# Monitoring surface water quality using social media in the context of citizen science

Hang Zheng[1], Hong Yang[2], Di Long[2], Hua Jing[2]

[1]School of Environment and Civil Engineering, Dongguan University of Technology, Dongguan, Guangdong, 523106, China; School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, 4072, Australia

[2]State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, China

*Correspondence to*: Hang Zheng (zhenghang@tsinghua.edu.cn)

**Abstract.** Surface water quality monitoring (SWQM) provides essential information for water environmental protection. However, SWQM is costly and limited in terms of equipment and sites. The global popularity of social media and intelligent mobile devices with GPS and photography functions allow citizens to monitor surface water quality. This study aims to propose a method for SWQM using social media platforms. Specifically, a WeChat-based application platform is built to collect water quality reports from volunteers, which have been proven valuable for water quality monitoring. The methods for data cleaning and volunteer recruitment are discussed based on the collected reports. The proposed methods provide a framework for collecting water quality data from citizens and offer a primary foundation for big data analysis in future research.

**Keywords.** water quality, monitoring, social media, volunteered geographic information

## 1 Introduction

Surface freshwater is a finite resource that is necessary to the survival of mankind and the ecosystem. An adequate quantity and quality of water are also essential for sustainable development (Khalil and Ouarda, 2009). However, many surface water systems have been contaminated by treated or untreated wastewater that have been discharged by domestic, industrial, and agricultural water users. Water quality has also become an important component of the global water scarcity crisis.

The degradation of the surface water system emphasizes the need to determine the status of water quality in detecting water pollution and in providing scientific guidance for water resources management (Wang et al., 2014). Water quality monitoring refers to the acquisition of quantitative and representative information on the physical, chemical, and biological characteristics of water bodies over time and space (Sanders et al., 1983; Strobl and Robillard, 2008). A water quality monitoring network requires monitoring sites, frequency, variables, and instruments as well as trained/educated field personnel. However, establishing a surface water quality monitoring (SWQM) network in a broad area entails huge costs

(Horowitz, 2013). For example, the US Geological Survey runs the Mississippi River Basin monitoring network to address land loss and hypoxia in the Gulf Coast. However, collecting a single sample from this site costs between US$4,000 and US$6,000, while analyzing various physical/chemical parameters costs an additional US$1,500 to US$2,000 per sample (Horowitz, 2013). These costs reduce the number of samples and sites that can be monitored, thereby necessitating the

5    installation of several monitors on various sites and samples at regular temporal intervals. These limitations hinder the monitoring program from detecting illegal polluting activities, such as hidden sewage dumping, which tend to occur in areas that are located far from the monitoring sites or at a time when no sampling has been conducted. For example, many Chinese industrial facilities dump their sewage water discharge in rivers in the middle of the night to avoid detection (Wei, 2013).

10   A participatory monitoring approach by citizens can fill the spatiotemporal gaps of the current monitoring network. According to Wei (2013), two ordinary citizens in China observed that someone has been dumping sewage into a river near their home from 2004 to 2007. By collecting samples and taking photos throughout the years, these two citizens have documented changes in the color, smell, and temperature of the river water. These voluntary records imply that those citizens who directly are affected by sewage discharge are strongly motivated to monitor and report polluting activities. If more

15   volunteer reporters come forward, then we can detect hidden sewage dumping or polluting activities.

Water quality can be defined in terms of anything from one variable to hundreds of compounds and for multiple usages (Khalil et al., 2010). Taking photographs effectively offers evidence of hidden or midnight sewage dumping activities. Citizens without professional equipment for water quality analysis can describe the physical characteristics of the water (e.g.,

20   color, smell, and temperature) to assess its quality and degree of pollution. Voluntary reporting is more flexible, effective, and inexpensive than the traditional monitoring programs being operated by the government.

Therefore, volunteered geographic information (VGI) in the citizen science context provides a proximate sensing solution for water conservation issues. VGI has been recently introduced as an alternative to the traditional authoritative information

25   provided by mapping agencies and corporations (Goodchild and Glennon, 2010). VGI has been defined as "collaboratively contributed geographic information" (Bishr and Mantelas, 2008) in the context of participatory geographic information systems (GIS) (McCall, 2003), crowdsourcing GIS (Goodchild and Glennon, 2010), participatory planning (Seeger, 2008), and citizen science (Tulloch, 2008). Citizen science, which is an indispensable means of combining ecological research with environmental education and natural history observation, ranges from community-based monitoring to the use of the Internet

30   to "crowd source" various scientific tasks (i.e., from data collection to discovery) (Dickinson et al., 2012). Citizen science is a process whereby citizens are involved in science as researchers (Linda E. Kruger, Margaret A. Shanno, 2000). A citizen scientist voluntarily collects or processes data as a component of a scientific enquiry. These scientists participate in projects related to climate change, invasive species, conservation biology, ecological restoration, water quality monitoring, population ecology, and other types of monitoring (Silvertown, 2009). VGI in the context of citizen science can be produced

immediately and may determine environmental changes as soon as they occur. Therefore, VGI offers an innovative approach for improving environmental governance by fostering accountability, transparency, legitimacy, and other dimensions of governance (McCall, 2003).

5   As a new approach, VGI has recently attracted the attention of researchers. VGI has been applied to numerous research and business domains, particularly in detecting, reporting, and geo-tagging disasters, including earthquakes (Kim, 2014), floods (Perez et al., 2015), hurricanes (Bunce et al., 2012; Virtual Social Media Working Group, 2013), wildfires (Slavkovikj et al., 2014), tsunamis (Mersham, 2010), and storms (Lwin et al., 2015). VGI has successfully increased public knowledge on emergency situations and provided a novel and effective approach for disaster warning and management. Sakaki (2009) built

10 an earthquake detection system in Japan by monitoring reports submitted by citizens through tweets. This system promptly detects earthquakes and sends e-mails to registered users within a minute (occasionally within 20 seconds) after detecting earthquakes. The notifications from this system are delivered much faster than the announcements of the Japan Meteorological Agency, which are broadcasted six minutes after an earthquake. Tang et al. (2015) descriptively evaluated the strengths, weaknesses, opportunities, and threats of VGI in managing the California drought in 2014 and provided an

15 overall description of the role of this system in disaster management. Apart from offering a practical tool for events detection, VGI provides a new level of interaction, participation, and engagement to citizens for environmental governance (Werts et al., 2012). VGI also creates a new paradigm to investigate the self-aware, self-adapting, and self-organizing socio-technical system that combines people, mobile technology, and social media in a complex network of information (Perez et al., 2015).

20   One of the major obstacles in using VGI lies in its unknown quality. The general population is not trained to make specific observations necessary in environmental management and may either intentionally or unintentionally supply erroneous information. Data quality is often unknown, and data sampling is frequently dispersed and unstructured. Other types of data provided by amateurs have attracted similar concerns, which reflect the profound association among qualifications, institutions, and trust (Goodchild and Glennon, 2010). Nevertheless, several grounds show that the quality of VGI can

25 approach and even exceed that of authoritative sources (Goodchild and Glennon, 2010). Fore et al. (2001) trained volunteers to collect benthic macro invertebrates using professional protocols, and found no significant difference between the field samples collected by volunteers and professionals. Citizen volunteers with proper training can collect reliable data and make stream assessments that are comparable with those made by professionals. The data collected by volunteers can also supplement the information being used by government agencies to manage and protect rivers and streams (Fore et al., 2001).

30

Community-based water quality monitoring has been conducted in several countries, such as the Secchi Dip-In program in the US (Lee et al., 1997), the Waterwatch program in Australia (Kingham, 2002), and the Open Air Laboratories Water Survey in the UK (Rose et al., 2015). The Australian Waterwatch program is a national community-based monitoring network that aims to involve community groups and individuals in the protection and management of waterways (Nicholson

et al., 2002). Devlin et al. (2001) analyzed the movement of nutrients and sediments into the Great Barrier Reef during high flow events using the community-based data from the Waterwatch program. Metzger and Lendvay (2006) applied the well-demonstrated benefits of community-based monitoring to the struggle for environmental justice of the low-income, minority residents of the Bayview Hunters Point community in San Francisco, California. These aforementioned programs provide volunteers with the protocols, guidelines, equipment, and trainings necessary for water quality monitoring. The volunteers also collect water samples and measure their quality through test strips and apparatus. Nevertheless, community-based monitoring still entails a high economic cost and much inconvenience, thereby limiting the application of this program.

