Grey Relational Analysis And Its Application On Multivariate Time Series.

Roselina Sallehuddin, Siti Mariyam Hj. Shamsuddin, Siti Zaiton Mohd Hashim

Fakulti Sains Komputer dan Sistem Maklumat,

Universiti Teknologi Malaysia, 81300

Skudai, Johor

Roselina, Mariyam, SitiZaiton @utm.my

Abstract

Recently, grey relational analysis (GRA) has been widely applied in analyzing multivariate time series data (MTS); as a solution to the traditional statistical limitations. GRA, which has been proven to be simple and accurate method for analysing relationship in MTS are broadly used in various disciplines such as economics, sociology and engineering. GRA is employed to search for grey relational grade (GRG), which can be used to describe the relationships among the factors and to determine the important factors that significantly influence some defined objectives. Consequently, the aim of this paper is to demonstrate the application of GRA in determining the most influential affecting factors that affect the grain crop yield in China from 1990 to 2002. Additionally, a list of affecting factors that ordered according to their importance to crop yield were also determined. Analysis results of sample data showed that the main factor that affect the trend of crop yield are the consumption of pesticide and chemical fertilizer; while the least important factors need to be considered is the agriculture labour. Thus, by properly adjusting the significant affecting factors, the China's crop yield performance can be improved efficiently. Furthermore, GRA can provide a ranking scheme that rank the order of the grey relationship among dependent and independent factors, and this allow us to decide which input factors need to be considered to forecast grain crop yield more precisely.

Keywords: Grey relational analysis, Grey relational grade, affecting factors, multivariate time series, forecast

1.Introduction

In a complex and multivariate time series system, such as social system and economic system, many factors simultaneously influence the system to determine the state of development of the system Usually, we want to know which factors influences the system more and which factor influences the system little. However, the relationship between various factors is usually grey where the information is unclear, incomplete and uncertain. Moreover, practical and experimental data was difficult to obtain and too much scatter to analyze. Two conventional statistical methods frequently used on the relationship between independent and dependent factors were factor analysis and regression analysis methods. However, this analysis prescribes that there must be relationship of mutual influence between variables, and the function relationship will only be worked out under the condition of large quantities of data, which should conform to the typical distribution, like the normal distribution. For instance, to extract the significant financial ratio variables and financial indicators which affect the financial performance of venture capital in Taiwan [1]. However, normally, it is difficult to obtain the interior data from the venture capital enterprises. Therefore, traditional multivariate statistical methods could have a hard time obtaining a persuasive explanation. As a result, a new analysis method needs to be located.

To overcome the shortage of regression analysis and factor analysis, multi-attribute method, Grey relational analysis (GRA) has been proposed to solve the problem [1, 2, 3]. GRA is a kind of effective tool to make system analysis, and also lays a foundation for modelling, forecasting, clustering of grey systems. Comparing to regression analysis and factor analysis in mathematic statistics, grey relational has some merits such as small sample, having no use for typical distribution, no requirement for independency and small amount of calculation. Additionally, GRA analysis is already proved to be simple and accurate method for selecting factors especially for those problems with unique characteristic [4].Therefore, this study will utilize the GRA to establish GRG; a ranking scheme that rank the order of the grey relationship among dependent and independent factors. This scheme is called as an evaluation model for selecting significant factors in multivariate time series; where, GRG is rearranged according to the order of their magnitude.

Then, these selected significant factors will be used as a foundation for the next data mining process. In this paper, these significant factors will be used as input node in artificial neural network (ANN) forecasting model. As a case study, the grain crop yield of China and its affecting factors will be employed. GRA will search the relationships among the affecting factors and then determine the most and least important factor that significantly influences the trend of grain crop yield. Subsequently, GRA will recommend the number of input node for ANN forecasting model.

The rest of the paper is organized as following. In section 2, the grey theory and grey relational analysis are described. The method of GRA is presented in Section 3. Experimental data that used to evaluate the GRA model is given in section 4. In section 5, the experimental results obtained from GRA are discussed, and finally section 6 provides the concluding remarks.

