

GIS-BASED SPATIAL STATISTICAL ANALYSIS OF COLLEGE GRADUATES EMPLOYMENT

Ruidai Tang

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University,
Wuhan 430079, P.R. China- ruidaihaohao@whu.edu.cn

Commission II, II/3

Keywords: GIS; Spatial Distribution; College Graduate; Employment; Spatial Analysis; The Stepwise Multiple Linear Regression

ABSTRACT:

It is urgently necessary to be aware of the distribution and employment status of college graduates for proper allocation of human resources and overall arrangement of strategic industry. This study provides empirical evidence regarding the use of geocoding and spatial analysis in distribution and employment status of college graduates based on the data from 2004-2008 Wuhan Municipal Human Resources and Social Security Bureau, China. Spatio-temporal distribution of employment unit were analyzed with geocoding using ArcGIS software, and the stepwise multiple linear regression method via SPSS software was used to predict the employment and to identify spatially associated enterprise and professionals demand in the future. The results show that the enterprises in Wuhan east lake high and new technology development zone increased dramatically from 2004 to 2008, and tended to distributed southeastward. Furthermore, the models built by statistical analysis suggest that the specialty of graduates major in has an important impact on the number of the employment and the number of graduates engaging in pillar industries. In conclusion, the combination of GIS and statistical analysis which helps to simulate the spatial distribution of the employment status is a potential tool for human resource development research.

1. INTRODUCTION

With the development of Chinese higher education from elite education to mass education, the employment problem of college graduates is always the focus of the whole society. It is difficult to clearly grasp the distribution and transfer of professionals through present human resource management, accordingly hardly to make correct decisions for professional demand, especially for determining whether professionals required for a region of pillar industry are appropriately allocated and for talent introduction in the future. Well master of spatial distribution and mobility of professionals plays an important role in human resource management, decision-making of government and career planning for college graduates. The mobility and dynamic distribution of college graduates is of great significance to reasonable allocation and import of professionals as well as career planning. Therefore, it is urgently necessary to be aware of the spatio-temporal distribution and employment status of college graduates for proper allocation of human resources and overall arrangement of strategic industry.

GIS (geographic information system) is powerful tool for the storage and management, as well as analysis of natural and cultural data, increasing studies have applied the spatial statistical method of GIS in social sciences, for example, some research explores the spatio-temporal distribution of population (Zhang and Qi 2007; Chen, Peng et al. 2010; Ousley 2010; Zhang, Zhu et al. 2010) and student persistence (Ousley 2010). However, this depends on the geocoding of GIS and statistical analysis. Geocoding helps to associate spatial information with non-spatial information and integrate the information in different spatial domain (Grayson 2000; Jiang and Li 2003). Thus, geocoding is widely used in the spatial distribution and location analysis of population or enterprises

(Jiang and Li 2003). Statistical analysis can establish prediction models for college graduates employment and identify the most contributive factors for the prediction. Zhang and Qi (2007) investigated several population density models for the population distribution and found the Inverse distribution functions was the most optimal population density.

As mentioned above, in this paper the geocoding of GIS and statistical analysis (stepwise multiple linear regression) are employed to explore the distribution of college graduates employment.

Geography information system (GIS) has recently been increasingly applied in the humanities. The spatialization of employment relates employment data to geographic space with the combination of spatial statistics and GIS technique. It reflects the characteristics of spatial distribution of talents, and plays an important role in the optimization of talent structure, the reasonable allocation of talents, the predictive requirement of talents and in the decision making of human resource.

This study aimed to (i) dissect the major technical problem of the spatialization of employment and (ii) to analyze the distributive characteristics and change of college graduates.

2. DATA AND METHODS

2.1 Data description and pre-processing

In this study, the data source is from 2004-2008 Wuhan Municipal Human Resources and Social Security Bureau, China. The number of college graduates employment (male, female, specialty), the number of and address of enterprises where graduates employed in 2004-2008, and the number of graduates engaging in photoelectron information industry, modern

equipment manufacturing industry, bio-pharmaceutical industry, energy-saving and environment protection industry, modern agriculture industry in Wuhan east lake high and new technology development zone were extracted.

