Global Gridded Precipitation and Temperature Datasets Uncertainty in Climate Change Impact Studies

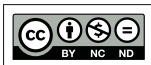
by

Mostafa Tarek Gamaleldin Galal IBRAHIM

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, DECEMBER 15TH, 2020

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

M. François Brissette, Thesis Supervisor
Departement of Civil Engineering, École de technologie supérieure

M. Richard Arsenault, Co-supervisor
Departement of Civil Engineering, École de technologie supérieure

M. Hany Moustapha, President of the Board of Examiners
Departement of Mechanical engineering, École de technologie supérieure

Mrs. Annie Poulin, Member of the jury
Departement of Civil Engineering, École de technologie supérieure

Mrs. Mélanie Trudel, External Independent Examiner
Departement of Civil and Building Engineering, Université de Sherbrooke

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON DECEMBER 7TH, 2020

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

First of all, I would like to start by expressing my profound thanks to my Thesis supervisor François Brissette and my co-supervisor Richard Arsenault who made all of this work possible.

I am extremely grateful to Professor François Brissette for always being there with valuable guidance and encouragement. Your constant support, helpful advice and patience played a key role enabling me to reach this point. I have loved the way you have been teaching me during the three years spent under your supervision. You made everything easy throughout my journey in completing this work. No words can explain how much I have appreciated your support and time in the lab.

I would like to express my gratitude to my co-supervisor, Professor Richard Arsenault for his support, valuable advice and especially his quick response to my late-time bothering emails throughout my time at ETS. Whenever I needed help, you were there.

Having two highly accomplished researchers in the field of climate change as my supervisors enabled me to learn a lot either in the field or on how to tackle research problems in general. I have never felt that I was working under any pressure. I can say that I was fortunate to have such pleasant and highly accomplished supervisors like you.

I am deeply grateful to the jury members; Professor Hany Moustapha, Professor Annie Poulin and Professor Mélanie Trudel for taking the time to review my doctoral thesis.

Last but not least, I would like to thank my family, whose support has contributed so much to the completion of this project. My wife, Samar Reda, has been an infinite source of inspiration, encouragement and understanding during my time as a Ph.D. student, and this has made me realize that together, anything is possible.

Contribution des données climatiques de référence à l'incertitude des études sur les impacts du changement climatique

Mostafa Tarek Gamaleldin Galal IBRAHIM

RÉSUMÉ

Les études d'impact du changement climatique nécessitent un ensemble de données climatologiques de référence fournissant une période de référence par rapport à laquelle évaluer les changements futurs et post-traiter les biais des modèles climatiques. Des ensembles de données climatiques à haute résolution, interpolées sur une grille à partir des stations météorologiques, sont disponibles pour la précipitation et la température dans la plupart des régions ayant des réseaux d'observations bien développés, comme l'Europe et les États-Unis. Cependant, dans de nombreuses régions du monde, la faible densité des réseaux d'observation rend les jeux de données basés sur ces observations très incertains. Des ensembles de données de satellites, de réanalyses et de produits fusionnés ont été utilisés pour surmonter cette lacune. Cependant, on ignore le degré d'incertitude que le choix d'un ensemble de données de référence peut apporter aux études d'impact.

Pour répondre à cette question, cette étude compare les ensembles de données globales / quasi globales de précipitations et de température sur 3138 bassins versants nord-américains (haute densité de stations) et 1145 bassins versants africains (faible densité de stations) pour évaluer la contribution de l'incertitude de l'ensemble de données aux études de changement climatique. Ces ensembles de données couvrent tous une période commune de 30 ans, ils pourraient donc tous être utilisés comme ensembles de données de référence pour les études d'impact du changement climatique. Les ensembles de données sur les précipitations comprennent deux produits uniquement basés sur les observations des stations météorologiques (GPCC, CPC Unified), deux produits satellites (CHIRPS et PERSIANN-CDR) corrigés à l'aide d'observations au sol, quatre produits de réanalyse (JRA55, NCEP-CFSR, ERA-I et ERA5) et un produit fusionné de stations, de satellites et de réanalyses (MSWEP). Les jeux de données de température comprennent un produit station (CPC Unified) et deux produits de réanalyse (ERA-I et ERA5).

Toutes les combinaisons des 9 jeux de précipitations et des 3 jeux de températures ont été comparées et utilisées comme données d'entrée à des modèles hydrologiques globaux pour évaluer leur performance. Ils ont également été utilisés pour évaluer les changements dans les débits futurs et pour évaluer l'incertitude liée aux jeux de données par rapport à celle provenant d'autres sources d'incertitude. L'étude d'impact du changement climatique a utilisé une chaîne classique de modélisation hydroclimatique utilisant 10 MCG CMIP5 sous RCP8.5 et deux modèles hydrologiques globaux (HMETS et GR4J) pour générer des débits futurs sur la période 2071-2100. Une décomposition de la variance a été effectuée pour comparer la contribution à l'incertitude des diverses sources d'incertitude.

Sur l'Amérique du Nord, les résultats ont montré que les trois ensembles de données de température performent de manière similaire, bien que les performances du CPC soient systématiquement inférieures aux deux autres. Des différences significatives de performance ont

cependant été observées entre les jeux de données de précipitations. MSWEP a obtenu les meilleurs résultats, suivi par les deux ensembles de données basés sur les stations, des réanalyses et des données satellite. ERA5 était la réanalyse la plus performante, tandis que CHIRPS était le meilleur produit satellite. Il a également été constaté que les performances relatives des ensembles de données dépendent de la région. Les résultats montrent que les jeux de données basés sur des stations devraient être préférés dans les régions où la densité du réseau météorologique est bonne, mais que CHIRPS et ERA5 seraient de bonnes alternatives dans les régions où les données de stations sont plus rares.

Pour l'étude d'impact du changement climatique en Afrique, les résultats montrent que toutes les combinaisons de jeux de données de précipitations et température fournissent de bonnes simulations d'écoulement fluvial sur la période de référence. Toutefois, 4 jeux de données de précipitations ont surpassé les autres pour la plupart des bassins versants: ils sont dans l'ordre MSWEP, CHIRPS, PERSIANN et ERA5. Ces ensembles de données les plus performants diffèrent de ceux identifiés en Amérique du Nord, démontrant l'impact de la densité des stations météorologiques.

Pour l'étude d'incertitude sur les changements climatiques, les deux jeux de données de température ont fourni un niveau d'incertitude négligeable. Cependant, l'ensemble des neuf jeux de données de précipitations a fourni une incertitude égale ou supérieure à celle liée aux MCG pour la plupart des métriques d'écoulement et sur la plupart des bassins versants. La sélection des 4 jeux de données de précipitation les plus performants sur l'Afrique (ensemble de crédibilité) a considérablement réduit l'incertitude attribuée aux précipitations pour la plupart des métriques, mais est toujours restée la principale source d'incertitude pour certaines autres. Le choix d'un ensemble de données de référence peut donc être critique pour les études d'impact du changement climatique, car de petites différences apparentes entre les jeux de données sur une période de référence commune peuvent se propager et générer de grandes incertitudes dans les débits futurs.

Mots-clés: Changements climatiques, Incertitude, Ensemble de données climatiques de référence, Précipitation, Température, Modélisation hydrologique, Modèle climatiques globaux

Global Gridded Precipitation and Temperature Datasets Uncertainty in Climate Change Impact Studies

Mostafa Tarek Gamaleldin Galal IBRAHIM

ABSTRACT

Climate change impact studies require a reference climatological dataset providing a baseline period against which to assess future changes and post-process climate model biases. High-resolution gridded precipitation and temperature datasets interpolated from weather stations are available in regions of high-density networks of weather stations, as is the case in most parts of Europe and the United States. In many of the world's regions, however, the low-density of observational networks renders gauge-based datasets highly uncertain. Satellite, reanalysis and merged products dataset have been used to overcome this deficiency. However, it is not known how much uncertainty the choice of a reference dataset may bring to impact studies. To tackle this issue, this study compares global/near-global precipitation and temperature datasets over 3138 North American catchments (high station-density), and 1145 African catchments (low station-density) to evaluate the dataset uncertainty contribution to the results of climate change studies. These datasets all cover a common 30-year period, so they could all potentially be used as reference datasets for climate change impact studies. The precipitation datasets include two gauged-only products (GPCC, CPC Unified), two satellite products (CHIRPS and PERSIANN-CDR) corrected using ground-based observations, four reanalysis products (JRA55, NCEP-CFSR, ERA-I, and ERA5) and one gauged, satellite, and reanalysis merged product (MSWEP). The temperature datasets include one gauged-only (CPC Unified) product and two reanalysis (ERA-I and ERA5) products.

All combinations of these 9 precipitation and 3 temperature datasets were compared and used as inputs to lumped hydrological models to evaluate their performance. They were also used to evaluate the changes in future streamflows and to assess dataset uncertainty against that of other sources of uncertainty. The climate change impact study used a top-down hydroclimatic modeling chain using 10 CMIP5 GCMs under RCP8.5 and two lumped hydrological models (HMETS and GR4J) to generate future streamflows over the 2071-2100 period. Variance decomposition was performed to compare how much the different uncertainty sources contribute to actual uncertainty.

Over North-America, the results showed that all three temperature datasets performed similarly, albeit with the CPC performance being systematically inferior to the other two. Significant differences in performance were, however, observed between the precipitation datasets. The MSWEP dataset performed best, followed by the gauge-based, reanalysis and satellite datasets categories, respectively. ERA5 was the best-performing reanalysis, while CHIRPS was the best satellite product. Relative dataset performance was also found to be region-dependent. Results show that gauge-based datasets should be preferred in regions with good weather network density, but that CHIRPS and ERA5 would be good alternatives in data sparse regions.

For the climate change impact study over Africa, results show that all combination of precipitation and temperature datasets provide good streamflow simulations over the reference period, but 4 precipitation datasets outperformed the others for most catchments: they are in order MSWEP, CHIRPS, PERSIANN, and ERA5. These best-performing datasets differ from the ones identified over North-America, demonstrating the impact of the density of weather stations.

For the climate change uncertainty study, the 2-member ensemble of temperature datasets provided negligible levels of uncertainty. However, the ensemble of nine precipitation datasets uncertainty that was equal to or larger than that related to GCMs for most of the streamflow metrics and over most of the catchments. A selection of the best 4 performing reference datasets over Africa (credibility ensemble) significantly reduced the uncertainty attributed to precipitation for most metrics, but still remained the main source of uncertainty for some streamflow metrics. The choice of a reference dataset can therefore be critical to climate change impact studies as apparently small differences between datasets over a common reference period can propagate to generate large amounts of uncertainty in future climate streamflows.

Keywords: Datasets, Precipitation, Temperature, Hydrological model, Hydrology, GCMs, Uncertainty, Climate changes

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

BC	Bias correction
CANOPEX	The Canadian model parameter experiment database
CF	Change Factor
CFSR	Climate Forecast System Reanalysis
CHC	The Climate Hazards Center
CHIRPS	Climate Hazards group Infrared Precipitation with Stations
CHPclim	The Climate Hazards Group Precipitation Climatology
CMAES	Covariance Matrix Adaptation Evolution Strategy
CMIP5	The Coupled Model Intercomparison Project Phase 5
CMORPH	The CPC MORPHing technique
CPC Unified	Climate Prediction Center Unified Gauge
DJF	The winter season months (December, January and February)
DM	Downscaling Method
ECCC	The Environment and Climate Change Canada
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	The European Centre for Medium-Range Weather Forecast's 5th generation reanalysis
ERA-Interim	European Centre for Medium-Range Weather Forecasts Interim reanalysis
ESM	Earth System model
FAO	the Food and Agriculture Organization of the United Nations

GCM	General circulation model
GEO	Geostationary Satellites
GHCN	The Global Historical Climatology Network
GHG	Greenhouse Gases
GHGES	Greenhouse Gas Emission Scenario
GIS	The Geographic Information System
GPCC	Global Precipitation Climatology Center
GPCC	Global Precipitation Climatology Centre
GPCP	The Global Precipitation Climatology Project
GPM	The Global Precipitation Measurement
GR4J	Modèle du Génie Rural à 4 paramètres Journalier
GRDC	The Global Runoff Data Centre
GriSat	The Globally Gridded Satellite
GSMaP-MVK	Global Satellite Mapping of Precipitation moving vector with Kalman filter
GSOD	Global Summary of the Day
GTS	Global Telecommunication System
GTS	The Global Telecommunication System
GTS	The Global Telecommunication System
HMETS	Hydrological Model - École de technologie supérieure
HydroSHEDS	The Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales database

IMERG	The Global Precipitation Measurement Integrated Multisatellite Retrievals
IMH	The Institute of Meteorology and Hydrology
IPCC	The Intergovernmental Panel on Climate Change
IR	Infrared
JJA	The summer season months (June, July and August)
JMA	Japan Meteorological Agency
JRA-25	The 25-year Japan Reanalysis
JRA-55	Japanese 55-year ReAnalysis
KGE	The Kling-Gupta Efficiency metric
LEO	Low Earth Orbiting satellites
MAE	Mean Absolute Error
MBCn	N-dimensional Multivariate quantile mapping Bias Correction
ME	Mean Error
MOS	Model Output Statistics
MSWEP	The Multi-Source Weighted-Ensemble Precipitation
NA	North America
NAC^2H	The North American Climate Change and hydroclimatology database
NCAR	The National Center for Atmospheric Research
NCDC	The National Climatic Data Center
NCDC	The National Climatic Data Center

NCEP	National Centers for Environmental Prediction
NCEP-CFSR	The National Centers for Environmental Prediction Coupled Forecast System Reanalysis
NESDIS	The National Climatic Data Center
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NSE	The Nash-Sutcliffe Efficiency metric
NWIS	The National Water Information System
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
PERSIAN-CCS	PERSIANN-Cloud Classification System
PERSIAN-CDR	PERSIANN Climate Data Record (CDR)
PET	Potential EvapoTranspiration
PIP	Precipitation Intercomparison Projects
PMW	Passive Microwave
RCM	Regional Climate Model
RCPs	Representative Concentration Pathways
RIHMI	The Russian Research Institute for Hydrometeorological Information
RMSE	Root Mean Square Error
SASSCAL	The Southern African Science Service Centre for Climate Change and Adaptive Land Management

SH	Southern Hemisphere
TF	Transfer function
TIR	Thermal InfraRed
TMPA	The Tropical Rainfall Measuring Mission (TRMM) Multi-Satellite Precipitation Analysis
UCAR	The USA from the University Corporation for Atmospheric Research
USA	United States of America
USGS	The United States Geological Survey
WG	Weather generators
WMO	World Meteorological Organization
WWF	The World Wildlife Fund

LISTE OF SYMBOLS AND UNITS OF MEASUREMENTS

CO_2	Carbon Dioxide
°C	Degree Celsius
ha	Hectare
hPa	Hecto-Pascal
hr	Hour
km	Kilometer
km^2	Square Kilometers
mm	Millimeter
m	Meter
m^3/sec	Cubic meter per second
N	North
r	Correlation coefficient
S	South
β	Bias component
γ	Variability component

INTRODUCTION

General Circulation Models/Earth System models (ESM) /Global Climate Models (GCMs) are the primary tools used to simulate the response of the global climate system to increases in greenhouse gas concentrations and to generate future climate projections. GCMs are complex mathematical representations of the physical and dynamical processes governing atmospheric and oceanic circulations as well as the interactions with the land surface. In order to reduce the computation burden, which can be considerable, GCMs represent the earth with a grid having a relatively coarse spatial resolution (IPCC, 2001). Consequently, GCM projections cannot be used directly for fine scale climate impact studies. Statistical/empirical or dynamical downscaling techniques have thus commonly been used to address this scale mismatch. In addition, climate model outputs are always biased, and the extent of these biases can be evaluated through a comparison against observations over a common reference period. A bias correction procedure is therefore generally performed in addition to the downscaling step, and biases are assumed to be invariant in time when the correction is applied to future climate projections (Velázquez *et al.*, 2015). Although a two-step downscaling bias correction approach is preferable in most cases, a single instance of bias correction is sometimes used to account for both scale mismatch and GCM biases. While this may be acceptable when the scale difference is small (e.g., when using catchment averaged values), recent studies have shown that bias correction has limited downscaling skills (Maraun, 2016).

Statistical downscaling, bias correction approaches as well as the calibration of hydrological model primarily rely on hydrometeorological observations over a historical reference period. It is therefore primordially important that the observed reference dataset represents the true climate state as closely as possible. For this task, ground stations remain the standard and most accurate/trusted source of weather data (New *et al.*, 2001; Nicholson, 2013). However, the spatial distribution of these stations varies widely across the globe, and coverage is often sparse and even deficient in many parts of the world outside of Europe and the US. Even in well-

covered regions, gauge data is subject to many problems, such as missing data, precipitation undercatch and inhomogeneities related to a variety of issues such as equipment change, station relocation and land surface modifications near each station (Kidd *et al.*, 2017; Peterson *et al.*, 1998).

0.1 Problem Statement

In recent decades, extensive efforts have been devoted to the development and improvement of gridded global and quasi-global climate datasets to overcome the limitations of gauge stations. These datasets provide meteorological record time series with continuous spatiotemporal coverage, and typically, no missing data. However, various error sources are inherent in these datasets, thus also bringing uncertainty to the data (Voisin *et al.*, 2008). Thus, choosing an appropriate reference dataset for climate change impact studies is an important concern, and especially so in regions with sparse ground station coverage.

According to Huth (2004): "For estimates based on downscaling of General Circulation Model (GCM) outputs, different levels of uncertainty are related to: (1) GCM uncertainty or inter-model variability, (2) scenario uncertainty or inter-scenario variability, (3) different realizations of a given GCM due to parameter uncertainty (inter-model variability) and (4) uncertainty due to downscaling methods". In most climate change impact studies, it is generally assumed that GCMs are the major source of uncertainty (Mpelasoka & Chiew, 2009; Kay *et al.*, 2009; Vetter *et al.*, 2017). Rowell (2006) compared the effect of different sources of uncertainty using the initial condition ensembles of different General Circulation Models (GCMs), Greenhouse Gas Emission Scenarios (GHGESs) and Regional Circulation Models (RCMs) on changes in seasonal precipitation and temperature in the United Kingdom. The results indicated that the largest uncertainty comes from the GCM choice. Minville *et al.* (2008) used ten equally-weighted climate projections derived from a combination of five GCMs, two GHGES and the change factor approach for downscaling to investigate the uncertainty envelope of future

hydrologic variables. Their results showed that the uncertainty related to the GCM choice is dominant. These results have also been confirmed by several studies (Prudhomme & Davies, 2009; Nóbrega *et al.*, 2011; Dobler *et al.*, 2012)).

Other studies have assessed other sources of uncertainty such as Greenhouse Gases Emission Scenarios (GHGESs) (Prudhomme *et al.*, 2003; Kay *et al.*, 2009; Chen *et al.*, 2011a), the downscaling method (Wilby & Harris, 2006; Khan *et al.*, 2006) and hydrological modeling (Bae *et al.*, 2011; Vetter *et al.*, 2017). Recent studies have also looked at the uncertainty related to the choice of the impact model (Giuntoli *et al.*, 2018; Krysanova *et al.*, 2018). From these studies, a more complex picture emerges, in which the main source of uncertainty may vary, depending on geographical location and metric under study. Dataset uncertainty has been assessed in numerous studies either by direct inter-comparison between datasets (Vila *et al.*, 2009; Andermann *et al.*, 2011; Romilly *et al.*, 2011; Jiang *et al.*, 2012; Chen *et al.*, 2014; Prakash *et al.*, 2018; Nashwan & Shahid, 2019) or by using hydrological modeling (Behrangi *et al.*, 2011; Beck *et al.*, 2017b; Wu *et al.*, 2018; Zhu *et al.*, 2018; Tarek *et al.*, 2019). However, to the best of our knowledge, the uncertainty of gridded datasets has not been evaluated against other sources of uncertainties when performing climate change impact studies.

0.2 Study Objectives

The main objective of this study is therefore to ***assess the impact of the choice of a given reference dataset on the global uncertainty chain of climate change impact studies in hydrology***. Since this is of particular concern to regions with sparse weather station coverage, this study is conducted over Africa. To achieve this main objective, this work is further divided into **three secondary objectives**:

1. Evaluate the performance of relevant precipitation and temperature datasets in North America over 3138 catchments, where the gauge-station networks are high;

2. Re-evaluating the datasets performance in data-sparse regions. This step will be performed over 350 African catchments;

3. Assess the uncertainty of the reference dataset compared to other sources of uncertainty.

0.3 Thesis Organization

This thesis consists of six chapters organized as follows: Literature review, Study region and Data, Methodology, Results, Discussion, and finally Conclusions and recommendations.

Chapter one is devoted to the literature review. It introduces the global warming phenomena and the key indicators of climate changes. It also discusses how to evaluate the impacts of climate changes by using climate projections from general circulation models (GCMs). It covers the issues of downscaling to address the scale mismatch between the model and the regional study, as well as the assessment of uncertainty. Then, the different types of gridded precipitation and temperature datasets that can be used as potential reference datasets are described. Finally, it covers the topic of hydrological models which are used in the study of climate change impacts on water resources.

Chapter two first presents the target study area of Africa. It covers the geography, climate, physiographic characteristics and vulnerability of the main African catchments and rivers. The precipitation and temperature datasets are then described in four main sections: Section 1 describes the criteria which were used to select the participating precipitation and temperature datasets. A brief description of each of these datasets follows. Section 2 presents the observed streamflow data, which was obtained from two data sources: The Global Runoff Data Centre (GRDC) database for the African catchments, and the North American Climate Change and Hydroclimatology (NAC^2H) database for the North American catchments. Section 3 describes the ten used GCMs that were obtained from the Coupled Model Intercomparison Project Phase

5 (CMIP5). Finally, section 4 describes the HydroSHEDS database that was used to provide the watershed boundaries of the African catchments.

Chapter three describes the methodology used to conduct this research, which is divided into three steps. In the first step, the performance of the chosen precipitation and temperature datasets is assessed over North American watersheds which are known to generally have a good coverage of weather stations. The analysis is conducted in two parts: (1) a statistical evaluation comparing the Mean Error, Mean Absolute Error, Root Mean Square Error and correlation coefficient, as well as an intercomparison between the selected gridded datasets at the seasonal and annual scales, (2) an indirect evaluation using hydrological modelling to simulate streamflows over 3138 catchments. In the second step, the performance of the same datasets is re-evaluated over Africa using a similar methodology. Two lumped hydrological models are implemented for this evaluation. In the third and final step, a climate change impact study over 1145 African catchments is performed using a top-down hydroclimatic modeling chain. A variance decomposition analysis is performed to compare the uncertainty related to reference datasets against that of GCMs and hydrological models for 51 streamflow metrics over 1145 African catchments.

Chapter four outlines the main findings of the study in a systematic and detailed way. The results are presented following the same sequence as described in the methodology. The results of the gridded datasets intercomparison and the hydrological modelling analysis in North America are shown first. Then, the re-evaluation results of the same datasets over Africa are presented. Finally, the results of the variance decomposition analysis are presented.

Chapter five presents a general discussion on the obtained results.

Finally, **Chapter six** presents the main conclusions and summarizes the main contributions of this Thesis, as well as providing recommendations for future research work. The four appen-

dices that follow, each presents a published/submitted paper. These papers are briefly described in the contribution to science section.

CHAPTER 1

LITERATURE REVIEW

1.1 Global warming

Global warming has become one of the most important environmental challenges for humanity. In its Fifth Assessment Report, the Intergovernmental Panel on Climate Change (IPCC), a group of 1,300 independent climate scientists from countries all over the world, agreed that industrial activities, that our modern civilization depends upon, considered the main cause of the current global warming trend (IPCC, 2014). The IPCC stated that: “Human influence on the climate system is clear. This is evident from the increasing greenhouse gas concentrations in the atmosphere, positive radiative forcing, observed warming, and understanding of the climate system”. No one knows how much warming is “safe” but it is known that climate change is already disrupting the climate system and water cycle processes.

Key indicators of global climate change are attributed to the steady increase of atmospheric trace gases, such as carbon dioxide (CO_2), methane (CH_4), nitrous oxides (NO_2), and chlorofluorocarbons (CFC or "Freon"). These gases are commonly referred to as "greenhouse gases". These gases absorb sunlight and infrared radiation that have reflected from earth’s surface and block the heat from escaping to space. As the concentration of greenhouse gases increases, less long-wave radiation escapes, and the earth’s surface temperature has to increase to balance its radiative budget. This phenomenon, which is known as the “Greenhouse Effect” has been well-represented and studied through climate models.

There are many indicators of global warming such as changes in surface temperature, precipitation, atmospheric water vapor, warmer oceans, severe events, higher sea levels and reduced amounts of snow and ice (IPCC, 2001; Fowler *et al.*, 2007). For engineers, the impacts to temperature and precipitation are key since those variables control the distribution of water resources in both time and space. Distribution and availability of water is a key concern for the health and safety of many countries.

To better evaluate the impacts of future climate change, this requires the development of high quality climate projections with the best possible estimation of uncertainty. General circulation models (GCMs) and Earth System models (ESMs) are the major tools that are used to project the earth's climate in the future.

1.2 General Circulation Models and climate scenarios

1.2.1 General Circulation Models (GCMs)

General Circulation Models or Global Climate Models (GCMs) are complex mathematical representations of physical and dynamical processes to simulate the interaction between the atmosphere, land surface, oceans and sea ice. Moreover, they are used to simulate the response of the global climate system to the increase of greenhouse gas concentration. GCMs are physically based on the principle of fluid dynamics and describe the entire globe using a three-dimensional grid. In order to reduce the huge calculation requirements, general circulation models usually have to use coarse spatial resolutions (a few hundreds of kilometers) to simulate the globe, typically having a horizontal resolution between 110 and 500 km. Moreover, vertical layers range from 10 to 20 layers in the atmosphere and sometimes as many as 30 layers in the oceans (IPCC, 2001).

To depict long-term climate change, the model has to be provided with inputs on how the environment is changing. One of the most critical variables is fossil-fuel emissions. Greenhouse gases were represented by injecting carbon dioxide to the model. In recent years, more detailed models have used methane and other key gases (Henson, 2011). Two major types of simulations used to study the effect of greenhouse gases on climate:

1. Equilibrium: runs by instant and massive injection of carbon dioxide; i.e. the amount is twice the pre-industrial level, which is expected to be reached by the end of the century.

2. Transient: runs more closely representing reality; i.e. the carbon dioxide is added in smaller amounts every year to represent the actual amounts added (and predicted for the

future) by human activities. Scientists are more interested in this type of simulation as this scenario is closer to reality and shows a good agreement against past trends.

1.2.2 Greenhouse Gases Emission Scenarios (GHGESs)

The estimation of future greenhouse gases (GHG) emissions is recognized with high uncertainty (Hegerl *et al.*, 2007; Min *et al.*, 2011). In 1990, the IPCC produced a first set of global GHG emission (SA90) which consisted of four scenarios (A-D) (Houghton *et al.*, 1990). Then, new information regarding the increasing population and the economic growth rates became available. Consequently, these scenarios were updated in 1992 and a new set of 6 scenarios (IS92) was (IS92a-f) (Leggett *et al.*, 1992).

Much additional knowledge was gained since then, which prompted the IPCC in 1996 to provide another set of scenarios for its Third Assessment Report. This new set of scenarios was divided into four different narrative story-lines (A1, A2, B1 and B2) called "families". Each family has one group, with the A1 family containing three groups based on future energy technologies; A1F1 (fossil fuel intensive), A1B (balanced), and A1T (predominantly non-fossil fuel). Overall, 40 different scenarios were in the Special Report on Emissions Scenarios (SRES), which described the relationship between the emissions and the latest available information on technology change, social and economic development as well as the population projections which are the main driving forces of future greenhouse gas trajectories (Nakicenovic *et al.*, 2000).

Last but not least, the IPCC decided to update its emission scenarios in its fifth assessment report (AR5). For this report, the IPCC produced a standard set of scenarios called the Representative Concentration Pathways (RCPs), where population, economic activities, energy consumption, lifestyle and climate policies are factored into. The RCPs consist of four scenarios (RCP2.6, RCP4.5, RCP6.0 and RCP8.5), each representing a specific radiative forcing pathway (Pachauri *et al.*, 2014). For example, RCP8.5 refers to a radiative forcing that reaches > 8.5 W/m^2 by 2100. As a result of these different radiative forcings, the average increase in

the global surface temperatures varies between each pathway. These variations are likely to be between (0.3 - 1.7) °C for RCP2.6, (1.1 - 2.6) °C for RCP4.5, (1.4 - 3.1) °C for RCP6.0 and (2.6 - 4.8) °C for RCP8.5 (Stocker *et al.*, 2013).

1.2.3 General Circulation Models and climate scenarios Uncertainties

The processes to describe natural phenomena are always incomplete. Science tries to represent things and events in a way that is “very close to reality”. Scientifically speaking, the gap between reality and science representation is known as "Uncertainty". In other words, the natural phenomena is very complex to be accurately represented throughout the modeling process. This incomplete scientific knowledge results in Uncertainty.

Different climate models may use different computational approaches to represent the climate processes. Because of this, different climate models may react differently to an identical forcing. This different climate sensitivity is the main source of GCMs uncertainty. Through the use of a large number of GCMs, GHGESs, downscaling methods (DMs) and different hydrological model structures, it has become increasingly challenging to assess the uncertainties resulting from their combination (Chen *et al.*, 2011a).

Generally, it is considered that the largest sources of uncertainty are linked to General Circulation Models (GCMs) and Greenhouse Gases Emissions Scenarios (GHGESs) for quantifying the impacts of climate change (Mpelasoka & Chiew, 2009; Kay *et al.*, 2009; Vetter *et al.*, 2017). GCM simulations of precipitation and temperature have to be used with caution as they often show significant biases (Déqué *et al.*, 2007; Teutschbein & Seibert, 2012). However, other sources of uncertainty have been given less attention (Chen *et al.*, 2011a).

Rowell (2006) compared the effect of different sources of uncertainty using different GCMs, emission scenarios and Regional Climate Model (RCM) initial condition ensembles on the changes in seasonal precipitation and temperature for the United Kingdom, and showed that of all the sources of uncertainty, that related to choice of GCM is normally dominant. These

results have been also confirmed by several other studies (Prudhomme & Davies, 2009; Kay *et al.*, 2009; Green *et al.*, 2014).

Minville *et al.* (2008) used ten equally weighted climate projections from a combination of five general circulation models (GCMs), two greenhouse gas emission scenarios (GHGESs) and the change factor approach for downscaling to investigate the uncertainty envelope of future hydrologic variables. The results indicated that the largest uncertainty comes from the choice of GCM.

Prudhomme & Davies (2009) investigated the uncertainty in the impact of climate change on water resources for four catchments in Britain. They concluded that the largest source of uncertainty comes from GCMs, likewise the downscaling uncertainty is significant. While, hydrological modeling uncertainty was found to vary significantly between catchments.

Vetter *et al.* (2017) evaluated the main contributors to uncertainty on twelve river catchments located in six continents. Their study included nine hydrological models (HMs), four Representative Concentration Pathways (RCPs) and five General Circulation Models (GCMs). The analysis of variance (ANOVA) was used to quantify uncertainties from the three different sources (e.g., GCMs, RCPs and HMs). Their results showed that GCMs dominated the uncertainty followed by RCPs and HMs, respectively.

Several studies worked to assess the greenhouse gas emission scenarios uncertainty in climate change impact studies. Prudhomme *et al.* (2003) used four GHGES to evaluate the output of several GCMs on five catchments in Great Britain. The results showed little uncertainty among the different emission scenarios. Kay *et al.* (2009) used a range of GHGESs and applied them on two catchments in England. The results showed that the GHGESs uncertainty for the larger and flatter catchments was very low, while it was more important for the smaller and steeper catchments. However, Chen *et al.* (2011a) compared the uncertainty of two emission scenarios, A2 and B1, against other sources of uncertainty. The results showed that emission scenarios had the least uncertainty compared to other sources.

1.3 Downscaling

Generally, GCMs are run at an average spatial resolution between 150-300 km, while regional studies (such as hydrological studies), require a finer resolution and even in some cases less than 10 km. To overcome this problem, downscaling techniques are commonly used to address the scale mismatch between the coarse resolution outputs of global climate models (GCMs) and the finer catchment scale required for climate change impact assessment and hydrological modeling.

There are two main types of downscaling methods; dynamical and statistical methods; and each has its advantages and drawbacks. A growing number of studies have compared statistical and dynamical methods (Boé *et al.*, 2007; Pierce *et al.*, 2013; Huth *et al.*, 2015; Ayar *et al.*, 2016). These two methods are discussed in more detail in the next sections.

1.3.1 Statistical downscaling

Traditionally, statistical downscaling has been used as an alternative to dynamical downscaling. These methods fall into three categories: transfer function, weather generator and weather typing (Chen *et al.*, 2011a). No one can say that there is a particular type of downscaling that is superior to all others and many climate change impact studies use more than one technique (Chen *et al.*, 2012; Onyutha *et al.*, 2016; San-Martín *et al.*, 2017). However, statistical downscaling has drawbacks that must be taken into account in practical applications. First, the spatial pattern of climate will remain constant under different future climate conditions and ignoring change in variability (Fowler *et al.*, 2007). In addition, data needed to develop those relationships may not be available in regions with complex topography. Furthermore, an ideal statistical downscaling needs a strong statistical relationship explaining completely the variability of the local scale variable, whereas this is almost never the case as the predictors never describe all the variability of the local variable which is also affected by local factors not accounted by the large scale fields (Wilby *et al.*, 2004).

Another widely used popular downscaling method is the change factor (CF) method (Diaz-Nieto & Wilby, 2005; Minville *et al.*, 2008; Anandhi *et al.*, 2011; Goodarzi *et al.*, 2015). This method works as it adjusts the observed reference series by multiplying the ratio (for precipitation) and adding the difference (for temperatures) between future and present climates as simulated by GCMs. The most significant drawback of this method is the inability in changing the temporal sequence of wet and dry days.

1.3.1.1 Transfer function

Transfer function (TF) approaches are based on establishing statistical linear or nonlinear relationships between GCM large-scale outputs (predictors) and the observed regional or local climatic variables (predictands) (Fowler *et al.*, 2007). This method is relatively easy to apply, but its main obstacle is the probable lack of a stable relationship between predictors and predictands, such that climate variables may change from one year to another within the same circulation regime.

1.3.1.2 Weather generators

Weather generators (WG) develop synthetic time series of weather data based on the statistical characteristics of observed weather. Two steps are typically used to predict stochastic weather data; the first is to model daily precipitation, and then to model the remaining variables of interest, such as daily maximum and minimum temperature, solar radiation, humidity and windspeed conditional on precipitation occurrence. To reflect seasonal variability different model parameters are usually required for each month. The most likely feature of using a weather generator approach is its ability to rapidly produce sets of climate scenarios for studying the impacts of rare climate events and investigating natural variability (Chen *et al.*, 2011a).

1.3.1.3 Weather typing

Weather typing schemes group local meteorological variables in relation to different classes of atmospheric circulation. The main advantage of this approach is the stationary relationship between local variables and global circulation. However, its reliability depends on the relationship between local climate and large-scale circulation (Chen *et al.*, 2011a). More recently, Model Output Statistics (MOS) (Pinto *et al.*, 2010) methods have been largely favored by the climate change impact community. MOS methods aim at directly correcting climate model outputs (most typically temperature and precipitation) rather than trying to rely on large-scale atmospheric variables that are normally better represented by regional and global climate models. Bias correction (BC) is the class of MOS now used in most climate change impact studies. While there are many possible variants of bias correction, most use some form of quantile mapping, in which the distributions of climate variables are mapped onto the distribution of observations over the reference period. These methods can be univariate (one variable corrected at a time) or multivariate (multiple variables corrected at the same time). The work of Cannon (Cannon, 2016, 2018; Cannon *et al.*, 2015) present good examples of both cases.

1.3.2 Dynamical downscaling

Dynamical downscaling consists of using Regional climate models to downscale General Circulation models to a finer scale. Dynamical downscaling is computationally very expensive to run, and still contain biases inherited from their parent GCM and originating from their own formulation and parameterization. They therefore still require a bias correction step to make them amenable to be used in impact models.

The dynamical approach is also constrained by the availability of RCM simulations and driving GCMs to adequately cover climate model uncertainty. Accordingly, statistical downscaling (and MOS methods in particular) remain the most popular alternative for climate change impact studies due to their relative ease of use and general performance (Eden & Widmann, 2014).

In particular, some downscaling methods are unable to capture the extremes of climate events that are often of particular concern in hydrology. For example, changes in precipitation extremes would have very important impacts. Therefore, it is important to study the patterns of such events. Changes of extremes are evaluated at a range of temporal and spatial scales, i.e., from globally extremely warm years to locally peak rainfall intensities. To span this entire range, data is required at a daily or sub-daily time step (Singh & Patwardhan, 2012).

1.3.3 Downscaling Uncertainty

Downscaling methods add uncertainty due to the limitations of each technique. Accordingly, the choice of a downscaling method is important when studying the uncertainty. Wilby & Harris (2006) studied the impact of climate change on low flows in the Thames river using four GCMs, two emissions scenarios, two statistical downscaling techniques, two hydrological model structures and two sets of hydrological model parameters. They concluded that the results are most sensitive to the choice of GCM as well as the downscaling method, and less sensitive to the choice of hydrological model parameters or emissions scenario.

Khan *et al.* (2006) studied the downscaling uncertainty by comparing the performance of three statistical downscaling methods using 40 years of precipitation, maximum and minimum daily data. They investigated that the performance of the three methods would remain the same under future climate forcing, as the uncertainty of their results would be mostly correlated by the GCM output uncertainty.

Chen *et al.* (2011a) studied the uncertainty using six dynamical and statistical downscaling methods in quantifying the impacts of climate change on the hydrology of a Canadian province (Quebec) river basin. The results indicate that studying climate change impacts based on only one downscaling method should be avoided.

1.4 Precipitation and temperature Gridded Datasets

Notwithstanding the limitations associated with meteorological stations as reference datasets, such as missing records, inhomogeneity, short temporal coverage, sparse spatial coverage and the inability to adequately represent the climate variability in all topographic and climatic zones, the stations are still considered to constitute the most accurate source of climate data (Tapiador *et al.*, 2012; Nicholson, 2013; Colston *et al.*, 2018).

In recent decades, to overcome some of the limitations of station data, several global and regional gridded datasets have been developed with different input data sources (gauges, radar, satellite, reanalysis or combinations thereof), spatial resolutions (0.05° to 2.5°), spatial coverage (continental to global), temporal scales (30 minutes to annual) and temporal coverage (from 1 to several years) (Henn *et al.*, 2018). Such gridded datasets provide continuous spatial and temporal coverage, and typically, with no missing data.They allow some information from regions with good network coverage to be extended, to some extent, towards areas with less information. Interpolated datasets, however, do not create new information, no matter how complex and how much additional information is used in the interpolation schemes (Essou *et al.*, 2016a; Newman *et al.*, 2015).

1.4.1 Ground-based datasets

Gridded datasets can be classified as a function of their data source. Gauge-based gridded datasets are obtained by interpolating and mapping the information measured at a small scale (typically, a point measurement at a weather station) onto a predefined spatial and temporal resolution grid. Gridded datasets are much simpler to use than their direct station data. However, variations in gauge types or instrument replacements affect error characteristics in long-term records. In addition, observations are affected by systematic biases due to evaporation and wind effects, as well as the elevation of gauges in mountainous regions (Isotta *et al.*, 2014). Gauges are also typically placed in regions allowing easier access for station maintenance and troubleshooting, meaning that the gauges do not necessarily reflect the actual climatic condi-

tions of their surroundings. Interpolated station gridded climate data products are thus subject to these limitations and many integrate adiabatic lapse rates and elevation/precipitation relationships using terrain elevations in a bid to correct some of these shortcomings. The main issue is how to map the gauge data onto the grid. One approach is to average the entire gauge records within each grid box, and, if there is no available information, the grid box is left empty. Another approach is to fill the grid boxes based on some distance measured between the closest gauges and the grid box (Kidd & Huffman, 2011).

Gauge-based datasets include the Global Precipitation Climatology Centre (GPCC), which uses data from the Global Telecommunication System (GTS) network of approximately 64,000 gauges, improved with other national network data to generate daily and monthly global land-products used as a reference dataset for numerous studies (Rudolf *et al.*, 2010). The Global Historical Climatology Network (GHCN) is another notable database with 31,000 stations (Vose *et al.*, 1992).

Another method used to measure precipitation is using the ground weather radars that become more common and available at an even higher resolution (Beck *et al.*, 2019). They partially address the issue of rain gauge spatial coverage since each radar site covers a relatively large area. Moreover, it provides much larger spatial coverage to measure precipitation than the point measurements provided by gauges. However, radar coverage is limited to developed regions that have a high population. In addition, they provide estimates of the rainfall rate at certain observational levels above the ground and cannot detect surface precipitation. Therefore, the presence of weather stations is required for the calibration and correction processes between surface measurements and atmospheric precipitation estimates (Martens *et al.*, 2013).

1.4.2 Satellite remote sensing datasets

Remotely sensed datasets have long carried the hope of bringing relevant hydrometeorological information over large swaths of land, up to the global scale, and over regions with absent or low-density observational networks (Lettenmaier *et al.*, 2015). There are now several global or

near-global precipitation datasets derived from various satellites, with high spatial and temporal resolutions and near real time coverage (Sun *et al.*, 2018). They are mainly suitable for rainfall estimation in the tropics and data-scarce regions. Given this advantage, satellite products have been used in water resource management studies (Giardino *et al.*, 2010; Nishat & Rahman, 2009; Siddique-E-Akbor *et al.*, 2014), hydroclimatological studies (Khan *et al.*, 2011; Jutla *et al.*, 2015) and in extreme event analysis (Lockhoff *et al.*, 2014; Boers *et al.*, 2015). However, satellites are relatively insensitive and generally miss a significant quantity of light precipitation and tend to fail over snow- and ice-covered surfaces (Tian *et al.*, 2009; Laviola *et al.*, 2013).

Satellite precipitation estimates can be derived from a range of observations from many different sensors. Two major types of satellites are used to measure precipitation; Geostationary (GEO) satellites and Low Earth Orbiting (LEO) satellites (Kidd *et al.*, 2012). GEO satellites orbit the Earth above the Equator at about 35,800 km. They orbit at the same rate as the Earth turns, thus appearing stationary relative to the Earth's surface. Each GEO satellite is able to view about one third of the Earth's surface. However, five operational GEO satellites are required to ensure full East–West and 60° N/S coverage due to the increasing scan angle towards the edges of the imagery. LEO satellites orbit the Earth once about every hour and a half at an altitude of about 850 km (Klaes *et al.*, 2007). Observations from LEO orbiting satellites complement the observations from GEO-based instrumentation. Sensors typically include both multi-channel IR sensors and passive microwave (PMW) sounders and imagers that are sensitive to precipitation. The NASA WetNet project organized a series of Precipitation Intercomparison Projects (PIP) concentrating on global monthly estimates and regional-scale performance. The main conclusion from those studies was that PMW techniques were clearly better than IR techniques for precipitation estimates, primarily due to their direct observation of rainfall.

All remotely sensed precipitation datasets do however only provide indirect measurements of the target variable. They typically provide biased estimates, and ground stations are often needed to correct the remotely sensed estimates (Fortin *et al.*, 2015). Some studies evaluated the uncertainties of these datasets and showed that high resolution satellite products perform

better when bias corrected using gauge observations (Xie *et al.*, 2007; Awange *et al.*, 2016). Most studies evaluated accuracy using independent gauge observations (Hirpa *et al.*, 2010; Buarque *et al.*, 2011; Alijanian *et al.*, 2017) or gauge-adjusted radar fields (AghaKouchak *et al.*, 2011; Islam *et al.*, 2012), while others just compared their spatiotemporal patterns (Kidd *et al.*, 2012, 2013). Awange *et al.* (2016) evaluated the uncertainties of these products without being dependent on the choice of a reference dataset. The results showed that the satellite merged products with rain gauge observations contain smaller error amplitudes compared to the satellite only products. Finally, other studies quantified the performance of different remote sensing datasets using hydrological modeling, by comparing simulated and observed values of river discharge (Collischonn *et al.*, 2008; Cole & Moore, 2009; Behrangi *et al.*, 2011; Bastola & François, 2012; Falck *et al.*, 2015). Generally, results of these works showed that remote sensing data are not accurate enough to consistently allow hydrological models to adequately simulate river flows.

1.4.3 Reanalysis datasets

Retrospective-analysis / reanalysis is another product that has generated interest increasingly in the recent decade and provide global four dimensional earth system data including many physical and dynamical processes (Bosilovich *et al.*, 2008). They are vital sources of data in weather and climate studies. A typical reanalysis system consists of two main components, the forecast model and the data assimilation system. Data assimilated in a reanalysis consist mostly of atmospheric and ocean data and do not typically rely on surface data, such as measured by weather stations. The role of the data assimilation system is to integrate observed databases of many sources of observations with the numerical weather forecast models to produce consistent gridded datasets (Seyyedi *et al.*, 2015). Reanalysis assimilates data from measurements derived from different sources. The main sources are floats, aircrafts, satellites and global measurement networks (Rienecker *et al.*, 2008). Floats mainly measure real time temperature and the salinity of the first 2000 meters of ocean water, while Aircrafts and satellites provide sev-

eral atmospheric data, such as radiance, wind, humidity and pressure, at different atmospheric heights (Essou *et al.*, 2016a).

Reanalysis combine a wide array of measured and remotely sensed information within a dynamical–physical coupled numerical model. They use the analysis part of a weather forecasting model, in which data assimilation forces the model toward the closest possible current state of the atmosphere. A reanalysis is a retrospective analysis of past historical data making use of the ever-increasing computational resources and more recent versions of numerical models and assimilation schemes. Reanalyses have the advantage of generating a large number of variables not only at the land surface, but also at various vertical atmospheric levels.

Although reanalysis are not direct observations and are not directly dependent on the density of surface observational networks, they provide variables throughout the world, including in areas where weather stations are non-existent or scattered (Bosilovich, 2013). These products are of specific interest especially for estimating snow and rain on snow, which are often poorly measured by satellite precipitation products (Clifford, 2010). Additionally, these models are ideally suitable for simulating the evolution of large-scale weather systems (Roads, 2003). For the previous reasons, datasets were designed for different applications and provide widely varying precipitation estimates. One of the key utilities in a reanalysis is that the output generated from the model physics provides data not easily observed, but consistent with the analyzed observed data.

Reanalysis have increasingly been used in various environmental and hydrological applications Chen *et al.* (2018); Ruffault *et al.* (2017); Emerton *et al.* (2017); Di Giuseppe *et al.* (2016). They are commonly used in regional climate modelling, weather forecasting and, more recently, as substitutes for surface precipitation and temperature in various hydrological modelling studies (Chen *et al.*, 2018; Essou *et al.*, 2016a, 2017; Beck *et al.*, 2017b). They have been shown to provide good proxies to observations and even to be superior to interpolated (from surface stations) datasets in regions with sparse network surface coverage (Essou *et al.*, 2017). Precipitation and temperature outputs from reanalysis have, however, been shown to

be inferior to observations in regions with good weather station spatial coverage (Essou *et al.*, 2017). The relatively coarse spatial resolution of reanalyses is thought to be partly responsible for this.

Widely used reanalysis products include the 44-year reanalysis from the National Centers for Environmental Prediction Coupled Forecast System Reanalysis (NCEP-CFSR) (Saha *et al.*, 2006), a 40-year reanalysis (ERA-40) from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Uppala *et al.*, 2005) and the 41-years ERA-Interim (Dee *et al.*, 2011). Numerous reanalysis validation studies have been carried out to provide a better and deeper understanding about these products. Fekete *et al.* (2004) computed runoff from observed and reanalysis precipitation, and stated that the largest errors from the reanalysis are found in arid and semiarid regions. Moreover, many studies have shown ERA-Interim (European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis) to be the best or amongst the best performing reanalysis products (Sun *et al.*, 2018; Beck *et al.*, 2017a; Essou *et al.*, 2016a, 2017), arguably the result of its sophisticated assimilation scheme, and despite a spatial resolution inferior to that of most other modern reanalyses.

In March 2019, ECMWF released the fifth generation of its reanalysis (ERA5) over the 1979–2018 period (Hersbach & Dee, 2016). ERA5 incorporates several improvements over ERA-Interim. Of particular interest to the hydrological community are the largely improved spatial (30 km) and temporal (1-hour) resolutions. The spatial resolution is now similar to or better than that of most observational networks in the world, with the exception of some parts of Europe and the United States. The hourly temporal resolution matches that of the best observational networks. In the United States and Canada, for example, there are currently no readily available observation derived precipitation and temperature datasets at the sub-daily timescale, and sub-daily records are not consistently available for weather stations. In particular, the hourly temporal resolution, if proven accurate, could open the door to many applications, and notably for modelling small watersheds for which a daily resolution is not adequate. Such watersheds are expected to be especially impacted by projected increases in extreme convective events resulting from a warmer troposphere in a changing climate. Some early results

from ERA5 have shown that it outperforms other reanalysis sets and its predecessor ERA-I (Balsamo *et al.*, 2018; Olauson, 2018; Urraca *et al.*, 2018; Tarek *et al.*, 2019). Overall, no single precipitation product could be considered ideal for measuring precipitation. In fact, all precipitation products tend to miss a significant volume of rainfall (Behrangi *et al.*, 2011).

1.4.4 Gridded datasets uncertainty

Appropriate dataset selection is a key issue in climate studies. High uncertainty is found across most gridded datasets, coming from multiple sources, such as using different data sources, merging and interpolation algorithms or quality control techniques (Vogel & Vogel, 2013; Prakash *et al.*, 2015b,a; Prein & Gobiet, 2017; Nashwan *et al.*, 2019). Moreover, the number and the accuracy of observations used to correct these products typically vary. However, some products are calibrated to the observations, thus making annual biases minimal, while their daily patterns are significantly different from the observations (Sylla *et al.*, 2013). Therefore, gridded datasets should be comprehensively evaluated before they are used.

Several studies have assessed the performance, advantages and limitations of gridded datasets. Most of these studies focus solely on precipitation datasets and evaluate the accuracy of these products through a straightforward comparison against ground weather stations (Vila *et al.*, 2009; Romilly *et al.*, 2011; Jiang *et al.*, 2012; Prakash *et al.*, 2018; Nashwan *et al.*, 2019; Andermann *et al.*, 2011) evaluated five remote sensing and gauge-based gridded datasets with ground-based measurements in the Himalayan region. The results showed that the satellite products underestimate the precipitation at both the annual and seasonal scales. The authors reported that the findings likely resulted from the bias correction techniques applied to correct the datasets using the Global Telecommunication System (GTS) rain gauge network, which has a poor spatial coverage in the study region; in addition, 0 mm precipitation is used to compensate for missing values in the database. Moreover, there is a lag experienced by the remote sensors in precisely capturing the snowfall, which is the major contributor of precipitation in the Himalayas. The conclusion that satellite approaches tend to fail in snow-dominant regions has also been reported in other studies (Kidd *et al.*, 2012; Laviola *et al.*, 2013). Chen

et al. (2014) also evaluated two satellite-based products, (CMORPH) (Joyce *et al.*, 2004) and PERSIANN-CCS (Ashouri *et al.*, 2015), to capture the rainfall in the mountainous zones located west and north of Beijing. The study showed that both datasets failed to capture the spatial pattern and the temporal variation of precipitation.

Other studies have used hydrological modelling as an indirect method to evaluate the performance of these datasets in forcing hydrological models (Zhu *et al.*, 2018; Duan *et al.*, 2019; Tarek *et al.*, 2019). Hydrological modelling offers an interesting perspective since results depend on the coherence between precipitation and temperature datasets and on an accurate representation of the annual cycle of both variables. Hydrological modelling is also not overly sensitive to biases present in every dataset, as these are typically removed during the calibration process (Essou *et al.*, 2016a).

Behrangi *et al.* (2011) evaluated five satellite-based products to force a hydrological model and simulate streamflows. The results showed that the bias-corrected datasets captured streamflow patterns well. However, the non-bias-corrected products overestimated the streamflow over warm seasons and underestimated it in cold seasons. Wu *et al.* (2018) evaluated the Multi-Source Weighted-Ensemble Precipitation (MSWEP V2.1) and three satellite-based precipitation products with rain gauge observations to simulate streamflows on the upper Huaihe River Basin in China. The results showed that the merged precipitation product (MSWEP V2.1) generally outperformed the other satellite datasets, although significant uncertainty existed in mountainous regions.

Various studies have shown that in the Northern Hemisphere (NH), the older reanalysis products such as the National Center for Atmospheric Research (NCAR), the 15-year reanalysis (ERA-15), 40-year reanalysis (ERA-40), and the 25-year Japan Reanalysis (JRA-25), generally do a good job at reproducing the spatial distribution of precipitation (Hodges *et al.*, 2003; Wang *et al.*, 2006). In the Southern Hemisphere (SH) larger differences were found, indicating a higher degree of uncertainty. This is related to how the available observations in the SH are assimilated (Hoskins & Hodges, 2005).

Beck *et al.* (2017b) evaluated 23 gridded precipitation datasets over the 2000-2016 period. Thirteen daily uncorrected datasets (non-dependent on gauges for correction) were compared with observations from gauges, and the other ten gauge-corrected datasets were evaluated using hydrological modelling. Among the uncorrected datasets, the merged-products datasets (MSWEP-ng) generally performed the best, followed by the reanalysis and then the satellite products. For the corrected datasets, results showed that datasets integrating daily gauge data (CPC Unified and MSWEP products) generally outperformed the other datasets.

Finally, precipitation datasets have also been evaluated using the surface water budget (Getirana *et al.*, 2011; Lorenz *et al.*, 2014; Munier & Aires, 2018; Sheffield *et al.*, 2009; Smith & Kummerow, 2013; Song *et al.*, 2016), as well as using surface water and energy budgets (Kang & Ahn, 2015; Hobeichi *et al.*, 2020b,a; Yang *et al.*, 2015). Despite the growing literature on the subject, the question regarding the most accurate dataset for capturing the spatio-temporal variability of weather events or driving hydrological models for climate change impact studies remains unanswered.

Near-surface air temperature is a key variable for meteorological monitoring and forecasting services (Nieto *et al.*, 2011), as well as for climate and hydrological studies. In hydrological modelling, the air temperature is the main driving variable for the evapotranspiration and snowmelt processes. Hence, accurate temperature data is vital when driving hydrological models in historical and future climate periods. However, the lack of an adequate gauge network can result in improper temperature estimations. Therefore, gridded temperature datasets are also crucial in many fields. Temperature products are generally thought to be less complex than precipitation datasets due to the much smaller spatial and temporal temperature variability in the former. Therefore, significantly fewer studies have compared and evaluated the uncertainty of using different temperature datasets in hydrological impact models (Essou *et al.*, 2016b).

1.5 Hydrological modelling

Moradkhani & Sorooshian (2009) defined a "model" as a simplified representation of a real world system. The best model is the one that represents the reality as closely as possible using the least parameters and model complexity (Devia *et al.*, 2015). The sensitivity of the hydrological cycle to diverse climate conditions makes climate change projections essential for the assessment of future variations (Teutschbein & Seibert, 2012). Extracting climate variables (temperature and precipitation) from global climate models (GCMs) and simulating using hydrological models is the most commonly used method to estimate future climate change impacts on hydrology. Therefore, the choice of the suitable hydrological model is essential.

1.5.1 Different types of hydrological models

Hydrological models represent the hydrological cycle using mathematical equations and are available under a wide range of model structures. The most important model classifications are lumped, distributed and semi-distributed models, and empirical, conceptual and physically-based models.

In lumped hydrologic models, the entire watershed is described as one unit. Generally, a lumped model can be expressed by empirical or differential equations with uniform spatial characteristics (i.e the spatial heterogeneity of most or all catchment attributes is disregarded). At the opposite, a distributed model is structured to consider these spatial variations to some extent by dividing the basin into smaller units or sub-basins. Consequently, model complexity rises rapidly, which increases the computational effort needed to run this class of model. Semi-distributed models try to navigate the middle ground to allow for the representation of spatial heterogeneity while keeping the computational budget reasonable (Orellana *et al.*, 2008).

Empirical models are observation models that only use mathematical relationships between input and output time series without considering the physical processes of the catchment. Unit hydrographs are an example of this method (Devia *et al.*, 2015). Conceptual models use simple mathematical equations to describe the main physical elements of the hydrological process

(i.e rainfall, infiltration, evapotranspiration, runoff, etc.). However, while some of the model parameters can be extracted from field data, in most cases, calibration is needed to define an optimal parameter set. Calibration requires large amounts of hydrological and meteorological data (Wheater, 2002). The physical models use more advanced equations such as mass conservation and momentum transfer equations to describe the hydrological processes of water movement (Aghakouchak & Habib, 2010). The parameters of such models have, in principle, a direct physical significance and can vary as a function of both time and space. They do not require an extensive amount of data for calibration, but the initial state of the model such as the soil moisture content and initial water depth, for example, as well as the morphology of the catchment are required (Devia *et al.*, 2015).

Model selection is dependent on data availability (dos Santos *et al.*, 2018). Data requirements are much less demanding lumped models compared to their distributed counterparts. It is often argued that the more complex model provides better results (Devia *et al.*, 2015). However, Wheater (2002) stated that "simple model structure does not reflect the complexity of the rainfall-runoff response and a complex model structure is not always supported by the available data. A balance between the complexity of the model and available information is crucial for successful model identification". In this context, several studies have shown that a well calibrated lumped conceptual model may perform similarly or better than the more complex distributed model (Ghavidelfar *et al.*, 2011; Lobligeois *et al.*, 2014; Sayama *et al.*, 2012). However, in some cases, both the simple and complex models perform badly (Ajami *et al.*, 2004). Vansteenkiste *et al.* (2014) used five hydrological models; three lumped and two distributed models, to assess the differences in the runoff and extremes simulations. The results showed that, overall, the lumped hydrological models performance is higher than the distributed models with much less computational time required. Distributed models have however the distinct advantage of being able to provide information within the catchment, such as intermediate streamflows.

1.5.2 Hydrological model calibration

In order for the hydrological model to accurately simulate streamflows, the model parameters have to be adjusted to the studied basin. This is generally needed for models of all complexity, from a simple lumped model to a very complex physical-distributed model (Moradkhani & Sorooshian, 2009). Generally, all hydrological models are dependent on the set of parameters that are used to control the model. Therefore, these parameters should be tuned precisely through a calibration process to match the observed basin runoff (Arsenault *et al.*, 2014). This calibration process could be either manual or automatic. Manual calibration depends on iterative trial and error adjustment. Consequently, it is a time-consuming task, even with a model with few parameters. When using a complex model, it is nearly impossible to get the best calibrated parameter set using the manual method. Automatic calibration algorithms are generally considered better and a much faster alternative (Moradkhani & Sorooshian, 2009). In automatic calibration, the set of parameters are adjusted automatically using mathematical and statistical methods; mainly optimization algorithms, according to a specific search scheme to minimize the difference between simulated and observed data. Typically, an objective function (also called error-measure function) is used to measure that difference.

1.5.3 Hydrological modeling Uncertainty

The typical way to evaluate the hydrological model uncertainty is to compare the performance of different models to represent the basin hydrology. Most studies showed that the arising uncertainty from using different hydrological models is relatively small compared to GCMs and GHGESs uncertainties (Vetter *et al.*, 2017). While others pointed out that the role of the hydrological model in generating a large amount of uncertainty should be considered (Bastola *et al.*, 2011; Asseng *et al.*, 2013).

The hydrological models uncertainty is mainly related to the model structure and model parameters (Koskela *et al.*, 2012; Ajami *et al.*, 2007; Liu & Gupta, 2007). Some studies showed that model parameters led to significant uncertainty (Wilby, 2005). Considerable concerns have

been given to the automatic calibration methods intending to find the best fitting parameters set. Finding the best single set of parameters is, however, unrealistic owing to the existence of many available objective functions that can be used in model calibrations. Consequently, bad model-parameter identification results in considerable uncertainty in model outputs (Moradkhani & Sorooshian, 2009).

