Citizen science pioneers in Kenya – a crowdsourced approach for hydrological monitoring

Authors

B. Weeser[1,2,3*], J. Stenfert Kroese[3,6], S. R. Jacobs[1,3,4], N. Njue[1,3], Z. Kemboi[5], A. Ran[5], M. C. Rufino[6], L. Breuer[1,2]

Affiliations

[1] Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (IFZ), Justus Liebig University Giessen, Heinrich-Buff-Ring 26, 35392 Giessen, Germany.
[2] Centre for International Development and Environmental Research, Justus Liebig University Giessen, Senckenbergstr. 3, 35390 Giessen, Germany.
[3] Centre for International Forestry Research (CIFOR), c/o World Agroforestry Centre, United Nations Avenue, Gigiri, P.O. Box 30677, 00100 Nairobi, Kenya.
[4] Karlsruhe Institute of Technology – Institute of Meteorology and Climate Research, Atmospheric Environmental Research (KIT/IMK-IFU), Kreuzeckbahnstr. 19, 82467 Garmisch-Partenkirchen, Germany.
[5] Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Ngong Road P.O. Box 19512, 00202 Nairobi, Kenya.
[6] Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, United Kingdom.
*b.weeser@uni-giessen.de

Abstract

Although water is involved in many ecosystem services, the absence of monitoring data restricts the development of effective water management strategies especially in remote regions. Traditional monitoring networks can be expensive, with unaffordable costs in many low-income countries. Involving citizens in monitoring through crowdsourcing has the potential to reduce these costs but remains uncommon in hydrology. This study evaluates the quality and quantity of data generated by citizens in a remote Kenyan basin and assesses whether crowdsourcing is a suitable method to overcome data scarcity. We installed thirteen water level gauges equipped with signboards explaining the monitoring process to passers-by. Results were sent via a text-message-based data collection framework that included an immediate feedback to citizens. A public web interface was used to visualize the data. Within the first year, 124 citizens reported 1,175 valid measurements. We identified 13 citizens as active observers providing more than ten measurements, whereas 57% only sent one record. A comparison between the crowdsourced water level data and an automatic gauging station revealed high data quality. The results of this study indicate that citizens can provide water level data of sufficient quality and with high temporal resolution.

Key words

water resources management; hydrology; water level; East Africa; Sondu catchment; text message; Kenya

Highlights

- Hydrological monitoring is costly and often not achievable for low-income countries
- Involving citizens in the monitoring process can increase the amount of data
- Citizens reported water level for a remote catchment regularly and with high quality
- Crowdsourced data can be a valuable additional data source

1 Introduction

Water provides crucial ecosystem services for human beings and comprehensive hydrological knowledge is essential to manage this resource sustainably (Buytaert et al., 2014). However, water management strategies can only be effective if they are based on reliable monitoring. The absence of long-term data makes it difficult to develop sustainable management practices (Gilbert, 2010). While the available water data pool is arguably sufficient in developed countries, low-income countries are constrained by scarce data, restricting sustainable development (Buytaert et al., 2014). Ongoing climate and land use change processes influence water availability and, as a result, regional and local changes become more variable and difficult to predict (Jackson et al., 2001). Climate variability will increase pressure on the development of sustainable water resource management strategies, especially on the African continent (UNESCO, 2015). In addition, empirical evidence is required to advance our understanding of hydrological processes, e.g. observations are necessary to improve hydrological models (Royem et al., 2012). Fast developing African nations with an increasing water demand face the largest constraints to acquire and manage water data (UNESCO, 2003). However, the installation of comprehensive monitoring networks raise costs for technical equipment, personnel, management, and maintenance (Mazzoleni et al., 2017), especially in remote areas, where accessing the sensors for maintenance and data collection becomes a time-consuming task. In low-income countries, these installations and running costs may prevent the establishment and maintenance of water monitoring networks. The use of remote sensing technology to gain hydrological information as it is used to monitor large waterbodies is also not suitable for small streams due to the spatial and vertical resolution of the available data.

Citizen science projects have the potential to be a cost-effective way of gathering data and can reduce laborious or costly research problems (Bonney et al., 2014; Gura, 2013; Pocock et al., 2014; Tweddle et al., 2012). This seems to motivate decision-makers and non-governmental organizations worldwide, who are engaging volunteers for various monitoring responsibilities. In general, citizen science is described as a practice in which volunteers with no science background assist in conducting research (Raddick et al., 2010), generating new scientific knowledge (Buytaert et al., 2014), or collecting data without a direct integration into the scientific process (often referred to as crowdsourcing). Besides reducing costs, citizen science projects are an opportunity to link scientific work to the broader community. Involving the general public may increase public awareness and the public's attitude towards the topic investigated (Chase and Levine, 2017). Referring to the US National Science Foundation, citizen science projects are more readily funded, because they satisfy the requirement for "broader impact on society" of research grants (Gura, 2013). Consequently, citizen science publications have increased more than 10-fold within the last fifteen years (Tipaldo and Allamano, 2016).

Incorporating the general public in data assimilation has a long history in science. For example, the Christmas Bird Count by the National Audubon Society has been using eyewitness accounts to discover the distribution and abundance of birds in the United States for over 100 years (Audubon, 2017). Lowry and Fienen established a crowdsourcing approach to collect water level data in the U.S (Lowry and Fienen, 2013) by setting up a software called "Social.Water" (Fienen and Lowry, 2012). Starting with nine sites in 2011, their project monitors now more than 100 water level stations in lakes and streams over the United States. Breuer et al. conducted a crowdsourcing campaign to determine the spatial distribution of nitrogen solutes in German surface waters (Breuer et al., 2015). Especially low-income countries in Africa, like Kenya, can profit from this method of data collection to extend the spatial and temporal resolution of their monitoring networks. A wide range of actors, including NGOs and scientific organisations are engaged in in citizen science studies and citizen science increased its popularity in the media, with policymakers and the scientific community

(Pettibone et al., 2017). We chose Kenya to test this innovative way of data collection considering that Kenya is recognized as the economic hub of East Africa. The fast economic growth in this region will bring about new environmental concerns, challenging natural resource managers to adapt and to implement appropriate mitigation strategies. However, investments in a monitoring infrastructure are essential to make robust management decisions, but these investments are currently implemented at a relatively low speed in Kenya. Nevertheless, integrating the general public in collecting hydrological measurements is still an uncommon practice, since the measurements are more complex and often require expensive techniques (Buytaert et al., 2016). To support efficient use of water resources, sustainable water management and allocation plans have to be developed and implemented, thus requiring effective and reliable monitoring data. However, the Kenyan water sector of Kenya does not have the financial capacity to monitor natural resources with expensive high-tech equipment. New and affordable technologies have the potential to engage new actors in the monitoring process, transforming data collection from few data collectors toward a dynamic and decentralized network of citizens scientists (Buytaert et al., 2016).

