From the Department of Landscape Ecology and Resources Management

A dissertation to obtain the degree of Doctor of Natural Sciences
(Dr. rer. nat.) in the Faculty of Agricultural Science, Nutritional Science and
Environmental Management of the Justus Liebig University Gießen

MODELING HYDROLOGICAL FLUXES OF TROPICAL MOUNTAINOUS WATERSHEDS IN KENYA USING CROWDSOURCED WATER LEVEL DATA

BJÖRN WEESER

Referees from the Justus Liebig University Gießen:

Prof. Dr. Lutz Breuer

Prof. Dr. Martin Petrick

Submitted: 16th October 2020

Abstract

Climate change and a growing population alter established water usage pathways in Eastern Africa and create an urgent need for effective and sustainable water management strategies. However, required data to develop such strategies is often missing, especially in remote regions. This dissertation examines (1) whether water level data collected by citizens can improve the hydrological database, (2) how this data can be used to establish rainfall-runoff models, and (3) the socio-economic background and motivation of citizens to participate in data collection or reasons that prevent them from continuing.

First, a crowdsourced water level monitoring network was established at thirteen locations within the Sondu-Miriu River basin located in Western Kenya. Interested citizens were invited to record water level data and report these values by sending a simple text message using their cellphone. Over a period of 3.5 years 258 citizens reported 3,480 valid data points. Validation against water level data collected by an automatic radar station at one of the sites revealed high data quality.

In a second step, a conceptual rainfall-runoff model was calibrated on water level data collected by citizens using Spearman-Rank-Coefficients between the simulated discharge and the water levels. Considering a water balance filter derived from measured precipitation and remotely sensed evapotranspiration, the model calibrated on crowdsourced data reached a model efficiency close to values obtained from a benchmark model that was built using automatically measured discharge data (Nash-Sutcliffe-Efficiency of 0.69 compared to 0.88).

Finally, a telephone survey among the participants in the monitoring project revealed that those who submitted data over a long period were generally between 30 and 50 years old and hold a primary or secondary school diploma. Many participants stated that helping water management and conservation purposes were their primary motivation of involvement. Sensitization meetings were mentioned as the main source of information about the project by long-term participants.

This reserach shows that crowdsourced monitoring approaches are a promising additional tool for water resources management, particularly in ungauged or poorly gauged catchments and under limited financial resources. These findings can be used to support the development for sustainable community-based water monitoring programs.

Table of content

List of tables

List of figures

1 Extended summary

1.1 Introduction

Water provisioning is one of the most fundamental ecosystem services for human beings (Buytaert et al. 2014). However, stressors such as climate change, population growth, and land use changes put pressure on this resource and jeopardize established water usage pathways both for human society and for nature itself (Johnson et al. 2007, Everard 2012). Falcone et al. (2010) postulated that these stressors directly affect the hydro-biogeochemical processes of ecosystems and impairs their resilience to extreme events or other disruptive factors. Land use change and climate variability further alter the availability of water in catchments and make it difficult to predict local and regional changes (Jackson et al. 2001). Buytaert et al. (2014) suggested that insufficient water supply often represents a significant bottleneck for sustainable development and poverty alleviation. Consequently, the changes and effects caused by stressors obstruct the achievement of the Sustainable Development Goals defined by the United Nations in the Agenda 2030, which are intended to ensure a better and more sustainable future (United Nations 2015).

Sustainable water management strategies are crucial to minimizing the impact of negative effects on water availability. The evidence-based decision making that is needed for sustainable water management requires dense hydrological monitoring networks with a high temporal and spatial resolution (Mishra and Coulibaly 2009, Ochoa-Tocachi et al. 2018). Grab sampling approaches are often too expensive for regional or national monitoring programs (Hildebrandt et al. 2006) and can miss short hydrological events (Jacobs et al. 2018b). Permanently installed automatic monitoring stations, like river gauging stations, are prone to corrosion, vandalism, and theft and therefore require routine site maintenance and security (Gomani et al. 2010, Hannah et al. 2011, van Overloop et al. 2014). In addition, remote locations are often inaccessible, which further limits the amount of data that can be collected with available resources (Zheng et al. 2018). Data restriction policies delay data release (Vörösmarty et al. 2001) and limit the use of data for water resources management, especially when up-to-date information is required

(Wagner et al. 2009). Hence, substantial costs and challenges in practical implementation lead to sparse data collection and irregular monitoring. While the available data pool is frequently sufficient in developed countries, low-income countries are often constrained by data scarcity which complicates or prevents the implementation of sustainable management practices and sustainable development (Gilbert 2010, Buytaert et al. 2014, Jacobs et al. 2018b, Rufino et al. 2018).

Recent literature underlines the fact that hydrological data in large parts of the world is incomplete and the lengths of the time series are insufficient to characterize and adequately manage water resources (Mishra and Coulibaly 2009, Chacon-Hurtado et al. 2017). As a result, research increasingly focuses on alternative data collection methods. Besides the use of remote sensing technology for meteorological and discharge data (Smith et al. 1996), studies also investigated the use of cameras (Le Coz et al. 2016, Jiang et al. 2019), social media (Le Boursicaud et al. 2016, Chaudhary et al. 2019), cell phone networks (Gosset et al. 2016) or privately operated weather stations (Bell et al. 2013) to gather additional information.

Relatively new are citizen science methods for monitoring environmental data, which have received increasing attention from the scientific community and the public during the last years (Njue et al. 2019). Citizen science is described as a practice in which volunteers are involved in the scientific research process such as collecting, categorizing, transcribing, or analyzing scientific data (Bonney et al. 2009). The European Commission (2013) defined citizen science as a "general public engagement in scientific research activities where citizens actively contribute to science either with their intellectual effort, or surrounding knowledge, or their tools and resources". In the literature, common terms like volunteer-based monitoring (Deutsch and Ruiz-Córdova 2015), crowdsourcing (Howe 2006), community-based monitoring (Palmer Fry 2011), citizen observatories (Liu et al. 2014), or participatory monitoring (Danielsen et al. 2005) are used to describe different forms of public participation in scientific processes.

Over the last twenty years, citizen science projects are considered as a promising approach for long-term monitoring of local and global environmental change (Danielsen et al. 2005, Silvertown 2009, Johnson et al. 2014, McKinley et al. 2017). These projects can be a cost-effective way of data collection and support the implementation of otherwise labor-intensive or expensive research problems (Tweddle et al. 2012, Gura 2013, Bonney et al.

2014, Pocock et al. 2014). Consequently, decision-makers and non-governmental organizations increasingly cooperate with volunteers for monitoring tasks. Besides reducing expenses, citizen science projects link scientific work to the broader community which may raise public awareness and the public's attitude towards the topic investigated (Chase and Levine 2018). Overdevest et al. (2004) reported that locals, who are involved in citizen science activities are more likely to protect environmental resources and participate in community services or socio-political activities.

The data collected by citizens can be used in a wide range of research scenarios. While high quality and high frequent discharge data remain complex to measure, the literature suggests that especially water level data can easily be collected with high accuracy by citizen either using physical (Weeser et al. 2018, Lowry et al. 2019) or virtual staff gauges (Seibert et al. 2019). This data can then be, for example, used in hydrological models and offer a way to assess the behavior of catchments to climate change and land use scenarios, which allows the development and evaluation of sustainable management strategies. In order to do so, these models require data like precipitation, temperature, and discharge. Nevertheless, using crowdsourced data for hydrological modeling is still in its infancy. Data collected by citizens differ from traditionally collected data in their temporal- and spatial coverage, quantity, and accuracy (Assumpção et al. 2018). Until now, only a few studies investigated how these characteristics influence the model calibration process using synthetic datasets derived from traditionally measured discharge (Mazzoleni et al. 2017, Mazzoleni et al. 2018), water levels (Seibert and Vis 2016, Weeser et al. 2019) or water levels measured by volunteers, which were converted into discharge using site-dependent stage-discharge relationships (Avellaneda et al. 2020).

This dissertation aims to further evaluate the potential of crowdsourcing approaches to contribute to hydrological research, particularly for low-income countries where experience with crowdsourced projects are limited and required hydrological data is often not available (Buytaert et al. 2014, Njue et al. 2019). First, a crowdsourced water level monitoring network with thirteen stations was designed and implemented in western Kenya to investigate with which temporal resolution and accuracy volunteers contribute data in such a setting (Weeser et al. 2018). In a second step, the data generated within the network was used to assess whether the data was suitable to run a hydrological model and how the model efficiency differs between a model calibrated on crowdsourced water level data against a model calibrated on traditional discharge measurements (Weeser et al. 2019). One key factor to ensure a successful and sustainable citizen science project that contributes to an abiding data collection are motivated and long-term engaged volunteers. Thus, understanding the socio-economic background and reasons why volunteers participate can support the design of an effective program. Hence, in a final step of this dissertation, citizen scientists in the aforementioned water level monitoring program were interviewed through a telephone survey to investigate their socio-economic background, understand their motivations and identify potential obstacles that might hinder them from turning into a long-term engaged volunteer (Weeser et al. under review).

1.2 Research Questions

The main aim of this dissertation is:

To rigorously test the feasibility to collect water level data by citizen scientists in a low-income country, evaluate the potential use of crowdsourced data for hydrological studies and modeling, and assess the background and motivations of participating citizens.

Three research questions will be addressed in separate chapters:

Chapter 2: Can the involvement of citizens in a water level monitoring be an appropriate way to overcome data scarcity in remote catchments like the Sondu-Miriu River basin in Kenya?

Chapter 3: Is water level data collected by citizen scientists suitable to calibrate a rainfall-runoff model and how do model uncertainties differ in comparison to a model calibrated with conventional data sources?

Chapter 4: What is the socio-economic background and motivation of citizen scientists in this project, and which challenges or opportunities exist for improving their engagement?

In addition to addressing the scientific knowledge gap on the use of crowdsourced data in hydrological research, the expertise gained from this study can be used to address data scarcity in remote catchments and support evidence-based decision making for sustainable water resources management and associated land use planning.

1.3 Methods & Approaches

1.3.1 Study Area

The data used for this dissertation was acquired in the Sondu-Miriu River basin (3,450 km²) located in western Kenya. Citizens measured water levels at thirteen locations distributed over the entire catchment. Additionally, data from an automatic measuring system installed next to a crowdsourced station at one headwater catchment was used (Jacobs et al. 2018b, Jacobs et al. 2020). The automatic system provided water levels with a high temporal resolution (ten minutes interval), high measurement accuracy (±2 mm) and served as a benchmark for the crowdsourced measurements and the hydrological model (see chapter 1.3.3). The location of the basin, all thirteen crowdsourced water level stations, the automatic station, a weather monitoring station, and tipping buckets used for climatic input data can be found in Figure 1.

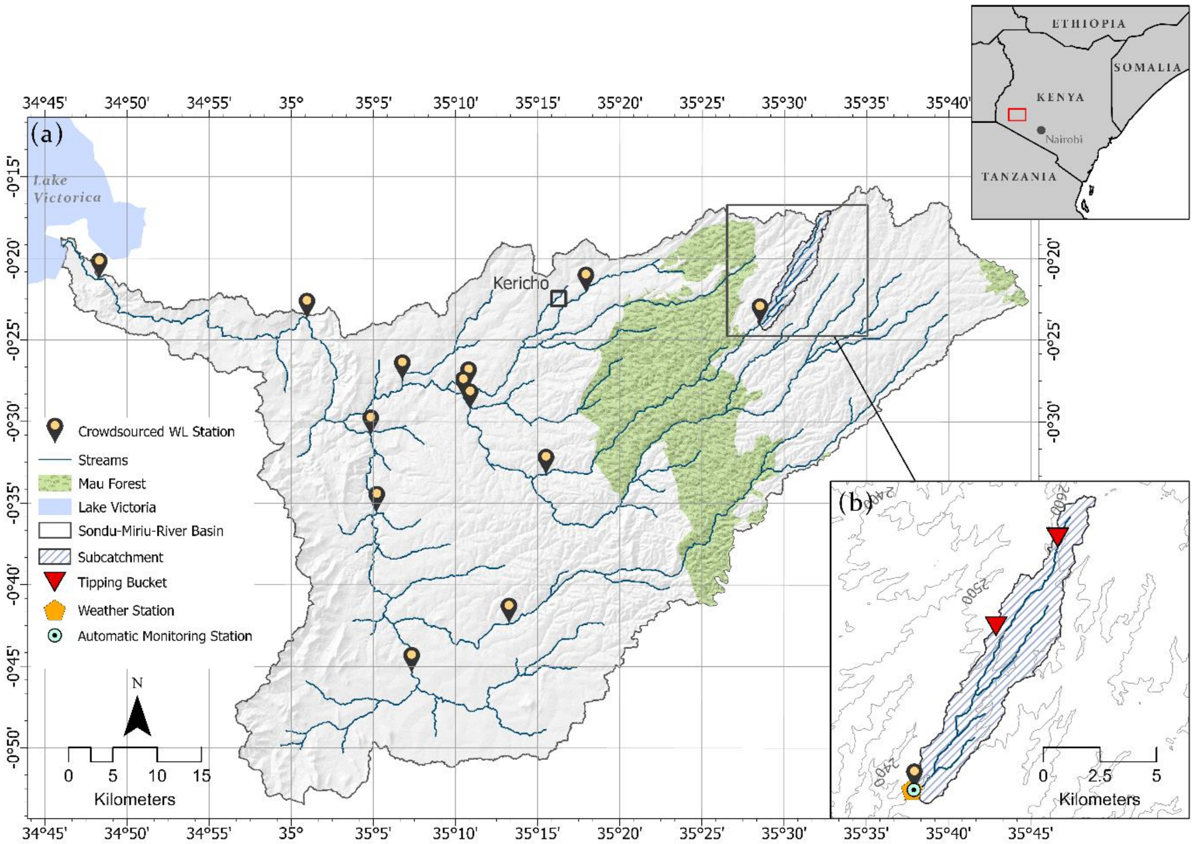

Figure 1: (a) Map of the Sondu-Miriu-River Basin showing the thirteen crowdsourced water level stations. Reference grid displays coordinates in WGS 1984. (b) Subcatchment for which the model was set-up including the position of the automatic measuring system, the weather station and the tipping buckets used to collect model input data.

The elevation in the Sondu-Miriu River basin ranges from 1,140 meters above sea level (m a.s.l.) at the outlet at Lake Victoria up to 2,900 m a.s.l. in the north-east region. Land use is dominated by smallholder agriculture and subsistence farming, with cultivation of maize, beans, cabbage, and potatoes in the eastern region. The central part of the basin is covered by the Mau Forest, Kenya's largest indigenous closed-canopy forest. Commercial tea and tree (mainly eucalyptus) plantations prevail in the northern parts around Kericho town. A mixed land use dominated by smallholder agriculture and small settlements can be found towards Lake Victoria.

The climate is influenced by the Intertropical Convergence Zone, which leads to 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 varies from 1,300 mm yr^{-1} at the lower altitudes of the study area, to 1,900 mm yr^{-1} in the north-eastern 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 in the lower parts of the basin (Vuai and Mungai 2012).

The Mau Forest Complex provides critical water-related ecosystem services such as water storage, river flow, flood mitigation, groundwater recharge, and micro-climate regulation (Benn and Bindra 2011). Poor implementations of land use policies in combination with a growing population and the need for agricultural land as well as settlement have resulted in a rapid forest degradation. More than one-quarter (100,000ha) of the native forest has been lost within the last few decades (Khamala 2010). This land use change seems to affect the hydrological cycle and lead to a decline in discharge (Olang and Kundu 2011) but comprehensive data to further investigate the land use change effects is still absent. Therefore, are clear need of data, particularly in remote and understudied locations like the Mau Forest Complex exists.

1.3.2 Crowdsourced water level collection framework

To investigate if engagement of citizens in a water level monitoring project can help to overcome data scarcity in remote catchments, we designed a crowdsourced water level monitoring network in the Sondu-Miriu River basin. Thirteen water level gauges were installed at easily accessible locations. Each gauge was equipped with a signboard (Figure 2) that explained with pictures and written instructions in English as well as Swahili how to read the water level, transmit the data, and hence, how to participate in the project. Following an approach described by Fienen and Lowry (2012), participants first read the water level and then sent a text message to a central phone number, containing their measurement and the station-ID as indicated on the signboard. The simple method allowed volunteers to participate without requiring special equipment such as a smartphone or a mobile internet connection. Text messages are a common way of communication in East Africa, which are inexpensive (~0.01 USD), easy to use, and of high availability. Besides, using a text message-based system allows providing real-time feedback to the volunteers, which enable the user to immediately detect and rectify incorrect inputs.

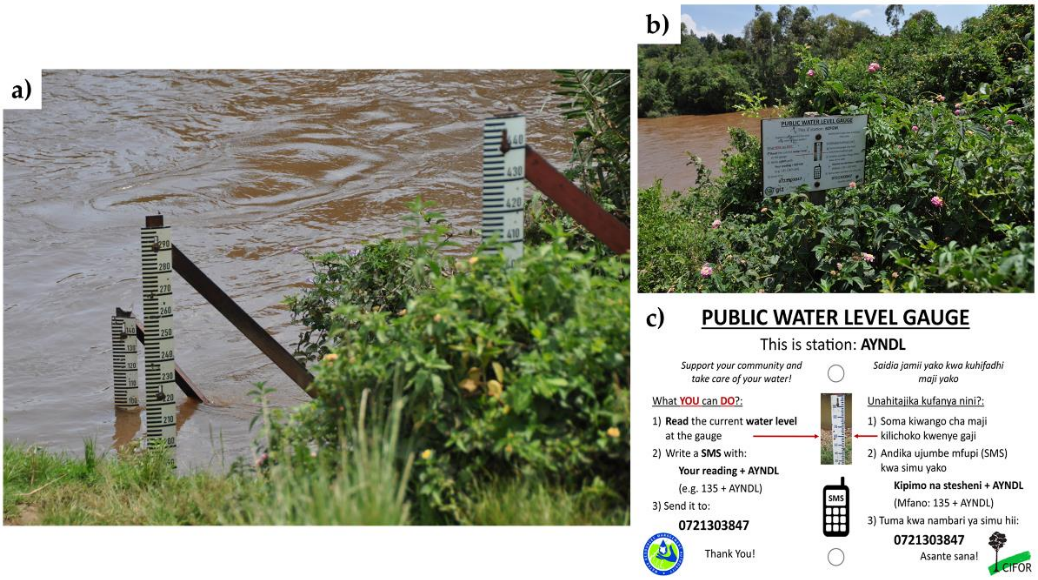

Figure 2: Example of a crowdsourced water level monitoring station (a) with a sign board (b+c) holding simple and precise instructions that make it easy for interested citizens to participate.

A SMS-server handling the incoming messages was built with a Raspberry Pi 2 Model B combined with a GSM-modem providing a local cell phone connection. A python script processed the incoming data. All data underwent a plausibility check whereby implausible data was flagged for further manual checking. The processed data was stored in a database and feedback based on the plausibility check was sent to the observer within seconds after the initial transmission. All data was accessible through a website. An interactive plot allowed interested citizens and authorities to view the water level curve at each site and to download data for further use.

Sensitization meetings with interested citizens were arranged at each site to promote the project idea, train the citizens, and assess its acceptance. During the meetings, it became evident that citizens, especially in the remote areas of the basin, might have issues sending the data due to a lack of cell phone credit. Hence, a reimbursement system for participants was tested at one station where the transmission costs (1 KES ≈ 0.01 USD) were reimbursed twofold for every valid observation sent. The amount was automatically calculated and disbursed at the end of each month using the SMS-server as described below. All other stations operated without reimbursement.

The costs for the crowdsourced monitoring network, including the hardware, were low with approximately 6,000 USD for all thirteen gauges. Additional minor costs were caused by on-site meetings with observers, the SMS-response, and the webpage.

1.3.3 Crowdsourced enhanced rainfall-runoff modeling

To assess if crowdsourced collected data can be applied to run a conceptual rainfall-runoff-model, a lumped model using the Catchment Modelling Framework (CMF) (Kraft et al. 2011) was developed. CMF operates with building blocks to construct hydrological models (Jehn et al. 2017), which allows a flexible model set-up.

The conceptual rainfall-runoff processes that were represented by the model are based on the results of Jacobs et al. (2018a). As input, daily timeseries of precipitation and potential evapotranspiration derived from temperature and extraterrestrial radiation applying the Hargreaves equation (Hargreaves and Samani 1985) were used. Within the model, precipitation was first divided by saturation excess, where water that was not able to infiltrate (q_{inf}) was directly transported to the outlet (q_{surf}) (Figure 3). Infiltrated water was stored in a single storage box that loses water either due to evapotranspiration (ET) or outflow (q_{out}). Five parameters were needed to calibrate the model. Three parameters (β, Q_0, V_0) controlled a power-law equation that determined the outflow of the storage box, the parameter $fETV_1$ modified the evapotranspiration and, the parameter $W_{1/2}$ represented the saturation at which half of the incoming water infiltrates and the other half is directed to the outlet.

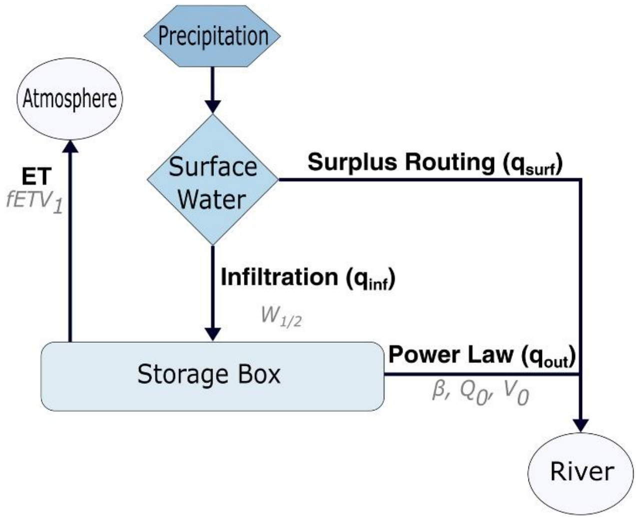

Figure 3: Schematic model structure. Processes from the Catchment Modeling Framework (Kraft et al. 2011) are given in bold (ET=Evapotranspiration) and their parameters in italic letters. Oval structures represent sinks, the hexagon an input flux, the box a storage and the rhombus a distribution node without storage functionality.

The available data was split in a warm-up period (1 January 2016 to 31 March 2016), a calibration period (April 1, 2016 to March 31, 2017), and a validation period (April 1, 2017 to March 31, 2018). The open-source python package SPOTPY (Houska et al. 2015) was applied to calibrate the model using a Monte Carlo based calibration using Latin Hyper Cube sampling (McKay et al. 1979). A total of 10^6 parameter sets were generated within predefined (a priori) parameter ranges. The calibration efficiency was evaluated with two objective functions, the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970) and percentage bias (PBIAS). While the NSE is mainly influenced by peak values and therefore ensures an acceptable model fit under high runoff conditions, the PBIAS shows the model tendency to over- or underestimate the runoff over the whole simulation period.

In addition to measured discharge, water level data reported by the citizens was used as an alternative source to calibrate the model. This approach is applicable since the water levels are dynamically linked to the discharge variation, which allows a comparison of modeled discharge against measured water levels using the Spearman rank correlation coefficient (R_{Spear}) (Seibert and Vis 2016). A benefit of using water level data over converting the water levels to discharge is that the uncertainty introduced by using a stage-discharge relation is avoided (Jian et al. 2017). Admittedly, water levels do not contain information on the total water volume, which can lead to a systematical bias. Since the R_{spear} only reflects the similarity of the dynamics between the observed discharge and water level data and not the absolute volumes a high agreement does not ensure a perfect fit in the modeled water volume (Seibert and Vis 2016). The same dynamics of the modeled discharge and the measured water levels consequently lead to a perfect fit even if the model over- or underestimates the absolute volume.

To overcome this problem, the literature suggests several approaches to reduce the risk of bias. Seibert and Vis (2016) assumed that information on the annual streamflow volume is available and used this information to filter acceptable model parameters after calibrating the model on water level data. Jian et al. (2017) proposed the integration of regionalized runoff coefficients from similar catchments to account for and to reduce the volume bias. While annual streamflow information remains difficult to obtain, especially for remote or ungauged catchments like the Sondu-Miriu river, regionalized runoff coefficients might not fit for a specific study area, even if they are obtained from similar catchments. To adress these issues, a new Water-Balance-Filter approach, which only relies on measured precipitation and actual evapotranspiration derived from MODIS (Moderate Resolution Imaging Spectroradiometer) data was developed within this thesis. The annual water balance in the catchment was calculated based on the mean actual evapotranspiration (ET_{act}) of 1,055 mm yr^{-1} provided by MODIS for the two-year simulation period and subtracting the mean observed precipitation of 1,422 mm yr^{-1} for the same period. We applied a confidence interval of +/-30% to the retrieved MODIS value to compensate for measurement errors and unknown uncertainties as well as possible storage changes within the catchment area, resulting in an ET_{act} between 738 and 1,371 mm yr^{-1}. Consequently, model runs which resulted in a simulated specific discharge of >684 mm yr^{-1} or <51 mm yr^{-1} were discarded within the Water-Balance-Filter based calibration routine.

Six independent calibration schemes were designed to assess the contribution of crowdsourced data to the model uncertainty. The model calibrated on daily discharge data, using either the NSE or the Spearman-Rank coefficient (schemes Q-NSE and Q-SR), served as a benchmark assuming that the models using these calibration schemes represent the best possible results. In a second step, the model was calibrated using crowdsourced water level measurements only (scheme CS-SR). The last three calibration schemes used all accepted runs from the first three calibration schemes (Q-NSE, Q-SR, and

CS-SR) and filtered the accepted runs further using the Water-Balance-Filter (resulting in schemes Q-NSE$_F$, Q-SR$_F$, CS-SR$_F$). Behavioral parameter sets that can give a good prediction of the discharge were selected through ranking all model runs by their associated objective function value taking the best 0.25 % of all 10^6 runs, resulting in 2,500 parameter sets.

1.3.4 Telephone survey

For obtaining information about the socio-economic background and motivation of the volunteers participating in the water level monitoring, standardized telephone interviews were conducted in the first week of July 2017. For this purpose, the telephone numbers from all messages received between the start of the project in April 2016 and the June 30, 2017 were extracted from the SMS-server, excluding numbers that were related to project staff or commercial and other non-project related purposes. A team of trained interns from the Water Resources Authority office in Kericho speaking Swahili and English, as well as one of the local languages (Luo or Kalenjin), conducted the interviews to overcome potential language barriers. Each phone number was called three times at different times and days until the respective person was reached. If none of the attempts to get in touch was successful, a text message informing about the survey was sent, inviting the person to arrange a suitable time if interested in participating in the survey. No in-kind or monetary compensation was offered for participation. The survey contained open as well as pre-coded questions to assess the motivation, possible obstacles, and socio-economic background information. The manifest message method described by Weisberg et al. (1996) was used to code the open questions.

