Information Extraction and Spatial Distribution of Research Hot Regions on Rocky Desertification in China

Yunfeng Hu [1,2,*] , **Yueqi Han** [1,2] **and Yunzhi Zhang** [1,2]

[1] State Key Laboratory of Resources and Environmental Information System,
Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;
hanyq.16s@igsnrr.ac.cn (Y.H.); zhangyunzhi@lreis.ac.cn (Y.Z.)

[2] University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: huyf@lreis.ac.cn; Tel.: +86-10-64888020

check for updates

Abstract: Rocky desertification is an important type of ecological degradation in southwest of China. The author uses the web crawler technology and obtained 9345 journal papers related to rocky desertification from 1950s to 2016 in China. The authors also constructed a technological process to extract research hot regions on rocky desertification and then a spatial distribution map of research hot regions on rocky desertification was formed. Finally, the authors compared the spatial distribution using the sensitivity map of rocky desertification to find the differences between two maps. The study shows that: (1) rocky desertification research hot regions in China are mainly distributed in Guizhou, Yunnan and Guangxi, especially in Bijie, Liupanshui, Guiyang, Anshun, Qianxinan Autonomous Prefecture, QianNan Autonomous Prefecture, Qiandongnan Autonomous Prefecture in Guizhou Province, Hechi, Baise, Nanning, Guilin in Guangxi Zhuang Autonomous Region and Zhaotong in Yunnan Province. (2) The research hot regions on rocky desertification have good spatial consistency with the sensitivity regions of rocky desertification. At the prefecture level, the overlap rate of the two regions reached 85%. Because of the influence of topography, vegetation coverage, population distribution, traffic accessibility and other factors, there were regions that consisted of combinations of high sensitivity but low research popularity regarding rocky desertification; these sites included Qionglai Mountain-Liangshan Area of Sichuan, Wushan-Shennongjia Area of Hubei, Hengduan Mountain Area of western Yunnan and Dupangling Area of southern Hunan. (3) The research hot regions and sensitive regions cannot be matched completely in time, space and concept. Therefore, we can use their spatial distribution differences to improve the pertinence of planning, governance and study of rocky desertification.

Keywords: karst; rocky desertification; hot regions analysis; information extraction; spatial distribution; comparison analysis

1. Introduction

Rocky desertification refers to a phenomenon of land degradation, which includes the destruction of surface vegetation, serious soil loss, the large-scale deposit of exposed bedrock or gravel; these processes are affected by human activities in tropical and subtropical areas with humid and semi-humid climatic conditions and extremely developed karst environment [1,2]. The rocky desertification regions of China are mainly distributed in southwestern China with a wide coverage and a large population of residents; this region represents an ecologically fragile area with the largest karst landform and the most seriously impacted region affected by rocky desertification worldwide [3–5]. Severe rocky desertification has affected the economic and social development of the region and has

reduced the level of well-being of the local people [6]. Under the major goals of the central government and the 17 major goals of the United Nations SDG in 2020, it is of great significance to implement targeted ecological control and anti-poverty work in the rocky desertification areas.

In traditional studies, to master the spatial distribution of rocky desertification, researchers must first rely on geological surveys and geographical national surveys to obtain the exposed location of ground rock and the information related to agricultural habitat. Then, based on a combination of field survey sample information and satellite remote sensing data, an indoor mapping operation can be conducted using artificial visual interpretation and multi-spectral and hyperspectral spatial clustering [7–9]. Finally, the spatial distribution of regional rocky desertification and intensity can be obtained. A large number of studies have been conducted on the rocky desertification formation and developmental evolution mechanisms, the methods for the restoration and reconstruction of degraded ecological environments and the evaluation of their effects [10–12]. These studies have formed a rich text database that includes research reports, papers, statistical yearbooks and government bulletins. These studies reflect the status and locations of rocky desertification.

With the rapid development of Internet search engines and big data technologies, the methods of extracting geographic information from text have been transformed from manual readings to methods that use web crawlers, text mining, machine learning, data modeling to conduct data retrieval, information extraction, geospatial analysis and spatial mapping from online or offline big data and then the results are analyzed in combination with the models and methods in specialized fields [13–17]. There are many related works about text mining with explicit or implicit geographic messages. In the field of public health detection, Adam Sadilek and his colleagues detected illness-related messages and predicted real-time infection rates of influenza from geo-tagged Twitter data [18]. For implicit geographic information in text, Mark Dredze and his colleagues developed a geolocation system called Carmen to assign a location to each tweet and applied the results to improve influenza surveillance [19]. In the field of nature disaster detection, Takeshi Sakaki and colleagues tried to estimate the location of an earthquake event using tweets with locations through the use of spatial models [20]. Mark A. Cameron and his colleagues developed a platform called the ESA-AWTM system to detect, assess, summarize and report messages of interest for crisis coordination published by Twitter and formed a Tweet Map [21]. In addition to social media data, the literature can also be used as a data source. Many tools with similar technical processes of data mining algorithms have been developed to extract gene expression, protein messages or other biological information with locations, such as BioCreative, PubGene, MEDIE and Pubmatrix [22–25].

In the process of information extraction, there are some mature technological methods for retrieving data and crawling for Internet information [26,27], such as word segmentation and entity recognition technology [28–30]. However, it should be noted that most of the past research has focused on specific events, such as epidemics and natural disasters. These events usually explode in the Internet news media in a short period of time. Studies focused on natural and ecological evolution over a long-term scale are relatively rare. In addition, because most research has focused on specific events, the geographic information is usually presented as points after geocoding [31–34]. However, for natural and ecological evolution, these processes occur in a region rather than a point. In other words, when text mining is combined with research on the physical geography and ecological environment, how to recognize natural geographical divisions, ecological degradation and ecological support for professional issues such as engineering science layouts is important. Especially in terms of the identification of ecological degradation and spatial mapping, the construction of scientific and reasonable models for degradation type identification and spatial location inference is key to the study of ecological degradation data.

