Analyzing windstorm pattern in Malaysia based on extracted Twitter data.

ISA, N.A., SALLEH, S.A., CHAN, A., ZAKARIA, N.H., SUIF, Z. and HALIM, M.A.

2022

PAPER • OPEN ACCESS

Analyzing Windstorm Pattern in Malaysia based on Extracted Twitter Data

To cite this article: N A Isa *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1019** 012011

View the article online for updates and enhancements.

Analyzing Windstorm Pattern in Malaysia based on Extracted Twitter Data

N A Isa[1], S A Salleh[2,3]*, A Chan[4], N H Zakaria[2], Z Suif[5] and M Abdul Halim[2]

[1]Faculty of Architecture, Surveying and Real Estate, Geomatika University College, Setiawangsa, 54200 Kuala Lumpur, Malaysia.

[2]Centre of Studies for Surveying Science and Geomatics, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

[3]Institute of Biodiversity and Sustainable Development, Universiti Teknologi MARA, Shah Alam, Malaysia

[4]Department of Civil Engineering, Faculty of Engineering, University of Nottingham, Semenyih, Selangor, Malaysia

[5]Department of Civil Engineering, Universiti Pertahanan Nasional Malaysia, Sungai Besi, Kuala Lumpur, Malaysia

*aekbal@uitm.edu.my

Abstract. Wind-rain interactions often lead to severe windstorm events and consequently cause damages and fatal destructions. The increase in frequency of recent windstorm events overwhelmed the nation. Thus, efforts in obtaining and recording these events are intensified with the help of current technology. This study aims to analyze the pattern of recent windstorm events by utilizing big data and GIS. In this study, the reported windstorm events in Twitter application were extracted using R-programming. Prior to analyses, the extracted data were screened to remove any outliers found. The extracted data were selected based on the credibility of its sources to ensure the accuracy and quality. These selected data were extracted from trusted users such as Meteorological Department of Malaysia (MMD), Berita Harian, Bernama and others. This study has demonstrated the possibility of Twitter data as an alternative data source in windstorm studies based on its reasonable findings. It is exhibited that there is drastic increased of windstorm events frequency in years 2018-2020, especially in the northern and west-coast regions of Peninsular. The highest frequency was recorded in April (inter-monsoon season) while the lowest is in February and December (northeast monsoon). The increase of frequency in several locations in the Peninsular is very alarming especially in the Klang Valley since this region is highly populated and serves as Malaysia's important economic zones. Hence, risk control should be considered in this region to reduce the negative impacts as suggested in SDG11 and SDG13.

Keywords: Windstorm, big data, GIS, R-programming, SDG11, SDG13

1. Introduction

Wind is a significant component of the climate in the tropics. Wind has a sizable influence on daily weather in this region [1]. Recent changes in wind behavior have sparked global concern due to the occurrence of numerous destructive phenomena. Windstorms have increased in frequency in the majority of tropical regions, including Malaysia [2]. Windstorms are a relatively recent phenomenon in

Malaysia. Since its occurrences caused no or minor damage in the past, until recently, the phenomenon became more frequent, causing serious damage and fatalities [3].

Due to a lack of information and appropriate instruments, windstorm occurrences in Malaysia are not effectively observed and documented [2]. Furthermore, because the occurrences of windstorms are unpredictable in terms of location, capturing the events on the ground is extremely difficult. The advancement of technology has made research in this field less challenging [4]. Several methods and techniques to obtain the data for windstorm events are employed, especially big data analytics. Nowadays, big data analytics often preferred by many in obtaining various data for numerous studies worldwide [5]. Because of its rich content, several researches have used big data as one of the prospective data sources to address the scarcity of well-documented data. However, to assure the quality of the data sources and the precision of the analysis undertaken, they should be managed [6].

This research aims to analyze the windstorm events by utilizing big data analytics for windstorm research in Malaysia. Given the benefits of big data analytics, this study chose this strategy to address the windstorm incidents. The approach is exceedingly cost-effective, simple to use, and produces extremely rich content [7]. Data can be extracted using numerous ways utilising specified keywords and is freely available online.

