Assessing the social impacts of extreme weather events using social media

Submitted by Michelle Dawn Spruce, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Computer Science in March 2022.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

(Signature) ……M.D.Sp……

Abstract

The frequency and severity of extreme weather events such as flooding, hurricanes/storms and heatwaves are increasing as a result of climate change. There is a need for information to better understand when, where and how these events are impacting people. However, there are currently limited sources of impact information beyond traditional meteorological observations.

Social sensing, which is the use of unsolicited social media data to better understand real world events, is one method that may provide such information. Social sensing has successfully been used to detect earthquakes, floods, hurricanes, wildfires, heatwaves and other weather hazards. Here social sensing methods are adapted to explore potential for collecting impact information for meteorologists and decision makers concerned with extreme weather events.

After a review of the literature, three experimental studies are presented. Social sensing is shown to be effective for detection of impacts of named storms in the UK and Ireland. Topics of discussion and sentiment are explored in the period before, during and after a storm event. Social sensing is also shown able to detect high-impact rainfall events worldwide, validating results against a manually curated database. Additional events which were not known to this database were found by social sensing.

Finally, social sensing was applied to heatwaves in three European cities. Building on previous work on heatwaves in the UK, USA and Australia, the methods were extended to include impact phrases alongside hazard-related phrases, in three different languages (English, Dutch and Greek).

Overall, social sensing is found to be a good source of impact information for organisations that need to better understand the impacts of extreme weather. The research described in this project has been commercialised for operational use by meteorological agencies in the UK, including the Met Office, Environment Agency and Natural Resources Wales.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Hywel Williams for the continuous support of my PhD study, related research and all the other opportunities given to me in my time as a research student at the University of Exeter. Thank you for your patience, motivation, and unfailing positivity when I needed it most.

I would also like to thank my second supervisor Rudy Arthur for his insightful comments and encouragement with my work, but also for the hard questions which inspired me to widen my research from various perspectives. I am also grateful to Joanne Robbins at the Met Office for her time, discussions and enthusiasm for the wider context of my work.

My PhD was only possible thanks to an EPSRC studentship (grant number EP/N509656/1). I greatly appreciate all of the opportunities that this funding gave me.

I thank my fellow SEDALab colleagues for the informative discussions, both in person and online. Particular thanks to Tristan Cann, Josh Buxton and James Young for your help, advice and support.

Also, I thank my long-standing friend, Becky Turner, for her encouragement to start the PhD journey in the first place, and for being a voice of reason and understanding in my life. Thanks as well to my Mum and Dad who have always taught me not to be afraid to seize opportunities and try new things. Like stars, I know that you are all always there, even if I don’t see you as often as I would like.

Last, but by no means least (I really couldn’t have got to the end without you!), extra special recognition to my husband Richard and children Emma and Charlotte for their love, encouragement and constant support. Team Spruce!

Table of Contents

List of Figures

List of Tables

Author’s Declaration

The following chapter is currently being prepared for submission to a relevant journal. Expected submission in September 2022:

Chapter 2 - Using social media to detect severe weather events and evaluate impacts: a systematic literature review

Additionally, some of the chapters in this thesis have been previously published and contain work by co-authors.

Chapter 4: Using social media to measure impacts of named storm events in the United Kingdom and Ireland

Chapter 4 has been published in the journal Meteorological Applications in 2020 (Spruce *et al.*, 2020) and was completed with Rudy Arthur[1] and Hywel T.P. Williams[1].

For this paper I:

- Designed the experimental setup.
- Generated the experimental results.
- Interpreted the results.
- Designed the visualisations of all included figures.
- Wrote the first draft of the manuscript, and edited it in response to comments from my co-authors and reviewers.

The version printed in this thesis is as published in Meteorological Applications. This content is reproduced under the Creative Commons Attribution licence.

Chapter 5: Social sensing of high-impact rainfall events worldwide: A benchmark comparison against manually curated impact observations

Chapter 5 has been published in the journal Natural Hazards and Earth System Sciences in 2021 (Spruce *et al.*, 2021) and was completed with Rudy Arthur[1], Hywel T. P. Williams[1] and Joanne Robbins[2].

[1] College of Engineering, Maths and Physical Sciences, University of Exeter, Exeter EX4 4SB, UK
[2] Met Office, Exeter EX1 3PB, UK

For this paper I:

- Designed the experimental setup.
- Generated the experimental results.
- Interpreted the results.
- Designed the visualisations of all included figures.
- Wrote the first draft of the manuscript, and edited it in response to comments from my co-authors and reviewers.

The version printed in this thesis is as published in Natural Hazards and Earth System Sciences with minor amendments due to a change in referencing style and reproduced under the Creative Commons Attribution licence.

Chapter 6: Social sensing of heatwaves in cities

I acknowledge some of the work included in the section titled “Previous work” in Chapter 6 was taken from the publication “Social Sensing of Heatwaves” published in the journal Sensors in 2021 (Young *et al.*, 2021) in which I worked with James Young[3], Rudy Arthur[3] and Hywel T. P. Williams[3].

For this paper I:

- Assisted with the experimental design and data curation stages.
- Advised on the interpretation and visualisation of results.
- Contributed to the first draft of the manuscript, and advised on edits in response to comments from co-authors and reviewers.

The text relating to content from this journal article and printed in this thesis is paraphrased by myself. Figures from this journal article and printed in this thesis are as published in Sensors and reproduced under the Creative Commons Attribution licence.

[3] College of Engineering, Maths and Physical Sciences, University of Exeter, Exeter EX4 4SB, UK

Chapter 1 - Introduction

Extreme weather events, such as heavy rainfall, strong winds and extreme temperatures, are becoming more frequent and impacting more communities across the world. The Intergovernmental Panel on Climate Change (IPCC) report from 2014 (IPCC, 2014) suggested that as a result of climate change these types of weather events are likely to increase in their frequency and intensity in the years to come. According to data collected by the United Nations Office for Disaster Risk Reduction (UNISDR) and the Centre for Research on the Epidemiology of Disasters (CRED) the overwhelming majority (90%) of disasters in the last twenty years has been as a result of floods, storms, heatwaves and other weather-related events (UNISDR and CRED, 2015). Therefore, it is important that a good understanding of the social impacts of extreme weather are understood and that the risks are forecast well by National Meteorological and Hydrological Services (NMHS).

Weather forecasting has made great strides in recent years with NMHS using bigger and more complex computing power, improved observations and models, and better data assimilation techniques to understand and predict weather patterns across the world (Alley *et al.*, 2019). However, while weather forecasts focus on the meteorological conditions (e.g. precipitation, temperature, wind speed), which are measured using traditional observational methods, the human and material impacts of the weather (i.e. flooding, landslides, damage, disruption, etc) remain difficult to quantify and to assess. The World Meteorological Organisation (WMO) recommend that NMHS across the world should move towards multi-hazard impact-based forecast and warning services – where weather forecasts and warnings focus on not just the risk of certain meteorological conditions, but also on the impacts that such conditions may bring (Campbell *et al.*, 2018; Taylor, 2018). Impacts are defined as the negative outcomes of an event and "primarily refers to a loss of life and injuries, damage to the environment, infrastructure, and private property, often followed by secondary effects like psychological trauma, or disruption of workflow and traffic" (Casteel, 2016; Kox *et al.*, 2018). The fundamental distinction between a general weather warning and an impact-based warning is that the impact of the weather drives the messaging, rather than the weather itself. Impact-based weather warnings include information on the potential social, economic and environmental hazard impacts. They will also include information about location and timing of such impacts, considering the vulnerability of the area affected (i.e. a more densely populated area versus a rural environment) (WMO, 2015, 2021a). Impact-based forecasts and warnings are already changing the way that organisations and individuals respond to weather and climate events across the globe. They enable them to make decisions and take early action before disasters, which minimises the socio-economic costs of weather and climate hazards, and ensures sufficient resources and supplies are in place. This saves lives and protects property and livelihoods (climatecentre.org, 2020).

This shift in forecasting approach also requires a change in forecast verification and evaluation techniques that will now need to assess both the actual weather hazards and the impacts experienced. Consequently, this presents a need for data on the social impacts of weather, which lie beyond traditional meteorological observations, and is therefore a new challenge for NMHS across the world. Verification and evaluation of these forecasts requires reliable impact data which provides details of how and where people and property were affected. Information on social and economic impacts in the wake of an extreme weather event is typically not as easily accessible as traditional meteorological observations, not widely dispersed, not in the correct format and is often only available with a significant time lag after a weather event has occurred (Vieweg *et al.*, 2014). The majority of NMHSs find it challenging to access impact information and do not have processes in place to systematically collect impacts in terms of an in-house database (Kaltenberger *et al.*, 2020). Furthermore, the process of impact data collection, evaluation and validation against weather models, forecasts and warnings has been found to be manual, complex and time consuming (Hemingway & Robbins, 2020). A study conducted by researchers in New Zealand found that some of the challenges of verifying impact-based forecasts included questions over the audience for which such forecasts should be presented; the potential for conflicting messages; conflicts with roles, responsibilities and increased burden on agencies; and finally verification of warnings based on impacts with a lack of impact data (Potter *et al.*, 2021). This presents an opportunity for exploring other data types that focus on the social effect of weather, which could be accessed in an automated or semi-automated way, and move beyond traditional meteorological observations of hazards experienced. In addition, having information on impacts more readily available close to the time of a weather event occurring would help to improve situational awareness about where and how impacts of a weather event are being experienced.

1.1 Weather impact data

Sources of impact data relating to weather events include news reports, social media, citizen data collated by government or official agencies, insurance data and eye-witness accounts. However, this data can be hard to access or is not always available until a significant time after a weather event has occurred (Hemingway & Robbins, 2020). Accessing appropriate impact data easily and in a timely manner is a major challenge for NMHSs implementing impact-based forecasts and warnings (Potter *et al.*, 2021; Harrison *et al.*, 2022).

Databases containing impact information already exist or are being developed by groups using a citizen science approach. The European Severe Storms Laboratory (ESSL) have developed the European Severe Weather Database (ESWD) which collects and provides detailed and quality controlled information about the impacts of storm events over Europe via an interactive mapping tool (Dotzek *et al.*, 2008). The ESWD contains information about when and where convective storm events took place as well as details of damage and disruption caused as a result. However, this database is limited to certain types of weather events, geographic locations in Europe and relies heavily on third parties to manually input information into the database. While a very credible source of information, it is also not available in real-time due to rigorous quality controls and may have gaps for smaller, less impactful events. In the USA, a similar database to ESWD, the NOAA Storm Events Database (NOAA, 2021), documents the occurrence of storms and other significant weather events in the USA causing significant social impacts, including rare/unusual weather phenomena which generate media attention. While a highly credible database due to the depth of information available, robust data collection techniques and processing procedures, data is often not available in the database for some months after an event has occurred. There are also databases such as NatCatSERVICE which records insurance loss (material or human) as a result of natural catastrophes (Climate ADAPT, 2004). Again, this database is curated using manual inputs from third parties with a focus on impacts that generated insurance claims. Therefore, this database may not include lower impact events, or events which occurred in less affluent locations in the world. The Natural Hazards Assessment Network (NATHAN) (Munich RE, 2021) is a subset of NatCatSERVICE for the insurance industry which includes global hazard data that has been systematically recorded from a range of sources including insurance losses and news media by Munich Re for the last 40 years. A further example is the EM-DAT International Disaster Database which includes impact information collated from a range of sources globally and includes “all” disaster events resulting in human impacts from 1900 to the present day (CRED, 2009). This database is not just limited to severe weather events, however, it does focus on the most high-impact disaster events and therefore, like the NatCatSERVICE, is likely to not include the more localised, lower impact weather events from which impact information for forecast warning and validation is still required and is also not available in real-time for situational awareness. While each of these databases has its advantages in terms of the impact data available, they lack consistency in terms of spatial and temporal coverage, unit of analysis, loss information, or other parameters.

Studies which have explored the social impacts of previous weather events (mainly flooding events) have provided information about the highest risk areas and times of year for specific hazards, impact trends as well as the extent of economic losses. Impact information sources used in these studies include: news media sources (Llasat *et al.*, 2009); insurance loss data (Changnon, 2003; Crompton & McAneney, 2008); vulnerability and exposure data to estimate the socio-economic impacts of weather events on a particular area or country (e.g. Italy (Farinosi *et al.*, 2012), India (Raghavan & Rajesh, 2003) and USA (Pielke Jr. *et al.*, 2008)); existing impact databases such as EM-DAT and NATHAN (Barredo, 2009); or a range of these sources (Lastoria *et al.*, 2006; Llasat *et al.*, 2010; Doocy *et al.*, 2013). These studies have identified a number of useful sources of impact data sources which can be used *after* a weather event has occurred for the collation of impact information. However, all of these studies required a substantial amount of manual effort to pull the information together and there is also a significant time lag after an event has occurred before the information sources become available.

The need to address the collation of impact information more systematically and closer to the time of an event has also been explored. Papagiannaki et al. (2013)

developed a database containing event information including spatial, temporal and community impact following high impact weather events in Greece from 2001-2011. Robbins and Titley (2018) outline an impact-based evaluation approach to identify how well Met Office Global Hazard Map forecasts relate to community impacts. In this study, the authors created a global impact database which includes a collation of impact information during extreme rainfall events across the world. Using a collection of different news and media sources they run semi-automated periodic keyword searches to extract relevant news articles and then manually review and categorise this information to understand the severity of impacts during the rainfall event. Events are categorised on an impact severity scale based on the number of people affected by the event, fatalities and damage caused. This approach provides a comprehensive database of past weather events and their impacts, however the intensive manual curation of information from news and other media sources is time consuming and may lack consistency in terms of impact detection.

Looking at existing work on the collation of impact data has highlighted the complexity of collating impact information. None of the approaches examined would be suitable for use in situational awareness or real-time verification of weather forecasts and warnings as the event unfolds, which NMHSs are also interested in. In a survey carried out with individuals in organisations providing weather hazard forecast or warning information, Harrison *et al.* (2021) found that NMHSs generally use social media for situational awareness and real-time verification during weather events. Social media can be used to identify potential 'hot spots' of weather impacts as an event unfolds. However, while social media is a rich source of information, it can be difficult to verify, can include a lot of irrelevant detail and risks overlooking or missing those impacted by a weather event who are not on social media. Therefore, the outcomes of this survey concluded that social media data for impact collation is incredibly useful, however it should be used comparatively with other data.

What is notable from looking at the impact information currently available and being collated by other organisations is that all data sources and databases rely to a large extent on manual input and filtering for relevant information, collecting information from a range of news and media sources as well as third parties to interact with the database and input data. This presents an opportunity to develop a more automated approach for collecting weather impact information both in real-time for situational awareness and for post event forecast validation and evaluation.

1.2 News articles

The drama and disruption caused by severe weather events are of significant human interest and therefore often reported by newspapers, particularly for events with the most severe impacts. The digitisation of news media sources means that this type of information is readily accessible from a wide range of online sources for the verification of weather forecasts and the collation of impact information. For example, FloodList (FloodList, 2021) provides a platform which collates relevant news articles from a range of news media sources relating to flooding events worldwide.

Some research which uses news articles to derive impact information has already been undertaken. For example, news articles have been successfully used in the development and expansion of the BGS National Landslide Dataset (Taylor *et al.*, 2015), research into observed impact databases for Scotland (Gunawan & Aldridge, 2018), development of a community impacts database for rainfall events worldwide (Robbins & Titley, 2018) and extraction of impact data from news media sources to assess flood impacts (Escobar *et al.*, 2016). While some of this research presents semi-automated methods for searching for relevant news articles, the acquisition of impact information from news media sources still relies on the time-consuming manual effort of reading each report to extract the relevant details.

Additionally, news reporting can be subject to biases and events may only be reported if considered 'newsworthy' enough to warrant the news outlet covering the story (Harcup & O'Neill, 2001). This bias towards 'newsworthiness' could lead to lower impact weather events, which may still lead to social impacts, not being reported. Another issue with using news articles to derive impact information is that the specific details of time, location, and who/what was affected can be missing or be reported ambiguously (Robbins & Titley, 2018). It can also be challenging to geolocate news reports of hazard impacts as it relies on mention of place names in the news report, which will, for obvious reasons, not contain precise coordinates of the location of the impacts. Therefore, news articles have their limitations in terms of their use in situational awareness and verification of forecasts during weather events.

1.3 Social Media

Another significant source of social impact data that could be explored for the purpose of impact-based forecast verification is social media data. Social media is an internet-based form of communication which allows users to have conversations, share information and create web content. There are many forms of social media ranging from blogs, micro-blogs and social networking sites, to podcasts, photo-sharing and instant messaging. Users engage with social media via computers, tablets or smartphone devices. Social media has increased in use across the world since smartphones became available in 2006, with more than 4.5 billion users in 2021, amounting to more than half the world's population (DataReportal, 2021). It is used in almost all countries across the world. For example, in 2021 the Philippines had the highest social media usage rate in the world, closely followed by Colombia. The most popular social media platforms across the world in 2021 include YouTube[4], Facebook[5], Twitter[6], Instagram[7] and TikTok[8] (Digital Marketing Institute, 2021). Social media users write posts about what is happening around them in near real-time, therefore this makes it a good source of impact information for situational awareness and real-time verification, as well as post-event analysis.

Social media began as predominantly a platform for socialisation and congregation. However, user generated content is now far ranging and includes information posted by businesses, news media sources, official bodies, as well as individuals (Kapoor *et al.*, 2017). Therefore, it is subject to a number of challenges and biases which should be recognised when using social media as a data source. While social media has been around in various formats for many years, in its current form, with such a huge user base, it is still a relatively new type of communication and information source. It is rapidly changing data source with new features (e.g. the addition of hashtags, like buttons, etc) being added all the time. This changes the way in which users might interact with a social media

[4] https://www.youtube.com (Accessed: 17 March 2022)
[5] https://www.facebook.com (Accessed: 17 March 2022)
[6] https://twitter.com (Accessed: 17 March 2022)
[7] https://www.instagram.com (Accessed: 17 March 2022)
[8] https://www.tiktok.com (Accessed: 17 March 2022)

platform and therefore may change the way that information is presented or spread on social media (Weller, 2015). Additionally, one major challenge which has come to light in recent times is the spread of misinformation on social media. This was made particularly evident in recent events such as Covid-19, Australian bushfires and US elections where the spread of false information on social media became news in itself (Muhammed & Mathew, 2022). Social media users are also greatly influenced by the behaviour of others on social media (Jain, 2018). Therefore, an event, such as a severe storm, which gets a lot of media attention, and therefore discussed a lot on social media, may lead others to be more likely to post about it.

While these challenges and biases in social media exists, it remains an invaluable trove of information and is therefore worthy of exploration as an information source during severe weather events.

1.3.1 Social Sensing

'Social sensing' utilising social media has been widely used in knowledge discovery in fields relating to public health, human behaviour, social influence and market analysis (Wang *et al.*, 2016). Social sensing broadly refers to a set of sensing and data collection models where data are collected from humans or devices on their behalf (Wang *et al.*, 2015). The social sensing approach has been developed in recent years to detect and analyse real-world events. In the case of social media, when a user publicly posts an item to a social media platform, they are providing a piece of sensory data. Therefore, each individual posting on social media plays the role of a sensor. When a large volume of social data is categorised and spatio-temporally tagged, social sensing provides an observatory for human behaviour (Liu *et al.*, 2015). Therefore, grouping relevant social media posts by topic or location may be useful in developing an understanding of a range of issues, including weather event detection and weather impacts. Social sensing has already proved successful in the detection of wildfires (Boulton *et al.*, 2016), floods (Arthur *et al.*, 2018), pollen/hayfever (Cowie *et al.*, 2018), named storms (Spruce *et al.*, 2020, Chapter 4), wind (Weaver *et al.*, 2021), global rainfall (Spruce *et al.*, 2021, Chapter 5) and heatwaves (Young *et al.*, 2021).

1.3.2 Types of Social Media

Social media is a rich source of information about where and when people are affected by events happening around them, as users can easily share their thoughts, feelings and insights instantaneously. One of the most useful types of social media in terms of finding out information about how and where people are being affected by events is microblogging on sites, such as Facebook, Twitter and Instagram. Twitter is a micro-blogging site where users post short messages publicly. Data is made available to developers and researchers via the Twitter Developer API[9]. From 2021, Twitter have also released a new 'academic research product track' allowing researchers easy access to the whole historical dataset of tweets going back to 2006, making Twitter the most easily accessible source of social media data for researchers (Ahmed, 2021). Users' posts are also all visible via the Twitter website, contributing to an overall global 'feed' of information and discussion. However, user privacy rules can make it difficult for researchers to access data on other microblogging sites. For example, though Facebook does enter into some research collaborations (e.g. https://dataforgood.fb.com/research/) it does not have an open API like Twitter, making it difficult to obtain Facebook data. Also, unlike Twitter, rather than one continuous global 'feed' of information, Facebook has numerous groups and personal pages with differing privacy settings, which can also make accessing information relevant to a particular event difficult. It is generally not possible to automatically access Facebook and Instagram data for routine monitoring, and is in fact disallowed explicitly by Facebook (https://www.facebook.com/robots.txt). Instagram and WhatsApp are both owned by the Facebook group and subject to similar privacy constraints.

1.3.3 Twitter

With the exception of China, where Twitter has been officially blocked in the country since 2009 (Branigan, 2009), Twitter is a globally accessible social media platform and is very popular with researchers as a source of social information. Data can be retrieved from the Twitter Developer API using keywords or 'hashtag' references which relate to specific topics or events, or by selecting a geographic bounding box of coordinates from which user posts originate. However, all Twitter posts (tweets) that meet these search criteria will be returned, regardless of

[9] https://developer.twitter.com/en/products/twitter-api (Accessed: 17 March 2022)

relevance to a particular event or not. Therefore, suitable algorithms to filter the data are required to ensure only relevant information is taken forward for analysis (Spence *et al.*, 2016).

Another challenge with working with Twitter data is locating the user and/or location to which the post relates. A limitation of using tweets for spatial analysis is that at present only 1-2% of tweets carry a GPS location or specific location coordinates (Dredze *et al.*, 2013). Therefore, statistical methods also need to be employed to infer the place of origin of the tweet or location of an event being discussed. In addition, the specific content of social media posts can be categorised and further analysed to better understand the type of information being shared or emotional response to natural hazard events.

Access to Twitter data using the Developer API has been made available since 2006, when the Twitter platform first launched. There are many free options for developers to access tweet data, which require authenticated access tokens which researchers can apply for. However, prior to 2021, most free options for accessing tweet data were rate limited (e.g. at the time of writing, for the Standard v1.1 API search tweets function, which allows the download of tweets in real-time, only up to 1% of all tweets at any point in time or 180 requests per user can be made available). The Standard v1.1 API also provides historical access to tweets for only up to 7 days in the past. Therefore, access to tweets by researchers for specific time periods in the past required either tweets to be downloaded in real-time during the period of interest, or to be purchased from Twitter retrospectively for significant sums of money. In response to the increasing use of tweets by researchers, Twitter launched their API v2 Academic Research access[10] in 2021. This provides academic researchers with access to the entire historical dataset of tweets going back to 2006 when the platform first launched. This has a monthly cap of 10 million tweets, however, still allows tweets to be queried based on specific keywords, language and other attributes and provides access to all historical tweets. An application must be made to Twitter and approved to access this version of the API. This provides researchers now and in the future with a much more accessible option to Twitter data.

[10] https://developer.twitter.com/en/products/twitter-api/academic-research (accessed: 1st July 2022)

In this thesis, the studies outlined in Chapters 4 and 5 were conducted before Twitter launched their academic research access API option, therefore tweets were downloaded using the Standard v1.1 API. The study in Chapter 6 was carried out after the academic access option was made available. Therefore, as the tweets used in the study had not already been downloaded in real-time, the academic research access API option was used so that a historical dataset of tweets could be accessed without cost to the researcher.

1.4 Use of social media for monitoring weather impacts

Social media has already been found by to be useful to NMHSs and emergency management agencies for building situational awareness quickly, real-time verification of forecasts and updating warnings (Harrison & Johnson, 2017; Harrison *et al.*, 2021). There are a number of approaches which have already been taken to understanding the social impacts of natural hazards using social media. Previous literature review studies which have explored the landscape of existing research into the use of social media during natural hazard events have found Twitter to be the main social media platform researchers use (Reuter & Kaufhold, 2018; Reuter *et al.*, 2018; Zhang *et al.*, 2019). The first stage in all studies is to find only those social media posts of relevance to the natural hazard event being examined. The second stage is to categorise the 'relevant' information so that the specific impacts both temporally and spatially can be determined. Wang & Ye (2018) suggest there are roughly three types of categorisation for social media posts: location inference, topic-based classification, and sentiment-based classification. Location inference of social media posts provides a spatial profile of when and where the impacts of weather events are being experienced (Schulz *et al.*, 2013); topic-based classification provides detail of what the impacts are (Alam *et al.*, 2018b); and sentiment-based classification gives an indication of public reaction to an event (Harrison & Johnson, 2017; Harrison *et al.*, 2021). Each of these categorisations is useful in isolation for providing situational awareness and impact data, however combining these classifications would provide a much more holistic overview of when, where and how individuals are being impacted when a significant weather event occurs.

Chapter 2 will discuss this in more detail, providing a review of previous studies which have explored the use of social media as a source of impact data during and following weather events.

1.5 Research questions

Having explored the issues surrounding the collation of impact data to support impact-based weather forecasts and warnings, there is a recognisable need to explore the usefulness of social media data as a source of impact information for NMHSs. Having considered the impact data collection already in use by NMHSs and other organisations, social media is most certainly a potential source of impact information that should be investigated. However, most existing impact databases rely on manual data collection and information extraction from a range of online news and media sources, as well as reliance on third parties to provide the information. Additionally, as will be discussed further in Chapter 2, previous studies examining social media during weather events to date tend to focus on the detection of events, rather than categorisation of impacts. Therefore, it is clear that in the context of impact-based forecasts and warnings, the need for impact data to support situational awareness and forecast verification requires further exploration.

Therefore, in this study the following research questions will be explored:

> *RQ1. How useful is social media as a source of impact information during and after weather events?*
>
> *RQ2. What tools and methods can be successfully applied to extract relevant social media data during weather events?*
>
> *RQ3. What are the limitations of social media as an impact data source?*

The importance of effective communication of impact-based warnings using social media and the response of individuals and organisations to them is also an area of significant interest and importance to NMHSs and emergency management organisations (e.g. to better understand crisis communication throughout the crisis lifecycle of a major hurricane event (Stewart & Wilson, 2016) and exploring the types of information communicated by state emergency management accounts to better understand the flow of risk communication during a crisis (Rainear *et al.*, 2018a)). However, exploring the effectiveness of disaster and risk communication on social media is beyond the scope of this thesis, which will be focused on the curation of impact data from social media, rather than the use of social media during impactful weather events.

1.6 Thesis plan

This thesis will provide information on previous studies and original experimental work to explore the use of social media as a source of impact information during and following weather events. Chapter 2 discusses a systematic literature of studies which have already explored the use of social media as a source of impact data during weather events. Chapter 3 will discuss emerging themes from the literature review in Chapter 2, ethical considerations for using social media as a data source and revisit the research questions this thesis will try to address. Chapter 4 and Chapter 5 describe two experimental studies that develop methods and tools for extracting relevant information from social media during particular weather events (UK/Ireland named storms and global rainfall). These chapters are based on published articles in peer-reviewed literature (Spruce *et al.*, 2020, 2021). Chapter 6 explores the feasibility of applying these methods for heatwave events, and more specifically in three European cities. The final Chapter 7 discusses the effectiveness of social sensing as a tool for the extraction of relevant impact information from social media for impact-based warnings and services, limitations for using social media as a data source, and the next steps for future research.

Chapter 2 - Using social media to detect severe weather events and evaluate impacts: a systematic literature review

2.1 Introduction

The impacts of extreme weather on society both socially and economically are well known (IPCC, 2014). As a result, there has been a shift by weather forecasting agencies across the world towards providing a more impact-based forecast that focuses on both the likelihood of weather occurring as well as the impacts to society (WMO, 2015). However, the evaluation of impact-based forecasts remains difficult to assess and quantify due to the lack of definitive sources of information about impacts experienced beyond more traditional meteorological observations (Potter *et al.*, 2021).

Social media has become an information dissemination and communication tool that is now part of everyday life. It is used for communication between individuals, communication between agencies and the public, and for reporting of events as and when they occur. This makes it an ideal source of information on social impacts during a significant event, such as an extreme weather event. ‘Social sensing’ is the systematic analysis of unsolicited social media data to observe and characterise real-world events. It refers to using a set of sensing and data collection techniques where data originates from humans or their personal devices (Wang *et al.*, 2015). In the context of this literature review, social sensing refers to the analysis of user generated social media content.

The increase in the use of social media over the last 10 years (Clement, 2020a) has led to an increase in the amount of research undertaken which uses social media as a data source. This holds true for research which uses social media to better understand the impacts of weather events and as a result, there have been a number of studies conducted in recent years that explore the use of social media for detecting and better understanding the impacts of severe weather events (e.g. Arthur *et al.*, 2018; Li *et al.*, 2019; Sit *et al.*, 2019; de Bruijn *et al.*, 2019). However, studies originate from many different academic disciplines across both the physical and social sciences, as well as from agencies and policy makers interested in better understanding the impacts of extreme weather. This makes collating information about the breadth and type of research in this area difficult. Therefore, the aim of this literature review is to examine a comprehensive list of relevant literature related to the use of social media in the study of social impacts as a result of weather events. The review will exclude studies where the focus of the research is not related to event impacts (e.g. studies which focus on the effectiveness of communications during a natural disaster). By bringing together these studies in a systematic review the aim will be to find out the following:

Q1. What research has already been undertaken that uses social media to better understand the impacts of weather events?

Q2. Within this research, which weather types have been explored?

Q3. What social media platforms have researchers explored as a source of data on the social impacts of weather events?

Q4. What methods and tools have been used to analyse social media data for the purpose of understanding the impacts of weather?

Q5. Based on a review of the literature, what are the challenges and future direction for the use of social media in understanding weather impacts?

2.2 Methods

To create a comprehensive list of relevant literature, a systematic literature search was carried out. Literature in this area originates from many different disciplines, therefore a systematic approach will ensure that a full appraisal of existing literature can be carried out, avoiding any bias towards any particular subject area. This type of review requires the researcher to apply a methodical approach to their literature search which includes the following: search/inclusion criteria, literature identification, screening for inclusion, quality and eligibility assessment, data extraction and analysis/synthetisation of results (Xiao & Watson, 2017). A systematic search and appraisal of relevant literature was determined to be the best type of approach to take for searching the literature on the social sensing of weather. This aim of this approach is to identify the research already undertaken in this area; what remains unknown; limitations of the approach; and recommendations for future research. An accompanying table of studies will also be produced to support the narrative. The approach followed in this review is in line with the systematic search and review described by Grant and Booth (2009).

For this literature review, all research published between 2010 and 2020 was searched using three literature searching platforms, using pre-defined search terms included in the study title. 2010 was used as a starting year for the search because the first major papers in the area of using social media to enhance situational awareness during natural disasters were published in 2010. For example, Vieweg *et al.* (2010) analyse social media (Twitter) posts during the 2009 Red River floods and Oklahoma grassfires, providing methods for extracting information useful for situational awareness from social media; and Sakaki *et al.* (2010) analyse Twitter posts during earthquake events in Japan, finding that they could detect earthquakes using social media before official reports from the Japan Meteorological Agency. These papers precipitated further research into using social media for event detection or as a source of impact information during natural hazards.

2.2.1 Searching the literature

The first step in the review was to conduct a title-based search of the Google Scholar[11] database. The titles of selected papers should contain at least one keyword from both the social media/social sensing category and the weather event/impact category. It is recognised that using only weather hazards in the title search may exclude papers which consider natural hazards more generally. However, not all natural hazards are related to weather, e.g. earthquakes, wildfires, etc. As this review is focused on the social sensing of weather events, using weather hazard keywords in the search string will ensure only studies focused on weather will be returned in the search.

Social media platforms included in the search string were taken from the currently most used social media platforms published by Statista (Statista, 2021). To account for social media platforms which have been in use over the last 10 years, but which have fallen in popularity, a further search was carried out to identify

[11] https://scholar.google.com (Accessed: 17 March 2022)

platforms which are now less popular, a comprehensive list of these was found using Global Change Data Lab (2019). These social media platforms were then also added to the search string.

Weather types to include in the search string were based on meteorological conditions most likely to result in human or material impacts included as hazards by the Natural Hazards Partnership[12]. This returned 560 studies from the Google Scholar database using the following search criteria:

(“social media” OR “social sensing” OR twitter OR tweets OR facebook OR instagram OR whatsapp OR flickr OR weibo OR wechat OR youtube OR telegram)

AND

(weather OR flood OR floods OR flooding OR rain OR rainfall OR hurricane OR hurricanes OR tornado OR storm OR storms OR lightning OR ice OR snow OR precipitation OR wind OR landslide OR landslides OR heatwave OR heatwaves OR “heat wave” OR “heat waves” OR forecast OR forecasts)

The following social media platforms were also searched for in study titles but returned no/no relevant results: *TikTok, QQ, Douyin, QZone, Snapchat, Reddit, Kuaishou, Pinterest, Google+, Quora, Tumblr, MySpace, Hi5, Friendster*.

2.2.2 Screening for inclusion

The first stage in the screening process was to check the title of studies for relevance to social media and weather, as the search terms used also returned studies that did not relate to weather (e.g. analysis of ‘Twitter storms’ or use of social media to ‘forecast’ economic activity, to name a few). This left 417 studies from the Google Scholar database with a title appearing to be relevant to the use of social media to better understand the impacts of weather events.

To ensure that any important studies not included in the Google Scholar database were not excluded from the review, the same search criteria used to query the Google Scholar database was used to search both the Web of Science[13] and

[12] http://www.naturalhazardspartnership.org.uk/natural-hazards/ (Accessed: 17 March 2022)
[13] https://www.webofscience.com/wos/woscc/basic-search (Accessed: 17 March 2022)

Scopus[14] databases. After checking the results of these searches for studies that were not already included in the Google Scholar results, and also checking titles for relevance to the social sensing of weather, a further 20 studies were found in the Web of Science database and a further 21 studies found in the Scopus database. This left 458 studies to examine in the next step of the inclusion screening phase.

The second step in the screening process was to begin to apply exclusion criteria beyond the title search. Initially the type of publication was checked and the following categories removed: thesis/dissertation (as these are not likely to be peer reviewed), journal preprint, abstract only, poster. Publications that were not peer reviewed (e.g. newspaper/blog articles, papers posted online by researchers but not peer reviewed) were also excluded at this stage. This left 320 studies to be further checked for relevance.

2.2.3 Assessing quality

The next stage of filtering for inclusion in the review involved reading each paper's abstract to determine if the research was focused on assessing the impacts of weather using social media or not.

Inclusion criteria was therefore as follows:

1. The aim of the study was to understand the social impacts of weather event(s) using social media; or
2. The aim of the study focused on the detection of weather event(s) temporally and/or spatially using social media; or
3. The aim of the study focused on the analysis of or changes in social media activity during a weather event, with a focus on the information available using social media.

If the paper did not meet this inclusion criteria, then it was disregarded. For example, some papers returned in the search were more focused on the effectiveness of communication by agencies during a weather event or analysis of social media volume/content for purposes not related to the impacts of the weather event. These studies were therefore excluded from the final review.

[14] https://www.scopus.com (Accessed: 17 March 2022)

This final stage of quality assessment left 108 papers to review.

Each paper was then reviewed in turn to determine the purpose, methods used and findings from the research.

2.2.4 Limitations to scope

In terms of limitations to the systematic literature review method used, the exclusion of terms such as “natural disaster”, “disaster” and “natural hazard” in the title search of scholarly databases will have led to some relevant studies being excluded from the analysis. Initially it was decided to exclude these terms as many papers referring to ‘disasters’ focus on non-weather-related events, such as earthquakes, or consider disaster management more generally. However, despite this limitation in the search, this study has provided a good overview of the literature relating to the social sensing of weather events carried out to date.

2.3 Distribution of papers by year, platform, weather type and geography

To aid with addressing questions 1-3 in the aim of this literature review, an analysis of the number and type of papers is first discussed.

Figure 2.1 shows the number of papers published as journal articles or as conference proceedings found in this literature search that relate to the social sensing of the impacts of weather events. There has been a sharp increase in the number of papers published in this area, particularly in the last 3 years. There were no publications returned in the literature review search with a publication year of 2010 or 2011. This may be because, as previously noted, excluding terms relating to ‘natural hazards’ and ‘natural disasters’ from the title search may have excluded studies which explored the impacts of a weather hazard, but did not include the specific hazard in the title of the publication (e.g. Vieweg *et al.*, 2010; Imran *et al.*, 2013; Alam *et al.*, 2018a; Niles *et al.*, 2019). Also, studies that did not focus on a *weather* hazard were excluded (e.g. Sakaki *et al.*, 2010).

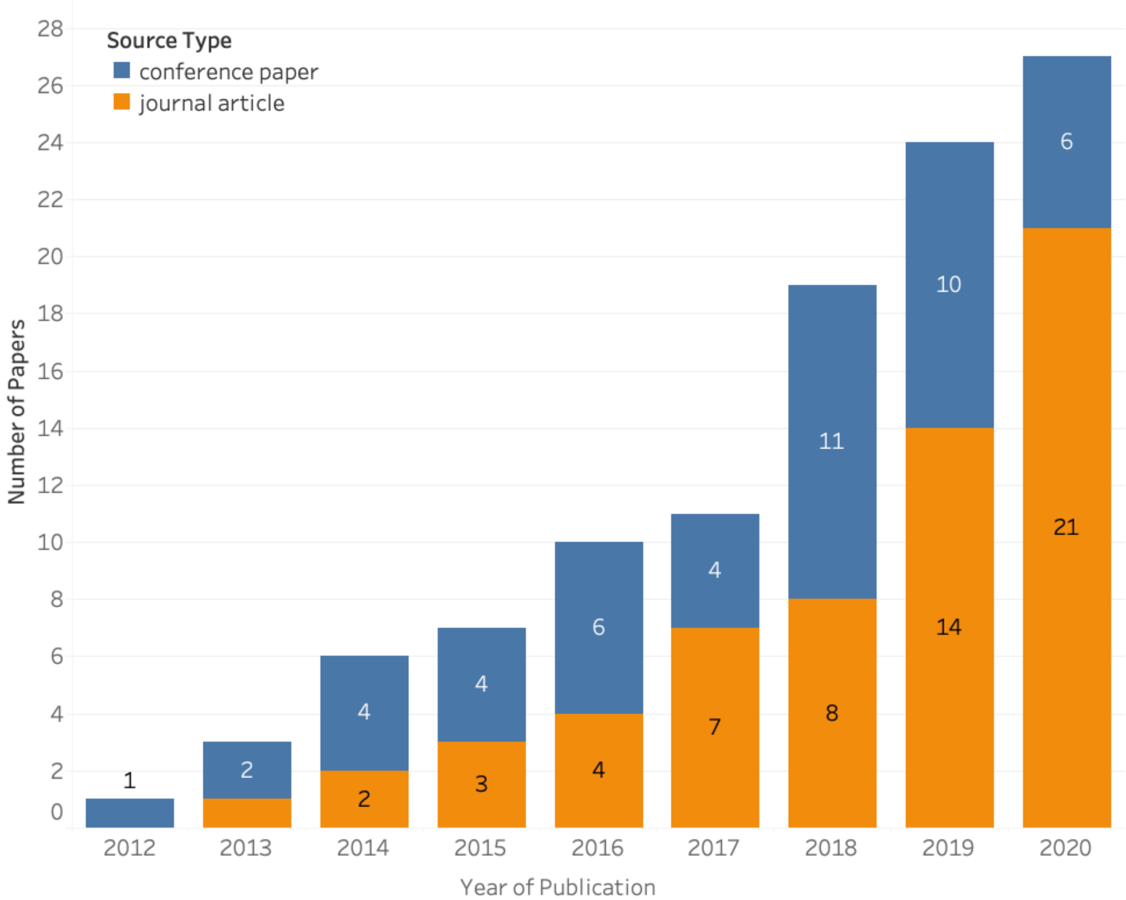

Figure 2.1 - Number of papers (published as journal article or conference proceedings) related to the social sensing of weather events from 2012-2020.

By examining the social media platform used in each publication we can see that the majority of research has used the social media platform Twitter as its data source. Figure 2.2 shows the number of papers split by the social media platform used in the research. Twitter is clearly the most used by researchers, with 82% of the papers examined in this literature review using Twitter as at least one data source (89 out of 108 papers). This does not come as a surprise as the availability of Twitter data through the Twitter Developer API makes it the most easily accessible data source to researchers. Sina Weibo (the Chinese social media platform) is the next most used social media platform, with 8% of papers in this literature review using data from this platform. As Twitter is not available for use in China, these 9 papers account for studies which focus on Chinese case studies. Other social media platforms used as data sources include Flickr, Instagram, Facebook and WhatsApp. These platforms are difficult for researchers to use as a data source due to rules around privacy and data being less easily accessible (Social Media Research Group, 2016). 'Other' accounts for other data sources, such as crowdsourcing applications that users interact with to provide data on weather events.

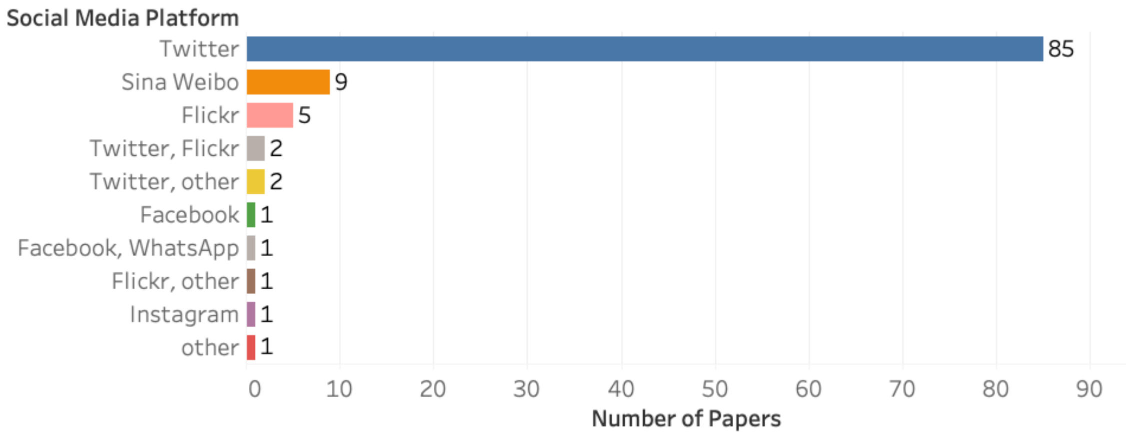

Figure 2.2 - Number of papers related to the social sensing of weather from 2012-2020 split by social media platform used in the research

Another finding from this literature review is the wide range of subject areas in which research on the social sensing of weather is published. Publication subject areas range from earth and planetary sciences to computer science and the social sciences. Figure 2.3 shows the number of papers in this literature review by subject area of the publication. The publication subject areas for each paper reviewed in this literature review was taken from Scimago journal and publication rank website (https://www.scimagojr.com/). Many publications resided in more than one subject area, therefore Figure 2.3 may count the same research publication in more than one subject area. Computer Science, Social Sciences and Engineering are the subject areas associated with the largest number of publications in this literature review. However, Earth and Planetary Sciences and Environmental Science also make up a significant proportion of subject areas in which research in this area is published. Many research publications in which papers included in this review reside are interdisciplinary in nature. 74% (80 out of 108 papers) were published in a research publication which had two or more subject areas. The most common combination of subject areas were computer science and engineering (25%), and earth & planetary sciences and social sciences (16%) (*data not shown*).

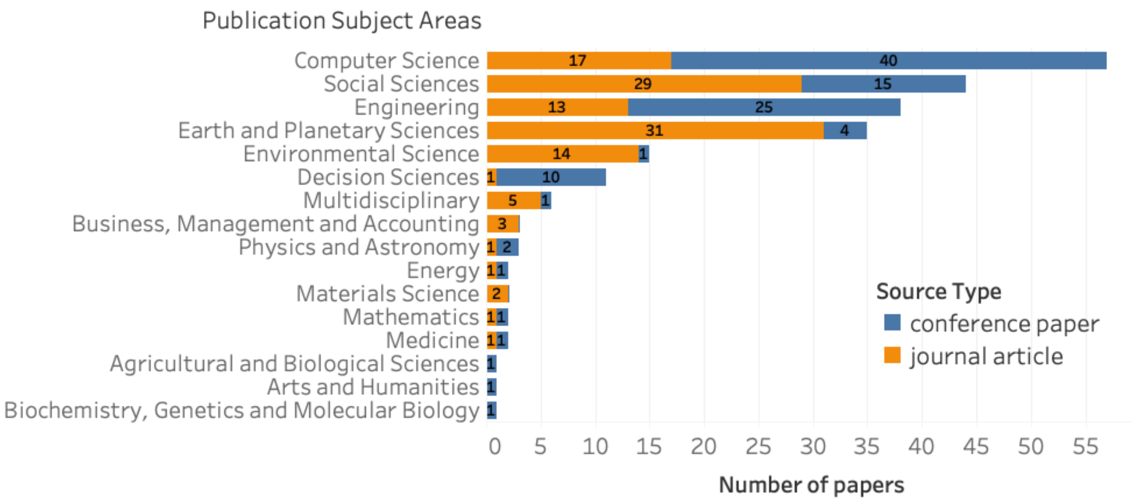

Figure 2.3 - Number of papers related to the social sensing of weather from 2012-2020 split by subject area of the research publication (determined from scimagojr.com) and type.

The type of weather event also appears to greatly influence researchers in the case studies they have used. Figure 2.4 shows the type of weather that has been the focus of papers in this literature review. The predominant weather impact type of interest to researchers appears to be flooding with 47% of papers focusing on the social sensing of flood events (51 out of 108 papers). Hurricanes are also of great interest to researchers with 29% of papers looking at the social sensing of hurricane events (31 out of 108 papers). The ‘mixed’ category was used for papers which look at multiple weather types or the impacts of weather more generally.

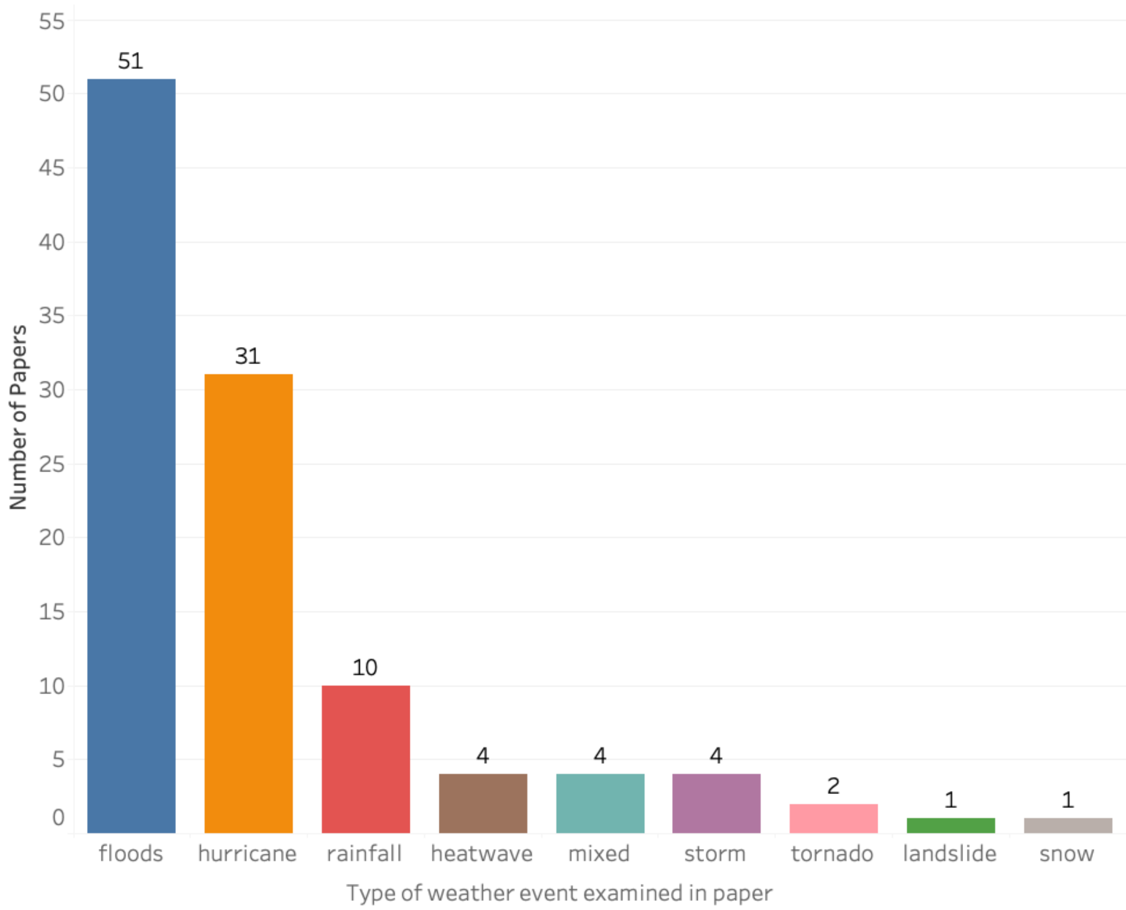

Figure 2.4 - Number of papers related to the social sensing of weather from 2011-2020 split by the type of weather examined in the research

Figure 2.5 shows which countries have been the location of interest in papers examined for this literature review. There were 7 papers (6%) where the country was not specified or where the research considered weather events across the world/with no specific location, rather than in one specific country. The country with the vast majority of research into the social sensing of weather events is USA (43% of papers), followed by China (8%) and the UK (7%). The number of papers related to each country are further divided by weather type examined in the study. 64% of papers relating to weather in the USA focus on hurricane events (30 out of 47 papers), which shows that the impacts of hurricane events are likely to be of most interest to researchers in the USA. However, many of these studies also focus on flooding as a result of hurricanes, which will be discussed later in this chapter. It can also be seen that the impact of flooding is the weather event of interest to most countries outside of the USA.

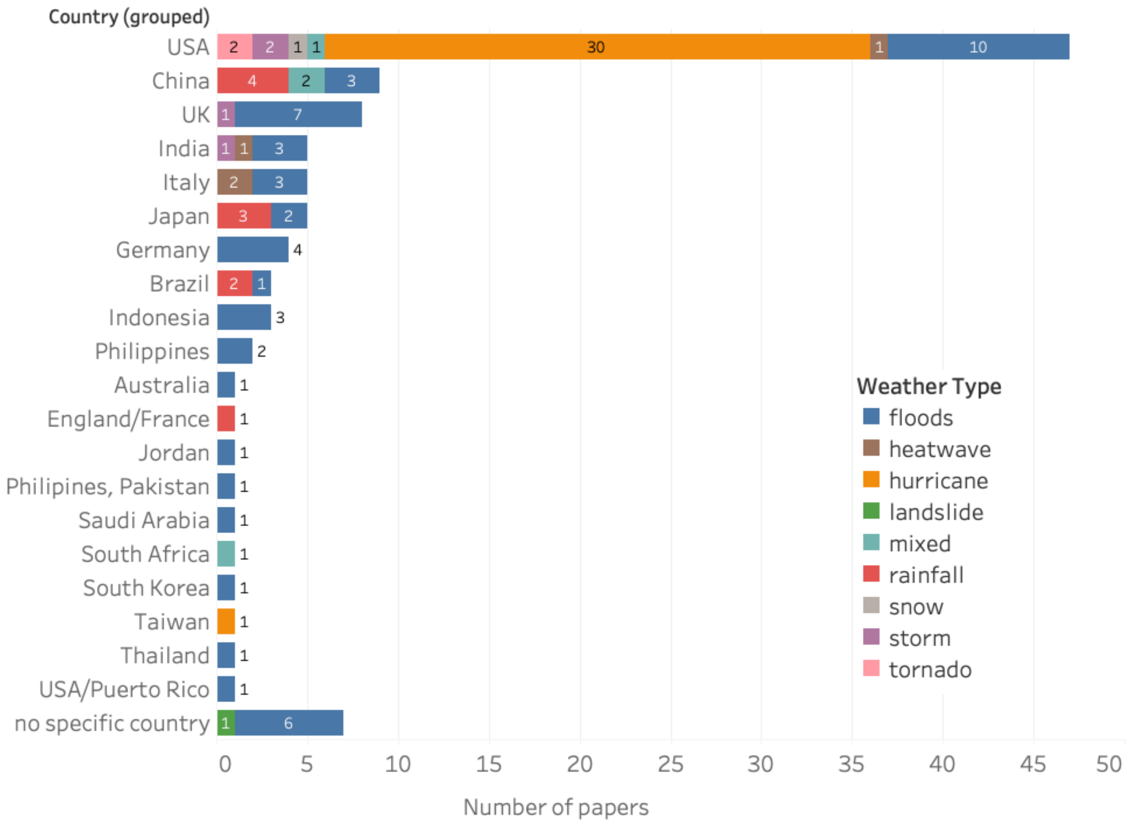

Figure 2.5 - Number of papers relating to the social sensing of weather from 2012-2020 split by country and weather type of interest.

2.4 Purpose and methodology of social sensing studies

Each paper was reviewed to determine its purpose and methods used. A critical review of all papers is included in the main text, split into two sections: purpose, aims and main ideas; and methods and tools. Table A.1 in the Appendix provides a summary of all papers reviewed including the social media platform, weather type and country of interest; main aims and ideas; methods and tools; advantages and disadvantages.

2.4.1 Purpose, aims and main ideas of studies

In terms of purpose, papers largely fell into one of the following three categories, although some studies may fall into more than one category:

1. Event Detection – main aim of study is spatial and/or temporal detection of weather events using social media (this includes papers focusing on inundation mapping as a result of flooding).
2. Impact Assessment – main aim of study is to improve situational awareness of the impacts of a weather hazard using information from social media content. This may include identifying topics of discussion,

extracting relevant social media posts, calculation of sentiment, or a more general analysis of volume/attributes/content of social media as a proxy for impact information.

3. Other – anything that does not fall into the above categories. This includes research focused on effects on health due to weather and how the public responds to weather and perception of risk.

There were 36 papers (33%) in the event detection category, 69 papers (64%) in the impact assessment category, and 3 papers (3%) in the other category.

Figure 2.6 provides the number of papers in each category, for each year of publication. The number of papers focused on impact assessment has increased since 2018.

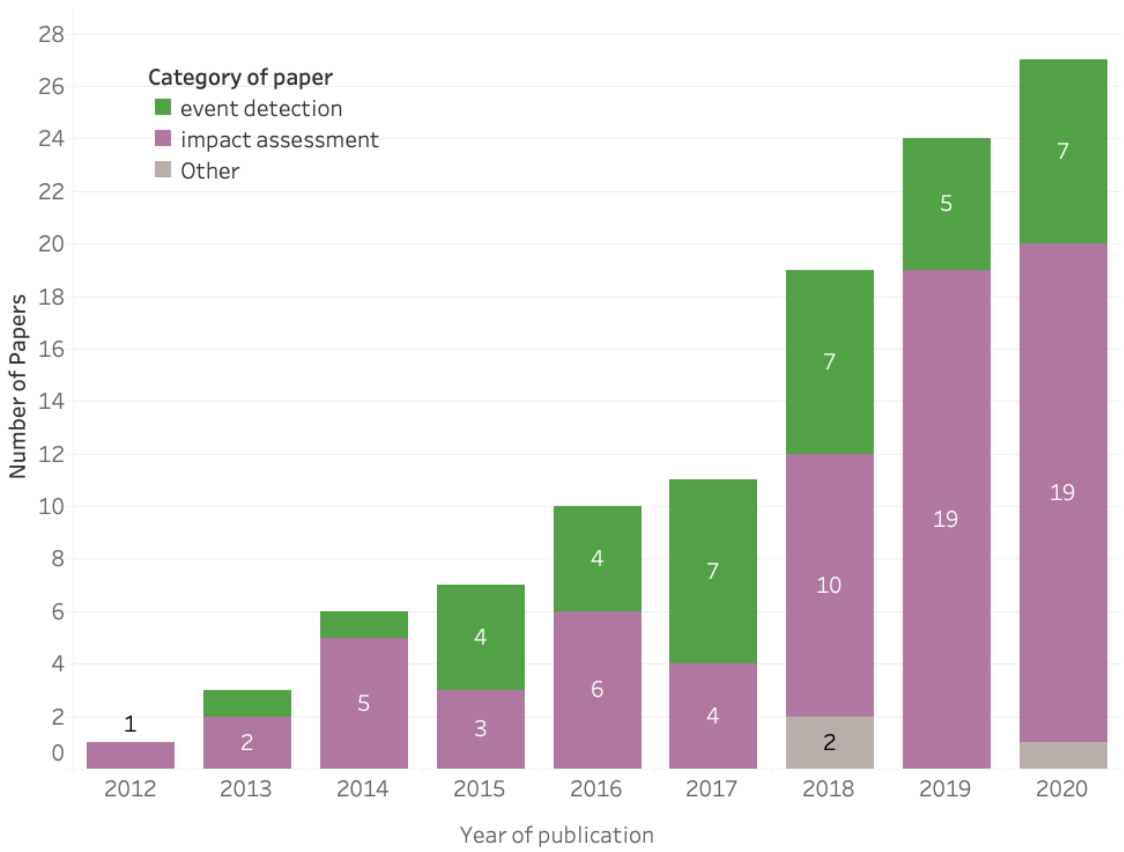

Figure 2.6 - Number of papers relating to the social sensing of weather from 2012-2020 split by category of paper.

Figure 2.7 shows the number of papers in each category, split by weather type of interest. The majority of papers in the event detection category were focused on event detection during flooding events.

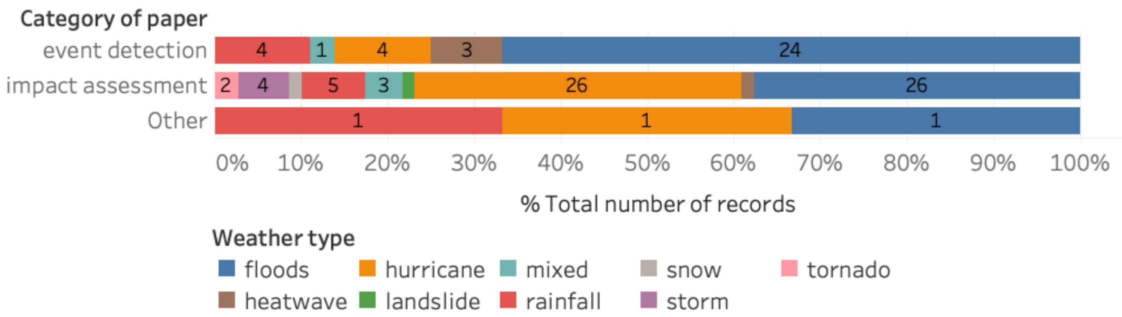

Figure 2.7 - Number of papers relating to the social sensing of weather from 2012-2020 split by category of paper and weather type of interest.

Event Detection

Using social media to determine when and where an event is taking place is helpful for meteorologists to understand the temporal and spatial impacts as a result of weather impacts. The majority of papers in this review which were focused on event detection related to the detection of flooding impacts. Some studies also focused on the development of a real-time application so that the impacts of weather can be understood at the time of an event, as well as retrospectively.

Feasibility studies

Most studies use social media data collected after a weather event has occurred to explore the feasibility of event detection in future events. For example, Crisci & Grasso (2013) use heat-related tweets to find areas most impacted during a heatwave; Saravanou *et al.* (2015) explore the use of Twitter to identify the location of flooding in the UK by comparing the location of flood-related tweets with locations of known floods; Cerutti *et al.* (2016) explore the spatial extent of flooding using the location of flood-related tweets; Tkachenko *et al.* (2017) analyse image tags on Flickr to find flood-related posts and explore its use an early warning indicator of flooding; Jitkajornwanich *et al.*, (2019) propose an early-flood warning system using flood-related tweets to build maps of flood activity; Farnaghi *et al.* (2020) develop a method to detect peaks in tweets activity as a proxy for the location of impacts during Hurricane Florence; and Wani *et al.* (2020) examine the location of flood-related tweet activity during 2018 Kerala floods in India.

Combining social media data with remote sensing or weather observation data is another event detection approach. For example, Jongman *et al.* (2015) combine satellite observations and flood-related tweets to detect flood events in the

Philippines and Pakistan; Lwin *et al.* (2015) compare Japanese Meteorological Agency (JMA) rainfall data and rainfall-related tweets during a rainfall event in Japan; and Restrepo-Estrada *et al.* (2018) use flood/rainfall-related tweets and official rainfall data to detect the spatial impacts of flooding.

The use of crowdsourcing data, in which people can volunteer specific weather impact information, alongside social media data was also explored by Wang *et al.* (2018). They use both flood-related tweets and crowdsourced data from the MyCoast app[15] to find flood information at the metropolitan scale.

Real-time application

Detecting weather events in real-time to aid situational awareness has also been explored using social media data from previous events. Arthur *et al.* (2018) detected floods in the UK using flood-related tweets. Their findings are compared with a database of actual flood events and good accuracy is found. Barker & Macleod (2019) also develop a method to find flood-related tweets in the UK using national flood warnings to identify known locations at risk of flooding and river level data. De Bruijn *et al.* (2019, 2020) apply a similar approach, creating a database of flood events across the world. Results are compared with an existing database of known flood events (NatCatSERVICE), finding good accuracy. Findings from these studies offer potential to develop real-time applications to detect flood events using Twitter data.

Other studies which discuss the development of real-time applications to aid in situational awareness during weather events include: Fitrianah *et al.* (2020), who develop a real-time application to show users flood-related tweets based on user inputted keywords; and Khaleq & Ra (2019) develop a real-time, cloud-based application to monitor impacts using social media data during hurricane events.

Flood inundation mapping

The use of social media to assist with flood inundation mapping has also been explored by a number of researchers. Fohringer *et al.* (2015) produce flood inundation maps using ‘flood’ images posted on Twitter and Flickr; Eilander *et al.* (2016) produce flood inundation probability maps using the location of flood-

[15] https://mycoast.org (Accessed: 17 March 2022)

related tweets; and Brouwer *et al.* (2017) also produce flood inundation maps using Twitter data.

Combining flood-related tweet activity with other remote sensing data has also been proposed to produce flood inundation maps. For example, Huang *et al.* (2018) produce flood inundation maps using flood-related tweets and other remote sensing data; Li *et al.* (2018) create flood inundation maps using flood-related tweets and stream gauge data; Rosser *et al.* (2017) estimate probability of flood inundation using remote sensing, images from Flickr and topographic data sources; and Scotti *et al.* (2020) combine satellite images, hydraulic models and flood-related Twitter posts to produce flood inundation maps.

Social media as a proxy for observations

Using social media data as a proxy for weather observations was explored by Butgereit (2014), finding good accuracy in detecting weather conditions when the weather is extreme or rapidly changes. Rainfall-related social media content has also been found to be a good a proxy for rainfall observations, which is particularly useful in areas of the world where there are limited meteorological observations or other remote sensing resources (de Vasconcelos *et al.*, 2016; Andrade *et al.*, 2017; Feng & Sester, 2017). Additionally, Sun *et al.* (2016) use flood-related images from Flickr to explore its use as a complementary data source for areas with limited remote sensing data.

During heatwaves, Jung & Uejio (2017) analyse tweets in US cities, finding a relationship between heat-related tweets and heat exposure metrics. Cecinati *et al.* (2019) also examine heatwave-related tweet activity as indicator of mortality rates in India as a result of excess heat.

Using images from social media, some studies have developed methods to use these to estimate flood water depth. For example, Chaudhary *et al.* (2019) use flood related images to estimate flood depth based on how far objects in the image are sunk into the water; and Pereira *et al.* (2020) estimate flood depth using flood-related images from Flickr.

Studies have also compared tweet activity with third party data sources as a method for exploring its use as an observational tool. For example, for monitoring the track of a hurricane, Yang *et al.* (2019b) assessed credibility and assigned an overall credibility score to tweets relating to Hurricane Harvey, finding the location of relevant tweets correlate to the progression of the hurricane. Owuor *et al.* (2020) also compare tweets with the track of Hurricane Dorian using the location of hurricane mentions on Twitter. They find a good match between peaks in hurricane-related tweet activity and the hurricane track. In the UK, Smith *et al.* (2017) compare flood-related tweet activity with locations of known floods, finding correlation between peaks in tweet volume and the extent and depth of the flood level.

Impact Assessment

As well as detecting when and where an impactful weather event may be taking place, researchers are also interested in the content of the tweets and what this can tell them about how people are being impacted by the weather. Accessing the social media content relating to a particular weather event, rather than event detection alone, can aid in situational awareness during a weather event. Research found in this review ranged from extracting the specific, relevant, social media content which can then be manually reviewed, to the automatic categorisation of tweets into types of impact.

Extracting social media content relating to an event

To aid situational awareness, one approach to using social media to obtain impact information is to extract the relevant tweet content relating to the event for further review. This includes extracting the relevant text and/or images, relating to reports of flooding (Oktafiani *et al.*, 2012; Herfort *et al.*, 2014b,a; Rossi *et al.*, 2018; Moumtzidou *et al.*, 2018; Rodavia *et al.*, 2018; Huang *et al.*, 2019b,a; Shi *et al.*, 2019; Jony *et al.*, 2019; Wang *et al.*, 2020a); damage assessment during floods (Brovelli *et al.*, 2014; Assis *et al.*, 2015); and creating a dataset of flood-related tweets in Arabic languages (Shannag & Hammo, 2019; Hamoui *et al.*, 2020). Accessing tweets relating to reports of landslides has also been explored by Musaev & Hou (2017).

Social media activity as an indicator of impact

Using social media activity as a measure for impact as a result of a weather event has also been explored by some authors. Preis *et al.* (2013) use the volume of hurricane-related images on Flickr as proxy for hurricane impact, finding a correlation between the number of photos and falling atmospheric pressure;

Grasso *et al.* (2017) explore heatwave impacts in urban areas using the volume of heatwave-related tweets; Fang *et al.* (2019) use rainfall-related Weibo posts to find impact information; and Spasenovic *et al.* (2019) explore the spatial distribution of hurricane-related tweets during Hurricane Michael to identify where impacts during the hurricane occurred. Grace (2020a) also analyses toponym usage in storm and non-storm related tweets to identify if particular patterns in toponyms are a likely indicator of impacts.