The objective of this study was to determine whether engaging the citizens in a water level monitoring project is a suitable way to overcome data scarcity in remote catchments like the Sondu-Miriu River basin in Kenya. There are three research questions framing this study:

(1) Is citizen science a suitable approach to gather water levels in a remote tropical region?
(2) Is a text-message-based monitoring platform sufficiently user-friendly to be accepted by participants?
(3) Is the water level data gathered by the general public robust and trustworthy?

2 Materials and Methods

2.1 Study area

The study was conducted in the Sondu-Miriu River basin (3,450 km^2) located in Western Kenya (Figure 1). Elevation ranges from 1,140 m a.s.l. at the outlet of the basin at the Lake Victoria up to 2,900 m a.s.l. in the north-east region. The land use in the eastern region is dominated by smallholder agriculture and subsistence farming cultivating e.g. maize, beans, cabbage and potatoes. The central part of the basin is covered by the Mau Forest, Kenya's largest indigenous closed-canopy forest. Commercial tea and eucalyptus plantations, established in the first half of the 20th century (Binge, 1962) characterize the overall landscape in the north around the town of Kericho. A mixed land use pattern, consisting of smallholder agriculture and small settlements prevails towards Lake Victoria.

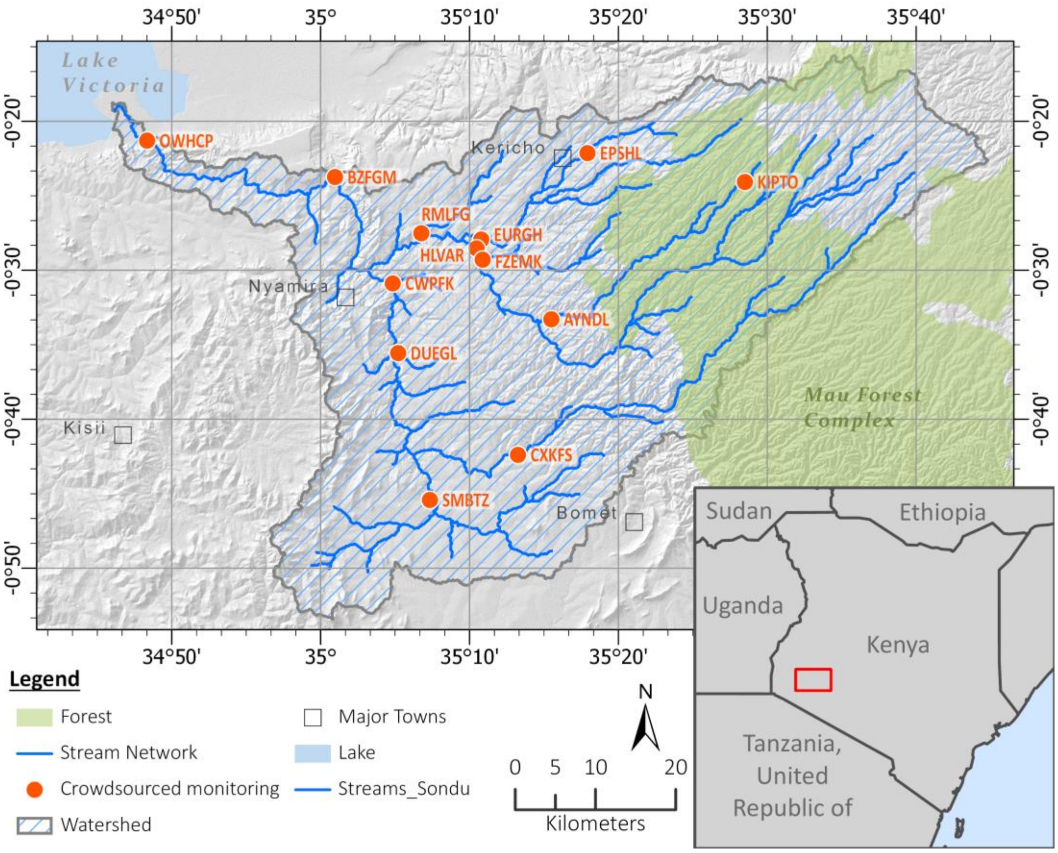

Figure 1: The Sondu-Miriu River basin in Kenya, including the stream network, major towns, natural forest areas, and the location of the crowdsourced monitoring stream gauging stations. The coordinates of the stations and additional information can be found in Table 1. Reference grid displays coordinates in WGS 1984.

The climate is influenced by the Intertropical Convergence Zone, resulting in a bimodal rainfall pattern with longer rainy seasons from April to July and a shorter rainy season between October and December. Monthly rainfall ranges from about 20 mm during the dry season to 180 mm during the rainy season (Olang and Kundu, 2011). Annual rainfall ranges from 1,300 mm yr^{-1} at the lower altitudes of the study area, to 1,900 mm yr^{-1} in the north-east region (Krhoda, 1988). The temperature does not show significant seasonality, but correlates with altitude. Highest temperatures, with an annual mean of 23°C have been recorded close to Lake Victoria (Vuai and Mungai, 2012), whereas the upland area around Kericho has a mean annual temperature of about 16°C (Stephens et al., 1992). Potential evapotranspiration rates range from 1,800 mm yr^{-1} at the lower altitudes to 1,400 mm yr^{-1} in elevated areas (Krhoda, 1988). Nitisols are common at the higher altitudes, whereas Acrisols are prevailing in the middle, and Regosols are mainly found at the lower parts of the basin (Vuai and Mungai, 2012).

The Mau Forest Complex provides critical water related ecosystem services e.g. water storage, river flow, flood mitigation, groundwater recharge, and micro-climate regulation (Benn and Bindra, 2011). Poor implementation of land use policies have resulted in a rapid forest degradation. More than one-quarter (100,000 ha) of the native forest have been lost within the last few decades (Khamala, 2010). This land use change had a negative impact on the hydrological cycle, resulting in an noticeable decline of discharge (Olang and Kundu, 2011).

2.2 Data collection

For this study, we installed thirteen locally-manufactured water level gauges at easily-accessible locations selected in agreement with the local water management authority, e.g. at public bridges (

Table *1*). Each monitoring site was equipped with a signboard placed next to the water level gauge (Figure 2) explaining the monitoring process using pictures and instructions in English as well as Swahili to invite passers-by to send data. Similar to the approach described by Fienen and Lowry (2012), participants read the water level and sent a text message, containing their record and the station-ID, which was indicated on the signboard. We aimed at keeping the method as simple as possible to minimise barriers for participation. Neither special equipment (like a smartphone with a camera) nor a mobile Internet connection or registration was required. The text message service is an easy to use, stable, inexpensive (0.01 USD each message) and established method of communication in East Africa. In addition, the system was designed to allow real-time feedback by sending response text messages to the observer.

Table 1. Station, site-ID, and geographical coordinates of the water level stations monitored in the Sondu-Miriu River basin, Kenya. Number of observations, the number of participants and the percentage of days with data for the period between April 2016 and March 2017 are given for every station.