Furthermore, if we consider the literature to reflect the status and locations of rocky desertification, whether the reflection is comprehensive, detailed and accurate is critical. There may be a region with severe rocky desertification that has not received any scholarly attention; in contrast, there may be a region of slight rocky desertification that has received too much attention. To ascertain the above

problems, in this paper, we propose a method for extracting rocky desertification research hot regions and construct a hot regions spatial distribution database of rocky desertification in China based on the China National Knowledge Infrastructure (CNKI) and the algorithms of keyword searches, Chinese word segmentation, address recognition and toponym matching. Then, we compare the spatial distribution of rocky desertification research hot regions and the spatial distribution of sensitive regions of rocky desertification based on multi-source satellite remote sensing. Finally, we analyze and discuss the effectiveness and accuracy of research hot regions extraction methods and provide a scientific basis for the selection of rocky desertification research areas and engineering planning.

2. Materials and Methods

2.1. Data and Preprocessing

In this study, the basic knowledge database for extracting the spatial distribution information of rocky desertification is the Chinese periodical database provided by the CNKI. The CNKI is a database system based on various literature and documents and it has become the world's largest Chinese knowledge engine and library. By 31 August 2017, the China Academic Journals full-text database (CAJ) in the CNKI included 10,898 journals, 1,700,252 articles and 60,496,267 articles. The authors compiled a special crawler system of the CNKI in JAVA and retrieved all 9345 documents related to rocky desertification from the 1950s to 2016 using the three keywords of 'rocky desertification', 'karst desertification' and 'karst degradation.' The authors also downloaded the title, abstract, keywords, author and other information and saved these data to the local database. The local database was implemented on SQLite. SQLite is an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine. These merits of SQLite make our information extraction method of research hot regions can be migrated to different computing platform.

The basic spatial data used for address recognition, toponym matching and spatial mapping in the study came from the 2012 version of the Chinese administrative division map provided by China Cartographic Publishing House. By considering the abbreviation, alias or historical name of administrative divisions, the study also used the 1:250,000 basic geographic database provided by the National Administration of Surveying Mapping and Geoinformation of China and the China's historical-level administrative division database provided by the National Earth System Science Data Sharing Infrastructure to revise partial toponyms and supplement some aliases of toponyms. All of the above databases comprised a standard toponym database (STD) in this study.

To analyze the sensitivity to rocky desertification and to compare the sensitivity and hot regions of rocky desertification, this study also uses (Figure 1)

1. The relief amplitude spatial distribution data calculated from a digital elevation model (DEM) with a 1-km resolution;
2. The vegetation coverage spatial distribution data calculated using the dichotomy model with MODIS NDVI data [35,36];
3. The 1-km grid population dataset of China (2010) based on the data of the Fifth National Census [37];
4. The road network density above the county level, which is based on the 1:25 million basic geographic database provided by the National Geomatics Center of China.
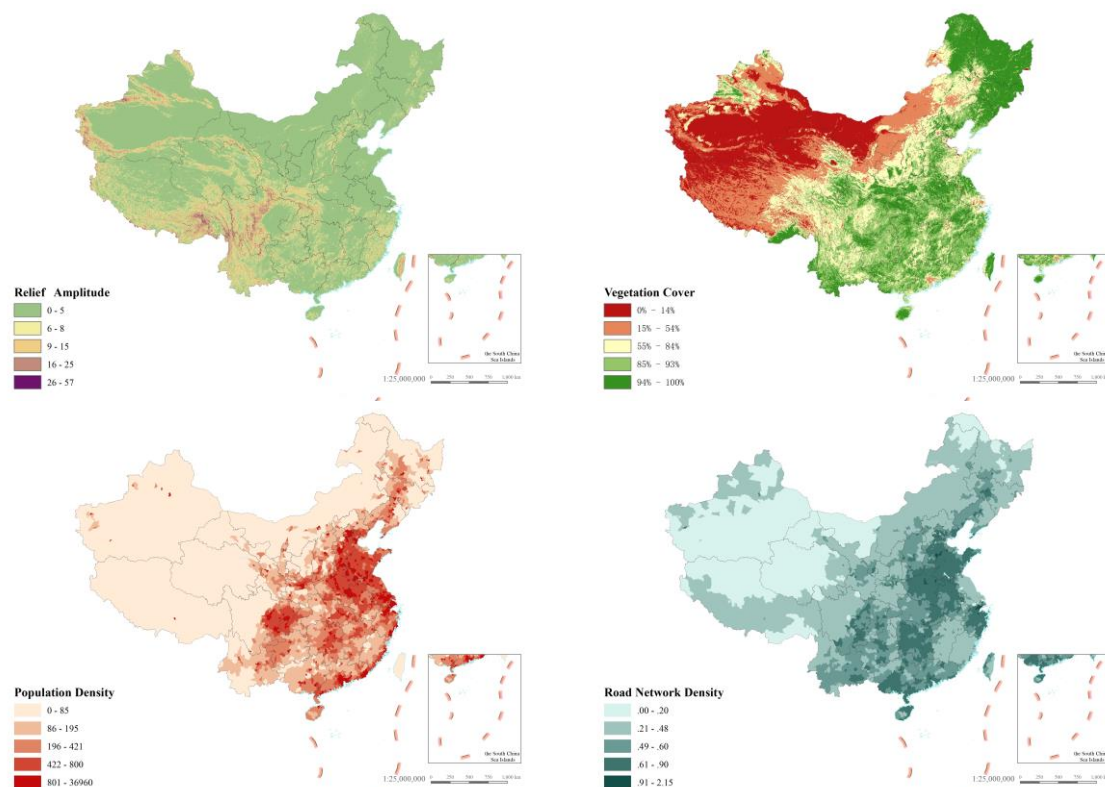
Figure 1. The Relief Amplitude, Vegetation Coverage, Population Density and Road Density Maps for China.

2.2. Analysis of Research Hot Regions

After retrieval, the study processes the Chinese word segmentation using the Ansj segmentation model for the title, keywords and abstract content in the rocky desertification article database and then uses the HanLP entity word recognition module to identify toponyms from the word segmentation results [38,39]. Then, a specific algorithm is used to perform matching between the above identified toponym and the standard toponym in the STD.