Topics of discussion

Exploring the type of impact information which may be available from social media was a focus for some studies. This included studies which explore the topics of discussion in social media posts and assign posts to categories of impact, as well as studies which provide a more general overview of the types of impact information which might be determined from social media. For example, Congjuico (2015) and Bhuvana & Arul Aram (2019) explore the type of impact information available on Facebook/Whatsapp community groups; Pourebrahim *et al.* (2019) explore the types of communication on Twitter during Hurricane Sandy to identify impacts; Yamada *et al.* (2019) explore types of information and patterns of discussion within Japanese language rainfall-related tweets during heavy rainfall event in Japan in 2018; Dalela *et al.* (2020) explore the categories of impact information available from storm-related tweets in storm events; and Liu *et al.* (2020) identify flood impacts using flood-related tweet content. All authors found that relevant social media content has the potential to provide impact information during each of the weather events analysed.

Topics of discussion on social media during a natural hazard event vary depending on the stage of the disaster (Imran *et al.*, 2016). Therefore, some studies examined how the topics of discussion on social media changes before, during and after a weather event, with the purpose of understanding how social media behaviour changes and the type of impact information during this time. For example, the type of discussion before, during and after Hurricane Sandy was explored by Lachlan *et al.* (2014), finding that many used Twitter as an emotional release during the worst stages of the disaster. Wang & Ye (2018b) also examine the topics of discussion before, during and after Hurricane Sandy in New York City, finding that topics on impacts to infrastructure and utilities remain a dominant topic of discussion throughout the event. Topics of discussion during the different stages of hurricane events were also examined by Alam *et al.* (2018b), Anam *et al.* (2019) and Sovacool *et al.* (2020). In all cases, authors found that topics of discussion relating to impacts as a result of the hurricane, such as damage and disruption, are more likely to be posted on social media when the most severe impacts are being experienced by users and therefore provide their methods to support its use as an impact information tool.

Changes in topics of discussion during other weather events have also been investigated. Wu *et al.* (2020) analyse rainfall-related Weibo tweets before, during and after the 2016 Hefei rainstorm in China to identify impacts from topics of discussion; Kankanamge *et al.* (2020) examine the change in topics of discussion during floods in Australia before, during and after the event, with aim of determining disaster severity using topics of discussion and sentiment; and Spruce *et al.* (2020) (Chapter 4) examine the changes in topics of discussion during Storm Brian in the UK in 2017. A more general analysis of topics of discussion during weather events which can be found using social media content relating to hurricanes (Dong *et al.*, 2013; Mukkamala & Beck, 2016; Chien *et al.*, 2017; Xin *et al.*, 2019; Vayansky *et al.*, 2019); storms (Grace, 2020b); floods (Nair *et al.*, 2017; Han & Wang, 2019b); tornadoes (Ukkusuri *et al.*, 2014; Halse *et al.*, 2018); and heavy rainfall (Wang *et al.*, 2016) has also been explored by a number of authors. In addition, Halse *et al.* (2018) compare tornado-related tweet topics of discussion with weather sensor data finding tweet activity correlates with wind speed.

Damage assessment

Focusing on identifying specific types of impact using social media data, Yuan & Liu (2018b) find damage-related tweet content during Hurricane Matthew, then in a further study combine these hurricane-related tweets and Unmanned Aerial Vehicle (UAV) data to conduct a rapid damage assessment during the hurricane (Yuan & Liu, 2018a). Furthermore, Ma & Surakitbanharn (2019) identify hurricane damage using damage-related tweets, socio-economic data and insurance claim information; Sit *et al.* (2019) identify areas impacted by Hurricane Irma and with infrastructure damage; and Roy *et al.* (2020) identify types of infrastructure damage and disruption as a result of case studies Hurricane Irma and Michael using hurricane-related tweets.

Impact on roads/travel disruption

The impact on roads and travel disruption is another impact topic explored in some studies. For example, Lin *et al.* (2015) predict the impact of snow on traffic flow using weather data, traffic information and snow-related tweets; Tse *et al.* (2017) compare Weibo posts with observed weather data to detect if there is a relationship between weather and traffic congestion in Beijing city; Lu *et al.* (2018) identify traffic disruption in Beijing using weather-related tweets, also providing a prototype real-time application to monitor traffic impacts; Ahmad *et al.* (2019) examine flood-related images from Twitter to detect passable roads during flooding; Chen *et al.* (2020) examine hurricane-related tweet activity located around highways in Houston, USA to find the impact on roads during a hurricane; and Yang *et al.* (2020) use peaks in rainfall-related Weibo activity to detect areas with traffic impacts during 2018 Beijing rainstorm.

Changes in behaviour

Changes in behaviour during a weather event have also been examined using social media content. For example, evacuation response during hurricanes (Martín *et al.*, 2017; Stowe *et al.*, 2018); and the behaviour of people around rivers during floods (Anzai & Kazama, 2018).

Extent/severity of impacts

Using impact information extracted from social media, the extent and/or severity of a weather event can also be determined. Cervone *et al.* (2015) examine images from Twitter and Flickr to show the extent of damage and disruption as result of flooding; Kwon & Kang (2016) and Bai *et al.* (2020) assess flood-related social media posts to determine the severity of risk and vulnerability to flooding in particular locations; and Yue *et al.* (2018) compare hurricane-related tweets with the severity level of a hurricane, finding a positive relationship between tweet activity and severity of the event.

Sentiment of social media posts during an event

A number of studies consider how the sentiment of social media posts changes during a severe weather event, finding in almost all cases that negative sentiment is generally a sign of being impacted by severe weather. Vayansky *et al.* (2019) explore changes in sentiment during Hurricane Irma, finding an inverse relationship between sentiment score and wind speed; Giuffrida *et al.* (2020) compare weather-related tweets with weather observation data finding that changes in sentiment are an indicator of 'human comfort'; and Alam *et al.* (2018b) and Spruce *et al.* (2020) (Chapter 4) examine how sentiment changes before, during and after significant weather events, finding more negative sentiment during the most severe periods of weather. Therefore, suitable methods for calculating changes in sentiment of social media posts during adverse weather may act as a proxy for impact. To explore this, Yao & Wang (2020) develop a method for calculating sentiment during a hurricane event to explore its use as a proxy for impact; Yuan *et al.* (2020) calculate 'weighted sentiment' of hurricane-related tweets, taking user post frequencies into account which avoids bias in sentiment from any particular user; and Yum (2020) examine sentiment of hurricane-related tweets during Hurricane Florence to explore its use as an indicator of disaster impact.

Other studies

The focus of other studies relating to the social sensing of weather events includes: Demuth *et al.* (2018), who explore hurricane-related tweets to understand the perception of risk during a hurricane; Wang *et al.* (2020b), who explore the public response to a flooding event by examining spatio-temporal patterns of flood-related Weibo activity; and Sato (2019), who examine tweets containing "#rescue" during period of heavy rainfall in Japan to explore its use for identifying people in need of aid during extreme weather events.

2.4.2 Methods and Tools

The methods and tools used by researchers in each paper were also reviewed. Methods for each stage in the process of extracting social media content and exploring it for impact information are discussed below.

Data collection

The first stage in the process of using social media as an information source is to access the data. The vast majority of papers in this review examined social media posts from Twitter (Figure 2.2) which has an API that researchers and developers can easily use to extract the data. The majority of studies used the Twitter standard API[16] which allows a user to download tweets containing all metadata in a JSON file format (up to 1% of all tweets at any point in time) using keywords

[16] https://developer.twitter.com/en/docs/twitter-api/v1 (Accessed: 17 March 2022)

or a bounding box of co-ordinates and for a specific period of time. For example, Arthur *et al. (*2018) describe the process of extracting tweets using the Twitter API using flood-related keywords, and Chen *et al.* (2020) describe how they used a bounding box of co-ordinates to extract tweets originating from a specific location of interest.

Each approach has its advantages and limitations. Using keywords to collect data from social media means that there is increased likelihood that all tweets relating to a specific event will be collected. However, while this is likely to lead to more relevant posts in the dataset, location will not always be known. Additionally, where a weather-related keyword used in the search is also used in other contexts (e.g. "*floods* of tears", "Twitter *storm*", "*wind* up"), this could lead to irrelevant posts, not related to a weather event, in the dataset. Therefore, further processing techniques to filter the data for both relevance and location will be required.

Using a bounding box of co-ordinates to collect data provides a more robust approach for obtaining data from a specific location. However, only 1-2% of Twitter posts contain geo-location co-ordinates (Dredze *et al.*, 2013), therefore this will limit search results quite considerably. The dataset will also require further filtering for relevance as not all tweets from a particular location at a particular time will relate to the event of interest. One workaround, suggested by some authors, is to include place names of interest in the keyword search, as well as keywords related to the weather event of interest (e.g. Cerutti *et al.*, 2016; Eilander *et al.*, 2016; Cecinati *et al.*, 2019; Shannag & Hammo, 2019; Grace, 2020; Hamoui *et al.*, 2020). However, this assumes that a user will post the name of the location affected, therefore this approach will also limit the relevant posts included in the dataset.

Some studies, using Twitter, obtained data via other means. For example, Scotti *et al.* (2020) use tweet data from the Evolution of Emergency Copernicus Database (E2mC) (Havas *et al.*, 2017), which has already been filtered for relevance to natural disasters and location; and Dalela *et al.* (2020) analyse tweets from a human annotated tweet corpora, which was previously collated by Imran *et al.* (2016). While a robust data source, unfortunately pre-existing processed databases of social media content, such as these examples, are not suitable for real-time monitoring or situational awareness during weather events.

Sina Weibo (which is a Chinese social media platform, very similar to Twitter) makes its data available like Twitter, via a dedicated API which can be queried using keywords or bounding box co-ordinates[17]. For example, Bai *et al.* (2020) describe the process of extracting Weibo posts using keywords relating to flooding; Han & Wang (2019) and Tse *et al.* (2017) use a place name keyword to extract posts in a particular area of interest; and Fang *et al.* (2019) extract Weibo posts using both rainfall-related keywords as well as the place name of interest. None of the studies in this review extracted Weibo posts using location co-ordinates.

Flickr was also used in a small number of studies reviewed. The platform mostly contains photos and images that users share. Data can be extracted from Flickr, again using an API[18], using keywords or locations of interest which are included in the 'tag' associated with the image posted (e.g. Preis *et al.*, 2013; Chien *et al.*, 2017; Rosser *et al.*, 2017). However, Flickr datasets returned in searches generally have quite low volumes of posts (e.g. Cervone *et al.* (2015)).

Instagram is a more popular social media platform than Flickr and has been increasing in use across the world in recent years (Clement, 2020b). However, despite its popularity, Instagram data was only used in one study reviewed (Figure 2.2). Like Flickr, this platform mainly contains images with a small amount of text shared by users and data can be accessed via an API. For example, Anzai & Kazama (2018) use flood-related keywords as well as the name of a particular river of interest to extract flood related images from the Instagram API. However, the main limitation with using Instagram data is the more stringent user privacy rules, which means researchers do not have as full access to data on the platform as for Twitter (Instagram, 2020).

WhatsApp and Facebook are more difficult to access data from due to more strict user privacy rules[19], in spite of this, there were 2 studies in this review exploring data from these platforms. For example, Bhuvana & Arul Aram (2019) make use

[17] https://open.weibo.com/wiki/API 文档/en (Accessed: 17 March 2022)
[18] https://www.flickr.com/services/api/ (Accessed: 17 March 2022)
[19] https://www.facebook.com/robots.txt (Accessed: 17 March 2022)

of information from both Facebook and WhatsApp posts during a flooding event in India. However, the authors needed to be members of specific Facebook and WhatsApp groups in the community to be able to access the information. Congjuico (2015) examine Facebook posts during a flooding event in the Philippines; again, the posts were accessed by being members of specific community Facebook groups. Using these platforms as a means of curating impact information is therefore likely to be quite labour intensive and not possible to use in an automated/semi-automated approach.

Relevance filtering

Once social media data has been sourced, the next stage in the process of using social media as an information source is to remove posts which are not relevant to the weather type or impacts of the weather. As discussed above, most researchers obtained their raw dataset of social media posts using weather-related keywords, e.g. flood, rain, storm, wind, etc., and this can lead to posts including these terms but from contexts unrelated to the weather being included in the search (for example, “floods of tears”, “cook up a storm”, “wind up”). Researchers also need to consider if all posts relevant to the weather event are useful for impact information (i.e. observations of the weather and warnings about weather yet to happen could arguably not be useful as sources of information about impacts experienced). Therefore, an additional stage of filtering for relevance is required.

Keyword and/or bounding box coordinates filtering

Some researchers rely only on keywords contained within the text of the social media post as a way of filtering for relevance, accepting that there are likely to be a number of irrelevant posts, not related to the weather event, in the final dataset analysed (e.g. Herfort *et al.* (2014a); Jongman *et al.* (2015); Smith *et al.* (2017); Li *et al.* (2018); Wani *et al.* (2020)). Other researchers use a more refined method of filtering for keywords once the raw dataset is obtained. For example, Anzai & Kazama (2018) initially obtain their dataset of Instagram posts using mention of a particular river name, but then further filter their data based on if the post describes or shows impacts of flooding; Fang *et al.* (2019), after obtaining Weibo posts containing rainfall-related keywords, further filter posts based on specific impact terms and phrases contained within the message; Yuan & Liu (2018a) apply a similar impact keyword filtering method to hurricane-related Twitter posts;

and Saravanou *et al.* (2015) build a custom lexicon using most commonly found terms in flood-related tweet posts to filter tweets related to flooding.

Another method used to filter posts for relevance, particularly where the research is focused on a particular event in a particular location, is to extract social media posts using a bounding box of co-ordinates (Herfort *et al.*, 2014b; Barker & Macleod, 2019; Chen *et al.*, 2020; Grace, 2020a) or to include place names in the keyword search (Brouwer *et al.*, 2017; Han & Wang, 2019b; Fitrianah *et al.*, 2020; Grace, 2020b). While this offers many advantages, particularly if spatial impacts are of importance, it will result in social media posts which do not contain such geographic information being excluded from the search. For example, Cecinati *et al.* (2019) were interested in heatwave impacts within the country of India, therefore they used the keyword 'India' when querying the Twitter search API. However, they acknowledge that this approach limited the posts included in their results, as not all users posting about the heatwave would have included the word 'India' in their post.

Some studies use a combination of the above approaches using both keywords and locations (either co-ordinates or place names) to filter for relevance (Jongman *et al.*, 2015; Cerutti *et al.*, 2016; Eilander *et al.*, 2016; Jitkajornwanich *et al.*, 2019; Pourebrahim *et al.*, 2019).

Manual filtering

A labour-intensive, yet robust approach, is to manually check each post included in the social media dataset for relevance. For example, Fohringer *et al.* (2015) manually review images posted on Twitter and Flickr for relevance to flooding; Grace (2020b) manually review more than 20,000 tweets containing storm-related keywords, place names or users known to be in a specific location during storm events in the USA for relevance to storm impacts; and Demuth *et al.* (2018) manually examined tweets posted by 53 specific Twitter users during the period of Hurricane Sandy to check for relevance to hurricane impacts. Andrade *et al.* (2017) first filter tweets within a particular bounding box of co-ordinates for rainfall-related terms and then manually read thousands of tweets to remove those not relevant to the impacts of rainfall. A similar approach is carried out by Huang *et al.* (2018) who manually review tweets initially filtered using keywords for relevance to flooding. In total there were 12 studies in this review which used a manual relevance filtering approach (Lachlan *et al.*, 2014; Fohringer *et al.*, 2015; Lin *et al.*, 2015; Mukkamala & Beck, 2016; Andrade *et al.*, 2017; Demuth *et al.*, 2018; Wang & Ye, 2018b; Huang *et al.*, 2018; Shi *et al.*, 2019; Sato, 2019; Grace, 2020b; Liu *et al.*, 2020). While a robust approach, this is a time-consuming process, particularly for events with severe impacts, such as hurricanes in the USA, which can generate thousands of social media posts.

Filtering using machine learning methods

Other research has used a more complex relevance filtering process, using supervised machine learning methods. Supervised machine learning methods produce excellent results in terms of automating the filtering process and extracting social media posts most relevant to the impacts of weather events. However, this approach relies on a robust, sizeable, training corpus including examples of both relevant and irrelevant social media posts, which requires a manual process to create. It also requires Natural Language Processing (NLP) techniques to process the text of a social media post ready for classification. For example, removing punctuation, stop words and emojis, as well as vectorisation of the text of a post into single word or bi-gram occurrences. Some different approaches for vectorisation of the data have been explored. For example Yue *et al.* (2018) use 'Bag-of-Words' and 'Word2Vec' to prepare data for classification; Barker & Macleod (2019) and Rossi *et al.* (2018) use a trained 'Doc2Vec' model along with a logistic regression machine learning algorithm to filter tweets for relevance. De Bruijn *et al.* (2019) use BERT (Devlin *et al.*, 2019), which is a a deep learning-based natural language processing (NLP) model that learns relations between words and sub-words in a text (i.e. word embeddings) and uses these to encode text and Dalela *et al.* (2020) describe the use of TF-IDF[20] to vectorise the tweet text prior to the application of various supervised machine learning methods to filter tweets for relevance.

Once the text of a social media post is prepared for classification, there are a number of supervised machine learning approaches that researchers have explored for filtering posts for relevance. For example, De Bruijn *et al.* (2019) design a supervised classification model to classify tweets as relevant or not relevant to a flooding event based on a pre-labelled dataset of tweets, followed

[20] http://www.tfidf.com/ (Accessed: 17 March 2022)

by a 'burst detection' algorithm to determine likelihood of a flood event occurring; Arthur *et al.* (2018) and Spruce *et al.* (2020) (Chapter 4) use a Naïve Bayes machine learning algorithm to filter tweets for relevance to the impacts of floods and storms in the UK using a pre-trained dataset of labelled tweets; the same approach was also used by Kankanamge *et al.* (2020). A similar approach using a Support Vector Machines (SVM) machine learning algorithm is used by Feng & Sester (2017), Musaev & Hou (2017) and Wang *et al.* (2016); and a logistic regression machine learning approach is used by Rossi *et al.* (2018), Barker & Macleod (2019) and Bai *et al.* (2020). A combination of machine learning algorithms including SVM, Naïve Bayes, Random Forest, Decision Trees is also used by a number of authors to find the most suitable approach (Nair *et al.*, 2017; Stowe *et al.*, 2018; Yue *et al.*, 2018; Alam *et al.*, 2018b; Yuan & Liu, 2018b, 2019; Dalela *et al.*, 2020). And finally, Ma & Surakitbanharn (2019) use a Bidirectional Long Short-Term Memory (LSTM) recurrent neural network to classify tweets as relevant if the tweet text is both related to the weather event and informative about how an individual is affected.

Third-party data

Another approach to relevance filtering found in some studies, was to combine social media data with other third-party data, such as rainfall or satellite data, in order to improve relevance filtering. This was predominantly for studies focused on flooding and with the purpose of providing flood inundation maps. For example, Rosser *et al.* (2017) use remote sensing information and topographic data to aid with the filtering of relevant images relating to flooding from Flickr; Restrepo-Estrada *et al.* (2018) find combining social media data with observed rainfall improves the relevance of flood related tweets; de Bruijn *et al.* (2020) develop a multilingual multimodal neural network which uses both textual and hydrological information to filter tweets for relevance to flooding; and Scotti *et al.* (2020) combine satellite images and hydraulic models and flood-related tweets to improve the accuracy of flood inundation maps created using social media content. However, the use of third-party data relies on having sufficient coverage of remote sensing information.

Filtering using images

Additionally, a number of studies focused on the use of images in social media posts for identifying impacts of weather hazards. Again, this was predominantly relating to identifying flooding impacts. Filtering images for relevance was carried out manually by some authors to identify flood-related images (Fohringer *et al.*, 2015; Anzai & Kazama, 2018; Liu *et al.*, 2020), and via supervised machine learning (mainly neural network) approaches (Cervone *et al.*, 2015; Wang *et al.*, 2018, 2020a; Moumtzidou *et al.*, 2018; Huang *et al.*, 2019b,a; Ahmad *et al.*, 2019; Jony *et al.*, 2019; Pereira *et al.*, 2020). For example, Moumtzidou *et al.* (2018), Wang *et al.* (2018), Jony *et al.* (2019) and Huang *et al.* (2019b,a) use a Convolutional Neural Network (CNN) approach to classify images posted on Twitter as relating to flooding; Ahmad *et al.* (2019) use a pre-trained dataset of flood images to identify flooded roads; and Wang *et al.* (2020a) use Computer Vision (CV)-based image classification to detect flood related images from Twitter, although find that their results still need an element of manual checking.

Location Inference

Many papers in this review were focused on the spatial extent of impacts during a weather event, particularly in relation to flooding inundation. Therefore, the need for location information from social media posts has led many researchers to use only those posts with exact geo-location information. Nearly half the studies (49%) in this review, and which also considered the spatial impact of weather events, only used social media posts that contained geo-location information (53 out of 108 papers). Whilst social media posts which contain geo-location coordinates provide accurate locations for mapping, there are a very low proportion of posts containing specific location co-ordinates. For example, Sina-Weibo encourages users to share location information, however there are still less than 10% of posts which contain GPS information (Wang *et al.*, 2016). For researchers using Twitter as their data source, the proportion of posts containing GPS co-ordinates is even lower at 1-2% of posts (Dredze *et al.*, 2013). This means that many tweets which carry useful information, but do not contain location co-ordinates will have been discarded in some studies (e.g. Butgereit (2014); Rossi *et al.* (2018); Barker & Macleod (2019)).

To overcome this limitation, some researchers have highlighted location inference methods which address the issue of lack of accurate geo-coordinates in the social media post. For research which is focused on one particular location or one particular weather event case study, a number of authors found that looking for specific place names in the tweet text aided both filtering tweets and plotting aggregated tweet activity on a map more generally (Jongman *et al.*, 2015; Kwon & Kang, 2016; Eilander *et al.*, 2016; Jung & Uejio, 2017; Fang *et al.*, 2019; Han & Wang, 2019b; Shannag & Hammo, 2019; Grace, 2020a; Fitrianah *et al.*, 2020; Grace, 2020b; Liu *et al.*, 2020). De Vasconcelos *et al.* (2016) apply the same approach but finding specific location mentions (e.g. city names) in the user location field of a tweet and match these to location co-ordinates using Google Maps. Brouwer *et al.* (2017) used Google Maps and Google StreetView to identify locations in the area of interest and then manually looked for mentions of these locations in the tweet text. Demuth *et al.* (2018) also manually determine locations in the area of interest mentioned in the tweet text.

A less labour intensive and more comprehensive automated location inference approach has been developed by a number of other authors. For example, Arthur et al. (2018) find location mentions in the tweet text and/or user location information to infer location and then look up coordinates in gazetteer databases such as Geonames[21] and DBpedia[22]. Geonames is a geographical database, which is available for download free of charge, and contains over 27 million placenames across the world and their associated latitude/longitude co-ordinates, population, elevation and other data (Geonames, 2020). DBpedia is a community effort to extract structured information, such as place names, from Wikipedia (https://www.wikipedia.org/). It provides a knowledge base of over 2.6 million entities, including geographic information such as geo-coordinates for identified place names (Bizer *et al.*, 2009). This increased the proportion of tweets which can be located during an event to around 70% as opposed to 1-2%. Very similar location inference methods were also employed by Spruce *et al.* (2020) (Chapter 4). These were the only studies which used both tweet text and user location information to locate tweets. A geoparsing technique was also developed by de Bruijn *et al.* (2019, 2020) who use the TAGGS algorithm (de Bruijn *et al.*, 2018), which was developed to find locations in the tweet text and look up co-ordinates in gazetteer databases. Wang *et al.* (2020b) find place names in the tweet text using a deep learning-based toponym recognition NER tool, NeuroNER, which trains a long short-term memory (LSTM) model, a variant of

[21] https://www.geonames.org/ (Accessed: 17 March 2022)
[22] https://wiki.dbpedia.org/ (Accessed: 17 March 2022)

the recurrent neural network based on the CoNLL 2003 dataset (Tjong *et al.*, 2003). They then link identified place names to co-ordinates using Geonames and TIGER data (which is a US database of road names). Other authors, focused on a particular case study or location of interest, also develop automated/semi-automated methods to extract location mentions from the tweet text and then look up the coordinates for these places in other databases (Cervone *et al.*, 2015; Grasso *et al.*, 2017; Wang *et al.*, 2018; Lu *et al.*, 2018; Rodavia *et al.*, 2018; Khaleq & Ra, 2019; Jitkajornwanich *et al.*, 2019; Yang *et al.*, 2019b; Roy *et al.*, 2020). Wani *et al.* (2020) and Yang *et al.* (2020) apply similar approaches, however used place names mentioned in the user location, rather than the tweet text, to locate tweets.

Using place names mentioned in social media meta-data, in addition to the text of the post, may have limitations that should be acknowledged. For example, place names mentioned in the user profile location may not be relevant to the weather event being discussed, as the user may not be in their ‘home’ location; the information in the user profile may not be current; or the user could be posting about an event in a different location to where they are based. However, overall, approaches which use other information in the tweet besides just the geo-location information are more likely to obtain a more complete spatial dataset for determining the location of impacts.

Tweet topic categorisation

As well as the spatial dimension, the content of social media posts can also provide insight into the social impact of weather events. Topic-based classification focuses on mining what people talk about in natural disaster situations (Wang & Ye, 2018a). There were 30 papers in this review which describe methods for categorising tweets into topics. The majority of these studies focused on the text of the social media post and categorised the text into a small number of impact topics such as damage reports, observations, disruption, etc. Methods ranged from manual categorisation of social media posts into topics (e.g. Spruce *et al.* (2020) (Chapter 4)), to more complex machine learning algorithms using natural language processing techniques to automatically assign posts into categories (e.g. Alam *et al.* (2018b)).

Manual categorisation of social media content

Where the social media data sample is small, then manual categorisation provides a robust and accurate method for assigning posts to categories. For example, Shi *et al.* (2019) had a very small dataset of tweets (just 109 posts) and were therefore able to easily manually code these into five categories based on the content of the dataset. However, for larger datasets, then an interim review stage may be required to aid with categorisation. For example, Ukkusuri *et al.* (2014) find the ten most frequently used words in tweet text to identify topics of discussion then manually categorise tweets into eight disaster-related categories based on these keywords including damage and injury reports, fundraising and support and consolation.

Lachlan *et al.* (2014) provide a coding strategy for manually identifying hurricane-related topics of discussion from tweets once they have been reviewed for purpose, primarily related to tweets which provide impact information such as evacuation, food/shelter, loss of assets, etc. They raise the important point that tweets which do not provide information (e.g. humorous tweets written to entertain, or emotional tweets written to convey fear, anxiety or excitement about an event) are not useful for information about how people are being affected. This point was further verified by Halse *et al.* (2018) who manually code tweets into four categories, including weather warning/retweet, informational, related to event but containing no information and unrelated, finding again that the informational category of tweet is most useful in terms of understanding impacts.

In another study, Herfort *et al.* (2014a) manually classify tweets into thematic categories, initially based on categories proposed by previous studies (Vieweg *et al.*, 2010; Imran *et al.*, 2013), however they found that once inspecting their tweet dataset, these pre-defined categories needed to be adapted to suit their particular flooding case study. Similarly, Wang & Ye (2018b) and Sovacool *et al.* (2020) manually review the dataset for their particular case study, finding that previous classification schema from other studies were not suited and therefore developed their own set of topics of discussion. This suggests that categorisation of social media content depends on the hazard and the severity of the impacts and therefore a dataset may need to be reviewed for suitable topics of discussion before applying categorisation schemes. For example, Spruce *et al.* (2020) (Chapter 4) break down the tweet timeseries during Storm Brian in the UK in 2017 into categories, having first manually reviewed a subset of tweets, finding six topics of discussion which they were able to apply to the remainder of the dataset. While a labour-intensive process, this provided a good insight into how the type of tweet changes during a significant storm event. In a similar study, Grace (2020b) provide a detailed breakdown of their process for manually assigning categories to storm-related tweets to establish the type of information they provide. They find very similar categories of discussion in their dataset to Spruce et al (2020) (Chapter 4). This similarity in categories found could be attributed to the fact that the localities of each study (United Kingdom and Pennsylvania, USA) were likely similar in terms of vulnerability and exposure to storms. Therefore, while it has been shown that there is no 'one size fits all' categorisation schema for all weather hazard events, there may be some evidence that impact categories could be shared across comparable events and localities.

A further consideration for categorisation of tweets is the purpose for which the information is required. For example, Sato (2019) examine tweets posted during a heavy rain disaster in Japan in 2017, specifically for the purpose of finding people requiring aid. They manually review tweets containing the hashtag '#rescue' for relevance to people needing aid and the type of aid required. Yuan & Liu (2018b) are primarily focused on damage assessment using information from social media, therefore they manually review filtered tweets, looking for specific damage-related keywords and then assign a relevant damage category to each tweet. Yang *et al.* (2019) also use pre-defined keywords to classify social media posts into themes relating specifically to impacts (i.e. sheltering, casualty, damage, flood, power/electricity).

NLP methods for categorisation of social media content

As with filtering tweets for relevance, manually assigning social media content into topics of discussion is a robust approach. However, for large datasets this is a time-consuming process and would be difficult to sustain in real-time or for numerous events. Therefore, some authors have considered more automated approaches to topic modelling of social media content.

A semi-automated approach to categorising tweets into topics is to look for specific keywords in the text of social media posts and automatically assign a category based on the presence of one or more particular keywords. For example, Lu *et al.* (2018) use ‘Word2Vec’ with tweets related to traffic incidents to create lists of both traffic-related and weather-related keywords. These are then used to automatically assign tweets containing a combination of words from each of these lists to various categories of weather-induced traffic incidents.

Where the impact information required from social media is known (e.g. damage assessment, travel disruption) or if researchers have already determined a pre-defined classification schema, then a supervised machine learning approach (which uses a training corpus of manually categorised content) may be appropriate. Nair *et al.* (2017) explore three different supervised machine learning classifiers (Random Forest, Decision Tree, Naïve Bayes) to assign categories to flood-related tweets using pre-defined categories and a training corpus of labelled tweets. They find the best performance using a Random Forest classifier. Dalela *et al.* (2020) also find good categorisation results using a supervised machine learning approach, finding the best results from a Linear SVC algorithm with a set of manually categorised training data.

One issue encountered with supervised classification methods is that the content of a social media post may belong to multiple categories. For example, a post could talk about different aid needs like food, water, shelter. Depending on the coding scheme, this can therefore make assigning one category to a social media post difficult. To overcome this, Bai *et al.* (2020) classify the text of flood-related Weibo posts with several different pre-defined ‘event-meta’ labels. They acknowledge the issue that some social media posts can be assigned to multiple categories (labels) and therefore propose a multi-classification approach using the machine learning logistic regression multi-classification algorithm ML-KNN (Zhang & Zhou, 2007), which can apply multiple labels to the text. Roy *et al.* (2020) also developed a multi-label Logistic Regression (LR) machine learning classification method to categorise hurricane-related tweets into different categories of damage and disruption, finding that their LR model is more successful at categorising text than using keywords and sentiment.

Unsupervised topic modelling

It is not always possible to apply a pre-defined classification of topics to unsolicited social media data, therefore rather than manually reviewing the dataset to define topics, some researchers have used automated topic modelling techniques to derive categories of discussion from social media posts. The advantage of unsupervised topic modelling is that it enables clustering of a large volume of tweets into different groups. For example, to understand topics of discussion in their hurricane-related social media dataset, Alam *et al.* (2018b) use Latent Dirichlet Allocation (LDA) (Blei *et al.*, 2003) which is a well-known topic modelling technique to generate topics from large amounts of textual data. These topics are then used with a Random Forest (RF) machine learning classifier to assign each tweet into a category of discussion (e.g. affected individual, infrastructure and utilities damage, injured or dead people, etc). Wang *et al.* (2016) also use LDA to assign topics to the text of Weibo posts and use this as a training sample to train a Support Vector Machines (SVM) machine learning classifier to classify tweets into topics in real-time. Han & Wang (2019) use a very similar approach, using the 'Gensim' package in Python to implement an LDA model to identify topics and assign the most likely topic to the text of each Weibo post. These are then used to build a training corpus of annotated Weibo text to, again, build a Random Forest (RF) machine learning classifier to further categorise the text into specific categories. They find that this approach has good performance. The LDA model is also used by a number of other researchers (Xin *et al.*, 2019; Vayansky *et al.*, 2019; Sit *et al.*, 2019; Wu *et al.*, 2020; Yuan *et al.*, 2020) to find topics of discussion from the text of social media posts and is the most common unsupervised topic modelling approach used in studies within this review.

Other unsupervised topic modelling approaches explored include: Oktafiani *et al.* (2012) who use a graph-based concept, complex network analysis and TF-IDF term weighting to identify clusters of similar words within tweet text, which identifies topics of discussion; Pourebrahim *et al.* (2019) apply a similar approach to Oktafiani *et al*. (2012), using TF-IDF to calculate term frequency and bi-grams in the tweet text, which are then used in a network structure to carry out word co-existence analysis to identify clusters of similar use of terms; Dong *et al.* (2013) used the Latent Semantic Indexing (LSI) algorithm in genism (Řehůřek & Sojka, 2010) to produce a range of topics from tweet text; and Anam *et al.* (2019) use their previously developed Continuous Wavelet Transform approach (Anam *et al.*, 2018) to identify word clusters in hurricane-related tweets. This method looks for word signals in the tweet text both in time and frequency, and therefore is useful for both event detection and topic discovery. However, while a comprehensive methodology, due to its complexity it may be difficult for other researchers or practitioners to apply in a real-time application.

One limitation of some of the unsupervised topic modelling approaches outlined in the papers reviewed is the allocation of only one category of discussion to an individual social media post. Therefore, unsupervised methods, such as LDA, which are able to assign a social media post to multiple categories (e.g. a social media post which describes damage observed AND travel disruption) may be worthy of further exploration.

Categorising images

A small number of papers discuss methods for classifying images in social media posts to improve situational awareness. All of these studies were focused on flood-related impacts. For example, to detect passable roads, Ahmad *et al.* (2019) manually annotated images to extract relevant images and classify as passable or not using a pre-trained neural network; Alam *et al.* (2018b) use a pre-developed image classification model (Nguyen *et al.*, 2017) to assess the severity of damage using images; and Wang *et al.* (2020a) manually classify 6542 images from Twitter with four categories which is then used to develop a "TensorFlow" based classification scheme using two neural networks (CNN and ResNet) to assign a category to flood-related images. Additionally, Liu *et al.* (2020) take two approaches to classifying images for flood-related impacts: they manually examine 308 flood related images from Twitter for flooded roads/streets and property damage; these same images are then uploaded to the Google Cloud Vision API which classifies and assigns categorical labels using Google's pre-trained machine-learning models and results compared with the manual approach.

Still related to the impacts of flooding, a small number of studies have also used flood-related images from social media posts to estimate flood depth. For example, Chaudhary *et al.* (2019) use a deep learning model to estimate water level from flood-related images; and Pereira *et al.* (2020) also estimate the flood depth of flood-related images using DenseNet and EfficientNet neural network architectures, which is then used as an indication of the severity of a flooding event in a particular location.

Sentiment analysis

An additional impact measure that can also be determined from social media posts is sentiment. This is where the text is used to calculate a positive or a negative opinion by the user. Aggregated together, the sentiment value of a group of event-related posts temporally or spatially can infer a positive/negative reaction to that event. In the case of weather events, then some researchers have used the sentiment score as another measure of impact as a result of adverse weather conditions. Yum (2020) argue that some statistical programs with pre-trained sentiment classifiers often struggle to interpret certain human emotions, such as sarcasm and irony. Therefore, they manually tag 1000 hurricane-related tweets with a sentiment category (very negative, negative, neutral, positive, very positive) to analyse changes in sentiment both temporally and spatially during Hurricane Florence in which they demonstrate the importance of human sentiment as an indicator of disaster impact. Therefore, a robust automated method for calculating the sentiment of social media posts would provide a good measure of impact.

There are some pre-trained packages which researchers can use to calculate the sentiment of short text documents, such as social media posts. Of the studies examined in this review, the following packages were used with good success: SentiStrength[23] (Ukkusuri *et al.*, 2014); Stanford sentiment analysis classifier[24] (Alam *et al.*, 2018b); TextBlob[25] (Ma & Surakitbanharn, 2019; Spruce *et al.*, 2020 (Chapter 4)); VADER(Hutto & Gilbert, 2014) (Roy *et al.*, 2020). All of these packages can be used to calculate a sentiment polarity score to reflect positive and negative use of words in posts.

Other researchers have developed their own methods. For example, Vayansky *et al.* (2019) develop a sentiment library in which words are assigned a score representing their positive or negative polarity. The score of words in a tweet are then averaged to give an overall sentiment score. Their method finds an inverse relationship between sentiment score and wind speed. Supervised machine learning classification methods have also been developed. Pourebrahim *et al.* (2019) calculate sentiment score using SVM to assign a polarity score (positive,

[23] http://sentistrength.wlv.ac.uk (Accessed: 17 March 2022)
[24] https://nlp.stanford.edu/sentiment/ (Accessed: 17 March 2022)
[25] https://textblob.readthedocs.io/en/dev/ (Accessed: 17 March 2022)

negative, neutral) to tweets; Giuffrida *et al.* (2020) train a classifier using weather-related tweets, which is then used to assign a positive, neutral, negative sentiment classification to tweets; Kankanamge *et al.* (2020) use the Weka 2.0 software to apply a decision tree algorithm to assign a sentiment score to tweets which have been trained for positive and negative sentiment. Other methods developed include: Yao & Wang (2020), who develop a domain-specific sentiment analysis approach specifically for tweets posted during hurricanes (DSSA-H) based on a recurrent neural network, finding that their model outperforms other general short-text sentiment analysis packages; and Yuan *et al.* (2020), who calculate weighted sentiment using a lexicon-based approach (AFINN), which takes user post frequencies into account to avoid bias by any particular user and provides a potentially more accurate method for calculating sentiment during a weather event.

Some researchers classify the sentiment of social media posts with a type of emotion, rather than a sentiment polarity score. For example, Demuth *et al.* (2018) manually review hurricane-related tweets for the type of emotion they express (e.g. worry, fear, anger), examining how the types of emotions of tweets vary during the period before, during and after a hurricane. Yang *et al.* (2020) use a previously developed deep learning model based on a Convolutional Neural Network (Yang *et al.*, 2019a) to classify Twitter posts into six different emotional categories and use this as an additional measure about weather-related traffic impacts. However, as outlined by Yum (2020), many automated sentiment classification approaches have less success classifying social media posts containing humour, irony and sarcasm.

2.5 Discussion

This systematic literature review provided an overview of how social media data has been used to understand impacts during significant weather events from 2010-2020. The majority of studies used Twitter data and focused on the impacts of flooding or hurricanes. However, other social media platforms (Sina Weibo, Flickr, Facebook, WhatsApp and Instagram) were used in some studies and other weather types explored included storms, heavy rainfall, heatwaves, tornadoes, landslides and snow. The publication of studies in this area was also found to be very interdisciplinary with researchers publishing their research across a range of subject areas from computer science to earth and planetary sciences and social sciences. The USA was also found to be the main country of interest to researchers, followed by China and the UK, with a smaller number of studies in other countries across the world.

The purpose and aims of studies largely fell into two categories, event detection and impact assessment. Event detection studies included work which explored the potential of detecting events with social media using social media data from previous events as a case study. Some authors also considered the potential of real-time applications to aid situational awareness. Flood inundation mapping and using social media as a proxy for observations in areas with limited remote sensing data were also proposed. Additionally, many studies focused on one particular weather event or case study, and therefore event detection methods proposed would need further testing with other similar events.

Impact assessment from social media ranged from using volumes of tweet activity in a particular time and place as an impact measure, to identifying topics of discussion from the content of text/images contained with social media posts. Researchers have found that the type of impact information from social media changes before, during and after an event, with the most impact information being available in the ‘during’ phase of a weather event. Studies have also explored the use of social media for damage assessment, impact on roads/travel and changes in behaviour. The calculation of sentiment using social media content has also been shown to be an indicator of impact.

In terms of methods and tools used, reviewing the literature has shown that there are a number of stages in the process of utilising social media data. The first stages of data collection and identifying relevant tweets can be achieved using a number of relevance filter methods. Some studies collected social media posts using a specific hashtag or keyword(s), whereas others searched for keywords and then applied relevance filters using supervised machine learning methods. Methods for filtering social media content for relevance ranged from manually checking each social media post, the use of particular keywords in the text of the post, to the application of supervised machine learning methods. Some studies also used third party remote sensing data, such as observed rainfall or hydrological data, to assist with filtering social media content for relevance, finding that this yielded better results in terms of relevant content. However, this relies on having sufficient third-party data available.

It is also clear that methods for better determining the location of social media posts in the absence of specific co-ordinates or location fields is required to utilise the full content of information available via social media. Many of the studies reviewed here used only social media posts containing GPS co-ordinates (geo-tags) or specific place co-ordinates. As only a small proportion of social media posts have specific location co-ordinates included in the metadata, much useful content is therefore being overlooked. Some studies in this review proposed methods to utilise place name mentions in the text of a social media post and/or the user location, to be able to better locate information posted about a weather event. Further studies aiming to locate tweets should therefore ensure that multiple methods are employed for locating tweets and look to refine and improve these methods in order to further increase the amount of spatially useful content that can be used relating to natural hazard events.

Utilising social media content to find impact information employed similar methods to relevance filtering. For text categorisation, some studies manually categorised social media content to find particular impacts. However, this is a time-consuming and labour-intensive process, and therefore not sustainable as an approach in real-time situational awareness. A semi-automated approach which employed the inclusion of particular keywords in social media posts to categorise the content was suggested by some authors. While reasonably easy to apply, this could lead to some content being mis-labelled. Supervised machine learning methods were also suggested in some studies. This included the use of a training corpus of examples of social media content labelled with a relevant category. Methods included neural network machine learning architectures or other supervised methods such as decision tree or SVM approaches. Unsupervised methods for topic discovery were mainly using LDA techniques to identify topics of discussion or to create a labelled training corpus which can then be used with supervised methods to categorise social media content. For image categorisation, supervised neural network methods were proposed by most researchers as the most suitable approach.

A further method for determining impact information was to calculate the sentiment of social media posts. Changes in sentiment were found to be a useful indicator of impact both spatially and temporally. Methods to calculate sentiment ranged from pre-trained packages which can be applied to social media text to bespoke supervised machine learning methods developed by authors to categorise social media content.

2.5.1 Future directions

As indicated by this literature review, future work to determine impact information relating to weather events from social media should consider the application of many of the proposed methods described here to develop real-time applications which can assist agencies with situational awareness. Methods which proved to be most successful were those which considered both spatial and temporal patterns in social media activity, as well as making use of the tweet content. Methods should therefore use both a robust, automated relevance filter to exclude social media content not related to the weather event and use location inference techniques to locate social media posts using more than just geo-location co-ordinates contained within the metadata of a post. Automated methods for relevance filtering using a supervised machine learning approach should be further explored and refined. Location inference techniques which build on and improve existing work should also be explored. Use of social media content to identify topics of discussion, impact types and sentiment have all been found to provide useful impact information. Many studies focused on use of text or images from social media, but rarely both together, therefore future studies may also wish to consider using both text and images posted on social media to improve situational awareness and impact information curation.

Another important consideration for future studies is that of bias towards more densely populated areas of the world or demographic bias (e.g. people more likely to be using Twitter in a particular country). How salient or familiar the hazard is will also affect the degree of reporting. An unusual (but small) event may end up being reported on social media significantly more than a familiar / common (but larger) event. This could skew interpretations of impact or severity if based on the volume of social media posts alone. Therefore, methods which mitigate these potential biases by not just relying on the volume of social media activity, but providing some sort of normalisation measure would also be beneficial. Other factors which could affect the volume of social media posts collected include time of day (e.g. a weekday morning) and time of year (e.g. Christmas), media coverage of the event and countries where that particular social media platform is more or less likely to be used.

For researchers using Twitter data, other social media platforms, for example Sina Weibo, could be explored to increase the global coverage of their models. Additionally, while Twitter and Weibo were the most commonly used social media platforms used by researchers in the studies reviewed, other useful impact information may be available on other social media platforms. For example, community response on WhatsApp or Facebook groups (Congjuico, 2015; Bhuvana & Arul Aram, 2019). However, general privacy constraints for these platforms, such as the need to be a member of particular community groups to access content, may make this data more difficult for researchers to access. Telegram[26] is an example of a social media platform which has been increasing in use in recent years, particularly in certain countries (e.g. India) (Dean, 2022) and which is yet to be explored fully by researchers as a source of impact information. Therefore, further work to explore other social media platforms could therefore be another avenue of future research.

2.6 Conclusion

What is clear from the review of these case studies however, is that social media (and in particular Twitter) can be used to provide additional information about the social impacts of weather events be that spatial, temporal or emotional and should continue to be explored to aid both validation of forecast models and to support situational awareness during severe weather.

Social media is only one of many information sources and therefore future studies may also wish to consider combining other data sources with social media to give a full picture of the impact of a severe weather event.

[26] https://telegram.org (Accessed: 17 March 2022)

Chapter 3 - Emerging themes and research questions

After reviewing previous studies which have used social media to better understand the impacts of weather events, there is clear scope for further work with this data source. In terms of motivation, Chapter 1 has provided the context behind the need for information about weather impacts and Chapter 2 examined previous work which used social media as a data source to begin to address this need.

In this chapter the themes which have emerged from examination of the literature and wider context will be discussed. Themes identified include: accessing only relevant information relating to the impacts of weather from social media; locating social media data which does not have geospatial information; impact assessment using social media; and development of real-time applications to assist with situational awareness. While each theme requires a separate approach in terms of methods and exploration, all these themes are interlinked, with each theme relying on the others to be achieved.

As social media originates from individuals, it is also important to consider the ethics behind using social media data for both research and as an information source. This will be discussed in Section 3.3.

3.1 Emerging themes

Each theme identified for further examination will now be discussed.

3.1.1 Accessing relevant information from social media

From examination of previous studies on the social sensing of weather events, there are a number of different approaches which have been explored. Manual methods, while accurate, are time consuming and labour intensive. The use of weather-related keywords presents a good initial stage for filtering social media data for relevance, however this can still result in social media content which is irrelevant in the context of a weather event (e.g. "*floods* of tears", "the streets were *flooded* with people") (Arthur *et al.*, 2018). Additionally, for curation of information about impacts, it is also important to consider whether the information included in the social media posts is simply providing a report that the weather event is occurring (e.g. “the rain is very heavy today”), or whether it is providing information about the specific impacts being experienced (e.g. “my house is flooded”). Therefore, an automated relevance filter, using machine learning techniques trained with examples of social media posts that describe impacts is most likely to provide timely access to relevant impact information for situational awareness.

Also, many of the previous studies considered in Chapter 2 focused on a single case study of a historic weather event. While this allows researchers to test out their methods for extracting relevant social media information in the context of a weather event, applying methods to multiple events, in multiple locations is an obvious next step in developing methods which can be applied to future events (e.g. de Bruijn *et al.* (2020)). It is not known where and when future weather events will occur, therefore methods need to be more generally applicable.

The social media platform used as an information source by most researchers in previous studies was Twitter due to its accessibility via the developer APIs (or Sina Weibo for studies focused on China, where Twitter is disallowed). Impacts are reported by people on other social media platforms during a weather event and capturing this data is therefore another area worthy of further exploration. However, limitations in privacy rules for other social media platforms are an issue for some platforms.

Verification of events detected using social media has been explored by some authors (e.g. Arthur *et al.* (2018); de Bruijn *et al.* (2019)). This is an important step when developing methods which can be used to interrogate social media data for future events and may also provide evidence of how social media can be used to detect events which would otherwise not have been detected.

Most previous studies examined also tended to focus on impactful events such as flooding and hurricanes. Only a small number of studies focused on other events which also cause significant impacts, for example, extreme temperature events such as heatwaves and snow. Therefore, extending some of the well-developed methods focused on flooding and hurricanes to other weather events is another area for further investigation.

3.1.2 Locating social media data which does not have geo-location information

The spatial extent of impacts as a result of weather events are an important aspect of impact information curation. The location of impacts is important for situational awareness in terms of *where* people are being impacted at a particular point in time and also to evaluate the effectiveness of forecasts and warnings issued for a particular place or area. Therefore, another major finding from exploring the literature is the usefulness of being able to locate social media posts which do not contain specific geospatial information. For example, in the case of Twitter, only 1-2% of tweets contain “geotags” which are point co-ordinates from where the user was when they posted their content, or “place” coordinates which provide the bounding box of a user specified location attached to a tweet. Therefore, there is a significant amount of social media content that cannot easily be located and therefore much potential spatial impact information not available. Despite this limitation, nearly half of the studies examined in the literature review only used social media posts containing geospatial coordinates in their impact assessment. However, a number of researchers have begun to explore location inference techniques that can determine place name mentions in social media content, determine the location coordinates for this place and therefore infer the location of the social media post (e.g. Grasso *et al.* (2017); Arthur *et al.* (2018); de Bruijn *et al.* (2018); Jitkajornwanich *et al.* (2019)). Refining location inference methods is therefore an important area of consideration and further work for curating impact information relating to a particular weather event and location from social media data.

3.1.3 Impact assessment using social media

It seems obvious that when a significant weather event occurs, people will talk about it on social media. However, there is a difference between discussion about occurrence of the event and discussion about impacts as a result of the event. Both types of discussion may be useful in different scenarios. For example, peaks in social media activity relating to a particular event provides a signal that the event is occurring. However, researchers looking to develop methods to use social media data as a source of impact information during weather events need to consider whether social media content provides information about the impacts of the event, not just that the event is occurring. As impact information is difficult to obtain from other sources, particularly in real-time, using social media for impact assessment is where social sensing methods may be most useful.

Previous studies have provided a number of methods and approaches for impact assessment using text and/or images posted on social media. Many researchers have considered the temporal and spatial variations in the volume and type of social media content as an indicator of impact (e.g. Spasenovic *et al.* (2019); Farnaghi *et al.* (2020)). However, the assessment of impact also needs to consider the type of impacts being reported (e.g. damage to property, disruption to transport, etc). Therefore, identifying topics of discussion during weather events has begun to be explored by some researchers (e.g. Yamada *et al.* (2019); Liu *et al.* (2020)). The temporal variations of the type of tweet content before, during and after a weather event can also provide information about how the type of impact information available changes during the event (e.g. Alam *et al.* (2018)). Additionally, calculating the sentiment of social media posts can also provide an indication of the severity of impacts during a weather event (e.g. Giuffrida *et al.* (2020); Yao & Wang (2020)). Each of these approaches provide an aspect of impact assessment using social media content, however few researchers have used multiple approaches to assess impact, tending to focus on one method of impact assessment only.

3.1.4 Development of real-time applications to assist with situational awareness

Impact assessment of weather events using relevant, timely social media content has been shown to have real potential as an information source, however it relies on applying suitable methods to filter content for relevance to weather impacts and determine its location. Furthermore, to be able to use this information for situational awareness requires it to be available to those managing the result of impacts or providing impact-based warnings and forecasts in a timely and accessible manner. Therefore, this research has the potential to move beyond exploring what is possible, to developing these methods and findings into applications which can be used in real-time to improve situational awareness during weather events and enable action to be taken - be that improving communications and warning information issued by National Meteorological and Hydrological Services (NMHS) or enabling organisations on the ground to direct efforts and resources to the most impacted people and locations.

3.2 Research questions

Revisiting the research questions proposed in Chapter 1, each will now be discussed in turn, relating the review of previous work to each to determine how these might be addressed.

RQ1. How useful is social media as a source of impact information during and after weather events?

In Chapter 2, a systematic literature review of previous work relating to the social sensing of weather, and in particular that relating to determining the impacts of weather from social media, was undertaken. This review found that many previous studies have proved the utility of social media as a source of impact information during different weather events, both for event detection, situational awareness and a retrospective review of impacts as a result of the weather event. However, many studies considered only one event case study, or focused on one aspect of impact. Therefore, further work which explores the usefulness of social media as a source of impact information across multiple events will provide a more holistic view of how social media might be used for impact assessment.

RQ2. What tools and methods can be successfully applied to extract relevant social media data during weather events?

As discussed in Section 3.1, previous studies have outlined a wide range of methods and tools which can be used to extract, filter and interrogate social media data relating to weather events. In particular, processes which included an automated machine learning approach for filtering data for relevance, and which applied location inference for posts which do not include geospatial information, are likely to yield the most accurate and complete results. Additionally, analysing the content of relevant social media posts for impact assessment can be achieved in a number of ways, including manual review and applying natural language processing techniques. Calculating the sentiment of social media posts was also found to be a useful measure for assessing the likely severity of impact for different weather events. Many studies reviewed focused on one or two of these aspects, for example, event detection only, or topic analysis. Therefore, further work which applies multiple methods and tools to extract, filter and analyse social media data for a particular weather event is recommended.

RQ3. What are the limitations of social media as an impact data source?

There are limitations of social media as a data source. One key limitation is that it contains a lot of irrelevant content, therefore suitable methods and tools must be employed to access the most relevant information. This may be difficult for operational staff in NMHS and other organisations monitoring the impacts of weather to apply, particularly in real-time for situational awareness.

Additionally, access to the data for certain social media platforms can be limited due to user privacy restrictions (e.g. Facebook, Instagram, WhatsApp). Twitter and Sina Weibo are easily accessible data sources for researchers and have been widely used. This means that information which is shared on other social media platforms, but that provides detailed impact information, is not accessible.

Having reviewed the literature, the global extent that social media data can provide impact information for is not clear. Is this data source only suitable for use in more densely populated, well developed countries that have been selected as case studies by researchers so far? Or is there sufficient content available in more sparsely populated areas of the world to be useful? This is an additional area for further exploration.

3.3 Ethical considerations for using social media data

Social media data provides an excellent opportunity for researchers to gather information that would otherwise have taken much time and resource to obtain. However, given that this data originates from individuals who may not be aware of the onward use of their social media posts, it is important to consider the responsibility to ensure that this data is obtained and used in an ethical way. Despite the vast amount of social media data that researchers have already gathered during crisis events, there has been some limited discussion about the ethics of using this data, but the practice of using this data has overtaken the embedding of embedding ethical principles in social media research (Crawford & Finn, 2015; Leonelli *et al.*, 2021). Furthermore, in the review of literature relating to the social sensing of weather events detailed in Chapter 2, only 5 out of the 108 studies reviewed made any mention of ethical issues that shaped how the social media data was collected and managed. Despite users' posts on social media platforms being seemingly public and therefore freely available for use, researchers should consider whether or not their research using this data is being conducted in an ethical way.

Legal considerations in social media research should ensure that any use of data from a social media platform is in line with the terms and conditions of the platform. Additionally, legislation surrounding European General Data Protection Regulation (GDPR, 2019) should also be taken into consideration, as social media data can include personal information. Personal data can be processed for research under Article 6 (1) (e) of the GDPR: "*Processing is necessary for the performance of a task carried out in the public interest"* and Article 9 (2) (j) of the GDPR: "*Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes in accordance with Article 89 (1).*" However, researchers should consider the ways in which personal data can be protected in a research project in line with GDPR requirements. For example, this could include using aggregate-level data to minimise the risk of an individual being identifiable.

As well as legal considerations, some of the key ethical areas of concern with regard to using social media data in research include user privacy (is a user's post private or public?), informed consent (users are not aware of their information being used in research), anonymity (it is difficult to anonymise individual user posts) and risk of harm (a user's identity could be determined from their social media activity). Traditional ethical frameworks provide some guidance that can assist researchers, however social media data brings new contextual challenges which the more traditional approaches to ethical standards may not be equipped to deal with and there is, as yet, no clear ethical framework for researchers using this data source (Townsend & Wallace, 2016).

When using social media data researchers should therefore consider the fairness of using social media for research, the FAIR principle of effective data management (see below), and good research practice for using social media data in an ethical way (Leonelli *et al.*, 2021). While based on a review of ethical considerations for using social media in health-related research, these principles are based on a review of research ethics across disciplines and therefore transferable to everyone using social media for research. In particular, with regard to fairness, the guiding principle to consider is whether or not a social media user would consider it reasonable that their information be used for research or, in the context of social sensing, to provide information that may support situational awareness during a crisis situation (Kennedy, 2016; Galbraith, 2017). GDPR legislation also recognises that it is not always practicable to contact each social media user to seek consent (UK Research Integrity, 2018).

The FAIR data principles, relate to the management and handling of research data and provides a guide for researchers working with datasets. To work with data in a FAIR way, data should be: easily *Findable*; *Accessible* to as many as possible, in ways that are user-friendly and machine-readable; *Interoperable* to foster links with other data; and *Reusable* (i.e. easy to re-purpose) (Mons *et al.*, 2017). As a result of social media platforms' data privacy rules and the right for a user to delete their posts and information on a platform, it is not always possible to for research using social media data to follow the traditional principles of FAIR data management. This is because social media data may no longer be available to view online or download via an API making it difficult for studies to be scrutinised and replicated. Researchers should also take care when storing social media data for this reason.

Moving towards an ethical framework for using social media data, there are certain measures proposed by Townsend & Wallace (2016) that should be considered when using this data source:

- Do the terms and conditions of the social media platform allow researchers to access the data?
- Do the terms and conditions that users sign up to when joining a social media platform include the fact that their information can be accessed by researchers? And therefore, can the social media user reasonably expect to be observed by strangers?
- Will the social media user be anonymised in published outputs?
- Does the social media platform allow the publishing or sharing of the dataset?

Additionally, to ensure that use of social media data is in line with fairness data principles set out by Leonelli *et al.* (2021), researchers should also consider:

- Demonstrable understanding of the populations from which data are sampled so as to understand any potential bias in results obtained from the research;

- Is there an appropriate data management plan in place to preserve the dataset for later re-use in with FAIR data principles?

Therefore, to ensure that ethical issues are addressed, researchers and those wishing to use social media as an information source should ensure that research and use of social media data is in line with the above proposed framework.

3.4 Discussion

Having reflected upon the emerging themes from previous research around the social sensing of weather events, a number of gaps in the literature and themes for further exploration have been identified. In particular, building on previous work to refine machine learning methods for assessing social media content for relevance to impacts; testing methods for multiple events, rather than a single event case study; validation of methods to detect impactful weather events using social media data by comparing results with existing databases; and applying methods for other weather hazards where social media data has not been as extensively explored (e.g. temperature extremes). Furthermore, developing methods to infer location from social media posts which do not contain geospatial information will be important for using social media content for impact assessment and situational awareness. Methods to determine the types and severity of impacts as a result of weather events using social media content have already begun to be explored by a number of researchers. However, further development of these methods, in particular for determining impact severity, will be another important feature for application of using social media data as a source of impact information.

The ethical considerations for using social media data in research have also been explored. Suggestions have been made for ensuring that research using social media data build such considerations into the research design.

Testing and validating methods to interrogate social media data for impact information during weather events is therefore an extremely important focus for future research. The aim being to be able to apply these methods in an operational setting for situational awareness and impact forecast validation. The following experimental chapters (Chapters 4-6) will therefore explore the identified themes and research questions. Chapter 4 explores the effectiveness of social sensing methods when applied to multiple events by applying processes to filter social media content for relevance and location, and then analysing this content for its usefulness as a source of impact information for multiple storm events in the 2017/18 UK/Ireland storm season. Chapter 5 tests and validates social sensing methods by comparing social media activity for multiple rainfall events across the world in 2017 against a manually curated impact database for the same period of time. Chapter 6 considers the application of social sensing methods for a less extensively explored weather hazard by examining social media content during heatwave events in 2019 for three European cities

Introduction to Chapter 4

Having established the need for impact observations relating to weather events and the potential of social media as a source for this information, Chapter 4 considers these issues in a paper published in the Meteorological Applications journal in 2020 (Spruce *et al.*, 2020). This research focused on determining the type of impact information available from Twitter data during named storm events in the UK and Ireland during the 2017/18 storm season. A storm is named when it has a high likelihood to cause moderate to severe impacts and therefore provides a good example of an extreme weather event for the UK/Ireland from which to assess impacts from social media.

The motivation behind this study was to explore the social impacts of named storm events which can be determined from social media. This required the application of automated methods to extract relevant social media data relating to this particular weather hazard and the use of a location inference process for social media posts where geospatial information is not available. Further analysis of the relevant social media content would then need to be examined to determine how the type of information available from this data source changes in the period before, during and after the weather event. The type of impacts observed using social media may include, for example, reports of property damage or disruption to travel. To confirm this, topics of discussion in the social media content would therefore need to be explored. The severity of impacts may also be determined by examining the sentiment of social media content. Therefore, application of a suitable methodology to calculate sentiment will also provide some insight into how this is affected on social media during an extreme weather event..

Successful relevance filtering and location inference methods were outlined in a study by Arthur *et al.* (2018). In this study the authors used Twitter data to find the spatial impacts of flooding in the UK, applying various stages of filtering to tweets for relevance to flooding, including the utilisation of a supervised machine learning process to further refine tweets for relevance. Location inference methods described in this study found that the location of approximately 70% of tweets could be inferred using place name mentions in tweet text and/or the user profile location. They also provided a measure to normalise tweet activity in relation to the propensity for tweets relating to flooding in that particular location and population. Their results were verified against a database of known flood activity, finding that their methods successfully detected flood events in the UK using tweet data. Therefore, the success of this study provided motivation to apply similar methods to other weather hazards which cause significant impact to the UK, such as named storms.

However, a weather hazard, such as a named storm is likely to cause more widespread and wider ranging impacts than flooding. Therefore, analysing the content of storm-related tweets may also provide useful impact information. Revisiting the literature, there are a number of studies focusing on the type of impact information available in tweets throughout the lifecycle of a significant storm event. In particular, studies relating to the information available from social media during hurricanes in the USA provided some inspiration for a study focused on storm events in the UK.

Lachlan *et al.* (2014) examined the volume and content of tweets during the period to Hurricane Sandy finding that tweet activity increases significantly in the days leading up to and during the hurricane event. Building on this work, Spence *et al.* (2015) consider Fink’s model (Fink, 1986) which categorises the “crisis life cycle” into four stages: the prodromal stage (period of build up to an event), acute stage (when the event takes place), chronic stage (directly after the event), and termination stage (when an event terminates). The authors consider the prodromal stage of Hurricane Sandy and analyse tweets during this period to understand how information circulated via Twitter changes as a natural disaster moves from a prodromal to an acute stage. Categorising tweets by impact type based on its content they analyse the change in type of tweets over the period of the prodromal stage of the hurricane. The authors find that the proportion of tweets containing humorous content increases substantially as the prodromal stage moves towards the acute stage, however this then turns to expressions of genuine fear as the acute stage nears and the threat is recognised as real. However, interspersed within these emotional responses are also important information being circulated by official agencies which can often be lost amongst the number of other types of commentary on Twitter about the event. What this study also demonstrates is that an understanding of change in sentiment leading up to and at the onset of a natural disaster can be determined using Twitter posts. Therefore, examining information from the content of social media, such as sentiment or topics of discussion, during the period before, during and after a natural hazard event can also provide an insight into the social impacts over and above quantifying the damages or area affected by the event.

Identifying topics of discussion from the content of social media during natural hazard events was also explored by Imran *et al.* (2016). By examining and manually categorising the type of information available in tweets during a number of different natural hazard events they provide both a methodology for categorising tweets and examples of types of impact information that may be shared by users on Twitter. Furthermore, Baylis *et al.* (2018) explore the relationship between the sentiment of Twitter posts and weather conditions (such as temperature, precipitation, cloud cover and humidity), finding that worsening weather conditions result in more negative sentiment of tweets. Therefore, calculating sentiment of tweets during a significant weather event, such as a named storm, may also provide a measure of impact.

Considering the previous work outlined above provides the motivation to apply similar methods and approaches for determining impacts as a result of named storm events in the UK using social media (Twitter) data, which will now be outlined in Chapter 4.

Chapter 4 - Using social media to measure impacts of named storm events in the UK and Ireland

Abstract

Despite increasing usage of impact-based weather warnings, the social impacts of extreme weather events lie beyond the reach of conventional meteorological observations and remain difficult to quantify. This presents a challenge for validation of warnings and weather impact models. This study considers the application of social sensing, the systematic analysis of unsolicited social media data to observe real-world events, to determine the impacts of named storms in the UK and Ireland during the winter storm season 2017-2018. User posts on Twitter are analysed to show that social sensing can robustly detect and locate storm events. Comprehensive filtering of tweets containing weather keywords reveals that ~3% of tweets are relevant to severe weather events, and for those, locations could be derived for about 75%. Impacts of storms on Twitter users are explored using the text content of storm-related tweets to assess changes in sentiment and topics of discussion over the period before, during and after each storm event. Sentiment shows a consistent response to storms, with an increase in expressed negative emotion. Topics of discussion move from warnings as the storm approaches, to local observations and reportage during the storm, to accounts of damage/disruption and sharing of news reports following the event. There is a high level of humour expressed throughout. This study demonstrates a novel methodology for identifying tweets which can be used to assess the impacts of storms and other extreme weather events. Further development could lead to improved understanding of social impacts of storms and impact model validation.

4.1 Introduction

It is well known that extreme weather events such as strong winds, heavy rain and snow cause impact and disruption to our daily lives (IPCC, 2014). However, there is little observational record of the specific impacts (e.g. damage to property, disruption to travel, danger to life, stress and anxiety) that occur as a result of these weather events. This information lies beyond the scope of traditional meteorological observations. The frequency and intensity of extreme weather events has increased over recent years and is predicted to continue to increase (IPCC, 2014). Meanwhile, there has been a shift from forecasts that focus on meteorological conditions alone to forecasts that incorporate information about their associated impacts (Taylor, 2018). This impact-based forecasting strategy is endorsed by the World Meteorological Organisation (WMO), who have produced guidance to support its development (WMO, 2015). Together, these trends create an urgent need to understand the ways in which extreme weather events affect people and property, to validate forecast models and warning systems.

