PLOS Water

RESEARCH ARTICLE

Evaluating hydrological model performance using varying amounts of participatory monitoring water level data

Fabian Mitze[1*], **Suzanne Robin Jacobs**[1,2], **Lutz Breuer**[1,2], **Jazmin Campos Zeballos**[1,3], **Björn Weeser**[1]

1 Centre for International Development and Environmental Research, Justus Liebig University, Giessen, Germany, **2** Institute for Landscape Ecology and Resources Management, Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University, Giessen, Germany, **3** Faculty of Agriculture, University of Bonn, Bonn, Germany

* fabian.mitze@zeu.uni-giessen.de

Citation: Mitze F, Jacobs SR, Breuer L, Zeballos JC, Weeser B (2025) Evaluating hydrological model performance using varying amounts of participatory monitoring water level data. PLOS Water 4(9): e0000405. https://doi.org/10.1371/journal.pwat.0000405

Editor: Daniele Secci, University of Parma: Universita degli Studi di Parma, ITALY

Received: February 17, 2025

Accepted: September 10, 2025

Published: September 23, 2025

Data availability statement: All data and scripts and data used in the study can be obtained from Zenodo (https://doi.org/10.5281/zenodo.14772421).

Funding: This work was supported by the Kurt Eberhard Bode Foundation (grant number 0122401952022 to all authors) and the

Abstract

In the context of participatory monitoring projects in hydrology, the collection of water level data by laypersons is used as a simple and cost-effective alternative to automatic water level sensors, especially in poorly gauged catchments in remote areas of countries in the Global South. Such data can be used for the development of hydrological models to support water resources management. However, a common problem with participatory monitoring approaches is the irregularity of data collection and its decreasing frequency over time. Determining the amount and timing of data collection required for satisfactory model calibration is critical. To investigate this further, we examined daily water levels from a four-year project in western Kenya. We set up scenarios that represented datasets of different lengths and seasonal starting points for measurements. These scenarios were then used to calibrate a simple rainfall-runoff model. The data were supplemented with satellite data on evapotranspiration to improve the simulated water balance. The model runs were filtered using a water balance filter, and the Kling-Gupta-Efficiency (KGE) was used to compare model efficiencies. While a single month of water level data collected during the rainy seasons was sufficient to achieve good model performance (KGE ≥ 0.75), several months of data were required for simulations starting in the dry seasons. Similar results were found for the validation period, with lower overall model performance (KGE ≥ 0.6). Water level data collected during high flows generally led to an improvement in model performance compared to data collected during low flows. However, after a certain threshold, more water level data did not lead to further substantial model improvement. Based on the analysis of different datasets, this study demonstrated that short-term participatory monitoring programs that collect water level data during the wet season have the potential to provide sufficient input to calibrate a hydrological model.

Deutsche Forschungsgemeinschaft (DFG) (grant number BR2238/23-2 to all authors). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

1. Introduction

Climate and land use changes can greatly impact the hydrological cycle [1]. The consequences could threaten established water use pathways for both society and nature and affect water-related ecosystem services. Efforts to achieve the Sustainable Development Goals (SDGs) defined in the United Nations' Agenda 2030 [2], such as SDG 6 on ensuring water availability for all, could be compromised as a result. This could be a particular threat to countries in the Global South [3], which lack the resources to ensure that the goals are met [4–7].

Modelling catchment behavior provides a means of assessing the impacts of global change. This in turn enables the development of sustainable water management strategies. However, hydrological models require data to operate. Such data can be scarce, as data collection is costly. While low-income countries in particular suffer from inadequate data availability, a global decline in monitoring networks has been observed [5–7]. A recent analysis of the occurrence of river gauges showed that there is a strong bias towards an abundance of gauging stations in North America, Europe, Australia, Chile, Brazil, Japan and parts of India, and large, almost blank areas in Africa, Latin America and Asia [8].

As a result, alternative methods for collecting the necessary hydro-meteorological data have been increasingly explored recently, but the focus of these studies has been predominantly in high-income countries such as North America or Europe [7,9]. In addition to using remotely sensed data for meteorological and discharge data [10–13], studies have also investigated the use of cameras [14,15], social media [16], the mobile phone network [17], or private web-connected weather stations [18,19] to collect the required information.

Another alternative is the use of "citizen science", in which the general public is involved in scientific research in various ways [7,20]. Participatory monitoring (PM) can be considered as a subcategory of this, where laypersons only collect data, while data analysis remains with experts [21,22]. The involvement of the general public shows great potential and there has been increased research into whether PM data can improve the data availability in poorly gauged basins [23–26]. Potential challenges include the complexity of the measurements [27] or temporal and spatial coverage of PM data, which may vary due to different sampling designs and willingness to participate [9]. This may result in irregular measurement frequency and overall fewer measurements compared to, for example, automatic sensors.

While high-quality discharge measurements remain complex to carry out for the layperson, it has been shown that water level data in particular can be easily collected by citizens with high accuracy either using physical staff gauges [24,27,28] or virtual staff gauges using a smartphone application [29–31]. As discharge is directly related to water level, these data can also be used for hydrological modelling [25,29,32–34]. For example, Weeser et al. [25] were able to calibrate a simple rainfall-runoff model in a headwater catchment in western Kenya using one year of PM water level data. The PM data covered 75% of all days in the calibration period and were characterized by a low measurement error. This demonstrates a common feature of many PM projects focusing on water level measurements: participants measure less frequently and more irregularly compared to automatic sensors. However, the data is often of high quality [9].

Several studies have shown that model calibration does not necessarily require continuous data [25,34–36]. Therefore, it would be beneficial to determine the amount of data required to ensure reliable model calibration. Different studies have investigated the importance of spatial and temporal resolution of hydrological data for model calibration and validation and concluded that, while higher resolution can improve model results, this enhancement is not guaranteed [37–40]. Unusual events or high flow have been described as more valuable for calibration than base flow conditions [32,34,39,41,42]. Artificial reduction of datasets has shown that smaller subsets of the whole data set often provide enough information for satisfying model results [29,32,35,39,43–45]. It should be noted, however, that the results varied widely depending on the choice of model parameters [35]. Therefore, not all data in a time series contribute equally to determining the most behavioral model parameters for a given model structure.

As water level data are considered a valuable alternative when discharge data are not available [33,34], the same is true for PM water level data [25]. So far, only artificial PM data sets derived from automatic sensor data have been used for such analyses [29,46]. Hence, this study aims to test how many daily PM water level measurements are required to achieve a good calibration performance with a particular hydrological model applied in a small catchment located in Kenya. For this several scenarios were developed to represent different amounts of PM water level data at different times of the hydrological year. With this, a critical contribution can be made to the design of efficient future PM projects under similar conditions.

2. Materials and methods

2.1. Study area and data

During a four-year PM field campaign in the Sondu-Miriu River basin in western Kenya (0.35°S; 35.5°E WGS 1984), water level measurements were collected through PM at 13 gauging stations [24]. Kenya's largest indigenous closed-canopy forest, the Mau Forest, makes up a large proportion of the area [24]. This area provides critical water-related ecosystem services, such as water storage and river flow [47] which are threatened by continued forest loss and degradation [48]. This could potentially jeopardize the achievement of the SDG 6 targets in the region. For this study, we used data from the station with the highest temporal coverage of water level data (January 2016 – April 2020) covering a catchment area of 27.4 km^2 [25].

Most of the catchment is characterized by smallholder farms with a mix of annual crops, woodlands, forests, and pastures [25]. The soils in the study area are characterized as deep and well drained and fall under the classification of Humic Nitisols and Mollic Andosols [49]. Due to the movement of the Intertropical Convergence Zone, the rainfall pattern of the area is characterized by two rainy seasons per year [50].

Air temperature and precipitation were collected at 10-minute intervals from an automatic weather station (ECRN-100 high resolution rain gauge and VP-3 sensor, Decagon Devices, Pullman WA, USA) located 100 m northwest of the catchment outlet. In addition, precipitation was recorded at two other sites located in the center and upper parts of the catchment using tipping buckets (Theodor Friedrichs, Schenefeld, Germany). Area-weighted precipitation was calculated using Thiessen polygons excluding stations with data gaps by adjusting the weights of the remaining stations. Precipitation and air temperature were then aggregated in daily steps. Smaller gaps in the air temperature dataset up to seven days (1% of the data), were filled using linear interpolation. Two larger gaps from 18th June 2016–18th July 2016 and 12th to 31st March 2020 were filled by averaging values of the previous and following years. Daily minimum, maximum and mean temperatures plus the extraterrestrial radiation were used to calculate annual potential evapotranspiration (ET_{pot}) was calculated using the Hargreaves equation [51]. At the outlet, participants manually measured the water level using a physical water level gauge [24]. A signboard within sight of the water level gauge explained the monitoring and data submission process using a combination of pictures and instructions in English and

Swahili [24]. Measurements were sent by participants to a local server via SMS. The measurement campaign at this site resulted in 819 valid individual measurements (covering 56% of all days in the simulation period from April 2016 to April 2020). Two obvious outliers were removed.

To evaluate the quality of laypersons' measurements, water level was additionally measured every ten minutes with a radar sensor (VEGAPULS WL61, VEGA Grieshaber KG, Schiltach, Germany) 20 m upstream of the staff gauge. The radar-based water level data and the measurements by the participants show a very high correlation with Pearson r = 0.96 (S1 Table). Together with manual discharge measurements (n = 86) using the salt dilution method and an Acoustic Doppler Current Profiler (RiverSurveyor S5, SonTek, San Diego CA, USA), these values were used to develop a rating curve (Equation 1) to calculate discharge as described by Jacobs et al. [52].

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

Where Q = measured discharge ($m^3\ s^{-1}$) and h = water level (m).

Small data gaps, up to two days, in discharge and radar-based water level data were filled using linear interpolation (0.1% of the data). A larger gap of 32 days between 12th September 2019 and 14th October 2019 was filled using a random forest regression algorithm [53,54], using the open source Python package Scikit-learn [55]. This was possible because both the discharge and radar-based water level data and the PM water level data show a high degree of agreement ($\text{Pearson}_{q\text{-}wl} = 0.96$ and $\text{Pearson}_{wl\text{-}pm\_wl} = 0.96$). The hydro-meteorological data of the simulation period are shown in Fig 1.

