A New 60-Year 1940/1999 Monthly-Gridded Rainfall Data Set for Africa

Claudine Dieulin [1,*], Gil Mahé [1,*], Jean-Emmanuel Paturel [1], Soundouss Ejjiyar [2], Yves Tramblay [1], Nathalie Rouché [1] and Bouabid EL Mansouri [2]

[1] HydroSciences Montpellier, Université de Montpellier–CC 57, 163 rue Auguste Broussonnet, 34090 Montpellier, France; jean-emmanuel.paturel@ird.fr (J.-E.P.); yves.tramblay@ird.fr (Y.T.); nathalie.rouche@ird.fr (N.R.)

[2] Faculty of Sciences, Hydro-Informatic Section, Campus Maamora, Ibn Tofail University, Kenitra BP 133–14000, Morocco; ejjiyar-soundouss@hotmail.com (S.E.); b_elmansouri@yahoo.fr (B.E.M.)

* Correspondence: claudine.dieulin@umontpellier.fr (C.D.); gil.mahe@ird.fr (G.M.); Tel.: +33-467-149-025 (C.D.); Fax.: +33-467-144-774 (C.D. & G.M.)

check for updates

Abstract: The African continent has a very low density of rain gauge stations, and long time-series for recent years are often limited and poorly available. In the context of global change, it is very important to be able to characterize the spatio-temporal variability of past rainfall, on the basis of datasets issued from observations, to correctly validate simulations. The quality of the rainfall data is for instance of very high importance to improve the efficiency of the hydrological modeling, through calibration/validation experiments. The HydroSciences Montpellier Laboratory (HSM) has a long experience in collecting and managing hydro-climatological data. Thus, HSM had initiated a program to elaborate a reference dataset, in order to build monthly rainfall grids over the African continent, over a period of 60 years (1940/1999). The large quantity of data collected (about 7000 measurement points were used in this project) allowed for interpolation using only observed data, with no statistical use of a reference period. Compared to other databases that are used to build the grids of the Global Historical Climatology Network (GHCN) or the Climatic Research Unit of University of East Anglia, UK (CRU), the number of available observational stations was significantly much higher, including the end of the century when the number of measurement stations dropped dramatically, everywhere. Inverse distance weighed (IDW) was the chosen method to build the 720 monthly grids and a mean annual grid, from rain gauges. The mean annual grid was compared to the CRU grid. The grids were significantly different in many places, especially in North Africa, Sahel, the horn of Africa, and the South Western coast of Africa, with HSM_SIEREM data (database HydroSciences Montpellier_Système d'Information Environnementales pour les Ressources en Eau et leur Modélisation) being closer to the observed rain gauge values. The quality of the grids computed was checked, following two approaches—cross-validation of the two interpolation methods, ordinary kriging and inverse distance weighting, which gave a comparable reliability, with regards to the observed data, long time-series analysis, and analysis of long-term signals over the continent, compared to previous studies. The statistical tests, computed on the observed and gridded data, detected a rupture in the rainfall regime around 1979/1980, on the scale of the whole continent; this was congruent with the results in the literature. At the monthly time-scale, the most widely observed signal over the period of 1940/1999, was a significant decrease of the austral rainy season between March and May, which has not earlier been well-documented. Thus, this would lead to a further detailed climatological study from this HSM_SIEREM database.

Keywords: Africa; rainfall; monthly grids; database; inverse distance weighted

1. Introduction

On a global scale, the first climate observations began during the second part of the 19th century. Among these climate variables, rainfall data sets have been the most complete, since the beginning of the 1950s [1]. In Africa, the most well-known change in rainfall regime occurred in Sahel, but all of Western and Central Africa experienced an abrupt decrease in mean annual rainfall, at the end of the 1960s or towards the beginning of the 1970s [2–4]. To study rainfall changes at the scale of the whole African continent, rainfall grids can be downloaded from several institutions, but it is well-known that the African continent is less documented than other parts of the world [5,6]. This has led to some discrepancies between rainfall calculated from different databases, as has been shown for Western Africa, by Mahé et al. [7]. Among the available data sets, Climatic Research Unit of University of East Anglia, UK (CRU) [8,9] is used very often and seems to be the most accurate, compared to other sources. However, for instance, for Western and Central Africa, rainfall data are not very well-documented [10,11].

The first objective of the present work is to elaborate monthly grids of the best possible quality, over the African continent. For this, it was necessary to build the most exhaustive rainfall database to help improve understanding of the climatic processes and other applications, such as hydrological modeling. Hydrologists at the Institut de Recherche pour le Développement (IRD) or the French National Research Institute for Sustainable Development, previously ORSTOM (Office de Recherche Scientifique Outre-Mer), had historically participated in the acquisition of climatological data and built up a digital database, as early as the end of the 1960s. However, by the end of the 1980s, the IRD stopped collecting and managing rainfall data for Western and Central Africa. The researchers of the HSM decided to valorize this unique set of data for Africa, first, in the search of additional data to cover the whole continent, and to update the database up to 1999. This project was an opportunity to create a reference database, called HSM_SIEREM database, which is still managed by the HSM. The HSM_SIEREM [12] database has been used for sub-continental studies of climate change, like those by Paturel et al. [10,11], many regional studies of river-runoff relationships [13,14], and studies of the impact of climate change on hydrological regimes in Western Africa [15]. These data were first used at a continental scale, for a study of rainfall-runoff variability over the past decades [16].

In this article, we have presented, for the first time, the methods through which the data were selected and evaluated, the process leading to the creation of the monthly-gridded dataset over Africa (at a 0.5 × 0.5 degree resolution), and a first assessment of the content of this new database. In the second part, we have compared this new HSM_SIEREM grid, with the CRU grid and have pointed out the similarities and the differences to help improve the future uses of both bases. In the third part, we have presented a first study of rainfall variability, over the 1940/1999 time period, for the whole continent, to check the general quality of the dataset, by searching if the main rainfall signals were visible in large-scale regional averaged time-series. This first study will be followed by further detailed studies on the spatio-temporal variability of rainfall over the African continent. The monthly grids we created have been provided, for free, on the website of the HydroSciences Montpellier SIEREM project (http://www.hydrosciences.org/spip.php?article1387), as zipped ASCII grids and in the NetCDF format.

2. Materials and Methods

2.1. History of the HSM_SIEREM Database

Created in 1943, ORSTOM/IRD's mission was to promote scientific cooperation and education in Western and Central Africa. In this framework, hydrologists installed and managed most of the hydro-meteorological stations in French speaking countries—the former colonies. The records were noted in paper booklets and a long process of data entry began, as early as 1967, on punched cards, first, and were then transferred onto magnetic tapes of the CNRS (Centre National de la Recherche Scientifique) computers in Orsay/Paris. This transfer was taken into account to run a first control on

the data quality check for duplicate cards, removal of data for non-existent days (such as, 31st of April and 30th of February), before building up the first database. These cards were individually controlled and were modified, kept or removed. During this process, every measurement point was linked to only one data set. Due to the state of technology at the time the data was collected, some measurement stations had no geographical referencing, and it was added later, when possible.

