THE TYPHOON DISASTER ANALYSIS BASED ON WEIBO TOPIC HEAT

J. Yuan [1, 3], A. Gong [1, 2, 3], J. Wang [1, 2, 3], J. Li [1, 2, 3]

[1] Key Laboratory of Environmental Change and Natural Disaster, MOE, Beijing Normal University, Beijing, China - gad@bnu.edu.cn
[2] Beijing Key Laboratory of Environmental Remote Sensing and Digital City, Beijing Normal University, Beijing, China - gad@bnu.edu.cn
[3] Faculty of Geographical Science, Beijing Normal University, XinJieKouWai Street, Beijing, China - gad@bnu.edu.cn

Commission III, WG III/IVa

KEY WORDS: Typical cyclone disaster, Weibo data, Topic heat, Quantitative relationship, Damage assessment, Hazard management

ABSTRACT:

Could social media data be utilized in hazard evaluation? Typhoon disaster as one of the costly disaster has become devastating threats for human. Moreover, social media change the communication way of human and citizens can turn to this platform to express disaster-related information at real time. Therefore, social media improves situational awareness and widens the method of hazard information acquiring. With more and more studies investigating in relationship between social media response and degree of damage, the strong correlation has been proved. Weibo as one of the most popular social media in China can provide data with posted text, location, user identification and other additional information. Combining with 10 tropical cyclones and Weibo data in 2013, We perform a quantitative analysis between the grade of hazard situation and Weibo related topic heat in province scale. We provide a new model of Weibo topic heat to evaluate the Weibo activity in study area. Also we demonstrate the hazard assessing formula is $H=1.8845\ln(\alpha) + 15.636$ in tropical cyclone disaster. High level goodness of curve fitting also suggest that this equation can be used for rapid assessment of hazard caused by tropical cyclones.

1. INTRODUCTION

Typhoon disaster represents costly, which is the one of the most devastating threats for human nowadays (Hsiang et al., 2014; Smith et al., 2013). In the last 10 years, the direct economic loss of tropical cyclone around the world evenly reaches up to 50 billion US dollars every year, and the loss of China represents roughly 15% of the world (Chen et al., 2017). Due to the growing of population and assets continuously, there is increasing risking exposure to coastal areas of China (Pielke et al., 1998; Raghavan et al., 2003). Among such hazards, one of the challenges of emergency management is lack of the real-time status-information from affected areas.

Social media, including Twitter, Facebook, weibo and etc., has changed the way of communication for people (Kimanuka, 2015). Weibo, as one of the most famous social media platform in China, can provide real-time and round-the-clock situational awareness from people. In 2012, the daily active number of weibo users are up to 46.2 million, and the amount of posted messages expend 100 million (Tang et al., 2014). More over the number of Weibo users continues to be lager. Especially, as most texts are posted through mobile phone, weibo and other social media can offer data with location-based information and improve the hazard management research and practice (Roick et al., 2013).

Based on the advantages of social media data, more and more natural hazard studies make a good use of it in emergency situation. As for Chinese researchers, they are mainly attracted to dig the relationship between the disaster status presented from weibo text and the real. For instance, comparing the weibo data on the events Olypic Games and typhoon disaster was shown to strong correlation of the location distribution between affected areas and weibo texts, while Olympic event does not have (Chen et al., 2017). At abroad, existing researches on the use of Twitter in natural hazard is more manifold (Kryvasheyeu et al., 2016). Some researchers study its contribution to situational awareness (Vieweg et la., 2010; Power et al., 2014), and some pay attention to practical field of classifying disaster message, detecting events and identifying risking areas (Earle et al., 2011; Sakaki et al., 2010; Kumar et al., 2014; Lmran et al., 2013; Caragea et al., 2011). Especially for tropical cyclone disaster, Kryvasheyeu and others presented a multiscale analysis of Twitter activity and Hurricane Sandy and found a strong relationship between Sandy's path and topic related Twitter activity (Kryvasheyeu et al., 2016). Also, Yago's research prove that Twitter peak occurred preparedness phase and decreased abruptly after the storm (Martín et al., 2017). However, most studies on the use of weibo data are qualitative research.

Here, we can, for the first time, present a quantitative analysis of disaster related to Weibo topic heat. We start at provincial level, and progressively use spatial information of Weibo data to analyze the relationship between different factors and topic heat. Finally, based on the Weibo topic heat (H) and association degree (α), we find the equation to describe the quantitative correlation of H-α.

