Towards a historical precipitation database for West Africa: overview, quality control and harmonization

Jan Bliefernicht, Seyni Salack, Moussa Waongo, Thompson Annor, Patrick Laux, Harald Kunstmann

Angaben zur Veröffentlichung / Publication details:

Bliefernicht, Jan, Seyni Salack, Moussa Waongo, Thompson Annor, Patrick Laux, and Harald Kunstmann. 2022. "Towards a historical precipitation database for West Africa: overview, quality control and harmonization." *International Journal of Climatology* 42 (7): 4001–23. https://doi.org/10.1002/joc.7467.

Received: 5 May 2021 | Revised: 13 October 2021 | Accepted: 31 October 2021 | Published on: 29 December 2021

DOI: 10.1002/joc.7467

RESEARCH ARTICLE

International Journal of Climatology RMetS

Towards a historical precipitation database for West Africa: Overview, quality control and harmonization

Jan Bliefernicht[1] | Seyni Salack[2] | Moussa Waongo[3] | Thompson Annor[4] | Patrick Laux[1,5] | Harald Kunstmann[1,5]

[1]Institute of Geography, University of Augsburg, Augsburg, Germany

[2]West African Climate Service Centre on Climate Change and Adapted Land Use, Competence Center, Ouagadougou, Burkina Faso

[3]Training and Research Department, AGRHYMET Regional Centre, Niamey, Niger

[4]Department of Physics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

[5]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence
Harald Kunstmann, Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany.
Email: harald.kunstmann@kit.edu

Funding information
Federal Ministry of Education and Research, Grant/Award Numbers: 01LG1202C1, 01LG1202C; German Research Foundation, Grant/Award Number: KU/2090 14-1

Abstract
Reliable long-term observations from precipitation stations are often required for climatological studies but are strongly limited in many regions of the world. To improve this limitation for West Africa, we compiled daily and monthly observations from more than 20 national, continental and global databases, to establish a historical precipitation archive with a focus on four countries (Burkina Faso, Ghana, Benin and Togo). The new archive contains long-term daily and monthly precipitation measurements from 1819 to 2013 for more than 1,000 sites. It is, therefore, the most comprehensive historical dataset with daily and monthly precipitation observations for this region. To produce a quality-controlled and harmonized precipitation dataset for the focal region, various statistical algorithms have been implemented. These algorithms rely on straightforward geostatistical approaches (e.g., spatial correlograms) and corresponding statistical tests for identification and elimination of unreliable time series, in addition to various standard approaches used by global data centers. Although the quality control revealed various data errors and uncertainties for measurements and meta-information (e.g., unit conversion errors, temporal offsets, frequent and long data gaps), a spatial interpolation using the quality-controlled and harmonized dataset produced relatively reliable precipitation patterns for different target variables (e.g., monthly precipitation amount and daily precipitation probability). A major remaining challenge is providing free access to this database for research and other noncommercial purposes, due to national data protection regulations. However, several further tasks have been initiated and implemented (e.g., free provision of gridded precipitation datasets and point statistics) to improve the access and availability of station-based precipitation observations and related data products for this climatologically challenging region.

KEYWORDS
climate observations, geostatistics, precipitation, quality control, West Africa

1 | INTRODUCTION

Reliable long-term observations from precipitation stations are the basis for many investigations in climatology and related disciplines. This information is required for the analysis of climate variability and change (Trenberth, 2011; O'Gorman, 2015), for application of hydrological models and evaluation of climate models (Dai, 2006; Tapiador *et al.*, 2017), for ground-truthing of satellite and radar precipitation products (Goudenhoofdt and Delobbe, 2009; Huffman *et al.*, 2009), and for many other climate and weather services. Although more than 100,000 precipitation stations are operated by national hydrological and meteorological services (NHMS) and other institutions worldwide, long-term observations from precipitation gauges on the ground are often difficult to obtain from NHMS. Many climatological studies and related disciplines therefore use information from station-based global and continental climate databases and their associated gridded precipitation products in data-scarce regions (e.g., Moron *et al.*, 2010; Ceccherini *et al.*, 2017; Wan *et al.*, 2021). The information in these archives is usually free-of-charge for noncommercial purposes and can be easily accessed via FTP servers or comfortable web interfaces.

Two of the most commonly used station-based global databases are provided by the Global Historical Climate Network (GHCN, Menne *et al.*, 2012) and the Global Surface Summary of Day (GSOD, NOAA, 2020a) of NOAA. They have been used in many regional studies around the world (Moron *et al.*, 2009; Moron *et al.*, 2010; Ermert *et al.*, 2011; Ceccherini *et al.*, 2017; Zhang *et al.*, 2019). These archives are also a crucial source to generate a number of gridded products used for climate analysis and monitoring (Yatagai *et al.*, 2009; Wong *et al.*, 2011; Becker *et al.*, 2013; Schamm *et al.*, 2014), and to evaluate climate model outputs and seasonal forecast products (Diallo *et al.*, 2012; Nikulin *et al.*, 2012; Siegmund *et al.*, 2015; Heinzeller *et al.*, 2018; Bliefernicht *et al.*, 2019). Although GHCN and GSOD databases contain a huge amount of daily and monthly precipitation records collected worldwide, most stations are located in North America and Europe. Lorenz and Kunstmann (2012) showed that the number of measurements in different global climatological products decreases sharply for Africa. An example of overcoming data availability problems for this region is given in Barry *et al.* (2018) for the use of climate extremes and indices.

To provide reliable precipitation observations for subsequent analysis in climatology and other disciplines, quality assurance techniques must be applied to measurements and metadata (Aguilar *et al.*, 2003). Many different algorithms are used to identify unreliable measurements of climatological variables and for the homogenization of long-term time series. Overviews are provided by Peterson *et al.* (1998a), Aguilar *et al.* (2003), Costa and Soares (2009) and Ribeiro *et al.* (2016). Recent examples are shown by Delvaux *et al.* (2019), Skrynyk *et al.* (2019) and Coll *et al.* (2020) for different regions of the world. However, many of these approaches were primarily developed for annual or monthly time series of climatological variables, often with an emphasis on temperature measurements due to the importance of this variable for climate change studies. Vicente-Serrano *et al.* (2010) and Costa and Soares (2009) noted that there are only a few algorithms for automatic quality assessment of sub-monthly precipitation information. Examples are the quality control algorithms for daily precipitation (and other meteorological variables) used by GHCN and GSOD (Smith *et al.*, 2011, Durre *et al.*, 2010 and Lott, 2004) and MeteoSwiss (Scherrer *et al.*, 2011). Further approaches for daily precipitation are presented by Feng *et al.* (2004), Vicente-Serrano *et al.* (2010) and Boulanger *et al.* (2010).

The objective of this work is to introduce a novel station-based precipitation database for West Africa, named West African Historical Precipitation Database (WAHPD). The database consists of daily and monthly observations from precipitation stations with first records starting in 1819. The dataset was collected from global, continental and national data archives as part of the WASCAL (West African Science Services Centre on Climate Change and Adapted Land Use) Observation Network (WASCAL ON, Salack *et al.*, 2019) with a specific focus on four West African countries (Burkina Faso, Ghana, Togo and Benin) to support novel meso-scale observation networks (Bliefernicht *et al.*, 2018; Salack *et al.*, 2019) in this region with climate data. WASCAL ON is a joint collaboration between the West African NHMS, WASCAL and partner institutions to strengthen the observational infrastructure of the West African NHMS and to improve the availability of hydro-meteorological observations for this region (Salack *et al.*, 2019). Subsets of the WAHPD and related data products were already used by various investigations, for example, for evaluation of regional climate model simulations (Dieng *et al.*, 2017) and seasonal forecasts products (Bliefernicht *et al.*, 2019), analysis of precipitation extremes and comparison with satellite products (Engel *et al.*, 2017), as input information for groundwater reconstruction (Ascott *et al.*, 2020) and as climate background data of the WASCAL hydro-meteorological observatories (Bliefernicht *et al.*, 2018; Salack *et al.*, 2018b; Berger *et al.*, 2019).

In addition, an overview of statistical algorithms used for quality control and harmonization of the different data archives is given, in order to produce a quality-controlled and joint precipitation database. This work is carried out for the focal region (Burkina Faso, Ghana, Benin and Togo) with a specific focus on the daily database. In addition to other quality algorithms (such as Durre *et al.*, 2010), the presented algorithms rely on geostatistical approaches, such as spatial correlograms and statistical tests for the identification of unreliable time series. Geostatistical approaches are seldom used for quality assurance (Costa and Soares, 2009), but have the advantage that measurements and important meta-information (e.g., station coordinates) can be treated together. One of the first approaches was presented by Costa and Soares (2009). Another approach was presented by Ribeiro *et al.* (2017), which was applied to monthly precipitation.

The manuscript has the following structure. In Section 2, a brief description of the status quo of climatological observations and several reasons for the limited data availability in West Africa are presented. A detailed description of the precipitation database can be found in Section 3. An overview of the statistical algorithms used for quality control and the detected data limitations are shown in Section 4. Section 5 gives an overview of the harmonization procedure used for merging the different data sources and shows an application of the joint database. The outcomes of the study are discussed in Section 6. Section 7 summarizes the main findings of this work and ends with a conclusion.

2 | AVAILABILITY AND ACCESS TO CLIMATOLOGICAL OBSERVATIONS IN WEST AFRICA

Climatological observations are often not readily available for many subregions in West Africa. Figure 1 shows this problem for monthly precipitation for two different years using the Global Precipitation Climatology Centre (GPCC) gridded data set (Becker *et al.*, 2013). This data set is one of the world's most important precipitation products used for validation of satellite precipitation products and regional climate models in, for instance, West Africa (Nicholson *et al.*, 2003; Panitz *et al.*, 2014; Dosio *et al.*, 2015; Annor *et al.*, 2018). In August 1982, approximately 960 measurements were used for interpolation. However, 30 years later, in August 2012, the interpolated precipitation field was only based on 134 measurements (86% fewer measurements). Moreover, the network of precipitation stations for 2012 shows large spatial data gaps for several countries in the southwest of West Africa (Sierra Leone, Guinea and Liberia) and in Nigeria. In addition, the low measurement density is not only a problem for a specific month or year. The number of measurements peaked in the late 1980s and has steadily declined since then (as shown in Figure S1 of the supporting information). In 2013, an average of 150 measurements was used for the interpolation of the GPCC product, comparable with the situation in the 1920s.

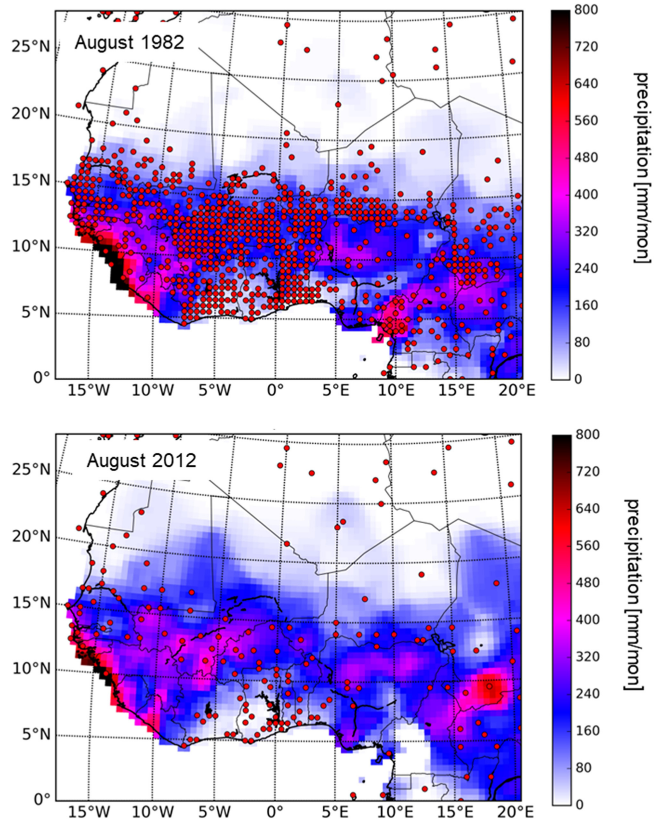

FIGURE 1 Monthly precipitation amounts for August 1982 and 2012. The red dots indicate grid points at which measurements from precipitation stations were available for interpolation. Thus, the figure also indicates the spatial coverage of the measurements used for interpolation. GPCC reanalysis version 7, 0.5° resolution [Colour figure can be viewed at wileyonlinelibrary.com]

The most important reason for the limited availability of climatological observations in West Africa is the weak observational infrastructure of NHMS (Galle *et al.*, 2018). For instance, the network density for six West African countries was 7–50 times smaller in comparison to the rainfall observation network operated by the German weather service in 2010 (Table 1). In the case of Togo, more than 75% (50 out of 210) of the precipitation stations did not work. In a recent evaluation of the West African observation networks, Salack *et al.* (2019) argued that many stations do not transfer information to the global transmission system because they are located in conflict zones. Other stations have neither measured nor reported, and in some cases, the stations were simply abandoned due to the retirement or death of voluntary observers. Moreover, due to the limited financial resources of many NHMS, modernization and

Characteristics	GH	BF	CD	BN	ML	TO
Country area (10^3 km^2)	238	274	322	112	1240	57
Synoptic stations (–)	22	10	14	6	19	9
Climatological stations (–)	61	9 (12)	5 (6)	0	0	2 (19)
Rainfall stations (–)	156	130	41 (186)	59	200	50 (210)
Agrometeorological stations (–)	54	14 (20)	14 (26)	29	50	0
Network density (1/10^3 km^2)	0.65	0.47	0.13	0.52	0.16	0.88
Density index (–)	9.6	13.3	49.5	12.0	39.0	7.1

Note: The information is based on a questionnaire made within the WASCAL (West African Science Service Centre on Climate Change and Adapted Land Use) programme. The network density (ND) is based on the number of functional rainfall stations. The total number of rainfall stations is given in the brackets. The density index compares the network density of West African NHMS with the network density of the rainfall network operated by the *Deutscher Wetterdienst* (DWD) in Germany. A density index of 10 indicates that the NHMS needs 10 times more stations to have the same nationwide coverage as the DWD in Germany.

TABLE 1 The status quo of the meteorological networks in Ghana (GH), Burkina Faso (BF), Côte d'Ivoire (CD), Benin (BN), Mali (ML) and Togo (TO) in 2010

densification of hydrometeorological networks seems like a never-ending challenge.