A tripartite agreement between the Comité Interafricain d'Etudes Hydrauliques (CIEH) or the Interafrican Hydraulic studies Committee, the Agence pour la SECurité de la Navigation Aérienne (ASECNA) or Agency for the Safety of Aerial Navigation in Africa and Madagascar, and the ORSTOM, began in 1973, for the collection of data from thirteen countries of Western and Central Africa (Benin, Burkina Faso, Cameroon, Congo, Ivory Coast, Gabon, Mali, Mauritania, Niger, Central Africa, Senegal, Chad, and Togo) and ended in 1989, with the edition of the two rainfall data books, for each country, covering the years up to 1965 and 1966, to 1980 [5]. This agreement allowed to update the database at a daily time-step, up to 1980. Figure 1 [17] shows the benefits of this agreement in the evolution of the rainfall database. From this set of data, a map of the mean annual rainfall over Western and Central Africa was drawn by L'Hote & Mahé [18].
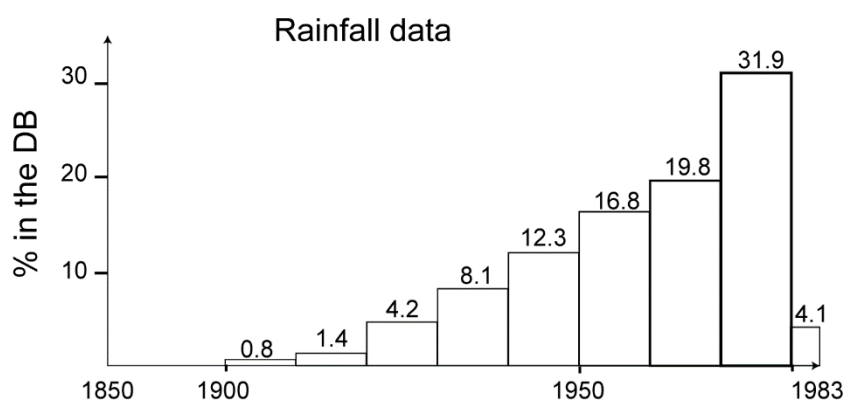

Figure 1. Distribution of data in the ORSTOM rainfall database from the origin of the observation to 1983 [17].

The Water Assessment [19] financed by the World Bank, the UNDP (United Nations Development Program), the African Bank of Development, and the French ministry of Cooperation, initiated a program to collect monthly rainfall data over the whole African continent, for the period of 1981/1990. This program was carried out by Mott MacDonald International, the BCEOM (Bureau Central d'études pour les Equipements Outre-Mer) and the SOGREAH (Société Grenobloise d'Etudes et d'Applications Hydrauliques) offices for the Western and Central part of Africa, and the hydrologists of ORSTOM, who participated to the gathering of data in the Sub-Saharan area.

Up to 1999, the IRD Hydrology Laboratory in Montpellier gathered and managed all the climatological data archived in every country where ORSTOM hydrologists were present, mainly in Africa, as well as the rainfall data from all past programs. Since 2000, the HSM_SIEREM database was only used and enriched by the teams of HSM in Montpellier, with no institutional mandate for commitment to the management or the dissemination of this base. For Eastern and Southern Africa, HSM gathered data through its commitment in the IHP (International Hydrological Program) of UNESCO and mainly in the FRIEND programs (Flow Regimes from International Experimental and Network Data), as HSM managed the implementation and hosting of the databases of the different African FRIEND groups. Many other international programs that the HSM collaborated on in Africa, allowed the IRD hydrologists access to more recent data.

The CRU is specifically acknowledged, as they had developed a collaboration with the hydrologists of HSM which let them access the rainfall data in the areas where HSM did not have any access. However, both CRU and HSM databases where kept independent.

The HSM_SIEREM database, thus, contains rainfall data for more than 6000 stations across Africa.

2.1.1. The Reference Database

The development of a continuous, quality-checked and reliable long-term database of monthly rainfall across Africa, was a key component of this project. The database contained numerous records, due to the fact that any incoming rainfall datum was always stored. According to Rouché et al. [20], the result was that, for some stations, up to ten different data sets were stored (collected by different people, at different time-steps, data corrected for specific purposes, etc.). Therefore, it was essential to eliminate redundant sets and choose the better series.

Constitution of the Reference Data Set

The dataset was built for every African country. Records with missing geographical coordinates were removed, daily, and ten-day data were aggregated at the monthly time-step. Then, if several series for the same station still remained in the database, a "quality label" was given to the series (depending on the origin of the data, the reliability of the provider, and the length of the observation period), to select which one was to be kept. When the retained series had missing periods, it was filled with data from a different station of the same measurement point, when available.

The aim of this process was to keep only one station per measurement point. The graph in Figure 2 shows the number of sets before and after the creation of the reference database, for every country.
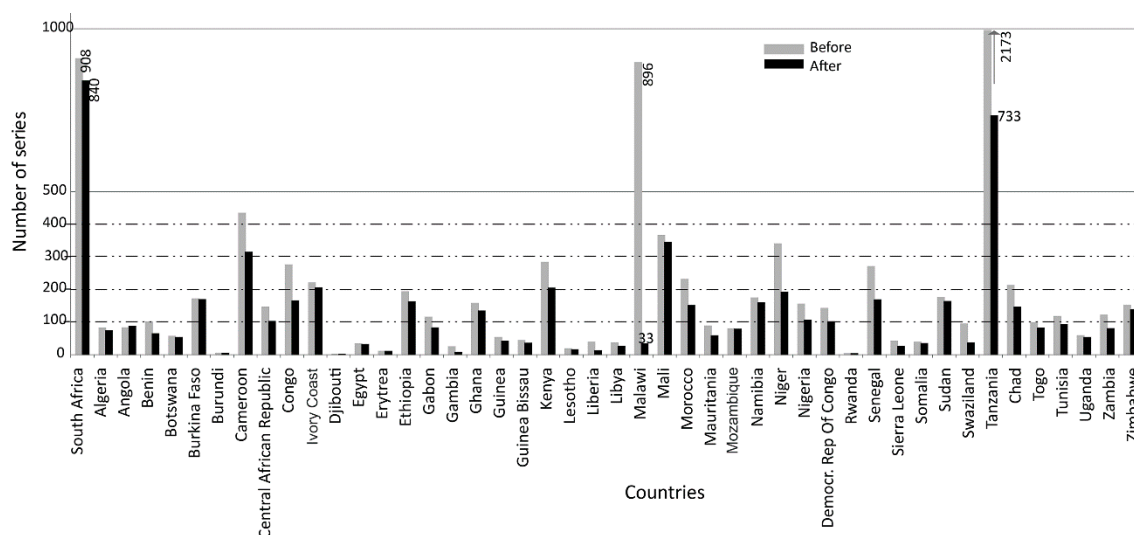

Figure 2. Result of the selection process between redundant time-series—in grey, the number of series at the beginning, and in black, the final number of reference series for the period 1900/1999, where each series corresponds to a unique station. Note: Due to insufficient data availability, Madagascar has not been processed in this project.

Criticism of the Reference Data Set

For every country, patterns were established to define:

- the period of the rainy and dry seasons; and
- a range of monthly rainfall values (minimum and maximum) that were not to be crossed.

During this process, a series of automatic tests were run, and a flag was inserted if any value:

- exceeded the previously defined values for the country;
- was identical to the value of the same month of the previous or next year at the station; or
- was identical to the value of the previous month of the same year.

The flagged months were printed and checked by the HSM hydrologists, who determined if the flag led to the removal of the data or of part of the time-series. This step was the longest but the most

important for the quality of the database. Thus, the time-series of each station was both automatically and manually checked, to reduce the risk of erroneous data in the reference database.

2.1.2. The Period

The HSM_SIEREM base was exported in the form of one file per country and per year, over the period 1900/1999. The first goal was to create monthly grids for the whole of the 20th century. However, when plotting the measurement points, the maps showed that before 1940 (Figure 3), the data were too sparse and too heterogeneously distributed to be spatialized over the whole continent. The decision was then made to limit the interpolation process to the period 1940/1999. Contrary to CRU, we decided not to create a reference period like the ones the CRU used to establish patterns to create interpolated rainfall measurement points [9]—even for the years where the distribution of data was heterogeneous or density of stations was low—enabling them to start the interpolated time-series in 1900. This limited our possibility to extend the grid to a period prior to 1940. However, this is also what makes a fundamental difference between the HSM_SIEREM and the CRU grids, i.e., in the HSM_SIEREM database, each month, the grid was calculated only from the available observed data.
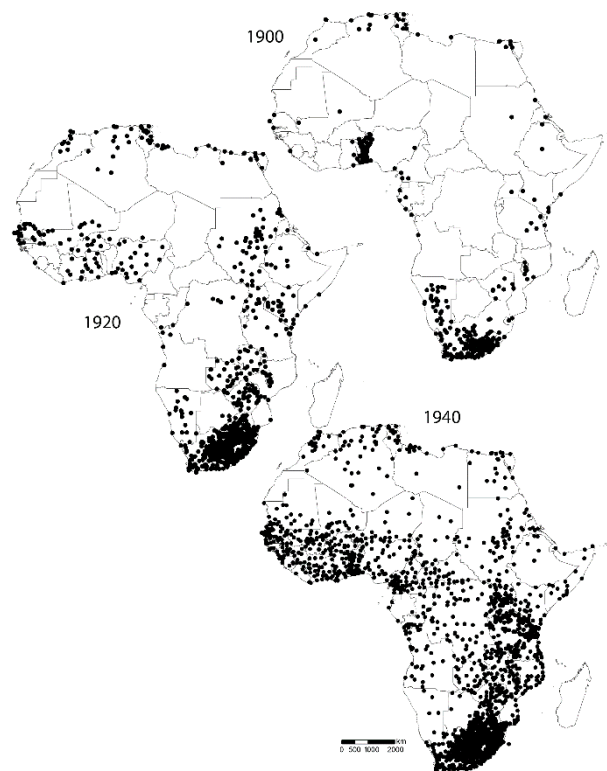

Figure 3. Evolution of the number of measurement stations in 1900, 1920, and 1940, in the HSM_SIEREM rainfall database.

