Pan African University
Institute of Water
and Energy Sciences

PAN-AFRICAN UNIVERSITY
INSTITUTE FOR WATER AND ENERGY SCIENCES
(Including CLIMATE CHANGE)

Master Dissertation

Submitted in partial fulfillment of the requirements for the Master degree in

WATER ENGINEERING

Presented by:

Yasmine ZIANI

The comparison of historical time series of gridded monthly precipitation data from different sources CRU, GPCC, and HSM_SIEREM over Africa for the period (1940-1999).

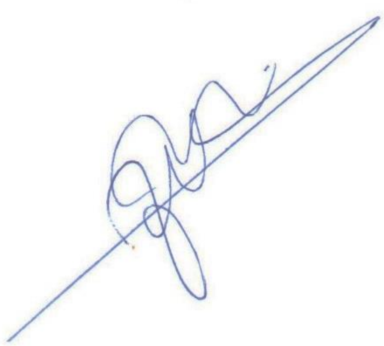
Defended on 04/09/2019 Before the Following Committee:

Chair	Nesrin Sahelian	Dr	University Of Mashhad, Iran
Supervisor	Karima BENHATTAB	Dr	University of Oran, Algeria.
Co. Supervisor	Gil MAHE	Prof	IRD, France.
External Examiner	Bernhard TISCHEBIEN	Dr	ZEF, Germany
Internal Examiner	Abdelhadi AMMARI	Dr	ENSH, Algeria.

Academic year: **2018/2019**

I, **ZIANI Yasmine** hereby declare that this thesis represents my personal work, realized to the best of my knowledge. I also declare that all information, material and results from other works presented here, have been fully cited and referenced in accordance with the academic rules and ethics.

Signed **Date__31/08/2019__________**



Dedication

To my parents,

For the rising generation who will persevere for the pursuit of science before honors,

For us, African students, make sure we change Africa with our own hands…

Ziani Yasmine.

Acknowledgements

My thanks go first to my supervisors ***Ms. BENHATTAB Karima*** *for her motivations and encouragement that she showed from the beginning of the project,*

And

Mr MAHE Gil *for his righteousness throughout this project.*

I also thank ***Mrs DIEULIN Claudine*** *Engineer at the research institute for the development with who I learned a lot during my internship by sharing her irreproachable experience in Africa and her particular instructions.*

To the head department of Climatology and data treatment of Hydrometeorological Institute of Training and Research,Oran ***Mr TAIBI Slimane*** *who had the pleasure of welcoming me to answer some of my questions.*

I thank ***Ms Louise*** *for her help during my first steps on the software R*

My thanks in the end go to HSM general staff.

Abbreviation

ARC	African Rainfall Climatology
CHIRPS	Climate Hazards Group Infrared Precipitation with Station
CNRS	Centre National de la Recherche Scientifique
CRU	Climatic Research Unit
GIS	Geographical Information Analysis
GEOSS	Global Earth Observation System of Systems
GPCC	Global Precipitation Climatology Centre
ENSO	El Niño Southern Oscillation
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IDW	Inverse Distance Weighting
IRD	Institute of Research for development
MEA	The Millennium Ecosystem Assessment
MSE	Mean-square error
NESDIS	National Environmental Satellite data and Information Service
NOAA	National Oceanic and Atmospheric Administration
NOAA-CPC	National Oceanic and Atmospheric Administration Climate Prediction Center
ODI	Open Data Institute
ORSTOM	Office de la recherche scientifique et technique outre-mer
TRMM	Tropical Rainfall Measuring Mission
TIN	Triangulated Irregular Network
SIEREM	Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation.
UNEP	United Nations Environment Programme
WAM	West African Monsoon
WeMO	Western Mediterranean Oscillation
WMO	World Meteorological Organization

Table of Contents

List of tables

List of Figures

LITERRATURE REVIEW

METHODOLOGY AND MATERIALS

RESULTS

Abstract

Nowadays, many precipitation gridded datasets are used for climatic research in Africa, a way for scientists to complement their observed data for several studies as river-runoff relationship, impact of climate change or hydrological modelling. These gridded datasets supposedly based on the same gauging stations delivered by any National Meteorological Institution seem to differ greatly according to different causes: the number of stations used, the quality check of data, the length of the time series, the interpolation method and lack of data for several periods and countries.

In our study, we compared a number of available precipitation gridded dataset which are GPCC, CRU and HSM-SIEREM at a resolution of 0.5 x 0.5 degrees, during period between 1940 and 1999, and over Africa, then, we checked their correlation with ground based data chosen randomly to conclude the best gridded data product which appears to be closest to the reality.

The results shows HSM-SIEREM presents best correlation coefficient with 0.9761, followed by GPCC with an insignificant difference by a correlation coefficient 0.9734 and a little behind CRU with a coefficient of correlation 0.9389.

Keywords

Africa, Gridded data, Precipitation, GPCC, CRU, HSM-SIEREM, Comparison, Cartography.

Résumé

De nos jours, de nombreux jeux de données maillés sur les précipitations sont utilisés pour la recherche climatique en Afrique, ce qui permet aux scientifiques de compléter les données observées pour plusieurs études telles que la relation débit-ruissellement, l'impact du changement climatique ou la modélisation hydrologique. Ces ensembles de données quadrillées supposément basés sur les mêmes stations de jaugeage délivrées par une institution météorologique nationale semblent différer considérablement selon les causes: le nombre de stations utilisées, le contrôle de la qualité des données, la longueur de la série chronologique, la méthode d'interpolation et le manque de données sur des périodes et pays précis.

Dans notre étude, nous avons comparé un ensemble de données grilles de précipitation disponibles, qui sont GPCC, CRU et HSM-SIEREM à une résolution de 0,5 x 0,5 degrés, entre 1940 et 1999, et sur l'Afrique. Nous avons ensuite vérifié leur corrélation avec les données sol choisies au hasard pour conclure le meilleur produit de données maillé qui semble le plus proche de la réalité.

Les résultats montrent que HSM-SIEREM présente le meilleur coefficient de corrélation avec 0,9761, suivi de GPCC avec une différence non significative d'un coefficient de corrélation de 0,9734 et un peu derrière CRU avec un coefficient de corrélation de 0,9389.

Mots clés :

Afrique, données grilles, précipitations, GPCC, CRU, HSM-SIEREM, comparaison, cartographie.

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Precipitation is one of the most basic meteorological elements. It is thus crucial to be able to forecast precipitation variability according to climate change. Recent computer resources and techniques have been used in simulating hydrological changes, such as changes in river runoff and extreme events resulting from changes in precipitation, climatic research in Africa; such as regional studies of river-runoff relationship, impact of climate change on hydrological regimes in Western and Center of Africa, and for sub-continental studies of climate change **(**Dieulin et al., 2019). For such purposes, data on precipitation are needed supposedly based on gauging stations observations delivered by the National Meteorological Institutions, they may be insufficient or missing in specific localization of period, therefore a gridded precipitation dataset is necessary (Yatagai et al., 2012**)**. These gridded data are accessible and delivered by many environmental data information systems which propose free access to their data online, including TAMSAT (Tropical Applications of Meteorology using SATellite data and ground-based observations) daily rainfall dataset based on high-resolution thermal-infrared observations, available from 1983, CRU (Climatic Research Unit) developed by University of East Anglia (UK) and considered one of the oldest gauged-based database available since the end of the 19th century, GPCC (Global Precipitation Climatology Center) which provides gridded gauge-analysis products derived from quality controlled station data[1] and SIEREM (Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation) which developed a database of monthly gridded rainfall data for Africa at 0.5 º x 0.5 º resolution and over the period 1900-2000, But they seem to differ greatly according to several causes: the number of stations used, the quality check of data, the length of the time series, the interpolation method and lack of data for several periods and countries.

Many studies have been conducted to compare different datasets for many purposes as modelling and correcting algorithms. Using HSM-SIEREM data, number of papers have been published which describe the content of the base, the quality check and the techniques used. A comparison between kriging and inverse distance methods using a cross validation method have been done, and both give very similar accuracy according to the ground based data, Moreover,

[1] (https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre)

Mahe and al, 2008 published a study on the evaluation of data from three gridded rainfall data sets of monthly rainfall available in 5 basins across Burkina-Faso over the period 1922 to 1998. Their analysis has enabled them to investigate the sensitivity of river flows simulated by gridded hydrological modelling, to the quality of available gridded monthly rainfall data sets.

Besides, by using monthly grids of rainfall data provided by the Climate Research Unit (CRU), Paturel and al, 2010 have undertaken analysis of the main features of the changes in rainfall regime that they highlight in West and Central Africa during the 20th century. From these data grids, and using robust statistical analysis tools, it is shown that during the last century West and Central Africa experienced alternations of dry and rainy seasons with very variable spatial and temporal extent.

Furthermore, Paturel and al, 2009 published a research note on the monthly rainfall grids in West and Central Africa. They compared gridded rainfall data, with those provided by the Climate Research Unit of University of East Anglia (CRU-UEA).

The main goal of this study is to compare a number of available gridded data set from different sources, which propose rainfall grids during period between 1940 and 1999, resolution of 0.5^{o} x 0.5^{o} and for spatial coverage of whole Africa, suggesting in the end the best representative specialize data to use to conduct eventual researches.

The study in general includes comparison on methodology of interpolation used by each dataset to specialize data and an approach on the influence of atmospheric teleconnection patterns on the changes in rainfall variability over Africa and a validation of the dataset in the end with some gauge stations.

The work plan was organized as follows:

A literature review will be firstly on status of satellites precipitation dataset, their systematic and random errors and Satellites earth observations and environmental open data policies. Second part will be a presentation on different Africa gridded datasets and their comparison;

A chapter on methodologies and materials which describe first the study area “The continent” Africa in general (localization, main apparent elevations, climate zones and water resources), secondly, all steps that occurred from downloading gridded data, to extraction using the statistical and programming software “R”, the export grids datasets in ASCII format for comparison using the software ArcGIS, and all calculations that have been made to obtain annual precipitation and inter-annual mean precipitation for the period between 1940 and 1999 for the three datasets (HSM-SIEREM, GPCC and CRU). This chapter will contain also a part on the interpolation methods used to specialize each dataset since the methodology is very important to make difference between datasets, and the correlation method to correlate observed data with gridded data in some points over Africa.

Results and discussion on maps generated and for the validation of gridded data with the observed data will be developed in the last chapter.

LITERATURE REVIEW

CHAPTER I

CHAPTER I

I.1 Part one: Satellites precipitation dataset

I.1.1 Introduction

Precipitation is one of the most basic meteorological elements and a highly dynamic phenomenon and changing its intensity, frequency and duration with influence of topographical parameters such as elevation, vegetation, land use cover, anthropogenic activity and many more fact(C, Shetty, & Kiros, 2016). It is thus crucial to be able to forecast precipitation variability according to climate change. Recent computer resources and techniques have been used in simulating hydrological changes, such as changes in river runoff and extreme events resulting from changes in precipitation. (Yatagai et al., 2012) However, there are two main problems with regards to access to rainfall data over most of Africa. The first problem is that the available station network is very sparse. Second, access to the available data is very limited particularly outside the individual countries. (Abtew & Melesse, 2014) besides, the availability of historical precipitation data sets can also be problematic, varying in availability, completeness and consistency as well availability for near real-time analysis (Kidd & Levizzani, 2011) As a result, a gridded precipitation dataset is necessary and many researchers use either satellite data or gridded rainfall products, which are freely available from different providers. However, the qualities of these products need to be evaluated.

I.1.2 Status of satellites precipitation dataset

The high spatial and temporal resolution and the continuous, timely, and public availability of satellite datasets are essential requisites for downstream predictions with enough lead time to implement management and response actions, especially in poorly gauged basins (Serrat-Capdevila et al. 2014) (Cattani, Merino, & Levizzani, 2016)

Satellites can offer an unrivalled vantage point to observe and measure Earth system processes and parameters. Precipitation in particular, benefit from such observations since precipitation is spatially and temporally highly variable. Satellites are characterized by the need to provide frequent large-scale measurements resulting in large swath widths, resolutions finer than 1km and systems capable of regular/frequent observations. (Kidd & Levizzani, 2011).

I.1.2.1 Geostationary (GEO) satellites

Geostationary satellites are in an orbit and they rotate around the Earth at the same speed as the Earth rotates (west to east), directly over the equator. Thus appearing stationary relative to a location on the Earth. These satellites have limitations which is the exact position of a geostationary satellite, relative to the surface, it varies slightly over the course of each 24-hour period because of gravitational interaction among the satellite (Margaret Rouse, 2008)

A total of five geostationary satellites provide operational imagery: these currently include the Meteosat Second Generation satellites (MSG) from the European Organization for the Exploitation of Meteorological Satellites (EU- METSAT), two Geostationary Operational Environmental Satellite (GOES) and the Japanese Multifunctional Transport Satellites (MTSAT) series. (Cattani et al., 2016).

I.1.2.2 Low-Earth orbiting (LEO) satellites

This type of system employs a swarm of satellites, each in a polar orbit at an altitude of a few hundred kilometers. And generally cross the Equator at the same local time on each orbit. Current operational polar-orbiting satellites include the National Oceanic and Atmospheric Administration (NOAA) series of satellites with NOAA-19 and EUMETSAT's MetOp series (A, B and C). (Margaret Rouse, 2008), (Cattani, Merino, & Levizzani, 2016).

The Tropical Rainfall Measuring Mission (TRMM), as joint mission of NASA and the Japan Aerospace Exploration Agency (JAXA), launched in 1997, was the first dedicated precipitation satellite designed to improve understanding of the distribution and variability of precipitation (from official site of NASA[2]). TRMM can provide information on the occurrence of

[2] https://trmm.gsfc.nasa.gov

rainfall extremes and characteristics of the local meteorology that can help improve short-term forecasting of such events.

I.1.3 Precipitation products validation

Validation data sets, or surface reference data sets, generally rely upon available gauge data or radar depending upon the temporal and spatial scales over which the estimates were being evaluated. In all cases, the surface data sets are not necessarily the "truth" since many factors affect the true value (exploitation of the data, measuring equipment). However, through the careful selection of data, inconsistencies or fault can be reduced (Villarini et al., 2008). Specific ground validation campaigns identify specific locations for the collection of (primarily) radar and gauge data sets: these campaigns are generally designed for improving the physical retrievals of the satellite estimates. However, more recent and on-going validation rely upon mainly statistical inter-comparisons with surface data sourced from routinely available data sets. (Cattani, Merino, & Levizzani, 2016).

I.2 Part two: Satellites precipitation data and random/systematic errors

I.2.1 Introduction

Satellite precipitation data have error estimates. They are of crucial importance in many studies which contribute to characterize these errors to develop uncertainty models and algorithms (Aghakouchak, Mehran, Norouzi, & Behrangi, 2012), However, evaluating satellite precipitation errors is a very challenging task because it relates to multiple factors, including the natural temporal and spatial variability of precipitation, measurement errors, and sampling uncertainties, especially at fine temporal and spatial resolutions.

I.2.2 Random/Systematic errors

Uncertainty associated with satellite precipitation products includes both systematic and random errors. The random component mainly depends on the sensor sampling design, whereas biases arise from systematic problems such as the inclusion of gauge information that is only available over land (Maggioni, Sapiano, & Adler, 2016)

An example of a methodology in high-resolution satellite precipitation was developed to discern random errors from systematic ones by the scientific Willmott in 1981, and was applied by Agha Kouchak in 2012 to satellite precipitation products.

Willmott suggested the following formulation for the systematic and random components of the mean-square error (MSE) in numerical weather prediction models:

$$\mathrm{MSE} = \mathrm{MSE}_{\mathrm{Syst}} + \mathrm{MSE}_{\mathrm{Rand}} \quad \text{and} \qquad (1)$$

$$\frac{\sum_{n}(x-y)^2}{n} = \frac{\sum_{n}(\hat{x}-y)^2}{n} + \frac{\sum_{n}(x-\hat{x})^2}{n}, \qquad (2)$$

Where:

x: the satellite precipitation.

y: is the reference precipitation.

n: is the number of time steps

x: is defined as $\hat{x} = ay + b$, (3)

a and b are parameters (slope and intercept, respectively) to be calibrated. The systematic error is the part of the error to which a linear function can be fitted.

This component represents systematic average deviations of satellite estimates with respect to the corresponding reference rainfall over the pixel size (Willmott, 1981).and the procedure is introduced to be applied to future realizations of satellite precipitation and for real-time adjustments.

I.2.3 Remarks and conclusion on Random/Systematic errors

- ✓ Studies reveal that the systematic biases of satellite precipitation were distinctively different in the summer and winter. The results revealed that the systematic error was remarkably higher during the winter. The main reasons behind this increase may be problems with high latitude precipitation detection and winter precipitation estimation. (Sorooshian et al., 2011) (Aghakouchak et al., 2012)
- ✓ Quantification of systematic and random error components of precipitation may lead to a major advancement in the development of next generation bias removal algorithms. Currently, most adjustment algorithms are based on correcting the volume of rain rate over a certain period of time (e.g., monthly). (Sorooshian et al., 2011)
- ✓ In addition to variability of errors in space and time, systematic error is proportional to the intensity of rain rate. In a fact many studies showed that the systematic error increases as the rain rate increases**.** (Habib et al., 2009)

I.3 Part Three: Satellites earth observations and environmental open data policies:

I.3.1 Introduction

Open data access debate is very important to satellite earth observation. For decades the earth observation has seen different views from those who want data that are free of charge at the point to use in order to maximize scientific benefits of data. (Harris & Baumann, 2015) and for environmental purposes in collecting climate data, satellites earth observation play a vital role offering comprehensive global coverage that cannot be coordinated by in situ observation which make open data very requested.

I.3.2 Open data definition

The UK Open Data Institute (ODI) has defined open data as information that is available for everyone to use for any purpose, at no cost. Open data has to have a license that says it is open data. Without license the data cannot be reused. In another definition, given by Panton Principles[3] on open data in science, it means that it is freely available on public internet permitting any user

[3] Panton Principles, Principles for open data in science. Murray-Rust, Peter; Neylon, Cameron; Pollock, Rufus; Wilbanks, John; (19 Feb 2010). Retrieved [March 2019]

to download, copy, analyze, re-process, and pass them to software, for any other purpose without financial, legal, or technical barriers.

The term “Free” thus, should be reserved for data that are free of charge because the mean “Open” covers already the meaning of freely available.

I.3.3 Open access policies and legal instruments

Some of US Open data policies manage information as an asset and it is based on the presumption of openness. For them, open data refers to publicly available data that are structured in a way that enables the data to be fully discoverable and usable by end users. Especially the need of data to be machine readable and interoperable. Exceptions concerns confidentiality, security, trade secret etc.

About environmental satellite data and metadata, NOAA (National Oceanic and Atmospheric Administration) and the (National Environmental Satellite data and Information Service) NESDIS policies refers to twelve authorities and reference documents that recognizes the need for full and exchange of data. Furthermore, the access of these environmental data is subject to relevant international, national laws and agreement. The following link summarize the access and Distribution of Environmental Satellite Data and Products (http://www.ospo.noaa.gov/Organization/Documents/PDFs/NESDIS_Data_Access_Distribution_Policy.pdf).

Concerning worldwide organizations as GEOSS (Global Earth Observation System of Systems) which regroup 105 members (in 2019, http://www.earthobservations.org) share principles require full and open exchange of data, metadata and products but shall be made available with minimum time delay and minimum cost. Another principal expect that all shared data should be free of charge for educational and research purposes.

The corporation of International disaster charter regrouping 13 organizations allows the use of space facilities during event of disasters and make their earth observation data available.

The data is free of charge for the time of disasters such floods or storms. No matter what is pricing policy of the provider organisation.

European organizations as Copernicus data policy and ESA data policy make data available and free of charge subject to limitations concerning registration, dissemination formats, and access restrictions. (Harris & Baumann, 2015)

I.3.4 Conclusion

Very little systematic and structured research has been done on the issues that are covered by environmental open data policies, their intent and actual impact and no suitable framework for comparing open data policies is available since open data is a new phenomenon, a comparison was made by two scientific from Delft university of technology including factors as context of policies, policy content, performance indicators and public values, the comparison suggest more collaborations with national and international organizations, focusing on the impact of the policy, stimulating the use of open data and looking at the need to create a culture in which publicizing data is incorporated in daily working processes. (Zuiderwijk, 2013)

CHAPTER II

CHAPTER II

II.1 Introduction

The need for high resolution gridded data is decisive and not only limited to hydrological applications. Their Availability differ in depending on domain size, spatial, temporal resolution, and originate from different sources and methods (Ahmed, Shahid, Wang, Nawaz, & Najeebullah, 2019). Gridded data is defined as a geographically large distribution of variable (any meteorological "data") over a region or area, these gridded data take a certain resolution which can be a half square degrees 0, 5º x 0, 5 º or another. Resolutions are calculated from altitudes and longitudes lines where distance remain approximately 111Km between lines but varies frequently by longitude and the poles.

Thereby dimensions X and Y are given:

X = (0.50 * [Distance between longitude lines at a given location]), Y = (0.50 * [Distance between latitude lines at a given location]). (Michael G, 2017. University of Missouri).

II.2 Satellites gridded precipitation products over the world including the continent Africa

The assessment of satellites gridded precipitation data is usually done by comparing with reliable observed data. (Ahmed et al., 2019). Nonetheless, several studies are conducted to evaluate their reliability by comparing them to other global dataset product when gauge data is missing. Comparison has the majority of time the aim to assess the performance of dataset sources. And more interesting when it concerns historical gridded data, for an inter-annual period, the study for any variability and average of climate system is much more significant**.** (Eric S, Masocha, & Dube, 2019)

II.2.1 GPCC (Global Precipitation Climatology Centre)

Available for the whole world, the Global Precipitation Climatology Centre provides gridded gauge-analysis products derived from quality controlled station data. [4] It has been established in 1989 on request of the World Meteorological Organization (WMO) and it is operated by Deutscher Wetterdienst (DWD, National Meteorological Service of Germany) as a German contribution to the World Climate Research Program (WCRP) (Schneider, Becker, Ziese, & Rudolf, 2014).

The aim of the GPCC is to serve user requirements with regard to accuracy of the gridded precipitation analyses and timeliness of the product availability. GPCC products, gauge-based gridded precipitation data sets for the global land surface, are available in spatial resolutions of 1.0° latitude by longitude, depending on the product additional spatial resolutions of 0.25°, 0.5° and 2.5° are available. GPCC's new global precipitation climatology V.2018 (available in 2.5°, 1.0°, 0.5° and 0.25° resolution) is based on data from more than 79,000 stations. (Schneider et al., 2014). In Africa, GPCC collects additional monthly precipitation data from national meteorological and hydrological services or research projects. The total of the data used for Africa for each month is, therefore, typically about 2600 grid points with about 540 contributing stations (in 2000). The data compiled from different sources at GPCC are methodically and rigorously checked, analyzed, and merged before being gridded**.** (Adeyewa & Nakamura, 2003)

[4] National Center for Atmospheric Research Staff (Eds). Last modified 20 Sep 2018. "The Climate Data Guide: GPCC: Global Precipitation Climatology Centre." Retrieved from https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre.

