Food Control 99 (2019) 48–56

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

An improved ISM method based on GRA for hierarchical analyzing the influencing factors of food safety

Xiaoyong Lin[a,b], Shiying Cui[a,b], Yongming Han[a,b,*], Zhiqiang Geng[a,b,**], Yanhua Zhong[c]

[a] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education in China, Beijing, China
[c] Jiangmen Polytechnic, Jiangmen, Guangdong, 529020, China

A R T I C L E I N F O

Keywords:
Grey relational analysis (GRA)
Interpretative structural modeling (ISM)
Food safety
Risk early warning
Infant formula
Sterilized milk

A B S T R A C T

Food safety is closely related to economic development and daily life of people. The valid food safety and risk warning contribute to the social health and sustainable development. However, the inspection data of food safety is characterized by high complexity, high dimension and non-linearity. Therefore, an improved interpretative structural modeling (ISM) method based on the grey relational analysis (GRA) (GRA-ISM) is proposed to hierarchical analyze influencing factors of food safety. The correlation coefficient between the influencing factors is calculated by the GRA. Then the ISM is used for stratifying and establishing the multi-hierarchical structure of influencing factors of food safety. Finally, the infant formula data and the sterilized milk data of food safety in China are hierarchical analyzed by the GRA-ISM. The multi-hierarchical structure model of different factors affecting food safety can be obtained. The Student's t-test (*t*-test) is used to verify the validity of the threshold selection and the result of the GRA-ISM. Meanwhile, through the early warning analysis of the major factors, the proposed method can guide relevant departments to strengthen supervision and urge enterprises to work safely.

1. Introduction

Food safety affects people's health, which has profound social significance. In recent years, food safety accidents have happened frequently all around the world (Chiou, Leung, Lee, & Wong, 2015; Lam, Remais, Fung, Xu, & Sun, 2013). The outbreak of toxic eggs in Europe in 2017 has been reported in nearly 20 countries and regions, including the Netherlands, Belgium and Germany (Liu, 2017). And a series of food safety incidents have happened in China, such as Sudan red (Liu, Hei, He, & Li, 2011), melamine (Xiu & Klein, 2010) and gutter oil (Lu & Wu, 2014). The accidents happened have hampered social stability and economic development. The national food unqualified situation report of 2012–2017 is issued by the China State Bureau of Technical Supervision as shown in Fig. 1, which shows that the number and proportion of unqualified food batches in China are on the rise from 2012 to 2017 with a small fluctuation. Hence, it is critical to find the major factors affecting food safety and food safety problems in advance.

By analyzing the influencing factors of food safety, the source control of major factors can be realized. Consider the grim situation of dairy food safety in China (Yan, 2012), a new interpretative structural modeling (ISM) method based on the grey rational analysis (GRA) (GRA-ISM) for hierarchical analyzing the influencing factor of food safety is proposed in this paper. Firstly, the correlation coefficient between influencing factors is calculated by the GRA, and the strength and weakness of the mutual relations between the factors is quantified simultaneously. Then the ISM is used to stratify the influencing factors of food safety, and the main influencing factors are obtained. Finally, the infant formula data and the sterilized milk data of food safety in China are hierarchical analyzed by the GRA-ISM. Though the Student's t-test (*t*-test), the validity of the threshold selection and the result of the GRA-ISM is verified. Meanwhile, the experimental results show that the GRA-ISM can analyze the factors affecting food safety better.

The organizational structure of this paper is as follows. The current research of the ISM method and the GRA are introduced in section2. Section 3 is the detailed description of the GRA-ISM method. Section 4 is the case study for hierarchical analyzing the influencing factors of the infant formula and the sterilized milk. Section 5 is the discussion of the experiment results and the proposed method. Finally, the conclusion is given in Section 6.

* Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, China.
** Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing, China.
E-mail addresses: hanym@mail.buct.edu.cn (Y. Han), gengzhiqiang@mail.buct.edu.cn (Z. Geng).

https://doi.org/10.1016/j.foodcont.2018.12.020
Received 8 November 2018; Received in revised form 13 December 2018; Accepted 17 December 2018
Available online 18 December 2018
0956-7135/

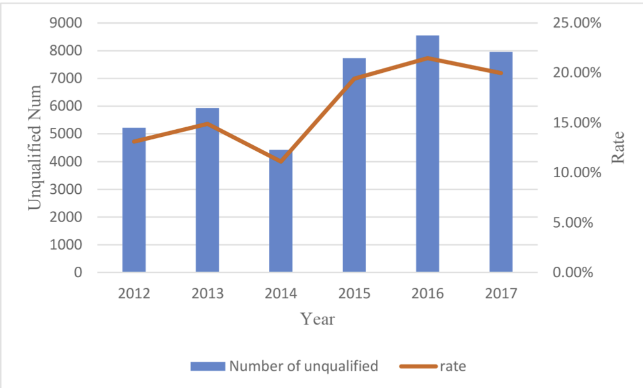

Fig. 1. Statistics of China's food failures from 2012 to 2017.

2. Relate work

Due to food safety incidents occur frequently, food safety research has become the focus of the global research (Perrot, Trelea, Baudrit, Trystram, & Bourgine, 2011). At present, the Artificial Neural Network (ANN) (Geng, Shang, Han, & Zhong, 2019) and the Bayesian Network (BN) (Greco, Landoni, Biondi-Zoccai, D'Ascenzo, & Zangrillo, 2016) are the most mature research methods on food safety. The ANN is an effective computing model and widely used in classification of food species (da Silva, Filardi, Pepe, Chaves, & Santos, 2015) and food quality (Argyri, Panagou, Tarantilis, Polysiou, & Nychas, 2010), the element content detection (Özbalci, Boyaci, Topcu, Kadılar, & Tamer, 2013), risk assessment (Oscar, 2009) and risk early warning (Koyuncugil & Ozgulbas, 2012). However, the modeling ability of the ANN algorithm is limited. The accuracy of image recognition and other complex tasks is poor (Hajihassani, Armaghani, Marto, & Mohamad, 2015). The BN is a kind of probability network (Williams, Ebel, & Vose, 2011), which is widely used in risk assessment (Smid, Verloo, Barker, & Havelaar, 2010) and food fraud type predication (Bouzembrak & Marvin, 2016; Marvin et al., 2016). However, due to the food safety data is characterized by high dimension and high complexity, the analysis ability of the BN is limited (Ercsey-Ravasz, Toroczkai, Lakner, & Baranyi, 2012).