The degree of engagement was classified according to the number of valid measurements reported to the SMS-server. Persons with 0-1 readings were classified as low engaged, persons with 2-9 or 10 or more readings as medium or high engaged, respectively. As a second classification, persons that continue sending data after the survey were classified as long-term participants.

All valid survey responses were analyzed using R studio 1.2.1335. The explanatory variables source of information about the project, frequency of passing the station, distance to station, type of phone, age class, highest completed level of education, and Water Resource Users Association membership were used to identify the driving factors why participants are low, medium or high engaged or why they were short-term or long-term engaged using random forest classification (Breiman 2001). The out-of-bag (OOB) error rate calculated by the R package 'randomForest' (Breiman et al. 2018) was used as an indicator for model accuracy after running the model 5,000 times for each measure of engagement (low/medium/high and long-term/short-term). The mean decrease in Gini index served as a measure to assess the relative importance of each explanatory variable.

1.4 Main results

1.4.1 Crowdsourced water level collection framework

Between April 1, 2016 and October 31, 2019, 3,480 valid and 304 (8.75%) invalid measurements were reported by 258 different participants for all thirteen stations. Invalid readings were mainly caused by misuse (e.g., citizens trying to apply for a job) or missing information (no station-ID or no water level). Around half of the participants (53%) submitted only one record, the most active participant reported 542 valid measurements. The majority of data was generated by participants who highly committed themselves to the project, sending several readings each month. Participants who sent more than ten valid readings during the project period contributed to 91% of the overall valid data. Only little data (9%) was generated by random passers-by sending in total less than ten valid values during the entire project period.

One station got damaged during a flood event within the first months and was excluded from further analysis. Most measurements were transmitted within the first year after the installation of the gauges, when the citizens showed high interest in the project and the functionality of the system (Figure 4). In the further course of time, the participation decreased at most stations, which can be attributed to declining interest, reduced communication between project staff and participants and in addition to more difficult conditions to read the gauges and sign-boards due to weathering processes or vandalism. The most active station KIPTO received 1,081 valid measurements reported by 31 different observers.

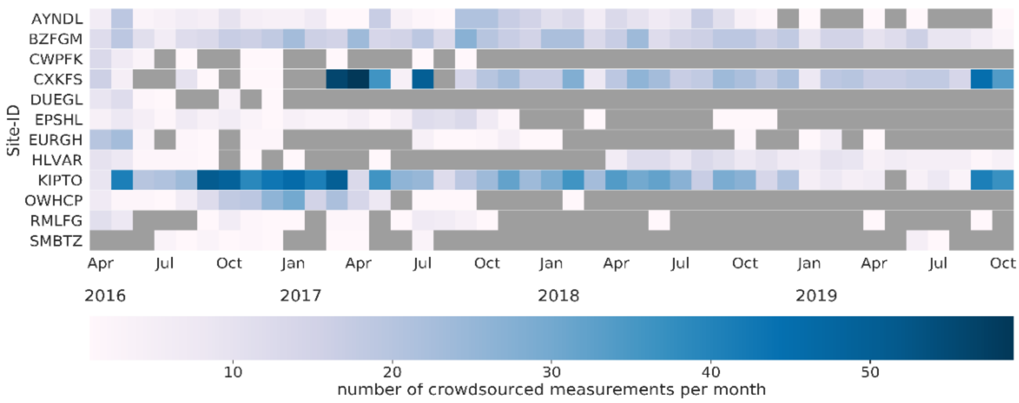

Figure 4: Monthly aggregated valid data for each crowdsourced monitoring station in the Sondu-Miriu River basin, Kenya, between April 2016 and October 31, 2019. Dark blue indicates a high activity, light blue less active months. Months without crowdsourced data are marked grey.

The comparison of automatically collected water level data recorded by a radar (VEGAPULS WL61, VEGA Grieshaber KG, Schiltach, Germany) and the crowdsourced data revealed similar trends in both data sets (Figure 5). The visible deviation between the two datasets during high- and low-flow conditions is mainly caused by a slightly different cross-section, as the radar was installed 20 m upstream from the water level gauge were citizens did the measurements. Overall, the crowdsourced monitoring framework proved to be a very cost-efficient and robust approach to monitor water levels at thirteen stations within the basin.

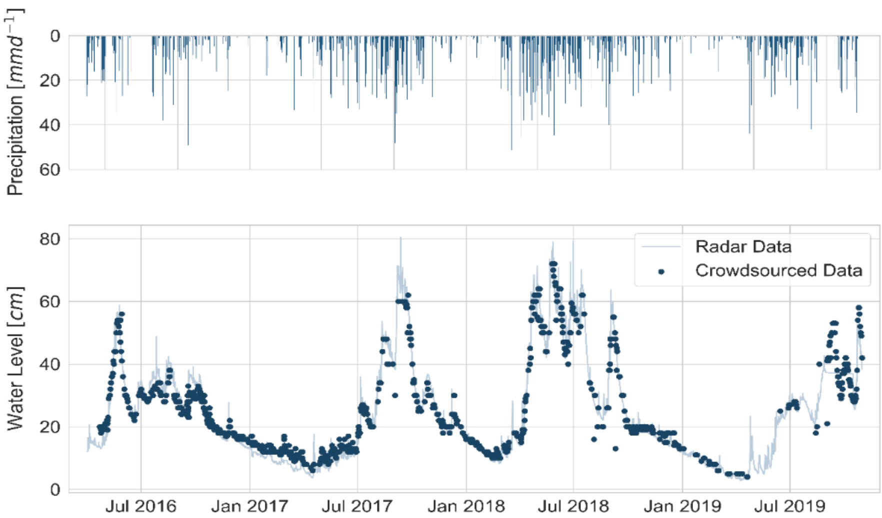

Figure 5: Time series of data collected by citizens and validation data from April 2016 to October 31, 2019. Validation data generated by a radar sensor is displayed as a light blue line, the citizen science data as blue dots. The blue bars on top of the graphic show daily rainfall data measured by an ECRN-100 tipping bucket 120 meters to the north-west of the gauge.

1.4.2 Crowdsourced enhanced rainfall-runoff modeling

A conceptual rainfall-runoff model was set-up to investigate if water level data collected by citizen scientists is suitable to calibrate such a model and how the uncertainty differ in comparison to a model calibrated with conventional discharge measurements.

Calibrated on the conventional observed discharge (Q-NSE scheme) the model simulated the discharge well and reached a mean NSE of 0.88 during calibration and a mean NSE of 0.86 during validation when considering the best 0.25% (equals 2,500) runs as behavioral. If the model was calibrated on discharge but using runs that achieved the best 0.25% of the R_{Spear} values (Q-SR scheme) the model performance decreased, achieving a mean NSE of 0.43 (0.69 during validation). When the model was calibrated and validated against the crowdsourced water level data without applying the Water-Balance-Filter (CS-SR scheme) the model still predicted the discharge within appropriate ranges. With the CS-SR scheme, the mean NSE reached comparable values than during the Q-SR calibration scheme. It is worth noting that the mean PBIAS was >0 in all R_{Spear} calibrated cases indicating that the R_{Spear}-based schemes tend to overestimate the total discharge.

The Water-Balance-Filter, which discarded model runs which violated a specific range of annual discharge based on a water balance calculated on actual evapotranspiration and precipitation, remarkably improved the model performance for all R_{Spear}-based calibration schemes. The model was able to predict the discharge almost as good as during the reference calibration scheme Q-NSE when using a calibration based on crowdsourced water level data and the Water-Balance-Filter (CS-SR_F scheme) reaching a NSE of 0.69 during calibration and 0.82 during validation. All calibration schemes tended to marginally overestimate the base flow conditions but yielded similar lower discharge bands. Only the upper discharge band deviated clearly for the CS-SR scheme compared to the CS-SR_F and Q-NSE scheme. Figure 6 represents the modeled discharge time series during calibration and validation for the Q-NSE scheme and the two crowdsourced based calibration schemes CS-SR and CS-SR_F.

The simulated processes operated within realistic boundaries. The simulated flows under the various calibration schemes did not differ substantially from each other. The variability in the fluxes of different components of the model was smallest for the Q-NSE scheme and elevated for the filtered and unfiltered schemes. However, the distribution within the unfiltered and filtered schemes were comparable, which indicated that the objective function might have a bigger influence on the model than the type of data (i.e. discharge or crowdsourced water level) used for calibration. Surface runoff was low during all calibration schemes which is in accordance with findings reported by Jacobs et al. (2018a) for the same catchment.

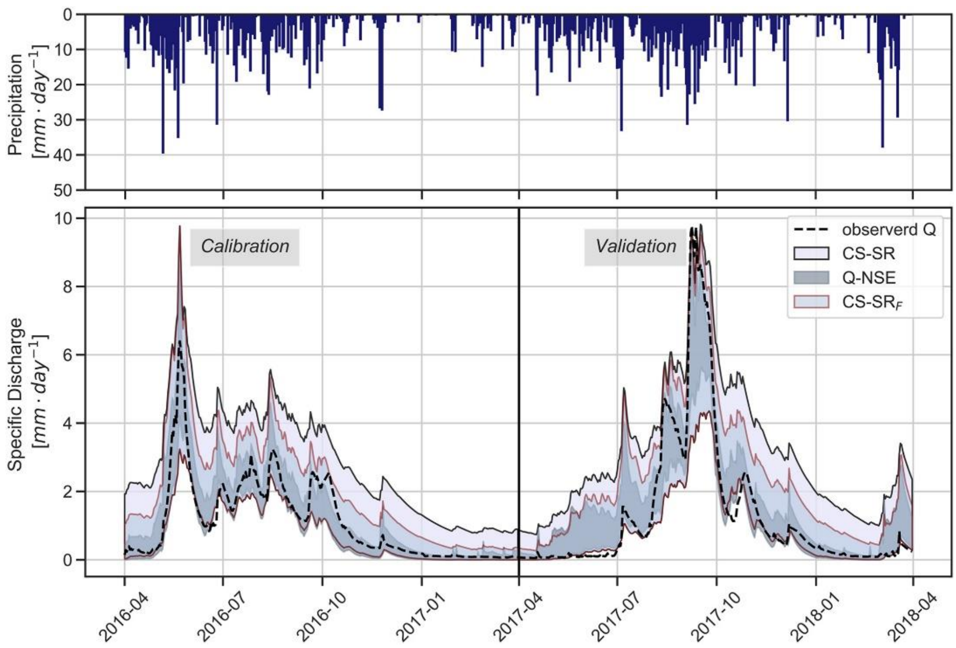

Figure 6: Observed precipitation (top) and discharge (black dashed line in the lower box) from April 2016 to March 2018. The simulated discharge is displayed for three different calibration schemes during calibration and validation. Q-NSE indicates a traditional calibration against observed discharge data. CS-SR a calibration against 2500 runs with the highest Spearman-Rank-Coefficient obtained during a calibration against crowdsourced water level data. CS-SR$_F$ a calibration using the same runs obtained from CS-SR but filtered for a maximum yearly runoff based on an estimated water balance using observed precipitation and actual evapotranspiration derived from MODIS.

To conclude, simple to obtain crowdsourced monitoring data can be combined with a modeling approach to improve the knowledge of available water resources and process understanding in otherwise understudied catchments. Hence, the approach presented here could be considered as an additional tool for water resources management, particularly in elsewise ungauged catchments and under limited financial resources.

1.4.3 Telephone survey

In total 87 persons (5 females, 78 male) participated in the telephone survey representing a response rate of 56%. Most participants were younger than 50 years (90%). The educational level was uniformly distributed among primary, secondary, and higher education. Two third (67%) were classified as low engaged participants sending zero or one valid message during the evaluation time.

The signs at the monitoring stations and sensitization meetings were mentioned most frequently as the main source of information. Highly engaged participants were mainly reached through sensitization meetings while participants with a low level of engagement became aware of the project mainly through the sign. Most respondents were aware of the purpose of the monitoring program stating that they contribute to water level observations (*n*=46). Less frequently mentioned were "monitoring for management and conservation purposes" (*n*=27) or other purposes like flood monitoring, rainfall measurements or water quality assessment (Figure 7).

Most of the respondents state that managing and conserving water as an important resource is their main reasons for participation in the project, followed by curiosity to test the system or willingness to volunteer (Figure 7).

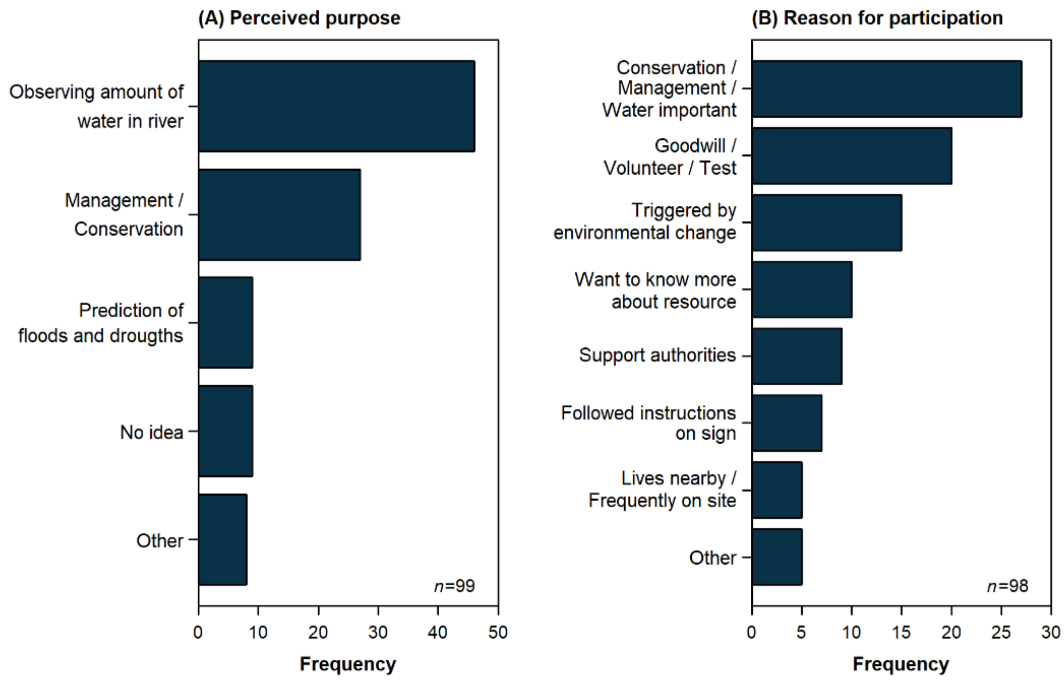

Figure 7: Respondents' answers on (a) the perceived purpose of project and (b) their reason to participate as citizen scientist. n indicates the number of total responses.

More than half of the respondents (62%) stated that they did not experienced any challenges during their participation. Respondents who mentioned challenges mostly described a lack of cellphone credit (*n*=10) or difficult access to the station (*n*=10). Asked about what could improve the engagement most respondents indicated a need for more training, education, and sensitization meetings (*n*=58).

Based on the random forest models, the highest completed level of education turned out to be the most important variable that determines the level of engagement measured by the amount of data points reported. The age class had the least influence on the engagement level. Contrary to these results, the age class was the most important variable when long-term commitment was the target variable (Figure 8).

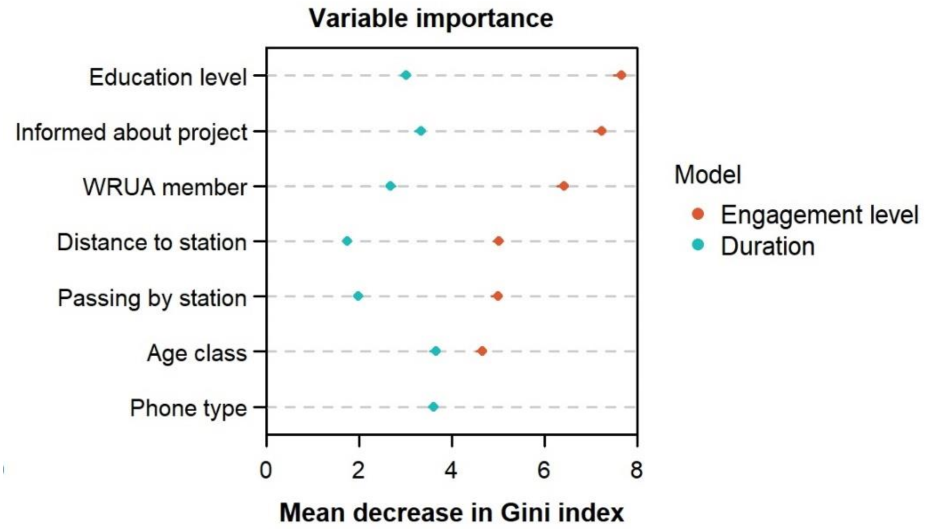

Figure 8: Relative importance of the included variables in the random forest models to predict the level and duration of engagement of the participants in the citizen science water monitoring project. The dots indicate the median value, the segments the range of values from all 10,000 runs. WRUA = Water Resource Users Association.

Summing up, identifying the target groups for citizen science projects and understanding the socio-economic background as well as motivations of volunteers is crucial to implement a successful project. Understanding these variables allows overcoming potential challenges that might hamper a long-term engagement. Particularly sensitization meetings turned out to be a powerful tool to reach out the community and increase the likelihood of participation within a setting like the remotely located study area in Kenya.

1.5 Implications & Outlook

Global annual water usage is expected to increase by more than two trillion cubic meters until 2030, leading to a global water crisis unless efforts to monitor and manage this resource are strengthened (Gilbert 2010). Particularly in low-income countries, the required data is regularly not available, even though the pressure on water resources in these countries is high (Hannah et al. 2011, Buytaert et al. 2016). The situation is further complicated by the fact that long-term monitoring networks using classical methods (e.g., automatic sampling, gauging stations, or weather stations) cause substantial costs during installation, management, and maintenance (Lowry and Fienen 2013, Buytaert et al. 2014, Mazzoleni et al. 2017). Resulting data gaps impede the assessment of temporal and spatial changes of environmental variables, which is an essential prerequisite to avoid natural disasters and for sound decision making (Davids et al. 2017). In addition, empirical evidence is required to advance our understanding of hydrological processes which is the basis to characterize catchment behavior (Royem et al. 2012). A profound process understanding is also essential for model-based future projections, which are crucial to implement mitigation measures and to meet policy needs (Tetzlaff et al. 2017).

One possible solution to improve the hydrological data pool is the implementation of citizen science based monitoring frameworks in which participants voluntarily contribute to data collection or any scientific process. The number of studies reporting citizen science approaches to measure hydrological data increased rapidly in the last decade, especially since 2014 (Njue et al. 2019). The steady increase in the number of citizen science studies in hydrology over the last decade coincides with emerging technologies like low-cost sensor equipment, better phone coverage, and a growing interest in sustainable water resource management. The rapid technological advances in sensors and the massive spread of mobile communication technologies combined with an increased computational power further support the use of alternative data collection methods or data analysis, particularly in low-income countries.

This thesis demonstrated that citizens were able to collect water level data in a remote catchment in western Kenya. The citizens consequently contributed valuable data to an otherwise understudied basin. The reported water level data was of high quality and showed a good agreement with reference measurements. Several studies that evaluated the role of citizens in reporting environmental data came to similar conclusions. Fienen and Lowry (2012), for example, reported a good match between crowdsourced water level data and data recorded by a pressure transducer in the United States of America and concluded that the observation of relatively simple parameters can be efficiently conducted by citizen scientists. Strobl et al. (2019) demonstrated that citizens can estimate stream level classes sufficiently well.

Beyond water level measurements citizens can report additional relevant hydrological parameters. Especially parameters that can be acquired without special effort or specific hardware are promising to be integrated into citizen-based data collection frameworks. Large-Scale Particle Image Velocimetry (Fujita et al. 1998) can, for example, be used to estimate discharge. As smartphone technology becomes more powerful, Large-Scale Particle Image Velocimetry can be applied on-site on commercially available smartphones (Lüthi et al. 2014). Combined with decreasing prices of smartphones, the development of smartphone-based measurement methods will provide easier access to such techniques in the future, allowing citizens to easily contribute to the data collection. Seibert et al. (2019) introduced a simple way to avoid the installation of physical water level gauges by using virtual ones generated within a smartphone application. Such techniques allow a fast and easy upscaling of monitoring programs and decrease implementation costs. The project *Soda Bottle Science* (Davids et al. 2019a) showed that citizens can improve precipitation observations by complementing ground-based and remotely-sensed precipitation in Nepal. On a nationwide scale, the *CoCoraHS* project (the Community Collaborative Rain, Hail, and Snow network) collected more than 31 million daily precipitation values by 37,500 participants in the United States (Reges et al. 2016).

The study presented in Chapter 3 demonstrated that crowdsourced water level data can be used to calibrate a conceptual rainfall-runoff model and consequently that citizen science based monitoring contribute to a better process understanding in catchments that have so far been understudied. However, water levels were the only crowdsourced data used for the modeling approach. Professionally collected high-resolution rainfall and temperature data was used as model input. Since several studies suggest that citizen can collect these types of data a rainfall-runoff processes modeling based on data collected only by citizens, eventually combined with freely available remotely sensed data, seems promising and should further be investigated. During this follow-up work, special attention should be given to how the different temporal resolutions and uncertainties in the individual measurements relate to each other and whether such a purely citizen-supported format can provide useful results.

In addition, future work could evaluate how many crowdsourced data points are needed to ensure reliable model calibration. Given that the available data pool generated by citizens in this thesis is dense, a gradual reduction of crowdsourced measurements used during model calibration could identify to which degree the model efficiencies depend on the amount of data. If this is done systematically by e.g., reducing only peak flow or base flow measurements these findings could contribute to answering the question when it might be most crucial to engage citizens to measure data. This conclusion is supported by a study by Pool et al. (2017) which revealed that only twelve strategically sampled runoff measurements can be sufficient to calibrate a runoff model. Once these strategically important sampling points are identified, the results can be used to communicate to participants of a crowdsourced monitoring network when measurements are most crucial. Such a communication strategy would, in turn, lead to more efficient monitoring and avoids an unnecessary burden on the participants.

A limitation of the modeling approach presented in Chapter 3 is that only one catchment was available to compare the value of crowdsourced data versus automatically recorded data from a fully automatic classic gauging station to calibrate the model. Hence, future work should additionally focus on testing the approach under multiple catchment conditions with different climatic and environmental settings and include, as indicated above, crowdsourced measurements not only for water level but also for input parameters like precipitation. However, the datasets required for such a comparison are currently not available. A comparative study using synthetically derived data from 671 catchments by Seibert and Vis (2016) revealed that the general approach might be transferable to various catchments. Similar behavior can be expected when using real crowdsourced data.

In practice, a successful citizen science project relies on motivated volunteers that are willing to commit their time to contribute to the goals of the project. To be able to address potential volunteers, the correct identification of target groups is essential. A proper determination of the target groups could increase the probability of success of a citizen science project (Parrish et al. 2018, Füchslin et al. 2019). Differences in the socio-economic and cultural background between geographic regions make the characterization of citizen scientists difficult. The telephone survey conducted within this thesis revealed, for example, that particularly participants between 30 and 50 years turn into long-term or highly engaged participants for water level monitoring in Kenya. These findings differ from the results described by Füchslin et al. (2019), who found that people with an age of 50 and above showed a higher willingness to participate in a Swiss citizen science project than younger people. However, the discrepancy of around 17 years in life expectancy between the two countries may also have contributed to these trends. Unlike in many developed countries, where people have more time during retirement, many people in this rural setting in Kenya are committed to lifelong full-time farming activities. Apart from targeting specific socio-economic groups, targeting members from existing groups that voluntarily work on water conservation strategies, like members from the Water Resource Users Associations in Kenya, is promising to increase the overall engagement of volunteers since crowdsourced monitoring projects could address their needs (Golumbic et al. 2020). However, structural problems like a low rate of acceptance of such associations by governmental water management authorities may impede a better integration of such associations.

The sustainability of a citizen science project is another important indicator when examining the relevance of such projects. This is linked to the question of how volunteers can be kept involved for a longer period. The participation rate in the crowdsourced monitoring network presented in this thesis decreased towards the end. Active management of citizen science projects seems to be crucial to ensure a long-term commitment and sustainability of monitoring networks. Similar hypotheses were reported by San Llorente Capdevila et al. (2020) who identified the interaction between citizens and institutions as one key factor for successful citizen science project implementations. After collecting data at 120 locations in the United States over eight years, Lowry et al. (2019) concluded that a strong citizen science network is maintained by a core group of engaged individuals where 0.1% of the total number of participants contributed almost 20% of all observations. These results correlate well with the findings within the monitoring network in Kenya (Chapter 2).

The importance of a reliable communication strategy was also highlighted by the participants of the telephone survey presented in Chapter 4. Regular feedback could show participants the impact of their contributions and help them to understand the importance of their voluntary contribution (San Llorente Capdevila et al. 2020). Even though the data collection framework used in this thesis included a simple feedback loop by sending a confirmation of reception to each participant and presenting the data on an interactive webpage, the communication method might have not addressed the participants' needs. Due to a lack of internet access, it is likely that most participants were not able to obtain further information about the project or display the data they reported. Future citizen science projects in similar settings should, therefore, explicitly incorporate communication strategies tailored to the needs of the potential citizens. Ideally, they would further investigate the impact of such strategies on the overall participation rate.

As illustrated earlier, the governmental structure, the motivations and challenges described by participants, and the target groups in the water level monitoring projects described in this thesis depend on the local conditions and could differ per region. These differences might limit a direct transferability of the methods and particular the results presented in this thesis. Consequently, the local situation should thoroughly be considered, and the study design accordingly adapted to the local circumstances before a citizen science project is implemented. Future studies are necessary to investigate the transferability of hydrological related citizen science projects between locations with different environmental, socio-cultural, or governmental conditions. The outcome of such studies could facilitate the design and implementation of future citizen science monitoring approaches worldwide. Such studies should, as mentioned above, integrate a comprehensive communication strategy adapted to local conditions to allow for cost-effective, science-based and successful crowdsourcing schemes for monitoring water resources particularly in understudied catchments.

2 Citizen science pioneers in Kenya – A crowdsourced approach for hydrological monitoring

This chapter is published in *Science of the Total Environment* as:

Weeser, B., J. Stenfert Kroese, S. R. Jacobs, N. Njue, Z. Kemboi, A. Ran, M. C. Rufino, and L. Breuer. 2018. Citizen science pioneers in Kenya – A crowdsourced approach for hydrological monitoring. *Science of The Total Environment* 631-632:1590–1599.