All African National Meteorological services were then contacted to update the data series at a monthly time-step, especially from the last 20 years of the 20th century, but only a few of them sent recent monthly data. The number of months (Figure 4) with available data, therefore, varies from 4061 stations in 1975 to 1464 stations in 1999. This shows a decrease of the number of observed rain gauges; this trend started as soon as the 1980s and has been observed worldwide, but it affects Africa more severely. However, the HSM_SIEREM database still has a sufficient number of points, compared to the Global Historical Climatology Network (GHCN) values—2500 gauges, worldwide, for the year 2000 [5], compared to a little less than 1,500 for Africa alone, in our database, for December 1999.
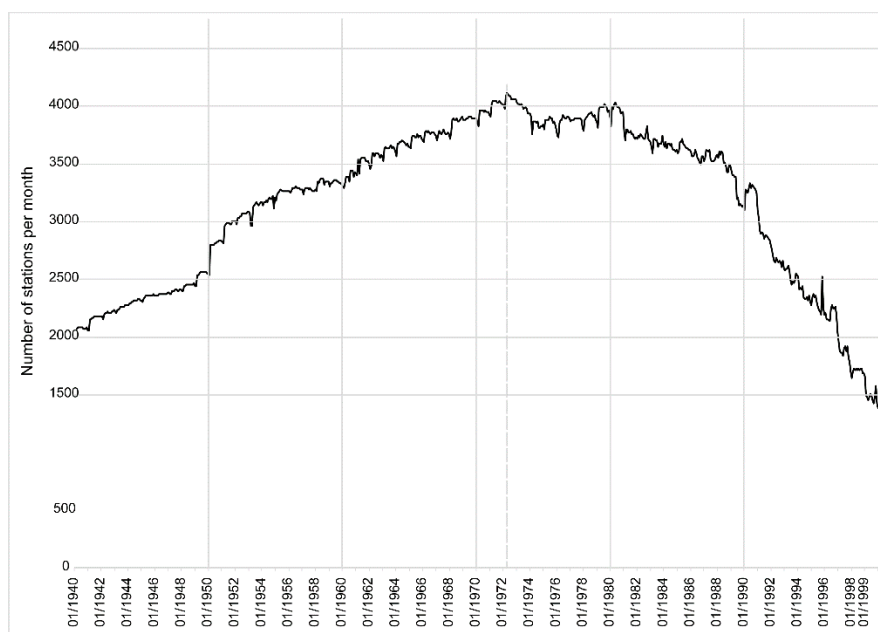
Figure 4. Number of monthly data observed in the HSM_SIEREM base, per year, from January 1940 to December 1999.

The scarcity of rainfall data was a recurrent problem, whatever the purpose [5,9,21]. Due to its history, the HSM laboratory collected a large amount of data. However, it is not allowed, either by the National Meteorological Services or the WMO (World Meteorological Organization) to disseminate raw rainfall data, as they are kept as the property of these Services. It was then decided that interpolated grids would be created, which would be disseminated for free to the community as the HSM_SIEREM database. This database has already been compared to the CRU for hydrological modeling in Western and Central Africa, and had given better results [7].

2.2. Geographical Processing

2.2.1. Reference of the CRU Grid Constitution

Our references for rainfall grids over the African continent were the CRU grids. These were built at a resolution of 0.5 × 0.5 degree. They concern seven climatological variables, namely precipitation, wet-day frequency, mean temperature, diurnal temperature range, vapor pressure, cloud cover, and ground frost frequency. For all data types, the construction of a monthly surface climate over global land areas, excluding Antarctica, covers the 1901/1996 time period (period of the first version, later update extended this period up to the 21st century). An "anomaly" approach was used, this technique attempted to maximize the available station data in space and time. "In this technique, grids of monthly anomalies relative to a standard normal period (1961/1990) (. . .) were first derived. The anomaly grids were then combined with a high-resolution mean monthly climatology to arrive at fields of estimated monthly surface climate." [8]. Rainfall anomalies are expressed in percentage. The standard period was chosen because of the good quantity of data available for this period.

The consistency checks were somewhere similar to the ones we applied on our data sets, defining monthly minimum, mean, and maximum values, and checking the values out of these ranges for confirmation or removal.

The anomalies over the standard normal period of rainfall were calculated, the CRU used tiles; 29 tiles were defined worldwide, with 4 tiles being defined for the African continent: tile 17 (20°W to 20°E, 38°N to 0°), tile 18 (20°E to 52°E, 38°N to 0°), tile 19 (0° to 52°E, 0° to 20°S), and tile 20 (10° to 52°E, 20°S to 36°S).

Three global station datasets were compiled by the CRU, from the basis of the construction of the gridded anomalies of primary variables. The precipitations and mean temperatures were compiled by the CRU, the diurnal temperature range dataset was based on the GHCN maximum and minimum temperature data.

The interpolation method adopted to process the tiles was a thin-plate spline, a function of latitude, longitude, and elevation. "The technique is robust in areas with sparse or irregularly spaced data points. The main advantage of splines over many other geostatistical methods is that prior estimation of the spatial autocovariance structure is not required" [8].

The process was computed per tile at a step of 0.5×0.5 degrees, with the tiles overlapping by at least 5×5 degrees (the number of stations varied from 200 to 1000 per tile). On the overlapping areas, the grid values were calculated as a weighted average of the contributing tiles; the weights were simply grid points between a particular point and the edge of its tile. The grids were constrained to avoid unrealistic values for rainfall, and the negative values were converted to zero.

"A GCV and its square root (RTGCV) provide an estimate of the mean predictive error (and hence power) of the fitted surface and as such permit an assessment of the relative accuracy of a fitted surface" [8].

The second step was to collect mean temperature, diurnal temperature, and precipitation over the 1901/1996 period. Any series with less than 20 years of data during the 1961/1990 period were excluded from the analysis. Rainfall data were expressed in percentages. Each station time-series was converted to anomalies, relative to the 1961/1990 mean.

The interpolation of the monthly anomaly method chosen for the 1901/1996 series was ADW (Angular Distance-Weighted), using the eight nearest stations, except when there were more than eight stations within a single $0.5°$ grid cell (only three cells were used in this case). The ADW gridding employed in this study did not permit the inclusion of elevation as a predictor, it was, thus, not included in a tri-variate interpolation technique [9]. "To prevent extrapolation to unrealistic values, the interpolated anomaly fields were forced towards zero at grid points beyond the influence of any stations. This was accomplished by creating synthetic stations with anomaly values of zero in regions where there were no stations within a predefined distance chosen to be equal to the global-mean CDD"; (450 km for precipitation) [9].

2.2.2. Choice of the Interpolation Process Options

Compared to the CRU standard normal reference period, we had a sufficient number of observations for a longer period and could interpolate rainfall from the observed values at each monthly time-step (Figure 5).

