m-1}\sum_{j=1}^{n} r'_{pq} + \frac{1}{2m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n-1} r''_{pq} \quad (2)$$

where

$$r'_{pq} = sgn((a_{i+1,j} - a_{ij})(b_{i+1,j} - b_{ij})) \cdot \frac{1 + \left|a_{i+1,j} - a_{ij}\right| + \left|b_{i+1,j} - b_{ij}\right|}{1 + \left|a_{i+1,j} - a_{ij}\right| + \left|b_{i+1,j} - b_{ij}\right| + \left|\left|a_{i+1,j} - a_{ij}\right| - \left|b_{i+1,j} - b_{ij}\right|\right|},$$

$$r''_{pq} = sgn((a_{i,j+1} - a_{ij})(b_{i,j+1} - b_{ij})) \cdot \frac{1 + |a_{i,j+1} - a_{ij}| + |b_{i,j+1} - b_{ij}|}{1 + |a_{i,j+1} - a_{ij_1}| + |b_{i,j+1} - b_{ij}| + ||a_{i,j+1} - a_{ij}| - |b_{i,j+1} - b_{ij}||}.$$

2.3 Modified Absolute GRA Model

Absolute GRA model is first proposed by Zhang et al. [24], and is applicable to modeling of monotonic or independent disjoint panel data. To overcome this limitation, Liu et al. [13] extended the absolute GRA model, named as the modified absolute GRA model.

The modified absolute GRA model mainly adopted the geometric volume characteristics of the panel data to measure the incidence degree between samples. And the modified three-dimensional grey absolute incidence degree of X_p and X_q is defined as

$$r_{pq} = \frac{1 + s_p + s_q}{1 + s_p + s_q + s_{pq}} \tag{3}$$

where

$s_p = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} [f(a_{ij}, a_{i+1,j}, a_{i,j+1}) + f(a_{i+1,j+1}, a_{i+1,j}, a_{i,j+1})]$,

$s_q = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} [f(b_{ij}, b_{i+1,j}, b_{i,j+1}) + f(b_{i+1,j+1}, b_{i+1,j}, b_{i,j+1})]$,

$s_{pq} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} f(a_{ij} - b_{ij}, a_{i+1,j} - b_{i+1,j}, a_{i,j+1} - b_{i,j+1}) + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} f(a_{i+1,j+1} - b_{i+1,j+1}, a_{i+1,j} - b_{i+1,j}, a_{i,j+1} - b_{i,j+1})$,

$$f(x,y,z) = \begin{cases} |1/6(x+y+z)| & x \geq 0, y \geq 0, z \geq 0 \quad or \quad x \leq 0, y \leq 0, z \leq 0 \\ \left|1/6(x+y+z) - 1/3 \cdot \dfrac{x^3}{(y-x)(z-x)}\right| & x > 0, y < 0, z < 0, or \\ & x < 0, y > 0, z > 0 \\ \left|1/6(x+y+z) - 1/3 \cdot \dfrac{y^3}{(x-y)(z-y)}\right| & x \leq 0, y > 0, z \leq 0, or \\ & x \leq 0, y < 0, z \leq 0 \\ \left|1/6(x+y+z) - 1/3 \cdot \dfrac{z^3}{(x-z)(y-z)}\right| & x < 0, y \leq 0, z > 0, or \\ & x > 0, y \geq 0, z < 0 \end{cases}.$$

2.4 Curvature GRA Model

Curvature GRA model [14] used the discrete curvature of the panel data projected on the time dimension and the index dimension for cluster analysis of a system. Accordingly, the three-dimensional grey curvature incidence degree of X_p and X_q is defined as

$$r_{pq} = \frac{1}{2m(n-2)} \sum_{i=1}^{m} \sum_{j=1}^{n-2} \frac{1}{1 + |r'_{pq}|} + \frac{1}{2(m-2)n} \sum_{i=1}^{m-2} \sum_{j=1}^{n} \frac{1}{1 + |r''_{pq}|} \tag{4}$$

where

$$r'_{pq} = sgn(a_{i,j+2} - 2a_{i,j+1} + a_{ij}) \cdot \frac{a_{i,j+2} - 2a_{i,j+1} + a_{ij}}{(1 + (a_{i,j+1} - a_{ij})^2)^{3/2}} - sgn(b_{i,j+2} - 2b_{i,j+1} + b_{ij}) \cdot \frac{b_{i,j+2} - 2b_{i,j+1} + b_{ij}}{(1 + (b_{i,j+1} - b_{ij})^2)^{3/2}},$$

$$r''_{pq} = sgn(a_{i+2,j} - 2a_{i+1,j} + a_{ij}) \cdot \frac{a_{i+2,j} - 2a_{i+1,j} + a_{ij}}{(1 + (a_{i+1,j} - a_{ij})^2)^{3/2}} - sgn(b_{i+2,j} - 2b_{i+1,j} + b_{ij}) \cdot \frac{b_{i+2,j} - 2b_{i+1,j} + b_{ij}}{(1 + (b_{i+1,j} - b_{ij})^2)^{3/2}}.$$

Obviously, the four ICGRA models mentioned above account for the coupling between cross-sectional data, but they do not satisfy the permutation of indicators. That is, the grey relational order would change if reordering the indicators of panel data.

2.5 Tetrahedral GRA Model

Tetrahedral GRA model [15] adopted the volume of tetrahedroid and the sub-matrixes composed of binary indicators in panel data to calculate the grey tetrahedral incidence degree between samples.

