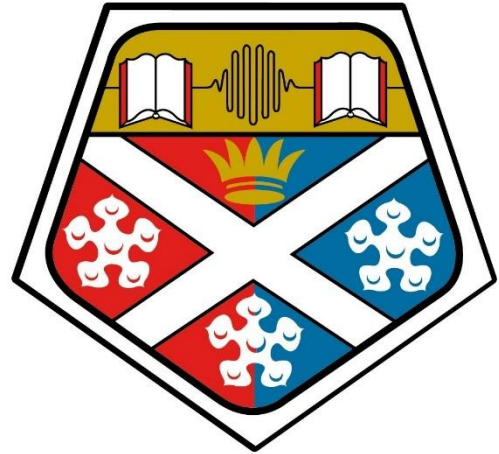
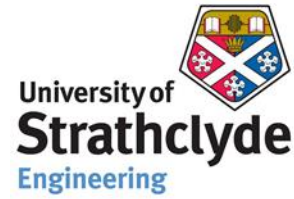


**Integrated modelling of water security in  
data-sparse regions under uncertainty.**



**University of  
Strathclyde  
Glasgow**

**November 2023**



# **Integrated modelling of water security in data-sparse regions under uncertainty**

**Ibrahim Mohammed Lawal**

A thesis presented for the degree of Doctor of Philosophy

School of Engineering,  
Department of Civil and Environmental Engineering,  
University of Strathclyde,  
James Weir Building, 75 Montrose Street, Glasgow, G1 1XJ.

Funded by the Petroleum Technology Development Fund (PTDF), Nigeria

## **DEDICATION**

This thesis is dedicated to Allah (SWT) who granted me the wisdom, knowledge and discipline to pursue my life goals, and in loving memory of my Late Dad (Mohammed Lawal Ibrahim) who is my inspiration for the lessons thought, and guidance provided, values instilled in me necessary to navigate through life challenges and may you have eternal bliss.

## DECLARATION OF AUTHORSHIP

This thesis is the result of the author's original research work. It was composed by the author and has not been previously submitted for oral examination leading to the award of a degree.

All consulted literature from published works of others have been clearly and appropriately cited.

All sources of help rendered during the research were clearly acknowledged.

Where the research is based on work done jointly with others, I have made clear exactly the contribution by others and what I have contributed myself.

A part of this work has been published in reputable journals before submission.

The copyright of this research work belongs to the author under the terms of the United Kingdom Copyright Act as qualified by the University of Strathclyde Regulation 3.50 contained in the PGR guidelines. An acknowledgement must always be made to use any part or whole material contained in or derived from the thesis.



---

Signature

24<sup>th</sup> November 2023

---

Date

## **ACKNOWLEDGEMENTS**

I would foremostly thank Allah (SWT) for the gift of life, to my family, friends, and colleagues for their prayers and moral support throughout the PhD research journey.

I would like to express my profound gratitude and appreciation to my Ph.D. advisor Dr. Douglas Bertram, for all the academic guidance, moral, and psychological support to navigate the hurdles and hassles of Doctoral Degree studies. I appreciate your thoughtful guidance and shared precious life-changing experiences, especially around work-life balance, which have made me a better husband, father and human being. I will eternally remain grateful and indebted to him.

I would also like to thank my second PhD advisor, Dr Christopher John White for all the support and guidance through thought-provoking positive criticism and insight to improve this research work. I will remain eternally indebted and appreciative for making me improve and be an effective researcher.

I would like to extend my gratitude to all academic and non-academic staff members of the Department of Civil and Environmental Engineering, and my PGR colleagues for their support, cheerfulness and hospitality. You made my journey and stay at the University of Strathclyde a memorable and wholesome experience.

This PhD research would not have been possible without the financial support from the Petroleum Technology Development Fund (PTDF), Nigeria and the University of Strathclyde. My gratitude also goes to Abubakar Tafawa Balewa University for the support to pursue my studies.

Last, but not least, I would like to thank my lovely wife Halima and my kids Hauwa, Aisha and Halima at whose expense I had to spend long hours away from their attention to pursue my lifelong goal and complete my thesis. I remain highly indebted for your support and understanding. To my Friends, Colleagues and Well-wishers thank you for the support. Finally, thank you to my mum, brother and sisters who are my bedrock and source of spiritual support throughout my life, and I will remain forever indebted. May Allah (SWT) bless you all immensely.

## ARTICLES PUBLISHED FROM THE RESEARCH

1. Lawal, I.M., Bertram, D., White, C.J., Jagaba, A.H., Hassan, I., Shuaibu, A., 2021. Multi-criteria performance evaluation of gridded precipitation and temperature products in data-sparse regions. *Atmosphere* 12, 1597. <https://doi.org/10.3390/atmos12121597>
2. Lawal, I.M., Bertram, D., White, C.J., Kutty, S.R.M., Hassan, I., Jagaba, A.H., 2023. Application of Boruta algorithms as a robust methodology for performance evaluation of CMIP6 general circulation models for hydro-climatic studies. *Theoretical and Applied Climatology* 152, 04466–5. <https://doi.org/10.1007/s00704-023-04466-5>
3. Lawal, I.M., Bertram, D., White, C.J., Jagaba, A.H., 2023. Integrated framework for hydrologic modelling in data-sparse watersheds and climate change impact on projected green and blue water sustainability. *Frontiers in Environmental Science* 11, 1 – 26. <https://doi.org/10.3389/fenvs.2023.1233216>

**Disclaimer:** The published articles were entirely my effort to objectively provide a solution to the questions from the identified research gap with guidance and additional insights from my supervisors Dr. Douglas Bertram and Dr. Christopher John White in conceptualization and formulation of the theoretical research background, numerical simulation of models, analysis of research results, and writing of the draft manuscript. However, assistance was also rendered through intellectual engagement for additional critical review and editing of the manuscript from a senior colleague through collaboration (Prof. S. R. M. Kutty) and my PhD peers within (A. Shuaibu, I. Hassan) and outside (A. H. Jagaba) the university.

## ABSTRACT

Freshwater scarcity and sustainability is one of the most complicated and difficult issues the world is currently facing, and it has been identified as a global concern. According to expert studies, 80% of the world's population is projected to live in freshwater threats due to a plethora of factors viz., rapid population growth, urbanization, global climate change resulting from spatial and temporal changes in magnitude, frequencies and intensity of precipitation and temperature which leads to the transformation of the hydrologic cycle.

Recent initiatives, including sustainable development goals, have been made to address these problems and offer solutions. However, the quantity and quality of freshwater systems and resources must be objectively and comprehensively understood and assessed at the scale of river basins to provide sufficient mitigation and resilience planning.

Hydrologic modelling has been one the most suitable and efficient strategies for basin-scale assessment of freshwater dynamics to current and projected climate change and the focus has been on the application of traditional modelling framework which is tenable where data requirements are sufficient to couple hydrologic models with atmospheric data to account for climate change.

The aforementioned strategy is a challenge in regions with inadequate ground-based observations necessary for climate and hydrologic modelling. The rarely available data in such regions may have repetitive gaps of missing data points with negative consequences including biased statistical representation of basin climatic features, ineffective model calibration and unreliable timing of peak flows which may amplify the uncertainties of the hydrologic dynamics leading to flawed depictions of watershed responses.

Recently, integrated strategies are evolving that couple hydrologic models with climate data in water resource studies to account for uncertainties through the use of alternative data

sources of many spatial climate data products from climate research centres to overcome the identified challenges.

This research developed and applied a multi-criteria approach to examine the efficacy of gridded climate products using different performance metrics, a machine learning-based approach, Boruta random forest (BRF) to assess multiple GCM datasets required for hydro-climatic studies and an integrated BRF-SWAT technique to define the relationship between the hydrologic variables and improve rainfall-runoff modelling in a data-sparse and climate-sensitive watersheds.

The developed model was applied to assess the projected green and blue water dynamics and sustainability in the Yobe-Komadugu basin of the greater Lake Chad, a watershed that is prone to extreme events (SPEI of flood and drought hazards). The results demonstrate that though the performance of the gridded data varies in space and time, multi-criteria assessment enhances the choice of a product with reduced uncertainty for climate modelling.

The incorporation of the BRF approach in GCM evaluation indicates a consistent spatial and temporal representation of the climatological features with suitable mean correlation ( $R^2 = 0.95$ ), reduced mean annual precipitation bias of 0.69 mm/year and enhanced statistical trend and magnitude of the SPEI drought and flood hazards relative to identified and tested approaches from the literature.

The integrated framework of the rainfall-runoff modelling strategy indicated that the hydrologic fluxes can be simulated fairly accurately with varying degrees of acceptability, irrespective of the watershed morphological properties, although there are significant trade-offs in model parameter sensitivity.

The availability of satellite-based measurements of hydrologic fluxes and states, coupled with a machine learning feature selection and data refinement process has made integrated



water balance modelling widely seen as a viable alternative for improving watershed hydrologic processes in data-sparse regions within acceptable uncertainty limits.

Furthermore, the sub-watershed assessment of the projected changes in spatial and temporal green and blue water sustainability status has shown that the sub-basins will be ecologically fragile, and the identified freshwater geographic hotspots may be beyond restoration without adequate long-term river basin water resources plans. The modelling framework developed is, however, independent of the model and data type and can be applied to watersheds with similar modelling challenges.

This study has provided a pathway or methods for managing and securing water resources information as a decision support tool to guarantee ongoing watershed monitoring and assessment of water security even in the face of increasingly unpredictable future circumstances in data-sparse watersheds that take into account uncertainty and that a course for prospective risk assessment or the possibility and understanding that a certain effect brought on by climate-induced hazards would prevail in watershed freshwater sustainability.

Therefore, It is essential to comprehend the constraints associated with forecasting changes in the water cycle to improve the climate and hydrologic modelling process, which is required to create effective strategies for adapting to climate change-related water resource hazards. Even in the face of severe uncertainty about the future, this will be essential in addressing concerns related to water security and management and promoting the climatic resilience of ecosystems and society.

## TABLE OF CONTENTS

TITLE PAGE .....	ii
DEDICATION .....	iii
DECLARATION OF AUTHORSHIP .....	iv
ACKNOWLEDGEMENTS .....	v
ARTICLES PUBLISHED FROM THE RESEARCH .....	vi
ABSTRACT .....	vii
TABLE OF CONTENTS .....	x
LIST OF FIGURES .....	xv
LIST OF TABLES .....	xviii
ABBREVIATIONS AND NOTATIONS .....	xx
CHAPTER 1: INTRODUCTION .....	1
1.1 Research background .....	1
1.1.1 Water footprint concept as a tool for water security governance. ....	3
1.1.2 Watershed modelling in ungauged basins .....	4
1.1.3 Uncertainty management in watershed modelling .....	6
1.2 Overview of research questions .....	7
1.3 Aim of the research .....	8
1.4 Objectives of the research .....	8
1.5 Scope of the research.....	9
1.6 Rationale for study area selection and research motivation .....	10
1.7 Thesis Structure.....	12
CHAPTER 2: LITERATURE REVIEW .....	14
2.1 Climate change and the environment – Key Insights.....	14
2.1.1 Climate Change and the hydrologic cycle .....	14
2.1.2 Climate Models and climate modelling .....	17

2.1.3	Uncertainty in climate modelling .....	22
2.1.4	Climate change scenarios .....	25
2.2	Watershed hydrology and the water balance dynamics .....	30
2.2.1	Hydrologic dynamics and climate change.....	30
2.2.2	Hydrologic modelling and process optimization.....	32
2.2.3	Uncertainty in hydrologic modelling.....	36
2.2.4	Challenges of hydrologic modelling.....	39
2.3	Water resources assessment .....	41
2.3.1	Water footprint concept .....	41
2.3.2	Watershed water resources sustainability .....	44
2.4	Summary and conclusion .....	46
<b>CHAPTER 3: RESEARCH DESIGN FRAMEWORK.....</b>		<b>49</b>
3.1	Introduction .....	49
3.2	Reliability analysis of gridded and observation data .....	51
3.3	Statistical downscaling and bias correction of climate models.....	53
3.3.1	Delta change method .....	53
3.3.2	Quantile Mapping method .....	54
3.3.3	Empirical quantile mapping method.....	55
3.4	Evaluation of the effectiveness of bias correction method .....	56
3.5	Multi-model assessment for hydrologic application .....	56
3.6	Watershed modelling and water security assessment under uncertainty .....	58
<b>CHAPTER 4: MULTI-CRITERIA PERFORMANCE EVALUATION OF GRIDDED PRECIPITATION AND TEMPERATURE PRODUCTS IN DATA SPARSE REGION. ....</b>		<b>61</b>
4.1	Abstract of paper .....	62
4.2	Introduction .....	63
4.3	Study Area and data .....	68
4.3.1	Study area .....	68

4.3.2 Observation data and sources .....	71
4.3.3 Gridded data and sources.....	73
4.4 Research Methodology.....	75
4.4.1 Symmetric Uncertainty .....	75
4.4.2 Statistical Metrics .....	77
4.4.3 Trend Analysis.....	78
4.5 Results .....	80
4.5.1 Assessment of gridded data using symmetric uncertainty.....	80
4.5.2 Statistical metric efficiency .....	81
4.5.3 Trend Analysis of gridded data .....	87
4.6 Discussion .....	93
4.7 Conclusion.....	96
4.8 Afterward .....	98
<b>CHAPTER 5: APPLICATIONS OF BORUTA ALGORITHMS AS A ROBUST METHODOLOGY FOR PERFORMANCE EVALUATION OF CMIP6 GENERAL CIRCULATION MODELS FOR HYDRO-CLIMATIC STUDIES.....</b>	<b>99</b>
5.1 Abstract of paper .....	101
5.2 Introduction .....	102
5.3 Case Study Area and Data.....	110
5.3.1 Case study area .....	110
5.3.2 Gridded data and sources.....	112
5.3.3 General circulation model data and sources .....	113
5.4 Research Methodology.....	115
5.4.1 Methodology flow chart .....	115
5.4.2 Statistical downscaling and bias correction of GCM .....	115
5.4.3 Performance evaluation and selection of appropriate GCM output by Machine learning technique.....	116
5.4.4 Performance evaluation of selected GCM ensemble mean. ....	119

5.4.5 Validation of evaluation approach based on SPEI drought and flood hazard assessment. ....	120
5.4.6 Standardized precipitation evapotranspiration index .....	120
5.5 Results and Discussion.....	122
5.5.1 Spatial and temporal downscaling and bias correction of CMIP6 GCMs.....	122
5.5.2 Performance evaluation and selection of CMIP6 GCMs .....	124
5.5.3 Identification and evaluation of multi-model ensemble mean of GCMs. ....	131
5.6 Conclusion.....	141
5.7 Afterward .....	142
<b>CHAPTER 6: INTEGRATED FRAMEWORK FOR HYDROLOGIC MODELLING IN DATA-SPARSE WATERSHEDS AND CLIMATE CHANGE IMPACT ON PROJECTED GREEN AND BLUE WATER SUSTAINABILITY. ....</b>	<b>144</b>
6.1 Abstract of paper .....	145
6.2 Introduction .....	147
6.3 Case Study area and data.....	153
6.3.1 Case study Area .....	153
6.3.2 Dataset Description.....	155
6.4 Research Methodology.....	157
6.4.1 Pre-processing of input data .....	157
6.4.2 Integrated modelling framework .....	160
6.4.3 Integrated model set-up, calibration, validation and uncertainty analysis. ....	164
6.4.4 Assessment of water footprint environmental sustainability.....	167
6.5 Results and discussions .....	170
6.5.1 Calibration and Validation of the integrated model .....	170
6.5.2 Assessment of climate change impact on projected green and blue water resources. ....	177
6.5.3 Climate change impact and socioeconomic drivers on spatial variation of projected green and blue water sustainability.....	188

6.5.4 Climate change impact and socioeconomic drivers on the temporal variability of projected changes of green and blue water sustainability.....	192
6.6 Conclusions and future work.....	196
6.7 Afterward .....	199
CHAPTER 7: SUMMARY OF FINDINGS AND DISCUSSIONS .....	201
7.1 Assessment of quality-controlled observation data for consistency .....	201
7.2 Reliability of gridded climate data using multi-criteria approach.....	202
7.3 Data pruning as an effective approach for GCM uncertainty reduction in modelling .....	204
7.4 Spatial and temporal assessment of GCM ensembles for hydrologic modelling...	207
7.5 Modelling evapotranspiration using integrated SWAT-BRF framework.....	208
7.6 Impact of climate change on projected green and blue water resources .....	210
7.7 Impact of climate change on projected green and blue water sustainability.....	212
7.8 Overview of integrated framework for water security modelling.....	214
CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH .....	215
8.1 Conclusions .....	215
8.2 Recommendations to current practice .....	217
8.3 Recommendations for future research.....	217
REFERENCES .....	220
APPENDICES .....	250
Appendix 1: Multivariate Imputation and Chained Equation (MICE) missing data script.....	250
Appendix 2: Homogeneity test results of station data.....	250
Appendix 3: Climate data extraction R-script.....	251
Appendix 4: Mann-Kendall trend analysis R-script.....	253
Appendix 5: Random forest algorithms for data pruning R-script.....	254

## LIST OF FIGURES

<b>Figure 2.1:</b> Hydrologic cycle dynamics at watershed scale (Quinteiro et al., 2018a) .....	16
<b>Figure 2.2:</b> General overview of the main climate modelling uncertainty sources (Martin et al., 2020) .....	23
<b>Figure 2.3:</b> Ratio of GCM variance to total variance as a measure of uncertainty. In red areas, GHM uncertainty predominates, Greenland has been masked (Schewe et al., 2014)...	24
<b>Figure 2.4:</b> Schematic diagram of the steps in developing SSPs, with narratives, socioeconomic scenario drivers, SSP baselines and mitigation scenarios (Riahi et al., 2017). .....	27
<b>Figure 2.5:</b> Phases in water footprint assessment (Hoekstra et al., 2011) .....	43
<b>Figure 3.1:</b> Schematic summary of research methodology flowchart (a) consistency and reliability of observations and gridded climate datasets using MCA (b) assessment of efficacy of downscaling and bias corrections and data pruning approaches and validations for hydroclimatic studies (c) integrated framework of modelling using selected data pruning method and SWAT model in data sparse regions for water security assessment. ....	50
<b>Figure 4.1:</b> Map of Lake Chad Basin showing Elevation, Lake, Climate Stations, major river networks and Sub-basins.....	70
<b>Figure 4.2:</b> Double-mass curves for Lake Chad basin. (a): Cumulative annual precipitation at all stations against base station. (b): Cumulative annual temperature at all stations against base station.....	73
<b>Figure 4.3:</b> Boxplot of distribution of the similarity coefficients across the Lake Chad basin (a): variation of similarity coefficient of gridded precipitation against observed station data. (b) variation of similarity coefficient of gridded temperature against observed station data. .	81
<b>Figure 4.4:</b> Boxplot of statistical metrics in the Lake Chad basin (a): KGE and md of gridded precipitation against observed station data. (b): KGE and md of gridded temperature against observed station data.....	85
<b>Figure 4.5:</b> Taylor diagrams for time series data (1979 – 2012). (a): Annual precipitation of gridded and observed station data (b): Annual temperature of gridded against observed station data. (c): Monsoon precipitation of gridded against observed station data. (d): Monsoon temperature of gridded against observed station data. (e): Premonsoon precipitation of gridded against observed station data. (f): Premonsoon temperature of gridded against observed station data. Blue line is Normalized station deviation, Green line is Pearson correlation coefficient and Red line is Normalized root mean square error. ....	87

<b>Figure 4.6:</b> Magnitude of linear trend of Annual gridded and observed precipitation for Chad Basin (1979-2012) .....	92
<b>Figure 4.7:</b> Magnitude of linear trend of Monsoon gridded and observed precipitation for Chad Basin (1979-2012).....	92
<b>Figure 4.8:</b> Magnitude of linear trend of Annual gridded and observed Temperature for Chad Basin (1979-2012) .....	93
<b>Figure 4.9:</b> Magnitude of linear trend of Monsoon gridded and Observed Temperature for Chad Basin (1979 – 2012) .....	93
<b>Figure 5.1:</b> Lake Chad basin showing elevation, climate stations and climatic zones for the proposed study. ....	112
<b>Figure 5.2:</b> GCM evaluation and selection for hydro-climatic study. ....	115
<b>Figure 5.3:</b> Plot of bias-corrected GCMs and observed climate data using delta change, quantile mapping and empirical quantile mapping method in the Lake Chad basin. <b>(a):</b> variation of GCM precipitation relative to observed CPC data. <b>(b):</b> Variation of GCM maximum temperature relative to PGF data. <b>(c)</b> Variation of GCM minimum temperature relative to observed PGF data. ....	123
<b>Figure 5.4:</b> Box plot of variable importance score of GCMs relative to observed climate data using BRF. <b>(a):</b> relative importance of GCM precipitation to observed CPC data. <b>(b):</b> relative importance of GCM maximum temperature to observed PGF data. <b>(c):</b> relative importance of GCM minimum temperature to observed PGF data. ....	129
<b>Figure 5.5:</b> Spatial spread of GCMs performance relative to observed climate data using BRF. <b>(a):</b> ranking of spatial spread GCM precipitation relative to observed CPC data. <b>(b):</b> ranking of spatial spread GCM maximum temperature relative to observed PGF data. <b>(c):</b> ranking of spatial spread GCM minimum temperature relative to observed PGF data. ....	131
<b>Figure 5.6:</b> Comparison of spatial correlation of GCM ensemble mean performance relative to observed climate data <b>(a)</b> Annual precipitation <b>(b)</b> Annual temperature. ....	134
<b>Figure 5.7:</b> Comparison of the spatial pattern of GCM ensemble performance relative to observed climate data <b>(a)</b> Annual precipitation <b>(b)</b> Annual temperature. ....	135
<b>Figure 5.8:</b> Variation of the temporal pattern of mean annual precipitation across grid points for 1979 – 2012.....	136
<b>Figure 5.9:</b> Temporal variations climate extreme event of GCMs ensemble mean performance of relative to observation using BRF, SU and AME approach in the Lake Chad basin. <b>(a):</b> Saharan zone <b>(b):</b> Sahelo-Saharan zone <b>(c):</b> Sahelo-Sudanian zone <b>(d):</b> Sudano-Guinean zone. ....	140



<b>Figure 6.1:</b> Lake Chad Basin showing sub-basins, lake, major river networks and MODIS Evapotranspiration (ET) data points. ....	155
<b>Figure 6.2:</b> Description of morphological data in the study (a) digital elevation model and meteorological points, (b) soil types (c) land use and cover data.....	160
<b>Figure 6.3:</b> Schematic overview of the integrated SWAT and BRF modelling framework (IMF) for reliable water balance modelling in data-sparse regions. ....	161
<b>Figure 6.4:</b> Plot of comparison of the observed and simulated results (95% prediction uncertainty band) of actual evapotranspiration between 1983 – 2006) in the Lake Chad basin. (a): Yobe-Komadugu Watershed (b): Magay-Ngadda Watershed (c): Chari-Logone Watershed (d): Bodou-Dillia .....	175
<b>Figure 6.5:</b> Delineated Yobe Komadugu Watershed with sub-basin locations .....	178
<b>Figure 6.6:</b> Variations in (mm) of spatial distribution of annual green water flow in the Yobe-Komadugu Watershed.....	180
<b>Figure 6.7:</b> Variations (mm) in the temporal distribution of mean monthly green water flow (a) 2021 – 2050 (b) 2051 – 2080 in Yobe-Komadugu Watershed. ....	181
<b>Figure 6.8:</b> Variations in (mm) of spatial distribution of green water storage in the Yobe-Komadugu Watershed.....	183
<b>Figure 6.9:</b> Variations (mm) in the temporal distribution of mean monthly green water storage (a) 2021 – 2050 (b) 2051 – 2080 in Yobe-Komadugu Watershed .....	184
<b>Figure 6.10:</b> Variations (mm) in the spatial distribution of blue water flow in the Yobe-Komadugu watershed.....	187
<b>Figure 6.11:</b> Variations (mm) in the temporal distribution of mean monthly blue water flow (a) 2021 – 2050 (b) 2051 – 2080 in Yobe-Komadugu Watershed .....	188
<b>Figure 6.12:</b> Spatial hazards map of changes of baseline and projected green water environmental sustainability in Yobe-Komadugu Watershed .....	190
<b>Figure 6.13:</b> Spatial hazards map of changes of baseline and projected blue water environmental sustainability in Yobe-Komadugu Watershed .....	191
<b>Figure 6.14:</b> Heat Map showing temporal changes of mean monthly baseline and projected green water environmental sustainability in Yobe-Komadugu Watershed .....	193
<b>Figure 6.15:</b> Heat Map showing temporal changes of mean monthly baseline and projected blue water environmental sustainability in Yobe-Komadugu Watershed. ....	194

## LIST OF TABLES

<b>Table 2.1:</b> Climate model intercomparison project phase six forcing emission thresholds and concentrations. ....	30
<b>Table 4.1:</b> List of reliable Observed temperature stations. location and temporal span in Chad Basin .....	72
<b>Table 4.2:</b> List of Reliable Observed Precipitation station, location and temporal span in Chad Basin .....	72
<b>Table 4.3:</b> Summary of gridded dataset considered in this study. ....	74
<b>Table 4.4:</b> Similarity score of gridded precipitation and temperature against observed datasets estimated by Symmetric Uncertainty .....	81
<b>Table 4.5:</b> Summary of performance of statistical metrics of gridded monthly precipitation data against observed data in Chad Basin .....	83
<b>Table 4.6:</b> Summary of performance of statistical metrics of gridded temperature data against observed data in Chad Basin .....	84
<b>Table 4.7:</b> Mann Kendall Z-Statistics values of linear trend of Annual gridded and observed precipitation for Chad Basin (1979-2012) .....	89
<b>Table 4.8:</b> Mann Kendall Z-Statistics values of linear trend of Monsoon gridded and observed precipitation for Chad Basin (1979-2012).....	90
<b>Table 4.9:</b> Mann Kendall Z-Statistics values of linear trend of Annual gridded and observed temperature for Chad Basin (1979-2012). ....	90
<b>Table 4.10:</b> Mann Kendall Z-Statistics values of linear trend of Monsoon gridded and observed temperature for Chad Basin (1979-2012).....	91
<b>Table 4.11:</b> Summary of the best two (or three when the metrics have the same performance) gridded precipitation and temperature dataset across the performance metrics considered in this study. ....	94
<b>Table 5.1:</b> Summary of CMIP6 models considered in this study. ....	114
<b>Table 5.2:</b> Summary of GCM SU coefficients of precipitation, maximum and minimum temperature relative to observation data .....	126
<b>Table 5.3:</b> Summary of GCM importance score of precipitation, maximum and minimum temperature relative to observation data .....	127
<b>Table 5.4:</b> Summary of GCM overall ranking based on aggregated SU and IS of climate variables relative to observation data.....	132

<b>Table 5.5:</b> Mann Kendall Z-statistic of linear trend of extreme event for the period 1980 - 2012.....	138
<b>Table 6.1:</b> Input data required for hydrologic model development. ....	157
<b>Table 6.2:</b> SWAT model basin delineation.....	164
<b>Table 6.3:</b> Sectoral Water Use Information in Chad and Nigeria.....	169
<b>Table 6.4:</b> Model sensitive parameters, ranges and best-fitted values at sub-watersheds ....	172
<b>Table 6.5:</b> Median of the projected changes in annual precipitation and temperature in the Yobe-Komadugu Watershed.....	177

## ABBREVIATIONS AND NOTATIONS

AET	Actual Evapo-Transpiration
AerChemMIP	Aerosols and Chemistry Model Intercomparison Project
ALPHA_BF.gw	Baseflow alpha factor
ALPHA_BNK.rte	Baseflow alpha factor (Bank Storage)
AME	All Model Ensembles
BRF	Boruta Random Forest
BWF	Blue Water Footprint
°C	Degrees centigrade
CDF	Cumulative Distribution Function
CH_K2.rte	Effective Hydraulic Conductivity in Main Channel Alluvium
CH_N2.rte	Manning's Roughness Coefficient (Main Channel)
CIESIN	Centre for International Earth Science Information Network
CMIP3	Coupled Model Intercomparison Project Phase 3
CMIP5	Coupled Model Intercomparison Project Phase 5
CMIP6	Coupled Model Intercomparison Project Phase 6
CN2.mgt	SCS curve number
CO <sub>2</sub>	Carbon-dioxide
CORDEX	Coordinated Regional Climate Downscaling Experiment
CPC	Climate Prediction Centre
CRU TS V4.04	Climate Research Unit Time-Series Version 4.04
DEM	Digital Elevation Model
EFR	Environmental Flow Requirements
ENSO	El Niño-Southern Oscillation
ENVISAT	Environmental Satellite

EPCO.hru	Plant Uptake Compensation Factor
EQM	Empirical Quantile Mapping
ESCO.hru	Soil Evaporation Compensation Factor
ET	Evapo-Transpiration
FAO	Food and Agricultural Organisation
GCMs	General Circulation Models
GHCN-D	Global Historical Climatology Network Daily
GHCN2	Global Historical Climatology Network Version 2
GHGs	Greenhouse Gases
GHMs	Global Hydrologic Models
GPCC v.2018	Global Precipitation Climatology Centre Version 2018
GTS	Global Telecommunication System
GW_DELAY.gw	Groundwater Delay
GWF	Green Water Footprint
GW_REVAP.gw	Groundwater “REVAP” Coefficient
GWP	Global Water Partnership
GW-Q	Groundwater Flow (Main Channel)
GWQMN.gw	Threshold Depth of Water in Shallow Aquifer (Return flow)
GW_RCHG	Groundwater Recharge (Aquifer)
GWS	Green Water Storage
HRU	Hydrologic Response Units
HRU_SLP.hru	Average Slope Steepness
HWSD	Harmonized World Soil Database
IAMs	Integrated Assessment Models
IAV	Impact Adaptation and Vulnerability

IMF	Integrated Modelling Framework
IPCC	Intergovernmental Panel on Climate Change
ISI-MIP6	Inter-Sectoral Impact Model Intercomparison Project 6
IS92	IPCC Scenario 1992
ITCZ	Inter-Tropical Convergence Zone
KGE	Kling Gupta Efficiency
km <sup>2</sup>	Square kilometre
LCBC	Lake Chad Basin Commission
LHS	Latin Hypercube Sampling
LULC	Land Use and Land Cover
LUMIP	Land Use Model Intercomparison Project
m	metre
MBE	Mean Bias Error
md	Modified Index of Agreement
MERIS	Medium Resolution Imaging Spectrometer
MI	Mutual Information
MICE	Multivariate Imputation by Chained Equation
MIPs	Model Intercomparison Projects
mm	millimetre
MODIS	Moderate Resolution Imaging Spectroradiometer
m <sup>3</sup> /s	Cubic metre per second
mm/yr.	millimetre per year
NASA	National Aeronautics and Space Administration
NCAR	National Centre for Atmospheric Research
NCEP	National Centres for Environmental Prediction

NIMET	Nigerian Meteorological Agency
NOAA	National Oceanic and Atmospheric Administration
NRMSE	Normalised Root Mean Square Error
NSE	Nash Sutcliffe Efficiency
NTCF	Near-Term Climate Forcings
OV_N.hru	Overland Flow (Manning's Roughness Coefficient)
PDF	Probability Distribution Function
PGF v.2	Princeton University Global Meteorological Forcings Version 2
QM	Quantile Mapping
R <sup>2</sup>	Correlation Coefficient
RCHRG_DP.gw	Deep Aquifer Percolation Fraction
RCPs	Representative Concentration Pathways
REVAPMN.gw	Threshold Water in the Shallow Aquifer (return flow)
RMSE	Root Mean Square Error
SA90	Scientific Assessment 1990
SD-GCM v2.0	Statistical Downscaling of General Circulation Models Version 2
SI	Sustainability Indices
SLSUBBSN.hru	Average Slope Length
SNHT	Standard Normal Homogeneity Test
SO <sub>2</sub>	Sulphur dioxide
SOL_K.sol	Soil Saturated Hydraulic Conductivity
SOL_AWC	Available Water Capacity (Soil layer)
SOL_BD.sol	Soil Bulk Density
SOL_Z.sol	Soil Depth
SPEI	Standardized Precipitation Evapotranspiration Index

SRES	Special Report on Emission Scenarios
SSPs	Shared Socioeconomic Pathways
SU	Symmetrical Uncertainty
SUFI-2	Sequential Uncertainty Fitting Tool Version 2
SW	Soil Water
SWAT	Soil and Water Assessment Tool
SWAT-CUP	Soil and Water Assessment Tool Calibration and Uncertainty Programs
UDel V5.01	University of Delaware Version 5.01
USDA	United State Department of Agriculture
UTM	Universal Transverse Mercator
VIM	Variable Importance Measure
WA	Water Availability
WF	Water Footprint
W/m <sup>2</sup>	Watt per square meter
WMO	World Meteorological Organization
WSI	Water Security Indicator
WSS	Wavelet-based skill score
WYLD	Water Yield
$\Delta$	Delta (bias correction factor)
$\mu$	Mu (transfer function)
$\sigma$	Sigma (standard deviation)
95PPU	95 Percent Prediction Uncertainty



## CHAPTER 1: INTRODUCTION

### 1.1 Research background

Worldwide, increased population and economic activity have been correlated with growing water scarcity (Wang et al., 2015). Water shortages have become more common in recent years, even in areas with abundant rainfall, little seasonal change in rainfall, and rather dense river networks. One of the main causes of the extended drought and water scarcity in recent years is the large seasonal variation of rainfall in a given year caused by global climate change (Diffenbaugh et al., 2015).

According to several studies, the global spatiotemporal distribution of rainfall, including magnitudes (Guerreiro et al., 2018; Neelin et al., 2017), frequencies (Benestad et al., 2019; Fischer and Knutti, 2016), and intensities (Fowler et al., 2021; Harp and Horton, 2022; Konapala et al., 2017), has increased due to global warming, which has affected how the hydrologic cycle functions (Douville et al., 2021).

Over the past three decades, the world has experienced rapid population and economic growth, which has led to high demand for water across a variety of industries. Between one to two billion people are already affected by water scarcity worldwide, especially for people who, for the most part, live in drylands, where there is the greatest global disparity between water supply and demand (Wang et al., 2016).

This barrier suggests that decisions made about water management as a result of climate change will have a substantial impact on drylands and the inhabitants of such areas. Climate change projections indicate that within a few decades, more people roughly half of the world's population will be living in situations with increased water stress (Byers et al., 2018).

It was suggested in previous studies that around 80% of the world's population is projected to face freshwater hazards due to a variety of issues, including population expansion and climate change (Vörösmarty et al., 2010), which have been predicted to have

an adverse effect on factors of production and in turn, affects agricultural and industrial development (Parish et al., 2012).

When there is a concentration of people or economic activity in one area, the demand for water is great and frequently exceeds the supply, especially when there is little rainfall, and the temperature is only gradually rising. However, as a result of increased temperatures and rainfall brought on by global climate change, catchment hydrology and the quantity of freshwater resources are significantly altered. These have caused water scarcity in areas with high water footprints, which is problematic for developing economies like those in Africa, a continent that is thought to be extremely vulnerable with little capacity for adaptation.

Additionally, a significant contributor to water scarcity is the drastically uneven spatial distribution of water demand. A key factor in the ecosystem's complexity is precipitation. Therefore, even minor changes in the climate would have a big impact on the local rainfall patterns, which would subsequently have an impact on the hydrological regimes that might lead to the extinction of species and a decrease in biodiversity (Chase et al., 2000; Rashid et al., 2015).

These factors such as climate change, population growth, rapid urbanisation, and economic growth will continue to put pressure on the available water resources under hydrological uncertainty, and research achieving water security in ungauged basins in developing regions remains a significant developmental challenge (Flörke et al., 2018; Hirpa et al., 2019).

Studies have shown that it is preferable to establish adaptation measures to reduce the adverse effects of climate change because there are considerable uncertainties in the development of water resources for the future (Sýs et al., 2021).

### **1.1.1 Water footprint concept as a tool for water security governance.**

The efficient management of the currently available water resources is of the utmost importance, particularly in watersheds where freshwater resources are scarce or limited, and rivers might potentially provide household, industrial, and agricultural water. In these areas, modelling the temporal fluctuations in river flow is a requirement for the efficient and long-term management of the river basins (Esmaeili-Gisavandani et al., 2021).

Although it is crucial for policymakers and governmental organisations to address specific issues related to water and food security governance, it is also crucial to understand the factors that have an impact on these issues. This can be done by conducting evidence-based studies on the impact of watershed socioeconomic dynamics and climate change on hydrological drivers like streamflow, soil moisture, evapotranspiration, aquifer recharge etc., on freshwater sustainability and how these factors are spatially and temporally distributed.

This necessitates an urgent need to develop new strategies for reliable assessment of historical, current and projected changes in water resources sustainability to mitigate hydrological hazards and enhance water management at the local basin scale. Water resource sustainability in this context, refers to the availability of adequate or sufficient quantity of water for human, industrial and agricultural use for ecosystem sustainability.

Water footprint studies have become a technique for identifying regional and temporal water use patterns, which helps with global water governance and sustainable water management (Galli et al., 2012; Hoekstra et al., 2012; Vanham and Bidoglio, 2013).

By coupling producer and consumer perspectives of efficient water resource allocations through a shift from limitations to consumptions, distribution, and ways to address the sustainability of the resource at local catchment and global scale, the water footprint assessment provides a basis to complement traditional approaches of water demand and supply (Galli et al., 2012; Senbel et al., 2003). This approach according to Hoekstra, (2009)

“...show the importance of human consumption and global dimensions in good water governance”.

Furthermore, long-term model simulation of blue, green and greywater facilitates the identification of water security hotspots indicating both resource sustainability and quality that shows abrupt temporal change points and locations at sub-basins which, it is likely to occur using the water footprint concept (Schuol et al., 2008; Zang and Liu, 2013).

As suggested by Rodrigues et al., (2014), that water resource managers and other stakeholders must identify these hotspots where surface or groundwater needs to be abstracted with the possibility of reducing supply without impairing the need of the downstream consumers where this may create inter-basin and transboundary water resource conflicts.

However, in developing a strategy using hydrologic model to manage the freshwater resources, one of the most crucial hydrological factors supporting the sustainability of aquatic ecosystems, flood forecasting, and drought warnings at various basin scales is river discharge (Couasnon et al., 2020; McNally et al., 2017), and this important hydrological data is inadequate spatially where reliable and critical river basin water management decisions and planning are a necessity.

This is usually the situation in developing arid/semi-arid regions of the world where hydrometeorological gauging stations are sparse (Krabbenhoft et al., 2022; van de Giesen et al., 2014), and the number and quality of data from such stations are declining (Rodríguez et al., 2020).

### **1.1.2 Watershed modelling in ungauged basins**

The task of developing watershed hydrologic models in ungauged regions will be challenging and may amplify the uncertainties of the basin hydrologic dynamics leading to a

flawed depiction of responses and subsequently lead to inadequate water resource policy decisions and adaptation measures.

Additionally, the gridded land surface hydrologic model dataset for food and water security monitoring would be fuelled by rainfall products that excel over data-sparse regions and are accessible over long historical records and near-real-time for contextualising current events and initialising forecasts. This dataset will be helpful for both drought monitoring and the hydrological science community by providing estimates of land surface states and fluxes that can be used (McNally et al., 2017).

These hydro-meteorological station datasets are important components of hydrologic model analysis, and their spatial and temporal dynamics is essential. Unfortunately, this observation data is rarely available and contains missing data points due to systematic errors in their measurements making hydrologic studies difficult in data sparse watersheds.

Alternatively, gridded datasets are adopted as primary inputs in hydrologic modelling studies. However, they are limited to the assessment of historical and current watershed processes and requires objective assessment of their reliability and latency using in situ measurements at local basin scale to properly capture the spatiotemporal variability which may affect water resource availability and the behaviour of the hydrologic responses at basic scale (Panda et al., 2022; Pang et al., 2020).

Advancement in the use of hydrologic models and computational resources in water resources research has led to several large-scale modelling studies using the water footprints concept to further the course of intelligent allocation of the different components of freshwater resources at the regional and global scale that allows for the investigation of the role of climate change and land use in the dynamic changes of projected freshwater resources (Abbaspour et al., 2015; Zang and Liu, 2013).

However, these studies may be difficult at local watershed scales, due in part by the necessity to rescale general circular models by accurate downscaling, re-gridding and correcting the biases in the dataset necessary to condition the models to reproduce reliably of not just the past and present, but also the projected watershed hydrology based on anticipated carbon dioxide (CO<sub>2</sub>) emission scenarios within an acceptable uncertainty limit to overcome the limitations of the gridded datasets (Nkiaka et al., 2022; Shiru and Chung, 2021).

### **1.1.3 Uncertainty management in watershed modelling**

The initial source of uncertainty for modelling the effects of climate change on hydrology and water resources is the choice of the control emission scenario, which stipulates the estimated development of the amount of greenhouse gases in the atmosphere based on the projection of the social and economic development of society (Holtanova et al., 2014; Sýs et al., 2021).

The scale and structure of the general circulation models (GCMs) are additional sources of uncertainty, and choosing the GCM has a bigger impact on hydrological changes than choosing the emission scenarios (Velázquez et al., 2013), while Wilby and Dessai, (2010) argued that the capability of the downscaling of the GCM to a finer scale spatially may not necessarily result in a reliable prediction and is significantly limited by available meteorological data and their quality.

Despite this, there is still a great deal of uncertainty, particularly about the GCM simulations that were used to create the hydrologic system. The baseline and future uncertainty of hydrological simulations can therefore be affected by the modeler's subjective choice of meteorological data.

It is also crucial for all water accounting studies to comprehend how the combined parameter transferability concept behaves when trying to quantify the effect of input choices on blue and green water computation in a river basin.

Current assessment strategies in literature require further review and reintroduction of new and objective data refinement processes and procedures to further manage and improve modelling to minimize inherent uncertainties especially in data-sparse regions through accurate parameterization before their application in model-based water resources assessment.

## **1.2 Overview of research questions**

The challenges faced with watershed modelling at local basin scale will require an objective and careful understanding of the various phases of model development from data choice, parameterization, assimilation techniques to overcome the limitations of local watershed hydrologic responses.

These processes can be achieved by using alternative datasets, approaches like machine learning integrated into traditional modelling processes to improve water resource assessment for effective policy decisions. However, the research questions that are pertinent to achieve our desired goals based on identified gaps in literatures are listed below.

1. How reliable are available gridded climate datasets and interpolation techniques to spatially replicate the historical basin climate in data-sparse regions?
2. What evidence-based reliability and performance evaluation of global climate models using gridded datasets can reproduce and improve the climatology of a basin in data-sparse regions to further reduce the transfer of uncertainty between the climate model and hydrologic modelling process?
3. How effectively, and objectively will hydrologic modelling, integrated with machine learning approach efficiently simulate the target basin hydrologic fluxes within the acceptable uncertainty band?

4. Will satellite-based remote sensing data effectively improve the hydrologic modelling process to overcome the shortcomings of the inadequate land-based observations in data sparse regions?
5. What are the impacts of climate change on projected changes of hydrologic variables that influence the availability of blue and green water resources at spatial and temporal levels?
6. How do the dynamic changes of catchment hydrology due to climate change and anthropogenic activities influence the risk of water scarcity and vulnerability at the basin scale to influence the environmental sustainability of freshwater resources?

### **1.3 Aim of the research**

This research is aimed at understanding and improving the climatological and hydrologic process representation in data-sparse regions by critically and objectively model water demand and supply within an integrated assessment framework that is homogeneous, internally consistent and captures the interactions and feedback with other natural and human systems.

This can be incredibly helpful in accurately assessing the present and future picture of basic-scale water sustainability. By explicitly simulating the effects and feedback of both natural processes (climate model and land use model) and human systems (anthropogenic forcing, land use change, and socioeconomics changes).

This integrated modelling framework will make it easier to estimate water resource demands and supplies on a river basin scale and improve future local water policy decision for adequate mitigation measures and resilience to hydrologic hazards.

### **1.4 Objectives of the research**

The stated research aim can be achieved through the following outlined objectives:



1. To investigate the reliability and performance of various gridded climate datasets in data-sparse regions using a refined quality controlled observed data and multicriteria decision approach.
2. To investigate and develop a new framework for effective and efficient evaluation of the global climate model suitable for the hydrologic modelling process to reliably represent historical and projected basin hydrology for accurate water resource assessment.
3. To investigate the efficacy of satellite-based products to reliably represent the hydrologic fluxes of the integrated model framework for the assessment of blue and green water resources in data-sparse regions.
4. To investigate and propose a new water footprint accounting framework for reliable assessment of blue water resources using model-based parameters, available water use information and gridded population data in data-sparse regions.
5. To investigate the influence of climate change and anthropogenic activities on the historical and projected changes of basin hydrology and their impact on temporal and spatial variation of green and blue water flow and storage respectively.
6. To investigate the historical and projected changes in spatial and temporal patterns of basin-scale hazards to green and blue water sustainability at the local basin scale using the proposed framework.

### **1.5 Scope of the research**

This research work was conducted to provide a holistic approach to small scale watershed modelling where traditional approaches may be difficult to apply due to limitations of data requirement. However, hydrologic information will be required to provide policy direction for proper river basin planning to mitigate the impacts of hazards from extreme events and provide adequate resilience and adaptation measures or coping mechanisms to projected

water sustainability concerns due to climate change and socioeconomic activities that may exacerbate transboundary water conflicts. The identified limitations are listed below.

- The research conducted does not consider the effects local watershed practices such as irrigation scheduling and water withdrawal, reservoir regulation due to unavailability of data.
- The outcome does not take account future land use and topographic (terrain and slopes) changes which might be significant driving factors governing the responses to future basin water resource sustainability dynamics.
- A conservative estimate of monthly blue water footprint was used without accounting for monthly variation due to unavailability of actual sectoral water use information.
- The study does not consider greywater in the modelling process due lack of observation data to reliably calibrate and validate the hydrologic fluxes for greywater quantification.

## **1.6 Rationale for study area selection and research motivation**

The Lake Chad basin is one of the biggest endorheic (landlocked) hydrologic basins in the world, and situated in Sub-Saharan Africa, with unfavourable hydro-climatic conditions that may have contributed to a drop in primary production, widespread desertification, and land degradation (Shiferaw et al., 2014), and the well-known "Sahelian paradox," which describes how the extreme drought conditions in the 1970s and 1980s caused the huge Lake Chad to shrink, is just one example of the intricate hydrological dynamics at play (Ndehedehe et al., 2016).

The high spatiotemporal variability of rainfall and soil moisture in the basin are some of the notable challenges that call for further understanding of the regions' hydrological processes, especially with the evident impact of changes in global climate. Hydrologic

variability of the basin was found to be highly sensitive to rainfall fluctuation, and this has a direct impact on the lake level and size.

Since the lake is closed, a decrease in precipitation events in the basin may ultimately result in a net decrease in inflow from the tributaries and a subsequent decrease in the lake level and size. This decrease may have a significant negative impact on the local population, whose source of livelihood depends on this natural resource.

Furthermore, regions with similar features are common around the world and are typically termed “data-sparse” because of inadequate ground-based information, like traditional gauge measurements of hydrometeorological data are inconsistent, difficult to retrieve due to government bureaucracies, and high cost of logistics and the management of reliable in-situ stations over large heterogeneous landscapes has made traditional approaches to hydrologic modelling a challenge.

The results from global or regional basin scale assessments tend to be unrealistic and quite uncertain for such locations with high spatiotemporal variability. The regional-based assessment as demonstrated by Cook, (2008), states that “...the uncertainties are likely overstated, as most models simulate small changes in rainfall over the Sahel and the projected increases are likely influenced by a few outlier models” and the flaws in the input rainfall data might be amplified by the nonlinearity of the hydrologic process (Maggioni and Massari, 2018).

Therefore, it is important to provide alternative solutions that may sufficiently describe the climatic and hydrologic phenomena like flood, drought, and water resource status etc., in terms of their propagation and characteristics such as duration, severity, onset, intensity and frequency with their projections at the desired watersheds levels for adequate planning and management.

## 1.7 Thesis Structure

This thesis comprises eight chapters. A brief description of the individual chapters is given below:

**Chapter 1:** This chapter is a short background that discusses the research concept, aim and objectives, research questions identified based on the research gap, limitations of the research, rationale for study area adoption and motivation of the research and the overall thesis structure.

**Chapter 2:** This chapter is a detailed background and reviewed literature supporting the research concept. This is detailed understanding around the concept of climate and hydrology research around water resource issues at global, regional and local basin scale. Climate and hydrologic modelling tools and strategies previously employed are discussed.

**Chapter 3:** This chapter discussed summary of the research methods employed and study. The methods employed are related to climate modelling and simulation such as downscaling and bias correction strategies, machine learning approaches for data refinement and selection. Methods employed in surface water hydrology and alternative data selection and integration in hydrologic modelling, uncertainty evaluation and watershed water resources assessment and sustainability status for baseline and projected changes due to climate change and socioeconomic activities.

**Chapter 4:** This chapter discusses the use of systematic approaches to justify the capability of gridded climate product for reliable representation of local basin features for applications in hydrologic impact studies. This has been published as a research paper titled “multi-criteria performance evaluation of gridded precipitation and temperature products in data-sparse regions” in 2021 (Atmosphere, MDPI).

**Chapter 5:** This chapter discusses the development of a systematic methodology for combined GCM downscaling and performance evaluation using machine learning technique

for improved climate models' parameterization to enhance its input for hydrologic models' development for projected climate change impact studies. This have been published as a research paper titled "application of Boruta algorithms as a robust methodology for performance evaluation of CMIP6 general circulation models for hydro-climatic studies" in 2023 (Theoretical and applied climatology, Springer Nature).

**Chapter 6:** This chapter discusses an integrated modelling strategy/framework by coupling the machine learning approach from chapter 5 and the SWAT hydrologic model to investigate the efficacy of the modelling strategy in four sub-watershed of the study area with distinct morphological and climate dynamic features and also analyse the dynamic changes and sustainability of basin-scale projected blue and green water resources due to climate change and its implication for integrated water resource management in one of the sub-watershed associated with recurring incidence of extreme flood events and intense water use in agricultural activities.

This have been published as a research paper titled "Integrated framework for hydrologic modelling in data-sparse watersheds and climate change impact on projected green and blue water sustainability" in 2023 (Frontiers in Environmental Sciences).

**Chapter 7:** This chapter discussed the summary of findings of the entire research and recommendations for improved water policy direction at local watershed scale.

**Chapter 8:** This chapter discussed and summarized the inferences derived from the findings and future work to improved watershed modelling at data-sparse regions.

