Spatial information mining and visualization for Qinghai-Tibet Plateau's literature based on GIS

Wang Xuemei[a, b], Ma Mingguo[c]
[a]The Scientific Information Center for Resources and Environment , CAS, Lanzhou, 730000, China
[b]The Lanzhou Branch of the National Science Library, CAS, Lanzhou, 730000, China
[c]Cold and Arid Regions Environmental and Engineering Research Institute(CAREERI), Chinese Academy of Sciences, 260 Donggang West Road, Lanzhou, China 730000

ABSTRACT

The subject intersection becomes a hot research topic recently. This paper tried to couple the Bibliometrics and Geographical Information System (GIS) technologies for studying on the spatial information mining and visualization from the Qinghai-Tibet Plateau's literature. All the literatures about Qinghai-Tibet Plateau research were indexed in the ISI Web of Knowledge. The statistical tables about the authors were extracted from the papers by using the method of bibliometrics. The spatial information of the author's countries was linked with the GIS database. The spatial distribution was presented by the format of maps based on the GIS technologies. Comparing with the regular presentation forms of the bibliometrical analysis, the spatial distribution maps can afford more abundant and intuitive senses for the users.

Keywords: spatial information mining, spatial visualization, Bibliometrics, GIS, Qinghai-Tibet Plateau

1. INTRODUCTION

The Qinghai-Tibet Plateau is so important for the global climatic change, which is considered as one of the drivers and ampilfers of the global climatic change (Pan et al., 1996). Its temperature rising is earlier and higher than the global average level. Therefore, the researches on the Qinghai-Tibet Plateau have been a hotspot of geosciences since several decades. A lot of papers were published in these researches. Normally the research status and trends could be quantificationally analyzed based on bibliometrical methods (Nourbala, et al., 2008).

The bibliometrical analysis can be defined as the statistical exploitation of scientific publications in order to extract some available information on the research situations, which mainly include the activity of the producers (researcher, laboratory, institute, country, etc.) or distributors (periodical, publisher, etc.) of scientific information, both in quantitative and in qualitative terms (http://wiki.epfl.ch/infoscience/bibliometrics-en). Therefore, the bibliometrical analysis is one of typical information mining methods. It is widely used to assess the status and trends of a subject or research area.

The regular methods of bibliometrics mainly include information gathering, publication enumeration and statistical analysis. The normal presentation forms include statistical tables and statistical charts (Wang, *et al.*, 2007). Recently new presentation forms come forth continuously, such as network mapping, knowledge mapping (Hou *et al.*, 2009). Normally, the information extracted from the literature contains a lot of spatial position information. For example, the author belongs to a city of a country or region. For the geosciences, one or some study areas or sampling points during the field campaign are normally involved. If the information is spatially associated with the position coordinate, the spatial distribution could be mapped by using Geographical Information System (GIS) technologies. In this paper, the Qinghai-Tibet Plateau's literature was indexed. The spatial information involved in the authors was mined and analyzed. The spatial distribution within the information was presented by several maps, which aimed to afford more intuitive visualization for the users.

2. METHODOLOGY

International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, edited by Yaolin Liu, Xinming Tang, Proc. of SPIE Vol. 7492, 74920T · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.839525

The American Institute for Scientific Information (ISI) maintains the most comprehensive, multidisciplinary, bibliographic database of research information in the world. The ISI database covers over 16,000 international journals, books, and proceedings. It has better representation when it is used to search scientific papers and analyze the status and trends of some specific subject areas. In this paper, the scientific papers including article, proceedings paper, and review were indexed in the ISI Science Citation Index Expanded (SCIE). The publication year is from 1900 to end of April 2009. The indexed fields include title, keywords, and abstract. The theme words include tibet* or himalaya* or xizang* or qinghai* or qilian* or kunlun* or hengduan*. There are total 14017 papers indexed. The Thomson Data Analyzer (TDA) developed by Thomson Corporation was used for the data mining and analysis. The ArcView developed by ESRI Inc. was used to realize the spatial allocation, spatial analysis, and spatial mapping. The processing flowchart is shown in Fig. 1.

