PSE FOR REGIONAL ECONOMY SPATIAL ANALYSIS BASED ON SPATIAL STATISTICAL MODEL

Song Quanhong[a, b,]*, Dongchun[b]

[a] Chinese Academy of Surveying and Mapping, Beijing,100039, China, 2894053@163.com
[b] School of Geomatics, Liaoning Technical University, Fuxin, 123000,China

KEY WORDS: PSE, Spatial Statistics, Grey Clustering, Regional Economy, Spatial Analysis, Visualization, Model Verification.

ABSTRACT:

Analyzing the law of economic development and estimating the regional economy development's level to discover the relation between the economy development and area diversity, which can provide the reference gist for the decision-maker in the government economic department. Therefore, this paper set up regional economy spatial analysis PSE based on spatial statistical model, take the time serial model as complementarity, and through the friendly interface and the visualized analytical result to provide a fast and exact problem solving environment for users. Taking the spatial statistical model as the approach to solve the problem and the way to dig implied meaning of data, which increases the result's reliability, at the same time, fetch up the GIS software's disadvantag in spatial statistics, and offer a reference for GIS to boost up its spatial analysis abilities.

1. INTRODUCTION

According with the quick development of the data acquirement technology, the phenomenon of "abundant data but poor information "is more and more obvious , Spatial data mining technology enables extract the implicit information from the mass of spatial data. The data mining technology integrates many mature tools and technology which from various subject domain, including databases, statistics, pattern recognition, artificial intelligence, neural networks and so on[1]. Spatial statistics has strong theoretical foundation and a large number of sophisticated algorithms, it can effectively deal with numerical data and give the reality mode of spatial phenomenon (Zeitouni, 2002), at the same time, spatial statistics can improve the GIS' capability which deal with the random processing, analyze the error propagation low of spatial model, and analyze the spatial process to forecast future[2]. Ultimately provide a theoretical basis and quantitative tool for analyzing the spatial relativity of the seriate area. Thus, spatial statistics is the basic data mining technology, especially the multiple spatial statistics analysis such as principal component analysis, correlation analysis, multiple regression analysis, etc. (Li Deren, 2005) [3]

In recent years, China has put forward a series of regional development policies, from the eastern coastal opening to the western exploitation, from the northeast flourish to the revitalization of mid area, all that illuminate the importance to develop the regional economy harmony. Therefore, in the condition of different resource and different social circumstance, analyzing whether different area's economy development level has much differentia and evaluating the differentia objectively to research its rule and reasons will have important significance for the guidance of regional economic development strategy[4]. Some mature GIS software has powerful spatial analysis function. Whereas, even though the most powerful GIS software – ArcGIS is weak in spatial statistical analysis, at the same time, the complex operation is not suitable for economic analyst. At present, there are some popular statistical software such as SPSS, SAS and State in china. However, SAS is too professional to master, State is weak in multivariate analysis, and SPSS can't update the statistical method. Based on the above, this paper put forwards a PSE for regional economy which bases the spatial statistic model, it is the supplement for GIS' spatial analysis function, besides, it wraps the algorithms and the details of operation that will not only facilitate the non-professionals but also free the researchers from complex programme to concentrate on their own domain's research[5].

2. REGIONAL ECONOMY SPATIAL ANALYSIS PSE CONSTRUCTION

2.1 Architecture of regional economy spatial analysis PSE

PSE (Problem solving environment) is a computing environment, for the special problems, PSE will build a database management system, model system and visual display platform. PSE is characterized by problem-oriented and computing ability [6].This paper construct the regional economy PSE serve the decision-maker in national economy department to make decision by the statistical analytical result. This paper uses the gray clustering theory to analyze the economy development level of the 31 counties in Lancang River basin, and take it as the example to bring forth the architecture of Regional economic spatial analysis PSE.