## **CHAPTER 2: LITERATURE REVIEW**

### **2.1 Climate change and the environment – Key Insights**

#### **2.1.1 Climate Change and the hydrologic cycle**

The strength of the hydrological cycle and its fluctuations through time is very significant in light of the current status of the climate. Due to this phenomenon, the hydrological cycle is primarily made up of moisture evaporation in one location and precipitation in other locations. In particular, when evaporation exceeds precipitation over oceans, the atmosphere can transport moisture to land where precipitation exceeds evapotranspiration and the runoff runs into streams and rivers before discharging into the ocean to complete the cycle (Trenberth et al., 2011).

Solar radiation is the main energy source driving the hydrologic cycle. There has been a net increase in radiation input as a result of the atmosphere's emission of greenhouse gases (GHGs) rising due to human activity (Fawzy et al., 2020). In comparison to pre-industrial times (1850 – 1900), For example, the decade (2011 – 2020), the increase in global temperatures is assessed to have reached the range of 0.95°C – 1.20°C and similarly, hot temperature extremes that occurred once in 50 years on average in a climate without human influence have reached +1.2°C and is projected to reach a warming level of 2.0°C, 2.7°C and 5.3°C respectively (Chen et al., 2021).

Natural cycles, such as seasonal variations or recurrent changes in solar radiation, or unique climate occurrences, like the El Nino phenomenon, define climate variability. The water cycle is anticipated to intensify as a result of a warmer temperature, which will result in more energy in the hydro-climatic system due to an increase in evapotranspiration (Kundzewicz and Schellhuber, 2004).

Key water cycle properties including precipitation intensity, duration, and intermittency change as a result of global energy budget restrictions and regional moisture budgets. These

changes occur as a result of climate change (Döll et al., 2018; Pendergrass and Hartmann, 2014).

Strong evidence suggests that small- and large-scale weather patterns have already been impacted by the global temperature increase, which is greater over continents, in high latitudes, and in high mountains. As a result, the increase in global precipitation is not distributed evenly across the continents; in fact, many regions now frequently experience even less precipitation than they did previously due to long term climate variability and change (Coumou et al., 2015; Di Capua and Coumou, 2016).

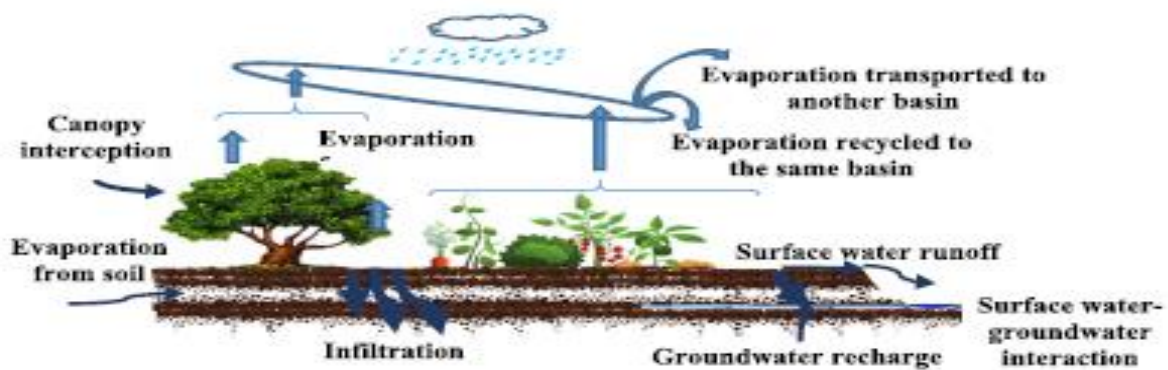
Precipitation, evapotranspiration, local runoff, and river discharge, infiltration, groundwater recharge make up the bulk of the hydrologic cycle (**Figure 2.1**). Any variations in precipitation and evapotranspiration due to increasing GHGs in the atmosphere is expected to modify and have a significant impact on the terrestrial hydrological cycle and the accompanying water flows and storages (Prudhomme et al., 2014), and very few portion of precipitation makes it to ground and surface water storage, particularly in arid and semi-arid locations where actual evapotranspiration is approximately equal to or even higher than precipitation.

Increases in atmospheric water vapour on a regular basis trigger strong amplifying feedback that modify the surface and atmospheric energy balance, exacerbate heavy precipitation events and atmospheric moisture transport, and impact variations in global evaporation and precipitation (Douville et al., 2021). Evapotranspiration is less sensitive to trends in wind speed and is partly controlled by vegetation greening (Zeng et al., 2018), and the increasing trend is attributed to internal variability (Zhang et al., 2016).

Furthermore, actual evapotranspiration can be utilized as a stand-in moisture availability for plant in a given area and, consequently, for their productivity (Martin et al., 2020). Variations in evaporation, which are influenced by vegetation limits on evaporative losses

and are generally driven by an increase in the atmospheric evaporative demand, can affect future water availability (Lemordant et al., 2018; Milly and Dunne, 2016; Scheff and Frierson, 2014; Vicente-Serrano et al., 2020).

Additionally, anthropogenic climate change has altered regional and local streamflow, although a significant trend has not been observed in the global average, however, warming temperatures have contributed to streamflow reductions since at least the late 20<sup>th</sup> century (Martin et al., 2020).



*Figure 2.1: Hydrologic cycle dynamics at watershed scale (Quinteiro et al., 2018a)*

Therefore, the management and protection of water supply in the future depends on an understanding of how anthropogenic and climatic changes will impact the basin hydrologic dynamics. The primary determinants of water balance dynamics are climatic trends and changes in land use and cover (Neupane and Kumar, 2015), and these hydrologic dynamics and drivers have undergone changes globally during the past few decades, and they will probably undergo much more changes in the ensuing decades (IPCC, 2018).

These hydrologic drivers are usually applied in studies to investigate and understand the trends and changes in water resource availability (Fazeli Farsani et al., 2019; Sýs et al., 2021; Zhang et al., 2022), sustainability (Gesualdo et al., 2019; Keys and Falkenmark, 2018; Nkiaka et al., 2022), hydrologic extremes (flood and drought hazards) (Kim et al., 2023; Rodríguez et al., 2020) and the benefit is that they are generally monitored easily. These



indicators are frequently utilized as the results of hydrologic simulations of the effects of climate change.

Data quality and availability, however, can differ. Furthermore, some of the indicators can be used to assess the effects on hydrologic cycle elements for which there is less data (such as recharge for renewable groundwater resources) and where it is more difficult to estimate the effects of climate change (such as actual evapotranspiration as an indicator for plant productivity) (Martin et al., 2020).

The magnitude of the impact of climate change on these basin hydrologic drivers is tied to the emissions levels of GHGs in terrestrial ecosystems and efforts to quantify the effects are evolving through the application of computer (hydrologic) models. The assessment of the hydrologic effects of the alterations is complicated by the inherent uncertainties generated along the impact model chains (Martin et al., 2020).

Global climate models (GCMs), greenhouse gas emissions, and the creation of socioeconomic scenarios are only a few examples of the sources of uncertainty in the process (e.g., shared socioeconomic pathways (SSPs)), downscaling techniques and type of watersheds models and their limitations (Gidden et al., 2019).

### **2.1.2 Climate Models and climate modelling**

Climate datasets required for impact analysis varies enormously depending on the type, scale and purpose of study. Two essential weather variables used as inputs to watershed models are precipitation and air temperature. For effective modelling and prediction of extreme events and hydrological processes from models, an accurate depiction of the temporal and geographical variability of the key climate features is crucial (Duan et al., 2019; Laiti et al., 2018).

Consistent monitoring of precipitation and air temperature in sufficient quantities to accurately depict the weather at the basin, gauge stations should ideally be connected in a

relatively dense network. In practice, there is frequently a lack of gauge stations, and point-based observations, which cover a small region, are insufficient to capture the spatial and temporal variability of meteorological variables (Roth and Lemann, 2016).

Several studies for example Awange et al., (2016); Gampe and Ludwig, (2017); Prein and Gobiet, (2017); Tan and Duan, (2017) have evaluated basin scale performance of alternative gridded precipitation data at the local basin scale using different single performance metrics such as statistical metrics, machine learning approach, trend analysis etc. because the accuracy of the datasets varies from one region to the other as acknowledged.

However, the application of any chosen dataset based on its superior performance from a singular metric may be subjective especially when the re-parameterization process is utilized, which may distort some watershed representation of climatic features, and this has been identified as a weakness (Shiru et al., 2019b).

Therefore, a robust approach is required involving evaluation using multiple metrics to establish and reach a baseline decision that allows for the choice of a climate product that will reliably represent the climatic features within the acceptable uncertainty limit.

But also has the requisite scale and timestep with high spatial variability and favors its application in impact study. See **4.2** for specific literature that addressed this identified weakness and related to research question and objective number 1 respectively.

Furthermore, the application of gridded dataset in impact study can only address the understanding of historical and current watershed hydrologic processes where data is available. However, in order to predict the potential effects of climate change and its significant impact more accurately on the natural environment, decision-makers and managers of water resources need information relating to future changes in temperature and hydrologic variability. These shortcomings in the use of gridded datasets can be addressed by the use of GCMs.

The atmosphere, land surface, ocean, cryosphere are the physical processes of the climate systems represented by GCMs, which are sophisticated numerical climate models. They are the sole reliable instrument available and required for modelling the manner in which the global climate system would react to rising GHGs concentrations (Santoso et al., 2008).

The changes in the quantitative estimates of future climate have gained confidence (Tokarska et al., 2020) and the capability to simulate important basin features such as extreme events, El Nino-Southern Oscillation etc., has improved (Hughes et al., 2014; Seneviratne et al., 2021).

Even though GCMs are thought to be the best at predicting future climate changes brought on by anthropogenic forcing, many impact studies find them to be too coarse (Weigel et al., 2010). The realism of the data is impacted by how the coarse scale GCM outputs are treated before being utilized as inputs to many impact evaluations and research, particularly downscaling techniques (Knutti et al., 2013).

The ability to depict a realistic future climate, ease of use, and the type of climatic information or data needed for impact assessments are just a few of the numerous considerations when choosing a downscaling technique and the model differences lead to a range of climate sensitivities that are most likely between 2°C and 4.5°C, with a best of 3°C (Boko et al., 2007; Niang et al., 2014).

Furthermore, guidance was provided when creating the scenarios for sensitivity assessments (Carter et al., 2001; van Vuuren et al., 2014), it is frequently beneficial to take the output of multiple models into account. Users may find it challenging to choose the best models, especially when there are numerous models available with a wide range of projected results (some models may produce inconsistent results) (Duan et al., 2019).

Some of the criteria for the selection of climate models suggested by Smith and Hulme, (1998) are:

- Vintage: recency of the climate models, which are more likely to incorporate new knowledge in their construction.
- Resolutions: the resolution of earlier versions is coarser than that of newer variants. More spatial information is present in finer resolutions, and important climate variability processes, such as ENSO episodes, are clearly represented. He countered, however, that improved model performance is not always a result of greater resolution.
- Validity: that is how well the climate models simulate the current climate relative to the observation data.
- Representativeness of results: a variety of changes in important climate variables in the research region can be illustrated by choosing sample GCM from the available data.

Furthermore, Collins, (2007), suggested it is essential to remember that uncertainty cascades from broad GCM scales to local scales and ultimately to impact variables of interest. For this reason, it is crucial to adequately account for uncertainty at every stage of impact modelling.

Whilst Knutti et al., (2017), made the case that treating all models equally is pointless and that the increasing number of models with unique traits and significant interdependence finally justifies giving up on strict model democracy. They also suggested a weighting scheme for multimodel climate projections that takes into account both significant variances in model performance and interdependencies.

The basis for this is the massive model spread in current climatology, which means that biases in some models will be so great that model democracy will be hard to defend. Working with projected anomalies relative to today can be problematic in scenarios where processes

are sensitive to the base state, making scaling methods and bias correction impractical (Weigel et al., 2010).

In certain instances, emergent constraints can improve projections and are relevant for model evaluation. Large initial-condition ensembles can be challenging to combine with single runs from other models. Lastly, model dependence becomes more significant as modelling institutions replicate more code (Knutti et al., 2013).

Collins, (2017), however, emphasised that caution must be used to take into account model drifts, flaws that are present in all models, and the more difficult-to-understand problems with the physical realism of all models. In addition, he suggested that more work be done to enhance models and comprehend how to interpret their projections in order to fully utilise this new technique, which includes assessing robustness and sensitivity and displaying weighted and unweighted projections.

Additionally, previous literatures that assess the capability of GCM applications in impact studies has revealed that various techniques have been used such as statistical metrics (Gu et al., 2015; Wu et al., 2017), evolving machine learning approaches (Ahmed et al., 2019b; Pour et al., 2018; Shiru et al., 2019b).

However, several approaches used in literature may be promising but the scale of data (timestep) and purpose of data used in impact study have not been justified and the weighing scheme of GCM performance based on plausible criteria can yield contradictory results (Chandler, 2013; Raju et al., 2017).

While other drawbacks alluded to machine learning approaches is data overfitting and the inability to capture the temporal variability such as trends and the frequency of climate extremes which is an important factor in model performance in the successes of impact study (Shiru et al., 2019b).

In many applications of climate modelling for simultaneous projections of several climate variables, the choice and selection of a simulator is sometimes the main source of uncertainty in such projections, despite the fact that these simulators represent our greatest understanding of the climate system at this moment (Hawkins and Sutton, 2009).

Therefore, seeking improvement in quantitative climate prediction at regional and local basin scale that will impact economies, people, and ecosystems require a methodology in GCM selection that narrows the uncertainty and address the drawbacks in model (response) and scenario uncertainty by providing an approach that will assess the capability of GCMs ensembles to realistically represent spatial and temporal variability and projections that are consistent with local climate after parameterization to finer scales. See 5.2 for specific literature related to research question and objective number 2 respectively.

### **2.1.3 Uncertainty in climate modelling**

Understanding and measuring climate variability, climate change, and their effects requires the use of climate models (Flato et al., 2013). In the process of simulating the regional effects of climate change, a series of computer models translate global projections of GHG emissions and atmospheric concentrations into effects on local water supplies, hydrological processes, and extreme events (such as floods and droughts).

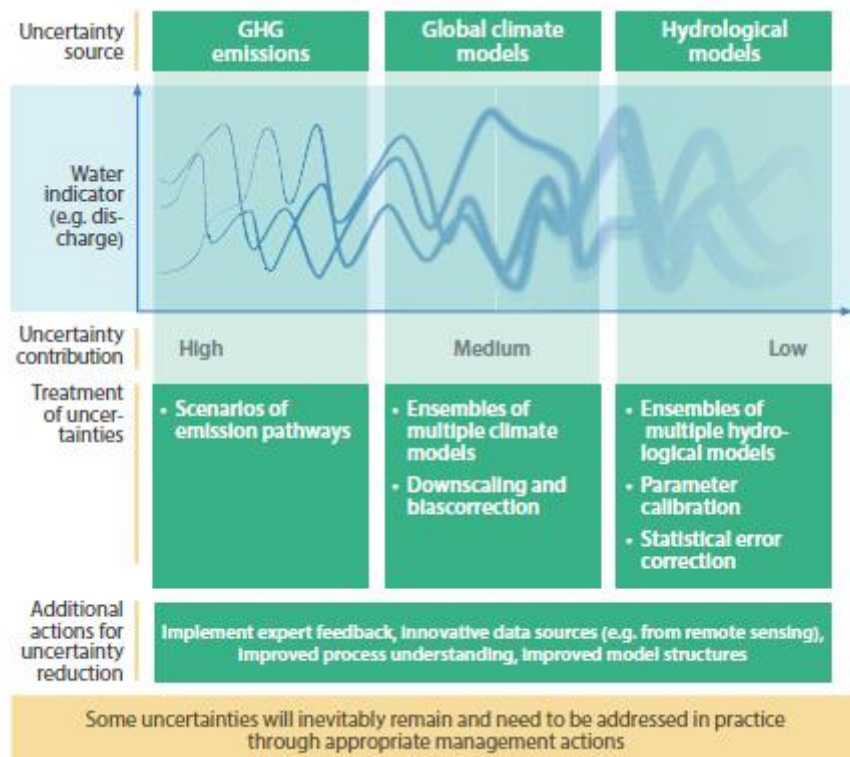
The process begins with the application of physically-based GCMs that are transformed into regional to local climate and weather simulations through downscaling to correct systematic errors or biases relative to observations. This transformation introduces a great deal of uncertainties that move from one layer to the next and are then picked up by the individual uncertainty of the subsequent layer, which ultimately results in a great deal of combined uncertainties at the bottom of the cascade (Wilby and Dessai, 2010).

As a result, decision-makers frequently struggle to understand the consequences of projections of the future impacts of climate change. Such uncertainties are challenging to

account for. So, in order to aggregate information about uncertainty, researchers employ a variety of strategies to narrow their effect in impact studies (Smith et al., 2018).

GCMs and global hydrologic models (GHMs) can also be significant causes of uncertainty. During application, each model group adds its own components of innate uncertainty. Instead of using a single model to address this situation, researchers frequently use ensembles of models. However, it is frequently still challenging to determine which model stage (i.e., GCMs or GHMs) provides the majority of the overall uncertainty in a given scenario (Martin et al., 2020).

A summary of the various sources of uncertainty in the climate impact models for water resources, including GHG concentration pathways, GCM, and GHM, is shown in **Figure 2.2**, along with some recommendations for reducing uncertainty.

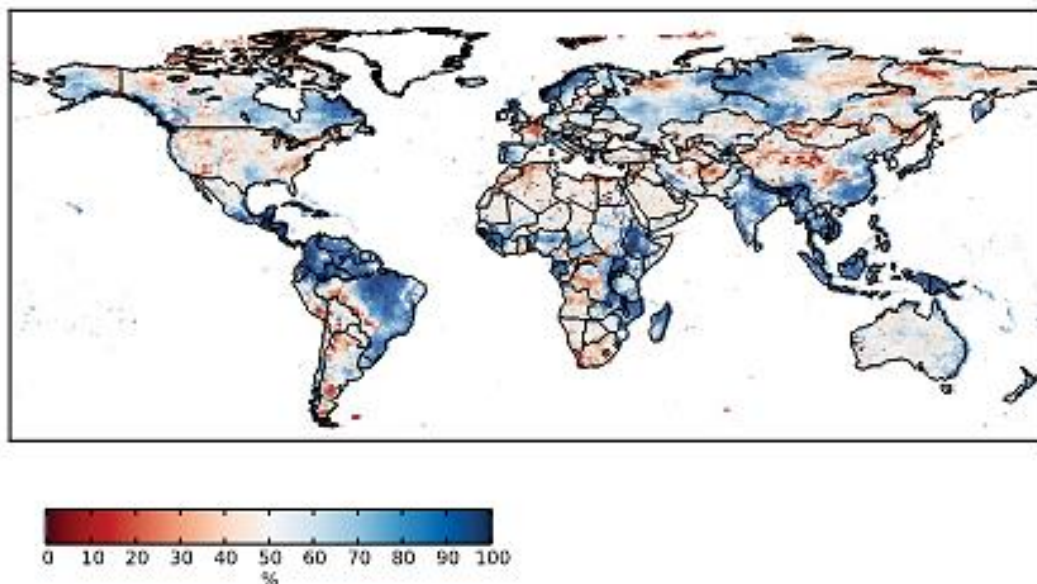


**Figure 2.2:** General overview of the main climate modelling uncertainty sources (Martin et al., 2020)

The interlinkages of these source of uncertainties between the GCM and GHM and the regional application of these models, according to Schewe et al., (2014), may vary across the

different part of the globe. Findings from their study (**Figure 2.3**), showed that GCM uncertainty is notably significant in tropical and northern regions, which are marked by substantial levels of precipitation, and GHMs account for the majority of projections uncertainty in relatively dry sub-tropical and arid regions.

While (Hattermann et al., 2018), argued that uncertainty associated with GCMs is frequently much greater than the impact of choosing a particular GHG concentration scenario. However, this assertion may be significant based on the time period considered since the impact of scenario uncertainty makes a little difference only up to 2050s (IPCC, 2018)



**Figure 2.3:** Ratio of GCM variance to total variance as a measure of uncertainty. In red areas, GHM uncertainty predominates, Greenland has been masked (Schewe et al., 2014)

Therefore, the potential for uncertainty reductions depends on the source of uncertainty, data type, expert understanding of the complexity of the model structures, parameterization process and exploration of evolving and effective strategies to limit uncertainty propagation of historical and current climate, while understanding the potential of quantifying uncertainty in the field of GHG emission and concentration scenario is low, due to lack of clarity concerning future development.



Climate models assessment based on historical and current basin features will have a profound potential in achieving uncertainty reduction and certainly both modelling process will benefit from an improved understanding and implementation of downscaling and bias correction techniques, innovative capability assessment of the choice of GCMs that are capable of detecting basin scale hazards such as climate extreme values (return period of drought and flood hazards), which are often poorly reflected by GCMs due to distortion of climate signals by insufficient bias correction schemes (See 5.5.3).

However, even with an improved dataset, there will obviously still be some uncertainties about future development and projections. This poses difficulties for the creation of adaptation plans in the water-related industries and necessitates effective management measures.

These challenges can be reduced by adopting the right type of climate models that accurately reproduce projections that is acceptable within the scientific research community to improve hazard and adaptation planning at local basin scale where hydrologic modelling information is predicated on the accuracy of the climate data. Refer to 6.2 for additional literature.

#### **2.1.4 Climate change scenarios**

The study and assessment of climate change must include scenarios. They help us comprehend the long-term implications of short-term decisions by enabling researchers to have a sufficient grasp of the intricate connections between the climate system, ecosystems, and human activities. They also allow us to analyse several prospective futures in the light of significant future uncertainty (Moss et al., 2010).

Most importantly, scenarios have traditionally been crucial for fostering integration amongst numerous research communities, for instance by providing a standard platform for

the assessment of mitigation strategies, impacts, alternative adaptation strategies, and changes to the physical earth system (Riahi et al., 2017).

Established scenarios include the earlier one's developed and adopted by the Intergovernmental Panel on Climate Change (IS92, SA90, and SRES, and RCPs) and the more recent is the share socioeconomic pathways (SSPs) (van Vuuren et al., 2017).

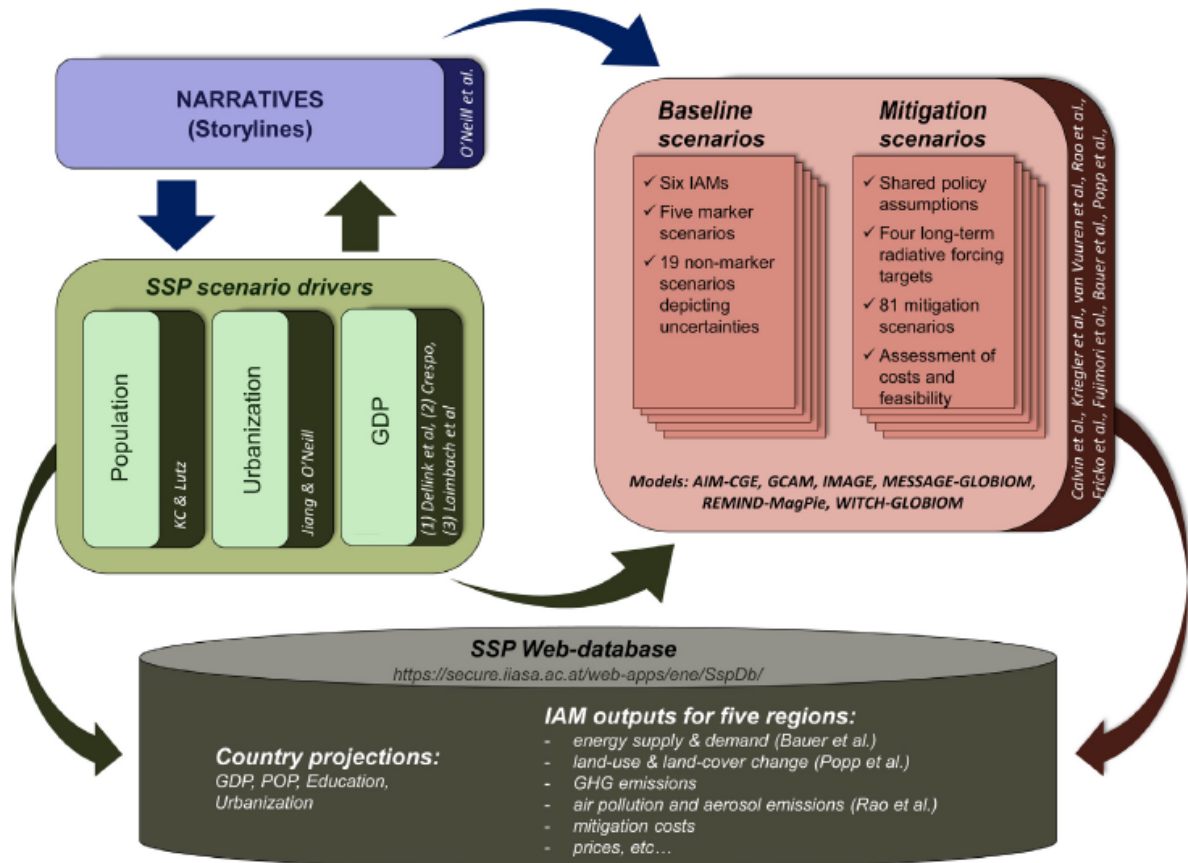
The SSPs are a collection of five hypothetical storylines that describe potential future developments of the anthropogenic climate change drivers namely, greenhouse gases, chemically reactive gases, aerosols, and land use in a way that is consistent with socioeconomic developments, which are crucial to climate research (O'Neill et al., 2016).

The forcing levels covered by the RCPs were the primary focus of the present SSPs mitigation scenarios development. From the standpoint of emissions mitigation, the resulting pairing of SSPs and RCPs is the scenario matrix's first thorough implementation (van Vuuren et al., 2014).

The SSPs were created to present five clearly different scenarios for potential future socioeconomic changes in the absence of explicit extra policies and actions to reduce climate forcing or improve adaptive capacity (Riahi et al., 2017). The development comprise the five steps illustrated in **Figure 2.4**:

1. Each SSP's core underlying logic is provided by the storytelling design, which focuses on those aspects of socioeconomic change that are frequently outside the purview of formal models.
2. Extension of narratives in terms of "input tables" for the model, describing qualitatively the primary SSP characteristics and scenario assumptions.
3. Using mathematical models to elaborate on the fundamental SSP components in terms of economic and demographic dynamics.

4. The SSP baseline scenarios' changes to the energy system, land use, and emissions of greenhouse gases and air pollutants are elaborated using a collection of Integrated Assessment Models (IAMs).
5. IAMs' elaboration of these components in relation to SSP mitigation scenarios.



**Figure 2.4:** Schematic diagram of the steps in developing SSPs, with narratives, socioeconomic scenario drivers, SSP baselines and mitigation scenarios (Riahi et al., 2017).

The scenarios covered here are only applicable to tier 1 components, where each experimental design's detailed description, rationale, and key characteristics are briefly reviewed. Tier 1 components in this context refers to scenarios that prioritized a wide range of uncertainty in future forcings pathways essential to research in climate science, IAM, and IAV studies (O'Neill et al., 2016), that provides accurate understanding of physical climate system consequence and the impact of the scenarios on the society, as well as informed

mitigation and adaptation policy considerations. The explanation for the selection of the driving SSP and narratives, as well as the relevance of the forcing paths, are given.

**SSP5-8.5:** The RCP8.5 route is updated in this scenario, which is meant to be used by a number of other CMIP6-Endorsed MIPs to help them with their scientific research. It represents the most advanced possible future pathways as stated in the IAM literature.

The SSP5 scenario was chosen for this forcing path because it is the only SSP scenario with emissions high enough to provide an  $8.5\text{Wm}^2$  radiative forcing in 2100 (O'Neill et al., 2016). In this scenario, the push for economic and social development is prioritised along with high exploitation of abundant fossil fuel resources and the adoption of resource- and energy-intensive lifestyles globally. Climate mitigation challenges are high, while adaptation challenges are low (van Vuuren et al., 2017).

**SSP3-7.0:** In terms of potential future forcing paths, this scenario reflects the middle to high end. It closes a particularly important gap in the CMIP5 forcing pathways by reflecting a level of force comparable to that in the SSP2 baseline scenario. SSP3-7.0 will play a significant role in LUMIP and AerChemMIP in addressing scenario-relevant queries regarding the sensitivity of regional climate to land use and aerosols.

This is due to the fact that SSP3-7.0 is a scenario with significant land use change, specifically decreased global forest cover, and high near-term climate forcing (NTCF) emissions, mainly  $\text{SO}_2$  (O'Neill et al., 2016). In this scenario, countries prioritise regional to national energy and food security concerns over larger development aspirations, posing medium to high hurdles for climate mitigation and adaptation (Riahi et al., 2017).

**SSP2-4.5:** This scenario updates the RCP4.5 pathway and represents the middle of the possible future forcing pathways. The Coordinated Regional Climate Downscaling Experiment (CORDEX), which will also employ SSP5-8.5 for regional downscaling, is one

of the CMIP6-Endorsed MIPs that will use it as a reference experiment. This product will be useful to the IAV community.

SSP2 was chosen because it is a scenario that is relevant to IAM/IAV research and combines intermediate societal vulnerability with an intermediate forcing level, as well as because its land use and aerosol routes are not excessive in contrast to other SSPs (O'Neill et al., 2016).

This scenario/pathway illustrates moderate obstacles to climate mitigation and adaptation where income growth and development occur unevenly, with some countries making reasonably excellent progress while others fall short of expectations in achieving sustainable development targets (van Vuuren et al., 2014).

**SSP1-2.6:** This scenario updates the RCP2.6 pathway and is the least force-producing pathway that could possibly occur in the future according to IAM research. It can enable evaluations of this policy objective because it is predicted to result in a multi-model mean warming of much less than 2 °C by 2100.

SSP1 was picked because it involves a significant change in land use, and this scenario is very important because it has minimal vulnerability, few problems for mitigation, and little forcing signal (O'Neill et al., 2016).

This narrative considers gradual shift towards sustainable and more inclusive development in line with perceived environmental boundaries through increased commitment to achieving sustainable development goals where consumptions is oriented towards lower resource, material growth and energy intensity (Riahi et al., 2017).

The SSPs are designed to serve as a vital tool for tying together research on climate change from various disciplinary perspectives, including the physical climate system, hydrology and water resources planning, impacts of climate change, and solutions for

adaptation and mitigation. They can be utilised at several geographic scales, including global, regional, and local scales.

The details of scenario model intercomparison project experiment design is given in

**Table 2.1.**

*Table 2.1: Climate model intercomparison project phase six forcing emission thresholds and concentrations.*

Scenario Name	Forcing threshold	Forcing concentration (Wm <sup>-2</sup> )
SSP5-8.5	High	8.5
SSP3-7.0	High	7.0
SSP2-4.5	Medium	4.5
SSP1-2.6	Low	2.6

Source: (Riahi et al., 2017)

In this proposed research work, climate change impacts on green and blue water resource availability and sustainability at local basin scale and their deviation from baseline using scenarios SSP5-8.5 and SSP2-4.5 were tested. These two scenarios were quite necessary for adaptation planning at local basin scale.

**2.2 Watershed hydrology and the water balance dynamics**

**2.2.1 Hydrologic dynamics and climate change**

Hydrological model simulations driven by individual and combined forcing demonstrate that decreasing precipitation can result in higher deficits in soil moisture, streamflow, and water table depth than other forcings, but they also demonstrate that these factors are not linearly cumulative when applied in combination (Hein et al., 2019).

The decrease in precipitation which is largely related to climate change may alter the ecosystems water and energy balance, affecting the hydrologic basin dynamics thereby affecting water resources and current and projected ecosystem services (Martin et al., 2020).

Precipitation is very important and serves as the watershed’s sole source of water supply (Zhu et al., 2022) and its combination with air temperature dynamics in the hydrologic process define the water balance variables characteristic such as surface runoff, soil water

content, subsurface runoff, evaporation, and water vapour movement realizing the hydrologic cycle responses.

A number of studies, for example Eromo et al., (2016); Ranzani et al., (2018); Savelsberg et al., (2018); and Wagena et al., (2018) have confirmed the existence and continuous effects of climate change on the entire planet in the future. In particular the influence is tangible on hydrology, water resources and terrestrial environment, and analysis of the scope of the alleged impact on different component of the earth's system has proven to be important (Fatahi et al., 2021).

The response of the hydrologic dynamics is not basin-specific; for instance, the frequency of rainfall, the rate of soil infiltration, the vertical profile of soil moisture, and the level of the water table all affect highly nonlinear processes like land surface runoff and groundwater recharge.

There is a nonlinear relationship between precipitation and groundwater recharge in the tropics, where large seasonal rainfalls linked with internal climatic variability contribute disproportionately to recharging (Cuthbert et al., 2019; Taylor et al., 2013).

However, groundwater fluxes in arid areas are frequently less susceptible to climate change than in humid areas. This can either result in a long-lasting, first undetectable, hydrological response to global warming or serve to temporarily mitigate the consequences of climate change on water resources (Cuthbert et al., 2019).

Furthermore, Hein et al. (2019) demonstrated that simulations of hydrological models driven by individual and combined forcing show that decreasing precipitation can cause greater deficits in soil moisture, streamflow, and water table depth than other forcings, but they also show that these factors are not linearly cumulative when combined.

Therefore, model variations under the same forcing scenario, as shown by CMIP6 models, continue to be the primary cause of uncertainty for projection of changes in regional

precipitation to achieve a consistent modelling (Lehner et al., 2020), and reliable hydrologic response dynamics due to climate change at local basin scale necessitates the benefits of uncertainty analysis to allow for consistent comparisons of current and projected GCM outputs which are important to improve hydrologic modelling decisions to increase confidence in impact analysis.

Because according to Fatichi et al., (2016), projections at local scales and finer resolutions are subjected to similar uncertainties as those at the global scale and this allows for a robust conclusions for proper mitigation and adaptation planning.

## **2.2.2 Hydrologic modelling and process optimization**

### **2.2.2.1 Hydrologic models**

A hydrologic model is a set of computer codes that are compiled and run with a suitable sets of initial boundary conditions, parameters, and forcings that is capable of simulating the spatial and temporal evolution of hydrologic fluxes and states of a time series of hydrologic responses of a watershed (Ogden, 2021).

These models are quite important in projecting or forecasting future states, hypothesize and design of watershed dynamics for integrated management of ecosystem services. Early modelling methods estimated runoff by using conceptualizations. To reduce the complexity of the problem to one that could be addressed arithmetically, pre-computer approaches used sound reasoning supported by appropriate simplifying assumptions (Cherif et al., 2023).

Different computational hydrologic models continue to use each of these conceptualizations. The techniques usually tout their simplicity as a positive, despite the fact that these assumptions about heterogeneity, time, and dynamics can offer significant challenges. It is accurate to say that many hydrologic models are overparameterized or depend too heavily on a large number of parameters (Ogden, 2021).



This suggests that they might be overfit, which suggests that the user may find a lot of other parameter values that yield similar outcomes, a phenomenon known as equifinality (Beven and Cloke, 2012), and according to Kirchner et al., (2021) It is exceedingly challenging to impossible to prove that a hydrologic model is accurately simulating the real system.

#### **2.2.2.2 Characteristics of hydrologic models**

In order to be able to express the physical process using mathematical equations, hydrologic models are built on simplifying assumptions. They typically mimic the behaviours of processes from the water cycle, including evapotranspiration, precipitation, infiltration, interception, subsurface flows, etc., and their interactions (Ogden, 2021).

They take into account the different free parameters to generate forecasts that precisely imitate the behaviour of the observable variables in a certain circumstance. These parameters, which are acquired via the calibration procedure, aid in effectively simulating the phenomena at the model's output (Cherif et al., 2023). Thus, the model's feature consists of the following components:

- The watershed and its characteristics (e.g., morphological features)
- Observed variables serve as model inputs.
- The process equations, which may include the calibrated parameters, describe how the modelling system behaves.
- Model initial boundary conditions.
- The variable outputs

The model components are generally design differently and its processes include the water cycle, natural and anthropogenic phenomena which all interact at the watershed system to determine the model outputs based on its intended usage. However, the model also has

internal state variables and changes over time which generally includes reservoirs filling levels, such as production and routing, snow etc.

According to Ogden, (2021), how a model integrates with other applications is described by its external structure. Does the model fit within a bigger system? Does the model provide fast interfaces for parameter estimation, assignment, and input? Is there a command-line interface on the model that other software programmes can use to call it?

This is related to the kind and amount of data that are available to power the model and evaluate its performance. Any successfully run hydrologic model requires spatial discretization, which ranges in complexity from a simple lumped watershed representation to a hyper-resolution three-dimensional unstructured mesh.

### **2.2.2.3 Hydrologic model calibration and validation**

The skills and experience of the modeller are quite important and used to create hydrological models. The range of variables covered by this knowledge is constrained by the finite number of observed data. As a result, the model's predictions for its assumptions, inputs, and parameters are only as trustworthy as those projections (Singh, 1995).

The hydrological model chosen needs to be calibrated to match the characteristics of the watershed under study, but because reality has been simplified, some or possibly all of the parameters used to characterise the model can't be directly linked to field observations, necessitating for a gradual mathematical calibration.

The objective function, a similarity measurement criterion, can be used to compare the simulated values to the observed data when model output data are available. The goal of calibration is to find the model parameters' numerical values that will most accurately represent the observed response. It is the process of choosing groups of parameters' ideal values and an important step in modelling because it affects reliability of the outcome (Cherif et al., 2023).

Model parameterization in hydrologic process is important to provide solutions to equations nonlinearity and scaling transformation between observation and modelled parameter process and this is done to control point data transfer in process-based models which helps in representing model state and behaviour.

The process involved three parts namely a calibration approach, an objective function and observed data for model calibration. The objective function was described by Yapo et al., (1998) as a mathematical equation that links the observed and model simulated outputs and measures the relationship between the parameters.

The choice of an objective function and the observation parameter must be made carefully to preserve the model performance in impact study. Some notable objective functions used in literature are Nash Sutcliff Efficiency (NSE), Correlation coefficient ( $R^2$ ), Root mean square error (RMSE), etc., (Abbaspour, 2015).

The hydrologic model validation process verifies the effectiveness of the chosen parameters after the best ones have been determined. In this step, the model is tested using a set of data that are distinct from those used during the calibration phase and it is often the last stage of the modelling process. In literature, two approaches are generally used.

- A conventional method that entails selecting a portion of the data series used to calibrate the model parameters at random. There must be a difference between the calibration and validation times. This step of internal validation examines the model's sensitivity to the values of the parameters that make it up.
- The entire basin underwent a multicriteria, multiscale validation method. It compares the model's forecasts to data from observations (hydrographs measured at intermediate stations and at piezometric levels) that were not utilised to create the model.

#### **2.2.2.4 Hydrologic model process optimization**

Gaining proximity to the optimum points is the goal of optimization, which aims to enhance model performance in hydrologic modelling. The pursuit of model perfection and the approach to an ideal point are the two components of this concept. Therefore, it's crucial to explicitly separate the improvement process from its endpoint or the ideal state (Cherif et al., 2023).

However, when assessing optimization techniques, convergence is frequently the only thing that is considered, and intermediate performances are ignored. However, enhancing the modelling outcomes is the optimization's primary goal (Abbaspour et al., 2017; Cherif et al., 2023). Since perfection is challenging to obtain, achieving a sufficient level of performance should be prioritised in all modelling processes (Simon, 1996).

#### **2.2.3 Uncertainty in hydrologic modelling**

The demand to consider model reliability and their capability to effectively respond to operational difficulties has been linked to the growing usage of hydrological modelling findings in decision support systems. As the models' limitations and flaws became more apparent, uncertainty analysis was eventually incorporated into the modelling process, (e.g., Abbaspour et al., 2017; Bennett et al., 2013), and this can be characterized based on their type and sources.

There are two distinct types namely, epistemological (reducible) uncertainty, which are associated to lack of knowledge of the process, and stochastic (irreducible) uncertainty which are related to the natural variability of variables in the modelling process. However, this notable sources are uncertainties about the quality of the input data and initial boundary conditions of the model, uncertainty introduced in parameterization and estimation and the uncertainties related to model structure and limitations (Bennett et al., 2013; Cherif et al.,

2023; Gal et al., 2014). A brief description of sources of hydrologic models' uncertainties are given below.

### **2.2.3.1 Data uncertainty**

These uncertainties cover both the response data, such as streamflow, and the uncertainties in the model input data, such as precipitation, air temperature and potential evapotranspiration. They are the result of measurement inaccuracies as well as how this input data was processed (Gal et al., 2014), and in most cases, especially in data-sparse regions only a few rain gauges are used to determine the average rainfall-runoff relationship over a watershed, data errors can also result from spatial precipitation field sampling or from the rating curve representing the relationship between water level and discharge.

In order to minimize this type of uncertainties while initialising a hydrological model, additional spatial data are required to account for certain regional properties of the catchment area, such as land use, soil and geological characteristics, surface elevation, and water management practices (Martin et al., 2020).

However, Arhonditsis and Brett, (2004) argued that a high degree and number of state variables required to build a complex hydrologic model does not guarantees a superior model performance and often times this processes may lead to model over-parameterization which results in misinterpretation and poor predictions of watershed features (Jakeman et al., 2006).

This climate data uncertainty is too complex to be represented in numerical models due in part to imperfect conceptualization and discretization of the climate models and spatial averaging within the grid cells, and the assumptions and parameterization of the models are approached distinctly by the modelling centres and their projections of regional and local basin features differ across GCMs for the same global mean values (Weigel et al., 2010).

The application of these models at regional and local scale will amplify the uncertainty levels of hydrologic models' feedback and due diligence is required by applying evolving and

improved techniques to limit the propagation of data uncertainty in watershed modelling (Lawal et al., 2023b).

### **2.2.3.2 Parameter estimation uncertainty**

A hydrological model's reaction to changes in rainfall-runoff dynamics and processes at the watershed scale is altered by changing its parameters. These variables give a fixed mathematical model whose flexibility needs to be adjusted to achieve accurate predictions in different watersheds (Ogden, 2021).

Given the challenges in determining the values of parameters based on the available data, it is usual practise to alter the parameters to correctly replicate the observed model variables of interests. For example, flow at the watershed's outlet when flow rate measurements are available. This estimating procedure leaves the parameters uncertain. It is most frequently represented by getting many sets of parameters that produce reliable runoff estimation results (Nkiaka et al., 2022).

Researchers have proposed a number of automated strategies that are frequently employed since parameter estimate is essential. By comparing a time-series of model outputs against observations, these techniques often produce objective error measurements that are used to determine a cost function (Ogden, 2021). It is crucial to choose acceptable objective error metrics carefully in order to select those that are most representative of the desired model behavior.

A section of a record is typically used to calibrate a model, and the model's calibration is subsequently evaluated on the remaining piece of the record. This is split-sample testing, which may result in erratic findings depending on the calibration and validation data series selected (Arsenault et al., 2018).

### **2.2.3.3 Structural model uncertainty**

Structural (conceptual) model uncertainty is a reflection of a model's imperfect capability to capture the dynamics of the modelled system due to the model's oversimplified structure in comparison to the modelled system's complexity, the corresponding resolution, or the digital implementation (Cherif et al., 2023). These uncertainties are generally due to representation of processes within the model structure that is not consistent with the true watershed features (Abbaspour, 2015).

In environmental modelling process, structural model uncertainty can be minimized for improved prediction by ensuring the accuracy of the input data was assessed, inadequacy of the data highlighted and improvements to the model structure routinely done to reflect the drawbacks, careful assessment and reduction of uncertainty propagation at the different stage of the source etc. (Martin et al., 2020).

### **2.2.4 Challenges of hydrologic modelling**

Since hydrologic predictions are challenging to produce in all but the most straightforward situations, the model parsimony paradigm is often advised. This paradigm states that the simplest model with the fewest parameters should be chosen to accurately forecast the variable(s) of interest for a given set of relevant inputs (Ogden, 2021).

Empiricism's pervasiveness and tenacious persistence in hydrology are evidence of how challenging hydrologic forecasts are; and as Kirchner, (2006) alluded, that making hydrologic predictions can be done in a variety of ways, but it can be challenging to get the right model for the right reasons.

The majority of river basins worldwide are essentially ungauged with regard to the most important hydrological and water management variables, and these regions are associated with sparse or non-existent hydrologic gauging networks, which may lead to three peculiar

challenges, including data scarcity, data quality, and hydrologic non-stationarity in the modelling process (Visessri and McIntyre, 2016).

Earlier approaches from literature to deal with such challenges, is through the use of regionalization method which arises when modelling on a big scale (e.g., on a regional to continental scale), making it the only way to avoid the enormous regions with no data. Some make the supposition that regions with comparable physiography and climates ought to behave similarly (Visessri and McIntyre, 2016).

Some of the regionalisation concepts generally used in literature are by calibrating hydrological model of gauged basins and transposed the model parameter estimates to the ungauged basin known as spatial proximity analysis (Zhang and Chiew, 2009).

Other methods for example, in the review by Razavi et al., (2013) considered model similarities based on the watershed attributes by developing a relationship between model optimization parameters and watershed attributes of the gauged basin to predict the watershed response in the ungauged watershed.

While some considered regionalisation by homogenisation which involves identifying homogeneous zones by grouping basins with similar physiographic or climatic characteristics between the gauged and ungauged basins. The effectiveness of the homogenization approach according to Sivapalan, (2003) depends on the capability of the technique in identifying and delimiting the homogeneous zones.

However, these techniques are quite challenging in hydrological science especially in data-sparse regions due to inadequate runoff data and studies on different site as posited by Oudin et al., (2008) produced contradictory results and concluded that this may be related to variation of catchment characteristics across the different cases.



Furthermore, coarse resolution of input climate data and base maps of certain morphological properties of the watershed land surface can hinder the application of regionalisation methods at local basin scale (Olden et al., 2012).

It is, therefore, necessary to investigate and incorporate multi-disciplinary strategies between the climate and integrated assessment modelling frameworks to overcome the knowledge gaps identified in modelling issues in data-sparse watersheds to advance the understanding of water-related assessments.

This will further strengthen the process through the inclusion of a mechanism for ongoing coordination and information exchange through data and information integration systems to coordinate and improve user support among researchers. Section 6.2 and 6.4.2 – 6.4.4 provides specific literature and concept to address research question and objectives 3 and 4 respectively.

## **2.3 Water resources assessment**

### **2.3.1 Water footprint concept**

Water supplies are heavily impacted by population increase, brisk economic development, and rising household, industrial, and agricultural usage (Launiainen et al., 2014). Freshwater supply and quality problems already have an effect on people's quality of life, economic development, and biodiversity loss in many watersheds around the world (Vörösmarty et al., 2010).

Effective sustainability indicators have been a source of concern in water resource management and governance as water is becoming a more limited resource. As a measure to assess how effectively water resources are being used in relation to human consumption, the term "water footprint" (WF) was introduced (Hoekstra and Hung, 2002).

WF assessment supports management of water resources by providing information on water consumption and pollution and is quickly rising to the top of the list of priorities for

water sustainability at watershed scale (Čuček et al., 2015), and is an ambitious concept for monitoring and assessing human appropriation of freshwater resources and encouraging sustainable use (Launiainen et al., 2014).

This concept has been used to address water security at watershed scale by defining and apportioning water resources into green, blue and grey water (Schneider, 2013), and adjudged as an appropriate and effective methodology for river basin management.

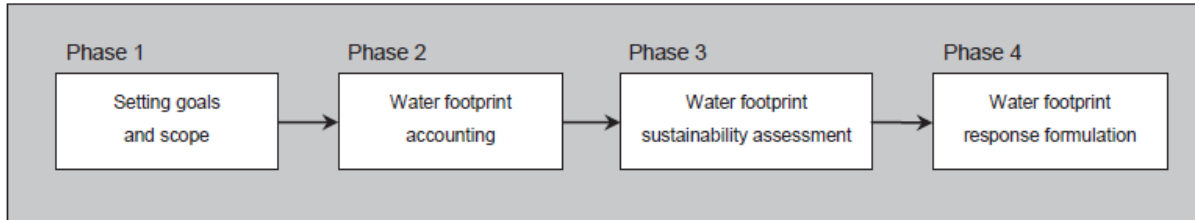
The methodology defined green water as moisture available in the soil layer for plants and microbes absorbed by the root zone layer, used up by plants, and release to the atmosphere through transpiration (Rockström et al., 2009; Rodrigues et al., 2014; Schneider, 2013; Veetil and Mishra, 2016).

Blue water is the water that flows through the surface and subsurface layer and generally stored in aquifers, and also in rivers, lakes and reservoirs for human use in various sectors such as domestic, industrial, irrigation agriculture etc. (Hoekstra et al., 2011; Quinteiro et al., 2018a; Schneider, 2013; Yuan et al., 2019).

Given the concentration and current ambient water quality regulations, the amount of freshwater resources needed to absorb the load of contaminants is referred to as "grey water" or "pollution water." (Hoekstra et al., 2011). A brief theoretical background for blue and green water footprint and availability considered within the scope of the research is given in **6.4.**

This concept has been very helpful in determining how activities and products related to water scarcity and pollution affects human health as well as what can be done to prevent unsustainable use of freshwater resources, for example in a geographic perspective, will be water footprint within a delineated area such as province, municipality, watershed or basins.

An objective evaluation of the concept will be initiated by setting a goal and scope for the study before other phases of the assessment based on the intended purpose. The phases as defined by Hoekstra et al., (2011) is given in **Figure 2.5**.



*Figure 2.5: Phases in water footprint assessment (Hoekstra et al., 2011)*

The concept has been used in water security assessments at various watershed levels to show water consumption patterns in terms of space and time, and it offers knowledge and recommendations for water governance and sustainable water policy at a global (Galli et al., 2012; Liu et al., 2017; Mekonnen and Hoekstra, 2012; Rodell et al., 2018), regional (Masud et al., 2018; Naderi and Parsa, 2022; Xie et al., 2020) and local (Aghakhani Afshar et al., 2018; Liu et al., 2023; Mao et al., 2020; Veettil and Mishra, 2018; Zhang et al., 2022; Zhu et al., 2018) watershed scale.

However, such studies at global and regional scale may not provide accurate water footprint accounting and estimates of local watershed hydrologic features in some parts of the world due to scale and resolution of data used in the assessment for effective local basin scale water policy decisions.

Watersheds where studies are conducted at local scale, were possible due in part to availability of accurate and robust long-term records of sectoral water use information across various sectors. Therefore, basins with inadequate data are not adequately studied and effort to develop alternative strategy and framework for improved water footprint assessment is a necessity for adequate river basin water policy planning in the face of projected climate change and socioeconomic changes at local watershed scale.

### **2.3.2 Watershed water resources sustainability**

The main goal of climate policy and river basin water resource management as contained in the United Nations' framework on climate change convention is to avert the negative anthropogenic interference of the system and state of the climate (Arnell et al., 2011). One of the important areas where impacts will be anticipated is the availability of freshwater resources (Alcamo et al., 2007).

