International Journal of Geo-Information

MDPI

Article

Assessing the Intensity of the Population Affected by a Complex Natural Disaster Using Social Media Data

Changxiu Cheng [1,2,3,4], Ting Zhang [1,2,3,4], Kai Su [1,2,3,4], Peichao Gao [1,2,3,4] and Shi Shen [1,2,3,4,*]

1 Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing 100875, China
2 State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
3 Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
4 Center for Geodata and Analysis, Beijing Normal University, Beijing 100875, China
* Correspondence: shens@mail.bnu.edu.cn

Received: 28 June 2019; Accepted: 11 August 2019; Published: 13 August 2019



Abstract: Complex natural disasters often cause people to suffer hardships, and they can cause a large number of casualties. A population that has been affected by a natural disaster is at high risk and desperately in need of help. Even with the timely assessment and knowledge of the degree that natural disasters affect populations, challenges arise during emergency response in the aftermath of a natural disaster. This paper proposes an approach to assessing the near-real-time intensity of the affected population using social media data. Because of its fatal impact on the Philippines, Typhoon Haiyan was selected as a case study. The results show that the normalized affected population index (NAPI) has a significant ability to indicate the affected population intensity. With the geographic information of disasters, more accurate and relevant disaster relief information can be extracted from social media data. The method proposed in this paper will benefit disaster relief operations and decision-making, which can be executed in a timely manner.

Keywords: social media; natural disasters; emergency response; affected people intensity

1. Introduction

Natural disasters are some of the most complex phenomena in the world [1]. The impacts of a natural disaster on society, the community, and the population are also complex [2,3]. Assessing the complex effect on a population is a critical part of the emergency response and relief efforts. Because of their significant impact on infrastructure, natural disasters often lead to a lack of clean water, power shortages, the demolition of numerous houses, and even a number of casualties [4]. People who live in post-disaster areas are vulnerable and need appropriate relief and support from emergency responders (e.g., governmental agencies and non-government organizations). Hence, although the intensity of the population affected by disasters (i.e., the severity of the disaster's impact on the population) is a general concept and without a rigorous definition, it plays a crucial role in implementing proportionate and timely rescue and relief operations.

However, traditional approaches to assessing the affected population intensity have several shortcomings related to timeliness and accuracy. For example, census data, which are fundamental information in traditional methods, are relatively static rather than timely because the data are obtained by surveying the population at regular intervals (usually years) [5]. Therefore, it is unlikely to represent the population at the time of a disaster. Moreover, when disaster areas include resorts and tourist destinations, a large number of non-native people may be affected [6], and census data may not reflect this segment of the population. Moreover, reliance on remote-sensing-based methods causes delays

that result from the time-consuming processing of remote sensing products [7]. In contrast, social media provide timely and active communication and near-real-time firsthand insight during natural disasters [8].

In addition, social media data have proved to be capable of revealing the disaster-affected population [9]. First, data derived from social media have been shown to have strong relationships with the extent of damage from natural disasters [10–13]. Second, combined with other disaster-related data, such as geographical information and remote sensing images, social media data can enhance and improve useful information extracted for emergency response [14–17]. Last, but important, smart mobile devices allow users to report information about disasters (e.g., locations with specific names), which is likely to be useful for estimating disaster-related information, including the affected area, damaged infrastructure, affected people, and evacuation zones [18–21].

Aiming to improve the situation awareness in emergency response and relief operations, we present an approach that uses social media data to assess the affected population intensity in post-disaster areas. On the basis of our previous research [22,23], the proposed approach leverages the spatial extent and temporal range of a disaster event to extract and filter actual disaster-related social media data. Then, a new proxy index is introduced that indicates the affected population intensity. Afterward, Typhoon Haiyan, which occurred in 2013, is used as a case study to verify the proposed approach and compare its efficiency with the result of the traditional method and the activity-based index.

The rest of this paper is organized as follows. The next section provides a general review of the research on the use of social media data to assess the damage caused by disasters. Section 3 demonstrates the methodology for collecting and preparing social media data and leveraging the information to assess the affected population intensity. Section 4 presents the result of Typhoon Haiyan as a case study. This paper also discusses issues and concludes with challenges and future work on leveraging social media data to assess the intensity of populations affected by natural disasters.

2. Related Work

2.1. Traditional Approaches to Assessing Affected Population

Estimation of the disaster-affected population is a crucial issue in the disaster research literature. Emergency response with the traditional approach depends primarily on historical data, observational data, or both. Jaiswal and Wald [24] developed an empirical model to estimate earthquake fatalities in a country or region using the global mortality rates of earthquakes. Aghamohammadi et al. [25] proposed a back-propagation neural network for modeling and estimating the severity and distribution of human loss in earthquakes, and their work was supported by building and population survey data. Bhatt et al. [26] used multitemporal satellite images, hydrological observations, and gridded population data to estimate the population affected by a flood. Raza et al. [27] superimposed the flood extent and gridded population data to count the number of people in the affected population in the flooded area. Ozcelik et al. [28] proposed a GIS-based methodology to model the population affected by storm surges. The data in their approach included the gridded population of the world and the surge profiles obtained by linear regression.

Regardless of whether an empirical approach or an event-based method is used to assess the affected population, there is a heavy dependence on the underlying population data. Therefore, the timeliness and accuracy of their estimations are primarily determined by the population data that they use. Unfortunately, these data are relatively static and time-delayed. As a result, especially for areas with large population movements and rapid changes, traditional approaches are unlikely to rapidly assess the affected population. However, social media data can provide timely information about the disaster situation and population to compensate for the pitfalls of census data, which is leveraged in our method.