2. Grey Theory and Grey Relational Analysis

Grey System (GS) is the system of which part information is known and part information is unknown. Up to now, GS theory has been developing a set of theories and techniques including grey mathematics, Grey relational analysis, grey modelling, grey clustering, grey forecasting, grey decision making, grey programming and grey control, and has been applied successfully in many engineering and managerial fields such as industry, ecology, meteorology, geography, earthquake, hydrology, medicine and military [6]. The major advantage of grey theory is that it can handle both incomplete information and unclear problems very precisely. It serves as an analysis tool especially in cases where there is insufficient data [3].

GRA is a new analysis method, which has been proposed in the Grey system theory and it is founded by Professor Deng Julong from Huazhong University of Science and Technology, People's Republic of China. GRA is based on geometrical mathematics, which compliance with the principles of normality, symmetry, entirety, and proximity. GRA is suitable for solving complicated interrelationships between multiple factors and variables and has been successfully applied on cluster analysis, robot path planning, project selection, prediction analysis, performance evaluation, and factor effect evaluation and multiple criteria decision [7,8,9]. Detailed explanation about GRA method is presented in the following section.

3. Grey Relational Analysis Method

There are 3 main steps in GRA. The first step is data pre-processing. Data pre-processing is usually required when the range or unit in one data sequence is different from others or the sequence scatter range is too large. Data pre-processing is a method of transferring the original data sequence to a comparable sequence. Therefore, data must be normalized, scaled and polarized first into a comparable sequence before proceed to other steps. The processing is called generation of grey relation or standard processing.

There are two process involves in this step; data representative and data normalization. First represent the original data series (X) as reference (x_0) and comparative series (x_i). In this paper, reference series, (x_0) represent the China's crop yield and comparative series (x_i) represent the ten affecting factors that influent the production of China's crop.

Then implement the data normalization. There are a few formulas of data preprocessing available for the GRA such as equation 1(a) and equation 1(b), [1]. The determination of which formula to be employed for data normalization is based on the characteristics of a data sequences, for example:
If the expectancy is the higher-the-better, then it can be expressed by

$$x^*(k) = \frac{x_i^0(k) - \min x_i^0(k)}{\max x_i^0(k) - \min x_i^0(k)} \text{ ,Eq. (1(a))}$$

If the expectancy is the lower-the-better, then it can be expressed by

$$x^*(k) = \frac{\max x_i^0(k) - x_i^0(k)}{\max x_i^0(k) - \min x_i^0(k)} \text{ Eq. (1(b))}$$

Where

$i = 1,...m;\quad k = 1,...n.$
m is number of experimental data items,
n is the number of parameters;
$x_i^O(k)$ is the original sequence,
$x_i^*(k)$ is the sequences after data preprocessing,
$\min x_i^0(k)$ and $\max x_i^0(k)$ are the smallest and the largest value of $x_i^O(k)$.

In this study, Equation (1(a)) is employed, since output of this study has the characteristic of the “higher is better”, means that if the value of grey relational grade is higher, than there is a strong relationship between comparative and reference series. The range of data is adjusted so as to fall within [0,1] range.

The second step is to locate the grey relational coefficient by using Eq. (2a), [10]:

$$\xi_i(k) = \frac{\Delta\min + \zeta\Delta\max}{\Delta_{0,i}(k) + \zeta\Delta\max}, \quad \text{Eq.(2a)}$$

where,

$\Delta_{0i} =$ deviation sequences of the reference sequence and comparability sequence

$\Delta_{0,j} = \left\| x_0^*(k) - x_i^*(k) \right\|,$

$\Delta\min = \min_{\forall j\in i}\min_{\forall k}\left\| x_0^*(k) - x_j^*(k) \right\|,$

$\Delta\max = \max_{\forall j\in i}\max_{\forall k}\left\| x_0^*(k) - x_j^*(k) \right\|,$

$x_0^*(k) =$ the reference sequence, and

$x_i^*(k) =$ the comparative sequence

ζ is known as identification coefficient with $\zeta \in [0,1]$, which can be adjusted to help make better distinction between normalized reference series and normalized comparative series. Normally ζ =0.5 is used because it’s offers moderate distinguishing effect and stability [15]. Furthermore, based on mathematic proof, the value change of ζ will only change the magnitude of the relational coefficient but it won’t change the rank of the grey relational grade [16]. From the calculation, the values of $\Delta\min$ and $\Delta\max$ are respectively 0 and 1. Then replace these values in Eq. (2a), and we obtained:

$$\xi_i(k) = \frac{0 + \zeta(1)}{\Delta_{0,i}(k) + \zeta(1)} = \frac{0.5}{\Delta_{0,i}(k) + 0.5} \quad \text{Eq.(2b)}$$