* Corresponding author. 2894053@163.com

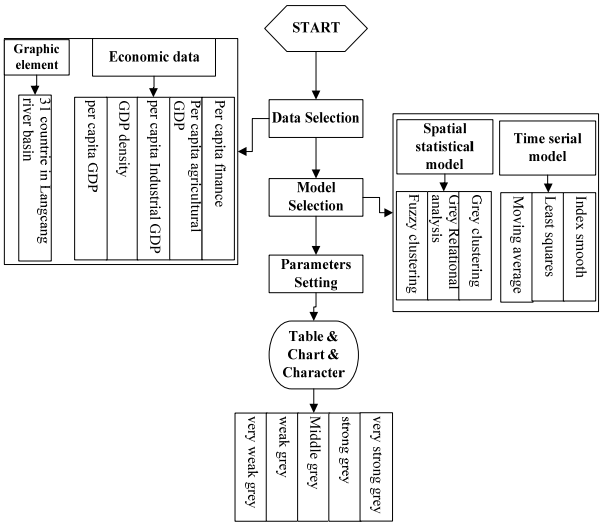

Figure 1. PSE Architecture

	MODEL	MODEL PRINCIPLE	EXAMPLES
Spatial statistical model	Gray Clustering	Based on gray correlation matrix and whiting function of ashen number , making some observation indicators or observation object assembled to generate some well-defined class	ascertain regional industrial pillar industries; the evaluation of the quality of the environment;
	Gray Relational analysis	Take the elements' degree of resemble and dissimilitude as "the grey relational degree", and use the relational degree to weight the element's relational degree.	Gray correlation analysis of science and technology investment and economic growth
	Fuzzy Clustering	Cluster analysis is a multivariate analysis method, it is a mathematical method to quantificationally ascertain the samples' affiliation, fuzzy clustering usually clusters via fuzzy similarity.	IT market forecast analysis; the weighted fuzzy cluster of the developing situation of the characteristic county .
	Spatial Autocorrelation Analysis	The similarity of the same attribution between some region and its neighborhood.	Research on Fujian regional economic diversity based on Spatial autocorrelation analysis .
`	Seemingly Unrelated Regression Model	Research on the relation among multi-time index and multi-spatial index indicators, and it only fit the instance when the time indexes are less than spatial index.	
	Moving Average Forecast	Research on the relation among multi-time index and multi-spatial index indicators, and it only fit the instance when the time indexes are more than spatial index.	

Table 1. Statistical model

2.2.2 Model Verification: Model verification in this paper may say as statistical verification, it aims at verifying the models' statistical character. The statistical verification use the mathematical statistics methods to verify the equation, the evaluation of model's parameters and the reliability. This methods verification includes fitting optimization verification, equal significance verification and variable verification. For clearly, this paper gives the model flow chart as Figure 2:

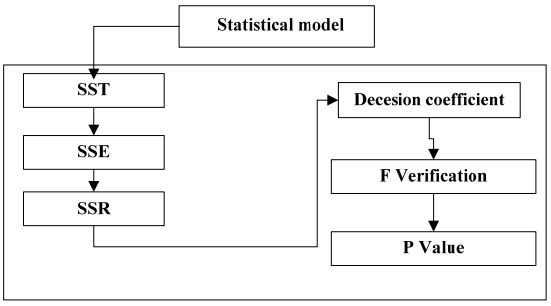

Figure2. Model verification

2.2 Key technology of regional economy spatial analysis PSE

2.2.1 The choice of statistical model : In many statistical algorithm, how to select the most fit algorithm to analyze the regional economy development is an important point, that may

determine the results' reliability. In this paper we choose some mature models which have been verified by practice. In the talbe1, briefly show the selected models' work principium, and give an example for models to illuminate the models' application area.

SST(Square Sum of Regression), denotes the summation of indicator's changes; SSE(Square Sum of Error), denote the residual error's magnitude that is not be interpreted by independent variables. In general, above the residual error summation ,the smaller the better; SSR(Square Sum of Totality) display the ability of independent variables to explain the dependent variable, usually the bigger the better; R^2 means determined coefficient whose value equals SSR / SST, is located between 0 and 1 , and the greater the regression model the stronger whose explanatory power to explain practice; F - test, that is F value test, reflects the model's explanatory power, also the bigger the better ; P value calculated in accordance with F value is a probability value between 0-1 , the smaller P value the more notable of the models. For example, the P value is less than 0.01,said that model is significant under the probability of 0.01.

2.2.3 Result Visualization: The advent of the analytical results is not only limited to the accuracy of the results, the open-and-shut is more and more attached importance to by people, a good expression of the results can easily dig the implied information from original data , and then infer the inherent law of the business' development. Based on above, this paper combined with the means of charts, tables and characters to display the analysis results, the expression of table can show the every index's detailed analytical result, the expression of chart can easily illuminate the general situation , at last, the characteristic may clearly display some interpretative information to users , it is the supplement for the analytic result.

3. REALIZATION OF REGIONAL ECONOMY SPATIAL ANALYSIS PSE

The PSE is developed in the VS.NET development environment and programmed by the C # programming language ,the PSE is under the B / S structure, whose data stores in Microsoft SQL Server database, based on the statistical model to do some analysis, at the same time , proof-test the model by mathematical means. Next, take the analysis of the counties' economy development in Lancang river basin in Yunnan province as the example to show the PSE's working.