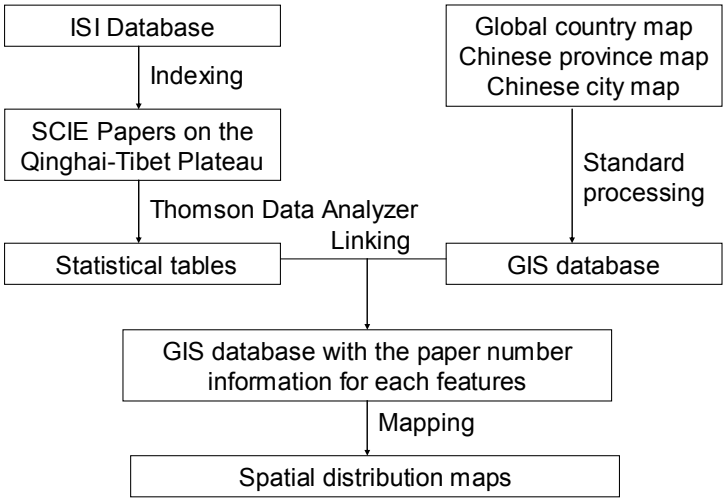

Fig. 1 Flowchart of spatial information mining and visualization for Qinghai-Tibet Plateau's literature based on GIS

3. RESULTS

3.1 Global spatial distribution of the countries with more publication papers

Table 1 shows the top 20 countries and paper numbers on Qinghai-Tibet Plateau for all authors. Fig. 2 shows the global spatial distribution of paper numbers on Qinghai-Tibet Plateau for all authors. From the Table 1 the countries with higher publication papers can be indicated clearly. But the position relationship of the spatial distribution can not be indicated. From the Fig. 2 both the countries with higher publication papers and their spatial distribution can be presented intuitively. Two formats were used to presented the numbers of the papers for each country. One is the graduated points. The other is polygon filling with the graduated color. These countries mainly distributed in Asia, Europe, and North America. In Asia, the countries (China, India, Nepal, Pakistan, and Japan) with more publication papers mainly distribute close to Qinghai-Tibet Plateau. In Europe, the countries (UK, France, Germany, Italy, and Switzerland) with more publication papers indicate centralized distribution characteristics. USA and Canada have more publication papers in North America.

3.2 Global spatial distribution of cooperation papers

International cooperation and exchange is the current developing trend in the global scientific research. Cooperation papers between countries on Qinghai-Tibet Plateau were shown in Table 2. China and USA have more cooperation papers with other countries. Even though India has the second paper number in the world, its cooperation records are a lot less than China and USA. The spatial distribution of the cooperation relationship was shown in the Fig. 3. The line links two countries which have cooperation researches. The line thickness shows the cooperation paper number. The cooperation relationships among countries can be shown directly from this map.

Table 1 List of the top 20 countries and papers on Qinghai-Tibet Plateau for all authors

ID	Country	Papers	ID	Country	Papers	ID	Country	Papers	ID	Country	Papers
1	China	4613	6	Japan	730	11	Nepal	243	16	Netherlands	112
2	India	2752	7	Germany	657	12	Italy	238	17	Norway	95
3	USA	2569	8	Canada	370	13	Russia	172	18	New Zealand	82
4	UK	853	9	Australia	300	14	Pakistan	146	19	Belgium	80
5	France	750	10	Switzerland	273	15	Austria	129	20	Spain	79