After the grey relational coefficient is derived, grey relational grade (GRG) is calculated by averaging the value of the grey relational coefficients. [10,11,12]. GRG is defined as the numerical measure of the relevancy between two systems or two sequences such as the reference sequence and the comparability sequence. The existing GRG between two series is always distributed between 0 and 1. Grey relational grade can be calculated using formula below [10]:

$$\gamma_i = \frac{1}{n}\sum_{k=1}^{n}\xi_i(k) \quad \text{Eq. (3)}$$

where γ_i represents GRG; the level of correlation between the reference sequence and the comparability sequence. In this study, GRG is used to indicate the degree influence that the comparability sequence (10 affecting factors) could exert over the reference sequence (China’s crop yield). Therefore, if a particular comparability sequence is more important than the other comparability sequences to the reference sequence, than the GRG for that comparability sequences and reference sequence will be higher than other GRG [13]. For example, if $\gamma(x_0, x_i) > \gamma(x_0, x_j)$, then the element x_i is closer to the reference element x_0 than the element x_j .Generally, γ_i > 0.9 indicates a marked influence, γ_i > 0.8 a relatively marked influence, γ_i > 0.7 a noticeable influence and γ_i < 0.6 a negligible influence [14].

4. Experimental Data Setup

To facilitate this study, a dataset obtained from journal [5] are used as experimental data. This sample contains 13 observations and it had specified annual China gross grain crop yields with their affecting factors. **Table 1** shows yearly data for gross grain crop yields and its affecting factors in China during 1990 till 2003. There are ten factors that affect the production of gross grain crop in China, and these include (a) total power of agricultural, (b) electricity consumed in rural areas, (c) irrigation area, (e) consumption of chemical fertilizer, (f) areas affected by natural disaster, (g) budgetary expenditure for agriculture, (h) sown area of grain crops, (l) consumption of pesticide, (m) consumption of agricultural film, and (n) agriculture laborers. The total production of grain crop yield is denoted by (d) .

In this study, data is divided into two parts. First part (1990-2001) is used to construct the evaluation model and to train the ANN. Data from 2002 to 2003 are used to validate the performance of ANN forecasting model. Here, forecasting one step ahead is implemented.

Table 1: Affecting factors for China grain crop yield(1990-2003)

Year	a	b	c	e	f	g	h	l	m	n	d
1990	28707.7	844.5	47403.1	2590.3	17819	9.98	113466	73.3	481982	33336.4	44624
1991	29388.6	963.2	47822.1	2805.1	27814	10.26	112314	76.1	642145	34186.3	43529
1992	30308.4	1106.9	48590.1	2930.2	25859	10.05	110560	79.5	780611	34037.2	44265.8
1993	31816.6	1244.8	48727.9	3151.9	23133	9.49	110509	84.5	707321	33258.2	45648.8
1994	33802.5	1473.9	48759.1	3317.9	31383	9.2	108544	97.9	887064	32690.3	44510.1
1995	36118.1	1655.7	49281.2	3593.7	22267	8.43	110060	108.7	915487	32334.5	46661.8
1996	38546.9	1812.7	50381.4	3827.9	21233	8.82	112548	114.1	1056151	32260.4	50453.5
1997	42015.6	1980.1	51238.5	3980.7	30309	8.3	112912	119.5	1161532	32434.9	49417.1
1998	45207.7	2042.1	52295.6	4083.7	25181	10.69	113787	123.2	1206867	32626.4	51229.5
2000	48996.1	2173.4	53158.4	4124.3	26731	8.23	113161	132.2	1258674	32911.3	50838.6
2001	52573.6	2421.3	53820.3	4146.4	34374	7.75	108463	128	1335446	32797.5	46217.5
2002	55172.1	2610.8	54249.4	4253.8	31743	7.71	106080	127.5	1449286	32451	45263.7
2003	57929.9	2993.4	54354.8	4339.4	27319	7.17	103891	131.2	1539484	31990.6	45705.8

5. Results and discussion

In GRA evaluation model, experimental data are first normalized in the range between zero and one, which also called grey relational generation. Subsequently, the grey relational coefficient is calculated from the normalised experimental data to express the relationship between the desired and actual experimental data. The Grey relational grade (GRG) is then computed by averaging the grey relational coefficient corresponding to each response. The overall evaluation of the multiple process response is based on the GRG.

