

王雪梅, 李新, 马明国 等. 青藏高原科研文献地理信息空间分析研究[J]. 地球科学进展, 2012, 27(11): 1 288-1 294. [Wang Xuemei, Li Xin, Ma Mingguo, et al. Spatial analysis on the geographical information of the scientific literatures for Qinghai-Tibet plateau[J]. Advances in Earth Science, 2012, 27(11): 1 288-1 294.]

青藏高原科研文献地理信息空间分析研究*

王雪梅^{1,2} 李 新² 马明国² 张志强¹

(1. 中国科学院国家科学图书馆兰州分馆/中国科学院资源环境科学信息中心, 甘肃 兰州 730000;
2. 中国科学院寒区旱区环境与工程研究所, 甘肃 兰州 730000)

摘 要: 多学科的协同、渗透、交叉已成为当代科学发展的重要趋势, 文献计量学和地理信息系统 (GIS) 技术的集成是拓展 GIS 应用领域的一个新兴方向。许多与地理空间位置有关的信息包含在大量地学科研文献中。从 CNKI 文献数据库收录的研究青藏高原的科研文献中提取研究区、样带和采样点等位置相关信息, 运用 GIS 技术进行空间分析和展现。结果显示, 青藏高原研究的主要学科领域集中在生态环境及其变化、地质与地球物理特征及其演化、资源及其勘探、地质工程等方面。有关青藏高原的研究在空间分布上存在较大的差异, 对青藏高原的研究从西南、西北到东南、东北区域, 其受关注的程度逐渐增加。祁连山脉、羌塘高原、青海湖、青藏公路、青藏铁路和川藏公路等是研究的热点区域。沿经度和纬度方向不同区间研究的分布也不同, 研究热点区域呈现出从青藏高原中部向东北部转移的趋势。多数采样点和观测点沿交通干线分布, 并且随着与交通干线距离的增加数量明显减少。GIS 技术在关于青藏高原的文献计量分析中得到了有效应用, 有助于挖掘出大量与空间位置相关的信息, 从而深入分析青藏高原研究的发展态势。

关 键 词: 青藏高原; 研究区; 采样点; 空间分析; 文献计量学

中图分类号: P9; G250.252 文献标志码: A 文章编号: 1001-8166(2012)11-1288-07

1 引 言

随着人为活动产生的全球温室气体排放的增加, 全球气候变化日益显著。许多气候指标显示地球正在变暖, 近百年来地球平均气温升高了 0.4 ~ 0.8 °C^[1,2]。青藏高原对外界因素的扰动具有高度的脆弱性和敏感性, 是全球变化研究的热点区域之一, 被称为全球气候变化的“驱动器”和“放大器”^[3]。青藏高原由于其独特的形成、演化历史及其在区域乃至全球地球系统演化中的特殊地位与作用, 是国际地球科学、生命科学、资源环境科学方面

具有独特优势的研究区域和难得的天然实验室, 青藏高原研究对了解岩石圈地球动力学和全球环境变化有着重要的意义, 对高原区域可持续发展也有广阔的应用前景^[4]。

对该地区的研究已形成了相当的科学积累并发表了大量研究论文。传统的文献计量学主要借助文献的各种特征的数量, 采用数学与统计学方法来描述、评价和预测科学技术的现状与发展趋势^[5]。多学科的协同、渗透、交叉已成为当代科学发展的重要趋势, 文献计量学和地理信息系统 (GIS) 技术的集成是一个新的发展方向。GIS 在科研文献空间分析

* 收稿日期: 2012-04-07; 修回日期: 2012-06-27.

* 基金项目: 中国科学院资源环境科学与技术局委托任务“资源环境科技发展态势监测分析与战略研究”(编号: 2011-2015); 中国科学院“西部之光”项目“我国资源环境科学发展的空间演替研究”; 国家自然科学基金项目“基于 GIS 的青藏高原科研文献知识发现研究”(编号: 40701133) 资助.