Social media is increasingly used across the world (Clement, 2017) and this presents an opportunity to utilise the rich social information it creates to inform preparedness and response to natural hazard events. Many people routinely use social media to discuss weather conditions, particularly when weather patterns are unusual. During crisis events, such as periods of extreme weather, technological challenges in affected areas may slow official news correspondent reports, while social media reports may be more swiftly distributed (Spence *et al.*, 2015). The public availability of data from some social media platforms, notably Twitter, opens the possibility to use social media data to understand how human activity is affected during an extreme weather event.

'Social sensing' utilising social media has been widely used for knowledge discovery in fields relating to public health, human behaviour, social influence and market analysis (Wang *et al.*, 2016). Social sensing broadly refers to a set of sensing and data collection models whereby data are collected from humans or personal devices (Wang *et al.*, 2015). In this paper, social sensing using unsolicited social media data is distinguished from solicited crowd-sourcing, where users voluntarily participate and report observations in a structured or semi-structured manner. Examples of solicited crowd-sourcing include the UK Met Office 'Weather Observations Website' ("WOW. Met Office Weather Observations Website," 2019), where the public can provide amateur weather observations, and the UKSnowMap ("UK Snow Map," 2010), where Twitter users are asked to report snowfall observations using a particular hashtag (#uksnow). While solicited crowd-sourcing offers benefits in that data is more reliable and can be provided in a structured form by a set of dedicated volunteers, the volumes of data generated are typically low relative to the high volumes seen in unsolicited social media use; this can limit the usefulness of solicited data for understanding of wider impacts.

For social sensing using unsolicited social media, each individual user plays the role of a sensor. When a user publicly posts an item to a social media platform, they are providing a piece of sensor data. When grouped together by topic or location, large numbers of social media posts can therefore be used to develop an understanding of a range of issues. Social sensing of this nature has already been successfully used to detect natural hazards such as earthquakes (Sakaki *et al.*, 2010), wildfires (Boulton *et al.*, 2016) and floods (Tkachenko *et al.*, 2017; Brouwer *et al.*, 2017; Rossi *et al.*, 2018; Arthur *et al.*, 2018). A number of studies have used social media to understand impacts of hurricanes in the USA (Guan & Chen, 2014; Cervone *et al.*, 2015; Kryvasheyeu *et al.*, 2016; Morss *et al.*, 2017; Wu & Cui, 2018; Kim & Hastak, 2018).

This study explores whether social sensing can help meteorologists to understand how human activity is affected during extreme weather events, in terms of both emotional impacts and other social impacts (e.g. disruption, damage) revealed by the topics of conversation during storm events. Some weather-related studies have begun to explore this opportunity. The effects of weather on mood have been shown using sentiment expressed in tweet text linked to weather conditions (Hannak *et al.*, 2012; Caragea *et al.*, 2014; Li *et al.*, 2014; Baylis *et al.*, 2018; Li Hu and Jadidi, 2019). The categorisation of tweet content related to weather and natural hazards has also been explored both using manual methods (Spence *et al.*, 2015; Halse *et al.*, 2018) and automated methods (Alam *et al.*, 2018b). However, to date there has been little exploration of social sensing focused on social impacts of weather for the purposes of impact-based forecast validation.

In this study, data from the social media platform Twitter was collected during the 2017/2018 UK and Ireland storm season (approximately October-March) to explore social sensing as a methodology for assessing the social impacts of storms. The research uses and builds on the social sensing methods described by Arthur et al., (2018) to extract, filter, locate and get useful meaning from social media data collected during this storm period. Sentiment analysis is used to look at the aggregated emotional response to storms and how this changes during the period of a storm event. Categorisation of storm-related tweet content provides an indication of what kind of information can be determined from tweets, looking in particular for content related to social impacts. The aims of the study are to: (i) establish a methodology for social sensing that can provide useful information about social impacts of storms; (ii) apply the methodology to explore the impact of storms in the UK & Ireland during winter 2017/2018. These objectives are intended to help develop social sensing as a source of impact observations suitable for validation of impact-based weather forecasting systems.

The paper is split into the following sections: Data Collection & Methods outlines the methods used for data collection, filtering, and content analysis; Results reports the main findings of the analysis, focusing on sentiment and categorised impacts observed during storm events; finally the Discussion summarises the main benefits and limitations of the social sensing approach as demonstrated in this study, and makes some suggestions for future research.

4.2 Data Collection & Methods

This study uses a hybrid approach of methods from previous studies which successfully collected and found useful meaning from Twitter data relating to weather events or natural hazards (Lachlan *et al.*, 2014; Halse *et al.*, 2018; Arthur *et al.*, 2018; Cowie *et al.*, 2018). Social media data was collected, filtered for relevance and geo-located. The content of the resulting dataset was then analysed using sentiment analysis and automated categorisation.

4.2.1 UK/Ireland Storm Season 2017/2018

Since 2015 the Met Office in the UK and Met Éireann in Ireland have used a storm naming system to raise public awareness of the effects of stormy weather with the public and to increase preparedness in response to weather extremes. A storm is named if it is expected to cause 'medium' or 'high' impacts from wind and/or precipitation, i.e. storms will be named for weather systems which are expected to have an Amber or Red weather warning issued by Met Éireann and/or the Met Office's National Severe Weather Warning Service (NSWWS) (https://www.metoffice.gov.uk/news/releases/2017/storm-names-for-2017-18-announced). Weather warnings are colour coded in response to their potential impact and likelihood; amber and red warnings are therefore issued for weather events which are both probable and likely to cause significant disruption.

In the 2017/2018 UK storm season, which generally runs from autumn to early spring, there were a number of named storms which affected the UK with expected medium or high impacts from wind and/or rain/snow (Table 4.1). The reason for naming storms is to improve public communication about weather events likely to cause significant impacts. Named storms are likely to attract attention from social media users because of their severity and the use of the names in official communication and forecasts. Named storms are also useful from a technical point of view, as one can search directly for the storm's name. Therefore, this study mainly focuses on named storms and the impacts associated with them. As the Twitter data was collected from 16th October 2017 (when news of ex-hurricane Ophelia hitting the UK was reported in the media) for named storms for the duration of the 2017/2018 UK storm season up until 10th March 2018, post Storm Emma. Tweets containing keywords for weather related to a storm (e.g. wind, rain, etc) were also collected during this period. This was so that tweet activity which included weather terms only could be compared with tweets relating specifically to named storms.

Storm Name	Date Named	Date of Impact on UK
Aileen*	12 September 2017	12 - 13 September 2017
Ex-Hurricane Ophelia	11 October 2017 (named by NHC)	16 - 17 October 2017
Brian	19 October 2017	21 October 2017
Caroline	5 December 2017	7 December 2017
Dylan	29 December 2017	30 - 31 December 2017
Eleanor	1 January 2018	2 - 3 January 2018
Fionn	16 January 2018	16 January 2018
David*	17 January 2018 (named by Méteo France)	18 January 2018
Georgina	23 January 2018	24 January 2018
Emma	1 March (named by the Portuguese Met Service)	28 February – 3 March 2018

*Table 4.1 - Met Office record of storm names during the 2017/2018 storm season. Information taken from: https://www.metoffice.gov.uk/barometer/uk-storm-centre (*Please note: Twitter data for Storm Aileen and Storm David were not collected for this study)*

Other countries' meteorological services may also name storms, using similar naming systems, so that some storms are already named before hitting the UK/Ireland. If a weather system has previously been named by another meteorological service, then it retains this name when it reaches the UK/Ireland. For example Ex-Hurricane Ophelia was named by the US National Hurricane Centre (NHC), Storm David by Méteo-France and Storm Emma by the Portuguese Met Service.

4.2.2 Social Media and Twitter Data Collection

At the end of 2017 it was estimated that there were 2.46 billion social media users around the world, reflecting the global usage of smartphones and mobile devices. The social media platform Twitter, having 330 million monthly active users (Clement, 2017), is a social networking and microblogging service that allows registered users to interact via short published messages (tweets) up to 280 characters in length. Twitter makes user posts freely available via the Twitter API, making Twitter a popular source of observational data for both social and natural scientists (Williams *et al.*, 2013). Data collection using Twitter can be achieved using keywords or 'hashtag' references to specific topics or events.

However, suitable algorithms must be applied to filter the data to ensure only relevant information is then taken forwards for analysis (Spence *et al.*, 2015). Locating the user who has posted an item to a social media platform is another challenge. At present only 1-2% of Twitter posts, for example, carry a GPS location or specific location coordinates (Dredze *et al.*, 2013), therefore other methods must be employed to infer the place of origin.

Using the methods outlined by Arthur et al., (2018), tweets relating to named storms and storm-associated weather conditions were collected using the Twitter Streaming API (via a Python script utilising the Twython package (McGrath, 2013)). This API returns all tweets up to a limit of 1% of the total volume of tweets at any point in time. Search keywords were used as an initial filter applied by the API to identify and download relevant tweets (Table 4.2). As tweets using these keywords are unlikely to reach the API limit, it is believed that most, if not all relevant tweets are downloaded using this method (Morstatter *et al.*, 2013). Some storm names were prone to typing errors in tweets, therefore some common variants were accounted for in the search terms used. Only tweets in the English language were collected, since the geographical areas of interest in this study (UK and Ireland) are majority English speaking. Tweets were collected over the time period 16th October 2017 – 10th March 2018. Each tweet was saved as a JSON object which is a lightweight data-interchange format often used for transmitting data from a server to a web application (https://www.json.org). Each JSON object contains the tweet text as well as a number of meta-data fields relating to each tweet (i.e. timestamp, username, user location, geotag, etc).

Collection	**Keywords**
Wind	*wind, gale, windstorm, hurricane*
Precipitation	*rain, raining, rainy, rainstorms, rainstorm, hail, hailstones, hailstorm, hailing, hale, snow, blizzard, snowstorm,*
Storm Names	*storm, ophelia, ofelia, opelia, opehlia, opheliaireland, brian, caroline, dylan, eleanor, fionn, fion, georgina, emma*

Table 4.2 - Twitter collections referred to in this study. Only tweets containing one or more of the keywords shown were added to the initial unfiltered dataset for each collection.

The storm name collection keywords are shown in Table 4.2. Storm names were added to the "Storm Names" data collection in the days leading up to each storm event and therefore collections for each storm name do not cover the whole of the study period. As wind is the main weather type to cause impacts during a storm event, tweets relating to wind were collected as well as storm names. Precipitation also causes impacts during a storm event, however weather warnings relating to each of the named storms predominantly related to the impact of winds, rather than precipitation. It is also likely that there were precipitation events (snow or heavy rain) not related to storm activity which makes the precipitation dataset less comparable with the storm dataset. Therefore, while tweets relating to precipitation were also collected and filtered for relevance, the crucial comparison is between the storm tweet collection and the wind tweet collection. More than 100 million tweets were collected from the API during the 2017/2018 storm season (see Table 4.3).

	All Tweets in raw data collection (unfiltered)	**Tweets remaining after filtering for relevance**		**Tweets remaining after filtering for relevance AND location inference**		
Tweet Collection	**Number of Tweets**	**Number of Tweets**	**% (of All Tweets)**	**Number of Tweets**	**% (of All Tweets)**	**% (of Tweets after filtering for relevance)**
1. Precipitation	**67,448,047**	**3,264,573**	**4.8%**	**1,982,378**	**2.9%**	**60.7%**
2. Wind	**26,298,449**	**831,076**	**3.2%**	**472,586**	**1.8%**	**56.9%**
3. All Storm names	**8,101,901**	**278,412**	**3.4%**	**214,220**	**2.6%**	**76.9%**
ophelia	897,054	214,730	23.9%	167,369	18.7%	77.9%
brian	2,037,045	12,970	0.6%	9,439	0.5%	72.8%
caroline	1,199,149	8,552	0.7%	4,993	0.4%	58.4%
dylan	2,504,264	3,907	0.2%	2,410	0.1%	61.7%
eleanor	555,433	11,872	2.1%	9,761	1.8%	82.2%
fionn	43,936	1,260	2.9%	878	2.0%	69.7%
georgina	104,327	894	0.9%	650	0.6%	72.7%
emma	760,693	24,227	3.2%	18,720	2.5%	77.3%

Table 4.3 - Total number of tweets collected within each collection and remaining after applying both relevance filter and location inference.

Figure 4.1 shows time series of the numbers of tweets containing the specified keywords collected per day during the period 16/10/2017 – 10/03/2018. This includes all tweets (including retweets) in the raw dataset prior to any filtering for relevance to named storms. The time period of each named storm in the collection period is shown by the grey bars. There appear to be associated peaks in Twitter activity relating to Wind discussion. Peaks in the Storm Name collection are less obviously associated with storm events, but inspection suggested that this collection contained some highly relevant content amongst a lot of irrelevant content, which is likely to confound the association. The Precipitation collection has some storm-associated peaks but also many peaks not associated with storm events.

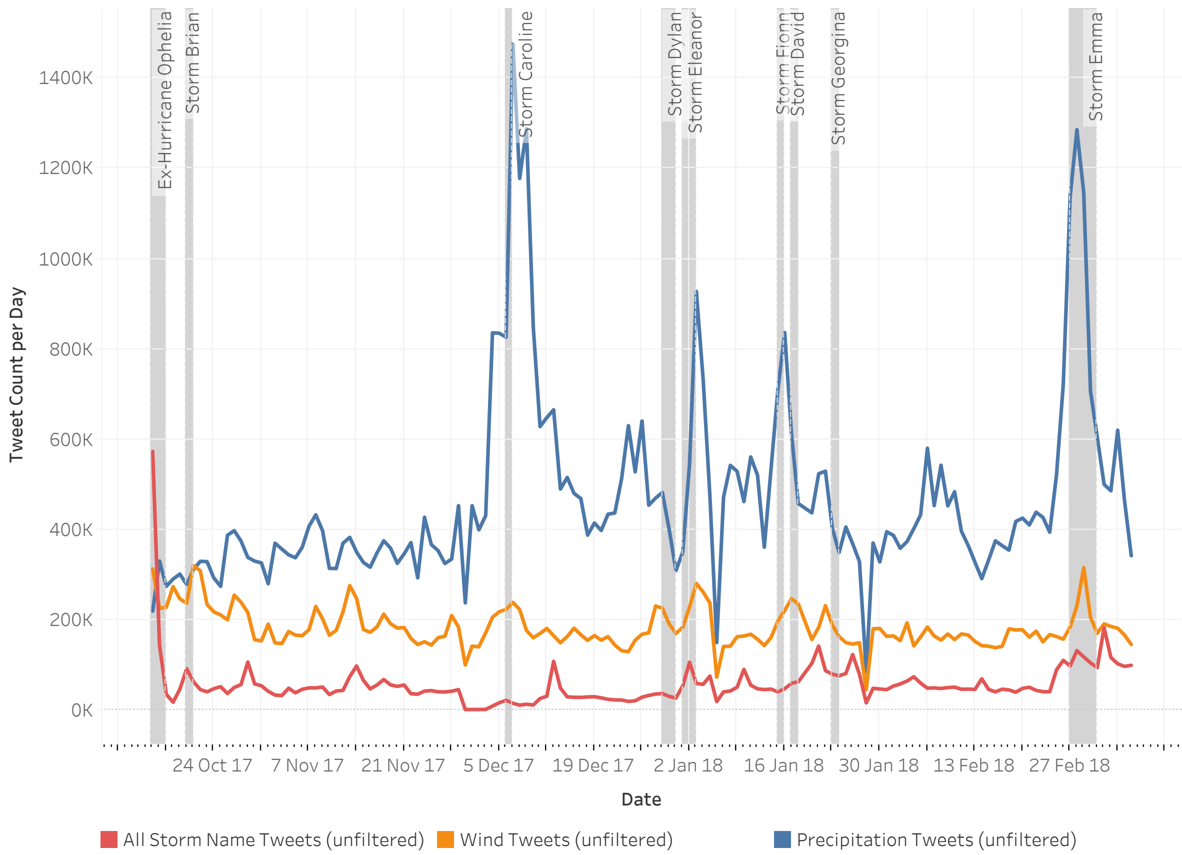

Figure 4.1 - Number of tweets collected per day during collection period 16/10/2017 to 10/03/2018. The period of each named storm is shown by a grey bar.

This study is concerned with the social impact of storms as experienced by social media users. For this purpose, retweets are retained in most parts of the analysis, including counts and timeseries measuring total activity around storms, and sentiment analysis (where it is asserted that retweeting implies endorsement, approval or agreement with the sentiment expressed in the original tweet). For purposes of observing social impacts, retweets and “quote” tweets are removed as they do not represent original observations. This removal was performed using tweet metadata. Retweets made up 63% of the dataset after filtering for relevance and quote tweets 6%.

4.2.3 Filtering and Location Inference

After data collection, the first stage in processing the Twitter data was to apply a suitable relevance filter to remove any obviously irrelevant data. The various filters applied can be split into the following stages which are described in the order in which they were applied:

Timezone Filter

The raw data collection contains tweets from all global locations including the US and other countries. Only tweets which relate to weather activity in the UK and Ireland are of interest for this study, therefore the dataset is first filtered based on the timezone entity of each tweet to remove international tweets. The use of timezone as a proxy for the country level location of a tweet is discussed by Schulz et al. (2013) who found that over 80% of tweets can be accurately localized to a country using the timezone entity. Tweets with the following timezones are therefore kept in the dataset: *GMT, London, Europe/London, UTC, BST, GMT+1, Dublin, Europe/Dublin, Edinburgh.*

As of May 2018, in order to comply with General Data Protection Regulation (GDPR) requirements, Twitter has removed the timezone field from tweet metadata (Cowie *et al.*, 2018). Other methods for location inference (as described below) remain effective in the absence of timezone information. This filter removes approximately 90% of tweets in the raw data collection and therefore makes later processing steps more computationally efficient.

Bot Filter

\`Bots' are automated user accounts that are set up to perform a particular function, such as collate/spread content from a set of sources, promote a particular view, or deliver advertising. Automated tweets from bot accounts are highly unlikely to contain information relating to social impacts of weather activity, but the presence of this kind of content can distort the dataset. To remove bot content, the number of tweets by each user account was calculated for the entire dataset. User accounts with a disproportionately high number of tweets (in this case >1% of the total volume of tweets in the dataset) were identified as bot accounts; automated accounts tend to create significantly more tweets than human users. All tweets posted by bot accounts were then removed from the dataset. A further manual review of the remaining users generating a high proportion of tweets found some additional bot accounts which were also removed. This filter removes approximately 1% of tweets in the raw data collection.

Weather Station Filter

Data collections containing weather-related terms include a high number of tweets automatically posted by amateur weather stations. As this study is focused on social impacts, these tweets are deemed irrelevant since they are not directly related to social impacts. A process was developed to remove them. Tweets from weather stations typically follow a fixed structure, for example: '*Wind 2.0 mph E Barometer 30.10 in Falling slowly Temperature 68.5 F Rain today 0.00 in Humidity 55'.* Here these were identified using a script that searches the text of a tweet and counts weather-related terms; if there were more than 2 weather-related terms the tweet was identified as a weather station tweet. This method was shown to work well by manual inspection. Tweets identified as being from weather stations using this method were removed from the dataset. This filter removes a very small number of tweets in the raw data collection for named storms, however removes approximately 1% of tweets in the raw data collections for wind and precipitation.

Irrelevant Term Filter

As for the weather station filter, this filter is more relevant to the data collections containing weather related terms, rather than storm names. There are many phrases in the English language which use weather-related terms but do not relate to weather, as well as some homographs for weather-related words; these are irrelevant to this study so tweets that contain them were removed using a look-up table method. A list of common terms or phrases which use weather-related terminology but are clearly not referring to a weather event (such as: '*wind up*', '*throw caution to the wind*', '*cook up a storm'*, *etc)* were identified in tweet text and those tweets removed from the dataset. This filter removes a very small proportion of tweets from the remaining raw data collection.

Machine Learning Relevance Filter

Although the previous stage removed much irrelevant content, an additional stage of filtering was still necessary to remove tweets which included the search keywords but were not relevant to wind, precipitation and storms. These included (e.g.) business advertising, links to articles on other topics, references to people and places who shared a name with the storm, and various other irrelevant content. Tweets in the Storm Names collection were particularly in need of additional filtering, since there are many celebrities or other individuals that share the same names as the storms studied here. To achieve this the methods used successfully in previous studies (Arthur *et al.*, 2018; Cowie *et al.*, 2018) were employed.

A set of 6000 tweets were randomly selected from the tweet collections. Each tweet in this set was then manually labelled as relevant or irrelevant. Manual coding was conservative, labelling as irrelevant tweets that were obviously un-related to the study topic and also tweets which were ambiguous (i.e. providing insufficient information to decide on relevance). In total there were 1495 tweets in the dataset labelled as relevant and 4505 tweets labelled as irrelevant. The labelled dataset was then used as training data for a Multinomial Naïve Bayes classifier. As a first validation test for this approach, 25% of the data was held back as a validation set and a classifier was trained on the remaining 75% of cases; this classifier had accuracy (i.e. correctly identified the relevance/irrelevance) of 92% on the held-back validation tweets, with an F1 score of 0.84. As a second test, to confirm the robustness of the approach, the same training/validation test was repeated with 6-fold cross-validation. The results of each test were combined to give an overall mean F1 score of 0.80 and the summed confusion matrix (also known as ‘contingency table’) below (where True is relevant and False is irrelevant):

$$\begin{pmatrix} & & \multicolumn{2}{c}{Predicted} \\ & & False & True \\ Actual & False & 4301 & 204 \\ & True & 274 & 1221 \end{pmatrix}$$

This confusion matrix shows overall accuracy of 92%, with most tweets in the filtered dataset classified as not relevant. Accuracy was higher on the False class (4301/4505=95%) than on the True class (1221/1495=82%), with a slight tendency to mis-classify relevant tweets as irrelevant. This could be attributed to the training dataset being unbalanced and biased towards irrelevant tweets. However, this is a conservative error that ensures tweets that are retained are highly likely to be relevant. This is likely due to the wide variety of tweets in the

Storm Names collection which were not related to named storm discussion. The Multinomial Naive Bayes classification approach was deemed to be accurate enough and sufficient for the purposes of this study based on the results discussed above. A new classifier was then trained on the entire set of manually coded tweets to take forward as the relevance filter for this study. As an additional check of the performance of this classifier, random manual checks of the data after this filter was applied to the whole tweet dataset confirmed that it was performing well.

The Bayesian filter described above removes a further 4-5% of tweets in the data collection for named storms and approximately 2% of tweets in the wind and precipitation data collection.

Table 4.3 shows the number and percentage of tweets remaining for each tweet collection after the stages of relevance filtering described above have been applied. Overall, there are 3-4% of tweets remaining after relevance filtering. Supplementary Table B.2 provides a more detailed breakdown of the number and percentage of tweets removed at each stage of relevance filtering for each tweet collection.

Location Inference

After relevance filtering is completed, each tweet in the dataset is also processed to identify if it can be located using information contained within the tweet. The spatial distribution of tweets relating to the weather would also give an indication of social impacts in particular locations.

As found in other studies, this study also finds that only ~1% of tweets contain geo-coordinates of the tweet origination. Therefore, a location inference method is required. Using the same location inference approach as the one outlined by Arthur et al. (2018), the filtered tweet dataset was examined for different kinds of geographical information: geo-coordinates (geotag), the place a user designated in the Twitter application when posting (place), the location given in the user profile (user location), and place names mentioned in the tweet text. This method is based on the location inference method validated by Schulz, Hadjakos, Paulheim, Nachtwey, & Uhlhäuser (2013) who found 92% accuracy when inferred location was compared against tweets for which a geotag is known.

Thus, there were 4 tweet elements examined for location information in the following order:

- ***Geotag***: Locate tweets using Geotag (GPS coordinates)
- ***Place*****:** Location of tweet (polygon coordinates)
- ***User location***: (GPS coordinates, if not lookup text with Geonames db)
- ***Place names mentioned in tweet text***: (dbpedia spotlight (Mendes *et al.*, 2011) lookup identifies place names and coordinates)

It was found that the most useful elements of a tweet which can be used to determine a location are the user location and place name mentioned in the tweet text. Table 4.3 shows the number and percentage of tweets in the filtered dataset for which a location can be found for each tweet collection. On average 77% of filtered tweets could be located using this inference method. Here “located” means that a tweet was allocated to a defined spatial area with high confidence. Supplementary Table B.3 provides more detail on the specific numbers and proportion of tweets located by each tweet element for each tweet collection.

Further details on the processes and packages used in the social sensing code can be found in Appendix C.

4.2.4 Results of filtering and location inference

After applying the above methods of relevance filtering the number of tweets retained for analysis was substantially reduced. Figure 4.2 shows an example of this reduction for Storm Brian. Compared with the unfiltered data, the filtered dataset contains far fewer tweets. However, there is now a clear peak of Twitter activity of relevance to Storm Brian which coincides with the period of the storm (shown by the grey bar in the figure). The same is found for each of the named storms in the dataset (data not shown). Figure 4.3 shows tweets that were both located (using location inference) and relevant (passed the relevance filters). All other analysis uses all relevant tweets that are located to the UK and Ireland by timezone, but not necessarily precisely located using the inference process.

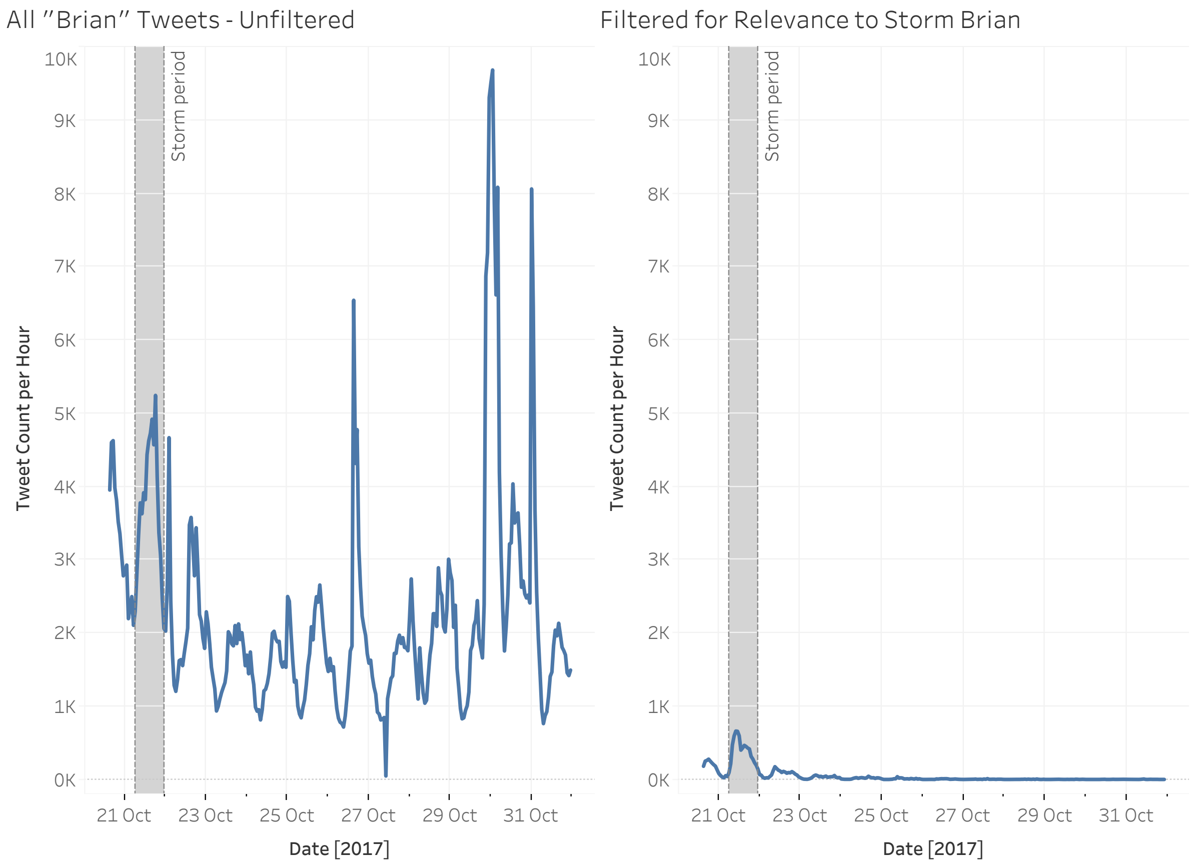

Figure 4.2 - "Brian" tweets unfiltered (i.e. all tweets containing the word "brian") versus post filtering for tweets relevant to Storm Brian.

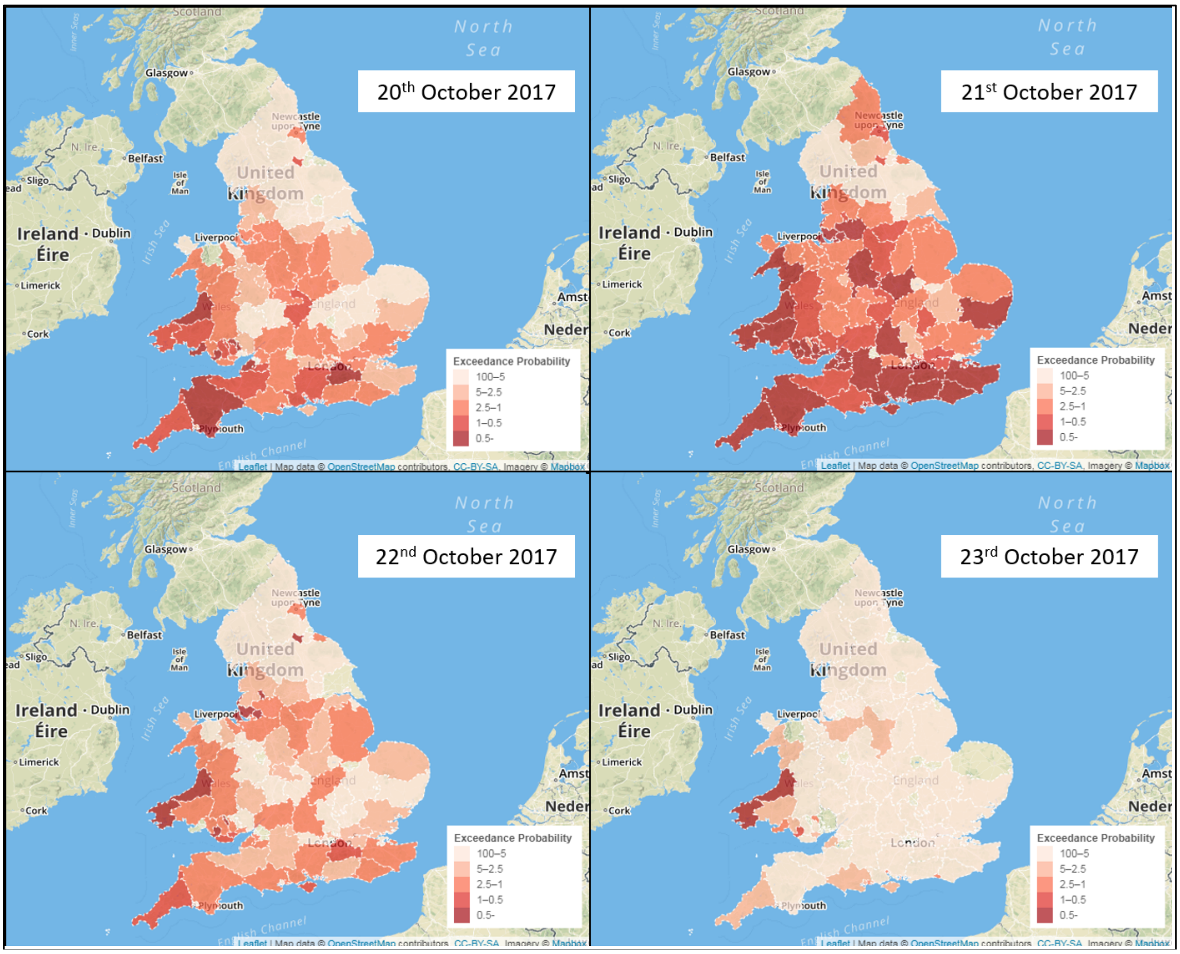

Figure 4.3 - Storm Brian tweets (after filtering for relevance) located in England/Wales and grouped by county for each day of the storm period. Storm Brian hit the UK on 21/10/2017. Shading indicates the exceedance probability for the number of tweets observed by county (i.e. likelihood of that activity level accounting for prevalence of tweet activity in that particular location). Data shown in this visualisation is restricted to England and Wales only, but data analysed in this study extends to Scotland and Ireland

Results for the Precipitation, Wind and Storm Name collections, pre- and post-filtering and after location inference, can be found in Table 4.3. Typically, <5% of tweets are retained after filtering for relevance. Interestingly this was much higher (~24%) for the dataset relating to Ex-Hurricane Ophelia. This is most likely because Ophelia is an uncommon name. Where a storm is named with a more common name (i.e. Brian, Caroline, etc) the percentage of tweets retained after filtering for relevance is much smaller because there is a higher background level of Twitter activity. Of the relevant tweets, typically 55-80% could be successfully geolocated using the inference method outlined above.

Figure 4.3 presents a case study of located tweets in England and Wales by county, as an example of the social sensing technique. This case study shows the spatial extent of tweet activity in England and Wales for Storm Brian following application of location inference. Tweets located in Scotland, Northern Ireland and Ireland are not shown in this figure, but were included in other analyses. Darker shading indicates where there was more Twitter activity for a particular area than average for that location, plotted as an exceedance probability. Probability of exceedance is a statistical metric describing the probability that a particular value will be met or exceeded (McMahan *et al.*, 2013). In this example, this provides the likelihood of recording a given number of tweets about storms in this particular location, based on the frequency distribution of observed counts across the whole storm collection dataset. This provides geographical information on where the storm is being most discussed on Twitter and therefore an indication of which areas of the country are likely to be most affected by the storm. In this example for Storm Brian, more significant tweet activity can be seen in the West, South and Southwest of England and Wales. It also shows how the spatial pattern of tweets changes over time during the period leading up to, during and after the storm. As anticipated, there is a peak of activity on the day of the storm, which quickly reduces in the days afterwards.

Once both relevance filtering and location inference were completed, the dataset was then prepared for further analysis to determine information on social impact from the tweet data. All filtered tweets' text was used for sentiment and content analysis.

4.2.5 Sentiment Analysis

The 'sentiment' of a tweet measures the net level of positive or negative emotion it expresses. In this case, following various studies that use sentiment analysis with tweets to examine collective mood related to weather conditions (Baylis et al., 2018; Caragea et al., 2014; Hannak et al., 2012; Li Hu and Jadidi, 2019; J. Li et al., 2014) sentiment analysis is used to infer the mood of Twitter users. By analysing the collective sentiment of tweets during the period of a storm event, the aim is to get an indication of the emotional impact of the storm.

Tweet text was analysed using the sentiment analysis package TextBlob (Loria, 2010). This Python package is a popular lexicon-based sentiment analysis tool well-suited to the relatively short text strings found in tweets. In preliminary work, Textblob was tested against another leading sentiment package, VADER (Hutto

& Gilbert, 2014), which gave comparable results. Since there was no substantive difference, Textblob was preferred for ease of use with this dataset.

The TextBlob package returns a sentiment polarity value between -1 and 1, where <0 implies negative sentiment and >0 implies positive sentiment. The value returned is based on a sentiment classifier trained on a large dataset of text relating to movie reviews tagged as positive or negative. The sentiment polarity score for each tweet is based on all words in the tweet text. Figure 4.4 provides examples of tweets with sentiment score calculated using TextBlob.

Category	**Tweet Text Examples**	**Sentiment Score**
Humour	*"Brian? What kind of name is that for a storm? Everyone knows Brian is a snail."*	0.60
	"Am I the only one to find it really hard to take a storm called #Brian seriously?	-0.21
	"And Brian? Really? Storm Rambo or Terminator would be far better than #StormBrian"	0.27
Damage	*"This is the scene this morning as the waves have damaged the Harbour Office during Storm Brian."*	-0.26
	"Storm Brian damage causes floodlight damage. Revised home game vs @ChesterCityFC"	-0.40
	"Scaffolding in Helsby High Street BLOWN OVER by #StormBrian high winds"	0.00
Disruption	*"Train delay: National Rail have warned of delays due to high winds from Storm Brian"*	-0.25
	"Storm Brian latest - tree blocks railway lines and hovercraft suspended"	-0.41
	"Major motorway was CLOSED after Storm Brian floods carriageway"	-0.02
Warnings	*'#StormBrian could lead to travel disruption this weekend.'*	-0.06
	'Storm Brian set to batter UK with heavy rain and 70mph winds.'	-0.20
	'Take care on the coast folks. Waves are quite high with #StormBrian'	0.16
Observations	*"It's really windy out there!"*	0.20
	"Storm Brian seems to have arrived now…"	0.00
	"Storm Brian just brought in the heaviest rain shower I've ever seen.....it really scared our 2 cats."	-0.29
News	*"Storm Brian makes landfall on west coast"*	0.00
	"Storm Brian: 'Weather bomb' storm set to batter Britain and Ireland - Earth News"	-0.11
	"Live updates on Storm Brian across Somerset at https://t.co/r49v2aUKFc"	0.00
Other	*"Will I be ok putting my bin out tomorrow or will Brian still be raging ?"*	-0.05
	"Had afternoon tea with this cutie in a cosy cottage as Storm Brian rattled the windows. #friends #halftermbreak"	0.00
	"No sign of Storm Brian up here, thankfully! We have a lovely day, a great one for a long walk on the beach"	0.42

Figure 4.4 - Example of the types of tweets included in each category with sentiment score calculated using the TextBlob package. These are synthetic tweets rather than actual tweets, in order to protect user privacy.

4.2.6 Content Analysis

Filtered tweets in the Storm Brian dataset at times of peak activity (20/10/2017 - 22/10/2017) were manually analysed and placed into one of seven categories based on their content. Only tweets containing original content (i.e. excluding retweets and quotes) were analysed for their content. Categories were determined after an initial inspection of a subsample of filtered tweets, using a similar approach to a study on the volume and content of Tweets associated with Hurricane Sandy (Lachlan *et al.*, 2014). The categories used were:

- **Humour** – Tweet contains a joke, sarcastic remark, or light-hearted commentary on experience of the storm event; does not provide any information about any impact as a result of the storm.
- **Damage** – Tweet contains information about damage to persons or property.
- **Disruption** – Tweet contains information about disruption to daily life e.g. train delays, road closures, not able to go to work.
- **Observations** – Tweet contains commentary on the weather occurring e.g. ‘wind is very strong’, ‘Storm Brian has arrived here in Balamory’.
- **Warnings** – Tweet contains information and advice about the forthcoming storm, or a warning about danger to persons or property due to the storm.
- **News** – Tweet contains reference to a media report on the storm event.
- **Other** – Tweet content relating to the storm that does not fit into the above categories.

Figure 4.4 provides examples of the types of tweets used in each category.

Categorisation of tweets was performed manually by two human coders after initial discussion and agreement of the coding scheme. In total 5961 tweets relating to Storm Brian were manually categorised. A subsample of 100 randomly chosen tweets from the filtered tweet data was used for an inter-coder reliability check. Cohen’s kappa (κ) was used to determine the agreement between the two coders’ judgement on the category of each tweet in the subsample. There was near perfect agreement between the two coders with $\kappa = .889$, $p < .0005$. This provided confidence in the categorisation coding scheme used.

Note that both text and pictures in tweets were used to assign a category, but not emojis as these were removed from the dataset to simplify text analysis processes.

4.3 Results

4.3.1 Combined Time series Plot

Tweet counts in the filtered datasets for wind and storm names were plotted over time (Figure 4.5). The time period for each storm is also shown. Peaks in the volume of tweets coincide with the (UK Met Office recorded) date of impact of storms shown in Table 4.1. Peaks in the volume of wind tweets also coincide with peaks in the volume of storm name tweets.

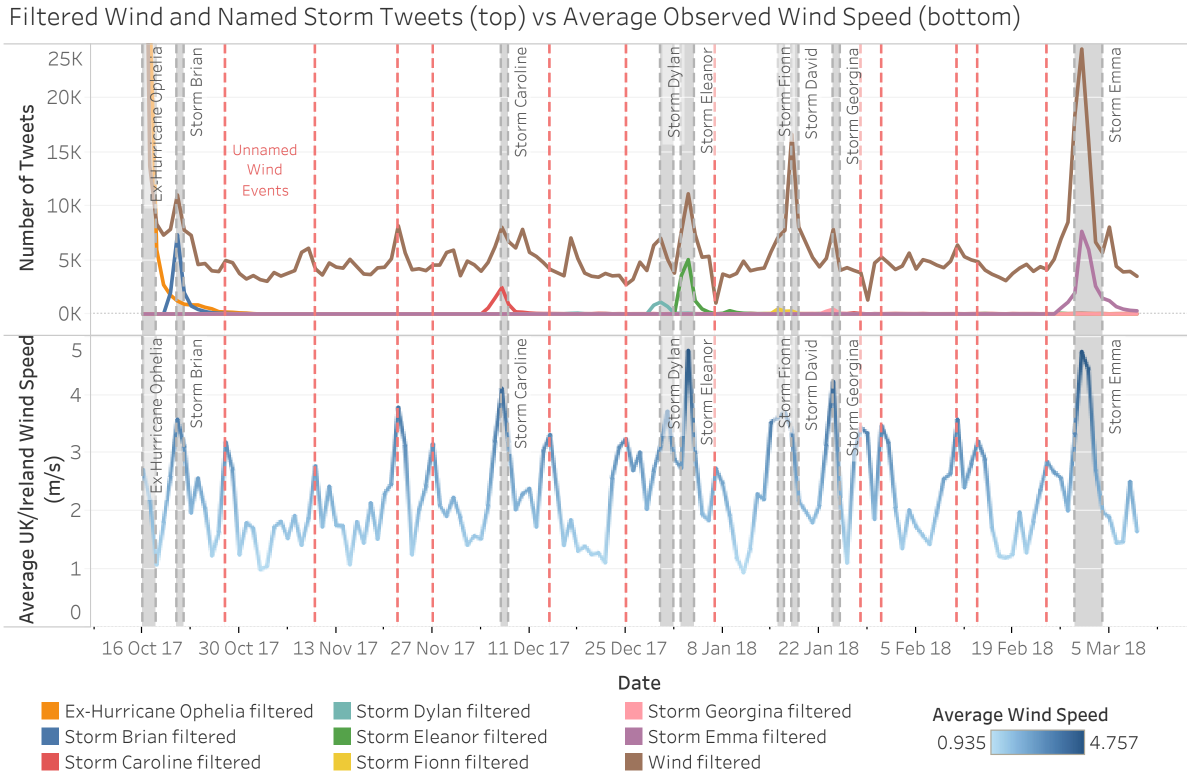

Figure 4.5 - [Top] Timeseries of the number of tweets per day for Named Storm events (after filtering for relevance) versus the number of Wind tweets per day for the 2017/2018 Storm period. Ex-Hurricane Ophelia produced very high numbers of tweets in the Named Storm and Wind collections for 16/10/2017; that is why plotted counts are truncated for display. Tweet counts for each collection on this date are ~170k ("Ophelia") and ~60k ("wind") respectively. [Bottom] Timeseries of the average UK/Ireland wind speed for the same period. Peaks in wind speed are identified by dashed lines between the two plots to allow visual comparison of wind speed and peaks in wind tweet activity.

Figure 4.5 also shows that there were peaks in tweets relating to wind events which occurred at a time when there was not a named storm event (indicated by 'Unnamed Wind Event(s)' on the figure). Of the 12 peaks in wind speed not attributed to a named storm event, manual inspection of the timeseries identifies that 4 of these peaks correspond to peaks in wind tweet volume while 8 appear not to. This shows that there were wind related events being talked about on Twitter at these times and could suggest that the weather was sufficiently windy to generate discussion on Twitter, however not enough for a named storm event. This shows that social media may have some success in detecting smaller wind events that are not named storms.

There is one peak in wind tweet activity on 17 December 2017 which does not appear to correspond to a peak in wind speed. On inspection of the tweet content in the filtered wind dataset for this date, there were a large number of irrelevant tweets containing the term ‘Invisible Wind Factory’, which is a location in which a music concert took place and prompted discussion on Twitter. This is an example of where relevance filters could be improved. However, overall, there was good correlation between peaks in wind related tweet activity and peaks in wind speed.

The storms which saw the greatest wind speed and impacts (Brian, Caroline, Eleanor, Emma) also appear to have the largest volumes of tweets than the lesser known/less impactful storms (Dylan, Fionn, Georgina).

4.3.2 Sentiment

To understand the emotional response to storm events during the period of the storm, the average sentiment by hour was plotted against the tweet volume over time (Figure 4.6). For Ex-Hurricane Ophelia there is a very clear drop in sentiment (i.e. tweets become less positive and even negative) during and following the peak of tweet activity, before rising again after the storm has passed.

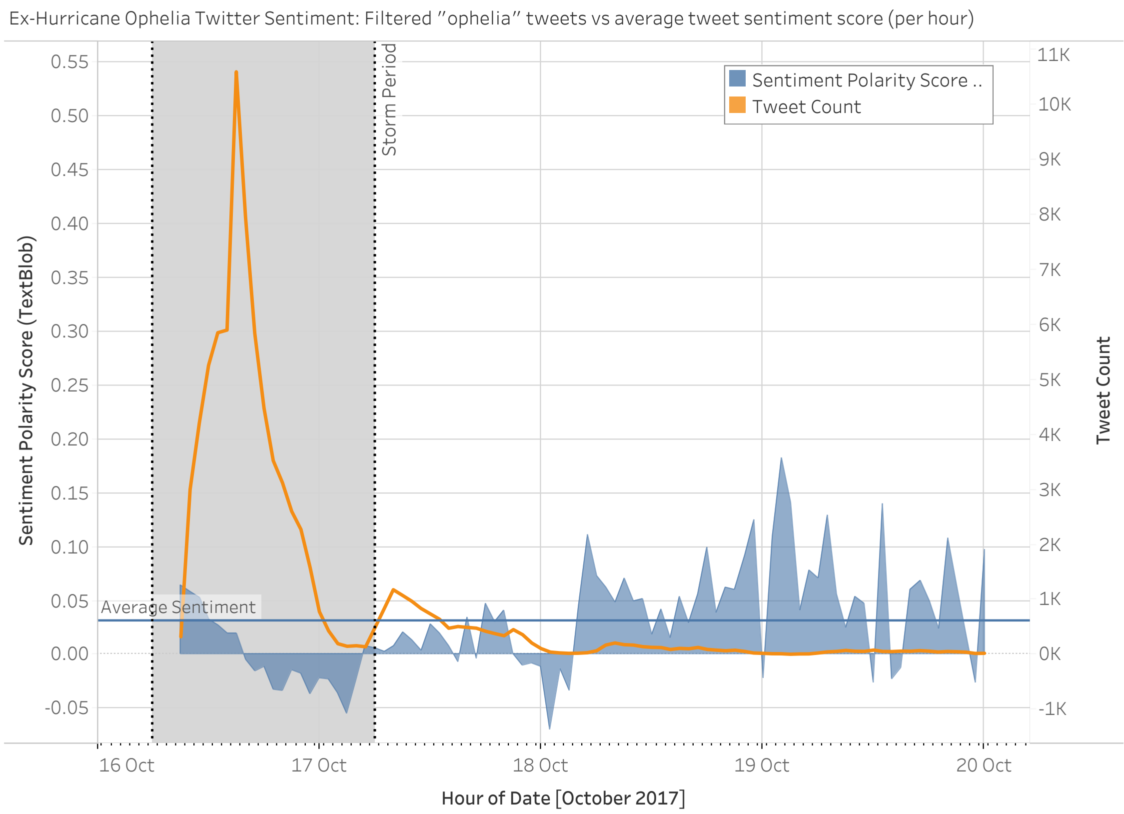

Figure 4.6 - Sentiment polarity score for "Ophelia" tweets vs tweet count – line graph shows tweet count, area graph shows sentiment polarity score, aggregated over 2-hour windows. The period of the storm is shown by the grey shaded bar. There is a clear trend in sentiment, which drops during the storm period and then rises following the storm; see also Figure 4.7. Due to the smaller numbers of tweets, the variance of the sentiment scores increases significantly with time.

The distribution of sentiment in filtered tweets is shown as a histogram of average hourly sentiment in each of the Twitter collections (Figure 4.7). Average sentiment of tweets in the UK during 2017 was shown in another study (using the same sentiment analysis methods) to be 0.13 (Arthur & Williams, 2018); this reference value is shown in Figure 4.7 for comparison. For each tweet collection the distribution of tweet sentiment peaks around an average sentiment score lower than the UK average sentiment. The tweet collection with the lowest average sentiment is the Storm Names collection, with the Wind and Precipitation collections showing relatively higher values, albeit still below the UK baseline. This suggests that wind and rain have an adverse effect on sentiment, with more extreme weather (storms) associated with more extreme low sentiment.

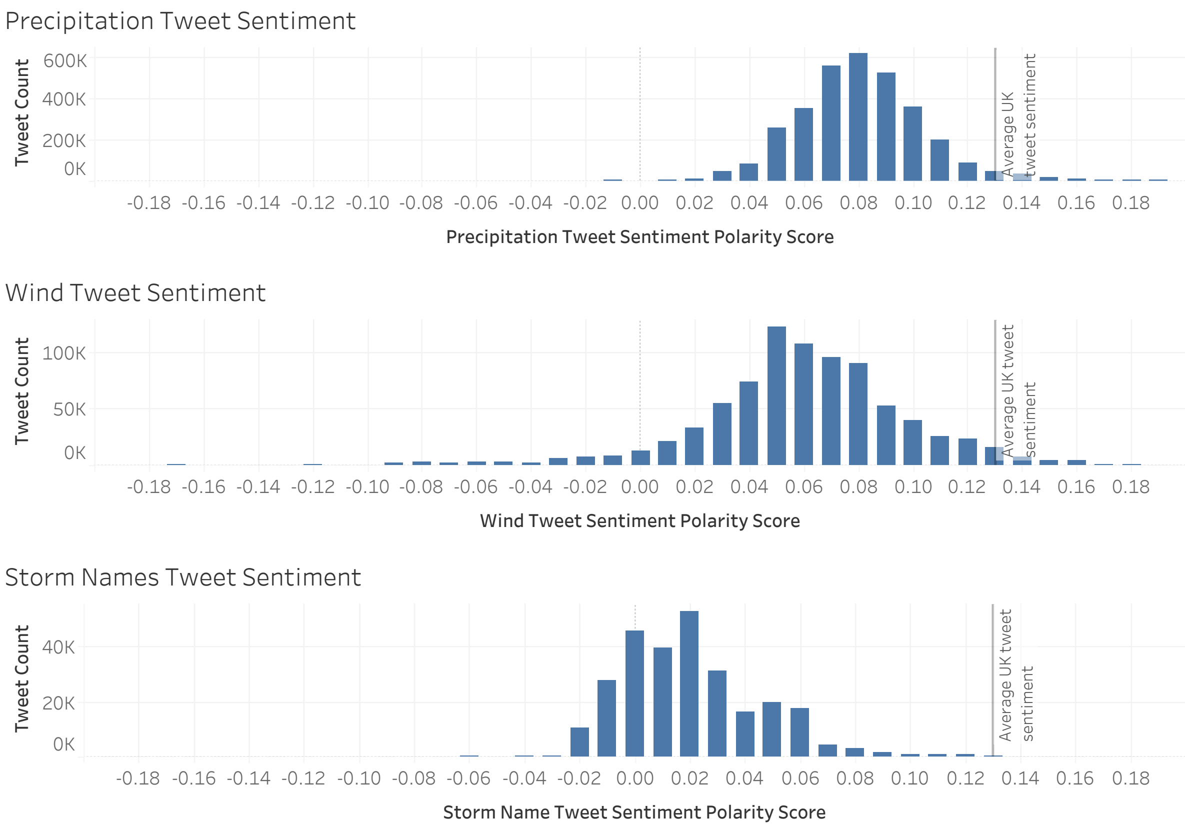

Figure 4.7 - Average tweet sentiment score per hour histograms for precipitation, wind and storm name collections. Average Sentiment score is normally distributed around a mean average sentiment score. The average UK tweet sentiment (Arthur & Williams, 2018) *is shown for comparison.*

4.3.3 Content Analysis

For each storm, filtered named storm tweets in the day before, during and after each named storm event were manually reviewed and categorised. The results for Storm Brian from 20/10/2017 - 22/10/2017 are shown in Figure 4.8. Similar patterns were observed for other named storms (data not shown). There is a clear temporal trend to the types of content posted by Twitter users as the storm passes through. In early stages, warnings are prevalent, but these show a distinct drop in volume as the main effects of the storm begin to be felt (in the early hours of 21/10/2017). In contrast, tweets relating to observations of the weather occurring and reports of damage/disruption begin to increase as the storm passes through. News reports also increase in frequency in the day after the storm. The level of humour expressed throughout the storm period is somewhat more consistent, remaining around 25% of tweets. Tweets categorised as ‘other’ include tweets which cannot be categorised under any of the other headings, e.g. commentary on sports results, business advertising, very short tweets with no information. There appears to be no obvious trend in volumes of these tweets.

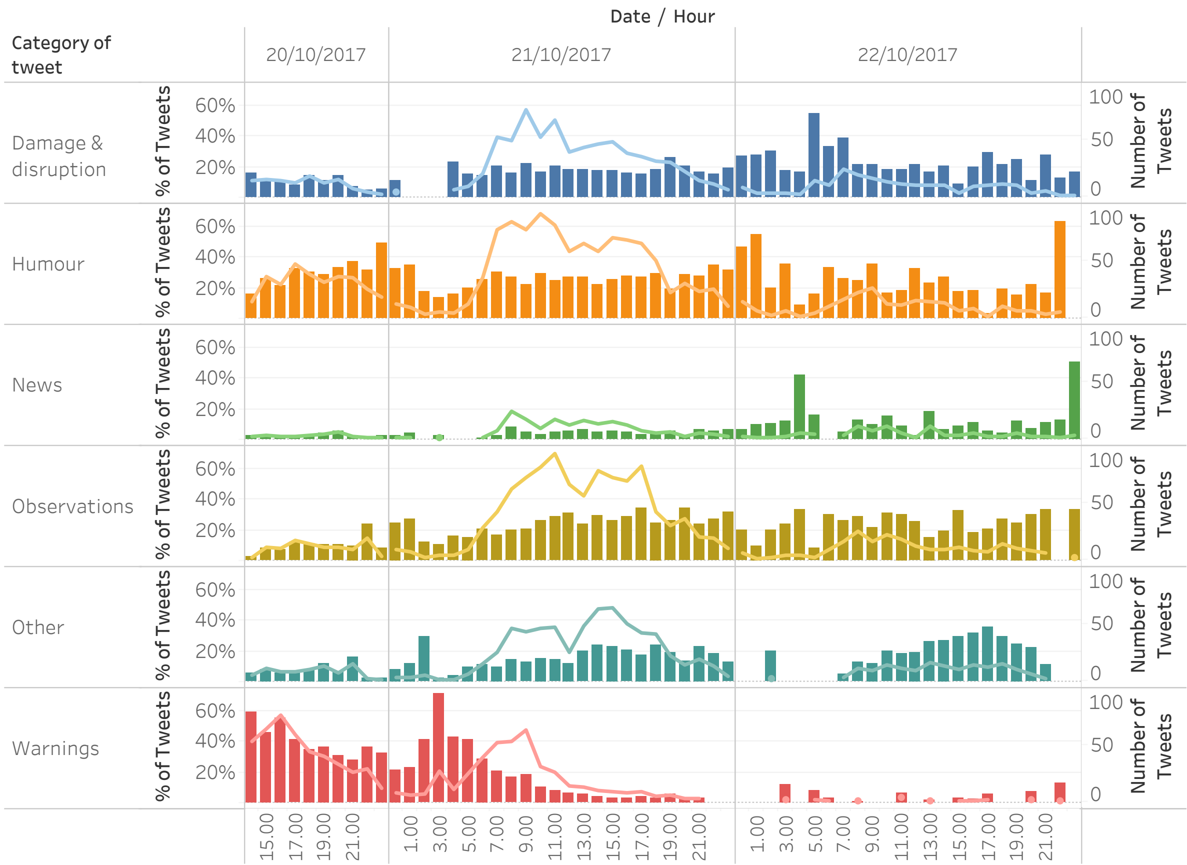

Figure 4.8 - Categorised tweets by date/hour for Storm Brian in the period leading up to, during and the day after the storm passed over the UK and Ireland. Tweets are categorised and plotted as a percentage of all tweets in that hour to account for the expected variation in tweet volumes over each 24-hour period. The number of tweets in that hour is also shown by the line graph.

In terms of tweets providing information on social impacts of the storm, those tweets categorised as damage or disruption are likely to provide us with information on the specific impacts experienced by Twitter users. For the example of Storm Brian in Figure 4.8, 1020 tweets were categorised as damage or disruption. This means that approximately 17% of filtered tweets for Storm Brian provide information on impacts ranging from damage to property, road closures and power outages.

4.4 Discussion

The widespread use of Twitter during extreme weather events, such as named storms in the UK and Ireland, has created an opportunity to use this rich data source to find useful information. In particular, it offers a potential "social sensing" mechanism by which observations of social impacts of extreme weather can be gathered and measurements which are not available from traditional meteorological observations. The demand for such information is evidenced by the recent rise in impact-led forecasting across the meteorological sciences.

This study presents an analysis of data collected from Twitter during the 2017/18 storm season in the UK and Ireland. Various computational techniques were used to filter and extract only those tweets of relevance to wind, precipitation and named storm events. The volume of storm-related weather (wind/rain) tweets increases substantially during storm events. Tweets referring to named storms, after careful filtering to exclude irrelevant content, show clear spikes of activity corresponding to the storm event. Analysis of content shows systematic trends in both sentiment and topics expressed in tweets relating to storms.

Sentiment analysis of tweet content showed clear and consistent emotional impacts of named storms. Average sentiment in weather-related tweets during a named storm event was much less positive than the expected baseline for “normal” Twitter activity. Consistent across multiple storms, collective sentiment was shown to fall significantly as the extreme weather associated with the storm begins to be experienced, before recovering after the storm passes. Furthermore, sentiment is consistently lower in tweets relating to storms than in tweets about wind or rain; however, sentiment for all these weather conditions is lower than the baseline expectation. While sentiment analysis is a crude measure of the psychological aspects of extreme weather, the strength and consistency of the results shown here suggest that these weather events have a substantive adverse impact on social wellbeing.

Categorisation of filtered tweets based on their topic and/or content showed another consistent pattern in the type of information being posted on Twitter during the period of a named storm weather event. In the period leading up to a storm it was found that tweets were mainly giving warnings and information about potential impacts. During the storm, tweets contain information about how people are being affected by the storm, such as tweets on disruption and damage. After the storm, tweets continue to report observations and damage/disruption, but also begin to share links to news reports covering the storm. Surprisingly, the proportion of tweets categorised as ‘humour’ remains quite consistently large throughout the period of a storm, with many tweets making light of the given name of each storm and sharing humorous comments about its impacts, rather than commenting directly on the weather. The patterns shown here suggest that further investigation of content might allow robust measurements of damage and disruption associated with storm events, with some refinements to the method to control for noise and bias. Common sources of noise and bias in social media data include linguistic variation (e.g. regional dialect, slang), tangential content (e.g. tweets related to the storm but not its direct impact, i.e. humour, other) and tweets providing mis-leading or false information. This kind of impact measurement is hard to obtain by other methods and has clear value for validation of weather hazard impact models. Combined with the location inference method this could be developed to provide information of both how and where the biggest impacts as a result of the storm are experienced.

An interesting finding of this study is the existence of peaks of Twitter activity relating to wind and precipitation that are not related to named storm events. With the exception of one peak in wind related tweets, inspection shows that these peaks reflect genuine discussion of weather conditions, showing high levels of public engagement and concern with weather, similar in some cases to those observed for named storms. This finding may have implications for the design of storm-naming systems and wider understanding of when public information should be issued by meteorological agencies.

There are a number of methodological caveats and limitations to this study. After filtering tweets for relevance to storm events, there were relatively small numbers of tweets retained in the data collections for some of the named storms. The relatively small size of the dataset in these cases makes it difficult to confidently identify patterns in tweet discussion.

With regard to sentiment analysis, the tool used in this study (Textblob) has a predefined training corpus based on a dataset of movie reviews. Therefore, it is likely that there may be some uncertainty over the accuracy of some of the sentiment scores assigned to tweets in the storm dataset. To enhance the sentiment analysis of tweets relating to an extreme weather event, it is suggested that a bespoke training corpus based on example tweets from the filtered dataset in this study be created to identify positivity and negativity in tweets relating to the weather. This would provide more confidence in the relevance of the data being used for sentiment scoring.

Aside from improvements to the methods used here, future work might increase understanding of the power and scope of social sensing for weather hazard/impact monitoring by looking at content in different ways. An obvious extension to the work performed in this study is to go into further depth regarding the identification of particular kinds of hazard and/or impact, e.g. by separating travel disruption from damage to property from risks to health. Whether this approach can provide accurate quantification in terms of counting instances of particular impacts is an open research question. The results reported here suggest that clear patterns can be obtained at a reasonable level of granularity. An extension might consider validation of each tweet against the observed weather conditions for that date/time and grid square; this might allow epidemiological study of how different weather conditions (both chronic and episodic) affect behaviour and wellbeing, alongside the more straightforward opportunity to validate the accuracy of individual users as social sensors. Related to impact-based weather forecasting, the volume of activity generated by events categorised as red/amber/yellow might be analysed to study the match between severity judged by meteorological organisations and severity as reported by the general population.

What this study has shown is how social media can be used to provide another layer of information about the social impacts of extreme weather, both emotionally and physically, spatially and temporally, in a way that has not been available before. Being able to determine more specific information about social impacts not available in weather observation data means that impact-based warnings for the public can be tailored towards high impact events. It also provides a method of validation of information provided by meteorological agencies in weather warnings for the public.

Introduction to Chapter 5

In Chapter 4, analysing the content of tweets relevant to named storm events in the UK and Ireland showed that useful impact information can be determined from social media. The methods used to filter and locate the Twitter data also appeared to be successful, as the content of the tweets after the relevance filtering and location inference process were found to be relevant to storms in the UK. However, one issue to consider when developing methods to use social media data as a source of impact information is the need to verify findings from social media with a record of actual impacts. The lack of observations for the type and location of impacts because of weather events is one issue that social sensing methods may be able to address. However, to verify the impacts determined from social media using these methods, it is necessary to compare outputs against other impact data sources, where these are available.

As discussed in Chapter 1, there are a number of databases that have been developed to collate impact information after a significant weather event has occurred (e.g. European Severe Weather Database (ESWD)[27], NatCatSERVICE Database [28] and the EM-DAT International Disaster Database [29]). These databases may be able to provide a source of verification of findings from social media. For example, de Bruijn *et al.* (2019) compare flood events detected using Twitter with flood events in the NatCatSERVICE Database with good accuracy; and Ma & Surakitbanharn (2019) compare hurricane discussion in Twitter with socio-economic and insurance claim data to verify their results. However, while each of these databases has a robust methodology for collating impact information, they also have their limitations in terms of the reliance on third parties to provide information, types of information included, and temporal and spatial coverage. Therefore, there may be gaps in the comparison between events detected on social media and events logged in the databases.

It is also worth noting that the 'threshold of concern' that prompts people to post about the weather on Twitter may vary depending on regional perceptions on what constitutes an impact or a particularly exceptional weather event (e.g. a

[27] https://eswd.eu (Accessed: 17 March 2022)
[28] https://www.munichre.com/en/solutions/for-industry-clients/natcatservice.html (Accessed: 17 March 2022)
[29] https://www.emdat.be (Accessed: 17 March 2022)

person in a location which often experiences high winds may be less likely to post on social media about the wind or its impacts unless it is particularly note-worthy, compared to a person who is in a location less susceptible to high winds). This is important to consider when comparing activity on social media in a particular location with other impact information.

The Met Office Community Impacts database (Robbins & Titley, 2018) is one proposed solution for curating weather impact information, which may provide a good source of data to compare findings relating to weather events from social media against. The database is manually maintained, sources information relating to weather events worldwide using a semi-automated search process of online media, such as news articles and online records of weather events (e.g. FloodList). It also provides a consistent set of information relating to impacts as a result of the event, as well as an impact severity measure based on a set of categorisation criteria. The database is also global in its scope and therefore includes information about impacts as a result of rainfall events worldwide.

Extending the methods in Chapter 4 to detect weather-related social media discussion across the world, Chapter 5 compares weather-related social media activity with the Met Office Community Impacts database. The work was published in a paper in Natural Hazards and Earth System Sciences in 2021 (Spruce *et al.*, 2021). This research focused on using previously explored event detection methods, detailed in Chapter 4, for extracting and locating relevant Twitter data but extended to explore rainfall events worldwide. As well as comparison of tweet activity with the manually curated Community Impacts database provided by the Met Office, the aim was to also explore an appropriate event detection method for rainfall events worldwide using social media, and verify findings using a database of known impactful rainfall events.

Chapter 5 - Social sensing of high-impact rainfall events worldwide: A benchmark comparison against manually curated impact observations

5.1 Abstract

Impact-based weather forecasting and warnings create the need for reliable sources of impact data to generate and evaluate models and forecasts. Here we compare outputs from social sensing -- analysis of unsolicited social media data, in this case from Twitter -- against a manually curated impact database created by the Met Office. The study focuses on high-impact rainfall events across the globe between January-June 2017.

Social sensing successfully identifies most high-impact rainfall events present in the manually curated database, with an overall accuracy of 95%. Performance varies by location, with some areas of the world achieving 100% accuracy. Performance is best for severe events and events in English-speaking countries, but good performance is also seen for less severe events and in countries speaking other languages. Social sensing detects a number of additional high-impact rainfall events that are not recorded in the Met Office database, suggesting that social sensing can usefully extend current impact data collection methods and offer more complete coverage.

This work provides a novel methodology for the curation of impact data that can be used to support the evaluation of impact-based weather forecasts.