Another important problem for the limited data availability is the national data regulation, which NHMS underlie. Hence, climate observations from West African NHMS are usually not free of charge for research, education or other noncommercial use, unlike other countries like the United States or Germany. In some cases, data can only be shared if there are bilateral agreements between NHMS and the data acquiring institution. However, even in a comfortable situation of a cooperation agreement, it can be difficult to obtain data due to slow and complicated bureaucracy (Salack *et al*., 2019). In some instances, data requests can incur a high fee, even for research or educational purposes. Moreover, accessing data can be very difficult due to specific technical limitations. Measurements may only be available on paper, as they are not yet digitized. Other data limitations are nonstandardized data formats, frequent and long data gaps, and less quality-controlled data, which are addressed in the next sections in more detail.

3 | OVERVIEW OF THE WEST AFRICAN HISTORICAL PRECIPITATION DATABASE

The WAHPD consists of daily (WAHPD-D) and monthly measurements (WAHPD-M, Figure 2). The backbone of the WAHPD-D are 13 national datasets with long-term measurements for a current period (NHMS) and a historical period (AMMA-P), compiled from NHMS and the African Monsoon Multidisciplinary Analysis program (AMMA; Redelsperger *et al*., 2006; Lebel *et al*., 2010). In addition, two continental (regional) and global archives (GHCN-D and GSOD) are used. The continental data were compiled from a second data archive of the AMMA project (AMMA-S) and the GLOWA initiative (Van De Giesen *et al*., 2002). WAHPD-M is based on monthly values computed from WAHPD-D, complemented by four continental (regional) and global databases with monthly records.

3.1 | Daily database

The NHMS database consists of four national datasets (Burkina Faso, Ghana, Benin and Senegal, Figure 3) with measurements ranging from 1960 to 2010 and good data availability between 1970 and 2010 of approximately 85% (Figure 4). The biggest dataset of the NHMS database is the subset of Burkina Faso (NHMS-BF), with 142 precipitation time series. Unlike the other three NHMS subsets, this dataset contains information from synoptic (10), climatic (10), agrometeorological (17) and standalone rain gauges (105). The Senegalese subset (NHMS-SN) is the most complete dataset of the NHMS-P database with almost 100% data availability for 20 stations over 50 years. In addition, 22 precipitation time series from rainfall gauges in Ghana (NHMS-GH) and 24 from Benin (NHMS-BN) are used.

The AMMA-P database is a big archive of historical daily precipitation measurements from before 1981, which consists of 1,058 time series from locations in nine West African countries (Figures 3 and 4). The data were extracted from the Pluvio dataset of the AMMA database (Fleury *et al*., 2011). First daily measurements are already available from 1891, but more than 90% of the precipitation sites did not contain any information before 1940 (Figure 4). Although this dataset provides a large amount of data, the Pluvio datasets from the AMMA database

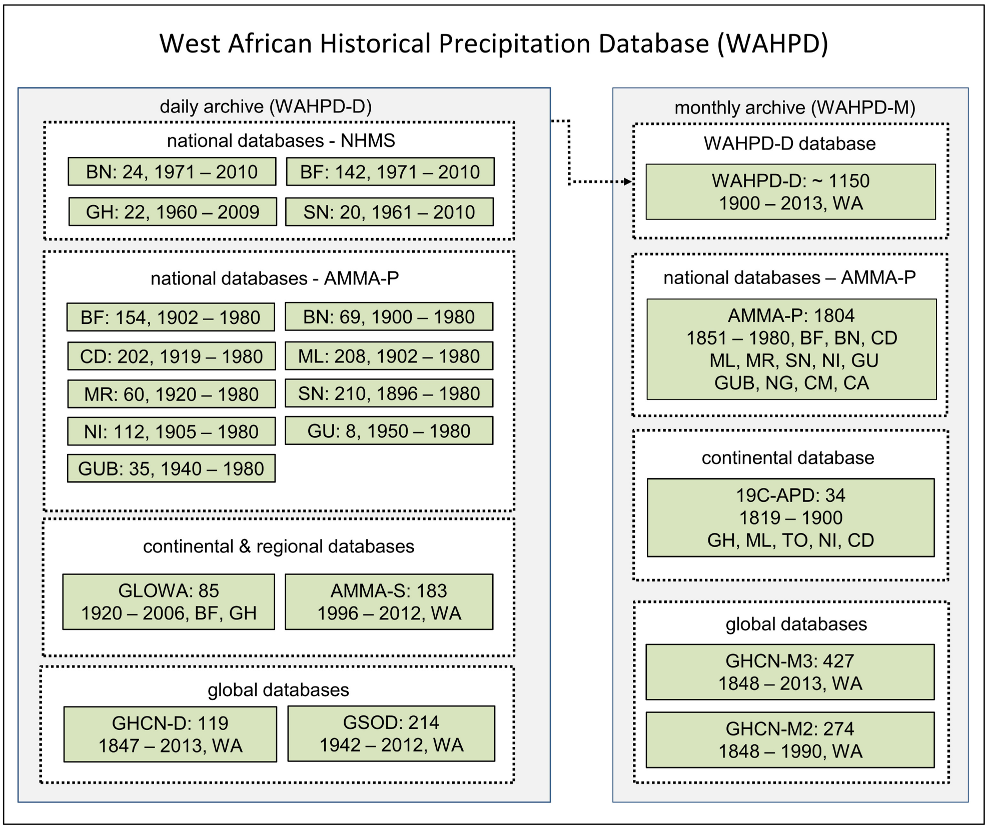

FIGURE 2 Overview of the West African Historical Precipitation Database (WAHPD) with its daily (WAHPD-D) and monthly archive (WAHPD-M) and the data subsets used for the compilation. In addition, the country code, the number of time series and the data period are shown for each subset. NHMS = national hydrological and meteorological services, AMMA-P = precipitation database of the AMMA programme, GLOWA = precipitation database of the GLOWA-Volta project, AMMA-S = precipitation measurements from the synoptical database of the AMMA programme, GSOD = global summary of surface day, GHCN-D = daily database of global historical climate network, GHCN-M = monthly databases of the global historical climatology network, D = daily, M = monthly, 19C-APD = 19th Century African Instrumental and Documentary Precipitation Data, BN = Benin, BF = Burkina Faso, CD = Côte d'Ivoire, GH = Ghana, ML = Mali, MR = Mauretania, SN = Senegal, NI = Niger, GU = Guinea, GUB = Guinea-Bissau, TO = Togo, NG = Nigeria, CM = Cameroon, CA = Chad, WA = West Africa [Colour figure can be viewed at wileyonlinelibrary.com]

seem to be rarely used. To our knowledge, there is no study that refers to these datasets.

The GLOWA subset consists of 85 time series of precipitation measurements, mainly from 1950 to 2006, with 42 in Burkina Faso and 43 in Ghana. This dataset is a good addition for Ghana since the NHMS database contains only 22 precipitation sites for this country and the AMMA-P database has no site. The origin of the GLOWA dataset is the geoportal of the Volta basin authority, established as part of the GLOWA-Volta project. The data were originally obtained from the NHMS in Ghana and Burkina Faso and digitized from weather reports. Precipitation subsets of the GLOWA database were already used by Neumann *et al.* (2007), Jung and Kunstmann (2007) and Laux *et al.* (2008).

The AMMA-S subset contains measurements for all West African countries. The data is available for 183 sites from 1996 to 2012 (Figures 3 and 4). It was extracted from an archive of daily and sub-daily measurements from synoptic stations provided by the AMMA database. Since the data extraction was done for a rectangular domain that covers all West African countries, the AMMA-S also contains time series from countries bordering West Africa such as Cameroon (Figure 3).

The GHCN-D dataset extracted for the study region consists of 119 stations, 88 of which are located in 10 West African countries (Figure 3). The dataset was taken from the GHCN daily database (NOAA, 2020b). Figure 4 shows that most of the GHCN-D measurements are available from 1945 to 1980, and that the data coverage has been relatively poor (approximately 20 measurements) since the 1980s. In addition to the GHCN-D, 214 time series of the GSOD database are used. GSOD is a global archive of daily weather

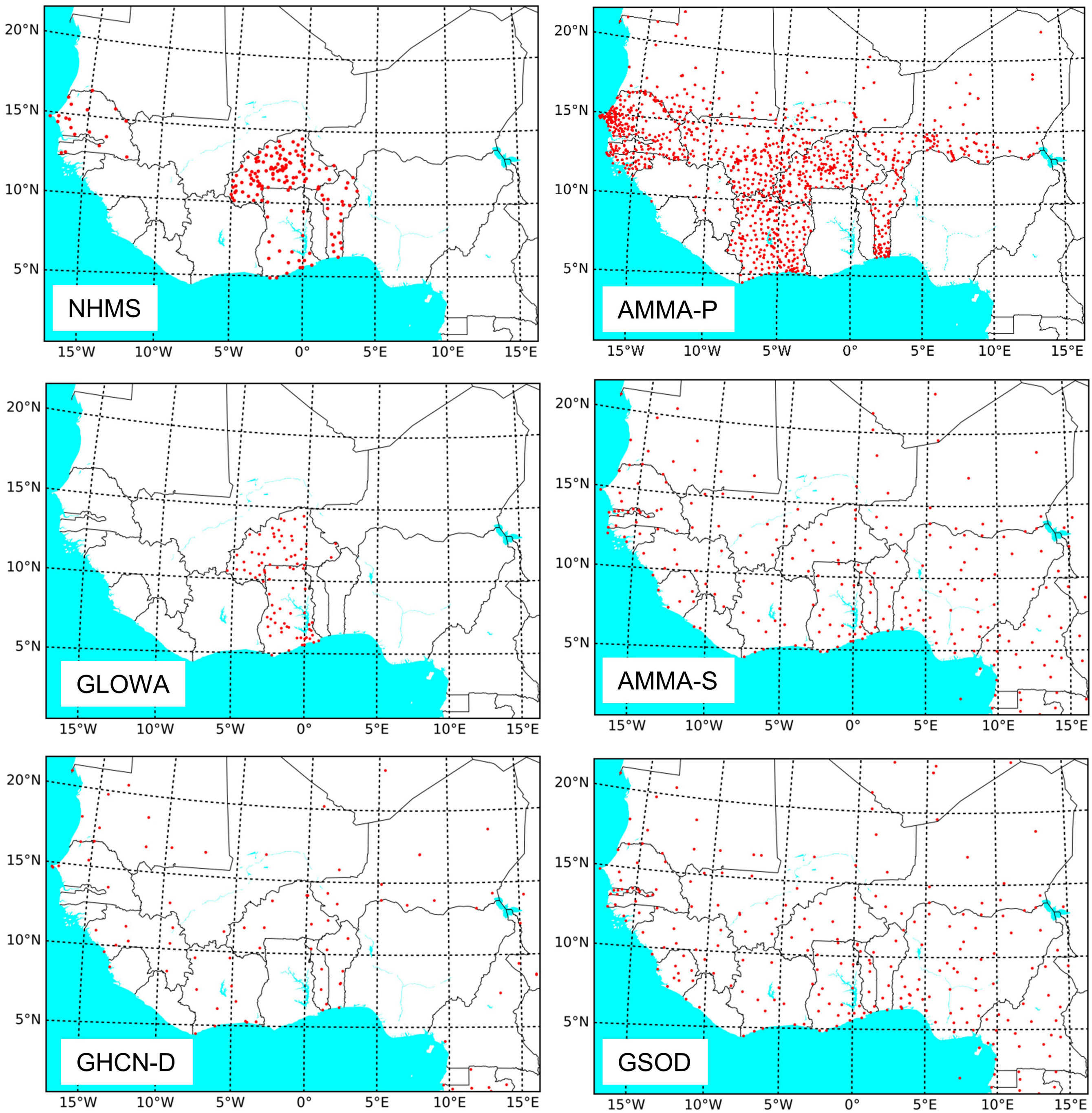

FIGURE 3 Spatial coverage of the rainfall sites with daily measurements from the different data sources (NHMS, AMMA-P, GLOWA, AMMA-S, GHCN, GSOD) used for the compilation of WAHPD-D [Colour figure can be viewed at wileyonlinelibrary.com]

observations for 18 meteorological surface variables from more than 9,000 land stations worldwide, with the best coverage since the early 1970s (NOAA, 2020c). In the past 40 years (1974–2013), approximately 80 measurement values per day have been provided for 112 precipitation stations located in West Africa. Compared to GHCN, the data coverage of the GSOD database has been much better since the 1980s.

The different subsets of the WAPD-D database contain 1,867 daily time series in total, with more than 11.9 million daily precipitation measurements that require quality control. However, due to the same data sources, GHCN-D, GCOS, AMMA-S and the other subsets of WAHPD-D are not independent of each other and may therefore contain many measurements from the same site or measurement device.

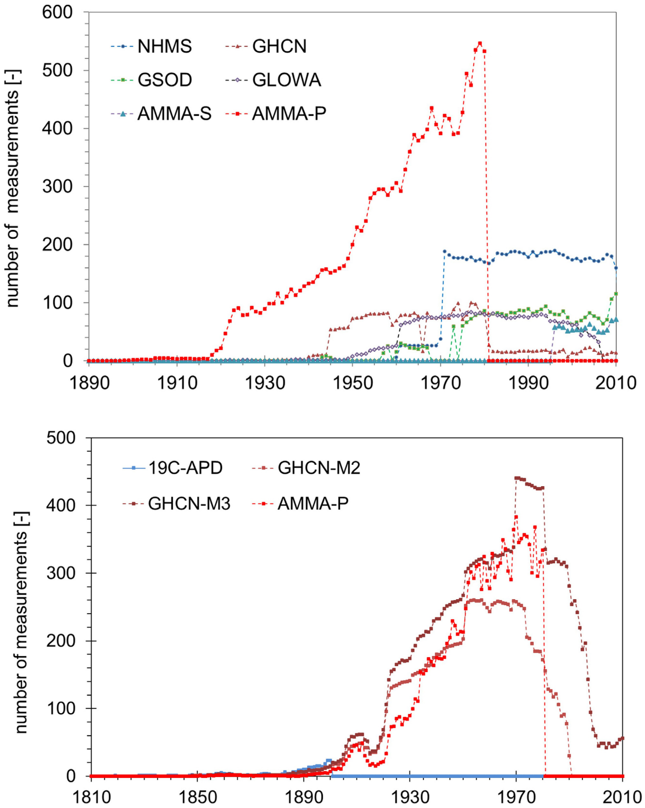

FIGURE 4 Temporal data coverage in terms of the mean annual number of daily (upper panel) and monthly measurements (lower panel) for the different daily (NHMS, AMMA-P, GLOWA, AMMA-S, GHCN-D and GSOD) and monthly archives (AMMA-P, GHCNM-2, GHCNM-3 and 19C-APD) of WAHPD [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Monthly databases

The monthly values of WAHPD-D with approximately 1,150 time series are the most important data set of WAHPD-M. In addition, 1,804 monthly time series of the AMMA-P database with measurements from 1851 to 1980 are used, but only 28% of the time series contained data for over 10 years. This archive also contains monthly measurements for locations that are not part of the daily database, for example, for Nigeria and Guinea (Figure 5) and from two West African neighbouring countries (Cameroon and Chad). Furthermore, the monthly measurements of GHCN version 2 (GHCN-M2, Peterson *et al.*, 1998b) and version 3 (GHCN-M3, Lawrimore *et al.*, 2011) are applied. The data were extracted for sites located in a rectangular domain covering all West African countries (Figure 5). Since GHCN-M2 contains time series that are not part of GHCN-M3, both archives are used. In addition, a small subset of 34 precipitation time series for the 19th century is used (19C-APD, NOAA, 2020a) compiled from a database of historical precipitation records for locations in Africa (Nicholson, 2001). According to this data set, the first precipitation measurements in West Africa were carried out in 1819.