2.2. Model setup

The rainfall-runoff model with daily time steps follows the concept of Weeser et al. [25], who used the Python package "Catchment Modelling Framework" (CMF) [56] to develop a 5-parameter lumped conceptual model (Fig 2) based on the understanding of the hydrological processes in the catchment described in Jacobs et al. [57].

Incoming precipitation either infiltrates into a storage box or forms direct runoff by saturation excess. From the storage box, water is released either to the outlet using the normalized water volume raised to a power (Equation 2) or to the atmosphere by evapotranspiration calculated based on potential evapotranspiration. Surface runoff and water discharging from the storage box sums up to streamflow.

$$q_{out} = Q_0 \left(\frac{V}{V_0} \right)^{\beta} \tag{2}$$

where q_{out} = outflow from storage box into outlet, Q_0 = reference runoff when storage contains the reference volume V_0, V = actual water volume stored in the box, V_0 = Reference volume of the box representing maximum storage capacity and β = kinematic flow curve shape exponent.

A set of a priori model parameter ranges (Table 1), originally defined by Weeser et al. [25] and based on expert knowledge from previous comparable model applications and exploratory data analysis was used for this study.

Model parameter uncertainties were evaluated using a Latin Hypercube-based approach [58] with 10^5 parameter sets using the open source Python package SPOTPY [59].

2.3. Model calibration and validation

The dataset was divided into a warm-up period (1st January 2016–31st March 2016), a three-year calibration period (1st April 2016–31st March 2019) and a one-year validation period (1st April 2019–31st March 2020) (Table 2).

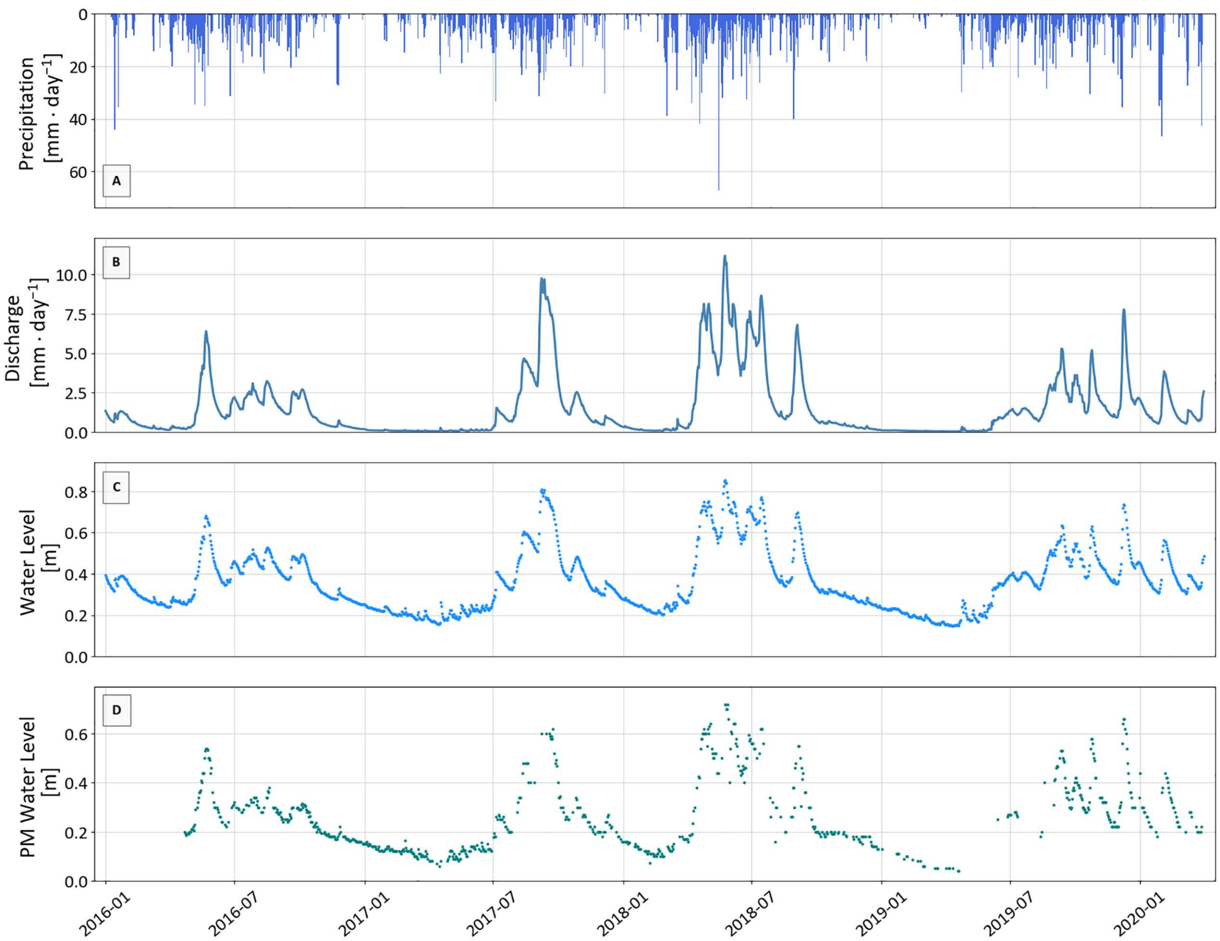

Fig 1. Measured hydro-meteorological data from January 2016 to April 2020: Daily mean areal weighted precipitation using Thiessen-Polygons (A), discharge (B), radar-based water level (C) and participatory monitoring (PM) water level (D).

https://doi.org/10.1371/journal.pwat.0000405.g001

The Kling-Gupta-Efficiency (KGE, Equation 3) [60] was used to evaluate model performance.

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \tag{3}$$

where r = linear correlation between observations and simulations, σ_{sim}= standard deviation in simulations, σ_{obs} = standard deviation in observations, μ_{sim} = simulation mean and μ_{obs} = observation mean.

KGE is a decomposition of the commonly used Nash-Sutcliffe-Efficiency (NSE) [61] to overcome problems such as volume balance errors in basins with high runoff variability [60]. To calibrate the model on water level data, the Spearman Rank correlation (R_{spear}) [62] was used [33] (S1 Table). R_{spear} ranges between -1 and 1 and reflects the similarity between the dynamics of observed discharge and water levels [33]. Since R_{spear} values close to 1 indicate a strictly monotonous relationship between observed discharge and water level, it does not ensure correct total discharge volumes [33]. Therefore, even a perfect fit of the model with R_{spear} = 1 using water level readings to calibrate the model may result in an over- or underestimation of the respective discharge [33]. This means that setting a threshold for behavioral parameter sets

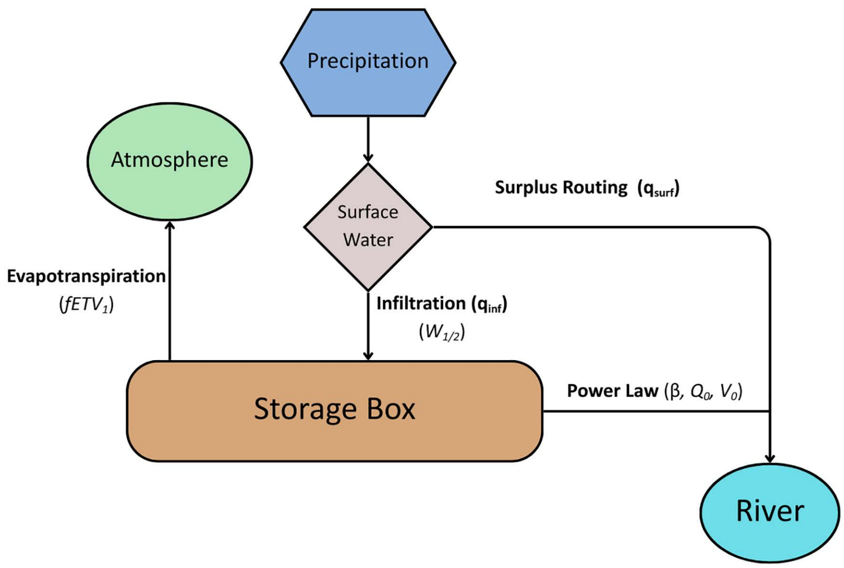

Fig 2. Schematic overview of the model structure, where CMF processes are written in bold and their corresponding parameters in italic letters (own illustration adapted from [25]).

https://doi.org/10.1371/journal.pwat.0000405.g002

Table 1. Model parameters and a priori ranges.

Name	Meaning	Unit	a priori range	
			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_0	Reference volume where storage runoff is equal to Q_0	–	100	3,000
$fETV_1$	Scaling factor for potential evapotranspiration	–	0.01	0.8
$W_{1/2}$	Saturation, where half of the catchment is saturated	–	0.1	0.9

https://doi.org/10.1371/journal.pwat.0000405.t001

Table 2. Averaged annual hydro-meteorological data for the calibration and validation period in the study area. ET_{pot} = potential evapotranspiration calculated using the Hargreaves equation.

Period	Discharge (mm)	Precipitation (mm)	Mean Air Temperature (°C)[a]	ET_{pot} (mm)
01 April 2016–31 March 2019	579	1,494	14.5	1,538
01 April 2019–31 March 2020	566	1,904	14.5	1,488

a) measured at the weather station at the catchment outlet, std = 1.45 (°C)

https://doi.org/10.1371/journal.pwat.0000405.t002

cannot be defined as in a KGE-based calibration. Instead, we ranked all model runs with regard to their R_{spear} and selected the top 0.5%, similar to Weeser et al. [25]. For the validation period, all accepted parameter sets from the calibration period were used.

2.4. Water balance filter

To compensate for possible over- or underestimation of the modelled discharge and to ensure that the models are not completely out of range, Weeser et al. [25] introduced an annual water balance filter. This was considered necessary, because R_{spear} does not reflect absolute discharge volumes [33].

The filter was applied according to the three-year calibration period (1st April 2016–31st March 2019). The balance was calculated from observed precipitation minus mean actual evapotranspiration (ET_{act}) retrieved from the MODerate Resolution Imaging Spectroradiometer (MODIS). An ET_{act} of 2,995 mm for the three years was calculated using the MOD16A2 Collection 6 Global Evapotranspiration Product [63]. This dataset provides evapotranspiration data in the form of scaled tiles, with an eight-day temporal resolution and a spatial resolution of 500 meters [63]. To calculate ET_{act}, the tiles within the three-year calibration period were clipped and averaged to the catchment area. As the values of the image tiles have been scaled, they need to be multiplied by a scaling factor of 0.1 in order to restore them to their original values. The values of the tiles are then summed up.