In the process of matching toponyms, it is necessary to judge the spatial inclusion relationship of the administrative regions and count the frequency of toponyms in all literature. The ultimate goal is to unify multi-level, non-standard, ambiguous toponyms into county-level, canonical and unique toponyms and to identify the frequency of occurrence of each county unit in the literature—both scientifically and rationally. Therefore, the study establishes a toponym matching process of "step-by-step coverage and accumulative statistics"; thus, the different expressions (e.g., full name, abbreviation, or alias) and the different levels (e.g., province, prefecture, or county level) of the same place name can be identified and counted accurately and finally, the statistical results are normalized.

The process of toponym matching is shown in Figure 2. First, we divided all toponyms in the STD into three levels—province, prefecture, or county. When we obtained a toponym from the literature, we determined whether this toponym was at the province level. If it was, then the frequency of all counties belonging to this toponym will be increased. If not, we continued to determine whether the toponym contained a province-level toponym. If it did, we matched the toponym with all prefectures belonging to the province using the Knuth-Morris-Pratt (KMP) algorithm to perform character-level fuzzy matching [40]. We used this step to void the mistakes that occur when there are two identical prefecture names in different provinces. If matching has still not occurred, we continued to determine whether the toponym was at the prefecture level and all of the other steps were carried out as listed above; however, the steps were carried out to determine the frequency of all counties belong to this

prefecture toponym. The determination process will continue until it cannot be matched with every toponym at the county level.
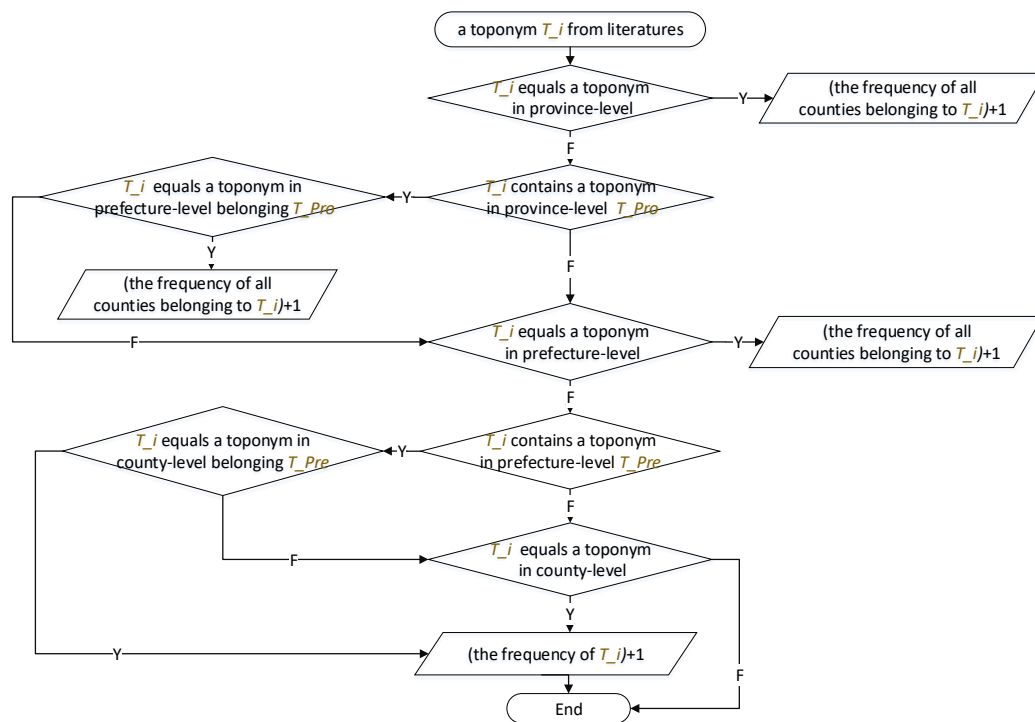

Figure 2. The Flowchart of the Chinese Toponym Matching Algorithm.

2.3. Sensitivity to Rocky Desertification

Eco-environmental sensitivity refers to the degree of ecosystem sensitivity to the disturbances caused by natural and human activities within a region [41]. This sensitivity reflects the degree of difficulty and the likelihood of ecological environmental problems occurring when regional ecosystems encounter disturbances and the value is used to characterize the possible consequences of external disturbances. Considering the systematisms of indicators and the availability of indicator data, researchers have constructed a sensitivity assessment model for rocky desertification [42–44]. The sensitivity assessment model method has been adopted for national ecological zoning.

The specific process of the evaluation of sensitivity to rocky desertification is as follows: based on the percentages of exposed carbonate area, vegetation coverage and terrain slope, the spatial distributions of the key influencing factors of rocky desertification sensitivity are obtained by the comprehensive dividing method based expert knowledge. Then, using the calculation method shown below, we obtained the rocky desertification sensitivity index [43] (Figure 3).

