



Social weather: A review of crowdsourcing-assisted meteorological knowledge services through social cyberspace

Yifan Zhu¹ | Sifan Zhang¹ | Yinan Li¹ | Hao Lu^{1,2} | Kaize Shi¹ | Zhendong Niu^{1,3,4}

¹School of Computer Science & Technology, Beijing Institute of Technology, Beijing, China

²The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

³Library, Beijing Institute of Technology, Beijing, China

⁴School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA

Correspondence

Zhendong Niu, School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China.
Email: zniu@bit.edu.cn

Funding information

China's National Strategic Basic Research Program (973 Program), Grant/Award Number: 2012CB720700; Public Weather Service Center of China Meteorological Administration; National Natural Science Foundation of China, Grant/Award Number: 61370137; Ministry of Education-China Mobile Research Foundation Project, Grant/Award Number: 2016/2-7; Postgraduate Education Research Project of Beijing Institute of Technology, Grant/Award Number: 2017JYYJG-004

Abstract

Crowdsourcing has significantly motivated the development of meteorological services. Starting from the beginning of 2010s and highly motivating after 2014, crowdsourcing-driven meteorological services have evolved from a single collection and observation of data to the systematic acquisition, analysis and application of these data. In this review, by focusing on papers and databases that have combined crowdsourcing methods to promote or implement meteorological knowledge services, we analysed the relevant literature in three dimensions: data collection, information analysis and meteorological knowledge applications. First, we selected the potential data sources for crowdsourcing and discussed the characteristics of the collected data in four dimensions: consciousness, objectiveness, mobility and multidisciplinary. Second, based on the purpose of these studies and the extent of utilizing data as well as knowledge, we categorize the crowdsourcing-based meteorological analysis into three levels: relationship discovery, knowledge generalization and systemized service. Third, according to the application scenario, we discussed the applications that have already been put into use, and we suggest current challenges and future research directions. These previous studies show that the use of crowdsourcing in social space can expand the coverage as well as enhance the performance of meteorological

Dataset

Identifier: <https://drive.google.com/open?xml:id=179C9ADzH-UeIEmYPeHyOOGZxxTb1VATY>

Creator: [Y. Zhu, S. Zhang, Y. Li, H. Lu, K. Shi and Z. Niu]

Title: Studies on Crowdsourcing-Assisted Meteorological Knowledge Services (2011-2018)

Publisher: Institute of Software Intelligence and Software Engineering, Beijing Institute of Technology

Publication year: 2019

Resource type: Metadata document

Version: 1.0

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

service. It was also evident that current researches are contributing towards a systemic and intelligent knowledge service to establish a better bridge among academic, industrial and individual community.

KEYWORDS

crowdsourcing, data-driven, knowledge services, meteorological services, social space

1 | INTRODUCTION

Weather not only directly affects the physical activities of human but can influence human society in an imperceptible way (Adger *et al.*, 2013). The mechanisms behind different types of weather are very complicated and almost beyond the control of human beings. Thus, the observation, forecasting as well as alerting and responding to these weather situations (particularly severe weather) play significant roles in meteorological services (Schafer, 2012; Veltri & Atanasova, 2017). To help measure and predict weather and climate change, vast amounts of manpower and funds have been invested into meteorological and other related research areas, which has brought meteorological research into an era of ‘big data’ (Greengard, 2014). A large number of meteorological satellites have been launched and densely concentrated stations have also been established to collect meteorological data. Powerful computational resources, such as high-performance supercomputers, are utilized to calculate sophisticated meteorological models (Spiegelhalter *et al.*, 2011). These meteorological data can then be processed and utilized by meteorological and emergency planning departments.

Recently, the emergence of meteorological and climatic researches that utilizes social and public power has brought a new perspective towards more intelligent meteorological knowledge services. In other words, the original client and the information receptor (i.e. the public) have also become the information providers and research promoters. The rich amount of information perceived by them with their equipped smart devices has gradually formed a space containing these social signals. With the rapid advances in social and mobile computing, the Internet of Things (IoT) and big data analysis, this space has been acknowledged as a new dimension of the real world in addition to existing physical spaces, which has been named the ‘social space’ (Zhang *et al.*, 2018a). In this social space, people are not only regarded as message receivers but also as perceivers and producers of weather information. In other words, every person is considered to be a ‘social sensor’ that together can perceive and sense the real world, harmonize tasks and data sources: the greater the public focus is, the more data that could be extracted (Zheng *et al.*, 2016). For

instance, people can express their opinions and discuss weather freely on social network platforms such as Twitter, Facebook and Sina Weibo, and subsequently, data collectors can integrate the multidomain information through social platforms, collect and analyse subjective responses, and finally generate knowledge by using data mining and knowledge engineering technologies while considering social significance (Wang, 2014). This can allow timely acquisition of feedback and has helped to guide policy enactment, decision-making, emergency responses and public relations by government departments (Goncalves *et al.*, 2014; Xiong *et al.*, 2015). However, extracting information that current research can benefit from this social space that is filled with unstructured and complex data requires a unique set of methods for data collection, processing, analysis and application.

From the perspective of a practical investigation process on the social space, data and processing work can be evenly distributed between a large, relatively open and often rapidly evolving group of online users, whether they are aware or not. This method, known as crowdsourcing, greatly liberates economic and human constraints and provides a convenient bridge for low-cost, fast, dense, and diverse data collection and processing (Prpić *et al.*, 2015; Taeihagh, 2017).

Specifically concerning the application of the social space and crowdsourcing into meteorological service, Muller *et al.* have reported on the progress of crowdsourcing in atmospheric science based on the situation as it was before 2014 (Muller *et al.*, 2015). However, compared with the initial period, relevant research and industrial applications have since made great progress over the following years. As a typical feature, the current connotation of meteorological knowledge service promoted by crowdsourcing has been derived from simple data collection to systematic information collection, fusion, analysis and application. Compared with traditional meteorological knowledge service, Figure 1 shows the differences in focus and coverage between conventional approach and this novel approach. In other words, the two approaches exactly represent top-bottom and bottom-up ideas. Based on this, we believe that this novel way of collecting, analysing, and applying this social space-based information provides a

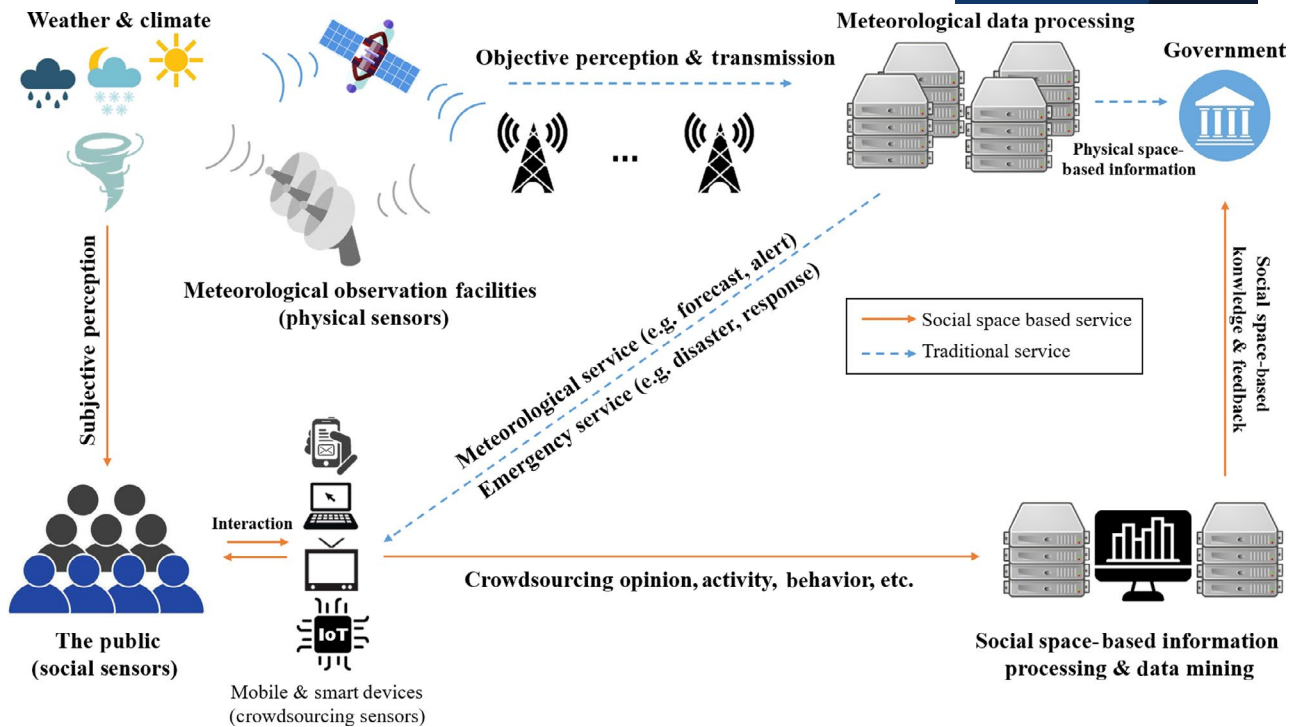


FIGURE 1 Differences between traditional and social space-based meteorological knowledge services. The social space-based meteorological knowledge service (orange) shows an outstanding characteristic of gathering public perception and feedback to official departments/agencies (bottom-up), while the traditional meteorological knowledge service (blue) shows a clear characteristic of getting conclusion from meteorological agencies and delivering notification to the public (top-down)

new perspective for future meteorology research. Here, we use ‘social weather’ to roughly represent novel methods of collecting, processing, analysing and applying multisource, multimodal, massive meteorological data with the help of society and the public, that is a crowdsourcing-assisted approach (Figure S1).

To briefly summarize, this social space-based meteorological research benefits from the following: First, in addition to the objective data able to be collected from the real world, social weather research focuses more on the public's subjective opinions, reactions, feelings and feedback, which can strongly influence emergency planning and decision-making (Grasso & Crisci, 2016; Sharpe & Bennett, 2018). Second, although social weather utilizes the advantages of data fusion, fusing this large heterogeneous dataset efficiently and sparingly between meteorological agencies and other departments in various domains is difficult (e.g. transportation, agriculture and tourism) and is known as the so-called ‘data islands’ problem (Dey *et al.*, 2015; Zipper, 2018; Padilla *et al.*, 2018). Third, the crowdsourcing data collection of social weather provides opportunities for more precise and individual-level data collection, thereby relieving the dependency on deploying a large number of sensors which are costly to deploy over dense ground even-space grids (Strangeways, 2018).

In this paper, we survey existing social space-based meteorological research powered by crowdsourcing. To be specific, we first present the current state of research and report

their bibliometric characteristics. Then, we briefly introduce the data sources that use crowdsourcing to promote meteorological knowledge services in the social space and discussed in depth the commonalities and characteristics of these data sources. On the data basis, we further analyse the specific analytical methods used in the current progress of research and divided them into three levels according to the roles and purposes of the methods. We also discuss the application areas of meteorological services by investigating typical and deployed online weather service applications. Finally, we discuss the current social weather challenges and the issues that need to be addressed and predict the future value of social weather research.