This MODIS-derived ET_{act} is comparable (+ 8.3%) to the ET_{act} sum of 2,745 mm for the three-year calibration period, which was calculated from the difference of measured precipitation and calculated runoff. Possible changes in water balance storage were considered to be small enough to be ignored [64].

A subjective confidence interval of +/-25% was added to account for possible measurement errors, storage changes and unknown uncertainties. This confidence interval is comparable to other studies [25,65,66]. Based on this assumption, the true ET_{act} therefore likely ranged from 2,245 mm to 3,742 mm. Subtracting this range from the measured precipitation of 4,482 mm resulted in a discharge range of 739 mm to 2,236 mm for the three years. Therefore, all model runs that resulted in total discharge values outside of this range were discarded.

2.5. Data availability scenarios

To be able to assess the influence of the duration and timing of data collection on model performance, we developed scenarios representing datasets of different length and seasonal starting points of measurements. To avoid the potential influence of the irregularity of the citizen science measurements, resulting in differences in the number of measurements per month over time, and to ensure comparability, we needed to reduce data asynchrony and establish a baseline PM dataset of daily measurements. To accomplish this, we took advantage of the high correlation between PM and radar-based water level data (Pearson r = 0.96). We used a random forest based regression model [53–55] to generate this baseline dataset with daily measurements. The model was trained with the available PM and radar-based water level data. We then used this model to predict the remaining 44% of days without PM measurements during the simulation period. The random forest regression model achieved a prediction accuracy of 0.96 (Fig 3). Given this, we consider the predicted values to be comparable to the original measurements provided by the participants.

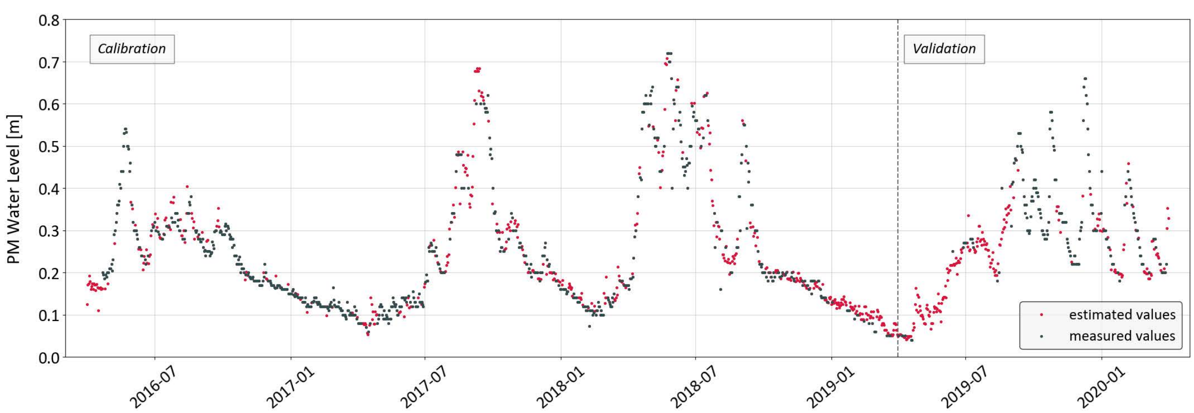

Fig 3. Baseline PM dataset of daily water level data in the simulation period (April 2016 – April 2020), showing the measured PM values and values estimated through random forest regression.

https://doi.org/10.1371/journal.pwat.0000405.g003

This baseline PM dataset was used to develop different scenarios of varying data availability to analyze the influence of varying amounts of water level data and the timing in the hydrological cycle on model credibility. Because high-flow conditions (higher discharge and higher water levels) appear to be more valuable for calibration than baseflow events [32,34,39,41,42], the scenarios were grouped based on seasonal weather conditions at the start of the dataset. For comparability and to limit random effects, six starting points, three in wet and three in dry conditions (one per season per year) were defined by analyzing rainfall, discharge and water level during the calibration period (Fig 4).

A wet period is characterized by frequent rainfall, high discharge and high water levels. In this study, the onset of a wet period was determined to begin when discharge exceeds 4 mm day^{-1}. Dry periods are characterized by minimal to no rainfall, low discharge and low water levels. The starting point was set following the last local peak in discharge (b: 0.47 mm day^{-1}, d: 0.91 mm day^{-1} and f: 0.36 mm day^{-1}) of a wet period (see Fig 4).

Data availability scenarios were developed by increasing the length of the dataset from each starting point (a-f) with monthly increments. For example, scenario a1 is a 1-month dataset starting at starting point a, while scenario c4 is a 4-month dataset start at starting point c. The number of scenarios per starting point corresponds to the maximum number of months from the respective starting point to the end of the calibration period (Table 3). For ease of reference scenarios based on a, c and e are also referred to below as “wet scenarios” because they start under wet conditions, while b, d and f are referred to as “dry scenarios” accordingly. Due to the increasing amount of PM water level data considered for each scenario, “wet scenarios” with more data may also include data from dry periods (lower water levels) and “dry scenarios” vice versa.

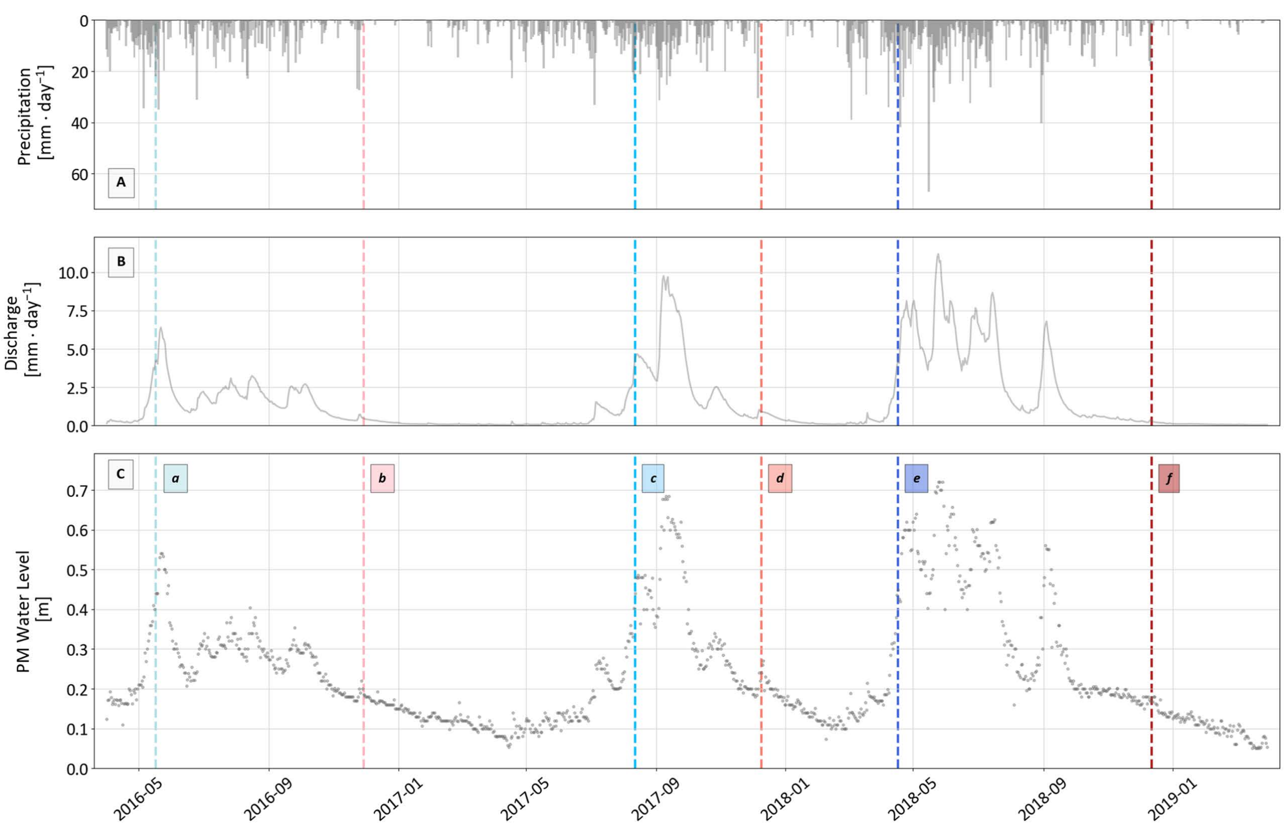

Fig 4. Overview of A) precipitation, B) discharge and C) baseline PM water level data in the calibration period (April 2016 – April 2019) with six different scenario starting points, n = 3 in wet seasons (a, c and e) and n = 3 in dry seasons (b, d and f).

https://doi.org/10.1371/journal.pwat.0000405.g004

Table 3. Scenario groups and number of scenarios based on the different starting points.

Scenario group	Starting month	Conditions	Number of scenarios
a	05/2016	wet	34
b	11/2016	dry	28
c	08/2017	wet	19
d	12/2017	dry	15
e	04/2018	wet	11
f	12/2018	dry	3
Total number of scenarios			**110**

https://doi.org/10.1371/journal.pwat.0000405.t003

A scenario was considered suitable for model calibration if KGE_{mean} (average of the 500 best KGE values within the applied water balance filter) ≥ 0.75 [60,67–69]. As a benchmark, the model was calibrated with the original PM dataset, the baseline PM dataset and the radar water level dataset. Calibration was carried out for these datasets with and without applying the water balance filter to assess the added value of the filter.

In order to compare the differences and similarities between the different scenario groups, the uncertainty of each model parameter and its contribution to the individual model performance (KGE) was analyzed.

3. Results

3.1. Comparison with original and radar-based dataset and impact of the water balance filter

Using the original PM dataset ($KGE_{mean} = 0.52$), the baseline PM dataset ($KGE_{mean} = 0.51$) and the radar-based water level dataset ($KGE_{mean} = 0.53$) for model calibration revealed similar model performances. By applying the water balance filter on these calibration schemes the model performance improved considerably for all schemes at a similar magnitude (original PM dataset = 0.73, PM baseline dataset = 0.75, radar-based dataset = 0.74). Considering the positive effect of the water balance filter and the marginal difference in calibration performance between the original and baseline dataset, only filtered scenario results (KGE and parameter values) are presented in the following sections, which are all based on the gap-filled baseline PM dataset.