5.2 Introduction

Impact-based weather forecasts are increasingly used by National Meteorological and Hydrological Services (NMHS) to provide advice and warnings about both the likelihood and potential impacts of weather events (Campbell *et al.*, 2018). However, methods to evaluate these forecasts are currently limited due to a lack of reliable, quality controlled and sustainable sources of impact data. Meteorological agencies have long-established systems to measure and monitor weather variables, which have allowed weather forecasting to develop to its current high level of performance. But evaluating weather impacts depends on measurements of social activities, health and wellbeing, socioeconomic processes, and other `human factors'; this kind of measurement lies beyond the scope of traditional meteorology. In this paper, we compare two approaches to the evaluation of weather impacts: manual curation of impact databases based on news media and direct reporting, and `social sensing' of impacts based on social media.

Robbins and Titley (2018) made some initial steps to develop an impact-based evaluation methodology by collating information of global socio-economic impacts related to heavy rainfall events. These impacts represent the direct and tangible impacts of high-impact weather (e.g. damage to property, loss of life, evacuation and injury, and restricted or delayed access to essential services). The Community Impacts Database was developed to enable the evaluation of high-impact weather forecasts that are available from the Met Office Global Hazard Map (GHM). The Met Office is the national meteorological service for the UK, providing weather services and contributing to climate science research worldwide (https://www.metoffice.gov.uk/about-us/who). The GHM summarises the risk of high impact weather across the globe for the next 7-days (i.e. weather which can result in significant impacts on safety, property or socio-economic activity). The Community Impacts Database includes information on when and where an impactful rainfall event occurred, as well as a description of the impacts observed, with each event then assigned to an impact severity category. The impact severity category ranges from 1 to 4, where 4 is the most impactful and 1 is the least impactful. There are certain criteria that the impacts of the event must meet for each severity category. Data contained within the database is obtained from a range of online sources across the world, including news, humanitarian and natural hazard websites, in the English language. Collation of the database was labour intensive and required a significant level of manual inspection to extract the relevant temporal, spatial and impact information for each weather event. The data was standardised so that the impact information could be compared with the high-impact weather forecasts provided by the GHM in an automated way. Despite the labour-intensive nature of the process, the authors found the database a good solution to enable impact-based evaluation of high-impact weather forecasts.

There are limited options available for other global databases containing weather impacts with which to compare our methodology against. There are databases such as NatCatSERVICE, produced to record insurance loss as a result of natural catastrophes. However, we would like to consider impacts of extreme weather (i.e. disruption to daily life) which don't necessarily lead to financial loss which could be missing from this kind of record. ReliefWeb, which is a humanitarian information source on global crises and disasters, is another possible database from which to compare our results, however this is filtered for disaster events which are most relevant to global humanitarian workers and decision-makers, rather than all impactful events. Other available databases rely on citizen input (e.g. the European Severe Weather Database (ESWD)), may be limited to certain geographical areas, and are unlikely to contain the same level of rigour as the Community Impacts Database in terms of criteria for inclusion. Considering the options available to us, the Community Impacts Database therefore provides the most comprehensive database for comparing our methodology against.

5.3 Related Work

A number of studies have explored the use of social media as a source of information about the impacts of extreme weather. Social sensing is an approach developed in recent years to analyse unsolicited social media data to detect real-world events of interest.

While social sensing is not specific to natural hazards and can be applied in a variety of contexts (Wang *et al.*, 2012, 2019; Liu *et al.*, 2015), social sensing has demonstrated usefulness for natural hazard events.

Twitter data was used by Sakaki et al. (2010) to detect earthquakes in Japan, with reports arriving in some locations before the shock had been detected by conventional seismography. Many studies have followed, using a number of different approaches to explore the use of social media as an information source during and following natural hazard events. Some studies have focused on the use of social media to better understand risk communication during an extreme natural hazard event. For example, Stewart and Wilson, (2016) explore the use of social media throughout the crisis lifecycle during Hurricane Sandy in the USA, building the STREMII model to better understand crisis communication during an extreme weather event; Rainear et al., (2018) used Twitter data collected during Hurricane Joaquin to explore the types of information communicated by state emergency management accounts to better understand the flow of risk communication during a crisis; Bossu et al., (2020) explored the use of crowdsourced information, along with Twitter data, in a bespoke application during the 2019 earthquake in Albania, finding that engagement of users with the app provided much more information about the damage caused as a result of the earthquake than was available using conventional methods.

Other studies have explored the use of social media to better understand the impacts of extreme weather events. Many studies focus on individual events. For example Fang et al., (2019) use data from the Chinese social media platform, Sina Weibo, during the 2016 Beijing rainstorm, finding a positive correlation between social media activity and precipitation intensity; Sit et al., (2019) examine Twitter data collected during Hurricane Irma, using geo-located tweets to identify locations with a high density of affected individuals and infrastructure damage; and Han and Wang, (2019) use data from Sina Weibo during the 2018 Shouguang flood to analyse the changes in sentiment of social media users during the different development stages of the flood. Further examples of other studies examining the impacts of individual weather events at one particular location include: studies relating to specific hurricanes in the United States (Guan & Chen, 2014; Lachlan *et al.*, 2014; Morss *et al.*, 2017; Wu & Cui, 2018; Kim & Hastak, 2018; Zou *et al.*, 2018; Niles *et al.*, 2019) and specific flooding events (Cervone *et al.*, 2015; Aisha *et al.*, 2015; Brouwer *et al.*, 2017; Rossi *et al.*, 2018; Li *et al.*, 2018; Kankanamge *et al.*, 2020).

Some authors have begun to explore the use of Twitter for more wide-scale specific weather event detection, Arthur et al., (2018) use Twitter data to detect and locate flood events in the UK to produce maps of flood activity. de Bruijn et al., (2019) compare Twitter activity relating to flooding and hydrological information with flood events in the NatCatSERVICE disaster database, finding a good comparison between these data sources. Boulton et al., (2016) use Twitter data collected during several time periods to detect and locate wildfires in the USA. Cowie et al., (2018) find that user reports on Twitter during the year can help to locate peaks in hayfever symptoms as a result of pollen levels in the UK. Furthermore, Spruce et al., (2020) (Chapter 4) examine Twitter data relating to named storms, wind and precipitation in the UK finding that it is possible to identify tweets which can be used to assess the impact of storms both temporally and spatially.

In social sensing, each individual in a social network acts as a sensor and their posts provide pieces of sensor data which can be used to better understand what is happening to or near that individual at a given place and time. Filtering and grouping this information by topic, time or location provides a better understanding of an event through the eyes of a social network. In the context of weather, social sensing can therefore be used to determine where, when and how individuals are being impacted by a specific weather event.

This study seeks to build on and expand the scope of previous work to determine if high impact weather events can be detected without prior knowledge of when or where an event happened. We use the social media platform Twitter to extract tweets from across the world containing key words relating to heavy rainfall and its secondary hazards (flooding/landslides). We then examine peaks in Twitter activity (relative to the normal level of tweet activity for each location) relating to mentions of heavy rain, flooding or landslides. This is then compared with the Met Office Community Impacts Database (Robbins & Titley, 2018) for the same period and hazard focus, to assess the value of socially-sensed tweets for impact database development. Rainfall, and its associated secondary hazards, is a good weather type for this kind of evaluation because it occurs in many places across the globe, with relatively high frequency. In comparison with other hazards, rainfall-related impacts are generally more widely documented (Robbins

& Titley, 2018). This can be attributed to the fact that over the past 50 years rainfall-related impacts from storms and flooding have accounted for the majority of economic losses and deaths worldwide (WMO, 2021b).

The paper is split into several sections. The Methods section gives detail of social sensing methods used, followed by the Results section which compares outputs of social sensing to the manually curated Met Office database. The Discussion section gives some interpretation of the findings and places the work in a broader context.

5.4 Methods

Most social sensing studies have made use of Twitter data and we follow this pattern here. Twitter is an online social networking service that enables users to send short 280-character messages called tweets. It is currently one of the leading social media platforms worldwide based on active users (Clement, 2020a). It provides a platform for users to share and exchange information and news about current events as they unfold in a faster way than traditional media sources (Wu & Cui, 2018). It also encourages the use of text in messages and data is made freely available via the Twitter developer API. There are still some countries where use of the internet is not as widespread or where social media is limited to certain platforms. Despite this limitation, however, Twitter is still one of the most prevalent social media platforms across the world and therefore likely to be a good source of information for understanding where people are being affected by extreme weather, and how they are being impacted by it.

The methods used in this paper to gather, filter and locate the Twitter data follow a similar approach to that used in previous social sensing studies (Arthur *et al.*, 2018; Cowie *et al.*, 2018; Spruce *et al.*, 2020 (Chapter 4)). New methods were developed to compare the results of the social sensing of Twitter data with the Met Office Community Impacts data.

5.4.1 Data Collection

Met Office Community Impacts Database

The extract of the Met Office Community Impacts Database provided for this study included records of high impact rainfall events from 01/01/2017 -

30/06/2017. The database was provided as an Excel spreadsheet which included the following information about each event: impact record date; country in which impact occurred along with nominal location (state/province) provided by latitude/longitude; description of impacts observed; media source of information. Additional information was provided where known: start and end dates for heavy rainfall events; higher resolution location (lower administrative division) provided by latitude/longitude; additional hazard information. Each event was also assigned an impact severity category from 1 to 4 to reflect the severity of impacts experienced during the event. Table 5.1 provides a breakdown of the criteria used for each severity category. As described by Robbins and Titley (2018), the information contained in the database was predominantly obtained from online news and social media, personal correspondence with National Meteorological and Hydrological Services, and existing hazard and impact databases. These included specific known sources (e.g. http://floodlist.com) and news/social media via internet searches including terms such as “heavy rainfall”, “flooding”, “landslide”, etc. The dataset used in this study contained 519 entries (135 unique events) in the period January-June 2017. Unique events refers to the fact that a single rainfall event can lead to impacts in multiple locations.

Severity Category	Description of impacts
1 - Low	Some roads and (< 10) properties inundated over a small area; 1 or 2 localized assets affected/damaged; No fatalities/injuries or hospitalizations; Low-level disruption to daily life (e.g. delays in transport, services shut for short periods).
2 - Moderate	Multiple assets affected (transport, business, residential) over a moderately large area (e.g. multiple districts); > 1,000 homes damaged and/or destroyed; > 1,000 minor injuries and hospitalizations; Wider-scale and prolonged disruption to daily life and services; > 1,000 people displaced/evacuated and/or receiving aid.
3 - High	>= 1 fatalities (but < 50); > 1,000 people displaced/evacuated and/or receiving aid; Multiple assets affected (transport, business, residential) over a large area (e.g. province or state); > 1,000 homes damaged and/or destroyed.
4 - Severe	> 50 fatalities; > 50,000 people displaced/evacuated and/or receiving aid; Extensive damage to multiple assets causing prolonged disruption, inaccessibility and hardship.

Table 5.1: Descriptions of impacts required for each impact severity category related to a heavy-rainfall event (adapted from Robbins and Titley, 2018)

Twitter Data

To gather the tweet data, English-language key words relating to rainfall and impacts of heavy rainfall were used to query the Twitter Streaming API. This API returns all tweets containing the key words from the query, up to a limit of 1% of the total volume of tweets worldwide at any point in time. The key words used to identify and download relevant tweets using the API were: rain, rainfall, raining, rainstorm, flood, flooding, landslide. It is unlikely that tweets using these keywords will have reached the global API limit, since rainfall events tend to be widely dispersed in time and space. Based on these considerations and the absence of any obvious artefacts in our time series we are confident that the API rate limit does not affect our collection (Morstatter *et al.*, 2013).

Tweets were collected during the period 01/01/2017 to 30/06/2017 in line with the time period of the sample of the Met Office Impact Database data used for comparison in this study. Each tweet was saved as a JSON object containing the tweet text as well as a number of meta-data fields relating to each tweet (e.g. timestamp, username, user location, geotag, retweet status, etc). The Twitter

Streaming API searches the whole of the tweet metadata for the search terms requested in the search including tweet text, urls, and usernames. Therefore, collected tweets were filtered to extract only those with one or more of the selected keywords in the tweet text and to remove any duplicate tweet IDs. In total 44.7 million tweets were collected using this method.

5.4.2 Filtering Twitter data

Once all tweet data collected using the API for the study period had been extracted, the raw unfiltered data was then passed through a number of filtering steps to remove irrelevant data. Filters were applied in the following order:

Retweets and quotes

Tweets that were duplicates of an original tweet authored by another user and re-distributed to their own followers (retweets) and tweets which were posted as a quote from another user's tweet (quotes) were removed using tweet metadata relating to 'retweeted status' or 'quoted status'. These tweets do not represent original observations therefore removing them from the dataset prevents any bias in the volume of tweet activity because of secondary public interest in a specific event or location. Though retweets and quotes could provide additional information, their frequency is controlled to a large extent by social network effects, which will be different in different regions depending on local popularity and differences in the use of Twitter. This filter removed 20.7 million tweets (46%) from the raw unfiltered collection leaving 24 million tweets to be passed to the next stage of filtering.

Bot filter

Twitter has many automated user accounts (bots) which are set up to perform a particular function. For example, to collate and post content from a set of sources outside of Twitter, deliver advertising or to promote a particular issue. These types of tweets are unlikely to contain information relating to the impacts that users have experienced from heavy rainfall and may therefore distort the dataset. Therefore, where possible, bot content was removed from the dataset. As bot accounts tend to create many more tweets than human users, simple bot filtering was achieved by identifying user accounts which had a disproportionately high number of tweets (using a threshold of >1% of the total number of tweets in the dataset). Any tweet in the dataset which was posted by an identified bot account was removed. Manual inspection of tweets during the development of the filtering process identified a number of other bot accounts which were also removed. The bot filter removed 2.7 million tweets (6% of the total unfiltered dataset), leaving 21.3 million tweets to be passed to the next stage of filtering.

Weather Station Filter

As the tweet collection in this study is focused on weather-related terms, a high number of weather station tweets were also present in the dataset. Some amateur weather stations are set up to automatically post observations to Twitter. As for Twitter bots, weather station tweets, while containing information on the weather conditions at a particular location and time (such as the amount of rainfall), are unlikely to provide any relevant information on the impacts from heavy rainfall (e.g. damage, disruption). Therefore, any weather station tweets not picked up by the bot filter described above required an additional weather station filter to remove them from the dataset. Many of these tweets follow a fixed structure (for example: *'06:30 AM Temp: 53.0oF Hum: 91% Wind: 7.0 mph N Bar: 29.530 in. Rain: 0.09 in'*) and therefore the majority can be identified by searching for multiple occurrences of meteorological terms and units. Any tweet with 3 or more of any combination of weather terms and/or units was therefore removed from the dataset. A randomised sample of tweets removed using this filter was checked to ensure no tweets that were not weather stations were removed using this filter. The weather station filter removed 4.7 million tweets (11% of the total unfiltered dataset), leaving 16.6 million tweets to be passed to the next stage of filtering.

Phrase Filter

Another issue with the collection of tweets containing weather related keywords is the use of weather terms in phrases and figures of speech which are not related to the weather. For example: '*floods of tears*', '*rain check', 'raining offers*', '*winning by a landslide'*, etc. Other terms found to be present in irrelevant tweets are also removed. These are generally political in nature and include terms such as *election, vote, trump, labour, migration*, etc. Song titles containing the key words were also removed, for example *'Purple Rain', 'Singing in the Rain'*, etc. Applying the phrase filter removed 1.3 million tweets (3% of the total unfiltered dataset), leaving 15.3 million tweets to be passed to the final stage of filtering.

Machine learning filter

Although the previous stages of filtering removed many irrelevant tweets, manual inspection of remaining tweets found that there were still a large number that contained the keywords but that were not relevant to rainfall or the impacts of heavy rainfall. These included warnings about forecasts of rainfall, business advertising, links to articles on other topics, and various other irrelevant content. Therefore a Naïve Bayes classifier, found to be successful in other studies (Arthur *et al.*, 2018; Cowie *et al.*, 2018; Spruce *et al.*, 2020 (Chapter 4)) for the filtering of tweet content, was employed.

A set of 5434 tweets were randomly selected from the filtered dataset of tweets remaining after the phrase filter described above. Each tweet in this random set of tweets was manually inspected and labelled as relevant or irrelevant. A tweet was marked as relevant based on the criteria that the tweet had to be relating to rainfall that was currently happening, had happened recently or was about the impacts of rainfall experienced recently. Everything else was marked as irrelevant. For example, *‘Rain destroys 60 buildings in Ondo’* would be marked as relevant whereas *‘Rain expected in Ondo tomorrow’* would be marked as irrelevant. In total there were 1316 tweets marked as relevant and 4118 tweets marked as irrelevant.

The labelled dataset was then used as training data for a Multinomial Naïve Bayes classifier. As a first validation test for this approach, 25% of the data was held back as a validation set and a classifier was trained on the remaining 75% of cases; this classifier had accuracy (i.e. correctly identified the relevance/irrelevance) of 90% on the held-back validation tweets, with an F1 score of 0.88 As a second test, to confirm the robustness of the approach, the same training/validation test was repeated with 6-fold cross-validation. The results of each test were combined to give an overall mean F1 score of 0.89 and the summed confusion matrix (also known as ‘contingency table’) shown below (where True is relevant and False is irrelevant):

$$\begin{pmatrix} & & Predicted & \\ & & False & True \\ Actual & False & 3966 & 152 \\ & True & 140 & 1176 \end{pmatrix} \tag{1}$$

This confusion matrix shows overall accuracy of 95%, with most tweets in the filtered dataset classified as not relevant. Accuracy was higher for the False class (3966/4118 = 96%) than the True class (1176/1316=89%). This could be attributed to the training dataset being unbalanced and biased towards irrelevant tweets. Overall the results of the machine learning filter testing indicate good performance.

The machine learning filter removed 10.4 million tweets (23% of the total unfiltered dataset), leaving 4.9 million tweets (11% of the total unfiltered dataset) for further analysis.

5.4.3 Location inference

Typically, only ~1% of tweets collected using the Twitter developer API using keywords contain the geo-coordinates needed to determine the specific location of a tweet, while a further 2-3% contain specific place coordinates (Dredze *et al.*, 2013). Therefore, even after filtering for relevance, determining the location of a tweet collected in this way requires further processing to determine where in the world it originated from or relates to, in a process of location inference.

The 4.9 million tweets remaining after the relevance filtering stages were further processed to see if location could be identified using information contained within the tweet. The location of the tweet is important in understanding where in the world the rainfall event had/was taking place. We chose to work at a geographic resolution of GADM Level 1 units, which are sub-national administrative regions (e.g. US states, UK countries, Australian states). This choice is a balance between fine-scale resolution and having enough tweet data in each unit to give meaningful outputs; it is also the resolution at which the Met Office impact database was aggregated for evaluation against weather forecasts.

We found that 2% of tweets contained specific geo coordinates of the tweet origination (geotag) and a further 5% contained the coordinates for the place a user designated in the Twitter application when posting the tweet (place).

However, this left 3.7 million tweets without specific location coordinates. As these tweets would very likely contain relevant information relating to the impacts of a rainfall event, it was important to try to determine the location of the tweet so that the information contained within the tweet could be used. Therefore, a location inference process was used for each remaining tweet to see if location could be determined either from the location given in the user profile (user location) or place name detected in the tweet text. The steps taken in the location inference process are as follows:

Country filter

Place names alone without any other information, such as country or state name can often apply to more than one country. For example, York (UK and Canada), London (UK and Canada), Pasco (USA and Peru), etc. Therefore, an initial filter was created to identify the country associated with a place name. For some countries, place names in text commonly follow a specific pattern or use certain abbreviations. For example, in the USA, Canada and Australia, users often put a place name followed by a 2-character or 3-character abbreviation for the state (e.g. Los Angeles, CA; Vancouver, BC; Sydney, NSW). Text scanning for place names was extended to look for the 'place name, state abbreviation' template, as well as the names/abbreviations of states and/or country name for USA, Canada or Australia. Where a country or state could be identified in this way, any further location inference steps only checked for place names in that particular country. This disambiguation step gave much better location performance overall, as well as computational efficiency benefits.

Gazetteer look-up

This filter checked the tweet to determine if a discernible place name could be detected from the user location and/or the tweet text using gazetteers including Geonames (Geonames, 2020) and DBPedia (DBpedia, 2020). The following methodology was applied to each tweet which did not contain geo or place coordinates as described above:

- Geonames was used as our primary source of gazetted features as it is a geographical database with information about all countries with over eight million places, such as cities and points of interest. Where there was no match found in the Geonames database, the DBpedia database was used.

- Where a match to a place name is found, a set of co-ordinates or bounding boxes from the gazetteer database is returned.
- Where locations were found in both the user profile and tweet text, place names in the tweet text are preferred as they are more likely to relate to the subject of the tweet.
- In a small number of cases, the user profile location and tweet text locations may differ; in that case, the place determined from the tweet text is given more weight during the location inference process.
- Where multiple matches to a place name were found in Geonames (i.e. where a place name exists in more than one country), then if there was no reference to the country elsewhere in the tweet or the country had not already been determined by the country filter described above, the place with the largest population (which has been found in previous studies to be the most likely location for the tweet (Schulz *et al.*, 2013; Arthur *et al.*, 2018) was logged and coordinates returned.
- In addition, where multiple place names are determined from a tweet, to infer the most probable location, areas of overlap between the matching location polygons are detected before a final coordinate or bounding box is returned. This assumes that polygon overlaps are the highest likelihood locations.

Since some place names are also commonly used to denote something other than a location (Liu *et al.*, 2011), a database of words which are also places was used to remove apparent locations which were more likely to be a word than a place (e.g. dew, aka, var, etc).

Validation

The method described above is based on the location inference method validated by Schulz et al. (2013) who found 92% accuracy when inferred location from user location/place name mentioned in tweet was compared against tweets for which a geotag was known. The method was also used successfully by Arthur et al. (2018) and Spruce et al. (2020) (Chapter 4).

To validate the location inference approach for this study, a random sample of 100 tweets, including the tweet metadata, was taken after the filtering and location inference stage had taken place from the whole dataset for all dates. Each tweet's metadata was examined for location references and this was cross-referenced with the GADM Level 1 location(s) that the tweet was assigned to using the social sensing location inference method. We found that 93 out of 100 tweets in this sample were assigned to the correct location(s) which shows that the location inference method was working well. This is also in line with previous studies' validation of this location inference approach. Applying this location inference approach on a global scale carries more potential for place names used in multiple countries being mis-assigned their geographical coordinates than if working with tweets for a single country. Therefore, locating tweets with a 93% accuracy in this study is considered a good success rate given the potential ambiguities.

Matching to GADM Level 1

Once a place is identified it is matched to the GADM Level 1 Administrative area polygon that contains it. If a tweet's location spans multiple GADM Level 1 areas then the contribution of that tweet to the total count is split proportionally between each area. After processing the location for all tweets, the overall counts of tweets within each GADM level 1 are then collated for each day within the period of study (1/1/2017 – 30/6/2017).

5.4.4 Metrics for comparison of social sensing and Met Office Community Impact Database

The number of relevant tweets in each GADM level 1 area for each day was used to calculate a ranking for all days in the study period for each location, given as a tweet count percentile e.g. day X is in the Yth percentile of tweet counts at location Z. This metric tells us how the number of tweets on a specific day in that location compares with 'normal' tweet activity in that place. We use percentiles in preference to absolute counts of tweets to account for varying prevalence of tweets in different locations due to either the size of population or propensity of the local population for using Twitter. If the number of tweets in a particular location on a particular day is low for that location, the percentile will be low, if the number of tweets is high for that location, the percentile will be high. We are interested in locations and days where the percentile of tweets is particularly high as this indicates that there is unusually high Twitter discussion about rainfall that particular day, which in turn suggests that there is more likely to be a rainfall event taking place. We might also infer that the higher the percentile (i.e. the more extreme the number of tweets for that place), the more impactful the event.

To test our theory that a higher percentile of rainfall-related tweets in a location implies that a rainfall event, or the impacts of a rainfall event, are being experienced, we compare our percentile calculations with the events logged in the Met Office Community Impact Database. For each day in the study period and location included in the Met Office database, we compare the percentile of tweets with whether or not an event is logged in the database on that day, in that place. As we do not currently know the percentile threshold that implies an impactful rainfall event is taking place, we repeat this comparison for different tweet percentile thresholds between the 65th and 99th percentiles. Where a rainfall event spans multiple days in the database we compare the percentile of tweets for each day of the event. The results of these comparisons are discussed below.

It is also worth noting the limitations of the Met Office impact database as a validation source for our Twitter data. As noted by Robbins and Titley (2018), the methods used to create the records in the Met Office database use manual searches of news and social media sources written in English, which does not necessarily lead to an exhaustive list of all high impact rainfall events that have occurred across the world. This means that this study is not necessarily a validation of `ground truth' event detection using Twitter but instead is a triangulation between identified impact events using Twitter and the Met Office impact database. In the results that follow, we present outcomes as if the Met Office data were ground truth, i.e. where we find a false negative it indicates a case where social sensing does not find an event that is found in the Met Office data. The true number of false negatives (events that occurred in reality but are not detected by social sensing OR by Met Office data) is unknown.

5.5 Results

In this Results section, we first analyse the coverage of the two datasets (social sensing and manually curated Met Office database). Then we present some illustrative examples to show the properties of the two data sources, before a sensitivity analysis on factors affecting the performance of social sensing, assuming that the Met Office data represents “ground truth” (note that this is not necessarily the case - we return to this assumption in the Discussion). The final set of results shown is an assessment of local/global performance of the social sensing method.

5.5.1 Data coverage

Figure 5.1 shows a timeseries of the number of tweets collected per day and the number of tweets retained after filtering the raw dataset for relevance. There was unfortunately some server downtime between 16/03/17 and 18/03/17 resulting in missing tweets for this time period (grey bar in Figure 5.1). These dates are therefore excluded from all further analysis and comparisons between the Twitter data and the Met Office database.

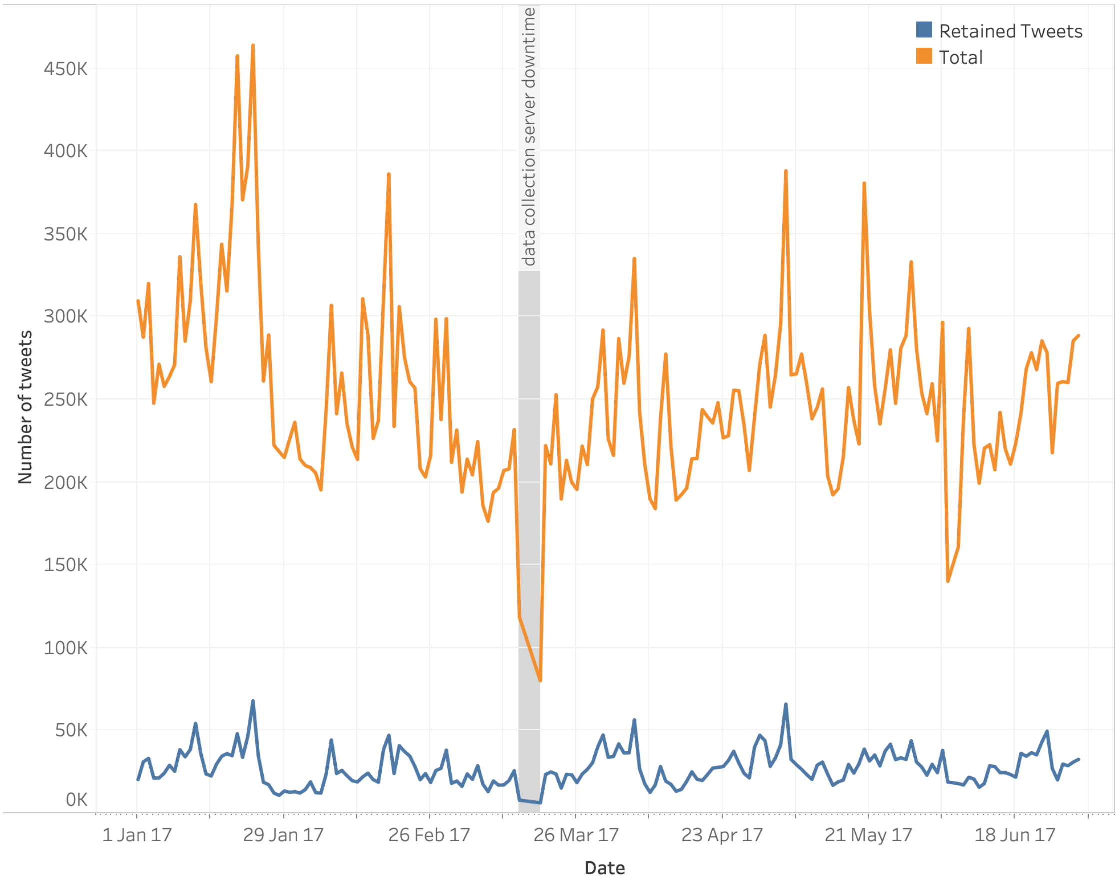

Figure 5.1: Number of tweets collected per day between 01/01/2017 and 30/06/2017. Data shown for both the total number of tweets collected (top line) and the number of tweets retained after filtering for relevance (bottom line). The period where the tweet collection failed (16/03/2017–18/03/2017) is shown by a grey bar.

Figure 5.2 shows the number of tweets in each GADM Level 1 area across the world for the whole study period. The majority of tweets are located within the USA, UK and Australia. This is not surprising given that we have collected tweets containing English language terms and these are English-speaking countries with a very large number of Twitter users. Any areas without any tweets during the study period are shaded white on the map. The figure shows that we have good global coverage of discussion about rainfall on Twitter, with at least some tweets in most areas.

Figure 5.2 also shows the locations of high impact rainfall events recorded in the Met Office database. Again, there is a good global spread of events both in English-speaking and other language speaking countries. The relevance filters are likely to remove other language tweets.

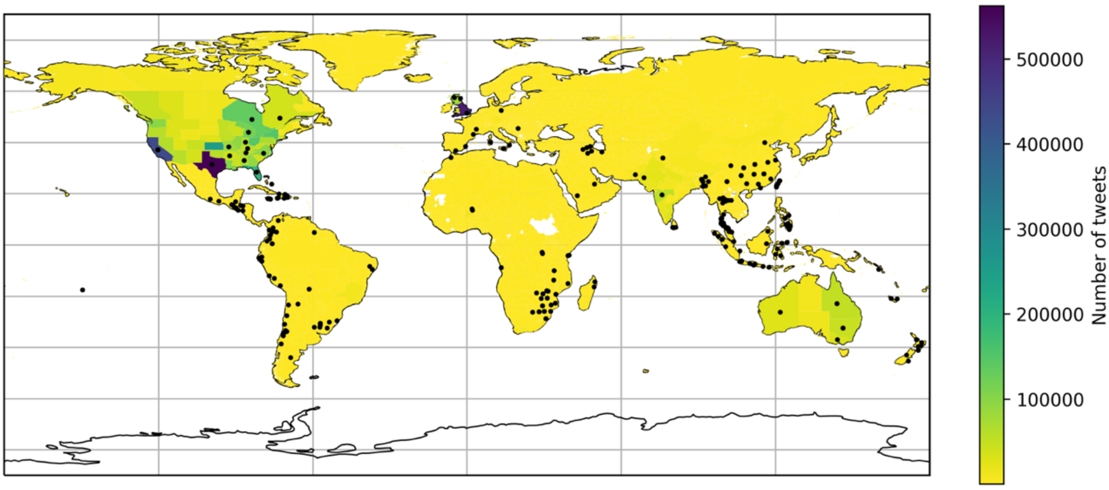

Figure 5.2: Global map showing the number of filtered heavy rainfall tweets located in each GADM level 1 administrative area during the period of study (01/01/2017–30/06/2017). Areas with white shading had no located tweets during the period of study; shaded areas had at least 1 tweet. Locations of impact events recorded in the Met Office database are shown by black points.

Figure 5.3 shows the number of GADM level 1 areas which had at least 1 tweet recorded in the filtered dataset (3379/3491 areas) and the number without tweets (112/3491 areas). GADM areas without tweets were found to be predominantly areas within countries with a low population density (e.g. Angola, Laos, Svalbard) or island nations (e.g. the Bahamas, Nauru, Seychelles, Vanuatu). The areas with and without tweets are also compared with the number of GADM level 1 areas with an event in the Met Office database (224/3491 areas). All GADM level 1 areas with an event in the Met Office database had tweets recorded. None of the areas with zero tweets recorded had an event in the Met Office database. It is striking how many GADM Level 1 regions have some tweets recorded that talk about extreme rainfall or flooding, compared to the number that have verified high-impact rainfall events (floods and landslides) recorded in the Met Office database. We will return to the reasons for this disparity in the discussion.

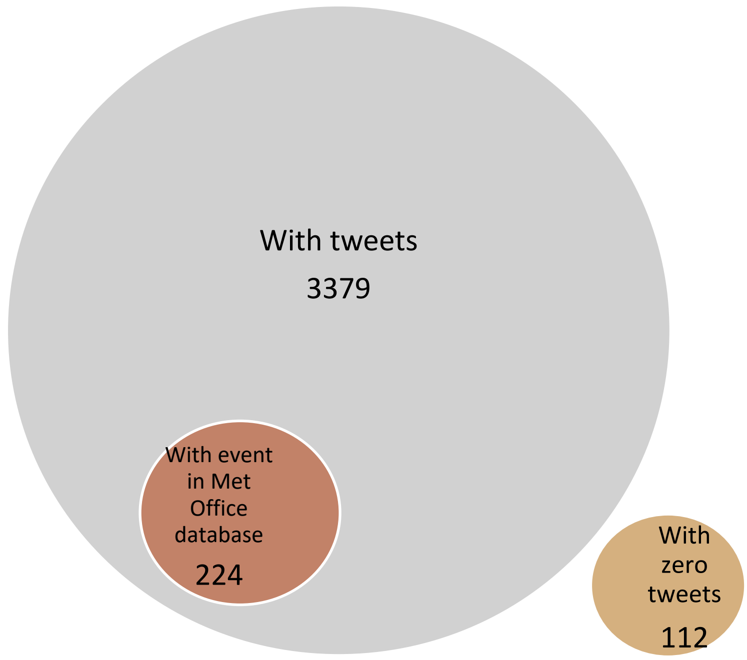

Figure 5.3: Venn diagram showing the number of GADM Level 1 areas (from a total of 3491 areas) with tweets and without tweets compared with the number of areas with at least one event in the Met Office database.

5.5.2 Comparison between social sensing and the Met Office database

The following are illustrative examples that demonstrate the properties of the two data sources.

Spatial correspondence between social sensing outputs and precipitation observations

For each day in the study period, the percentile of tweets for each GADM Level 1 area was mapped. A visual inspection of each map identified a number of examples of peaks in Twitter activity that correlate with observed rainfall. Figure 5.4 shows an example of a particularly impactful rainfall event in the USA on 30th April 2017. The areas with the highest percentile of tweets appear to correlate well with areas of significant rainfall. This provides some confidence that the spatial distribution of peaks in Twitter data correspond to areas of observed rainfall.

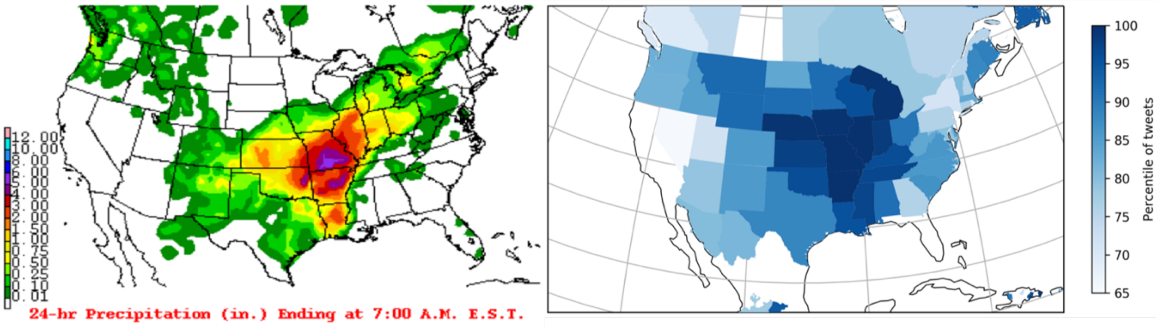

Figure 5.4: (LEFT) 24-hour precipitation (inches) for USA on 30th April 2017 (http://www.wpc.ncep.noaa.gov). (RIGHT) Map of North America showing the percentile of tweet activity for each GADM level 1 administrative area on 30th April 2017.

Temporal correspondence between social sensing and event database outputs

Time series of the volume of tweets for each GADM Level 1 area which had an event recorded in the Met Office database were examined to determine whether spikes of Twitter activity correspond to event dates in the Met Office database. Figure 5.5 shows an example of this for GADM Level 1 areas in Australia. Events in the Met Office database (shown by the vertical dashed lines in Figure 5.5) largely correspond with peaks in tweet activity for these regions. It also appears that there may be at least one high impact rainfall event detected by social sensing that is not included in the Met Office database. Looking at 9th April 2017 there is a significantly high number of tweets in Victoria which do not correspond to an event in the Met Office database. Investigation of news articles and weather reports for this date identified that there was a significant rainfall event on this date that would have met the criteria for inclusion in the Met Office database. Therefore, this provides an example where the use of social sensing could aid with impact event detection and provide an additional source of impact information. Other peaks in tweet activity where the volume of tweets is above the 95th percentile for the region are also labelled as possible high-impact events which might have met the criteria for inclusion in the Met Office impact database, but were missed in the original creation.

It can also be seen from Figure 5.5 that for some events there is a time lag between the rainfall event occurring (dashed vertical lines) and the peak in the tweet activity (line graph). For example, in New South Wales (top plot in Figure 5.5) a rainfall event occurred on 19th March 2017, however a peak in Twitter discussion was not observed until 2 days later, on 21st March 2017. Additionally, for Western Australia (bottom plot in Figure 5.5) a peak in tweet activity was observed and correlated with the rainfall event on 30th March 2017, however there was an additional peak in tweet activity in the following days on 3rd April 2017 when extensive news coverage about the rainfall and its impacts occurred. Therefore, potential time lags between a phenomenon occurring (i.e. heavy rain), the hazard produced (e.g. a riverine flood) and the impact (e.g. bridge destroyed) and then people posting on Twitter about it need to be taken into consideration as a factor which may affect the performance of social sensing.

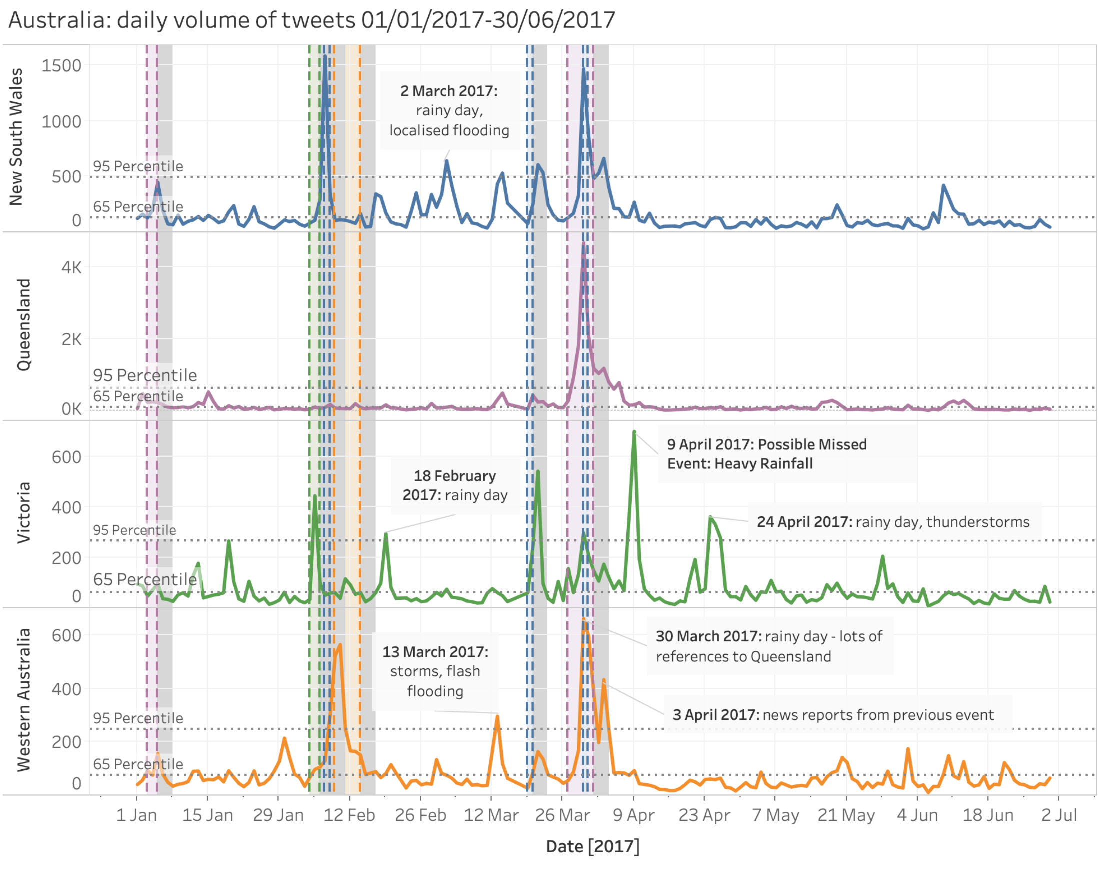

Figure 5.5: Timeseries of filtered tweet counts per day for each of the Australian administrative areas with events in the Met Office database. The period of each heavy rainfall event in the Met Office database is shown by the vertical dashed lines containing a shaded bar colour coded to the administrative area. The 3 days after each event is shown by a grey shaded bar. Social sensing "events" that are not present in the Met Office database are labelled.

Figure 5.6 shows a similar plot to Figure 5.5, but for the United Kingdom (UK). In this example, there are greater disparities between events identified in the Met Office database and those identified using the social sensing method.

There are a number of rainfall events identifiable from the tweet time series in Figure 5.6 which are absent from the Met Office database: 12/13th January; 23rd February; 17th May; 27th June 2017. A significant peak in tweet activity (above the 95th percentile) is noted for each of these dates and further investigation of news media and weather reports shows that there were rainfall impacts in the UK on or around these dates. However, not all of the peaks in tweet activity can be attributed to genuine high impact rainfall events. For example, the peak in tweet activity seen around the 27th-29th May 2017 coincided with a Bank Holiday weekend in the UK with a weather forecast for bad weather. This generated a large amount of news and social media discussion on cancelled events and holiday plans, as well as some travel disruption, not all of which was related to the weather. This provides an example where social sensing can provide a false positive result. False positives could occur for a number of reasons: For example, do smaller, less impactful rainfall events in the UK generate more discussion than in other countries given that rainfall is quite common here? Or being a relatively small country, impacts due to the weather have potential to be more localised, affect less people and therefore not as high a severity on the global impact scale used for the curation of the Met Office database. In this particular example there is also a question regarding the relevance of a bank holiday in affecting people's perception of risk and impact. Additionally, wet weather during an otherwise quiet dry summer could generate more discussion on Twitter as it is more noticeably out of the ordinary and therefore more attention grabbing that if the same rainfall fell during a wet and stormy winter.

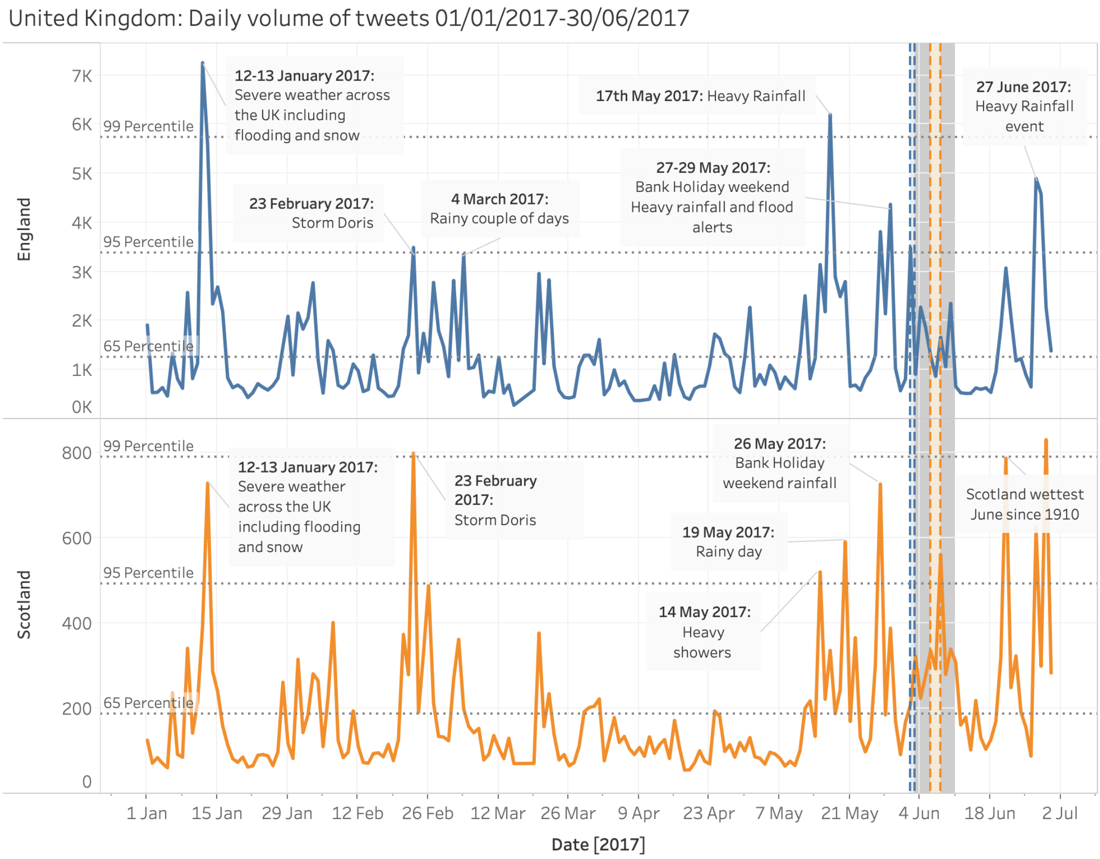

Figure 5.6: Timeseries of filtered tweets per day for each of the UK administrative areas with events in the Met Office database. The period of each heavy rainfall event in the Met Office database is shown by a shaded bar colour coded to the administrative area. The 3 days after each event is shown by a grey shaded bar. Potential missed events in the Met Office database, which are identified in the Twitter data are labelled.

Examining the illustrative examples above as well as time series for other areas (not shown) we found there was a good match between areas with recorded heavy rainfall events and a high percentile of tweet activity relating to rain and the impacts of rain. We also found a good match between peaks in tweet activity and events in the Met Office database for some areas (e.g. Australia, some parts of the USA, Malaysia, Saudi Arabia, Angola) and a poorer match for others (e.g. UK, India, Haiti). Investigating peaks in tweet activity which do not correspond to a recorded event in the Met Office database, we found that most of these peaks refer to genuine high-impact rainfall events. These findings suggest that social sensing of rainfall events can be a useful addition to current manual methods of impact data collection, helping to identify a wider variety and greater number of high-impact events.

5.5.3 Factors affecting social sensing performance

Performance metrics

To understand how the social sensing method is working in terms of links between peaks in Twitter activity (i.e. percentile of tweets for a particular area) and events logged in the Met Office database, we tested the social sensing method as an event detector, assuming that the Met Office events database represents ground truth. To quantify performance and account for the various methodological factors (for example, the tweet activity percentile threshold used to decide when an event had occurred), we plotted precision/recall curves.

Recall is used to show the ability of a model to find all of the relevant cases in a dataset (Koehrsen, 2018). In this study, calculating recall indicates how well the social sensing method finds events in the Met Office database. Recall is calculated by taking the number of true positives divided by the number of true positives + the number of false negatives (Eq. (2)). For each day in the study period, a true positive would be counted if there is an event in the Met Office database AND the percentile of tweets is greater than or equal to the chosen percentile threshold (meaning the social sensing method correctly detects the event). A false negative would be counted if there is an event in the Met Office database but the percentile of tweets is less than the chosen percentile threshold (i.e. the event was not detected using tweets).

$$recall = \frac{[true\ positives]}{[true\ positives] + [false\ negatives]} = \frac{[events\ correctly\ detected\ using\ tweets]}{[events\ correctly\ detected] + [events\ not\ detected]} \quad (2)$$

Precision is used to show the proportion of data points a model says are relevant compared to those which are actually relevant (Koehrsen, 2018). In this study, precision shows how accurately the social sensing method finds events in the Met Office database – i.e. if there is a peak in Twitter activity in a particular place on a particular day, does this correspond to an event in the Met Office database? Precision is calculated by taking the number of true positives divided by the number of true positives + the number of false positives (Eq. (3)). For each day in the study period, a true positive would be counted as described for recall above, whereas a false positive would be counted where the percentile of tweets is greater than or equal to a given percentile threshold but there is NOT an event in the Met Office database (event detected but not actually an event).

$$precision = \frac{[true\ positives]}{[true\ positives]+[false\ positives]} = \frac{[events\ correctly\ detected\ using\ tweets]}{[events\ correctly\ detected]+[events\ incorrectly\ detected]} \quad (3)$$

Plotting precision and recall against each other shows how well (or not) the social sensing method is replicating the Met Office database of recorded events. Recall and precision were therefore calculated for each GADM level 1 administrative areas with an event in the Met Office database. As we do not know the optimum percentile threshold that would achieve the best social sensing performance, recall and precision were calculated using tweet percentile thresholds between the 65th and 99th percentiles. This will help to determine which percentile threshold is optimal for signalling that an impactful rainfall event is occurring.

Further to precision and recall, we also calculated the *f-score* - a metric which takes both precision and recall into account. This is a single score that indicates how well the social sensing method is working and can be used to find the optimal percentile threshold to signal a rainfall event is occurring. The F1 score is defined as the harmonic mean of precision and recall and aids in tuning a model to be optimised for both of these metrics (Koehrsen, 2018). In this study, we calculate a variation of the F1 score, the F2 score, which gives a higher weight to recall in its calculation (Eq. (4)).

$$F2\ Score = 5 * \frac{Precision * Recall}{(4 * Precision) + Recall} \quad (4)$$

For reference, F2 scores fall in the range [0,1], with a score of 1 being perfect recall and perfect precision. As used here, we are interested mainly in the change in F2 as different parameters are varied, rather than its absolute value.

We choose to favour recall here as we are most interested in how well the social sensing method detects events in the Met Office database; furthermore, calculations of precision are somewhat less reliable due to the lack of genuine ground truth data. While the accuracy of the event detection is important, we prefer to detect as many events as possible and tolerate occasional peaks in Twitter activity that do not match an event in the Met Office database. As previously noted, the Met Office database does not provide a definitive list of all high impact rainfall (and secondary hazard) events that have occurred and there may well be events missing from this database that Twitter can help us detect. In other words, neither dataset is perfect but utilising the positive attributes of both methods could lead to an enhanced approach for sustainable and robust impact data collection.

Sensitivity of social sensing performance to event detection window

Figure 5.7 shows precision and recall calculated for all GADM Level 1 areas where an event was recorded in the Met Office database. Each plotted point shows precision and recall for a given tweet percentile threshold for event detection. Initially, precision and recall were calculated requiring that a peak in tweet activity must exactly match the day of the heavy rainfall event (Day 0). However, as identified by Robbins and Titley (2018), there can sometimes be a time lag between a rainfall event and impacts of the event being experienced or reported. Time lags were also observed between a rainfall event occurring and peaks in tweet activity in Figure 5.5 above. Therefore, precision and recall calculations were repeated for event detection windows of varying duration: Day 0 only; Day 0 + Day 1 (Day +1); Day 0 + Day 1 + Day 2 (Day +2); Day 0 + Day 1 + Day 2 + Day 3 (Day +3). Longer time windows were trialled in preliminary work, but showed no additional benefit; also, longer time windows reduce the ability to locate events in time. Figure 5.7 shows precision/recall curves for each of these scenarios, showing that the 3-day window (Day +3) yields the best results.

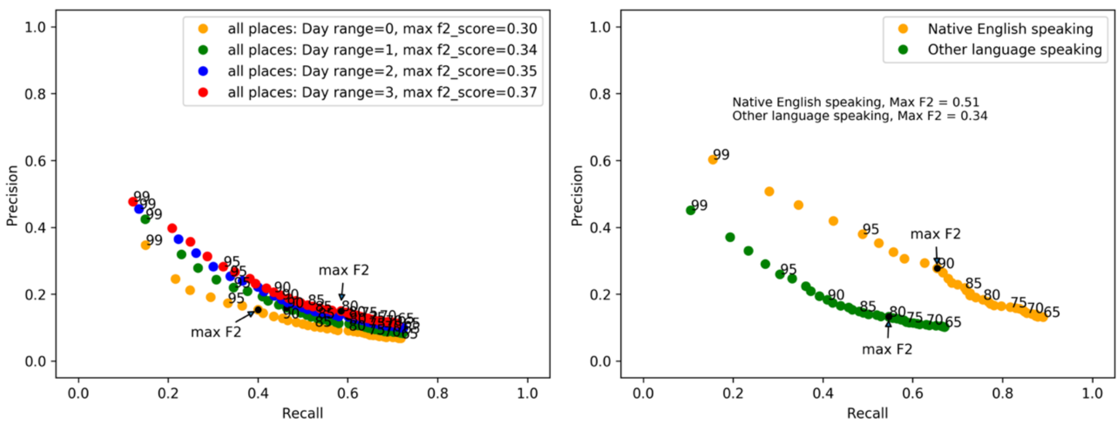

Figure 5.7: [LEFT] Precision and recall values when comparing tweet data with the Met Office impact Database for Day 0 only, Day +1, Day +2 and Day +3 from the impact event date. Each point represents the tweet percentile threshold used to signal true and false positive values for an event taking place in the Twitter data. Tweet percentile thresholds tested range from the 65th percentile to the 99th percentile (step size 1). [RIGHT] Precision vs Recall plot for matches (within 3 days of event) to Met Office impact event database vs tweet percentile thresholds 65–99 (step size 1) for native English-speaking countries vs other language speaking countries

Social sensing performance in English-speaking and other language speaking countries

As the tweets collected were in the English language only, we are also interested in whether the social sensing method works better for native English-speaking countries. Using the precision/recall calculations described above and for day range +3, a precision/recall curve was plotted for tweets from native English-speaking countries versus other language speaking countries. Figure 5.7 shows the results of this comparison and that the social sensing method yields much better results for native English-speaking countries with a maximum F2 score of 0.51 compared with 0.34 for other language speaking countries. The difference in performance is perhaps not surprising given that tweets were collected with English-language keywords, but it is interesting to note that reasonable performance is still achieved in countries speaking other languages.

Social sensing performance at different event impact levels

A further consideration for impact-based forecast evaluation is the severity of impacts associated with different (in this case, hydro-meteorological) events. Each event logged in the Met Office impact database is assigned a category from 1 (least severe) to 4 (most severe) (Table 5.1). To see how effective the social sensing method is for events with different levels of impact, we plot recall (the number of events in the Met Office database that are matched by peaks in Twitter activity) for different impact severity categories. Figure 5.8 shows recall across a range of percentile thresholds for each impact severity category. This shows that events with the most severe impacts (severity category 4) are more likely to be picked up by the social sensing method. Surprisingly, the least impactful events (severity category 1) achieve the next best recall. This plot also shows us that as the percentile threshold is increased, recall decreases (i.e. more events are missed at the higher percentile thresholds). More on finding the optimum tweet percentile threshold for the social sensing method will be discussed later in Section 5.5.4.

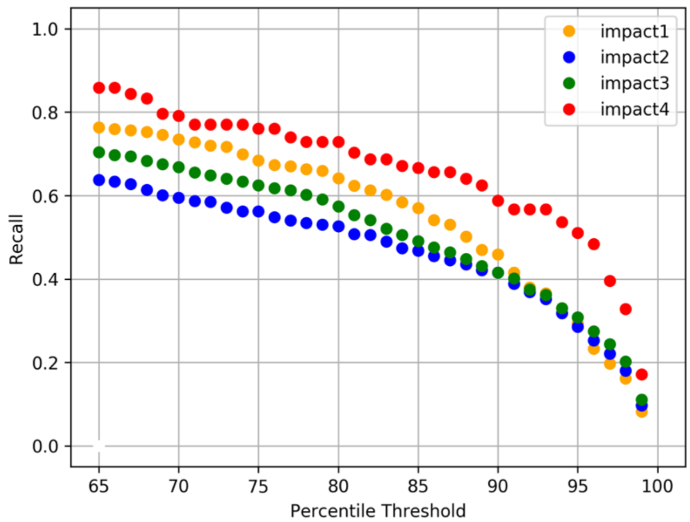

Figure 5.8: Recall versus tweet percentile threshold for matches (within 3 days of event) to the Met Office impact event database for each category of impact severity (where impact severity category 4 represents the most impactful events).

5.5.4 Social sensing performance around the world

Having considered some of the factors which affect performance of the social sensing methodology, we now examine how well social sensing performs in different geographic regions around the world. To do this, we first look at the choice of percentile threshold for different places, then the dependence of social sensing on tweet volumes, before finally examining performance in different GADM Level 1 regions. Again, we assume that the manually curated Met Office impact database is "ground truth", while acknowledging that the actual ground truth is unknown.

Choice of percentile threshold

The optimal tweet percentile threshold overall (yielding the highest F2 score) was found to be around the 80th percentile, however this varies by location. Figure 5.9 plots the optimal tweet percentile threshold for every GADM Level 1 region in which a Met Office impact event was recorded. Where the plot is white in colour, no events were recorded; these regions are not considered in our analysis. The plot shows that the optimal percentile threshold for social sensing performance varies by country (at least, in terms of recovering the known events recorded in the Met Office database). Therefore, the social sensing method may need to use a different percentile threshold for different locations to achieve its best performance.

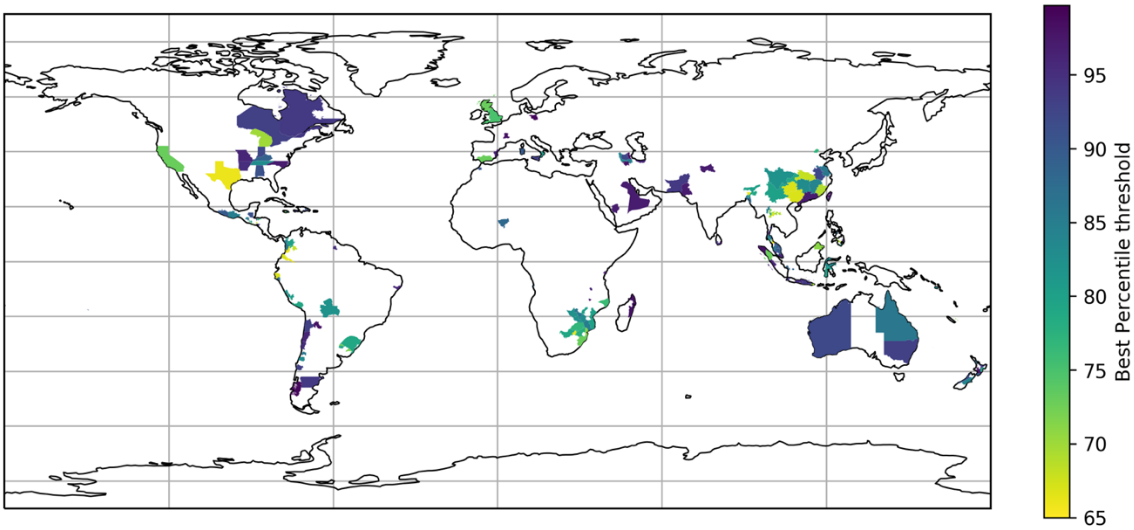

Figure 5.9: Global map showing the tweet percentile threshold which yielded the highest F2 score of precision/recall between filtered heavy rainfall tweet activity and events in the Met Office impact database for each GADM level 1 administrative area with an event recorded in the Met Office database during the study period.

Dependence on tweet volume

It is reasonable to assume that the volume of tweet activity might affect social sensing performance. This leads to an expectation that social sensing will work best in locations with large user populations and resulting large data volumes. To test this assumption, we examined the relationship between F2 scores and tweet volumes for each GADM Level 1 region for which an event was recorded in the Met Office database. Figure 5.10 plots the average tweet count and the maximum F2 score for each location with an event recorded in the Met Office database. The plot shows no obvious relationship between the two variables; this is confirmed by a weak correlation (Pearson's r=0.11, p=0.10). This finding demonstrates that (perhaps unexpectedly) a greater number of tweets does not necessarily mean that the social sensing method will be more accurate. Good performance can be achieved with any volume of tweets, so long as there is temporal variation in volume driven by rainfall events.

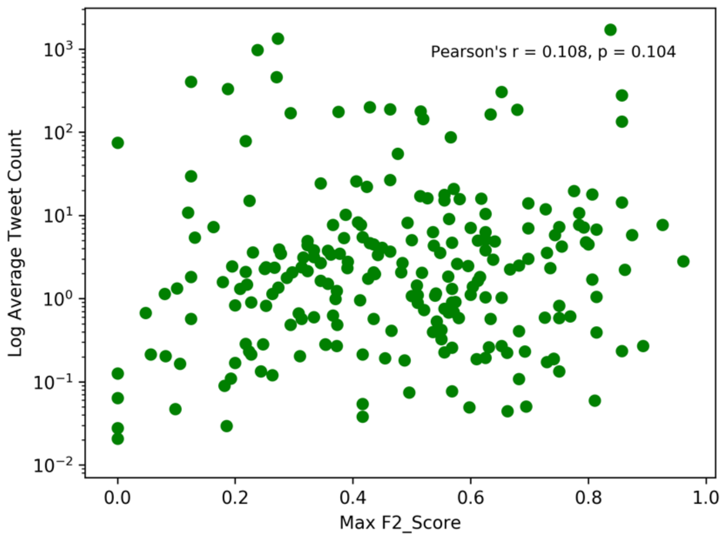

Figure 5.10: Log Average number of tweets versus maximum F2 score for each location with an event in the Met Office database.

Performance of social sensing around the world

The performance of social sensing in different locations across the world was also examined. Figure 5.11 shows the maximum accuracy for each GADM Level 1 administrative area with an event recorded in the Met Office database. Accuracy is calculated based on the proportion of true results among the total number of cases examined with 1 being 100% accuracy, i.e. no false positive or negatives, and 0 being 0% accuracy, i.e. no true events found. Figure 5.11 shows how the accuracy is high for all areas where social sensing was compared to the Met Office database. The maximum accuracy achieved for each area ranges from 86% to 99%. The high accuracy achieved suggests that the social sensing method detected almost all events in the Met Office database. However, as we are also interested in how well our social sensing method detects high impact rainfall events which are not in the Met Office database, the F2 score (which also takes this into account) is likely to provide a more realistic measure of how well, or otherwise the social sensing method detected events in the database.

Figure 5.11 also shows the maximum F2 score for the GADM Level 1 administrative areas with an event recorded in the Met Office database. It is clear from this figure that there are some places where the method works particularly well (e.g. Australia, some parts of the USA, Saudi Arabia) and others where the method doesn't work as well (e.g. Europe, India). This may be in part due to language limitations, as only English language tweets were analysed. It may also be due to some parts of the world where rainfall is more common or the time frame of the study being only 6 months meaning some areas' heavy rainfall (e.g. Indian monsoon) are not included.

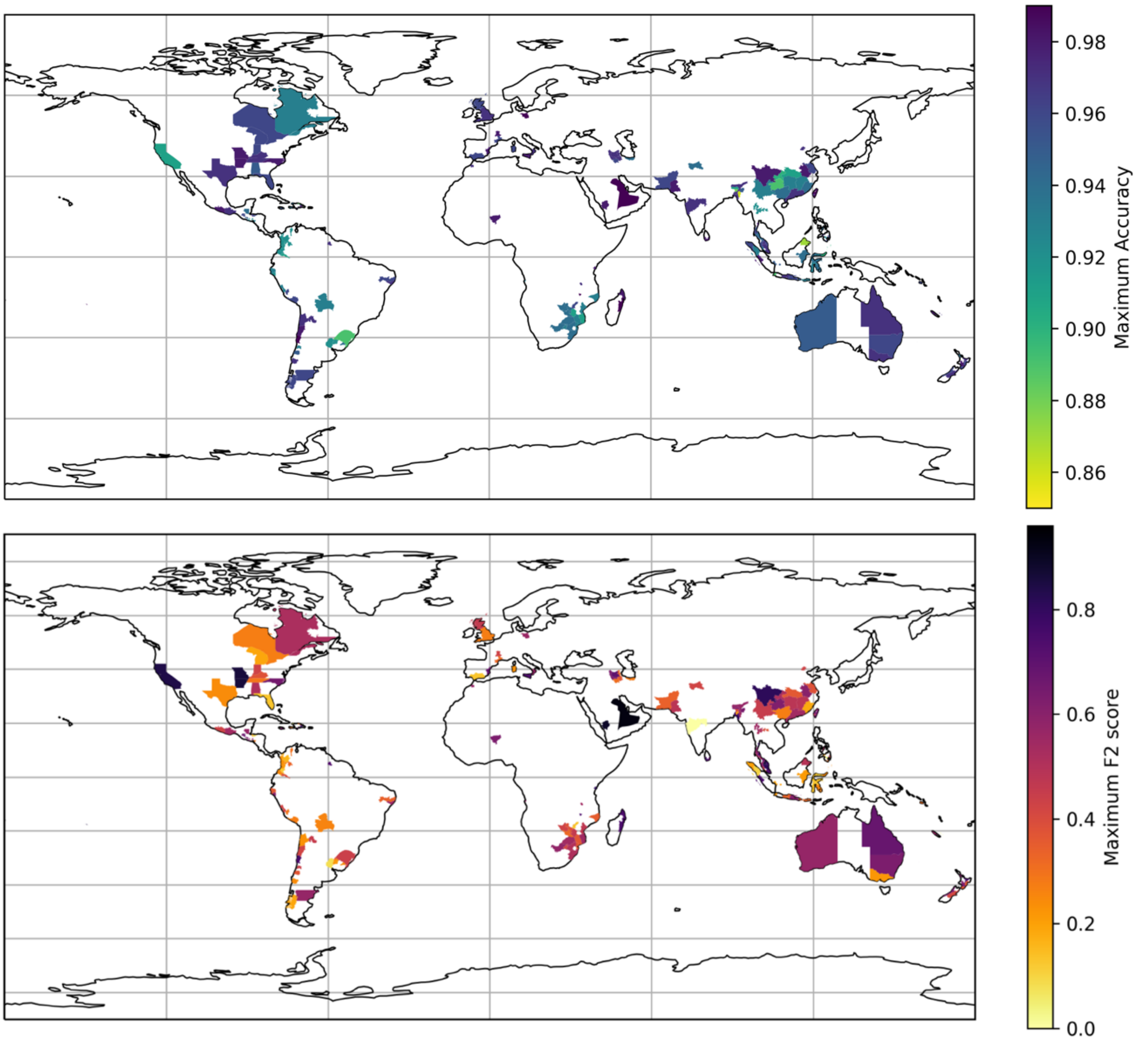

Figure 5.11: [TOP] Global map showing the average accuracy of true positives between filtered heavy rainfall tweet activity and events in the Met Office impact database for each GADM level 1 administrative area with an event recorded in the Met Office database during the period. [BOTTOM] Global map showing the maximum F2 score of precision/recall between filtered heavy rainfall tweet activity and events in the Met Office impact database for each GADM level 1 administrative area with an event recorded in the Met Office database during the period.

