Journal of Modern Mathematics and Statistics 2 (1): 12-17, 2008
ISSN: 1994-5388

The Forecast of Chinese Three Industrial Structures Under Grey Combination Data Model

[1,2]Niannian Jia and [1]Yanyou Chai
[1]College of Science, Harbin Engineering University, Harbin, P.R. China 150001
[2]School of Management, Harbin Institute of Technology, Harbin, P.R. China 150001

Abstract: At present, China is at a crucial changing period from the second step strategic target to the third step strategic target. To forecast the future trend of the three industrial structures scientifically and to draw up the development target of industry rationally have great significance in optimizing the allocations of resources and achieving sustained economic growth. In this study, we combine the spherical projection with the grey model and present a new method to forecast Chinese 3 industrial structures. Compared with the actual data, we attest to that the model presented by this study can forecast Chinese 3 industrial structures quite accurately, reflect the Chinese three industrial structures' mechanism and have rational and effective functions for analyzing trend and forecasting.

Key words: Forecast, grey model, spherical projection, three industrial structures

INTRODUCTION

Industrial structure means the composition and correlation of each industrial department of the national economy and the internal sections of each industry department. The method of the three industrial classification is presented by Australia economist Fisher (1933). The British economist Clark (1940) applies the method in the research of economic development and summarizes the general regulations of industrial structure evolution in the process of economic development. The history of economic development has indicated that the industrial structure need take a variable process from an unreasonable and lower level to a reasonable and higher level along with the economic development (Poul *et al*., 2006; Kim and Moon, 2001; David, 1996). As a result of the past foundation, the industrial structure has become a key factor of influencing the economic growth.

At present, China is at a crucial changing period from the second step strategic target to the third step strategic target. To forecast the future trend of the three industrial structures scientifically and to draw up the development target of industry rationally have great significance in optimizing the allocations of resources and achieving sustained economic growth.

The traditional industrial structure forecast and analysis are mainly paid attention to the time series analysis of one certain industry, then build an appropriate mathematical and economic model to forecast the future changing trend (Tang, 2005). However, many constraint conditions are existed among the 3 industrial structures. It means that the sum of the proportion of three industrial structures must equal to 1. For these special characters' combination data, how to build an appropriate mathematical model is being one of the most important problems of the adjustment and forecast of the three industrial structures.

In statistics the combination of each share data of a group of variable under the constraint is called composition data. We suppose that the sum of each share is equal to 1. The definition of composition data is presented by Ferers (1866), but the properties of this kind of data are very complicated. In 1978 Pearson points out the difficulties while explaining the relativity of each composition. In Aitchison (1986) a book named the statistics and analysis of composition data is published and the theories of logical normal distribution and the algorithm of logarithmic transform are considered. The proportion of three industrial structures is just like the composition data. We can use the composition data method to process the proportions of Chinese three industrial structures.

Meanwhile, we still have much unknown about Chinese 3 industrial structures. So we can consider it as a grey system (Liu *et al*., 2004). In this study, we combine the spherical projection with the GM (Deng, 2005) model and present a new forecast method to forecast Chinese 3 industrial structures (Wang *et al*., 2003;

Corresponding Author: Niannian Jia, College of Science, Harbin Engineering University, Harbin, P.R. China 150001

Shi and Chai, 2007). Compared with the actual data, the model presented by this study can forecast Chinese 3 industrial structures quite accurately, reflect the Chinese three industrial structures' mechanism and have rational and effective functions for analyzing trend and forecasting.

MATHEMATICAL MODELS

GM (1, 1) model in grey system

GM (1, 1) model: Grey models play the core function in grey system theory. For a full name, they should be called grey differential equation models. The basic model is GM (1.1). The first 1 in the parentheses means that the equation is a one-rank equation and the second 1 means that there is only one variable in the equation.

Suppose

$$X^{(0)} = \{ x^{(0)}(1), x^{(0)}(2),.....x^{(0)}(T)\}$$

be the primary data sequence, in order to establish GM (1, 1) model, the steps are given as follows:

- Accumulating generation:

$$x^{(1)}(t) = \sum_{k=1}^{t} x^{(0)}(k) \quad t = 1, 2, ..., T \tag{1}$$

where $x^{(1)}(t)$ is the t-th datum in the accumulating generation number, $x^{(0)}(k)$ is the k-th datum in sequence of $X^{(0)}$.