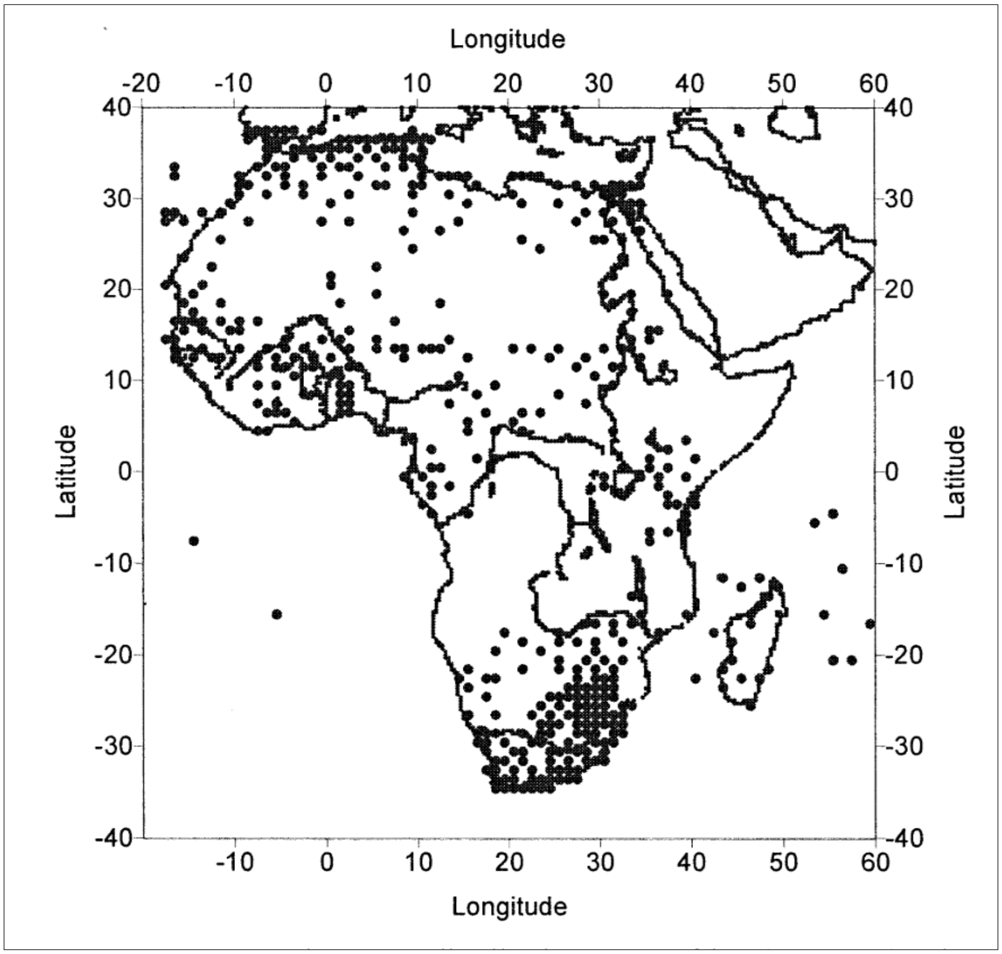


Figure II 1 *: GPCC rain gauge distribution over Africa (Jan 2000) in 1.0º x1.0º grid cell (Adeyewa & Nakamura, 2003)*

II.2.2 TRMM (The tropical rainfall measuring mission)

From the official website of NASA[5] The Tropical Rainfall Measuring Mission (TRMM) is a joint mission between NASA and the Japan Aerospace Exploration (JAXA) Agency to study rainfall for weather and climate research. TRMM delivered a unique 17 years dataset of global tropical rainfall and lightning since its launching in 1997.

TRMM observed rainfall rates over the tropics and subtropics principally, where two-thirds of the world's rainfall occurs and carried the first precipitation radar flown in space, which returned

[5] https://trmm.gsfc.nasa.gov

data that were made into 3-D imagery, enabling scientists to see the internal structure of storms for the first time.

TRMM's original flight altitude was optimized for the precipitation radar. To obtain precipitation profiles through the depth of the lower atmosphere and to concentrate the measurements in the tropics, the orbit was confined to 35 degrees north to 35 degrees south latitude at an altitude of 350 km.

TRMM is designed to measure rain rates from space using a combination of high resolution radar (PR), passive microwave radiometer (TMI) and Visible-Infrared Radiometer Scanner (VIRS)) measurements from a spacecraft in a rapid precession (Adeyewa & Nakamura, 2003) about 350 km orbit inclined at 35 deg. (Figure 1) These measurements, averaged over a 500 km grid for a month, are expected to provide monthly mean rainfall to an accuracy of 10 – 15 percent (Theon, Aeronautics, & Administration, 1994).

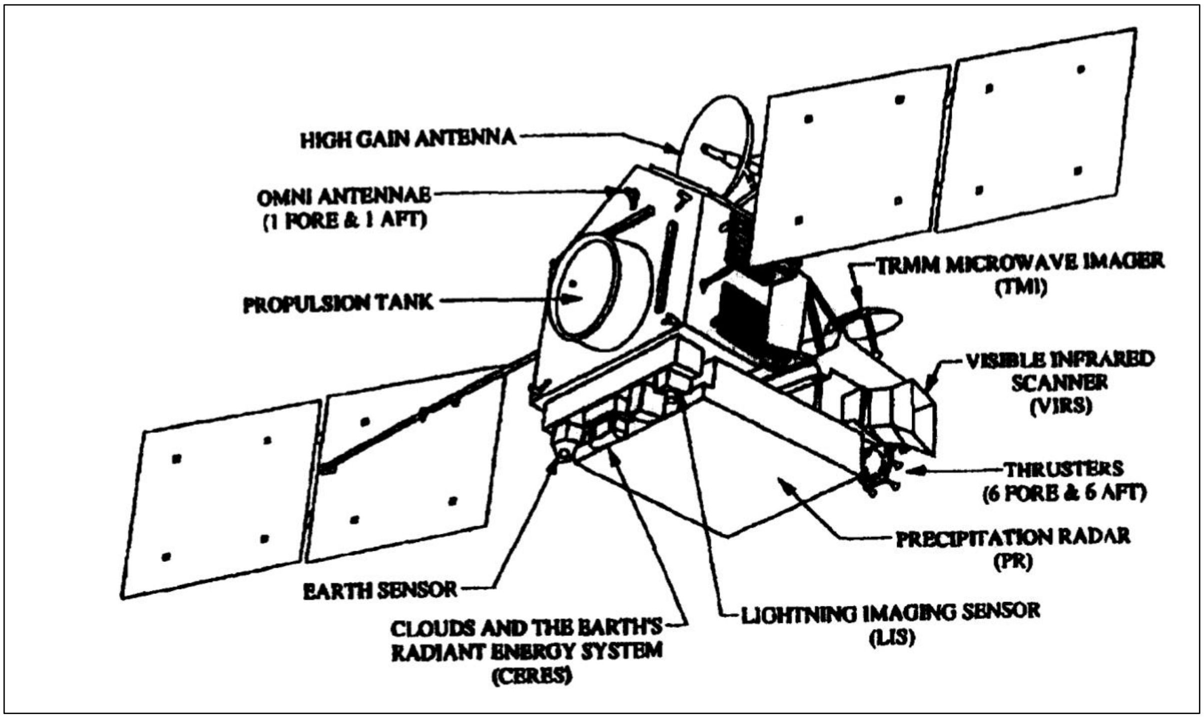


***Figure II 2**: TRMM sensors locations (Theon, 1994)*

.

II.2.3 TAMSAT (Tropical Applications of Meteorology using SATellite data and ground-based observations)

Established by the University of Reading in 1977, TAMSAT enhances the capacity of African meteorological agencies and other organizations by providing and supporting the use of satellite-based rainfall estimates and related data products. [6] All TAMSAT data are released for operational, research and commercial use under a creative commons license and it produces daily rainfall estimates for all Africa at 4km resolution based on high-resolution thermal-infrared observations, and available from 1983 to the present. The datasets are based on the disaggregation of 10-day version (v2.0) and 5-day version (v3.0) total TAMSAT rainfall estimates to a daily time-step using daily cold cloud duration. (Black et al., 2017) ,both TAMSAT-2 and TAMSAT-3 are derived from the TAMSAT rainfall estimation algorithm which is based on two primary data inputs: Meteosat TIR imagery provided by The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) and rain gauge observations for calibration.

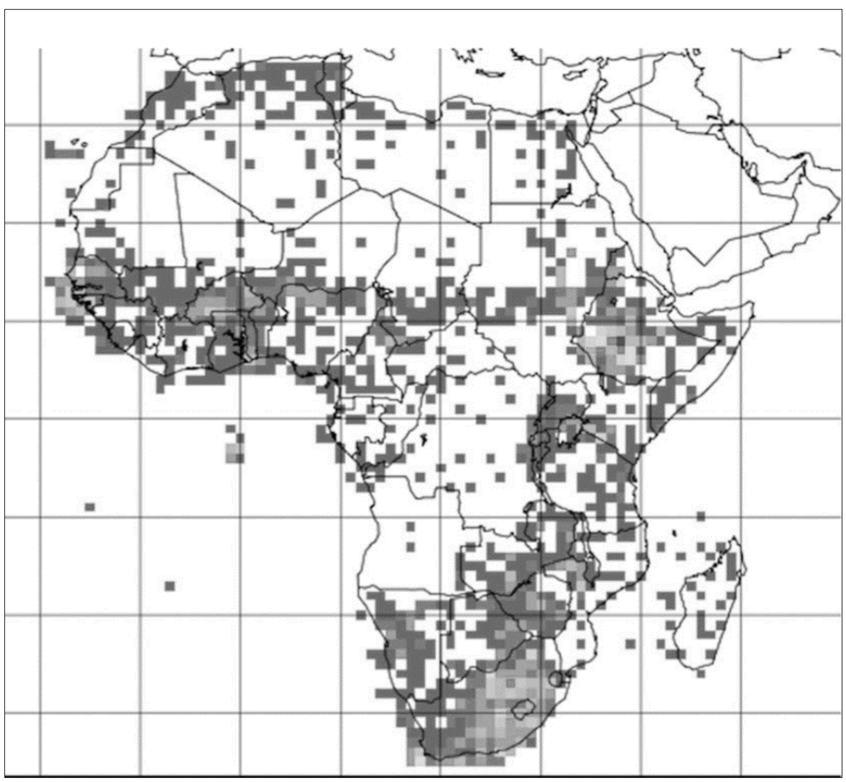


Figure II 3: *TAMSAT rain gauge distribution over Africa in 1.0º x1.0º grid cell (Black et al., 2014)*

[6] Official website to TAMSAT dataset https://www.tamsat.org.uk

Besides TAMSAT there are similar satellites products as (National Oceanic Administration Climate Prediction Center (NOAA-CPC), African Rainfall Climatology (ARC) and Climate Hazards Group Infrared Precipitation with Station (CHIRPS) are specifically designed for rainfall estimation over Africa (**Table II.1**).

Table II 1*: Satellite-based rainfall datasets covering 30years and issued operationally over Africa*

Dataset	Spatial extent	Temporal extent	Data input	Spatial resolution	Temporal resolution
GPI	40°N–40°S	1986–present	TIR	2.5°	Monthly
GPCP	Global	1979–present	TIR, PMW, gauge	2.5°	Pentad, monthly
NOAA-CPC ARC	40°N–40°S 20°W–55°E	1983–present	TIR, gauge	0.1°	Daily
CHIRPS	50°N–50°S	1981–near present	TIR, gauge	0.05°	Pentad
TAMSAT	Africa	1983–present	TIR, gauge	0.0375°	Dekadal

II.2.4 CRU (The Climatic Research Unit)

The Climatic Research Unit (CRU) was established in the School of Environmental Sciences (ENV) at the University of East Anglia (UEA) in Norwich in 1972.[7] Their datasets are used world-wide for monitoring climate change, understanding processes, evaluating climate models, and the study of natural and anthropogenic climate.

Climate data (temperature, precipitation and droughts, Pressure and Circulation Indices) provided by CRU are obtainable in different high-resolution gridded dataset, any format , updated in many versions (direct link to access) https://crudata.uea.ac.uk/cru/data/hrg/ ,accessible for free of charge, covering the majority part of the world and available since 19th century which make CRU datasets very important for long term and historical studies.

Historical gridded dataset are calculated taking into consideration different references period where gauge stations are interpolated via anomaly interpolation method which required the station values to be converted into percentage anomalies. In the methodology part of this project the interpolation method will be more detailed.

[7]Official web site to CRU (http://www.cru.uea.ac.uk/web/cru/about-cru/history)

II.2.5 HSM-SIEREM (HydroSciences Montpellier- Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation)

Montpellier Laboratory of HydroSciences, HSM, has developed a database of monthly gridded rainfall data for Africa at 0.5 º x 0.5 º resolution over the period 1900-2000, available for free on the SIEREM site [8] (Rouché et al., 2010)**.** Those grids were interpolated from several thousand identifications rainfall stations (Paturel, Boubacar, L'Aour, & Mahe, 2010) which represent approximately 6 204 pluviometric stations over 46 African countries **(Figure II.5)** (excluding Madagascar and the islands) (Rouché et al., 2010) taking into consideration also that in 2010 Soudan was not separated with South Soudan).

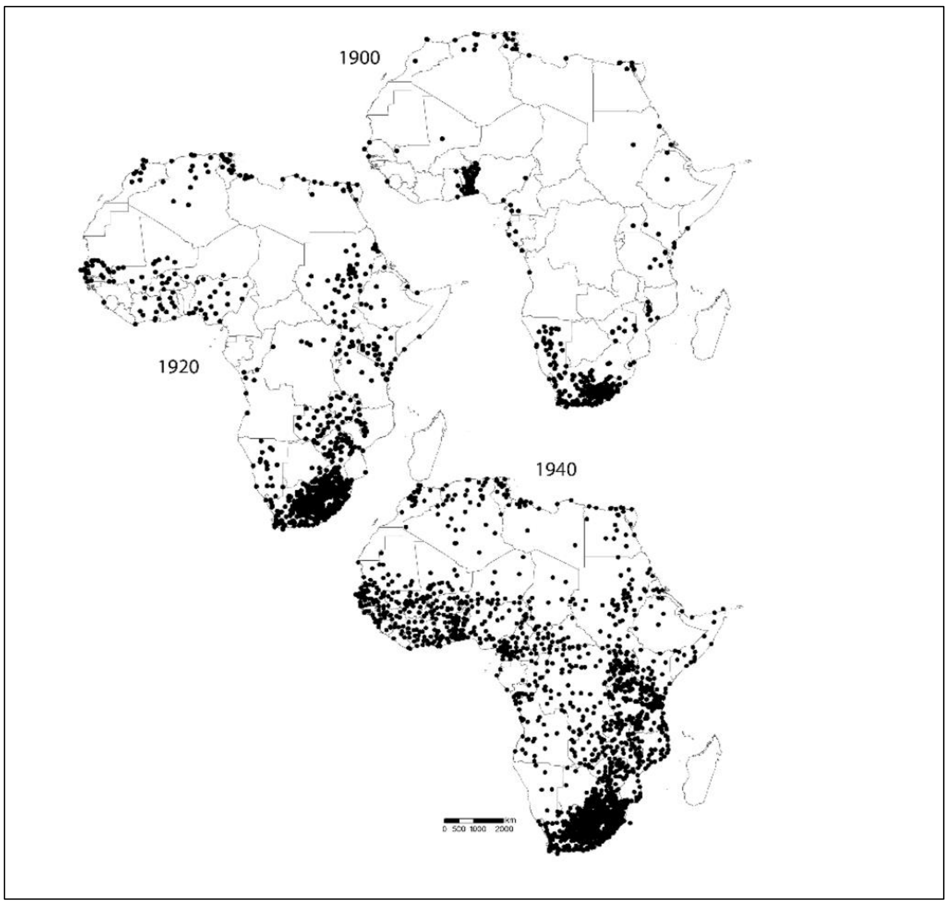


Figure II 4 *: Evolution of the number of measurement stations in 1900, 1920, and 1940, in the HSM_SIEREM rainfall database.(Dieulin et al., 2019)*

[8] (http://www.hydrosciences.fr/sierem/)

HSM database was created in 1943, ORSOM/IRD's mission was to promote scientific cooperation and education in Western and Central Africa. In this framework, hydrologists installed 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**.**(Dieulin et al., 2019)

The development of HSM-SIEREM started in 1999, where hydrologists gathered and managed all the climatological data archived in every country, and it was done with two simultaneous phases: data collection and the system analysis and design. The challenge was to build a system with both chronological data and spatialized information (such as soil and DEM layers). A team of engineers starts three tasks: collection of hydro-meteorological data and metadata, collection and building-up of geographical information, and the required data homogenization and integration into an environmental system built using a specific method of system analysis (Boyer et al., 2006)**.**

The chronological data are measurements done at one time in one place. All the object modelling which describes such type of data is defined and the description of the place, the date of the measurement, the person who measured out, and the owner of the data. (Boyer et al., 2006)

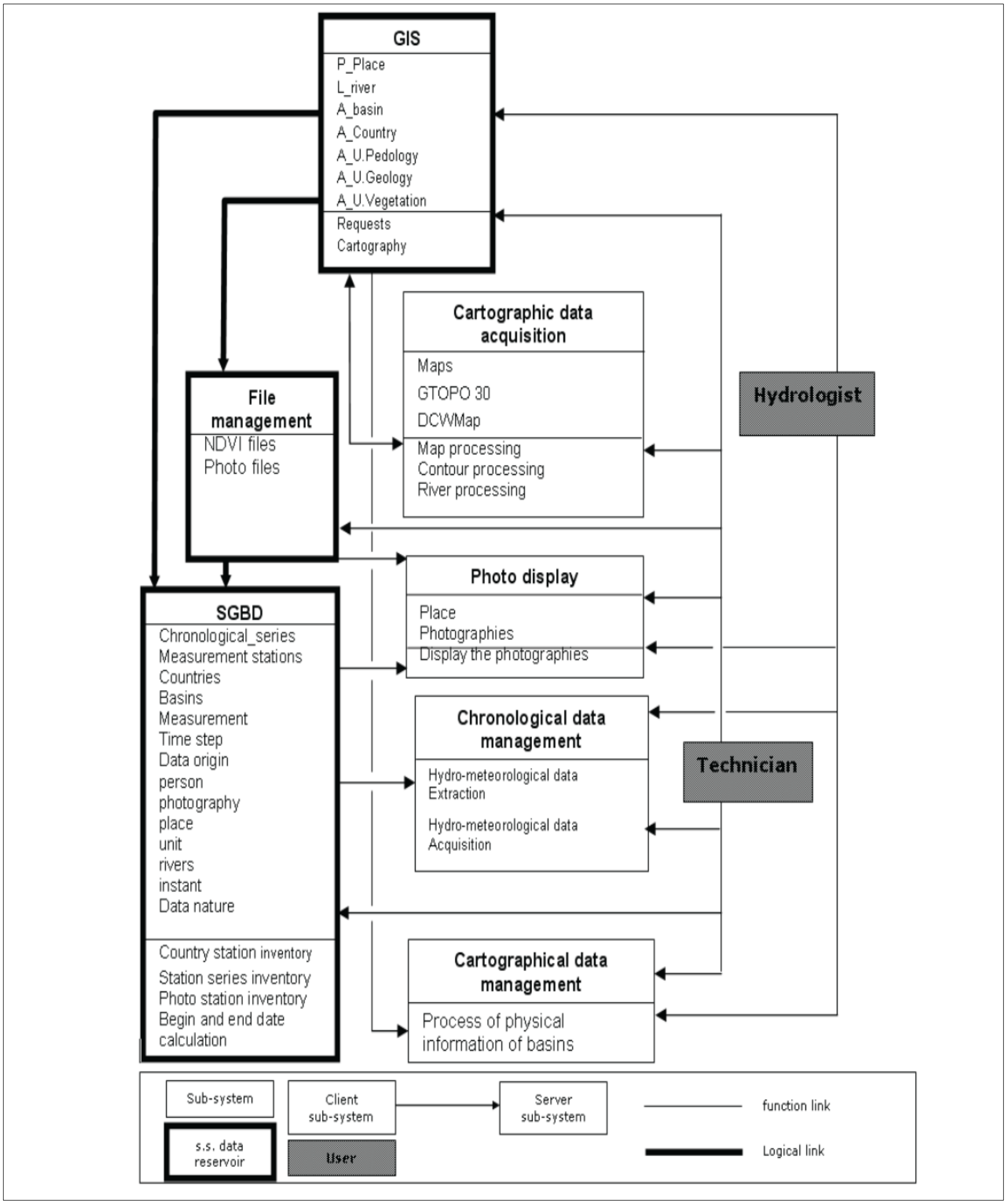


Figure II 5: *SIEREM system diagram (Boyer et al., 2006)*

II.2.6 University of Delaware Air Temperature & Precipitation

Two scientists Cort Willmott & Kenji Matsuura from the University of Delaware put data together from a large number of stations, both from the GHCN2 (Global Historical Climate Network) and, more extensively, from the archive of Legates & Willmott.

With support from NASA's Innovation in Climate Education (NICE) Program, The result is a monthly climatology of precipitation and air temperature, both at the surface, and a time series, spanning 1900 to 2010, of monthly mean surface air temperatures, and monthly total precipitation. It is land-only in coverage, and complements the ICOADS (International Comprehensive Ocean-Atmosphere Data Set) data set well [9] . Spatial coverage of the gridded data is at 89.75N - 89.75S, 0.25E - 359.75E with a resolution of 0.5 degree latitude x 0.5 degree longitude.

All data are available in NOAA website, referenced and detailed.

For the climate data “precipitation” the source of data are station data, compiled from several updated sources including a recent version of the Global Historical Climatology Network (GHCN2); the Atmospheric Environment Service/Environment Canada; the Hydrometeorological Institute in St. Petersburg, Russia (courtesy of Nikolay Shiklomanov); GC-Net data (Steffen et al., 1996); Greenland station records from the Automatic Weather Station Project (courtesy of Charles R. Stearns at the University of Wisconsin-Madison); the National Center for Atmospheric Research (NCAR) daily India data Reanalysis 1; Sharon Nicholson’s archive of African precipitation data (2001); Webber and Willmott’s (1998) South American monthly precipitation station records and the Global Surface Summary of Day (GSOD) (NCDC). (Willmott, C. J. and K. Matsuura, 2001)

About the process of compilation station records that had the same geographical coordinates were interleaved or blended to create a single, station time series for that location. For some data sets, monthly values were derived from daily values; that is, when the number of missing days was no more than five days during a month. If there were two or more station observations (monthly totals) for a given month, the median of these observations was taken as *P* for that month. When there was only one station observation for a month, it was taken as *P* for that month. This

[9] (https://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html#detail)

was done to make use of all available data. Observations from stations which had different geographical coordinates were assumed to belong to different station records, although sometimes parts of nearby station records were extremely similar. The resultant number of stations used for estimating monthly total precipitation ranges from about 4,100 to 22,000 globally. (Willmott, C. J. and K. Matsuura, 2001)

As mentioned above Sharon Nicholson's archive of African precipitation data was used to collect data for the African continent, The most of data collected are from regular reporting stations established by colonial governments but generally abandoned with the departure of the colonial powers early in the twentieth century, Larger numbers of such stations existed in countries such as Togo and Tanzania, under German control, and Malawi, under British control Other type of data collection consists of temporary nongovernmental observations, usually made at mission stations, explorers' camps. (Nicholson, Dezfuli, & Klotter, 2012).

II.3 Comparisons and Conclusions

Despite all these datasets, there are many difference which exist between them:

1. Difference in the type of providing data :

The gridded datasets can be divided into three categories: (Gauge-based as HSM-SIEREM, satellite-based TRMM, merged satellite-gauge products as CRU).

2. Difference in temporal resolution:

Some of dataset products produce timely scale and others daily or monthly data.

3. Difference in spatial resolution:

Some of them produce data at 0.25 degree, others are 0.5 degree and 1 degree.

4. Various weather variables:

Some of dataset products just provide precipitation data.

5. The period of recording or data collection:

Mainly datasets products using gauge-based have history gridded data from the end of 19^{th} century, first meteorological satellites were launched towards the end of the fifties (1959).

METHODOLOGY AND MATERIALS

METHODOLOGY AND MATERIALS

III.1 Presentation of study area

III.1.1 Introduction

Most of the information for the description of the study area below about main apparent elevations, rivers, and climate zones in Africa were extracted from two atlas named “Africa: Atlas of our changing environment” and “Africa: Water Atlas realized by (UNEP) United Nations Environment Programme in 2008 and 2010.

Provided by IRD[10] and HSM a map of annual rainfall over Africa for a period between 1940 and 1999 (60 years) will be presented in this chapter.