One file per month was created in the period 1940/1999 (i.e., 720 files). All files of monthly data were imported in the ArcMap and a point shapefile was created. The following processes were computed, using the ArcInfo Workstation macro-command language (AML). This ensured a homogeneous process.

Considering the size of the region to process and the heterogeneity of the distribution of rain gauge stations, and taking advantage of the CRU experience, the IDW (Inverse Distance Weighted) method was chosen.

With the ArcInfo Workstation, the IDW interpolation method does not allow to take into account the altitude of the measurement points. A few things can be discussed about this point. Due to the very different geographical situations across Africa, the question of altitude influence on precipitation cannot be dealt with one simple approach for the whole continent. For instance, altitude has no influence on rainfall over Western Africa (except in few specific locations, like near the Atakora mountains and close to the Guinea Mounts) and in the Sahel, covering several thousands of kilometers, therefore, the spline method gives the same results as the kriging method [22]. The coast of Cameroon is particularly wet, and the highest rainfall amounts in Africa are recorded at the foot of the Mount Cameroon; the HSM_SIEREM database gives much higher rainfall in this area than the CRU one—an

example of the kind of differences between the two grids. In this area the rainfall decreases with altitude, from 10,000 mm per year on the coast of Debundsha (altitude 26 m), to 2000 mm near the top of Mount Cameroon (4040 m). The situation was roughly identical on the coast of the Guinean Mounts, with very high annual amounts on the coast, and decreasing values with elevation. Contrarily, the rainfall amount increased in other hilly areas, over the continent, like the Atlas Mountains in Maghreb, or in the hills of East Africa. In this context, it could be seen that the spatial gradient of precipitation with altitude was not homogeneous over the continent, and to consider altitude in the interpolation framework would require a more regional approach.
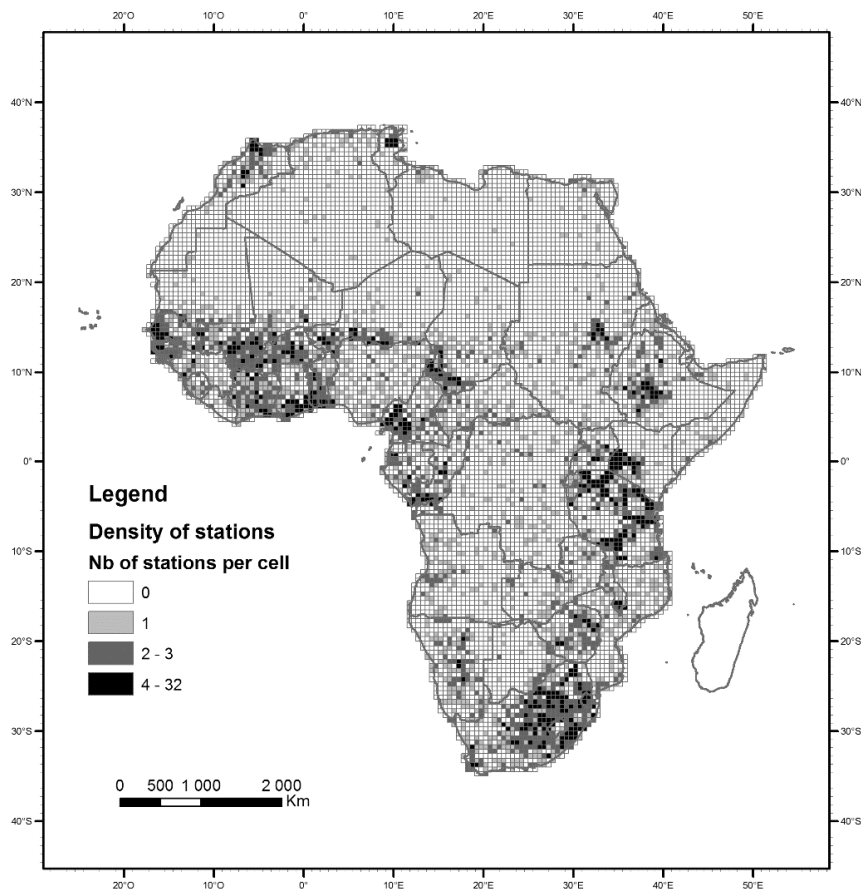

Figure 5. Density of the stations in the HSM_SIEREM rainfall database.

2.2.3. Options for Spatialization

The IDW process with ArcInfo lets the user choose between different options of interpolation. Among these, the options used for this study were:

- The default power value was used: 2.
- Priority was given to the radius around the interpolated point, with a value of six times the cell size of the output grid (in this case, each cell was half a square degree; the radius is then three decimal degrees). All the points within the radius were used for the interpolation.
- A minimum of four observation points were required to spatialize. If this minimum was not filled, the radius was extended until the program found four points.
- The output grid cell size was half a square degree. This was the common resolution of the grids used for regional modeling.
- The crest line of the Atlas Mountains is known to stop the humid Atlantic winds. Therefore, a barrier-line was created along this crest line, which modified the spatialization so that the interpolation did not use the values of points that were beyond the barrier line, on both sides.

- The extent of the spatialization process was 18° west, 51.5° east, 37.5° north, and 35° south. If this option was not present, the spatialization stopped at the stations located at the four edges—west, east, north, and south.

The raster grids at a step of 0.5 degree were converted into vector and then intersected to a regular grid at the step of half a square degree. The ArcInfo covers were converted into shape files and the raster grids (the first step of the process) were exported as ASCII files. Both the shape files and the ASCII grids being the products that would be available as free downloads as well as NetCDF arrays.

2.2.4. Validation of the IDW Interpolation

To evaluate the robustness of the interpolation method considered here, the interpolation methods of inverse distance weighting (IDW) and ordinary kriging (OK) were validated, using a cross-validation, as in [23]. Each rain gauge was in turn removed and the monthly precipitation was estimated with the remaining stations using the different methods. Only stations with at least 30 years of monthly precipitation data, representing 2912 rain gauges, were considered to validate the methods, to ensure robust estimates of the validation metrics; the Pearson correlation coefficient (r) and the relative bias. For both methods, IDW and OK, a search neighborhood with a 300 km radius was considered to perform the validation. The variograms required for the OK method were fitted with a spherical variogram model for each time-step, when rain was measured for at least four stations, otherwise, the IDW interpolation was performed. The spherical variogram model is convenient for precipitation, as it is not a spatially continuous field like temperature, since it provides a value of the decorrelation distance [24].

The validation results indicated very similar performances between the IDW and the OK, with an average correlation coefficient of validation of 0.86 for the IDW and 0.85 for the OK. The relative bias over all stations was equal to 7.8% with IDW and 7.1% with OK. The spatial patterns of the validation results are very similar, as shown in Figure 6, for the correlation coefficients of the validation samples being obtained with IDW or OK. The areas with the lowest density of stations, such as East Africa or south of the Maghreb countries, had the lowest performances in validation. Therefore, the interpolated precipitation in these areas must be interpreted with care.
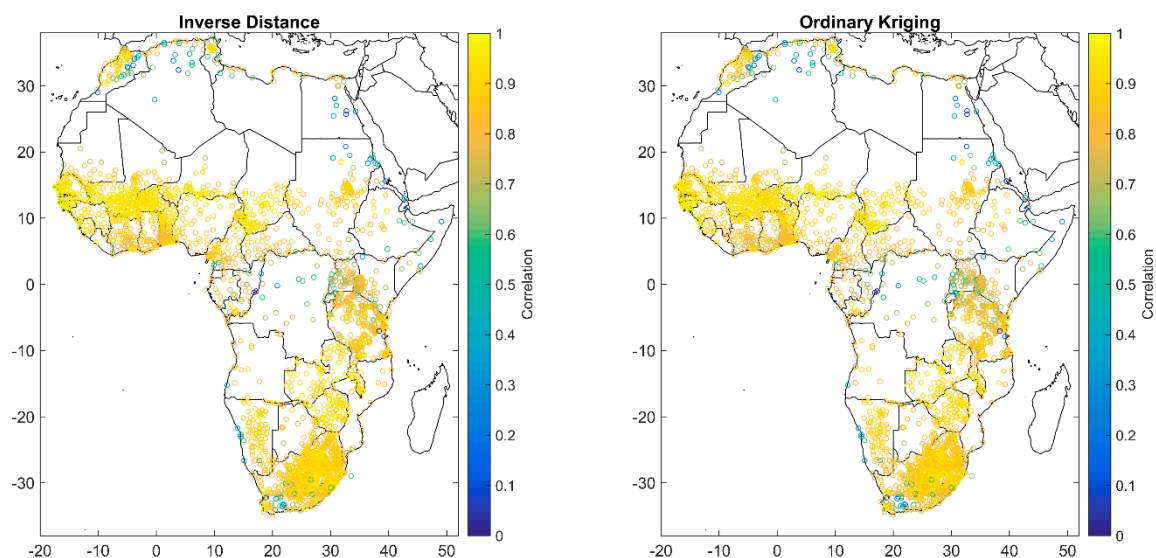

Figure 6. Validation results with independent station data for Inverse Distance (**left**) and Ordinary Kriging (**right**).