Therefore, the reliability and accessibility of freshwater supplies from water reservoirs are both directly impacted by ongoing changes to our climate, which go beyond just the fields of climatology and hydrology, and studies on the assessment of its sustainability at local watershed's scale required basic understanding of sectoral dynamics of water usage for domestic, agricultural and industrial purposes ensured by allowing water to accumulate in active storage capacity in reservoirs (Sýs et al., 2021).

Additionally, changes in the watershed's hydrologic processes, such as the timing and volume of basin hydrologic drivers and events, could have an effect on other environmental factors like the flux of nutrients and sediment into water sources (Simonovic and Li, 2004).

Water availability for domestic, agricultural, and industrial use has grown in importance as a subject of international and interdisciplinary research as a result of the significant water challenges that some parts of the world, particularly those in the Arid, Semi-Arid, and Mediterranean basins, are currently experiencing.

Although there is enough water in theory to sustain practically everyone on the earth (Savenije, 2000), however, the distribution and flows of this resource spatially and temporally in combination with increasing demands for example, in developing nations in practice is faced with severe shortages or water scarcity conditions in more areas around the globe.

These problems will continue to pose a serious threat to the future sustainability of water resources across sectors which may further be exacerbated by an increase in population, anticipated future projections of climate change due to sporadic redistribution of rainfall intensity and volumes and earth surface temperature increase (Arnell, 2004).

The quantification of human appropriation of the watershed freshwater resources is a necessity and several approaches and projects have been developed as a result of the societal realization of the significance of sustainable water usage (Launiainen et al., 2014). A minimum of two approaches can be used to define sustainability, which are:

- Freshwater availability, implying that available i.e., renewable supplies should not be exceeded by sustainable usage.
- Freshwater quality, implying that possible deterioration of water quality or detrimental effects on the provision of ecosystem services.

A number of aggregate indices to gauge water sustainability have been developed since the turn of the 20th century in response to concerns about monitoring sustainable development initiatives. Additionally, there are some that are explicitly designed to assess water security, such as those put out by Assefa et al., (2019); Jensen and Wu, (2018); Shrestha et al., (2018) in urban, or (Zhou et al., 2021) in rural settings.

The sustainability indices are quite important in watershed assessment especially when availability of freshwater has been seriously harmed by irresponsible management of water resources and rising abstraction rates.

For instance, by 2014, the average amount of renewable freshwater available globally has dropped precipitously by almost 40% since the 1970s, and the issue is made worse by the world's uneven distribution of freshwater resources and their pronounced seasonality (de Castro-Pardo et al., 2022). This has resulted in conflicts especially when the management involves numerous jurisdictions or countries (Chellaney, 2016).

This additionally highlights the growing demand for a new framework and the application of water security indicators (WSI) in order to conduct an accurate monitoring and produce the data or information required to enhance water management decision-making in data-sparse watersheds.

The hydrologic modelling framework developed, and water footprint sustainability indices adopted were applied to the Yobe-Komadugu watershed located in the west of Lake Chad, a transboundary and semi-arid basin that lies between northeastern, Nigeria and southeastern, Niger.

The basin is characterized with high rainfall variability, inadequate streamflow monitoring stations and often water stressed especially in the post monsoon season as a decision support tool to derive water resource information in the face of climate change and population growth to address research questions and objectives 4, 5 and 6 See (8,9) within the context of the research design respectively.

## **2.4 Summary and conclusion**

Appropriate application of climate data in impact studies at watersheds with sparse observational data is a source of concern and has limited the advancement of water resources studies that provides information as a decision-support tool for river basin management. In an effort to address this challenges, alternative data provided by various modelling centers are used, however, they differ in spatial and temporal resolution as acknowledged in (Gampe and Ludwig, 2017; Henn et al., 2018), and their application at the local basin through parameterization (downscaling and bias correction) requires objective critical analysis.

This is quite necessary to provide basin-scale features that objectively represent the current and projected changes in climate and hydrologic dynamics in watershed modelling within the acceptable limit of observational uncertainty.

There have been efforts and frameworks to advance the requirement for an improved understanding of the data assimilation, parameterization and statistical approaches for the determination of basin-scale climatological features and the interaction between input data and hydrologic models appropriate for the assessment of the current and projected changes in water resources and hydrologic hazards especially, in trends and magnitudes spatially and temporally. However, there is still some weaknesses in some of the applied techniques and often produce contradictory outcomes as alluded in (Oudin et al., 2008; Razavi et al., 2013).

Some of the methods deployed in the literature may have led to several assumptions that can only be applied to watersheds where traditional modelling approaches are plausible, for example earlier studies using regionalization modelling approach by Faramarzi et al., (2013); Schuol et al., (2008) exhibited a poor relationship between simulation and observation hydrologic responses and uncertainty ranges based on calibration and validation data statistics in the data-sparse Lake Chad regions which might lead to unreliable watershed representation of water resources and security indices.

Therefore, where water resource information is sought for at local watersheds with sparse data, there is limited research to establish the successful application of this technique to generate a coherent methodology for the modelling of basin hydrology has been identified as a research gap.

At present, there are few studies that demonstrate the use of alternative datasets and sophisticated approaches to limit the identified uncertainty sources in modelling especially, the input variables and scenario uncertainty, due in part to the complexity of basin characteristics, data quality and non-stationarity of the modelling process (Nkiaka et al., 2022; Visessri and McIntyre, 2016).

This research aims to incorporate the existing approach with machine learning technique to improve model selection after parameterization and investigate the viability of the

framework in watershed modelling in data-sparse watersheds and further advance the course of blue water footprint assessment at basins with limited sectoral water use information to aid water sustainability analysis as a decision support tool for enhanced river basin water policy.



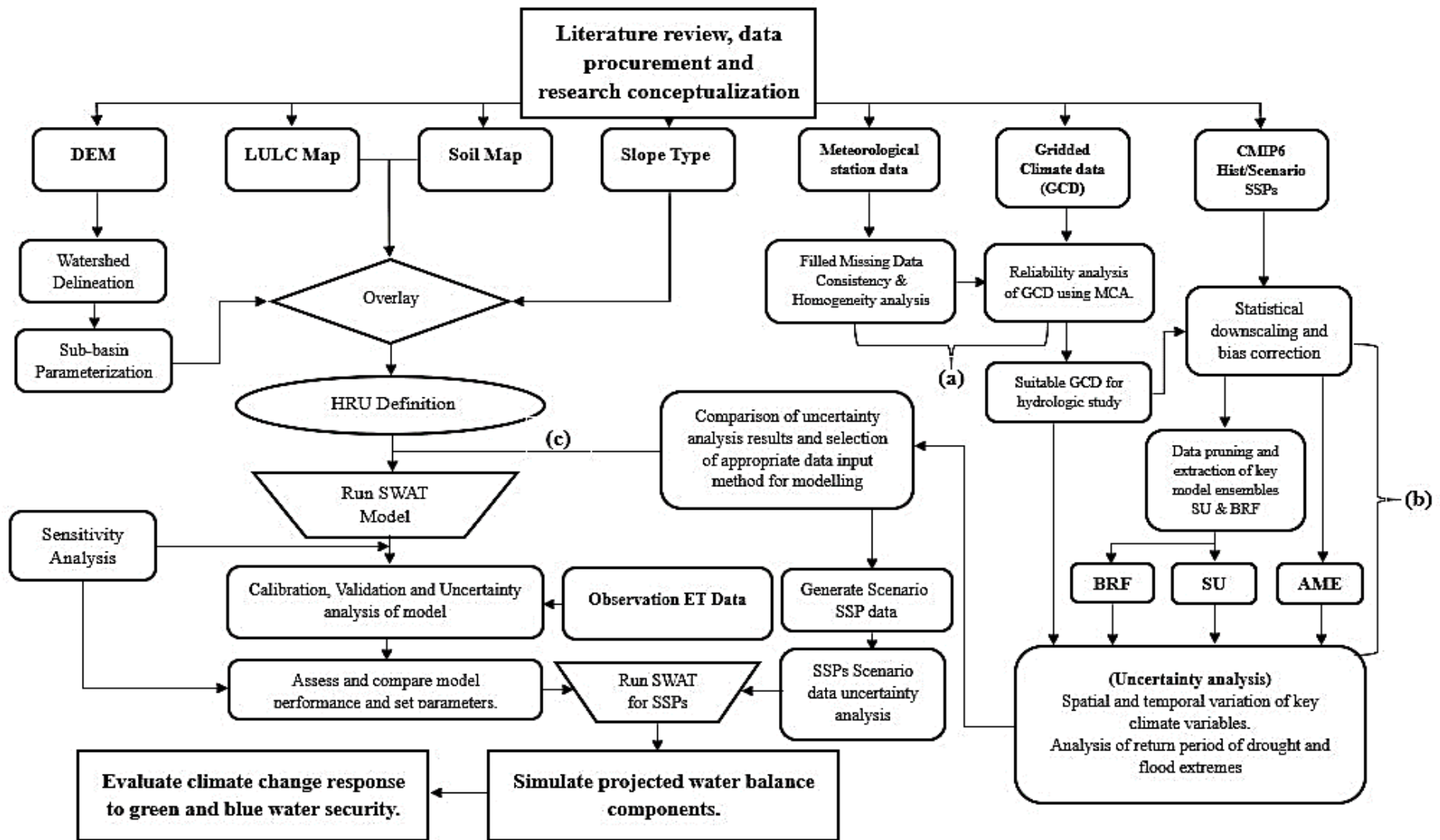
## CHAPTER 3: RESEARCH DESIGN FRAMEWORK

### 3.1 Introduction

This chapter discusses the summary of the research software tools and methods employed and how it fits the study area characteristics. The methods are outlined in a conceptual research design flowchart (**Figure 3.1**), that expressed climate modelling and simulation strategies, surface water hydrologic modelling framework related to data preparation as a solution to limit input and scenario uncertainty especially in regions where robust conventional modelling and water security assessment data are not adequate for objective and accurate model development.

However, the description of the methodology in this chapter focuses on the brief conceptual research design and discussion of methods that are not captured in the subsequent core chapters (**61, 99, and 144**), and references were made to approaches in these discussions where necessary for clarity. The research is formulated by scholarly review of relevant literatures and opportunities were identified for additional inquiries related to uncertainty sources, data type, quality and complexity, data assimilation and parameterization techniques and leverage on expert understanding of hydrologic model structure and their limitations in watershed modelling.

In this thesis, a quantitative research method is employed using numerical systems (Gay et al., 2009), to measure and analyse watershed climate (**3.2 – 3.4**) and hydrologic (**3.5 - 3.6**) changes through a variety of statistical models and establish the nature of their associations (Creswell, 2009). In this context, established theories, refined modelling approach and the use of quantitative data were explored (**3.6**) to understand and improve rainfall-runoff model development, assess, and predict watershed changes to green and blue water security (availability and sustainability) in a data-sparse region (Lawal et al., 2023). A summary of the quantitative methodology flowchart used in this thesis is presented in **Figure 3.1**.



**Figure 3.1:** Schematic summary of research methodology flowchart (a) consistency and reliability of observations and gridded climate datasets using MCA (b) assessment of efficacy of downscaling and bias corrections and data pruning approaches and validations for hydroclimatic studies (c) integrated framework of modelling using selected data pruning method and SWAT model in data sparse regions for water security assessment.

### 3.2 Reliability analysis of gridded and observation data

The practical application of data series in climate and hydrologic modelling is necessary for adequate planning, research design and operation of complex water resource systems. In this research, 36 climate stations were identified that contained data series for both precipitation and temperature or one of the variables at different spans from different sources (See **Figure 4.1**).

After careful assessment of the data length (30 years as recommended by WMO) and missing data points (< 40%) of the series, 12 precipitation (**Table 4.2**) and 15 temperature (**Table 4.1**) stations were adjudged reliable based on the guidance of Aguilera et al., (2020) for infilling the discontinuities without compromising prediction accuracy and increase confidence in climate data assessment used in impact study.

Data imputation technique was used to fill up the missing data point in the R-software domain by multivariate imputation by chained equations (MICE) i.e., a sequence of iterative predictive models used to "fill in" (impute) missing data in a dataset (van Buuren and Groothuis-Oudshoorn, 2011).

Every iteration uses the other variables in the dataset to impute each specified variable. The predictive model run these iterations until convergence has been reached (Luken et al., 2021). Variable data are modelled based on their distribution, i.e., binary variables are modelled using logistic regression and continuous variables are modelled using linear regression (Azur et al., 2011).

Generally, all missing values are replaced with the mean observed values and then changed back to the missing values for the first variable that will be imputed to account for statistical uncertainty in the imputations (De Carvalho et al., 2017; Gibson et al., 2022) (**See 250**) and this was due the effectiveness of the technique to explore and provide a continuous two-level (i.e., variable data with repeated measured values taken across time and nested

within a cluster or group) data and maintain consistency (van Buuren and Groothuis-Oudshoorn, 2011).

The technique was quite relevant for data series with gaps in different duration, climatic regions and at a different season (Gyau-Boakye and Schultz, 1994). The infilled data series at different stations were tested for consistency and homogeneity by relative (**Figure 4.2**) and absolute homogeneity (**See 250 and 251**) test results.

However, some obvious potential limitations that was not explored was the effect of the characteristic of the period of missingness (i.e., cooler, warmer, wetter, drier) of the climate data relative to average values due to naturally climate variability and this could lead to imputation biases. Although, the effect is likely to be insignificant due to the fact that the percentage of the missing data point (< 40%) was kept below the recommended threshold to effectively provided a reliable prediction as supported by Aguilera et al., (2020); Gibson et al., (2022).

Following the careful assessment of the reliability of quality-controlled precipitation and temperature data series for consistency and homogeneity (Lawal et al., 2021), the gridded datasets adopted for the study were extracted from various sources (**Table 4.3**) at the observation station resolutions by codes developed in the R-software domain (**See 251**) for the period 1979 - 2012.

A plethora of performance metrics such as machine learning-based filters i.e., symmetric uncertainty (Nashwan and Shahid, 2019), statistical performance indicators and the observed trend and magnitude (**See 4.4**) of the Spatio-temporal changes in precipitation and temperature at the local station points at annual and seasonal scale (Shiru et al., 2019a).

For all the conditions considered, the gridded dataset's derived ranking and relative performance make it possible to select the most suitable products by multi-criteria decision for a hydrologic impact study at the local basin scale.

The work detailed above is found in **CHAPTER 4**:

### **3.3 Statistical downscaling and bias correction of climate models**

Three statistical downscaling and bias correction techniques (delta change, Quantile mapping and Empirical Quantile mapping) were employed after 16 GCM data series were extracted that are essential for the accurate analysis of the historical and projected changes of watershed hydro-climatic features especially in a topographically complex terrain at the local scale.

The GCM models (**Table 5.1**) were first extracted and interpolated to a  $2^{\circ} \times 2^{\circ}$  common grid and spatial downscaled by bilinear interpolation for smooth transformation as recommended by Fischer et al., (2014) and bias corrected within the domain of statistical downscaling of general circulation models (SD-GCM v2.0) software (**See 5.4**). The theoretical summary of the downscaling methods adopted are discussed below.

#### **3.3.1 Delta change method**

The approach used in this research as applied to the GCMs, delta or change factor was derived at each grid point and then added that onto the observation measurement. The idea is to generate high-resolution and bias-corrected representation of the mean climates. The derived delta is the variation between the long term mean (usually 30 years) of climate variable between future and historical period (Navarro-Racines et al., 2020). The bias-corrected historical and future data for the grid points were calculated as follows:

- Determine the 30-year average of present-day simulations and future period.
- Determine the absolute variation between future and present-day period for temperature and proportional difference in precipitation.
- Interpolate the anomalies by the centroid of GCM grid cells.
- Add the interpolated anomalies and gridded climate data.

The technique for bias-corrected temperature (absolute difference) and precipitation (relative change) respectively was represented in equation (3.1) and (3.2) below.

$$\Delta Y_i = Y_{fi} - Y_{ci} \quad (3.1)$$

$$\Delta Y_i = \frac{Y_{fi} - Y_{ci}}{Y_{ci}} \quad (3.2)$$

Where,  $\Delta Y_i$  defined the delta change,  $Y_{ci}$  is the long-term mean (30 years) of the variable current climate, and  $Y_{fi}$  long-term mean of the variable in the projected climate of each GCM considered in the  $i^{th}$  time-step.

The projected temperature and precipitation anomalies were used to generate the future climate at each grid point based on equation (3.4) and (3.3) respectively.

$$Y_{DCi} = Y_{obsi} + \Delta Y_i \quad (3.3)$$

$$Y_{DCi} = Y_{obsi} * (1 + \Delta Y_i) \quad (3.4)$$

Where,  $Y_{obsi}$  is the current climate gridded climate observations;  $\Delta Y_i$  is the interpolated anomaly (delta change); and  $Y_{DCi}$  is the downscaled projected climate of each GCM data at the interpolated surface.

### 3.3.2 Quantile Mapping method

The quantile mapping method was applied to all the grid points with the aim to equate the cumulative distribution functions (CDFs) or probability distribution functions (PDFs)  $F_{o,c}$  and  $F_{m,c}$  as the case may be, of the respective observed  $x_{o,c}$  and modelled  $x_{m,c}$  data (Cannon et al., 2015). The transformation of the data series was represented by the transfer function in the equation below.

$$\hat{Y}_{m,p}(t) = F_{o,c}^{-1}\{F_{m,c}[x_{m,p}(t)]\} \quad (3.5)$$

The transfer functions are deemed to map the observations accurately if the CDFs of the projected bias corrected  $x_{m,p}(t)$  and observation  $x_{m,c}(t)$  data have similar distribution, and it was generally constructed from the information of the historical climate period exclusively. It was noted that the approach assumed that the postprocessing algorithms strongly relied on the fact that climate models' biases corrected are stationary.

### 3.3.3 Empirical quantile mapping method

This approach was applied across the GCM grids in the basin due to its flexibility such that distributional assumptions made in quantile mapping are non-existent here (Holthuijzen et al., 2022) and to also evaluate the effectiveness of the variations of mapping the modelled data relative to the observation. The transfer functions used in this approach are expressed based on the empirical cumulative distribution functions ( $\mu$ ) and its inverse ( $\mu^{-1}$ ). The long-term daily transfer function calculated for a common grid is given in the equation below.

$$\hat{x}_{m,p}(t) = \mu_{o,c}^{-1}\{\mu_{m,c}[x_{m,p}(t)]\} \quad (3.6)$$

Where,  $\hat{x}_{m,p}(t)$  is the corrected model value at time t,  $\mu_{o,c}^{-1}$  and  $\mu_{m,c}$  is the inverse empirical cumulative distribution function of the observation and modelled empirical cumulative distribution functions respectively, and  $x_{m,p}(t)$  is the raw model value at the daily time step. The shape of the quantile-quantile map or notable statistical performance metrics can be used to provide an understanding of the type and magnitude of the model bias (Ghimire et al., 2019).

The work detailed above is found in **CHAPTER 5**:

### **3.4 Evaluation of the effectiveness of bias correction method**

Base tests on the success and applicability of the bias correction methods were developed on a grid-by-grid basis using Climate prediction centre (CPC) precipitation and Princeton university global meteorological forcings (PGF) maximum and minimum precipitation data for the period 1979 – 2014 and 1979 – 2012 respectively.

The performance of the methods was evaluated based on statistical mean bias error, correlation coefficient and modified index of agreement to make inferences to the applicability of the tested approach in a basin with highly variable climatology (Lake Chad). As such the method that provided the best fit for the entire duration of the series were selected for further investigation.

The work detailed above is found in **CHAPTER 5**:

### **3.5 Multi-model assessment for hydrologic application**

Model reliability is directly related to the quality and resolution of climate data, and this is an area that is quite important for ecological and hydrologic study at the local watershed scale. Providing a solution to deal with GCM evaluation and selection of an appropriate set of datasets to generate a multi-model ensemble that resolved the challenges in managing simultaneously the input and scenario uncertainty in data-sparse regions spatially and temporally has become a challenge.

In this study, two random forest algorithms (**See 254**) were explored to investigate their capability of evaluating and selecting appropriate downscaled and bias-corrected GCM across the grids points to provide accurate basin historical and future climatology through data pruning (**See 5.4**) (Lawal et al., 2023b).

The evaluation and selection of GCMs from the two machine learning approaches were adopted from the methodology presented in Ahmed et al., (2019b); Raju and Kumar, (2016) by information aggregation i.e., a payoff matrix was formulated for the 16 GCMs and their



associated performance metrics (attribute importance score (IS) for BRF and similarity coefficients (SC) for SU approach), for each grid point across the basin.

The group decision making approach was followed in ranking the GCMs performance across the 54 grid points of the basin as given below:

- Each GCM was ranked 1 to 16 based on their performance and assigned a weight as the reciprocal of the rank (weight =  $1/\text{rank}$ ) at each grid point.
- The number of grid points where a GCM attained a common rank was computed as the payoff matrix to signify the frequency of occurrence.
- The frequency of the GCM ranks was multiplied by the corresponding weights at each grid point.
- The total weight of a GCM was calculated as the sum of weight across the 54 grid points for precipitation, maximum and minimum temperatures.
- The overall performance of the GCMs was determined and ranked based on spatial mean value of IS for BRF and SC for SU multiplied by the overall weights. (Note: the value of zero was assigned to a GCM at any grid point where its important score is less than the shadow attributes in BRF simulation).
- Four GCMs making up the top 75<sup>th</sup> Percentile of the ranked GCMs was adopted to generate the multimodel ensemble mean appropriate for further validation.

The multi-model ensemble generated from these techniques were further examined for the representation of spatial and temporal pattern of annual precipitation and temperature across the four climatic zones relative to observations along with the traditional approach of multi-model mean of all the GCMs referred in this research as all model ensemble (AME) for clarity.

The approaches were further used to examine the influence of the ensemble mean precipitation, maximum and minimum temperature on the severity, trend and magnitude of

extremes related to return period of drought and flood events across the four climatic zones as this can have a disproportionate effect on models spatially at the local watershed scale (Lanzante et al., 2021).

This further validation is necessary to investigate the viability of the techniques generating multi-model ensemble members capable of reducing significant distortions of climate signals, exaggerate or underestimate extreme events trends and magnitudes which is the cause additional source of uncertainty in bias-corrected GCMs (Maraun et al., 2017; Tani and Gobiet, 2019) and may amplify the effects in hydrologic models due to non-linearity of the process.

The work detailed above is found in **CHAPTER 5:**

### **3.6 Watershed modelling and water security assessment under uncertainty**

Traditional approaches in empirical watershed modelling in data-sparse regions like regionalisation (spatial proximity or homogenisation) method has proven to be difficult to apply in basins with complex morphology and insufficient runoff data (Oudin et al., 2008).

Here, A modelling framework was proposed by incorporating data pruning using BRF (Lawal et al., 2023b), to SWAT hydrologic model process to investigate the efficacy of simulating ET at four sub-basins with varied soil and land use features at a wider spatial scale aggregated with 59 observation points as recommended in Abbaspour et al., (2019).

The detailed methodology related to conceptualization, data parameterization and assimilation, basin discretization, model development with respect to HRU definition, simulation, calibration and validation, sensitivity and uncertainty analysis has been discussed in section **6.4**.

The modelling results has been compared to notable studies that relied on regionalisation methods in the basin (Faramarzi et al., 2013; Schuol et al., 2008). Although, expert

assumptions were made in the model development where conventional model data is inadequate without compromising model parsimony paradigm (Abbaspour et al., 2017).

The model outputs were further used to analyse basin scale changes in green and blue water availability and sustainability for the period 2021 – 2050 and 2051 – 2080 in response to projected climate change related shared socioeconomic pathways SSP2-4.5 and SSP5-8.5.

This two selected SSPs in this study were quite necessary and accurately depict level of vulnerability and socioeconomic dynamics of the basin and the need for efficient adaptation planning and may help to facilitate actions on the consequences for poverty and the most susceptible areas for immediate policy decisions, as well as help determine the scope and spatial patterns of predicted hazards and exposure to different water resource threats in developing economies for sustainable development (Gidden et al., 2019; van Vuuren et al., 2017).

However, inadequate or lack thereof basin wide sectoral water uses information, the analysis was restricted to Yobe-Komadugu watershed. Additionally, model geometry, projected population data, conservative per capita water use information were exploited to develop analytical blue water footprint accounting that objectively ensured sustainability assessment.

The water footprint methodology adopted here is necessary because it measures the different water needs for various sectors in a given geographic area and is crucial for estimating how much human activity will affect sustainable production (Pellicer-Martínez and Martínez-Paz, 2016), unlike the traditional method of water security assessment that depends on demand-supply approach that relied on streamflow data for modelling to examine low flows, high flows, hydrograph analysis etc. for water security assessment (Pushpalatha et al., 2012).

Other advantages of the methodology are that it is adaptable and can guide a wide range of strategic initiatives and policies from the viewpoints of the environment, society, and economy to ensure proper governance of water security.

This will assist water resource managers and policymakers in making decisions on the distribution and management of sustainable water resources on a local, national, or regional level and establish sector-specific water consumption and pollution benchmarks (Hoekstra et al., 2011; Quinteiro et al., 2018b). Finally, It increases awareness of concerns that pertain to water sustainability and the dynamics of water consumption.

The work detailed above is found in **CHAPTER 6:**

## **CHAPTER 4: MULTI-CRITERIA PERFORMANCE EVALUATION OF GRIDDED PRECIPITATION AND TEMPERATURE PRODUCTS IN DATA SPARSE REGION.**

### **Preamble**

This chapter aims to evaluate the use of systematic approaches to justify the capability of gridded climate product for reliable representation of local basin features for applications in hydrologic impact studies and in addressing this aim, the challenges of observational climate data usage at local basin studies is reviewed.

This limits the prospects of their robust application due to limited gauging stations and the data were marred with gaps (incomplete records) that tends to affect objective assessment and monitoring of the condition of watershed hydrology. However, the identified challenges can be ameliorated by available climate products in the form of model generated products or satellite observations, but objective validation is a requirement before their application.

A multi-criteria approach was applied to investigate and justified their capability with a focus on the understanding and localisation of these multiple climate dataset to replicate the spatial and temporal dynamics of the station and to a large extent the watershed climatology using multiple performance metrics.

Multi-criteria approach used in this study, offers an organised and methodical approach to assist in making complex decisions following preset standards and goals of the assessment of the climate data. When used to evaluate gridded climate data, it may yield the best information about how various performance measures will affect the choice of the dataset that most accurately replicates the temporal and spatial basin climatic feature.

The Paper following was published in Atmosphere (MDPI) titled “Multi-criteria performance evaluation of gridded precipitation and temperature products in data-sparse regions (2021) <https://doi.org/10.3390/atmos12121597>.

#### 4.1 Abstract of paper

Inadequate climate data stations often make hydrological modelling a rather challenging task in data-sparse regions. Gridded climate data can be used as an alternative; however, their accuracy in replicating the climatology of the region of interest with low levels of uncertainty is important to water resource planning. This study utilised several performance metrics and multi-criteria decision-making to assess the performance of the widely used gridded precipitation and temperature data against quality-controlled observed station records in the Lake Chad basin. The study's findings reveal that the products differ in their quality across the selected performance metrics, although they are especially promising with regard to temperature. However, there are some inherent weaknesses in replicating the observed station data. Princeton University Global Meteorological Forcing precipitation showed the worst performance, with Kling–Gupta efficiency of 0.13–0.50, a mean modified index of agreement of 0.68, and a similarity coefficient  $SU = 0.365$ , relative to other products with satisfactory performance across all stations. There were varying degrees of mismatch in unidirectional precipitation and temperature trends, although they were satisfactory in replicating the hydro-climatic information with an acceptable level of uncertainty. Assessment based on multi-criteria decision-making revealed that the Climate Research Unit, Global Precipitation Climatology Centre, and Climate Prediction Centre precipitation data and the Climate Research Unit and Princeton University Global Meteorological Forcing temperature data exhibit acceptable performance in terms of similarity, and are recommended for application in hydrological impact studies—especially in the quantification of projected climate hazards and vulnerabilities for improved water policy decision making in the Lake Chad basin.

**Keywords:** gridded climate data, performance metrics, regional modelling, climate, Lake Chad Basin.

## 4.2 Introduction

Accurate climate data are critical to the success of modelling processes in order to reduce uncertainty and achieve accurate prediction in hydrological impact studies. Unfortunately, reliable and long-term observed meteorological datasets are sparse and unavailable in some regions especially sub-Saharan Africa and the Mediterranean making hydrological studies a challenging task (Flato et al., 2013; Hassan et al., 2020). Alternatively, high-resolution gridded data have been developed to address these shortcomings. However, an understanding of their limitations in terms of observational uncertainties and reliability is important to address the twin issues of choice of dataset and suitability.

Some climate data products are more appropriate than others in their applications for climate change impact studies across different regions; therefore, careful and adequate assessment of their strengths and limitations is required in order to provide guidance for future climate and hydrological studies, especially in data-sparse basins. An accurate hydro-climatic impact study requires climate data at high temporal and spatial resolutions.

The most accurate measurement devices are rain gauges, and although these are often situated on land and in populated areas for ease of measurement (Ouallouche et al., 2018), there are a limited number of ground-based rain gauge stations in most parts of the world for effective and efficient hydro-climatic studies with reduced uncertainty in spatial climate prediction.

However, weather station records are typically site-specific, while most hydrological studies in environmental sciences research require areal observations of climate data in order to achieve accurate modelling processes with minimal uncertainty in impact studies (Berndt and Haberlandt, 2018).

Climate data have been seen to be an important component of hydrologic cycle analysis over time and space. The knowledge and understanding of their spatiotemporal dynamics are

essential, and provide useful information for their practical applications in the field of agriculture, aquaculture, water resource and river basin management, and hazard and flood disaster warnings and management (Beck et al., 2019; Sehad et al., 2017).

Climate and hydrologic studies require complete and reliable rainfall and temperature records with good spatial and temporal resolutions (Aieb et al., 2019). Unfortunately, climate records from various databases contain gaps or missing data points due to systematic errors which are prevalent in the Mediterranean and Sub-Saharan African countries and makes hydrologic studies difficult (Aieb et al., 2019; Gyau-Boakye and Schultz, 1994; Shiru et al., 2019c).

Several gridded climate data developed by various modelling centres are used as an alternative owing to their reliability and generated from the observed climate station data after quality control, enhanced reliability analysis and long-term temporal and spatial coverage. (Shiru et al., 2019b).

In a hydro-climatic study, the choice of gridded data for the process of bias correction of general circulation models in the data-sparse region indicates an essential procedure for climate change impact assessment studies (Gampe et al., 2019).

However, the choice of reference dataset that is available in either station data or gridded products derived from observations (Isotta et al., 2014; Schamm et al., 2014), reanalysis data (Bosilovich et al., 2016; Landelius et al., 2016; Poli et al., 2016), or from remote sensing data (Ashouri et al., 2015; Kummerow et al., 1998), is critical in the overall uncertainty associated in projected climate change impact studies.

These datasets form the primary input in hydrologic modelling studies and climate change for the accurate assessment of hydrologic variables such as streamflow, runoff, soil moisture, evapotranspiration etc., in order to manage hydropower operations, irrigation scheduling and



early warning systems for landslides and changes in future water availability due to climate change and basin hydrologic cycle assessment (Tapiador et al., 2012).

A study conducted over Africa has shown varying degrees of spatial mismatch between observed weather stations and reanalysis data (Zhang et al., 2013). The techniques and efforts in the analysis vary based on temporal coverage, climate variables involved and region of interest (Salih et al., 2018).

However, gridded and reanalysis data are being updated due to advances in the improved understanding of the knowledge climate science over time, and detailed evaluation of their performance at catchment scale in Africa is rarely found in literature, although this may be attributed to limited availability of reliable long term climate records, expertise and ease of access to data (Brunet and Jones, 2011; Washington et al., 2006).

Studies have shown that high resolution gridded data have been developed to provide valuable information on disaster management, initialization and validation of numerical models and resolving the diurnal global cycle of precipitation (Hijmans et al., 2005; Joyce et al., 2004).

There are many gridded data products available at different timescale (hourly, daily and monthly) with a finer resolution of  $0.5^\circ \times 0.5^\circ$ , which can provide insight that relates but not limited to a model forecast of hydrologic cycles, climate change such as trend analysis, climate downscaling etc. (Beharry et al., 2014; Chen et al., 2002; Faiz et al., 2018; Feng et al., 2011; Marengo et al., 2008; Nashwan et al., 2019; Pour et al., 2018; Schamm et al., 2014; Shirvani and Landman, 2016; Xie et al., 2007).

The downside in the use of gridded and reanalysis climate data lies in the fact that gridded climate data are a combination of observed station data and quality controlled statistical interpolation that could result in the attenuation of local climate signals, while reanalysis data is model-based forecasts that requires parameterization of the model, good assimilation

technique and the quality of the observation (Salih et al., 2018; Schoof and Pryor, 2003; Szczypta et al., 2011).

Gridded climate products are known to differ in their source, spatial and temporal resolutions, domain size (Global Coverage), available timescales and also exhibit different error bands due to interpolation procedures and considerable differences in general climatology, which are well known and acknowledged, (Gampe and Ludwig, 2017; Henn et al., 2018; Isotta et al., 2014; Palazzi et al., 2013).

The choice and selection of reference datasets at the catchment scale should be based on observational uncertainty and purpose through critical analysis of the gridded data products. The development and application of gridded climate data is growing rapidly, especially due to the advances and knowledge of their spatiotemporal resolution, latency and reliability.

However, the uncertainty associated with their application across local and regional catchments is still a cause for concern that led to some studies related to the ability of the gridded data to replicate or mimic reliable but sparse ground-based data across the globe (Eum et al., 2014; Kyselý and Plavcová, 2010; Manatsa et al., 2008; Nashwan and Shahid, 2019; Prakash et al., 2015b, 2015a; Prein and Gobiet, 2017; Sylla et al., 2013).

Furthermore, the performance of the gridded data is predicated on using individual or a combination of statistical metrics to replicate some particular characteristics of the observed data and often times exhibit contradictory results and makes the decision making difficult (Ahmed et al., 2019; Beck et al., 2019; Salman et al., 2019).

It has been posited that some gridded data has shown to be appropriate compared to others in specific applications in certain regions around the world (Tanarhte et al., 2012), and a single statistical metric cannot justify the performance or suitability of a particular gridded data. Therefore, it is important to use various metrics to obtain an ideal solution, based on

optimal performance across all the metrics especially in the data-sparse region for prediction efficiency (Salman et al., 2019; Xu et al., 2015).

The objective of this study was to employ multi-criteria decision making to assess the performance of five widely used and recently updated gridded precipitation datasets and four temperature datasets in replicating the total and average monthly precipitation and temperature of available gauge-based records in the Lake Chad hydrological basin.

This study was necessary to provide guidance on the choice of reference dataset(s) to be adopted for future research in the basin, depending on performance and purpose in hydro-climatic studies, in order to reduce uncertainty in predictions as well as computational time and resource costs.

Furthermore, the choice of the gridded dataset in previous climate studies found in literature in the basin for example (Adeyeri et al., 2017; Mahmood and Jia, 2019a; Nkiaka et al., 2018b; Pattnayak et al., 2019), has been based on their successful applications in other basins, without a proper justification of their suitability compared to other available products for improving the reliability of predictions and reducing model biases to provide an accurate representation of basin-scale hydrological features.

Additionally, some of the products are only available in monthly time steps, which may not be suitable for downscaling of GCMs with daily time steps as input requirements in some hydrological models and climate change impact studies.

The study will employ entropy-based symmetric uncertainty (William et al., 1996), a machine learning approach that has been found to be an efficient tool for the assessment of agreement in data that measure the shapes and patterns of data sequences via the concept of mutual information theory, by comparing the similarity between two long time-series climate datasets, and has found its application in various fields (Saleem A. Salman et al., 2018; Shiru et al., 2019b).

The benefit of this method is that it does not depend on the data distribution, unlike the statistical metrics used in other studies (Nashwan and Shahid, 2019; Pour et al., 2018; Saleem A. Salman et al., 2018). Four statistical metrics were used in this study—namely, Taylor diagrams, modified index of agreement (md), Kling–Gupta efficiency (KGE), and normalised root-mean-square error (NRMSE)—and then finally trend analysis of the precipitation and temperature data at the annual and seasonal scales of the gridded and observed station records was compared for mean variability and temporal homogeneity across the basin.

### **4.3 Study Area and data**

#### **4.3.1 Study area**

The Lake Chad Basin is one of the largest endorheic basins in the world and occupies an estimated area of  $\sim 2,500,000 \text{ km}^2$ , approximately 8% of Africa (Coe and Foley, 2001; Gao et al., 2011). The basin cuts across the whole or part of Algeria, Cameroon, Central Africa Republic, Chad, Libya, Niger, Nigeria and Sudan in Central Africa. The basin is geographically located at latitudes of  $5.2^\circ - 25.3^\circ \text{ N}$  and longitudes  $6.9^\circ - 24.5^\circ \text{ E}$ , right at the transition zone of the Sahara region and the tropics of Sudano-Sahelian region of West Africa (Ndehedehe et al., 2018), (**Figure 4.1**).

The basin is characterised to have a vast and shallow freshwater lake located at the centre, with inflows from Chari-Logone rivers ( $\sim 90 - 95\%$ ) from the southern pool, Yobe-Komadugu rivers ( $\sim 2.5 - 5\%$ ) in the western region and enters the lake through the northern pool (Coe and Birkett, 2004; Magrin, 2016; Sarch and Birkett, 2000), and other minor rivers such as Gubio, Ngadda, Yesderam and Elbeid, which supplied only ( $\sim 1 - 2\%$ ) inflow to the lake through the southwestern part of the basin between 1961 - 2013 (Mahmood and Jia, 2019b). The basin serves as the main source of fresh water that supports livelihood across pastoral land, agricultural land and fish farming (Buma et al., 2016), with irrigation

agriculture as the major user of the resource that supports a majority (~60%) of the population (UNEP, 2006).

The basin is divided into several climatic zones, namely the Saharan zone located in the north of the basin, the Sahelo-Saharan zone located in the central part of the basin, which covers the North of Diffa, Niger and Lake Chad; the Sudano-Sahelian zone, which covers Ndjamena in the Chad Republic and the northern part of Cameroon and Nigeria and the Sudano-Guinean zone located in the south, which covers the south of Chad and the Central African Republic, with annual precipitation of, < 100 mm, between 100 - 400 mm, 400 - 600 mm and 600 - 1500 mm, respectively.

The average annual temperature in the basin ranges from 35°C to 40°C in the northern part of the basin to as low as 26.5°C in the southern part (Nkiaka et al., 2018b) characterised by hot and dry, wet and dry and cool weather during March to June, June to October and November to February respectively (Mahmood et al., 2019).

The basin is located in a region that is characterised to have little relief, no surface outlet and a spatial extent that is quite sensitive to climatic variability and the elevation of the basin ranges from – 330 m to 3446 m (**Figure 4.1**). However, according to (Coz et al., 2009; Nkiaka et al., 2018a), the basin is a relatively flat area with an average slope of < 1.3%, except for some local hills, plateau and mountains in southern and northern parts of the basin.

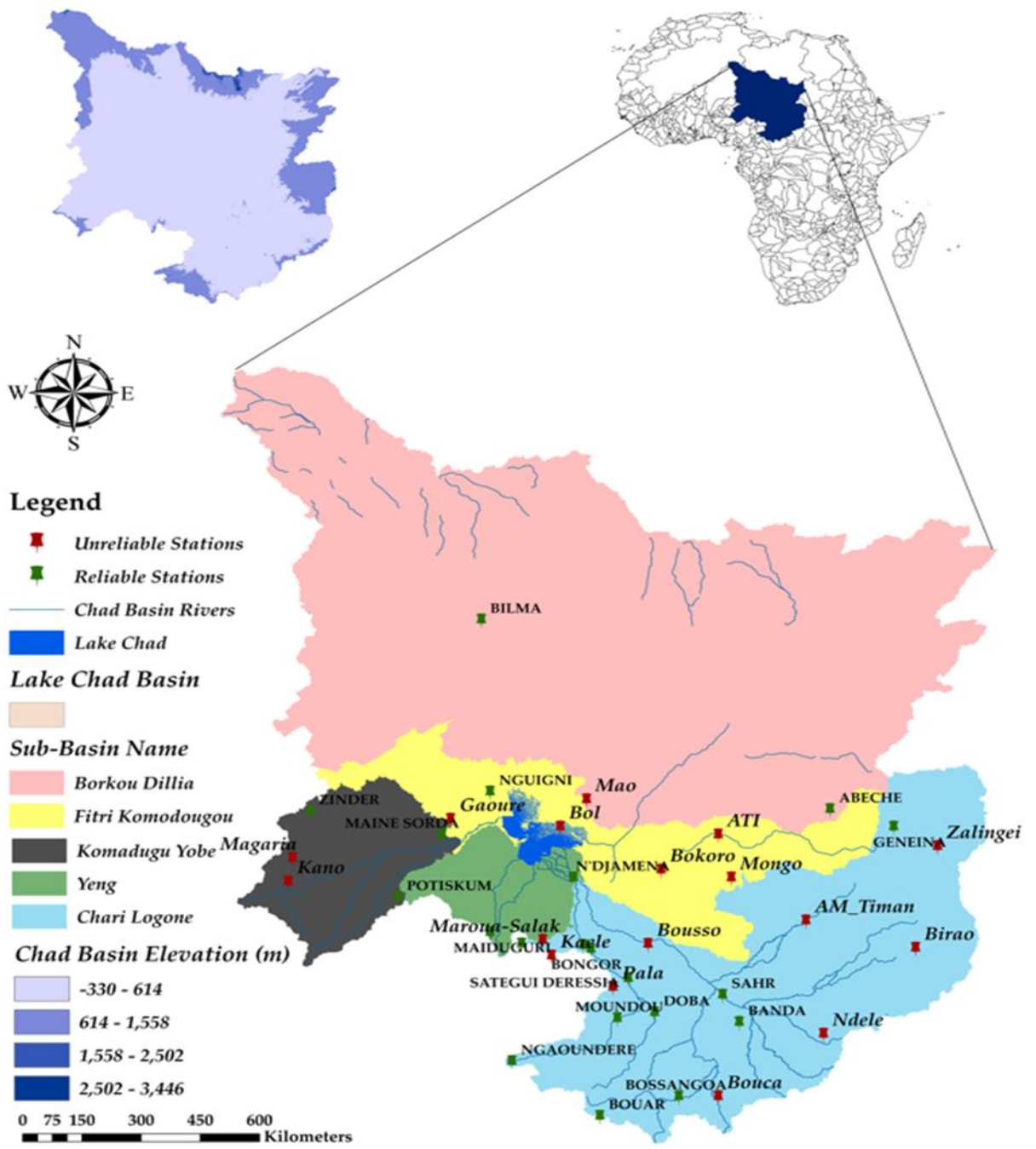


Figure 4.1: Map of Lake Chad Basin showing Elevation, Lake, Climate Stations, major river networks and Sub-basins.

### 4.3.2 Observation data and sources

The observed climate data used in this study were acquired from a number of sources; for example, 12 average monthly observed temperature data were obtained from the global historical climatology network monthly temperature dataset version 4 (**Table 4.1**) <http://www.ncdc.noaa.gov/ghcnm/v4.php> accessed on 2 March 2020 (Menne et al., 2018), while 11 total monthly observed precipitation data were obtained from Lake Chad Basin Commission and can be found as supplementary dataset online in Mahmood and Jia, (2019a). We also used 2 Station records from Nigerian Meteorological Agency (NIMET) and the NOAA Global Historical Climatology Network Daily (GHCN-D) version 3.23 (<http://www.ncdc.noaa.gov/pub/data/ghcn/daily>) (**Table 4.2**).

The observed station data considered in this study were carefully selected based on the condition of having fewer missing records and an acceptable temporal span for hydro-climatic analysis in order to achieve effective and reliable predictions. The observed missing climate data records were filled using multivariate imputation by chained equations (MICE) package (**See 250**), due to its ability to impute continuous two-level data and maintain consistency between imputations while employing passive imputation (van Buuren and Groothuis-Oudshoorn, 2011).

The data was checked for 100% completeness after imputation and assessed for comparison using a double mass curve approach for subjective evaluation of non-homogeneity in the datasets (Kohler, 1949), and finally subjected to absolute homogeneity tests namely standard normal homogeneity test (SNHT), Pettitt and Von Neumann ratio test (**See 250 and 251**) (Wijngaard et al., 2003; Yozgatligil and Yazici, 2016).

The double mass curve showed an almost straight line at all stations without breakpoints (**Figure 4.2**), and absolute homogeneity test results were all less than the critical values. The null hypothesis that all the station climate data tested were homogenous at a 95% level of

confidence cannot be rejected; and therefore, the quality-controlled data for precipitation and temperature are suitable for performance evaluation of the gridded climate data.

**Table 4.1:** List of reliable Observed temperature stations, location and temporal span in Chad Basin

S/No	Station Name	Data Range	Missing Data (%)
1	Bilma	1950–2019	8.2
2	Bossangoa	1954–2016	35.0
3	Bouar	1951–2019	34.7
4	Geneina	1951–2019	11.5
5	Maiduguri	1910–2012	14.0
6	Maina sorda	1951–2019	11.5
7	Moundou	1951–2016	34.3
8	N'Djamena	1951–2019	26.3
9	Ngaoundere	1951–2019	35.2
10	Nguigni	1953–2019	5.8
11	Sahr	1941–2018	39.3
12	Zinder	1923–2019	3.5

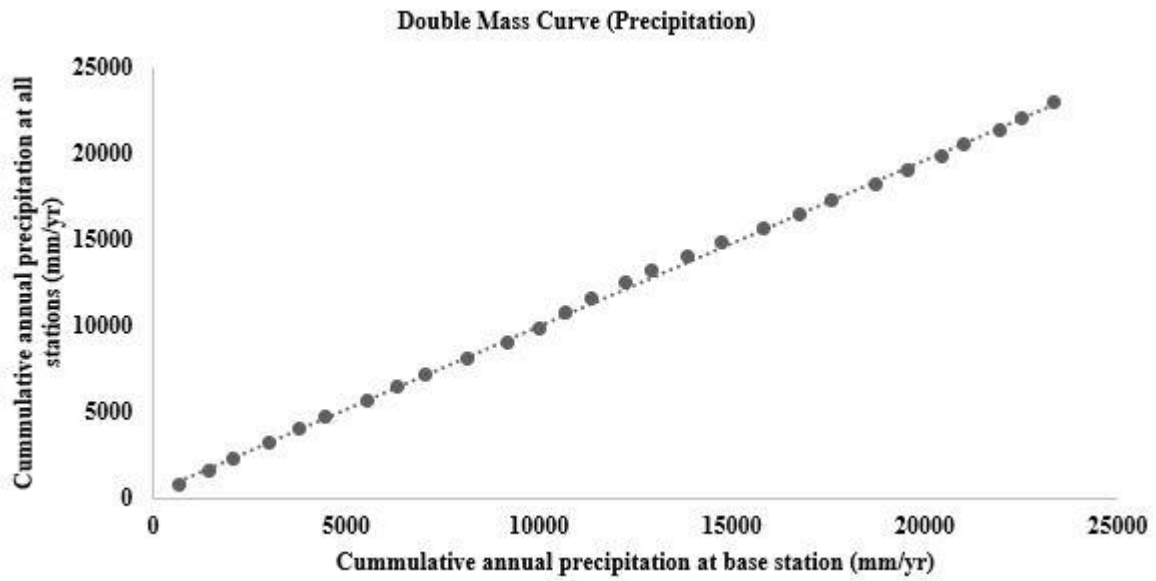
*Source:* (Menne et al., 2018) <http://www.ncdc.noaa.gov/ghcnm/v4.php> .

**Table 4.2:** List of Reliable Observed Precipitation station, location and temporal span in Chad Basin

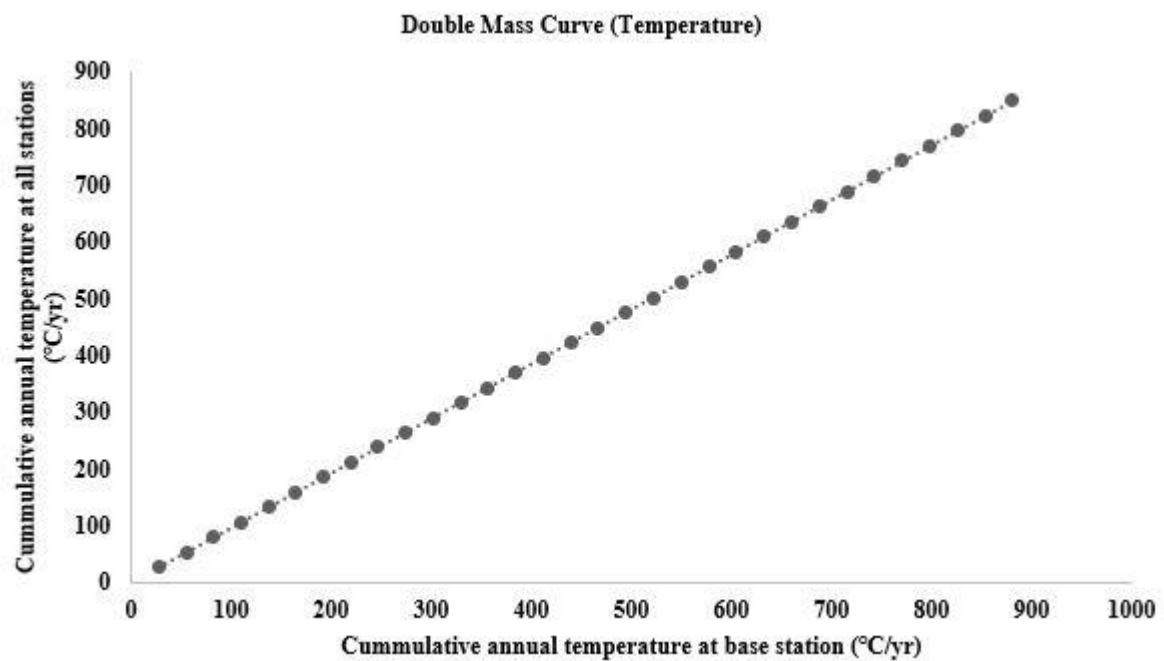
S/No	Station Name	Data Range	Missing Data (%)
1	Abeche	1985–2015	0.0
2	Banda	1950–2013	0.0
3	Bongor	1950–2013	0.0
4	Bossangoa	1950–2013	0.0
5	Doba	1950–2013	0.0
6	Maiduguri	1979–2010	0.0
7	Moundou	1985–2015	0.0
8	N'Djamena	1985–2013	0.0
9	Nguigni	1968–2020	15.3
10	Potiskum	1980–2010	0.0
11	Sahr	1950–2013	0.0
12	Samry-I	1950–2013	0.0
13	Sategui Deressia	1950–2013	0.0
14	Tsanaga	1950–2013	0.0
15	Zinder	1906–2020	17.1

*Source:* ((LCBC) (Mahmood and Jia, 2019a), NIMET, GHCN-D (<http://www.ncdc.noaa.gov/pub/data/ghcn/daily>))





(a)



(b)

**Figure 4.2:** Double-mass curves for Lake Chad basin. (a): Cumulative annual precipitation at all stations against base station. (b): Cumulative annual temperature at all stations against base station.