Social media, including Twitter, Facebook, Sina Weibo, and Weixin (the popular Chinese version of Twitter), can guide and offer incentives to volunteers through real-time online communication. Social media has recently become a major communication channel in our society (Jiang et al., 2015). Internet-based applications allow people to conduct online communications intended for interaction, community input, and collaboration (Lindsay, 2011). Social media also allow multiple parties to share information using their computers or mobile devices, specifically through social networking sites (e.g., Facebook, YouTube, and Twitter), SMS, chatrooms, discussion forums, and blogs (Tang et al., 2015). Social media builds on the ideological and technological foundations of Web 2.0, and enables the creation and exchange of user-generated content (Kaplan and Haenlein, 2010). Social media has several major functions in environmental management processes, including one- and two-way information sharing, situational awareness, rumor control, reconnection, and decision making (Tang et al., 2015). Jiang et al. (2015) monitored the dynamic changes of air quality in large cities by analyzing the spatiotemporal trends in geo-targeted social media messages using comprehensive big data filtering procedures. Werts et al. (2012) launched the AbandonedDevelopments.com website to collect VGI, monitored the sediment pollution of abandoned structures in upstate South Carolina, and combined Web-GIS technologies, data sources, and social media for future applications in soil and water conservation.

The advertising, instruction, and guidance for water quality monitoring can be spread extensively and directly delivered to the mobile devices of potential volunteers through social media platforms. The observed sewage dumping or water pollution activities can be disseminated rapidly in social media networks and call the attention of the government. Social media provides a platform for volunteers to present, discuss, and communicate their criticism, anger, and solutions to the water pollution issues that they observe. Communication and mutual encouragement strongly motivate volunteers to monitor water quality and share their observations. Discussing pollution activities in social media networks encourages public opinion and pressures the government to solve these problems. Government feedback can also be promptly disseminated to volunteers through social media. The timely dissemination of government feedback motivates volunteers to monitor water quality continuously. Volunteers own smartphones that are equipped with digital cameras, GPS, digital maps, and other resources, which grant each empowered citizen in a densely populated city with the ability to create and share information (Goodchild and Glennon, 2010).

This study aims to establish an approach for SWQM through citizen scientists. A social-media-based application is built to collect water quality information and to monitor water pollution using VGI and social media. The findings highlight the feasibility of using VGI in monitoring water quality. The effects of photographed function, anonymous submission, and economic incentives on increasing data credibility and volunteer motivation are also analyzed. This paper is organized as follows. Section 2 presents the methodology. Section 3 presents the monitoring reports that are obtained across China. Section 4 discusses the data quality and motivation of volunteers. Section 5 draws the conclusions.

## 2 Methodology

A methodological framework is established to collect sensory surface water quality data from volunteer citizens who describe and take photographs of the water bodies that they pass by or are located nearby. These citizens send their descriptions and photographs to a data center through a social media application installed on their mobile devices. The framework also discloses illegal/hidden sewage dumping activities that are very difficult to detect through official monitoring networks. Upon receiving sewage dumping reports from the volunteers, the data center will transfer these reports to the mass media and water authorities after performing credibility checks.

### 2.1 Data type

Four indicators are adopted to describe the physical characteristics of water quality (Figure 1). The volunteer citizens choose from 11 water colors, including red, orange, yellow, green, cyan, blue, purple, milky, pink, black, and crystal. Smell is quantified based on the scores given by the volunteers; each volunteer is asked to rate the smell of the sample from 0 to 10, with 0 implying a lack of odor and 10 implying a foul odor. Turbidity is scored between 0 and 10, where 0 implies transparency and 10 implies non-transparency. A higher score also suggests the presence of more contaminants in the water. The presence of floating objects or materials on the water is rated on a scale of 0 to 10, where 0 indicates the absence of any floating objects, while 10 indicates that the water is completely covered with oil, plastics, and rubbish, among others. This item offers an integrated assessment of water quality, which can be ranked as worst, very bad, bad, good, or excellent. Volunteers evaluate the water quality based on their perception.

### 2.2 Application in the social media platform

The Tsinghua Environment Monitoring Platform (TEMP, http://www.thuhjjc.com/) is built based on public WeChat accounts. WeChat is a mobile text and voice messaging communication service that was released by Tencent in China in January 2011. WeChat eventually became one of the largest messaging applications in China, with over a billion existing accounts and 700 million active users (Intelligence, 2016). WeChat provides text messaging, hold-to-talk voice messaging, broadcast messaging, video conferencing, photo and video sharing, and location sharing functions (Tencent, 2016). Users have to register for a public account, which allows them to push their feeds, interact, and offer their services to their

subscribers. These public accounts create a platform for various services, such as hospital pre-registrations, visa renewals, or credit card services (Wikipedia, 2016). WeChat also allows users to post images and texts, share music and articles, and comment or "like" posts in the Moments section of other profiles. The contents and comments in the Moments section can only be viewed by the friends of a particular user. WeChat also supports payment and money transfer, thereby offering its users with peer-to-peer transfer and electronic bill payment services (Tencent, 2016).

Volunteer reporters log into TEMP through their WeChat accounts and then submit their reports together with the GPS position of the reported water body. The location can either be automatically extracted from the devices or manually inputted by the reporters. Volunteers can tweet their reports to their friends or post and comment on them in Moments. TEMP also ranks the volunteers based on their contributions, with the top-ranking reporters receiving awards, such as cash delivered through WeChat Payment. TEMP also has a computer-based website where the public can view and download reports (see TEMP, http://www.thuhjjc.com.)

### 2.3 Volunteer recruitment

The volunteers are recruited in two modes. In Mode 1, TEMP is expanded from a central group to the general public (Figure 1). Specifically, the university students recruited for this study post a link or a TEMP two-dimension QR code in their Moments and chat groups after logging into the platform through their WeChat accounts. Their friends who are interested in SWQM can either click the link or scan the code to be directed to TEMP. The platform then contacts these people upon their log in and submission of reports. TEMP cannot control the time of submission and the origin of the monitoring reports. The data are scattered under this mode.

In Mode 2, a group of professional citizens are recruited to monitor the quality of water in targeted sites. Those professionals who are working for environmental authorities and organizations are invited and motivated to register in TEMP. They are required to monitor those water bodies that they are familiar with and regularly submit their reports through the platform. In this mode, TEMP can recruit specific volunteers and collect data much faster than in Mode 1.

### 2.4 Data analysis

A new method is established to analyze the monitoring reports quantitatively. The smell, turbidity, floats, and integrated assessments reported by the volunteers are quantified and normalized between 0.0 and 1.0 according to their ranking scores. Water color is used for data cleansing and rumour control through a cross validation between the submitted descriptions and photos. Table 1 shows the indicators and their value ranges.

**Table 1: Normalization of indicators**

| Report items | Indicator | Data type | Qualifications |
|---|---|---|---|
| Water color | $C$ | Text | None |
| Smell | $S$ | Score from 0 to 10 | 1.0–0.0 |
| Turbidity | $T$ | Score from 0 to 10 | 1.0–0.0 |
| Floats | $F$ | Score from 0 to 10 | 1.0–0.0 |
| Integrated assessment | $I_a$ | Grand from 1 to 5 | 1.0, 0.75, 0.5, 0.25, 0.0 |

Note: The value of smell, turbidity, and floats ranges from 1.0 to 0.0. The indicator of integrated assessment is normalized across 1.0, 0.75, 0.5, 0.25, and 0.0, which correspond to the five grades of water quality assessment (i.e., excellent, good, bad, very bad, and worst).

## 2.5 Validation

The VGI data are validated by comparing the citizen-based reports with the gauged data. The reported turbidity data at Huayuankou station on the Yellow River in China are compared with the gauged data from the Yellow River Conservation Commission (YRCC). The Huayuankou station is one of the key stations along the main reach of Yellow River and is located where the middle and lower reaches are divided. The hydrological regime at this station presents an overview of the hydrological regime of the entire river basin.