This validation confirmed that in the presence of a dense network, different interpolation methods (such as IDW or OK) performed with a similar efficiency.

2.2.5. Comparison between the CRU and HSM_SIEREM Grids

The grids from the CRU are widely used in models. Paturel et al. [11] compared the results of the GR2M hydrological model, using the rainfall grids from CRU and the ones from the previous versions of the HSM_SIEREM, over Western and Central Africa. The HSM_SIEREM database gave better results. Two mean annual grids over the 1940/1999 period were built, one with the HSM_SIEREM database and one with the CRU grids. The CRU monthly historical climate database, converted to the ESRI ASCII raster format by Jawoo Koo (HarvestChoice/IFPRI—Raw data), was downloaded from http://badc.nerc.ac.uk/data/cru). Figure 7 shows that the main features of the African rainfall distribution were similar.
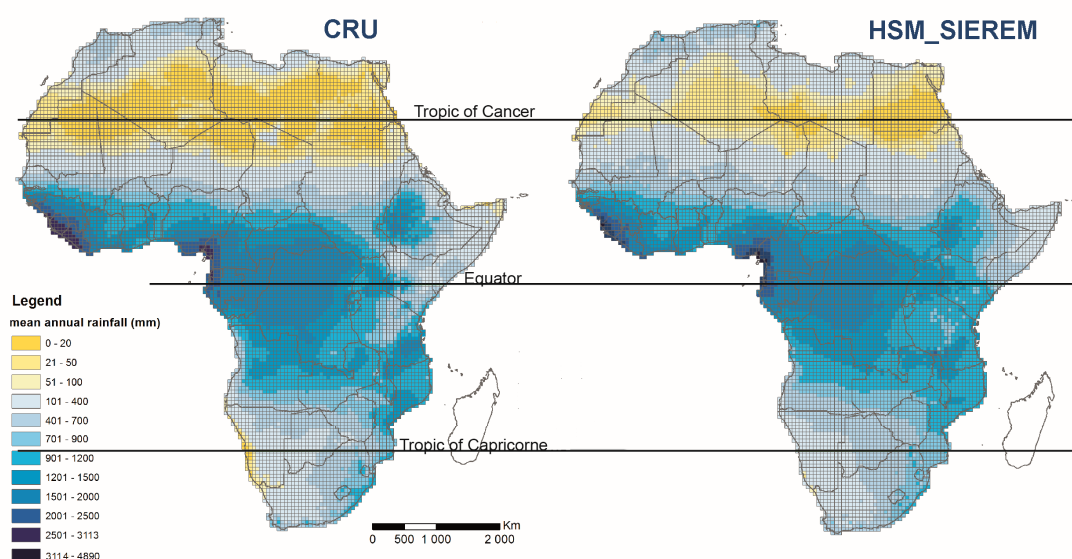

Figure 7. Mean annual Climatic Research Unit of University of East Anglia, UK (CRU) and HSM_SIEREM data base over the period of 1940/1999.

We checked the differences between the two mean inter-annual grids of the African continent, over the period of 1940 to 1999 (Figure 8).

The complete grid contained 10,380 cells, among which:

- A total of 4402 cells (in brown and yellow), i.e., 43%, had values ranging from −10% to −15,640%, compared to the HSM_SIEREM grid. They were mostly located in the Sahara and overlapped with the Sahel. In this area, the CRU grid had wide areas with values of mean annual rainfall between 0 and 20 mm, while the HSM_SIEREM ranged from 10 to 50 mm.
- A total of 4708 cells (in white), i.e., 45%, had values between −10% and +10%, compared to the HSM_SIEREM grid, which is very close.
- A total of 1270 cells (in green), i.e., 12%, had higher values, between +10% and +92%, compared to the HSM_SIEREM grid cells, the highest values being located on the southern slope of the Atlas Mountains (due to the barrier-line created in the HSM_SIEREM spatialization) and in Eastern Egypt (Dubief [22] drew a 5 mm isohyet in this area for the 50 first years of the 20th century).

The map of differences in percent showed two main features—in sub-Saharan Africa, the difference between the two grids mainly ranged between −9% and +10%, which could seem quite low, while the difference was much greater over Sahara and most of Northern Africa. In sub-Saharan Africa, some areas of low rainfall also showed a greater difference, as in the South Western coast of Africa and most of the horn of Africa. The CRU grid showed a higher rainfall over the Guinean mounts, the central part of the Congo basin, and the South of Angola. The HSM_SIEREM grid showed a higher rainfall in the South of Ghana, most of Nigeria, Cameroon and Gabon, along the South Western coast of

Africa, and over most of East and South Africa. About the Sahara region, the CRU grid showed values very different from that of the HSM_SIEREM grid. In most of the Sahara the CRU grid showed near 0 values, while the HSM_SIEREM grid showed a significant amount of rainfall, which were coherent with the Dubief rainfall map [25]. The CRU grid also showed too high a rainfall over the Saharan areas, in Egypt, North Chad, South-East Algeria and the Saharan border of the Atlas Mountains.
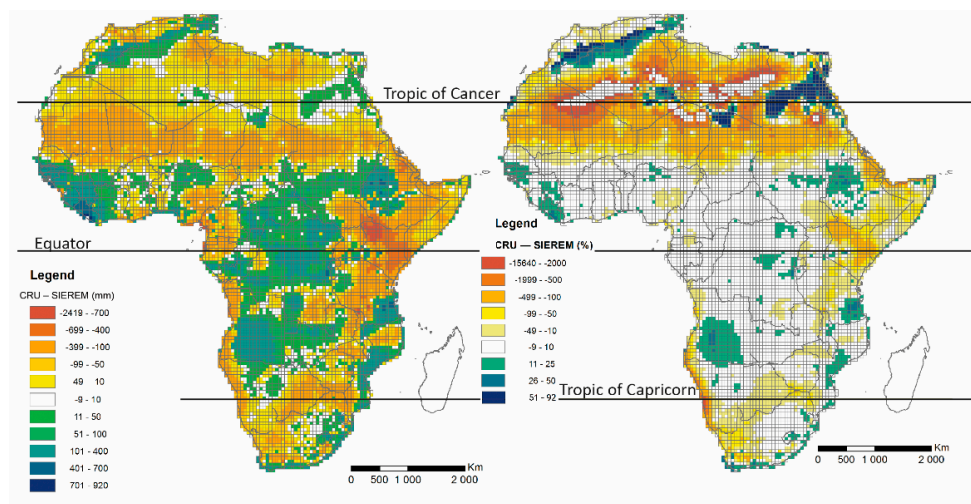

Figure 8. Differences of the CRU grid minus the HSM_SIEREM grid, in millimeters on the left and in percent on the right, over the period 1940/1999.

Even in the regions where the difference in percent was the lowest between the two grids (−9% to +10%), there were many areas where the difference was over 50 or 100 millimeters per year, which could have a significant impact on water-balance issues.

