

Editor Decision: Reconsider after major revisions (further review by Editor and Referees) (14 Dec 2016) by Pieter van der Zaag

Comments to the Author:

I find the topic of the paper very interesting and potentially very valuable. However, the paper needs a major revision because of the following four comments, whereby my Comment 1 is the most important.

Comment 1

I agree with the two reviewers that on the one hand, the validation of the methodology in the original manuscript lacked robustness, but that on the other hand, without such a validation the manuscript does not have sufficient content to be publishable in HESS.

I therefore encourage the authors to see whether there are any gauged water quality data that can be linked to either one of the three parameters that have been reported by citizens, i.e. either C (water colour), or S (smell) or T (turbidity). If so, then the gauged data could be compared with the citizen contributed data.

I wasn't convinced by the method adopted namely to compare the three citizen reported parameters (S, T, F) with the combined indicator Q that collapses the scores for E, N and D through averaging.

I would think that turbidity would often be routinely measured; if so that could then be a way to validate, albeit on only one parameter, the VGI method discussed in the paper.

So just throwing out the validation section altogether is not the way forward to get the paper published in HESS.

Comment 2

I do not understand the usefulness of equation 1, which combines all information on different parameters into one aggregate indicator Q through averaging. The question is: does this create additional information or does this in fact result in a loss of information. Also, why average the scores for C, S, and T? One could also aggregate in a different way, such as Min (S,T,F) or Max (S,T,F). And why combine the, arguably more "subjective", indicator for integrated assessment with that of the three parameters? And if we have to combine them, again why take the average?

I would find it much more interesting to compare the scores for S with Ia, F with Ia and F with Ia, and see whether there are differences in correlations, and whether there can be explanations for these differences (or similarities).

Comment 3

In my opinion Figure 9 adds little information, and can in my view be omitted.

Comment 4

There is need for strict language editing by a native English speaker.

I hope the authors consider my decision in a constructive light. I encourage them to find a way to validate the claim that citizens can indeed contribute data that are valuable in terms of water quality monitoring.

Responses to the comments:

Thanks a lot for the comments. We truly agree that it is important to increase the robustness of validation. We have provided more data and analysis for method validation.

To comment 1:

We tried to improve the validation by providing more data, comparing the reported water quality between upstream and downstream regions, analyzing the correlations between S , T , F and Ia , and comparing the reported turbidity with the gauged data.

- (1) The collected data has been extended to 15 December 2016. The TEMP is continuously receiving the water quality reports during 2016. Comparing with the former manuscript, more data is provided for validation. 219 reports are analyzed after data cleansing.
- (2) The Figure 4 in the revised manuscript presents the average value of Ia in each province of China. 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 fits the general situation of water quality along a river. This finding also illustrates the water quality situation in China and the reasonableness of the VGI reports.
- (3) The correlations between S , T , F and Ia are presented in the revised manuscript. Figures 5, 6, and 7 analyze the smell (S), turbidity (T), floats (F), and integrated assessment (Ia) of the water bodies presented in 107 reports with photographs. These reports are divided to five groups according to Ia , namely, 0.00, 0.25, 0.50, 0.75, and 1.00. The reports in the same group have the same value of Ia . Figures 5, 6, and 7 plot the minimum, maximum, and average T , F , and S for each group, respectively. Ia is highly correlated with T , 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 Ia than F . Water turbidity can be easily observed and greatly influence the judgment of volunteers compared with the other indicators. People tend to rate the quality of muddy water as bad despite the absence of floating objects.
- (4) The reported turbidity is compared with the gauged data at Huayuankou station on the Yellow River. The reported and gauged data show similar temporal variations of T , which indicates the effectiveness of the VGI data to some extent.

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.

To comment 2:

Thanks again for the comment. We totally agree that it can provide more information to compare the scores of S with Ia , T with Ia and F with Ia . We analyze the correlations between $\text{Min}(S)$, $\text{Average}(S)$, $\text{Max}(S)$ and Ia ; $\text{Min}(T)$, $\text{Average}(T)$, $\text{Max}(T)$ and Ia ; $\text{Min}(F)$, $\text{Average}(F)$, $\text{Max}(F)$ and Ia . See Figures 5, 6, and 7 in the revised manuscript. The results show that the F is highly correlated with Ia . This suggests that the turbidity is the key element effecting people's sense on water quality. Water turbidity can be easily observed and greatly influence the judgment of volunteers compared with the other indicators. This also demonstrates that the VGI reports are reasonable, because the volunteers indeed completed the reports based on their actual observations of the water.

To Comment 3

We have removed the Figure 9 from the former manuscript. .

To Comment 4

The revised manuscripts is edited by a native English speaker.

The department of author Hang Zheng, has change to:

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

This change has been accepted by all the other authors.

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

Hang Zheng¹, Hong Yang¹, Di Long¹, Hua Jing¹

¹State Key Lab of Hydrosience and Engineering, Dept. of Hydraulic Engineering, Tsinghua Univ., Beijing 100084, China

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

10 **Abstract.** Surface water quality monitoring (SWQM) provides essential information for water environmental protection. However, SWQM is costly and is limited in terms of equipment and sites. The global popularity of social media and intelligent mobile devices with GPS and photograph functions provide exciting opportunity for citizens to monitor surface water quality. This study aims to establish a method to monitor surface water quality using social media platforms. A WeChat-based application platform is built to collect water quality reports from volunteers. Results show that the citizen-based data is valuable for water quality monitoring. The key application functions and the methods for data cleaning and volunteer recruitment are discussed basing on the reports. The proposed framework and methods can provide a framework mechanism to collect water quality data from citizens and offer a primary foundation for big data analysis in future research.

15 **Keywords.** Water quality; monitoring; social media; volunteered geographic information.

1 Introduction

Surface freshwater is a finite resource that is essential for human and ecosystem existence. An adequate quantity and quality of water are essential for sustainable development (Khalil and Ouarda, 2009). However, many surface water systems have been contaminated by treated or untreated wastewater discharged from domestic, industrial, and agricultural water users.

20 Water quality has become an important component of the global water scarcity and water crisis.

The degradation of the surface water system has increased the need to determine the status of water quality to detect water pollution, thereby 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). Monitoring sites, frequency, variables and instruments, and trained/educated field personnel are required in a water quality monitoring network. Evidently, establishing a surface water quality monitoring (SWQM) network in a broad area is costly (Horowitz, 2013). For example, the US Geological Survey runs the Mississippi River Basin monitoring network to address the gulf coast land loss and hypoxia. Currently, collecting a single sample at a site costs between US\$4,000 and US\$6,000. Analyses

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font: (Default) Times New Roman, Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

of various physical/chemical parameters add US\$1,500 to US\$2,000 more per sample (Horowitz, 2013). These costs limit the sample numbers/sites that can be monitored. The limited resources necessitate the monitoring program's installation of a limited number of monitors on sites and samples at regular temporal intervals. These limitations reduce the capability of the monitoring program to detect illegal polluting activities, such as hidden sewage dumping, which tend to occur in areas that are distant from the monitoring sites or at a time when no sampling has been conducted. For example, many Chinese industrial facilities dump sewage water discharge in rivers in the middle of the night to avoid detection (e.g. Wei, 2013).