Thus the differential equation of GM (1,1) model can be written as follows:+

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{2}$$

The coefficient vector is $\hat{a} = (a\ b)^T$. -a reflects the developmental situation of $\hat{x}^{(1)}$ and $\hat{x}^{(0)}$. b is obtained from the primary data sequence, which reflects the relationship of the data' changing (Deng, 2002).

- Construct data matrix:

$$B = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \cdots & \vdots \\ -\frac{1}{2}[x^{(1)}(T-1) + x^{(1)}(T)] & 1 \end{bmatrix} \tag{3}$$

$$Y_n = [x^{(0)}(2)\ x^{(0)}(3)...\ X^{(0)}(T)] \tag{4}$$

- Solve out the coefficient by using least square method:

$$\hat{a} = (B^T B)^{-1} B^T Y_n \tag{5}$$

Then we can establish the model as follows:

$$\hat{x}^{(1)}(t+1) = (\hat{x}^{(1)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a} \tag{6}$$

- Revert number sequence:

$$\begin{aligned} \hat{x}^{(0)}(t+1) &= \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \\ &= (1 - e^{a})\left(x^{(0)}(1) - \frac{b}{a}\right)e^{-at} \end{aligned} \tag{7}$$

GM(1,1) model's verify: Suppose the primary sequence is $x^{(0)}$and the fitting sequence is $\hat{X}^{(0)}$, then the residual error sequence is

$$\varepsilon^{(0)}(t) = x^{(0)}(t) - \hat{x}^{(0)}(t) \quad (t = 1, 2, ..., T) \tag{8}$$

Let S_1 be the mean square deviation of primary sequence, be mean square deviation of residual error sequence, namely:

$$S_1^2 = \frac{1}{n-1}\sum_{t=1}^{T}[x^{(0)}(t) - \bar{x}^{(0)}]^2 \tag{9}$$

$$S_2^2 = \frac{1}{n-1}\sum_{t=1}^{T}[\varepsilon^{(0)}(t) - \bar{\varepsilon}^{(0)}]^2 \tag{10}$$

Where,

$$\bar{x}^{(0)} = \frac{1}{n}\sum_{t=1}^{T} x^{(0)}(t),\ \bar{\varepsilon}^{(0)} = \frac{1}{n}\sum_{t=1}^{T}\varepsilon^{(0)}(t) \tag{11}$$

The posterior variance ratio is $C = \frac{S_2}{S_1}$. The probability of little error is

$$P = \{\left|\varepsilon^{(0)}(t) - \bar{\varepsilon}^{(0)}\right| < 0.6745 S_1\} \tag{12}$$

If p>0.95 and C<0.35, the model is tried, then we can use this model to forecast, else we should do residual error revision.

Residual error revision model: While $p \leq 0.95$ or $C \geq 0.35$, we choose the appropriate caudal segment from the residual error sequence to establish the GM (1,1) model, then we add the forecasting value of the residual error to the primary forecasting value in order to compensate the primary forecasting value and enhance the precision (Deng, 2002).

Suppose the residual error sequence be

$$\varepsilon_1^{(0)}(t) = (\varepsilon^{(0)}(t_0), \varepsilon^{(0)}(t_0+1), \ldots, \varepsilon^{(0)}(T)),$$

then we choose the appropriate caudal segment of the residual error to establish the GM(1,1) model,

$$\hat{\varepsilon}_1^{(0)}(t+1) = (-a_\varepsilon)\left(\varepsilon^{(0)}(t_0) - \frac{b_\varepsilon}{a_\varepsilon}\right) \exp[-a_\varepsilon(t-t_0)]\, t \geq t_0 \qquad (13)$$

If there are negatives in the chosen residual error caudal segment sequence, we must turn the negatives into the positive numbers first. The detailed steps can be listed as follows:

First we find the least negative $\varepsilon_{1,\min}^{(0)}$ in $\varepsilon_1^{(0)}$, then let

$$\varepsilon_2^{(0)}(i) = \varepsilon_1^{(0)}(i) + 2\left|\varepsilon_{1,\min}^{(0)}\right| \qquad (14)$$