Other studies have shown that important uncertainty is related to the model structure (Chen *et al.*, 2011b; Poulin *et al.*, 2011). Poulin *et al.* (2011) applied one lumped conceptual model and one spatially-distributed physically-based model, calibrated using multiple automatic calibration algorithms to analyse the effect of model parameters and model structure on the hydrological modeling uncertainty in climate change studies. The results showed the significant effect of hydrological model structure on the modeling uncertainty. Troin *et al.* (2018) used seven snow models, five potential evapotranspiration methods, and three hydrological model structures to evaluate the uncertainties linked to the hydrological model components over two catchments in Canada. Their results show that most of the contribution to uncertainty was related to the hydrological model structure. Brissette *et al.* (2020) stated that selecting an appropriate hydrological model may be even more crucial than climate model selection in certain cases. Therefore, quantifying and reducing uncertainties are an important but very challenging task in hydrologic modeling studies.

CHAPTER 2

STUDY REGIONS AND DATA

2.1 Study Regions

2.1.1 Geographic Situation

North America is the third largest continent with a surface area of about 21.34 million km^2. It is bounded by the Arctic Ocean in the north, South America in the south, and by the Atlantic and Pacific Oceans respectively on its east and west sides. North America can be divided into five main physiographic regions: 1) the mountainous west, where all of the largest mountain chains are located (e.g., the Rocky Mountains); 2) the Great Plains which are located in the middle of the continent; 3) the Canadian Shield that extends over eastern, central, and north-western Canada; 4) the eastern region, which includes the Appalachian Mountains; and 5) the Caribbean islands, which are normally considered as part of North-America.

Africa is the second largest and second most-populous continent in the world. It covers a land area of about 30.3 million km^2, including adjacent islands, which represents 6% of Earth's total surface area and 20.4% of its total land area (Mawere, 2017). Deserts and dry lands cover 60% of its entire surface (Prăvălie, 2016). The average elevation of Africa is almost 600 m above sea level, roughly close to the average elevations of North and South America (Atrax, 2016). Generally, higher-elevation areas lie to the east and south, while a progressive decrease in altitude towards the north and west is apparent.

The African continent can be divided into 25 major hydrological basins (Karamage *et al.*, 2018). Generally speaking, the main drainage for all of the continent's basins is towards the north and west, and ultimately, into the Atlantic Ocean. About 95% of its streams are drained through permanent rivers. In some arid areas (i.e., Northwest Sahara Desert), drainage is sometimes absent or masked by sand seas (Karamage *et al.*, 2018). Roughly, 60% of the African

continent is drained by 10 large rivers (Congo, Limpopo, Niger, Nile, Ogooue, Orange, Senegal, Shebelle, Volta and Zambezi) and their tributaries (Paul *et al.*, 2014).

2.1.2 Climate profiles

The North American continent has a wide variety of climate zones ranging from harsh winters to moderate summers (James Wreford Watson & Others, 2020). Canada's climate varies greatly across different geographical regions. The temperature ranges between -25°C in winter and 35°C in summer. The precipitation characteristics differ in both space and time. The snow falls almost all over Canada in winter, except on the west coast. The west coast is characterized by a temperate climate resulting from the warm air streams from the Pacific Ocean. The Canadian Prairies, extending from the Rocky Mountains to the Great Lakes, are characterized by cold snowy winters and dry summers. The Great Lakes see wet summers and snowy winters. Northern Canada is covered with snow for a large part of the year and temperatures rise above freezing only for a few months during summer.

In the United States, the climate varies considerably across the different regions. Generally, the climate in the southern and western regions is warmer than regions located in the northern and eastern regions. Hot summers and mild winters can be expected in the western and southern regions, while harsh winters with heavy snowfall are typical of the north-eastern regions (Faridi, 2016). Six main climate regions describe the climate in the USA: Northwest Pacific; Mid/-South Pacific; Midwest; Northeast; Southeast and Southwest. The Northwest Pacific is the wettest region with warm summer temperatures and scattered rainfall almost throughout the year. The Mid/South Pacific zone is characterized by nice winter weather with limited snowfall. The summers are generally dry with cooler weather in the northern part. The weather in the Midwest zone has reasonable summer temperatures and harsh winters with lots of snow. The Northeast has warm summers and very cold winters with heavy snowfall. The Southeast zone has good weather all around the year with moderate rainfalls and some snow during winter. Finally, the Southwest zone is considered the hottest region of the US. It is characterized by heavy rainfall accompanied with thunderstorms and tornadoes.

Africa is the hottest continent on earth, and is the area that has seen the highest ever-recorded land surface temperature (58°C in Libya; El Fadli *et al.* (2013)). The continent is characterized by highly variable climates that range from tropical to subarctic on its highest peaks. According to the Koppen climate classification (Köppen, 1900) as shown in Figure 2.1, the northern region, the Sahara Desert, is mainly classified as dry (red) whereas the central and western areas contain both savannah plains and dense forests with tropical and humid subtropical climates (blue) with a semi-arid climate in-between (El Fadli *et al.*, 2013). The southern region transitions toward semi-tropical (green) and semi-arid climates (orange and red). These wide climate ranges are characterized by a wide variety of precipitation extremes, including droughts and floods. Droughts occur mostly in the Sahel and in some parts of Southern Africa, whereas flooding is most prevalent in the southern and eastern regions.

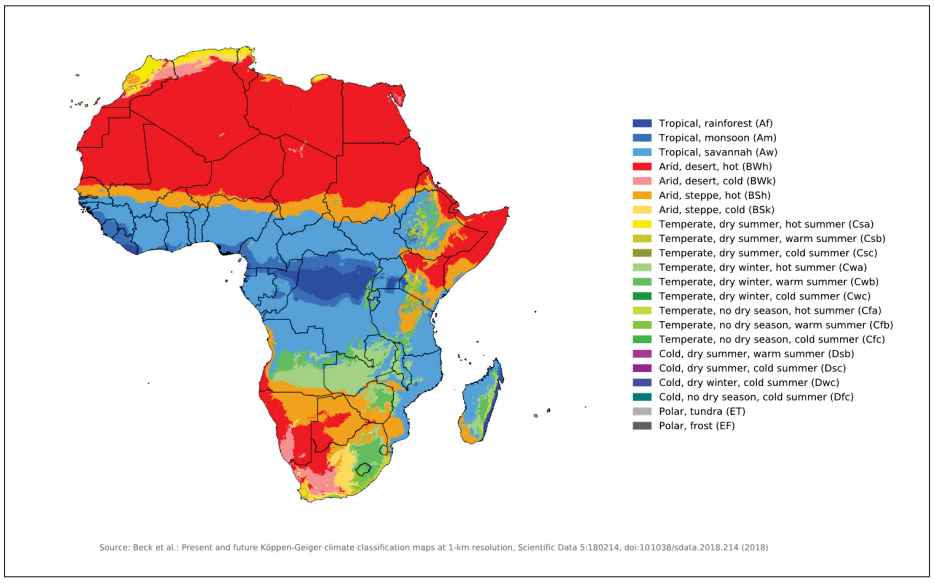


Figure 2.1 Koppen-Geiger climate zones classification map for Africa in the (1980-2016) period
Taken from (Beck *et al.*, 2018)

Precipitation varies widely across the African continent and spans some of the earth's driest and wettest climates. While most of the precipitation is liquid, some mountain ranges in North-Africa and, to a lesser extent in South-Africa experience regular snowfall. The three vol-

canic cones of Mount Kilimanjaro also experience frequent snowfall and hosts several glaciers, which are however rapidly shrinking. Rainfall patterns are dominated by the intertropical convergence zone (ITZ) which generate intense precipitation along the Equatorial band. This band experiences intense thunderstorms and convective activity. The movement of the ITZ dominates the seasonality of precipitation with well-defined wet and dry seasons along Central African countries. The West coast of Africa experiences some relatively small hurricane activity during the months of August and September. The most destructive hurricanes that travel across the Atlantic Ocean are typically formed near the Cape Verde islands.

2.1.3 Climate change in Africa

The African ecosystems are already being affected by the consistent anthropogenic climate change with evidence of warming over land regions. Over the past century, minimum temperatures have increased more rapidly than maximum temperatures. It is very likely that over most of Africa, the mean annual temperature has increased by 0.5°C or more over the past century (IPCC, 2001). Moreover, average temperatures in Africa are expected to rise faster than the global increase during the 21^{st} century (IPCC, 2001; James & Washington, 2013). Warming projections under medium greenhouse gases emission scenarios indicate an increase in the mean annual temperature exceeding 2°C by the end of this century and between 3 °C and 6 °C under a high Representative Concentration Pathway (RCP) (IPCC, 2014).

Precipitation projections are more uncertain than temperature projections with higher spatial and seasonal dependence (Orlowsky & Seneviratne, 2012). Changes in mean seasonal precipitation coupled with temperature increases will result in significant impacts to river streamflow and reservoir levels that will amplify the existing water-stressed availability problems in Africa. Climate projections point to a likely reduction in precipitation over the Mediterranean region of northern Africa as well as the western parts of Africa by the end of the 21^{st} century (Giorgi & Lionello, 2008; Shongwe *et al.*, 2009; Kotsias *et al.*, 2020). On the contrary, expected increases in mean and extreme rainfall by the end of the century for regions of high or complex topography such as in western regions (Akinsanola & Zhou, 2019; Dinku *et al.*, 2007),

and parts of eastern and southern Africa (IPCC, 2001), likely resulting in additional flooding risk. In western Africa, the impact of climate change on water resources is restricted by the significant climate model uncertainties. Some studies expect an increase in precipitation, while others predict a likely decreasing trend in precipitation by the end of the century (IPCC, 2014).

Several studies point to the future decrease in water abundance. Abouabdillah *et al.* (2010) estimated that higher temperatures and declining rainfall would reduce the water resources in Tunisia. Droogers *et al.* (2012) estimated that 78% of future water stress could be attributed to socioeconomic factors, while climate change will account for the other 22% of future water shortage in Northern Africa by 2050. In eastern Africa, potential climate change impacts on the Nile Basin are of concern for its socioeconomic importance. Kingston & Taylor (2010) showed in their studies an initial increase between (2010–2039), followed by a decline in surface water discharge in the Blue Nile by late century as a result of both declining precipitation and increased evaporating demand.

2.2 Data

2.2.1 Precipitation and temperature datasets

As discussed earlier, there is now a rather large number of gridded datasets that have been stockpiled from stations, satellites, reanalysis or a combination thereof. However, not all these datasets can be used for climate change impact studies. Dataset selection should be driven by the availability of high spatial and long temporal resolutions. The chosen datasets have to meet the minimum requirement of having at least 30-years of data series, as defined by the World Meteorological Organization (WMO) for climatological studies (Burroughs & Burroughs, 2003), covering the same time period, so they can show a high-frequency internal variability as well as they could potentially be used as reference datasets in climate change impact studies. To be useful in this study, a dataset must have the following characteristics: 1) spatial resolution (finer than 1° for example to be used in local hydrological studies); 2) daily scale or finer temporal resolution; 3) long temporal coverage (at least 30 years to establish

robust statistics for downscaling and bias correction); and 4) for an inter-comparison study, all datasets should cover approximately the same time period.

Table 2.1 The selected global gridded datasets

No.	Short Name	Data Source	Spatial Resolution	Spatial Coverage	Temporal Resolution	Temporal Coverage
1- Precipitation						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	GPCC	Gauge	1.0°	Global	Daily	1982-2016
3	PERSIANN-CDR (V1R1)	Gauge, Satellite	0.25°	±60° Lat.	6 hourly	1983-2012
4	CHIRPS V2.0	Gauge, Satellite	0.05°	±50° Lat.	Daily	1981-Present
5	NCEP-CFSR	Reanalysis	0.5°	Global	6 hourly	1979-2012
6	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
7	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017
8	JRA-55	Reanalysis	0.5625°	Global	3 hourly	1959-Present
9	MSWEP V1.2	Gauge, Satellite and Reanalysis	0.25°	Global	3 hourly	1979-2015
2- Temperature						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
3	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017

According to above criteria, nine precipitation and three temperature gridded datasets are included in this study, and are presented in Table 2.1. The precipitation datasets are classified based on their respective data sources. Two datasets are based solely on gauge data: CPC Unified (Climate Prediction Center Unified Gauge) and GPCC V2018 (V2) (Global Precipitation

Climatology Center); two combine gauge and satellite data: CHIRPS V2.0 (Climate Hazards group Infrared Precipitation with Stations) and PERSIANN-CDR V1R1 (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) Climate Data Record (CDR)); four are derived from reanalysis: ERA5 (The European Centre for Medium-Range Weather Forecast's 5^{th} generation reanalysis), ERA-Interim (European Centre for Medium-range Weather Forecasts ReAnalysis Interim), JRA55 (Japanese 55-year ReAnalysis) and NCEP-CFSR (National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR), while the last is a multi-source dataset integrating gauge, satellite and reanalysis data (MSWEP V1.2 (Multi-Source Weighted-Ensemble Precipitation)). In terms of temperature, three datasets are included in this study: the gauge-based CPC unified dataset, the ERA-Interim and ERA5 reanalysis products. Properties of the temperature datasets are also provided in Table 2.1.

CPC Unified is a gauge-based dataset launched at the NOAA Climate Prediction Center (CPC) to provide daily precipitation data over the global land areas (Chen *et al.*, 2008). Gauged data was collected from the Global Telecommunication System (GTS), national and international agencies with more than 30,000 stations. The CPC Unified dataset provides rainfall data for the 1979-2016 period with a 0.5° x 0.5° spatial resolution. Data quality control was performed through direct comparisons against nearby stations, satellite records and numerical model forecasts.

GPCC V2018 (V2) is a global land-surface precipitation dataset that provides daily rainfall data for the 1982-2016 period on a regular grid with a spatial resolution of 1.0° x 1.0° based on in-situ rain gauge data (Schneider *et al.*, 2018). The data was provided by the national meteorological and hydrological services, global and regional data collections and the World Meteorological Organization (WMO) GTS data, and the GPCC's database incorporates data from more than 116,000 different stations.

PERSIANN-CDR is a near-global (60°N - 60°S) satellite-based precipitation dataset developed by the National Climatic Data Center (NCDC) of NOAA. It is generated using the archives of

historical GridSat-B1 infrared data to produce long-term rainfall estimates for the period 1983 to near present with a 0.25° x 0.25° spatial and 6-hours temporal resolutions (Ashouri *et al.*, 2015; Miao *et al.*, 2015). PERSIANN-CDR was bias-corrected using the Global Precipitation Climatology Project (GPCP) monthly 2.5° product (version 2.2), which includes the Global Precipitation Climatology Centre (GPCC) gauge information.

CHIRPS V2.0 is a quasi-global (50°N - 50°S) rainfall dataset that was released by the U.S. Geological Survey (USGS) and the Climate Hazards Center (CHC) scientists. It provides daily rainfall estimates for the period 1981 to near present with a high spatial resolution of 0.05° x 0.05° (Funk *et al.*, 2014). CHIRPS utilizes two global Thermal InfraRed (TIR) archives; the Globally Gridded Satellite (GriSat) archive of the National Climate Center (NOAA) for the period 1981 to 2008, and the NOAA Climate Prediction Center dataset (CPC TIR) for the period 2000 to present. The 2000-2008 overlap period was used to adjust the systematic bias although the slight differences found between both sources (Funk *et al.*, 2015b). Several public and private gauged observation archives are incorporated in the development of the dataset. The public data has been provided by GHCN daily, GHCN monthly, Global Summary of the Day (GSOD), World Meteorological Organization's Global Telecommunication System (GTS) and Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL). Additional observations come from national meteorological agencies in Mexico, Central America, South America, and sub-Saharan Africa (Funk *et al.*, 2015b).

The NCEP Climate Forecast System Reanalysis (CFSR) was designed to be a global and high resolution coupled atmosphere-ocean-land surface-sea ice system and could be used in climate studies. It was generated on a 6-hours temporal resolution and a spatial resolution of 0.5° x 0.5° (Saha *et al.*, 2010). The main historical data used in this product came from the National Center for Atmospheric Research (NCAR), the National Climatic Data Center (NCDC), and the National Climatic Data Center (NESDIS) archives.

ERA-I is a global atmospheric reanalysis that was released by the ECMWF in 2006 (Dee *et al.*, 2011) in replacement of ERA40. ERA-I introduced an advanced four-dimensional variational

(4D-var) analysis assimilation scheme with a 12-hour time step. It computes 60 vertical levels from the surface up to 0.1 hPa. Its horizontal resolution is 0.75° x 0.75° (approximately 80 km). Precipitation and temperature are available at a 12-hour time step and were aggregated to the daily scale in this work. The production of ERA-I was ceased in August 2019, thus providing temporal coverage from January 1979 until August 2019.

ERA5 is the fifth generation reanalysis from ECMWF. It provides several improvements compared to ERA-I, as detailed by Hersbach & Dee (2016). The analysis is produced at a 1-hourly time step using a significantly more advanced 4D-var assimilation scheme. Its horizontal resolution is 0.25° x 0.25° and it computes atmospheric variables at 139 pressure levels up to a height of 80 km. Data for the 1979–2018 period was released in March 2019, while the entire dataset from 1950 to present is still expected to be released in 2020. This work only looks at the 1979–2018 period because outputs of reanalysis prior to 1979 have been put into question due to the more limited availability of data to be assimilated, and notably from earth-observing satellites (e.g. Bengtsson *et al.* (2004)). While ERA5 may solve some of these problems, it is believed that a careful evaluation of inhomogeneity in ERA5 time series would be needed before using pre-1979 data.

JRA-55 is the second Japanese global atmospheric reanalysis that was conducted by the Japan Meteorological Agency (JMA) as of December 2009. JRA-55 covers the period starting from 1959 on a global basis with a spatial resolution of 0.5625° x 0.5625° and 3-hours time step (Kobayashi *et al.*, 2015). JRA-55 has solved the two major problems found in the previous product of JRA-25; which are the dry bias in the Amazon basin and the cold bias in the lower stratosphere. Observations used in ERA-40 (Uppala *et al.*, 2005) and those that were archived by the Japan Meteorological Agency (JMA) were the primary source of data. Snow depths for Russia were provided by the Russian Research Institute for Hydrometeorological Information (RIHMI), in the USA from the University Corporation for Atmospheric Research (UCAR) and in Mongolia from the Institute of Meteorology and Hydrology (IMH). A detailed list of data suppliers could be also found in (Kobayashi *et al.*, 2015).

MSWEP V1.1 is a global precipitation dataset designed mainly for hydrological modeling. The dataset provides data for the 1979–2015 period at a 0.25° x 0.25° spatial resolution and 3-hours temporal resolution (Beck *et al.*, 2017a). The long-term mean of MSWEP was provided from the Climate Hazards Group Precipitation Climatology (CHPclim V1.0) dataset (Funk *et al.*, 2015a). (CHPclim) is a global climatology database based on satellite data and corrected by gauge observations. The correction for gauge under-catch was performed using streamflow records from 13,762 stations across the globe. MSWEP incorporates inputs from seven datasets; two gauge-based products (CPC Unified and GPCC), three satellite-based (CMORPH, GSMaP-MVK, and TMPA 3B42RT), and two reanalysis (ERA-Interim and JRA-55). The reanalysis datasets performed better in mid to high latitudes as well as for snowfall estimates, while the satellite products performed better in the tropics. For each grid cell, the weight of gauge-based products was calculated based on the gauge network density, and for the satellite and reanalysis products the weight was assigned relative to their performance.

There are additional satellite products that provide global rainfall information at finer resolutions than PERSIANN-CDR, which has been selected in this work. Of particular interest is the Global Precipitation Measurement (GPM) mission, designed to further precipitation monitoring from an array of microwave sensors. It was launched to provide a new generation of precipitation datasets with an improved accurate measurement for light rainfall and snow precipitation as well as more frequent observations over the medium and high latitudes (Hou *et al.*, 2014). GPM utilizes passive microwave sensors in addition to the infrared measurements from geostationary satellites, providing rainfall monitoring around the globe with higher spatial and temporal resolutions than the previously widely used TMPA products (Yong *et al.*, 2015). These improvements are likely to provide significant advantages for hydrometeorological studies, weather forecasting, water budget studies and many other applications. In particular, the GPM Integrated Multisatellite Retrievals (IMERG), provides data at 0.1° and half-hourly spatial and temporal scales (Huffman *et al.*, 2015) and the Global Satellite Mapping of Precipitation (GSMaP) provides hourly rainfall data also at a 0.1° resolution (Okamoto *et al.*, 2005). These products have been evaluated against gauge measurements over different

regions (Aslami *et al.*, 2019; Asong *et al.*, 2017; Chen *et al.*, 2016; Lu & Yong, 2018; Mazzoglio *et al.*, 2019) and showed satisfactory results. However, these state-of-the-art products do not cover a long-enough time period to be used for the evaluation of a climatic base-line period required for climate change impact studies. They were therefore not chosen for this study.

2.2.2 Observed hydrometeorological data

2.2.2.1 GRDC database (Africa)

The Global Runoff Data Centre (GRDC) archive is arguably the most complete global discharge database providing free access to river discharge data (Fekete & Vörösmarty, 2007). The database provides streamflow records collected from 9213 stations across the globe, with an average temporal coverage of 42 years per station (Do *et al.*, 2017). It is operated under the World Meteorological Organization (WMO) umbrella to provide broad hydrological data to support the scientific research community. GRDC data has been widely used in various hydrological studies, such as those examining hydrological model calibrations (Milliman *et al.*, 2008; Hunger & Döll, 2008; Donnelly *et al.*, 2010; Haddeland *et al.*, 2011), or as a benchmark to compare simulated streamflows (Trambauer *et al.*, 2013; Zhao *et al.*, 2017).

2.2.2.2 NAC^{2}H database (North-America)

The observed data (OBS) in this study was taken from the North American Climate Change and hydroclimatology (NAC^{2}H) database (Arsenault *et al.*, 2020), which is a hydrology and climate change impact dataset developed to study the impacts of different components of the modelling chain on hydrological indices over a collection of 3540 North American catchments. It includes hydrometeorological data such as precipitation values (mm/d) on a daily time step, maximum and minimum temperature (°C) and streamflow at the daily scale for each of the catchments. Climate data for the 698 catchments in Canada were taken from the CANOPEX database (Arsenault *et al.*, 2016), whereas the data for the 2842 United States catchments were collected from the United-States Geological Survey's (USGS) National Water Informa-

tion System (NWIS) (US Geological Survey, 2016) and National Hydrography Dataset (U.S Geological Survey, 2019).

For Canada, the data is sourced from Environment and Climate Change Canada (ECCC) that were post-processed and basin-averaged using Thiessen Polygon weighting. Any missing values were replaced by the NRCan interpolated climate data product (Hutchinson *et al.*, 2009). While the Canadian streamflow data is provided by Water Survey Canada, the hydrometric branch of ECCC. For the United States, NAC^{2}H uses the Livneh gridded dataset for meteorological data (Livneh *et al.*, 2015), whereas streamflows are provided by the United States Geological Survey (USGS) National Water Information Service. The data cover the period 1951–2010 for Canada and 1950-2014 for the United States. NAC^{2}H data is open source and available on the Open Science Foundation data repository at the following website: https://osf.io/s97cd/. More details can be found in (Arsenault *et al.*, 2020).

2.2.3 General Circulation Models (GCMs)

All GCMs used in this study were part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor *et al.* (2012)). Long historical climate simulations (1850–2005) and future climate projections (up to 2100 and beyond) for 4 Representative Concentration Pathways (RCPs) are included in the CMIP5 database.

Ten CMIP5 GCMs from 10 different modeling centers were selected for this study, as shown in Table 2.2. They were selected as a subset of the GCMs used to set up the NAC^{2}H database (Arsenault *et al.*, 2020). The number of GCMs (10) was selected as a compromise between having an accurate representation of GCM climate sensitivity variability and keeping the large computational burden of this project reasonable. All GCM data was extracted over the 1983-2012 and 2071-2100 future periods under the (RCP8.5) emission scenario.

Table 2.2 List of chosen GCMs, research centres and spatial resolutions

No.	Models	Research Center	Spatial Resolution
1	BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration, China	2.79° x 2.81°
2	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China	2.79° x 2.81°
3	CanESM2	Canadian Center for Climate Modeling and Analysis, Canada	2.79° x 2.81°
4	CCSM4	National Center of Atmospheric Research, USA	0.94° x 1.25°
5	CMCC-CESM	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	3.44° x 3.75°
6	CNRM-CM5	National Center of Meteorological Research, France	1.40° x 1.40°
7	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, China	2.79° x 2.81°
8	INMCM4	Institute for Numerical Mathematics, Russia	1.5° x 2.0°
9	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	1.40° x 1.40°
10	MRI-CGCM3	Meteorological Research Institute, Japan	1.12° x 1.125°

2.2.4 African watersheds boundaries data

HydroSHEDS (the Hydrological data and maps based on the SHuttle Elevation Derivatives at multiple Scales database) is a freely available global archive, developed through a World Wildlife Fund (WWF) program, that uses a hydrologically-corrected digital elevation model to provide hydrographic information for regional and global studies (Lehner *et al.*, 2008). In addition, it applies a consistent methodology using Geographic Information System (GIS) tech-

nology to provide watershed polygons for more than 7000 GRDC gauging stations. Figure 2.2 shows watershed polygon layers at different spatial scales for the African continent. HydroSHEDS consists of four main databases; 1) HydroRIVERS, 2) HydroLAKES, 3) HydroATLAS and 4) HydroBASINS.

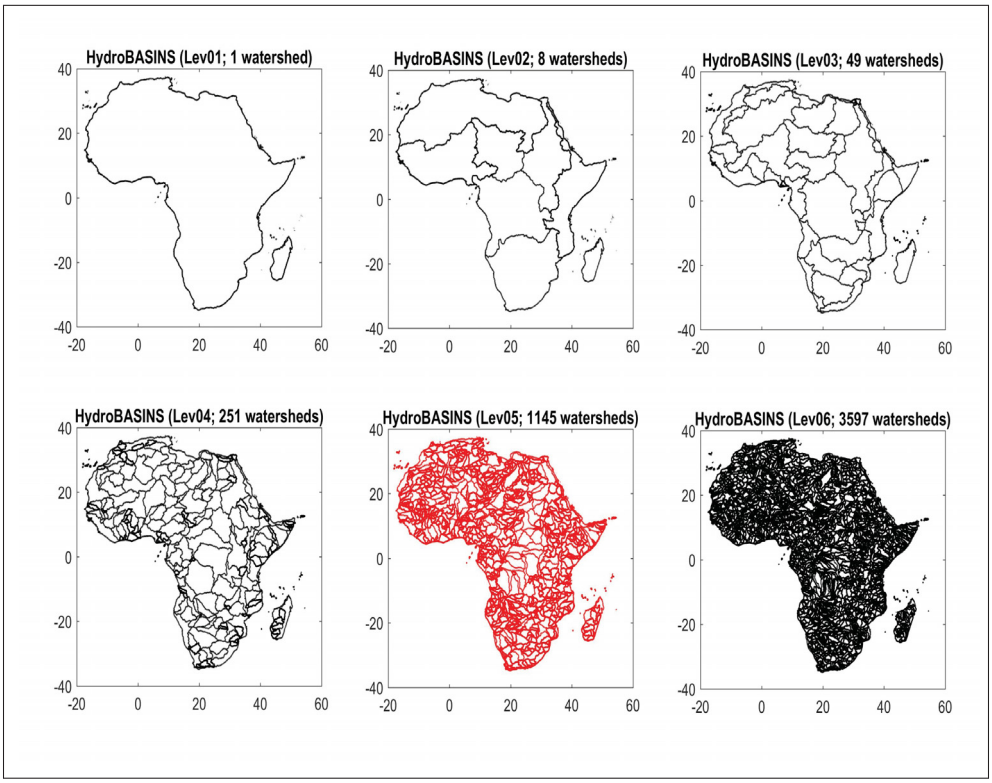


Figure 2.2 Sample of the different vector layers of watersheds on the African continent. Each layer has a different number of watersheds, depending on the required scale

HydroRIVERS database provides a global rivers network data in a vector format for all rivers that have either a basin area of at least 10 km^2 or/and an average flow of at least 0.1 m^3/sec. HydroLAKES (Messager *et al.*, 2016) is another database that offers polygon shapes of the 1.4 million global lakes that have, at least, a surface area of 10 ha. HydroATLAS (Linke *et al.*, 2019) is a comprehensive database presenting extensive series of hydro-environmental information such as hydrology, physiography, climate, land cover and use, soils and geology, and anthropogenic influences for all global watersheds and rivers. Finally, HydroBASINS

(Lehner & Grill, 2013) database offers various data layers such as void-filled elevations, drainage directions, flow accumulations, streamflow networks and basin boundaries, in raster and vector formats, to support hydrological modeling and watershed studies. These data are available with different spatial resolution ranges from 3 arc-second (approximately 90 m at the equator) to 5 minute (approximately 10 km at the equator) with a quasi-global extent between $\pm 60^{\circ}$ latitude (Lehner, 2014). The vector layer (lev05), which consists of 1145 watersheds, was chosen to be used in this study.

CHAPTER 3

METHODOLOGY

The methodology structure is divided into three main parts which follow the secondary objectives of the study. All of the chosen datasets are first explored and evaluated over North America, where there is a good network of weather stations and reference datasets to assess their advantages and limitations. The nine global/near-global precipitation and three global temperature datasets are compared over 3138 catchments relative to a reference dataset; in this case, the NAC^{2}H observation dataset. Then to further assess dataset performance, the nine precipitation and 3 temperature datasets under evaluation, in addition to the two reference datasets from NAC^{2}H are combined in their 40 possible arrangements and used as inputs to the two lumped conceptual hydrological models; GR4JCN and HMETS.

A second evaluation of these gridded precipitation and temperature datasets is then performed over Africa, which displays a very sparse network of weather stations. As is the case over North America, all gridded datasets combinations are then used to drive the two hydrological models over 850 African catchments. These catchments were chosen based on the availability of at least 5 consecutive years of GRDC streamflow records during the study period. Since there is no reference dataset over Africa, the analysis is based on an intercomparison of all dataset combinations.

Not all these 850 catchments were compatible with the HydroSHEDS database. However, the reason for using these catchments in that part of the study was to run the hydrological models over the largest number of available catchments. These 850 stations from the GRDC database were re-filtered again by choosing only the stations that were compatible with the selected HydroSHEDS layer (e.g., lev05). As a result, the 1145 African catchments were divided into 350 gauged (which correspond to GRDC streamflow data) and 795 ungauged (no GRDC hydrometric station) catchments such that the gauged catchments were used to predict the streamflow at the ungauged sites in an additional regionalization step."

In a last step, a large-sample hydrological climate change impact study is performed over 1145 African catchments. It uses the standard top-down approach in a modeling chain, which consists of 10 GCMs, 2 hydrological models, 2 temperature and 9 precipitation datasets, for a total of 360 possible combinations. A single GHGES (RCP8.5), a single climate projection for each GCM and a single downscaling method (see below) are used, since the focus of this work is not on conducting a complete uncertainty chain study. All precipitation and temperature datasets combinations were used as reference datasets to bias-correct the climate projections. Both hydrological models were calibrated on all catchments for all 18 combinations of reference datasets (2 temperature datasets x 9 precipitation datasets), for a total of 41,220 independent hydrological model calibrations. The climate projections from each combination were used to feed the hydrological model using the relevant set of parameters obtained during the hydrological model calibration over the same reference period employed in the bias-correction step. For each catchment, 360 30-year streamflow time series are generated for both the reference (1983-2012) and future (2071-2100) time periods. Fifty-one streamflow metrics are computed for each of these time series. An n-dimensional analysis of variance is performed to partition the uncertainty linked to the four groups of components of the uncertainty modeling chain. The uncertainty related to the reference dataset will therefore be compared to that of the climate model ensemble and against that of both hydrological models. Figure 3.1 describes the main methodological steps for this work.

3.1 Intercomparison of gridded climate products and statistics

The study regions are composed of 3138 North American and 850 African catchments. Figure 3.2 presents the geographic distribution of these catchments. The comparison in North America was done at the catchment scale since the available catchments almost cover the entirety of the continent. In Africa, the first part of the analysis: datasets inter-comparison, was done at the grid scale since the spatial distribution of gauged catchments is more heterogeneous than over North-America. The second part of the analysis: the hydrological modeling evaluation, was conducted at the catchment scale.

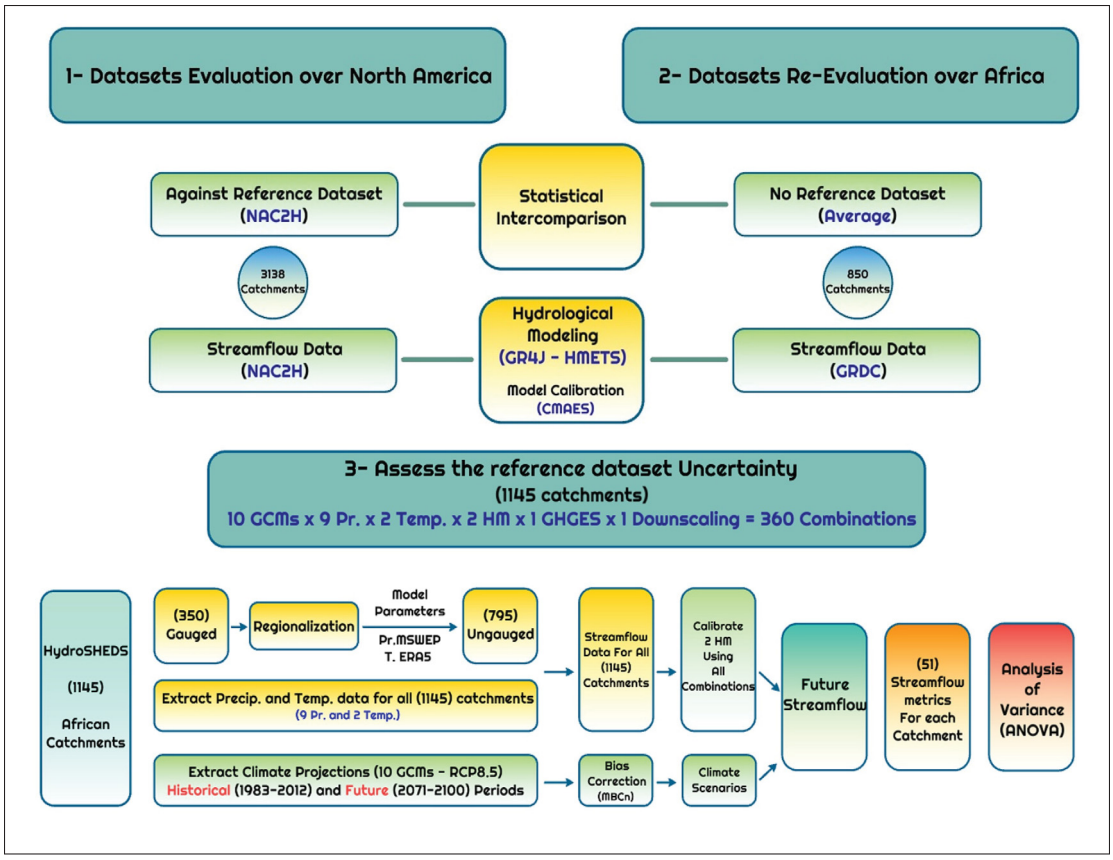


Figure 3.1 Overview of the various methodological steps implemented in this study

3.1.1 Datasets evaluation over North America

While the NAC^{2}H database was considered as the reference dataset, there was no underlying assumption that it is of higher quality or more accurate than any of the gridded products. Rather, it simply served as a baseline against which the other data products were compared. Analyses were performed by comparing annual and seasonal means of the gridded climate variables to the reference dataset. This allowed finding spatial patterns of differences in average precipitation and temperatures to obtain an evaluation of the regional differences between the products. A similar analysis was then performed to investigate the differences in variability within these datasets on a daily time step. This also allowed evaluating the properties at a time scale that is more difficult to manage for gridded datasets compared to aggregated values at the annual or seasonal scales. The tests were performed because gridded datasets presenting

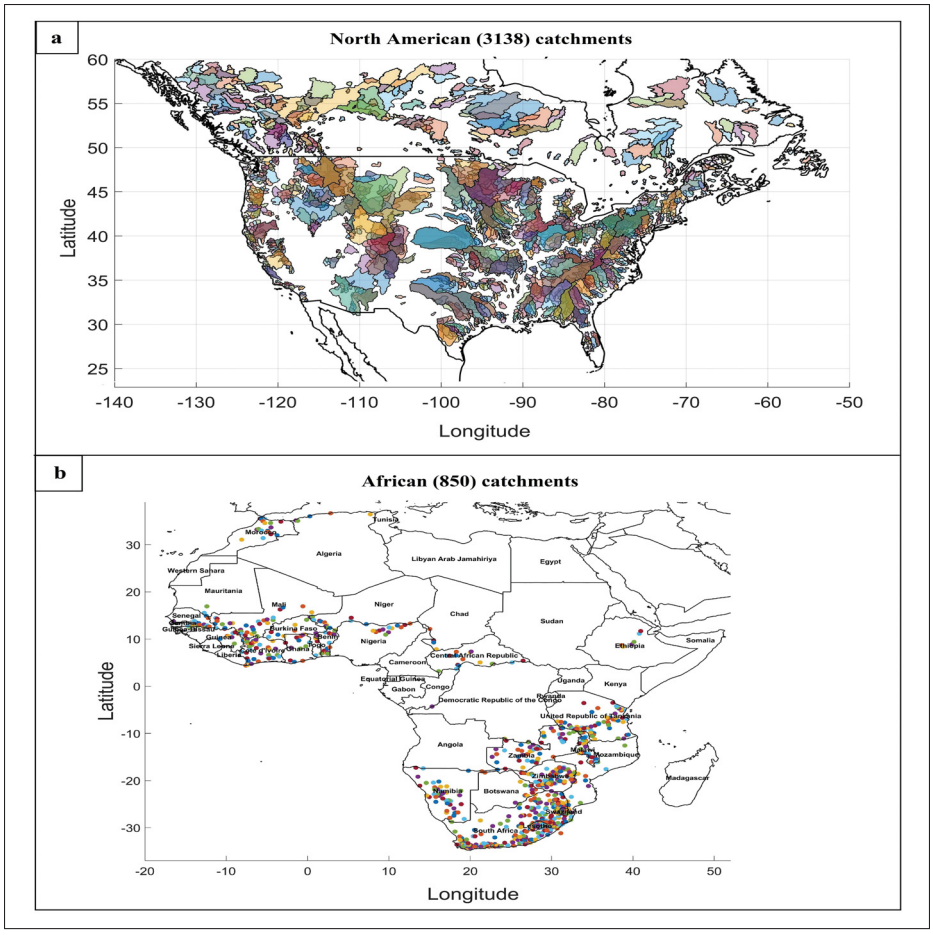


Figure 3.2 Spatial distribution of: a) the 3138 North American watersheds and b) the 850 African watersheds (each dot represents the watershed centroid)

little to no precipitation difference at the annual scale could still be largely underestimating the variance found in observational records. The Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and correlation coefficient (r) were used to compare the annual and seasonal precipitation, and temperature values to the reference dataset (the article is presented in APPENDIX II). The main characteristics of North American basins were extracted from the NAC^{2}H database as well as the main characteristics of the African basins were extracted from the GRDC database and are presented in Table 3.1.

Table 3.1 Main characteristics of the study basins

Basin attribute	Minimum	Maximum	Median	Mean
North America				
Elevation (m)	7.3	3585	380	692
Area (km^2)	302	179150	1803	6317
Mean annual precipitation (mm/year)	307	3895	993	981
Mean annual discharge (m^3/sec.)	0.048	1584	17.60	56.75
Temporal coverage (years)	5	30	30	27.60
Africa				
Elevation (m)	22	2897	598	803
Area (km^2)	1.18	$3.65x10^6$	16268	40471
Mean annual discharge (m^3/sec.)	0.0014	2955	11.47	208.86
Temporal coverage (years)	5	30	30	25.70

3.1.2 Datasets evaluation over Africa

Following the North-American evaluation, the chosen datasets were re-evaluated over Africa (see APPENDIX III). For the temperature datasets, Tarek *et al.* (2019) showed significant relative improvement of ERA5 with respect to its predecessor ERA-Interim (The article is presented in APPENDIX I). ERA5 systematically reduced biases present in ERA-Interim for the temperature variables. Therefore, the ERA-I temperature dataset was excluded and only two temperature datasets were kept: ERA5 and CPC. The same nine precipitation datasets and hydrological models were included.

3.2 Evaluation using hydrological modeling

The quality and performance of the climate variable datasets were evaluated indirectly through an independent measure, namely, the watershed observed streamflow. The hypothesis posed here is that climate datasets that allow for more accurate hydrological modelling with respect to the observed streamflow should be considered as being of higher quality. Of course, the choice of a hydrological model does influence performance. However, this should be seen as a first attempt at finding inconsistencies within the climate datasets. Hydrological modelling is sensitive to the annual cycle of precipitation and temperature, as well as to the coherency

between both variables, so it can therefore be seen as a good evaluator of dataset overall quality. This approach has been used in several other studies (Beck *et al.*, 2017b; Essou *et al.*, 2016a; Tarek *et al.*, 2019).

In the course of this study, two lumped hydrological models were implemented and calibrated over each of the available catchments because the large-scale nature of this study precluded the widespread implementation of distributed models. The two hydrological models selected to evaluate the performance of the various climate datasets, GR4JCN and HMETS, are flexible and adaptable and have been shown to perform well in a wide range of climates and hydrological regimes (Arsenault *et al.*, 2015, 2018; Martel *et al.*, 2017; Valéry *et al.*, 2014; Perrin *et al.*, 2003).

3.2.1 The GR4J hydrological model

The GR4J hydrological model (Perrin *et al.*, 2003) is a lumped and conceptual model that is based on a cascading reservoir production and routing scheme. Water is routed from these reservoirs to the outlet in parameterized unit hydrographs. While the original GR4J model includes four calibration parameters, the version used in this study had six calibration parameters in order to include a snow-accumulation and snowmelt routine, namely CEMANEIGE (Valéry *et al.*, 2014). This GR4J-CEMANEIGE (GR4JCN) combination has shown excellent results in studies across the globe (Raimonet *et al.*, 2017, 2018; Youssef *et al.*, 2018; Riboust *et al.*, 2019; Wang *et al.*, 2019), including in Canada and the United States. It requires daily precipitation, temperature and potential evapotranspiration (PET) as inputs. The PET was computed in the present study using the Oudin formulation (Oudin *et al.*, 2005) as it was shown to be simple yet efficient when used in GR4JCN. Furthermore, the choice of PET is more sensitive than in other simple hydrological models because GR4J does not scale the input PET to adjust its overall mass balance. Instead, a parameter is included that allows exchanges between underground reservoirs of neighbouring catchments.

3.2.2 The HMETS hydrological model

HMETS (Hydrological Model - École de technologie supérieure) is a lumped and conceptual model used in many research applications and as a component of operational multi-model hydrological studies and forecasting (Martel *et al.*, 2017). It was selected due to its good performance in the study domain in previous studies and because a lumped model was required to simulate discharge over the large number of catchments in the study.

The HMETS hydrological model (Martel *et al.*, 2017) is more complex than GR4JCN, and as such has more calibration parameters (21). The HMETS hydrological model has four reservoirs instead of two (surface runoff, hypodermic flow from the vadose zone reservoir, delayed runoff from infiltration and groundwater flow from the phreatic zone reservoir), allowing for finer adjustments to the runoff and routing schemes. Its snowmelt module requires 10 of the 21 parameters and was selected specifically to be more robust in Nordic catchments with specific routines for snow accounting, snowmelt, snowpack refreezing, ice formation and soil freezing and thawing. It requires daily maximum and minimum temperature as well as daily rainfall and snowfall amounts. All the chosen temperature datasets provide daily minimum and maximum temperature, with the exception of ERA5, which provides mean hourly air temperature. The minimum and maximum hourly temperatures for each day were therefore used as being representative of the daily maximum and minimum values. For ERA5, the mean daily temperature was computed as the average of the 24 hourly temperature values.

HMETS starts by computing the potential evapotranspiration using the Oudin formulation, which is scaled through a calibration parameter, and then computes snow accumulation and melt with a 10-parameter degree-day-based snow module developed by Vehviläinen (1992). Rainfall is then added to the runoff generated by snowmelt to obtain total water production. Potential evapotranspiration is then subtracted from the total water production to obtain the final runoff depths. The water then infiltrates into one of three underground soil layers modelled as reservoirs (the aquifer, the vadose zone and the delayed surface runoff zone), using six calibrated parameters. Some of the water is also kept above the soil as the surface runoff

reservoir. Water from these reservoirs is routed using two independent 2-parameter gamma distribution unit hydrographs for the surface unit hydrograph and the delayed unit hydrograph, respectively.

As will be detailed in the following section, the precipitation and temperature datasets were combined in their all possible arrangements for analysis purposes. It follows that the sheer number of calibrations to be performed in this study required implementation of automatic model parameter calibration methods.

3.2.3 Hydrological model calibration

For this study, the automatic Covariance Matrix Adaptation Evolution Strategy (CMAES) optimization algorithm was implemented because of its flexibility (Hansen *et al.*, 2003). Indeed, it performs well for small and large parameter spaces such as the 6-parameter and 21-parameter spaces in this study. It was also shown to be robust and is considered to be one of the best auto-calibration algorithms for hydrological modelling (Arsenault *et al.*, 2014).

For the African catchments, streamflow records from the GRDC database were used to calibrate the hydrological models and to evaluate the hydrological modeling performance. The GRDC database contains streamflow data from 1150 African stations. Only 850 stations were chosen based on two criteria. First, stations should have data during the 1983-2012 study period. Second, stations with less than five consecutive years of data during this period were excluded. For North American 3138 catchments, the streamflow records were extracted from the NAC^2H database and used in the hydrological model calibration.

The hydrological model parameters were calibrated on the entire available record of data for each catchment, foregoing the usual model validation step. This method was chosen for two reasons. First, calibrating on all years ensures that the maximum amount of information from the climate data is present in the parameter set and thus that there is no added uncertainty from choosing calibration and validation years. Second, Arsenault *et al.* (2018) have shown that the

model performance is statistically better when more years are added to the dataset and that validation and calibration skills are not necessarily correlated.

Both precipitation and temperature datasets were averaged at the catchment scale before being fed to the hydrological model. Each catchment was calibrated by letting CMAES converge over 15,000 model evaluations and repeating the process twice. The calibration was performed on the entire length of the available data as recommended in (Arsenault *et al.*, 2018). The best calibration score from the two generated sets was used to reduce the chance of considering a parameter set that had not properly converged to an acceptable minimum. This calibration procedure was repeated for each combination of precipitation and temperature datasets, including the reference datasets, for each watershed, in order to allow their objective comparison (Essou *et al.*, 2016a).

The objective function used to calibrate the parameters was the Kling-Gupta Efficiency (KGE) metric, which was introduced by Gupta *et al.* (2009) and modified by Kling *et al.* (2012), and which is an equally-weighted bias, variance and correlation aggregate metrics. KGE corrects the fact that the Nash-Sutcliffe Efficiency (Nash & Sutcliffe, 1970) metric (NSE) underestimates variability in the goodness-of-fit function. It is defined as a combination of three elements:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{3.1}$$

Where r is the correlation component represented by Pearson's correlation coefficient, β is the bias component represented by the ratio of estimated and observed means, and γ is the variability component represented by the ratio of the estimated and observed coefficients of variation. The KGE values theoretically range from negative infinity, implying an extremely poor performance of the model, all the way to one, suggesting a perfect performance. The performance ratings used in this study are defined based on the work of Gutenson *et al.* (2019) and Pech-

livanidis & Arheimer (2015) who divided the KGE values into three modeling-performance groups: Poor (KGE < 0.4), Acceptable ($0.4 \leq$ KGE < 0.7) and Good (KGE ≥ 0.7).

3.3 Regionalization

The 350 gauged catchments were used to predict the streamflow at the 795 ungauged sites. The transfer of hydrological information (i.e., model parameters or streamflow) from one catchment (gauged) to another (ungauged) is known as "regionalization" (Razavi & Coulibaly, 2013). Regionalization can be conducted using two methods: 1) rainfall-runoff models/model-dependent method, which typically transfers the model parameters from one or more gauged watersheds to an ungauged watershed, and 2) hydrological model-independent methods, which transfer the streamflow directly from gauged to ungauged watersheds. In this study, the model-dependent method was applied as it has been used in several studies and has shown acceptable results (Merz & Blöschl, 2004; McIntyre *et al.*, 2005; Boughton & Chiew, 2007; Cutore *et al.*, 2007; Samaniego *et al.*, 2010; Beck *et al.*, 2016; Arsenault & Brissette, 2014b; Saadi *et al.*, 2019).

The three most prominent approaches, namely, the spatial proximity (S.P), physical similarity (P.S) and multi-linear regression (MLR) methods (Oudin *et al.*, 2008), have been used with varying degrees of success to estimate the hydrological model parameters in ungauged catchments. The spatial proximity approach transfers the model parameters based on the spatial distance (geographic location) between the gauged and the ungauged basins. The physical similarity approach depends on the similarity of the physiographic attributes (e.g. area, elevation, slope, soil type, land cover, etc.) between the gauged and the ungauged basins. Finally, the multi-linear regression approach estimates a linear relationship between model parameters and catchment attributes on the gauged catchments, which can then be used to estimate via regression each of the hydrological model parameters at the ungauged site.

First, the climatological data from 9 precipitation and 2 temperature datasets were extracted for each of the 1145 catchments. Then, the three approaches were tested on the 350 gauged catchments to find the best method to apply, using a leave-one-out cross-validation frame-

work. Finally, the best-performing precipitation-temperature datasets combination based on the streamflow simulations (e.g., MSWEP precipitation and ERA5 temperature datasets) were used to feed the hydrological models and simulate the streamflow at the ungauged locations. Based on the hydrological modeling performance on the 350 gauged catchments, the MSWEP precipitation and ERA5 temperature datasets were found to be the best combination used in computing the streamflow for the 795 ungauged catchments.

3.4 Bias correction

In this study, the N-dimensional multivariate bias correction algorithm (MBCn) was used (Cannon, 2018). MBCn is an image processing technique extension that preserves the change between the historical and projected periods for all quantiles of the distribution. The algorithm consists of three main steps: (1) application of an orthogonal rotation to both model and observational data; (2) correction of the marginal distributions of the rotated model data using quantile mapping, and (3) application of an inverse rotation to the results. These three steps are repeated until the model distribution matches the observational distribution. MBCn is arguably the best-performing quantile-based method available (Adeyeri *et al.*, 2020; Meyer *et al.*, 2019).

3.5 Variance analysis

An n-dimensional analysis of variance theory (ANOVA) was used to quantify the contribution of the different uncertainty sources to the overall uncertainty (Von Storch & Zwiers, 2001). This method has been applied in many previous studies for this purpose (Addor *et al.*, 2014; Bosshard *et al.*, 2013; Trudel *et al.*, 2017). The ANOVA considers the interactions between the different sources of uncertainty. Without considering these interactions, the uncertainty of each individual source is overestimated. Therefore, testing all possible combinations/interactions of the different uncertainty sources is crucial. However, one limitation of this method is the difficulty to interpret how these interactions contribute to the overall uncertainty. Another disadvantage is that it requires a normal (Gaussian) distribution of the projected values to provide good results. Consequently, the results are vulnerable to outliers (Kim *et al.*, 2019).

Analysis of variance was performed for the 51 streamflow metrics defined in Arsenault *et al.* (2020) for each of the 1145 catchments (the article is presented in APPENDIX IV). These metrics cover a wide range of streamflow conditions: mean annual, seasonal and monthly values, distribution quantiles, as well as low- and high-flow extreme metrics. A variance was attributed to each of the four groups under study, namely, GCM, precipitation dataset, temperature dataset and hydrological model. A total of 11 variance components were computed: 4 main effect components, 6 first-order, 3 second-order, and 1 third-order interaction components.

CHAPTER 4

RESULTS

4.1 Evaluating the gridded datasets in North America

The first part of the study was to compare precipitation and temperature values over North American catchments. Precipitation and temperature was first averaged at the catchment scale in order to preserve the consistency between the climate data and the hydrological modelling results presented later in this chapter.

4.1.1 Analysis of precipitation and temperature in North America

The results in Figure 4.1 shows the mean annual temperature of the reference dataset (upper left) and differences between each of the three chosen temperature datasets. The term $difference$ is used below, instead of *bias*, since our reference dataset is not a true representation of the population, and is not inherently better than other datasets. On average, the three datasets are warmer than the observations, with ERA-I being the warmest. The warm difference is particularly clear in the western United States. Overall, it can be seen that ERA5 is the closest to observations, with small differences across central and eastern North America, and reduced differences on the West Coast.

The differences between ERA-I and ERA5 temperature datasets were studied in more deep details (see APPENDIX I). Figure 4.2 shows the mean annual temperatures for the observations and the ERA5 and ERA-Interim reanalysis products for the catchments in this study (top row). It also shows the mean absolute differences between the datasets for the winter (centre row) and summer seasons (bottom row). It can be seen that ERA5 sees a strong reduction in biases compared to those in the ERA-Interim dataset. The western coast of North America clearly still shows some important biases of up to 3 °C in summer and -2 °C in winter, although for most catchments the bias amplitude is smaller.

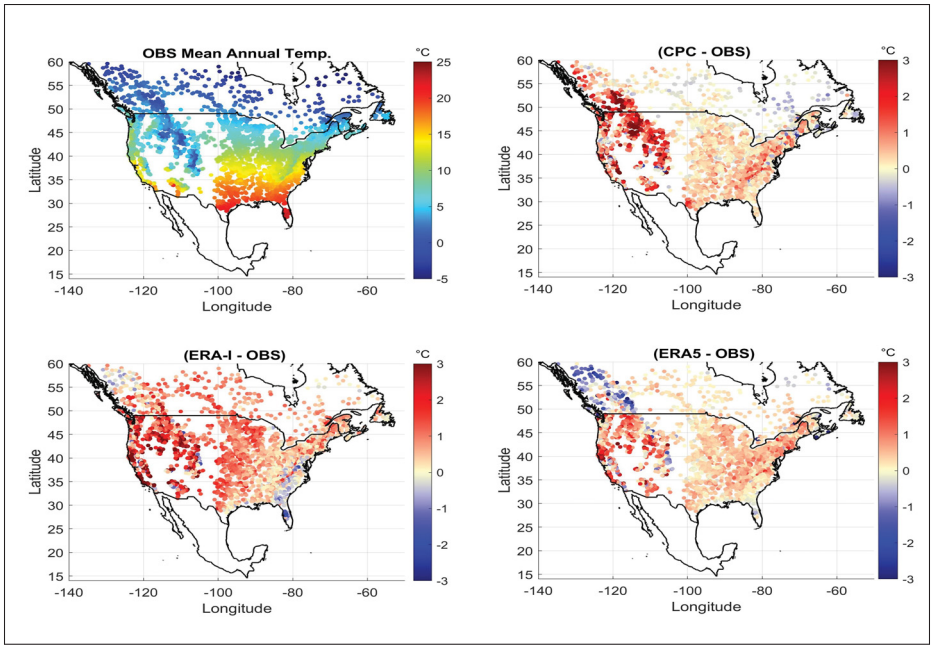


Figure 4.1 Difference maps for the mean annual temperature (dataset-OBS). The (top-left) plot shows the observed mean annual temperature extracted from the NAC^{2}H database

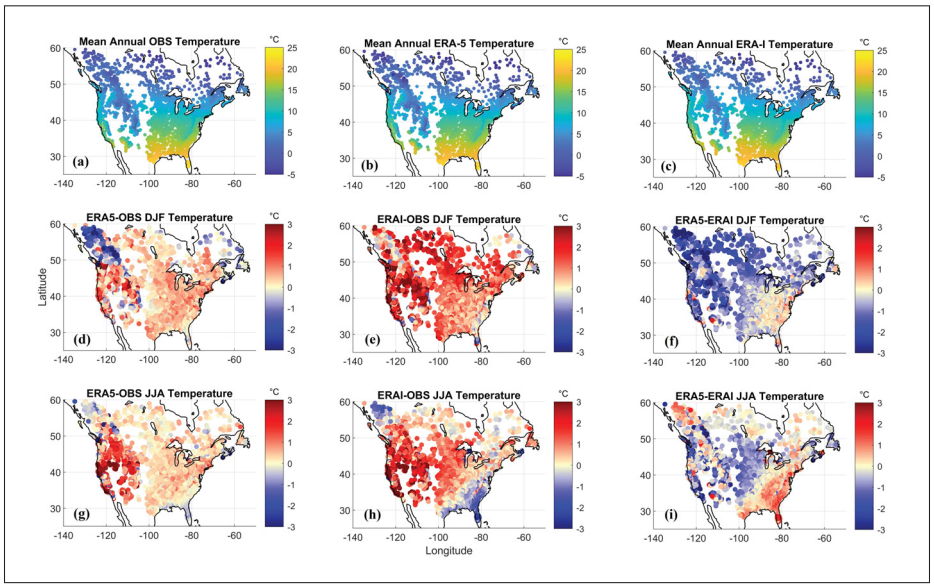


Figure 4.2 Mean annual temperature for the three datasets (a, b, c), winter (d, e, f) and summer (g, h, i) differences. All values are in degrees Celsius

It should be noted that most of the large differences are observed in mountainous areas, where observation networks are generally considered less robust. In the panels representing the differences between ERA5 and ERA-Interim in Figure 4.2, it can be seen that the ERA5 product corrects the biases in ERA-Interim; i.e. the areas that were too hot in ERA-Interim are colder in ERA5 and vice versa. Based on these results, the ERA-I temperature dataset will be excluded from any analysis over Africa.

Figure 4.3 presents a similar analysis for precipitation. It shows observed mean annual precipitation (upper left) and differences of the 9 studied precipitation datasets. It can be seen that the precipitation products differ widely, depending on the source and type of processing of data. Datasets that integrate observations (first row: CPC, GPCC, MSWEP and last row: PERSIANN and CHIRPS) show much smaller differences in general, as compared to the four reanalysis products (central row). ERA5 is the best-performing reanalysis, followed by its predecessors, ERA-I, JRA55 and CFSR products, which are wetter over most of North America. Figure 4.3 also shows large differences in the western mountain ranges for all datasets, outlining limitations for all gridded precipitation datasets under study.

To further investigate precipitation seasonality, Figures 4.4 and 4.5 present seasonal precipitation differences for winter (DJF) and summer (JJA).

Results for winter (figure 4.4) are very similar to those obtained at the annual scale. It seems that ERA-Interim and ERA5 are very similar, as the differences between those datasets are small. One exception is the western coast, where a dry difference persists although it has been reduced in ERA5 as compared to ERA-Interim. Summer differences (figure 4.5) do, however, display important differences. These differences are smaller for the gauge-based and satellite-based datasets, and larger for the reanalysis datasets in general. The CFSR reanalysis is particularly dry in the central USA. The differences between ERA5 and ERA-I are also larger, with ERA5 having smaller differences all across North America. Additionally, ERA5 has a strong reduction in differences for the eastern half of the United States, where ERA-Interim was problematic.