The reports from all volunteers cannot be easily validated. These volunteers are distributed all over the country and submit their reports randomly upon seeing dirty water nearby. The reporting points rarely have an official gauge site nearby. TEMP employs a group of trained volunteers to report the water quality at Huayuankou station for every two or three days between March and April 2016, and these reports are used to provide site-specific VGI data for validation.

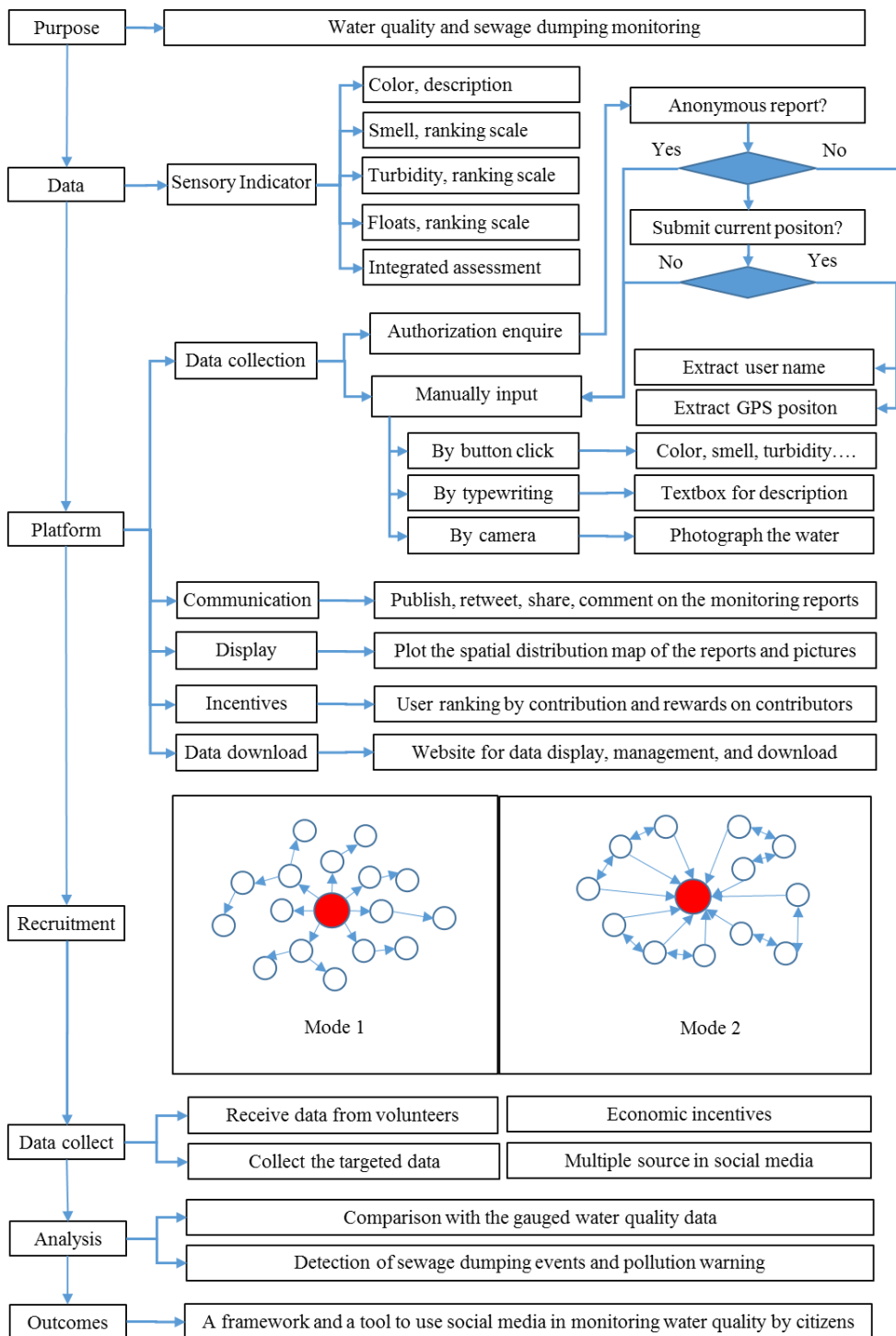



**Figure 1: Framework of monitoring water quality by VGI and social media.**

**3 Results**

**3.1 Water quality reports across China**

TEMP received 324 reports from volunteers in China between 12 October 2015 and 15 December 2016, among which 219 are used after data cleansing.

5   Table 2 presents an overview of the 219 reports that are collected from 30 provinces and municipalities in China. Approximately 30 reports are obtained from Beijing and Henan. The reports from Beijing are mainly submitted by the students of Tsinghua University, where the research group is based. The reports from Henan Province are mostly contributed by Henan-based professional volunteers working in the YRCC. Six provinces have submitted more than 10 reports. No reports have been collected from Xinjiang, Hainan, Taiwan, Macau, and the South China Sea.

10  **Table 2: Number of reports collected across China**

| No. | Provinces | Number of reports | | | Integrated assessment ($Ia$) | No. | Provinces | Number of reports | | | Integrated assessment ($Ia$) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Total | Anonymous | Photographed | | | | Total | Anonymous | Photographed | |
| 1 | Beijing | 41 | 26 | 9 | 0.53 | 18 | Hebei | 3 | 1 | 0 | 0.75 |
| 2 | Henan | 37 | 19 | 17 | 0.74 | 19 | Guangdong | 3 | 3 | 2 | 0.50 |
| 3 | Shanghai | 23 | 2 | 10 | 0.71 | 20 | Qinghai | 2 | 1 | 1 | 0.88 |
| 4 | Shandong | 17 | 1 | 17 | 0.49 | 21 | Sichuan | 2 | 2 | 0 | 0.75 |
| 5 | Chongqing | 12 | 6 | 11 | 0.52 | 22 | Neimeng | 2 | 1 | 0 | 0.38 |
| 6 | Jiangxi | 12 | 3 | 4 | 0.71 | 23 | Heilongjiang | 1 | 1 | 0 | 0.50 |
| 7 | Hubei | 8 | 3 | 3 | 0.63 | 24 | Xianggang | 1 | 1 | 1 | 1.00 |
| 8 | Fujian | 8 | 3 | 3 | 0.78 | 25 | Shannxi | 1 | 1 | 0 | 0.75 |
| 9 | Gansu | 7 | 2 | 6 | 0.75 | 26 | Ningxia | 1 | 1 | 0 | 0.75 |
| 10 | Tibet | 6 | 0 | 4 | 0.83 | 27 | Jilin | 1 | 1 | 0 | 0.75 |
| 11 | Yunnan | 5 | 1 | 4 | 0.85 | 28 | Hunan | 1 | 1 | 1 | 0.50 |
| 12 | Shanxi | 5 | 3 | 3 | 0.55 | 29 | Guangxi | 1 | 1 | 1 | 0.75 |
| 13 | Jiangsu | 4 | 1 | 3 | 0.88 | 30 | Anhui | 1 | 1 | 0 | 0.75 |
| 14 | Zhejiang | 4 | 1 | 3 | 0.63 | 31 | Xinjiang | / | / | / | |
| 15 | Guizhou | 4 | 1 | 0 | 0.88 | 32 | Hainan | / | / | / | |
| 16 | Tianjin | 3 | 3 | 2 | 0.75 | 33 | Taiwan | / | / | / | |
| 17 | Liaoning | 3 | 1 | 2 | 0.58 | 34 | Macau | / | / | / | |
| Total | | | | | | | | 219 | 92 | 107 | |

A total of 92 submitted reports do not indicate the names of the submitting WeChat users, which suggests that some volunteers prefer to submit anonymous reports to protect their privacy. Therefore, an anonymous function must be installed in TEMP to guarantee the privacy of the volunteers when they disclose the water pollution activities in their locations. A

total of 107 reports include photos of water, which significantly increase the credibility of these reports by providing substantial information for water quality analysis. However, 50% of the reports do not include any photos. The volunteers are also concerned about the charges for mobile Internet data usage, through which they upload their photos without Wi-Fi connection. Additional incentive measures must be implemented to encourage volunteers to upload photos of water. Figure 2 shows the total number of reports, while Figure 3 shows the number of anonymous reports and reports with photographs.