2.2 The computing configuration

Hardware: the operating system: Microsoft Windows XP Professional (5.1, version 2600) Chinese simplified edition
Memory: 1912 MB RAM
The processor: Intel (R) Core (TM) 2 Duo CPU T6670 @ 2.20 GHz (2 CPUs)
Software: ArcMap 9.3 SPSS Statistics of 17

2.3 Vectorization of spatial data

The vectorization was implemented using ArcMap 9.3. The administrative boundaries were first extracted and sketched in a new layer against the raster map of Wuhan east lake high and new technology development zone through visual interpretation and analysis. The vectorization of the development zone boundary was carried out after conscientious comparison and necessary corrections and saved as SHP format. Similarly, the road, lake, green-land and industrial park were vectorized.

2.4 Geocoding

Geocoding associates spatial address data with spatial coordinates, so that the location of geographical entity (i.e. enterprises, schools, hospitals) can be identified, facilitating spatial analysis and statistical analysis of entities. The process of geocoding consists of address standardization and address matching (Grayson 2000). Address standardization makes disorganized street address into organized and common format. Address matching determines the spatial location of geographical entity which is then drawn in the vector map. The geocoding of enterprises where graduates employed in 2004-2008 was then performed by ArcMap 9.3.

2.5 Statistical analysis

Statistical analysis, i.e. stepwise multiple linear regression, helps to build a model between the dependent variable and independent variables. In this study, stepwise multiple linear regression (SMLR) was used to establish relationships between the number of employment, the number of graduates engaging in five pillar industry and the number of graduates majoring in different specialties. The selection of independent variables depends on the F test and adjusted determination of coefficients. In this study, SMLR was performed using SPSS software version 17.0. The independent variable was deleted when the F test value was more than 0.1.

Apart from the social survey, this study employed geocoding, spatial analysis of GIS and statistical analysis (stepwise multiple linear regression) to investigate systematically the spatial distribution and configuration of college graduates.

(i) This study utilized the social survey such as investigation and questionnaire methods to set the table of social characteristics of college graduates, e.g., sex, age, education, profession and work unit, and to present the employment and to

provide precondition for the investigation of the spatialization of employment.

(ii) This study employed geocoding and spatial analysis of GIS to relate the employment data to the geographic space, and to explore the spatial distribution of college graduates in the Optical Valley of Wuhan from 2004 to 2008.

(iii) Stepwise multiple linear regression was used to study the distributive characteristics of college graduates.

3. RESULTS AND ANALYSIS

3.1 Spatio-temporal variation of enterprises

Fig. 1-5 illustrates the spatial distribution of enterprises where graduates employed in 2004-2008. The enterprises were drawn on the map by geocoding. By and large, the enterprises were distributed in the north of the development zone due to convenient transportation and closed to the downtown of Wuhan City. In 2004, there were only 15 enterprises scattered in the industrial parks and around the main road (Fig. 1). The number of enterprises dramatically increased to 91 in 2005, 168 in 2006, and 173 in 2007. These enterprises were densely distributed in the industrial parks (Fig. 2, Fig. 3, Fig. 4). The enterprises in 2008 decreased to 147, but the enterprises were gradually expanded to the east of the development zone (Fig. 5). With the galloping development of economy and investment, more and more enterprises were located in the development zone. Furthermore, the number of graduates and employment in Wuhan has been steadily increased during the last decade. Accordingly, the enterprises where graduated employed tended to rapidly increased in 2004-2007. The decline of enterprises may result from the financial crisis which led to the decreasing demand of employment in 2008.

From 2004 to 2005, the centre of spatial distribution of enterprises moved northwestward about 1786 m (Fig. 6). From 2005 to 2006, the centre shifted south about 142 m, and moved southeastward about 144 m in 2007. From 2007 to 2008, the centre shifted southeastward about 568 m, indicating a centripetal tendency of developing southeastward in the development zone in the years that followed. This may result from the location advantage and industrial cluster in the south and east of the development zone.