The ISM is often used to hierarchical analyze the influencing factors of complex system, which is decomposed into several subsystem elements to reflect the structural relationship between elements (Shen, Song, Wu, Liao, & Zhang, 2016). In 1973, Warfelt designed and developed the ISM method to analyze the problems related to complex systems. In other words, a multi-hierarchical structure model was constructed by using people's practical experience and knowledge. Singh and Kant (Singh & Kant, 2008) used the ISM method to analyze the relationship between the identified barriers of knowledge management. Sagheer et al. (Sagheer, Yadav, & Deshmukh, 2009) determined and analyzed the key factors affecting the compliance of standards in the food industry by using the ISM method. This ISM used the expert experience method to construct key influencing factors, which is less objective. Han et al. (Han, Geng, Zhu, & Lin, 2015) built the correlation coefficient matrix of the ISM by using the partial correlation analysis method, and the feasibility of these methods was verified with the ethylene data. However, the ISM method which the correlation coefficient matrix established by using expert experience method or partial correlation analysis has poor accuracy and is not suitable for analyzing the nonlinear data.

The GRA method belongs to geometric processing (Liu, Cai, Yang, & Cao, 2013), which refers to the geometric comparison of data series reflecting the changing characteristics of various factors. Thus the GRA is suitable for analyzing and processing the nonlinear data (Mu & Zhang, 2008). The GRA was proposed by Deng in 1985 (Tan & Deng, 1995). Chan and Tong (Chan & Tong, 2007) proposed an integrated method for selecting material. Due to non-linear constraints, this method ranked the materials by using the GRA. Hashemi et al. (Hashemi, Karimi, & Tavana, 2015) established the rational coefficient matrix using the GRA and proposed a green supplier selection method to deal with the supplier selection uncertainties. Therefore, the GRA method can also effectively realize the correlational analysis of the nonlinear data.

Hence, due to the characteristics of high dimension and high complexity in the food safety data, the GRA-ISM is proposed to hierarchical analyze the influencing factors of food safety.

3. Interpretative structural modeling based on grey relational analysis (GRA-ISM)

3.1. The grey relational analysis (GRA)

Correlation coefficient can measure the degree of linear correlation between two random factors (Puth, Neuhäuser, & Ruxton, 2014). Generally, the more consistent the change tendency of the two variables, the higher the degree of correlation between the two variables. On the contrary, the degree is low. The correlation coefficient matrix is established by using the GRA.

Let the reference sequence be $x_0 = \{ x_0 (1), x_0 (2), \ldots, x_0(n)\}$, where n is the number of samples, and the comparison sequence is $x_i = \{x_i(1), x_i(2), \ldots, x_i(n)\}$ $(1 = 0,1,2, \ldots, m\text{-}1)\}$, where m is the number of all indicators. Every indicator is served as a reference sequence and the rest as a comparison sequence.

The m group indicators are normalized for eliminating the effect of dimensions, which is shown in Eq. (1).

$$y_i(k) = \frac{x_i(k)}{\frac{1}{n}\sum_{m=1}^{n} x_i(m)} \tag{1}$$

Then the grey correlation coefficients of $y_i(k)$ and $y_0(k)$ at the moment k are calculated as follows:

$$\xi_i(k) = \frac{\min_i \min_k |y_0(k) - y_i(k)| + \rho \max_i \max_k |y_0(k) - y_i(k)|}{|y_0(k) - y_i(k)| + \rho \max_i \max_k |y_0(k) - y_i(k)|} \tag{2}$$

Where, $\xi_i(k)$ is a gray correlation coefficient, and the adjustment parameter ρ can make the difference of each coefficient enhanced, and$\rho \in (0,1)$. The correlation coefficient between the two sequences y_0 and y_i can be expressed in Eq. (3):

$$\gamma(y_0, y_i) = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k) \tag{3}$$

Each indicator is served as a reference sequence, and the correlation coefficient matrix of all indicators can be obtained as Eq. (4):

$$\gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{bmatrix}_{n*n} \tag{4}$$

Generally, the absolute value of the correlation coefficient corresponds to the strength of the correlation, and the larger the value, the stronger the correlation, and conversely, the weaker the correlation (Adler & Parmryd, 2010). The relationship between correlation coefficient values and correlation degrees is shown in Fig. 2.

3.2. The interpretative structural modeling (ISM)

The complex system is decomposed into several subsystems and forms the multi-hierarchical structure by using the ISM. The ISM can transform vague thoughts and views into intuitive models with good structural relations (Poduval & Pramod, 2015). It is especially suitable for the system analysis with many variables, complex relations and unclear structures.

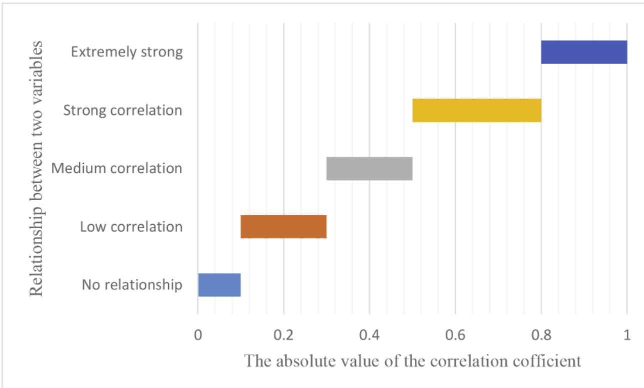

Fig. 2. The relationship between correlation coefficient values and correlation degrees.

3.2.1. The adjacent matrix

The adjacency matrix is obtained by comparing the magnitude of correlation coefficient with the intensity threshold. When γ_{ij} is larger than the threshold value, the adjacency value of x_i to x_j is $a_{ij} = 1$ and $a_{ji} = 0$ (i = 0,1,2, …,n; j = 0,1,2, …,n). Otherwise, $a_{ij} = 0$ and $a_{ji} = 1$. The adjacency matrix is shown in Eq. (5).