In this paper, the studying hazards were triggered by the 10 hurricanes landing in China of the 2013 season. Also, the Weibo data, collected from crisis provinces in the hurricane phases, are used to analyze the topic heat.

2. DATA COLLECTION AND PREPROCESSING

2.1. Tropical cyclone Information

There are 10 tropical cyclones of the 2013 season selected from CMA Tropical Cyclone Database from 'tcdata.typhoon.org.cn'

(Ying et al., 2014). Their names are Bebinca, Rumbia, Soulik, Cimaron, Jebi, Utor, Trami, Usagi, Fitow, and Haiyan, as shown in Figure 1. Also, we acquire more information of the tropical cyclones, including number, path, time, strength of wind. And all the selected made their landfall on the coastal area of China and caused different levels of disaster.
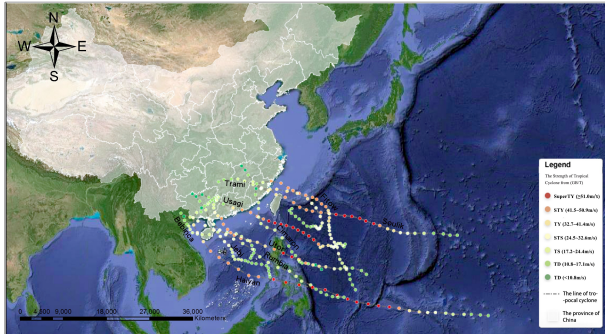

Figure 1. The path of 10 selected tropical cyclones in 2013

2.2. Damage Assessment: Tropical Cyclone

Because the tropical cyclone damage was mostly confined through different characteristics of hazard bearing body, to several provinces, we perform hazard analysis at a comprehensive perspective. Specifically, we examine it through the model for general grade division of typhoon disaster which is conducted by affected agricultural land, number of death, number of collapsed houses, and direct economic losses (Wang et al. 2014).

The primary data source contributes to the Yearbook of Meteorological Disaster in China (Song et al., 2014), including the needed characteristics of hazard bearing body for affected provinces of each tropical cyclone.

There are two steps for the above model applying. First, based on the graded standard of tropical cyclone damage in China, as shown in Table 1, we can establish the non-dimension function transform (Wang et al. 2014).

Item	Very Severe	Severe	Moderate	Slight	Negligible
Affected Agricultural Land (hm^2)	$(10^6,+\infty)$	$(10^5,10^6)$	$(10^4,10^5)$	$(10^3,10^4)$	$(10^2,10^3)$
Number of Death	$(10^2,+\infty)$	$(30,100)$	$(10,30)$	$(3,10)$	$(1,3)$
Number of Collapsed Houses	$(2*10^5,+\infty)$	$(10^5,2*10^5)$	$(3*10^4,10^5)$	$(3*10^3,3*10^4)$	$(1,10^3)$
Direct Economic Losses (RMB)	$(10^9,+\infty)$	$(10^8,10^9)$	$(10^7,10^8)$	$(10^6,10^7)$	$(10^5,10^6)$

Table 1. Graded standard of single disaster item in China

Second, by using the theory of grey association analysis (Yang, 1997; Wang et al. 2014; Fu, 1992), the association degree can be calculated and the general disaster level can be confined. The primary data and assessed results are summarized in Table 2. And the grading results covers all degree classes.