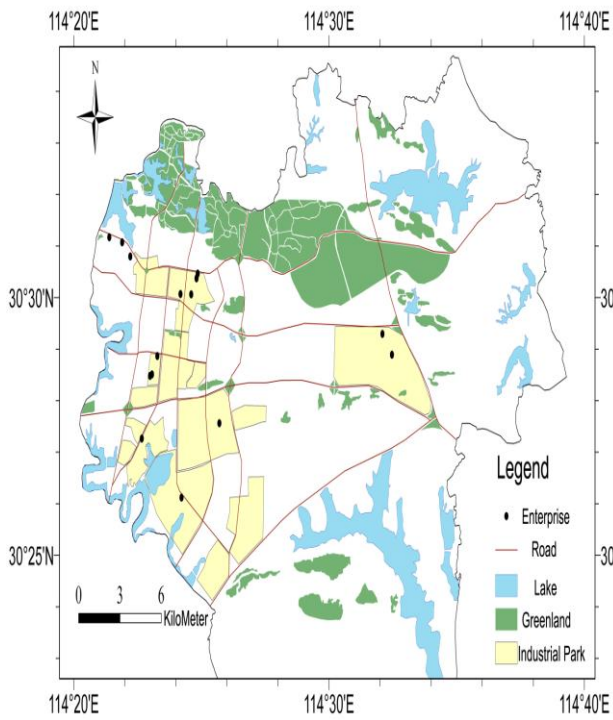


Fig. 1 Spatial distribution of enterprises in 2004

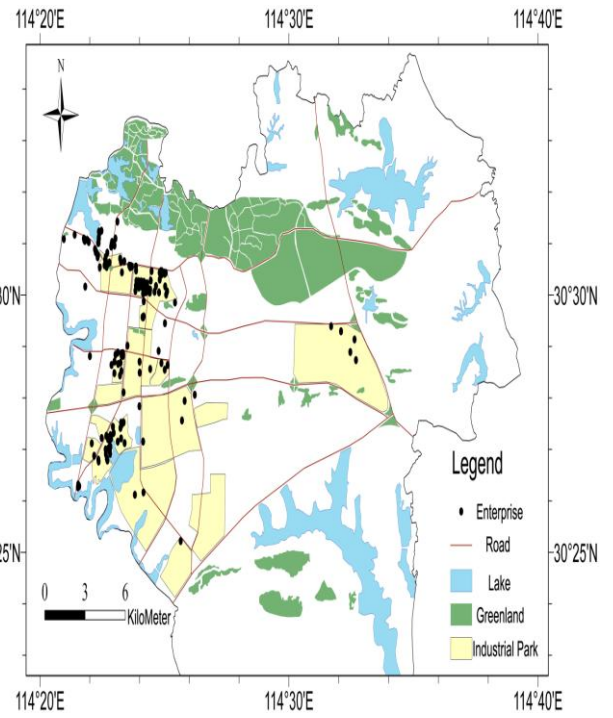


Fig. 3 Spatial distribution of enterprises in 2006

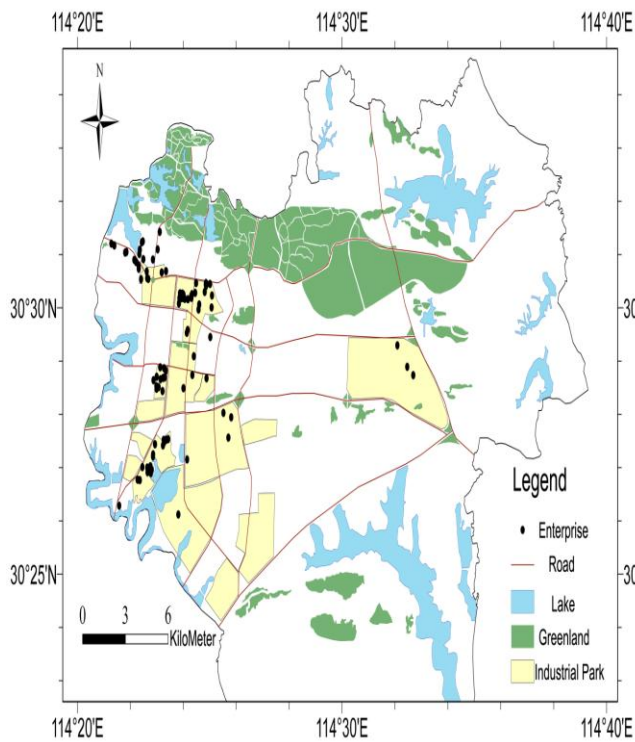


Fig. 2 Spatial distribution of enterprises in 2005

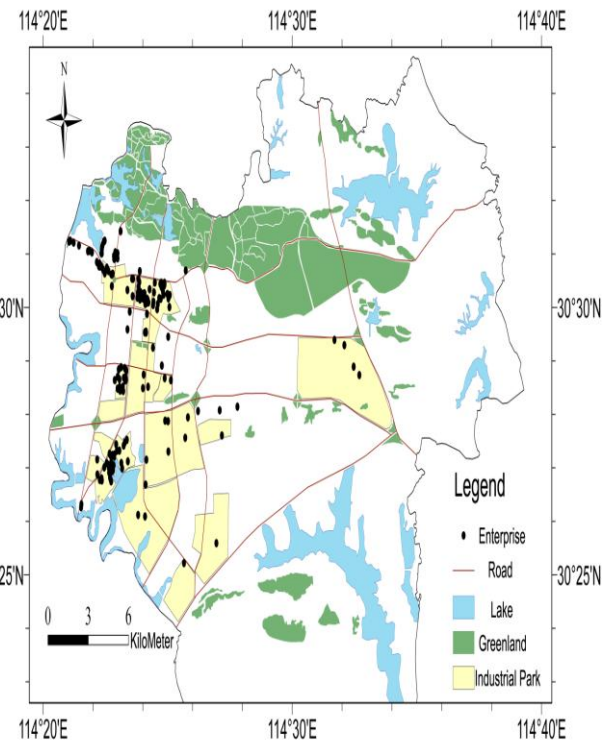


Fig. 4 Spatial distribution of enterprises in 2007

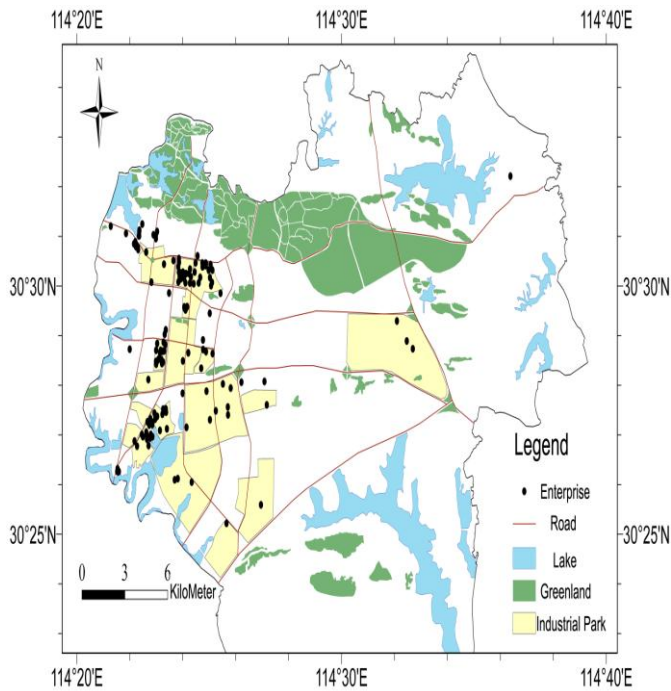


Fig. 5 Spatial distribution of enterprises in 2008

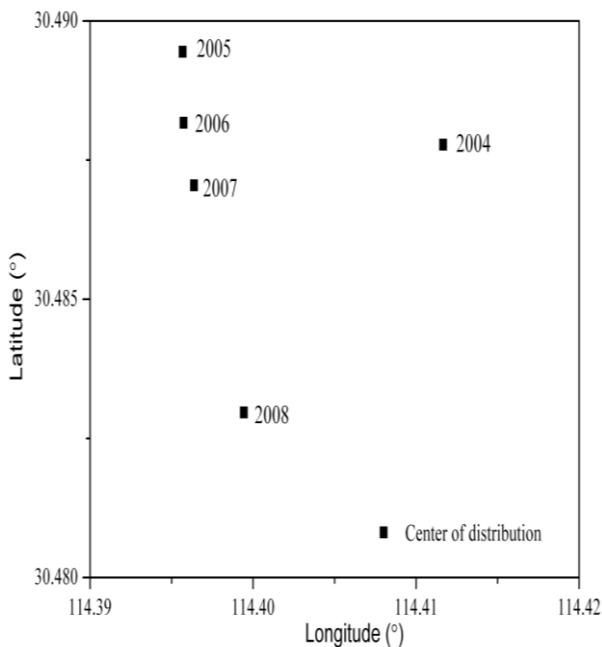


Fig. 6 The center of spatial distribution of enterprises in 2004-2008

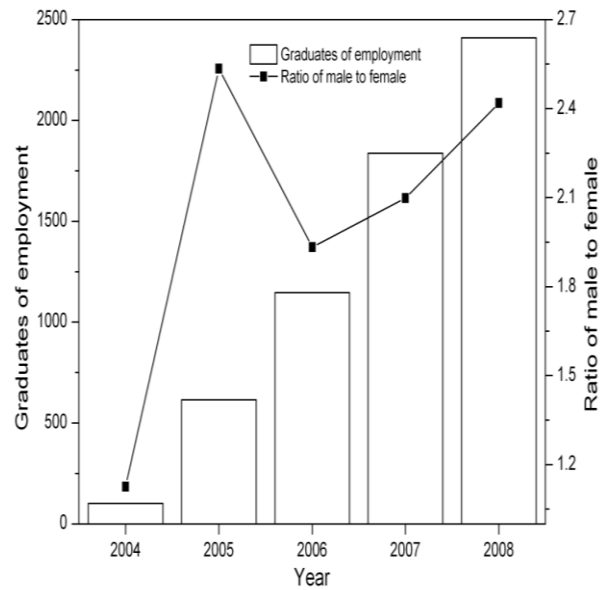


Fig. 7 The number of graduates of employment and ratio of male to female