Grey tetrahedral incidence degree of X_p and X_q is defined as

$$r_{pq} = \frac{2}{m(m-1)(n-1)} \sum_{s_{i_1}, s_{i_2} \in M_s^2} \sum_{j=1}^{n-1} \frac{1}{1 + \left| V_p(i_1, j) - V_q(i_2, j) \right|} \tag{5}$$

where

$$M_s^{\,2} = \left\{ \{s_{i_1}, s_{i_2}\} \,|\, i_1 \neq i_2 \in \{1, 2, 3, \cdots, m\} \right\},$$

$$V_p(i,j) = \begin{cases} 1/6 \left| a_{i_1,j} + a_{i_2,j+1} - a_{i_2,j} - a_{i_1,j+1} \right| & a_{i_1,j} + a_{i_2,j+1} \geq a_{i_2,j} + a_{i_1,j+1} \\ -1/6 \left| a_{i_1,j} + a_{i_2,j+1} - a_{i_2,j} - a_{i_1,j+1} \right| & a_{i_1,j} + a_{i_2,j+1} < a_{i_2,j} + a_{i_1,j+1} \end{cases},$$

$$V_q(i,j) = \begin{cases} 1/6 \left| b_{i,j} + b_{i+1,j+1} - b_{i+1,j} - b_{i,j+1} \right| & b_{i,j} + b_{i+1,j+1} \geq b_{i+1,j} + b_{i,j+1} \\ -1/6 \left| b_{i,j} + b_{i+1,j+1} - b_{i+1,j} - b_{i,j+1} \right| & b_{i,j} + b_{i+1,j+1} < b_{i+1,j} + b_{i,j+1} \end{cases}.$$

In essence, the tetrahedral GRA model decomposes the sample matrix into sub-matrices composed of binary indicators, calculates the grey incidence degree of all sub-matrices, and then takes their average value. Therefore, the tetrahedral GRA model belongs to the ICGRA model, and it considers the impact on the grey relational order between samples when reordering the indicators of panel data. Moreover, the tetrahedral GRA model keeps the consistency of grey relational order between samples.

3 Numerical Simulation of Indicators-Coupled Grey Relational Analysis Models for Panel Data

3.1 Dataset Description

REF dataset is an important component of the UCI machine learning database, and represents the small-scale multivariate sample data. Specific data is provided by the open dataset website (http://kdd.ics.uci.edu/databases/robotfailure/robotfailure.html) (accessed on 3 March 2025). REF dataset arises from the error tracking in robots by using six sensors, and results in five separate sub-datasets. Here, the initial subset LP1 is utilized to perform experiments. The overall sample size is 88, and it involves four categories, consisting of 21 normal, 16 front collision, 17 collision and 34 obstructions. Each sample is represented by a 6×15 matrix, encompassing 6 indexes and 15 consecutive acquisition instances.

Considering the specifics of REF dataset LP1, here assume $X_p = (a_{ij})_{6\times15}$ represents the system feature behaviors, as shown in Table 1, and $1 \leq i \leq 6$ is the indexes, $1 \leq j \leq 15$ is the consecutive acquisition instances. In detail, rows one to three of the matrix show the change in force, whereas rows four to six show the change in torque. Meanwhile, the entire set of 88 samples within REF dataset LP1 is supposed to be the factor interaction matrix. Specifically, all $X_q = (b_{ij})_{6\times15}$ are the factor interaction matrix, here $1 \leq q \leq 88$.

Table 1: System feature behaviors matrix $\boldsymbol{X_p}$

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F_x	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
F_y	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1

(Continued)

Table 1 (continued)

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F_z	63	63	63	63	63	63	63	63	63	63	63	63	63	63	63
T_x	−3	−3	−3	−3	−3	−3	−3	−3	−3	−3	−3	−3	−3	−3	−3
T_y	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
T_z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.2 Data Normalization

Firstly, we dimensionless the REF dataset LP1 before testing as there is an inconsistency between the units of force and torque.

Due to the fact that each sample within REF LP1 dataset is collected in a very short period of time, and each indicator can be approximated as continuously changing. In other words, the 88 factor interaction matrix are correlated across the timeline. Therefore, we will perform global normalization on the test dataset by using the interval operator ϕ. And the definition of interval operator is as follows [25]:

Let $A_k = (a_{ij})_{6\times 15}$ be the system behavior matrix, here $1 \le i \le 6$, $1 \le j \le 15$, and $1 \le k \le 88$, if ϕ meets the condition $\phi A_k = (\phi a_{ij})_{m\times n}$, here $\phi a_{ij} = \dfrac{a_{ij} - \min\limits_{1\le k\le 88}\min\limits_{1\le j\le 15} a_{ij}}{\max\limits_{1\le k\le 88}\max\limits_{1\le j\le 15} a_{ij} - \min\limits_{1\le k\le 88}\min\limits_{1\le j\le 15} a_{ij}}$, then ϕ is named the interval operator.

3.3 Testing Process

In this section, we will describe the detailed processes for the numerical simulation of ICGRA models for panel data. The calculation steps can be summarized as follows:

Step 1: Normalize the LP1 using the interval operator, that is, $LP1 = \{X_1, X_2, \cdots, X_{88}\}$, where $X_q = (b_{ij}^{(q)})_{6\times 15}$,$1 \le q \le 88$.

Step 2: Select a test model (such as, Convex GRA model) and a test sample (such as, $X_1 = (b_{ij}^{(1)})_{6\times 15}$), and then calculate the incidence degree between the test sample $X_1 = (b_{ij}^{(1)})_{6\times 15}$ and the reference sample $X_p = (a_{ij})_{6\times 15}$ based on the selected test model.

Step 3: Exchange the columns of the test sample $X_1 = (b_{ij}^{(1)})_{6\times 15}$, and recalculate the incidence degree between the test sample and the reference sample based on the test model selected in Step 2.

Step 4: Continue to exchange the columns of test sample $X_1 = (b_{ij}^{(1)})_{6\times 15}$ (there are a total of 6! exchanges here), and repeat Step 3.

Step 5: Replace the test sample $X_1 = (b_{ij}^{(1)})_{6\times 15}$ of Step 2 with the next test sample (there are a total of 88 samples), and repeat Step 3 and Step 4.