### 4.3.3 Gridded data and sources

This study analysed five gridded precipitation and four temperature products, namely the University of East Anglia, Climate Research Unit CRU TS V4.04, German weather service Global Precipitation Climatology Centre, GPCC v.2018 (precipitation only), US Climate

Prediction Centre CPC, Princeton University Global Meteorological Forcing PGF v.2 and University of Delaware UDel V5.01. **Table 4.3** summarises the climatic variables, temporal and spatial resolution, temporal span and sources of the gridded data.

*Table 4.3: Summary of gridded dataset considered in this study.*

<b>Data Product</b>	<b>Variable</b>	<b>Temporal Resolution</b>	<b>Data Span</b>	<b>Spatial Resolution</b>	<b>Source</b>
CPC	P, Tmax, Tmin	Daily	1979 - 2020	0.5°	Precipitation <a href="https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globalprecip.html">https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globalprecip.html</a> ) and temperature <a href="https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html">https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html</a>
CRU TS v.4.04	P, Tmax, Tmin	Monthly	1901 - 2016	0.5°	<a href="https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.04">https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.04</a>
PGF v.2	P, Tmax, Tmin	Daily	1901 - 2012	0.5°	<a href="http://hydrology.princeton.edu/data/pgf.php">http://hydrology.princeton.edu/data/pgf.php</a>
GPCC v.2018	P	Monthly	1901 - 2016	0.5°	<a href="http://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html">http://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html</a>
UDeI V5.01	P T <sub>ave</sub>	Monthly	1900 - 2017	0.5°	<a href="https://psl.noaa.gov/data/gridded/data.UDeI_AirT_Precip.html">https://psl.noaa.gov/data/gridded/data.UDeI_AirT_Precip.html</a>

The gridded datasets were developed from different modelling centres using different interpolation techniques, for example CPC gridded data was developed by optimal interpolation of station or gauged based records of GTS (Xie et al., 2007), CRU data were developed via angular distance weighing of monthly observed station data from World Meteorological Organisation, National Oceanic and Atmospheric Administration (NOAA) database, and Climate Records from National Meteorological Agencies across the globe (New et al., 2000), PGF dataset was developed based on forcings from NCEP-NCAR reanalysis and other global data via bilinear interpolation from their native gridded scale (Sheffield et al., 2006), GPCC datasets were developed by the combination of monthly gauged and quality controlled records from 7000 stations around the world, along with GTS synoptic weather reports, interpolated to regular grid using ordinary point kriging method (Schamm et al., 2014) and UDeI precipitation and temperature data are interpolated using shepherd algorithms to grid based data from various sources such as GHCN2, NCAR, GHCN-Daily dataset, GHCN-monthly version 3, and records from national meteorological agencies (Lawrimore et al., 2011; Menne et al., 2012).

## 4.4 Research Methodology

The performance of the gauged based gridded precipitation and temperature data was evaluated by multiple approaches including, machine learning filter-based symmetric uncertainty, statistical metrics (Index of agreement, Kling Gupta Efficiency and Normalized Root Mean Square Error and Taylors Diagram) and time series analysis of trends exhibited at an annual and seasonal timescale in order to assess the gridded data in terms of their reliability in mimicking the observed data across all the stations in the study area for the period 1979-2012. The gridded datasets were sourced from the websites of the providers as shown in **Table 4.3**, extracted by Raster and ncdf4 R packages (**See 251**) and interpolated to the observed data station resolution using inverse distance weighting method. The detailed methodology of the study is outlined below.

### 4.4.1 Symmetric Uncertainty

Symmetric Uncertainty is an entropy-based machine learning algorithm (Filter Method) used in assessing the pair-wise agreement between long time-series data. The method utilises information entropy through the concept of mutual information (MI), which measures the commonality between two variables. For example, if  $p(x)$  and  $p(y)$  are considered probability density functions of observed variable ( $x$ ) and the gridded variable ( $y$ ), and  $p(x, y)$  is the mutual probability distribution functions of  $x$  and  $y$ , and therefore, mutual information can be evaluated as follows:

$$MI(x, y) = p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)} \quad (4.1)$$

The mutual information shown in Equation (4.1) can also be evaluated as the difference between the mutual entropy of two-time series variables, and in this case, taking the observed data as  $H(x)$ , the gridded data as  $H(y)$  and the mutual entropy of the observed and gridded data time series as  $H(y, x)$ , MI can be written as:

$$MI(x, y) = H(y) - H(y, x) \quad (4.2)$$

Thus,  $H(y)$  and  $H(y, x)$  indicates the amount of uncertainty inherent in the gridded and the joint gridded and observed probability density functions of precipitation and temperature time series data. The two independent variables in Equation (4.2) can be expressed as:

$$H(y, x) = \sum_{i=1}^n p(x, y) \log \frac{p(x, y)}{p(x) \times p(y)} \quad (4.3)$$

$$H(y) = - \int p(y) \log(p(y)) dx \quad (4.4)$$

The entropy estimated in Equation (4.3), implies the extent of mutual information between the gridded and observed precipitation/temperature data. The mutual information tends to be zero in the absence of common information and has a value of unity when the model data series can depict the complete information associated with the observed data series. However, biases are inherent when using time series with larger values if there are less similar values between the two variables (Yu and Liu, 2003); This drawback has been addressed through the concept of Symmetric Uncertainty by dividing the value of mutual information gain and the sum of entropies of  $y$  and  $x$  as shown in Equation (4.5):

$$SU(x, y) = 2 \times \frac{MI(x, y)}{H(x) + H(y)} \quad (4.5)$$

The value of Symmetric Uncertainty ranges from 0 to 1, where 0 indicates poor similarity and 1 indicate high similarity between the gridded and observed precipitation/temperature time-series data (Shreem et al., 2016). This study utilized the Fselector package of R software (Romanski and Lars, 2018), to assess the similarity between the monthly gridded and

observed precipitation/temperature data (**Table 4.4**). **Figure 4.3** shows the distribution of gridded precipitation and temperature data of all available stations within the study area.

#### 4.4.2 Statistical Metrics

In this study, four statistical metrics were used to evaluate the ability of the selected monthly gridded precipitation and temperature datasets to replicate the observed station time series. Further details of the metrics are outlined below.

##### 4.4.2.1 Kling Gupta Efficiency (KGE)

This is a metric that highlight three components namely correlation, bias and ratio of variances between gridded and observed time series data as proposed by Gupta et al., (2009); the values of KGE varies between 0 and 1 which indicates no agreement and perfect agreement between gridded ( $x_g$ ) and observed ( $x_{obs}$ ) data respectively. The coefficient can be computed as given in Equation (4.6):

$$KGE = 1 - \sqrt{(\gamma - 1)^2 + \left(1 - \frac{\mu_g}{\mu_{obs}}\right)^2 + \left(\frac{\sigma_g/\mu_g}{\sigma_{obs}/\mu_{obs}}\right)^2} \quad (4.6)$$

##### 4.4.2.2 Modified Index of Agreement

This is a statistical metric that evaluates the standardized measure of the degree of model prediction error. This metric was modified from its original form proposed in Willmott, (1981), which has shown to be less sensitive to extreme values. it is defined as the ratio of mean square error and potential error and can detect additive and proportional differences in two long time series data. It has a value that varies between 0 and 1 indicating no agreement and perfect agreement between gridded ( $x_g$ ) and observed ( $x_{obs}$ ) data respectively. The coefficient can be computed as given in Equation (4.7):

$$md = 1 - \frac{\sum_{i=1}^n (x_{obs,i} - x_{g,i})^i}{\sum_{i=1}^n (|x_{g,i} - \bar{x}_{obs}| + |x_{obs,i} - \bar{x}_{obs}|)^i} \quad (4.7)$$

#### 4.4.2.3 Normalized Root Mean Square Error

This is a statistical metric that facilitates and summarises the magnitude of errors between a model ( $x_g$ ) and observed ( $x_{obs}$ ) data with different scales and is defined by the ratio of root mean square error and standard deviation of the data. This metric is considered to be a great measure of precision, and the predictive ability of the model is considered to be accurate with values closer to zero (Chen and Liu, 2012; Willmott, 1982). The value of NRMSE ranges between  $-\infty$  and 1. The optimal value is 1 and can be computed by equation (4.8):

$$NRMSE = \frac{\left[ \frac{1}{n} \sum_{i=1}^n (x_{g,i} - x_{obs,i})^2 \right]^{1/2}}{\frac{1}{n} \sum_{i=1}^n x_{g,i}} \quad (4.8)$$

#### 4.4.2.4 Taylor Diagram

These are a graphical representation that summarise proximity or similarity between model and observed long time series data. The similarity is quantified based on their correlation, centred root means square difference, and the amplitude of variations (Standard deviations). The models are quite useful in gauging the relative performance of models as compared to the observed data (Taylor, 2005, 2001). In this study, the Taylor diagram was used to examine the relative capability of gridded precipitation and temperature data at annual, pre-monsoon and monsoon seasons relative to the station observation data for the period 1979 – 2012.

#### 4.4.3 Trend Analysis

In this study, trend analysis was carried out using the nonparametric Mann-Kendall and modified Mann-Kendall test (Henry, 1945; Kendall, 1948), where significant autocorrelation was observed in the time series data, trend free pre-whitening was applied to correct the anomaly (Yue et al., 2002), and Sen's slope estimator was used calculate the statistically significantly increasing or decreasing trends and magnitude of the trends respectively, at 5%

level of significance for the period 1979 – 2012 (See 253). The Mann-Kendall statistics are given as:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (4.9)$$

Where  $x_j$  and  $x_k$  are sequential data values for n number of time series. The sgn of the series is defined as:

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{if } x_j > x_k \\ 0 & \text{if } x_j = x_k \\ -1 & \text{if } x_j < x_k \end{cases} \quad (4.10)$$

The mean  $E(S)$ , Variance  $V(S)$  and the Z statistics can be computed as:

$$E(S) = 0 \quad (4.11)$$

$$V(S) = \frac{1}{18} \left\{ n(n-1)(2n+5) - \sum_{i=1}^p t_i(t_i-1)(2t_i+5) \right\} \quad (4.12)$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{for } S < 0 \end{cases} \quad (4.13)$$

In the equation above, p represents the number of tied groups in the series and each of the tied groups was indicated by  $t_i$  . all positive and negative values of the Z statistics represent statistically increasing and decreasing trends in the time series data.

The magnitude of the detected trends in the time series data was computed by the nonparametric Sen's slope estimator because the method is robust against outliers in time series analysis and given below:

$$SS = \text{median} \left[ \frac{x_j - x_i}{j - i} \right] \text{ for all } i < j \quad (4.14)$$

Where  $x_i$  represent the value of the data at a time step  $i$  and  $x_j$  for time step  $j$ .

## 4.5 Results

### 4.5.1 Assessment of gridded data using symmetric uncertainty

The performance of the gridded monthly total precipitation and average temperature data for the period 1979 – 2012 was individually assessed at each station location downloaded at the station resolutions. The symmetric uncertainty (SU) score obtained at the Zinder stations located at Latitude 13.8° N, Longitude 8.9° E is given in **Table 4.4**. The result from the table showed that CRU and PGF were found to have the highest estimated similarity score of precipitation and temperature across the station.

However, there are inconsistencies in the skill of the gridded data at different stations in replicating the observed data. The CRU, GPCP and PGF dataset showed an effective skill in 53.3%, 33.3% and 13.3% and CRU, PGF and UDel also showed an improved skill at 41.7%, 33.3 and 25.0% of the precipitation and temperature stations in the study area respectively.

Observed precipitation stations located in the Sudano-Guinean zone of the basin, for example, Bossangoa, Samry-I, Sategui, Tsanaga, Doha and Moundou ( $SU \geq 0.478$ ) has shown to have an improved similarity coefficient compared to stations located in the Sahelo-Saharan zone i.e., Nguigni, Zinder ( $SU \leq 0.395$ ) etc. However, the temperature stations located in the Saharan and Sahelo-Saharan zone (Bilma, Maina-Sorda, Nguigni and Zinder) has shown an efficient skill in simulating the observed temperature ( $SU \geq 0.692$ ) and worst skill in the Sudano-Guinean zone ( $SU \leq 0.483$ ) in the study area respectively. The

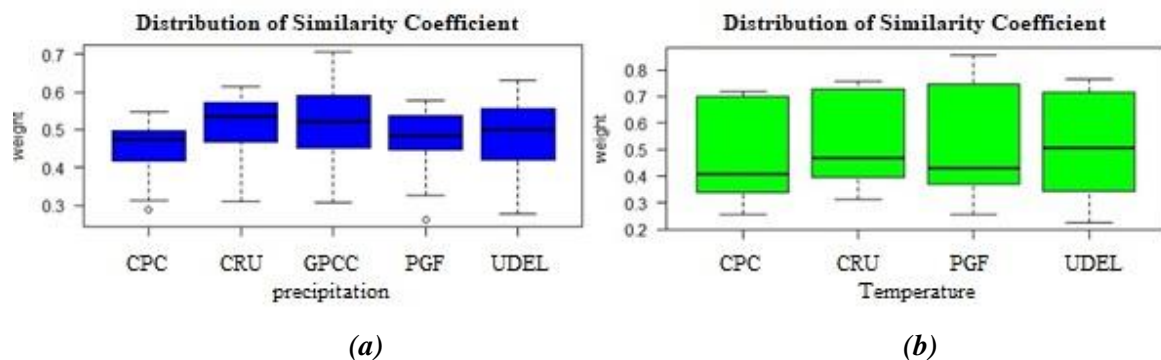


consistency in the precipitation data in the Sudano-Guinean zone may be attributed to more and accurate station records in the southern zone of the basin and reduction and abandonment of ground-based gauged data and errors in taking records resulting in systematic errors in the northern part of the basin due to migration, political instability etc.

The mean similarity coefficient of the gridded precipitation across the stations in the basin for the CPC, CRU, GPCC, PGF and UDeI dataset was 0.448, 0.504, 0.512, 0.472 and 0.480 and temperature for CPC, CRU, PGF and UDeI was 0.482, 0.532, 0.523 and 0.519 respectively. The results showed that there is a slight variation in the skill across the gridded dataset assessed in this study. **Figure 4.3** showed the distribution of the similarity coefficient of the dataset across the stations considered.

**Table 4.4:** Similarity score of gridded precipitation and temperature against observed datasets estimated by Symmetric Uncertainty

Rank	Precipitation Dataset	SU	Temperature Dataset	SU
1st	CRU	0.395	PGF	0.765
2nd	GPCC	0.387	CRU	0.725
3rd	UDeI	0.369	UDeI	0.722
4th	CPC	0.367	CPC	0.709
5th	PGF	0.365		



**Figure 4.3:** Boxplot of distribution of the similarity coefficients across the Lake Chad basin (a): variation of similarity coefficient of gridded precipitation against observed station data. (b) variation of similarity coefficient of gridded temperature against observed station data.

#### 4.5.2 Statistical metric efficiency

The results of the statistical metrics of gridded precipitation and temperature data for the period 1979 – 2012 considered in this study, showed that there is a good agreement between the gridded precipitation dataset and observed data across the station within the basin with

mean Kling Gupta Efficiency (KGE) and Index of Agreement (md) coefficient in the range of ~ 0.7 and above (**Table 4.5**) for CRU, GPCC, CPC and UDel except for PGF data that performed the least across the stations with a mean Kling Gupta Efficiency and Index of Agreement coefficient of 0.33 and 0.68 respectively.

The results from the Normalized Root Mean Square Error (NRMSE) are consistent with all the gridded dataset but has a higher value in PGF data with a mean value 1.07. However, the results for the temperature products showed an improved performance or mean similarity coefficient of all the gridded data products with values generally above 0.85 for both KGE and md (**Table 4.6**) and exhibit almost similar NRMSE across the stations in the study area.

*Table 4.5: Summary of performance of statistical metrics of gridded monthly precipitation data against observed data in Chad Basin*

**TOTAL MONTHLY PRECIPITATION**

STATION	KGE					INDEX OF AGREEMENT					NRMSE				
	CPC	CRU	GPCC	PGF	UDEL	CPC	CRU	GPCC	PGF	UDEL	CPC	CRU	GPCC	PGF	UDEL
Abeche	0.64	0.63	0.52	0.24	<b>0.71</b>	0.89	0.87	0.85	0.62	<b>0.91</b>	<b>1.058</b>	1.272	1.237	1.308	1.093
Banda	0.84	0.84	0.80	0.44	<b>0.87</b>	0.93	<b>0.96</b>	0.93	0.73	0.95	0.615	<b>0.455</b>	0.639	0.926	0.508
Bongor	0.71	<b>0.83</b>	0.78	0.32	0.81	0.89	0.95	<b>0.96</b>	0.67	0.95	0.706	0.527	<b>0.468</b>	1.073	0.505
Bossangoa	0.80	<b>0.94</b>	0.87	0.38	0.89	0.94	<b>0.98</b>	0.97	0.69	0.97	0.409	<b>0.244</b>	0.312	0.818	0.298
Doba	0.75	0.77	0.76	0.31	<b>0.78</b>	0.94	<b>0.96</b>	<b>0.96</b>	0.70	<b>0.96</b>	0.481	0.405	<b>0.394</b>	0.911	0.401
Maiduguri	0.70	0.83	<b>0.85</b>	0.37	0.44	0.84	<b>0.93</b>	<b>0.93</b>	0.66	0.90	0.949	<b>0.668</b>	0.701	1.188	0.888
Moundou	0.86	0.91	<b>0.94</b>	0.41	<b>0.94</b>	0.95	0.96	<b>0.97</b>	0.71	<b>0.97</b>	0.513	0.443	0.403	0.979	<b>0.392</b>
Ndjamena	0.72	0.66	0.61	0.27	<b>0.73</b>	<b>0.90</b>	0.86	0.86	0.62	<b>0.90</b>	0.827	0.928	0.91	1.273	<b>0.79</b>
Nguigni	<b>0.48</b>	0.45	0.43	0.13	0.47	0.82	<b>0.84</b>	0.82	0.57	0.81	1.097	<b>1.034</b>	1.066	1.421	1.095
Potiskum	<b>0.84</b>	0.67	0.62	0.26	0.69	0.75	0.90	0.88	0.62	<b>0.91</b>	1.045	0.728	0.795	1.209	<b>0.711</b>
Sahr	0.58	<b>0.65</b>	0.53	0.23	0.60	0.83	<b>0.86</b>	0.80	0.65	0.83	0.826	<b>0.711</b>	0.901	0.965	0.802
Samry-I	0.74	0.68	<b>0.78</b>	0.50	0.72	0.89	0.95	<b>0.96</b>	0.81	<b>0.96</b>	0.908	0.706	<b>0.601</b>	0.944	0.617
Sategui Deressia	<b>0.70</b>	0.60	0.64	0.49	0.51	<b>0.91</b>	0.90	<b>0.91</b>	0.86	0.89	0.777	0.893	0.81	<b>0.706</b>	0.975
Tsanaga	0.81	0.87	<b>0.90</b>	0.40	0.67	0.91	0.96	<b>0.99</b>	0.70	0.96	0.704	0.503	<b>0.293</b>	1.069	0.549
Zinder	0.64	0.63	0.52	0.24	<b>0.71</b>	0.89	0.87	0.85	0.62	<b>0.91</b>	0.837	0.899	0.918	1.248	<b>0.772</b>
<b>Mean Value</b>	<b>0.72</b>	<b>0.73</b>	<b>0.70</b>	<b>0.33</b>	<b>0.70</b>	<b>0.89</b>	<b>0.92</b>	<b>0.91</b>	<b>0.68</b>	<b>0.92</b>	<b>0.78</b>	<b>0.69</b>	<b>0.70</b>	<b>1.07</b>	<b>0.69</b>

*Note: Bold values indicate stations where gridded datasets have an improved skill.*

*Table 4.6: Summary of performance of statistical metrics of gridded temperature data against observed data in Chad Basin*

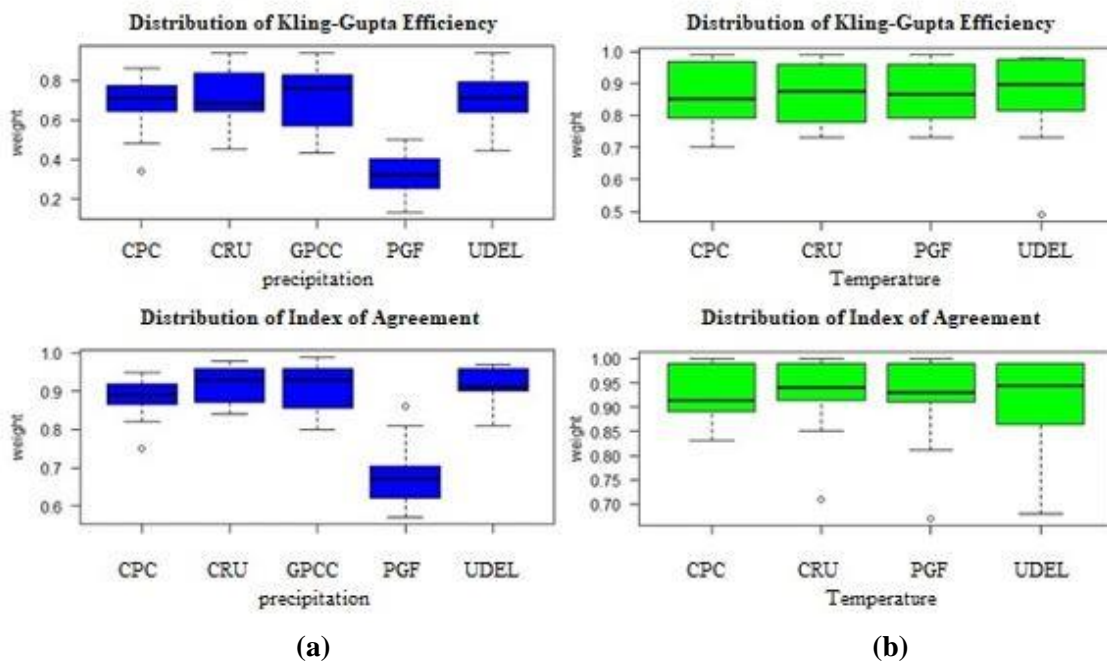
AVERAGE MONTHLY TEMPERATURE												
STATION	KGE				INDEX OF AGREEMENT				NRMSE			
	CPC	CRU	PGF	UDEL	CPC	CRU	PGF	UDEL	CPC	CRU	PGF	UDEL
Bilma	0.97	0.97	<b>0.98</b>	<b>0.98</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.038</b>	0.041	0.042	<b>0.038</b>
Bossangoa	0.81	0.80	0.83	<b>0.86</b>	0.90	0.91	<b>0.92</b>	0.91	0.037	0.035	<b>0.033</b>	0.038
Bouar	0.77	0.73	0.74	<b>0.85</b>	0.83	0.71	0.67	<b>0.89</b>	0.054	0.08	0.091	<b>0.043</b>
Geneina	0.81	0.82	0.84	<b>0.93</b>	0.88	0.93	0.93	<b>0.97</b>	0.076	0.06	0.057	<b>0.043</b>
Maiduguri	<b>0.77</b>	0.76	0.75	0.49	0.92	<b>0.93</b>	0.92	0.68	0.064	<b>0.059</b>	0.063	0.12
Maina Sorda	<b>0.97</b>	<b>0.97</b>	0.96	<b>0.97</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	0.033	<b>0.026</b>	<b>0.026</b>	<b>0.026</b>
Moundou	0.83	<b>0.85</b>	0.84	0.78	0.90	<b>0.92</b>	0.90	0.84	0.052	<b>0.046</b>	0.054	0.074
Ndjamena	0.92	<b>0.94</b>	<b>0.94</b>	0.93	0.96	<b>0.97</b>	0.96	<b>0.97</b>	0.047	<b>0.04</b>	0.045	0.042
Ngaoundere	0.70	<b>0.75</b>	0.73	0.73	0.84	<b>0.85</b>	0.81	0.84	0.052	<b>0.051</b>	0.061	0.053
Nguigni	<b>0.98</b>	0.95	0.96	<b>0.98</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	0.031	<b>0.024</b>	0.025	<b>0.024</b>
Sahr	0.87	<b>0.90</b>	0.89	0.86	0.91	<b>0.95</b>	0.93	0.92	0.046	<b>0.031</b>	0.039	0.045
Zinder	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	0.98	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.99	0.019	<b>0.017</b>	0.018	0.022
<b>Mean value</b>	<b>0.87</b>	<b>0.87</b>	<b>0.87</b>	0.86	<b>0.93</b>	<b>0.93</b>	0.92	0.92	0.046	<b>0.043</b>	0.046	0.047

*Note: Bold values indicate stations where the gridded dataset has an improved skill.*

The results from **Table 4.5** and **Table 4.6** revealed that CPC, CRU, GPCC and UDel gridded precipitation products has exhibited an improved skill in about 20%, 20%, 26.7% and 40% in terms of KGE and 13.3%, 40%, 46.7% and 46.7% in terms IA in all the stations respectively.

The temperature data has shown a more suited similarity with the observed station data with CPC, CRU, PGF, and UDel recorded an acceptable skill in 33.3%, 50%, 25% and 50% in terms of KGE and 33.3%, 75%, 41.7% and 50% in terms of md respectively.

The metrics used in this study are presented using box plots in **Figure 4.4** and revealed a consistent variation in their skill to replicate the observed precipitation and temperature data except for PGF gridded precipitation that may pose a large uncertainty in climate variable prediction. This may be due to, but not limited to interpolation technique, source and quality of observed data used in its development covering the entire Chad basin.



**Figure 4.4:** Boxplot of statistical metrics in the Lake Chad basin (a): KGE and md of gridded precipitation against observed station data. (b): KGE and md of gridded temperature against observed station data.

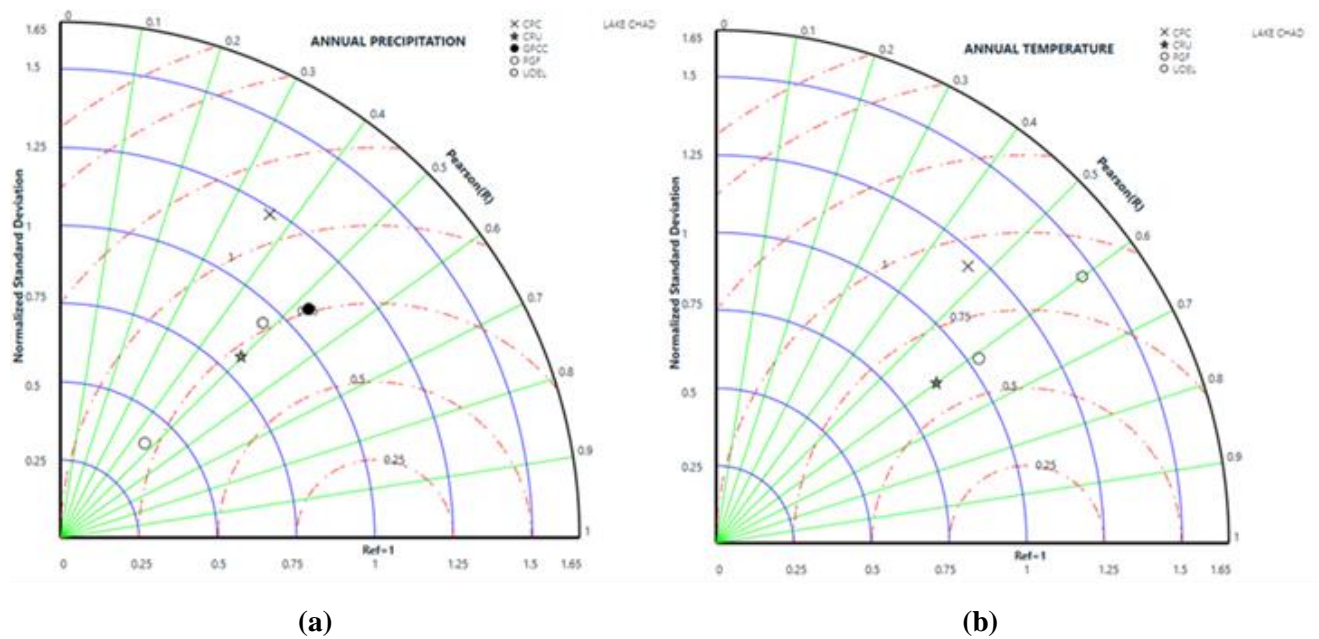
The similarity or agreement between gridded precipitation/temperature dataset and observed data for the period 1979 – 2012 at an annual, pre-monsoon and monsoon scale was

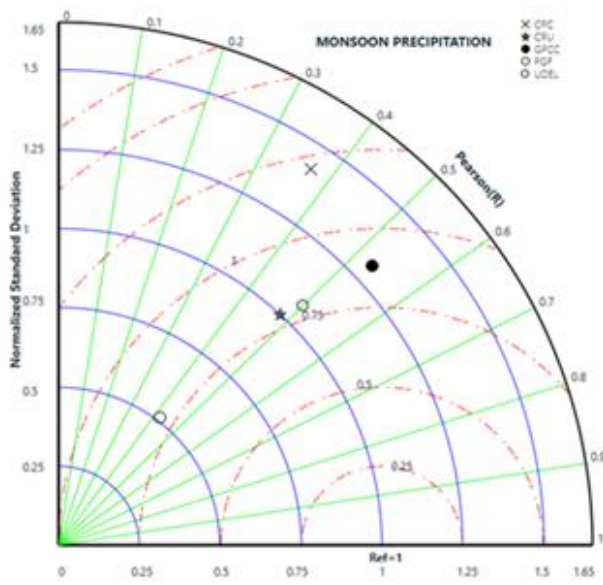
evaluated using Taylor diagram (Figure 4.5). The result indicates that the polar plots lie in the first quadrant which revealed that all correlation values are positive.

The results revealed a notable variability in normalized standard deviation in the gridded precipitation products in the three timescales. Although GPCC and CPC datasets recorded the best performance with a Pearson correlation coefficient greater than 0.5 and PGF has the least performance due to wider variability and lower Pearson correlation coefficient across the annual and seasonal timescales.

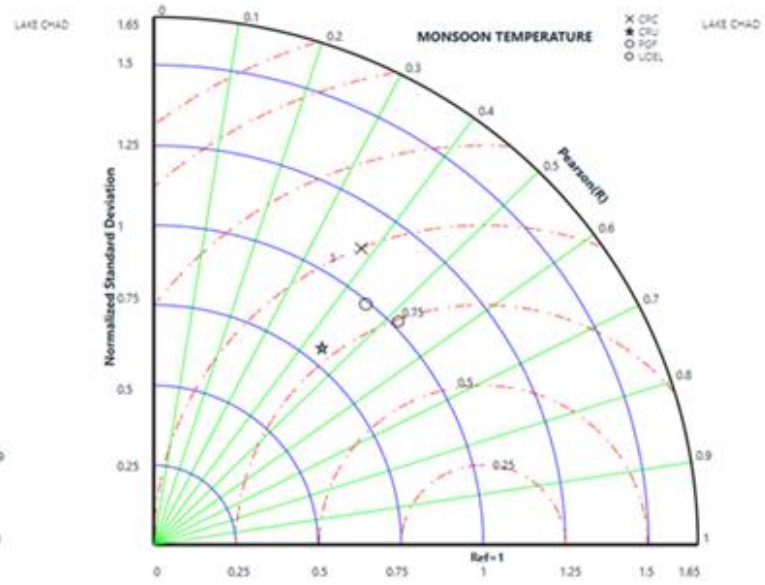
However, PGF data showed an effective performance in the temperature products with normalized standard deviation (Variability)  $\sim 1.0$  in the annual and monsoon season and an acceptable correlation to the observed data.

Although the gridded temperature products in general exhibit a more suitable skill with low observational uncertainty in replicating the observed data at annual timescale relatively well compared to monsoon and pre-monsoon season. The differences in their performance were apparent but seemingly a quite similar root mean square difference was observed in the precipitation products at all time scale.

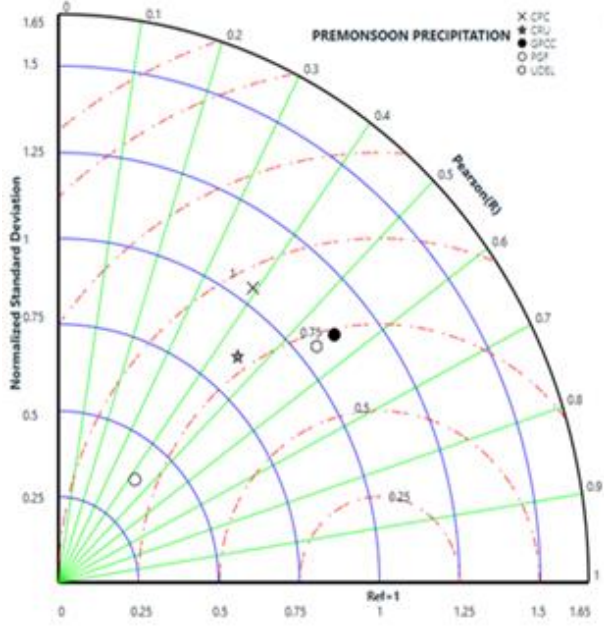




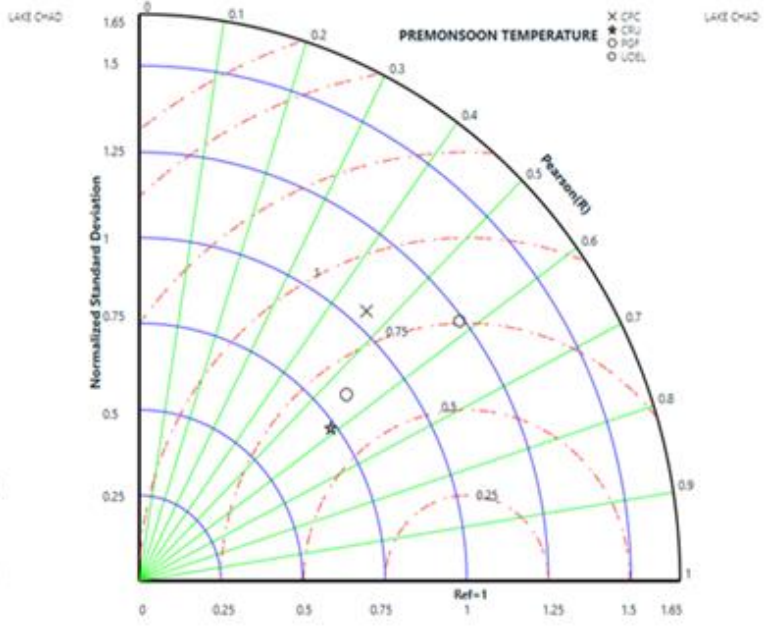
(c)



(d)



(e)



(f)

**Figure 4.5:** Taylor diagrams for time series data (1979 – 2012). (a): Annual precipitation of gridded and observed station data (b): Annual temperature of gridded against observed station data. (c): Monsoon precipitation of gridded against observed station data. (d): Monsoon temperature of gridded against observed station data. (e): Premonsoon precipitation of gridded against observed station data. (f): Premonsoon temperature of gridded against observed station data. Blue line is Normalized station deviation, Green line is Pearson correlation coefficient and Red line is Normalized root mean square error.

**4.5.3 Trend Analysis of gridded data**

The trend of precipitation and temperature were evaluated at annual and monsoon seasons for the gridded dataset and observed station records for the period 1979 – 2012. For the sake

of clarity, the performance of the datasets was assessed based on statistically increasing trends, decreasing trends or no trends respectively, arising due to limited station records to ease the complexity of analysis.

The Z-statistic value, (**Table 4.7 - Table 4.10**) showed the unidirectional trend of precipitation and temperature at the Annual and Monsoon season respectively. The results revealed that 20% and 26.7% of the observed station data showed a statistically decreasing trend of precipitation at annual and monsoon season, while 8.3% and 33.3% of the observed station data showed a statistically decreasing trend of temperature at annual and monsoon season respectively.

The stations where the analysis recorded a declining precipitation trend, for example, Maiduguri, Nguigni and Bongor are situated in the Semi-arid and Sudano zone. Declining temperature trend was observed at Sahr, Moundou, Ndjamena, Zinder are in the Sahelo-Sudan and Guinean zone, while other stations indicate statistically increasing unidirectional trends in the study area consistent with the findings in (Conway et al., 2009; Nkiaka et al., 2017; Sarr, 2012).

However, the result revealed some varying degree of mismatch between the gridded and observed station records across the basin, for example, CPC, CRU, GPCC, PGF and UDel data showed 66.67%, 73.33%, 60.0%, 66.67% and 73.33% agreement in unidirectional trend at annual time scale and 73.33%, 73.33%, 46.67%, 66.67% and 66.67% agreement in unidirectional trend in monsoon season against the observed data respectively.

The evaluated trend indicated that the CRU and CPC datasets have the ability to replicate the observed station records with a higher percentage agreement compared to other products. The temperature products showed a higher degree of agreement of unidirectional trend against the station records with fairly similar results - for example, CRU, PGF and UDel showed an improved performance in annual (91.67%) and monsoon (66.67%) seasons.



Although the CRU gridded datasets showed strong increasing trends and may tend to overestimate the temperature of the basin.

Therefore, the choice of the dataset adopted for hydro-climatic studies is critical for accurate assessment in order to reduce uncertainty in the prediction of hydrologic variables in modelling studies to achieve good policy planning in water resource management.

*Table 4.7: Mann Kendall Z-Statistics values of linear trend of Annual gridded and observed precipitation for Chad Basin (1979-2012)*

Stations	Precipitation Trend					
	OBSERVED	CPC	CRU	GPCC	PGF	UDEL
Abeche	-1.60	-0.85	<b>2.59</b>	<b>1.96</b>	<b>2.23</b>	<b>2.43</b>
Banda	<b>2.02</b>	0.15	1.07	<b>2.05</b>	0.00	1.10
Bongor	-0.71	-1.63	1.17	0.15	1.42	1.63
Bossangoa	0.33	-0.82	0.60	-0.53	0.44	-0.53
Doba	0.16	-1.01	0.21	0.59	0.50	0.33
Maiduguri	<b>-2.19</b>	0.79	<b>2.25</b>	<b>2.16</b>	<b>2.45</b>	-1.90
Moundou	0.69	0.26	1.21	0.18	1.01	0.34
Ndjamena	0.67	<b>2.29</b>	<b>2.02</b>	0.29	1.42	<b>2.54</b>
Nguigni	-0.20	0.71	<b>2.31</b>	0.92	<b>2.02</b>	<b>2.31</b>
Potiskum	1.82	1.51	<b>2.51</b>	0.19	<b>1.96</b>	1.28
Sahr	0.06	0.98	1.19	<b>2.58</b>	1.81	<b>2.11</b>
Samry-I	1.91	-1.57	1.16	0.15	1.63	1.63
Sategui D.	0.74	0.77	1.10	0.36	1.41	1.30
Tsanaga	1.13	0.27	1.39	1.75	1.69	1.39
Zinder	1.60	1.07	<b>2.13</b>	0.65	<b>2.16</b>	<b>2.16</b>
<b>% Agreement</b>	<b>-</b>	<b>66.67</b>	<b>73.33</b>	<b>60.00</b>	<b>66.67</b>	<b>73.33</b>

*Note: Highlighted bold values indicate significant trends*

*Table 4.8: Mann Kendall Z-Statistics values of linear trend of Monsoon gridded and observed precipitation for Chad Basin (1979-2012).*

Stations	Precipitation Trend					
	OBSERVED	CPC	CRU	GPCC	PGF	UDEL
Abeche	-1.24	-0.38	0.24	1.68	<b>2.39</b>	<b>2.07</b>
Banda	1.07	1.17	0.50	1.07	-0.24	-0.27
Bongor	-0.74	-0.95	1.51	-0.09	1.42	1.50
Bossangoa	0.01	-0.47	0.53	-0.92	0.12	-0.83
Doba	0.46	0.83	0.24	-0.06	0.71	0.53
Maiduguri	<b>-1.99</b>	1.18	<b>2.38</b>	<b>2.25</b>	<b>2.94</b>	-1.80
Moundou	0.38	1.44	1.32	-0.14	1.24	0.34
Ndjamena	0.75	<b>2.46</b>	1.88	-0.08	1.51	<b>2.42</b>
Nguigni	-0.47	0.53	<b>2.16</b>	0.62	<b>1.96</b>	<b>2.25</b>
Potiskum	1.40	1.70	<b>2.69</b>	0.54	<b>2.22</b>	1.28
Sahr	0.95	1.84	0.67	1.60	1.60	0.39
Samry-I	<b>2.11</b>	-0.95	1.51	-0.09	1.43	1.50
Sategui D.	0.36	1.33	1.90	0.24	<b>2.16</b>	0.85
Tsanaga	0.98	0.53	1.71	1.33	1.90	1.07
Zinder	1.10	1.39	<b>2.08</b>	0.62	<b>2.11</b>	<b>1.96</b>
% Agreement	-	<b>73.33</b>	<b>73.33</b>	<b>46.67</b>	<b>66.67</b>	<b>66.67</b>

*Note: Highlighted bold values indicate significant trends*

*Table 4.9: Mann Kendall Z-Statistics values of linear trend of Annual gridded and observed temperature for Chad Basin (1979-2012).*

Stations	Temperature Trend				
	OBSERVED	CPC	CRU	PGF	UDEL
Bilma	<b>3.21</b>	<b>3.18</b>	<b>4.28</b>	<b>3.33</b>	<b>3.39</b>
Bossangoa	1.36	<b>2.53</b>	<b>3.75</b>	<b>3.30</b>	<b>2.25</b>
Bouar	0.50	<b>2.58</b>	<b>3.43</b>	<b>2.65</b>	1.90
Gencina	<b>2.80</b>	-1.79	<b>2.46</b>	<b>3.21</b>	<b>2.00</b>
Maiduguri	0.37	<b>2.55</b>	<b>3.85</b>	<b>3.07</b>	<b>3.03</b>
Maina S.	<b>3.35</b>	<b>3.67</b>	<b>3.61</b>	<b>3.11</b>	1.60
Moundou	-0.37	<b>2.65</b>	<b>3.49</b>	<b>3.30</b>	<b>3.86</b>
Ndjamena	<b>2.37</b>	<b>5.19</b>	<b>3.33</b>	<b>2.93</b>	1.75
Ngaoundere	1.60	-0.02	<b>3.64</b>	<b>3.05</b>	<b>3.83</b>
Nguigni	<b>2.99</b>	<b>2.96</b>	<b>3.42</b>	<b>2.95</b>	<b>2.03</b>
Sahr	0.50	<b>3.25</b>	<b>3.08</b>	<b>3.92</b>	<b>3.41</b>
Zinder	<b>2.76</b>	<b>2.92</b>	<b>3.73</b>	<b>3.14</b>	<b>3.83</b>
% Agreement	-	<b>75.00</b>	<b>91.67</b>	<b>91.67</b>	<b>91.67</b>

*Note: Highlighted bold values indicate significant trends*

**Table 4.10:** Mann Kendall Z-Statistics values of linear trend of Monsoon gridded and observed temperature for Chad Basin (1979-2012).

Stations	Temperature Trend				
	OBS	CPC	CRU	PGF	UDEL
<b>Bilma</b>	<b>2.90</b>	<b>2.77</b>	<b>2.71</b>	<b>3.28</b>	<b>2.71</b>
<b>Bossangoa</b>	1.46	1.29	<b>3.33</b>	<b>3.46</b>	1.45
<b>Bouar</b>	0.79	1.13	<b>3.31</b>	<b>2.15</b>	1.11
<b>Geneina</b>	1.41	<b>-2.62</b>	1.69	<b>3.38</b>	1.72
<b>Maiduguri</b>	1.59	0.78	<b>2.57</b>	1.45	<b>2.21</b>
<b>Maina S.</b>	1.60	1.53	1.89	0.92	0.47
<b>Moundou</b>	-1.29	1.07	<b>2.90</b>	<b>3.04</b>	<b>2.98</b>
<b>Ndjamena</b>	-0.45	<b>3.27</b>	<b>2.61</b>	<b>2.02</b>	0.62
<b>Ngaoundere</b>	0.50	-0.82	<b>2.66</b>	<b>1.98</b>	<b>3.33</b>
<b>Nguigni</b>	<b>2.15</b>	1.76	<b>2.15</b>	1.42	0.52
<b>Sahr</b>	-0.82	0.77	<b>3.25</b>	<b>3.31</b>	1.66
<b>Zinder</b>	-0.33	-0.50	1.41	0.62	<b>2.17</b>
<b>% Agreement</b>	-	<b>58.33</b>	<b>66.67</b>	<b>66.67</b>	<b>66.67</b>

*Note: Highlighted bold values indicate significant trends*

The results of Sen’s slope of the trend analysis of annual and monsoon season at 95% level of confidence revealed that there are mismatches or difficulty of gridded precipitation and temperature dataset in reproducing a similar magnitude of trend relative to the station records (**Figure 4.6 - Figure 4.9**).

For example, the gridded datasets over-estimated and under-estimated the annual and monsoon precipitation and temperature in most of the stations. However, the mismatch has been acknowledged as one of the sources of uncertainty in modelling processes.

Although, in this assessment, the median magnitude of precipitation trend for CPC (1.200) and GPCC (1.000), and temperature trend of CRU (0.0185) and CPC/UDel (0.0075) has shown to be more effective relative to other datasets relative to the observed station data at annual and monsoon season across the basin.

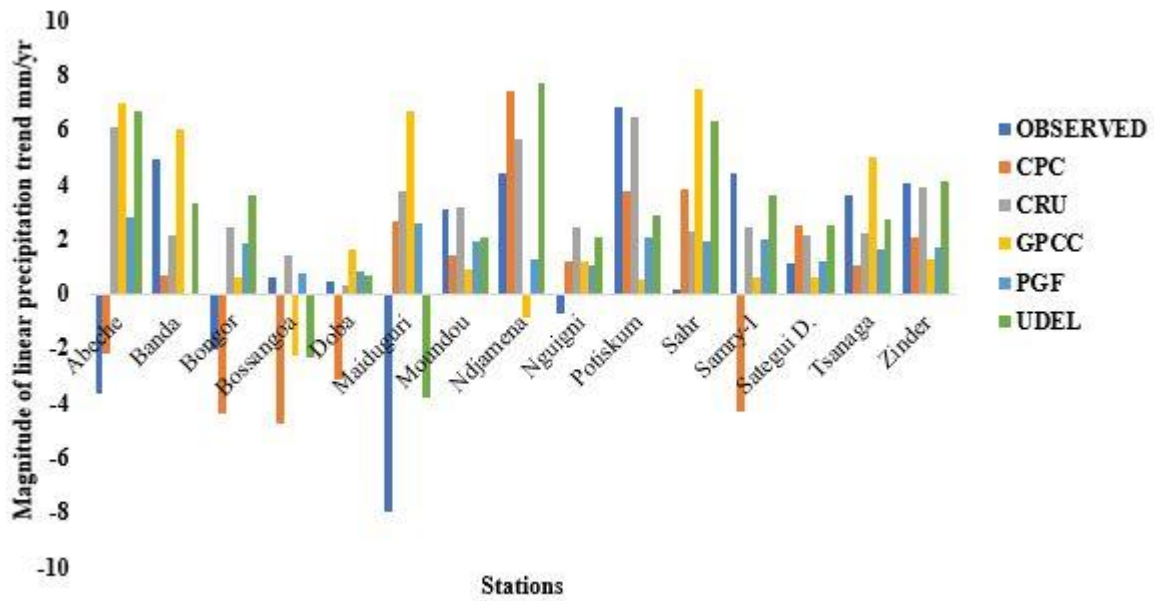


Figure 4.6: Magnitude of linear trend of Annual gridded and observed precipitation for Chad Basin (1979-2012)

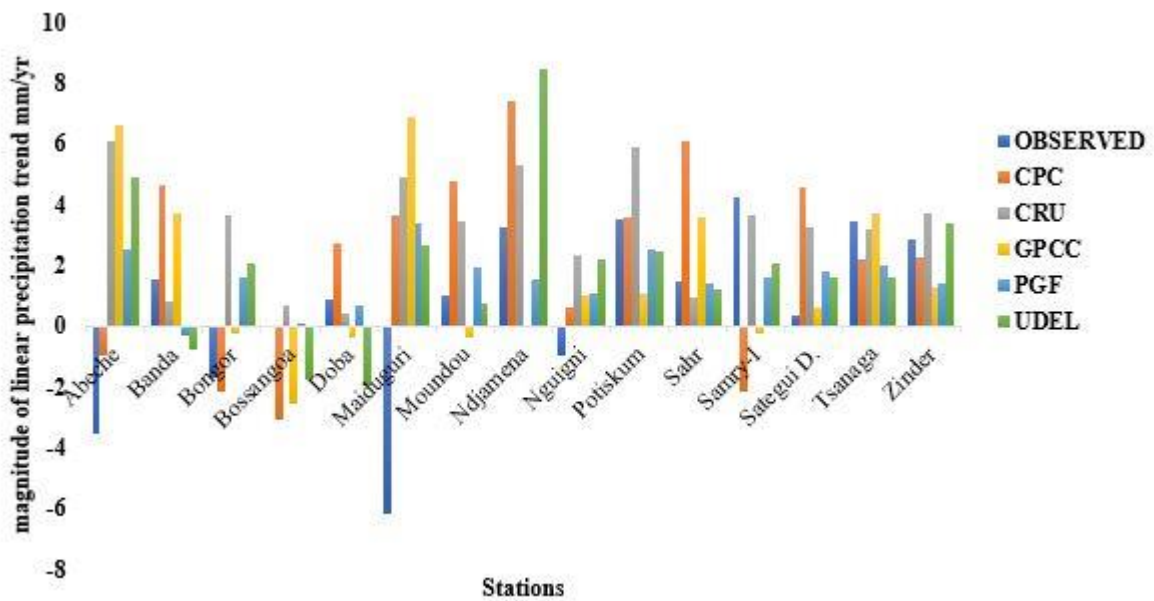


Figure 4.7: Magnitude of linear trend of Monsoon gridded and observed precipitation for Chad Basin (1979-2012)

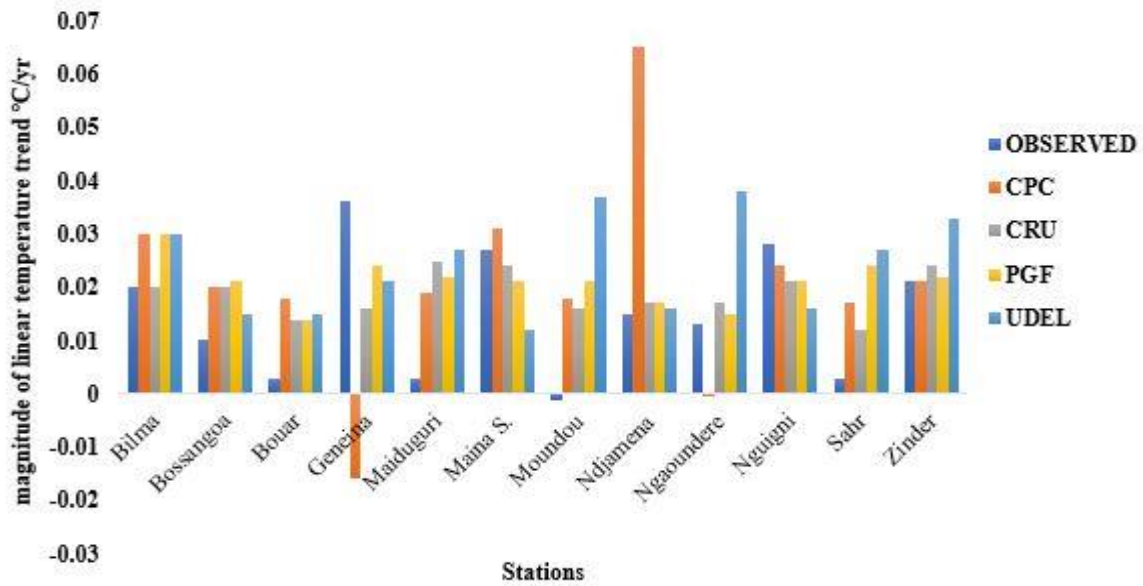


Figure 4.8: Magnitude of linear trend of Annual gridded and observed Temperature for Chad Basin (1979-2012)

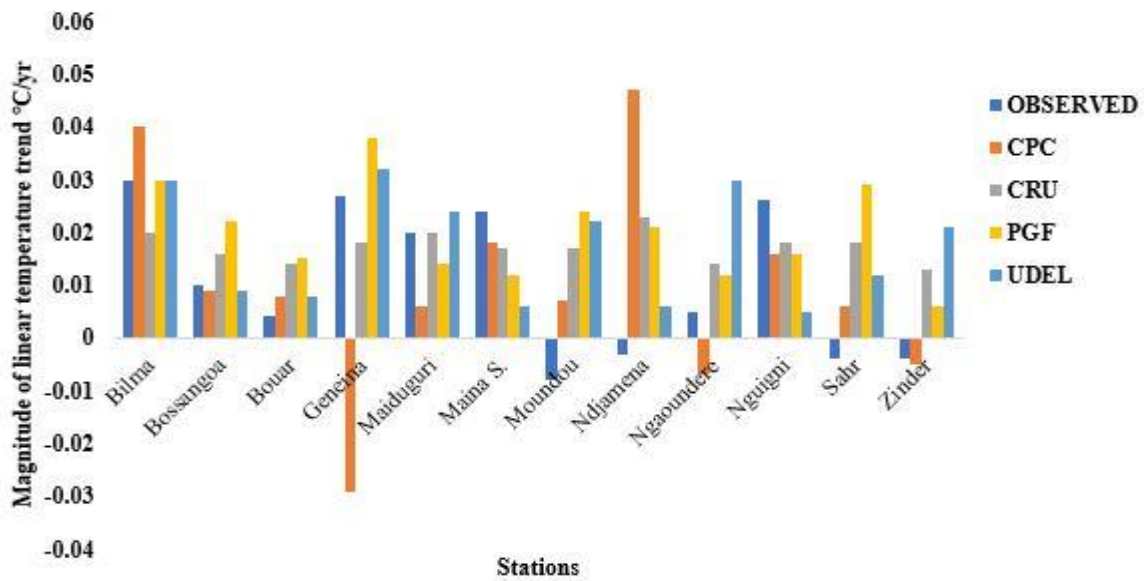


Figure 4.9: Magnitude of linear trend of Monsoon gridded and Observed Temperature for Chad Basin (1979 – 2012)

#### 4.6 Discussion

The performance of widely used, updated and recently available high resolutions gridded precipitations and temperature datasets were assessed over the available, quality controlled and reliable 15 observed precipitations and 12 temperature station records over the Lake Chad basin. The study temporal span was set to be 1979 – 2012 for consistency across all the datasets and met the requirement for a hydrologic impact study.