2.   Study Area

This study focuses on Malaysia environment only. Any windstorm events mentioned in tweets were extracted in all states in Malaysia. Since Malaysia is located in the tropical region, the country is hot and humid throughout the year with little variations in temperature and humidity. It experiences the hottest month in March where the highest temperature is approximately 36°C while the coldest month is in January with the lowest temperature of approximately 22°C. The annual mean temperature is between 22.5°C to 34°C [8].
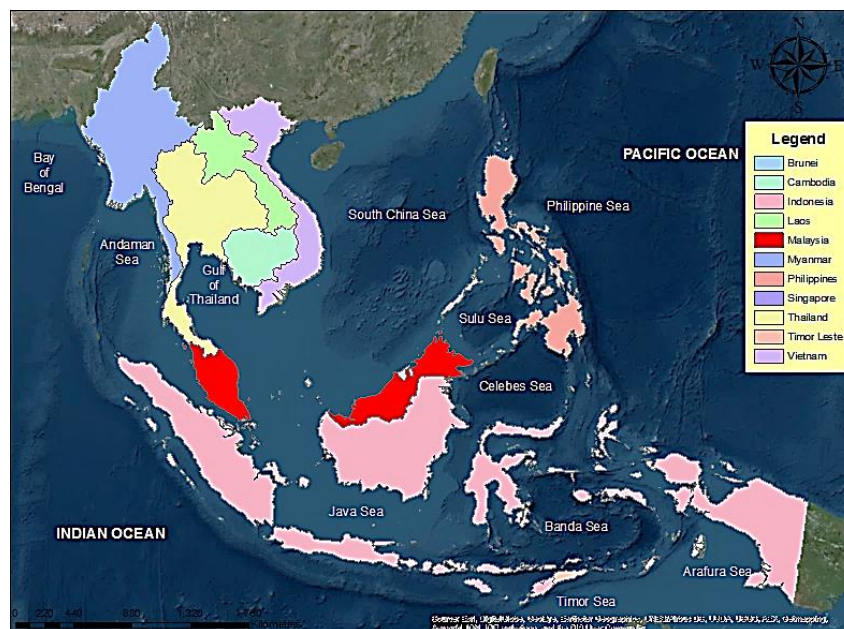
Figure 1 shows the location of Malaysia within Southeast Asia [9]

Situated in Southeast Asia, this country is surrounded by Thailand (north), Singapore and Java Island of Indonesia (south), Philippines (east) and Sumatera Island of Indonesia (west) [9]. Figure 1 shows the location of Malaysia and its neighbouring countries [9]. This location gives Malaysia a common tropical experience with monsoon seasons, El Nino–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD) and Madden–Julian Oscillation as well as extreme events such as floods and droughts [10]. Due to these,

the wind within this region is unstable and unpredictable leading to lack of understanding of wind-related phenomena especially windstorms.

3.   Methods

This study is conducted in two main stages. Figure 2 shows the workflow and research activities conducted during the study.
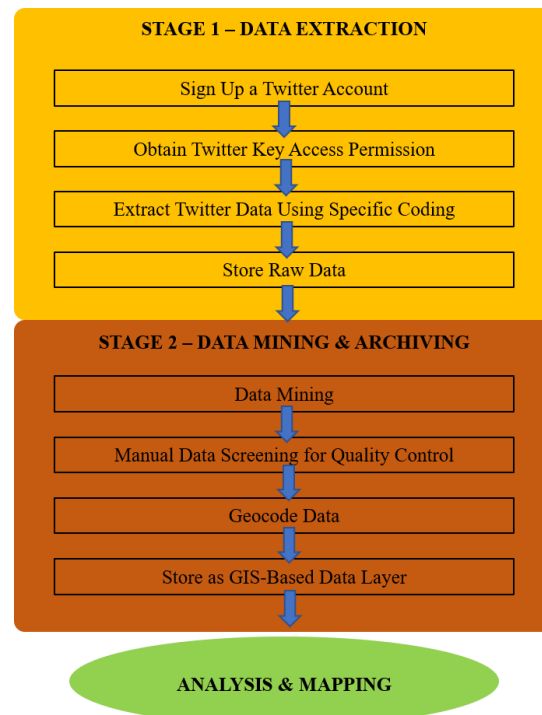
Figure 2 Study Workflow

3.1 Data Extraction

This study employed R-programming to extract the data from Twitter. The application is freely available online with numerous packages to extract big data from various applications. Prior to data extraction, Application Programming Interface (API) key and access token were obtained from Twitter to ensure that the permission to extract the data is given. The API key and access token can be applied through Twitter website. Twitter package should be embedded in the R-Studio so that the data extraction can be conducted.