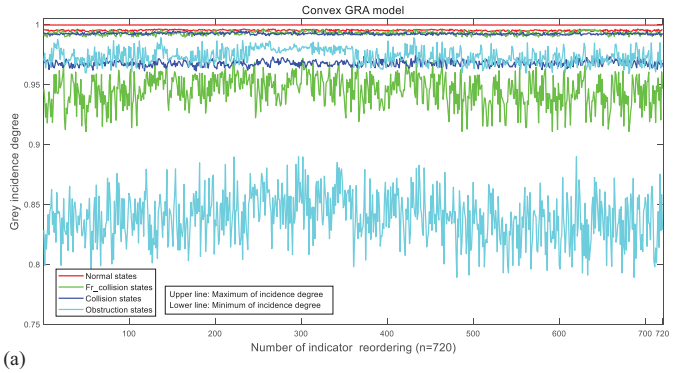
Step 6: Replace the test model of Step 2 with the next test model (there are a total of 4 models to be tested), and repeat Step 2 to Step 5 in sequence, and finally obtain $4 \times 720 \times 88$ values of incidence degree.

Step 7: Perform a data analysis and a graphical display based on the $4 \times 720 \times 88$ values of the incidence degree.

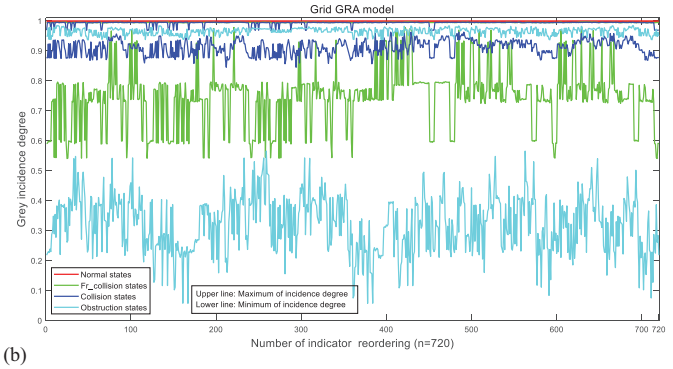
The above steps need to be performed with some computer software. In this paper, all the procedure is realized by programming in Python.

3.4 Test Reports

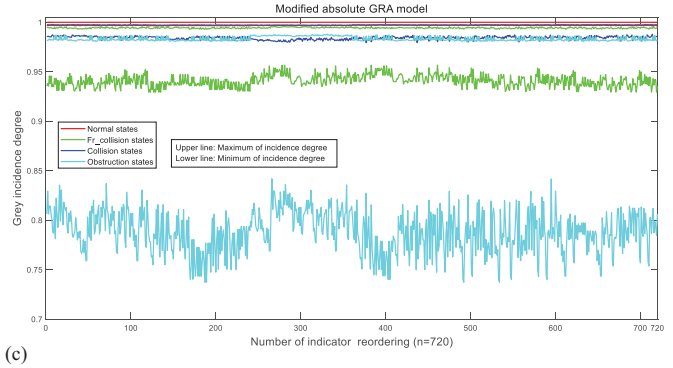
According to the previous description of the REF dataset, there are six indicators for each sample in the LP1 dataset, and there are a total of 6! (that is, 720) arrangements of indicators in panel data for each sample. Furthermore, there are a total of 88 samples in the LP1 dataset, so there will yield 720×88 grey incidence degree for each ICGRA model. By programming in Python, the 720×88 grey incidence degree of each ICGRA model was obtained. For illustration, Fig. 1 plotted the interval graph of grey incidence degree of four states in 720 arrangements of indicators for each model.



(a)

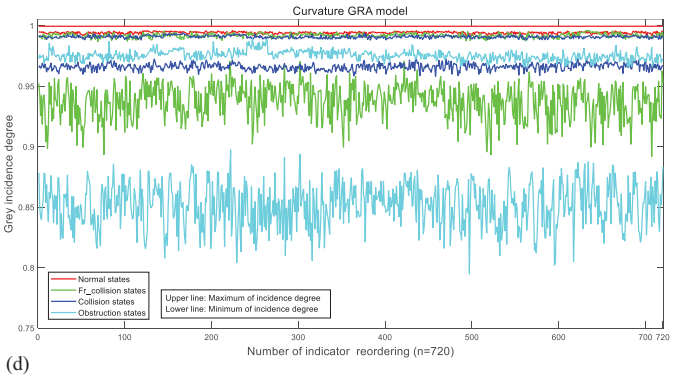


(b)



(c)

Figure 1: (Continued)



(d)

Figure 1: Interval graph of incidence degree of four states in each indicator sequence provided by four ICGRA models. (a) Convex GRA model; (b) Grid GRA model; (c) Modified absolute GRA model; (d) Curvature GRA model

As depicted in Fig. 1, the grey incidence degree provided by the four ICGRA models fluctuates irregularly with changes in the arrangement of indicators. In some arrangements, the grey incidence degree maintains a consistent change across the four states. However, in most arrangements, the grey incidence degree shows a distinct pattern of change, exemplified by the incidence degree interval transitioning from non-intersecting to intersecting states. Clearly, there is an inconsistency in the grey relational order at the juncture between the upper and lower bounds of the incidence degree interval.

In this way, if these ICGRA models are directly used to process cross-sectional data, the grey relational order generated may become unstable due to indicator reordering, which may have a negative impact on decision reference. Such as frequent changes in relational order, unreliable decision references, and limited applicability of the model itself. Therefore, when using these ICGRA models for cross-sectional data analysis, it is necessary to consider using other methods to stabilize decision-making.

4 Optimization of Indicators-Coupled Grey Relational Analysis Models for Panel Data

4.1 Mean Value Theory for Optimization

According to the previous discussions, for n cross-sectional data, an ICGRA model can yield $n!$ Grey relational orders. Therefore, when n is slightly larger, such as $n = 5$, then $n! = 120$; if $n \geq 6$, then $n! \geq 720$, then there will generate a large sample due to the reordering of the indicator. If so, we can consider using mean value theory to optimize the stability for the above ICGRA models.