Name of Tropical cyclone	Affected Province of China	Affected Agricultural Land (10^4*hm^2)	Number of Death	Number of Collapsed Houses (10^4)	Direct Economic Losses (10^8RMB)	Association Degree (α)	Disaster Grade
Bebinca	Hainan	0.000	0.000	0.000	0.100	0.100	Ne
Rumbia	Guangdong	17.100	0.000	0.200	10.600	0.404	Sl
	Guangxi	1.400	0.000	0.100	1.000	0.283	Ne
	Synthesis	18.500	0.000	0.300	11.600	0.417	Sl
Soulik	Fujian	4.100	0.000	0.000	17.600	0.343	Ne
	Zhejiang	2.700	0.000	0.000	3.500	0.299	Ne
	Guangdong	1.200	5.000	0.200	5.800	0.398	Ne
	Jiangxi	2.200	2.000	0.100	4.900	0.353	Ne
	Anhui	0.000	0.000	0.000	0.100	0.100	Ne
	Synthesis	10.200	7.000	0.300	31.900	0.504	Mo
Cimaron	Fujian	2.200	4.000	0.200	19.800	0.430	Sl
Jebi	Hainan	2.000	0.000	0.000	3.100	0.290	Ne
	Guangdong	1.200	5.000	0.200	5.800	0.398	Ne
	Guangxi	1.700	0.000	0.100	1.300	0.293	Ne
	Synthesis	4.900	5.000	0.300	10.200	0.449	Sl
Utor	Guangdong	47.000	55.000	3.100	168.600	0.702	VS
	Guangxi	6.100	22.000	1.700	21.700	0.574	Mo
	Hunan	3.800	9.000	0.500	24.300	0.502	Mo
	Hainan	0.300	0.000	0.000	0.400	0.204	Ne
	Synthesis	57.200	86.000	5.300	215.000	0.744	VS
Trami	Fujian	3.000	0.000	0.100	19.200	0.364	Ne
	Zhejiang	3.900	0.000	0.100	6.100	0.345	Ne
	Hunan	3.700	0.000	0.100	7.700	0.349	Ne
	Guangxi	1.700	2.000	0.200	0.900	0.325	Ne
	Jiangxi	0.200	0.000	0.000	0.300	0.189	Ne
	Synthesis	12.500	2.000	0.500	34.200	0.468	Sl
Usagi	Guangdong	25.000	30.000	1.100	235.500	0.648	Se
	Hunan	10.100	4.000	0.100	6.700	0.425	Sl
	Fujian	6.700	0.000	0.100	20.700	0.383	Ne
	Guangxi	0.400	0.000	0.000	0.900	0.228	Ne
	Jiangxi	0.000	0.000	0.000	0.200	0.115	Ne
	Synthesis	42.200	34.000	1.300	264.000	0.666	Se
Fitow	Zhejiang	54.700	9.000	0.500	599.400	0.591	Mo
	Fujian	4.000	0.000	0.100	25.300	0.376	Ne
	Jiangsu	3.200	0.000	0.000	3.000	0.299	Ne
	Shanghai	2.800	2.000	0.000	3.700	0.326	Ne
	Synthesis	64.700	11.000	0.600	631.400	0.608	Se
Haiyan	Hainan	13.400	13.000	0.100	30.500	0.514	Mo
	Guangxi	34.000	7.000	0.400	14.400	0.519	Mo
	Guangdong	8.100	0.000	0.000	0.900	0.293	Ne
	Synthesis	55.500	20.000	0.500	45.800	0.606	Se

[a]Ne=Negligible, Sl=Slight, Mo=Moderate, Se=Severe, VS=Very Severe

Table 2. The tropical cyclone damage and general disaster grading

2.3. Raw Weibo Data and Filtering

The row data set for tropical cyclones in 2013 was collected through Weibo Application Programming Interface (API). And

the time interval of Weibo data was based on the duration of each tropical cyclone. Considering the post-evacuation process after the passage of storm, the time interval of each event extended three days after the ending of storm. Previous studies (Kryvasheyeu et al., 2016) suggested that there is a strong inverse relationship between distance to path hurricane and related social media activity, because the closer area can be directly hit by or close to. Therefore, the Weibo data gained from affected province can better contribute to the assessment of disaster situation.

All the Weibo data contain the text of the message, location of posting, and a range of additional significant information, such as time, user identification, follower counts, and so on. Due to repost messages in Weibo without location information, we only obtained the original to perform finer spatial analysis.

There has been study (Martín et al., 2017;) proved that tropical cyclone related social media activity can reach peak during pre-impact and preparedness phase and decrease after passage of storm. Moreover, previous study (Kryvasheyeu et al., 2016) also suggested that per-capita activity closely correlates with the per-capita economic damage. Therefore, we'd like to directly filter tropical cyclone related data to do further study in quantity relationship.

The Filtered data set should include some words which can convey the theme relevant to tropical cyclone event. There are Chinese keywords shown in Table 3, like "storm", "super-storm", which are selected from news, Weibo data samples and other text corpus. In total, for 10 tropical cyclones in 2013, we obtained 35,991 messages from unique 32,586 users.