The study adopted a 0.5° resolution dataset and finally extracted and interpolated the gridded dataset to station resolution using the inverse distance weighting method for consistency and adequate for regional studies.

In summary, the evaluated gridded climate products have shown to differ in their skills across the selected metrics, although the skill or performance exhibited across all the gridded climate products are promising especially with regards to temperature because climate models simulate the variable effectively relative to precipitation as acknowledged in Barron and Moore, (1994). However, these notable differences observed in the performance of the gridded dataset may be misleading in the selection or choice of a dataset based on evaluation using a single metric.

Therefore, a good, gridded dataset should have the ability to accurately replicate the climate pattern and amplitude of spatial and temporal variability across different performance metrics which is important and critical for the thorough and accurate hydro-climatic application. This however has been acknowledged in Gampe and Ludwig, (2017), stating that a single performance metric cannot adequately be relied upon for the selection of gridded datasets for application globally. The merit of this methodology was predicated on the fact that the combination of multiple metrics for selection; reduces the chance of underperforming models indicating a reliable performance for the wrong reasons (Flato et al., 2013).

*Table 4.11: Summary of the best two (or three when the metrics have the same performance) gridded precipitation and temperature dataset across the performance metrics considered in this study.*

Performance Metrics	Precipitation					Temperature			
	CPC	CRU	GPCC	PGF	UDEL	CPC	CRU	PGF	UDEL
<b>Symmetric Uncertainty</b>	-	√	√	-	-	-	√	√	-
<b>Kling-Gupta Efficiency</b>	√	√	-	-	-	√	√	√	-
<b>Index of Agreement</b>	-	√	√	-	√	√	√	-	-
<b>NRMSE</b>	-	√	-	-	√	√	√	√	-
<b>Taylor's Diagrams</b>	√	-	√	-	-	-	-	√	-
<b>Trend Analysis</b>	√	√	-	-	-	-	√	√	√
<b>Metrics Agreement (%)</b>	<b>50</b>	<b>83.3</b>	<b>50</b>	<b>0</b>	<b>33.3</b>	<b>50</b>	<b>83.3</b>	<b>83.3</b>	<b>16.7</b>

The gridded dataset has been shown to exhibit some inherent weaknesses across the performance metrics, especially in simulating the trend and magnitude of precipitation and temperature across the annual and monsoon season, for example, the CRU data has shown to be efficient in replicating the observed station data relative to the other dataset but over-estimate the temperature trend across annual and monsoon season.

However, the skills across the performance metrics considered in this study revealed that CRU data is more reliable relative to observation data in 5 out of 6 (83.3%) performance metrics considered for both precipitation and temperature, while CPC and GPCC have a relatively effective performance, with 3 out of 6 (50%) for precipitation and PGF had 5 out of 6 (83%) skill for temperature in terms of the performance metrics respectively.

The superiority of the CRU data may point to the fact of its development employing a large of station data than other gridded data products do. However, the products were able to exhibit a satisfactory performance and represent the variability reasonably well and indicate that they possessed the ability to provide reliable hydro-climatic information with a lower level of uncertainty in their prediction.

The findings of this study, using multi-criteria decisions based on the assessment of the dataset by several performance metrics revealed that CRU, GPCC and CPC precipitation dataset and CRU and PGF temperature dataset showed an optimum performance and were recommended for application for hydro-climatic studies in the Lake Chad basin to enhance prediction and achieve a low level of uncertainty in terms of similarity to the observation station precipitation and temperature data respectively.

Furthermore, this finding has been the first study to evaluate the performance and uncertainty associated with gridded precipitation and temperature dataset in Lake Chad hydrologic basin and could not be compared to previous studies. However, findings from other studies in the regions with similar climatology have shown to be consistent with the

present findings. For example, a study in the arid region of Pakistan, indicates that GPCC was found to have an acceptable agreement with the observed dataset (Ahmed et al., 2019), CRU dataset was also found to have effectively reproduce the climatology of the Niger Delta region of Nigeria (Ibrahim Hassan et al., 2020) and also GPCC, CRU and CPC was consistent with the observed climate data in the mountainous region of South Africa (Manatsa et al., 2008).

#### **4.7 Conclusion**

This paper evaluated the performance of long-time series of high-resolution gridded precipitation and temperature datasets and their suitability for hydro-climatic studies in the data-sparse Lake Chad basin. The emphasis in this assessment was to employ multiple performance metrics to evaluate the skill exhibited of the selected dataset to replicate the quality-controlled observed meteorological station records available and provide guidance and methodology based on the multi-criteria decision on the choice of reference dataset suitable for climate and hydrologic impact studies in a data-sparse region.

The results of the analysis revealed that all the gridded precipitation datasets have the ability to replicate the observed climate record with varying levels of uncertainty, except the PGF dataset that exhibits unsatisfactory performance in terms of KGE, md and NRMSE. However, all the temperature datasets showed a strong agreement that was consistent with the observed data.

The result from the Taylor diagram indicates a notable variability in normalized standard deviation across the annual, pre-monsoon and monsoon season but have an acceptable Pearson correlation coefficient relative to the observed data record.

The trend analysis results have shown that the temperature dataset exhibits a suitable skill in replicating the trend over precipitation data i.e., the measure of the overall accuracy of the forecast, i.e., the degree of variability of the gridded dataset from the observation data across



all metrics of the assessment adopted is suitably represented by the temperature dataset. This may be related to the mechanism and representation of temperature in development of the dataset is simple while the precipitation is variable in its real state and very difficult to predict or model accurately.

Although there is a varying degree of mismatches in the magnitude of the trend across the stations, the CRU data exhibited a strong increasing temperature trend across all the stations at annual and monsoon season and tend to over-estimate the temperature of the basin relative to the observed gauged data.

However, the multi-criteria decision-making approach applied based on the skill exhibited across the performance metrics in this study, revealed that the CRU, GPCC and CPC datasets are appropriate for precipitation and CRU and PGF dataset for temperature respectively, to accurately replicate the basin climatology with an acceptable and low level of uncertainty in the prediction of climate and hydrologic variables for improved policy planning and water resource management.

Furthermore, the results from this study highlighted that the choice of gridded data is critical for fair representation of historical and presumably reduces the propagation and transfer of uncertainty to projected future climate as acknowledged in (Flato et al., 2013).

Therefore, due to differences in temporal resolution of gridded datasets, the daily CPC precipitation and PGF temperature and monthly CRU precipitation/temperature dataset is recommended for the downscaling and bias correction of global climate models in hydrologic modelling processes depending on the input data requirements of hydrologic models in Lake Chad Basin and the need for improvement in climate models development to effectively reproduce satisfactory temporal and spatial variability in climate indices and limits uncertainties and improve prediction accuracy in the quantification of projected climate hazards for efficient water policy decision making.

## 4.8 Afterward

The findings from this study can only address the uncertainties related to the historical and current climate and watershed assessment of basin-scale features in hydrologic modelling. It is worth noting that insufficient dense network of station observations and information may impede the appropriateness of the gridded dataset's ability to reproduce the observed trends because the limited available station data may not be representative of the wider basin.

Perhaps extending the study to examine the appropriateness of the gridded datasets to reproduce extremes e.g., SPEI drought and flood hazards will add value to the reliability assessment because the different gridded datasets used may have been generated using different source gauges at different times and may not necessarily be expected to reliably reflect trends and magnitude of change of the basin climate.

However, in order to investigate and understand the potential for the reduction of input and scenario uncertainty in hydrologic modelling concerning future developments, the recommendations from this study were applied afterwards in the next phase of the research to investigate and understand effectively the implementation of downscaling and bias correction techniques and innovative capability assessment of GCM models by data pruning for reliable application in hydrologic model development at local basin-scale.

## **CHAPTER 5: APPLICATIONS OF BORUTA ALGORITHMS AS A ROBUST METHODOLOGY FOR PERFORMANCE EVALUATION OF CMIP6 GENERAL CIRCULATION MODELS FOR HYDRO-CLIMATIC STUDIES**

### **Preamble**

This chapter outline key issues related to the application of CMIP6 climate data in hydrologic impact study and proffer solution to improve their evaluation framework objectively at regional and local basin-scale. This is a source of concern in watershed with sparse observational data to validate the alternative and improve key climatological basin features.

Recent research on the application general circulation models at regional and local basin using advanced methodology tend to overlook the data scale and timestep necessary for hydrologic modelling and the after effect of multi-model ensemble performance to prevent inclusion of models that compromise accuracy and increase uncertainty in watershed modelling.

This anomaly may be introduced by inadequate internal parameterization through downscaling and bias correction to finer resolution of the climate model structure causing distortions in watershed hydrologic dynamics such as trend and magnitude of SPEI drought and flood hazards and flawed depiction of the basin hydrology.

In an effort to address this shortcoming, a methodology was proposed to reduce climate data uncertainty in numerical modelling process and provide an objective evaluation framework that is independent of assumptions and parameterization methods of GCM model development.

The methodology was objectively applied to analyse models distinctly from the different modelling centres and their successful applications at regional and local watersheds especially in data-sparse regions where the uncertainty levels can be amplified in hydrologic

models with average confidence in the validation process. It is worth noting that, the assessment was limited to 16 GCM models based on two criteria.

(1). Models that possess data at daily step necessary for hydrologic modelling.

(2). Models that possess all climate change scenario (SSPs) data relevant for hydroclimatic impact study.

Fitting quantitative correlations between large-scale GCM and local (Observations) climate variables may provide several avenues for additional improvements in multimodel ensembles. The main benefit of using machine learning methodologies in climate model simulations is that they can take advantage of significant investments already made in sizable ensembles of climate models with various boundary conditions and the non-linear nature of the climatic data series of the prediction process.

They also offer opportunities for training and testing the models, to reduce the complexity of the model data, minimises over-fitting by choosing the best features, and iteratively eliminates features that are statistically less relevant in order to improve predictions due to non-stationarity and large spatiotemporal scale of the climate data. This will facilitate the comprehension of various structural uncertainties that could impact the precision of the multimodel ensemble produced for hydrologic modelling.

Similarly, the evaluation of the efficacy of the methodology was validated by assessing the severity, trend and magnitude of drought and flood hazards across the four climatic zone of the basin using standardized precipitation evapotranspiration index (SPEI) relative to observation.

The SPEI drought monitoring index was adopted in favour of other indicators because of its ability to identify various drought types and impacts in the context of climate change, thanks to its multi-scalar properties and its ability to take the effect of ET on drought severity especially in water resource assessment.

Finally, “trend envelope” as used in the assessment of drought and flood event in this chapter entails the scale of change in the severity of drought and flood indicators at basin scale.

The Paper following was published in “Theoretical and applied climatology” titled “Application of Boruta algorithms as a robust methodology for performance evaluation of CMIP6 general circulation models for hydro-climatic studies (2023) <https://doi.org/10.1007/s00704-023-04466-5>.

### **5.1 Abstract of paper**

Regional climate models are essential for climate change projections and hydrologic modelling studies, especially in watersheds that are overly sensitive to changes in climate. Accurate hydrologic model development is a daunting task in data-sparse regions where climate change’s impact on hydrologic and water quality processes is necessary for a well-informed policy decision on adaptation and hazard mitigation strategies.

Novel approaches have been evolving that evaluated GCMs with the objective of improved parameterization to limit uncertainty and improve hydrologic model development. However, conclusions drawn should be purpose-driven based on intended usage.

This study provides an overview of the state-of-the-art Boruta random forest as a robust methodology in the performance evaluation of GCMs models for hydroclimatic study. Highlights from the assessment indicate that: (1) there is consistency in replicating the three observed climate variables of daily precipitation, maximum and minimum temperature respectively, (2) improved temporal correlation ( $R^2 = 0.95$ ) in annual precipitation with a mean bias of 0.638mm/yr.; when compared to symmetrical uncertainty (SU) ( $R^2 = 0.82$ ), and all models ensembles (AME) ( $R^2 = 0.88$ ) with associated biases of 68.19mm/yr. and 10.57mm/yr. respectively.

Evaluation of the multi-year climate extreme indices, trends and magnitude reveal that there is a fair representation of basin-scale observed climate extreme events. However, the Boruta random forest approach exhibited an accurate statistical trend and magnitude of the extreme event in the basin. The findings of the study revealed that Boruta random forest, enhanced GCM dataset evaluation can present a simple and efficient methodology to examine the limitations associated with selected GCM ensemble for impact study in hydrology.

**Keywords:** CMIP6 models, climatology, performance metrics, model uncertainty, regional modelling, Lake Chad Basin.

## 5.2 Introduction

To reduce observational uncertainty and the impact of projected changes in climate and catchment hydrologic variables on the availability of water resources for an accurate assessment of sustainability concerns at local, regional, and global scales, a successful hydro-climatic study requires accurate data with high temporal and spatial resolution. Therefore, future variations of global surface air temperature and precipitation are deemed important for climate change policy formulation, and hazard assessment which may affect the human livelihood and regional economy, especially in developing economies (Akhter et al., 2017) and ensure proper mitigation and adaptation strategies are adopted for integrated water resources management (Ahmed et al., 2019a).

Global energy balance has undergone periodic changes due to increased emission of atmospheric greenhouse gas owing to widespread fossil fuels usage and industrial activities across the globe (Chu et al., 2010; Huang et al., 2011; Salman et al., 2018). The rise in the concentration of greenhouse gases has been noted to increase the Earth's temperature at a rate of 0.15 °C/decade (Flato et al., 2013), and has potentially a severe effect on the earth's ecosystem and most notably in the tropical regions (Khan et al., 2019; Mohsenipour et al., 2018; Wang et al., 2014).

Mishra and Liu, (2014) observed that the impact of the variations of rainfall-induced by projected climate change would be more intense in the tropical regions of the world. This phenomenon has become a significant socio-economic and political issue (Alamgir et al., 2019).

General circulation models are the main tool for predicting future changes in the global climate (Maraun et al., 2010), because they possessed the potential in replicating historical climatic changes as well as future changes considering the greenhouse gas concentration (Goyal et al., 2012), and other shared socio-economic pathways integrated into the recently developed coupled model inter-comparison project phase six models.

Evaluating the performance of general circulation models (GCM) outputs is instrumental for simulating the historical and future basin scale hydrologic cycles using hydrological models. The accuracy of the precipitation and temperature outputs from GCMs, when used as an input data for hydrological modelling, affects the reliability of hydrologic variables prediction.

Therefore the spatial and temporal performances, as well as their seasonal variations, need to be evaluated arising from natural and climate-induced radiative forcing in a multi-model context (Eyring et al., 2016; Gusain et al., 2020; Wang et al., 2021) to reduce uncertainties and enhance prediction accuracy for reliable policy planning of basin scale hydrology.

GCMs are developed to produce projections at a coarse spatial scale and could not resolve finer scale features such as clouds, land use change etc. Studies at the regional scale require that the output be downscaled to finer a resolution. The mismatch between the hydro-climatic information required and the GCM output is a major cause for concern or obstacle in hydrologic impact studies (Willems and Vrac, 2011).

Careful assessment of various downscaling techniques is important to enable accurate prediction and a solution to the mismatch between regional hydro-climatic information and GCM outputs at a spatial scale of between 5 – 50 km (Yang and Delsole, 2012). However, McSweeney et al., (2015) argued that downscaling the complete ensemble may not be desirable or necessary to produce a meaningful range of future climate conditions relevant to evaluate hazards associated with future climate change because high resolution downscaling is labour and computing resource intensive.

Therefore, various strategies need to be explored to sample from the available CMIP6 GCMs and their shared socio-economic pathways (SSPs) scenarios to generate projections relevant for water policies at catchment scale. The assessment process is to identify the challenges and opportunity to exclude models deemed unsatisfactory in the representation of key climate features.

These models have been utilized to simulate the historical and projected changes in climate at global (Sachindra et al., 2014; Wright et al., 2015), regional (Abbasian et al., 2019; Ahmed et al., 2019a; Salman et al., 2018) and local basin scale (Akhter et al., 2017; Hassan et al., 2020).

However, over the years, six phases of the Coupled Model Inter-comparison Projects (CMIPs) have been developed by various modelling centres for climate change studies. Previous studies utilizing the fifth phase (CMIP5) GCMs indicated some significant improvements in simulated climate variables relative to the latter generation (CMIP3) models (L. Wang et al., 2016), due to improved knowledge of climate science.

Studies that utilized CMIP5 GCMs have identified the Sahel region, tropical West Africa and Southern part of Africa as hotspot for severe impact of regional climate change (Diffenbaugh and Giorgi, 2012; Niang et al., 2014). Although studies are evolving utilizing



CMIP6 models recently to understand their ability in replicating historical and future climate at the regional and global scale.

There are notable differences between the CMIP6 GCMs and the earlier phases which integrate new specification for greenhouse gas concentration and emission scenarios, as well as land use scenarios (Gidden et al., 2019).

A limited number of studies utilizing the CMIP6 GCMs indicated improvement and its robustness over earlier phases in some regions, for example Australia (Grose et al., 2020); Africa (Almazroui et al., 2020; Ayugi et al., 2021; Sian et al., 2021); Nigeria (Shiru and Chung, 2021); South Korea (Song et al., 2020).

The performances may vary from one region to another and therefore, necessitate the evaluation of their performance at any region of interest, especially in data sparse catchments before adoption for proper representation of the climatic feature to avoid projections with large uncertainty range which may result to over confidence and poor adaptation (McSweeney et al., 2015).

Several techniques have been used to assess the performance of climate models such as ensemble averaging (Giorgi and Mearns, 2002), combined statistical measures like root mean square error, mean bias error, mean absolute error, correlation coefficient into one performance index (Gu et al., 2015; Wu et al., 2017), Relative entropy (Shukla et al., 2006), Probability Density Functions (PDF) (Perkins et al., 2007), Bayesian regression (Chandler, 2013), Clustering (Knutti et al., 2013), Correlation (Li et al., 2016; Xuan et al., 2017), and Symmetrical uncertainty (Salman et al., 2018).

Nevertheless, often times the methods are quite cumbersome or produce inconsistent performance across climate variables that will require a multi-criteria decision tool for selection of appropriate GCMs for hydrologic impact studies.

However, based on previous studies; no consensus was made regarding the choice of GCM selection approach and McMahon et al., (2015) posited that no single criterion is accepted universally for GCM performance, although assessment at multiple time scales and the ability of a GCM capturing the spatial structure of a catchment's key climate feature may give crucial information for water resource management accurately, especially at basins with high climate variability (Ahmed et al., 2019; Gleckler et al., 2008).

The downside of the statistical metric is that they only assessed certain features of the time series data when compared to the observed data (McSweeney et al., 2015; Weigel et al., 2010), and often provide contradictory results across different metrics (Nashwan and Shahid, 2019; Raju et al., 2017).

Studies on GCM selection are grouped into the past performance method (Biemans et al., 2013), where the models are selected based on their capability to replicate the historical climate and the envelope method (Warszawski et al., 2014) where the ensemble of GCMs are selected based on the ability to encompass the whole range of future projections.

However, both of these approaches have weaknesses; for example, the former method neglects agreement between GCMs to simulate projected future climate, while the latter method ignores the ability of GCMs to replicate the historical observed climate (Ahmed et al., 2019b).

Machine learning algorithms are gaining a lot of attention, especially filters and wrappers such as Clustering (Knutti et al., 2013; Raju and Kumar, 2016), Bayesian weighing (Min and Hense, 2006), Weighted skill score (Maxino et al., 2008; Perkins et al., 2007) etc. Furthermore, these techniques are used to evaluate and select the GCMs using a single performance index (Ahmadalipour et al., 2017; Fu et al., 2013; Raju and Kumar, 2016; Wójcik et al., 2014).

However, the techniques are found to have some inherent weaknesses such as the inability to capture the temporal variability of climate, and, the variation of frequency of climate extreme which is found to be an important factor in the assessment of model performance (Salman et al., 2018).

The major advantages of machine learning techniques are that, in feature selection where the dependent variables are analysed and ranked based on their importance or impact on the independent variable, and where features that are likely to decrease the efficiency of a model are screened out to reduce uncertainty in model development.

It is important in machine learning application to have the observed data well represented and the information within the data series possessed the ability to be learnable for the modelling process, as this is critical for the final performance (Kursa, 2016).

In GCM selection for hydro-climatic study, the information entropy based filter referred to as symmetrical uncertainty (SU) (Witten et al., 2005), has gained the attention of researchers due to its ability to select variable without bias and reliably.

The technique was used to rank GCMs according to their degree of similarity or otherwise with the observations for the entire time series data and it has the advantage of giving a universal metric for the relationship between dependent and independent features irrespective of the shape of the underlying distributions (Wu and Zhang, 2004) and has been used in various studies (Ahmed et al., 2019b; Nashwan and Shahid, 2019; Pour et al., 2018; Salman et al., 2018).

The technique has shown to be promising and possessed a similar or more improved skill for performance evaluation and selection of appropriate GCM dataset for hydro-climatic study as compared to other available methods such as compromise programming, wavelet-based skill score (WSS), statistical metrics which sometimes exhibits contradictory results and makes decisions on performance and selection difficult (Ahmed et al., 2019b).

Although most filter based classifiers are known to use single algorithms to integrate variable selection and modelling and often evaluate univariate or very simple interactions between attributes and decisions, which affects the outcomes or performances of features (Kursa, 2016).

However, machine learning approaches are evolving, For example, the wrapper-based Boruta algorithms for feature selection in model building and has been applied in other disciplines such as, Kursa, (2016) showed the relative skill or effectiveness of the methodology in feature selection of random ferns classifier, Ahmed et al., (2021) applied the technique for soil moisture estimation under global warming scenarios.

Advancement in computational capabilities has ensured that machine and deep learning techniques are useful for accurate variable prediction due to their ability to extract, process and handle relatively large amount of complex data with high degree of variable mapping skills and efficiency (Gong et al., 2019).

Numerous studies have successfully implemented different feature selection algorithms such as Information entropy (Shukla et al., 2006), Bayesian weighting (Min and Hense, 2006), Elastic net and ridge regression (Hammami et al., 2012), Artificial Neural Networks (Hajnayeb et al., 2011), Support Vector Machine (Maldonado and Weber, 2009), Random forest (Genuer et al., 2010), Neighbourhood component analysis for regression (Ghimire et al., 2019), iterative input selection algorithm (Prasad et al., 2017) for hydro-climatic study.

Among the techniques, Symmetrical Uncertainty (SU), that is based on information entropy (William et al., 1996) is widely used and has gain the attention of researchers (Ahmed et al., 2019b; Kannan and Ramaraj, 2010; Nashwan and Shahid, 2019; Sa'adi et al., 2020).

However, there are identified shortcomings of the various method such as, inter-dependencies among the models are ignored for a known feature which may result in the

selection of inappropriate GCMs due to overfitting and Some performance indices are based on the state of the climate as a whole and temporal variability is not considered which is critical in model performance assessment (Ahmed et al., 2019b; Reichler and Kim, 2008).

This study employed and pilot the use of Boruta-random forest algorithms (BRF) developed by Kursa et al., (2010) for performance evaluation and selection of appropriate CMIP6 GCM models to examined their individual capability to accurately simulate observed daily historical climate variables in the basins due to its high sensitivity to climate change.

In this case, Lake Chad basin was adopted to examine the robustness of the methodology due in part to its highly skilled variable mapping as a requirement for input parameters required in hydrologic study, uniqueness and high climate variability across the basin of interest and this technique has been used and recommended based on previous study.

For example, Prasad et al., (2019) utilised BRF to predict soil moisture, Christ et al., (2016) applied BRF for industrial big data application in distributed and parallel time series feature extraction, Leutner et al., (2012) predict forest biodiversity and Lyu et al., (2017) applied the concept to forecast air quality, where these algorithms was used to define significant input parameters by comparing the real features or variables to those of random probes and all the studies has suggested a convincing outcome of model accuracy.

However, the authors acknowledged that, this is first attempt to have applied the proposed technique for GCM evaluation and selection and to validate the methodology, an information aggregation approach was adopted to combine the ranks of the GCMs across the grid points in the entire basin to identify the best ensemble of the CMIP6 GCM for simulation of the above variables. the result of this approach will be compared with the well-established symmetrical uncertainty technique to understand the efficacy, applicability and the inherent uncertainties of the model evaluation and selection approach.

Furthermore, earlier studies that utilized simulation based studies to investigate the impact of climate change on freshwater hydrology required data in the form of daily precipitation and temperature series to drive various hydrological watershed models; however, the performance of the model is dependent upon the driving general circulation models, internal parameterizations and model domain configuration (Déqué, 2007).

Therefore, it is important to note that; conclusions drawn from studies that evaluated and select GCM performance at monthly time scale (Abbasian et al., 2019; Ahmed et al., 2019b; Salman et al., 2018) may well not be relied upon for studies that requires climate data at daily time step for hydrologic modelling processes, in order to reduce biases and observational uncertainties for realistic predictions of climate change impact, adaptation and resilience in hazard assessment. The performance of GCMs from BRF, SU and an all-ensemble average approach adopted in this study will be evaluated for a realistic assessment of basin scale features and climate variable dynamics.

Finally, this study intends to propose and provide a robust approach that enhance and preserve climate signals' internal parameterization exerted by re-gridding, downscaling and bias correction to realistically utilize the optimal amount of GCM dataset capable of assessing the complex interactions within the earth system (hydrologic) models.

This is essential for accurate understanding and forecasting of long-term changes of basin hydrology resulting from external forcing which are not adequately addressed in previous research especially in data-sparse regions within an acceptable level of uncertainty.

### **5.3 Case Study Area and Data**

#### **5.3.1 Case study area**

The Lake Chad Basin is one of the world's largest endorheic basins, with an estimated area of ~2,500,000 km<sup>2</sup>, (Coe and Foley, 2001; Gao et al., 2011). It is situated at latitudes of

5.2° - 25.3° N and longitudes of 6.9° - 24.5° E, in the transition zone between the Sahara and the Sudano-Sahelian regions of West Africa (Ndehedehe et al., 2018), (**Figure 5.1**).

The basin provides the primary source of freshwater for livestock grazing, agricultural production, and fish farming (Buma et al., 2016). The basin is divided into four climatic zones namely: the Saharan, Sahelo-Saharan, Sahelo-Sudanian, and the Sudano-Guinean zone, with an annual precipitation range of, < 100 - 1500 mm respectively.

The basin's average annual temperature is between 35°C to 40°C in the northern part to as low as 26.5°C in the southern part of the basin (Nkiaka et al., 2018b) characterised by hot, wet and dry weather condition from March to June, from June to October, and from November to February, respectively (Mahmood et al., 2019).

The basin is situated in an area with minimal relief, no surface outflow, and a spatial extent that is highly vulnerable to climate change, with elevations ranging from -330 m to 3446 m (**Figure 5.1**), However, according to (Coz et al., 2009; Nkiaka et al., 2018a), With the exception of few isolated hills, plateaus, and mountains in the southern and northern regions, the basin is generally flat ground with an average slope of < 1.3%.

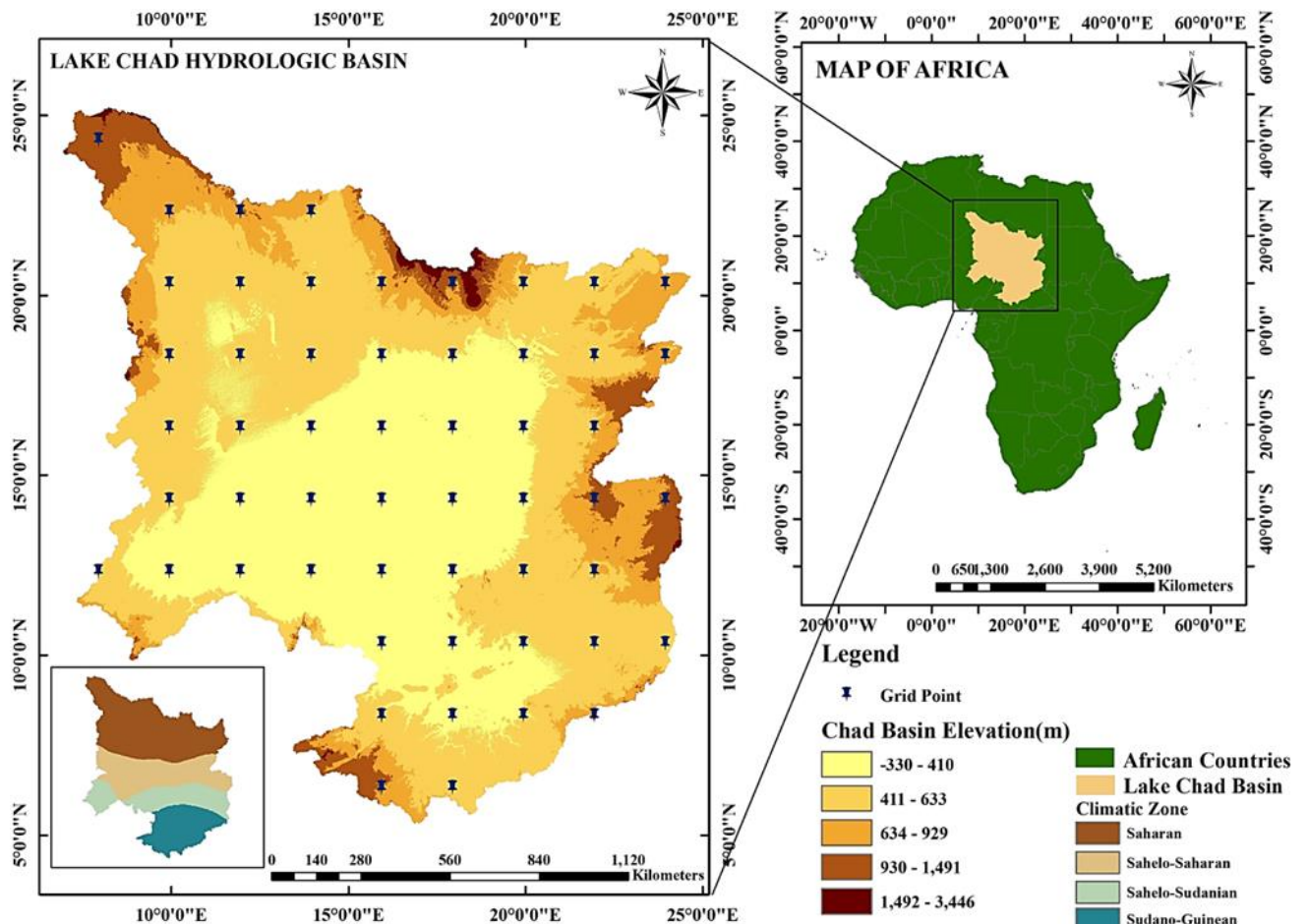


Figure 5.1: Lake Chad basin showing elevation, climate stations and climatic zones for the proposed study.

### 5.3.2 Gridded data and sources

Several gridded climate data sources were explored and analysed for suitable application in climate studies in the Chad Basin due to inadequate gauged based meteorological data in most part of the world Sub-Saharan Africa and the Mediterranean, in particular.

This study utilized the daily gridded precipitation data of the US Climate Prediction Centre (CPC), optimal interpolation of station or gauged based records of GTS (Xie et al., 2007) and daily maximum and minimum temperature data of the Princeton University Global Meteorological Forcing PGF v.2 from forcing's of NCEP-NCAR reanalysis and other global data by bilinear interpolation (Sheffield et al., 2006), from 1979 – 2012 and available at <https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globalprecip.html> and <http://hydrology.princeton.edu/data.pgf.php> respectively, as recommended by Lawal et al., (2021) to represent adequately the true climatic features of the selected basin.



### 5.3.3 General circulation model data and sources

In this study 16 coupled inter-comparison project phase 6 (CMIP6) general circulation models at daily time resolution were considered. These are available at <https://esgf-node.llnl.gov/projects/esgf-llnl/CMIP6> for the period 1979 – 2012 and consistent with the gridded data time scale for realistic performance evaluation. Details of the GCMs name, spatial resolutions, models' development centres and country of origin are provided in **Table 5.1.**, below.

The first ensemble members of the GCM were adopted for fair assessment. The models were chosen because they have the requisite historical and future climate change emissions scenario data (SSP1 – SSP5) at a daily timestep as an important requirement for a hydrologic modelling study.

This is essential for understanding watershed projected hazards due to climate change and provide valuable information to policy makers for informed decision on integrated water resource management.

Table 5.1: Summary of CMIP6 models considered in this study.

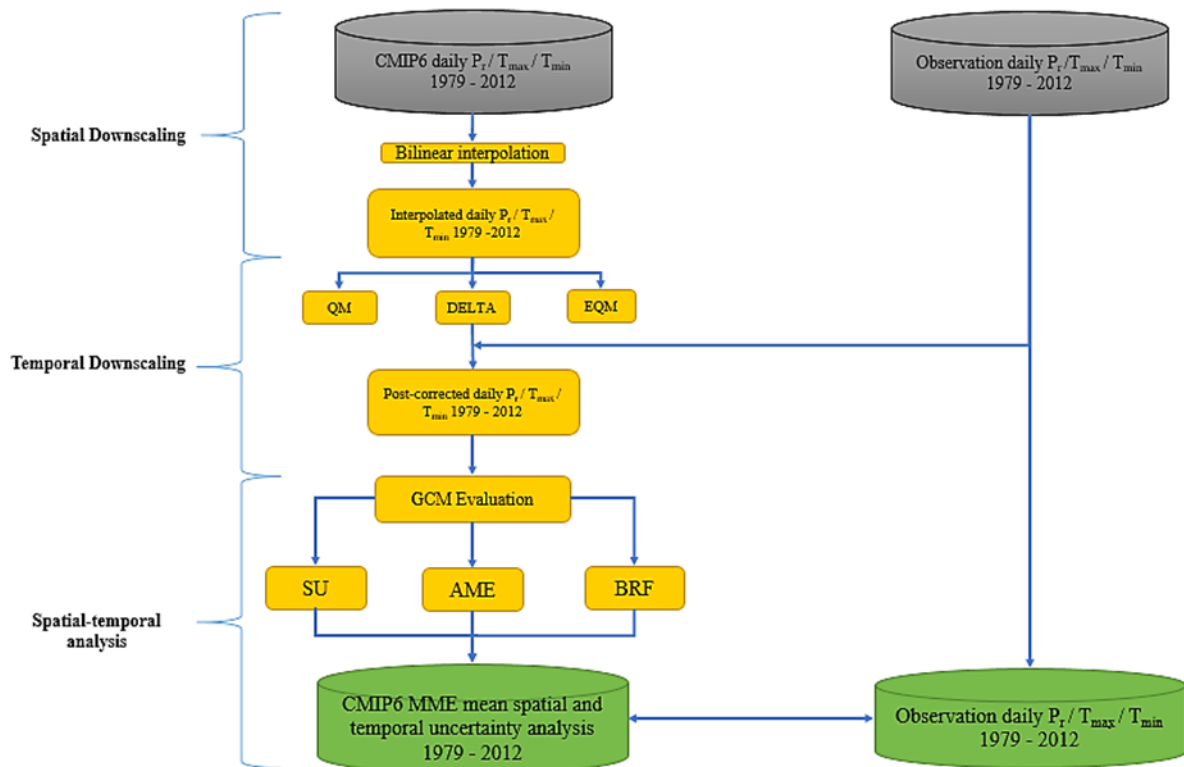
Country	Modelling Centre	Model Name	Resolution in arc Deg. (Lon × Lat)	Model Number
<b>Australia</b>	Commonwealth Scientific and Industrial Research Organisation; ARCCSS (Australian Research Council Centre of Excellence for Climate System Science)	ACCESS CM2	1.875×1.25	M1
<b>Australia</b>	Commonwealth Scientific and Industrial Research Organisation; ARCCSS (Australian Research Council Centre of Excellence for Climate System Science)	ACCESS ESM1-5	1.875×1.25	M2
<b>China</b>	Beijing Climate Centre	BCC CSM2-MR	1.10×1.10	M3
<b>Canada</b>	Canadian Centre for Climate Modelling and Analysis	CanESM5	2.81×2.81	M4
<b>Sweden</b>	EC-Earth consortium	EC-Earth3	0.70×0.70	M5
<b>Sweden</b>	EC-Earth consortium	EC-Earth3-Veg	0.70×0.70	M6
<b>China</b>	Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) modelling (Chinese Academy of Sciences)	F-Goals-g3	2.81×2.78	M7
<b>China</b>	Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) modelling (Chinese Academy of Sciences)	GFDL-ESM4	1.25×1.00	M8
<b>Russia</b>	Institute of Numerical Mathematics of the Russian Academy of Sciences	INM-CM4.8	1.12×1.12	M9
<b>Russia</b>	Institute of Numerical Mathematics of the Russian Academy of Sciences	INM-CM5.0	1.87×1.25	M10
<b>France</b>	Institut Pierre-Simon Laplace	IPSL-CM6A-LR	2.50×1.27	M11
<b>Japan</b>	Centre for Climate System Research; Japan Agency for Marine-Earth Science and Technology; National Institute for Environmental Studies	MIROC6	1.40×1.40	M12
<b>Germany</b>	Max-Planck-Institut für Meteorologie	MPI-ESM1.2-LR	1.875×1.875	M13
<b>Japan</b>	Meteorological Research Institute (MRI)	MRI-ESM2.0	1.125×1.125	M14
<b>Norway</b>	Norwegian Climate Consortium	NorESM2-LM	2.50×1.875	M15
<b>Norway</b>	Norwegian Climate Consortium	NorESM2-MM	1.25×0.9375	M16

(Source: <https://esgf-node.llnl.gov/projects/esgf-llnl/CMIP6> )

## 5.4 Research Methodology

### 5.4.1 Methodology flow chart

The general procedure used for the evaluation of GCMs using gridded CPC and PGF climate data as surrogate for observed, selection, and further analysis of GCM ensemble is outlined in the methodology flow chart as shown in **Figure 5.2**.



*Figure 5.2: GCM evaluation and selection for hydro-climatic study.*

### 5.4.2 Statistical downscaling and bias correction of GCM

A multi-point statistical downscaling and bias correction approach was considered. The CMIP6 daily precipitation, maximum temperature and minimum temperature GCM models were interpolated to a common  $2^\circ \times 2^\circ$  grid and spatially downscaled using bilinear interpolation approach for smooth transformation of coarse to fine resolution as recommended by Fischer et al., (2014), resulting into 54 grid point (**Figure 5.1**), that covered the entire study area.

The deviation of the interpolated climate data was corrected to improve the GCM models agreement with the observed data (CPC precipitation and PGF temperature), and to provide enhancement in model output because direct GCM output cannot be relied upon for accurate assessment of watershed climate features at local and regional scale required for impact studies due to their coarse resolution.

This is achieved by combining features of local observation and simulations resulting in insignificant biases and higher resolution climate projections using three well known bias correction approaches and the detailed methodologies can be found in the cited references namely delta change (Navarro-Racines et al., 2020), Quantile mapping (Cannon et al., 2015), and Empirical quantile mapping technique (Ghimire et al., 2019).

The grid based bias correction performance was evaluated using three statistical metrics e.g. correlation coefficient, mean bias error and index of agreement to understand the limitations of the various techniques to prevent misuse in selecting the best suited output relative to other method for further analysis and their detailed methodology can be found in (Taylor, 1997; Willmott, 1981; Willmott and Matsuura, 2005).

#### **5.4.3 Performance evaluation and selection of appropriate GCM output by Machine learning technique**

The daily bias corrected CMIP6 GCM outputs adopted were evaluated by two machine learning approaches to discover the significant features of the GCMs relative to the observation data. This section gives a brief overview of the machine learning based symmetric uncertainty and Boruta random forest algorithms for performance assessment and ensemble projections.

#### 5.4.3.1 Symmetric Uncertainty (SU)

Symmetric uncertainty was developed by William et al., (1996), and an entropy based filter that assess pair-wise similarity between dependent and independent attribute irrespective of their probability distribution and interdependency (Wu and Zhang, 2004). It measures the information gain of the response random variable relative to the predictor and the lesser the entropy the greater the association of the data.

The bias in the measurement is corrected by dividing the information gain with the sum of entropies of the random features of the observation and GCMs precipitation, maximum and minimum temperature respectively. The SU was estimated according to Equation (5.1). Detailed methodology of the approach can be found in (Ahmed et al., 2019b; Lawal et al., 2021) (See 254).

$$SU(x, y) = 2 \times \frac{MI(x, y)}{H(x) + H(y)} \quad (5.1)$$

In Equation (5.1),  $MI(x, y)$  is the mutual information gain,  $H(x)$  and  $H(y)$  represent the entropies of the GCM and observation respectively. Symmetric uncertainty values range from 0 to 1 indicating no similarity to perfect similarity respectively.

#### 5.4.3.2 Boruta random forest algorithm (BRF)

Boruta feature selection was developed based on random forest algorithm (Breiman, 2001). The approach was introduced by Stoppiglia et al., (2003) to identify significant input parameters from a host of many dependent features to match the attributes of an independent feature.

The algorithm computes the Z-score of all input predictors and the distribution determines the shadow features as well as the important variables of the predictors based on the Z-score metrics (Kursa and Rudnicki, 2010). First, the 16 GCM (predictors) and observation (target)

variables were divided into 70% and 30% as training and test datasets. This is necessary because the accuracy of the prediction depends on multiple runs for convergence. Also, there are situations where unimportant attributes may achieve a higher Z score than the shadow attributes. Therefore, multiple random forest runs were executed to arrive at a significant decision.

The methodology adopted are shown below and the source code is available in **254**.

- A duplicate or shadow random variable,  $q'_t$  for a particular vector,  $q_t$  to increase randomness and correlation between predictors and target variable ( $P_t$ ), for all group of discrete inputs ( $q_t \in R^n$ ), T and target variables ( $P_t \in R$ ) for all inputs (n) and  $t = 1, 2, 3, \dots \dots \dots T$ . The target variable is evaluated for association using the shadow variable  $q'_t$  and actual variable  $q_t$ .
- The variable importance measures or scores (mean decrease accuracy) for each predictor  $q_t$  and shadow  $q'_t$  attributes were computed for 500 iterations based on equation (5.2)., to achieve accurate predictions (Hur et al., 2017).

$$VIM = \frac{1}{m_{tree}} \sum_{m=1}^{m_{tree}} \frac{\sum_{t \in OOB} I(P_t = f(q_t)) - \sum_{t \in OOB} I(P_t = f(q^n_t))}{|OOB|} \quad (5.2)$$

$I(\bullet)$  is an indication function, OOB is out-of-bag predicted error in the training samples;  $P_t = f(q_t)$  and  $P_t = f(q^n_t)$  are predicted values before and after permutation.

- The standard score (Z-score) of the predictor and shadow attributes relative to the observation are computed as:

$$Z - score = \frac{VIM}{\sigma} \quad (5.3)$$

Where,  $\sigma$  is the measured standard deviation of accuracy losses, the minimum, median and maximum Z-score of the shadow features is computed and analysed relative to the predictor variable importance distribution of all dependent features. When all input features are confirmed to be important, or the iteration limit is achieved, the algorithm ends.

- The attribute is deemed important for a set iteration, if its variable importance score is higher than the maximum importance score of the shadow attributes.

To check the efficacy of the feature selection approach, a cross-validation procedure was run on the remaining (30%) test data to obtain a classification error and a repeat BRF procedure was run on the selected attributes from the group ranking approach.

The OOB error and model accuracy for the two procedure was 0.231 and 0.130 and 0.7692 and 0.88 respectively. This implied that consecutive reduction of features that are weakly important can eliminate noise and increase prediction accuracy.

#### **5.4.4 Performance evaluation of selected GCM ensemble mean.**

The weight or performance score of the two machine learning technique was used to rank the bias corrected precipitation, maximum and minimum temperature at individual grid point and the ranking were aggregated based on multi-criteria rating metrics using numerical averaging as recommended by Raju and Kumar, (2016) for the entire study area.

Multi-model ensemble mean of the GCMs was developed by selecting four best GCM relative to the observation based on the two machine learning approaches (BRF and SU) and that of the 16 GCMs here referred as all model ensembles mean (AME) to further evaluate spatial and temporal changes in annual precipitation, maximum and minimum temperature changes, and extreme conditions of the basin climate.

Analysis of individual model bias as seen in other studies may not necessarily translate to accurate performance. Here, the study focused on the ensemble mean of the GCM to

effectively understand the uncertainty (spread) of the combined model ensemble mean features, which is important and sensitive in their usage for hydrologic impact studies and reliable future projections of water resources.

#### **5.4.5 Validation of evaluation approach based on SPEI drought and flood hazard assessment.**

The evaluation approaches considered i.e., BRF, SU and AME were further validated by examining the influence of the ensemble mean GCM precipitation, maximum and minimum temperature on the severity of drought relative to observation across the four climatic zone of the basin by estimating the 12-monthly standardized precipitation evapotranspiration index (SPEI) for the period of observation (1979 – 2012). The estimated SPEI time series were further used to assess the temporal pattern in the trend and their significance by Mann Kendall trend test and Sen's slope estimator at 95% level of confidence.

#### **5.4.6 Standardized precipitation evapotranspiration index**

Standardized precipitation evapotranspiration index is a phenomenon that depicts water surplus or deficit within long climatic time scale that the difference between precipitation and potential evapotranspiration is calculated and then fitted with probability density function to estimate the severity and frequency of flood or droughts in a catchment.

The severity was classified as  $SPEI \geq 2.0$ , 1.5 to 1.99, 1.0 to 1.49, 0.99 to  $-0.99$ ,  $-1.0$  to  $-1.49$ ,  $-1.5$  to  $-1.99$  and  $SPEI \leq -2.0$  to indicate extremely wet, severely wet, moderately wet, normal, moderate drought, severe drought, and extreme drought event respectively (Shekhar and Shapiro, 2019).

Validation by SPEI captures the impact of temperature increase on water demand and natural variability of climate and its influence on climate change studies. Further evaluation to observe the trend and its significance was based on the methodology by Henry, (1945); Kendall, (1948) and Sen, (1968) given in the equations below.