5.6 Discussion

This study has shown the potential of social sensing of Twitter data to identify and locate high impact rainfall events across the world. Social sensing can help to support the curation of impact data following extreme weather events, which may in turn support better evaluation of impact-based forecasts and the development of new impact models. The process used to generate the Met Office impact database can produce high quality and detailed records, with few if any false positives. However, manual collection is extremely laborious, resource intensive and ultimately unsustainable for many Meteorological Services. This could be improved by developing automated procedures which accomplish the same goal. Social sensing is one automated approach which could be used to automatically identify events breaching a predetermined threshold. We have seen that social sensing achieves high coverage (few false negatives) thus the addition of a social sensing tool to enhance impact data collection as part of a semi-automated process is very promising and would allow high quality impact data to be collected with significantly reduced manual work.

Comparison of social sensing results with the Met Office impact database identified a number of surprising results which may highlight both limitations in the design of the Met Office database and also opportunities for the two approaches to complement one another. In particular we found that there were a number of events identified in the Twitter data which were not included in the Met Office database. While recorded as false positives when calculating the precision and recall of the social sensing approach, many of these peaks in tweet activity were found to be true events after further investigation. On closer inspection these events would have met the criteria for being assigned an impact severity category and are therefore genuine omissions from the Met Office database. There are a number of possible reasons for this disparity. Firstly, we speculate that there are a number of high-impact rainfall events that occurred but were not captured by Met Office data collection methods, e.g. due to the focus on English-language news sources, or because they did not meet the inclusion criteria of that database. The Met Office database does not include news reports which did not make clear reference to the cause of the impacts. For example, if flooding and associated impacts were reported but did not make clear reference to heavy rainfall as the trigger, then the report would not have been included in the Met Office database. There were also temporal and spatial constraints on report inclusion into the Met Office database so that flood events associated with groundwater or significant fluvial flooding (caused by long-term rainfall over a season for example) were not included. This was because the Met Office Global Hazard Map (GHM) focuses on forecasting daily heavy rainfall events and therefore the impact database was generated with evaluation of those forecasts in mind. By contrast, in the Twitter data an event would be inferred by the volume of discussion about rainfall/flooding alone, without this context. Therefore, differences between the two datasets in this case would be expected. Second, there is a difference in style of reporting between Twitter, which typically provides an individual's identification of a single high-impact event based on their own experience and subjective perception of impact, compared with the dominant sources used to produce the Met Office impact database, which typically try to be objective and tend to aggregate impacts (e.g. news media often report aggregated impacts associated with an event). This means that Twitter data may pick up a greater number of smaller-scale, localised impacts, which are often missed in broader, aggregated sources (e.g. FloodList). Third, we note that the presence of tweets relating to rainfall in a region does not indicate that a major rainfall event occurred. It is likely that many tweets are written in reference to minor or normal rainfall and not in response to an extreme event. However, the disparity in coverage between Met Office data and Twitter data does suggest that the social sensing approach may facilitate more effective wide-scale observation of high-impact rainfall events.

The performance of social sensing was found to be affected by a number of factors, including the event detection time window and location in the world. Time lags of a few days were observed for some events between the rainfall event occurring and the peak in Twitter discussion relating to it. Better precision/recall performance was also observed when the event detection window was extended up to 3 days after the first date of the rainfall event occurring. Therefore, it will be important to take potential time lags into consideration when using social sensing as a potential event detection tool.

It was also found that events in the Met Office impact database were more likely to correlate with events detected using social sensing for English-speaking countries. This is not surprising given that the data collected from Twitter was in the English language and the methods used to collate the records of impact events in the Met Office database also relied on news and media sources in English. While the limitations on language would lead to a clear English language bias in terms of performance, it was encouraging to find that social sensing with

English tweets does still work well in some other-language speaking countries and also that the number of tweets in a location does not adversely affect the social sensing method.

The most impactful events in the Met Office database (impact severity category 4) also returned better success using the social sensing approach than the lower severity categories, which is not an unexpected result given that events of this magnitude are likely to generate more interest in social media channels. What was surprising, however, was that events in severity category 1 had better recall than severity categories 2 or 3. One possible reason for the strong performance of severity category 1 events is because of the style of reporting by Twitter users. Category 1 includes localised impacts and low-level disruption (i.e. disruption to daily life, delays and short-term in-accessibility to services). Given the individualistic nature of Twitter reporting, it is likely that these types of impacts are registered more routinely, while such events have to reach an undetermined significance (in terms of interest) threshold to be reported in the media or in other aggregated data sources. It should also be noted that the frequency of events in each severity category, within the Met Office database, is uneven, with events assigned to severity category 3 far outweighing the number of category 4 events.

5.6.1 Limitations and further work

The main limitation to studies of this type is the lack of data to confirm the absolute truth for validating our findings. In this case there is no definitive list of all impactful heavy rainfall events across the world that we can refer to. While the Met Office database was laborious and time consuming to collect, it is very useful because it pulls information from a wide range of sources; includes all events found, regardless of location in the world; and has clear and consistent criteria for events to be included within it. We have also shown that Twitter is a good source of data for event detection. Therefore, what has been presented in this study is a comparison of two datasets, which if combined together could help to provide a more holistic view of heavy rainfall impacts across the world.

Another limitation for this study is that only 6 months of data was examined. This means that locations which experience high rainfall at different times of the year to the period of this study (e.g. the Indian Monsoon season) would have been under-represented. Any further work in this area should consider extending the timeframe to include all likely weather extremes across the year. This would be important as it will support improved understanding of tweet behaviour between wet and dry seasons where these occur. The underlying tweet counts which were used to calculate percentiles would also benefit from being calculated for a longer time frame (e.g. 3-5 years) rather than just the period of this study. This would likely yield better results in terms of identifying peaks in Twitter data.

Tweaks to the underlying method may also benefit the performance of social sensing for both similar studies to this one and other studies comparing Twitter data with other datasets. In relation to this study, the terms included in the Twitter API search could be extended to be wholly in line with terms used to find news and media sources for the Met Office database. For example, the tweet collection only included the word '*landslide*', however the Met Office database would have also included other terms such as '*mudslide*' and '*landslip*' in searches for news reports. The development of libraries of suitable search terms can be considered somewhat easier for hazards, which often have well defined usage, compared with terms that aim to identify socio-economic impacts. This work has focussed on identifying impacts based on the occurrence of tweets with specific hazard phrases, rather than socio-economic impact phrases. Further analysis of tweet text from filtered tweets to extract information about the types of impacts being experienced by Twitter users would be an obvious next step. This could then be used to further classify the events in line with the Met Office impact severity category criteria or to help to refine impact severity categorisation. It is likely that a combination approach could yield additional insights into the details of high-impact events, but further work would be required to fully establish the utility of Twitter for providing detailed impact assessment.

Extending this study to investigate if tweet activity relating to heavy rainfall (or other weather types) could be monitored globally in real-time would greatly add weight to its long-term utility as a source of impact data. One of the primary limitations of our method is the exclusive use of English. We have demonstrated in Section 5.5.1 that we achieve good global coverage despite this restriction but as shown in Figure 5.7 our ability to detect events is lower in countries where English is not a native language. Applying this methodology in real-time and as a source of impact data on a global scale would require a similar list of key words to be generated in a number of other major languages, especially those popular on Twitter. The subsequent location inference and relevance filtering steps would also have to be optimised to be language agnostic. Though English is the most popular language on Twitter (Mocanu *et al.*, 2013) the majority of tweets are in other languages, with Spanish, Malay and Indonesian making up a significant proportion. We have demonstrated that there is significant benefit to this methodology working with English tweets only, but we must keep in mind this bias and look to add other major languages in future work.

Despite the acknowledged limitations and the recommendations for further methodological work, this study shows that it is possible to use Twitter data to identify high-impact rainfall events and their impacts, globally. Furthermore, the type of record that Twitter provides (i.e. eye-witness accounts, individual reports of events taking place), is different in nature to the aggregated sources that the Met Office database and other similar databases use. Therefore, Twitter data can be used as a ‘first pass’ event detection tool, largely automating the difficult manual curation task. Prototyping this methodology in ‘real-time’ to generate an automated Twitter-based impact database would be the next step. It would also be interesting to repeat the impact-based evaluation methodology described in Robbins and Titley (2018) using a Twitter-based impact database. Based on the findings from this work, we believe that a method that utilises the strengths of both methods (social sensing methodology and media/aggregated data collection from trusted sources) could lead to an enhanced approach for sustainable and robust impact data collection. The generation of a framework to bring these data together would allow the impact-based evaluation method to migrate away from its original, semi-automated approach to a fully automated impact-based evaluation methodology.

5.7 Summary and Conclusion

In this study, data was collected from Twitter in the first half of 2017 relating to mentions of rainfall and the impacts of rainfall across the world. This data was analysed and compared with a manually-curated database of global rainfall events that caused socio-economic impacts collated by the Met Office for the same period of time. The aim was to assess the potential of using Twitter as a source of impact data following a significant weather event. A 'social sensing' methodology was used to apply various computational techniques to filter and extract only those tweets from the dataset of relevance to the impacts of a heavy rainfall event. Tweets without geo-located coordinates were then further processed to infer the location of the tweet, or event mentioned in the tweet, so that the location of the rainfall event could also be determined. Using the percentile of the number of tweets for a particular day and location as a proxy for the likelihood of an impactful event taking place, this accounted for the prevalence of tweets in each location. Comparison of these spikes of activity within the filtered Twitter data with the Met Office database of high impact rainfall events finds that the majority of events recorded by the Met Office were also detected using social sensing. Interestingly, the social sensing approach also found additional impactful rainfall events within the Twitter data which were not recorded in the Met Office database. It was also encouraging to find that social sensing with English tweets still worked well in some other language speaking countries and also that the number of tweets in a location does not adversely affect the social sensing method. This suggests that social sensing of Twitter data would be a useful addition to current impact data collection processes.

Chapter 6 - Social sensing of heatwaves in European cities

6.1 Introduction

Summer 2019 was one of the hottest summers on record in Europe with many places in Western Europe, including the UK and the Netherlands, breaking temperature records. Two record-breaking heatwaves occurred in June and July 2019 (Vautard *et al.*, 2020). These heatwaves were recognised as the deadliest natural disaster in the world for 2019 with an excess mortality toll of 2500 people (CRED, 2020). During the July heatwave, the Netherlands recorded, for the first time ever, a temperature above 40^{o}C and in the UK a new highest ever temperature of 38.7^{o}C was measured in Cambridge (Vautard *et al.*, 2020). In Greece, two heatwaves were also experienced in 2019, in early July and early August. Heatwaves have severe impacts on air quality, economy and the ecosystem and are listed as the leading cause of weather-related mortality (Pyrgou & Santamouris, 2018; Xu *et al.*, 2020). Increased temperatures have been shown to exacerbate drought, increase the likelihood of wildfires, and to impact agriculture and therefore food security. Heat also has health implications, such as increasing the risk of cardiovascular and respiratory complications as well as kidney disease. Certain demographic groups, such as those on low incomes, are also more vulnerable to heat stress (where the body becomes unable to cool itself) as they have less access to cooling activities (Dean, 2021). The extreme impacts of heatwaves, such as mortality rates, wildfires and economic losses are reasonably well documented, however the less extreme social impacts are likely under-reported and not fully understood (Red Cross Red Crescent Climate Centre, 2020).

Heatwaves are experienced differently depending on local climate and preparedness. Towns and cities are more likely to experience higher temperatures than rural areas due to the Urban Heat Island (UHI) effect (Ward *et al.*, 2016; Zhao *et al.*, 2018). More than half of the world's population resides in urban areas (Ritchie & Roser, 2018). The world's population has increased by almost 6 times from 751 million in 1950 to 4.2 billion in 2018 (Corneille, 2020). This has led to the intensification of urbanisation and therefore the increased impacts of heat on cities. Some of the reasons for higher temperatures in cities include increased land coverage using man-made materials such as concrete, tarmac and asphalt. These man-made materials generally absorb and store solar energy during the day and release it at night. The lack of vegetation in urban areas also reduces evaporation rates so that local air temperature is increased through heating from the sun. Furthermore, the geometry of buildings in cities means that heat gets trapped close to the surface and is not easily radiated to space. Higher population density in cities also produces more heat via human activities, such as heat produced by car engines and air-conditioners. All of these factors contribute to turning cities into “urban heat islands” (UHIs) (Ramamurthy *et al.*, 2017; Heaviside, 2020; Dean, 2021). Due to the UHI effect, cities are more likely to experience higher temperatures, particularly at night. The UHI effect is therefore likely to bring more severe impacts during a heatwave event (Kong *et al.*, 2021).

Resilience planning is a major consideration for large cities, whose civil authorities must respond to ongoing events and also plan for a changing climate (UNDRR, 2013; Resilient Cities Network, 2020). Extreme weather events, such as heatwaves, are increasingly challenging and costly to manage. For heatwaves, city authorities currently estimate vulnerability based on demographics and built infrastructure, but beyond health data on hospitalisations/ambulance call-outs, they have little information about where and when heat is a problem in the city (Kong *et al.*, 2021). Obtaining relevant impact data during heatwaves is therefore slow and costly, with city officials and organisations turning to news reports, citizen data and insurance data (Young *et al.*, 2021). However, during the 2019 European heatwave news media sources tended to focus on creating a good story and linking soaring temperatures to climate change, making local and national news articles unreliable as a source of information about social impacts as a result of the heatwave (Strauss *et al.*, 2021). Furthermore, there can often be a time lag before sufficient robust data from these sources is available. Therefore, there is a need for other sources of information to support city officials. Redesigning cities to be resilient towards heatwaves is a huge and costly undertaking (Dean, 2021) and therefore city governments and local organisations need timely access to data which enables them to fully understand the social impacts of heat in cities (Victorian Auditor-General, 2014; Kong *et al.*, 2021).

Social sensing is the systematic analysis of unsolicited social media data to observe and characterise real-world events (Wang *et al.*, 2019). Social sensing using Twitter data to detect and locate social impacts of natural hazards has been successfully used to detect impacts of a variety of weather hazards, including wildfires (Boulton *et al.*, 2016), floods (Arthur *et al.*, 2018), pollen (Cowie *et al.*, 2018), storms (Spruce *et al.*, 2020 (Chapter 4)), high winds (Weaver *et al.*, 2021), extreme rainfall (Spruce *et al.*, 2021 (Chapter 5)) and heatwaves (Young *et al.*, 2021). In particular, the latter study by Young et al. (2021) provides good evidence that social sensing during heatwaves provides data which may be useful to city officials. The social sensing methodology applies programming and data science techniques in several stages:

1. Collection of social media data (primarily Twitter) relating to extreme weather hazards;
2. Filtering the data using machine learning techniques to retain only relevant data;
3. Location inference to determine the geographic areas associated with each data point;
4. Mapping and visualisation using geospatial tools.

Social sensing has been shown to be able to provide real-time information about unfolding events, which may support better civil responses. Therefore, based on findings from previous studies (Arthur *et al.*, 2018; Spruce *et al.*, 2020, 2021 (Chapter 4, Chapter 5)), it has potential to be able to provide information on the health & wellbeing impacts of heatwaves, transport disruption from heatwaves, and the effect of city responses and communications. This information could be used to better manage situations and direct emergency responses. Additionally, data gathered over time from social sensing can be used to guide resilience policy-making and hazard mitigation strategies, e.g. identifying which parts of a city are most vulnerable to impacts as a result of heatwaves; understanding the disruptive impacts of heatwaves on daily life and economic activity; and tracking incidence of extreme weather impacts over long timescales, including variation caused by climate change. This information can be used to evaluate historic events and improve policies for responding to extreme weather events, such as heatwaves.

A number of approaches and platforms have been studied for use in the social sensing of weather, e.g., image tags on Flickr (Tkachenko *et al.*, 2017), Facebook and WhatsApp (Bhuvana & Arul Aram, 2019), custom mobile apps (Wang *et al.*, 2018), among many other studies related to the social sensing of weather events, which were discussed in Chapter 2. However, Twitter is the platform that has been used most extensively for the social sensing of weather events (Reuter & Kaufhold, 2018; Reuter *et al.*, 2018; Zhang *et al.*, 2019). Twitter has a public API, global user base and focus on current events (the Twitter prompt for a user post is “What’s happening?”), which makes it an ideal platform for social sensing. While Twitter can be used to build a map of where people are talking about a particular topic, it is somewhat less ‘social’ than other social networks as it does not enable the formation of online groups to chat or discuss local issues. This means that people are unlikely to use it to coordinate community response or offer aid. For this, other social networks like Facebook and WhatsApp are likely to be crucial, though for researchers these platforms are inaccessible without data sharing agreements with the companies. However, given Twitter’s data accessibility, availability in real-time, and propensity for Twitter data to originate from more densely populated areas, such as cities (Arthur & Williams, 2019), it provides the most obvious choice in terms of exploring the social impacts of heatwaves in cities.

In this study, discussion on social media (Twitter) will be explored using social sensing techniques to reveal the social response to the 2019 summer heatwave in three European cities (London, The Hague and Athens). Each of these cities was chosen as resilience planners and city officials there have expressed an interest in using social sensing to better understand the impacts of heatwaves in their respective cities. London in the UK and The Hague in The Netherlands, both experienced unusually high temperatures as a result of the summer 2019 heatwave (Vautard *et al.*, 2020). Athens is well known as a ‘hot spot’ with respect to high temperatures and heat-related risk, with an increase in the frequency and intensity of heatwaves there in recent years (Katavoutas & Founda, 2019). Therefore, this study will analyse Twitter data from summer 2019 as a case study to explore the potential for developing a tool for city resilience planners to access spatial/real-time social media information about heatwave impacts as they are occurring.

6.2 Related Work

This section provides a brief literature review of social sensing applied to heatwaves. It then considers previous studies that have applied social sensing techniques in the UK, The Netherlands and Greece.

6.2.1 Social sensing of heatwaves

There are a small number of studies which have explored social media use during heatwaves. Watson & Finn (2014) examined the use of social media during the UK heatwave of 2013 to better understand who and what was being communicated on social media during a heatwave event. Some studies have focused on exploring people's perspective on the weather during heatwaves or periods of high temperature (Austin, 2014; Jung & Uejio, 2017). The effect of weather on people's mood has also been explored by extracting emotions from a set of general tweets and then relating them to historical weather data (Hannak *et al.*, 2012). Previous studies have also found a positive relationship between maximum air temperature and tweet activity in cities relating to heat-related themes (Jung & Uejio, 2017; Grasso *et al.*, 2017). Cecinati *et al.* (2019) use social media posts from Twitter to complement climatic data in tracking heatwave events in real-time and to identify heatwave mortality in India. More recently, another study explored the concept of human comfort in outdoor spaces (HCOS) comparing the sentiment of tweets with related atmospheric data (Giuffrida *et al.*, 2020). The social sensing of the 2018 heatwaves in the UK, Australia and the USA was also successfully explored by Young *et al.* (2021), identifying the relationship between tweet activity and temperature. The authors also further explored the Urban Heat Island concept in London, Sydney and New York. Therefore, there is good evidence that social sensing of heat-related tweets is likely to provide information about the social impacts of heatwaves in cities.

6.2.2 Social sensing of natural hazards in the United Kingdom, The Netherlands and Greece

Social sensing of natural hazards in the United Kingdom has been carried out with good success. The social sensing of floods has been shown to work very well in the UK context for flood inundation mapping (Rosser *et al.*, 2017; Brouwer *et al.*, 2017; Smith *et al.*, 2017), flood event detection (Arthur *et al.*, 2018; Barker & Macleod, 2019) and to support situational awareness (Saravanou *et al.*, 2015).

The use of social media to support the curation of the National Landslides database has also been explored (Foster *et al.*, 2012). The social sensing of storms/high winds in the UK has also been carried out with good success (Gray *et al.*, 2016; Spruce *et al.*, 2020 (Chapter 4); Weaver *et al.*, 2021). Furthermore, as already discussed, Young *et al.* (2021) found that the social sensing of heatwaves in the UK is also possible with good success.

In the Netherlands, there do not appear to be any specific social sensing studies relating to natural hazards. However, de Bruijn *et al.* (2019) have used social sensing with multi-language tweets, including the Dutch language, for flood detection and to build a global database of flood events. These authors have shown that the use of social media, in particular Twitter, works well in the Dutch context for flooding. Therefore, there is a good expectation that social sensing would also prove successful for heatwave events.

There is also currently limited academic literature related to the social sensing of natural hazards in the Greek context. The main studies which have explored the use of social media to detect natural hazard events in Greece have been produced by one group of researchers centred on Stathis Arapostathis. Using a dataset of tweets containing relevant keywords from the period around the Messinia flooding event in September 2016, good agreement is found between peaks in the Twitter activity and the location of flooding, as well as using tweet classification to identify the severity of impact (Arapostathis, 2019, 2021). Other case studies relating to other natural hazards have also been explored using Twitter data in Greece, e.g. earthquakes (Arapostathis *et al.*, 2018) and wildfires (Arapostathis & Karantzia, 2019).

6.3 Previous work

Previous work (Young *et al.*, 2021) showed that social sensing of heatwaves using English-language tweets successfully detected heatwave events during the summer of 2018 in the UK, USA and Australia. In each country, there was a slightly different response, with more negative sentiment found in Australia (where heatwaves are associated with significant drought and other environmental impacts) and positive sentiment in the UK (where heatwaves are often seen by the public as an opportunity for outdoor leisure) (Figure 6.1).

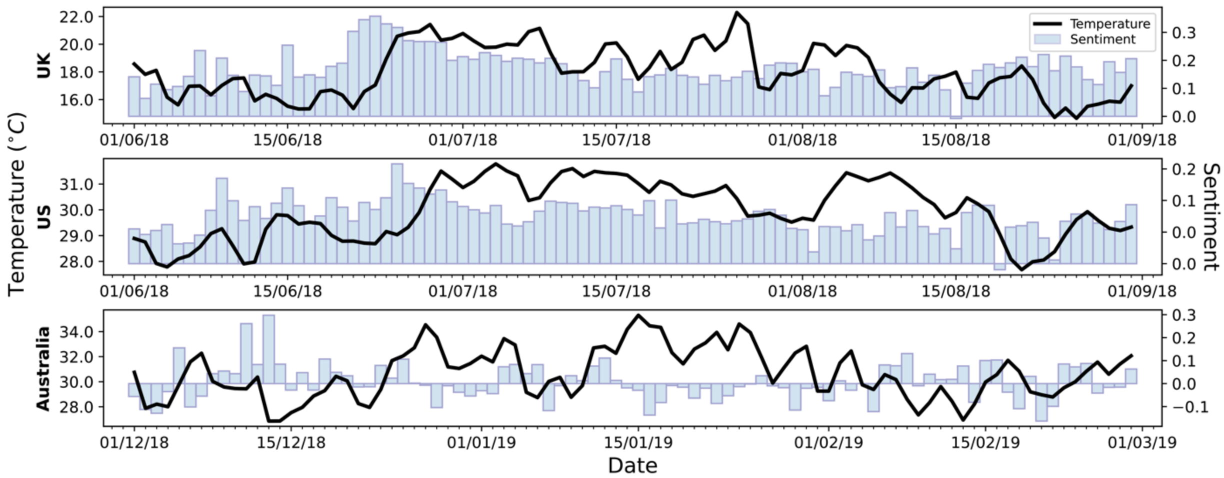

Figure 6.1 - Bar chart of sentiment changes throughout the summer months from tweets within each country, with the daily average maximum temperature plotted in black. Please note: the dates on the plot for Australia reflect the different time of year for the Australian summer season (Young et al., 2021).

Across all countries, the volume of heatwave-related tweets increased as temperature increased, showing public attention to the topic (Figure 6.2).

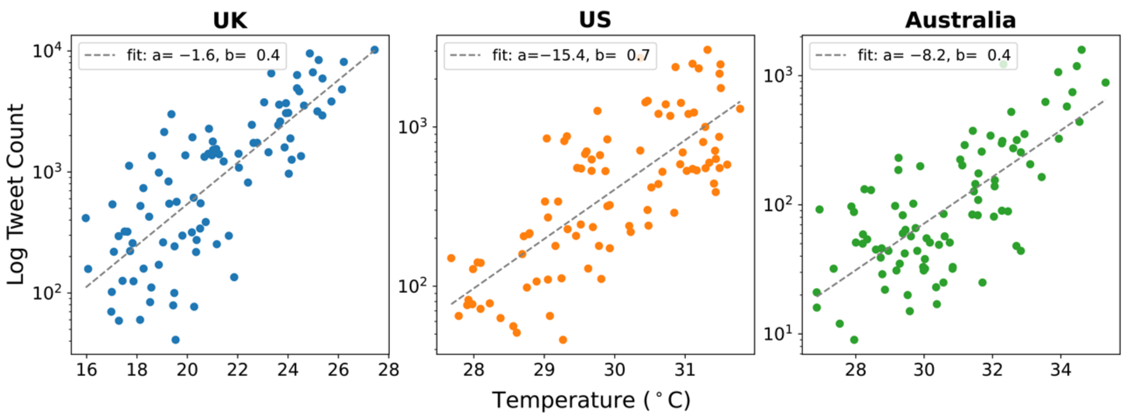

Figure 6.2 - Scatter plot comparing the log daily tweet count to the average daily maximum temperature for each country. The line of best fit is modelled by log(y) = a + bx (Young et al., 2021).

The content of tweets varied by country, with the UK discussion tending to be focused on heatwave activities, such as ‘ice-cream’, ‘world cup’, ‘beach’, ‘sunshine’ and ‘garden’, as well as personal discomfort and inconvenience such as ‘sleep’ and ‘hosepipe ban’; Australia discussion was more focused on national issues, such as ‘climate change’, ‘temperature records’ and ‘childrens futures’; and the US discussion focused again on national issue, such as ‘natural disaster’, ‘climate change’ and ‘fire’.

At the city-scale, this work also provided good evidence that the methods can be effective in large cities, including London, New York and Sydney (Figure 6.3). However, it is unclear whether the same social sensing methods will be as effective in determining heatwave impacts for The Hague and Athens due to tweets being in different languages and the smaller population size of these locations.

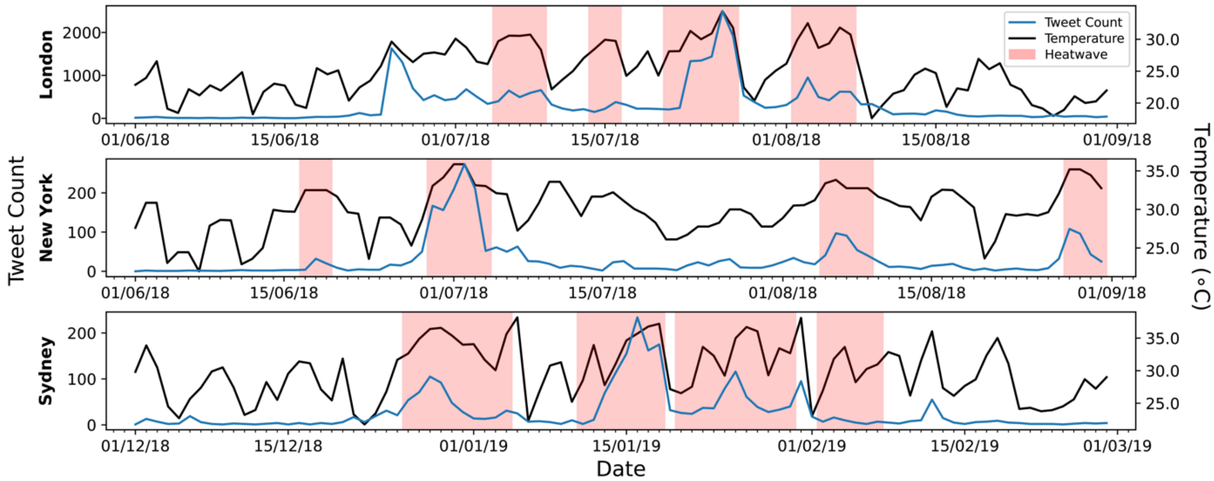

Figure 6.3 - Daily tweet count in the investigated cities, overlaid with the government defined heatwave days and the average daily maximum temperature (Young et al., 2021).

6.4 Research Aims

This study will explore discussion on social media (Twitter) during the 2019 summer heatwaves in three European cities (London, The Hague and Athens). The general aim is to build on previous work on the social sensing of heatwaves (Young *et al.*, 2021), discussed above, to evaluate the potential for city-scale analysis of Twitter data during heatwaves and/or periods of high temperature. Three cities (London, The Hague and Athens) were used as case studies for the project as resilience planners in these cities have expressed a particular interest in using social media to understand social impacts during heatwave events. Even though record-breaking temperatures were also experienced in the European heatwaves of 2020, the 2019 heatwave was chosen as a case study because it pre-dates the Covid-19 pandemic and therefore social media discussion is less affected by the discussion of impacts related to Covid-19.

The aim is to understand the effectiveness of social sensing at the geographic scale of a city and to determine whether there is useful information available for city officials to identify impacts of the heatwave and help manage the heatwave response. Social sensing depends on high volumes of social media data from accessible platforms, i.e. platforms where data can be easily obtained and analysed without violating user expectations around privacy. It is currently unclear whether sufficient volumes of data can be obtained in the context of The

Hague and/or The Netherlands as well as Athens and/or Greece. Therefore, this study will focus on the following:

1. Check the availability of useful social media data during heatwaves in London, The Hague and Athens.
2. Perform some exploratory analysis to determine the type of information about social behaviour during heatwaves that social media may provide.
3. Illustrate how social media during heatwaves may be useful to local authorities and policy decision makers in cities.

6.5 Methods

This study will use and build on the methods from the previous study by Young *et al.* (2021) and also uses methods from previous work on the social sensing of UK/Ireland storms discussed in Chapter 4 (Spruce *et al.*, 2020) and global rainfall discussed in Chapter 5 (Spruce *et al.*, 2021).

6.5.1 Data collection

Social media data during the summer of 2019 was collected using the Twitter API V2 Academic Track[30]. Twitter data collected included the English, Dutch and Greek terms shown in Table 6.1. A wider variety of heatwave/heat activity-related terms were collected than in previous work (Young *et al.*, 2021) based on a review of tweets collected in that study. For each language collection, the Twitter API search query was also restricted to tweets where the language field was populated with the corresponding language code (as shown in the language column of Table 6.1). The use of the language field, particularly for Dutch and Greek, helped to restrict search results to the countries of interest. This is more difficult with English language tweets as English is used by many Twitter users in many countries. Therefore, the English language dataset was very large in comparison to the Dutch and Greek datasets.

[30] https://developer.twitter.com/en/products/twitter-api/academic-research (Accessed: 17 March 2022)

Language	Search Terms
English (EN)	heatwave, heat, warm, warmest, hottest, "too hot", sun, sunny, boiling, roasting, melting, burn, sunburn, heatstroke, sunstroke, humid, humidity, sweat, sweaty, sleep, tired, "keep cool", suncream, aircon, "air conditioning", ac, fan, "ice cream", bbq, barbeque, beach, swimming, "cool down", "hot weather", "hot day", "so hot", "very hot", summer, "how hot"
Dutch (NL)	hittegolf, hitte, gloed, zoonegloed, warmte, warmste, heetste, "te heet", zon, zonnig, kook, smeltend, zonnebrand, zonnesteek, vochtig, nattig, benauwd, vochtigheid, zweet, zweterig, slaap, moe, "blijf kalm", zonnecreme, airconditioning, ijsje, barbecue, strand, scheveningen *(most famous beach area of The Hague)*, zwemmen, afkoelen, "warm weer", "hete dag", "zo heet", "heel heet", zomer, "hoe heet", rokjesdag *(dutch saying 'skirt day' referring to the first hot day of the summer)*, bloedheet, "de mussen vallen van het dak" *(expression that it's really hot)*, "grote vakantie" *(summer holidays)*, komkommertijd *(period in which media/tv makes use of a lot of reruns and daily shows are on a break)*
Greek (EL)	"θερμό κύμα", θερμότητα, ζέστη, θερμότης, καύσωνας, ζεστός, "πιο ζεστό", "πιο καυτά", "πάρα πολύ καυτό", ήλιος, ηλιόλουστος, ευήλιος, ηλιοφώτιστος, βρασμός, ψήσιμο, τήξη, "ηλιακό έγκαυμα", ηλιοκαίω, θερμοπληξία, ηλίαση, υγρός, υγρασία, ιδρώτας, ιδρωμένος, ύπνος, κουρασμένος, "κρατήστε δροσερό", "αντηλιακή κρέμα", κλιματισμός, αιρ-κοντίσιον, "τεχνητός αερισμός", ανεμιστήρας, παγωτό, μπαρμπεκιου, ψησταριά, παραλία, κολύμπι, κρυώνω, "ζεστός καιρός", "ζεστή μέρα", "τόσο καυτό", "πολύ ζεστό", καλοκαίρι, θέρος, "πόσο ζεστό"

Table 6.1 - Keywords included in the 'heatwave' tweet collection for each language of interest: English, Dutch and Greek. Google translate was used to translate English search terms to their corresponding Dutch/Greek equivalent, where one exists. Additional Dutch colloquialisms were also added following advice from colleagues in The Hague.

Heat-related tweets were collected in the Dutch and Greek language for the period 1st May 2019 – 30th September 2019 to cover the entire summer period. Due to the large volume of heat-related tweets returned in the English language, the time period was restricted to only the official heatwave periods in the United Kingdom, 28-30th June and 24-26th July 2019.

6.5.2 Relevance filtering and location inference

Using the same social sensing methods successfully used in previous studies (Arthur *et al.*, 2018; Cowie *et al.*, 2018; Young *et al.*, 2021) and outlined in detail in both Chapter 4 (Spruce *et al.*, 2020) and Chapter 5 (Spruce *et al.*, 2021), the Twitter data was first filtered for relevance to discussion about and/or the impacts of heatwaves and then information within the tweet metadata was used to infer the location of the tweet.

Filtering for relevance to heatwaves

Relevance filters were applied to each tweet dataset (English, Dutch, Greek) separately, removing the following from each dataset:

- retweets and quotes (which are duplications of original content posted by other users)
- tweets posted by identified 'bot' accounts and automatic weather stations (which automatically create Twitter posts without human input)
- tweets containing irrelevant phrases or references (e.g. song titles, pop music such as references to BTS, sexual and advertising terms)
- tweets that were identified as irrelevant using a Multinomial Naïve Bayes machine learning classifier (see below)

For the final stage of relevance filtering, a sub-sample of 1000 tweets from each language dataset (after removing retweets/quotes, bots and irrelevant terms) was used to create a training corpus for a Multinomial Naïve Bayes machine learning classifier, which was found to perform well in previous work (Young *et al.*, 2021). This tweet extract was tagged for relevance to discussion of heatwaves and/or the impact of heatwaves. However, as the researcher is not a Greek or Dutch speaker, it was necessary to translate the text field of tweets not in the English language to English using Google Translate[31] in order to tag a tweet as relevant or not. Using the tweet ID, the original tweet text of the Greek/Dutch tweets was returned to the training corpus and tweets were therefore filtered for relevance using the original language of the tweet text, not the English translation.

The performance metrics, using the training dataset, for each set of language tweets is shown in Table 6.2.

Language	Accuracy	Precision	Recall	F1 Score
English	0.854	0.877	0.922	0.898
Dutch	0.868	0.622	0.691	0.654
Greek	0.885	0.456	0.602	0.519

Table 6.2 - Performance metrics for each language dataset trained using a Multinomial Naive Bayes classifier

As can be seen in Table 6.2, the machine learning classifier performed much better for English tweets, than for the Dutch and Greek datasets. This could be

[31] https://translate.google.co.uk/ (Accessed: 17 March 2022)

because the English dataset contained tweets mainly from days in June and July 2019 with a high temperature and therefore contained a higher proportion of tweets relevant to the heatwave, whereas the Greek and Dutch datasets spanned a larger time period and therefore contained more irrelevant tweets. The different structure and alphabet used for the Greek language is also a likely reason why a classifier trained using an English translation may not have worked as successfully.

Recall is also noticeably higher for each training set, suggesting the classifier is more likely to tag a tweet as relevant to heatwave discussion than irrelevant. Given the need to obtain as much impact information as possible from the tweet data, this is a preferable result. Even though F1-score for the Dutch and Greek datasets is quite low, for the purposes of this investigation, the performance metrics were deemed good enough to proceed with the next steps of the study. However, fine-tuning the classifier for Dutch and Greek language tweets is an obvious area of further development. Table 6.3 provides the number of tweets in each dataset before the filtering stages and after filtering for relevance. The Dutch dataset retains the largest proportion of tweets after the relevance filtering steps.

Language	**Total unfiltered tweets**	**Retained after relevance filtering**	**%Retained**	**Location information in relevant tweet**	**%With Location**
English	10,085,286	2,430,597	24.1%	1,281,484	52.7%
Dutch	5,215,084	2,190,831	42.0%	1,088,798	49.7%
Greek	597,725	154,519	25.9%	64,532	41.8%

Table 6.3 - Number of tweets in each dataset before relevance filtering, after relevance filtering and with location information contained within tweet.

Location inference

For tweets identified as relevant using the above filtering steps, and where geo-location coordinates were not present in the tweet, place names in the tweet metadata (user location and/or tweet text) were identified and coordinates of these places returned by linking the place name to an external database of place names including Geonames [32] and DBpedia [33]. Location information at the country, county/municipality and city scale were determined using location inference methods also used in the previous studies mentioned above, with the

[32] http://www.geonames.org/ (Accessed: 17 March 2022)
[33] https://wiki.dbpedia.org (Accessed: 17 March 2022)

addition of variations of place names in the relevant Dutch and Greek language, as well as just the English variation of place names. Table 6.3 provides the number and percentage of tweets in each dataset with location information detected. The English dataset has the highest number of tweets where a location could be identified. The Greek dataset has the lowest number of tweets with location information. However, this is not surprising given the difficulties of identifying place names in the Greek language using methods previously set up using English variations of place names.

6.5.3 Comparison of results to other data sources

For comparison of tweet activity with meteorological conditions during the heatwave, the maximum daily temperature for each city was used. This was taken from the NOAA 'Global Surface Summary of the Day (GSOD)' dataset[34].

6.6 Results

6.6.1 Tweet timeseries

English heat-related tweets

Once the location of each heat-related tweet in the English dataset was determined, it was checked to see what proportion were located in the UK and then how many were located in London. Of the heat-related English language tweets, 36% were located in the United Kingdom (463,127 out of 1,281,484 tweets). Examining the location information for the remaining tweets in the dataset, the majority of other English language tweets were located in the United States and Australia, which is not surprising given these are both native English language speaking countries, with a huge volume of Twitter users. The number of tweets located in London accounts for 18% of the total tweets located in the United Kingdom (83,080 out of 463,127 tweets). The population of London is approximately 14% of the overall population of the United Kingdom (2019 population: UK 67.5 million; London 9.1 million (United Nations, 2022)) therefore, a slightly greater proportion of tweets in London compared with the population of the country seems appropriate, particularly given that greater Twitter use is generally found in cities (Arthur & Williams, 2019).

[34] https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day (Accessed: 17 March 2022)

Due to the large size of the English tweets dataset, and given that good success was found for social sensing heatwaves in London in previous work (Young *et al.*, 2021), it was decided that it would be most efficient to only collect and process tweets for days during official heatwave periods. The number of relevant heat-related tweets after filtering for relevance and location is shown in Table 6.4. Higher temperatures were experienced in the UK heatwave from 24th-26th July 2019, peaking on 25th July, therefore it is not surprising to see the largest number of heat-related tweets in this heatwave period.

Date	Number of tweets located in the UK	Number of tweets located in London	Minimum daily temperature (LONDON) °C	Average daily temperature (LONDON) °C	Maximum daily temperature (LONDON) °C
28/06/2019	47,345	6,165	13.4	17.9	24.0
29/06/2019	60,957	11,891	14.0	24.4	33.9
30/06/2019	36,265	2,590	14.6	19.5	24.8
24/07/2019	97,728	13,078	20.0	25.9	32.8
25/07/2019	147,874	36,018	19.7	28.6	37.2
26/07/2019	72,958	13,338	18.0	22.1	25.4

Table 6.4 - Number of English language heat-related tweets located in the United Kingdom for each heatwave day, after filtering for relevance and location, with average daily temperature.

Dutch heat-related tweets

Of the total number of heat-related Dutch tweets filtered for relevance to heatwave discussion and containing location information, 23% were located in the Netherlands (247,122 tweets out of 1,088,798 tweets). This suggests a high number of tweets in the Dutch language are posted by users in other countries. Examining the location information determined, the majority of other tweets were posted by users in the United States, Australia, Belgium (which also has Dutch as a native language) and the United Kingdom.

Figure 6.4 shows the number of heat-related Dutch tweets filtered for relevance to heatwaves and located in the Netherlands for the entire summer period 1st May to 30th September 2019, with the periods of official heatwaves for the country also highlighted. The number of tweets located in The Hague accounts for 6.3% of the total tweets located in The Netherlands (15,496 out of 247,122 tweets). As the population of The Hague is approximately 4% of the overall population of the Netherlands (2019 population: Netherlands 17.1 million; The Hague 0.7 million

(United Nations, 2022)) this seems about right in terms of the proportion of tweets.

There are also clear peaks in heat-related tweet activity in line with daily temperature, with the most activity being at the same time as the official heatwave periods. There are also peaks in tweet activity in line with higher temperatures on 2nd June and 26/27th August. The tweets located in The Hague show the same pattern of peaks in tweet activity as the overall tweets for the Netherlands.

The average daily sentiment score is also shown in Figure 6.4. While there is no clear tendency for higher or lower sentiment score in line with higher temperatures, it is noticeable that the second heatwave of the period (24-26th July) and the days afterwards saw a lower average daily sentiment score than the overall average. This was also the period with the highest temperatures, which may therefore suggest that excessive heat is likely to result in lower sentiment score.

Figure 6.4 - [TOP] Number of Dutch tweets containing heat related keywords (after filtering for relevance and location in The Netherlands) for the period 1st May 2019 - 30th September 2019. [MIDDLE TOP] Of Dutch tweets shown in Top plot, number of tweets located in The Hague. [MIDDLE BOTTOM] Average daily sentiment score for the same period. [BOTTOM] Average, maximum and minimum daily temperature for Hoek Van Holland (closest weather station data to The Hague).

Greek heat-related tweets

Of the total number of heat-related Greek tweets filtered for relevance to heatwave discussion and containing location information, 45% were located in Greece (29,030 out of 64,532 tweets). While a greater proportion than was found for Dutch tweets, this suggests a high number of tweets in the Greek language are also posted by users in other countries. Examining the location information of tweets not located in Greece, the majority of other tweets in the Greek language were posted by users in Cyprus (where Greek is also a native language), United States, United Kingdom and Australia.

Figure 6.5 shows the number of heat-related Greek language tweets filtered for relevance and located in Greece for the entire summer period from 1st May to 30th September 2019, with the periods of official heatwaves for the country also highlighted. The number of these tweets specifically located in Athens is also shown. There are a large proportion of the overall tweets for Greece located in Athens (66.1%), which is quite high even though Athens has approximately a third of the Greek population (2019 population: Greece 10.5 million; Athens 3.1 million (United Nations, 2022)). This could be indicative of the fact that Athens is considered a 'hot-spot' during periods of higher temperature (Katavoutas & Founda, 2019), therefore the impacts of heat are more likely to be felt by people there.

On visual inspection, there do appear to be higher volumes of heat-related tweet activity in line with daily temperature, with the main peaks in tweet activity being at the same time as the official heatwave periods in Greece (7th-10th July and 31st July-2nd August). The peak in tweet activity on 1st June was actually due to many Greek users on Twitter posting good wishes about the official first day of summer, which were not removed by the relevance filter. This would also explain the higher sentiment score observed on this day. There are also peaks in tweet activity in line with higher temperatures on 4/5th July, 12/13th August and 26th August. The tweets located in Athens show the same pattern of tweet activity as the overall tweets for Greece.

The average daily sentiment score is also shown in Figure 6.5. Sentiment does appear to be lower during the period of higher temperatures in July and August which is in line with findings from examining the sentiment of Dutch tweets during periods of higher temperature.

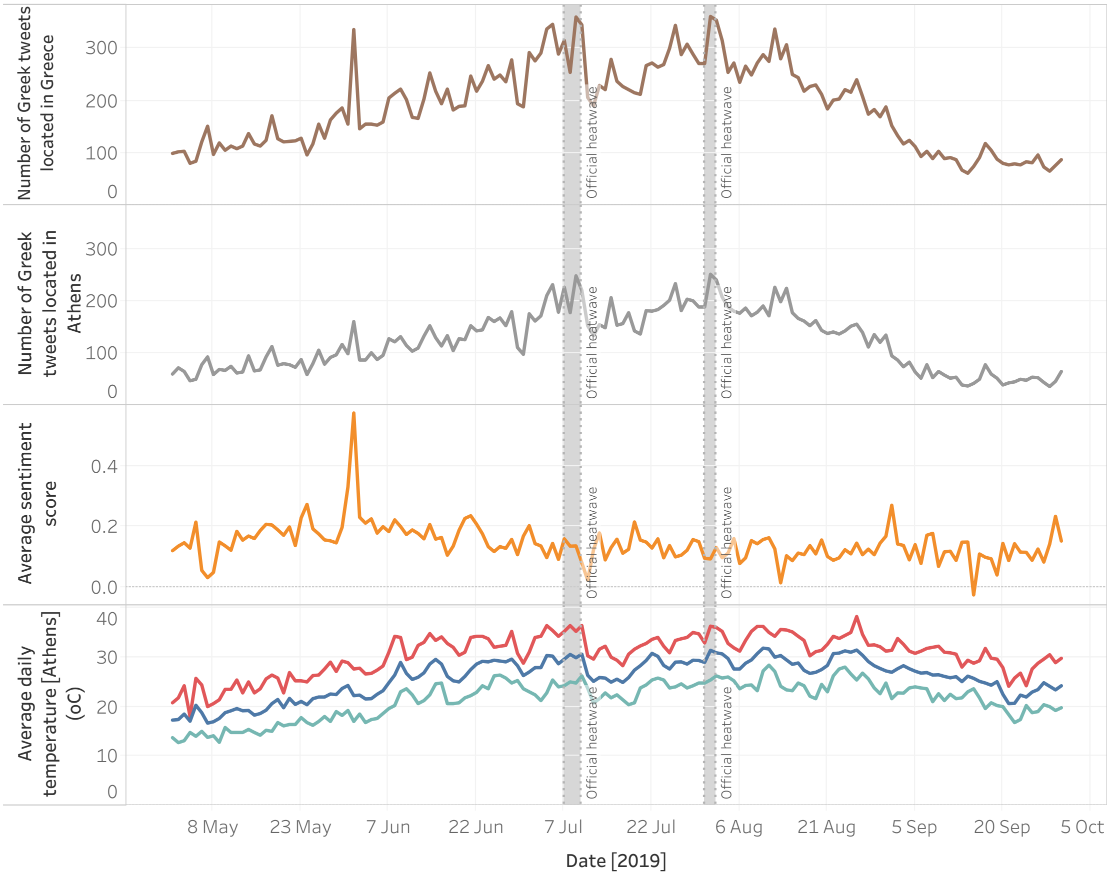

Figure 6.5 - [TOP] Number of Greek tweets containing heat related keywords (after filtering for relevance and location in Greece) 1 May to 30 September 2019. [MIDDLE TOP] Of Greek tweets shown in Top plot, number of tweets located in Athens. [MIDDLE BOTTOM] Average daily sentiment score for the same period. [BOTTOM] Average, maximum and minimum daily temperature in Athens for the same period. The time period for the two official heatwaves in Greece in 2019 are shown by the grey bars.

6.6.2 Temperature vs Number of Tweets

Visual inspection of Figure 6.4 and Figure 6.5 shows that there are peaks in tweet activity during periods of higher temperatures. To confirm this, daily tweet count was plotted against average daily temperature for each day between 1st May and 30th September for both Dutch and Greek tweets (Figure 6.6). There is a very clear positive relationship between daily tweet count and average daily temperature. For Dutch tweets there appears to be a non-linear relationship between tweet activity and higher temperatures, with a particular tendency for a high volume of tweets once temperatures go above around 23^{o}C. For this case study in summer 2019, the Netherlands experienced much higher than average temperatures whereas, while temperatures were high in Greece, they were in line with what is to be expected in a Greek summer. Therefore, this may suggest that users are more likely to tweet about heat and the impacts of heat if it is out of the ordinary to what is usually experienced at that time of year. This is in line with findings from previous studies (Jung & Uejio, 2017; Grasso *et al.*, 2017; Young *et al.*, 2021).

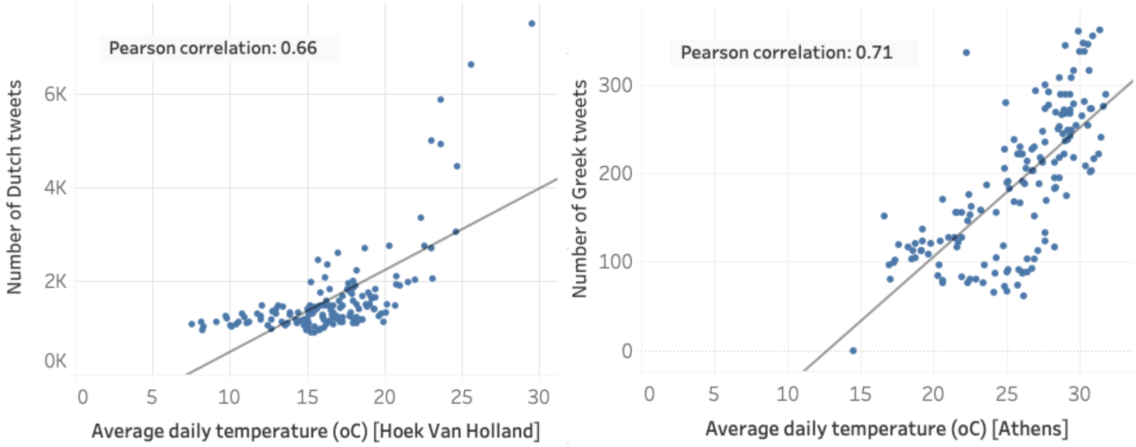

Figure 6.6 – Average daily temperature from 1st May to 30th September 2019 vs the daily tweet count for heat-related tweets located in The Netherlands and Greece respectively.

6.6.3 City-level heat-related tweets

Tweets located at the city-level scale for the cities included in this case study (London, The Hague, Athens) were explored to determine their suitability as a source of impact information in cities during heatwaves.

As tweet sentiment score has been found to be an indicator of heat impacts in previous studies (Giuffrida *et al.*, 2020), the sentiment of tweets on “heatwave days” (days which fall within an official heatwave period) for each city in the case study was plotted for each hour of the day (Figure 6.7). In London, the average hourly sentiment score clearly falls during the day and becomes most negative at night-time, during the early hours of the day on heatwave days. This is likely due to the Urban Heat Island effect which causes temperatures during the night-time to be higher than average night-time temperatures during heatwaves in large cities like London. The lower sentiment score in London in the evening/overnight could also be explained by decreased access to air conditioning, which is not typically available in homes in the UK, compared with, for example, shops and offices, where people may be during the day.

In The Hague, sentiment score was found to be lower during the day on heatwave days than at night-time. This could be because the average number of tweets during night-time hours is quite low and therefore this may not be indicative of actual sentiment experienced. Another reason for higher sentiment in the evening could be because those in The Hague may be more likely to have air conditioning in the home, compared with during the day where they may be outdoors or travelling on public transport without air conditioning, for example. In Athens, the average sentiment score is lower (negative) during the night-time in the early hours. Like London, this could be due to the UHI effect, meaning that the heat is more uncomfortable and users are more likely to post negative tweets. However, there are also a low number of tweets at this time which, as for The Hague, could mean that sentiment calculations are unreliable.

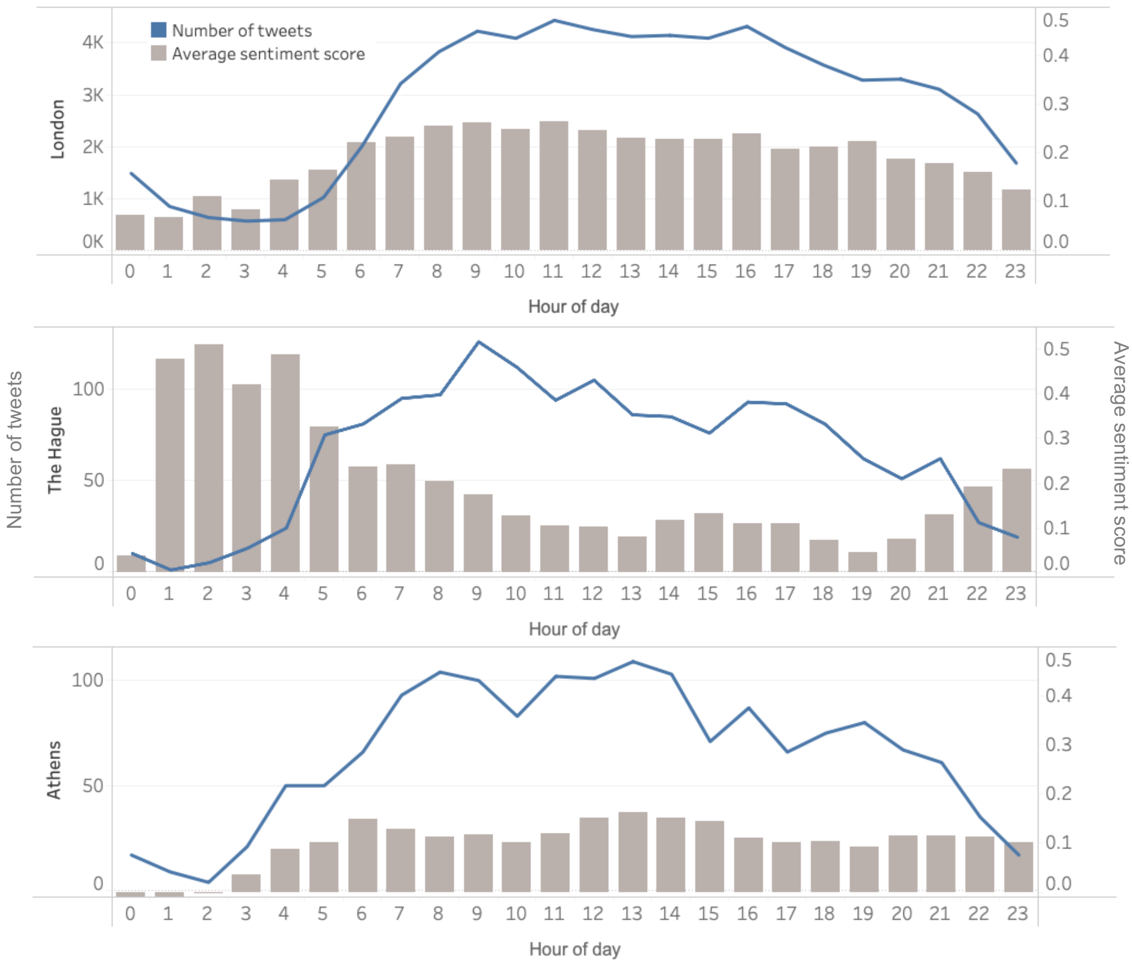

Figure 6.7 - Bar chart of average sentiment score for each hour of the day on heatwave days for each city. Line chart shows the average number of tweets per hour.

London

Further exploring heat-related tweets located in London, some of the tweets posted on heatwave days were reviewed to understand the type of information being shared and therefore what impact information may be available to city planners.

To identify the most prominent topics of interest, Figure 6.8 shows a word cloud of the most frequently used words in London heat-related tweets on heatwave days. Ignoring the most common visible terms which were part of the Twitter API search (e.g. Summer, Hot, Heat, Heatwave), the most common topics of discussion appear to be around comments on it being the hottest day (i.e. #hottestdayoftheyear was found in a large number of tweets in London), complaints about lack of sleep, disruption to rail and train services as a result of the heat, wanting to get 'cool' and mentions of air conditioning, and complaints about difficulties working in the heat.

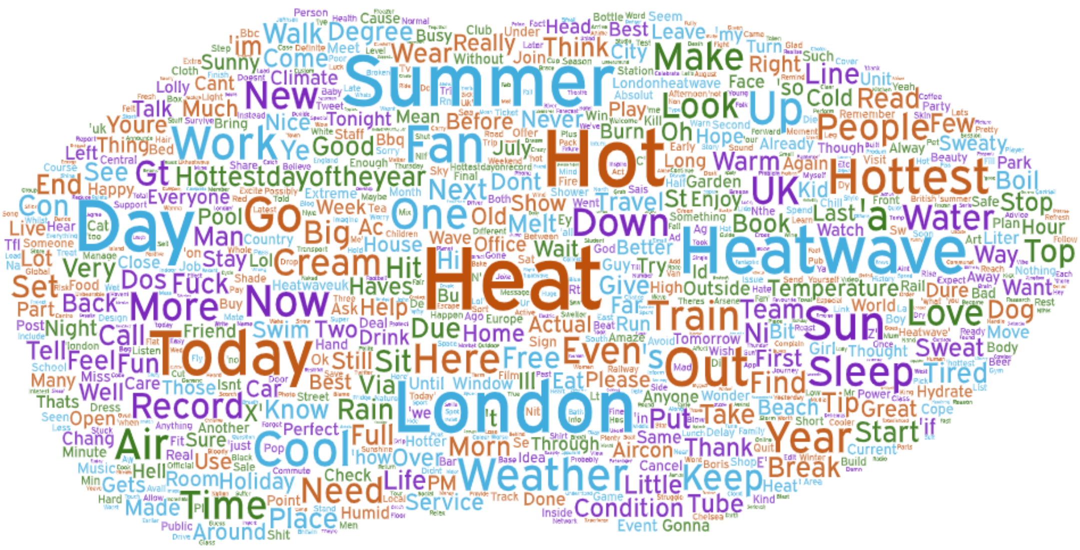

Figure 6.8 - Word cloud of most frequently used terms in heat-related tweets located in London during heatwave days in 2019. The larger the word, the more frequently it appeared in tweet text.

Some example tweets from across the range of sentiment scores are shown in Table 6.5. These are tweets which have been manually chosen to demonstrate the typical types of information being shared on Twitter in London during the summer 2019 heatwave. There is a mix of discussion about enjoying the high temperatures and complaining about them. Positive comments include references to ice-cream, air-conditioning and generally enjoying the good weather. Negative comments are more focused on travel disruption on the railways, as well as lack of sleep due to the heat. Indicating that these are the main impacts of heat that people are likely to discuss on Twitter in London, during a heatwave.

Tweet text	**Sentiment score**
Been sat in front of the fan all day and had 2 ice-creams finally starting to feel cool again xx	0.8908

Finally home time, love the A/C on the train. May try and sleep here tonight, although based on recent trips with @SW_Help there is a strong chance I wont get home anytime soon anyway. #LongDay	0.8591
Yes, it's been super hot but I am enjoying this heat wave. Proper summer I'd say	0.836
Its the hottest day but it still drizzled during a mini thunderstorm! I do love the UK weather! #hottestdayonrecord #londonheatwave	0.8118
Good luck sleeping tonight everyone!! #makeitstop !! #hottestdayoftheyear #hottestdayonrecord #hot #heatwaveuk #heatwave	0.7946
Too hot for work today! #hottestdayoftheyear #sunshine #workingfromhome	0.5411
Today's record temperatures have only strengthened my belief that people who fan themselves furiously in the heat just make themselves hotter.	0.296
The one day I want the ice cream man and he decides not to turn up	0.0772
London is not equipped to deal with this 37 degree weather. Currently melting on this train..	0
Its officially too hot to parent! #toohot #hottestdayoftheyear #parenting #summerholidays	0
I'm melting	0
Heatwave could lead to UK shortage of Christmas turkeys	-0.25
@XXXXX no aircon on in the 1st carriage of the delayed 18.24 to Littlehampton, currently stuck just outside of East Croydon. Carriage number 78726, people starting to feel unwell.	-0.5859
Commuter chaos as heatwave sparks trackside FIRES with police called in to rescue people trapped on sweltering trains #hottestdayoftheyear	-0.5859
Complete misery for London rail passengers today in the sweltering heat. As well as severe delays, there was air con not working on the train leading to further misery.	-0.765
Thank you London for making what was already a crappy and VERY hot day inconveniently worse. Every train at Waterloo delayed. #Seriously	-0.7783
Im seriously struggling to sleep. Im so uncomfortable	-0.7824
Am I in hell or just on a London bus in 40 degree heat with no air con??	-0.7998
The worst part about this Heat?? Trying to get to sleep... It isn't fun	-0.7999
Passengers soaked in sweat after being trapped on inhumane train in 40C heat. The London North Eastern Railway (LNER) train was severely delayed before British Transport Police helped the dripping-wet passengers from carriages	-0.8074
Ok. I'm done (for today) with complaining about this dirty heat I've been working 12 hour shifts in with no air con. No air con in my car. No air con in my gaff... I'll now go sweat to death in silence	-0.9217

Table 6.5 – Example tweet text of typical heat-related tweets located in London on heatwave days, with sentiment score shown.

The location of heat-related tweets within the London area was also plotted for the hottest day of the summer (25th July 2019) (Figure 6.9). Tweets during daytime hours (6am to 6pm) and evening/night-time hours (Midnight to 6am, 6pm-Midnight) are shown separately with sentiment also indicated by colour. As well as the noticeable difference in volume of tweets between day and night, the sentiment of tweets is also visibly lower during evening/night-time than during the day. This is in line with what is shown in Figure 6.7. There are a greater volume of heat-related tweets located within the central London area during the daytime. This would be expected as many people commute to work or visit the central London area during the day but live outside of the city.

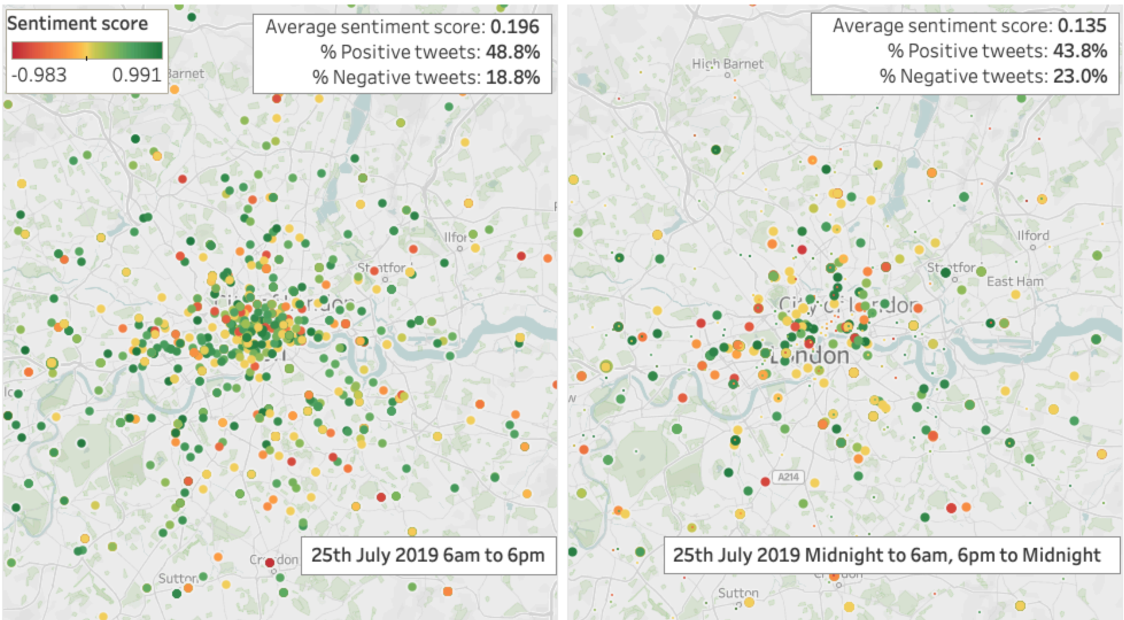

Figure 6.9 - Heat-related tweets in London on 25th July 2019 (hottest day of the year) coloured by sentiment score [LEFT] Daytime tweets from 6am to 6pm; [RIGHT] Evening/night-time tweets from Midnight to 6am and 6pm to Midnight.

The Hague

A similar analysis of heat-related tweets located within the area of The Hague was undertaken. Figure 6.10 shows a word cloud of the most frequently used terms included in tweet text (translated to English using Google Translate) of heat-related Dutch tweets located in The Hague on heatwave days. Topics of discussion focus around comments on the heat and how hot it is, observations on the temperature, wanting to be cool and air conditioning, and going to the beach (which is a very dominant topic). Despite the translation of tweet text into English, some Dutch words still remain (e.g. “hitte meaning heat, “hitegolf” meaning heatwave, “hitterecord” meaning heat record), however these are mainly used in hashtag references to the fact that it’s very hot!