4 | QUALITY CONTROL OF THE PRECIPITATION DATABASES

The quality control system of the WAHPD consists of three components. The first component (Q1) uses several automatic algorithms and manual steps to control the reliability and consistency of data format and meta-information, such as dates and station coordinates (Section 4.1). The second component (Q2) applies straightforward quality control algorithms and visual checks for data screening and elimination (or correction) of unreliable precipitation measurements and time series (Section 4.2). Similar approaches are used for quality control of daily precipitation time series at GHCN (Durre *et al.*, 2010; Menne *et al.*, 2012). The third component (Q3) consists of a geostatistical-based algorithm to detect unreliable precipitation time series in comparison to their neighbourhood (Section 4.3).

4.1 | Quality control of meta-information and data format

The quality control of meta-information and data format consists of the following steps:

1. Control, processing and conversion of data formats into a uniform data format.
2. Check the completeness of dates (e.g., leap year, missing dates, extra days).
3. Visual control of station coordinates and correction of nonplausible coordinates.
4. Check and convert missing values (e.g., −9, 9,999) and other entries (e.g., tr, xxx) to a consistent missing value format (−999).
5. Calculation of data availability (number of observations, average lengths) and gap statistics (e.g., number of data gaps/mean duration of data gaps).
6. Elimination of precipitation stations without any daily or monthly measurements.

This part of data control contains many manual and subset-specific steps, especially for the processing and conversion of data formats. For instance, raw data from the AMMA-P database is a mixture of daily and monthly values for the same precipitation station. A particularly large number of months contained only a single entry (a zero), probably to save space and time during data collection and

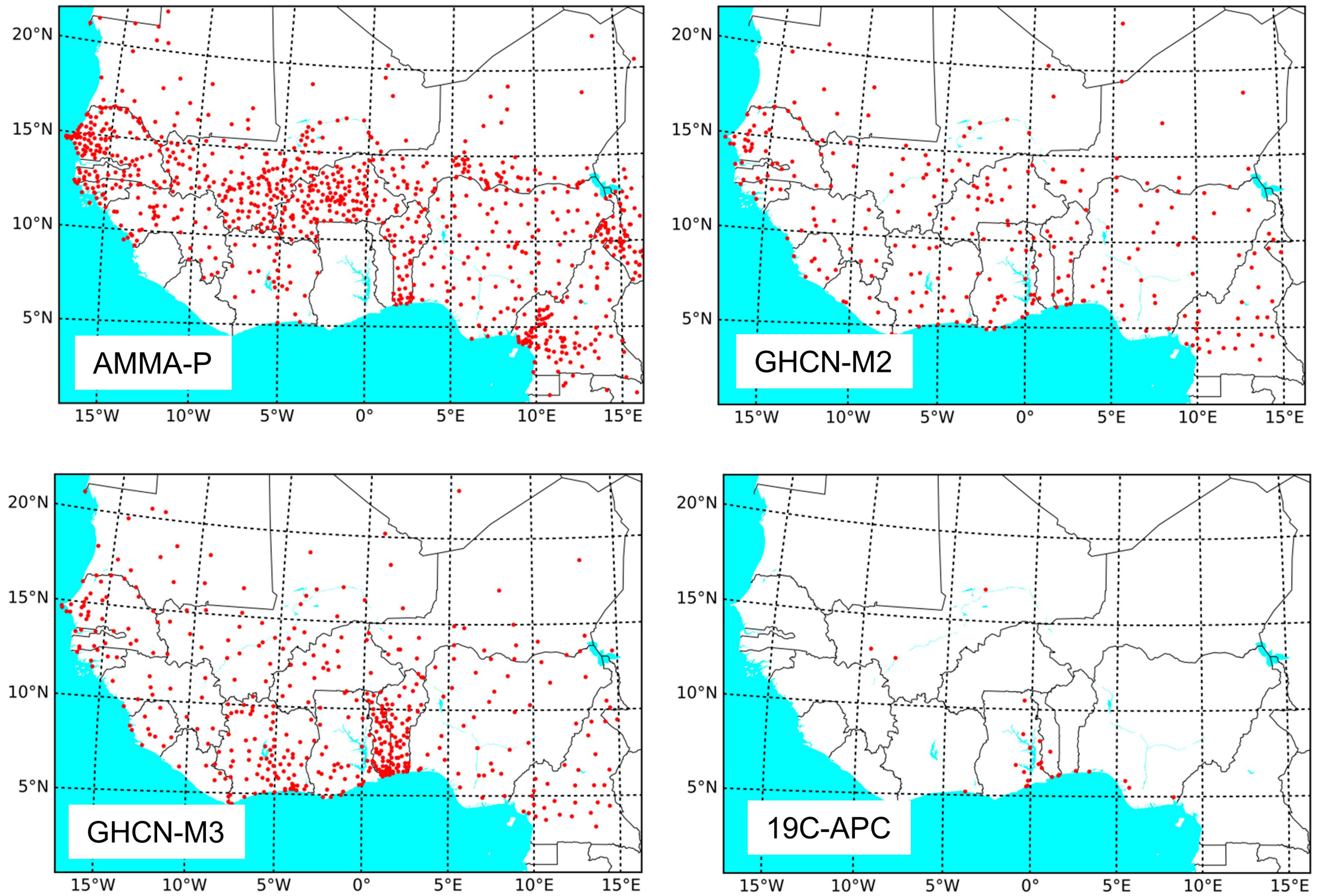

FIGURE 5 Spatial coverage of the rainfall sites with monthly measurements for the different data sources (AMMA-P, GHCN-M1, GHCN-M2, 19C-APD) used for the compilation of the WAHPD-M database. AMMA-P only shows stations with at least 3 years of measurements [Colour figure can be viewed at wileyonlinelibrary.com]

archiving. These zeroes occurred mainly in the dry season as shown for four different AMMA-P subsets in Figure S2. Thus, the monthly zeroes indicate dry months in a relatively plausible way. We used, therefore, the monthly zeros to complete the daily time series with zeros. Based on this procedure, more than 2.8 million data entries were infilled with zeroes. However, there is still the possibility that less or no measurements were carried out during the dry season. Non-automatic precipitation stations may not be permanently manned during this period. This uncertainty should be taken into account when historical precipitation measurements from the rainfall network in West Africa are used for data analysis, especially for questions where the exact timing of the rainy season is of interest, such as the estimation of the onset dates (Laux *et al.*, 2008; Rauch *et al.*, 2019).

Another serious limitation was the NHMS-GH archive's data format. The raw data were provided using a spreadsheet software with 31 × 12 data entries for each year and station (see Figure S3). However, many months contained more days (extra days) than allowed (see Table S1 of the supporting information). Another problem was that the data input (precipitation record or missing value) for February 29 was not available in some leap years or on other days (e.g., December 31). In total, four different documentation errors (extra days, missing days, typing errors and missing unit conversion) and two further documentation limitations (unknown identifiers and missing entries) were identified for this subset (Table S1 of the supporting information). In particular, the extra days were present at almost all sites and occurred in other datasets, as well (Table S2). Extra and missing days are extremely problematic as they lead to a temporal shift of the measurements, affecting the synchronicity of the precipitation measurements in a network. The example also demonstrates the importance of the quality control of data format and dates, the use of standardized data formats, and an appropriate training of staff to avoid these simple documentation errors. To create a consistent data format, the information of all extra-days was neglected, and missing values were added for dates for which no measurements (e.g., blanks, identifiers) were available.

Frequent and long data gaps are a fundamental limitation for almost all precipitation datasets of the WAHPD-D database, but there are strong differences between the subsets (Figure 6, left panel). For instance, the GHCN-D subset has a relatively small average number of data gaps (20.0) per time series with a mean duration of 138.2 days, while the GSOD stations have many more data gaps (711.6) but with a shorter duration (29.4 days). There is only one dataset with almost no data gaps (NHMS-SN). Another problem is that almost all precipitation time series of a subset contains data gaps, as shown in the right panel of Figure 6. It should be also noted that the data gap statistics in Figure 6 were only calculated from the first to the last measurement of the precipitation time series and not over the entire period, and data gaps due to late starting and early ending are also relevant (Bárdossy and Pegram, 2014). Thus, the data gap situation is often much worse, especially if long-term observations from multiple sites are needed.

The outcomes of the Q1 algorithms are listed in Table 2. It shows that a correction of the time series or corresponding meta-information is dataset dependent and was not performed for each step (e.g., data gap statistics). However, for certain subsets like NHMS-GH, the original data series were changed for more than 80% of sites due to extra days or other limitations.

4.2 | Standard quality control of the precipitation measurements

The quality control of the precipitation measurements consists of following steps:

1. Check physical limits of daily and monthly measurements (0–750 $\text{mm}\cdot\text{day}^{-1}$ and 0–5,000 $\text{mm}\cdot\text{month}^{-1}$).
2. Detect and eliminate measurement sequences with the same nonzero measurement value (maximum two repetitions are allowed).
3. Eliminate time series with a low sample size ($n < 365$ for daily and $n < 36$ for monthly).
4. Check for inhomogeneities using double sum analysis and correction of unit conversion errors.
5. Calculation of basic precipitation statistics (e.g., rainfall probability, mean-wet day amount, maximum value, minimum precipitation amount).
6. Qualitative analysis of rainfall statistics in relation to site factors like latitude and height to determine unreliable time series.
7. Check time series repetition within a precipitation subset and merging these to a single time series.

The physical limits for the daily and monthly precipitation amount were subjectively defined based on an iterative procedure following the WMO archive of world weather records (WMO, 2021) and visual analysis of the precipitation extremes concerning site factors (e.g., latitude) to identify unreliable data clusters (e.g., AMMA-S database). Finally, relatively high thresholds were selected to eliminate only a small number of extremes to better keep the original data structure. The results of this quality check are listed jointly with the other Q2 outcomes in Table 3.

The quality control of daily precipitation times series showed that time series can contain repetitions of the same nonzero measurement value for several consecutive days or even longer (as shown in Figure S4). Frequent observations of the same measurement value are fairly unrealistic for moderate or high daily precipitation amounts, due to the high stochastic nature of precipitation in monsoonal regions, but occur in several datasets as listed in Table 3. A similar problem was also shown by Durre *et al.* (2010) for daily precipitation of the GHCN database. According to them, these errors can be related to a malfunction of the precipitation devices or setting inadequate missing values.

Other data problems discovered are inconsistencies due to an incorrect conversion of the precipitation amount from inches to mm. Figure 7 shows that the annual

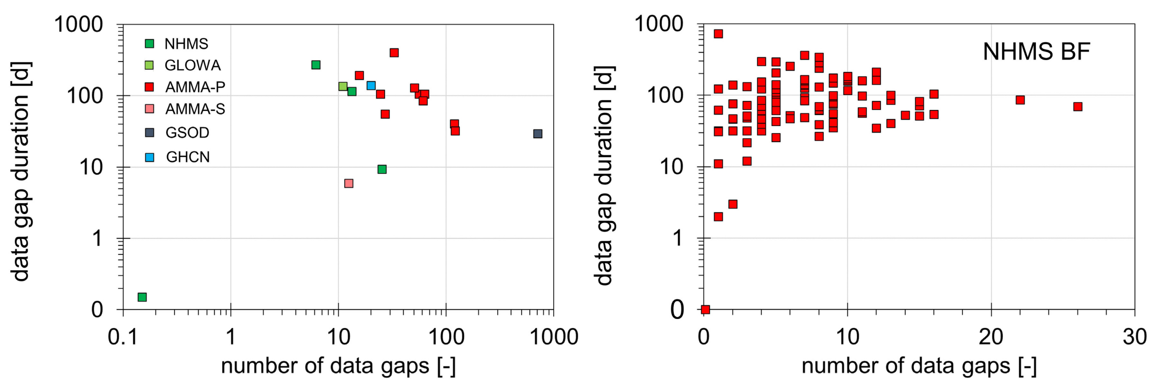

FIGURE 6 Data gap statistics (mean number of data gaps and mean data gap duration) for the precipitation subsets of WAHPD-D (left) and for the individual precipitation stations of the NHMS dataset for Burkina Faso (right) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Outcomes of the quality control (Q1) applied for the control of meta information and data format of the daily subsets of WAHPD (see Figure 2)

			Q1.1	Q1.2	Q1.3	Q1.4	Q1.5	Q1.6
Subset	**nst**	***n***	—	ncm (−)	ncs (−)	ncm (−)	—	ncs (−)
MS-BF	142	1,672,549	b	1.500	1	0	o	1
MS-GH	22	295,704	a,b	0.012	0	0	o	0
MS-BN	23	293,997	b	<100	0	12,505	o	0
MS-SN	20	365,237	b	<10	0	3	o	0
GSOD	214	1,206,198	c	nq	0	0	o	0
GHCN-D	119	1,353,245	c	nq	1	0	o	0
GLOWA	85	1,272,259	—	nq	1	61	o	0
AMMA-S	183	315,702	c	nq	0	0	o	18
AMMA-P BF	154	1,295,564	c	d	0	0	o	0
AMMA-P BN	69	695,807	c	d	0	0	o	0
AMMA-P CD	202	1,235,851	c	d	0	0	o	0
AMMA-P GU	43	128,759	c	d	0	0	o	20
AMMA-P GUB	35	255,939	c	d	0	0	o	0
AMMA-P ML	208	1,815,278	c	d	0	0	o	0
AMMA-P MR	60	470,929	c	d	0	0	o	0
AMMA-P NI	112	799,839	c	d	0	0	o	0
AMMA-P SN	210	1,403,627	c	d	2	0	o	0

Note: ns = total number of stations, n = total number of records (sample size), Q1.1–Q1.6 refers to the six steps of Q1. Q1.2 and Q1.4 shows the number of corrected measurements ncm, Q1.3 and Q1.6 lists the number of corrected (eliminated) stations ncs, a = time consuming manual processing of the data format, b = daily measurements were given as line entries for each month and site, c = measurement series was not complete (e.g., months/days were missing), d = infilling of missing days in the dry period using monthly zeroes, nq = not quantified, o = quality check was performed but no corrective action was done.

precipitation of Kpeve in Ghana from 1981 to 1991 is higher compared to a neighbouring location. A double-sum analysis (Buishand, 1982), in which the accumulated precipitation of Kpeve is compared with those of a neighbouring station, clearly indicates two breakpoints (1981 and 1991). If the annual amounts are divided by a factor of 2.54 a much more reliable annual precipitation time series is obtained and the double sum curve is improved (Figure 7, lower right panel). In total, 5 daily precipitation time series (between 5 and 30% of the measurement values) of the NHMS-GH and GLOWA dataset were affected by this measurement problem and daily measurements of approximately 40 years were corrected (Table 3). Moreover, unit conversion errors are not only a problem for annual precipitation. The affected period can also be much shorter, and for some stations, a sequence of only a few rainfall events is affected. This problem is shown for the precipitation station in Accra for the GLOWA and NHMS datasets in Figure 8. Dividing the precipitation amount of the largest event (95.3 vs. 37.5 mm·day^{-1}) and for several smaller events yields the conversion factor of millimetres to inches. Another problem is a temporal offset of one or multiple days (Figure 8, lower panels).