Station name	site-ID	Coordinates[a]		Observations	Participants	Coverage[b]
		Latitude	Longitude			%
Kiptiget 1JA02	AYNDL	-0.554822	35.258283	74	10	18.6
Sondu 1JG05	BZFGM	-0.395118	35.015983	178	18	44.9
Kipsonoi 1JF08	CWPFK	-0.514703	35.080172	27	8	7.1
Kipsonoi 1JF06	CXKFS	-0.708547	35.221307	90	12	15.1
Kipsonoi 1JF07	DUEGL	-0.592747	35.086642	29	11	7.9
Kimugu 1JC03	EPSHL	-0.368775	35.298784	50	24	12.1
Ainabkoi 1JD04	EURGH	-0.465570	35.179745	53	12	13.2
Itare 1JB05	FZEMK	-0.488137	35.181330	9	5	1.9
Chemosit 1JB03	HLVAR	-0.475725	35.174287	27	12	6.0
Kuresoi	KIPTO	-0.401145	35.475240	434	15	74.2
Sondu 1JG04	OWHCP	-0.354440	34.805502	160	8	42.7
Lisere-Ainapkoi	RMLFG	-0.458506	35.112567	32	7	7.4
Lower Sisei	SMBTZ	-0.757450	35.122997	12	11	2.5

[a] WGS 1984 UTM Zone 36 S
[b] Percentage of the days between Apr 2016 and Mar 2017 with ≥1 observation per day

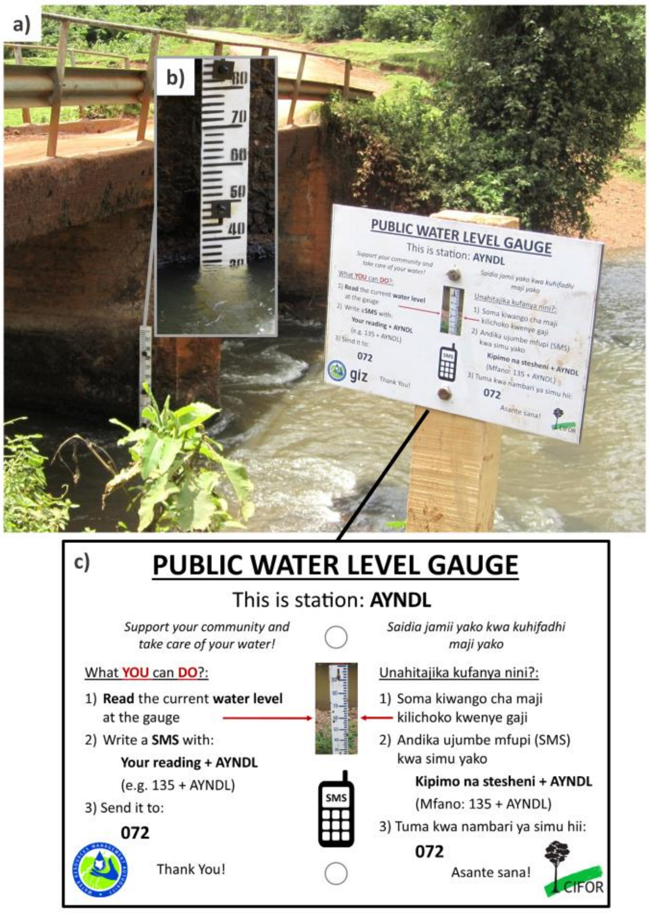

Figure 2: Example of the signboard (c) placed next to a water level gauge (b) (station AYNDL) (a). Simple and precise instructions make it easy for interested citizens to participate. Every gauge has an individual sign showing the station-ID.

To promote the project idea and assess its acceptance, several meetings were arranged with interested citizens at each site at the beginning of the project. These meetings were used to explain the measurement process and to train potential participants. It became evident that citizens, especially in the remote areas of the basin, had issues raising the money to send the data using their cell phones. To investigate if the lack of cash limits participation, we tested a reimbursement system for participants at the KIPTO station. The transmission costs (1 KES ≈ 0.01 USD) were reimbursed twofold for every valid observation sent. This payment was completed by transferring an aggregated monthly amount as cell phone credit to each observer and was limited to a maximum of 60 KES (i.e. thirty observations). The amount was automatically calculated and disbursed using an SMS-server as described in the section below. All other stations were operated without any reimbursement. The initial costs for the full monitoring network were low with approximately 6,000 USD for the gauges, mounting and signboards. Minor running costs were caused by on-site meetings with observers, the SMS-response and the webpage. The initial costs for simple pressure transducer to collect water level data automatically are substantially higher and need a regular maintenance and data collection, which causes further costs.

2.3 Description of the SMS-Server

2.3.1 General Approach

To collect and process the observations made by the citizens, we developed a software and hardware framework based on the general approach described by Fienen and Lowry (2012). Both approaches used text messages send by the observers to transmit the collected data and signboards placed next to the water level gauges explained the system for interested passers-by. Furthermore, both systems could handle spelling mistakes in the transmitted data using a text matching approach as described below. To adapt the idea to the local requirements in Kenya, we extended and changed the general approach. In contrast to the approach described by Fienen and Lowry (2012), where Google Voice is used to receive the text messages, we developed our own server infrastructure based on a Raspberry Pi 2 Model B. This allowed us to use the server outside the U.S., where Google Voice is not available, to avoid any dependency to the Google infrastructure and to provide a local cell phone number to ensure low transmission costs for participants. Furthermore, this approach allowed us to extend the functionality of the framework. We provided a real-time plausibility check of the data combined with a direct feedback to the participant by sending a text message fully automated by the server and imbedded a SQLite-database for data storing. In addition, we tested an automatic reimbursement system, where observers at one station received a cost compensation depending on the amount of valid data they sent. Further information regarding the technical implementation can be found in

Appendix 1.

2.3.2 Software

From the moment of sending an observation until the online presentation of the data, all transmitted messages underwent a process described schematically in Figure 3 and

Appendix 1. Based on the result of the plausibility check, the Python script automatically sent a feedback to the participant. Implausible data was flagged for further manual checking and the processed data was stored in the database. If a reading was valid, the participant received an SMS confirming the detected water level value and the station name associated with the site-ID. Furthermore, the number of previously reported values for the same site was given with an acknowledgment for the participation. If the water level sent was too high for the site, the participant was informed that the reading is above the maximum gauge height. Similarly, the participant was informed if the submitted site-ID did not coincide with a valid site-ID. Providing an immediate feedback using the same communication channel had several advantages. First, the participants were able to evaluate whether their contribution had the proper format or if they should check and resubmit the observation. Second, giving feedback about the number of collected data at the site could be an additional incentive and motivation to continue participating. The server was also used to calculate the amount of monthly reimbursement based on the amount of valid measurements per month for every participant were applicable. The reimbursement was then transferred automatically to the cell-phone of each participant using an interface provided by the Kenyan network operator. A website (www.uni-giessen.de/hydro/hydrocrowd_kenya) was created to publish the crowdsourced data. On the website, all processed data could be accessed with information about the individual monitoring sites. An interactive plot allowed interested citizens and authorities to view the hydrograph at each site and to download data for further use.