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n-1} & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{n-1n-1} & a_{nn} \end{bmatrix}_{n\times n} \tag{5}$$

3.2.2. The reachable matrix

Let

$$E = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{n\times n} \tag{6}$$

be the identity matrix, then

$$A + E = (A + E)^2 = \dots = (A + E)^{n-1} = (A + E)^n \tag{7}$$

And $R = A + E^{n-1}$ is the reachable matrix of A.

$$R = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n-1} & R_{1n} \\ R_{21} & R_{22} & \dots & R_{2n-1} & R_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ R_{n1} & R_{n2} & \dots & R_{n-1n-1} & R_{nn} \end{bmatrix}_{n\times n} \tag{8}$$

3.2.3. Modeling the ISM

Definition 1. In the reachable matrix R, the set of variables corresponding to the column in which the element R_i having all elements in the i-th row with 1 is a reachable set, and the symbol is represented asS_i.

Definition 2. In the reachable matrix R, the set of variables corresponding to the row in which the element R_j having all the elements in the j-th column with 1 is a leading set, and the symbol is represented as B_j.

According to the judgment rule of $S_i \cap B_j = S_i$, the highest layer factor set L_1 is determined by the hierarchical division of the influencing factors. After obtaining the highest layer element, the corresponding rows and columns of the highest layer element L_1 in the reachable matrix are temporarily deleted. The same decision rule determines the layer elements L_2, L_3,...,L_k. The ISM hierarchical structure of the influencing factors is obtained from the obtained hierarchical elements L.

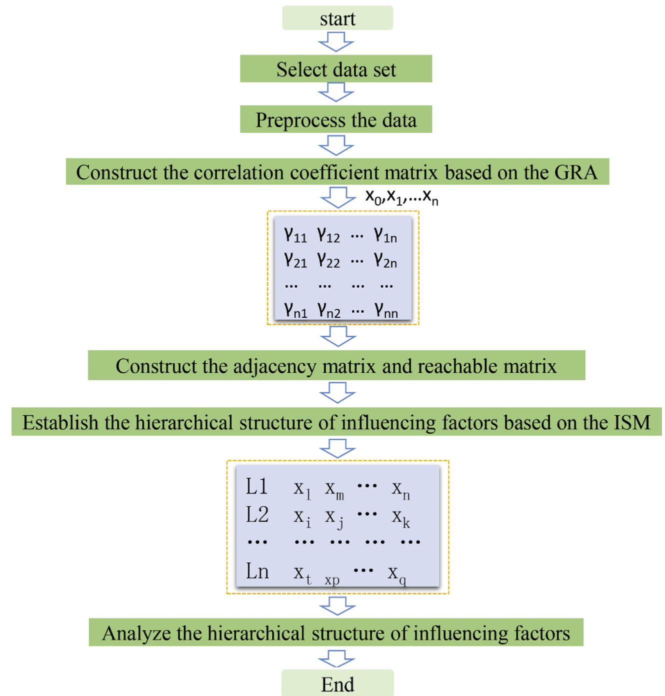

Fig. 3. The flow chart of the GRA-ISM.

3.3. The structural analysis based on the GRA-ISM

The structural analysis steps of the GRA-ISM method is shown as follows.

Step 1: Input and preprocess the data.
Step 2: Construct the correlation coefficient matrix based on Eqs. (1)–(4) by using the GRA.
Step 3: Construct adjacency matrix A and reachable matrix R based on Eqs. (5)–(8).
Step 4: Obtained the elements of every level by the ISM method.

The flow chart of the GRA-ISM is shown in Fig. 3.

4. Case study: the influencing factors analysis of food safety

This paper analyzes the influencing factors of food safety. The used data is from a food inspection agency in a province in China. Firstly, the GRA is used to establish the correlation coefficient matrix of each factor. Then the ISM is used to stratify the influencing factors, and the multilevel hierarchical structure of the factors affecting the food safety is established. The inspection data of infant formula and sterilized milk are used in this paper.

4.1. Data preprocessing

4.1.1. Data integration

Raw data gathered by the food inspection agency are adopted. The inspection data of infant formula data and sterilized milk is selected as the research object. Technical requirements of national food safety standard of the infant formula GB 10765–2010 (National food safety standards of China, 2010a, 2010b) include ten points, with raw material requirements, sensory requirements, essential ingredients, optional ingredients, other indicators, limit of pollutants, limit of fungal toxins, limit of microorganisms, food additives and nutritive fortifier and urease activity. Technical requirements of national food safety standard of the sterilized milk GB 25190–2010 (National food safety standards of China, 2010a, 2010b) include six points, with raw material

Table 1
The evaluation index of infant formula.

Item	Index				Testing method
	Per 100 kJ		Per 100 kcal		
	Minimum	Maximum	Minimum	Maximum	
Protein/(g)	0.45	0.70	1.88	2.93	GB 5009.5
Fat/(g)	1.05	1.40	4.39	5.86	GB 5413.3
Vitamin A/(μg RE)	14	43	59	180	GB5413.9
Vitamin E/(mg α-TE)	0.12	1.20	0.50	5.02	
Vitamin C/(mg)	2.5	17.0	10.5	50.2	GB5413.18
Pantothenic acid/(μg)	96	478	402	2000	GB5413.17
Sodium/(mg)	5	14	21	59	GB5413.21
Iron/(mg)	0.10	0.36	0.42	1.51	
Copper/(μg)	8.5	29.0	35.6	121.3	
Magnesium/(mg)	1.2	3.6	5.0	15.1	
Zinc/(mg)	0.12	0.36	0.50	1.51	
Calcium/(mg)	12	35	50	146	
Phosphorus/(mg)	6	24	25	100	GB5413.22
Iodine/(μg)	2.5	14.0	10.5	58.6	GB5413.23
Chloride/(mg)	12	38	50	159	GB5413.24
Moisture content/(%) ≤	5.0				GB 5009.3
Ash content/(%) ≤	4.0				GB 5009.4

Table 2
The evaluation index of sterilized milk.