Next, we will investigate the 720 incidence degree of four states in different arrangements of indicators, including the mean value and the standard deviation. For an intuitive comparison, Table 2 lists the maximum and minimum of mean values of incidence degree for each state. Meanwhile, Figs. 2 and 3 depict the scatter plots of mean and standard deviation of the incidence degree offered by these ICGRA models, respectively. Besides, for comparison, the four states are still displayed in the previous four colors similar to Fig. 1.

Table 2 shows that, except for the modified absolute GRA model, the other three ICGRA models can all successfully distinguish between normal and failure states based on the average of 720 incidence degree. Compared to Fig. 1, we can clearly find that the original convex GRA model and the original curvature GRA model can only distinguish between normal and failure states in some cases. Therefore,

averaging the incidence degree for all arrangements of indicators in panel data can effectively improve the consistency of relational orders offered by the existing ICGRA models.

Table 2: Maximum and minimum of mean values of incidence degree for four states calculated by four ICGRA models

Models	21 normal states	16 front collision states	17 collision states	34 obstruction states
Convex GRA	[0.9956, 0.9995]	[0.9459, 0.9907]	[0.9679, 0.9916]	[0.8581, 0.9702]
Grid GRA	[0.9970, 0.9997]	[0.7706, 0.9945]	[0.9199, 0.9856]	[0.3357, 0.9532]
Modified absolute GRA	[0.9967, 0.9999]	[0.9407, 0.9941]	[0.9839, 0.9971]	[0.7918, 0.9831]
Curvature GRA	[0.9946, 0.9995]	[0.9391, 0.9894]	[0.9667, 0.9900]	[0.8643, 0.9734]

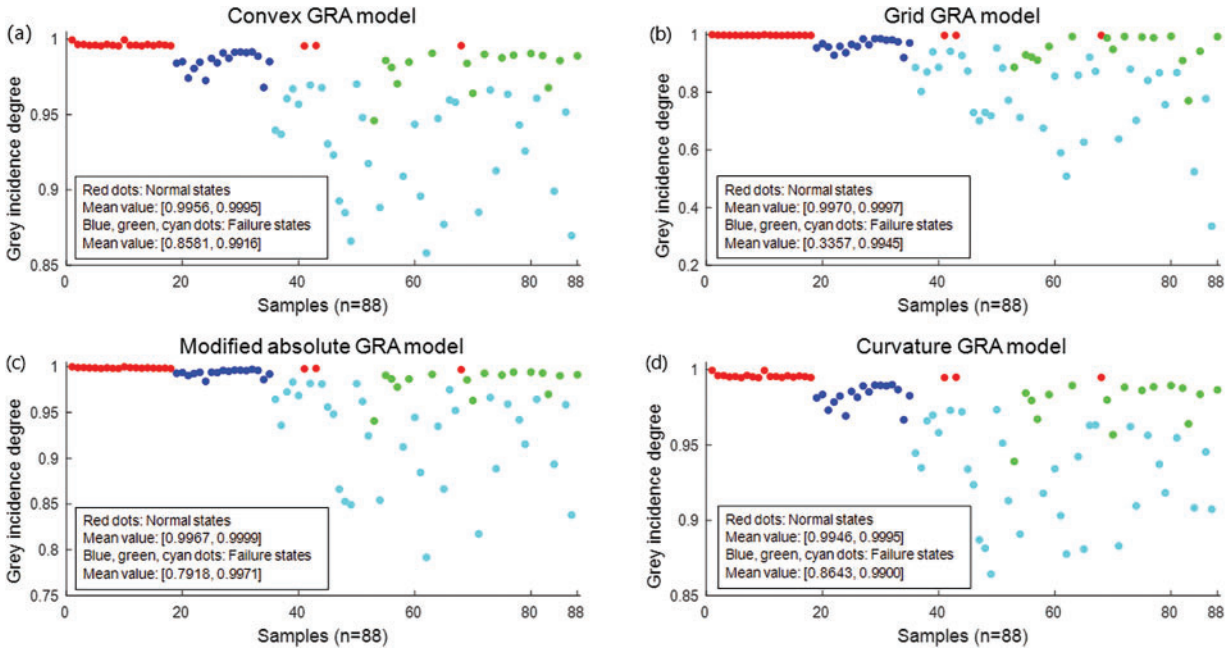


Figure 2: Scatter plot of 88 mean values of incidence degree provided by four ICGRA models. (a) Convex GRA model; (b) Grid GRA model; (c) Modified absolute GRA model; (d) Curvature GRA model

Fig. 2 indicates that, except for the collision state, the mean values of incidence degree of the other three states show a clear hierarchical distribution, with the normal state having the highest mean values, followed by the front collision state, and the obstruction state having the lowest mean values. This implies that the ICGRA models can effectively distinguish different states by averaging the incidence degree of indicators of all arrangements in panel data.

On the other hand, Fig. 3 shows that the standard deviation of the incidence degree is very small (less than 0.03) except for the grid GRA model. This indicates that the average of the incidence degree of all arrangements of indicators can also represent the central trend of the sample. At the same time, it

also demonstrates that the mean value theory provides a reliable optimization approach for the existing ICGRA models.