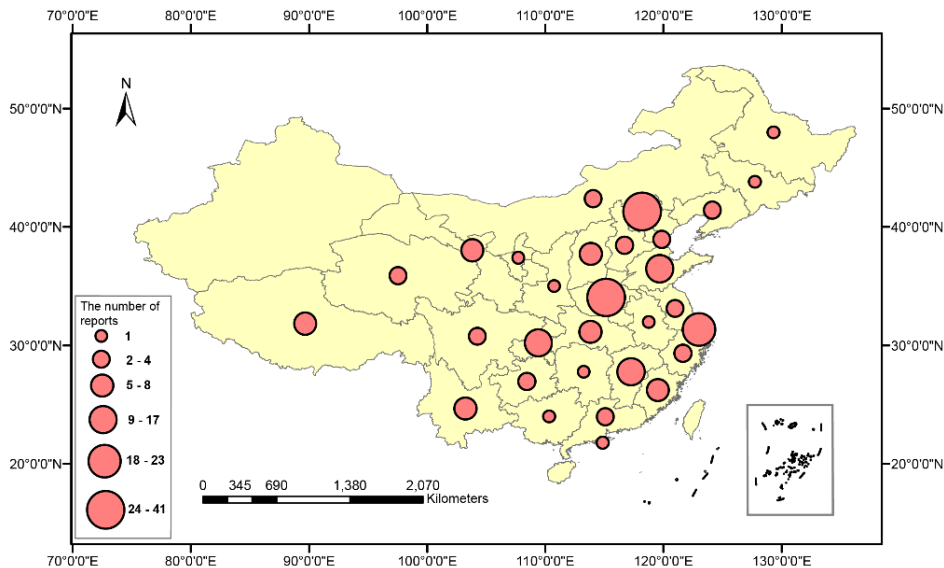



**Figure 2: Distribution of reports across the provinces and cities of China.**

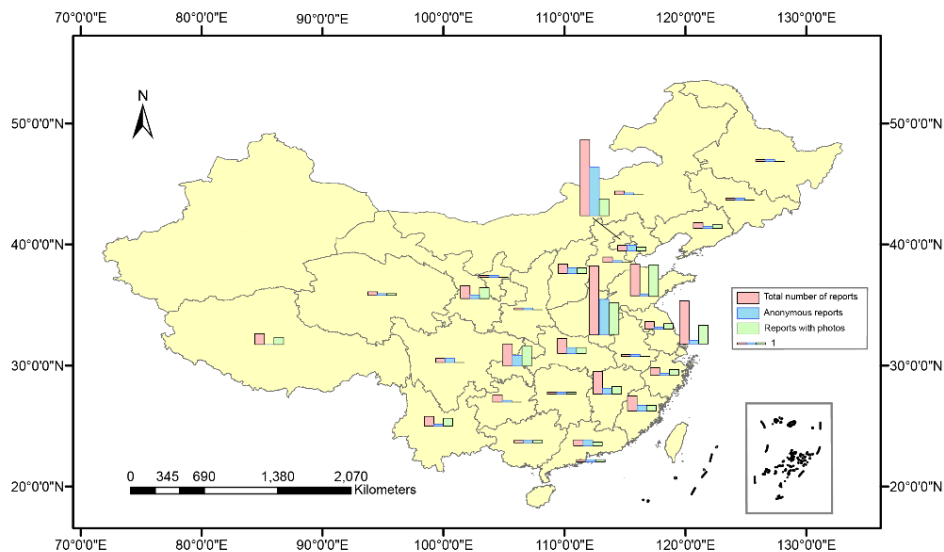



**Figure 3: Distribution of anonymous reports and reports with photographs.**

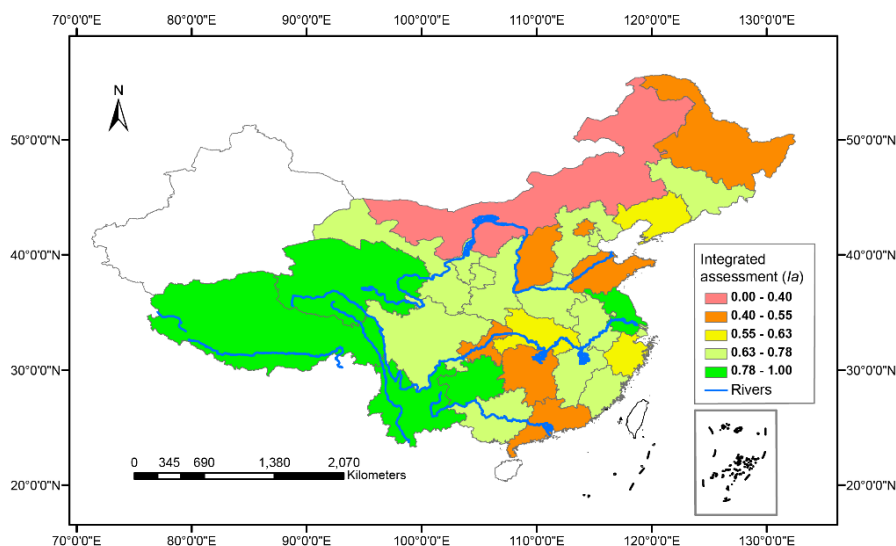
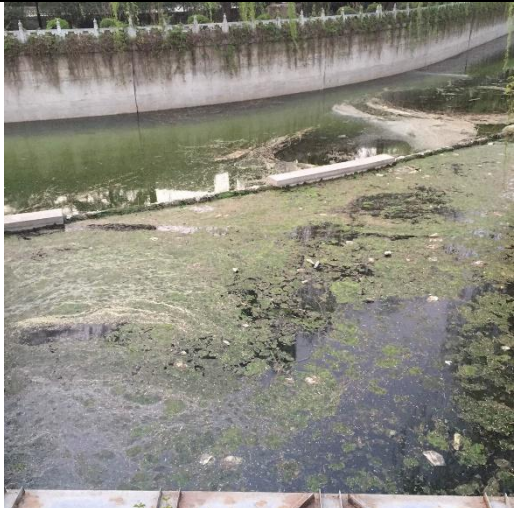



**Figure 4: The integrated assessment ($I_a$) results of China.**

5    Table 2 and Figure 4 present the average value of $I_a$ in each province. The reported water quality in the provinces located upstream Yellow River, Yangzi River, and Pearl River, such as Tibet, Qinghai, and Yunnan, is better than those observed at downstream provinces, such as Shandong and Guangzhou. This finding illustrates the water quality situation in China and the reasonableness of the VGI reports. However, citizen assessment cannot accurately represent the overall surface water quality in a region because the reports have insufficient coverage and frequency. The surface water quality of a whole region

10  can only be depicted if enough volunteers are involved and a sufficient number of reports is provided.

### 3.2 Examples of reports for pollution disclosure

Table 3 presents three reports with photographs that show algal blooms and water surface foam. Report 1 includes a photo of the river in Tsinghua University in Beijing. The river was polluted by domestic sewage and suffered from eutrophication. Report 2 describes the water quality in an unidentified river located in Fei Town, Linyi City, Shandong Province. The

15  accompanying photo shows that the water surface is covered by algae and rubbish. The reporter rated the water quality as very bad. Report 3 describes the water quality in Tianjing City. As shown in the accompanying photo, the water in the city had a black color and a bad quality.

None of the 107 photos collected by TEMP during the research period shows sewage water flowing into a river or lake. The sewage water dumping activities in China generally occur at night and in hidden locations that are rarely found by volunteers.

20  If a sufficient number of volunteers disclose the hidden sewage dumping activities in a region, then the water polluting activities can be determined by TEMP.