2.1 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. Remote sensing technologies can be a potential source to gain hydrological information, but are limited by the spatial resolution (spaceborne measurements) or temporal resolution and costs (airborne or unmanned aerial vehicle-borne measurements). Thus, the use of remote sensing techniques to obtain comprehensive datasets of water level in medium (100-1000 m) and small (< 100 m) rivers with sufficient accuracy, spatial resolution and temporal dynamics remains challenging (Bandini et al. 2017).

Citizen science projects have the potential to be a cost-effective way of gathering data and can reduce laborious or costly research problems (Tweddle et al. 2012, Gura 2013, Bonney et al. 2014, Pocock et al. 2014). 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 2018). 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. (2015) conducted a crowdsourcing campaign to determine the spatial distribution of nitrogen solutes in German surface waters. 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?

2.2 Materials and Methods

2.2.1 Study area

The study was conducted in the Sondu-Miriu River basin (3,450 km^2) located in Western Kenya
(Figure 9). Elevation ranges from 1,140 meters above sea level (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.

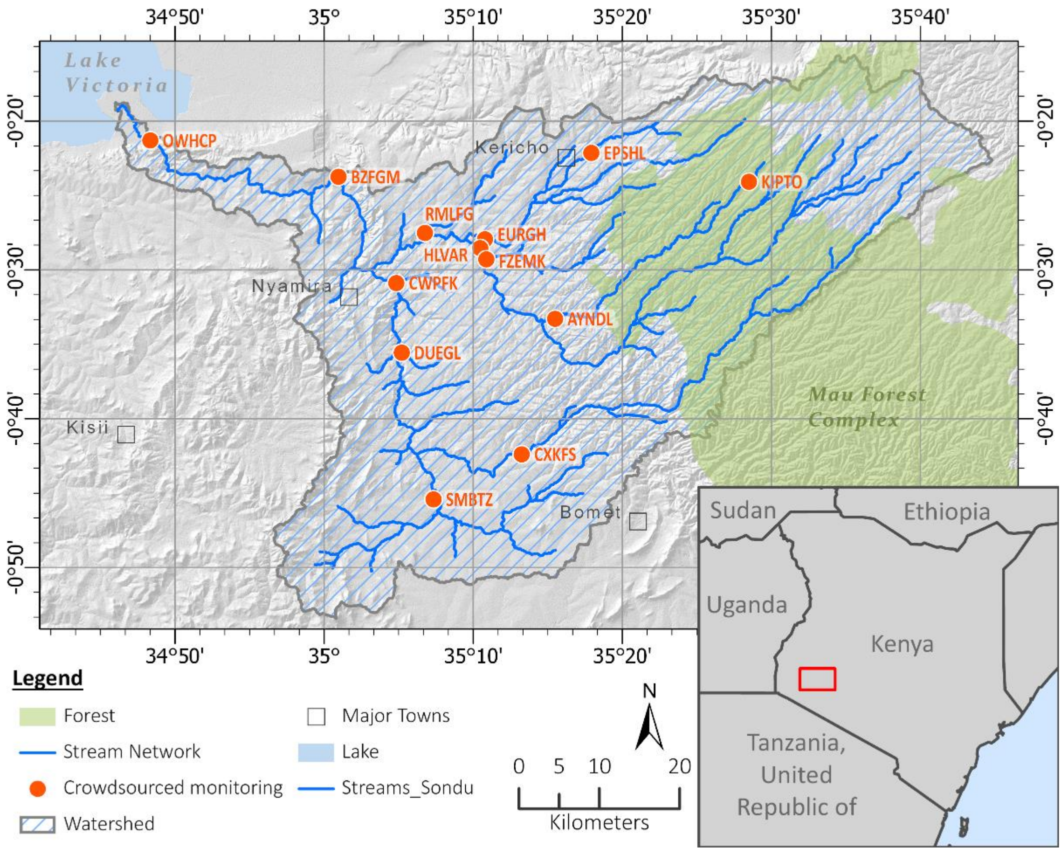

Figure 9: The Sondu-Miriu River basin in Kenya, including the stream network, major towns, natural forest areas, and the location of the crowdsourced monitoring stream gauging stations. The coordinates of the stations and additional information can be found in Table 1. Reference grid displays coordinates 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).

2.2.2 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 10) 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

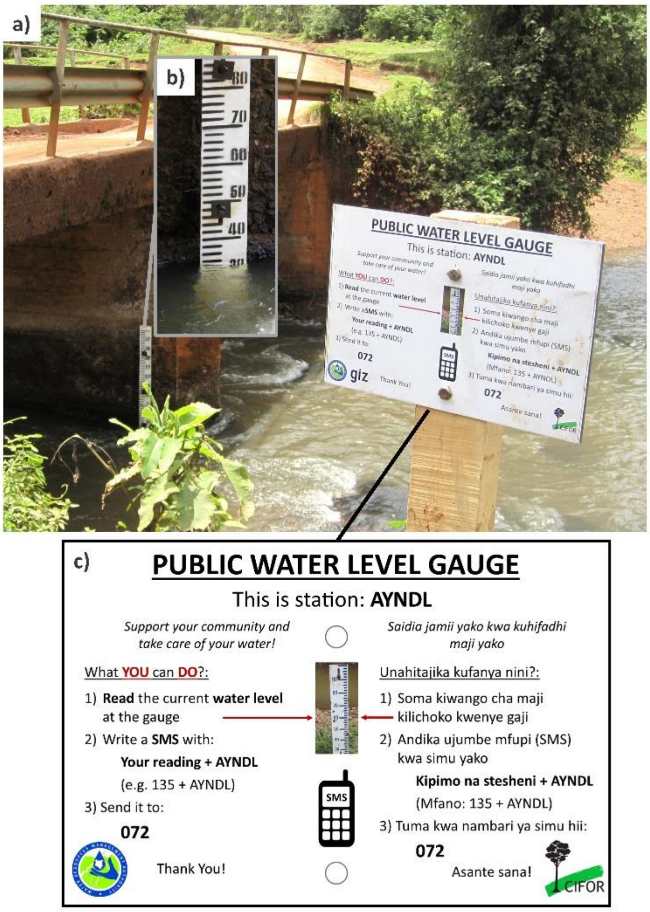

Figure 10: Example of the signboard (c) placed next to a water level gauge (b) (station AYNDL) (a). Simple and precise instructions make it easy for interested citizens to participate. Every gauge has an individual sign showing the station-ID.

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 sign-boards. 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.

2.2.3 Description of the SMS-Server

2.2.3.1 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 sent by the observers to transmit the collected data and signboards placed next to the water level gauges explained the system for interested potential participants. 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 2-1.

2.2.3.2 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 11 and Appendix 2-1. 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.

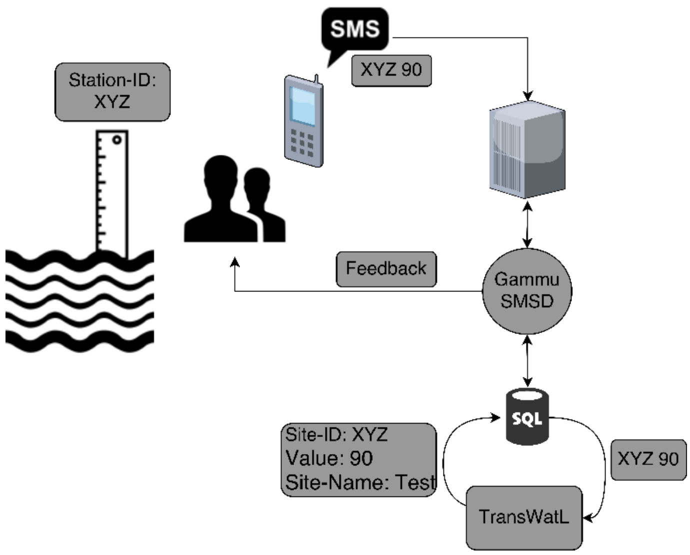

Figure 11: Schematic view of the crowdsourced data collection process. Observers read the water level and send a text message containing the value and a specific site-ID to a central server. The server stores the data received in a SQLite-database and an algorithm programmed in Python further processes the raw data and gives individual real-time feedback to observers.

2.2.2.4 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.

2.2.2.5 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.

2.3 Results

2.3.1 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 12).

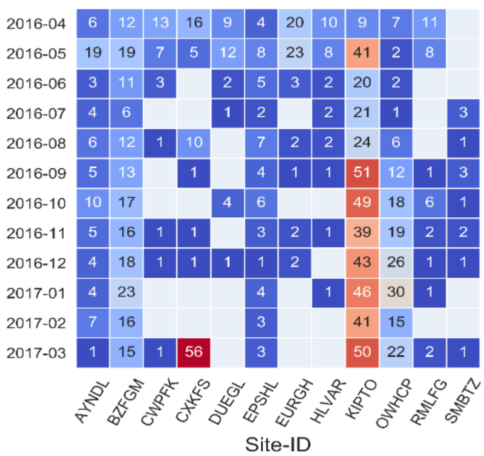

Figure 12: Monthly aggregated valid data for each station in the Sondu-Miriu River basin, Kenya, between April 2016 and March 2017. Dark blue indicates low activity, dark red very active months, and months without data received are grey.

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 13 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.

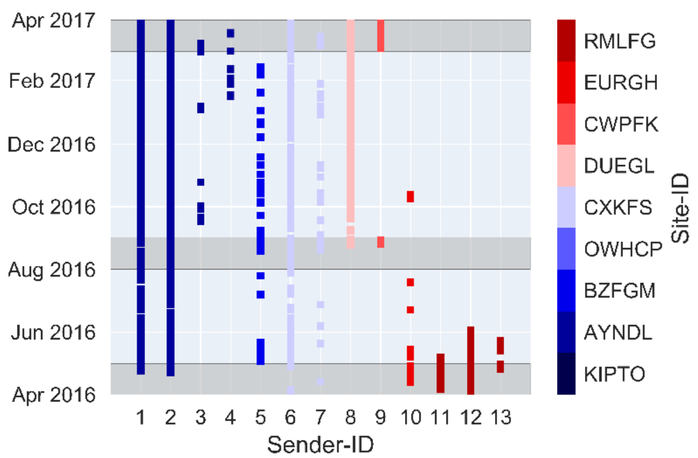

Figure 13: Temporal resolution of water level data in the Sondu-Miriu River basin in Kenya reported by active observers (more than ten observations during the observation period) in the period from April 2016 to March 2017. Every dot represents a measurement from the observer (Sender-ID). The related 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 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

2.3.2 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 14). 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.

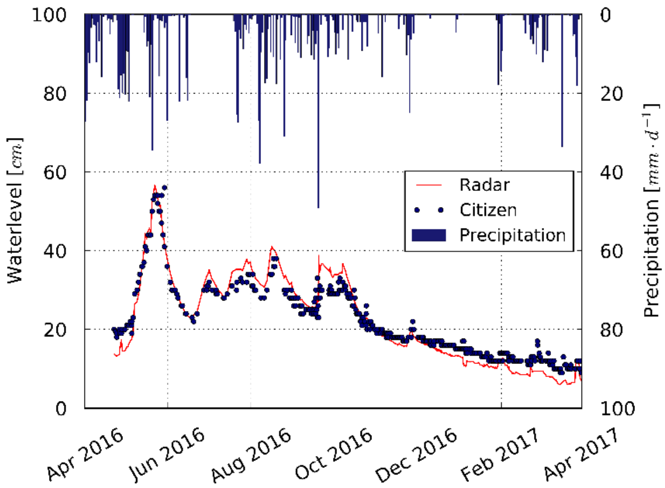

Figure 14: Time series of citizen-transmitted and validation data at the KIPTO catchment in the period from April 2016 to March 2017. Validation data generated by a VEGA radar sensor is displayed as a red line, the citizen science data is displayed usng blue dots. The blue bars show daily rainfall data 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 15). With this transformation we are now able to compare the water level changes of both stations taking into account that the riverbed between these 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.

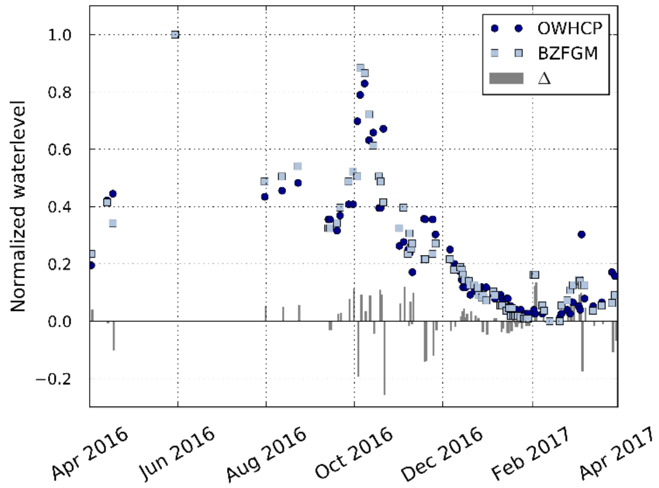

Figure 15: Standardized water level data and their differences (Δ) observed on the same day for two nearby stations (OWHCP and BZFGM) close to the outlet of the Sondu-Miriu river basin in Kenya between April 2016 and Mach 2017. The water levels transmitted for both stations follow the same trend and do not show a deviation over the time indicating reliable data reported by citizens.

2.3.3 Socioeconomic background of the participants

During the telephone survey in July 2017, 87 observers were reached and agreed to participate. Seven interviewed persons (8%) were female and 80 persons (92%) male. 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. The survey showed that the AOs in our study were in general older and of lower educational background.

Table 3: 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)
Mean Age		40	33,5
Education [%]	Primary	50	20
	Secondary	42	36
	High	8	37
	No Answer	0	7

2.4 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.

2.4.1 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 12). 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.

2.4.2 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 potential to use crowdsourced water level data to extend conventional monitoring networks. However, the integration of crowdsourced data in hydrology is still evolving, and more research is needed to unravel its full advantages and disadvantages.

2.4.3 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.

2.5 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 2-1

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.

3 Rainfall-Runoff Modelling Using Crowdsourced Water Level Data

This chapter is published in *Water Resources Research* as:

Weeser, B., S. Jacobs, P. Kraft, M. C. Rufino, and L. Breuer. 2019. Rainfall-Runoff Modeling Using Crowdsourced Water Level Data. Water Resources Research 55(12):10856–10871.

3.1 Introduction

Increasing human population and climate change increase the pressure on water resources, make society more dependent on this resource, and require practitioners to be better prepared to manage this scarce resource more efficiently (Rodda 2001, Montanari et al. 2013). Sustainable and effective water resource management decisions can only be made if reliable spatial and temporal water balance information is available. The performance of hydrological models, which can offer central support in the decision-making process, depends on sound hydro-meteorological input (Wagner et al. 2009). In contrast, the engagement and investment of environmental agencies in hydrological and meteorological monitoring effort is decreasing worldwide (Vörösmarty et al. 2001, van de Giesen et al. 2014). Particularly large tropical basins are suffering from this decrease, many remain poorly gauged or were never gauged, often due to poor accessibility (Getirana et al. 2009). In addition, data restriction policies can lead to a delay on data release (Vörösmarty et al. 2001), which makes data use for immediate water resources management difficult especially when recent information is needed (Wagner et al. 2009). This data gap prevents the investigation of temporal and spatial changes of relevant parameters for water resources management, which are critical to support decision-making and the design of for example mitigation actions to prevent natural disasters (Davids et al. 2017).

Hydrological models can be used to investigate land use or climate change impacts on basins and to predict and assess the effects of management decisions on water resources. The level of complexity and the required amount of input data vary between different models. Nevertheless, all models need input and calibration data and require a monitoring network, which can be difficult and costly to establish and maintain. In the recent past, attempts have been made to obtain necessary data using novel ways. The increasing availability of remotely sensed data provides scientists with some of the important water balance variables in regions where monitoring networks are scarce (Montanari et al. 2013). While remote sensing provides spatial data of variable resolution, hydrologists are still looking for ways to obtain direct hydrometric information such as on water levels or discharge at higher temporal and spatial resolution. So far, spaceborne remote sensing methods provide information like water level data in a sufficient resolution for large to medium-sized catchments (Yan et al. 2015), but these methods are still not operational for narrow rivers (<100-m width) (Bandini et al. 2017).

Besides remotely sensed data, crowdsource approaches have recently become attractive in research and capacity building campaigns from nongovernment institutions and agencies to fill hydrometeorological monitoring gaps (Walker et al. 2016, Davids et al. 2017). Data collected by citizens can help to create new hydrological knowledge and may support the efforts to identify the human impacts on the water cycle (Buytaert et al. 2014, Njue et al. 2019). The fast developments of communication technology will further increase the potential for citizen scientists to collect, submit, store, and process relevant data more easily (Buytaert et al. 2012, Montanari et al. 2013). In order to ensure a smooth and widespread implementation, the tasks assigned to citizens should be quick to perform and should not require special equipment. Davids et al. (2019b) showed that undergraduate researchers can conduct discharge measurements using, among others, the salt dilution streamflow measurement method within reasonable ranges when compared against professional measurements. Other studies revealed that collecting simple parameters such as water levels is straightforward and that citizens can perform this task successfully (Fienen and Lowry 2012, Weeser et al. 2018) and over long periods (Lowry et al. 2019).

Recent studies proved that water level instead of discharge data can be used for model calibration by using the monotonic relationship between water level and discharge mapped by the Spearman-Rank-Coefficient (Seibert and Vis 2016, Jian et al. 2017, van Meerveld et al. 2017). This step avoids the need to convert water levels to discharge and potentially reduces the uncertainty introduced by this conversion (Jian et al. 2017). However, this step can also lead to a systematic bias since no information on the total water volume is taken into account. Therefore, there have been different attempts to include additional data during model calibration, for example, by filtering acceptable model parameters using annual streamflow volume (Seibert and Vis 2016) or regionalized runoff coefficients from similar catchments (Jian et al. 2017). In this study, we tested a simple Water-Balance-Filter, which does not rely on any previous hydrometric information other than measured precipitation and actual evapotranspiration derived from MODIS (Moderate Resolution Imaging Spectroradiometer) data.

Using crowdsourced data for hydrological modeling is still in its infancy, and the value of this data source has not been comprehensively tested yet. Data collected by citizens differ from traditionally collected data in being irregular and of unknown quality and uncertainty. A few studies investigated the impact of these before-mentioned characteristics on the model calibration process using synthetic data sets derived from traditionally measured discharge (Mazzoleni et al. 2017, Mazzoleni et al. 2018), water levels (Seibert and Vis 2016), or discharge combined with an error term generated from discharge estimates by citizens (Etter et al. 2018). However, none of these studies used real crowdsourced data.

Besides the potential use of crowdsourced data, involving the community brings additional benefits. Locals, who are supporting citizen science projects are more likely to protect environmental resources and participate in community services or sociopolitical activities (Overdevest et al. 2004). Linking this to the fact that especially low-income countries face pressing challenges in the water sector, it is attractive to test the integration of crowdsourced data for water resources management. To address this need, we established a comprehensive monitoring network based on crowdsourcing in the Sondu-Miriu River basin in Kenya in 2016 (Weeser et al. 2018). To date, the implementation of this approach has yielded more than 5,000 records of water levels.

This study aimed at rigorously testing the potential use of crowdsourced data for hydrological modeling, which could support the assessment of management practices in tropical environments. It was designed to answer the question of (1) whether water level data collected by citizen scientists are suitable for calibrating a rainfall-runoff model with an uncertainty similar to the uncertainty resulting from a calibration with conventional data sources and (2) if the model uncertainties can be reduced by using a simple to obtain Water-Balance-Filter as an additional criterion.

3.2 Materials and Methods

3.2.1 Study Area

The study was conducted in a headwater catchment (27.4 km²) in the north-western part of the Sondu-Miriu-River Basin in Western Kenya (0.35°S, 35.5°E WGS1984) (Figure 16). Smallholder agriculture dominates the land use including annual crops, grazing lands, woodlands, and forests and led to a degraded land cover (Olang and Kundu 2011). Increasing human population has resulted in rapid forest cover loss and forest degradation in the last decades (Brandt et al. 2018) and physical evidence has revealed a noticeable discharge decline for major rivers in the region (Olang and Kundu 2011). The soils are in general deep and well-drained classified as Humic Nitisols and Mollic Andosols (ISRIC - World Soil Information 2007)

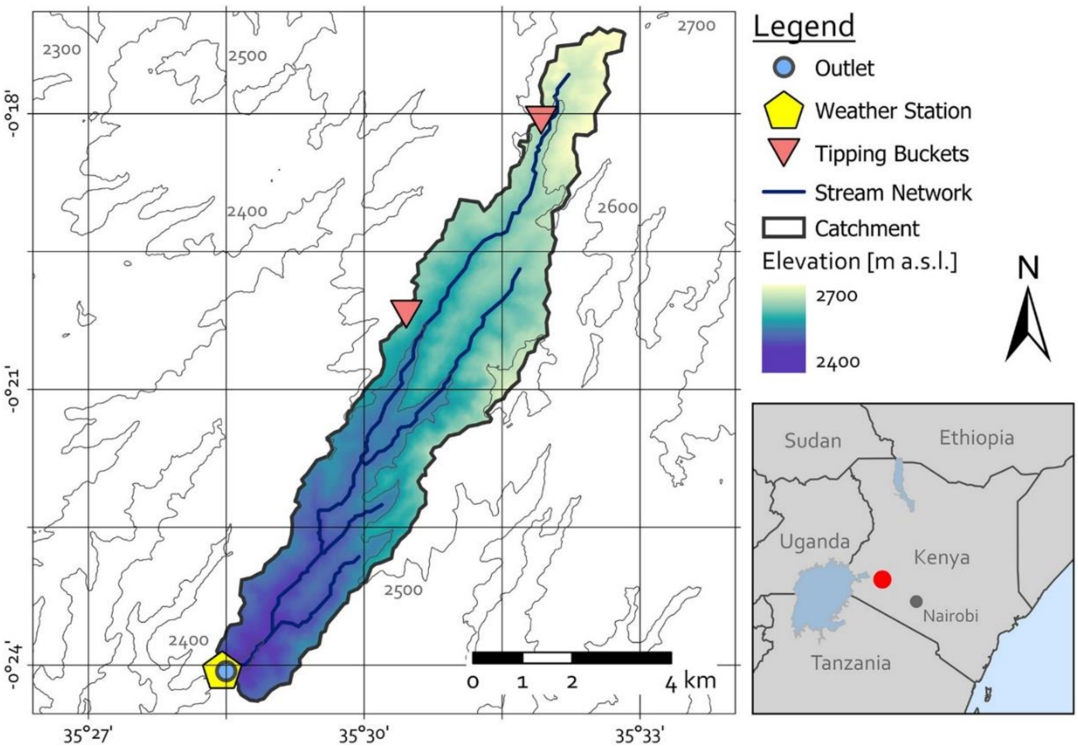

Figure 16: Location of the Sondu-Miriu-River Basin in Western Kenya (red dot in the overview map) and a map of the study area, including the stream network, outlet, weather station, and tipping buckets. The reference grid displays coordinates in WGS 1984.

The climate is influenced by the Intertropical Convergence Zone, resulting in a bimodal rainfall pattern with a longer rainy season from April to July and a shorter rainy season between October and December (Figure 17). Temperature and precipitation (Table 4) were recorded by a weather station (ECRN-100 high-resolution rain gauge and VP-3 sensor, Decagon Devices, Pullman WA, USA) located 100 m northwest of the outlet measuring at a 10-min resolution. The installation of the instruments was carried out as far as possible according to the WMO guidelines, whereby local conditions had to be taken into account. The resolution of the rain gauge is 0.2 mm per tip, the accuracy of the VP-3 sensor depends on the temperature and humidity but lies in most cases within ~0.25°C and 2-5% humidity. Precipitation was measured at two additional sites located in the center and the upper part of the catchment using tipping buckets (Theodor Friedrichs, Schenefeld, Germany). Thiessen-Polygons were used to calculate the area-weighted precipitation. If precipitation data gaps existed, the weights were adjusted by omitting the tipping bucket where no data was available. Data gaps in the temperature and precipitation time series were scarce (precipitation: tipping buckets <0.1%, precipitation from the weather station 5.5%; temperature: 7.2%) and filled with a linear interpolation after the data were aggregated to daily time steps. The yearly potential evapotranspiration (ET_{pot}) using grass as a reference crop was calculated based on the daily minimum, maximum and mean temperatures and the extraterrestrial radiation using the Hargreaves equation (Hargreaves and Samani 1985). These values were in line with long term, altitude depending, ET_{pot} of 1,400 to 1,800 mm reported for this area (Nyenzi et al. 1981, Krhoda 1988).

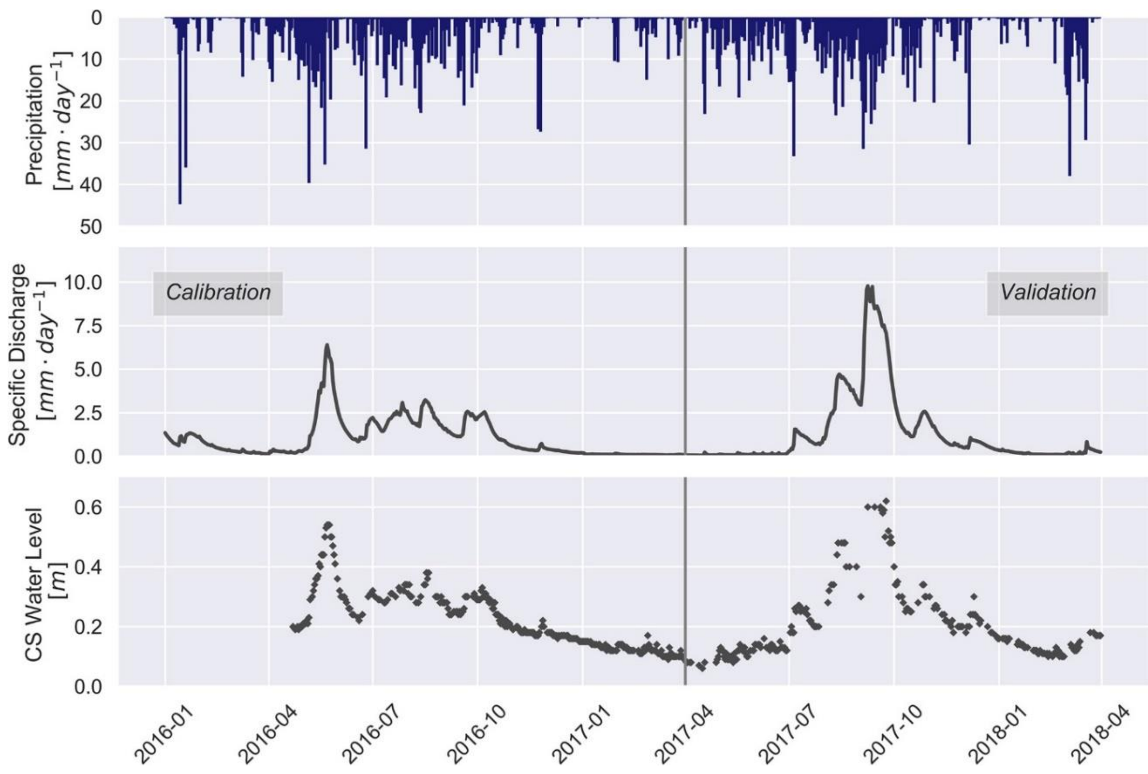

Figure 17: Daily mean areal weighted precipitation using Thiessen-Polygons (upper panel), specific discharge (middle panel) and crowdsourced (CS) reported water level data (bottom panel) of the catchment for January 2016 - April 2018.