The differences between both grids could be due to the difference in the statistical approach of data processing and due to the density of observed stations. According to the figures given in Eklund et al. [26], the number of stations in Africa in the CRU data set was twice lower than that in the HSM_SIEREM data set, between 1940 and 1985, and three times lower after 1985. The maximum number of stations was over 4,000 for the HSM_SIEREM and less than 2000 for the CRU in the mid-seventies. It is quite likely that a greater number of stations will undergo a better precision in rainfall interpolation, at a grid scale.

3. Grid Production and Quality Assessment

The interpolation process was run over 720 months; every month was exported as an ASCII grid format file, containing one rainfall value at every 0.5 degree over the African continent; 6 zipped files, each containing 10 years of grids. Additionally, the NetCDF format array containing the 720 months of interpolated values, can be freely downloaded from the HSM website.

This first assessment of the HSM_SIEREM grid content by comparison with the CRU grid gives interesting details about where the use of HSM_SIEREM grid led to substantial differences with the CRU grids. This comparison would worth being completed by a first look at the rainfall variability over Africa, to check whether the main climatic signals are well-depicted by the HSM_SIEREM database. To assess the overall quality of this new dataset, we performed a first climatological analysis of rainfall over the whole continent, over the years 1940/1999.

3.1. Analysis of Rainfall Inter-Annual Variability

The standardized precipitation index (SPI) (Figure 9) was calculated, both, with the observed data (Figure 9a) and gridded data (Figure 9b), over the 60-year 1940/1999 period. For the observed data, the SPI was calculated with all stations having at least one complete year of observations. The SPI was

calculated as SPI = (Xi-X)/S; where: Xi is the mean annual rainfall of year i, X is the mean annual rainfall over the reference period, and S is the standard deviation of the rainfall set over the reference period.
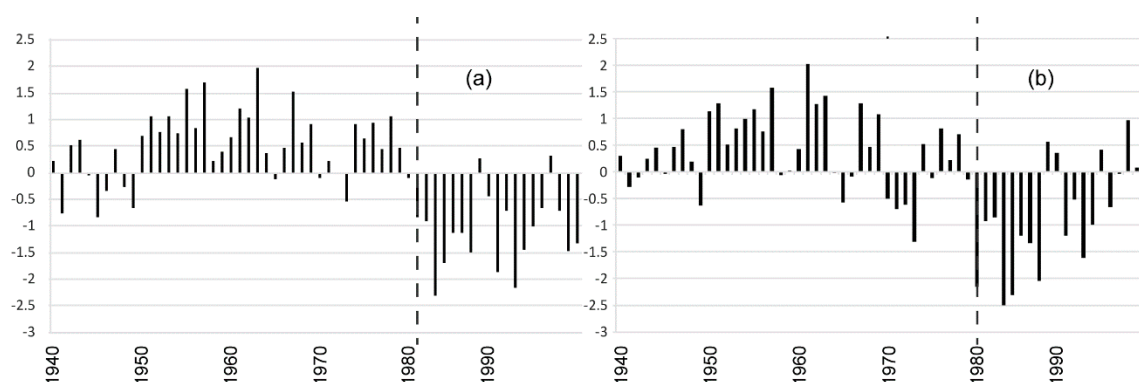

Figure 9. Mean annual rainfall for Africa, standardized precipitation index (SPI) on the observed data (**a**) and on the HSM_SIEREM gridded values (**b**).

SPI calculated with both observed and gridded data sets gave very similar time-series variations. It showed that, for the whole period, 1940/1999, and the whole continent, the SPI was almost permanently negative after 1981. This was coherent with most previous results [2,27]. There were some minor discrepancies between the two signals during the 1940s and at the end of the 1990s, which could, first, be linked to the lower number of stations in Africa during these periods, and, second, to their uneven distribution across the continent.

3.2. Statistical Tests on Mean Annual Rainfall

These tests, checking the random character of the series, were gathered using the Khronostat software [28], which can be downloaded, free of charge, at http://www.hydrosciences.org/spip.php?article1000. The most widely used tests deal with the stationarity of the mean of the series, throughout its observation period [29]. This test applied on the observed data confirmed a rupture in the series at three confidence levels, 99%, 95%, and 90% (Table 1).

Table 1. Results of the Mann–Whitney test, modified by Pettitt [29].

Observed Data	Gridded Data
Null hypothesis rejected at confidence levels of 99%, 95%, and 90%	Null hypothesis rejected at confidence levels of 99%, 95%, and 90%
Probability of exceeding the critical value 1.93×10^{-6} in 1979	Probability of exceeding the critical value 5.44×10^{-3} in 1969

The results of the statistical analysis of the inter-annual observed rainfall time-series over Africa was slightly surprising, as the 1970s rupture in the rainfall time-series is well known and has been largely described in Western and Central Africa [2,30,31] This meant that what happened in Western Africa and pro parte in the Western part of Central Africa, was more or less specific to this area and not to the whole continent. This should inspire further research on regional variables to explain these climatic features. The fact that the rupture date was found to be in 1969, with the gridded, data might be due to two causes—the first being that the rupture in 1969 might not have been the most prominent signal everywhere in Africa, even if it was present in many places. The second points out the impact of the interpolation method in regions with a high heterogeneity of the distribution of stations, and also the impact of the difference of length of a time-series. Indeed, this rupture in 1969 on the gridded data might be mostly driven by the extension of this signal that was most prominent in Western and Central Africa, due to the lack of data in many other parts of sub-tropical Africa, after 1990. This has already been discussed by Singla et al. [32], for Northern Africa.

3.3. Monthly Time-Series

The previous results led to the construction of two new data sets of monthly rainfall for two distinct periods: 1940/1979 and 1980/1999. Then a graph was drawn for the three-monthly time-series (before and after the rupture of 1979, and the entire 60-year period) (Figure 10), for observed and gridded data.

For observed data (in the left, in Figure 10), the graph showed two rainy seasons, February/ay and June–October. The first rainy season occurred in the austral hemisphere and the second occurred in the boreal hemisphere. The second rainy season remained the rainiest and the differences between the three periods were limited. Concerning the first rainy season during the month of March, the monthly values reached during the years 1940/1979 were clearly much higher than those during the years 1980/1999—the mean monthly rainfall value decreased from 90 mm to 64 mm, by almost 30% for the observed values. For the gridded data set (in the right, in Figure 10), values showed the same interdecadal variability, but with lower monthly values and a lower difference between the periods. The difference for March was still significant, with a reduction of 12% after 1980, from 62 mm to 55 mm.
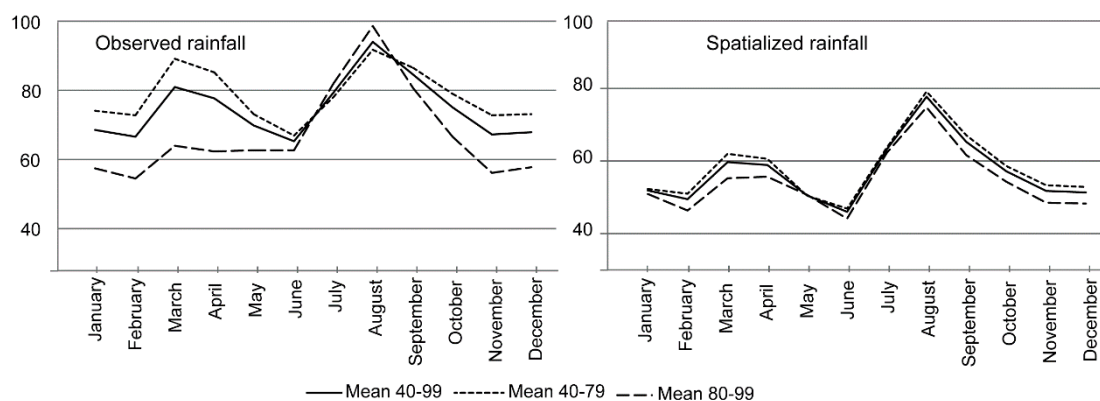

Figure 10. Seasonal variation of the rainfall for the periods 1940/1999, 1940/1979, and 1980/1999, for observed and gridded data.