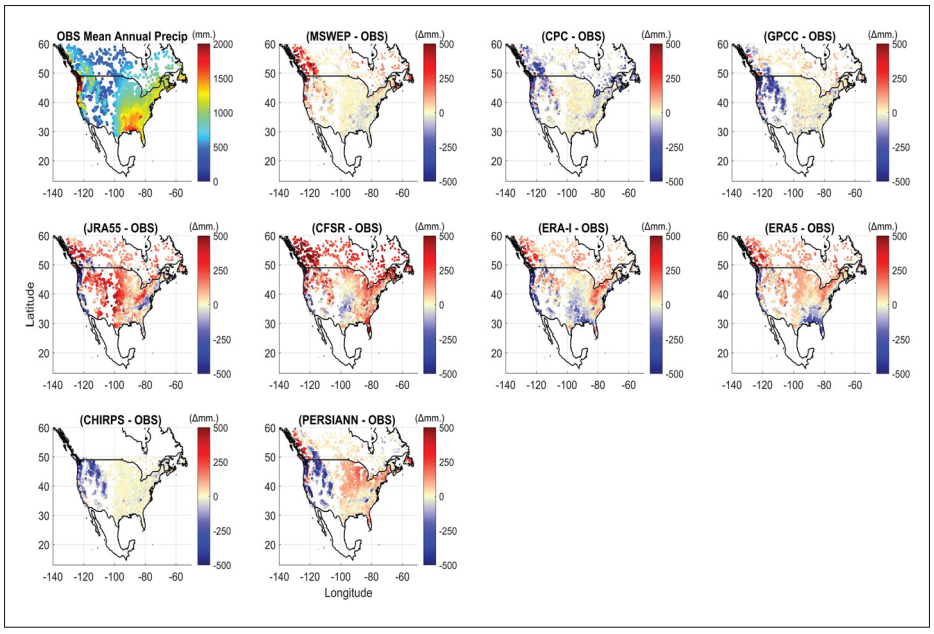


Figure 4.3 Difference maps of mean annual precipitation (dataset-OBS). Note that CHIRPS does not provide data beyond $\pm 50^{\circ}$ Latitude

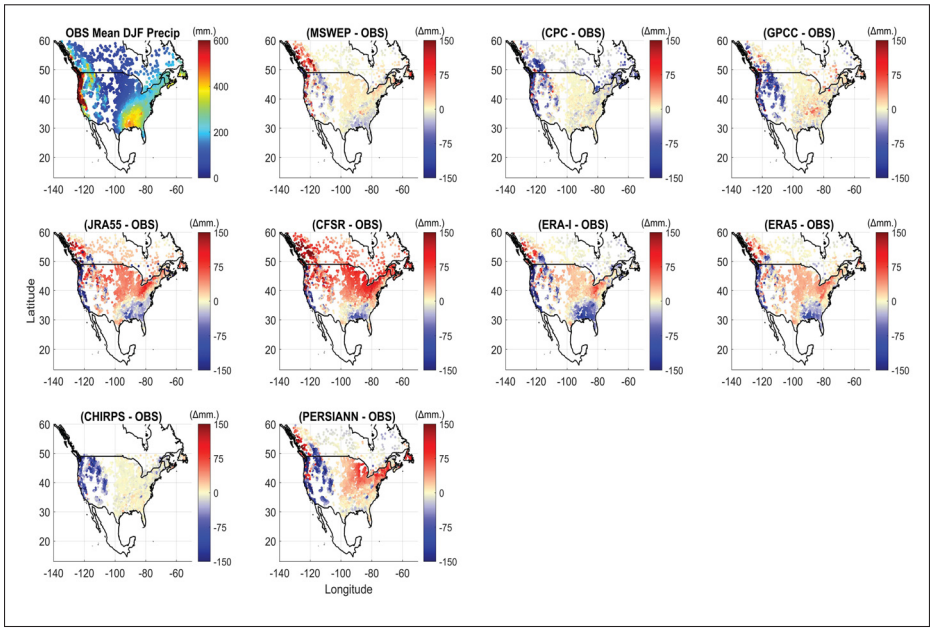


Figure 4.4 Mean winter (DJF) Precipitation difference maps for the 1983-2012 period

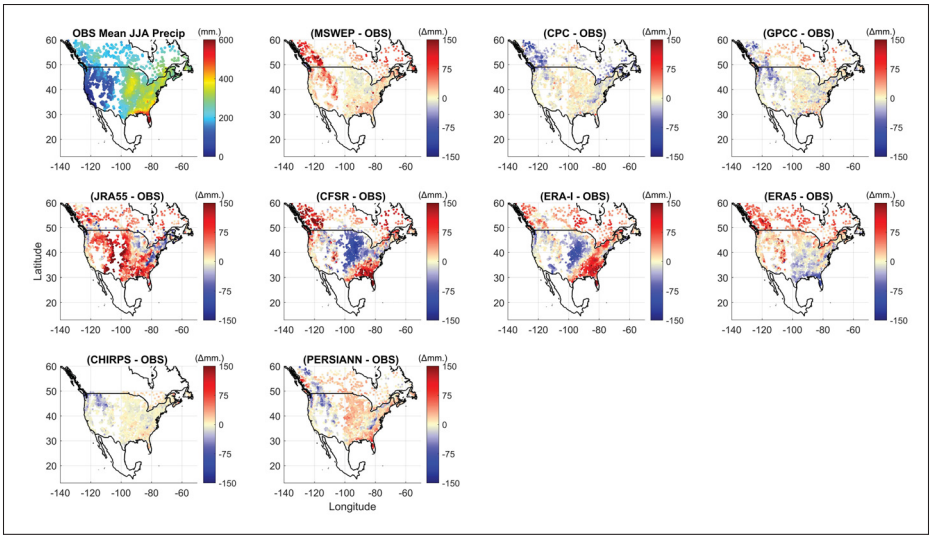


Figure 4.5 Mean summer (JJA) Precipitation difference maps for the 1983-2012 period

To analyze the results at a more localized scale, mean annual precipitation statistics were computed for each catchment and compared to the reference precipitation dataset. Figure 4.6 presents boxplots for annual precipitation Mean Error (ME), Mean Absolute Error (MAE), RMSE and correlation coefficient. The spatial distribution of the last two metrics is also presented in Figures 4.7 and 4.8. The boxplots are built from 3138 values, one from each individual catchment. The central boxes show the 25^{th} and 75^{th} quantiles (bottom and top), with the median in red. The whiskers display the smallest and largest values. Red crosses are considered statistical outliers. Overall, when compared to the reference dataset, we see that MSWEP is consistently the closest across all metrics. The two gauge-based products (CPC and GPCC) and CHIRPS follow. The median difference of ERA-I is close to zero, but otherwise displays a large spread. Surprisingly, ERA5 shows a relatively large positive mean difference, as do the other reanalyses (JRA55, CFSR). Correlation coefficients tell a similar story, with the main differences being that ERA5 clearly outperforms the other reanalysis. RMSE distributions also follow a similar pattern.

Figure 4.7 presents the spatial distribution of mean annual precipitation RMSE values between each precipitation dataset and the reference dataset. JRA55 and CFSR are clearly the worst-

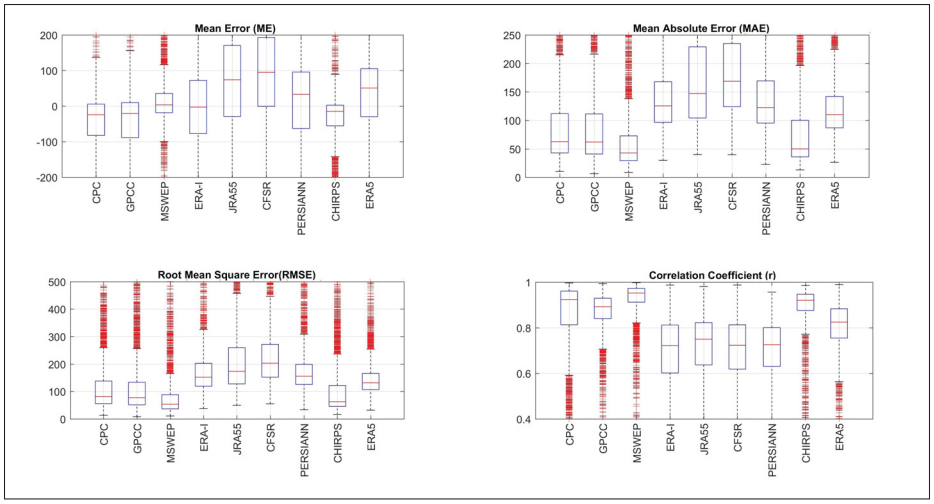


Figure 4.6 Boxplots comparing ME, MAE, RMSE and r for 9 precipitation datasets at the annual scale

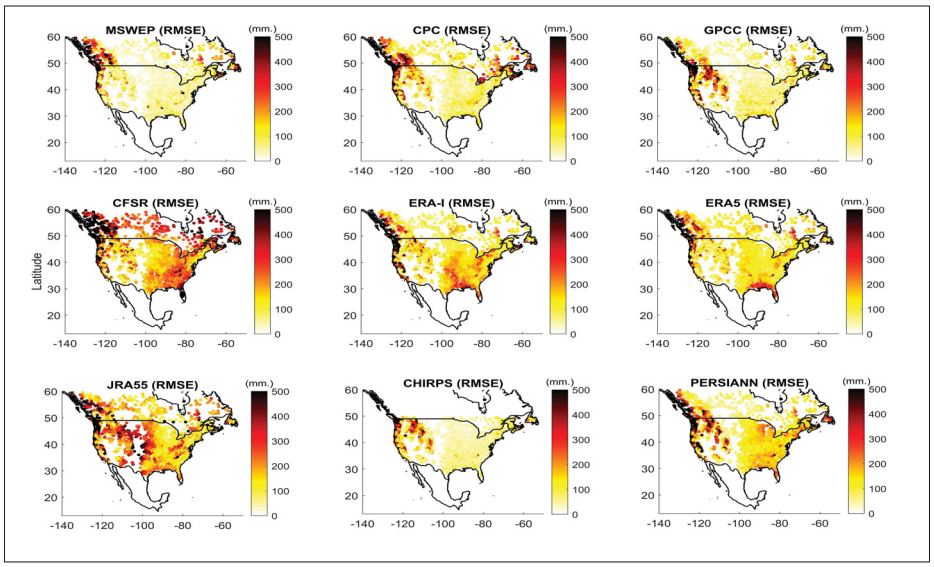


Figure 4.7 Root Mean Square Error (RMSE) of mean annual precipitation for 9 precipitation datasets

performing datasets. MSWEP performs the best everywhere, with the exception of Western Canada, where ERA-5 and GPCC perform best.

Figure 4.8 presents the spatial distribution of correlation coefficients calculated for daily precipitation. MSWEP, ERA5, GPCC and CHIRPS are clearly the best-performing datasets. CPC

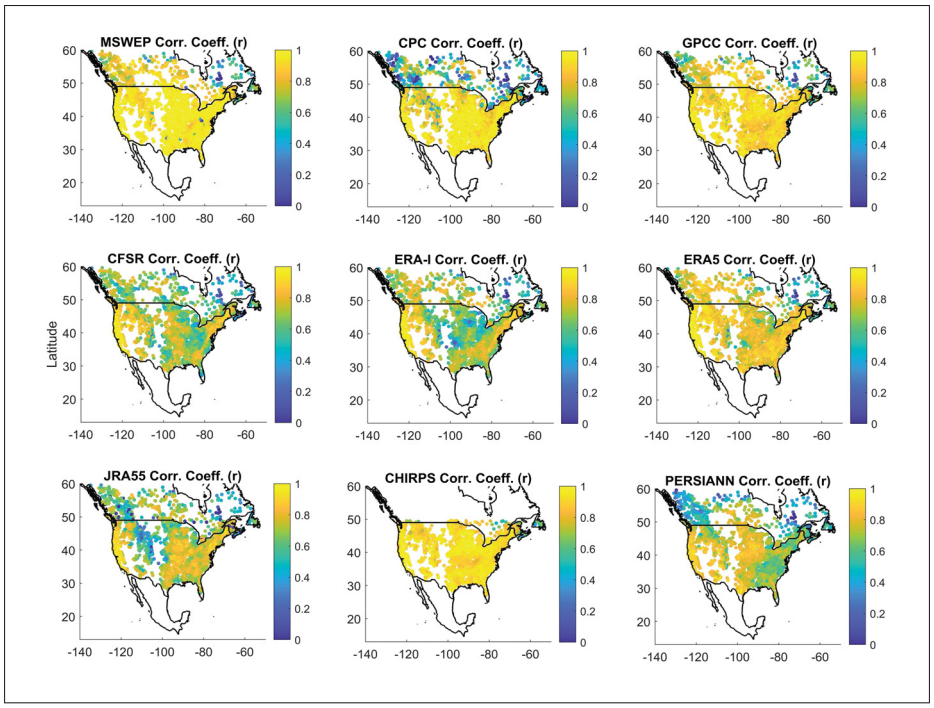


Figure 4.8 Spatial distribution of correlation coefficients computed at the daily time step

performs well over the USA, but quite badly in Canada, where weather station density gets lower.

4.1.2 Hydrological model simulations over North America

This section presents the results of the hydrological model simulations using all possible combinations of precipitation and temperature datasets. The results shown in this section are for the HMETS model only. Both models display very similar spatial patterns and HMETS was chosen as it generally outperforms the GR4J model. Figures 4.9 and 4.10 show the distributions of KGE scores for all catchments below (figure 4.9) and above 50°N (Figure 4.10). The separation at 50°N was made for two reasons: the unavailability of data for CHIRPS, and to investigate the impact of the much lower resolution of observation networks in the North, which should technically affect gauge-based datasets.

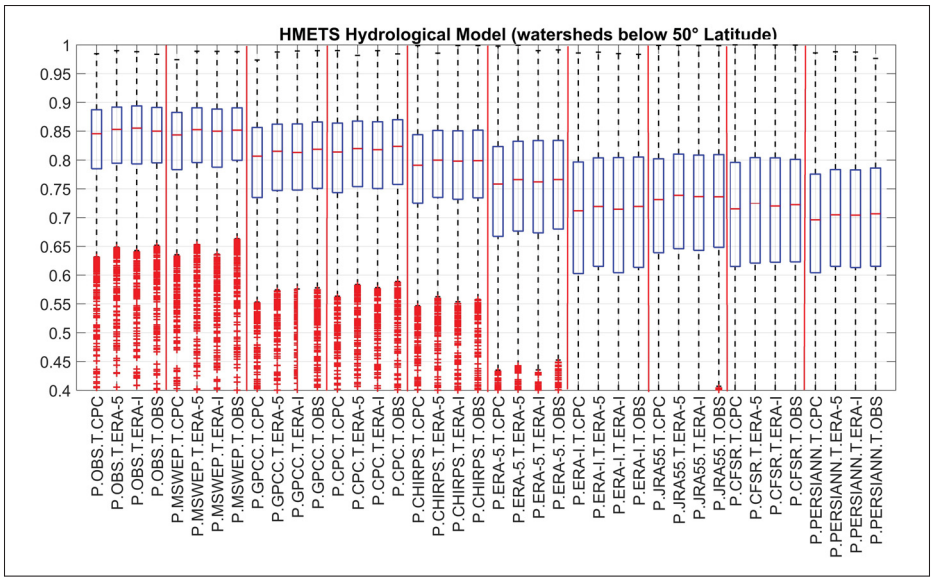


Figure 4.9 KGE boxplots of simulated streamflows (below 50°N Latitude) from 10 precipitation datasets and 4 temperature datasets (40 combinations) using the HMETS hydrological model

Figure 4.9 shows that all datasets can be used to generate good hydrological modelling, with all combinations generating median KGE values larger than 0.7. There are, however, large performance variations across datasets and some outliers are present for all combinations. These outliers reflects the lower performance of these datasets in some watersheds, either due to regional dataset issues, streamflow records deficiencies, or a combination thereof. It can be seen that the main driver of the modelling skill is the precipitation dataset. All four temperature datasets offer a nearly equal performance below 50°N, although CPC is consistently the worst of the four. The reference and MSWEP datasets clearly outperform all other precipitation datasets. These are then followed by the GPCC and CPC gauge-interpolated datasets, and then by CHIRPS and ERA5. The other satellite (PERSIANN) and reanalysis (ERA-I, JRA55 and CFSR) products perform clearly worse than their best counterparts (CHIRPS and ERA5).

The results for catchments north of 50°N (figure 4.10) are markedly different. The differences between all datasets is much smaller, with the exception of CPC, which is the worst-performing dataset. This is consistent with results presented in Figure 4.8 showing that CPC precipitation behaves quite differently over Canada. The lower density of the station network is an equalizer,

preventing gauge-based datasets from outperforming their counterparts. MSWEP and ERA5 are the best-performing datasets, and are slightly better than using the regional gridded dataset used as a reference.

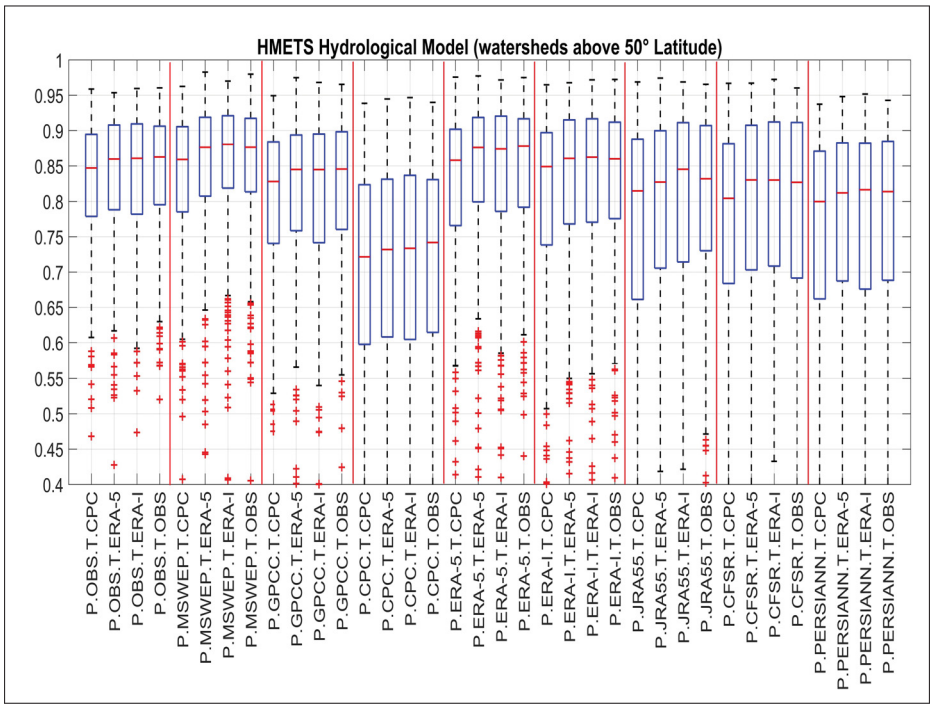


Figure 4.10 KGE boxplots of simulated streamflow for 9 precipitation datasets and 4 temperature datasets (36 combinations) using the HMETS hydrological model (above 50°N). Note that CHIRPS V2.0 does not provide data beyond $\pm 50^{\circ}$ latitude, and is excluded from this comparison

Overall, ERA5 and the observations provide very similar results, whereas ERA-Interim and CPC temperatures lag slightly behind. In this sense, the temperature data from ERA5 are marginally more accurate for hydrological modelling at the catchment scale than the other two temperature datasets and are similar to that of the observed temperature dataset. Hydrological modelling performance is very good, and, for most datasets, better than below 50°N. This is very likely a combination of watersheds being larger, thus producing smoother, less reactive and easier to model hydrographs, and because snow-dominated catchments typically have relatively simple hydrographs, with a long winter recession curve and a strong spring snowmelt

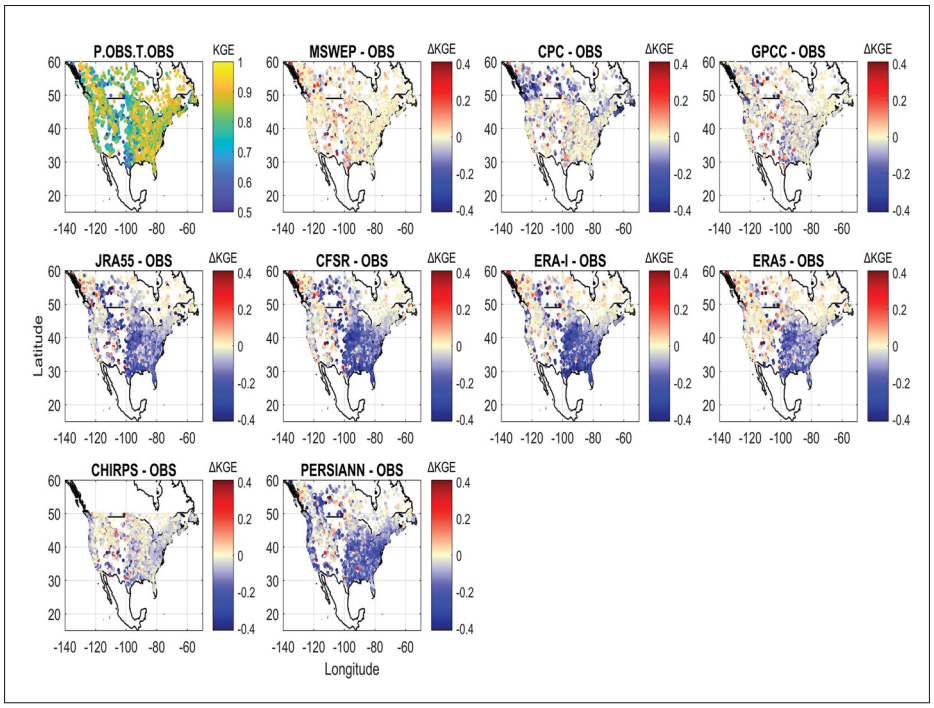


Figure 4.11 Spatial distribution maps for the KGE difference between the observed precipitation dataset combined with the observed temperature dataset (top left plot) and the 9 precipitation datasets combined with the observed temperature dataset

signature. The gain in performance is notable for all reanalyses, which are less affected by a deficient local observational network.

Figure 4.11 shows the spatial distribution of KGE scores for the different precipitation datasets combined with the reference temperature dataset. The upper left graph shows the KGE values for the reference dataset, whereas all the other graphs display the difference in KGE values for each precipitation dataset. A red colour indicates a better performance, and blue, a worse one. It can be seen that MSWEP, GPCC, and CHIRPS to some extent, compare favourably with the reference dataset, and that CPC is affected by the lack of stations in the northern parts of North America. Also of note is the strong negative score associated with some of the reanalysis and satellite datasets in the eastern United States. Outside of this zone, ERA5 performs extremely well, as noted by (Tarek *et al.*, 2019).

Finally, Figure 4.12 presents the aggregate mean KGE score over all catchments for all precipitation temperature pairs (first row), as well as for the catchments below (second row) and above 50°N (third row). The first two rows are nearly identical due to the much larger number of stations located below 50°N. The third row displays warmer colours related to the prevalence of snowmelt dominated watersheds, which are easier to model. Otherwise, these results confirm those of Figure 4.11, and underline the relatively poor performance of CPC above 50°N for precipitation, and to a lesser extent, for temperature. Reanalysis datasets perform comparatively much better with both ECMWF reanalysis (ERA5 and ERA-I) among the best products.

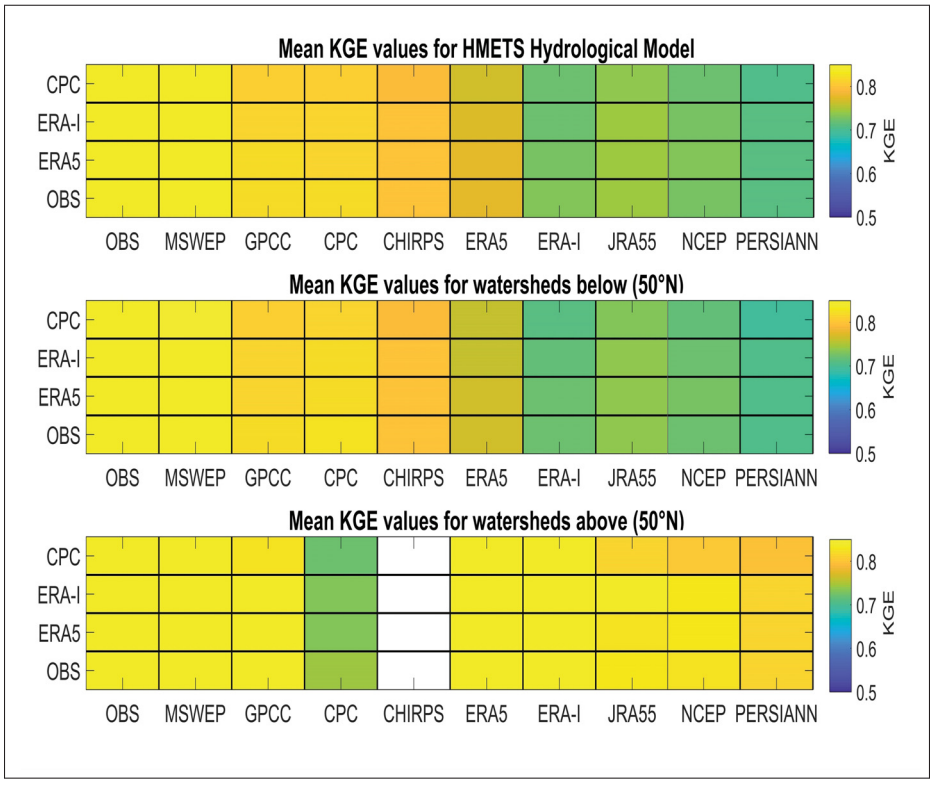


Figure 4.12 Mean KGE values for all catchments (top panel), catchments below 50°N Latitude (center panel) and catchments above 50°N Latitude (bottom panel) for 10 precipitation datasets and 4 temperature datasets. CHIRPS does not provide data beyond ±50° Latitude, and is left in white

4.2 Evaluating the gridded datasets in Africa

4.2.1 Analysis of precipitation and temperature in Africa

Figure 4.13 presents mean annual temperature over the 1983-2012 period for the two selected temperature datasets. Both datasets display the same temporal patterns. ERA5 is however significantly warmer than CPC with a typical warm difference of 5-6 degrees over most of Africa. This difference is very large and can potentially affect evapotranspiration. However, the specific calibration of the hydrological model to each dataset has the potential to take this into account.

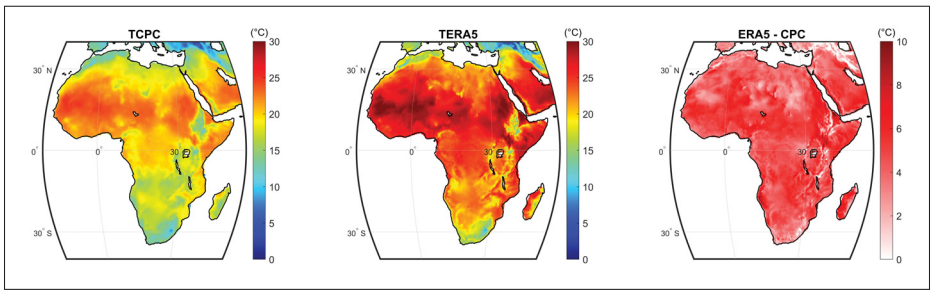


Figure 4.13 Mean annual temperature maps and the difference between the two datasets for the period (1983-2012) over Africa

Regarding the precipitation datasets, to better present the differences between the products, the difference in the mean annual precipitation was calculated between each individual dataset and the average of all datasets as shown in Figure 4.14. The average here is considered as the reference benchmark. Since all the gridded datasets have different spatial resolution, the datasets were first interpolated to the finest grid scale. A red color indicates that the dataset is wetter than the average, while the blue color indicates it is dryer. Results show important differences between the different precipitation datasets. All the datasets are generally similar in the desert and semi-desert regions but large differences are obvious in the tropical western and central regions.

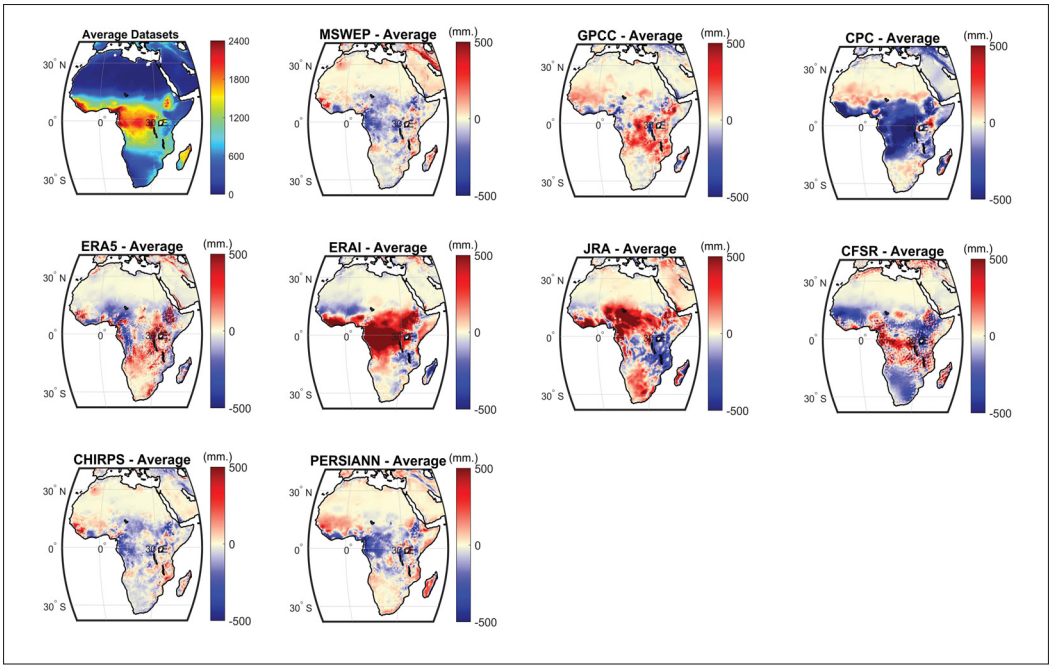


Figure 4.14 Difference maps for mean annual precipitation (dataset-Average) in Africa. The (top-left) plot is the average of all the 9 precipitation datasets

Overall, the reanalysis (middle row) are wetter over the intertropical zone, with ERA5 being much closer than the other three considered reanalysis. The CPC gauge-based dataset is much drier than all other datasets, whereas at the opposite end, the ERA-I reanalysis dataset is much wetter. The large differences between both gauge-based datasets (CPC and GPCC) outline the complexity of interpolation in data-sparse regions. Differences between the other datasets are comparatively smaller. In the absence of any reliable reference datasets, it is difficult to interpret the differences observed here. While an outlier dataset (e.g., CPC) may lead to suspicion, the limitations associated with each dataset do not allow for any firm conclusion to be drawn. This is why hydrological modeling is used as an indirect validation method in this study. Even though streamflow gauges records do contain errors (Di Baldassarre & Montanari, 2009), in the context of this study, they are considered as the most reliable source for validation of the precipitation and temperature datasets.

Figures 4.15 and 4.16 present seasonal precipitation differences for summer (JJA) and winter (DJF). It can be seen that the merged product (MSWEP), the two satellite products (PERSIANN and CHIRPS) and the reanalysis product (ERA5) generally provide the best representation, as the differences between these datasets and the average are much smaller compared to the other products in both seasons. The reanalysis datasets have significant differences with the exception of ERA5, which is therefore considered the best-performing reanalysis. The two satellite datasets; CHIRPS and PERSIANN, perform similarly for both seasons. Both gauge-based precipitation datasets display large differences for both seasons. The GPCC is dry over the intertropical zone, while CPC is much wetter over the same region.

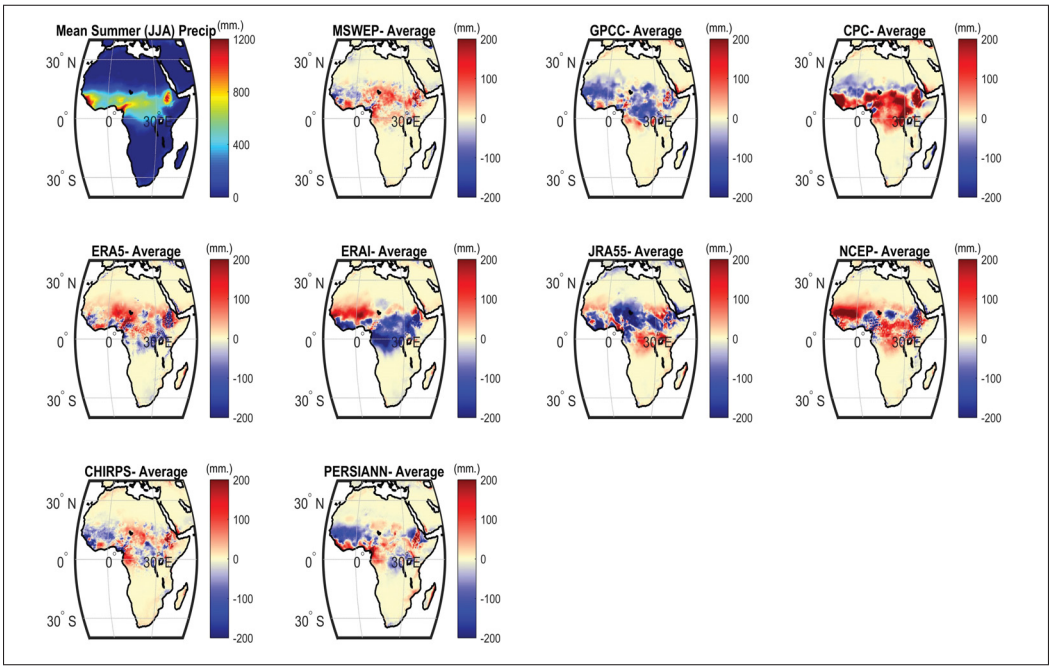


Figure 4.15 Mean summer (JJA) Precipitation difference maps for the 1983-2012 period in Africa. The (top-left) plot is the mean summer precipitation of all the 9 datasets

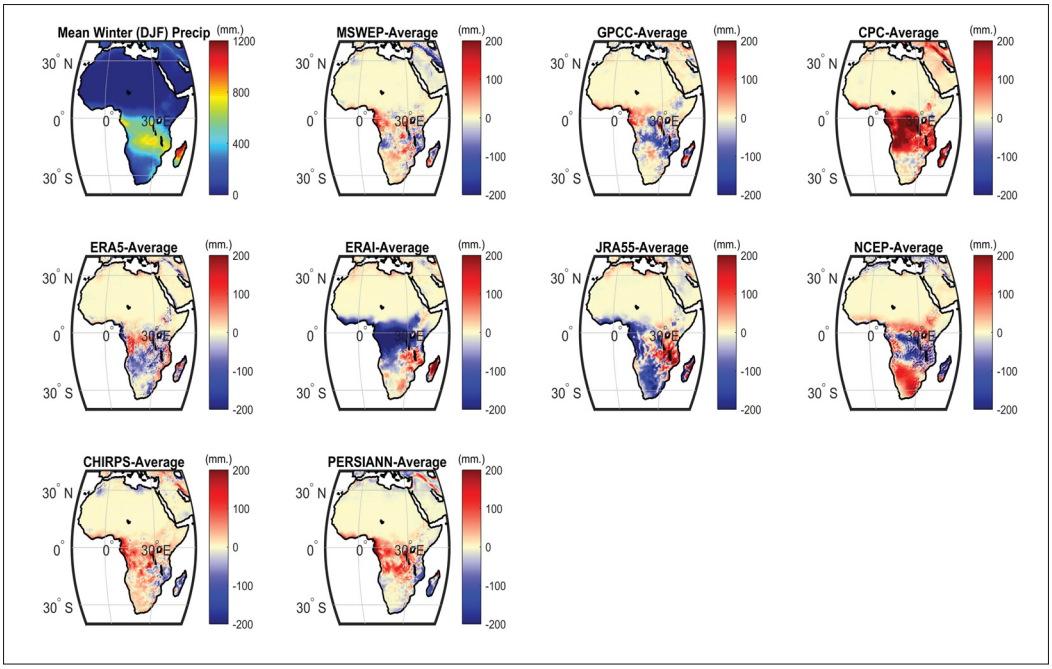


Figure 4.16 Mean winter (DJF) Precipitation difference maps for the 1983-2012 period in Africa. The (top-left) plot is the mean winter precipitation of all the 9 datasets

4.2.2 Hydrological model simulations over Africa

This section presents the results obtained from the hydrological modelling simulations. Figure 4.17 shows the distribution of KGE scores for each of the 18 combinations of the 9 precipitation and 2 temperature datasets. Results are only shown for the HMETS model since both hydrological models display the same patterns. In addition, HMETS generally outperforms the GR4J model in terms of KGE scores. Each boxplot in Figure 4.17 contains the KGE scores of the 850 catchments provided from the GRDC database.

Many conclusions can be drawn from Figure 4.17. Both temperature datasets perform very similarly across all precipitation datasets, although ERA5 (blue boxplots) gives very small but consistently better results. Most of the differences observed therefore originate from the precipitation datasets.

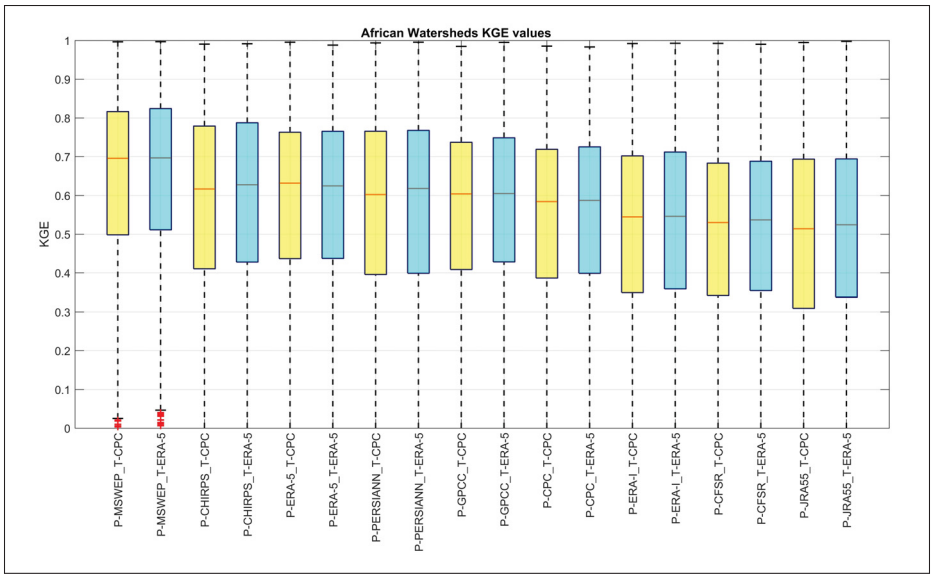


Figure 4.17 KGE boxplots of simulated streamflows from 9 precipitation datasets and 2 temperature datasets (18 combinations) using the HMETS hydrological model

All precipitation datasets result in an acceptable KGE median value larger than 0.5, showing they can all be used for hydrological modeling. There are however large differences with some datasets clearly outperforming others. The CPC and GPCC gauge-based datasets are outperformed by, respectively, five and four datasets. The MSWEP merged-product is quite clearly the best-performing precipitation dataset, followed by the CHIRPS satellite and the ERA-5 reanalysis datasets. The ERA-I, CFSR and JRA reanalysis are the least-performing datasets in this study.

In order to study the impact of spatial variability, Figure 4.18 presents the spatial distribution of KGE values for all nine precipitation datasets used in conjunction with ERA5 temperature. The spatial patterns are consistent for all precipitation datasets. Hydrological modelling performance is generally quite good everywhere with the exception of South Africa. This could either be due to less reliable streamflow records in this region or more likely to the hydrological model difficulties in dealing with the arid climate of south Africa. Rainfall-runoff models have long been known to have difficulties in such climates (Wheater *et al.*, 2007).

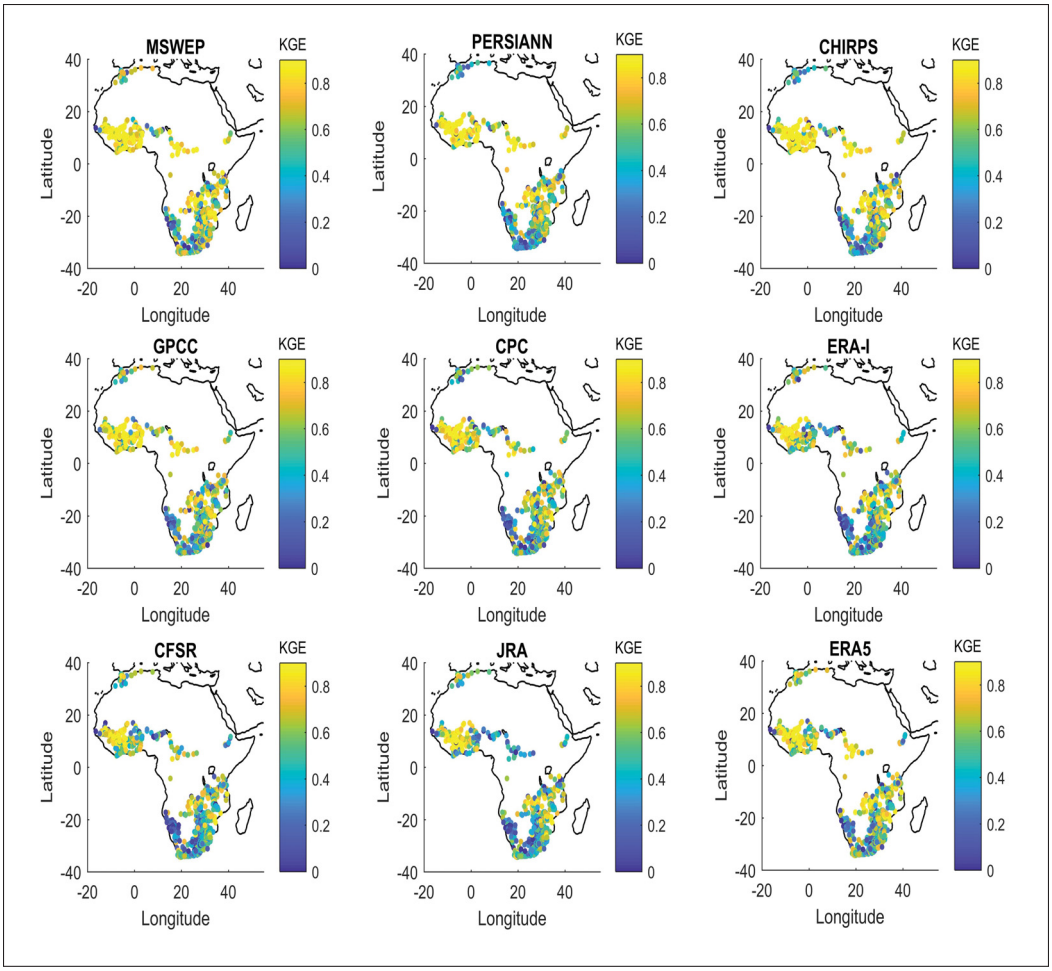


Figure 4.18 Spatial distribution of Kling-Gupta efficiency metrics for nine precipitation datasets and ERA5 temperature dataset using the HMETS hydrological model

4.3 Contribution to uncertainty - Variance analysis in African catchments

Figure 4.19 presents the calibration results for both hydrological models using all possible combinations of the 9 precipitation and 2 temperature datasets. Each boxplot consists of 350 KGE values corresponding to the calibration results for each of the 350 selected gauged African catchments. Each box extends from the 25^{th} quantile to the 75^{th} quantile, with the median displayed as the red line within that range. The top and bottom whiskers (where shown) represent highest and lowest values. Red crosses are considered statistical outliers.

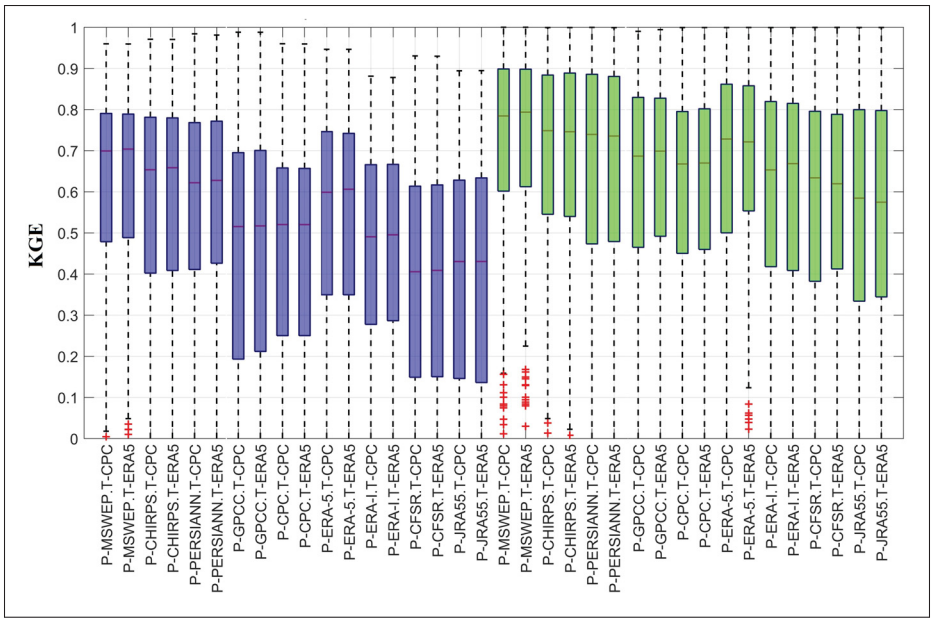


Figure 4.19 KGE calibration values using the 18 possible combinations of precipitation and temperature datasets, for both hydrological models (GR4J in blue and HMETS in green) for each of the 350 selected gauged catchments

Results show that both hydrological models perform well, but that there are important differences between datasets. HMETS performs better than GR4J, with respective overall mean KGEs of 0.58 and 0.41. All the precipitation and temperature datasets result in acceptable median KGE values (Pechlivanidis & Arheimer, 2015).

Both temperature datasets perform very similarly across all combinations, with ERA5 generally slightly outperforming CPC. Figure 4.19 clearly shows that most of the variability seen originates from the precipitation datasets. Four precipitation datasets are ahead of the field. They are in order of performance: the merged product MSWEP, followed by the two satellite datasets; CHIRPS and PERSIANN, and the ERA5 reanalysis dataset. The gauge-based precipitation datasets (e.g., GPCC and CPC), and the ERA-I reanalysis follow with a similar performance. Finally, the CFSR and JRA55 reanalysis are the worst-performing products for hydrological model calibration.

Table 4.1 presents the main results of the analysis of variance for the 2071-2100 period for the gauged catchments. It shows the relative variance for all main effect (precipitation datasets (P), General Circulation Models (GCM), temperature datasets (T) and hydrological modeling (HM)), first-order interactions of the four components of uncertainty under study, and for 6 streamflow metrics (shown in rows 5 to 10). The variance originating from second- and third-order interactions are summed up and presented in the last row. QQ5 and QQ95 are respectively the 5^{th} and 95^{th} quantiles of streamflow distribution. QX1 is the 30-year mean of the annual daily maximum streamflow value. Results show that most of the variance consistently comes from 5 sources, for all 6 streamflow metrics. They are: precipitation datasets (P), GCMs, hydrological models (HM), interactions between precipitation datasets and GCMs (P-GCM) as well as interactions between precipitation datasets and hydrological models (P-HM). The colored-rows outline the main contributors to variance.

Table 4.1 Mean percentage of variance for 6 streamflow metrics for 350 gauged catchments

	Mean Q	**Winter Q**	**Summer Q**	**QQ5**	**QQ95**	**QX1**	**Average**
P	21.62	24.12	28.54	34.38	23.17	22.36	25.70
GCM	39.71	24.93	27.29	4.39	39.56	25.82	26.95
T	0.17	0.12	0.09	0.02	0.15	0.04	0.09
HM	5.18	8.43	19.99	21.96	5.59	5.50	10.11
P-GCM	21.55	25.19	10.20	3.42	16.01	26.33	17.12
P-T	0.02	0.01	0.02	0.01	0.02	0.01	0.015
P-HM	7.38	9.72	14.69	31.12	8.17	8.78	12.31
GCM-T	0.01	0.01	0.006	0.0018	0.017	0.005	0.008
GCM-HM	1.30	2.13	1.44	1.36	2.49	3.49	2.04
T-HM	0.0087	0.0098	0.0069	0.0041	0.0189	0.0058	0.009
Others	2.78	5.20	3.46	2.99	4.60	7.58	4.43

Table 4.1 indicates that both the precipitation datasets and GCMs are the main contributors to variance, including through interactions (P-GCM). The hydrology models also generate some uncertainty, and in particular, through interaction with the precipitation datasets. All metrics exhibit a similar pattern, with the exception of the low-flow metric (QQ5), where precipitation, hydrological models and their interaction components (P-HM) are dominant, and for which

GCM uncertainty is minimal. In almost all cases, the five highlighted components represent approximately 92% of the total variance. The average amount of variance introduced by both temperature datasets is less than 0.20% for all 6 different streamflow metrics.

To show cross-catchment variability, Figure 4.20 shows boxplots of the relative variance attribution results for the 5 main contributors to variance, as identified in Table 4.1, and for the same 6 streamflow metrics. The results are also decomposed into three parts: all 1145 catchments (A), as well as the 350 gauged (G) and 795 ungauged (U) catchments, in order to ensure that the regionalization process does not introduce undesirable effects on the results.

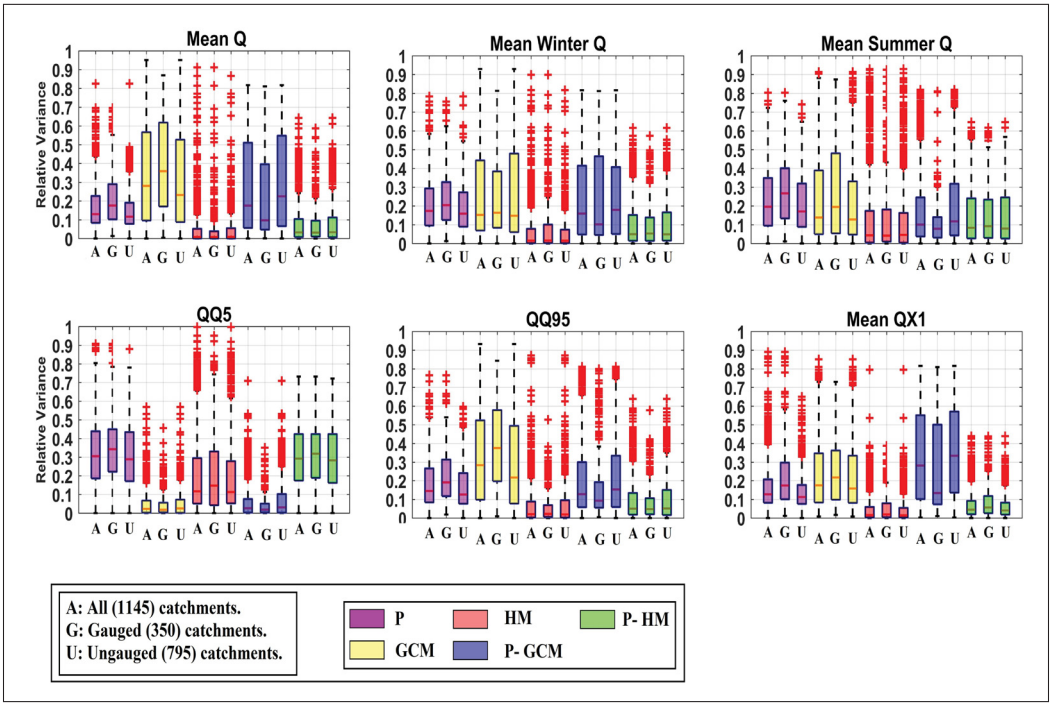


Figure 4.20 Boxplots of the relative variance attribution results. For each component, variance is shown for all (1145), gauged (350) and ungauged (795) catchments

Figure 4.20 shows that the response of the gauged and ungauged catchments is very similar across all variance components and streamflow metrics, and that no major variance artifact is introduced by the regionalization step. Consequently, all further results will only be shown for all 1145 catchments, with no differentiation made between the gauged and ungauged ones.

The results show that there is considerable across-catchment variability, as shown by the extent of the boxplots, with GCM and P-GCM interaction being the most important, and most variable contributors to variance. As was shown in Table 4.1, the low-flow metric displays a pattern that is much different from the other five metrics, with HM being important and GCM, being the lowest. There is a relatively large difference between the two metrics representing high flows (Q95 and Q1X). While GCM dominates the former, a much larger part of the uncertainty is transferred to the precipitation dataset (P and P-GCM) for the latter.

In order to study the impact of spatial variability, Figure 4.21 presents the spatial distribution of the relative variance attribution for the five main contributors to variance of Table 4.1 and all 6 streamflow metrics. Mean Q, Winter Q, QX1 and QQ95 display somewhat similar spatial patterns. Summer Q and QQ5 metrics display somewhat similar spatial patterns. The largest precipitation uncertainty (P and P-GCM interactions) is found in the northern parts of Sub-Saharan Africa, between 0 and 30°N. GCM uncertainty appears to be larger all around the coastlines of Africa. Hydrological model uncertainty is strongest for QQ5, but spatial patterns are fairly consistent across all 6 streamflow metrics. GCM uncertainty is strongly different for both Summer Q and Winter Q, likely because of the monsoon pattern. Above 20°N, there is generally less than 100 mm of total annual precipitation, and some level of care should therefore be taken when analyzing results in relative contribution to variance.

In other words, a variance analysis of a metric with very little absolute variance could be misleading. Consequently, Figure 4.22 displays the standard deviation of the 360 streamflow values computed for each streamflow metric and for each watershed. The streamflow value for each metric is normalized per unit area to allow for a comparison of large and small watersheds in the same figure. Not surprisingly, the results demonstrate a larger variance along the equatorial band, where precipitation is largest. This pattern is particularly clear for the QQ95 high-flow metric. The catchment database is, however, large enough to show some catchments which exhibit a large variance, even in arid regions above 20°N and below 20°S.

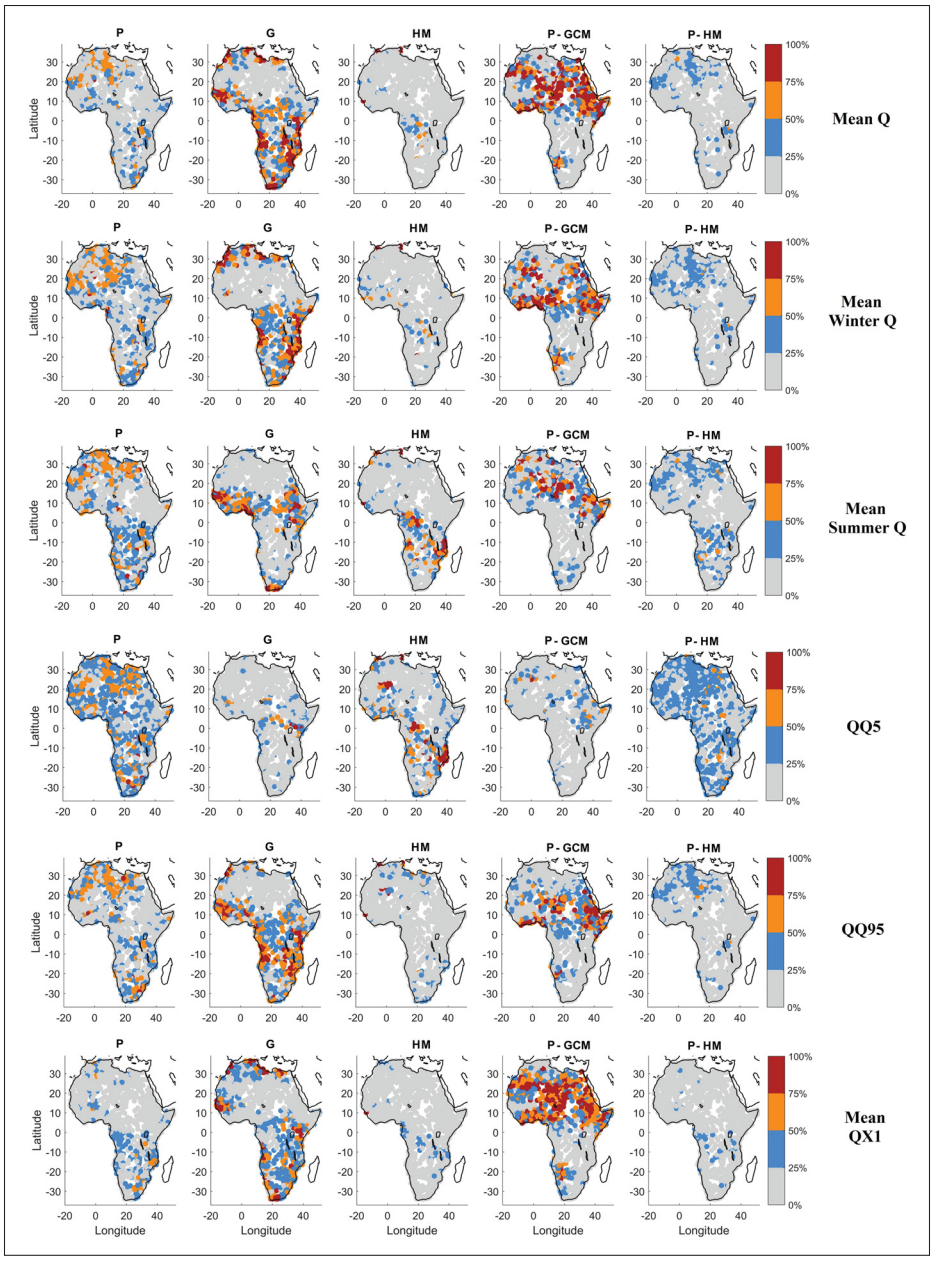


Figure 4.21 Spatial distribution of the five main contributors to variance for each of the 6 streamflow metrics

Since some precipitation datasets are clearly better than others based on the hydrological model calibration results, it may not be entirely fair to compare precipitation uncertainty to GCM uncertainty. To investigate this further, the uncertainty contribution obtained when using all 9 precipitation datasets is compared to that of 3 sub-ensembles, as presented in Table 4.2. While

ensemble 4 is composed of the clearly best-performing datasets for model calibration, the main goal here is to investigate the impact of dataset selection, not the definition of a credibility ensemble, as will be further discussed later.

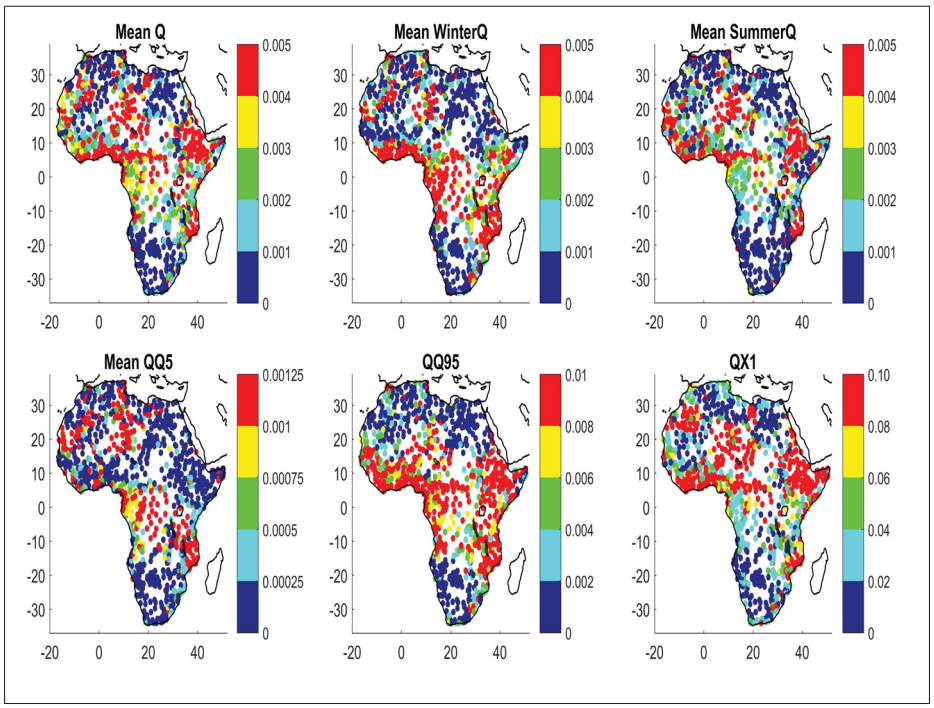


Figure 4.22 Standard deviation of discharge per unit area (in $m^3/sec/km^2$), constructed from 360 values for each catchment and streamflow metric

Figure 4.23 presents the boxplots of percentages of variance for each catchment, for the five main contributors to variance for all 4 precipitation dataset ensembles of Table 4.2. Unsurprisingly, it shows that reducing the size of the precipitation ensemble results in a consistent decrease in the variance attributed to precipitation. Most of this reduction in variance comes from the P-GCM interaction term, although there is also a noticeable decrease in the main effect P component. The lost precipitation variance is transferred mostly to GCMs, and to a lesser extent, to hydrological modeling. The exception is the low-flow QQ5, where most of the variance is transferred to HM. Most of the drop observed is obtained by dropping the five worst precipitation datasets, as no significant difference is observed between precipitation ensembles

Table 4.2 List of ensembles of precipitation datasets

Ensemble	**Number of precipitation datasets**	**Rationale for selection**	**Datasets included**	**Datasets excluded**
1	9	All 9	All	None
2	7	Mean KGE$\geq$0.65	MSWEP, CHIRPS, PERSIANN, ERA5 GPCC, CPC, and ERA-I	CFSR, and JRA55
3	4	Best of each category (merged, satellite, gauge, and reanalysis	MSWEP, CHIRPS, GPCC, and ERA5	PERSIANN, CPC, ERA-I, CFSR, and JRA55
4	4	Best 4	MSWEP, CHIRPS, PERSIANN, and ERA5	GPCC, CPC, ERA-I, CFSR, and JRA55

3 and 4. Even in a reduced ensemble, precipitation datasets still provide between 10 to 20% of median variance, and more than 30% for the low-flow metric (QQ5) when taking into account the main effect and first-order interaction term.