2 | OVERVIEW OF LITERATURE SELECTION AND ANALYSIS ON BIBLIOMETRICS

By using the combination between the words in meteorological-related theme words (weather, climate, atmospheric science, meteorology) and crowdsourcing-related words (crowdsourcing, social media) as the query, we search the relevant journals for the conference papers in the Web of Science and Engineering Village indexing system, respectively. As a result, we selected a total of 97 related journal and conference articles which published from 2011 to

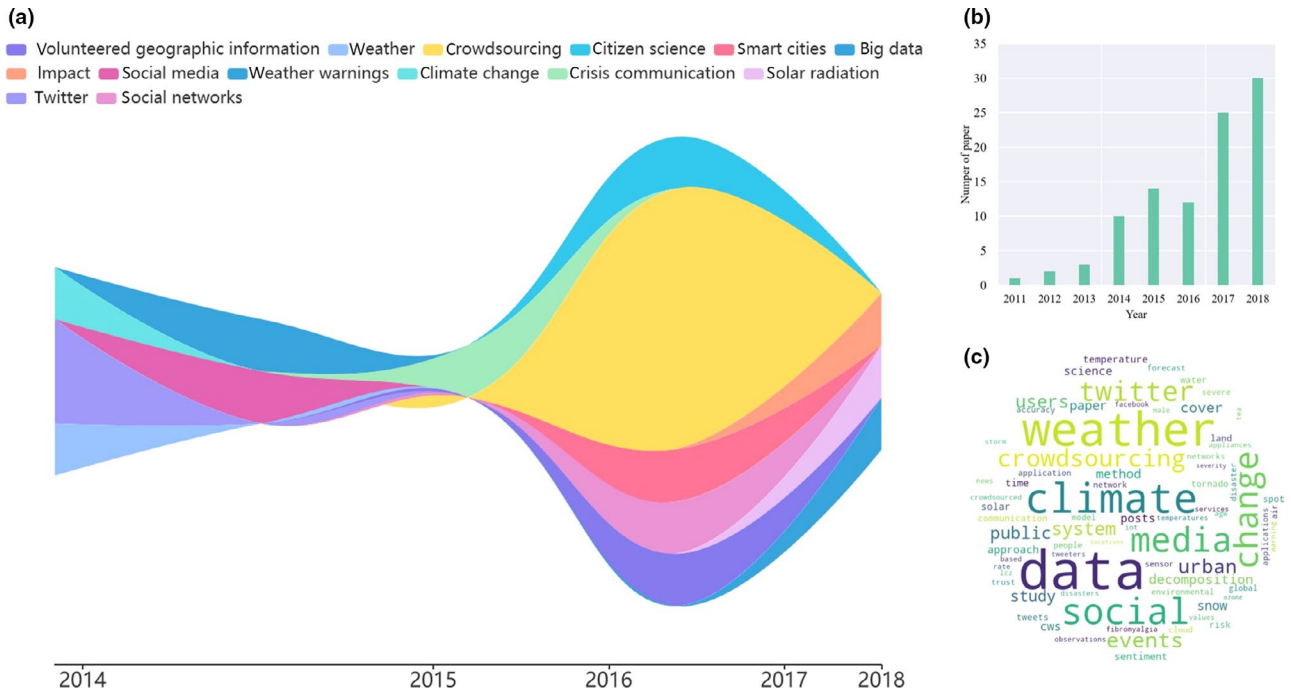


FIGURE 2 Bibliometric characteristics of the surveyed literature. (a) The evolution of frequent typical keywords based on research focused on social weather (2014–2018). Note that the width of each color band represents a keyword and the frequency at which the keyword appears in papers published each year. (b) The distribution of related papers from 2011–2018. (c) A word cloud consisting of the topical words among all related papers’ abstracts, in which the size of word is determined by the word frequency

2018 (Figure 2b), the list of these papers are provided in Supporting Information). For an intuitive overview of this social weather-related research, we present visualizations based on their basic bibliometric factors.

First, as the simplest but most topical form, the keywords of each paper are collected and counted to determine the evolution of research focus in this area. The 14 most frequently appearing keywords were recorded and displayed in Figure 2a. In general, the most common keywords consist of domain-specific topical words (e.g. climate change, weather), methodological words (e.g. crowdsourcing, citizen science, big data), words describing data sources (e.g. social media, Twitter) and targeted areas (e.g. smart cities, crisis communication) that have been present throughout the whole process of this research. In addition, the term ‘crowdsourcing’, which is also used as a part of the title of this paper, shows its first peak of appearance approximately 2015 and has received widespread attention since 2016. Obviously, from the view of the use of keywords, it is suggested that the importance of data sources is equal to their methodology and purpose. In other words, a data-driven approach is widely used in social weather-related studies, which also reflects the basic features of big data analysis in this area.

Then, the abstracts are utilized to find themes that are more detailed. However, to avoid the influence of common but meaningless words in a language, we use the latent Dirichlet allocation (LDA) model to extract topical words of all abstracts from all related papers. Figure 2c shows the

word cloud of topical words extracted from abstracts, with the font size reflected by the frequency of these words. In addition to the words we observed in typical keywords, more detailed theme and objective words, such as sentiments, sensor, disaster, communication, system, etc., also occur in the word cloud with significant appearances in published papers. Furthermore, more detailed research objects (e.g. temperature, forecast, water, network, events, etc.) also appear with a relatively larger frequency, which indicates the information that the current meteorological knowledge service focuses on.

3 | REVIEW OF PREVIOUS WORK AND ANALYSIS

3.1 | Crowdsourcing-based data sources and attributes

As predicted, data derived from social weather are playing an increasingly important role in meteorological research, especially in public discourse (Kirilenko & Stepchenkova, 2014; Auer *et al.*, 2014; Krennert *et al.*, 2018) and in intensive and cost-efficient data acquisition. Also known as ‘crowdsourcing for atmospheric sciences’, social weather data consist of temporal, spatial, textual, behavioural and other features. Recorded contents are specific to one or more aspects of people’s activities and are specific to an individual

or a community. The specific features and their structures and resolutions as well as the objectivity determine the strengths, weaknesses and scopes of further application. In other words, different fusions and combinations of data are selected to solve different tasks and problems. Despite previous research (Muller *et al.*, 2015) suggesting that there are many types of crowdsourcing information, from the view of data-driven information science, in this review, we tagged all data sources as one of four types based on the characteristics of data itself, such as consciousness, objectiveness, mobility and discipline.

3.1.1 | Potential data providers

In crowdsourcing-based data collection, the idea of ‘humans as sensors’ plays a significant role. The role of people in this process can either be as the operators of data collection or the generators of data. After years of development of crowdsourcing-based data collection in meteorology, there

are many data sources continuously providing rich information. Here, we briefly introduce these sources, since previous research has already discussed and listed them (Muller *et al.*, 2015):

- Smart embedded sensors: Sensors that are embedded into smart phones and other IoT devices have been used to estimate fine-grained meteorological parameters. Temperature, pressure and light intensity collected by huge crowds of smart phone owners tagged with high-precision GPS depict an exquisite map (Overeem *et al.*, 2013; Sosko & Dalyot, 2017). This smart phone-embedded device-based collection method benefits from promising geographic resolution, deployment and mobility.
- Social networks: People's evolution of attitude towards long-term meteorological phenomena and short-term emergency communications can be extracted by mining their opinions from social network sites (SNS) (Liu & Zhao, 2017; Sisco *et al.*, 2017; Veltri & Atanasova, 2017; Rossi *et al.*, 2018). As free communication platforms,

TABLE 1 Typical public meteorological-related datasets generated via a crowdsourcing approach

Research or platform	Content	Usage	Other potential domain for application
Henneken, Helm, & Menzel, (2012)	Honey bee swarm events created by beekeepers, public institutions and members of the public and analysed with respect to prevailing weather	Weather influence on animal behaviour	Phenology
Local Environmental Observer Network (LEO, 2012, Mosites <i>et al.</i> , 2018)	670 environmental observations were posted by 1,870 citizen scientists in 35 countries from February 2012 to July 2017	Climate assessment	Environment observation
CrowdFlower sentiment data (Crowdsourcing at Scale, 2013)	The data are comprised of 98,979 tweets. Each tweet was evaluated by at least 5 raters, for a total of approximately 500,000 answers including 5 types of emotion	Meteorological-related sentiment analysis	Social computing
Cyclone Center (Hennon <i>et al.</i> , 2015)	Global, homogeneous, tropical cyclone (TC)-centric IR satellite imagery labelled and analysed by citizen scientists	Tropical cyclone analysis	Remote sensing
Venanzi, Teacy, Rogers, & Jennings, (2015)	Dataset recollected using Amazon Mechanical Turk from an original dataset provided by CrowdFlower as part of the 2013 Crowdsourcing at Scale shared task challenge. The dataset contains 6,000 classifications of the sentiment of 300 tweets, with gold standard sentiment labels, provided by 110 workers	Meteorological-related sentiment analysis	Social computing
Seimon <i>et al.</i> , (2016)	A video imagery database about the 31 May 2013 El Reno, Oklahoma, tornado crowdsourced from storm chasers	Tornado research	Computer vision, social computing, etc.
Kadlec and Ames, (2017)	Snow cover assessment enriched by combining volunteer snow reports, cross-country ski track reports and station measurements to fill cloud gaps in the Moderate Resolution Imaging Spectroradiometer	Snow cover assessment	Remote sensing, social computing, etc.
Fritz <i>et al.</i> , (2017)	Global land cover and land use reference data derived from the Geo-Wiki crowdsourcing platform	Climate assessment	Remote sensing, environmental research, etc.
mPING (NSSL <i>et al.</i> , 2019)	A project to collect weather information from the public through their smart phone or mobile device	Meteorological data process	

social networks accommodate different angles and attitudes expressed by people of different backgrounds and roles. Thus, this makes the data collection of cross-domain opinion and knowledge possible and convenient (Lu *et al.*, 2018b).

- Sensors provided by weather-sensitive departments: Weather and climate influence many industrial sectors. Conversely, using these weather-sensitive industry data in meteorology also has an advantageous influence. This data source can be of great help to interdisciplinary research by virtue of the cross-disciplinary nature of its birth.
- Volunteer and citizen science: Unlike government or agency-led meteorological information gathering, enabling volunteers and citizen science enthusiasts to collect relevant data spontaneously in the scheme of a collaborative research approach is also considered a source of crowdsourcing (Lee *et al.*, 2012; Tiplado & Allamano, 2017; Irwin, 2018).

3.1.2 | Processed public datasets

Due to the results of previous efforts, the generation of crowdsourcing datasets provides more opportunities and accessibilities to academic as well as industrial communities, thereby further benefitting from a collective effort to validate or verify the reliability of these solutions with a single standard. It should also be acknowledged that the production of datasets sometimes requires enormous human and economic costs. Therefore, the free disclosure of the dataset requires great courage and determination for the provider.