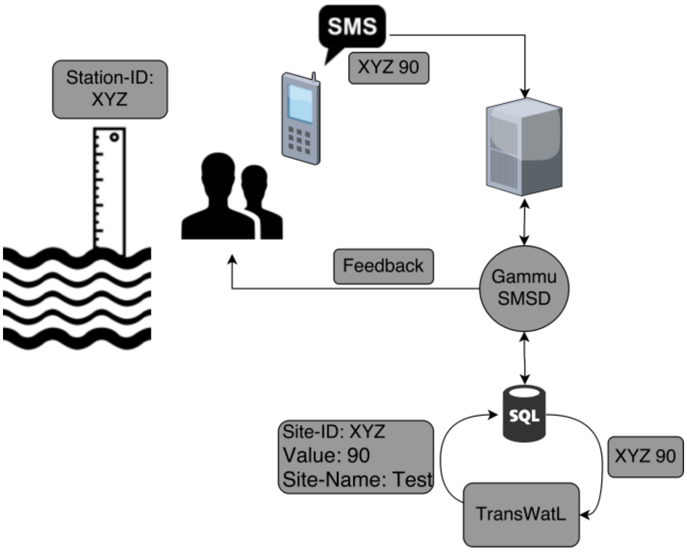

Figure 3: Schematic view of the crowdsourced data collection process. Observers read the water level and send a text message containing the value and a specific site-ID to a central server. The server stores the data received in a SQLite-database and an algorithm programmed in Python further processes the raw data and gives individual real-time feedback to observers.

2.4 Validation of data transmitted

To validate the crowdsourced data, a radar-based sensor (VEGAPULS WL61, VEGA Grieshaber KG, Schiltach, Germany) was placed twenty meters upstream of the KIPTO site, measuring water level data at ten-minute intervals. The hydrograph was inspected visually to estimate the quality of the crowdsourced collected data. Furthermore, the water levels at stations OWHCP and BZFGM, both located in the Sondu River, were evaluated and compared by assessing the difference of all standardized water levels collected on the same days for both stations.

2.5 Telephone survey

A telephone survey was carried out to obtain information about the socio-economic background of the participants. All participants were contacted using the phone number provided during the data transmission and asked to answer questions related to the project. This survey enabled us to give an overview about the gender, age and educations status of the volunteers.

3 Results

3.1 Received data

Between April 1st, 2016 and March 31th, 2017, 124 different participants reported 1,175 valid measurements. The amount of observations for each person varied from one (56.8% of the observers) to 224 transmitted values for the most active participant. Apart from station FZEMK, which was damaged during a flood event and therefore excluded from the analysis, citizens regularly reported measurements for most of the stations (Figure 4).

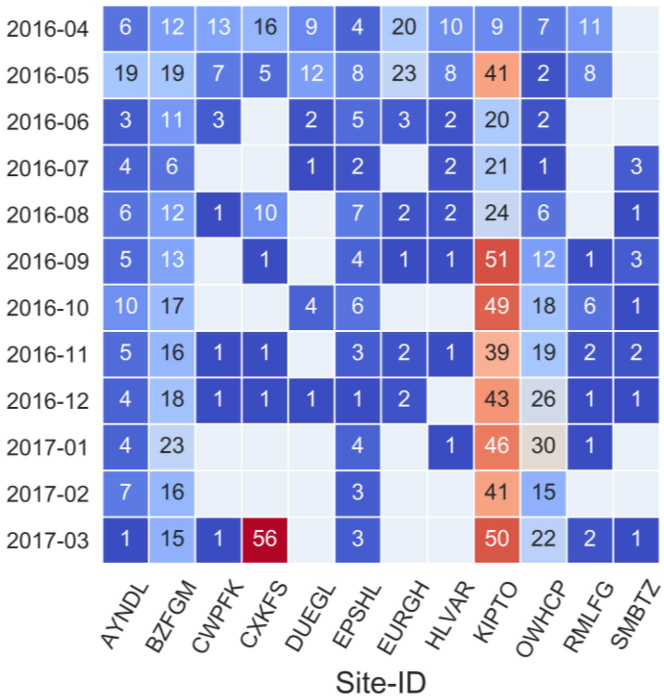

Figure 4: Monthly aggregated valid data for each station in the Sondu-Miriu River basin, Kenya, between April 2016 and March 2017. Dark blue indicates low activity, dark red very active months, and months without data received are grey.

It is noteworthy that even when some stations did not receive data for two or three months, these stations became active again (e.g. CXKFS, RMLFG). Most observations were reported after installing the gauges, when the citizens showed high interest in the project and the functionality of the system. Station KIPTO received the most measurements with 434 valid readings reported by fifteen different observers, followed by BZFGM and OWCHP with 178 and 160 observations, respectively. The station with the lowest amount of data was SMBTZ with only twelve received measurements (

Table 1*). The number of participants at each station did not vary greatly and ranged from seven individual observers at RMLFG to 24 observers at EPSHL.

Observers who reported more than ten water level records during the project period were considered active observers (AOs). Figure 5 gives an overview of the temporal resolution and the behaviour of the 13 identified AOs. Six observers continued transmitting values throughout the entire observation period, whereas the other seven AOs only sent messages for a certain period.

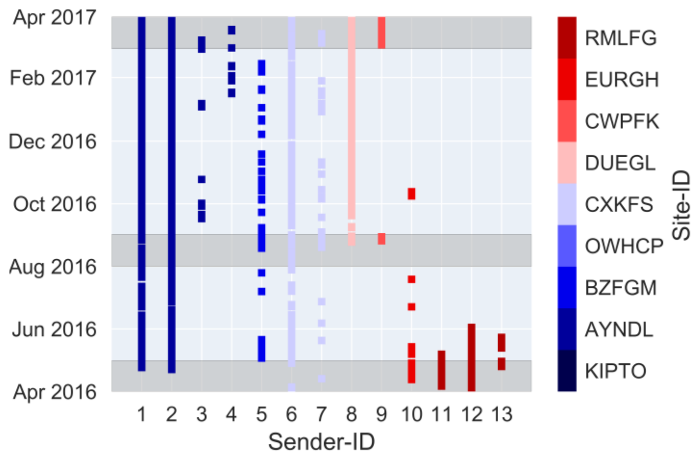

Figure 5: Temporal resolution of water level data in the Sondu-Miriu River basin in Kenya reported by active observers (more than ten observations during the observation period) in the period from April 2016 to March 2017. Every dot represents a measurement from the observer (Sender-ID). The related station is indicated by the colour as described in the colour ramp to the right. Grey rows mark wet periods with more than 120 mm precipitation per month.