**Table 3: Examples of reports for pollution disclosure**

| No. | 1 | | | | | 2 | | | | | 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Report** | **Color** | **Smell** | **Turbidity** | **Floats** | $I_a$ | **Color** | **Smell** | **Turbidity** | **Floats** | $I_a$ | **Color** | **Smell** | **Turbidity** | **Floats** | $I_a$ |
| | Green | 3.0 | 5.0 | 7.0 | VB | Green | 3.0 | 6.0 | 5.0 | VB | Black | 0.0 | 7.0 | 7.0 | B |
| **Photo** |  | | | | |  | | | | |  | | | | |
| **Date** | 10 April 2016 | | | | | 15 April 2016 | | | | | 1 May 2016 | | | | |
| **Location** | Tsinghua University, Beijing City, China | | | | | Fei Town, Linyi City, Shandong Province | | | | | New coastal area, Tianjin City, China | | | | |

Note: B means bad, while VB means very bad.

### 3.3 Correlation between *Ia* and *S*, *T*, and *F*

Figures 5, 6, and 7 analyze the smell (*S*), turbidity (*T*), floats (*F*), and integrated assessment ($I_a$) of the water bodies presented in 107 reports with photographs. These reports are divided to five groups according to $I_a$, namely, 0.00, 0.25, 0.50, 0.75, and 1.00. The reports in the same group have the same value of $I_a$. Figures 5, 6, and 7 plot the minimum, maximum, and average *T*, *F*, and *S* for each group, respectively.

$I_a$ is highly correlated with *T* and *S*, which suggests that the VGI reports are reasonable. The volunteers have completed the reports based on their actual observations of the water. *T* has a higher correlation with $I_a$ than *F*. Water turbidity can be easily observed and greatly influence the judgement of volunteers compared with the other indicators. People tend to rate the quality of muddy water as bad despite the absence of floating objects.

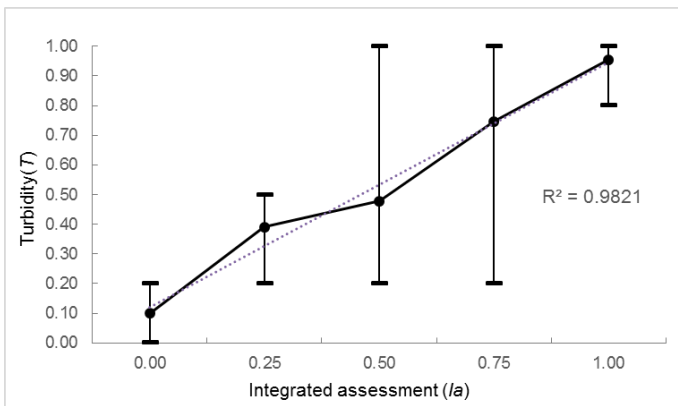



**Figure 5: The correlation between *Ia* and *T*.**          **Figure 6: The correlation between *Ia* and *F*.**

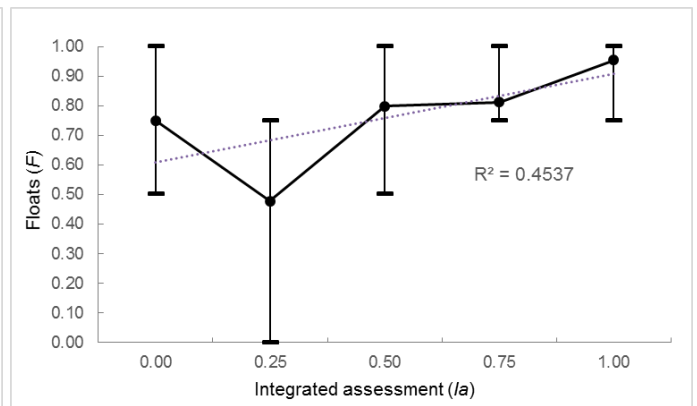
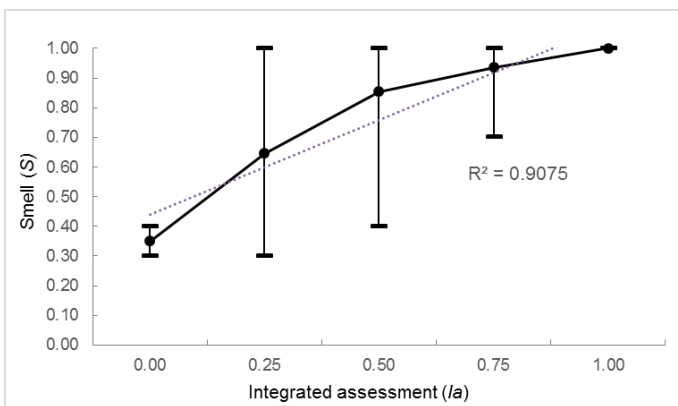



**Figure 7: The correlation between *Ia* and *S*.**

### 3.4 Validation with the gauged data

The *T* from the reports collected at Huayuankou station is compared with the gauged data (Figure 8). The reported and gauged data show similar temporal variations of *T*, which indicates the effectiveness of the VGI data to some extent. All these reports have accompanying photos and can be regarded as highly credible, thereby implying that the water quality assessments based on the perceptions of citizens are compatible with the actual situations.

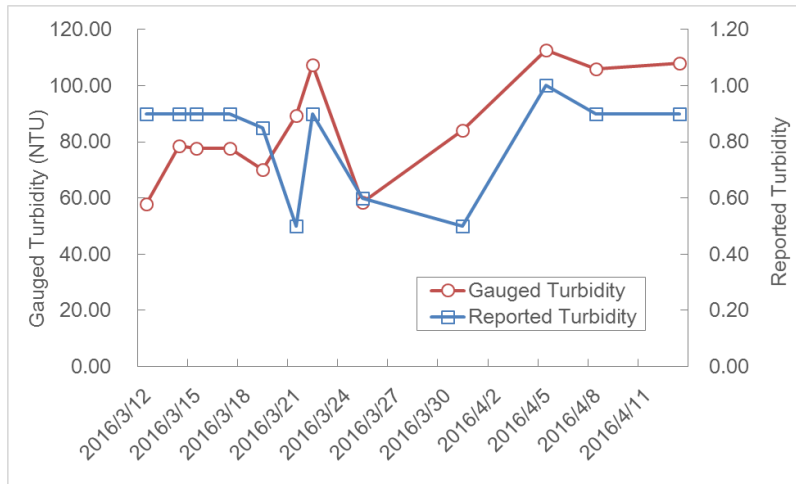



**Figure 8: The water turbidity from the reports and gauged data.**

### 5 Discussion

This study aims to develop a method for SWQM through volunteer citizens using a social media application. A framework is also established to guide the application design, volunteer recruitment, data collection, and report analysis. The TEMP application is built based on WeChat, through which TEMP users can describe and take photos of river and lake waters following the TEMP instructions. Users can also report the surface water pollution activities that affect their living and health through this platform.

A total of 219 validated reports are analyzed in this study. These reports are collected from 140 volunteers across 30 provinces and cities in China. These volunteers assess water quality based on their sensory organs, particularly through their observations on the smell, turbidity, and floating matters on the water. However, people may have varying perceptions on water quality and may provide different assessment reports on the water from the same site. The water assessment results from different sites cannot be easily compared because the reports from citizens are subjective to some degree.

The credibility of the reports presents a major concern for this study. Therefore, identifying whether the reports are real and whether the volunteers have generated these reports based on their observations is necessary. This study follows three criteria to wash the data. First, a report that specifies the exact GPS location of the assessed water body is considered credible. The GPS information of a water body is automatically abstracted from the mobile devices of the reporters upon the submission of their reports. Second, if the reports are submitted several times within a short period and most of these reports have come from the same volunteer located on the same site, then these reports have low credibility and are mostly assumed to be test reports submitted by new volunteers. Third, those reports with accompanying photos are considered the most credible. A total of 324 reports are cleansed following these criteria, and 219 reports have been validated. A total of 107 photograph reports have been rated as highly credible.