2.2. Social Media-Based Approach for Disaster Situational Awareness

With the development of location-enabled mobile devices and the prosperity of social networks, social media provide an additional vital data source for emergency responses [15,29]. Much of the existing natural disaster research on using social media data have focused on various aspects, such as damage assessment [11,13,14,30] and event detection [31–35]. However, the application of social media data to improve the situation awareness of emergency responders is still in an early stage, and more efforts on this topic are needed [36]. De Albuquerque et al. [15] introduced an approach that leverages authoritative data to identify disaster-related tweets (firsthand short text messages posted on Twitter). Using the River Elbe flood in 2013 as a case study, they found that tweets near the severely flooded area were more associated with firsthand observations, official actions, and infrastructure damages. However, their research did not apply the content of the tweets. Resch et al. [7] used the Latent Dirichlet Allocation (LDA) model to extract semantic information and map locations in which the population was significantly affected. Shan et al. [5] developed a quantitative framework for disaster damage assessment, and the LDA model was used to extract disaster-related topics for evaluating the affected population. However, the specific meaning of the topic extracted by the LDA model is still not very clear and relies extensively on the background knowledge of the researchers. Fang et al. [37] developed a framework to assess a disaster's impact through keyword frequency analysis of impact-related topics by using social media data from the Weibo platform. Their work showed that the keyword-based method is able to evaluate the impact of the disasters, but the method does little with the geolocation information contained in these messages. Additional and detailed research on applying social media to disaster situation awareness can be found in the review by Imran et al. [8].

Three major issues exist in the discussed research on disaster assessment using social media data. (1) Most of the studies depend on tweets with accurate geolocations (i.e., latitudes and longitudes), which are unable to overcome the donut effect caused by a lack of power supply, Internet connections, and computing abilities in disaster areas [4]. (2) It is difficult for machine learning-based topic models (e.g., LDA) to assign semantic meaning to an extracted topic, leading to uncertainty in the interpretation of topics using the proposed LDA-based methods [7,38]. (3) The geographical information contained in tweets has not always been considered or utilized [7,16,39].

3. A Social Media-Based Approach to Assess the Disaster-Affected Population

Figure 1 depicts the social media-based assessment approach. First, this approach employs the affected ranges of disasters and a classifier (e.g., support vector machine (SVM)) to extract disaster-related hashtags from the 1% of Twitter that is in the Internet Archive database. The detailed procedure of the five steps is introduced in Section 2.1. Afterward, the disaster-related hashtags resulting from the first step are used in the application of a crawler program to download disaster-related tweets from Twitter. Then, according to the geolocation of disaster-related tweets and the locations of their users' accounts, a geodecoding service (e.g., Google Geocoder API (Application Programming Interface)) is leveraged to find the location from which the tweets originate, and spatial analysis is used to remove tweets that are outside the disaster areas. Finally, a self-constructing gazetteer of the disaster area is used to retrieve the relevant geographical entities from the tweets and construct the normalized affected population index (NAPI) to estimate the affected population intensity.

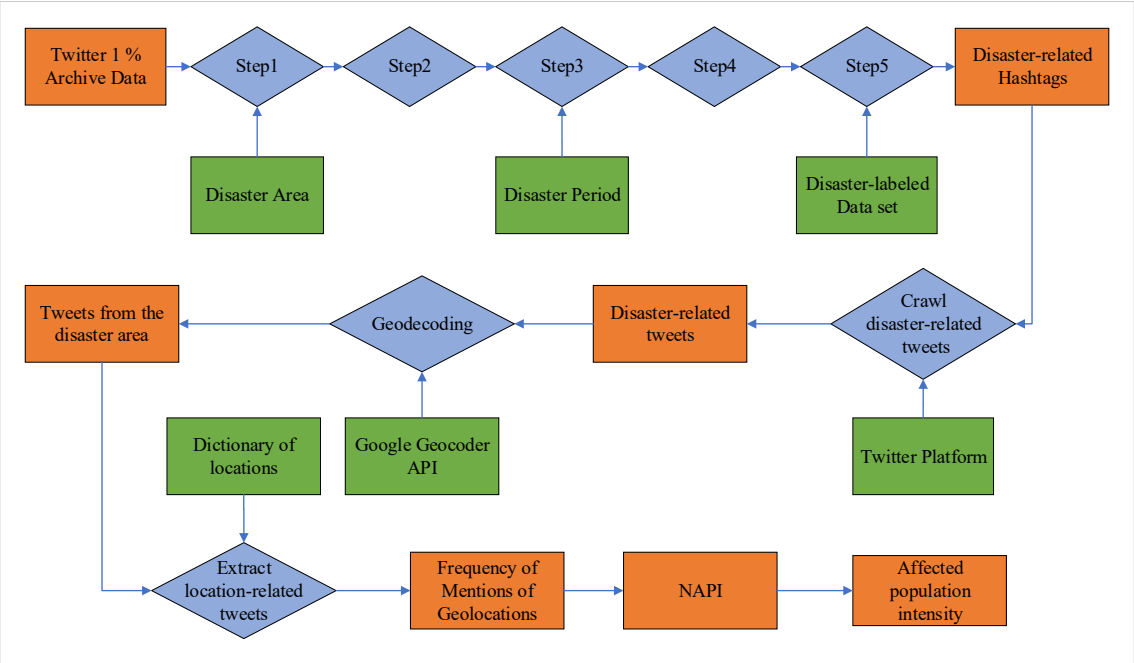


Figure 1. Flowchart of the assessment of the affected population intensity using social media.

3.1. Tweet Collection and Preprocessing

Numerous methods have been proposed for extracting disaster-related tweets. In this study, we followed the method presented by Murzintcev and Cheng [22] and Shen et al. [23]. Before we acquired tweets related to disasters, the first and critical step was to retrieve disaster-related hashtags from social media. This was achieved in five steps:

- Step 1: The tweets and hashtags located in the disaster area were extracted from the Internet Archive database, which contains 1% of Twitter;
- Step 2: The hashtags that appeared only once were filtered out;
- Step 3: The hashtags that appeared before the disaster warning were filtered out;
- Step 4: Labeled disaster datasets (e.g., CrisisLexT6 [40], CrisisLexT26 [41], AIDR2015Q2 [42]) were used to build an SVM classification model to filter out the hashtags with less than a 0.5 probability of being related to disasters;
- Step 5: Expert knowledge can also be used to further improve the accuracy of the obtained disaster-related hashtags by removing irrelevant ones. However, this step is not always necessary.