The representation of Africa’s continent and topographic elevation maps were obtained using online data available in software Global Mapper (V18.02.0) or Google earth Pro (7.3.2.5776), using coordinates (in decimal degrees) for the African continent which are: North: 38º South: -35º West: -18 º East: 52º.

III.1.2 Africa the Geography

Africa is the second-largest continent in the earth, at about 30.3 million including adjacent islands, it covers 6% of Earth's total surface area and 20% of its land area. Africa is surrounded by the Red sea to the northeast, The Indian Ocean the southeast, the Atlantic Ocean to the west and the Mediterranean Sea to the north (Sayre, April Pulley 1999**)**. There are 54 countries located in Africa and by land area, Algeria is the largest country on the continent. By population, Nigeria is the largest [11]

As a continent, Africa held world record with the oldest continent in the Earth, second large continent after Asia, longest river which is The Nile (about 6693 km by length), and largest desert (with about 9 million Km square) which dominates the Northern of Africa.

---

[10] http://www.cartographie.ird.fr/images/carte_pluie/PLUVIO_AFR_web.jpg

[11]From site WorldAtlas https://www.worldatlas.com/articles/how-many-countries-are-in-africa.html (07/2019)

Figure III 1: *Representation of the Africa continent on Google Earth Pro (7.3.2.5776)*

.

III.1.3 Africa (Elevation and Mountains)

Mountains in Africa generally occur as widely scattered exceptions to the plateaus and plains that dominate the landscape (Taylor 1996). The Atlas Mountains (Shown in the Figure **III.2** with red frame) are the most northerly of Africa's mountain chains, extending 1 610 km across Morocco, Algeria, and Tunisia (UNEP. 2008) Extending northeast to southwest, they rise to a maximum height of 4 167 m.

Mount Kilimanjaro is a volcano in East Africa. At 5 895 m, Uhuru Point on the mountain's summit is the highest point on the African continent (Shown in the Figure **III.2** with yellow frame)

The Drakensberg Mountains ("Dragon's Mountains") are the highest in southern Africa, rising to an elevation of 3 482 m at Thabana Ntlenyana.

In blue frame The Ethiopian Highland (Rift Valley).

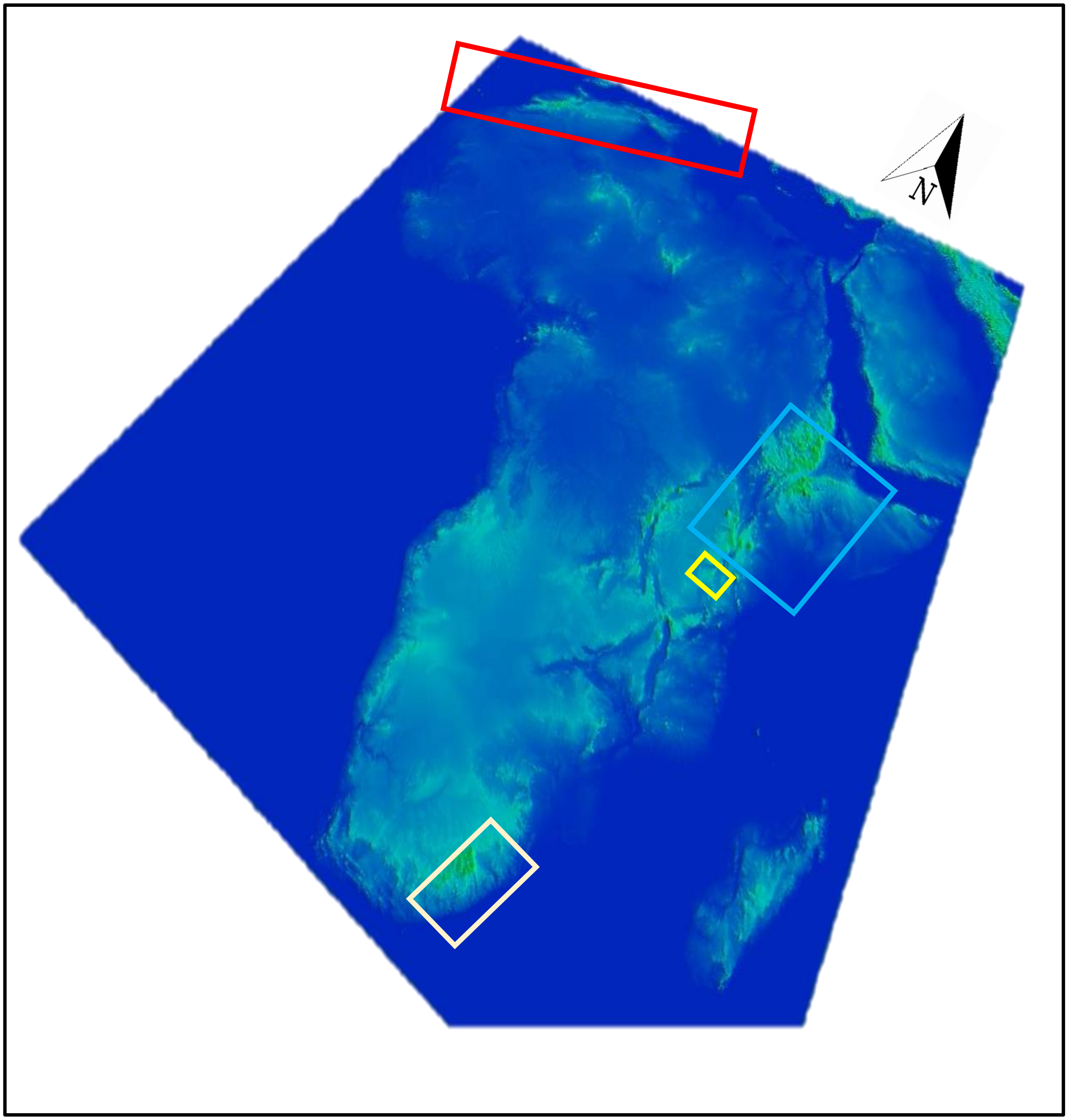


***Figure III 2**: 3D Representation of Topographic Elevation in Africa made on Global mapper (V18.02.0)*

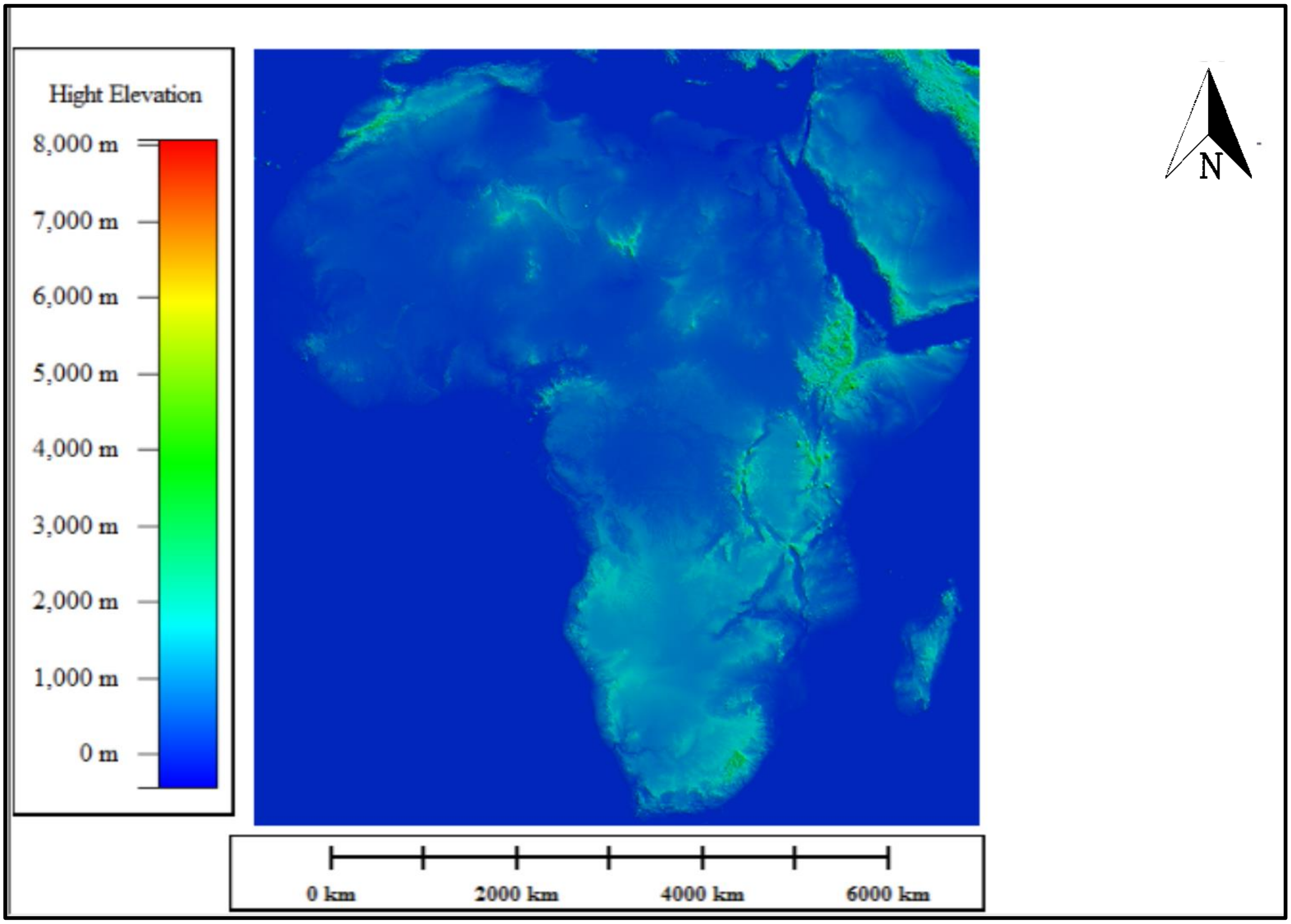


Figure III 3*: Map of height Elevation in Africa made on Global mapper (V18.02.0) using ASTER GDEM v2 World Wide Elevation Data (1 arc-second resolution)*

III.1.4 Africa Water Towers and their distribution

Mountainous and other elevated areas in several African watersheds contribute disproportionately to the total stream flow of Africa's major rivers. These areas generally receive more rainfall than their lower surroundings. They also usually lose less water to evapotranspiration because temperatures are lower. Downstream areas often benefit from the abundant runoff. Rivers such as the Nile, the Niger, the Senegal and the Orange flow from relatively rain-abundant areas to areas that would otherwise be too arid to support much life. These important, high-elevation watersheds have been referred to as "the water towers of Africa" for the role they play in supplying millions with life-giving water. The Millennium Ecosystem Assessment (MEA) states that mountains act as water towers by storing water in glaciers, snow-packs, soil or groundwater (MA 2005) **(UNEP, 2010).**

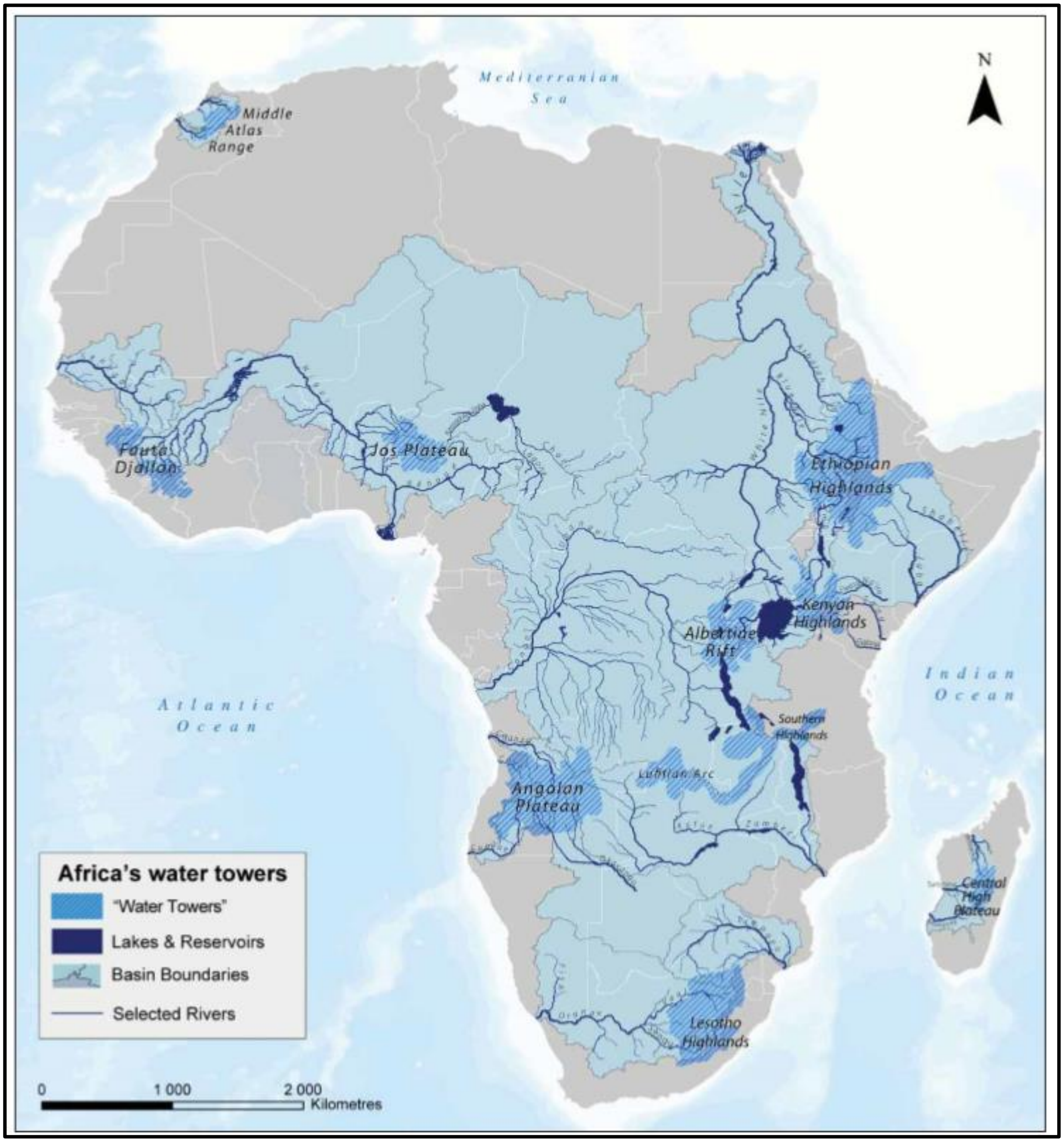


Figure III 4: *Map taken from (UNEP, 2010) Representing Africa's "water towers." Identified by relative elevation (generally 200–800 m above the surrounding area); precipitation above 750 mm; and runoff above 250 mm. They were also selected for the contribution they make to water resources for populations beyond their delineated boundaries*

.

There are wide differences in natural water distribution within Africa's sub-regions. Central and Western Africa are endowed with the highest proportions at 51 and 23 per cent respectively, while the share is as low as 3 per cent for Northern Africa.

III.1.4 Africa annual rainfall variability

Monthly and daily data from about 6000 rain gauge stations were collected and processed to establish a first Africa annual rainfall map between period of 1940 and 1999 carried out at the joint research Unit HSM for the data processing and the cartographic department of IRD center of Bondy for the map. Those data are part from historical database of IRD, additional data were collected from National Meteorological Services, From FRIENDS Program of Unesco (international Hydrological Program), from CRU (Climatic Research University of Anglia). (Mahé &.al 2012)

The map (**Figure III.5)** present variability of precipitation over Africa (period between 1940 and 1999) for a minimum of 200 mm to a maximum of approximately 4000 mm of precipitation.

Precipitation distribution over Africa match mainly to climatic zones where humid tropical zones and equatorial zones have the highest precipitation rate, while Sahelian and Mediterranean zones received for the same period less rate of precipitation.

Another interesting study was carried out recently by **(Nicholson, Funk, & Fink, 2018)** about Inter-annual-rainfall over the African continent from the 19th through the 21st century, once resuming the process used for the collection of data, the sources of data, periods observed depending on availability of continuous time series and the distribution of sectors (**Figure III.6)** Over Africa or as named in the article "the regionalization process" used as part of the quality control procedure, Results showed that rainfall is highly seasonal in each of the sectors, with a unimodal distribution (one clear peak) in extra-tropical and subtropical sectors and a bimodal distribution in the low latitudes. In the Mediterranean sectors rainfall peaks in the boreal winter. In the sector just south of the Sahara rainfall is similarly unimodal but with a peak in the boreal summer, in August on average. In the four equatorial regions rainfall tends to peaks in the boreal spring and autumn. In most of southern Africa, rainfall peaks in the austral summer/boreal winter.

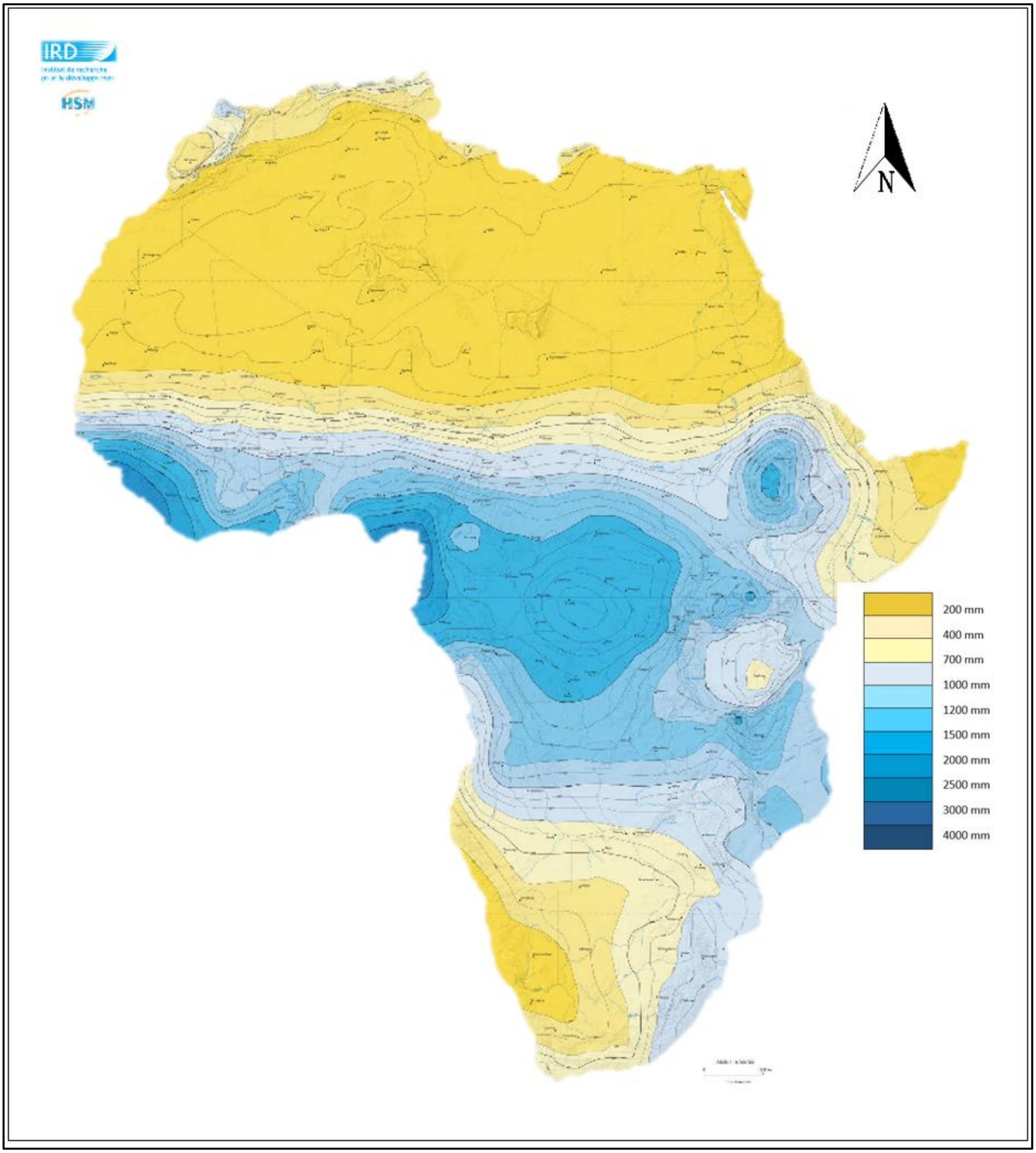


Figure III 5:*Annual rainfall map of Africa for the period between 1940 and 1999 (Mahé &.al, 2012)*

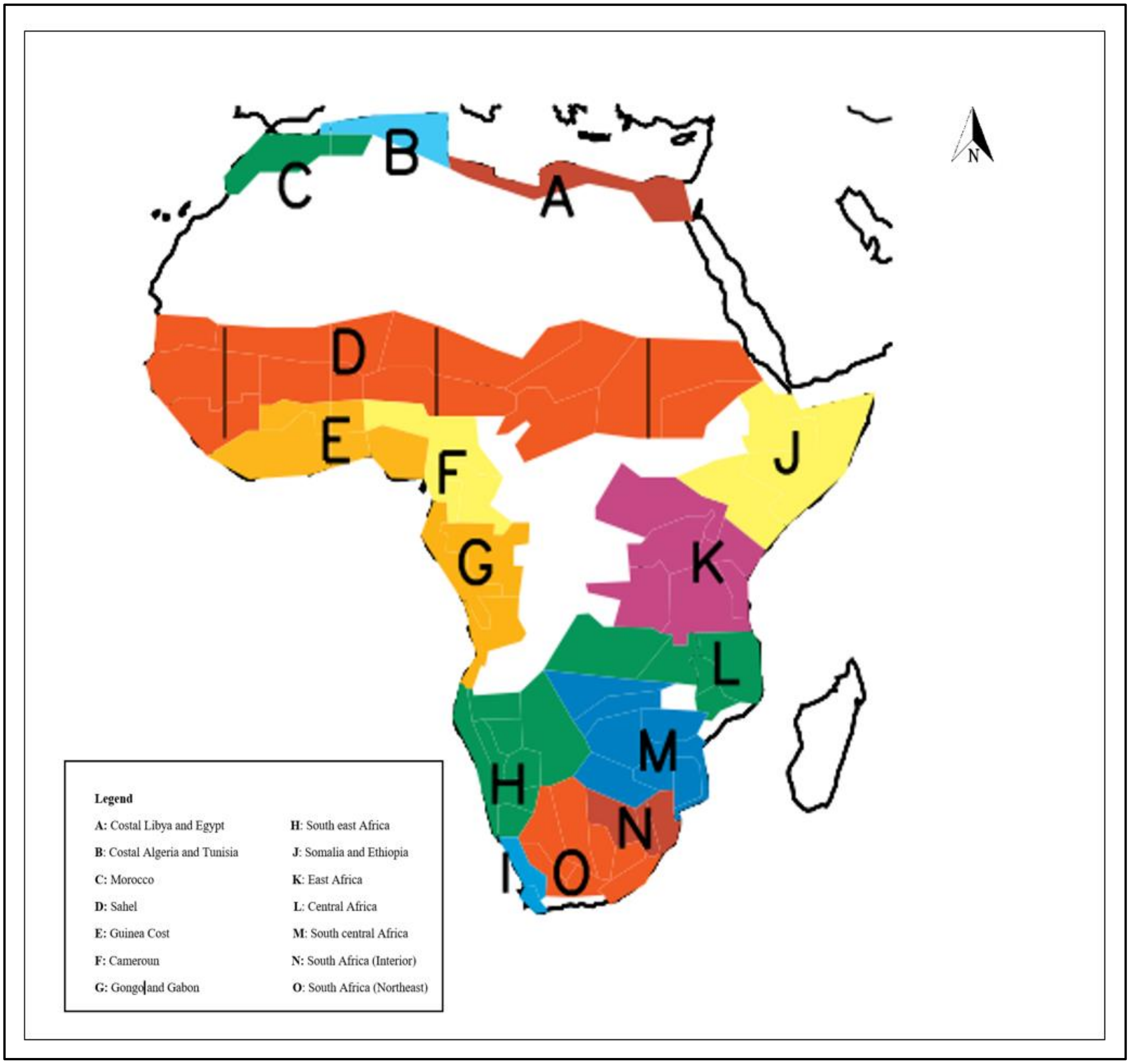


Figure III 6: *Distribution of sectors and sub-sectors over Africa for the inter-annual rainfall study (by (Nicholson et al., 2018)*

