From the Department of Landscape Ecology and Resources Management
Chair of Land Use Systems with Focus on Agroforestry
of the Justus-Liebig University Giessen

Utilization of Hydrometeorological Participatory Monitoring in Data-Scarce Remote Tropical Mountainous Regions

Cumulative dissertation

to obtain the degree of Doctor of Natural Sciences (Dr. rer. nat.)

submitted by

Fabian Mitze

from Frankenberg (Eder)

Gießen, January 2026

With the permission of the Faculty of Agricultural Sciences, Nutritional Science, and Environmental Management of the Justus-Liebig University Giessen

The research reported in this dissertation was carried out at the Center for International Development and Environmental Research (ZEU)
Justus Liebig University Giessen

Examination committee:

1st Supervisor:	Prof. Dr. Suzanne Robin Jacobs
2nd Supervisor:	Dr. Björn Weeser
Examiner:	Prof. Dr. Lutz Breuer
Examiner:	Prof. Dr. Till Kleinebecker
Chairperson:	Prof. Dr. Jan Siemens

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

- Sir Arthur Conan Doyle, A Study in Scarlet (1887)

Table of contents

List of abbreviations

CoCoRaHS .... *Community Collaborative Rain, Hail and Snow Network*
CV .... *Coefficient of Variation*
EPAs .... *Environmental Protection Agencies*
ERA5-Land .... *Land component of the European ReAnalysis dataset*
ET_{pot} .... *Potential Evapotranspiration*
GHCN .... *Global Historical Climatology Network*
GRDC .... *Global Runoff Data Centre*
KGE .... *Kling-Gupta-Efficiency*
m.a.s.l. .... *meters above sea level*
MAE .... *Mean Absolute Error*
MODIS .... *Moderate Resolution Imaging Spectroradiometer*
NCEI .... *National Centers of Environmental Information*
NOAA .... *National Oceanic and Atmospheric Administration*
R^2 .... *Coefficient of determination*
RMSE .... *Root Mean Squared Error*
r_{spear} .... *Spearman Rank correlation*
SDGs .... *Sustainable Development Goals*
T_{as} .... *automatically measured air temperature*
T_{pm} .... *participatory monitoring air temperature*
T_{rs} .... *ERA5-Land 2m air temperature*

List of figures

List of tables

Abstract

Climate change is an ever-present anthropogenic problem that affects all regions of the earth. Reliable hydrometeorological data are essential for the evaluation of the impacts of climate change, with a robust data basis providing a foundation for climate research and the development of mitigation and adaptation measures. Ground-based monitoring stations that provide such data are distributed very unevenly across the globe, and remote sensing methods may be limited in terms of spatial or temporal resolution for specific applications. This is why there is clearly a need for alternative monitoring approaches, particularly in remote tropical mountainous regions that may lack the necessary resources. Participatory monitoring, in which the public is involved in data collection, might be part of a cost-effective solution. Therefore, this work aims to investigate, how hydrometeorological participatory monitoring can be used for (1) efficient hydrological model calibration, (2) alternative observation of different hydrometeorological parameters and (3) enhancing the accuracy of other observational datasets. This methodology was tested in remote tropical mountainous regions in Ecuador, Honduras, Kenya and Tanzania, where conventional data collection methods are severely limited or unavailable.

The first chapter of this dissertation analyzes how many daily participatory monitoring water level measurements at what stages of the hydrological cycle are required to achieve satisfactory hydrological model performance. A simple rainfall-runoff model was used to test this in a tropical mountainous catchment in Kenya. The analysis shows that the starting conditions of the monitoring (wet vs. dry season) influence the required amount of data. Good model performance was achieved within one month when monitoring started in the wet season, while multiple months were required when starting in the dry season.

The subsequent chapter evaluates (1) how the ability of measuring air temperature, relative humidity, rainfall and water level using simple analog sensors differs between frequent and non-frequent participants and (2) how suitable the analog sensors are in terms of accuracy compared to automatic sensors. Between May 2023 and May 2025, 2,982 hydrometeorological observations were received from 52 low-cost stations in

Ecuador, Honduras and Tanzania, of which frequent participants submitted the majority (84.4%), with slightly better accuracy. The analog sensor measurements showed mixed results when compared to automatically measured data. Air temperature and water level performed best with the lowest mean absolute errors (0.74 – 1.65 °C; 0.04 – 0.08 m) while relative humidity data required correction to obtain moderate accuracy (5.45 – 9.50 %) and rainfall was generally underestimated (2.55 to 3.10 mm).

In the final chapter, the use of the participatory monitoring air temperature data from the previous chapter to bias-correct a large-scale continuous air temperature data was tested. Simple linear regression models were trained using ERA5-Land reanalysis and participatory monitoring air temperature measurements from one station in Ecuador (n = 67), Honduras (n = 23) and Tanzania (n = 275). The models were then used to correct the ERA5-Land air temperature bias over a period for which automatically measured hourly reference air temperature data was available (up to one and a half years). ERA5-Land air temperature bias was reduced at all stations, but with varying degrees. The most significant reduction was obtained in Ecuador, with the mean absolute error decreasing from 5.49°C to 1.76°C.

To conclude, this research shows that hydrometeorological participatory monitoring has the potential to improve overall data availability in remote tropical regions with limited financial resources. The knowledge gained from this research could help to inform future participatory monitoring programs and strengthen the acceptance of data collected using such alternative approaches.

Zusammenfassung

Der Klimawandel ist ein allgegenwärtiges anthropogenes Problem, das alle Regionen der Erde betrifft. Zuverlässige hydrometeorologische Daten sind für die Bewertung der Auswirkungen des Klimawandels unerlässlich, wobei robuste Daten auch eine Basis für die Klimaforschung und die Entwicklung von Maßnahmen zur Eindämmung und Anpassung bilden. Bodengestützte Messstationen, die solche Daten liefern, sind global sehr ungleichmäßig verteilt und Fernerkundungsmethoden können hinsichtlich ihrer räumlichen oder zeitlichen Auflösung für bestimmte Anwendungen eingeschränkt sein. Aus diesem Grund besteht ein klarer Bedarf für alternative Monitoring-Ansätze, insbesondere in abgelegenen, bergigen, tropischen Regionen, in denen möglicherweise die erforderlichen Ressourcen fehlen. Partizipative Überwachung, bei der die Öffentlichkeit in die Datenerhebung einbezogen wird, könnte Teil einer kosteneffizienten Lösung sein. Diese Arbeit zielt daher darauf ab, zu untersuchen, wie partizipative hydrometeorologische Überwachung für (1) eine effiziente Kalibrierung hydrologischer Modelle, (2) die alternative Erfassung verschiedener hydrometeorologischer Parameter und (3) die Verbesserung der Genauigkeit anderer Beobachtungsdatensätze genutzt werden kann. Diese Methodik wurde in abgelegenen, tropischen Bergregionen in Ecuador, Honduras, Kenia und Tansania getestet, wo herkömmliche Datenerfassungsmethoden stark eingeschränkt oder nicht verfügbar sind.

Das erste Kapitel dieser Dissertation analysiert, wie viele tägliche partizipative Wasserstandsmessungen in welchen Phasen des hydrologischen Zyklus erforderlich sind, um eine zufriedenstellende Leistung des hydrologischen Modells zu erzielen. Zur Überprüfung dieser Frage wurde ein einfaches Niederschlag-Abfluss-Modell in einem tropischen Berggebiet in Kenia verwendet. Die Analyse zeigt, dass die Ausgangsbedingungen des Monitorings (Regenzeit vs. Trockenzeit) Einfluss auf die erforderliche Datenmenge haben. Eine gute Modellleistung wurde bereits mit einem Monat täglicher Wasserstandsdaten erreicht, sofern die Überwachung in der Regenzeit begann. Begann die Überwachung hingegen in der Trockenzeit, waren mehrere Monate an Wasserstandsdaten erforderlich.

Das folgende Kapitel bewertet (1) inwiefern sich regelmäßige und gelegentliche Teilnehmer hinsichtlich ihrer Fähigkeit unterscheiden, Lufttemperatur, relative

Luftfeuchtigkeit, Niederschlag und Wasserstand mit einfachen analogen Sensoren zu messen, und (2) wie geeignet die analogen Sensoren hinsichtlich ihrer Genauigkeit im Vergleich zu automatischen Sensoren sind. Zwischen Mai 2023 und Mai 2025 gingen 2.982 hydrometeorologische Beobachtungen von 52 kostengünstigen Stationen in Ecuador, Honduras und Tansania ein, von denen regelmäßige Teilnehmer den Großteil (84,4 %) mit etwas besserer Genauigkeit einreichten. Die Messungen mit analogen Sensoren zeigten im Vergleich zu automatisch gemessenen Daten gemischte Ergebnisse. Lufttemperatur und Wasserstand schnitten mit den niedrigsten mittleren absoluten Fehlern (0,74 – 1,65 °C; 0,04 – 0,08 m) am besten ab. Messungen zur relativen Luftfeuchtigkeit mussten korrigiert werden, um eine moderate Genauigkeit (5,45 – 9,50 %) zu erreichen und die Niederschlagsmenge wurde im Allgemeinen unterschätzt (2,55 bis 3,10 mm).

Im letzten Kapitel wurde die Verwendung der partizipativen Lufttemperaturdaten aus dem vorherigen Kapitel zur Fehlerkorrektur großräumiger kontinuierlicher Lufttemperaturdaten getestet. Einfache lineare Regressionsmodelle wurden unter Verwendung der ERA5-Land-Reanalyse und partizipativen Lufttemperaturmessungen von einer Station in Ecuador (n = 67), Honduras (n = 23) und Tansania (n = 275) trainiert. Die Modelle wurden dann verwendet, um die ERA5-Land-Lufttemperaturfehler über einen Zeitraum zu korrigieren, für den stündliche, automatisch gemessene Referenzlufttemperaturdaten verfügbar waren (bis zu eineinhalb Jahre). Die Fehler in den ERA5-Land-Lufttemperaturdaten wurden an allen Stationen reduziert, jedoch in unterschiedlichem Maße. Die signifikanteste Reduktion wurde in Ecuador erzielt, wo der mittlere absolute Fehler von 5,49 °C auf 1,76 °C sank.

Insgesamt zeigt diese Forschung, dass die partizipative hydrometeorologische Überwachung das Potenzial hat, die allgemeine Datenverfügbarkeit in abgelegenen tropischen Regionen mit begrenzten finanziellen Ressourcen zu verbessern. Die aus dieser Studie gewonnenen Erkenntnisse könnten dazu beitragen, künftige partizipative Überwachungsprogramme zu gestalten und die Akzeptanz von Daten zu stärken, die mit solchen alternativen Ansätzen erhoben werden.

Zusammenfassung in einfacher Sprache

Der Klimawandel ist ein von Menschen verursachtes Problem, der jede Region der Erde betrifft. Um die Folgen vom Klimawandel besser verstehen und sich daran anpassen zu können, braucht man Daten über Wetter und Wasser. In vielen entlegenen tropischen Gebieten gibt es jedoch zu wenig Messstationen und auch Satellitenmessungen helfen nicht immer weiter. Für diese Regionen sind deshalb neue Ansätze notwendig um verlässliche Daten zu erheben.

Eine Möglichkeit ist, die Menschen vor Ort an der Datenerfassung zu beteiligen, ein Konzept, das man „partizipatives Monitoring“ nennt. In dieser Arbeit wurde deshalb folgendes getestet:

(1) Wie viele solcher Messungen notwendig sind, um ein hilfreiches Wasser-Simulationsmodell zu bauen.

(2) Wie gut verschiedene Teilnehmer, Dinge wie z.B. die Lufttemperatur oder den Regen mit einfachen Messinstrumenten messen können.

(3) Ob solche Temperaturmessungen helfen können, andere aber ungenauere Daten zu verbessern.

Bergregionen in verschiedenen in tropischen Ländern – Ecuador, Honduras, Kenia und Tansania – wurden dafür ausgewählt. Dort gibt es wenige bis gar keine Messtationen, teilweise auch, weil diese zu teuer sind. Hier wurde das partizipative Monitoring getestet.

Im ersten Kapitel der Arbeit wurde untersucht, wie viele Wasserstandsmessungen zu verschiedenen Zeiten des Jahres nötig sind, um eine gute Leistung mit einem Wasser-Simulationsmodell zu erreichen. Im Ergebnis reicht ein Monat an täglichen Wasserstandsmessungen für eine gutes Modell aus, wenn es in dieser Zeit viel geregnet hat. War es jedoch sehr trocken, hat das Modell mehrere Monate an täglichen Daten gebraucht, um auch ähnlich gut zu funktionieren.

Im zweiten Kapitel wurden regelmäßige und gelegentliche Teilnehmer verglichen, wie gut sie die Lufttemperatur, relative Luftfeuchtigkeit, Regen und Wasserstand mit einfachen analogen Sensoren messen können. Zwischen Mai 2023 und Mai 2025 haben die Teilnehmer an dafür kostengünstig gebauten Stationen in Ecuador, Honduras und

Tansania fast 3.000 Beobachtungen gesammelt. Regelmäßige Teilnehmer haben über 80 % der Daten gesendet und maßen etwas genauer. Allgemein hat sich herausgestellt, dass die Lufttemperatur und der Wasserstand sich sehr gut mit den einfachen Instrumenten messen lassen. Die relative Luftfeuchtigkeit war jedoch ungenau und musste korrigiert werden. Die Regenmessungen waren im Durchschnitt zu niedrig.

Im letzten Kapitel wurde getestet, ob die zuvor gesammelten Lufttemperaturdaten genutzt werden können, um andere, bereits verfügbare Lufttemperaturdaten zu verbessern. Hierfür wurden Messungen aus Ecuador, Honduras und Tansania genutzt, um einfache Rechenmodelle anzupassen. Mithilfe dieser Modelle wurden dann längere Zeiträume der großräumig verfügbaren Daten versucht zu verbessern. An allen drei Standorten haben sich die Daten verbessert, insbesondere in Ecuador hat sich die Genauigkeit deutlich erhöht.

Insgesamt zeigt diese Arbeit, dass die Einbindung von Menschen vor Ort in die Messung von Wetter- und Wasserparametern ein vielversprechendes Mittel ist, um die Datenqualität in abgelegenen tropischen Regionen, in denen finanzielle Mittel knapp sind, deutlich zu verbessern. Diese Erkenntnisse können helfen, solche Programme noch besser zu machen und die Akzeptanz solcher Daten zu verbessern.

Publications connected with this thesis

Accepted articles in peer-reviewed journals

Campos Zeballos, J., Valencia, J., Codalli, F., Mitze, F., Shagega, F., Weeser, B., Jacobs, S. (2025). Evaluating participatory monitoring in mountainous tourist regions. Frontiers in Environmental Science, 13, 1537278. https://doi.org/10.3389/fenvs.2025.1537278

Mitze, F., Jacobs, S. R., Breuer, L., Zeballos, J.C., 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

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Codalli, F., Shagega, F. P., Weeser, B., (2026). Validation of analog sensor measurements in hydrometeorological participatory monitoring in various tropical countries. Frontiers in Earth Science, 14. https://doi.org/10.3389/feart.2026.1721642

Submitted article to peer-reviewed journal

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Dowling, T. P. F., & Weeser, B. (under review). Bias correction of ERA5-Land air temperature in remote tropical regions using participatory monitoring data.

Selected conference contributions

Mitze F., Breuer L., Jacobs S., Weeser B. 2024, Hydrometeorological participatory monitoring in data-poor tropical mountain regions, Water Security and Climate Change Conference, Giessen, Germany, October 9 – 11, 2024

Mitze F., Jacobs S., Breuer L., Campos Zeballos J., Weeser B. 2025, Combining participatory monitoring and remote sensing data with machine learning for high-resolution air temperature estimation, Citizen Science for Water Conference, Delft, The Netherlands, June 3 – 5, 2025

Mitze F., Jacobs S., Breuer L., Campos Zeballos J., Weeser B. 2025, Validation of participatory monitoring using analog hydrometeorological sensors in different tropical countries, Water Research and Horizon Conference, Bochum, Germany, September 29 – 30, 2025

1. Extended summary

1.1 Introduction

Climate change is an omnipresent, man-made problem affecting all regions of the Earth (IPCC, 2023a). Reliable environmental and climatologic data are required for climate research and the development of climate protection and adaptation measures. Over the past few decades, the overall availability of hydrometeorological data has improved significantly, where especially the increased deployment of satellites and radars has led to an unprecedented increase of available data (Margulis et al., 2006). A review of literature from 2015 to 2024 revealed substantially increasing research in the field of environmental monitoring, indicating more data is available from year to year (lubis et al., 2025). Several terabytes of new data are generated every day by the National Oceanic and Atmospheric Administration (NOAA) and by the end of 2023 the overwhelming amount of data managed by the National Centers of Environmental Information (NCEI) of the United States, reached 60 Petabytes, with an exponential upward trend (NCEI, 2023; Willett et al., 2023). Nevertheless, the increase in data availability doesn't apply to all types of hydrometeorological datasets. An analysis of Ruhi et al. (2018) using data from the Global Runoff Data Centre (GRDC) showed, that discharge monitoring worldwide peaked in 1979 and declined afterwards. Similar patterns were found in the Global Historical Climatology Network (GHCN) database (Menne et al., 2012). While the number of stations measuring air temperature increased until the late 1960s and remained roughly the same since then, rainfall monitoring reached a peak back then and declined afterwards, especially outside North America (Menne et al., 2012).

Although the amount of environmental data is generally increasing, its availability is unevenly distributed around the world. An analysis by Kidd et al. (2017), for instance, showed that rainfall monitoring is highly disproportionate on a worldwide scale. It is also very likely that there is more data available in areas with high population density, such as cities and surrounding areas reflecting the distribution of the population (Kidd et al., 2017; Menne et al., 2012; Taylor et al., 2024; Zhan et al., 2023). As a consequence, remote or rural areas might face lower availability of precise environmental data due to lower gauging density.

Furthermore, the distribution is quite uneven between high- and low-income countries, with many regions of the world, especially in the Global South (United Nations, 2022) facing incomplete hydrometeorological datasets (Chacon-Hurtado et al., 2017; Krabbenhoft et al., 2022; Menne et al., 2012; Mishra & Coulibaly, 2009; Walker et al., 2016). The roughly 80,000 weather stations included in the GHCN database show a highly disproportionate distribution between countries of different backgrounds (Menne et al., 2012). For instance, Germany, a central-European and high-income country, has a current number of 1,069 stations listed, while Honduras, a central-American and low-income country, has only 7 (NOAA, 2025).

This unequal data distribution may also be a substantially limiting factor for sustainable development, as insufficient data could diminish the basis for evidence-based strategic decisions, such as those related to climate change adaptation (Nilashi et al., 2023). Sustainable development is assessed by the United Nations for all countries of the world by the monitoring of the degree of achievement of the seventeen Sustainable Development Goals (SDGs) (United Nations, 2015). These goals are political objectives to ensure sustainable development on economic, social and environmental scale (United Nations, 2015). Moyer and Hedden (2020) analyzed to what extent the SDGs have been achieved and found that 28 particularly vulnerable countries are not likely to achieve most of the goals in a middle-of-the-road scenario. 26 of them are tropical sub-Saharan African countries (Moyer & Hedden, 2020). Following on from this, analysis of climate change research data revealed that climate modelling is primarily designed and verified in high-income countries, where more data is available (James et al., 2018; Otto et al., 2020). This further emphasizes the need for reliable hydrometeorological data for countries in the Global South, which may also support the achievement of sustainable development. Therefore, alternatives must be explored for these regions.

Remote sensing (RS) based monitoring, such as satellite or radar applications, has great potential for closing gaps in environmental data because it enables data collection where ground-based measurements are unavailable. Despite covering larger and inaccessible areas and growing accuracy over the last decades, it often does not provide the same spatiotemporal precision as ground-based measurements for all required parameters (A. P. Barros & Arulraj, 2020; Kimani et al., 2017; Levizzani & Cattani, 2019; Mao et al., 2021; Marra et al., 2019). Apart from that, all remotely sensed data requires ground truth in the form of ground-based measurements for verification (Hoffer, 1972; Nagai et al., 2020).

Furthermore, some parameters, such as water levels in small rivers, cannot be measured with sufficient precision using remote sensing (Bandini et al., 2017; Grimaldi et al., 2016; Musa et al., 2015). This is why the golden standard for hydrometeorological data collection at a local level remains the use of automated sensors, such as automatic weather stations (Sene, 2024; Valipour et al., 2020). The main challenge for this traditional data collection are primarily the costs for sensors and their maintenance, which is most likely a main reason for the uneven distribution. This disadvantages countries of the Global South, and particularly remote mountainous regions, where financial resources and access to technology and spare parts may be limited (Buytaert et al., 2014; Fankhauser & McDermott, 2014; Ruhi et al., 2018).

Due to all these previously mentioned limitations, alternative approaches for data collection have been investigated in recent years. One possible alternative to hydrometeorological stations equipped with automatic sensors could be participatory monitoring (PM), where citizens collect data (Danielsen et al., 2010). This is a subcategory of citizen science, which is defined as the involvement of the general public in the scientific process (Dickinson et al., 2012). Multiple successful approaches have been tested in different parts of the world. For example, volunteers in the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) collected weather data in the United States, Canada and the Bahamas (Reges et al., 2016). In the CrowdWater project (Seibert et al., 2022), participants were included in the collection of water level, soil moisture and temporary stream data in various countries around the globe. While having an established status in Europe, North America or other parts of the Global North, PM programs are still not as common in countries of the Global South (Buytaert et al., 2014; Njue et al., 2019; Requier et al., 2020). In the last couple of years, there have been different initiatives for PM, such as water levels in Kenya (Weeser et al., 2018), rainfall in Nepal and Vietnam (Davids, Devkota, et al., 2019; Duwal et al., 2025; Tran et al., 2025) or Mexico (Shinbrot et al., 2020a) or various water related parameters, such as soil moisture or groundwater depth, in Ethiopia (Rigler et al., 2022).

The implementation of such projects can also help to enhance the understanding and public awareness of science (Bonney et al., 2016). Personal development benefits for participants beyond helping with data collection were observed as well, such as increased knowledge and interest regarding environment issues (Peter et al., 2021; Phillips et al., 2019; Wilson et al., 2025). There is also multiple evidence that participants in such PM projects are more likely to show pro-environmental behavior (Branchini et al., 2015; Meschini, Prati, et al., 2021; Overdevest et al., 2004; Srisathan et al., 2024). Also several Environmental Protection Agencies (EPAs) have already recognized the value of PM for environmental protection (Rubio-Iglesias et al., 2020).

All these PM programs showed that it is possible to mobilize people to collect environmental data. Nevertheless, the acceptance of PM data for management and research purposes is still low considering temporal resolution, uncertainty of measurement accuracy as well as further application of the data or long-term participation (Buytaert et al., 2014; Njue et al., 2019; Pandeya et al., 2021). Hence, further research exploring to what extent PM data can supplement traditionally collected data or provide an alternative is required.

This dissertation therefore aims to contribute to a better understanding of the utilization of hydrometeorological participatory monitoring by demonstrating the quality and use of PM data. This is done with PM applied in remote tropical mountainous regions in Kenya, Ecuador, Honduras and Tanzania where traditional data collection is severely limited or not done at all. The insights obtained from this dissertation should help to develop effective future PM projects to increase data availability in regions with limited monitoring resources. This is done in several steps, which are 1) a local application to test how much PM data is required to successfully calibrate a hydrological model, 2) a validation of PM for different hydrometeorological parameters using simple analog sensors and the suitability of different participants and 3) the application of such data to bias-correct large scale remote sensing data.

Furthermore, the research findings obtained in this dissertation might also contribute to the following SDG targets (United Nations, 2015):

- 6.a Expand international cooperation and capacity-building support to developing countries.
- 6.b Support and strengthen the participation of local communities in improving water and sanitation management.
- 13.3 Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning.

- 13.b Promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing States, including focusing on women, youth and local and marginalized communities.
- 15.4 By 2030, ensure the conservation of mountain ecosystems, including their biodiversity, in order to enhance their capacity to provide benefits that are essential for sustainable development.

1.2 Objective and research questions

The main objective of this dissertation is to demonstrate the quality and further usability of hydrometeorological participatory monitoring data in various remote tropical mountainous regions. Three separate chapters will address different research questions, with each chapter focusing on a specific aspect of the overall objective:

Chapter 2: How many PM measurements at what stages of the hydrological cycle are required to achieve satisfactory performance of a simple hydrological model applied to a tropical mountainous catchment?

Chapter 3: How does participation frequency influence the participants ability to measure hydro-meteorological parameters using simple analog sensors and how good are the measurements in terms of accuracy?

Chapter 4: Can PM data be used to successfully bias correct large-scale air temperature data in remote tropical regions?

In addition to addressing the scientific knowledge gap regarding the application of PM in the field of hydrometeorology, the knowledge gained from this research could help to inform future participatory monitoring programs and strengthen the acceptance of data collected using such alternative approaches.

1.3 Study areas

The datasets presented in this dissertation were acquired in different tropical mountainous regions in Africa, Central- and South America (Figure 1.1), which are presented in more detail in the following.

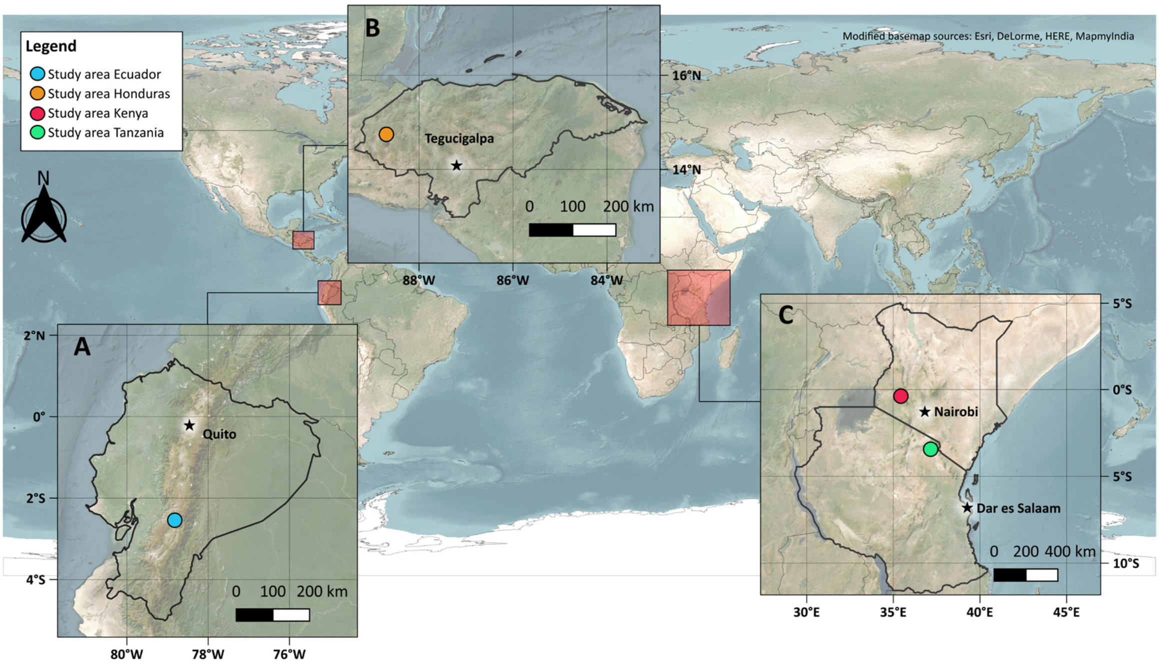


Figure 1.1: *Overview of the study regions in southern Ecuador (A), western Honduras (B) and southwest Kenya (upper) & northern Tanzania (lower) (C) with the black stars showing the capitals of the countries.*

The majority of the study regions are within or around protected national parks, which are considered crucial for providing water-related ecosystem services such as water purification, retention, and climate regulation (Benn & Bindra, 2011; Grizzetti et al., 2016; Portalanza et al., 2025; Shagega et al., 2025; Southworth et al., 2004). Ultimately, these services are crucial for the provision of clean drinking water for the people living in immediate vicinity. This makes these regions particularly important for local implementation of SDG 6, ensuring access to water and sanitation for all (United Nations, 2015).

1.3.1 Kenya

The study region of Kenya, which is the basis for Chapter 2, is located in the southwestern part of the country about 170 km northwest of the capital Nairobi, in a headwater catchment of the Sondu-Miriu River basin (Figure 1.2).

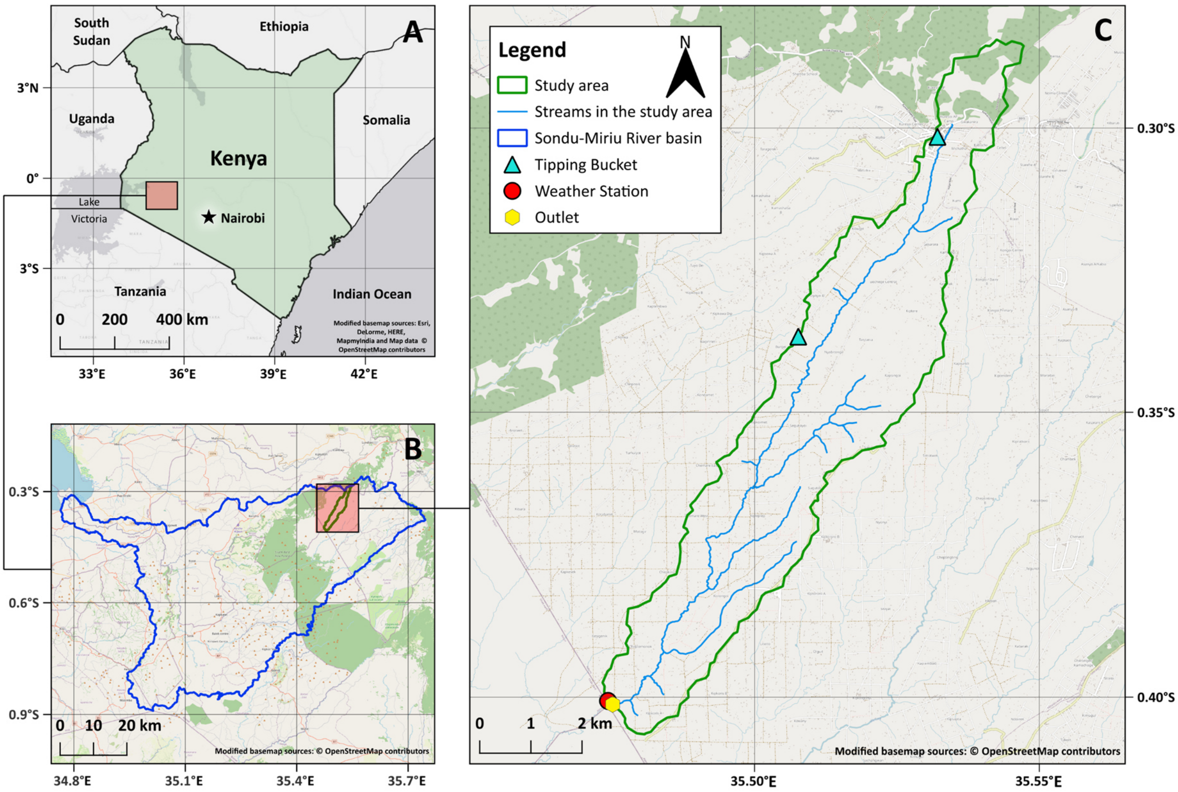


Figure 1.2: *Overview of the study region in Southwestern-Kenya (A), the location within the Sondu-Miriu River catchment (B) and the headwater catchment with the installed sensors (C).*

The catchments topography ranges from 2,380 up to 2,691 meters above sea level (m.a.s.l.) (Jacobs et al., 2020). The soils are Humic Nitisols and Mollic Andosols and the landscape is characterized by a mix of smallholder farms with annual crops, woodlands, forests and pastures (ISRIC, 2007; Weeser et al., 2019). Due to the elevation a moderate tropical climate can be found in the area with a bimodal rainfall pattern with rainy seasons from April to July and October to December caused by the movement of the Intertropical Convergence Zone (Hills, 1979). As precise climate records are missing for this area, the Climate Change Knowledge Portal of the World Bank (2025b) was used to determine the average air temperature of the region Nakuru in which the catchment is located, to be 15.63 °C. The mean annual precipitation from 2015 to 2018 was approximately 1,568 mm (Jacobs et al., 2020).

1.3.2 Ecuador

The second case study region is located in the province Azuay, Southern Ecuador. It is divided into three different study sites (Figure 1.3).

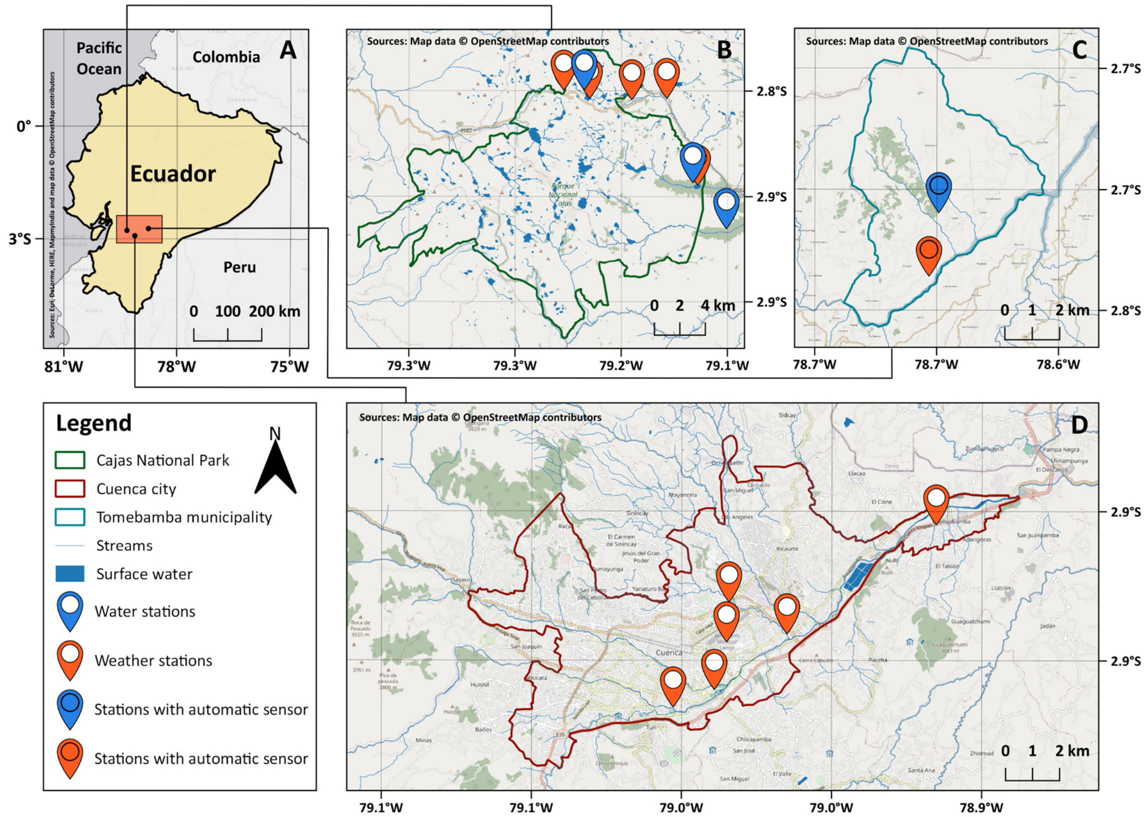


Figure 1.3: *Study area in Southern Ecuador (A) with the locations of the HydroCrowd stations in the three sites within and around the Cajas National Park (B), within the municipality of Tomebamba (C) and within the city of Cuenca (D) (adapted from Mitze et al. (2026)).*

The first site is in the northeastern part of the Cajas National Park (Figure 1.3B) with an elevation between 3,144 to 4,429 m a.s.l., which is about 25 km northwest of the city of Cuenca (Pesántez et al., 2018). This area is part of the Andean Páramo, a tundra-like, grassland-dominated ecosystem (Correa et al., 2016). The landscape is characterized by Holocene Andosols and Histosols where tussock grass, low shrubs, and glacial lakes dominate, though there are also small areas of pine forest and pasture (Bandowe et al., 2018; Buytaert et al., 2006; Carrillo-Rojas et al., 2016; Harden, 2006). The local climate is classified as a tropical high-mountain climate, influenced by atmospheric mechanisms. The most prominent are the Intertropical Convergence Zone and the El Niño-Southern Oscillation and the Pacific Decadal Oscillation (Buytaert et al., 2006; Morán-Tejeda et al., 2016; Poveda et al., 2006). The seasonal variability is minimal, due to its equatorial location with daytime temperatures between 12 - 18°C and nighttime temperatures that can drop below zero (Buytaert et al., 2006; Hansen et al., 2003). Precipitation may vary depending on the location, but various studies observed an annual precipitation between 1,000 and 1,300 mm (Buytaert et al., 2006; Celleri et al., 2007; Padrón et al., 2020).

The second site is within the boundaries of the city of Cuenca (Figure 1.3C), with a population of approximately 596,000 (Instituto Nacional de Estadística y Censos, 2022). The city is located in a basin with an altitude of 2,350 to 2,550 m a.s.l. (Gianoli & Bhatnagar, 2019) with a temperate Andean climate with an average air temperature of 16.3°C and an annual precipitation of 876 mm (WMO, 2025b).

The third site is located in the municipality of Tomebamba, which is around 37 km north-east of the city of Cuenca. According to the World Bank, the climate is comparable to Cuenca (World Bank, 2025c). Due to its location at the edge of a valley, the area covers a significant elevation gradient (Figure 1.4), with the municipality of Tomebamba being located at approximately 2,380 m a.s.l.

Figure 1.4: *Landscape of the study site in Tomebamba with smallholder agriculture, pasture and shrubs in the lower and forest in the higher area (own photograph from November 6, 2023).*

1.3.3 Honduras

The study region in Honduras encompasses mainly the Cacique Lempira Señor de las Montañas Biosphere Reserve, which is a greater area around the Celaque National Park (Figure 1.5).