$$R = \sqrt[3]{M \times S \times C} \tag{1}$$

In Equation (1), R is the rocky desertification sensitivity index, M is the percentage of the area exposed by carbonate, S is the topographic gradient of the region and C is the regional vegetation coverage.
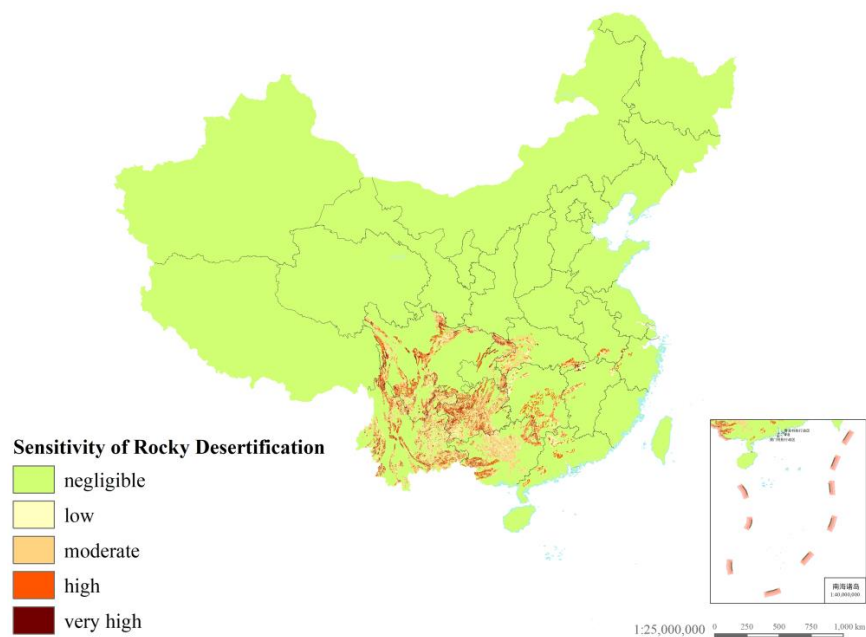
Figure 3. The Grid Map of the Rocky Desertification Sensitivity Index in China.

3. Results

3.1. Research Hot Regions Mapping

Figure 4 shows the spatial distribution of rocky desertification research hot regions in China based on the literature from the 1950s to 2016. The research hot regions on rocky desertification are mainly concentrated in three provinces in southwest China, that is, Guizhou, Guangxi and Yunnan. There are also a few rocky desertification studies in Sichuan, Hunan, Hubei, Chongqing and other provinces. Among them, the hot regions of rocky desertification with a moderate degree or greater of research are mainly distributed in western and southern Guizhou, eastern Yunnan and northern and western Guangxi, including 108 counties and a total area of 245.7 thousand km^2.
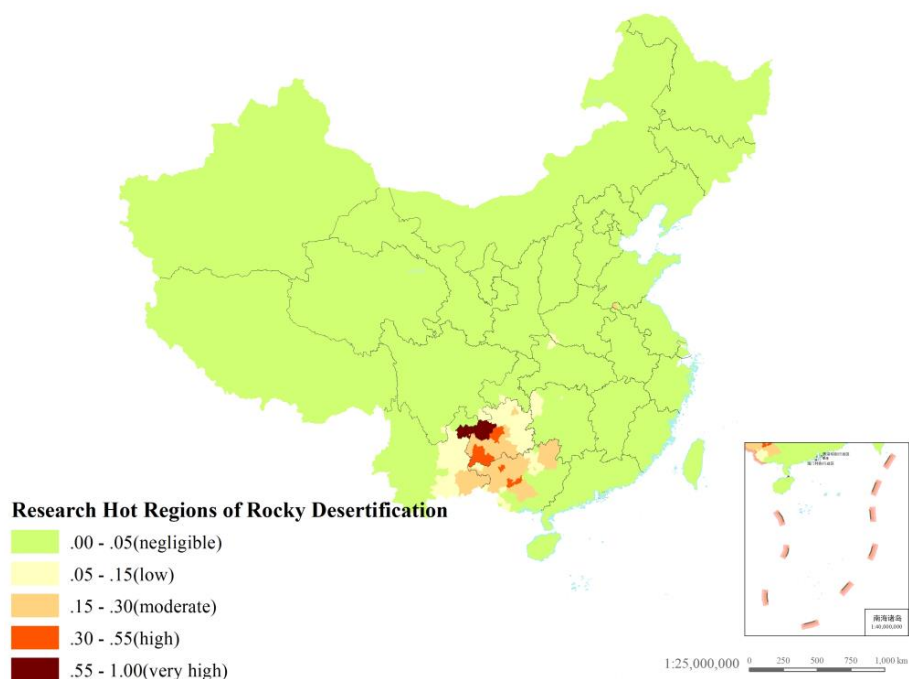

Figure 4. The Map of Hot Regions of Rocky Desertification Research in China.

Specifically, in terms of rocky desertification, the most literature was found in relation to Bijie, western Guizhou Province, which belong to the area with very high research hotness of rocky desertification. The rocky desertification phenomenon is also serious with high research hotness in Guiyang in central Guizhou Province, Southwest Guizhou Autonomous Prefecture in southwestern Guizhou Province and in parts of northwestern Guangxi (e.g., Fengshan County, Pingguo County and Mashan County). In Liupanshui, Anshun of western Guizhou, Buyi and Miao Autonomous Prefecture of QianNan of southern Guizhou, Hechi, Baise, Nanning and Guilin in western and northern Guangxi and Wenshan Zhuang and Miao Autonomous Prefecture in southeastern Yunnan, the research hotness of rocky desertification is lower than those of the previous two types and these regions belong to moderate research hot regions of rocky desertification.