III.1.5 Africa Climatic zones

Africa's climate is predominantly tropical, with the majority of the continent having mean temperatures above 21 degrees Celsius for nine months of the year (Goudie 1996). Moving away from the equator, climate zones vary in nearly mirror-image patterns[12] to the north and south. These patterns are not interrupted by the climatic influence of long mountain ranges (Goudie 1996). (UNEP, 2010)

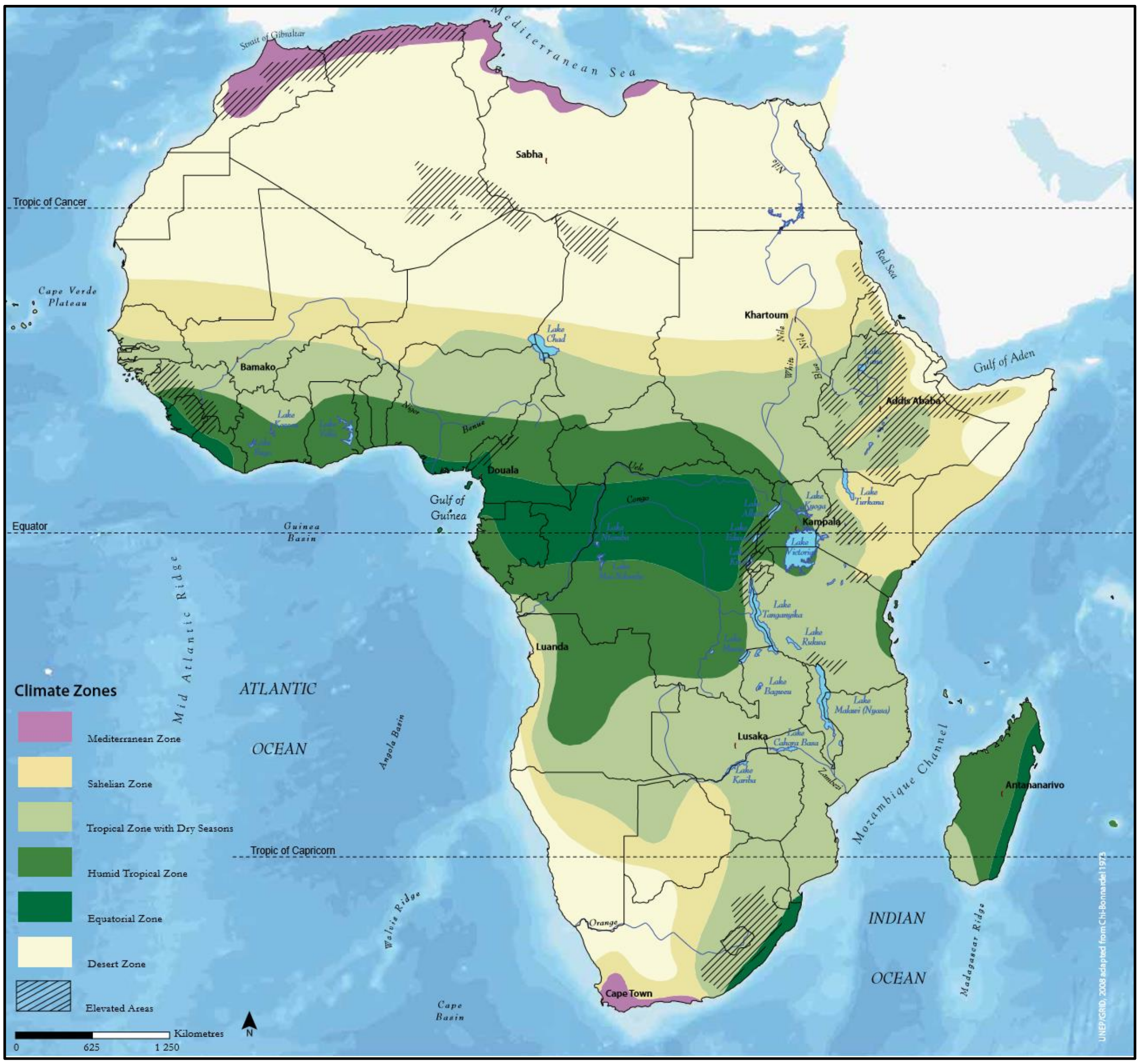


Figure III 7: *Different climatic zones in Africa (UNEP, 2008).*

[12] It is reversed with reference to an intervening axis or plane (which is Equator)

III.1.5.1 Mediterranean Zone:

At the northern and southern extremes of Africa, there are zones of Mediterranean climate with hot dry summers and wet mild winters (ChiBonnardel 1973). A few locations receive as much as 700 mm of precipitation, but most receive less than 500 mm. In the summer, temperatures typically average around 25 degrees Celsius; however more inland locations often see freezing temperatures in the winter, especially at higher elevations (ChiBonnardel 1973). **(UNEP, 2010)**

III.1.5.2 Sahelian Zone

Only about 250 to 500 mm of rain falls in the Sahelian climate zone (Stock 2004; FAO 2001). With considerable seasonal and inter-annual variation in rainfall, before the spring rains, daily maximum temperatures often reach 40 degrees Celsius (Chi-Bonnardel, 1973). Average annual temperatures in the Sahelian climate zones adjacent to the Namib Desert are several degrees cooler (CRES 2002).

III.1.5.3 Humid Tropical

The humid tropical zone exhibits peaks in precipitation and a short dry season. The average annual rainfall generally ranges between 1 100 mm and 1 800 mm in this zone (FAO 2001). Temperatures are relatively high, but with somewhat more seasonal variation than temperatures in the equatorial zone (Goudie, 1996)

III.1.5.4 Equatorial Zone

Africa's equatorial climate zone is found along the equator from Gabon to Uganda, as well as in coastal Liberia and Sierra Leone and in eastern Madagascar. In this zone, rain falls throughout the year; if there is a dry period, it is very brief (Goudie 1996). Average annual rainfall generally exceeds 1 700 mm and reaches 3 000 mm at points along the Liberian and Sierra Leone coasts and in eastern Madagascar (FAO 2001). Mean annual temperatures are high, around 25 degrees Celsius, with very small variation throughout the year (Stock 2004)

III.1.5.5 Desert Zone

Africa’s desert climates receive little precipitation and in the case of the Sahara, daytime temperatures can be extremely high. Average annual precipitation is scant, exceeding 100 mm only in a few areas and tending to be below 25 mm for much of the Sahara and the western edge of the Namib Desert in southern Africa. (World Meteorological Organization).

III.2 Data sources for the study

III.2.1 Criteria of selection of Gridded data

Comparison of different gridded data was for a selected historical period of 60 years (from 1940 to 1999), resolution of cells 0.5º x 0.5º degrees, for monthly time step, spatial coverage of whole Africa and gridded data available for the variable “precipitation”.

Furthermore HSM-SIEREM, three different datasets were selected for the comparison which are GPCC (Global Precipitation Climatology Centre), University of Delaware Air Temperature & Precipitation, and CRU (The Climatic Research Unit) see (**Table IV.1**).

Gridded data were downloaded online taking the uploaded one (latest version) if available and in format (NetCDF) except for HSM-SIEREM where they were already available in ASCII format from their official website.

- Link to download data from University of Delaware Air Temperature & Precipitation: https://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html#detail (April 2019)
- Link to download data from CRU (The Climatic Research Unit): https://crudata.uea.ac.uk/cru/data/hrg/ (April 2019)
- Link to download data from GPCC (Global Precipitation Climatology Centre): https://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html (April 2019)
- Link to download data from HSM-SIEREM (Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation): http://www.hydrosciences.fr/sierem/produits/Grilles/60PM.asp (March 2019)

NetCDF format is a file format intended primarily for storing arrays of multidimensional numbers ie big data (unlimited size).

III.2.2 Extraction of Gridded Data

The extraction of data was done by using the statistical analyst and free software and programming language "R" Version 1.2.1335. As the extraction requests specify input data (attributes and variables) from the metadata of each downloaded dataset the first step then consisted of displaying it , see **(Appendix 1,2,3).**

Some functions and loops were needed to select data for the period between 1940 and 1999, and for the latitudes and longitudes which correspond to the continent of Africa (longitude between -18° and 52 °), (latitudes between -35 ° and 37, 5 °).

(Figure III.8 and Figure III.9) shows an example of commands, functions, loops used for displaying metadata, extracting gridded data from CRU, and importing them to csv format to the specific repertoire.

During the extraction some issues were faced:

- New loop was used to get negative longitudes before meridian line for (GPCC dataset)
- Latitudes for CRU gridded data after importing them and brief check were found reversed.
- Gridded data from University of Dalawere were excluded from comparison in a fact for the same period we got had gridded data each two years and not every year over same period (1940- 1999).

Data were imported then in Comma-separated values format "CSV" for each month for selected period 1940-1999 and a total of 720 excel sheets (GPCC and CRU).

PRODUCT NAME	DATA SOURCE	VERSION	TEMPORAL AND SPACIAL RESOLUTION	COVERAGE	REFERENCE
HSM –SIEREM *(Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation)*	**GAUGE BASED DATA** Different observed data with no statistical use of reference period. From African National Services in charge of collection of hydro meteorological data in each African country.	/	Monthly grid over (period of 60 years) 1940-1999 Interpolated into spatial resolution of 0.5° x 0.5°, Using IDW (Inverse distance weighed).	**Africa**	BOYER- J F, DIEULIN C , ROUCHE N, CRES A, SERVAT E, PATUREL J E & MAHÉ G - 2006 SIEREM an environmental information system for water resources. 5th World FRIEND Conference, La Havana - Cuba, November 2006 in Climate Variability and Change – Hydrological Impacts IAHS Publ. 308, p19-25, 2006.
CRU *(Climatic Research Unit)*	**SATELLITE- GAUGE PRODUCT** CLIMAT, Monthly Climatic Data for the World(MCDW), World Weather Records(WWR), National Meteorological Services, Australian Bureau of Meteorology(BoM)	**CRU TS v4.01** support from a number of funders	Monthly grid over period (of 95 years) 1901-2016 spatial resolution of 0.5° x 0.5° Angular Distance Weighting (ADW) for gridding the	**Global**	University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D. (2017): CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016). Centre for Environmental Data Analysis, *04 December 2017*. doi:10.5285/58a8802721c94c66ae45c3baa4d814d0. http://dx.doi.org/10.5285/58a8802721c94c66ae45c3baa4d814d0

			monthly anomalies		
GPCC *Global Precipitation Climatology Centre*	**GAUGE PRODUCT** In situ rain gauge data. (Precipitation data provided by the NOAA/OAR/ESRL PSD)	**Full Data Product (V2018)**	Monthly grid over (period of 115years) 1891 to 2016 Interpolated into spatial resolution of 0.5° x 0.5°, Using SPHEREMAP method	**Global**	Schneider, Udo; Becker, Andreas; Finger, Peter; Meyer-Christoffer, Anja; Rudolf, Bruno; Ziese, Markus (2011): GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. DOI: 10.5676/DWD_GPCC/FD_M_V7_050
University of Delaware Air Temperature & Precipitation	**GAUGE PRODUCT** Recent version of the Global Historical Climatology Network. The Atmospheric Environment Service/Environment Canada The Hydro meteorological Institute in St.	**Version 3.01**	Monthly grid over (period of 110 years) 1900-2010 Interpolated into spatial resolution of 0.5° x 0.5° Spherical version of Shepard's algorithm, which employs an enhanced	**Global**	Willmott, C. J. and K. Matsuura (2001) Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999), http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html. (with support from NASA's Innovation in Climate Education (NICE) Program)

Petersburg, Russia (courtesy of Nikolay Shiklomanov) Greenland station records from the Automatic Weather Station Project (courtesy of Charles R. Stearns at the University of Wisconsin-Madison. The Global Surface Summary Day (GSOD) The National Center for Atmospheric Research (NCAR) daily India data Sharon Nicholson's archive of African precipitation data (2001) Webber and Willmott's (1998) South American monthly precipitation station records		distance-weighting method		

Table III 1*: Details about datasets extracted for the comparison.*

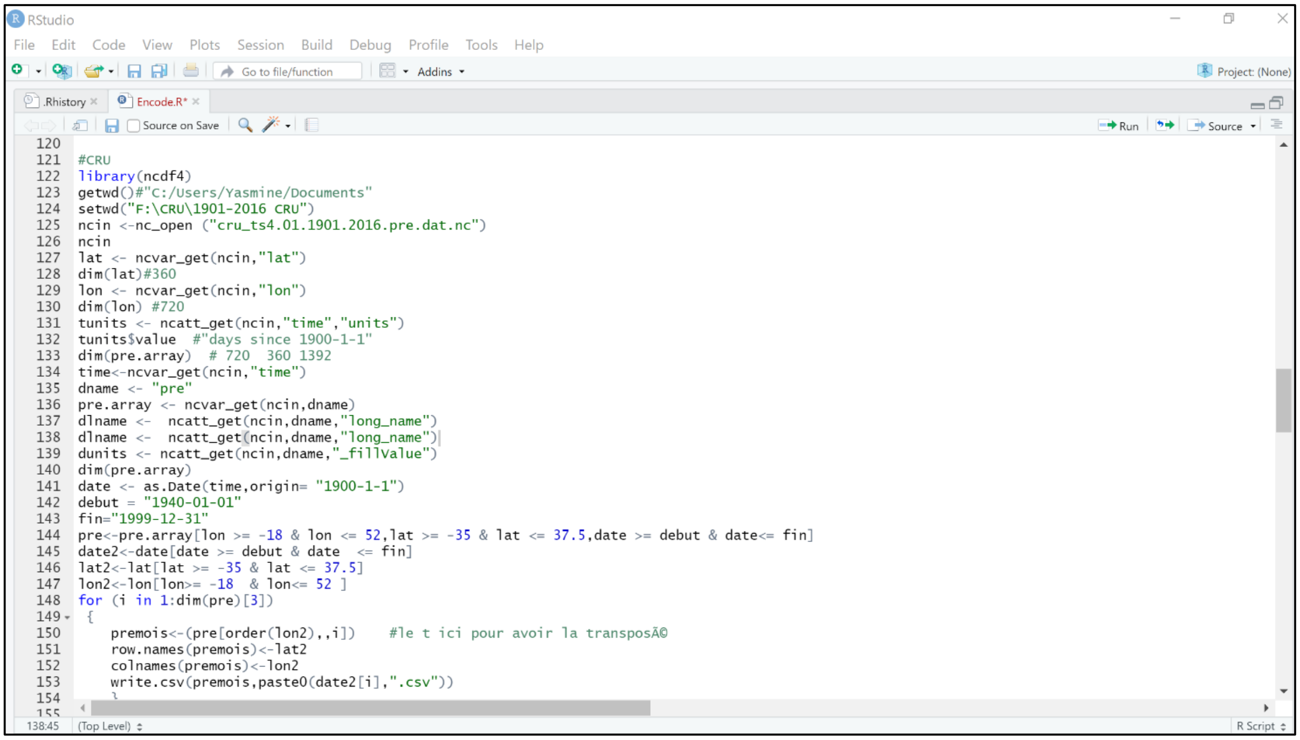
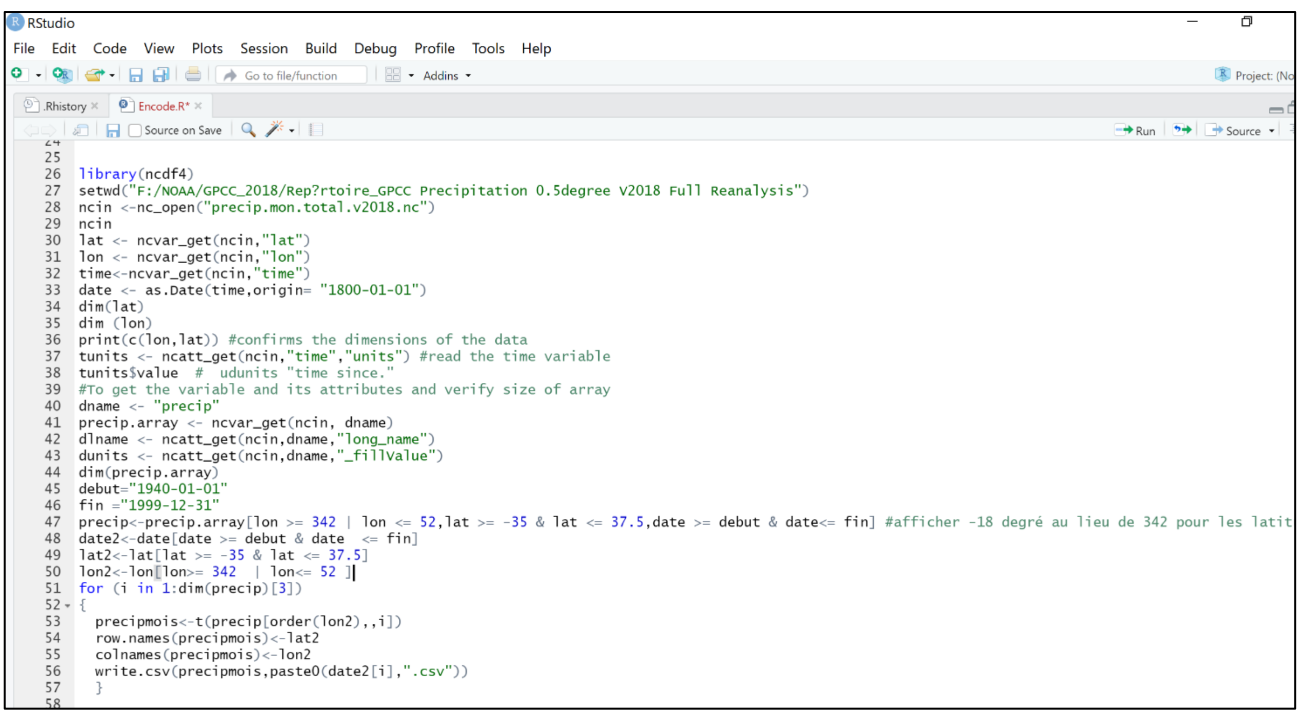


Figure III 8: *Screenshot of script used in R to extract gridded data from file CRU*



.

Figure III 9: *Screenshot of script used in R to extract gridded data from file GPCC.*

III.2.3 Choice of observed data

The choice of stations for the observed data was made only after the comparison of grid data (Chosen randomly) over Africa, indeed for a resolution of 0.5 ºx 0,5 º degrees we can find surfaces that they are not covered by stations and other surfaces with a density of stations and from which an average between values is requested.

In the other hand data have been retrieved for 60 years exclusively and only from HSM (collection of hydro meteorological data was proved by African National Services of hydro meteorological data) these stations (with a resolution of 0.5 x 0.5) do not cover each cell.

62 stations were used, which complete the longest time series during our study period between 1940 -1999 and the map below made on ArcGIS (10.3.0.43.22) shows their position over the continent.

Details about these stations (serial number of each station, name of the station, start and end date of precipitation recording, station type and latitude, longitude position) are available in **(appendix 7)**.

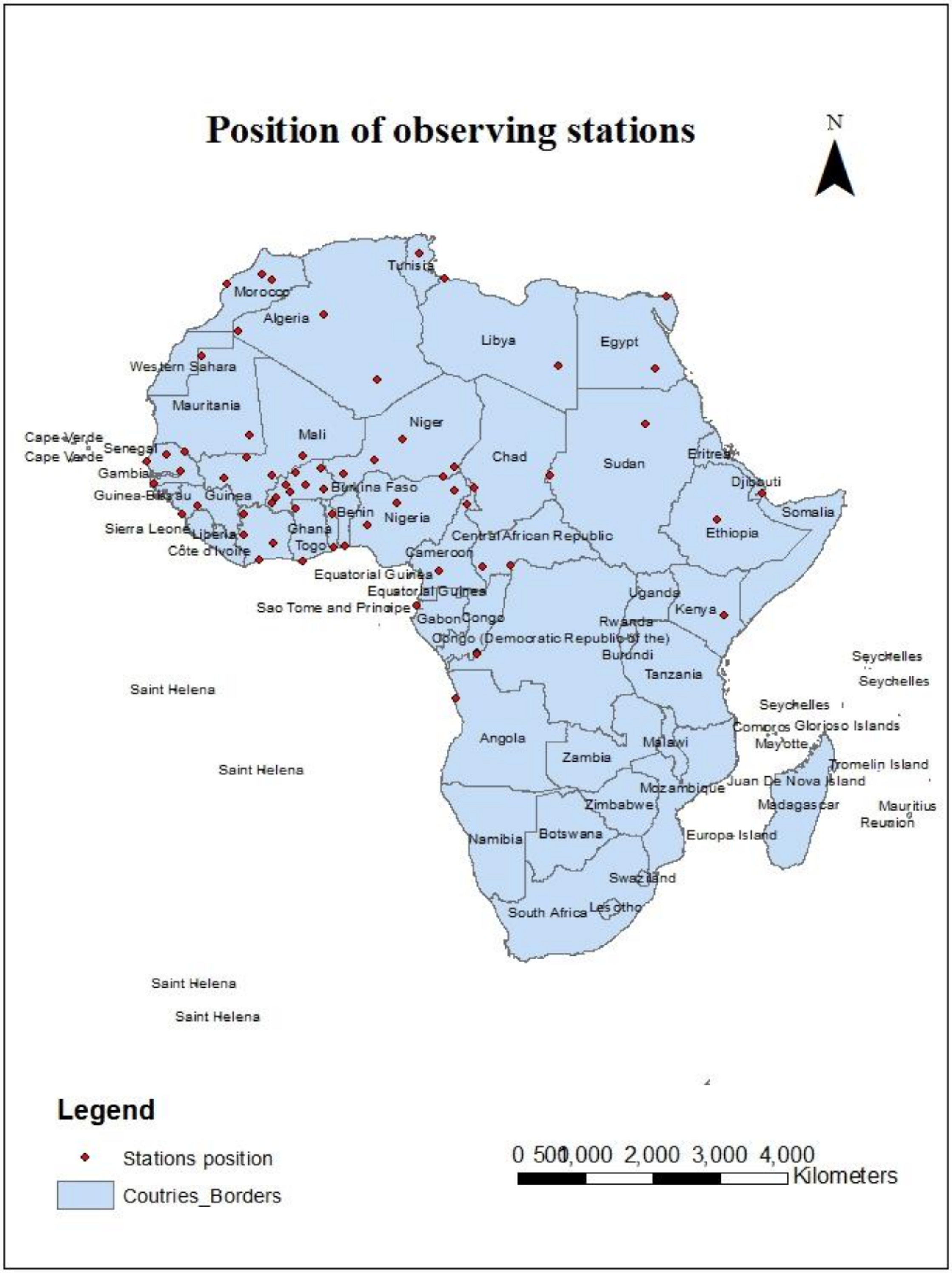


Figure III 10: *Position of observing stations over Africa.*

III.3 Method of interpolation

III.3.1 Spatial interpolation definition

Spatial interpolation technique have wide application in the design, construction and application of a GIS. It has continued to be an important tool for estimating continuous spatial environmental variables for effective decision making (Collins, F.C. and Bolstad, P.V, 1996). It is defined as predicting the values of a primary variable at points within the same region of sampled locations, while predicting the values at points outside the region covered by existing observations is called extrapolation (Li & Heap, 2014), In another definition it is to " find the function that will best represent the whole surface and that will predict values at points or for other subareas." (Lam, 1983) .The correct determination of the spatial distribution of meteorological variables is as important as their measurement. Depending on the spatial attributes of the data, the accuracy of the results may vary widely among spatial interpolation methods (Apaydin, Kemal Sonmez, & Yildirim, 2004).