3.1 The gray clustering analysis of regional economy in Lancang River Basin in Yunnan province

Analytical data is the economic data from the counties and cities in Yunnan Lancang River Basin ,such as the per capita GDP, GDP density, per capita agricultural output value, per capita industrial output value and per capita revenue, gray clustering of cities and counties aimed at analysis the level of economic development. The Analysis operation as follows:

1. Select Data table, and then select the analysis indicators, as Figure 3 :

Regional Economy Analysis ① data selection ② model analysis ③ model validation ④ resual expression

数据库名: TimeSpaceData

表名: 40 countries' economy development index in Lancang River o

	1	2	3	4	5	6	7
	NO.	county name	PGDP	GDP	per capita ag production value	per catita industrial value	per capita finance
1	1	Cuiyun District	8041	46.84	1318.42	4642.92	558
2	2	Pu'er Hani&Yi Autonomous Country	4843	23.15	1285.38	2071.77	203
3	3	Jingdong Yi Autonomous Country	3401	26.82	1106.75	643.42	225
4	4	Jinggu Dai Autonomous Country	4462	16.79	1430.03	3319.02	249
5	5	Zhenyuan Yiand Hani and Lagu Autonomous Country	2559	12.41	1094.15	961.27	122
6	6	Jiangcheng Hani Autonomous Country	4318	13.75	1915.21	986.91	203
7	7	Menglian Dai and Lagu and Wa Autonomous Country	3778	22.3	1859.15	1718.46	196
8	8	Lancna Lagu Autonomous Country	2334	12.53	897.7	899.09	81
9	9	Ximeng Wa Autonomous Country	2569	15.35	707.14	666.9	66
10	10	Jinghong city	10326	55.68	2824.15	1495.21	363
11	11	Menghai Country	4343	23.36	1566.4	1968.62	133
12	12	Mengla Country	8182	23.24	3839.31	1062.61	173

Figure 3. Data selection

2. Enter into the model analysis process, combined with the means of tables, charts and character to express the analytical result, such as figure 4.

GREY CLUSTERING ANALYSIS RESULT

No	County Name	Very Strong Grey Class	Strong Grey Class	Middle Grey Class	Weak Grey Class	Very Weak Grey Class	Grey Class
1	Cuiyun District	0.42	0.419	0.354	0.306	0.038	Very Strong
2	Pu'er Hani&Yi Autonomous County	0.236	0.869	0.819	0.598	0.061	Strong
3	Jingdong Yi Autonomous County	0.184	0.704	0.816	0.82	0.27	Weak
4	Jinggu Dai Autonomous County	0.271	0.847	0.688	0.536	0	Strong
5	Zhenyuan Yiand Hani and Lagu Autonomous County	0.15	0.57	0.743	0.881	0.426	Weak
6	Jiangcheng Hani Autonomous County	0.24	0.705	0.697	0.514	0	Strong
7	Menglian Dai and Lagu and Wa Autonomous County	0.246	0.755	0.765	0.493	0.023	Middle
8	Lancna Lagu Autonomous County	0.125	0.477	0.622	0.752	0.706	Weak
9	Ximeng Wa Autonomous County	0.111	0.422	0.55	0.681	0.84	Very Weak
10	Jinghong city	0.439	0.304	0.111	0.019	0	Very Strong
11	Menghai County	0.229	0.818	0.75	0.619	0	Strong
12	Mengla County	0.404	0.312	0.347	0.232	0	Very Strong