Several authors [21,33,34] have already noticed a substantial decrease in rainfall over the southern part of Africa, between the equator and 10° South, but none were found to be as significant as the one shown here.

This first result showed that rainfall had decreased on a continental scale, for Africa, for several decades, and that the monthly rainfall regime did not change much between June and September, i.e., it showed a boreal summer tropical rainy season but changed much more during November to April/May, with a strong decrease that took place during the austral summer and autumn tropical rainfall period. This meant that it was mostly the Southern Hemisphere that seemed to have been affected by this decrease. One must note, however, that due to the rupture being detected in 1979/1980 for the whole continent, the decrease of rainfall that occurred in the Western and in part of Central Africa, since 1969/1970, did not appear clearly at the scale of the continent average. It was nonetheless one of the strongest climatic signals ever recorded in Africa.

4. Discussion and Conclusion

This paper is the result of a long duration of work on the collection, criticism, and assessment of a very rich and high-quality database of rainfall. This study could be carried out due to the quantity of data collected during a long period, mainly in Western and Central Africa, by the ORSTOM and the IRD, and due to the long and thorough criticism process that was applied to the data sets. It is especially precious, as the African continent has a particularly poor observation network, compared to the rest of the world.

The HSM_SIEREM rainfall gridded data set showed a very good correlation with the observed data, whatever the interpolation method used. It was compared to the widely used CRU rainfall grid, and their differences have been mapped, to help determine their best use for future studies. The HSM_SIEREM grid was built from twice the number of stations than the CRU grid, the number of stations was even three times higher for the period of 1985/1999, and the distribution of stations better covered the Sahara and some other parts of Africa. This gave the HSM_SIEREM grid more confidence in describing rainfall over Africa, with some regions showing very important differences, especially in the low-rainfall regions, in the Sahara and the Sahel, the Southwestern coast of Africa, and the horn of Africa.

The results confirmed that the African continent was seriously affected by a rainfall deficit and a change in the rainfall regime at the end of the 20th century. The first assessment showed that the two main periods of rupture in the time-series was between 1969/1970 and 1979/1980, which are depicted in the global African rainfall times-series issued from the whole data set, in both, the gridded and observed time-series. It seems that it was the first part of the year, from February to May, which registered the highest rainfall reduction, especially in the austral part of Africa (Figure 11). This reduction was several times higher than the rainfall reduction in the boreal part of Africa, at the same period. Even if part of these figures might be linked to the interpolation of unevenly distributed observed rainfall stations, this could have had strong incidences on the global climate of the area and requires further studies to analyze the regional variability of rainfall over the continent, between different databases.
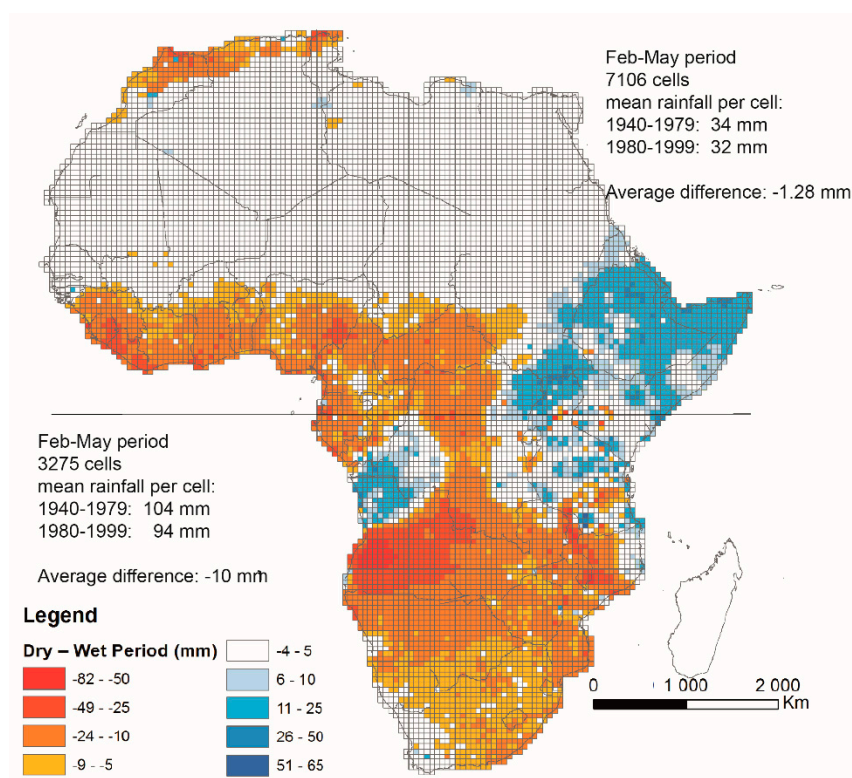

Figure 11. Map of the anomalies, Globally Dry Period (1980/1999)–Globally Wet Period (1940/1979). Average period for the months of February to May.

This study also showed that the use of direct observed data for interpolation gave different results and spatial representations than the use of rainfall patterns and anomalies to generate annual maps, which were used by the CRU and many other data centers. The comparison between the spatially-averaged rainfall data and observed data, over several regions of Morocco [32], showed that averaging th data over large surfaces led to a reduced visibility of climatic variability and the

main climatic signals, but the main signals remained visible whatever the size of the region used for averaging. This supports the idea that the results visible with, both, observed and gridded data sets are quite consistent.

All monthly grids built during this study are available for free on the HydroSciences Montpellier website and our aim is to contribute to the maintenance of an accurate database for climatic scenarios (http://www.hydrosciences.org/spip.php?rubrique1387/). These grids are already available on the Researchgate website: https://www.researchgate.net/project/Monthly-rainfall-gridded-data-set-for-Africa.

As it is very difficult to gather observed data covering the whole continent since 2000, it is improbable that an update of this study might be performed, knowing that, since then, a number of rainfall gridded product have been released on the basis of satellite-derived rainfall estimations [35], which help fill the gap between the needs of researchers for recent data and the unavailability of a free international observed rainfall database. The length of the time-series was, thus, adapted for a comparison of historical time-series and rainfall time-series from the GCM/RCM reanalysis, or for study of time-series for regions where data are difficult to obtain, even for historical periods.

References

1. IPCC. Summary for Policymakers. In *Climate Change 2013: The Physical Science Basis. Contribution of Working*; Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013.
2. Mahé, G.; Olivry, J.C. Variations des précipitations et des écoulements en Afrique de l'Ouest et Centrale de 1951 à 1989. *Sécheresse* **1995**, *6*, 109–117.
3. Servat, E.; Paturel, J.-E.; Kouamé, B.; Travaglio, M.; Ouedraogo, M.; Boyer, J.F.; Lubès-Niel, H.; Fritsch, J.M.; Masson, J.M.; Marieu, B. *Identification, Caractérisation et Conséquences d'une Variabilité Hydrologique en Afrique de l'Ouest et Centrale. 'Abidjan'98', Abidjan, Côte d'Ivoire, 16–19 Novembre 1998*; IAHS-AISH Publication: Wallingford, UK, 1998; Volume 252, pp. 323–337.
4. Mahé, G.; L'Hote, Y.; Olivry, J.C.; Wotling, G. Trends and discontinuities in regional rainfall of West and Central Africa—1951–1989. *Hydrol. Sci. J.* **2001**, *46*, 211–226. [CrossRef]
5. New, M.; Todd, M.; Hulme, M.; Jones, P. Precipitation measurements and trends in the twentieth century. *Int. J. Climatol.* **2001**, *21*, 1899–1922. [CrossRef]
6. Washington, R.; Harrison, M.; Conway, D.; Black, E.; Challinor, A.; Grimes, D.; Jones, R.; Morse, A.; Kay, G.; Todd, M. African Climate Change. *Bull. Am. Meteorol. Soc.* **2006**, *87*, 1355–1366. [CrossRef]
7. Mahé, G.; Girard, S.; New, M.; Paturel, J.E.; Cres, A.; Dezetter, A.; Dieulin, C.; Boyer, J.F.; Rouché, N.; Servat, E. Comparing available rainfall gridded datasets for West Africa and the impact on rainfall-runoff modelling results, the case of Burkina-Faso. *Water SA* **2008**, *34*, 529–536.
8. New, M.; Hulme, M.; Jones, P. Representing Twentieth-Century Space-Time Climate Variability. Part I: Development of a 1961–90. Mean Monthly Terrestrial Climatology. *J. Clim.* **2000**, *12*, 829–856.
9. New, M.; Hulme, M.; Jones, P. Representing Twentieth-Century Space-Time Climate Variability. Part II: Development of a 1901–96. Monthly grids of Terrestrial Surface Climate. *J. Clim.* **2000**, *13*, 2217–2238.
10. Paturel, J.E.; Boubacar, I.; L'Aour-Cres, A.; Mahé, G. Monthly rainfall grids in West and Central Africa. *Revue des Sciences de l'Eau* **2010**, *23*, 325–333. [CrossRef]