A participatory monitoring approach by citizens could fill the spatiotemporal gaps of the current monitoring network. Wei (2013) explained that two ordinary citizens recorded sewage dumping in a river near their home in China from 2004 to 2007. Collecting samples and taking photos for four years enabled the two citizens to document dozens of instances of discoloured, foul-smelling, or warm water in the river. These voluntary records imply that citizens who are affected by the sewage discharge have a strong motivation to monitor and report the pollution. If there are more volunteer reporters (e.g., the two citizens) come forward, the more possibility to detect the hidden sewage dumping or pollution.

Water quality can be defined in terms of anything from one variable to hundreds of compounds and for multiple usages (Khalil et al., 2010). To collect evidence of hidden or midnight sewage dumping, what citizens see and photograph appears to be effective. Citizens without professional equipment for water quality analysis can describe the physical characteristics of the water (e.g., colour, smell, and temperature) to assess the water quality and disclose the pollution. Voluntary reporting is more flexible and effective and less costly compared with the traditional monitoring program operated by the government.

Thus, we argue that the volunteered geographic information (VGI) in the citizen science context provides a proximate sensing solution for some water conservation issues. VGI has recently provided an interesting alternative to traditional authoritative information from mapping agencies and corporations (Goodchild and Glennon, 2010). VGI is also identified as "collaboratively contributed geographic information" (Bishr and Mantelas, 2008) in the context of participatory 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 to "crowd source" various scientific tasks (i.e., from data collection to discovery) (Dickinson et al., 2012). Citizen science is the process whereby citizens are involved in science as researchers (Linda E. Kruger, Margaret A. Shanno, 2000). A citizen scientist is a volunteer who collects or processes data as a component of a scientific enquiry. Currently, citizen scientists participate in projects on climate change, invasive species, conservation biology, ecological restoration, water quality monitoring, population ecology, and different types of monitoring (Silvertown, 2009). VGI in the context of citizen science can be produced immediately and may determine environmental changes nearly as rapidly as they occur. VGI is recognized

as an innovative approach for an improving environmental governance by fostering accountability, transparency, legitimacy, and other dimensions of governance (McCall, 2003).

5 Several recent studies have provided the beginnings of the literature on this new approach. 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 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 earthquakes are detected. Notifications are delivered considerably faster than the announcements that are broadcasted by the Japan Meteorological Agency (JMA). On average, a JMA announcement is broadcasted 6 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 overall description of VGI's role in disaster management. Apart from being 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 for using VGI is its unknown quality. The general population is untrained to make the specific observations necessary in environmental management, and may intentionally or unintentionally supply erroneous information. Data quality is often unknown, and data sampling is frequently dispersed and unstructured. Similar concerns are often expressed regarding many other types of information provided by amateurs, thereby reflecting the profound association that society makes between qualifications, institutions, and trust (Goodchild and Glennon, 2010). Nevertheless, several grounds are used as basis for believing that the VGI quality can 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 detected no significant difference between the field samples collected by volunteers and those collected by professionals. Citizen volunteers who are properly trained can collect reliable data and make stream assessments that are comparable with those made by professionals. Data collected by volunteers can supplement information used by government agencies to manage and protect rivers and streams (Fore et al., 2001).

Community-based water quality monitoring has been conducted in several countries, e.g. Secchi Dip-In Program in U.S. (Lee et al., 1997), Waterwatch Program in Australia (Kingham, 2002) and the Open Air Laboratories (OPAL) Water Survey in UK (Rose et al., 2015). 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 analysed 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 a struggle for environmental justice in low-income, minority residents of the Bayview
5 Hunters Point (BVHP) community in San Francisco, California. These programs provide the protocols, guidelines, equipment and trainings to volunteers for water quality monitoring. The volunteers sample the water and measure the quality indicators through test strips and apparatus. The economic cost and inconvenience of community-based monitoring are still high and it limits the application of the program.

10 Social media, such as Twitter, Facebook, Sina Weibo, and Tencent Weixin (the popular Chinese version of Twitter), is capable of guiding and providing incentives to volunteers through real-time online communication among volunteers. In recent years, social media has become a major communication channel in our society (Jiang et al., 2015). Social media refers to Internet-based applications that enable people to conduct online communications intended for interaction, community input, and collaboration (Lindsay, 2011). Social media enables information sharing from multiple parties on computers and
15 mobile devices, particularly through social networking sites (e.g., Facebook, YouTube, and Twitter), texting, chat rooms, 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). The major functions of social media in the environmental management processes include one-way information sharing, two-way information sharing, situational awareness, rumor control, reconnection, and decision-making (Tang et al., 2015). Jiang et al. (2015)
20 effectively monitored the dynamic changes of air quality in large cities by analysing the spatiotemporal trends in geo-targeted social media messages with comprehensive big data filtering procedures. Werts et al. (2012) established a website called AbandonedDevelopments.com to collect VGI and detect the sediment pollution of abandoned structures in upstate South Carolina. It combined Web-GIS technologies, data sources, and social media for future use in soil and water conservation.

25 The advertising, instruction, and guidance for water quality monitoring can be spread extensively and delivered to the potential volunteers' mobile devices directly through social media platforms. The observed sewage dumping or water pollution activities can be disseminated rapidly in social media networks and call the government's attention. Social media provides the platform for volunteers to present, discuss, and communicate their criticism, anger, and solutions to the water
30 pollution they observe. Communication and mutual encouragement provide a strong motivation for volunteers to monitor water quality and persistently share their observations. Furthermore, the discussion on the pollution activities reported in social media networks increases the opinion's pressure on the government to solve the problem. Government feedback can also be promptly disseminated to volunteers through social media. Timely dissemination of government feedbacks also
| motivates volunteers to continuously monitor water quality (see Fig. 1). Today, volunteers are equipped with digital cameras,

GPS, digital maps, and other resources through their smartphone. Multiplying the resources of the average empowered citizen by the population of the city results in an astounding ability to create and share information (Goodchild and Glennon, 2010).



Figure 1: Positive feedback ring for volunteer's motivation.

This study aims to establish an approach to monitor surface water quality through volunteered citizen scientists. An App platform basing on social media is built to collect the water quality information. The application of the App suggests new methods to monitor water pollution using VGI and social media in the future. Results reveal the feasibility of VGI in monitoring water quality. The impacts of photographed function, anonymous submission and economic incentives on increasing data credibility and volunteer motivation are analysed. This paper is organized as follow. Methodology is presented in Section 2. Monitoring reports obtained across China are displayed in Section 3. The data quality and motivation of volunteers are discussed in Section 4. Conclusions drawn from this study are given in Section 5.

2 Methodology

A methodological framework is established to: (1) Collect sensory data of surface water quality from the volunteer citizens, who describe and photograph the water that they pass by or are close to. Descriptions and photos are sent to a data centre from their mobile devices of volunteers through a specific social media application in the social media platform. (2) Detect/Disclose the illegal/hidden sewage dumping that is very difficult to detect through the official monitoring network. Once sewage dumping reports from the volunteers are submitted, the data centre will transfer them to the mass media and water authorities after undertaking credibility checks.

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

2.1 Data type

Four indicators ~~were-are~~ adopted to describe the physical characteristics of water quality (see Figure 21). A total of 11 water color options ~~were-are~~ provided for volunteers to choose from, including red, orange, yellow, green, cyan, blue, purple, milky, pink, black, and crystal. The second indicator is smell, which is quantified by the ranking scores made by volunteers.