Item	Index	Testing method
Fat/(g/100g) ≥	3.1	GB5413.3
Protein/(g/100g) ≥	2.9	GB5009.5
Solids-non-fat/(g/100g) ≥	8.1	GB5413.39
Acidity/(°T)	12–18	GB5413.34
Lead/(mg/kg)	0.05	GB2762
Mercury/(mg/kg)	0.01	GB2762
Arsenic/(mg/kg)	0.1	GB2762
Chromium/(mg/kg)	0.3	GB2762
Aflatoxin M1/(μg/kg)	0.5	GB2761

Table 3
The data after preprocessing of the infant formula.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Protein/(g/100 kJ)	0.6	0.597	0.61	0.61	0.62
Fat/(g/100 kJ)	1.22	1.22	1.22	1.23	1.21
Vitamin A/(μg RE/100 kJ)	35	23	25	21	28
Vitamin E/(mg α-TE/100 kJ)	0.54	0.653	0.63	0.6	0.58
Vitamin C/(mg/100 kJ)	5.3	7.8	8	7.9	6.8
Pantothenic acid/(μg/100 kJ)	242	274	156	141	166
Sodium/(mg/100 kJ)	7	9	10	8	9
Iron/(mg/100 kJ)	0.26	0.248	0.23	0.24	0.23
Copper/(μg/100 kJ)	16.4	19.4	11.6	12.3	13.9
Magnesium/(mg/100 kJ)	2.2	2.45	2	1.7	1.8
Zinc/(mg/100 kJ)	0.22	0.191	0.2	0.18	0.23
Calcium/(mg/100 kJ)	21	20	17	17	19
Phosphorus/(mg/100 kJ)	13	13.5	16	13	13
Iodine/(μg/100 kJ)	5.6	5	4	5.4	5.3
Chloride/(mg/100 kJ)	16	18	19	19	20
Moisture content/(%)	2.3	2.61	1.6	2	1.9
Ash content/(%)	2.8	2.8	2.8	2.7	2.9

requirements, sensory requirements, physical and chemical indicators, limit of pollutants, limit of fungal toxin and requirements of microorganisms.

Because the collected data contains a large amount of product-related information, critical data is dispersed across all database tables.

Table 4
The data after preprocessing of the sterilized milk.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Fat/(g/100g)	3.73	3.88	3.71	3.64	3.94
Protein/(g/100g)	3.22	3.26	3.13	3.36	3.18
Solids-non-fat/(g/100g)	8.81	8.66	8.75	9.08	8.51
Acidity/(°T)	13.23	13.54	12.83	14.4	13.9
Lead/(mg/kg)	0	0	0.008	0	0.01
Mercury/(mg/kg)	0	0	0	0.00121	0
Arsenic/(mg/kg)	0	0	0.11	0	0
Chromium/(mg/kg)	0.019	0.022	0.0027	0.0037	0.00093
Aflatoxin M1/(μg/kg)	0.0385	0.025	0	0	0

Information about attributes that are not closely related, such as production date, origin, and inspector information, should be ignored while integrating key attributes. According to the key attributes combined with the above national standards, influencing factors of the quality of infant formula and sterilized milk are obtained as shown in Table 1 and Table 2. There are 16 influencing factors of the infant formula, and fat, protein, vitamin, pantothenic acid and the minerals belong to the essential ingredients, and moisture content and ash content belong to the other indicators. There are 9 influencing factors of the sterilized milk, and fat, protein, solids-non-fat and acidity are physical and chemical indicators, and aflatoxin M1 belongs to mycotoxins, and the rest belongs to pollutants.

4.1.2. Data integration

In the process data of dairy products, the data appears as a complex nonlinear time series relationship. Due to the particularity of the food data, and the data including missing value data and outlier data, the data needs to be integrated.

The missing value data is handled. Data of missing value types are mainly represented by low level censored data. Low level censored data is produced when the pollutant content is sometimes below the detection limit (LOD) and quantitative limit (LOQ) of the method or equipment, and the measured data showed "not detected". This data mainly exists in the monitoring of chemical contaminants in food. For low-level censored data, the methods of replacement, maximum likelihood estimation, and order statistic regression are used.

The outlier data is processed. The cause of outlier data includes three aspects, with the data from different sources, data measurement

Table 5
The correlation coefficient matrix of infant formula data in group a.

	Protein	Fat	Vitamin A	Vitamin E	Vitamin C	Pantothenic acid	Sodium	Iron	Copper	Magnesium	Zinc	Calcium	Phosphorus	Iodine	Chloride	Moisture content	Ash content
Protein	1.00	0.77	0.68	0.85	0.79	0.75	0.76	0.85	0.73	0.77	0.74	0.88	0.87	0.72	0.78	0.79	0.80
Fat	0.77	1.00	0.69	0.85	0.80	0.81	0.71	0.73	0.75	0.75	0.75	0.78	0.71	0.81	0.72	0.79	0.73
Vitamin A	0.68	0.69	1.00	0.79	0.76	0.76	0.70	0.72	0.70	0.71	0.72	0.78	0.72	0.76	0.69	0.78	0.71
Vitamin E	0.85	0.85	0.79	1.00	0.81	0.81	0.78	0.81	0.78	0.79	0.77	0.84	0.79	0.76	0.75	0.85	0.81
Vitamin C	0.79	0.80	0.76	0.81	1.00	0.75	0.68	0.72	0.68	0.66	0.67	0.77	0.71	0.76	0.78	0.75	0.72
Pantothenic acid	0.75	0.81	0.76	0.81	0.75	1.00	0.70	0.73	0.79	0.79	0.80	0.80	0.75	0.73	0.68	0.86	0.74
Sodium	0.76	0.71	0.70	0.78	0.68	0.70	1.00	0.82	0.81	0.81	0.72	0.85	0.81	0.75	0.73	0.79	0.76
Iron	0.85	0.73	0.72	0.81	0.72	0.73	0.82	1.00	0.80	0.82	0.75	0.91	0.86	0.74	0.73	0.81	0.80
Copper	0.73	0.75	0.70	0.78	0.68	0.79	0.81	0.80	1.00	0.82	0.75	0.82	0.75	0.71	0.67	0.81	0.69
Magnesium	0.77	0.75	0.71	0.79	0.66	0.79	0.81	0.82	0.82	1.00	0.80	0.86	0.82	0.73	0.67	0.83	0.74
Zinc	0.74	0.75	0.72	0.77	0.67	0.80	0.72	0.75	0.75	0.80	1.00	0.79	0.75	0.72	0.68	0.82	0.73
Calcium	0.88	0.78	0.78	0.84	0.77	0.80	0.85	0.91	0.82	0.86	0.79	1.00	0.85	0.71	0.70	0.81	0.78
Phosphorus	0.87	0.71	0.72	0.79	0.71	0.75	0.81	0.86	0.75	0.82	0.75	0.85	1.00	0.66	0.74	0.78	0.75
Iodine	0.72	0.81	0.76	0.76	0.76	0.73	0.75	0.74	0.71	0.73	0.72	0.71	0.66	1.00	0.76	0.76	0.71
Chloride	0.78	0.72	0.69	0.75	0.78	0.68	0.73	0.73	0.67	0.67	0.68	0.70	0.74	0.76	1.00	0.74	0.72
Moisture content	0.79	0.79	0.78	0.85	0.75	0.86	0.79	0.81	0.81	0.83	0.82	0.81	0.78	0.76	0.74	1.00	0.75
Ash content	0.80	0.73	0.71	0.81	0.72	0.74	0.76	0.80	0.69	0.74	0.73	0.78	0.75	0.71	0.72	0.75	1.00