The general interpolation equation is given by:

$$Z_0' = \sum_{i=1}^{n} Wi * \ Zi \qquad \text{With} \qquad \sum_{i=1}^{n} Wi = 1$$

Z_0': is the attribute value to be predicted at un-sampled site.

Zi : is the attribute value at the *i* point of the nearby locations.

Wi : is the weight assigned to the attribute at point *i*, should sum up to 1 (to be unbiased).

n : is the total number of nearby locations involved.

III.3.2 Key Spatial Interpolation methods

Different methods produce different spatial representations in different datasets; also, in-depth knowledge of the phenomenon in question is necessary in evaluating which of the interpolation methods produces results closest to reality (Ikechukwu, Ebinne, Idorenyin, & Raphael, 2017)

Spatial interpolation has many features 15 according to Goovaerts in 1997 but can be considered in brief, in point interpolation exact or approximate, local or global, gradual or abrupt and deterministic or stochastic, one method can fit several of these categories in the same time as Thiessen Polygons. Recently several novel hybrid methods have been developed from the machine learning field (Li & Heap, 2014) they are sometimes a combination of two interpolation methods. As a key of spatial interpolation methods, Waters in 1989 gives all details about their categorization (where figure below resume) see **Figure III.11**. Although, recently scientists have been able to classify several interpolation methods according to the field of need. The exact point interpolation methods are summarized as follows:

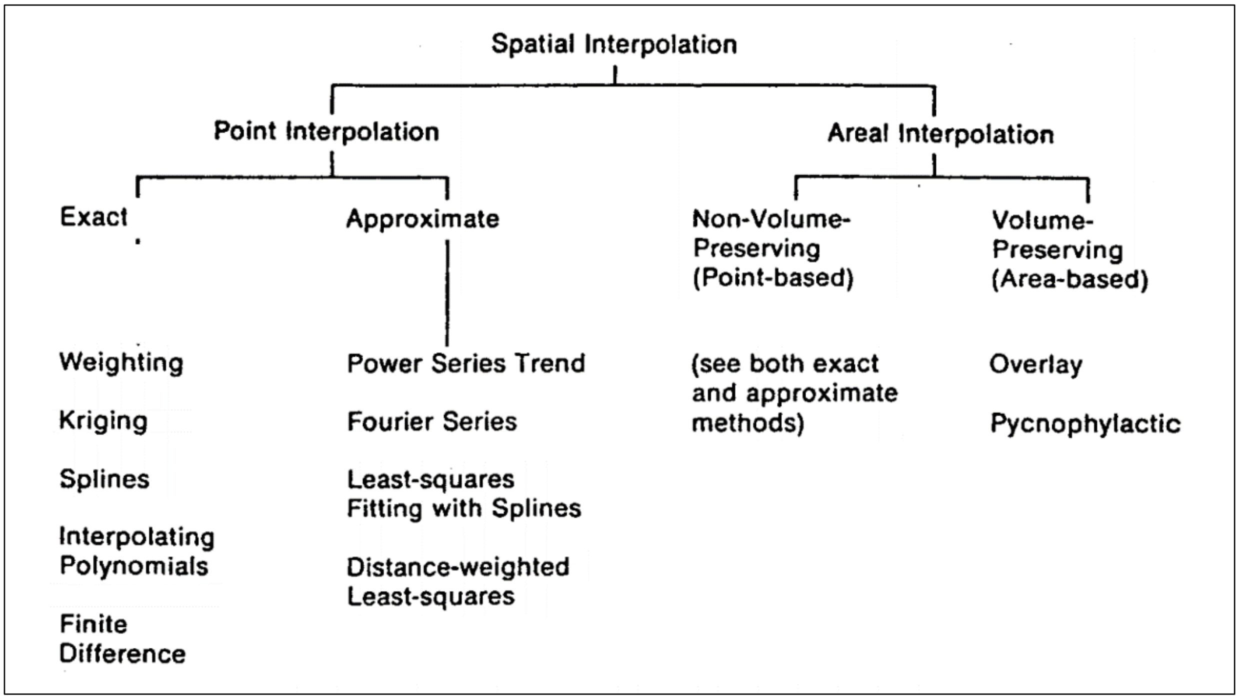


Figure III 11: *Diagram summarizing the different interpolation methods (Waters, 1989)*

III.3.2.1 Thiessen Polygons

The Theissen polygon technique was originally used for computing areal estimates from rainfall data, it is an exact method of interpolation that assumes the unknown values of the points on a surface to be equal to the value of the nearest known point. As a robust technique it will always produce the same surface from the same set of data points. This is a disadvantage and the technique is unintelligent and unable to respond to external knowledge about factors which may influence the values recorded at the observed data points**.**

III.3.2.2 the Triangulated Irregular Network (TIN):

The TIN technique is an exact and deterministic interpolation method, useful technique for irregularly spaced points, the known data points are connected by lines to form a series of triangles, and a simple linear equation can be used to calculate an interpolated value for any position within the boundary of the TIN. That methodology calculate real 3D distance between data points along vertices using trigonometry. The estimated surface is continuous but with abrupt changes at the margins of the triangles (Webster and Oliver, 2001).

III.3.2.3 Kriging:

Kriging is a geostatistical, exact, local and stochastic method for interpolation, it is known by the acronym BLUE which stands for" Best Linear Unbiased Estimator". (Rishikeshan, Katiyar, & Vishnu Mahesh, 2014) The term is derived from the name of D. G. Krige, who introduced the use of moving aver ages to avoid systematic overestimation of reserves in the field of mining (Krige, 1976) (Lam, 2008), this technique is used when data are sufficient to compute variograms, kriging provides a good interpolator for sparse data. Binary and nominal data can be interpolated with Indicator kriging.(Li & Heap, 2014).

III.3.2.4 Inverse Distance Weighting (IDW):

The IDW is an exact local deterministic interpolation technique, works well with regularly spaced data, and the procedure doesn't generate error assessment (Li & Heap, 2014). The IDW assumes that the value at an unsampled location is a Distance-weighted average of values at sampled points within a defined neighborhood surrounding the unsampled point (Burrough, P.A., McDonnell, R.A., 1998)

III.3.3 Selection of Spatial interpolation method

There is no simple answer regarding the choice of an appropriate spatial interpolation method, in a fact it depends on many factors regarding sometimes to influences which affect the performance of spatial interpolation method (data density, data nature and quality, spatial correlation, and distribution), many scientists proposed guidelines by developing classification tree of many interpolation method based on their features, others applicate two or three methods on the same data sample and arrange best interpolation method with reference to results, or take advantage on experience of studies made in the same topic field.

III.3.4 Interpolation methods used for datasets in our study

- ✓ The interpolation method completed by HSM-SIEREM is IDW, taking into consideration size of the region and heterogeneity of the distribution of rain gauge stations (Dieulin et al., 2019). In a previous study a comparison was carried out between two methodologies which are the IDW and the kriging on (dry and wet years only) over Africa for annual grid and observed data, the conclusion was that the IDW gap is not as higher as with the method of interpolation "kriging".
- ✓ GPCC employs a robust empirical interpolation method SPHEREMAP. The method constitutes a spherical adaptation (Willmott et al., 1985) of Shepard's empirical weighting scheme (Shepard, 1968) and takes into account: (a) the distances of the stations to the grid point (for a limited number of nearest stations), (b) the directional distribution of stations in relation to the grid point (in order to avoid an overweight of clustered stations), and (c) the gradients of the data field in the grid point environment.

  GPCC's monthly precipitation analysis products used for this study (V.2018) are based on anomalies from the climatological normals at the gauge stations on the period 1951-2000 which consists of normals from more than 79,000 stations. The climatology comprises normals collected by WMO, delivered by the countries to GPCC or calculated from time-series of monthly data (with at least 10 complete years of data). (University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D. 2017)
- ✓ The method of interpolation used by CRU for the latest version (CRU TS 4.0) knows main process change which is the move to Angular Distance Weighting (ADW) for

gridding the monthly anomalies compared to previous versions which used triangulated linear interpolation.

ADW allows a total control over how station observations are selected for gridding, and complete traceability for every datum in the output files.

The configuration for ADW in v4.00 is largely based on some aspects of that configuration are: Between 1 and 8 stations contributing to a grid cell at any time step

- ✓ The exponent in the distance weighting calculation is 4
- ✓ Observations take precedence over synthetic data (where both present)
- ✓ Synthetic data stations that lie within 45Â° of a 'real' data station are not used
- ✓ Grid cells with no stations in range are set to 0 anomaly (ie the climatology)

(Ian Harris, NCAS-Climate, 2017, Climatic Research Unit, School of Environmental Sciences, University of East Anglia)

III.4 Annual sum and inter-annual mean of precipitation

III.4.1 Calculations of annual sum and inter-annual mean in "R"

After importing all CSV files from the two datasets (GPCC and CRU), all sheets were organized per year from 1940 to 1999 for a total of 720 sheets.

The calculation of annual sum of each year was made by using "R Version 1.2.1335". As a programming software many methods are possible, one of them consist to create a repertoire for each year, to declare each file (month), then to sum whole data frames (cell per cell) for a total of 20 155 cells each data frame, with a simple addiction since our data frames have same dimension which are 145 rows and 139 columns, finally to import the data frame obtained as csv file.

Another way is to sum using the third dimension (required new commands and package) which is time from the first array created to extract data, and then to import the result. Since checking data is essential during each step, first method was used.

Script used for summation and mean (summation obtained for each dataset were divi ded cell per cell on 60 years (1940-1999)) (example used for GPCC) is represented in **(Figure III.13)**

Concerning comparison between inter-annual mean, see **(Figure III.14).**

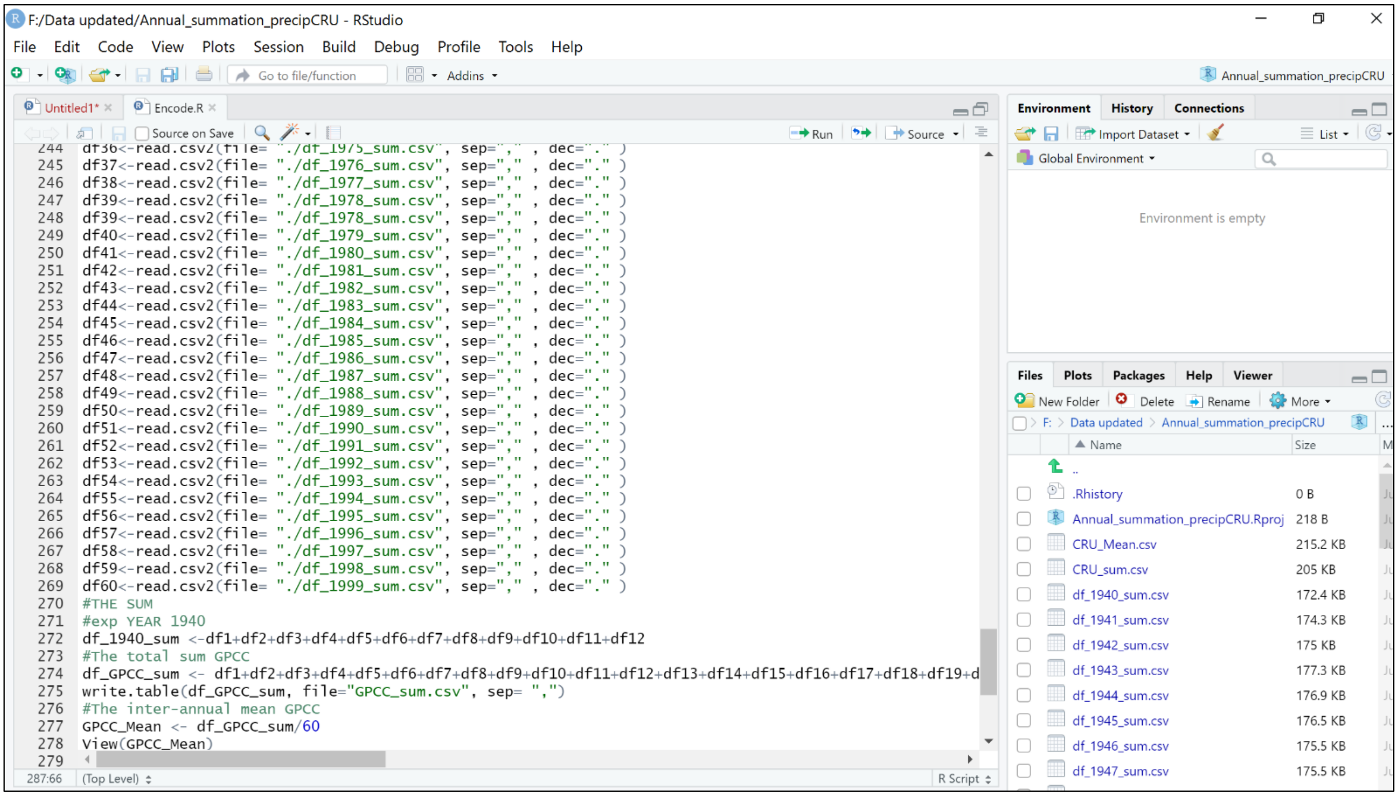


Figure III 12: *Screenshot of script used in R to sum GPCC files.Figure III.12: Screenshot of script used in R to sum GPCC files.*

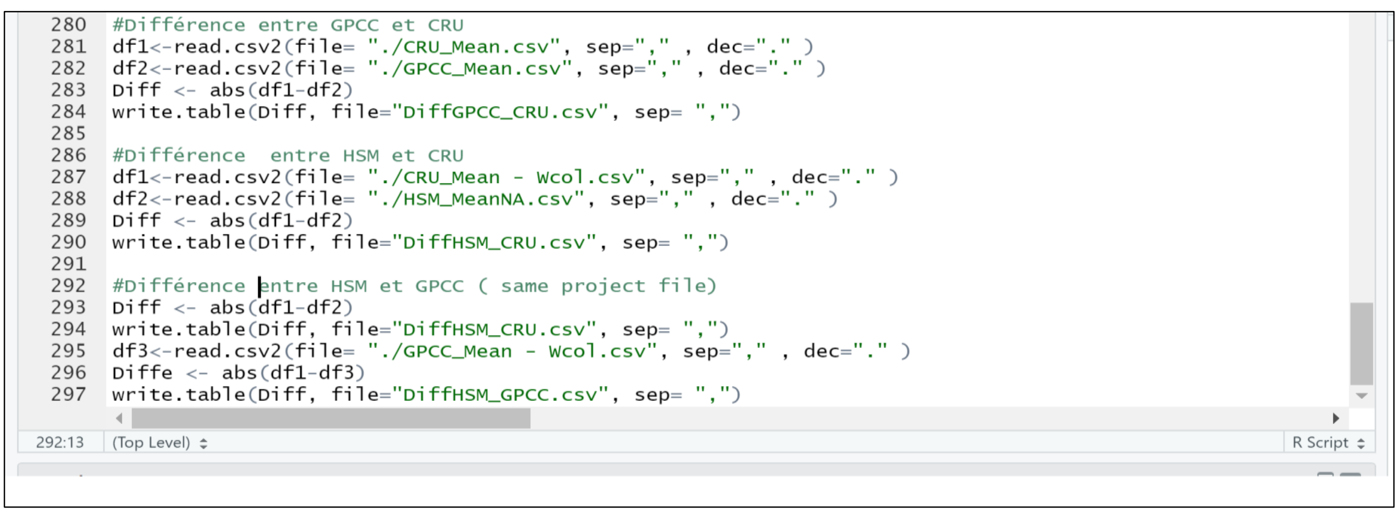


Figure III 13:*Screenshot of script used in R for inter-annual comparison between datasets.*

III.4.2 Display of inter-annual maps on ArcGIS

The display of the first maps on ArcGIS requires ASCII format, then, and by using Notepad ++ all Not available (NA) values have been changed to (-9999). Besides, separators in csv format have been removed, after that we added a header that defines the properties of the raster, such as CELL_SIZE, number of rows and columns, or coordinates, YLLCORNER, XLLCORNER and NODATA_VALUE.

Table III 2*:ASCII header information and example of header HSM.*

Parameter	Description	Required condition	HEADER (HSM)
NCOLS	Number of cell columns	Integer greater than 0	139
NROWS	Number of cell rows	Integer greater than 0	145
XLLCORNER	X-coordinate of the origin (lower left corner of the cell)	Match with x-coordinate type.	-18
YLLCORNER	Y-coordinate of the origin (lower left corner of the cell)	Match with y-coordinate type.	-35
CELLSIZE	Cell size	Greater than 0.	0.5
NODATA_VALUE	The input values to be NoData in the output raster	Optional. Default is -9999	-9999

III.4.3 Validation of spatialized datasets with observed data

Having the decimal coordinates of the stations to compare with, the cells that corresponded to each position were extracted in ArcGIS point by point for each dataset, after that the validation between observed data and specialize data extracted were performed using a linear correlation which is Pearson coefficient. Pearson's correlation coefficient is a statistical measure of the strength of a linear relationship between paired data. In a sample it is denoted by r and is by design constrained as follows: **$-1 \leq r \leq 1$**

Furthermore:

- Positive values denote positive linear correlation;
- Negative values denote negative linear correlation;
- A value of 0 denotes no linear correlation;
- The closer the value is to 1 or –1, the stronger the linear correlation.[13]

In the other hand, we can verbally describe the strength of the correlation using the guide that Evans (1996) suggests for the absolute value of r:

.00-.19 "very weak" / .20-.39 "weak" / .40-.59 "moderate" / .60-.79 "strong" / .80-1.0 "very strong"

Below script used for the correlation in R:

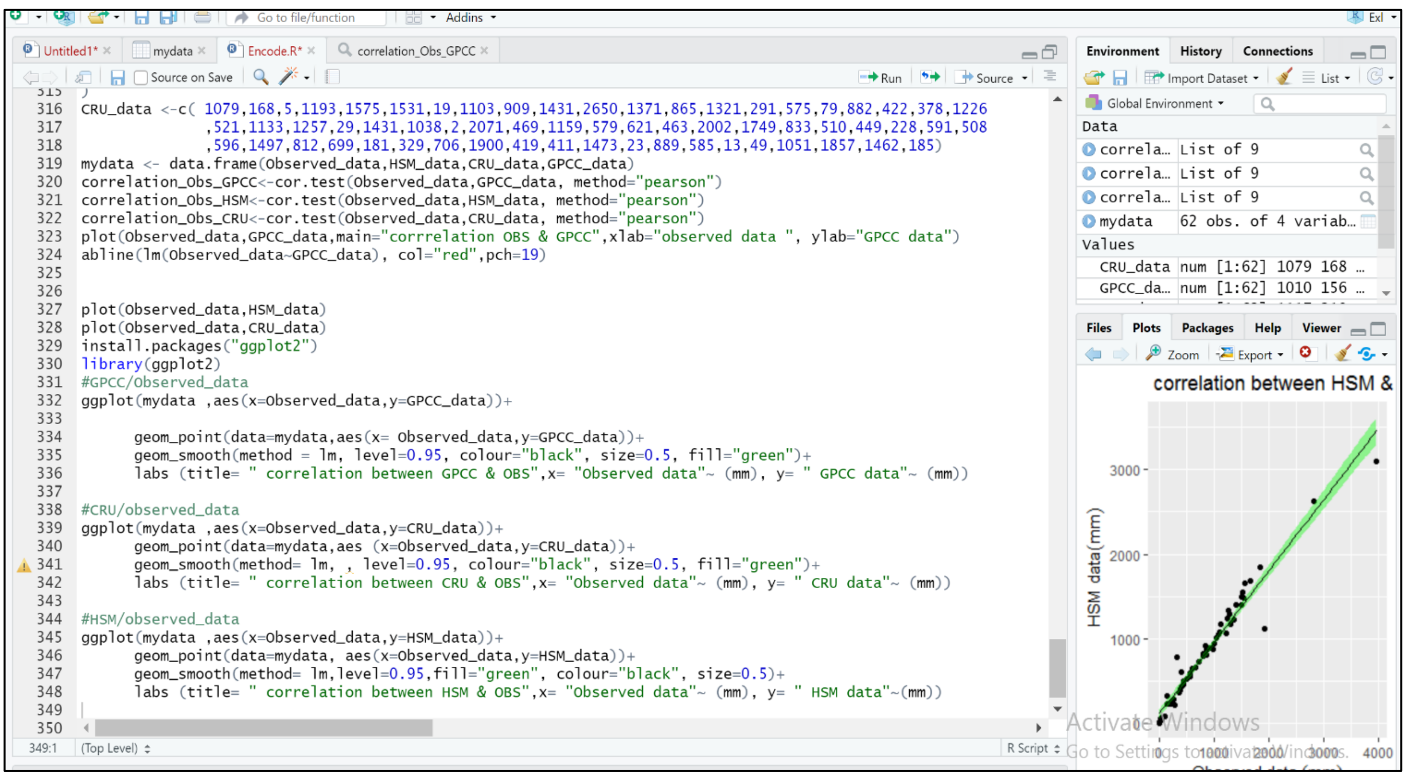


***Figure III 14**: Screenshot of script used in R to correlate observed data with datasets.*

[13] http://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf

RESULTS

RESULTS

IV.1 Inter-annual mean

Results of the various calculations performed on programming software “R” for the 60-year inter-annual period generated three different maps (HSM, GPCC and CRU), displayed using ArcGIS (10.3.0.43.22).

Since data on Madagascar and other islands in Africa are not produced by HSM, comparison with other datasets in these islands are not included.

The classification of precipitation adopted for the three inter-annual maps have homogeneous ranges so that we can make a difference during interpretation.

IV.1.1 Inter-annual mean HSM-SIEREM (Hydroscience-Montepellier)

Figure IV.1 below shows the inter-annual map obtained for HSM during period between (1940 and 1999). The lowest recorded precipitation value is "5.84 (mm) " for 60 years and it is situated in the Egyptian desert position (long = 30.71, lat = 24.91) and the highest value “4890 mm” is recorded for the same period in the Cameroonian coast position (long= 8.96, lat= 4.35).

As we can see in the figure VI.1 the lowest recorded precipitation during period between 1940 and 1999 is 5 (mm) to 100 (mm), they are located firstly in the Great African desert “Sahara”, it stretches from the Red Sea in the east and the Mediterranean in the north to the Atlantic Ocean in the west, Besides, the desert of Namibia which is located in the southern of Africa. In the other hand, the highest recorded precipitation “2900 (mm) to 4890 (mm)” are located especially in the western of Africa: coastal Liberia and Sierra, Cameroun, Gabon and Equatorial Guinea.

We have also zones named “The humid tropical zone” where the average of precipitation between is (1200 and 1700 mm), as we can see in the map, this area include countries like Congo, Central Africa, a part of Republic democratic of Congo, and Uganda.

In North Africa the average of precipitation is comprised between 350 (mm) – 500 (mm) and 500(mm) – 700 (mm) in some part like eastern of Algeria, North Morocco and Tunisia. This area is known as Mediterranean zone.