3.2. Scenario calibration and validation

The KGE_{mean} of the scenarios showed noticeable differences with wet scenarios ranging from 0.51 to 0.85, while the dry ones ranged from 0.24 to 0.82 (Fig 5).

The mean values differed significantly between wet and dry scenarios with 0.75 (a), 0.81 (c) and 0.82 (e) compared to 0.58 (b), 0.66 (d) and 0.67 (f). Differences between the scenario groups were evaluated using with the Mann-Whitney U test [70], which showed a statistically significant difference between groups a-b ($p < 0.001$), c-d ($p < 0.01$) and e-f ($p = 0.048$).

Looking at the individual values for each scenario (Fig 6), it was observed that the wet scenarios already meet the condition ($KGE_{mean} \geq 0.75$) when calibrating on one month of available data (with a1 = 0.79, c1 = 0.83 and e1 = 0.80). The highest overall calibration efficiency was achieved with scenarios c5 and c6 (0.85), and the lowest with b7 (0.24).

Using more data did not necessarily improve model performance, which can be observed for all scenario groups (a), (c) and (e). For example, scenarios based on group (a) maintain model performances around 0.8 when calibrated on data for up to 10 months (a10). The subsequent decline coincides with the onset of the dry season and decreasing water levels. Results for scenarios groups (c) and (e) were more stable around 0.8 and showed smaller decreases when water levels fell around 12/2017 and 10/2018, respectively. Scenarios (b), (d) and (f), which started under dry conditions did not show such a clear trend. All scenario groups were characterized by different efficiencies with a substantial dynamic of increasing

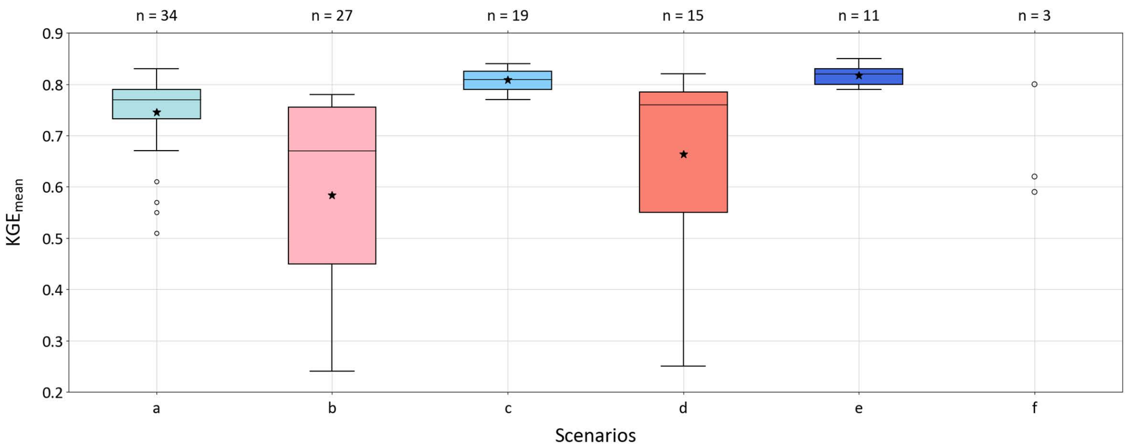

Fig 5. Boxplots comparing KGE_{mean} values of every scenario group (a-e) for calibration, where the black star is showing the mean. For scenario group f the individual KGE_{mean} values are shown due to the small sample size. Note: all model runs for scenario b1 were discarded, as none fulfilled the water balance filter.

https://doi.org/10.1371/journal.pwat.0000405.g005

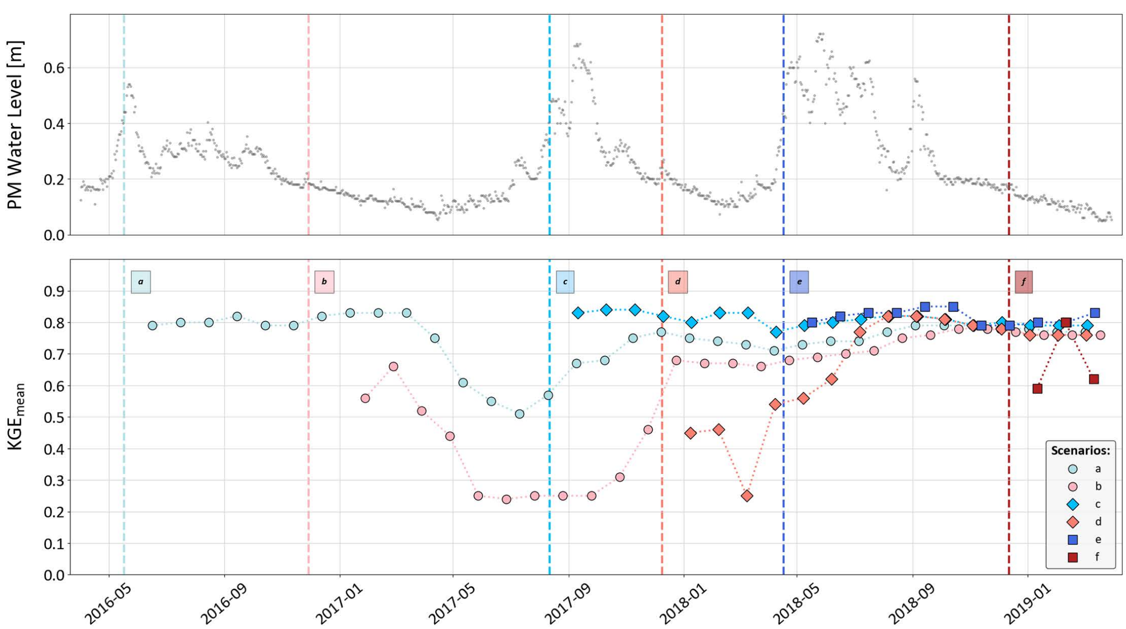

Fig 6. Results for each scenario calibration (lower panel) and the corresponding baseline PM water level data (upper panel). Each scenario used the baseline PM water levels from at the specific threshold (a-f) until the circle of the corresponding scenario. This circle represents the KGE_{mean} of each scenario. Note: model runs for scenario b1 were discarded due to not fulfillment of the water balance filter.

https://doi.org/10.1371/journal.pwat.0000405.g006

and decreasing values. Most striking was the long-lasting low model performance of scenario group (b), induced by the long dry spell from 12/2016–08/2017. Also noticeable are the outliers of the efficiency in d3 as well as f2. Overall, it can be observed that the model efficiency (for both wet and dry) saturates over time at around 0.76 - 0.79 and even long datasets (> 24 months) did not lead to a clear improvement of the model performance.

The validation results showed a similar pattern to the calibration results (Fig 7). While the overall KGE_{mean} values were lower, there was again a clear difference between wet and dry scenarios, although less pronounced than in the calibration.

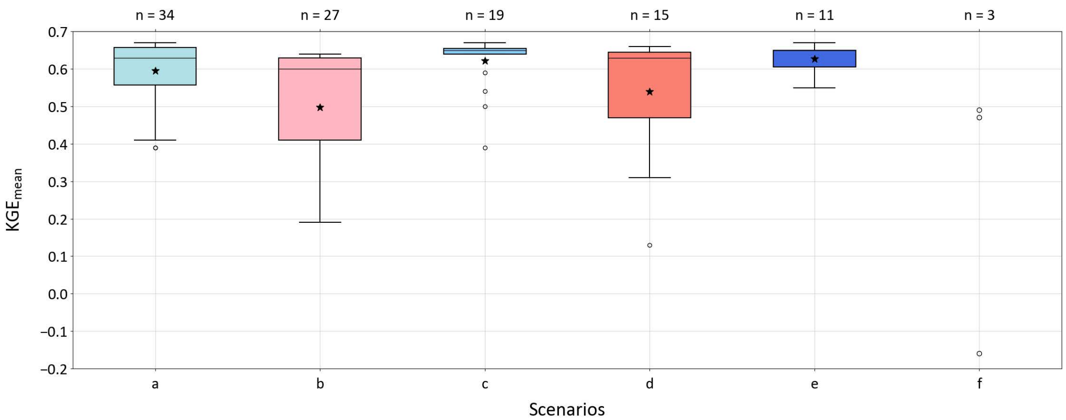

Fig 7. Boxplots comparing KGE_{mean} values of every scenario setting (a-e) for validation, where the filled black dot symbolizes the mean. For scenario group f the individual KGE_{mean} values are shown due to the small sample size. Note: As for calibration all model runs for scenario b1 were discarded due to not fulfillment of the water balance filter.

https://doi.org/10.1371/journal.pwat.0000405.g007

Wet scenarios ranged from 0.39 to 0.67, while dry scenarios ranged from -0.16 to 0.66. The means differed slightly less between wet and dry scenarios at 0.6 (a), 0.62 (c) and 0.63 (e) in comparison to 0.5 (b), 0.54 (d) and 0.27 (f). The Mann-Whitney U test again showed a statistically significant difference between the groups of a-b (p-value = 0.002), c-d (p-value = 0.03) and e-f (p-value = 0.01). The individual KGE_{mean} values of the scenarios confirm this (Fig 8).

Wet scenarios with few months of data again had higher values than the corresponding dry scenarios. Similar to the calibration, it can be observed that the KGE_{mean} values did not improve further after a certain point (12 months) and remained around 0.65. Again, an increasing amount of PM water level data could only lead to a significant improvement in model performance for the dry season scenarios.

3.3. Parameter uncertainty

To assess parameter uncertainty and their influence on model performance, individual parameter values of each scenario were plotted against their corresponding KGE values using dotty plots. Note that only filtered model parameters are displayed.