In this way, we get a new sequence $\varepsilon_2^{(0)}$

- We establish a GM (1, 1) model by using the sequence $\varepsilon_2^{(0)}$ then we get $\hat{\varepsilon}_2^{(0)}(t+1)$
- We subtract

$$2\left|\varepsilon_{1,\min}^{(0)}\right|$$

from

$$\hat{\varepsilon}_2^{(0)}(t+1),$$

then we get

$$\hat{\varepsilon}_1^{(0)}(t+1)$$

Thus we get a new model:

$$\hat{x}_*^{(0)}(t+1) = \begin{cases} \hat{x}^{(0)}(t+1) & t < t_0 \\ \hat{x}^{(0)}(t+1) + \hat{\varepsilon}_1^{(0)}(t+1) & t \geq t_0 \end{cases} \qquad (15)$$

Compositional data: The pie chart is one kind of very commonly used statistical graph (Ferrers, 1866) which can express that each composition accounts for the proportion in some thing (Aitchison, 1986). In pie chart, the combination of each share is called compositional data (Zhang, 2000). For a list of compositional data collected according to the time order

$$P(t) = \{(p_1(t), \ldots, p_n(t)) \in R^n \mid \sum_{j=1}^{t} p_j(t) = 1, 0 \leq p_j(t) < 1\}\ t = 1, 2, \ldots, T$$

- Makes a simple nonlinear transformation to the primary compositional data (Wang *et al*., 2007)

$$q_j(t) = \sqrt{p_j(t)}, \quad j = 1, 2, \ldots, n; t = 1, 2, \ldots, T$$

Let, $O(t) = (q_1(t), \ldots, q_n(t)^T$, $t = 1, 2, \ldots, T$ apparently we have

$$\|Q(t)\|^2 = \sum_{j=1}^{n} (q_j(t))^2 = 1 \qquad (16)$$

- For any t $(t = 1, 2, \ldots, T)$,

$$Q(t) = (q_1(t), \ldots, q_n(t))^T \in R^n$$

distribute in a 1-radius and n-dimension hypersphere. Transform:

$$Q(t) = (q_1(t), \ldots, q_n(t))^T, \ t = 1, 2, \ldots, T$$

from the rectangular coordinate system to the spherical coordinate system $(r(t), \theta_2(t), \ldots, \theta_n(t) \in \Theta^n$. Because of $(r(t))^2 = \|Q(t)\|^2 = 1$, we have the mapping of $R^n \rightarrow \Theta^{n-1}$ as follows:

$$q_1(t) = \sin\theta_2(t) \sin\theta_3(t) \sin\theta_4(t) \ldots \sin\theta_n(t)$$

$$q_2(t) = \cos\theta_2(t) \sin\theta_3(t) \sin\theta_4(t) \ldots \sin\theta_n(t) \qquad (17)$$

$$q_3(t) = \cos\theta_3(t) \sin\theta_4(t) \ldots \sin\theta_n(t)$$

$$\vdots$$

$$q_{n-2}(t) = \cos\theta_{n-2}(t) \sin\theta_{n-1}(t) \sin\theta_n(t)$$

$$q_{n-1}(t) = \cos\theta_{n-1}(t) \sin\theta_n(t)$$

$$q_{n-1}(t) = \cos\theta_{n-1}(t) \sin\theta_n(t)$$

$$q_n(t) = \cos\theta_n(t) \quad 0 < \theta_j(t) \leq \frac{p}{2}, j = 2, 3, \ldots, n$$

- In 1st~2nd step transformation, the compositional data is fallen from the n-dimension space to (n-1)-dimension space. Therefore, the primary n linear correlated variables are transformed to n-1 independent angles: θ_j (t), j = 2,3,....n. By using the recursive algorithm, we can get:

$$\theta_n(t) = \arccos q_n(t)$$

$$\theta_{n-1}(t) = \arccos\left(\frac{q_{n-1}(t)}{\sin\theta_n(t)}\right)$$

$$\theta_{n-2}(t) = \arccos\left(\frac{q_{n-2}(t)}{\sin\theta_n(t)\,\sin\theta_{n-1}(t)}\right)$$

$$\vdots$$

$$\theta_2(t) = \arccos\left(\frac{q_2(t)}{\sin\theta_n(t)\,\sin\theta_{n-1}(t)...\sin\theta_3(t)}\right)\; t = 1,2,...,T \quad (18)$$