While most of the AOs began participating during the initial project phase, some AOs joined after the project was already in progress. AOs were consistently sending data from one station, i.e. they did not move within the study area. The majority of AOs transmitted data for the full observation period. Some of them also resumed their work after long intervals without any transmission. Only a few AOs left the project after six to eight weeks. The wet periods, defined as months with more than 120 mm precipitation, did not influence the behaviour of the AOs, i.e. the amount of observations neither increased nor decreased during wet periods. Even though new participants joined in from time to time, most data was generated by AOs sending several readings each month. Only the minority of data (17%) was generated by random passers-by sending less than ten values.

Even though we aimed at keeping the system as simple as possible, not every text message provided by the citizens contained valid or interpretable data. Fifty-nine messages were marked as invalid (5%). Most of these errors were induced by misuse (e.g. citizens trying to apply for a job as regular gauge readers), mistyping as well as omitting the station-ID or the value. While the latter type of error can be handled by the system providing an immediate response to the observer, the first type of error causes unusable data, which were excluded from further analysis. Table 2**Fehler! Verweisquelle konnte nicht gefunden werden.** shows typical text messages containing invalid data detected and marked by the system.

Table 2. Examples for typical text messages containing errors or invalid readings. All messages have been automatically marked as invalid by the SMS-server. Some sentences have been partly corrected for spelling and grammar.

No.	Message	Problem
1	The level of water is 155	Station-ID missing
2	Wish to work with you. Kindly consider me when a chance arise. Thanks in advance	Applying for a job
3	What do you give me if I am sent the waterlevel everyday?	Applying for a job
4	Chemosit bridge 135+160=295	Real name of the site. Two readings at once (-> Invalid time stamp)
5	176	Station-ID missing
6	30 ml	Station-ID missing
7	Hi I’m Vincent, I am at KUREXOI NORTH. I am happy to express your support for water as source of life	Requested further information about the project
8	When you will be back again? I want to join you as an environmental volunteer	Requested information about the project

3.2 Data quality and validation

Comparison of data recorded by the radar sensor and the crowdsourced data at Station KIPTO showed similar trends in both datasets (Figure 6). Given that the radar was installed upstream, the observations from the radar and from the participants cannot be compared precisely, even when the shape and condition of the riverbed was almost similar. The citizen reported water levels systematically deviate from the water levels recorded by the radar during high-flow and low-flow conditions was related to the different cross-sections between the two locations. The visual comparison of the radar data with the crowdsourced water levels depicted a good agreement. Both datasets showed similar behaviour to rainfall events in terms of rising and falling water levels. Both high flow and base flow conditions were measured accurately by the citizens.

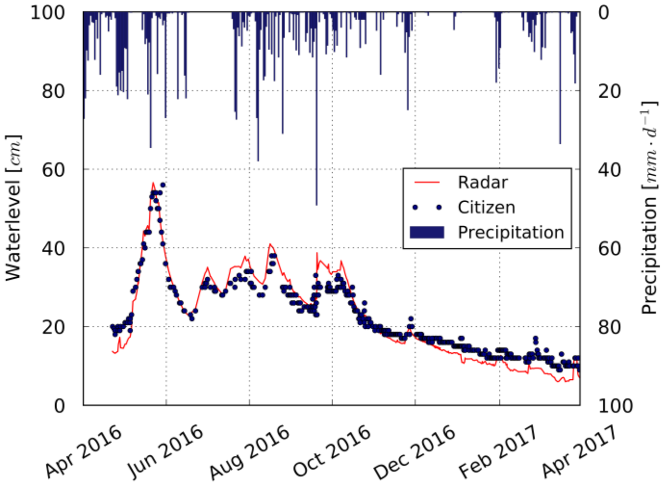

Figure 6: Time series of citizen-transmitted and validation data at the KIPTO catchment in the period from April 2016 to March 2017. Validation data generated by a VEGA radar sensor is displayed as a red line, the citizen science data is displayed usng blue dots. The blue bars show daily rainfall data measured by an ECRN-100 tipping bucket 120 meters to the north-west of the gauge.

As a second benchmark, we compared the data of two stations: BZFGM and OWHCP, which is located 35.5 km downstream of station BZFGM, both within the Sondu River. Because of the proximity of the stations without significant tributaries flowing into the river between these stations, we expected a uniform trend for both hydrographs when comparing measurements recorded on the same day. Due to the distance between stations, we assume that the observers did not know one another. Therefore, we considered the samples independent. Data collected by the citizens would be reliable if the measurements reported were correlated. In contrast, we would expect a weak correlation if the crowdsourced data contained large random errors. To make the data of both stations comparable, we normalized the water level readings and plotted them together with the differences between both observations (Figure 7). With this transformation we are now able to compare the water level changes of both stations taking into account that the riverbed between this two stations is different (and therefore give a systematically bias of the absolute values). Both stations clearly followed the same trend and did not show a distinctive drift over the year. The difference between the normalized water level of the two stations moved around the zero line suggesting a reliable and unbiased data acquisition for these stations.

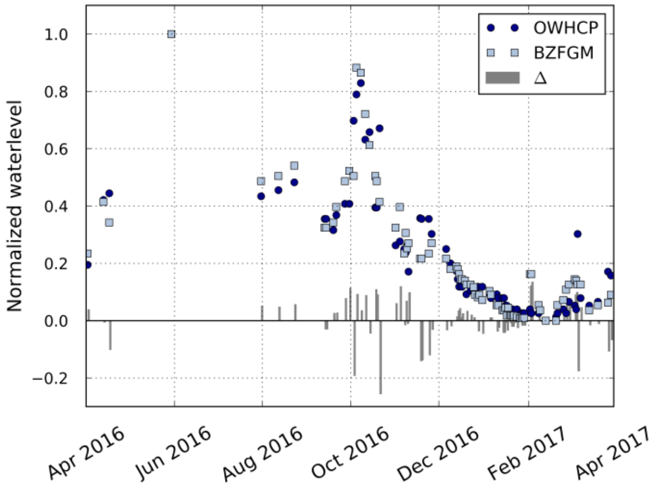

Figure 7: Standardized water level data and their differences (Δ) observed on the same day for two nearby stations (OWHCP and BZFGM) close to the outlet of the Sondu-Miriu river basin in Kenya between April 2016 and Mach 2017. The water levels transmitted for both stations follow the same trend and do not show a deviation over the time indicating reliable data reported by citizens.

3.3 Socioeconomic background of the participants

During the telephone survey, 87 observers were reached and agreed to participate. From thirteen identified AOs, twelve could be contacted by phone. One AO, who was active from January to March 2017 was not reachable and the phone number was not online anymore. *Table 3* shows the distribution of gender, age and education of the twelve AOs in comparison to 75 observers which contributed less than ten values.