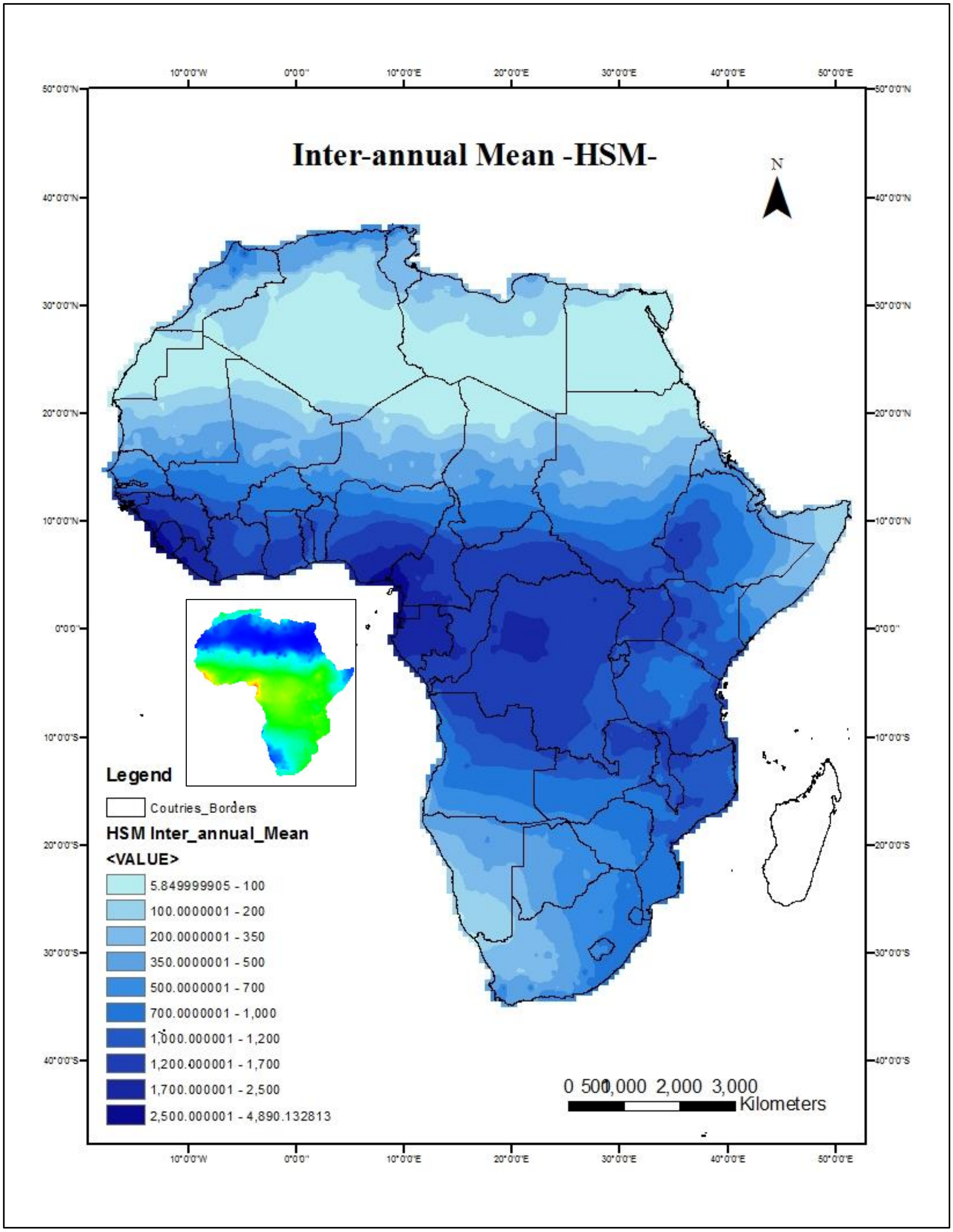


Figure IV 1:*Inter-annual mean for 60 years over Africa (HSM)*

IV.1.2 Inter-annual mean GPCC (Global Precipitation Climatology Centre)

Figure IV.2 represent the inter-annual mean obtained for dataset GPCC during same reference period (60 years). The lowest value of precipitation is 0 (mm) and more than 58 cells take this value, all located in the Egyptian desert. On the other hand, the highest value “3890 (mm)” is recorded on the Guinean coast position approximately (long= -13.356, lat = 9.614).

As distinguished in the previous map, the lowest precipitation values between (0 and 100 mm) are located in the Great African Desert and in the desert of Namibia where we can see that it is more pronounced then HSM. In the other hand, we can distinguish highest precipitation values “2500 (mm) to 3890 (mm)” recorded particularly in the eastern of Madagascar , and Western Africa including countries as Cameroun, Equatorial Guinea, Liberia and sierra Leone.

Also, Precipitation between “1700 (mm) and 2500 (mm)” seems to take a large part in Equatorial countries such Republic Democratic of Congo another time where it is more pronounced than HSM, a little bit in Ethiopia and in western of Madagascar.

With an inter-annual mean between 200 to 700 (mm) of precipitation, Sahelian zone stretches across the south-central latitudes of Northern Africa between the Atlantic Ocean and the Red Sea including countries like Niger, Mali, Chad , South Soudan and Mauritania. This zone is known having a very irregular precipitation that vary from season to another.

In the north of Africa the Mediterranean climate shows precipitation going from 200 to 1000 (mm) marked especially on the heights "mountains of Atlas” where Mountain peaks can reach 4000 meters altitude in Morocco.

In South Africa, precipitations are more accentuated in the East then in the west (we can see it in the map) and it gradually decreases with a range between 200 and 500 (mm).

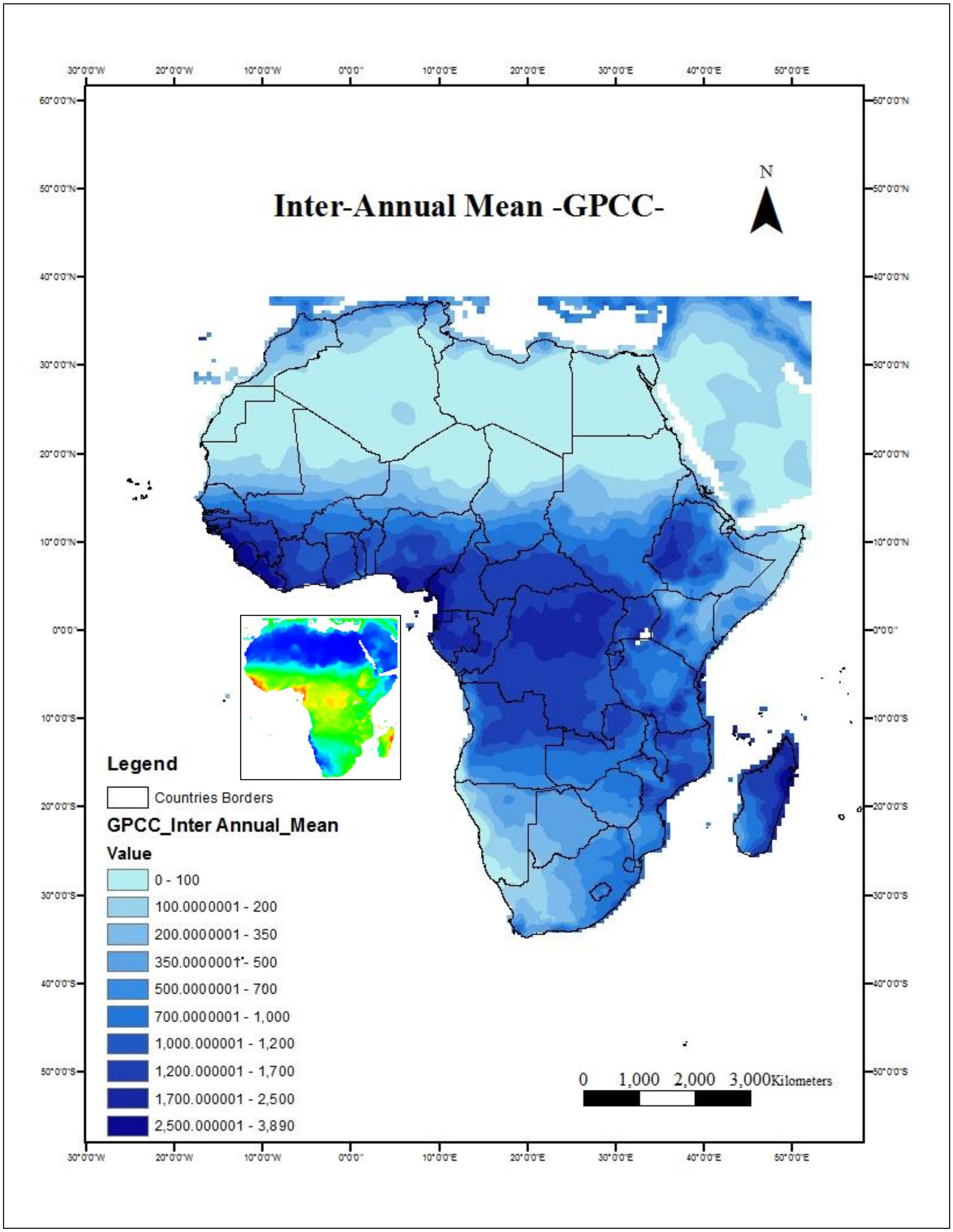


Figure IV 2 *: Inter-annual mean for 60 years over Africa (GPCC)*

IV.1.3 Inter-annual mean CRU (Climatic Research Unit)

Figure IV.3 shows the inter-annual map achieved using ArcGIS for the dataset CRU during the period (1940-1999), the lowest value of precipitation “ 0 (mm) “ is recorded in many places over Egyptian desert, North Soudan, Libya, North of Niger, South of Algeria, until north of Mali for a total of 293 cells of resolution 0.5 x 0.5 degree. On the other hand, the highest value of precipitation recorded is located in Sierra Leone coast position (long = -12.544 m lat= 8.015).

East Africa manifest tropical zones that do not reach more then 1700 (mm) of precipitation including countries like East Mozambique, Southern Tanzania, Malawi and North east Zimbawe. These countries have two distinct seasons (dry and wet season).

Most of the western Ethiopia also lies in the tropics, in a fact, we have high-altitude Mountains that reach 4000 meter altitude, in the North and western of the country.

With some differences, the same other areas of precipitation are marked as GPCC and HSM in general, but as we can see areas look very smoothed, which will tell us more about the interpolation method used, the difference will be more explicit in the following part.

IV.1.4 Summary on the three inter-annual maps

In Summary, the three datasets have different minimum and maximum inter-annual mean values, HSM with the highest value up to 4890 (mm). Maximums on the inter-annual (60 years) are all situated in West Africa, minimums on the same period seem to be all in the Egyptian desert except "CRU" which presents a larger area in the great desert stretched from Egypt to Mali (0 mm). To overview, precipitations are distributed on the continent according to the climatic zones, the three datasets distinguish these zones with some differences, and the comparison which will follow will give us more details on these differences.

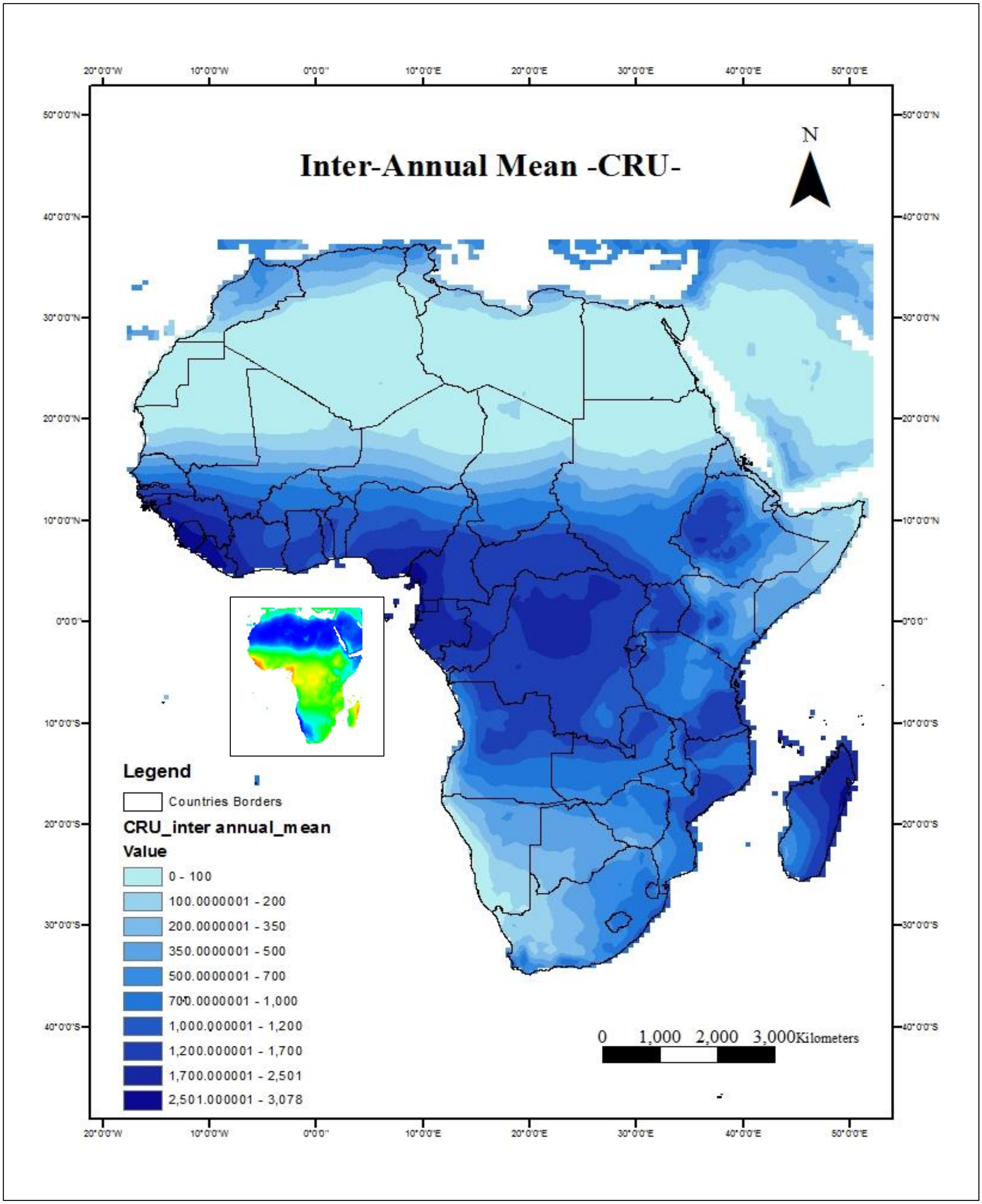


Figure IV 3:*Inter-annual mean for 60 years over Africa (CRU)*

IV.2 Comparison between inter-annual mean datasets

Three maps were shaped for comparison which are: (HSM_SIEREM-GPCC), (HSM_SIEREM-CRU) and (GPCC-CRU). The value ranges were created therefore, positive differences represent a dataset, and the negative differences represent the other dataset so that we can know who has more or less value of precipitation recorded or spatialized than the other one for each cell. Besides, knowing that the differences are not homogeneous indeed, the differences are accented on ranges and others not, so a geometric interval has been used, useful for visualizing data that is not distributed normally.

IV.2.1 Comparison between GPCC and CRU

Figure IV.5 shows the map obtained from cell-to-cell comparison made on arcGIS between the two datasets (GPCC and CRU), positive values represent the cells where GPCC has more values of precipitation than the CRU, as long as the negative values are those where CRU has precipitation values more that the GPCC.

The statistical analysis generated the histogram in Figure IV 4 which shows the percentage of cells that take diverse differences of precipitation.

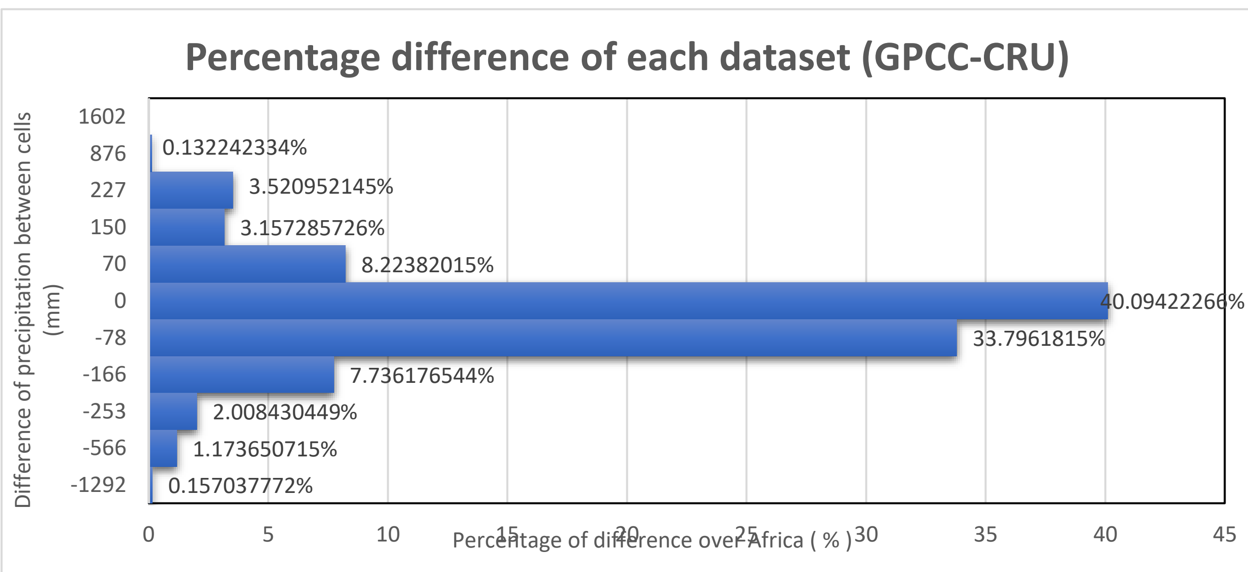


Figure IV 4: *Percentage of cells that take diverse differences of precipitation (GPCC-CRU)*

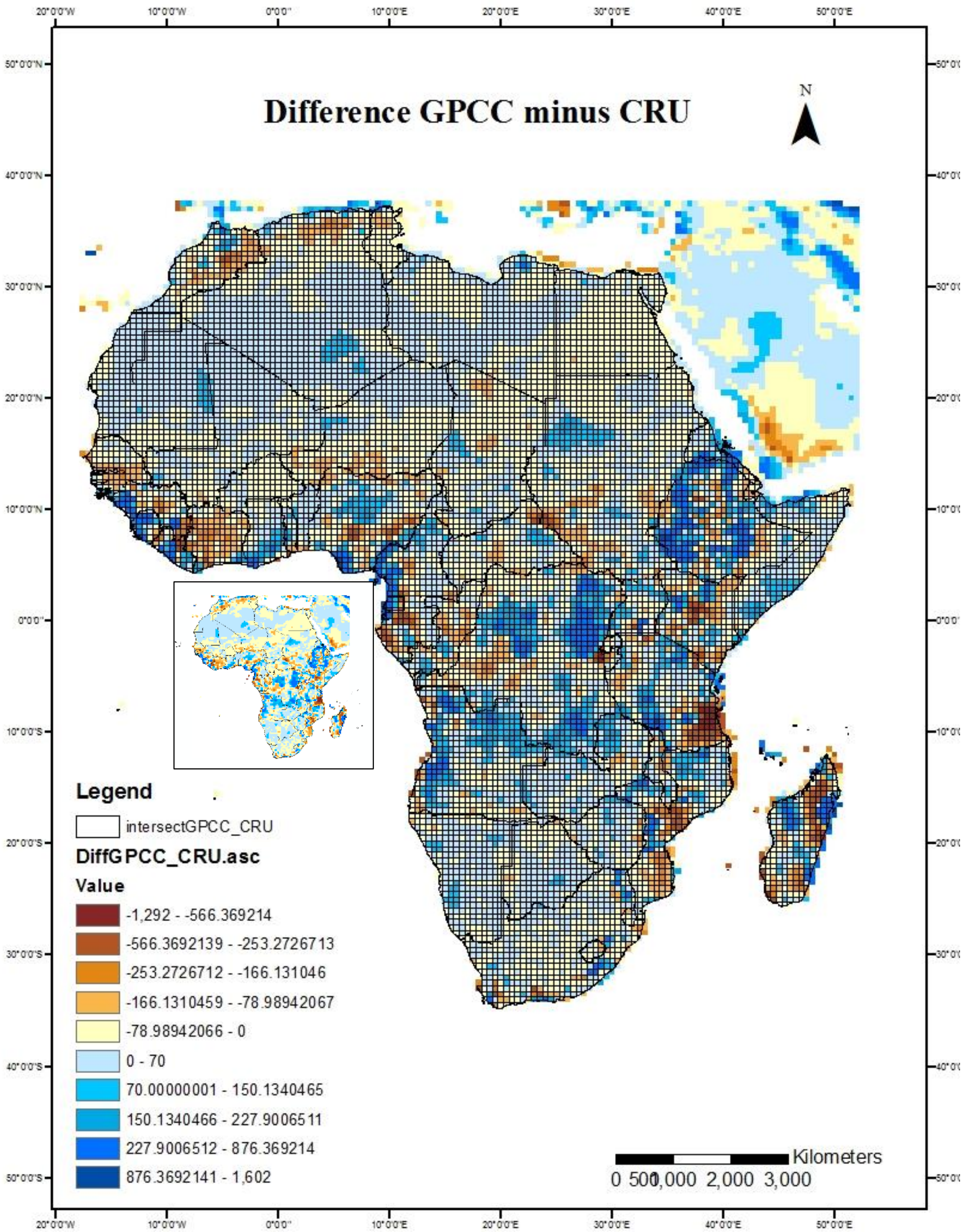


Figure IV 5: *Cartographic representation of difference in cells values of precipitation between GPCC and CRU for the reference period 1940-1999*

The results show that the large number of cells (+ 40% (GPCC and CRU)) are seen in the difference values between 0 and 70 mm of precipitation. On the map these differences (surface in light blue and light yellow) are generally found in the Great Desert, Somalia, desert of Namibia, and the Sahelian countries, in general, in areas that do not record a high values of precipitations.

The large difference in measured precipitation values (up to 1000 mm of difference) which take the smallest surface percentage (less than 1%) over Africa are located for GPCC in Ethiopian highlands, along the Congo River in East part situated in the Democratic Republic of the Congo, the coast of Cameroon, Nigeria and Equatorial Guinea, besides, the very east coast of Madagascar. About CRU, precipitation values that are greater than those of the GPCC are located all along the mountainous massif of Madagascar, The Atlas Mountains in North Africa, the largest game reserve in Africa covering surface of 50 000 km square (Selous Game Reserve, Tanzania) and also as we can see in the map at the very west border between Sudan and South Sudan at (Random National Park, where Rivers, streams, and permanent pools cover much of the park).

Other differences between GPCC and CRU are seen all over Africa where sometimes the variability of precipitation at the same time is huge (up to 1000 mm) in a small area or region (GPCC more than CRU or the opposite) and can be seen especially in countries with tropical and equatorial climate (wet countries).

IV.2.2 Comparison between HSM-SIEREM and CRU

Figure VI.7 shows the map obtained from cell-to-cell comparison made between the two datasets (HSM-SIEREM and CRU), positive values represent the cells where HSM has more values of precipitation recorded than the CRU, as long as the negative values are those where CRU has precipitation values more that HSM-SIEREM.

The statistical analysis generated the histogram in Figure IV.6 which shows the percentage of cells that take diverse differences of precipitation.