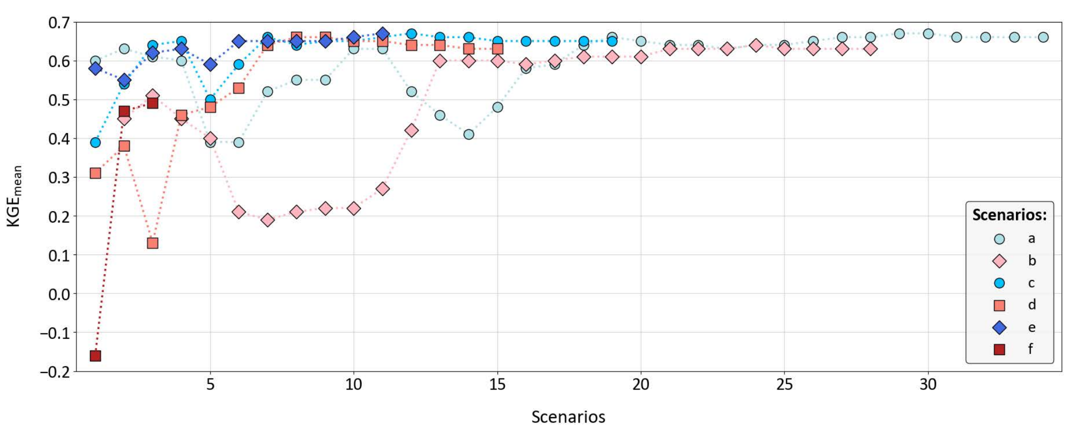

Fig 8. Validation results for the different scenario parameter sets and the resulting KGE_{mean} value. Note: model runs for scenario b1 were discarded due to not fulfillment of water balance filter.

https://doi.org/10.1371/journal.pwat.0000405.g008

For wet scenarios, parameters V_0 and $fETV_1$ show clear optimal ranges with 500–1000 mm and 0.1 to 0.2, respectively (Fig 9).

Contrasting, Q_0 and $W_{1/2}$ were not constrained at all. The behavior of β was somewhat intermediate, depicting a tendency towards constraint at values of 5–6. However, the patterns among the scenario families of (a, c, f) were somewhat different. Scenarios based on group (a) showed a generally larger variability of KGE going down as low as 0.1, while scenario families (c) and (e) outperformed (a) with KGE ranging from 0.6 to 0.95.

When comparing the parameter sensitivity of the dry scenarios (b, d, f), some similar patterns could be recognized, but there were also clear differences (Fig 10).

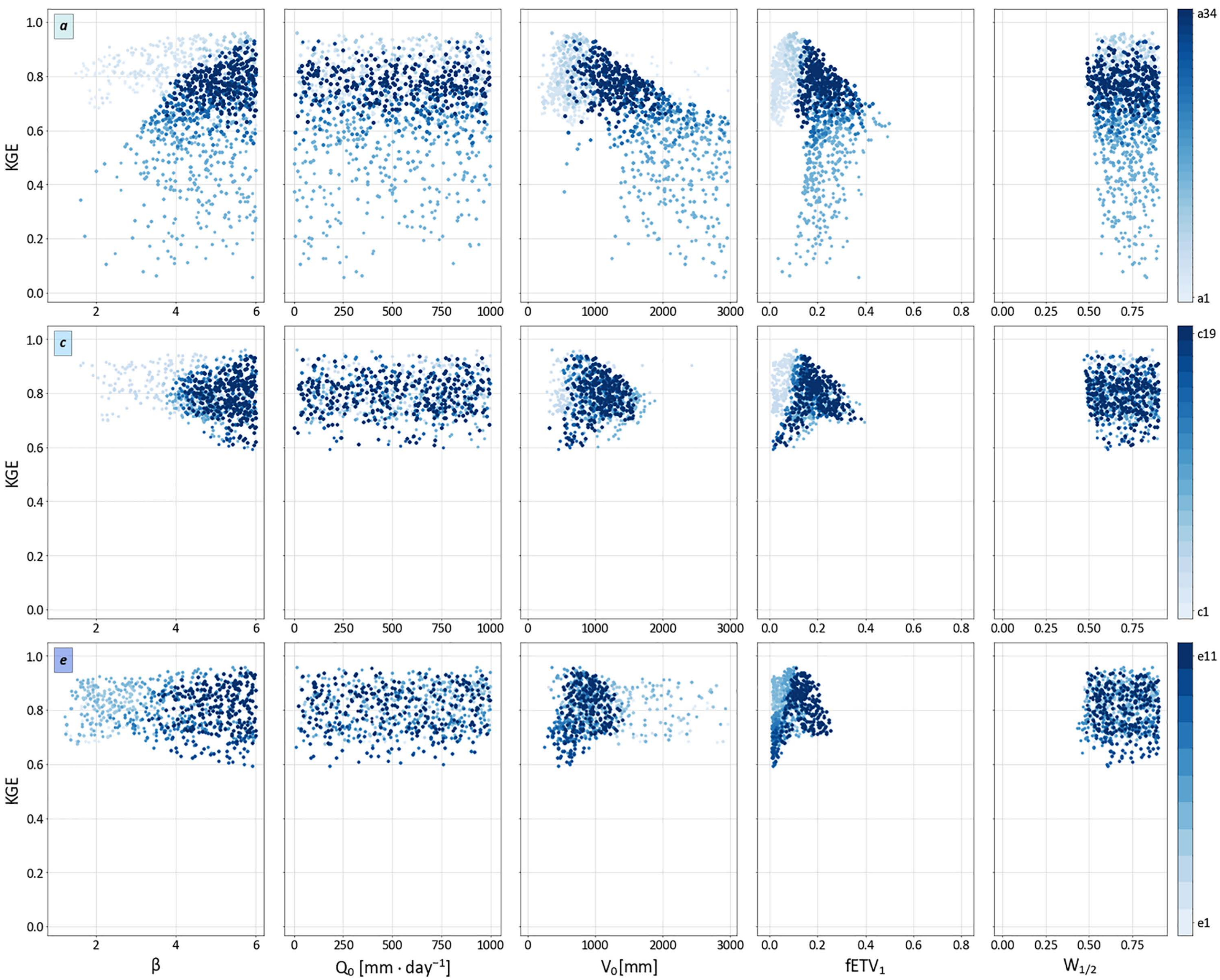

Fig 9. Parameter values and the corresponding KGE values of the filtered respective model runs for wet scenarios based on threshold a, c and e. Different shades indicate the different scenarios.

https://doi.org/10.1371/journal.pwat.0000405.g009

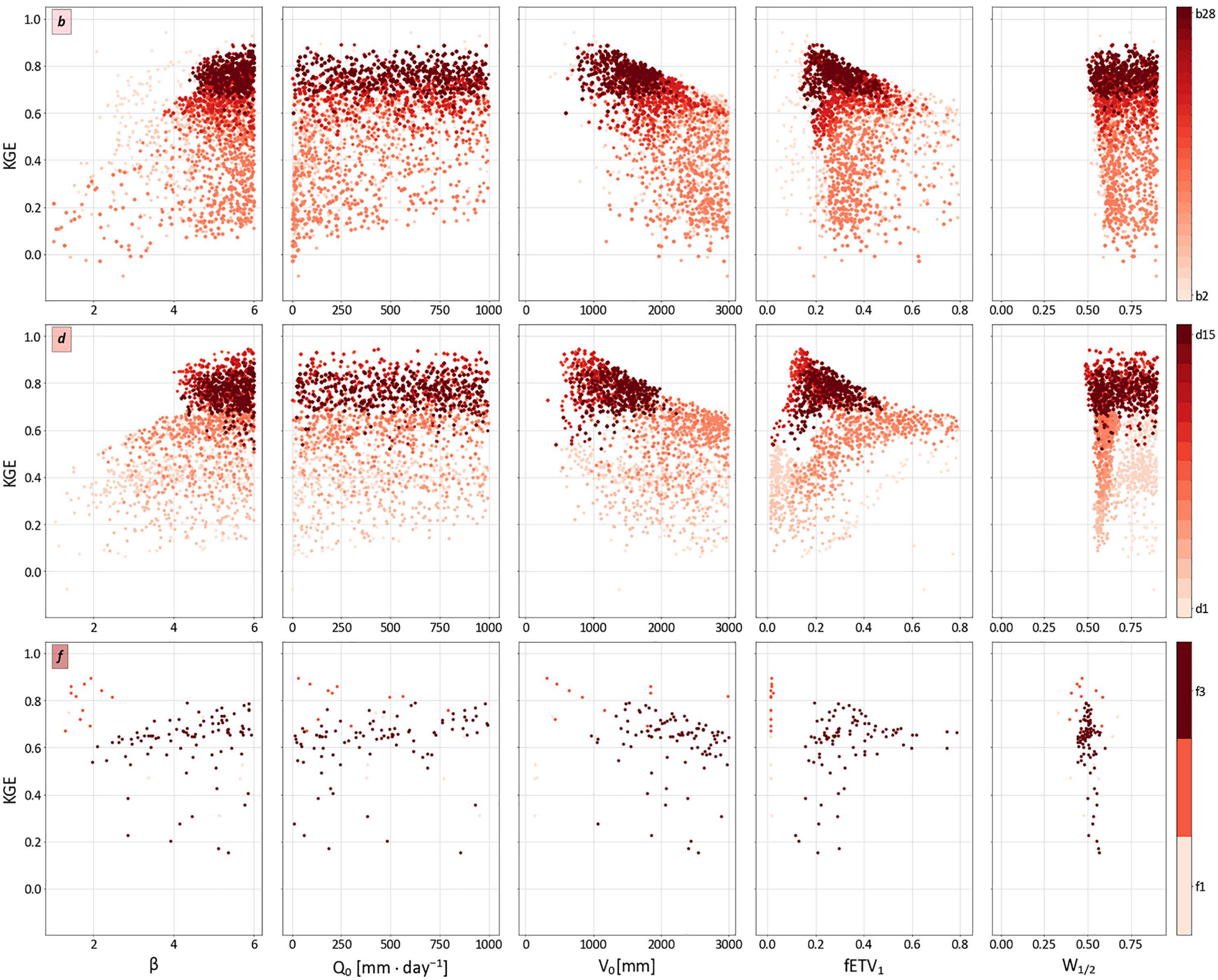

Fig 10. Parameter values and the corresponding KGE values of the filtered respective model runs for dry scenarios based on threshold b, d and f. Different shades indicate the different scenarios.

https://doi.org/10.1371/journal.pwat.0000405.g010

Most striking was the larger range of KGE for similar parameter values, while for the wet scenarios this was only found for (a). In general, dry scenarios which considered fewer PM water level data (light red), showed wider parameter ranges and possible KGE values, than those containing more data (dark red).

For scenario groups (b) and (d) parameter β showed a higher tendency towards constraint with higher values than scenarios (a, c, e). Parameter values for scenario group (f) did not show such a clear trend. For parameter Q_0, there was no trend at all, while for V_0 and $fETV_1$, low parameter values led to higher KGE. Scenario groups (f) only included medium (around 0.5) values for parameter $W_{1/2}$ and groups (b) and (d) only included higher values, but the corresponding KGE values covered a wide range. Scenario f2, which can be considered an outlier in model performance, showed significantly different parameter values than the other two scenarios of (f).