After acquiring the disaster-related hashtags, tweets associated with a specific disaster event were searched from the Twitter platform using these hashtags and grabbed by a crawler program. Although the used hashtags were obtained from the 1% of Twitter in the Internet Archive database, the Twitter platform was directly searched for tweets about disasters. In this manner, the method is capable of gaining more accurate disaster-related information than keyword-based methods.

It is worth noting that the obtained tweets could have been posted from anywhere in the world. Many tweets are not geotagged or geolocated. Here, to shrink the geographic range of the tweet sources to the disaster area, we used the location that the users linked with their accounts during registration. First, we used the Google geocoder API to decode the registered locations of the users who sent disaster-related tweets. Other methods, such as machine learning algorithms [43], can also be employed to link the tweets to geographic information. After locating the tweets, we used spatial analysis with GIS software to remove the tweets outside the disaster areas. Consequently, the relevance and consistency of the content of the tweet were improved once again.

3.2. The Proxy Index of the Affected Population Intensity

Various proxy indices that indicate and evaluate disaster-induced damages have been proposed in previous studies [5,11]. Most of them can be grouped into two types: activity-based and sentiment-based indices. In general, disaster research regarding activity or sentiment indices has implied the following hypothesis: The behavior and emotions of people on social media are affected by natural disasters. Therefore, the activities and sentiment of people's posts and expressions can indicate the damage and impact of natural disasters. A typical activity index is the number of posted tweets per unit period and area, which has been proved to have a significant relationship with the per-capita economic damage caused by a hurricane [11]. In this paper, we introduce a new proxy index of the affected population intensity using the geolocation information contained in tweets. As shown in Figure 1, after extracting the tweets located within the disaster area, we built and employed a gazetteer that contains the administration zones and other geo-entity names (or both) to calculate the frequency of the places mentioned in tweets. Because Twitter limits the number of characters in a tweet (to a maximum of 140 characters at the time of this research), a large number of place name deformations, abbreviations, and aliases are present in tweets. Therefore, the gazetteer includes the administration zones and their corresponding formal names, abbreviations, aliases, and deformations to improve the representativeness. Hence, the normalized affected population index (NAPI)—the proxy index of a disaster used in this approach—can be described as follows. We denote a place series by $p(i), i = 1, 2, \ldots, n$, where $p(i)$ represents one of n areas that is affected by a disaster. The series $q(i), i = 1, 2, \ldots, n$ accordingly indicates the frequency of the $p(i)$ mentioned, that is, how many times $p(i)$ was retrieved in the tweets from the disaster area using the dictionary of locations. Hence, the NAPI for a place $p(l)$ is calculated by

$$NAPI(p(l)) = \frac{q(l)}{\max\limits_{1 \le i \le n} (q(i))} \quad (1)$$

For example, 44 tweets mention Antique Province, and Cebu Province is the most frequently mentioned by tweets (756 times). Hence, the NAPI of Antique Province is 0.582 (44/756).

4. Case Study and Dataset

4.1. Typhoon Haiyan in 2013

From 6 November to 9 November 2013, Typhoon Haiyan, which is also named Yolanda, hit the Philippines and became one of the strongest and most destructive tropical cyclones in Philippine history. Before its first landing on Guiuan, Eastern Samar, the wind speed near the center of Typhoon Haiyan was 215 kph. According to the final report by Philippines' National Disaster Risk Reduction and Management Council (NDRRMC), more than 16 million persons were affected in 44 provinces, and the total number of deaths, injuries, and missing persons were 6300, 28,688, and 1062, respectively [44].

Figure 2 shows the distribution of the affected population intensity, which is the number of people affected in each region divided by the largest number of people affected, according to the NDRRMC final report about Typhoon Haiyan. The blue dashed lines indicate the impact range of the cyclone. The color represents the affected population intensity: the redder the color, the greater the intensity. It is clear that the administration zones located within the range of the cyclone had a high affected population intensity. According to the affected population intensity, the natural breaks method was used to classify the affected areas into five degrees of intensity: very severe, severe, moderate, low, and very low. As a result, there were 13 administration zones, labeled 1–13, with a higher than moderate severity of the affected population intensity, and these areas were ranked in descending order by the actual intensity of the affected population. Among them, zones 1–4 have the largest affected population intensities.