3.2. Sensitive Region Mapping

According to the spatial distribution of the sensitivity to rocky desertification (1-km grid), the sensitivity index of rocky desertification can be obtained for the county unit, as shown in Figure 5. The sensitive areas of rocky desertification in China are mainly distributed in Guizhou Province, Guangxi Province and Yunnan Province. There are also a small number of rocky desertification-sensitive areas in Sichuan, Chongqing, Hubei, Hunan, Jiangxi and Guangdong. Among them, areas with moderate or greater sensitivity to rocky are mainly distributed in central and western Guizhou, eastern Yunnan, northwest and western Guangxi, northern Guangxi, western Sichuan and other regions. At the same time, there are a small number of distributions in the border areas between Hubei and Chongqing and between Hunan and Yunnan. Areas with moderate or greater sensitivity to rocky desertification included 117 counties or cities, with a total area of 273.2 thousand km^2.
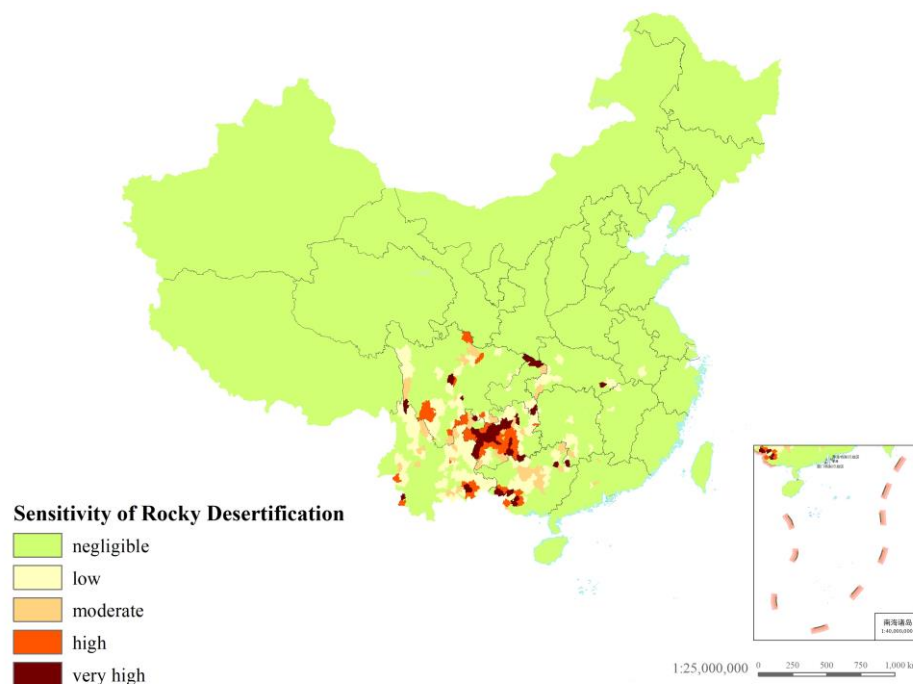

Figure 5. The County-based Rocky Desertification Sensitivity Index in China.

Specifically, the areas with very high sensitivity were mainly distributed in some counties in Bijie, Liupanshui and Guiyang of Guizhou Province and some counties of Baise, Chongzuo and Guilin in Guangxi Zhuang Autonomous Region. High sensitive areas were mainly distributed on the periphery of very high sensitive areas and were often interwoven with them. They were mainly distributed in Bijie, Liupanshan, Anshun, Southwest Guizhou Autonomous Prefecture, Buyi and Miao Autonomous Prefecture of QianNan, some counties of Qiandongnan Miao and Dong Autonomous Prefecture in

Guizhou Province, as well as some counties in Hechi, Chongzuo and Guilin of Guangxi Province. The moderately sensitive areas were distributed in Zhaotong, Kunming and Hani-Yi Autonomous Prefecture of Honghe in eastern Yunnan, some counties of central and western Guizhou (Gan Liang, Shi Bing and Tantan) and some counties and cities in northwestern Guangxi Province.

3.3. Comparison of Research Hot Regions and Sensitive Regions

Comparing the spatial distribution map of the rocky desertification research hot regions (Figure 4) and the spatial distribution map of the rocky desertification sensitive regions (Figure 5), we found that on the national scale, two maps have high overlap in the core area of rocky desertification; however, there are differences in the outer space distribution map based on the intensity classification.

First, two maps indicate that the central and western regions of Guizhou, eastern Yunnan and northwestern Guangxi are sensitive areas and research hot regions on rocky desertification. In the case of regions that have moderate or greater sensitivity and moderate or greater levels of research hotness, there are 52 county units or 11 prefecture units that have collectively been identified. The county unit coincidence rate was 48% and the prefecture unit coincidence rate was 85%. The high degree of coincidence among the geographic units indicates that the rocky desertification sensitivity model roughly integrates the driving factors of rocky desertification and can better reflect the actual rocky desertification intensity level; on the other hand, it also shows that in the past 60 years, the research on rocky desertification has had clear targets, an outstanding focus and a good level of geographical matching. In addition, the research area is focused on the areas that are sensitive or prone to rocky desertification and on degraded areas.

Second, in general, the distribution of areas that are sensitive to rocky desertification is larger than the areas of research hot regions on rocky desertification. Specifically, in Qionglai Mountain—Liangshan Area of Sichuan, Wushan-Shennongjia Area of Hubei, Hengduan Mountain Area of Western Yunnan and Dupangling Area of Southern Hunan, the sensitivity to rocky desertification is obviously high but research popularity is not high enough. To analyze the "decoupling of space" between sensitivity and research popularity in the above four areas, further statistics on the differences between the typical rocky desertification areas and the above 4 "space decoupling" areas in terms of topography, vegetation coverage, population distribution, traffic accessibility were analyzed (Figure 6 and Table 1). Hotness Index, Sensitivity Index, Environmental and Social Backgrounds Index in Key Regions.

Comparative analysis shows that compared with typical rocky desertification areas (e.g., Bijie, Liupanshan and Guiyang in the central and western Guizhou), the Qionglai Mountain—Liangshan Area in Sichuan is located in the eastern margin of the Hengduan Mountains and has large fluctuations in its types of landforms. The average slope of the terrain in the region is much higher than that in the typical rocky desertification regions, which means that people that people are relatively difficult to active here. At the same time, the population density in the region is much lower than that in the typical rocky desertification areas. This result means that rocky desertification has a minor impact on people's lives in this region; additionally, the road network density is lower and the vegetation coverage is higher in this region, which means that it is relatively difficult for mankind to be active, too. Therefore, given the higher terrain slope, higher vegetation coverage, lower road network density and low social impact in this area, the study of rocky desertification in the Qionglai Mountain—Liangshan Area is naturally low.

Similar to the Qionglai Mountain—Liangshan Area in Sichuan, the Wushan-Shennongjia Area in Hubei and the Hengduan Mountain Area in western Yunnan also have a higher terrain slope, higher vegetation coverage, lower population density and lower road network density and these factors lead to a significantly lower study of rocky desertification in the above two regions. For the Dupangling Area in southern Hunan, although the terrain conditions and accessibility are slightly better than those in the typical rocky desertification areas, due to the higher vegetation coverage and lower population density in this area, the amount of research in this area is not high.