In this study, the data were extracted specifically within Malaysia along with the coordinate of each tweet (if available) for mapping purposes. Using the following command line, tweets made within specific radius in desired locations were extracted:

```
tweets <- searchTwitter("ribut", n = 1000, since = "2020-03-01", until = "2020-03-02", geocode =
'2.830417752634409, 101.72736752017727,100km', retryOnRateLimit = 10000)
```

This command line is able to extract the text, location, user ID and coordinate (if available) of each tweet. The extracted data were saved for screening to eliminate irrelevant entries. Several keywords were used which indicate windstorm events occurrences such as "ribut" and "angin kencang".

3.2 Data Mining and Archiving

Since the extracted tweets may contain irrelevant entries, data mining was conducted to ensure that the tweets contents are related to windstorm events. Any irrelevant tweets were excluded to ensure the

precision of further analyses. This technique was also used to eliminate repetitive tweets for the same events to be archived. This also can affect the outcomes from this study. Thorough screening was conducted to ensure the archived data were in good quality.

The extracted data were archived in GIS data layer formats. Each tweet was assigned to the state it belongs to. For every entry, four important information were documented which are the location, month, year and the damage caused.

4.   Result and Discussions

In this study, the windstorm events tweeted within Malaysia were extracted and recorded according to the states. The patterns of the events within the last decade were documented, analyzed and mapped. The following subsections explained the study findings:

4.1   Windstorm Events in Malaysia

The study has identified an increasing pattern of windstorm events throughout the country. A drastic increase was found during 2018 and kept increasing onwards. The highest windstorm event was recorded in 2020 while the lowest was in 2011. This finding is parallel with previous studies which have reported an alarming increase of windstorm occurrences [2]-[3],[10]. Even though this is the case, several months were found to have calmer wind especially during November to January. The highest event counts were recorded in April (inter-monsoon), while the lowest numbers were recorded in December and February (Northeast monsoon). This could be related to inclement weather and unpredictably strong winds during the inter-monsoon season, as well as climate change [11]. Figure 3 and Figure 4 shows the trend of windstorm occurrences during the last decade in Malaysia.
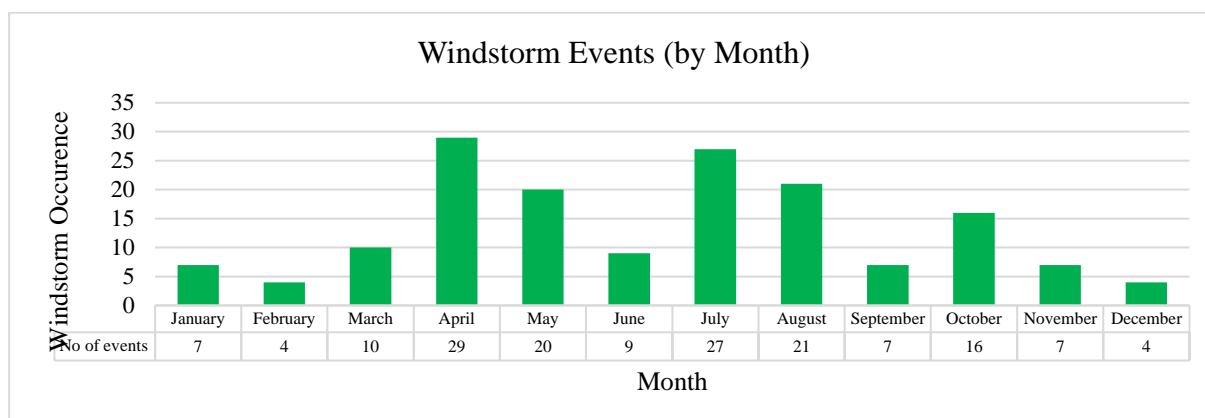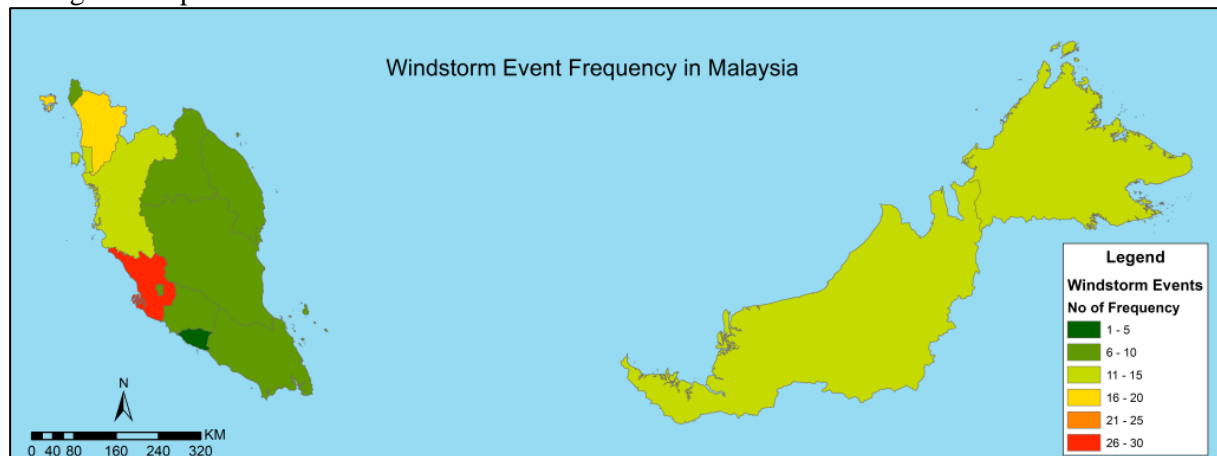