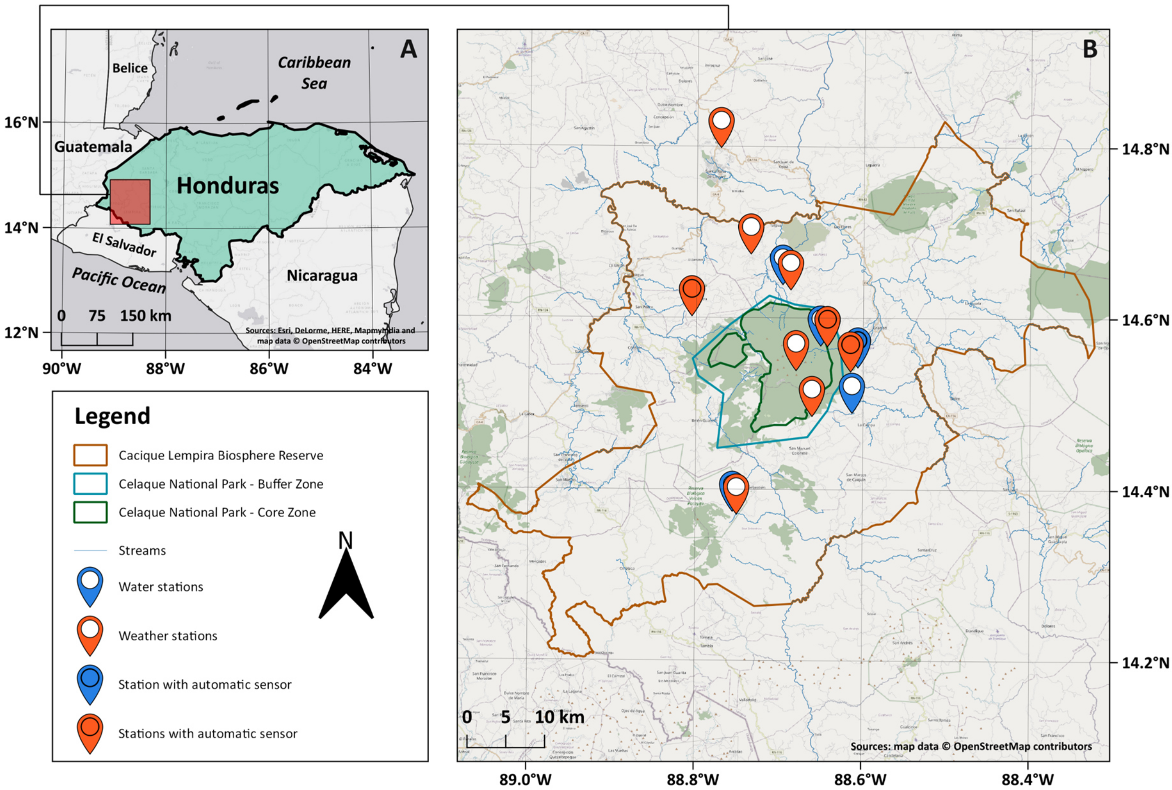


Figure 1.5: *Study area in Western Honduras (A) with the locations of the HydroCrowd stations in the Cacique Lempira Biosphere Reserve (B; adapted from Mitze et al. (2026)).*

The region has a mountainous landscape, up to 2,870 m a.s.l. with rocky, thin Lithosols (D. L. Anderson & Devenish, 2009; FAO, 1966; Southworth et al., 2004). Most of the area is considered unsuitable for most agricultural activities, anyway agricultural and industrial activities have been prohibited in the area above 1,800 m a.s.l. since 1987 (Pfeffer et al., 2001; Southworth et al., 2004). While in the protected area above 1,800 m a.s.l. consist of a transition from pine-oak forest into a mixed broad-leaf/ pine montane forest, broadleaf forests dominate above 2,200 m a.s.l. (Southworth et al., 2004). Below 1,800 m a.s.l. land use is a mixture of coniferous forests, coffee plantation, shrublands and a small share of annual crops (see also Figure 1.6) outside of the National Park (Valdez et al., 2017).

Figure 1.6: *Example of the study region in Honduras at approx. 1,181 m a.s.l. with pasture and mixed forests. The mountains of the Celaque National Park are visible in the background (own photograph from April 22, 2024).*

Climate in the region can be described as tropical savannah climate with average precipitation around 1,600 mm per year and air temperature around 24 °C in the lower areas and up to 2,400 mm per year at higher elevations of the park (Aguilar, 2005; Valdez et al., 2017). As it is typical for savannah climate in the region, the precipitation of the area predominantly occurs in a wet phase from May to October, while the rest of the year is dry (Aguilar, 2005).

1.3.4 Tanzania

The study area in Tanzania is predominantly focused on the southern slopes of the Kilimanjaro National Park (Figure 1.7).

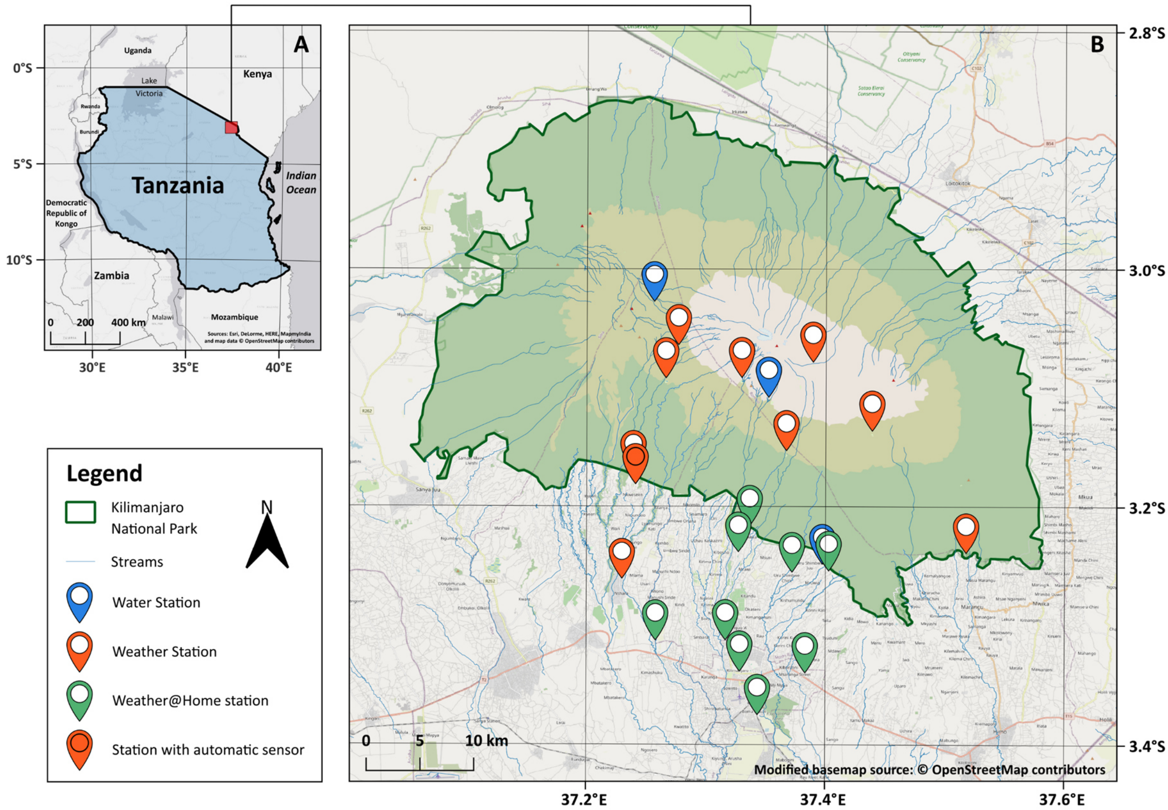


Figure 1.7: *Study area in Northern Tanzania (A) with the locations of the HydroCrowd stations at the southern slopes of Mt. Kilimanjaro (B; adapted from Mitze et al. (2026)).*

Significant differences in climate, dominating vegetation and land use can be found due to the high elevation gradient (from 770 to 5,895 m a.s.l.) in the area (Hemp et al., 1998), which results in multiple ecological zones (Hemp et al., 1998). An example of the diverse landscape in the area is shown in Figure 1.8.

Figure 1.8: *Landscape of the study region in Tanzania at approximately 1,700 m a.s.l. with agroforestry in the foreground and Kibo (highest point of Kilimanjaro) slightly visible behind the clouds in the left corner of the picture (photograph by Suzanne Jacobs from October 29, 2021).*

These differences are also reflected in the distribution and amount of rainfall, where 900 mm of precipitation are reached in the lowlands at 800 m a.s.l., while it peaks at around 2,700 mm at 2,200 m a.s.l. (Hemp, 2006; Røhr & Killingtveit, 2003). Above that, rainfall decreases down to 750 mm at 3,750 m a.s.l. and continues to decrease beyond this point (Appelhans et al., 2016). Seasonal rainfall distribution follows a bi-modal pattern, with peaks in March to May and November and December (Shagega et al., 2025). The dominating soil types are Andosols (Zech, 2006).

1.4 Methodology

1.4.1 Evaluation of hydrological model performance using PM

To analyze how much PM water level data is required for a satisfactory hydrological model performance, data collected during a PM project in a sub-catchment of the Sondu-Miriu River basin in south-western Kenya from 2016 to 2020 was used (Weeser et al., 2018, 2019). The water level data was collected by participants using an analog water level gauge and the measured value was sent via SMS to a central database (Weeser et al., 2018). A comparison with radar-based water level measurements revealed a high Pearson correlation coefficient of 0.96. The dataset was split into a three-year calibration (April 2016 – April 2019) and a one-year validation period (April 2019 – April 2020).

In a next step, gaps in the PM water level dataset were filled to create a baseline dataset and to ensure comparability between the water level availability scenarios to be compared. For this, a random forest regression was used with an accuracy of 0.96. To test different scenarios of PM water level data availability, six starting points in wet and dry conditions were defined. To ensure comparability, three starting points were selected in wet conditions with high water levels/discharge values (a, c, e) and three in dry conditions with low water levels/discharge values (b, d, f) (see Figure 1.9).

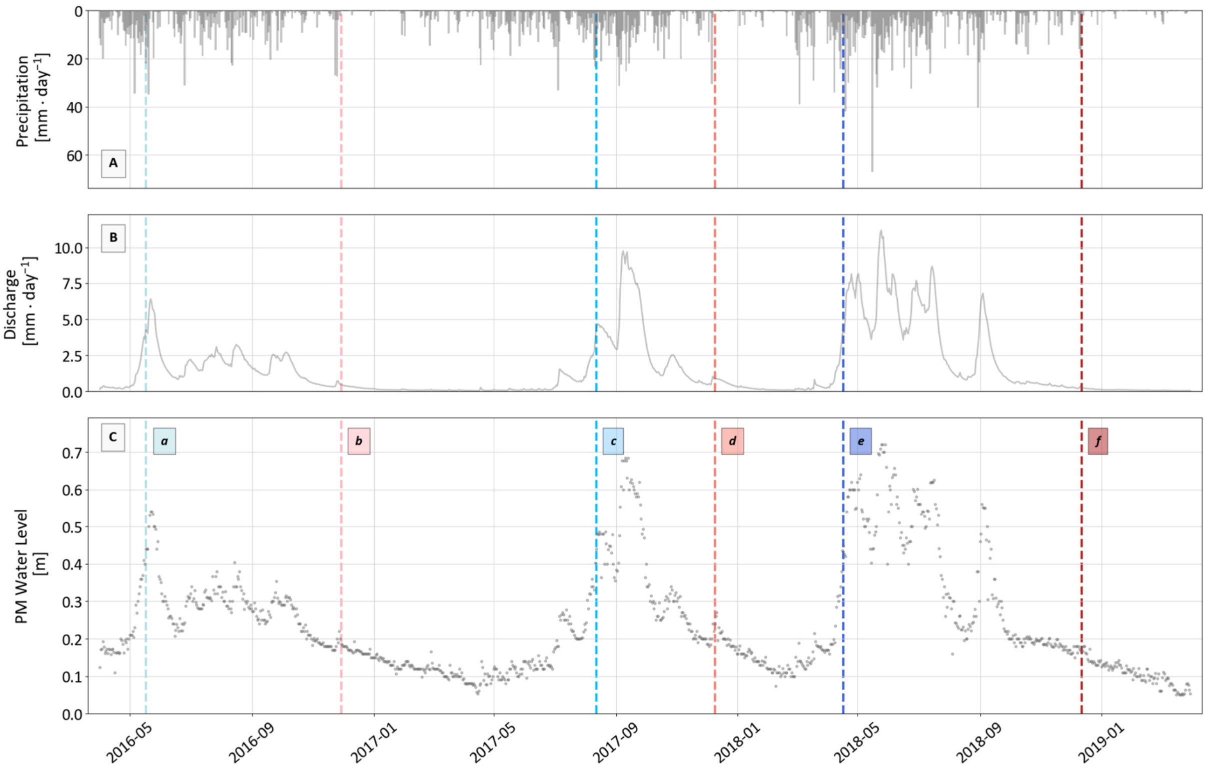


Figure 1.9: *Overview of A) precipitation, B) discharge and C) baseline PM water level data in the calibration period (April 2016 – April 2019) with six different scenario starting points, n = 3 in wet seasons (a, c and e) and n = 3 in dry seasons (b, d and f) (adapted from Mitze et al. (2025)).*

From each scenario starting point, scenarios were sampled from the baseline PM water level dataset, where the first scenario of each group contained one month of daily water level measurements (n = 30), the next one two months and so on. This led to overall 110 different scenarios with differing amounts of PM water level data from different periods.

A simple rainfall-runoff model developed by Weeser et al. (2019) was then calibrated on each water level scenario instead of runoff, using Spearman Rank correlation (r_{spear}) (Seibert & Vis, 2016). A Latin Hypercube-based approach (McKay et al., 1979) was used to evaluate 10^5 model parameter sets. As a benchmark the model was additionally calibrated on the automatically measured radar-based water level data, the whole original (with gaps) and the baseline (without gaps) PM water level dataset. Model performance was measured using Kling-Gupta-Efficiency (KGE) (Gupta et al., 2009), where a scenario was considered good, when KGE was ≥ 0.75. As r_{spear} does not reflect absolute discharge volumes (Seibert & Vis, 2016), only the best 0.5% of all tested model runs were used (KGE_{mean}). On top of each calibration (benchmark and each scenario), a Water Balance filter (Weeser et al., 2019) was applied to only select calibration schemes, that are within a reasonable range of discharge when subtracting potential evapotranspiration from rainfall. To test model performance in the validation period, all accepted parameter sets from the calibration of the different scenarios were used.

1.4.2 Validation of hydrometeorological PM

To validate how participation frequency influences the ability to measure hydrometeorological parameters using simple analog sensors and how good the accuracy of the analog sensor measurements is, data from the PM project HydroCrowd between May 2023 and May 2025 was used. In the project, simple and low-cost hydrometeorological measurement networks for air temperature, relative humidity, rainfall, water level and turbidity were set up in remote mountainous regions in Ecuador, Honduras and Tanzania (Campos Zeballos et al., 2025).

An Android and iOS smartphone application was developed to enable participants to upload their measurements. The participants had to enter the values of their hydrometeorological measurements and were asked to upload a photo for validation purposes. In addition, every station offered a QR-Code to upload measurements directly, if downloading the application was not possible or not desired (https://www.spotteron.com/hydrocrowd/). Two main types of simple stations were developed for the different parameters: weather stations (including a thermometer for air temperature, a hygrometer for relative humidity and a rain gauge for rainfall; Figure 1.10) and water stations (including a water level gauge for water levels and a turbidity tube for turbidity; Figure 1.11) (Campos Zeballos et al., 2025; Mitze et al., 2026).

Figure 1.10: *Example of a HydroCrowd weather station from Honduras with the three analog sensors: a hygrometer, a thermometer and a rain gauge (upper right side of the wooden panel from left to right; adapted from Mitze et al. (2026)).*

Figure 1.11: *Example of a HydroCrowd water station (left) from Honduras with the corresponding water level gauges (right) mounted on a big rock in the water (own photograph from April 30, 2024).*

Additionally, a few more stations with only a small simple signpost and a rain gauge were established in Tanzania, which are only located on private property and are called Weather@home stations.

The station network in Ecuador, consisting of twelve weather and four water stations, was established in the Cajas National Park, the city of Cuenca and the community of Tomebamba (see Figure 1.3). The majority of the stations, which were installed in the Cajas National Park were placed alongside hiking routes. In Honduras, nine weather and five water stations were installed at various sites within the Cacique Lempira Biosphere Reserve area (see Figure 1.5). The stations are located alongside hiking routes, municipal areas such as schoolyards or gardens or on private property. Ten weather, three water and nine weather@home stations were installed in the Kilimanjaro National Park and slightly outside on the southern slopes of Kilimanjaro in Tanzania (see Figure 1.7). The predominant placement of the weather and water stations was also alongside popular hiking routes. Overall, a total number of 31 weather, twelve water and nine Weather@home stations were installed between May 2023 and May 2025.

By checking the unique ID of each user of the smartphone application it was analyzed how often participants measured and in what temporal pattern. If they submitted a maximum of six observations within the first seven days after their first contribution, they were assigned as non-frequent users. All anonymous contributions, submitted via the web application, were assigned to the non-frequent users as well. If users submitted more, they were assigned as frequent participants. The participant's measurement accuracy was then validated by comparing the submitted values to the corresponding photos of each measurement to check if regular participation influences the accuracy of measurements.

To validate the analog sensors accuracy, one station per country was selected to additionally install automatic sensors (see Figure 1.12) for air temperature, relative humidity, rainfall, and water level (only in Honduras and Ecuador). The first automatic temperature/ humidity sensor was stolen in Honduras. This sensor was replaced and a second one was installed and shifted between the weather stations to ensure data collection and validation for as long of the study period as possible.

Figure 1.12: *Example of a HydroCrowd weather station in Honduras with a hidden temperature/humidity sensor behind the station (right) and a tipping bucket for automated rainfall monitoring (left), which was accessed during fieldwork (photograph by Suzanne Jacobs from the June 11, 2023).*

The accuracy of the originally submitted (not corrected with the photos) analog measurements was assessed by using the automatic measurements closest in time applying mean absolute error (MAE), root mean squared error (RMSE), coefficient of variation (CV) and r_{spear}. For rainfall, the amount of automatically measured rainfall since the last analog measurement was used for comparison. To support the comparison using the different metrics, the analog measurements were checked for statistically significant differences using the Wilcoxon signed-rank test.

1.4.3 Bias correction using PM data

To analyze the suitability of PM data for bias correction of large-scale air temperature datasets, outlier-corrected PM air temperature (see Chapter 3) measurements from the weather stations at Tomebamba (Ecuador), Don Tito (Honduras) and Nkweseko (Tanzania) were selected. These stations differed in amounts of available measurements (Tomebamba = 67, Don Tito = 23 and Nkweseko = 275). For each analog measurement, the corresponding 2m air temperature value closest in time of the land component of the large-scale European ReAnalysis (ERA5-Land) dataset (Muñoz-Sabater et al., 2021) was retrieved. Since each pixel is approximately 9 km in size, the point value was extracted using the coordinates of the PM stations (Mitze et al., under review). A descriptive statistical analysis was conducted to analyze the differences between PM (T_{pm}) and ERA5-Land (T_{rs}) air temperature. Subsequently, T_{rs} was used to train a linear regression model on the T_{pm} for each location.

For model validation, T_{rs} was also retrieved for the same periods for which automatically measured air temperature data (T_{as}) was available (table 1.1).

Table 1.1: *Period monitored for stations with automatically measured air temperature (n = number of individual hourly measurements available, adapted from Mitze et al. (under review)).*

Station name	Country	n	Period monitored
Nkweseko	Tanzania	12,912	26.08.2023 - 14.02.2025
Tomebamba	Ecuador	4,680	07.12.2023 - 19.06.2024
Don Tito	Honduras	10,080	08.05.2024 - 02.07.2025

Subsequently, the trained models were used to correct these T_{rs} values. The bias corrected values (T_{bc}) were then compared against the automatically measured values (T_{as}). As a baseline, the T_{rs} was compared directly to T_{as}. The comparison was evaluated using MAE, RMSE and the coefficient of determination (R^2). Furthermore, the structure of the bias correction datasets was analyzed for temporal patterns and differences between the ERA5-Land and bias corrected values were checked for statistical significance.

1.5 Summary of results

The objective of this doctoral thesis was to demonstrate the quality and further usability of hydrometeorological PM data in in different remote tropical mountainous regions. The results for every chapter are presented below.

1.5.1 Objective 1: Evaluation of hydrological model performance using PM

To evaluate how much PM water level measurements from what stage of the hydrological cycle are required to achieve satisfactory performance of a simple hydrological model applied in a mountainous catchment in Kenya, different scenarios of differing availability of water level data were tested (Chapter 2). In the first place, the comparison of the calibration with the radar-based, the original and baseline (gap filled) PM water level data did not reveal structural differences in model performance, with the KGE_{mean} varying from 0.51 to 0.53. Applying the Water Balance filter on these three calibration schemes, the performance was significantly improved for all to a similar extent with KGE_{mean} ranging from 0.73 to 0.75.

A comparison of the 110 scenario results showed that scenarios starting in wet conditions performed substantially better and needed less data to achieve a satisfactory model performance (KGE_{mean}: 0.51 – 0.85). On the other side, those starting under dry conditions needed considerably more data, to reach acceptable performance (KGE_{mean}: 0.24 – 0.82). The wet scenarios reached $KGE_{mean} \geq 0.75$ already with one month of data, while the dry ones usually needed more than half a year, including data from following wet seasons to achieve similar performance. Overall, it could be observed that using more PM water level data for model calibration did not necessarily improve model performance. In fact, model performance saturated for all scenario groups at around 0.76 – 0.79 after different periods of time. Validation results showed similar patterns where wet scenarios performed better in general, with a lower required amount of water level data. However, the model's performance was lower overall, with the KGE_{mean} reaching a maximum of 0.67. Similar to the calibration, the model's performance reached a saturation point after around 12 months, stabilizing at approximately 0.65.

The case study presented in chapter 2 showed on the example of a small tropical mountainous catchment in Kenya that a simple hydrological model can be calibrated successfully using only a low number of PM water level measurements. One month of data, i.e. 30 daily measurements, has proven to be sufficient if collected within a wet period with high water levels where higher discharge occurs. This is in line with previous research using smaller samples of discharge or regularly measured water level (Jian et al., 2017; Juston et al., 2009; Seibert & Beven, 2009).

1.5.2 Objective 2: Validation of hydrometeorological PM

To validate the measurement accuracy of different participants and the accuracy of the analog sensors used, a hydrometeorological participatory monitoring was conducted in remote tropical mountainous regions in Ecuador, Honduras and Tanzania (Chapter 3). 2,982 observations by participants in the hydrometeorological PM project HydroCrowd were received in the study period from May 2023 to May 2025. Due to the lower number of water stations in the study regions, fewer water level measurements were conducted (n = 230), while for rainfall, air temperature and relative humidity a comparable number of around 1,950 measurements were received (without weather@home measurements). All individual parameter measurements from Tanzania totaled in 2,668, including Weather@home stations, while for Ecuador (n = 2,323) and Honduras (n = 1,804) fewer measurements were received. The vast majority of 84.4% of all measurements were conducted by frequent participants, while non-frequent participants submitted 15.6%. Comparing measurements where photos for validation were available, both groups of participants had low error rates (i.e. erroneous measurements), which were below 10% for rainfall, air temperature and relative humidity. For water level, frequent participants had an error rate of 11.7 % and non-frequent participants 23.5%. Overall, frequent participants had a lower error rate for all parameters, albeit not statistically significant.

A moderate share (8.2 – 20%) of all analog air temperature and relative humidity measurements that were compared to automatic measurements to assess their accuracy, were classified as outliers due to overheating of the sensors due to intense sunlight exposure. These were removed prior to further validation. The comparison with automatic sensor measurements revealed high agreement of air temperature (MAE 0.74 – 1.65 °C) and water level (0.04 – 0.08 m; Figure 1.13).

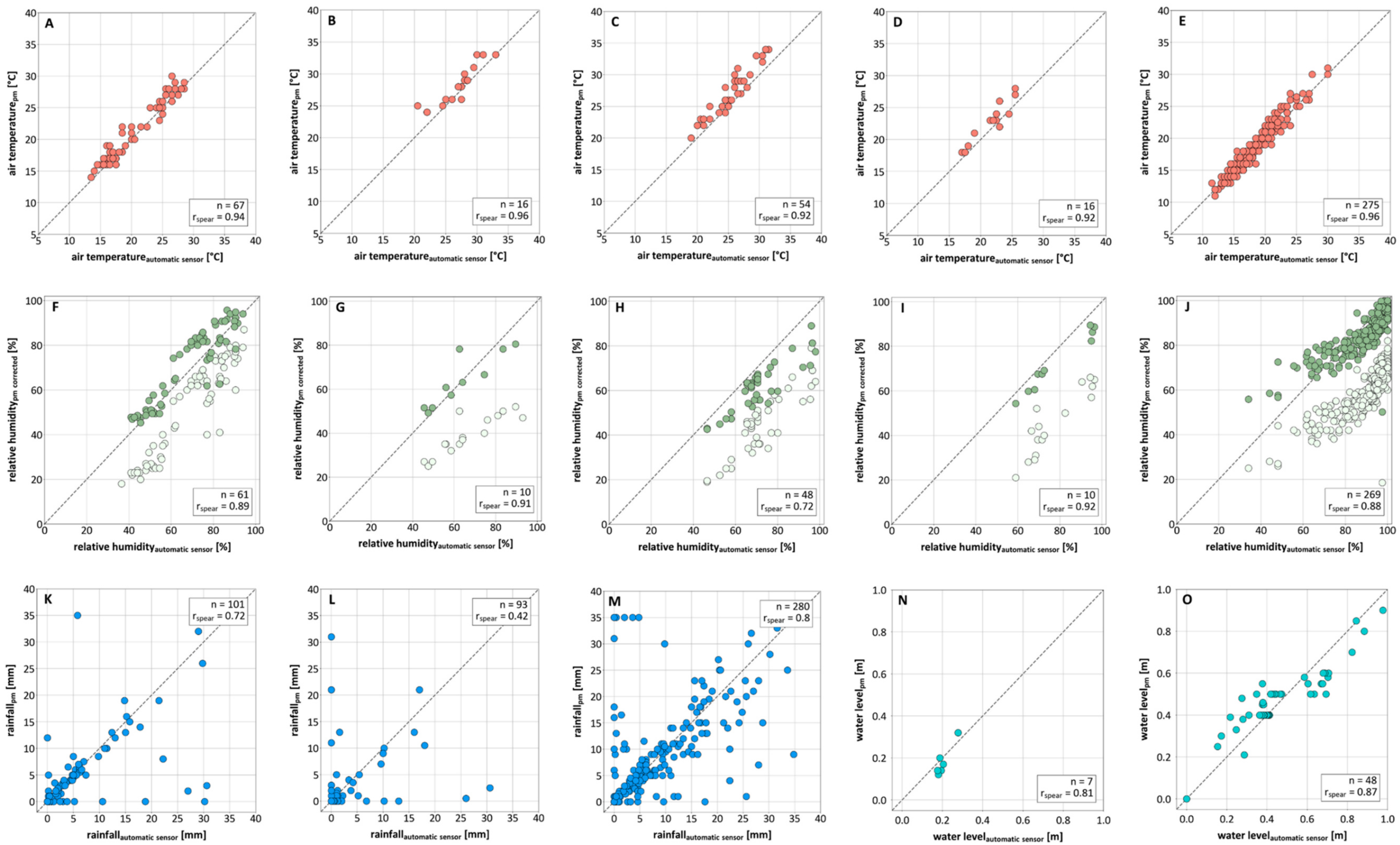


Figure 1.13: *Distribution of PM versus automatic sensor data (n = number of measurements compared where for F - J only the corrected numbers are presented, rspear = Spearman Rank correlation) for air temperature (Tomebamba = A, Don Tito =B, Finca El Nogal = C, Parque Celaque = D and Nkweseko = E), raw (light green) and corrected (dark green) relative humidity (Tomebamba = F, Don Tito = G, Finca El Nogal = H, Parque Celaque = I and "Nkweseko = J), rainfall (Tomebamba = K, Don Tito = L and Nkweseko = M) and water level (Quebrada Santul = N and Rio Arcilaca = O) (adapted from Mitze et al. (2026)).*

For relative humidity a high deviation up to MAE = 31.69 % was detected, which was corrected using linear regression, which in turn resulted in moderate deviation with MAE ranging from 5.45 - 9.50 %. Errors for rainfall were also high with MAE ranging from 2.55 to 3.10 mm, where rainfall was predominantly underestimated.

These findings of chapter 3 were more ambivalent, revealing successful, moderate and rather unsuitable methods for measuring various hydrometeorological parameters within a PM framework. Looking at the users, frequent participants, measured more precisely and their overall contribution was more critical as they submitted more than 80% of all measurements. Looking at the individual parameters, air temperature was the most promising, as it was measured with the lowest deviation. Relative humidity on the other hand was only measured in moderate quality and required correction, as all analog hygrometers used, had a substantial offset. Apart from that, both sensors, analog thermometers and hygrometers, have proven to be sensitive to intense sunlight. Measuring water levels was slightly more complicated for both participant groups. Nonetheless, measurements taken with analog water level gauges were of good quality. This mostly aligns with results from previous projects (Seibert et al., 2022; Weeser et al., 2018). Rainfall could not be measured in satisfactory quality, as it was often underestimated, which could be linked to two main reasons: the limited capacity of the rain gauges used (35 mm) and the irregular measurement frequency which may have increased the error due to evaporation. For rainfall in the context of PM, other similarly simple and cheap, but more accurate alternatives to regular rainfall monitoring were found, for instance the soda bottle-based measurements by the Smartphones4Water network with the self-made rain gauges having a measurement error of only -2.9% (Davids, Devkota, et al., 2019; Tran et al., 2025). In general, no analog sensor for air temperature and relative humidity were tested with a PM approach before, only simple electrical sensors, for example in school contexts, with similar or slightly better overall accuracies (N. Barros et al., 2024; Loglisci et al., 2024).

1.5.3 Objective 3: Bias correction using PM data

In order to analyze if PM data can be used to successfully bias correct large-scale air temperature data in remote tropical mountainous regions, analog air temperature measurements from the HydroCrowd project were applied to train regression models (Chapter 4). T_{rs} data was trained on the corresponding T_{pm} data for each of the three selected stations with varying amounts of available measurements (Tomebamba = 67, Honduras = 23 and Nkweseko = 275) using linear regression.

At all stations, T_{pm} was unevenly distributed throughout the day. There were no measurements available for model training, particularly between midnight and 6 a.m. A statistical analysis revealed significant differences between T_{pm} and T_{rs} for all stations ($p < 0.001$). While for Nkweseko the range of the measured T_{pm} was slightly wider than for the corresponding T_{rs} data, range and mean values for Tomebamba and Don Tito differed more clearly from one another. For instance, the maximum value of T_{pm} and T_{rs} for Tomebamba differed by almost 13 °C (Nkweseko: 3°C and Don Tito: 6°C). However, all stations showed a comparable high degree of correlation: Tomebamba ($r_{spear} = 0.83$), Don Tito ($r_{spear} = 0.85$) and Nkweseko ($r_{spear} = 0.86$).

The validation using T_{as} showed that, the bias corrected air temperature (T_{bc}) had a significantly lower deviation from automatic sensor measurements for all locations ($p > 0.001$) than the original values (T_{rs}). Overall, bias-correction resulted in lowered air temperature for Nkweseko, while temperatures were substantially increased for Tomebamba and Don Tito (Figure 1.14).

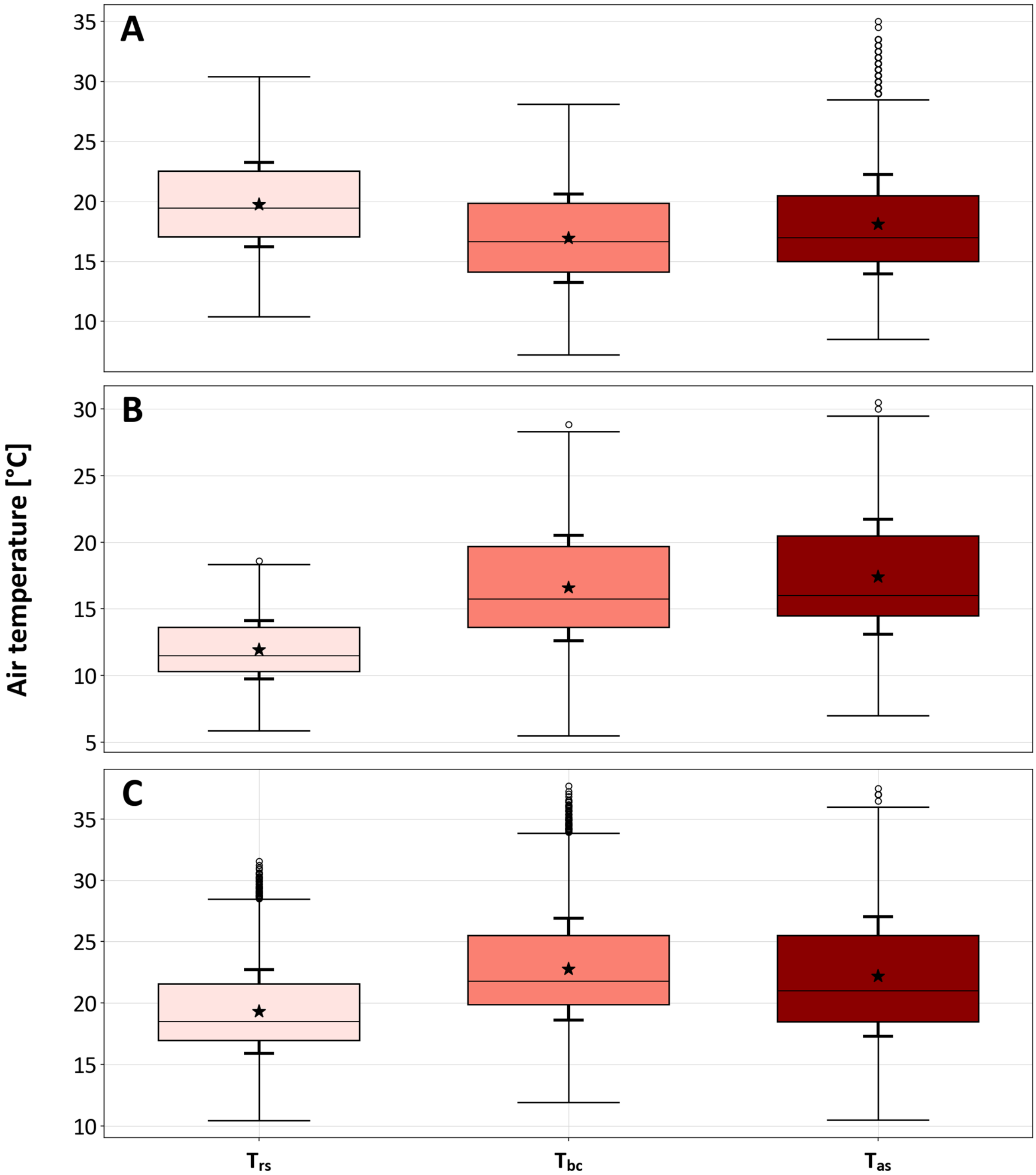


Figure 1.14: *Boxplots for the validation data at the three stations (A) Nkweseko, (B) Tomebamba and (C) Don Tito (T_{rs} = ERA5-Land 2m air temperature data, T_{bc} = bias corrected ERA5-Land air temperature and T_{as} = automatically measured air temperature). The black star indicates the mean value and the thick smaller whiskers standard deviation (adapted from Mitze et al. (under review)).*

Furthermore, validation showed a substantial improvement for Tomebamba and Don Tito, while it was only moderate for Nkweseko (table 1.2) with the lowest deviation and highest R^2 was found for Don Tito with MAE = 1.54 °C and R^2 = 0.83.

Table 1.2: *Validation metric results for the three stations, with n = number of compared values, $R^2_{rs.}$ = baseline results using T_{rs} and R^2_{bc} = bias correction results using T_{bc} (adapted from Mitze et al. (under review)).*

Station	n	R^2_{rs}	MAE_{rs} (°C)	$RMSE_{rs}$ (°C)	R^2_{bc}	MAE_{bc} (°C)	$RMSE_{bc}$ (°C)
Nkweseko	12,934	0.55	2.27	2.77	0.62	1.98	2.54
Tomebamba	4,682	-0.98	5.49	6.09	0.73	1.76	2.27
Don Tito	10,080	0.46	2.95	3.57	0.83	1.54	1.97

A further analysis of the individual hours of the day showed that during the daytime, especially in the late afternoon, the deviations were still higher at all stations, while during the night the bias correction worked better.

The results of Chapter 4 demonstrated with examples from Ecuador, Honduras and Tanzania that PM air temperature data can be used to bias correct large-scale remote sensing data, such as ERA5-Land air temperature. As this dataset offers global coverage, the approach using PM air temperature data provides a considerable alternative for bias correction in regions where the coverage with automatic hydrometeorological stations, is low. Despite the varying accuracy of the bias correction in the different study regions, this research provides a novel approach, as so far only professionally measured air temperature data was used to bias correct ERA5-Land data and no study region in mountainous tropical areas were examined (Dhawan et al., 2024; Niazkar et al., 2023).

1.6 Findings and synthesis

Under the current circumstances of increasing amounts of environmental data (lubis et al., 2025) with unequal spatial distribution worldwide (Buytaert et al., 2016; Chacon-Hurtado et al., 2017; Krabbenhoft et al., 2022; Mishra & Coulibaly, 2009; Walker et al., 2016), the urgency of alternative data sources for regions with inadequate coverage is emphasized. As PM usually does not provide continuous data at regular temporal intervals, it is necessary to explore how much data is required for further applications and which fraction of this data is more useful (Chapter 2) or even suitable to improve existing datasets that provide continuous data but might lack in accuracy (Chapter 4). Also, it is necessary to examine how participants differ in their ability to measure with low-cost sensors and how suitable these sensors are in terms of measurement accuracy (Chapter 3). The findings of the individual chapters demonstrate that increased research into participatory monitoring led to valuable insights of 1) a better understanding of how much PM data is required for further hydrological assessment, 2) which parameters can be measured accurately using simple analog sensors considering different participants and 3) how only a few PM measurements may be sufficient to improve the accuracy of existing large-scale remote sensing datasets. All these novel insights might help to improve the hydrometeorological data situation in the Global South and consequently could also contribute to various sustainable development targets.

1.6.1 Recommendations for further research

While this dissertation provides relevant insights that can be used to develop effective PM projects, there is scope for further research and development of methods for PM.

In the study in chapter 2 only one catchment in a specific tropical mountainous environment in Kenya was used to analyze how much water level data is required for satisfactory model calibration. Further verification from other, particularly more complex catchments, is strongly recommended to confirm the suitability of a low number of PM water level measurements for reliable hydrological model calibration. As similar characteristics haven been indicated in other studies with multiple catchments using data strongly correlated with PM water level data such as discharge or regularly measured water levels (Jian et al., 2017; Juston et al., 2009; Seibert & Beven, 2009), it is likely that similar results will be found in other catchments. Since calibration with water level data brings certain limitations, such as unclear water volumes, it is also highly recommended to use tools, such as the simple water balance filter introduced by Weeser et al. (2019) applied in this study, in order to reduce uncertainty of the modelling approach and ensure realistic absolute water volumes. Given the fact that only a low number of measurements from a specific period is required, this research provides advantageous information for planning efficient future participatory monitoring approaches which offer alternative ways to analyze water flows if regular measurements, such as automatically measured water levels or discharge are not available.

Several aspects of the use of analog sensors in chapter 3 showed that some sensors require special precautions while others are not suitable at all for PM. Shielding and ventilation against overheating were used in the construction of the wooden signposts for the analog thermometers and hygrometers. The shielding proved to be inadequate for proper protection against intense sunlight, which lead to distorted values for up to 20 % of the measurements at some stations. It is therefore crucially for both sensors, to ensure even better shielding against sunlight. Other hygrometers should be tested as well, as the hygrometers used in this research provided only moderate accuracy after correction of a high deviation that has not been conclusively explained on all hygrometers. When considering the main problems using the analog rain gauges it is also highly recommended to use bigger rain gauges, which in the best case also offer some kind of protection against evaporation, like the ones successfully tested by Smartphones4Water in Nepal and Vietnam (Davids, Devkota, et al., 2019; Tran et al., 2025). Therefore, the use of simple analog sensors for water level and air temperature should be expanded in future PM projects, while for relative humidity other sensors should be tested and for rainfall bigger rain gauges should be preferred. When considering the participation of different users, it becomes clear that frequent participants are of outstanding importance. They provide more accurate measurements and, more importantly, measure much more frequently. Since the success of a PM project depends on participation, future projects should be tailored to the local population, for instance to local schools, who are more likely to participate frequently.