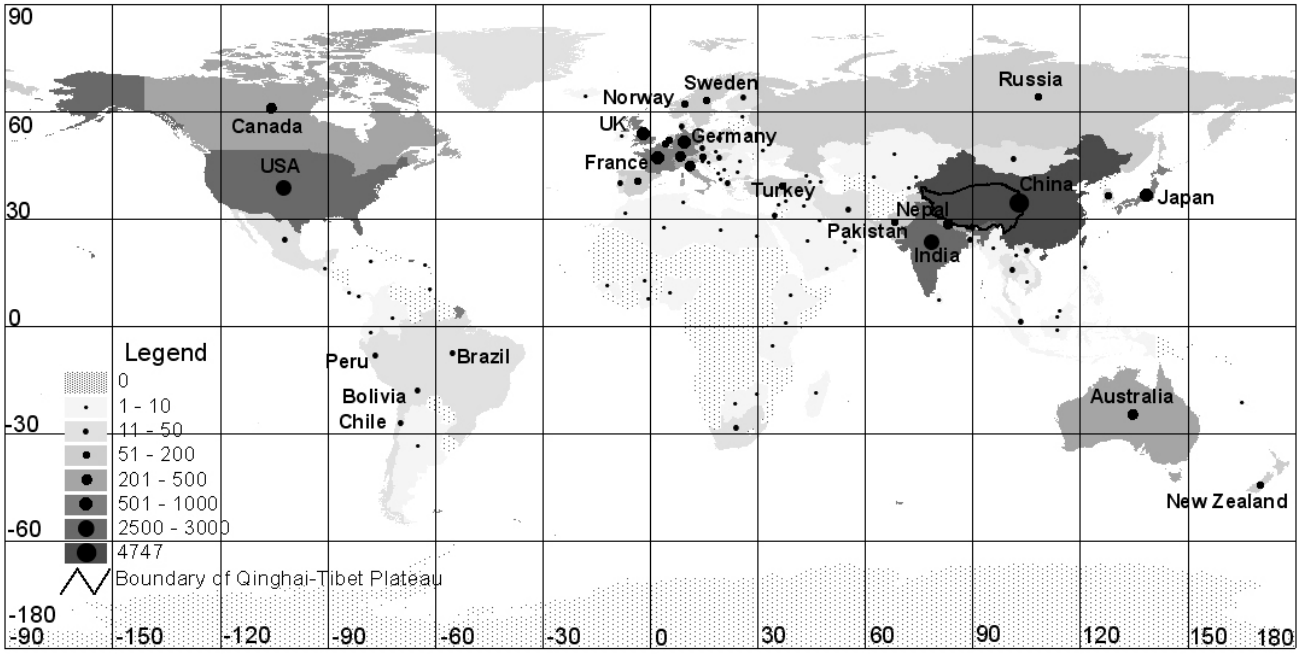

Fig. 2 Global spatial distribution of papers on Qinghai-Tibet Plateau for all authors

Table 2 Cooperation papers between countries on Qinghai-Tibet Plateau for all authors

	China	India	USA	UK	France	Japan	Germany	Canada	Australia
China									
India	22								
USA	788	141							
UK	198	58	155						
France	184	26	128	46					
Japan	313	40	74	23	17				
Germany	185	67	118	53	22	18			
Canada	108	16	81	25	23	4	19		
Australia	106	20	60	29	14	12	9	7	
Switzerland	68	6	42	28	49	7	37	9	4
Nepal	12	19	68	18	26	18	21	9	0
Italy	24	13	35	19	23	4	19	8	7
Russia	25	9	30	6	15	2	11	4	1
Pakistan	3	3	43	22	12	12	9	3	5
Netherlands	37	9	14	20	14	11	8	1	1