作者简介: 王雪梅(1976-), 女, 重庆永川人, 副研究员, 主要从事资源环境科学计量评价研究. E-mail: wxm@lzb.ac.cn

方面的应用主要体现在以下方面: ①将馆藏文献资源的分布表现在电子地图上,使读者对文献资源的空间分布状态有直观的了解,有利于读者快速定位资源、提高资源的检索效率^[6]。②把科研论文、科研成果、历史资料、图片等信息与空间数据相关联,使用户可以方便地查询、获取相关文献信息。③基于 Google Earth 等 WEBGIS 资源开发文献信息的空间显示和查询系统。④分析文献信息的空间分布特征以及相互之间的关联性。Hengl 等^[7]利用地统计分析方法从文献信息中获取热点区域。Agnieszka 等^[8]利用 GIS 技术分析波兰不同地区 WoS 论文发文数量、合作论文数和合作类型的空间分布特征。⑤利用 GIS 分析和展现历史事件或遗址。Ma 等^[9]把来自文献的红军长征路线、时间和队伍人数等信息分布在地理三维空间上。Wang^[10]基于 GIS 中国北方全新世环境考古研究揭示了早期的人类迁移。

与地球科学相关的科研文献常会涉及空间信息,如研究区、采样点、观测点、抽样点、样带等。作者曾经对科研文献中著者地址项的国家地区信息进行分析,并基于 GIS 技术以多种形式的地图展现了其空间分布,但分析中没有涉及研究区域^[11]。本文对研究区、采样点和观测点的信息进行提取,并运用 GIS 技术进行空间分析和展示。研究的目的在于:集成 GIS 技术和文献计量学,发展新的信息挖掘方法,通过空间分析为基于文献计量的学科发展态势分析提供更丰富的信息。

2 数据与方法

2.1 研究对象

青藏高原(25°~40°N,74°~104°E)是中国最大、世界平均海拔最高的高原,总面积约 2.6×10^6 km²(图 1)。平均海拔超过 4 500 m,全球海拔超过 8 000 m 的 14 座高峰全分布在该地区,有“世界屋脊”和“第三极”之称。青藏高原由一系列高大山脉组成,青藏铁路、青藏公路、青康公路、川藏公路是穿越青藏高原东部和西部地区的主要通道。青藏高原具有地球上海拔最高、数量最多、面积最大的高原湖群区,面积 1 km² 以上的湖泊有 1 091 个,占中国湖泊总面积的 49.4%^[12],包括我国最大的内陆湖泊——青海湖。

2.2 数据采集

该研究主要针对青藏高原的中文文献,学科主题涉及地质学、气象学、生物学、地球物理学和环境科学。在中国知网全文数据库 CNKI 中检索期刊论

文和学位论文,检索词包括“青藏 OR 青海湖 OR 纳木错 OR 色林错 OR 当惹雍错 OR 阿牙克库木湖 OR 扎陵湖 OR 鄂陵湖 OR 羊卓雍错 OR 赤布张错 OR 乌兰乌拉湖 OR 昂拉仁错 OR 可可西里盆地 OR 沱沱河盆地 OR 羌塘盆地 OR 措勤盆地 OR 岗巴一定日盆地 OR 囊谦盆地 OR 昌都盆地 OR 比如盆地 OR 日喀则盆地 OR 拉萨盆地 OR 波林盆地 OR 江孜盆地 OR 伦坡拉盆地 OR 札达盆地 OR 风火山 OR 梅里雪山 OR 纳木那尼峰 OR 希夏邦马峰 OR 唐古拉山 OR 雀儿山 OR 南迦巴瓦峰 OR 巴颜喀拉山 OR 喜马拉雅山 OR 昆仑山 OR 冈底斯山 OR 祁连山”。数据库收录时间为 1994—2011 年,检索时间为 2011 年 7 月。

检索获得 5 211 篇全文数据,遴选出 3 970 篇含有研究区、采样点或观测点信息的有效数据。研究区类型包括区域(如羌塘高原)、交通线路(如青藏公路和铁路)、山脉(如祁连山、唐古拉山)、湖泊(如青海湖)、山峰、冰川等。其中,有 1 846 篇论文的研究对象为整个青藏高原,以下的空间分析未将其纳入。线状目标主要包括公路和铁路,通过计算 10 km 缓冲区将其转换为面状地物。山脉沿着主山脊线计算 20 km 缓冲区将其转换为面状地物。最终从 1 923 篇论文中提取出 2 013 个有效的研究区。