The comparison of PM and automatically measured air temperature (see chapter 3 and 4) showed that PM data deviate considerably from automatically measured data. This is one explanation why there is likely still a deviation in the bias corrected data presented in chapter 4. The effect of the deviation should be further analyzed testing both, PM and automatically measured data for the bias correction model training. While linear regression performed well in this study, the performance of other bias correction methods, especially machine learning, should be explored further, too. Other research obtained even better accuracies with more complex models, whereby more precise automatically measured data was used (Dhawan et al., 2024; Niazkar et al., 2023). The influence of the location of the PM station within the ERA5-Land pixel (9 km) on the performance of the bias correction and methods to address this could be further investigated as well. Nevertheless, considering that significant improvements have already been achieved with just a few measurements in two of three stations, there is reason to continue testing and developing this approach to improve existing large-scale datasets for air temperature, such as ERA5-Land. Since the air temperature worked predominantly well, a similar approach for large-scale bias correction could be used to test other hydrometeorological parameters in the next step. For example, relative humidity could be tested, despite its lower accuracy, which it is also available in the ERA5-Land dataset.

1.6.2 Outlook and conclusions

Based on these findings, the following conclusions can be drawn:

1. Some hydrometeorological parameters, such as air temperature and water level can be measured by participants using analog low-cost sensors with good accuracy…
2. …but not all sensors are suitable for precise data collection within PM approaches as the moderate quality of relative humidity and inaccurate precision of rainfall measurements showed.
3. Overall, participants in PM projects, such as HydroCrowd, are able to measure different hydrometeorological parameters accurately for the most part…
4. …however, frequent participants such as residents can be considered to be more crucial for successful data collection in hydrometeorological PM as they collect more and slightly more precise data.
5. Only a few PM water level measurements are sufficient for hydrological model calibration if collected at the right time of the hydrological cycle, preferably the wet season with high water level and high discharge variability.
6. Using PM data for bias correction of large-scale remote sensing air temperature data, such as ERA5-Land, might be a reasonable method to correct large-scale values if no automatic weather station data is available.

Finally, the benefits of PM depend to a high degree on the overall participation. This is why strategies need to be developed or applied to ensure consistent and long-term participation, even beyond short-term research projects.

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

This chapter is published in *PLOS Water* as:

Mitze, F., Jacobs, S. R., Breuer, L., Zeballos, J. C., 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

2.1 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 of data 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.

2.2 Introduction

Climate and land use changes can greatly impact the hydrological cycle (IPCC, 2023a). 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 (United Nations, 2015), 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 (United Nations, 2022), which lack the resources to ensure that the goals are met (Buytaert et al., 2014; Daw et al., 2011; Fankhauser & McDermott, 2014; Ruhi et al., 2018).

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 (Fankhauser & McDermott, 2014; Ruhi et al., 2018; Buytaert et al., 2014). 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 (Krabbenhoft et al., 2022).

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 (Buytaert et al., 2014; Njue et al., 2019). In addition to using remotely sensed data for meteorological and discharge data (Gleason & Durand, 2020; Mu et al., 2007; Schmugge et al., 2002; X. Xu et al., 2014), studies have also investigated the use of cameras (Lee et al., 2010; R. Meier et al., 2022), social media (Rosser et al., 2017), the mobile phone network (Gosset et al., 2016), or private web-connected weather stations (de Vos et al., 2020; Hahn et al., 2022) 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 (Buytaert et al., 2014; Dickinson et al., 2012). Participatory monitoring (PM) can be considered as a subcategory of this, where laypersons only collect data, while data analysis remains with experts (Danielsen et al., 2009; Villaseñor et al., 2016). 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 (Ochoa Tocachi, 2019; Uprety et al., 2019; Weeser et al., 2018, 2019). Potential challenges include the complexity of the measurements (Lowry et al., 2019) or temporal and spatial coverage of PM data, which may vary due to different sampling designs and willingness to participate (Njue et al., 2019). 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 (Fienen & Lowry, 2012; Lowry et al., 2019; Weeser et al., 2018) or virtual staff gauges using a smartphone application (Etter, Strobl, Seibert, et al., 2020; Scheller et al., 2024; Seibert et al., 2019). As discharge is directly related to water level, these data can also be used for hydrological modelling (Etter, Strobl, Seibert, et al., 2020; Jian et al., 2017; Seibert & Vis, 2016; van Meerveld et al., 2017; Weeser et al., 2019). For example, Weeser et al. (2019) 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 (Njue et al., 2019).

Several studies have shown that model calibration does not necessarily require continuous data (N. R. McIntyre & Wheater, 2004; Seibert & Beven, 2009; van Meerveld et al., 2017; Weeser et al., 2019). 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 (Ficchì et al., 2016; Huang et al., 2019; Juston et al., 2009; Krebs et al., 2014). Unusual events or high flow have been described as more valuable for calibration than base flow conditions (Jian et al., 2017; Juston et al., 2009; Singh et al., 2012; Singh & Bárdossy, 2012; van Meerveld et al., 2017). Artificial reduction of datasets has shown that smaller subsets of the whole data set often provide enough information for satisfying model results (Etter, Strobl, Seibert, et al., 2020; Jian et al., 2017; Juston et al., 2009, 2009; N. McIntyre et al., 2005; Perrin et al., 2007; Seibert & Beven, 2009; Seibert & McDonnell, 2015). It should be noted, however, that the results varied widely depending on the choice of model parameters (Seibert & Beven, 2009). 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 (Seibert & Vis, 2016; van Meerveld et al., 2017), the same is true for PM water level data (Weeser et al., 2019). So far, only artificial PM data sets derived from automatic sensor data have been used for such analyses (Etter, Strobl, Seibert, et al., 2020; Mazzoleni et al., 2017). 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.

2.3 Material and methods

2.3.1 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 (Weeser et al., 2018). Kenya’s largest indigenous closed-canopy forest, the Mau Forest, makes up a large proportion of the area (Weeser et al., 2018). This area provides critical water-related ecosystem services, such as water storage and river flow (Benn & Bindra, 2011) which are threatened by continued forest loss and degradation

(Brandt et al., 2018). 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 (Weeser et al., 2019).

Most of the catchment is characterized by smallholder farms with a mix of annual crops, woodlands, forests, and pastures (Weeser et al., 2019). The soils in the study area are characterized as deep and well drained and fall under the classification of Humic Nitisols and Mollic Andosols (ISRIC, 2007). Due to the movement of the Intertropical Convergence Zone, the rainfall pattern of the area is characterized by two rainy seasons per year (Hills, 1979).

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 to 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 (Hargreaves & Samani, 1985). At the outlet, participants manually measured the water level using a physical water level gauge (Weeser et al., 2018). 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 (Weeser et al., 2018). 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 2.1) to calculate discharge as described by Jacobs et al. (Jacobs, Weeser, et al., 2018).

$$\boldsymbol{Q} = \begin{cases} 0.0973 - 1.892 \cdot h + 6.923 \cdot h^2, & h \geq 0.236\, m \\ 0.651 \cdot h^2, & h < 0.236\, m \end{cases} \qquad R^2 = 0.98, \qquad \text{(Equation 2.1)},$$

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 (Breiman, 2001; Geurts et al., 2006), using the open source Python package Scikit-learn (Pedregosa et al., 2011). 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 ($Pearson_{q-wl}$ = 0.96 and $Pearson_{wl-pm\_wl}$ = 0.96). The hydro-meteorological data of the simulation period are shown in Figure 2.1.

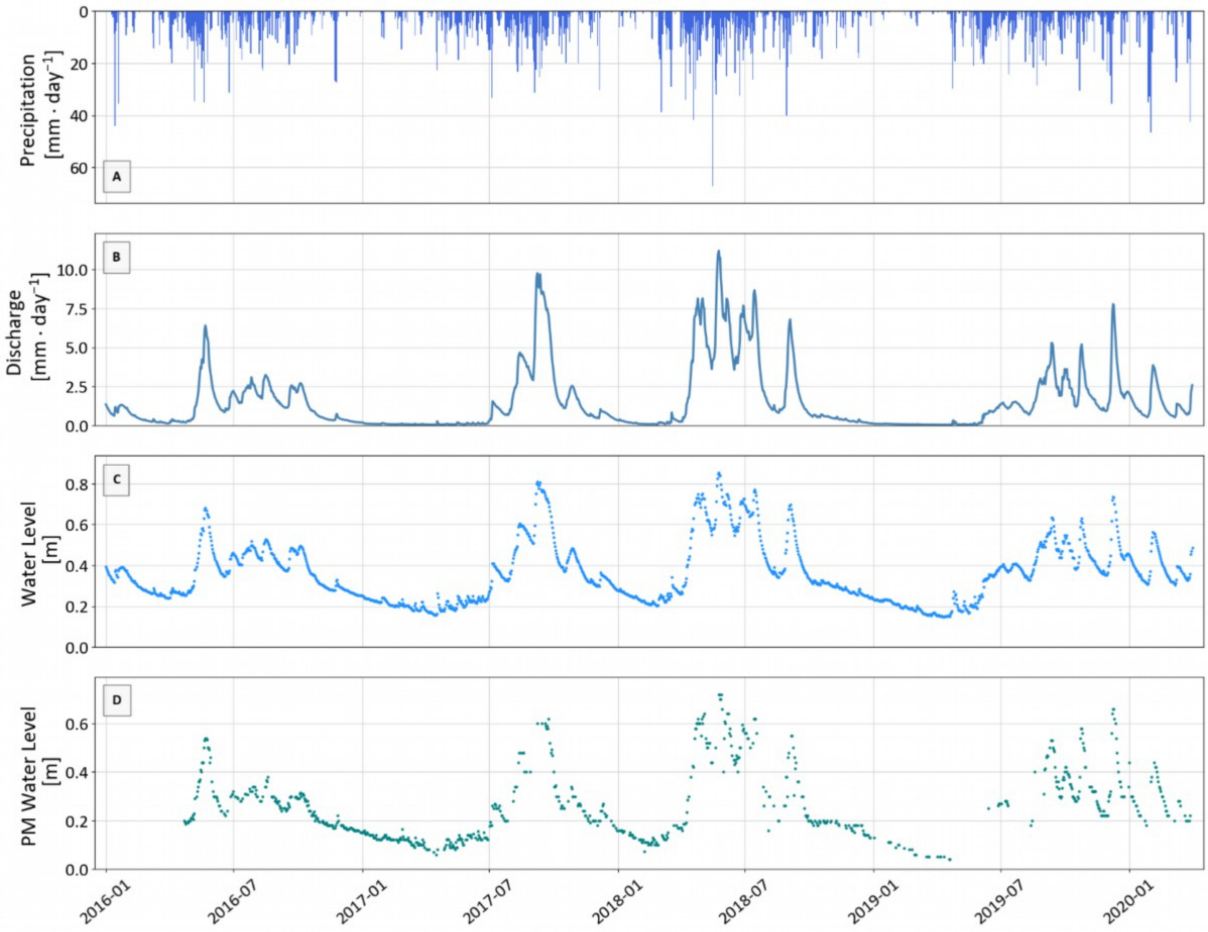


***Figure 2.1**: Measured hydro-meteorological data from January 2016 to April 2020: Daily mean areal weighted precipitation using Thiessen-Polygons (A), discharge (B), radar-based water level (C) and participatory monitoring (PM) water level (D).*

2.3.2 Model setup

The rainfall-runoff model with daily time steps follows the concept of Weeser et al. (2019), who used the Python package "Catchment Modelling Framework" (CMF) (Kraft et al., 2011) to develop a 5-parameter lumped conceptual model (Figure 2.2) based on the understanding of the hydrological processes in the catchment described in Jacobs et al. (2018).

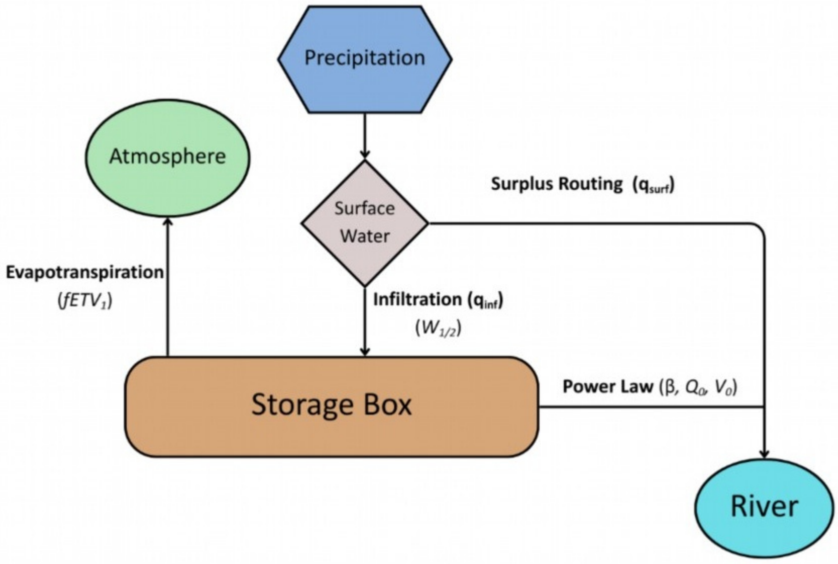


Figure 2.2: *Schematic overview of the model structure, where CMF processes are written in bold and their corresponding parameters in italic letters (own illustration adapted from Weeser et al. (2019)).*

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 2.2) or to the atmosphere by evapotranspiration calculated based on potential evapotranspiration. Surface runoff and water discharging from the storage box sums up to streamflow.

$$\boldsymbol{q_{out}} = Q_0(\frac{V}{V_0})^{\beta} \qquad \text{(Equation 2.2),}$$

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 2.1), originally defined by Weeser et al. (Weeser et al., 2019) and based on expert knowledge from previous comparable model applications and exploratory data analysis was used for this study.

Table 2.1: *Model parameters and a priori ranges.*

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

Model parameter uncertainties were evaluated using a Latin Hypercube-based approach (McKay et al., 1979) with 10^5 parameter sets using the open source Python package SPOTPY (Houska et al., 2015).

2.3.3 Model calibration and validation

The dataset was divided into a warm-up period (1st January 2016 to 31st March 2016), a three-year calibration period (1st April 2016 to 31st March 2019) and a one-year validation period (1st April 2019 to 31st March 2020) (Table 2.2).

Table 2.2: *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 to 31 March 2019	579	1,494	14.5	1,538
01 April 2019 to 31 March 2020	566	1,904	14.5	1,488

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

The Kling-Gupta-Efficiency (KGE, Equation 2.3) (Gupta et al., 2009) was used to evaluate model performance.

$$\boldsymbol{KGE} = 1 - \sqrt{(r-1)^2 + (\frac{\sigma_{sim}}{\sigma_{obs}} - 1)^2 + (\frac{\mu_{sim}}{\mu_{obs}} - 1)^2} \quad \text{(Equation 2.3)},$$

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) (Nash & Sutcliffe, 1970) to overcome problems such as volume balance errors in basins with high runoff variability (Gupta et al., 2009). To calibrate the model on water level data, the Spearman Rank correlation (R_{spear}) (Spearman, 1904) was used (Seibert & Vis, 2016) (S1

Table). R_{spear} ranges between -1 and 1 and reflects the similarity between the dynamics of observed discharge and water levels (Seibert & Vis, 2016). 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 (Seibert & Vis, 2016). 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 (Seibert & Vis, 2016). This means that setting a threshold for behavioral parameter sets 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. (2019). For the validation period, all accepted parameter sets from the calibration period were used.

2.3.4 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. (2019) introduced an annual water balance filter. This was considered necessary, because R_{spear} does not reflect absolute discharge volumes (Seibert & Vis, 2016).

The filter was applied according to the three-year calibration period (1st April 2016 to 31st March 2019). The balance was calculated from observed precipitation minus mean actual evapotranspiration (ET_{act}) retrieved from the MODerate Resolution Imaging Spectroradio-meter (MODIS). An ET_{act} of 2,995 mm for the three years was calculated using the MOD16A2 Collection 6 Global Evapotranspiration Product (Running et al., 2021). This dataset provides evapotranspiration data in the form of scaled tiles, with an eight-day temporal resolution and a spatial resolution of 500 meters (Running et al., 2021). 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 (Senay et al., 2011).

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 (Alemayehu et al., 2017; Velpuri et al., 2013; Weeser et al., 2019). 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.

2.3.5 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 (Breiman, 2001; Geurts et al., 2006; Pedregosa et al., 2011) 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 (Figure 2.3). Given this, we consider the predicted values to be comparable to the original measurements provided by the participants.

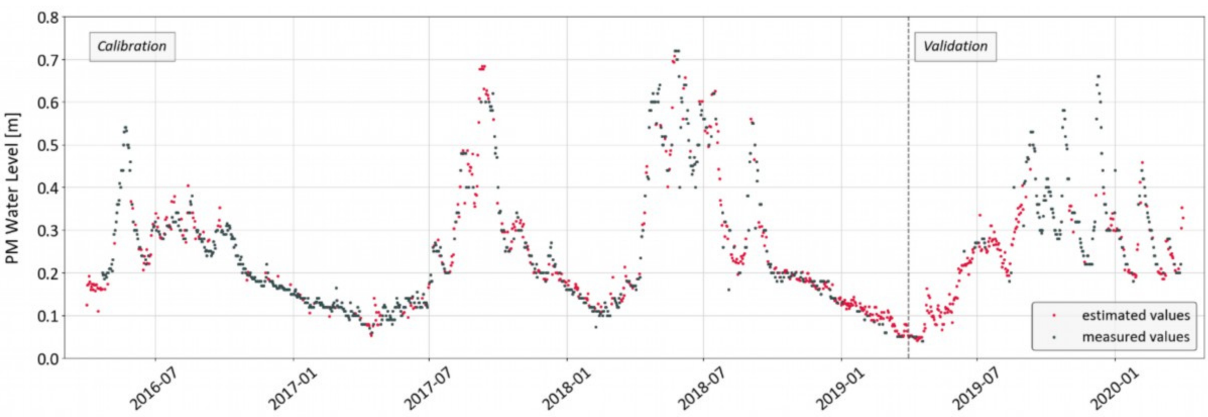


Figure 2.3: *Baseline PM dataset of daily water level data in the simulation period (April 2016 – April 2020), showing the measured PM values and values estimated through random forest regression.*

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 (Jian et al., 2017; Juston et al., 2009; Singh et al., 2012; Singh & Bárdossy, 2012; van Meerveld et al., 2017), 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 (Figure 2.4).

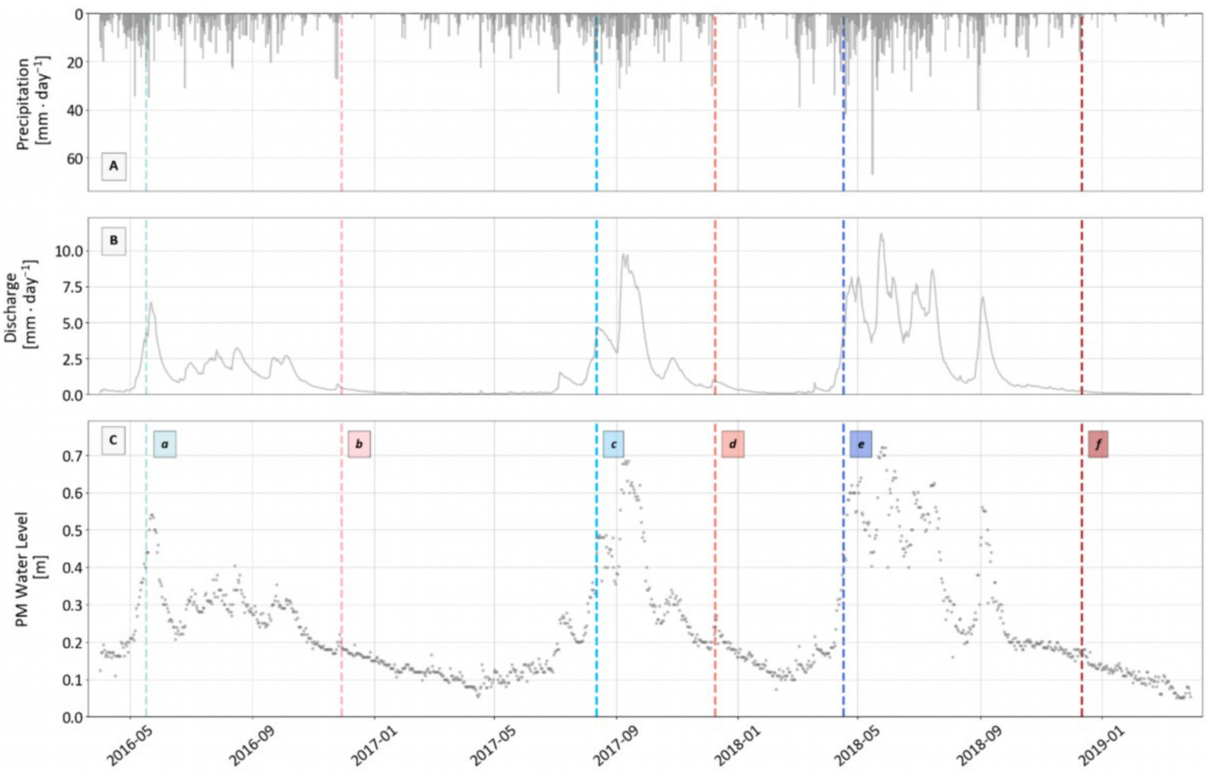


Figure 2.4: *Overview of A) precipitation, B) discharge and C) baseline PM water level data in the calibration period (April 2016 – April 2019) with six different scenario starting points, n = 3 in wet seasons (a, c and e) and n = 3 in dry seasons (b, d and f).*

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 Figure 2.4).

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 2.3). 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.

Table 2.3: *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**

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 (Clark et al., 2021; Gupta et al., 2009; Knoben et al., 2019; Moriasi et al., 2007). 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.

2.4 Results

2.4.1 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.

2.4.2 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 (Figure 2.5).

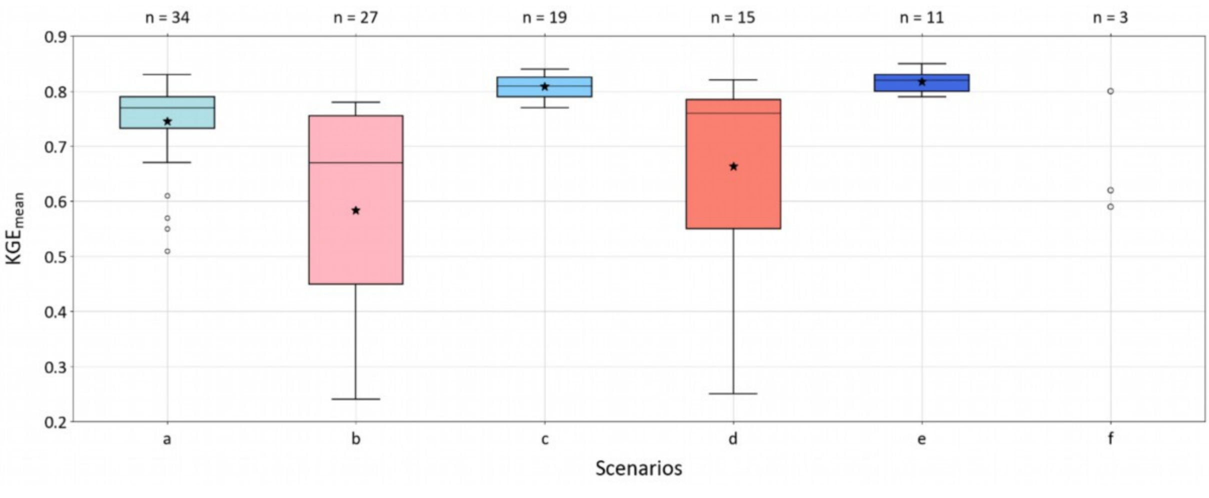


Figure 2.5: *Boxplots comparing KGE_{mean} values of every scenario group (a-e) for calibration, where the black star is showing the mean. For scenario group f the individual KGE_{mean} values are shown due to the small sample size. Note: all model runs for scenario b1 were discarded, as none fulfilled the water balance filter.*

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 (Mann & Whitney, 1947), 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 (Figure 2.6), 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).

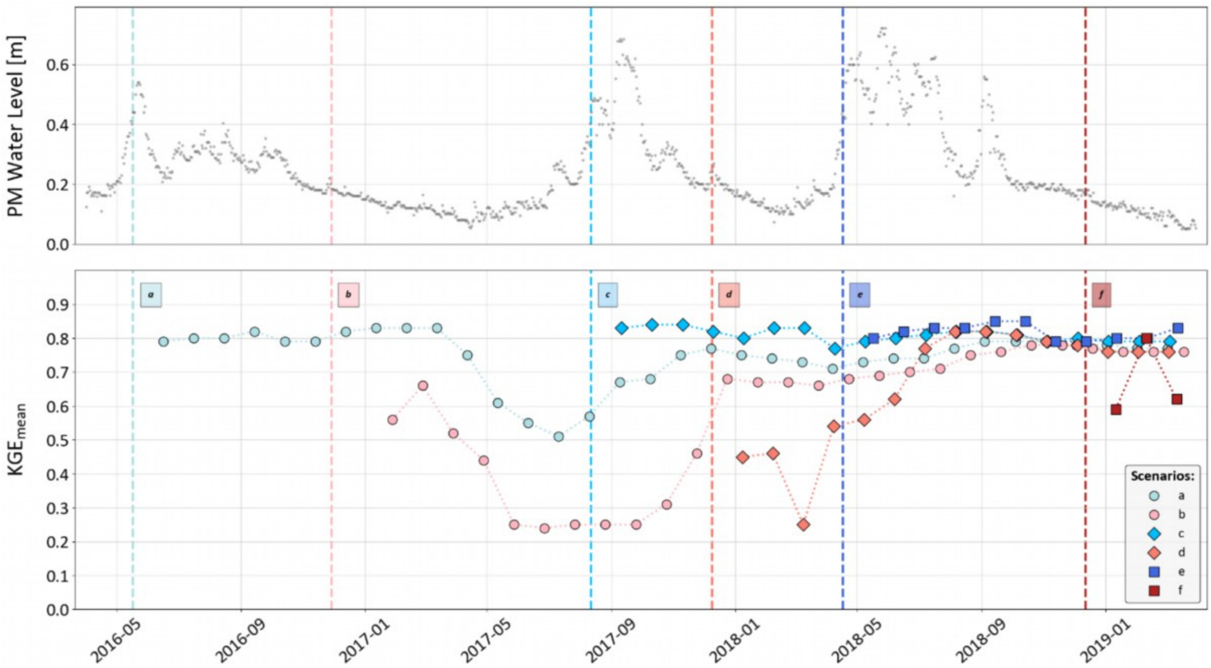


Figure 2.6: *Results for each scenario calibration (lower panel) and the corresponding baseline PM water level data (upper panel). Each scenario used the baseline PM water levels from at the specific threshold (a-f) until the circle of the corresponding scenario. This circle represents the KGE_{mean} of each scenario. Note: model runs for scenario b1 were discarded due to not fulfilment of the water balance filter.*

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 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 to

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 (Figure 2.7). 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.

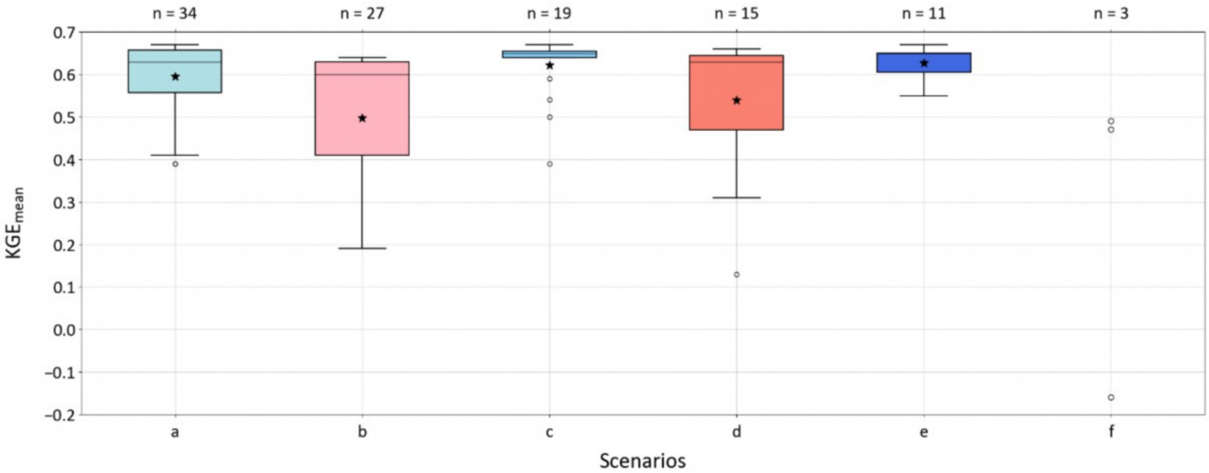


Figure 2.7: *Boxplots comparing KGE_{mean} values of every scenario setting (a-e) for validation, where the filled black dot symbolizes the mean. For scenario group f the individual KGE_{mean} values are shown due to the small sample size. Note: As for calibration all model runs for scenario b1 were discarded due to not fulfilment of the water balance filter.*

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 (Figure 2.8).

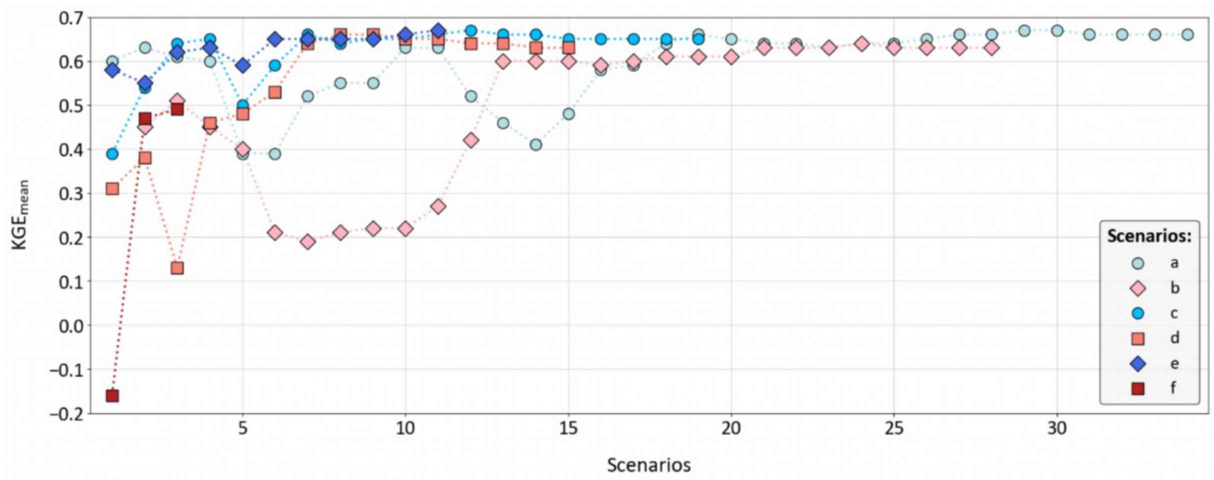


Figure 2.8: *Validation results for the different scenario parameter sets and the resulting KGEmean value. Note: model runs for scenario b1 were discarded due to not fulfillment of water balance filter.*

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.

2.4.3 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.

For wet scenarios, parameters V_0 and $fETV_1$ show clear optimal ranges with 500-1000 mm and 0.1 to 0.2, respectively (Figure 2.9).

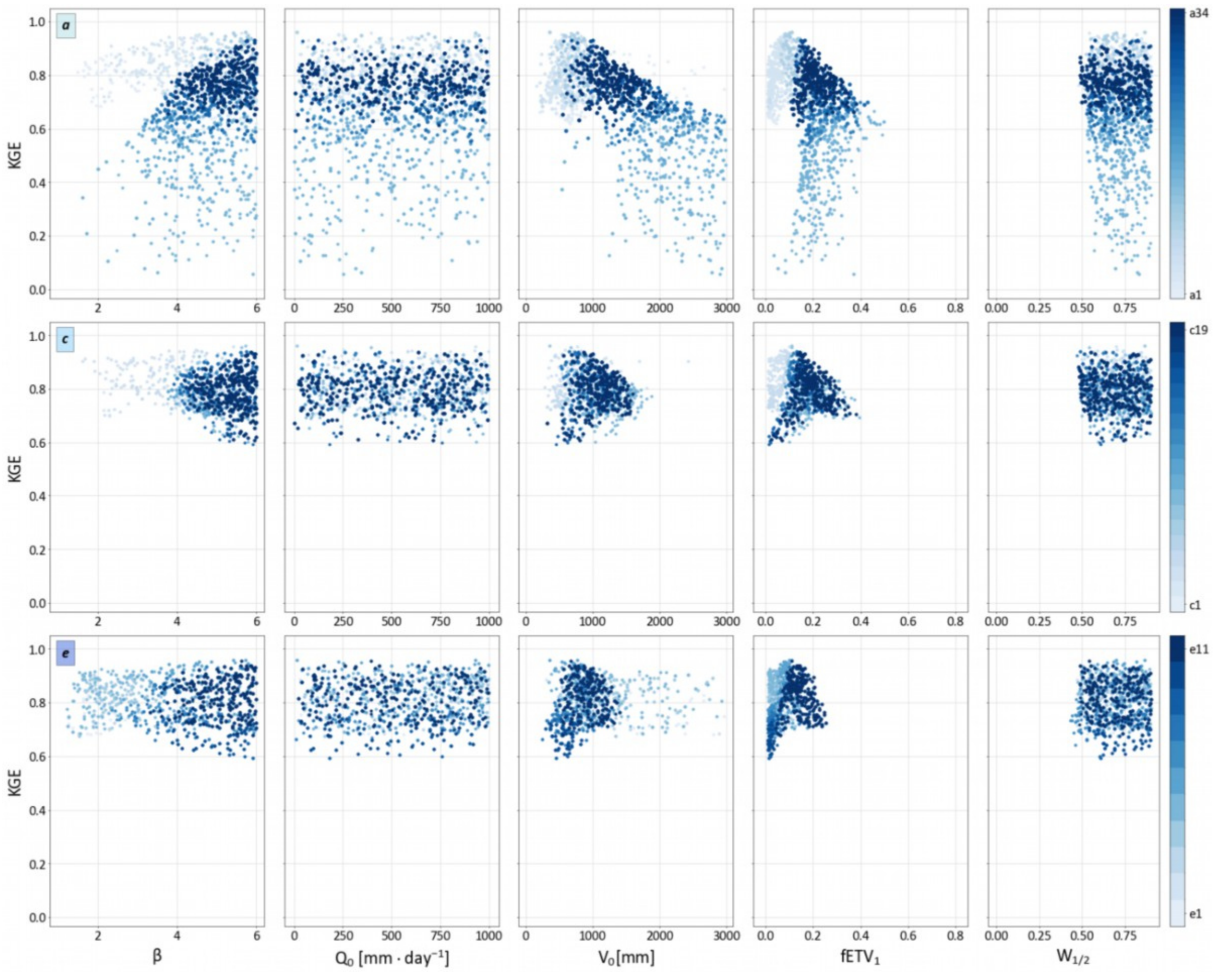


Figure 2.9: *Parameter values and the corresponding KGE values of the filtered respective model runs for wet scenarios based on threshold a, c and e. Different shades indicate the different scenarios.*

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 to 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 (Figure 2.10).

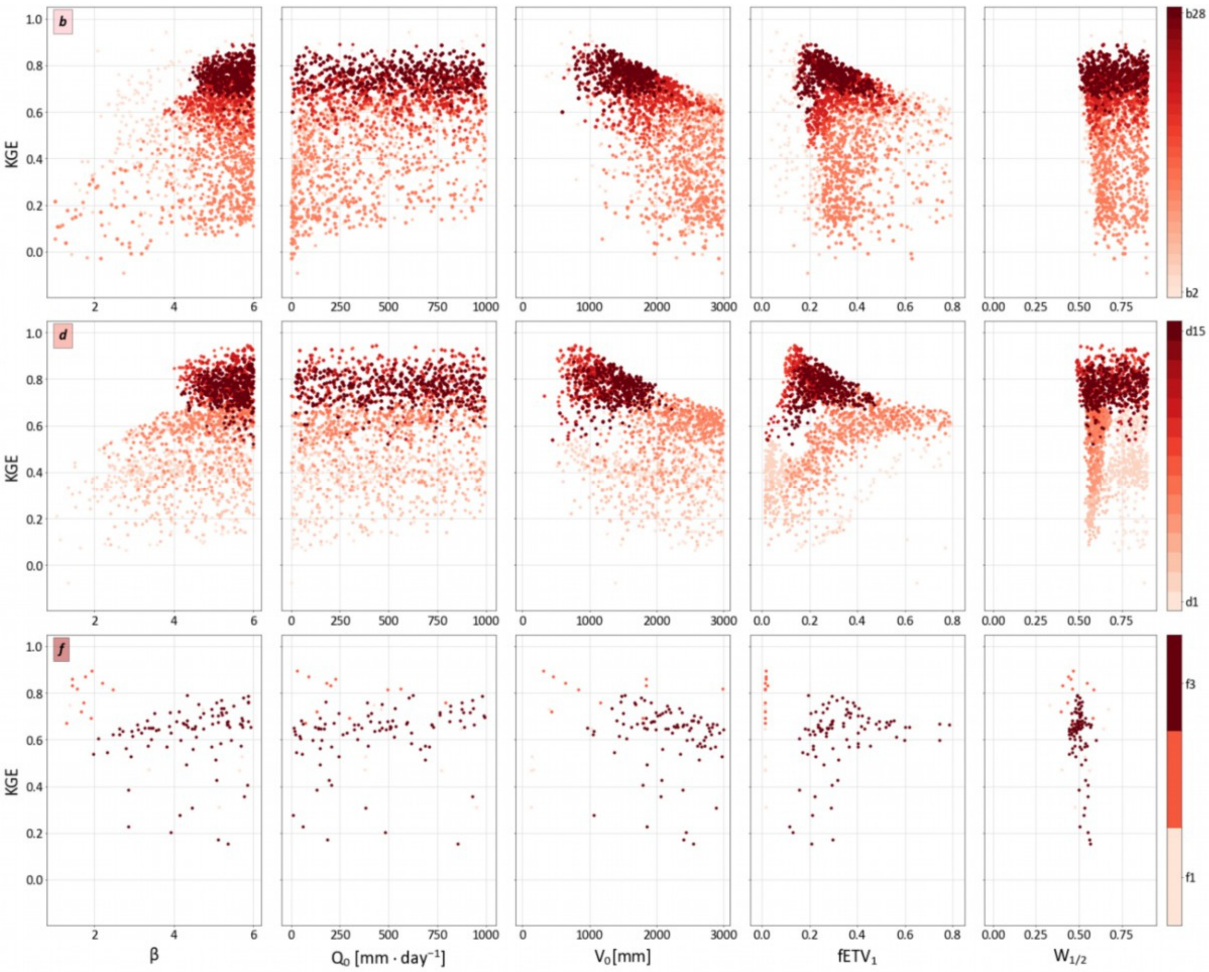


Figure 2.10: *Parameter values and the corresponding KGE values of the filtered respective model runs for dry scenarios based on threshold b, d and f. Different shades indicate the different scenarios.*

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 to 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.

2.5 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 (Njue et al., 2019; United Nations, 2022). There is also still considerable potential (Sachs et al., 2024) 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) (United Nations, 2015), as these countries might lack the resources to ensure it (Buytaert et al., 2014; Daw et al., 2011; Fankhauser & McDermott, 2014; Ruhi et al., 2018). 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.

2.5.1 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 (Etter, Strobl, Seibert, et al., 2020; Jian et al., 2017; Juston et al., 2009; Seibert & Beven, 2009; van Meerveld et al., 2017). For example, Seibert & Beven (2009) 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. (2020), 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 (Luo et al., 2012).