The calculation of the rainfall statistics revealed relatively reliable precipitation characteristics for many subsets, although more detailed investigations are necessary to confirm these findings. For instance, both global databases (GSOD and GHCN-D) show relatively similar latitudinal changes in rainfall probability for their station network (Figure 9). In the Sahelian region above 17°N, the daily rainfall probability is less than 5%. Between 17°N and 10°N, the rainfall probability strongly increases equatorward from approximately 5% to 20%. This characteristic is also shown in more detail for the large and relatively dense network of more than 200 precipitation stations in Mali from the AMMA-P database. Below 10°N, the rainfall probability is more diverse, and several stations show relatively low rainfall probabilities for this latitude (<15%). The local rainfall along the coast of the Gulf of Guinea anomaly is a typical feature of the West Africa climate, as a result of a strong weakening of the West African Monsoon (WAM) in this region (Acheampong, 1982; Vollmert and Fink, 2003; Aryee *et al.*, 2018). This feature is also shown for several stations in the GLOWA database. Another important characteristic is the extremely high precipitation amounts (>3,000 mm·a^{-1}) along the mountain

TABLE 3 Outcomes of the quality control algorithms for each step of Q2 and the geostatistical algorithm (Q3), daily precipitation subsets of WAHPD (see Figure 2)

	Q2.1	Q2.2	Q2.3	Q2.4	Q2.5	Q2.6	Q2.7	Q3
Subset	**ncm (–)**	**ncm (–)**	**ncs (–)**	**ncs (–)**	**—**	**—**	**ncs (–)**	**ncs (–)**
MS-BF	0	0	1	x	o	o	2	1
MS-GH	0	3	0	3	o	(o)	0	0
MS-BN	0	0	0	x	o	(o)	0	0
MS-SN	0	0	0	x	o	(o)	0	0
GSOD	10	335	47	(o)	o	(o)	2	8
GHCN-D	3	31	0	(o)	o	(o)	1	4
GLOWA	2	28	0	2	o	o	1	2
AMMA-S	17	0	66	(o)	o	o	2	20
AMMA-P BF	0	2	2	x	o	o	2	3
AMMA-P BN	0	12	1	x	o	o	1	3
AMMA-P CD	0	31	5	x	o	(o)	6	40
AMMA-P GU	2	3	1	x	o	o	x	x
AMMA-P GUB	2	259	2	x	o	o	x	x
AMMA-P ML	0	13	7	x	o	o	8	25
AMMA-P MR	0	0	15	x	o	(o)	x	x
AMMA-P NI	0	1	0	x	o	o	5	14
AMMA-P SN	0	14	1	x	o	(o)	x	x

Note: ncm = number of corrected (eliminated) measurement values, ncs = number of corrected (eliminated) sites, x = no quality check was performed, o = quality check was performed but no corrective action was done, (o) = quality check was only partially performed (because site specific information was missing).

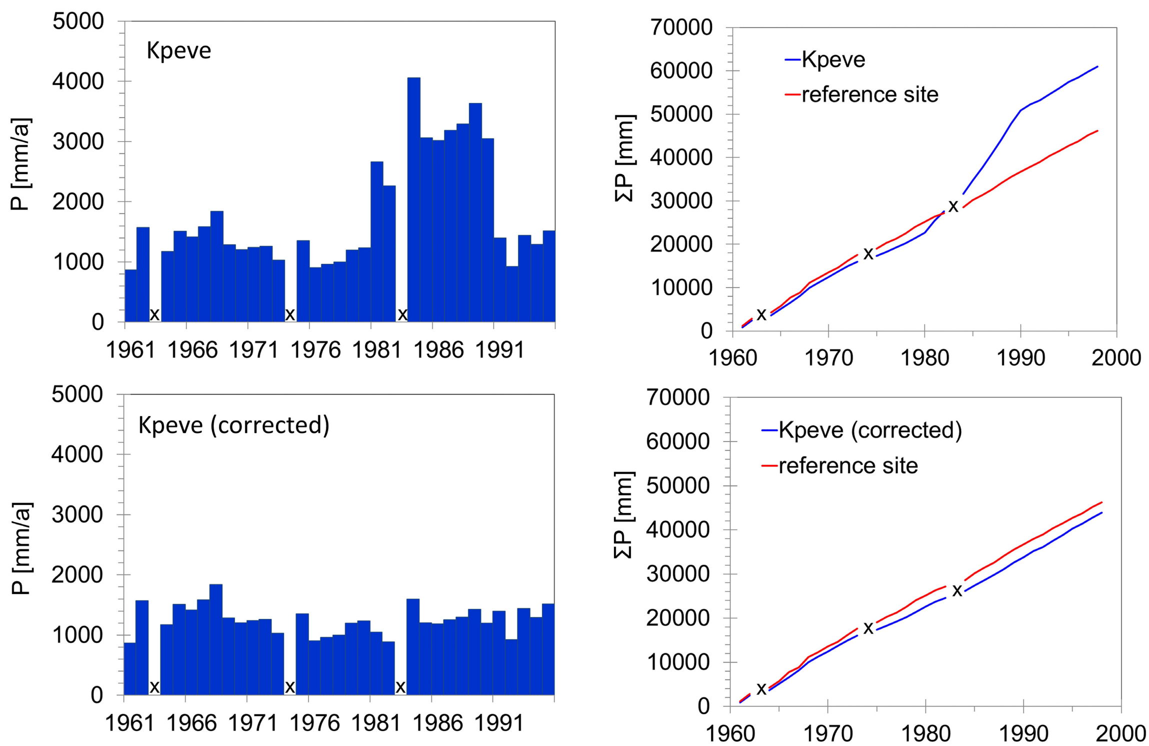

FIGURE 7 Annual precipitation amounts at Kpeve (Ghana) taken from the GLOWA database (upper left); double-mass-analysis for Kpeve in comparison to the reference station Kumasi (upper right); corrected annual precipitation amounts for Kpeve in Ghana (lower left); double-mass-curve for the corrected annual precipitation time series for Kpeve in comparison to the reference site (lower right); missing values are indicated by a cross, 1961–1999 [Colour figure can be viewed at wileyonlinelibrary.com]

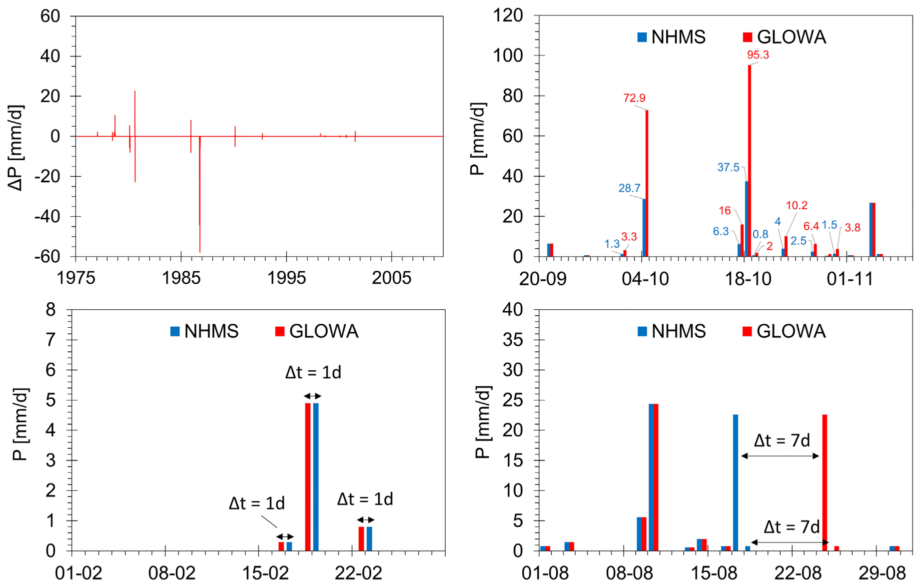

FIGURE 8 Precipitation differences (ΔP) between two daily time series from the same precipitation station in Accra (Ghana) but from two different subsets, namely GLOWA and NHMS, 1975–2009 (upper left). Comparison between both precipitation time series used for the calculation of ΔP for the period between September 19, 1986 and November 9, 1986 showing several unit conversion errors (upper right). Comparison between both daily precipitation time series indicating a temporal shift of one (bottom left, February 1990) and 7 days (bottom right, August 1980) [Colour figure can be viewed at wileyonlinelibrary.com]

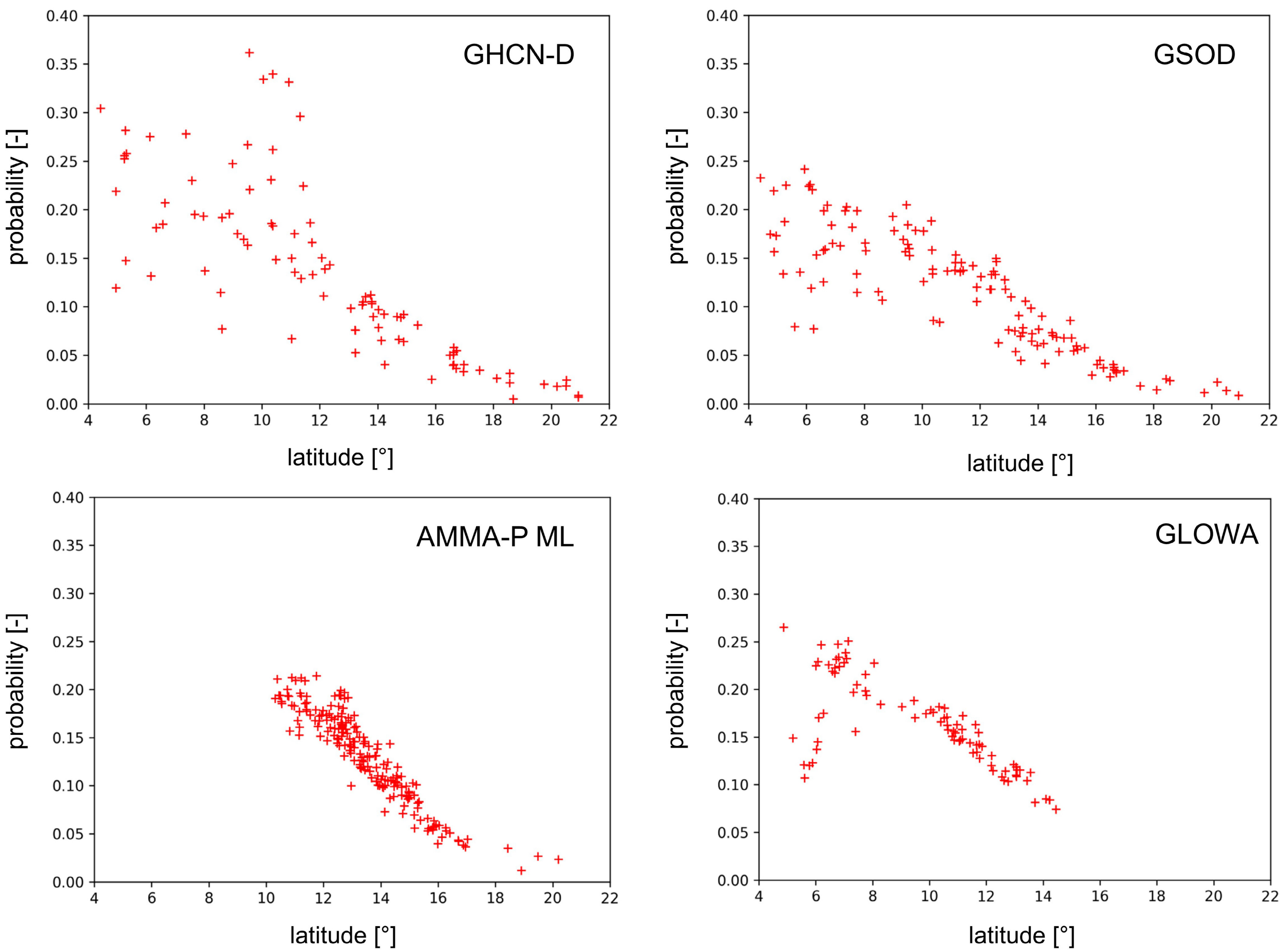

FIGURE 9 Latitudinal changes of the mean daily rainfall probability for four selected subsets (GHCN-D, GSOD, AMMA-P ML, GLOWA) of WAHPD-D [Colour figure can be viewed at wileyonlinelibrary.com]

ranges in Southwest West Africa due to orographic lifting, as described for the Fouta Djallon Highlands in Northwest Guinea by Sall *et al.* (2007). The isolated GHCN-D data points with the highest (30–45%) and relatively unusual rainfall probabilities between 9°N and 12°N lie precisely in this region. Based on this additional information, these time series cannot be excluded. However, this cluster is not shown in the GSOD subset, although the same rainfall sites are part of this database. This outcome indicates that there are substantial differences between both global precipitation subsets for certain regions, and that more detailed investigations are necessary to clarify how reliable these global datasets are for West Africa.