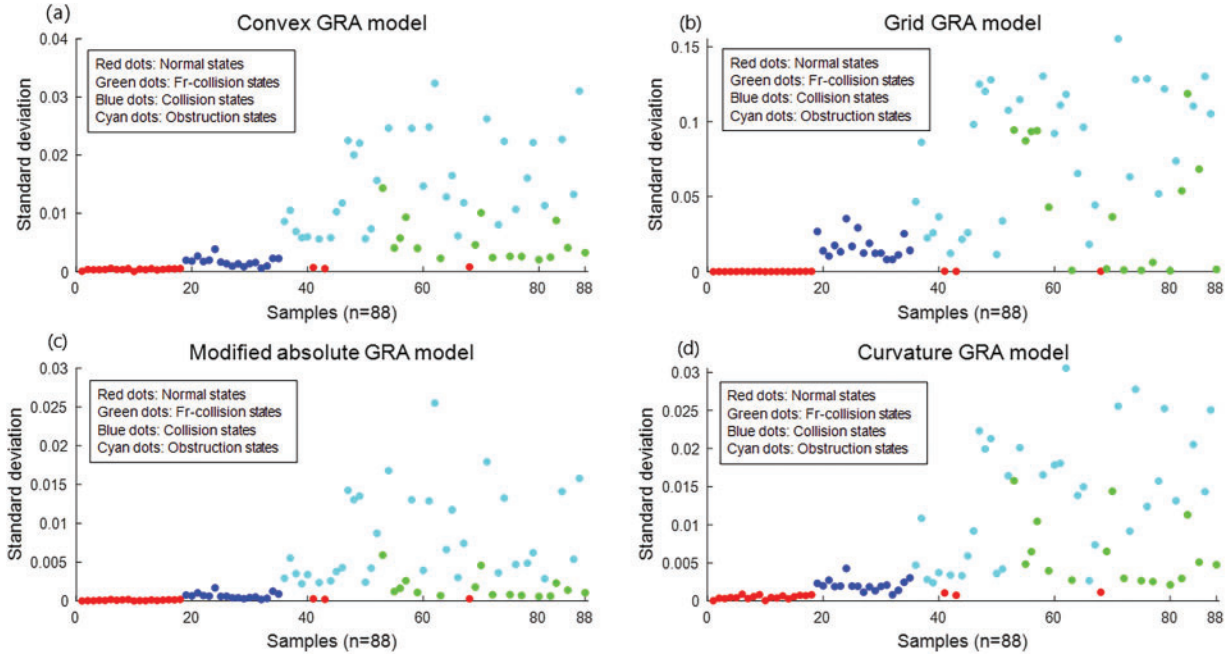


Figure 3: Scatter plot of 88 standard deviation of incidence degree provided by four ICGRA models. (a) Convex GRA model; (b) Grid GRA model; (c) Modified absolute GRA model; (d) Curvature GRA model

4.2 Deep Modeling for Optimization

Note that the grid GRA model in Section 2.2 adopts the absolute horizontal distance of the indicators as a measure of similarity in the time dimension. Due to the spatiotemporal characteristics of the panel data, it is also necessary to measure the dynamic horizontal distance of the indicators during modeling. Therefore, in order to better describe panel data, the dynamic horizontal distance of indicators will be added to the original grid GRA model to measure the similarity, and a deep modeling for optimization of grid GRA model is presented as follows:

Let

$$r'_{pq} = sgn((a_{i+1,j} - a_{ij})(b_{i+1,j} - b_{ij})) \cdot \frac{1 + |a_{i+1,j} - a_{ij}| + |b_{i+1,j} - b_{ij}|}{1 + |a_{i+1,j} - a_{ij}| + |b_{i+1,j} - b_{ij}| + ||a_{i+1,j} - a_{ij}| - |b_{i+1,j} - b_{ij}||},$$

$$r''_{pq} = sgn((a_{i,j+1} - a_{ij})(b_{i,j+1} - b_{ij})) \cdot \frac{1 + |a_{i,j+1} - a_{ij}| + |b_{i,j+1} - b_{ij}|}{1 + |a_{i,j+1} - a_{ij}| + |b_{i,j+1} - b_{ij}| + ||a_{i,j+1} - a_{ij}| - |b_{i,j+1} - b_{ij}||},$$

$$r''_{pq} = sgn(\frac{a_{i,j+1} - a_{ij}}{a_{ij}} \cdot \frac{b_{i,j+1} - b_{ij}}{b_{ij}}) \cdot \frac{1 + \left|\frac{a_{i,j+1} - a_{ij}}{a_{ij}}\right| + \left|\frac{b_{i,j+1} - b_{ij}}{b_{ij}}\right|}{1 + \left|\frac{a_{i,j+1} - a_{ij}}{a_{ij}}\right| + \left|\frac{b_{i,j+1} - b_{ij}}{b_{ij}}\right| + \left|\left|\frac{a_{i,j+1} - a_{ij}}{a_{ij}}\right| - \left|\frac{b_{i,j+1} - b_{ij}}{b_{ij}}\right|\right|}.$$

Then, $r_{pq} = \frac{1}{2(m-1)n}\sum_{i=1}^{m-1}\sum_{j=1}^{n} r'_{pq} + \frac{1}{4m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n-1} r''_{pq} + \frac{1}{4m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n-1} r'''_{pq}$ is defined as the deep grid grey incidence degree of X_p and X_q.

Next, the first sub LP1 of the REF dataset will continue to be used to test the deep grid GRA model for panel data. Table 3 displays the upper and lower bounds of mean values of incidence degree offered by the deep grid GRA model in all arrangements of indicators. Meanwhile, Fig. 4 presents the mean and the standard deviation of the incidence degree for the deep grid GRA model, depicted in the form of scatter plots, where the four states are still displayed by four colors.