For example, Meteorological Phenomena Identification Near the Ground (mPING) makes users of smart phones able to share current weather situations with apps in United States. The collected structural data can be accessed by calling the official application programming interface (API). We also list several typical public datasets for social weather in Table 1.

3.1.3 | Attributes of social weather data

In general, social weather collects quantitative meteorological and climatic signals, geographic coordinates, texts, images and web usage-based behavioural information. In other words, different problems require different data sources with different levels of precision, content and timeliness. Considering the inefficiencies of counting the new applications and platforms that are constantly emerging from the view of information science (the view of data itself, in other words), we use four aspects to label current and potential data sources.

- Consciousness: In this aspect, we mainly examine whether the data were deliberately observed and recorded by

people during acquisition, or whether the weather data were generated during the user's daily life and other inadvertent behaviours. In other words, the question 'is the research-demanded data provided actively by a crowd or is it data collected passively by people?' is asked to judge the consciousness of the data. As an example, Kirilenko *et al.* compared the data processing work between for-profit businesses and volunteer amateur citizen scientists. No matter how the approach is, this kind of work requires users to actively observe, generate and upload information (Kirilenko *et al.*, 2017). Another example is Morita *et al.*, who developed a system that can collect and analyse disaster-related information posted on social media by residents in real time (social sensor system) and implemented it in society. The system can pick up information that is useful for grasping disaster situations even though the intention of these online posts' authors was not to provide information to government agencies (Morita *et al.*, 2018). The advantages and disadvantages of these two types of data are obvious: conscious acquisition can collect in-demand and noiseless data, but the collection density and amount of data are still subject to the recruitment of volunteers and citizen scientists. In contrast, unintentional data acquisition enables large-scale and high-density information collection, but it is likely to be collected less urgently than required and to provide noisy data. There is also research that focuses on fusing these two types of data collection to utilize the advantages of both. For instance, Niforatos *et al.* developed a crowdsourcing weather app that not only periodically samples smartphones' sensors for weather measurements but also allows users to enter their own estimates of both current and future weather conditions. The results from a 32-month public deployment show that this combination results in more accurate temperature estimates and features an average error rate of 2.7°C (Niforatos *et al.*, 2017).

- Objectiveness: Another important characteristic to describe the crowdsourcing-based data is whether the content is a result of subjective feeling and opinion or objective observation. Indeed, the most modern natural sciences such as meteorology relies on objective data, which makes the research quantifiable and measurable. However, subjective opinions and feedback reflect the purpose of meteorological science research: science and technology are developed to meet the needs of humanity and ultimately serve human development. Therefore, how to objectively study these subjective opinions has opened a subfield termed 'opinion mining' in information science. For example, insight from personal experiences and the Twitter feed was extracted during the Pan American Games in 2015 by Herdt *et al.* to aid weather stations in evaluating thermal discomfort and the heightened risk of heat-related illnesses. These subjective opinions are ultimately more densely deployed in the target area than weather stations are and provide the most

direct feelings of the end users that weather stations cannot provide (Herdt *et al.*, 2018). In addition, the SNS platforms provide different data according to their role and functions. For example, a weak relationship social platform such as a microblog seeks the most spread while strong relationship social platform retains privacy for individuals. Currently, major related research on social media is more focused on follower-based platforms (e.g. Twitter, Weibo, etc.) than friend-based one (such as Facebook). The SNS platform distribution used in relevant meteorological-related research papers using social media as a data source can be found in Supporting Information (Figure 3c).

- **Mobility:** Crowdsourcing-based data collection provides a significant attribute that the bottomed sensors can be movable. The mobility makes the collection more intelligent and customizable. Specifically, this attribute fits well with the need for meteorological services that are more concerned with the concerns of people. Combined with the geographic coordinates of different precision levels, mobility data collection effectively improves the efficiency of the use of sensor resources. For example, a typical application scenario is to evaluate meteorological conditions through current vehicular sensing signals and road network information in transportation systems. These approaches take the advantages of regular and structural information gathering especially alongside the daily service work of vehicles (like Uber, taxi, etc.) (Dey *et al.*, 2015; Tomas *et al.*, 2016).
- **Multidisciplinarity:** Meteorological data collection is not limited to meteorological and atmospheric science, but many related fields affected by meteorology can provide data support for social weather through crowdsourcing collection in their industry. In other words, multidisciplinarity enables problem solving assisted by knowledge from other domains. In the field of hydrology, for instance, Lowry & Fienen compared stream gauge measurements from volunteers to official measurements, who collected data via a conscious approach (Lowry and Fienen, 2013). Similarly, phonological change in terrestrial, freshwater and marine organisms shows that many of the climatic events can be reflected, which suggests shared large-scale drivers (Thackeray *et al.*, 2016; Willis *et al.*, 2017; Strangeways, 2018). From natural to artificial fields in agriculture, crop planting records and remote sensing data have also been used to measure key areas for meteorological observations (Dell'Acqua *et al.*, 2018).

To summarize, as a typical characteristic of reviewed papers, all experiments and work are built on the basis of fusing or aggregating different types of data. Meteorology usually refers to a real-world phenomenon, which always has a comprehensive dimension to describe itself. In addition to meteorological data in this area, data from other fields are

important sources of data supplementation, especially when the actual problem is a cross-disciplinary task (e.g. tourism (Padilla *et al.*, 2018), agriculture (Minet *et al.*, 2017; Zipper, 2018), etc.). The advantages and limitations of different social weather reveal a huge demand for advanced data fusion approaches and techniques. However, current research in this field is still at a preliminary stage. The fusion of cross-domain and multisource heterogeneous data still has plenty of space for further breakthroughs. Shown in Table 2, we list the relevant data sources and the data specialties that can be generated by them.

3.2 | Crowdsourcing-motivated analytics on social weather

Data analysis plays a crucial role in social meteorological research. Through data analysis and mining techniques, data-driven public deep-seated behaviours towards weather and climate can be analysed and acquired (including conscious or unconscious ones). Especially when the collected data are semistructured or even unstructured with more or less noise, a powerful analysis and process method is vitally needed. In addition, due to the huge volume and different noises of data, the performance and efficiency of analysis techniques are also important factors that should be taken into consideration. In this review, from the perspective of the discovery, mastery and application of data as well as knowledge, we categorize these research efforts into three levels: correlation and relationship discovery in acquired data, generalization and abstraction of discovered knowledge, and current effort to a more systemic and intelligent knowledge service. However, it should be noted that this stage-based grading method should not be considered as having either technical or research value.

3.2.1 | Correlation and relationship discovery in acquired data

The first level is to discover basic correlations and relationships between human and meteorological-related data. In this stage, statistical methods are utilized to calculate the correlation, optimize the process, map the distribution, etc. In this stage, an elemental knowledge is acquired and generalized. The target of this level is to find the pointcut of a specific problem and to attempt to ensure this pointcut and the correlation are stable. To acquire such knowledge, methods to capture, clean, and extract features of data are proposed. Feature extraction and correlation calculation are the most common topics in the related research.

On the premise that human behaviour (including behaviour on the Internet) is affected by the weather and climate (Li *et al.*, 2014), the regular patterns between

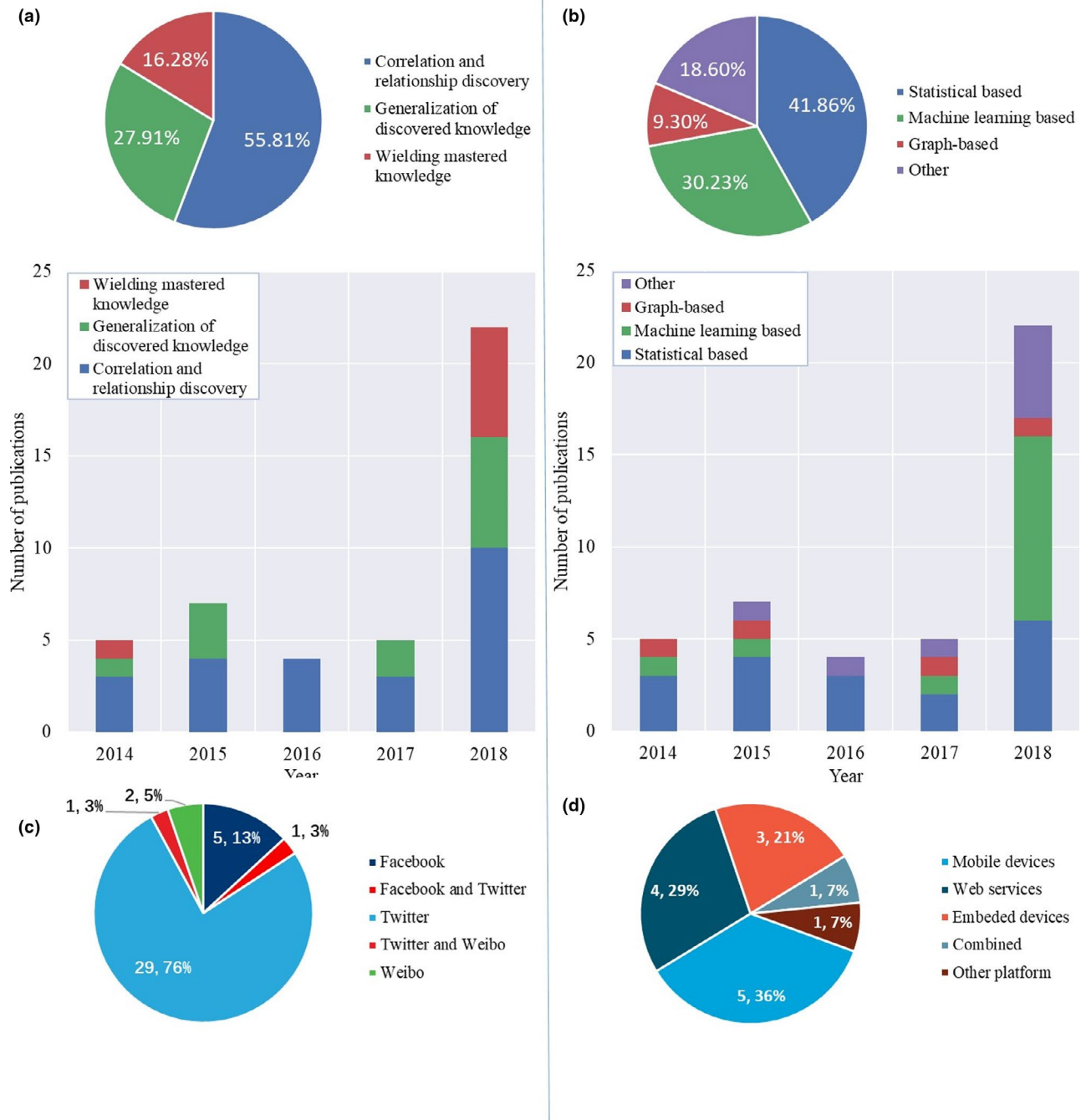


FIGURE 3 Partial results of literature analysis. (a) Technical level of analytics in each year and overall proportion (2014–2018). (b) The proportion of the type of analysis methods in each year and the overall proportion (2014–2018). (c) The distribution of social media platforms used by research articles that use social media as a data source. (d) Distribution of equipment types and platforms deployed by meteorological knowledge service application in recent 5 years

behaviour and atmospheric condition can be feasibly explored (Dutta *et al.*, 2017). For example, it has been found that mental well-being deteriorates during warmer periods according to an investigation of depressive language in over 600 million social media posts, potentially leading to a change in suicide rates comparable to the estimated impact of economic recessions, suicide prevention programmes or gun restriction laws (Burke *et al.*, 2018). It is not the only instance of using social media to assist correlation

investigation by combining medical illnesses and meteorological information. A series of studies have reported the potential for meteorological and disease relationships through crowdsourcing approaches, such as the varying degrees of correlation between local weather conditions and the subjective text on Twitter attributable to discussions of Fibromyalgia (Haghighi *et al.*, 2017) and depression (Yang *et al.*, 2015; Modoni & Tosi, 2016). In the field of behaviour differences among different groups, Holmberg and