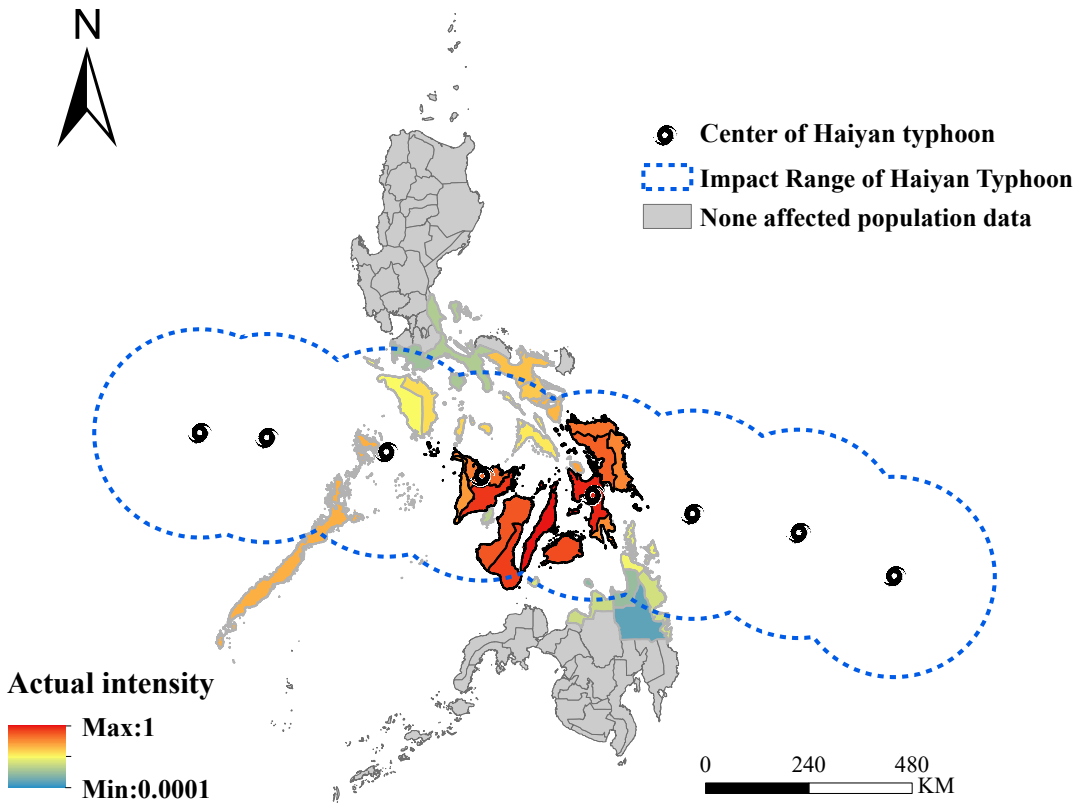


Figure 2. The actual affected population intensity by Typhoon Haiyan.

4.2. Datasets

In this paper, we only focus on the Twitter platform because our search results using relevant hashtags and keywords (e.g., Haiyan, Youlanda, Typhoon Haiyan, Typhoon Yolanda) revealed that it had much more information about Typhoon Haiyan than another platform. Using the method proposed by Murzintcev and Cheng [22], we obtained 21 hashtags (Figure 3) associated with Typhoon Haiyan from the Twitter Internet Archive database. Then, a web crawler program was employed to grab the original tweets (excluding reposted tweets) and the authors of these tweets containing these hashtags. The resultant 411,738 tweets from 2013 to 2016 contained the following information: the identification number of a tweet, the time at which a tweet was posted, the number of times a tweet was liked, the number of reposts of the tweet, the number of replies to the tweet, the identification number of the tweet poster, the location registered by the user, and the textual content of a tweet. Finally, after filtering and geodecoding, 32,048 tweets associated to Typhoon Haiyan from 6 November to 10 November 2013 in the Philippines were extracted and used in the following analysis and NAPI calculation. In 32,048 tweets, only 3023 mentions disaster areas.Table 1 list the number of tweets in each procedure of the proposed method.

In addition to the tweet data, the latest gridded population data [45] of the Philippines (2010) was used to evaluate the affected population by the traditional assessment method. The tracks and impact range of the typhoon were identified with the IBTrACS database [46].

#bangonvisayas, #haiyan, #ormoc, #prayforphilippines, #prayfortacloban, #prayfortheph, #prayforthephilippines, #reliefph, #rescueph, #safenow, #strongerph, #supertyphoon, #surigaodelnorte, #tacloban, #tracingph, #typhoonhaiyan, #typhoonyolanda, #ugc, #yolanda, #yolandaph, #yolandaupdates

Figure 3. Hashtags related to Typhoon Haiyan.

Table 1. The number (N) of tweets in each procedure.

N of Tweets Related to Typhoon Haiyan	N of Tweets from the Philippines	Province	N of Tweets Mentioned the Province
411,738	32,048	Cebu	756
		Leyte	510
		Samar	304
		Iloilo	199
		Bohol	188
		Eastern Samar	125
		Palawan	97
		Albay	80
		Capiz	68
		Aklan	64
		Southern Leyte	64
		Masbate	53
		Romblon	52
		Batangas	46
		Negros Occidental	45
		Antique	44
		Biliran	43
		Quezon	43
		Surigao del Sur	38
		Oriental Mindoro	21
		Sorsogon	19
		Occidental Mindoro	17
		Dinagat Islands	16
		Marinduque	16
		Negros Oriental	16
		Northern Samar	16
		Guimaras	14
		Camarines Sur	12
		Misamis Oriental	12
		Surigao del Norte	12
		Siquijor	11
		Agusan del Sur	9
		Agusan del Norte	7
		Camiguin	6

5. Results

Figure 4 shows the actual affected population intensity and the estimated affected population intensity resulting from the three methods (NAPI (Figure 4b), activity proxy index (Figure 4c), and the traditional population-density-based estimation method (Figure 4d)). The values of the proposed NAPI are calculated by the frequency of mentions in the tweets. The activity proxy index is calculated using the number of active Twitter users, and the traditional population-density-based estimation method employs a GIS spatial analysis method according to the gridded population data.

Figure 4b shows the NAPIs in the Philippines after Typhoon Haiyan. The blue dashed line indicates the impact range of the cyclone. The color is proportional to the NAPI: the redder the color, the higher the NAPI. In NAPIs of the 13 areas that were actually severely affected have high values, especially zones 1–3, which have larger NAPIs than the other zones. These values indicate that zones 1–3 had higher affected population intensities than other zones.

Figure 4c shows the distribution of activities after the cyclone when Twitter users' activities are used as a proxy index of the affected population intensity. In this paper, the Twitter users activities are the number of local users associated with Haiyan Typhoon divided by the local population in the post-disaster area. In Figure 4c, of the 13 severe zones, only zones 1 and 3 show a very high affected population intensity using the activity proxy index. Moreover, zones 7, 11, 12, and 13 show a lower activity proxy index than the actual intensity of the affected population.