4.3 | Geostatistical quality control using spatial correlograms

Precipitation measurements and meta-information such as station coordinates can be wrong. To tackle these problems jointly, a geostatistical algorithm was developed to determine precipitation stations whose time series are characterized by an unusual spatial correspondence compared to their neighbourhood. The basis of this algorithm is a pairwise comparison of a precipitation time series $x=(x_1,x_2, \ldots ,x_n)$ of a site i and a precipitation time series $y=(y_1,y_2, \ldots ,y_n)$ of a site j using correlation measures like Pearson correlation r_{ij} or the Spearman rank correlation ρ_{ij}. The correlation measure is related to the separation distance h_{ij} between both stations, which is usually based on the Euclidean distance. The measure is then calculated for each data pair and is plotted in relation to the corresponding separation distance. These scatter plots (spatial correlogram clouds) are created in geostatistics as the basis for calculating empirical correlogram functions (Lorenz *et al.*, 2018). Examples of spatial correlograms for precipitation are given for example, by Ciach and Krajewski (2006), Bliefernicht *et al.* (2008) and Schroeer *et al.* (2018).

The graphical analysis of the spatial dependence structure of the daily precipitation measurements is combined with a statistical procedure for eliminating unreliable precipitation time series from the original database. This algorithm consists of the following steps:

1. Selection of station i and its closest neighbour stations. The closest stations are defined using the Euclidean distance.
2. Calculation of the mean correlation $\overline{r}_i$ and mean distance $\overline{d}_i$ between station i and its neighbours:

$$\overline{r}_i=\frac{1}{n_c}\sum_{i=1}^{n_c} r_{ij} \quad (1)$$

$$\overline{d}_i=\frac{1}{n_c}\sum_{i=1}^{n_c} d_{ij} \quad (2)$$

3. Standardization of the correlation measure using a z-score transformation:

$$z_i=\frac{\overline{r}_i(d)-\overline{r}(d)}{s_r(d)} \quad (3)$$

   Large negative (positive) values indicate stations with much lower (higher) spatial correspondence in comparison to their neighbourhood.
4. Selection of rejection threshold z_t and comparison with standardized values z_i. If $z_i>|z_t|$, the precipitation time series of station i is removed from the original database, otherwise it is accepted. We used a threshold value of $z_t=1.96$ related to a 95%-confidence interval for rejection.
5. Repeat steps 1–4 for all stations.

This quality control is performed globally, so that the entire precipitation time series is excluded from the dataset if it fails the test. An example of a correlogram cloud is shown for four databases in Figure 10 using Pearson correlation. The correlogram clouds for AMMA-P, GLOWA and the NHMS Burkina Faso subset show that spatial correspondence increases with decreasing distance and that there are no strong outliers. However, the correlogram cloud of the GHCN-D database indicates many unreliable data points. The problems presented by the GHCN-D subset were a one-day time lag for several time series in Nigeria and incorrect coordinates (1.52°E instead of 1.52°W) for the precipitation station at the Ouagadougou international airport.

The geostatistical check was performed for the daily datasets needed to set up the harmonized database for the focal region (see column Q3 of Table 3). In total, 1,237 time series were used and 119 time series (roughly 10%) were excluded from the datasets by this algorithm.

5 | HARMONIZATION OF WAHPD DATASETS

In this section, the harmonization of the different data archives to a joint database is described. In addition, an application of the joint database is briefly illustrated using a common interpolation algorithm.

5.1 | Initial analysis for Ouagadougou

The harmonization of the different data archives into a single archive is not straightforward. A basic limitation is

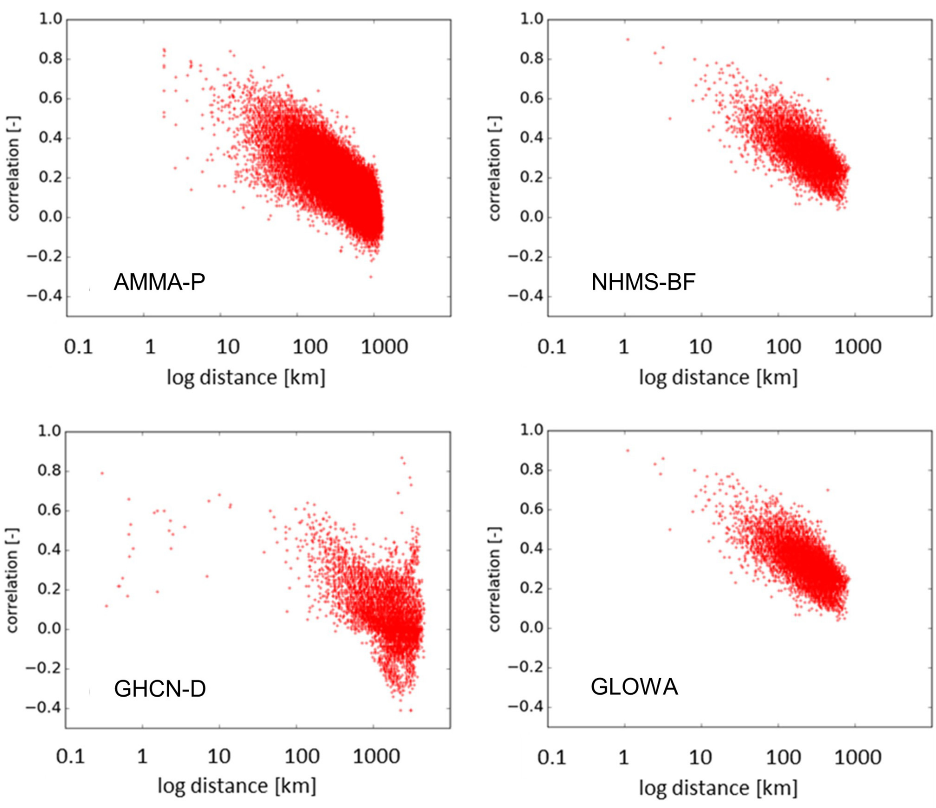

FIGURE 10 Spatial correlogram cloud for four selected subsets (AMMA-P [BF/BN], NHMS-BF, NHMS-BN, GHCN) of the WAHPD-D database [Colour figure can be viewed at wileyonlinelibrary.com]

that precipitation time series are often not identical, although the data comes from the same precipitation station. This problem is shown in Figure 11 for the station located at the international airport in Ouagadougou (Burkina Faso). The precipitation time series at this site is part of five precipitation subsets (NHMS, AMMA-P, GSOD, GHCN-D and AMMA-S). Figure 11 shows an inter-comparison of these subsets, in which the measurements from the meteorological service of Burkina Faso are chosen as reference values. In addition, a scatter plot is shown for a neighbouring location (d_{ij} = 2.2 km). Several standard measures, such as the mean absolute error (MAE), are calculated to compare the time series (Table 4). In addition, two binary association measures (PSS and PCM) are calculated. PSS is based on the Peirce skill score (Hogan and Mason, 2012). PCM is the proportion of close measurements to determine the frequency of data pairs that are close to the 1:1 line of a scatterplot and are above a precipitation threshold ($p_t = 1.0\,\text{mm}\cdot\text{day}^{-1}$). The PCM value ranges between 0 and 1 and perfect match is indicated by 1. Figure 11 shows a very high correspondence for AMMA-P and GHCN-D. Only a few measurements differ from the NHMS observations, leading to a very low MAE and very high r and PCM, respectively. A much poorer association is shown for the other two datasets. The correspondence of the GSOD subset (r = 0.526) is even lower in comparison to the neighbouring station (r = 0.860). The scatter plot also shows that many GSOD-NHMS pairs are relatively close to the 1:1 line. This leads to a relatively high PCM value of 0.748, which is much higher in comparison to the neighbouring site (PCM = 0.134). Thus, a substantial part of the GSOD measurements seems to have the same origin as the NHMS observations.

A second limitation of the different data sources is shown in Table 5. The meta-information of the stations can also be different, such as the geographical coordinates. Because of both limitations (different precipitation values and coordinates), adding a precipitation time series to a reference database is tedious, and automatic algorithms are required to facilitate this task.

5.2 | Semi-automatic algorithm for harmonization

For harmonization of the individual database, a semi-automatic algorithm is used, which searches for a time series in a new dataset that is part of the reference dataset. This algorithm consists of the following steps:

1. Select station j from the new dataset.
2. Calculation of the distance d_{ij} between station i (reference dataset) and station j.
3. Calculation of the correspondence measures, for example, r_{ij} or PCM_{ij}.

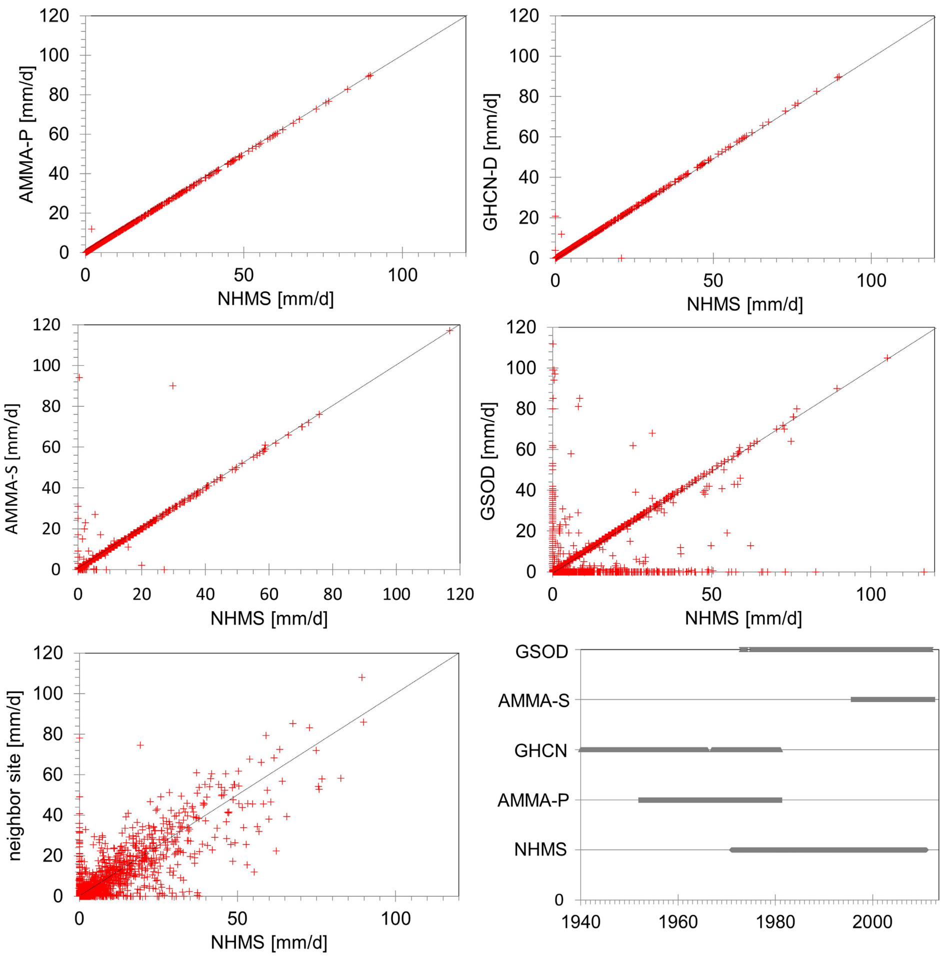

FIGURE 11 Inter-comparison of the daily precipitation amounts for the precipitation station located at the international airport Ouagadougou of the NHMS subset in comparison to four precipitation subsets (NHMS, AMMA-P, GHCN-D, AMMA-S, GSOD) of the WAHPD-D database and a neighbour site (bottom left). The neighbour site (Ouagadougou mission) is located very close (approx. 2 km) to the precipitation station at the airport. The diagram in the lower right corner shows the timeline of the measurement data for the different precipitation subsets [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Quantitative comparison of four precipitation subsets (AMMA-P, GHCN-D, AMMA-S, GSOD) with NHMS observations for the precipitation stations at the Ouagadougou airport using several correspondence measures

Measure	AMMA-P	GHCN-D	AMMA-S	GSOD	REF
n_t (−)	670	668	838	1,646	992
k (−)	669	666	730	1,232	133
MAE (mm·day^{-1})	0.015	0.016	0.569	2.236	5.330
r (−)	1.000	0.998	0.971	0.526	0.860
PSS (−)	0.998	0.993	0.910	0.784	0.816
PCM (−)	0.999	0.997	0.871	0.748	0.134

Note: MAE = mean absolute error, r = Pearson correlation, PSS = binary discrimination measure based on the Peirce skill score measuring the difference between the rate of right wet days minus the rate of false dry days, PCM = proportion of close precipitation amounts, PCM = k/n_t, k = number of close measurements, n_t = number of joint precipitation measurements above a precipitation threshold $p_t = 1.0\,\text{mm}/d$, ref = reference site, NHMS = National Hydrological Meteorological Services, AMMA-P = AMMA Pluvio database, AMMA-S = AMMA-SYNOP database, GSOD = global surface summary of the day, GHCN-D = Global Historical Climate Network daily database, REF = neighbouring stations close (<2.2 km) to the Ouagadougou Airport.

4. Repeat steps 2 and 3 for all stations in the reference dataset.
5. Select station i with the maximum correspondence measure and do infilling with the measurement from station j for selected station i if conditions for an infilling are fulfilled (e.g., $PCM_{ij} > 0.2$ and $d_{ij} < 10$ km).
6. If conditions are not fulfilled, station j is accepted as a new station in the reference database.
7. Repeat steps 1–6 for all further stations in the new dataset.

The harmonization algorithm was tested for Burkina Faso, Ghana, Benin and Togo using 4 quality-controlled daily archives (GLOWA, NHMS, AMMA-S and GHCN-D) and 3 monthly archives (WAHPD-D, GHCN-M2 and GHCN-M3). The archives were added

TABLE 5 Meta information of the different precipitation subsets for the station located at the Ouagadougou international airport

Subset	Lat. (°E)	Lon. (°N)	z (m)	Name
NHMS	−1.5167	12.35	303	Ouaga_aero
AMMA-P	−1.5167	12.35	304	Ouagadougou_aero
AMMA-S	−1.5167	12.35	306	Ouagadougou
GSOD	−1.517	12.35	306	Ouagadougou
GHCN-D	−1.52[a]	12.35	304	Ouagadougou

Note: Lat. = latitude, lon. = longitude, z = elevation height, NHMS = National Hydrological Meteorological Services, AMMA-P = AMMA Pluvio, AMMA-S = AMMA-SYNOP, GSOD = global surface summary of the day, GHCN-D = daily database of Global Historical Climate Network.
[a]Corrected value due to quality routines used in Section 4.3; the original value was 1.52.