Table 3: Maximum and minimum of mean values of incidence degree for four states calculated by the deep grid GRA model

Model	21 normal states	16 front collision states	17 collision states	34 obstruction states
Deep grid GRA	[0.9967, 0.9997]	[0.7699, 0.9925]	[0.9171, 0.9848]	[0.3334, 0.9512]

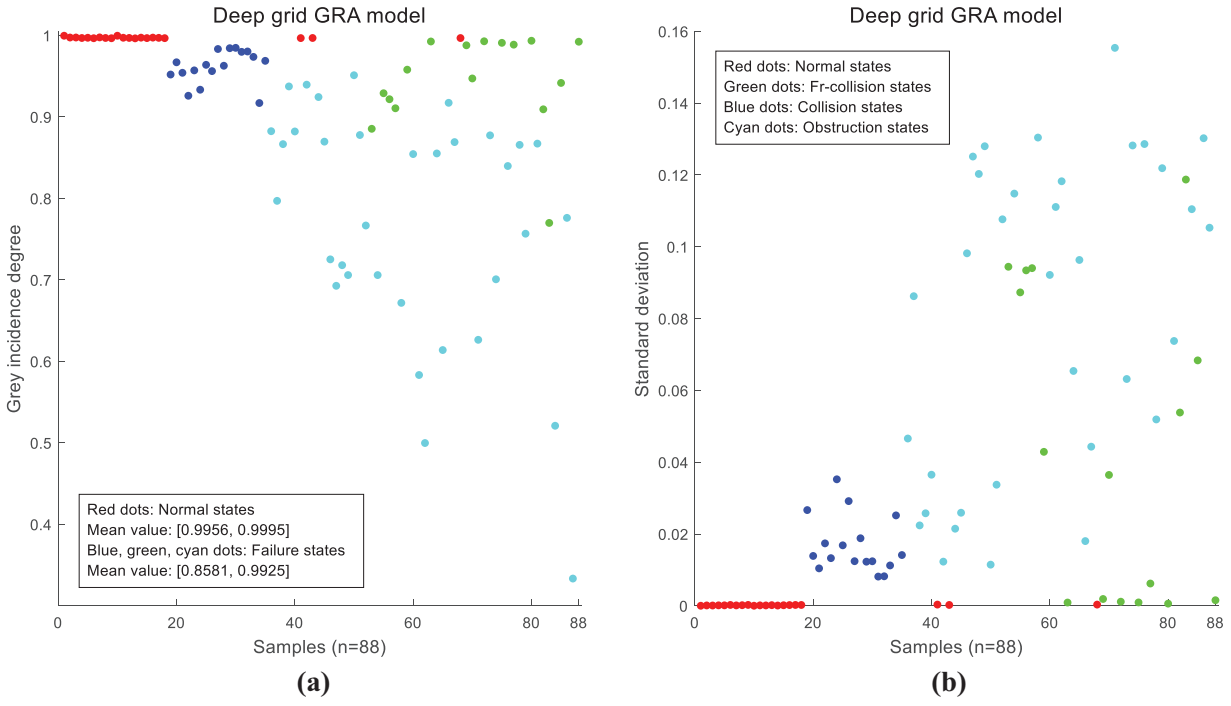


Figure 4: Scatter plot of 88 mean values and standard deviation of incidence degree provided by the deep grid GRA model. (a) Deep grid GRA model; (b) Deep grid GRA model

Table 3 confirms the effectiveness of the deep grid GRA model, which successfully separate normal states from failure states based on the average of 720 incidence degree. In addition, Fig. 4 implies that the deep grid GRA model shows more consistent in the mean and standard deviation of incidence degree than the original grid GRA model. This indicates that the deep grid GRA model not only inherits the advantages of the original model, but also provides an optimization for the existing grid GRA model.

5 Comparison and Analysis

In order to comprehensively verify the effectiveness of optimization for ICGRA models discussed in Section 4, two comparisons were constructed in this section: one was compared with the latest model in grey system theory, and the other was compared with the k-nearest neighbor (KNN) algorithm in machine learning theory.

5.1 Comparison in Grey System Theory

As mentioned in Section 2.5, the tetrahedral GRA model is one of the few ICGRA models that considers the impact of reordering indicators in panel data on the incidence degree of samples. For direct comparison, we will continue to use the first sub LP1 of REF dataset to test the tetrahedral GRA model. Table 4 lists the maximum and minimum of the incidence degree obtained by the tetrahedral GRA model for each state. Meanwhile, Fig. 5 presents the mean and standard deviation of the incidence degree of the tetrahedral GRA model in the form of scatter plots, where the four states are still displayed by four colors.

Table 4: Maximum and minimum of incidence degree for four states calculated by tetrahedral GRA model

Model	21 normal states	16 front collision states	17 collision states	34 obstruction states
Tetrahedral GRA	[0.9959, 0.9994]	[0.9927, 0.9970]	[0.9639, 0.9946]	[0.9243, 0.9976]