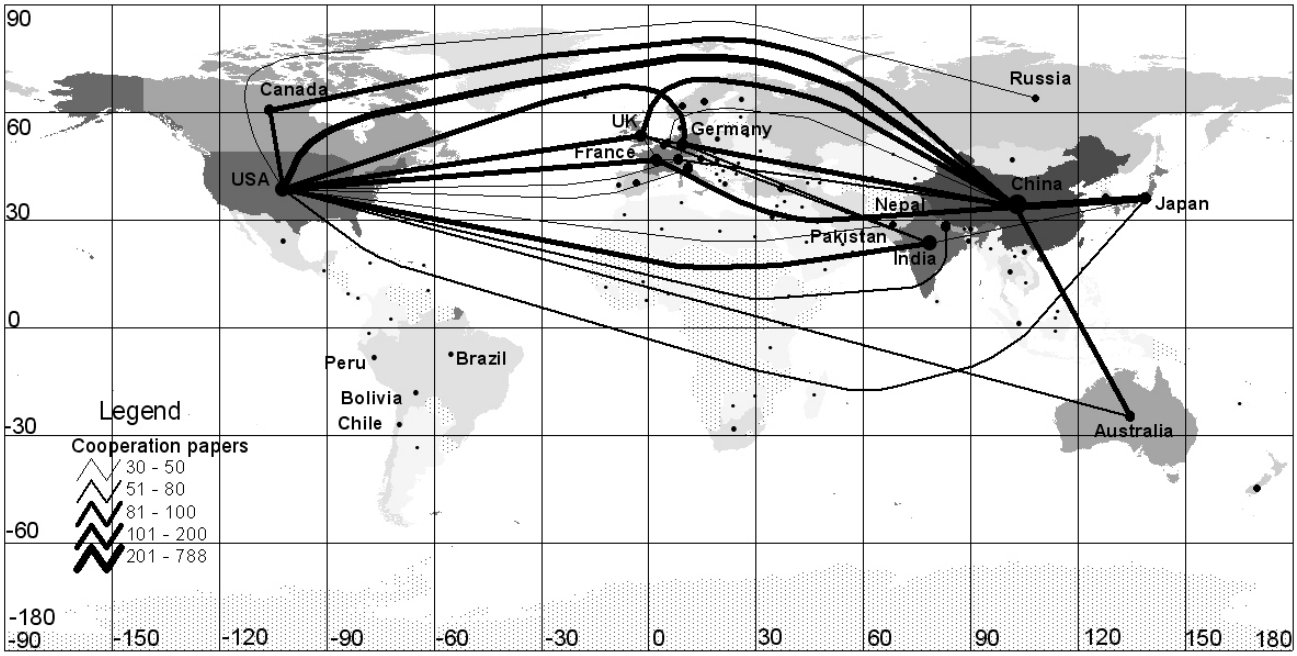

Fig. 3 Global spatial distribution of cooperation papers (≥ 30) between countries on Qinghai-Tibet Plateau for all authors

3.3 Global spatial distribution of publication year

The paper numbers during different periods were shown in the Fig. 4. From this map the countries with higher paper numbers can be presented clearly, such as China, India, and USA. But some countries can not be presented because of the much less papers. Therefore the Publication paper ratios during different periods were mapped in the Fig. 5. In this map, the change trend can be presented obviously for each country. The papers fast increase in recent decade for all the countries in the world. It also indicates an exponential increase trend. But there some differences among the countries. For example, China and Pakistan have the biggest increase speed among the recent three 5-years. On the contrary, France, UK, and Russia have gently increase trend.

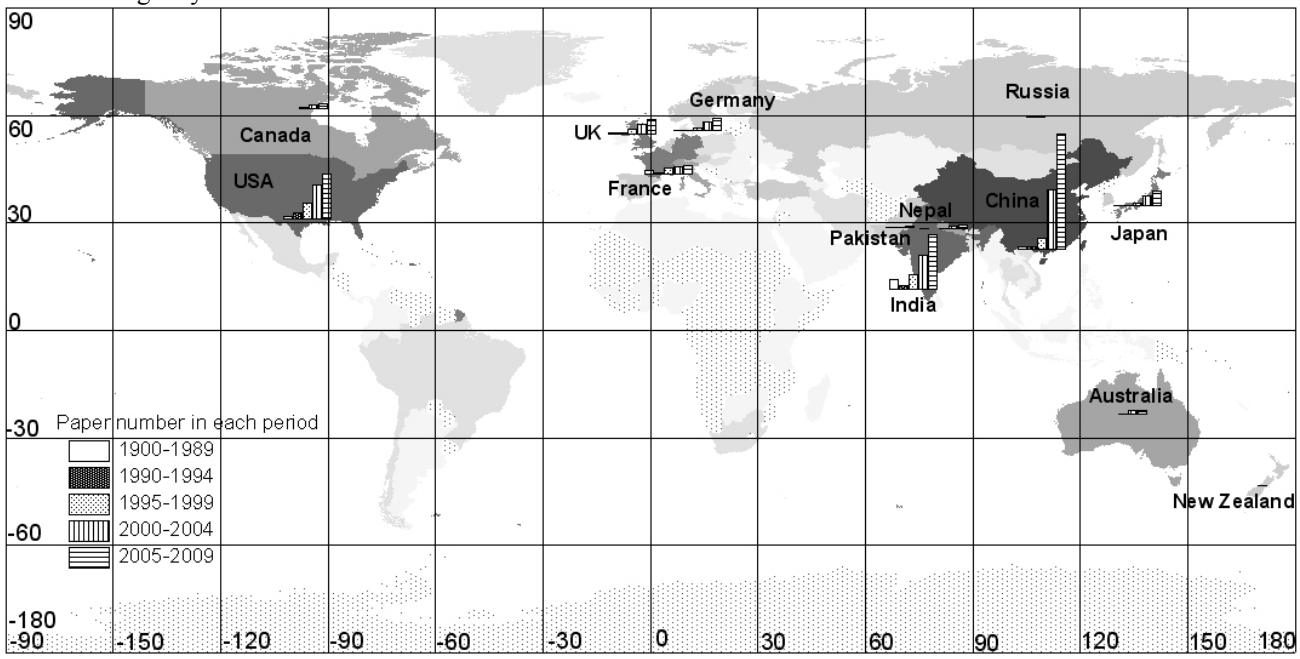

Fig. 4 Paper number during different periods on Qinghai-Tibet Plateau for all authors