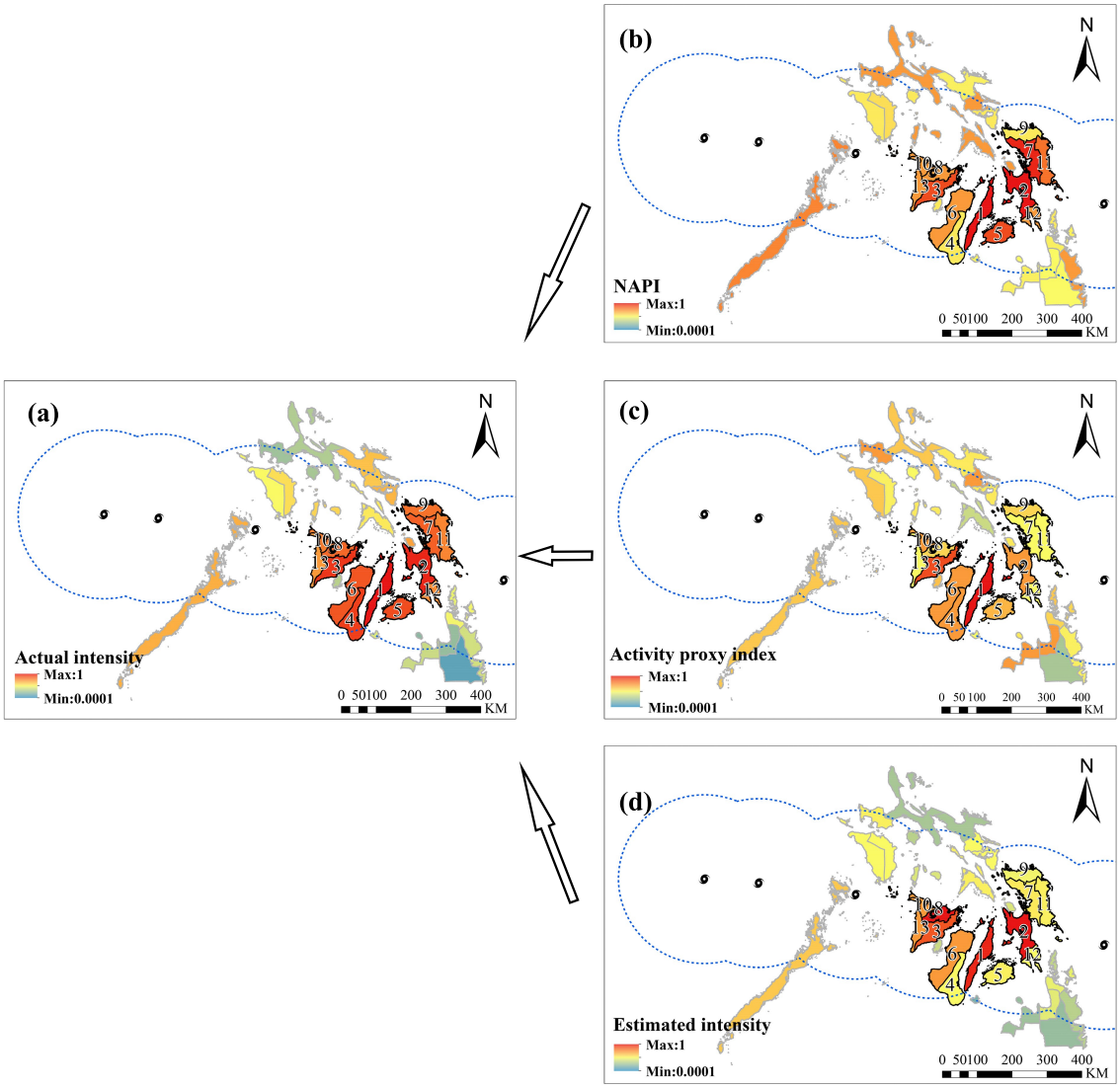


Figure 4. The affected population intensity due to Typhoon Haiyan. (**a**) The distribution of the actual intensity of the affected population and the estimated distribution of the intensity of the affected population resulting from (**b**) the normalized affected population index (NAPI), (**c**) the activity proxy index, and (**d**) the traditional population-density-based method.

The estimated affected population intensity using the traditional method is shown in Figure 4d. According to disaster system theory, the affected population can be considered as the coupling between the impact of the cyclone and the population exposed to the cyclone. Therefore, the affected population intensity can be estimated by the gridded population density and its inverse distance to the center of the cyclone. Figure 4d shows that zones 1 and 2 have a high affected population intensity. However, zone 8 presents a much higher estimated intensity than its actual intensity, and zones 4, 5, 7, 9, 11, and 12 have a lower estimated intensity than their actual intensity.

To compare the overall effectiveness of these three methods (NAPI (Figure 4b), the activity proxy index (Figure 4c), and the traditional population-density-based estimation method (Figure 4d)), a regression analysis was conducted. Figure 5 shows the results of performing linear regression between each of the three methods and the actual affected population intensity. All of them have a significant correlation with the actual affected population intensity. However, the coefficient of determination between the NAPI (Figure 5a) and the actual intensity is the highest (0.86). The coefficients of determination from regressing the actual affected population intensity on the

activity proxy index (Figure 5b) and the traditional population-based method (Figure 4c) are 0.64 and 0.54, respectively. These regression results suggest that the NAPI is more representative of the actual affected population intensity than the other two methods. Moreover, the obtained regression model can be used to determine the actual intensity of the affected population.