Figure 3 Windstorm events in Malaysia
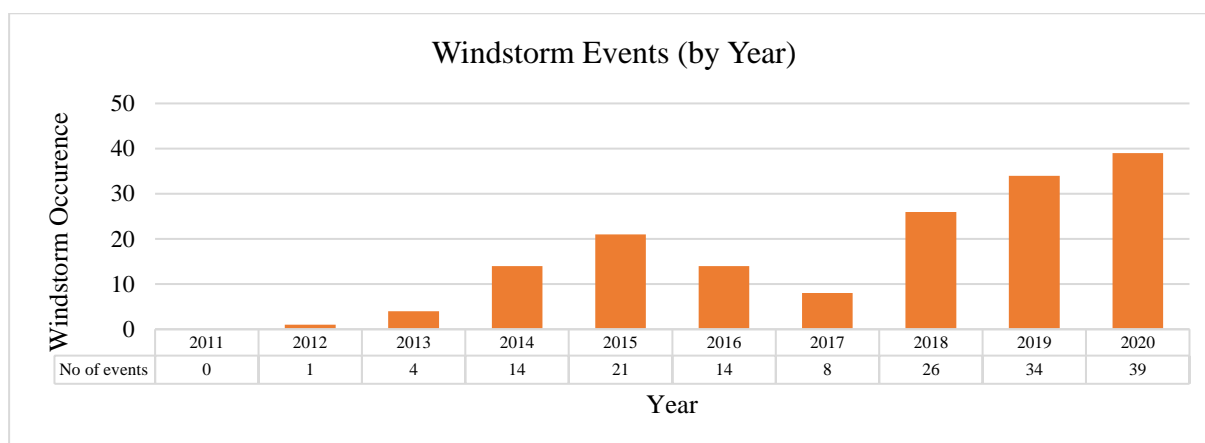
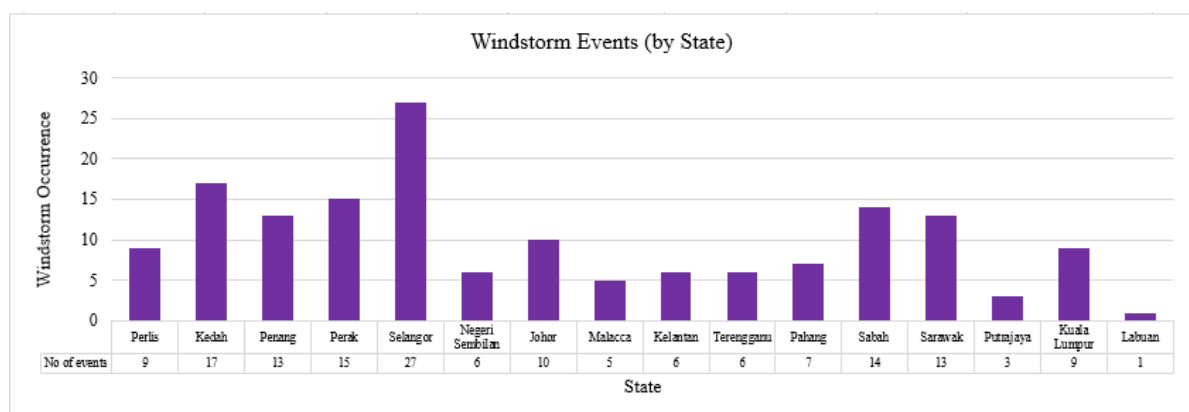**Figure 4** Windstorm events in Malaysia