TABLE 2 Sources and characteristics of available social weather data

Modality Of data	Observation from citizen sciences	Social media	Smart devices and embedded sensors	Sensors provided by other observation departments
Numerical	Time Observed parameters	Time	Time	Time
Textual	Incident reports	Tweets Online posts and articles	Observed parameters	Observed parameters
Graphical	Images Videos	Images Videos	--	--
Geographic	Geographic coordinates Locations	Location	Images Videos Geographic coordinates Wireless grid positioning Base station positioning	Images Videos Geographic coordinates Location
Behavioural	--	Interaction data Information dissemination	--	Incident responses
Characteristic of data	Structural feature Precision/resolution Timeliness/regularity Consciousness Objectiveness Mobility Multidisciplinarity	Semistructured or unstructured Low Low and irregular No Subjective Yes Yes	Structured High High Mixed Objective Yes No	Structured High High Yes Objective No Yes
Examples	Labelling flow- ering time Providing extra weather observation Wood burning emissions Flood report Tropical cyclone data labelling	Willis <i>et al.</i> , (2017) Quarati <i>et al.</i> , (2017), Fenner, Meier, Bechtel, Otto, & Scherer, (2017) Lopez-Aparicio, Vogt, Schneider, Kahila-Tani, and Broberg, (2017) Naik, (2016) Hennon <i>et al.</i> , (2015)	Chatfield and Brajawidagda, (2014) (Haghighi <i>et al.</i> , 2017, Verma, Park, & Kim, 2017) Short-term warning Weather-effected illness analysis Attitude analysis towards climate change Sentiment analysis towards weather	(Overseem <i>et al.</i> , 2013) Igoe <i>et al.</i> , (2018) Tiwari and Kaushik, (2015) Pankratius, Lind, Coster, Erickson, and Semeter, (2014) Igoe, Amar, Parisi, and Turner, (2017) Solar ultraviolet B rays evaluation Zimmerman and Kruschwitz, (2017), Baylis <i>et al.</i> , (2018) Surface Pressure Mass and Madaus, (2014)

Hellsten found that there are more women that have a firm attitude towards the human impact of climate change, while men are more sceptical in their daily discussions on Twitter (Holmberg & Hellsten, 2014; Holmberg & Hellsten, 2015). Another scenario is that risk perception may affect users' trust towards severe natural hazards, such as the increased estimated likelihood of severe weather and the expectation of danger and harm, which were also associated with greater trust and are helpful for advancing risk mitigation actions (Losee & Joslyn, 2018).

Rather than quantitative correlation, other studies linking user online behaviour to meteorology have also reported several interesting phenomena. For instance, Anderson and Huntington explored sarcasm and incivility usage in Twitter towards climate change during 4,094 extreme weather events. Their result showed that the sarcasm and incivility in discussions about climate change are overall rarely seen and more related to right-leaning politics profiles (Anderson & Huntington, 2017; Anderson & Becker, 2018). Facing severe weather, such as smog and storms, differences in reposting microblogs between oriental- and occidental-backgrounded users are discussed: less humour or increase in affective outpouring as the crisis developed was found on the Chinese Sina Weibo than on Twitter (Lin *et al.*, 2016). Similarly, Stewart and Wilson tracked the entirety of Hurricane Sandy's lifecycle in social media with their proposed model based on a crisis communication theory (Stewart & Wilson, 2016). By contrast, information spread through interpersonal social networks such as Facebook also contributes to the awareness and popularization of meteorological information such as climate change (Ali, 2011; Sharpe & Bennett, 2018). Connor *et al.* analysed messages spreading through the communication chain in social network on Facebook, finding that statements centred on conventional climate change topics survived longer in communication chains than those with less conventional topics (Connor *et al.*, 2016). In political use in the United States, climate change-related tweets also exhibit the difference in use by people with different political beliefs: when the tweets are classified according to the location of states, red states prefer 'climate change' rather than 'global warming', and blue states show the opposite preference (Jang & Hart, 2015). A similar phenomenon could also be observed between left- and right-leaning news agencies (Saunders *et al.*, 2018). Pearce *et al.* analysed the nature of the debate about the 2013 Intergovernmental Panel on Climate Change on Twitter and investigate three user communities to try to find significant links between the climate-convinced and critical users (Pearce *et al.*, 2014).

While the core of crowdsourcing is to capture, analyse and apply data for each terminal user as much as possible, the roles of professional and authoritative meteorological services providers in the process of information dissemination are important as well (Fownes *et al.*, 2018). For example,

Kim *et al.* identified that such professional news and weather agencies played a dominant role as information sources and information diffusers to the public (Kim *et al.*, 2018). However, blog owners that deny or downplay global warming attempted to disregard the overwhelming scientific evidence, such as Arctic sea-ice loss and polar bear vulnerability, to cast doubt on other established ecological consequences of global warming, thereby aggravating the consensus gap (Harvey *et al.*, 2018). A case study on the question-answer community Quora also suggests that emphasis on specific subjects but not popular knowledge of climate change could draw more public attention on the issue of climate change (Jiang *et al.*, 2018). Brandt *et al.* examined another case including before, during and after historic rainfall and flooding in the Midlands region of the greater Columbia, South Carolina, area in October 2015 to distinguish the characteristic of tweets of each period in the life cycle of an emergency event on social network (Brandt *et al.*, 2018). The crowdsourced opinion can also be used as a measurement of accuracy of forecasting. Wiwatwannana *et al.*, (2015) reported that their case study in Bangkok, Thailand, examined that the social response can serve as an alternative validation for weather forecasting, and such responses were also able to produce bias and weakness in forecast services more prominent.

3.2.2 | Generalization of discovered knowledge

The second level establishes generalization of acquired knowledge. The evidence-based knowledge and data found in the previous phase were abstracted at this stage. In other words, researchers found different types of regular patterns at the first level and concluded them to be laws for predicting unseen new situation and instances at this stage. Common tasks include the classification, regression, clustering and combination of them. Powered by artificial intelligence (e.g. machine learning) tools, such predictions and estimations can be implemented and driven by preprocessed data. The target of this stage is to abstract knowledge and examine the soundness and accuracy of this abstraction and to further visualize it to be easier to understand by the public.

For example, generating weather reports from collected tweets is one kind of abstraction (Butgereit, 2014). When applying them to a practical use, it is able to generate weather reports from new unseen tweets, which is a type of generalization. As another typical example of generalization, the sentiment tendency on subject texts abstracts emotional factors from various texts that people publish on the Internet on a daily basis. Applied in the meteorological field, people's feelings about weather and climate are also an important issue worth studying (Lachlan *et al.*, 2014). Cody *et al.*, (2015) use their previously implemented tool named the

'Hedonometer' (Dodds *et al.*, 2011) to measure the happiness of tweets containing the word 'climate', extracting lexical patterns which can be further used for automatically judging a new text's sentimental tendency as it appears on the social network. However, the form of tweets that contains keywords still highlights more weakly than hashtags (e.g. '#ClimateChange'). Hashtags significantly reduce and even eliminate subjectivity errors associated with humans and provide an inexpensive solution for event detection on Twitter, thus revealing a promising perspective for detecting emergent weather events (Hamed *et al.*, 2015). Not limited to English, the current academic circles have also proposed a variety of meteorological emotion calculation methods for different language usage habits (e.g. Arabic (Al-Kabi *et al.*, 2018)).

Predicting new samples in the future through existing data and scenarios is also a typical task of generalization of knowledge. Benefiting from a global-scale sensor network crowdsourcing 15+ billion transponder messages per day from aircraft, Trub *et al.* proposed a method to estimate meteorological parameters in upper air conditions with a relatively low error (Trub *et al.*, 2018). Data subject to timing constraints tend to be more affected when the authoritative data are scarce, or unavailable for some periods. Thus, this makes the crowdsourcing work more significant in missing data imputation. Restrepo-Estrada *et al.* utilized georeferenced authoritative tweets on Twitter as an alternative data source to replenish timing such sequential feature and established a rainfall-runoff estimation and flood forecasting in Brazil (Restrepo-Estrada *et al.*, 2018), and a similar task has been reported by Arthur *et al.* by using a text filter on relevant tweets in the UK (Arthur *et al.*, 2018).

In the task of abstracting data to higher level understandable information, it is generally believed that directly collected data only provide information at the record level (signal, posts, etc.), while information better for human understanding requires an event level in the meteorological field (Ali & Ogie, 2017). Zhu *et al.*, (2019) proposed a severe weather event extraction method to compile online news and tweets on Sina Weibo into weather event by considering their temporal, spatial and thematic feature.

Not only can external data be used to solve questions in meteorological research, but meteorological data can also be used for fixing problems in other fields. In transportation research, for instance, modelling and trying to predict traffic situations require the aspects of features. Lin *et al.* extracted weather events captured by hashtag from Twitter to predict the traffic speed (i.e. smoothness) on freeways with a linear regression model (Lin *et al.*, 2015). A city-level traffic awareness alerting model was also proposed by Lu *et al.*, (2018b), and this model extracted opinions about severe weather and traffic from Weibo and validated the prediction and alert by referring to news reports, thereby saving large costs on deploying physical sensors and integrating data from different

sources and departments. In hydrological and agricultural research, Ravazzani *et al.*, (2017) proposed a prediction model to estimate soil moisture and crop water requirements by combining ground observed data from space and reporting information from cyberspace with a crowdsourcing approach.

3.2.3 | Wielding mastered knowledge: towards a systemic and intelligent knowledge utilization

The final level cultivates knowledge service by aggregating knowledge and information derived by the aforementioned two stages. This stage aims to provide systemic and robust service and framework to support high-level human activities such as decision-making advice, expert systems, emergency response, and information dissemination. In addition to complex intrinsic models and algorithms, this level of analysis also has very high requirements for visualization on results of analysis. These requirements ask for readable, informative-rich, and intelligible information due to the involvement of cross-domain people, machines, data, methods and processes (Kox *et al.*, 2018).