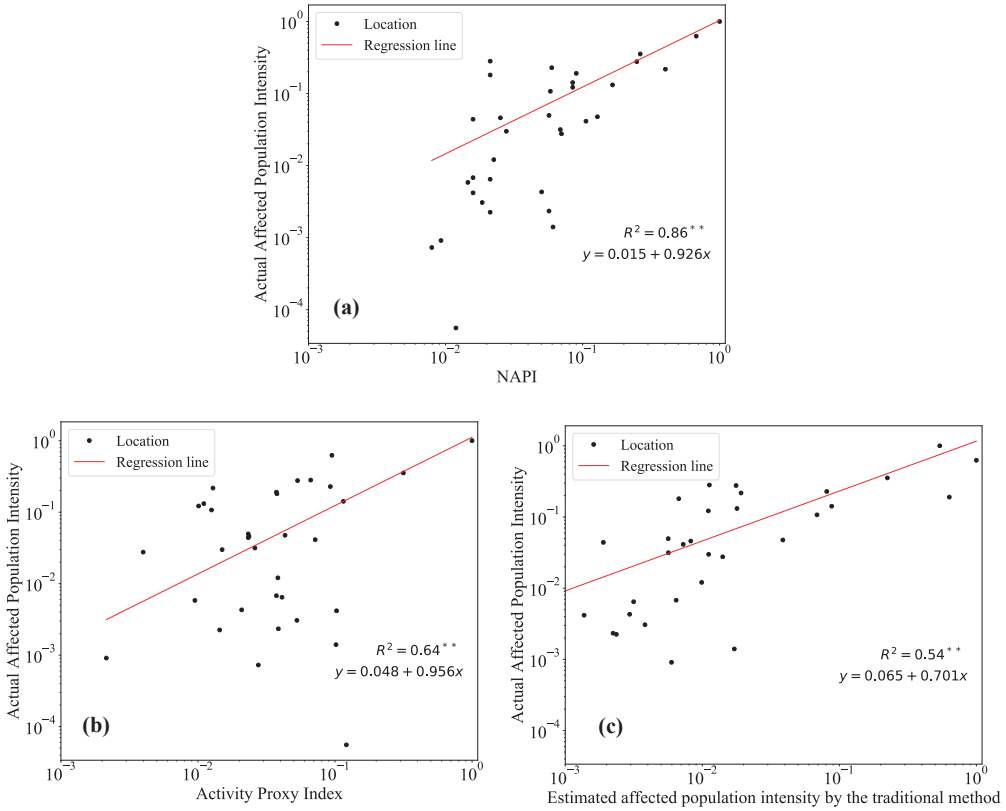


Figure 5. Regression analysis between each of the three methods and the actual affected population. (**a**) Relationship between NAPIs and the actual affected population intensity. (**b**) Relationship between the activity proxy index and the actual affected population intensity. (**c**) Relationship between the affected population intensity estimated by the traditional population-density-based method and the actual affected population intensity.

6. Discussion

The approach addressed in this paper presents a geographic method and efficient index for assessing the affected population intensity using social media data. The results of comparing the proposed NAPI index, the activity proxy index, and the traditional population-density-based method with the actual affected population intensity show that the NAPI has the greatest ability to assess the actual affected population intensity. Before computing the NAPI, calculation of the geographical attributes, including the spatial extent and period of a disaster, plays a critical role in improving the relevance of the social media data to natural disasters.

The method proposed in this paper can counteract the shortcomings of using census data in three ways. First, if a tourist posts a geolocated or geotagged tweet or other social media data, we can identify them as part of the disaster area. In this study, unfortunately, most of the tweets had no geolocation or geotagged information. Hence, in this case, we could only use the location registered with the user account to locate tweets. Second, the proposed method can, in fact, extract and use tweets from users located in other countries. However, to improve the quality and relevance, we conservatively removed tweets from outside the Philippines. Finally, the NAPI depends on the content

that people post in their tweets or other social media information. This information then illustrates the near-real-time and relatively realistic situation as perceived, witnessed, or heard by people in the disaster area.

However, these results should be considered within the scope and limitations of the present study. Only Typhoon Haiyan was used to test the approach. Typhoon Haiyan was chosen because it is the deadliest disaster to have occurred in the Philippines and one of the most destructive disasters in the world in the most recent decade. Moreover, the people in the Philippines form one of the most active social media populations in the world, which provides ideal conditions for studies that use social media data. Typhoon Haiyan has also been frequently chosen as a case study in previous research. Finally, the NDRRMC provided more detailed surveys on the post-Haiyan affected population than other disaster events, which enabled the comparison between our results and the actual affected population intensity.

Although the case study is about a cyclone, its results are representative because it is a typical world-influential disaster in the age of social media. Additionally, because of the lack of geolocation and geotagged tweets, we were unable to identify tourists' tweets in this case. To further deal with this limitation, additional techniques, such as geoparsing, are potentially useful solutions.

Compared with the research of Middleton, S.E. et al. [18], Yin, Jie, et al. [19], and Avvenuti, Marco, et al. [21], our research slightly differs in the starting point of the study. First, although the techniques presented in this article may not be the state-of-the-art, the results show the great capability and potential of applying simple geographical information that is present in social media. Second, we preferred the use of existing disaster information (e.g., disaster area, temporal range) as references to filter related tweets, while other studies developed methods to identify the disaster area and form situation awareness according to social media data. Third, the aim of our work—the evaluation of an affected population—is quite a specific problem and important for practical disaster assessment and relief operations, but this has received little discussion in previous studies. Nevertheless, the methods and models developed in these previous studies could also be used to our advantage in future works.

Moreover, the information contained in tweets can be leveraged for other emergency response purposes, as well. The NAPI is only constructed according to the geolocation information in tweets. Because of the complexity of natural disasters, the entropy-based method is likely to be an efficient means of evaluating the impact of disasters [47,48]. Furthermore, information about survival needs, disaster damage, and other data can also be extracted from tweets and provide emergency responders with useful guidance and support.

7. Conclusions

Assessing the affected population intensity is a critical mission for relief operations. This paper proposes a new approach and an index (NAPI) to leverage social media data for the assessment. Combined with geographic information, disaster-related tweets were efficiently extracted from social media data. The NAPI index, whose R^2 (0.86) from the linear regression with the actual affected population intensity was much higher than those of the activity proxy index and the traditional method, can accurately project the affected population intensity distribution by utilizing the geolocation information contained in tweets. Hence, more timely and accurate disaster information can be provided for emergency response by using near-real-time social media data and the NAPI index. For future work, more cases of other disasters and regions can be studied to verify this approach further. Furthermore, other information contained in tweets will be leveraged to provide diverse information to improve emergency response efforts.