有的文献中包含 1 个或多个采样点(如钻孔点或测量点)、气象站点或地震监测点。这些点有的包含坐标信息,可以直接提取,有的只有站点名称需要人工赋予坐标值。最终从 462 篇论文中提取出 988 个有效的点数据。

3 分析结果

3.1 研究区的空间分布

图 2 展现了青藏高原不同区域论文数量的空间分布,从中可见其空间分布存在较大差异。对青藏高原的研究从西南、西北到东南、东北区域,其受关注的程度依次递增。祁连山脉是最重要的研究区之一,研究方向主要包括降水、土壤、水文、地质和冰川等。祁连山是西北地区石羊河、黑河和疏勒河 3 个内陆河流域水源的源头所在,这 3 个流域当前面临着严重的生态环境问题,受到国家和地方政府的高度重视,近期开展了一系列治理和保护项目,因此产生了一系列研究论文。其次是羌塘高原,研究方向主要包括地质构造(石油和天然气资源、地层、岩石)和地球物理(地震、构造运动)等,因在该区域发现丰富的石油储量而受到地学界关注。对青海湖的

研究主要包括水文和湖泊学、气候变化、古气候和同位素地质学。对青藏公路和青藏铁路沿线的研究很多,研究方向包括冻土、气候变化、工程地质和地质勘探等。川藏线也是热点区之一,研究方向主要集中在地质灾害方面。其他一些研究热点区主要分布在唐古拉山、巴颜喀拉山、昆仑山、喜马拉雅山、若尔盖盆地等山脉和盆地地区。

3.2 经纬度方向上的论文分布

图3显示的是沿纬度和经度方向的论文数量分布。沿纬度方向,关于青藏高原北部研究的论文数量明显多于南部地区,其中有2个高值区:32°~35°N,该区间包括羌塘高原和青藏高原东北部;37°~40°N,该区间包括祁连山、青海湖和青藏高原东北部。