Figure 4.24 presents the spatial distribution of the relative variance attribution for each of the 6 streamflow metrics after including only the four best overall precipitation datasets (Ensemble 4 of Table 4.2). This is the same as Figure 4.21 but with a reduced precipitation ensemble. Results outline that GCM uncertainty is the dominant source of uncertainty when using the reduced precipitation ensemble, with the exception of the low-flow metric, for which precipitation uncertainty remains dominant. There are, however, significant interactions between GCM and precipitation for all metrics, especially in the Northern half of the continent. Otherwise, the observed spatial patterns are similar to the ones presented in 4.21.

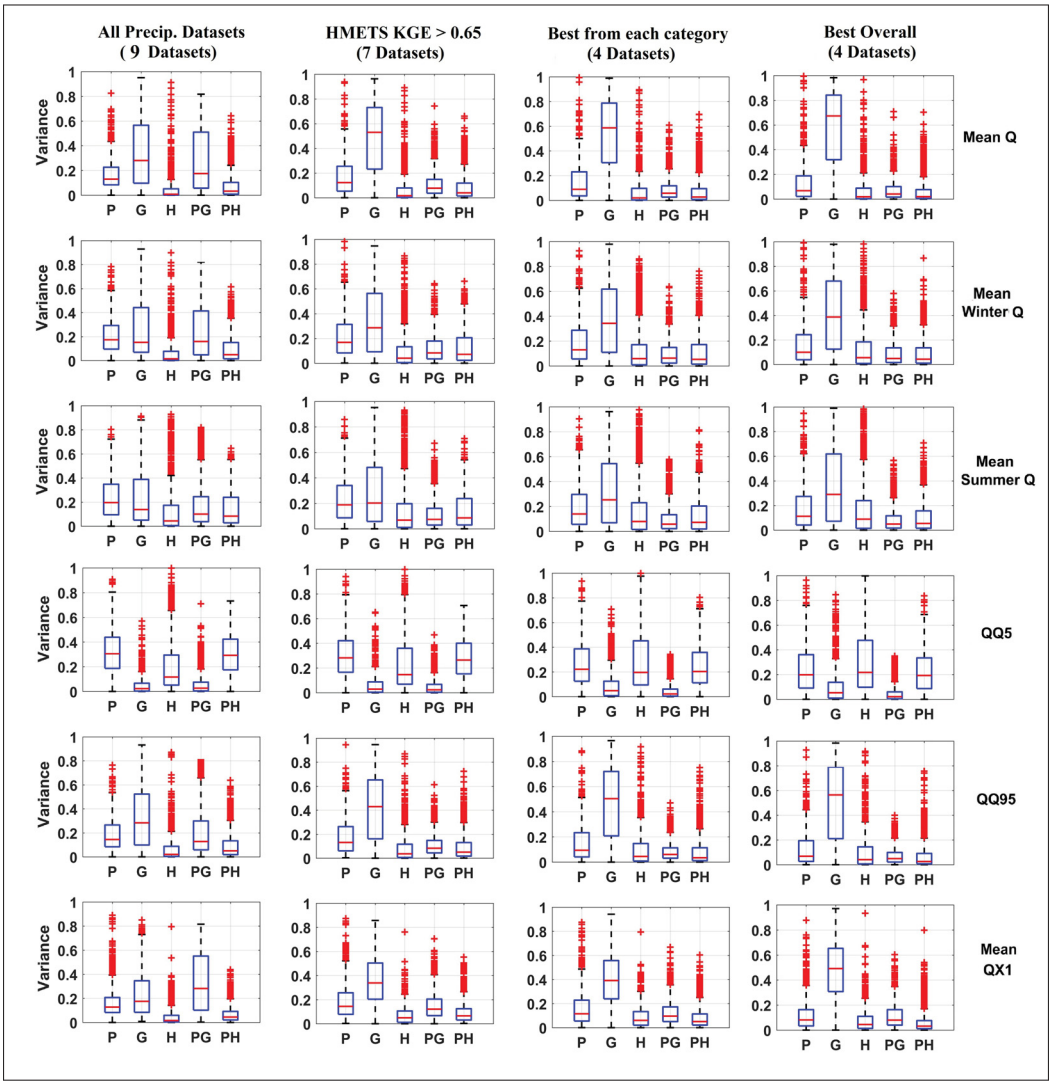


Figure 4.23 Boxplots of the five main components of the variance attribution: precipitation (P), GCMs (G), hydrological models (H), interaction between precipitation datasets and GCMs (PG) and interaction between precipitation datasets and hydrological models (PH). Columns represent the four precipitation ensembles of Table 4.2, while rows represent the 6 hydrological indices investigated in this study

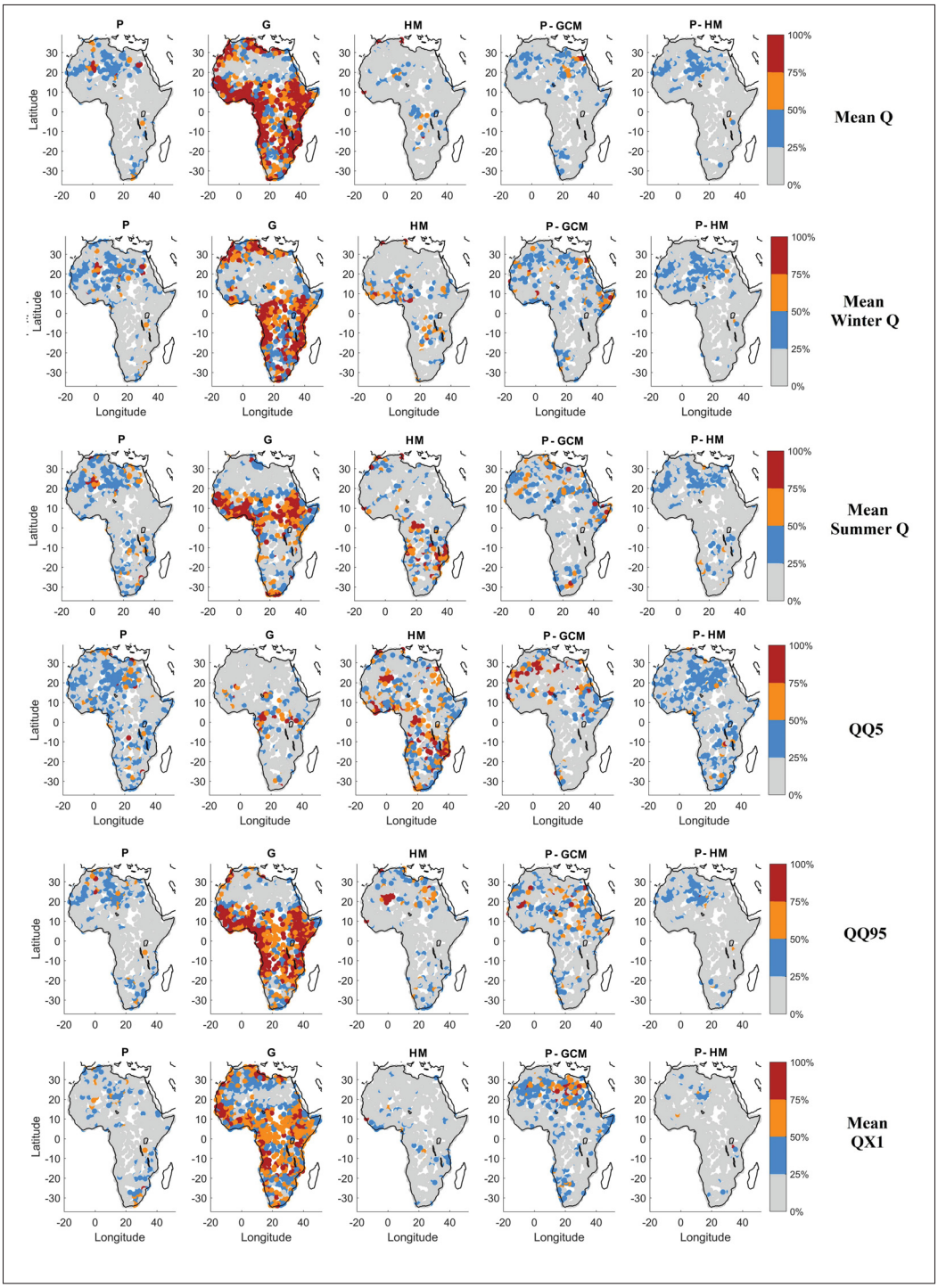


Figure 4.24 Spatial distribution of the five main contributors to variance for each of the 6 streamflow metrics, using the 4 best precipitation datasets (Ensemble 4 of Table 4.2)

CHAPTER 5

DISCUSSION

Impact models strongly rely on hydrometeorological information. The performance of such models (stochastic and deterministic) is fundamentally dependent on the quality of input data. Defining a reference climate dataset is an important but difficult task. A reference climate dataset is used as a benchmark for monitoring environmental changes and correcting climate model biases of future climate projections to assess future impacts of a changing climate. Weather station observations are still considered as the most accurate representation of the current climate, but are limited in both time and space. Time series of relevant hydrometeorological variables are plagued with problems such as short temporal horizons, missing data, measurement errors, instrument biases and discontinuities introduced through equipment change and modification of the environment of weather stations, including their displacement. To allow for regular data coverage and remove missing data, it is now a common practice to interpolate station data onto a regular grid. Such gridded datasets greatly simplify the processing of meteorological data for environmental studies at the regional, continental and global scales. In regions with a good weather station coverage, gridded datasets using the same underlying data differ due to the different interpolation methods (Essou *et al.*, 2016a), and typically see an increase in the number of wet days and a decrease in the frequency of extreme events (Ensor & Robeson, 2008). In regions with scarce weather stations coverage (such as Africa), interpolation becomes extrapolation, and is therefore potentially highly unreliable. Additionally, the slow but steady decreasing trend in the number of weather stations around the world (Lawrimore *et al.*, 2011) compounds the problem. Gridded datasets are created to try to overcome many of the above problems. While it is likely that multi-source merged gridded products are the way of the future, it is not clear how good and reliable the many currently available gridded products are.

Several inter-comparison studies have been done (Beck *et al.*, 2017b; Essou *et al.*, 2017) including over Africa (Satgé *et al.*, 2020; Dembélé *et al.*, 2020). These studies outline a complex

picture in which performance depends on scale, climate and data source, and for which no dataset consistently outperforms all of the others. Because of this, in data-sparse regions such as Africa, there is not only no commonly agreed upon reference dataset, but even no agreement on the optimal source of climate data (e.g., satellite vs. reanalysis), and different environmental studies have used completely different datasets. This work sheds some light on this issue by comparing nine global or near-global precipitation datasets and three temperature datasets first over North America, and then over Africa, therefore combining regions with high and low densities of weather stations. For climate change impact studies, there is no knowledge on how dataset uncertainty may propagate in the typical hydroclimatic modeling chain. The results presented in this study attempt to answer this question by comparing dataset uncertainty to other sources of uncertainty, such as that derived from GCMs.

5.1 Gridded datasets evaluation

The results show important differences between all the datasets, as well as within categories of datasets (gauge-based, reanalysis and satellite-based). All the datasets were shown to be adequate for driving a hydrological model. However, some datasets were clearly better than others in various circumstances. This is in agreement with the results of Beck *et al.* (2017b) and Beck *et al.* (2019). A first conclusion was that most of the dataset uncertainty originates from precipitation. Temperature displays much smaller spatial and temporal variability than precipitation, and can therefore be a lot more reliably interpolated using the adiabatic lapse rate to account for elevation and terrain orientation in mountain areas. There was little difference between the four selected temperature datasets (NAC^2H observations, CPC, ERA-I and ERA5), even though CPC performed slightly worse than the selected reanalyses (ERA5 and ERA-I) and the reference gridded dataset. The equal performance of both reanalyses, when compared to the reference gridded datasets, could likely be explained by the fact that they assimilate the surface temperature from weather stations (in addition to a plethora of other data sources) and by the relatively small spatial and temporal variability of temperature, at least when compared to precipitation. Our evaluation of temperature is, however, based solely on hydrological modelling.

Hydrological models have the ability to filter out some level of variability in driving inputs. Many other levels of validation still need to be performed (e.g., extremes) to determine if these alternative products are able to represent specific types of events in the hydrologic cycle. These results are nonetheless very encouraging for reanalyses, which are now available in near real-time and at spatial and temporal resolutions matching or exceeding those of most observational networks.

Comparatively, precipitation datasets provided a much more challenging problem, which explains why most dataset intercomparison work has focused on this variable. The selection of the best-performing precipitation dataset was evaluated over a reference period using the single metric of the KGE criterion. The criterion is considered to be a good metric as it weights bias, correlation and RMSE between simulation and observations, all rightfully considered to be important attributes of a good hydrological simulation. There are, however, many other metrics that could have been chosen to perform this comparison, some of which might be even more important for specific applications such as floods. Based on KGE performance over a common reference period, all nine precipitation datasets performed adequately in terms of hydrological modeling performance, but some clearly performed much better than others. One important conclusion of this work is that the relative performance of precipitation datasets in North America below 50°N, which includes the contiguous United States and southern Canada, and above 50°N. Above 50°N, the density of the Canadian observational network is much lower. Results imply that a low-density station network narrows the gap between the reanalyses and gauge-based products. Reanalyses do not assimilate surface precipitation in their analysis scheme, and are therefore much less affected by a lack of ground precipitation measurements (either sparse station network or precipitation undercatch in gauged locations). This suggests that ERA5 precipitation is as robust as the best gauged products above 50°N and reanalyses should therefore be considered as good candidates in regions with deficient observational networks, confirming the conclusions of Tarek *et al.* (2019).

The spatial and temporal resolution of the datasets reviewed in this work differ widely. The temporal resolution itself (hourly to sub-daily) was not investigated. The spatial resolution

of the above products, which varies from 0.05° to 1°, was summarily evaluated by analyzing hydrological modelling performance with respect to watershed size and by elevation, on the basis that higher-resolution datasets would perform better on smaller watersheds, or for high elevation watersheds, where the topography is more complex. No clear link was found between dataset performance and either size or elevation (results not shown). The main notable result was a clear improvement of ERA5 over ERA-I for high elevation catchments. This suggests that the resolution difference between the highest and lowest resolution datasets is not large enough to make a difference in this type of application, or that the use of a global hydrological model (which requires the averaging of the contributing grid points irrespective of their resolution) is not ideal to investigate the impact of resolution. Most of the selected watersheds are relatively large and therefore have a response time larger than one day. On those watersheds, the averaging of input data coupled with the smoothed hydrographs from the global hydrological models result in differences that are very difficult to see when using a criterion such as the KGE metric. Other metrics (e.g. peak flow reproduction, streamflow variance) may have been better suited to study the impact of dataset resolution (Kokkonen & Jakeman, 2001). The conceptualized nature of the hydrological models used in this study may also not be best suited to outline such differences. For the smaller watersheds in our database, sub-daily modeling would be better suited (Bevelhimer *et al.*, 2015), but was not feasible since most datasets are limited to the daily time step. The use of a distributed hydrological model may be preferable to study the impacts of data resolution.

The results show that amongst all datasets tested in this study, MSWEP either is the best dataset, or is tied for best. The performance of MSWEP demonstrates the potential of merged products in providing high quality outputs, by utilizing and integrating all available information. In high network density regions, MSWEP weighs observations heavily, but also relies heavily on reanalysis when weather station observation networks are less dense, such as in Northern Canada and Africa. We can expect an increasing number of datasets to rely on multi-source information, at the regional and global scales. At the regional scale, for example, high-resolution datasets can be obtained by combining ground-based radars and weather stations (Lespinas

et al., 2015; Shen *et al.*, 2018). In addition, there are other potential reasons for the excellent MSWEP results in North America. MSWEP is the closest relative (in terms of construction and resolution) to the chosen reference dataset (NAC^2H) and especially over the US. Over Africa and Canada, MSWEP relies to a much larger extent on reanalysis as well as using streamflow data in its merging scheme, which may give it an advantage over the other datasets in terms of long-term biases.

CHIRPS performed very well for most of the comparison criteria and in both study regions; North America and Africa. It performed better during the warm seasons, owing to its limitation in terms of detecting snowfall in North America. CHIRPS, which integrates satellite and gauge stations data on a high spatial resolution grid of 0.05°, has been shown to be a viable choice in climatological studies. Other studies have indeed mentioned its quality in this regard (Toté *et al.*, 2015; Duan *et al.*, 2016; Poméon *et al.*, 2017; Beck *et al.*, 2017b; Duan *et al.*, 2019).

For hydrological modelling in Africa, the MSWEP merged-product dataset was clearly the best performing one, followed by CHIRPS and ERA5 products, respectively. The performance of all other reanalysis datasets (ERA-I, CFSR and JRA55) was inferior. However, in other studies, JRA55 was shown to provide the best reanalysis (Odon *et al.*, 2019), while CFSR was successfully used for precipitation modeling (Khedhaouiria *et al.*, 2018). Clearly, the results presented in this study should only be used as intended (i.e., to study uncertainty related to the choice of a reference climate dataset), and not as a judgment of the absolute performance of each dataset. As mentioned earlier, it is important to keep in mind that all of the datasets used in this study generate adequate streamflow simulations. In North America, the results have shown that, in general, gauge-based datasets perform better than reanalyses, whereas the performance of the two selected satellite products differ widely, with CHIRPS clearly outperforming PERSIANN. CPC is the worst gauge-only product, especially so over Canada. A few relevant studies have assessed the influence of gauge-density on climate data (Arsenault & Brissette, 2014a; Gubler *et al.*, 2017; Hofstra *et al.*, 2010; Janis *et al.*, 2004). In particular, Janis *et al.* (2002) evaluated the required station-density to capture the regional climate variability in the United-States. The

study reported that a station-density of 1 station per 180 km^2 would be needed to adequately monitor the climate variability.

ERA5 presents clear improvements over its predecessor (ERA-I), and is the best reanalysis product for hydrological modelling amongst those used in this study. ERA5 shines brightly, particularly in North America above 50°N. The high spatial (0.25°) and temporal (1 hour) resolutions of ERA5 and the fact that it is available in near real-time lends it a significant advantage over most of the other datasets. The ECMWF recently launched the ERA5-LAND reanalysis at a 0.1° resolution. It uses the same assimilation process as ERA5, but is run at a finer resolution over land. Reanalyses could be considered as extremely complex multi-source merged products, and are likely to gain in importance in the near future. Their main limitation, when compared to MSWEP, for example, is that they do not integrate precipitation gauge data into the assimilation scheme. Reanalyses are very likely to be supported and improved in the future, as compared to the other datasets used in this study, which do not rely on recurrent national funding, and which often result from the efforts of small teams. Reanalysis performed generally worse than gauge-based products, and particularly so in Africa and the eastern half of the U.S. Essou *et al.* (2016a) showed that reanalysis had difficulties reproducing the seasonal cycle of precipitation over this region in the U.S. Reanalysis precipitation could easily be post-processed at the monthly scale using observations to palliate this problem, as was previously done on older reanalysis products (Weedon *et al.*, 2014).

Overall, results show that gauge-based datasets should be preferred in regions with good weather network density, with MSWEP being clearly the best performing dataset as represented by its results. In regions where observational network density is much poorer, CHIRPS and ERA5 datasets perform relatively well. This indicates that ERA5, and potentially CHIRPS would be good choices as reference datasets for climate change impact studies in data sparse regions.

5.2 Uncertainty of gridded datasets

The uncertainty contribution of datasets to future streamflow uncertainty was first evaluated using all 9 precipitation datasets, in conjunction with 2 temperature datasets, a sample of 10 GCMs and two hydrological models, for a total of 360 possible element combinations. While this is a relatively large sample, not all sources of uncertainty were accounted for. In particular, GHGESs, downscaling and bias correction were not included in the analysis. In comparison, the North American Climate Change and Hydroclimatology Dataset (NAC^2H) database (Arsenault *et al.*, 2020) offers 16,000 combinations allowing examining future streamflow uncertainty. In this regard, the relative variance contribution of the climate dataset is best examined in comparison to that of GCMs, the most studied source of climate change impact uncertainty. Results outline the important, and in some cases, dominant contribution of the precipitation dataset to the overall uncertainty of future streamflows. For all 6 streamflow metrics presented here, the precipitation dataset uncertainty was comparable and sometimes larger than that of GCMs.

Uncertainty contribution was then studied by retaining subsets of precipitation datasets, eliminating the least performing ones with respect to the chosen KGE metric. This follows the concept of a credibility ensemble based on carefully selecting the best/most robust components of the hydroclimatic modeling chain, in order to obtain the most credible uncertainty range (Brissette *et al.*, 2020; Giuntoli *et al.*, 2018; Maraun *et al.*, 2017). Results demonstrate a large decrease in contribution to uncertainty for 5 of 6 streamflow metrics. The precipitation dataset remained the largest contributor to uncertainty for the low-flow metric, and still accounted for 10 to 20% of the total variance for the other metrics. Most of the decrease in uncertainty was obtained by dropping the worst-performing datasets, rather than keeping the best-performing in each category.

The results presented here indicate that hydrological model uncertainty is relatively small, with the exception of the low-flow metric. These results should be taken with caution because only two hydrological models were used, and also because they both share the same potential

evapotranspiration (PET) formula. For climate change impact studies, the climate sensitivity of PET is now thought to be an important source of uncertainty for impact studies (Clark *et al.*, 2016; Brissette *et al.*, 2020), and the importance of hydrological model uncertainty has been outlined in many studies (Vetter *et al.*, 2017; Krysanova *et al.*, 2018; Giuntoli *et al.*, 2018). It is therefore likely that the contribution of hydrological models is underestimated here.

It is recommended that reference dataset uncertainty be included in climate change impact studies, and especially so in regions with a sparse network of weather stations. We believe that climate dataset uncertainty can be minimized for most streamflow metrics using a careful validation and selection of the best-performing ones. A dataset ensemble should nonetheless be retained to assess the sensitivity of the impact study to the choice of a reference dataset. As is the case for most other elements of the hydroclimatic modeling chain of future climate change impacts, there is '*no free lunch*' in the sense that there is no single recipe, which will be applicable in all cases. Climate dataset performance is spatially-dependent, as shown here and in other studies, and will depend on the criteria used to assess said performance. In addition, the relative uncertainty contribution also depends on the catchment location and streamflow metric under study. The importance of first-order interactions in variance analysis, and especially of interactions between precipitation datasets with GCMs and with the hydrology models testify to the complex nature of the propagation of uncertainties in the hydroclimatic modeling chain. The use of an appropriate credibility climate dataset ensemble is therefore more than likely to be catchment-related and metric-dependent, and some minimum level of upstream validation would be needed to select the best components.

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

The main objective of this study was to assess the uncertainty related to the choice of a reference dataset against that of other sources of uncertainty in climate change impact studies. This was achieved in a three-step process. First, the performance of precipitation and temperature global gridded dataset products was assessed over 3138 North American catchments, where the underlying network of weather stations is dense and of good quality. The same datasets were then re-evaluated over Africa, where the station-network is sparse and of lower quality. Finally, a large-sample hydrological climate change impact uncertainty study over 1145 African catchments was performed.

In North America, the datasets were compared against high-resolution regional gridded dataset (NAC2H). Performance was evaluated using annual and seasonal biases, mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and coefficient of correlation (r). Streamflows were simulated using all 40 possible combinations of precipitation and temperature datasets, and compared against data from gauging stations. In comparison, there is no equivalent high-quality reference dataset available in Africa as a base for comparison. Therefore, the comparison was done against the average of all datasets. The performance was evaluated statistically using the annual and seasonal differences. Hydrological models were used to simulate the streamflows and to compare the datasets performance against streamflow records from the Global Runoff Data Centre (GRDC).

Results showed that precipitation datasets are the main driver of uncertainty due to the relatively large differences between the datasets. Comparatively, differences between temperature datasets played a much smaller role as all four products behave very similarly. Temperature derived from the ERA-5 reanalysis provided consistently better results than the other tested

temperature datasets. For precipitation, overall, the merged-product MSWEP consistently performed best over both Africa and North America. Both gauge-based global products performed well over the U.S., but their performance decreased over Canada (and particularly in the case of CPC Unified), where observations are based on a less dense observational network. The ERA5 reanalysis performed really well over Africa, Canada and Western U.S., but its overall performance was affected by a relatively poorer performance over the Eastern U.S. It clearly outperformed the other three tested reanalyses. CHIRPS was found to be easily the best-performing satellite precipitation dataset over Africa and North America, outperforming PERSIANN and all reanalysis.

For the uncertainty contribution of datasets to future streamflow, the study used 9 precipitation and 2 temperature datasets, along with 10 GCMs and 2 hydrological models, for a total of 360 possible combinations. Results showed that temperature dataset-related uncertainty was minimal, with a median relative contribution to uncertainty less than 0.20% for all 6 presented streamflow metrics. On the other hand, the nine precipitation dataset ensembles generated a future uncertainty equal to or larger than that related to GCMs. Using a reduced ensemble of the best-performing precipitation datasets systematically reduced the precipitation dataset uncertainty, but still accounted for 10 to 20% of the total variance for 5 of the 6 streamflow metrics, and still remained the main source of uncertainty for the low-flow metric. The main conclusion of this study is that the choice of a climate reference dataset can induce significant uncertainty in climate change impact studies, at least in regions with a sparse weather station coverage.

Limitations and recommendations for future work

As is the case with any large-scale comparison studies, some methodological limitations may potentially impact conclusions drawn from the presented results. This subsection outlines the limitations of this research and the recommendations for further work.

1- This study does not calculate the relative future changes for either the climate variables or the expected streamflows. Therefore, the obvious next step is to generate a climate change impact study over Africa.

2- In terms of hydrological modeling, this study uses only two lumped conceptual models sharing the same potential evapotranspiration (PET) formula. It is therefore very likely that hydrological modeling uncertainty is underestimated in this study. Therefore, using more complex hydrological models (including more physically-based distributed models) would be important to figure out the impact of physically based processes and distributed inputs on the uncertainty. Using additional PET formulas would also be useful for a better understanding of the temperature sensitivity of such formulas, and particularly in the case of commonly used formulas which are mostly empirical and temperature-based only.

3- A single objective function (e.g., KGE) is used in this study. It is likely that other objective functions would return different results, and should be tested in further work.

4- There are several other streamflow criteria that could shed light on differences between datasets, such as extremes. In particular, high-flow extremes have the potential to outline other improvements that we have not analyzed here. In this sense, sub-daily components should be investigated, since most of the datasets studied here have data at the sub-daily time scale. Access to climate models at the sub-daily time scale would however be required and the issue of sub-daily bias correction would have to be considered.

5- Regarding the reference dataset uncertainty, this work evaluates the uncertainty of datasets in regions with low-density station networks. It is suggested to assess the dataset uncertainty in high station network regions as well, such as in North America or Europe. It would also be interesting to consider the uncertainty generated by the datasets over the reference period, and how it propagates in the future.

6- While there is a large range of uncertainty sources that were considered in the study, others were not accounted for in the variance analysis and should be studied. In particular, greenhouse gas emission scenarios and bias correction methods should be explored in more detail in future work, to complete the hydroclimatic modeling chain.

Main scientific contributions of this Thesis

The work presented in this Thesis offers several contributions to scientists concerned with the impact of climate change on water resources. In particular:

1- This Thesis has shown that the choice of a reference dataset is a very important issue and that precipitation dataset uncertainty, in some cases, can be the main source of uncertainty for future streamflows.

2- This Thesis offers a valuable contribution to the study of 'alternative' datasets for hydrological studies. For example,in regions where the weather station network is deficient or sparse, like Africa, MSWEP, CHIRPS and ERA5 were shown to be very valuable sources of data, typically outperforming gauge-based interpolated gridded datasets. While gauge-based gridded dataset remains the best source of data when station network density is good, there are many other products which perform very well and will only get better in the future due to improving computation power and technological advances.

3- The ERA5 temperature dataset provides very similar results compared to those obtained with gauge-based temperature datasets. It also provides much improved precipitation compared to its older version ERA-I. When replacing observed temperatures with ERA5 temperatures, the hydrological modelling performance marginally improves. While it is not a significant difference, this attests to the quality of the ERA5 temperatures in general for hydrological modeling. The results from our work indicate that reanalysis, which will only improve in the future, with

even better temporal and spatial resolutions, could become an important contributor of data for environmental studies.

APPENDIX I

EVALUATION OF THE ERA5 REANALYSIS AS A POTENTIAL REFERENCE DATASET FOR HYDROLOGICAL MODELING OVER NORTH-AMERICA

Mostafa Tarek[1,2], François P. Brissette[1], Richard Arsenault[1]

[1] École de technologie supérieure,
1100 Notre-Dame West, Montréal, Québec, Canada, H3C 1K3

[2] Department of Civil Engineering, Military Technical College, Cairo, Egypt.

Paper published in *Hydrology and Earth System Sciences (HESS)*, May 2020

1. Introduction

Hydrological science knowledge has long been anchored in the need for observations (Wood (1998)). Observations and measurements of all components of the hydrological cycle have been used to gain a better understanding of the physics and thermodynamics of water and energy exchange between the land and the atmosphere (e.g. Luo *et al.* (2018); McCabe *et al.* (2017); Siegert *et al.* (2016); Zhang *et al.* (2016); Stearns & Wendler (1988)). In particular, measurement of precipitation and temperature at the earth's surface has been a critical part of the development of various models describing the vertical and horizontal movements of water. Hydrological models, for example, are routinely used to transform liquid and solid precipitation into streamflows, using other variables such as temperature, wind speed and relative humidity to increase their predictive skill (Singh & Woolhiser (2002)). Throughout the last several decades, such data has essentially been provided by surface weather stations (Citterio *et al.* (2015)). However, and despite the utmost importance of observed data for hydrological sciences, a net decline in the number of stations in the historical climatology network of monthly temperature datasets has been observed since the beginning of the 21st century (Menne *et al.* (2018); Lins (2008)). Perhaps more importantly, data from the NASA-GISS surface temperature analysis shows a particularly large decrease in the number of stations with a long record, a decline starting in 1980. Stations with long records are critical for monitoring

trends in hydroclimatic variables (Whitfield *et al.* (2012); Burn *et al.* (2012)). In addition, the GISS data documents a slow but consistent decrease in the percent of hemispheric area located within 1200km of a reporting station since the middle of the 20th century.

On the upside, other sources of data have steadily appeared to compensate for this worrisome diminishing trend in surface weather stations (e.g. Beck *et al.* (2019); Sun *et al.* (2018); Beck *et al.* (2017a);Beck *et al.* (2017b); Lespinas *et al.* (2015)). Interpolated gridded datasets of precipitation and temperature are now common. They allow some information from regions with good network coverage to be extended, to some extent, towards areas with less information. Interpolated datasets, however, do not create new information, no matter how complex and how much additional information is used in the interpolation schemes (Essou *et al.* (2016a); Newman *et al.* (2015)). Remotely sensed datasets have long carried the hope of bringing relevant hydrometeorological information over large swaths of land, up to the global scale, and over regions with absent or low-density observational networks (Lettenmaier *et al.* (2015)). There are now several global or near global precipitation datasets derived from various satellites with spatial resolutions varying between 0.05° to 1° (Sun *et al.* (2018)). Ground radar based products are also becoming more common and are available at an even higher resolution (Beck *et al.* (2019)). All remotely sensed precipitation datasets do however only provide indirect measurements of the target variable. They typically provide biased estimates, and ground stations are often needed to correct the remotely sensed estimates (Fortin *et al.* (2015)).

Atmospheric reanalysis is another product that has generated interest increasingly in the recent decade. Reanalysis combine a wide array of measured and remotely sensed information within a dynamical-physical coupled numerical model. They use the analysis part of a weather forecasting model, in which data assimilation forces the model toward the closest possible current state of the atmosphere. A reanalysis is a retrospective analysis of past historical data making use of the ever-increasing computational resources and more recent versions of numerical models and assimilation schemes. Reanalysis have the advantage of generating a large number of variables not only at the land surface, but also at various vertical atmospheric levels. Data assimilated in a reanalysis consist mostly of atmospheric and ocean data and do not typically

rely on surface data, such as measured by weather stations. Reanalysis outputs are therefore not directly dependent on the density of surface observational networks and have the potential to provide surface variables in areas with little to no surface coverage. Several modelling centres now provide reanalysis with varying spatial and temporal scales (Lindsay *et al.* (2014); Chaudhuri *et al.* (2013)). Reanalysis and observations share similarities and differ in other aspects (Parker (2016)). Reanalysis have increasingly been used in various environmental and hydrological applications (e.g. Chen *et al.* (2018); Ruffault *et al.* (2017); Emerton *et al.* (2017); Di Giuseppe *et al.* (2016)). They are commonly used in regional climate modeling, weather forecasting and, more recently, as substitutes for surface precipitation and temperature in various hydrological modeling studies (Chen *et al.* (2018); Essou *et al.* (2017); Beck *et al.* (2017b); Essou *et al.* (2016a)). They have been shown to provide good proxies to observations and even to be superior to interpolated (from surface stations) datasets in regions with sparse network surface coverage (Essou *et al.* (2017)). Precipitation and temperature outputs from reanalysis have, however, been shown to be inferior to observations in regions with good weather station spatial coverage (Essou *et al.* (2017)). The relatively coarse spatial resolution of reanalysis is thought to be partly responsible for this. Amongst all available reanalysis, many studies have shown ERA-Interim (European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis) to be the best or amongst the best performing reanalysis products (e.g. Sun *et al.* (2018); Beck *et al.* (2017b); Essou *et al.* (2017); Essou *et al.* (2016a)), arguably the result of its sophisticated assimilation scheme, and despite a spatial resolution inferior to that of most other modern reanalysis. In March 2019, ECMWF released the fifth generation of its reanalysis (ERA5) over the 1979-2018 period (Hersbach & Dee (2016)). ERA5 incorporates several improvements over ERA-I (see section 3 of this paper).

Of particular interest to the hydrological community are the largely improved spatial (30-km) and temporal (1-hour) resolutions. The spatial resolution is now similar or better than that of most observational networks in the world, with the exception of some parts of Europe and the United-States. The hourly temporal resolution matches that of the best observational networks. In the United-States and Canada, for example, there are currently no readily avail-

able observation-derived precipitation and temperature datasets at the sub-daily time scale, and sub-daily records are not consistently available for weather stations. In particular, the hourly temporal resolution, if proven accurate, could open the door to many applications, and notably for modeling small watersheds for which a daily resolution is not adequate. Such watersheds are expected to be especially impacted by projected increases in extreme convective events resulting from a warmer troposphere in a changing climate. Some early results from ERA5 have shown that it outperforms other reanalysis sets and its predecessor ERA-I (Balsamo *et al.* (2018); Olauson (2018); Urraca *et al.* (2018)).

2. Study objectives

This work aims at providing a first evaluation of the ERA5 reanalysis over the 1979-2018 period with an emphasis on hydrological modeling at the daily scale. Even though the hourly temporal scale brings a lot of many potential applications for hydrological studies, a first step in the evaluation of ERA5 precipitation and temperature datasets is performed at the daily scale. The daily scale allows for a comparison against other North-American datasets available at the same temporal resolution, as well as against results from previous studies. In addition, validation at the hourly scale over North-America presents additional difficulties, as discussed above, due to the absence of US or Canadian datasets at this resolution, and to the absence of recorded hourly precipitation for many weather stations. In Canada, for example, fewer than 15% of weather stations have archived hourly variables, and hourly precipitation records contain particularly large ratios of missing data, thus complicating the validation at the regional scale. Consequently, the objectives of this study are to:

1- Provide a first assessment of the potential of ERA5 at providing an accurate representation of precipitation and temperature fields at the daily temporal scale;

2- Evaluate the hydrological modeling potential of ERA5 precipitation and temperature datasets over a large set of hydrologically heterogeneous watersheds using two lumped hydrological models;

3- Based on the above results, document any spatial variability in dataset performance and quantify improvements compared to ERA-I.

3. Methods and data

3.1 Data and study area

The goal of this study is to evaluate the ERA5 reanalysis product as a substitute for observed data and to compare its properties to those of the older ERA-Interim reanalysis for hydrological modelling uses. Therefore, the ERA5, ERA-Interim and observed (weather station) meteorological datasets were used and basin-averaged over 3138 catchments over Canada and the United-States, whose locations and average elevations are shown in Figure I-1. It can be seen that there is a good coverage of the entire domain, although some sparsely populated areas in Northern Canada and in the United-States Midwest have a lower density of hydrometric gauges.

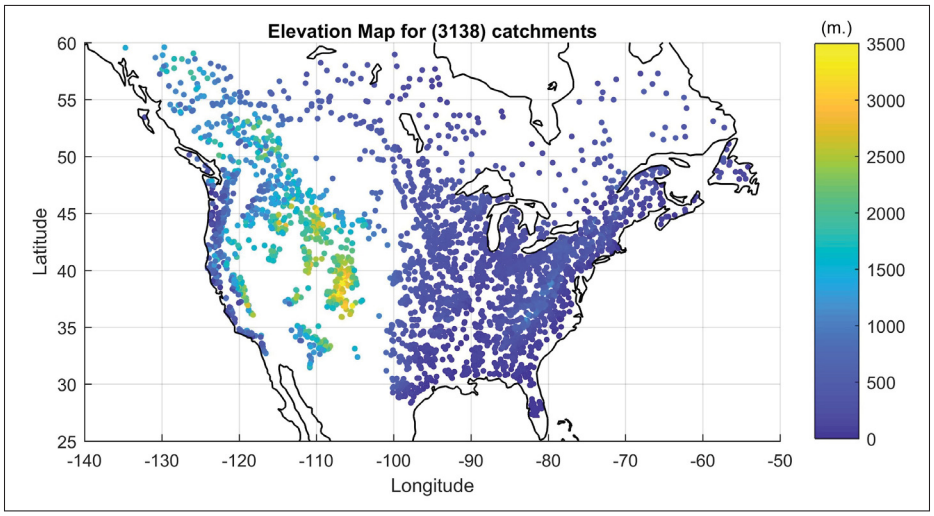


Figure-A I-1 Watershed locations and their mean elevations in North America (each dot represents the watershed centroid)

The hydrological models used in this study required minimum and maximum daily temperature as well as daily precipitation amounts. ERA-Interim and the observed datasets were already

on a daily time step, however ERA5 is an hourly product and as such, it was necessary to derive daily values from the hourly data by summing precipitations and taking the maximum and minimum one-hour temperatures of the day.

3.1.1 ERA-Interim

ERA-Interim (ERA-I) is a global atmospheric reanalysis which was released by the ECMWF in 2006 (Dee *et al.* (2011)) in replacement of ERA40. ERA-I introduced an advanced 4-dimensional variational (4D-var) analysis assimilation scheme with a 12-hour time step. It computes 60 vertical levels from the surface up to 0.1 hPa. Its horizontal resolution is approximately 80km. Precipitation and temperature are available at a 12-hour time step and were aggregated to the daily scale in this work. The production of ERA-I will cease in August 2019, thus providing temporal coverage from January 1999 until August 2019.

3.1.2 ERA5

ERA5 is the fifth generation reanalysis from ECMWF. It provides several improvements compared to ERA-I, as detailed by Hersbach & Dee (2016). The analysis is produced at a 1-hourly time step using a significantly more advanced 4D-var assimilation scheme. Its horizontal resolution is approximately 30km and it computes atmospheric variables at 139 pressure levels. Data for the 1979-2018 period was released in March 2019. The 1950-1978 period is expected to be released in the summer of 2019. This paper only looks at the 1979-2018 because outputs of reanalysis prior to 1979 have been put into question due to the more limited availability of data to be assimilated, and notably from earth-observing satellites (e.g. Bengtsson *et al.* (2004)). While ERA5 may solve some of these problems, it is believed that a careful evaluation of inhomogeneity in ERA5 time series would be needed before using pre-1979 data. ERA5 precipitation and temperature was downloaded and aggregated to the daily time step for this work.

3.1.3 Observed weather data

The observed weather data come from multiple sources due to the transboundary component in this study. Climate data for catchments in Canada were taken from the CANOPEX database (Arsenault *et al.* (2016)), which includes weather stations from Environment Canada that were post-processed and basin-averaged using Thiessen Polygon weighting. The data cover the period 1950-2010. Any missing values were replaced by the NRCan interpolated climate data product (Hutchinson *et al.* (2009)).

For the United-States, historical weather data was taken from the Santa-Clara gridded data product (Maurer *et al.* (2002)) as it was shown to be as good as observations for hydrological modelling in a previous study (Essou *et al.* (2016a)) and covers a long time period (1949-2010). The data is interpolated along a regular 0.125°x0.125° grid, and is then averaged at the catchment scale.

3.1.4 Observed streamflow data

Streamflow records from the United States Geological Survey (USGS) and Environment Canada were used to calibrate the hydrological models at each of the 3138 catchments and evaluate the hydrological modelling performance. The availability of streamflow data was the limiting factor for the simulation length of many catchments, as it varied from 20 years (minimum amount used in these databases) to over 60 years of streamflow records. Missing data were left as-is and were simply not included in the computation of the evaluation metrics.

3.2 Hydrological models

In the course of this study, two lumped hydrological models were implemented and calibrated over each of the available catchments because the large-scale aspect of this study precluded the widespread implementation of distributed models. Although ERA5's spatial resolution is more refined than ERA-Interim (31km vs. 79km), it is still coarse enough that a distributed model would not have changed the results dramatically in this regard. The two hydrological models

selected to evaluate the performance of the various climate datasets, GR4J and HMETS, are flexible, adaptable and have shown to perform well in a wide range of climates and hydrological regimes (Arsenault *et al.* (2018); Arsenault *et al.* (2015); Martel *et al.* (2017); Valéry *et al.* (2014); Perrin *et al.* (2003)). It was decided to perform the study using two hydrological models in order to assess the impacts of the climate data selection on the overall uncertainty of the hydrological modelling simulations.

3.2.1 The GR4J hydrological model

The GR4J hydrological model (Perrin *et al.* (2003)) is a lumped and conceptual model that is based on a cascading-reservoir production and routing scheme. Water is routed from these reservoirs to the outlet in parameterized unit hydrographs. While the original GR4J model includes 4 calibration parameters, the version used in this study had 6 calibration parameters in order to include a snow-accounting and melt routine, namely CEMANEIGE (Valéry *et al.* (2014)). This GR4J-CEMANEIGE (GR4JCN) combination has shown excellent results in studies across the globe (Raimonet *et al.* (2017); Raimonet *et al.* (2018); Youssef *et al.* (2018); Riboust *et al.* (2019); Wang *et al.* (2019)), including in Canada and the United-States. It requires daily precipitation, temperature and potential evapotranspiration (PET) as inputs. The PET was computed using the Oudin *et al.* (2005) as it was shown to be simple yet efficient when used in GR4JCN. Furthermore, the choice of PET is more sensitive than in other simple hydrological models because GR4J does not scale the input PET to adjust its overall mass-balance. Instead, a parameter is included that allows exchanges between underground reservoirs of neighboring catchments.

3.2.2 The HMETS hydrological model

The HMETS hydrological model (Martel *et al.* (2017)) is more complex than GR4JCN, and as such has more calibration parameters (21). While it is similar conceptually to GR4JCN, it has four reservoirs instead of two (surface runoff, hypodermic flow from the vadose zone reservoir, delayed runoff from infiltration and groundwater flow from the phreatic zone reservoir)

allowing for finer adjustments to the runoff and routing schemes. Its snowmelt module requires 10 of the 21 parameters and was selected specifically to be more robust in Nordic catchments with specific routines for snow accounting, melt, snowpack refreezing, ice formation and soil freezing and thawing. As for PET, it uses the same Oudin formulation as GR4JCN but HMETS includes a scaling parameter on PET to control mass-balance. It has also been used in large-scale hydrological studies and has shown overall good performance and robustness in a myriad of climates and hydrological conditions.

3.3 Hydrological model calibration

As will be detailed in the following section, the three precipitation and three temperature datasets were combined in their 9 possible arrangements for analysis purposes. It follows that the sheer number of calibrations to be performed (3 precipitation datasets x 3 temperature datasets x 2 hydrological models x 3138 catchments) in this study required implementing automatic model parameter calibration methods. For this study, the CMAES algorithm was implemented because of its flexibility (Hansen *et al.* (2003)). Indeed, it performs well for small and large parameter spaces such as the 6-parameter and 21-parameter spaces in this study. It was also shown to be robust and is considered as one of the best auto-calibration algorithms for hydrological modelling (Arsenault *et al.* (2014)).

The hydrological model parameters were calibrated on the entire available record of data for each catchment, foregoing the usual model validation step. This method was chosen for two reasons. First, calibrating on all years ensures that the maximum amount of information from the climate data is present in the parameter set, and thus that there is no added uncertainty from choosing calibration and validation years. Second, Arsenault *et al.* (2018) have shown that the model performance is statistically better when more years are added to the dataset, and that validation and calibration skills are not necessarily correlated.

Finally, the calibration objective function was the Kling-Gupta Efficiency (KGE) metric, which is a modified version of the Nash-Sutcliffe Efficiency metric that was introduced by Gupta *et al.*

(2009) and Kling *et al.* (2012). KGE corrects the fact that NSE underestimates variability in the goodness of fit function. It is defined as a combination of three elements:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad \text{(A I-1)}$$

Where r is the correlation component represented by Pearson's correlation coefficient, β is the bias component represented by the ratio of estimated and observed means, and γ is the variability component represented by the ratio of the estimated and observed coefficients of variation:

A perfect fit between observed and simulated flows will return a KGE of 1. Using the mean hydrograph as a predictor returns a KGE of 0, and a KGE inferior to 0 implies that the simulated streamflow is a worse predictor of the observed flows than taking the mean of the observed values. KGE values above 0.6 are generally considered good, however this is a subjective quantification of the quality of the goodness of fit.

3.4 Evaluation of the ERA5, ERA-I and observed datasets

The next steps following the calibration of the hydrological models on the 3138 catchments were to analyze the raw climate data (precipitation and temperature) at the catchment scale. This analysis was performed by generating the 9 possible arrangements of 3 precipitation and 3 temperature datasets and comparing their relative differences. Then, after performing the model calibration and hydrological simulation steps, the same type of comparison was performed using the calibration KGE metric as a proxy to the quality of the climate dataset. For example, if a certain combination of precipitation and temperature datasets generates higher KGE calibration scores, it is assumed that the climate data are more likely to be accurate than another dataset that returns lower KGE scores.

The various analyses were conducted on the yearly scale as well as for winter (December, January and February, or DJF) and summer (June, July and August, or JJA) seasons. The

results were then analyzed according to their respective catchment locations, climates and sizes in an effort to explain any relationships or differences between the dataset characteristics (i.e. resolution, physics) and their performance (i.e. KGE scores).

4. Results

4.1 Analysis of precipitation and temperature

The first part of the study was to compare precipitation and temperature values averaged at the catchment scale. Figure I-2 shows the mean annual temperatures for the observations, the ERA5 and the ERA-Interim reanalysis products for the catchments in this study (top row). It also shows the mean absolute differences between the datasets for the winter (center row) and summer seasons (bottom row).

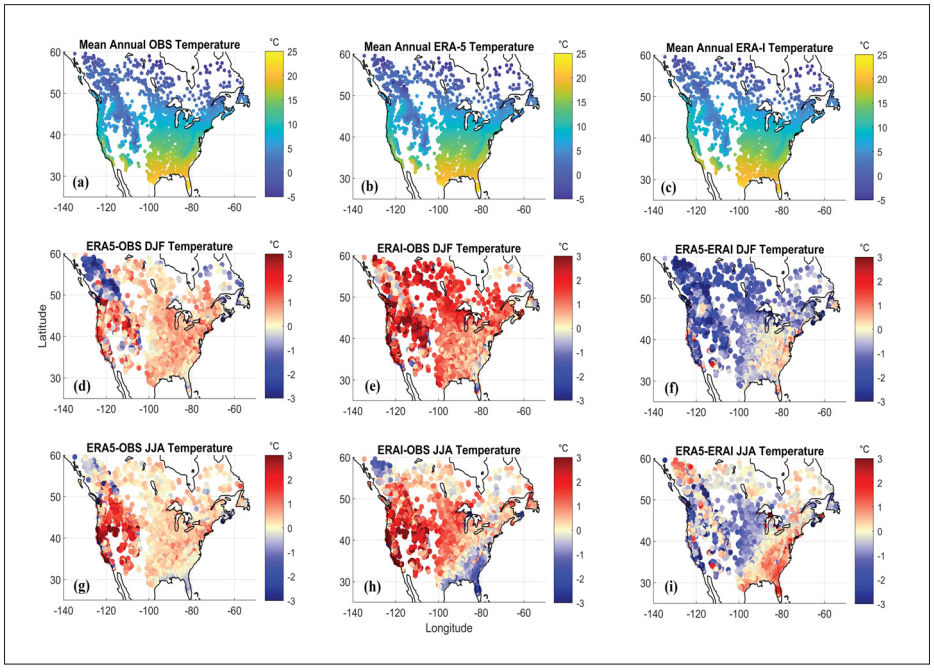


Figure-A I-2 Mean annual temperature for the three datasets (a, b, c), winter (d, e, f) and summer (g, h, i) differences. All values are in degrees Celsius

The results in Figure I-2 are averaged at the catchment scale in order to preserve the consistency between the climate data and the hydrological modelling results presented further in this paper. It can be seen that the ERA-Interim and ERA5 temperatures are generally similar to the observations, although ERA-Interim displays a warm bias almost everywhere except for the southeastern United-States and a few catchments in Canada, where it has a cold bias.

On the other hand, ERA5 sees a strong reduction in biases compared to those in the ERA-Interim dataset. The west coast of North America clearly still shows some important biases of up to 3 °C in summer and -2 °C in winter, although for most catchments the bias amplitude is smaller. It should be noted that most of the large biases are observed in mountainous areas, where observation networks are generally considered less robust. In the panels representing the differences between ERA5 and ERA-Interim in Figure I-2, it can be seen that the ERA5 product corrects the biases in ERA-Interim, i.e. the areas that were too hot in ERA-Interim are colder in ERA5 and vice-versa. The southeast USA was particularly problematic for ERA-Interim in the context of hydrological modelling (Essou *et al.* (2016a)), and it will therefore be explored further with ERA5 in the rest of this study.

The precipitation time series from the three datasets in this study were compared in a similar manner to the temperature data, with Figure I-3 showing the mean annual precipitation for the observations, the ERA5 and the ERA-Interim reanalysis products for the catchments in this study (top row). Figure I-3 also shows the mean absolute differences between the datasets for the winter (center row) and summer seasons (bottom row).

From Figure I-3, it is clear that there is a good representation of mean seasonal and annual precipitation values across the study domain. For winter, it seems that ERA-Interim and ERA5 are very similar as the differences between those datasets are small. One exception is the west coast, where a dry bias persists although it has been reduced in ERA5 as compared to ERA-Interim. For the summer period, there is a strong reduction in biases for the eastern half of the United-States where ERA-Interim was problematic. The dry/wet bias pattern of ERA-Interim is strongly reduced in ERA5. However, both reanalysis products are wet in the North, although

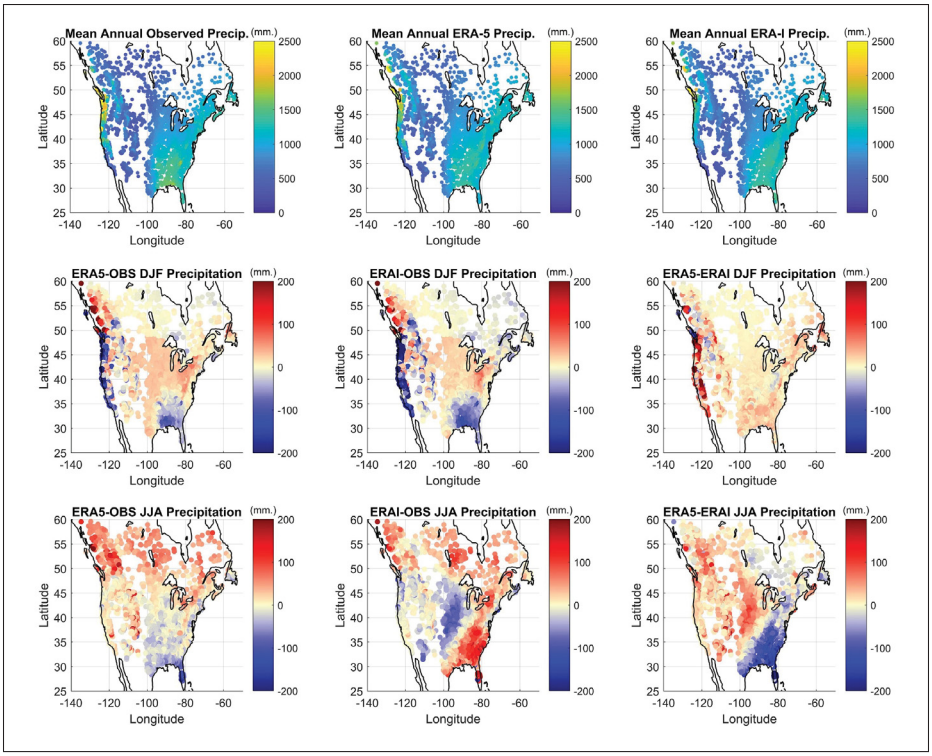


Figure-A I-3 Mean annual precipitation for the three datasets (a, b, c), winter (d, e, f) and summer (g, h, i) differences. All values are in mm.yr^{-1}

as will be discussed in section 5.1, this might be related to the quality of the observation datasets in the remote Northern catchments.

4.2 Hydrological model simulations

The first results obtained in the hydrological modelling portion of this study was the performance of the hydrological models in calibration when driven by the various combinations of precipitation and temperature data. Figure I-4 shows the calibration KGE scores for the HMETS (left panel) and GR4JCN (right panel) for the 9 combinations of precipitation (3 sets) and temperature (3 sets). Each boxplot in Figure I-4 contains the KGE scores of all of the catchments in this study.

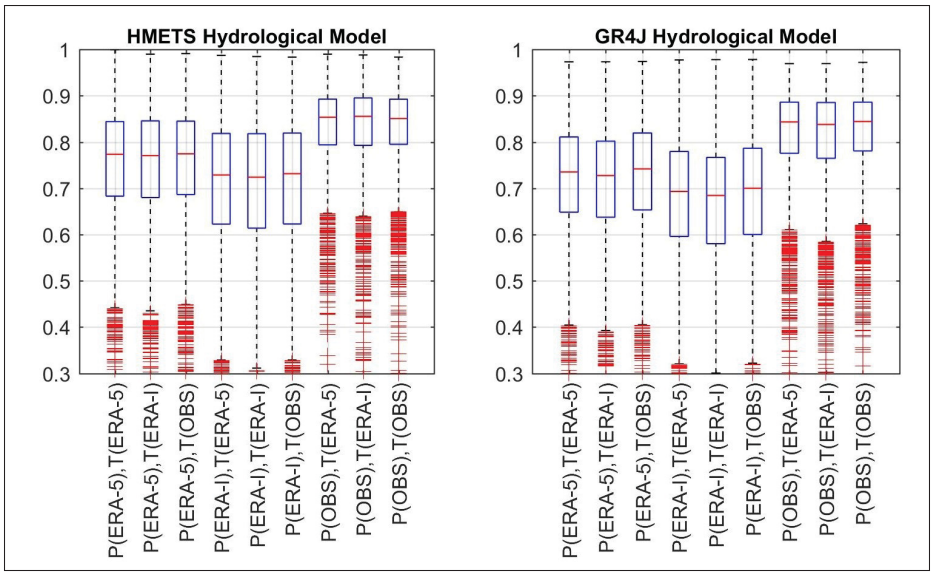


Figure-A I-4 Distribution of calibration KGE scores for all watersheds as a function of meteorological inputs for (a) HMETS and (b) GR4JCN

From Figure I-4, it seems clear that the observations remain the best source of precipitation data for hydrological modelling. It is clear that for hydrological modelling, the ERA5 dataset is a net improvement over the ERA-Interim reanalysis ranking second after the observations. For the catchments in this study, using ERA5 precipitation allows reducing the median gap between the older ERA-Interim reanalysis and the observations by approximately 40%. The precipitation data is the main driver behind the differences observed between the datasets as it can also be seen that the variability linked to the temperature dataset is minimal.

Regarding temperature, ERA5 and the observations provide very similar results, whereas ERA-Interim temperature lags slightly behind. In this sense, the temperature data from ERA5 is marginally more accurate for hydrological modelling at the catchment scale than ERA-Interim, and is similar to that of the observed temperature dataset.

From Figure I-4, it is also interesting to note that the hydrological models respond similarly to the various inputs, indicating that the improvements seen with ERA5 are due to the dataset rather than the choice of hydrological model. In general, it can also be seen that HMETS

performs better than GR4JCN when using the reanalysis datasets (with a median 0.04 KGE improvement) that is modest but statistically significant using a Kruskal-Wallis nonparametric test (McKight & Najab (2010)). HMETS and GR4JCN are statistically equivalent in terms of KGE when using the observed meteorological data.

The hydrological modelling KGE metrics were next analyzed with respect to the catchment locations, as seen in figures I-5 and I-6. Figure I-5 presents absolute values of KGE efficiency metrics for all three datasets and both hydrological models. The differences between hydrological models (first vs second row) are generally small, although the better performance of HMETS is particularly clear over the Rocky Mountains, and especially in the case of both reanalyses. Both hydrological models perform similarly when using observations as inputs compared to reanalysis.