Table 1. Hotness Index, Sensitivity Index, Environmental and Social Backgrounds Index in Key Regions.

Mode	Area	Research Hoeness Index	Sensitivity Index	Slope (°)	Vegetation Coverage (%)	Population Density (People/km^2)	Road Network Density (km/km^2)
Highly Matched Area (High Sensitivity High Research Hotness)	Typical Rocky Desertification Area	4.7	5.04	3.13	0.82	573.21	0.56
Space Decoupling Area (High Sensitivity Low Research Hotness)	Qionglai Mountain—Liangshan Area, Sichuan	1	4.11	10.91	0.86	139.66	0.44
	Wushan-Shennongjia Area, Hubei	1	4.17	6.7	0.93	144.82	0.50
	Hengduan Mountain Area, Western Yunnan	1	3.90	8.66	0.84	94.76	0.46
	Dupangling Area, Southern Hunan	1	3.30	2.92	0.85	435.46	0.62

[1] The major counties of all regions: (1) typical rocky desertification area includes Bijie, Liupanshui and Guiyang in central and western Guizhou; (2) Qionglai mountain-liangshan area includes Deong, Batang, Baiyu and other counties in Sichuan; (3) Wushan-shennongjia area includes Yi'an and other counties in Hubei; (4) Hengduan mountain area includes Zhenkang, Shidian, Ximeng and other counties in western Yunnan; (5) Dupangling area includes Yonju and other counties in southern Hunan.
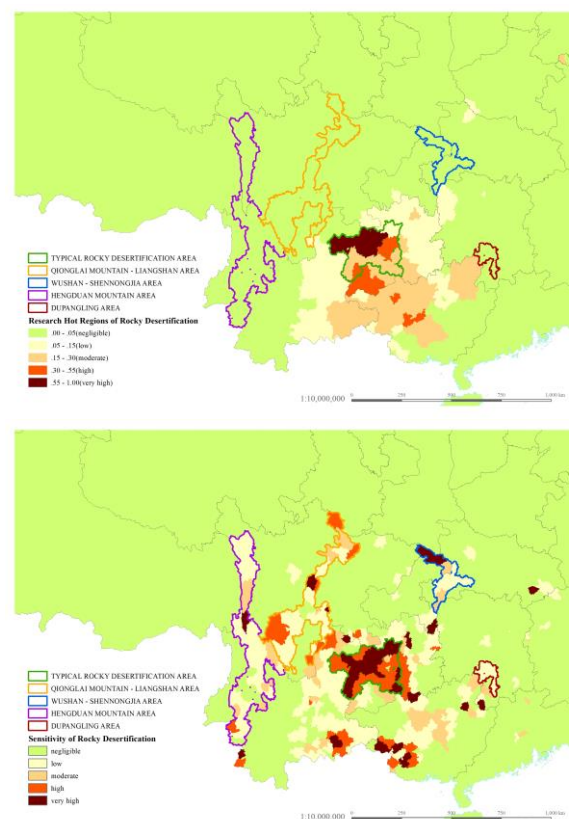
Figure 6. Comparison of Research Hot Regions and Sensitive Regions.

4. Discussion

4.1. The Distinct Characteristics of Research Hot Regions and Sensitive Regions

Rocky desertification is a natural geographic process. The sensitivity of rocky desertification reflects the proneness of rocky desertification processes under realistic natural and geographical conditions. The research hots regions on rocky desertification reflects the regions used for the study of rocky desertification that have been selected by people under realistic natural geographical conditions and based on economic and social development needs. Strong rocky desertification processes are mostly found in areas with high sensitivity to rocky desertification and in easily distorted ecosystems. Such areas are usually regions with a high level of research due to the attention of the government and the researchers and vice versa. Therefore, it is reasonable to deduce the spatial distribution rules of the rocky desertification process by analyzing the rock desertification research hot regions.

However, in essence, there is still a clear difference in the conceptual definitions of the rocky desertification hot regions and the rocky desertification area. Therefore, there are naturally incomplete spatial and temporal matching features. First, the process of rocky desertification in nature is a long-term and objective process. However, rocky desertification research is constrained by the level of economic and social development, discipline construction, terminology demarcation, research and project establishment. It is a subjective exploration of the rocky desertification process in human society at a certain stage. The inconsistent conceptual connotations of the rocky desertification areas and the rocky desertification research hot regions will undoubtedly cause the two regions to be incompletely matched in time and space. Second, the rocky desertification process in nature is based on geographic ecological units and is distributed in patches. However, rocky desertification researches usually describe the rocky desertification process based on administrative divisions. Therefore, the inconsistency of the object's target on the spatial scale and the types of spatial features also cause deviations in the spatial orientation, spatial reasoning and spatial statistics. Third, rocky desertification

studies can reflect the strength and weakness of the rocky desertification process and the size of the area. However, the importance of the regional location, regional economic and social development goals and planning and other factors may have a more important impact on the formation of research hot regions.

4.2. The Potential Applications of Hot Regions Analysis

The difference in concept, time and space between rocky desertification and rocky desertification research hot regions is of great significance for the comparative analysis of the relationship between real rocky desertification and human society's response to rocky desertification; additionally, it can effectively support the development of rocky desertification planning, governance and research work. In particular, by comparing the spatial distribution differences between regional rocky desertification spatial distribution maps based on satellite remote sensing and ground surveys and rocky desertification research hot regions maps based on web retrieval and text mining, it can be determined which areas have rocky desertification and have received attention from the government, researchers and engineering management, or which areas also have rocky desertification but have not received enough attention from the government, researchers and engineering management. By comparing the spatial disparity and dynamic matching of the above regions, it is possible to examine the prevention and control planning, engineering management, technical research and evaluate the effectiveness of various ecological degradations, including rocky desertification attention from the government and the research community. The role of remnants and shortages and targeted promotion will greatly improve the temporal and spatial pertinence of the rocky desertification control plan, project management and scientific research.