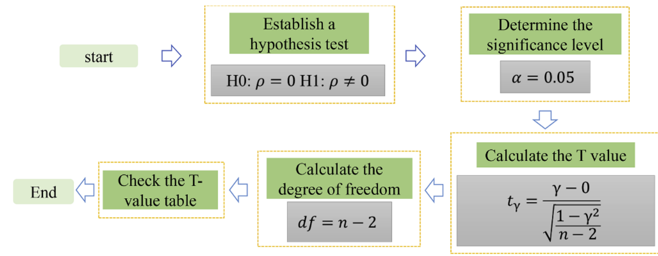

Fig. 4. The calculation process of *t*-test.

Table 6
The calculation results of *t*-test.

	γ	n	df	t_γ	P value
Infant formula [a]	0.82	99	97	14.111	$1.52*10^{-25}$
Infant formula [b]	0.82	87	85	13.209	$1.32*10^{-22}$
Sterilized milk	0.90	62	60	15.994	$5.56*10^{-24}$

and collection error. The preprocessing of the outlier includes two steps with detecting the outliers existing in the record and processing the known outliers. The box line diagram is used to detect the outliers. The methods for dealing with outliers include deleting, retaining, replacing with mean or other statistics, and treating them as missing values. The specific processing method is judged by the specific situation.

Part of the data after preprocessing of the infant formula and the sterilized milk are shown in Tables 3 and 4.

4.2. Hierarchical analysis of the food safety factors

4.2.1. Construct correlation coefficient matrix

In this paper, 2 groups of the infant formula data and 62 detection data sets of the sterilized milk after preprocessing are used. Group a includes the infant formula data contains 99 sets of data, and group b contains 87 sets of data, group a and group b are the inspection data from different brands and batches. Each infant formula data set includes 17 attributes, and each sterilized milk data set includes 9 attributes. The normalized data and unified dimensionality are used to obtain the standardized data. And then the gray correlation coefficient of each factor with respect to other indicators is calculated by the GRA. Each indicator serves as a reference sequence, namely x_0, and the remaining indicators serve as a comparison sequence, namely $x_i (i = 0,1,2,...,m - 1)$. The matrices of correlation coefficients between indicators are obtained, and the correlation coefficient matrix of the detection data of group a of the infant formula is shown in Table 5.

As shown in Table 5, all values on the principal diagonal of the matrix are 1, that is, the correlation coefficient of each factor with respect to itself is 1. Taking the protein of Table 5 as an example, the correlation coefficient of the protein with respect to the calcium is 0.88, which is the largest compared with the residual factors, namely the strongest correlation between the protein and the calcium.

4.2.2. Modeling the ISM

As shown in Table 5, 17 factors of infant formula are correlated, so the threshold is set to 0.82. And the correlation of sterilized milk data is stronger than the infant formula data, so the threshold of the sterilized milk data is set to 0.90. And the rationality of threshold selection is verified by *t*-test.

Taking the group a of the infant formula data as an example, the hypothesis test is established, H0: $\rho = 0$, H1: $\rho \neq 0$; And the significance level value $\alpha = 0.05$ is determined. Then T value of correlation coefficient 0.82 is calculated. Given that the number of samples is

Table 7
The adjacency matrix of the infant formula data in group a.

	Protein	Fat	Vitamin A	Vitamin E	Vitamin C	Pantothenic acid	Sodium	Iron	Copper	Magnesium	Zinc	Calcium	Phosphorus	Iodine	Chloride	Moisture content	Ash content
Protein	1	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0
Fat	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Vitamin A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vitamin E	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
Vitamin C	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Pantothenic acid	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
Sodium	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
Iron	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
Copper	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Magnesium	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0
Zinc	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Calcium	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Phosphorus	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Iodine	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Chloride	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Moisture content	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Ash content	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 8
The reachable matrix of the infant formula data in group a.

	Protein	Fat	Vitamin A	Vitamin E	Vitamin C	Pantothenic acid	Sodium	Iron	Copper	Magnesium	Zinc	Calcium	Phosphorus	Iodine	Chloride	Moisture content	Ash content
Protein	1	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0
Fat	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0
Vitamin A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vitamin E	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0
Vitamin C	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Pantothenic acid	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
Sodium	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
Iron	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
Copper	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0
Magnesium	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0
Zinc	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Calcium	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Phosphorus	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Iodine	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Chloride	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Moisture content	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Ash content	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 9
Levels of the influencing factors of infant formula data in group a.