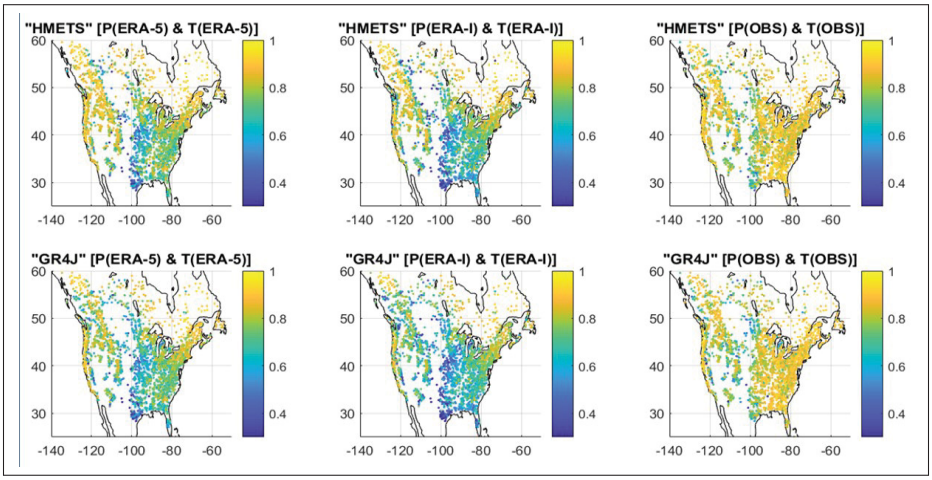


Figure-A I-5 Spatial distribution of Kling–Gupta efficiency metrics for all 3138 watersheds for the HMETS model (a, b, c) and GR4J model (d, e, f), and for ERA5 (a, d), ERA-I (b, e) and observations (c, f)

Focusing on the best performing hydrological model results (first row), two major observations can be made. First, hydrological modeling with observations is clearly superior to using both reanalysis datasets for the eastern part of the US but not so much for Western US and Canada. Second, hydrological modelling performance using ERA5 appears to be consistently superior

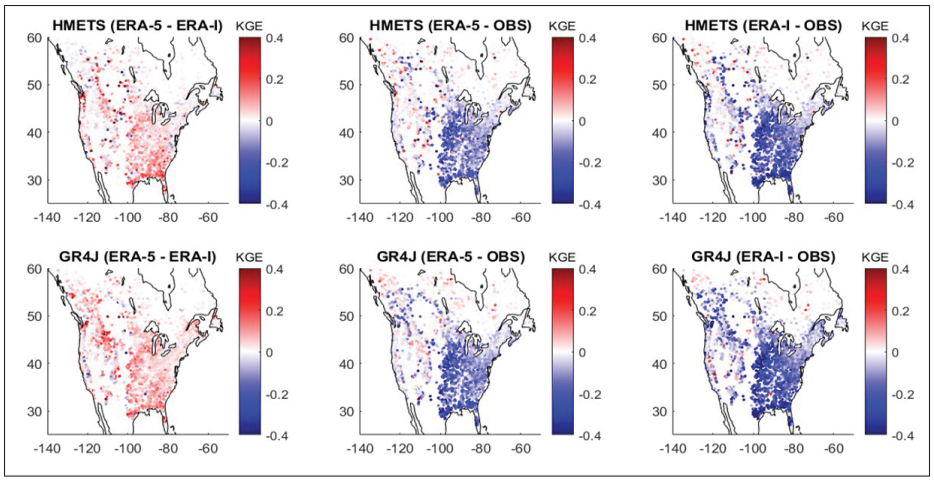


Figure-A I-6 Spatial distribution of the difference of Kling–Gupta efficiency metrics between the three datasets for all 3138 watersheds, for the HMETS model (a, b, c) and the GR4J model (d, e, f)

to ERA-I. To better emphasize these conclusions, Figure I-6 presents differences in KGE efficiency metrics between all three datasets. The maps in Figure I-6 are therefore obtained by subtracting the maps from Figure I-5, two at a time. The middle (ERA5) and right (ERA-I) columns present differences in hydrological modeling performance when using reanalyses compared to observations. A blue colour indicates that observations are superior for hydrological modeling, the reverse being true for red colours. This figure provides a clear view of the spatial patterns of hydrological modeling performance. Observations are clearly superior to reanalyses for the eastern half of the US. This corresponds to the zone with relatively large summer precipitation biases presented earlier in Figure I-3. Outside of this zone, both reanalyses perform similarly to observations, and especially so for ERA5. The left side of Figure I-6 testifies to the uniform and significant improvement in hydrological modeling performance when using ERA5 compared to its predecessor ERA-I.

To gain a better understanding of the reasons behind these observations, hydrological modeling performance was analyzed by looking at watershed size (Figure I-7), elevation (Figure I-8) and climate zone (figures I-9 and I-10). In those three cases, the results are only shown for

the HMETS hydrological model, since the results for GR4J are similar, albeit with a small degradation in modeling performance, as shown in the preceding figures.

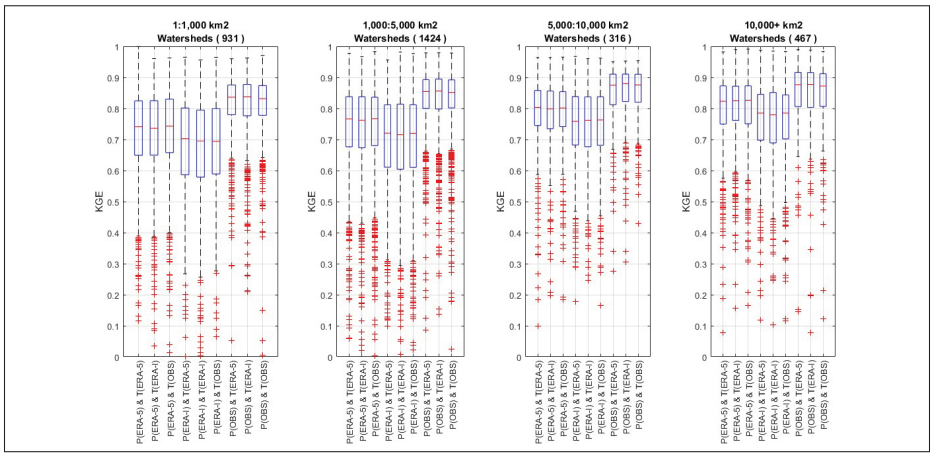


Figure-A I-7 Distribution of the Kling–Gupta efficiency metrics for various watershed surface areas, for HMETS model

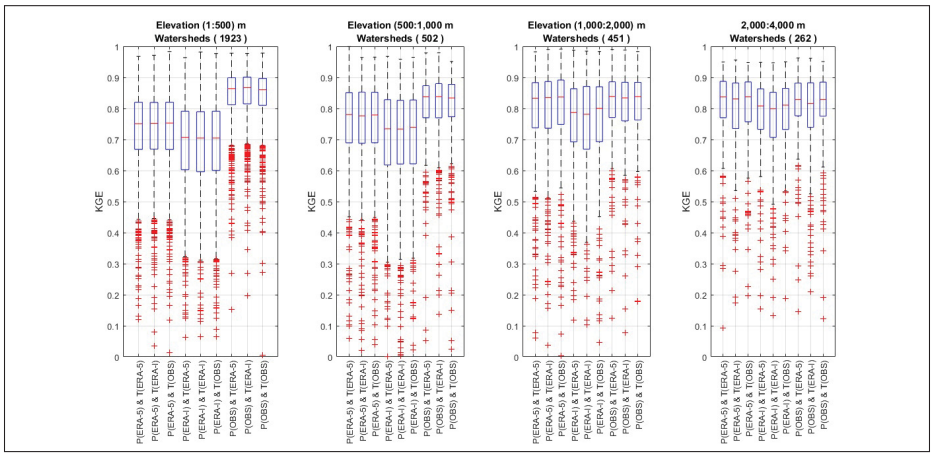


Figure-A I-8 Distribution of the Kling–Gupta efficiency metrics for various elevation bands, for HMETS model

Since all three gridded datasets have different spatial resolutions, Figure I-7 looks at modeling performance for watersheds grouped under 4 different size classes. The patterns are consistent across all four size classes, and similar to those of Figure I-4, with observations being best for all classes, followed by ERA5 and then ERA-I. However, it can be seen that hydrological

modelling performance gets progressively better for larger watersheds for all three datasets. This is particularly clear for both reanalysis. While observations perform better at all scales, the gap with reanalysis gets smaller as catchment size increases. The interquartile range (defined by the solid rectangle of the boxplot) is roughly constant for observations but consistently decreases for both reanalysis. Therefore, a larger proportion of smaller size watersheds are challenging for hydrological modeling than for larger size watersheds. Differences between ERA5 and ERA-I stay constant across all size classes.

Figure I-8 presents the same data but as a function of watershed elevation, separated once again in four classes. Mean watershed elevation is mapped in Figure I-1. Figure I-8 shows a strong dependence of hydrological modeling results on watershed elevation. Observations clearly perform better for the low elevation (< 500 m) watersheds, but differences rapidly shrink with ERA5 actually performing as strongly and even better than observations for the last two elevation classes. It is relevant to stress that over 60% of all watersheds are included in the first elevation class, and that most of the Eastern US watersheds are within the first two elevation classes. Results from Figure I-7 could therefore be influenced by watershed location in addition to elevation. It is also clear that ERA-Interim temperature gets progressively less competitive as the elevation rises, being significantly less efficient than ERA5 and the observations in the high-elevation groups.

The data was finally analyzed by climate zone groupings. Figure I-9 presents North-America's climate classes from the Koppen-Geiger classification (Peel et al., 2007). It can be seen that North-America displays 4 of the 5 main climate zones, with the exception of the Equatorial climate. In total, 13 classes were kept for this analysis. Figure I-10 presents hydrological modeling results for each of those 13 zones.

Results indicate that dataset performance and relative performance strongly depends on the climate zone. This is not surprising since performance was already shown to display spatial patterns. From figures I-9 and I-10, it is apparent that the ERA5 dataset is systematically better than ERA-Interim for all climate zones and that the observations are clearly superior

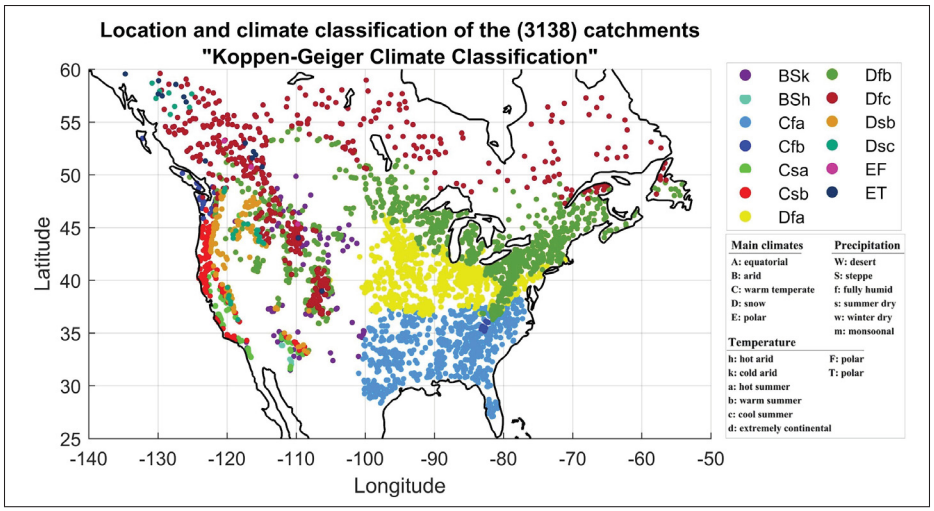


Figure-A I-9 Köppen–Geiger climate classification of the North American watersheds presented in this study

to ERA5 for the Cfa and Dfa climate zones. Elsewhere, the differences are less pronounced. The Cfa and Dfa climate zones are the two main climate zones in the eastern US, which were shown to be problematic for the reanalysis datasets. Furthermore, ERA5 fares better than the observations in the Northern parts of Canada and in the mountainous regions with climate zones Dfc and BSh, respectively. This observation will be discussed further, in section 5.2. Figure I-11 summarizes these results with the use of the Kruskal-Wallis statistical significance test to determine the best dataset for each climate zone. The Kruskal-Wallis hypothesis test is a non-parametric test to evaluate if two samples originate from the same distribution. In Figure I-11, the green, yellow and red colors respectively indicate the best, second best and worst datasets for each climate zone. If two datasets share a color for the same climate zone, the distribution of KGE values is considered to not be statistically different. Results indicate that there are no differences in hydrological modelling performance between ERA5 and observations over 9 of the 13 climate zones. For the other 4 regions (all in the easternUnited States - Bsk, Cfa, Dfa, Dfb), using observations will result in a statistically significant better hydrological modelling performance. ERA-I is the worst performing dataset over 8 climate zones. In the remaining 5 zones: Bsh (3), Csa (53), Dsc (33), EF (3) and ET (15), all three datasets perform identically

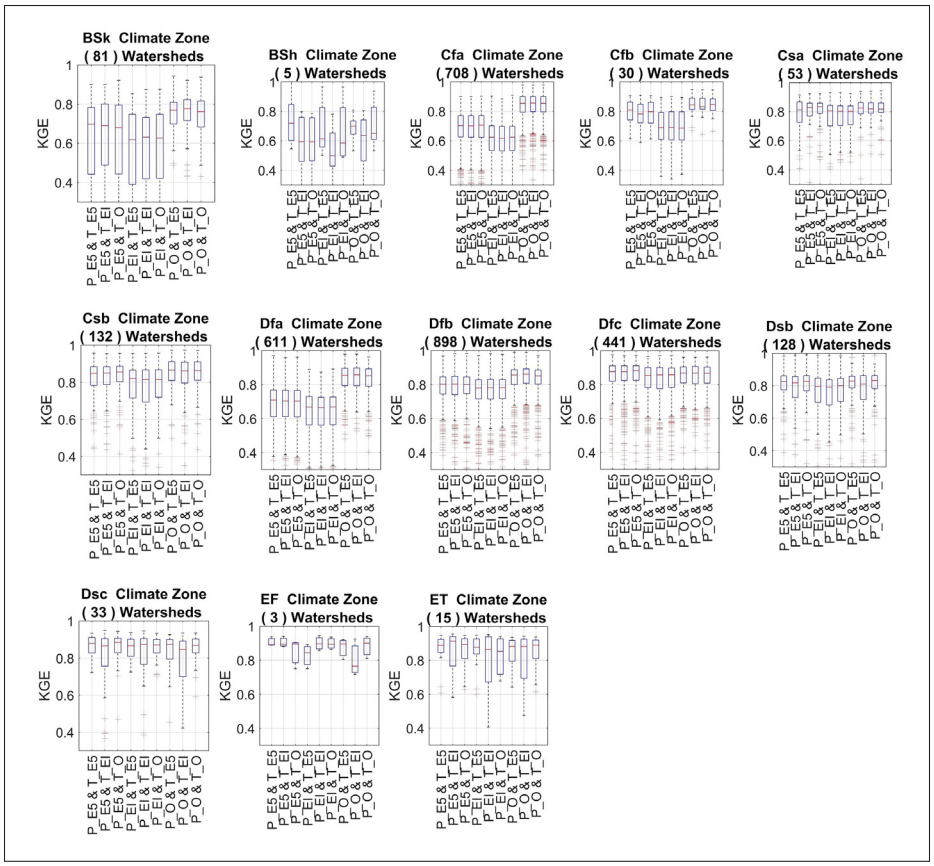


Figure-A I-10 Distribution of the Kling–Gupta efficiency metrics for the 13 climate zones of Fig. 9, for the hydrological model HMETS

from a statistical viewpoint. These zones share in common having the fewest watersheds and most extreme climates (arid and polar).

In order to better explore the differences related to the watershed locations and properties, three catchments of different hydrological regimes were analyzed in depth. Figure I-12 presents the hydrological modelling KGE difference for HMETS between ERA5 and the observation dataset (first column) along with the mean monthly precipitation (second column), mean monthly temperature (third column) and mean annual hydrograph (fourth column). Results are presented for the Ouiska Chitto Creek Near Oberlin, Louisiana USA (First row), the Grande Rivière à la Baleine in Quebec, Canada (center row) and the Cosumnes River at Michigan Bar, California, USA (bottom row). Table I-1 shows summarized statistics for the three catchments.

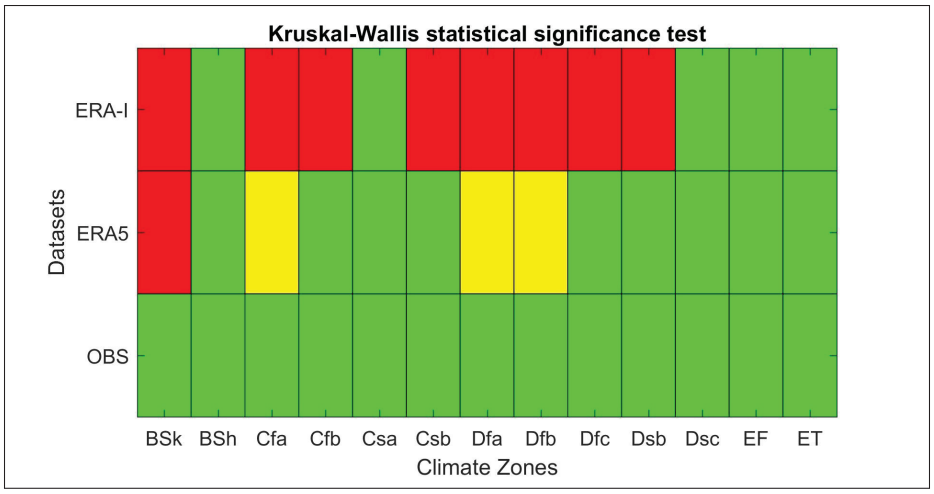


Figure-A I-11 Results of the Kruskal–Wallis statistical significance test to determine the best dataset for hydrological modelling as observed through the KGE metric, for each climate zone. The green, yellow and red colours, respectively, indicate the best, second best and worst datasets for each climate zone

Table-A I-1 Main properties of the study basins

Catchment	**Outlet Lat.**	**Outlet Long.**	**Elevation (m)**	**Area (km^2)**	**KGE in calibration**		
					ERA5 dataset	**ERA-I dataset**	**OBS dataset**
Ouiska Chitto (Southeast USA)	30.93°	-92.98°	53	1320	0.65	0.49	0.87
Grande Baleine (Northern Canada)	55.08°	-73.10°	389	36300	0.94	0.94	0.92
Cosumnes River (Western USA)	38.60°	-120.68°	696	1388	0.87	0.83	0.90

The first row in Figure I-12 presents a catchment in the southeastern United-States, which is a region in which the reanalysis-driven hydrological models are unable to perform as well as the observation-driven models. ERA-Interim has a clear precipitation seasonality problem, being too dry except for the summer months where there is a large overestimation of precipitation compared to the observations. This seasonality problem is mostly solved by ERA5, but a dry bias persists all year, as shown in Figure I-3. The temperatures between the three datasets are practically identical, which means that evapotranspiration should be relatively constant

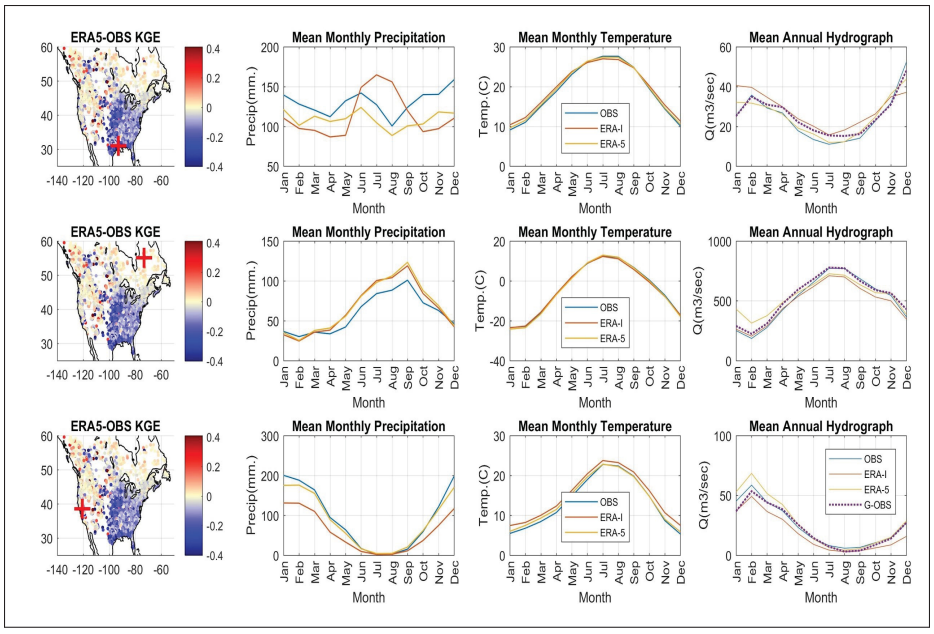


Figure-A I-12 Difference in hydrological modelling performance, mean monthly precipitation and temperature and mean annual hydrograph using ERA-I and ERA5 observations (OBS) and streamflow observations (G-OBS) on three dissimilar catchments: Ouiska Chitto Creek (top row), Grande Rivière à la Baleine (centre row) and Cosumnes River (bottom row)

between the products. The lack of precipitation should therefore become apparent in the simulated hydrograph, however the streamflow is higher for ERA5 than for the observations when the opposite would normally be expected. It is important to note that the hydrological model can adapt its mass balance by adjusting the potential evapotranspiration scaling, which it has clearly done in this case. The difference in hydrological modelling then comes from the temporal distribution of precipitation, and it can be seen that the ERA5 winter precipitations are relatively lower in winter than for the rest of the year. The PET scaling therefore attempts to reduce evaporation for the entire year but does not compensate enough to account for this difference in winter. Indeed, it can be seen that the observed hydrograph is underestimated by ERA5 and ERA-Interim for that period in the southeastern United-States.

The second catchment is located in Northern Quebec, Canada, and as such is in a remote and sparsely gauged region. In this case, it can be seen that the ERA5-driven KGE metric is

superior to that obtained using the observations. One key difference between the reanalysis and observed datasets is the precipitation, where ERA5 and ERA-Interim both show more precipitation than the observations. Again, the temperatures are practically identical, meaning that the potential evapotranspiration, although weak in that region, are very similar. The mean annual hydrograph is also very similar between ERA-Interim and the observations, but it can be seen that the ERA5 model overestimates streamflow in winter while matching the snowmelt peak flows more closely than the other datasets. The difference in KGE in this case comes from a better matching of peak flows, which counts more heavily towards the KGE than the low-flows.

The third catchment, located in the west, is characterized by large precipitation systems in fall and winter, with a months-long dry spell in summer. ERA5 mostly corrected ERA-Interims' strong underestimation of precipitation for that catchment, as is the case for most West-coast catchments as seen in Figure I-3. ERA5 temperatures are slightly cooler and are more in-line with the observations. In terms of hydrological modelling, ERA-Interim underestimates the average streamflows year-round while ERA5 slightly overestimates them in winter. As seen in Table I-1, the ERA5 dataset managed to improve the KGE from 0.83 (ERA-Interim) to 0.87, as compared to the reference of 0.90 obtained with the observed data. The improvements in precipitation in ERA5 for this region thus seem to translate to improved hydrological modelling compared to using ERA-Interim, which confirms the findings of Figure I-6.

5. Discussion

This study aims to evaluate the ERA5 reanalysis product as a potential reference dataset for hydrological modelling. The ERA5 reanalysis was compared to the ERA-Interim and observation datasets when used in two hydrological models covering 3138 catchments in North America. This section aims to analyze and explain the results obtained in light of the literature and properties of the ERA5 reanalysis. First, differences in climate and hydrological data will be investigated, followed by an analysis based on climate classifications and catchment size. Finally, limitations of the study and recommendations for future work will be provided.

5.1 Differences in temperature and precipitation between the ERA5, ERA-I and observation datasets

In this study, the observations are taken as the reference dataset and ERA5 is compared to both the observations and ERA-Interim. This allows validating both the improvement in ERA5 with respect to ERA-Interim, as well as evaluating the possibility of using ERA5 reanalysis data as inputs to hydrological models to overcome potential deficiencies of observation networks, related to either quality and/or availability.

The evaluation of ERA5 temperature and precipitation variables compared to ERA-Interim and the observation datasets showed that ERA5 systematically reduced biases present in ERA-Interim for the temperature variables, whereas precipitation was generally also less biased, although to a lesser degree. There are remaining precipitation biases on the West coast of North-America with ERA5, but from Figure I-2 it can bee seen that the scale of these biases is dependent on the season. In the Southeast United-States, ERA5 largely corrects biases that were present in ERA-Interim dataset and led to relatively poor hydrological modelling in a few studies (e.g. Essou *et al.* (2016a)). As for temperature, Figure I-2 shows that summer temperatures in ERA5 are mostly too high for the catchments west of the Rocky Mountains but are improved over the ERA-Interim data. There is also an interesting pattern of biases between the East and West coasts (figures I-2 and I-3), which could be partly explained by some processes not being accounted for in ERA5, notably the high-amplitude ridge trough wave patterns which have seen a recent increase allowing severe weather in both the East and West simultaneously (Singh *et al.* (2016); Raymond *et al.* (2017)), although ERA5 did improve the representation of many processes since ERA-I (Hoffmann *et al.* (2019)).

It is important to note that these perceived biases suppose that the observation data is perfect. In reality, at the catchment scale, one would expect that the observations would be far from perfect and contain errors due to location representativeness, precipitation undercatch, and missing data due to station malfunction or instrument replacement, for example. However, the observation data are the best estimates available which makes them the de facto reference dataset. This means that although figures I-2 and I-3 show ERA5 and ERA-Interim as con-

taining some important biases on western North America, it is possible that these biases are caused by biases in the station data relative to the catchment size. The reanalysis products also have the advantage of being driven by spatialized sources such as satellites, which can help in estimating precipitation and temperature data in regions where the weather station network is deficient or sparse.

5.2 Differences in hydrological simulations using ERA5, ERA-I and observation data as inputs to hydrological models

One way to evaluate the quality of the observation and reanalysis data is to use hydrological models as integrators to compare simulated and observed streamflow, which can act as an independent validation variable. In an attempt to independently assess precipitation and temperature data for each dataset, all possible combinations of precipitation and temperature were fed to two hydrological models, which were then calibrated for each combination. This was to remove any bias caused by parameter sets calibrated on one single dataset, which would obviously be favored in the resulting analysis. As was the case for the climatological variables, the observed streamflows act as the reference hydrometric data and are considered as unbiased. Of course, in reality streamflow gauges contain various sources of errors (Di Baldassarre & Montanari (2009)), but for this study they are the best available estimates. This hypothesis could have a small effect on the conclusions of this study. For example, if a certain combination of precipitation and temperature datasets generate higher KGE calibration scores, it is assumed that the climate data are more likely to be correct than another dataset that returns lower KGE scores. This could be incorrect in some instances, where the error actually comes from the streamflow data; however, on average over the 3138 catchments this effect should not influence the results.

The results in Figure I-4 showed that the hydrological models driven with the observed precipitation generally provide the most representative simulated hydrographs, with KGE values exceeding those of the ERA5-precipitation driven hydrological models by 0.1 on average, which is a significant difference. ERA5 precipitation is also shown to be clearly better than ERA-

Interim precipitation on average for the catchments in this study. Another interesting aspect is that in Figure I-4, replacing observed temperatures with ERA5 temperatures marginally improves the hydrological modelling skill. While not a significant difference, this attests to the quality of the ERA5 temperatures in general for hydrological modelling. Therefore, the differences observed in the hydrological modelling performance are almost entirely due to the precipitation data quality. The rest of this study will thus focus on the precipitation and hydrological modelling and forego further analysis on temperature data.

Also of note is that in general, ERA5-driven hydrological simulations are less skillful than those driven by observations. However, there are some catchments - mostly in the mountainous regions of western United-States and in Northern Canada - where use of ERA5 leads to improved hydrological simulations. This is probably due to the difficulty in installing weather stations and obtaining representative observation data in those regions, but it shows that reanalysis data can be used as a replacement to observations for hydrological modelling in these regions, as previously reported by Essou *et al.* (2016a)).

The more detailed spatial (Figure I-6) and climate zone (figures I-10 and I-11) analysis outlined the strong spatial dependence on dataset performance. Observations clearly outperformed ERA5 over the Eastern half of the US, where a larger portion of the watersheds used in this study are located. To illustrate this point, Figure I-13 presents modelling performance over the Eastern US (grouping climate zones Cfa, Dfa, and Dfb) against that of the other 10 climate zones.

Figure I-13 paints a much different picture than Figure I-6 since it shows that hydrological modeling with ERA-5 precipitation and temperature is as good as observations everywhere in North-America, with the exception of the Eastern US. The disproportionate number of watersheds in this region may overemphasize the performance differential between ERA5 and observations as seen in Figure I-6. An interesting fact is that the Eastern US is the North-American region having by far the highest density of weather stations, as reported by Janis *et al.* (2002). Theoretically, this could explain why observation-based modeling performs bet-

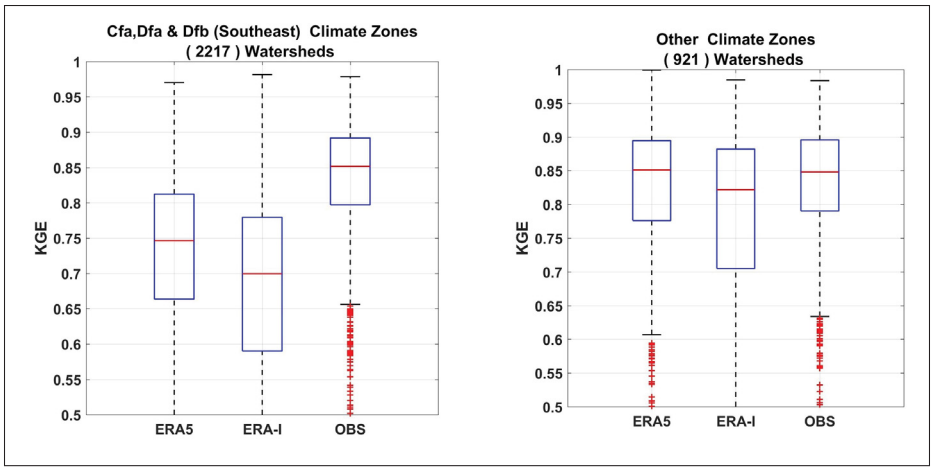


Figure-A I-13 Distribution of the Kling–Gupta efficiency metrics for the 3 north-eastern US climate zones (Cfa, Dfa, Dfb) and for all the other 10 climate zones grouped together, for hydrological model HMETS

ter in this region. However, Figure I-13 shows that observation-based modelling performance is not different in the other regions, whereas reanalysis-based modeling clearly suffer over the Eastern US. This was also noted in Essou *et al.* (2016a). It could mean that reanalysis face a harder challenge in the Eastern US, further away from the Pacific Ocean control on atmospheric circulation. A large proportion of summer and fall precipitation in these zones come from convective storms. Eastern Canadian watersheds are well modelled using reanalysis, but the hydrological behaviour of most of those watersheds is dominated by the spring flood which is largely controlled by temperature, which is very well reproduced by both reanalysis.

Alternatively, this could also mean that Eastern US watersheds are in fact more difficult to hydrologically model and that differences are therefore directly linked to network density. Equal performance of ERA5 and observations elsewhere would therefore be the result of the improved process representation of ERA5 coupled with some degradation of observations due to the gridded interpolation process between more distant stations. As discussed below, a more precise investigation of modeling performance as a function of station density could shed light on this issue.

5.3 Differences between the HMETS and GR4J hydrological models

In this study, two hydrological models were selected to perform the hydrological evaluation of the reanalysis and observation datasets. While both models are conceptually similar, GR4J is simpler than HMETS (two routing processes instead of four, non-scalable PET, much simpler snow model, less than half the number of parameters, etc.). They were shown to perform generally well over all climate zones represented by the catchments used in this study, as can be seen in Figure I-4. Interestingly, both GR4J and HMETS return similar results for any given driving climate dataset. HMETS performs slightly better than GR4J almost everywhere, although that can be attributed to its more flexible model structure and parameterizations that can better adapt to various hydrological conditions.

Since the main objective of this study was to evaluate the ERA5 dataset for hydrological modelling, the interest is not to compare the hydrological model performances, but to compare the ERA5-driven simulations to the others for each model. In both cases, as can be seen in figures I-4, I-6 and I-8, ERA5-driven hydrological models clearly outperforms the ERA-Interim-driven models, which shows that the precipitation scheme in ERA5 is superior to that in ERA-Interim for hydrological modelling purposes. As stated in section 5.2, temperature seems to play only a minor role in the differences in hydrological modelling.

Furthermore, the observation-driven hydrological models generally perform better than the ERA5-driven models, which confirms that station data should be prioritized when possible. The main caveat to this point is that when the observation station network is of poor quality or too sparse, then ERA5 can be used to fill the voids and get an acceptable hydrological response, as discussed in section 5.2.

5.4 Analysis of the impacts of catchment size and elevation on the hydrological simulation performance using the ERA-I and ERA5 reanalyses

One of the major differences between ERA-Interim and ERA5 is the horizontal resolution, improving from 79km to 31km. This finer resolution should allow for more precise estimations of precipitations and temperatures over smaller catchments that were not adequately represented by ERA-Interim. This logic should apply even though the hydrological models are lumped models. Larger catchments could also see some improvements, namely in a better estimation of the terrain elevation, but it is expected that the gain would not be as large as for smaller catchments.

In order to test this hypothesis, the improvements between ERA5 and ERA-Interim in hydrological modelling were sorted according to catchment size, as shown in Figure I-7. It is clear from Figure I-7 that the catchment size is not a good predictor of hydrological simulation improvement. While most catchments see improvements with ERA5 over ERA-Interim, the catchment size does not seem to affect the rate of improvement. This suggests that the improvements do not come from the higher spatial resolution, lending credence to the hypothesis that the enhancements are due to ERA5's improved physics and process representations.

A similar analysis was performed to evaluate the impact of catchment elevation on hydrological modelling skill. It can be seen from Figure I-8 that the elevation plays a significant role in the hydrological model's ability to estimate streamflow. For example, the median and interquartile ranges increase for all datasets as elevation increases. This could be caused by a more rapid hydrological response in higher-elevation and steeper catchments, compared to the slow runoff schemes often found in flat lowlands. The hydrological models being lumped models could contribute to this as large and flat catchments would be more affected by the location of rainfall events compared to steeper ones, especially in the timing of the hydrograph peaks. For the Northern catchments, the peaks are caused by snowmelt which is much more uniform than rainfall events, which would minimize this effect.

Another, a more probable reason for the reanalysis datasets being stronger in mountainous regions is simply because there are fewer weather stations set up in those areas due to difficulties in accessing and maintaining them. The density of weather stations in the eastern part of the US is typically at least twice as large than for the western part (Janis *et al.* (2002)). In such cases, a reanalysis would provide information that is not conveyed by station data, making it a de-facto best estimation of precipitation. In essence, the ERA5 data are not yet as accurate as observations, however they are able to perform very well in their absence.

Finally, in all the analyzed scenarios in this study, ERA5 has always been either at least as good as ERA-Interim in terms of hydrological performance. The same is true for the precipitations and temperatures at the catchment scale. From all the results in this study, there does not seem to be any reason or indication that ERA-Interim should continue to be used for hydrological modelling applications, at least in North America. This is not to say that ERA5 is perfect, but it should become the reference for the time being.

5.5 Limitations

As is the case with any large-scale comparison studies, some methodological limitations may potentially impact conclusions drawn from the presented results. In terms of hydrological modeling, this study only uses two lumped conceptual models and one flow criteria (KGE). Both models are lumped, which limits the assessment of the horizontal resolution component of the three datasets. This aspect was however indirectly assessed by looking at the impact of watershed size. Both hydrological models are conceptually similar but HMETS is more flexible and has more hydrological processes (and parameters). Accordingly, this study was able to look at the impact of parametric space flexibility in dealing with various datasets biases, but not at other issues such as the impact of physically-based processes and distributed inputs. A study looking at the latter points would require more complex hydrological models, but at the expense of having to look at much fewer watersheds.

The single streamflow criteria and objective function (KGE), like its Nash-Sutcliffe relative, is weighted towards higher flow events. Other objective functions would return different results, however the fact that ERA5 climate data is generally improved in all areas is an indicator that other metrics could potentially see improved results as well, although no test has been performed to that effect in this study. There are several other streamflow criteria which could shed light on differences between datasets, such as extremes. In particular, high flow extremes have the potential to outline improvements in ERA5 compared to its predecessor ERA-I because of improved resolution and processes. Low flows may also be of interest, although the are typically less well-modelled by conceptual hydrological models, and more strongly dependent on temperature, which is very comparable across all three datasets. Finally, there are now several potential other precipitation datasets that could have been included in the comparison (see for example Beck *et al.* (2017b)). However, the goal of this work was a first evaluation of the 1979-2019 ERA5 dataset, because of the potential linked to its spatial and temporal resolutions.

5.6 Recommendations

One of the main reasons for the interest in the ERA5 reanalysis resides with its hourly temporal resolution. Therefore, the obvious next step is to investigate sub-daily components, and particularly for precipitation. Sub-daily precipitation is key to investigating the hydrological response of smaller watersheds. However, sub-daily studies raise another set of challenges, notably the absence of a robust baseline hourly meteorological dataset. MSWEP (Beck *et al.* (2017a)) is the best potential candidate at the sub-daily time scale (3-hourly), but the reliability of its sub-daily component is largely unknown. Reliance on hourly weather station data will therefore be required, meaning additional problems including having to deal with missing data.

The differences noted in Eastern USA raised the question of the potential impact of the density of the station network on the absolute and relative performance of the various datasets. This could be better studied by assigning a network density index to each watershed. This could ultimately lead to a better understanding of the role of station density, and provide guidance

on network improvements or rationalization. It could also be envisioned to extend this work to underdeveloped countries where there is a fewer number of observational gauges, where a good quality reanalysis might allow for improved hydrological simulations and better understanding of the regional weather characteristics.

The hydrological performance of ERA5 opens specific avenues of research for streamflow forecasting using ECMWF forecasts. Calibrating hydrological models with ERA5 data could potentially reduce streamflow forecasts biases since the reanalysis and forecasts essentially originate from the same model.

6. Conclusion

The main objective of this study was to evaluate the ERA5 reanalysis as a potential reference dataset for hydrological modeling over North-America, by performing a large-scale hydrological modelling study using ERA5, ERA-Interim and observations as forcing data to two hydrological models. The first assessment showed that ERA5 precipitation and temperature data were greatly improved compared to its predecessor ERA-Interim, although some significant biases remain in the southeast United-States and North-American West coast. These improvements were then shown to translate well to the hydrological modelling results, where both hydrological models showed significant increases in skill with ERA5 as opposed to ERA-Interim. In all cases, ERA5 was consistently better than ERA-Interim for hydrological modelling, and as good as observations over most of North-America, with the exception of the Eastern half the US. The lesser performance of reanalyses in this region may reflect some deficiencies at representing precipitation seasonality accurately, and may also result from the higher-density network over Eastern USA, thus favoring observations, or a combination thereof. We also showed that the catchment size did not impact the hydrological modelling performance, thus the improvements are not linked to ERA5's model resolution but to its improved internal physics and assimilation. While some limitations apply to ERA5, it seems that this reanalysis is significantly improved compared to ERA-I and that is should definitely be considered as a high-

potential dataset for hydrological modelling in regions where observations are lacking either in number or in quality.

Future work should focus on evaluating the sub-daily performance of hydrological modelling with ERA5, testing its quality on other continents, integrating ERA5-based model calibration for hydrological forecasting applications and evaluating its potential for weather network augmentation and rationalization.

Finally, it is important to state that this paper does not advocate for the replacement of observed data from weather stations by products such as reanalysis, nor should it be interpreted as providing justification to pursue the current trend of decommissioning additional stations. Weather stations will continue to provide the best estimate of surface weather data at the local and regional scales and there are many fundamental reasons to keep on supporting a strong network of quality weather stations. The results provided in this study for ERA5 show that atmospheric reanalysis have likely reached the point where they can reliably complement observations from weather stations, and provide reliable proxies in regions with less dense station networks, at least over North America.

7. Code and data availability

The gridded observed weather data was downloaded from the Santa Clara repository, available here: http://hydro.engr.scu.edu/files/gridded_obs/daily/ncfiles_2010.

The Canopex climate and streamflow data can be downloaded from the official data repository available here: http://canopex.etsmtl.net/.

The USGS streamflow data can be downloaded from the USGS Water Data for the Nation repository, available here: https://waterdata.usgs.gov/nwis/sw.

ERA-Interim data are available through the ECMWF servers at: https://apps.ecmwf.int/datasets/data/interim-full-daily/.

ERA5 data is available on the Copernicus Climate Change Service (C3S) Climate Data Store: https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset.

The HMETS hydrological model is available on the Matlab File Exchange: https://www.mathworks.com/matlabcentral/fileexchange/48069-hmets-hydrological-model.

Finally, the GR4J model and Cemaneige snow module are made available by the IRSTEA: https://webgr.irstea.fr/en/models/.

APPENDIX II

LARGE-SCALE ANALYSIS OF GLOBAL GRIDDED PRECIPITATION AND TEMPERATURE DATASETS FOR CLIMATE CHANGE IMPACT STUDIES

Mostafa Tarek[1,2], François P. Brissette[1], Richard Arsenault[1]

[1] Departement of Civil Engineering, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Québec, Canada, H3C 1K3

[2] Department of Civil Engineering, Military Technical College, Cairo, Egypt.

Paper accepted in *Journal of Hydrometeorology*, September 2020

Abstract. Currently, there are a large number of diverse climate datasets in existence, which differ, sometimes greatly, in terms of their data sources, quality control schemes, estimation procedures, and spatial and temporal resolutions. Choosing an appropriate dataset for a given application is therefore not a simple task. This study compares nine global/near-global precipitation and three global temperature datasets over 3138 North American catchments. The chosen datasets all meet the minimum requirement of having at least 30-years of available data, so they could all potentially be used as reference datasets for climate change impact studies. The precipitation datasets include two gauged-only products (GPCC and CPC-Unified), two satellite products corrected using ground-based observations (CHIRPS V2.0 and PERSIANN-CDR V1R1), four reanalysis products (NCEP-CFSR, JRA55, ERA-Interim and ERA5) and one merged product (MSWEP V1.2). The temperature datasets include one gauge-based (CPC-Unified) and two reanalysis (ERA-Interim and ERA5) products. High-resolution gauge-based gridded precipitation and temperature datasets were combined as the reference dataset for this inter-comparison study. To assess dataset performance, all combinations were used as inputs to a lumped hydrological model. The results showed that all temperature datasets performed similarly, albeit with the CPC performance being systematically inferior to that of the other three. Significant differences in performance were, however, observed between the precipitation datasets. The MSWEP dataset performed best, followed by the gauge-based, reanalysis and satellite datasets categories. Results also showed that gauge-based datasets should be pre-

ferred in regions with good weather network density, but CHIRPS and ERA5 would be good alternatives in data sparse regions.

1. Introduction

Climate change impact assessments require future climate scenarios developed at adequately high spatial and temporal resolutions. Although General Circulation Models (GCMs) are typically sourced for future climate projections, their spatial resolution is often too coarse for fine-scale climate studies (Ahmed *et al.* (2013)). Consequently, spatial downscaling and/or bias correction approaches are normally used to bring bias-corrected GCM simulation information to the appropriate scale before it is used in impact models. This procedure requires a reference climate dataset generally consisting of precipitation and temperature data to ensure that the downscaling and/or bias correcting steps preserve the main characteristics of the reference climate at the finer scale. An often limiting additional requirement of reference datasets is that they should cover a time-period long-enough to filter out high-frequency internal variability (e.g. Deser *et al.* (2012)). A 30-year period is generally favored as defined by the World Meteorological Organization, even though periods as short as 20 years have been used in the literature (e.g. Martel *et al.* (2018)). It is important that the chosen dataset represent the true state of the reference climate as closely as possible, since any deficiency in the reference dataset would be transferred to the future climate scenario.

Notwithstanding the limitations associated with meteorological stations as reference datasets, such as missing records, inhomogeneity, short temporal coverage, sparse spatial coverage and the inability to adequately represent the climate variability in all topographic and climatic zones, the stations are still considered to constitute the most accurate source of climate data (Tapiador *et al.* (2012); Nicholson (2013); Colston *et al.* (2018)). In recent decades, to overcome some of the limitations of station data, several global and regional gridded datasets have been developed with different input data sources (gauges, radar, satellite, reanalysis or combinations thereof), spatial resolutions (0.05° to 2.5°), spatial coverage (continental to global), temporal scales (30 minutes to annual) and temporal coverage (from 1 to several years) (Henn

et al. (2018)). Such gridded datasets provide continuous spatial and temporal coverage, and typically, with no missing data.

Gridded datasets can be classified as a function of their data source. Gauge-based gridded datasets are obtained by interpolating and mapping the information measured at a small scale (typically, a point measurement at a weather station) onto a predefined spatial and temporal resolution grid. However, variations in gauge types or instrument replacements affect error characteristics in long-term records. In addition, observations are affected by systematic biases due to evaporation and wind effects, as well as the elevation of gauges in mountainous regions (Isotta *et al.* (2014)). Gauges are also typically placed in regions allowing easier access for station maintenance and troubleshooting, meaning that the gauges do not necessarily reflect the actual climatic conditions of their surroundings. Interpolated station gridded climate data products are thus subject to these limitations and many integrate adiabatic lapse rates and elevation/precipitation relationships using terrain elevations in a bid to correct some of these shortcomings.

A different approach to measuring precipitation uses ground weather radars, as they partially address the issue of rain gauge spatial coverage since each radar site covers a relatively large area. However, radar coverage here is limited to high population developed regions. In addition, they provide estimates of the rainfall rate at certain observational levels above the ground and cannot detect surface precipitation. Therefore, the presence of weather stations is required for the calibration and correction processes between surface measurements and atmospheric precipitation estimates (Martens *et al.* (2013)).

Nowadays, satellite products are available at the global scale, and can cover large areas at high spatial and temporal resolutions with near real-time coverage. They are mainly suitable for rainfall estimation in the tropics and data-scarce regions. Given this advantage, satellite products have been used in water resource management studies (Giardino *et al.* (2010); Nishat & Rahman (2009); Siddique-E-Akbor *et al.* (2014)), hydroclimatological studies (Khan *et al.* (2011); Jutla *et al.* (2015)) and in extreme event analysis (Lockhoff *et al.* (2014); Boers

et al. (2015)). However, satellites are relatively insensitive and generally miss a significant quantity of light precipitation and tend to fail over snow- and ice-covered surfaces (Tian *et al.* (2009)). Some studies have evaluated the uncertainties of these datasets and shown that high-resolution satellite products perform better when bias is corrected using gauge observations (Awange *et al.* (2016)).

Retrospective analysis/reanalysis systems represent vital sources of data in weather and climate studies. A typical reanalysis system consists of two main components: the forecast model and the data assimilation system. The role of the data assimilation system is to integrate many sources of observations to provide the forecast model with the most accurate representation of initial atmospheric states. Then, the numerical weather forecast models are executed for a given time-step to produce consistent gridded datasets (Di Luzio *et al.* (2008)). Although reanalysis are not direct observations, they nevertheless provide analyzed variables, even in areas with minimal or nonexistent stations (Bosilovich (2013)).

Overall, no single precipitation product could be considered ideal for measuring precipitation. In fact, all precipitation products tend to miss a significant volume of rainfall (Behrangi *et al.* (2011)).

Near-surface air temperature is a key variable for meteorological monitoring and forecasting services (Nieto *et al.* (2011)), as well as for climate and hydrological studies. In hydrological modelling, the air temperature is the main driving variable for the evapotranspiration and snowmelt processes. Hence, accurate temperature data is vital when driving hydrological models in historical and future climate periods. However, the lack of an adequate gauge network can result in improper temperature estimations. Therefore, gridded temperature datasets are also crucial in many fields. Temperature products are generally thought to be less complex than precipitation datasets due to the much smaller spatial and temporal temperature variability in the former. Therefore, significantly fewer studies have compared and evaluated the uncertainty of using different temperature datasets in hydrological impact models.

Appropriate dataset selection is a key issue in climate studies. High uncertainty is found across most gridded datasets, coming from multiple sources, such as using different data sources, merging and interpolation algorithms or quality control techniques (Vogel & Vogel (2013); Prakash *et al.* (2015b); Prakash *et al.* (2015a); Prein & Gobiet (2017); Nashwan *et al.* (2019)). Moreover, the number and the accuracy of observations used to correct these products typically vary. However, some products are calibrated to the observations, thus making annual biases minimal, while their daily patterns are significantly different from the observations (Sylla *et al.* (2013)). Therefore, gridded datasets should be comprehensively evaluated before they are used.

Several studies have assessed the performance, advantages and limitations of gridded datasets. Most of these studies focus solely on precipitation datasets and evaluate the accuracy of these products through a straightforward comparison against ground weather stations (Vila *et al.* (2009); Romilly *et al.* (2011); Jiang *et al.* (2012); Prakash *et al.* (2018); Nashwan *et al.* (2019)). Andermann *et al.* (2011) evaluated five remote sensing and gauge-based gridded datasets with ground-based measurements in the Himalayan region. The results showed that the satellite products underestimate the precipitation at both the annual and seasonal scales. The authors reported that the findings likely resulted from the bias correction techniques applied to correct the datasets using the Global Telecommunication System (GTS) rain gauge network, which has a poor spatial coverage in the study region; in addition, 0 mm precipitation is used to compensate for missing values in the database. Moreover, there is a lag experienced by the remote sensors in precisely capturing the snowfall, which is the major contributor of precipitation in the Himalayas. The conclusion that satellite approaches tend to fail in snow-dominant regions has also been reported in other studies (Kidd *et al.* (2012); Laviola *et al.* (2013)). Chen *et al.* (2014) also evaluated two satellite-based products, CMORPH (Joyce *et al.* (2004)) and PERSIANN-CCS (Ashouri *et al.* (2015)), to capture the rainfall in the mountainous zones located west and north of Beijing. The study showed that both datasets failed to capture the spatial pattern and the temporal variation of precipitation.

Others studies have used hydrological modelling as an indirect method to evaluate the performance of these datasets in forcing hydrological models (Zhu *et al.* (2018); Duan *et al.* (2019);

Tarek *et al.* (2019)). Hydrological modelling offers an interesting perspective since results depend on the coherence between precipitation and temperature datasets and on an accurate representation of the annual cycle of both variables. Hydrological modelling is also not overly sensitive to biases present in every dataset, as these are typically removed during the calibration process (Essou *et al.* (2016a)). Behrangi *et al.* (2011) evaluated five satellite-based products to force a hydrological model and simulate streamflows. The results showed that the bias-corrected datasets captured streamflow patterns well. However, the non-bias-corrected products overestimated the streamflow over warm seasons and underestimated it in cold seasons. Wu *et al.* (2018) evaluated the Multi-Source Weighted-Ensemble Precipitation (MSWEP V2.1) and three satellite-based precipitation products with rain gauge observations to simulate streamflows on the upper Huaihe River Basin in China. The results showed that the merged precipitation product (MSWEP) generally outperformed the other satellite datasets, although significant uncertainty existed in mountainous regions. Beck *et al.* (2017b) evaluated 23 gridded precipitation datasets over the 2000-2016 period. Thirteen daily uncorrected datasets (non-dependent on gauges for correction) were compared with observations from gauges, and the other ten gauge-corrected datasets were evaluated using hydrological modelling. Among the uncorrected datasets, the merged-products datasets (MSWEP-ng) generally performed the best, followed by the reanalysis and then the satellite products. For the corrected datasets, results showed that datasets integrating daily gauge data (CPC Unified and MSWEP products) generally outperformed the other datasets. Finally, precipitation datasets have also been evaluated using the surface water budget (Getirana *et al.* (2011); Lorenz *et al.* (2014); Munier & Aires (2018); Sabarly *et al.* (2016); Sheffield *et al.* (2009); Smith & Kummerow (2013); Song *et al.* (2016)), as well as using surface water and energy budgets (Kang & Ahn (2015); Hobeichi *et al.* (2020b); Hobeichi *et al.* (2020a); Yang *et al.* (2015)).

Despite the growing literature on the subject, the question regarding the most accurate dataset for capturing the spatio-temporal variability of weather events or driving hydrological models for climate change impact studies remains unanswered. The main objective of this study is therefore to establish a large-scale comparison of available temperature and precipitation

datasets covering a time-period long-enough to define the regional climate and therefore having the potential to be used as reference datasets to bias-correct climate model outputs for climate change impact studies. Secondary objectives include a comparison of families of climate data sources (gauges, gridded products, satellite products and reanalysis products), a quantification of the variability between the different datasets and providing recommendations on their applicability depending on the region of interest and physiographic characteristics.

2. Study Region and Data

2.1 Study Region

The study region is composed of 3138 catchments distributed over North America. The catchments were selected from the NAC^{2}H database (Arsenault et al., 2020a), which is a set of pre-processed catchments for climate-change studies that is a subset of the larger HYSETS database (Arsenault *et al.* (2020)). Figure II-1 presents the geographic distribution of the catchments, and Table II-1 presents the main statistics of the set of catchments.

Table-A II-1 Main properties of the study basins

Basin attribute	**Minimum**	**Maximum**	**Median**	**Mean**
Elevation (m)	7.3	3,585	380	692
Area (km^2)	302	179,150	1803	6,317
Mean Annual Precip. (mm/yr)	307	3,895	993	981
Mean Annual discharge (m^3/sec)	0.048	1,584	17.60	56.75
Temporal coverage (years)	5	30	30	27.6

2.2 Data

2.2.1 Gridded precipitation and temperature datasets

Currently, a significant number of gridded datasets have been stockpiled from stations, satellites, reanalysis or a combination thereof. However, not all these datasets can be used for

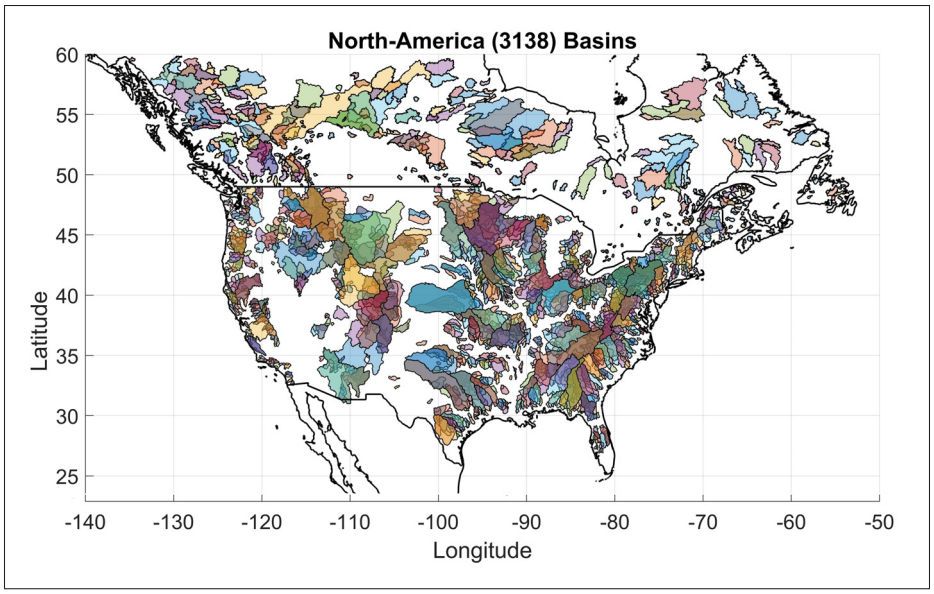


Figure-A II-1 Spatial distribution of the North American basins

climate change impact studies. To be useful, a dataset must have the following characteristics: 1) spatial resolution (finer than 1° for example to be used in local hydrological studies); 2) daily scale or finer temporal resolution; 3) long temporal coverage (at least 30 years to establish robust statistics for downscaling and bias correction); and 4) for an inter-comparison study, all datasets should cover approximately the same time period.

Based on these criteria, nine precipitation and three temperature gridded datasets are included in this study, and are presented in Table II-2. The precipitation datasets are classified based on their respective data sources. Two datasets are based solely on gauge data: CPC Unified (Climate Prediction Center Unified Gauge) and GPCC (Global Precipitation Climatology Center); two combine gauge and satellite data: CHIRPS V2.0 (Climate Hazards group Infrared Precipitation with Stations) and PERSIANN-CDR V1R1 (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) Climate Data Record (CDR)); four are derived from reanalysis: ERA5 (The European Centre for Medium-Range Weather Forecast's 5^{th} generation reanalysis), ERA-Interim (European Centre for Medium-range Weather Forecasts ReAnalysis Interim), JRA55 (Japanese 55-year ReAnalysis) and NCEP-CFSR (National Centers for Environmental Prediction (NCEP) Climate

Forecast System Reanalysis (CFSR), while the last is a multi-source dataset integrating gauge, satellite and reanalysis data (MSWEP V1.2 (Multi-Source Weighted-Ensemble Precipitation)). In terms of temperature, three datasets are included in this study: the gauge-based CPC unified dataset, the ERA-Interim and ERA5 reanalysis products. Properties of the temperature datasets are also provided in Table II-2.

Table-A II-2 The selected global gridded datasets

No.	Short Name	Data Source	Spatial Resolution	Spatial Coverage	Temporal Resolution	Temporal Coverage
1- Precipitation						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	GPCC	Gauge	1.0°	Global	Daily	1982-2016
3	PERSIANN-CDR (V1R1)	Gauge, Satellite	0.25°	±60° Lat.	6 hourly	1983-2012
4	CHIRPS V2.0	Gauge, Satellite	0.05°	±50° Lat.	Daily	1981-Present
5	NCEP-CFSR	Reanalysis	0.5°	Global	6 hourly	1979-2012
6	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
7	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017
8	JRA-55	Reanalysis	0.5625°	Global	3 hourly	1959-Present
9	MSWEP V1.2	Gauge, Satellite and Reanalysis	0.25°	Global	3 hourly	1979-2015
2- Temperature						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
3	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017

There are additional satellite products that provide global rainfall information at finer resolutions than PERSIANN-CDR, which has been selected in this work. Of particular interest is the Global Precipitation Measurement (GPM) mission, designed to further precipitation monitoring from an array of microwave sensors. It was launched to provide a new generation of precipitation datasets with an improved accurate measurement for light rainfall and snow precipitation as well as more frequent observations over the medium and high latitudes (Hou *et al.* (2014)). GPM utilizes passive microwave sensors in addition to the infrared measurements from geostationary satellites, providing rainfall monitoring around the globe with higher spatial and temporal resolutions than the previously widely used TMPA products (Yong *et al.* (2015)). These improvements are likely to provide significant advantages for hydrometeorological studies, weather forecasting, water budget studies and many other applications. In particular, the GPM Integrated Multisatellite Retrievals (IMERG), provides data at 0.1° and half-hourly spatial and temporal scales (Huffman *et al.* (2015)) and the Global Satellite Mapping of Precipitation (GSMaP) provides hourly rainfall data also at a 0.1° resolution (Okamoto *et al.* (2005)). These products have been evaluated against gauge measurements over different regions (Aslami *et al.* (2019); Asong *et al.* (2017); Chen *et al.* (2016); Lu & Yong (2018); Mazzoglio *et al.* (2019)) and showed satisfactory results. However, these state-of-the-art products do not cover a long-enough time period to be used for the evaluation of a climatic base-line period required for climate change impact studies. They were therefore not chosen for this study.

2.2.2 Observed hydrometeorological data

The observed data (OBS) over North America was taken from the North American Climate Change and hydroclimatology (NAC^2H) database, which is a hydrology and climate change impact dataset developed to study the impacts of different components of the modelling chain on hydrological indices over a collection of 3540 North American catchments. It includes hydrometeorological data such as precipitation, maximum and minimum temperature and streamflow at the daily scale for each of the catchments. Observed meteorological data come from regional datasets interpolated from station data. For Canada, the meteorological data is sourced

from Environment and Climate Change Canada (ECCC), whereas Canadian streamflow data is provided by Water Survey Canada, the hydrometric branch of ECCC. In the United States, NAC^{2}H uses the Livneh gridded dataset for meteorological data, whereas streamflows are provided by the United States Geological Survey (USGS) National Water Information Service. NAC^{2}H data is open source and available on the Open Science Foundation data repository at the following website: https://osf.io/s97cd/. More details can be found in (Arsenault *et al.* (2020)).

3. Methodology

3.1 Intercomparison of gridded climate products and statistics

The first step in this study was to evaluate the different products relative to a reference dataset; in this case, the NAC^{2}H observation dataset. While it was considered as the reference dataset, there was no underlying assumption that it is of higher quality or more accurate than any of the gridded products. Rather, it simply served as a baseline against which the other data products were compared. Analyses were performed by comparing annual and seasonal means of the gridded climate variables to the reference. This allowed finding spatial patterns of differences in average precipitation and temperatures to obtain a first feeling on the regional differences between the products. Then, a similar analysis was performed to investigate the differences in variability within these datasets on a daily time step. This also allowed evaluating the properties at a time scale that is more difficult to manage for gridded datasets than are aggregated annual or seasonal scales. The tests were performed because gridded datasets presenting no annual differences in precipitation versus the reference could still be largely underestimating the variance found in observational records. Metrics such as the Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and correlation coefficient (r) were used to compare the annual and seasonal precipitation and temperature values to the reference datasets, allowing to quantify the level of departure from the reference data.