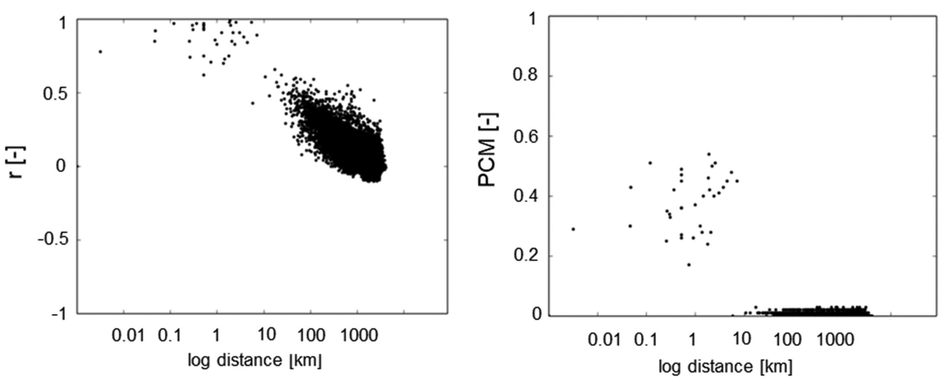

FIGURE 12 Inter-comparison of the AMMA-SYNOP database with WAHPD-D (consisting of NHMS, GHCN-D and GLOWA) using the Pearson correlation r (left) and the proportion of close measurements PCM (right)

iteratively based on a reference data set. In each step, a scatter plot is created, which relates a given correspondence measure to the distance in order to visualize the harmonization process. An example of this scatterplot is shown in Figure 12 when the AMMA-S database was compared with the merged daily database (GLOWA and NHMS). The figure shows that there are no perfect matches. However, several data pairs have a relatively high agreement and a small distance in comparison to many other data pairs. An even better separation between the two groups is achieved in the PCM diagram.

5.3 | Application of the harmonized precipitation database

The final database of the harmonization algorithm consists of 413 daily time series (1940–2013) and 687 monthly time series (1847–2013). Around 50% of the stations are located in the four countries. This database was used for a spatial interpolation using Ordinary Kriging (Atkinson and Lloyd, 1998; Tobin *et al.*, 2011) to check whether important rainfall features of the WAM can be reliably reproduced for this region. Figure 13 shows the interpolation outcomes for four target variables with a focus on the monsoon peak in August. The precipitation band with the highest intensities and probabilities is located at approximately 11°N. This outcome confirms Nicholson (2013), in that the main precipitation band lies around 10°N during the Sahelian phase. Another important feature is the very low precipitation amounts (< 50 mm·day^{-1}) and probabilities (<10%) along the coast of the Guinean Gulf due to a strong and unusual WAM weakening for this latitude (see also Section 4.2). Moreover, typical interpolation artefacts like ‘bull eyes’ are still visible in the rainfall probability patterns, especially in Northwest Benin. However, these are much less pronounced in comparison to former interpolation works done for this region (Laux *et al.*, 2009). Thus, the outcomes of the spatial interpolation show that relatively reliable spatial patterns for the different precipitation variables can be produced with the quality-controlled and harmonized database.

6 | DISCUSSION

The outcomes of the previous sections showed the importance of a comprehensive quality control of the precipitation time series and their meta-information for the different datasets used by this study. It was possible to identify very simple errors in the data records, such as the conversion error from millimetres to inches or the extra days, which could have been avoided through good data documentation. Although we applied several different controls of metadata and precipitation time series,

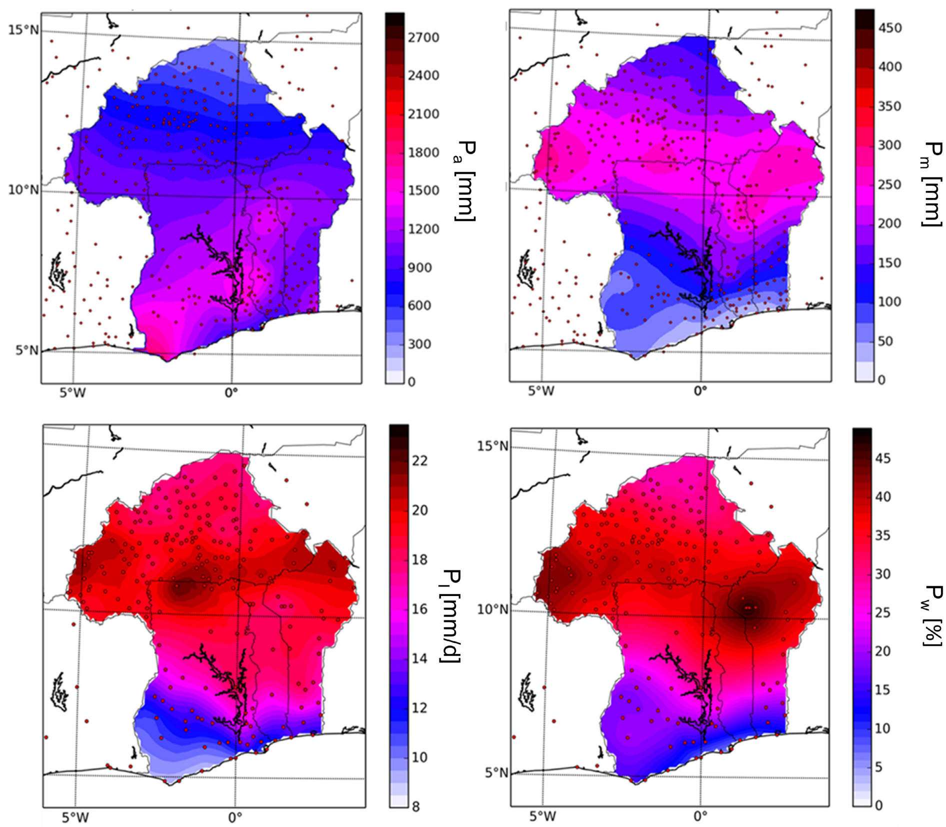

FIGURE 13 Interpolated precipitation patterns using the quality-controlled and harmonized WAHPD database for the focal region (Burkina Faso, Ghana, Benin and Togo) for a period ranging from 1960 to 2010. P_a = mean annual precipitation amount, P_m = mean monthly precipitation amount, P_I = mean precipitation intensity in terms of the mean wet day amount and P_w = mean daily precipitation probability (%). P_m, P_I and P_w are shown for August, the peak period of the West African Monsoon. The red dots indicate the location of the precipitation stations used for a spatial interpolation of the variable of interest [Colour figure can be viewed at wileyonlinelibrary.com]

some data limitations could only be discovered by the geostatistical approaches (as shown for the GHCN database in Figure 10). This makes it clear how important a joint analysis of the precipitation time series and the station coordinates is. One limitation of the current quality assessment is that the geostatistical approach is applied globally. However, Costa and Soares (2009) noted that quality-control approaches on finer temporal scales are generally lacking. The same applies to standard methods like the double sum analysis, which is usually carried out with monthly or annual precipitation amounts (Buishand, 1982; Bickici Arikan and Kahya, 2019). Another limitation of the quality controls performed in this study is that no flag system is used, although this is very common (Durre *et al.*, 2010; Menne *et al.*, 2012). A flag-based system has the advantage that information about data quality is available for each precipitation value or time series. A further limitation is the subjective evaluation of the proposed algorithms. In the case of the harmonization algorithm, a manual inspection of data protocols and added time series was performed to minimize the error that stations are incorrectly added to other stations. However, future investigations should focus on the development of evaluation procedures to better determine the quality of the applied algorithms in terms of the number of type I and II errors and related quality measures like the false positive rate, as proposed by Durre *et al.* (2008) for the GHCN database.

The analysis in Figure 6 also revealed that 17 out 18 daily archives applied in this study contain many time series with long (>10 days) and frequent (>10) data gaps. However, time series with no data gaps are needed for many different applications in climate sciences and in many other disciplines. Moreover, infilled time series are essential for a better quality analysis (Costa and Soares, 2009), in particular for joint quality control at multiple sites. However, closing data gaps for daily precipitation in a reliable way is extremely difficult in our study region due to the high stochastic nature of this variable and the low station density. Moreover, common infilling approaches like inverse distance or regression approaches (Di Piazza *et al.*, 2011; Campozano *et al.*, 2014) are only applied deterministically, thereby ignoring the inherent uncertainty of the infilling problem and not maintaining the variability of the target variable. There are more sophisticated methods for infilling precipitation time series, such as those presented by Bárdossy and Pegram (2014) using Copulas. Due to their complexity, however, these approaches cannot be easily integrated within the quality control algorithms and were therefore out of the scope of this study.

The investigation also indicated that many precipitation subsets used in this study have a moderate to good quality after passing the different steps (Q1–Q3) of the quality control algorithm, although many different errors and uncertainties were identified in the data. For instance, the spatial interpolation of the four West African countries using the quality-controlled and harmonized precipitation database showed relatively reliable patterns for monthly precipitation and the precipitation probability for the peak period of the West African Monsoon. However, a more detailed analysis is needed to confirm these initial findings. These investigations should also assess rainfall characteristics on other time scales (e.g., seasonal, decadal) as those presented here. In addition, rainfall indices like intra-seasonal rainfall characteristics (e.g., dry-spells) need to be used to assess whether this information can be reproduced reliably, as well.

The free access to this database for research and other noncommercial purposes is a remaining challenge. Most of the data are owned by the NHMS and is, therefore, based on national data protection regulations that usually guide the free provision of daily meteorological time series. Selected precipitation time series from the WAHPD database can be obtained for research and other noncommercial purposes, if corresponding formal agreements are made between the involved institutions (data user, WASCAL and NHMS) meeting the ongoing country-specific memorandum of understandings for data use and sharing established within the framework of WASCAL (Salack *et al*., 2019). However, the current data archive can be used to provide related precipitation products such as interpolated data sets in high spatial resolution for research and other noncommercial purposes. For example, a first version of a historical daily gridded data set with a spatial resolution of 10 km is provided via the WASCAL database (WASCAL, 2020). In addition, point and areal precipitation statistics can be computed and freely provided for precipitation sites or ungaged areas. This was realized by Salack *et al*. (2018a, 2018b) for locations in the Sahel region, focusing on precipitation extremes. Thus, several alternative steps have been initiated and implemented to overcome the current limitations of data provision and to improve the provision of station-based precipitation products for this region.

There are still many ways to expand the current database for historical periods (before 2010). An extension of this database to cover other West African countries that were not the focus of the current study (like BF, GH, TO and BN) is also planned for future studies. One example is the AMMA-DClim database, which contains climate data from 143 synoptic stations in 14 French-speaking countries, with daily measurements ranging from 1854 to 1983. Another example is the historical rainfall databases of the NHMS. As shown in the previous section, the daily measurements of the national rainfall networks are missing in WAHPD for many West African countries since 1980, and there is practically no daily (monthly) information for Nigeria, Sierra Leone and Liberia. However, an extension of the database is not straightforward. A basic problem is that precipitation measurements at NHMS are still available only in hard archives (e.g., paper or micro-fiches) and are, therefore, not directly accessible. Future initiatives should focus on the collection and digitalization of these historical databases to conserve this valuable information, as recently shown for Ghana by Israelsson *et al*. (2020), and to make the data available for research.

There are also many possibilities to update the WAHPD database with more recent precipitation measurements (>2013) and in a much higher temporal resolution (<1 day). The global database of GHCN and GSOD is regularly updated with new measurements (Menne *et al*., 2018). Although the number of updated daily time series is relatively low (<100) and can contain data gaps, this information can serve as initial observed precipitation information for specific sites and regions in West Africa. In addition, precipitation networks that are not directly operated by the NHMS can be used for the expansion of WAHPD like the meso-scale rainfall networks of the AMMA programme (Galle *et al*., 2018) and the DACCIWA (Dynamics-Aerosol-Chemistry-Cloud Interactions) project (Maranan *et al*., 2020). WASCAL also established a new transnational hydro-meteorological network for West Africa in 2017 jointly with the NHMS with 50 new or upgraded weather stations located in 10 West African countries (Salack *et al*., 2019) complemented by three meso-scale hydro-meteorological networks with more than 30 rainfall sites (Bliefernicht *et al*., 2018; Salack *et al*., 2019). Another important example is the Trans-African Hydro-Meteorological Network (TAHMO, 2020) of low cost weather sensors established within the last few years over Africa (van de Giesen *et al*., 2014). It contains more than 200 sensors in eight West African countries (TAHMO, 2020).

7 | SUMMARY AND CONCLUSIONS

In this study, a new station-based precipitation dataset, the West African Historical Precipitation Database, was designed for West Africa with a focus on Burkina Faso, Ghana, Benin and Togo. The precipitation data were compiled from more than 20 national, continental and global data archives and contain long-term daily and monthly precipitation measurements for more than 1,000 measurement locations over a period from 1819 to 2013. It is, therefore, the most comprehensive dataset with long-term daily and monthly precipitation observations for the West African region. The free access to this database for research and other noncommercial purposes remains a challenge due to national data protection regulations. However, several further tasks have been initiated and implemented (e.g., initiation of data sharing policies, provision of gridded precipitation products and statistics) to improve the access and availability of station-based precipitation observations and related data products for this challenging region.

We also established a semi-automatic geostatistical approach in addition to the application of various standard algorithms for control of meta-information, data format and precipitation measurements. This new algorithm allows a joint quality control of precipitation time series and meta-information (station coordinates) to eliminate unreliable time series in comparison to their neighbourhood. In addition, a geostatistical-based algorithm for harmonizing the different precipitation subsets was developed and applied to generate a joint database for the focal region. A spatial interpolation of this new database illustrated that relatively reliable precipitation patterns can be generated for this challenging region. However, a basic shortcoming of the applied quality algorithms is that no flag-based system is used and not detailed validation has been carried out so far to better assess their quality.

The data screening using various quality algorithms revealed different data limitations, which were present in the measurement values (unit conversion errors, temporal offsets), meta-information (wrong station coordinates) and data format (incorrect dates and poor missing value coding). We assume that many of these limitations can be avoided through better data documentation and processing. According to our current state of knowledge, these limitations seem to be rarely documented in the scientific literature but are important for a better understanding of the uncertainties and errors coming from station-based precipitation databases. It also gives insights into how quality algorithms for daily and monthly precipitation can be advanced in the future for this climatologically challenging region.

We also highlighted in this study that there are many ways to extend the current database for historical and present periods. They range from digitizing national data archives, embedding new transnational and research observation networks to using networks based on low-cost weather stations. Future initiatives should closely work with NHMS and partner institutions to integrate this information in existing databases to improve the availability of station-based precipitation observations for the different countries. This is of great importance for providing improved precipitation products for the West African region. We are sure that in the long-term run not only climate research and services, but also many other disciplines will profit from these enhanced precipitation databases.