The general findings of the scenario analysis were also confirmed by the parameter uncertainties. For example, parameters for (a) show high KGE values for the first scenarios (a1 - a10), while the parameter ranges were much wider (including lower KGE values) for scenarios that contain the period 05/2017–09/2017 (e.g., a11 to a17). The parameters of the final scenarios (a18 - a34) clustered in a specific range of parameter values with higher associated KGE values. Parameters of dry scenarios with high KGE showed a similar accumulation. Scenario f2, on the other hand, had an outlier tendency, with clearly distinct parameter ranges compared to scenarios f1 and f3, which showed a distribution similar to those of the lower (b) and (d) scenarios.

4. Discussion

Based on the results of this study, it could be argued that short-term PM projects aimed at collecting water level data during the wet season can be sufficient for the development of reliable hydrological models. Although most PM studies in hydrology have been conducted in North America and Europe, this could also be a feasible approach for countries in the Global South [3,9]. There is also still considerable potential [71] for regions of the Global South, such as the study area, to use a PM approach to address subobjectives of SDG 6, such as 6.5 (Implement integrated water resources management), 6.a (Expand Water and sanitation support to developing countries) and 6.b (support local engagement in water and sanitation management) [2], as these countries might lack the resources to ensure it [4–7]. This work particularly represents an important contribution to subobjective 6.b, as it shows how such PM projects can help to involve the local community. Nevertheless, this will not fully address the overall shortage of hydrometeorological data in many of these countries.

4.1. How much PM data is required for model calibration?

The analyses of different scenarios of PM water level data for model calibration and validation provided valuable information on how much water level data is required for a well-calibrated hydrological model for applications in similar small tropical catchments. The results showed that a small number of daily measurements collected by laypersons can already be sufficient for high model performance if collected at the right time of the hydrological year. Based on our results, the best model performance is obtained when data is collected during a wet period with high flow conditions, i.e., high variability of water levels.

These general findings of our study are in line with previous studies that have investigated how many discharge or water level measurements are required for model calibration [29,32,34,35,39]. For example, Seibert & Beven [35] concluded that about 32 discharge measurements were sufficient for model calibration in several Swedish catchments, which strongly agrees with our results. In the case of Etter et al. [29], PM data were artificially calculated from real water level measurements in several Swiss catchments. Here, weekly values for a 1-year period were sufficient. In contrast, a study conducted in several catchments in eastern Australia using automatic sensor discharge data revealed a more complex picture. While for some winter rainfall dominant regions 20 years of data prior to the prediction lead to best results, catchments with more uniformly distributed rainfall data from individual months was sufficient [72].

4.2. Influence of data from dry and wet seasons

In particular, data from wet periods provide more valuable information for parameter selection because they cover more dynamics of the water cycle and represent different hydrological processes and catchment behavior. In our case study, only one month of data collected during wet periods was sufficient for good model performance. Similarly, Van Meerveld et al. [34] converted continuous discharge data from the United States into discrete water level classes and concluded that a few water level classes are sufficient for discharge simulation. For scenarios starting in the dry season, considerably more measurements were required to achieve similar results. However, adding more daily water level measurements was not automatically associated with better model performance. In fact, including more measurements taken during low flows actually worsened model performance. The reduced value of lower water levels (without strong dynamics) for model calibration was already pointed out by Seibert and Vis [33] who concluded that water levels alone may not be sufficient for model calibration in drier catchments.

The importance of high discharge or water level data for calibration has also been found by others [32,34,35,73]. For example, a study by Huang and Bardossy [73] investigated several catchments in the eastern United States and found a higher correlation between peak flow and high flow data and higher model performance. However, in contrast to our findings, they determined that approximately 10 years of data are required for calibration in order to achieve very good results for validation [73].

Jian et al. [32], identified that water level measurements at the 95th quantile are more suitable for identifying dynamics of discharge. Furthermore, a study by Rahbeh et al. [74], which analyzed a multi-year dataset, concluded that the use of discharge from wetter years leads to overall better results for hydrological model performance. Considering that other studies have come to the same conclusions about the importance of high flow conditions for better model performance, it can be assumed that our chosen thresholds for identifying the onset of wet and dry periods were appropriate. When comparing the results of the scenario groups with respect to which time period is included in the scenarios, general patterns in terms of the required amount of PM water level data for good model performance can be observed. While most of the scenarios based on wet starting points (a, c, e) show comparable high model performances, those including water level data from the period 05/2017–09/2017 experienced a sharp decrease in model performance, where long dry spells led to decreased or in general low model performance. After adding data from a period with higher water levels, like scenarios b11 – b28 or d3 - d15, a significant increase in model performance was experienced. The only outlier was f2, where a comparable calibration model performance to the wet scenarios was reached with two months of water level data from a dry season. However, this clarity was not evident in the validation with the parameters found for f2. A general substantial increase from the first to the second dry scenario was also found for b2 and d2, however, less pronounced and with and lower KGE_{mean} for calibration and validation.

The better model performance of scenarios where data collection started in the wet period compared to scenarios starting in the dry season found for model calibration were also confirmed for model validation, despite slightly lower overall KGE_{mean}. This lower model performance during the validation period can be mainly attributed to the overestimation of discharge at the beginning of the validation period between 05/2019 and 08/2019. It is likely that contribution of local precipitation recorded at individual rainfall stations during this period was overestimated by the Thiessen polygon method we used for spatial interpolation. The problem of scaling precipitation measured at individual points is a long-known problem in hydrological modeling [75,76]. However, we can still see that most of the measured discharge values fall within the uncertainty bands spanned by accepted $KGE_{mean} > 0.75$ (S1 and S2 Figs). Nevertheless, our results strongly indicate that water level measurements from wet seasons are more important for model performance than during dry seasons.

4.3. Model parameter uncertainty

A general analysis of the five model parameters revealed that the selected parameter ranges were within a space where a high KGE was achievable for all parameters. Most of them, except Q_0, showed a clear accumulation of good model performances at specific points within the selected ranges. When compared with other studies where one-dimensional dotty plots were used to assess parameter value distributions [77–79], similarities but also differences were found. For instance, Shin et al. [77] compared a four-parameter IHACRES model [80] and a nine-parameter SIMHYD model [81]. While the analysis for the four-parameter model showed clear optima for all parameters, the distributions for the nine-parameter model showed comparable distributions to our model, especially those for infiltration.

However, for our model it was noticeable that the parameters V_0 and $fETV_1$ showed a predominant accumulation in lower parameter values (V_0 = around 1000 mm and for $fETV_1$ between 0 and 0.25), while all other scenario groups showed a wider spread in possible parameter values. Also, parameter $W_{1/2}$ showed an accumulation of both, low and high model performances only for medium to higher parameter values (roughly 0.5 – 0.9). This indicates that model performance depends mainly on V_0, $fETV_1$ and to a certain degree on $W_{1/2}$, while Q0 and β are less constrained. As different parameter sets led to good model performance, any future model applications should also be carried out within a suitable uncertainty framework and consider uncertainty in the model outputs.

4.4. Utility of water balance filter

Since the calibration on water level data does not guarantee correct absolute discharge volumes [33] the application of a water balance filter is a simple way to avoid over- or underestimation of the modelled discharge. The water balance filter used in this study considerably improved the model performance (KGE_{mean}) of all scenarios by discarding simulations which were not within the set uncertainty of ±25%. However, no sensitivity analysis of the selected confidence interval has been carried out. Future research applying a water balance filter could investigate the influence of the confidence interval on the modelling results. The benefit of a simple water balance filter has already been observed in other studies where the filter was applied in a similar manner [25,82] or the hydrological model was calibrated directly on the remote sensing derived data [83].

A major advantage of posterior constraining models using a water balance filter is its ease of application. As stated by Weeser et al. [25], the filter does not affect the modelling, but subsequently only selects those model runs which are within the set boundaries. In general, the use of such a simple water balance filter could be a helpful tool for the analysis of other ungauged catchments, as it provides a seemingly accurate enough estimation of the water balance.

4.5. Research limitations

Our scenario approach used data from one catchment, which is similar to most other studies using PM water level data for hydrological modelling [25,84–86]. However, the complexity of our model should be taken into account when evaluating the scenario analysis. Outliers like scenario f2, where a high model performance (KGE_{mean} around 0.8) with two months of low water level data was achieved, could be a random result of the simple model structure. Testing the approach in other catchments with different characteristics is therefore recommended.

To increase comparability of the presented scenarios, the gaps in the available water level data were closed using the correlation with the radar-based measurements. In fact, the original data showed a mostly even distribution over the entire calibration period, even though PM measurements were available for less than half of the days. Considering the high quality of predictions and the marginal difference between the baseline calibration and the calibration using the original dataset, the potential impact of this step on the overall model performance can be disregarded.

5. Conclusions

This study analyzed the influence of PM data availability, expressed as length of the dataset and period of the hydrological year in which water level data collection started (dry vs. wet period), on the calibration and validation performance of a simple rainfall-runoff model. To increase comparability of the presented scenarios the gaps in the available water level data were closed using the correlation with the radar-based measurements. In fact, the original data showed a mostly even distribution over the entire calibration period, even though PM measurements were available for less than half of the days. Considering the high quality of predictions and the marginal difference between the baseline calibration and the calibration using the original dataset, the potential impact of this step on the overall model performance can be disregarded.

Results showed that one month of data collected during a wet (rainy) period is already sufficient to achieve good model efficiency, while the same amount of data collected during a dry period was less suitable. Including additional low water level measurements even reduced model efficiency. In most cases, scenarios where data collection started in the dry season required substantially more data to approach similar calibration model efficiencies compared to scenarios where data collection started in the wet season. Models also showed a saturation effect over time where additional data did not improve or decrease model efficiency, regardless of the starting period. Validation results support these observations but showed lower overall model performances.

In this model application, water level data measured by participants proved to be a reliable alternative to sensor-based discharge measurements for model calibration. Since the model relied on automatic sensor data for precipitation and air temperature, low-cost monitoring alternatives for these parameters could be investigated as a next step. Here, for example, the focus could be set on the possible measurement irregularity of precipitation and its suitability for hydrological modelling.