Figure 4-a. Analytical result expressed by table

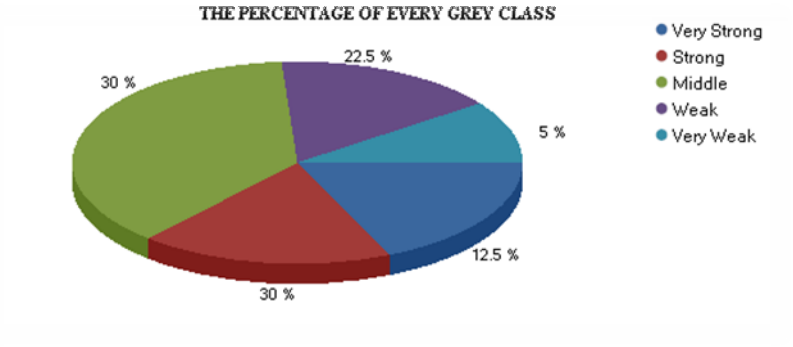

Figure 4-b. Analytical result expressed by chart

GREY CLUSTERING ANALYSIS RESULT

Very Strong Grey Class: (5)
Cuiyun District、Jinghong city、Mengla County、Dali、Gucheng District
Strong Grey Class : (12)
Pu'er Hani&Yi Autonomous County、Jinggu Dai Autonomous County、Jiangcheng Hani Autonomous County、Menghai County、Yangbi Yi Autonomous County、Xiangyun County、Binchuan County、Nanjian Yi Autonomous County、Er'yuan County、Lanping Bai and Pumi Autonomous County、Yun county、Gengma Dai and Dai Autonomous County
Middle Grey Class: (12)
Menglian Dai and Lagu and Wa Autonomous Country、Weishan Yi Autonomous County、Yongping County、Yunlong County、Jianchuan County、Heqing County、Changning County、Lushui County、Gongshan Dulong and Nu Autonomous County、Lincang County、Shuangjiang Lagu and Wa and Bulang and Dai Autonomous County、Cangyuan Wa Autonomous County
Weak Grey Class: (9)
Jingdong Yi Autonomous County、Zhenyuan Yiand Hani and Lagu Autonomous County、Lancna Lagu Autonomous County、Midu County、Yulong County、Deqin County、Weixi Lisu Autonomous County、Fengqing County、Yongde County
Very Weak Grey Class : (2)
Ximeng Wa Autonomous County、Fugong County

Figure 4-c. Analytical result expressed by character

3.2 Model verification

in the aspect of model analysis and forecast the time serial model is same as the spatial statistical model ‘s operation process, but the time serial model has accomplished model verification, based on the residual error summation computer can give user the best analysis model for a given case. Take the Guangdong province’s GDP forecast analysis as example to show the model verified result , the result as figure 5:

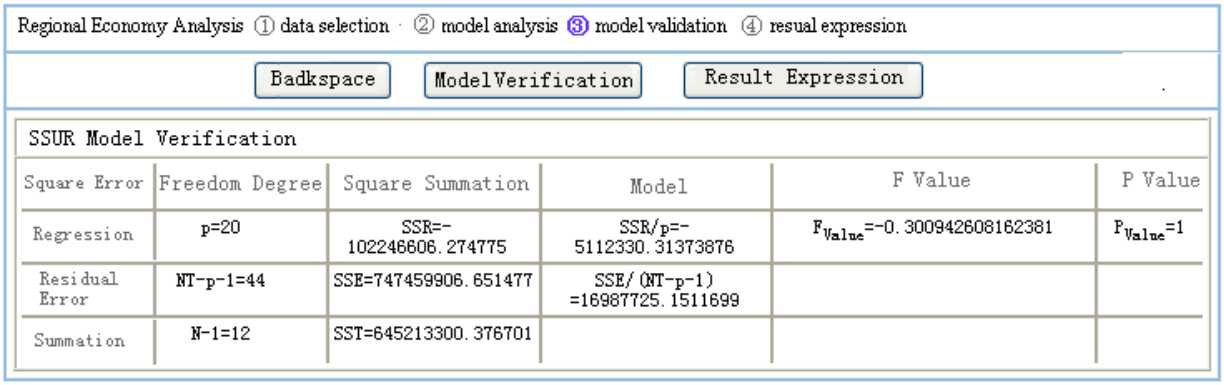

Figure 5. Model verified result

4. CONCLUSIONS

In this paper, though the regional economy spatial analysis PSE does not have the powerful capability like these software a certain extent, but it makes up these software's shortcoming and has its own advantage when faces the specific subject, these merit as follows:

1. Wrapped algorithm and data reduces the complex degree of operation largely;
2. The multiform expression can easily to show the low of the analysis result;
3. The model verification increases the reliability of the analysis result.

However, the PSE has its own insufficiency in a way, the PSE doesn't have model management system, that weaken its ability to manage models such as add models, delete models and modify models and so on. So, the next task is to construct a model management system to improve the models' communication in different research domain.

REFERENCE

[1] Wu Lianggang, Zhou Haitao, J. 2002. Research on data mining technique based on consider. *Guangxi University transaction(physical science edition)*, 27(1), pp. 67~70.

[2] Sun Weibin, J. 2004. The statistical methods' function and issue in data mining area. *Mathematical statistics and management*, 23(5), pp.78-81 .

[3] Li Deren, Wang Shuliang, Li Deyi, J. 2008. *Theory and application of spatial data mining*. Beijing China, Science publishing company, pp.265-267.

[4] Dong Chun, J. 2008. Research on Evaluating Urban Comprehensive Competitiveness during Multiyear and the Spatial Characters of Important Cities in China. *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*. pp.498-503.

[5] Feng Shengzhong, T. 2008, Some consideration on the web problem solving environment. Beijing, China, http://www.ict.ac.cn/diffusive/channel/detail382.asp.

[6] Wang Renchong, Xu Tao, and so on, J. 2008. Research on the problem solving environment based on the internet. *Computer Engineering*,32(15). pp.99-101.

.