To serve the emergency response mechanism, Ripberger *et al.*, (2014) proposed a measurement on public attention to severe weather risk communication based on the growing stream of data that individuals publish on social media platforms, thus providing a better understanding of the relationship between risk communication, attention and public reactions to severe weather. Another example is making the decision between a safe but costly option (e.g. spending on protecting from a storm) and a risky option (e.g. spending less under the estimation that the storm will not attack). A contrasting experiment conducted by Mu *et al.*, (2018) suggests that although providing more evidence and information to users can make the judgement more convincing, the best decision is not always made using a massive amount of information.

On the issue of situational awareness and decision support, challenged but motivated by large heterogeneous data streams, powerful and adaptable situational awareness systems are highly demanding. Akbar *et al.* proposed an architecture for fusing and analysing traffic, weather and social media streams, thereby predicting the probability of congestion in real time (Akbar *et al.*, 2018). Adjusted by the feedback of traffic administrators, their case in Madrid reaches a promising predicting accuracy and similar results are also reported by Lu *et al.* on cases in Beijing (Lu *et al.*, 2018b) and Qingdao (Lu *et al.*, 2018a). From the view of individualized services, the knowledge service is usually located in personalized advising and recommendations. Chen *et al.* implemented and provided a health guidance system for urban residents on respiratory diseases, travel

and sleep quality based on fusing urban air quality data collected through meteorological sites, mobile crowdsourcing and IoT sensing and even users' body signals (Chen *et al.*, 2018).

3.2.4 | Methodologies of analytics utilized in social weather research

To summarize, these analytical methods translate the data discussed in the previous subsection into knowledge. To further analyse the details of these analytical methods, we divide the articles in the above review by the proportion of various types of research in recent five years and present the result in Figure 3a,b. As a result, Figure 3a shows that nearly five years of research have focused primarily on discovering the relevance and relationships associated with atmospheric science, while knowledge extraction and decision-making respond which leverage existing knowledge to newly unseen data will have gained considerable attention in the near future. We also classify these papers according to their theoretical tools: statistical based, machine learning based, graph based and other methods. In this paper, the statistical based method refers to a method of calculating the intrinsic relationship by statistical methods after the existing data are processed, while machine-learning-based approaches originated from statistical learning methods, using existing data to perform a specific task effectively without using explicit instructions, relying on patterns and inference of the data instead. The graph-based approach uses graph mining algorithms to derive conclusions by constructing network relationships on existing studied data. In Figure 3b, the paper based on statistical methods has been in a stable publication for nearly five years and accounts for about 41% of the total number of papers published in analytical methods. On the other hand, articles based on machine-learning methods have experienced extremely significant growth in the recent year. Articles belonging to machine-learning-based analytic methods in 2018 accounted for 76.92% of the articles in the past five years (10 in 13 papers). This result also shows that the research hotspots on analytical methods in this area are moving towards a goal of systematic and intelligent knowledge services.

3.3 | Crowdsourcing-driven applications

Focusing on applications and platforms, in this section, we analyse proposed systems that have already been deployed. Additionally, we also discuss the characteristic as well as problems of these crowdsourced supported meteorological knowledge service oriented systems. We divide these

applications into five categories based on the role of crowdsourcing in their research. In addition, user interfaces of typical online applications can be found at Figure S2.

3.3.1 | Crowdsourcing for dense high-resolution meteorological parameter

As discussed above, the crowdsourcing approach makes the collection with universal sensors possible, thereby sharply reducing the cost of data acquisition and releasing the constraint of observation equipment. Thus, the most direct function from which the meteorological research can benefit is to collect and analyse perceived parameters with a precise geotemporal tag. The systems collecting temperature by using the batteries of smart phones established by Overeem *et al.* (2013) and Chau (2018) are typical examples of dense high-resolution meteorological parameter collection. Public pictures with geo-tagged records are also exploited to monitor situation of snow (Giuliani *et al.*, 2016) and ultraviolet (Wilkes *et al.*, 2016). Similarly, a series of platforms that collect meteorological parameters by end users actively clicking and uploading observations through apps on smart phones achieved this goal with more diverse but less data due to the requirement for user labour (Niforatos *et al.*, 2015; Niforatos *et al.*, 2016; NSSL *et al.*, 2019). When the observation turns to emergent, Frigerio *et al.*'s app, named MAppERS (Frigerio *et al.*, 2018), provides not only the observation but also emergency requests to authorities.

3.3.2 | Crowdsourcing for feedback and public concerns

Inaccurate forecasting and failure of focusing public concerns may cause a serious trust crisis between meteorological departments and the public. Therefore, the evaluation and opinion by the public are valuable for judging the service of meteorological agencies. In addition, the timely identification of hot events of concern is also conducive to enhancing the public's goodwill and satisfaction with the meteorological service sector. As a most direct application on this basis, Zhu *et al.* proposed a meteorological public opinion monitoring the platform by concentrating related news and online posts into events with focused heat values (Zhu *et al.*, 2019).

3.3.3 | Crowdsourcing for individualized service

With the rapid development of machine intelligence, the understanding of human preferences makes it possible to personalize services for individuals. By discovering these rules

based on commonalities and characteristics in big data, personalized services provide great stimulation for smart cities, smart medical and intelligent education. In conjunction with the meteorological field, the system called 'UH-BigDataSys' proposed by Chen *et al.* combines climate, air and personal-level health data to achieve personalized guidance services for respiratory diseases, outdoor travel and sleep quality (Chen *et al.*, 2018).

3.3.4 | Crowdsourcing for infrastructure

Infrastructure and ad hoc physical collectors also bring an extra opportunity for establishing meteorological knowledge service (Chapman *et al.*, 2015). Vishwarupe *et al.*, (2015) proposed an embedded system to collect real-time weather data and deploy them under cooperation with telecom infrastructure. To relieve the bottleneck of transmitting large volumes of unstructured data by using data centres, Du *et al.*, (2018) proposed a novel strategy for creating an efficient and low-latency-distributed message delivery system with operational connected vehicles applications. In addition to data, crowdsourcing can focus on computational resources. Chen *et al.* used a virtual cloud host to build a meteorological computing model, thereby realizing the crowdsourcing of computing resources (Chen *et al.*, 2017).

3.3.5 | Crowdsourcing for interdisciplinary

One of the great advantages of crowdsourcing is that regardless of the purpose of crowdsourcing, the methods and means of implementing crowdsourcing often have similarities. Therefore, the use of crowdsourcing can effectively advance the meteorologically related interdisciplinary research progress. For instance, Aihara *et al.* proposed a smart phone-based driving recorder. The recorder records a video of the front environment while driving and provides a series of upload interfaces for users to report weather, traffic and other conditions to collect and aggregate information in the above fields (Aihara *et al.*, 2016).

3.3.6 | Platform and scenario of applications

Based on the analysis of the above-mentioned articles on the application of meteorological knowledge services, we further divide them according to the applicable equipment and applying platforms which is shown in Figure 3d. Different applicable devices often correspond to different application scenarios. For example, mobile device-based applications are directed to the services of user terminals that are capable of providing information to end users while collecting

data through these terminal devices. On the other hand, the data management-based web system usually processes and visualizes the aggregated data to server end user such emergency responder and decision-makers. As can be observed from Figure 3d, these two categories account for the majority of current application research (65%). Another research that has received considerable attention is the use of embedded devices to achieve low-energy, high-efficiency meteorological knowledge service data transmission (21%).

Considering these current applications in meteorological service domain, we can observe that since 2010, especially in the past five years, these applications have rapidly evolved from appearance to maturity with external technology environments such as mobile computing, social networking and machine intelligence. However, compared to the current research work in this crowdsourcing-assisted field, these applications still need to be further upgraded to achieve the 'smart' level. In other words, the current service is still unified or depends on the user's active query of information. How to provide personalized and differentiated knowledge services and recommendations to the end users intelligently by the system is still needed to further explore and implement.

4 | DISCUSSION: CHALLENGES AND FUTURE OF SOCIAL WEATHER

Despite the increasing attention given to this area, research on social weather still has prominent room for improvement. These progressive spaces are still rooted in the various aspects we have outlined in the previous sections, but the motivation for such improvement is to change or even replace the existing schemes.

In the area of data collection, two crucial questions still remain to be answered: First, with the next-generation mobile computing technologies such as 5G and the increasing access of mobile and IoT devices, how can we distribute sensors more efficiently during crowdsourcing data collection (e.g. rewards), and how can we resolve the transmission of massive amounts of collected data efficiently in the near future to avoid data bottlenecks? The second question is that multisource data are difficult to collect, aggregate and match. For example, as a widely acknowledged difficult problem, author attribution and name disambiguation among different platforms require complicated processing models, even manpower especially, when the name structures of the users are different than each other (Wang *et al.*, 2012; Tang *et al.*, 2012; Zhang *et al.*, 2018b). However, the uncertainty inherent with crowdsourcing-based collection such as inconsistency observation standards, subjective basis and data errors of missing data still must be solved (Khan *et al.*, 2015; Tiplado & Allamano, 2017).

When processing and analysing the collected data, challenges also exist: First, feature engineering extracts features from large volumes of data with noise and poor structure is a classic but vital step in data preprocessing. However, the emergence of deep learning has incorporated this so-called feature engineering into the intelligent model (Lecun *et al.*, 2015). By learning from huge volumes of data, the state-of-the-art model acquires a special representation of data and knowledge automatically, which makes more intelligent tasks in meteorology possible. In addition, pretrained language representation models based on prior knowledge (e.g. BERT (Devlin *et al.*, 2018), Word2vec (Mikolov *et al.*, 2013) and Glove vector (Pennington *et al.*, 2014)) better organize and represent the semantic information of text by learning from large texts in the general domain. This greatly assists the analysis of the subjective text information and content, making weather-related text processing able to be explored in a more complex and connotative direction. So far, the application of deep learning in this domain is still at a very preliminary stage. Second, the joint analysis of multimodal data is also expected to advance. Obviously, social weather contains data of different kinds of structures, such as images, texts, entities and relationships. Such heterogeneous data make unifying the model difficult. Therefore, generalized, comprehensive and informative data fusion has far-reaching implications for meteorological knowledge services (Lin *et al.*, 2018).

To date, social weather-based service applications are still mostly in the ‘Query-Results’-based passive service. However, smarter ‘active services’, recommendation systems, automated briefings and even personalized climate-related education and healthcare are more important and more challenging directions for the integrated use of social meteorological knowledge. Another active service also required to establish a crowdsourcing scheme and strategy (e.g. reward scheme) to optimize the efficiency, range and goal during the whole process of collection, analysis and application (dos Santos, 2017; Chapman *et al.*, 2017). Additionally, from the perspective of meteorological teaching, how to organize knowledge efficiently to establish accurate push for professionals and students in this field is worthy of further study (Tarus *et al.*, 2017; Wan & Niu, 2019). Similarly, how to mine new knowledge in the measurement of scientific literature, and even predict new hotspots is also an important direction of meteorological knowledge services (Tarus *et al.*, 2018; Nie *et al.*, 2019; Yousif *et al.*, 2019).