11. Paturel, J.E.; Ouedraogo, M.; Mahé, G.; Servat, E.; Dezetter, A.; Ardoin, S. The influence of distributed input data on the hydrological modelling of monthly river flow regimes in West Africa. *Hydrol. Sci. J.* **2003**, *48*, 881–890. [CrossRef]

12. Boyer, J.F.; Dieulin, C.; Rouché, N.; Cres, A.; Servat, E.; Paturel, J.E.; Mahé, G. SIEREM: An environmental information system for water resources. In *Climate Variability and Change—Hydrological Impacts, Proceedings of the Fifth FRIEND World Conference, Havana, Cuba, November, 26th 2006*; IAHS Publication: Wallingford, UK, 2006; pp. 19–25.

13. Paturel, J.E.; Boubacar, A.; L'Aour, A.; Mahé, G. Analyses de grilles pluviométriques et principaux traits des changements survenus au 20ème siècle en Afrique de l'Ouest et Centrale. *Hydrol. Sci. J.* **2010**, *55*, 1281–1288. [CrossRef]

14. Paturel, J.E.; Ouedraogo, M.; Servat, E.; Mahé, G.; Dezetter, A.; Boyer, J.F. The concept of rainfall and streamflow normals in West and central Africa in a context of climatic variability. *Hydrol. Sci. J.* **2003**, *48*, 125–137. [CrossRef]

15. Ardoin-Bardin, S.; Dezetter, A.; Servat, E.; Paturel, J.E.; Mahé, G.; Niel, H.; Dieulin, C. Using general circulation model outputs to assess impacts of climate change on runoff for large hydrological catchments in West Africa. *Hydrol. Sci. J.* **2009**, *54*, 77–89. [CrossRef]

16. Conway, D.P.; Persechino, A.; Ardoin-Bardin, S.; Hamandawana, H.; Dieulin, C.; Mahé, G. Rainfall and river flow variability in sub-Saharan Africa during the 20th century. *J. Hydrometeorol.* **2009**, *10*, 41–59. [CrossRef]

17. Jaccon, G.; Travaglio, M. La banque de données hydropluviométriques du laboratoire d'hydrologie. In *Systèmes d'information pour l'environnement*; Colloques et Séminaires; ORSTOM: Bondy, France, 1990; pp. 49–59.

18. L'hote, Y.; Mahé, G. *Afrique de l'Ouest et Centrale, Précipitations Moyennes Annuelles (Période 1951–1989)*; ORSTOM/IRD: Bondy, France, 1996.

19. Water Assessment. In *Evaluation Hydrologique de l'Afrique Sub-Saharienne. Pays de l'Afrique de l'Ouest*; Mott MacDonald International, BCEOM, SOGREAH, ORSTOM, Eds.; Sci. Financement Banque Mondiale, PNUD, BAD, MFC; ORSTOM Editions: Montpellier, France, 1992.

20. Rouché, N.; Mahé, G.; Ardoin-Bardin, S.; Brissaud, B.; Boyer, J.F.; Cres, A.; Dieulin, C.; Bardin, G.; Commelard, G.; Paturel, J.E.; et al. Constitution d'une grille de pluies mensuelles pour l'Afrique, période 1900–2000. *Secheresse* **2010**, *21*, 336–338.

21. Nicholson, S.E. The nature of rainfall fluctuations in subtropical West Africa. *J. Am. Meteorol. Soc.* **1980**, *108*, 473–487. [CrossRef]

22. Wotling, G.; Mahé, G.; L'Hôte, Y.; Le Barbé, L. Analysis of the space-time variability of annual rainfall linked to the African monsoon using regional vectors. *Veill. Clim. Satellitaire* **1995**, *52*, 58–73.

23. Tramblay, Y.; Thiemig, V.; Dezetter, A.; Hanich, L. Evaluation of satellite-based rainfall products for hydrological modelling in Morocco. *Hydrol. Sci. J.* **2016**, *61*, 2509–2519. [CrossRef]

24. Lebel, T.; Bastin, G.; Obled, C.; Creutin, J.D. On the accuracy of areal rainfall estimation. *Water Resour. Res.* **1987**, *23*, 2123–2134. [CrossRef]

25. Dubief, J. *Le Climat du Sahara. Mémoire de Recherche de l'Institut de Météorologie et de Physique du Globe de l'Algérie*; Université d'Alger: Alger, Algérie, 1963.

26. Eklund, L.; Romankiewicz, C.; Brandt, M.S.; Doevenspeck, M.; Samimi, C. Data and methods in the environment-migration nexus: A scale perspective. *Die Erde J. Geogr. Soc. Berl.* **2016**, *147*, 139–152. [CrossRef]

27. Nicholson, S.E.; Fink, A.H.; Funk, C. Rainfall over the African continent from the 19th through the 21st century. *Glob. Planet. Chang.* **2018**, *165*, 114–127. [CrossRef]

28. Lubes-Niel, H.; Masson, J.M.; Servat, E.; Paturel, J.E.; Kouame, B.; Boyer, J.F. *Caractérisation de Fluctuations dans une série Chronologique par Application de Tests Statistiques. Etude Bibliographique*; Programme ICCARE. Rapport n°3; ICCARE; ORSTOM: Montpellier, France, 1994.

29. Pettitt, A.N. A non-parametric approach to the change-point problem. *Appl. Stat.* **1979**, *28*, 126–135. [CrossRef]

30. Hulme, M.; Doherty, R.; Ngara, T.; New, M.; Lister, D. African climate change: 1900–2100. *Clim. Res.* **2001**, *17*, 145–168. [CrossRef]

31. Mahé, G.; Olivry, J.C. Changements climatiques et variations des écoulements en Afrique occidentale et centrale, du mensuel à l'interannuel. In Proceedings of the Hydrology for the Water Management of Large River Basins Congrès AISH, Vienne, Autriche, 13–15 août 1991; van de Ven, F.H.M., Gutknecht, D., Loucks, D.P., Salewicz, K.A., Eds.; IAHS Publications: Wallingford, UK, 1991; Volume 201, pp. 163–172.

32. Singla, S.; Mahé, G.; Dieulin, C.; Driouech, F.; Milano, M.; El Guelai, F.Z.; Ardoin-Bardin, S. Evolution des relations pluie-débit sur des bassins versants du Maroc. In *Global Change: Facing Risks and Threats to Water Resources, Proceedings of the Sixth World FRIEND Conference, Fez, Morocco, October 2010*; IAHS Publications: Wallingford, UK, 2010; Volume 340, pp. 679–687.

33. Fauchereau, N.; Trzaska, S.; Rouault, M.; Richard, Y. Rainfall Variability and Changes in Southern Africa during the 20th Century in the Global Warming Context. *Nat. Hazards* **2003**, *29*, 139–154. [CrossRef]

34. Morishima, W.; Akasaka, W. Seasonal trends of Rainfall and Surface Temperature over Southern Africa. *Afr. Study Monogr.* **2010**, *40*, 67–76.

35. Adeyewa, Z.D.; Nakamura, K. Validation of TRMM rainfall data over major climatic regions in Africa. *J. Appl. Meteorol.* **2003**, *42*, 331–347. [CrossRef]