In this section, discussion on the final result given by GRA analysis only will be presented. Thus, Table 2 demonstrates the calculated value of grey relational grade (GRG) for each affecting factors (x_i) used in this study to China's crop yield (x_0). From Table 2, we can see that GRG value are in the range of [0,1] and each affecting factors and crop yield have positive correlation. This indicates that each affecting factors influence the production of China grain crop.

Table 2 : Grey Relational Grade for each affecting factor (non- ordered)

Affecting factors (x_i)	(x_0)	GRG : $\gamma_i = \gamma(x_0, x_i)$	orders	Affecting factors (x_i)	(x_0)	GRG : $\gamma_i = \gamma(x_0, x_i)$	orders
a	***Crop Yield (d)***	0.788	7	g	***Crop Yield (d)***	0.741	8
b		0.801	5	h		0.809	3
c		0.803	4	l		0.830	1
e		0.822	2	m		0.800	6
f		0.737	9	n		0.717	10

Based on the calculated value of grey relational grade, the grey relational order of each affecting factors can be determined. Therefore, a ranking scheme based on the size of γ_i is constructed as shown in Table 3 and Figure 1. Each γ_i is ordered to the increasing grey relational coefficient. This derived order then gives the priority list in choosing the affecting factors that are closely related to the crop yield, x_o. Since the GRG represent the level of correlation between the crop yield sequence and affecting factors sequence, the greater value of GRG means that the affecting factor sequence has a stronger correlation to the crop yield sequence [14]. Therefore, from Table 2, we can rank the values of $\gamma(x_0, x_i)$ as in Table 3.

Table 3: Grey relational grade for each affecting factors (ordered sequence)

Affecting factors (x_i)	(x_0)	GRG : $\gamma_i = \gamma(x_0, x_i)$	orders	Affecting factors (x_i)	(x_0)	GRG : $\gamma_i = \gamma(x_0, x_i)$	orders
l	***Crop Yield (d)***	0.830	1	m	***Crop Yield (d)***	0.800	6
e		0.822	2	a		0.788	7
h		0.809	3	g		0.741	8
c		0.803	4	f		0.737	9
b		0.801	5	n		0.717	10

Or it can be written as ranking scheme (Figure 1):

$0.830(l) > 0.822(e) > 0.809(h) > 0.803(c) > 0.802(b) > 0.800(m) > 0.788(a) > 0.741(g) > 0.737(f) > 0.717(n)$

Less important

Figure 1: A ranking scheme for China Crop Yield

Thus based on the GRG values that present the relevancy of each affecting factors for crop yield we can said that the consumption of pesticide (l) affects the crop production the most, followed by the consumption of chemical fertilizer (e) and sown area of grain crops (h) accordingly. Subsequently, the irrigation area (c), electricity consumed in rural areas (b), consumption of agricultural film (m), total power of agricultural (a), and budgetary expenditure for agriculture (g) gave the moderate influence to the crop yield. While, the areas affected by natural disaster (f), and agriculture labourers (n) gave the least effect to the pattern of crop yield.

Hence the China agriculture department should focus their attention more on these influential factors in order to monitor the performance of China grain crop production more effectively. These influential factors also can be employed in forecasting model for more accurate result. On the contrary, the forecaster can pay less attention to the less influential factors as these elements contribute relatively little to the variation of grain crop yield. Therefore, we excluded all the input factors that less than 0.8 because they are less important. As the result, the application of GRA has reduced the number of affecting factors that affect the production of China grain crop from 10 to 6.