2.5.2 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. (2017) 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 (2016) 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 (Huang & Bardossy, 2020; Jian et al., 2017; Seibert & Beven, 2009; van Meerveld et al., 2017). For example, a study by Huang and Bardossy (2020) 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 (Huang & Bardossy, 2020).

Jian et al. (2017), identified that water level measurements at the 95th quantile are more suitable for identifying dynamics of discharge. Furthermore, a study by Rahbeh et al. (2011), 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 to 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 modelling (Arnaud et al., 2002; Chaubey et al., 1999). 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 Figure and S2 Figure). Nevertheless, our results strongly indicate that water level measurements from wet seasons are more important for model performance than during dry seasons.

2.5.3 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 (Shin et al., 2015; Silvestro et al., 2015; Wagener et al., 2001), similarities but also differences were found. For instance, Shin et al. (2015) compared a four-parameter IHACRES model (Croke & Jakeman, 2004) and a nine-parameter SIMHYD model (Chiew et al., 2002). 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.

2.5.4 Utility of water balance filter

Since the calibration on water level data does not guarantee correct absolute discharge volumes (Seibert & Vis, 2016) 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 (Weeser et al., 2019; Yu et al., 2023) or the hydrological model was calibrated directly on the remote sensing derived data (Meyer Oliveira et al., 2021).

A major advantage of posterior constraining models using a water balance filter is its ease of application. As stated by Weeser et al. (2019), 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.

2.5.5 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 (Avellaneda et al., 2020; Dasgupta et al., 2022; Starkey et al., 2017; Weeser et al., 2019). 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.

2.6 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.

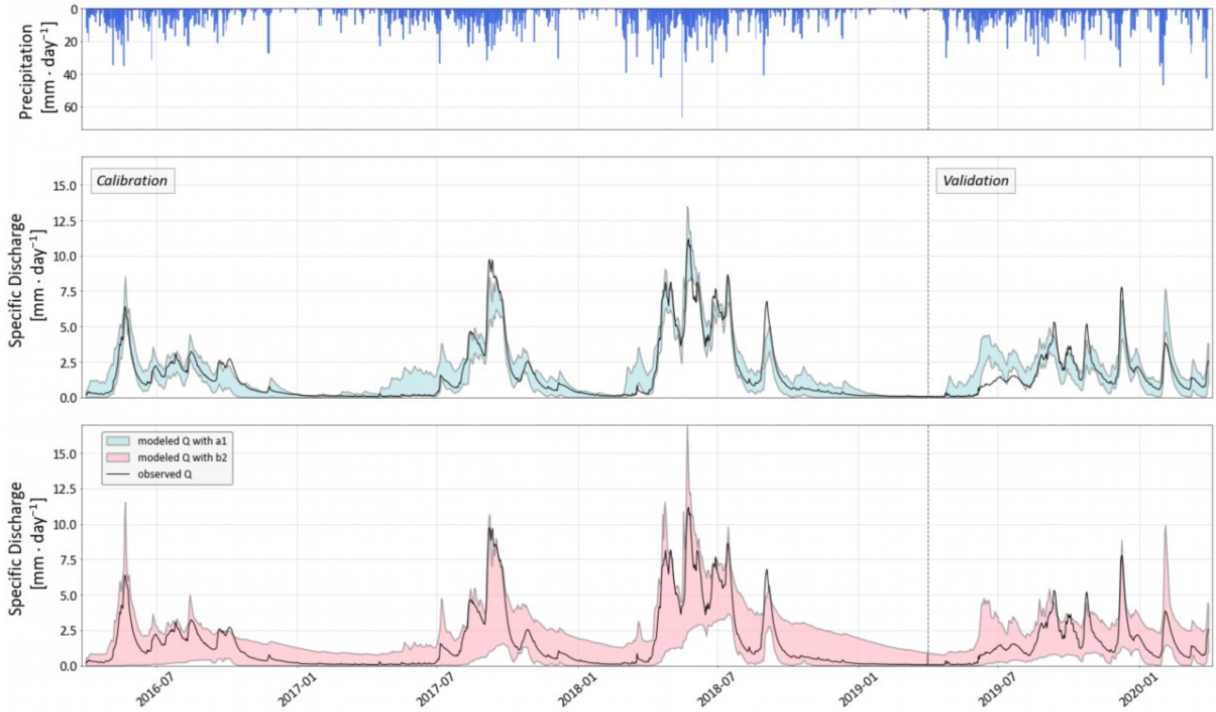
2.7 Acknowledgements

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.

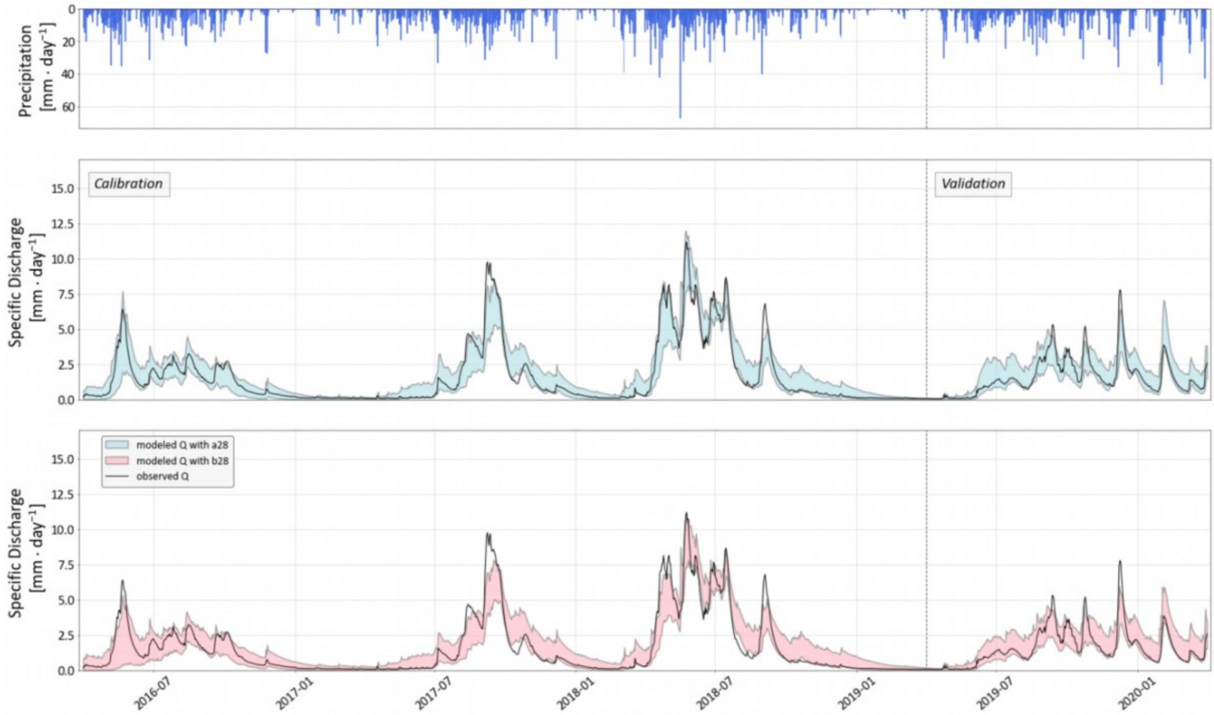
2.8 Data Availability Statement

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

2.9 Supporting information



S1 Figure: *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}.



***S2 Figure:** 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}.*

S1 Table: *Statistical metrics used in the manuscript.*

name	abbreviation	equation	definition	application
Pearson correlation	Pearson r	$Pearson\ r = \frac{cov(X,Y)}{\sigma_X\ \sigma_Y}$	cov = covariance of X and Y, σ_X is the standard deviation for X, σ_Y is the standard deviation for Y	correlation between variables
Kling-Gupta-Efficiency	KGE	$KGE = 1 - \sqrt{(r-1)^2 + (\frac{\sigma_{sim}}{\sigma_{obs}} - 1)^2 + (\frac{\mu_{sim}}{\mu_{obs}} - 1)^2}$	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	Compare model efficiency
Spearman correlation	R_{spear}	$R_{spear} = \frac{cov[R[X], R[Y]]}{\sigma_{R[X]}\ \sigma_{R[Y]}}$	cov = covariance of the rank variables of X and Y, σ_X is the standard deviation for the rank variable of X, σ_Y is the standard deviation for the rank variable of Y	Calibrate model with water level data

3. Validation of analog sensor measurements in hydrometeorological participatory monitoring in various tropical countries

This chapter is accepted for publication in *Frontiers in Earth Science* as:

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Codalli, F., Shagega, F. P., Weeser, B. (2026). Validation of analog sensor measurements in hydrometeorological participatory monitoring in various tropical countries. Frontiers in Earth Science, 14. https://doi.org/10.3389/feart.2026.1721642

3.1 Abstract

As remote tropical mountain regions often lack open data and traditional methods of collecting hydrometeorological data are not always feasible, this study validates an alternative participatory monitoring approach for collecting hydrometeorological data in mountainous regions in Ecuador, Honduras and Tanzania.

Volunteers used analog low-cost sensors to measure air temperature, relative humidity, rainfall and water level. The measurements were validated with photos taken alongside the measurements. Data from selected stations were additionally validated against automatic sensor data using different metrics, such as the mean absolute error (MAE). In addition, errors made by frequent and non-frequent participants were compared, assessing the performance of these two target groups.

In the period between May 2023 and May 2025 a total of 2,982 observations were received, whereby the majority were submitted by frequent participants (84.4 %). A comparison between frequent and non-frequent users showed that the former measured with higher accuracy. The comparison with automatic sensor data showed a correlation for all parameters ranging from 0.42 to 0.96. The best results in terms of accuracy were achieved for air temperature (MAE: 0.74 - 1.65 °C) and water level (MAE: 0.04 - 0.08 m). On the other hand, a high deviation was found for relative humidity (MAE: 16.76 - 31.69 %). This deviation was corrected by applying linear regression, resulting in moderate

deviation (MAE: 5.45 - 9.50 %). Rainfall had a MAE ranging from 2.55 to 3.10 mm. This was mainly attributed to the low measurement frequency and the limited capacity of the rain gauges.

Overall, the study showed ambivalent results, where analog thermometers and water level gauges can be considered the most promising alternatives to automatic sensor measurements. However, the hygrometers only provided moderate measurement quality, while the rain gauges used were too small to cover all rainfall in the periods analyzed.

3.2 Introduction

The Sustainable Development Goal 6 (SDG 6), of the United Nations 2030 Agenda for Sustainable Development, is aimed at ensuring availability and sustainable management of water and sanitation for all until 2030 (United Nations, 2015). In 2022, approximately 2.2 billion people worldwide lacked access to safely managed drinking water, and around half of the global population faced severe water scarcity (United Nations Department of Economic and Social Affairs, 2024). These figures highlight the urgent need for sustainable water management strategies, which are essential now and will become increasingly critical in the future. The majority of critically affected people live in the so-called “Global South” (United Nations, 2022), where climate change already has significant impacts, especially on vulnerable populations (Ngcamu, 2023a; Sen Roy, 2018).

Strategies to cope with climate change impacts on water resources require reliable hydrometeorological data like rainfall, air temperature or water levels. Such data are commonly recorded by automatic weather and gauging stations, but these are costly and complex to maintain. The global distribution of gauging and precipitation monitoring stations is highly disproportionate and is decreasing (Kidd et al., 2017; Krabbenhoft et al., 2022; Ruhi et al., 2018). This particularly affects remote regions in the Global South, often because of the lack of financial resources (Buytaert et al., 2014; Fankhauser & McDermott, 2014; Ruhi et al., 2018). Satellite-based monitoring becomes increasingly relevant due to the possibility of obtaining information where ground observations are complex or not possible. While satellite remote sensing products, for instance for precipitation (Q. Sun et al., 2018; Tang et al., 2020) or land surface temperatures (Li et al., 2023b), have improved over the last decades, the spatial and temporal resolution remain limiting factors (Levizzani & Cattani, 2019; Mao et al., 2021; Marra et al., 2019). Especially orographic precipitation in mountainous areas remains complex to sense (A. P. Barros & Arulraj, 2020; Hemp & Hemp, 2024; Kimani et al., 2017). Collecting water level data in smaller rivers with satellite-based products is challenging or even not possible due to the spatial resolution and accuracy (Bandini et al., 2017; Grimaldi et al., 2016; Musa et al., 2015). Therefore, it is important to explore suitable, cost-effective alternatives to ground observations.

One possible approach is the integration of the general public into the scientific process, which is commonly known as citizen science (Buytaert et al., 2014; Dickinson et al., 2012). A sub-category of this is called participatory monitoring (PM), in which volunteers only participate in order to collect data and information (Danielsen et al., 2009). Various projects have demonstrated PM potentials with different approaches for measuring water level, precipitation and other hydrometeorological parameters (Arienzo et al., 2021; Buytaert et al., 2014; Davids, Devkota, et al., 2019; Scheller et al., 2024; Weeser et al., 2018). While most PM projects have focused on North America or Europe, fewer have been implemented in countries in the Global South (Buytaert et al., 2014; Njue et al., 2019).

In this study, the data collection of the HydroCrowd project is validated. Here, hydrometeorological observations were collected using a novel PM approach in remote mountainous regions in Ecuador, Honduras and Tanzania. A network of weather and water monitoring stations equipped with analog low-cost sensors was installed in these areas. The approach combines different elements of various PM projects, such as rainfall monitoring like in the CoCoRaHS (Reges et al., 2016) and Smartphones4Water (Davids, Devkota, et al., 2019) project, or water level monitoring and smartphone based data transmission previously demonstrated in the CrowdWater (Seibert et al., 2022) project. To obtain an even more comprehensive suite of hydrometeorological parameters, the monitoring is extended to also include air temperature, relative humidity and turbidity. At weather stations, participants measured air temperature, relative humidity and rainfall. At water stations at rivers, the water level and turbidity of the river water were measured. Similar to CoCoRaHS, but unlike Smartphones4Water, we used standardized, commercially available rain gauges for rainfall collection. Unlike CrowdWater, we used physical rather than virtual water level gauges. While local community members (from here on referred to as frequent participants) were targeted as possible participants in Honduras, the main focus in Ecuador and Tanzania was on local and international tourists (non-frequent participants), respectively. For validation purposes, participants were asked to upload photos of the sensors when sending their observation which could be done using the projects smartphone application or the web interface. Additionally, selected stations were equipped with automatic sensors to test the accuracy of the analog sensors. The data collection approach is visualized in figure 3.1.

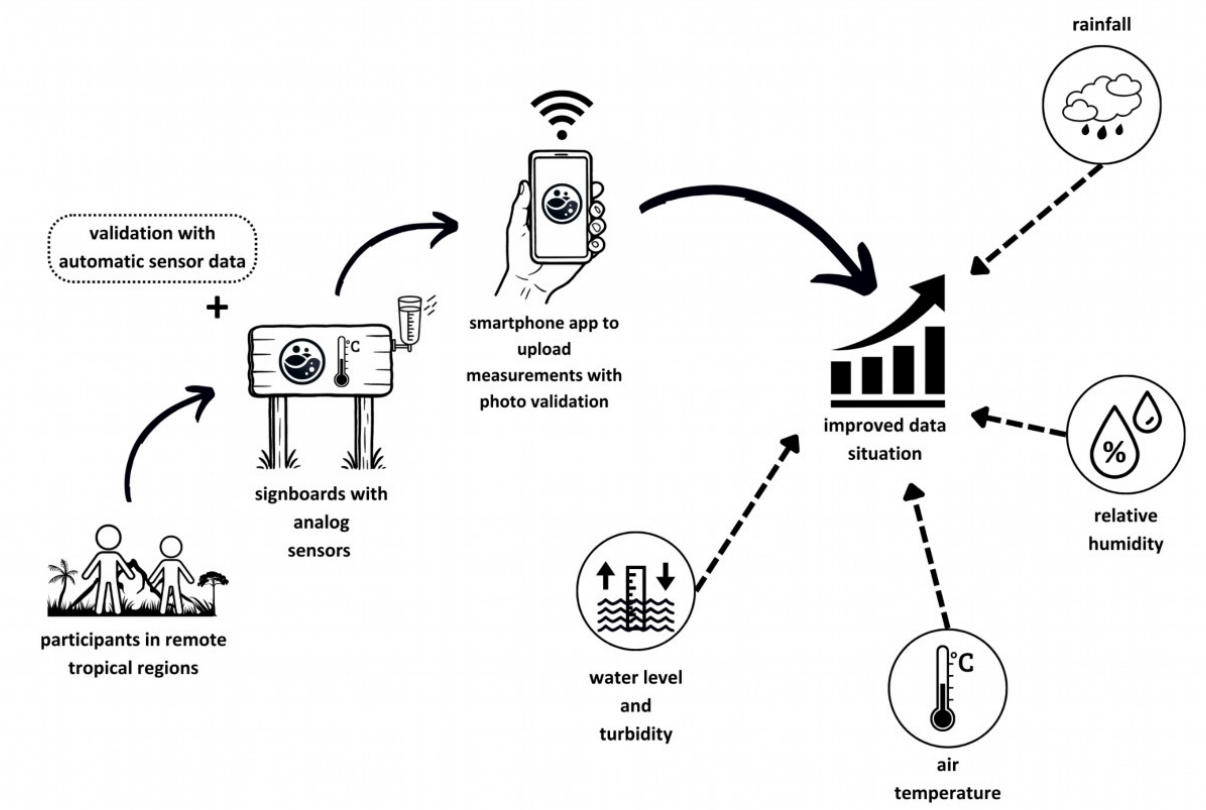


Figure 3.1: *Schematic overview of the participatory monitoring approach used in HydroCrowd.*

The main objective of this study was to validate the project's participatory approach for collecting hydrometeorological data. The results of this analysis can be used to inform future projects about the suitability of PM approaches for hydrometeorological monitoring in similar contexts, depending on the project purpose and associated data accuracy requirements. The following research questions were addressed in this study:

1) How well can frequent and non-frequent participants measure hydrometeorological parameters using simple analog sensors?
2) How good are the analog sensor measurements by participants compared to automatic sensors in terms of accuracy?

3.3 Material and methods

3.3.1 Study areas

For the project, study areas in three different tropical mountain settings in Ecuador (ECU), Honduras (HND) and Tanzania (TNZ) were selected, with the majority of the stations located in protected national parks. The areas differ in dominating land use and local climatological conditions and are described in the following sections.

3.3.1.1 Ecuador

Three sites in the province of Azuay, Southern Ecuador (ECU), were selected as study areas (Figure 3.2). The first site is located in the Andean Páramo ecosystem in the North-Eastern part of the Cajas National Park (3,144 - 4,429 m above sea level (a.s.l.)) (Arcusa et al., 2020; Pesántez et al., 2018).

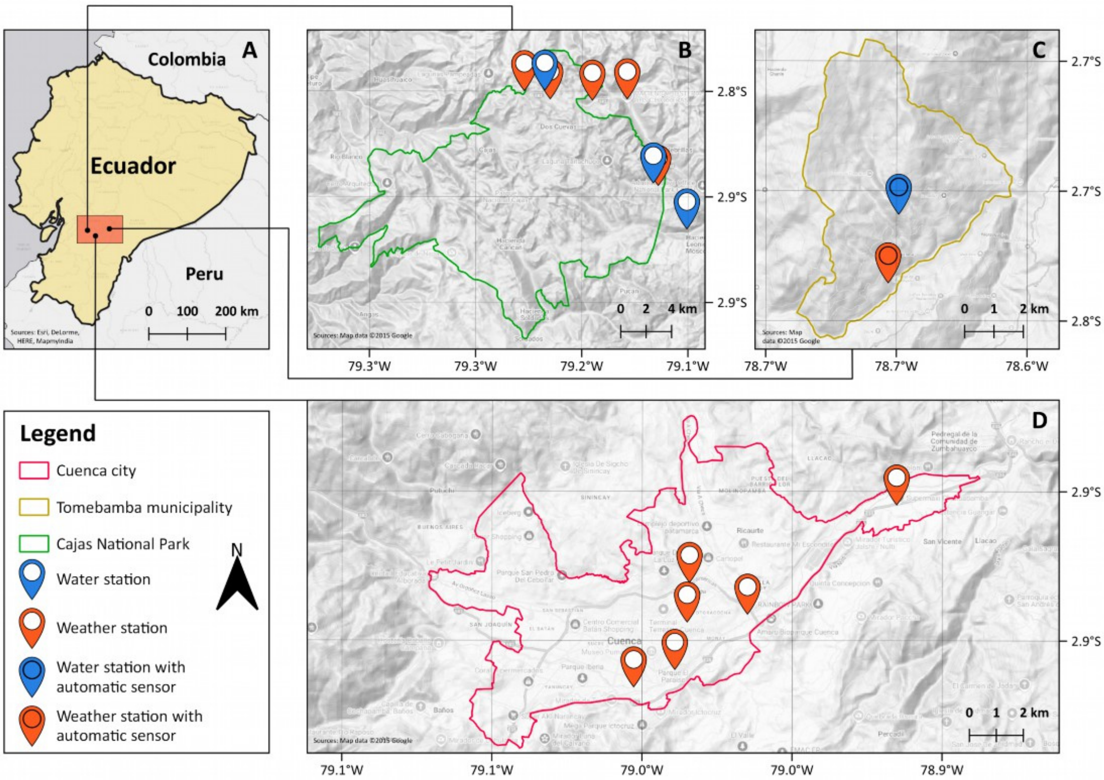


__Figure 3.2:__ Study area in Southern Ecuador (A) with the three sites within and around the Cajas National Park (B), within the municipality of Tomebamba (C) and within the city of Cuenca (D).

The climate in the area is influenced by various atmospheric mechanisms, including the Intertropical Convergence Zone and the El Niño-Southern Oscillation and the Pacific

Decadal Oscillation (Morán-Tejeda et al., 2016; Poveda et al., 2006). The climate of the region is typically classified as a tropical high-mountain climate, characterized by minimal seasonal variation attributable to its equatorial location with daytime temperature between 12 - 18°C and nighttime temperatures that can drop down to -8°C (Buytaert et al., 2006; Hansen et al., 2003). Various studies characterized an annual precipitation between 1,000 and 1,300 mm (Buytaert et al., 2006; Celleri et al., 2007; Padrón et al., 2020). The landscape is dominated by tussock grass, low shrubs and glacial lakes and a small share of pine forest and pasture on Holocene Andosols and Histosols (Bandowe et al., 2018; Buytaert et al., 2006; Carrillo-Rojas et al., 2016; Harden, 2006). The city of Cuenca, the second study site, is the third largest city of the country with a population of approx. 596,000 (Instituto Nacional de Estadística y Censos, 2022). It is located in a basin around 30 km east to the Cajas National Park, with an altitude spanning between 2,350 to 2,550 m a.s.l. (Gianoli & Bhatnagar, 2019). It has a temperate Andean climate averaging with 16.3°C and an annual precipitation of 876 mm (WMO, 2025b). The last site in this study area is the municipality of Tomebamba, located around 37 km north-east of the city of Cuenca. The site has a comparable climate to Cuenca with an elevation of approx. 2,380 m a.s.l. (World Bank, 2025c).

HydroCrowd stations at sites in the Cajas National Park and Tomebamba were installed in June 2023 while stations in Cuenca were installed in August 2024 and January 2025. There are currently twelve weather and four water stations in use. Five weather stations and three water stations have been installed in the Cajas National Park site, one weather station and one water station in Tomebamba, and six weather stations in Cuenca. Due to theft and vandalism, six additional water stations and one weather station in Cuenca city were damaged and unavailable for this study.

3.3.1.2 Honduras

The Cacique Lempira Señor de las Montañas Biosphere Reserve (from here referred as Cacique Lempira Biosphere Reserve) which encompasses the Celaque National Park, is located in Western Honduras (HND) (Figure 3.3). HydroCrowd stations were installed in May 2023, May 2024 and August 2024, with a total of nine weather stations and five water stations.

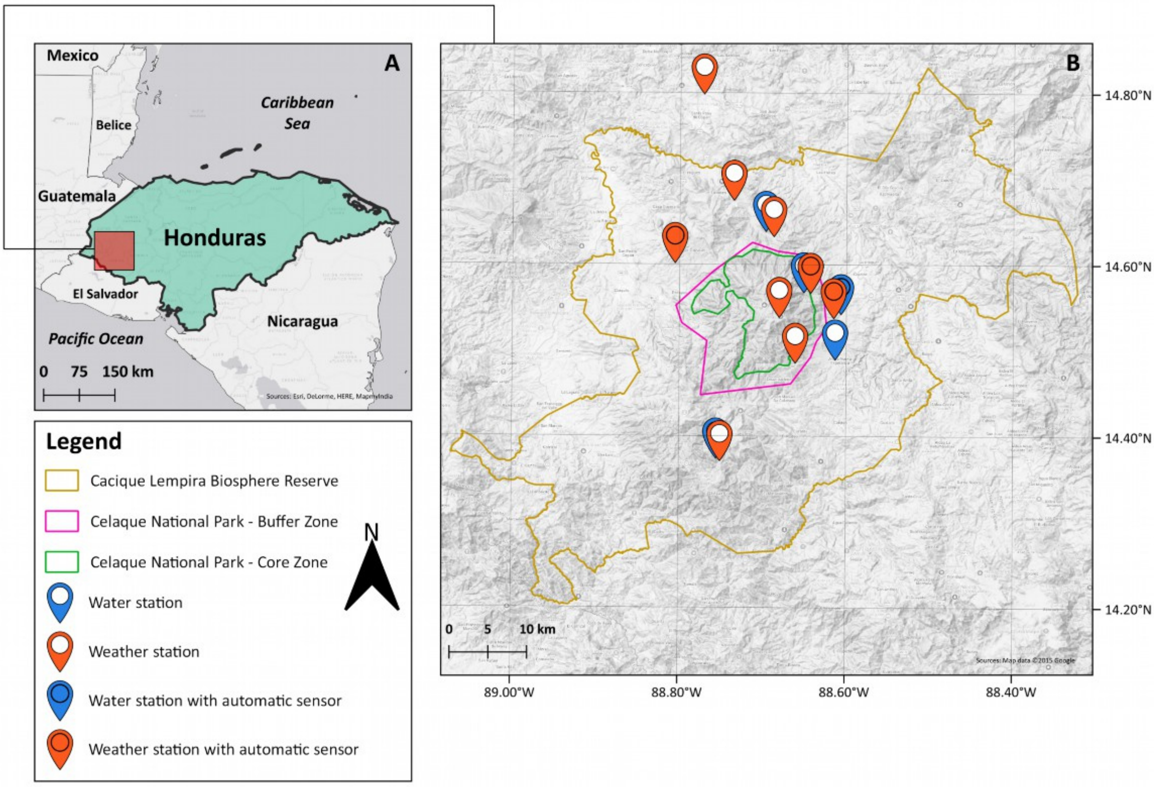


Figure 3.3: Study area in Honduras (A) and (B) in the Cacique Lempira Biosphere Reserve.

The region is characterized by mountainous topography (highest point at Cerro Las Minas with 2,870 m a.s.l) and rocky thin lithosols, where most parts of the area is considered unsuitable for intensive agriculture (D. L. Anderson & Devenish, 2009; FAO, 1966; Southworth et al., 2004). The area above 1,800 m (including the National Park) has been protected since 1987, with agricultural and industrial activities prohibited. (Pfeffer et al., 2001). Land use can be described as a combination of coniferous forests, coffee plantation, shrubland and a small share of annual crops outside the national park (Valdez et al., 2017). The region has a tropical savannah climate with an average precipitation of 1,600 mm/yr and average temperatures of 24°C in the lower areas, and with up to 2,400 mm/yr precipitation at higher elevations (Aguilar, 2005; Valdez et al., 2017). The largest proportion of annual precipitation falls between May and October, while the rest of the year is dry (Aguilar, 2005).

3.3.1.3 Tanzania

The study area in Tanzania (TNZ) is located on the southern slopes of the Kilimanjaro National Park (Figure 3.4). The area has a sharp elevation gradient ranging from around 770 m a.s.l. in the lowlands up to 5,895 m a.s.l., with Kibo (the highest volcanic cone).

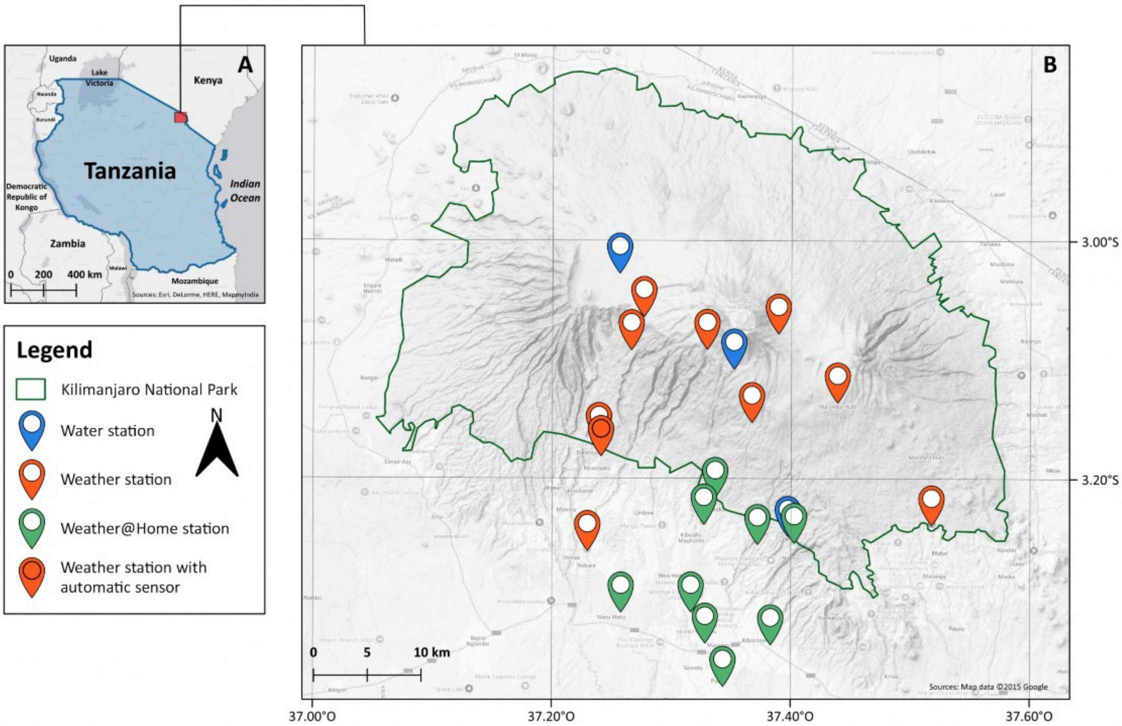


Figure 3.4: *Study area in Tanzania (A) at the southern slopes of Mt. Kilimanjaro (B).*

Climate, vegetation types and land use differ significantly with changing elevation (Hemp et al., 1998). Rainfall amounts to 900 mm at 800 m above sea level in the lowlands, peaking at around 2,700 mm at 2,200 m a.s.l. (Hemp, 2006; Røhr & Killingtveit, 2003) before dropping sharply to 750 mm at 3,750 m a.s.l. and continuing to decrease beyond this point (Appelhans et al., 2016). Precipitation follows a bi-modal annual pattern, with a first peak from March to May and a second one from November to December (Shagega et al., 2025). The area can be divided into multiple ecological zones which also differ greatly in land use and are described in more detail in Hemp et al. (1998). Andosols are the dominant soil type in this area (Zech, 2006).

HydroCrowd stations were installed in August 2023, with a total of ten weather and three water stations. Additionally, nine "Weather@home" stations were installed on private property of interested locals as well as hotels and guest houses, which are reduced variant of weather stations consisting only of a metal holder with the rain gauge and a smaller panel with a description of the approach. An overview of all stations installed for the project can be found in Appendix 1.

3.3.2 Data collection

3.3.2.1 Station types

All weather stations are made up of a roofed wooden signpost with a printed PVC banner and are equipped with three sensors: an analog hygrometer, thermometer and a plastic rain gauge. Figure 3.5 shows an example of a weather station.

Figure 3.5: *Weather station in Honduras, consisting of a wooden structure, a signpost with general information (left side) and instructions how to upload data (lower right side), and three analog sensors: a hygrometer for relative humidity, a thermometer for temperature and a rain gauge for rainfall (upper right side from left to right).*

In order to engage participants, the panel incorporates general information regarding the project, the importance of hydrometeorological data for measurement, and the reason behind the selection of the station's location in the region. The submission of observations was facilitated by the development of a smartphone application, "HydroCrowd", for Android and iOS operating systems. This application was developed by SPOTTERON, a company based in Vienna, Austria, and was released in May 2023. It is available in English, Spanish and Swahili. The panel also gives instructions on how to use the app. Alternatively, the provided QR-Code can be used to access a web application for data submission (www.spotteron.com/hydrocrowd). Participants are asked to submit a photo in addition to the actual observations. The sensors were installed in such a manner that all three sensors are clearly visible in a single photograph.

For the water stations, similar signposts were used with customized information content (Figure 3.6). The water level gauges were installed in sight in the river (Figure 3.6B). Additionally, turbidity can be measured at the water stations using a jug for taking a water sample and a turbidity tube for measurements (Figure 3.6A, Table 1). However, the latter data were not analyzed, as no validation photos nor automatic sensor data were available.

Figure 3.6: *Water station (A) with general information and instructions, a turbidity tube (white box), a jug for taking water samples connected with a chain to the station, and water level gauges (B) mounted on a rock in the river. A number on top of the gauge indicates the maximum height for each staff gauge reading.*

When possible, water level gauges were installed on stable, immovable structures like bridges or big rocks. If no such option was available, the gauges were mounted on metal structures with spikes which were hammered into the riverbed and stabilized with tension cables to avoid movement.

3.3.2.2 Analog sensors

Various criteria were taken into account for the selection of analog sensors for the weather and water stations. Firstly, the sensors needed to be user-friendly and easy to read. To this end, they needed to have a simple, clear scale, whereby measurement accuracy should be affected as little as possible. An easy-to-read scale was also a prerequisite for another criterion: the ability to verify the observations with a photograph. The final important criterion was commercial availability and low cost of the sensors. The low cost was also intended to minimize the probability of theft and enable a cost-effective replacement if a sensor is lost. These criteria were used to select the analog sensors listed in Table 3.1, which are also described in more detail below.

Table 3.1: *Station types with installed analog sensors and sensor specifications.*

Type	Parameter	Range	Sensor type	Model	Cost
Weather station	air temperature	-20 to 50°C[a]	bimetallic spiral spring	TFA Dostmann K1.100273	7 €
	relative humidity	0 – 100 %[b]	bimetallic spiral spring	TFA-Dostmann K1.100445	7 €
	rainfall	0 – 35 mm	plastic funnel	TFA-Dostmann 47.1013 & Relaxdays 10025387[d]	6 €
Water station	water level	0 – 100 cm[c]	rigid foam staff gauge	Nestle 18500000	42 €
			metal staff gauge	Locally made	150 €
	turbidity	5 – 1000 NTU	plastic turbidity tube	Trace2o Turbidity Tube	52 €

[a] Accuracy according to manufacturer's specifications ± 1°C, [b] Accuracy according to manufacturer's specifications ± 5%, [c] per unit installed – some sites required more than one staff gauge to cover the full range of water levels, [d] the two models are identical in design and volume.

For relative humidity and air temperature, analog hygrometers and thermometers were used (Figure 3.5). Both sensor types were tested by conducting parallel measurements (n = 100) using three hygrometers and three thermometers in Germany (appendix 2). These were compared to automatic sensor measurements (Lascar Electronics, EL-USB-2+, Whiteparish, United Kingdom) using the root mean square error (RSME - Equation 3.1).

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \text{(Equation 3.1),}$$

where n = number of observations, y_i = automatic sensor value and $\hat{y}_i$ = PM value. The test confirmed the manufacturer's confidence intervals (RMSE 0.57 - 0.86°C for air temperature and RMSE 2.63 - 4.66 % for relative humidity).

Due to the design of the weather station, simple funnel-shaped rain gauges (35 mm capacity) with an easy-to-read millimeter scale were selected for the PM rainfall collection (see Figure 3.5). With a rotating holder on the station, they can be emptied by simply turning them upside down after the measurement. It was expected that the majority of the rainfall would be measured, however, its volume might render it inadequate for inclusion in the analysis of heavy rainfall events.

For water level, 1 m rigid foam water level gauges were installed in Honduras and Tanzania. Due to import regulations, metal staff gauges sourced locally and printed with the same measurement scale were used in Ecuador. Depending on the expected maximum water level, several staff gauges were installed. On top of each gauge, a red number indicated the maximum height for each staff gauge (e.g. “1” for 1 m).

For comparability, all analog sensors except the metal staff gauges were purchased in Germany, while all station materials were sourced locally. Further information and a detailed cost breakdown are available in Campos Zeballos et al. (2025). Since hydrometeorological measurements may be affected by various interferences, the wooden station panels and attached sensors were installed according to WMO guidelines, where possible. However, this could not always be ensured, consequently the location of the stations was sometimes a compromise between accessibility, participants safety, station security and ease of installation.

3.3.2.3 Automatic sensors

Automatic sensors were installed at selected stations to collect reference data to validate the participatory observations. Table 3.2 provides an overview of the stations at which sensors were installed, as well as the periods covered.

Table 3.2: *Stations with automatic sensors and their measurement periods.*

Parameter and station name	Country	Period monitored
Air temperature and Relative humidity		
Nkweseko	TNZ	26.08.2023 - 15.02.2025
Tomebamba	ECU	07.12.2023 - 19.06.2024
Don Tito	HND	08.05.2024 - 01.05.2025
Finca El Nogal	HND	17.05.2024 - 02.12.2024
Parque Celaque	HND	03.12.2024 - 01.05.2025
Rainfall		
Don Tito	HND	25.05.2023 - 01.05.2025
Nkweseko	TNZ	27.08.2023 - 01.05.2025
Tomebamba	ECU	07.12.2023 - 18.12.2024
Water level		
Rio Arcilaca	HND	29.05.2023 - 01.05.2025[a]
Quebrada Santul	ECU	28.02.2024 - 18.06.2024

[a] No data was recorded from 19.12.2023 to 08.05.2024 due to incorrect reinstallation of the sensor.

Overall, air temperature and relative humidity were automatically monitored at five locations: one in Ecuador and Tanzania, and three in Honduras. Rainfall was monitored at one location in each country, and water levels were monitored at one station in Ecuador and in Honduras.

Air temperature and relative humidity measured by automatic sensors at stations in Ecuador and Honduras were collected in 15-minute (Tanzania: 1 h) intervals using stand-alone loggers (Lascar Electronics, EL-USB-2+, Whiteparish, United Kingdom; accuracy: ± 0.3 °C and ± 2.0 %), which were mounted to the weather stations hidden under the roof. For rainfall, automatic tipping buckets (Metek, Rain gauge 7043.0100, Elmshorn, Germany; accuracy: 4 % for 0 - 50 mm/h and 5 % for ≥ 50 mm/h) were installed near the weather stations measuring as well in 15-minute (Tanzania: 10 minutes) intervals. Water levels were measured at 15-minute intervals by pressure transducers and corrected for air pressure. For this, a pressure transducer (Diver, Driesen & Kern, P-Log3020-PA-INT, Bad Bramstedt, Germany; accuracy: ± 10 mm) was installed in a perforated pipe near the water level gauge. A corresponding barometer (Baro, Driesen & Kern, P-Log3020-baro, Bad Bramstedt, Germany; accuracy: ± 0.6 hPa) was mounted nearby (e.g., on the water stations panel). All automatic sensors were installed according to WMO guidelines, where possible, taking local conditions into account. Water level was subsequently calculated according to hydrostatic pressure (Equation 3.2), assuming a constant water density:

$$\boldsymbol{h} = \frac{(p_{abs} - p_{bar}) * 100}{\rho * g} \quad \text{(Equation 3.2)},$$

where h = water level in m, p_{abs} = absolute pressure in mbar, p_{bar} = barometric pressure in mbar, ρ = density of the water (997 kg m^{-3} at 25°C), and g = acceleration due to gravity (9.81 m s^{-2}).