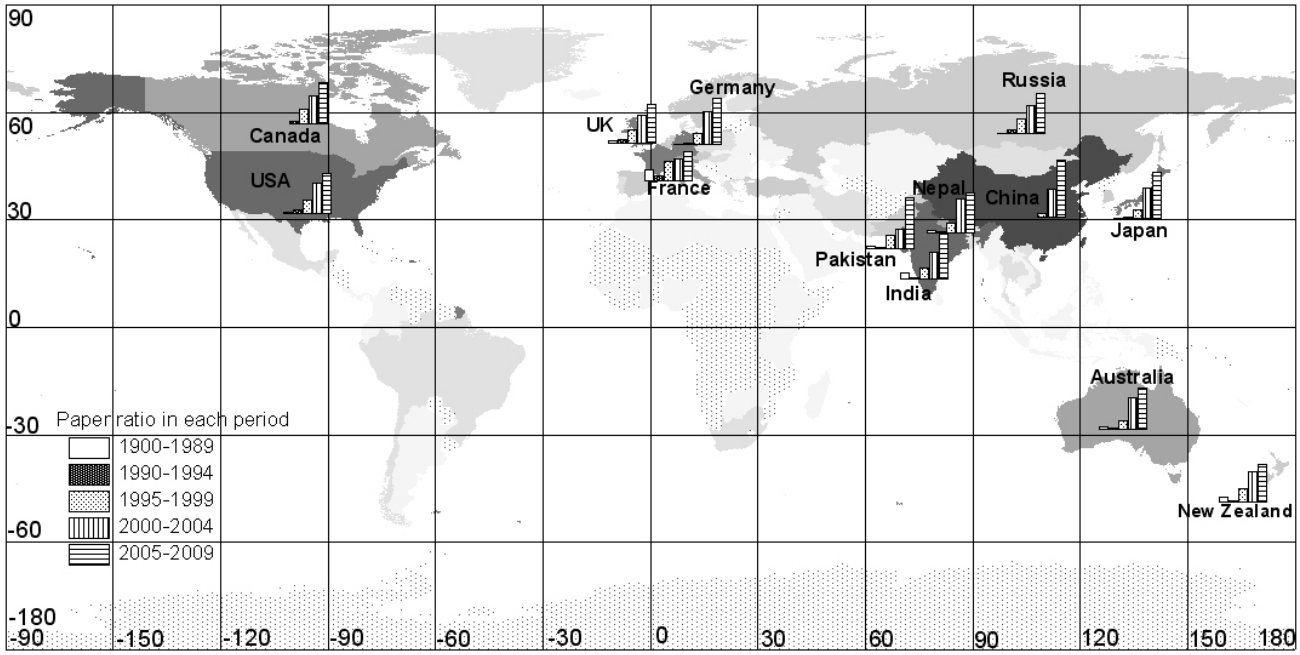

Fig. 5 Publication paper ratio during different periods on Qinghai-Tibet Plateau for all authors

3.4 Global spatial distribution of subject ratio

The spatial distribution of the subject ratio of the top 7 subjects and top 16 countries was shown in Fig. 6. The subject of Geosciences, Multidisciplinary occupies the biggest percentage in 11 countries. The subject of the Geochemistry & Geophysics occupies the biggest percentage in the other 5 countries. The second big percentage in most countries is often one of these two subjects. But in some countries other subjects occupy the second big percentage. For example, the subject of the second big percentage in Japanese is Meteorology & Atmospheric Sciences.