ACKNOWLEDGEMENTS

This work was part of the WASCAL phase I programme (2010–2017, www.wascal.org) granted by the Federal Ministry of Education and Research in Germany (grant numbers 01LG1202C and 01LG1202C1) and German Research Foundation (grant number KU/2090 14-1). We would like to thank the meteorological services of Burkina Faso, Ghana, Benin and Senegal for providing the NHMS data as part of the WASCAL observation network. We also acknowledge the AMMA programme for providing the meteorological data sets used in this study. We also thank the NOAA's National Centers for Environmental Information and the Volta Basin Authority for providing meteorological data via their databases and two anonymous reviewers for their valuable comments. Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

Jan Bliefernicht: Analysis, conceptualization, first draft writing. **Seyni Salack**: Editing, reviewing. **Moussa Waongo**: Editing, data curation, reviewing. **Thompson Annor**: Editing, data curation, reviewing. **Patrick Laux**: Editing, data curation, reviewing. **Harald Kunstmann**: Funding acquisition, reviewing.

ORCID

Jan Bliefernicht https://orcid.org/0000-0002-8591-6231
Seyni Salack https://orcid.org/0000-0002-1308-6742
Patrick Laux https://orcid.org/0000-0002-8657-6152
Harald Kunstmann https://orcid.org/0000-0001-9573-1743

REFERENCES

Acheampong, P.K. (1982) Rainfall anomaly along the coast of Ghana. Its nature and causes. *Geografiska Annaler*, 64(3), 199–211. Available at:. http://www.jstor.com/stable/520646.

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. and Wieringa, J. (2003) Guidance on metadata and homogenization. *World Meteorological Organization*, 1186, 1–53.

Annor, T., Lamptey, B., Wagner, S., Oguntunde, P., Arnault, J., Heinzeller, D. and Kunstmann, H. (2018) High-resolution long-term WRF climate simulations over Volta Basin. Part 1: validation analysis for temperature and precipitation. *Theoretical and Applied Climatology*, 133(3), 829–849. https://doi.org/10.1007/s00704-017-2223-5.

Aryee, J.N.A., Amekudzi, L.K., Quansah, E., Klutse, N.A.B., Atiah, W.A. and Yorke, C. (2018) Development of high spatial resolution rainfall data for Ghana. *International Journal of Climatology*, 38(3), 1201–1215. https://doi.org/10.1002/joc.5238.

Ascott, M.J., Macdonald, D.M.J., Black, E., Verhoef, A., Nakohoun, P., Tirogo, J., Sandwidi, W.J.P., Bliefernicht, J., Sorensen, J.P.R. and Bossa, A.Y. (2020) In situ observations and lumped parameter model reconstructions reveal intra-annual to multidecadal variability in groundwater levels in sub-Saharan Africa. *Water Resources Research*, 56(12), e2020WR028056. https://doi.org/10.1029/2020WR028056.

Atkinson, P.M. and Lloyd, C.D. (1998) Mapping precipitation in Switzerland with ordinary and indicator kriging. Special issue: spatial interpolation comparison 97. *Journal of Geographic Information and Decision Analysis*, 2(1–2), 72–86.

Bárdossy, A. and Pegram, G. (2014) Infilling missing precipitation records—a comparison of a new copula-based method with other techniques. *Journal of Hydrology*, 519, 1162–1170. https://doi.org/10.1016/j.jhydrol.2014.08.025.

Barry, A.A., Caesar, J., Klein Tank, A.M.G., Aguilar, E., McSweeney, C., Cyrille, A.M., Nikiema, M.P., Narcisse, K.B., Sima, F., Stafford, G., Touray, L.M., Ayilari-Naa, J.A., Mendes, C.L., Tounkara, M., Gar-Glahn, E.V.S., Coulibaly, M. S., Dieh, M.F., Mouhaimouni, M., Oyegade, J.A., Sambou, E. and Laogbessi, E.T. (2018) West Africa climate extremes and climate change indices. *International Journal of Climatology*, 38(S1), e921–e938. https://doi.org/10.1002/joc.5420.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U. and Ziese, M. (2013) A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. *Earth System Science Data*, 5(1), 71–99. https://doi.org/10.5194/essd-5-71-2013.

Berger, S., Bliefernicht, J., Linstädter, A., Canak, K., Guug, S., Heinzeller, D., Hingerl, L., Mauder, M., Neidl, F., Quansah, E., Salack, S., Steinbrecher, R. and Kunstmann, H. (2019) The impact of rain events on CO2 emissions from contrasting land use systems in semi-arid West African savannas. *Science of the Total Environment*, 647, 1478–1489. https://doi.org/10.1016/j.scitotenv.2018.07.397.

Bickici Arikan, B. and Kahya, E. (2019) Homogeneity revisited: analysis of updated precipitation series in Turkey. *Theoretical and Applied Climatology*, 135(1), 211–220. https://doi.org/10.1007/s00704-018-2368-x.

Bliefernicht, J., Berger, S., Salack, S., Guug, S., Hingerl, L., Heinzeller, D., Mauder, M., Steinbrecher, R., Steup, G., Bossa, A.Y., Waongo, M., Quansah, E., Balogun, A.A., Yira, Y., Arnault, J., Wagner, S., Klein, C., Gessner, U., Knauer, K., Straub, A., Schönrock, R., Kunkel, R., Okogbue, E.C., Rogmann, A., Neidl, F., Jahn, C., Diekkrüger, B., Aduna, A., Barry, B. and Kunstmann, H. (2018) The WASCAL hydrometeorological observatory in the Sudan savanna of Burkina Faso and Ghana. *Vadose Zone Journal*, 17(1), 1–20. https://doi.org/10.2136/vzj2018.03.0065.

Bliefernicht, J., Waongo, M., Salack, S., Seidel, J., Laux, P. and Kunstmann, H. (2019) Quality and value of seasonal precipitation forecasts issued by the West African regional climate outlook forum. *Journal of Applied Meteorology and Climatology*, 58(3), 621–642. https://doi.org/10.1175/JAMC-D-18-0066.1.

Bliefernicht, J., Bárdossy, A. and Ebert, C. (2008) Stochastic simulation of hourly precipitation fields for extreme events on the rivers Freiberger Mulde, Oberer Main and Fränkische Saale. *Hydrologie und Wasserbewirtschaftung*, 52(4), 1–20.

Boulanger, J.P., Aizpuru, J., Leggieri, L. and Marino, M. (2010) A procedure for automated quality control and homogenization of historical daily temperature and precipitation data (APACH): part 1: quality control and application to the argentine weather service stations. *Climatic Change*, 98(3), 471–491. https://doi.org/10.1007/s10584-009-9741-9.

Buishand, T.A. (1982) Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, 58(1), 11–27. https://doi.org/10.1016/0022-1694(82)90066-X.

Campozano, L., Sánchez, E., Avilés, Á. (2014) Evaluación de métodos de relleno para series temporales de precipitación y temperatura diarias: el caso de los Andes ecuatorianos. *Maskana*, 5(1), 99–115. https://doi.org/10.18537/mskn.05.01.07.

Ceccherini, G., Russo, S., Ameztoy, I., Marchese, A.F. and Carmona-Moreno, C. (2017) Heat waves in Africa 1981–2015, observations and reanalysis. *Natural Hazards and Earth System Sciences*, 17(1), 115–125. https://doi.org/10.5194/nhess-17-115-2017.

Ciach, G.J. and Krajewski, W.F. (2006) Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma. *Advances in Water Resources*, 29(10), 1450–1463. https://doi.org/10.1016/j.advwatres.2005.11.003.

Coll, J., Domonkos, P., Guijarro, J., Curley, M., Rustemeier, E., Aguilar, E., Walsh, S. and Sweeney, J. (2020) Application of homogenization methods for Ireland's monthly precipitation records: comparison of break detection results. *International Journal of Climatology*, 40(14), 6169–6188. https://doi.org/10.1002/joc.6575.

Costa, A.C. and Soares, A. (2009) Homogenization of climate data: review and new perspectives using geostatistics. *Mathematical Geosciences*, 41(3), 291–305. https://doi.org/10.1007/s11004-008-9203-3.

Dai, A. (2006) Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, 19(18), 4605–4630. https://doi.org/10.1175/JCLI3884.1.

Delvaux, C., Ingels, R., Vrábeĺ, V., Journée, M. and Bertrand, C. (2019) Quality control and homogenization of the Belgian historical temperature data. *International Journal of Climatology*, 39(1), 157–171. https://doi.org/10.1002/joc.5792.

Diallo, I., Sylla, M.B., Giorgi, F., Gaye, A.T. and Camara, M. (2012) Multimodel GCM-RCM ensemble-based projections of temperature and precipitation over West Africa for the early 21st century. *International Journal of Geophysics*, 2012, 1–19. https://doi.org/10.1155/2012/972896.

Dieng, D., Smiatek, G., Bliefernicht, J., Heinzeller, D., Sarr, A., Gaye, A.T. and Kunstmann, H. (2017) Evaluation of the COSMO-CLM high-resolution climate simulations over West Africa. *Journal of Geophysical Research*, 122(3), 1437–1455. https://doi.org/10.1002/2016JD025457.

Dosio, A., Panitz, H.J., Schubert-Frisius, M. and Lüthi, D. (2015) Dynamical downscaling of CMIP5 global circulation models over CORDEX-Africa with COSMO-CLM: evaluation over the present climate and analysis of the added value. *Climate Dynamics*, 44(9–10), 2637–2661. https://doi.org/10.1007/s00382-014-2262-x.

Durre, I., Menne, M.J., Gleason, B.E., Houston, T.G. and Vose, R.S. (2010) Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, 49(8), 1615–1633. https://doi.org/10.1175/2010JAMC2375.1.

Durre, I., Menne, M.J. and Vose, R.S. (2008) Strategies for evaluating quality assurance procedures. *Journal of Applied Meteorology and Climatology*, 47(6), 1785–1791. https://doi.org/10.1175/2007JAMC1706.1.

Engel, T., Fink, A.H., Knippertz, P., Pante, G. and Bliefernicht, J. (2017) Extreme precipitation in the west African cities of Dakar and Ouagadougou: atmospheric dynamics and implications for flood risk assessments. *Journal of Hydrometeorology*, 18(11), 2937–2957. https://doi.org/10.1175/JHM-D-16-0218.1.

Ermert, V., Fink, A.H., Jones, A.E. and Morse, A.P. (2011) Development of a new version of the Liverpool malaria model. II. Calibration and validation for West Africa. *Malaria Journal*, 10, 1–19. https://doi.org/10.1186/1475-2875-10-62.

Feng, S., Hu, Q. and Qian, W. (2004) Quality control of daily meteorological data in China, 1951–2000: a new dataset. *International Journal of Climatology*, 24(7), 853–870. https://doi.org/10.1002/joc.1047.

Fleury, L., Boichard, J.L., Brissebrat, G., Cloché, S., Eymard, L., Mastrorillo, L., Moulaye, O., Ramage, K., Asencio, N., Coppeaux, J., Devic, M.P., Favot, F., Ginoux, K., Lafore, J.P., Polcher, J., Redelsperger, J.L., Roussot, O. and Tytéca, M. (2011) AMMA information system: an efficient cross-disciplinary tool and a legacy for forthcoming projects. *Atmospheric Science Letters*, 12(1), 149–154. https://doi.org/10.1002/asl.303.

Galle, S., Grippa, M., Peugeot, C., Moussa, I.B., Cappelaere, B., Demarty, J., Mougin, E., Panthou, G., Adjomayi, P., Agbossou, E.K., Ba, A., Boucher, M., Cohard, J.M., Descloitres, M., Descroix, L., Diawara, M., Dossou, M., Favreau, G., Gangneron, F., Gosset, M., Hector, B., Hiernaux, P., Issoufou, B.A., Kergoat, L., Lawin, E., Lebel, T., Legchenko, A., Abdou, M.M., Malam-Issa, O., Mamadou, O., Nazoumou, Y., Pellarin, T., Quantin, G., Sambou, B., Seghieri, J., Séguis, L., Vandervaere, J.P., Vischel, T., Vouillamoz, J.M., Zannou, A., Afouda, S., Alhassane, A., Arjounin, M., Barral, H., Biron, R., Cazenave, F., Chaffard, V., Chazarin, J.P., Guyard, H., Koné, A., Mainassara, I., Mamane, A., Oi, M., Ouani, T., Soumaguel, N., Wubda, M., Ago, E.E., Alle, I.C., Allies, A., Arpin-Pont, F., Awessou, B., Cassé, C., Charvet, G., Dardel, C., Depeyre, A., Diallo, F.B., Do, T., Fatras, C., Frappart, F., Gal, L., Gascon, T., Gibon, F., Guiro, I., Ingatan, A., Kempf, J., Kotchoni, D.O.V., Lawson, F. M.A., Leauthaud, C., Louvet, S., Mason, E., Nguyen, C.C., Perrimond, B., Pierre, C., Richard, A., Robert, E., Román-Cascón, C., Velluet, C. and Wilcox, C. (2018) AMMA-CATCH, a critical zone observatory in West Africa monitoring a region in transition. *Vadose Zone Journal*, 17(1), 180062. https://doi.org/10.2136/vzj2018.03.0062.

Van De Giesen, N., Kunstmann, H., Jung, G., Liebe, J., Andreini, M. and Vlek, P.L.G. (2002) 'The GLOWA Volta project: integrated assessment of feedback mechanisms between climate, landuse, and hydrology', pp. 151–170. doi: https://doi.org/10.1007/0-306-47983-4_9.

van de Giesen, N., Hut, R. and Selker, J. (2014) The trans-African hydro-meteorological observatory (TAHMO). *Wiley Interdisciplinary Reviews: Water*, 1(4), 341–348. https://doi.org/10.1002/wat2.1034.

Goudenhoofdt, E. and Delobbe, L. (2009) Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrology and Earth System Sciences*, 13(2), 195–203. https://doi.org/10.5194/hess-13-195-2009.

Heinzeller, D., Dieng, D., Smiatek, G., Olusegun, C., Klein, C., Hamann, I., Salack, S., Bliefernicht, J. and Kunstmann, H. (2018) The WASCAL high-resolution regional climate simulation ensemble for West Africa: concept, dissemination and assessment. *Earth System Science Data*, 10(2), 815–835. https://doi.org/10.5194/essd-10-815-2018.

Hogan, R. and Mason, I. (2012) Deterministic forecasts of binary events. In: I.T. Jolliffe, & D.B. Stephenson (Eds.), *Forecast verification: a practitioner's guide in atmospheric science* (pp. 31–59). John Wiley & Sons. https://doi.org/10.1002/9781119960003.ch3.

Huffman, G.J., Adler, R.F., Bolvin, D.T. and Gu, G. (2009) Improving the global precipitation record: GPCP version 2.1. *Geophysical Research Letters*, 36(17), 1–5. https://doi.org/10.1029/2009GL040000.