As the water level gauges and pressure transducers could not be installed directly next to each other at the same height, there is a constant offset of a few centimeters in the automatic measurements. To correct this, a HydroCrowd team member carried out a reference measurement at both stations. The difference between the analog and automatic water level of these reference measurements was then used to correct the offset in all subsequent automatic measurements at both stations.

Due to installations at different times, relocation of sensors, repeated theft and vandalism in Honduras and Ecuador, it was not possible to maintain similar measurement periods for all study areas.

3.3.3 Data validation

Data validation was conducted in two independent parts in accordance with the two research questions of this study for the period from May 2023 to May 2025 for observations with corresponding 1) photos and 2) automatic sensor data. Observations by HydroCrowd team members were excluded.

3.3.3.1 Validation of different participants using photos

Participants were given a unique user ID when registering for the smartphone application. This ID was used to divide them into two groups of users: frequent and non-frequent users, similar to the analysis of Campos Zeballos et al. (2025). The former are people who often submit readings, such as community members, farmers or park rangers, while the latter are only on site occasionally, such as tourists who only measure once or a few times, for instance during a hike. The groups were defined as follows: non-frequent participants submitted a maximum of six observations within the first seven days after their first contribution. If they submit more or over a longer period of time, they are considered to be frequent participants. Participants who uploaded observations without registration were always indicated as “anonymous”, received the same ID and were considered as non-frequent participants.

To address the first research question, all analog observations were validated using the photos of the observations, if submitted alongside the observation. The submitted observations were compared with photos of the measuring devices submitted at the same time and the mean absolute error (MAE, equation 3.3), the RMSE (equation 3.1) and the coefficient of variation (CV, equation 3.4) were calculated. As a normal distribution, tested with the Shapiro-Wilk test (Shapiro and Wilk, 1965), could not be confirmed for all parameters, the Spearman Rank correlation (r_{spear}, Equation 3.5) was used to determine the degree of correlation between analog and automatic sensor measurements. Additionally, a Wilcoxon signed-rank test (Wilcoxon, 1945) was performed to statistically compare frequent and non-frequent users' data (alpha level = 0.01).

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad \text{(Equation 3.3),}$$

where n = number of observations, y_i = automatic sensor value, and $\hat{y}_i$= PM value.

$$\boldsymbol{C_V} = \frac{\sigma}{\mu} \qquad \text{(Equation 3.4),}$$

where σ = standard deviation of the data, μ = arithmetic mean of the data.

$$\boldsymbol{r_{spear}} = \frac{\mathrm{cov}[\,\mathrm{R}[X],\mathrm{R}[Y]]}{\sigma_{R[X]}\sigma_{R[Y]}} \qquad \text{(Equation 3.5),}$$

where cov = covariance of the rank variables, σ_X, σ_Y =standard deviations of the rank variables.

3.3.3.2 Validation with automatic sensor data

To address the second research question, the PM observations were validated against the automatically measured reference data (Table 3.2). To compare the data, the following steps were carried out:

- To ensure comparability between the study regions all analog observations were assigned to the nearest full hour. For instance, a measurement taken at 10:47 a.m. was assigned to 11:00 a.m. The air temperature, relative humidity and water level observations were then compared to the corresponding automatic measurements.
- Automatic rainfall measurements were summed up for the period between two analog observations and compared with the second analog observation. Evaporation of water from the analog rain gauge was not considered.

- All rainfall totals compared where the automatically measured rainfall was higher than 35 mm were excluded from further analysis.
- Despite the roof, analog hygrometers and thermometers were not always completely shielded from sunlight. This may have caused observations to be distorted due to overheating of the bimetal spiral spring inside the sensors. Assuming generally correct observations by participants and only a slight deviation of the analog sensors compared to the automatic sensors, outliers were eliminated by calculating the interquartile range (IQR) of the difference between the analog and automatic sensor measurements. An observation is considered an outlier if the value is 1.5 times greater or 1.5 times less than the IQR (Equation 3.6). If a PM air temperature observation was filtered out as an outlier, the corresponding relative humidity value was removed too as the analog hygrometers are also bimetal based and seemed to be affected by intense exposure to sunlight as well.

  $$(Q_1 - 1.5\ IQR) > \boldsymbol{x} > (Q_3 + 1.5\ IQR) \qquad \text{(Equation 3.6)},$$

  where x = individual air temperature measurement, Q_1 = first quartile of the air temperature difference, Q_3 = third quartile of the air temperature difference and IQR = $Q_3 - Q_1$.
- As the barometer from the station Rio Arcilaca was temporarily moved to a different location, the air pressure values in the periods from 04.06.2023 to 12.06.2023 and from 28.09.2023 to 20.11.2023 were corrected for the difference in altitude.

For all stations with automatic sensor reference data, the analog observations were compared by calculating the range, the arithmetic mean, MAE, RMSE and CV. Similar to the photo validation, data was not normally distributed based on the Shapiro-Wilk test. Therefore, again r_{spear} was calculated for all stations to determine the correlation with the automatically measured data. To evaluate the influence of measurement frequency on the accuracy of rainfall observations, the effect of the time between individual analog observations on the difference in measured rainfall was also analyzed. The Wilcoxon signed-rank test was used to statistically compare analog and automatic data with an alpha level of 0.01. Due to the findings of this analysis the analog relative humidity observations were first validated following the procedure introduced above, then corrected using linear regression and validated again afterwards.

3.4 Results

In the period from May 2023 to May 2025 a total of 2,982 PM observations were received. Table 3.3 provides an overview of the distribution of these by country and parameter.

Table 3.3: *Distribution by parameter and number of observations per country.*

Country	Rainfall	Air temperature	Relative humidity	Water level
ECU	733	750	751	89
HND	568	551	552	133
TNZ	675 (1,322[a])	669	669	8
Total	1,976 (2,623[a])	1,970	1,972	230

[a] Including measurements carried out at Weather@home stations.

The most individual measurements were collected in Tanzania (2,668 including Weather@home measurements) followed by Ecuador (2,323) and Honduras (1,804). Excluding observations from Weather@home stations, the number of rainfall, air temperature and relative humidity measurements were comparable for the three countries. The number of observations from water stations (water level measurements) differed substantially between the countries.

3.4.1 Frequent and non-frequent participants and photo validation

When dividing all participants into the two user groups, a total of 2,518 observations (84.4 %) were submitted by frequent participants, while non-frequent participants submitted a total of 464 observations (15.6 %). From the frequent participants' observations, 2,292 were from weather and Weather@home stations and 226 from water stations. Non-frequent participants submitted 439 weather and Weather@home station and 25 water station observations. Table 3.4 provides the validation results of the PM observations using the submitted photos, separated into frequent and non-frequent participants.

Table 3.4: *Validation results for the photo validation divided by "frequent" and "non-frequent" participants (n = number of specific measurements, MAE = mean absolute error, RMSE = root mean squared error, rspear = Spearman Rank correlation, CV = coefficient of variation for submitted and with photos corrected values and percentage = percentage share of the differing values from n).*

Parameter	n	Differing values	Percentage	MAE	RMSE	r_{spear}	$CV_{submitted}$	$CV_{corrected}$
Frequent								
rainfall	1,621	49	3.0 %	0.17 mm	1.78 mm	0.99	1.70	1.71
air temperature	1,608	44	2.7 %	0.14 °C	1.04 °C	0.99	0.28	0.28
relative humidity	1,609	100	6.2 %	0.48 %	3.25 %	0.99	0.35	0.35
water level	206	24	11.7 %	0.01 m	0.02 m	0.99	0.65	0.65
Non-frequent								
rainfall	333	15	4.5 %	0.49 mm	2.53 mm	0.93	1.65	1.74
air temperature	343	12	3.5 %	0.29 °C	2.03 °C	0.97	0.34	0.32
relative humidity	344	17	7.6 %	0.85 %	4.62 %	0.98	0.43	0.41
water level	17	4	23.5 %	0.03 m	0.12 m	0.90	0.74	0.65

For all parameters, except water level, less than 10 % of the observations were classified as incorrect and could be corrected using the submitted photos. Observations from frequent participants are characterized by a lower error rate across all parameters. The smallest error rate was observed for air temperature measured by frequent participants with 2.7 % while water level measured by non-frequent participants showed the largest error rate with 23.5 %. The correlation between submitted and corrected observations was very high with r_{spear} = 0.99 for all frequent participant's observations and r_{spear} ranging from 0.90 - 0.98 for non-frequent participants. The CVs revealed similar patterns of variation where the difference between $CV_{submitted}$ and $CV_{corrected}$ was higher for non-frequent participants. The Wilcoxon signed-rank test showed that there were no statistically significant differences in the distribution of the two user groups for any of the four parameters ($p\text{-value}_{air\ temperature}$ = 0.128, $p\text{-value}_{relative\ humidity}$ = 0.506, $p\text{-value}_{rainfall}$ = 0.143 and $p\text{-value}_{water\ level}$ = 0.738).

3.4.2 Validation using automatic sensor data

For the validation using automatic sensor data, the IQR analysis enabled the detection of potential outliers in the PM air temperature and relative humidity data. Between 8.2 and 20.0 % of the measurements were identified as outliers and were excluded from further analysis (Table 3.5). An additional 4.7 - 17.0% of all rainfall events were excluded where the automatically measured rainfall was higher than 35 mm.

Table 3.5: *Stations with original (= n) and filtered (excluding outliers) (= m) air temperature, relative humidity data and rainfall data.*

Parameter	Station name	n	m	Excluded	Excluded share (%)
air temperature and relative humidity	Tomebamba	73	67	6	8.2
	Don Tito	20	16	4	20.0
	Finca El Nogal	59	54	5	8.5
	Parque Celaque	19	16	3	15.8
	Nkweseko	314	275	39	12.4
rainfall	Tomebamba	106	101	5	4.7
	Don Tito	112	93	19	17.0
	Nkweseko	315	280	35	11.1

Table 3.6 gives an overview of the validation results for the four parameters at all stations equipped with automatic sensors.

Table 3.6: *Validation results for air temperature, relative humidity, rainfall and water level (n = number of measurements compared, pm = participatory monitoring, aut. = automatic sensor, std = standard deviation, CV = coefficient of variation, MAE = mean absolute error, RMSE = root mean squared error, r_{spear} = Spearman Rank correlation).*

Parameter	Station	Country	n	$Range_{pm}$	$Range_{aut.}$	$Mean_{pm}$	$Mean_{aut.}$	Std_{pm}	$Std_{aut.}$	CV_{pm}	$CV_{aut.}$	MAE	RMSE	r_{spear}
air temperature (°C)	Tomebamba	ECU	67	14 – 30	13.5 – 28.5	20.7	20.0	4.7	4.6	0.22	0.23	0.95	1.27	0.94
	Don Tito	HND	16	24 – 33	20.5 – 33.0	28.1	26.9	3.2	3.4	0.11	0.12	1.44	1.83	0.96
	Finca El Nogal	HND	54	20 – 34	19.0 – 31.5	26.7	25.1	3.5	3.1	0.13	0.12	1.65	1.93	0.92
	Parque Celaque	HND	16	18 – 28	17.0 – 25.5	22.2	21.2	3.3	3.0	0.14	0.14	1.28	1.46	0.92
	Nkweseko	TNZ	275	11 – 33	11.5 – 30.0	17.8	17.8	3.8	3.4	0.21	0.19	0.74	0.95	0.96
relative humidity (%)	Tomebamba	ECU	67	18 – 87	36.5 – 94.5	51.7	68.7	20.8	16.7	0.40	0.24	16.96	18.75	0.91
	Don Tito	HND	16	25 – 52	45.5 – 93.0	39.0	66.4	8.9	14.9	0.22	0.22	27.44	28.57	0.87
	Finca El Nogal	HND	54	19 – 79	46.5 – 98.0	45.8	72.9	13.6	13.0	0.29	0.18	27.04	27.72	0.76
	Parque Celaque	HND	16	21 – 66	59.0 – 96.5	45.4	77.1	14.5	13.0	0.31	0.16	31.69	32.22	0.85
	Nkweseko	TNZ	275	18.5 – 82	34.0 – 100	58.2	88.4	10.3	11.7	0.18	0.13	30.21	30.93	0.88
rainfall (mm)	Tomebamba	ECU	101	0 – 35	0 – 30.6	4.3	5.4	6.5	7.5	1.49	1.37	2.55	6.50	0.72
	Don Tito	HND	93	0 – 35	0 – 30.6	2.3	2.3	5.5	5.5	2.41	2.37	2.56	6.58	0.42
	Nkweseko	TNZ	280	0 – 35	0 – 34.8	6.4	6.1	8.6	8.1	1.34	1.33	3.10	7.03	0.80
water level (m)	Quebrada Santul	ECU	7	0.12 – 0.32	0.18 – 0.28	0.20	0.21	0.08	0.04	0.39	0.19	0.04	0.04	0.81
	Rio Arcilaca	HND	48	0 – 0.90	0 – 0.98	0.46	0.46	0.17	0.21	0.39	0.46	0.08	0.10	0.87

A high correlation was revealed for all stations and parameters (0.72 – 0.96) except for rainfall at station Don Tito (0.42). For air temperature, the results were overall the best, with the lowest deviation (MAE and RMSE) and the highest correlation (r_{spear}). Similar results but with lower r_{spear} and slightly higher deviation were found for water level. Analog rainfall measurements were characterized by high deviation (MAE = 2.55 – 3.10) compared to the mean values (2.3 – 6.1) and low to good correlation (0.42 – 0.80). The coefficients of variation of PM and automatically measured data were comparable for water level data at Rio Arcilaca and all air temperature and rainfall data, but not for water levels at Quebrada Santul and all relative humidity data.

A systematic underestimation was determined for PM relative humidity, resulting in substantial deviations with MAE ranging from 16 to 31%. Nevertheless, an acceptable to high r_{spear} was found (0.76 – 0.91). This significant deviation was only detected during validation, as the test measurements carried out before deployment of the sensors (see Section 2.2.2) showed no major deviation from the automatic sensor measurements. Given the relatively high r_{spear}, we corrected the values using linear regression. As the deviations differed per station, a separate model was developed for each station. Considering the low total number of measurements and for keeping the correction comparable, each model was trained using six randomly selected pairs of values from each station, after which the remaining values were estimated and the comparison metrics recalculated (Table 3.7). For all stations the deviation was substantially reduced (MAE: 5.45 - 9.50 %) while r_{spear} remained at a similarly high value.

Table 3.7: *Validation results for corrected relative humidity (n = number of measurements compared, pm = participatory monitoring, std = standard deviation, CV = coefficient of variation, MAE = mean absolute error, RMSE = root mean squared error, r_{spear} = Spearman Rank correlation).*

Station	**n**	**Coefficient**	**$Range_{pm}$**	**$Mean_{pm}$**	**Std_{pm}**	**CV_{pm}**	**MAE**	**RMSE**	**r_{spear}**
Tomebamba	61	0.95*	45.3 – 95.7	71.5	16.7	0.23	5.60	7.05	0.89
Don Tito	10	0.71*	49.2 – 80.5	63.7	11.9	0.18	5.45	6.97	0.91
Finca El Nogal	48	0.90*	42.6 – 89.0	63.1	10.5	0.16	9.50	11.24	0.72
Parque Celaque	10	0.94	54.3 – 89.4	72.6	13.1	0.17	6.14	6.83	0.92
Nkweseko	269	0.84*	50.1 – 100	84.9	8.9	0.11	5.92	7.53	0.88

* slope is significant.

For all stations and parameters, a statistical test was carried out to compare analog and automatic measurements. The Wilcoxon signed-rank test revealed significant differences in distribution for air temperature at station Finca El Nogal (p = 0.008) and corrected relative humidity at Nkweseko ($p < 0.001$) and Finca El Nogal ($p < 0.001$). For all other stations and parameters no significant differences were observed ($p > 0.01$). The distribution of the data from the various stations is shown in figure 3.7, comparing automatic and PM measurements.

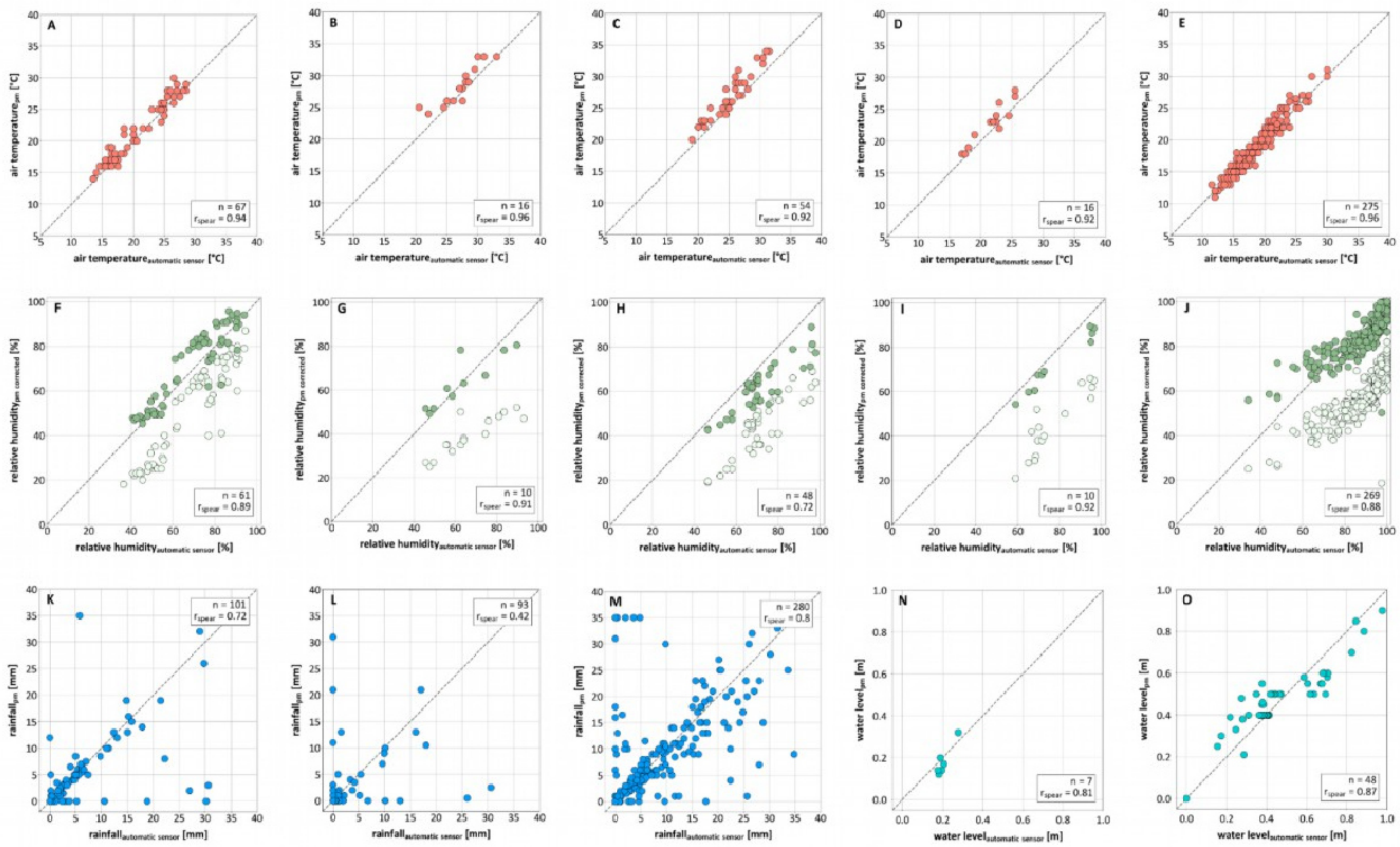


Figure 3.7: *Distribution of PM versus automatic sensor data (n = number of measurements compared where for F - J only the corrected numbers are presented, rspear = Spearman Rank correlation) for air temperature (Tomebamba = A, Don Tito =B, Finca El Nogal = C, Parque Celaque = D and Nkweseko = E), raw (light green) and corrected (dark green) relative humidity (Tomebamba = F, Don Tito = G, Finca El Nogal = H, Parque Celaque = I and "Nkweseko = J), rainfall (Tomebamba = K, Don Tito = L and Nkweseko = M) and water level (Quebrada Santul = N and Rio Arcilaca = O).*

For air temperature, a slight overestimation of most values can be noted at station Finca El Nogal (Figure 3.7C), which corresponds with the relatively high MAE and RMSE (Table 6). As shown in Figure 3.7 (F-J), the values for corrected relative humidity demonstrate a wider variation yet also exhibit a linear trend. It is evident that the values of Finca El Nogal (3.7H) exhibit a marginally higher underestimation in comparison to the other values. A strong accumulation of measurements of 0 mm can be noted for rainfall (Figure 3.7 K-M), and it is visible that the deviation in both directions, under- and overestimation, between PM and automatically measured rainfall increases with higher rainfall amounts. Looking at the plots for water level data, a monotonic trend and low deviation from the sensor values can be recognized for both stations. Considering the comparatively low number of measurements from station Quebrada Santul (Figure 3.7N), only a slight hint for a linear trend is visible. Participants' measurements at the Rio Arcilaca station also showed that lower water levels (< 0.6 m) were slightly overestimated and higher levels (above 0.6 m) underestimated (Figure 3.7O).

To understand whether the length of the interval between subsequent measurements influences the deviation from 'true' rainfall, the difference between each PM rainfall measurement and the corresponding automatically measured rainfall was plotted against the number of days since the last PM measurement (Figure 3.8). The total rainfall for both analog and automatic measurements was also calculated.

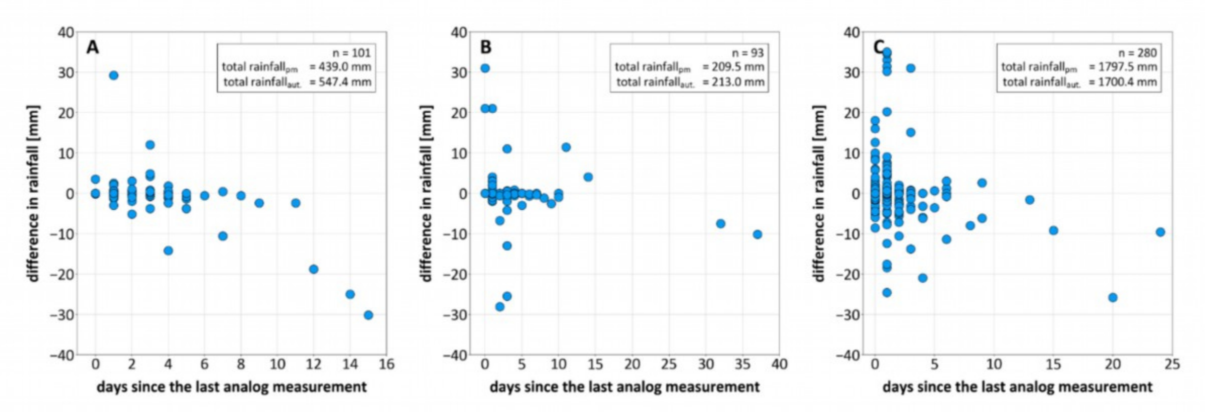


Figure 3.8: *Difference (PM – automatic sensor) in rainfall between PM and automatic sensor measurements (with n = number of PM measurements) depending on the time between each analog measurement at Tomebamba from 07.12.2023 to 17.12.2024 (A), at Don Tito from 11.07.2023 to 26.04.2025 (B) and at Nkweseko from 28.08.2023 to 03.02.2025 (C).*

Substantial under- and overestimation of rainfall can already be noted at all stations for periods ranging from less than a day up to five days, suggesting that increasing the measurement interval does not necessarily substantially improve the accuracy of the PM measurements. The total rainfall amounts show an underestimation for Tomebamba (19.80 %) and Don Tito (1.64 %), and an overestimation for Nkweseko (5.40 %).

3.5 Discussion

Different groups of participants were able to successfully measure hydrometeorological parameters with predominantly low errors using simple analog sensors. However, not all analog sensor measurements proved to be a good alternative for automatic sensors regarding measurement accuracy. This provides scope for critical discussion as to whether PM and the use of analog sensors are suitable for monitoring these parameters.

3.5.1 Performance of target groups

The integration of photo-based validation of analog measurements was found to be a suitable method for cross-checking analog measurements and for comparing the quality of submissions from frequent and non-frequent participants. The photo feature of the smartphone application can also be employed for long-term monitoring, thereby allowing the verification of stations independent of physical visits. In this manner, it is possible to verify the status of the sensors, including whether they are still present, whether there has been possible damage or other impairments, which are factors that compromise the readability of the sensors. The use of this method of photo validation was previously applied by other research projects for analyzing hydrometeorological observations (Davids, Devkota, et al., 2019; Eisma et al., 2023; Thatoe Nwe Win et al., 2019). Davids et al. (2019a), analyzed rainfall data collected via PM in Nepal with DIY rain gauges made of repurposed soda bottles in the context of the SmartPhones4Water monitoring network. Participants were asked to submit a photo of the bottles alongside their measurements, similar to HydroCrowd. From May to September 2018, 9% of all observations were classified as faulty and corrected using the values recorded by photos which is nearly three times more than for the rainfall observations analyzed in our study. Another example is a PM approach in Myanmar where photos were used to verify participants performance in water quality monitoring (Thatoe Nwe Win et al., 2019). Considering the greater uncertainty of the methods used, about half of the data was correctly assigned to the correct classes by participants revealing no systematic errors for the parameters compared.

The comparison between frequent and non-frequent participants showed clear differences in data collection frequency and data quality between the two groups. Frequent participants in our study not only collected significantly more data with a share of 84.4 %, but also data of better quality. However, a Wilcoxon signed-rank test did not identify statistically significant differences in the distribution of the compared data. These results correspond to the observations of Campos Zeballos et al. (2025) for the period from May 2023 to July 2024, where contribution relied mainly on frequent users and higher errors could predominantly be linked to non-frequent users. Other studies revealed clear differences in different user groups within PM projects. For instance, Ali et al. (2019) assessed the accuracy of participants measuring agrochemical contaminants in river water in the United States. Participants were grouped into different groups depending on how often they conducted such measurements: inexperienced, experienced and experts. Experienced and expert participants achieved better overall accuracy in correctly measuring the contaminants, while inexperienced participants required additional training. They concluded, based on their results and prior research (Sauermann & Franzoni, 2015; Scott & Frost, 2017) that higher participant's experience level might not only lead to higher accuracy but also affect motivation of participants regarding further participation. Another study by Meschini (2021) analyzed a PM program between 2007 and 2015, in which over 16,000 participants collected marine biodiversity data during dives in the Red Sea. Well known species could be identified better than rare species and participants' accuracy improved with experience gained.

However, the complexity of measuring a parameter can also be linked to the resulting accuracy. For instance, the errors of water level measurements in our study were higher, and participants described this parameter as more difficult to measure than air temperature, relative humidity or rainfall (Campos Zeballos et al., 2025). This might also be linked to the fact that the three weather station sensors could be read directly in front of the panel, while the stream gauges for water level are a little further away in the river and must be read from a certain distance. These results suggest that easier methods are more suitable for PM projects, although this might lead to a trade-off with data accuracy, which is also supported by previous PM research (Davids, Rutten, et al., 2019; Ramírez et al., 2023). Ramírez et al. (2023) reviewed different water quality monitoring methods in various PM projects. They concluded that, while simpler, low-cost methods were more user-friendly and enabled wider participation, they produced lower-quality data. In contrast, more complex methods that required more training produced better, more accurate data. Different streamflow PM methods were evaluated by Davids et al. (2019b). While the salt dilution method was recognized as the most accurate, it was perceived as the least favorable in terms of required training, cost and accuracy by the participants involved. Overall, a human factor (= participants measurement uncertainty) contributed more to the error for water level than for the other parameters.

3.5.2 Suitability and accuracy of analog sensors for hydrometeorological monitoring

As our results indicated, participants generally appeared to be able to read the analog sensors well. However, the comparison between analog and automatic sensor measurements revealed different accuracy issues for the different parameters. Here the question arises as to whether all analog sensors used in this study are an adequate choice for PM.

Several analog measurements of air temperature and relative humidity showed extraordinary deviations. These were not caused by wrong readings of the sensors and could therefore be considered possible outliers due to exposure to intense sunlight at certain times of day. The use of the IQR to identify outliers in air temperature and relative humidity datasets, as applied previously in other studies (Fredianto & Putri, 2023; Zafeirelli & Kavroudakis, 2024), proved to be a suitable approach to remove these erroneous values from the datasets. Based on the cleaned dataset, participants achieved the lowest measurement errors with analog air temperature measurements, and these also performed best overall in terms of deviation from and correlation with automatic sensors. Several other projects implemented a crowdsourced approach to measure air temperature (N. Barros et al., 2024; Beele et al., 2022; Leichtle et al., 2023; F. Meier et al., 2017; Peerlings et al., 2024; Rajagopalan et al., 2020; Sampson et al., 2025; Weyhenmeyer et al., 2017), but to the authors' knowledge, there are currently no PM studies that consider similar analog sensors like those used in HydroCrowd. Those that are most comparable used simple digital sensors (Beele et al., 2022; Loglisci et al., 2024; F. Meier et al., 2017; Rajagopalan et al., 2020; Sampson et al., 2025). However, the PM measurements across all these studies are of good to very good quality compared to regular weather stations, with better (± 0.5 °C) or comparable accuracies (Loglisci et al., 2024; Rajagopalan et al., 2020) than the analog sensors used for HydroCrowd (± 1°C).

The comparison with PM relative humidity and automatic sensor measurements revealed a high correlation but also substantial deviation and therefore did not provide accurate data. A correction through linear regression substantially reduced the high deviation of up to 30% down to 5.45 - 9.50 %. Nevertheless, a statistical test revealed significant differences between the distributions of the compared data at two out of five stations, which indicates that the selected hygrometers do not have sufficient accuracy to replace automatic sensors at all locations investigated. As the preceding test of the hygrometers confirmed the manufacturer's confidence intervals and the reading errors by participants are quite low, the high error rate is open to speculation. External circumstances, such as vibration during transportation from Europe to the study regions or the high altitude of the stations, might possibly have caused the deviation. Similar to air temperature, there are no PM studies comparing analog sensors with automatic sensors for relative humidity, but some where participants used professional handheld sensors (N. Barros et al., 2024; Loglisci et al., 2024; Rajagopalan et al., 2020). Compared to the analog hygrometers used in HydroCrowd (accuracy = ± 5%), the digital sensors used in these studies have comparable (Rajagopalan et al., 2020) or higher accuracies specified by the manufacturers with ± 2% (N. Barros et al., 2024), and did not show general under- or overestimation for different locations like observed in our study.

Analog rainfall measurements were characterized by a comparable error rate as air temperature, which indicates that it is easy to measure. Similarly, previous research concluded that PM holds a large potential for rainfall collection (Buytaert et al., 2014; Elmore et al., 2014; Njue et al., 2019). Nevertheless, while correlation with automatically measured rainfall in the same period varied from low to high, the deviation was quite high for all stations. Substantial deviations were observed for both low and high precipitation amounts. Furthermore, there was a trend observed for all stations analyzed that an increasing number of days between the measurements leads to higher deviations. While results for Tomebamba and Don Tito also showed that only a low number of days can already lead to high deviation, results for Nkweseko showed the opposite where many high deviations were found for already only a few days. This indicates the analog sensors cannot match the measuring accuracy of an automatic sensor. A possible explanation for this could be evaporation, against which the rain gauges offer no protection. Davids et al. (2019a) collected rainfall data via PM in Nepal with DIY rain gauges made of repurposed soda bottles with a total capacity of 200 mm. The rainfall measured with these bottles was compared to standard 203 mm rain gauges, resulting in a low error of -2.9%. The lowest error was revealed for the version of the bottle rain gauge where the upper part of the bottle was used as a funnel and protection against evaporation, suggesting that this could have significant influence on the accuracy of the data. Therefore, the irregularity of the measurements is problematic, since it cannot be ensured that correct rainfall amounts will be measured after several days. Since substantial deviations could be observed for both low and high rainfall amounts, it was not possible to determine an ideal measurement frequency with a low error rate. Nevertheless, from a hydrometeorological perspective and from practical feasibility within a PM approach, a daily measurement frequency should be pursued. Despite the observed deviations, the total amount of rainfall at the three stations only differed between 1 and 20% from those recorded by the automatic stations. This is similar to results obtained by Shinbrot et al. (2020). They conducted a rainfall collection study across two watersheds in Veracruz, Mexico. Trained participants used 250 mm rain gauges (WeatherYourWay/CoCoRaHS, https://weatheryourway.com/) to collect rainfall over a period of 1.5 years in two catchments with monitoring points between altitude range between 1,309 and 2,581 m a.s.l. (Shinbrot et al., 2020b) which is comparable to HydroCrowd. Despite a high measurement frequency, resulting in data for of 91 % of the days in the study period, the study revealed a general underestimation of rainfall by participants (12 %), especially in wet periods (16 %). Another core issue is the limited capacity of the rain gauge, due to which the analysis was limited to rainfall less than or equal to 35 mm. This is a substantial weakness of the analog sensor, which is further emphasized as up to 17% of the rainfall events for the individual stations were excluded due to exceeding 35 mm. This means that a substantial proportion of total rainfall cannot be measured, which leads to the follow-up question of how informative the analog measurements with relatively small rain gauges can be. Considering the high error rate the rainfall data measured with these rain gauges is not recommended for direct practical application. Subsequent enhancements, such as the installation of larger rain gauges and the ensuring of at least daily measurement frequencies, could prove advantageous in this context.

Although water level showed the highest error rate of all parameters, this did not dramatically affect the accuracy of the measurements. This may also be attributable to the fact that a significantly smaller number of water level measurements were available in comparison with the other parameters. The results for water level were not as good as for air temperature but still good with a low deviation from and high correlation with automatic sensor measurements. Other participatory projects focusing on water level monitoring showed similar validation results to the current study. In the CrowdWater project (Etter, Strobl, van Meerveld, et al., 2020; Seibert et al., 2022; Strobl et al., 2019; van Meerveld et al., 2017), a smartphone application was used to submit water level measurements using virtual water level gauges. Participants submitted pictures where water level was expressed in up to 10 water level classes rather than actual measurements (Etter, Strobl, van Meerveld, et al., 2020). Most of the measurements analyzed revealed good agreements between water level classes and water level measurements (Etter, Strobl, van Meerveld, et al., 2020). In another research by Weeser et al. (2018), water level was measured at 13 stations across a river basin in western Kenya, using comparable water level gauges similar to the ones used in HydroCrowd. At one station a radar-based sensor was installed to validate measurements by participants. Despite an absolute deviation of the water level values between PM and radar-based sensor measurements caused by an offset, an even higher correlation ($R = 0.98$) was found (Weeser et al., 2019).

Overall, not all analog sensors proved to be as accurate as automatic sensors. Air temperature data could be measured in very good quality, and the use of analog thermometers can be considered a suitable alternative to automatic sensors. Water level could also be measured with good accuracy using analog water level gauges, but frequent users can be considered more valuable for accurate readings. On the other hand, relative humidity showed moderate results, which might be improved by choosing other analog sensors than the ones used in this study. Ultimately, the rain gauges used did not prove to be suitable for accurate collection of actual rainfall, which can be mainly attributed to the irregular measurement frequency and most likely to the limited capacity of the rain gauges. In terms of sources of error for each parameter, the thermometers and hygrometers used are predominantly affected by environmental influences, such as intense sunlight. Rainfall is mainly affected by the capacity of the rain gauges, which most likely lack protection against evaporation. For water levels, a higher contribution to the error can be attributed to the reading errors by participants.

3.6 Conclusions

The objective of this study was to validate analog measurements by participants collected within the PM project HydroCrowd. Over a period of two years (May 2023 – May 2025), 2,982 analog hydrometeorological observations, consisting of air temperature, relative humidity, rainfall and water level were recorded. Considering these results of this study, we conclude the following:

- The slightly more accurate measurements by frequent participants for most parameters and their far better participation suggest that integration of local communities into PM programs is the most promising way to collect high-quality data instead of relying on untrained, non-frequent participants, such as tourists. The slightly higher complexity of specific parameters could be tackled through even more targeted recruitment of regular participants which are more experienced and therefore achieve higher accuracy measuring different parameters.
- Air temperature and water level data can be collected by participants in a high quality using simple analog sensors. For relative humidity further investigation under similar conditions is recommended to ensure accuracy of analog sensors. As both the thermo- and hygrometers seem quite sensitive to overheating, even better shielding against intense sunlight should be considered for PM projects using such sensors. As the 35 mm rain gauges cannot be considered reliable enough for continuous daily rainfall monitoring in areas where intense rainfall is expected, for instance, bigger rain gauges, ideally with sufficient protection against potential evaporation, should be used in future PM projects.

The next steps should consider the potential further use of the collected hydrometeorological data. As, for instance, air temperature proved to be a considerable alternative for automatic measurements in terms of accuracy, future PM projects should focus on expanding the density of the PM station network to further increase the availability of this data. A practical application could be a combination with other data sources, such as remote sensing datasets to further improve the spatial resolution of data. Since the air temperature data are of high quality and can be considered precise ground-based measurements, they could be combined for downscaling coarser remote sensing data. Considering the high correlation with automatic data, the relative humidity data could be used for that as well, despite the poorer accuracy. The rainfall data on the other hand could be combined with remote sensing rainfall products, such as CHIRPS or IMERG (Funk et al., 2015; Huffman et al., 2020) to test whether gaps where no measurements were taken or where the gauges' capacity was exceeded could be filled. As point observations have a limited spatial representativeness, further work could also analyze how different remote sensing datasets would be influenced by differing spatial distribution of the PM stations. Another conceivable application could be the integration of the PM data into hydrological modeling. As PM water level data has already been used successfully for different hydrological model applications (Mitze et al., 2025; Weeser et al., 2019), a next step could be the integration of other PM parameters. It could be investigated how much automatically measured data can be replaced with PM data, for example using the air temperature and relative humidity data.

As air temperature and water level monitoring worked out well, it should also be considered to scale the approach up to other remote regions of the world, while alternatives for relative humidity and rainfall monitoring require further investigation. It is most likely possible to source the station panel material locally in other regions as well, thus making it easy to establish a basis for the measurement network. The smartphone-based data collection worked well, and the ability to verify measurements with photos is useful. A critical and probably the most restricting aspect is the limited internet access in many remote regions, which was also a considerable barrier for participation in some of the HydroCrowd study regions (Campos Zeballos et al., 2025). This is why other data transmission methods should be investigated for remote areas with limited internet access. For site selection it is highly recommendable to include local authorities and institutions, to find the best locations for successful monitoring campaigns. Overall, local water resource management could benefit from increased data availability from PM projects and therefore, similar PM programs should be implemented in other countries of the Global South to address the general data scarcity in these regions.

3.7 Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.8 Author contributions

FM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing, Field work and project set-up; **SRJ**: Conceptualization, Methodology, Formal analysis, Funding acquisition, Project administration, Supervision, Writing – review & editing, Field work and project set-up; **LB**: Conceptualization, Funding acquisition, Writing – review & editing; **JCZ**: Conceptualization, Data curation, Formal analysis, Writing – review & editing, Field work and project set-up; **FC**: Writing – review & editing, Field work and project set-up; **FPS**: Writing – review & editing, Field work and project set-up; **BW**: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing, Field work and project set-up.