Variables	Reachable set	First set	intersection
Protein	protein, vitamin E, iron, calcium, phosphorus, moisture content	protein	protein
Fat	fat, vitamin E, calcium, phosphorus, moisture content	fat	fat
Vitamin A	vitamin A	vitamin A	vitamin A
Vitamin E	vitamin E, calcium, phosphorus, moisture content	protein, fat, vitamin E	vitamin E
Vitamin C	vitamin C	vitamin C	vitamin C
Pantothenic acid	Pantothenic acid, moisture content	Pantothenic acid	Pantothenic acid
Sodium	sodium, calcium, phosphorus	sodium	sodium
Iron	iron, calcium, phosphorus	protein, iron	iron
Copper	copper, calcium, phosphorus	copper	copper
Magnesium	magnesium, calcium, phosphorus, moisture content	magnesium	magnesium
Zinc	zinc, moisture content	zinc	zinc
Calcium	calcium, phosphorus	protein, fat, vitamin E, sodium, iron, copper, magnesium, calcium	calcium
Phosphorus	phosphorus	protein, fat, vitamin E, sodium, iron, copper, magnesium, calcium, phosphorus	phosphorus
Iodine	iodine	iodine	iodine
Chloride	chloride	chloride	chloride
Moisture content	moisture content	protein, fat, vitamin E, Pantothenic acid, magnesium, zinc, moisture content	moisture content
Ash content	ash content	ash content	ash content

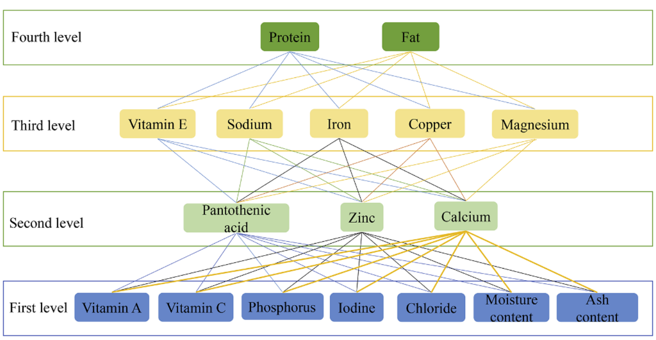

Fig. 5. The hierarchical structure of the infant formula data in group a.

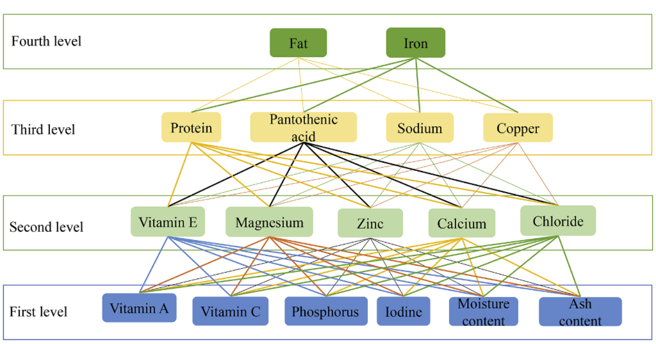

Fig. 6. The hierarchical structure of the infant formula data in group b.

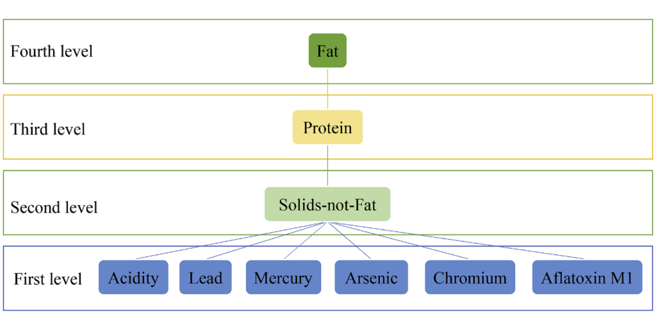

Fig. 7. The hierarchical structure of the sterilized milk data.

n = 99, t_γ is calculated, $t_\gamma = \frac{0.82-0}{\sqrt{\frac{1-0.82^2}{99-2}}} = 14.111$.Then the degree of freedom is calculated, namely $df = 99 - 2 = 97$. According to the $t_\gamma$and freedom value, T value table is checked to determine the $P = 1.52 * 10^{-25}$. Because $P < 0.05$, H0 is rejected, H1 is accepted, the threshold 0.82 is statistically significant. The calculation process of *t*-test is shown as Fig. 4. The calculation results of *t*-test of infant formula and sterilized milk are shown in Table 6, and the P values of each group are all less than 0.05, which proved that the selected threshold value is reasonable.

The infant formula data is taken as an example. When the correlation coefficient is greater than 0.82, the value of adjacency matrix is 1. And when it is less than 0.82, the value is 0. The adjacency matrices of the infant formula data and the sterilized milk data are obtained according to the above rules. The adjacency matrix of the infant formula in group a is shown in Table 7. The adjacency matrix describes the direct relationship between the elements of the system. Taking the protein in Table 7 as an example, the protein to protein, vitamin E, iron, calcium and phosphorus can be reached directly, and it is not directly accessible to the remaining indicators.

According to Eqs. (6) and (7), the reachable matrices of the infant formula and the sterilized milk data can be established. The reachable matrix of the infant formula in group a is shown in Table 8. By analyzing the reachable matrix of Table 8, taking the protein as an example, the reachable set of the protein includes protein, vitamin E, iron, calcium, phosphorus and moisture content, and the remaining indicators do not belong to its reachable set.

Through the first set B_j and the reachable set S_i and $S_i \cap B_j$ of each element, the first level can be obtained by the reachable matrix. The levels of the influencing factors are obtained by the first set, reachable set and intersection. The levels of infant formula in group a is shown in Table 9. The vitamin E in Table 9 is taken as an example, and the reachable set of vitamin E includes vitamin E, calcium, phosphorus and moisture content, and the first set of vitamin E includes protein, fat and vitamin E, and its intersection is vitamin E.

From the $S_i \cap B_j = S_i$ in Table 8, the first level elements affecting the quality of the infant formula are obtained. Then the elements included in the first level are deleted. The first sets, the reachable sets and the intersections are established again. By using the GRA-ISM, the hierarchical structures of two groups of the infant formula are shown in Figs. 5 and 6, and the hierarchical structure of the sterilized milk data is shown in Fig. 7.