5 The volunteer is asked to rank the smell of the sample from 0 to 10, with 0 implying a lack of odour and 10 ~~implying representing~~ a foul-smelling sample. The third indicator is turbidity in which a score of 10 means that the water is non-transparent and a score of 0 represents transparency. ~~The higher score claims that there are more contaminants in the water.~~
The fourth indicator represents floating objects or floating material on the water. If the water is completely covered by oil, plastics, and rubbish, among others, the floats score is 10. By contrast, a score of 0 means no floating object is present.
10 The last item is an integrated assessment on the water quality, ranking from worst, very bad, bad, and good, to excellent. Volunteers evaluate the water quality based on their perception.

Formatted: Font: (Default) Times New Roman

2.2 App in social media platform

The Tsinghua Environment Monitoring Platform (TEMP, <http://www.thuhjc.com/>) application ~~was-is~~ built based on WeChat public accounts. WeChat is a mobile text and voice messaging communication service developed by Tencent in China, and was first released in January 2011. WeChat is one of the largest messaging applications in China. As of May 15 2016, WeChat has 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 (Tencent, 2016). Moreover, ~~WeChat enables users to register a public account, thereby enabling them to push feeds to subscribers, interact with subscribers, and provide them with services.~~ Public accounts can be used as a service platform, such as hospital pre-registrations, visa renewal, or credit card service (Wikipedia, 2016). WeChat provides functions for 20 users to post images and texts, share music and articles, as well as comment and “like” in the Moments. Only the users' friends will be able to view the contents and comments of their Moments. In addition, WeChat supports payment and money transfer, thereby enabling users to use peer-to-peer transfer and electronic bill payment (Tencent, 2016).

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

25 Volunteer reporters ~~install WeChat on their mobile devices and~~ log into TEMP through their WeChat accounts. Through TEMP, volunteers are able to submit reports together with the GPS position where the reported water is located. The location can either be automatically extracted from the devices or be input manually by the reporters. Volunteers can tweet the reports to their friends, and post and comment on them in the Moments. TEMP also provides a function for ranking volunteers according to their contributions. Awards, such as cash through WeChat Payment, ~~will-can~~ be offered to top-ranking reporters. A computer-based website is also provided for the public to view and download the reports (See TEMP, 30 <http://www.thuhjc.com/>)

2.3 Volunteer recruitment

Two modes ~~were-are~~ used for volunteer recruitment. In Mode 1, TEMP ~~iswas~~ popularized from a central group to the general public (see [Figure- 12](#)). The university students recruited for the present study post the address of the TEMP two-dimension code in Moments and Chat Groups after they log in the TEMP through their WeChat accounts. Their friends who are interested in SWQM will click the address or scan the two-dimension code (QR code), and then be directed to TEMP. If these people log in and submit reports, they will be involved and contacted by the TEMP. ~~They can receive the message from TEMP and be encouraged to submit more reports and attract numerous volunteers to be involved.~~ TEMP is not able to control when and where the monitoring reports come from. Data ~~was-is~~ scattered in Mode 1.

In Mode 2, a group of professional citizens ~~were-are~~ recruited to monitor the water quality in targeted sites. Professionals who work for environmental authorities and organizations ~~were-are~~ interviewed and convinced to register in TEMP. They ~~were-are~~ required to monitor water bodies that they are familiar with and regularly submit the reports through TEMP. In Mode 2, The TEMP can recruit specific volunteers and collect a number of data more quickly than in Mode 1.

2.4 Data analysis

A method ~~was-is~~ established to quantitatively analyse monitoring reports. Smell, turbidity, floats, and integrated assessment reported by the volunteers ~~were-are~~ quantified and normalized between 0.0 and 1.0 according to their ranking scores. Water colour ~~was-is~~ used for data cleansing and rumour control through cross validation between the 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	<i>C</i>	Text	None
Smell	<i>S</i>	Score from 0 to 10	1.0–0.0
Turbidity	<i>T</i>	Score from 0 to 10	1.0–0.0
Floats	<i>F</i>	Score from 0 to 10	1.0–0.0
Integrated assessment	<i>I_a</i>	Grand from 1 to 56	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 integrated assessment's indicator 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).

An overall water quality assessment, *Q*, was calculated as,

$$Q = (P_a + I_a)/2 = [(S + T + F)/3 + I_a]/2, \quad (1)$$

A preliminary assessment (*P_a*) is calculated through the average of *S*, *T*, and *F*. The average value of *S*, *T* and *F* (*P_a*) provides an indicator-based assessment of the water quality. *I_a* is an integrated assessment in the reports based on reporters' overall perception of the water. *I* is higher than the individual sense-based indicators *S*, *T*, and *F*. The water sample's smell, turbidity, and floats were used in the calculation to enhance the assessment's credibility.

2.5 Validation

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

It should be pointed out that it is quite difficult to validate all the volunteers' reports. Volunteers distribute all over the country and they submit the reports randomly once they see the dirty water nearby. There is rarely an official gauge site just locating on a volunteer's report point. The TEMP employed a group of trained volunteers to report the water quality at Huayuankou station each 2 or 3 days during March and April 2016. This provides the site-specific VGI data for validation.

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

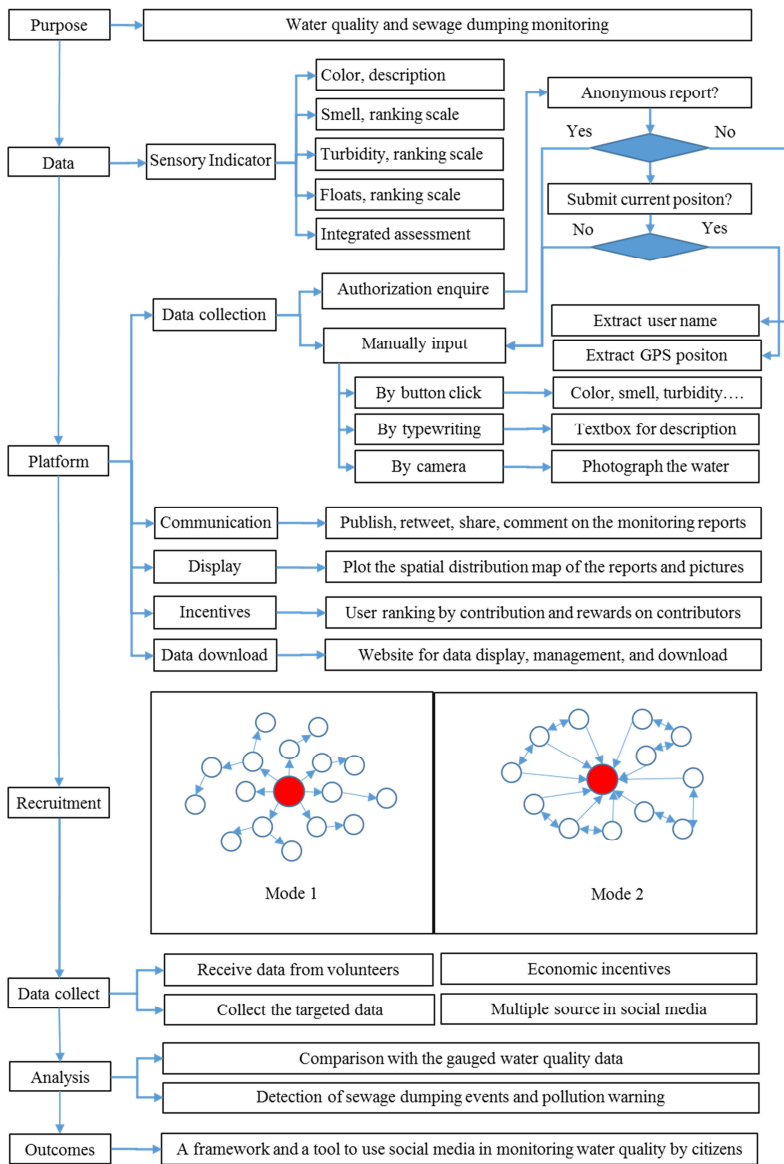


Figure 2: Framework of monitoring water quality by VGI and social media.