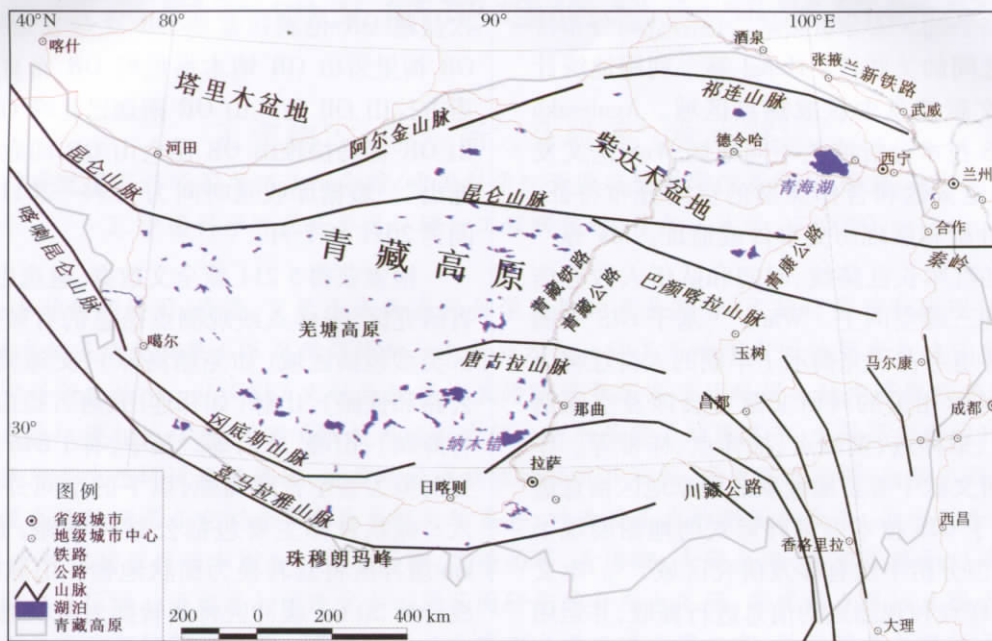


图1 青藏高原研究区

Fig.1 Study area of the Qinghai-Tibet Plateau

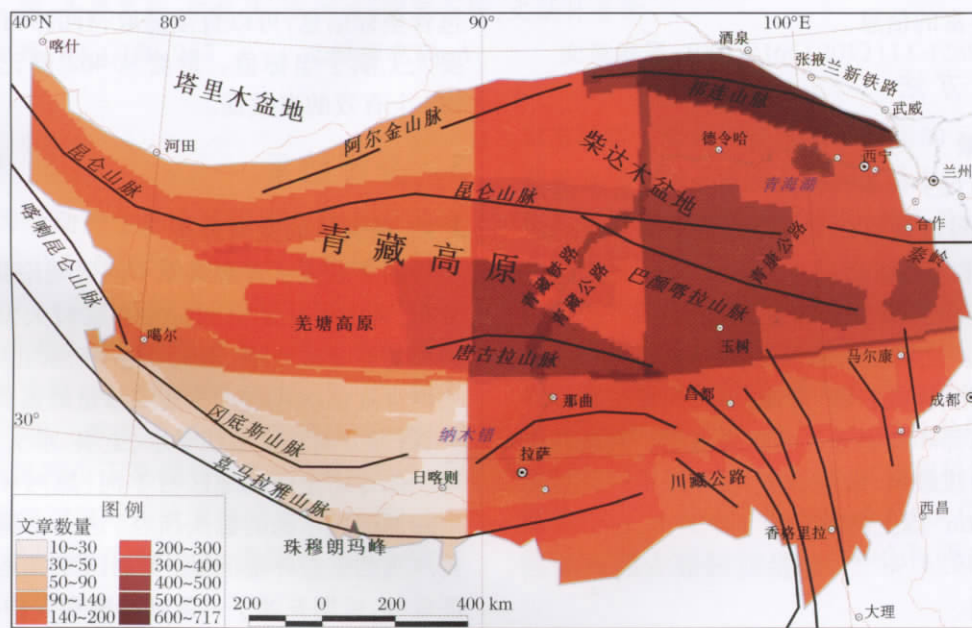


图2 青藏高原研究区论文数量的空间分布

Fig.2 Spatial distribution map of the paper numbers of Qinghai-Tibet Plateau research

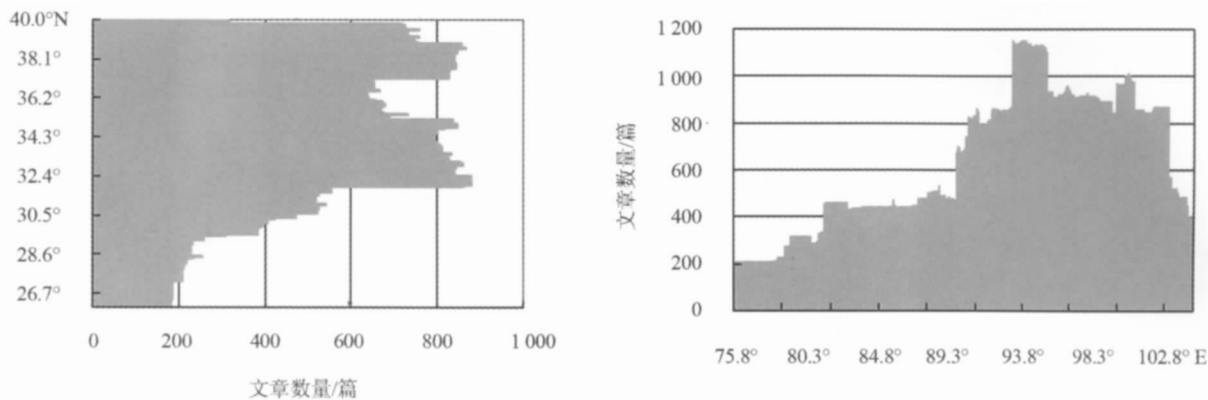


图 3 沿纬度(左)和经度(右)方向的论文数量

Fig. 3 Paper numbers along latitudinal direction (left) and longitudinal direction (right)

沿经度方向,关于青藏高原东部研究的论文数量多。在 90° ~ 103°E 的高值区间分布着许多热点研究区,包括祁连山、青海湖、青藏公路和铁路、羌塘高原东部、青藏高原东北和东部地区。尤其是在 93.25° ~ 95.35°E 区间论文数量最多,主要涉及唐古拉山、沱沱河、五道梁和青藏高原中部等地区。