As shown in Table 1, the Guizhou Plateau, which is a typical rocky desertification area, has been fully studied by researchers. However, in the Qionglai Mountain—Liangshan Area of Sichuan, the Wushan-Shennongjia Area of Hubei and the Hengduan Mountain Area of western Yunnan, the sensitivity to rocky desertification is high or very high; however, the amount of research conducted here is very low. Therefore, more researchers can choose these regions as a new study area to detect occurrence regularity and the evolutionary process of rocky desertification.

5. Conclusions

In view of the spatial distribution of rocky desertification, this paper studies the information extraction technology of rocky desertification research hot regions based on literature text mining and then analyzes the spatial distribution of research hot regions on rocky desertification in China and the applicable scenarios of the analysis methods of research hot regions. The main conclusions are as follows:

1. Based on the CNKI knowledge database, using the web crawler, research information can be obtained quickly and easily in a short time; then, through text parsing, toponym identification and matching, hotness metering and other technologies and algorithms, the research hot regions on rocky desertification in China can be identified effectively.
2. The research hot regions of rocky desertification in China are mainly distributed in Guizhou, Yunnan and Guangxi Province. There are 108 counties with 245.7 thousand km^2 that belong to moderate, high and very high level of research hotness, especially in Bijie, Liupanshui, Guiyang, Anshun, Southwest Guizhou Autonomous Prefecture, Buyi and Miao Autonomous Prefecture of QianNan, Qiandongnan Autonomous Prefecture, Qiandongnan Miao and Dong Autonomous Prefecture in Guizhou Province, Hechi, Baise, Nanning, Guilin in Guangxi Province and Zhaotong in Yunnan Province.
3. The research hot regions on rocky desertification have good spatial consistency with the sensitive areas of rocky desertification. The two spatial distributions show that the areas of central Guizhou, eastern Yunnan and northern Guangxi are rocky desertification-prone areas and research hot

regions. The county unit coincidence rate of the two spatial distributions is 48% and the prefecture unit coincidence rate is 85%. A combination of high-sensitivity and low-intensity research on rocky desertification has emerged in the four regions—Qionglai Mountain-Liangshan Area of Sichuan, Wushan-Shennongjia Area of Hubei, Hengduan Mountain Area of western Yunnan and Dupangling Area of southern Hunan—because of topography, vegetation coverage, degree of accessibility, human influence and other factors.

4. Strong rocky desertification processes are mostly found in areas with high sensitivity to rocky desertification and easily disproportionate ecosystems and such areas usually have higher research frequency and vice versa. Although there is a close connection between the rocky desertification area and the rocky desertification research hot regions, there were similar characteristics in terms of incomplete matching concepts, space and time. The rational application of the differences in the spatial distribution of the two can improve the spatial and temporal pertinence of rocky desertification prevention and control planning, engineering governance and scientific research.

Author Contributions: Y.H. (Yunfeng Hu) designed the idea and methodology; Y.H. (Yueqi Han) and Y.Z. implemented the idea, performed the experiments and preliminarily analyzed the results; and all authors drafted and revised the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lan, A.; Xiong, K.; An, Y. Analysis on driving factors of karst rock-desertification—With a special reference to guizhou province. *Bull. Soil Water Conserv.* **2001**, *6*, 19–23.
2. Wang, S.; Liu, Q.; Zhang, D. Karst rocky desertification in southwestern china: Geomorphology, landuse, impact and rehabilitation. *Land Degrad. Dev.* **2004**, *15*, 115–121. [CrossRef]
3. Su, W.; Zhou, J. Rocky desertification in Guizhou karst region and its preventive strategy. *Resour. Environ. Yangtza Val.* **1995**, *4*, 177–182.
4. Su, W. Controlling model for rocky desertification of karst mountainous region and its preventing strategy in southwest, China. *J. Soil Water Conserv.* **2002**, *16*, 29–32.
5. Sen, L.; Dong, Y.; Wang, J. Re-discussion on the concept and classification of rocky desertification. *Carsologica Sin.* **2007**, *26*, 279–284.
6. Su, W.; Xiao, Z.; Ning, X. Stone desertification and eco-economics improving model in Guizhou karst mountain. *Carsologica Sin.* **2002**, *21*, 19–24.
7. Zhou, Z. Application of remote sensing and GIS technology for land desertification in Guizhou karst region. *Bull. Soil Water Conserv.* **2001**, *66*, 52–54.
8. Wang, S. Concept deduction and its connotation of karst rocky desertification. *Carsologica Sin.* **2002**, *21*, 101–105.
9. Pan, Y.; Shi, P.; Zhu, W.; Gu, X.; Fan, Y.; Li, J. Measurement of ecological capital of chinese terrestrial ecosystem based on remote sensing. *Sci. China* **2005**, *48*, 786–796. [CrossRef]
10. Wang, S. Karst rocky desertification:Formation background, evolution and comprehensive taming. *Quat. Sci.* **2003**, *23*, 657–666.
11. Qin, X.; Zhu, M.; Jiang, Z. A review on recent advances in rocky desertification in Southwest China Karst Region. *Carsologica Sin.* **2006**, *25*, 234–238.
12. Dai, Q.; Peng, X.; Yang, Z.; Zhao, L. Runoff and erosion processes on bare slopes in the karst rocky desertification area. *Catena* **2017**, *152*, 218–226. [CrossRef]
13. Galaz, V.; Crona, B.; Daw, T.; Bodin, Ö.; Nyström, M.; Olsson, P. Can web crawlers revolutionize ecological monitoring? *Front. Ecol. Environ.* **2010**, *8*, 99–104. [CrossRef]

14. Hu, Y.; Han, Y.; Zhang, Y.; Zhuang, Y. Extraction and dynamic spatial-temporal changes of grassland deterioration research hot regions in china. *J. Resour. Ecol.* **2017**, *8*, 352–358.

15. Li, Y.; Feng, L.; Liu, X.; Cheng, S.; Zhang, X. A method of context enhanced keyword extraction for sparse geo-entity relation. *J. Geo-Inf. Sci.* **2016**, *18*, 1465–1475.