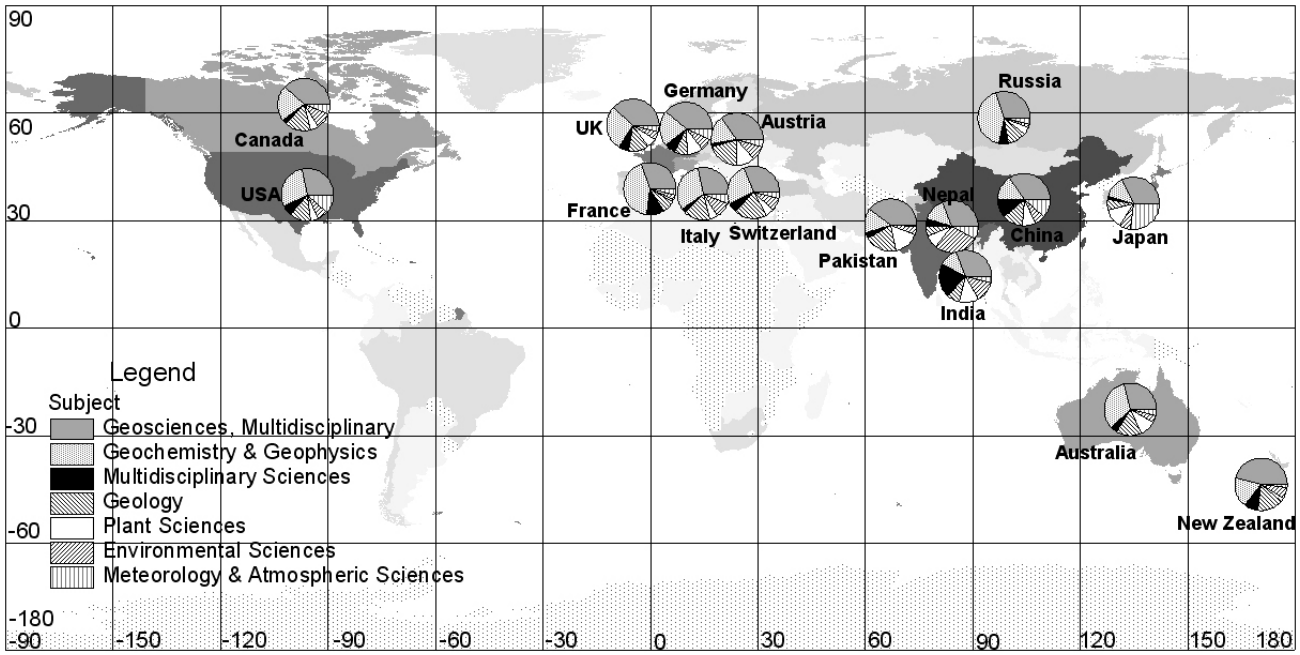

Fig. 6 Global spatial distribution of main subject ratio on Qinghai-Tibet Plateau for all authors

3.5 Chinese spatial distribution characteristics of papers

The most paper number on the Qinghai-Tibet Plateau is China. Therefore, the Chinese spatial distribution characteristics of papers were analyzed in detail (Table 3, Fig. 7). The papers of cities show that the provincial center cities have more papers normally. The cities with more than 300 papers include Beijing, Lanzhou, Chengdu, and Nanjing, which are the main research centers of geosciences in China. They are the branches of Chinese Academy of Sciences and have a lot of famous universities with directions of geosciences. The cities with 100 to 300 papers include Xining, Xi'an, Kuming, Wuhan, Guangzhou, Lhasa, Shanghai, and Hongkong. These cities all have more institutes and universities, which work on the research of geosciences. The spatial distribution of these cities shows that the cities closed to Qinghai-Tibet Plateau have more papers commonly. The other cities with higher paper number all distribute near the coast.

The sum of the paper number of the cities which belong to a province were calculated and the provincial paper numbers were achieved (Fig. 6). Then a 3-dimension visualization of provincial paper numbers on Qinghai-Tibet Plateau was made (Fig. 8). The map shows that a continuous distribution region exists close to the Qinghai-Tibet Plateau. This region includes six provinces of Gansu, Sichuan, Qinghai, Yunnan, Shaaxi, and Hubei. The other provinces with higher paper number are fragmentary in distribution.