图 4 和图 5 是 3 个不同时期沿纬度和经度方向的论文数量分布。这 3 个时间段分别是: 1994—1999 年、2000—2005 年和 2006—2011 年。从 2 幅图可以看出,与 1994—1999 年相比,进入 21 世纪以来关于青藏高原研究的论文产出明显增多,21 世纪初 2 个时间段的论文数量随着经纬度的变化而有所不同。

在纬度方向上,低于 32°N 的区间里,2000—2005 年和 2006—2011 年 2 个时期的论文数量接近。2000—2005 年期间论文涉及的研究区较多分布在

32° ~ 37°N 范围内,而 2006—2011 年期间论文的研究区更多地分布在 37° ~ 40°N,表明研究热点区域呈现出从青藏高原中部向北部地区转移的趋势。

在经度方向上,76° ~ 93°E 21 世纪初 2 个时间段的论文数量大致相同。2000—2005 年期间研究涉及 103° ~ 104.5°E 的论文数量多;2006—2011 年期间论文涉及的研究区更多地分布在 93° ~ 103°E,一定程度上表明关于青藏高原东部地区的研究有增长趋势。

3.3 采样点和观测点的空间分布

图 6 展示了采样点和观测点的空间分布。从中可见,很多样点都靠近交通干线,这些交通线路包括青藏公路、青藏铁路、青康公路、川藏公路、中尼公路等。还有些样点沿祁连山、喜马拉雅山、昆仑山等山脉分布。羌塘高原也有较多样点分布。