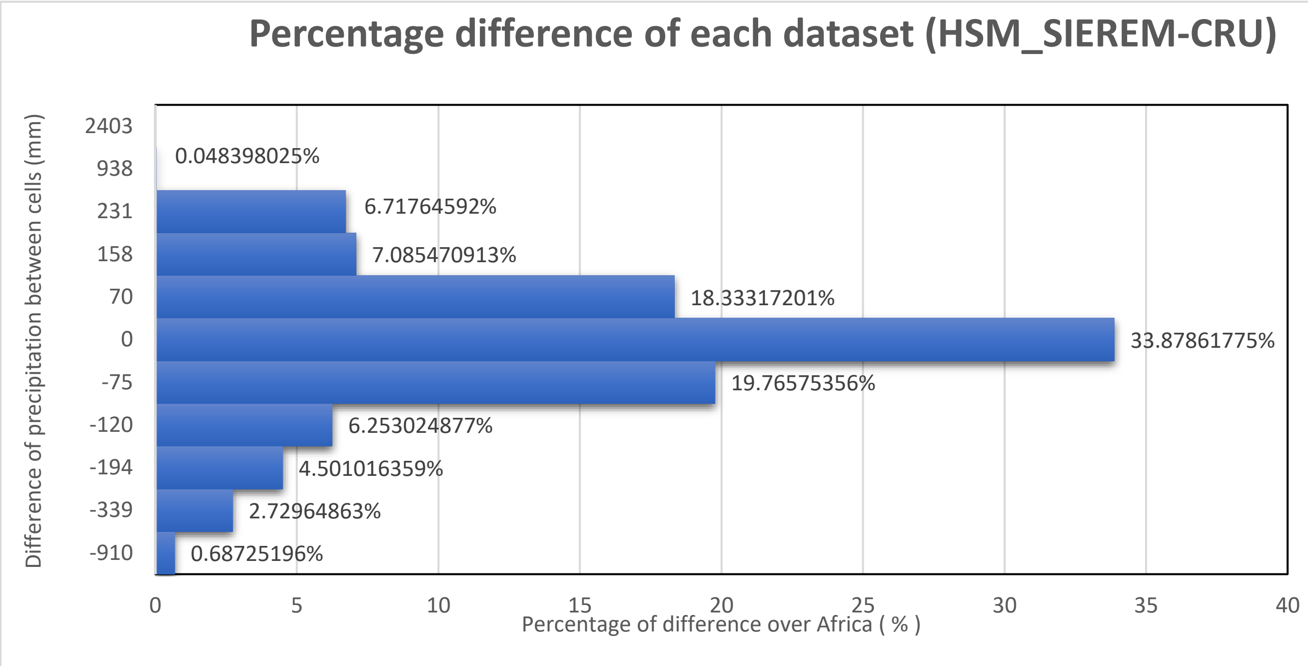


Figure IV 6*: Percentage of cells that take diverse differences of precipitation (HSM-CRU)*

During same period between 1940 –1999, and same as the first comparison, the large number of cells (+ 33 %) are seen in the difference values between 0 and 75 mm of precipitation for both HSM and CRU, concerning HSM, these differences are located in the great desert and Namibia where HSM has higher values of precipitation, same in the Sahelian countries where difference can reach 200 mm, we can see it in the map over West Tchad, Niger, East Mali and Mauritania.

Concerning CRU and for similar difference of precipitation (0 to 75 mm), they are distributed heterogeneously everywhere in Africa (particularly in Central countries of Africa).

Largest differences in precipitation from 300 (mm) to 900 (mm) which represent 3% of total cells (characterized in dark brown in the map for CRU) are recorded in many regions, along Atlas Mountains in north Africa, along coastal Eastern countries (Liberia, Sierra Leone, Guinea and Guinea Bissau), Congo Basin, Ethiopian Highlands, and Selous Game Reserve in south east of Tanzania. On the other hand, HSM has its highest difference of precipitation (6% of total cells represented in Dark blue) situated as well some parts of Sahelian countries, in Eritrea, West Ethiopia, Kenya, Tanzania, Zambia to Botswana in south of Africa and in the west Africa all over coast Cameroun, Nigeria and Gabon.

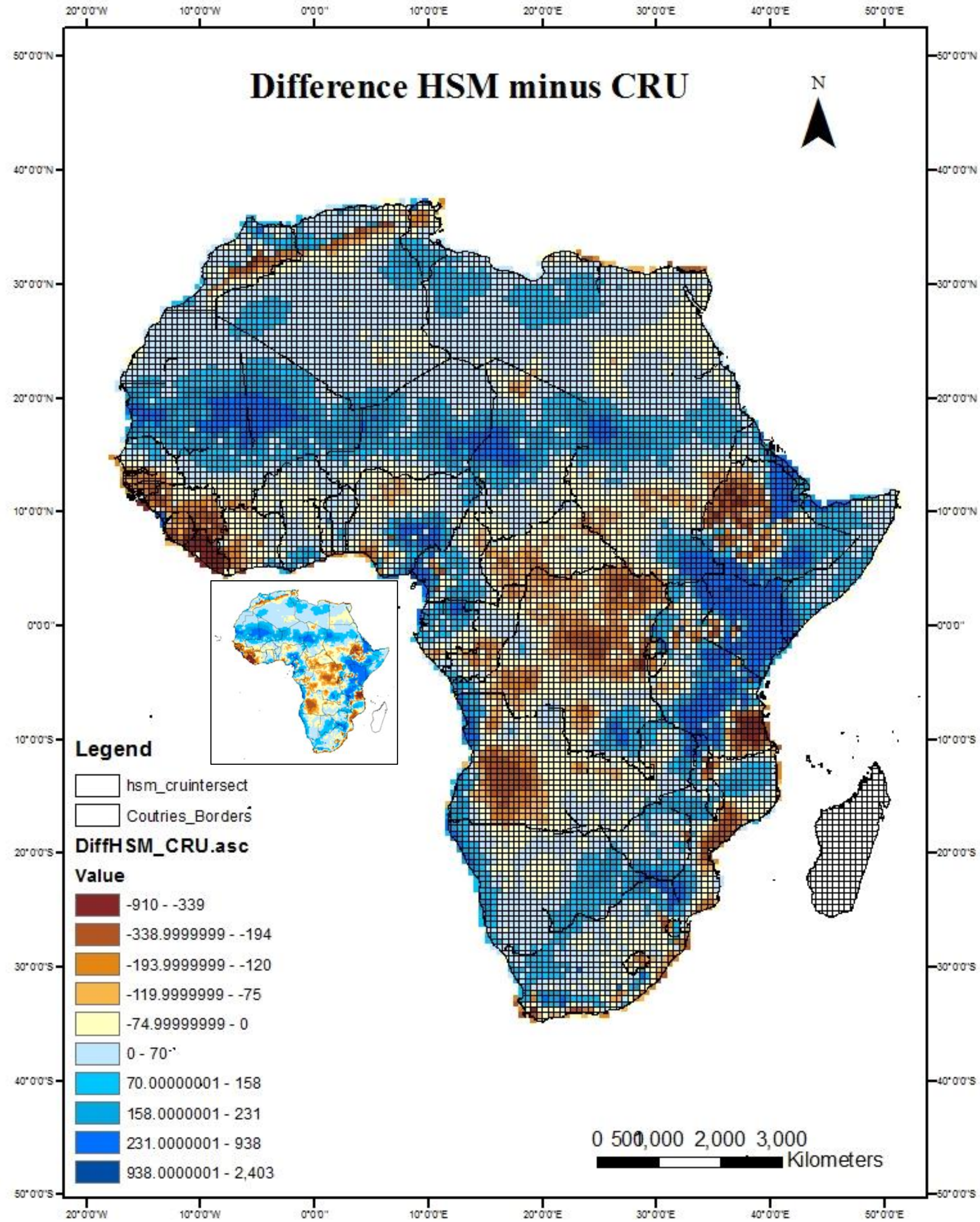


Figure IV 7*: Cartographic representation of difference in cells values of precipitation between HSM-SIEREM and CRU for the reference period 1940-1999*

IV.2.3 Comparison between HSM and GPCC

Figure VI.9 shows the map obtained from cell-to-cell comparison made between the two datasets (HSM-SIEREM and GPCC), positive values represent the cells where HSM has more values of precipitation recorded than GPCC, as long as the negative values are those where GPCC has precipitation values more that HSM-SIEREM.

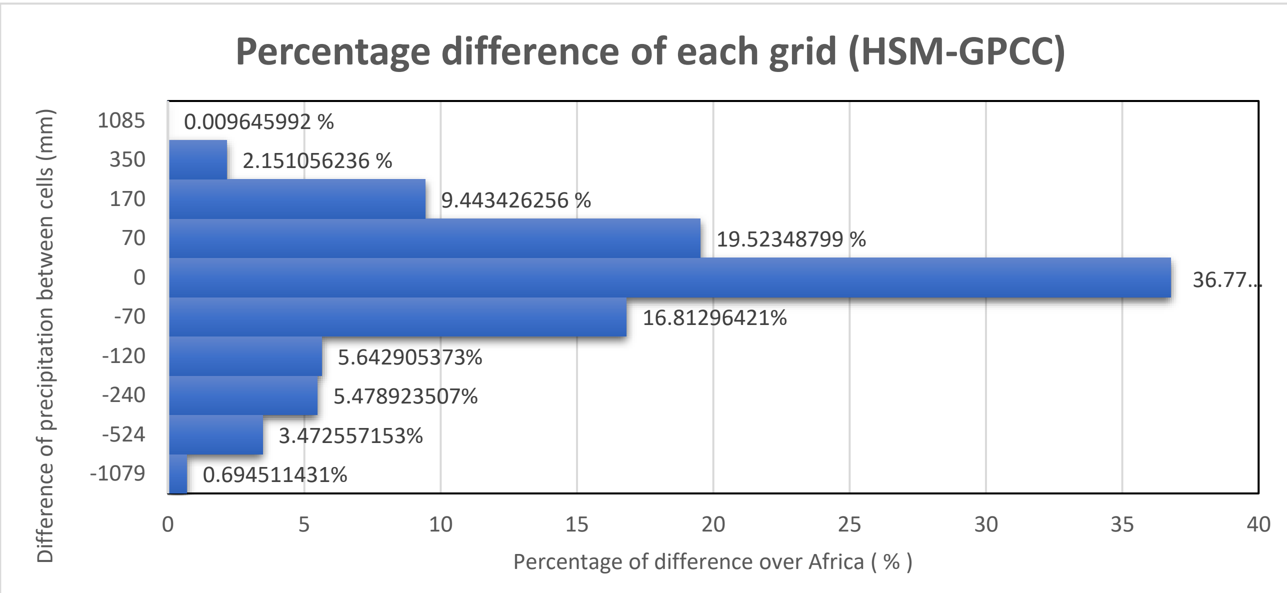


Figure IV 8: *Percentage of cells that take diverse differences of precipitation (HSM-GPCC).*

The largest number of cells have differences in values between 0 and 100 mm and still predominantly with a percentage reaching 36% of total cells. As noted on the map in figure IV.7 it appears the same difference by place between HSM-SIEREM_ GPCC and HSM-SIEREM _CRU where, HSM -SIEREM has higher precipitation values recorded in sub-Saharan countries, the Great Desert and Southern Africa countries, the opposite in central African countries (DRC, Angola) and west Africa (Liberia, Sierra Leone, guinea) where GPCC records larger values of precipitation. However differences up to 1000 mm represents the lowest number of cells with less than 1% of total cells.

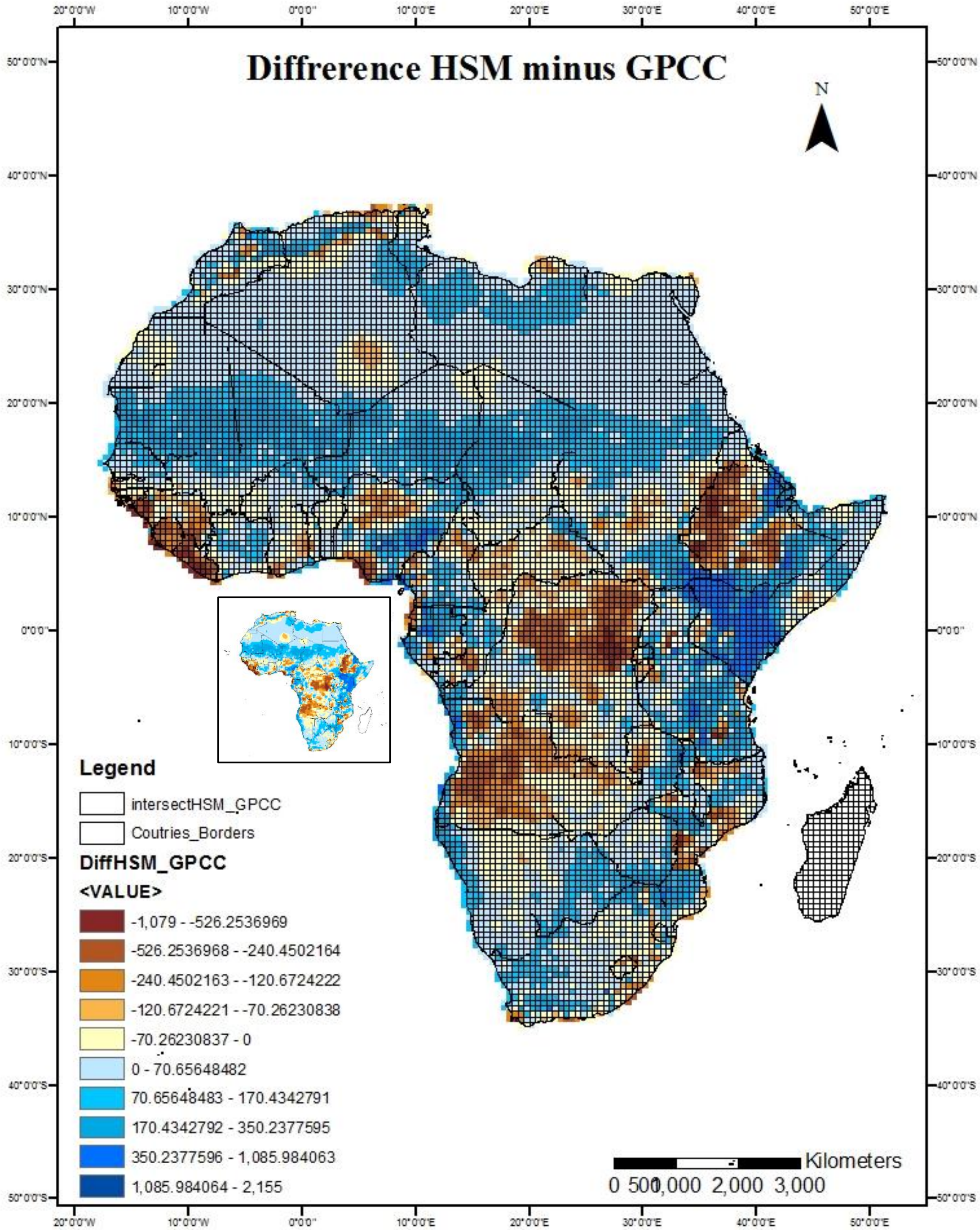


Figure IV 9: *Cartographic representation of difference in cells values of precipitation between HSM-SIEREM and GPCC for the reference period 1940-1999*

IV.2.4 Discussion on comparisons

Results showed differences in precipitation values between one dataset and another, these differences could be very small over a large area or very high difference but represent only individual cells. These small differences between dataset and another represent in most cases arid areas in Africa (deserts), firstly, relative to the climate there where rains do not fall much, secondly, the gauge stations are almost inexistent, or gauge stations do not give long time series during 60 years, Therefore, each dataset use a method to generate gridded data (Interpolation, acquisition of environmental satellite data, or Climate Normals). Hence the variance between the datasets.

We have also to remind that data provided by satellite in these areas where rainfall is very small give more errors as already mentioned in the literature review.

HSM-SIEREM seem in general to provide much more values of precipitation in Northern Africa and in Sahelian zones, in the other hand GPCC and CRU give more values of precipitation in Central Africa. In South Africa, the differences are rather neutral, we have to mention that South Africa seems to be the country with a very high density of station for several years. The differences are in the majority of the cases because of the number of stations which cover or not the surfaces, in a fact, some datasets which have a denser number of gauge stations in some cells do not have in other cells, and here the difference.

Differences have also been noticed in dense regions by vegetation cover or large wetlands, these areas sometimes inaccessible for several reasons including the factor of safety.

More technically, the method of interpolation used make also the difference, as we could see in highlands, where the most differences appear, and some methods do not take into consideration the altitude while it has an effect on the variability of precipitation.

In the end we will try to proceed to a validation of some cells thanks to data on the ground, the number of observed data is very limited at the continental level, but time series are long and representative for 60 years.

IV.3 Validation of datasets

As mentioned in the methodology, Pearson's linear correlations were carried out between observed data and specialized data (HSM-CRU and GPCC), a very important step to conclude the study.

The correlation between observed data and GPCC spatialized data in "R" gave the following graph:

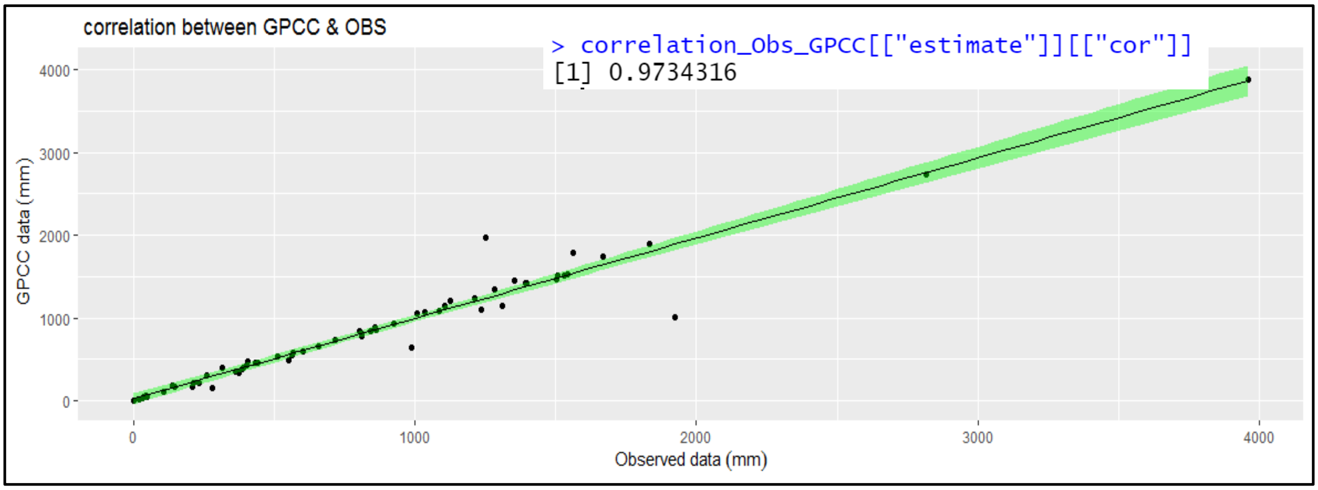


Figure IV 10: *Correlation between GPCC and observed data.*

The result shown a coefficient of correlation **r = + 0.9734316** which is positive correlation between -1 and + 1 and very strong coefficient closed to 1 accordingly to Evans (1996). Therefore, we have positive linear bond between Observed data and GPCC data that evolve in the same direction.

Added to that the 95% confidence interval (green area in the graph) which represents the range of values within which we are 95% certain to find closed values to the regression line, we can see that some values are outside confidence interval.

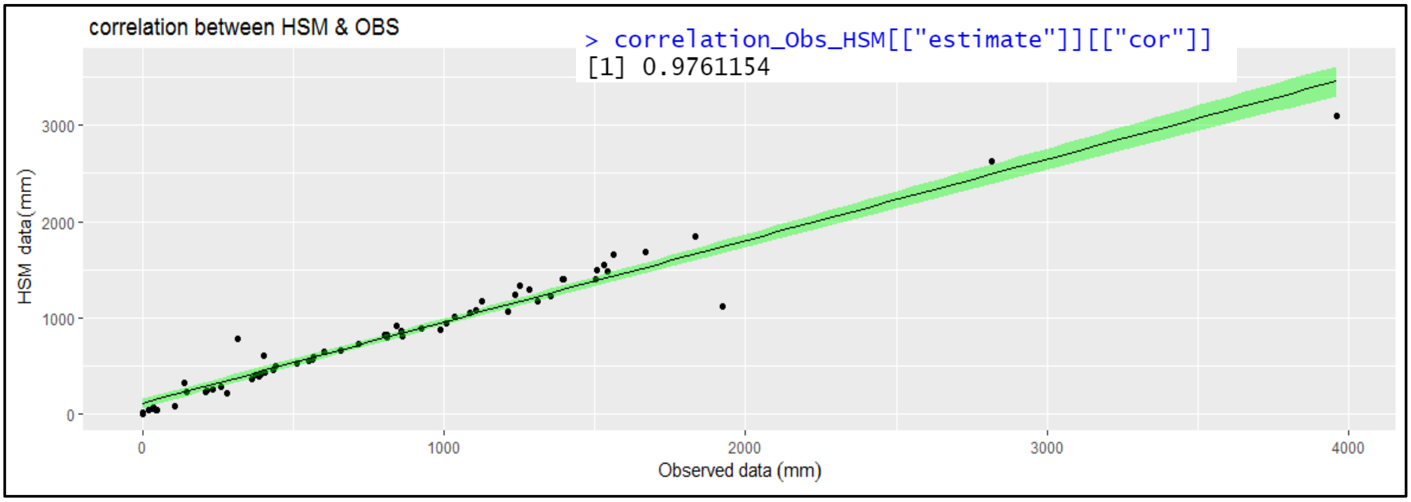


***Figure IV 11:** Correlation between HSM and observed data.*

A bit more than coefficient of correlation of GPCC and Observed data, Results from graph above show a positive coefficient of correlation **r = + 0.9761154** between -1 and + 1 and very strong coefficient closed to 1 accordingly also to Evans (1996).Thus, we have positive linear bond between Observed data and HSM data that evolve in the same direction.

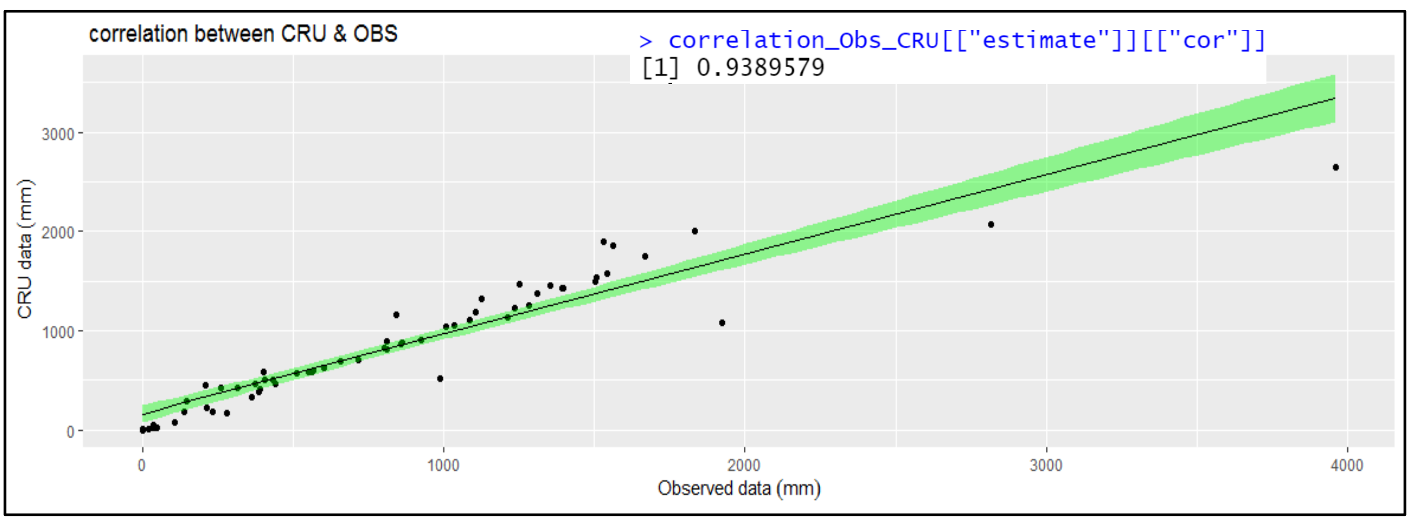


***Figure IV 12:** Correlation between CRU and observed data.*

With a correlation coefficient **r= +0.9389579**, correlation between CRU and observed data seems to be the weakest compared to GPCC and HSM correlations with observed data.

Nevertheless, it stills a strong positive correlation which means there is linear bond between Observed data and CRU data that evolve in the same direction.

Among all these results there is one very important thing in common which is the very strong positive correlation with the data observed for the three datasets (GPCC, HSM and CRU), even if the number of stations is not high at continental level to compare with more cells. The random choice of the stations shows nevertheless that the specialized data can replace or add to the data on the ground when it is necessary.