Author Contributions: Conceptualization, Changxiu Cheng and Shi Shen; Data curation, Ting Zhang, Peichao Gao and Shi Shen; Formal analysis, Shi Shen; Funding acquisition, Changxiu Cheng; Investigation, Changxiu Cheng, Ting Zhang and Peichao Gao; Methodology, Changxiu Cheng, Kai Su and Shi Shen; Project administration, Changxiu Cheng; Resources, Changxiu Cheng and Shi Shen; Software, Ting Zhang and Kai Su; Supervision, Changxiu Cheng and Shi Shen; Validation, Ting Zhang, Kai Su, Peichao Gao and Shi Shen; Visualization, Ting Zhang, Kai Su and Shi Shen; Writing—original draft, Shi Shen; Writing—review & editing, Changxiu Cheng, Ting Zhang, Peichao Gao and Shi Shen.

Funding: This research was funded by National Key Research and Development Plan of China (Grant No. 2017YFB0504102), National Natural Science Foundation of China (Grant No. 41771537), and the Fundamental Research Funds for the Central Universities.

Acknowledgments: We are very grateful to the three reviewers for their comments and suggestions, which have greatly helped to improve the quality of this paper. We would also like to thank the high-performance computing support from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (https://gda.bnu.edu.cn/)

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

Abbreviations

The following abbreviations are used in this manuscript:

NAPI	Normalized Affected Population Index
LDA	Latent Dirichlet Allocation
SVM	Support Vector Machine
GIS	Geographic Information System

References

1. Shen, S.; Cheng, C.; Song, C.; Yang, J.; Yang, S.; Su, K.; Yuan, L.; Chen, X. Spatial distribution patterns of global natural disasters based on biclustering. *Nat. Hazards* **2018**, *92*, 1809. [CrossRef]
2. Ho, H.C.; Knudby, A.; Chi, G.; Aminipouri, M.; Lai, D.Y.F. Spatiotemporal analysis of regional socio-economic vulnerability change associated with heat risks in Canada. *Appl. Geogr.* **2018**, *95*, 61–70. [CrossRef] [PubMed]
3. Ho, H.C.; Wong, M.S.; Yang, L.; Chan, T.C.; Bilal, M. Influences of socioeconomic vulnerability and intra-urban air pollution exposure on short-term mortality during extreme dust events. *Environ. Pollut.* **2018**, *235*, 155–162. [CrossRef] [PubMed]
4. Goodchild, M.F.; Glennon, J.A. Crowdsourcing geographic information for disaster response: A research frontier. *Int. J. Digit. Earth* **2010**, *3*, 231–241. [CrossRef]
5. Shan, S.; Zhao, F.; Wei, Y.; Liu, M. Disaster management 2.0: A real–time disaster damage assessment model based on mobile social media data—A case study of Weibo (Chinese Twitter). *Saf. Sci.* **2019**, *115*, 393–413. [CrossRef]
6. Heir, T.; Rosendal, S.; Bergh-Johannesson, K.; Michel, P.; Mortensen, E.L.; Weisæth, L.; Andersen, H.S.; Hultman, C.M. Tsunami-affected Scandinavian tourists: Disaster exposure and post-traumatic stress symptoms. *Nord. J. Psychiatry* **2011**, *65*, 9–15. [CrossRef] [PubMed]
7. Resch, B.; Usländer, F.; Havas, C. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 362–376. [CrossRef]
8. Imran, M.; Castillo, C.; Diaz, F.; Vieweg, S. Processing social media messages in mass emergency: A survey. *ACM Comput. Surv.* **2015**, *47*, 1–38. [CrossRef]
9. Yuan, F.; Liu, R. Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study. *Int. J. Disaster Risk Reduct.* **2018**, *28*, 758–767. [CrossRef]
10. Gao, H.; Barbier, G.; Goolsby, R. Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *IEEE Intell. Syst.* **2011**, *26*, 10–14. [CrossRef]
11. Kryvasheyeu, Y.; Chen, H.; Obradovich, N.; Moro, E.; Van Hentenryck, P.; Fowler, J.; Cebrian, M. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2016**, *2*, e1500779. [CrossRef] [PubMed]
12. Tkachenko, N.; Jarvis, S.; Procter, R. Predicting floods with Flickr tags. *PLoS ONE* **2017**, *12*, e0172870. [CrossRef] [PubMed]