3.2 Distribution of college graduates employment

The increasing enterprises and college graduates play an important part in graduate employment in development zone. Furthermore, the distribution of employment reflects the distribution of industry and the types of talents. Fig. 7 demonstrates the number of employment and the ratio of male to female in 2004-2008. The number of employment increased rapidly from 102 in 2004 to 2410 in 2008. In 2004, the ratio of male to female was 1.125. Then the ratio dramatically increased to 2.534 in 2005, double than that in 2004. In 2006, the ratio decreased to 1.933. In the years that followed, the ratio grew fast (2.098 and 2.418 in 2007 and 2008 respectively).

The variation of the ratio relates to the specialty of graduates and the industrial layout. Different specialties have different ratios of male to female, for example, more men major in engineering than women, however, the situation performs oppositely for Arts and History. In addition, the industrial layout influences the type of specialty, for example, the photoelectron information industry requires graduates majoring in Engineering.

SPECIALTY	YEAR				
	2004	2005	2006	2007	2008
PHILOSOPHY	0	0	1	0	1
ECONOMICS	5	11	37	78	77
LAW	4	5	16	32	42
EDUCATION	8	4	13	13	20
ARTS	11	28	59	109	137
HISTORY	0	1	1	0	1
SCIENCE	1	54	126	257	272
ENGINEERING	49	390	700	1083	1416
AGRICULTURE	4	50	46	22	90
MEDICINE	1	1	18	35	51
MANAGEMENT	19	71	130	208	303