3.2 Evaluation using hydrological modeling

The quality and performance of the climate variable datasets were evaluated indirectly through an independent measure, namely, the watershed observed streamflow. The hypothesis posed here is that climate datasets that allow for more accurate hydrological modelling with respect to the observed streamflow should be considered as being of higher quality. Of course, the choice of a hydrological model does influence performance. However, this should be seen as a first attempt at finding inconsistencies within the climate datasets. Hydrological modelling is sensitive to the annual cycle of precipitation and temperature, as well as to the coherency between both variables, so it can therefore be seen as a good evaluator of dataset overall quality. This approach has been used in several other studies (e.g. Beck *et al.* (2017b); Essou *et al.* (2016a); Tarek *et al.* (2019)).

3.2.1 The HMETS hydrological model

HMETS (Hydrological Model - École de technologie supérieure) is a lumped and conceptual model used in many research applications and as a component of operational multi-model hydrological studies and forecasting (Martel *et al.* (2017)). It was selected due to its good performance in the study domain in previous studies and because a lumped model was required to simulate discharge over the large number of catchments in the study.

The HMETS hydrological model is a 21-parameter, reservoir-based model that simulates water balance, including snow processes and the horizontal hydrological fluxes, using a series of unit hydrographs. It requires daily maximum and minimum temperature as well as daily rainfall and snowfall amounts. All the chosen temperature datasets provide daily minimum and maximum temperature, with the exception of ERA5, which provides mean hourly air temperature. The minimum and maximum hourly temperatures for each day were therefore used as being representative of the daily maximum and minimum values. When needed by the hydrological model, mean daily temperature was computed as the average of daily minimum and maximum temperatures. HMETS starts by computing the potential evapotranspiration using the Oudin

formulation (Oudin *et al.* (2005)), which is scaled through a calibration parameter, and then computes snow accumulation and melt with a 10-parameter degree-day-based snow module developed by Vehviläinen (1992). Rainfall is then added to the runoff generated by snowmelt to obtain total water production. Potential evapotranspiration is then subtracted from the total water production to obtain the final runoff depths. The water then infiltrates into one of three underground soil layers modelled as reservoirs (the aquifer, the vadose zone and the delayed surface runoff zone), using six calibrated parameters. Some of the water is also kept above the soil as the surface runoff reservoir. Water from these reservoirs is routed using two independent 2-parameter gamma distribution unit hydrographs for the surface unit hydrograph and the delayed unit hydrograph, respectively.

3.2.2 Hydrological model calibration

HMETS was calibrated using the automatic Covariance Matrix Adaptation Evolution Strategy (CMAES) optimization algorithm (Hansen *et al.* (2003)), which was shown to perform well with such optimization problems (Arsenault *et al.* (2014)). The objective function used to calibrate the parameters was the Kling-Gupta Efficiency (KGE) metric, which was introduced by (Gupta *et al.* (2009)) and modified by (Kling *et al.* (2012)), and which is an equally weighted bias, variance and correlation aggregate metric. It has been shown to be more easily interpreted than the Nash-Sutcliffe Efficiency (Nash & Sutcliffe (1970)) metric from which it is derived. The KGE values theoretically range from negative infinity (extremely poor performance) all the way to one (perfect performance). The performance ratings used in this study are defined based on the work of Gutenson et al. (2019) and Pechlivanidis and Arheimer (2015) who divide KGE values into three modeling-performance groups: Poor performance (KGE < 0.4), acceptable ($0.4 \leq$ KGE < 0.7) and good (KGE $\geq$ 0.7). Both precipitation and temperature datasets were averaged at the catchment scale before being fed to the hydrological model. Each catchment was calibrated by letting CMAES converge over 15,000 model evaluations and repeating the process twice. The calibration was performed on the entire length of the available data as recommended in (Arsenault *et al.* (2018)). The best calibration KGE score from the two generated sets was used to reduce the chance of considering a parameter set that had not

properly converged to an acceptable minimum. This calibration procedure was repeated for each combination of precipitation and temperature datasets, including the reference datasets, for each watershed, in order to allow their objective comparison (Essou *et al.* (2016a)).

4. Results

The results of the climatic variables analysis are first presented, followed by an analysis of the performance of the precipitation and temperature data when used in hydrological modelling.

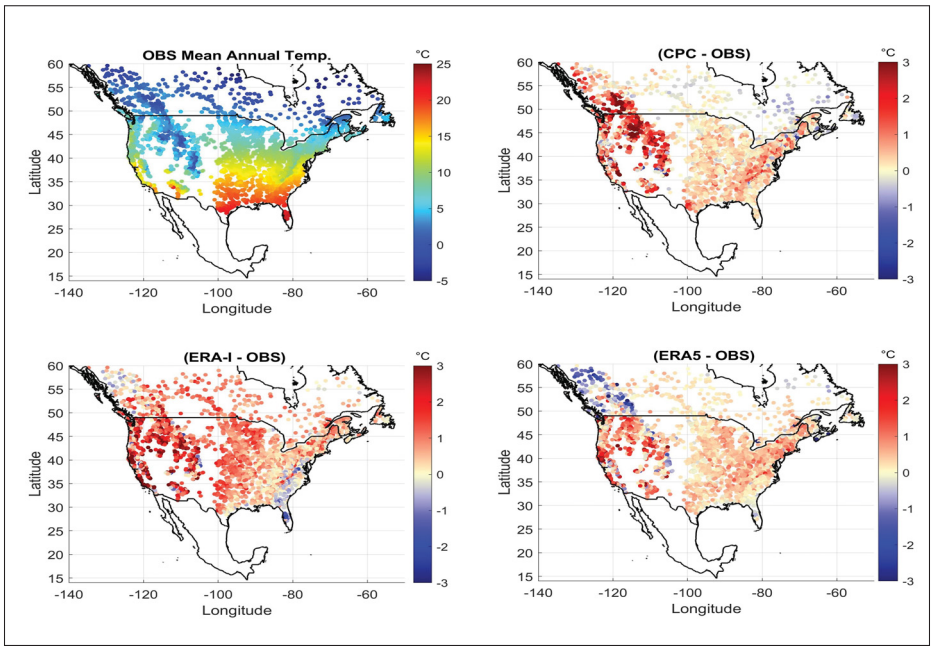


Figure-A II-2 Difference maps for the mean annual temperature (dataset-OBS) for the period (1983-2012)

4.1 Analysis of precipitation and temperature

Figure II-2 shows the mean annual temperature of the reference dataset (upper left) and differences between each of the three chosen temperature datasets. The term $difference$ is used below, instead of *bias*, since our reference dataset is not a true representation of the population, and is not inherently better than other datasets.

On average, the three datasets are warmer than the observations, with ERA-I being the warmest. The warm difference is particularly clear in the western United States. Overall, ERA5 is the closest to observations, with small differences across central and eastern North America, and reduced differences on the West Coast. Figure II-3 presents a similar analysis for precipitation. It shows observed mean annual precipitation (upper left) and differences of the 9 studied precipitation datasets.

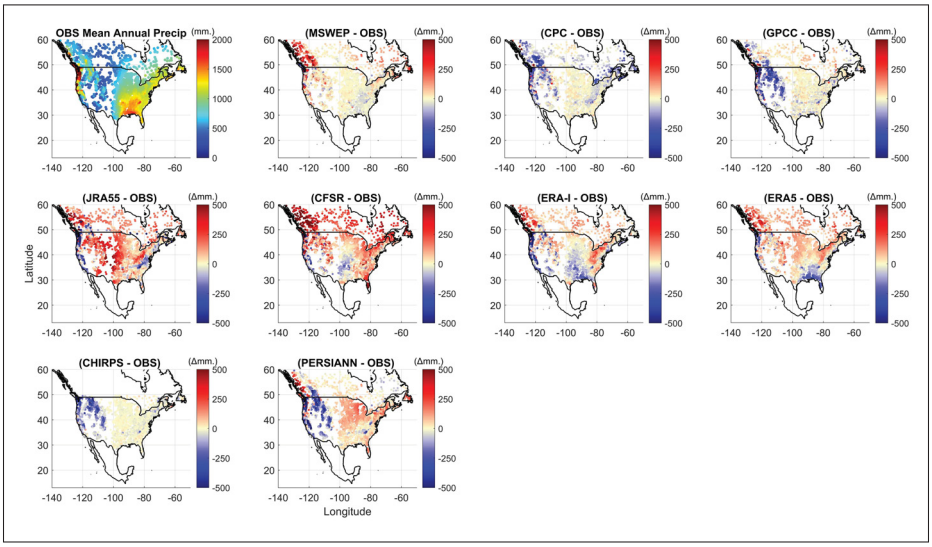


Figure-A II-3 Difference maps for mean annual precipitation (dataset-OBS). Note that CHIRPS does not provide data beyond $\pm 50^{\circ}$ Latitude

It can be seen in (Figure II-3) that the precipitation products differ widely, depending on the source and type of processing of data. Datasets that integrate observations (first row: CPC, GPCC, MSWEP and last row: PERSIANN and CHIRPS) show much smaller differences in general, as compared to the four reanalysis products (central row). ERA5 is the best-performing reanalysis, followed by its predecessors, ERA-I, JRA55 and CFSR products, which are wetter over most of North America. Figure II-3 also shows large differences in the western mountain ranges for all datasets, outlining limitations for all gridded precipitation datasets under study.

To further investigate precipitation seasonality, figures II-4 and II-5 present seasonal precipitation differences for winter (DJF) and summer (JJA).

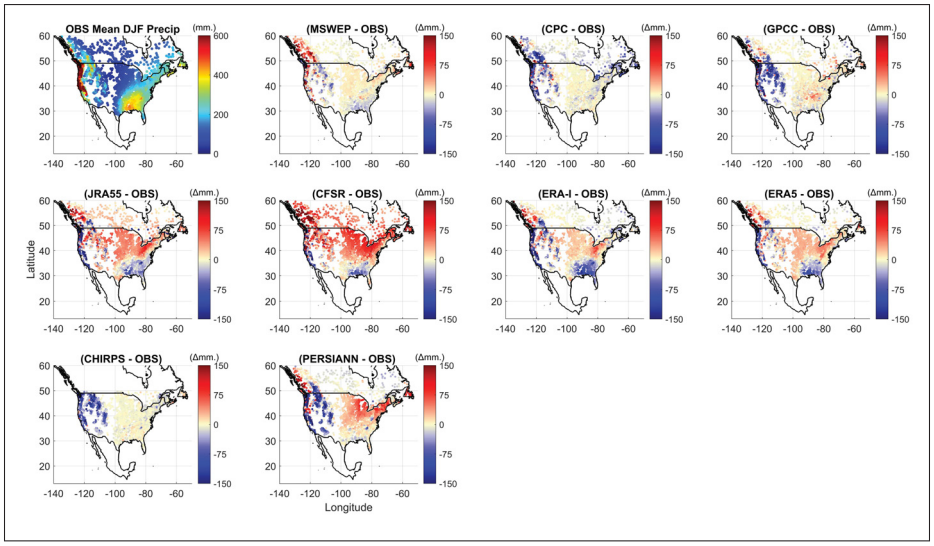


Figure-A II-4 Mean winter (DJF) Precipitation difference maps for the 1983-2012 period

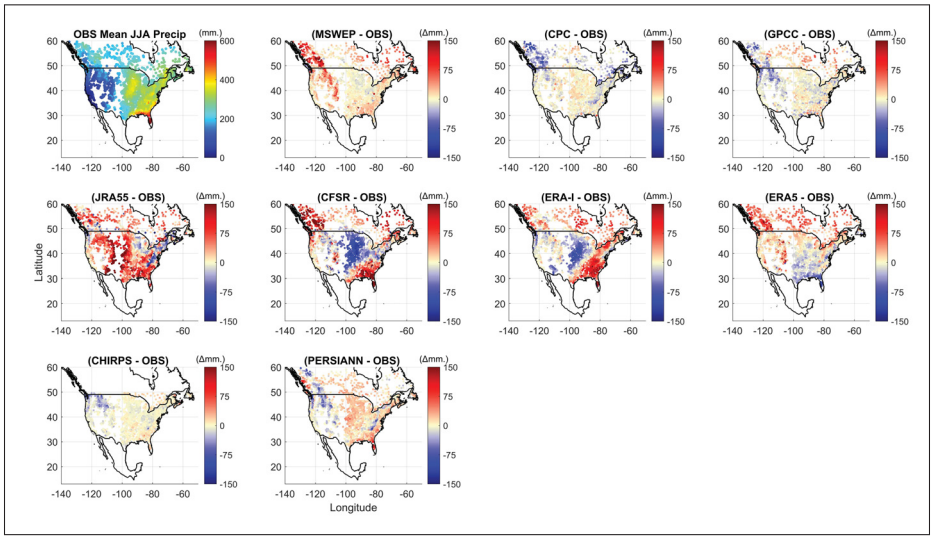


Figure-A II-5 Mean summer (JJA) Precipitation difference maps for the 1983-2012 period

Results for winter are very similar to those obtained at the annual scale. Summer differences (Figure II-5) do, however, display important differences. These differences are smaller for the satellite-based datasets and larger for the reanalysis datasets in general. The CFSR reanalysis is particularly dry in central USA. The differences between ERA5 and ERA-I are also larger, with ERA5 having smaller differences all across North America.

To analyze the results at a more localized scale, mean annual precipitation statistics were computed for each catchment and compared to the reference precipitation dataset. Figure II-6 presents boxplots for annual precipitation Mean Error (ME), mean absolute Error (MAE), RMSE and correlation coefficient. The spatial distribution of the last two metrics is also presented in figures II-7 and II-8. The boxplots are built from 3138 values, one from each individual catchment. The central boxes show the 25^{th} and 75^{th} quantiles (bottom and top), with the median in red. The whiskers display the smallest and largest values. Red crosses are considered statistical outliers. Overall, when compared to the reference dataset, we see that MSWEP is consistently the closest across all metrics. The two gauge-based products (CPC and GPCC) and CHIRPS follow. The median difference of ERA-I is close to zero, but otherwise displays a large spread. Surprisingly, ERA5 shows a relatively large positive mean difference, as do the other reanalyses (JRA55, CFSR). Correlation coefficients tell a similar story, with the main differences being that ERA5 clearly outperforms the other reanalysis. RMSE distributions also follow a similar pattern.

Figure II-7 presents the spatial distribution of mean annual precipitation RMSE values between each precipitation dataset and observations. JRA55 and CFSR are clearly the worst-performing datasets. MSWEP performs the best everywhere, with the exception of Western Canada, where ERA-5 and GPCC perform best. Figure II-8 presents the spatial distribution of correlation coefficients calculated for daily precipitation. MSWEP, ERA5, GPCC and CHIRPS are clearly the best-performing datasets. CPC performs well over the USA, but quite badly in Canada, where weather station density gets lower.

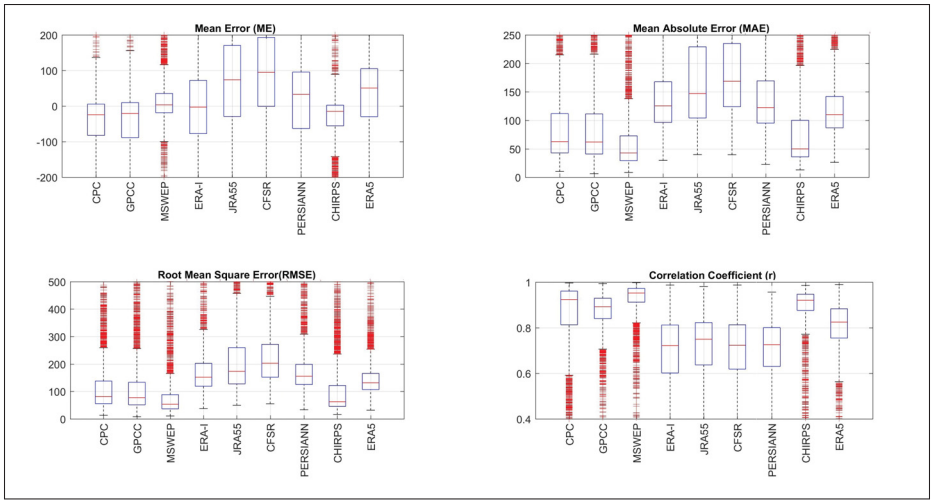


Figure-A II-6 Boxplots comparing ME, MAE, RMSE and r for 9 precipitation datasets at the annual scale

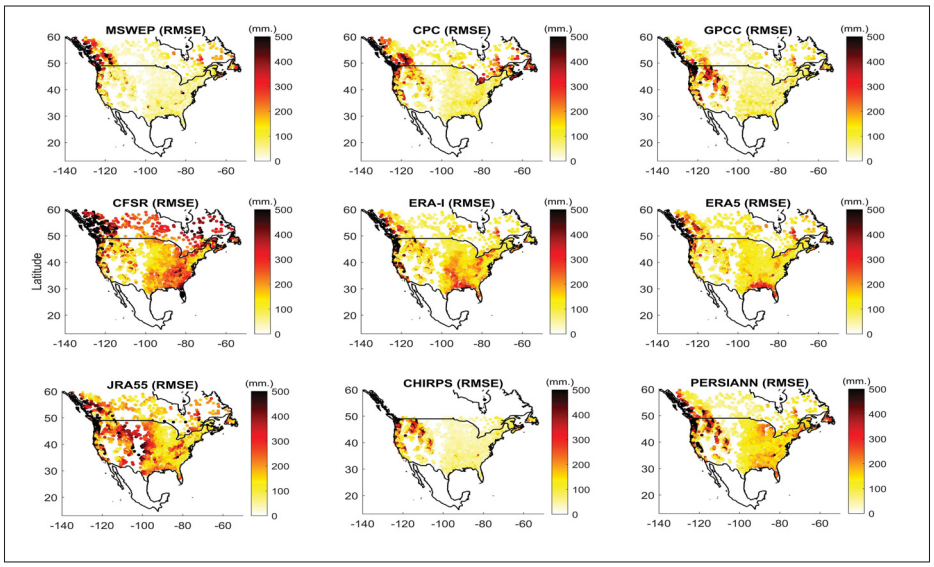


Figure-A II-7 Root Mean Square Error (RMSE) of mean annual precipitation for 9 precipitation datasets

To investigate in more details, the distribution of precipitation amounts at the watershed scale was examined. Figure II-9 presents quantile-quantile plots for the Saint Louis River watershed. This catchment was chosen as its behaviour is typical of most other watersheds, even though there is some level of regional control on quantile biases (not shown). The median quantile is represented by the solid red circle on each graph. This Figure therefore emphasises the larger

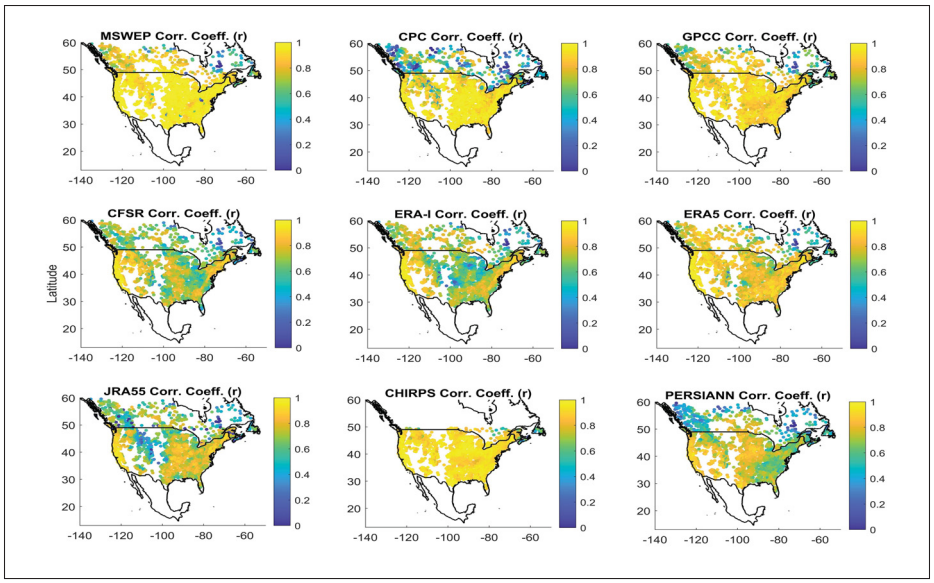


Figure-A II-8 Spatial distribution of correlation coefficients computed at the daily time step

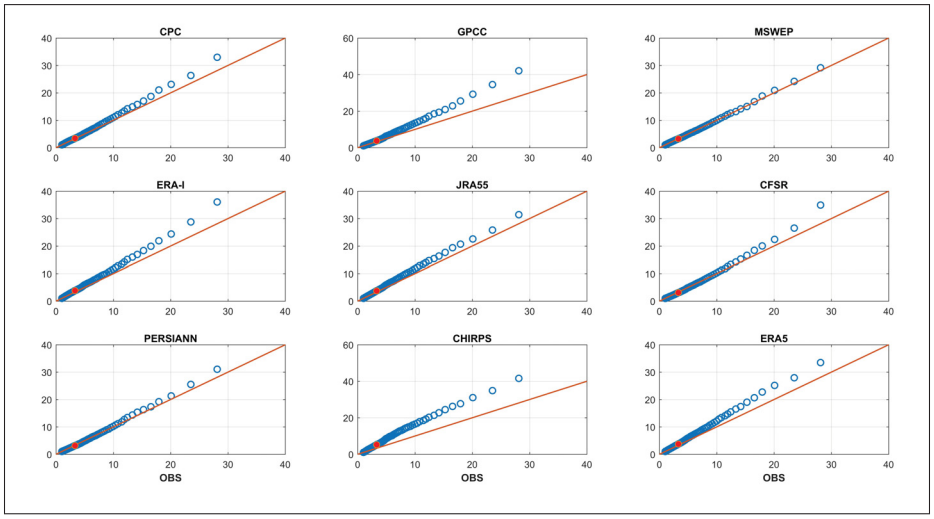


Figure-A II-9 Quantile-Quantile precipitation plots (0.01 to 0.99 at a 0.01 interval) for the Saint Louis River catchment in Minnesota State (USA). The median (quantile 0.5) is represented by the solid red circle

quantiles. The most striking feature of Figure II-9 is the overestimation of the larger quantiles at the catchment level for most datasets. This is particularly the case for all reanalysis, CHIRPS and GPCC. The lower quantiles are generally well represented by all datasets.

4.2 Hydrological model simulations

This section presents the results of the hydrological model simulations using all possible combinations of precipitation and temperature datasets. Figures II-10 and II-11 show the distributions of KGE scores for all catchments below (Figure II-10) and above 50°N (Figure II-11). The separation at 50°N was made for two reasons: the unavailability of data for CHIRPS, and to investigate the impact of the much lower resolution of observation networks in the North, which should technically affect gauge-based datasets.

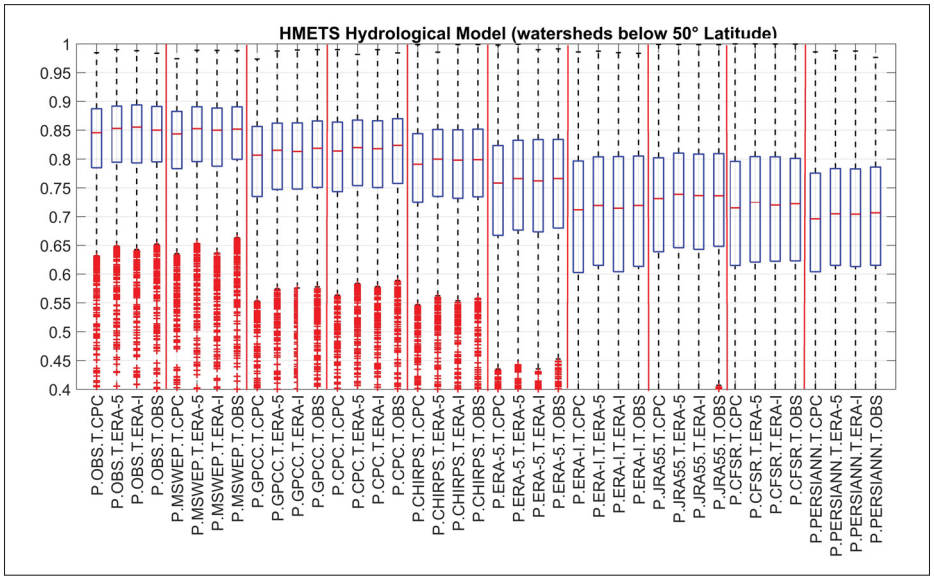


Figure-A II-10 KGE boxplots of simulated streamflows (below 50°N Latitude) from 10 precipitation datasets and 4 temperature datasets (40 combinations) using the HMETS hydrological model

Figure II-10 shows that all datasets can be used to generate good hydrological modelling, with all combinations generating median KGE values larger than 0.7. There are, however, large performance variations across datasets. The main driver of the modelling skill is the precipitation dataset. All four temperature datasets offer a nearly equal performance below 50°N, although CPC is consistently the worst of the four. The reference and MSWEP datasets clearly outperform all other precipitation datasets. These are then followed by the GPCC and CPC gauge-interpolated datasets, and then by CHIRPS and ERA5. The other satellite (PERSIANN)

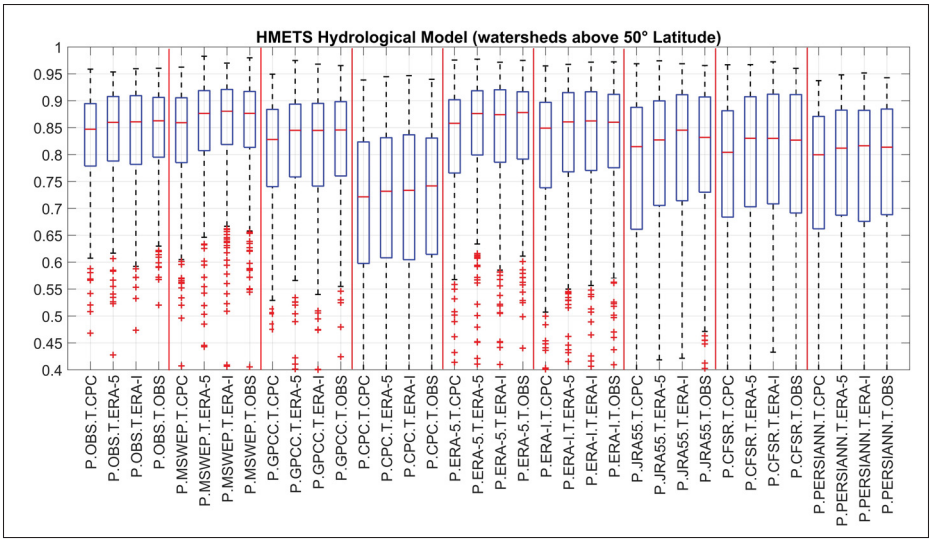


Figure-A II-11 KGE boxplots of simulated streamflow for 9 precipitation datasets and 4 temperature datasets (36 combinations) using HMETS hydrological model (above 50°N). Note that CHIRPS V2.0 does not provide data beyond ±50° latitude, and is excluded from this comparison

and reanalysis (ERA-I, JRA-55 and CFSR) products perform clearly worse than their best counterparts (CHIRPS and ERA5).

The results for catchments north of 50°N (Figure II-11) are markedly different. The differences between all datasets is much smaller, with the exception of CPC, which is the worst-performing dataset. This is consistent with results presented in Figure 8 showing that CPC precipitation behaves quite differently over Canada. The lower density of the station network is an equalizer, preventing gauge-based datasets from outperforming their counterparts. MSWEP and ERA5 are the best-performing datasets, and are slightly better than using the regional gridded dataset used as a reference. Overall, hydrological modelling performance is very good, and, for most datasets, better than below 50°N. This is very likely a combination of watersheds being larger, thus producing smoother, less reactive and easier to model hydrographs, and because snow-dominated catchments typically have relatively simple hydrographs, with a long winter recession curve and a strong spring snowmelt signature. The gain in performance is notable for all reanalyses, which are less affected by a deficient local observational network.

Figure II-12 shows the spatial distribution of KGE scores for the different precipitation datasets combined with the reference temperature dataset. The upper left graph shows the KGE values for the reference dataset, whereas all the other graphs display the difference in KGE values for each precipitation dataset. A red colour indicates a better performance, and blue, a worse one. It can be seen that MSWEP, GPCC, and CHIRPS to some extent, compare favourably with the reference dataset, and that CPC is affected by the lack of stations in the northern parts of North America. Also of note is the strong negative score associated with some of the reanalysis and satellite datasets in the eastern United States. Outside of this zone, ERA5 performs extremely well, as noted by Tarek *et al.* (2019).

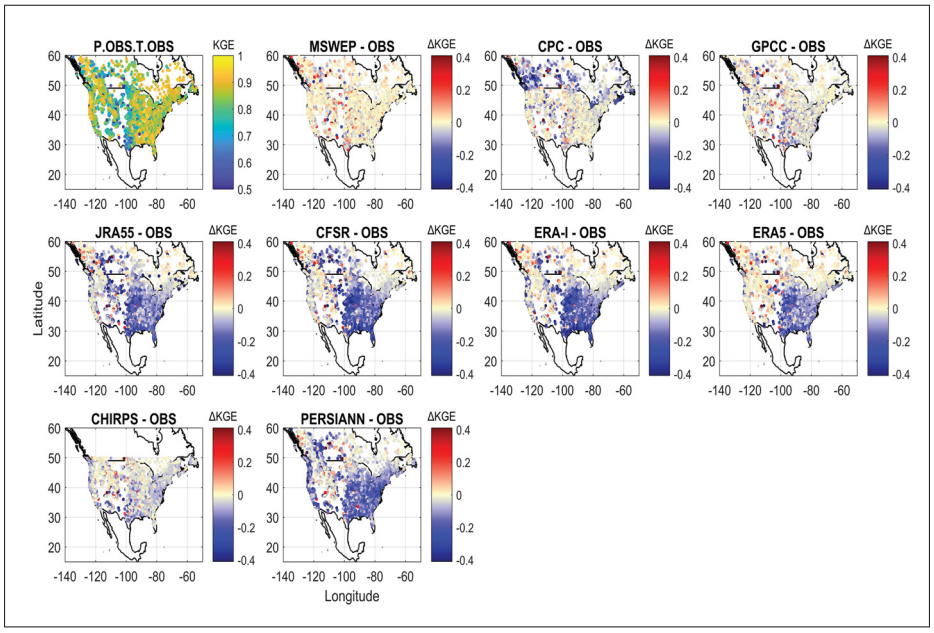


Figure-A II-12 Spatial distribution maps for the KGE difference between the observed precipitation dataset combined with the observed temperature dataset (top left plot) and the 9 precipitation datasets combined with the observed temperature dataset

Finally, Figure II-13 presents the aggregate mean KGE score over all catchments for all precipitation/temperature pairs (first row), as well as for the catchment below (second row) and above 50°N (third row). The first two rows are nearly identical due to the much larger number of stations located below 50°N. The third row displays warmer colours related to the preva-

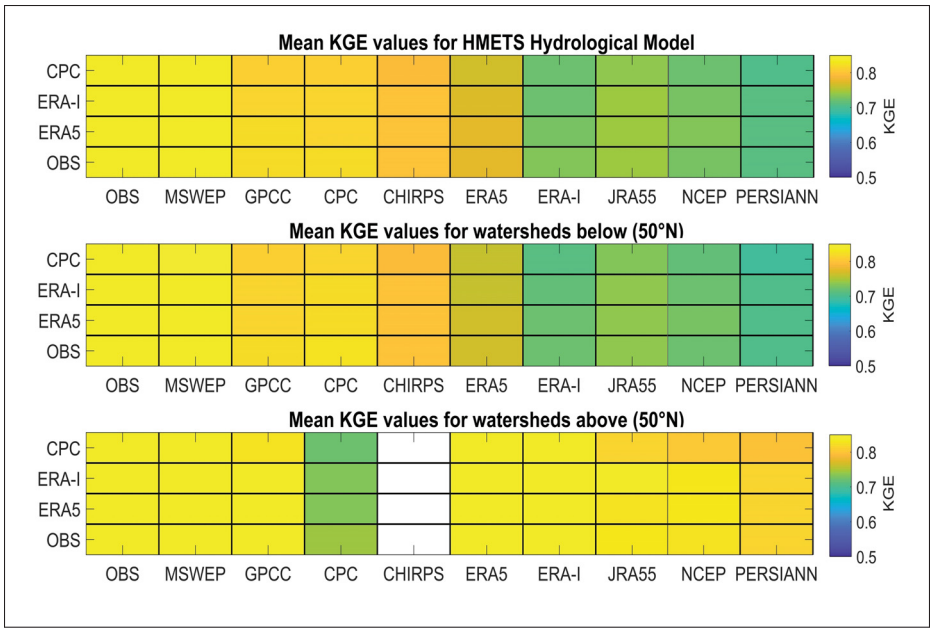


Figure-A II-13 Mean KGE values for all catchments (top panel), catchments below 50°N Latitude (center panel) and catchments above 50°N Latitude (bottom panel) for 10 precipitation datasets and 4 temperature datasets. CHIRPS does not provide data beyond ±50° Latitude, and left in white

lence of snowmelt-dominated watersheds, which are easier to model. Otherwise, these results confirm those of Figure II-12, and underline the relatively poor performance of CPC above 50°N for precipitation, and to a lesser extent, for temperature. Reanalysis datasets perform comparatively much better with both ECMWF reanalysis (ERA5 and ERA-I) products.

5. Discussion

Impact models strongly rely on hydrometeorological information. The performance of such models (stochastic and deterministic) is fundamentally dependent on the quality of input data. Weather station observations are considered as key information for most applications, but are limited in both time and space. Time series of relevant hydrometeorological variables are plagued with problems such as short temporal horizons, missing data, measurement errors, instrument biases and discontinuities introduced through equipment change and modification of the environment of weather stations, including their displacement. The low spatial density

of stations in many parts of the world, as well as the slow but steady decreasing trend in the number of weather stations around the world (Lawrimore *et al.* (2011)), compounds the problem. Gridded datasets are created to try to overcome many of the above problems. While it is likely that multi-source merged gridded products are the way of the future, it is not clear how good and reliable the many currently available gridded products are. This paper sheds some light on this issue by comparing nine global or near-global precipitation datasets and three temperature datasets over North America, therefore combining regions with high and low densities of weather stations.

The results showed important differences between all the datasets, as well as within categories of datasets (gauge-based, reanalysis and satellite-based). All the datasets were shown to be adequate for driving a hydrological model over 3138 catchments across North America. However, some datasets were clearly better than others in various circumstances. A first conclusion was that precipitation datasets are the main drivers of uncertainty. There was little difference between the four selected temperature datasets (NAC^2H observations, CPC, ERA-I and ERA5), even though CPC performed slightly worse than the selected reanalyses (ERA5 and ERA-I) and the reference gridded dataset. The equal performance of both reanalyses, when compared to the reference gridded datasets, could likely be explained by the fact that they assimilate the surface temperature from weather stations (in addition to a plethora of other data sources) and by the relatively small spatial and temporal variability of temperature, at least when compared to precipitation. Our evaluation of temperature is, however, based solely on hydrological modelling. Hydrological models have the ability to filter out some level of variability in driving inputs. Many other levels of validation still need to be performed (e.g., extremes) to determine if these alternative products are able to represent specific types of events in the hydrologic cycle. These results are nonetheless very encouraging for reanalyses, which are now available in near real-time and at spatial and temporal resolutions matching or exceeding those of most observational networks.

Comparatively, precipitation datasets provided a much more complex comparison picture. One important conclusion of this work is that the relative performance of precipitation datasets

below 50°N, which includes the contiguous United States and southern Canada, and above 50°N. Above 50°N, the density of the Canadian observational network is much lower. Results imply that a low-density station network narrows the gap between the reanalyses and gauge-based products. Reanalyses do not assimilate surface precipitation in their analysis scheme, and are therefore much less affected by a lack of ground precipitation measurements (either sparse station network or precipitation undercatch in gauged locations). This suggests that ERA5 precipitation is as robust as the best gauged products above 50°N and reanalyses should therefore be considered as good candidates in regions with deficient observational networks, confirming the conclusions of Tarek *et al.* (2019).

The spatial and temporal resolution of the datasets reviewed in this work differ widely. The temporal resolution itself (hourly to sub-daily) was not investigated. The spatial resolution of the above products, which varies from 0.05° to 1°, was summarily evaluated by analyzing hydrological modelling performance with respect to watershed size and by elevation, on the basis that higher-resolution datasets would perform better on smaller watersheds, or for high elevation watersheds, where the topography is more complex. No clear link was found between dataset performance and either size or elevation (results not shown). The main notable result was a clear improvement of ERA5 over ERA-I for high elevation catchments. This suggests that the fourfold resolution difference between the highest and lowest resolution datasets is not large enough to make a difference in this type of application, or that the use of a global hydrological model (which requires the averaging of the contributing grid points irrespective of their resolution) is not ideal to investigate the impact of resolution. Most of the selected watersheds are relatively large and therefore have a response time larger than one day. On those watersheds, the averaging of input data coupled with the smoothed hydrographs from the global hydrological models result in differences that are very difficult to see when using a criterion such as the KGE metric. Other metrics (e.g. peak flow reproduction, streamflow variance) may have been better suited to study the impact of dataset resolution (e.g. Kokkonen & Jakeman (2001)). The conceptualized nature of the hydrological models used in this study may also not be best suited to outline such differences. For the smaller watersheds in our database, sub-

daily modeling would be better suited (Bevelhimer *et al.* (2015)), but was not feasible since most datasets are limited to the daily time step. The use of a distributed hydrological model may be preferable to study the impacts of data resolution.

The results show that amongst all datasets tested in this study, MSWEP either is the best dataset, or is tied for best. In addition, MSWEP provides the second-highest spatial and temporal resolutions of all datasets. The performance of MSWEP demonstrates the potential of merged products in providing high quality outputs, by utilizing and integrating all available information. In high network density regions, MSWEP weighs observations heavily, but also relies heavily on reanalysis when weather station observation networks are less dense, such as in Northern Canada. We can expect an increasing number of datasets to rely on multi-source information, at the regional and global scales. At the regional scale, for example, high-resolution datasets can be obtained by combining ground-based radars and weather stations (Lespinas *et al.* (2015); Shen *et al.* (2018)). In addition, there are other potential reasons for the excellent MSWEP results. MSWEP is the closest relative (in terms of construction and resolution) to the chosen reference dataset (NAC^2H) and especially over the US. Over Canada, MSWEP relies to a much larger extent on reanalysis and is therefore not as closely related to NAC^2H. In addition, MSWEP uses streamflow data in its merging scheme, which may give it an advantage over the other datasets in terms of long-term biases. The use of streamflow data by MSWEP is however limited to long-term mean streamflow corrections, and only in regions with snowfall and/or complex topography. It is unlikely that this procedure has noticeable impacts on hydrological modeling performance at the daily scale.

CHIRPS performed very well for most of the comparison criteria. It performed better during the warm seasons, owing to its limitation in terms of detecting snowfall. It had a high correlation value (0.93) with the reference dataset as well as the second lowest mean absolute and root mean square errors. CHIRPS, which integrates satellite and gauge stations data on a high spatial resolution grid of 0.05°, has been shown to be a viable choice in climatological studies. Other studies have indeed mentioned its quality in this regard (Toté *et al.* (2015); Duan *et al.* (2016); Poméon *et al.* (2017); Beck *et al.* (2017b); Duan *et al.* (2019)).

For hydrological modelling, the results in this study have shown that, in general, gauge-based datasets perform better than reanalyses, whereas the performance of the two selected satellite products differ widely, with CHIRPS clearly outperforming PERSIANN. CPC is the worst gauge-only product, especially so over Canada. A few relevant studies have assessed the influence of gauge-density on climate data (Arsenault & Brissette (2014a); Gubler *et al.* (2017); Hofstra *et al.* (2010); Janis *et al.* (2004)). In particular, Janis *et al.* (2002) evaluated the required station-density to capture the regional climate variability in the United-States. The study reported that a station-density of 1 station per 180 km^2 would be needed to adequately monitor the climate variability.

ERA5 presents clear improvements over its predecessor (ERA-I), and is the best reanalysis product for hydrological modelling amongst those used in this study. ERA5 shines brightly, particularly above 50°N. The high spatial (0.25°) and temporal (1 hour) resolutions of ERA5 and the fact that it is available in near real-time lends it a significant advantage over most of the other datasets. The ECMWF recently launched the ERA5-LAND reanalysis at a 0.1° resolution. It uses the same assimilation process as ERA5, but is run at a finer resolution over land. Reanalyses could be considered as extremely complex multi-source merged products, and are likely to gain in importance in the near future. Their main limitation, when compared to MSWEP, for example, is that they do not integrate precipitation gauge data into the assimilation scheme. Reanalyses are very likely to be supported and improved in the future, as compared to the other datasets used in this study, which do not rely on recurrent national funding, and which often result from the efforts of small teams. Reanalysis performed worse than gauge-based products below 50°N, and particularly so in the eastern half of the U.S. Essou *et al.* (2016a) showed that reanalysis had difficulties reproducing the seasonal cycle of precipitation over this region. Reanalysis precipitation could easily be post-processed at the monthly scale using observations to palliate this problem, as was previously done on older reanalysis products (Weedon *et al.* (2014)).

Overall, results show that gauge-based datasets should be preferred in regions with good weather network density, with MSWEP being clearly the best performing dataset as repre-

sented by its results below 50°N. Above this latitude, where observational network density is much poorer, ERA5 performs just as well as MSWEP and the reference dataset. This indicates that ERA5, and potentially CHIRPS would be good choices as reference datasets for climate change impact studies in data sparse regions.

6. Conclusion

The performance of nine precipitation and three temperature global gridded dataset products was assessed in this paper in a two-step process. The datasets were first compared against two high-resolution regional gridded datasets over the U.S. and Canada. Performance was evaluated over 3138 North American catchments using annual and seasonal biases, mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and coefficient of correlation (r). In a second step, streamflows were simulated using all 40 possible combinations of precipitation and temperature datasets over 3138 North American catchments, and compared against data from gauging stations. Results showed that precipitation datasets are the main driver of uncertainty due to the relatively large differences between the datasets. Comparatively, differences between temperature datasets played a much smaller role as all four products behave very similarly. Temperature derived from observations and from the ERA-5 reanalysis provided marginally, but consistently better, results than the other two tested temperature datasets. For precipitation, overall, the merged-product MSWEP consistently performed best. Both gauge-based global products performed well over the U.S., but their performance decreased over Canada (and particularly in the case of CPC Unified), where observations are based on a less dense observational network. The ERA5 reanalysis performed really well over Canada and Western U.S., but its overall performance was affected by a relatively poorer performance over the Eastern U.S. It clearly outperformed the other three tested reanalysis. CHIRPS was found to be easily the best-performing satellite precipitation dataset, outperforming PERSIANN and all reanalysis, with the exception of ERA5.

7. Datastatement

The CPC, GPCC and NCEP datasets can be downloaded from the Earth System Research Laboratory (ESRL), available here:
https://www.esrl.noaa.gov/psd/data/gridded/tables/precipitation.html.

ERA-Interim, ERA5 and JRA55 datasets are available on the Research Data Archive:
https://rda.ucar.edu/datasets/ds628.0/.

MSWEP data are available through the PCA servers at:
https://platform.princetonclimate.com/PCA_Platform/mswepRetro.html.

The CHIRPS satellite dataset can be downloaded from the Climate Hazards Center:
https://www.chc.ucsb.edu/data/chirps.

Finally, the HMETS hydrological model is available on the MATLAB File Exchange:
https://www.mathworks.com/matlabcentral/fileexchange/48069-hmets-hydrological-model.

APPENDIX III

COMPARISON OF GRIDDED DATASETS FOR THE SIMULATION OF STREAMFLOW IN AFRICA

Mostafa Tarek[1,2], François P. Brissette[1], Richard Arsenault[1]

[1] Departement of Civil Engineering, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Québec, Canada, H3C 1K3

[2] Department of Civil Engineering, Military Technical College, Cairo, Egypt.

Article presented in *the 13th International Conference on Civil and Architecture Engineering (ICCAE-13), Cairo, Egypt*, 7-9 July 2020

Abstract. In recent decades, many parts of the African continent have experienced high precipitation variability with periodic drought and flood events. However, the network of streamflow gauges is too sparse in most countries to adequately capture these variations. In addition, no observed reference climatological dataset exists to adequately represent precipitation and temperature changes within all topographic and climatic zones. Consequently, the use of global gridded datasets needs to be considered. This paper aims to use the different available gridded datasets as inputs to a hydrological model to evaluate dataset performance. Nine precipitation and two temperature gridded datasets are used to this effect. The precipitation datasets include two gauged-only products, two satellite products corrected using ground-based observations, four reanalysis products and one merged product of gauge, satellite, and reanalysis. The two temperature datasets include one gauged-only and one reanalysis product. The nine precipitation and two temperature datasets were combined in their 18 possible arrangements for analysis purposes. Each combination was used to force the HMETS lumped hydrological model. The model parameters were calibrated individually for each combination against the streamflow records of 850 African catchments. The Kling-Gupta Efficiency (KGE) was used to evaluate the simulation performance. Results show that both temperature datasets performed equally well. Large differences were however observed between precipitation datasets. The MSWEP merged-product was the best-performing precipitation dataset, followed by CHIRPS satellites and ERA5 reanalysis products, respectively. The performance of both gauged-only

datasets (CPC and GPCC) was inferior, outlining the limitations of extrapolating information in data-sparse regions.

Keywords: precipitation datasets, gridded datasets, reanalysis products, streamflow simulation, hydrological modeling, African catchments.

1. Introduction

Ground meteorological stations are considered the most accurate source of climate data, as they offer physical record of data in a specified area. However, stations may suffer from many limitations such as missing measurements or short temporal coverage (Tapiador *et al.* (2012)). In recent decades, many regions have experienced high variability in precipitation with periodic drought and flood events (Tschakert *et al.* (2010); Rojas *et al.* (2011); Nicholson (2013); Omondi *et al.* (2014)). However, the spatial coverage of station networks is not sufficient to adequately represent these changes within all topographic and climatic zones (Nicholson (2013)). In addition, a gradual but steady decrease in the number of weather stations with long record listed in the Global Historical Climatology Network (GHCN) has started in the early 1990. To resolve all these problems, a large effort has been put into producing global gridded meteorological datasets. Such datasets provide continuous spatial and temporal coverage and, typically, with no missing data.

Over recent decades, several precipitation products have been produced with different spatial and temporal characteristics. These datasets differ in terms of data sources (gauge, radar, satellite, reanalysis or combinations thereof), spatial resolution (0.05° to 2.5°), spatial coverage (continental to global), temporal scale (30 minutes to annual) and temporal coverage (from 1 to several years). Several studies addressed the importance of evaluating these datasets to stand on their advantages and limitations. Most studies quantified the accuracy of these products through a direct comparison against data from weather stations (Vila *et al.* (2009); Andermann *et al.* (2011); Jiang *et al.* (2012); Prakash *et al.* (2018); Romilly *et al.* (2011); Chen *et al.* (2014)), while others assessed the performance indirectly using a hydrological model to compare

against observed streamflow (Behrangi *et al.* (2011); Tarek *et al.* (2019)Wu *et al.* (2018); Duan *et al.* (2019); Zhu *et al.* (2018)).

2. Study Area

In this study, the African continent was chosen as the main research area. Africa is considered the second largest continent with an area of 30.3 million km^2 covering about 20% of the global land area (Sayre (1999)). Africa is considered to be the hottest continent on Earth. The northern half is mostly covered by drylands and desert, while the central and southern parts contain savanna and rainforests (Hulme *et al.* (2001)). Based on the combination of temperature, precipitation and evapotranspiration, Africa can be divided into four main climatic zones; 1) arid and semi-arid, 2) tropical, 3) equatorial, and 4) temperate (Ngaira (2007)).

3. Data and Methods

3.1 Data

For many African regions, observed meteorological data are not easily available, either due to the lack of weather stations or the high fees to access the data. Most studies in Africa therefore depend on using satellite-derived data as a reference dataset (Skinner *et al.* (2015); Adjei *et al.* (2015); Koriche & Rientjes (2016); Bâ *et al.* (2018)). Hence, this paper aims to evaluate the performance of other several types of gridded datasets over a large set of hydrologically heterogeneous watersheds. The dataset performance is assessed through their ability as generating accurate streamflow through the use of a hydrological model.

3.1.1 Precipitation gridded datasets

Gridded datasets can be classified as a function of their data source. Gauge-based gridded datasets are obtained by interpolating the information measured at a small scale (typically a point measurement at a weather station) and mapped onto a predefined spatial and temporal resolutions grid. However, variation in gauge types or instrument replacements affect error

characteristics on the long-term records. In addition, observations are affected by systematic biases from evaporation and wind effect or due to, for example, elevation placement of gauges in mountainous regions (Isotta *et al.* (2014)).

A different approach to measure precipitation is using ground weather radars, as it partially addresses the issue of rain gauge coverage. Moreover, it provides much larger spatial coverage to measure precipitation than the point measurements provided by gauges. However, radar coverage is limited to developed regions that have a high population. In addition, they sense the real rainfall rate at a certain observational level above ground. Therefore, the presence of weather stations is required for the calibration and correction processes (Martens *et al.* (2013)).

Nowadays, satellite products are available at the global scale and can cover large areas at high spatial and temporal resolutions and near real time coverage. They are mainly suitable for rainfall estimation in the tropics and data sparse regions. However, satellites are relatively insensitive and generally miss a significant quantity of light precipitation and tend to fail over snow and ice-covered surfaces (Tian *et al.* (2009)). Some studies evaluated the uncertainties of these datasets and showed that high resolution satellite products perform better when bias corrected using gauge observations (Xie *et al.* (2007); Awange *et al.* (2016)).

Retrospective-analysis / reanalysis systems are vital sources of data in weather and climate studies. A typical reanalysis system consists of two main components, the forecast model and the data assimilation system. The role of the data assimilation system is to integrate observed databases of many sources of observations with the numerical weather forecast models to produce consistent gridded datasets (Di Luzio *et al.* (2008)). Although reanalysis are not direct observations, they provide analyzed variables in areas where stations are minimal (Bosilovich (2013)). Overall, no single precipitation product could be considered ideal for measuring precipitation. In fact, all precipitation products tend to miss a significant volume of rainfall (Behrangi *et al.* (2011)).

As discussed earlier, there is now a rather large number of gridded datasets from stations, satellites, reanalysis or a combination thereof. However, not all those datasets can be used

for climate change impact studies. Appropriate datasets would have the following desirable characteristics: 1) spatial resolution (between 0.05° to 1°); 2) daily scale or finer temporal resolution, 3) long temporal coverage (30 years), and 4) all datasets should cover approximately the same time interval. Based on those criteria, nine precipitation and two temperature gridded datasets were chosen in this study as shown in Table III-1. The precipitation datasets include two gauged-only products (GPCC and CPC), two satellite products corrected using ground-based observations (CHIRPS and PERSIANN), four reanalysis products (JRA55, NCEP-CFSR, ERA-Interim and ERA5) and one merged product of gauge, satellite, and reanalysis (MSWEP).

3.1.2 Temperature gridded datasets

Land surface temperature is a key variable for meteorological monitoring and forecasting services (Nieto *et al.* (2011)). It is also a key variable for climate and hydrological studies. In hydrological modelling, the air temperature is the key driving variable for the evapotranspiration and snowmelt processes. Hence, accurate temperature data is a vital issue. However, the lack of adequate gauge network can result in improper estimates of temperature. Therefore, temperature gridded datasets are also crucial in many fields. Temperature products are generally thought to be less complex than precipitation datasets due to its much smaller spatial and temporal variability. Therefore, much fewer studies have compared and evaluated the uncertainty of using different temperature datasets. On this study, two temperature datasets have been included: the gauge-based CPC dataset, and the ERA5 reanalysis.

3.1.3 Observed streamflow data

Streamflow records from the Global Runoff Data Centre (GRDC) were used to calibrate the hydrological models and evaluate the hydrological modelling performance. The GRDC database has streamflow data from 1150 African stations. In this study, 850 stations were chosen based on two criteria. First, stations should have data during the 1983-2012 study period. Second,

Table-A III-1 The selected global gridded datasets

No.	Short Name	Data Source	Spatial Resolution	Spatial Coverage	Temporal Resolution	Temporal Coverage
1- Precipitation						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	GPCC	Gauge	1.0°	Global	Daily	1982-2016
3	PERSIANN-CDR (V1R1)	Gauge, Satellite	0.25°	±60° Lat.	6 hourly	1983-2012
4	CHIRPS V2.0	Gauge, Satellite	0.05°	±50° Lat.	Daily	1981-Present
5	NCEP-CFSR	Reanalysis	0.5°	Global	6 hourly	1979-2012
6	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
7	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017
8	JRA-55	Reanalysis	0.5625°	Global	3 hourly	1959-Present
9	MSWEP V1.2	Gauge, Satellite and Reanalysis	0.25°	Global	3 hourly	1979-2015
2- Temperature						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
3	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017

stations that have less than five years of consecutive data during this period were excluded. The spatial distribution of these stations is shown in Figure III-1.

3.2 Hydrological model

In this study, the use of a distributed model was discarded due the scale of the study. The lumped hydrological model HMETS (Martel *et al.* (2017)) was used to evaluate the performance of the various climate datasets. This model has shown an overall good performance in a wide range of climates and hydrological studies (Martel *et al.* (2017); Perrin *et al.* (2003); Ar-

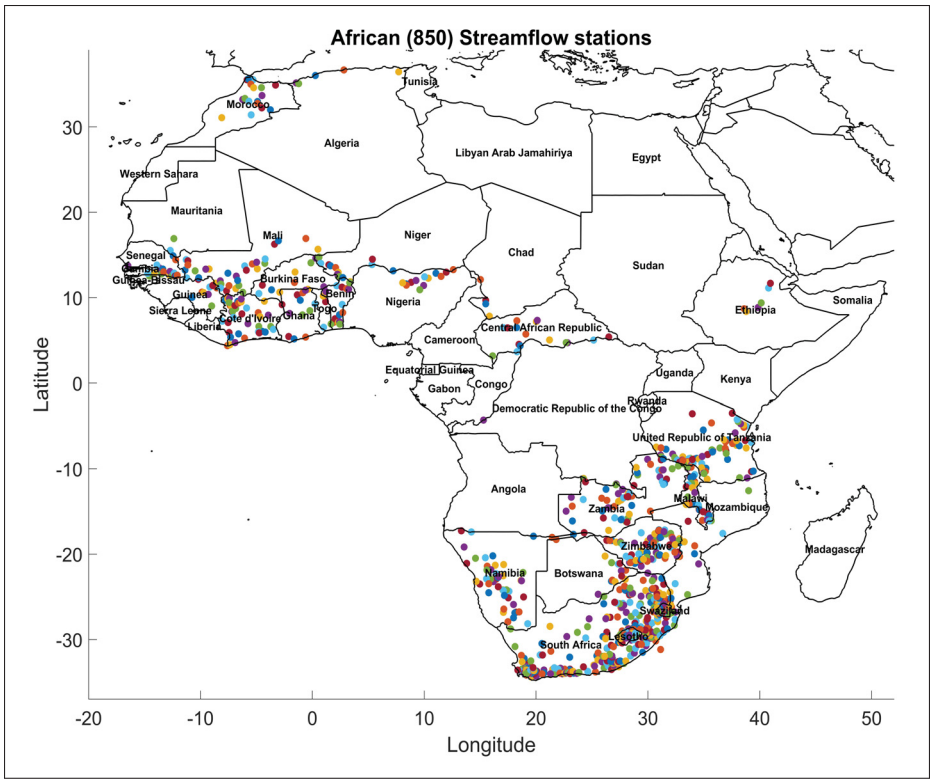


Figure-A III-1 Spatial distribution of the African 850 streamflow stations

senault *et al.* (2018)). The model requires daily precipitation, temperature and potential evapotranspiration (PET) as inputs. The Oudin's temperature-based formula (Oudin *et al.* (2005)) was used to calculate PET as it has shown an overall good performance and robustness on large-scale hydrological studies (Baguis *et al.* (2010)).

3.3 Hydrological model calibration

As will be detailed in the following section, the nine precipitation and two temperature datasets were combined in their 18 possible arrangements for analysis purposes and the hydrological model parameters were calibrated for each catchment and each dataset combination. The 15300 calibrations to be performed (9 precipitation datasets x 2 temperature datasets x 850 catchments) required the application of an automatic model parameter calibration method. For this study, the CMAES algorithm was applied because of its flexibility (Hansen *et al.* (2003)).

Moreover, it is considered as one of the best auto-calibration algorithms for hydrological modelling (Arsenault *et al.* (2014)).

The Kling-Gupta Efficiency (KGE) calibration objective function was used to evaluate the simulation performance. KGE is a modified version of the Nash-Sutcliffe Efficiency (NSE) metric that was introduced by Gupta (Gupta *et al.* (2009)) and modified by Kling (Kling *et al.* (2012)). It is defined as a combination of three elements; correlation, bias and variability as shown in (equation III-1). Pearson's correlation coefficient used to represent the correlation component (r), the ratio of estimated and observed means used to calculate the bias component (β) and the ratio of the estimated and observed coefficients of variation represent the variability component (γ).

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \qquad \text{(A III-1)}$$

The theoretical value for KGE to be equal 1 means that there is a perfect fit between the observed and simulated flows. Generally, KGE values above 0.6 are considered good.

4. Results and discussion

4.1 Analysis of precipitation and temperature

Figure III-2 presents mean annual temperature over the 1983-2012 period for the two selected temperature datasets. Both datasets display the same temporal patterns. ERA5 is however significantly warmer than CPC with a typical warm bias of 5-6 degrees over most of Africa. This difference is very large and can potentially affect evapotranspiration. However, the specific calibration of the hydrological model to each dataset has the potential to take this into account.

Regarding the precipitation datasets, to better present the differences between the products, the bias in the mean annual precipitation was calculated between each individual dataset and the average of all datasets as shown in Figure III-3. The average here is considered as the reference

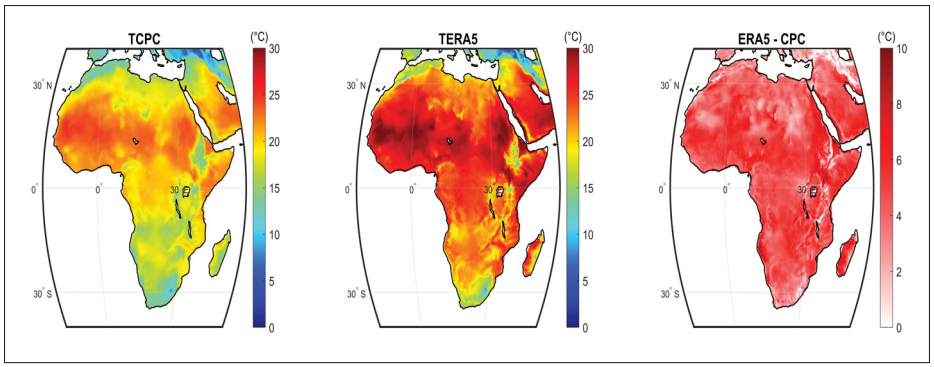


Figure-A III-2 Mean annual temperature for the two datasets and the bias between them

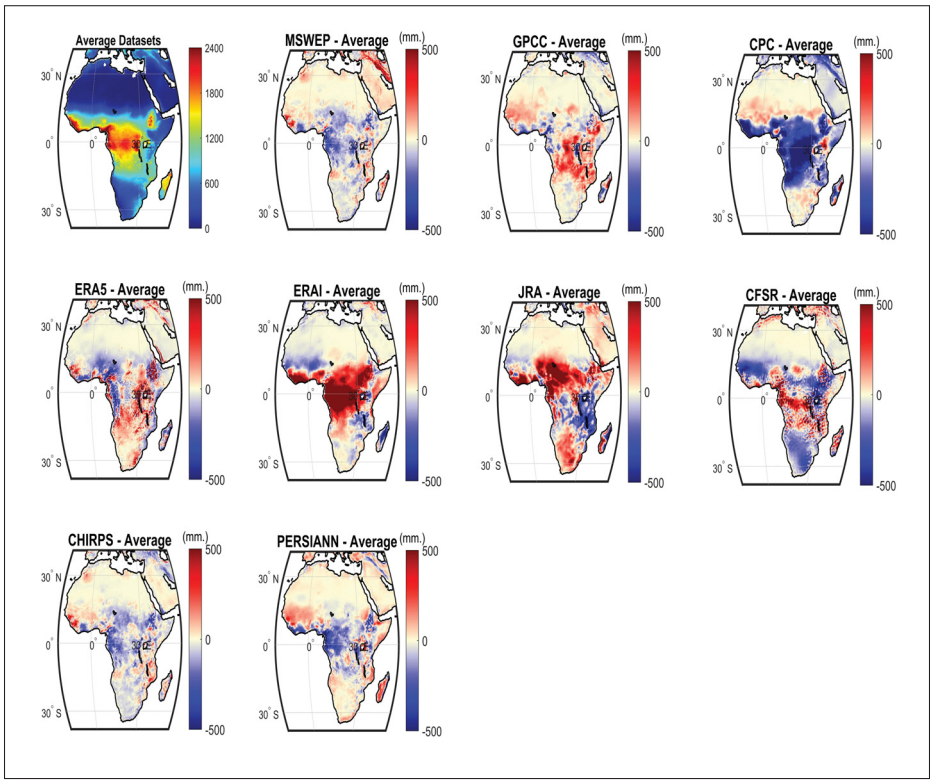


Figure-A III-3 Mean annual precipitation for the average of the all datasets (top left) and the bias from the average

bench mark. Since all the gridded datasets have different spatial resolution, the datasets were first interpolated to the finest grid scale. A red color indicates that the dataset is wetter than the average, while the blue color indicates it is dryer. Results show important differences between

the different precipitation datasets. All the datasets are generally similar in the desert and semi-desert regions but large differences are obvious in the tropical western and central regions.