13. Mendoza, M.; Poblete, B. Valderrama, I. Nowcasting earthquake damages with Twitter. *EPJ Data Sci.* **2019**, *8*, 3. [CrossRef]
14. Barrington, L.; Ghosh, S.; Greene, M.; Har-Noy, S.; Berger, J.; Gill, S.; Lin, A.Y.; Huyck, C. Crowdsourcing earthquake damage assessment using remote sensing. *Ann. Geophys.* **2012**, *54*. [CrossRef]
15. De Albuquerque, J.; Herfort, B.; Brenning, A.; Zipf, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *Int. J. Geogr. Inf. Sci.* **2015**, *8816*, 1–23. [CrossRef]
16. Wu, D.; Cui, Y. Disaster early warning and damage assessment analysis using social media data and geo–location information. *Decis. Support Syst.* **2018**, *111*, 48–59. [CrossRef]
17. Schnebele, E.; Cervone, G. Improving remote sensing flood assessment using volunteered geographical data. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 669–677. [CrossRef]
18. Middleton, S.E.; Middleton, L.; Modafferi, S. Real-Time Crisis Mapping of Natural Disasters Using Social Media. *IEEE Intell. Syst.* **2014**, *29*, 9–17. [CrossRef]
19. Yin, J.; Lampert, A.; Cameron, M.; Robinson, B.; Power, R. Using Social Media to Enhance Emergency Situation Awareness. *IEEE Intell. Syst.* **2012**, *27*, 52–59. [CrossRef]
20. Huang, Q.; Xiao, Y. Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1549. [CrossRef]
21. Avvenuti, M.; Cresci, S.; Del Vigna, F.; Fagni, T.; Tesconi, M. CrisMap: A big data crisis mapping system based on damage detection and geoparsing. *Inf. Syst. Front.* **2018**, *20*, 993. [CrossRef]
22. Murzintcev, N.; Cheng, C. Disaster Hashtags in Social Media. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 204. [CrossRef]
23. Shen, S.; Murzintcev, N.; Song, C.; Cheng, C. Information retrieval of a disaster event from cross-platform social media. *Inf. Discov. Deliv.* **2017**, *45*, 220–226. [CrossRef]
24. Jaiswal, K.; Wald, D. An Empirical Model for Global Earthquake Fatality Estimation. *Earthq. Spectra* **2010**, *26*, 1017–1037. [CrossRef]
25. Aghamohammadi, H.; Mesgari, M.S.; Mansourian, A.; Molaei, D. Seismic human loss estimation for an earthquake disaster using neural network. *Int. J. Environ. Sci. Technol.* **2013**, *10*, 931–939. [CrossRef]
26. Bhatt, C.M.; Srinivasa Rao, G.; Asiya, B.; Manjusree, P.; Sharma, S.V.S.P.; Prasanna, L.; Bhanumurthy, V. Satellite images for extraction of flood disaster footprints and assessing the disaster impact: Brahmaputra floods of June–July 2012, Assam, India. *Curr. Sci.* **2013**, *104*, 1692–1700.
27. Raza, S.F.; Ahsan, M.S.; Ahmad, S.R. Rapid assessment of a flood-affected population through a spatial data model. *J. Flood Risk Manag.* **2017**, *10*, 219–225. [CrossRef]
28. Ozcelik, C.; Gorokhovich, Y.; Doocy, S. Storm surge modelling with geographic information systems: Estimating areas and population affected by cyclone Nargis. *Int. J. Climatol.* **2012**, *32*, 95–107. [CrossRef]
29. Roick, O.; Heuser, S. Location Based Social NetworksDefinition, Current State of the Art and Research Agenda. *Trans. GIS* **2013**, *17*, 763–784.
30. Cervone, G.; Sava, E.; Huang, Q.; Schnebele, E.; Harrison, J.; Waters, N. Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *Int. J. Remote Sens.* **2016**, *37*, 100–124. [CrossRef]
31. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users: Real–time Event Detection by Social Sensors. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 851–860.
32. Earle, P.S.; Bowden, D.C.; Guy, M. Twitter earthquake detection: Earthquake monitoring in a social world. *Ann. Geophys.* **2011**, *54*, 708–715.
33. Crooks, A.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. #Earthquake: Twitter as a Distributed Sensor System. *Trans. GIS* **2013**, *17*, 124–147.
34. Kent, J.D.; Capello, H.T. Spatial patterns and demographic indicators of effective social media content during theHorsethief Canyon fire of 2012. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 78–89. [CrossRef]
35. Yang, J.; Yu, M.; Qin, H.; Lu, M.; Yang, C. A Twitter Data Credibility Framework–Hurricane Harvey as a Use Case. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 111. [CrossRef]
36. Ragini, J.R.; Rubesh, A.P.M.; Vidhyacharan B. Mining crisis information: A strategic approach for detection of people at risk through social media analysis. *Int. J. Disaster Risk Reduct.* **2018**, *27*, 556–566. [CrossRef]
37. Fang, J.; Hu, J.; Shi, X.; Zhao, L. Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm. *Int. J. Disaster Risk Reduct.* **2019**, *34*, 275–282. [CrossRef]

38. Chang, J.; Boyd-Graber, J.; Gerrish, S.; Wang, C.; Blei, D.M. Reading tea leaves: How humans interpret topic models. In Proceedings of the 22nd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009; Curran Associates Inc.: Vancouver, BC, Canada, 2009; pp. 288–296.
39. Spence, P.R.; Lachlan, K.A.; Rainear, A.M. Social media and crisis research: Data collection and directions. *Comput. Hum. Behav.* **2016**, *54*, 667–672. [CrossRef]
40. Olteanu, A.; Castillo, C.; Diaz, F.; Vieweg, S. CrisisLex: A lexicon for collecting and filtering Microblogged communications in crises. In Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, MI, USA, 1–4 June 2014; pp. 376–385.
41. Olteanu, A.; Vieweg, S.; Castillo, C. What to expect when the unexpected happens: Social media communications across crises. In Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing, Vancouver, BC, Canada, 14–18 March 2015; ACM: New York, NY, USA, 2015; pp. 994–1009.
42. Imran, M.; Castillo, C.; Lucas, J.; Meier, P.; Vieweg, S. AIDR: Artificial intelligence for disaster response. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; ACM: New York, NY, USA, 2014; pp. 159–162.
43. Avvenuti, M.; Cresci, S.; Nizzoli, L.; Tesconi, M. Gsp (Geo-Semantic-Parsing): Geoparsing and Geotagging with Machine Learning on Top of Linked Data. In *European Semantic Web Conference (ESWC 2018)*; Springer: Cham, Switzerland, 2018.
44. National Disaster Risk Reduction and Management Council (NDRRMC). *Final Report Effects of Typhoon Yolanda (Haiyan)*; NDRRMC: Quezon City, Philippines, 2014.
45. Center for International Earth Science Information Network—CIESIN—Columbia University. *Gridded Population of the World, Version 4 (GPWv4): Administrative Unit Center Points with Population Estimates, Revision 11*; CIESIN: Palisades, NY, USA, 2018.
46. Knapp, K.R.; Kruk, M.C.; Levinson, D.H.; Diamond, H.J.; Neumann, C.J. The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Am. Meteorol. Soc.* **2016**, *91*, 363–376. [CrossRef]
47. Cushman, S.A. Calculating the configurational entropy of a landscape mosaic. *Landsc. Ecol.* **2016**, *31*, 481–489. [CrossRef]
48. Cushman, S.A. Thermodynamics in landscape ecology: The importance of integrating measurement and modeling of landscape entropy. *Landsc. Ecol.* **2015**, *30*, 7–10. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).