Table 1 The number of graduates majoring in different specialties in 2004-2008

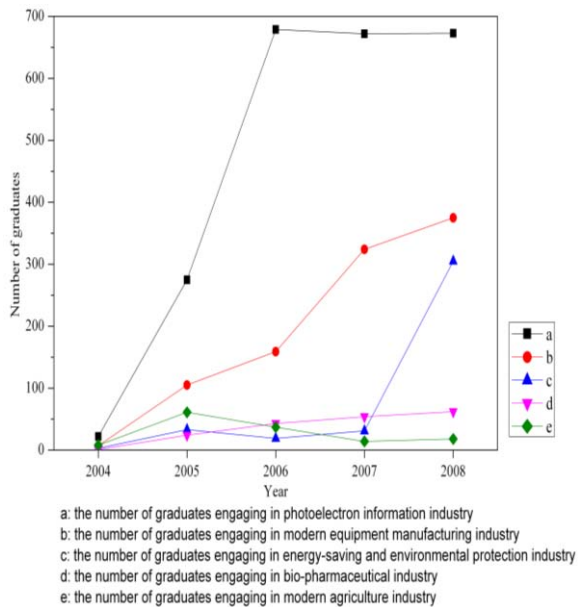


Fig. 8 The number of graduates engaging in five pillar industries

Table 1 lists the number of graduates majoring in different specialties in 2004-2008. It can be observed that engineering is the most significantly growing specialty, from 49 in 2004 to 1416 in 2008. The other specialties increased at different levels in 2004-2008. Fig. 8 shows the number of graduates engaging in five pillar industry (photoelectron information industry, modern equipment manufacturing industry, bio-pharmaceutical industry and modern agriculture industry). The number of graduates engaging in photoelectron information industry increased rapidly from 2004 to 2006, but kept stable in the years that followed. The number of graduates engaging in modern agriculture industry associates with the number of graduates majoring in Agriculture (Table 1), for example, from

2004 to 2005, they both increased, but decreased in 2006 and 2007, and both increased in 2008. This can suggest that specialty of graduates and industry in which graduates engage matter to each other to some extent. Furthermore, they have an impact on the number of college graduate employment.

In this study, stepwise multiple linear regression method was used to investigate what factors contribute to the college graduates employment. There were no appropriate models for predicting the number of graduates engaging in modern agriculture (Y6) and energy-saving and environment protection industry (Y7), because no significant correlation existed between Y6, Y7 and the number of specialties graduates majored in. Models were built to establish the relationship between the number of the employment (Y1), the number of graduates engaging in photoelectron information industry (Y2), the number of graduates engaging in bio-pharmaceutical industry (Y3), the number of graduates engaging in modern equipment manufacturing industry (Y4), the number of enterprises (Y5) and the specialties that graduates majored in from 2004 to 2008 (Table 2). In addition, the relationships between the number of the employment (Y1), the number of enterprises (Y5) and the number of graduates engaging in five pillar industry (Y2, Y3, Y4, Y6, Y7) were explored (Table 2).

MODEL	R ²	RMSE	F	P
$Y_1 = 1.738 + 1.471X_1 + 7.736X_2$	1.000	5.463	57253.38	0.000
$Y_2 = 107.894 + 0.490X_1$	0.774	165.708	10.265	0.049
$Y_3 = 4.671 + 0.044X_1$	0.949	6.398	55.988	0.005
$Y_4 = 14.709 + 1.263X_3$	0.986	20.846	212.800	0.001
$Y_5 = 18.278 + 0.217Y_2$	0.965	14.371	82.712	0.003
$Y_1 = 61.489 + 5.983Y_4$	0.982	142.782	164.661	0.001

Table 2 Prediction results using multiple linear regression method*

Y₁, Y₂, Y₃, Y₄, Y₅ represent the number of the employment, the number of graduates engaging in photoelectron information industry, the number of graduates engaging in bio-pharmaceutical industry, the number of graduates engaging in modern equipment manufacturing industry and the number of enterprises from 2004 to 2008 respectively; X₁, X₂, X₃ represent the number of graduates majoring in Engineering, Economics and Science from 2004 to 2008 respectively.

The number of graduates majoring in Engineering (X₁) and Economics (X₂) contributed to the prediction of the number of the employment (Table 2), the determination of coefficients (R²) was 1, indicating a good predictive accuracy. The number of graduates majoring in Engineering (X₁) affected the models of predicting the number of graduates engaging in photoelectron information industry (Y₂) and bio-pharmaceutical industry (Y₃), providing a R² of 0.774 and 0.949 respectively. The number of graduates majoring in Science (X₃) significantly associated with the number of graduates engaging in modern equipment manufacturing industry (Y₄) with a R² of 0.986. In addition, the number of graduates engaging in photoelectron information industry (Y₂) contributed to the prediction of the number of enterprises (Y₅) with a R² of 0.965. The number of graduates engaging in modern equipment manufacturing industry (Y₄)

also associated with the number of employment (Y_1) with a R^2 of 0.982.