3 Results

3.1 Water quality reports across China

TEMP received a total of ~~265-324~~ reports from volunteers in China between 12 October 2015 and ~~30-15 May-December~~ 2016. Out of ~~the 265 those~~ reports, ~~472-219 were validated~~ ~~are adopted after data cleansing in the present study.~~

Table 2 shows the overview of ~~219-172 validated~~ reports across ~~29-30~~ provinces and municipalities in China. ~~Henan-Beijing and Henan Province and Beijing municipality~~ have about 30 reports. ~~The reports from Beijing were~~ ~~are mainly from the students of Tsinghua University, where the research group of the present study is located.~~ The reports from Henan Province ~~were~~ ~~are mostly~~ contributed by Henan-based professional volunteers working in the ~~Yellow River Conservation Commission~~ YRCC. ~~There are~~ ~~The reports from Beijing were mainly from the students of Tsinghua University, where the research group of the present study is located.~~ A total of ~~23-6~~ provinces ~~where the number of reports is more than~~ ~~have under~~ 10 reports ~~10~~. A total of 14 provinces have only 1 report during the period. Meanwhile, ~~n~~ No reports ~~are~~ were provided from ~~Tibet~~, Xinjiang, Hainan, Taiwan, Macau, and the South China Sea.

Table 2: Number of reports across China

No. Provinces	Number of reports			Integrated	No. Provinces	Number of reports			Integrated
	Total	Anonymous	Photographed	assessment(<i>I_a</i>) Assessment		Total	Anonymous	Photographed	assessment(<i>I_a</i>) Assessment
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	

Formatted Table
Formatted: Font: Italic
Formatted: Font: Italic

Table 2 shows that 8292 out of 172 reports were submitted without the WeChat users' names indicated. TEMP users were allowed to submit their reports anonymously. This represents that there are indeed a number of volunteers preferring anonymous reports due to the privacy. were concerned about their privacy; thus, they submitted reports anonymously. The number of anonymous reports implies the necessity of setting up an anonymous function in TEMP because volunteers care about their privacy when they disclose the water pollution activities around them. A total of 83-107 reports out of 172 had photos of water. Photos significantly increased the credibility of the reports by providing substantial information for water quality analysis. However, 50% of the reports had no photos. Volunteers were also concerned about charges for mobile Internet data usage, through which they upload the photos without Wi-Fi. Additional incentive measures are required to encourage volunteers to upload the photos of water. Figure 3-2 and Figure 4-3 show the number of reports and anonymous reports and photograph reports, respectively.

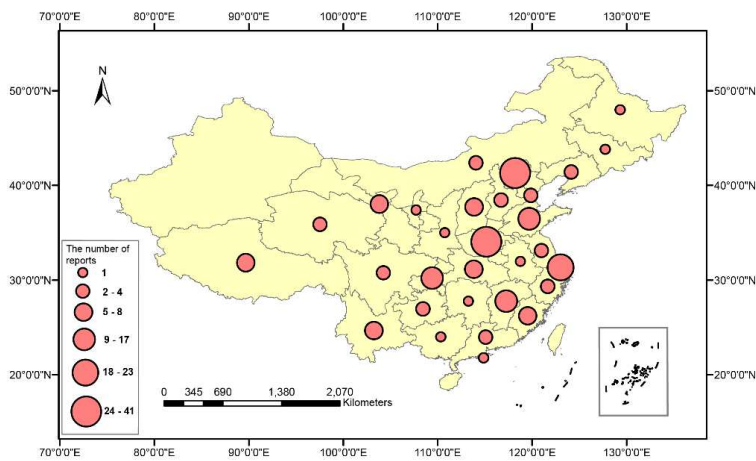


Figure 32. Distribution of the reports across the provinces and cities in China.

Note: No reports were provided from Tibet, Xinjiang, Hainan, Macau, Taiwan, and South China Sea.

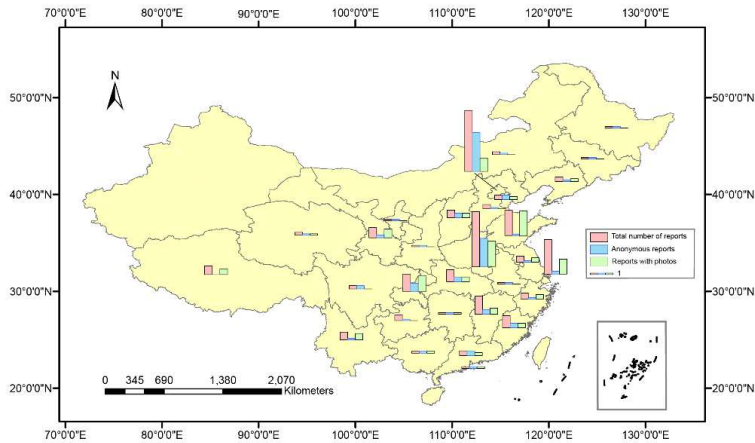


Figure 43. Distribution of anonymous reports and photograph reports.

Note: No reports were provided from Tibet, Xinjiang, Hainan, Macau, Taiwan, and South China Sea.

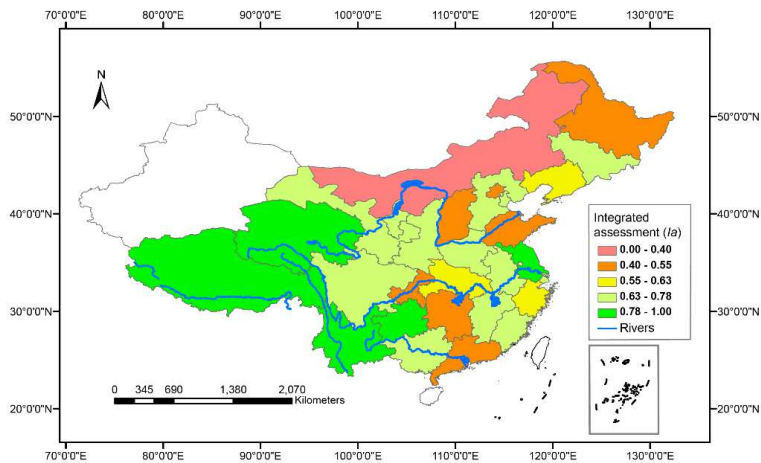


Figure 44. The integrated assessment (I_a) results of China

Note: No reports were provided from Tibet, Xinjiang, Hainan, Macau, Taiwan, and South China Sea.