Table 4: Averaged annual hydrometeorological data for the study area.

Period	Specific Discharge	Precipitation	Mean daily temperature	ET_{pot}
	mm	mm	°C	mm
01. Apr 2016 - 31. Mar 2017	413	1,287	14.9	1,596
01. Apr 2017 - 31. Mar 2018	485	1,557	14.4	1,522

Note. ET_{pot} = potential evapotranspiration calculated using the Hargreaves equation. Temperature represents the mean daily temperature measured at the weather station at the catchment outlet

Water levels were measured in two different ways. A radar-based sensor (VEGAPULS WL61, VEGA Grieshaber KG, Schiltach, Germany) automatically collected water level data 20 m upstream of the outlet at 10-min intervals. Data collected by citizens (n=271 during the calibration period) were recorded at the outlet (Weeser et al. 2018). The crowdsourcing-monitoring station was installed in April 2016 and equipped with a sign-board that explains to locals how they can participate in the monitoring. A small reward of 0.02 USD per measurement is paid automatically to compensate for the transmission costs. The coverage in the observation period was high with typically more than 16 observations per month covering 75% of all days during the calibration period. The data were not further filtered. Only one obvious outlier caused by a misinterpretation of the received text message was removed after checking the original text message associated with the doubtful data point. A comparison between the crowdsourced data and the automatically measured water levels showed a high agreement between both data types resulting in a Pearson correlation coefficient of 0.98 for the 271 measurements during the calibration period (Figure 18a). The high value of the correlation coefficient indicates that the crowdsourced data only differs slightly from professional measurements. Note that intercept and slope deviate from 0 and 1, respectively, due to the fact that readings from citizens were conducted 20 m upstream from where the professional reading took place resulting in slightly different cross sections.

A rating curve and the catchment area were used to convert the automatically measured water levels into daily specific discharge, which was the basis for model testing and evaluation (Figure 18b). To develop the rating curve (Eq 1), 86 manual discharge measurements using the salt dilution method (n=82) and an Acoustic Doppler Current Profiler (RiverSurveyor S5, SonTek, San Diego CA, USA) (n=4) over a wide range of water levels (h) were conducted. Extrapolation below the water level of 0.236 m was done using a quadratic function through the lowest measured discharge and zero discharge (Jacobs et al. 2018b). For water levels above the highest measured water level used to develop the rating curve (0.66 m), we extrapolated the discharge using the same rating curve (3.3% of the time). To assess the discharge uncertainty we followed the procedure described in Jacobs et al. (2018b), where the uncertainty was estimated based on the standard deviation (SD) of repeated measurements (SD water level: 1 mm, SD ADCP: 6.2%, SD Salt Dilution: 6.9%). We assumed that the true values were within 3*SD and generated 10,000 random samples for each water level/discharge combination. Figure 18b shows the uncertainty 95% confidence interval for the rating curve.

$$Eq\ 1 \qquad Q = \begin{cases} 0.0973 - 1.892 \cdot h + 6.923 \cdot h^2, & h \geq 0.236\,m \\ 0.651 \cdot h^2, & h < 0.236\,m \end{cases} \qquad R^2 = 0.98$$

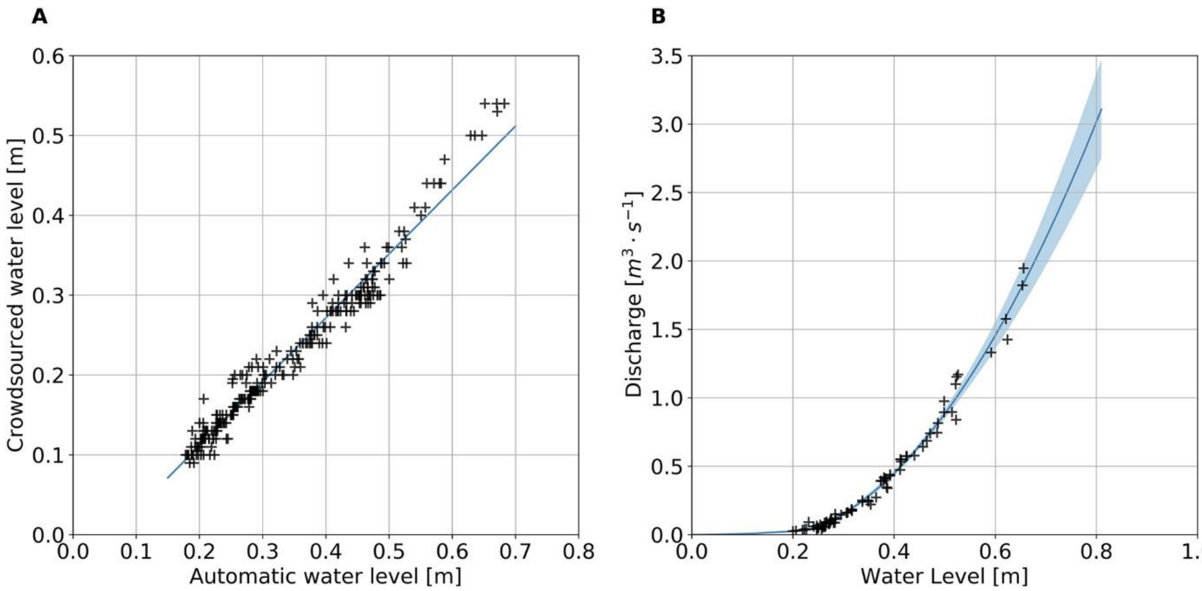

Figure 18: (a) Correlation between automatic water level measurements and crowdsourced water level data (Pearson correlation r=0.98, n=271) for the calibration period and (b) Rating curve (solid blue line) with 95% confidence interval (blue shaded band) for the outlet of the study area based on 86 water level discharge pairs (black crosses).

3.2.2 Model Setup

We developed a lumped conceptual rainfall-runoff-model (Figure 19) using the Catchment Modelling Framework (Kraft et al. 2011) Version 1.4.1 (Kraft et al. 2018) for Python 3. Catchment Modelling Framework is a python based programing library to build hydrological models from building blocks (Jehn et al. 2017). This framework has been applied to a variety of catchments and is considered to describe the underlying hydrological processes sufficiently well (Windhorst et al. 2014, Jehn et al. 2017, Maier et al. 2017). The model structure represents the conceptual understanding of the rainfall-runoff processes reported by Jacobs et al. (2018a). Daily precipitation and ET_{pot} are the only model inputs required.

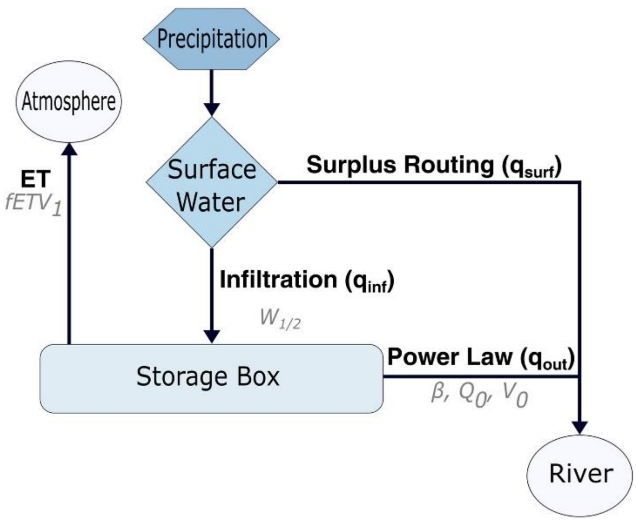

Figure 19: Schematic model structure. CMF processes are given in bold and their parameters in italic letters. Oval structures represent sinks, the hexagon an input flux, the box a storage and the rhombus a distribution node without storage functionality.

In the model, water storage is represented as a single storage volume (V), which receives water from infiltration and loses water to the catchment outlet and actual evapotranspiration (Eq 2). Precipitation (P) is partitioned into infiltration (q_{inf}) and direct runoff (q_{surf}) by saturation excess (e_{sat}) (Eq 3). Since the soil saturation happens only in parts of the catchment area, depending on the stored volume, the saturated area is modeled with the Boltzmann sigmoidal function based on the soil water storage and a parameter $W_{1/2}$. The parameter $W_{1/2}$ represents the saturation at which half of the incoming water infiltrates and the other half is routed to the outlet without timelag. Water in the storage box is released to the outlet using the normalized water volume raised to a power. This power-law equation (Eq 4) determines the outflow (q_{out}) based on the actual water volume stored in the box (V) and the parameters V_0 (reference volume), Q_0 (outflow from the source when V equals V_0), and β (shape of the response curve). Water in the storage box is subject to evapotranspiration where the daily ET_{pot} is limited by the available water in the storage box. Actual evapotranspiration is assumed to be equal to the potential evapotranspiration if the water volume in the box is greater than the factor $fETV_1$

multiplied by the box capacity. If the volume is lower, the actual evapotranspiration is linearly scaled down to 0.

We used the implicit, error controlled CVode solver (Hindmarsh et al. 2005) to integrate the differential equation (Eq 2) to prevent numerical problems (Kavetski and Clark 2011). *A priori* model parameter ranges (Table 5) were chosen based on expert knowledge from previous applications of comparable model types and an exploratory analysis of a large parameter space.

Eq 2 $$\frac{dV}{dt} = q_{inf}(P,V) - q_{out}(V) - fET(V) \cdot ET_{pot}(t)$$

Eq 3 $$q_{surf}(V,P) = e_{sat}\left(V,W_{1/2}\right)P$$
$$q_{inf} = P - q_{surf} = \left(1 - e_{sat}\left(V,W_{1/2}\right)\right)P$$

Eq 4 $$q_{out} = Q_0\left(\frac{V}{V_0}\right)^{\beta}$$

3.2.3 Model Calibration and Validation

The time series were split-up in a warm-up period (1 January 2016 to 31 March 2016), a calibration period (1 April 2016 to 31 March 2017) and a validation period (1 April 2017 to 31 March 2018). We followed a Monte Carlo based calibration approach and quantified the model parameter uncertainty using the open-source python package SPOTPY (Houska et al. 2015). We evaluated the calibration efficiency using two objective functions, that is, Nash-Sutcliffe-Efficiency (NSE, Eq 5 where e_i is the i-th observation, s_i is the i-th simulation and $\bar{e}$ is the mean of the observations) (Nash and Sutcliffe 1970) and percent bias (PBIAS, Eq 6).

Eq 5 $$NSE = 1 - \frac{\sum_{i=1}^{N}(e_i - s_i)^2}{\sum_{i=1}^{N}(e_i - \bar{e})^2}$$

Eq 6 $$PBIAS = 100 * \frac{\sum_{i=1}^{N}(e_i - s_i)}{\sum_{i=1}^{N}(e_i)}$$

While the NSE is mainly influenced by peaks and therefore ensures an acceptable model fit during high flow conditions, the PBIAS indicates the tendency of overestimation or underestimation of the discharge through the model over the full period. In total, 10^6 parameter sets were generated for the calibration process within predefined (*a priori*) parameter ranges (Table 5). Instead of sampling the entire parameter space, we used Latin Hypercube Sampling (McKay et al. 1979).

Table 5: Model parameters and a priori ranges.

Name	**Meaning**	**Unit**	***A priori* parameter ranges**	
			Min	Max
β	Kinematic flow curve shape exponent	[-]	1	6
Q_0	Reference runoff when storage contains the reference volume V_0	[mm day^{-1}]	0.01	1,000
V_o	Reference volume where storage runoff is equal to Q_0	[mm]	100	3,000
$fETV_1$	Scaling factor for potential evapotranspiration	[-]	0.01	0.8
$W_{1/2}$	Saturation, where half of the catchment area is saturated	[-]	0.1	0.9

3.2.3.1 Using the Spearman-Rank-Coefficient to Calibrate on Water Level Data

To calibrate the model on water level data we took advantage of the fact that water levels are dynamically linked to discharge variation and that they can, therefore, be compared against modeled discharge by using the Spearman rank correlation coefficient (R_{Spear}) (Seibert and Vis 2016). Ranging from -1 to 1, an R_{Spear} close to 1 indicates that the simulated discharge and the measured water levels reproduce the same dynamics and that the water level and discharge values are strictly monotonically related (Seibert and Vis 2016). The R_{Spear} is not affected if the data is transformed using a strictly monotonically increasing or decreasing function as done by the rating curve in this study. In this case, the R_{Spear} values will be similar regardless if the automatically measured water level data (which was converted into discharge data using the rating curve) or the discharge data itself is used. Consequently, we do not show a calibration based on the automatically measured water level data since the results are the same as obtained from a discharge-based calibration.

Since the R_{spear} only reflects the similarity of the dynamics between the observed discharge and water level data and does not reflect the absolute volumes, a value of 1 does not ensure a perfect fit (Seibert and Vis 2016). Therefore, a threshold for behavioral parameter sets cannot be defined similarly to a calibration based on objective functions like the NSE. Instead, we propose to select behavioral parameter sets by ranking all model runs by their associated R_{Spear} value and take the top set. In this study, we defined behavioral parameter sets by taking the best 0.25% of the 10^6 runs, resulting in 2,500 parameter sets. The same procedure of taking the best 0.25% was applied when the model was calibrated on discharge data and the NSE to ensure the comparability of the different calibration schemes.

3.2.3.2 Water-Balance-Filter

As stated above, utilizing water level readings for calibrating a model to calculate discharge can lead to overestimation or underestimation. Therefore, we tested a simple annual Water-Balance-Filter to obtain acceptable model outputs by selecting only those parameters sets where model runs resulted in a high R_{Spear} and additionally matched a simplified water balance. The annual water balance was calculated from observed precipitation minus mean actual evapotranspiration (ET_{act}). ET_{act} can be retrieved from spaceborne remote sensing data sets obtained from the MODIS. For the study area, a mean ET_{act} of 1,055 mm yr^{-1} was derived from data provided by the MOD16A2 Collection 6 Global Evapotranspiration Product from MODIS imagery based on land surface temperature and albedo and the Penman-Monteith equation (Running et al. 2017) for the 2-yr simulation period. These values are close to the estimation from our measured data when we subtract runoff from precipitation assuming that possible storage changes can be considered small enough to be ignored for the 2-yr period and the remaining water consequently represents the ET_{act} (Senay et al. 2011). From our measured data we derived an ET_{act} of 973 mm on average for the 2 yr, which is 7.7% less than the value determined using the MODIS dataset. The MOD16A2 Collection 6 dataset, obtained from the satellite Terra, contains composite evapotranspiration data with 500-m pixel resolution for 8-day periods. In order to calculate the annual ET_{act} each satellite image was cropped to the catchment area, fill values without calculated ET were set to unavailable, the result was multiplied by 0.1 (scale factor after Running et al. (2017)) and a mean value for the catchment area was calculated. In order to determine the annual value, all individual values were summed up. To compensate measurement errors, unknown uncertainties and possible storage changes we added a (subjective) confidence interval of +/-30%, resulting in an ET_{act} between 738 and 1,371 mm yr^{-1} (ET_{act}/ET_{pot}-ratio = 48-88%) for the study area. This value is in line with a study of Velpuri et al. (2013), which reported mean uncertainties up to 25% for MOD16 datasets at basin scale. Given the average annual precipitation over the 2-yr observation period of 1,422 mm (Table 4), model runs were discarded if the simulated specific discharge was >684 or <51 mm yr^{-1}.

3.2.3.3 Calibration Schemes

Six independent calibration schemes were carried out to evaluate the value of crowdsourced water level data for model calibration. As a benchmark, we first calibrated the model on the discharge data using both the Nash-Sutcliffe-Efficiency and the Spearman-Rank coefficient (schemes *Q-NSE* and *Q-SR*). After that, we calibrated the model on the crowdsourced water level measurements and did not consider any automatically measured water level or discharge data (*CS-SR*). Finally, all accepted parameters from the different calibration schemes were filtered using the Water-Balance-Filter ($Q\text{-}NSE_F$, $Q\text{-}SR_F$, $CS\text{-}SR_F$).

3.2.4 Model Comparison (Benchmark)

The model was validated by conducting runs for the validation period using the *a posteriori* parameter sets, comparing the modeled with observed discharge. To compare the model efficiencies between the different calibration schemes, we defined a lower benchmark (R_{lower}) following an approach described by Seibert et al. (2018). For this, we run the model 2,500 times with random parameter sets within the *a priori* parameter ranges (Table 5). From these 2,500 model runs, a mean discharge time series was calculated and compared against the observed discharge for both, the calibration and validation period. The upper benchmark (R_{upper}) was defined as the best efficiency obtained during the discharge-based calibration assuming that this value reflects the best possible calibration of the model for the given data set. The relative performance ($R_{Relative}$) of each calibration scheme can then be determined following Eq 7, whereby R_x indicates the performance reached for each individual calibration scheme.

$$\text{Eq 7} \qquad R_{Relative} = \frac{R_x - R_{lower}}{R_{upper} - R_{lower}}$$

3.3 Results

3.3.1 Lower and Upper Benchmark

To compare the efficiencies of the different model calibration scenarios a lower benchmark was defined by randomly selecting 2,500 model runs and calculating a mean discharge time series, which was compared against the observed discharge values. A NSE of -0.56 and a PBIAS of 97.51% was found as a lower benchmark for the calibration period. For the validation using the same 2,500 random parameter sets, the NSE was 0.13 and the PBIAS 111.31%. The highest performance measure within the *Q-NSE* scheme defined the upper benchmark, resulting in an upper NSE benchmark of 0.93 and a PBIAS of 0%. All schemes resulted in at least one parameter set with similar best performance measures for all schemes.

3.3.2 Discharge-Based Calibration (Q-NSE and Q-SR)

The model simulated observed discharge reasonably well when calibrated against discharge using the Q-NSE scheme (Table 6). Under this scheme the model achieved a mean NSE of 0.88 and a relative NSE performance for the mean of all runs of 96.6% (relative performance of PBIAS 99.1%) when compared against the upper and lower benchmark of the *Q-NSE* scheme. The parameter sets which achieved the best 0.25% (equals 2500) NSE values (*Q-NSE*) or R_{Spear} values (*Q-SR*) were considered as behavioral and were accepted for further analysis. When testing the behavioral parameter sets of *Q-NSE* against the validation time series the model performance was only marginally lower achieving a mean NSE of 0.86 and a relative NSE performance 91.3% (relative performance of PBIAS 93.8%).

Calibrated on discharge but using 0.25% of all parameter sets with the highest R_{Spear} instead of NSE (*Q-SR*) the model performance decreased achieving a mean NSE of 0.43 and a relative performance of 66.4%. The mean PBIAS increased from -0.88% (*Q-NSE*) to 52% during calibration. The same trend was followed during validation with similar performance measures.

Table 6: Relative performance and model efficiency measures Nash-Sutcliffe-Efficiency (NSE) and percent bias (PBIAS) during calibration and validation of the different calibration schemes using discharge observations (Q) and the crowdsourced data (CS) without and with a Water-Balance-Filter (Filter) for the best 0.25% of all 10^6 model runs calibrated on the NSE or R_{Spear}. Heat map indicates best (green) to worst (yellow) model performance.

					NSE						**PBIAS**					
					Calibration			**Validation**			**Calibration**			**Validation**		
Dataset	Calibrated with	Filter	ID	n-Runs	mean [-]	best [-]	$R_{relative}$ [%]	mean [-]	best [-]	$R_{relative}$ [%]	mean [-]	range [%]	$R_{relative}$ [%]	mean [-]	range [%]	$R_{relative}$ [%]
Q	NSE	No	Q-NSE	2500	0.88	0.91	96.6	0.86	0.93	91.3	-0.88	[-23,16]	99.1	6.88	[-17,29]	93.8
		Yes	Q-NSE_F	2500	0.88	0.91	96.6	0.86	0.93	91.3	-0.88	[-23,16]	99.1	6.88	[-17,29]	93.8
	R_{Spear}	No	Q-SR	2500	0.43	0.91	66.4	0.69	0.93	70.0	51.95	[-36,133]	46.7	51.8	[-28,76]	53.4
		Yes	Q-SR_F	1539	0.70	0.91	84.6	0.80	0.93	83.8	28.48	[-36,65]	70.8	30.4	[-28,76]	72.7
CS	R_{Spear}	No	CS-SR	2500	0.36	0.91	61.7	0.70	0.93	71.3	58.27	[-30,142]	40.2	53.9	[-27,124]	51.6
		Yes	CS-SR_F	1408	0.69	0.91	83.9	0.82	0.93	86.3	32.5	[-30,65]	66.7	30.9	[-27,70]	72.2

3.3.3 Crowdsourced Calibration (CS-SR)

The model predicted the observed discharge within acceptable ranges when calibrated and validated against the crowdsourced water level data without applying the Water-Balance-Filter. The mean NSE performance decreased by 34.9% during calibration in comparison to the *Q-NSE* scheme to similar values than the ones achieved with the *Q-SR* calibration scheme. However, the *CS-SR* scheme outperformed the lower benchmark model. The PBIAS revealed a decrease of the relative performance of 58.9% in relation to the relative performance of the PBIAS during calibration for the *Q-NSE* scheme. A comparable decrease could be observed during validation. Since the mean PBIAS is >0 in all cases, the CS-SR schema tends to overestimate the overall discharge.

3.3.4 Water-Balance-Filter Effects on the Calibration ($Q\text{-}NSE_F$, $Q\text{-}SR_F$, $CS\text{-}SR_F$)

No differences were observed between the *Q-NSE* and the $Q\text{-}NSE_F$ scheme since all accepted parameter sets within the *Q-NSE* scheme already matched the water balance and subsequently no parameter set was discarded. For all R_{Spear}-based calibration schemes, the filter improved the model performance notably. This holds regardless of the data set used for both the discharge-based calibration (*Q-SR*) and the crowdsourced water level data calibration (*CS-SR*). The relative performance for these calibration schemes increased to comparable values between 84% and 86% during calibration and validation for NSE and between 66% and 72% for PBIAS. Hence, calibrated with crowdsourced water level data combined with the Water-Balance-Filter ($CS\text{-}SR_F$), the model predicted the discharge almost as well as if calibrated on the observed discharge (*Q-NSE*). This applies for the behavior of both model efficiency measures, the NSE and the PBIAS.

Figure 20 shows the modeled discharge time series during calibration and validation for the *Q-NSE* scheme and the crowdsourced-based calibration scheme (*CS-SR* and $CS\text{-}SR_F$). This figure underlines the similarities and differences between the different calibration methods. In general, all calibration schemes tended to slightly overestimate base flow conditions. Remarkably, all schemes resulted in similar lower discharge bands and only the upper discharge band deviated for the scenario *CS-SR* compared to the scenarios $CS\text{-}SR_F$ and *Q-NSE*, which was also reflected in the PBIAS.

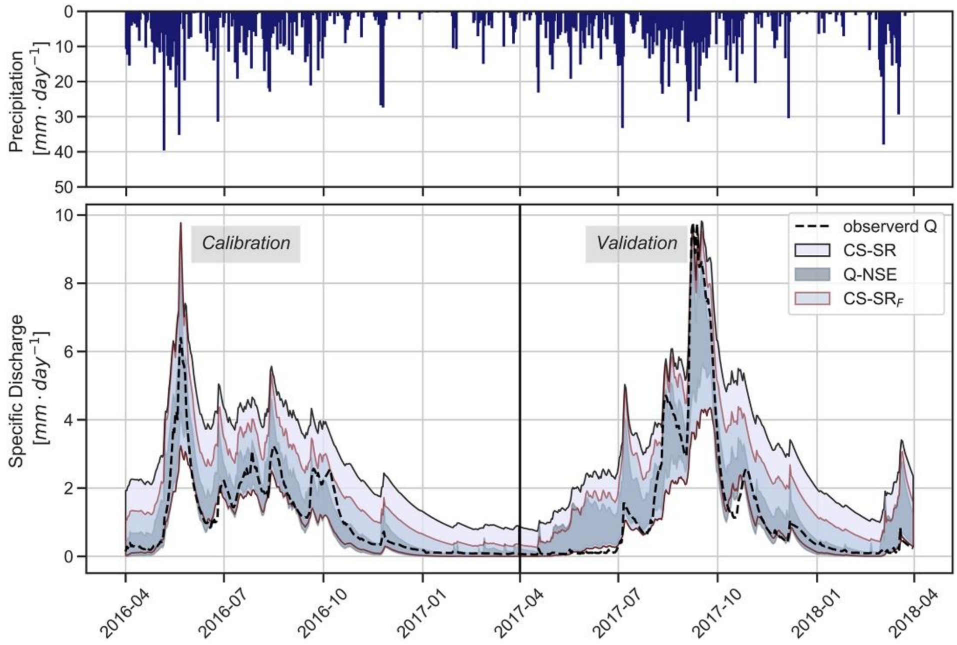

Figure 20: Observed precipitation (top) and discharge (black dashed line in the lower box) in the study area from April 2016 to March 2018. Simulated discharge for three different calibration schemes during calibration and validation (from light blue to dark blue: CS-SR, CS-SR$_F$ and Q-NSE), where Q-NSE indicates a traditional calibration against observed discharge data, CS-SR a calibration against 2500 runs with the highest Spearman-Rank-Coefficient when calibrated against the crowdsourced water level data and CS-SRF a calibration using the same runs obtained from CS-SR but filtered for a maximum yearly runoff based on an estimated water balance using observed precipitation and actual evapotranspiration derived from MODIS.