- Calculate the gained angel's data

$$\{\theta_1(t),\, t = 1,2,3,...,T\},\; j = 2,3,...,\, n-1,$$

and established (n-1) GM(1,1) model:

$$\hat{\theta}_j^{(0)}(t) = \hat{\theta}_j^{(1)}(t+1) - \hat{\theta}_j^{(1)}(t),\; j = 2,3,...,n,\; t = 1,...,T,T+1,...$$

If $p \le 0.95$ or $C \ge 0.35$, we ought do residual error revision.

- The fitting and forecasting value at the time of t+1 is

$$\hat{Q}(t+1) = (\hat{q}_1(t+1),...,\hat{q}_n(t+1))^T$$

apparently

$$\sum_{j=1}^{n}(\hat{q}_j(t+1))^2 = 1,\, t = 1,...,T,T+1,... \quad (20)$$

- The fitting and forecasting compositional data's value at the time of t+1 is

$$\hat{p}_j(t+1) = (\hat{q}_j(t+1))^2,\, j = 1,2,...,n,\, t = 1,...,T,T+1,... \quad (21)$$

APPLICATION

We adopt Chinese 3 industrial structures from 1999-2004 as the original data. The original data are listed in the following Table 1.

The angels' fitting and forecasting: We use the compositional data forecasting method introduced in 2.2 and take the data in Table 1 as the primary data. Here we take:

$P(1) = (0.162,0.458,0.380)^T$
$P(2) = (0.148,0.459,0.393)^T$
$P(3) = (0.141,0.452,0.407)^T$
$P(4) = (0.135,0.448,0.417)^T$
$P(5) = (0.125,0.460,0.415)^T$
$P(6) = (0.131,0.462,0.407)^T$

then we get the two angels' values (listed in Table 2) through the non-linear mapping.

We establish a GM (1, 1) model on $\theta_2$and we have (t+1) = 0.525023 $e^{-0.01669t}$. Its fitting results are listed in Table 3.

Then we test the model: The posterior variance ratio C = 0.284843 <0.35 and the probability of little error p = 1>0.95. So we can use this model to forecast θ_2'S value from 2005-2007(listed in Table 4).

Likewise, we establish a GM (1, 1) model on θ_3, we have (t+1) = 0.88928 $e^{-0.004207t}$. Its fitting results are listed in Table 5.

As we can see from Table5, the relative residual error is small. However, because of C = 0.474041>0.35 and

Table 1: Chinese three industrial structures from 1999-2004 (http://www.stats.gov.cn/tjdt/zygg/t20060109_ 402300176.htm.)

Year	1999	2000	2001	2002	2003	2004
The first industry	0.162	0.148	0.141	0.135	0.125	0.131
The second industry	0.458	0.459	0.452	0.448	0.460	0.460
The third industry	0.380	0.393	0.407	0.417	0.415	0.407

Table 2: The angels' values from 1999-2004

Year	1999	2000	2001	2002	2003	2004
θ_2	0.536540	0.516435	0.509363	0.502032	0.480531	0.489304
θ_3	0.906581	0.893232	0.878943	0.868784	0.870813	0.878943

Table 3: The fitting results of θ_2 from 1999-2004

Year	Original value	Fitting value	Residual error	Relative residual error (%)
2000	0.516435	0.516333	0.000102	0.02
2001	0.509363	0.507787	0.001576	0.31
2002	0.502032	0.499382	0.002650	0.53
2003	0.480531	0.491116	-0.010585	-2.20
2004	0.489304	0.482987	0.006317	1.29

Table 4: The forecasting value of θ_2 from 2005-2007

Year	2005	2006	2007
θ_2	0.474993	0.467131	0.459399

Table 5: The fitting results of θ_3 from 1999-2004

Year	Original value	Fitting value	Residual error	Relative residual error (%)
2000	0.893232	0.885546	0.007686	0.86
2001	0.878943	0.881828	-0.002885	-0.33
2002	0.868784	0.878126	-0.009342	-1.08
2003	0.870813	0.874439	-0.003626	-0.42
2004	0.878943	0.870768	0.008175	0.93