However, there are differences between the datasets, where some grids do not give much precision, they can be seen in the main highlands in Africa, example of the gauge which is in Addis Ababa (more than 2000 m of altitude) where the three datasets do not show a good correlation with the observed value. Moreover, in arid regions where gauge stations record very less values of precipitation we can find some cells that are overestimated.

.

CONCLUSION AND RECOMMANDATIONS

CONCLUSION AND RECOMMANDATIONS

The comparison of the historical precipitation gridded data of several datasets (HSM-SIEREM, GPCC, CRU) for a period from 1940 to 1999 and over the African continent was the result of several months of work which required a lot of time for the treating of all these data, including the use of different software, for the process and mapping, it demonstrates the real need nowadays for data scientist and analyst skills to conduct researches, particularly on climate change.

Through this project, several conclusions have been issued:

- ✓ The important role of historical gridded data to replace missing observed data in order to establish researches that requires data for very long periods, indeed, strong correlations have been generated, but still depend on several factors (study area, resolution, period of study, method of interpolation).
- ✓ Gauge stations on the ground remain the most credible than satellite data or radars, to compare with, their density in a region increase the chances of bringing spatialized data closer to them and the opposite is true example of deserts (great desert and desert of Namibia) where the majority of cells have given differences between the three datasets, where observation stations are in majority of case non-existent.
- ✓ The interpolation method used is a major factor in generating the best spatialized data, they are now subject to research and comparisons for their improvements by adopting algorithmic models, some methods do not consider Altitude, but remain very effective, other methods that take into consideration altitudes for comparisons on the variable "precipitation" can give negative values which is not very logical compared to examples for a study on the temperature as a variable, and this is how and this is how big differences appeared on the major parts of highlands in Africa.
- ✓ The acquisition of gauge station data by each dataset is not only a question of availability of gauge station in an area or another, it can sometimes be influenced by policy reasons, partnerships, historical reasons as for example HSM-SIEREM which holds the largest historical data of the formerly colonized French-speaking countries in Africa, where the difference.

During this project, many constraints were encountered because of the time, tools to use, lack of observed data to validate more cells but also the study area (at the continental level), therefore it will require more precision (larger resolution) for research that has a specific goal such as hydrological modeling, or climate change impact study on floods in a region for example.

At our objective of study we could generate climate zones with temperature and pressure gridded data, and push the comparison towards climate models. We could also treat or make more statistical analysis on each dataset than to compare. We are recommending also a comparison towards interpolation methods used only in highlands since differences there, are highly perceived. And making research partnerships between institutions to share observed data.

In the end this project is the beginning of many and several other project, since historical study period is very important for climate change research.

REFERENCES

Abtew, W., & Melesse, A. M. (2014). The Nile River Basin. *Nile River Basin: Ecohydrological Challenges, Climate Change and Hydropolitics*, *9783319027*, 7–21. https://doi.org/10.1007/978-3-319-02720-3_2

Adeyewa, Z. D., & Nakamura, K. (2003). Validation of TRMM Radar Rainfall Data over Major Climatic Regions in Africa. *Journal of Applied Meteorology*, *42*(2), 331–347. https://doi.org/10.1175/1520-0450(2003)042<0331:votrrd>2.0.co;2

Aghakouchak, A., Mehran, A., Norouzi, H., & Behrangi, A. (2012). Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, *39*(9), 3–6. https://doi.org/10.1029/2012GL051592

Ahmed, K., Shahid, S., Wang, X., Nawaz, N., & Najeebullah, K. (2019). Evaluation of gridded precipitation datasets over arid regions of Pakistan. *Water (Switzerland)*, *11*(2). https://doi.org/10.3390/w11020210

Apaydin, H., Kemal Sonmez, F., & Yildirim, Y. E. (2004). Spatial interpolation techniques for climate data in the GAP region in Turkey. *Climate Research*, *28*(1), 31–40. https://doi.org/10.3354/cr028031

Black, E., Allan, R. P., Tarnavsky, E., Nkonde, E., Alcántara, E. M. U., Senkunda, S., … Greatrex, H. (2017). A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Scientific Data*, *4*, 170063. https://doi.org/10.1038/sdata.2017.63

Black, E., Stringer, M., Allan, R. P., Tarnavsky, E., Chadwick, R., Kayitakire, F., … Grimes, D. (2014). Extension of the TAMSAT Satellite-Based Rainfall Monitoring over Africa and from 1983 to Present. *Journal of Applied Meteorology and Climatology*, *53*(12), 2805–2822. https://doi.org/10.1175/jamc-d-14-0016.1

Boyer, J. F., Dieulin, C., Rouche, N., Cres, A., Servat, E., Paturel, J. E., & Mahé, G. (2006). SIEREM: an environmental information system for water resources, *308*(November), 19–25.

C, V. D., Shetty, A., & Kiros, G. (2016). LongTermRainfallTrendUsingHighResolutionGridded, *3*(1).

Cattani, E., Merino, A., & Levizzani, V. (2016). Evaluation of Monthly Satellite-Derived

Precipitation Products over East Africa. *Journal of Hydrometeorology*, *17*(10), 2555–2573. https://doi.org/10.1175/jhm-d-15-0042.1

Dieulin, C., Mahé, G., Paturel, J.-E., Ejjiyar, S., Tramblay, Y., Rouché, N., & EL Mansouri, B. (2019). A New 60-year 1940/1999 Monthly-Gridded Rainfall Data Set for Africa. *Water*, *11*(2), 387. https://doi.org/10.3390/w11020387

Eric S, S., Masocha, M., & Dube, T. (2019). Evaluation of TAMSAT satellite rainfall estimates for southern Africa: An inter-product comparison study. *Physics and Chemistry of the Earth, Parts A/B/C*. https://doi.org/10.1016/J.PCE.2019.02.008

Harris, R., & Baumann, I. (2015). Open data policies and satellite Earth observation. *Space Policy*, *32*(January), 44–53. https://doi.org/10.1016/j.spacepol.2015.01.001

Ikechukwu, M. N., Ebinne, E., Idorenyin, U., & Raphael, N. I. (2017). Accuracy Assessment and Comparative Analysis of IDW, Spline and Kriging in Spatial Interpolation of Landform (Topography): An Experimental Study. *Journal of Geographic Information System*, *09*(03), 354–371. https://doi.org/10.4236/jgis.2017.93022

Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, *15*(4), 1109–1116. https://doi.org/10.5194/hess-15-1109-2011

Lam, N. S.-N. (2008). Spatial Interpolation Methods: A Review. *The American Cartographer*, *10*(2), 129–150. https://doi.org/10.1559/152304083783914958

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling and Software*, *53*, 173–189. https://doi.org/10.1016/j.envsoft.2013.12.008

Maggioni, V., Sapiano, M. R. P., & Adler, R. F. (2016). Estimating Uncertainties in High-Resolution Satellite Precipitation Products: Systematic or Random Error? *Journal of Hydrometeorology*, *17*(4), 1119–1129. https://doi.org/10.1175/jhm-d-15-0094.1

Nicholson, S. E., Dezfuli, A. K., & Klotter, D. (2012). A two-century precipitation dataset for the continent of Africa. *Bulletin of the American Meteorological Society*, *93*(8), 1219–1231. https://doi.org/10.1175/BAMS-D-11-00212.1

Nicholson, S. E., Funk, C., & Fink, A. H. (2018). Rainfall over the African continent from the 19th through the 21st century. *Global and Planetary Change*, *165*(May 2017), 114–127. https://doi.org/10.1016/j.gloplacha.2017.12.014

Paturel, J., Boubacar, I., L'Aour, A., & Mahe, G. (2010). Note de recherche : Grilles mensuelles de pluie en Afrique de l ' ouest et centrAle. *Revue Des Sciences de l'eau*, *23*(4), 325–333.

Rishikeshan, C. A., Katiyar, S. K., & Vishnu Mahesh, V. N. (2014). Detailed evaluation of dem interpolation methods in GIS using DGPS data. *Proceedings - 2014 6th International Conference on Computational Intelligence and Communication Networks, CICN 2014*, 666–671. https://doi.org/10.1109/CICN.2014.148

Rouché, N., Mahé, G., Ardoin-bardin, S., Dieulin, C., Bardin, G., & Paturel, J. (2010). pour l ' Afrique (pe. *Office*, *21*(4), 336–338.

Schneider, U., Becker, A., Ziese, M., & Rudolf, B. (2014). Global Precipitation Analysis Products of the GPCC. *Internet Publication*, (May), 1–13. Retrieved from ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf

Theon, J. S., Aeronautics, N., & Administration, S. (1994). •• w, *14*(3), 159–165.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., & Kitoh, A. (2012). Aphrodite constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, *93*(9), 1401–1415. https://doi.org/10.1175/BAMS-D-11-00122.1

APPENDIX

Appendix 1:

GPCC Full Data Reanalysis Version 2018 0.5x0.5 Monthly Total

```
1 variables (excluding dimension variables):

        float precip[lon,lat,time]
            missing_value: -9.96920996838687e+36
            var_desc: Precipitation
            level_desc: Surface
            statistic: Nobs
            parent_stat: Observations
            long_name: GPCC Monthly total of precipitation: nobs
            valid_range: 0
             valid_range: 8000
            units: observations
            level: Surface
            actual_range: 0
             actual_range: 112
            dataset: GPCC Precipitation 0.5degree V2018 Full Reanalysis

     4 dimensions:
        lat  Size:360
            long_name: Latitude
            units: degrees_north
            standard_name: latitude
            axis: Y
            coordinate_defines: point
            actual_range: 89.75
             actual_range: -89.75
        lon  Size:720
            long_name: Longitude
            units: degrees_east
            standard_name: longitude
            actual_range: 0.25
             actual_range: 359.75
            axis: X
            coordinate_defines: point
        nbnds  Size:2
        time  Size:1512   *** is unlimited ***
            long_name: Time
            units: days since 1800-1-1 00:00:00
            delta_t: 0000-01-00 00:00:00
            avg_period: 0000-01-00 00:00:00
            standard_name: time
            axis: T
            coordinate_defines: start
            actual_range: 33237
             actual_range: 79227

    10 global attributes:
        Original_Source: http://www.dwd.de/en/FundE/Klima/KLIS/int/GPCC/GPCC.htm
 is the webpage and the data is at ftp://ftp.dwd.de/pub/data/gpcc/download.html
        Reference: Users of the data sets are kindly requested to give feed back and to refer to GPCC publications on this webpage: http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop/?_nfpb=true&_pageLabel=_dwdwww_klima_umwelt_datenzentren_wzn&T12404518261141645246564gsbDocumentPath=Content%2FOeffentlichkeit%2FKU%2FKU4%2FKU42%2Fteaser__product__access.html&_state=maximized&_windowLabel=T12404518261141645246564&lastPageLabel=_dwdwww_klima_umwelt_datenzentren_wzn
```

```
                original_source: ftp://ftp-anon.dwd.de/pub/data/gpcc/html/fulld
ata_download.html
                Conventions: CF 1.0
                dataset_title: Global Precipitation Climatology Centre (GPCC)
                References: https://www.esrl.noaa.gov/psd/data/gridded/data.gpc
c.html
                title: GPCC Full Data Reanalysis Version 2018 0.5x0.5 Monthly T
otal
                history: Created 09/2018  based on V2018 data obtained via ftp
                data_modified: 2019-03-12
                _NCProperties: version=2, netcdf=4.6.3,hdf5=1.10.5
```

Appendix 2:

File cru_ts4.01.1901.2016.pre.dat.nc (NC_FORMAT_CLASSIC):

```
 2 variables (excluding dimension variables):
    float pre[lon,lat,time]
        long_name: precipitation
        units: mm/month
        correlation_decay_distance: 450
        _FillValue: 9.96920996838687e+36
        missing_value: 9.96920996838687e+36
    int stn[lon,lat,time]
        description: number of stations contributing to each datum
 3 dimensions:
    lon  Size:720
        long_name: longitude
        units: degrees_east
    lat  Size:360
        long_name: latitude
        units: degrees_north
    time  Size:1392   *** is unlimited ***
        long_name: time
        units: days since 1900-1-1
        calendar: gregorian
8 global attributes:
    Conventions: CF-1.4
    title: CRU TS4.01 Precipitation
    institution: Data held at British Atmospheric Data Centre, RAL,
UK.
    source: Run ID = 1709081022. Data generated from:pre.1704241136
.dtb
    history: Fri  8 Sep 2017 12:54:11 BST : User ianharris : Progra
m   makegridsauto.for called by update.for
    references: Information on the data is available at http://badc
.nerc.ac.uk/data/cru/
    comment: Access to these data is available to any registered CE
DA user.
    contact: BADC badc@rl.ac.uk
```

Appendix 3:

File precip.mon.total.v301.nc (NC_FORMAT_NETCDF4_CLASSIC):

```
           1 variables (excluding dimension variables):
              float precip[lon,lat,time]
                  missing_value: -9.96920996838687e+36
                  units: cm
                  var_desc: Precipitation
                  level_desc: Surface
                  statistic: Total
                  parent_stat: Other
                  long_name: Monthly total of precipitation
                  cell_methods: time: sum
                  avg_period: 0000-01-00 00:00:00
                  dataset: Univ. of Delaware Precipitation and Air Temp. v3.0
1
                  actual_range: 0
                   actual_range: 765.690002441406
                  valid_range: 0
                   valid_range: 1200

           3 dimensions:
              lat  Size:360
                  actual_range: 89.75
                   actual_range: -89.75
                  long_name: Latitude
                  units: degrees_north
                  axis: Y
                  standard_name: latitude
                  coordinate_defines: center
              lon  Size:720
                  long_name: Longitude
                  units: degrees_east
                  axis: X
                  standard_name: longitude
                  actual_range: 0.25
                   actual_range: 359.75
                  coordinate_defines: center
              time  Size:1332   *** is unlimited ***
                  long_name: Time
                  units: hours since 1900-1-1 0:0:0
                  delta_t: 0000-01-00 00:00:00
                  avg_period: 0000-01-00 00:00:00
                  axis: T
                  standard_name: time
                  coordinate_defines: start
                  actual_range: 0
                   actual_range: 972264

          8 global attributes:
              history: created 8/2012 by CAS NOAA/ESRL PSD
              title: Univ. Delaware Monthly Precipitation, 1900-2010 V3.01
              Conventions: CF-1.0
              Source: http://climate.geog.udel.edu/~climate/html_pages/Global
2_Ts_2009/README.global_p_ts_2009.html
              version: 3.01
              dataset_title: Terrestrial Air Temperature and Precipitation: 1
900-2014 Gridded Monthly Time Series
              source: http://climate.geog.udel.edu/~climate/
              References: https://www.esrl.noaa.gov/psd/data/gridded/data.UDe
l_AirT_Precip.html
```

Appendix 4:

Meta data of shapefiles HSM-SIEREM

FID: id of every cell, automatically managed by ArcGis
Shape: polygon
AREA: 0.25 decimal degree, half a square degree
PERIMETER: 2 degrees
GRID_CODE: rainfall value in millimetres of the cell, calculated with IDW **method**
CELLID: unique value identifying the cell it is made of the longitude /latitude value of the centroid of the cell in the form of 5 digits for the longitude (3 digits for the degrees and 2 digits for the minutes in decimal base, with no separator) then the letter E o W ((East or West) ant 4digits for the latitude (2 digits for the degrees and 2 digits for the minutes in decimal base) then a letter N or S (North or South)
XG: longitude value in decimal degrees of the centroid of the cell
YG latitude value in decimal degrees of the centroid of the cell

All the layers are in geographical coordinate system,
datum is D_WGS_1984,
central meridian is Greenwich,
angular unit is decimal degrees

Appendix 5: "inventory of observed data"

Number of stations: 63 stations (synoptic and rainfall stations)

Variable: Precipitation

Time step: monthly

Source: Hydroscience Montpellier

Série_identifiant	Station_identifiant	Nom_station	Date_de_début	Date_de_fin	Type_de_station	Longitude	latitude
ETPR019	6345000	ADDIS ABABA	01/01/1898	1/5/2000	Station synoptique	38.75	9.03
NEPR003	1320000400	AGADEZ	1/1/1921	1/12/2000	Station synoptique	7.9889	16.975
EGPR021	6241400	ASSWAN	1/1/1935	1/5/2000	Station synoptique	32.8	24
BFPR008	1200001900	BANFORA AGRICULTURE	1/1/1922	1/12/2001	Station pluviometrique	-4.7667	10.6167
CFPR100	1060150000	BANGUI ORSTOM	1/4/1907	1/12/2007	Station climatologique	18.55	4.4337
CFPR019	1060005500	BERBERATI	1/1/1938	1/12/2007	Station synoptique	15.8	4.25
MRPR009	1300001500	BIR MOGHREIN	1/1/1942	1/10/2003	Station synoptique	-11.6167	25.2333
BFPR019	1200004000	BOBO - DIOULASSO	1/1/1907	1/12/2001	Station synoptique	-4.3167	11.1667
BFPR023	1200005200	BOROMO	1/1/1922	1/12/2001	Station synoptique	-2.9333	11.75
CGPR000	1070000100	BRAZZAVILLE (MAYA-MAYA)	1/1/1932	1/5/2000	Station synoptique	15.2544	-4.2572
GNPR000	1170000100	CONAKRY-PORT - AERO dep.1950	1/1/1905	1/12/2000	Station synoptique	-13.6167	9.5667
BJPR000	1110000100	COTONOU AERO	1/1/1910	1/5/2000	Station synoptique	2.3833	6.35
BFPR034	1200007900	DEDOUGOU	1/1/1922	1/12/2001	Station synoptique	-3.4667	12.4667
CIPR052	1090009100	DIMBOKRO	1/10/1921	1/12/2000	Station synoptique	-4.7	6.65

00	DJPR0	6312500	DJIBOUTI AIRPORT	1/1/1901	1/9/1999	Station synoptique	43.164503	11.544824
045	BFPR	1200010000	DORI	1/6/1920	1/12/2001	Station synoptique	-0.0333	14.0333
012	EGPR	6237000	EL ARISH	1/3/1907	1/5/2000	Station pluviometrique	33.82	31.08
025	SUPR	6277000	EL GENEINA	1/5/1928	1/10/1999	Station synoptique	22.4	13.4
046	BFPR	1200010300	FADA N GOURMA	1/8/1919	1/12/2001	Station synoptique	0.3667	12.0333
132	KEPR	6372300	GARISSA	1/1/1931	1/4/2000	Station synoptique	39.5	-0.5
089	MLPR	1270016900	HOMBORI	1/1/1920	1/5/2000	Station synoptique	-1.7	15.2833
043	NGPR	1330040000	ILORIN	1/1/1906	1/5/2001	Station climatologique	4.5833	8.4833
007	MAPR	1280018600	Ifrane	1/10/1934	1/12/2008	Station pluviometrique	-5.7	33.32
048	NGPR	1330044500	KADUNA	1/1/1913	1/4/2001	Station synoptique	7.45	10.6
048	TOPR	1470013300	KARA VILLE (LAMA-KARA VILLE)	1/4/1931	1/4/2000	Station pluviometrique	1.1833	9.55
006	SUPR	6266000	KARIMA	1/1/1936	1/5/2000	Station synoptique	31.8	18.5
000	CDPR	1080000100	KINSHASA-N'DOLO(LEOPOLDVILLE)	1/1/1931	1/3/2000	Station pluviometrique	15.3167	-4.3167
112	MLPR	1270022900	KITA	1/1/1931	1/12/2000	Station synoptique	-9.4667	13.0667
25	LYPR0	6227100	KUFRA	1/1/1933	1/7/1999	Station synoptique	23.3	24.2

000	GAPR	1140000100	LIBREVILLE AERO	9	01/05/185	1/9/2001	Station synoptique	9.4167	0.45
060	SNPR	1380015100	LINGUERE		1/1/1933	1/12/2004	Station synoptique	-15.1167	15.3833
000	TOPR	1470000100	LOME AERO	2	01/03/189	1/12/1999	Station synoptique	1.2519	6.1675
000	AOPR	1020000100	LUANDA Observatoire	9	01/01/187	1/12/1989	Station synoptique	13.2167	-8.8167
060	NGPR	1330056500	MAIDUGURI		1/1/1909	1/9/2000	Station synoptique	13.0833	11.85
064	NEPR	1320012100	MAINE SOROA		1/1/1936	1/12/2002	Station synoptique	11.9833	13.2333
029	GNPR	1170618000	MAMOU		1/1/1921	1/12/2000	Station synoptique	-12.0833	10.3667
96	CIPR0	1090014200	MAN-AERO		1/1/1922	1/12/2000	Station synoptique	-7.5167	7.4
120	CMPR	1050032000	MAROUA SALAK		1/8/1926	1/12/2001	Station synoptique	14.2572	10.4603
068	SNPR	1380016300	MATAM		1/1/1918	1/11/2003	Station synoptique	-13.25	15.65
011	MAPR	6019500	MIDELT		1/6/1931	1/12/2008	Station synoptique	-4.73	32.68
076	NEPR	1320014500	N GUIGMI		1/1/1921	1/12/2002	Station synoptique	13.1167	14.25
000	TDPR	1460000100	N'DJAMENA AERO (FORT LAMY)		1/9/1904	1/9/2003	Station synoptique	15.0333	12.1333
160	MLPR	1270032800	NARA		1/1/1921	1/12/2000	Station synoptique	-7.2833	15.1667
000	NEPR	1320000100	NIAMEY AERO		1/1/1905	1/12/2000	Station synoptique	2.1667	13.4833

CIPR1 11	1090016000	ODIENNE	1/1/1921	1/12/2000	Station synoptique	-7.564765	9.539887
BFPR 000	1200000100	OUAGADOUGOU AERO	1/1/1902	1/12/2001	Station synoptique	-1.5167	12.35
BFPR 116	1200024100	OUAHIGOUYA	1/1/1920	1/12/2001	Station synoptique	-2.4333	13.5833
MRPR 033	1300005200	OUALATA	1/6/1936	1/10/2003	Station pluviometrique	- 7.0167	17.2833
MAPR 015	6018500	SAFI	1/1/1900	1/12/2008	Station synoptique	-9.23	32.28
MLPR 195	1270040000	SAN	1/1/1921	1/12/2001	Station synoptique	-4.8333	13.3333
CIPR1 21	1090017200	SASSANDRA	1/1/1922	1/12/2000	Station synoptique	-6.0833	4.95
TNPR 071	1486658811	SIDI SAAD GARE ou JAUGEAGE	1/1/1923	1/8/2006	Station pluviometrique	9.696389	35.39333
NEPR 090	1320016900	TAHOUA	1/1/1921	1/12/2000	Station synoptique	5.25	14.9
GHPR 114	1160083500	TAKORADY METEO	1/1/1939	1/5/2000	Station synoptique	-1.7667	4.8833
DZPR 070	6068000	TAMANRASSET	1/1/1925	1/4/2000	Station synoptique	5.5	22.8
SNPR 103	1380025300	TAMBACOUNDA	1/6/1919	1/12/2003	Station synoptique	-13.6833	13.7667
SNPR 107	1380026500	THIES	1/1/1918	1/12/2003	Station pluviometrique	-16.95	14.8
DZPR 059	6060700	TIMIMOUN	1/1/1939	1/4/2000	Station synoptique	0.28	29.25
DZPR 067	6065600	TINDOUF	1/1/1935	1/4/2000	Station synoptique	-8.13	27.67

121 GHPR	1160092000	WA METEO	1/1/1939	1/12/2005	Station synoptique	-2.5	10.0667
230 CMPR	1050056400	YAOUNDE (AERO)	01/04/1889	1/7/2003	Station synoptique	11.5269	3.84
114 SNPR	1380028600	ZIGUINCHOR	1/5/1918	1/12/2003	Station synoptique	-16.2667	12.55
01 LYPR0	6200700	ZUARA	1/1/1920	1/1/2000	Station synoptique	12.08	32.88