3.3.5 Comparison of Different Calibration Schemes

We analyzed specific flux components simulated by the model to further understand and evaluate the model behavior regarding the different calibration schemes. This allowed us to assess whether the simulated processes are within realistic boundaries and whether the different calibration schemes influence the hydrological fluxes. A large discrepancy between the individual fluxes would be questionable and indicate a mismatch between the model simulations and the underlying processes. The same applies to abnormally large or small values for the actual evapotranspiration. In addition, the analysis provides insights into the range of the simulated flows under the various calibration schemes and thus into the related model uncertainties. Figure 21 shows the distribution of the sums of each flux for every model run within the calibration schemes for the validation period (the figure for the calibration is similar and not shown) excluding the Q-NSEF scheme, because of its redundancy to the Q-NSE scheme. The results reveal an equal distribution of the modeled flux components for all five calibration schemes. The variability in fluxes is smallest for the Q-NSE scheme and increases for the filtered (Q-SRF, CS-SRF) and unfiltered (Q-SR, CS-SR) schemes. For example, the range of simulated ET values under Q-SR was largest (359-1,076 mm), and the contribution to the total water balance was on average (mean 693 mm) lower than for Q-NSE (mean 940 mm). Consequently, more water left the system from the storage box to the outlet in the Q-SR scheme compared to Q-NSE. This can also be seen in the time series where the Q-SR scheme (similar to the CS-SR scheme) tends to overestimate the flow (Figure 20). The distributions within the unfiltered or filtered calibration schemes are comparable. Consequently, the R_{Spear} calibrated data sets show a similar distribution regardless of whether they were calibrated to the discharge or the citizen-based water levels. The proportion of surface runoff (SW) was low for all three methods. This is in line with the general process understanding for this catchment and its environmental conditions (Jacobs et al. 2018a). Surface runoff can occur during heavy rain events but remains low. A high fraction of surface runoff would, therefore, not be realistic.

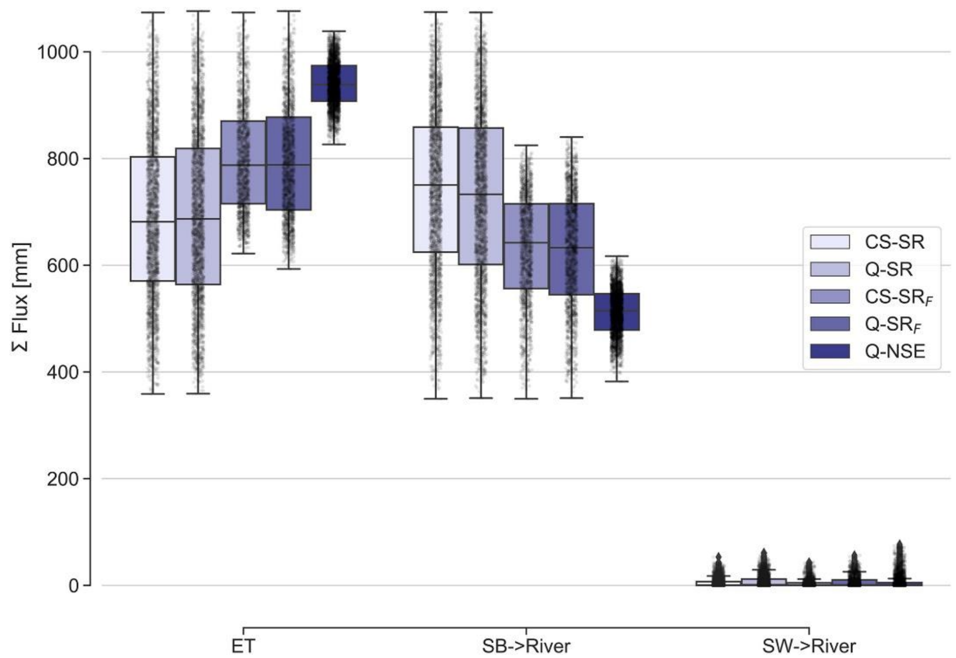

Figure 21: Boxplots of the sum of fluxes released by the different model components (ET = actual Evapotranspiration, SB->River = Water released from the Storage Box to the Outlet, SW->River = Water released from the Surface Water Storage to the Outlet) under different model calibration schemes (CS-SR = calibration based on crowdsourced water level data, Q-SR = calibration based on discharge and the Spearman-Rank-Coefficient, CS-SRF = calibration based on crowdsourced data in combination with a Water-Balance-Filter, Q-SRF = calibration based on discharge and the Spearman-Rank-Coefficient data in combination with a Water-Balance-Filter, Q-NSE = traditional calibration process based on discharge data and the Nash-Sutcliffe model efficiency coefficient) during the validation period.

3.4 Discussion

To date, integrating crowdsourced data into hydrological monitoring remains uncommon with only a few studies that examined the value of these data for model calibration (Mazzoleni et al. 2017, Starkey et al. 2017, Etter et al. 2018). All these studies, however, lack a direct comparison between a model driven by real-world crowdsourced data and a model calibrated using conventionally collected discharge measurements. Some of the studies used synthetically generated data to mimic crowdsourcing. In contrast, our study compares the efficiencies of a model calibrated on crowdsourced water levels, combined with or without a Water-Balance-Filter, against results from a model calibrated on measured daily discharge. Projects like *CrowdHydrology*, which resulted in more than 16,000 observations since 2010 ongoing, proved that it is possible to gather this type of data, also over extended periods (Lowry et al. 2019). We found that the model calibrated with crowdsourced water levels combined with a Water-Balance-Filter performs similarly in terms of model efficiencies but results in greater model uncertainties and overestimations of discharge.

3.4.1 Assessing Model Performance Through Spearman-Rank-Coefficient

The Spearman-Rank-Coefficient (R_{Spear}) used to identify behavioral parameter sets during calibration allowed a comparison of the dynamics of the simulated discharge and measured water levels but revealed no information on total discharge volumes (Jian et al. 2015). There are no previous reports as to which parameter sets should be declared as behavioral when using the R_{Spear} as an objective function. Consequently, we ranked all parameter sets by their R_{Spear} value and chose a certain percentage (0.25% of 10^6 runs) as behavioral. An advantage of using the R_{Spear} is that for example no extra Inverse Rating Curve function with additional uncertainties needs to be estimated (Jian et al. 2015) and that water levels can directly be used to calibrate the model. This avoids the tedious process of discharge measurements, which also requires special equipment and expert knowledge, often not available particularly in remote places in low-income countries. This might change in the future when new methods to determine discharge like particle-image-velocimetry (Adrian 1991) produce sound results, especially when such approaches become operational on consumer electronic devices like smartphones as shown by Lüthi et al. (2014). We believe, however, that water level measurements carried out by citizen scientists remains easier, reduces measurement errors, requires no expert knowledge, delivers reliable measurements and can promote community participation (Lowry and Fienen 2013, Weeser et al. 2018).

3.4.2 The Value of a Water-Balance-Filter

As stated above, a risk exists that a hydrological model might be biased when only calibrated with water level data. Seibert and Vis (2016) addressed this issue when calibrating a model for more than 600 catchments in the United States using daily water level and discharge data. Their study revealed that models that were calibrated on water levels performed well in wet catchments where the precipitation input was higher than the potential evapotranspiration. Seibert and Vis (2016) related this to the fact that the actual evapotranspiration in these catchments was close to the potential evapotranspiration which diminished the influence of different parameter sets on this term of the water balance. Our results confirmed these findings by showing acceptable results in a catchment with precipitation values close to potential evapotranspiration. At the same time, these results indicate that a more intense testing of the approach under different environmental conditions is needed. The hydrological behavior of different catchments might or might not have a further impact on the transferability of our approach, which we finally cannot decide based on a single catchment study. Seibert and Vis (2016) indicated that some volume information might improve the results for drier catchments and the authors stressed the need for further research on this field. In our study, we tested the added value of a Water-Balance-Filter on the parameter set selection to reduce the risk of selecting parameter sets that result in biased model calibration. However, we have to point out that the uncertainty of the actual evapotranspiration derived from the MODIS data set cannot be determined precisely since it depends on various local factors. Mu et al. (2011) identified uncertainties in the used algorithm input data (such as the daily meteorological data), inaccuracy of the measured eddy covariance flux tower data, the scaling from the flux tower point measurements to the landscape and algorithm limitations as main factors, which influence the bias between estimated and measured ET_{act}. When we compared the derived ET_{act} from MODIS with our measured precipitation minus the measured discharge and neglected storage changes (Senay et al. 2011), we found an overestimation for the remotely sensed actual evapotranspiration of 7.7%. After applying the uncertainty compensation of ±30%, the resulting Water-Balance-Filter range falls within the measured ET_{act} value. Consequently, our *CS-SR$_F$* and *Q-SR$_F$*

results showed that the model efficiencies improved when those parameter sets, which were selected as behavioral in the first step using the R_{Spear}, were further filtered. The filter effectively removed model runs that resulted in a discharge overestimation. In contrast to that, these runs were accepted within the unfiltered R_{Spear}-based calibration schemes (*CS-SR*, *Q-SR*) since no volume information was considered.

All schemes resulted in fluxes that were in line with the general process representation. The analysis of the individual fluxes showed that the different schemes did not change the general process understanding of the model. Evapotranspiration was calibrated differently, which resulted in more water draining into the river in the crowdsource-based model schemes. Having in mind that the approach should be applicable under remote conditions or in understudied catchments, we developed a filter that can be easily derived from publicly available data sources rather than aiming for a high precision of the filter itself. The uncertainty factor (30%) we used to define the Water-Balance-Filter based on the measured precipitation and remotely sensed evapotranspiration might deviate for other input data or catchments. We, therefore, argue that a wide range should be chosen. Since the filter only reduces the previously selected parameter sets but does not affect the calibration process itself, the filter has no negative influence on the results. Our results show that including such a simple filter in the *a posteriori* model selection process reduces effectively the bias that is inherent when calibrating the model using R_{Spear} as an objective function.

In general, the increasing availability of remotely sensed data brings new opportunities to obtain relevant water balance variables, particularly in regions where in situ monitoring networks are sparse (Montanari et al. 2013), although the spatial resolution is coarse and ground-truthing often is required. The sparse repeat cycle of satellite data hampers the measurement of daily or weekly changes further (Jian et al. 2017) making it impossible to detect or quantify short events which are typical for tropical catchments. Therefore, the combination of crowdsourced observations with remotely sensed data could be a way to support hydrological modeling in areas where no or only limited hydrometric information is available.

3.4.3 The Role of Input Data and Innovative Input Data Sources

Beside water levels, we used precipitation and temperature-based calculated evapotranspiration as inputs for our model. The quality and resolution of these data influence the model performance. We used precipitation and temperature data from automatic meteorological stations, with a controlled quality to demonstrate the feasibility of calibrating a model using crowdsourced water levels. However, these data might not be available in all cases and can become an additional error source. For larger catchments, where the spatial resolution might be less important which can lead to smoothing effects, these data could be derived from remote sensing or interpolated using measurements from existing meteorological stations. Beyond that, it is possible that precipitation and temperature measurements are performed by citizen scientists. Starting in 1998 as a local project the CoCoraHS (the Community Collaborative Rain, Hail, and Snow network) became the largest provider of daily manual rainfall measurements in the United States with 37,500 participants and over 31 million crowdsourced daily precipitation reports (Reges et al. 2016). A study by Walker et al. (2016) showed that an Ethiopian community monitored precipitation sufficiently for 18 months resulting in a high correlation between the crowdsourced data and data from a national station. By using a community-based rain gauge network, a high spatial resolution might compensate a potentially lower data precision since local rain events can be captured, which cannot be detected by coarser professional networks (Kirchner 2006). Technical development opens the potential for new and alternative data collection methods which could contribute to improved availability of data. Overeem et al. (2013), for example, showed the possibility to estimate daily mean air temperatures from smartphone battery temperatures, while Messer et al. (2006) described a method how the signal levels of cellular networks can represent precipitation amounts. Gosset et al. (2016) claimed that this technique is particularly suitable for areas that lack precipitation measurement infrastructure including large parts in Africa. Linking data from different and innovative methods together may have great potential for hydrological modeling.

3.4.4 Model Structure and Data Resolution

The conceptual model used in this study involved only five parameters, which allowed a consistent calibration and avoided over-parameterization (Kirchner 2006). Furthermore, since few parameters are involved, the model can be easily applied in data scarce regions. However, a more complex physically based and/or spatially distributed model might have benefits by providing the opportunity to use observed data from various sources and locations and integrate them into the model approach (Starkey et al. 2017). Mazzoleni et al. (2017) demonstrated the use of synthetically generated crowdsourced streamflow observations in a spatially distributed model to improve flood predictions. These authors showed that the temporal variability of data influenced the results less than their accuracy, which confirms the usefulness of crowdsourced data given that their accuracy is assured. However, even the resolution of the water level scale (vertical resolution) is not an exclusion criterion. For example, van Meerveld et al. (2017) demonstrated that the vertical resolution of water level measurements is less critical. These authors used a time series of only two stream level classes to calibrate a conceptual model successfully. These findings may further increase the applicability of crowdsourced data as it allows the use of data with reduced vertical resolution and hence reduced accuracy and temporal resolution. A study by Seibert et al. (2019) showed that virtual water level gauges, generated by a mobile application, can be used to monitor water levels in any stream without physical installations, which can make the approach scalable. These results indicate a promising way to increase the spatial coverage of crowdsourced measurements in future.

3.4.5 Crowdsourced Versus a Discharge-Based Calibration

The often-expressed concern that data irregularity induces problems can therewith be mitigated. Our study confirms this assumption since no evidence was found that data irregularity within the crowdsourced data affected the model performance and the model could be calibrated using the crowdsourced data which had a variable temporal resolution and measurement uncertainty. The crowdsourced-based calibration schemes led to comparable results as the discharge-based calibration when using the R_{Spear} performance measure. The increased uncertainty is therefore mainly induced by using the R_{Spear} and only marginally by the crowdsourced data itself. The crowdsourced data only led to a decrease of the relative performance of around 5% for both the NSE and PBIAS during calibration (*CS-SR*) in comparison to the discharge-based calibration (*Q-SR*). Compared to the NSE-based calibration (*Q-NSE*) the relative performance decreased by 30-35% under the R_{Spear}-based schemes regardless of the model was calibrated on discharge or crowdsourced water level data (*Q-SR* and *CS-SR*).

3.5 Conclusions

Based on our results, we suggest crowdsourced monitoring approaches as an additional tool for water resources management, particularly in ungauged or poorly gauged catchments and under limited financial ressources. Combining simple measurement carried out by citizen scientists with a modeling approach could be a way to improve our knowledge of available water resources and process understanding in catchments that have so far been understudied. This approach may be an alternative in places where observational gaps are caused by a lack of hydrometeorological gauging networks. However, some limitations are worth noting. Although our findings provide evidence that crowdsourced water levels can be used to calibrate hydrological models, the outcome also depends on the quality of other input data and the general catchment behavior. Our study area only had a few flooding events and the water balance seems to be fairly simple. Consequently, the observed discharge could be modeled well using a simple model structure. The crowdsourced data we used in this study is from outstanding quality with high temporal coverage and low measurement errors. Future work should, therefore, investigate the behavior of crowdsourced calibrated models for catchments of different land use and climatic conditions, test the implementation of crowdsourced climate data and investigate the impact of crowdsourced data of various quality.

Based on our evidence, we provide the following answers to the research questions raised in section 1:

(1) Are water levels collected by citizen scientists suitable for calibrating a rainfall-runoff model with an uncertainty similar to the uncertainty resulting from a calibration with conventional data sources?

A conceptual rainfall-runoff model can be calibrated on crowdsourced water level data. The combination of crowdsourced data and a rainfall-runoff-model might solve an often-raised critical point when using crowdsourcing in hydrology, that is, data irregularity. After a 1-yr calibration, the model transforms the community-based collected data into a continuous time series. This is particularly valuable when one considers that only water levels were used for calibration and no discharge measurements had to be carried out, which would have required special equipment and training of the citizen scientists. The model, which was only calibrated against water level data, predicts the observed discharge in acceptable ranges, but the efficiencies were lower than the efficiencies of a model that was calibrated on conventional discharge data. However, the lower efficiencies are mainly introduced by using the R_{Spear} as a performance measure, which leads to an overestimation of the discharge.

(2) Can the model uncertainties be reduced by using a simple to obtain Water-Balance-Filter as an additional criterion?

By applying a simple Water-Balance-Filter it was possible to achieve model efficiencies similar to those obtained from traditional calibration against streamflow. We used a parsimonious water balance derived from measured precipitation and remotely sensed evapotranspiration data, avoiding data-intense estimates. Similar water balances can be established in other data-sparse regions. Combining the filter with the rainfall-runoff model increased the model reliability. The filter can compensate the effect of keeping parameter sets that result in unrealistic high or low volumes which can occur when using the Spearman-Rank-Correlation as an objective function. To achieve this effect, the uncertainty of the remotely sensed actual evapotranspiration data should not have led to values that are too far from the real actual evapotranspiration in the respective catchment.

3.6 Acknowledgments

All data and scripts (model and data analysis) used in the study can be obtained from Zenodo (DOI: 10.5281/zenodo.3341592).

We would like to express our thanks to the citizens of the Sondu-Miriu River basin who participate in our program. We also thank the Deutsche Forschungsgemeinschaft DFG (BR2238/23-1), the Deutsche Gesellschaft für Internationale Zusammenarbeit GIZ, and the 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 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. This work was partially funded by the CGIAR program on Forest, Trees and Agroforestry led by the Centre for International Forestry Research (CIFOR).

4 Crowdsourced Water Level Monitoring in Kenya's Sondu-Miriu Basin – Who is "the crowd"?

This chapter is submitted to Frontiers in Earth Science as:

Weeser, B., J. Gräf, N. K. Njue, P. Cerutti, M.C. Rufino, L. Breuer, and S. R. Jacobs Crowdsourced Water Level Monitoring in Kenya's Sondu-Miriu Basin – Who is "the crowd"? *Frontiers in Earth Science* (under review).

4.1 Introduction

Currently, two billion people live in an environment where recurrent water stress is expected, hindering sustainability and limiting social and economic development (United Nations 2018). An increasing water demand from the growing human population will further exacerbate water stress, particularly in certain parts of sub-Saharan Africa (Le Blanc and Perez 2008). To meet this growing water demand and to allocate water equitably, especially in the context of climate change, improvement of water management practices is crucial. To develop such plans and practices, comprehensive and expensive monitoring approaches as well as sound data are needed. However, studies show that the amount of water resource monitoring networks is actually declining worldwide (Vörösmarty et al. 2001, Ruhi et al. 2018). Low-income countries, where improved water management is particularly urgent, often lack the necessary infrastructure and financial capacities. Despite the increased availability of low-cost sensors, remote locations, vandalism, and limited capacity building impede the use of advanced technical devices. Remote sensing approaches have become increasingly available but are still not operational for small catchments. Yet, appropriate monitoring and management of small headwater catchments is crucial to ensure water supply to local communities and downstream regions.

As conventional monitoring approaches are not always adequate or feasible to implement in low-income countries, new ways of data collection need to be explored. Such methods should not rely on major investments, specialized equipment, and highly trained personnel. In recent years, citizen science has increasingly been used for hydrological data collection (Njue et al. 2019). Participation of members of the local community in environmental monitoring offers the possibility to strengthen local stakeholder cooperation, while the data collected by the community members supports evidence-based management decisions (Overdevest et al. 2004, Domroese and Johnson 2017). Additionally, community members can provide valuable local knowledge to support the development of management plans (Whitelaw et al. 2003, Nare et al. 2006). Numerous studies have shown the successful integration of citizens into, for example, water level monitoring (Lowry and Fienen 2013, Weeser et al. 2018, Lowry et al. 2019, Seibert et al. 2019), precipitation measurements (Reges et al. 2016, Davids et al. 2019a) and water quality assessments (Toivanen et al. 2013, Breuer et al. 2015). Furthermore, Weeser et al. (2019) demonstrated that data collected by citizens (crowdsourcing) were valuable for hydrological modelling. Supported by technological developments and the growing use of smartphones, there is an increasing number of environmental variables that can be monitored by citizen scientists (Newman et al. 2012). Citizen science approaches have the additional advantage that they can easily be scaled and, therefore, generally have a better spatial coverage than conventional measurement approaches.

Citizen science has been identified as a highly promising tool for monitoring the sustainable development goals (Quinlivan et al. 2020) and for sustainable development in low-income countries (Pocock et al. 2019). However, the success of any citizen science project depends on the willingness of volunteers to invest their time and knowledge (Parrish et al. 2018). Therefore, knowing what motivates and challenges people's participation can help to design a successful citizen science project (Shirk et al. 2012). Furthermore, characterizing the socio-economic background of highly motivated participants is important to target the right people (Etter 2020). Although few studies have made an attempt at analyzing the motivation and methods of engagement of citizen scientists (e.g., Aoki et al. 2017, Rutten et al. 2017, Phillips et al. 2019, Golumbic et al. 2020), most of these studies have focused on western countries, where the majority of the citizen science-based hydrological monitoring programs have been implemented (Njue et al. 2019). Because these findings might not apply to low-income countries due to socio-economic and cultural differences (Hacker et al. 2017), we conducted a telephone survey to explore the motivation, challenges and socio-economic background of participants in a citizen science water level monitoring project in the Sondu-Miriu basin, western Kenya. Based on this case study, we identified socio-economic characteristics of the participants with sustained long-term engagement in crowdsourced water monitoring. Finally, we identified motivations, challenges, and opportunities for improving the engagement of the local community in water monitoring to support sustainable water management.

4.2 Material and Methods

4.2.1 Study area and project background

The Sondu-Miriu basin (3,450 km²) in western Kenya is one of the many river basins contributing water to Lake Victoria and the river Nile. Its headwaters lie in the Mau Forest Complex. With more than 40,000 ha, this is East Africa's largest remaining tropical montane forest and an important 'water tower', providing numerous water-related ecosystem services, such as water storage and supply, groundwater recharge, flood mitigation and micro-climate regulation (Benn and Bindra 2011). Large-scale conversion of forest to agricultural land, particularly smallholder agriculture, and forest degradation have supposedly led to changes in water quality and flow (Mango et al. 2011, e.g. Defersha and Melesse 2012, Jacobs et al. 2017, Jacobs et al. 2018a). The Sondu-Miriu basin reaches from 1,400 m a.s.l. at the outlet to 2,900 m a.s.l. on the Mau Escarpment. Whereas the upper part of the basin receives 1,900 mm rainfall per year, the lower part is a lot drier (1,300 mm y^{-1}) and regularly experiences flood events during the rainy season. In addition to the challenging climatic variation within the basin, sustainable water management is further hampered by the lack of data of sufficient quality and spatiotemporal resolution.

To improve the data availability and coverage in the Sondu-Miriu basin, a citizen science water level monitoring project was implemented in 2016. Together with the local Water Resources Authority, 14 monitoring stations were selected and gauges restored or installed (Figure 22). A sign with instructions and station code was installed at each site. In principle, this would allow any interested citizen to participate in the project. Data was submitted by sending a simple text message (SMS) with the water level and the station code to a local phone number, provided on the sign. Messages were processed by a server infrastructure based on a Raspberry Pi 2 Model B developed specifically for the project (full details available in Weeser et al. (2018)) The participant received a response message, thanking the sender for its contribution and repeating the value and site. The sending time, message and sender number were stored in the server.

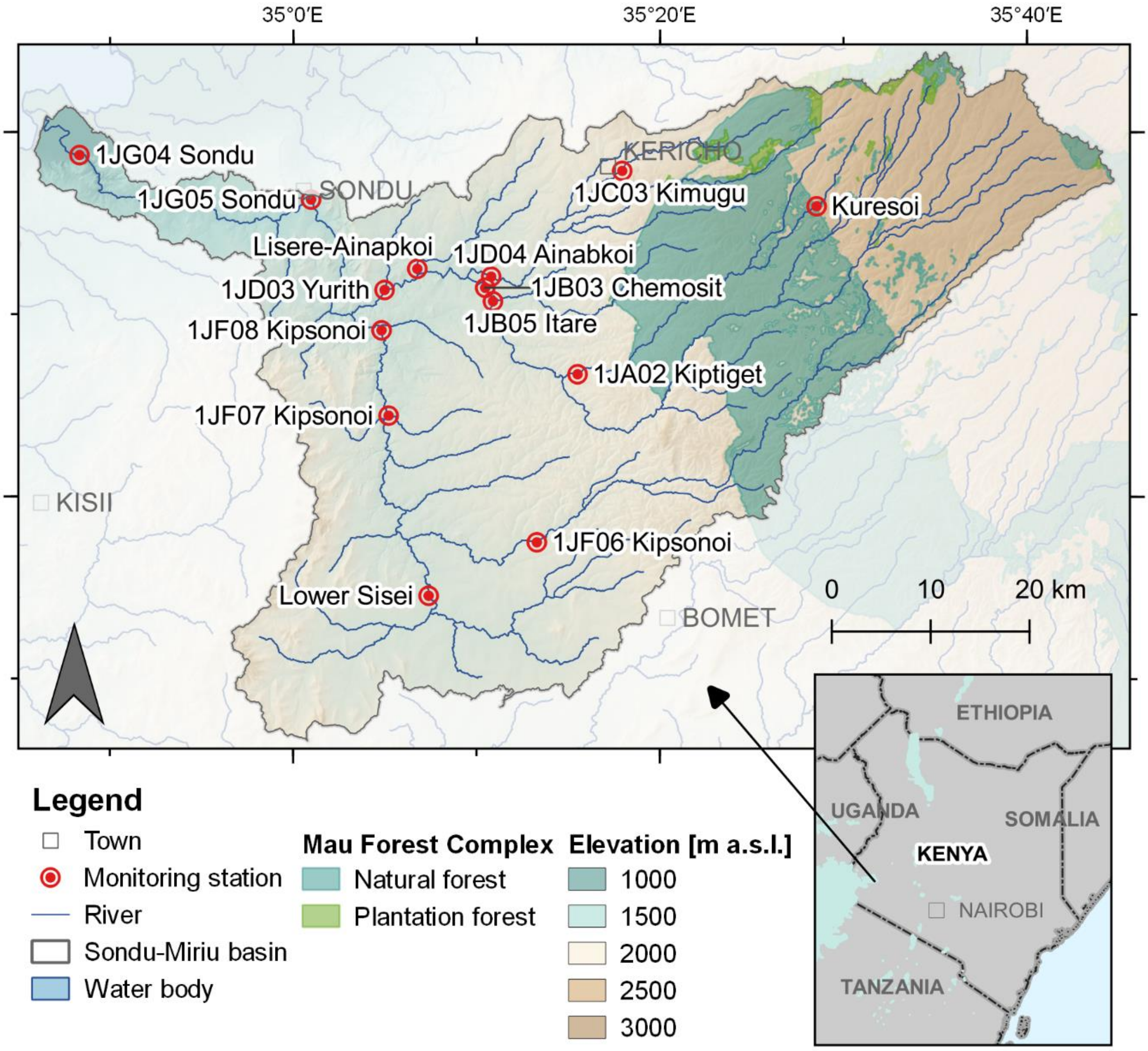

Figure 22: Map of the Sondu-Miriu basin in western Kenya, indicating the citizen science water level monitoring stations. Coordinates are displayed in WGS 1984.