Summarized, calibrating hydrological models on PM water level data instead of automatic sensor discharge seems to be an economic and easy to establish way to analyze water flows. This is especially the case in poorly or ungauged catchments often found in the Global South. Local residents can contribute to monitoring SDG 6 and addressing sub indicator 6.b by participating in measurement campaigns. However, further confirmation from other catchments with more complex hydrological processes is recommended.

Supporting information

S1 Fig. Rainfall during calibration and validation period (upper panel), modelled discharge when calibrating and validating on scenario a1 (middle panel) and on b2 (lower panel). Shaded areas represent uncertainty of model simulations constrained by all simulations with accepted $KGE_{mean.}$
(TIF)

S2 Fig. Rainfall during calibration and validation period (upper panel), modelled discharge when calibrating and validating on scenario a28 (middle panel) and on b28 (lower panel). Shaded areas represent uncertainty of model simulations constrained by all simulations with accepted KGE_{mean}.
(TIF)

S1 Table. Statistical metrics used in the manuscript.
(DOCX)

Acknowledgments

We would like to express our thanks to the citizens of the Sondu-Miriu River basin who participated in our program, as well as Jaqueline Stenfert Kroese, Naomi Njue and Zacchaeus Kemboi for the implementation of the monitoring network on the ground.

Author contributions

Conceptualization: Fabian Mitze, Lutz Breuer, Björn Weeser.

Data curation: Fabian Mitze.

Formal analysis: Fabian Mitze, Björn Weeser.

Funding acquisition: Suzanne Robin Jacobs, Björn Weeser.

Methodology: Fabian Mitze, Lutz Breuer, Björn Weeser.

Project administration: Suzanne Robin Jacobs, Björn Weeser.

Supervision: Suzanne Robin Jacobs, Lutz Breuer, Björn Weeser.

Visualization: Fabian Mitze.

Writing – original draft: Fabian Mitze.

Writing – review & editing: Fabian Mitze, Suzanne Robin Jacobs, Lutz Breuer, Jazmin Campos Zeballos, Björn Weeser.

References

1. IPCC. Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 1st ed. Cambridge University Press; 2023. https://doi.org/10.1017/9781009325844
2. United Nations. Resolution adopted by the General Assembly on 25 September 2015: Transforming our world: The 2030 Agenda for Sustainable Development. 2015. Available from: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf
3. United Nations. UNCTAD Handbook of Statistics 2022. 2022. https://www.un-ilibrary.org/content/books/9789210021784
4. Daw T, Brown K, Rosendo S, Pomeroy R. Applying the ecosystem services concept to poverty alleviation: the need to disaggregate human well-being. Environ Conserv. 2011;38(4):370–9. https://doi.org/10.1017/s0376892911000506
5. Fankhauser S, McDermott TKJ. Understanding the adaptation deficit: Why are poor countries more vulnerable to climate events than rich countries? Glob Environ Change. 2014;27:9–18. https://doi.org/10.1016/j.gloenvcha.2014.04.014
6. Ruhi A, Messager ML, Olden JD. Tracking the pulse of the Earth's fresh waters. Nat Sustain. 2018;1(4):198–203. https://doi.org/10.1038/s41893-018-0047-7
7. Buytaert W, Zulkafli Z, Grainger S, Acosta L, Alemie TC, Bastiaensen J, et al. Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. Front Earth Sci. 2014;2. https://doi.org/10.3389/feart.2014.00026
8. Krabbenhoft CA, Allen GH, Lin P, Godsey SE, Allen DC, Burrows RM, et al. Assessing placement bias of the global river gauge network. Nat Sustain. 2022;5:586–92. https://doi.org/10.1038/s41893-022-00873-0 PMID: 36213515
9. Njue N, Stenfert Kroese J, Gräf J, Jacobs SR, Weeser B, Breuer L, et al. Citizen science in hydrological monitoring and ecosystem services management: state of the art and future prospects. Sci Total Environ. 2019;693:133531. https://doi.org/10.1016/j.scitotenv.2019.07.337
10. Gleason C, Durand M. Remote sensing of river discharge: a review and a framing for the discipline. Remote Sens. 2020;12(7):1107. https://doi.org/10.3390/rs12071107
11. Mu Q, Heinsch FA, Zhao M, Running SW. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens Environ. 2007;111(4):519–36. https://doi.org/10.1016/j.rse.2007.04.015
12. Schmugge TJ, Kustas WP, Ritchie JC, Jackson TJ, Rango A. Remote sensing in hydrology. Adv Water Resourc. 2002;25(8–12):1367–85. https://doi.org/10.1016/s0309-1708(02)00065-9
13. Xu X, Li J, Tolson BA. Progress in integrating remote sensing data and hydrologic modeling. Progr Phys Geogr. 2014;38:464–98.
14. Lee M-C, Leu J-M, Chan H-C, Huang W-C. The measurement of discharge using a commercial digital video camera in irrigation canals. Flow Measure Instrum. 2010;21(2):150–4. https://doi.org/10.1016/j.flowmeasinst.2010.02.002
15. Meier R, Tscheikner-Gratl F, Steffelbauer DB, Makropoulos C. Flow measurements derived from camera footage using an open-source ecosystem. Water. 2022;14(3):424. https://doi.org/10.3390/w14030424
16. Rosser JF, Leibovici D, Jackson M. Rapid flood inundation mapping using social media, remote sensing and topographic data. Nat Hazards. 2017;87:103–20.
17. Gosset M, Kunstmann H, Zougmore F, Cazenave F, Leijnse H, Uijlenhoet R, et al. Improving rainfall measurement in Gauge Poor Regions thanks to mobile telecommunication networks. Bull Am Meteorol Soc. 2016;97(3):ES49–51. https://doi.org/10.1175/bams-d-15-00164.1
18. de Vos L, Droste A, Zander M, Overeem A, Leijnse H, Heusinkveld B, et al. Opportunistic weather sensors: an Amsterdam case study of private weather stations, commercial microwave links and smartphones. 2020;10868. https://doi.org/10.5194/egusphere-egu2020-10868
19. Hahn C, Garcia-Marti I, Sugier J, Emsley F, Beaulant A-L, Oram L, et al. Observations from personal weather stations—EUMETNET interests and experience. Climate. 2022;10(12):192. https://doi.org/10.3390/cli10120192
20. Dickinson JL, Shirk J, Bonter D, Bonney R, Crain RL, Martin J, et al. The current state of citizen science as a tool for ecological research and public engagement. Front Ecol Environ. 2012;10(6):291–7. https://doi.org/10.1890/110236
21. Danielsen F, Burgess ND, Balmford A, Donald PF, Funder M, Jones JPG, et al. Local participation in natural resource monitoring: a characterization of approaches. Conserv Biol. 2009;23(1):31–42. https://doi.org/10.1111/j.1523-1739.2008.01063.x PMID: 18798859

22. Villaseñor E, Porter-Bolland L, Escobar F, Guariguata MR, Moreno-Casasola P. Characteristics of participatory monitoring projects and their relationship to decision-making in biological resource management: a review. Biodivers Conserv. 2016;25(11):2001–19. https://doi.org/10.1007/s10531-016-1184-9
23. Uprety M, Ochoa-Tocachi BF, Paul JD, Regmi S, Buytaert W. Improving water resources management using participatory monitoring in a remote mountainous region of Nepal. J Hydrol: Reg Stud. 2019;23:100604. https://doi.org/10.1016/j.ejrh.2019.100604
24. Weeser B, Stenfert Kroese J, Jacobs SR, Njue N, Kemboi Z, Ran A, et al. Citizen science pioneers in Kenya - a crowdsourced approach for hydrological monitoring. Sci Total Environ. 2018;631–632:1590–9. https://doi.org/10.1016/j.scitotenv.2018.03.130 PMID: 29727983
25. Weeser B, Jacobs S, Kraft P, Rufino MC, Breuer L. Rainfall-Runoff modeling using crowdsourced water level data. Water Resourc Res. 2019;55(12):10856–71. https://doi.org/10.1029/2019wr025248
26. Ochoa Tocachi B. Participatory hydrological monitoring to support sustainable water resources management. Imperial College London; 2019. https://doi.org/10.25560/75557
27. Lowry CS, Fienen MN, Hall DM, Stepenuck KF. Growing pains of crowdsourced stream stage monitoring using mobile phones: the development of crowdhydrology. Front Earth Sci. 2019;7. https://doi.org/10.3389/feart.2019.00128
28. Fienen MN, Lowry CS. Social.Water—A crowdsourcing tool for environmental data acquisition. Comput Geosci. 2012;49:164–9. https://doi.org/10.1016/j.cageo.2012.06.015
29. Etter S, Strobl B, Seibert J, van Meerveld HJI. Value of crowd-based water level class observations for hydrological model calibration. Water Resourc Res. 2020;56(2). https://doi.org/10.1029/2019wr026108
30. Scheller M, van Meerveld I, Seibert J. How well can people observe the flow state of temporary streams? Front Environ Sci. 2024;12. https://doi.org/10.3389/fenvs.2024.1352697
31. Seibert J, van Meerveld HJ, Etter S, Strobl B, Assendelft R, Hummer P. Wasserdaten sammeln mit dem Smartphone – Wie können Menschen messen, was hydrologische Modelle brauchen? Hydrologie Wasserbewirtschaftung. 2019;63:74–84. https://doi.org/10.5675/HyWa_2019.2_1
32. Jian J, Ryu D, Costelloe JF, Su C-H. Towards hydrological model calibration using river level measurements. J Hydrol: Reg Stud. 2017;10:95–109. https://doi.org/10.1016/j.ejrh.2016.12.085
33. Seibert J, Vis MJP. How informative are stream level observations in different geographic regions? Hydrological Processes. 2016;30(14):2498–508. https://doi.org/10.1002/hyp.10887
34. van Meerveld HJI, Vis MJP, Seibert J. Information content of stream level class data for hydrological model calibration. Hydrol Earth Syst Sci. 2017;21(9):4895–905. https://doi.org/10.5194/hess-21-4895-2017
35. Seibert J, Beven KJ. Gauging the ungauged basin: how many discharge measurements are needed? Hydrol Earth Syst Sci. 2009;13(6):883–92. https://doi.org/10.5194/hess-13-883-2009
36. McIntyre NR, Wheater HS. Calibration of an in-river phosphorus model: prior evaluation of data needs and model uncertainty. J Hydrol. 2004;290(1–2):100–16. https://doi.org/10.1016/j.jhydrol.2003.12.003
37. Ficchì A, Perrin C, Andréassian V. Impact of temporal resolution of inputs on hydrological model performance: an analysis based on 2400 flood events. J Hydrol. 2016;538:454–70. https://doi.org/10.1016/j.jhydrol.2016.04.016
38. Huang Y, Bárdossy A, Zhang K. Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data. Hydrol Earth Syst Sci. 2019;23(6):2647–63. https://doi.org/10.5194/hess-23-2647-2019
39. Juston J, Seibert J, Johansson P. Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. Hydrol Process. 2009;23(21):3093–109. https://doi.org/10.1002/hyp.7421
40. Krebs G, Kokkonen T, Valtanen M, Setälä H, Koivusalo H. Spatial resolution considerations for urban hydrological modelling. J Hydrol. 2014;512:482–97. https://doi.org/10.1016/j.jhydrol.2014.03.013
41. Singh SK, Liang J, Bárdossy A. Improving the calibration strategy of the physically-based model WaSiM-ETH using critical events. Hydrol Sci J. 2012;57(8):1487–505. https://doi.org/10.1080/02626667.2012.727091
42. Singh SK, Bárdossy A. Calibration of hydrological models on hydrologically unusual events. Adv Water Resourc. 2012;38:81–91. https://doi.org/10.1016/j.advwatres.2011.12.006
43. McIntyre N, Lee H, Wheater H, Young A, Wagener T. Ensemble predictions of runoff in ungauged catchments. Water Resourc Res. 2005;41(12). https://doi.org/10.1029/2005wr004289
44. Perrin C, Oudin L, Andreassian V, Rojas-Serna C, Michel C, Mathevet T. Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. Hydrological Sciences Journal. 2007;52(1):131–51. https://doi.org/10.1623/hysj.52.1.131
45. Seibert J, McDonnell JJ. Gauging the Ungauged Basin: relative value of soft and hard data. J Hydrol Eng. 2015;20(1). https://doi.org/10.1061/(asce)he.1943-5584.0000861
46. Mazzoleni M, Verlaan M, Alfonso L, Monego M, Norbiato D, Ferri M, et al. Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? Hydrol Earth Syst Sci. 2017;21(2):839–61. https://doi.org/10.5194/hess-21-839-2017
47. Benn J, Bindra S. UNEP annual report 2010 a year in review. Nairobi: United Nations Environmental Programme (UNEP); 2011.