Table 3: Gender, age and education level of 87 observers contacted during a telephone-survey campaign. The data was divided in answers provided by active observers, which transmitted more than ten values (AO) and observers which reported ten or less observations (Other)

		AO (n = 12)	Other (n = 75)
Gender [%]	Female	25	3
	Male	75	97
Mean Age		40	33,5
Education [%]	Primary	50	20
	Secondary	42	36
	High	8	37
	No Answer	0	7

The survey showed that the AOs in our study were older and of lower educational background. Three out of five women became an AO, while two reported less than ten observations.

4 Discussion

In this study, we tested whether involving citizens in the monitoring process could help to overcome the low spatial and temporal resolution of water level data. After one year of water level monitoring conducted by volunteers, we were able to assess the overall performance of this innovative data collection method in a remote tropical catchment.

4.1 Motivation and participation of citizens

High enthusiasm was shown by participants, which resulted in more than 1,100 valid data points for thirteen monitoring sites within the observation period from April 2016 to March 2017. The thirteen most AOs reported 83% of all data. Only 17% were reported by citizens, which sent ten or less values. This indicates that especially some persons identify themselves with the project and the idea of monitoring their environment. Whereas most of the AOs participated over the full project period, some new observers joined the project later. We attribute the increase in participation to a recruitment by other motivated observers, who were positive about the project. In combination with the socioeconomic background of the AOs and all participants we conclude that the active participation is not depending on the actual education level but rather induced by their personal perception of and dependency on their environment. Especially citizens who depend on local water resources are expected to be interested in increasing their understanding of their environment and to participate in local political decisions to ensure a sustainable use of their resources (Overdevest et al., 2004). We experienced a similar behaviour during our field campaigns, where especially farmers of smallholder areas were interested in monitoring their water resources. Besides the increment of data, the participation of citizens can potentially lead to other positive side-effects. It has been observed that participants who increase their understanding of local resources, motivate neighbours and form opinions to support local policies (Overdevest et al., 2004). At the same time, low participation rates at some stations can be attributed partly to the transmitting cost of 0.01 USD per text message, which was paid by the volunteers. Especially in rural areas, participants expressed that they might be unable to participate due to costs. Buytaert et al. (2014) described that observers in low-income countries often derive an income from their engagement in citizen-science projects. These authors argue, that the concept of sending data voluntarily is not well developed, and that it may be necessary to reward people at local wages for motivation. We found that paying a small reward that covers the costs significantly increases the overall participation rate. In comparison to the other stations, the amount of data reported for station KIPTO, where a reimbursement system was set up, is seven times larger than the average of reported data from stations without reimbursement system and 2.5 times larger in comparison with the second most active station BZFGM. By paying back the transmission costs twofold, the motivation of the observers may remain strong over a longer period. The same behaviour was observed for station OWHCP, where the amount of data transmitted significantly increased after August 2016 (Figure 4). Instead of a reimbursement centrally paid by the project, interested water users organized an own reward system by collecting a contribution from several users to reimburse one person recording the water level data. However, a real payment or reward was not necessary, since the intrinsic motivation of the participants seemed to be sufficient when lack of money was overcome.

Transmitting the observations using simple cell phones and text messages turned out to be stable and reliable without major technical problems. Text messages are a common way of communication and significantly lowered the technical barrier to contribute and send data. The use of this communication channel was widely accepted. Furthermore, the participants were able to send text messages without additional training. The SMS-server was available most of the time. Only during the initial phase we faced minor problems caused by unstable drivers of the GSM-modem used, resulting in a loss of data for some transmitted values. This issue was fixed by changing the GSM-modem. Furthermore, the feedback loop allows participants to identify whether their observation was correctly received. We occasionally faced phone network coverage issues. Due to the location of the water level gauges in valleys, mostly in remote areas, the network coverage at the monitoring point was sometimes weak. However, those stations with restricted network availability did not turn out as a limited factor for data contribution. Observers took the readings of the water level and waited until they reached an area with network coverage to send their messages. This led to a minor deviation of the time of the record since the time stamp is generated from the text message header. However, we expect that the observers sending messages after a couple of minutes rather than waiting several hours. In comparison to more sophisticated methods, like using smartphones, we believe that this approach produces more and, in turn, more reliable results in a low-income country because wrong data and outliers become obvious.

4.2 Data accuracy and suitability

The quality and temporal resolution of the crowdsourced data is important to assess their usefulness. The comparison of the citizen data with data measured by an automatic radar sensor at station KIPTO revealed a high correlation between these datasets. Intensive training of the participants was not necessary to ensure high quality data. Fienen and Lowry (2012) obtained a RMSE (4.88×10^{-3} m) between crowdsourced data and a pressure transducer, from which the authors concluded, that the observations of relatively simple parameters can be efficiently conducted by citizen scientists. From 83 citizen science studies evaluated by Aceves-Bueno et al. (2015), only one study reported an insufficient data quality. Our results showed that citizens provided data comparable to conventional data loggers. From over 1,000 recorded data points, less than 5% were invalid and therefore not useable for further analysis. In most cases, these errors were caused by participants trying to submit or inquire additional information that cannot be handled automatically by the system. In these cases, a personal interaction with the participants is necessary. The research team or data managers of citizen science projects should evaluate this additional information to recognize further demands of the participants. Regarding the temporal resolution, we observed a large variability between the stations. While some stations have data for 50, and even up to 75% of the days per year, other stations only received data for less than 15% of the days per year.

It seems that citizens cannot deliver the same temporal resolution as modern automated monitoring equipment. However, hydrological models can play an important role to fill gaps in irregular measurements taken by citizens. Seibert and Vis (2016) evaluated whether stream level data without an established rating curve would be sufficient to calibrate a simple hydrological model using the Spearman rank correlation coefficient. The authors observed, that a water level time series is already sufficient to obtain a good model performance in wet catchments where precipitation is higher than the potential evapotranspiration. The Sondu-Miriu River basin has both: wet areas in the elevated parts and dry areas towards Lake Victoria, making it a good place to test this approach. In a recent study van

Meerveld et al. (2017) demonstrated, that this approach is applicable also with a reduced vertical resolution of stream level data. Seibert and Beven (2009) demonstrated, that a few discharge observations were already sufficient to calibrate a model for several catchments in Sweden. After adding 32 observations, the authors did not obtain an improvement of the average model performance. In a follow up study Pool et al. (2017) showed, that already twelve strategically sampled discharge measurements have the potential to calibrate simple hydrological models across the eastern US. Mazzoleni et al. (2017) demonstrated, that (synthetic) crowdsourced discharge data complements traditional monitoring networks when used for flood forecasting even when the crowdsourced data were characterized as asynchronous. In a review written by Assumpção et al. (2017) the authors concluded that crowdsourced data can be integrated in hydrological models and improve their overall performance. Other studies reveal that citizen are particularly interested in monitoring extreme events, which could be a valuable support in the flood risk assessment (Le Coz et al., 2016). Based on our experience and that of others in different regions, we see a great potential to use crowdsourced water level data to extend conventional monitoring networks.