The major influencing factors of the infant formula safety are obtained by analyzing the data of group a and group b as shown in Figs. 5 and -6. And the major factors affecting the two groups of data are similar. The major factors of group a include vitamin A, vitamin C, phosphorus, iodine, chloride, moisture content and ash content. And the major influencing factors of group b include all the major influencing factors excepting phosphorus in group a. In the data of group a, pantothenic acid, zinc, calcium, vitamin E, sodium, iron, copper and magnesium play a connecting role, and protein and fat are the basic influencing factors. In the data of group b, vitamin E, magnesium, zinc, calcium, chloride, protein, pantothenic acid, sodium and copper play a connecting role, and fat and iron are the basic influencing factors.

According to the hierarchical structure based on the infant formula data, 6 major factors influencing the safety of infant formula are obtained. The vitamin A and vitamin C are vitamins, and the phosphorus and iodine are minerals. Those factors are both essential ingredients in infant formula, and the quality of the infant formula can be controlled by monitoring those factors. The moisture content and ash content belong to the other indicators, so the monitoring can be enhanced by improving the detection methods of moisture content and ash content, and the quality of infant formula is guaranteed.

The major influencing factors affecting the sterilized milk safety are acidity, lead, mercury, arsenic, chromium and aflatoxin M1 by analyzing Fig. 7. The basic influencing factors are fat, protein and non-fat milk solids, which play a supporting role.

According to the hierarchical structure based on the sterilized milk data, the acidity in heavy metal pollution and mycotoxin limit and physical and chemical indicators is the most important indicator in the evaluation of sterilized milk. The residual causes of heavy metal pollution in sterilized milk include two aspects. First, the environment contains heavy metal elements for natural or human reasons, resulting in contamination of animal feed, and there are excessive levels of heavy metals in raw milk. The second reason is accidental contamination of dairy products during collection, transportation, and processing. The contamination of mycotoxins in sterilized milk mainly includes the contamination of aflatoxin M1. The aflatoxin M1 contained in the dairy products mainly comes from the feed of dairy cows. When the edible feed of cows contains aflatoxin B1, milk and dairy products will Contains aflatoxin M1. From the above analysis results, it is known that in the sterilized milk processing, the quality of raw milk needs to be strictly controlled. The environment of the pasture planting base is monitored to ensure that the air, water quality and soil of the pastoral area are not polluted, so that the planted forage do not contain pollution by heavy metals, and the quality of raw milk should be ensured.

5. Discussion

First, The GRA-ISM is proposed. The correlation coefficient matrix of the non-linear data is calculated by using the GRA. Then major factors affecting food safety can be found based on the ISM. By focusing on major factors, the proposed method can guide relevant departments to strengthen supervision and urge enterprises to work safely.

Second, this proposed method is used for effectively analyzing the major factors affecting the safety of infant formula and sterilized milk. The *t*-test is used to verify the threshold setting and the validity of the threshold is proved. Though the hierarchical structure analysis of the ISM, the evaluation index of infant formula and sterilized milk are both divided into 4 layers, and the first layer is the main influencing factor.

The main reasons affecting the quality of infant formula and sterilized milk can be found by analyzing the major factors. At the same time, similar influencing factors are obtained by comparing the data of different infant formula, which can prove the effectiveness of the results. The main factors can be provided for risk assessment and risk warning, and data basis can be provided for determining the index weight. The proposed method can reduce the occurrence of food safety accidents.

Third, there are some drawbacks in the proposed method. The threshold of the GRA-ISM is determined through *t*-test. Hence, the self-adaption artificial neural network approach will be used to optimize the threshold of the proposed method in the future work.

6. Conclusion

The improved GRA-ISM method for hierarchical analyzing the influencing factor of food safety is proposed in this paper. Taking the inspection data of infant formula and sterilized milk as examples, the correlation coefficient matrix of the factors is calculated by using the GRA. And then the ISM method is used to stratify the influencing factors and establish the multi-hierarchical structure of the influencing factors of infant formula and sterilized milk. The *t*-test is used to verify the validity of the threshold selection and prove the validity of the result. Finally, the 99 and 87 sets of the infant formula data and the 62 sets of the sterilized milk data are selected as sample sets, 17 factors of the infant formula and 9 factors influencing the safety of the sterilized milk are analyzed. By analyzing the major influencing factors and the actual production situation, the improved solution to improve the quality of infant formula and sterilized milk is obtained. Meanwhile, this proposed method can well reflect the hierarchical structure of the factors influencing the food safety, and can also be used as a basis for risk assessment and risk warning assessment criteria and early warning.

In our further work, the self-adaption artificial neural network method will be used to adaptively adjust the threshold of the GRA-ISM. Moreover, this proposed method can be widely used in other complex food safety fields.

Acknowledgement

This work is partly financial supported by National Key Research and Development Program of China (2017YFC1601800), National Natural Science Foundation of China (61673046 and 61374166), and Fundamental Research Funds for the Central Universities (XK1802-4).

References

Adler, J., & Parmryd, I. (2010). Quantifying colocalization by correlation: The pearson correlation coefficient is superior to the mander's overlap coefficient. *Cytometry, Part A, 77*(8), 733–742.

Argyri, A. A., Panagou, E. Z., Tarantilis, P. A., Polysiou, M., & Nychas, G. J. (2010). Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. *Sensors and Actuators B: Chemical, 145*(1), 146–154.

Bouzembrak, Y., & Marvin, H. J. (2016). Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling. *Food Control, 61*, 180–187.

Chan, J. W., & Tong, T. K. (2007). Multi-criteria material selections and end-of-life product strategy: Grey relational analysis approach. *Materials & Design, 28*(5), 1539–1546.

Chiou, J., Leung, A. H. H., Lee, H. W., & Wong, W. T. (2015). Rapid testing methods for food contaminants and toxicants. *Journal of integrative agriculture, 14*(11), 2243–2264.

Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z., & Baranyi, J. (2012). Complexity of the international agro-food trade network and its impact on food safety. *PLoS One, 7*(5), e37810.