Table 2 shows the water quality assessment results. The average value of I_a in each province is represented in Table 2 and Figure 3. The reported water quality in the provinces locating at the source of Yellow River, Yangzi River and Pearl River, such as Tibet, Qinghai and Yunnan is better than those at downstream, e.g. Shandong and Guangzhou. This is consistent with the water quality situation in China and demonstrates the Reasonableness of the VGI reports. However, the assessment

Formatted: Font: Italic

Formatted: Font: Italic, Subscript


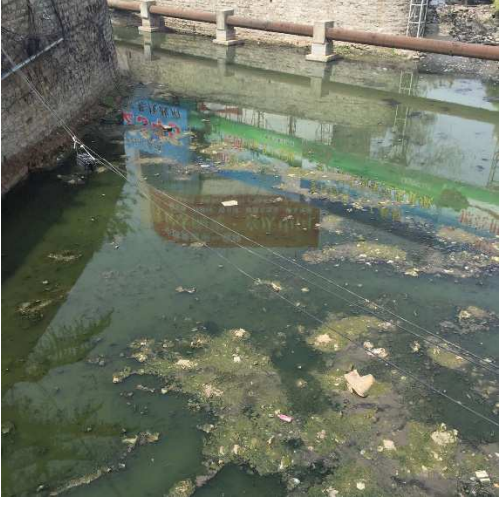

is unable to accurately represent the overall situation of the surface water quality in a region because the coverage and frequency of the reports are insufficient. It is possible to depict ~~The assessment results for each province were calculated based on the volunteered reports from that province (see Equation (1)). The assessment is, however, unable to represent the overall situation of the surface water quality in a region because the coverage and frequency of the reports is insufficient at the current stage.~~ Meanwhile, TEMP provides a practical tool for citizens to monitor the rivers and the lakes around them. An increasing number of volunteers will be involved in the future. The surface water quality of a whole region or a river basin can be depicted once enough volunteers and sufficient reports and photos are provided.

3.2 Examples of reports for pollution disclosure

Table 3 shows three photograph reports. Algal blooms and water surface foam are shown in the photos. In Report 1, the volunteer submitted a photo of the river in Tsinghua University, Beijing, China. The river ~~is~~ was polluted by the domestic sewage and ~~suffers~~ suffered from eutrophication. Report 2 represents the water quality in an unidentified river located in Fei Town, Linyi City, Shandong Province. Photo 2 shows that the water surface ~~is~~ was covered by algae and rubbish. The reporter assessed the water quality as very bad. Report 3 is from a coastal area, namely, Tianjing City. The reporter identified the color of the water as black and assessed its quality as bad. ~~White foam was observed on the water surface, thereby implying that the quality of water is not good.~~

Among the ~~83~~ 107 photos collected by TEMP during the period, the volunteers failed to monitor the scene, wherein the dumped sewage water ~~was~~ is just flowing to the river or lake. Sewage water dumping in China generally occurs at night and in a hidden place, which is rarely found by volunteers. If there are sufficient volunteers aiming to disclose the hidden sewage dumping in a region, ~~then~~ the pollution activities can be determined by TEMP ~~based on its current functions.~~

Table 3: Examples of reports for pollution disclosure

No.	1					2					3				
Report	Color	Smell	Turbidity	Floats	Assess-_L	Color	Smell	Turbidity	Floats	Assess-_L	Color	Smell	Turbidity	Floats	Assess-_L
	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	April 10, 2016					April 15, 2016					May 1, 2016				
Location	Tsinghua University, Beijing City, China					Fei Town, Linyi City, Shandong Province					New coastal area, Tianjin City, China				

Note: B means bad; VB means very bad.

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

Formatted: Font: Italic

Formatted: Centered

3.3 Correlation between I_a and S, T, F

The smell (S), turbidity (T), floats (F) and integrated assessment (I_a) from 107 photograph reports are analysed and shown in Figure 5, 6 and 7. The reports are divided to five groups according to I_a , including 0.00, 0.25, 0.50, 0.75 and 1.00. The value of I_a is equal among the reports in the same group. For each group, the minimum, maximum and average value of turbidity (T) is plotted in Figure 5. So does the floats (F) in Figure 6 and the smell (S) in Figure 7.

I_a is highly correlated with T and S . The correlation demonstrates that the VGI reports are reasonable. The volunteers indeed complete the reports basing on actual observation on the water. The turbidity (T) is higher correlated with I_a than floats (F). Turbidity is easier to be observed and is more influencing on volunteers' judgement. People trend to regard the water as in good quality if it is crystal, despite there may be objects, e.g. leaves or branches floating on it.

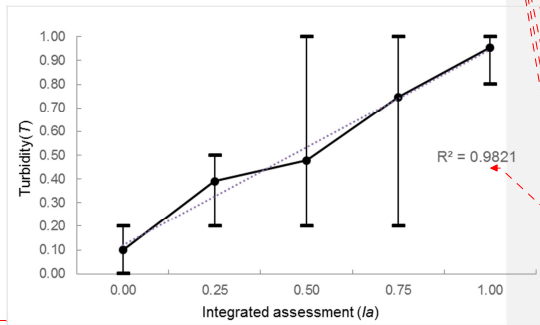
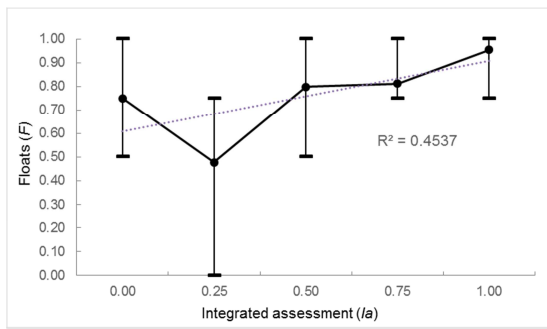


Table 3



- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic, Subscript
- Formatted: Normal
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Normal, Right -2.98 ch

Figure 5. The correlation between L_a and T

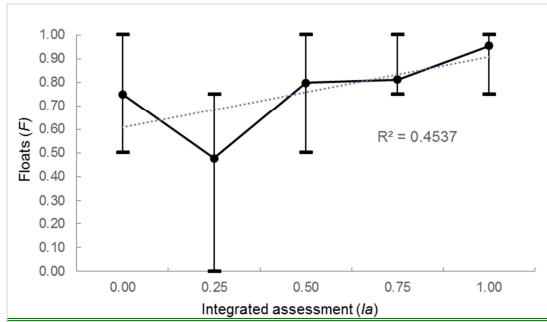


Figure 6. The correlation between L_a and F

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

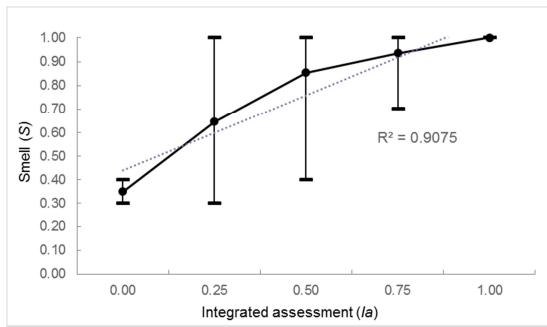


Figure 7. The correlation between L_a and S

3.4 Validation with the gauged data

The turbidity (T) from the reports at Huayankou station of the Yellow River is compared with the gauged data (see Figure 8). The temporal variation of T shows to be similar between the reported and the gauged data. It demonstrates the effectiveness of the VGI data to some extent. All these reports have photos and can be regarded as high credibility. The results imply that the water quality assessment through the citizens' sensory is able to be compatible with the reality.