3.9 Funding

LB and SRJ acknowledge funding of the DFG project "Biodiversity and the supply of water-related NCPs" (BR 2238/35-1/-2 and JA 3059/4-2) as part of the DFG Research Unit 5064 "The role of nature for human well-being in the Kilimanjaro Social-Ecological System" (Kili-SES). FM, JCZ, SRJ and BW would like to thank the Kurt Eberhard Bode Foundation for funding the junior research group "HydroCrowd" (grant number 0122/40195/2022).

3.10 Generative AI statement

The authors declare that generative AI was used during the preparation of the manuscript in order to correct grammar and spelling of the text using the online version of DeepL Write (DeepL SE, Cologne, Germany). After using this tool, the authors reviewed and edited the content as needed. Additionally, some icons of figure 1 were created using ChatGPT 4.0 (OpenAI Inc., San Francisco, USA). The authors take full responsibility for the content of the publication.

3.11 Acknowledgements

We would like to thank Mancomunidad de Municipios del Parque Nacional Montaña de Celaque (Mapance) and Alliance of Bioversity International and CIAT in Honduras; Empresa Pública Municipal de Telecomunicaciones, Agua Potable, Saneamiento y Gestión Ambiental del cantón Cuenca (ETAPA) and Universidad de Cuenca in Ecuador; the Kilimanjaro National Park Authority (KINAPA) and the DFG Research Unit "The role of nature for human well-being in the Kilimanjaro Social-Ecological System (Kili-SES)" in Tanzania, for their support of the HydroCrowd project. Special thanks go to Jefferson Valencia Gómez, José Miguel del Cid and Jensen Mauricio Bautista-Perdomo, Jürgen Baumann, Nicolás Zúñiga, Pedro Alexander Sanchez, María Augusta Bermeo and all the others for their help in implementing the project. Finally, we would like to say thank you to all volunteers for their invaluable contribution, without their dedication and commitment, the success of this project would not have been attainable.

3.12 Data availability statement

The datasets and scripts used for this study can be obtained from Zenodo [https://doi.org/10.5281/zenodo.17712676].

4. Bias correction of ERA5-Land air temperature in remote tropical regions using participatory monitoring data

This chapter was submitted to Frontiers in Environmental Science as:

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Dowling, T. P. F., & Weeser, B. (under review). Bias correction of ERA5-Land air temperature in remote tropical regions using participatory monitoring data.

4.1 Abstract

Many tropical remote mountainous regions lack precise, spatially distributed air temperature data. This limits both scientific research and applied solutions in fields such as hydrology, meteorology, health and agriculture. As regular monitoring networks do not provide sufficient spatial coverage in these regions, alternatives are required. To address this issue, this study proposes a novel approach using participatory monitoring (PM) data for bias correction of ERA5-Land air temperature data. ERA5-Land data was trained on PM data from different remote mountainous regions in Ecuador (n = 67), Honduras (n = 23) and Tanzania (n = 275) using linear regression. Validation was conducted using automatically measured air temperature, applying different metrics, including mean absolute error (MAE) and coefficient of determination (R^2). Validation demonstrated an improvement across all stations, with the most significant reduction in MAE, from 5.49°C to 1.76°C. Concurrently, R^2 increased across all stations up to 0.83. The decrease in deviation was statistically significant for all stations. The study demonstrated that incorporating PM data into simple linear regression bias correction can enhance the accuracy of ERA5-Land air temperature data. This can be regarded as a pragmatic solution for remote regions, where the absence of weather stations is a common issue.

4.2 Introduction

The World Meteorological Organization (WMO) has confirmed that 2024 was the warmest year on record with the global air temperature 1.55°C above the pre-industrial level (WMO, 2025a). In less than ten years since the 1.5 degree limit that was set out in the Paris Agreement of the United Nations has been exceeded (COP 21: The Paris Agreement, 2025). This demonstrates that climate change is a matter of pressing concern. It has far-reaching consequences for human health, the global economy, and agriculture, necessitating significant adaptation measures (Anwar et al., 2013; IPCC, 2023b; Patz et al., 2014; Romanello et al., 2025). For countries belonging to the "Global South" that might lack resources for climate change adaptation and mitigation the effects are felt sooner and more intensely (Abraham, 2018; Cavazos et al., 2024; Ngcamu, 2023b; Sen Roy, 2018; United Nations, 2022). Given that all these challenges are directly or indirectly linked to rising air temperature (Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)], 2023), reliable data on this parameter can be considered critical for many kinds of climate change research.

Traditionally, hydrometeorological data such as air temperature has been collected using professional manual weather stations managed by meteorological authorities (WMO, 2017). Today, many of these are replaced by automatic weather stations which are considered the state of the art for this type of data collection (WMO, 2017). A common problem with these stations is their complexity of maintenance, costs and their highly uneven distribution around the globe. For instance, the distribution of the approximately 80,000 weather stations included in the Global Historical Climatology Network (GHCN) daily database (Menne et al., 2012), highlight this issue: Relatively few stations are located in countries like Honduras (n =7), Ecuador (n =15) or Tanzania (n = 14) compared to Western European countries like the Netherlands (n = 353) or Germany (n = 1069) (NOAA, 2025). This becomes more evident when considering the weather station density, i.e. the number of GHCN stations per unit area of the country. For instance, Honduras has a density of 0.062 stations * 1,000 km-^{2}, whereas the Netherlands has a density of 8.497 stations * 1,000 km-^{2}.

In recent decades, remote sensing-based monitoring has become a powerful alternative to ground-based data collection for estimation of large-scale air temperature (Hooker et al., 2018; Prihodko & Goward, 1997; Y.-J. Sun et al., 2005). Possible limitations of this type of data can be, however, the often coarse spatial or temporal resolution. A current state-of-the-art air temperature dataset is ERA5-Land, which has a global coverage with 9 km spatial- and a one hour temporal resolution (Muñoz-Sabater et al., 2021). Nevertheless, this spatial resolution may still be too coarse for specific applications, for instance the calculation of evapotranspiration in mountainous areas where air temperature varies greatly over short distances at different elevations.

There have been multiple attempts to adapt large-grid cell model outputs to make them more useful for the local scale (Dhawan et al., 2024; Niazkar et al., 2023; Sebbar et al., 2023; L. Xu et al., 2024; Zhang et al., 2024; Zhu et al., 2021). One commonly used approach is known as "bias correction", which removes a systematic error or deviation in the data to match a specific location. Some studies explicitly focused on bias correction for air or land surface temperature in mountainous regions in different parts of the globe (Dhawan et al., 2024; Sebbar et al., 2023; Zhu et al., 2021), whilst others tested different methods, such as statistical methods (Dhawan et al., 2024; Lompar et al., 2019), machine learning (Dhawan et al., 2024; Niazkar et al., 2023; Sebbar et al., 2023; Zhang et al., 2024) and deep learning (Niazkar et al., 2023; L. Xu et al., 2024). However, the dataset on which the correction is based is usually retrieved from automatic weather stations or other professional automatic sensors, which are not available everywhere. This raises the question of whether data from non-traditional sources in remote, mountainous regions can also be used for correction. Participatory monitoring (PM), where volunteers participate in the data collection process (Beele et al., 2022; Campos Zeballos et al., 2025; Danielsen et al., 2009; Mitze et al., 2026; Rajagopalan et al., 2020) could be such other source of data. Instruments provided to participants for this kind of approach are often cheap often have low prices and are easy for lay people to use (Buytaert et al., 2014; Davids, Devkota, et al., 2019). This makes it a viable alternative for regions with limited financial resources (Buytaert et al., 2014; Fankhauser & McDermott, 2014; Mitze et al., 2026). Depending on the intended use of the collected data, this kind of approach might improve the spatial resolution in comparison to traditional monitoring. However, temporal resolution is often the most significant challenge, as participants do not measure in a continuous and established mode as automatic sensors (Njue et al., 2019). A recent example by Mitze et al. (2026) showed, the quality of air temperature data measured by participants is of considerable quality compared to regular automatic sensors, despite outliers, which could be excluded. Nevertheless, the temporal resolution of the measurements might remain the limiting factor for further advanced application or long-term use.

In order to overcome the temporal limitations of PM data and the spatial limitations of remote sensing data, a combination of both types of data may be utilized, thus compensating for the inherent limitations of each data type. Only a few studies are available that cover such corrections on an hourly timescale using automatically measured data (Dhawan et al., 2024; Niazkar et al., 2023). Moreover, to date no research has been conducted on the use of PM data for the purpose of bias correction of ERA5-Land air temperature data. For this reason, this study will utilize PM air temperature data retrieved from different tropical remote regions in Ecuador, Honduras and Tanzania within the HydroCrowd project (Campos Zeballos et al., 2025; Mitze et al., 2026). The objective of this study is to analyze whether using PM data for bias correction can improve the accuracy of ERA5-Land air temperature in the regions under analysis.

As the spatiotemporal availability of local-scale air temperature data is not ensured in remote tropical mountainous regions, the results of this study can be beneficial for future applications of PM approaches and the local availability of continuous air temperature data, if no other options are feasible. Furthermore, the improved availability of such data can be vital for local adaptation to climate change, agriculture or health care.

4.3 Material and methods

4.3.1 Study regions and climate

In this study, air temperature data collected between May 2023 and August 2025 in the hydrometeorological PM project HydroCrowd was used. The study regions are located in different tropical mountain settings predominantly in and around national parks in Ecuador, Honduras and Tanzania (Campos Zeballos et al., 2025; Mitze et al., 2026). One station from each country was selected where air temperature was additionally monitored using an automatic sensor for validation purposes (Figure 4.1). Further information on the participatory approach and the data collection can be found in Campos Zeballos et al. (2025) and Mitze et al. (2026).

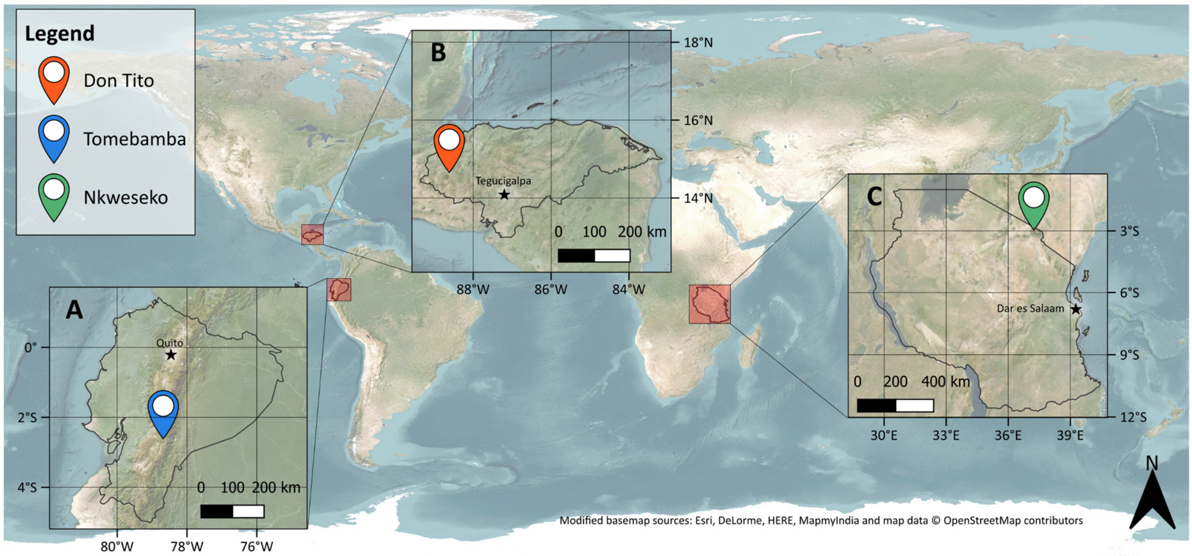


Figure 4.1: *Locations of the stations used for the analysis in the study regions in (A) Southern Ecuador, (B) Western Honduras and (C) Northern Tanzania.*

4.3.1.1 Ecuador

From the study region in Ecuador the station "Tomebamba" was selected for this analysis. The station is located in the municipality of the same name, province Azuay, Southern Ecuador (Figure 4.1A). It was installed in the center of the town, around 37 km north-east of the city of Cuenca, at an elevation of approx. 2,380 m a.s.l. Due to the absence of a climatological station in the immediate vicinity, data from the nearest WMO station in Cuenca is used to describe the prevailing climatic conditions. This station is at a comparable elevation (2,530 m a.s.l.) and according to the World Bank, it has the same temperate Andean climate (temperate oceanic climate) (World Bank, 2025a). The annual average temperature is 16.3 °C, with low seasonal variation (WMO, 2025c). November had the highest average temperature (22.2 °C), while July had the lowest (18.6 °C) (WMO, 2025c). Annual rainfall was 876 mm, with peak rainfalls in April, with over 120 mm (WMO, 2025c).

4.3.1.2 Honduras

The station "Don Tito" is located on a pasture in the vicinity of Celaque National Park in Western Honduras (Figure 4.1B). It was installed at an elevation of approx. 1,182 m a.s.l. about 5 km southwest of the city of Gracias. The region has a tropical savannah climate with an average temperature of 24 °C and annual average precipitation of 1,600 mm (Valdez et al., 2017). The area has a distinct rainy season from May to October, while the rest of the year is dry (Aguilar, 2005).

4.3.1.3 Tanzania

From the study region in Tanzania, the station "Nkweseko" was selected. It is located close to the border of Kilimanjaro National Park at an elevation of approx. 1,700 m a.s.l. (Figure 4.1C). The area has a typical equatorial daytime climate with varying temperature and rainfall depending on elevation (Hemp, 2005). Rainfall has a bimodal pattern with two rainy seasons from October to December and March to May (Shagega et al., 2025) resulting in approx. 2,616 mm per year (Gütlein et al., 2018). According to Gütlein et al. (Gütlein et al., 2018) and estimations of Hemp (Hemp, 2005), mean annual air temperature is around 18.4 - 18.8 °C.

4.3.2 Data sources

4.3.2.1 Participatory monitoring data and automatic sensor data

For participatory monitoring, analogue thermometers (TFA-Dostmann, K1.100273, Wertheim, Germany; accuracy: ± 1 °C) were used. The participants read the value from the sensor and submitted the measurements via a smartphone application or a web interface to the HydroCrowd database (Mitze et al., 2026). As outliers due to exposure to intense sunlight during certain hours of the day were detected in the data, the same interquartile range outlier elimination approach as presented by (Mitze et al., 2026) was used to remove outliers from the air temperature data. This resulted in a moderate reduction of the PM dataset (Tomebamba: -8.2%, Don Tito: -17.9% and Nkweseko: -12.4%).

At the same stations, stand-alone loggers (Lascar Electronics, EL-USB-2+, Whiteparish, United Kingdom; accuracy: ± 0.3 °C) for continuous measurements of air temperature and relative humidity were installed. It was not possible to provide continuous data for the full measurement period for all stations due to repeated theft and vandalism (Mitze et al., 2026). Table 4.1 gives an overview of the period and number of PM and automatic measurements monitored. Please note that PM temperature data will be referred to as T_{pm} and the data from the automatic sensors as T_{as}, in the following.

***Table 4.1**: Stations with participatory monitoring (pm) and automatically (as) measured data and the corresponding measurement periods for the stations (n = number of T_{pm} and m = number of T_{as} and coordinates in WGS1984).*

Station name	Country	Latitude	Longitude	n	m	Period monitored
Nkweseko	Tanzania	-3.184554	37.240911	275	12,912	26.08.2023 - 14.02.2025
Tomebamba	Ecuador	-2.742007	-78.682567	67	4,680	07.12.2023 - 19.06.2024
Don Tito	Honduras	14.532804	-88.61398	23	10,080	08.05.2024 - 02.07.2025

4.3.2.2 ERA5-Land

The land component of the fifth generation of European ReAnalysis (ERA5-Land) was used as a continuous data set for which deviations from automatic sensor measurements were to be corrected using T_{pm} (Muñoz-Sabater et al., 2021). It was developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and offers a global dataset from 1950 to the present with 50 variables describing the water and energy cycles over land (Muñoz-Sabater et al., 2021). It is based on a numerical integration of the ECMWF land surface model using the ERA5 climate reanalysis dataset (Muñoz-Sabater et al., 2021). This dataset relies on various data sources like meteorological station, weather ballons, satellites and other remote sensing data sources data (Copernicus, 2025; Muñoz-Sabater et al., 2021). As the same spatial and temporal resolution of data is not available for all regions in the world, interpolation was used to complete the reanalysis dataset (Muñoz-Sabater et al., 2021).

The global ERA5-Land dataset is available at 9 km spatial and 1-hour temporal resolution (Copernicus Climate Change Service, 2019). This product was preferred over other remote sensing datasets, due to its gapless continuous temperature values dataset, at global range. The 2 m air temperature of this dataset (referred to as T_{rs} in the following) was retrieved and processed via the Python earthengine-api of Google Earth Engine (Gorelick et al., 2017).

4.3.3 Bias correction

4.3.3.1 Data preprocessing

To match the timestamps of the automatically and ERA5-Land retried data, all PM data was assigned to the nearest full hour. As ERA5-Land data is available in coordinated universal time (UTC), time stamps were corrected to the local time zones (Ecuador: + 5, Honduras: + 6 and Tanzania: -3). For the model training, T_{pm} timestamps were used to retrieve the corresponding T_{rs} values for the three stations (training datasets). Similarly, the T_{rs} values were also retrieved for all hourly T_{as} in order to test the model on a larger time series later on (validation datasets). From each 9 km T_{rs} grid pixel a point value was extracted using the coordinates of the PM station (Table 1). Only the pixel value containing the station was used because ERA5-Land grid pixel values already represent a spatial average of 9 km. Incorporating surrounding pixels, for instance, via buffering or neighborhood averaging would introduce unwarranted spatial assumptions and could obscure meaningful gradients in the data, which we are trying to avoid.

4.3.3.2 Model training

Empirical quantile mapping is currently the most widely used method for bias correction in climate research, e.g. for air temperature data (Copernicus Climate Change Service, 2025). However, the small dataset size, especially for Tomebamba (n = 67) and Don Tito (n = 23), was considered insufficient to determine representative quantiles. Therefore, simple linear regression was applied to carry out bias correction. For this, the corresponding linear regression function from the machine learning Python package Scikit-learn (Pedregosa et al., 2011) was used to train a model for each of the three locations. The T_{rs} of the training datasets of each location were used as predictor variables for the regression models, while the corresponding T_{pm} were used as the target variables. As model testing is conducted using solely on the validation datasets, no train-test-split was performed for model training, i.e. all T_{pm} measurements of each station were used for model training. The trained models were then used to bias correct the hourly air temperature values (the bias corrected values are referred to as T_{bc} in the following) using the validation datasets.

4.3.4 Validation and statistical analysis

4.3.4.1 Descriptive analysis

A descriptive analysis was conducted for the training and validation datasets, to provide a more extensive foundation for the discussion of bias correction. For that, different summary statistics (range, median, mean and standard deviation) were calculated for T_{pm}

and T_{rs} of the training datasets. Boxplots were created for validation datasets to facilitate comparison of the different distributions.

To further analyze the correlation and differences between T_{rs} and T_{pm} of the training data, the normal distribution was tested using a Shapiro-Wilk test (Shapiro & Wilk, 1965). As normal distribution could not be confirmed ($p < 0.001$), Spearman Rank correlation (rspear, Equation 4.1) was used to determine the correlation.

$$\boldsymbol{r_{spear}} = \frac{\mathrm{cov}[\,\mathrm{R}[X],\mathrm{R}[Y]]}{\sigma_{R[X]}\sigma_{R[Y]}} \qquad \text{(Equation 4.1)},$$

where cov = covariance of the rank variables, σ_X, σ_Y =standard deviations of the rank variables.

Potential statistical differences in the training datasets were tested using a two-sided Wilcoxon signed-rank test (Wilcoxon, 1945). Assuming that participants were more likely to have taken measurements during the daytime, the temporal patterns of T_{pm} (i.e. the distribution of the actual T_{pm} values) were analyzed by examining the times of the day at which PM data were collected.

4.3.4.2 Model validation

To validate the bias correction approach, T_{rs} and T_{bc} of the validation datasets, were compared to the corresponding T_{as} for each station, whereby the comparison to T_{rs} is considered as a baseline. The mean absolute error (MAE - equation 4.2), the root mean squared error (RMSE - equation 4.3) and the coefficient of determination (R^2, equation 4.4) were used to determine the bias correction performance.

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|\,, \qquad \text{(Equation 4.2)},$$

where n = number of observations, y_i is T_{as} value and $\hat{y}_i$ is the T_{rs} or T_{bc} value.

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\,, \qquad \text{(Equation 4.3)},$$

where n = number of observations, y_i is T_{as} value and x_i is the T_{rs} or T_{bc} value.

$$\boldsymbol{R^2} = 1 - \frac{SS_{res}}{SS_{tot}}, \qquad \text{(Equation 4.4)},$$

where SS_{res} = residual sum of squares and SS_{tot} = total sum of squares.

4.3.4.3 Statistical and data structure analysis

The bias correction was further analyzed by providing statistical supporting evidence. In a first step, normal distribution of the validation data (T_{rs}, T_{as} and T_{bc}) was checked. Here, the Anderson-Darling-test (T. W. Anderson & Darling, 1952) was used as it is considered more robust for larger datasets. As the normal distribution could not be confirmed for all stations ($p < 0.001$), a one-sided Wilcoxon signed-rank test (Wilcoxon, 1945) was performed to analyze whether the distribution of the residuals of T_{bc} is significantly smaller (i.e. better) than the residuals of T_{rs} (alpha level = 0.01).

Further analysis was then conducted to understand the strengths and limitations of the bias correction using PM data. For this, samples of the validation datasets are plotted as time series to analyze the temporal pattern of the bias corrections in comparison to the ERA5-Land and automatic sensor data. By comparing this to the temporal distribution analysis of the training datasets (section 4.3.4.1), any influence of temporal patterns in training data collection were investigated. In order to proceed with this analysis, the deviation from T_{as} was grouped for each hour of the day of each test dataset. Thereafter, the Anderson-Darling-test was applied to all groups to verify normal distribution. As it could not be confirmed for all groups, the Friedman test (Friedman, 1937) was then conducted to compare the distribution of all samples against each other.

If this test was positive, indicating statistically significant differences between the individual hours of the day, a two-sided Wilcoxon signed-rank test (Wilcoxon, 1945) was conducted as a post-hoc test in order to ascertain whether there are statistically significant differences from zero (i.e. no significant deviation from T_{as}) for each hour, with an alpha level of 0.01.

4.4 Results

4.4.1 Descriptive analysis

The training data (T_{pm} and T_{rs}) from the three stations is characterized by substantial differences which are highlighted by the summary statistics in Table 4.2.

Table 4.2: *Summary statistics of the training data (where pm refers to T_{pm} and rs to T_{rs} data) for the three stations used in this study (n = number of Tpm measurements used, Std = Standard deviation).*

Station name	n	$Range_{pm}$	$Range_{rs}$	$Median_{pm}$	$Median_{rs}$	$Mean_{pm}$	$Mean_{rs}$	Std_{pm}	Std_{rs}
Nkweseko	275	11.0 – 31.0	12.3 - 28.2	17.0	20.2	17.8	20.1	3.8	3.1
Tomebamba	67	14.0 – 30.0	9.4 - 17.1	18.0	12.7	20.7	13.1	4.7	1.8
Don Tito	23	24.0 – 33.0	18.6 - 27.3	28.0	21.9	28.3	22.4	2.8	2.5

An unbalanced number of measurements was available, leading to Nkweseko having more than 10 times more measurements available for training than Don Tito. In addition, substantial differences in the mean between T_{pm} and T_{rs} can be observed. A statistical analysis revealed significant differences in the distribution for all stations ($p < 0.001$). While for Nkweseko the range of the measured T_{pm} was slightly wider than for the corresponding T_{rs} data, range and mean values for Tomebamba and Don Tito differed more clearly from one another. For instance, the maximum value of T_{pm} and T_{rs} for Tomebamba differed by almost 13 °C (Nkweseko: 3°C and Don Tito: 6°C). Nevertheless, all stations showed a comparable high degree of correlation: Tomebamba (r_{spear} = 0.83), Don Tito (r_{spear} = 0.85) and Nkweseko (r_{spear} = 0.86).

The temporal pattern (i.e. number of measurements collected for each hour of the day by participants) of T_{pm} shows large differences between Nkweseko and Tomebamba on the one hand and Don Tito on the other hand (figure 4.2).

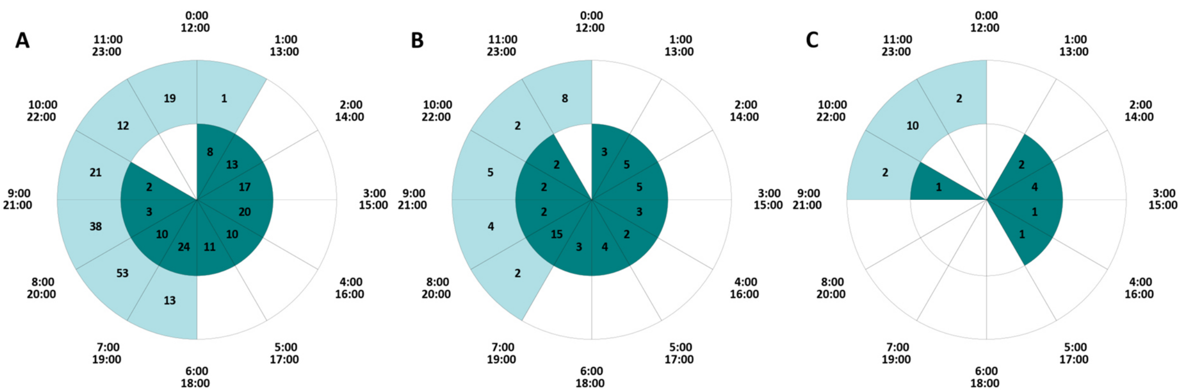


Figure 4.2: *Availability of participatory monitoring data for the three stations (A) Nkweseko, (B) Tomebamba and (C) Don Tito in the morning (light blue shading) and afternoon (petrol shading), with black numbers indicating how many measurements are available for each hour.*

Although Nkweseko and Tomebamba have high temporal coverage, no measurements were conducted in the late evening and at night. In sharp contrast to that, at Don Tito measurements were only taken during the late morning and predominantly during the afternoon.

Figure 4.3 displays the distribution of the air temperature validation data (T_{as} and T_{rs}) and the bias correction (T_{bc}) using boxplots.

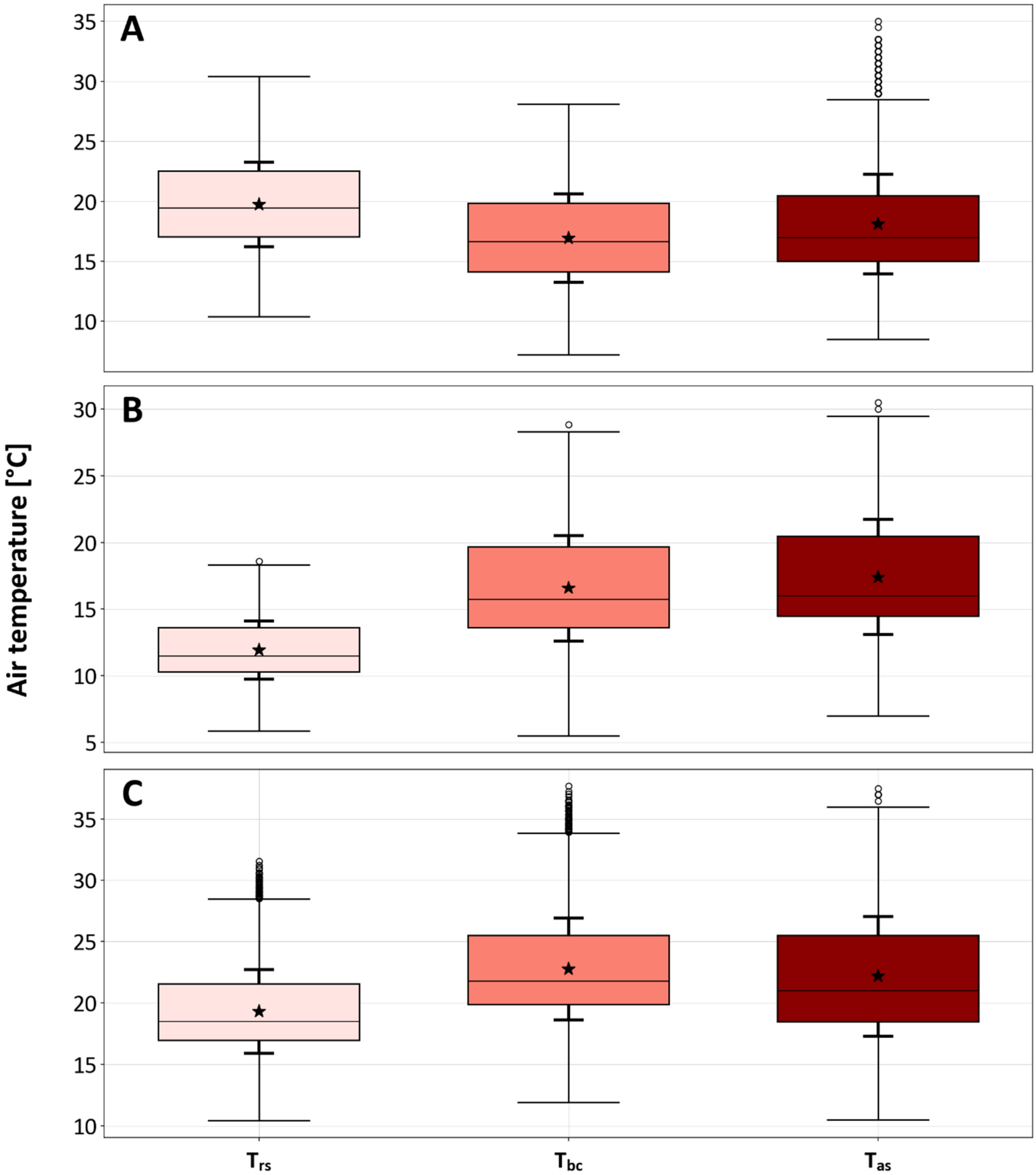


Figure 4.3: *Boxplots for the validation data at the three stations (A) Nkweseko, (B) Tomebamba and (C) Don Tito. The black star indicates the mean value and the thick smaller whiskers standard deviation.*

A comparison of the ranges of T_{bc} and T_{as} reveals that the bias correction adjusts the entire range predominantly good to the automatically measured data (T_{as}) for Nkweseko and Tomebamba. While the range for Nkweseko was only shifted slightly, it was extended for Don Tito at the lower and upper boundary. However, the small range for T_{rs} was drastically improved for Tomebamba. A visual comparison of medians and means indicates that for Nkweseko the bias correction adjusted air temperature downwards while for Tomebamba and Don Tito it was adjusted upwards.

4.4.2 Model validation

For a first overview of the validation of the bias correction the distribution of T_{rs} and T_{bc} is plotted against T_{as} (Figure 4.4).

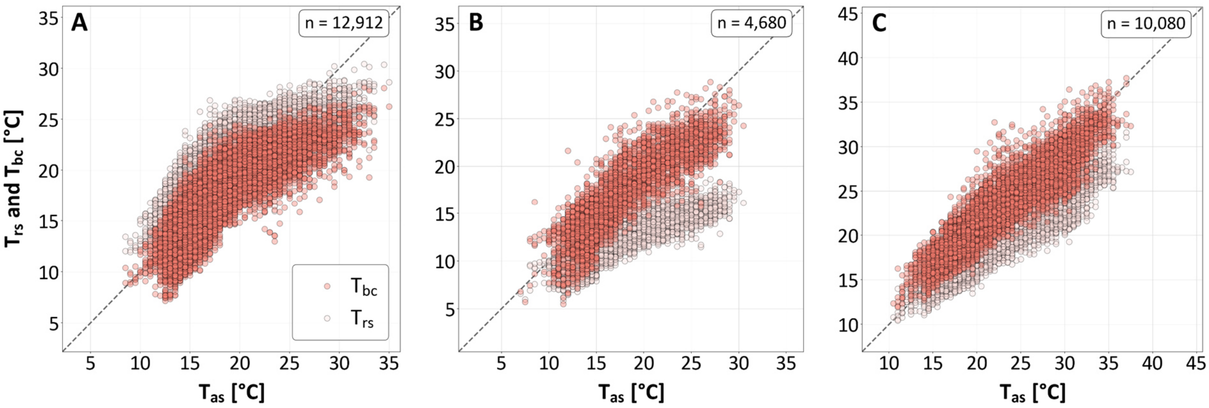


Figure 4.4: *Scatterplots comparing the validation data distribution of Trs (pale red) and Tbc (red) at the three stations (A) Nkweseko, (B) Tomebamba and (C) Don Tito.*

Clearly, the bias-corrected data for Tomebamba and Don Tito align more closely with the automatically measured data than the original ERA5-Land (Trs). For Nkweseko this is not so clear as the difference is substantially smaller than for the other two stations. Nevertheless, the data from Nkweseko and Tomebamba are also not evenly distributed along the 1:1 line, which would indicate a perfect fit. Especially higher temperatures are more likely underestimated by both Trs and Tbc. The data from Don Tito on the other hand shows a very good fit to the 1:1 line with slightly more scattering between 25 and 30°C. These impressions are confirmed by the validation metric results (Table 4.3).

Table 4.3: *Validation metric results for the three stations, with n = number of compared values, $R^2_{rs.}$ = baseline results using T_{rs} and R^2_{bc} = bias correction results using T_{bc}.*

Station	n	R^2_{rs}	MAE_{rs} (°C)	$RMSE_{rs}$ (°C)	R^2_{bc}	MAE_{bc} (°C)	$RMSE_{bc}$ (°C)
Nkweseko	12,934	0.55	2.27	2.77	0.62	1.98	2.54
Tomebamba	4,682	-0.98	5.49	6.09	0.73	1.76	2.27
Don Tito	10,080	0.46	2.95	3.57	0.83	1.54	1.97

For all three stations MAE and RMSE were reduced, where the reduction for Nkweseko was the smallest with 0.29 °C, while the value at Don Tito nearly halved to 1.41 °C. The most substantial improvement was achieved for Tomebamba, where MAE was the highest and was also reduced the most, by almost 4°C. Simultaneously, R^2 was improved for all stations, where the improvement was the smallest for Nkweseko (+ 0.07) and the highest for Tomebamba (+1.71). Overall, the mean represented by the MAE deviation at all stations could be reduced to less than 2°C.

4.4.3 Statistical and data structure analysis

These results are supported by statistical evidence. A one-sided Wilcoxon signed-rank test was performed on the test data from all stations, which was significant in all cases ($p < 0.001$). This indicates that the distributions of the residuals of T_{bc} (bias correction) are significantly smaller (i.e. better) than those of T_{rs} (baseline).

To understand and analyze the structure of the bias correction further, as an example, the first month of each test dataset including the bias correction is displayed in a timeseries plot figure 4.5.

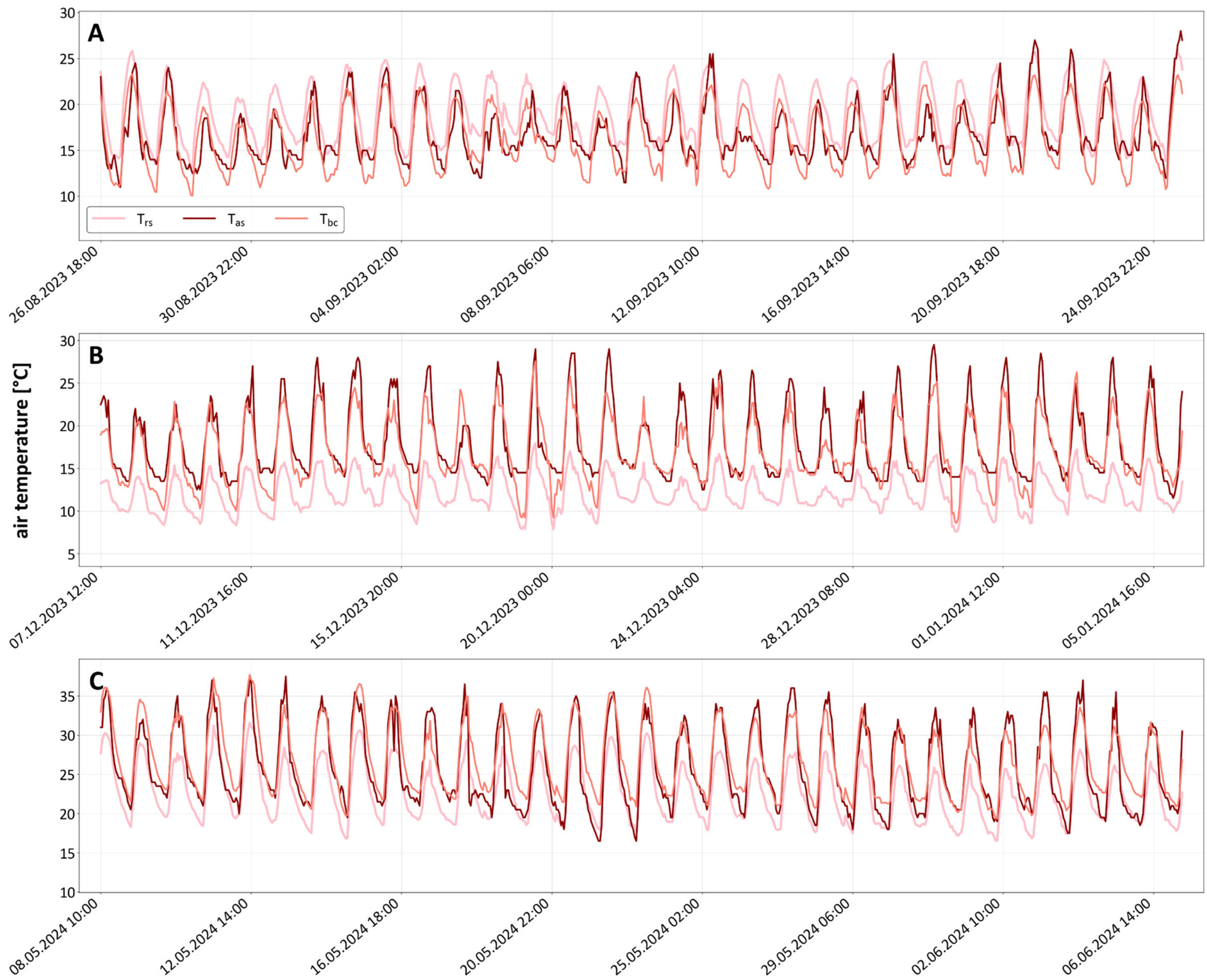


Figure 4.5: *Comparison of the first month (n = 720) of the test data (Tas, Trs and Tbc) for (A) Nkweseko, (B) Tomebamba and (C) Don Tito.*

Examining the time series for the first month of each station in Figure 4.5 gives an impression of which time of the day is corrected well and which not. While daily maximum temperatures are often overestimated by ERA5-Land at Nkweseko, the bias correction corrected this for multiple days. Lower temperatures on the other hand became slightly worse after the bias correction. Rising temperatures in the morning and falling temperatures in the evening, on the other hand, are corrected better. At Tomebamba (Figure 4.5B) both maximum and minimum temperatures were adjusted and fit better to the actual temperatures (T_{as}), while some maximum temperatures are still underestimated. Similar results could be observed for air temperature at Don Tito, where especially underestimated daily maximum temperatures were corrected in most cases (Figure 4.5C).