The Mann-Kendall statistics is:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (5.4)$$

For a n number of time series,  $x_j$  and  $x_k$  are consecutive data values. The series sgn is as follows:

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{if } x_j > x_k \\ 0 & \text{if } x_j = x_k \\ -1 & \text{if } x_j < x_k \end{cases} \quad (5.5)$$

The mean  $E(S)$ , Variance  $V(S)$  and the Z statistics is evaluated as:

$$E(S) = 0 \quad (5.6)$$

$$V(S) = \frac{1}{18} \left\{ n(n-1)(2n+5) - \sum_{i=1}^p t_i(t_i-1)(2t_i+5) \right\} \quad (5.7)$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{for } S < 0 \end{cases} \quad (5.8)$$

Because the nonparametric Sen's slope estimator is resilient against outliers in time series analysis, it was used to calculate the size of the discovered trends in the time series data:

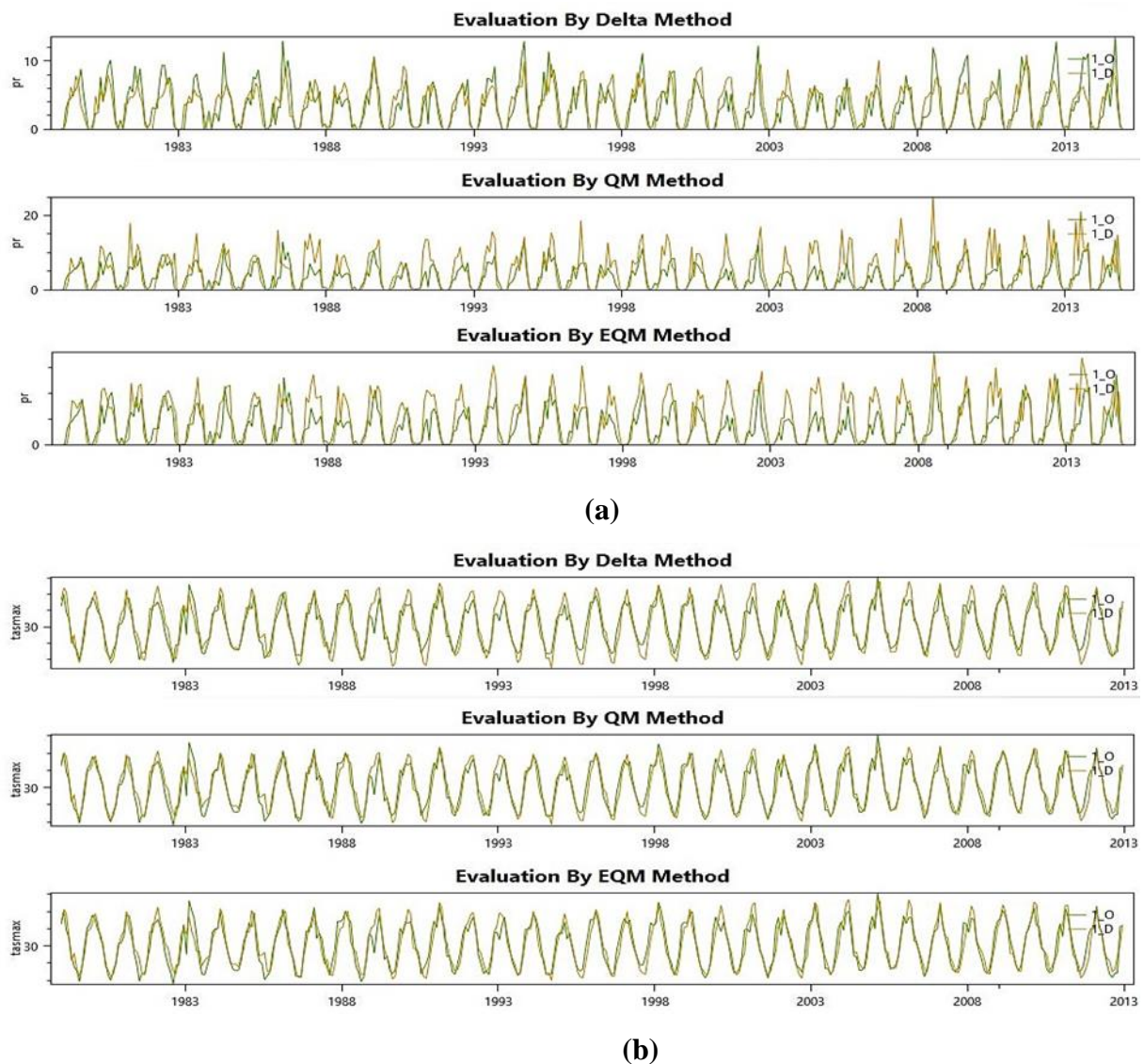
$$SS = \text{median} \left[ \frac{x_j - x_i}{j - i} \right] \quad \text{for all } i < j \quad (5.9)$$

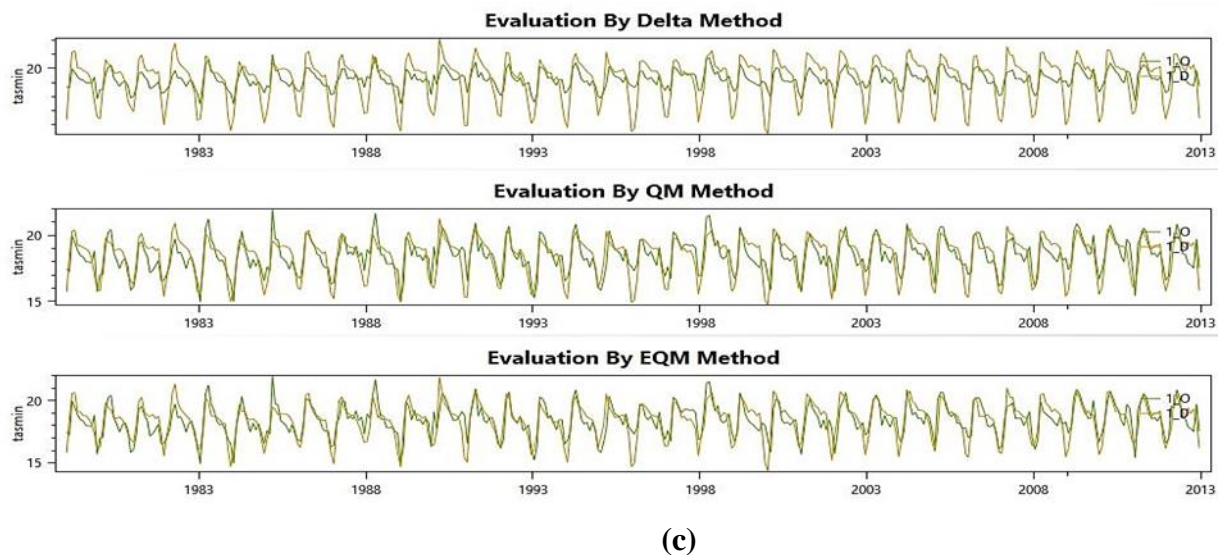
Where  $x_i$  represents the data value at a time step  $i$  and  $x_j$  represents the data value at time step  $j$ .

## 5.5 Results and Discussion

### 5.5.1 Spatial and temporal downscaling and bias correction of CMIP6 GCMs

Figure 5.3 shows the example of post-corrected precipitation and maximum and minimum temperature data of GCM output ACCESS CM2 relative to the observations at grid point 1 (15.95°, 6.31°); for 1979-2014 and 1979-2012 respectively, using delta change, quantile mapping and empirical quantile mapping technique.





**Figure 5.3:** Plot of bias-corrected GCMs and observed climate data using delta change, quantile mapping and empirical quantile mapping method in the Lake Chad basin. (a): variation of GCM precipitation relative to observed CPC data. (b): Variation of GCM maximum temperature relative to PGF data. (c) Variation of GCM minimum temperature relative to observed PGF data.

The evaluation results across the 54 grid point that covers the study area of the downscaled and the daily bias corrected GCM outputs of precipitation, maximum and minimum temperature indicated that delta change and empirical quantile mapping method is the most suitable for daily precipitation and maximum and minimum temperature with a recorded mean bias error,  $MBE = 0$ , mean correlation coefficient  $R^2 = 0.8$  and modified index of agreement,  $md = 0.86$  at 94% and a  $MBE$  range of  $-0.01 - 0$ , mean correlation coefficient  $R^2 = 0.92$ , and modified index of agreement  $md = 0.96$  respectively, relative to other evaluation method across all the grid points.

However, the results according to **Figure 5.3c**, there are difficulties of downscaling models to capture peak values of minimum temperature across all evaluation methods, especially in the Sahelo-Saharan zone but this might be attributed to scale gap between GCM outputs and observations, and such variations cannot be accounted for by the GCM outputs as corroborated in (Sachindra et al., 2014).

Hence, GCM outputs from this evaluation technique were adopted for the next phase of analysis. The evaluation was limited to statistical downscaling because it was considered to

be more flexible than dynamical technique and its projections can be downscaled to point specific locations as corroborated in Martinez-García et al., (2021).

The merit of this approach is to provide a reference for evaluating the accuracy of precipitation and temperature, which are the two critical hydrologic model input parameters, to improve the predictions of future river basin hydrologic cycles. The bias correction technique performed efficiently in simulating the observation in the Guinean-Sudanian zone, with a cluster of grid points with almost perfect correlation relative to the observation. However, Saharan zone exhibit some inadequacies in replicating the observed climate features which may be due in part to poor climate signals of sparse precipitation events.

In general, the result indicates some improvements of the CMIP6 GCM in capturing the observation signals of the climate variables and has proven to restore the inter-station dependencies. The multi-site approach illustrated the capability in addressing the inequities of transitioning and interpolation of GCMs from coarse to finer resolution and vice versa to accurately reproduce observed multiple statistical properties of climate variable for improved hydrological climate change impact studies.

### **5.5.2 Performance evaluation and selection of CMIP6 GCMs**

The individual models were coded  $M_1, M_2, \dots, M_{16}$  to represent ACCESS CM2, ACCESS ESM1-5, ..... NorESM2-MM respectively, as appeared in table 5.1., for ease of identification. The performance of the model across individual grid point was analysed based on the two machine learning approaches SU and BRP to investigate their significance in replicating the basin climatic features of precipitation and maximum and minimum temperature for the period 1979 – 2014 and 1979 – 2012 respectively, which was a compromise and limitations between CPC, PGF and historical CMIP6 GCM based on world meteorological organisation climatological standard normal period of data range and availability.

The GCMs were ranked based on their symmetric uncertainty (SU) coefficient (**5.5.2.1**) and variable importance score (**5.5.2.2**) at each grid point for all the climate variable to understand the degree of association or relative skill (importance) of the models with insignificant bias across the basin respectively. However, in BRF analysis, a zero score was recorded for GCMs whose variable importance score is less than that of the shadow attributes at any grid before aggregation and ranking for consistency.

#### **5.5.2.1 Evaluation based on symmetric uncertainty.**

The result of symmetric uncertainty for all the climate variables were aggregated across the 54 grid points considered, which was a fair representation of the basin as shown in **Table 5.2**. The rankings based on the coefficients indicated that GFDL-ESM4, MIROC6, INM-CM4.8, ACCESS ESM1-5, and MRI-ESM2.0 exhibit an improved skill for precipitation with coefficient in the range of 0.94 – 0.81, while MPI-ESM1.2-LR, INM-CM4.8, EC-Earth3-Veg, and EC-Earth3 are relatively effectively in simulating maximum temperature with coefficient in the range of 0.98 – 0.92 and finally, MPI-ESM1.2-LR, MIROC6, NorESM2-MM and CanESM5 are relatively suitable in simulating minimum temperature with a recorded coefficients in the range of 0.84 – 0.85. However, all the CMIP6 GCMs exhibit an above average skill except for F-Goals-g3 which was quite poor for precipitation.

Varied level of GCM performance were noticed which further complicates evaluation and has been acknowledged by McMahon et al., (2015), that GCMs have strength and weaknesses in simulating different climate variables.

Simulation result of CMIP6 GCM by symmetric uncertainty exhibit some improvement in contrast to earlier phase as in Ahmed et al., (2019b), which may be attributed to differences in timescale and timesteps of the chosen GCMs, improvement in model parameterizations and development and quality of observation data. This is corroborated in studies by (Ayugi et al., 2021; Grose et al., 2020; Wang et al., 2021).

**Table 5.2:** Summary of GCM SU coefficients of precipitation, maximum and minimum temperature relative to observation data

<b>SYMMETRIC UNCERTAINTY</b>					
<b>Precipitation</b>		<b>Maximum Temperature</b>		<b>Minimum Temperature</b>	
<b>Model</b>	<b>SC</b>	<b>Model</b>	<b>SC</b>	<b>Model</b>	<b>SC</b>
<b>M8</b>	<b>0.94</b>	<b>M13</b>	<b>0.98</b>	<b>M13</b>	<b>0.85</b>
<b>M12</b>	<b>0.88</b>	<b>M9</b>	<b>0.94</b>	<b>M12</b>	<b>0.85</b>
<b>M9</b>	<b>0.81</b>	<b>M6</b>	<b>0.93</b>	<b>M16</b>	<b>0.84</b>
<b>M2</b>	<b>0.81</b>	<b>M5</b>	<b>0.92</b>	<b>M4</b>	<b>0.84</b>
<b>M14</b>	<b>0.81</b>	M11	0.90	M6	0.83
M16	0.74	M16	0.90	M8	0.82
M13	0.72	M2	0.90	M5	0.81
M4	0.72	M12	0.89	M11	0.80
M3	0.71	M8	0.85	M9	0.80
M15	0.71	M4	0.83	M2	0.79
M11	0.65	M10	0.79	M15	0.77
M1	0.65	M14	0.78	M1	0.76
M10	0.65	M1	0.78	M14	0.75
M6	0.65	M15	0.76	M10	0.74
M5	0.62	M7	0.72	M7	0.73
M7	0.34	M3	0.70	M3	0.72

### 5.5.2.2 Evaluation based on Boruta random forest algorithm.

The performance of the GCM models were assessed based on its ability to iteratively identify the importance of the original attributes (CMIP6 GCMs) with their randomised sets (shadow attributes) to truly replicate the observation data. The simulation and importance measure were generated, and the variables were ranked as shown in **Figure 5.4** for grid 1.

The ranking indicates that model M9, M13, M10, M14, M4, M11 and M7 are quite important in simulating GCM daily precipitation with variable importance score in the range of 4 – 22, while others whose important score are below that of the maximum importance score of the shadow attributes are considered poor in simulating the observed (CPC) daily precipitation. However, all GCMs exhibit a good skill in simulating maximum and minimum temperature with a significant difference in variable importance score in the range of 18.0 54.2 and 19.1 – 44.6 respectively.

The procedure was repeated across all grid points and attributes were filtered or rejected with a performance score below the shadow attributes. The variable importance score was

aggregated across all points and were ranked based on descending order of mean importance score for all the variables as shown in **Table 5.3**.

The evaluation of the climate variables by the novel Boruta random forest revealed a consistent performance of the GCMs across the grid points which makes the technique quite efficient in the ranking process and necessary for holistic assessment where the minimal optimal set of GCMs might be more useful rather than the application of the entire set of available models which may be computationally intensive, require more resources and time and decrease model proficiency.

*Table 5.3: Summary of GCM importance score of precipitation, maximum and minimum temperature relative to observation data*

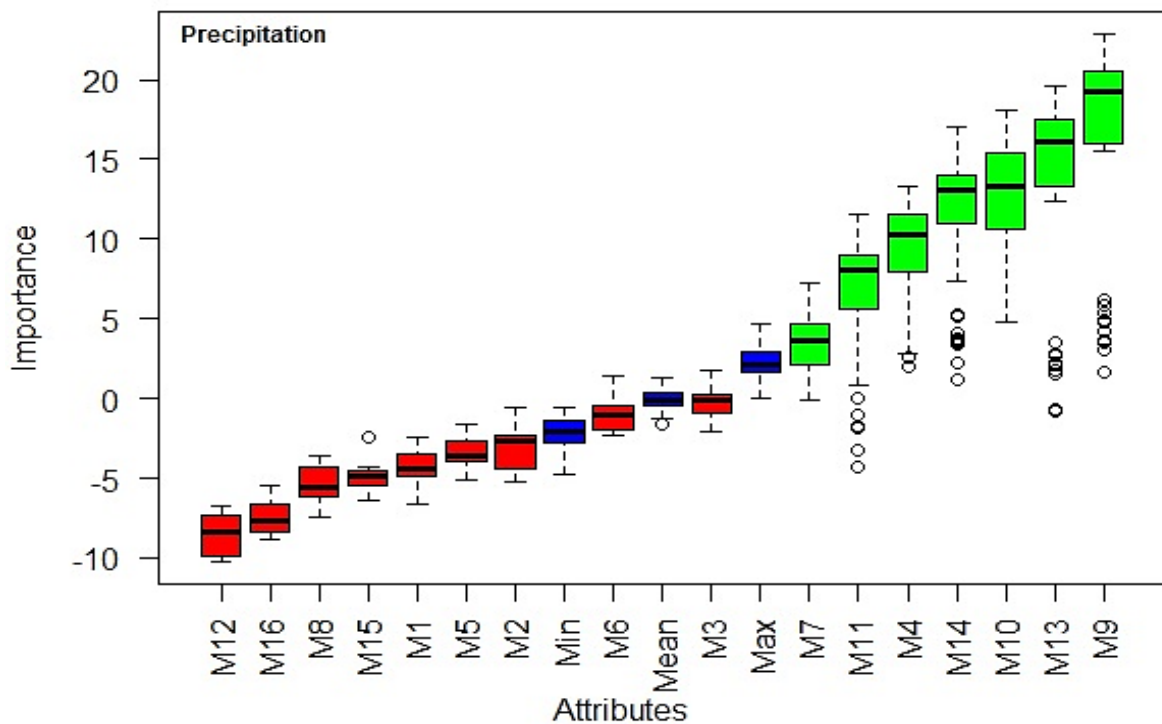
<b>BORUTA ALGORITHMS</b>					
<b>Precipitation</b>		<b>Maximum Temperature</b>		<b>Minimum Temperature</b>	
<b>Model</b>	<b>IS</b>	<b>Model</b>	<b>IS</b>	<b>Model</b>	<b>IS</b>
<b>M9</b>	<b>4.68</b>	<b>M9</b>	<b>41.07</b>	<b>M10</b>	<b>32.03</b>
<b>M14</b>	<b>3.20</b>	<b>M13</b>	<b>39.44</b>	<b>M9</b>	<b>28.96</b>
<b>M10</b>	<b>2.88</b>	<b>M10</b>	<b>36.81</b>	<b>M13</b>	<b>28.86</b>
<b>M7</b>	<b>2.82</b>	<b>M12</b>	<b>34.10</b>	<b>M14</b>	<b>27.32</b>
M4	2.68	M14	33.43	M16	27.18
M8	2.53	M2	32.25	M5	26.57
M13	2.12	M11	32.25	M7	26.50
M6	1.81	M7	31.36	M15	26.24
M12	1.33	M6	30.93	M6	26.17
M5	1.29	M16	30.87	M2	25.82
M1	1.20	M1	30.79	M1	25.52
M2	0.96	M3	29.89	M4	25.00
M11	0.75	M8	29.66	M11	24.92
M3	0.72	M5	29.62	M12	24.79
M16	0.31	M4	29.59	M8	24.53
M15	0.10	M15	27.56	M3	24.18

The rankings based on Boruta algorithm indicated the difficulty of a single GCM to reliably simulate the daily observed precipitation across all grid points satisfactorily, although some GCM have a relatively improved performance and are quite consistent across the grid points for maximum and minimum temperature.

**Figure 5.5a-c** showed the spatial spread of the GCM performance and the aggregated importance score in **Table 5.3**, indicated that INM-CM4.8, MRI-ESM2.0, INM-CM5.0 and

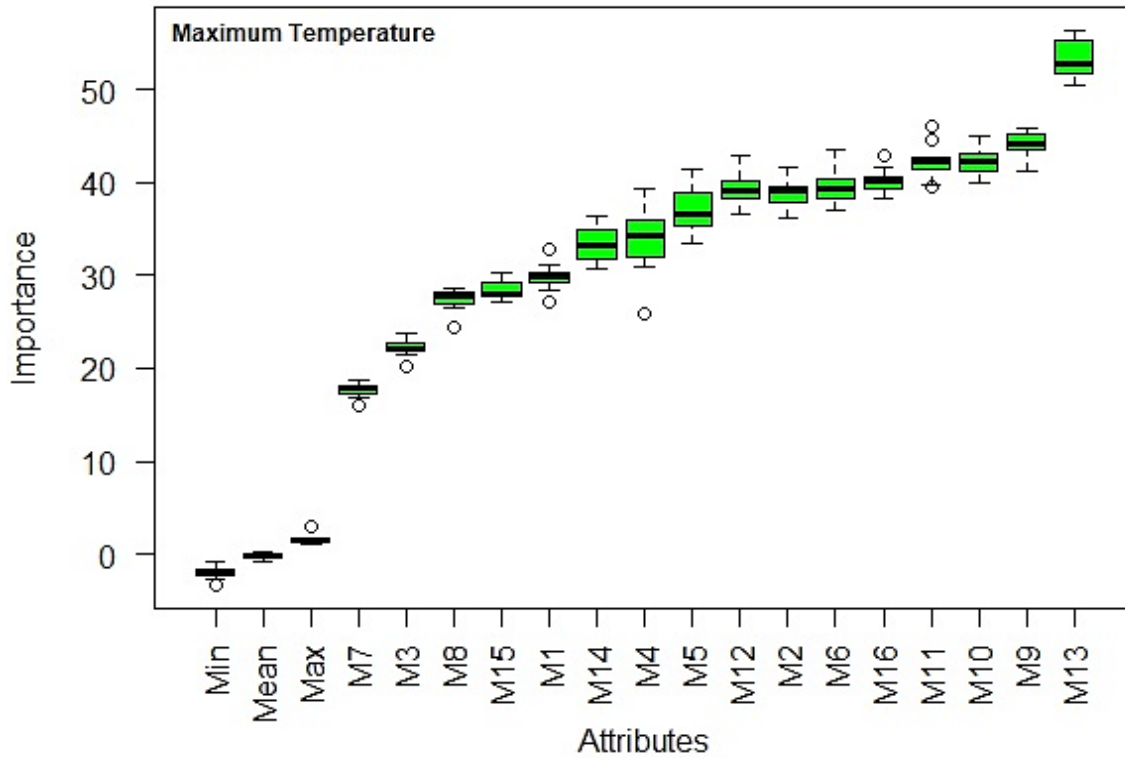
F-Goals-g3 exhibit a relatively suitable skill in replicating observed precipitation, while INM-CM4.8, MPI-ESM1.2-LR, INM-CM5.0 and MIROC6 for maximum temperature and INM-CM5.0, INM-CM4.8, MPI-ESM1.2-LR and MRI-ESM2.0 for minimum temperature respectively.

GCM Models evaluated across the grid points indicated that a few are quite consistent in replicating the observation of the three climate variables; for example, INM-CM4.8 and INM-CM5.0 have shown to exhibit a statistically significant skill in simulating the observed daily precipitation at 24.1% and 18.52% of the grid points, while INM-CM4.8 and MPI-ESM1.2-LR at 35.19% and 16.67% for daily maximum temperature and finally INM-CM5.0 and MPI-ESM1.2-LR at 53.70% and 25.93% for daily minimum temperature respectively.

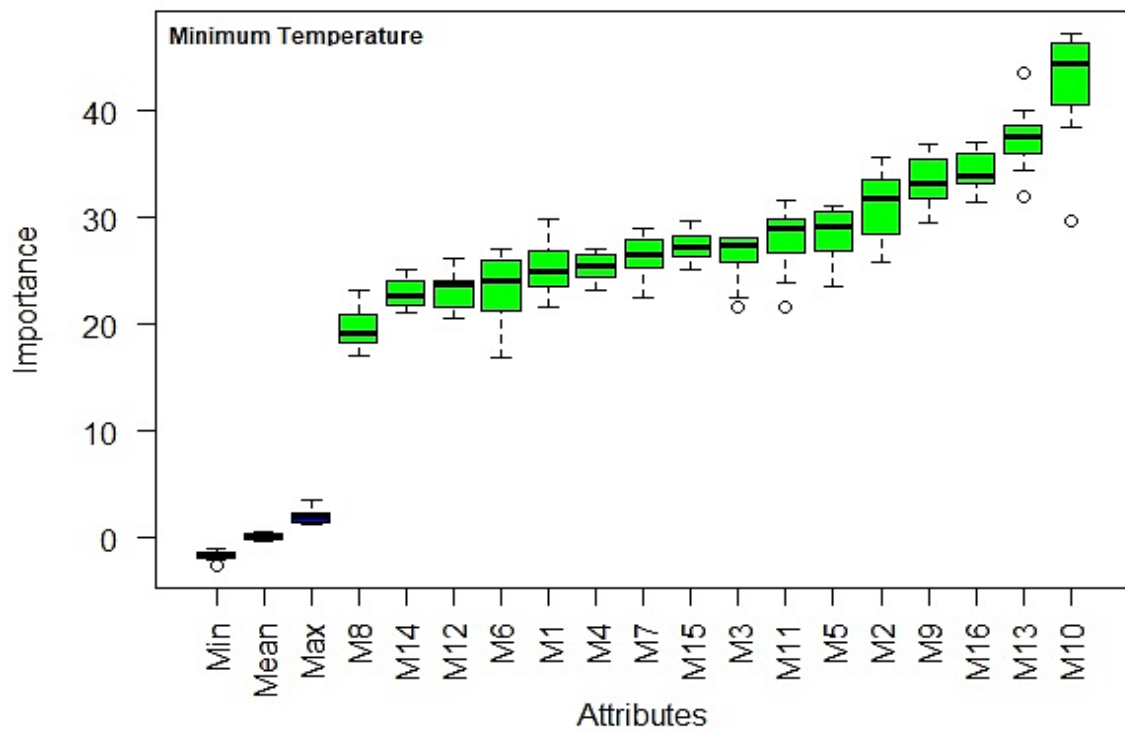


(a)



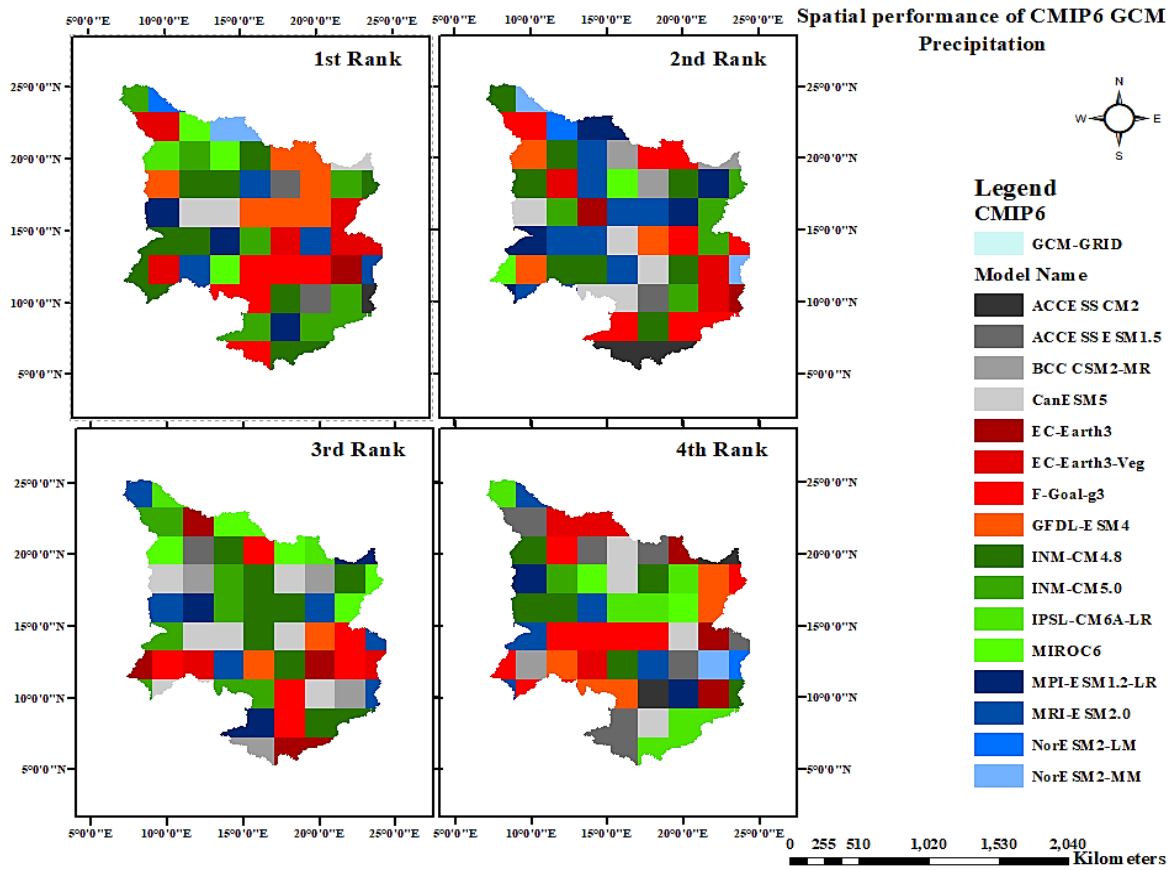


(b)

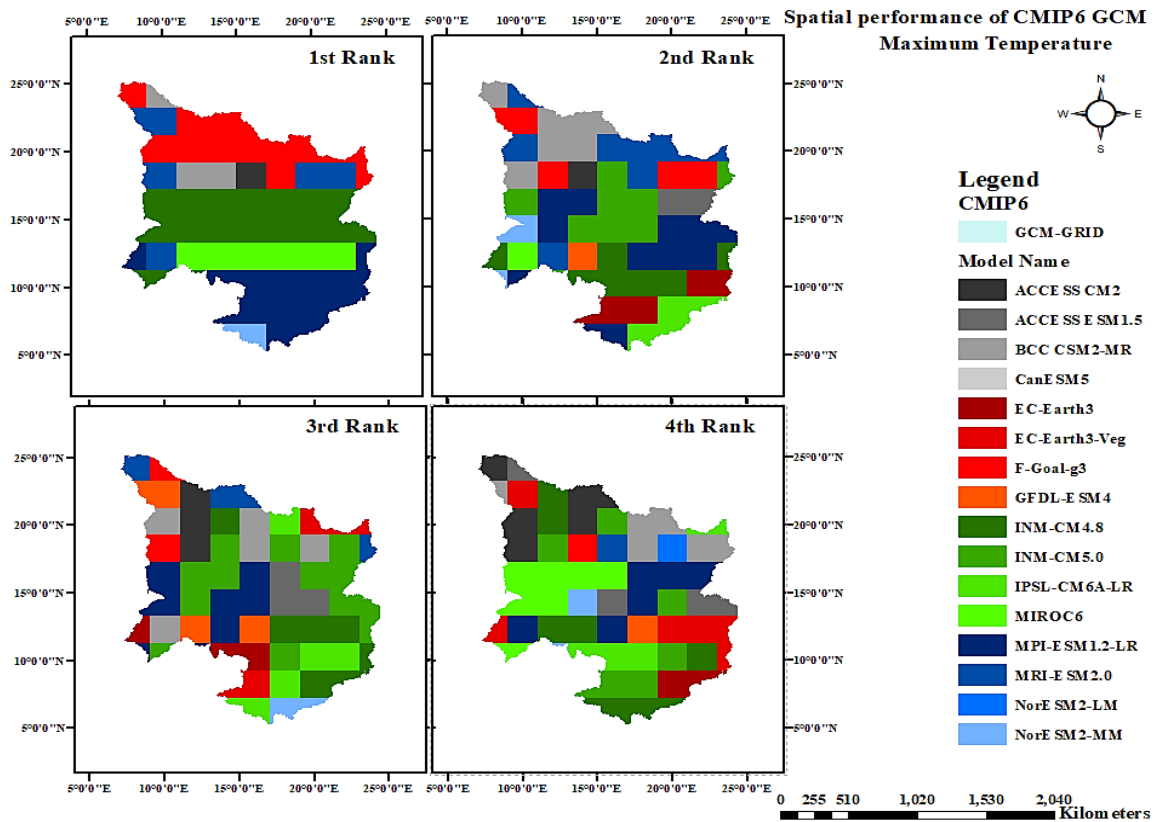


(c)

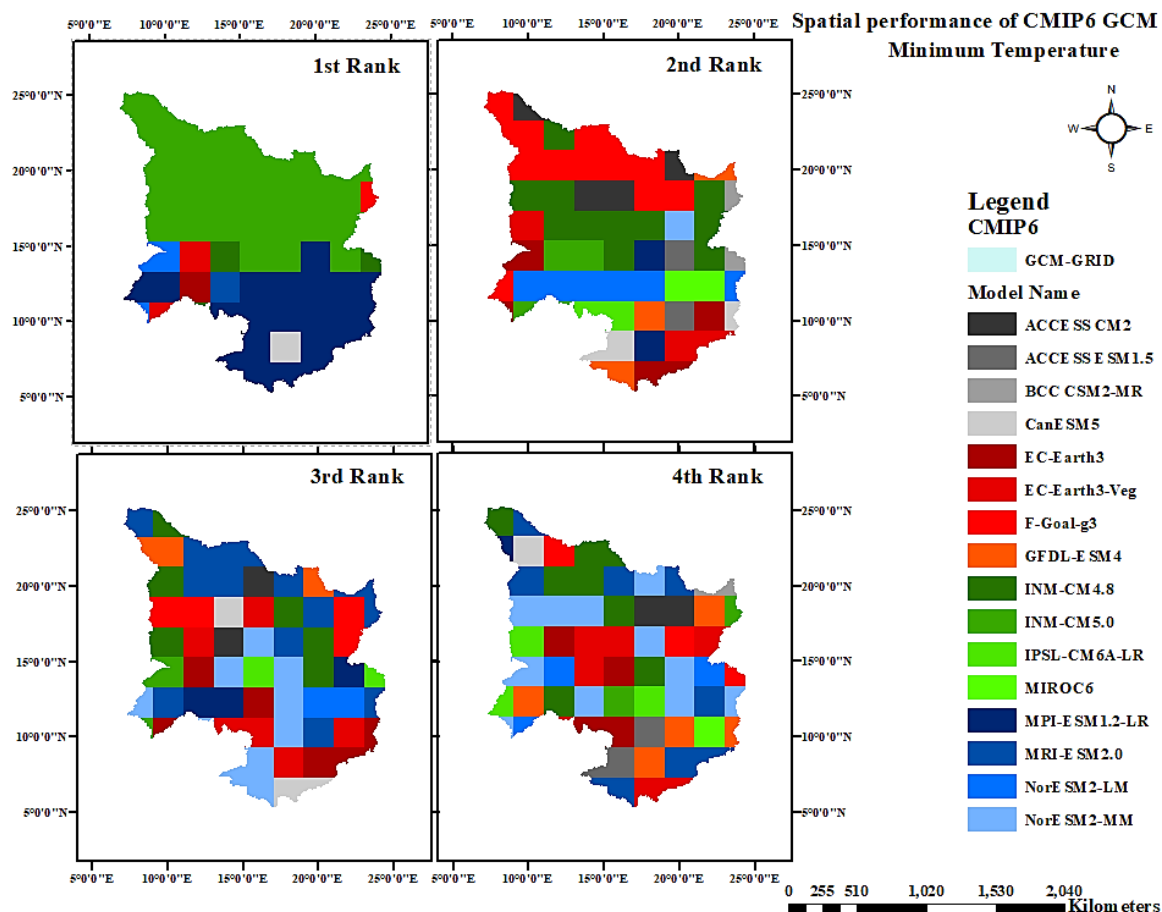
**Figure 5.4:** Box plot of variable importance score of GCMs relative to observed climate data using BRF. (a): relative importance of GCM precipitation to observed CPC data. (b): relative importance of GCM maximum temperature to observed PGF data. (c): relative importance of GCM minimum temperature to observed PGF data.



(a)



(b)



**Figure 5.5:** Spatial spread of GCMs performance relative to observed climate data using BRF. (a): ranking of spatial spread GCM precipitation relative to observed CPC data. (b): ranking of spatial spread GCM maximum temperature relative to observed PGF data. (c): ranking of spatial spread GCM minimum temperature relative to observed PGF data.

### 5.5.3 Identification and evaluation of multi-model ensemble mean of GCMs.

The GCMs evaluated by the approaches considered were a precursor in understanding their skills necessary to match observations. However previous studies, for example Kim et al., (2016) have shown that uncertainties in climate projection can be reduced by identifying and adopting GCMs with improved performance for impact assessment studies.

Earlier literature such as Weigel et al., (2010) and Miao et al., (2012) recommended the use of a collection of GCMs ensemble mean to optimize reliability in prediction and minimize uncertainty in climate variable assessment.

In this study, four best GCM were selected after re-aggregation (**Table 5.4**), due in part to the significant difference observed in the aggregate importance score value between model M14 and M12 in Boruta random forest evaluation to form the multi-model ensemble mean herein referred to as SU and BRF and a combination of the 16 GCM model referred as AME and were further analysed and validated for spatial pattern of precipitation and temperature and SPEI drought and flood hazard for the study period and their implication for hydrologic modelling.

**Table 5.4:** Summary of GCM overall ranking based on aggregated SU and IS of climate variables relative to observation data.

SYMMETRIC UNCERTAINTY		BORUTA ALGORITHMS	
Model	SC	Model	AIS
<b>M13</b>	<b>0.89</b>	<b>M9</b>	<b>60.22</b>
<b>M12</b>	<b>0.87</b>	<b>M10</b>	<b>45.13</b>
<b>M9</b>	<b>0.85</b>	<b>M13</b>	<b>29.54</b>
<b>M16</b>	<b>0.83</b>	<b>M14</b>	<b>15.11</b>
M6	0.80	M12	10.43
M8	0.80	M16	8.54
M4	0.80	M7	8.49
M5	0.78	M2	8.14
M11	0.78	M5	6.75
M2	0.78	M6	6.52
M14	0.78	M11	6.49
M15	0.75	M1	5.18
M1	0.73	M15	5.15
M10	0.73	M4	4.52
M3	0.71	M8	4.36
M7	0.60	M3	3.97

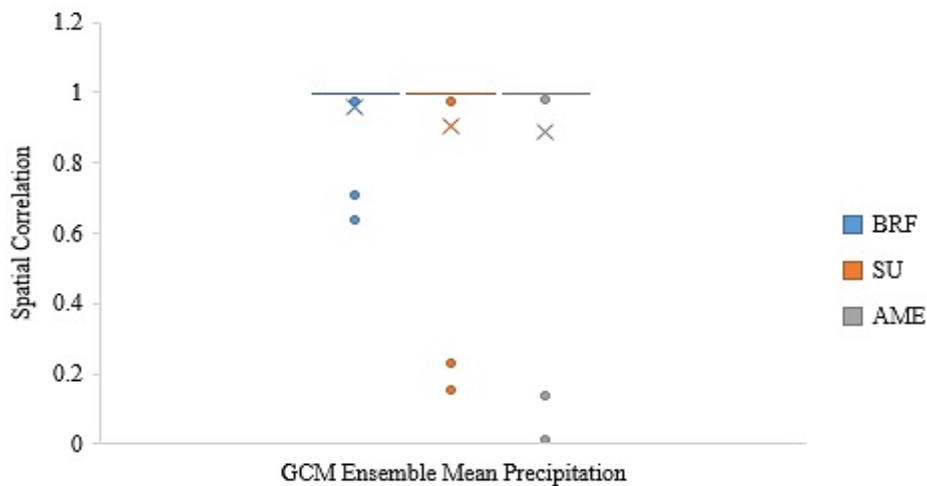
The result of the overall ranking indicated that MPI-ESM1.2-LR, MIROC6, INM-CM4.8 and NorESM2-MM are quite suitable from symmetric uncertainty approach, while INM-CM4.8, INM-CM5.0, MPI-ESM1.2-LR and MRI-ESM2.0 for Boruta random forest approach and are limited to four GCMs to others, due in part to the significant difference in their skills from the rating metrics scores in **Table 5.4**.

### 5.5.3.1 Spatial and temporal pattern of precipitation and temperature of GCM ensemble mean.

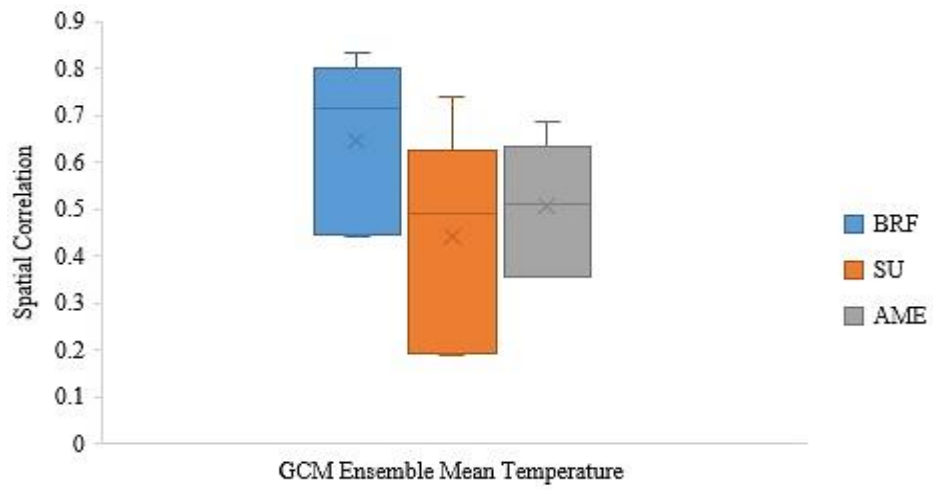
The spatial correlation and pattern of the GCM ensemble mean annual precipitation and temperature was used to validate and measure uncertainty range of the three different approaches relative to the observation as in **Figure 5.6a-b** and **Figure 5.7a-b** for the period 1979 – 2012.

The result of the spatial correlation between the mean annual precipitation and temperature indicated that BRF is consistent with the observation having a correlation value in the range of 0.641 – 0.9995 (0.9991) and 0.4423 – 0.8345 (0.7136) respectively.

Evaluation from SU and AME are quite satisfactory, however, there are mismatches or poor correlation observed, especially in the Sahelo-Sudanian zone with a recorded spatial correlation as low as 0.15 and 0.19 for SU and 0.01 and 0.36 for AME relative to the observations for annual mean precipitation and temperature respectively.

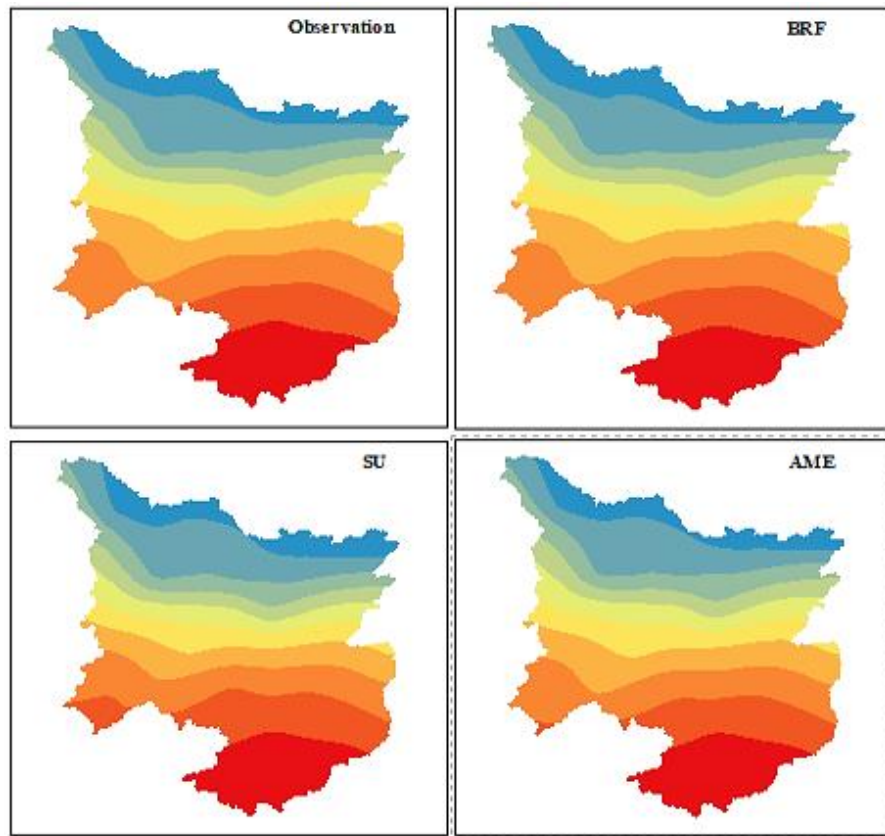


(a)

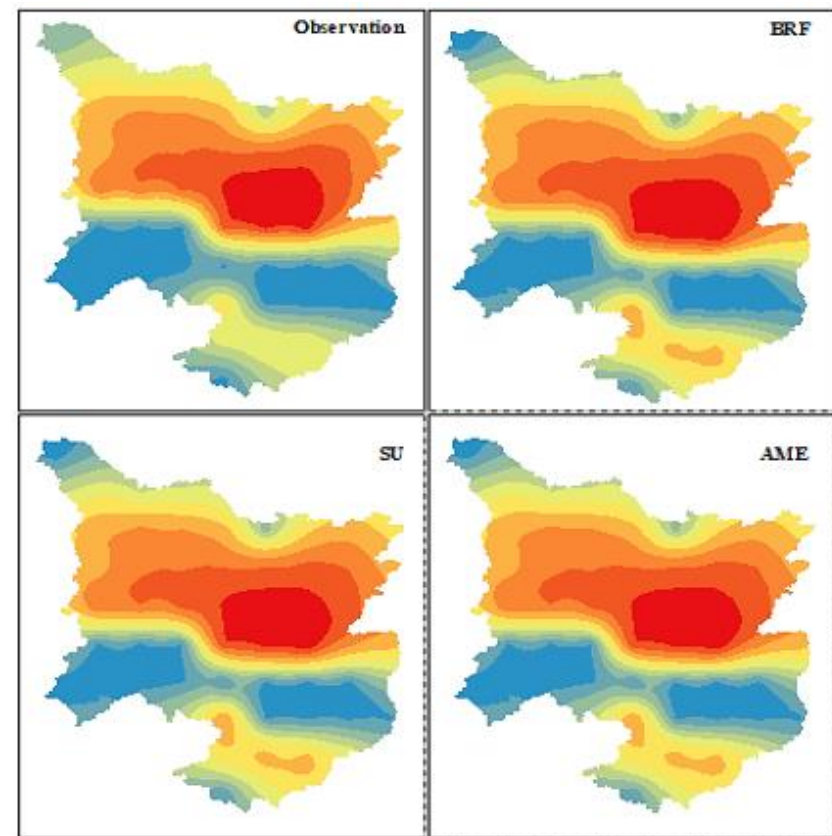


(b)

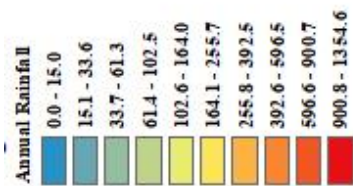
**Figure 5.6:** Comparison of spatial correlation of GCM ensemble mean performance relative to observed climate data (a) Annual precipitation (b) Annual temperature.



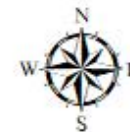
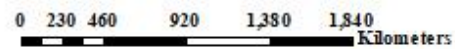
Mean Annual Precipitation 1979 - 2012



Mean Annual Temperature 1979 - 2012



(a)



(b)

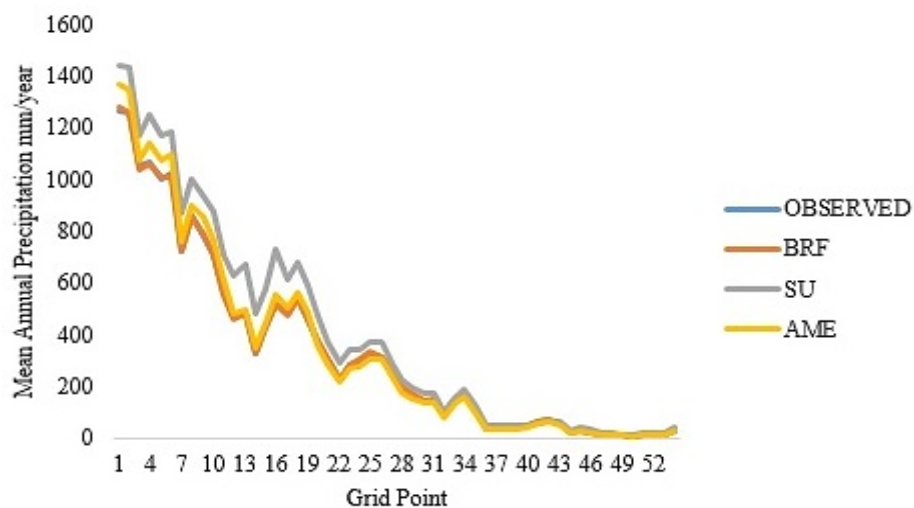


Figure 5.7: Comparison of the spatial pattern of GCM ensemble performance relative to observed climate data (a) Annual precipitation (b) Annual temperature.

The results obtained in the temporal assessment of the evaluated approaches have shown that the correlation between the ensemble mean temperature are similar and quite skilful with  $R^2 = 0.984$  and a mean bias of  $0.49^\circ\text{C}$ ,  $0.49^\circ\text{C}$  and  $0.50^\circ\text{C}$  for BRF, SU and AME respectively.

However, the BRF approach indicated an effectively significant temporal correlation of 0.95 and an annual mean bias of  $0.638\text{mm/year}$  as compared to SU and AME with spatial correlation of 0.82 and 0.88 and annual mean bias precipitation of  $68.19\text{mm/year}$  and  $10.57\text{mm/year}$  respectively.

The biases are found to be significant and visible in the south-western part of the basin as seen in the grid-based analysis of the annual precipitation in **Figure 5.8** shown below, where significant deviations are noticed between grid point 1 to 10 for SU and 1 to 26 for AME, while BRF showed almost a perfect match relative to the observation across the grid points.



*Figure 5.8: Variation of the temporal pattern of mean annual precipitation across grid points for 1979 – 2012.*

### 5.5.3.2 Assessment of climate extremes (SPEI drought and flood hazard) events of GCM ensemble mean.

The temporal analysis of climate extreme indices was assessed at 12-month time step by the BRF, SU and AME and compared to the observation at the four climatic zone of the basin



to understand the relative skills in predicting the pattern and frequency of extreme event (SPEI drought and flood hazard) and shift in trend within the study period in the four climatic zones.

The result indicates an inherent shift (**Figure 5.9a.**), relatively in wet climate (1980 – 1998) to a transition from moderate to extreme droughts (1999 – 2012) in the Saharan zone with the frequency of SPEI values of the BRF, SU and AME approach consistent with the observation at 53.3%, 45.5% and 48.9% of the time respectively.

Results of the SPEI values of the approaches relative to the observation **Figure 5.9b – d**, in the Sahelo-Saharan zone is 48.99%, 53.28% and 47.22%, Sahelo-Sudanian zone is 45.2%, 45.96%, and 43.43% and Sudano-Guinean zones is 48.23%, 50.25% and 39.14% respectively.

However, the statistical trends based on multi-year SPEI indices for the period 1980 – 2012 in the four climatic zones of the basin indicated that the BRF approach captured the extreme event direction quite accurately relative to the observation as seen in the z-statistic values within the same trend envelope (**Table 5.5**).