Based on these current researches and zooming out from microscopical procedures and technics to the macroscopic framework and structure of social weather-related research, we propose to extend the concept of cyber-physical-social systems (CPSS) to cover all related research. Originated in intelligent system society, CPSS provides a framework with

which to depict the relationship among real-world situations (physical), sensing and computing (cyber) and human factors (social), which has been used in enterprises and transport domains (Wang, 2010; Zheng *et al.*, 2016; Tse *et al.*, 2017). In the CPSS model of meteorology, the composition of the triad is human/public, meteorological department and weather/climate. The common intersection of any two and three subfields conceptually includes the research scope of social weather, which could be found in the Figure S3. In other words, the intersection of subjects is the main research area of current social weather, and the specific problems studied can be classified into one of the subjects.

5 | CONCLUSION

Crowdsourcing-assisted meteorological service plays an important role in addressing the problem of harmonizing the public feedback and data sources as well as providing the information support to both official departments and individuals in a more affordable, dense and public concern-centred way. Compared with conventional meteorological approach generally providing service based on physical sensor, transmission and model processing, the social-based meteorological service on the other hand provides a new dimension of collecting, analysing and applying data provided and benefited from the public.

In this paper, we have presented a review of crowdsourcing-assisted meteorological service which we term ‘social weather’. The contribution of our work in this survey is two-fold. First, we have summarized the research achievements in the class of crowdsourcing-assisted meteorological service from its early beginning of 2010 to 2018 by classifying the papers with a focus on attributes of data collection, stages of information analysis and scope of practical applications. Secondly, we have given an insight on current challenges as well as the future trends on crowdsourcing-assisted meteorological knowledge service.

This survey has revealed that crowdsourcing-assisted meteorological services in combination with other data-driven and intelligent techniques are getting more extensive acceptance and application. Furthermore, it is evident from the papers reviewed that current researches are contributing towards a systemic and intelligent knowledge service to establish a better bridge among academic, industrial and individual community. We hope that this review study will widen the frontiers of knowledge and provide useful literature for researchers interested in advancing this field of research.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61370137, the Ministry of

Education-China Mobile Research Foundation Project under Grant 2016/2-7, the Postgraduate Education Research Project of Beijing Institute of Technology under Grant 2017JYYJG-004, China's National Strategic Basic Research Program (973 Program) under Grant 2012CB720700 and also supported by The Public Weather Service Center of China Meteorological Administration. The authors would like to thank all the editor and anonymous reviewers for their suggestions that have helped us to improve the work.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Yifan Zhu  <https://orcid.org/0000-0002-7695-1633>

REFERENCES

- Adger, W.N., Barnett, J., Brown, K., Marshall, N., & O'Brien, K. (2013) Cultural dimensions of climate change impacts and adaptation. *Nature Climate Change*, 3, 112–117.
- Aihara, K., Bin, P., Imura, H., Takasu, A., & Tanaka, Y. (2016). On feasibility of crowdsourced mobile sensing for smarter city life. *Distributed, Ambient and Pervasive Interactions, (Dapi 2016)*, 9749, 395–404.
- Akbar, A., Kousiouris, G., Pervaiz, H., Sancho, J., Ta-Shma, P., Carrez, F. *et al.* (2018) Real-time probabilistic data fusion for large-scale IoT applications. *Ieee Access* 6, 10015–10027.
- Ali, M.S.S. (2011). *The Use of Facebook to Increase Climate Change Awareness among Employees*. 2011 International Conference on Social Science and Humanity. Singapore: IACSIT Press.
- Ali, A.U., & Ogie, R. (2017) Social media and disasters highlighting some wicked problems. *IEEE Technology and Society Magazine*, 36, 41–+.
- Al-Kabi, M., Alsmadi, I., Khasawneh, R., & Wahsheh, H. (2018) Evaluating social context in arabic opinion mining. *International Arab Journal of Information Technology*, 15, 974–982.
- Anderson, A.A., & Becker, A.B. (2018) Not just funny after all: Sarcasm as a catalyst for public engagement with climate change. *Science Communication*, 40, 524–540.
- Anderson, A.A., & Huntington, H.E. (2017) Social media, science, and attack discourse: how Twitter discussions of climate change use sarcasm and incivility. *Science Communication*, 39, 598–620.
- Arthur, R., Boulton, C.A., Shotton, H., & Williams, H.T.P. (2018) Social sensing of floods in the UK. *PLoS ONE*, 13, e0189327.
- Auer, M.R., Zhang, Y.M., & Lee, P. (2014) The potential of microblogs for the study of public perceptions of climate change. *Wiley Interdisciplinary Reviews-Climate Change*, 5, 291–296.
- Baylis, P., Obradovich, N., Kryvasheyev, Y., Chen, H., Coviello, L., Moro, E., ..., Fowler, J.H. (2018) Weather impacts expressed sentiment. *PloS one*, 13(4), e0195750. <https://doi.org/10.1371/journal.pone.0195750>
- Brandt, H.M., Turner-McGrievy, G., Friedman, D.B., Gentile, D., Schrock, C., Thomas, T. *et al.* (2018) Examining the role of Twitter in response and recovery during and after historic flooding in South Carolina. *Journal of Public Health Management and Practice*, 25, E6–E12.
- Burke, M., Gonzalez, F., Bayliss, P., Heft-Neal, S., Baysan, C., Basu, S. *et al.* (2018) Higher temperatures increase suicide rates in the United States and Mexico. *Nature Climate Change*, 8, 723–+.
- Butgereit, L. (2014) *Crowdsourced weather reports: An implementation of the μ model for spotting weather information in Twitter*. Ist-africa Conference, 7-9 May 2014 Le Meridien Ile Maurice, Mauritius. IEEE.
- Chapman, L., Muller, C.L., Young, D.T., Warren, E.L., Grimmond, C.S.B., Cai, X.M. *et al.* (2015) The birmingham urban climate laboratory an open meteorological test bed and challenges of the smart city. *Bulletin of the American Meteorological Society*, 96, 1545–1560.
- Chapman, L., Bell, C., & Bell, S. (2017) Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *International Journal of Climatology*, 37, 3597–3605.
- Chatfield, A. T., & Brajawidagda, U. (2014). Crowdsourcing hazardous weather reports from citizens via twittersphere under the short warning lead times of EF5 intensity tornado conditions. In 2014 47th Hawaii International conference on system sciences (pp. 2231–2241). IEEE. <https://doi.org/10.1109/HICSS.2014.281>
- Chau, N.H. (2018) A new approach to estimate urban air temperature using smartphones. *Intelligent Information and Database Systems, Aciids 2018. Pt I*, 10751, 633–641.
- Chen, X.H., Huang, X.L., Jiao, C.Y., Flanner, M.G., Raeker, T., & Palen, B. (2017) Running climate model on a commercial cloud computing environment: A case study using Community Earth System Model (CESM) on Amazon AWS. *Computers & Geosciences*, 98, 21–25.
- Chen, M., Yang, J., Hu, L., Hossain, M.S., & Muhammad, G. (2018) Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. *IEEE Communications Magazine*, 56, 14–20.
- Cody, E.M., Reagan, A.J., Mitchell, L., Dodds, P.S., & Danforth, C.M. (2015) Climate change sentiment on twitter: An unsolicited public opinion poll. *PLoS ONE*, 10, e0136092.
- Connor, P., Harris, E., Guy, S., Fernando, J., Shank, D.B., Kurz, T. *et al.* (2016) Interpersonal communication about climate change: how messages change when communicated through simulated online social networks. *Climatic Change*, 136, 463–476.
- Crowdsourcing at Scale (2013) *Sentiment Analysis Judgment Dataset- Judge emotions about weather from Twitter*. Retrieved from <https://sites.google.com/site/crowdscale2013/shared-task/sentimentanalysis-is-judgment-data>
- Dell'Acqua, F., Iannelli, G.C., Torres, M.A., & Martina, M.L.V. (2018) A Novel Strategy for Very-Large-Scale Cash-Crop Mapping in the Context of Weather-Related Risk Assessment, Combining Global Satellite Multispectral Datasets, Environmental Constraints, and In Situ Acquisition of Geospatial Data. *Sensors*, 18, 591.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018) *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. ArXiv e-prints [Online]. Available: <https://ui.adsabs.harvard.edu/#abs/2018arXiv181004805D> [Accessed October 01, 2018].
- Dey, K.C., Mishra, A., & Chowdhury, M. (2015) Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review. *IEEE Transactions on Intelligent Transportation Systems*, 16, 1107–1119.
- Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A., & Danforth, C.M. (2011) Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE* 6: e26752.