Tables 4 illustrate the affecting factors that yield by applying GRA, which have the greatest influence for annual grain crop yield. Based on the calculated value of grey relational grade, only six factors; a, b, c, e, h and l are selected as the inputs to ANN to predict the grain crop yield. Then these factors are fed up in ANN forecasting model to forecast the annual crop yield for year 2002 and 2003.

Table 4: Affecting factors for China's Grain Crop Yield selected by GRA

Year	a	b	c	e	h	l	d
1990	1.0000	1.0000	1.0000	1.0000	0.0324	1.0000	0.8578
1991	0.9767	0.9448	0.9397	0.8772	0.1488	0.9525	1.0000
1992	0.9452	0.8779	0.8293	0.8057	0.3261	0.8947	0.9043
1993	0.8936	0.8137	0.8094	0.6789	0.3312	0.8098	0.7247
1994	0.8257	0.7071	0.8049	0.5840	0.5298	0.5823	0.8726
1995	0.7464	0.6225	0.7298	0.4263	0.3766	0.3990	0.5932
1996	0.6633	0.5494	0.5716	0.2924	0.1252	0.3073	0.1008
1997	0.5446	0.4715	0.4483	0.2051	0.0884	0.2156	0.2354
1998	0.4354	0.4427	0.2962	0.1462	0.0000	0.1528	0.0000
2000	0.3057	0.3816	0.1721	0.1230	0.0633	0.0000	0.0508
2001	0.1833	0.2662	0.0769	0.1103	0.5380	0.0713	0.6509
2002	0.0944	0.1780	0.0152	0.0489	0.7788	0.0798	0.7747
2003	0.0000	0.0000	0.0000	0.0000	1.0000	0.0170	0.7173

The network structure used to train ANN forecasting model is 6:12:1, where 6 represent six input nodes (six affecting factors: a, b, c, e, h and l), 12 represents twelve hidden nodes and 1 represent one output node; in this case, grain crop yield.

Table 5 and Figure 2 illustrate the forecasting values produce by ANN model.

Table 5: Forecasting values produce by ANN model

Predicted Year	Actual data	ANN	Error (ANN)
2002	45263.70	45088.35	175.35
2003	45705.80	45142.26	563.55
		RMSE	417.33

Root mean square error (RMSE) is used to evaluate the forecasting performance. In this experiment, RMSE for ANN model is 417.33. From Table 5 and Figure 2, the production of China crop yield is increased about 0.9% from year 2002 to year 2003. Meanwhile, ANN forecasting model predict there is a slight increment (0.2%) from 2002 to 2003.

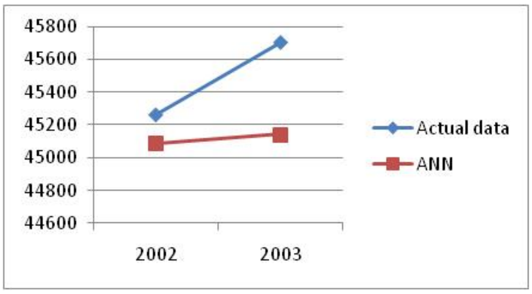

Figure 2: China's grain crop yield forecasted by ANN model.

However, these forecasting values are acceptable since ANN model has more than 95% accuracy performance. Moreover, ANN model can predict the movement of grain crop yield.

6.0 Conclusion

GRA is part of grey system theory, which is suitable for solving complicated interrelationships between multiple factors and variables; in this case, how influential, affecting factors such as the consumption of pesticide (l), the consumption of chemical fertilizer (e), sown area of grain crops (h), the irrigation area (c), electricity consumed in rural areas (b), consumption of agricultural film (m), total power of agricultural (a), budgetary expenditure for agriculture (g), the areas affected by natural disaster (f), and agriculture labourers (n) gave effect to control variables; China's crop yield.

This approach is based on the level of similarity and variability among all factors to establish their relation. The GRA suggest how to make prediction and decision, and generate reports that make suggestions how to choose the affecting factors. This analytical model magnifies and clarifies the grey relation among all factors. It also provides data to support quantification and comparison analysis. In other words, the GRA is a method to analyze the relational grade for discrete sequences.

This is unlike the traditional statistic analysis handling the relation between variables. Some of it defects are: it must have plenty of data, data distribution must be typical, a few factors are allowed and can be expressed functionally. But the GRA require less data and can analyze many factors that can overcome the disadvantages of statistics methods.