4.3 Towards citizen-based monitoring

One of the two most commonly cited reasons for unsuccessful management strategies is the lack of proper monitoring data (Aceves-Bueno et al., 2015). We argue that the simplicity and cost-effectiveness of our method has the potential to create new insights in the hydrological cycle and can support the decision process of local water managers. We agree with Buytaert et al. (2014), that data collected by citizens can create new hydrological knowledge and help to identify the human impacts on the water cycle, especially in remote regions. Involving the general public in monitoring can increase drastically the amount of environmental observations. It is necessary that scientists and resource managers accept the data collected by the general public to use them for further analysis (Freitag et al., 2016). Based on 83 peer-reviewed published papers on citizen science case studies in natural resource management settings, Aceves-Bueno et al. (2015) concluded, that in 41% of the studies the data gathered by the general public was used to make management decisions. We conclude that using data collected by citizens for simple measurements should be taken into account as a valuable data source. Moreover, citizen science projects should not only be considered as possible data source, but also as a great opportunity to support citizens in generating further knowledge about their environment and, additionally, to bring often complex research projects closer to the communities. It has been observed, that crowdsourced based monitoring increases the volunteers' awareness of their local resources and a multiplier effect, where volunteers share the knowledge gained with other community members (Storey et al., 2016). We also noticed these multiplier effects in our projects where new volunteers stepped in and actively contributed data, most likely after being motivated by other observers.

Overall, the results of our study indicate that citizens have the ability to record water level data of a sufficient quality and quantity. However, prospective experiments should be conducted to analyse further the precision of the citizen science data. We plan to install additional automatic water level sensors next to the citizen monitoring stations to investigate the long-term precision and accuracy of the crowdsourced data. As a next step, we will test the usefulness of the crowdsourced data for hydrological modelling and upscaling purposes. We plan to set up and run simple models and compare if the increased spatial resolution of the data collected by citizens has the potential to increase the model performance. Furthermore, we plan to assess if only the water level data is useful to calibrate models in a tropical catchment using the method described by Seibert and Vis (2016) To overcome poor participation due to text message costs that have to be covered by observers, we suggest to establish a toll-free number, which allows observers to transmit their data without any costs. Alternatively, if a toll-free number cannot be established, the influence of a reward system on the data quality and quantity should be systematically tested. Finally, we plan to investigate whether the framework presented in the study can be used to collect more sophisticated data like water quality parameters.

5 Conclusion

The increasing demand for water makes it necessary to use this resource more efficiently based on sustainable management strategies and monitoring solutions. Citizen science programs are promising cost-efficient methods to monitor environmental resources, which make them especially suitable for low-income countries to overcome their sparse data resolution. Since today's citizen science studies are mostly located in high-income countries, we are enthusiastic to motivate the scientific community to conduct citizen science studies in low-income countries. Overall, our study shows that involving the local community in the water level data collection in a remote Kenyan basin generates good quality data and is promising to deliver new insights into the hydrological processes. It is important to understand the driving factors that keep participants motivated. Giving feedback to the participants is necessary, since it keeps the participants updated and prevents raising unrealistic expectations associated with the monitoring, management plans or rewards. By using the text message system for the data collection, we were able to give fast and individual feedback.

We conclude that:

(1) The interest and motivation of the citizens can be considered as one of the leading reasons to decide whether a citizen science approach is applicable. Our research has shown that it is possible to engage community members to conduct water level monitoring resulting in more than 1,000 measurements within the first year.

(2) Text messages are a common way of communication in Kenya and were accepted as a method to contribute data. Since this method does not rely on expensive smartphones or an Internet connection, this approach lowers the technical barrier of participation. A small reimbursement covering the costs has the potential to improve participation.

(3) Crowdsourced data can be a valuable additional data-source to monitor water resources. Data delivered by citizens is reliable, consistent and of similar quality to data collected by an automatic radar.

For the Sondu-Miriu River basin in particular the collected water level data has the potential to support the development of water allocation plans, which becomes evermore essential due to the increasing water demand in this region. The basin currently does not have a sufficient water allocation plan, which can be attribute to the data scarcity in this region. Local Water Resource User Associations could profit from additional data to develop small-scale sub-catchment management plans, which are part of their assignment. Members of Water Resource User Associations expressed their interest in the data for this purpose during personal talks with the authors. Coupled with river discharge data, this data can furthermore be used to develop strategies to prevent or mitigate flood-related disasters, which affects people living in the lower part of the basin in particular. This population suffers from floods and droughts and it can be expected that these effects will increase with ongoing climate change.

Acknowledgments

We would like to thank the citizens of the Sondu Basin who participate from our programme, and the Water Resources Management Authority (WRA) of Kenya for supporting the research team to conduct this research.We also thank the Deutsche Forschungsgemeinschaft DFG (BR2238/23-1), and the Deutsche Gesellschaft für Internationale Zusammenarbeit GIZ, German Federal Ministry for Economic Cooperation and Development (Grants 81195001 "Low Cost methods for monitoring water quality to inform upscaling of sustainable water management in forested landscapes in Kenya" and Grant 81206682 "The Water Towers of East Africa: policies and practices for enhancing co-benefits from joint forest and water conservation") for generously providing financial support.

Appendix 1

This appendix gives further information about the technical implementation of the developed SMS-server handling the data reported by citizens. The server was connected to the local cell-phone network using a mobile broadband modem (ZTE MF 190) and a SIM-card from a local mobile network operator. The power supply was ensured by connecting the server to the local electricity network. Additionally, a 10,000 mAh powerbank was connected, acting as an uninterrupted power supply. In case of power cuts, the powerbank was able to provide electricity for another 24 hours. To handle the incoming text messages we used the Gammu SMS Deamon (Gammu SMSD), which collected the text messages from the modem and storeed them in a SQLite database using the 'libdbi backend'. SQLite was chosen because of its high performance and the absence of multi-user-access needs on the server. However, more complex database systems, like MySQL or PostgreSQL, could be easily integrated if required. After receiving and storing the raw data, data was further processed to ensure consistentcy using a Python script developed for this project. This script retrieved the raw data from the database, extracted the specific site identifier (site-ID) as well as the transmitted water level value and verified the data plausibility. Data became implausible if the new water level value was higher than the gauge height at the associated site or if the submitted site-ID did not match any of the existing site-IDs. If the script detected questionable data, the observation was flagged to allow a manual correction where applicable. To avoid errors caused by mistyping, the submitted site-ID was extracted and compared with all existing site-IDs using the Levenshtein Distance. As a result, the most likely site-ID was returned with a matching factor ranging from zero (no similarity) to one-hundred (perfect match). We used the python package “fuzzywuzzy“ (Cohen, 2016), to implement the Levenshtein distance calculation and to determine the differences between the string sequences of the incoming station name and the existing stations. A regular expression (\d+[\.,]?\d*) was applied to extract the water level value from the text message. If a message contained more than one value, only the first value was extracted for further analysis.