Israelsson, J., Black, E., Neves, C., Torgbor, F.F., Greatrex, H., Tanu, M. and Lamptey, P.N.L. (2020) The spatial correlation structure of rainfall at the local scale over southern Ghana. *Journal of Hydrology: Regional Studies*, 31, 100720. https://doi.org/10.1016/j.ejrh.2020.100720.

Jung, G. and Kunstmann, H. (2007) High-resolution regional climate modeling for the Volta region of West Africa. *Journal of Geophysical Research Atmospheres*, 112(23), 1–17. https://doi.org/10.1029/2006JD007951.

Laux, P., Wagner, S., Wagner, A., Jacobeit, J., Bárdossy, A. and Kunstmann, H. (2009) Modelling daily precipitation features in the Volta Basin of West Africa. *International Journal of Climatology*, 29(7), 937–954. https://doi.org/10.1002/joc.1852.

Laux, P., Kunstmann, H. and Bárdossy, A. (2008) Predicting the regional onset of the rainy season in West Africa. *International Journal of Climatology*, 28(3), 329–342. https://doi.org/10.1002/joc.1542.

Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. and Rennie, J. (2011) An overview of the global historical climatology network monthly mean temperature data set, version 3. *Journal of Geophysical Research Atmospheres*, 116(19), 1–18. https://doi.org/10.1029/2011JD016187.

Lebel, T., Parker, D.J., Flamant, C., Bourlès, B., Marticorena, B., Mougin, E., Peugeot, C., Diedhiou, A., Haywood, J.M., Ngamini, J.B., Polcher, J., Redelsperger, J.L. and Thorncroft, C. D. (2010) The AMMA field campaigns: multiscale and multidisciplinary observations in the West African region. *Quarterly Journal of the Royal Meteorological Society*, 136(SUPPL. 1), 8–33. https://doi.org/10.1002/qj.486.

Lorenz, C. and Kunstmann, H. (2012) The hydrological cycle in three state-of-the-art reanalyses: intercomparison and performance analysis. *Journal of Hydrometeorology*, 13(5), 1397–1420. https://doi.org/10.1175/JHM-D-11-088.1.

Lorenz, M., Bliefernicht, J., Haese, B. and Kunstmann, H. (2018) Copula-based downscaling of daily precipitation fields. *Hydrological Processes*, 32(23), 3479–3494. https://doi.org/10.1002/hyp.13271.

Lott, J.N. (2004) The quality control of the integrated surface hourly database. *Conference proceedings (extended abstract), 84th AMS Annual Meeting*, pp. 5039–5045. Seattle, WA. https://ams.confex.com/ams/pdfpapers/71929.pdf. Accessed February 25, 2021.

Maranan, M., Fink, A.H., Knippertz, P., Amekudzi, L.K., Atiah, W. A. and Stengel, M. (2020) A process-based validation of GPM IMERG and its sources using a mesoscale rain gauge network in the west African forest zone. *Journal of Hydrometeorology*, 21(4), 729–749. https://doi.org/10.1175/JHM-D-19-0257.1.

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E. and 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.

Menne, M.J., Williams, C.N., Gleason, B.E., Rennie, J.J. and Lawrimore, J.H. (2018) The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, 31(24), 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1.

Moron, V., Robertson, A.W. and Boer, R. (2009) Spatial coherence and seasonal predictability of monsoon onset over Indonesia. *Journal of Climate*, 22(3), 840–850. https://doi.org/10.1175/2008JCLI2435.1.

Moron, V., Robertson, A.W. and Qian, J.H. (2010) Local versus regional-scale characteristics of monsoon onset and post-onset rainfall over Indonesia. *Climate Dynamics*, 34(2), 281–299. https://doi.org/10.1007/s00382-009-0547-2.

Neumann, R., Jung, G., Laux, P. and Kunstmann, H. (2007) Climate trends of temperature, precipitation and river discharge in the Volta basin of West Africa. *International Journal of River Basin Management*, 5(1), 17–30. https://doi.org/10.1080/15715124.2007.9635302.

Nicholson, S.E. (2001) A semi-quantitative, regional precipitation data set for studying African climates of the nineteenth century, part I. Overview of the data set. *Climatic Change*, 50(3), 317–353. https://doi.org/10.1023/A:1010674724320.

Nicholson, S.E. (2013) The West African Sahel: A review of recent studies on the rainfall regime and its interannual variability. *International Scholarly Research Notices*, 2013, 453521. https://doi.org/10.1155/2013/453521

Nicholson, S.E., Some, B., McCollum, J., Nelkin, E., Klotter, D., Berte, Y., Diallo, B.M., Gaye, I., Kpabeba, G., Ndiaye, O., Noukpozounkou, J.N., Tanu, M.M., Thiam, A., Toure, A.A. and Traore, A.K. (2003) Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part I: validation of GPCC rainfall product and pre-TRMM satellite and blended products. *Journal of Applied Meteorology*, 42(10), 1337–1354. https://doi.org/10.1175/1520-0450(2003)042<1337:VOTAOR>2.0.CO;2.

Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensen, O.B., Déqué, M., Fernandez, J., Hänsler, A., van Meijgaard, E., Samuelsson, P., Sylla, M.B. and Sushama, L. (2012) Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *Journal of Climate*, 25(18), 6057–6078. https://doi.org/10.1175/JCLI-D-11-00375.1.

NOAA (2020a) 19th century African instrumental and documentary precipitation data. Available at: https://www.ncdc.noaa.gov/paleo-search/study/12201 [Accessed 14 July 2020].

NOAA (2020b) Global historical climate network daily—description. Available at: https://www.ncdc.noaa.gov/ghcn-daily-description [Accessed 14 July 2020].

NOAA (2020c) Global surface summary of the day—GSOD. Available at: https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00516 [Accessed 13 July 2020].

O'Gorman, P.A. (2015) Precipitation extremes under climate change. *Current Climate Change Reports*, 1(2), 49–59. https://doi.org/10.1007/s40641-015-0009-3.

Panitz, H.J., Dosio, A., Büchner, M., Lüthi, D. and Keuler, K. (2014) COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-interim driven simulations at 0.44° and 0.22° resolution. *Climate Dynamics*, 42(11–12), 3015–3038. https://doi.org/10.1007/s00382-013-1834-5.

Peterson, T.C., Vose, R., Schmoyer, R. and Razuvaëv, V. (1998a) Global historical climatology network (GHCN) quality control of monthly temperature data. *International Journal of Climatology*, 18(11), 1169–1179. https://doi.org/10.1002/(SICI)1097-0088(199809)18:11<1169::AID-JOC309>3.0.CO;2-U.

Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. and Parker, D. (1998b) Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, 18(13), 1493–1517. https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Di Piazza, A., Lo Conti, F., Noto, L.V., Viola, F. and La Loggia, G. (2011) Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation*, 13(3), 396–408. https://doi.org/10.1016/j.jag.2011.01.005.

Rauch, M., Bliefernicht, J., Laux, P., Salack, S., Waongo, M. and Kunstmann, H. (2019) Seasonal forecasting of the onset of the rainy season in West Africa. *Atmosphere*, 10(9), 1–21. https://doi.org/10.3390/atmos10090528.

Redelsperger, J.L., Thorncroft, C.D., Diedhiou, A., Lebel, T., Parker, D.J. and Polcher, J. (2006) African monsoon multidisciplinary analysis: an international research project and field campaign. *Bulletin of the American Meteorological Society*, 87(12), 1739–1746. https://doi.org/10.1175/BAMS-87-12-1739.

Ribeiro, S., Caineta, J., Costa, A.C. and Henriques, R. (2017) Gsimcli: a geostatistical procedure for the homogenisation of climatic time series. *International Journal of Climatology*, 37(8), 3452–3467. https://doi.org/10.1002/joc.4929.

Ribeiro, S., Caineta, J. and Costa, A.C. (2016) Review and discussion of homogenisation methods for climate data. *Physics and Chemistry of the Earth, Parts A/B/C*, 94, 167–179. https://doi.org/10.1016/j.pce.2015.08.007.

Salack, S., Saley, I.A., Lawson, N.Z., Zabré, I. and Daku, E.K. (2018b) Scales for rating heavy rainfall events in the west African Sahel. *Weather and Climate Extremes*, 21, 36–42. https://doi.org/10.1016/j.wace.2018.05.004.

Salack, S., Bossa, A., Bliefernicht, J., Berger, S., Yira, Y., Sanoussi, K.A., Guug, S., Heinzeller, D., Avocanh, A.S., Hamadou, B., Meda, S., Diallo, B.A., Bado, I.B., Saley, I.A., Daku, E.K., Lawson, N.Z., Ganaba, A., Sanfo, S., Hien, K., Aduna, A., Steup, G., Diekkrüger, B., Waongo, M., Rogmann, A., Kunkel, R., Lamers, J.P.A., Sylla, M.B., Kunstmann, H., Barry, B., Sedogo, L.G., Jaminon, C., Vlek, P., Adegoke, J. and Savadogo, M. (2019) Designing transnational hydroclimatological observation networks and data sharing policies in West Africa. *Data Science Journal*, 18(1), 1–15. https://doi.org/10.5334/dsj-2019-033.

Salack, S., Saley, I.A. and Bliefernicht, J. (2018a) Observed data of extreme rainfall events over the west African Sahel. *Data in Brief*, 20, 1274–1278. https://doi.org/10.1016/j.dib.2018.09.001.

Sall, S.M., Viltard, A. and Sauvageot, H. (2007) Rainfall distribution over the Fouta Djallon — Guinea. *Atmospheric Research*, 86(2), 149–161. https://doi.org/10.1016/j.atmosres.2007.03.008.

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M. and Stender, P. (2014) Global gridded precipitation over land: a description of the new GPCC first guess daily product. *Earth System Science Data*, 6(1), 49–60. https://doi.org/10.5194/essd-6-49-2014.

Scherrer, S.C., Frei, C., Croci-Maspoli, M., van Geijtenbeek, D., Hotz, C. and Appenzeller, C. (2011) Operational quality control of daily precipitation using spatio-climatological plausibility testing. *Meteorologische Zeitschrift*, 20(4), 397–407. https://doi.org/10.1127/0941-2948/2011/0236.

Schroeer, K., Kirchengast, G. and Sungmin, O. (2018) Strong dependence of extreme convective precipitation intensities on gauge network density. *Geophysical Research Letters*, 45(16), 8253–8263. https://doi.org/10.1029/2018GL077994.

Siegmund, J., Bliefernicht, J., Laux, P. and Kunstmann, H. (2015) Toward a seasonal precipitation prediction system for West Africa: performance of CFSv2 and high-resolution dynamical downscaling. *Journal of Geophysical Research*, 120(15), 7316–7339. https://doi.org/10.1002/2014JD022692.

Skrynyk, O., Aguilar, E., Skrynyk, O., Sidenko, V., Boichuk, D. and Osadchyi, V. (2019) Quality control and homogenization of monthly extreme air temperature of Ukraine. *International Journal of Climatology*, 39(4), 2071–2079. https://doi.org/10.1002/joc.5934.

Smith, A., Lott, N. and Vose, R. (2011) The integrated surface database: recent developments and partnerships. *Bulletin of the American Meteorological Society*, 92(6), 704–708. https://doi.org/10.1175/2011BAMS3015.1.

TAHMO (2020) *Trans-African hydro-meteorological observatory*. Available at: https://tahmo.org/ [Accessed 14 July 2020].

Tapiador, F.J., Navarro, A., Levizzani, V., García-Ortega, E., Huffman, G.J., Kidd, C., Kucera, P.A., Kummerow, C.D., Masunaga, H., Petersen, W.A., Roca, R., Sánchez, J.L., Tao, W. K. and Turk, F.J. (2017) Global precipitation measurements for validating climate models. *Atmospheric Research*, 197, 1–20. https://doi.org/10.1016/j.atmosres.2017.06.021.

Tobin, C., Nicotina, L., Parlange, M.B., Berne, A. and Rinaldo, A. (2011) Improved interpolation of meteorological forcings for hydrologic applications in a Swiss alpine region. *Journal of Hydrology*, 401(1), 77–89. https://doi.org/10.1016/j.jhydrol.2011.02.010.

Trenberth, K.E. (2011) Changes in precipitation with climate change. *Climate Research*, 47(1–2), 123–138. https://doi.org/10.3354/cr00953.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., García-Vera, M.A. and Stepanek, P. (2010) A complete daily precipitation database for Northeast Spain: reconstruction, quality control, and homogeneity. *International Journal of Climatology*, 30(8), 1146–1163. https://doi.org/10.1002/joc.1850.

Vollmert, P. and Fink, A.H. (2003) 'Ghana Dry Zone' und 'Dahomey Gap': Ursachen für eine Niederschlagsanomalie im tropischen Westafrika. *Die Erde*, 134(4), 375–393.

Wan, Y., Chen, J., Xie, P., Xu, C.Y. and Li, D. (2021) Evaluation of climate model simulations in representing the precipitation non-stationarity by considering observational uncertainties. *International Journal of Climatology*, 41(3), 1952–1969. https://doi.org/10.1002/joc.6940.

WASCAL (2020) Daily Observed Gridded Precipitation for Burkina Faso, Ghana, Benin and Togo from 1970 to 2009 in 0.1° x 0.1° Resolution. Available at: https://wascal-dataportal.org/2.0/?q=dataset/daily-observed-gridded-precipitation-burkina-faso-ghana-benin-and-togo-1970-2009-01°-x-01° [Accessed 20 November 2020].

WMO (2021) World Meteorological Organization global weather & climate extremes archive. Available at: https://wmo.asu.edu/content/world-meteorological-organization-global-weather-climate-extremes-archive [Accessed 20 September 2021].

Wong, C.L., Venneker, R., Jamil, A.B.M. and Uhlenbrook, S. (2011) Development of a gridded daily hydrometeorological data set for peninsular Malaysia. *Hydrological Processes*, 25(7), 1009–1020. https://doi.org/10.1002/hyp.7654.

Yatagai, A., Arakawa, O., Kamiguchi, K., Kawamoto, H., Nodzu, M. I. and Hamada, A. (2009) A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Scientific Online Letters on the Atmosphere*, 5(1), 137–140. https://doi.org/10.2151/sola.2009-035.

Zhang, W., Villarini, G. and Wehner, M. (2019) Contrasting the responses of extreme precipitation to changes in surface air and dew point temperatures. *Climatic Change*, 154(1), 257–271. https://doi.org/10.1007/s10584-019-02415-8.

SUPPORTING INFORMATION

Additional supporting information is found in the online version of the article at the publisher's website.

How to cite this article: Bliefernicht, J., Salack, S., Waongo, M., Annor, T., Laux, P., & Kunstmann, H. (2022). Towards a historical precipitation database for West Africa: Overview, quality control and harmonization. *International Journal of Climatology*, *42*(7), 4001–4023. https://doi.org/10.1002/joc.7467