Overall, the reanalysis (middle row) are wetter over the intertropical zone, with ERA5 being much closer than the other three considered reanalysis. The CPC gauge-based dataset is much drier than all other datasets. The large differences between both gauge-based datasets (CPC and GPCC) outline the complexity of interpolating in data-spare regions. Differences between the other datasets are comparatively smaller. In the absence of any reliable reference datasets, it is difficult to interpret the differences observed here. While an outlier dataset (e.g. CPC) may lead to suspicion, the limitations associated with each dataset does not allow for any firm conclusion. This is why hydrological modeling is used as an indirect validation method in this study. Even though streamflow gauges records do contain errors (Di Baldassarre & Montanari (2009)), in the context of this study, they are considered as the most reliable source for validation of the precipitation and temperature datasets.

4.2 Hydrological model simulations

This section presents the results obtained from the hydrological modelling simulations. Figure III-4 shows the distribution of KGE scores for each of the 18 combinations of precipitation (9 sets) and temperature (2 sets). Each boxplot in Figure III-4 contains the KGE scores of all of the catchments in this study.

Many conclusions can be drawn from Figure III-4. Both temperature datasets perform very similarly across all precipitation datasets, although ERA5 gives very small but consistently better results. Most of the differences observed in Figure III-4 therefore originate from the precipitation datasets.

All precipitation datasets result in acceptable KGE median value larger than 0.5, showing they can all be used for hydrological modeling. There are however large differences with some datasets clearly outperforming others. The CPC and GPCC gauge-based datasets are outperformed by five datasets. The MSWEP merged-product is quite clearly the best-performing

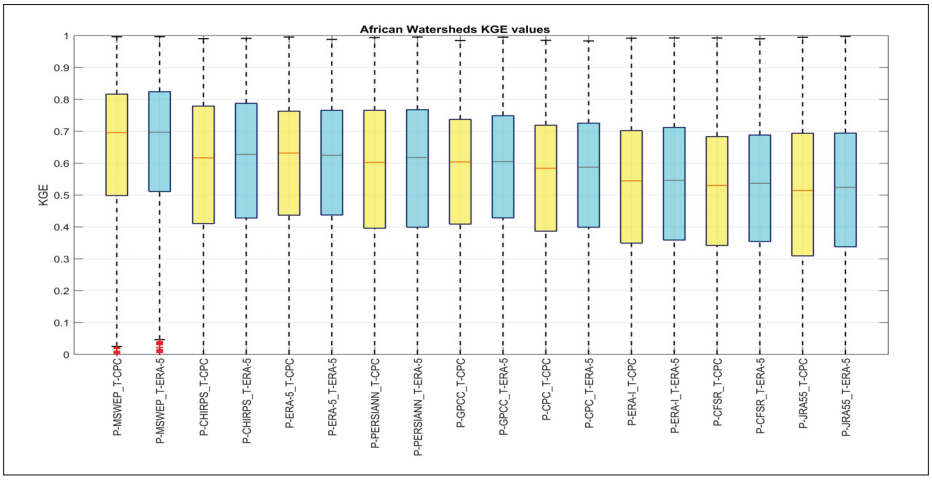


Figure-A III-4 KGE boxplots of simulated streamflows from 9 precipitation datasets and 2 temperature datasets (18 combinations) using the HMETS hydrological model

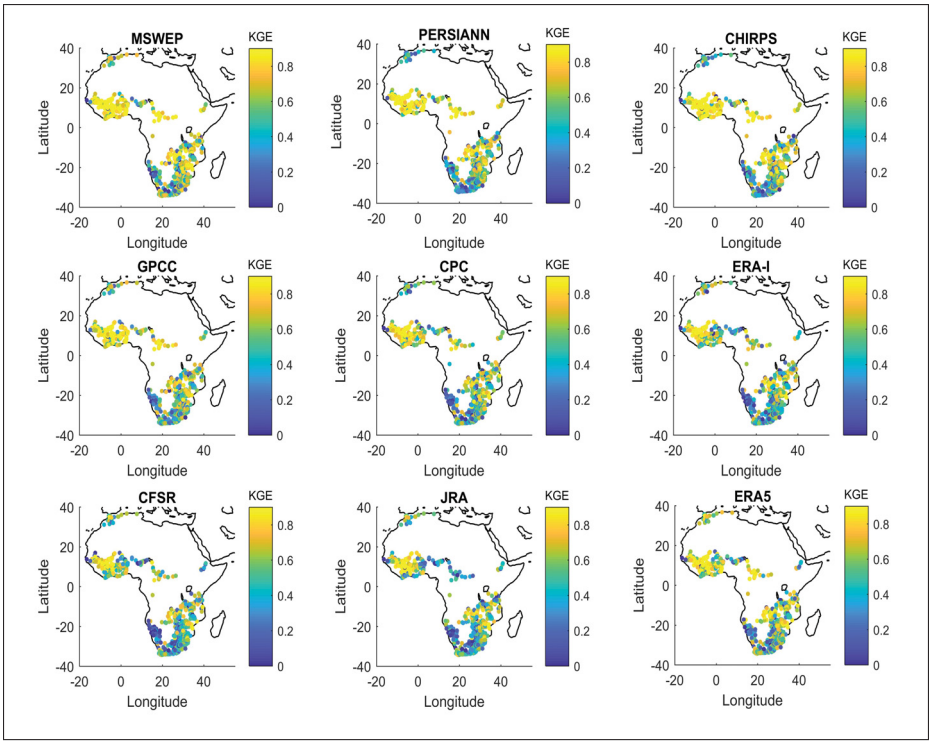


Figure-A III-5 Spatial distribution of KGE for nine precipitation datasets and ERA5 temperature dataset

precipitation dataset, followed by the CHIRPS satellite and the ERA-5 reanalysis datasets. The ERA-I, CFSR and JRA reanalysis are the least-performing datasets in this study.

In order to study the impact of spatial variability, Figure III-5 present the spatial distribution of KGE values for all nine precipitation datasets used in conjunction with ERA5 temperature.

The spatial patterns are consistent for all precipitation datasets. Hydrological modelling performance is general quite good everywhere with the exception of South Africa. This could either be due to less reliable streamflow records in this region or more likely to the hydrological model difficulties in dealing with the arid climate of south Africa. Rainfall-runoff models have long been known to have difficulties in such climates.

5. Conclusion

The main objective of this study was to evaluate the performance of nine precipitation and two temperature datasets to simulate streamflows of 850 African catchments over the 1983-2012 period. The MSWEP merged-product dataset was clearly the best performing one, followed by CHIRPS and ERA5 products, respectively. The performance of both gauged-only datasets (CPC and GPCC) was inferior, outlining the limitations of extrapolating point-based measurement in data-sparse regions. Both temperature datasets performed similarly.

6. Data Access

The CPC, GPCC and NCEP datasets can be downloaded from the Earth System Research Laboratory (ESRL), available here:
https://www.esrl.noaa.gov/psd/data/gridded/tables/precipitation.html.

ERA-Interim, ERA5 and JRA55 datasets are available on the Research Data Archive:
https://rda.ucar.edu/datasets/ds628.0/.

MSWEP data are available through the PCA servers at:
https://platform.princetonclimate.com/PCA_Platform/mswepRetro.html.

The CHIRPS satellite dataset can be downloaded from the Climate Hazards Center: https://www.chc.ucsb.edu/data/chirps.

Finally, the HMETS hydrological model is available on the MATLAB File Exchange: https://www.mathworks.com/matlabcentral/fileexchange/48069-hmets-hydrological-model.

APPENDIX IV

UNCERTAINTY OF GRIDDED PRECIPITATION AND TEMPERATURE REFERENCE DATASETS IN CLIMATE CHANGE IMPACT STUDIES

Mostafa Tarek[1,2], François Brissette[1], Richard Arsenault[1]

[1] Departement of Civil Engineering, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Québec, Canada, H3C 1K3

[2] Department of Civil Engineering, Military Technical College, Cairo, Egypt.

Paper submitted in *Hydrology and Earth System Sciences (HESS)* journal, October 2020

Abstract. Climate change impact studies require a reference climatological dataset providing a baseline period to assess future changes and post-process climate model biases. High-resolution gridded precipitation and temperature datasets interpolated from weather stations are available in regions of high-density networks of weather stations, as is the case in most parts of Europe and the United States. In many of the world's regions, however, the low density of observational networks renders gauge-based datasets highly uncertain. Satellite, reanalysis and merged products dataset have been used to overcome this deficiency. However, it is not known how much uncertainty the choice of a reference dataset may bring to impact studies. To tackle this issue, this study compares nine precipitation and two temperature datasets over 1145 African catchments to evaluate the dataset uncertainty contribution to the results of climate change studies. These datasets all cover a common 30-year period needed to define the reference period climate. The precipitation datasets include two gauged-only products (GPCC, CPC Unified), two satellite products (CHIRPS and PERSIANN-CDR) corrected using ground-based observations, four reanalysis products (JRA55, NCEP-CFSR, ERA-I, and ERA5) and one gauged, satellite, and reanalysis merged product (MSWEP). The temperature datasets include one gauged-only (CPC Unified) product and one reanalysis (ERA5) product.

All combinations of these precipitation and temperature datasets were used to assess changes in future streamflows. To assess dataset uncertainty against that of other sources of uncertainty, the climate change impact study used a top-down hydroclimatic modeling chain using

10 CMIP5 GCMs under RCP8.5 and two lumped hydrological models (HMETS and GR4J) to generate future streamflows over the 2071-2100 period. Variance decomposition was performed to compare how much the different uncertainty sources contribute to actual uncertainty.

Results show that all precipitation and temperature datasets provide good streamflow simulations over the reference period, but 4 precipitation datasets outperformed the others for most catchments: they are, in order: MSWEP, CHIRPS, PERSIANN, and ERA5. For the present study, the 2-member ensemble of temperature datasets provided negligible levels of uncertainty. However, the ensemble of nine precipitation datasets provided uncertainty that was equal to or larger than that related to GCMs for most of the streamflow metrics and over most of the catchments. A selection of the best 4 performing reference datasets (credibility ensemble) significantly reduced the uncertainty attributed to precipitation for most metrics, but still remained the main source of uncertainty for some streamflow metrics. The choice of a reference dataset can therefore be critical to climate change impact studies as apparently small differences between datasets over a common reference period can propagate to generate large amounts of uncertainty in future climate streamflows.

Keywords: Gridded datasets, precipitation, temperature, uncertainty, reanalysis products, streamflow simulation, hydrological modeling, African catchments.

1. Introduction

General Circulation Models/Earth System models (ESM) /Global Climate Models (GCMs) are the primary tools used to simulate the response of the global climate system to increases in greenhouse gas concentrations and to generate future climate projections. GCMs are complex mathematical representations of the physical and dynamical processes governing atmospheric and oceanic circulations as well as the interactions with the land surface. In order to reduce the computation burden, which can be considerable, GCMs represent the earth with a grid having a relatively coarse spatial resolution (IPCC (2001)). Consequently, GCM projections cannot be used directly for fine scale climate impact studies. Statistical/empirical or dynam-

ical downscaling techniques have thus commonly been used to address this scale mismatch. In addition, climate model outputs are always biased, and the extent of these biases can be evaluated through a comparison against observations over a common reference period. A bias correction procedure is therefore generally performed in addition to the downscaling step, and biases are assumed to be invariant in time when the correction is applied to future climate projections (Velázquez *et al.* (2015)). Although a two-step downscaling bias correction approach is preferable in most cases, a single instance of bias correction is sometimes used to account for both scale mismatch and GCM biases. While this may be acceptable when the scale difference is small (e.g., when using catchment averaged values), recent studies have shown that bias correction has limited downscaling skills (Maraun (2016)).

Statistical downscaling and bias correction approaches primarily rely on hydrometeorological observations over a historical reference period. It is therefore primordially important that the observed reference dataset represents the true climate state as closely as possible. For this task, ground stations remain the standard and most accurate/trusted source of weather data (New *et al.* (2001); Nicholson (2013)). However, the spatial distribution of these stations varies widely across the globe, and coverage is often sparse and even deficient in many parts of the world outside of Europe and the US. Even in well-covered regions, gauge data is subject to many problems, such as missing data, precipitation undercatch and inhomogeneities related to a variety of issues such as equipment change, station relocation and land surface modifications near each station (Kidd *et al.* (2017); Peterson *et al.* (1998)).

In recent decades, extensive efforts have been devoted to the development and improvement of gridded global and quasi-global climate datasets to overcome the limitations of gauge stations. These datasets provide meteorological record time series with continuous spatiotemporal coverage, and typically, no missing data. However, various error sources are inherent in these datasets, thus also bringing uncertainty to the data (Voisin *et al.* (2008)). Thus, choosing an appropriate reference dataset for climate change impact studies is an important concern, and especially so in regions with sparse ground station coverage.

According to Huth (2004): “For estimates based on downscaling of General Circulation Model (GCM) outputs, different levels of uncertainty are related to: (1) GCM uncertainty or inter-model variability, (2) scenario uncertainty or inter-scenario variability, (3) different realizations of a given GCM due to parameter uncertainty (inter-model variability) and (4) uncertainty due to downscaling methods”. In most climate change impact studies, it is generally assumed that GCMs are the major source of uncertainty (Mpelasoka & Chiew (2009); Kay *et al.* (2009); Vetter *et al.* (2017)). (Rowell (2006)) compared the effect of different sources of uncertainty using the initial condition ensembles of different General Circulation Models (GCMs), Greenhouse Gas Emission Scenarios (GHGESs) and Regional Circulation Models (RCMs) on changes in seasonal precipitation and temperature in the United Kingdom. The results indicated that the largest uncertainty comes from the GCM choice. (Minville *et al.* (2008)) used ten equally-weighted climate projections derived from a combination of five GCMs, two GHGESs and the change factor approach for downscaling to investigate the uncertainty envelope of future hydrologic variables. Their results showed that the uncertainty related to the GCM choice is dominant. These results have also been confirmed by several studies (Prudhomme & Davies (2009); Nóbrega *et al.* (2011); Dobler *et al.* (2012)). Other studies have assessed other sources of uncertainty such as Greenhouse Gases Emission Scenarios (GHGESs) (Prudhomme *et al.* (2003); Kay *et al.* (2009); Chen *et al.* (2011a)), the downscaling method (Wilby & Harris (2006); Khan *et al.* (2006)) and hydrological modeling (Bae *et al.* (2011); Vetter *et al.* (2017)). Recent studies have also looked at the uncertainty related to the choice of the impact model (Giuntoli *et al.* (2018); Krysanova *et al.* (2018)). From these studies, a more complex picture emerges, in which the main source of uncertainty may vary, depending on geographical location and metric under study. Dataset uncertainty has been assessed in numerous studies either by direct inter-comparison between datasets (Vila *et al.* (2009); Andermann *et al.* (2011); Romilly *et al.* (2011); Jiang *et al.* (2012); Chen *et al.* (2014); Prakash *et al.* (2018); Nashwan & Shahid (2019)) or by using hydrological modeling (Behrangi *et al.* (2011); Beck *et al.* (2017b); Wu *et al.* (2018); Zhu *et al.* (2018); Tarek *et al.* (2019)). However, to the best of our knowledge, the uncertainty of gridded datasets has not been evaluated against other sources of uncertainties when performing climate change impact studies. The objective of this study is

therefore to assess the impact of the choice of a given reference dataset on the global uncertainty chain of climate change impact studies. Since this is of particular concern to regions with sparse weather station coverage, this study is conducted over Africa.

2. Study Region and data

2.1 Study Region

2.1.1 Geographic situation

Africa is the second largest and second most-populous continent in the world. It covers a land area of about 30.3 million km^2, including adjacent islands, which represents 6% of Earth's total surface area and 20.4% of its total land area (Mawere (2017)). Deserts and dry lands cover 60% of its entire surface (Prăvălie (2016)). The average elevation of Africa is almost 600 m above sea level, roughly close to the average elevations of North and South America (Atrax, 2016). Generally, higher-elevation areas lie to the east and south, while a progressive decrease in altitude towards the north and west is apparent.

The African continent can be divided into 25 major hydrological basins (Karamage *et al.* (2018)). Generally speaking, the main drainage for all of the continent's basins is towards the north and west, and ultimately, into the Atlantic Ocean. About 95% of its streams are drained through permanent rivers. In some arid areas (i.e., Northwest Sahara Desert), drainage is sometimes absent or masked by sand seas (Karamage *et al.* (2018)). Roughly, 60% of the African continent is drained by 10 large rivers (Congo, Limpopo, Niger, Nile, Ogooue, Orange, Senegal, Shebelle, Volta and Zambezi) and their tributaries (Paul *et al.* (2014)).

2.1.2 Climate profile

Africa is the hottest continent on earth, and is the area that has seen the highest ever recorded land surface temperature (58 °C in Libya; El Fadli *et al.* (2013)). The continent is characterized by highly variable climates that range from tropical to subarctic on its highest peaks. According

to the Koppen climate classification (Köppen (1900)), the northern half is mainly classified as dry (group B) whereas the central and southern areas contain both savannah plains and dense forests with tropical and humid subtropical climates (groups A and C) with a semi-arid climate in-between (El Fadli *et al.* (2013)). These wide climate ranges are characterized by a wide variety of precipitation extremes, including droughts and floods. Droughts occur mostly in the Sahel and in some parts of Southern Africa, whereas flooding is most prevalent in the southern and eastern regions.

2.2 Data

This project used several datasets built from climate models, observed precipitation, temperature and streamflow, as well as catchment boundaries. These are described in the following four sub-sections.

2.2.1 General Circulation Models (GCMs)

All GCMs used in this study were part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor *et al.* (2012)). Long historical climate simulations (1850–2005) and future climate projections (up to 2100 and beyond) for four Representative Concentration Pathways (RCPs) are included in the CMIP5 database.

Ten CMIP5 GCMs from 10 different modeling centers were selected for this study, as shown in Table IV-1. They were selected as a subset of the GCMs used to set up the NAC^{2}H database (Arsenault *et al.* (2020)). The number of GCMs (10) was selected as a compromise between having an accurate representation of GCM climate sensitivity variability and keeping the large computational burden of this project reasonable. All GCM data was extracted over the 1983-2012 and 2071-2100 future periods under the (RCP8.5) emission scenario.

Table-A IV-1 List of chosen GCMs, research centres and spatial resolutions

No.	Models	Research Center	Spatial Resolution
1	BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration, China	2.79° x 2.81°
2	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China	2.79° x 2.81°
3	CanESM2	Canadian Center for Climate Modeling and Analysis, Canada	2.79° x 2.81°
4	CCSM4	National Center of Atmospheric Research, USA	0.94° x 1.25°
5	CMCC-CESM	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	3.44° x 3.75°
6	CNRM-CM5	National Center of Meteorological Research, France	1.40° x 1.40°
7	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, China	2.79° x 2.81°
8	INMCM4	Institute for Numerical Mathematics, Russia	1.5° x 2.0°
9	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	1.40° x 1.40°
10	MRI-CGCM3	Meteorological Research Institute, Japan	1.12° x 1.125°

2.2.2 Gridded precipitation and temperature datasets

The precipitation and temperature dataset selection was made on the basis of a high spatial resolution, daily (or better) temporal resolution, and of the availability of at least 30 years of data covering the same time period, in order to properly define the reference climate. Some recent datasets that provide global/near-global rainfall information at finer spatial and temporal resolutions (e.g. the GPM Integrated Multisatellite Retrievals (IMERG) (Huffman *et al.* (2015)),

Table-A IV-2 The selected global gridded datasets

No.	Short Name	Data Source	Spatial Resolution	Spatial Coverage	Temporal Resolution	Temporal Coverage
1- Precipitation						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	GPCC	Gauge	1.0°	Global	Daily	1982-2016
3	PERSIANN-CDR (V1R1)	Gauge, Satellite	0.25°	±60° Lat.	6 hourly	1983-2012
4	CHIRPS V2.0	Gauge, Satellite	0.05°	±50° Lat.	Daily	1981-Present
5	NCEP-CFSR	Reanalysis	0.5°	Global	6 hourly	1979-2012
6	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
7	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017
8	JRA-55	Reanalysis	0.5625°	Global	3 hourly	1959-Present
9	MSWEP V1.2	Gauge, Satellite and Reanalysis	0.25°	Global	3 hourly	1979-2015
2- Temperature						
1	CPC Unified	Gauge	0.5°	Global	Daily	1979-Present
2	ERA-Interim	Reanalysis	0.75°	Global	3 hourly	1979-8/2019
3	ERA5	Reanalysis	0.25°	Global	hourly	1979-2017

and the Global Satellite Mapping of Precipitation (GSMaP) (Okamoto *et al.* (2005)) were left out because their temporal coverage was too short to properly represent the mean climate over the reference period.

According to above criteria, nine precipitation and two temperature datasets were selected for this study. The precipitation datasets include two gauged-only products, two satellite products corrected using ground-based observations, four reanalysis products and one gauge, satellite,

and reanalysis merged product. The temperature datasets include one gauged-only and one reanalysis product as shown in Table IV-2.

2.2.3 Observed streamflow data

The Global Runoff Data Centre (GRDC) archive is arguably the most complete global discharge database providing free access to river discharge data (Fekete & Vörösmarty (2007)). The database provides streamflow records collected from 9213 stations across the globe, with an average temporal coverage of 42 years per station (Do *et al.* (2017)). It is operated under the World Meteorological Organization (WMO) umbrella to provide broad hydrological data to support the scientific research community. GRDC data has been widely used in various hydrological studies, such as those examining hydrological model calibrations (Milliman *et al.* (2008); Hunger & Döll (2008); Donnelly *et al.* (2010); Haddeland *et al.* (2011)), or as a benchmark to compare simulated streamflows (Trambauer *et al.* (2013); Zhao *et al.* (2017)).

2.2.4 Watersheds boundaries data

HydroSHEDS (the Hydrological data and maps based on the SHuttle Elevation Derivatives at multiple Scales database) is a freely available global archive, developed through a World Wildlife Fund (WWF) program, that uses a hydrologically-corrected digital elevation model to provide hydrographic information for regional and global studies (Lehner *et al.* (2008)). In addition, it applies a consistent methodology using Geographic Information System (GIS) technology to provide watershed polygons for more than 7000 GRDC gauging stations. Figure IV-1 shows watershed polygon layers at different spatial scales for the African continent. The vector layer (lev05), which consists of 1145 watersheds, was chosen to be used in this study.

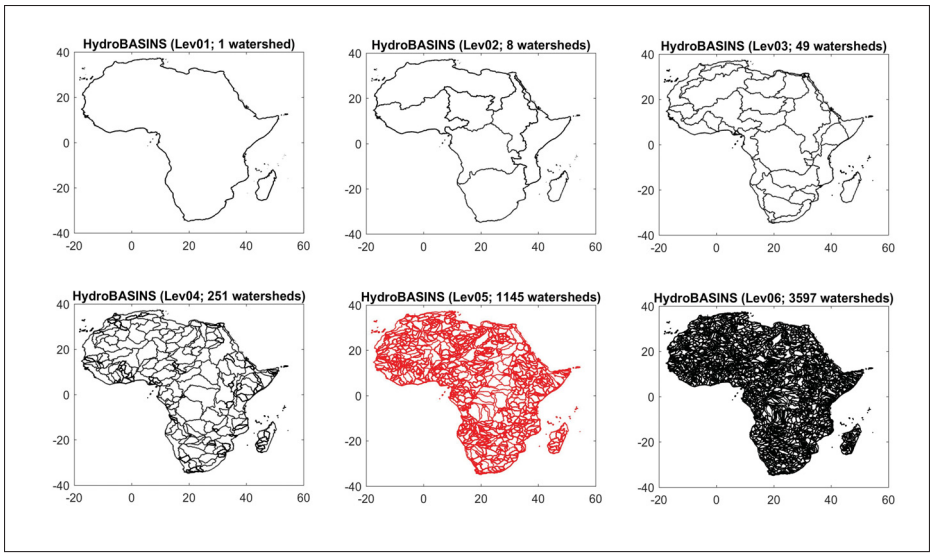


Figure-A IV-1 Sample of the different vector layers of watersheds on the African continent. Each layer has a different number of watersheds, depending on the required scale

3. Methodology

Figure IV-2 presents the methodological framework for this study. A large-sample hydrological climate change impact study is performed over 1145 African catchments. It uses the standard top-down approach in a modeling chain, which consists of 10 GCMs, 2 hydrological models, 2 temperature and 9 precipitation datasets, for a total of 360 possible combinations. A single GHGES (RCP8.5), a single climate projection for each GCM and a single downscaling method (see below) are used, since the focus of this work is not on conducting a complete uncertainty chain study. The uncertainty related to the reference dataset will therefore be compared to that of the climate model ensemble and against that of both hydrological models. For each catchment, 360 30-year streamflow time series are generated for both the reference (1983-2012) and future (2071-2100) time periods. Fifty-one streamflow metrics are computed for each of these time series. An n-dimensional analysis of variance is performed to partition the uncertainty linked to the 4 groups of components of the uncertainty modeling chain, as well as their interactions.

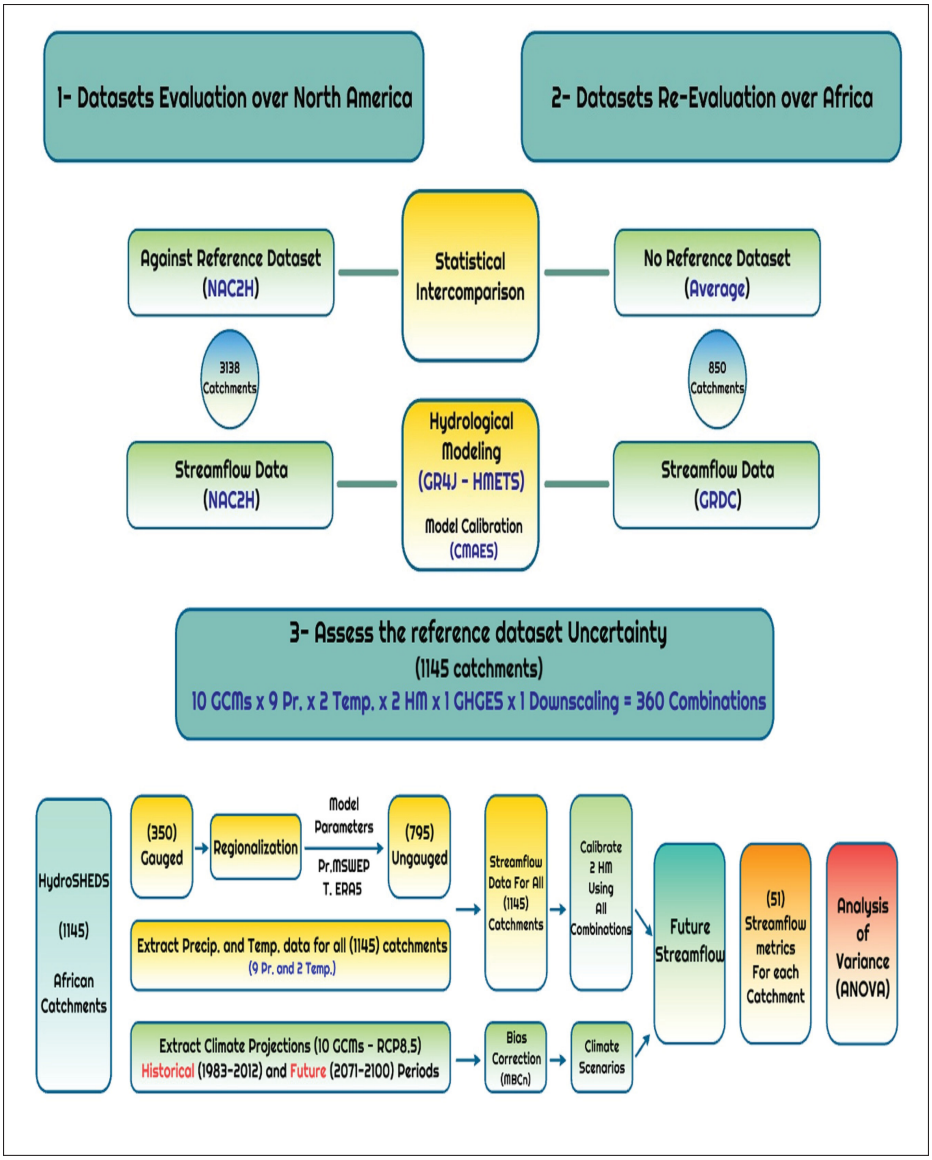


Figure-A IV-2 Overview of the various methodological steps implemented in this study

A first set of 350 catchments were selected based on the availability of long-enough gauged streamflow series. In order to include additional catchments to allow a better spatial coverage of the African continent, an additional 795 catchments were selected and an additional regionalization step was performed to generate streamflows.

Both hydrological models were calibrated on all catchments for all 18 combinations of reference datasets (2 temperature datasets x 9 precipitation datasets), for a total of 41,220 independent hydrological model calibrations. The main methodological steps are described below in Figure IV-2.

The watershed boundaries for the African continent were extracted from the HydroSHEDS database. Streamflow records from the GRDC database were used to calibrate the hydrological models and to evaluate the hydrological modeling performance. The GRDC database contains streamflow data from 1150 African stations. In this study, only 350 stations were chosen based on three criteria. First, stations should have data for the 1983-2012 study period. Second, stations that have less than five consecutive years of data during this period were excluded. Finally, all the stations should be compatible with the selected HydroSHEDS catchments. Consequently, the 1145 catchments were divided into 350 gauged and 795 ungauged catchments. The climatological data from 9 precipitation and 2 temperature datasets were then extracted for each of the 1145 catchments.

3.1 Hydrological modeling

Given the large-scale nature of this study, distributed and physically-based models were not considered. Two lumped hydrological models, GR4J and HMETS, were selected and calibrated over each of the 350 gauged catchments. The two hydrological models have been shown to perform well in a wide range of studies and over a wide range of climate zones (Arsenault *et al.* (2018); Martel *et al.* (2017); Tarek *et al.* (2019); Valéry *et al.* (2014)).

3.1.1 The GR4J hydrological model

The GR4J (Génie Rural à 4 paramètres Journalier) model is a four-parameter lumped and conceptual rainfall-runoff model (Perrin *et al.* (2003)). This model has shown overall good performance in several studies across the globe (Aubert *et al.* (2003); Raimonet *et al.* (2018); Riboust *et al.* (2019); Westra *et al.* (2014); Youssef *et al.* (2018)). The model requires daily precipitation, temperature and potential evapotranspiration (PET) as inputs to simulate the streamflow.

The Oudin formulation (Oudin *et al.* (2005)) was used in the present study to compute the daily PET series as it was shown to be simple and efficient.

3.1.2 The HMETS hydrological model

The HMETS hydrological model (Hydrological Model – École de technologie supérieure; Martel *et al.* (2017)) is more complex than GR4J, with 21 model parameters. It has four reservoirs (surface runoff, hypodermic flow from the vadose zone reservoir, delayed runoff from infiltration and groundwater flow from the phreatic zone reservoir). HMETS uses the same Oudin PET formulation, but with scaling parameters to control the mass balance.

3.1.3 Hydrological model calibration

The nine precipitation and two temperature datasets were combined in their eighteen possible arrangements for analysis purposes. Due to the large number of calibrations to be performed (41,220 model calibrations), an automatic model parameter calibration approach was selected. The Covariance Matrix Adaptation Evolution Strategy (CMAES) algorithm was chosen because of its flexibility and robustness (Hansen *et al.* (2003)). CMAES has been shown to be one of the best and fastest automatic calibration algorithms available (Arsenault *et al.* (2014); Yu *et al.* (2013)).

All 30 years were used for calibration, and no validation step was performed following the work of Arsenault *et al.* (2018). They showed that validation and calibration skills are not necessarily correlated, and that adding more years to the calibration dataset improves the hydrological model performance and robustness. The calibration objective function was the Kling-Gupta efficiency (KGE) metric, introduced by Gupta *et al.* (2009) and modified by Kling *et al.* (2012). It is defined as a combination of equally-weighted bias, variance and correlation aggregate metrics. The KGE values theoretically range from negative infinity, implying an extremely poor performance of the model, all the way to one, suggesting a perfect performance. Pechlivani-

dis & Arheimer (2015) divided the KGE values into three performance groups: Bad (KGE < 0.4), Acceptable ($0.4 \leq$ KGE < 0.7) and Good (KGE ≥ 0.7).

3.2 Regionalization

The transfer of hydrological information (i.e., model parameters or streamflow) from one catchment (gauged) to another (ungauged) is known as "Regionalization" (Razavi & Coulibaly (2013)). Regionalization can be conducted using two methods: 1) rainfall-runoff models/model-dependent method, which typically transfers the model parameters from one or more gauged watersheds to an ungauged watershed, and 2) hydrological model-independent methods, which transfer the streamflow directly from gauged to ungauged watersheds (Razavi & Coulibaly (2013)). In this paper, the model-dependent method was applied as it has been used in several studies and has shown acceptable results (Merz & Blöschl (2004); McIntyre *et al.* (2005); Boughton & Chiew (2007); Cutore *et al.* (2007); Samaniego *et al.* (2010); Beck *et al.* (2016); Arsenault & Brissette (2014b); Saadi *et al.* (2019)).

The three approaches, namely, the spatial proximity (S.P), physical similarity (P.S) and multilinear regression (MLR) methods (Oudin *et al.* (2008)), have been used to estimate the model parameters in ungauged catchments. First, the three approaches were tested to find the best method to apply. Then, the best-performing precipitation-temperature datasets combination were used to feed the hydrological models and simulate the streamflow of the ungauged catchments. Based on the hydrological modeling performance on the 350 gauged catchments, the MSWEP precipitation and ERA5 temperature datasets were found to be the best combination used in computing the streamflow for the 795 ungauged catchments.

3.3 Bias correction

In this study, the N-dimensional multivariate bias correction algorithm (MBCn) was used (Cannon (2018)). MBCn is an image processing technique extension that preserves the change between the historical and projected periods for all quantiles of the distribution. The algorithm consists of three main steps: (1) application of an orthogonal rotation to both model

and observational data; (2) correction of the marginal distributions of the rotated model data using quantile mapping, and (3) application of an inverse rotation to the results. These three steps are repeated until the model distribution matches the observational distribution. MBCn is arguably the best-performing quantile-based method available (Adeyeri *et al.* (2020); Meyer *et al.* (2019)).

3.4 Variance analysis

An n-dimensional analysis of variance was performed for the 51 streamflow metrics defined in Arsenault *et al.* (2020) for each of the 1145 catchments. These metrics cover a wide range of streamflow conditions: mean annual, seasonal and monthly values, distribution quantiles, as well as low- and high-flow extreme metrics. A variance was attributed to each of the four groups under study, namely, GCM, precipitation dataset, temperature dataset and hydrological model. A total of 11 variance components were computed: 4 main effect components, 6 first-order, 3 second-order, and 1 third-order interaction components.

4. Results

This section outlines the main findings of the work. Figure IV-3 presents the calibration results for both hydrological models using all possible combinations of the 9 precipitation and 2 temperature datasets. Each boxplot consists of 350 KGE values corresponding to the calibration result for each of the 350 selected gauged catchments. Each box extends from the 25^{th} quantile to the 75^{th} quantile, with the median displayed as the red line within that range. The top and bottom whiskers (where shown) represent highest and lowest values. Red crosses are considered statistical outliers.

Results show that both hydrological models perform well, but that there are important differences between datasets. HMETS performs better than GR4J, with respective overall mean KGEs of 0.58 and 0.41. All the precipitation and temperature datasets result in acceptable median KGE simulations Pechlivanidis & Arheimer (2015).

Both temperature datasets perform very similarly across all combinations, with ERA5 generally slightly outperforming CPC. Figure 3 clearly shows that most of the variability seen originates from the precipitation datasets. Four precipitation datasets are ahead of the field. They are, in order of performance: the merged product MSWEP, followed by the two satellite datasets; CHIRPS and PERSIANN, and the ERA5 reanalysis dataset. The gauge-based precipitation datasets (e.g., GPCC and CPC), and the ERA-I reanalysis follow with a similar performance. Finally, the CFSR and JRA55 reanalysis are the worst-performing products for hydrological model calibration.

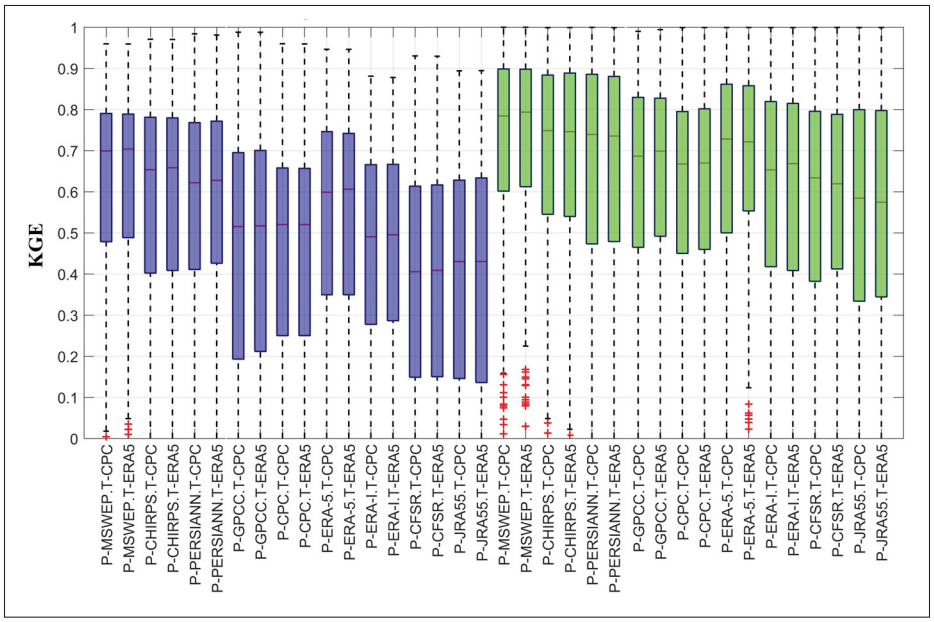


Figure-A IV-3 KGE calibration values using the 18 possible combinations of precipitation and temperature datasets, for both hydrological models (GR4J in blue and HMETS in green) for each of the 350 selected gauged catchments

Table IV-3 presents the main results of the analysis of variance for the 2071-2100 period for the gauged catchments. It shows the relative variance for all main effect (P, GCM, temperature (T) and HM), first-order interactions of the four components of uncertainty under study, and for 6 streamflow metrics (shown in rows 5 to 10). The variance originating from second- and third-order interactions are summed up and presented in the last row. QQ5 and QQ95 are respectively

the 5^{th} and 95^{th} quantiles of streamflow distribution. QX1 is the 30-year mean of the annual daily maximum streamflow value. Results show that most of the variance consistently comes from 5 sources, for all 6 streamflow metrics. They are: precipitation datasets (P), GCMs, hydrological models (HM), interactions between precipitation datasets and GCMs (P-GCM) as well as interactions between precipitation datasets and hydrological models (P-HM). The colored-rows outline the main contributors to variance.

Table-A IV-3 Mean percentage of variance for 6 streamflow metrics for 350 gauged catchments

	Mean Q	**Winter Q**	**Summer Q**	**QQ5**	**QQ95**	**QX1**	**Average**
P	21.62	24.12	28.54	34.38	23.17	22.36	25.70
GCM	39.71	24.93	27.29	4.39	39.56	25.82	26.95
T	0.17	0.12	0.09	0.02	0.15	0.04	0.09
HM	5.18	8.43	19.99	21.96	5.59	5.50	10.11
P-GCM	21.55	25.19	10.20	3.42	16.01	26.33	17.12
P-T	0.02	0.01	0.02	0.01	0.02	0.01	0.015
P-HM	7.38	9.72	14.69	31.12	8.17	8.78	12.31
GCM-T	0.01	0.01	0.006	0.0018	0.017	0.005	0.008
GCM-HM	1.30	2.13	1.44	1.36	2.49	3.49	2.04
T-HM	0.0087	0.0098	0.0069	0.0041	0.0189	0.0058	0.009
Others	2.78	5.20	3.46	2.99	4.60	7.58	4.43

Table IV-3 indicates that both the precipitation datasets and GCMs are the main contributors to variance, including through interactions (P-GCM). The hydrology models also generate some uncertainty, and in particular, through interaction with the precipitation datasets. All metrics exhibit a similar pattern, with the exception of the low-flow metric (QQ5), where precipitation, hydrological models and their interaction components (P-HM) are dominant, and for which GCM uncertainty is minimal. In almost all cases, the five highlighted components represent approximately 85% of the total variance. The average amount of variance introduced by both temperature datasets is less than 0.25% for all 6 different streamflow metrics.

To show cross-catchment variability, Figure IV-4 shows boxplots of the relative variance attribution results for the 5 main contributors to variance, as identified in Table IV-3, and for the

same 6 streamflow metrics. The results are also decomposed into three parts: all 1145 catchments (A), as well as the 350 gauged (G) and 795 ungauged (U) catchments, in order to ensure that the regionalization process does not introduce undesirable effects on the results.

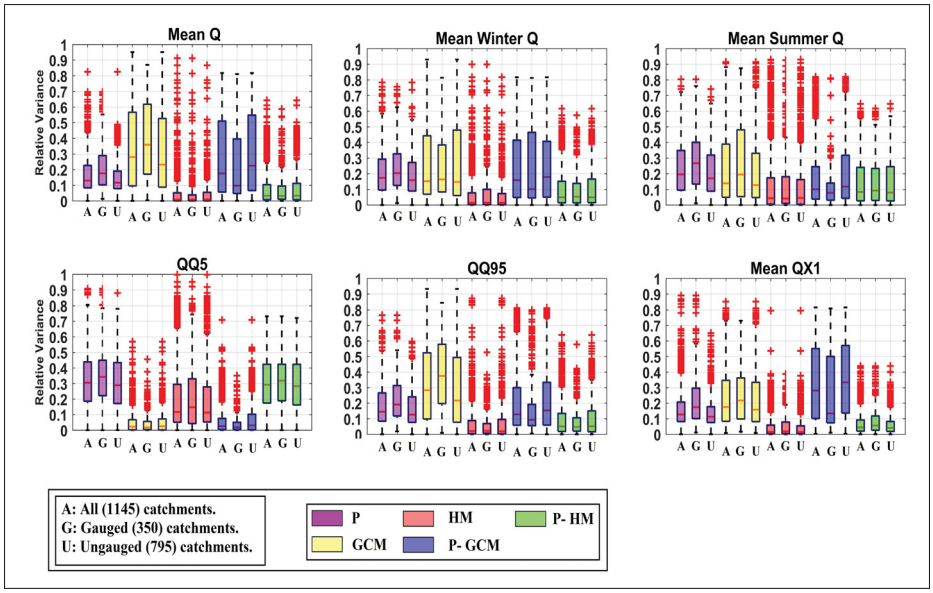


Figure-A IV-4 Boxplots of the relative variance attribution results for the five main contributors to overall variance (P, GCM, HM, P-GCM and P-HM) and 6 streamflow metrics. Relative variance is shown for all 1145 catchments: (A), 350 gauged (G) and 795 ungauged (U) catchments

Figure IV-4 shows that the response of the gauged and ungauged catchments is very similar across all variance components and streamflow metrics, and that no major variance artifact is introduced by the regionalization step. Consequently, all further results will only be shown for all 1145 catchments, with no differentiation made between the gauged and ungauged ones.

The results show that there is considerable across-catchment variability, as shown by the extent of the boxplots, with GCM and P-GCM interaction being the most important, and most variable contributors to variance. As was shown in Table IV-3, the low-flow metric displays a pattern that is much different from the other five metrics, with HM being important and GCM, being the lowest. There is a relatively large difference between the two metrics representing high

flows (Q95 and Q1X). While GCM dominates the former, a much larger part of the uncertainty is transferred to the precipitation dataset (P and P-GCM) for the latter.

In order to study the impact of spatial variability, Figure IV-5 presents the spatial distribution of the relative variance attribution for the five main contributors to variance of Table 3 and all 6 streamflow metrics. Mean Q, Winter Q, QX1 and QQ95 display somewhat similar spatial patterns. Summer Q and QQ5 metrics display somewhat similar spatial patterns. The largest precipitation uncertainty (P and P-GCM interactions) is found in the northern parts of Sub-Saharan Africa, between 0 and 30 °N. GCM uncertainty appears to be larger all around the coastlines of Africa. Hydrological model uncertainty is strongest for QQ5, but spatial patterns are fairly consistent across all 6 streamflow metrics. GCM uncertainty is strongly different for both Summer Q and Winter Q, likely because of the monsoon pattern. Above 20°N, there is generally less than 100 mm of total annual precipitation, and some level of care should therefore be taken when analyzing results in relative contribution to variance.

In other words, a variance analysis of a metric with very little absolute variance could be misleading. Consequently, Figure IV-6 displays the standard deviation of the 360 streamflow values computed for each streamflow metric and for each watershed. The streamflow value for each metric is normalized per unit area to allow for a comparison of large and small watersheds in the same figure. Not surprisingly, the results demonstrate a larger variance along the equatorial band, where precipitation is largest. This pattern is particularly clear for the QQ95 high-flow metric. The catchment database is, however, large enough to show some catchments which exhibit a large variance, even in arid regions above 20°N and below 20°S.

Since some precipitation datasets are clearly better than others based on the hydrological model calibration results, it may not be entirely fair to compare precipitation uncertainty to GCM uncertainty. To investigate this further, the uncertainty contribution obtained when using all 9 precipitation datasets is compared to that of 3 sub-ensembles, as presented in Table IV-4. While ensemble 4 is composed of the clearly best-performing datasets for model calibration,

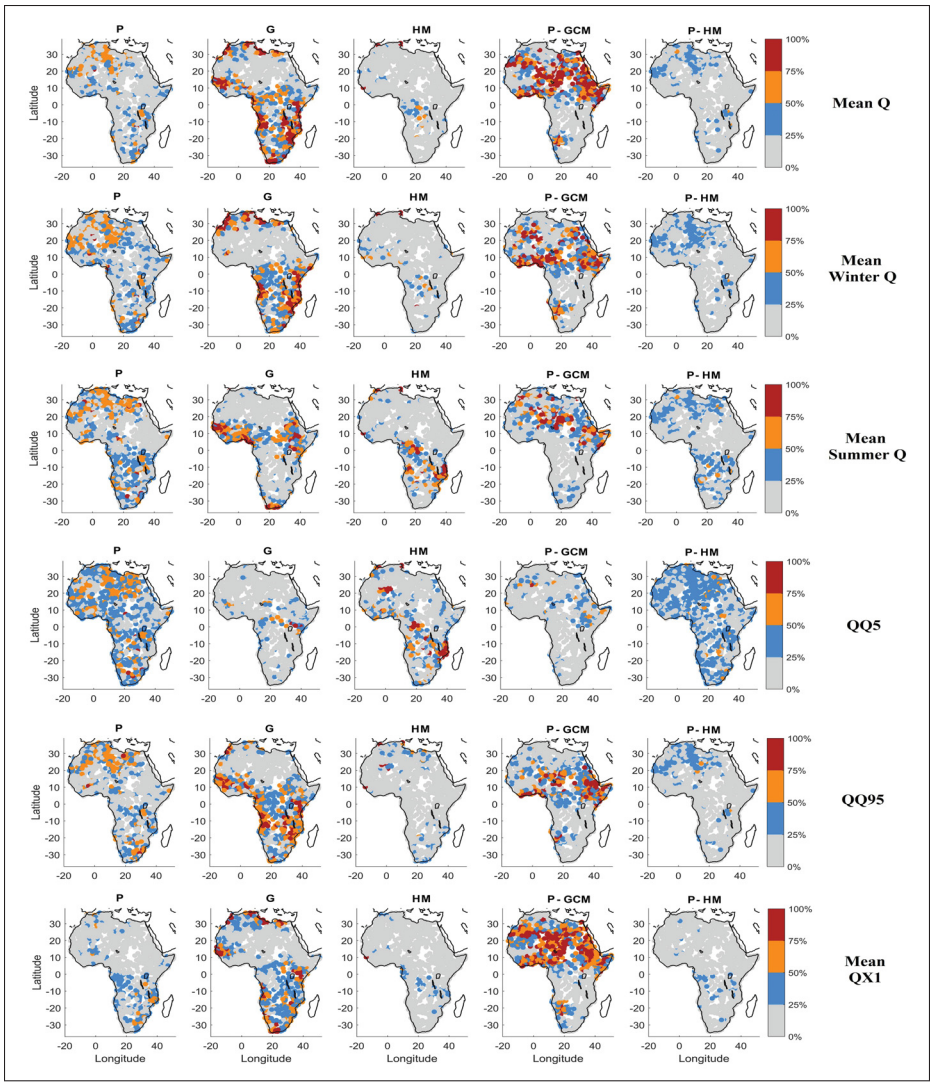


Figure-A IV-5 Spatial distribution of the five main contributors to variance for each of the 6 streamflow metrics

the main goal here is to investigate the impact of dataset selection, not the definition of a credibility ensemble, as will be further discussed later.

Figure IV-7 presents the boxplots of percentages of variance for each catchment, for the five main contributors to variance for all 4 precipitation dataset ensembles of Table IV-4. Unsurprisingly, it shows that reducing the size of the precipitation ensemble results in a consistent decrease in the variance attributed to precipitation. Most of this reduction in variance comes from the P-GCM interaction term, although there is also a noticeable decrease in the main

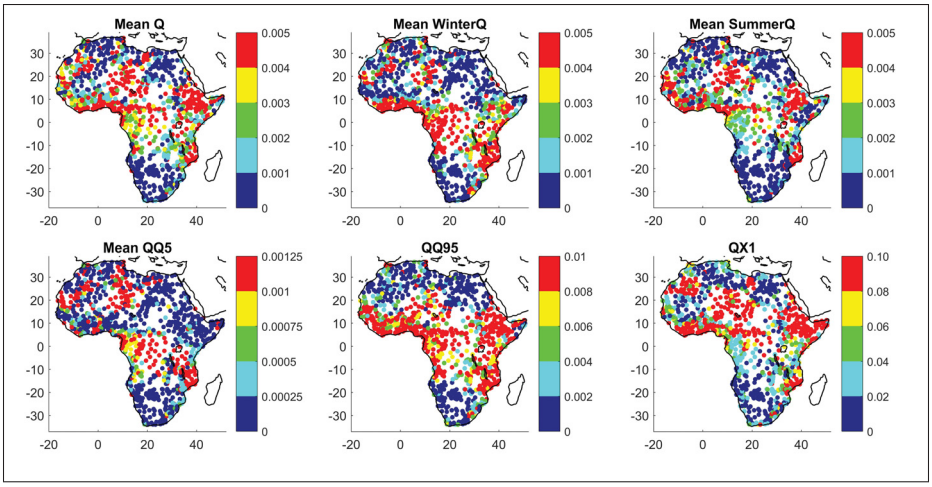


Figure-A IV-6 Standard deviation of discharge per unit area (in $m^3/sec/km^2$), constructed from 360 values for each catchment and streamflow metric

Table-A IV-4 List of ensembles of the different precipitation datasets

Ensemble	Number of precipitation datasets	Rationale for selection	Datasets included	Datasets excluded
1	9	All 9	All	None
2	7	Mean KGE≥0.65	MSWEP, CHIRPS, PERSIANN, ERA5 GPCC, CPC, and ERA-I	CFSR, and JRA55
3	4	Best of each category (merged, satellite, gauge, and reanalysis	MSWEP, CHIRPS, GPCC, and ERA5	PERSIANN, CPC, ERA-I, CFSR, and JRA55
4	4	Best 4	MSWEP, CHIRPS, PERSIANN, and ERA5	GPCC, CPC, ERA-I, CFSR, and JRA55

effect P component. The lost precipitation variance is transferred mostly to GCMs, and to a lesser extent, to hydrological modeling. The exception is the low-flow QQ5, where most of the variance is transferred to HM. Most of the drop observed is obtained by dropping the five worst

precipitation datasets, as no significant difference is observed between precipitation ensembles 3 and 4. Even in a reduced ensemble, precipitation datasets still provide between 10 to 20% of median variance, and more than 30% for the low-flow metric (QQ5) when taking into account the main effect and first-order interaction term.

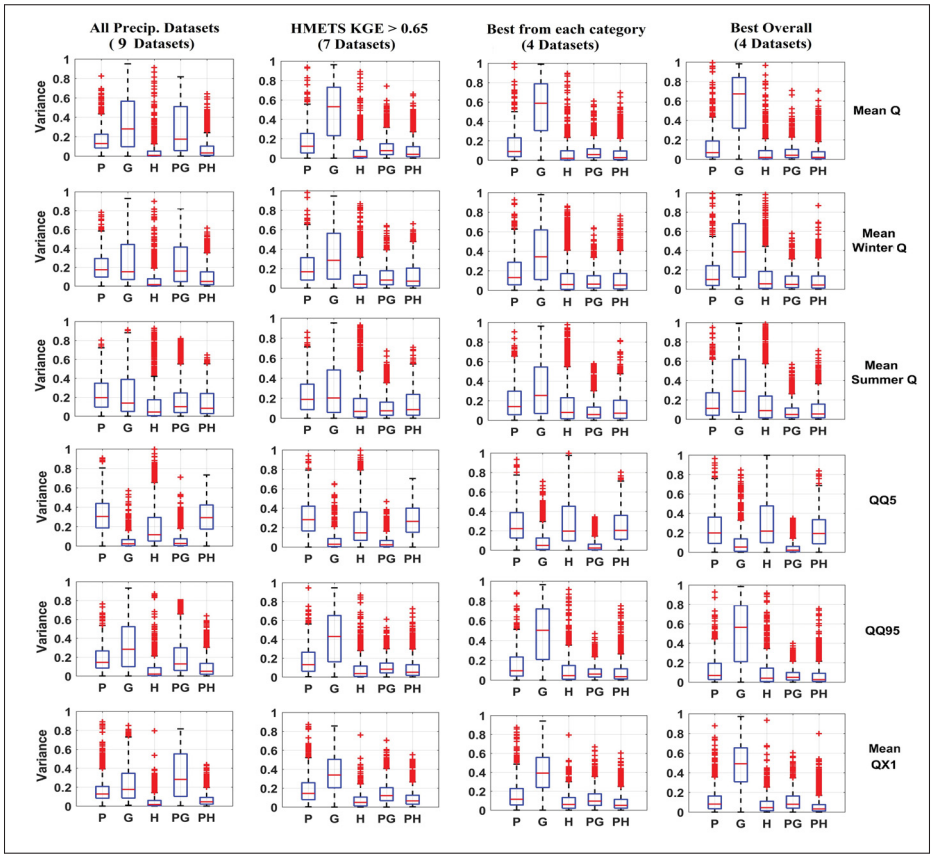


Figure-A IV-7 Boxplots of the five main components of the variance attribution: precipitation (P), GCMs (G), hydrological models (H), interaction between precipitation datasets and GCMs (PG) and interaction between precipitation datasets and hydrological models (PH). Columns represent the four precipitation ensembles of Table IV-4, while rows represent the 6 hydrological indices investigated in this study

Figure IV-8 presents the spatial distribution of the relative variance attribution for each of the 6 streamflow metrics after including only the four best overall precipitation datasets (Ensemble 4 of Table 4). This is the same as Figure IV-5, but with a reduced precipitation ensemble.

Results outline that GCM uncertainty is the dominant source of uncertainty when using the reduced precipitation ensemble, with the exception of the low-flow metric, for which precipitation uncertainty remains dominant. There are, however, significant interactions between GCM and precipitation for all metrics, especially in the northern half of the continent. Otherwise, the observed spatial patterns are similar to the ones presented in Figure IV-5.