of water

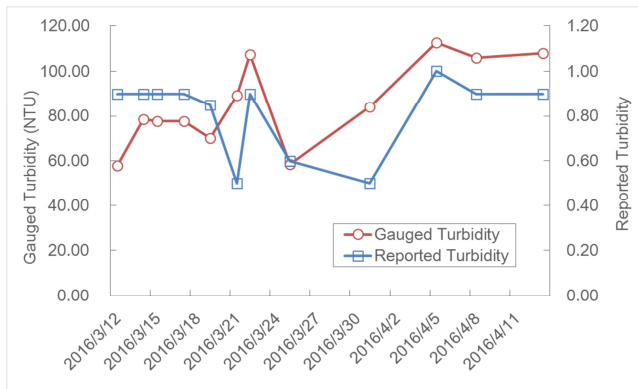


Figure 8. The turbidity from the reports and the gauged data

5 Discussion

The present study aimed to develop a method to monitor surface water quality through volunteer citizens using a social media application. A framework was established to guide the application design, volunteer recruitment, data collection, and report analysis. The TEMP application was built based on the social media platform called WeChat. Using the application, TEMP users can describe and photograph the water in rivers and lakes following the TEMP instructions.

Moreover, users can report the surface water pollution activities that affect their living and health.

A total of 219 validated reports were analyzed in this study. These reports are from 140 volunteers across 30 provinces and cities in China. The water quality on sites was assessed by the volunteers using their sense organs, particularly through their observations on water smell, turbidity, and floating matter on the water. People's perception on water quality varies, and different people may provide different assessment reports on water from the same site. Comparing the assessment results across different sites is difficult because the citizens' reports are subjective to some degree. This situation will change if numerous volunteers and extensive reports are obtained. Data could also potentially be supplemented

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Normal

Formatted: Font color: Auto

~~by satellite and aerial remote sensing and 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 data density is present. The present study provides 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).~~

The credibility of the reports is the major concern of the present study. Regardless of the water quality assessment by citizens, identifying whether the reports are real and whether the volunteers generated the reports based on their observations is necessary ~~rather than on false statements or rumours. Rumour control is significant when a water pollution activity is detected and reported in social media (Tang et al., 2015).~~ The current study applied three criteria to ~~validate~~ wash the ~~reports~~data, including: (1) a report with the exact GPS location information of the site is regarded as credible. The sites' GPS information is automatically abstracted from the reporters' mobile devices when they submit the reports. (2) If reports are submitted several times during a short period, and most of these reports are from the same volunteer on the same site, then the reports are deemed ~~to have~~as low credibility. These reports may be test reports from new volunteers. (3) Reports with photos are the most credible. A total of ~~265-324~~ reports are cleansed following these criteria, and ~~172-219~~ were validated in the present study. ~~107 photograph reports were regarded as high credibility. Further cross-validation among different reports can be applied if large amounts of data in a region are present.~~

~~The validation of the reports is a tricky issue. The reports are submitted from scattered sites, where there are insufficient gauged data for validation. The TEMP recruited a group of professional volunteers from the Yellow River Conservation Commission. They report the water quality continuously at the Huayuankou station where the gauged data is available. They are familiar with the water quality status on site but did not know the gauged data when submitting the reports. The gauged results are assessable at least one day after sampling at the station, because the water sample needs to be analysed in the laboratory. The volunteers can only assess to the gauged data after submitting the reports. Meanwhile, The TEMP records the time of the reports according to the clock on TEMP server. There is no way for the volunteer to modify the time of the report. The TEMP only received 13 reliable photograph reports at the station during March and April 2016. Although the data is limited, the validation represents that the VGI data is valuable for water quality monitoring to some extent. The citizens' assessment is effective in representing the water quality status if the reporter/citizen is relatively trained. Similar evidence was also obtained by Fore et al. (2001), Monk et al. (2008), Flanagan and Metzger (2008), and Koss et al. (2005).~~

~~There is an opportunity to develop image analysis to check if the photo matches the ratings of turbidity. E. Minkman, P.J. van Overloop, M.C.A. van der Sanden, 2015 explored the potential of citizen science in mobile crowd sensing in water quality monitoring, using mobile crowd sensing application for water quality measurements in the Netherlands. It consists of a colorimetric analysis using smartphone camera. The indicator test strips are photographed and the colour of the strips are analysed automatically by an App in the smartphone to obtain chemical water quality data. Snik et al., 2014 developed the~~

Formatted: Font: (Default) Times New Roman, (Asian) Chinese (PRC)

Formatted: Font: (Default) Times New Roman, (Asian) Chinese (PRC)

Formatted: Font: (Default) Times New Roman, (Asian) Chinese (PRC)

Formatted: Font: (Default) Times New Roman, (Asian) Chinese (PRC)

Formatted: Font: (Default) Times New Roman, (Asian) Chinese (PRC)

Formatted: Font: (Asian) Chinese (PRC)

Formatted: Font: (Default) Times New Roman, (Asian) Chinese (PRC)

~~iSPEX, a low-cost, mass-producible optical add-on for smartphones with a corresponding app. People could purchase the iSPEX smartphone accessory and place it on top of their iPhone. They are able to measure particulate matter in the atmosphere using a special smartphone application (Minkman, 2015). Such devices are meaningful in citizen-based water quality monitoring to increase data credibility. The devices should be smart, portable and convenient to use with smartphone, e.g. an external device using laser to detect water quality (Chen et al., 2015).~~

The volunteers' motivation to proactively submit data is another major concern of the present study. Various VGI studies reveal that the volunteers' motivation is a key factor to the VGI program's success (Werts et al., 2012). Coleman et al. (2009), as cited in (Werts et al., 2012, Page 817) revealed that the key motivators include 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, among others. Several 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, Page 817), regarded fun, learning, and instrumentality as primary motivators for geographic information contributors, thereby noting that "when contributors see their data appear visually on maps, they receive deep satisfaction."

In the present study, core members in the chat group who lead the communication and remind the volunteers are beneficial in motivating the contributors. Moreover, certain economic incentives are quite effective in increasing the number of contributors and their contributions. Figure 5-9 shows the number of volunteers involved in the present study. The TEMP application ~~was built and~~ has been tested since April 2015. After the application development and testing, TEMP was promoted by the faculty, staff, and students at Tsinghua University through their WeChat Moments. The number of users steadily increased until the first official launch of TEMP in the ceremony conference for the establishment of the Tsinghua Remote Sensing and Big Data Research Centre (<http://hydrosky.org/Newscon/index/id/581/aid/544445113>) in October 2015, when a sharp increase occurred. Professionals and the public related to the research centre were observed to have registered in TEMP, thereby facilitating the increase in the number of users in October 2015. After the increase in the number of users, TEMP users continuously promoted and advertised the platform in their WeChat Moments, and attracted the involvement of additional users. The number of reports increased gradually because of the lack of incentives to the volunteers before March 2016.