At the start of the project, sensitization meetings were conducted with the help of local administration or chairperson of the Water Resource Users Associations (WRUAs), depending on their presence near the station. The establishment of WRUAs was enabled by the Kenyan Water Act (2002) to support the implementation of water management at the grass root level (Omonge et al. 2020). They are considered important to reduce water allocation conflicts (Mutiga et al. 2010) and enhance the users' involvement and participation in setting goals and implementation of water management through the development of subcatchment management plans (Omonge et al. 2020). Through these subcatchment management plans, the WRUAs are supposed to promote sustainable and equitable water use, to safeguard water supply to fulfil ecological demands and basic human needs and to improve water quality and quantity through catchment conservation measures (Richards and Syallow 2018). Therefore, the WRUAs were considered as an essential stakeholder in the project and a good entry point to reach community members with an interest in environmental conservation and management. During the sensitization meetings, participants were informed about the importance of environmental monitoring and purpose of the project. Participants were also trained on how to read the water level gauge and how to send data to the SMS server.

4.2.2 Telephone survey

To obtain information about the background and motivation of the participating citizen scientists, standardized telephone interviews were conducted in the first week of July 2017. We decided to use a telephone survey, since the telephone numbers of all citizen scientists were available from the SMS server. To overcome potential language barriers, a team of interns of the Water Resources Authority office in Kericho was engaged as enumerators, being fluent in Swahili and English, as well as one of the local languages (Luo or Kalenjin). The enumerators received training and conducted test calls.

All telephone numbers from which at least one message was sent between the start of the citizen science water monitoring project in April 2016 and the 30th of June 2017 were extracted from the SMS server. The telephone members of project staff and employees of the Water Resources Authority were excluded, as well as numbers sending messages for commercial and other non-project related purposes. Each telephone number was called up to three times at different times and days until contact was established. If these attempts were unsuccessful, the person was informed via text message about the survey and asked to arrange a suitable time if interested to participate in the survey. No monetary incentive was offered for participation and each interview took about 10 minutes.

The survey consisted of open and pre-coded questions and was structured in two blocks (Appendix 4-1). The first block consisted of questions to assess the motivation, possible obstacles and background information, like what kind of phone the participant uses and distance to the gauge. These questions were structured in four open and five pre-coded questions. In the second block, three open and two pre-coded questions were asked to obtain socio-economic background information. Some pre-coded questions provided the possibility to give an open answer if the participant did not find an appropriate answer within the given possibilities. The answers of the participants were recorded on printed, standardized survey sheets.

4.2.3 Data analysis

The survey was digitized by entering all answers in Microsoft Excel. Answers from open questions were coded using the manifest message method (Weisberg et al. 1996), whereby obvious themes, messages and points were extracted from the answers and coded accordingly. For these questions, it was possible to have more than one answer per respondent. A dataset with only valid cases (i.e. survey respondents) was analyzed using R studio 1.2.1335.

4.2.3.1 Engagement

The degree of engagement of citizen scientists was assessed using two measures. Firstly, participants were classified according to the number of valid measurements sent to the SMS server between April 2016 and June 2017. A measurement was considered valid when the site and water level reading could be identified from the message, either by an algorithm implemented on the SMS server itself (Weeser et al. 2018) or through manual interpretation. After inspecting the dataset distribution, participants with 0 or 1, 2 to 9 or 10 or more valid measurements were classified as low, medium and high level of engagement, respectively.

As a second measure of engagement, telephone numbers from which water level measurements were submitted between the 1st of July 2017 and the 31st of December 2018 were extracted from the SMS database. If a participant continued to submit measurements after completion of the telephone survey, the participant was classified as a long-term participant.

4.2.3.2 Random forest

We used the random forest algorithm (Breiman 2001) to classify respondents as having a low, medium or high level of engagement and whether they were short-term or long-term engaged, based on several explanatory variables. The latter included source of information about the project, frequency of passing the station, distance to station, type of phone, age class, level of education and WRUA membership.

For the two indicators of engagement (level and duration), we ran the randomForest function from the R package 'randomForest' (Breiman et al. 2018), creating 5,000 decision trees using sampling with replacement and testing 2 variables at each node. The function calculated the out-of-bag (OOB) error rate, which we used as indicator for model accuracy. The importance of each explanatory variable was assessed with the mean decrease in Gini index by exclusion of the variable, as calculated by the randomForest function. The order of the explanatory variables based on decreasing Gini index value represents the relative importance of each variable to classify the dependent variable. Only cases without missing data were included in the analysis. The algorithm was run 10,000 times, following a Monte Carlo approach, whereby the median values for OOB and the importance of the explanatory variables over all runs was calculated, together with the minimum and maximum values as a measure of uncertainty.

4.3 Results

4.3.1 Engagement of project participants

Out of 155 phone numbers submitting a message (referred to as participants) between April 2016 and June 2017, 87 took part in the telephone survey (referred to as respondents), resulting in a response rate of 56%. Six respondents did not submit a valid measurement. For three of these, the site from which they sent the message could be identified. The remaining three respondents were classified to site "Unknown", together with other participants sending only invalid measurements without being able to identify the corresponding site (n=14; Figure 23).

Among all participants, 67% sent 0 to 1 valid message and were therefore classified as showing a low level of engagement (Figure 23). Nevertheless, 11 of these participants contributed additional measurements after June 2017, suggesting a long-term commitment to the project. In total, 83% of the citizen scientists showing long-term commitment (n=23) participated in the survey. The response rate was highest under participants with a high level of engagement (93%, n=14), followed by respondents with a medium (64%, n=23) and low level of engagement (48%, n=50).

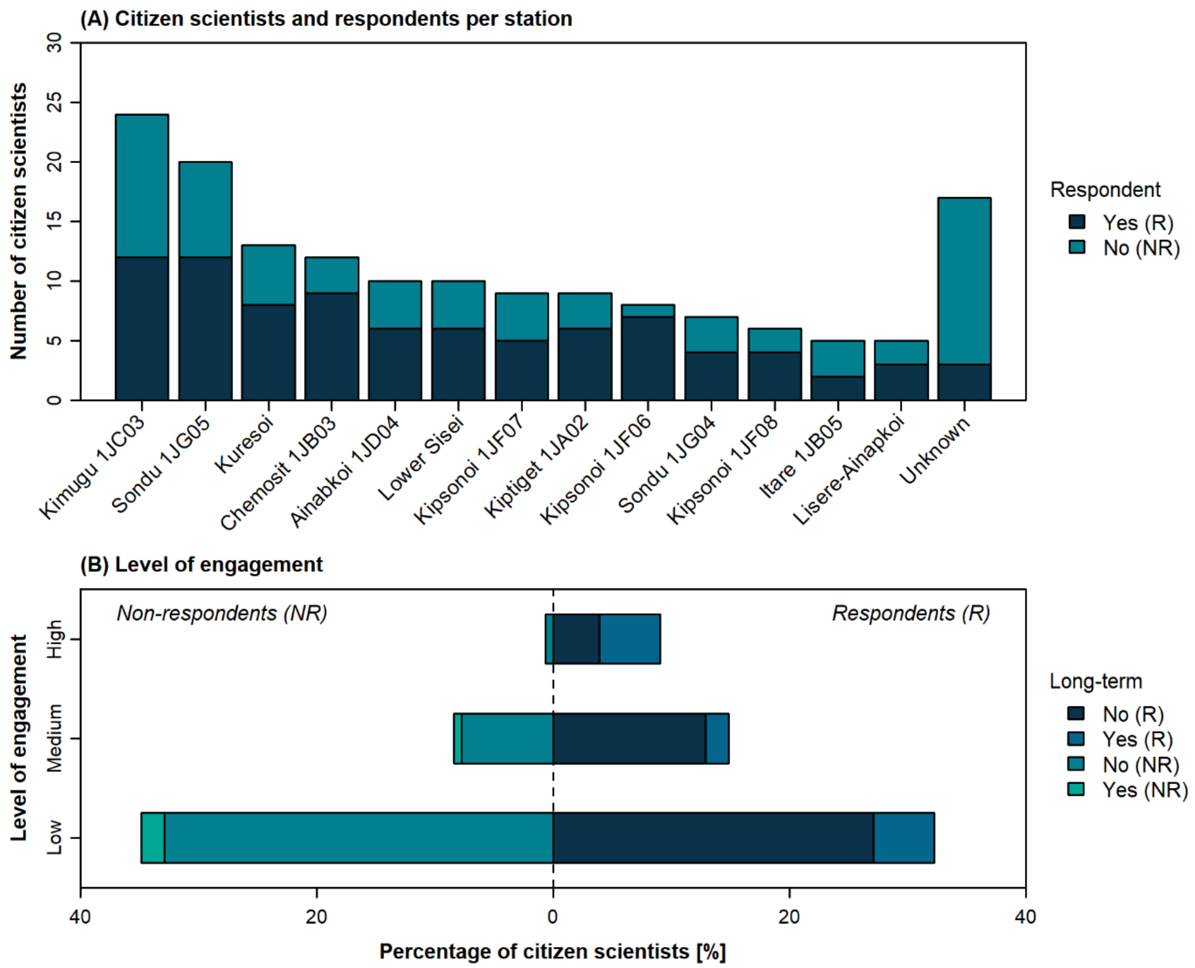

Figure 23: Number of citizen scientists (n=155), classified as survey respondent (R) or non-respondent (NR) of the survey, (a) per site and (b) by level of engagement in the crowdsourced water monitoring project in the Sondu-Miriu basin, Kenya between April 2016 and June 2017. Overall engagement was assessed by total number of valid measurements until June 2017 (Low = 0-1, Medium = 2-9, High = 10 or more). Long-term engagement was assessed based on the contribution of additional measurements after June 2017.

4.3.2 Characterization of participants

Only 5 women participated in the survey compared to 78 men (4 respondents did not provide an answer), which seems representative based on female participation in sensitization meetings at the start of the project. During the sensitization meetings a briefing about the project and a short training was conducted with interested citizens and, where available, the WRUA members. The majority of the respondents were under 50 years old (90%, *n*=76), with 34 of the respondents being 20-29 years old (Appendix 4-2). Three respondents did not complete any education, whereas the remaining participants were fairly equally distributed among primary, secondary and higher (e.g. vocational training, college or university degree) education.

The sign at the monitoring sites and sensitization meetings were the most effective methods to reach participants, with 69 respondents (79%) identifying these as their source of information. Eight out of 14 highly engaged respondents (57%) were informed through the sensitization meetings, whereas the majority of the respondents with a low level of engagement (60%) read the sign near the gauge (Figure 24). Other sources of information mentioned by the respondents included the Water Conservation Forum (*n*=1) and the project staff during installation of the gauges (*n*=1). Although only 6 of the respondents indicated to have been informed about the project through the Water Resources Users Association, 20 respondents stated they were members of the local WRUA. Of the non-members, 26 were aware of the WRUA, 28 had not heard about WRUAs before and 13 did not answer the question whether they had heard about the WRUA.

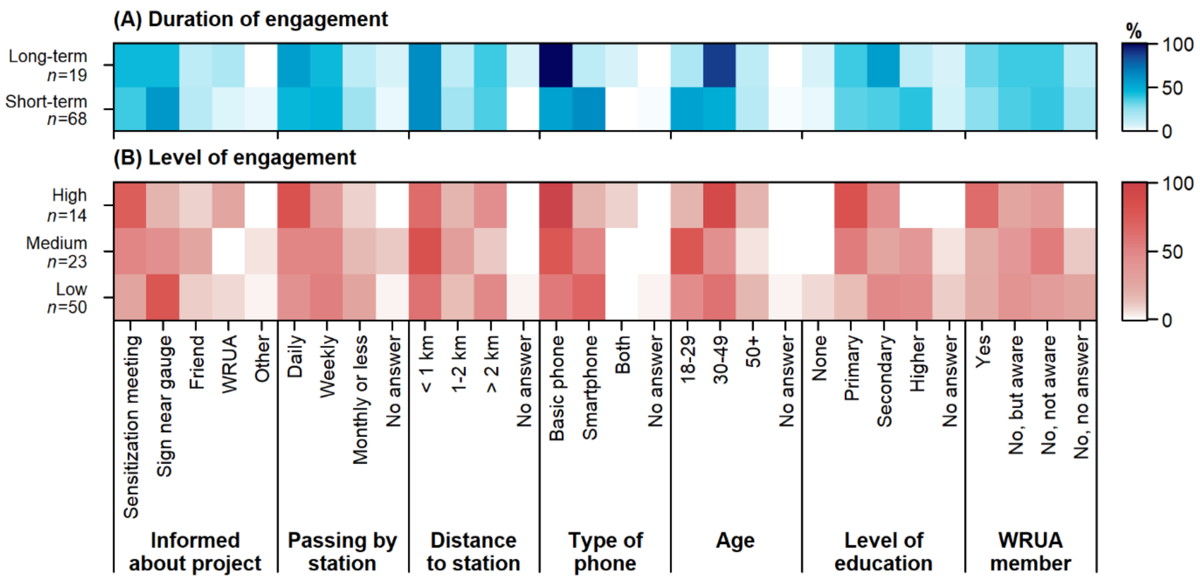

Figure 24: Characterization of the participants of the citizen science water monitoring project in the Sondu-Miriu basin, Kenya, according to different engagement classes, expressed as percentage of respondents within each class. The duration of engagement is based on whether the respondent continued sending data after June 2017. The level of engagement is based on the number of valid measurements contributed between April 2016 and June 2017 (Low = 0-1, Medium = 2-9, High = 10 or more). n = number of respondents within each engagement class; WRUA = Water Resource Users Association.

The observation of the water level ('amount of water' in the survey) was perceived most frequently as purpose of the water monitoring project by the respondents (*n*=46; Figure 25a). This was followed by monitoring for management and conservation purposes (*n*=27). Other perceived purposes included monitoring floods, rainfall patterns and the weather

(n=4), to determine water quality (n=2), to warn people for disasters (n=1) and to know about climate change (n=1).

When asked about their motivation for participation, the importance of water as a resource and the desire to assist in conservation and management of the resource was mentioned most frequently (n=27), followed by the willingness to volunteer or curiosity to test the system (n=20; Figure 25b). Fifteen respondents mentioned that they participated because they were triggered by the changes observed in the environment and water supply patterns over the years. Other reasons for participation included the expectation to be paid (n=2), to assist in monitoring the environment (n=1), a general concern about the environment (n=1) and because the respondent previously worked with project partner German Corporation for International Cooperation (GIZ) (n=1).

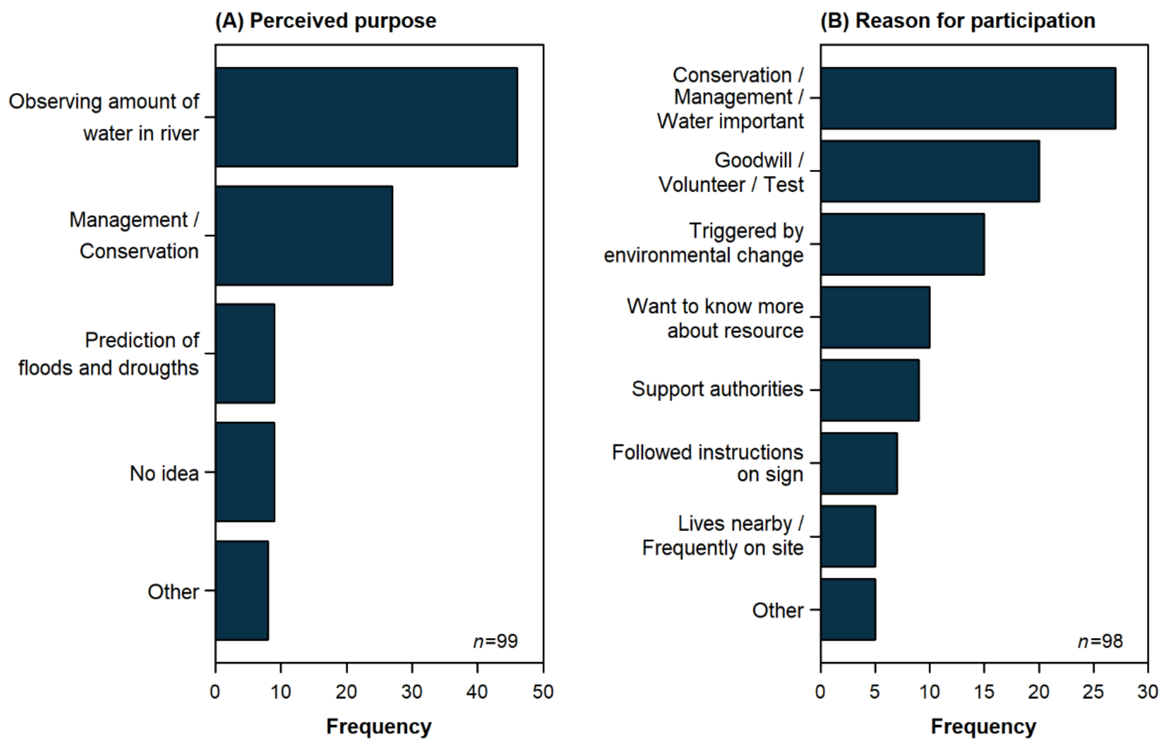

Figure 25: Respondents' answers on (a) the perceived purpose of the water monitoring project and (b) their reason to participate as citizen scientist. n = number of responses.

Roughly half of the respondents (n=46) estimated they lived within 1 km distance of the closest monitoring station, while 26 indicated they lived more than 2 km away. Those living closest passed by the site relatively more often, with daily visits by 24 respondents living <1 km from station, compared to 3 living 1-2 km away and 8 respondents living more than 2 km away. Normal cellphones without comprehensive mobile computing functionality (n=47, 55%) were more often used to participate in the water monitoring project than smartphones. This was especially common among the older (≥30 years old) respondents (n=33 for basic phone, n=17 for smartphone).

Based on the survey information, the highest level of engagement was found for people who were 30-49 years old, with a primary school education (Figure 24b). Highly engaged participants were also characterized by passing by the station on a daily basis, living with 1 km of the station and being a WRUA member. These highly engaged participants mainly learnt about the project through the sensitization meetings. Similar socio-economic background characterized respondents with a long-term engagement, although WRUA membership was less important for this group than for highly engaged participants and most respondents had secondary school education (Figure 24a). A high level of engagement did not result in long-term commitment, with 8 out of 14 highly engaged respondents also showing long-term engagement. Conversely, neither did a low level of engagement preclude long-term commitment: 8 out of 50 respondents with a low level of engagement continued sending data after June 2017.

4.3.3 Challenges and opportunities

In total, 54 respondents (62%) indicated that they experienced no challenges when participating in the project. Nevertheless, 25 of these respondents said they stopped participating. Of the 33 respondents (38%) that did experience challenges, 16 respondents indicated they stopped participating. Difficult access and lack of cellphone credit (each n=10) were mentioned most frequently (Figure 26a). The state of the gauge (either damaged or because the water level was below or above the gauge; n=6) and difficulty with making an accurate reading due to turbulence (n=5) were also mentioned as challenges to sending data. Only one respondent indicated that further training was necessary.

Eleven out of 25 respondents that stopped despite not experiencing challenges indicated to have no clear reason for stopping (Figure 26b). Absence from the area, due to, for example, living far from the station or temporary migration for education purposes, was another common reason (n=12). Other responses included a lack of communication about the project (n=2), loss of the phone or phone number (n=2) and having given the responsibility to another person (n=1). Challenges did not necessarily translate in a reason to stop participation. Note, for example, that out of 5 respondents mentioning lack of cellphone credit as challenge, only one respondent gave this as reason for stopping. In general, experiencing challenges to participate was not a determinant to stop participating (χ^2=0.012, p=0.913).

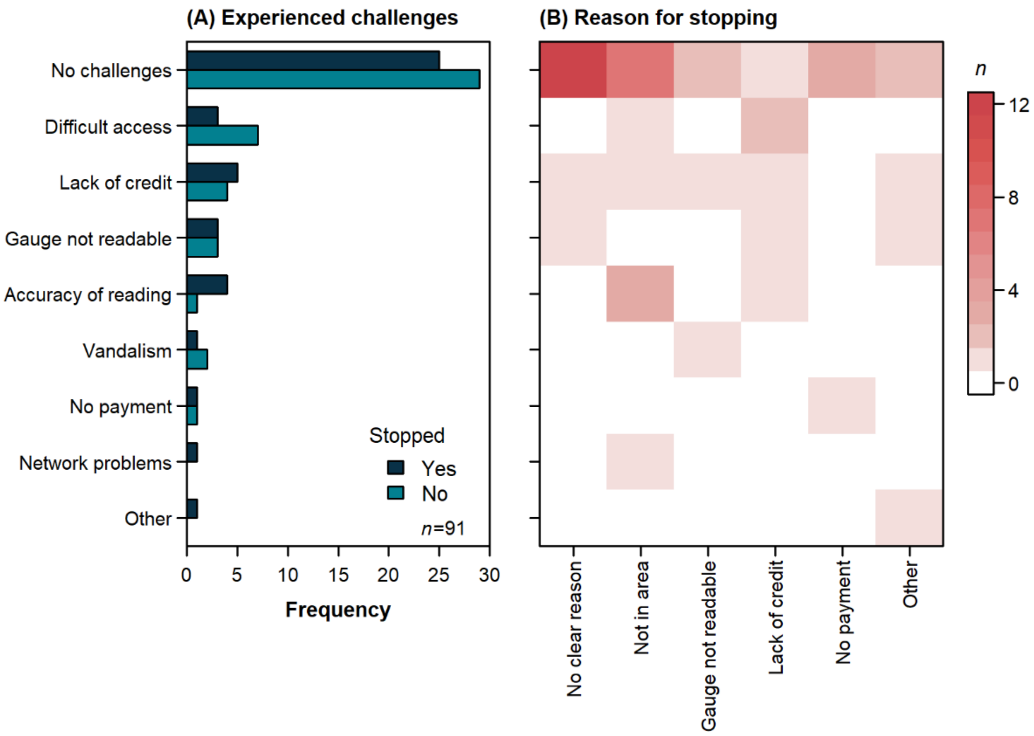

Figure 26: Respondents' answers regarding (a) type of challenges experienced, classified by whether the respondent indicated to have stopped participating, and (b) reason stated for stopping participation, grouped by challenge. n = number of responses.

The majority of the respondents indicated that more training, education and sensitization meetings were required to encourage more people to participate (n=58). Increased advertisement through, for example signs and social media was mentioned 12 times, whereas three respondents mentioned to encourage participation through word of mouth (e.g. community leaders, friends). Paying the participants was mentioned by 18 respondents. In addition, one respondent suggested to use a toll-free number, such that volunteers would not spend their own cellphone credit on sending data. Two respondents mentioned feedback to the community as a way of keeping volunteers engaged, as well as activities by project leads to maintain motivation after the start of the project (n=3). Better targeting of communities (those living next to the river) was mentioned once, as was better maintenance of the gauge.

4.3.4 Explanatory variables

The random forest models had a moderate explanatory power for both duration and level of engagement. The model for the level of engagement had a median out-of-bag error rate of 46.1% (range: 42.1-50.0%). This was caused by high error rates for the classes 'Medium engagement' and 'High engagement' (OOB error rate > 70%). Investigation of the mean decrease in Gini index showed that the variable 'Phone type' was not as important as the other variables. A second run without this variable yielded a lower OOB error rate (median: 40.3%, range: 37.7-45.5%) (Figure 27a). Especially the accuracy for the class 'High engagement' improved (median: 57.1%, range: 57.1-64.3%). Respondents in the category 'Low engagement' were best predicted, with an OOB error rate of 20.9% (16.3-25.6%), but 'Medium engagement' still had a high OOB error rate (70.0%; 65.0-85.0%). The algorithm performed better for the duration of engagement, with an OOB error rate of 26.3% (25.0-30.3%). However, long-term engagement had an OOB error rate of 93.8% (81.3-93.8%), while the OBB error rate for short-term engagement was only 8.3% (6.7-13.3%).

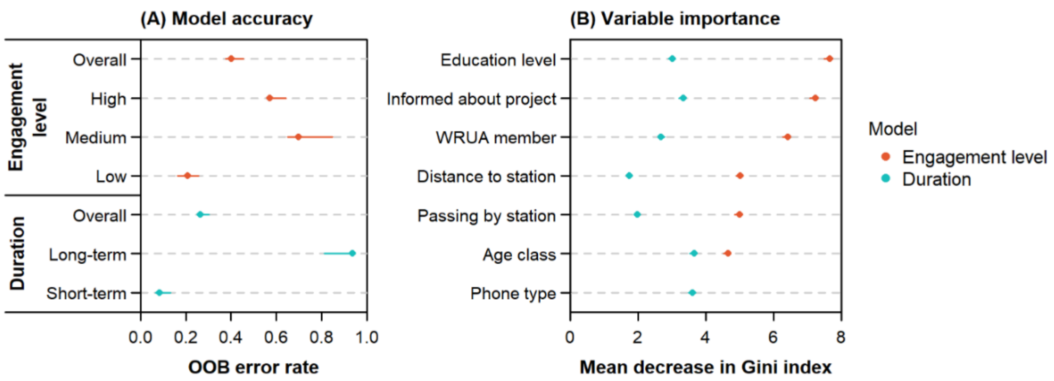

Figure 27: Performance of the random forest models to predict the level and duration of engagement of the participants in the citizen science water monitoring project: (a) model accuracy based on out-of-bag (OOB) error rate and (b) relative importance of the included variables in each model. The circles indicate the median value, the segments indicate the range of values across 10,000 runs.

Based on the mean decrease in Gini index, where a higher decrease indicate a higher variable importance, the highest level of education was the most important variable determining the level of engagement, followed by the source of information about the project and WRUA membership (Figure 27b). The frequency of passing by a monitoring station, the distance to the station and the age class of the participant were less important. Regarding the duration of the engagement, age class, phone type and source of information were the most important variables. As for level of engagement, frequency of passing by the station as well as distance to station were the least important to determine the duration of engagement.