References

Aceves-Bueno E, Adeleye AS, Bradley D, Brandt WT, Callery P, Feraud M et al. Citizen Science as an Approach for Overcoming Insufficient Monitoring and Inadequate Stakeholder Buy-in in Adaptive Management: Criteria and Evidence. Ecosystems 2015;18(3):493–506.

Assumpção TH, Popescu I, Jonoski A, Solomatine DP. Citizen observations contributing to flood modelling: Opportunities and challenges. Hydrol. Earth Syst. Sci. Discuss. 2017:1–26.

Audubon. Join the Christmas Bird Count | Audubon: Join the Christmas Bird Count, 2017.

Benn J, Bindra S. UNEP annual report 2010 a year in review. Nairobi: United Nations Environment Programme (UNEP); 2011.

Binge FW. Geology of the Kericho Area: Ministry of Commerce, Industry and Communications; 1962. Geological Survey of Kenya 50.

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ et al. Citizen science. Next steps for citizen science. Science (New York, N.Y.) 2014;343(6178):1436–7.

Breuer L, Hiery N, Kraft P, Bach M, Aubert AH, Frede H-G. HydroCrowd: a citizen science snapshot to assess the spatial control of nitrogen solutes in surface waters. Scientific Reports 2015;5:16503.

Buytaert W, Dewulf A, Bièvre B De, Clark J, Hannah DM. Citizen Science for Water Resources Management: Toward Polycentric Monitoring and Governance? Journal of Water Resources Planning and Management 2016;142(4):1816002.

Buytaert W, Zulkafli Z, Grainger S, Acosta L, Alemie TC, Bastiaensen J et al. Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. Frontiers in Earth Science 2014;2.

Chase SK, Levine A. Citizen Science: Exploring the Potential of Natural Resource Monitoring Programs to Influence Environmental Attitudes and Behaviors. Conservation Letters 2017.

Cohen A. FuzzyWuzzy. Fuzzy string matching in python. 0th ed; 2016.

Fienen MN, Lowry CS. Social.Water—A crowdsourcing tool for environmental data acquisition. Computers & Geosciences 2012;49:164–9.

Freitag A, Meyer R, Whiteman L. Strategies Employed by Citizen Science Programs to Increase the Credibility of Their Data. Citizen Science: Theory and Practice 2016;1(1):2.

Gilbert N. How to avert a global water crisis. Nature 2010.

Gura T. Citizen science: Amateur experts. Nature 2013;496(7444):259–61.

Houska T, Kraft P, Chamorro-Chavez A, Breuer L, Hui D. SPOTting Model Parameters Using a Ready-Made Python Package. PLOS ONE 2015;10(12):e0145180.

Jackson RB, Carpenter SR, Dahm CN, McKnight DM, Naiman RJ, Postel SL et al. Water in a changing world. Ecological Applications 2001;11(4):1027–45. http://onlinelibrary.wiley.com/doi/10.1890/1051-0761(2001)011[1027:WIACW]2.0.CO;2/full.

Khamala E, editor. Acting Today to Save our Tomorrow: The Case of Restoring the Mau Forest, Kenya; 14-17.06; 2010.

Krhoda GO. The Impact of Resource Utilization on the Hydrology of the Mau Hills Forest in Kenya. Mountain Research and Development 1988;8(2/3):193.

Le Coz J, Patalano A, Collins D, Guillén NF, García CM, Smart GM et al. Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. Journal of Hydrology 2016;541:766–77.

Lowry CS, Fienen MN. CrowdHydrology: Crowdsourcing Hydrologic Data and Engaging Citizen Scientists. Ground Water 2013;51(1):151–6.

Mazzoleni M, Verlaan M, Alfonso L, Monego M, Norbiato D, Ferri M et al. Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? Hydrology and Earth System Sciences 2017;21(2):839–61.

Olang L, Kundu P. Land Degradation of the Mau Forest Complex in Eastern Africa: A Review for Management and Restoration Planning. In: Ekundayo E, editor. Environmental Monitoring: InTech; 2011.

Overdevest C, Orr C, Stepenuck K. Volunteer Stream Monitoring and Local Participation in Natural Resource Issues. Research in Human Ecology 2004;11(2):177–85.

Pettibone L, Vohland K, Ziegler D. Understanding the (inter)disciplinary and institutional diversity of citizen science: A survey of current practice in Germany and Austria. PLOS ONE 2017;12(6):e0178778.

Pocock MJO, Chapman DS, Sheppard LJ, Roy HE. A Strategic Framework to Support the Implementation of Citizen Science for Environmental Monitoring: Final Report to SEPA. Wallingford, Oxfordshire: Centre for Ecology & Hydrology; 2014.

Pool S, Viviroli D, Seibert J. Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? Journal of Hydrology 2017;554:613–22.

Raddick MJ, Bracey G, Gay PL, Lintott CJ, Murray P, Schawinski K et al. Galaxy zoo: Exploring the motivations of citizen science volunteers. Astronomy Education Review 2010;9(1).

Royem AA, Mui CK, Fuka DR, Walter MT. Technical Note: Proposing a Low-Tech, Affordable, Accurate Stream Stage Monitoring System. Transactions of the ASABE 2012;55(6):2237–42.

Seibert J, Beven KJ. Gauging the ungauged basin: How many discharge measurements are needed? Hydrology and Earth System Sciences 2009;13(6):883–92.

Seibert J, Vis MJP. How informative are stream level observations in different geographic regions?: Value of Stream Level Observations. Hydrological Processes 2016;30(14):2498–508.

Stephens W, Othieno CO, Carr MKV. Climate and weather variability at the Tea Research Foundation of Kenya. Agricultural and Forest Meteorology 1992;61(3-4):219–35.

Storey RG, Wright-Stow A, Kin E, Davies-Colley RJ, Stott R. Volunteer stream monitoring: Do the data quality and monitoring experience support increased community involvement in freshwater decision making? E&S 2016;21(4).

Tipaldo G, Allamano P. Citizen science and community-based rain monitoring initiatives: An interdisciplinary approach across sociology and water science: Citizen science and community-based rain monitoring initiatives. Wiley Interdisciplinary Reviews: Water 2016:e1200.

Tweddle JC, Robinson LD, Pocock MJO, Roy HE. Guide to citizen science: Developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK. [Swindon]: UK Environmental Observation Framework; 2012.

UNESCO. Water for people, water for life: A joint report by the twenty-three UN agencies concerned with freshwater. New York: Unesco; 2003.

UNESCO. Water for a sustainable world includes data and indicators annex for water and energy. Paris: Unesco; 2015.

van Meerveld HJI, Vis MJP, Seibert J. Information content of stream level class data for hydrological model calibration. Hydrol. Earth Syst. Sci. 2017;21(9):4895–905.

Vuai SAH, Mungai N. Spatial Variation of Nutrients in Sondu-Miriu and Simiyu-Duma Rivers:Implication on Sources and Factors Influencing their Transportation into the Lake Victoria. Journal of Earth Science & Climatic Change 2012;03(02).