The study also discovered that two parts of the Malay Peninsula were particularly vulnerable to windstorms: the northern and western coasts, as shown in Figure 5. Selangor is the most afflicted state, while only one case reported in Labuan. Kedah was discovered to be the most afflicted state in the northern region, while Selangor was found to be the most damaged state on the Peninsula's west coast. The east coast of the Peninsular, as well as Sabah and Sarawak, had a lower number of instances. This outcome is also consistent with the findings of Zakaria et al. [2] and Majid et al. [3] that were based on the last few years of observations.

4.2 Windstorm Events in the states of Malaysia

According to the study, the two affected regions were hit by a windstorm during the inter-monsoon season (April). Selangor, Kedah, Penang, Perak, Kuala Lumpur, Johor, and Malacca were the states with the most cases. Because all of these states are located along the Malacca straits, common geographic-related issues should be examined. The monsoon season is prevalent in these areas, which may act as a catalyst for windstorms. Figure 6 shows the number of windstorm occurrences for each state in Malaysia. Further study on the prone areas should be conducted to identify the common factors that contribute to the incidences of windstorms. This information can be used to mitigate or at least reduce its negative impacts.

Figure 5 Prone areas of windstorm events in Malaysia

Figure 6 Number of windstorm events according to the states in Malaysia

Table 1 shows the windstorm events in each year for every state in Malaysia. From the table, it can be concluded that the windstorm struck almost the entire country in 2018-2020. The most affected year is 2020 with 14 states recorded. Selangor is the most affected state in 2018-2020 with highest case count (21 cases). A drastic increase was also recorded in Selangor during 2018-2020 as compared to the

previous years. This could be attributed to Selangor's built environment and landscape, which is dominated by buildings of various shapes, sizes, and heights [11]. Because of this, it is necessary to investigate the effects of urban geometry on wind behaviour in Selangor. Other than Selangor, Kedah, Perlis and Penang were also found to be badly affected by the phenomenon in the year of 2018-2020.

Table 1 Windstorm cases in each year for every state in Malaysia

State	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	TOTAL
Perlis	0	0	0	0	1	3	0	1	3	1	9
Kedah	0	1	0	2	0	2	1	1	6	4	17
Penang	0	0	2	0	0	0	2	2	4	3	13
Perak	0	0	0	2	4	0	1	2	3	3	15
Selangor	0	0	0	3	1	2	0	6	6	9	27
Negeri Sembilan	0	0	0	0	1	1	0	2	1	1	6
Johor	0	0	0	0	1	2	2	0	2	3	10
Malacca	0	0	1	0	0	1	0	0	1	2	5
Kelantan	0	0	0	0	0	1	0	4	0	1	6
Terengganu	0	0	0	0	2	0	1	0	1	2	6
Pahang	0	0	0	1	1	0	0	2	1	2	7
Sabah	0	0	0	3	4	0	1	1	1	4	14
Sarawak	0	0	0	1	5	0	0	3	2	2	13
Putrajaya	0	0	1	0	0	0	0	0	0	2	3
Kuala Lumpur	0	0	0	1	1	2	0	2	3	0	9
Labuan	0	0	0	1	0	0	0	0	0	0	1