Name of Tropical Cyclone	Formation of Tropical Cyclone	Tropical Cyclone Degree	Warning Notice	Disaster Situation		
贝碧嘉	热带气旋	超强台风	蓝色预警	山洪	大浪	大风
温比亚	低压涡旋	强台风	黄色预警	山崩	巨浪	强风
苏力	强对流	台风	红色预警	滑坡	强浪	烈风
西马仑	眼壁置换	强热带风暴		泥石流	狂涛	风暴潮
飞燕	眼壁结构	热带风暴		石流	洪水	暴雨
尤特	风眼	热带低压			洪灾	强降雨
潭美		热带低气压			洪涝	暴风雨
天兔						大暴雨
菲特						特大暴雨
海燕						

Table 3. The keywords for Weibo data filtering

2.4. Weibo Topic Heat

Weibo topic heat (H) here raised is one general comparison metrics to show Weibo activity. It can be conducted with amount of active users, number of text posted, number of active users' followers, and the classification of verified users. In general, posted messages from users with more followers have greater influence, and gain more attention from public. Also, messages from verified users mostly convey real, valid, and noticeable information.

Relying on the location information of filtered Weibo data, we did a spatial analysis through thermodynamic diagram shown in Figure 2. This diagram directly shows the tropical cyclone path, the situation of landform, and the Weibo activity in each storm event through the density of posted messages. As we can see, the most active areas were not the location closest to the tropical cyclone path, but the spaces were, in affected provinces, with flat terrain, large cities and inferred great population. Therefore, such areas with high-density fortune and population likely suffer more losses and have plenty active Weibo users to post related messages. Taking all the factors into consideration, the situation of underlying surface, the area of affected place, and the degree of hazard can be partly reflected from the number of messages and users.
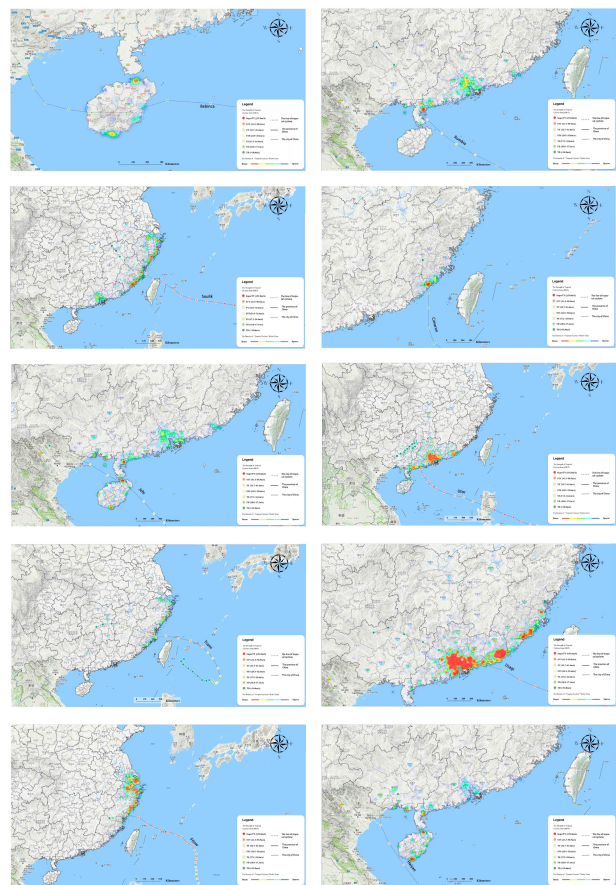

Figure 2. The thermodynamic diagram of tropical cyclone related Weibo activity in affected provinces

Relying on the relationship between above parameters and hazard situation, we put forward Equation (1) to calculate Weibo topic heat (H).

$$H = \ln \left(\frac{U}{M} * \sum_{m=1}^{M}(F * V) \right) \qquad (1)$$

where U = number of users
M = number of messages
F = number of user followers for each posted message
if the message from verified user, V = 1.5; if not, V=1

In Equation (1), the core thought is accumulating Weibo related messages to assess Weibo topic heat. However, there is different effect, when texts with the same theme are posted from different users. Therefore, user followers (F) and verified user (V=1.5) are

both made as weighting factors to show the impact of each event. Also, not only do number of messages (M) reflect weibo topic heat, but also the amount of active users (U) can do it as well. Therefore, we use the ratio between U and M, which has positive relationship with H, as another weighting factor.