Although all the approaches showed a satisfactory result in predicting the trend direction but there is under-estimation in the magnitude of the extreme event by the SU ( $-0.0013$ ) and AME ( $-0.0033$ ) approach in the Saharan zone which indicate an insignificant shift from wet to drought events (drying trend) as against the trend magnitude exhibited by the observation ( $-0.0057$ ) which is consistent with the BRF ( $-0.0058$ ) approach.

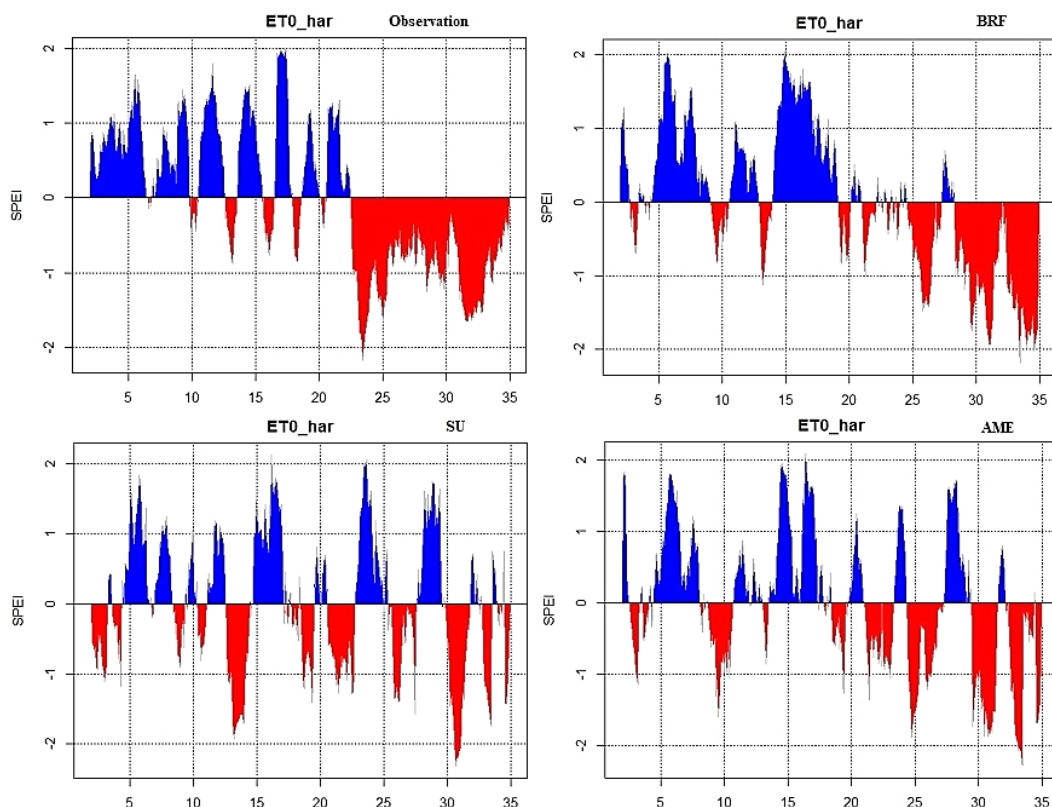
The approaches showed difficulty in predicting magnitude of the trend in the Sahelo-Sudanian zone which indicated an over-estimation relative to the observation with statistically significant wetting trend and the magnitude in the order of 0.0031, 0.0038 and 0.0057 for BRF, SU and AME respectively, as against observation (0.001).

However, the deviation is more pronounced in the AME approach and it is observed that there is a frequent and consistent shift in trend from wetting to drying period across the climatic zone and is consistently captured by the BRF approach relative to the observation. Bold values in **Table 5.5.**, indicate consistent agreement between simulated GCM ensemble mean and observation that efficiently represent the basin climatic features and the outcome will be suited to limit the magnitude of uncertainties and accurate hydroclimatic hazard representation in impact studies.

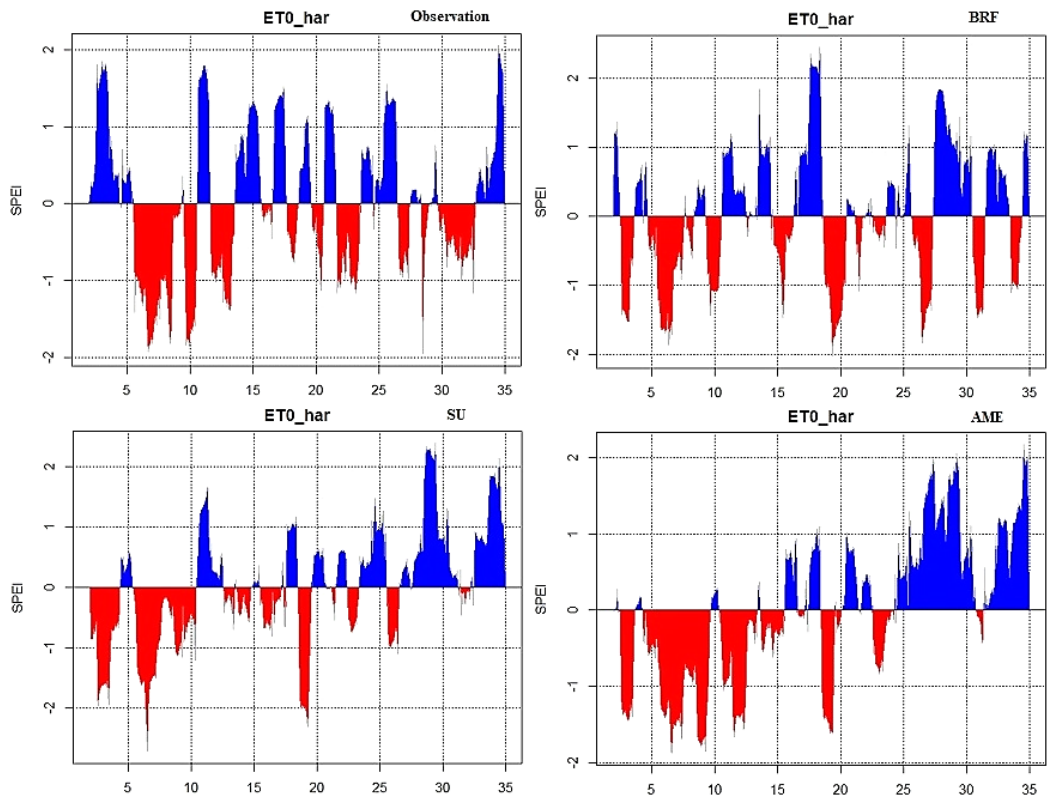
*Table 5.5: Mann Kendall Z-statistic of linear trend of extreme event for the period 1980 - 2012*

Climatic Zones	SPEI Trend			
	Observation	BRF	SU	AME
Saharan	<b>-2.44*</b>	<b>-2.25*</b>	-1.04	-1.33
Sahelo-Saharan	<b>1.39</b>	<b>1.04</b>	3.25*	5.78*
Sahelo-Sudano	0.38	2.27*	3.07*	4.94*
Sudano-Guinean	<b>0.69</b>	<b>0.74</b>	2.28*	2.28*

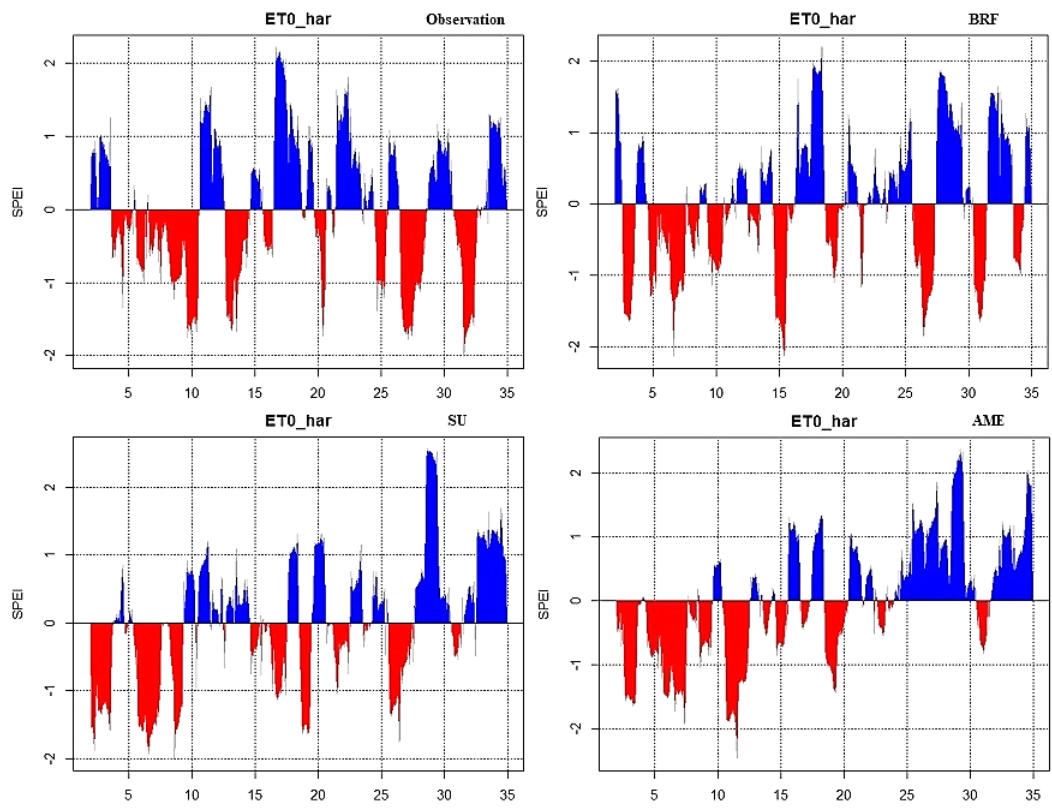
Note: \* indicates statistically significant trend at 0.05 level of significance and bold values a match of similar trend envelop.



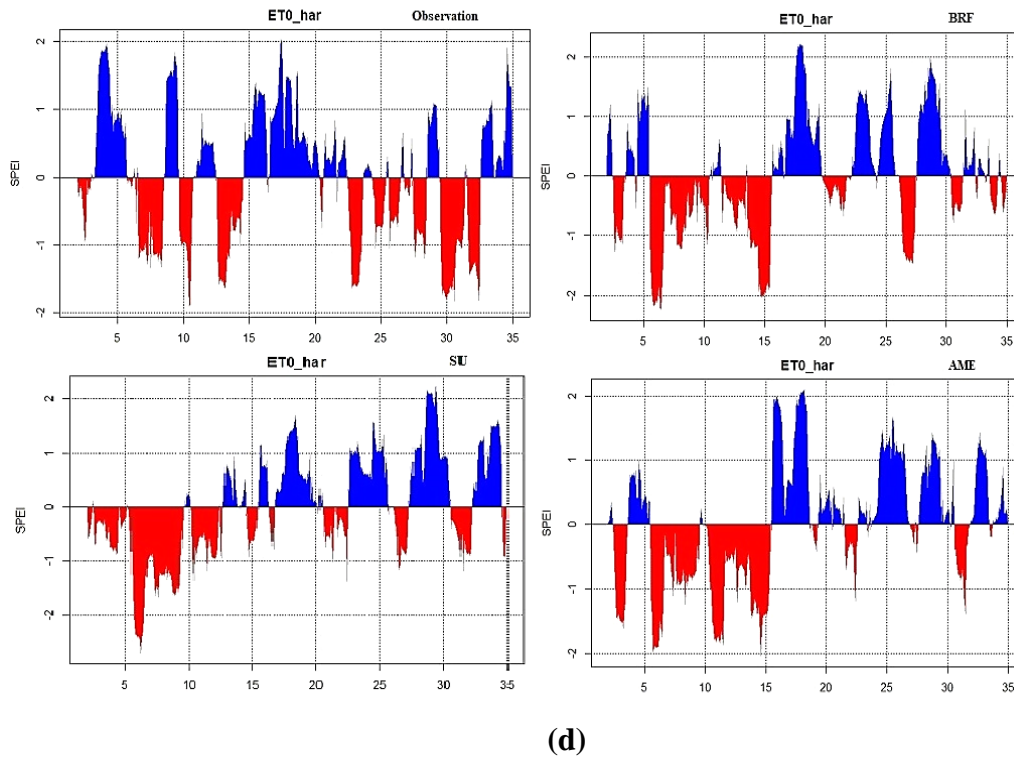
(a)



(b)



(c)



**Figure 5.9:** Temporal variations climate extreme event of GCMs ensemble mean performance of relative to observation using BRF, SU and AME approach in the Lake Chad basin. (a): Saharan zone (b): Sahelo-Saharan zone (c): Sahelo-Sudanian zone (d): Sudano-Guinean zone.

A trade off was created in the evaluation process to justify the selected model performance using the two techniques irrespective of climate variable of interest and a varied level of performance was noticed from one grid point to another.

The ensemble mean approach was quite essential because it led to reduced biases and their combinations emphasizing on few models with good performance are required and this is particularly important in watersheds with sparse observed climate data and high climate variability.

The BRF approach has shown to be promising in the evaluation with a recorded lower bias in the temperature and precipitation and a more accurate representation of the magnitude, pattern, and trends of extreme event. Overestimation or considerable bias were observed by the SU and AME approaches in the southwestern part of the basin.

The associated uncertainties can be evaluated further by considering the sensitivities of ensemble with alternative metrics or combination of approaches. However, the BRF approach

has shown to be quite robust in evaluating the integrity of the regionalisation of GCMs over different timescale that exhibit good performance during the baseline period and their combination may likely be valid to represent the future period under climate change scenarios with certainty. The weighing scheme developed in this study is an exploratory framework that can be tested in various watersheds of interest for improved water policy planning.

## **5.6 Conclusion**

General circulation models are important and provides a pathway for simulation and assessment of the perceived impact of climate change on local, regional, and global hydrology. However, the choice of GCM input data, interpolation and downscaling method, timescale and timesteps are essential and critical for effective and accurate representation of the past, present, and future basin hydrologic process for fair and equitable river basin management and policy planning for sustainability.

Earlier studies suggest that most evaluations ignored the inter-dependencies among models of a known variable and could create over-fitting problems. This study is based on an ensemble of 16 CMIP6 GCMs at daily timestep evaluating the efficacy and robustness of the state-of-the-art Boruta random forest algorithm technique has shown this to be a viable tool for selection of relevant models to reduce redundancy, complexity and over-fitting problems associated with climate models to ensure sufficient overlap of chosen models' ensemble mean with observations.

This seeks to limit the drawbacks encountered from existing techniques such as but not limited to inability to analyse complicated inputs, stochastic aspects, and climatic and hydrologic properties that are intricately interrelated, reduce the transfer of uncertainties into hydrologic processes and address critical temporal and spatial behaviour of the climate variables as a precursor for reliable and accurate predictions.

Highlights from the study revealed that there are inherent weaknesses associated temporal and spatial downscaling techniques and multiple techniques should be tried and examined to limit uncertainty range and inadequacies of GCMs, because exploring different multi-site downscaling techniques is very important in increasing the effectiveness of GCMs performance.

Therefore, combination of appropriate GCMs can enhance spatial and temporal variability to accurately reproduce observed multiple statistical properties of climate variables for improved output and reduced uncertainty in hydrologic modelling at regional and local basin scale.

The selected models from Boruta random forest technique adequately have the capability in reducing biases in precipitation compared to the other approaches, although similar performances were observed in terms temperature and can capture the trends, patterns, and magnitude of extreme events within the accepted confidence limit.

The findings associated with this study are generally not meant to be a process to identify viable GCM dataset suitable for hydroclimatic study, but also to present a simple and efficient methodology to examine the limitations associated with the selected GCM ensemble for impact study. Therefore, the methodology proposed is not unique and therefore be explored to other basins for reliable representation of catchment climatology, representing a key step forward in GCM ensemble impact research.

## **5.7 Afterward**

The successful implementation of the data pruning approach using BRF algorithms for GCMs models from this study has shown that efforts in building hydrologic models for water security assessment require exploring multi-disciplinary framework to improve the science of watershed representation and reduction in the propagation of the various sources of uncertainty in modelling.

The strategy developed in this study was infused in the workflow of SWAT modelling process in the next phase of the research to form an integrated framework. The framework was further used to investigate the efficacy of simulating ET within the context of the research objective for effective hydrologic modelling in four sub-watershed of data-sparse Lake Chad basin.

The output of the modelling process was also used to investigate the impact of climate change on projected green and blue water availability and sustainability at sub-watershed scale.

This exploratory approach is necessary especially in data-sparse watersheds and this has been applied in the next phase of the research to provide a solution for reliable water resources assessment in the context of climate change, rather than relying on outputs from regionalization concepts by spatial proximity or homogenization, i.e. the process of transferring model parameter values from neighboring gauged catchment to the ungauged catchments, the only rationale being that they share similar behavior because the climate and catchment conditions vary evenly across the basins.

This have been reported by Oudin et al., (2008) to produce unsatisfactory results and its application may be unreliable and capable of generating misleading water resource information that hinders adequate policy planning at local basin scale.

## **CHAPTER 6: INTEGRATED FRAMEWORK FOR HYDROLOGIC MODELLING IN DATA-SPARSE WATERSHEDS AND CLIMATE CHANGE IMPACT ON PROJECTED GREEN AND BLUE WATER SUSTAINABILITY.**

### **Preamble**

This chapter intends to proffer solution to modelling concerns in data-sparse watersheds by providing an assessment framework that can be relied upon especially in the management and protection of water resources in the future at the local watershed scale and understand how anthropogenic climate changes will impact the basin's water balance dynamics. This is important for assessing projected water resource availability and sustainability for watershed planning and water policy decisions.

Hydrologic modelling has been the key tool for understanding watershed response to projected climate change and its impact on water resource hazards. However, the development of such models requires ground observations for reliable predictions and the lack of observation data in most regional and local basins across the world especially sub-tropical and arid regions necessitates the use of alternative satellite-based measurements and their interlinkages with hydrologic models is a major source of uncertainty in regional applications.

Earlier approaches in the literature to deal with such challenges in data-sparse regions through regionalization is quite challenging and produced contradictory results as demonstrated by Oudin et al., (2008), where the modelling results of three regionalization approaches on 913 catchments in France produces varied performance efficiencies due to variations and effects in catchment gauge network quality, density and lack of key physical characteristics (Visessri and McIntyre, 2016).

This challenge was addressed by incorporating machine learning and the traditional SWAT model as an integrated framework to develop a prediction model that can deliver



actionable hydrologic information within an acceptable uncertainty in data-sparse regions for accurate water security assessment in response to climate change.

The CPC precipitation and PGF maximum and minimum temperature data at daily time step was adopted and used for preprocessing (i.e., downscaling and bias correction) of the GCM data because they are considered to reproduce the basin climatologic features fairly and accurately.

Similarly, the four GCM models were selected based on the superior performance of the multimodel ensemble average in representing the basin observational biases of annual precipitation and average temperature, trend and magnitude of return period of drought and flood hazards which is essential in hydroclimatic impact study.

However, it is worth noting that the study considered four sub-watersheds due to the variations in their morphological characteristics to investigate and understand the performance of the modelling strategy and finally lack of sectoral water resource information in some of the sub-watershed constrained the application of research objectives 5 and 6 to the Yobe-Komadugu watershed.

The Paper following was published in “Frontiers in Environmental Sciences” titled “Integrated framework for hydrologic modelling in data-sparse watersheds and climate change impact on projected green and blue water sustainability” (2023). <https://doi.org/10.3389/fenvs.2023.1233216>.

### **6.1 Abstract of paper**

Climate and hydrologic hazards threaten the distribution of watersheds’ water resources in time and space, necessitating planning for sustainable resilience and adaptation. Hydrologic modelling has emerged as a potential solution for understanding watershed responses to projected climate change and a prediction model that can deliver actionable information is necessary, although it requires basin-scale observations to calibrate the model

to reliably predict basin-scale water resources hazards. Such luxury is not always tenable in watersheds with inadequate ground-based observation.

However, satellite-based ET data coupled with a machine learning feature selection as a data refinement process has made integrated water balance modelling widely regarded as a viable alternative for improving the capability of watershed modelling processes in data-sparse regions.

This study developed a convincing hydrologic model framework to calibrate sufficiently and provide accurate behavioural solutions for all model responses. The framework was applied to four sub-watersheds that form the larger Lake Chad basin. The model results were applied to assess the dynamic changes in projected blue and green water resource sustainability in response to climate change in one of the subbasins.

Study findings indicate that hydrologic fluxes can be simulated accurately with varying degrees of acceptability with  $R^2$  and NSE values in the range of 0.69 – 0.88 and 0.45 – 0.77 for calibration and 0.69 – 0.79 and 0.34 – 0.63 for validation respectively, and captured within a satisfactory uncertainty range of P-factor and R-factor values of 0.68 – 0.93 and 0.73 – 1.31 in 83%, 67%, 85.7% and 81.3% of the sub-watersheds based on multi-site simulation in spite of distinct watershed morphology although there are significant trade-offs in parameter sensitivity.

Whilst green water is the dominant freshwater component across the basin relative to blue water, climate change may be a significant factor in changes in projected green water sustainability status, the combination of socioeconomic drivers (only considered for blue water sustainability assessment) and climate change may significantly impact the projected blue water sustainability status across the basin.

Projected changes in green and blue water sustainability status have shown that more than 50% of the sub-watershed will be ecologically fragile and identified freshwater geographic

sustainability hotspots may be beyond restoration without adequate long-term river basin water resources plans.

**Keywords:** Data uncertainty, Feature selection, integrated modelling, ungauged watershed, Climate change, Water footprint, Freshwater sustainability.

## 6.2 Introduction

Water resources planning must find a solution to the issue of achieving efficient and equitable water usage, particularly in light of the growing population, climate change, and depleting water supplies (Novoa et al., 2019). Water is the cornerstone of community development since it provides such a wide range of ecological functions. This allows for its effective, equitable, and sustainable allocation in order to reduce poverty, foster economic growth, and safeguard the environment (Hu et al., 2016).

The rate and amount of time that water spends in various storage reservoirs, including surface and groundwater, seas, atmosphere, snow, and ice, has been altered due to human use (Keys et al., 2016). Consequently, it is a part of the world's greatest difficulties in attaining water sustainability, which is defined as meeting everyone's present water needs without compromising the supply in the future while advancing societal goals and preserving the environment (Chouchane et al., 2018; Hu et al., 2016).

As a result, many administrative authorities have made the management of water resources sustainably a top priority to ensure that all residents and economic sectors have access to water sufficiently in the right quality and quantity (Martinsen et al., 2019; Tortajada et al., 2019).

The sustainability of water in a basin can only be achieved if it is possible to sustain ecosystems' hydrological, ecological, biological, and chemical processes while providing an equitable and effective water supply over time (Pfister et al., 2009; Wang et al., 2016). The water footprint (WF) concept addresses these needs by providing an assessment of water

resources that accounts for natural variability and demands that compete for their use (Hejazi et al., 2014).

A multi-dimensional indicator called the Water Footprint (WF) reveals characteristics of anthropogenic stresses on water supplies. From the perspective of river basin water accounting, WF is defined as freshwater consumed and contaminated by production activities of various economic sectors present in the river basin (Muratoglu et al., 2022; Xie et al., 2020).

This offers insights into water-related challenges and aids in understanding present patterns of water allocation across different river basin sectors (Muratoglu et al., 2022), and enables decision-makers to take advantage of the substantial data on water use supplied by the WF technique by improving water management, hotspot identification, and the development of appropriate responses to changes (Pellicer-Martínez and Martínez-Paz, 2018). The approach is excellent for comparing water resources across different administrative boundaries in terms of quantity and quality (Li et al., 2018).

Blue and green water are the two categories into which the freshwater cycle can be separated based on the hydrological processes and types of storage involved. Greenwater is the portion of precipitation that seeps into the ground and changes into soil moisture, or momentarily sits on top of the ground or vegetation, and subsequently evaporates and transpires back into the atmosphere. Bluewater is the term for precipitation that accumulates in aquifers, lakes, and reservoirs and flows through or below the land surface (Rockström et al., 2009; Rodrigues et al., 2014).

The consumption of both blue and green water by human activities is included in the water footprint concept, per Hoekstra et al., (2011), The result is that the green water footprint (GWF) represents the estimated amount of green water that is consumed by agricultural land (i.e., evapotranspiration from crop and pastureland), which is frequently

referred to as productive vapour flows), as opposed to the blue water footprint (BWF), which represents the consumptive use of blue water resources (surface water and groundwater).

Geographic hydrologic models created at various time resolutions and spatial scales have started to become more complicated as a result of the use of WF as a sustainability indicator in order to determine environmental water consumption restrictions (Shrestha et al., 2017). This indicator is particularly helpful in regions susceptible to water variabilities, such as basins in Mediterranean and tropical climates, where the demand for water for irrigation rises during decreased precipitation, limiting runoff and downstream flows (Novoa et al., 2019).

Previous studies have demonstrated that, in addition to its effects on precipitation, soil moisture, evapotranspiration, and runoff, climate change has an impact on the distribution of water resources both spatially and temporally (Montaldo and Oren, 2018; Sun et al., 2018). Extreme hydrological events are rising in frequency and importance due to the deepening of global climate change, creating new problems for managing water resources and the regional water cycle (Tabari, 2020; Vicente-Serrano et al., 2017).

Global attention has been drawn to climate change and its possible effects on water resources. However, there are many uncertainties in future climate change estimates (such as those for temperature and precipitation), making it difficult for planners to decide on appropriate adaptation measures (Dessai and Hulme, 2007; Gosling and Arnell, 2016).

The main sources of these uncertainties are changes in the initializations and parameterizations used in climate models to explain physical processes as well as downscaling methods (Zhuang et al., 2016).

It has been discovered that water resources are vulnerable to these uncertainties and challenging to anticipate with precision in a changing environment. Therefore, it is essential to develop water management plans in times of global change within a complex and uncertain environment (Wang et al., 2016).

Multiple climate models were used in an evaluation framework to find effective water resource management plans under the effects of climate change; the results showed that these plans are extremely vulnerable to climatic changes.

Some conclusions drawn from other studies revealed that water resources in various regions are sensitive to climate change, and the relative influence varies significantly around the world, according to the evidence, the sources and type of uncertainty affect how adaptation strategies are chosen (Arnell et al., 2011; Cai et al., 2015; Dessai and Hulme, 2007; Refsgaard et al., 2013; Sun et al., 2017; Tzabiras et al., 2016).

Identification of effective corporate strategies and policy actions requires uncertainty analysis (For example, climate adaptation, resilience and mitigation measures). Researchers, decision-makers, and stakeholders have more transparency and confidence in scientific analyses when they are informed about the locations, types, and nature of uncertainty (Gabbert et al., 2010; Kirchner et al., 2021). Uncertainty analysis is typically necessary for the scientific publishing of model-based quantitative assessments and is regarded as excellent modeling practice (Troost et al., 2015).

Observational data are the foundation of our understanding of environmental systems, but their scarcity and unpredictability, limit study and their practical applications. The accuracy of atmospheric data is crucial for the validity of hydro-meteorological and climatological investigations, among other things (Zandler et al., 2019).

The flaws in the input rainfall data utilised might be reduced or amplified by the nonlinearity of hydrological processes, which can lead to a good or bad depiction of the hydrological responses and consequently lead to inadequate water resource policy and adaptation measures (Maggioni and Massari, 2018).

In order to improve the spatiotemporal process representation, distributed observational datasets must be used to inform and assess distributed hydrological models, which have been

developed to enable large-scale forecasts (Baroni et al., 2019; Ocio et al., 2019). In this context, determining if meteorological data are adequate and coherent to accurately reproduce hydrological processes is a requirement before choosing data for managing water resources (Laiti et al., 2018).

Integrated modelling is useful in many areas of study on global climate change. Here, we define integrated modelling as an interdisciplinary system of connected empirical data and mathematical models that are based on disciplinary concepts in order to provide a more complete and accurate picture of interactions between people and their environment (Laniak et al., 2013; Moss et al., 2010).

Uncertainty can manifest and build up across an integrated modelling framework, which makes it a significant problem for integrated modelling. Uncertainty is mostly dealt with in two ways by existing integrated modelling frameworks: First, quantifying the uncertainty of future developments through scenarios, for example, description of alternatives of internally consistent possible future (Mitter et al., 2019; Reilly and Willenbockel, 2010). Secondly, contributions from research teams, systematic model comparisons across the scientific community, and other techniques are utilised to address uncertainty due to various data sources and model designs (Elliott et al., 2014; Folberth et al., 2019).

The full identification and tracking of uncertainties in integrated modelling i.e., the manner in which uncertainty spreads among climate models as applied to hydrologic modelling in data-sparse regions has received very little attention (Holzkämper et al., 2015; Karner et al., 2019; Mitter and Schmid, 2019). Such analysis was previously acknowledged as a serious research gap in the early phases of integrated modelling, particularly for the propagation of uncertainty from land use optimisation models to the construction of hydrologic models.

In connection with this effort, numerous modelling studies have been conducted, that provides a unique opportunity to define the various hydrological processes and the relationships between the various hydrological variables (Bierkens et al., 2015). However, it has also been emphasised that one of the major scientific difficulties is continually refining the depiction of hydrological processes in the model design (Clark et al., 2015).

The hydrologic community has agreed on the necessity for additional datasets in addition to observed streamflow and associated signature measurements to enhance the portrayal of the key physical processes (Clark et al., 2016). The fact that climate models still struggle to accurately replicate important climate processes, is of greater concern.

While precipitation estimates are widely variable, temperature projections are more similar across all climate models as they can still vary. Future hydro-meteorological conditions can be uncertainly predicted because of the significant degree of variability in GCM outputs.

In order to create an integrated modelling framework, it is often necessary to work on individual model modifications, model connections that are improved, and the application of the integrated modelling framework to particular research issues, whose outcome can be relied upon for basin-scale assessment of water security, sustainability, and other related applications to achieve effective water policy decision in response to projected climate change.

For a meaningful comprehension of basin weather patterns and their future trends based on feature extraction by training the historical dataset using artificial intelligence to track water resource indicators, a prediction model that can deliver actionable information is necessary (Ali et al., 2020; Kratzert et al., 2018).

In this study, we created a convincing framework or strategy to deal with the difficulties of modelling in areas with little or sparse data, appropriate ways to use alternative research



data to evaluate models, and considerations for data uncertainty and the incompatibility between models and measurements. The framework integrates a machine learning technique, Boruta random forest (BRF) optimizer and process-based hydrologic model referred to as soil and water assessment tool (SWAT) to refine the data input process mechanism for the development of a reliable hydrologic model for basin water resources assessment. The methodology will be applied to four sub-watersheds that encompass the Chad hydrologic basin with variable morphological properties, in Sub-Saharan Africa.

The objective is to provide a novel pathway to increase transparency and improve uncertainty communication of long-term water balance models in an easily understood way without compromising scientific accuracy in data-sparse watersheds which have not been adequately studied.

This concept aims to be generic and flexible enough within the acceptable uncertainty band to allow for its application in other basins with similar modelling issues. Finally, the integrated model framework will provide a crucial link between hydrology and human activities at local watershed levels to assess the implications and dynamic changes from baseline, the projected blue and green water resources and sustainability in response to climate change at annual and monthly timescale in Yobe-Komadugu sub-watershed.

### **6.3 Case Study area and data**

#### **6.3.1 Case study Area**

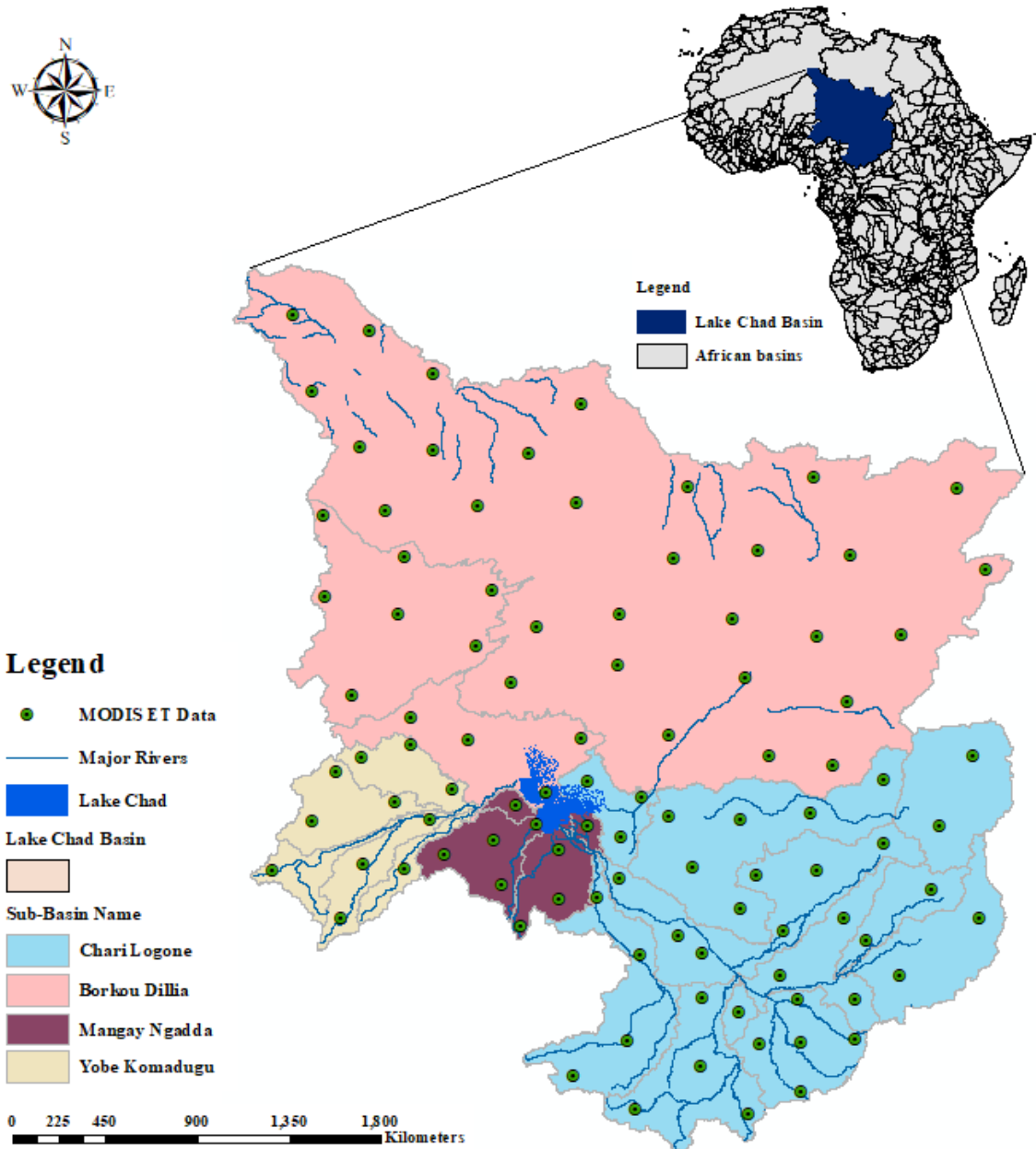
The Lake Chad Basin, with an estimated area of over 2,500,000 km<sup>2</sup>, is one of the largest endorheic basins in the world (Coe and Foley, 2001; Gao et al., 2011). It is located between the Sahara and the Sudano-Sahelian areas of West Africa, between latitudes of 5.2° - 25.3° N and longitudes of 6.9° - 24.5° E (**Figure 6.1**).

The basin receives the majority of its annual rainfall between July and September. The basin is known for being extremely vulnerable to the effects of climate variability, which

typically leads to significant drought and water shortage conditions (Ndehedehe et al., 2018), and the lake is the primary source of fresh water for livestock grazing, fish farming and other socioeconomic activities (Buma et al., 2016).

The major contributors of discharge to the lake are Chari River (~ 90%) with annual streamflow of 860 m<sup>3</sup>/s between 1960 – 2013 and Yobe River (~ 2 – 5%) with annual streamflow of 18 m<sup>3</sup>/s between 1961 – 2013 (Lemoalle, 2014). Other rivers that contribute supplies of between 1 – 2% are Gubio, Yesderam, Ngadda, and El-beid.

However, there are a few rivers like the Batha River and others situated in the Saharan zone that does not have an outlet to Lake Chad (**Figure 6.1**). The precipitation in the basin varies geographically and seasonally between < 100 – 1500 mm/yr. (Nkiaka et al., 2018b).



*Figure 6.1: Lake Chad Basin showing sub-basins, lake, major river networks and MODIS Evapotranspiration (ET) data points.*

## 6.3.2 Dataset Description

### 6.3.2.1 MODIS actual evapotranspiration data

The state of observed streamflow data is quite poor and inadequate with many missing data points, which undermines the confidence in the output of hydrologic modelling results temporally and spatially in the entire basin. Alternatively, the availability of high spatial variability of satellite-derived land surface MODIS-NASA Evapotranspiration (ET) data at

monthly timescale was extracted by overlaying the  $1.0 \times 1.0$  AET grids with the subbasin map of the ArcSWAT project and 100 observation points (**Figure 6.1.**) was generated and reaggregated to develop 59 simulation sub-basin points across the entire watershed.

This is achieved by overlaying the MODIS-AET regular grids that was extracted on the sub-basin map and the AET grids inside each sub-basin was averaged into a single grid point to represent the AET of the sub-basin as recommended by Abbaspour et al., (2019) and the data was divided into 1983 – 1998 for calibration and 1999 – 2006 for validation of the models.

A more accurate hydrologic model simulation may be obtained by taking into account the geographical distribution of the actual ground surface ET throughout the watershed. The Penman-Monteith equation is used by the MODIS-NASA ET technique, which also takes into account things like plant transpiration and the evaporation of soil moisture. The quality control parameters have been confirmed using global evaporation flux tower data (Autovino et al., 2016).

#### **6.3.2.2 Digital Elevation Model, Soil, Land Use and Land cover data**

The watershed was delineated using the ArcSWAT programme in the ArcMap 10.8 environment using the topography information from the basin that was collected from the Aster global digital elevation model version 3 with a spatial resolution of 30 m. The software extension can be found in (<https://swat.tamu.edu/software/arcswat/>).

The soil data were obtained from Harmonized World Soil Database (HWSD), with a 1 km resolution founded by the Food and Agricultural Organisation (FAO) and notable research centres (Abbaspour et al., 2019).

Land use and land cover (LULC) data was obtained from European Space Agency, which was an initiative that developed global composite land cover maps using observations from the 300 m MERIS sensor onboard the ENVISAT satellite mission. The GlobeCover map

contains 23 land cover types (Bontemps et al., 2011). The description, resolution and data source are shown in **Table 6.1**.

*Table 6.1: Input data required for hydrologic model development.*

<b>Data Used</b>	<b>Description</b>	<b>Resolution</b>	<b>Source</b>
Topography	Digital Elevation Model	30m × 30m	<a href="https://search.earthdata.nasa.gov/search/granules">https://search.earthdata.nasa.gov/search/granules</a> <i>ASTER Global Digital Elevation Model V003</i>
Land Use Data	GlobeCover land use Map	5° × 5°	<a href="http://due.esrin.esa.int/page_globcover.php">http://due.esrin.esa.int/page_globcover.php</a> <i>(GlobeCover)</i>
Soil Data	Digital Soil Map	1 km	<a href="http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database">http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database</a> <i>(HWSO v1.12)</i>
Meteorological Data	Precipitation, Maximum Temperature Minimum Temperature	1° × 1° Daily	<a href="https://esgf-node.llnl.gov/projects/esgf-llnl/CMIP6">https://esgf-node.llnl.gov/projects/esgf-llnl/CMIP6</a>
ET Data	Actual Evapotranspiration	1° × 1° mm/month	<a href="http://files.nts.gov.umt.edu/data/ET_global_monthly">http://files.nts.gov.umt.edu/data/ET_global_monthly</a> <i>MODIS-NASA Data</i>

### 6.3.2.3 Climate Data

The gridded precipitation data of the US Climate Prediction Centre (CPC) and maximum and minimum temperature data of the Princeton University Global Meteorological Forcing (PGF v.2) as recommended from a previous study by Lawal et al., (2021), at daily time step between 1979 – 2011 was adopted in this study.

The data was extracted at 1° × 1° grid resolution and used for pre-processing of CMIP6 general circulation models for baseline (1979 – 2011) and projected (2021 – 2080) climate change scenario data considering two shared socioeconomic pathways based on carbon dioxide emission scenarios SSP2(4.5) and SSP5(8.5), supported by Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP6), and the data is available and can be extracted from the source provided in **Table 6.1**.

## 6.4 Research Methodology

### 6.4.1 Pre-processing of input data

The dataset required for the integrated model framework needs to be checked and prepared to fit the model specifications for efficient and accurate output of the model

hydrologic variables. The primary input data required are pre-processed to depict the status of the watershed land management and vegetation properties.

#### **6.4.1.1 Climate models, downscaling and bias correction.**

The general circulation models (GCMs) are the primary source of information for the assessment of climate change impacts at the global and regional scales. The ensemble of four (INM-CM4.8, INM-CM5.0, MPI-ESM1.2-LR and MRI-ESM2.0) coupled model intercomparison project phase 6 (CMIP6), models were extracted at 210 data points (**Figure 6.2a**) of daily precipitation, the maximum and minimum temperature for the baseline/historical (1979 – 2011) and projected climate change scenarios (2021 – 2080) for two shared socioeconomic pathways (SSP4.5 and SSP8.5) corresponding to total radiative forcings of 4.5 and 8.5 W/m<sup>2</sup> (approximately equal to mean CO<sub>2</sub> emission concentrations of 650 and 1370 ppm), respectively in 2100.

Before predicting the future climate, it was necessary to correct the anomalies in the climate model's outputs, because they contain biases using bias correction techniques. The delta and quantile mapping techniques were used to correct the identified biases for precipitation and temperature respectively, based on a previous study conducted in the basin using CPC and PGF gridded data in line with study requirements Lawal et al., (2023).

The methods are non-parametric and corrected the predicted climate data based on point-wise empirical cumulative distribution functions. The downscaling strategies using bilinear interpolation were found to significantly improve the forms of the linked frequency distributions and minimise systematic biases and derived indices of extremes by around one order of magnitude (Thiemeßl et al., 2012).

#### **6.4.1.2 Land Use, Soil and DEM Data**

The soil characteristics for the entire watershed were extracted from the world HWSD dataset and include two soil profiles (0–30 cm and 30–100 cm depths), the available water

capacity, and the bulk density, along with the majority of the soil information needed for the SWAT model. Clay, loam, sand, clay-loam, sandy-clay, loamy-sand, sandy-loam, sandy-clay-loam, and rock make up the majority of the primary soil classifications that make up the watershed (**Figure 6.2b**).

The watershed's land use and land cover data were extracted, re-gridded to match the spatial resolution of the meteorological forcing data, and categorised into six different land uses that work with the SWAT model (**Figure 6.2c**), including artificial area (URMD) 0.013%, barren land (BARR) 52.873%, agricultural land (AGRL) 3.743%, forest land (FRST) 15.636%, vegetation (PAST) 27.842%, and water bodies (WATR) 0.166%.

In order to extract the topographic features of the terrain, which are a necessity for hydrological research, basin elevation information is crucial. The 30 m spatial resolution DEM was extracted from ASTER global digital elevation model version 3 (**Figure 6.2a**) and transformed into a Universal Transverse Mercator (UTM) coordinate system to aid the delineation of the watershed boundary.

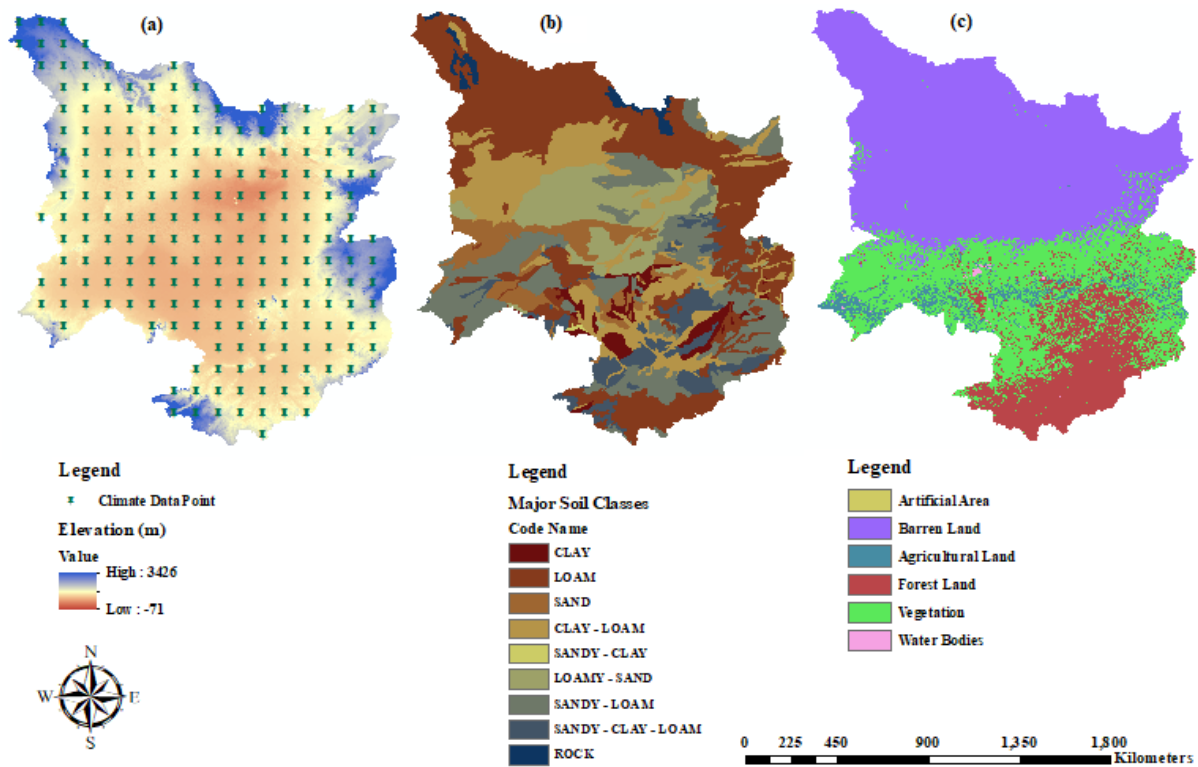


Figure 6.2: Description of morphological data in the study (a) digital elevation model and meteorological points, (b) soil types (c) land use and cover data.

### 6.4.2 Integrated modelling framework

We combined a statistical machine learning optimizer (BRF) and soil and water assessment tool (SWAT) model to refine the input process of predicted baseline and projected climate data to reduce uncertainties in model quantities such as input data and parameters (i.e., technical uncertainties), to enhance the simulation process and improve the confidence on the modelling output for the reliable assessment of basin-scale hydrologic features.

This approach is necessary to further lower the danger of misinterpreting climate signals and improve adaptation assessments. Our goal is to create an integrated modelling framework (IMF) that satisfies these criteria for evaluating the effects of anticipated regional water balance changes brought on by climate change scenarios on the sustainability of green and blue water resources in data-sparse regions under uncertainty. The two integrated processes



are described briefly in the following sub-section and the schematic overview is shown below (Figure 6.3).

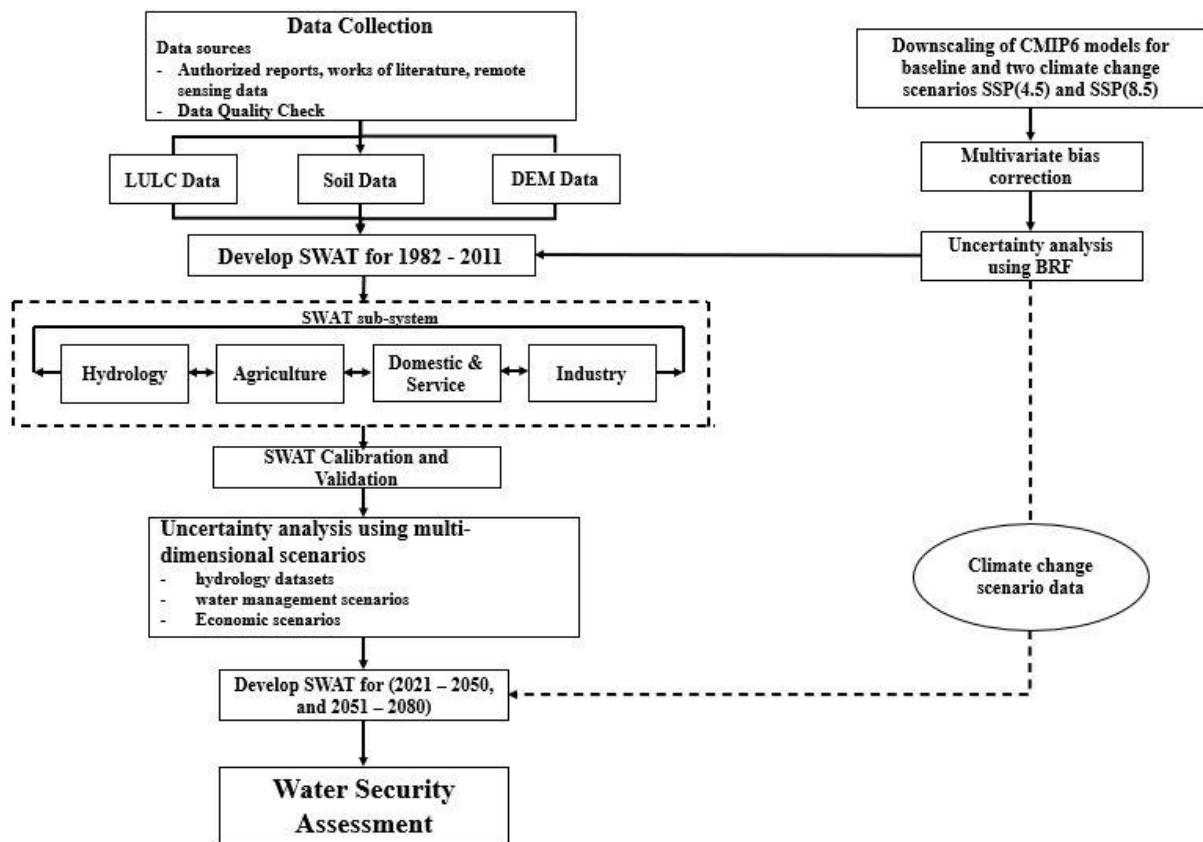


Figure 6.3: Schematic overview of the integrated SWAT and BRF modelling framework (IMF) for reliable water balance modelling in data-sparse regions.

#### 6.4.2.1 Boruta Random Forest Optimizer

The Boruta Feature selection method was created as a wrapper for the random forest algorithm, which is used to identify important features of the predictors. Every input predictor's Z-score distribution relative to the shadow property is calculated. The distribution of the Z-score metrics establishes the key components of the predictors (Kursa and Rudnicki, 2010).