The above models suggest the specialty of graduates major in has an important impact on the number of the employment and the number of graduates engaging in pillar industries, facilitating to regulate the industrial layout. Based on the number of specialty of graduates major in, we can predict the number of employment conveniently and accurately and get hold of the spatial distribution of college graduates employment and enterprises in development zone.

4. CONCLUSION

The enterprises where graduates employed in increased dramatically from 2004 to 2008, and tended to distribute southeastward.

The number of graduates majoring in Engineering and Economics contributed to the prediction of the number of employment, and the number of graduates majoring in Engineering relates closely with the number of graduates engaging in photoelectron information industry and biopharmaceutical industry, the number of graduates majoring in Science contributed to the prediction of the number of engaging in modern equipment manufacturing industry. This suggest that the specialty of graduates major in has an important impact on the number of the employment and the number of graduates engaging in pillar industries, facilitating to regulate the industrial layout.

To sum up, this is the first time to get such a comprehensive accurate results. The combination of GIS and statistical analysis which helps to simulate the spatial distribution of the employment status is a potential tool for human resource development research.

5. REFERENCE

Nicholas N. Nagle, Stuart H. Sweeney, Phaedon C. Kyriakidis(2010) "A Geostatistical Linear Regression Model for Small Area Data " *Geographical Analysis* ISSN 0016-7363

Grayson, T. H. (2000). *Address Matching and Geocoding*, Massachusetts Institute of Technology Department of Urban Studies and Planning.

Gerke J. Hoogstra, Jouke van Dijk, Raymond J. G. M. Florax (2010) "Determinants of Variation in Population–Employment Interaction Findings: A Quasi-Experimental Meta-Analysis" *Geographical Analysis* ISSN 0016-7363

Ousley, C. (2010). "A Geographic-Information-Systems-Based Approach to Analysis of Characteristics Predicting Student Persistence and Graduation." *Doctoral Dissertation The University Of Arizona*

C. Clark, "Urban population densities," *Journal of the Royal Statistical Society*. Vol. 144(4), pp. 490–496, 1951

Zhang, J., Y. Zhu, et al. (2010). "The Application in Spatial Distraction of Population of Gansu Province Based Oil Geostafisticai Analysis of ArcGIS (in Chinese)." *Northeast population* 31(3): 91-94.

Tzee Kiu Edwin Chow ·Yan Lin Niem Tu Huynh ·John M. Davis(2010) "Using web demographics to model population change of Vietnamese-Americans in Texas between 2000 and 2009" *GeoJournal* DOI 10.1007/s10708-010-9390-6

Zhou, Y., & Wu, J. (2005). *Modeling university enrollments with ArcGIS*. Proceedings of the 25th Annual ESRI International Users Conference (pp. 1-9). Redlands, California:ESRI.

Griffith, D. A., and L. J. Layne. (1996). "Uncovering Relationships Between Geo-Statistical and Spatial Autoregressive Models." In *Proceedings of the Section on Statistics and the Environment*, 91–96. American Statistical Association.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*.Cambridge: Cambridge University Press.

Schmitt, B., M. S. Henry, V. Pigué, and M. Hilal. (2006). "Urban Growth Effects on Rural Population, Export and Service Employment: Evidence from Eastern France." *Annals of Regional Science* 40, 779–801.

Zwick, R., & Sklar, J. G. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal*, 42 (3), 439-464.

Wang, F. (2006). *Quantitative methods and applications in GIS*. Boca Raton, Florida:Taylor & Francis Group.

Liu Dong, Feng Zhiming, Yang Yanzhao(2010)" Evaluation of Material-based Development of China Using GIS Methods"*Chinese Journal of Population , Resources and Environm ent* Vol. 8 No. 3 Septem ber 2010

Kyriakidis, P. C. (2004). "A Geostatistical Framework for Area-To-Point Spatial Interpolation." *Geographical Analysis* 36(3), 259–89.

Glavac, S. M., A. C. Vias, and G. F. Mulligan. (1998). "Population and Employment Interactions in the Growth of United States Micropolitan Centers." *Urban Geography* 19, 632–56.

6. ACKNOWLEDGEMENTS

The author would like to thank Prof. Zhu Qing from Wuhan University for his constant source of help.