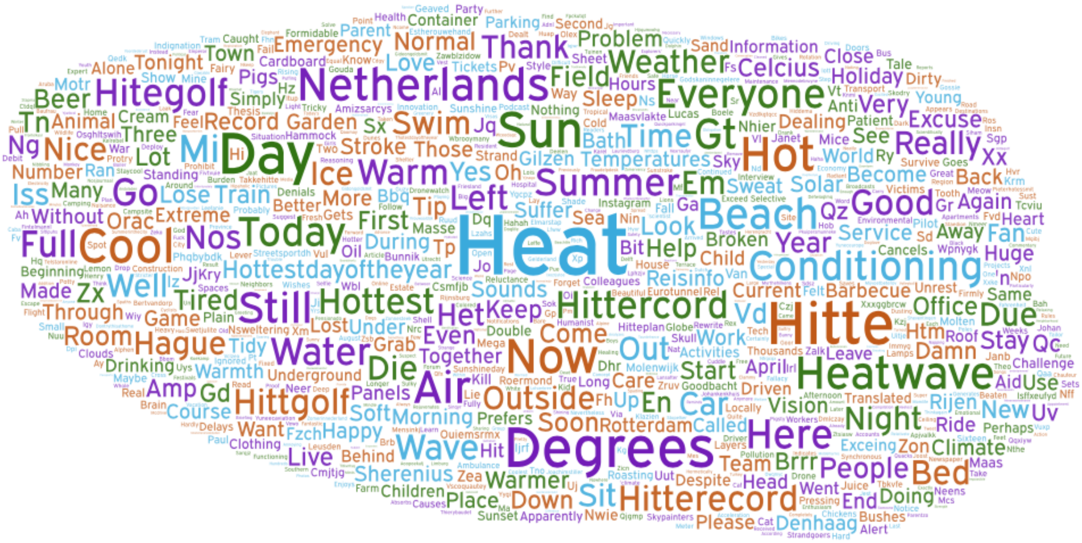

Figure 6.10 - Word cloud of most frequently used terms in heat-related Dutch tweets (translated into English using Google Translate) located in The Hague during heatwave days in 2019. The larger the word, the more frequently it appeared in tweet text.

As there are a relatively low number of Dutch heat-related tweets located in The Hague, compared to the English tweet dataset located in London, it was decided to manually tag the tweet text of heat-related tweets on heatwave days into categories, so that the most common themes could be more clearly identified than by using a word cloud. Figure 6.11 shows the volume of tweets in each category. Discussion about going to the beach and the beach area of The Hague (Scheveningen) is the most dominant topic, followed by general observations on the heat/temperature and general discussion about the nuisance of the heatwave. In terms of impact information, then the fact that there is such a high volume of discussion about the beach should aid city planners in managing the greater number of people wanting to go to this area during a heatwave. Some of this discussion also included tweets about “lost children”, “overcrowding” and difficulties driving around and parking in the beach area. All of which is helpful impact information for city officials. In terms of more general information about the impacts of the heatwave in The Hague, then there are a much smaller volume of tweets focusing specifically on disruption (e.g. shops/attractions closing as a result of the heat, travel disruption) and health impacts. Therefore, Twitter data may not be able to provide sufficient impact data on these topics.

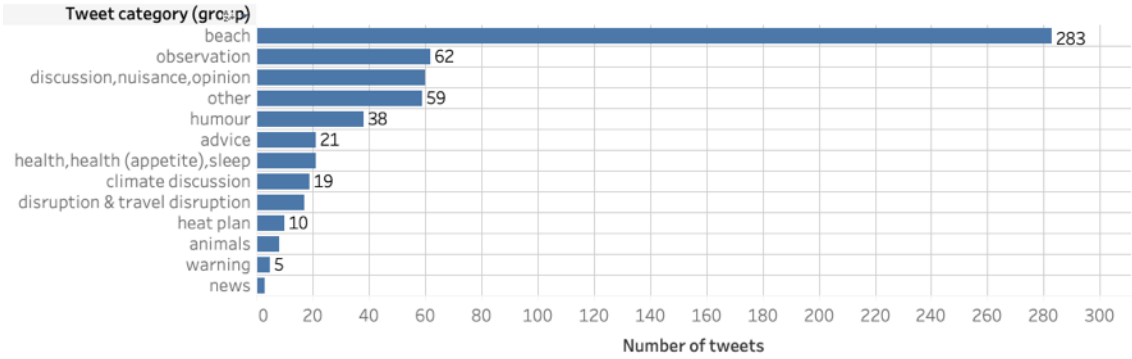

Figure 6.11 - Number of heat-related tweets by category 23rd July - 27th July 2019 and located in The Hague.

Some example tweets from across the range of sentiment scores are shown in Table 6.5. These are tweets which have been manually chosen to demonstrate the typical types of information being shared on Twitter in The Hague during the summer 2019 heatwaves. Many tweets appear to be observations of the temperature and how hot it is. There are also many tweets discussing going to, and issues at, the beach. There was also an issue in the news about some pigs dying due the heat, which raised a lot of discussion on Twitter during the July heatwave.

tweet text	**translated**	**sentiment score**
Eerst lekker gezond doen door koud citroensap te drinken en dan toch gezwicht voor een biertje. Excuus: de vakantie is einde middag begonnen! #hitte #hittegolf #gezondheid Iedereen fijne weken gewenst!	First doing well by drinking cold lemon juice and then nevertheless a beer. Excuse: The holiday started at the end of afternoon! #Hitte #Hittgolf #Health everyone wishes nice weeks!	0.8436
Welterusten allemaal! Het was een lange hete zomer dag. Hier is het tegen het einde van de avond nog steeds 34C. Wie is er nog meer gesmolten? #summer #sun #sunshine #sunshineday #hot #sand #sea #beach #zon #zee #heatwave #sky #clouds #skypainters #sunset	Good night everyone! It was a long hot summer day. Here is still 34C by the end of the evening. Who is more molten? # Summer #sun #sunshine #sunshineday #hot #sand #sea #beach #zon #zea #heatwave #sky #clouds #skypainters #sunset	0.75
@XXX @XXX Buiten nog vies heet, drukkend warm. Ging even naar buiten en bijna out... Bijna bevangen vd hitte tijdens water geven aan de tuin (hartpatient)	@XXX @XXX outside is still hot, pressing hot. Went out and almost out ... Almost lost vd heat during water to the garden (heart patient)	0.4927
Hitterecord vandaag was 40,7 graden. Maar lokaal (zoals in Roermond) is het ng warmer geworden! #hottestdayoftheyear #hitte	Hitterecord Today was 40.7 degrees. But locally (as in Roermond) the ng has become warmer! #Hottestdayoftheyear #Hitte	0.4753
Nog even afkoelen. #StilleStrand #DenHaag	Just cool down. # still strand # denhaag	0.3182

In de trein is het gevoelsmatig 45 graden. Toch bijzonder dat de treinstellen hermetisch zijn afgesloten voor frisse lucht #fail #hittegolf #hitterecord @XXX	In the train it is emotional 45 degrees. Still special that the train sets are hermetically closed to fresh air #Fail #Hittgolf #Hitterecord @XXX	0.2732
Extreme #heatwave #hittegolf here in the Netherlands. Current temperatures exceed 42 degrees Celcius, i.e.107 degrees Fahrenheit. Its hotter here than in most tropical Summer destinations.	Extreme #heatwave #HiteGolf here in the Netherlands. Current Temperatures Exceed 42 Degrees Celcius, I.E.107 Degrees Fahrenheit. It's Hotter Here Than in Most Tropical Summer Destinations.	0
Blijf drinken in deze hitte #hottestdayoftheyear #40graden #hitterecord #hitte #hittegolf	Keep drinking in this heat #HottestDayofTheyear # 40spading #HitterCord #Hitte #Hittegolf	0
Dacht haal even een ijsje op Scheveningen... #hitte #hittegolf	Thought a little ice cream at Scheveningen ... # heat #HiteGolf	0
Video: Heet weer betekent ook drukte op het strand. Michiel en zijn collegas werken op en rond het strand in Scheveningen en Kijkduin.	Video: Hot weather also means crowds on the beach. Michiel and his colleagues work on and around the beach in Scheveningen and Kijkduin.	-0.3089
Honderden varkens sterven van de hitte in varkensstal Middelharnis	Hundreds of pigs die from the heat in pigsty Middelharnis	-0.5994
@HTM_Reisinfo De bestuurder geeft aan dat de airco stuk is gegaan door de warmte. Dat klopt niet. Als je de meldingen doorneemt zie je dat vanaf april er gemeld wordt dat de airco van deze wagen het niet doet.	@Htm_reisinfo The driver indicates that the air conditioning has been broken by the heat. That is not true. If you read the notifications, you will see that from April that the air conditioning of this car does not do it.	-0.6632
@XXX Dit probleem speelt elke zomerse dag; uitpuilend OV, overvolle straten met geluids- en parkeeroverlast, uitverkochte supermarkten met rijen Duitsers, etc. maken Scheveningen in de zomer verre van een pretje voor bewoners.	@XXX This problem plays every summer day; Outpuffal OV, crowded streets with sound and parking nuisance, sold-out supermarkets with rows of Germans, etc. Make Scheveningen in the summer far from a fun for residents.	-0.7717

Table 6.6 - Example tweet text and English translation (using Google Translate) of typical heat-related tweets located in The Hague on heatwave days, with sentiment score shown.

Figure 6.12 shows the location of heat-related tweets located within The Hague area on the hottest day of the summer (25th July 2019). The map of tweets for daytime (6am-6pm) and evening/night-time (Midnight – 6am, 6pm-Midnight) is shown separately. There are tweets spread throughout the city, however there does appear to be a clustering of activity around the beach area of The Hague on the coast (Scheveningen beach). Interestingly, there are a greater proportion of heat-related tweets in the beach area in the evening/night-time plot, compared with during the day. This could suggest that people in the city move towards the coast to cool down/enjoy the hot weather in the evenings, which may be useful impact data for city officials – particularly as there are a number of heat-related tweets located in The Hague complaining about congestion and crowding at the beach (Figure 6.11). As Twitter posts about being at the beach tend to be quite positive in terms of sentiment score, it may also explain why sentiment of tweets in the evening/night-time is quite high (Figure 6.7). On visual inspection it also appears that tweets are more negative inland during the day, compared to the beach area.

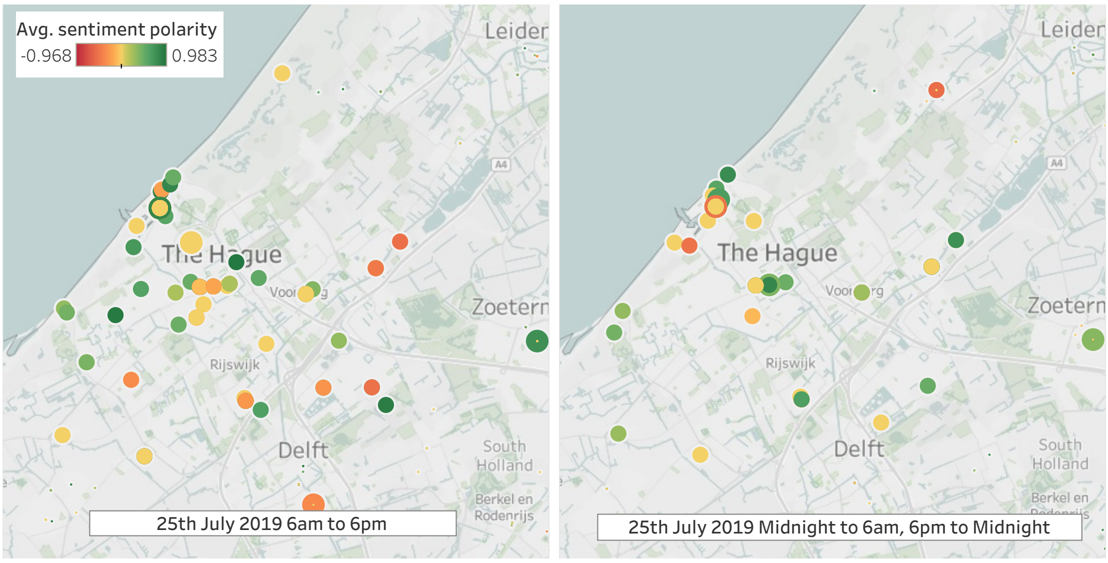

Figure 6.12 - Heat-related tweets in The Hague on 25th July 2019 (hottest day of the year) coloured by sentiment score [LEFT] Daytime tweets from 6am to 6pm; [RIGHT] Evening/night-time tweets from Midnight to 6am and 6pm to Midnight.

Athens

In Athens, there was a lower volume of tweet activity overall compared with London and The Hague. As Athens is the second largest city of the three cities in the case study, more Twitter discussion about the heat was expected. The low volume of tweets could be due to Twitter not being very popular as a social media platform in the country of Greece (Facebook is the predominant social media platform in use in Greece (Statista, 2021)). Another reason could be that high temperatures and the impacts of heat are more of a common occurrence in Greece, and in particular Athens. Air temperatures often rise above 37°C in the city centre in the summer months due to the topography of the local area around Athens and the Urban Heat Island effect (Keramitsoglou *et al.*, 2013). The temperatures experienced during the period of this study in Athens (30°C to 38°C) were not above normal expectation for the time of year and therefore may not have generated as much discussion as in the UK and Netherlands, where these temperatures are not a common occurrence. Normalisation Bias, in which people become accustomed to natural hazard events or assumed that they fared adequately in a previous event (Johnston *et al.*, 1999; Becker *et al.*, 2017), may therefore be a factor affecting the social response to high temperatures in Athens. In any case, similar analyses of heat-related tweets located in Athens were undertaken as for London and The Hague.

Figure 6.13 shows a word cloud of the most frequently used terms included in tweet text (translated to English using Google Translate) of heat-related Greek tweets located in Athens on heatwave days. Topics of discussion focus around comments on the heat and how hot it is, observations on the temperature, the beach and swimming. There are no obvious heat-related impact topics from examining the most frequent terms contained within translated tweet text.

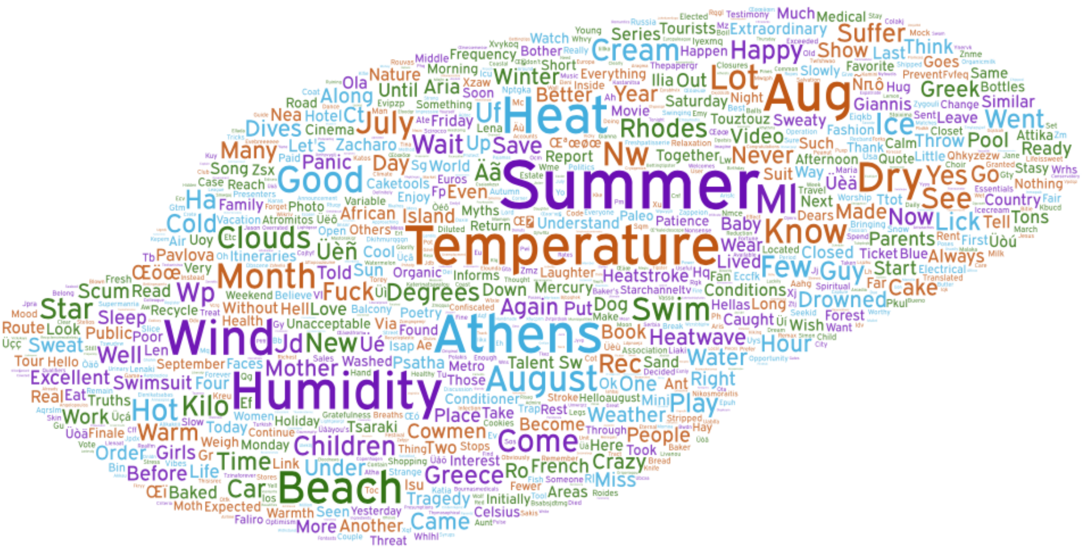

Figure 6.13 - Word cloud of most frequently used terms in heat-related Greek tweets (translated into English using Google Translate) located in Athens during heatwave days in 2019. The larger the word, the more frequently it appeared in tweet text.

On manual inspection of the heat-related Greek dataset, there were more irrelevant tweets included. This will most likely be as a result of the poorer performance of the machine learning relevance filter, which was discussed in the Methods section. However, Table 6.7 shows some example tweet text across the range of sentiment scores of typical heat-related Greek tweets located in Athens posted on heatwave days. These tweets represent the type of discussion that is most prevalent in relevant heat-related tweets. Many tweets focus on sharing advice about coping in the heat. There are also a number of tweets sharing observational information about the temperature and general discussion about the heat. There were no obvious reports about particular impacts (e.g. travel disruption, health issues) as a result of the heat.

tweet text	translated	sentiment score
Καλό μήνα αγαπημένες και αγαπημένοι μου!!! Μετά από ένα καλοκαίρι που, μέχρι τώρα, μας ταλαιπώρησε αρκετά, έρχεται ο Αύγουστος φέρνοντας ανάσες ηρεμίας, αισιοδοξίας, χαλάρωσης και έρωτα. Λίγη υπομονή και από την…	Happy month my dears and loved ones !!! After a summer that, so far, has bothered us a lot, comes August bringing breaths of calm, optimism, relaxation and love. A little patience from…	0.9726
4 τρόποι για να δροσιστείς μέσα στη ζέστη χωρίς να χαλάσεις μια περιουσία στο ηλεκτρικό ρεύμα	4 ways to cool off in the heat without damaging a fortune in electricity	0.466
Καύσωνας: Ο Ιατρικός Σύλλογος Αθηνών συνιστά προληπτικά μέτρα	Heatwave: The Medical Association of Athens recommends preventive measures	0.2263
Πρώτη μέρα του Αυγούστου Το λοιπόν ξεκινάμε σιγά σιγά την αντίστροφη μέτρηση γι' αυτές τις πιο δροσερές ημέρες... Τι εννοείς ότι κ τον Σεπτέμβρη έχει ζέστη δεν καταλαβαίνω	First day of August So we are slowly starting the countdown for these cooler days ... What do you mean that September is hot I do not understand	0
Καύσωνας με το θερμόμετρο ως 42 βαθμούς το επόμενο διήμερο	Heat with the thermometer up to 42 degrees the next two days	0
Καύσωνας: Πως θα καταλάβετε ότι κινδυνεύετε από τη ζέστη	Heat: How to understand that you are in danger from the heat	-0.2732
Θερμοπληξία: Ακόμα μια απειλή για τα παιδιά στο αυτοκίνητο	Heat Stroke: Another Threat to Children in the Car -	-0.5267
Σήμερα το απόγευμα στη παραλία η Ελληνίδα μάνα ξέπλενε το σκατομένο μαγιό του γιού της στα ρηχά...κατά τα άλλα σας ενοχλούν τα σκυλιά😠😠😠	This afternoon on the beach, the Greek mother washed her son's dirty swimsuit in the shallows ... otherwise the dogs bother you 😠😠😠	-0.6486
Θερμοπληξία: Ακόμα μια απειλή για τα παιδιά μέσα στο αυτοκίνητο: Ο θάνατος από θερμοπληξία…	Heatstroke: Another threat to children in the car: Death from heatstroke…	-0.8074

Table 6.7 - Example tweet text and English translation (using Google Translate) of typical heat-related tweets located in Athens on heatwave days, with sentiment score shown.

Figure 6.14 shows a map of heat-related tweets located in Athens on one of the hottest days of the summer (1st August 2019) during daytime hours (6am-6pm) and evening/night-time hours (Midnight-6am, 6pm-Midnight). During the day, the most tweets appear to originate from the west of the city centre, which is one of the main tourist areas. Tweets seem to have generally quite neutral or positive sentiment, with less tweets in the main city centre area in the evening/night-time, compared with during the day. Tweets appear to be more concentrated in the centre of the city, rather than in the wider area around Athens.

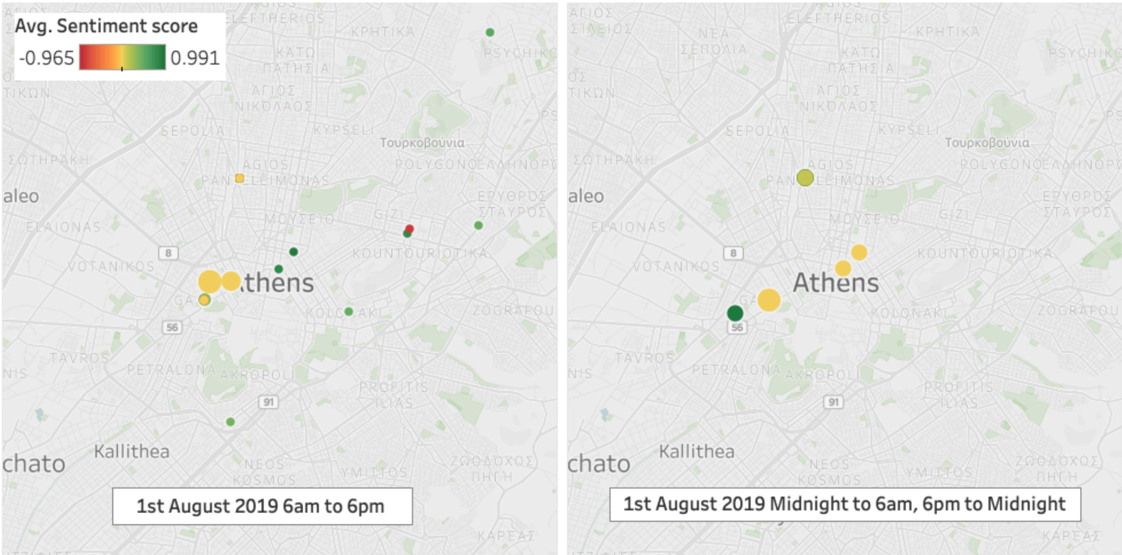

Figure 6.14 - Heat-related tweets in Athens on 1st August 2019 (hottest day of the year) coloured by sentiment score [LEFT] Daytime tweets from 6am to 6pm; [RIGHT] Evening/night-time tweets from Midnight to 6am and 6pm to Midnight.

6.7 Discussion

Social sensing of heatwaves has been proven to be successful in previous studies (Young *et al.*, 2021) and there was good evidence that information about the impacts of heat during heatwaves would found by examining social media (Twitter) data. Therefore, this study focused on the social sensing of heatwaves at the city scale in three different European cities (London, The Hague and Athens) during the summer of 2019, in which high temperatures and heatwaves were experienced. This study also included the analysis of tweets in different languages, which is an extension of previous studies which have focused on one language (mainly English).

The impacts of heatwaves found included an increase in tweet activity relating to heat, and the impacts of heat, as temperature increased in both The Hague and Athens (due to tweet download rate limits in the Twitter API, this was not checked for London due to the volume of English tweets returned in the API query search, therefore the collection was focused on heatwave days only). In particular, peaks in Twitter activity were observed during official heatwave periods. Tweet activity in particular areas of a city increased during heatwaves (e.g. more tweets found in the beach area of The Hague and tourist areas of Athens).

However, this study found that for smaller cities such as Athens and The Hague, social sensing of heatwaves at the city scale is not as clear cut as social sensing at a larger geographical scale. This could suggest that the performance of social sensing varies at different tweet volumes and/or geographical scales. Likewise, the poor performance of social sensing for finding heat-related impact information during heatwaves in Athens could have been affected by the fact that Twitter is not as widely used in Greece as it is in the UK and the Netherlands. Therefore, the propensity for using particular social media platforms may also be a factor in the performance of social sensing for collecting impact information.

Additionally, there may also be the factor of Normalisation Bias (Johnston *et al.*, 1999; Becker *et al.*, 2017) affecting the amount of heat-related tweets for the city of Athens during the period of high temperatures in this study. While temperatures were high in Athens compared with London and The Hague in the summer of 2019, they were not significantly above seasonal norms for the city. Therefore the lower volume of tweet activity seen may support previous studies' findings that Normalisation Bias has an impact on people's perceived level of risk related to the heat (Mishra *et al.*, 2009; Frondel *et al.*, 2017; Barrett, 2022) and therefore propensity to tweet about the high temperatures.

Overall, there were some impacts identified during heatwaves in the cities studied. In London, discussion about disruption to rail travel as a result of the heat was quite a prevalent impact topic. In The Hague, there was a clear signal for increased activity in the beach area of the city (Scheveningen). There was less in the way of impact information gathered for the city of Athens.

Therefore, the results of this study are mixed. The location of heat-related tweets within a city area could aid city planners in terms of where cooling facilities would be best placed within the city for use on days with a high temperature, for example. This could also assist with planning for where there are likely to be greater numbers of people on these days. In terms of using the tweet content for gaining information about the specific impacts experienced by people in the city, then in London, this would likely be a useful source of information. However, in

Athens and The Hague, the specific heat-related impacts experienced by people posting on Twitter was less obvious.

The advantages to using Twitter as a source of impact information during heatwaves is most obviously the timeliness of data being available, almost in real-time, as users post about their experiences of the heat during the day. This would give city officials a head-start on assessing the impacts of the heat before more official, or news media sources, provide information. The disadvantages are that determining the specific impacts is not as obvious, particularly in cities with a smaller volume of Twitter activity.

Revisiting the original aims of this study, to explore the feasibility of developing a tool to monitor heatwave impacts in real-time using social media (Twitter data), then findings are mixed. There is definitely scope to extract relevant information from Twitter relating to heatwave impacts, locate it within a city context and provide this for city resilience planners. For large population cities with a large volume of Twitter users, such as London, then there may be benefits to city planners managing heatwave response if a tool which provided heatwave-related tweets in real-time were available. However, for smaller population cities, such as The Hague, or cities with a lower volume of Twitter users, such as Athens, then there may not be a large enough volume of useful impact-related heatwave tweets for this kind of tool to be useful.

6.7.1 Limitations and further study

There are some limitations to this study which may well have affected results presented. The machine learning classifier did not perform as well for Dutch and Greek tweets. As discussed in the methods section, this is likely due to training the classifier using translated tweets. Therefore, having a Dutch and Greek speaker involved in the classification of tweets in these languages may be beneficial. Further work to improve the classifier for the Dutch and Greek language should yield better results in terms of the proportion of heat-related tweets in the final dataset. Also, issues with words in the Greek, rather than the Latin, alphabet could also have affected how well the relevance filters and the location inference performed. This could therefore have led to a number of Greek tweets not being available in the final dataset, and could be another reason why social sensing did not perform as well for Athens.

Additionally, the English dataset was only explored for specific days during an official heatwave due to size of the dataset. However, it may be valuable to examine heat-related English language tweets for a larger time period so that differences between tweet activity on “heatwave days” vs “non-heatwave days” could be explored.

6.8 Summary and conclusion

Using methods from previous work on the social sensing of heatwaves, Twitter activity in three European cities (London, The Hague and Athens) during the summer of 2019 was explored. The aim of the study was to determine if information on the impacts of heat in cities could be curated from social media (Twitter) data. Findings from the study were mixed, with a good signal for heat-related discussion in London and some information about specific impacts experienced found. In The Hague, the volume of tweet activity was much lower, however there was some useful information about where in the city people were likely to go during a heatwave. In Athens, tweet activity was quite low and impact information was less obvious. Therefore, social sensing of heatwave impacts in cities may work well for larger cities, with a good volume of Twitter data. However, it may not work as well for smaller cities, or cities with less Twitter users.

Chapter 7 - Discussion

Throughout the previous chapters, methods and tools for accessing weather-related impact information from social media and its potential to improve situational awareness during weather events have been explored. The following research questions were defined, based on a review of the motivation for the need for weather impact information and existing academic literature on the social sensing of weather events:

- *RQ1. How useful is social media as a source of impact information during and after weather events?*

- *RQ2. What tools and methods can be successfully applied to extract relevant social media data during weather events?*

- *RQ3. What are the limitations of social media as an impact data source?*

These questions have been explored through the application of social sensing methods to extract and analyse social media data for different weather events in Chapters 4-6 of this thesis. Each of these chapters will now be summarised in the context of these research questions:

7.1 Using social media to measure impacts of named storm events in the UK and Ireland

In Chapter 4, Twitter data relating to eight named storm events, precipitation and wind from October 2017 to March 2018 was extracted, filtered for relevance to the stormy weather and located within the UK/Ireland. The aim of the study was to both explore the use of Twitter data to detect named storm events and to further interrogate the filtered content for information relating to impacts of the storms. Previously tested methods to extract, filter and locate Twitter data in the context of floods (Arthur et al, 2018) were applied to named storm events in the UK and Ireland during the 2017/18 storm season. The text of relevant tweets was then analysed to examine how the sentiment and the topics of discussion change during the storm event. By examining the change in topics of discussion during Storm Brian it was found that in the day leading up to the storm, warnings and alerts were the predominant topic of discussion. As Storm Brian hit the UK and

Ireland, topics of discussion moved towards observation of the weather conditions and reports of damage and disruption. The day after the storm event found news articles and reports of damage/disruption continuing to be shared. The proportion of tweets containing humorous content remained constant throughout the storm period. In terms of sentiment, this was calculated for tweets relating to precipitation, wind and named storms, with the latter (and more severe weather event) having the most negative sentiment.

As the content of tweets during the most active period of the storm included observations and reports of damage and disruption, Twitter is shown to be a good real-time source of impact information. This could therefore provide NMHSs with timely impact information for situational awareness. The social sensing methods used also had good accuracy in terms of the proportion of relevant content extracted from the overall Twitter dataset. Referring back to research questions RQ1 and RQ2, this study therefore provided justification for the usefulness of social media as an impact information source, and suitable methods for filtering content for relevance to the weather hazard.

7.2 Social sensing of high-impact rainfall events worldwide: A benchmark comparison against manually curated impact observations

In Chapter 5, Twitter data relating to heavy rainfall (rain, rainfall, rainstorm) and the impacts of heavy rainfall (floods, landslides) were explored for its use as a global impact detection tool. Tweets containing these rainfall/rainfall impact related terms were collected from 1 January 2017 to 30 June 2017. Methods to filter tweets for relevance and infer location in the UK/Ireland in previous studies were extended and applied to this dataset in a global context. Good global coverage of relevant tweet activity relating to rainfall events was found, despite examining tweets in the English language only. Tweet activity was measured using a percentile measure, as opposed to actual counts of tweets, to normalise and account for the propensity for tweets and/or population in a particular location.

Different percentile tweet activity thresholds were then tested as an indicator of an event detection on Twitter and compared with the Met Office Community Impacts Database. This database contains information about the impacts of rainfall events across the world and was manually curated using a range of online news and media sources. Comparison of events detected on Twitter to the validated events in the Community Impacts database found over 90% accuracy. Good accuracy was found for countries with a higher propensity for posting content on Twitter (e.g. USA, Australia), as well as for countries with lower English-language tweet volumes (e.g. Malaysia, Saudi Arabia). Additionally, several events which caused significant impacts, but were not included in the impacts database, were also detected using Twitter data. Several reasons were suggested for why additional events were found using social media but not using the other online media sources used to curate the Met Office Community Impacts database. In particular, the type of record that Twitter presents (i.e. eye-witness accounts, individual reports of events taking place) is different to aggregated sources of information, such as news articles. Also, smaller events which resulted in less wide-spread impacts may be more likely to be detected using social media, rather than other online media sources, as they may be less ‘newsworthy’ and therefore less worthy of report. Therefore, this study showed that social media is a useful ‘first-pass’ event detection tool that can aid with impact information curation and assessment.

Another interesting finding from this study was that heavy rainfall events in countries where English is not a first language can also be detected using English language tweets. An obvious extension of this study would be to explore Twitter content in other languages to increase the volume of social media content in non-English language speaking countries. However, even with this limitation in language, good accuracy in event detection was found.

Revisiting research questions RQ1 and RQ2, this study provides an example of the usefulness of social media as an impact assessment tool and further tested social sensing methods for multiple events and locations, finding good accuracy. Considering research question RQ3, this study also suggests that while good global coverage and accuracy was found by using English language social media posts only, extending the dataset to include content in multiple languages would likely be advantageous in increasing information available from social media at a global scale.

7.3 Social sensing of heatwaves in European cities

The aim of this study was to explore the use of social media (Twitter data) related to heatwaves in three European cities as an impact assessment tool for use by city officials during heatwave events. Many previous studies by other researchers have focused on flooding or hurricanes, therefore this was an experiment to explore the use of existing social sensing methods for a lesser studied weather hazard. This study built on an earlier study (Young *et al.*, 2021), which explored the application of social sensing methods during heatwave events in the UK, USA and Australia, during their respective summer periods in 2018. In this study, it was found that social media provided a good indicator of a heatwave event, with content and sentiment analysis suggesting that the impact of heatwaves could be determined from social media. Chapter 6 therefore described an extension of this earlier study and explores Twitter data containing a wider range of heat and heat impact related terms collected during the summer 2019 period, in which Europe experienced some of its highest recorded temperatures. As each of the three cities in the study were in different European countries, Twitter data was also collected in three different languages (English, Dutch and Greek) to extend the volume of tweets in the dataset beyond those in the English language only. Additionally, the focus on tweet content at the city, rather than a larger country or administrative area scale, further tested the social sensing methodology to determine its usefulness at a smaller geographical scale.

Revisiting research question RQ2, it was found that social sensing methods perform well at detecting heatwave-related tweet content during heatwaves, and a positive correlation between tweet activity and average daily temperature was found. In particular, during periods of highest temperatures significant tweet volume was found. However, at the city scale, the volume of tweet activity was too low to provide a reliable measure of impacts. Examining the content of relevant tweets did provide some information about the behaviour of individuals and impacts during heatwaves (e.g. travel disruption in London, people visiting the beach in The Hague, and increased tweet activity in tourist areas of Athens). However, the usefulness of this information for situational awareness and for impact assessment is uncertain.

7.4 Critical reflection

The result of this thesis is an improved understanding of the methods required to interrogate social media data and its potential as a source of impact information relating to weather hazard events. Building on previous work, social sensing methods to detect impactful weather events on social media were applied for different weather hazard events and types. This provided evidence that social sensing is a robust approach with which to filter social media for relevance and infer location, and which can be used to aid in the curation of impact information during weather events.

Reflecting on some of the findings from this thesis, there are a number of observations for both the potential and limitations of using social media as an impact information source.

In terms of potential, with further refinement of training classifiers, filtering social media for relevance using supervised machine learning methods have been shown to work well, with good accuracy. Additionally, locating social media posts using discernible place names within the text and/or user location of the post greatly improves the volume of spatial impact information available. Location inference has its limitations for dealing with locations with the same name in multiple locations (e.g. London, UK and London, Canada). However, continued refinement of these methods will improve this approach. For impact assessment, temporal, spatial and content analysis of relevant posts can provide a good indicator of the type and severity of impacts during weather events.

Social sensing has been shown to be a good method for extracting relevant and timely impact information from social media. However, there are no pre-defined quantitative metric thresholds (e.g. specific volume of social media posts) which can be used as a proxy for an impactful event to be identified. For an event to be detected on social media, it depends on a number of factors: the propensity for use of a specific social media platform in a particular place; whether or not the weather event being experienced is extreme for that particular place or not; and the geographical scale.

Salience and attention to a weather event is also likely to affect concern and motivation for a person to post on social media about it. This could be affected by whether or not the weather event being experienced was typical or not for the location. As seen in Chapter 6, Normalisation Bias was likely to have affected the amount of Twitter activity for the city of Athens during high temperatures as these were not out of the ordinary for the time of year. However, for similar temperatures experienced in London and The Hague, where these temperatures were much more uncommon, much more heat-related discussion on Twitter was identified. What is considered as extreme or hazardous in one place, may not be in another. A number of days' rain during the winter in a location such as the United Kingdom may not necessarily be out of the ordinary. However, in a place with a drier climate in the summer months, a period of rain like this may be more of an extreme event and therefore generate more discussion on social media. Therefore, the importance of context, local conditions and seasonality when comparing discussion on social media in different places in the world needs to be considered. It may also be necessary to redefine the definition of impact for different locations as the impacts from extreme weather in one place may not be thought of as extreme in another.

Additionally, a weather event which gets a lot of media attention (for example, named storms in the UK, as discussed in Chapter 4) may skew peoples' sensitivity to the event and encourage more people to post on social media about it. Studies have examined the influence of news media sources on public attention to an event (García-Perdomo *et al.*, 2017; Araujo & van der Meer, 2018). Therefore, a weather event with a lot of news coverage is likely to feature more heavily in social media feeds, encouraging more attention from social media users. As found by Abdullah *et al.* (2017), this, in turn, could generate more social media discussion and retweeting about the event than a weather event with less media coverage. Likewise, if another significant event is occurring at the same time as the weather event (e.g. national election, Covid-19, celebrity news) this may detract attention away from the impacts of the weather being experienced.

It is also important to understand any potential biases which may affect what impacts are reported on social media during a weather event. The demographics of who is reporting the impacts may well affect the type of impacts being reported on social media. Research has shown that participants on social media are often from Western, Educated, Industrialised, Rich and Democratic (WEIRD) societies (Azar, 2010). People from these societies represent only 12% of the world's population, however make up the majority of social media users worldwide.

Therefore, access to social media due to socio-economic status could skew they type of impacts being reported during weather events and this limitation needs to be recognised when using social media as a source of impact information.

As mitigation and adaptation measures are taken as a result of significant weather events, visible and tangible impacts should, in theory, decrease. This may therefore have some effect on the impacts that might be determined from social media in locations where such measures are implemented in the future. Likewise, if one significant weather event follows immediately after another, then the impacts of the second event may be skewed accordingly and may appear lower impact in places where damage has already been experienced. Both of these additional factors on the impacts that can be determined from social media during weather events should also be taken into account.

Considering the effect of geographical scale on the usefulness of information on social media, in this thesis, social sensing was shown to work well at the country/state/county level in Chapters 4 and 5, but was found to be less useful at a more granular scale (e.g. city) in Chapter 6. Furthermore, the level of impact information available from social media during weather events also appears to vary by weather hazard type (i.e. flooding is well discussed on social media, heatwaves less so). Therefore, the effectiveness of social sensing and applications developed using its principles need to take these factors into consideration. Social sensing will never be able to provide a ‘perfect’ record of impacts and its effectiveness does depend on the context of the particular hazard, severity of the impacts experienced and how much the events are ‘out of the ordinary’.

The types of impact information which can be determined from social media was found to range from specific information about who and what was impacted (e.g. travel disruption, damage to buildings, etc) to more subjective indicators of impact, such as sentiment analysis. For specific impact information, it is important to consider the challenges of defining what is meant by impact as a result of weather events if social sensing is to be operationalised for use as a situational awareness tool. As discussed in Chapter 1, impacts as a result of the weather are generally considered to be socio-economic impacts such as disruption to transport and travel, damage to homes and businesses, people displaced/evacuated, injuries, hospitalisations and loss of life. Some of these impacts are more easily detected from social media posts than others. For example, a blocked road, delayed train or damaged roof may be more likely to posted by an individual on social media. However, more widespread disruption such as hospitalisations, people displaced, etc are less discernible from social media posts and this information may need to be obtained via other sources.

Sentiment analysis of the text of tweets during named storm events in the UK was explored in Chapter 4 as another potential indicator of impact. Sentiment was also calculated for heat-related tweets during heatwaves in Chapter 6. While it was found that sentiment of tweets was more negative overall during these severe weather events, it is important to recognise that not all text in social media posts may be expressing emotion. For example, “Storm Brian causes floodlight damage” was assigned a negative sentiment score of -0.4, likely due to the use of the word “damage”, which is a negative word. However, this tweet was not necessarily expressing emotion, just providing information. Additionally, a tweet designed to be humorous: “Brian? What kind of name is that for a storm? Everyone knows Brian is a snail” was assigned a positive sentiment score of +0.6. In the context of determining impact, the usefulness of sentiment analysis with social media posts about the weather needs to be considered. As discussed in Appendix C, it is also important to note the limitations of the use of the python package TextBlob to calculate sentiment in Chapter 4, which was trained on movie reviews and may therefore not have been the most appropriate package to use to calculate the sentiment of weather tweets. However, in Chapter 6 the Vader sentiment tool, which is specifically tuned to use with short text strings in social media posts, was used to calculate sentiment which provided the sentiment of language used, but not necessarily in the context of weather. Therefore, the question of the usefulness of sentiment in the context of weather tweets and what it is actually measuring still needs to be considered. It may therefore be appropriate to explore the development of a sentiment analysis tool specifically tuned to the context of the impacts of weather events to improve the usefulness of this measure.

In Chapter 5, time lags of a few days were observed for some events between a rainfall event occurring and the peak in Twitter discussion about it as impacts of the event are realised. In some geographic locations for example, riverine flood impacts may occur days after the upstream rainfall event. This was observed in flooding events in Queensland, Australia (Kankanamge *et al.*, 2020). Therefore, for the social sensing method to be applied as an impact detection tool, these potential time lags between phenomenon occurring (e.g. rainfall event), the impacts being realised (e.g. riverline flooding leading to bridge collapse) and people posting on social media about it, need to be taken into consideration.

Social sensing certainly has potential to be used to improve situational awareness, warnings and communication by NMHSs and other organisations during significant weather events, and prevent additional impacts. As discussed in Chapter 1, one of the main challenges of organisations accessing information about impacts during severe weather events is the lack of an operational tool and robust source of reliable impact information (Vieweg *et al.*, 2014). The use of social media as a 'first pass' source of impact information during weather events has been widely considered to be something that organisations would find useful as individuals can report what is happening in real-time before other official sources of information, such as news media, become available (Harrison *et al.*, 2021). However, the collation of information from social media has been found to be manual, complex and time consuming (Hemingway & Robbins, 2020) and difficult with no systematic processes in place (Kaltenberger *et al.*, 2020).

Social sensing has been shown in this thesis to be a process which could provide filtered relevant information to highlight the socio-economic impacts during weather events from social media. This could be in the form of a feed of filtered social media posts related to the impacts of a weather event (i.e. with irrelevant posts removed) and specific to a particular location. This would greatly reduce the manual effort and time spent searching social media pages by those in organisations interested in situational awareness and managing the communication of warnings during severe weather events. The location inference aspect of social sensing also means that social sensing could provide heat maps of where there are peaks in social media discussion about a weather event and therefore provide a quick overview of where the most significant impacts of a weather event are being realised. This would help to target relief efforts and messaging by agencies during the event. Social sensing outputs could also be used to help with the retrospective evaluation of impact-based forecasts, both using the volume of social media activity and the types of impacts determined from social media posts.

Development of a suitable tool to access the outputs of social sensing would therefore greatly aid organisations improve their situational awareness, etc during weather events. Tools developed in previous studies for flood risk management (Fitrianah *et al.*, 2020) and situational awareness during hurricanes (Khaleq & Ra, 2019) are examples of how social sensing could be operationalised. It will be important to work with end users, risk communicators and agency message providers to ensure that any tool developed provides information in a suitable format.

Academic research into the use of social media as an information source for both event detection and impact assessment during weather events is beginning to be applied in an operational context. During the course of this PhD study, the researcher was directly involved in the formation of Social Sensing Ltd, a small UK company, which builds on the research detailed in this thesis. This has resulted in the development of an online application which operational meteorologists and flood forecasters can use to access relevant social media (Twitter) data relating to floods, wind/storms and snow in real-time (Figure 7.1). The Social Sensing tool is now in operational use in the Met Office, Environment Agency and Natural Resources Wales and is an excellent example of the application of research by practitioners. The potential of social media as a tool for situational awareness and information source for impact assessment is therefore being realised.

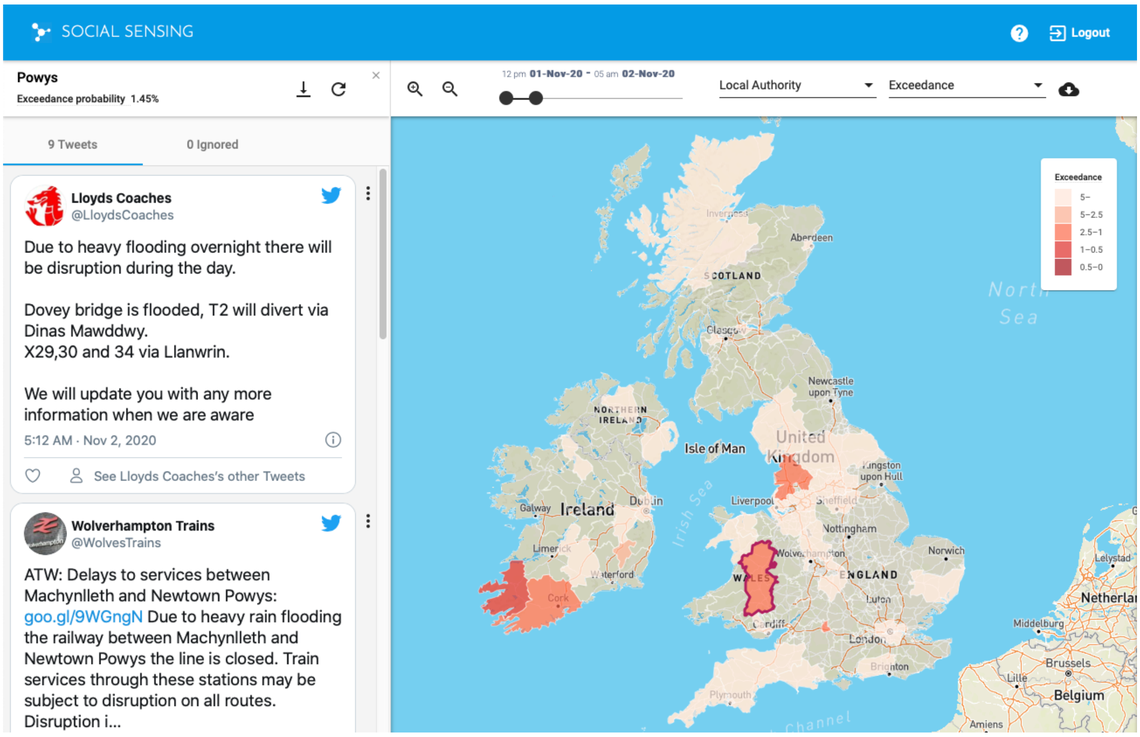

Figure 7.1 - example of operationalised social sensing output during a flood event in the UK from Social Sensing Ltd.

7.5 Future work

There are a number of potential areas for future work identified from the research detailed in this thesis. As the motivation for using social media is to gather impact information, extending the data collection keywords from weather hazard-related terms to also include weather impact keywords might offer greater volumes of data. For example, the text of a social media post reporting impacts of wind (such as trees falling down, road closures or damage to buildings) may not use weather-related keywords (such as 'wind', 'storm', etc) in the text of the social media post but instead only make mention of the resulting impact. This information is directly related to the impact of the weather event, however the lack of mention of the specific weather hazard in the text of the social media post precludes its inclusion in the social media dataset extracted from the social media platform based on weather-related keywords only. Therefore, expanding the social media search to include impact terms as well as weather-related terms could increase and improve the amount of impact information that can be determined from social media. Furthermore, this may also provide the potential to study impacts of an event independently of the weather hazard being experienced. Chapter 6 began to explore this approach by including both heatwave and heatwave impact related terms. However, further work in this area is recommended.

Another identified limitation in some of the studies highlighted is that many studies rely on social media datasets in one particular language, usually the native language of the country within which the impact data is required (i.e. English for studies focused on UK, USA, Australia, etc; Japanese for studies focused on Japan; etc). However, for countries with multiple languages in use (e.g. India) or to expand the use of social media for impact data collection beyond one particular country, being able to access and analyse multi-language social media data would be an advantage. In Chapter 5, social sensing rainfall events globally using tweets in the English-language only yielded good global coverage and results. However, this could have been expanded much further if multi-language tweets were added to the dataset. In Chapter 6, social sensing heatwaves in different countries incorporated tweets in English, Dutch and Greek, providing an example of working with tweets in more than one language. Therefore, further work to expand this work to include multiple language tweets is recommended.

De Bruijn *et al.* (2019, 2020) include multiple languages in their methods and there may be much to learn from this example in terms of developing the analysis of tweets beyond single language studies. Methods which employ machine translation techniques to filter social media content for relevance will need to be explored, as will the correct terminology and colloquialisms relating to weather and weather impacts in different languages. However, use of machine translation techniques (e.g. using Google Translate) needs to be approached carefully as direct translation of localised terminology and colloquialisms may yield different results in terms of what constitutes a weather event. For example, in Dutch, the expression “de mussen vallen van het dak” has a direct translation into English using Google Translate of “the sparrows fall from the roof” which appears to have little relevance to the weather. Yet in The Netherlands this expression is used when the weather is particularly hot and therefore would be useful to include in any assessment of discussion during heatwaves and periods of high temperature. Developing an understanding of the use of language in different languages and contexts, potentially working with residents and/or native speakers of different countries, is another important future development for social sensing.

Expanding the use of social sensing more broadly to incorporate multiple languages internationally or more targeted regions that may have different cultural groups domestically may require some further work to understand cultural contexts. This is particularly important when analysing humour in social media posts. Humour has been found to be used as a coping mechanism to manage fear or to cope with impacts during or the risk of impacts from extreme events (Parkhill *et al.*, 2011; Knox *et al.*, 2016; Demuth *et al.*, 2018). Therefore, social media posts conveying humour may be important in understanding the response to extreme weather events. In Chapter 4, tweets conveying humour were found to make up a significant proportion of the response to Storm Brian in the UK in 2017, many of which conveyed quite sarcastic humour. In addition, Arapostathis (2021) finds tweets containing expressions of irony when examining flood related tweets in Greece in 2016. Different cultural contexts of what constitutes humour (e.g. satire, deadpan, sarcasm, irony) may therefore need to be considered in future research.

Most studies rely on Twitter (or Sina Weibo in China where Twitter is disallowed) due to its ease of access via their dedicated APIs and more public privacy settings which enable user content to be shared and interrogated. However much useful information on other social media platforms may also be useful, e.g. community response on WhatsApp or Facebook groups. However, general privacy constraints for these platforms, such as the need to be a member of particular community groups to access content, may make this data more difficult for researchers, National Meteorological and Hydrological Services (NMHS) and other organisations interested in the impacts of weather events to access. Telegram[35] and other social media platforms which have begun to increase in use in recent years may provide further sources of impact information and have begun to be explored by researchers (e.g. Telegram use during Kerala floods in India (Young *et al.*, 2022). Further work to explore other social media platforms could therefore be another avenue of future research.

The use of other online sources of impact information, such as news and other media sources (e.g. local news online pages, FloodList[36]) as well as crowdsourcing applications (e.g. “Flooded Streets” (Naik, 2016), “MyCoast”

[35] https://telegram.org (Accessed: 17 March 2022)
[36] https://floodlist.com (Accessed: 17 March 2022)

(Wang *et al.*, 2018)) can work alongside social media data to provide a more comprehensive overview of impacts. As shown in Chapter 4, news items are more likely to be shared after the weather event and resulting impacts have occurred. Therefore, social media may be more suitable as a ‘citizen sensor’ for real-time situational awareness. However, examining other online media sources may improve and/or provide validation for findings on social media. Incorporating other media sources into impact information curation may also be useful for the verification and evaluation of impact-based forecasts after the event has occurred.

Some studies in the literature review showed that social media volume in a particular location alone may be an indicator of severity of impacts. However, this could be further explored by identifying the use of language during different weather and identification of key terms relating to the severity of impacts (e.g. fatalities, level and type of damage or disruption, etc) may be possible using social media content. For example, Weaver *et al.* (2021) find changes in the use of particular terms in wind-related tweets as wind speed increases.

Managing the signal to noise when examining peaks in social media discussion related to weather events is another important area to consider in future research. For example, discussion about a historical weather event on social media may be misinterpreted as a current event. Some studies have started to consider this issue by combining weather forecast data or other contextual information with social media data as an additional filter for relevance to a current weather event (Rossi *et al.*, 2018; de Bruijn *et al.*, 2020). Approaches like this could therefore also be explored.

The use of percentile as a measure of Twitter activity for a particular region, rather than the number of tweets, in Chapter 5 was one method which was found to be successful for normalising social media activity across regions with different populations and propensity for posting on social media. However, other events (e.g. elections, global pandemic, etc) have the potential to result in less weather-related social media content being generated than it might otherwise have, due to public attention being focused elsewhere. Future research on social sensing could therefore consider how to normalise the weather-related social media activity against overall activity for a particular region to try to account for when people’s attention might be ‘divided’.

Images and videos are widely shared on social media, however many studies focus mainly on temporal and/or spatial volume of social media posts and text content, as natural language processing methods using text are generally more straightforward to apply. However, developing machine learning methods that can identify impacts from images or video content on social media (e.g. images of floods, videos of wind causing damage, etc) may provide further information and context beyond text content alone. Some studies (e.g. Chaudhary *et al.* (2019); Wang *et al.* (2020a)) have begun to develop methods to interrogate image content. Video content is another avenue that could be explored.

It is also important to consider how the use of social media platforms, such as Twitter, and access to the underlying data may change in the future. For example, in 2018, Twitter removed the ‘timezone’ field from the data available via the API due to GDPR (Cowie *et al.*, 2018). Timezone had previously been used as an additional measure to determine location in the social sensing code, however the removal of this field from downloaded tweet data meant that adaptations to the location inference methodology were required. Similar changes to the code were also required in 2019 when Twitter changed the way in which geolocation coordinates were made available in the data (Kruspe *et al.*, 2021). Therefore, in order to future proof social sensing, it will be important monitor changes to the availability of social media data going forwards. It is reassuring, however, that Twitter have recently made changes to their API with the introduction of the API v2 academic access to enable more users, in particular researchers, to access tweet data more easily. So, the future of being able to access the necessary data to support the social sensing approach looks promising.

As found in the literature review in Chapter 2, many studies related to the use of social media for weather impact assessment originate from a wide range of disciplines. Therefore, it is recommended that future work should incorporate research from across the disciplines and focus should be on using research to improve situational awareness, improve warnings and communications, prevent further impacts to homes and infrastructure, and ultimately to save lives!

7.6 Conclusion

In this thesis, a review of the literature related to the social sensing of weather and three novel experimental pieces of work were conducted to assess the usefulness of social media as a source of impact information during extreme weather events. Social sensing was found to be a good approach to accessing relevant impact information from social media and a number of methods and tools for social sensing were explored. There is definitely scope for the development of social sensing methods. Therefore, suggestions for future work to advance and extend the curation of impact information from social media have been proposed. Social sensing could also be developed into real-time operational tools to improve situational awareness during extreme weather events. Additionally, it could be used in the evaluation of impact-based forecasts and warnings issued by organisations such as NMHS and policy makers concerned with the impacts of extreme weather.

Appendices

A. Supporting table to Using social media to detect severe weather events and evaluate impacts: a systematic literature review

	Ref. **- Social Media platform** **- Weather type** **- Country of interest**	**Main Aims and Ideas**	**Methods & Tools**	**Location Inference**	**Advantages**	**Disadvantages**
1.	Ahmad *et al.* (2019) - Twitter - floods/hurricane - USA/Puerto Rico	Used flood related images (manually annotated) to automatically identify passable roads as a result of flooding during a number of hurricane events in the United States. The authors compare four models in their analysis. They further examine satellite imagery for the same purpose.	Use neural network method using pre-trained datasets to extract relevant images and identify if road is passable or note. Then use convolutional neural networks and transfer learning-based classification approach with satellite images for the same purpose.	N/A	Found that a model trained on places, rather than an objects dataset achieves better results in terms of identifying if a road is passable or not.	Relies on a large training set of images.

2.	Alam *et al.* (2018b) - Twitter - hurricane - USA	Collected tweets containing specific references to three named hurricanes (Harvey, Irma and Maria). The sentiment of tweets was calculated to understand emotional response throughout the period of the hurricane. Then used text and image analysis to understand topics of discussion pre, during and post each hurricane event in the study.	Used Random Forest machine learning classifier to filter tweet text for relevance and into categories of discussion. Use LDA for topic modelling of filtered tweet content. Sentiment calculated using Stanford sentiment analysis classifier. Classify images for relevance and impact severity using a pre-developed image classification model (Nguyen *et al.*, 2017).	N/A	Provide a method for automatically categorising tweets into topics of discussion during a hurricane event. Found how sentiment and topics of discussion change on Twitter during the lifecycle of a hurricane.	Approach analysed the tweet dataset as a whole and did not consider the location of the tweet.
3.	Anam *et al.* (2019) - Twitter - hurricane - USA	Build a methodology to analyse time-frequency features of words on social media to identify context before, during and after a hurricane event.	Tweets filtered based on keywords relating to Hurricane Michael and Florida. Use Continuous Wavelet Transform (Anam *et al.*, 2018) to create word features and clusters in both time and frequency. Identify themes to conversations associated with stages of a disaster.	N/A	Find that wavelet features reflect the topic that exists in both time and frequency of their occurrence.	This is a complex methodology, therefore application will rely on adding efficient input from domain experts.
4.	Andrade *et al.* (2017) - Twitter - rainfall - Brazil	Compare rainfall data with tweet data to determine if social media can be used as a proxy for rainfall observations.	Filter tweets using keywords relating to rainfall then manually check tweets for relevance to rainfall.	No - use only geo-tagged tweets	Provides a novel approach to understand the correlation between rainfall data and rainfall related social media messages.	Different time lags between peaks in rainfall and tweet data are found. Manual filtering of tweets in a labour-intensive process.

5.	Anzai & Kazama (2018) - Instagram - floods - Japan	Using Instagram data seek to understand the behaviour of individuals around rivers during flooding events.	Instagram data filtered using hashtag reference to river names. Conducted time series, text and image analysis. Keywords used to identify post relevant to the impacts of flooding.	Some – manually determined based on observation of photographs.	Provides a method for using Instagram data for the analysis of impacts after a flood event, which is less 'noisy' than Twitter data and therefore more likely to be relevant.	Relies on collecting data for a known event. Did not account for population surrounding rivers increasing the number of posts in results.
6.	Arthur *et al.* (2018) - Twitter - floods - UK	Developed an automated process to plot maps of flood activity in the UK using social media.	Tweets filtered using pre-trained data and a Naïve Bayes machine learning algorithm.	Yes – use both place names in user location and tweet text	Provides an automated process for flood event detection. Verified against known flood data finding a good correlation between peaks in tweet activity and flood events.	Location of tweet depends on location information being discernible in the tweet at a suitable administrative level.
7.	Assis *et al.* (2015) - Twitter - floods - Germany	Use hydrological data to prioritize social media messages in areas where flooding is anticipated.	Filter tweets using location and distance to known flooded area.	No - use only geo-tagged tweets	Provide a method for examining tweets in real-time where flooded areas are already known.	Will not detect floods discussed on Twitter if they are not in an area already known to be flooded/at risk of flooding. Limited by only using geo-located tweets.
8.	Bai *et al.* (2020) - Sina Weibo - floods - China	Selected Weibo posts based on keywords relating to flooding and which contained specific location information. Classified posts into four topic categories to determine the severity of flood risk.	A machine learning logistic regression ordered multi-classification algorithm was developed to filter for relevance and classify posts.	N/A	Level of flood risk is mapped and compared with the actual precipitation rainfall and finds good agreement.	The authors acknowledge different environmental factors outside of text of social media post which may affect their severity calculation - i.e. elevation, high tide, etc.

9.	Barker & Macleod (2019) - Twitter - floods - UK	Developed a pipeline for retrieval of flood social data by first obtaining locations at risk of flooding using national flood warning and river level web data sources. Then combine these locations with geo-tagged tweets which have been filtered for relevance to flooding.	Use a trained Doc2Vec model and logistic regression machine learning algorithm.	No - use only geo-tagged tweets	Provides a method for obtaining and filtering flood tweets in real-time which could be used to develop an application to monitor social media discussion about flooding. Use of existing information about flood activity helps to filter tweets for relevance.	Limited by only using geo-located tweets.
10.	Bhuvana & Arul Aram (2019) - Facebook & WhatsApp - floods - India	In addition to conducting a survey of Facebook and WhatsApp users satisfaction levels, the authors also examined messages posted on these platforms during the Chennai floods in India in 2015 to determine what users post and therefore how this information could help with disaster relief efforts.	Looked at keywords and hashtags within posts during the period of the disaster	N/A – looked only at specific Facebook and WhatsApp groups for the location of interest	Provides an alternative source of social media information to Twitter for disaster relief efforts.	Need to have access to specific Facebook and/or WhatsApp groups to get to data. Size of dataset is small.
11.	Brouwer *et al.* (2017) - Twitter - floods - UK	Present a method for creating deterministic and probabilistic flood maps from Twitter messages that mention locations of flooding.	Filter tweets based on keywords relating to flooding and the location of the case study (York, UK).	Yes – use place names in the tweet text	Find good agreement between flood estimates and known information about the flood extent.	This method has a tendency to over-estimate the extent of flooding.
12.	Brovelli *et al.* (2014) - Twitter - floods - Italy	Analysed Twitter data to determine its use in flood damage assessment. Focus of paper is method for searching for relevant tweet data and what information is contained within tweet data.	Tweets filtered using keywords (in Italian) related to flooding.	No – Used only geo-located tweets	Find information on damage as a result of flooding mainly in photographs included in tweets.	Size of dataset is small.
13.	Butgereit (2014) - Twitter - weather - South Africa	Examined tweets containing keywords relating to weather to determine if this could be used as a proxy for summarised weather reports.	Use a 'mu' model to identify keywords from tweets filtered for keywords related to the weather.	No - use only geo-tagged tweets	Found that weather events could be determined with >90% accuracy using tweets when the weather is extreme or rapidly changing.	Less success with method for smaller, or less exceptional weather events. Limited by only using geo-located tweets.

14.	Cecinati *et al.* (2019) - Twitter - heatwave - India	Compare heatwave-related social media data with heatwave mortality in India to explore its use as a source of heatwave impact data, particularly for developing countries.	Tweets filtered for relevance using heatwave keywords only.	No – use tweets containing the word 'India' only.	Comparison of tweet activity with other climate- related data. Find a positive correlation between increased heatwave-related tweet activity and heatwave mortality rates.	Tweet dataset is limited by restricted keyword search. Location information limited to 'India' only.
15.	Cerutti *et al.* (2016) - Twitter - floods - Italy	Use geo-referenced tweets to find the spatial extent of a past flood event in Italy.	Filter tweets based on keywords relating to flooding and location in Italy. Use clustering algorithms to detect peaks in tweet activity relating to flooding.	No - use only geo-tagged tweets	Provides a method for the identification of flood events at a regional scale with compares well with ground truth data.	There are a number of false positives found in the results.
16.	Cervone *et al.* (2015) - Twitter/Flickr - floods - USA	Use images collected in real-time from Twitter relating to flooding to assess the extent of damage and disruption of transport infrastructure during the Boulder, Colorado floods in 2013.	Develop the 'CarbonScanner' application to identify 'hotspots' of tweets activity with flood-related keywords and collect relevant tweets in the location of the flooded area. Relevant images showing flood extent were fused with satellite data to identify travel infrastructure flooding.	Yes – use place names identified in tweet text and locate using gazetteer database. Use only geo-located Flickr posts.	Provide a methodology for use of images in flood damage assessment – in particular for travel infrastructure disruption. Incorporating other sources of information improves accuracy of results beyond just looking at Twitter images in isolation.	Only able to geo-locate 8% of Twitter data collected.
17.	Chaudhary *et al.* (2019) - other - floods - not specific	Quantify flood water depth using images by estimating how far objects in the images are sunk into the water.	Use a pre-trained flood image dataset to train a deep learning model to estimate the water level.	N/A	Proved the ability of the trained model to effectively predict water level from images within an acceptable error.	Some objects are wrongly classified which impacts flood depth estimates.

18.	Chen *et al.* (2020) - Twitter - hurricane - USA	Analysed tweets in a bounding box for 'Houston' in the United States during Hurricane Harvey with the aim of determining the impacts on the region's roads during the event.	Tweets were filtered for relevance using pre-defined 'highway' terms. Data was normalised using 'pre-peak' phase data as a baseline and then calculated intensity of tweet activity above this baseline.	No – use of geo-tagged tweets only	Higher intensity tweet activity on a highway proved to be an indicator of more severe impacts on that highway. Found that users mostly talk about 'delays' and 'flooding' when a road was impacted.	Limited by only using geo-located tweets.
19.	Chien *et al.* (2017) - Flickr - hurricane - Taiwan	Examine spatial and sentiment values of Flickr posts during a Typhoon Morakot in Taiwan to explore its use in facilitating disaster management.	Flickr posts filtered for relevance using location of user profile and keywords. Sentiment calculated using Simple Sentiment Word-Count Method (SSWCM).	Some – use location information in user profile	Provides a novel method for using Flickr posts to highlight areas where disaster response may be required.	Some irrelevant posts in final dataset. Do not take account for propensity for Flickr use or population in particular locations.
20.	Congjuico (2015) - Facebook - floods - Philippines	Examine community Facebook posts and comments during flooding event in the Philippines for the purpose of flood risk management by local government officials.	Facebook posts and comments for particular government officials in area of interest were collected.	N/A	Information shared by communities on Facebook gives good detail on when and where impacts as a result of flooding are caused. Much of the useful information was contained within Facebook comments	Requires access to specific community Facebook groups (which are often private) in area of interest where information on flooding impacts shared.
21.	Crisci & Grasso (2013) - Twitter - heatwave - Italy	Using Twitter data, identify areas where people are most likely to be impacted by heat during a heatwave event in Italy. Combine tweet data with other meteorological data.	Filtered tweets based on keywords (in Italian language) relating to heat and the impacts of heat.	No – use of tweet being in Italian language used as a proxy for tweet originating from Italy.	Provides a methodology for identifying locations where people are suffering from 'heat stress' during a heat wave event. Use of other meteorological data confirms findings from Twitter.	No relevance filtering is conducted on tweets collected therefore dataset may have contained irrelevant tweets.

22.	Dalela *et al.* (2020) - Twitter - storms - India	Examine existing tweet datasets for a number of storm events in India, Pakistan and Nepal to identify categories relating to flooding impacts. Then develop a method to automatically classify tweets related to Storm Fani in India. Results are compared using different machine learning models.	Training dataset which contains tweets from previous storms is manually tagged with 9 categories of impact. This is then used to compare different machine learning models to automatically filter for relevance and classify tweets relating to Storm Fani. A linear SVC approach is found to be the best method for the classification task.	N/A	Provide a method to automatically classify tweets using data from previous similar events with fairly good precision. This method could provide a real-time classification approach.	Do not take location into consideration in the filtering stage. Method is limited to storms in this particular region of the world.
23.	de Bruijn *et al.* (2019) - Twitter - floods - global	Develop a method for detecting flooding events in real-time using social media data. Validate events detected using known flood event data from NatCatSERVICE[37].	Use BERT to filter tweets for relevance and a 'burst detection' algorithm to determine likelihood of flood event.	Yes – use place name mentions in tweet text (using TAGGS algorithm (de Bruijn *et al.*, 2018))	Provide a method for detecting floods in real-time. Validate their approach using known flood event data and find that 90% of flood events are detected using their method.	Some events which were not flood events were also detected.
24.	de Bruijn *et al.* (2020) - Twitter - floods - global	Describe a multilingual multimodal neural network which uses both textual and hydrological information to filter tweets for relevance to flooding.	Use a multimodal neural network to filter tweets for relevance to flooding.	Yes – use place name mentions in tweet text (using TAGGS algorithm (de Bruijn *et al.*, 2018))	Including hydrological features in the model produces better results than without.	Differentiating between man-made and natural events is sometimes difficult.