Table 6: The revised fitting results of θ_3 from 1999-2004

Year	Original value	Fitting value	Residual error	Relative residual error (%)
2000	0.893232	0.885546	0.007686	0.86
2001	0.878943	0.882328	-0.003385	-0.39
2002	0.868784	0.881926	-0.013142	-1.51
2003	0.870813	0.882639	-0.011826	-1.36
2004	0.878943	0.884768	-0.005825	-0.66

Table 7: The forecasting values of θ_3 from 2005-2007

Year	2005	2006	2007
θ_2	0.879642	0.886411	0.896696

p = 0.8<0.95, we must do residual error revision. According to the method introduced in 2.1.3, we do the residual error revision. Therefore, we get a new model:

$$\hat{\theta}_3(t+1) = 0.88928\, e^{-0.004207t} + \delta(t)(0.00735863\, e^{0.288479(t-1)} - 0.0186),$$

Where

$$\delta(t) = \begin{cases} 0 & t < 2 \\ 1 & t \geq 2 \end{cases}$$

By using this revised model, we get the new fitting results:

From Table 6, we can see that: The revised relative residual error is much more than the un-revised, but C = 0.321914 < 0.35 and p = 1>0.95. So we can use this model to forecast θ_3's value from 2005-2007 (Table 7).

The proportions' forecasting: Using the method introduced in 2.2, we can get the models as follows:

$$\hat{p}_1(t+1) = [\hat{q}_1(t+1)]^2 = [\sin\hat{\theta}_2(t+1)\sin\hat{\theta}_3(t+1)]^2$$
$$= \sin^2(0.525023e^{-0.01669t}) \times \sin^2[0.88928e^{-0.004207t} + \delta(t)(0.00735863e^{0.288479(t-1)} - 0.0186)]$$

$$\hat{p}_2(t+1) = [\hat{q}_2(t+1)]^2 = [\cos\hat{\theta}_2(t+1)\sin\hat{\theta}_3(t+1)]^2$$
$$= \cos^2(0.525023e^{-0.01669t}) \times \sin^2[0.88928e^{-0.004207t} + \delta(t)(0.00735863e^{0.288479(t-1)} - 0.0186)]$$

$$\hat{p}_3(t+1) = [\hat{q}_3(t+1)]^2 = [\cos\hat{\theta}_3(t+1)]^2$$
$$= \cos^2[0.88928e^{-0.004207t} + \delta(t)(0.00735863e^{0.288479(t-1)} - 0.0186)]$$

Table 8: The forecasting values of Chinese three industries from 2005- 2007

Year	2005	2006	2007
The first industry	0.124	0.122	0.120
The second industry	0.470	0.479	0.490
The third industry	0.406	0.400	0.390

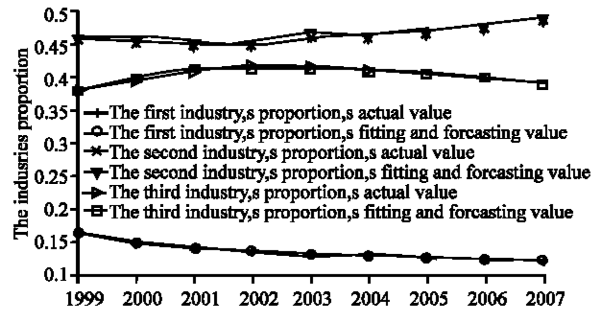

Fig.1: The proportions' fitting and forecasting values of the Chinese three industries

where ,

$$\delta(t) = \begin{cases} 0 & t < 2 \\ 1 & t \geq 2 \end{cases}$$

Let t = 1,...., 8,we can get the proportions' fitting and forecasting values of the Chinese three industries, which can been seen in Table 8 and Fig. 1.

While the 3 industries' actual proportions in 2005 is 0.125, 0.473 and 0.403 (http://www.stats.gov.cn/ tjsj/jdsj/t), so the relative error between the actual values and the forecasting values is 0.66, 0.74 and 0.82%. Meanwhile we can also see from Fig. 1, the fitting and forecasting values tally with the actual values quite well.