Geng, Z., Shang, D., Han, Y., & Zhong, Y. (2019). Early warning modeling and analysis based on a deep radial basis function neural network integrating an analytic hierarchy process: A case study for food safety. *Food Control, 96*, 329–342.

Greco, T., Landoni, G., Biondi-Zoccai, G., D'Ascenzo, F., & Zangrillo, A. (2016). A bayesian network meta-analysis for binary outcome: How to do it. *Statistical Methods in Medical Research, 25*(5), 1757–1773.

Hajihassani, M., Armaghani, D. J., Marto, A., & Mohamad, E. T. (2015). Ground vibration prediction in quarry blasting through an artificial neural network optimized by imperialist competitive algorithm. *Bulletin of Engineering Geology and the Environment, 74*(3), 873–886.

Han, Y., Geng, Z., Zhu, Q., & Lin, X. (2015). Energy consumption hierarchical analysis based on interpretative structural model for ethylene production. *Chinese Journal of Chemical Engineering, 23*(12), 2029–2036.

Hashemi, S. H., Karimi, A., & Tavana, M. (2015). An integrated green supplier selection approach with analytic network process and improved Grey relational analysis. *International Journal of Production Economics, 159*, 178–191.

Koyuncugil, A. S., & Ozgulbas, N. (2012). Financial early warning system model and data mining application for risk detection. *Expert Systems with Applications, 39*(6),

6238–6253.
Lam, H. M., Remais, J., Fung, M. C., Xu, L., & Sun, S. S. M. (2013). Food supply and food safety issues in China. *The Lancet, 381*(9882), 2044–2053.
Liu, G. (2017). The enlightenment of the "poisonous eggs" in Europe. *Chinese Journal of Animal Science, 18*, 64–65 (in Chinese).
Liu, S. F., Cai, H., Yang, Y. J., & Cao, Y. (2013). Research progress of grey relational analysis model. *System Engineering Theory and Practice, 33*(8), 2041–2046.
Liu, R., Hei, W., He, P., & Li, Z. (2011). Simultaneous determination of fifteen illegal dyes in animal feeds and poultry products by ultra-high performance liquid chromatography tandem mass spectrometry. *Journal of Chromatography B, 879*(24), 2416–2422.
Lu, F., & Wu, X. (2014). China food safety hits the "gutter". *Food Control, 41*, 134–138.
Marvin, H. J., Bouzembrak, Y., Janssen, E. M., van der Fels-Klerx, H. J., van Asselt, E. D., & Kleter, G. A. (2016). A holistic approach to food safety risks: Food fraud as an example. *Food Research International, 89*, 463–470.
Mu, R., & Zhang, J. T. (2008). Hierarchical comprehensive evaluation based on grey relational analysis. *Systems Engineering-theory & Practice, 28*(10), 125–130 (in Chinese).
National food safety standards of China (2010a). Infant formula (GB/10765-2010). http://www.nhfpc.gov.cn/sps/s7891/201004/7c44639d854445f1963df8d61533ddee.shtml.
National food safety standards of China (2010b). Sterilized milk (GB/25190-2010). http://www.nhfpc.gov.cn/sps/s7891/201004/7c44639d854445f1963df8d61533ddee.shtml.
Oscar, T. P. (2009). General regression neural network and Monte Carlo simulation model for survival and growth of Salmonella on raw chicken skin as a function of serotype, temperature, and time for use in risk assessment. *Journal of Food Protection, 72*(10), 2078–2087.
Özbalci, B., Boyaci, İ. H., Topcu, A., Kadılar, C., & Tamer, U. (2013). Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chemistry, 136*(3–4), 1444–1452.
Perrot, N., Trelea, I. C., Baudrit, C., Trystram, G., & Bourgine, P. (2011). Modelling and analysis of complex food systems: State of the art and new trends. *Trends in Food Science & Technology, 22*(6), 304–314.
Poduval, P. S., & Pramod, V. R. (2015). Interpretive structural modeling (ISM) and its application in analyzing factors inhibiting implementation of total productive maintenance (TPM). *International Journal of Quality & Reliability Management, 32*(3), 308–331.
Puth, M. T., Neuhäuser, M., & Ruxton, G. D. (2014). Effective use of Pearson's product–moment correlation coefficient. *Animal Behaviour, 93*, 183–189.
Sagheer, S., Yadav, S. S., & Deshmukh, S. G. (2009). An application of interpretative structural modeling of the compliance to food standards. *International Journal of Productivity and Performance Management, 58*(2), 136–159.
Shen, L., Song, X., Wu, Y., Liao, S., & Zhang, X. (2016). Interpretive Structural Modeling based factor analysis on the implementation of Emission Trading System in the Chinese building sector. *Journal of Cleaner Production, 127*, 214–227.
da Silva, C. E. T., Filardi, V. L., Pepe, I. M., Chaves, M. A., & Santos, C. M. S. (2015). Classification of food vegetable oils by fluorimetry and artificial neural networks. *Food Control, 47*, 86–91.
Singh, M. D., & Kant, R. (2008). Knowledge management barriers: An interpretive structural modeling approach. *International Journal of Management Science and Engineering Management, 3*(2), 141–150.
Smid, J. H., Verloo, D., Barker, G. C., & Havelaar, A. H. (2010). Strengths and weaknesses of Monte Carlo simulation models and Bayesian belief networks in microbial risk assessment. *International Journal of Food Microbiology, 139*, S57–S63.
Tan, X. R., & Deng, J. L. (1995). Grey relational analysis. New method for multivariate statistical analysis. *Statistical Research,* (3), 46–48 (in Chinese).
Williams, M. S., Ebel, E. D., & Vose, D. (2011). Framework for microbial food-safety risk assessments amenable to Bayesian modeling. *Risk Analysis: International Journal, 31*(4), 548–565.
Xiu, C., & Klein, K. K. (2010). Melamine in milk products in China: Examining the factors that led to deliberate use of the contaminant. *Food Policy, 35*(5), 463–470.
Yan, Y. (2012). Food safety and social risk in contemporary China. *Journal of Asian Studies, 71*(3), 705–729.