Considering these differences, especially for daily maximum temperatures, which are predominantly recorded at noon or early afternoon, boxplots averaging air temperature deviation (from T_{as}) for each hour of the day for all stations are presented in figure 4.6.

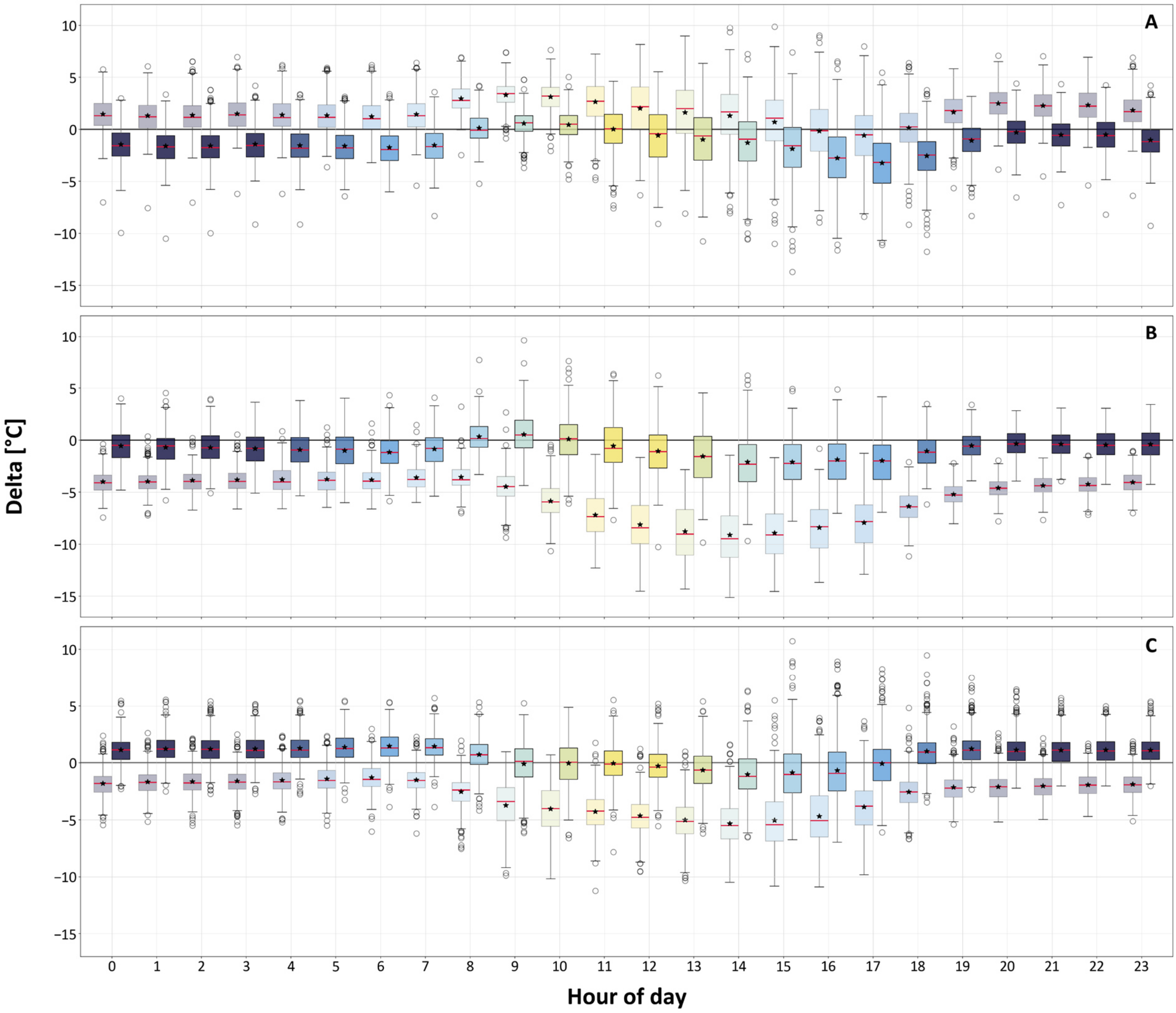


Figure 4.6: *Boxplots comparing delta (T_{bc} - T_{as}) for every hour of the day at (A) Nkweseko, (B) Tomebamba and (C) Don Tito. The pale boxplots in the background display delta for T_{rs} - T_{as}. The red line indicates the median, the black star, the mean value of each hour while the colors represent the time of the day.*

While for Nkweseko (Figure 4.6A) T_{rs} overestimated T_{as} for most of the day, except for the late afternoon where it was quite similar, it turns to the opposite for the bias correction. Here air temperature is underestimated at most hours of the day, with only low deviations during the morning from 8-10am, while the highest are observed in the late afternoon from 4 to 6pm. For Tomebamba (Figure 4.6B) the most substantial improvement can be recognized, similar to prior results. Deviation for all hours of the day is substantially reduced, only leaving higher deviations at noon to late afternoon (1-5pm). At Don Tito, however, the bias correction increases overall air temperature values excessively. This causes the slight underestimation to turn into a slight overestimation at night. Meanwhile, values before noon show low deviation, but values in the afternoon are still underestimated.

A statistical analysis using the Friedman test comparing deviation for all hours of the day was significant ($p < 0.001$) indicating statistically significant differences between the different times of the day. To follow this up, a post-hoc Wilcoxon signed-rank test was used to compare each hour of the day with 0 (i.e. no deviation). For Nkweseko the test indicated statistically significant deviations from zero for all hours of the day except 8 and 11 o'clock, while for Tomebamba only for 8 and 10 o'clock no significant differences were determined. On the other hand, for all hours of the day except 9-11 and 5 pm significant differences were determined for Don Tito.

4.5 Discussion

The utilization of participatory monitoring air temperature data for bias correction enhanced the accuracy of ERA5-Land 2m air temperature in various tropical remote locations. Simple linear regression was carried out for bias correction. However, it should be noted that the degree of correction exhibited by the stations was not uniform, with an average deviation of up to 2°C being observed. The performance of the stations was found to vary, thus providing a foundation for critical analysis of the approach and data utilized.

4.5.1 Using participatory monitoring for bias correction

Several previous studies analyzed bias correction or downscaling on ERA5-Land data in different study regions with varying spatiotemporal scales. Study regions covered a wide range of topography from the Yellow river basin in China (L. Xu et al., 2024; Zhang et al., 2024) to mountainous topography in the High Atlas Mountains (Sebbar et al., 2023) or the Southern Alps (Dhawan et al., 2024; Niazkar et al., 2023). None of them covered tropical mountainous regions. Among different metrics for validation, MAE, RMSE and R^2 were used most frequently, similar to our case (Dhawan et al., 2024; Niazkar et al., 2023; Sebbar et al., 2023; L. Xu et al., 2024; Zhang et al., 2024). While in our study only linear regression model was applied to correct T_{rs}, the methods varied in most investigations using different statistical methods, machine learning or machine learning ensembles (Dhawan et al., 2024; Niazkar et al., 2023; Sebbar et al., 2023; L. Xu et al., 2024; Zhang et al., 2024). Regularly, multiple weather stations were used as the main source of correction data (Dhawan et al., 2024; Niazkar et al., 2023; Sebbar et al., 2023; L. Xu et al., 2024; Zhang et al., 2024; Zhu et al., 2021) while for our analysis only one station for each location was used, what might be a limiting factor for our approach. In some cases additional data like digital elevation models, land surface temperature, land cover data or vegetation indices were used as well (Sebbar et al., 2023; L. Xu et al., 2024; Zhang et al., 2024).

On the other hand, the use of analogue participatory monitoring data for bias correction in ERA5-Land data remains unexplored prior to this work. PM datasets have characteristics like irregular measurement frequency or unequal coverage of different time periods and therefore do not replicate a full temporal cycle, as our analysis has shown. Furthermore, the PM temperature data used in our study is of good quality but still deviates slightly from automatically measured air temperature [30]. These characteristics are considerable weaknesses of PM data in comparison to automatically measured data and should be taken into account when comparing our results with those of previous studies. It is therefore logical that the accuracy of our approach does not fully match that of approaches that use regular weather station data. Hence, this might be the greatest limitation of our data. For instance, Dhawan et al. (2024) and Niazkar et al. (2023) achieved similar or even better results for MAE (0.7 – 2.0°), RMSE (1.0 – 2.5°C) and R^2 (up to 0.98) but also used more complex machine learning models and high-resolution data for air temperature correction.

While for most climate related bias corrections annual, monthly or daily temporal scales are used, these two recent examples also showed the applicability of bias correction of hourly data, which has not been investigated as frequently (Dhawan et al., 2024). While for the two aforementioned studies several years of data were used to train the models, in our case 23 to 275 single analog measurements were used, as not more data was available from the PM approach. This shows that even with a small training dataset substantial enhancements are already possible for single locations.

Overall, using PM data in combination with simple linear regression proved suitable for enhancing the accuracy of air temperature data for specific locations in our study regions. However, this approach may not achieve the same level of accuracy as other, more precise methods that use complex data and different bias correction techniques.

4.5.2 Varying model performance

Our results show that the bias correction performed quite differently for the three stations, with having achieved the best results (MAE = 1.54 °C) with the lowest number of training samples at Don Tito (n = 23). On the other hand, the station with the most samples, Nkweseko (n = 275) performed worse (MAE = 1.98 °C). None of the stations provide measurements for every hour of the day: during the night between 0 and 6am no measurements are available at all. The analysis of the temporal distribution of T_{pm} and the results for each hour of the day did not indicate that the incorrect prediction of T_{bc} can be linked to this incomplete temporal coverage, as higher deviations were observed predominantly during the hours for which most data were collected. However, it cannot be entirely ruled out that the correction could have been even better with a data set covering all times of day evenly. Also, the deviation of T_{pm} from the automatic sensor data (Mitze et al., 2026) can only partially explain the performance, as Nkweseko (MAE = 0.74°C) has the lowest and Don Tito (MAE = 1.44°C) has the highest deviation. As the varying performance and limitation of this approach cannot be entirely attributed to structure of the participatory monitoring data, the cause might also be linked to the underlying structure of the ERA5-Land data and the bias correction method used.

The model based on linear correlation can only work sufficiently with high correlation between the predictor (T_{rs}) and the target variable (T_{pm}). Although this is quite high in our study with r_{spear} ranging from 0.83 to 0.86, it does not establish a perfect monotonic relationship between the variables. There might be several explanations for that. Firstly, our PM data is of good quality but not a precise replication of automatically measured data which might affect the correlation with ERA5-Land data, which is partly based on automatically measured data, to some degree. A more critical point is probably the spatial reproduction of temperature variability, which the 9 km pixel of ERA5-Land offers for mountainous areas. There is likely a correlation between the ERA5-Land temperature values and ground-based measurements within a given pixel. However, a perfect fit may not be ensured for every location within a pixel because the ERA5-Land value is an average for the entire area. This is exemplified by the stations Nkweseko and Tomebamba, which are both located in terrain, where within a short distance the elevation increases sharply. The 9km pixel can therefore only offer a rough estimation for the area, which can be seen in the example of Nkweseko displayed in figure 4.7.

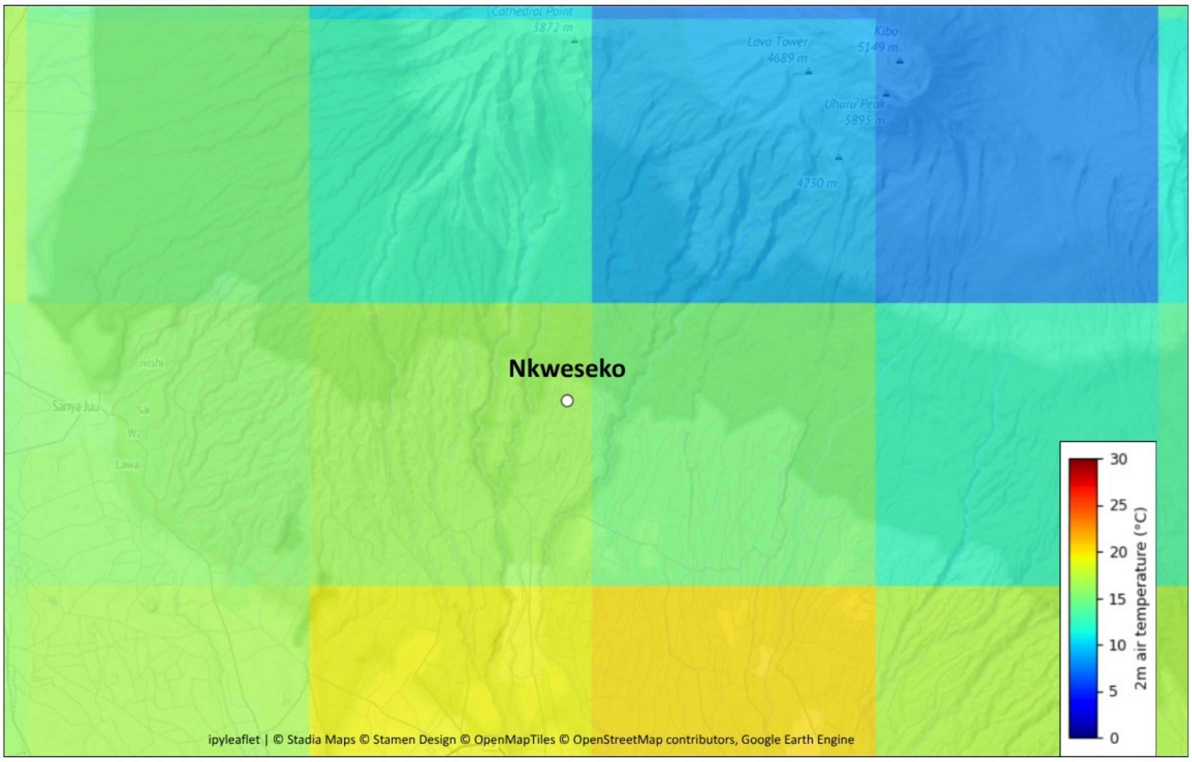


Figure 4.7: *Example of ERA5-Land 2m air temperature pixels at the southern slopes of Kilimanjaro around the Nkweseko station (white point) from the 30th of April 2024.*

While the deviation for Tomebamba was reduced from 5.49 °C to 1.76 °C, the highest of all stations in the raw ERA5-Land data, the deviation for Nkweseko was not that high and could not be reduced to a similar extent, from 2.27 °C to 1.98 °C. Therefore, it is highly possible that the approach might work better or worse depending on the location within the pixel.

Another reason might be the linear regression model used for the bias correction which can only transmit a linear relationship. Using more complex models that catch other characteristics apart from just linear relationships between variables could produce different results here. This limitation can be seen in the hourly based comparison of the deviation in figure 4.6. The bias correction improved overall accuracy of the air temperature by only shifting the overall dataset into a specific direction, with air temperature at Nkweseko air temperature being lowered, while being raised for Tomebamba and Don Tito. The original ERA5-Land trend used for prediction can still be recognized. As linear regression cannot be expected to considerably change the distribution of the data, other methods of bias correction besides linear regression should be tested in further research using PM data.

In summary, the limitations of our approach, which applies simple linear regression and participatory monitoring data, can most likely be linked to the original structure of the ERA5-Land air temperature data and the regression method used. The impact of restricted availability of PM data also remains a possible limitation, albeit not definitively established.

4.6 Conclusions

This analysis investigated how participatory monitoring could be used for the bias correction of ERA5-Land air temperature data in diverse remote tropical mountainous regions in Ecuador, Honduras and Tanzania where limited automatically measured hydrometeorological data are available. For bias correction, linear regression was used and for validation purpose ERA5-Land and the corrected data were compared to automatically measured air temperature using MAE, RMSE and R^2.

Statistically significant smaller residuals were detected for the bias-corrected air temperature at all locations. Validation results with different metrics showed varying degree of success of the bias correction with the lowest MAE at Don Tito (Honduras) with 1.54°C and the highest for Nkweseko (Tanzania) with 1.98 °C. A thorough examination of the test data revealed that the automatically measured data does not perfectly align with the bias-corrected data with only a few hours of the day having no significant deviations from zero.

It can be concluded that with small participatory monitoring datasets it is possible to significantly enhance ERA5-Land air temperature accuracy at remote tropical mountainous locations. Asymmetric distribution of the training data is most likely not leading to distorted prediction results, rather the ERA5-Land data and the bias correction method used can influence the results to a higher degree. Therefore, participatory monitoring in combination with ERA5-Land might be one solution for remote regions with insufficient automatic weather station density. As this data is required for a variety of applications, such as for agriculture, health protection or hydrology, it might be a crucial contribution to an improved data situation in such regions.

In the next step, other geographic regions should be tested using analog temperature measurements. Another possible next step could be an extended analysis of other ERA5-Land parameters that might have higher correlation with PM air temperature data. Here, for instance, earth skin temperature or soil temperature could be used to test if further accuracy-improvements are possible. Also, regressions for other parameters like relative humidity could be investigated, if such measurements are also available from PM projects. To check if other bias correction methods are better suited for PM data, we suggest testing this approach with other techniques. This could be traditional machine learning methods, like random forest regression or even deep learning models like Multilayer Perceptron's. As the location of the PM station within the ERA5-Land pixel seems to have an influence on the correction accuracy, other PM data at different locations and varying elevations within the same pixel should be tested as well to further analyze the influence of the location of the station.

4.7 Acknowledgments

The authors would like to thank all our project partners for their help in setting up the hydrometeorological monitoring network in Honduras, Ecuador and Tanzania. Special thanks go to Mancomunidad de Municipios del Parque Nacional Montaña de Celaque (Mapance) and Alliance of Bioversity International and CIAT in Honduras; Empresa Pública Municipal de Telecomunicaciones, Agua Potable, Saneamiento y Gestión Ambiental del cantón Cuenca (ETAPA) and Universidad de Cuenca in Ecuador; the Kilimanjaro National Park Authority (KINAPA) and the DFG Research Unit "The role of nature for human well-being in the Kilimanjaro Social-Ecological System (Kili-SES)" in Tanzania, for their support of the HydroCrowd project. Finally, we would like to say thank you to all volunteers for their invaluable contribution, without their dedication and commitment, the success of this project would not have been attainable.

4.8 Data availability statement

The datasets and scripts used for this study can be obtained from Zenodo [https://doi.org/10.5281/zenodo.17854458].

References

Abraham, E. M. (2018). Understanding climate change adaptation in Africa: Key considerations. *Management of Environmental Quality: An International Journal*, *29*(1), 165–179. https://doi.org/10.1108/MEQ-01-2017-0001

Aguilar, A. (2005). Remote Sensing of Forest Regeneration in Highland Tropical Forests. *GIScience & Remote Sensing*, *42*(1), 66–79. https://doi.org/10.2747/1548-1603.42.1.66

Alemayehu, T., Griensven, A. van, Senay, G. B., & Bauwens, W. (2017). Evapotranspiration Mapping in a Heterogeneous Landscape Using Remote Sensing and Global Weather Datasets: Application to the Mara Basin, East Africa. *Remote Sensing*, *9*(4), Article 4. https://doi.org/10.3390/rs9040390

Anderson, D. L., & Devenish, C. (2009). Áreas Importantes para la Conservación de las Aves AMÉRICA. In *Important Bird Areas Americas—Priority sites for biodiversity conservation: Honduras* (pp. 255–260). http://www.fecomol.org/pdf/%C3%81reas_Importantes_para_la_Conservaci%C3%B3n_de_las_Aves_America_Honduras.pdf

Anderson, T. W., & Darling, D. A. (1952). Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, *23*(2), 193–212. https://doi.org/10.1214/aoms/1177729437

Anwar, M. R., Liu, D. L., Macadam, I., & Kelly, G. (2013). Adapting agriculture to climate change: A review. *Theoretical and Applied Climatology*, *113*(1), 225–245. https://doi.org/10.1007/s00704-012-0780-1

Appelhans, T., Mwangomo, E., Otte, I., Detsch, F., Nauss, T., & Hemp, A. (2016). Eco-meteorological characteristics of the southern slopes of Kilimanjaro, Tanzania. *International Journal of Climatology*, *36*(9), 3245–3258. https://doi.org/10.1002/joc.4552

Arcusa, S. H., Schneider, T., Mosquera, P. V., Vogel, H., Kaufman, D., Szidat, S., & Grosjean, M. (2020). Late Holocene tephrostratigraphy from Cajas National Park, southern Ecuador. *Andean Geology*, *47*(3), 508. https://doi.org/10.5027/andgeoV47n3-3301

Arienzo, M. M., Collins, M., & Jennings, K. S. (2021). Enhancing Engagement of Citizen Scientists to Monitor Precipitation Phase. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.617594

Arnaud, P., Bouvier, C., Cisneros, L., & Dominguez, R. (2002). Influence of rainfall spatial variability on flood prediction. *Journal of Hydrology*, *260*(1), 216–230. https://doi.org/10.1016/S0022-1694(01)00611-4

Avellaneda, P. M., Ficklin, D. L., Lowry, C. S., Knouft, J. H., & Hall, D. M. (2020). Improving Hydrological Models With the Assimilation of Crowdsourced Data. *Water Resources Research*, *56*(5), e2019WR026325. https://doi.org/10.1029/2019WR026325

Bandini, F., Butts, M., Jacobsen, T. V., & Bauer-Gottwein, P. (2017). Water level observations from unmanned aerial vehicles for improving estimates of surface water–groundwater interaction. *Hydrological Processes*, *31*(24), 4371–4383. https://doi.org/10.1002/hyp.11366

Bandowe, B. A. M., Fränkl, L., Grosjean, M., Tylmann, W., Mosquera, P. V., Hampel, H., & Schneider, T. (2018). A 150-year record of polycyclic aromatic compound (PAC) deposition from high Andean Cajas National Park, southern Ecuador. *Science of The Total Environment*, *621*, 1652–1663. https://doi.org/10.1016/j.scitotenv.2017.10.060

Barros, A. P., & Arulraj, M. (2020). Remote Sensing of Orographic Precipitation. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds), *Satellite Precipitation Measurement: Volume 2* (pp. 559–582). Springer International Publishing. https://doi.org/10.1007/978-3-030-35798-6_6

Barros, N., Sobral, P., Moreira, R. S., Vargas, J., Fonseca, A., Abreu, I., & Guerreiro, M. S. (2024). SchoolAIR: A Citizen Science IoT Framework Using Low-Cost Sensing for Indoor Air Quality Management. *Sensors*, *24*(1), Article 1. https://doi.org/10.3390/s24010148

Beele, E., Reyniers, M., Aerts, R., & Somers, B. (2022). Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium. *Earth System Science Data*, *14*(10), 4681–4717. https://doi.org/10.5194/essd-14-4681-2022

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

Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Understanding of Science*, *25*(1), 2–16. https://doi.org/10.1177/0963662515607406

Branchini, S., Meschini, M., Covi, C., Piccinetti, C., Zaccanti, F., & Goffredo, S. (2015). Participating in a Citizen Science Monitoring Program: Implications for Environmental Education. *PLOS ONE*, *10*(7), e0131812. https://doi.org/10.1371/journal.pone.0131812

Brandt, P., Hamunyela, E., Herold, M., de Bruin, S., Verbesselt, J., & Rufino, M. C. (2018). Sustainable intensification of dairy production can reduce forest disturbance in Kenyan montane forests. *Agriculture, Ecosystems & Environment*, *265*, 307–319. https://doi.org/10.1016/j.agee.2018.06.011

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buytaert, W., Célleri, R., De Bièvre, B., Cisneros, F., Wyseure, G., Deckers, J., & Hofstede, R. (2006). Human impact on the hydrology of the Andean páramos. *Earth-Science Reviews*, *79*(1), 53–72. https://doi.org/10.1016/j.earscirev.2006.06.002

Buytaert, W., Dewulf, A., De Bièvre, B., Clark, J., & Hannah, D. M. (2016). Citizen Science for Water Resources Management: Toward Polycentric Monitoring and Governance? *Journal of Water Resources Planning and Management*, *142*(4), 01816002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000641

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., De Bièvre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D. M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., … Zhumanova, M. (2014). Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, *2*. https://doi.org/10.3389/feart.2014.00026

Campos Zeballos, J., Valencia, J., Codalli, F., Mitze, F., Shagega, F., Weeser, B., & Jacobs, S. (2025). Evaluating participatory monitoring in mountainous tourist regions. *Frontiers in Environmental Science*, *13*, 1537278. https://doi.org/10.3389/fenvs.2025.1537278

Carrillo-Rojas, G., Silva, B., Córdova, M., Célleri, R., & Bendix, J. (2016). Dynamic Mapping of Evapotranspiration Using an Energy Balance-Based Model over an Andean

Páramo Catchment of Southern Ecuador. *Remote Sensing*, *8*(2), Article 2. https://doi.org/10.3390/rs8020160

Cavazos, T., Bettolli, M. L., Campbell, D., Sánchez Rodríguez, R. A., Mycoo, M., Arias, P. A., Rivera, J., Reboita, M. S., Gulizia, C., Hidalgo, H. G., Alfaro, E. J., Stephenson, T. S., Sörensson, A. A., Cerezo-Mota, R., Castellanos, E., Ley, D., & Mahon, R. (2024). Challenges for climate change adaptation in Latin America and the Caribbean region. *Frontiers in Climate*, *6*. https://doi.org/10.3389/fclim.2024.1392033

Celleri, R., Willems, P., Buytaert, W., & Feyen, J. (2007). Space–time rainfall variability in the Paute basin, Ecuadorian Andes. *Hydrological Processes*, *21*(24), 3316–3327. https://doi.org/10.1002/hyp.6575

Chacon-Hurtado, J. C., Alfonso, L., & Solomatine, D. P. (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. https://doi.org/10.5194/hess-21-3071-2017

Chaubey, I., Haan, C. T., Grunwald, S., & Salisbury, J. M. (1999). Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology*, *220*(1), 48–61. https://doi.org/10.1016/S0022-1694(99)00063-3

Chiew, F. H. S., Peel, M. C., & Western, A. W. (2002). Pplication and testing of the simple rainfall-runoff model SIMHYD. *Water Resources Publications*, *Mathematical models of small watershed hydrology and applications*, 335–367.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., & Papalexiou, S. M. (2021). The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resources Research*, *57*(9), e2020WR029001. https://doi.org/10.1029/2020WR029001

COP 21: The Paris Agreement (2025). https://unfccc.int/sites/default/files/resource/parisagreement_publication.pdf

Copernicus. (2025). *Climate reanalysis | Copernicus*. https://climate.copernicus.eu/climate-reanalysis

Copernicus Climate Change Service. (2019). *ERA5-Land hourly data from 1950 to present* [Data set]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/CDS.E2161BAC

Copernicus Climate Change Service. (2025, November 4). *Copernicus Climate Change Service—Global Impacts* [Glossar]. What Is Bias Correction? https://climate.copernicus.eu/sites/default/files/2021-01/infosheet7.pdf

Correa, A., Windhorst, D., Crespo, P., Célleri, R., Feyen, J., & Breuer, L. (2016). Continuous versus event-based sampling: How many samples are required for deriving general hydrological understanding on Ecuador's páramo region? *Hydrological Processes*, *30*(22), 4059–4073. https://doi.org/10.1002/hyp.10975

Croke, B. F. W., & Jakeman, A. J. (2004). A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environmental Modelling & Software*, *19*(1), 1–5. https://doi.org/10.1016/j.envsoft.2003.09.001

Danielsen, F., Burgess, N. D., Balmford, A., Donald, P. F., Funder, M., Jones, J. P. G., Alviola, P., Balete, D. S., Blomley, T., Brashares, J., Child, B., Enghoff, M., Fjeldså, J., Holt, S., Hübertz, H., Jensen, A. E., Jensen, P. M., Massao, J., Mendoza, M. M., … Yonten, D. (2009). Local Participation in Natural Resource Monitoring: A Characterization of Approaches. *Conservation Biology*, *23*(1), 31–42. https://doi.org/10.1111/j.1523-1739.2008.01063.x

Danielsen, F., Burgess, N. D., Jensen, P. M., & Pirhofer-Walzl, K. (2010). Environmental monitoring: The scale and speed of implementation varies according to the degree of peoples involvement. *Journal of Applied Ecology*, *47*(6), 1166–1168. https://doi.org/10.1111/j.1365-2664.2010.01874.x

Dasgupta, A., Grimaldi, S., Ramsankaran, R., Pauwels, V. R. N., & Walker, J. P. (2022). A simple framework for calibrating hydraulic flood inundation models using Crowd-sourced water levels. *Journal of Hydrology*, *614*, 128467. https://doi.org/10.1016/j.jhydrol.2022.128467

Davids, J. C., Devkota, N., Pandey, A., Prajapati, R., Ertis, B. A., Rutten, M. M., Lyon, S. W., Bogaard, T. A., & van de Giesen, N. (2019). Soda Bottle Science—Citizen Science Monsoon Precipitation Monitoring in Nepal. *Frontiers in Earth Science*, *7*. https://doi.org/10.3389/feart.2019.00046

Davids, J. C., Rutten, M. M., Pandey, A., Devkota, N., van Oyen, W. D., Prajapati, R., & van de Giesen, N. (2019). Citizen science flow – an assessment of simple streamflow measurement methods. *Hydrology and Earth System Sciences*, *23*(2), 1045–1065. https://doi.org/10.5194/hess-23-1045-2019

Daw, T., Brown, K., Rosendo, S., & Pomeroy, R. (2011). Applying the ecosystem services concept to poverty alleviation: The need to disaggregate human well-being. *Environmental Conservation*, *38*(4), 370–379. https://doi.org/10.1017/S0376892911000506

de Vos, L., Droste, A., Zander, M., Overeem, A., Leijnse, H., Heusinkveld, B., Steeneveld, G.-J., & Uijlenhoet, R. (2020). *Opportunistic weather sensors: An Amsterdam case study of private weather stations, commercial microwave links and smartphones*. 10868. EGU General Assembly Conference Abstracts. https://doi.org/10.5194/egusphere-egu2020-10868

Dhawan, P., Torre, D. D., Niazkar, M., Kaffas, K., Larcher, M., Righetti, M., & Menapace, A. (2024). A comprehensive comparison of bias correction methods in climate model simulations: Application on ERA5-Land across different temporal resolutions. *Heliyon*, *10*(23). https://doi.org/10.1016/j.heliyon.2024.e40352

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Duwal, S., Prajapati, R., Upadhyay, S., Neupane, S., Lakhe, H., Thapa, B. R., Davids, J. C., & Talchabhadel, R. (2025). Leveraging a Citizen Science Approach for Rainfall Monitoring: Evaluating Performance and Reliability To Complement Standard Datasets. *Earth Systems and Environment*. https://doi.org/10.1007/s41748-025-00920-8

Eisma, J. A., Schoups, G., Davids, J. C., & van de Giesen, N. (2023). A Bayesian model for quantifying errors in citizen science data: Application to rainfall observations from Nepal. *Hydrology and Earth System Sciences*, *27*(19), 3565–3579. https://doi.org/10.5194/hess-27-3565-2023

Elmore, K. L., Flamig, Z. L., Lakshmanan, V., Kaney, B. T., Farmer, V., Reeves, H. D., & Rothfusz, L. P. (2014). MPING: Crowd-Sourcing Weather Reports for Research. *Bulletin of the American Meteorological Society*, *95*(9), 1335–1342. https://doi.org/10.1175/BAMS-D-13-00014.1

Etter, S., Strobl, B., Seibert, J., & van Meerveld, H. J. I. (2020). Value of Crowd-Based Water Level Class Observations for Hydrological Model Calibration. *Water Resources Research*, *56*(2), e2019WR026108. https://doi.org/10.1029/2019WR026108

Etter, S., Strobl, B., van Meerveld, I., & Seibert, J. (2020). Quality and timing of crowd-based water level class observations. *Hydrological Processes*, *34*(22), 4365–4378. https://doi.org/10.1002/hyp.13864

Fankhauser, S., & McDermott, T. K. J. (2014). Understanding the adaptation deficit: Why are poor countries more vulnerable to climate events than rich countries? *Global Environmental Change*, *27*, 9–18. https://doi.org/10.1016/j.gloenvcha.2014.04.014

FAO. (1966). *Inventario Nacional de Recursos Fisicos (AID/RIC GIPR No 5), (Soils-Agricultural—Honduras)* [Soil Map]. https://data.apps.fao.org/map/catalog/srv/eng/catalog.search#/metadata/d3ac2953-e097-43dc-8d8b-d80294fc7519

Ficchì, A., Perrin, C., & Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *Journal of Hydrology*, *538*, 454–470. https://doi.org/10.1016/j.jhydrol.2016.04.016

Fienen, M. N., & Lowry, C. S. (2012). Social.Water—A crowdsourcing tool for environmental data acquisition. *Computers & Geosciences*, *49*, 164–169. https://doi.org/10.1016/j.cageo.2012.06.015

Fredianto, & Putri, D. A. P. (2023). Comparison of the interquartile range algorithm and local outlier factor on Australian weather data sets. *AIP Conference Proceedings*, *2727*(1), 040010. https://doi.org/10.1063/5.0141897

Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*. (world). https://www.tandfonline.com/doi/abs/10.1080/01621459.1937.10503522

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gianoli, A., & Bhatnagar, R. (2019). Managing the Water-Energy Nexus within a Climate Change Context—Lessons from the Experience of Cuenca, Ecuador. *Sustainability*, *11*(21), Article 21. https://doi.org/10.3390/su11215918

Gleason, C. J., & Durand, M. T. (2020). Remote Sensing of River Discharge: A Review and a Framing for the Discipline. *Remote Sensing*, *12*(7), Article 7. https://doi.org/10.3390/rs12071107

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, Big Remotely Sensed Data: Tools, Applications and Experiences*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gosset, M., Kunstmann, H., Zougmore, F., Cazenave, F., Leijnse, H., Uijlenhoet, R., Chwala, C., Keis, F., Doumounia, A., Boubacar, B., Kacou, M., Alpert, P., Messer, H., Rieckermann, J., & Hoedjes, J. (2016). Improving Rainfall Measurement in Gauge Poor Regions Thanks to Mobile Telecommunication Networks. *Bulletin of the American Meteorological Society*, *97*(3), ES49–ES51. https://doi.org/10.1175/BAMS-D-15-00164.1

Grimaldi, S., Li, Y., Pauwels, V. R. N., & Walker, J. P. (2016). Remote Sensing-Derived Water Extent and Level to Constrain Hydraulic Flood Forecasting Models: Opportunities and Challenges. *Surveys in Geophysics*, *37*(5), 977–1034. https://doi.org/10.1007/s10712-016-9378-y

Grizzetti, B., Lanzanova, D., Liquete, C., Reynaud, A., & Cardoso, A. C. (2016). Assessing water ecosystem services for water resource management. *Environmental Science & Policy*, *61*, 194–203. https://doi.org/10.1016/j.envsci.2016.04.008

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gütlein, A., Gerschlauer, F., Kikoti, I., & Kiese, R. (2018). Impacts of climate and land use on N2O and CH4 fluxes from tropical ecosystems in the Mt. Kilimanjaro region, Tanzania. *Global Change Biology*, *24*(3), 1239–1255. https://doi.org/10.1111/gcb.13944

Hahn, C., Garcia-Marti, I., Sugier, J., Emsley, F., Beaulant, A.-L., Oram, L., Strandberg, E., Lindgren, E., Sunter, M., & Ziska, F. (2022). Observations from Personal Weather Stations—EUMETNET Interests and Experience. *Climate*, *10*(12), Article 12. https://doi.org/10.3390/cli10120192

Hansen, B. C. S., Rodbell, D. T., Seltzer, G. O., León, B., Young, K. R., & Abbott, M. (2003). Late-glacial and Holocene vegetational history from two sites in the western Cordillera of southwestern Ecuador. *Palaeogeography, Palaeoclimatology, Palaeoecology, Late-Quaternary Palaeoclimates of the Southern Tropical Andes and Adjacent Regions*, *194*(1), 79–108. https://doi.org/10.1016/S0031-0182(03)00272-4

Harden, C. P. (2006). Human impacts on headwater fluvial systems in the northern and central Andes. *Geomorphology*, *79*(3–4), 249–263. https://doi.org/10.1016/j.geomorph.2006.06.021

Hargreaves, G. H., & Samani, Z. A. (1985). Reference Crop Evapotranspiration from Temperature. *Applied Engineering in Agriculture*, *1*(2), 96–99. https://doi.org/10.13031/2013.26773

Hemp, A. (2005). Climate change-driven forest fires marginalize the impact of ice cap wasting on Kilimanjaro. *Global Change Biology*, *11*(7), 1013–1023. https://doi.org/10.1111/j.1365-2486.2005.00968.x

Hemp, A. (2006). Continuum or zonation? Altitudinal gradients in the forest vegetation of Mt. Kilimanjaro. *Plant Ecology*, *184*(1), 27–42. https://doi.org/10.1007/s11258-005-9049-4

Hemp, A., Hemp, C., & Winter, J. C. (1998). Der Kilimanjaro: Lebensräume zwischen tropischer Hitze und Gletschereis. *Natur und Mensch*, (1998), 5–28.