In March 2016, Mode 2 was adapted to recruit volunteers working in the ~~Yellow River Conservation Commission YRCC~~ and students in the universities in Beijing. Several volunteer chat groups were established in WeChat, and a core member was assigned in each group to lead the communication and remind users to submit reports. Consequently, the number of reports substantially increased after these activities. Since April 2016, economic incentives and rewards were adapted. The core members in the chat groups gave "Red Packet" money through WeChat Payment. The core members transferred the money in the chat groups, which was normally 100 RMB each time in a chat group. The "Red Packet" money can be sent from within a chat as sending photos. The "Red Packet" money can be obtained and shared by dozens of members, who point the

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

“Red Packet” money picture on the screen before others do. The winner receives a random share of the total “Red Packet” money. The members who win the money in the chat groups were considerably motivated to submit reports, tended to invite their friends in, and encouraged them to submit data. After April-May 2016, these economic incentives continuously motivated the volunteers and increased the number of reports.

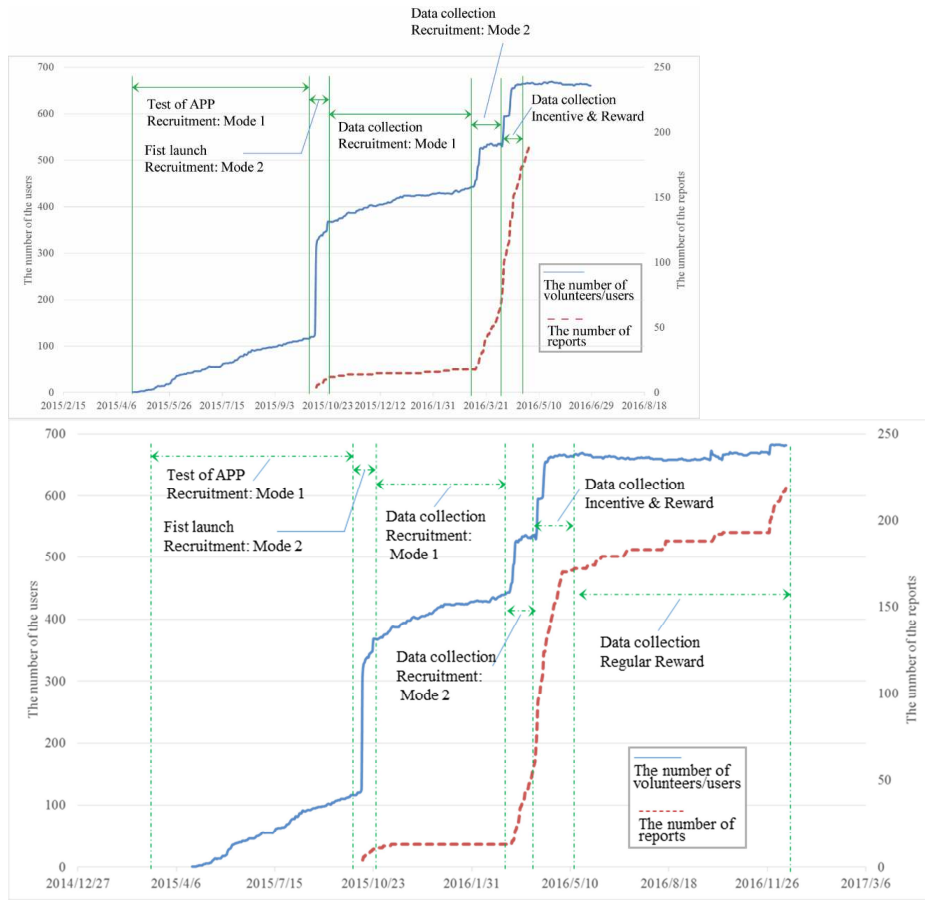


Figure-5-9: Increase in the number of reports and reporters.

The economic rewards were also provided in accordance with the volunteers’ contributions, rather than their finger speed of pointing the “Red Packet” money picture on their mobile devices’ screen. Figure-6-10 shows the distribution of the reports versus reporters. A total of 20 out of 39 reporters (excluding the anonymous reporters) About 50% of reporters submitted

Formatted: Font color: Red

only 1 report each during the period. There were 35% of reporters whose reports' number was larger than 10. A targeted reward by WeChat Payment was provided to the reporters with top-ranking submissions among all TEMP users since April 2016.

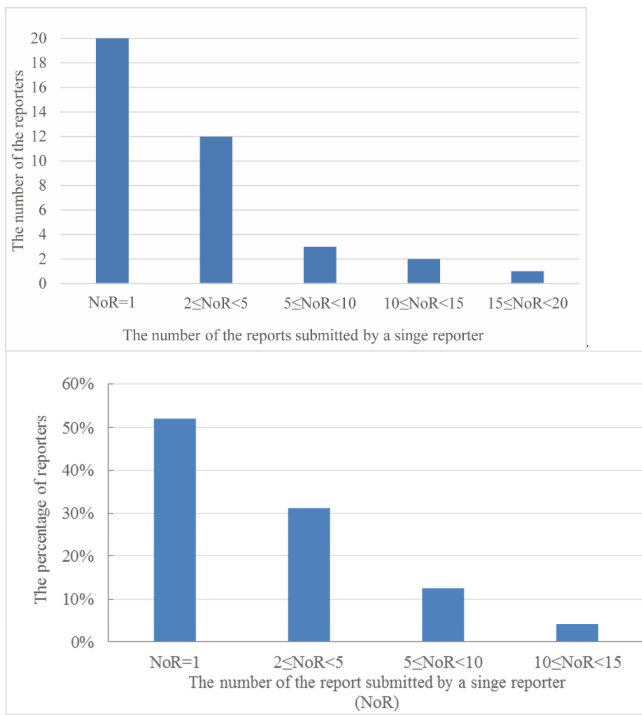


Figure 610: Distribution of the reports versus reporters.

Further research related to the present study may include efforts to collect additional data and recruit many active volunteers and develop methods for data validation and analysis. Additional data sources, such as Twitter, Facebook, and Sina Weibo, could be involved (Jiang et al., 2015). People tend to post or tweet a text or a picture in social media to complain water pollution, and represent their lives to their friends, if they see a dirty river or lake (Kaplan and Haenlein, 2010). In this case, social media users do not intend to report a water pollution activity. However, users inadvertently provide the data for water quality monitoring. The text and photos about water quality from Twitter, Facebook, and Sina Weibo can provide high density and massive data because of the presence of millions of active users in these social media platforms. This process saves the effort to recruit and motivate volunteers, although the data collected tend to be unstructured (Poser and Dransch, 2010).

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

~~This situation will change~~The data validation could be more robust if numerous volunteers and extensive reports are obtained. Data could also potentially be supplemented by satellite and aerial remote sensing and 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 data density is present. The present study provides 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 image analysis to check if the photo matches the ratings of turbidity. E. Minkman, P.J. van Overloop; M.C.A. van der Sanden, 2015 explored the potential of citizen science in mobile crowd sensing in water quality monitoring, using mobile crowd sensing application for water quality measurements in the Netherlands. It consists of a colorimetric analysis using smartphone camera. The indicator test strips are photographed and the colour of the strips are analysed automatically by an App in the smartphone 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 app. People could purchase the iSPEX smartphone accessory and place it on top of their iPhone. They are able to measure particulate matter in the atmosphere using a special smartphone application (Minkman, 2015). Such devices are meaningful in citizen-based water quality monitoring to increase data credibility. The devices should be smart, portable and convenient to use with smartphone, e.g. an external device using laser to detect water quality (Chen et al., 2015).