3. RESULTS

3.1. Weibo Topic Heat Response to the grade of damage situation

Weibo topic heat is put forward to quantitatively assess Weibo activity. For each tropical cyclone, we can calculate H results of each affected province and the synthesis based on Equation (1). Combining the data of H results above and the Association Degree (α) shown in Table 2, we analyze the quantitative relationship of H – α.

The result of analysis, presented in Figure 3, shows strong positive relationship between H and α. From a general view of the data points in Figure 3, Weibo topic heat of tropical cyclone grows gradually while association degree increases.

Relying on the scatter points, we create the fit curve to describe the quantitative relationship of H-α. In fact, Weibo topic heat can not expend unlimitedly, as the amount of Weibo users in affected area is invariable. Therefore, logarithmic curve always stands out in all curve choices to present the changing trend.
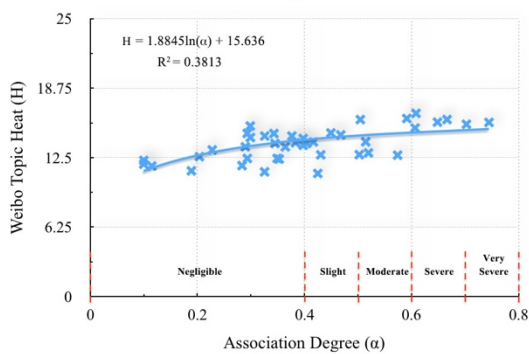

Figure 3. The fitting curve of Weibo Topic Heat and Association Degree

The equation of fitting result is as follow:

$$H = 1.8845\ln(\alpha) + 15.636 \qquad (2)$$

The significance of this regression is 0.000 presented in Table 4 and the R^2 is 0.3813, which both prove the high level of goodness of curve fitting.

	Quadratic Sum	df	Mean Square	F	Sig.
Regression	32.812	1	32.812	24.037	0.000
Residual	53.237	39	1.365		
Total	86.049	40			

Table 4. The ANOVA of H-α curve fitting

Arguably, the Weibo topic heat can be higher with the increasing of disaster degree. It also proves the result of previous study that

there is strong correlation with per-capital social media activity and the per-capital economic damage (Kryvasheyeu et al., 2016). This phenomenon can be explained by the dissemination of emergency information in different forms through media. More severe disaster, more people affected or close to, and finally more related information posted in Weibo.

4. DISCUSSION

4.1. Finer spatial analysis

There is limitation to the approach in this paper. In fact, finer spatial analysis can better assist governments in hazard management. They eager to grasp real-time information in each affected county or even finer scale to allocate resources.

However, when we pay attention to it, we not only need support in data source, but also should consider complex situation in real life. The density of related Weibo activity shown in Figure 2 has proved that there is no longer strong inverse relationship between the distance to path of cyclones and Weibo activity in county scale. There are a number of factors contribute to this result, including underlying surface, population, losses amount and others. For example, places with expanse of crop and large population can suffer more losses during one strong tropical cyclone. On the contrary, the losses at mountainous area with few people can be much less and negligible.

Further, we need to consider some situations in real-life, which may disturb the assessment by using Weibo topic heat. For example, people directly hit by very severe typhoon do not have mood and internet environment to post Weibo text, which may lead to lower Weibo heat in affected severely area. Moreover, there may be not great population to post Weibo in the countryside with badly affected agricultural land. Therefore, whether the conclusion of previous study (Kryvasheyeu et al., 2016), per-capita Twitter in terms of per-capita economic damage, can be adaptive finer scale in China is still need to be investigated.

4.2. The applicability of H-α equation

With time going, all factors can be changed. For example, the amount of Weibo users continues growing larger; the efficiency of hazard management becomes better; And the frequency of tropical cyclones may increase, which may lead more people pay attention to cyclones. Therefore, the applicability of H-α equation should be modified and verified by several-years data for further application.

5. CONCLUSION

This article did an analysis on the quantitative relationship between the grade of hazard situation and Weibo topic heat, which can be a new potential means to assess the situation in risk area at real time.

Considering the open source and tags of data, we finally utilize the Weibo API to acquire social media data to monitor social response to tropical cyclone events. Also, the hazard information and analysis here is based on the provincial scale.