Hemp, A., & Hemp, J. (2024). Weather or not—Global climate databases: Reliable on tropical mountains? *PLOS ONE*, *19*(3), e0299363. https://doi.org/10.1371/journal.pone.0299363

Hills, R. C. (1979). The Structure of the Inter-Tropical Convergence Zone in Equatorial Africa and Its Relationship to East African Rainfall. *Transactions of the Institute of British Geographers*, *4*(3), 329–352. https://doi.org/10.2307/622055

Hoffer, R. M. (1972, January 1). *The importance of ground truth data in remote sensing*. UN Panel Meeting on the Estab. and Implementation of Res. Programs in Remote Sensing. https://ntrs.nasa.gov/citations/19730007768

Hooker, J., Duveiller, G., & Cescatti, A. (2018). A global dataset of air temperature derived from satellite remote sensing and weather stations. *Scientific Data*, *5*(1), 180246. https://doi.org/10.1038/sdata.2018.246

Houska, T., Kraft, P., Chamorro-Chavez, A., & Breuer, L. (2015). SPOTting Model Parameters Using a Ready-Made Python Package. *PLOS ONE*, *10*(12), e0145180. https://doi.org/10.1371/journal.pone.0145180

Huang, Y., & Bardossy, A. (2020). Impacts of Data Quantity and Quality on Model Calibration: Implications for Model Parameterization in Data-Scarce Catchments. *Water*, *12*(9), Article 9. https://doi.org/10.3390/w12092352

Huang, Y., Bárdossy, A., & Zhang, K. (2019). Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data. *Hydrology and Earth System Sciences*, *23*(6), 2647–2663. https://doi.org/10.5194/hess-23-2647-2019

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F., Tan, J., Wolff, D. B., & Xie, P. (2020). Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds), *Satellite Precipitation Measurement: Volume 1* (pp. 343–353). Springer International Publishing. https://doi.org/10.1007/978-3-030-24568-9_19

Instituto Nacional de Estadística y Censos. (2022). *Censo de Ecuador 2022*. Censo de Ecuador a partir de 2022. https://www.censoecuador.gob.ec/resultados-censo/

IPCC. (2023a). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st edn). Cambridge University Press. https://doi.org/10.1017/9781009325844

IPCC. (2023b). *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]* (Climate Change 2023, pp. 35–115). IPCC. doi: 10.59327/IPCC/AR6-9789291691647

ISRIC. (2007). *Soil and Terrain Database for Kenya (KENSOTER), version 2.0*. ISRIC Data Hub. https://data.isric.org/geonetwork/srv/api/records/73e27136-9efe-49e4-af35-fd98b841d467

Jacobs, S. R., Timbe, E., Weeser, B., Rufino, M. C., Butterbach-Bahl, K., & Breuer, L. (2018). Assessment of hydrological pathways in East African montane catchments under different land use. *Hydrology and Earth System Sciences*, *22*(9), 4981–5000. https://doi.org/10.5194/hess-22-4981-2018

Jacobs, S. R., Weeser, B., Guzha, A. C., Rufino, M. C., Butterbach-Bahl, K., Windhorst, D., & Breuer, L. (2018). Using High-Resolution Data to Assess Land Use Impact on Nitrate Dynamics in East African Tropical Montane Catchments. *Water Resources Research*, *54*(3), 1812–1830. https://doi.org/10.1002/2017WR021592

Jacobs, S. R., Weeser, B., Rufino, M. C., & Breuer, L. (2020). Diurnal Patterns in Solute Concentrations Measured with In Situ UV-Vis Sensors: Natural Fluctuations or Artefacts? *Sensors*, *20*(3). https://doi.org/10.3390/s20030859

James, R., Washington, R., Abiodun, B., Kay, G., Mutemi, J., Pokam, W., Hart, N., Artan, G., & Senior, C. (2018). Evaluating Climate Models with an African Lens. *Bulletin of the American Meteorological Society*, *99*(2), 313–336. https://doi.org/10.1175/BAMS-D-16-0090.1

Jian, J., Ryu, D., Costelloe, J. F., & Su, C.-H. (2017). Towards hydrological model calibration using river level measurements. *Journal of Hydrology: Regional Studies*, *10*, 95–109. https://doi.org/10.1016/j.ejrh.2016.12.085

Juston, J., Seibert, J., & Johansson, P.-O. (2009). Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. *Hydrological Processes*, *23*(21), 3093–3109. https://doi.org/10.1002/hyp.7421

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, How Much of the Earth's Surface Is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kimani, M. W., Hoedjes, J. C. B., & Su, Z. (2017). An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa. *Remote Sensing*, *9*(5), Article 5. https://doi.org/10.3390/rs9050430

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., DelVecchia, A. G., Fritz, K. M., Shanafield, M., Burgin, A. J., Zimmer, M. A., Datry, T., Dodds, W. K., Jones, C. N., Mims, M. C., Franklin, C., Hammond, J. C., Zipper, S., Ward, A. S., … Olden, J. D. (2022). Assessing placement bias of the global river gauge network. *Nature Sustainability*, *5*(7), Article 7. https://doi.org/10.1038/s41893-022-00873-0

Kraft, P., Vaché, K. B., Frede, H.-G., & Breuer, L. (2011). CMF: A Hydrological Programming Language Extension For Integrated Catchment Models. *Environmental Modelling & Software*, *26*(6), 828–830. https://doi.org/10.1016/j.envsoft.2010.12.009

Krebs, G., Kokkonen, T., Valtanen, M., Setälä, H., & Koivusalo, H. (2014). Spatial resolution considerations for urban hydrological modelling. *Journal of Hydrology*, *512*, 482–497. https://doi.org/10.1016/j.jhydrol.2014.03.013

Lee, M.-C., Leu, J.-M., Chan, H.-C., & Huang, W.-C. (2010). The measurement of discharge using a commercial digital video camera in irrigation canals. *Flow Measurement and Instrumentation*, *21*(2), 150–154. https://doi.org/10.1016/j.flowmeasinst.2010.02.002

Leichtle, T., Helgert, S., Müller, M., Handschuh, J., Erbertseder, T., Wurm, M., & Taubenböck, H. (2023). Opposing land surface and air temperatures from remote sensing and Citizen Science for quantification of the Urban Heat Island effect. *2023 Joint Urban Remote Sensing Event (JURSE)*, 1–5. https://doi.org/10.1109/JURSE57346.2023.10144135

Levizzani, V., & Cattani, E. (2019). Satellite Remote Sensing of Precipitation and the Terrestrial Water Cycle in a Changing Climate. *Remote Sensing*, *11*(19), Article 19. https://doi.org/10.3390/rs11192301

Li, Z.-L., Wu, H., Duan, S.-B., Zhao, W., Ren, H., Liu, X., Leng, P., Tang, R., Ye, X., Zhu, J., Sun, Y., Si, M., Liu, M., Li, J., Zhang, X., Shang, G., Tang, B.-H., Yan, G., & Zhou, C. (2023). Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications. *Reviews of Geophysics*, *61*(1), e2022RG000777. https://doi.org/10.1029/2022RG000777

Loglisci, N., Milelli, M., Iurato, J., Galia, T., Galizia, A., & Parodi, A. (2024). Validation of Citizen Science Meteorological Data: Can They Be Considered a Valid Help in Weather Understanding and Community Engagement? *Sensors*, *24*(14), Article 14. https://doi.org/10.3390/s24144598

Lompar, M., Lalić, B., Dekić, L., & Petrić, M. (2019). Filling Gaps in Hourly Air Temperature Data Using Debiased ERA5 Data. *Atmosphere*, *10*(1), 13. https://doi.org/10.3390/atmos10010013

Lowry, C. S., Fienen, M. N., Hall, D. M., & Stepenuck, K. F. (2019). Growing Pains of Crowdsourced Stream Stage Monitoring Using Mobile Phones: The Development of CrowdHydrology. *Frontiers in Earth Science*, *7*, 128. https://doi.org/10.3389/feart.2019.00128

lubis, S., Nurmandi, A., Ahmad, J., Purnomo, E. P., & Purwaningsih, T. (2025). The use of big data in environmental quality monitoring and its impact on policy development: A bibliometric analysis. *Discover Applied Sciences*, *7*(9), 947. https://doi.org/10.1007/s42452-025-07265-x

Luo, J., Wang, E., Shen, S., Zheng, H., & Zhang, Y. (2012). Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity. *Hydrological Processes*, *26*(26), 3953–3961. https://doi.org/10.1002/hyp.8420

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60.

Mao, Q., Peng, J., & Wang, Y. (2021). Resolution Enhancement of Remotely Sensed Land Surface Temperature: Current Status and Perspectives. *Remote Sensing*, *13*(7), Article 7. https://doi.org/10.3390/rs13071306

Margulis, S. A., Wood, E. F., & Troch, P. A. (2006). The Terrestrial Water Cycle: Modeling and Data Assimilation across Catchment Scales. *Journal of Hydrometeorology*, *7*(3), 309–311. https://doi.org/10.1175/JHM999.1

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., Bárdossy, A., & Morin, E. (2019). Precipitation frequency analysis from remotely sensed datasets: A focused review. *Journal of Hydrology*, *574*, 699–705. https://doi.org/10.1016/j.jhydrol.2019.04.081

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., & Solomatine, D. P. (2017). Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? *Hydrology and Earth System Sciences*, *21*(2), 839–861. https://doi.org/10.5194/hess-21-839-2017

McIntyre, N., Lee, H., Wheater, H., Young, A., & Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, *41*(12). https://doi.org/10.1029/2005WR004289

McIntyre, N. R., & Wheater, H. S. (2004). Calibration of an in-river phosphorus model: Prior evaluation of data needs and model uncertainty. *Journal of Hydrology*, *290*(1), 100–116. https://doi.org/10.1016/j.jhydrol.2003.12.003

McKay, M. D., Beckman, R. J., & Conover, W. J. (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. https://doi.org/10.1080/00401706.1979.10489755

Meier, F., Fenner, D., Grassmann, T., Otto, M., & Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, *19*, 170–191. https://doi.org/10.1016/j.uclim.2017.01.006

Meier, R., Tscheikner-Gratl, F., Steffelbauer, D. B., & Makropoulos, C. (2022). Flow Measurements Derived from Camera Footage Using an Open-Source Ecosystem. *Water*, *14*(3), Article 3. https://doi.org/10.3390/w14030424

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Meschini, M., Machado Toffolo, M., Marchini, C., Caroselli, E., Prada, F., Mancuso, A., Franzellitti, S., Locci, L., Davoli, M., Trittoni, M., Nanetti, E., Tittarelli, M., Bentivogli, R., Branchini, S., Neri, P., & Goffredo, S. (2021). Reliability of Data Collected by Volunteers: A Nine-Year Citizen Science Study in the Red Sea. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.694258

Meschini, M., Prati, F., Simoncini, G. A., Airi, V., Caroselli, E., Prada, F., Marchini, C., Machado Toffolo, M., Branchini, S., Brambilla, V., Covi, C., & Goffredo, S. (2021).

Environmental Awareness Gained During a Citizen Science Project in Touristic Resorts Is Maintained After 3 Years Since Participation. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.584644

Meyer Oliveira, A., Fleischmann, A. S., & Paiva, R. C. D. (2021). On the contribution of remote sensing-based calibration to model hydrological and hydraulic processes in tropical regions. *Journal of Hydrology*, *597*, 126184. https://doi.org/10.1016/j.jhydrol.2021.126184

Mishra, A. K., & Coulibaly, P. (2009). Developments in hydrometric network design: A review. *Reviews of Geophysics*, *47*(2). https://doi.org/10.1029/2007RG000243

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Codalli, F., Shagega, F. P., & Weeser, B. (2026). Validation of analog sensor measurements in hydrometeorological participatory monitoring in various tropical countries. *Frontiers in Earth Science*, *14*. https://doi.org/10.3389/feart.2026.1721642

Mitze, F., Jacobs, S. R., Breuer, L., Campos Zeballos, J., Dowling, T. P. F., & Weeser, B. (under review). *Bias correction of ERA5-Land air temperature in remote tropical regions using participatory monitoring data*.

Mitze, F., Jacobs, S. R., Breuer, L., Zeballos, J. C., & 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

Morán-Tejeda, E., Bazo, J., López-Moreno, J. I., Aguilar, E., Azorín-Molina, C., Sanchez-Lorenzo, A., Martínez, R., Nieto, J. J., Mejía, R., Martín-Hernández, N., & Vicente-Serrano, S. M. (2016). Climate trends and variability in Ecuador (1966-2011): Climate trends and variability in Ecuador. *International Journal of Climatology*, *36*(11), 3839–3855. https://doi.org/10.1002/joc.4597

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Moyer, J. D., & Hedden, S. (2020). Are we on the right path to achieve the sustainable development goals? *World Development*, *127*, 104749. https://doi.org/10.1016/j.worlddev.2019.104749

Mu, Q., Heinsch, F. A., Zhao, M., & Running, S. W. (2007). Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, *111*(4), 519–536. https://doi.org/10.1016/j.rse.2007.04.015

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Musa, Z. N., Popescu, I., & Mynett, A. (2015). A review of applications of satellite SAR, optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation. *Hydrology and Earth System Sciences*, *19*(9), 3755–3769. https://doi.org/10.5194/hess-19-3755-2015

Nagai, S., Nasahara, K. N., Akitsu, T. K., Saitoh, T. M., & Muraoka, H. (2020). Importance of the Collection of Abundant Ground-Truth Data for Accurate Detection of Spatial and Temporal Variability of Vegetation by Satellite Remote Sensing. In

Biogeochemical Cycles (pp. 223–244). American Geophysical Union (AGU). https://doi.org/10.1002/9781119413332.ch11

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

NCEI. (2023, November 7). *Day In, Day Out, NCEI Takes Data Seriously*. National Centers for Environmental Information (NCEI). https://www.ncei.noaa.gov/news/NOAAArchive

Ngcamu, B. S. (2023a). Climate change effects on vulnerable populations in the Global South: A systematic review. *Natural Hazards*, *118*(2), 977–991. https://doi.org/10.1007/s11069-023-06070-2

Ngcamu, B. S. (2023b). Climate change effects on vulnerable populations in the Global South: A systematic review. *Natural Hazards*, *118*(2), 977–991. https://doi.org/10.1007/s11069-023-06070-2

Niazkar, M., Piraei, R., Menapace, A., Dhawan, P., Torre, D. D., Larcher, M., & Righetti, M. (2023). Bias correction of ERA5-Land temperature data using standalone and ensemble machine learning models: A case of northern Italy. *Journal of Water and Climate Change*, *15*(1), 271–283. https://doi.org/10.2166/wcc.2023.669

Nilashi, M., Boon, O. K., Tan, G., Lin, B., & Abumalloh, R. (2023). Critical Data Challenges in Measuring the Performance of Sustainable Development Goals: Solutions and the Role of Big-Data Analytics. *Harvard Data Science Review*, *5*(3). https://doi.org/10.1162/99608f92.545db2cf

Njue, N., Stenfert Kroese, J., Gräf, J., Jacobs, S. R., Weeser, B., Breuer, L., & Rufino, M. C. (2019). Citizen science in hydrological monitoring and ecosystem services management: State of the art and future prospects. *Science of The Total Environment*, *693*, 133531. https://doi.org/10.1016/j.scitotenv.2019.07.337

NOAA. (2025). *Search | Climate Data Online (CDO) | National Climatic Data Center (NCDC)*. Climate Data Online Search. https://www.ncei.noaa.gov/cdo-web/search

Ochoa Tocachi, B. (2019). *Participatory hydrological monitoring to support sustainable water resources management* [Dissertation, Imperial College London]. 10.25560/75557

Otto, F. E. L., Harrington, L., Schmitt, K., Philip, S., Kew, S., Oldenborgh, G. J. van, Singh, R., Kimutai, J., & Wolski, P. (2020). Challenges to Understanding Extreme Weather Changes in Lower Income Countries. *Bulletin of the American Meteorological Society*, *101*(10), E1851–E1860. https://doi.org/10.1175/BAMS-D-19-0317.1

Overdevest, C., Orr, C. H., & Stepenuck, K. (2004). Volunteer Stream Monitoring and Local Participation in Natural Resource Issues. *Human Ecology Review, Society for Human Ecology*, *11*(2), 77–185.

Padrón, R., Feyen, J., Córdova, M., Crespo, P., & Célleri, R. (2020). Rain Gauge Inter-Comparison Quantifies Deficiencies in Precipitation Monitoring. *La Granja*, *31*(1), 7–20. https://doi.org/10.17163/lgr.n31.2020.01

Pandeya, B., Buytaert, W., & Potter, C. (2021). Designing citizen science for water and ecosystem services management in data-poor regions: Challenges and opportunities. *Current Research in Environmental Sustainability*, *3*, 100059. https://doi.org/10.1016/j.crsust.2021.100059

Patz, J. A., Frumkin, H., Holloway, T., Vimont, D. J., & Haines, A. (2014). Climate Change: Challenges and Opportunities for Global Health. *JAMA*, *312*(15), 1565–1580. https://doi.org/10.1001/jama.2014.13186

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peerlings, E., Vranic, S., Ommer, J., Kalas, M., & Steeneveld, G.-J. (2024). Indoor heat in Amsterdam: Comparing observed indoor air temperatures from a professional network and from a citizen science approach. *City and Environment Interactions*, *24*, 100173. https://doi.org/10.1016/j.cacint.2024.100173

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serne, C., Michel, C., & Mathevet, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall—Runoff models. *Hydrological Sciences Journal*, *52*(1), 131–151. https://doi.org/10.1623/hysj.52.1.131

Pesántez, J., Mosquera, G. M., Crespo, P., Breuer, L., & Windhorst, D. (2018). Effect of land cover and hydro-meteorological controls on soil water DOC concentrations in a high-elevation tropical environment. *Hydrological Processes*, *32*(17), 2624–2635. https://doi.org/10.1002/hyp.13224

Peter, M., Diekötter, T., Höffler, T., & Kremer, K. (2021). Biodiversity citizen science: Outcomes for the participating citizens. *People and Nature*, *3*(2), 294–311. https://doi.org/10.1002/pan3.10193

Pfeffer, M. J., Schelhas, J. W., & Day, L. A. (2001). Forest Conservation, Value Conflict, and Interest Formation in a Honduran National Park. *Rural Sociology*, *66*(3), 382–402. https://doi.org/10.1111/j.1549-0831.2001.tb00073.x

Phillips, T. B., Ballard, H. L., Lewenstein, B. V., & Bonney, R. (2019). Engagement in science through citizen science: Moving beyond data collection. *Science Education*, *103*(3), 665–690. https://doi.org/10.1002/sce.21501

Portalanza, D., Morstadt, J. D.-C., Polhmann, V., Gallardo, G., Aguilera, K., Garcia, Y., Rodriguez-Jarama, F., Portalanza, D., Morstadt, J. D.-C., Polhmann, V., Gallardo, G., Aguilera, K., Garcia, Y., & Rodriguez-Jarama, F. (2025). Mapping Soil Erosion and Ecosystem Service Loss: Integrating RUSLE and NDVI Metrics to Support Conservation in El Cajas National Park, Ecuador. *Hydrology*, *12*(11). https://doi.org/10.3390/hydrology12110279

Poveda, G., Waylen, P. R., & Pulwarty, R. S. (2006). Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica. *Palaeogeography, Palaeoclimatology, Palaeoecology, Late Quaternary Climates of Tropical America and Adjacent Seas*, *234*(1), 3–27. https://doi.org/10.1016/j.palaeo.2005.10.031

Prihodko, L., & Goward, S. N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, *60*(3), 335–346. https://doi.org/10.1016/S0034-4257(96)00216-7

Rahbeh, M., Chanasyk, D., & Miller, J. (2011). Two-Way Calibration-Validation of SWAT Model for a Small Prairie Watershed with Short Observed Record. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *36*(3), 247–270. https://doi.org/10.4296/cwrj3603884

Rajagopalan, P., Andamon, M. M., & Paolini, R. (2020). Investigating thermal comfort and energy impact through microclimate monitoring- a citizen science approach. *Energy and Buildings*, *229*, 110526. https://doi.org/10.1016/j.enbuild.2020.110526

Ramírez, S. B., van Meerveld, I., & Seibert, J. (2023). Citizen science approaches for water quality measurements. *Science of The Total Environment*, *897*, 165436. https://doi.org/10.1016/j.scitotenv.2023.165436

Reges, H. W., Doesken, N., Turner, J., Newman, N., Bergantino, A., & Schwalbe, Z. (2016). CoCoRaHS: The Evolution and Accomplishments of a Volunteer Rain Gauge Network. *Bulletin of the American Meteorological Society*, *97*(10), 1831–1846. https://doi.org/10.1175/BAMS-D-14-00213.1

Requier, F., Andersson, G. K., Oddi, F. J., & Garibaldi, L. A. (2020). Citizen science in developing countries: How to improve volunteer participation. *Frontiers in Ecology and the Environment*, *18*(2), 101–108. https://doi.org/10.1002/fee.2150

Rigler, G., Dokou, Z., Khadim, F. K., Sinshaw, B. G., Eshete, D. G., Aseres, M., Amera, W., Zhou, W., Wang, X., Moges, M., Azage, M., Li, B., Holzer, E., Tilahun, S., Bagtzoglou, A., Anagnostou, E., Rigler, G., Dokou, Z., Khadim, F. K., … Anagnostou, E. (2022). Citizen Science and the Sustainable Development Goals: Building Social and Technical Capacity through Data Collection in the Upper Blue Nile Basin, Ethiopia. *Sustainability*, *14*(6). https://doi.org/10.3390/su14063647

Røhr, P. C., & Killingtveit, Å. (2003). Rainfall distribution on the slopes of Mt Kilimanjaro. *Hydrological Sciences Journal*, *48*(1), 65–77. https://doi.org/10.1623/hysj.48.1.65.43483

Romanello, M., Walawender, M., Hsu, S.-C., Moskeland, A., Palmeiro-Silva, Y., Scamman, D., Smallcombe, J. W., Abdullah, S., Ades, M., Al-Maruf, A., Ameli, N., Angelova, D., Ayeb-Karlsson, S., Ballester, J., Basagaña, X., Bechara, H., Beggs, P. J., Cai, W., Campbell-Lendrum, D., … Costello, A. (2025). The 2025 report of the Lancet Countdown on health and climate change. *The Lancet*, *0*(0). https://doi.org/10.1016/S0140-6736(25)01919-1

Rosser, J. F., Leibovici, D., & Jackson, M. (2017). Rapid flood inundation mapping using social media, remote sensing and topographic data. *Natural Hazards*, *87*, 103–120.

Rubio-Iglesias, J. M., Edovald, T., Grew, R., Kark, T., Kideys, A. E., Peltola, T., & Volten, H. (2020). Citizen Science and Environmental Protection Agencies: Engaging Citizens to Address Key Environmental Challenges. *Frontiers in Climate*, *2*. https://doi.org/10.3389/fclim.2020.600998

Ruhi, A., Messager, M. L., & Olden, J. D. (2018). Tracking the pulse of the Earth's fresh waters. *Nature Sustainability*, *1*(4), Article 4. https://doi.org/10.1038/s41893-018-0047-7

Running, S. W., Mu, Q., & Zhao, M. (2021). *MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V061* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD16A2.061

Sachs, J. D., Lafortune, G., & Fuller, G. (2024). *The SDGs and the UN Summit of the Future. Sustainable Development Report 2024*. Dublin: Dublin University Press. https://doi.org/10.25546/108572

Sampson, S. A., Abascal, A., Wang, J., Vanhuysse, S., Rodríguez Carreño, I., Garcia Ruiz, I., & Kuffer, M. (2025). Using Low-Cost Sensors and Citizen Science: Assessing Thermal Inequality in African Slums. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-7–2025*, 229–236. 44th EARSeL Symposium - 26–29 May 2025, Prague, Czech Republic. https://doi.org/10.5194/isprs-archives-XLVIII-M-7-2025-229-2025

Sauermann, H., & Franzoni, C. (2015). Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences*, *112*(3), 679–684. https://doi.org/10.1073/pnas.1408907112

Scheller, M., van Meerveld, I., & Seibert, J. (2024). How well can people observe the flow state of temporary streams? *Frontiers in Environmental Science*, *12*. https://doi.org/10.3389/fenvs.2024.1352697

Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J., & Rango, A. (2002). Remote sensing in hydrology. *Advances in Water Resources*, *25*(8), 1367–1385. https://doi.org/10.1016/S0309-1708(02)00065-9

Scott, A. B., & Frost, P. C. (2017). Monitoring water quality in Toronto's urban stormwater ponds: Assessing participation rates and data quality of water sampling by citizen scientists in the FreshWater Watch. *Science of The Total Environment*, *592*, 738–744. https://doi.org/10.1016/j.scitotenv.2017.01.201

Sebbar, B., Khabba, S., Merlin, O., Simonneaux, V., Hachimi, C. E., Kharrou, M. H., & Chehbouni, A. (2023). Machine-Learning-Based Downscaling of Hourly ERA5-Land Air Temperature over Mountainous Regions. *Atmosphere*, *14*(4), Article 4. https://doi.org/10.3390/atmos14040610

Seibert, J., & Beven, K. J. (2009). Gauging the ungauged basin: How many discharge measurements are needed? *Hydrology and Earth System Sciences*, *13*(6), 883–892. https://doi.org/10.5194/hess-13-883-2009

Seibert, J., Blanco, S., Scheller, M., Schwarzenbach, F., Ze, W., & van Meerveld, I. (2022). *Engaging the public for water data collection—Experiences from the CrowdWater project*. EGU22-6319. https://doi.org/10.5194/egusphere-egu22-6319

Seibert, J., & McDonnell, J. J. (2015). Gauging the Ungauged Basin: Relative Value of Soft and Hard Data. *Journal of Hydrologic Engineering*, *20*(1), A4014004. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000861

Seibert, J., van Meerveld, H. J., Etter, S., Strobl, B., Assendelft, R., & Hummer, P. (2019). Wasserdaten sammeln mit dem Smartphone – Wie können Menschen messen, was hydrologische Modelle brauchen? *Hydrologie und Wasserbewirtschaftung*, *63*(2), Article 2. https://doi.org/10.5675/HyWa_2019.2_1

Seibert, J., & Vis, M. J. P. (2016). How informative are stream level observations in different geographic regions? *Hydrological Processes*, *30*(14), 2498–2508. https://doi.org/10.1002/hyp.10887

Sen Roy, S. (2018). Climate Change in the Global South: Trends and Spatial Patterns. In S. Sen Roy (Ed.), *Linking Gender to Climate Change Impacts in the Global South* (pp. 1–25). Springer International Publishing. https://doi.org/10.1007/978-3-319-75777-3_1

Senay, G. B., Leake, S., Nagler, P. L., Artan, G., Dickinson, J., Cordova, J. T., & Glenn, E. P. (2011). Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. *Hydrological Processes*, *25*(26), 4037–4049. https://doi.org/10.1002/hyp.8379

Sene, K. (2024). *Hydrometeorology: Forecasting and Applications*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-58269-1

Shagega, F. P., Codalli, F., Jacobs, S., Munishi, S. E., Windhorst, D., & Breuer, L. (2025). Quantifying preferential flow occurrence in dependence of land cover on the southern slopes of Mount Kilimanjaro, Tanzania. *Journal of Hydrology: Regional Studies*, *58*, 102215. https://doi.org/10.1016/j.ejrh.2025.102215

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, *520*, 1–16. https://doi.org/10.1016/j.jhydrol.2014.11.040

Shinbrot, X. A., Muñoz-Villers, L., Mayer, A., López-Portillo, M., Jones, K., López-Ramírez, S., Alcocer-Lezama, C., Ramos-Escobedo, M., & Manson, R. (2020a). Quiahua, the First Citizen Science Rainfall Monitoring Network in Mexico: Filling Critical Gaps in Rainfall Data for Evaluating a Payment for Hydrologic Services Program. *Citizen Science: Theory and Practice*, *5*(1), Article 1. (12 volunteers perceptions of motivations and barriers to participation; 621 days of rainfall data from citizen scientist volunteer and automatic tipping gauge data). https://doi.org/10.5334/cstp.316

Shinbrot, X. A., Muñoz-Villers, L., Mayer, A., López-Portillo, M., Jones, K., López-Ramírez, S., Alcocer-Lezama, C., Ramos-Escobedo, M., & Manson, R. (2020b). Quiahua, the First Citizen Science Rainfall Monitoring Network in Mexico: Filling Critical Gaps in Rainfall Data for Evaluating a Payment for Hydrologic Services Program. *Citizen Science: Theory and Practice*, *5*(1). https://doi.org/10.5334/cstp.316

Silvestro, F., Gabellani, S., Rudari, R., Delogu, F., Laiolo, P., & Boni, G. (2015). Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrology and Earth System Sciences*, *19*(4), 1727–1751. https://doi.org/10.5194/hess-19-1727-2015

Singh, S. K., & Bárdossy, A. (2012). Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources*, *38*, 81–91. https://doi.org/10.1016/j.advwatres.2011.12.006

Singh, S. K., Liang, J., & Bárdossy, A. (2012). Improving the calibration strategy of the physically-based model WaSiM-ETH using critical events. *Hydrological Sciences Journal*, *57*(8), 1487–1505. https://doi.org/10.1080/02626667.2012.727091

Southworth, J., Nagendra, H., Carlson, L. A., & Tucker, C. (2004). Assessing the impact of Celaque National Park on forest fragmentation in western Honduras. *Applied Geography*, *24*(4), 303–322. https://doi.org/10.1016/j.apgeog.2004.07.003

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Srisathan, W. A., Malai, K., Narathawaranan, N., Coochampoo, K., & Naruetharadhol, P. (2024). The impact of citizen science on environmental attitudes, environmental knowledge, environmental awareness to pro–environmental citizenship behaviour. *International Journal of Sustainable Engineering*, *17*(1), 360–378. https://doi.org/10.1080/19397038.2024.2354269

Starkey, E., Parkin, G., Birkinshaw, S., Large, A., Quinn, P., & Gibson, C. (2017). Demonstrating the value of community-based (‘citizen science’) observations for catchment modelling and characterisation. *Journal of Hydrology*, *548*, 801–817. https://doi.org/10.1016/j.jhydrol.2017.03.019

Strobl, B., Etter, S., Meerveld, I. van, & Seibert, J. (2019). The CrowdWater game: A playful way to improve the accuracy of crowdsourced water level class data. *PLOS ONE*, *14*(9), e0222579. https://doi.org/10.1371/journal.pone.0222579

Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)] (pp. 1–34). (2023). IPCC. doi: 10.59327/IPCC/AR6-9789291691647.001

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Sun, Y.-J., Wang, J.-F., Zhang, R.-H., Gillies, R. R., Xue, Y., & Bo, Y.-C. (2005). Air temperature retrieval from remote sensing data based on thermodynamics. *Theoretical and Applied Climatology*, *80*(1), 37–48. https://doi.org/10.1007/s00704-004-0079-y

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., & Hong, Y. (2020). Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sensing of Environment*, *240*, 111697. https://doi.org/10.1016/j.rse.2020.111697

Taylor, J., Viitanen, A.-K., & Espinosa, A. (2024, September 20). *Community-based Monitoring of Urban Environmental Data*. Buildings and Cities. https://www.buildingsandcities.org/insights/commentaries/community-based-monitoring-urban-environmental.html

Thatoe Nwe Win, T., Bogaard, T., & van de Giesen, N. (2019). A Low-Cost Water Quality Monitoring System for the Ayeyarwady River in Myanmar Using a Participatory Approach. *Water*, *11*(10), 1984. https://doi.org/10.3390/w11101984

Tran, H. N., Duwal, S., Prajapati, R., Davids, J. C., Nguyen, D. T., & Miegel, K. (2025). Citizen science for monitoring rainfall in Vietnam. *E3S Web of Conferences*, *608*, 05014. https://doi.org/10.1051/e3sconf/202560805014

United Nations. (2015). *Resolution adopted by the General Assembly on 25 September 2015: Transforming our world: The 2030 Agenda for Sustainable Development*. https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf

United Nations. (2022). *UNCTAD Handbook of Statistics 2022*. https://www.un-ilibrary.org/content/books/9789210021784

United Nations Department of Economic and Social Affairs. (2024). *The Sustainable Development Goals Report 2024*. United Nations. https://doi.org/10.18356/9789213589755

Uprety, M., Ochoa-Tocachi, B. F., Paul, J. D., Regmi, S., & Buytaert, W. (2019). Improving water resources management using participatory monitoring in a remote mountainous region of Nepal. *Journal of Hydrology: Regional Studies*, *23*, 100604. https://doi.org/10.1016/j.ejrh.2019.100604

Valdez, M. C., Chang, K.-T., Chen, C.-F., Chiang, S.-H., & Santos, J. L. (2017). Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. *Geomatics, Natural Hazards and Risk*, *8*(2), 876–892. https://doi.org/10.1080/19475705.2016.1278404

Valipour, M., Bateni, S. M., Dalezios, N. R., Almazroui, M., Heggy, E., Şen, Z., Angelakis, A. N., Valipour, M., Bateni, S. M., Dalezios, N. R., Almazroui, M., Heggy, E., Şen, Z., & Angelakis, A. N. (2020). Hydrometeorology: Review of Past, Present and Future Observation Methods. In *Hydrology*. IntechOpen. https://doi.org/10.5772/intechopen.94939

van Meerveld, H. J. I., Vis, M. J. P., & Seibert, J. (2017). Information content of stream level class data for hydrological model calibration. *Hydrology and Earth System Sciences*, *21*(9), 4895–4905. https://doi.org/10.5194/hess-21-4895-2017

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., & Verdin, J. P. (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. https://doi.org/10.1016/j.rse.2013.07.013

Villaseñor, E., Porter-Bolland, L., Escobar, F., Guariguata, M. R., & Moreno-Casasola, P. (2016). Characteristics of participatory monitoring projects and their relationship to decision-making in biological resource management: A review. *Biodiversity and Conservation*, *25*(11), 2001–2019. https://doi.org/10.1007/s10531-016-1184-9

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, *5*(1), 13–26. https://doi.org/10.5194/hess-5-13-2001

Walker, D., Forsythe, N., Parkin, G., & Gowing, J. (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. https://doi.org/10.1016/j.jhydrol.2016.04.062

Weeser, B., Jacobs, S., Kraft, P., Rufino, M. C., & Breuer, L. (2019). Rainfall-Runoff Modeling Using Crowdsourced Water Level Data. *Water Resources Research*, *55*(12), 10856–10871. https://doi.org/10.1029/2019WR025248

Weeser, B., Stenfert Kroese, J., Jacobs, S. R., Njue, N., Kemboi, Z., Ran, A., Rufino, M. C., & Breuer, L. (2018). Citizen science pioneers in Kenya – A crowdsourced approach for hydrological monitoring. *Science of The Total Environment*, *631–632*, 1590–1599. https://doi.org/10.1016/j.scitotenv.2018.03.130

Weyhenmeyer, G. A., Mackay, M., Stockwell, J. D., Thiery, W., Grossart, H.-P., Augusto-Silva, P. B., Baulch, H. M., de Eyto, E., Hejzlar, J., Kangur, K., Kirillin, G., Pierson, D. C., Rusak, J. A., Sadro, S., & Woolway, R. I. (2017). Citizen science shows systematic changes in the temperature difference between air and inland waters with global warming. *Scientific Reports*, *7*(1), 43890. https://doi.org/10.1038/srep43890

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

Willett, D. S., Brannock, J., Dissen, J., Keown, P., Szura, K., Brown, O. B., & Simonson, A. (2023). NOAA Open Data Dissemination: Petabyte-scale Earth system data in the cloud. *Science Advances*, *9*(38), eadh0032. https://doi.org/10.1126/sciadv.adh0032

Wilson, C., Mills, S., & Wood, P. J. (2025). Children's environmental citizen science: Stakeholders' experiences and perceptions. *People and Nature*, *7*(2), 493–503. https://doi.org/10.1002/pan3.10773

WMO. (2017). *Challenges in the Transition from Conventional to Automatic Meteorological Observing Networks for Long-term Climate Records* (Vol. 12). https://library.wmo.int/idurl/4/55812

WMO,. (2025a). *WMO confirms 2024 as warmest year on record at about 1.55°C above pre-industrial level* [Press release]. Press Release. https://wmo.int/news/media-centre/wmo-confirms-2024-warmest-year-record-about-155degc-above-pre-industrial-level

WMO. (2025b). *World Weather Information Service*. World Weather Information Service. https://worldweather.wmo.int/en/city.html?cityId=2047

WMO. (2025c). *World Weather Information Service—Climate Report for Cuenca, Ecuador for the period 1981—2010*. World Weather Information Service. https://worldweather.wmo.int/en/city.html?cityId=2047

World Bank. (2025a). *EcuadorECU - Country Overview | Climate Change Knowledge Portal—Province Azuay*. https://climateknowledgeportal.worldbank.org/country/ecuador

World Bank. (2025b). *Kenya—Country Overview | Climate Change Knowledge Portal*. https://climateknowledgeportal.worldbank.org/country/kenya

World Bank. (2025c). *World Bank Climate Change Knowledge Portal*. https://climateknowledgeportal.worldbank.org/

Xu, L., Du, J., Ren, J., Hu, Q., Qin, F., Mu, W., & Hu, J. (2024). Improving the ERA5-Land Temperature Product through a Deep Spatiotemporal Model That Uses Fused

Multi-Source Remote Sensing Data. *Remote Sensing*, *16*(18), 3510. https://doi.org/10.3390/rs16183510

Xu, X., Li, J., & Tolson, B. A. (2014). Progress in integrating remote sensing data and hydrologic modeling. *Progress in Physical Geography*, *38*(4), 464–498.

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

Zafeirelli, S., & Kavroudakis, D. (2024). Comparison of outlier detection approaches in a Smart Cities sensor data context. *International Journal on Smart Sensing and Intelligent Systems*, *17*(1), 20240004. https://doi.org/10.2478/ijssis-2024-0004

Zech, M. (2006). Evidence for Late Pleistocene climate changes from buried soils on the southern slopes of Mt. Kilimanjaro, Tanzania. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *242*(3), 303–312. https://doi.org/10.1016/j.palaeo.2006.06.008

Zhan, C., Jiang, W., Zheng, Y., Lu, J., & Zhang, Q. (2023). A data-driven study of active meteorological stations and the factors motivating their establishment. *Sustainable Energy Technologies and Assessments*, *57*, 103147. https://doi.org/10.1016/j.seta.2023.103147

Zhang, Y., Li, J., & Liu, D. (2024). Spatial Downscaling of ERA5 Reanalysis Air Temperature Data Based on Stacking Ensemble Learning. *Sustainability*, *16*(5), 1934. https://doi.org/10.3390/su16051934

Zhu, J., Zhu, S., Yu, F., Zhang, G., & Xu, Y. (2021). A downscaling method for ERA5 reanalysis land surface temperature over urban and mountain areas. *National Remote Sensing Bulletin*, *25*(8), 1778–1791. https://doi.org/10.11834/jrs.20211257

Acknowledgements

Inspired by many other acknowledgements I have read, I tried to find the correct words for this last part of my dissertation. A successful doctorate depends on the people who have accompanied along the way - mentors and teachers, students and technicians, friends and colleagues.

First and foremost, I would like to thank my first supervisor and "Doktormutter", Suzanne, for patiently reviewing what felt like an endless string of drafts, for always listening when I encountered difficulties, for her encouraging words throughout my PhD, and for encouraging me to apply for funding for my research stay in New Zealand. I am also deeply grateful to Björn and Lutz for their relentless, constructive feedback and ideas, which continually sharpened my manuscripts and helped me distill and expand my own ideas. Their support was invaluable to my progress. Even though the three of you didn't always agree on the feedback you gave me on whatever it was, you helped me a lot to develop further. Special thanks go to the HydroCrowd colleagues, especially Jazmin, Fabia, Jefferson, and Frank, whose help and collaboration was essential to this work. I would also like to acknowledge the *Kurt-Eberhard-Bode-Stiftung für medizinische und naturwissenschaftliche Forschung* for financing our project and making this entire journey possible.

I also appreciate the people I shared my office (Room 113) in the Zeughaus with, as well as those who would pop up from behind the wall for a quick chat. My sincere thanks also extend to the ZEU team - Matthias, Petra, Iuliia, and, formerly, Sabine - for their administrative assistance and for being available for a conversation when no one else was around. I am also grateful to the International Giessen Graduate Centre for the Life Sciences (GGL) for its support within the structured doctoral program at JLU. A heartfelt thank you also goes to Thomas and Elizabeth in Auckland, who welcomed me so warmly and hosted me for several weeks in their working group. The time spent there remains one of the highlights of my academic journey. I also wish to thank all the wonderful people I met during fieldwork in Honduras and Ecuador, whose help in various forms made the project a reality.

I also would like to say thank you to the ArbeiterKind.de community, a big mutually supportive crowd of people for all those who are from a non-academic family background. Thanks to my local group in Gießen for reminding me that all of this, studying, graduating, and of course doing a PhD cannot be taken for granted by people like us.

Thanks also to my dear friend Hanna for being my study buddy since we did our master's program together. I hope that in a few months you will also have your PhD certificate in your hands. I also thank Tobi, who I studied with during my Bachelor's program and who became also a great study buddy, even though he completed the same master's degree somewhere else. Thank you both very much for proofreading this document.

At this point, I would like to express my deepest thanks to my family, especially my parents Karin and Karl-Heinz for always welcoming me warmly and believing in me, even though, what I was doing sometimes seemed very far away. Thanks to my three older siblings Marius, Sarah and Matthias for being people I could look up to. Also, thanks to all my close friends for their support and all the nice evenings together. Many of you have no academic background and I guess it was sometimes hard to grasp what I was doing, but nevertheless you always asked me questions in a motivated and interested manner. Thanks for that.

Last, but by no means least, Zoe. Thank you, for everything and for reminding me of the real priorities in life whenever needed. Thanks also to Link, who has been an excellent companion, enriching my life since the final phase of my PhD and reminding me to get up from my desk and go outside for some fresh air. Thank you both for giving me so much extra joy in life.

I hope no one reads this list and is sad they did not make it by name. I assure you this is a mistake on my part and I am grateful for you.

Formal 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 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.

……………………………………..

Gießen, 30th of January 2026