Validating the reports presents a tricky issue. The reports are submitted from scattered sites with insufficient gauged data for validation. TEMP recruits a group of professional volunteers from YRCC who continuously report the quality of water at the Huayuankou station where gauged data are available. These volunteers are familiar with the quality of water on the site, but are not oriented to the gauged data during their submission of reports. The gauged results can be assessed at least one day after sampling at the station because the water sample must be analyzed beforehand in a laboratory. The volunteers can only assess the gauged data after submitting their reports. TEMP records the time of the reports according to the clock on the TEMP server. The time of the report cannot be modified by the volunteer. TEMP has only received 13 reliable photograph reports at the station between March and April 2016. Despite the limited data, the validation indicates that the VGI data are valuable for water quality monitoring to some extent. The assessment by the citizens effectively indicates the water quality status if the reporter/citizen is relatively trained in water quality monitoring. Fore et al. (2001), Monk et al. (2008), Flanagin and Metzger (2008) reported similar findings.

The motivation of the volunteers to submit data proactively presents another major concern. Various VGI studies consider the motivation of volunteers as a key factor to the success of the VGI program (Werts et al., 2012). Coleman et al. (2009), as cited in (Werts et al., 2012, p. 817), identified altruism, professional or personal interest, intellectual stimulation, protection or enhancement of a personal investment, social reward, enhanced personal reputation, outlet for creative and independent self- expression, and pride of location as key motivators. Some citizens may be motivated by the perceived instrumentality in promoting change (Hertel et al., 2003). Budhathoki et al. (2010), as cited in (Werts et al., 2012, p. 817), identified fun, learning, and instrumentality as primary motivators for geographic information contributors, and noted that "when contributors see their data appear visually on maps, they receive deep satisfaction."

In this study, the core members in chat groups have key roles in motivating the contributors by leading the communication process and reminding the volunteers to submit their reports. Some economic incentives have also effectively increased the number of contributors and their contributions. Figure 9 shows the number of volunteers involved in this study. The TEMP

application has been tested since April 2015. After its development and testing, TEMP was promoted by the faculty, staff, and students of Tsinghua University though their WeChat Moments. The number of users steadily increased until the official launch of TEMP in the opening ceremony of the Tsinghua Remote Sensing and Big Data Research Centre (http://hydrosky.org/Newscon/index/id/581/aid/544445113) in October 2015, during which a sharp increase in the number of

5    TEMP users was observed. The professionals and other individuals related to the research center have registered in TEMP, thereby explaining the sudden increase in the number of users reported in October 2015. Since then, these TEMP users continuously promoted and advertised the platform in their WeChat Moments, thereby motivating more users to register. The number of reports gradually increased because of the limited incentives provided to the volunteers before March 2016.

10   In March 2016, Mode 2 was adopted to recruit volunteers from the YRCC and students from universities in Beijing. Several volunteer chat groups were established in WeChat, and a core member was assigned in each group to lead the communication process and remind the users to submit their reports. The number of reports substantially increased after implementing these initiatives. The platform also began to offer economic incentives and rewards in April 2016. The core members in the chat groups sent "Red Packet" money to contributors through WeChat Payment. The core members

15   transferred the money (usually 100 RMB) to the members of their chat groups. The "Red Packet" money can be sent within a chat group in a similar way as sending photos. This money can also be obtained and shared by dozens of members who must tap on an image of the "Red Packet" money on their screens as fast as they can. The first to tap on the image receives a random share of the total "Red Packet" money. Those members who receive the money were considerably motivated to submit more reports and invite more friends to register and submit data. Since May 2016, these economic incentives

20   continuously motivated the volunteers and increased the number of reports being submitted to the platform.

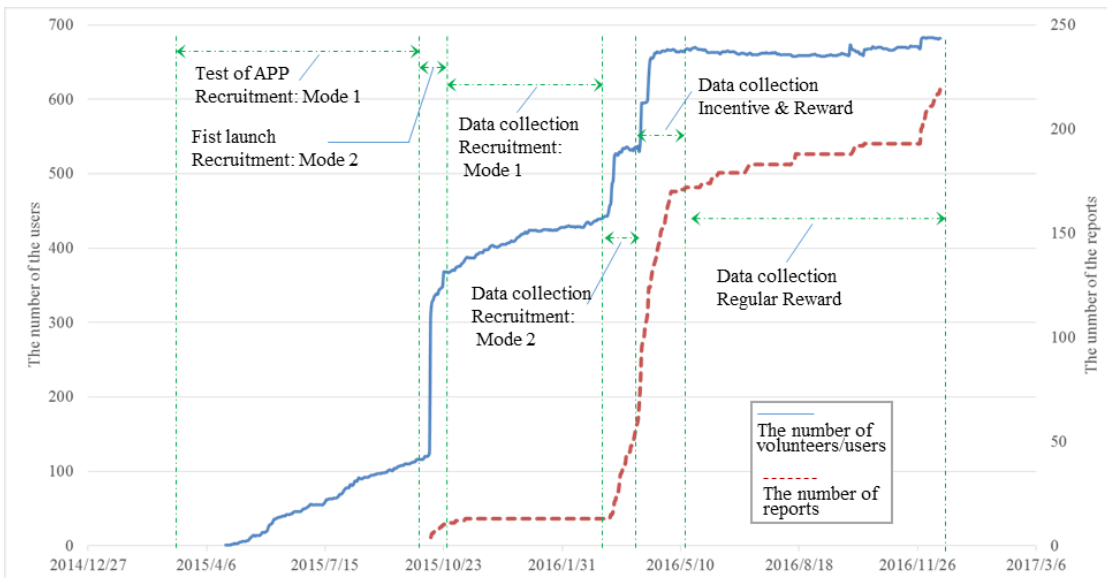



Figure 9: Increase in the number of reports and reporters.

The volunteers were also granted economic rewards based on their contributions rather than on how fast they tap the image of a "Red Packet" on the screen of their mobile devices. Figure 10 shows the distribution of reports versus reporters. About 50% of the reporters submitted only one report during the entire research period, while 5% have submitted more than 10 reports. Since 2016, TEMP offered those reporters with top-ranking submissions with monetary incentives sent through WeChat Payment.

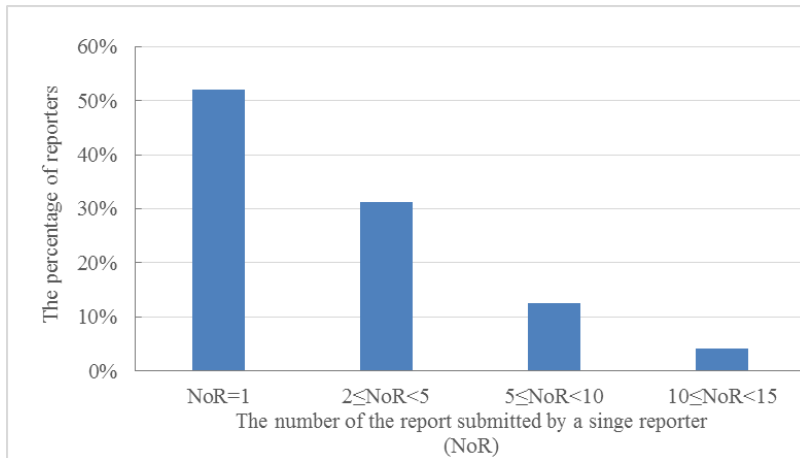



**Figure 10: Distribution of reports versus reporters.**

Future related research must develop other methods for data validation and analysis as well as collect data from other sources, including Twitter, Facebook, and Sina Weibo (Jiang et al., 2015). People tend to post or tweet a text or image in social media to complain about water pollution upon seeing a dirty river or lake (Kaplan and Haenlein, 2010). In this case, social media users do not intend to report water polluting activities, but inadvertently provide the necessary data for water quality monitoring. The water-quality-related text and photos collected from Twitter, Facebook, and Sina Weibo can provide high-density, massive information because of the millions of active users in these social media platforms. This process saves much effort in recruiting and motivating volunteers, although the collected data tend to be unstructured (Poser and Dransch, 2010). TEMP contacts and guides the volunteers in submitting structural reports and data. However, recruiting and motivating these volunteers require much effort. Although providing economic incentives can effectively motivate these volunteers, such method has been proven unsustainable. Cooperating with non-government organizations (NGOs) that are focused on environmental protection may present an alternative approach for recruiting volunteers. These NGO members must also be interested in TEMP and have an enduring and strong motivation to disclose water polluting activities.