48. Brandt P, Hamunyela E, Herold M, de Bruin S, Verbesselt J, Rufino MC. Sustainable intensification of dairy production can reduce forest disturbance in Kenyan montane forests. Agric Ecosyst Environ. 2018;265:307–19. https://doi.org/10.1016/j.agee.2018.06.011
49. ISRIC. Soil and Terrain Database for Kenya (KENSOTER), version 2.0. In: ISRIC Data Hub [Internet]. 2007 [cited 14 Feb 2025]. Available from: https://data.isric.org/geonetwork/srv/api/records/73e27136-9efe-49e4-af35-fd98b841d467
50. Hills RC. The Structure of the Inter-Tropical Convergence Zone in Equatorial Africa and Its Relationship to East African Rainfall. Trans Inst Br Geogr. 1979;4(3):329. https://doi.org/10.2307/622055
51. George GH, Zohrab SA. Reference crop evapotranspiration from temperature. Appl Eng Agric. 1985;1(2):96–9. https://doi.org/10.13031/2013.26773
52. Jacobs SR, Weeser B, Guzha AC, Rufino MC, Butterbach-Bahl K, Windhorst D, et al. Using high-resolution data to assess land use impact on nitrate dynamics in East African tropical montane catchments. Water Resour Res. 2018;54(3):1812–30. https://doi.org/10.1002/2017wr021592
53. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/a:1010933404324
54. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3–42. https://doi.org/10.1007/s10994-006-6226-1
55. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
56. Kraft P, Vaché KB, Frede H-G, Breuer L. CMF: a hydrological programming language extension for integrated catchment models. Environ Modell Softw. 2011;26(6):828–30. https://doi.org/10.1016/j.envsoft.2010.12.009
57. Jacobs SR, Timbe E, Weeser B, Rufino MC, Butterbach-Bahl K, Breuer L. Assessment of hydrological pathways in East African montane catchments under different land use. Hydrol Earth Syst Sci. 2018;22(9):4981–5000. https://doi.org/10.5194/hess-22-4981-2018
58. McKay MD, Beckman RJ, Conover WJ. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 1979;21(2):239–45. https://doi.org/10.1080/00401706.1979.10489755
59. Houska T, Kraft P, Chamorro-Chavez A, Breuer L. SPOTting model parameters using a ready-made python package. PLoS One. 2015;10(12):e0145180. https://doi.org/10.1371/journal.pone.0145180 PMID: 26680783
60. Gupta HV, Kling H, Yilmaz KK, Martinez GF. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J Hydrol. 2009;377(1–2):80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
61. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I — a discussion of principles. J Hydrol. 1970;10(3):282–90. https://doi.org/10.1016/0022-1694(70)90255-6
62. Spearman C. The proof and measurement of association between two things. Am J Psychol. 1904;15(1):72. https://doi.org/10.2307/1412159
63. Running SW, Mu Q, Zhao M. MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V061. NASA EOSDIS Land Processes DAAC; 2021. https://doi.org/10.5067/MODIS/MOD16A2.061
64. Senay GB, Leake S, Nagler PL, Artan G, Dickinson J, Cordova JT, et al. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. Hydrol Process. 2011;25(26):4037–49. https://doi.org/10.1002/hyp.8379
65. Alemayehu T, Griensven A van, Senay GB, Bauwens W. Evapotranspiration mapping in a heterogeneous landscape using remote sensing and global weather datasets: application to the Mara Basin, East Africa. Remote Sens. 2017;9(4):390. https://doi.org/10.3390/rs9040390
66. Velpuri NM, Senay GB, Singh RK, Bohms S, Verdin JP. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. Remote Sens Environ. 2013;139:35–49. https://doi.org/10.1016/j.rse.2013.07.013
67. Clark MP, Vogel RM, Lamontagne JR, Mizukami N, Knoben WJM, Tang G, et al. The abuse of popular performance metrics in hydrologic modeling. Water Resour Res. 2021;57(9). https://doi.org/10.1029/2020wr029001
68. Knoben WJM, Freer JE, Woods RA. Technical note: inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrol Earth Syst Sci. 2019;23(10):4323–31. https://doi.org/10.5194/hess-23-4323-2019
69. Moriasi DN, Arnold JG, Liew MWV, Bingner RL, Harmel RD, Veith TL. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transac ASABE. 2007;50:885–900. https://doi.org/10.13031/2013.23153
70. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist. 1947;18(1):50–60. https://doi.org/10.1214/aoms/1177730491
71. Sachs JD, Lafortune G, Fuller G. The SDGs and the UN Summit of the Future. Sustainable Development Report 2024. Dublin: Dublin University Press; 2024. https://doi.org/10.25546/108572
72. Luo J, Wang E, Shen S, Zheng H, Zhang Y. Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity. Hydrol Process. 2012;26(26):3953–61. https://doi.org/10.1002/hyp.8420
73. Huang Y, Bardossy A. Impacts of data quantity and quality on model calibration: implications for model parameterization in data-scarce catchments. Water. 2020;12(9):2352. https://doi.org/10.3390/w12092352
74. Rahbeh M, Chanasyk D, Miller J. Two-way calibration-validation of SWAT model for a small prairie watershed with short observed record. Canad Water Resour J/Revue Canadienne des Ressources Hydriques. 2011;36(3):247–70. https://doi.org/10.4296/cwrj3603884

75. Arnaud P, Bouvier C, Cisneros L, Dominguez R. Influence of rainfall spatial variability on flood prediction. J Hydrol. 2002;260(1–4):216–30. https://doi.org/10.1016/s0022-1694(01)00611-4

76. Chaubey I, Haan CT, Grunwald S, Salisbury JM. Uncertainty in the model parameters due to spatial variability of rainfall. J Hydrol. 1999;220(1–2):48–61. https://doi.org/10.1016/s0022-1694(99)00063-3

77. Shin M-J, Guillaume JHA, Croke BFW, Jakeman AJ. A review of foundational methods for checking the structural identifiability of models: results for rainfall-runoff. J Hydrol. 2015;520:1–16. https://doi.org/10.1016/j.jhydrol.2014.11.040

78. Silvestro F, Gabellani S, Rudari R, Delogu F, Laiolo P, Boni G. Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. Hydrol Earth Syst Sci. 2015;19(4):1727–51. https://doi.org/10.5194/hess-19-1727-2015

79. Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. A framework for development and application of hydrological models. Hydrol Earth Syst Sci. 2001;5(1):13–26. https://doi.org/10.5194/hess-5-13-2001

80. Croke B, Jakeman A. A catchment moisture deficit module for the IHACRES rainfall-runoff model. Environ Modell Softw. 2004;19(1):1–5. https://doi.org/10.1016/j.envsoft.2003.09.001

81. Chiew FHS, Peel MC, Western AW. Application and testing of the simple rainfall-runoff model SIMHYD. Mathematical models of small watershed hydrology and applications. Water Resources Publications; 2002. pp. 335–67.

82. Yu Z, Wu J, Yao H, Chen X, Cai Y. Calibrating a hydrological model in ungauged small river basins of the northeastern Tibetan Plateau based on near-infrared images. J Hydrol. 2023;618:129158. https://doi.org/10.1016/j.jhydrol.2023.129158

83. Meyer Oliveira A, Fleischmann AS, Paiva RCD. On the contribution of remote sensing-based calibration to model hydrological and hydraulic processes in tropical regions. J Hydrol. 2021;597:126184. https://doi.org/10.1016/j.jhydrol.2021.126184

84. Avellaneda PM, Ficklin DL, Lowry CS, Knouft JH, Hall DM. Improving Hydrological Models With the Assimilation of Crowdsourced Data. Water Resour Res. 2020;56(5). https://doi.org/10.1029/2019wr026325

85. Dasgupta A, Grimaldi S, Ramsankaran R, Pauwels VRN, Walker JP. A simple framework for calibrating hydraulic flood inundation models using Crowd-sourced water levels. J Hydrol. 2022;614:128467. https://doi.org/10.1016/j.jhydrol.2022.128467

86. Starkey E, Parkin G, Birkinshaw S, Large A, Quinn P, Gibson C. Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation. J Hydrol. 2017;548:801–17. https://doi.org/10.1016/j.jhydrol.2017.03.019