The Weibo topic heat was conducted by several factors, including the number of posted messages, the amount of active users, the follwers of user and whether the user being verified. Taking the relationships with each factor into consideration, we put forward to the Weibo topic heat equation.

Combining the results of Weibo heat and the association degree (α, utilized to describe the general degree of damage), we use logarithmic curve to fit it. Finally, we acquire one equation of H-α as followed $H = 1.8845\ln(\alpha) + 15.636$, which got the high level of goodness of curve fitting. This result suggested that this equation can be made a good use in further rapid assessment application.

ACKNOWLEDGEMENTS

REFERENCES

Hsiang, J., 2014. The causal effect of environmental catastrophe on long-run economic growth: Evidence from 6,700 cyclones. National Bureau Economic Research Working Paper Series, Cambridge, MA, USA.

Smith, K., 2013. US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Natural Hazards*, 67(2), pp. 387-410.

Chen, D., Lu, F., Shi, R., 2017. Review on Tropical Cyclone Risk Assessment. *Journal of Catastrophology*, 32(4), pp. 146-152.

Pielke, L., 1998. Normalized hurricane damages in the United States: 1952-95. *Weather and Forecasting*, 13(3), pp. 621-631.

Raghavan, R., 2003. Trends in tropical cyclone impact: A study in Andhra Pradesh, India. *Bulletin of the American Meteorological Society*, 84(5), pp. 635-644.

Kimanuka, 2015. Social media has changed the way we communicate. In: *The New Times*, http://www.newtimes.co.rw/section/read/188395.

Tang, X., 2014. Hotspot Mining Based on LDA Model and Microblog Heat. *Library and Information Service*, 58(5), pp. 58-63.

Roick, H., 2013. Location Based Social Networks - Definition, Current State of the Art and Research Agenda. *Transactions in GIS*, 17(5), pp. 763-784.

Chen, G., Luo, Z., 2017. Empirical discussion on relation between realistic disasters and social media data. *Science of Surveying and Mapping*, 42(8), pp. 44-48.

Kryvasheyeu, C., Obradovich, M., Hentenryck, F., Cebrian, 2016. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3), e1500779.

Vieweg, H., Starbird, P., 2010. Microblogging during two natural hazards events: What Twitter may contribute to situational awareness, Proceedings of the SIGHI Conference on Human Factors in Computing Systems(CHI'10), Atlanta, GA, USA.

Power, R., Colton, C., 2014. Emergency situation awareness: Twitter case studies. In: *International Conference on Information Systems for Crisis Response and Management in Mesiterranean*, pp. 218-231.

Earle, B., Guy, 2011. Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics*, 54(6), pp. 708-715.

Sakaki, O., Matsuo, 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In: Proceedings of the Nineteenth International World Wide Web Conference(WWW'10), pp. 851-860.

Kumar, H., Liu, 2014. A behavior analytics approach to identifying tweets from crisis regions. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media(HT'14).

Lmran, E., Castillo, D., Meier, 2013. Extracting information nuggets from disaster-related messages in social media. In: Proceedings of the Tenth International Conference on Information Systems for Crisis Response and Management (ISCRAM'13).

Caragea, M., Jaiswal, T., Kim, M., Wu, T., Giles, J., 2011. Classifying text messages for the Haiti earthquake. In: Proceedings of the Eighth International Conference on Information Systems for Crisis Response and Management (ISCRAM'11).

Martín, L., Cutter, 2017. Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew, 12(7), e0181701.

Ying, M., Zhang, Y., Lu, F., Fan, Z., Chen, 2014. An overview of the China Meteorological Administration tropical cyclone database. *Journal of Atmospheric and Oceanic Technology*, 31, pp. 287-301.

Wang, W., Ma, 2010. Model for General Grade Division of Typhoon Disasters and Application. *Meteorological Monthly*, 36(1), pp. 66-71.

Yang, 1997. Study on the Model of Grade Division of Natural Disaster and Comparision of Disasterous Conditions. *Meteorological Monthly*, 36(1), pp. 66-71.

Fu, 1992. *Theory of Grey Association and Application*. China Science and Technology Press, Beijing, pp. 191-199.

Song, F., 2014. *Yearbook of Meteorological Disaster in China*. China Meteorological Press, pp. 25-37.

Revised March 2018