- Dos Santos, M.M.L. (2017). *Holism, collective intelligence, climate change and sustainable cities*. 8th International Conference on Ambient Systems, Networks and Technologies (Ant-2017) and the 7th International Conference on Sustainable Energy Information Technology (Seit 2017), 109, 763–770.
- Du, Y.H., Chowdhury, M., Rahman, M., Dey, K., Apon, A., Luckow, A. *et al.* (2018) A distributed message delivery infrastructure for connected vehicle technology applications. *IEEE Transactions on Intelligent Transportation Systems*, 19, 787–801.
- Dutta, S., Ray, S., & Roy, S. (2017) *Correlation between weather and weather-related tweets — A preliminary study*. IEEE International Conference on Big Data.
- Fenner, D., Meier, F., Bechtel, B., Otto, M., & Scherer, D. (2017) Intra and inter 'local climate zone' variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. *Meteorologische Zeitschrift*, 26(5), 525–547. <https://doi.org/10.1127/metz/2017/0861>
- Fownes, J.R., Yu, C., & Margolin, D.B. (2018) Twitter and climate change. *Sociology Compass*, 12: e12587.
- Frigerio, S., Schenato, L., Bossi, G., Mantovani, M., Marcato, G., & Pasuto, A. (2018) Hands-on experience of crowdsourcing for flood risks. An android mobile application tested in Frederikssund, Denmark. *International Journal of Environmental Research and Public Health*, 15, 1926.
- Fritz, S., See, L., Perger, C., McCallum, I., Schill, C., Schepaschenko, D., ..., Lesiv, M. (2017) A global dataset of crowdsourced land cover and land use reference data. *Scientific data*, 4, 170075. <https://doi.org/10.1038/sdata.2017.75>
- Giuliani, M., Castelletti, A., Fedorov, R., & Fraternali, P. (2016) Using crowdsourced web content for informing water systems operations in snow-dominated catchments. *Hydrology and Earth System Sciences*, 20, 5049–5062.
- Goncalves, A., Silva, C., Morreale, P., & Bonafide, J. (2014) *Crowdsourcing for Public Safety*. 2014 8th Annual Ieee Systems Conference (Syscon). 50–56.
- Grasso, V., & Crisci, A. (2016) Codified hashtags for weather warning on Twitter: An Italian case Study. *PLoS Currents Disasters*, 8. <http://currents.plos.org/disasters/article/codified-hashtags-for-weather-warning-on-twitter-an-italian-case-study/>
- Greengard, S. (2014) Weathering a new era of big data. *Communications of the Acm*, 57, 12–14.
- Haghighi, P.D., Kang, Y.B., Buchbinder, R., Burstein, F., & Whittle, S. (2017) Investigating subjective experience and the influence of weather among individuals with fibromyalgia: A content analysis of Twitter. *JMIR Public Health & Surveillance* 3: e4.
- Hamed, A.A., Ayer, A.A., Clark, E.M., Irons, E.A., Taylor, G.T., & Zia, A. (2015) Measuring climate change on Twitter using Google's algorithm: perception and events. *International Journal of Web Information Systems*, 11, 527–544.
- Harvey, J.A., van den Berg, D., Ellers, J., Kampen, R., Crowther, T.W., Roessingh, P. *et al.* (2018) Internet blogs, polar bears, and climate-change denial by proxy. *BioScience*, 68, 281–287.
- Henneken, R., Helm, S., & Menzel, A. (2012). Meteorological Influences on Swarm Emergence in Honey Bees (Hymenoptera: Apidae) as Detected by Crowdsourcing. *Environmental Entomology*, 41(6), 1462–1465. <https://doi.org/10.1603/En12139>
- Hennon, C.C., Knapp, K.R., Schreck, C.J. III, Stevens, S.E., Kossin, J.P., Thorne, P.W., ..., Striegl, M. (2015) Cyclone center: can citizen scientists improve tropical cyclone intensity records? *Bulletin of the American Meteorological Society*, 96(4), 591–607. <https://doi.org/10.1175/BAMS-D-13-00152.1>
- Herd, A.J., Brown, R.D., Scott-Fleming, I., Cao, G.F., Macdonald, M., Henderson, D. *et al.* (2018) Outdoor thermal comfort during anomalous heat at the 2015 Pan American games in Toronto, Canada. *Atmosphere*, 9, 321.
- Holmberg, K., & Hellsten, I. (2014) *Analyzing the climate change debate on Twitter: content and differences between genders*. *Proceedings of the 2014 ACM conference on Web science*. Bloomington, Indiana, USA: ACM.
- Holmberg, K., & Hellsten, I. (2015) Gender differences in the climate change communication on Twitter. *Internet Research*, 25, 811–828.
- Igoe, D.P., Amar, A., Parisi, A.V., & Turner, J. (2017) Characterisation of a smartphone image sensor response to direct solar 305 nm irradiation at high air masses. *Science of the Total Environment*, 587, 407–413. <https://doi.org/10.1016/j.scitotenv.2017.02.175>
- Igoe, D.P., Parisi, A.V., Amar, A., Downs, N.J., & Turner, J. (2018) Atmospheric total ozone column evaluation with a smartphone image sensor. *International Journal of Remote Sensing*, 39(9), 2766–2783. <https://doi.org/10.1080/01431161.2018.1433895>
- Irwin, A. (2018) Citizen science comes of age efforts to engage the public in research are bigger and more diverse than ever. But how much more room is there to grow?. *Nature*, 562, 480–482.
- Jang, S.M., & Hart, P.S. (2015) Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. *Global Environmental Change-Human and Policy Dimensions*, 32, 11–17.
- Jiang, H.C., Qiang, M.S., Zhang, D.C., Wen, Q., Xia, B.Q., & An, N. (2018) Climate change communication in an online Q&A Community: A Case Study of Quora. *Sustainability* 10, 1509.
- Kadlec, J., & Ames, D.P. (2017) Using crowdsourced and weather station data to fill cloud gaps in MODIS snow cover datasets. *Environmental Modelling & Software*, 95, 258–270. <https://doi.org/10.1016/j.envsoft.2017.06.002>
- Khan, A., Imona, S.K.A., & Das, S.K. (2015) A novel localization and coverage framework for real-time participatory urban monitoring. *Pervasive and Mobile Computing*, 23, 122–138.
- Kim, J., Bae, J., & Hastak, M. (2018) Emergency information diffusion on online social media during storm Cindy in US. *International Journal of Information Management*, 40, 153–165.
- Kirilenko, A.P., Molodtsova, T., & Stepchenkova, S.O. (2015) People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Global Environmental Change-Human and Policy Dimensions*, 30, 92–100. <https://doi.org/10.1016/j.gloenvcha.2014.11.003>
- Kirilenko, A.P., & Stepchenkova, S.O. (2014) Public microblogging on climate change: One year of Twitter worldwide. *Global Environmental Change-Human and Policy Dimensions*, 26, 171–182.
- Kirilenko, A.P., Desell, T., Kim, H., & Stepchenkova, S. (2017) Crowdsourcing analysis of Twitter data on climate change: Paid workers vs. volunteers. *Sustainability*, 9, 2019.
- Kox, T., Kempf, H., Luder, C., Hagedorn, R., & Gerhold, L. (2018) Towards user-orientated weather warnings. *International Journal of Disaster Risk Reduction*, 30, 74–80.
- Krennert, T., Pistotnik, G., Kaltenberger, R., & Csekits, C. (2018) Crowdsourcing of weather observations at national meteorological and hydrological services in Europe. *Advances in Science and Research*, 15, 71–76.

- Lachlan, K.A., Spence, P.R., & Lin, X. (2014) Twitter use during a weather event: Comparing content associated with localized and nonlocalized hashtags. *Communication Studies*, 65, 519–534.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015) Deep learning. *Nature*, 521, 436.
- Lee, S.M., Olson, D.L., & Trimi, S. (2012) Innovative collaboration for value creation. *Organizational Dynamics*, 41, 7–12.
- LEO (2012) *Local Environmental Observer Network*. Retrieved from <https://www.leonetwork.org/en>
- Li, J., Wang, X., & Hovy, E. (2014) *What a Nasty Day: Exploring Mood-Weather Relationship from Twitter*. Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. Shanghai, China: ACM.
- Lin, L., Ni, M., He, Q., Gao, J., & Sadek, A.W. (2015) Modelling the impacts of inclement weather on freeway traffic speed exploratory study with social media data. *Transportation Research Record*, 2482(1), 82–89.
- Lin, X.L., Lachlan, K.A., & Spence, P.R. (2016) Exploring extreme events on social media: A comparison of user reposting/retweeting behaviors on Twitter and Weibo. *Computers in Human Behavior*, 65, 576–581.
- Lin, Q.K., Niu, Y.Q., Zhu, Y.F., Lu, H., Mushonga, K.Z., & Niu, Z.D. (2018) Heterogeneous knowledge-based attentive neural networks for short-term music recommendations. *IEEE Access*, 6, 58990–59000.
- Liu, J.C.E., & Zhao, B. (2017) Who speaks for climate change in China? Evidence from Weibo. *Climatic Change*, 140, 413–422.
- Lopez-Aparicio, S., Vogt, M., Schneider, P., Kahila-Tani, M., & Broberg, A. (2017) Public participation GIS for improving wood burning emissions from residential heating and urban environmental management. *Journal of Environmental Management*, 191, 179–188. <https://doi.org/10.1016/j.jenvman.2017.01.018>
- Losee, J.E., & Joslyn, S. (2018) The need to trust: How features of the forecasted weather influence forecast trust. *International Journal of Disaster Risk Reduction*, 30, 95–104.
- Lowry, C.S., & Fienen, M.N. (2013) CrowdHydrology: crowdsourcing hydrologic data and engaging citizen scientists. *Ground Water*, 51, 151–156.
- Lu, H., Shi, K., Zhu, Y., Lv, Y., & Niu, Z. (2018a) Sensing urban transportation events from multi-channel social signals with the Word2vec fusion model. *Sensors*, 18, 4093.
- Lu, H., Zhu, Y., Shi, K., Lv, Y., Shi, P., & Niu, Z. (2018b) Using adverse weather data in social media to assist with city-level traffic situation awareness and alerting. *Applied Sciences-Basel* 8, 1193.
- Mass, C.F., & Madaus, L.E. (2014) Surface pressure observations from smartphones: A potential revolution for high-resolution weather prediction? *Bulletin of the American Meteorological Society*, 95(9), 1343–1349. <https://doi.org/10.1175/Bams-D-13-00188.1>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013) *Efficient Estimation of Word Representations in Vector Space*. ArXiv e-prints [Online]. Available: <https://ui.adsabs.harvard.edu/#abs/2013arXiv1301.3781M> Accessed January 01, 2013.
- Minet, J., Curnel, Y., Gobin, A., Goffart, J.P., Melard, F., Tychon, B. et al. (2017) Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach. *Computers and Electronics in Agriculture*, 142, 126–138.
- Modoni, G.E., & Tosi, D. (2016) *Correlation of Weather and Moods of the Italy Residents through an Analysis of Their Tweets*. IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), 22–24 Aug. 2016 2016 Vienna, Austria. IEEE, 216–219.
- Morita, N., Hayakawa, M., & Takao, N. (2018) Disaster prevention system utilizing social media information. *Fujitsu Scientific & Technical Journal*, 54, 54–58.
- Mu, D., Kaplan, T.R., & Dankers, R. (2018) Decision making with risk-based weather warnings. *International Journal of Disaster Risk Reduction*, 30, 59–73.
- Mosites, E., Lujan, E., Brook, M., Brubaker, M., Roehl, D., Tcheripanoff, M., & Hennessy, T.W.J.O.H. (2018). Environmental observation, social media, and One Health action: A description of the Local Environmental Observer (LEO) Network. 6, 29–33. <https://doi.org/10.1016/j.onehlt.2018.10.002>
- Muller, C.L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G. et al. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35, 3185–3203.
- Naik, N. (2016). *Flooded streets—A crowdsourced sensing system for disaster response: A case study*. In 2016 IEEE International Symposium on Systems Engineering (ISSE) (pp. 1–3). IEEE. <https://doi.org/10.1109/SysEng.2016.7753186>
- Nie, Y., Zhu, Y., Lin, Q., Zhang, S., Shi, P., & Niu, Z. (2019) Academic rising star prediction via scholar's evaluation model and machine learning techniques. *Scientometrics*, 120, 461–476.
- Niforatos, E., Vourvopoulos, A., & Langheinrich, M. (2015) 'Weather With You': Evaluating Report Reliability in Weather Crowdsourcing. The 14th International Conference on Mobile and Ubiquitous Multimedia. Linz, Austria.
- Niforatos, E., Elhart, I., & Langheinrich, M. (2016) *WeatherUSI: User-Based Weather Crowdsourcing on Public Displays*. Cham: Springer International Publishing, 567–570.
- Niforatos, E., Vourvopoulos, A., & Langheinrich, M. (2017) Understanding the potential of human-machine crowdsourcing for weather data. *International Journal of Human-Computer Studies*, 102, 54–68.
- NSSL, Oklahoma, T. U. O. & Studies, T. C. I. F. M. M. (2019) *mPING: crowdsourcing weather reports* [Online]. Available: <https://mping.ou.edu/> [Accessed 2019-01-02 2019].
- Overeem, A., Robinson, J.C.R., Leijnse, H., Steeneveld, G.J., Horn, B.K.P., & Uijlenhoet, R. (2013) Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40, 4081–4085.
- Padilla, J.J., Kavak, H., Lynch, C.J., Gore, R.J., & Diallo, S.Y. (2018) Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PLoS ONE* 13, e0198857.
- Pankratius, V., Lind, F., Coster, A., Erickson, P., & Semeter, J. (2014) Mobile Crowd Sensing in Space Weather Monitoring: The Mahali Project. *Ieee Communications Magazine*, 52(8), 22–28. <https://doi.org/10.1109/Mcom.2014.6871665>
- Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014) Climate Change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group 1 Report. *PLoS ONE* 9(4): e94785.
- Pennington, J., Socher, R., & Manning, C.D. (2014) *GloVe: Global Vectors for Word Representation*. *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.
- Prpić, J., Taeiagh, A., & Melton, J. (2015) The Fundamentals of Policy Crowdsourcing. *Policy & Internet*, 7, 340–361.
- Quarati, A., Clematis, A., Roverelli, L., Zereik, G., D'Agostino, D., Mosca, G., & Masnata, M. (2017). Integrating heterogeneous weather-sensors data into a smart-city app. In 2017 International Conference on High Performance Computing &