对交通线路进行缓冲区分析,分别计算 20 40,

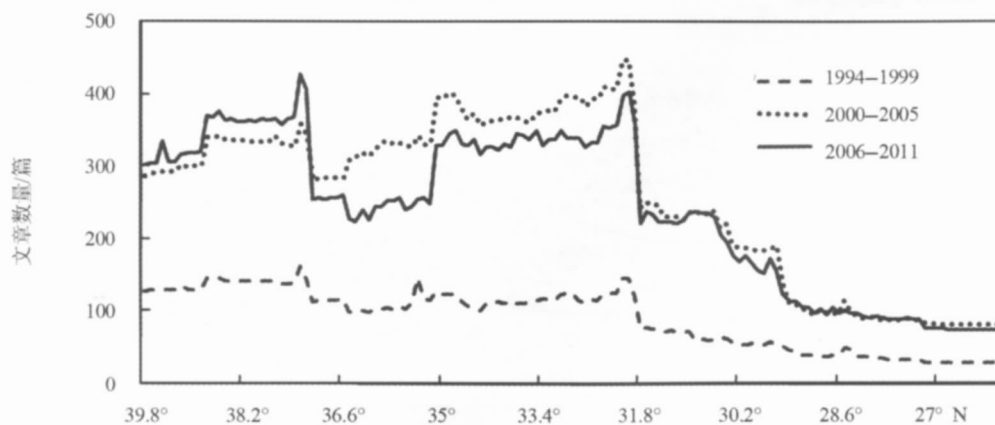


图 4 沿纬度方向 3 个时间段的论文数量分布

Fig. 4 Paper numbers along latitudinal direction during three periods

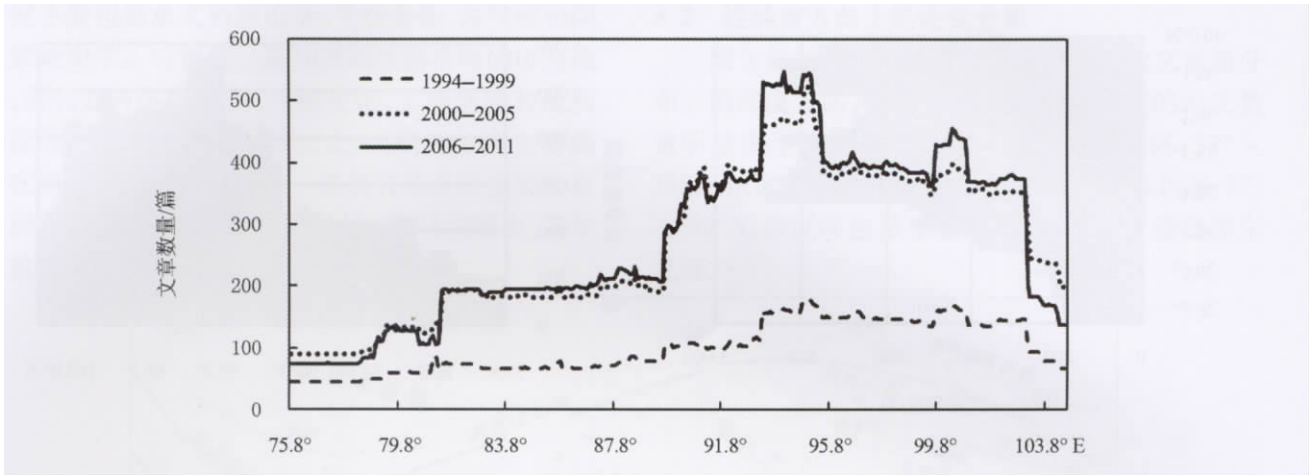


图 5 图沿经度方向 3 个时间段的论文数量分布

Fig. 5 Paper numbers along longitudinal direction during three periods

60 80 和 100 km 这 5 个缓冲区内的样点数量。图 7 显示了样点数量多少随缓冲区距离远近的变化趋势 随着与交通干线距离的增加样点数量明显减少, 统计结果显示有约 80% 的样点分布在沿交通干线 0 ~ 100 km 的区域内 表明交通线路是影响采样点和观测点分布的重要因素之一。

青藏高原上的道路修建本身存在大量科学问题 研究人员通过钻孔、气象站点、土壤和生态采样等手段获取大量点的信息 重点针对冻土、滑坡泥石流、生态保护等方面开展深入研究。另一方面 由于青藏高原大量区域属于无人区 环境恶劣 到达十分

困难 道路沿线开展采样相对容易。采样点的分布在青藏铁路和青藏公路沿线最为集中 而它们正是人员和物资进藏的关键线路 具有重要的战略意义, 科研人员通过数十年的研究积累解决了一系列世界屋脊上道路修建的科学难题。离交通线路较远的采样点分布在主要山脉、冰川、江河源、湖泊等周边 这些位置是研究地球系统演化和全球变化影响受关注的热点区域。