[37] https://climate-adapt.eea.europa.eu/metadata/portals/natcatservice-database-year-of-launch (Accessed: 17 March 2022)

25.	de Vasconcelos *et al.* (2016) - Twitter - rainfall - Brazil	Compare rainfall data with tweets related to rainfall to determine if social media data can complement existing rainfall observations.	Filter tweets based on keywords related to rainfall.	Some – use specific place names mentioned in user location	Find a close relationship between the number of tweets and the amount of rainfall.	Assume that all tweets containing keyword relating to rainfall are relevant.
26.	Demuth *et al.* (2018) - Twitter - hurricane - USA	With the aim of understanding the perception of risk during an extreme weather event, tweets for 53 Twitter users during the period of Hurricane Sandy were analysed. They also carried out sentiment analysis (types of emotion) for these tweets.	Tweets were manually reviewed for relevance and sentiment (type of emotion).	Yes – but location is manually determined from information contained within the tweet.	Provides a method to determine people's thoughts and behaviours during an extreme weather event. Found that perception of risk and severity of impacts from extreme weather varies from individual to individual and is also based on the forecast / information received beforehand.	Manual review of tweets is a labour-intensive process.
27.	Dong *et al.* (2013) - Twitter - hurricane - USA	Determine topics of discussion from tweets which have been filtered for relevance during Hurricane Sandy.	Use Latent Semantic Indexing (LSI) algorithm in genism to produce topic model of tweet content.	No – use only geo-located tweets.	Provide a method to determine topics of discussion from tweet text during a hurricane.	Limited by use of geo-located tweets only.
28.	Eilander *et al.* (2016) - Twitter - floods - Indonesia	Apply the ‘wisdom of the crowd’ principle to aggregate tweets relating to flooding, and which mention flood depth, in Jakarta (Indonesia) and produce flood probability maps.	Filter tweets based on keyword relating to flooding, flood depth mention and location in area of interest.	Some – use specific place names mentioned in the tweet text	Provides a method for producing flood inundation maps using social media data.	Method used only applicable to this particular location in the case study.

29.	Fang *et al.* (2019) - Sina Weibo - rainfall - China	Analyse rainfall-related Weibo posts to explore its use as a source of disaster impact data in China.	Posts were filtered for relevance to rainfall impacts using rainfall-related words contained within the message. Classify tweets into four categories of disruption. Conducted word frequency analysis and mapped peaks in Weibo activity.	Some – used specific place name mentions in tweet text.	Found that peaks in social media activity correlated with peaks in rainfall intensity. Word frequency analysis identified the main causes of disruption as a result of the rainfall as well as where in the Wuhan region the biggest impacts were realised.	Peaks in activity are skewed towards more populated areas of the region as did not consider population density or propensity to post to social media in certain areas.
30.	Farnaghi *et al.* (2020) - Twitter - hurricane - USA	Develop a methodology for detecting weather events using a case study for Hurricane Florence.	Detected clusters of tweet activity using content, time and location of tweets as a method for event detection before, during and after the hurricane event.	No - use only geo-tagged tweets	This provides a method for identifying the location of impacts during a hurricane event.	Limited by use of only geo-tagged tweets.
31.	Feng & Sester (2017) - Twitter/Flickr - rainfall - England/France	Use rainfall related tweets as a proxy for rainfall observations.	Use machine learning algorithm (SVM) to filter tweets for relevance to a rainfall event.	No - use only geo-tagged tweets	Provides a method for producing real-time rainfall monitoring using social media.	Limited by use of geo-located tweets only.
32.	Fitrianah *et al.* (2020) - Twitter - floods - Indonesia	Developed a real-time application which would show users all tweets based on user inputted keywords for Jakarta (Indonesia) with the aim of providing situational awareness during flood events.	Tweets filtered based on the presence of keyword ds and specific place name mentions in the tweet text.	Some – use geo-tags and specific place name mentions in the text.	Provides a method for monitoring tweets related to flooding in real-time.	Limited by a reliance on tweets with geo-tags and specific place name mentions.
33.	Fohringer *et al.* (2015) - Twitter, Flickr - floods - Germany	Produce flood inundation maps using images posted on Twitter and Flickr relating to the depths of flooding.	Filter posts based on keyword relating to flooding in the text. Use 'PostExplorer' to manually filter posts containing images for relevance to flooding.	No - use only geo-tagged tweets	The use of flood related images to identify locations impacted by flooding.	Limited by use of geo-located posts only.

34.	Giuffrida *et al.* (2020) - Weather - Twitter - USA	Compare tweets containing specific weather-related terms with actual weather data to see if changes in social media activity are an indicator of changes in human comfort due to weather conditions.	Tweets containing specific phrases are filtered as relevant. Crowdflower[38] used to make a training dataset, marking tweets as relevant or not, then further tagging the sentiment of the tweet as either positive, neutral or negative. A machine learning algorithm is then used to filter tweets for relevance and classify the sentiment.	No – use geo-tagged tweets only.	Found that Twitter is a suitable source of data to assess the effect of weather on human comfort.	Limited by use of only geo-located tweets.
35.	Grace (2020a) - Twitter - storms - USA	Analyses toponym usage and granularity across types of storm and non-storm-related information posted on Twitter.	Filter tweets using geo-location coordinates, tweets containing place names in the area of interest, and tweets which are part of a related network of users to the area of interest.	Some – use geo-tags and identify specific place names in tweet text.	Identify toponym usage and patterns that are likely to indicate impacts as a result of a storm. Find that users more likely to report place names of event in tweet during a crisis situation.	Findings limited to a single crisis event.
36.	Grace (2020b) - Twitter - storms - USA	Collected tweets using keywords, place names and tweets from users known to be in a particular location during storm events in the USA and categorise tweets into topics of discussion.	Tweets initially filtered for location using a range of methods to broaden the size of dataset available. Tweets then manually labelled into six categories of storm related information. Tweets then manually coded for relevance to a storm or not.	Yes – use place name mentions in the tweet text and network filtering based on known users in a location.	Provides a novel methodology for determining location of a tweet post during a storm event.	Manual labelling of tweets, while having high accuracy, is a labour-intensive process.

[38] https://visit.figure-eight.com/People-Powered-Data-Enrichment_T (Accessed: 17 March 2022)

37.	Grasso *et al.* (2017) - Twitter - heatwave - Italy	Analyse heatwave-related tweets in Italy to explore its use as a source of heatwave impact data in urban areas.	Tweets filtered for relevance using most common heat-related keywords.	Yes – use place name mentions in the tweet text.	Find an increase in tweet activity related to heatwave episodes in locations most prone to heat-related impacts.	Relevance filtering limited to keywords only.
38.	Halse *et al.* (2018) - Twitter - tornado - USA	Compare tweets with physical weather sensor data during a tornado event.	Tweets are collected within a 50-mile radius of the centre point of the tornado event and filtered for only those containing key weather-related terms. Tweets manually coded into 4 topic categories based on their content.	No – use only geo-tagged tweets.	Found that tweet activity correlates with wind speed. Categorisation of tweets shows how the discussion on Twitter changes during the tornado event and could be used as a model for future studies looking at tweets during tornadoes.	Due to limited filtering steps, there is a lot of unrelated data in the tweet dataset and many tweets are retweets of warnings already issued. Limited by use of geo-located tweets only.
39.	Hamoui *et al.* (2020) - Twitter - floods - Saudi Arabia	Explore the use of tweets in Arabic dialects to build a dataset (FloDusTA) of tweets for event detection during weather events in Saudi Arabia.	Tweets filtered for relevance based on keywords and location. Using a manually annotated tweet sample, use a Support Vector Machine algorithm to annotate tweets.	Some - use user location, hashtags, country and time zone to infer location	Provide a method for building a dataset of relevant tweets in the Arabic language during weather events.	Use of time zone filter will exclude tweets with no time zone data.
40.	Han & Wang (2019) - Sina Weibo - floods - China	Categorise Weibo tweets into topics of discussion during the 2018 Shouguang city flood.	Tweets filtered for relevance using keyword ‘Shouguang’. Combine LDA and Random Forest (RF) algorithm to classify tweets into topics.	Some – look for mention of specific place names in the tweet text.	Provide a method for categorising tweets into topics which performed well.	Use absolute number of tweets in analysis – do not account for propensity for tweets in particular locations or population.
41.	Herfort *et al.* (2014a) - Twitter - floods - Germany	Use geographical relations to prioritize social media messages relating to known flood events.	Filter tweets based on keywords relating to flooding in both English and German. Manually classify tweets into thematic topic categories.	No - use only geo-tagged tweets	Find that flood-related tweets that contain useful information for situation awareness are more likely to be closer to flood affected regions than others.	Will not detect floods discussed on Twitter if they are not in an area already known to be flooded/at risk of flooding. Limited by only using geo-located tweets.

42.	Herfort *et al.* (2014b) - Twitter - floods - Germany	Use hydrological data from other sources to determine locations of flood activity which is then used to filter tweets relevant to the flood based on geo location information in the tweet.	Use hydrological data to detect floods. Filter tweets using location of known floods.	No - use only geo-tagged tweets	Find that combining tweet location with hydrological information aids with the filtering of tweets.	Limited by only using geo-located tweets.
43.	Huang *et al.* (2018) - Twitter - floods - USA	Produce flood inundation maps using remote sensing, stream gauge and Twitter data relating to flooding.	Filtered tweets for location within the area of interest and containing flood related keyword. Manually filtered tweets for relevance.	No - use only geo-tagged tweets	Provides a method to quickly identify areas in need of urgent attention during flood events.	Small tweet sample size used in case study. Limited by use of geo-located tweets only.
44.	Huang *et al.* (2019a) - Twitter - floods - USA	Use a visual-textual approach to extract flood related Twitter data using both text and images.	Use convoluted neural network to filter tweets for relevance to flooding using both text and images manually labelled training set.	No - use only geo-tagged tweets	The addition of images in the relevance filtering process leads to better classification accuracy.	Neural network approach used applies equal weight to text and images in a post which may not be true in reality. Limited by use of geo-located tweets only.
45.	Huang *et al.* (2019b) - Twitter - floods - USA	Tweets were collected for the state of Texas in the USA only in the period before during and after a flood event to determine relevant flood keywords. Image analysis to find images relating to floods in the tweets conducted using training data containing images of floods previously collected and tagged from Twitter, Instagram and Flickr.	Developed a neural network methodology to process images extracted from the tweet dataset and corresponding tweet text to identify those of relevance to a flood event.	No - use only geo-tagged tweets	This study provides an approach for tagging flood tweets using both images and text.	Limited by only using geo-located tweets.
46.	Jitkajornwanich *et al.* (2019) - Twitter - floods - Thailand	Propose an early flood warning system using flood related Twitter data to produce maps of where floods are likely to be occurring.	Filter tweets based on keywords, location and NLP via string matching (Levenshtein's algorithms).	Yes – use place names in the tweet text.	The use of multiple filtering techniques and location inference from tweet text provides near accurate maps of flood activity.	Not all flood events were captured, therefore recommended as a complement rather than a standalone method.

47.	Jongman *et al.* (2015) - Twitter - floods - Philippines, Pakistan	Use satellite observations and tweets to improve situational awareness during flood events in Pakistan and the Philippines.	Tweets were filtered based on the presence of keywords and location names in the tweet text.	Some – tweets with specific place names in the tweet text were used.	Find the combination of satellite data and peaks in tweet activity around the area of a flood is helpful in terms of providing information about when and where floods are present.	Found that a more comprehensive approach for filtering of tweets for relevance was required.
48.	Jony *et al.* (2019) - Twitter, Flickr - floods - not specific	Develop a method to extract flood related posts using both text and images from social media.	Use convoluted neural network to filter tweets for relevance to flooding using both text and images manually labelled training set.	N/A	The combination of using both text and images when filtering social media posts for relevance yields better accuracy than looking at each separately.	Method needs to be tested using a suitable flood case study.
49.	Jung & Uejio (2017) - Twitter - heatwave - USA	Analyse tweets related to different heat-related themes for five US cities to determine if there is a relationship between tweet activity and heat-related activity/need in cities.	Tweets filtered for relevance using heat-related terms plus other terms indicating how/when/where affected. Irrelevant tweets identified using known irrelevant phrases.	Yes – use place name mentions in the user location field.	Find a positive relationship between heat-related tweets and heat exposure metrics in three out of five cities studied.	Relevance filtering limited by use of keywords only.
50.	Kankanamge *et al.* (2020) - Twitter - floods - Australia	Consider the three phases of a disaster event (pre-, during- and post-) and use a 'Capture-Understand-Present' framework to collect, process and analyse tweets with the aim of determining disaster severity. Categorise tweets into topics and calculate the sentiment of tweets during the event.	Tweets were collected based on geolocation, event dates and keywords. The same machine learning methods used by Arthur et al. (2018) were used to determine tweet relevance. Also used a decision tree machine learning algorithm to classify tweet text into topic clusters and sentiment.	No – use only geo-tagged tweets.	Show how tweet activity changes during the different stages of an extreme weather event and provide some methods that could be used to better understand where relief efforts are needed.	Limited by use of only geo-located tweets.

51.	Khaleq & Ra (2019) - Twitter - hurricane - USA	Develop a real-time, cloud-based, application which can be used to monitor impacts during a disaster event,	Use a cloud based machine learning service (Azure PaaS) to classify tweets for disaster relevance and then further categorise into topics. Use a microservice architecture application to develop their real-time application.	Yes - use place name mentions in the tweet text.	Provides a real-time application that could be used to monitor disaster impacts during an extreme weather event.	Small size of tweet dataset used to test the application.
52.	Kwon & Kang (2016) - Twitter - floods - South Korea	Assess tweets relating to flooding for 'level of risk' and 'vulnerability to flooding'. Results are plotted on a map to show areas most affected by flooding.	Filter tweets based on keyword related to flooding in the tweet text.	Some – use specific place name mentions in the tweet text.	Provides a method for both weather impact event detection and assigning a level of severity to the impact.	Reliance on keywords specific to the location and case study event makes this method difficult to use more widely.
53.	Lachlan *et al.* (2014) - Twitter - hurricane - USA	Analyse the type of content tweeted before, during and after Hurricane Sandy.	Tweets manually coded for relevance and into topics of discussion	N/A	Find how tweet content changes during the different stages of a hurricane. Provide a coding strategy for identifying topics of discussion in tweets during a hurricane.	Manual filtering and coding of tweets is a labour-intensive process.
54.	Li *et al.* (2018) - Twitter - floods - USA	Filter tweets containing keywords relating to a flood event. Comparing filtered tweets with stream gauge data, they created a model to produce inundation maps of the flood extent.	Extract key phrases from filtered tweets to provide qualitative information for forecasters during the event.	No - use only geo-tagged tweets	Their approach is able to produce flood inundation maps and is validated using official inundation maps, finding good agreement.	Limited by only using geo-located tweets.
55.	Lin *et al.* (2015) - Twitter - snow - USA	Compare weather data, Twitter data and traffic information to determine if Twitter can be used to predict the impact of the weather due to snow on traffic flow.	Filter tweets based on keyword relating to snow. Manually review filtered tweets for relevance. Use linear regression model to analyse data including and excluding tweets.	No - use only geo-tagged tweets	Find that Twitter has high sensitivity for predicting impacts due to snow.	Used small dataset of tweets covering short period of time and one location. Limited by use of geo-located tweets only.

56.	Liu *et al.* (2020) - Twitter - floods - USA	Used geo-located tweets which contained flooding impact keywords within the specific location of a flooding event in Colarado in 2013 to identify impacts during the event.	A manual analysis of the tweet text was undertaken to identify impact information. Images were also extracted from tweets and further analysed both manually and using a machine learning approach.	Some – initially only used geo-tagged tweets, then added further tweets using specific place name mentions in the tweet text.	Found that tweets contained reliable information about impacts during a flooding event.	Initial reliance on geo-tagged tweets only. Approach relied on a lot of manual effort.
57.	Lu *et al.* (2018) - Sina Weibo - weather - China	Use social media data to identify traffic disruption in Beijing as a result of adverse weather conditions.	Filter tweets using word2vec model trained on traffic-related tweets and news records. Use regression model to identify adverse weather and traffic incidents from filtered social media dataset.	Yes – use place names within tweet text	Methods used to build a prototype real-time warning system for traffic incidents during adverse weather.	The model does not exclude traffic incidents which are not as a result of adverse weather.
58.	Lwin *et al.* (2015) - Twitter - rainfall - Japan	Compare JMA rainfall data and rainfall-related tweets during a rainfall event in Japan.	Filtered tweets using multi-language keywords.	No – use geo-located tweets only	Use of multi-language tweets. Developed a tool and method for easily extracting tweets containing specific keywords.	A number of irrelevant tweets in final dataset. Limited by use of geo-located tweets only.
59.	Ma & Surakitbanharn (2019) - Twitter - hurricane - USA	Compare damage-related tweets, socio-economic data and insurance claim information in Florida during Hurricane Irma.	Machine learning filter (Bidirectional Long-Short Term Memory (LSTM) recurrent neural network) to identify and classify relevant tweets. Sentiment calculated using TextBlob.	No – use geo-located tweets only	Provides a method to potentially identify hurricane damage using social media.	Limited by use of geo-located tweets only.

60.	Martín *et al.* (2017) - Twitter - hurricane - USA	Examine tweet activity during Hurricane Matthew to assess the evacuation responses of residents.	Filter tweets for relevance using hurricane-related keywords.	No – use geo-tagged tweets only.	Provide a method for examining how the public respond to evacuation orders.	Do not account for the propensity for tweets or population in particular locations. Limited by use of geo-located tweets only.
61.	Moumtzidou *et al.* (2018) - Twitter - floods - not specific	Use text and images in social media posts to identify if content is relevant to flooding or not.	Use DCNN-based features and SVM classifier to classify images. Use TF-IDF method and Random Forests to filter tweets for relevance using text.	N/A	Use of both text and images to classify social media posts as relevant to flooding events increases the information available.	Images and text are treated separately in the proposed method. Method is tested on sample dataset only.
62.	Mukkamala & Beck (2016) - Twitter - hurricane - USA	Analyse nature and characteristics of tweets during Hurricane Sandy	Tweets manually filtered to exclude obviously irrelevant posts. Manually coded tweets based on information source and nature of message.	No – use geo-located tweets only.	A thorough filtering and coding scheme provides a useful insight into the nature of tweets posted during a Hurricane, finding that there is good real-time, first hand information about the event in Twitter data.	Manual filtering and labelling of tweets is a labour intensive process. Limited by use of geo-located tweets only.
63.	Musaev & Hou (2017) - Twitter - landslides - not specific	Provide a method for the classification of tweets relevant to a landslide event using both semantics from the tweet text and user influence.	Supervised machine learning filter (SVM) to classify relevant tweets.	No – use geo-tagged (or manually geo-tagged) tweets only.	Provides a novel approach to classifying tweet relevance to an event using both tweet text and user influence, achieving good accuracy.	Method does not distinguish between multiple events in the same location or 30-day time window.
64.	Nair *et al.* (2017) - Twitter - floods - India	Examine tweet activity and topics of discussion during the 2015 Chennai flood.	Tweets filtered for relevance using one specific hashtag. Find machine learning Random Forest classifier has best performance for tagging tweets with a particular topic	N/A	Provide a method for categorising tweets into topics during a flood event.	Do not filter tweets for relevance – assume all tweets containing search hashtag are relevant.

65.	Oktafiani *et al.* (2012) - Twitter - floods - Indonesia	Provide a method for extraction of Indonesian language tweets relevant to flooding in Jakarta (Indonesia).	Use graph concept, complex network analysis and RIDF term weighting to identify topics of discussion using tweet text.	N/A	Provides a system for topic identification using tweets.	Method does not exclude retweets or tweets relevant to other events (e.g. detects tweets about an election).
66.	Owuor *et al.* (2020) - Twitter - hurricane - USA	Compare the track of Hurricane Dorian with the location of event mentions on Twitter and GDELT (events from news feeds).	Filter tweets based on keywords related to Hurricane Dorian.	No - use only geo-tagged tweets	Find a good match between peaks in tweet activity and hurricane track.	Limited by use of geo-located tweets only.
67.	Pereira *et al.* (2020) - Flickr, other - floods - not specific	Developed an algorithm based on a neural network to predict if a social media image shows a flood event and then estimated the flood depth from the image. Images were taken from the social media platform Flickr and another bespoke European Flood 2013 dataset.	Used DenseNet and EfficientNet neural network architectures to classify images and estimate flood severity.	No - Images used had geo-location information.	Found good success with using neural network architectures to estimate severity of flooding using images on social media.	Limited by use of geo-located images only.
68.	Pourebrahim *et al.* (2019) - Twitter - hurricane - USA	Aim to understand the types of communication on Twitter during Hurricane Sandy by surveying Twitter users and analysing tweets from the period and location of the hurricane.	Tweets filtered based on Hurricane Sandy related keywords or geo-location in the area affected by the hurricane. Calculate term frequency (TF-IDF), Klout score (user influence) and sentiment score. Sentiment calculated using a training corpus of labelled tweets and SVM machine learning algorithm to assign a score.	No – used geo-tagged tweets only.	Provide a method for identifying impacts as a result of a hurricane using Twitter data.	Due to limited filtering steps, a large number of irrelevant tweets are included in the analysis. Limited by use of geo-tagged tweets only.

69.	Preis *et al.* (2013) - Flickr - hurricane - USA	Analyse the number of hurricane-related photos posted on Flickr during Hurricane Sandy as a proxy for attention/people impacted by the hurricane.	Filter posts for relevance using hurricane-related terms. Normalise data for time of day posted.	N/A	Provide a method that uses the number of photos posted on Flickr as a proxy for impact as a hurricane approaches and makes landfall. Find a correlation between number of photos posted and the atmospheric pressure.	Do not check photos for relevance to the hurricane event.
70.	Restrepo-Estrada *et al.* (2018) - Twitter - floods - Brazil	Use both flood/rainfall related tweets and official rainfall data to detect the spatial impacts of flooding.	Filter tweets using keywords related to rainfall and flooding.	No - use only geo-tagged tweets	Combining social media data with observed rainfall improves estimation of flooding.	There are a number of false positive tweets in the dataset analysed. Limited by use of geo-located tweets only.
71.	Rodavia *et al.* (2018) - Twitter - floods - Philippines	Selected tweets based on keywords relating to flooding in the Manila area (Philippines).	Used 'Association Rule Mining' to filter tweets for location and relevance to flooding	Yes - use place name mentions in tweet text	The study provides a novel methodology for filtering tweets for relevance and location.	
72.	Rosser *et al.* (2017) - Flickr - floods - UK	Provide a method for estimating flood inundation using remote sensing, images from social media (Flickr) and topographic data sources.	Flickr data including images collected using keywords related to flooding. Bayesian statistical model to estimate probability of flood inundation using weights-of-evidence approach.	No – use geo-tagged images only	Good accuracy when results are compared with ground-truth flood extent.	Social media sample size is small. Method cannot be applied automatically.
73.	Rossi *et al.* (2018) - Twitter - floods - Italy	Developed a methodology for detecting flood related tweets which uses both probabilistic weather forecasts and tweet data to detect an ongoing flood event.	Doc2Vec model and logistic regression machine learning algorithm used to filter tweets for relevance.	No - use only geo-tagged tweets	This study provides a machine learning methodology for filtering tweets for a current event using data from previous events.	Limited by only using geo-located tweets.

74.	Roy *et al.* (2020) - Twitter - hurricane - USA	For case studies Hurricane Irma and Michael, used multi-label classification to identify types of infrastructure damage and disruption and their co-occurrence in a tweet. They calculated sentiment as an additional measure of actual disruption or not.	Used a manually labelled training dataset to classify tweets into categories. Calculation of sentiment using VADER (Hutto & Gilbert, 2014).	Yes - used a geo-parsing method to extract location from tweet text.	For the case studies presented, able to plot damage / disruption occurrences on a map for each hurricane event.	Small training dataset for classification of tweets into damage / disruption type. Did not validate findings with known data about where damage and disruption occurred.
75.	Saravanou *et al.* (2015) - Twitter - floods - UK	Analyse tweets posted during a particular flood event in the UK in 2014 to identify locations impacted by flooding.	Build a custom lexicon to filter tweets using keywords related to flooding. Clustering algorithm used to identify areas with greater ratio of flood tweets.	No – use geo-tagged tweets only.	Find good agreement between areas of known floods and areas with a greater ratio of flood tweets.	Method produces poor results for some regions. Limited by use of geo-tagged tweets only.
76.	Sato (2019) - Twitter - rainfall - Japan	Examined tweets with the hashtag #rescue during a period of heavy rainfall in Japan. (#rescue proposed by Japanese subsidiary of Twitter as an alternative to calling emergency services).	Tweets are manually reviewed for relevance to people needing rescue/aid.	N/A	Use of specific hashtag in tweets aids with tweet relevance filtering.	Only a small proportion of these tweets were actually related to people needing rescue as a result of extreme weather event. Manual tagging is a labour-intensive process.
77.	Scotti *et al.* (2020) - Twitter, other - floods - USA	Combine satellite images, hydraulic models and social media posts related to flooding to produce flood inundation maps after a flooding event.	Relevant and located tweet data, including images, obtained from Evolution of Emergency Copernicus database (E2mC) (Havas, 2017).	No – use a pre-existing database where tweets have already been geo-located.	Post-event flooding maps were successfully produced, with social media data in particular providing well-defined spatial and temporal data about the flooding.	Found limitations with the use of satellite and rainfall data.

78.	Shannag & Hammo (2019) - Twitter - floods - Jordan	Develop two classifiers to filter social media data and detect flooding events using tweets in the Arabic language.	Tweets from most popular Arabic news accounts filtered based on the most popular hashtags and location of interest in the text. Tweets then further filtered based on keywords relating to the flooding event.	Some – used specific place names in the tweet text	Filtering process is simple and based on the most common terms relating to a specific event rather than pre-defined terms.	Social media sample size is small. Approach could be enhanced with use of NLP and machine learning techniques to filter social media data.
79.	Shi *et al.* (2019) - Twitter - floods - Japan	Collected tweets containing event and flooding impact keywords. Tweets were filtered for relevance to the flood event and were then further categorised into 5 topic categories.	Manual filtering and categorisation of tweets	No - use only geo-tagged tweets	The authors found that tweet content complemented existing flood inundation information. Manual filtering meant good accuracy for tweet relevance	Manual filtering of tweets, while accurate, is a labour-intensive approach. Only a small number of tweets (109) remained after filtering. Limited by only using geo-located tweets.
80.	Sit *et al.* (2019) - Twitter - hurricane - USA	For case study Hurricane Irma, used classification techniques and spatial clustering to identify areas impacted by the hurricane and with infrastructure damage.	Use Long Short-Term Memory (LSTM) network to classify tweets for relevance. Multi-label topic classification achieved using LDA. Spatial clustering used to identify impacted areas.	No – use tweets with geo-location coordinates only.	Find good agreement between known impacted areas and areas with a strong tweet signal. Provide a method with potential to be used in real-time to monitor the impacts of a hurricane.	Limited by use of geo-located tweets only.
81.	Smith *et al.* (2017) - Twitter - floods - UK	Collected tweets with keywords related to flooding and with a known geolocation in a particular place of interest to identify if peaks in tweet activity coincided with a known flood incident in the UK.	Filter tweets using keywords related to flooding.	No - use only geo-tagged tweets	They find a correlation between peaks in tweet data and the extent and depth of the flood level.	Tweets filtered using keywords only therefore dataset likely to include irrelevant tweets. Limited by only using geo-located tweets.
82.	Sovacool *et al.* (2020) - Twitter - hurricane - USA	Examine tweets before, during and after Hurricane Irma to identify topics of discussion throughout the stages of the storm.	Manual coding of a sample of tweets into categories. Use of NVivo to provide themes of discussion and keywords.	N/A	Find useful information about the impacts of hurricanes. Provide a dataset of topics and keywords contained within tweets related to a hurricane.	Manual coding of tweets into topics is a labour-intensive process.

83.	Spasenovic *et al.* (2019) - Twitter - hurricane - USA	Explore the spatial distribution of Twitter posts related to Hurricane Michael with the aim of gaining insight into the event impact.	Machine learning filter for tweet relevance using methods presented by Barozzi et al (2019). Use Kernel density estimation to explore the spatial distribution of Twitter posts. 'Hot Spot' analysis used to analyse the spatiotemporal distribution of the data.	No - use only geo-tagged tweets	Method provides fast information about when and where impacts during the event were occurring.	Some irrelevant data included in results and some locations flagged as high impact not always accurate. Limited by use of only geo-located tweets in analysis.
84.	Spruce *et al.* (2020) - Twitter - storms - UK	Analysed tweets during named UK/Ireland storms using sentiment analysis and identifying topics of discussion.	Tweets filtered using Naïve Bayes machine learning classifier. Sentiment calculated using TextBlob. Manual categorisation of tweets into topics.	Yes – look for place names in the tweet text and user location	Provide a method for identifying areas of impact during significant storm events. Also provide information about the types of impacts reported on social media and change in sentiment of users during storms.	Sentiment calculated using classifier trained on movie reviews. Manual categorisation of tweets is a labour-intensive process.
85.	Stowe *et al.* (2018) - Twitter - hurricane - USA	Provide a method for identifying hurricane-related tweets and classifying the content to better understand user evacuation behaviour during hurricanes.	Use both a SVM machine learning classifier and deep learning approaches to filter tweets for relevance. Tweets classified into topics using trained machine learning classifiers.	No – use geo-tagged tweets only.	Find that using both linguistic and geospatial features of filtered tweet dataset may provide some information about evacuation behaviour during a hurricane.	Achieve mixed results on performance of method - acknowledge that relevance classifier performance could be improved. Limited by use of geo-located tweets only.
86.	Sun *et al.* (2016) - Flickr - floods - USA	Compare flood-related geotagged Flickr photos with other remote sensing data to explore its use as a complementary data source to other data.	Filter posts for relevance using flood related keywords.	No – use geo-tagged posts only	Demonstrate the advantages of using social media as a complementary data source for remote sensing in areas with limited sensor data.	Small size of Flickr dataset.

87.	Tkachenko *et al.* (2017) - Flickr - floods - UK	Analyse image tags posted on Flickr during several flood events to determine if it can provide an early warning of a flood event.	Filter posts for relevance using flood/river-related keywords. Use a supervised learning method (Deconstructed Cascade Correlation Matrix (DCCM)) to identify Flickr tags with best flood forecasting capability.	No – use geo-tagged posts only	Provide a method for using Flickr posts for flood detection.	Limited by use of specific keywords in Flickr tags.
88.	Tse *et al.* (2017) - Sina Weibo - weather - China	Selected posts from Sina Weibo based on the location being in Beijing city. Posts were then compared with observed weather data to detect if there was a relationship between weather and traffic congestion in the city.	Posts filtered based on keywords related to weather conditions.	No – use geo-located posts only.	Analysis finds a relationship between good weather and traffic congestion.	Limited by use of geo-located posts only.
89.	Ukkusuri *et al.* (2014) - Twitter - tornado - USA	Analysed tweets posted during a tornado event in 2013 to explore the information available from social media on impacts and user behaviour.	Find most frequently used words to identify topics of discussion. Manual categorisation of individual tweets into topics. Sentiment calculated using SentiStrength.	N/A	Provide information about the types of impacts and change in sentiment of users during a tornado event. Simple method for identifying topics of discussion using Twitter data.	Findings limited by some missing data in analysis. Manual categorisation of tweets is a labour-intensive process.
90.	Vayansky *et al.* (2019) - Twitter - hurricane - USA	Analysed tweets containing relevant keywords during Hurricane Irma using sentiment analysis and topic modelling. Findings are compared with observed wind speed data.	Used LDA topic modelling to identify topics of discussion in the tweet data. Sentiment calculated using a bespoke library of words assigned negative/positive polarity scores.	No – used geo-tagged tweets only.	The sentiment analysis showed an inverse relationship between sentiment and wind speed, i.e. a decrease in sentiment score during the period of the hurricane. Topic modelling identified four topic groups relating to types of tweet during the hurricane period.	Limited by use of only geo-tagged tweets.

91.	Wang & Ye (2018b) - Twitter - hurricane - USA	Analyse tweets during Hurricane Sandy in New York City to find out how topics of discussion changed spatially and temporally during the hurricane.	Filter tweets containing keyword 'sandy'. Manually classify tweets for relevance and into topics. Use LQ and Markov transition probability matrix for space-time-content analysis of tweets.	No – use only geo-tagged tweets.	Provide a method for examining tweet activity and topics of discussion during a hurricane both spatially and temporally.	Manual classification of tweets is a labour-intensive process. Limited by use of geo-tagged tweets only.
92.	Wang *et al.* (2016) - Sina Weibo - rainfall - China	Analyse Weibo tweets during the 2012 Beijing Rainstorm to identify topics of discussion.	Use LDA and SVM algorithms to classify tweets for relevance and into topics of discussion.	No – use only geo-location in tweet.	Provide an unsupervised machine learning method for identifying topics of discussion from tweets.	Use of only tweets with geo-location for spatial analysis.
93.	Wang *et al.* (2018) - Twitter, other - floods - USA	Analyse Twitter data and crowdsourced data (form MyCoast app) to produce hyper-resolution flood information at the metropolitan scale. Findings are compared with precipitation and road closure reports.	Filter tweets based on flooding-related keywords. Flood depth information estimated using information in tweet text. Convolutional Neural Network (CNN) used to classify photos relating to flooding from crowdsourced application.	Yes – use location mentions in tweet text.	Find flood-related tweets are linearly correlated to precipitation departure. Provide a method to analyse flood-related images.	For smaller-scale flood event, unable to locate some tweets at the street-level scale required for road closure information.
94.	Wang *et al.* (2020a) - Twitter - floods - USA	Carry out image and text analysis of tweets related to flooding during Hurricane Harvey to develop a method for improving flooding situational awareness	Use toponym recognition tool to locate tweets using GeoNames and TIGER data. Use Convolutional Neural Networks (CNN) and Residual Neural Networks (ResNet) for image classification and topic filtering of tweet images.	Yes – use place names in the tweet text.	Provide a method for detecting flooding impact information using images posted on Twitter.	Results from image classification still required some manual input for checking results.

95.	Wang *et al.* (2020b) - Sina Weibo - floods - China	Examine spatio-temporal patterns of Weibo tweet activity alongside land-use data for a flooding case study in Nanjing city, China.	Use keywords related to flooding to filter tweets. Use a trained classifier to calculate sentiment of tweets. Plot tweet activity using public concern index (no. of flooding tweets/no. background tweets)	No – use geo-tagged (from “check-in” data) posts only.	Provide a method to identify public response to a flooding event.	Filter tweets using keywords only.
96.	Wani *et al.* (2020) - Twitter - floods - India	Used tweets containing specific hashtags related to floods in Kerala during 2018 to filter tweets relevant to the flooding event.	User locations were used to locate tweets so that the number of tweets in different locations during the flood could be monitored.	Yes – used place name mentions in tweet user location.	Provide a method for locating tweets during a specific flooding event.	No further processing of tweets was undertaken. Limited to this event by using only tweets containing specific hashtags.
97.	Wu *et al.* (2020) - Sina Weibo - rainfall - China	Analyse Weibo tweets before, during and after the 2016 Hefei rainstorm in China to identify impacts from topics of discussion.	Used DBSCAN spatial clustering and LDA to identify topics of discussion from tweets. Markov transition probability matrix used to measure changes in topics before, during and after the rainstorm.	No – use geo-tagged tweets only.	Provide a method for detecting topics of discussion during a rainstorm and identify impacts from social media.	Do not account for the propensity for tweets in more urban areas which may have introduced biases into topics of discussion found.
98.	Xin *et al.* (2019) - Twitter - hurricane - USA	Analysed tweets containing hurricane related keywords relating to five specific hurricane events (Bonnie, Sandy, Harvey, Lane, Florence, Michael). The main focus of the study was to refine the process of topic modelling and to determine relationships between the identified hurricane topic clusters.	Carried out topic modelling on the data using LDA and calculated the Hellinger distance to measure the connection between hurricane topics.	N/A	Provide an automated method which reduces the amount of human intervention required to determine relevant impacts from social media content	Use of hurricane names such as “Michael”, “Harvey” and “Florence”, and lack of relevance filtering stage, resulted in some irrelevant social media content being included in results.

99.	Yamada *et al.* (2019) - Twitter - rainfall - Japan	Obtained tweets in the Japanese language, containing the word for 'heavy rain' during a period of heavy rainfall in July 2018. The most frequent key words, hashtags and emojis in tweets were determined. They also analysed the number of retweets and tweets which shared a news URL.	Analysis of tweet characteristics in dataset during event including number of tweets, retweets and tweets sharing a news URL.	N/A	This study provides the characteristics of Twitter posts during a heavy rainfall event. Useful for future researchers wishing to understand patterns of discussion in Japan during a rainfall event.	Location of tweets was not considered, only the Japanese language as proxy for location to Japan.
100	Yang *et al.* (2019b) - Twitter - hurricane - USA	Assessed credibility of a tweet during Hurricane Harvey using tweet text and URL(s) contained within the tweet, as well as the number of tweets/retweets associated with the same event. The authors classified tweets into 5 different topic groups using pre-defined keywords. Tweets were assigned an overall credibility score based on classification and location relevant to event.	Filtered tweets for relevance to impacts and classified into 5 different topics using predefined keywords. Identified events by aggregating tweets based on their topics and location.	Yes - used place names detected in the tweet text and a manually created local gazetteer database.	Found that tweets correlated with the progression of the hurricane and therefore provide a potential method for real-time situational awareness during hurricane events. Also found that the topic of 'flood' was the most common during that particular event.	Do not account for population density / propensity for tweets in a particular location and therefore results are skewed to more densely populated locations.
101	Yang *et al.* (2020) - Sina Weibo - rainfall - China	Use Weibo tweet activity to detect traffic impact areas during the 2018 Beijing rainstorm.	Use hazard damage related keywords to filter tweets for relevance. Used a trained Convolutional Neural Network machine learning classifier to calculate sentiment of tweets.	Yes – use place names detected in user profile.	Find good results in detecting traffic impact areas and the severity of the impact using tweets.	Some irrelevant tweets were included in the analysis dataset.
102	Yao & Wang (2020) - Twitter - hurricane - USA	Propose a domain specific sentiment analysis approach (DSSA) to assess the sentiment of tweets during a hurricane event.	Sentiment score determined using a supervised neural network machine learning method and natural language processing.	No – use geo-located tweets only.	This study provides a method for classifying the sentiment of tweets during an extreme weather event with reasonable success.	Limited by only using geo-located tweets.

103	Yuan & Liu (2018a) - Twitter - hurricane - USA	Combine Twitter and unmanned aerial vehicles (UAV) data to conduct a rapid damage assessment during Hurricane Matthew in Florida.	Filter tweets using disaster-related keywords. Use ratio of disaster related tweets vs all tweets to indicate impacted area.	No – use geo-located tweets only.	Provide a method, using social media, which can aid in damage assessment and release of UAVs to affected areas during a hurricane.	Limited by use of only geo-located tweets.
104	Yuan & Liu (2018b) - Twitter - hurricane - USA	Test different machine learning algorithms to classify social media posts relating to the impacts of Hurricane Matthew. In particular to find damage-related social media posts	Tweets initially filtered based on location coordinates in Florida. Machine learning algorithms tested for classification of tweets: Naïve Bayes, support vector machines (SVM), decision tree.	No – use geo-located tweets only	Provides a method for classification of social media posts during a natural disaster. Find that Naïve Bayes model provides most reliable results.	Only classify posts relating to damage assessment - relies on large manually tagged training dataset. Limited by use of only geo-located tweets.
105	Yuan & Liu (2019) - Twitter - hurricane - USA	Builds on previous work which uses supervised machine learning approach to identify damage-related social media data. In addition, this study also looks at annual average sentiment as a baseline to calculate normalised sentiment, and compare findings with insurance claim data.	Tweets initially filtered based on location coordinates in Florida. Machine learning algorithms tested for classification of tweets: Naïve Bayes, support vector machines (SVM), decision tree.	No – use geo-located tweets only	Method has potential for use as a real-time damage assessment tool. Find decrease in normalised sentiment correlates with increase in insurance claim data.	Only classify posts relating to damage assessment - relies on large manually tagged training dataset. Sentiment analysis methods have varying levels of applicability to real world events. Limited by use of only geo-located tweets.
106	Yuan *et al.* (2020) - Twitter - hurricane - USA	Builds on previous work to incorporate social media users' post frequencies into analysis of tweet sentiment and tweet topics during Hurricane Matthew.	Weighted sentiment calculated using lexicon-based approach (AFINN) taking user post frequencies into account. Use LDA to identify topics of discussion from tweets.	N/A	Provides a more accurate method for calculating changes in public sentiment and topics of discussion during a hurricane event by weighting results to include user post frequencies.	Annual sentiment baseline calculated using one year of tweets only.

107	Yue *et al.* (2018) - Twitter - hurricane - USA	Find mappings between social media and severity level of a disaster using Twitter data during Hurricane Harvey and Irma.	Use Bag-of-Words and Word2Vec to prepare data for classification. Tweets filtered for relevance using 5 different machine learning classification algorithms.	No – use geo-located tweets only.	Provide a method for determining the severity of a hurricane event. Find a relation between impact related tweets and hurricane severity.	Method limited to use for specific hurricanes and areas of impact and based on small sample of Twitter data. Limited by use of only geo-located tweets.
108	Yum (2020) - Twitter - hurricane - USA	Examine the spatial patterns and sentiment of tweets during Hurricane Florence.	Filter tweets for relevance using hurricane-related keywords.	No – use geo-located tweets only.	Find tweet activity in affected and non-affected regions of a hurricane increases and importance of human sentiment as an indicator of disaster impact.	Limited by use of only geo-located tweets.

Table A.1 - Review and comparison of all papers relating to the social sensing of weather events which were assessed in this literature review

B. Supplementary tables for Using social media to measure impacts of named storm events in the UK and Ireland

		Number/% of tweets removed by each stage of filtering					
Tweet Collection	**All Tweets (unfiltered)**	**Timezone filter**	**Bot filter**	**Weather Station filter**	**Irrelevant Term filter**	**Bayesian filter**	**Tweets retained**
1. Precipitation	**67,448,047**	**62,305,788**	**292,884**	**545,828**	**5,454**	**1,033,519**	**3,264,573**
		92.38%	**0.43%**	**0.81%**	**0.01%**	**1.53%**	**4.84%**
2. Wind	**26,298,449**	**24,446,154**	**283,096**	**236,014**	**9,962**	**492,148**	**831,076**
		92.96%	**1.08%**	**0.90%**	**0.04%**	**1.87%**	**3.16%**
3. All Storm names	**8,101901**	**7,411,731**	**71,849**	**240**	**916**	**338,754**	**278,412**
		91.48%	**0.89%**	**0.00%**	**0.01%**	**4.18%**	**3.44%**
ophelia	897,054	614,034	8,034	71	127	60,058	214,730
		68.45%	0.90%	0.01%	0.01%	6.70%	23.94%
brian	2,037,045	1,908,520	24,726	55	358	90,417	12,970
		93.69%	1.21%	0.00%	0.02%	4.44%	0.64%
caroline	1,199,149	1,118,405	11,672	41	54	60,425	8,552
		93.27%	0.97%	0.00%	0.00%	5.04%	0.71%
dylan	2,504,264	2,427,862	8,673	2	112	63,709	3,907
		96.95%	0.35%	0.00%	0.00%	2.54%	0.16%
eleanor	555,433	514,949	5,465	37	91	23,020	11,872
		92.71%	0.98%	0.01%	0.02%	4.14%	2.14%
fionn	43,936	39,706	448	1	24	2,497	1,260
		90.37%	1.02%	0.00%	0.05%	5.68%	2.87%
georgina	104,327	97,268	789	12	17	5,347	894
		93.23%	0.76%	0.01%	0.02%	5.12%	0.86%
emma	760,693	690,988	12,043	21	133	33,281	24,227
		90.84%	1.58%	0.00%	0.02%	4.38%	3.18%

Table B.2 - Number and percentage of tweets removed by each stage of filtering for relevance

Tweet Collection	**Filtered for relevance AND located**	**Geo co-ordinates**	**Place co-ordinates**	**User location (co-ordinates)**	**User location (resolvable place name)**	**Place name mentioned in text**
1. Precipitation	**1,987,333**	**29,891**	**96,732**	**9,839**	**1,521,987**	**328,884**
		1.5%	**4.9%**	**0.5%**	**76.6%**	**16.5%**
2. Wind	**473,740**	**7,351**	**18,539**	**21,871**	**361,156**	**64,823**
		1.6%	**3.9%**	**4.6%**	**76.2%**	**13.7%**
3. All Storm names	**214,220**	**1349**	**7169**	**933**	**159207**	**45562**
		0.6%	**3.3%**	**0.4%**	**74.3%**	**21.3%**
ophelia	167,369	797	5,682	646	125,886	34,358
		0.5%	3.4%	0.4%	75.2%	20.5%
brian	9,439	175	476	42	6,744	2,002
		1.9%	5.0%	0.4%	71.4%	21.2%
caroline	4,993	52	178	36	4,017	710
		1.0%	3.6%	0.7%	80.5%	14.2%
dylan	2,410	17	84	10	1,958	341
		0.7%	3.5%	0.4%	81.2%	14.1%
eleanor	9,761	100	219	46	6,417	2,979
		1.0%	2.2%	0.5%	65.7%	30.5%
fionn	878	16	14	8	719	121
		1.8%	1.6%	0.9%	81.9%	13.8%
georgina	650	5	11	12	492	130
		0.8%	1.7%	1.8%	75.7%	20.0%
emma	18,720	187	505	133	12,974	4,921
		1.0%	2.7%	0.7%	69.3%	26.3%

Table B.3 - Results of location inference of filtered tweets. This shows the number and proportion of tweets located and which element of tweet metadata location is based on.

C. Documentation on the implementation of the code and data management

This document provides more detailed information about the social sensing code used in this thesis. The social sensing code was written in Python v.3.7[39].

C.1 Code implementation

The social sensing code used in this thesis is based on a number of processing stages as shown in Figure C.1. Each of these stages is described in more detail below. More detail on specific packages used in the social sensing code can be found in Sections C.2 and C.3.

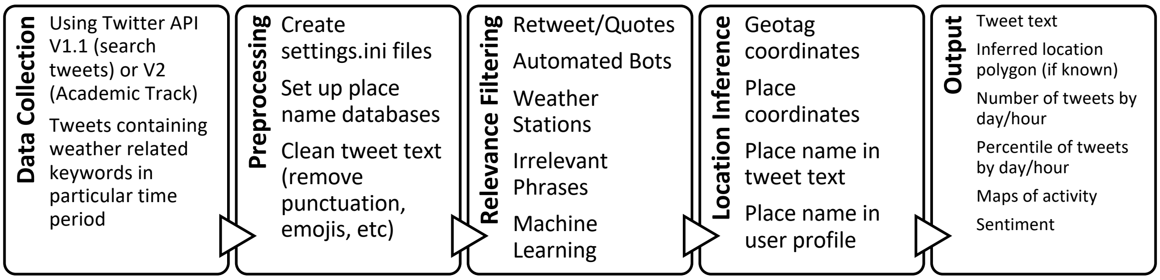

Figure C.1 - Flow diagram of Social Sensing process

C.1.1 Data Collection

Tweets are collected from the Twitter API using a Python script. There are different versions of the API, therefore each version required an alternative method to retrieve tweets.

In Chapters 4 and 5, the Twitter Standard v1.1[40] API Search Tweets functionality was used to stream tweets containing the relevant weather/hazard-related keywords in real time. The Streaming API allows up to 1% of all tweets at any point in time to be downloaded. Data was retrieved using the Twython package (see section C.2). The API provides tweet data in JSON format, which is then stored securely in a local database. With the Twitter v1.1 API, all available tweet fields are retrieved for each tweet. The full list of tweet fields retrieved using Twitter Standard API v1.1 can be found here: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet.

[39] https://www.python.org/downloads/release/python-370/ (accessed 12th July 2022)
[40] https://developer.twitter.com/en/docs/twitter-api/v1 (accessed: 12th July 2022)

In Chapter 6, the Twitter API v2 (Academic Research access)[41] was used. This allows academic researchers a full-archive search in which access to the entire database of tweets from when Twitter was first released in 2006 is made available. This enables tweets which contain relevant weather/hazard-related keywords for a particular time period to be downloaded retrospectively. With the Twitter v2 API, only the tweet ID and text is retrieved from the API by default, therefore other required fields must be specified when the request is made to the API. Table C.4 lists the fields chosen to be downloaded for the tweet dataset used in Chapter 6. Tweets were retrieved from the Twitter API v2 using the Python searchtweets-v2 package (see section C.2), specifying specific start and end dates, additional tweet fields to be downloaded, as well as the query string of keywords in tweets required.

Field	**Type**	**Description**	**Purpose**
Tweet Object			
id (default)	String	The unique identifier of the requested Tweet.	Uniquely identifies tweets – can be used to check for duplicates in the dataset, etc.
text (default)	String	The actual UTF-8 text of the Tweet.	Text can be used in relevance filtering and to search for place names in location inference
author_id	String	The unique identifier of the User who posted this Tweet.	Links tweet to user object
created_at	date (ISO 8601)	Creation time of the tweet	
geo	object	Contains details about the location tagged by the user in this Tweet, if they specified one	Provides coordinates of where tweet posted from (if provided) and link to place object with more information about location of tweet (if available)
lang	string	Language of the Tweet, if detected by Twitter. Returned as a BCP47 language tag.	Can be used for additional filtering for relevance (e.g. if just want English language tweets)
User object			
id (default)	string	The unique identifier of this user.	Provides link from Tweet object

[41] https://developer.twitter.com/en/products/twitter-api/academic-research (accessed 12th July 2022)

name (default)	string	The name of the user, as they've defined it on their profile. Not necessarily a person's name. Typically capped at 50 characters, but subject to change.	
username (default)	string	The Twitter screen name, handle, or alias that this user identifies themselves with. Usernames are unique but subject to change.	If user is a known bot, username can be used to identify as bot and filter from dataset
description	string	The text of this user's profile description (also known as bio) if the user provided one.	May provide information on location (not used)
location	string	The location specified in the user's profile if the user provided one. As this is a freeform value, it may not indicate a valid location, but it may be fuzzily evaluated when performing searches with location queries.	Searched for place names in location inference step.
Place object			
full_name (default)	string	A longer-form detailed place name.	Place name used to cross reference against place name databases
id (default)	string	The unique identifier of the expanded place, if this is a point of interest tagged in the Tweet.	Provides link from tweet object
country	string	The full-length name of the country this place belongs to.	Aids with location inference so only place names in this country are referenced.
country_code	string	The ISO Alpha-2 country code this place belongs to.	As above
geo	object	Contains place details in GeoJSON format.	Co-ordinates used for location inference.
name	string	The short name of this place.	Not used
place_type	string	Specified the particular type of information represented by this place information, such as a city name, or a point of interest.	May aid with location inference but not currently used.

Table C.4 - List of tweet fields chosen to be retrieved in JSON object when using Twitter API v2 (academic track) in Chapter 6.

C.1.2 Preprocessing

Prior to the filtering stage, the following is required to be set up:

- Settings.ini files which provides configuration settings to the social sensing code. This includes parameters such as the file location of databases and code, output prefixes, boolean statements to determine which stages in the filtering process should be applied, and geographical parameters (e.g. if location inference should be limited to a particular country).
- Creation of place name SQL tables using the external databases GADM, Geonames and DBpedia (see section C.3). These are used to retrieve point or polygon co-ordinates for identified placenames in tweet content.

C.1.3 Filtering for relevance

The next stage of the social sensing process requires processing each tweet through a number of routines to determine if it is relevant to the weather event/hazard. Tweets are read individually from the original JSON file line by line. Before each tweet is processed, emojis and punctuation are removed from the tweet text field. All text is also set to lowercase. Figure C.1 provides a simple flow chart to outline the relevance filter process in which tweets are removed from the dataset if they meet certain criteria. This is detailed as follows:

- *Retweet/quote* – this is an optional process – if a tweet is marked as a retweet and/or quote in the tweet metadata, and we have specified that these types of tweets should be removed in the configuration file, then they are excluded from the next step of filtering. If it is not a retweet/quote, it is passed to the next step of filtering.
- *Bot filter* – prior to running the relevance filtering process, the unfiltered dataset is checked for Twitter users which are either known automated accounts (bots) or if there are a disproportionate number of tweets from particular users (more than 1% of tweets). If a tweet in the dataset is from one of these users, it is excluded. If it is not, it is passed to the next step of filtering.
- *Weather station filter* – a table of terms which are common to tweets which are posted automatically by weather stations is referenced. The full list of terms excluded is detailed in Section C.4.1. If a tweet contains more than 2 of these terms, it is assumed to be a weather station and therefore excluded from the dataset. If it does not, it is passed to the next step of filtering.

- *Phrase filter* – the tweet text is checked for phrases which use the weather-related term but are not relevant to the weather event/hazard. E.g. “floods of tears”, “cook up a storm”, “wind up”. A full list of phrases which are excluded for the dataset in each chapter is detailed in Section C.4.2. If the tweet contains at least one of these terms, it is excluded from the dataset. If it does not, it is passed to the next step of filtering.
- *Machine learning filter* – training data is created using a random sample of tweet text from the dataset, after it has been passed through the above filters. Each tweet text in the training dataset is manually labelled as either 1 (relevant) or 0 (irrelevant). For example:

  [0 , 'So are journalists and news organisations guilty of getting caught up in the flood on social media and not taking a step back? #NX15']

  [1 , 'We have now cleared the flooding from Carrfield Rd.']

  At least 1000 example tweets are recommended in the training dataset for best performance. However, depending on the number of tweets being processed a much larger training sample is recommended for optimal results.

  Using this training data, a Multinomial Naïve Bayes classifier is used to determine if a tweet is relevant or irrelevant. If a tweet is marked irrelevant by the classifier, it is excluded from the dataset. If marked relevant, it is included in the dataset.

  More detail on machine learning step:

  As part of the classification process, a pre-processing step applies term frequency inverse document frequency (TF-IDF) to convert tweet text into a numerical input vector. TF-IDF highlights terms that are locally frequent but globally infrequent, increasing differentiation between vectors to improve classification. A Multinomial Naïve Bayes classifier with smoothing parameter $\alpha = 0.5$ is then applied (*the exact value of α is not crucial, the outputs from the social sensing code are very similar for a range of possible smoothing parameter values*).

  Each tweet dataset for different weather events/hazards will require a separate training dataset. The machine learning step must therefore be first validated using the training data to check that the classifier is not biased towards false negatives or false positives. Once a satisfactory accuracy is obtained using the training data it can then be applied in the social sensing process.

Please note: Multinomial Naïve Bayes was used because it yielded the best accuracy when compared with Support Vector Machine (SVM) and Logistic Regression when first tested with flood-related tweets (Arthur *et al.*, 2018)*. For consistency in applying the social sensing code to other weather events/hazards, Naïve Bayes continues to be used. The Naïve Bayes approach has been found to be a good approach for text classification due to its reduced complexity compared with other models and it not requiring to be continuously re-trained. It therefore has many advantages for the classification of short passages of text, such as tweets* (Tseng *et al.*, 2012)*.*

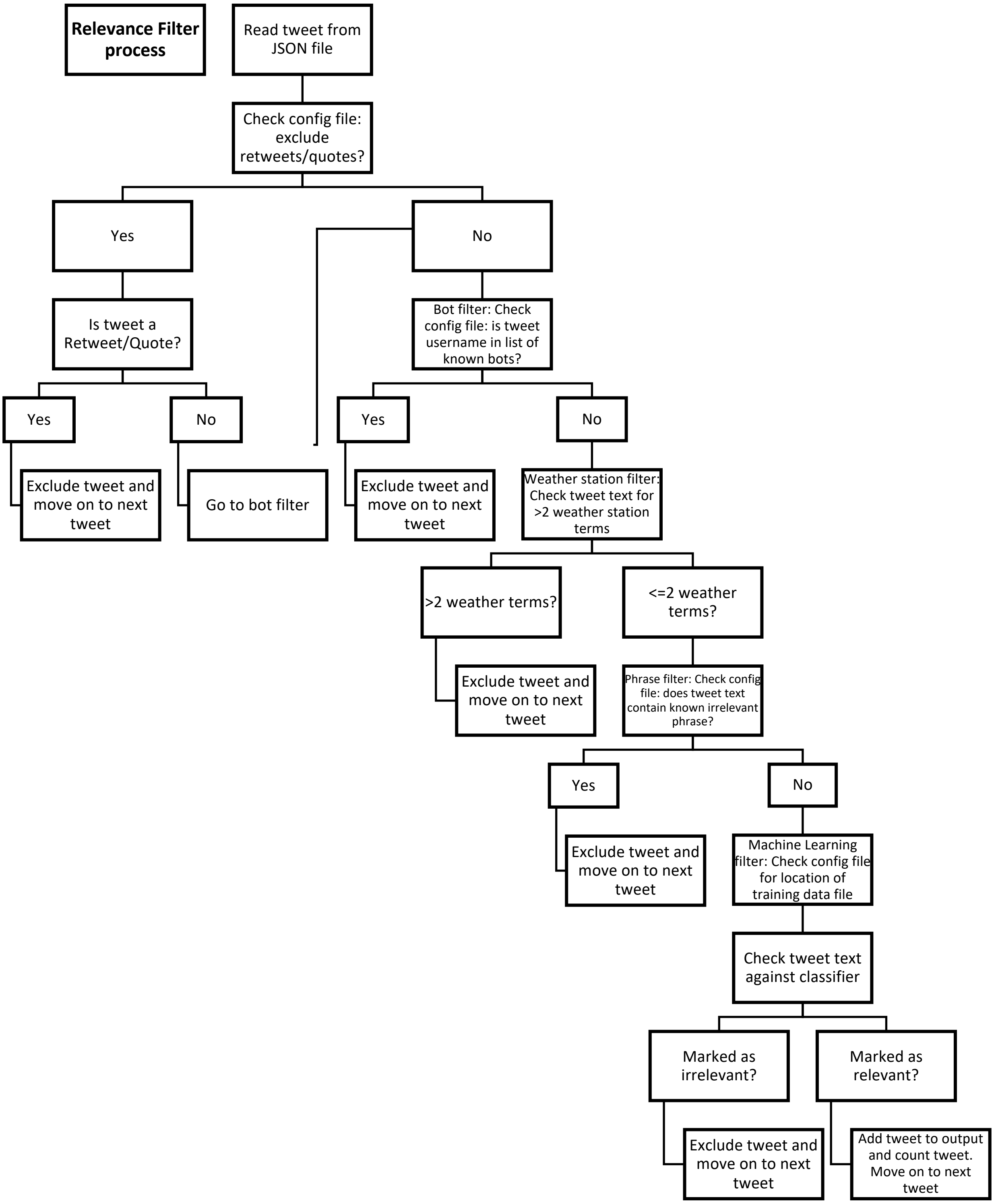

Figure C.2 - Flow chart of steps in the relevance filter code

If a tweet passes through all stages of the relevance filters it is counted and the tweet, including all tweet fields, is added to an output JSON file for each date in the dataset.

C.1.4 Location inference

If location of tweets is required, the next stage of the social sensing process requires checking each tweet for information which will determine the location to which it is referring to. Figure C.3 provides a flow chart of each step in the location inference process which is detailed as follows:

- *Geotag coordinates* – a small proportion (~1%) of tweets may contain the exact GPS coordinates from where it is posted from if a user has selected this option in their Twitter settings. A tweet is first checked to see if it contains these coordinates in the metadata. If it does, then these are used as the tweet location, otherwise the tweet is passed to the next location inference step.
- *Place coordinates* – a user may select a place when posting a tweet. If so, the point or polygon coordinates for this place will be included in the tweet metadata. The tweet is therefore next checked to see if it contains the place fields in the metadata. If it does, then the place coordinates are used as the tweet location, otherwise the tweet is passed to the next location inference step.
- *Place name in the tweet text* – if a user is posting about a particular weather event/hazard, then they may make mention of the specific place that this weather event relates to in the tweet text. Therefore, the tweet text is checked for placenames against first the Geonames database, and if no match is found here, it is then checked against the DBPedia database. If a match if found in the Geonames database and the place is marked as a ‘region’, ‘area’, ‘city’ or ‘state’, the location in the GADM database is accessed to obtain the polygon describing the location, otherwise we use the latitude and longitude provided by Geonames is used. If DBPedia is used and the place name is tagged as a place, then the GADM database is first checked for coordinates, if not found in GADM, then the latitude and longitude coordinates in the DBPedia database are used.

  If a matching place name and coordinates are found in this step, the coordinates found are used as the tweet location. Otherwise, the tweet is passed to the next and final location inference step.

Please note: If multiple placenames are found, then a number of sets of point and/or polygon coordinates will be returned. To determine the most likely location, each set of coordinates is checked for overlap and/or similarity. Those locations overlapping or geographically close to each other are given a greater weight and deemed the most likely location so used as the tweet location. If locations are geographically sparse (i.e. in different countries, then the placename with the largest population is used and assumed the most likely location).

If a placename is found which resides in multiple locations (e.g. Cambridge, UK, Cambridge, Massachusetts), then the tweet metadata is also checked for reference to a specific country – e.g. if user location field contains country, or timezone infers user's country location. If this information is not found, then the placename with the largest population is used and assumed the most likely location.

- *Place name in the user location* – the Twitter user profile includes a free text location field in which a user can put their home location. Some of these are irrelevant (e.g. Candyland, the Moon), however many users choose to put their home location in this field. While it is noted that a user's home location may not be the same as the location of the weather event/hazard taking place (e.g. if on holiday, travelling, commenting on an event elsewhere), it is assumed that in the majority of cases that a user will be in their home location when tweeting. The same process as described for 'Place name in the tweet text' above is followed to check the user location text field for place names. Coordinates returned are used as the tweet location. If no place name is found at this stage, then the tweet is assumed as not able to be located and removed from the dataset.

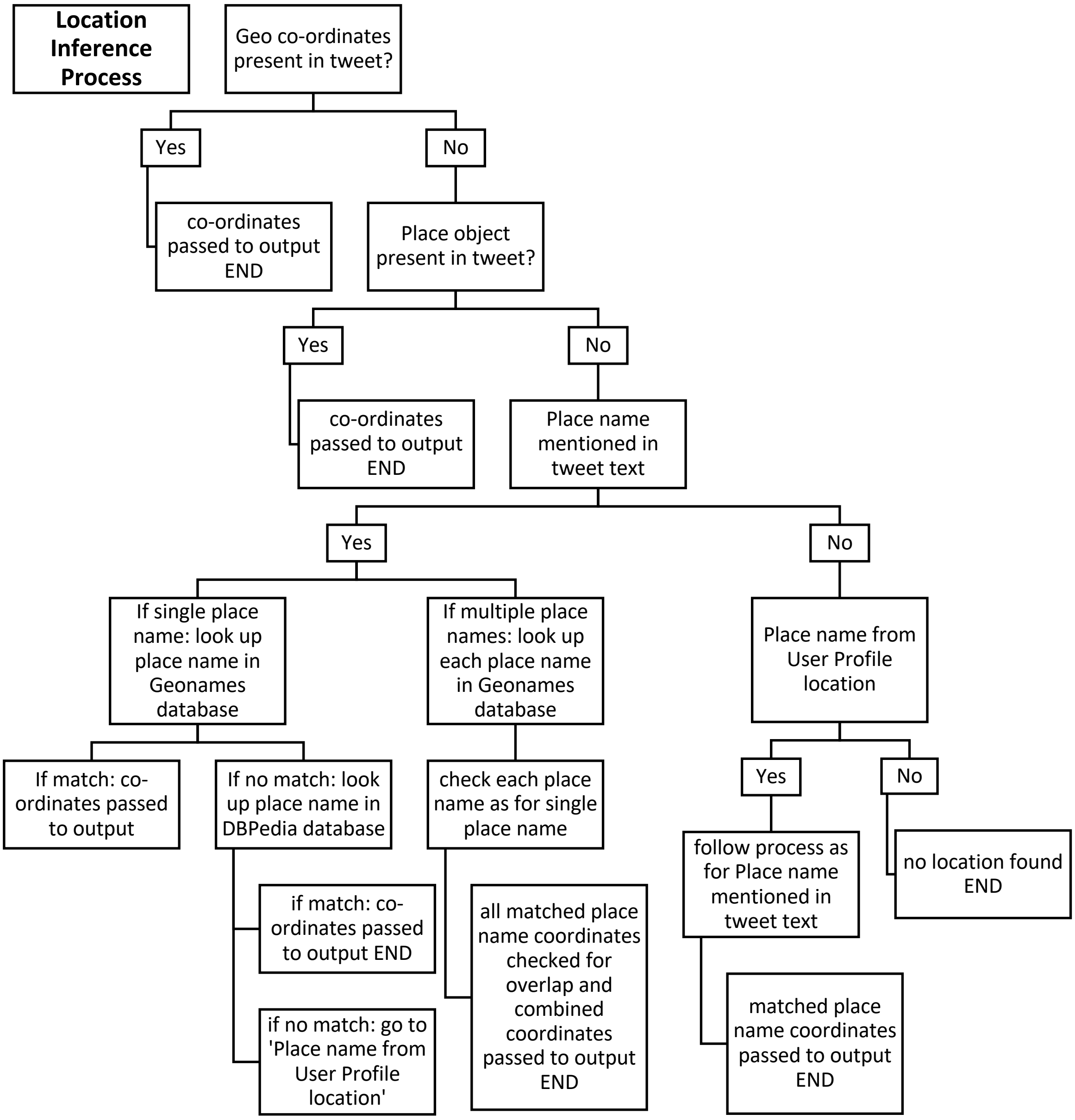

Figure C.3 - flow chart of steps in location inference code

Once all tweets have been checked for location information, remaining tweets are passed to an output JSON file containing the following fields: Tweet ID, Tweet text, Created date/time, Tweet location coordinates, type of location inferred (e.g. geotag, place, text, user location).