- Simulation (HPCS) (pp. 152–159). IEEE. <https://doi.org/10.1109/HPCS.2017.33>
- Ravazzani, G., Corbari, C., Ceppi, A., Feki, M., Mancini, M., Ferrari, F. *et al.* (2017) From (cyber)space to ground: new technologies for smart farming. *Hydrology Research*, 48, 656–672.
- Restrepo-Estrada, C., de Andrade, S.C., Abe, N., Fava, M.C., Mendiondo, E.M., & de Albuquerque, J.P. (2018) Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computers & Geosciences*, 111, 148–158.
- Ripberger, J.T., Jenkins-Smith, H.C., Silva, C.L., Carlson, D.E., & Henderson, M. (2014) Social media and severe weather: Do tweets provide a valid indicator of public attention to severe weather risk communication? *Weather Climate and Society*, 6, 520–530.
- Rossi, C., Acerbo, F.S., Ylisen, K., Juga, I., Nurmi, P., Bosca, A. *et al.* (2018) Early detection and information extraction for weather-induced floods using social media streams. *International Journal of Disaster Risk Reduction*, 30, 145–157.
- Saunders, C., Grasso, M.T., & Hedges, C. (2018) Attention to climate change in British newspapers in three attention cycles (1997–2017). *Geoforum*, 94, 94–102.
- Schafer, M.S. (2012) Online communication on climate change and climate politics: a literature review. *Wiley Interdisciplinary Reviews-Climate Change*, 3, 527–543.
- Seimon, A., Allen, J.T., Seimon, T.A., Talbot, S.J., & Hoadley, D.K. (2016) Crowdsourcing the El Reno 2013 Tornado A New Approach for Collation and Display of Storm Chaser Imagery for Scientific Applications. *Bulletin of the American Meteorological Society*, 97(11), 2069–2084. <https://doi.org/10.1175/Bams-D-15-00174.1>
- Sharpe, J.D., & Bennett, D.M. (2018) Use of Facebook for civilian-led disaster response after a winter storm: A "Snowed Out Atlanta" case study. *Journal of Emergency Management*, 16, 255–266.
- Sisco, M.R., Bosetti, V., & Weber, E.U. (2017) When do extreme weather events generate attention to climate change? *Climatic Change*, 143, 227–241.
- Sosko, S., & Dalyot, S. (2017) Crowdsourcing user-generated mobile sensor weather data for densifying static geosensor networks. *ISPRS International Journal of Geo-Information*, 6, 61.
- Spiegelhalter, D., Pearson, M., & Short, I. (2011) Visualizing uncertainty about the future. *Science*, 333, 1393–1400.
- Stewart, M.C., & Wilson, B.G. (2016) The dynamic role of social media during Hurricane #Sandy: An introduction of the STREMI model to weather the storm of the crisis life cycle. *Computers in Human Behavior*, 54, 639–646.
- Strangeways, I. (2018) Phenology: Plants and animals as meteorological sensors. *Weather*, 73, 86–89.
- Taeihagh, A. (2017) Crowdsourcing, sharing economies and development. *Journal of Developing Societies*, 33, 191–222.
- Tang, J., Fong, A.C.M., Wang, B., & Zhang, J. (2012) A unified probabilistic framework for name disambiguation in digital library. *IEEE Transactions on Knowledge and Data Engineering*, 24, 975–987.
- Tarus, J.K., Niu, Z.D., & Yousif, A. (2017) A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems-the International Journal of Esience*, 72, 37–48.
- Tarus, J.K., Niu, Z., & Mustafa, G. (2018) Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, 50, 21–48.
- Thackeray, S.J., Henrys, P.A., Hemming, D., Bell, J.R., Botham, M.S., Burthe, S. *et al.* (2016) Phenological sensitivity to climate across taxa and trophic levels. *Nature*, 535, 241–U94.
- Tipaldo, G., & Allamano, P. (2017) Citizen science and community-based rain monitoring initiatives: an interdisciplinary approach across sociology and water science. *Wiley Interdisciplinary Reviews-Water*, 4, e1200.
- Tiwari, S., & Kaushik, S. (2015). *Crowdsourcing based fuzzy information enrichment of tourist spot recommender systems. In International Conference on Computational Science and Its Applications* (pp. 559–574). Cham: Springer. https://doi.org/10.1007/978-3-319-21410-8_43
- Tomas, V.R., Pla-Castells, M., Martinez, J.J., & Martinez, J. (2016) Forecasting adverse weather situations in the road network. *IEEE Transactions on Intelligent Transportation Systems*, 17, 2334–2343.
- Trub, R., Moser, D., Schafer, M., Pinheiro, R., & Lenders, V. (2018) *Monitoring Meteorological Parameters With Crowdsourced Air Traffic Control Data*. 2018 17th Acm/Ieee International Conference on Information Processing in Sensor Networks (Ipsn), 25–36.
- Tse, R., Zhang, L.F., Lei, P., & Pau, G. (2017) Social network based crowd sensing for intelligent transportation and climate applications. *Mobile Networks and Applications*, 23, 177–183.
- Veltri, G.A., & Atanasova, D. (2017) Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science*, 26, 721–737.
- Venanzi, M., Teacy, W., Rogers, A., & Jennings, N. (2015). Weather Sentiment - Amazon Mechanical Turk dataset. Retrieved from, <https://eprints.soton.ac.uk/376543/>
- Verma, S., Park, Y., & Kim, M. (2017) Predicting Flu-Rate Using Big Data Analytics Based on Social Data and Weather Conditions. *Advanced Science Letters*, 23(12), 12775–12779. <https://doi.org/10.1166/asl.2017.10897>
- Vishwarupe, V., Bedekar, M., & Zahoor, S. (2015) *Zone specific weather monitoring system using crowdsourcing and telecom infrastructure*. 2015 International Conference on Information Processing (ICIP), 16–19 Dec. 2015 Pune, India. IEEE.
- Wan, S., & Niu, Z. (2019). A hybrid E-learning recommendation approach based on learners' influence propagation. *IEEE Transactions on Knowledge and Data Engineering*. Epub ahead of print.
- Wang, F.Y. (2010) The emergence of intelligent enterprises: From CPS to CPSS. *IEEE Intelligent Systems*, 25, 85–88.
- Wang, F.Y. (2014) Scanning the issue and beyond: Real-time social transportation with online social signals. *IEEE Transactions on Intelligent Transportation Systems*, 15, 909–914.
- Wang, X., Tang, J., Cheng, H., & Yu, P.S. (2012) *ADANA: Active Name Disambiguation*. IEEE International Conference on Data Mining.
- Wilkes, T.C., McGonigle, A.J.S., Pering, T.D., Taggart, A.J., White, B.S., Bryant, R.G. *et al.* (2016) Ultraviolet imaging with low cost smartphone sensors: Development and application of a raspberry Pi-based UV camera. *Sensors*, 16, 1649.
- Williams, H.T.P., McMurray, J.R., Kurz, T., Lambert, F.H. (2015) Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change- Human and Policy Dimensions*, 32, 126–138. <https://doi.org/10.1016/j.gloenvcha.2015.03.006>
- Willis, C.G., Law, E., Williams, A.C., Franzone, B.F., Bernardos, R., Bruno, L. *et al.* (2017) CrowdCurio: an online crowdsourcing platform to facilitate climate change studies using herbarium specimens. *New Phytologist*, 215, 479–488.
- Wiwatwattana, N., Sangkhatad, K., & Srivatanakul, T. (2015) *An analysis and case study of using mobile crowdsourcing for weather forecast verification*. IEEE International Conference on Software

- Engineering & Service Science, 23-25 Sept. 2015 Beijing, China. IEEE.
- Xiong, G., Zhu, F., Liu, X., Dong, X., Huang, W., Chen, S. *et al.* (2015) Cyber-physical-social System in Intelligent Transportation. *IEEE/CAA Journal of Automatica Sinica*, 2, 320–333.
- Yang, W., Mu, L., & Shen, Y. (2015) Effect of climate and seasonality on depressed mood among twitter users. *Applied Geography*, 63, 184–191.
- Yousif, A., Niu, Z., Chambua, J., & Khan, Z.Y. (2019) Multi-task learning model based on recurrent convolutional neural networks for citation sentiment and purpose classification. *Neurocomputing*, 335, 195–205.
- Zhang, J.J., Wang, F., Wang, X., Xiong, G., Zhu, F., Lv, Y. *et al.* (2018a) Cyber-physical-social systems: The state of the art and perspectives. *IEEE Transactions on Computational Social Systems*, 5, 829–840.
- Zhang, Y., Zhang, F., Yao, P., & Tang, J. (2018b) *Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop*. The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18), London, United Kingdom. ACM, 10 pages
- Zheng, X.H., Chen, W., Wang, P., Shen, D.Y., Chen, S.H., Wang, X. *et al.* (2016) Big data for social transportation. *IEEE Transactions on Intelligent Transportation Systems*, 17, 620–630.
- Zhu, Y., Chambua, J., Lu, H., Shi, K., & Niu, Z. (2019) An opinion based cross-regional meteorological event detection model. *Weather*, 74(2), 51-55.
- Zimmerman, S., & Kruschwitz, U. (2017). *Speaking of the weather: Detection of meteorological influences on sentiment within social media*. In *2017 9th Computer Science and Electronic Engineering (CEECE)* (pp. 1–6). IEEE. <https://doi.org/10.1109/CEECE.2017.8101590>
- Zipper, S.C. (2018) Agricultural research using social media data. *Agronomy Journal*, 110, 349–358.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zhu Y, Zhang S, Li Y, Lu H, Shi K, Niu Z. Social weather: A review of crowdsourcing-assisted meteorological knowledge services through social cyberspace. *Geosci Data J.* 2019;00:1–19. <https://doi.org/10.1002/gdj3.85>