The data validation process can become more robust if many volunteers are involved and extensive reports are obtained. The data can also be supplemented by satellite and aerial remote sensing or sensor system streaming. Thereafter, the big data method (Hampton et al., 2013) can be applied to improve the accuracy of water quality monitoring if high-density data are used. This study proposes an approach for collecting citizen reports on water quality, which is the first step in applying the big data method in environmental governance (Perez et al. 2015; McCall 2003).

There is also an opportunity to develop innovative devices for validation. Minkman,van Overloop, and van der Sanden (2015) explored a crowd sensing mobile applications for measuring water quality in the Netherlands. It consists of a camera- based colorimetric analysis of the test strips. The indicator test strips are photographed, and the color of the strips is analyzed automatically by a smartphone application to obtain chemical water quality data. Snik et al. (2014) developed the iSPEX, a low-cost, mass-producible optical add-on for smartphones with a corresponding application. People can purchase the iSPEX smartphone accessory that must be installed on top of their iPhones to measure particulate matter in the atmosphere (Minkman, 2015). These devices are crucial in citizen-based water quality monitoring and in enhancing the credibility of the data. The devices for water quality monitoring must be smart, portable, and convenient to be used with smartphones (e.g., an external device using laser to detect water quality) (Chen et al., 2015).

## 6 Conclusion

This study proposes a methodological framework for SWQM using social media. The selection of water quality indicators, the application design guide, the volunteer recruitment methods, and the data collection, cleansing, and analysis processes are discussed accordingly. The TEMP application is established based on WeChat, a popular social media platform in China. TEMP allows registered users to submit their descriptions and photos of rivers or lakes anonymously or non-anonymously. These photos are automatically geo-tagged with the GPS information of the sites as recorded by the mobile devices of the submitters.

TEMP received 324 reports from 30 provinces and cities in China between 12 October 2015 and 15 December 2016. Among these reports, 219 reports are used after data cleansing. The distribution analysis of these reports emphasizes the importance of installing privacy and photograph functions in TEMP. Over 42% of the 219 reports are submitted by anonymous users, which suggests that people care about their privacy when reporting water polluting activities within their vicinities. A total of 107 photos of rivers and lakes are collected through TEMP, and these photos provide extensive information for pollution detection. Thirteen reports with photographs are collected from the Huayuankou station in the Yellow River and have been validated by comparing the reported turbidity with the gauged value. These reports indicate that the citizen-based water quality data are relatively credible if the volunteers are trained on water quality monitoring.

This paper also discusses data quality and the motivation of the volunteers. The data are cleansed based on the location, time, and photos in the reports. Two modes for volunteer recruitment are adopted. Mode 2 can increase the number of volunteers within a short period. An economic incentive mechanism is also implemented to motivate the volunteers to contribute data under the guidance of the core members of their chat groups.

Future studies must collect additional data and validate the collected reports. The unpremeditated data on water quality that are collected from Twitter, Facebook, and Sina Weibo must be abstracted to increase the data volume. The big data method can be used when a sufficient amount of data is available.

### Data availability

10    The data underlying this research can be accessed publicly. All these data can be downloaded from www.thuhjjc.com, a website launched by the authors to display and download data from the TEMP platform.

### Author contribution

Dr. Hang Zheng designed the framework of this study, analyzed the data, and prepared the manuscript with contributions
15    from all co-authors. Professor Hong Yang designed the interface of the TEMP platform and contributed to the Discussion section. Dr. Di Long performed the data collection. Mr. Jing Hua developed the main functions of the TEMP platform.

### Competing interests

The authors declare no conflict of interest.

### Acknowledgement

20    This research is supported by the National Natural Science Foundation of China (51479089) and (51323014), the 2014 Chinese Ministry of Water Resource Program (201401031), the 13th Five-Year Research Program from the Ministry of Science and Technology, China (SQ2016YFSF020003), and the State Key Laboratory of Hydroscience and Engineering, Tsinghua University (2014-KY-04).

25

## References

Bishr, M. and Mantelas, L.: A trust and reputation model for filtering and classifying knowledge about urban growth, GeoJournal, 72(3–4), 229–237, doi:10.1007/s10708-008-9182-4, 2008.

Budhathoki, N. R., NedovićBudić, Z. and Bruce, B. C.: An Interdisciplinary Frame for Understanding Volunteered Geographic Informaion, Geomatica, 64(1), 11–26, 2010.

Bunce, S., Partridge, H. and Davis, K.: Exploring information experience using social media during the 2011 Queensland Floods: a pilot study, Aust. Libr. J., 61(1), 34–45, doi:10.1080/00049670.2012.10722300, 2012.

Chen, P., Pan, D., Mao, Z. and Tao, B.: Detection of water quality parameters in Hangzhou Bay using a portable laser fluorometer, Mar. Pollut. Bull., 93(1–2), 163–171, doi:10.1016/j.marpolbul.2015.01.023, 2015.

Coleman, D. J., Georgiadou, Y., Labonte, J., Observation, E. and Canada, N. R.: Volunteered Geographic Information : The Nature and Motivation of Produsers, Int. J. Spat. Data Infrastructures Res., 4, 332–358, doi:10.2902/1725-0463.2009.04.art16, 2009.

Devlin, M., Waterhouse, J. and Brodie, J.: Community and connectivity: summary of a community based monitoring program set up to assess the movement of nutrients and sediments into the Great Barrier Reef during high flow events, Water Sci. Technol., 43(9), 121–132, 2001.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T. and Purcell, K.: The current state of citizen science as a tool for ecological research and public engagement, Front. Ecol. Environ., 10(6), 291–297, doi:10.1890/110236, 2012.

E. Minkman; P.J. van Overloop; M.C.A. van der Sanden: Citizen science in water quality monitoring: mobile crowd sensing for water management in The Netherlands, in World Environmental and Water Resources Congress 2015: Floods, Droughts, and Ecosystems, pp. 1399–1408., 2015.

Flanagin, A. J. and Metzger, M. J.: The credibility of volunteered geographic information, GeoJournal, 72(3–4), 137–148, doi:10.1007/s10708-008-9188-y, 2008.

Fore, L. S., Paulsen, K. and O'Laughlin, K.: Assessing the performance of volunteers in monitoring streams, Freshw. Biol., 46(1), 109–123, doi:10.1046/j.1365-2427.2001.00640.x, 2001.

Goodchild, M. F. and Glennon, J. A.: Crowdsourcing geographic information for disaster response: a research frontier, Int. J. Digit. Earth, 3(3), 231–241, doi:10.1080/17538941003759255, 2010.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S. and Porter, J. H.: Big data and the future of ecology, Front. Ecol. Environ., 11(3), 156–162, doi:10.1890/120103, 2013.

Hertel, G., Niedner, S. and Herrmann, S.: Motivation of software developers in open source projects: An Internet-based survey of contributors to the Linux kernel, Res. Policy, 32(7), 1159–1177, doi:10.1016/S0048-7333(03)00047-7, 2003.

Horowitz, A. J.: A review of selected inorganic surface water quality-monitoring practices: Are we really measuring what we think, and if so, are we doing it right?, Environ. Sci. Technol., 47(6), 2471–2486, doi:10.1021/es304058q, 2013.

Intelligence, B.: WeChat pay fees kick in, Bus. Insid. [online] Available from: http://www.businessinsider.com/wechat-pay-fees-kick-in-2016-5/?r=AU&IR=T (Accessed 4 May 2016), 2016.

Jiang, W., Wang, Y., Tsou, M.-H. and Fu, X.: Using Social Media to Detect Outdoor Air Pollution and Monitor Air Quality Index (AQI): A Geo-Targeted Spatiotemporal Analysis Framework with Sina Weibo (Chinese Twitter), PLoS One, 10(10), e0141185, doi:10.1371/journal.pone.0141185, 2015.

Kaplan, A. M. and Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media, Bus. Horiz., 53(1), 59–68, doi:10.1016/j.bushor.2009.09.003, 2010.

Khalil, B. and Ouarda, T. B. M. J.: Statistical approaches used to assess and redesign surface water-quality-monitoring networks., J. Environ. Monit., 11(11), 1915–1929, doi:10.1039/b909521g, 2009.