C.1.5 Output

This stage of the social sensing process uses the output JSON file after tweets have been filtered for relevance and an inferred location has been assigned. If location is not required, then it is possible to use the output after the relevance filtering stage. However, both the temporal and spatial tweet activity is likely to be required, therefore the notes which follow assume location information is used.

Tweet activity varies by location due to population and propensity for using Twitter in certain locations. Therefore, it is important in the analysis of tweets to account for this and not use tweet counts alone as a measure of activity. Depending on the spatial area required (e.g. country scale, level 1 administrative areas (e.g. state), level 2 administrative areas (e.g. county), etc), a file which provides the percentile of tweet activity for each time period (e.g. day) and spatial area is generated. The tweet counts over time for each spatial area are first calculated using a simple routine which counts each tweet in the JSON file into a dictionary with keys for both date and spatial area. These are then used to convert tweet counts to percentile. A higher percentile is likely to show above normal tweet activity for that area at that particular point in time. This then infers that that area is being affected by the weather event/hazard of interest at that particular time.

Percentile can then be used in timeseries plots to show peaks in tweet activity for a particular location, or to produce (heat)maps of tweet activity using Cartopy (see Section C.2) or other similar geospatial and mapping packages.

Additionally, the sentiment of tweets (i.e. how positive or negative the text used in the tweet is) can be used to infer that an area is being adversely affected by a weather event/hazard. The tweet text in the JSON output file can therefore be used to calculate a sentiment polarity score for each tweet (see Section C.2). This can then be aggregated and averaged both temporally and spatially to give an indication of when and where the most negative sentiment occurs during a weather event.

C.2 Python packages used

The following list provides a brief description of some of the main Python packages used in the social sensing code, along with version numbers and a link to further information.

Twython v.3.7 (https://twython.readthedocs.io/en/latest/index.html)
Used in the Data collection stage. A Python wrapper for the Twitter API. It provides a way to access Twitter data including authentication with the API, querying the API using required Twitter API endpoints, receiving the tweet data in JSON format and deserializing it into a Python dictionary. Tweets can then be stored in an appropriate database.

Searchtweets-v2 v1.1.1 (https://github.com/twitterdev/search-tweets-python/tree/v2)

Used in the Data collection stage. A Python wrapper for the Twitter API v2 which supports all search end-points including the 'all' endpoint (i.e. search the entire historical tweet archive). It can be used to authenticate with the Twitter API v2 and return tweets or counts of tweets. Tweets are returned in JSON format which can then be stored in an appropriate database.

Scikit_learn v.0.21 (https://scikit-learn.org/dev/index.html)

Used in the Relevance filter stage to apply the Multinomial Naïve Bayes classifier to tweet text to determine if relevant or not using a training dataset. This package was chosen as it is a simple and efficient tool for building machine learning methods in Python in only a few lines of code. It is based on NumPy, SciPy and Matplotlib, which are commonly used mathematical packages in Python.

Shapely v.1.7.1 (https://shapely.readthedocs.io/en/latest/)

Used in the Location inference stage. Provides functionality to manipulate and analyse geospatial objects. In the social sensing code it is used to manipulate points, polygons and determine overlaps where multiple locations are found within a tweet.

SQLite3 v. (https://docs.python.org/3.8/library/sqlite3.html)

Used in the Location inference stage to link place names identified in tweet text to the relevant SQL placename databases. This package was chosen due to its ease of being able to interact with and query SQL databases in Python code without having to install any external software.

Cartopy v.0.20 (https://scitools.org.uk/cartopy/docs/latest/index.html)

Used in the Output stage to produce geospatial visualisation of tweet activity. Cartopy is a package designed for geospatial data processing and uses Matplotlib for visualisation. It allows the straightforward creation of maps with different cartographic transformations. It works similarly to the Basemap[42] feature of Matplotlib, however is a much more powerful tool with additional features and

[42] https://matplotlib.org/basemap/ (accessed 12th July 2022)

it is easy to transform latitude and longitude coordinates for different geospatial transformations.

VADER Sentiment v3.3.2 (https://github.com/cjhutto/vaderSentiment)
Used in the Output stage in Chapter 6, VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, such as emojis, repetitive words, capitalisation and punctuations (e.g. exclamation marks). It is therefore important to use the original tweet text that does not have these features removed when calculating sentiment using this tool. It outputs a sentiment polarity score of -1 to 1 where -1 represents highly negative sentiment and 1 shows the highly positive sentiment.

TextBlob v0.15 (https://textblob.readthedocs.io/en/dev/)
Used in the Output stage in Chapter 4, TextBlob is a Python library for processing textual data. It is useful for common natural language processing tasks (NLP) including part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, etc. The sentiment analysis feature of TextBlob is used to calculate the sentiment of tweets in Chapter 4 and provides a sentiment polarity score of -1 to 1, as for Vader sentiment. TextBlob was chosen over Vader sentiment in Chapter 4 because it yielded very similar results when the outputs of both packages were compared. However, TextBlob is limited in that the classifier used to assign sentiment scores is trained on a dataset based on movie reviews. This means it may not provide the best results when used on social media posts related to the weather. Improvements to the Vader sentiment package since the work in Chapter 4 was undertaken, and the fact that Vader is specifically tuned to work well with social media content, means that this is now the preferred sentiment package for social sensing and was therefore used in Chapter 6 to calculate sentiment.

Other commonly installed Python packages required include: Re, Ast, Json, Datetime, String, Numpy and Matplotlib.

C.3 Other tools and libraries used

The following are used in the Pre-processing stage to create SQL databases of placenames and associated attributes including place type, latitude/longitude co-ordinates, population, etc.

Geonames (https://www.geonames.org/)

The GeoNames geographical database covers all countries and contains over 27 million geographical names that are available for download free of charge. It integrates a number of features, such as placenames in different languages, elevation and population. The database is user-editable, is composed from a number of sources and is the result of a project founded in 2005[43].

The Geonames dataset containing placenames for all countries is used to create the Geonames SQL database used in the social sensing code to check for place names in tweet content.

DBPedia (https://www.dbpedia.org/)

DBPedia is community project which extracts information created in Wikipedia, including placenames and associated attributes, such as latitude/longitude coordinates. The first publicly available dataset was made available in 2007 (Auer *et al.*, 2007) and includes individual place names including cities, states and places of interest. This dataset is used to create the DBPedia SQL database used in the social sensing code when no place name match can be found in the Geonames database.

GADM (https://gadm.org/)

GADM (the database of Global Administrative Areas) provides a freely available dataset containing geospatial information in the form of shapefiles for countries split by different levels of administrative area (e.g. country, state, county, city). Data is provided at a high spatial resolution which makes it good for providing polygon coordinates required for mapping. The whole world GADM dataset is used to create the SQL GADM database used to retrieve polygon coordinates for place name mentions found in tweet content.

[43] https://www.red-gate.com/simple-talk/opinion/geek-of-the-week/marc-wick-geek-of-the-week/ (accessed 12th July 2022)

C.4 Parameters used

C.4.1 Terms used in weather station filter

Most common weather station terms used in the weather station filter. If the tweet text contains more than 2 of these terms, it is assumed to be a tweet from an automated weather station and therefore excluded from the dataset.

keywords = ['Temp', 'Temperature', 'Wind', 'Gust', 'Rain', 'Barometer', 'Baro', 'Pressure', 'Hum', 'Humidity', 'Dew', 'Dewpoint']

C.4.2 Phrases excluded in Phrase filter

Chapter 4:

Storms

cook up a storm, perfect storm, storm in a teacup, port in a storm, calm before the storm, up a storm, down a storm, storm out, storm into, storm off, by storm, tempest, twitter storm, storm at, storm around

Wind

wind up, wind power, wind of change, broken wind, catch wind, second wind, like the wind, sheets in the wind, get wind, gone with the wind, wind in sails, wind in ones sails, ill wind, wind lies, near the wind, into the wind, wind through, wind back, wind round, wind around

Precipitation (flood, rain, snow, hail)

flood with, flood of, flood in, flood it, floods of, flooded with, flooded by, flooding back, immigrant, migrant, market, migration, tears, purple rain, rain check, right as rain, rain or shine, rain on, rain in, snow bunny, snow under, driven snow, snow stuff, yellow snow, hail a cab, give hail, hail down

Chapter 5:

Rainfall/Flood

flood with, flood of, flood in, flood it, floods of, flooded with, flooded by, flooding back, immigrant, migrant, market, migration, tears, purple rain, rain check, right as rain, rain or shine, rain on, rain in

Chapter 6:

Heatwave

take the heat off, heat of the, heat on, in heat, hot mess, hot spot, hot under, hot tin, hot water, hot potato, k-pop, kpop, korean pop, blackpink, bts, jungkook, suga, j-hope, jimin, bts, exo, xiumin, suho, baekhyun, chanyeol, sehun, kyungsoo, album, beyonce, playoff, nfl, mlb, wwe, team, pussy, dick, sex, dorito, grab it while its hot, deal, deals, bargain, bargains, sale, sales, shopping

C.5 Data management

Due to the size of the dataset downloaded from the Twitter API (tens of millions of tweets), and the fact that data contains a great deal of personal information, it was necessary to store data on a secure server at the University of Exeter with significant storage. To assist with file storage space, the JSON files containing all original tweet information were compressed to .zip files, which greatly reduces their volume due to the structured nature of the JSON files and repeated fields. All these files remained on the secure server at the University.

Once the tweet data in the original JSON files has been processed, output files were carefully created so as not to include any more data for each tweet than was absolutely necessary. The unique tweet ID was kept so that should further information for a tweet be required, it could be retrieved if necessary.

Processed files containing tweet information were kept on a local storage device while the researcher was doing the analysis work. Once completed, these files were compressed, added to a secure location on the University's server and removed from the researcher's local storage device and computer. Should other researchers request this data for use in their own research or to replicate the results found in Chapters 4 to 6, only a file containing tweet IDs will be shared in line with Twitter Developer Terms and Conditions[44].

To protect personal information contained within tweets and maintain user anonymity, all outputs such as timeseries, maps, etc show aggregated information only, so it is not possible to identify individual users. Additionally,

[44] https://developer.twitter.com/en/developer-terms (accessed 12th July 2022)

example tweets shown in Chapters 4-6 were amended versions of the original content and no usernames were shown.

Bibliography

Abdullah NA., Nishioka D., Tanaka Y., Murayama Y. 2017. Why I Retweet? Exploring User's Perspective on Decision-Making of Information Spreading during Disasters. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE Computer Society, 432–441.

Ahmad K., Pogorelov K., Riegler M., Ostroukhova O., Halvorsen P., Conci N., Dahyot R. 2019. Automatic detection of passable roads after floods in remote sensed and social media data. *Signal Process. Image Commun.* **74**:110–118.

Ahmed W. 2021.Using Twitter as a data source an overview of social media research tools (2021) . *Available at https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/18/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-2021/* (accessed December 23, 2021).

Aisha TS., Wok S., Manaf AMA., Ismail R. 2015. Exploring the Use of Social Media During the 2014 Flood in Malaysia. *Procedia - Soc. Behav. Sci.* **211**:931–937.

Alam F., Ofli F., Imran M. 2018a. CrisisMMD: Multimodal Twitter Datasets from Natural Disasters. In: *12th International AAAI Conference on Web and Social Media, ICWSM*. AAAI Press, 465–473.

Alam F., Ofli F., Imran M., Aupetit M. 2018b. A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria. In: *Proceedings of the International ISCRAM Conference*. Information Systems for Crisis Response and Management, ISCRAM, 553–572.

Alley RB., Emanuel KA., Zhang F. 2019. Advances in weather prediction. *Science (80-.).* **363**:342–344.

Anam A., Gangopadhyay A., Roy N. 2018. Evaluating disaster time-line from social media with wavelet analysis. In: *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*. Institute of Electrical and Electronics Engineers Inc., 41–48.

Anam A., Gangopadhyay A., Roy N. 2019. Identifying the context of hurricane posts on twitter using wavelet features. In: *Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019*. Institute of Electrical and Electronics Engineers Inc., 350–358.

Andrade SC de., Estrada CR., Delbem ACB., Mendiondo EM., Albuquerque JP de. 2017. Mining Rainfall Spatio-Temporal Patterns in Twitter: A Temporal Approach. In: Bregt A, Sarjakoski T, Lammeren Van R, Rip F eds. *Societal Geo-innovation. AGILE 2017. Lecture Notes in Geoinformation and Cartography.* Springer, Cham, 19–37.

Anzai S., Kazama SO. 2018. Social media analysis of people's high-risk responses to flood occurance. *WIT Trans. Built Environ.* **184**:167–175.

Arapostathis SG. 2019. Tweeting About Floods of Messinia (Greece, September 2016) - Towards a Credible Methodology for Disaster Management Purposes. In: *IFIP Advances in Information and Communication Technology*. Springer, 142–154.

Arapostathis SG. 2021. A Methodology for Automatic Acquisition of Flood-event Management Information From Social Media: the Flood in Messinia, South Greece, 2016. *Inf. Syst. Front.*:1–18.

Arapostathis SG., Karantzia M. 2019. Mapping Information of Fire Events, from VGI Source (Twitter), for Effective Disaster Management (in Greece); The Fire of North-East Attica, August 2017, (Greece) Case Study. In: El-Askary H, Lee S, Heggy E, Pradhan B eds. *Advances in Remote Sensing and Geo Informatics Applications. CAJG 2018. Advances in Science, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development).* Springer, Cham, 257–260.

Arapostathis SG., Lekkas E., Kalabokidis K., Xanthopoulos G., Drakatos G., Spirou N., Kalogeras I. 2018. Developing seismic intensity maps from twitter data; the case study of Lesvos, Greece 2017 earthquake: Assessments, improvements and enrichments on the methodology. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII-3-W4**:59–66.

Araujo T., van der Meer TGLA. 2018. News values on social media: Exploring what drives peaks in user activity about organizations on Twitter:

https://doi.org/10.1177/1464884918809299 **21**:633–651.

Arthur R., Boulton CA., Shotton H., Williams HTP. 2018. Social sensing of floods in the UK. *PLoS One* **13**:e0189327.

Arthur R., Williams HTP. 2018. The Human Geography of Twitter.

Arthur R., Williams HTP. 2019. The human geography of Twitter: Quantifying regional identity and inter-region communication in England and Wales. *PLoS One* **14**:e0214466.

Assis L., Herfort B., Steiger E., Horita FEA., De Albuquerque JP. 2015. A geographic approach for on-the-fly prioritization of social-media messages towards improving flood risk management. In: *Proc. 4th Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*. Recife, Brazil,.

Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R., Ives Z. 2007. DBpedia: A nucleus for a Web of open data. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **4825 LNCS**:722–735.

Austin BJ. 2014. Perspectives of weather and sensitivities to heat: social media applications for cultural climatology. Kent State University.

Azar B. 2010.Are your findings “WEIRD”? *Available at https://www.apa.org/monitor/2010/05/weird* (accessed July 31, 2022).

Bai H., Yu H., Yu G., Huang X. 2020. A novel emergency situation awareness machine learning approach to assess flood disaster risk based on Chinese Weibo. *Neural Comput. Appl. 2020*:1–16.

Barker JLP., Macleod CJA. 2019. Development of a national-scale real-time Twitter data mining pipeline for social geodata on the potential impacts of flooding on communities. *Environ. Model. Softw.* **115**:213–227.

Barredo JI. 2009. Normalised flood losses in Europe: 1970–2006. *Nat. Hazards Earth Syst. Sci.* **9**:97–104.

Barrett E. 2022.Why it’s so hard to get people to care about heat waves | Fortune. *Available at https://fortune.com/2022/07/19/heatwave-uk-europe-climate-change-risk-perception-psychology-deaths/* (accessed August 3, 2022).

Baylis P., Obradovich N., Kryvasheyeu Y., Chen H., Coviello L., Moro E., Cebrian M., Fowler JH. 2018. Weather impacts expressed sentiment. *PLoS One* **13**:e0195750.

Becker JS., Paton D., Johnston DM., Ronan KR., McClure J. 2017. The role of prior experience in informing and motivating earthquake preparedness. *Int. J. Disaster Risk Reduct.* **22**:179–193.

Bhuvana N., Arul Aram I. 2019. Facebook and Whatsapp as disaster management tools during the Chennai (India) floods of 2015. *Int. J. Disaster Risk Reduct.* **39**:101135.

Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S. 2009. DBpedia - A crystallization point for the Web of Data. *J. Web Semant.* **7**:154–165.

Blei DM., Ng AY., Jordan MI. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **3**:993–1022.

Bossu R., Fallou L., Landès M., Roussel F., Julien-Laferrière S., Roch J., Steed R. 2020. Rapid Public Information and Situational Awareness After the November 26, 2019, Albania Earthquake: Lessons Learned From the LastQuake System. *Front. Earth Sci.* **8**.

Boulton CA., Shotton H., Williams HTP. 2016. Using social media to detect and locate wildfires. In: *Tenth International AAAI Conference on Web and Social Media*. North America: AAAI,.

Branigan T. 2009. China blocks Twitter, Flickr and Hotmail ahead of Tiananmen anniversary. *Guard.*

Brouwer T., Eilander D., Van Loenen A., Booij MJ., Wijnberg KM., Verkade JS., Wagemaker J. 2017. Probabilistic flood extent estimates from social media flood observations. *Nat. Hazards Earth Syst. Sci.* **17**:735–747.

Brovelli MA., Zamboni G., Arias Muñoz C., Bonetti A., Huerta Guijarro J., Schade S., Granell Canut. 2014. Exploring twitter georeferenced data related to flood events: an initial approach. In: Huerta, Schade, Granell eds. *Connecting a Digital Europe through Location and Place. Proceedings of the AGILE'2014 International Conference on Geographic Information Science*. Castellón:

AGILE Digital Editions,.

de Bruijn JA., de Moel H., Jongman B., de Ruiter MC., Wagemaker J., Aerts JCJH. 2019. A global database of historic and real-time flood events based on social media. *Sci. Data* **6**:311.

de Bruijn JA., de Moel H., Jongman B., Wagemaker J., Aerts JCJH. 2018. TAGGS: Grouping Tweets to Improve Global Geoparsing for Disaster Response. *J. Geovisualization Spat. Anal.* **2**:1–14.

de Bruijn JA., de Moel H., Weerts AH., de Ruiter MC., Basar E., Eilander D., Aerts JCJH. 2020. Improving the classification of flood tweets with contextual hydrological information in a multimodal neural network. *Comput. Geosci.* **140**:104485.

Butgereit L. 2014. Crowdsourced weather reports: An implementation of the µ model for spotting weather information in Twitter. In: *2014 IST-Africa Conference Proceedings*. IEEE Computer Society, 1–9.

Campbell R., Beardsley D., Sezin T. 2018.Impact-based Forecasting and Warning: Weather Ready Nations | World Meteorological Organization. *Available at https://public.wmo.int/en/resources/bulletin/impact-based-forecasting-and-warning-weather-ready-nations* (accessed January 19, 2020).

Caragea C., Squicciarini A., Stehle S., Neppalli K., Tapia A. 2014. Mapping Moods: Geo-Mapped Sentiment Analysis During Hurricane Sandy.

Casteel MA. 2016. Communicating Increased Risk: An Empirical Investigation of the National Weather Service’s Impact-Based Warnings. *Weather. Clim. Soc.* **8**:219–232.

Cecinati F., Matthews T., Natarajan S., McCullen N., Coley D. 2019. Mining Social Media to Identify Heat Waves. *Int. J. Environ. Res. Public Heal. 2019, Vol. 16, Page 762* **16**:762.

Cerutti V., Fuchs G., Andrienko G., Andrienko N., Ostermann F. 2016. Identification of disaster-affected areas using exploratory visual analysis of georeferenced Tweets: application to a flood event. In: *19th AGILE Conference on Geographic Information Science*.

Cervone G., Sava E., Huang Q., Schnebele E., Harrison J., Waters N. 2015. Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *Int. J. Remote Sens.* **37**:100–124.

Changnon SA. 2003. Shifting Economic Impacts from Weather Extremes in the United States: A Result of Societal Changes, Not Global Warming. *Nat. Hazards 2003 292* **29**:273–290.

Chaudhary P., D'Aronco S., Moy De Vitry M., Leitão JP., Wegner JD. 2019. Flood-water level estimation from social media images. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Enschede, The Netherlands: Copernicus GmbH, 5–12.

Chen Y., Wang Q., Ji W. 2020. Assessing Disaster Impacts on Highways Using Social Media: Case Study of Hurricane Harvey. In: *Construction Research Congress 2020: Infrastructure Systems and Sustainability*. American Society of Civil Engineers, 562–571.

Chien Y-MC., Comber A., Carver S. 2017. Does Flickr work in disaster management?-a case study of Typhoon Morakot in Taiwan. In: *Proceedings of GIS Research UK 2017*.

Clement J. 2017. Number of social media users worldwide 2010-2021.

Clement J. 2020a.Twitter - Statistics & Facts | Statista. *Available at https://www.statista.com/topics/737/twitter/* (accessed March 19, 2020).

Clement J. 2020b.Instagram users worldwide 2023. *Available at https://www.statista.com/statistics/183585/instagram-number-of-global-users/* (accessed February 11, 2022).

Climate ADAPT. 2004.NatCatSERVICE Database. *Available at https://climate-adapt.eea.europa.eu/metadata/portals/natcatservice-database-year-of-launch* (accessed December 17, 2021).

climatecentre.org. 2020.IMPACT-BASED FORECASTING FOR EARLY ACTION THE FUTURE OF FORECASTS

Congjuico TS. 2015. Social Media for Flood Risk Management: A Study on the

Use of Facebook by the First Class Municipality of Cainta, Rizal, Philippines. In: *International Conference on Information Technology & Society*. Kuala Lumpur, Malaysia, 433–462.

Corneille M. 2020.Why should cities aim for heat resilience? *Available at https://accelerator.copernicus.eu/cities-heat-resilience/* (accessed January 17, 2022).

Cowie S., Arthur R., Williams HTP. 2018. @choo: Tracking Pollen and Hayfever in the UK Using Social Media. *Sensors* **18**:4434.

Crawford K., Finn M. 2015. The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal* **80**:491–502.

CRED. 2009.EM-DAT The International Disaster Database. *Available at https://www.emdat.be/about* (accessed December 20, 2021).

CRED. 2020.CRED Crunch 58 - Disaster Year in Review (2019). *Available at https://www.emdat.be/cred-crunch-58-disaster-year-review-2019-0* (accessed January 10, 2022).

Crisci A., Grasso V. 2013. Social media and severe weather events: mapping the impact footprint. In: *Social Media and Semantic technologies in Emergencies Response*.

Crompton RP., McAneney KJ. 2008. Normalised Australian insured losses from meteorological hazards: 1967–2006. *Environ. Sci. Policy* **11**:371–378.

Dalela PK., Sharma S., Kushwaha NK., Basu S., Majumdar S., Yadav A., Tyagi V. 2020. Classification of Disaster-Related Tweets Using Supervised Learning: A Case Study on Cyclonic Storm FANI. *Smart Innov. Syst. Technol.* **195**:169–178.

DataReportal. 2021.Digital 2021 October Global Statshot Report. *Available at https://datareportal.com/reports/digital-2021-october-global-statshot* (accessed December 13, 2021).

DBpedia. 2020.A Public Data Infrastructure for a Large, Multilingual, Semantic Knowledge Graph. *Available at https://wiki.dbpedia.org/* (accessed October

13, 2020).

Dean Z. 2021.Building a Future Where Cities are Resilient to Extreme Heat. *Available at https://www.redseasearch.com/building-a-future-where-cities-are-resilient-to-extreme-heat/* (accessed January 17, 2022).

Dean B. 2022.How Many People Use Telegram in 2022? 55 Telegram Stats. *Available at https://backlinko.com/telegram-users* (accessed March 17, 2022).

Demuth JL., Morss RE., Palen L., Anderson KM., Anderson J., Kogan M., Stowe K., Bica M., Lazrus H., Wilhelmi O., Henderson J. 2018. "Sometimes da #beachlife ain't always da wave": Understanding People's Evolving Hurricane Risk Communication, Risk Assessments, and Responses Using Twitter Narratives. *Weather. Clim. Soc.* **10**:537–560.

Devlin J., Chang M-W., Lee K., Google KT., Language AI. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proc. 2019 Conf. North*:4171–4186.

Digital Marketing Institute. 2021.Social Media: What Countries Use It Most & What Are They Using? *Available at https://digitalmarketinginstitute.com/blog/social-media-what-countries-use-it-most-and-what-are-they-using* (accessed December 14, 2021).

Dong H., Halem M., Zhou S. 2013. Social media data analytics applied to Hurricane Sandy. In: *2013 International Conference on Social Computing*. Alexandria, VA, USA: IEEE, 963–966.

Doocy S., Daniels A., Murray S., Kirsch TD. 2013. The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review. *PLOS Curr. Disasters*.

Dotzek N., Groenemeijer P., Feuerstein B., Holzer AM. 2008. Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD.

Dredze M., Paul MJ., Bergsma S., Tran H. 2013. Carmen: A Twitter Geolocation System with Applications to Public Health.

Eilander D., Trambauer P., Wagemaker J., Van Loenen A. 2016. Harvesting Social Media for Generation of Near Real-time Flood Maps. *Procedia Eng.* **154**:176–183.

Escobar MP., Vinogradova M., Demeritt D. 2016. *Using newspapers as a source of data to assess flood impacts: methodology note*.

Fang J., Hu J., Shi X., Zhao L. 2019. Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm. *Int. J. Disaster Risk Reduct.* **34**:275–282.

Farinosi F., Carrera L., Maziotis A., Mysiak J., Eboli F., Standardi G. 2012. Policy-Relevant Assessment Method of Socio-Economic Impacts of Floods: An Italian Case Study. *SSRN Electron. J.*

Farnaghi M., Ghaemi Z., Mansourian A. 2020. Dynamic Spatio-Temporal Tweet Mining for Event Detection: A Case Study of Hurricane Florence. *Int. J. Disaster Risk Sci. 2020 113* **11**:378–393.

Feng Y., Sester M. 2017. Social media as a rainfall indicator. In: Bregt A. ST. LR van, RF ed. *Societal Geo-Innovation : short papers, posters and poster abstracts of the 20th AGILE Conference on Geographic Information Science*. Wageningen, The Netherlands: AGILE,.

Fink S. 1986. *Crisis management: Planning for the inevitable*. New York: American Management Association.

Fitrianah D., Jatikusumo D., Nurhaida I. 2020. D-Loc Apps: A Location Detection Application Based on Social Media Platform in the Event of A Flood Disaster. In: *ACM International Conference Proceeding Series*. Association for Computing Machinery, 41–45.

FloodList. 2021.FloodList – Floods and flooding news from around the world. *Available at https://floodlist.com/* (accessed February 24, 2022).

Fohringer J., Dransch D., Kreibich H., Schröter K. 2015. Social media as an information source for rapid flood inundation mapping. *Nat. Hazards Earth Syst. Sci* **15**:2725–2738.

Foster C., Pennington CVL., Culshaw MG., Lawrie K. 2012. The national landslide database of Great Britain: Development, evolution and applications. *Environ. Earth Sci.* **66**:941–953.

Frondel M., Simora M., Sommer S. 2017. Risk Perception of Climate Change: Empirical Evidence for Germany. *Ecol. Econ.* **137**:173–183.

Galbraith KL. 2017. Terms and Conditions May Apply (But Have Little to Do With Ethics). *Am. J. Bioeth.* **17**:21–22.

García-Perdomo V., Salaverría R., Kilgo DK., Harlow S. 2017. To Share or Not to Share. *https://doi.org/10.1080/1461670X.2016.1265896* **19**:1180–1201.

Geonames. 2020. *GeoNames*.

Giuffrida L., Lokys H., Klemm O. 2020. Assessing the effect of weather on human outdoor perception using Twitter. *Int. J. Biometeorol.* **64**:205–216.

Global Change Data Lab. 2019.The rise of social media - Our World in Data. *Available at https://ourworldindata.org/rise-of-social-media* (accessed November 5, 2021).

Grace R. 2020a. Hyperlocal Toponym Usage in Storm-Related Social Media. In: Hughes AL, McNeill F, Zobel C eds. *Proceedings of the 17th ISCRAM Conference*. 849–859.

Grace R. 2020b. Crisis social media data labeled for storm-related information and toponym usage. *Data Br.* **30**:105595.

Grasso V., Crisci A., Morabito M., Nesi P., Pantaleo G. 2017. Public crowdsensing of heat waves by social media data. *Adv. Sci. Res.* **14**:217–226.

Gray BJ., Weal MJ., Martin D. 2016. Social Media and Disasters: Applying a New Conceptual Framework to the Case of Storm Desmond. *Int. J. Inf. Syst. Cris. Response Manag.* **8**:41–55.

Guan X., Chen C. 2014. Using social media data to understand and assess disasters. *Nat. Hazards* **74**:837–850.

Gunawan O., Aldridge T. 2018. *Disaster Loss Data Management in Scotland. Technical Report DMS/2018/01 for the Scottish National Centre for*

Resilience, Research Contractor: Health & Safety Executive.

Halse S., Montarnal A., Tapia A., Bénaben F. 2018. Bad Weather Coming: Linking social media and weather sensor data. In: *Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2018*. Rochester Institute of Technology, 507–515.

Hamoui B., Mars M., Almotairi K. 2020. FloDusTA: Saudi Tweets Dataset for Flood, Dust Storm, and Traffic Accident Events. In: *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. 1391–1396.

Han X., Wang J. 2019a. Using social media to mine and analyze public sentiment during a disaster: A case study of the 2018 Shouguang city flood in china. *ISPRS Int. J. Geo-Information*.

Han X., Wang J. 2019b. Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China. *ISPRS Int. J. Geo-Information 2019, Vol. 8, Page 185* **8**:185.

Hannak A., Anderson E., Feldman Barrett L., Lehmann S., Mislove A., Riedewald M. 2012. Tweetin' in the Rain: Exploring Societal-scale Effects of Weather on Mood. In: *Sixth International AAAI Conference on Weblogs and Social Media*. 479–482.

Harcup T., O'Neill D. 2001. What Is News? Galtung and Ruge revisited. *Journal. Stud.* **2**:261–280.

Harrison SE., Johnson PA. 2017. Crowdsourcing the Disaster Management Cycle. *Int. J. Inf. Syst. Cris. Response Manag.* **8**:17–40.

Harrison SE., Potter SH., Prasanna R., Doyle EEH., Johnston D. 2021. 'Where oh where is the data?': Identifying data sources for hydrometeorological impact forecasts and warnings in Aotearoa New Zealand. *Int. J. Disaster Risk Reduct.* **66**:102619.

Harrison SE., Potter SH., Prasanna R., Doyle EEH., Johnston D. 2022. Identifying the Impact-Related Data Uses and Gaps for Hydrometeorological Impact Forecasts and Warnings. *Weather. Clim. Soc.* **14**:155–176.

Havas C., Resch B., Francalanci C., Pernici B., Scalia G., Fernandez-Marquez JL., Van Achte T., Zeug G., Mondardini MRR., Grandoni D., Kirsch B., Kalas M., Lorini V., Rüping S. 2017. E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time. *Sensors 2017, Vol. 17, Page 2766* **17**:2766.

Heaviside C. 2020. Urban Heat Islands and Their Associated Impacts on Health. *Oxford Res. Encycl. Environ. Sci.*

Hemingway R., Robbins J. 2020. Developing a hazard-impact model to support impact-based forecasts and warnings: The Vehicle OverTurning (VOT) Model. *Meteorol. Appl.* **27**:e1819.

Herfort B., de Albuquerque JP., Schelhorn SJ., Zipf A. 2014a. Exploring the Geographical Relations Between Social Media and Flood Phenomena to Improve Situational Awareness. In: Huerta J, Schade S, Granell C eds. *Connecting a Digital Europe Through Location and Place*. Springer International Publishing, Cham, 55–71.

Herfort B., Albuquerque JD., Schelhorn S-J., Zipf A. 2014b. Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe Flood in June 2013. In: *Proceedings of the 11th International ISCRAM Conference*. 747–751.

Huang X., Wang C., Li Z. 2018. Reconstructing flood inundation probability by enhancing near real-time imagery with real-time gauges and tweets. *IEEE Trans. Geosci. Remote Sens.* **56**:4691–4701.

Huang X., Wang C., Li Z. 2019a. Linking picture with text: tagging flood relevant tweets for rapid flood inundation mapping. *Proc. ICA* **2**:45.

Huang X., Wang C., Li Z., Ning H. 2019b. A visual–textual fused approach to automated tagging of flood-related tweets during a flood event. *Int. J. Digit. Earth* **12**:1248–1264.

Hutto CJ., Gilbert E. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *ICWSM.*

Imran M., Elbassuoni S., Castillo C., Diaz F., Meier P. 2013. Practical Extraction of Disaster-Relevant Information from Social Media. In: *Proceedings of the*

22nd International Conference on World Wide Web. Rio de Janeiro, Brazil: Association for Computing Machinery, New York, NY, USA, 1021–1024.

Imran M., Mitra P., Castillo C. 2016. Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages. *Proc. 10th Int. Conf. Lang. Resour. Eval. Lr. 2016*:1638–1643.

Instagram. 2020.Instagram Developer Documentation. *Available at https://www.instagram.com/developer/* (accessed February 11, 2022).

IPCC. 2014. *Climate Change 2014: Impacts, Adaptation and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jain P. 2018.Social Media: Impact on human behavior and society. *Available at https://www.linkedin.com/pulse/social-media-impact-human-behavior-society-piyushi-jain* (accessed August 16, 2022).

Jitkajornwanich K., Kongthong C., Khongsoontornjaroen N., Kaiyasuan J., Lawawirojwong S., Srestasathiern P., Srisonphan S., Vateekul P. 2019. Utilizing Twitter Data for Early Flood Warning in Thailand. In: *2018 IEEE International Conference on Big Data (Big Data)*. Institute of Electrical and Electronics Engineers Inc., 5165–5169.

Johnston DM., Bebbington MS., Lai CD., Houghton BF., Paton D. 1999. Volcanic hazard perceptions: comparative shifts in knowledge and risk. *Disaster Prev. Manag.* **8**:118–126.

Jongman B., Wagemaker J., Romero BR., Perez EC De. 2015. Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals. *ISPRS Int. J. Geo-Information 2015* **4**:2246–2266.

Jony RI., Woodley A., Perrin Di. 2019. Flood Detection in Social Media Images using Visual Features and Metadata. In: *2019 Digital Image Computing: Techniques and Applications, (DICTA)*. Institute of Electrical and Electronics Engineers Inc., 1–8.

Jung J., Uejio CK. 2017. Social media responses to heat waves. *Int. J. Biometeorol.* **61**:1247–1260.

Kaltenberger R., Schaffhauser A., Staudinger M. 2020. “What the weather will do” – results of a survey on impact-oriented and impact-based warnings in European NMHSs. *Adv. Sci. Res.* **17**:29–38.

Kankanamge N., Yigitcanlar T., Goonetilleke A., Kamruzzaman M. 2020. Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. *Int. J. Disaster Risk Reduct.* **42**:101360.

Kapoor KK., Tamilmani K., Rana NP., Patil P., Dwivedi YK., Nerur S. 2017. Advances in Social Media Research: Past, Present and Future. *Inf. Syst. Front. 2017 203* **20**:531–558.

Katavoutas G., Founda D. 2019. Response of Urban Heat Stress to Heat Waves in Athens (1960–2017). *Atmos. 2019, Vol. 10, Page 483* **10**:483.

Kennedy H. 2016. Fair Game? User Evaluations of Social Media Data Mining. In: *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary*. London, UK: Palgrave Macmillan, 159–187.

Keramitsoglou I., Kiranoudis CT., Maiheu B., De Ridder K., Daglis IA., Manunta P., Paganini M. 2013. Heat wave hazard classification and risk assessment using artificial intelligence fuzzy logic. *Environ. Monit. Assess.* **185**:8239–8258.

Khaleq AA., Ra I. 2019. Cloud-Based Disaster Management as a Service: A Microservice Approach for Hurricane Twitter Data Analysis. In: *2018 IEEE Global Humanitarian Technology Conference (GHTC)*. Institute of Electrical and Electronics Engineers Inc., 1–8.

Kim J., Hastak M. 2018. Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manage.* **38**:86–96.

Knox JA., Mazanec B., Sullivan E., Hall S., Rackley JA. 2016. Analysis of the Twitter Response to Superstorm Sandy: Public Perceptions, Misconceptions, and Reconceptions of an Extreme Atmospheric Hazard. In: Coleman JSM ed. *Atmospheric Hazards - Case Studies in Modeling,* *Communication, and Societal Impacts*. IntechOpen,.

Koehrsen W. 2018.Beyond Accuracy: Precision and Recall | by Will Koehrsen | Towards Data Science. *Available at https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c* (accessed October 2, 2020).

Kong J., Zhao Y., Carmeliet J., Lei C. 2021. Urban heat island and its interaction with heatwaves: A review of studies on mesoscale. *Sustain.* **13**:10923.

Kox T., Lüder C., Gerhold L. 2018. Anticipation and Response: Emergency Services in Severe Weather Situations in Germany. *Int. J. Disaster Risk Sci.* **9**:116–128.

Kruspe A., Häberle M., Hoffmann EJ., Rode-Hasinger S., Abdulahhad K., Zhu XX. 2021. Changes in Twitter geolocations: Insights and suggestions for future usage. *arXiv*:212–221.

Kryvasheyeu Y., Chen H., Obradovich N., Moro E., Van Hentenryck P., Fowler J., Cebrian M. 2016. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2**:e1500779----e1500779.

Kwon HY., Kang YO. 2016. Risk analysis and visualization for detecting signs of flood disaster in Twitter. *Spat. Inf. Res.* **24**:127–139.

Lachlan KA., Spence PR., Lin X., Del Greco M. 2014. Screaming into the Wind: Examining the Volume and Content of Tweets Associated with Hurricane Sandy. *Commun. Stud.* **65**:500–518.

Lastoria B., Simonetti MR., Casaioli M., Mariani S., Monacelli G. 2006. Socio-economic impacts of major floods in Italy from 1951 to 2003. *Adv. Geosci.* **7**:223–229.

Leonelli S., Lovell R., Wheeler BW., Fleming L., Williams H. 2021. From FAIR data to fair data use: Methodological data fairness in health-related social media research: *https://doi.org/10.1177/20539517211010310* **8**.

Li Z., Huang Q., Emrich CT. 2019a. Introduction to social sensing and big data computing for disaster management. *Int. J. Digit. Earth* **12**:1198–1204.

Li H., Jadidi Z., Chen J., Jo J. 2019b. The Use of Machine Learning for Correlation

Analysis of Sentiment and Weather Data. In: Kim J-H, Myung H, Kim J, Xu W, Matson ET, Jung J-W, Choi H-L eds. *Robot Intelligence Technology and Applications 5. RiTA 2017. Advances in Intelligent Systems and Computing*. Springer International Publishing, Cham, 291–298.

Li Z., Wang C., Emrich CT., Guo D. 2018. A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartogr. Geogr. Inf. Sci.* **45**:97–110.

Li J., Wang X., Hovy E. 2014. What a Nasty day: Exploring Mood-Weather Relationship from Twitter.

Lin L., Ni M., He Q., Gao J., Sadek AW. 2015. Modeling the Impacts of Inclement Weather on Freeway Traffic Speed: Exploratory Study with Social Media Data. *Transp. Res. Rec. J. Transp. Res. Board* **2482**:82–89.

Liu X., Kar B., Ishino FAM., Zhang C., Williams F. 2020. Assessing the Reliability of Relevant Tweets and Validation Using Manual and Automatic Approaches for Flood Risk Communication. *ISPRS Int. J. Geo-Information 2020* **9**:532.

Liu Y., Liu X., Gao S., Gong L., Kang C., Zhi Y., Chi G., Shi L. 2015. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Ann. Assoc. Am. Geogr.* **105**:512–530.

Liu X., Zhang S., Wei F., Zhou M. 2011. *Recognizing Named Entities in Tweets*. Association for Computational Linguistics.

Llasat MC., Llasat-Botija M., López L. 2009. A press database on natural risks and its application in the study of floods in Northeastern Spain. *Nat. Hazards Earth Syst. Sci.* **9**:2049–2061.

Llasat MC., Llasat-Botija M., Prat MA., Porcú F., Price C., Mugnai A., Lagouvardos K., Kotroni V., Katsanos D., Michaelides S., Yair Y., Savvidou K., Nicolaides K. 2010. High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database. *Adv. Geosci.* **23**:47–55.

Loria S. 2010.TextBlob: Simplified Text Processing. *Available at https://textblob.readthedocs.io/en/dev/* (accessed March 1, 2018).

Lu H., Zhu Y., Shi K., Lv Y., Shi P., Niu Z. 2018. Using Adverse Weather Data in

Social Media to Assist with City-Level Traffic Situation Awareness and Alerting. *Appl. Sci.* **8**:1193.

Lwin KK., Zettsu K., Sugiura K. 2015. Geovisualization and correlation analysis between geotagged Twitter and JMA rainfall data: Case of heavy rain disaster in Hiroshima. In: *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*. Institute of Electrical and Electronics Engineers Inc., 71–76.

Ma G., Surakitbanharn C. 2019. Predicting Hurricane Damage Using Social Media Posts Coupled with Physical and Socio-Economic Variables. In: Franco Z, González JJ, Canós JH eds. *Proceedings of the 16th ISCRAM Conference*. València, Spain,.

Martín Y., Li Z., Cutter SL. 2017. Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. *PLoS One* **12**:e0181701.

McGrath R. 2013.Twython — Twython 3.6.0 documentation. *Available at https://twython.readthedocs.io/en/latest/* (accessed March 26, 2019).

McMahan AC., Grover CN., Vignola FE. 2013.Exceedance - an overview | ScienceDirect Topics. *Available at https://www.sciencedirect.com/topics/engineering/exceedance* (accessed April 4, 2019).

Mendes PN., Jakob M., Garcia-Silva A., Bizer C. 2011. DBpedia Spotlight: Shedding Light on the Web of Documents. In: *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics)*.

Mishra S., Suar D., Paton D. 2009. Is externality a mediator of experience-behaviour and information-action hypothesis in disaster preparedness? *J. Pacific Rim Psychol.* **3**:11–19.

Mocanu D., Baronchelli A., Perra N., Gonçalves B., Zhang Q., Vespignani A. 2013. The Twitter of Babel: Mapping World Languages through Microblogging Platforms. *PLoS One* **8**:61981.

Mons B., Neylon C., Velterop J., Dumontier M., Da Silva Santos LOB., Wilkinson MD. 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Inf. Serv. Use* **37**:49–56.

Morss RE., Demuth JL., Lazrus H., Palen L., Barton CM., Davis CA., Snyder C., Wilhelmi O V., Anderson KM., Ahijevych DA., Anderson J., Bica M., Fossell KR., Henderson J., Kogan M., Stowe K., Watts J., Morss RE., Demuth JL., Lazrus H., Palen L., Barton CM., Davis CA., Snyder C., Wilhelmi O V., Anderson KM., Ahijevych DA., Anderson J., Bica M., Fossell KR., Henderson J., Kogan M., Stowe K., Watts J. 2017. Hazardous Weather Prediction and Communication in the Modern Information Environment. *Bull. Am. Meteorol. Soc.* **98**:2653–2674.

Morstatter F., Pfeffer J., Liu H., Carley KM. 2013. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In: *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*. AAAI Press, 400–408.

Moumtzidou A., Andreadis S., Gialampoukidis I., Karakostas A., Vrochidis S., Kompatsiaris I. 2018. Flood Relevance Estimation from Visual and Textual Content in Social Media Streams. In: *Companion Proceedings of the The Web Conference 2018*. Lyon, France: International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1621–1627.

Muhammed S., Mathew SK. 2022. The disaster of misinformation: a review of research in social media. *Int. J. Data Sci. Anal.* **13**:271–285.

Mukkamala A., Beck R. 2016. Enhancing disaster management through social media analytics to develop situation awareness what can be learned from twitter messages about Hurricane Sandy? In: *PACIS 2016 Proceedings*. 165.

Munich RE. 2021.NATHAN. *Available at https://www.munichre.com/us-non-life/en/solutions/reinsurance/nathan.html* (accessed February 24, 2022).

Musaev A., Hou Q. 2017. Gathering High Quality Information on Landslides from Twitter by Relevance Ranking of Users and Tweets. In: *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*. Pittsburgh, PA, USA: Institute of Electrical and Electronics Engineers (IEEE), 276–284.

Naik N. 2016. Flooded streets-A crowdsourced sensing system for disaster response: A case study. In: *ISSE 2016 - 2016 International Symposium on Systems Engineering - Proceedings Papers*. Institute of Electrical and Electronics Engineers Inc., 1–3.

Nair MR., Ramya GR., Sivakumar PB. 2017. Usage and analysis of Twitter during 2015 Chennai flood towards disaster management. *Procedia Comput. Sci.* **115**:350–358.

Nguyen DT., Ofli F., Imran M., Mitra P. 2017. Damage assessment from social media imagery data during disasters. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. Sydney, Australia: Association for Computing Machinery, Inc, 569–576.

Niles MT., Emery BF., Reagan AJ., Dodds PS., Danforth CM. 2019. Social media usage patterns during natural hazards. *PLoS One* **14**:e0210484.

NOAA. 2021.Storm Events Database. *Available at https://www.ncdc.noaa.gov/stormevents/* (accessed February 24, 2022).

Oktafiani PM., Jariyah A., Fitri SR., Takako H. 2012. Social media analysis for Indonesian language: case study flood in Jakarta. *2012 Int. Conf. Adv. Comput. Sci. Inf. Syst.*:161–166.

Owuor I., Hochmair HH., Cvetojevic S. 2020. Tracking Hurricane Dorian in GDELT and Twitter. *Agil. GIScience Ser.* **1**:1–18.

Papagiannaki K., Lagouvardos K., Kotroni V. 2013. A database of high-impact weather events in Greece: a descriptive impact analysis for the period 2001-2011. *Nat. Hazards Earth Syst. Sci* **13**:727–736.

Parkhill KA., Henwood KL., Pidgeon NF., Simmons P. 2011. Laughing it off? Humour, affect and emotion work in communities living with nuclear risk1. *Br. J. Sociol.* **62**:324–346.

Pereira J., Monteiro J., Silva J., Estima J., Martins B. 2020. Assessing flood severity from crowdsourced social media photos with deep neural networks. *Multimed. Tools Appl.* **79**:26197–26223.

Pielke Jr. RA., Gratz J., Landsea CW., Collins D., Saunders MA., Musulin R. 2008. Normalized Hurricane Damage in the United States: 1900–2005. *Nat. Hazards Rev.* **9**:29–42.

Potter S., Harrison S., Kreft P. 2021. The Benefits and Challenges of Implementing Impact-Based Severe Weather Warning Systems: Perspectives of Weather, Flood, and Emergency Management Personnel. *Weather. Clim. Soc.* **13**:303–314.

Pourebrahim N., Sultana S., Edwards J., Gochanour A., Mohanty S. 2019. Understanding communication dynamics on Twitter during natural disasters: A case study of Hurricane Sandy. *Int. J. Disaster Risk Reduct.* **37**:101176.

Preis T., Moat HS., Bishop SR., Treleaven P., Stanley HE. 2013. Quantifying the Digital Traces of Hurricane Sandy on Flickr. *Sci. Rep.* **3**:1–3.

Pyrgou A., Santamouris M. 2018. Increasing Probability of Heat-Related Mortality in a Mediterranean City Due to Urban Warming. *Int. J. Environ. Res. Public Heal. 2018, Vol. 15, Page 1571* **15**:1571.

Raghavan S., Rajesh S. 2003. Trends in Tropical Cyclone Impact: A Study in Andhra Pradesh, India: A Study in Andhra Pradesh, India. *Bull. Am. Meteorol. Soc.* **84**:635–644.

Rainear AM., Lachlan KA., Oeldorf-Hirsch A., ... 2018a. Examining Twitter content of state emergency management during Hurricane Joaquin. *Commun. …*.

Rainear AM., Lachlan KA., Oeldorf-Hirsch A., DeVoss CL. 2018b. Examining twitter content of state emergency management during Hurricane Joaquin. *Commun. Res. Reports* **35**:325–334.

Ramamurthy P., González J., Ortiz L., Arend M., Moshary F. 2017. Impact of heatwave on a megacity: an observational analysis of New York City during July 2016. *Environ. Res. Lett.* **12**:054011.

Red Cross Red Crescent Climate Centre. 2020.European summer heatwaves the most lethal disaster of 2019, says international research group. *Available at https://reliefweb.int/report/world/european-summer-heatwaves-most-lethal-disaster-2019-says-international-research-group* (accessed January

10, 2022).

Řehůřek R., Sojka P. 2010. Software Framework for Topic Modelling with Large Corpora. In: *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*. Valletta, Malta: University of Malta, 46–50.

Resilient Cities Network. 2020.What is urban resilience? *Available at https://resilientcitiesnetwork.org/what-is-resilience/* (accessed January 24, 2022).

Restrepo-Estrada C., de Andrade SC., Abe N., Fava MC., Mendiondo EM., de Albuquerque JP. 2018. Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Comput. Geosci.* **111**:148–158.

Reuter C., Hughes AL., Kaufhold M-A. 2018. Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research. *Int. J. Human–Computer Interact.* **34**:280–294.

Reuter C., Kaufhold MA. 2018. Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics. *J. Contingencies Cris. Manag.* **26**:41–57.

Ritchie H., Roser M. 2018.Urbanization. *Available at https://ourworldindata.org/urbanization* (accessed January 28, 2022).

Robbins JC., Titley HA. 2018. Evaluating high-impact precipitation forecasts from the Met Office Global Hazard Map (GHM) using a global impact database. *Meteorol. Appl.* **25**:548–560.

Rodavia MRD., Cuison LT., Barcelo A. 2018. Detecting Flood Vulnerable Areas in Social Media Stream Using Association Rule Mining. In: *2018 International Conference on Platform Technology and Service (PlatCon)*. Institute of Electrical and Electronics Engineers Inc., 1–6.

Rosser JF., Leibovici • D G., Jackson • M J. 2017. Rapid flood inundation mapping using social media, remote sensing and topographic data. *Nat. Hazards* **87**:103–120.

Rossi C., Acerbo FS., Ylinen K., Juga I., Nurmi P., Bosca A., Tarasconi F., Cristoforetti M., Alikadic A. 2018. Early detection and information extraction for weather-induced floods using social media streams. *Int. J. Disaster Risk Reduct.* **30**:145–157.

Roy KC., Hasan S., Mozumder P. 2020. A multilabel classification approach to identify hurricane-induced infrastructure disruptions using social media data. *Comput. Civ. Infrastruct. Eng.* **35**:1387–1402.

Sakaki T., Okazaki M., Matsuo Y. 2010. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In: *Proceedings of the 19th International Conference on World Wide Web*. Raleigh, North Carolina, USA: Association for Computing Machinery, New York, NY, USA, 851–860.

Saravanou A., Valkanas G., Gunopulos D., Andrienko G. 2015. Twitter floods when it rains: a case study of the UK floods in early 2014. In: *Proceedings of the 24th International Conference on World Wide Web*. Association for Computing Machinery, Inc, 1233–1238.

Sato S. 2019. Analysis of Tweets Hashtagged “# Rescue” in the 2017 North Kyushu Heavy Rain Disaster in Japan. In: *2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. Institute of Electrical and Electronics Engineers Inc., 1–7.

Schulz A., Hadjakos A., Paulheim H., Nachtwey J., Uhlhäuser M. 2013. A Multi-Indicator Approach for Geolocalization of Tweets. In: *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013.* AAAI Press, 573–582.

Scotti V., Giannini M., Cioffi F. 2020. Enhanced flood mapping using synthetic aperture radar (SAR) images, hydraulic modelling, and social media: A case study of Hurricane Harvey (Houston, TX). *J. Flood Risk Manag.* **13**:e12647.

Shannag FB., Hammo BH. 2019. Lessons learned from event detection from Arabic tweets: the case of Jordan flash floods near dead sea. In: *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. Institute of Electrical and Electronics Engineers Inc., 806–811.

Shi Y., Sayama T., Takara K., Ohtake K. 2019. Detecting flood inundation information through Twitter: The 2015 Kinu River flood disaster in Japan. *J. Nat. Disaster Sci.* **40**:1–13.

Sit MA., Koylu C., Demir I. 2019. Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. *Int. J. Digit. Earth* **12**:1205–1229.

Smith L., Liang Q., James P., Lin W. 2017. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *J. Flood Risk Manag.* **10**:370–380.

Social Media Research Group. 2016. *Using social media for social research: An introduction*.

Sovacool BK., Xu X., Zarazua De Rubens G., Chen CF. 2020. Social media and disasters: human security, environmental racism, and crisis communication in Hurricane Irma response. *Environ. Sociol.* **6**:291–306.

Spasenovic K., Carrion D., Migliaccio F., Pernici B. 2019. Fast insight about the severity of hurricane impact with spatial analysis of Twitter posts. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII-4/W14**:221–225.

Spence PR., Lachlan KA., Lin X., del Greco M. 2015. Variability in Twitter Content Across the Stages of a Natural Disaster: Implications for Crisis Communication.

Spence PR., Lachlan KA., Rainear AM. 2016. Social media and crisis research: Data collection and directions. *Comput. Human Behav.* **54**:667–672.

Spruce MD., Arthur R., Robbins J., Williams HTP. 2021. Social sensing of high-impact rainfall events worldwide: A benchmark comparison against manually curated impact observations. *Nat. Hazards Earth Syst. Sci. Discuss.* **21**:2407–2425.

Spruce M., Arthur R., Williams HTP. 2020. Using social media to measure impacts of named storm events in the United Kingdom and Ireland. *Meteorol. Appl.* **27**:e1887.

Statista. 2021.Most used social media 2021. *Available at* *https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/* (accessed February 14, 2022).

Stewart MC., Wilson GB. 2016. The dynamic role of social media during Hurricane #Sandy: An introduction of the STREMII model to weather the storm of the crisis lifecycle. *Comput. Human Behav.* **54**:639–646.

Stowe K., Anderson J., Palmer M., Palen L., Anderson K. 2018. Improving Classification of Twitter Behavior During Hurricane Events. In: *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*. Melbourne, Australia: Association for Computational Linguistics, 67–75.

Strauss N., Painter J., Ettinger J., Doutreix MN., Wonneberger A., Walton P. 2021. Reporting on the 2019 European Heatwaves and Climate Change: Journalists' Attitudes, Motivations and Role Perceptions. *Journal. Pract.*

Sun D., Li S., Zheng W., Croitoru A., Stefanidis A., Goldberg M. 2016. Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery. *Int. J. Digit. Earth* **9**:427–441.

Taylor AL. 2018. Communicating high impact weather: Improving warnings and decision making processes. *Int. J. Disaster Risk Reduct.* **30**:1–4.

Taylor FE., Malamud BD., Freeborough K., Demeritt D. 2015. Enriching Great Britain's National Landslide Database by searching newspaper archives. *Geomorphology* **249**:52–68.

Tjong EF., Sang K., De Meulder F. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 142–147.

Tkachenko N., Jarvis S., Procter R. 2017. Predicting floods with Flickr tags. *PLoS One* **12**:e0172870.

Townsend L., Wallace C. 2016. *Social Media Research: A Guide to Ethics*.

Tse R., Zhang LF., Lei P., Pau G. 2017. Crowd sensing of weather conditions and traffic congestion based on data mining in social networks. *Lect. Notes*

Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST **195 LNICST**:353–361.

Tseng C., Patel N., Paranjape H., Lin TY., Teoh S. 2012. Classifying twitter data with Naïve bayes classifier. *Proc. - 2012 IEEE Int. Conf. Granul. Comput. GrC 2012*:294–299.

UK Research Integrity. 2018. GDPR and Research-An Overview for Researchers What is GDPR?

UK Snow Map 2010. *Available at https://uksnowmap.com/* (accessed March 19, 2019).

Ukkusuri S V., Zhan X., Sadri AM., Ye Q. 2014. Use of Social Media Data to Explore Crisis Informatics: Study of 2013 Oklahoma Tornado. *Transp. Res. Rec. J. Transp. Res. Board* **2459**:110–118.

UNDRR. 2013.The TEN Essentials for Making Cities Resilient. *Available at https://www.unisdr.org/campaign/resilientcities/toolkit/article/the-ten-essentials-for-making-cities-resilient.html* (accessed February 4, 2022).

UNISDR and CRED. 2015. *The Human Cost of Weather Related Disasters 1995-2015*. Geneva.

United Nations. 2022.World Population Prospects. *Available at https://population.un.org/wpp/DataQuery/* (accessed January 30, 2022).

de Vasconcelos LEG., dos Santos ECM., Neto MLF., Ferreira NJ., de Vasconcelos LG. 2016. Using Tweets for Rainfall Monitoring. *Adv. Intell. Syst. Comput.* **448**:1157–1167.

Vautard R., Van Aalst M., Boucher O., Drouin A., Haustein K., Kreienkamp F., Van Oldenborgh GJ., Otto FEL., Ribes A., Robin Y., Schneider M., Soubeyroux JM., Stott P., Seneviratne SI., Vogel MM., Wehner M. 2020. Human contribution to the record-breaking June and July 2019 heatwaves in Western Europe. *Environ. Res. Lett.* **15**:094077.

Vayansky I., Kumar SAP., Li Z. 2019. An Evaluation of Geotagged Twitter Data during Hurricane Irma Using Sentiment Analysis and Topic Modeling for Disaster Resilience. In: *2019 IEEE International Symposium on Technology and Society (ISTAS)*. Institute of Electrical and Electronics Engineers Inc., 1–6.

Victorian Auditor-General. 2014. *Heatwave Management: Reducing the Risk to Public Health*.

Vieweg S., Castillo C., Imran M. 2014. Integrating social media communications into the rapid assessment of sudden onset disasters. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, 444–461.

Vieweg S., Hughes AL., Starbird K., Palen L. 2010. Microblogging during Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Atlanta, Georgia, USA: Association for Computing Machinery, 1079–1088.

Wang D., Abdelzaher T., Kaplan L. 2015. *Social Sensing: Building Reliable Systems on Unreliable Data*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

Wang RQ., Hu Y., Zhou Z., Yang K. 2020a. Tracking Flooding Phase Transitions and Establishing a Passive Hotline with AI-Enabled Social Media Data. *IEEE Access* **8**:103395–103404.

Wang D., Kaplan L., Le H., Abdelzaher T. 2012. On truth discovery in social sensing: A maximum likelihood estimation approach. In: *IPSN'12 - Proceedings of the 11th International Conference on Information Processing in Sensor Networks*. New York, New York, USA: ACM Press, 233–244.

Wang B., Loo BPY., Zhen F., Xi G. 2020b. Urban resilience from the lens of social media data: Responses to urban flooding in Nanjing, China. *Cities* **106**:102884.

Wang R-Q., Mao H., Wang Y., Rae C., Shaw W. 2018. Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data. *Comput. Geosci.* **111**:139–147.

Wang D., Szymanski BK., Abdelzaher T., Ji H., Kaplan L. 2019. The age of social sensing. *Computer (Long. Beach. Calif).* **52**:36–45.

Wang Y., Wang T., Ye X., Zhu J., Lee J. 2016. Using Social Media for Emergency Response and Urban Sustainability: A Case Study of the 2012 Beijing Rainstorm. *Sustainability* **8**:25.

Wang Z., Ye X. 2018a. Social media analytics for natural disaster management. *Int. J. Geogr. Inf. Sci.* **32**:49–72.

Wang Z., Ye X. 2018b. Space, time, and situational awareness in natural hazards: a case study of Hurricane Sandy with social media data. *Cartogr. Geogr. Inf. Sci.* **46**:334–346.

Wani S., Yadav D., Verma OP. 2020. Development of Disaster Management and Awareness System Using Twitter Analysis: A Case Study of 2018 Kerala Floods. *Adv. Intell. Syst. Comput.* **1053**:1165–1174.

Ward K., Lauf S., Kleinschmit B., Endlicher W. 2016. Heat waves and urban heat islands in Europe: A review of relevant drivers. *Sci. Total Environ.* **569–570**:527–539.

Watson H., Finn RL. 2014. Social media and the 2013 UK heat wave: Opportunities and challenges for future events. In: Hiltz SR, Pfaff MS, Plotnick L, Shih PC eds. *11th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2014)*. University Park, Pennsylvania, 757–761.

Weaver IS., Williams HTP., Arthur R. 2021. A social Beaufort scale to detect high winds using language in social media posts. *Sci. Rep.* **11**:3647.

Weller K. 2015. Accepting the challenges of social media research. *Online Inf. Rev.* **39**:281–289.

Williams SA., Terras MM., Warwick C. 2013. What do people study when they study Twitter? Classifying Twitter related academic papers. *J. Doc.* **69**:384–410.

WMO. 2015. *WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services*. Geneva, Switzerland.

WMO WMO. 2021a. *WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services (WMO-No. 1150), Part II: Putting Multi-Hazard IBFWS* *into Practice*. Geneva: WMO.

WMO. 2021b.Weather-related disasters increase over past 50 years, causing more damage but fewer deaths. *Available at https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer* (accessed July 29, 2022).

WOW. Met Office Weather Observations Website 2019. *Available at https://wow.metoffice.gov.uk/* (accessed March 19, 2019).

Wu D., Cui Y. 2018. Disaster early warning and damage assessment analysis using social media data and geo-location information. *Decis. Support Syst.* **111**:48–59.

Wu W., Li J., He Z., Ye X., Zhang J., Cao X., Qu H. 2020. Tracking spatio-temporal variation of geo-tagged topics with social media in China: A case study of 2016 hefei rainstorm. *Int. J. Disaster Risk Reduct.* **50**:101737.

Xiao Y., Watson M. 2017. Guidance on Conducting a Systematic Literature Review: *https://doi.org/10.1177/0739456X17723971* **39**:93–112.

Xin EZ., Murthy D., Lakuduva NS., Stephens KK. 2019. Assessing the Stability of Tweet Corpora for Hurricane Events Over Time: A Mixed Methods Approach. In: *Proceedings of the 10th International Conference on Social Media and Society*. Toronto, ON, Canada: Association for Computing Machinery, New York, NY, USA, 59–66.

Xu P., Wang L., Liu Y., Chen W., Huang P. 2020. The record-breaking heat wave of June 2019 in Central Europe. *Atmos. Sci. Lett.* **21**:e964.

Yamada S., Utsu K., Uchida O. 2019. An Analysis of Tweets Posted during 2018 Western Japan Heavy Rain Disaster. In: *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. Institute of Electrical and Electronics Engineers Inc., 1–8.

Yang T., Xie J., Li G., Mou N., Chen C., Zhao J., Liu Z., Lin Z. 2020. Traffic Impact Area Detection and Spatiotemporal Influence Assessment for Disaster Reduction Based on Social Media: A Case Study of the 2018 Beijing Rainstorm. *ISPRS Int. J. Geo-Information* **9**:136.

Yang T., Xie J., Li G., Mou N., Li Z., Tian C., Zhao J. 2019a. Social Media Big Data Mining and Spatio-Temporal Analysis on Public Emotions for Disaster Mitigation. *ISPRS Int. J. Geo-Information 2019, Vol. 8, Page 29* **8**:29.

Yang J., Yu M., Qin H., Lu M., Yang C. 2019b. A Twitter Data Credibility Framework—Hurricane Harvey as a Use Case. *ISPRS Int. J. Geo-Information* **8**:111.

Yao F., Wang Y. 2020. Domain-specific sentiment analysis for tweets during hurricanes (DSSA-H): A domain-adversarial neural-network-based approach. *Comput. Environ. Urban Syst.* **83**:101522.

Young JC., Arthur R., Spruce M., Williams HTP. 2021. Social Sensing of Heatwaves. *Sensors 2021, Vol. 21, Page 3717* **21**:3717.

Young J., Arthur R., Spruce M., Williams HTP. 2022. Social Sensing Kerala Floods. *Int. J. Disaster Risk Reduct.* **Submitted**.

Yuan F., Li M., Liu R. 2020. Understanding the evolutions of public responses using social media: Hurricane Matthew case study. *Int. J. Disaster Risk Reduct.* **51**:101798.

Yuan F., Liu R. 2018a. Integration of Social Media and Unmanned Aerial Vehicles (UAVs) for Rapid Damage Assessment in Hurricane Matthew. In: *Construction Research Congress 2018: Safety and Disaster Management*. New Orleans, Louisiana: American Society of Civil Engineers, 513–523.

Yuan F., Liu R. 2018b. Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study. *Int. J. Disaster Risk Reduct.* **28**:758–767.

Yuan F., Liu R. 2019. Identifying Damage-Related Social Media Data during Hurricane Matthew: A Machine Learning Approach. In: *ASCE International Conference on Computing in Civil Engineering 2019*. Atlanta, Georgia: American Society of Civil Engineers, 207–214.

Yue S., Kondari J., Musaev A., Smith RK. 2018. Using Twitter Data to Determine Hurricane Category: An Experiment Using AlgoTraffic Data to Improve Traffic Incident Management View project Computer Security View project. In: Boersma K, Tomaszewski B eds. *Proceedings of the 15th ISCRAM*

Conference. Rochester, NY, USA,.

Yum S. 2020. Mining Twitter Data to Understand the Human Sentiment on Hurricane Florence. *J. Disaster Emerg. Res.* **3**:74–86.

Zhang C., Fan C., Yao W., Hu X., Mostafavi A. 2019. Social media for intelligent public information and warning in disasters: An interdisciplinary review. *Int. J. Inf. Manage.* **49**:190–207.

Zhang M-L., Zhou Z-H. 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.* **40**:2038–2048.

Zhao L., Oppenheimer M., Zhu Q., Baldwin JW., Ebi KL., Bou-Zeid E., Guan K., Liu X. 2018. Interactions between urban heat islands and heat waves. *Environ. Res. Lett.* **13**:034003.

Zou L., Lam NSN., Cai H., Qiang Y. 2018. Mining Twitter Data for Improved Understanding of Disaster Resilience. *Ann. Am. Assoc. Geogr.* **108**:1422–1441.