From the results of the forecast model, we can conclude that the further readjustment of the three industrial structures in China is going to take in the future three years. With the promotion of production efficiency in primary industry and secondary industry, the demands of further growth of tertiary industry are increased and the growth proportion of tertiary industry indicates that the structures of economic growth are more reasonable and healthy than the original statistical state. According to the experience of industrialized countries, the proportion of three industrial with GDP are basically the same. For the proportion of GDP distribution, most of the developed countries maintain the proportion of primary industry less than 3%, generally not exceed 5%, meanwhile the proportion of secondary industry less than 30% and the proportion of tertiary industry are always greater than 65% (Markusen and Venables, 1999). Comparing the proportion of the three industrial structures in China with above data, it is shown that the percentage of primary industry and secondary industry are still larger than the percentage of tertiary industry (Zhang, 2000). In orde to achieve the sustained economic growth, we need to adjust the present economic structure through making great efforts to develop tertiary industry (Tao and Han, 2004; Yue, 2003).

CONCLUSION

Considering that compositional data is a special kind of statistical data, we present a new forecasting and model method. We use the GM (1, 1) model in the gray system theory to fit and forecast the compositional data, which have been reduced dimensions. Through the research of the actual case, it is shown that the model given by this study can reflect the compositional data's mechanism and have rational and effective functions in analyzing trend and forecasting.

ACKNOWLEDGMENT

The authors would like to thank Yan Li, a research student who always participated discussion and intercommunion with us and all the colleagues who had given us help in college of science, Harbin Engineering University.

REFERENCES

Aitchison, J., 1986. The statistical analysis of compositional data (book style), London: Chapman and Hall.

David, J. Teece, 1996. Firm organization, industrial structure and technological innovation. J. Econ. Behav. Organ., 31: 193-224.

Deng, J.L., 2002. Grey Forecasting and Grey Decision Making (book style, in Chinese), 1st Edn., Huazhong University of Science and Technology Press, Wuhan, China.

Deng, J.L., 2002. The Foundation of Grey Theory (book style, in Chinese). 2nd Edn., Huazhong University of Science and Technology Press, Wuhan, China.

Deng, J.L., 2005. The Primary Methods of Grey System Theory (book style, in Chinese). 2nd Edn., Huazhong University of Science Technology Press, Wuhan, China.

Kim, K. and S.W. Moon, 2001. Foreign reserve crisis and the Korean industrial structure-A CGE approach, Mathe. Comp. Modelling, 33: 577-596.

Liu, S.F., Y.G. Dang and Z.G. Fang, 2004. Grey System Theory and its Applications (book style, in Chinese). 3rd Edn., Sci. Press, Beijing, China.

Markusen, J. and Venables, 1999. A foreign direct investment as a catalyst for industrial development. Eur. Econ. Rev., 43: 335-356.

Paul Dewick, Ken Green, Toby Fleetwood and Marcela Miozzo, 2006. Modelling creative destruction: Technological diffusion and industrial structure change to 2050. Technological Forecasting and Soc. Change, 9: 1084-1106.

Shi, J.Y. and Y.Y. Chai, 2007. The Grey compositional data model and its application in the analysis and forecast of China's industrial structure (in Chinese). Stat. Infom. Forum, 22: 32-35.

Tang, X. J., 2005. The forecast method and case analysis of the 3 industrial structures (in Chinese). Commercial Research, 330: 146-148.

Tao, L.H. and J. Han, 2004. Study on the Synergetic Effect of Chinese Industrial Structure and Technical Structure (In Chinese) 26: 5.

Wang, H.W., W. Huang and Q. c Liu, 2003. Forecast analysis for Beijing's industry structure (in Chinese). Sys. Eng-theory and Practice, pp: 124-127.

Wang, H.W., Q. Liu, M.K. Henry, L. Mok, H. Fu and Wai Man Tse, 2007. A hyperspherical transformation forecasting model for compositionaldata. Eur. J. Oper, Res., 179: 459-468.

Yue, J., 2003. Institutional Innovation: The Way of Industrial Structure Preferencein China (In Chinese), 16: 5.

Zhang, J., 2006. Research on the adjustment effect of FDI on China's Industrial Structure (In Chinese). Asia-Pacific Econ. Rev., 16: 85-88.

Zhang, R.T., 2000. The theory of compositional data (book style, in Chinese), Science Press, Beijing, China.