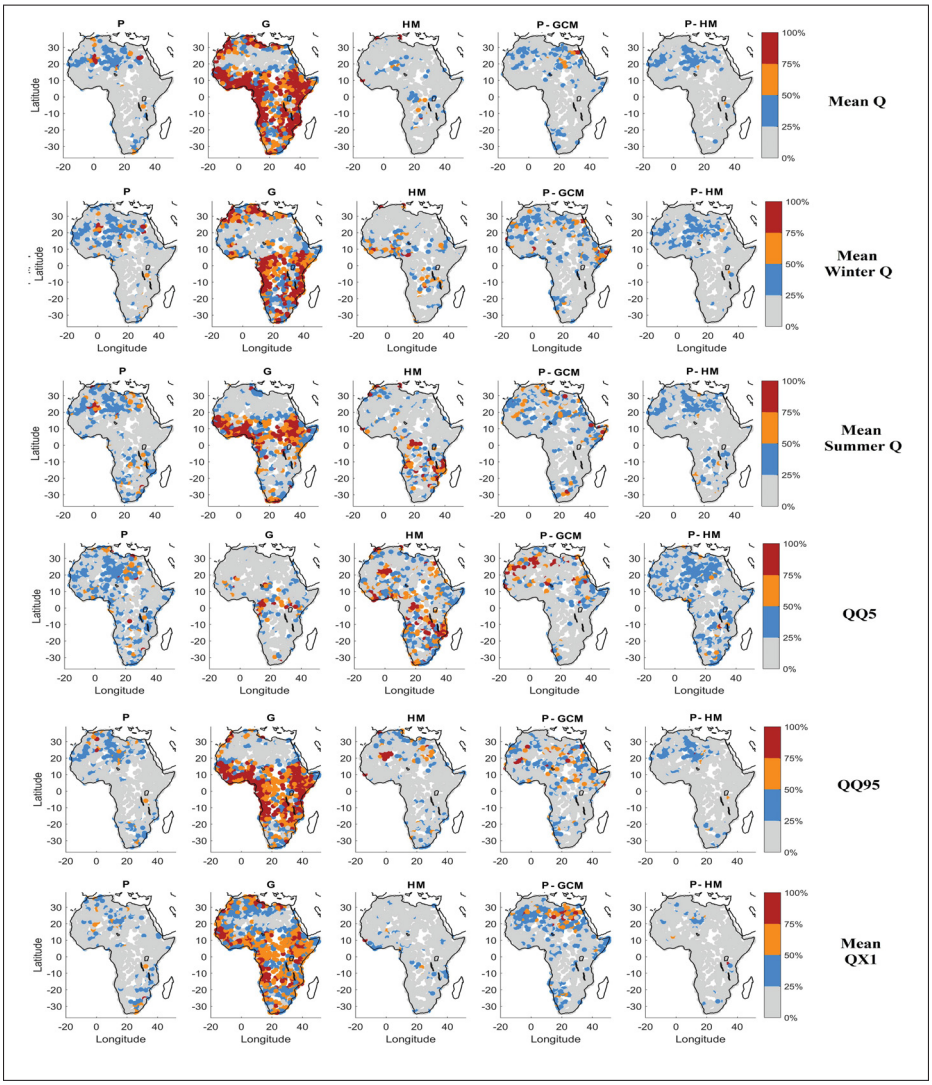


Figure-A IV-8 Spatial distribution of the five main contributors to variance for each of the 6 streamflow metrics, using the 4 best precipitation datasets (Ensemble 4 of Table IV-4)

5. Discussion

Defining a reference climate dataset is an important but difficult task. A reference climate dataset is used as a benchmark for monitoring environmental changes and correcting climate model biases of future climate projections to assess future impacts of a changing climate. Data from weather stations is still mostly considered to be the most accurate representation of the current climate, despite suffering from several important issues, such as precipitation under-catch and inhomogeneities (Peterson *et al.* (1998)). To allow for regular data coverage and remove missing data, it is now common practice to interpolate station data onto a regular grid. Such gridded datasets greatly simplify the processing of meteorological data for environmental studies at the regional, continental and global scales. However, even in regions with a good weather station coverage, gridded datasets using the same underlying data differ due to the different interpolation methods (Essou *et al.* (2016a)), and typically see an increase in the number of wet days and a decrease in the frequency of extreme events (Ensor & Robeson (2008)). In regions with scarce weather station coverage (such as Africa), interpolation becomes extrapolation, and is therefore potentially highly unreliable. In such cases, environmental studies have had to rely on additional sources of data, such as satellite and atmospheric reanalysis for environmental studies. Several inter-comparison studies have been done (e.g., Beck *et al.* (2017b); Essou *et al.* (2017)), including over Africa (Satgé *et al.* (2020); Dembélé *et al.* (2020)). These studies outline a complex picture in which performance depends on scale, climate and data source, and for which no dataset consistently outperforms all of the others. Because of this, in data-sparse regions such as Africa, there is not only no commonly agreed upon reference dataset, but even no agreement on the optimal source of climate data (e.g., satellite vs. reanalysis), and different environmental studies have used completely different datasets. This is particularly problematic for climate change impact studies since there is no knowledge on how dataset uncertainty may propagate in the typical hydroclimatic modeling chain. The results presented in this study attempt to answer this question by comparing dataset uncertainty to other sources of uncertainty, such as that derived from GCMs.

Results show that most of the dataset uncertainty originates from precipitation. Temperature displays much smaller spatial and temporal variability than precipitation, and can therefore

be a lot more reliably interpolated using the adiabatic lapse rate to account for elevation and terrain orientation in mountain areas. Precipitation interpolation is a much more challenging problem, which explains why most dataset intercomparison work has focused on this variable. Based on KGE performance over a common reference period, all nine precipitation datasets performed adequately in terms of hydrological modeling performance, but some clearly performed much better than others. This is in agreement with the results of Beck *et al.* (2017b) and Beck *et al.* (2019). The uncertainty contribution of datasets to future streamflow uncertainty was first evaluated using all 9 precipitation datasets, in conjunction with 2 temperature datasets, a sample of 10 GCMs and two hydrological models, for a total of 360 possible element combinations. While this is a relatively large sample, not all sources of uncertainty were accounted for. In particular, GHGESs, downscaling and bias correction were not included in the analysis. In comparison, the North American Climate Change and Hydroclimatology Dataset (NAC^2H) database (Arsenault *et al.* (2020)) offers 16,000 combinations allowing examining future streamflow uncertainty. In this regard, the relative variance contribution of the climate dataset is best examined in comparison to that of GCMs, the most studied source of climate change impact uncertainty. Results outline the important, and in some cases, dominant contribution of the precipitation dataset to the overall uncertainty of future streamflows. For all 6 streamflow metrics presented here, the precipitation dataset uncertainty was comparable and sometimes larger than that of GCMs.

Uncertainty contribution was then studied by retaining subsets of precipitation datasets, eliminating the least performing ones with respect to the chosen KGE metric. This follows the concept of a credibility ensemble based on carefully selecting the best/most robust components of the hydroclimatic modeling chain, in order to obtain the most credible uncertainty range (e.g., Brissette *et al.* (2020); Giuntoli *et al.* (2018); Maraun *et al.* (2017)). Results demonstrate a large decrease in contribution to uncertainty for 5 of 6 streamflow metrics. The precipitation dataset remained the largest contributor to uncertainty for the low-flow metric, and still accounted for 10 to 20% of the total variance for the other metrics. Most of the decrease in

uncertainty was obtained by dropping the worst-performing datasets, rather than keeping the best-performing ones.

The results presented here indicate that hydrological model uncertainty is relatively small, with the exception of the low-flow metric. These results should be taken with caution because only two hydrological models were used, and also because they both share the same potential evapotranspiration (PET) formula. For climate change impact studies, the climate sensitivity of PET is now thought to be an important source of uncertainty for impact studies (Clark *et al.* (2016); Brissette *et al.* (2020)), and the importance of hydrological model uncertainty has been outlined in many studies (Vetter *et al.* (2017); Krysanova *et al.* (2018); Giuntoli *et al.* (2018)). It is therefore likely that the contribution of hydrological models is underestimated here.

The selection of the best-performing precipitation dataset was evaluated over a reference period using the single metric of the KGE criterion. This criterion is considered to be a good metric as it weighs bias, correlation and RMSE between simulation and observations, all rightfully considered to be important attributes of a good hydrological simulation. There are, however, many other metrics that could have been chosen to perform this comparison, some of which might be even more important for specific applications such as floods or low flows. For example, the JRA55 and CFSR reanalyses were at the bottom of the list of the best-performing datasets presented here. However, in other studies, JRA55 was shown to provide the best reanalysis (Odon *et al.* (2019)), while CFSR was successfully used for precipitation modeling (Khedhaouiria *et al.* (2018)). Clearly, the results presented in this paper should only be used as intended (i.e., to study uncertainty related to the choice of a reference climate dataset), and not as a judgment of the absolute performance of each dataset. As mentioned earlier, it is important to keep in mind that all of the datasets used in this paper generate adequate streamflow simulations.

It is recommended that reference dataset uncertainty be included in climate change impact studies, and especially so in regions with a sparse network of weather stations. We believe that climate dataset uncertainty can be minimized for most streamflow metrics using a careful validation and selection of the best-performing ones. A dataset ensemble should nonetheless

be retained to assess the sensitivity of the impact study to the choice of a reference dataset. As is the case for most other elements of the hydroclimatic modeling chain of future climate change impacts, there is 'no free lunch' in the sense that there is no single recipe, which will be applicable in all cases. Climate dataset performance is spatially-dependent, as shown here and other studies, and will depend on the criteria used to assess said performance. In addition, the relative uncertainty contribution also depends on the catchment location and streamflow metric under study. The importance of first-order interactions in variance analysis, and especially of interactions between precipitation datasets with GCMs and with the hydrology models testify to the complex nature of the propagation of uncertainties in the hydroclimatic modeling chain. The use of an appropriate credibility climate dataset ensemble is therefore more than likely to be catchment-related and metric-dependent, and some minimum level of upstream validation would be needed to select the best components.

6. Conclusion

The main objective of this study was to assess the uncertainty related to the choice of a reference dataset against that of other sources of uncertainty in climate change impact studies. This was achieved by performing a large-sample hydrological climate change impact study over 1145 African catchments. The study used 9 precipitation and 2 temperature datasets, along with 10 GCMs and 2 hydrological models, for a total of 360 possible combinations. Temperature dataset-related uncertainty was minimal; with a median relative contribution to uncertainty, less than 0.25% for all 6 presented streamflow metrics. On the other hand, the nine precipitation dataset ensembles generated a future uncertainty equal to or larger than that related to GCMs. Using a reduced ensemble of the best-performing precipitation datasets systematically reduced the precipitation dataset uncertainty, but still accounted for 10 to 20% of the total variance for 5 of the 6 streamflow metrics, and still remained the main source of uncertainty for the low-flow metric. The main conclusion of this study is that the choice of a climate reference dataset can induce significant uncertainty in climate change impact studies, at least in regions with a sparse weather station coverage.

BIBLIOGRAPHY

Abouabdillah, A., Oueslati, O., De Girolamo, A. M. & Porto, A. L. (2010). Modeling the impact of climate change in a Mediterranean catchment (Merguellil, Tunisia). *Fresenius Environmental Bulletin*, 19(10), 2334–2347.

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R. & Seibert, J. (2014). Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resources Research*, 50(10), 7541–7562.

Adeyeri, O., Laux, P., Lawin, A. & Oyekan, K. (2020). Multiple bias-correction of dynamically downscaled CMIP5 climate models temperature projection: a case study of the transboundary Komadugu-Yobe river basin, Lake Chad region, West Africa. *SN Applied Sciences*, 2(7), 1–18.

Adjei, K. A., Ren, L., Appiah-Adjei, E. K. & Odai, S. N. (2015). Application of satellite-derived rainfall for hydrological modelling in the data-scarce Black Volta trans-boundary basin. *Hydrology Research*, 46(5), 777–791.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K. & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, 116(D2).

Aghakouchak, A. & Habib, E. (2010). Application of a conceptual hydrologic model in teaching hydrologic processes. *International Journal of Engineering Education*, 26(4 (S1)), 963–973.

Ahmed, K. F., Wang, G., Silander, J., Wilson, A. M., Allen, J. M., Horton, R. & Anyah, R. (2013). Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the US northeast. *Global and Planetary Change*, 100, 320–332.

Ajami, N. K., Gupta, H., Wagener, T. & Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of hydrology*, 298(1-4), 112–135.

Ajami, N. K., Duan, Q. & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water resources research*, 43(1).

Akinsanola, A. & Zhou, W. (2019). Projections of West African summer monsoon rainfall extremes from two CORDEX models. *Climate Dynamics*, 52(3-4), 2017–2028.

Alijanian, M., Rakhshandehroo, G. R., Mishra, A. K. & Dehghani, M. (2017). Evaluation of satellite rainfall climatology using CMORPH, PERSIANN-CDR, PERSIANN, TRMM, MSWEP over Iran. *International Journal of Climatology*, 37(14), 4896–4914.

Anandhi, A., Frei, A., Pierson, D. C., Schneiderman, E. M., Zion, M. S., Lounsbury, D. & Matonse, A. H. (2011). Examination of change factor methodologies for climate change impact assessment. *Water Resources Research*, 47(3).

Andermann, C., Bonnet, S. & Gloaguen, R. (2011). Evaluation of precipitation data sets along the Himalayan front. *Geochemistry, Geophysics, Geosystems*, 12(7).

Arsenault, R., B., F., C., J., D., G., G. & Q. (2020). NAC2H: The North-American Climate Change and hydroclimatology dataset. *Water Resources Research*.

Arsenault, R. & Brissette, F. (2014a). Determining the optimal spatial distribution of weather station networks for hydrological modeling purposes using RCM datasets: An experimental approach. *Journal of Hydrometeorology*, 15(1), 517–526.

Arsenault, R. & Brissette, F. P. (2014b). Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research*, 50(7), 6135–6153.

Arsenault, R., Poulin, A., Côté, P. & Brissette, F. (2014). Comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of Hydrologic Engineering*, 19(7), 1374–1384.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F. & Martel, J.-L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, 529, 754–767.

Arsenault, R., Bazile, R., Ouellet Dallaire, C. & Brissette, F. (2016). CANOPEX: A Canadian hydrometeorological watershed database. *Hydrological Processes*, 30(15), 2734–2736.

Arsenault, R., Brissette, F. & Martel, J.-L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of hydrology*, 566, 346–362.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R. & Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1), 69–83.

Aslami, F., Ghorbani, A., Sobhani, B. & Esmali, A. (2019). Comprehensive comparison of daily IMERG and GSMaP satellite precipitation products in Ardabil Province, Iran. *International Journal of Remote Sensing*, 40(8), 3139–3153.

Asong, Z., Razavi, S., Wheater, H. & Wong, J. (2017). Evaluation of integrated multisatellite retrievals for GPM (IMERG) over southern Canada against ground precipitation observations: A preliminary assessment. *Journal of hydrometeorology*, 18(4), 1033–1050.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P. J., Rötter, R. P., Cammarano, D. et al. (2013). Uncertainty in simulating wheat yields under climate change. *Nature climate change*, 3(9), 827–832.

Aubert, D., Loumagne, C. & Oudin, L. (2003). Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. *Journal of Hydrology*, 280(1-4), 145–161.

Awange, J., Ferreira, V., Forootan, E., Andam-Akorful, S., Agutu, N. & He, X. (2016). Uncertainties in remotely sensed precipitation data over Africa. *International Journal of Climatology*, 36(1), 303–323.

Ayar, P. V., Vrac, M., Bastin, S., Carreau, J., Déqué, M. & Gallardo, C. (2016). Intercomparison of statistical and dynamical downscaling models under the EURO-and MED-CORDEX initiative framework: present climate evaluations. *Climate Dynamics*, 46(3-4), 1301–1329.

Bâ, K. M., Diaz, V., Gómez-Albores, M. A., Díaz-Delgado, C., Nájera-Mota, N., Seidou, O. & Ortiz, F. (2018). Spatially distributed hydrological modelling of a Western Africa basin. *EPiC Ser. Eng*, 3, 343–350.

Bae, D.-H., Jung, I.-W. & Lettenmaier, D. P. (2011). Hydrologic uncertainties in climate change from IPCC AR4 GCM simulations of the Chungju Basin, Korea. *Journal of Hydrology*, 401(1-2), 90–105.

Baguis, P., Roulin, E., Willems, P. & Ntegeka, V. (2010). Climate change scenarios for precipitation and potential evapotranspiration over central Belgium. *Theoretical and applied climatology*, 99(3-4), 273.

Balsamo, G., Dutra, E., Albergel, C., Munier, S., Calvet, J.-C., Munoz-Sabater, J. & de Rosnay, P. (2018). ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? *Hydrology and Earth System Sciences*, 22(6), 3515–3532.

Bastola, S. & François, D. (2012). Temporal extension of meteorological records for hydrological modelling of Lake Chad Basin (Africa) using satellite rainfall data and reanalysis datasets. *Meteorological Applications*, 19(1), 54–70.

Bastola, S., Murphy, C. & Sweeney, J. (2011). The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Advances in Water Resources*, 34(5), 562–576.

Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J. & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52(5), 3599–3622.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Gonzalez Miralles, D., Martens, B. & De Roo, A. (2017a). MSWEP: 3-hourly 0.25 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), 589–615.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J. & Wood, E. F. (2017b). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12), 6201.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A. & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data*, 5, 180214.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., Huffman, G. J., Adler, R. F. & Wood, E. F. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23(1), 207–224.

Behrangi, A., Khakbaz, B., Jaw, T. C., AghaKouchak, A., Hsu, K. & Sorooshian, S. (2011). Hydrologic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology*, 397(3-4), 225–237.

Bengtsson, L., Hagemann, S. & Hodges, K. I. (2004). Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres*, 109(D11).

Bevelhimer, M. S., McManamay, R. A. & O'connor, B. (2015). Characterizing sub-daily flow regimes: implications of hydrologic resolution on ecohydrology studies. *River research and applications*, 31(7), 867–879.

Boé, J., Terray, L., Habets, F. & Martin, E. (2007). Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(12), 1643–1655.

Boers, N., Bookhagen, B., Marengo, J., Marwan, N., von Storch, J.-S. & Kurths, J. (2015). Extreme rainfall of the South American monsoon system: a dataset comparison using complex networks. *Journal of Climate*, 28(3), 1031–1056.

Bosilovich, M. G. (2013). Regional climate and variability of NASA MERRA and recent reanalyses: US summertime precipitation and temperature. *Journal of applied meteorology and climatology*, 52(8), 1939–1951.

Bosilovich, M. G., Chen, J., Robertson, F. R. & Adler, R. F. (2008). Evaluation of global precipitation in reanalyses. *Journal of applied meteorology and climatology*, 47(9), 2279–2299.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M. & Schär, C. (2013). Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resources Research*, 49(3), 1523–1536.

Boughton, W. & Chiew, F. (2007). Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model. *Environmental Modelling & Software*, 22(4), 476–487.

Brissette, F., Arsenault, R. & Chen, J. (2020). The potential pitfalls of hydrological climate change impact studies. *Geophysical Research Letters*.

Buarque, D. C., de Paiva, R. C. D., Clarke, R. T. & Mendes, C. A. B. (2011). A comparison of Amazon rainfall characteristics derived from TRMM, CMORPH and the Brazilian national rain gauge network. *Journal of Geophysical Research: Atmospheres*, 116(D19).

Burn, D. H., Hannaford, J., Hodgkins, G. A., Whitfield, P. H., Thorne, R. & Marsh, T. (2012). Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow. *Hydrological Sciences Journal*, 57(8), 1580–1593.

Burroughs, W. & Burroughs, W. S. (2003). *Climate: Into the 21st century*. Cambridge University Press.

Cannon, A. J. (2016). Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *Journal of Climate*, 29(19), 7045–7064.

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate dynamics*, 50(1-2), 31–49.

Cannon, A. J., Sobie, S. R. & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17), 6938–6959.

Chaudhuri, A. H., Ponte, R. M., Forget, G. & Heimbach, P. (2013). A comparison of atmospheric reanalysis surface products over the ocean and implications for uncertainties in air–sea boundary forcing. *Journal of Climate*, 26(1), 153–170.

Chen, H., Xu, C.-Y. & Guo, S. (2012). Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *Journal of hydrology*, 434, 36–45.

Chen, J., Brissette, F. P. & Leconte, R. (2011a). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of hydrology*, 401(3-4), 190–202.

Chen, J., Brissette, F. P., Poulin, A. & Leconte, R. (2011b). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research*, 47(12).

Chen, J., Brissette, F. P. & Chen, H. (2018). Using reanalysis-driven regional climate model outputs for hydrology modelling. *Hydrological Processes*, 32(19), 3019–3031.

Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R. & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research: Atmospheres*, 113(D4).

Chen, S., Liu, H., You, Y., Mullens, E., Hu, J., Yuan, Y., Huang, M., He, L., Luo, Y., Zeng, X. et al. (2014). Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations. *PloS one*, 9(4).

Chen, Z., Qin, Y., Shen, Y. & Zhang, S. (2016). Evaluation of global satellite mapping of precipitation project daily precipitation estimates over the Chinese mainland. *Advances in Meteorology*, 2016.

Citterio, M., van As, D., Ahlstrøm, A. P., Andersen, M. L., Andersen, S. B., Box, J. E., Charalampidis, C., Colgan, W. T., Fausto, R. S., Nielsen, S. et al. (2015). Automatic weather stations for basic and applied glaciological research. *Geological Survey of Denmark and Greenland Bulletin*, 69–72.

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R. & Brekke, L. D. (2016). Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, 2(2), 55–64.

Clifford, D. (2010). Global estimates of snow water equivalent from passive microwave instruments: history, challenges and future developments. *International Journal of Remote Sensing*, 31(14), 3707–3726.

Cole, S. J. & Moore, R. J. (2009). Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Advances in Water Resources*, 32(7), 1107–1120.

Collischonn, B., Collischonn, W. & Tucci, C. E. M. (2008). Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*, 360(1-4), 207–216.

Colston, J. M., Ahmed, T., Mahopo, C., Kang, G., Kosek, M., de Sousa Junior, F., Shrestha, P. S., Svensen, E., Turab, A., Zaitchik, B. et al. (2018). Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study. *Environmental research*, 165, 91–109.

Cutore, P., Cristaudo, G., Campisano, A., Modica, C., Cancelliere, A. & Rossi, G. (2007). Regional models for the estimation of streamflow series in ungauged basins. *Water resources management*, 21(5), 789–800.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P. et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656), 553–597.

Dembélé, M., Schaefli, B., van de Giesen, N. & Mariéthoz, G. (2020). Suitability of 17 rainfall and temperature gridded datasets for largescale hydrological modelling in West Africa. *Hydrology and Earth System Sciences Discussions*, 1–39.

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., De Castro, M. & van den Hurk, B. (2007). An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change*, 81(1), 53–70.

Deser, C., Phillips, A., Bourdette, V. & Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate dynamics*, 38(3-4), 527–546.

Devia, G. K., Ganasri, B. & Dwarakish, G. (2015). A review on hydrological models. *Aquatic Procedia*, 4, 1001–1007.

Di Baldassarre, G. & Montanari, A. (2009). Uncertainty in river discharge observations: a quantitative analysis. *Hydrology & Earth System Sciences*, 13(6).

Di Giuseppe, F., Pappenberger, F., Wetterhall, F., Krzeminski, B., Camia, A., Libertá, G. & San Miguel, J. (2016). The potential predictability of fire danger provided by numerical weather prediction. *Journal of Applied Meteorology and Climatology*, 55(11), 2469–2491.

Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K. & Arnold, J. G. (2008). Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology*, 47(2), 475–497.

Diaz-Nieto, J. & Wilby, R. L. (2005). A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. *Climatic Change*, 69(2-3), 245–268.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. & Ropelewski, C. (2007). Validation of satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*, 28(7), 1503–1526.

Do, H. X., Westra, S. & Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. *Journal of hydrology*, 552, 28–43.

Dobler, C., Hagemann, S., Wilby, R. & Stätter, J. (2012). Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed. *Hydrology and Earth System Sciences*, 16, 4343–4360.

Donnelly, C., Dahné, J., Rosberg, J., Strömqvist, J., Yang, W. & Arheimer, B. (2010). High-resolution, large-scale hydrological modelling tools for Europe. *IAHS Publ*, 340, 553–561.

dos Santos, F. M., de Oliveira, R. P. & Mauad, F. F. (2018). Lumped versus distributed hydrological modeling of the Jacaré-Guaçu Basin, Brazil. *Journal of Environmental Engineering*, 144(8), 04018056.

Droogers, P., Immerzeel, W., Terink, W., Hoogeveen, J., Bierkens, M., Van Beek, L. & Debele, B. (2012). Water resources trends in Middle East and North Africa towards 2050. *Hydrology and Earth System Sciences*, 16, 3101–3114.

Duan, Z., Liu, J., Tuo, Y., Chiogna, G. & Disse, M. (2016). Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. *Science of the Total Environment*, 573, 1536–1553.

Duan, Z., Tuo, Y., Liu, J., Gao, H., Song, X., Zhang, Z., Yang, L. & Mekonnen, D. F. (2019). Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia. *Journal of hydrology*, 569, 612–626.

Eden, J. M. & Widmann, M. (2014). Downscaling of GCM-simulated precipitation using model output statistics. *Journal of Climate*, 27(1), 312–324.

El Fadli, K. I., Cerveny, R. S., Burt, C. C., Eden, P., Parker, D., Brunet, M., Peterson, T. C., Mordacchini, G., Pelino, V., Bessemoulin, P. et al. (2013). World Meteorological Organization assessment of the purported world record 58 C temperature extreme at El Azizia, Libya (13 September 1922). *Bulletin of the American Meteorological Society*, 94(2), 199–204.

Emerton, R., Cloke, H., Stephens, E., Zsoter, E., Woolnough, S. & Pappenberger, F. (2017). Complex picture for likelihood of ENSO-driven flood hazard. *Nature communications*, 8(1), 1–9.

Ensor, L. A. & Robeson, S. M. (2008). Statistical characteristics of daily precipitation: comparisons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, 47(9), 2468–2476.

Essou, G. R., Arsenault, R. & Brissette, F. P. (2016a). Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of hydrology*, 537, 334–345.

Essou, G. R., Sabarly, F., Lucas-Picher, P., Brissette, F. & Poulin, A. (2016b). Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? *Journal of Hydrometeorology*, 17(7), 1929–1950.

Essou, G. R., Brissette, F. & Lucas-Picher, P. (2017). The use of reanalyses and gridded observations as weather input data for a hydrological model: Comparison of performances of simulated river flows based on the density of weather stations. *Journal of Hydrometeorology*, 18(2), 497–513.

Falck, A. S., Maggioni, V., Tomasella, J., Vila, D. A. & Diniz, F. L. (2015). Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the Tocantins–Araguaia basin in Brazil. *Journal of Hydrology*, 527, 943–957.

Faridi, R. (2016). Climate of North America: General Characteristics and Climate Regions. Consulted at https://rashidfaridi.com/2016/10/24/climate-of-north-america-general-characteristics/.

Fekete, B. M. & Vörösmarty, C. J. (2007). The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS publ*, 309, 129–136.

Fekete, B. M., Vörösmarty, C. J., Roads, J. O. & Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, 17(2), 294–304.

Fortin, V., Roy, G., Donaldson, N. & Mahidjiba, A. (2015). Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *Journal of Hydrology*, 531, 296–307.

Fowler, H. J., Blenkinsop, S. & Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(12), 1547–1578.

Funk, C., Verdin, A., Michaelsen, J., Peterson, P., Pedreros, D. & Husak, G. (2015a). A global satellite assisted precipitation climatology. *Earth System Science Data Discussions*, 8(1).

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. et al. (2015b). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, 2(1), 1–21.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., Verdin, A. P. et al. (2014). A quasi-global precipitation time series for drought monitoring. *US Geological Survey data series*, 832(4), 1–12.

Getirana, A. C., Espinoza, J., Ronchail, J. & Rotunno Filho, O. (2011). Assessment of different precipitation datasets and their impacts on the water balance of the Negro River basin. *Journal of Hydrology*, 404(3-4), 304–322.

Ghavidelfar, S., Alvankar, S. R. & Razmkhah, A. (2011). Comparison of the lumped and quasi-distributed Clark runoff models in simulating flood hydrographs on a semi-arid watershed. *Water resources management*, 25(6), 1775–1790.

Giardino, C., Bresciani, M., Villa, P. & Martinelli, A. (2010). Application of remote sensing in water resource management: the case study of Lake Trasimeno, Italy. *Water resources management*, 24(14), 3885–3899.

Giorgi, F. & Lionello, P. (2008). Climate change projections for the Mediterranean region. *Global and planetary change*, 63(2-3), 90–104.

Giuntoli, I., Villarini, G., Prudhomme, C. & Hannah, D. M. (2018). Uncertainties in projected runoff over the conterminous United States. *Climatic change*, 150(3-4), 149–162.

Goodarzi, E., Dastorani, M., Talebi, A. et al. (2015). Evaluation of the Change-factor and LARS-WG methods of downscaling for simulation of climatic variables in the future (Case study: Herat Azam Watershed, Yazd-Iran). *Ecopersia*, 3(1), 833–846.

Green, A., Thompson, J., Kingston, D. & Gosling, S. (2014). Quantifying the relative magnitude of sources of uncertainty in river flow projections under climate change: an assessment for the Mekong River. *EGUGA*, 15019.

Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C. & Rosas, G. (2017). The influence of station density on climate data homogenization. *International journal of climatology*, 37(13), 4670–4683.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1-2), 80–91.

Gutenson, J. L., Tavakoly, A. A., Wahl, M. D. & Follum, M. L. (2019). Comparison of Generalized Non-Data-Driven Reservoir Routing Models for Global-Scale Hydrologic Modeling. *Hydrology and Earth System Sciences Discussions*, 1–41.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D. et al. (2011). Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology*, 12(5), 869–884.

Hansen, N., Müller, S. D. & Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation*, 11(1), 1–18.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J., Nicholls, N., Penner, J. E. & Stott, P. A. (2007). Understanding and attributing climate change.

Henn, B., Newman, A. J., Livneh, B., Daly, C. & Lundquist, J. D. (2018). An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of hydrology*, 556, 1205–1219.

Henson, R. (2011). *The rough guide to climate change*. Dorling Kindersley Ltd.

Hersbach, H. & Dee, D. (2016). ERA5 reanalysis is in production. *ECMWF newsletter*, 147(7), 5–6.

Hirpa, F. A., Gebremichael, M. & Hopson, T. (2010). Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *Journal of Applied Meteorology and Climatology*, 49(5), 1044–1051.

Hobeichi, S., Abramowitz, G., Contractor, S. & Evans, J. (2020a). Evaluating precipitation datasets using surface water and energy budget closure. *Journal of Hydrometeorology*, 21(5), 989–1009.

Hobeichi, S., Abramowitz, G. & Evans, J. (2020b). Conserving Land–Atmosphere Synthesis Suite (CLASS). *Journal of Climate*, 33(5), 1821–1844.

Hodges, K. I., Hoskins, B. J., Boyle, J. & Thorncroft, C. (2003). A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves. *Monthly Weather Review*, 131(9), 2012–2037.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B. et al. (2019). From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmospheric Chemistry & Physics*, 19(5).

Hofstra, N., New, M. & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate dynamics*, 35(5), 841–858.

Hoskins, B. J. & Hodges, K. I. (2005). A new perspective on Southern Hemisphere storm tracks. *Journal of Climate*, 18(20), 4108–4129.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K. & Iguchi, T. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5), 701–722.

Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (1990). Climate change: the IPCC scientific assessment. *American Scientist;(United States)*, 80(6).

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P. & Yoo, S.-H. (2015). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). *Algorithm Theoretical Basis Document (ATBD) Version*, 4, 26.

Hulme, M., Doherty, R., Ngara, T., New, M. & Lister, D. (2001). African climate change: 1900-2100. *Climate research*, 17(2), 145–168.

Hunger, M. & Döll, P. (2008). Value of river discharge data for global-scale hydrological modeling.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E. & Papadopol, P. (2009). Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003. *Journal of Applied Meteorology and Climatology*, 48(4), 725–741.

Huth, R. (2004). Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. *Journal of Climate*, 17(3), 640–652.

Huth, R., Mikšovskỳ, J., Štěpánek, P., Belda, M., Farda, A., Chládová, Z. & Pišoft, P. (2015). Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature. *Theoretical and Applied Climatology*, 120(3-4), 533–553.

IPCC. (2001). Working Group I: The Scientific Basis. *Third Assessment Report of the Intergovernmental Panel on Climate Change. New York, Cambridge Univ. Press*.

IPCC. (2014). Mitigation of climate change. *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 1454.

Islam, T., Rico-Ramirez, M. A., Han, D., Srivastava, P. K. & Ishak, A. M. (2012). Performance evaluation of the TRMM precipitation estimation using ground-based radars from the GPM validation network. *Journal of Atmospheric and Solar-Terrestrial Physics*, 77, 194–208.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassegues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M. et al. (2014). The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, 34(5), 1657–1675.

James, R. & Washington, R. (2013). Changes in African temperature and precipitation associated with degrees of global warming. *Climatic change*, 117(4), 859–872.

James Wreford Watson, P. F. H. & Others. (2020). North America. Consulted at https://www.britannica.com/place/North-America.

Janis, M. J., Hubbard, K. G. & Redmond, K. T. (2004). Station density strategy for monitoring long-term climatic change in the contiguous United States. *Journal of climate*, 17(1), 151–162.

Janis, M., Hubbard, K. & Redmond, K. (2002). Determining the Optimal Number of Stations for the United States Climate Reference Network Final Report. *Southeast Regional Climate Center, Research Paper Series*.

Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F. & Ma, M. (2012). Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *Journal of Hydrology*, 452, 213–225.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of hydrometeorology*, 5(3), 487–503.

Jutla, A., Aldaach, H., Billian, H., Akanda, A., Huq, A. & Colwell, R. (2015). Satellite based assessment of hydroclimatic conditions related to cholera in Zimbabwe. *PloS one*, 10(9).

Kang, S. & Ahn, J.-B. (2015). Global energy and water balances in the latest reanalyses. *Asia-Pacific Journal of Atmospheric Sciences*, 51(4), 293–302.

Karamage, F., Liu, Y., Fan, X., Francis Justine, M., Wu, G., Liu, Y., Zhou, H. & Wang, R. (2018). Spatial relationship between precipitation and runoff in Africa. *Hydrology and Earth System Sciences Discussions*, 1–27.

Kay, A., Davies, H., Bell, V. & Jones, R. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic change*, 92(1-2), 41–63.

Khan, M. S., Coulibaly, P. & Dibike, Y. (2006). Uncertainty analysis of statistical downscaling methods. *Journal of Hydrology*, 319(1-4), 357–382.

Khan, S. I., Adhikari, P., Hong, Y., Vergara, H., Adler, R., Policelli, F., Irwin, D., Korme, T., Okello, L. et al. (2011). Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data. *Hydrology and Earth System Sciences*, 15(1), 107–117.

Khedhaouiria, D., Mailhot, A. & Favre, A.-C. (2018). Daily precipitation fields modeling across the great lakes region (Canada) by Using the CFSR reanalysis. *Journal of Applied Meteorology and Climatology*, 57(10), 2419–2438.

Kidd, C., Bauer, P., Turk, J., Huffman, G., Joyce, R., Hsu, K.-L. & Braithwaite, D. (2012). Intercomparison of high-resolution precipitation products over northwest Europe. *Journal of Hydrometeorology*, 13(1), 67–83.

Kidd, C. & Huffman, G. (2011). Global precipitation measurement. *Meteorological Applications*, 18(3), 334–353.

Kidd, C., Dawkins, E. & Huffman, G. (2013). Comparison of precipitation derived from the ECMWF operational forecast model and satellite precipitation datasets. *Journal of Hydrometeorology*, 14(5), 1463–1482.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G. & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, 98(1), 69–78.

Kim, Y., Ohn, I., Lee, J.-K. & Kim, Y.-O. (2019). Generalizing uncertainty decomposition theory in climate change impact assessments. *Journal of Hydrology X*, 3, 100024.

Kingston, D. & Taylor, R. (2010). Sources of uncertainty in climate change impacts on river discharge and groundwater in a headwater catchment of the Upper Nile Basin, Uganda. *Hydrology & Earth System Sciences*, 14(7).

Klaes, K. D., Cohen, M., Buhler, Y., Schlüssel, P., Munro, R., Luntama, J.-P., von Engeln, A., Clérigh, E. Ó., Bonekamp, H., Ackermann, J. et al. (2007). An introduction to the EUMETSAT polar system. *Bulletin of the American Meteorological Society*, 88(7), 1085–1096.

Kling, H., Fuchs, M. & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424, 264–277.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H. et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), 5–48.

Kokkonen, T. S. & Jakeman, A. J. (2001). A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resources Research*, 37(9), 2345–2352.

Köppen, W. (1900). Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift*, 6(11. H), 593–611.

Koriche, S. A. & Rientjes, T. H. (2016). Application of satellite products and hydrological modelling for flood early warning. *Physics and Chemistry of the Earth, Parts A/B/C*, 93, 12–23.

Koskela, J. J., Croke, B., Koivusalo, H., Jakeman, A. & Kokkonen, T. (2012). Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment. *Water Resources Research*, 48(11).

Kotsias, G., Lolis, C., Hatzianastassiou, N., Levizzani, V. & Bartzokas, A. (2020). On the connection between large-scale atmospheric circulation and winter GPCP precipitation over the Mediterranean region for the period 1980-2017. *Atmospheric Research*, 233, 104714.

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F. & Kundzewicz, Z. W. (2018). How the performance of hydrological models relates to credibility of projections under climate change. *Hydrological Sciences Journal*, 63(5), 696–720.

Laviola, S., Levizzani, V., Cattani, E. & Kidd, C. (2013). The 183-WSL fast rain rate retrieval algorithm. Part II: Validation using ground radar measurements. *Atmospheric research*, 134, 77–86.

Lawrimore, J. H., Menne, M. J., Gleason, B. E., Williams, C. N., Wuertz, D. B., Vose, R. S. & Rennie, J. (2011). An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *Journal of Geophysical Research: Atmospheres*, 116(D19).

Leggett, J., Pepper, W. J., Swart, R. J., Edmonds, J., Meira Filho, L., Mintzer, I. & Wang, M. (1992). Emissions scenarios for the IPCC: an update. *Climate change*, 69–95.

Lehner, B. (2014). HydroBASINS: Global watershed boundaries and sub-basin delineations derived from HydroSHEDS data at 15 second resolution—Technical documentation version 1. c.

Lehner, B. & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171–2186.

Lehner, B., Verdin, K. & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10), 93–94.

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P. & Stadnyk, T. (2015). Performance evaluation of the Canadian precipitation analysis (CaPA). *Journal of Hydrometeorology*, 16(5), 2045–2064.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M. & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51(9), 7309–7342.

Lindsay, R., Wensnahan, M., Schweiger, A. & Zhang, J. (2014). Evaluation of seven different atmospheric reanalysis products in the Arctic. *Journal of Climate*, 27(7), 2588–2606.

Linke, S., Lehner, B., Dallaire, C. O., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H. et al. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, 6(1), 1–15.

Lins, H. F. (2008). Challenges to hydrological observations. *WMO Bulletin*, 57(1), 55–58.

Liu, Y. & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water resources research*, 43(7).

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R. & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013. *Scientific data*, 2(1), 1–12.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P. & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events.

Lockhoff, M., Zolina, O., Simmer, C. & Schulz, J. (2014). Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations. *Journal of climate*, 27(2), 607–623.

Lorenz, C., Kunstmann, H., Devaraju, B., Tourian, M. J., Sneeuw, N. & Riegger, J. (2014). Large-scale runoff from landmasses: a global assessment of the closure of the hydrological and atmospheric water balances. *Journal of Hydrometeorology*, 15(6), 2111–2139.

Lu, D. & Yong, B. (2018). Evaluation and hydrological utility of the latest GPM IMERG V5 and GSMaP V7 precipitation products over the Tibetan Plateau. *Remote Sensing*, 10(12), 2022.

Luo, C., Wang, Z., Sauer, T. J., Helmers, M. J. & Horton, R. (2018). Portable canopy chamber measurements of evapotranspiration in corn, soybean, and reconstructed prairie. *Agricultural Water Management*, 198, 1–9.

Maraun, D. (2016). Bias correcting climate change simulations-a critical review. *Current Climate Change Reports*, 2(4), 211–220.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A. et al. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11), 764–773.

Martel, J.-L., Demeester, K., Brissette, F. P., Arsenault, R. & Poulin, A. (2017). HMET: a simple and efficient hydrology model forteaching hydrological modelling, flow forecasting and climate change impacts. *The International journal of engineering education*, 33(4), 1307–1316.

Martel, J.-L., Mailhot, A., Brissette, F. & Caya, D. (2018). Role of natural climate variability in the detection of anthropogenic climate change signal for mean and extreme precipitation at local and regional scales. *Journal of Climate*, 31(11), 4241–4263.

Martens, B., Cabus, P., De Jongh, I. & Verhoest, N. (2013). Merging weather radar observations with ground-based measurements of rainfall using an adaptive multiquadric surface fitting algorithm. *Journal of hydrology*, 500, 84–96.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P. & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of climate*, 15(22), 3237–3251.

Mawere, M. (2017). *Theorising development in Africa: Towards building an African framework of development*. Langaa RPCIG.

Mazzoglio, P., Laio, F., Balbo, S., Boccardo, P. & Disabato, F. (2019). Improving an extreme rainfall detection system with GPM IMERG data. *Remote Sensing*, 11(6), 677.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E., Franz, T. E. et al. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*, 21(7), 3879.

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

McKight, P. E. & Najab, J. (2010). Kruskal-wallis test. *The corsini encyclopedia of psychology*, 1–1.

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J. & Lawrimore, J. H. (2018). The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, 31(24), 9835–9854.

Merz, R. & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of hydrology*, 287(1-4), 95–123.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature communications*, 7(1), 1–11.

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J. & Cannon, A. J. (2019). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences*, 23(3), 1339–1354.

Miao, C., Ashouri, H., Hsu, K.-L., Sorooshian, S. & Duan, Q. (2015). Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China. *Journal of Hydrometeorology*, 16(3), 1387–1396.

Milliman, J. D., Farnsworth, K., Jones, P., Xu, K. & Smith, L. (2008). Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000. *Global and planetary change*, 62(3-4), 187–194.

Min, S.-K., Zhang, X., Zwiers, F. W. & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, 470(7334), 378–381.

Minville, M., Brissette, F. & Leconte, R. (2008). Uncertainty of the impact of climate change on the hydrology of a nordic watershed. *Journal of hydrology*, 358(1-2), 70–83.

Moradkhani, H. & Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In *Hydrological modelling and the water cycle* (pp. 1–24). Springer.

Mpelasoka, F. S. & Chiew, F. H. (2009). Influence of rainfall scenario construction methods on runoff projections. *Journal of Hydrometeorology*, 10(5), 1168–1183.

Munier, S. & Aires, F. (2018). A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water budget. *Remote Sensing of Environment*, 205, 119–130.

Nakicenovic, N., Alcamo, J., Davis, G., De Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. & Kram, T. (2000). A special report of Working Group III of the Intergovernmental Panel on Climate Change. *Emissions Scenarios*, 570.

Nash, J. E. & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3), 282–290.

Nashwan, M. S. & Shahid, S. (2019). Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmospheric Research*, 230, 104632.

Nashwan, M. S., Shahid, S. & Wang, X. (2019). Uncertainty in estimated trends using gridded rainfall data: A case study of Bangladesh. *Water*, 11(2), 349.

New, M., Todd, M., Hulme, M. & Jones, P. (2001). Precipitation measurements and trends in the twentieth century. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 21(15), 1889–1922.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L. & Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, 16(6), 2481–2500.

Ngaira, J. K. W. (2007). Impact of climate change on agriculture in Africa by 2030. *Scientific Research and Essays*, 2(7), 238–243.

Nicholson, S. E. (2013). The West African Sahel: A review of recent studies on the rainfall regime and its interannual variability. *ISRN Meteorology*, 2013.

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E. & Stisen, S. (2011). Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. *Remote Sensing of Environment*, 115(1), 107–116.

Nishat, B. & Rahman, S. M. (2009). Water Resources Modeling of the Ganges-Brahmaputra-Meghna River Basins Using Satellite Remote Sensing Data 1. *JAWRA Journal of the American Water Resources Association*, 45(6), 1313–1327.

Nóbrega, M., Collischonn, W., Tucci, C. & Paz, A. (2011). Uncertainty in climate change impacts on water resources in the Rio Grande Basin, Brazil. *Hydrology and Earth System Sciences*, 15(2), 585.

Odon, P., West, G. & Stull, R. (2019). Evaluation of Reanalyses over British Columbia. Part II: Daily and Extreme Precipitation. *Journal of Applied Meteorology and Climatology*, 58(2), 291–315.

Okamoto, K., Ushio, T., Iguchi, T., Takahashi, N. & Iwanami, K. (2005). The global satellite mapping of precipitation (GSMaP) project. *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.*, 5, 3414–3416.

Olauson, J. (2018). ERA5: The new champion of wind power modelling? *Renewable energy*, 126, 322–331.

Omondi, P. A., Awange, J. L., Forootan, E., Ogallo, L. A., Barakiza, R., Girmaw, G. B., Fesseha, I., Kululetera, V., Kilembe, C., Mbati, M. M. et al. (2014). Changes in temperature and precipitation extremes over the Greater Horn of Africa region from 1961 to 2010. *International Journal of Climatology*, 34(4), 1262–1277.

Onyutha, C., Tabari, H., Rutkowska, A., Nyeko-Ogiramoi, P. & Willems, P. (2016). Comparison of different statistical downscaling methods for climate change rainfall projections over the Lake Victoria basin considering CMIP3 and CMIP5. *Journal of Hydro-environment Research*, 12, 31–45.

Orellana, B., Pechlivanidis, I., McIntyre, N., Wheater, H. & Wagener, T. (2008). A toolbox for the identification of parsimonious semi-distributed rainfall-runoff models: Application to the Upper Lee catchment.

Orlowsky, B. & Seneviratne, S. I. (2012). Global changes in extreme events: regional and seasonal dimension. *Climatic Change*, 110(3-4), 669–696.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, 303(1-4), 290–306.

Oudin, L., Andréassian, V., Perrin, C., Michel, C. & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44(3).

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P. et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc.

Parker, W. S. (2016). Reanalyses and observations: What's the difference? *Bulletin of the American Meteorological Society*, 97(9), 1565–1572.

Paul, J. D., Roberts, G. G. & White, N. (2014). The African landscape through space and time. *Tectonics*, 33(6), 898–935.

Pechlivanidis, I. & Arheimer, B. (2015). Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case. *Hydrology and Earth System Sciences*, 19(11), 4559–4579.

Perrin, C., Michel, C. & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, 279(1-4), 275–289.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D. et al. (1998). Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 18(13), 1493–1517.

Pierce, D. W., Das, T., Cayan, D. R., Maurer, E. P., Miller, N. L., Bao, Y., Kanamitsu, M., Yoshimura, K., Snyder, M. A., Sloan, L. C. et al. (2013). Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dynamics*, 40(3-4), 839–856.

Pinto, J. G., Neuhaus, C. P., Leckebusch, G. C., Reyers, M. & Kerschgens, M. (2010). Estimation of wind storm impacts over Western Germany under future climate conditions using a statistical—dynamical downscaling approach. *Tellus A: Dynamic Meteorology and Oceanography*, 62(2), 188–201.

Poméon, T., Jackisch, D. & Diekkrüger, B. (2017). Evaluating the performance of remotely sensed and reanalysed precipitation data over West Africa using HBV light. *Journal of hydrology*, 547, 222–235.

Poulin, A., Brissette, F., Leconte, R., Arsenault, R. & Malo, J.-S. (2011). Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *Journal of Hydrology*, 409(3-4), 626–636.

Prakash, S., Gairola, R. & Mitra, A. (2015a). Comparison of large-scale global land precipitation from multisatellite and reanalysis products with gauge-based GPCC data sets. *Theoretical and Applied Climatology*, 121(1-2), 303–317.

Prakash, S., Mitra, A. K., Momin, I. M., Rajagopal, E., Basu, S., Collins, M., Turner, A. G., Achuta Rao, K. & Ashok, K. (2015b). Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season. *International Journal of Climatology*, 35(9), 2326–2338.

Prakash, S., Kumar, M. R., Mathew, S. & Venkatesan, R. (2018). How accurate are satellite estimates of precipitation over the north Indian Ocean? *Theoretical and Applied Climatology*, 134(1-2), 467–475.

Prăvălie, R. (2016). Drylands extent and environmental issues. A global approach. *Earth-Science Reviews*, 161, 259–278.

Prein, A. F. & Gobiet, A. (2017). Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, 37(1), 305–327.

Prudhomme, C. & Davies, H. (2009). Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate. *Climatic Change*, 93(1-2), 197–222.

Prudhomme, C., Jakob, D. & Svensson, C. (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. *Journal of hydrology*, 277(1-2), 1–23.

Raimonet, M., Oudin, L., Thieu, V., Silvestre, M., Vautard, R., Rabouille, C. & Le Moigne, P. (2017). Evaluation of gridded meteorological datasets for hydrological modeling. *Journal of Hydrometeorology*, 18(11), 3027–3041.

Raimonet, M., Thieu, V., Silvestre, M., Oudin, L., Rabouille, C., Vautard, R. & Garnier, J. (2018). Landward perspective of coastal eutrophication potential under future climate change: The Seine River case (France). *Frontiers in Marine Science*, 5, 136.

Raymond, C., Singh, D. & Horton, R. (2017). Spatiotemporal patterns and synoptics of extreme wet-bulb temperature in the contiguous United States. *Journal of Geophysical Research: Atmospheres*, 122(24), 13–108.

Razavi, T. & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of hydrologic engineering*, 18(8), 958–975.

Riboust, P., Thirel, G., Le Moine, N. & Ribstein, P. (2019). Revisiting a simple degree-day model for integrating satellite data: implementation of SWE-SCA hystereses. *Journal of hydrology and hydromechanics*, 67(1), 70–81.

Rienecker, M. M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R. et al. (2008). The GEOS-5 Data Assimilation System: Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0.

Roads, J. (2003). The NCEP–NCAR, NCEP–DOE, and TRMM tropical atmosphere hydrologic cycles. *Journal of Hydrometeorology*, 4(5), 826–840.

Rojas, O., Vrieling, A. & Rembold, F. (2011). Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. *Remote sensing of Environment*, 115(2), 343–352.

Romilly, T., G., G. & M. (2011). Evaluation of satellite rainfall estimates over Ethiopian river basins. *Hydrology and Earth System Sciences, 15(5), 1505*.

Rowell, D. P. (2006). A demonstration of the uncertainty in projections of UK climate change resulting from regional model formulation. *Climatic Change*, 79(3-4), 243–257.

Rudolf, B., Becker, A., Schneider, U., Meyer-Christoffer, A. & Ziese, M. (2010). The new "GPCC Full Data Reanalysis Version 5" providing high-quality gridded monthly precipitation data for the global land-surface is public available since December 2010. *GPCC Status rep*.

Ruffault, J., Moron, V., Trigo, R. & Curt, T. (2017). Daily synoptic conditions associated with large fire occurrence in Mediterranean France: evidence for a wind-driven fire regime. *International Journal of Climatology*, 37(1), 524–533.

Saadi, M., Oudin, L. & Ribstein, P. (2019). Random Forest Ability in Regionalizing Hourly Hydrological Model Parameters. *Water*, 11(8), 1540.

Sabarly, F., Essou, G., Lucas-Picher, P., Poulin, A. & Brissette, F. (2016). Use of four reanalysis datasets to assess the terrestrial branch of the water cycle over Quebec, Canada. *Journal of Hydrometeorology*, 17(5), 1447–1466.

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H., Pan, H.-L., Moorthi, S., Behringer, D. et al. (2006). The NCEP climate forecast system. *Journal of Climate*, 19(15), 3483–3517.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D. et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8), 1015–1058.

Samaniego, L., Bárdossy, A. & Kumar, R. (2010). Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. *Water Resources Research*, 46(2).

San-Martín, D., Manzanas, R., Brands, S., Herrera, S. & Gutiérrez, J. M. (2017). Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. *Journal of Climate*, 30(1), 203–223.

Satgé, F., Defrance, D., Sultan, B., Bonnet, M.-P., Seyler, F., Rouché, N., Pierron, F. & Paturel, J.-E. (2020). Evaluation of 23 gridded precipitation datasets across West Africa. *Journal of Hydrology*, 581, 124412.

Sayama, T., Tachikawa, Y. & Takara, K. (2012). Spatial lumping of a distributed rainfall-sediment-runoff model and its effective lumping scale. *Hydrological Processes*, 26(6), 855–871.

Sayre, A. P. (1999). *Africa*. Twenty-First Century Books.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A. & Ziese, M. (2018). GPCC Full Data Monthly Product Version 2018 at 0.5: Monthly land-surface precipitation from rain-gauges built on GTS-based and historical data. *Global Precipitation Climatology Centre*.

Seyyedi, H., Anagnostou, E. N., Beighley, E. & McCollum, J. (2015). Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin. *Atmospheric Research*, 164, 37–48.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F. & McCabe, M. F. (2009). Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, 36(7).

Shen, Y., Hong, Z., Pan, Y., Yu, J. & Maguire, L. (2018). China's 1 km merged gauge, radar and satellite experimental precipitation dataset. *Remote Sensing*, 10(2), 264.

Shongwe, M. E., Van Oldenborgh, G., Van Den Hurk, B., De Boer, B., Coelho, C. & Van Aalst, M. (2009). Projected changes in mean and extreme precipitation in Africa under global warming. Part I: Southern Africa. *Journal of climate*, 22(13), 3819–3837.

Siddique-E-Akbor, A., Hossain, F., Sikder, S., Shum, C., Tseng, S., Yi, Y., Turk, F. & Limaye, A. (2014). Satellite precipitation data–driven hydrological modeling for water resources management in the Ganges, Brahmaputra, and Meghna Basins. *Earth Interactions*, 18(17), 1–25.

Siegert, M. J., Ross, N. & Le Brocq, A. M. (2016). Recent advances in understanding Antarctic subglacial lakes and hydrology. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2059), 20140306.

Singh, A. & Patwardhan, A. (2012). Spatio-temporal distribution of extreme weather events in India. *APCBEE Procedia*, 1, 258–262.

Singh, D., Swain, D. L., Mankin, J. S., Horton, D. E., Thomas, L. N., Rajaratnam, B. & Diffenbaugh, N. S. (2016). Recent amplification of the North American winter temperature dipole. *Journal of Geophysical Research: Atmospheres*, 121(17), 9911–9928.

Singh, V. P. & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. *Journal of hydrologic engineering*, 7(4), 270–292.

Skinner, C. J., Bellerby, T. J., Greatrex, H. & Grimes, D. I. (2015). Hydrological modelling using ensemble satellite rainfall estimates in a sparsely gauged river basin: The need for whole-ensemble calibration. *Journal of hydrology*, 522, 110–122.

Smith, R. A. & Kummerow, C. D. (2013). A comparison of in situ, reanalysis, and satellite water budgets over the upper Colorado River basin. *Journal of Hydrometeorology*, 14(3), 888–905.

Song, C., Huang, B., Ke, L. & Ye, Q. (2016). Precipitation variability in High Mountain Asia from multiple datasets and implication for water balance analysis in large lake basins. *Global and Planetary Change*, 145, 20–29.

Stearns, C. R. & Wendler, G. (1988). Research results from Antarctic automatic weather stations. *Reviews of Geophysics*, 26(1), 45–61.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. M. et al. (2013). Climate change 2013: The physical science basis. *Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*, 1535.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. & Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1), 79–107.

Sylla, M., Giorgi, F., Coppola, E. & Mariotti, L. (2013). Uncertainties in daily rainfall over Africa: assessment of gridded observation products and evaluation of a regional climate model simulation. *International Journal of Climatology*, 33(7), 1805–1817.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J. et al. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104, 70–97.

Tarek, M., Brissette, F. P. & Arsenault, R. (2019). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modeling over North-America. *Hydrology and Earth System Sciences Discussions*, 1–35.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498.

Teutschbein, C. & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology*, 456, 12–29.

Tian, Y., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., Hsu, K.-l., Turk, F. J., Garcia, M. & Zeng, J. (2009). Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research: Atmospheres*, 114(D24).

Toté, C., Patricio, D., Boogaard, H., Van der Wijngaart, R., Tarnavsky, E. & Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sensing*, 7(2), 1758–1776.

Trambauer, P., Maskey, S., Winsemius, H., Werner, M. & Uhlenbrook, S. (2013). A review of continental scale hydrological models and their suitability for drought forecasting in (sub-Saharan) Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, 66, 16–26.

Troin, M., Arsenault, R., Martel, J.-L. & Brissette, F. (2018). Uncertainty of hydrological model components in climate change studies over two Nordic Quebec catchments. *Journal of Hydrometeorology*, 19(1), 27–46.

Trudel, M., Doucet-Généreux, P.-L. & Leconte, R. (2017). Assessing river low-flow uncertainties related to hydrological model calibration and structure under climate change conditions. *Climate*, 5(1), 19.

Tschakert, P., Sagoe, R., Ofori-Darko, G. & Codjoe, S. N. (2010). Floods in the Sahel: an analysis of anomalies, memory, and anticipatory learning. *Climatic Change*, 103(3-4), 471–502.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G. et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(612), 2961–3012.

Urraca, R., Huld, T., Gracia-Amillo, A., Martinez-de Pison, F. J., Kaspar, F. & Sanz-Garcia, A. (2018). Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. *Solar Energy*, 164, 339–354.

Valéry, A., Andréassian, V. & Perrin, C. (2014). 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2–Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of hydrology*, 517, 1176–1187.

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F. & Willems, P. (2014). Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation. *Journal of Hydrology*, 511, 335–349.

Vehviläinen, B. (1992). *Snow cover models in operational watershed forecasting*. National Board of Waters and the Environment Helsinki, Finland.

Velázquez, J. A., Troin, M., Caya, D. & Brissette, F. (2015). Evaluating the time-invariance hypothesis of climate model bias correction: implications for hydrological impact studies. *Journal of Hydrometeorology*, 16(5), 2013–2026.

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I. G., Plötner, S., Seidou, O. et al. (2017). Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Climatic Change*, 141(3), 419–433.

Vila, D. A., De Goncalves, L. G. G., Toll, D. L. & Rozante, J. R. (2009). Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. *Journal of Hydrometeorology*, 10(2), 533–543.

Vogel, R. & Vogel, R. (2013). *Quantifying the uncertainty of spatial precipitation analyses with radar-gauge observation ensembles*. Bundesamt für Meteorologie und Klimatologie, MeteoSchweiz.

Voisin, N., Wood, A. W. & Lettenmaier, D. P. (2008). Evaluation of precipitation products for global hydrological prediction. *Journal of Hydrometeorology*, 9(3), 388–407.

Von Storch, H. & Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge university press.

Vose, R. S., Schmoyer, R. L., Steurer, P. M., Peterson, T. C., Heim, R., Karl, T. R. & Eischeid, J. K. (1992). *The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data*.

Wang, H.-M., Chen, J., Xu, C.-Y., Chen, H., Guo, S., Xie, P. & Li, X. (2019). Does the weighting of climate simulations result in a better quantification of hydrological impacts? *Hydrology and Earth System Sciences*, 23(10), 4033–4050.

Wang, X. L., Swail, V. R. & Zwiers, F. W. (2006). Climatology and changes of extratropical cyclone activity: Comparison of ERA-40 with NCEP–NCAR reanalysis for 1958–2001. *Journal of Climate*, 19(13), 3145–3166.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50(9), 7505–7514.

Westra, S., Thyer, M., Leonard, M., Kavetski, D. & Lambert, M. (2014). A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resources Research*, 50(6), 5090–5113.

Wheater, H., Sorooshian, S. & Sharma, K. D. (2007). *Hydrological modelling in arid and semi-arid areas*. Cambridge University Press.

Wheater, H. (2002). Progress in and prospects for fluvial flood modelling. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360(1796), 1409–1431.

Whitfield, P. H., Burn, D. H., Hannaford, J., Higgins, H., Hodgkins, G. A., Marsh, T. & Looser, U. (2012). Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal*, 57(8), 1562–1579.

Wilby, R. L. (2005). Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological Processes: An International Journal*, 19(16), 3201–3219.

Wilby, R. L. & Harris, I. (2006). A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water resources research*, 42(2).

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P. & Mearns, L. O. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA*, 27.

Wood, E. F. (1998). 'Hydrologic measurements and observations: An assessment of needs. *Proc., 1997 Abel Wolman Distinglished Lecture and Symposium on Hydrologic Sciences*, pp. 67–86.

Wu, Z., Xu, Z., Wang, F., He, H., Zhou, J., Wu, X. & Liu, Z. (2018). Hydrologic evaluation of multi-source satellite precipitation products for the upper Huaihe River Basin, China. *Remote Sensing*, 10(6), 840.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y. & Liu, C. (2007). A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology*, 8(3), 607–626.

Yang, R., Ek, M. & Meng, J. (2015). Surface water and energy budgets for the Mississippi river basin in three NCEP Reanalyses. *Journal of Hydrometeorology*, 16(2), 857–873.

Yong, B., Liu, D., Gourley, J. J., Tian, Y., Huffman, G. J., Ren, L. & Hong, Y. (2015). Global view of real-time TRMM multisatellite precipitation analysis: Implications for its successor global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 96(2), 283–296.

Youssef, H., Simon, G., Younes, F., Ghani, C. & Vincent, S. (2018). Rainfall-Runoff modeling in a semi-arid catchment with presence of snow. The Rheraya wadi case study (Marrakech, Morocco). *EGUGA*, 5214.

Yu, X., Bhatt, G., Duffy, C. & Shi, Y. (2013). Parameterization for distributed watershed modeling using national data and evolutionary algorithm. *Computers & Geosciences*, 58, 80–90.

Zhang, Y., Zheng, H., Chiew, F. H., Arancibia, J. P. & Zhou, X. (2016). Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. *Journal of Hydrometeorology*, 17(3), 995–1010.

Zhao, F., Veldkamp, T. I., Frieler, K., Schewe, J., Ostberg, S., Willner, S., Schauberger, B., Gosling, S. N., Schmied, H. M., Portmann, F. T. et al. (2017). The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, 12(7), 075003.

Zhu, H., Li, Y., Huang, Y., Li, Y., Hou, C. & Shi, X. (2018). Evaluation and hydrological application of satellite-based precipitation datasets in driving hydrological models over the Huifa river basin in Northeast China. *Atmospheric Research*, 207, 28–41.