16. Sergio, M.; Agnese, M.; Dario, S.; Tullio, F. Opportunities for web-based indicators in environmental sciences. *PLoS ONE* **2012**, *7*, e42128.

17. Tang, X.; Chen, X.; Zhang, X. Research on toponym resolution in Chinese text. *Geomat. Inf. Sci. Wuhan Univ.* **2010**, *35*, 930–935, 982.

18. Sadilek, A.; Kautz, H.; Silenzio, V. Predicting Disease Transmission from Geo-Tagged Micro-Blog Data. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012; pp. 136–142.

19. Dredze, M.; Paul, M.J.; Bergsma, S.; Tran, H. Carmen: A Twitter Geolocation System with Applications to Public Health. In Proceedings of the AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI), Bellevue, WA, USA, July 2013; pp. 20–24.

20. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In Proceedings of the International World Wide Web Conference, Raleigh, NC, USA, 26–30 April 2010; pp. 851–860.

21. Cameron, A.M.; Power, R.; Robinson, B.; Yin, J. Emergency Situation Awareness from Twitter for Crisis Management. In Proceedings of the 21st International Conference on World Wide Web—WWW 7812 Companion, Lyon, France, 16–20 April 2012; pp. 695–698.

22. Krallinger, M.; Valencia, A.; Hirschman, L. Linking genes to literature: Text mining, information extraction, and retrieval applications for biology. *Genome Boil.* **2008**, *9* (Suppl. 2), S8. [CrossRef] [PubMed]

23. Jenssen, T.K.; Lægreid, A.; Komorowski, J.; Hovig, E. A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.* **2001**, *28*, 21–28. [CrossRef] [PubMed]

24. Ananiadou, S.; Kell, D.B.; Tsujii, J.I. Text mining and its potential applications in systems biology. *Trends Biotechnol.* **2006**, *24*, 571–579. [CrossRef] [PubMed]

25. Becker, K.G.; Hosack, D.A.; Glynn Dennis, J.; Lempicki, R.A.; Bright, T.J.; Cheadle, C.; Engel, J. Pubmatrix: A tool for multiplex literature mining. *BMC Bioinform.* **2003**, *4*, 61. [CrossRef] [PubMed]

26. Chakrabarti, S.; van den Berg, M.; Dom, B. Focused crawling: A new approach to topic-specific web resource discovery. In Proceedings of the WWW8 Conference, Toronto, ON, Canada, May 1999; pp. 1623–1640.

27. Thelwall, M. A web crawler design for data mining. *J. Inf. Sci.* **2001**, *27*, 319–325. [CrossRef]

28. Sproat, R.; Gale, W.; Shih, C.; Chang, N. A stochastic finite-state word-segmentation algorithm for Chinese. *Comput. Linguist.* **1996**, *22*, 377–404.

29. Levow, G.A. The Third International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition. In Proceedings of the Sighan Workshop on Chinese Language Processing, Sydney, Australia, 22–23 July 2006; pp. 108–117.

30. Wu, Y.; Zhao, J.; Xu, B. Chinese named entity recognition combining a statistical model with human knowledge. In Proceedings of the ACL 2003 Workshop on Multilingual and Mixed-Language Named Entity Recognition, Sapporo, Japan, 20–24 May 2003; pp. 65–72.

31. Doerr, M.; Papagelis, M. A method for estimating the precision of placename matching. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1089–1101. [CrossRef]

32. Liao, Y.; Wang, J.; Ma, J.; Qi, X. Placename data matching based on bpm-bm algorithm. *Bull. Surv. Mapp.* **2008**, *6*, 22–25.

33. Li, N. Chinese place name identification with Chinese characters features. *Comput. Eng. Appl.* **2009**, *45*, 230–232.

34. Ou, Y.E.; Zhang, J.; Li, J. Duplicate field matching for data cleaning of Chinese placenames. *J. Appl. Sci.* **2013**, *31*, 212–220.

35. Wang, X.; Hu, E.; Li, A.; Li, Q. Dynamic changes of vegetation coverage in daqingshan nature reserve based on modis ndvi image. *J. Arid. Land Resour. Environ.* **2014**, *28*, 61–65.

36. Jing, G.; Meng, B.; Yang, S.; Gao, J.; Feng, Q.; Liang, T.; Huang, X.; Gao, X.; Li, W.; Zhang, R. Dynamic monitoring of alpine grassland coverage based on uav technology and modis remote sensing data—A case study in the headwaters of the yellow river. *Acta Prataculturae Sin.* **2017**, *26*, 1–12.

37. Hu, Y.; Wang, Q.; Liu, Y.; Li, J.; Ren, W. Index system and transferring methods to build the national society and economy grid database. *J. Geo-Inf. Sci.* **2011**, *13*, 573–578. [CrossRef]

38. Hanlp: Han Language Processing. Available online: http://hanlp.linrunsoft.com/ (accessed on 1 October 2016).

39. Ansjdemo. Available online: https://github.com/NLPchina/ansj_seg (accessed on 1 October 2016).

40. Lu, H. An improved high-effective kmp pattern matching algorithm. *J. Huazhong Univ. Sci. Technol.* **2006**, *10*, 012.

41. Xiao, R.; Ouyang, Z.; Wang, X.; Zhao, T. Sensitivity of rocky desertification and its spatial distribution in southwestern China. *Chin. J. Ecol.* **2005**, *24*, 551–554.

42. Qin, X.; Jiang, Z. An assessment of eco-environmental vulnerability in karst counties of Guangxi. *Geol. Geochem.* **2005**, *33*, 45–51.

43. Liu, J.; Gao, J.; Ma, S.; Wang, W.; Zou, C. Evaluation of ecological sensitivity in China. *J. Nat. Resour.* **2015**, *10*, 1607–1616.

44. Shen, M.; Shen, L.; Zhong, S.; Zhang, C.; Kong, H. The spatial optimization of mineral industrial capacity reduction planning of resource-based cities in china based on ecological sensitive areas. *Resour. Sci.* **2016**, *38*, 1962–1974.