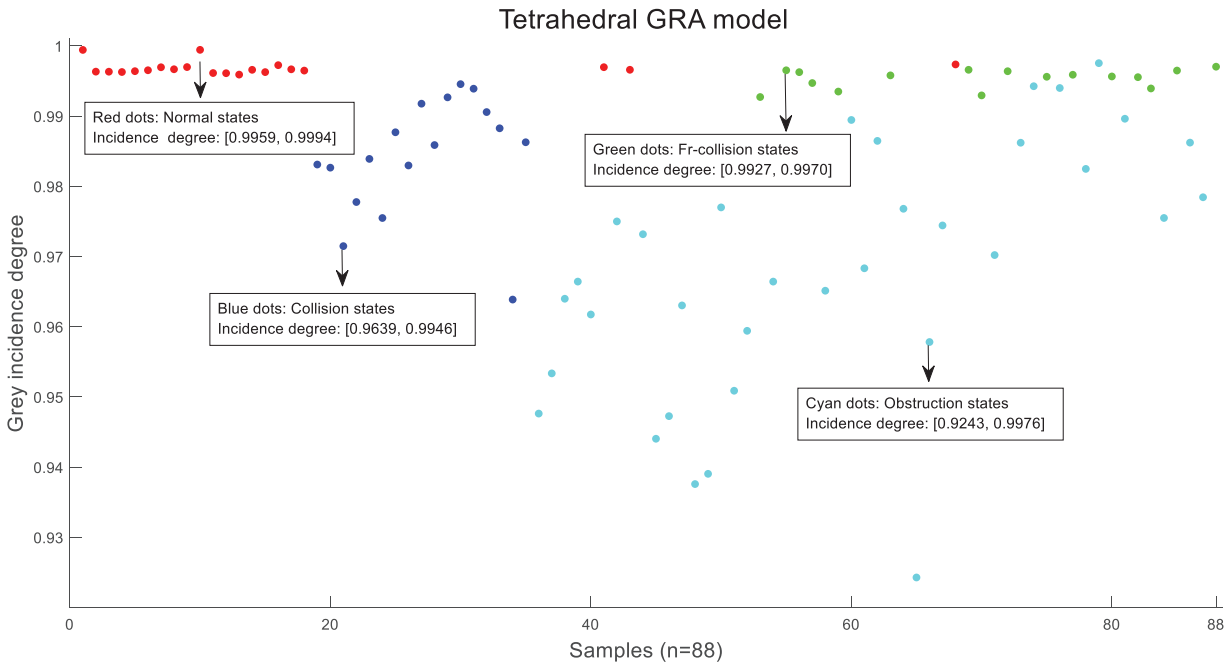


Figure 5: Scatter plot of 88 incidence degree provided by tetrahedral GRA model

Table 4 shows that the minimum of incidence degree of the 21 normal states is lower than the maximum of incidence degree of the 16 front collision states and 34 obstruction states. This implies that it is difficult for the tetrahedral GRA model to distinguish between normal and failure states.

Fig. 5 reveals a similarity between the tetrahedral GRA model and the optimization form of the modified absolute GRA model. As discussed in Section 4.1, even averaging the incidence degree for all arrangements of indicators in the panel data, the modified absolute GRA model cannot yet effectively distinguish between normal and failure states. The reasons for this may be the similarities in mathematical structures between the two ICGRA models, such as constructing models based on the volume of spatial geometry. Therefore, when coupling indicators in panel data at different time points, it is necessary to carefully consider the feasibility of the method.

5.2 Comparison with Machine Learning Theory

KNN algorithm is a simple and powerful machine learning algorithm, commonly used for classification [26]. It is to find K data points in the training data that are closest to the target data point, and predict the category to which the target data point belongs. For direct comparison, we will continue to use the first sub LP1 of REF dataset to test the KNN algorithm. Meanwhile, the classification evaluation results are used to compare with five grey clustering models discussed previously in this paper, including the averaged convex GRA model, the averaged grid GRA model, the averaged modified absolute GRA model, the averaged curvature GRA model, the averaged deep grid GRA model and the tetrahedral GRA model.

By programming in Python, the metrics of overall performance for all compared models are listed in Table 5. Where there are four classification evaluation metrics: accuracy, precision, recall (sensitivity) and F1 (score). These are quantitative measures used to assess the performance of a classification model, and provide insights into the accuracy and reliability of the model's predictions.

Table 5: Classification evaluation results of five optimized GRA models, tetrahedral GRA model and KNN method

Models	Accuracy	Precision	Recall	F1
Averaged convex GRA model	1	1	1	1
Averaged grid GRA model	1	1	1	1
Averaged modified absolute GRA model	0.9773	0.9687	0.9687	0.9687
Averaged curvature GRA model	1	1	1	1
Tetrahedral GRA model	0.8333	0.8438	0.9	0.871
KNN method	1	1	1	1
Averaged deep grid GRA model	1	1	1	1

Table 5 shows that the averaged convex GRA model, the averaged grid GRA model, the averaged curvature GRA model, the averaged deep grid GRA model and KNN method performed well, their classification accuracy is all 100%. The optimization of the absolute GRA model is still a little poor, and its classification accuracy is 97.73%. However, the latest tetrahedral GRA model is inadequacy, its classification accuracy is only 83.33%. This is consistent with the results shown in Fig. 5. This further proves that it is very important to carefully consider the coupling of indicators at different time points when clustering panel data with cross-sectional dependence.

6 Conclusion

In this paper, an advanced numerical simulation is achieved to investigate the performance of the four existing ICGRA models for panel data. Simulation results reveal that the grey incidence degree provided by the traditional ICGRA models fluctuates irregularly with changes in the arrangement of indicators, and there is an inconsistency in the grey relational order with the reordering of indicators in panel data. To resolve this problem, this paper proposed an averaging method on the grey incidence degree for all arrangements of indicators in panel data and added the dynamic horizontal distance of indicators to the original grid GRA model to obtain a deep modeling of it. Meanwhile, the optimized ICGRA models are compared with the tetrahedral GRA model to verify the effectiveness of optimization, and compared with the k-nearest neighbor (KNN) algorithm as well. Results show that the classification accuracy of the averaged absolute GRA model was 97.73%, the other optimized ICGRA models and the KNN method all achieved 100% accuracy, while the tetrahedral GRA model has an accuracy of 83.33%. Results also indicate that there is a similarity between the tetrahedral GRA model and the optimization form of the modified absolute GRA model, as their mathematical structures are based on the volume of spatial geometry and fail to distinguish between normal and failure states. This phenomenon indicates that, when coupling indicators in panel data at different time points, it is needed to carefully consider the feasibility of the method.

Therefore, numerical simulations based on the dataset can help validate the ICGRA models, and the average grey incidence degree for all arrangements of indicators and deep modeling significantly improves the stability of models and enhances the clustering accuracy in different cases.

Acknowledgement: The authors would like to thank Professor Naiming Xie for his suggestions on our manuscript, and the editors and reviewers for their effort in the work.

Funding Statement: This work was supported by the following funding: National Natural Science Foundation of China (62341308, Fanghong Jian); Natural Science Foundation of Jiangxi Province (20232BAB201020, Xiaomei Liu and Fanghong Jian); Natural Science Foundation of Jiangxi Province (20224BAB201010, Fanghong Jian); Natural Science Foundation of Jiangxi Province (20242BAB25095, Fanghong Jian); Talent Launch Project of Huangshan University (2023xskq006, Xiaozhong Tang).