Formatted: Font color: Red

TEMP facilitates the efficient method of contacting and guiding volunteers to submit the required structural report and data. However, a considerable effort is necessary to recruit and motivate them. Economic incentives are demonstrated to be effective in the present study; however, the use of incentives appears to be an unsustainable method. Cooperation with non-government organizations (NGOs) that are focused on environmental protection may provide another approach for volunteer recruitment. NGO members should be interested in TEMP and have an enduring and strong motivation to disclose water pollution activities.

6 Conclusion

A methodological framework to monitor surface water quality using social media is established in the present study, including the selection of the water quality indicators, application design guide, volunteer recruitment methods, data collection, cleansing, and analysis. The TEMP application ~~was~~is established based on a popular social media platform in China called WeChat. TEMP enables registered users to submit their descriptions and photos on rivers or lakes anonymously or non-anonymously with the automatically extracted GPS information of the sites from their mobile devices.

A total of ~~265-324~~ reports across ~~29-30~~ provinces and cities in China were received by TEMP between October 12, 2015 and ~~May-December 30~~ 15, 2016. Out of the ~~265-324~~ reports, ~~172-219~~ were-are displayed~~validated and analyzed after data cleansing~~. The distribution analysis of reports across China indicates that the anonymous and photograph functions are quite essential for TEMP. Over 42.8% of the ~~172-219~~ reports are from anonymous users. Accordingly, people care about their privacy when they try to disclose a water pollution activity occurring within their vicinity. A total of ~~83-107~~ photos of rivers and lakes ~~were-are~~ collected through TEMP, and these photos provide extensive information for pollution detection. 13 photograph reports at the Huayunkou station in the Yellow River is validated through comparing the reported turbidity with the gauged value. It shows the citizen-based water quality data is relatively credible, if the volunteers are trained.

Data quality and the users' motivation are also discussed. Three criteria ~~were-are~~ applied for data cleansing based on the location, time, and photos in the reports. Two modes for volunteer recruitment ~~were-are~~ used. Mode 2 is better for increasing the number of volunteers within a short period. An economic incentive mechanism to targeted volunteers ~~was-is~~ also implemented in the present study. Economic incentives are effective in motivating the volunteers to contribute data under guidance from the core members of the chat groups.

Further research should include collecting more data ~~and recruiting more motivated volunteers and validating the reports~~. The unpremeditated data referring to water quality from Twitter, Facebook, and Sina Weibo should be abstracted to increase the volume of data. The big data method is expected to be used with the availability of sufficient data.

Data availability

The data underlying this research can be accessed publicly. All the data can be download from the website www.thuhjic.com. This is a website established by the authors for displaying and downloading the data from TEMP platform.

Author Contribution

Dr. Hang Zheng designed the framework of this study, analysed the data and prepared the manuscript with contributions from all co-authors. Professor Hong Yang designed the interface of TEMP platform and contributed to the Discussion Section. Dr. Di Long carried out the data collection. Mr. Jing Hua developed main functions of the TEMP platform through computer programming.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgement

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

References

- 10 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., NedovicćBudić, 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.
- 15 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 Producers, *Int. J. Spat. Data Infrastructures Res.*, 4, 332–358, doi:10.2902/1725-0463.2009.04.art16, 2009.
- 20 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.
- 25 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.
- Fore, L. S., Paulsen, K. and O’Laughlin, K.: Assessing the performance of volunteers in monitoring streams, *Freshw. Biol.*, 30 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.
- 5 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.
- 10 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.*, 15 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, 20 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.
- 25 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/homesecc/R41987.pdf>, 2011.
- 30 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.
- 5 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.
- 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.
- 10 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.
- 15 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.
- 20 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.
- 25 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.*, 30 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.
- 5 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.
- 10 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/ijerph11111833, 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.
- 15 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.
- 20 Wikipedia: WeChat, [online] Available from: https://en.wikipedia.org/wiki/WeChat#cite_note-35 (Accessed 9 July 2016), 2016.

Figure 1: Positive feedback ring for volunteer's motivation.

5 **Figure 2: Framework of monitoring water quality by VGI and social media.**

Figure 3: Distribution of the reports across the provinces and cities in China.

Figure 4: Distribution of anonymous reports and photograph reports.

Figure 5: Increase in the number of reports and reporters.


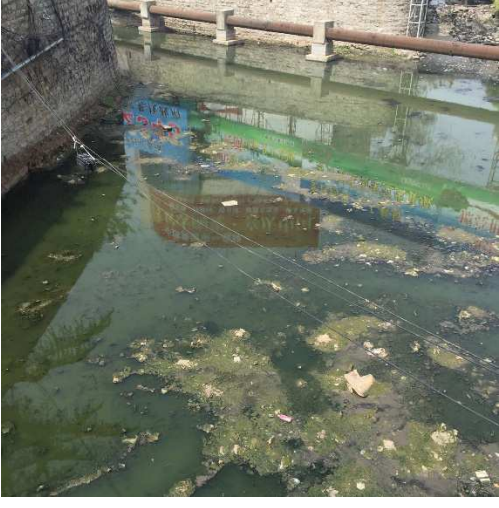

Figure 6: Distribution of the reports versus reporters.

10

Table 2: Number of reports across China

No.	Provinces	Number of reports			Assessment	No.	Provinces	Number of reports			Assessment
		Total	Anonymous	Photographed				Total	Anonymous	Photographed	
1	Henan	34	19	18	0.74	18	Guangdong	1	1	1	0.75
2	Beijing	29	23	4	0.59	19	Guangxi	1	1	1	0.77
3	Jiangxi	24	4	4	0.69	20	Guizhou	1	1	0	0.77
4	Shandong	16	1	16	0.61	21	Hebei	1	1	0	0.73
5	Shanghai	15	1	7	0.74	22	Hunan	1	1	1	0.58
6	Chongqing	12	6	11	0.57	23	Jilin	1	1	0	0.78
7	Fujian	6	3	2	0.88	24	Neimeng	1	1	0	0.33
8	Gansu	4	2	3	0.75	25	Ningxia	1	1	0	0.80
9	Yunnan	4	1	3	0.90	26	Shaanxi	1	1	0	0.75
10	Shanxi	3	2	2	0.74	27	Sichuan	1	1	0	0.78
11	Tianjin	3	3	2	0.80	28	Zhejiang	1	1	1	0.58
12	Hubei	2	0	2	0.64	29	Hongkong	1	1	1	1.00
13	Jiangsu	2	1	2	0.77	30	Xinjiang	/	/	/	/
14	Liaoning	2	1	1	0.70	31	Tibet	/	/	/	/
15	Qinghai	2	1	1	0.88	32	Hainan	/	/	/	/
16	Anhui	1	1	0	0.80	33	Taiwan	/	/	/	/
17	Heilongjiang	1	1	0	0.55	34	Macau	/	/	/	/
Total							172	82		83	

Table 3: Examples of reports for pollution disclosure

No.	1					2					3				
Report	Color	Smell	Turbidity	Floats	Assess.	Color	Smell	Turbidity	Floats	Assess.	Color	Smell	Turbidity	Floats	Assess.
	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	April 10, 2016					April 15, 2016					May 1, 2016				
Location	Tsinghua University, Beijing City, China					Fei Town, Linyi City, Shandong Province					New coastal area, Tianjin City, China				

Note: B means bad; VB means very bad.