4.4 Discussion

4.4.1 Who participated in monitoring water?

The first aim of this study was to characterize the participants who are likely to be engaged with the project in the long-term. Identification of target groups for citizen science projects could increase the likelihood of success of a project (Parrish et al. 2018, Füchslin et al. 2019). The majority of the respondents in our study was between 30 and 50 years old. This group also had the largest number of long-term or highly engaged participants. This evidence differs from the findings of Füchslin et al. (2019), who found that older people (aged 50 and above) showed a higher willingness to participate in citizen science projects than younger people in Switzerland. However, those who showed more willingness to participate often had more free time (e.g. retirement or part-time work) or had a higher proximity and trust in science (Füchslin et al. 2019). These characteristics are unlikely to apply to the participants in the Kenyan citizen science project, due to the rural setting where many people are dedicated to lifelong full-time farming activities, and many have relatively low education level (e.g. only 25 out 87 respondents reported having received education beyond secondary school).

Younger people (<30 years old) were found to participate less (low or medium engagement level) or short-term, as observed in other studies (e.g. Alender 2016, Beza and Assen 2016, Etter 2020). Nevertheless, age class was a poor predictor for the level of engagement, as demonstrated by the low variable importance (decrease in Gini index) in the random forest model (Figure 27b). A similar distribution of participants among age classes for the three levels of engagement (Figure 24) makes it difficult to use this variable to assess the likelihood of an individual to be highly engaged. On the other hand, there was a clear distinction in age distribution between those who were long-term and short-term engaged, with the majority of the long-term engaged respondents in the 30 to 49 years age class and most of the short-term engaged respondents under 30 years.

Due to the high gender imbalance among respondents (78 men, 5 women), we could not properly assess the effect of gender on participation rate. There is no clear trend on whether men or women are more likely to participate in citizen science projects (Phillips et al. 2019), and Füchslin et al. (2019) found that gender was not important in determining the likelihood of people to participate in a citizen science project. However, the dominance of male respondents in our telephone survey could reflect the gender ratio across all participants in our project. Although a study in the neighboring Nyando basin indicated that women were the most important collectors of water in 77% of the households (Onyango et al. 2007) and are thus more likely to visit monitoring stations frequently, their participation might be limited by a generally lower education level (Republic of Kenya 2019). Furthermore, in certain parts of the Sondu-Miriu basin phone ownership for women is still lower than that for men (Republic of Kenya 2019), which aligns with the overall gender gap in phone ownership in low-income countries (Rowntree et al. 2019). Furthermore, despite attempts by the Kenyan government to increase participation of women in water management by limiting the representation of men to two-thirds in any government arrangement including the Water Resource Users Associations (WRUAs), ensuring equal participation of men and women in barazas (community meetings organized by the area chief) and other events is still challenging (Ifejika Speranza and Bikketi 2018). Therefore, women could be less informed about projects such as our water level monitoring. The high importance of the variable 'Source of information about the project' for both level and duration of engagement and the majority of the highly motivated respondents having been informed through sensitization meetings, indicates that targeted communication could be an important entry point to sustain participation.

The highest completed level of education was also found to be an important characteristic to determine the level and duration of engagement. Unlike citizen science projects in India (Johnson et al. 2014) and the USA (Domroese and Johnson 2017), where the majority of the participants completed education beyond secondary school, respondents with a higher level of education in our study showed less long-term engagement and sent fewer data. A medium to high level of engagement was mainly found under respondents that had only completed primary school, whereas those educated up to secondary school level were more likely to be a long-term participant. As there are few job opportunities in rural areas in Africa, people with higher education diplomas likely move to towns and cities (Ginsburg et al. 2016), and are thus unable to contribute frequently or long-term to a citizen science project in their rural home.

Finally, distance to the station did not determine the level and duration of engagement of the citizen scientists as those living closest to the monitoring stations made up the highest number of respondents in each engagement class. Furthermore, distance to site and frequency of passing by both had a low importance (low decrease in the Gini index) in the random forest model for both indicators of engagement. Nevertheless, the majority of the respondents who were highly or long-term engaged visited the station on a daily basis, suggesting that targeting those people who live closest to the station could help to ensure a good pool of volunteers.

4.4.2 Why do citizens participate?

Knowing the socio-economic characteristic of highly motivated citizen scientists is not sufficient for successful engagement of community members in research or data collection. Adapting the rationale of a project to what motivates potential participants could ensure long-term participation. Following the framework of Batson et al. (2002) to classify motivations for community engagement, the majority of the respondents in our survey indicated altruistic and collectivistic motivations. The respondents hoped to contribute to water management and conservation from the viewpoint that water is an important resource for all and a concern triggered by changes in the environment (e.g. changing rainfall patterns).

Although citizen involvement in such projects is often seen as form of community empowerment (Aoki et al. 2017), none of the respondents indicated that they expected to actively participate in water management. Poor knowledge and information sharing on how communities can contribute to local water management, as enforced through the establishment of WRUAs could contribute to the lack of motivation to take action. For example, one third of the participants indicated not to be aware of the existence of WRUAs and thus of their roles and responsibilities. In addition, when asked what the purpose of sending water level data was, less than half of the respondents indicated purposes such as informing water management, conservation and flood prediction. Improved awareness about the relevance of monitoring the water and increased involvement of WRUAs in this process could help to increase the motivation to participate, as it will be clearer what the overall benefit of this collective action is.

Concern about the amount or quality of the water, as well as environmental changes over time seemed to have triggered the majority of participants to take part. This concern about the environment and altruistic behavior of participants was also found in nature conservation and water monitoring projects (Johnson et al. 2014, Alender 2016, Phillips et al. 2019). In the context of our project, this could also be considered self-interest, because the participant could eventually also benefit from improved water management leading to better access to clean water. Unlike other studies, where an interest to contribute to science was found to be an important motivation for participation (e.g. Füchslin et al. 2019, Vries et al. 2019), none of the respondents mentioned this. Also, motivations related to principlism, i.e. the upholding of some moral principle (Batson et al. 2002), were not mentioned. Etter (2020) found that such motivations were more relevant in a nature monitoring project than in a water level measurement project, highlighting that the subject and type of citizen science project also plays a role in the motivation of people to participate.

Fewer respondents mentioned self-interest motivations such as wanting to learn more about the water resources. Etter (2020) argues that learning is not as relevant in water monitoring projects, as there is less to learn from simply submitting water level data compared to, for example, identification of plants and animals. Furthermore, Aoki et al. (2017) found that participation out of concern for their own environment, which applies to the majority of the participants in our study, was a more important motivation in an air pollution project in the USA than the wish to learn about the environment. Nevertheless, people might expect to learn something from participating and the failure to fulfil this expectation might lead to low and short-term engagement of citizen scientists.

4.4.3 Why do participants withdraw?

Although the citizen science water monitoring project in the Sondu-Miriu basin managed to engage 155 people, only few of these participants were very active and kept on sending data for multiple years. Having a smaller group of very active contributors is not unusual in such projects (e.g. Domroese and Johnson 2017, Etter 2020), but tackling the challenges encountered by those who stop participating could boost the feasibility of sustainable citizen science-based data collection. Although citizens are able to participate and collect relevant data, they are not always motivated. Aoki et al. (2017) indicate that experiencing personal consequences from the environmental problem that is addressed by the project is more likely to motivate people than a more general environmental concern. Also intrinsic motivation, such as having an interest in the topic or willingness to learn, and the fulfilment of that expectation are very important for long-term commitment (Deci and Ryan 2000). Nevertheless, in our study, none of the motivations indicated by the respondents were characteristic for high or long-term engagement.

Awareness raising seemed to be important for long-term engagement of the participants in our project, as well as elsewhere (Hobbs and White 2012). The majority of the highly motivated and long-term engaged respondents in our study indicated they heard about the project through organized sensitization meetings. Although word of mouth is seen as an effective means to reach a wider audience (Johnson et al. 2014), only few respondents who were informed about the project by friends kept engaged for a long time. This suggests that simply knowing about the project and perhaps contributing a few messages is not sufficient to motivate volunteers long-term. The same applied to those who were informed through the instructional sign at the station, with 75% of this group of respondents sending only one message. Although the sign indicated that submitting water level information could help the community ("Support your community and take care of your water!"), the relevance was probably not clear enough to motivate people to continue sending data (Pocock et al. 2019). More than half of the respondents recommended more sensitization meetings even though no in-kind or monetary contribution was offered for participation. Additionally, they mentioned project feedback to participants and other project-related activities to encourage participation, which is a clear indication that active and continued communication with the volunteers is essential for a long-term project.

The relative simplicity of the measurement did not seem to form a barrier for long-term participation (Aoki et al. 2017). Although illiteracy could hinder participation, only one respondent indicated that further training was required. However, numerous respondents mentioned that readability of the gauge, vandalism and accuracy of the reading due to turbulence hindered participation (n=14) and three respondents mentioned these as reasons for stopping to participate. These respondents mainly fell in the low level of engagement class, suggesting that improved gauge maintenance could remove a barrier for long-term participation. On the other hand, 7 out of 20 medium or highly engaged respondents that experienced challenges mentioned lack of cellphone credit, although only three of these stopped participating for different reasons. The use of a toll-free number to submit measurements could address this challenge, as participants would not have to spend their own cellphone credit. At 1 KES (~ 0.01 USD) per message, this might be a barrier for participation by people from socio-economic deprived groups (Hobbs and White 2012).

A common reason for limited engagement of citizen scientist is a mismatch between data collection and the expectations that citizens have (Aoki et al. 2017, Etter 2020). Two respondents mentioned the expectation to be paid as a reason to participate, whereas four medium and highly engaged respondents indicated they stopped participating because they did not get paid. Furthermore, 18 respondents indicated that the project would be more successful if the volunteers would get paid, which goes against the principles of citizen science, whereby citizens voluntarily (i.e. without in-kind or monetary reward) participate in scientific activities. In addition to the expectation to be paid, participants might have gotten discouraged by the lack of other direct benefits. Those who hoped the project would lead to changes in the short-term, did not experience any change in water quality or supply as a consequence of improved management since the start of the project. Again, targeted and relevant communication could play a role here, as numerous studies found that citizen scientists appreciated communication of project findings more than receiving appreciation or recognition for their contribution (Alender 2016, Vries et al. 2019, Golumbic et al. 2020). Regular feedback through meetings or social media could keep participants updated about the impact of their contributions and help them to see why continuing sending data is important. This is supported by the feedback by some respondents who indicated that more motivation from authorities could help to increase participation in the citizen science project. WRUAs could play a big role in this, as they are most likely better embedded in local communities than high level authorities or international project staff. Also accessibility to the collected data is a good way to keep citizen scientists engaged (Vries et al. 2019). However, this is challenging in a setting whereby only few people have access to internet and in the absence of a suitable infrastructure (e.g. WRUA offices where data could be accessed). Nevertheless, a user-friendly platform to share data and inform participants could enhance the success of a citizen science project (Golumbic et al. 2020). Also showing appreciation through 'Thank you' messages, as was implemented in our project, could help citizen scientists to stay committed (Lowry et al. 2019, Vries et al. 2019).

4.5 Recommendations

Previous studies have shown that water level monitoring of sufficient spatial and temporal resolution can be achieved through citizen involvement, also in rural areas and low-income countries (Weeser et al. 2018, Weeser et al. 2019). In our study, we show that there are highly and long-term engaged citizens that are willing to participate, but there are still challenges to overcome. Long-term water level monitoring through citizen involvement does not necessarily require a few highly engaged citizens. A larger number of short-term participants or people with a low level of engagement could also make a valuable contribution. This is facilitated by the simplicity of the data collection method used in the project in the Sondu-Miriu basin and the fact that nothing but a simple mobile phone is required, especially since smartphone ownership in East Africa is still limited (Pocock et al. 2019). A toll-free number or reimbursement of cellphone credit used to submit data could lower the barrier for participation even further, and at the same time address some of the challenges mentioned by the respondents.

Based on the results of this study, sensitization meetings are a powerful means to reach out to the community and engage motivated volunteers. These meetings should be aimed at community members that frequently visit the site and are unlikely to move away for jobs or education. Those who depend on the river as source of water for domestic use or other activities (e.g. watering livestock) are also more likely to be concerned about their resource and have a higher incentive to participate. Specific targeting of WRUA members as existing community of people with an interest in water management is useful as well, as the project could address their needs (Golumbic et al. 2020). In general, active involvement of WRUAs in engaging volunteers and communicating results back to their members could increase the number of highly engaged volunteers. This requires recognition by the local and national water management authorities, who are there to support the WRUAs, as the establishment of WRUAs and development of subcatchment management plans is still in its infancy in many parts of Kenya. Embedding low-cost participatory approaches in water management practices can also empower the WRUAs, as it would give them a means to collect and access data which can help in the development of their subcatchment management plans. This would add a clear aim and benefit to all community members who depend on the local water resources, increase the awareness of the relevance of monitoring and thus motivate people to participate.

Appendix 4-1 Survey Sheet

Basic Data	
Telephone number:	Station:
Introduction	
Dear Madam/Sir, I'm an employee at Water resource authority in Kericho. You have participated in our volunteer water level observation network. Thank you very much for your commitment to protect our precious river. For an evaluation of this project we would like to ask you some questions.	
How were you informed about the project?	
	I participated in a sensitization meeting
	I read the sign nearby the bridge
	A friend informed me about
	The local administration informed me about the project
	A WRUA informed me about this project
	Other answer:
Why have you decided to participate?	
What do you think is the purpose of the data you send?	
	For prediction of floods and droughts
	For observing the amount of water in the river
	Other answer:
Do you still send data? If not, why have you stopped sending data?	

How often do you pass by the water level gauge?	
	Every day
	Once a week
	Once a
	Other answer:
How far is the water level gauge from your home?	
	I live nearby the gauge
	< 1 km
	< 2 km
	> 2 km
Do you use a smartphone or a normal phone?	
	Smartphone
	Normal phone
Did you face any challenges?	
What would you recommend that should be done to encourge more people to participate?	
Thank you very much for your feedback. Now we would like to ask you some domegraphic data.	
What is your age?	
What is your education level?	

What is your gender?	
	male
	female
Are you a WRUA member?	
	yes
	no
If not, have you heard about WRUAs before?	

Appendix 4-2

Number of respondents in each engagement class for different explanatory variables. The duration of engagement is based on whether the respondent continued sending data after June 2017. The level of engagement is based on the number of valid measurements contributed between April 2016 and June 2017 (Low = 0–1, Medium = 2–9, High = 10 or more). The contribution on measurements after June 2017 was seen as an indicator for long-term engagement.

		Level of engagement			Duration of engagement		
Variable	Class	Low	Medium	High	Short-term	Long-term	Total
Informed about project	Sensitization meeting	11	9	8	21	7	28
	Sign near gauge	31	8	2	34	7	41
	Friend	4	5	1	8	2	10
	WRUA	3	0	3	3	3	6
	Other	1	1	0	2	0	2
Passing by station	Daily	17	9	9	26	9	35
	Weekly	21	9	4	27	7	34
	Monthly or less	11	3	1	13	2	15
	No answer	1	2	0	2	1	3
Distance to station	<1 km	24	15	7	36	10	46
	1–2 km	6	6	2	12	2	14
	>2 km	19	2	5	20	6	26
	No answer	1	0	0	0	1	1
Type of phone	Basic phone	22	14	11	31	16	47
	Smartphone	27	9	2	36	2	38
	Both	0	0	1	0	1	1
	No answer	1	0	0	1	0	1
Age group	18–29	18	14	2	31	3	34

	30–49	24	8	10	28	14	42
	≥50	7	1	2	8	2	10
	No answer	1	0	0	1	0	1
Highest completed level of education	None	3	0	0	2	1	3
	Primary	6	10	9	19	6	25
	Secondary	19	5	5	20	9	29
	Higher	18	7	0	23	2	25
	No answer	4	1	0	4	1	5
WRUA membership	Yes	9	4	7	15	1	20
	No, but aware	16	7	3	20	6	26
	No, not aware	14	10	4	22	6	28
	No, no answer	11	2	0	11	2	13

References

Aceves-Bueno, E., A. S. Adeleye, D. Bradley, W. T. Brandt, P. Callery, M. Feraud, K. L. Garner, R. Gentry, Y. Huang, I. McCullough, I. Pearlman, S. A. Sutherland, W. Wilkinson, Y. Yang, T. Zink, S. E. Anderson, and C. Tague. 2015. Citizen Science as an Approach for Overcoming Insufficient Monitoring and Inadequate Stakeholder Buy-in in Adaptive Management: Criteria and Evidence. *Ecosystems* 18(3):493–506.

Adrian, R. J. 1991. Particle-Imaging Techniques for Experimental Fluid Mechanics. *Annual Review of Fluid Mechanics* 23(1):261–304.

Alender, B. 2016. Understanding volunteer motivations to participate in citizen science projects: a deeper look at water quality monitoring. *Journal of Science Communication* 15(3):A04.

Aoki, P., A. Woodruff, B. Yellapragada, and W. Willett. 2017. Environmental Protection and Agency: Motivations, Capacity, and Goals in Participatory Sensing. Pages 3138–3150 *in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.

Assumpção, T. H., I. Popescu, A. Jonoski, and D. P. Solomatine. 2017. Citizen observations contributing to flood modelling: Opportunities and challenges. *Hydrology and Earth System Sciences Discussions*:1–26.

Assumpção, T. H., I. Popescu, A. Jonoski, and D. P. Solomatine. 2018. Citizen observations contributing to flood modelling: opportunities and challenges. *Hydrology and Earth System Sciences* 22(2):1473–1489.

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

Avellaneda, P. M., D. L. Ficklin, C. S. Lowry, J. H. Knouft, and D. M. Hall. 2020. Improving hydrological models with the assimilation of crowdsourced data. *Water Resources Research* 56(5):e2019WR026325.

Bandini, F., M. Butts, T. V. Jacobsen, and P. Bauer-Gottwein. 2017. Water level observations from unmanned aerial vehicles for improving estimates of surface water-groundwater interaction. *Hydrological Processes* 31(24):4371–4383.

Batson, C. D., N. Ahmad, and J.-A. Tsang. 2002. Four Motives for Community Involvement. *Journal of Social Issues* 58(3):429–445.

Bell, S., D. Cornford, and L. Bastin. 2013. The state of automated amateur weather observations. *Weather* 68(2):36–41.

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

Beza, S. A., and M. A. Assen. 2016. Soil carbon and nitrogen changes under a long period of sugarcane monoculture in the semi-arid East African Rift Valley, Ethiopia. *Journal of Arid Environments* 132:34–41.

Binge, F. W. 1962. *Geology of the Kericho Area.* Geological Survey of Kenya 50.

Bonney, R., H. Ballard, R. Jordan, E. McCallie, T. B. Phillips, J. L. Shirk, and C. C. Wilderman. 2009. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. *Online Submission*. [online] URL: https://eric.ed.gov/?id=ED519688.

Bonney, R., J. L. Shirk, T. B. Phillips, A. Wiggins, H. L. Ballard, A. J. Miller-Rushing, and J. K. Parrish. 2014. Citizen science. Next steps for citizen science. *Science (New York, N.Y.)* 343(6178):1436–1437.

Brandt, P., E. Hamunyela, M. Herold, S. de Bruin, J. Verbesselt, and M. C. Rufino. 2018. Sustainable intensification of dairy production can reduce forest disturbance in Kenyan montane forests. *Agriculture, Ecosystems & Environment* 265:307–319.

Breiman, L. 2001. Random Forests. *Machine Learning* 45(1):5–32.

Breiman, L., A. Cutler, A. Liaw, and M. Wiener. 2018. Package 'randomForest'.

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

Buytaert, W., S. Baez, M. Bustamante, and A. Dewulf. 2012. Web-Based Environmental Simulation: Bridging the Gap between Scientific Modeling and Decision-Making. *Environmental Science & Technology* 46(4):1971–1976.

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

Buytaert, W., Z. Zulkafli, S. Grainger, L. Acosta, T. C. Alemie, J. Bastiaensen, B. De Bièvre, J. Bhusal, J. Clark, A. Dewulf, M. Foggin, D. M. Hannah, C. Hergarten, A. Isaeva, T. Karpouzoglou, B. Pandeya, D. Paudel, K. Sharma, T. Steenhuis, S. Tilahun, G. van Hecken, and M. Zhumanova. 2014. Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science* 2.

Chacon-Hurtado, J. C., L. Alfonso, and D. P. Solomatine. 2017. Rainfall and streamflow sensor network design: a review of applications, classification, and a proposed framework. *Hydrology and Earth System Sciences* 21(6):3071–3091.

Chase, S. K., and A. Levine. 2018. Citizen Science: Exploring the Potential of Natural Resource Monitoring Programs to Influence Environmental Attitudes and Behaviors. *Conservation Letters* 11(2):e12382.

Chaudhary, P., S. D'Aronco, M. Moy de Vitry, J. P. Leitão, and J. D. Wegner. 2019. Flood-Water level estimation from social media images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W5:5–12.

Cohen, A. 2016. *FuzzyWuzzy. Fuzzy string matching in python.* 0.14.0 edition.

Danielsen, F., N. D. Burgess, and A. Balmford. 2005. Monitoring Matters: Examining the Potential of Locally-based Approaches. *Biodiversity & Conservation* 14(11):2507–2542.

Davids, J. C., N. Devkota, A. Pandey, R. Prajapati, B. A. Ertis, M. M. Rutten, S. W. Lyon, T. A. Bogaard, and N. van de Giesen. 2019a. Soda Bottle Science—Citizen Science Monsoon Precipitation Monitoring in Nepal. *Frontiers in Earth Science* 7.

Davids, J. C., M. M. Rutten, A. Pandey, N. Devkota, W. D. van Oyen, R. Prajapati, and N. van de Giesen. 2019b. Citizen science flow – an assessment of simple streamflow measurement methods. *Hydrology and Earth System Sciences* 23(2):1045–1065.

Davids, J. C., N. van de Giesen, and M. M. Rutten. 2017. Continuity vs. the Crowd—Tradeoffs Between Continuous and Intermittent Citizen Hydrology Streamflow Observations. *Environmental Management* 60:12–29.

Deci, E. L., and R. M. Ryan. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11(4):227–268.

Defersha, M. B., and A. M. Melesse. 2012. Field-scale investigation of the effect of land use on sediment yield and runoff using runoff plot data and models in the Mara River basin, Kenya. *CATENA* 89:54–64.

Deutsch, W. G., and S. S. Ruiz-Córdova. 2015. Trends, challenges, and responses of a 20-year, volunteer water monitoring program in Alabama. *Ecology and Society* 20(3).

Domroese, M. C., and E. A. Johnson. 2017. Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation* 208:40–47.

Etter, S. 2020. *CrowdWater: Motivations of Citizen Scientists, the Accuracy and the Potential of Crowd-Based Data for Hydrological Model Calibration*. Dissertation. Universität Zürich, Zürich.

Etter, S., B. Strobl, J. Seibert, and H. J. I. van Meerveld. 2018. Value of uncertain streamflow observations for hydrological modelling. *Hydrology and Earth System Sciences* 22(10):5243–5257.

European Commission. 2013. *Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research*.

Everard, M. 2012. Safeguarding the provision of ecosystem services in catchment systems. *Integrated environmental assessment and management* 9(2):252–259.

Falcone, J. A., D. M. Carlisle, and L. C. Weber. 2010. Quantifying human disturbance in watersheds: Variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. *Ecological Indicators* 10(2):264–273.

Fienen, M. N., and C. S. Lowry. 2012. Social.Water—A crowdsourcing tool for environmental data acquisition. *Computers & Geosciences* 49:164–169.

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

Füchslin, T., M. S. Schäfer, and J. Metag. 2019. Who wants to be a citizen scientist? Identifying the potential of citizen science and target segments in Switzerland. *Public Understanding of Science* 28(6):652–668.

Fujita, I., M. Muste, and A. Kruger. 1998. Large-scale particle image velocimetry for flow analysis in hydraulic engineering applications. *Journal of Hydraulic Research* 36(3):397–414.

Getirana, A. C. V., M.-P. Bonnet, S. Calmant, E. Roux, O. C. Rotunno Filho, and W. J. Mansur. 2009. Hydrological monitoring of poorly gauged basins based on rainfall–runoff modeling and spatial altimetry. *Journal of Hydrology* 379(3-4):205–219.

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

Ginsburg, C., P. Bocquier, D. Beguy, S. Afolabi, O. Augusto, K. Derra, F. Odhiambo, M. Otiende, A. B. Soura, P. Zabre, M. White, and M. Collinson. 2016. Human capital on the move: Education as a determinant of internal migration in selected INDEPTH surveillance populations in Africa. *Demographic Research* 34(30):845–884.

Golumbic, Y. N., A. Baram-Tsabari, and B. Koichu. 2020. Engagement and Communication Features of Scientifically Successful Citizen Science Projects. *Environmental Communication* 14(4):465–480.

Gomani, M. C., O. Dietrich, G. Lischeid, H. Mahoo, F. Mahay, B. Mbilinyi, and J. Sarmett. 2010. Establishment of a hydrological monitoring network in a tropical African catchment: An integrated participatory approach. *Physics and Chemistry of the Earth, Parts A/B/C* 35(13-14):648–656.

Gosset, M., H. Kunstmann, F. Zougmore, F. Cazenave, H. Leijnse, R. Uijlenhoet, C. Chwala, F. Keis, A. Doumounia, B. Boubacar, M. Kacou, P. Alpert, H. Messer, J. Rieckermann, and J. Hoedjes. 2016. Improving Rainfall Measurement in Gauge Poor Regions Thanks to Mobile Telecommunication Networks. *Bulletin of the American Meteorological Society* 97(3):ES49-ES51.

Gura, T. 2013. Citizen science: Amateur experts. *Nature* 496(7444):259–261.

Hacker, E., A. Picken, and S. Lewis. 2017. Perceptions of Volunteering and Their Effect on Sustainable Development and Poverty Alleviation in Mozambique, Nepal and Kenya. Pages 53–73 *in* J. Butcher, and C. J. Einolf, editors. *Perspectives on Volunteering: Voices from the South*. Springer International Publishing, Cham.

Hannah, D. M., S. Demuth, H. A. J. van Lanen, U. Looser, C. Prudhomme, G. Rees, K. Stahl, and L. M. Tallaksen. 2011. Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes* 25(7):1191–1200.

Hargreaves, G. H., and Z. A. Samani. 1985. Reference Crop Evapotranspiration from Temperature. *Applied Engineering in Agriculture* 1(2):96–99.

Hildebrandt, A., S. Lacorte, and D. Barceló. 2006. Sampling of water, soil and sediment to trace organic pollutants at a river-basin scale. *Analytical and bioanalytical chemistry* 386(4):1075–1088.

Hindmarsh, A. C., P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software* 31(3):363–396.