4 结论与讨论

GIS技术能有效地应用于研究区、采样点和观

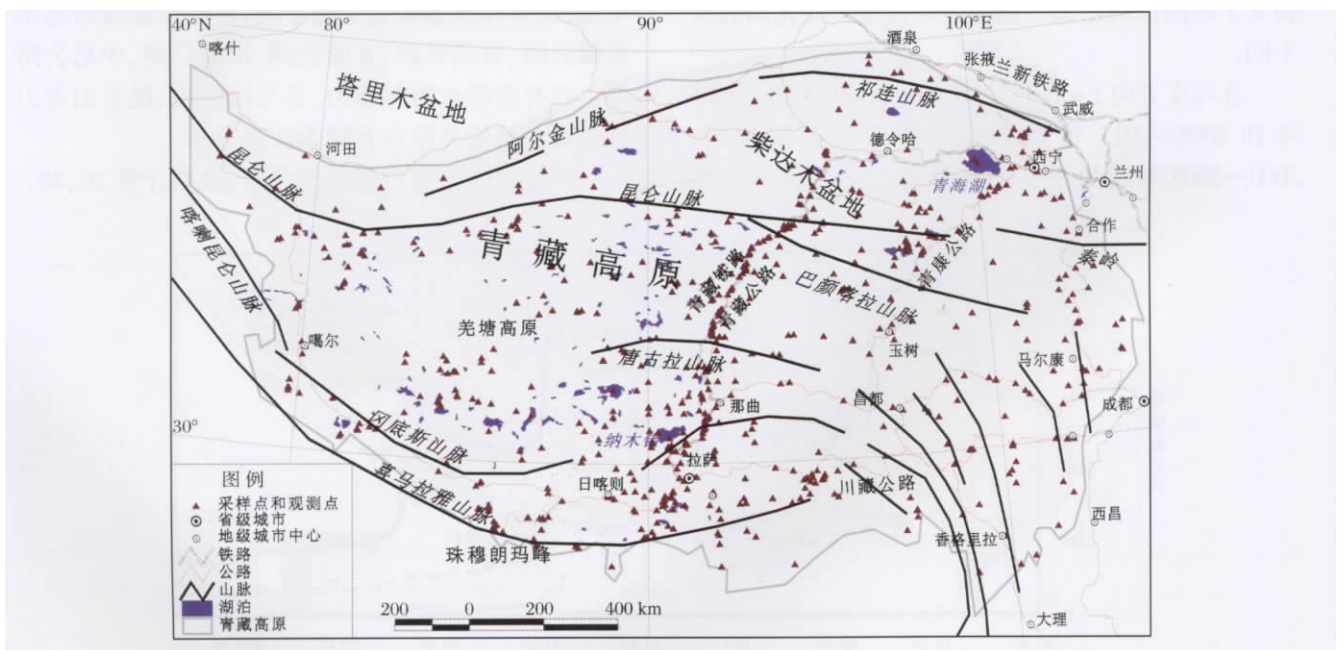


图 6 采样点和观测点的空间分布

Fig. 6 Spatial distribution map of the sampling and observing points

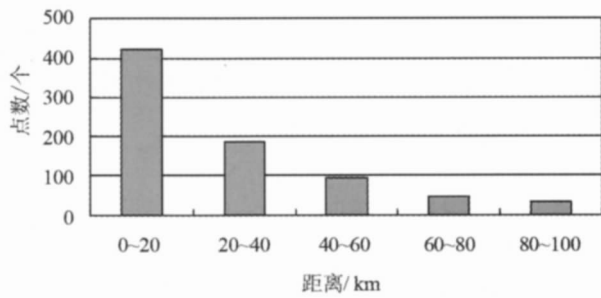


图 7 样点数量随缓冲区距离的变化

Fig. 7 Change trend of the sampling point numbers with the buffering distance

测点的空间分布特征及其发展趋势分析。对中国知网全文数据库 CNKI 中检索到的 3 970 篇带地理空间信息的青藏高原科研文献的分析表明, 这些论文研究区的空间分布存在明显差异, 在研究热点区的分布、经纬度方向的空间分布和缓冲区中的论文分布都有其规律和特征。

在这项研究中, 研究区、采样点和观测点的信息通过手工从文献中提取获得, 耗费了大量的人力物力, 因此有待发展自动化的信息提取方式。地名提取是目前最普遍的提取地理要素信息的工作之一, 这是因为地名相对规范和标准, 容易形成一个地理名称知识库。但是研究区、采样点和观测点千差万别, 很难形成地理名称知识库。有的文献中提供了地理坐标, 其信息易于识别提取, 但这些文献在总数中只占约 20%。有些地名可用于描述研究区、采样点和观测点的信息, 但如何从大量地名中提取少量有用信息是需要解决的难题。

该研究中还存在一些不确定性因素, 大多数研究区域的界线不是绝对的。例如, 通过缓冲区把有些研究区定义成一个规则的矩形, 但现实中真实的自然边界是不规则的形状。在我们的研究中, 自然边界尽量通过充分利用多源基础空间地理信息的方式进行采集, 如湖泊图、冰川图、道路图和铁路图。然而仍有许多研究区的边界难以准确界定, 如青藏高原北部、青藏高原南部、祁连山东部等。有些区域的范围是人为确定的, 例如把 32°N 作为青藏高原的南北界线, 这些界线还需要通过阅读大量参考文献进行细致讨论。

此外, 当前只对中文文献进行了代表性的空间分析, 英文文献的信息提取和分析还没有做。图 4 和图 5 中, 2006—2011 年间的论文总数与 2000—2005

年的接近, 一方面是因为 2011 年的论文检索不全, 另一方面也可能是近年来更多的文章发表在英文期刊上。因此, 需要进一步增加对英文文献的分析, 结果将会更加全面客观。

参考文献(References):