Outcome from the GRA has shown that, the order of importance of the affecting factors to the China's grain crop yield is the consumption of pesticide (l), the consumption of chemical fertilizer (e), sown area of grain crops (h), the irrigation area (c), electricity consumed in rural areas (b), consumption of agricultural film (m), total power of agricultural (a), budgetary expenditure for agriculture (g), the areas affected by natural disaster (f), and agriculture labourers (n). Experimental results have shown clearly that the consumption of pesticide (l) is the most influential factors to China's grain crop and the labour is the least influential factors that will affect the China's crop production. The GRA of GRG also provides us an alternative to decide which input factors show the crucial effect to the crop productions. In this paper, only six input factors; the consumption of pesticide (l), the consumption of chemical fertilizer (e), sown area of grain crops (h), the irrigation area (c), electricity consumed in rural areas (b), and consumption of agricultural film, are considered as the most influential factors and then they are being chosen as input node to ANN forecasting value.

Reference

[1]. Kung, C-F, Wen, K.-L, Applying GRA and Grey Decision Making to evaluate the relationship between company attributes and its financial performance- A case study of ventures capital in Taiwan, Decision Support Systems 43(2007), 842-852

[2]. J.Moran, E.Granada, J.L.Migiez, J.Porteiro, Use of grey relational analysis to asses and optimize small biomass boilers, Fuel Processing Technology 87(2006) 123-127

[3]Kun-lin Wen, The grey system analysis and its application in gas breakdown and var compensator finding(invited paper), Int. Journal of computational Cognition 2(1)(2004) 21-44.)

[4] Tsai,C-H, Chang,C-L, Chen,L, Applying grey relational analysis to the vendor evaluation model, Int. J. Of the computer; The Internet and management, Vol. 11, No. 3,2003, pp45-53.

[5]. Y. Zhang. &.Y. He, Study Of Prediction Model On Grey Relational BP Neural Network Based On Rough Set Proceedings Of The Fourth International Conference On Machine Learning And Cybernetics, Guangzhou, August 2005, pp 18-21.

[6].Ziliang.W, and Sifeng.L. Extension of Grey Superiority Analysis, Liu S.F. et al. Grey system Theory and its application, Science Press, Beijing, China,2004).

[7].K.C Chang and M-F, Yeh, Grey-relational analysis based approach for data clustering, IEEE Proc,_Vis. Image Signal Process., Vol.152,No,2, April 2005, pp165-172.

[8].Lu, H.C., Yeh,M-F.: Robot path planning based on modified grey analysis, Cybern. Syst., 2002, 33,(2), pp 129-159.

[9]. Yeh, M.-F., and Lu, H.C.: Evaluating weapon systems based on grey relational analysis and fuzzy arirhmetic operations, J. Chinese Inst. Eng., 2000, 23(2),pp.211-221

[10] Tosun N, Determination of optimum parameters for multi-performance characteristics in drilling by using grey relational analysis, Int. J. Adv. Manuf. Technol (2005), DOI 10.1007/s00170-004-2386-y

[11]Lin ZC, Ho CY, Analysis and application of grey relation and anova in chemical mechanical polishing process parameters. Int. J. Adv Manuf Technol 21,2003:10-14

[12].Lo SP (2002), The application of ANFIS and grey system method in turning failure detection. Int. J. Adv Manuf Technol 19:564-572

[13] Fung, CP (2003), Manufacturing process optimization for wear property of fibre reinforced polybutylene terephalate composites with grey relational analysis. Wear 254:298-306

[14]. Fu.C, Zheng.J, Zhao.J, and Xu.W, Application of grey relational analysis for corrosion failure of oil tubes, Corrosion Science 43(2001), 881-889

[15]. Lin, S.J., Lu,I.J, Lewis.L, Grey Relation performance Correlations Among Economics, energy use and carbon dioxide emission in Taiwan, Energy Policy 35(2007), 1948-1955

[16] Chiang B-C, Tsai S-L and Wang C-C, Machine Vision-Based Gray Relational Theory Applied to IC marking Inspection, IEEE Transactions on Semiconductor Manufacturing, Vol.15, N0.4, 2002