5.   Conclusion

This preliminary study has demonstrated the possibilities of utilizing Twitter as a data source to document windstorm events in Malaysia. It shows that, the technique able to return reasonable results in analyzing the pattern of windstorm events. However, integration and consolidation with other social media or application is highly recommended to ensure all or at least most of the events can be documented for further investigations. This is also to ensure the analyses conducted utilized adequate data sample, hence, the conclusions made were based on a firm ground.

From the study, a drastic increased in windstorm occurrences in several states is very alarming. The data showed that these events are frequent during the inter-monsoon season from March to April. It is found that, the northern and west-coast region of the Peninsula are prone to windstorm events in the last decade. In the last few years, windstorm events stroke the Klang Valley heavily with high damages and fatalities recorded. This is very distressing since the region is the most important economic zone in Malaysia. Thus, prevention and mitigation should be prioritized in this region.

Acknowledgements

References

[1] Chavas D R, Reed K A and Knaff J A 2017 Physical understanding of the tropical cyclone wind-pressure relationship *Nature Communication* **8** 1-11

[2] Zakaria N H, Salleh S A, Asmat A and Islam M A 2019 Seasonal windstorm pattern and damages in Peninsular Malaysia 2018 *IOP Conference Series: Earth and Environmental Science* **385** 1-12

[3] Majid T A, Zakaria S A S, Wan Chik F A, Deraman S N C, and Muhammad M K A 2016 Past windstorm occurrence trend, damage and losses in Penang, Malaysia *Journal of Engineering Science and Technology* **11** 397-406

[4] Wan Chik F A, Che Deraman S N A, Noram I R, Muhammad M K A, Majid T A and Zulkarnain N Z 2014 Development of windstorm database system for wind damages in Malaysia *Journal of Civil Engoneering Research* **4** 214-217

[5] Othman S, Abdul Halim M, Mohamed Saraf N and Salleh S A 2021 Investigation on accident hotspot area using social media data *IOP Conference Series: Earth and Environmental Science* **767** 1-9

[6] Arolfo F, Rodriguez K C and Vaisman A 2020 Analyzing the quality of Twitter data stream *Information Systems Frontiers* doi.org/10.1007/s10796-020-10072-x

[7] Abdul Halim M, Mohamed Saraf N, Hashim N I, Abdul Rasam A R, Idris A N and Md Saad N 2018 Discovering new tourist attractions through social media data: A case study in Sabah, Malaysia presented at *IEEE 8th International Conference on System Engineering and Technology (ICSET 2018)* in Bandung, Indonesia

[8] Ibrahim I and Samah A A 2011 Preliminary study of urban heat island: Measurement of ambient temperature and relative humidity in relation to landcover in Kuala Lumpur presented at *19th International Conference on Geoinformatics* in Shanghai, China

[9] Isa N A, Salleh S A, Wan Mohd W M N and Chan A 2018 Kuala Lumpur City of Tomorrow: Integration of Geospatial Urban Climatic Information in City Planning *Theoretical and Empirical Researches in Urban Management* **13** 5-27

[10] Tangang F T, Juneng L, Salimun E, Sei K M, Le L J and Muhamad H 2012 Climate change and variability over Malaysia: Gaps in science and research information *Sains Malaysiana* **41** 1355–1366

[11] Ooi M C G, Chan A, Subramaniam K, Morris K I and Oozeer M Y 2017 Interaction of urban heating and local winds during the calm intermonsoon seasons in the tropics: Urbanization effects on local circulation *Journal of Geophysical Research: Atmospheres* **122** 499-510