Table 3 Paper number of the cities on Qinghai-Tibet Plateau for all authors in China

ID	Country	Papers	ID	Country	Papers	ID	Country	Papers
1	Beijing	2700	11	Shanghai	146	21	Hefei	40
2	Lanzhou	859	12	Hong Kong	142	22	Zhengzhou	38
3	Nanjing	425	13	Taipei	88	23	Chongqing	36
4	Chengdu	410	14	Shijiazhuang	66	24	Changsha	34
5	Xining	298	15	Guiyang	64	25	Shenyang	29
6	Xian	268	16	Urumqi	63	26	Qingdao	29
7	Kunming	268	17	Jinan	62	27	Tainan	23
8	Wuhan	237	18	Changchun	53	28	Harbin	12
9	Guangzhou	219	19	Hangzhou	47	29	Yinchuan	12
10	Lhasa	178	20	Tianjin	45	30	Kuerle	12

4. CONCLUSIONS AND DISCUSSION

The spatial position information of the authors was mined from the Qinghai-Tibet Plateau's literature indexed from the ISI database. The spatial distribution was presented by the format of maps based on the GIS technologies. Comparing with the regular presentation forms of the bibliometrical analysis, the spatial distribution maps can afford more abundant and intuitive senses for the users. The analysis results indicate that the countries with higher paper numbers on the Qinghai-Tibet Plateau mainly distribute in Asia, Europe, and North America. The countries close to the Qinghai-Tibet Plateau have more publication papers. China and USA have more cooperation papers with other countries. In most of countries, the subjects of geosciences, geochemistry & geophysics occupy the most percentage of the papers. The papers fast increase in recent decade for the all the countries in the world. It indicates an exponential increase trend. There are the most publication papers in China. The papers of cities show that the provincial center cities have more papers normally. The spatial distribution of these cities shows that the cities closed to Qinghai-Tibet Plateau have more papers commonly. The other cities with higher paper number all distribute near the coast.

In this paper, the spatial visualization information mainly focuses on the author distribution. It is just a beginning and attempt of the spatial information mining and visualization from Qinghai-Tibet Plateau's literature based on GIS. The more complex spatial distribution characteristics could be mined and visualized. For example, study areas, sampling points during the field campaign. These elements have more spatial correlation. Some geospatial analysis tools can be used to mine deep level information. These tools include spatial interpolation, buffer analysis, spatial correlation analysis, spatial autocorrelation analysis, geostatistics analysis, and so on. About the spatial visualization methods, the independent software would be developed to afford some spatial operation functions, such as spatial data query,

mapping, layer controlling. More information following the temporal change would be presented dynamically. The user can obtain detail information based on more human computer interaction.