Khalil, B., Ouarda, T. B. M. J., St-Hilaire, A. and Chebana, F.: A statistical approach for the rationalization of water quality indicators in surface water quality monitoring networks, J. Hydrol., 386(1–4), 173–185, doi:10.1016/j.jhydrol.2010.03.019, 2010.

Kim, T.: Observation on copying and pasting behavior during the Tohoku earthquake: Retweet pattern changes, Int. J. Inf. Manage., 34(4), 546–555, doi:10.1016/j.ijinfomgt.2014.03.001, 2014.

Kingham, N.: Environmental action for community monitoring, Water Sci. Technol., 45(11), 177–184, 2002.

Lee, J., Binkley, M. R. and Carlson, R. E.: The Great American Secchi Dip-In: GIS contributes to national snapshot of lake-water transparency, GIS World, 10(8), 42–43, 1997.

Linda E. Kruger, Margaret A. Shanno: Getting to Know Ourselves and Our Places Through Participation in Civic Social Assessment, Soc. Nat. Resour., 13(5), 461–478, doi:10.1080/089419200403866, 2000.

Lindsay, B. R.: Social Media and Disasters: Current Uses, Future Options and Policy Considerations., Congr. Res. Serv. Reports, 13 [online] Available from: http://fas.org/sgp/crs/homesec/R41987.pdf, 2011.

Lwin, K. K., Zettsu, K. and Sugiura, K.: Geovisualization and correlation analysis between geotagged Twitter and JMA rainfall data: Case of heavy rain disaster in Hiroshima, Spat. Data Min. Geogr. Knowl. Serv. (ICSDM), 2015 2nd IEEE Int. Conf., 71–76, doi:10.1109/ICSDM.2015.7298028, 2015.

McCall, M. K.: Seeking good governance in participatory-GIS: A review of processes and governance dimensions in applying GIS to participatory spatial planning, Habitat Int., 27(4), 549–573, doi:10.1016/S0197-3975(03)00005-5, 2003.

Mersham, G.: Social Media and Public Information Management: The September 2009 Tsunami Threat to New Zealand, Media Int. Aust. Inc. Cult. Policy, (137), 130–143, doi:10.1177/1329878X1013700115, 2010.

Metzger, E. S. and Lendvay, J. M.: Seeking environmental justice through public participation: A community-based water quality assessment in Bayview Hunters Point, Environ. Pract., 8(2), 104–114, doi:doi:10.1017/S1466046606060133, 2006.

Minkman, E.: Citizen Science in Water Quality Monitoring: Developing Guidelines for Dutch Water Authorities for Contributory Mobile Crowd Sensing, Delft University of Technology., 2015.

Monk, J., Ierodiaconou, D., Bellgrove, A. and Laurenson, L.: Using community-based monitoring with GIS to create habitat maps for a marine protected area in Australia, J. Mar. Biol. Assoc. United Kingdom, 88(5), 865–871, doi:10.1017/s0025315408001835, 2008.

Nicholson, E., Ryan, J. and Hodgkins, D.: Community data - Where does the value lie? Assessing confidence limits of community collected water quality data, Water Sci. Technol., 45(11), 193–200, 2002.

Perez, P., Chemin, T. H. Du, Turpin, E. and Clarke, R.: Citizen-Driven Flood Mapping in Jakarta: A Self-Organising Socio-technical System, 2015 IEEE Int. Conf. Self-Adaptive Self-Organizing Syst. Work., 174–178, doi:10.1109/SASOW.2015.40, 2015.

Poser, K. and Dransch, D.: Volunteered geographic information for disaster management with application to rapid flood damage estimation, Geomatica, 64(1), 89–98, 2010.

Rose, N. L., Turner, S. D., Goldsmith, B., Gosling, L. and Davidson, T.: Quality control in public participation assessments of water quality: The OPAL Water Survey, BMC Ecol., 16(Suppl 1, S14, doi:10.1186/s12898-016-0063-2, 2015.

Sakaki, T.: Earthquake Shakes Twitter Users : Real-time Event Detection by Social Sensors, , 851–860, doi:10.1145/1772690.1772777, 2009.

Sanders, T. G., Ward, R. C., Loftis, J. C., Steele, T. D., Adrian, D. D. and Yevjevich, V.: Design of network for monitoring water quality, Water Resources Publication, Littleton, Colorado., 1983.

Seeger, C. J.: The role of facilitated volunteered geographic information in the landscape planning and site design process, GeoJournal, 72(3–4), 199–213, doi:10.1007/s10708-008-9184-2, 2008.

Silvertown, J.: A new dawn for citizen science., Trends Ecol. Evol., 24(9), 467–71, doi:10.1016/j.tree.2009.03.017, 2009.

Slavkovikj, V., Verstockt, S., Van Hoecke, S. and Van De Walle, R.: Review of wildfire detection using social media, Fire Saf. J., 68, 109–118, doi:10.1016/j.firesaf.2014.05.021, 2014.

Snik, F., Rietjens, J. H. H., Apituley, A., Volten, H., Mijling, B., Di Noia, A., Heikamp, S., Heinsbroek, R. C., Hasekamp, O. P., Smit, J. M., Vonk, J., Stam, D. M., van Harten, G., de Boer, J., Keller, C. U. and iSPEX Citizen Scientists, 3187: Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters, Geophys. Res. Lett., 41(20), 7351–7358, doi:10.1002/2014GL061462, 2014.

Strobl, R. O. and Robillard, P. D.: Network design for water quality monitoring of surface freshwaters: A review, J. Environ. Manage., 87(4), 639–648, doi:10.1016/j.jenvman.2007.03.001, 2008.

Tang, Z., Zhang, L., Xu, F. and Vo, H.: Examining the role of social media in California's drought risk management in 2014, Nat. Hazards, 79(1), 171–193, doi:10.1007/s11069-015-1835-2, 2015.

Tencent: WeChat - Features, [online] Available from: http://www.wechat.com/en/features.html#live_chat (Accessed 9 July 2016), 2016.

Tulloch, D. L.: Is VGI participation? From vernal pools to video games, GeoJournal, 72(3–4), 161–171, doi:10.1007/s10708-008-9185-1, 2008.

Virtual Social Media Working Group: Lessons learned: Social media and hurricane Sandy, , (June) [online] Available from: https://communities.firstresponder.gov/DHS_VSMWG_Lessons_Learned_Social_Media_and_Hurricane_Sandy_Formatted _June_2013_FINAL.pdf, 2013.

Wang, H., Wu, M., Deng, Y., Tang, C. and Yang, R.: Surface water quality monitoring site optimization for poyang lake, The largest freshwater lake in China, Int. J. Environ. Res. Public Health, 11(11), 11833–11845, doi:10.3390/ijerph111111833, 2014.

Wei, H.: Officials failing to stop textile factories dumping waste in Qiantang River, Chinadialogue [online] Available from: https://www.chinadialogue.net/article/show/single/en/5589-Officials-failing-to-stop-textile-factories-dumping-waste-in-Qiantang-River (Accessed 8 January 2013), 2013.

Werts, J. D., Mikhailova, E. a., Post, C. J. and Sharp, J. L.: An integrated WebGIS framework for volunteered geographic information and social media in soil and water conservation., Environ. Manage., 49, 816–832, doi:10.1007/s00267-012-9818-5, 2012.

Wikipedia: WeChat, [online] Available from: https://en.wikipedia.org/wiki/WeChat#cite_note-35 (Accessed 9 July 2016), 2016.

**Figure 1: Framework of monitoring water quality by VGI and social media.**

**Figure 2: Distribution of reports across the provinces and cities of China.**

**Figure 3: Distribution of anonymous reports and reports with photographs.**

**Figure 4: The integrated assessment ($I_a$) results of China.**

**Figure 5: The correlation between *Ia* and *T*.**

**Figure 6: The correlation between *Ia* and *F*.**

**Figure 7: The correlation between *Ia* and *S*.**

**Figure 8: The water turbidity from the reports and gauged data.**

**Figure 9: Increase in the number of reports and reporters.**

**Figure 10: Distribution of reports versus reporters.**