- [1] Hansen J, Sato M, Ruedy R, et al. Forcings and chaos in interannual to decadal climate change [J]. *Journal of Geophysical Research*, 1997, 102(D22): 25 679-25 720.
- [2] Konstantin Y V, Norman C G. Global warming trend of mean tropospheric temperature observed by Satellites [J]. *Science*, 2003, 302(5 643): 269-272.
- [3] Pan Baotian, Li Jijun. Qinghai-Tibetan Plateau: A driver and amplifier of the global climatic change—III. The effects of the uplift of Qinghai-Tibetan Plateau on climatic changes [J]. *Journal of Lanzhou University (Natural Science)*, 1996, 32(1): 108-115. [潘保田, 李吉均. 青藏高原: 全球气候变化的驱动力与放大器——III. 青藏高原隆起对气候变化的影响 [J]. 兰州大学学报: 自然科学版, 1996, 32(1): 108-115.]
- [4] Zheng Du, Yao Tandong. Progress in research on formation and evolution of Tibetan Plateau with its environment and resource effects [J]. *China Basic Science*, 2004, (2): 15-21. [郑度, 姚檀栋. 青藏高原形成演化及其环境资源效应研究进展 [J]. 中国基础科学, 2004, (2): 15-21.]
- [5] Pang Jing'an. *Scientometrics Research Methodology* [M]. Beijing: Science and Technology Literature Publishing House, 2002: 123-125. [庞景安. 科学计量研究方法论 [M]. 北京: 科学技术文献出版社, 2002: 123-125.]
- [6] Smith T R. A digital library for geographically referenced materials [J]. *Computer*, 1996, 29(5): 54-60.
- [7] Hengl T, Minasny B, Gould M. A geostatistical analysis of geostatistics [J]. *Scientometrics*, 2009, 80(2): 491-514.
- [8] Agnieszka O, Adam P. Mapping the regional science performance. Evidence from Poland [J]. *Collnet Journal of Scientometrics Information Management*, 2010, 4(1): 21-27.
- [9] Ma Ding, Meng L. The change of the number of troops in the Space Time Cube for the Red Army Long March [C] // Proceedings of SPIE. Wuhan, China, 2009, 74720X: 1-10.
- [10] Wang Lin. A GIS-Based Environmental Archaeology Study in Holocene Northern China [D]. Canberra: Australian National University, 2010.
- [11] Wang X M, Ma M G. Spatial information mining and visualization from Qinghai-Tibet Plateau's literature based on GIS [C] // Proceedings of SPIE. 2009, 747220: 1-8.
- [12] Jiang Jiahu, Huang Qun. Distribution and variation of lakes in Tibetan Plateau and their comparison with lakes in other part of China [J]. *Water Resources Protection*, 2004, (6): 24-27. [姜加虎, 黄群. 青藏高原湖泊分布特征及与全国湖泊比较 [J]. 水资源保护, 2004, (6): 24-27.]

Spatial Analysis on the Geographical Information of the Scientific Literatures for Qinghai-Tibet Plateau

Wang Xuemei^{1 2}, Li Xin², Ma Mingguo², Zhang Zhiqiang¹

(1. Lanzhou Branch of the National Science Library, Scientific Information Center for Resources and Environment, Chinese Academy of Sciences, Lanzhou 730000, China;

2. Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China)

Abstract: The multi-disciplinary cooperation and intersection becomes the main development tendency of the modern sciences. It is a new direction to expand the Geographical Information System (GIS) application areas by integrating the GIS technologies with Bibliometrics. A lot of geographic-related information is involved in the literatures concerned with geosciences normally involve some geo-science references. In this paper, the location-related information (*e. g.* study areas, sampling trips and sampling points) was extracted from the CNKI-indexed scientific literatures for Qinghai-Tibet Plateau. These data were spatially analyzed and presented by using the GIS technologies. The subject areas about the Qinghai-Tibet Plateau studies mainly focus on the ecological environment and its changes, geological-geophysical feature and evolution, natural resources and exploration, and geological engineering. The results indicate that there are big differences of the spatial distribution for Qinghai-Tibet Plateau researches. For the large scale, the degree of interest increases gradually as follows: southwest, northwest, southeast, and northeast. For the region scale, Qilian Mountains, Qiangtang Plateau, Qinghai-Tibet road and Qinghai-Tibet Railway, Qinghai Lake, and Sichuan-Tibet Highway are the hotspot regions. There are also big differences along the latitudinal direction and longitudinal direction, which indicates a transfer tendency from middle to northern Qinghai-Tibet Plateau. Most of the sampling and observing points are close to the traffic lines. The sampling point numbers decrease quickly with the increasing distance to the traffic lines. It is also indicated that the GIS technologies are effectively used in bibliometrical analysis of Qinghai-Tibet Plateau studies, which is of benefit to us for mining a great deal of spatial relative information and analyzing the developing trends.

Key words: Qinghai-Tibet Plateau; Study area; Sampling points; Spatial analysis; Bibliometrics.