Hobbs, S. J., and P. C. L. White. 2012. Motivations and barriers in relation to community participation in biodiversity recording. *Journal for Nature Conservation* 20(6):364–373.

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

Howe, J. 2006. The Rise of Crowdsourcing. *Wired Mag.*, 2006:1–4. [online] URL: https://www.wired.com/2006/06/crowds/.

Ifejika Speranza, C., and E. Bikketi. 2018. Engaging with Gender in Water Governance and Practice in Kenya. Pages 125–150 *in* C. Fröhlich, G. Gioli, R. Cremades, and H. Myrttinen, editors. *Water Security Across the Gender Divide*. Springer International Publishing, Cham.

ISRIC - World Soil Information. 2007. *Soil and Terrain Database for Kenya (KENSOTER) version 2.0*.

Jackson, R. B., S. R. Carpenter, C. N. Dahm, D. M. McKnight, R. J. Naiman, S. L. Postel, and S. W. Running. 2001. Water in a changing world. *Ecological Applications* 11(4):1027–1045.

Jacobs, S. R., L. Breuer, K. Butterbach-Bahl, D. E. Pelster, and M. C. Rufino. 2017. Land use affects total dissolved nitrogen and nitrate concentrations in tropical montane streams in Kenya. *Science of The Total Environment* 603-604:519–532.

Jacobs, S. R., E. Timbe, **B. Weeser**, M. C. Rufino, K. Butterbach-Bahl, and L. Breuer. 2018a. Assessment of hydrological pathways in East African montane catchments under different land use. *Hydrology and Earth System Sciences* 22(9):4981–5000.

Jacobs, S. R., **B. Weeser**, A. C. Guzha, M. C. Rufino, K. Butterbach-Bahl, D. Windhorst, and L. Breuer. 2018b. Using High-Resolution Data to Assess Land Use Impact on Nitrate Dynamics in East African Tropical Montane Catchments. *Water Resources Research*.

Jacobs, S. R., **B. Weeser**, M. C. Rufino, and L. Breuer. 2020. Diurnal Patterns in Solute Concentrations Measured with In Situ UV-Vis Sensors: Natural Fluctuations or Artefacts? *Sensors* 20(3).

Jehn, F. U., L. Breuer, T. Houska, K. Bestian, and P. Kraft. 2017. Incremental model breakdown to assess the multi-hypotheses problem. *Hydrology and Earth System Sciences Discussions*:1–22.

Jian, J., D. Ryu, J. F. Costelloe, and C.-H. Su. 2015. *Towards reliable hydrological model calibrations with river level measurements.* 21st International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand.

Jian, J., D. Ryu, J. F. Costelloe, and C.-H. Su. 2017. Towards hydrological model calibration using river level measurements. *Journal of Hydrology: Regional Studies* 10:95–109.

Jiang, S., V. Babovic, Y. Zheng, and J. Xiong. 2019. Advancing Opportunistic Sensing in Hydrology: A Novel Approach to Measuring Rainfall With Ordinary Surveillance Cameras. *Water Resources Research* 55(4):3004–3027.

Johnson, M. F., C. Hannah, L. Acton, R. Popovici, K. K. Karanth, and E. Weinthal. 2014. Network environmentalism: Citizen scientists as agents for environmental advocacy. *Global Environmental Change* 29:235–245.

Johnson, R. K., M. T. Furse, D. Hering, and L. Sandin. 2007. Ecological relationships between stream communities and spatial scale: implications for designing catchment-level monitoring programmes. *Freshwater Biology* 52(5):939–958.

Kavetski, D., and M. P. Clark. 2011. Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes* 25(4):661–670.

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

Kirchner, J. W. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* 42(3):2465.

Kraft, P., K. B. Vaché, H.-G. Frede, and L. Breuer. 2011. CMF: A Hydrological Programming Language Extension For Integrated Catchment Models. *Environmental Modelling & Software* 26(6):828–830.

Kraft, P., C. Weber, K. Bestian, F. U. Jehn, T. Houska, A. Karlson, and D. Windhorst. 2018. Cmf 1.4.

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

Le Blanc, D., and R. Perez. 2008. The relationship between rainfall and human density and its implications for future water stress in Sub-Saharan Africa. *Ecological Economics* 66(2):319–336.

Le Boursicaud, R., L. Pénard, A. Hauet, F. Thollet, and J. Le Coz. 2016. Gauging extreme floods on YouTube: application of LSPIV to home movies for the post-event determination of stream discharges. *Hydrological Processes* 30(1):90–105.

Le Coz, J., A. Patalano, D. Collins, N. F. Guillén, C. M. García, G. M. Smart, J. Bind, A. Chiaverini, R. Le Boursicaud, G. Dramais, and I. Braud. 2016. Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. *Journal of Hydrology* 541:766–777.

Liu, H.-Y., M. Kobernus, D. Broday, and A. Bartonova. 2014. A conceptual approach to a citizens' observatory—supporting community-based environmental governance. *Environmental health a global access science source* 13:107.

Lowry, C. S., and M. N. Fienen. 2013. CrowdHydrology: Crowdsourcing Hydrologic Data and Engaging Citizen Scientists. *Ground Water* 51(1):151–156.

Lowry, C. S., M. N. Fienen, D. M. Hall, and K. F. Stepenuck. 2019. Growing Pains of Crowdsourced Stream Stage Monitoring Using Mobile Phones: The Development of CrowdHydrology. *Frontiers in Earth Science* 7:36.

Lüthi, B., T. Philippe, and S. Peña-Haro. 2014. *Mobil Device App for Small Open-Channel Flow Measurement.* 7th International Congress on Environmental Modelling and Software, San Diego, California. [online] URL: https://pdfs.semanticscholar.org/fdef/e16d762c1ea1aab49ff41b610c625a388ffc.pdf.

Maier, N., L. Breuer, and P. Kraft. 2017. Prediction and uncertainty analysis of a parsimonious floodplain surface water-groundwater interaction model. *Water Resources Research* 53(9):7678–7695.

Mango, L. M., A. M. Melesse, M. E. McClain, D. Gann, and S. G. Setegn. 2011. Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: Results of a modeling study to support better resource management. *Hydrology and Earth System Sciences* 15:2245–2258.

Mazzoleni, M., V. J. Cortes Arevalo, U. Wehn, L. Alfonso, D. Norbiato, M. Monego, M. Ferri, and D. P. Solomatine. 2018. Exploring the influence of citizen involvement on the assimilation of crowdsourced observations: A modelling study based on the 2013 flood event in the Bacchiglione catchment (Italy). *Hydrology and Earth System Sciences* 22(1):391–416.

Mazzoleni, M., M. Verlaan, L. Alfonso, M. Monego, D. Norbiato, M. Ferri, and D. P. Solomatine. 2017. Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? *Hydrology and Earth System Sciences* 21(2):839–861.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21(2):239–245.

McKinley, D. C., A. J. Miller-Rushing, H. L. Ballard, R. Bonney, H. Brown, S. C. Cook-Patton, D. M. Evans, R. A. French, J. K. Parrish, T. B. Phillips, S. F. Ryan, L. A. Shanley, J. L. Shirk, K. F. Stepenuck, J. F. Weltzin, A. Wiggins, O. D. Boyle, R. D. Briggs, S. F. Chapin, D. A. Hewitt, P. W. Preuss, and M. A. Soukup. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation* 208:15–28.

Messer, H., A. Zinevich, and P. Alpert. 2006. Environmental monitoring by wireless communication networks. *Science (New York, N.Y.)* 312(5774):713.

Mishra, A. K., and P. Coulibaly. 2009. Developments in hydrometric network design: A review. *Reviews of Geophysics* 47(2).

Montanari, A., G. Young, H. H. G. Savenije, D. Hughes, T. Wagener, L. L. Ren, D. Koutsoyiannis, C. Cudennec, E. Toth, S. Grimaldi, G. Blöschl, M. Sivapalan, K. Beven, H. Gupta, M. Hipsey, B. Schaefli, B. Arheimer, E. Boegh, S. J. Schymanski, G. Di Baldassarre, B. Yu, P. Hubert, Y. Huang, A. Schumann, D. A. Post, V. Srinivasan, C. Harman, S. Thompson, M. Rogger, A. Viglione, H. McMillan, G. Characklis, Z. Pang, and V. Belyaev. 2013. "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022. *Hydrological Sciences Journal* 58(6):1256–1275.

Mu, Q., M. Zhao, and S. W. Running. 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment* 115(8):1781–1800.

Mutiga, J. K., S. T. Mavengano, S. Zhongbo, T. Woldai, and R. Becht. 2010. Water Allocation as a Planning Tool to Minimise Water Use Conflicts in the Upper Ewaso Ng'iro North Basin, Kenya. *Water Resources Management* 24(14):3939–3959.

Nare, L., D. Love, and Z. Hoko. 2006. Involvement of stakeholders in the water quality monitoring and surveillance system: The case of Mzingwane Catchment, Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C* 31(15):707–712.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* 10(3):282–290.

Newman, G., A. Wiggins, A. Crall, E. Graham, S. Newman, and K. Crowston. 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment* 10(6):298–304.

Njue, N., J. S. Kroese, J. Gräf, S. R. Jacobs, **B. Weeser**, L. Breuer, and M. C. Rufino. 2019. Citizen science in hydrological monitoring and ecosystem services management: State of the art and future prospects. *Science of The Total Environment*(639).

Nyenzi, B. S., P. M. R. Kiangi, and N. N. P. Rao. 1981. Evaporation values in East Africa. *Archives for Meteorology, Geophysics, and Bioclimatology Series B* 29(1-2):37–55.

Ochoa-Tocachi, B. F., W. Buytaert, J. Antiporta, L. Acosta, J. D. Bardales, R. Célleri, P. Crespo, P. Fuentes, J. Gil-Ríos, M. Guallpa, C. Llerena, D. Olaya, P. Pardo, G. Rojas, M. Villacís, M. Villazón, P. Viñas, and B. De Bièvre. 2018. High-resolution hydrometeorological data from a network of headwater catchments in the tropical Andes. *Scientific data* 5:180080.

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

Omonge, P., M. Herrnegger, G. Gathuru, J. Fürst, and L. Olang. 2020. Impact of development and management options on water resources of the upper Mara River Basin of Kenya. *Water and Environment Journal*.

Onyango, L., B. Swallow, J. Roy, and R. Meinzen-Dick. 2007. Coping with History and Hydrology: how Kenya's Settlement and Land Tenure Patterns Shape Contemporary Water Rights and Gender Relations in Water. Pages 173–195 *in* B. Koppen, M. Giordano, and J. Butterworth, editors. *Community-Based Water Law and Water Resource Management Reform in Developing Countries*. Centre for Agriculture and Biosciences International, Oxfordshire, UK.

Overdevest, C., C. Orr, and K. F. Stepenuck. 2004. Volunteer Stream Monitoring and Local Participation in Natural Resource Issues. *Research in Human Ecology* 11(2):177–185.

Overeem, A., J. C. R. Robinson, H. Leijnse, G. J. Steeneveld, B. K. P. Horn, and R. Uijlenhoet. 2013. Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters* 40(15):4081–4085.

Palmer Fry, B. 2011. Community forest monitoring in REDD+: the 'M' in MRV? *Environmental Science & Policy* 14(2):181–187.

Parrish, J. K., H. Burgess, J. F. Weltzin, L. Fortson, A. Wiggins, and B. Simmons. 2018. Exposing the Science in Citizen Science: Fitness to Purpose and Intentional Design. *Integrative and Comparative Biology* 58(1):150–160.

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

Phillips, T. B., H. L. Ballard, B. V. Lewenstein, and R. Bonney. 2019. Engagement in science through citizen science: Moving beyond data collection. *Science Education* 103(3):665–690.

Pocock, M. J. O., D. S. Chapman, L. J. Sheppard, and H. E. Roy. 2014. *A Strategic Framework to Support the Implementation of Citizen Science for Environmental Monitoring: Final Report to SEPA*, Wallingford, Oxfordshire. [online] URL: https://www.ceh.ac.uk/sites/default/files/hp1114final_5_complete.pdf.

Pocock, M. J. O., H. E. Roy, T. August, A. Kuria, F. Barasa, J. Bett, M. Githiru, J. Kairo, J. Kimani, W. Kinuthia, B. Kissui, I. Madindou, K. Mbogo, J. Mirembe, P. Mugo, F. M. Muniale, P. Njoroge, E. G. Njuguna, M. I. Olendo, M. Opige, T. O. Otieno, C. C. Ng'weno, E. Pallangyo, T. Thenya, A. Wanjiru, and R. Trevelyan. 2019. Developing the global potential of citizen science: Assessing opportunities that benefit people, society and the environment in East Africa. *Journal of Applied Ecology* 56(2):274–281.

Pool, S., D. Viviroli, and J. Seibert. 2017. Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? *Journal of Hydrology* 554:613–622.

Quinlivan, L., D. V. Chapman, and T. Sullivan. 2020. Applying citizen science to monitor for the Sustainable Development Goal Indicator 6.3.2: a review. *Environmental Monitoring and Assessment* 192(4):218.

Raddick, M. J., G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg. 2010. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review* 9(1).

Reges, H. W., N. Doesken, J. Turner, N. Newman, A. Bergantino, and Z. Schwalbe. 2016. CoCoRaHS: The evolution and accomplishments of a volunteer rain gauge network. *Bulletin of the American Meteorological Society*.

Republic of Kenya. 2019. *2019 Kenya Population and Housing Census. Volume IV: Distribution of Population by Socio-Economic Characteristics*, Nairobi, Kenya.

Richards, N., and D. Syallow. 2018. Water Resources Users Associations in the Mara Basin, Kenya: Pitfalls and Opportunities for Community Based Natural Resources Management. *Frontiers in Environmental Science* 6:138.

Rodda, J. C. 2001. Water under pressure. *Hydrological Sciences Journal* 46(6):841–854.

Rowntree, O., K. Bahia, H. Croxson, A. Delaporte, M. Meyer, M. Shanahan, and C. Sibthorpe. 2019. *GSMA Connected Women - The Mobile Gender Gap Report 2019*, London, UK.

Royem, A. A., C. K. Mui, D. R. Fuka, and M. T. Walter. 2012. Technical Note: Proposing a Low-Tech, Affordable, Accurate Stream Stage Monitoring System. *Transactions of the ASABE* 55(6):2237–2242.

Rufino, M. C., **B. Weeser**, Stenfert Kroese, N. Njue, J. Gräf, S. R. Jacobs, Z. Kemboi, A.M. Ran, P.O. Cerutti, C. Martius, and L. Breuer. 2018. *Citizen scientists monitor water quantity and quality in Kenya*. Center for International Forestry Research (CIFOR).

Ruhi, A., M. L. Messager, and J. D. Olden. 2018. Tracking the pulse of the Earth's fresh waters. *Nature Sustainability* 1(4):198–203.

Running, S. W., Q. Mu, M. Zhao, and A. Moreno. 2017. *User's Guide MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3) NASA Earth Observing System MODIS Land Algorithm*. [online] URL: http://files.ntsg.umt.edu/data/NTSG_Products/MOD16/MOD16UsersGuide2016_V1.5_2017Jul24.pdf.

Rutten, M. M., E. Minkman, and M. van der Sanden. 2017. How to get and keep citizens involved in mobile crowd sensing for water management? A review of key success factors and motivational aspects. *Wiley Interdisciplinary Reviews: Water* 4(4):e1218.

San Llorente Capdevila, A., A. Kokimova, S. Sinha Ray, T. Avellán, J. Kim, and S. Kirschke. 2020. Success factors for citizen science projects in water quality monitoring. *The Science of the total environment* 728:137843.

Seibert, J., and K. J. Beven. 2009. Gauging the ungauged basin: How many discharge measurements are needed? *Hydrology and Earth System Sciences* 13(6):883–892.

Seibert, J., B. Strobl, S. Etter, P. Hummer, and H. J. I. van Meerveld. 2019. Virtual Staff Gauges for Crowd-Based Stream Level Observations. *Frontiers in Earth Science* 7:18.

Seibert, J., and M. J. P. Vis. 2016. How informative are stream level observations in different geographic regions?: Value of Stream Level Observations. *Hydrological Processes* 30(14):2498–2508.

Seibert, J., M. J. P. Vis, E. Lewis, and H. J. I. van Meerveld. 2018. Upper and lower benchmarks in hydrological modeling. *Hydrological Processes*(32):1120–1125.

Senay, G. B., S. Leake, P. L. Nagler, G. Artan, J. Dickinson, J. T. Cordova, and E. P. Glenn. 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. *Hydrological Processes* 25(26):4037–4049.

Shirk, J. L., H. L. Ballard, C. C. Wilderman, T. B. Phillips, A. Wiggins, R. Jordan, E. McCallie, M. Minarchek, B. V. Lewenstein, M. E. Krasny, and R. Bonney. 2012. Public Participation in Scientific Research: a Framework for Deliberate Design. *Ecology and Society* 17(2).

Silvertown, J. 2009. A new dawn for citizen science. *Trends in Ecology & Evolution* 24(9):467–471.

Smith, L. C., B. L. Isacks, A. L. Bloom, and A. B. Murray. 1996. Estimation of Discharge From Three Braided Rivers Using Synthetic Aperture Radar Satellite Imagery: Potential Application to Ungaged Basins. *Water Resources Research* 32(7):2021–2034.

Starkey, E., G. Parkin, S. Birkinshaw, A. Large, P. Quinn, and C. Gibson. 2017. Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation. *Journal of Hydrology* 548:801–817.

Stephens, W., C. O. Othieno, and M.K.V. Carr. 1992. Climate and weather variability at the Tea Research Foundation of Kenya. *Agricultural and Forest Meteorology* 61(3-4):219–235.

Storey, R. G., A. Wright-Stow, E. Kin, R. J. Davies-Colley, and R. Stott. 2016. Volunteer stream monitoring: Do the data quality and monitoring experience support increased community involvement in freshwater decision making? *Ecology and Society* 21(4).

Strobl, B., S. Etter, H. J. I. van Meerveld, and J. Seibert. 2019. Accuracy of crowdsourced streamflow and stream level class estimates. *Hydrological Sciences Journal* 2(26):1–19.

Tetzlaff, D., S. K. Carey, J. P. McNamara, H. Laudon, and C. Soulsby. 2017. The essential value of long-term experimental data for hydrology and water management. *Water Resources Research* 53(4):2598–2604.

Tipaldo, G., and P. Allamano. 2016. 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*:e1200.

Toivanen, T., S. Koponen, V. Kotovirta, M. Molinier, and P. Chengyuan. 2013. Water quality analysis using an inexpensive device and a mobile phone. *Environmental Systems Research* 2(1):9.

Tweddle, J. C., L. D. Robinson, M. J. O. Pocock, and H. E. Roy. 2012. *Guide to citizen science: Developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK*. UK Environmental Observation Framework, [Swindon].

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

Unesco. 2015. *Water for a sustainable world includes data and indicators annex for water and energy*. Unesco, Paris.

United Nations. 2015. Transforming our world: The 2030 agenda for sustainable development. *General Assembley 70 session*.

United Nations. 2018. *SDG 6 Synthesis Report 2018 on Water and Sanitation*, Geneva, Switzerland. [online] URL: https://www.unwater.org/publications/highlights-sdg-6-synthesis-report-2018-on-water-and-sanitation-2/.

van de Giesen, N., R. Hut, and J. Selker. 2014. The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdisciplinary Reviews: Water* 1(4):341–348.

van Meerveld, H. J. I., M. J. P. Vis, and J. Seibert. 2017. Information content of stream level class data for hydrological model calibration. *Hydrology and Earth System Sciences* 21(9):4895–4905.

van Overloop, P. J., J. C. Davis, and M. M. Vierstra. 2014. *Mobile monitoring technologies: the mobiletracker and the remotetracker.* Proceedings of USCID Conference, Sacramento California.

Velpuri, N. M., G. B. Senay, R. K. Singh, S. Bohms, and J. P. Verdin. 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sensing of Environment* 139:35–49.

Vörösmarty, C., A. Askew, W. Grabs, R. G. Barry, C. Birkett, P. Doll, B. Goodison, A. Hall, R. Jenne, L. Kitaev, J. Landwehr, M. Keeler, G. Leavesley, J. Schaake, K. Strzepek, S. S. Sundarvel, K. Takeuchi, and F. Webster. 2001. Global water data: A newly endangered species. *Eos, Transactions American Geophysical Union* 82(5):54.

Vries, M. de, A. Land-Zandstra, and I. Smeets. 2019. Citizen Scientists' Preferences for Communication of Scientific Output: A Literature Review. *Citizen Science: Theory and Practice* 4(1):2.

Vuai, S. A. H., and N. Mungai. 2012. 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* 03(02).

Wagner, S., H. Kunstmann, A. Bárdossy, C. Conrad, and R. R. Colditz. 2009. Water balance estimation of a poorly gauged catchment in West Africa using dynamically downscaled meteorological fields and remote sensing information. *Physics and Chemistry of the Earth, Parts A/B/C* 34(4-5):225–235.

Walker, D., N. Forsythe, G. Parkin, and J. Gowing. 2016. Filling the observational void: Scientific value and quantitative validation of hydrometeorological data from a community-based monitoring programme. *Journal of Hydrology* 538:713–725.

Weeser, B., J. Gräf, N. Njue, P. Cerutti, M. C. Rufino, L. Breuer, and S. R. Jacobs. under review. Crowdsourced Water Level Monitoring in Kenya's Sondu-Miriu Basin – Who is "the crowd"? *Frontiers in Earth Science*.

Weeser, B., S. R. Jacobs, P. Kraft, M. C. Rufino, and L. Breuer. 2019. Rainfall-Runoff Modeling Using Crowdsourced Water Level Data. *Water Resources Research* 55(12):10856–10871.

Weeser, B., J. Stenfert Kroese, S. R. Jacobs, N. Njue, Z. Kemboi, A. Ran, M. C. Rufino, and L. Breuer. 2018. Citizen science pioneers in Kenya – A crowdsourced approach for hydrological monitoring. *Science of The Total Environment* 631-632:1590–1599.

Weisberg, H., J. A. Krosnick, and B. D. Bowen. 1996. *An introduction to survey research, polling, and data analysis*. SAGE Publications, Inc.

Whitelaw, G., H. Vaughan, B. Craig, and D. Atkinson. 2003. Establishing the Canadian Community Monitoring Network. *Environmental Monitoring and Assessment* 88(1):409–418.

Windhorst, D., P. Kraft, E. Timbe, H.-G. Frede, and L. Breuer. 2014. Stable water isotope tracing through hydrological models for disentangling runoff generation processes at the hillslope scale. *Hydrology and Earth System Sciences* 18(10):4113–4127.

Yan, K., G. Di Baldassarre, D. P. Solomatine, and G. J.-P. Schumann. 2015. A review of low-cost space-borne data for flood modelling: topography, flood extent and water level. *Hydrological Processes* 29(15):3368–3387.

Zheng, F., R. Tao, H. R. Maier, L. See, D. Savic, T. Zhang, Q. Chen, T. H. Assumpção, P. Yang, B. Heidari, J. Rieckermann, B. Minsker, W. Bi, X. Cai, D. P. Solomatine, and I. Popescu. 2018. Crowdsourcing Methods for Data Collection in Geophysics: State of the Art, Issues, and Future Directions. *Reviews of Geophysics* 56(4):698–740.

Acknowledgements

Throughout the writing of this dissertation, I received exceptional support and assistance. The research conducted was part of several joint research projects between the Centre for International Forestry Research (CIFOR) and the Justus Liebig University Giessen and was partially funded by the CGIAR program on Forest, Trees and Agroforestry led by CIFOR. Additional funding was supplied by the Deutsche Forschungsgemeinschaft (DFG) and the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). In addition, the assistance of the Kenyan Water Resources Authority (WRA) and of the Water Resources Users Associations (WRUAs) was greatly appreciated.

My sincere thanks go to my advisors Prof. Dr. Mariana Rufino and Prof. Dr. Lutz Breuer. Their expertise was invaluable in formulating the research questions and forming the methodology. Your insightful feedback and suggestions brought my work to a higher level. Thank you for always being there for me to answer questions and give advice in both a professional and personal context. Your trust enabled me to work with all the freedom to structure my work in the best feasible way. Special thanks go to Prof. Dr. Martin Petrick, who agreed to be the second supervisor of this dissertation.

This dissertation would not have been possible without the help of my colleague and friend Dr. Suzanne Jacobs, with whom I spent innumerable times in the field and who kept the project running on the ground over all the years. Not only did she provide endless support during work, be it in the field, or while discussing and analyzing the results, she also offered me hospitality as a friend during my stays in Kenya or accompanied me for delicious Indian dinners in Kericho. I would also like to thank Dr. Phillip Kraft for his substantial support while setting up and applying the model and the database and for all his help and explanations he provided to me during our coffee- or lunchbreaks.

Besides, I would like to acknowledge the students that help me with the tasks in the field, particularly Jaqueline Stenfert Kroese for her support while setting-up the crowdsourced monitoring network and Jan Gräf for assisting me in conducting the telephone survey. In addition, I would like to express my thanks to Naomi Njue who supported me with fieldwork during her own work conducted in the Sondu-Miriu River basin.

A further thank you goes do Dr. Matthias Höher, who supported me in his function as managing director of the Centre for International Development and Environmental Research (ZEU) in a lot of administrative tasks related to my work. Thank you, Ruth Strittmatter, Sabine Viehmann, and Petra Spamer-Trapp for helping me with administrative questions.

I am incredibly grateful for the unconditional support I received from my parents Dagmar and Thomas. You were always there for me, believed in me, and supported any of the ideas I ever came up with. This work is also in memory of my grandfather Heinz, who never got tired of teaching me the importance of our nature and who would have enjoyed philosophizing about the meaning of this work over a cup of tea from Kenya. My gratitude extends to other family and friends and their enthusiastic responses to my work including all my colleagues and friends from the Department of Landscape Ecology and Resources Management (ILR) and ZEU. Thank you for the pleasant time and the amusing discussions during all the lunch, coffee, and evening hours we had together Carla, Tobias, David, Uwe, Amir, Natalie, Amani, Florian, Lukas, Juliane, Ina, Pavlina, Nadine, Anja, Irmgard, Irina, Gracia, Konrad, …! I am truly grateful to all the friendly, warm, and helpful people I met, and that supported me and my work substantially!

Finally, I would like to thank Alicia for her unconditional love and support during and after work. The most important team.

Declaration

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an anti-plagiarism software to check my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen "Satzung der Justus-Liebig-Universität Giessen zur Sicherung guter wissenschaftlicher Praxis" in carrying out the investigations described in the dissertation.

Björn Weeser
Giessen, 16th of October 2020