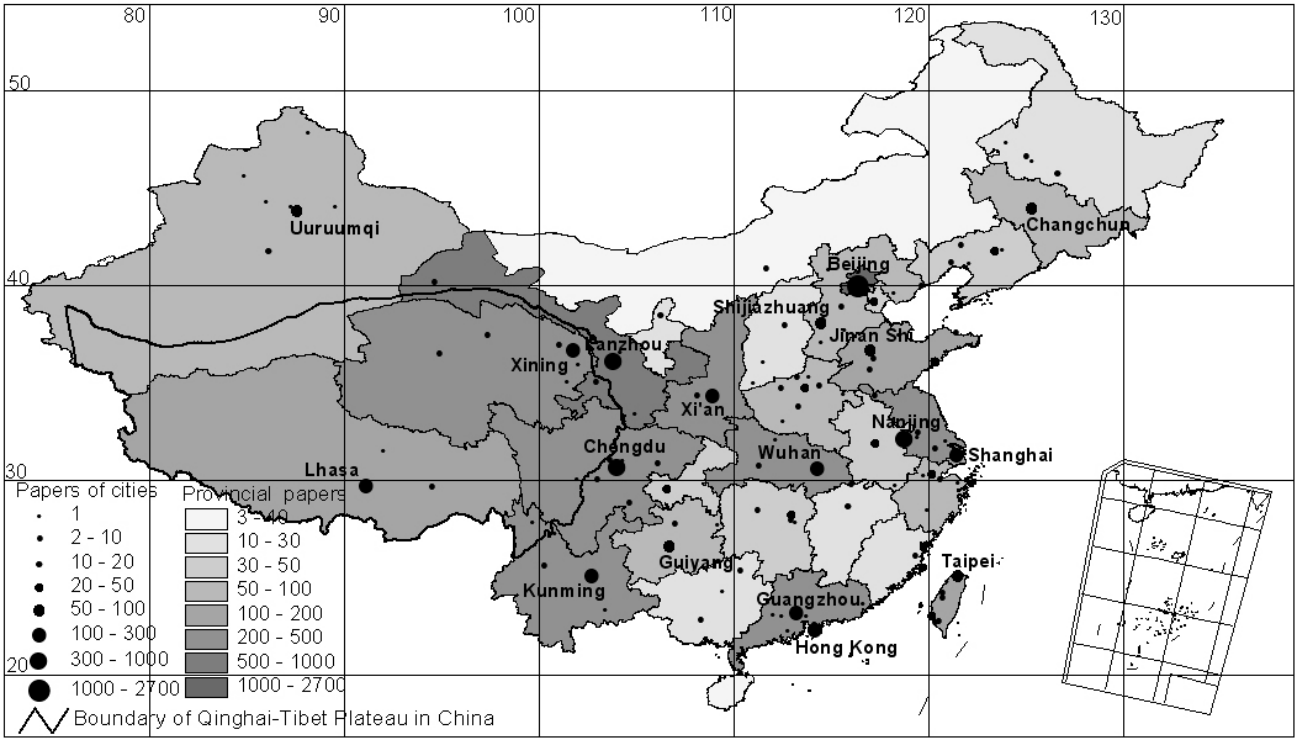

Fig. 7 Chinese spatial distribution of paper numbers on Qinghai-Tibet Plateau for all authors

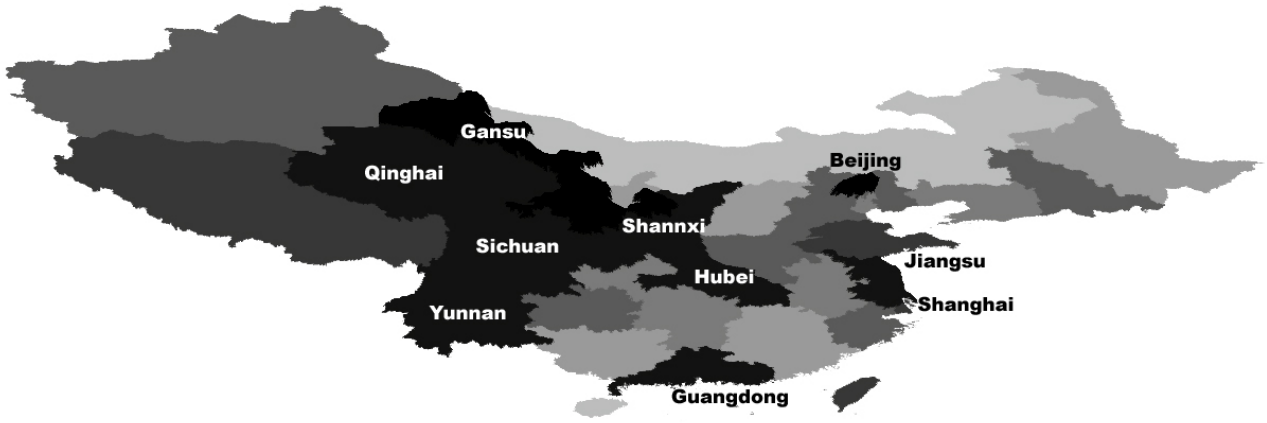

Fig. 8 3-dimension visualization of provincial paper number on Qinghai-Tibet Plateau for all authors in China

ACKNOWLEDGMENTS

This work is supported by the project of National Natural Science Foundation of China “Study on Qinghai-Tibet Plateau's literature based on GIS” (No. 40701133) and the West Light Foundation of The Chinese Academy of Sciences. The publication papers on the Qinghai-Tibet Plateau were indexed from SCIE database.

REFERENCES

1. Hou Haiyan, Liu Zeyuan, Luan Chunjuan. Quantitative analysis on the research front of international scientometrics based on mapping of knowledge. Science Research Management, 2009, 30(1): 164-170.
2. Nourbala MH, Taheri S, Habibi R, et al. "Transplantation" research output by Muslim Nations: Current status, trends and future outlook. Annals of Transplantation, 2008, 13(2): 21-27.
3. Pan Baotian, Li Jijun. Qinghai-Tibetan Plateau: A driver and ampilfer of the global climatic change. Journal of Lanzhou University (Natural Sciences), 1996, 32(1): 108-115. (in Chinese)
4. Wang Xuemei, Zhang Zhiqiang, Xiong Yonglan. A Bibliometrical Analysis of Status and Trends of International Ecological Footprint Research. Advance in Earth Science, 2007, 22(8): 760-765. (in Chinese)