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Xiaomei Liu, Fanghong Jian; Data collection: Xiaomei Liu; Analysis and interpretation of results: Xiaomei Liu, Fanghong Jian, Xiaozhong Tang; Draft manuscript preparation: Xiaomei Liu, Fanghong Jian. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Specific Robot Execution Failures (REF) dataset LP1 is provided by the open dataset website (http://kdd.ics.uci.edu/databases/robotfailure/robotfailure.htm) (accessed on 3 March 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Liu SF, Yang YJ, Forrest J. Grey data analysis: methods, models and applications. London, UK: Springer Press; 2016.
2. Ikram M, Zhang Q, Sroufe R, Ali Shah SZ. Towards a sustainable environment: the nexus between ISO 14001, renewable energy consumption, access to electricity, agriculture and CO_2 emissions in SAARC countries. Sustain Prod Consum. 2020;22(1):218–30. doi:10.1016/j.spc.2020.03.011.
3. Lin S, Wang Q, Deng M, Su L, Wei K, Guo Y, et al. Assessing the influence of water fertilizer, and climate factors on seed cotton yield under mulched drip irrigation in Xinjiang Agricultural Regions. Eur J Agron. 2024;152:127034. doi:10.1016/j.eja.2023.127034.
4. Rehman S, Rehman E, Mumtaz A, Zhang J. Cardiovascular disease mortality and potential risk factor in China: a multi-dimensional assessment by a grey relational approach. Int J Public Health. 2022;67:1604599. doi:10.3389/ijph.2022.1604599.
5. Lin M, Chen H, Jia L, Yang M, Qiu S, Song H, et al. Using a grey relational analysis in an improved Grunow-Finke assessment tool to detect unnatural epidemics. Risk Anal. 2023;43(7):1508–17. doi:10.1111/risa.14016.
6. Deng JL. The theory and methods of socio-economy grey system. Soc Sci China. 1984;6:47–60.
7. Jian F, Li J, Liu X, Wu Q, Zhong D. Modified Deng's grey relational analysis model for panel data and its applications in assessing the water environment of Poyang Lake. Processes. 2024;12(9):1935. doi:10.3390/pr12091935.
8. Gui YF, Xia YF, Deng LC. Grey incident grade in linear norm space. J Wuhan Univ Technol. 2004;28(3):399–412. (In Chinese).
9. Zhang M, Luo D, Su Y. Drought monitoring and agricultural drought loss risk assessment based on multisource information fusion. Nat Hazards. 2022;111(1):775–801. doi:10.1007/s11069-021-05078-w.
10. Sun DC, Luo D, Zhang HH. Modelling principles of grey matrix incidence analysis for panel data. J Grey Syst. 2021;33(3):16–30.
11. Wu LF, Liu SF, Yao LG, Yan SL. Grey convex relational degree and its application to evaluate regional economic sustainability. Sci Iran. 2013;20(1):44–9. doi:10.1016/j.scient.2012.11.002.
12. Liu Z, Dang YG, Qian WY, Zhou WJ. Grey grid incidence model for panel data. Syst Eng Theor Pract. 2014;34(4):991–6. (In Chinese).
13. Liu XM, Ke L, Yu JJ. An improved model of three-dimensional absolute degree of grey incidence. Stat Decis. 2018;34(22):20–4. (In Chinese). doi:10.13546/j.cnki.tjyjc.2018.22.004.
14. Wu HH, Qu ZF. A grey incidence model for panel data based on the curvature of discrete surface. J Grey Syst. 2022;34(2):75–87.
15. Wu HH, Liu SF, Fang ZG. Grey tetrahedral grid incidence analysis model based on panel data and its application. Control Decis. 2022;37(11):3033–41. (In Chinese). doi:10.13195/j.kzyjc.2021.0753.
16. Hu AQ, Xie NM. Construction and application of a novel grey relational analysis model considering factor coupling relationship. Grey Syst. 2024;15(1):1–20. doi:10.1108/GS-04-2024-0046.
17. Zeng S, Yang C. Risk evaluation of livestream e-commerce platforms based on expert trust networks and CODAS. Expert Syst Appl. 2025;260(1):125408. doi:10.1016/j.eswa.2024.125408.
18. Huang H, Kao C, Urga G. Copula-based tests for cross-sectional independence in panel models. Econ Lett. 2008;100(2):224–8. doi:10.1016/j.econlet.2008.01.017.
19. Zeng S, Ye A, Su W, Chen M, Llopis-Albert C. Site evaluation of subsea tunnels with sightseeing function based on dynamic complex MARCOS method. Technol Forecast Soc Change. 2024;199(8):123041. doi:10.1016/j.techfore.2023.123041.
20. Li X, Hipel KW, Dang Y. An improved grey relational analysis approach for panel data clustering. Expert Syst Appl. 2015;42(23):9105–16. doi:10.1016/j.eswa.2015.07.066.

21. Li B, Rong X, Li Y. An improved kernel based extreme learning machine for robot execution failures. Sci World J. 2014;2014(8):1–7. doi:10.1155/2014/906546.
22. Liu Y, Wang X, Ren X, Lyu F. Deep convolution neural networks for the classification of robot execution failures. In: 2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS); 2019 Jul 5–7; Xiamen, China: IEEE; 2019. p. 535–40. doi:10.1109/safeprocess45799.2019.9213393.
23. Liu XM, Yu JJ. Grey incidence analysis models for matrix data and matrix sequences data. J Grey Syst. 2019;31(3):59–70.
24. Zhang K, Liu SF. Extended clusters of grey incidences for panel data and its application. Syst Eng Theor Pract. 2010;30(7):1253–9. (In Chinese).
25. Deng JL. Grey control systems. Wuhan, China: Huazhong University of Science and Technology Press; 1985.
26. Uddin S, Haque I, Lu H, Moni M, Gide E. Comparative performance analysis of K-nearest neighbor (KNN) algorithm and its different variants for disease prediction. Sci Rep. 2024;12(1):1–12.