It involves developing a stepwise model using a minimal-optimal feature selection technique, that rates the salient important model features and the residual based on the Boruta-determined factors (Kursa, 2016). It is an effective feature selection technique that makes it easier to categorize high-dimensional data. Information gain is used to gauge each

feature's contribution and establish its association using a novel extension of balanced information gain. When analyzing a vast amount of data to get high generalization accuracy, this is very significant.

The methodology of the optimization process of the input dataset is discussed in Lawal et al., (2023). The proposed strategy is required to address potential shortcomings of the conventional modelling approach, such as their inability to analyze stochastic features, complex variable inputs features, and highly interconnected climatic and hydrological properties that limit their ability to address critical temporal behaviour (Adamowski et al., 2012).

This has been demonstrated in Lawal et al., (2023), that the BRF feature selection technique has proven to retain the climatic signals by filtering out the redundant downscaled GCMs that may create a dip in the capability of selected ensembles developed to accurately represent basin scales hydrologic features like extreme events (return period of flood and drought), their trends and magnitudes.

Thus, integrating the feature extraction algorithms with SWAT modelling may provide an improved output of calibrated and validated water balance models for a reliable and accurate prediction of baseline and projected hydrologic features in data-sparse watersheds for water security assessment.

#### **6.4.2.2 SWAT Hydrologic Model**

The US Department of Agriculture (USDA) created the physically-based, semi-distributed and continuous time-step hydrological model known as the SWAT. The model is utilized to study water quality (sediment load and nutrient flow), water quantity (streamflow, evapotranspiration, water yield, aquifer recharge etc) and crop growth processes in different landscape and management practices (Gesualdo et al., 2019; Veetil and Mishra, 2018).

Surface runoff in the model is generated by the use of the SCS curve number method (Abbaspour et al., 2015; Veetil and Mishra, 2018), and ET is estimated by various approaches based on the source of data and basin conditions such as Priestley-Taylor, Penman-Monteith, or Hargreaves methods.

Subsurface flow components including lateral flow, groundwater flow, and percolation are evaluated using the mass balance of the subsurface system (Swain et al., 2020). The water balance equation conceptualises SWAT's simulation of the hydrological processes as:

$$SW_t = SW_o + \sum_{i=0}^t (P_i - Q_i - ET_i - G_i - R_i) \quad (6.1)$$

Where  $SW_t$  is the residual water content in the soil (mm),  $SW_o$ , initial soil water content,  $t$ , time in days,  $P_i$ , precipitation,  $Q_i$ , surface runoff,  $ET_i$ , evapotranspiration,  $G_i$ , subsurface flow from the soil profile,  $R_i$ , return flow on  $i^{th}$  day all in (mm) respectively.

SWAT primarily analyses the water balance components of each Hydrological Response Unit (HRU), which is a different group of soil and vegetation types in the sub-basin, to estimate the water availability at each sub-main basin's channel for a particular time phase.

To regulate the flow of water, the water is subsequently channelled to the basin exit via the river and subsurface systems. SWAT's calibration, validation, sensitivity analysis, and uncertainty analysis are often performed using SWAT-CUP, or calibration and uncertainty programmes (Abbaspour, 2015).

The sequential uncertainty fitting tool version 2 (SUFI-2), an optimization algorithm based on stochastic procedures within the SWAT-CUP interface was utilized for adjusting independent parameter sets by Latin Hypercube Sampling (LHS). The interface used global or One-factor-At-a-time sensitivity analysis during calibration and validation.

The model performance is evaluated by several statistical metrics such as coefficient of determination ( $R^2$ ), Nash-Sutcliffe efficiency coefficient (NSE), etc. However, this project

will utilize the  $R^2$  and NSE for calibration and validation of model results whose equations are given below.

$$R^2 = \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (6.2)$$

$$NSE = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6.3)$$

Where  $\bar{x}$  is the observed mean values,  $x_i$  is the value of the  $i^{th}$  observation,  $y_i$  is the modelled value of the  $i^{th}$  observation,  $\bar{y}$  is the mean of the simulated model values and  $n$  is the total number of samples sets of the observation.

### 6.4.3 Integrated model set-up, calibration, validation and uncertainty analysis.

#### 6.4.3.1 Integrated model simulation

The model was set up by importing the DEM to the ArcSWAT interface and the watershed and sub-watershed boundary were delineated. However, the basin was divided into four major watersheds based on the climatic zones namely, Yobe-Komadugu, Magay-Ngadda, Chari-Logone and Bodou-Dillia sub-basins (**Figure 6.1**).

The main river networks and tributaries were generated based on threshold drainage area of 3000 km<sup>2</sup> and all connected to Lake Chad. The HRUs' adjusted threshold of soil type, land use and the slope were set at 15% respectively to fairly retained the characteristic of the land use features and slope classes of 0 – 2%, 2 – 8%, 8 – 15% and > 15% respectively. The catchment was discretized into 315 sub-basins, with a sub-division of 1702 HRUs' in **Table 6.2**.

*Table 6.2: SWAT model basin delineation*

Basin	Area (km <sup>2</sup> )	No of Sub-basin	No of HRU
Yobe-Komadugu	145908.9	30	160
Magay-Ngadda	84793.1	27	171
Chari-Logone	739129.4	91	572
Bodou-Dillia	1327055	167	799
<b>Total</b>	<b>2296886</b>	<b>315</b>	<b>1702</b>

The Boruta random forest filter was integrated to optimize the climate dataset used in this study. Initially, 16 GCMs' datasets at daily time step in **Table 5.1** were parsed through the algorithms by utilizing the observed gridded and GCM dataset as the target and dependent features respectively, at the 210 grid points considered, to screen and extract the significant input features (GCMs). The optimization process (i.e., inputs and target) features are considered to be statistically independent if the lagged values delay them.

The algorithm computes the Z-score of all input predictors and the distribution determines the shadow features generated from the target variable. At each grid point, an input feature is deemed important, if and only if the feature importance score (Z-score) is greater than the shadow attributes generated from the target feature after 500 iterations.

The ensemble of the four best GCMs at each grid point was formed for both baseline 1979 – 2011 and the projected climate change scenarios SSP2(4.5) and SSP5(8.5) at two-time slices of 2021 – 2050 and 2051 – 2080 and integrated into the hydrologic model. The optimization process is important to screen through antecedent lagged memories within the datasets (GCMs inputs) after the application of the algorithms could result in a potential correlation in time series arising from hydro-meteorological factors without necessarily misrepresenting the basin climate features.

The unavailability of observed wind speed, relative humidity, solar radiation and reservoir operation data, default model values were maintained, and the influence of the reservoir was neglected respectively. Hargreaves temperature-based approach was set up within the model in the simulation of evapotranspiration variable to prevent the influence of the aforementioned weather data in ET simulation.

#### **6.4.3.2 Model Calibration, Validation and Uncertainty Analysis**

The four watersheds were calibrated and validated using SUFI-2 optimization algorithms against the observed ET data extracted at 100 points and reaggregated to form 59 test points

based on the delineated watershed boundaries and with a balanced spatial distribution that covers the entire basin to increase confidence in the model output.

The primary objective was to identify sensitive model parameters in the watershed that controls the basin hydrology. Preselection of sensitive parameters was undertaken through a literature review (Abbaspour et al., 2017; Jiang et al., 2020; López et al., 2017), and one parameter at a time sensitive analysis using 5 simulation runs was conducted.

The observed ET data were sub-divided for calibration and validation at a point having a homogenous representation of ET characteristics capturing both wet, moderate and dry years across the available data period. The sensitivity analysis was used to assess the statistical significance of the model parameters estimated based on t-stat and p-value.

The uncertainty in the simulation was narrowed by identifying the range of parameters that reduced the effects in the model variable output. The model output was estimated by 95% prediction uncertainty (95PPU) determined at 97.5% and 2.5% levels of confidence. The performance of the integrated model was evaluated by the goodness of fit criteria as shown in Eqn. (6.2) and (6.3), and uncertainty range with P-factor and R-factor.

The P-factor is the percentage of observed data enclosed within the 95PPU band and R-factor is the ratio of the average thickness of the 95PPU band to the standard deviation of observed and simulated data. The optimum value of 1 and 0 respectively, indicated a perfect model.

The overall period used for analysis was (1979 – 1999) for calibration (the first three years were used as a warm-up period to prevent the effects of the unknown initial condition and (2000 – 2006) for validation. The model results of the calibrated and validated ET and the most sensitive parameter for each watershed are presented and explained in **6.5.1**.

#### 6.4.4 Assessment of water footprint environmental sustainability

The output of the hydrologic model was used to evaluate the impact of climate change on spatial and temporal variation of green and blue water footprint environmental sustainability of the Yobe-Komadugu watershed. The watershed is dominated by agricultural land and situated within the two-climate extreme of the basin.

Geographic hotspots that lead to water resource conflict was identified by defining the environmental sustainability of blue and green water at the basin size in relation to freshwater provision levels (threshold available water for human use).

To evaluate the environmental sustainability, we used a sustainability index (SI), which compared specific sub-basin WF to its corresponding water availability (WA) based on the equation below:

$$SI_{blue}^{i,j} = 1 - \frac{WF_{blue}^{i,j}}{WA_{blue}^{i,j}} \quad (6.4)$$

$$SI_{green}^{i,j} = 1 - \frac{WF_{green}^{i,j}}{WA_{green}^{i,j}} \quad (6.5)$$

Here,  $SI_{blue}^{i,j}$  and  $SI_{green}^{i,j}$  represents the indices that defines watershed blue and green water environmental sustainability in sub-basins  $i$  at time  $j$ ;  $WF_{blue}$ ,  $WF_{green}$ ,  $WA_{blue}$  and  $WA_{green}$  represents blue and green water footprint and availability respectively. When the blue and green water footprints exceed the availability, i.e., ( $SI_{blue}^{i,j} < 0$ ) and ( $SI_{green}^{i,j} < 0$ ), then water footprint is environmentally unsustainable because human water use violates environmental flow requirements and ecosystem needs respectively (Hoekstra et al., 2011).

Here, we categorize the green and blue water sustainability threshold into extremely (ES), ( $0.75 \leq SI \leq 1$ ), highly (HS), ( $0.5 \leq SI < 0.75$ ), and moderately (MS), ( $0.0 \leq SI < 0.5$ ) sustainable indices, referred to as viable water security points and extremely (EU), ( $SI <$

–1), highly (HU), ( $-1 \leq SI < -0.5$ ) and moderately (MU), ( $-0.5 \leq SI < 0.0$ ) unsustainable indices, which are referred and identified as high, medium and low-risk geographic water security hotspots respectively.

#### **6.4.4.1 Blue water footprint and availability assessment**

The blue water is determined from the output of the modelling framework (**Figure 6.3**). The blue water was estimated by combining groundwater storage and water yield (WYLD) referred to as blue water flow (BWF). The water yield (WYLD) defines the threshold amount of water that leaves the HRU and enters the main channel and groundwater storage is the difference between aquifer recharge (GW\_RCHG) and the main channel flow (GW-Q) (Rodrigues et al., 2014).

The basin blue water security evaluated by the sustainability indicators is in terms of blue water footprint or water abstractions restriction based on satisfying absolute environmental demand i.e., the concept of both abstraction (demand) and consumption (withdrawal minus return flow).

The blue water footprints were referred to as water appropriated or consumed by different sectors at the river basin scale and the spatial distribution of water uses was determined by sectoral water demand information (**Table 6.3**), 1 km gridded world population densities (CIESIN) data for baseline and projected future period consistent with the shared socioeconomic pathways of CO<sub>2</sub> emission scenarios related to the middle of the road (SSP2) and fossil-fuelled development (SSP5) available at <http://sedac.ciesin.columbia.edu/gpw> (Balk et al., 2006; Jones and O’Neill, 2016).

A conservative value of 92 L/capita/day was used to quantify absolute basic water consumption for domestic blue water footprint to meet minimum target during stringent water restrictions (Crouch et al., 2021). This concept was adopted here and can be applied to basins



where actual sectoral water demand information cannot be established or is inadequate for long-term water security assessment at the basin scale.

**Table 6.3:** Sectoral Water Use Information in Chad and Nigeria

Sector	Chad	Nigeria
Domestic Use	21%	58%
Industrial Use	0	4%
Agricultural Use	79%	39%

Source: (GWP, 2013)

The basin's annual blue water footprint was determined based on the equation below.

$$WF_{blue}^{i,j} = \sum_{i=1}^n 365C_iA_bP_dQ_w \times 1.15741 \times 10^{-8} \text{ (m}^3/\text{s)} \quad (6.6)$$

Where:  $C_i$  proportion of sectoral water use,  $A_b$  area of sub-basin (km<sup>2</sup>),  $P_d$  long term means population density per square km,  $Q_w$  per capita water use (L/capita/day), n number of sectors utilizing the freshwater resources. However, due to inadequate data, the annual water withdrawal figures were divided equally over twelve months without accounting for possible monthly variations for the assessment of blue water sustainability at a monthly scale.

The blue water availability was estimated by the methodology proposed by Hoekstra et al., (2011), where  $WA_{blue}$  was determined by considering the proportion of safe natural runoff (streamflow) that is available for consumptive use at each sub-basin as shown in Eqn. (6.7).

$$WA_{blue}^{i,j} = Q_{i,j} - EFR_{i,j} \quad (6.7)$$

Where Q represents the long-term sub-basin natural runoff (streamflow) (m<sup>3</sup>/s) and EFR is the environmental flow requirement to maintain healthy river ecosystems. In this case, EFR was estimated using the presumed standard method developed by Richter et al., (2012) that

20% of the long-term mean monthly natural runoff can be made available and considered appropriate for withdrawal.

$$EFR_{i,j} = 0.8Q_{mean(i,j)} \quad (6.8)$$

#### **6.4.4.2 Greenwater Footprint and Availability Assessment**

Green water has two components defined as green water flow (withdrawal) and green water storage (Availability). For evaporation, transpiration, or absorption, plants use the green water withdrawal that is held in the soil's root zone. According to the HRU output of the SWAT model, the green water withdrawal represents actual evapotranspiration and is defined as the green water footprint (Rodrigues et al., 2014; Veetil and Mishra, 2016).

The amount of soil moisture that can support crop development and soil evapotranspiration, which represents the original soil water content (SW) is referred to as "green water availability." It was acquired from the output of the SWAT model and applied to the water sustainability assessment (Abbaspour et al., 2015; Veetil and Mishra, 2018).

### **6.5 Results and discussions**

#### **6.5.1 Calibration and Validation of the integrated model**

The model calibration and validation process are challenging and to a certain extent subjective in complex hydrology, especially in a region with inadequate multi-variable observed data. We, therefore, aim to produce a model whose simulation reflects the natural conditions of the watershed. As a first step, we integrated the Boruta random forest feature selection approach as an interface to assess and filter out redundant downscaled GCM data across the 210 selected grid points of the entire watershed.

According to Lawal et al., (2023), this procedure was required to improve and preserve the internal variability of climate data signals that may be affected by reparameterization to utilise the right number of GCM ensembles capable of evaluating the complex interactions within hydrologic models and ensure all uncertainties (conceptual model, input data, and

parameters) ranges are mapped onto and bracketed by most of the observed data within the 95% prediction uncertainty range (Abbaspour et al., 2017), for an accurate understanding of long-term changes in baseline and projected watershed hydrology, especially in data-sparse and climate-sensitive regions which are not adequately studied.

The one-at-a-time-sensitive analysis adopted for the preselection of sensitive model parameters was relied on here, due in part to the use of different observed data for calibration and validations process from previous hydrologic studies from watersheds with similar features around the world (Abbaspour et al., 2017; Jiang et al., 2020; López et al., 2017), variations in watershed features and homogeneous representation of the evapotranspiration characteristics capturing both wet, moderated and dry years across the available data period.

The built-in sensitivity analysis tool utilized (SUFI-2) algorithms in SWAT-CUP identified 19 parameters in the four sub-watersheds analysed, with different levels of sensitivities outlined in **Table 6.4** below, and this may have alluded to the variations of land use and land cover, terrain and slope features across the watershed.

The result of the model global sensitivity analysis of the calibration process across the four sub-watersheds analysed indicated that the combination of the parameters rendered some less sensitive in the simulation run. Thus, we categorized the level of parameter sensitivity based on the p-value of the model run as (p-value = 0) highly sensitive (\*\*), ( $0 < \text{p-value} \leq 10^{-5}$ ) moderately sensitive (\*), and (p-value  $> 10^{-5}$ ) less sensitive respectively.

The sensitivity threshold applied indicated that SCS runoff curve number for average moisture condition (CN2.mgt), moist bulk density (SOL\_BD().sol), saturated hydraulic conductivity (SOL\_K().sol), and soil evaporation compensation factor (ESCO.hru) are the most important modelling parameters in the entire watershed as shown in **Table 6.4**.

**Table 6.4:** Model sensitive parameters, ranges and best-fitted values at sub-watersheds

Parameter Name	Sub-basin Fitted Parameter values.				Parameter range
	Yobe-Komadugu	Magay-Ngadda	Chari-Logone	Bodou-Dillia	
<b>r_CN2.mgt</b>	-0.01**	0.01**	-0.15*	-0.02**	- 0.2 - 0.2
<b>v_GW_DELAY.gw</b>	88.66**	182.56	63.28	88.66**	0.0 - 500.0
<b>v_ALPHA_BF.gw</b>	0.84*	0.50	0.71	0.84*	0.0 - 1.0
<b>r_GWQMN.gw</b>	-1.37	0.04	-0.38	-1.37	0.0 - 5000
<b>v_GW_REVAP.gw</b>	0.19	0.10	0.05	0.12	0.02 - 0.2
<b>v_REVAPMN.gw</b>	199.36	-	-	-	0.0 - 500
<b>r_RCHRG_DP.gw</b>	0.80	-	-	-	0.0 - 1.0
<b>r_SOL_Z().sol</b>	-0.03	-	-	-	0.0 - 3500
<b>v_SOL_BD().sol</b>	1.08**	0.60**	1.12**	1.08**	0.9 - 2.50
<b>v_SOL_AWC().sol</b>	0.58	0.58	0.42	0.39	0.0 - 1.0
<b>v_SOL_K().sol</b>	285.93**	4.36*	1435.35**	285.93**	0.0 - 2000
<b>v_CH_N2.rte</b>	0.22	0.18	0.27*	0.17	- 0.01 - 0.3
<b>v_CH_K2.rte</b>	367.65*	260.33	387.82	367.65*	- 0.01 - 500
<b>v_ALPHA_BNK.rte</b>	0.86	0.32	0.32	0.86	0.0 - 1.0
<b>r_SLSUBBSN.hru</b>	0.21*	-0.01	0.09**	0.21*	10.0 - 150
<b>v_OV_N.hru</b>	0.05	0.05**	0.02	0.05	0.01 - 1.0
<b>v_ESCO.hru</b>	0.92**	0.27**	0.52**	0.92**	0.0 - 1.0
<b>v_EPCO.hru</b>	0.24	0.14**	0.22**	0.24	0.0 - 1.0
<b>r_HRU_SLP.hru</b>	0.05	0.05	0.27**	-0.02**	0.0 - 1.0

Note: ‘v\_’ denotes that the existing parameter value was replaced by a given value and ‘r\_’ is a relative change and implied that the existing parameter value is multiplied by (1 + given value).

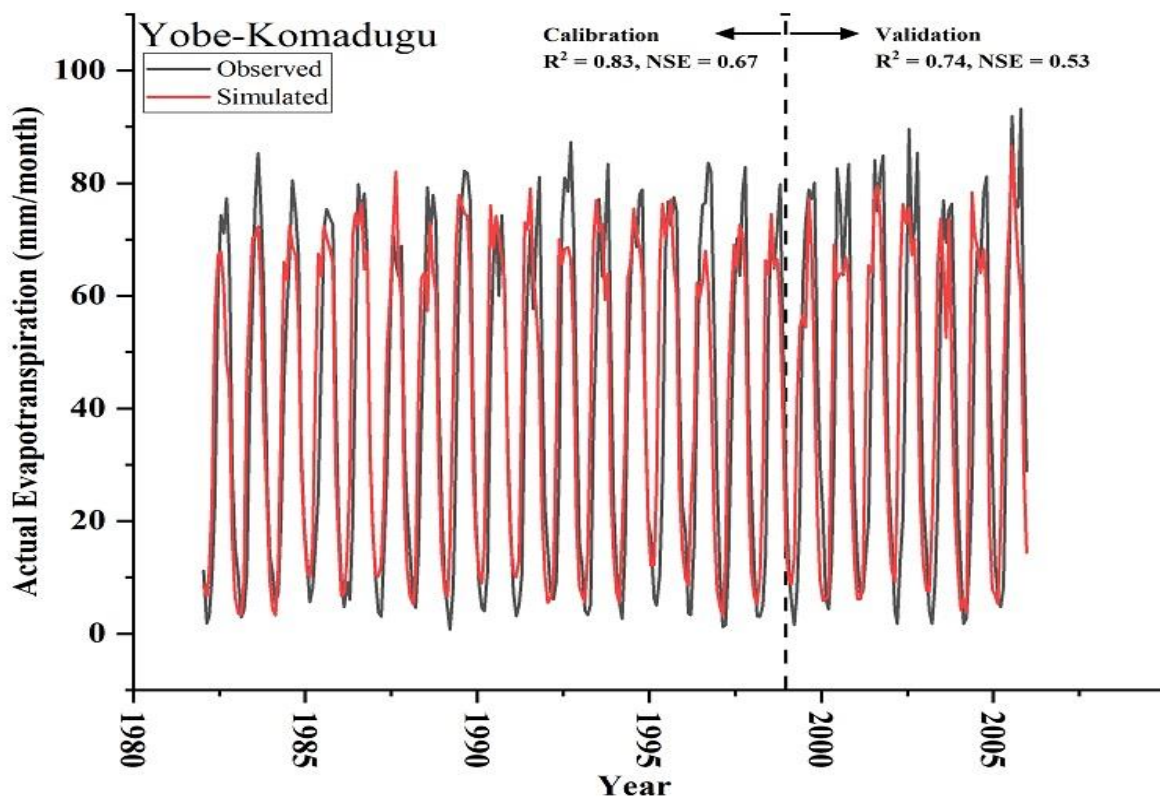
Other important sensitive parameters to note based on the sub-watershed modelling process are groundwater delay (GW\_DELAY.gw), baseflow alpha factor (ALPHA\_BF.gw), effective hydraulic conductivity in main channel alluvium (CH\_K2.rte), average slope length (SLSUBBSN.hru) in Yobe-Komadugu, manning’s roughness coefficient for overland flow (OV\_N.hru) and plant uptake compensation factor (EPCO.hru) in Magay-Ngadda, average slope steepness (HRU\_SLP.hru), baseflow alpha factor (ALPHA\_BF.gw), average slope length (SLSUBBSN.hru), plant uptake compensation factor (EPCO.hru), manning’s roughness for main channels (CH\_N2.rte) in Chari-Logone, and baseflow alpha factor (ALPHA\_BF.gw), effective hydraulic conductivity in main channel alluvium (CH\_K2.rte), average slope steepness (HRU\_SLP.hru), groundwater delay (GW\_DELAY.gw) in Bodou-Dillia watershed respectively.

However, the optimized watershed's sensitive parameter ranges were varied, and this lack of uniqueness is characteristic of the calibration of hydrologic models. This assertion was

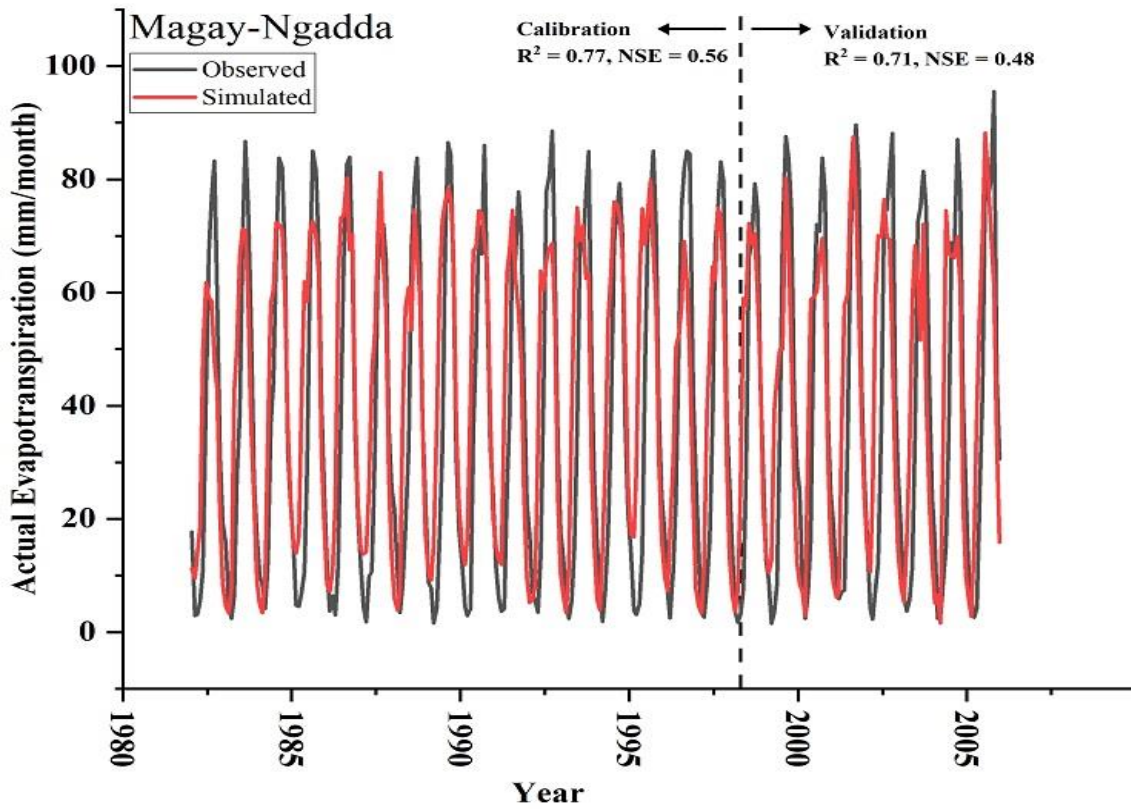
supported by Abbaspour et al., (2009), that there will be numerous such models with various parameter ranges if a model exists that fits the measurements.

The sub-watershed performance for calibration and validation respectively is shown in **Figure 6.4a-d**, calculated based on the observed and the “best” simulated monthly actual evapotranspiration value of the objective function across the 59 measured points spatially distributed across the basin.

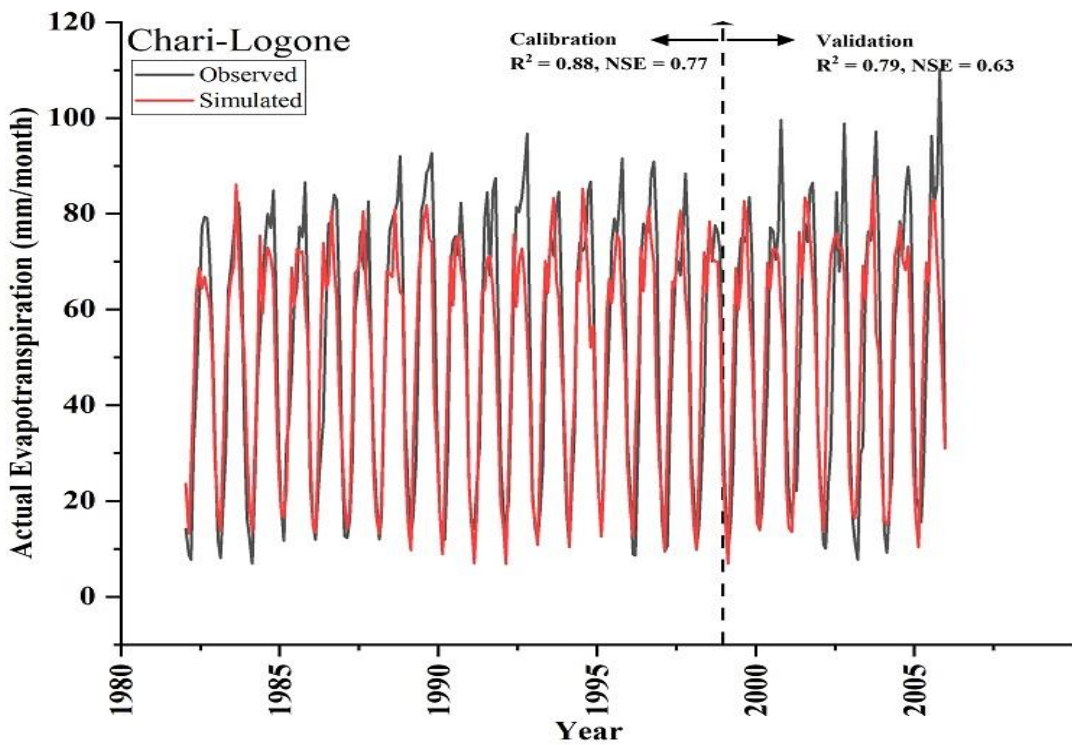
The calibrated and validated model results depicted by correlation coefficient ( $R^2$ ) and Nash-Sutcliffe efficiency criteria (NSE) were in the range of  $R^2 = 0.69 - 0.88$ ,  $NSE = 0.45 - 0.77$  and  $R^2 = 0.62 - 0.79$ ,  $NSE = 0.34 - 0.63$  across all the watersheds respectively. Moreover, a large number of the achieved model results fell within a satisfactory uncertainty range with P-factor and R-factor values in the range of 0.68 – 0.93 and 0.73 – 1.31 in 83%, 67%, 85.7% and 81.3% of the sub-watershed respectively.



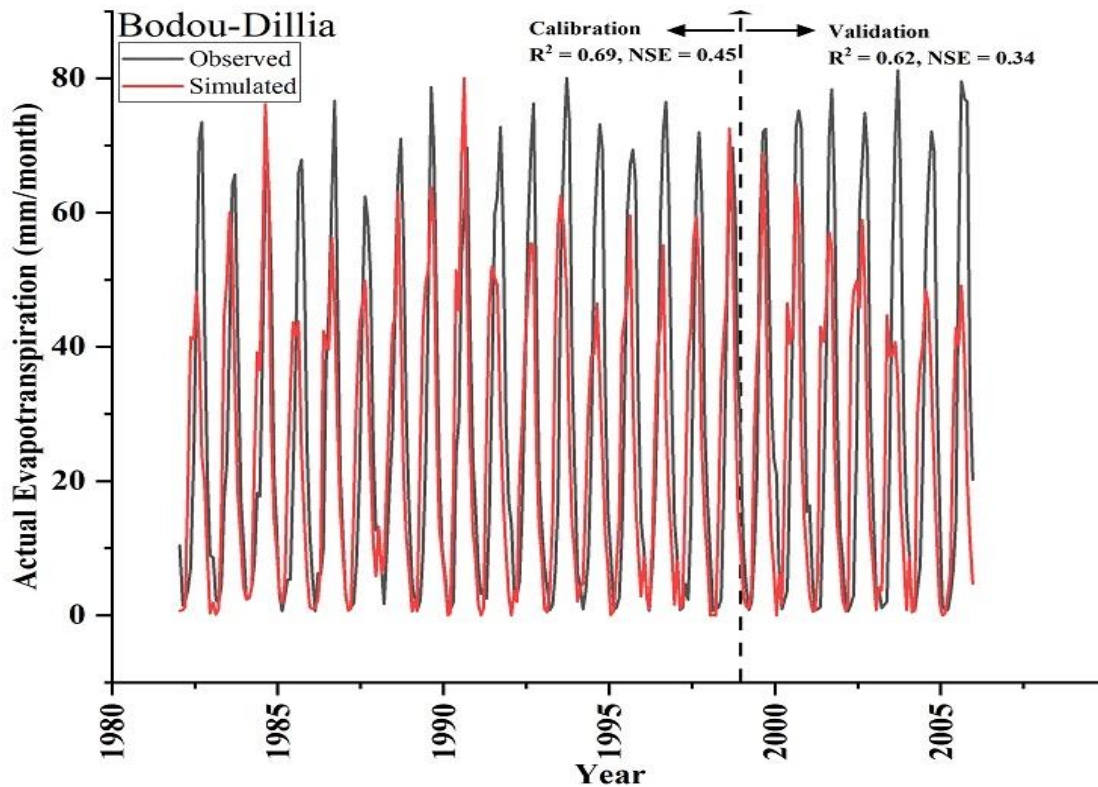
(a)



(b)



(c)



(d)

**Figure 6.4:** Plot of comparison of the observed and simulated results (95% prediction uncertainty band) of actual evapotranspiration against time between 1983 – 2006) in the Lake Chad basin. (a): Yobe-Komadugu Watershed (b): Magay-Ngadda Watershed (c): Chari-Logone Watershed (d): Bodou-Dillia.

There are a few sub-basins with poor simulated output whose  $R^2$  and NSE values are as low as 0.25 and 0.14 respectively, although has exhibited a good representation of data uncertainty band with encouraging P-factor and R-factor values in the range of 0.53 – 0.78 and 1.21 - 1.95 respectively.

Even the region with acceptable objective functions is faced with the difficulty simulating and matching the peak values of the observed evapotranspiration values and this may be alluded to simplification of the model by reaggregation of the land use features, inadequate data that accounts for some of the important basin-scale processes like lack of sufficient information such as reservoir operations, dams, water transfers, irrigation process etc.

This is generally classed as technical modelling uncertainties and natural heterogeneity in the hydrologic modelling process and has been corroborated in (Abbaspour et al., 2015;

Schuol et al., 2008). However, our results are generally quite realistic for basin-scale assessment of water-related hazards.

The obvious reason for the large variability in NSE estimates, across the four basins or model results, could be related to the “actual evapotranspiration only” calibrations. The modelling issue can be significantly “improved” by incorporating additional observation dataset in the distributed calibration modelling schemes (Kunnath-Poovakka et al., 2016; Rajib et al., 2016) where reliable data are made available. However, Koppa et al., (2019), argued that the ability of a model to simultaneously reproduce the included water balance components is not assessed by any limits of acceptability or error thresholds in multivariate calibration.

The result presented here, is a step forward and improvement to earlier studies by Faramarzi et al., (2013), and Schuol et al., (2008), using a stand-alone SWAT model with direct use of climate data, where the results from the studies indicate poor watershed representation of the portion of Lake Chad basin, which depicted large uncertainty range.

This may be attributed to the use of climate data with coarse resolution and distorted signals of watershed features where the complex orographic and land-sea distribution was not accounted for and may lead to local variation in basin water balance outputs and affects projected climate change assessment studies. However, there are differences in model variables and parameters adopted for calibration and these studies are on a wider scale.

Interestingly, the optimization approach used here by incorporating machine learning in the integrated modelling strategy could reduce large model uncertainty propagation and provide a new direction to modelling issues in data-sparse regions with variable morphological features by providing high-valued water resource information at the local basin scale to drive sustainable water policy decisions.



### 6.5.2 Assessment of climate change impact on projected green and blue water resources.

The assessment of climate change's impact on the spatial and temporal distribution of blue and green water resources will be of great significance at the sub-watershed level to provide the necessary information for decision support for water authorities.

The confidence in the output of the model results was reinforced by investigating variations in the projected mean changes in the near future (2021 – 2050) and far future (2051 – 2080) annual precipitation and average temperature from baseline (1982 – 2011) of the ensemble GCM refined by Boruta random forest feature selection approach of the Yobe-Komadugu watershed.

The results of the projected changes in annual precipitation and temperature for the two scenarios based on shared socioeconomic pathways are shown in **Table 6.5**.

*Table 6.5: Median of the projected changes in annual precipitation and temperature in the Yobe-Komadugu Watershed*

Variable	Precipitation (%)		Temperature (°C)	
	SSP2(4.5)	SSP5(8.5)	SSP2(4.5)	SSP5(8.5)
2021 - 2050	7.10	7.40	0.69	0.89
2051 - 2080	13.25	27.68	1.17	1.78

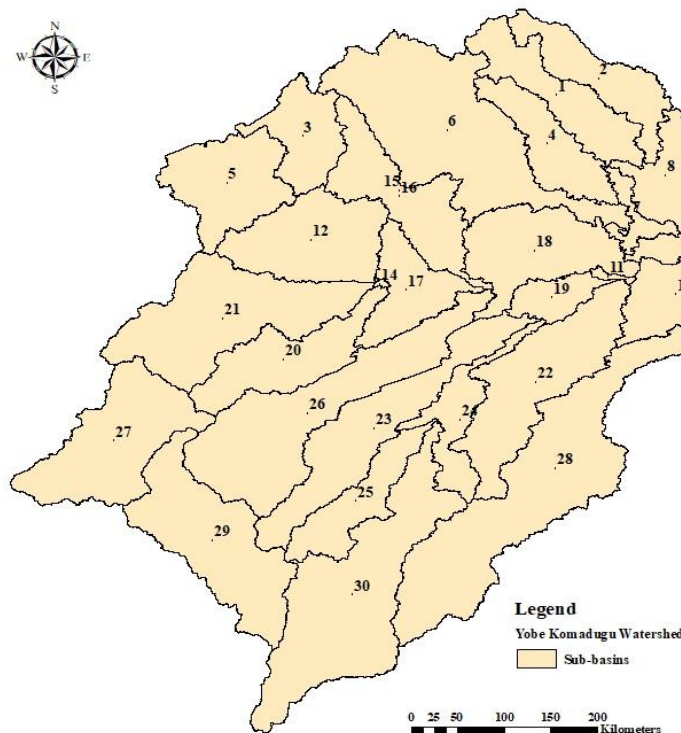
The results of the projected changes in precipitation indicated an increasing trend with an annual shift of 7.1 and 7.40% in the near future to 13.25 and 27.68% in the far future, with associated increased warming scenario of average temperature in the range of  $0.69 \pm 0.15^{\circ}\text{C}$  and  $0.89 \pm 0.11^{\circ}\text{C}$  between 2021 – 2050 and  $1.17 \pm 0.22^{\circ}\text{C}$  and  $1.78 \pm 0.24^{\circ}\text{C}$  between 2051 – 2080 for SSP2(4.5) and SSP5(8.5) respectively.

The range of projection here, is similar and consistent with reported findings of previous studies (Almazroui et al., 2020; Sylla et al., 2016; Vیزی et al., 2013), and the projected changes may be linked to variability and changes in West African Monsoon features, like changes in intensity and localisation of the African Easterly Waves and Jets, monsoon flows

as well as integrated moisture flux divergence and moist static energy (Mariotti et al., 2014; Sylla et al., 2015; Teichmann et al., 2013).

The result of the changes in spatial and temporal distribution from baseline of projected green water flow, green water storage and blue water flow under two climate change emission scenarios SSP2(4.5) and SSP5(8.5) for the near future (2021 – 2050) and far future (2051 – 2080) of Yobe-Komadugu watershed is displayed in **Figure 6.6 – Figure 6.11**, respectively.

The watershed was chosen because it is characterized by incidences of climate extremes, the most recent and notable event was the reported heavy windstorm in April 2022, and downpour in May 2022 that affected about 180 communities and resulted in the loss of lives, food, buildings, livestock, farmlands etc. (SEMA, 2022), and it is an important agriculture production regions. The delineated watershed boundary and the sub-basins are shown in **Figure 6.5**.



**Figure 6.5:** Delineated Yobe Komadugu Watershed with sub-basin locations

### 6.5.2.1 Spatial and temporal variation of green water flow under different climate change scenarios

The hydrological cycle is expected to intensify due to increased rainfall and a warmer atmosphere, as evidenced by the projected increase in atmospheric temperature as a result of CO<sub>2</sub> emissions, which indicates a greater evaporative demand and increases GWF and this is consistent with the findings of Ogutu et al., (2021); Pham-Duc et al., (2020); and Todzo et al., (2020).

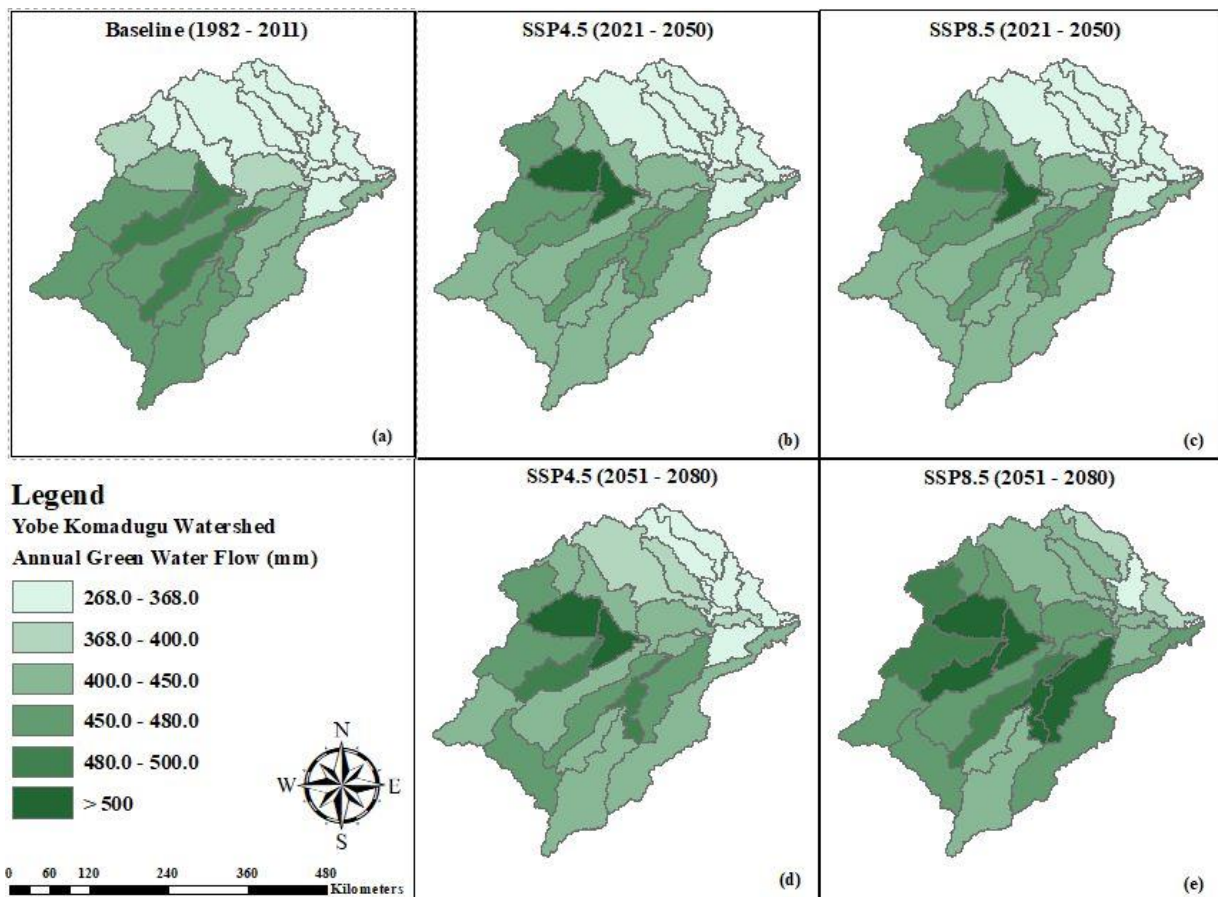
It is projected to increase at different levels based on the emission scenario and time slices. For example, the baseline period (1982 – 2011) depicted a mean annual GWF of 393.55 mm for the entire basin in **Figure 6.6a**. However, there is a marked increase in the spatial variations in GWF from 417.02 mm to 425.03 mm for SSP2(4.5) in **Figure 6.6b** and **d**, and from 418.75 mm to 457.86 mm in **Figure 6.6c** and **e**, for SSP5(8.5) emission scenario accounting for 6.0% and 8.0% in basin GWF for SSP2(4.5) and 6.4% and 16.34% increased basin GWF for SSP5(8.5) in near and far future time slices respectively.

Few exceptions were noted with contrasting GWF hydrologic features where declining GWF were predicted especially in the downstream (sub-basin 20, 21, 25 – 30) of the watershed. These changes are consistent with the studies indicating a general increase in evaporative demand due to increase air temperature in most part of the world ((Scheff and Frierson, 2014; Vicente-Serrano et al., 2020).

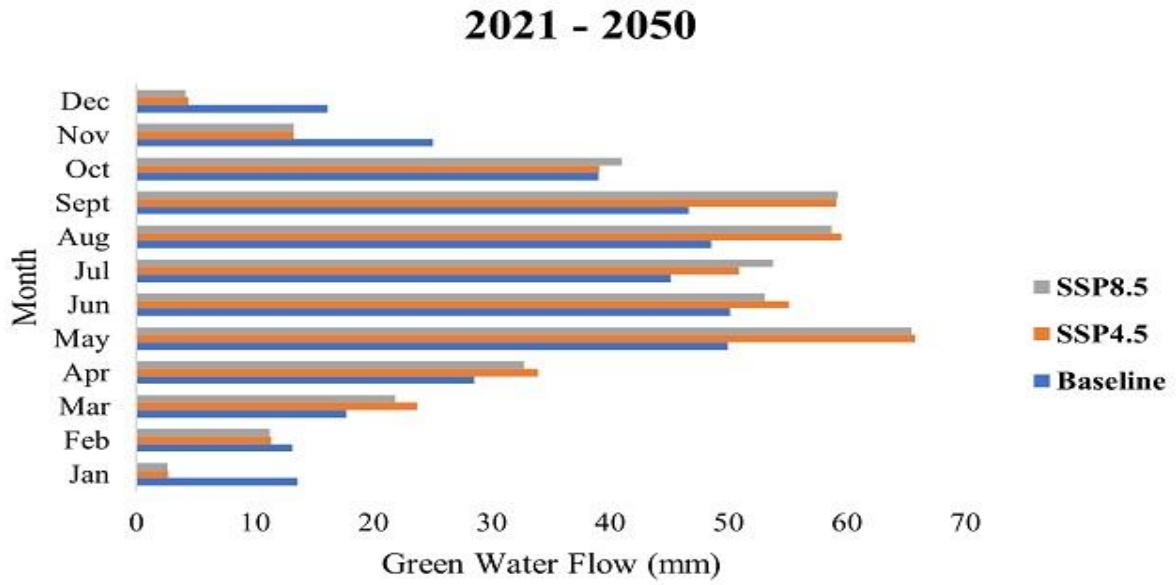
Analysis of the distribution and changes in the mean monthly variation of GWF of the near and far future relative to the baseline period, depicted in **Figure 6.7**, showed a consistent projected increase between spring and summer months in the range of 12.95 – 33.54% and 5.93 – 31.02% (**Figure 6.7a**), between 2021 – 2050 and 23.25 – 65.76% and 26.39 – 87.43% in **Figure 6.7b**, between 2051 – 2080 for SSP2(4.5) and SSP5(8.5) respectively.

However, there is a generally sharp decline in autumn and winter season across the basin projected to be around 53.38% and 54.10% in 2021 – 2050 and 54.72% and 36.0% based on the two emission scenarios within the time slices. The reason for the enhanced projected GWF may be related to the increased temperature in the tropical regions between April through September due to an increase in CO<sub>2</sub> emission concentration.

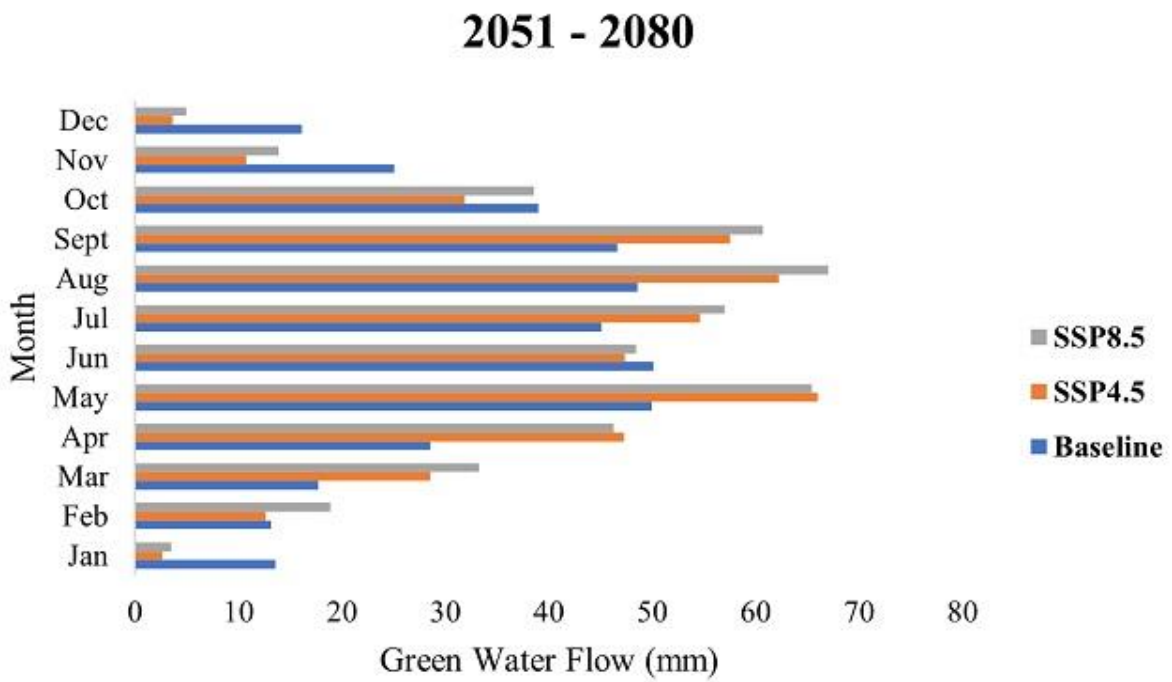
The projected increase in precipitation events also enhances vegetation cover and the activity of plant actual transpiration. The variation in GWF across the seasons is generally more pronounced in regions with high atmospheric evaporative demands that give rise to a pattern of seasonally variable regimes consistent with the findings of Konapala et al., (2020).



**Figure 6.6:** Variations in (mm) of spatial distribution of annual green water flow in the Yobe-Komadugu Watershed.



(a)



(b)

**Figure 6.7:** Variations (mm) in the temporal distribution of mean monthly green water flow (a) 2021 – 2050 (b) 2051 – 2080 in Yobe-Komadugu Watershed.

### 6.5.2.2 Spatial and temporal variation of green water storage under different climate change scenarios

The result of the mean annual GWS, represented by the soil moisture conditions, which changes over time and could be affected by the initial soil water content indicated a substantial projected decline in all sub-basins of the watershed from the baseline period with an annual average value of 341.89 mm (**Figure 6.8a**), to 324.79 mm and 302.43 mm for SSP2(4.5) in **Figure 6.8b** and d, and 299.45 mm and 293.45 mm for SSP5(8.5) in **Figure 6.8c** and e. This accounts for the projected decline of 4.99% and 11.54% in basin GWS for SSP2(4.5) and 12.41% and 14.17% for SSP5(8.5) in the near and far future period respectively.

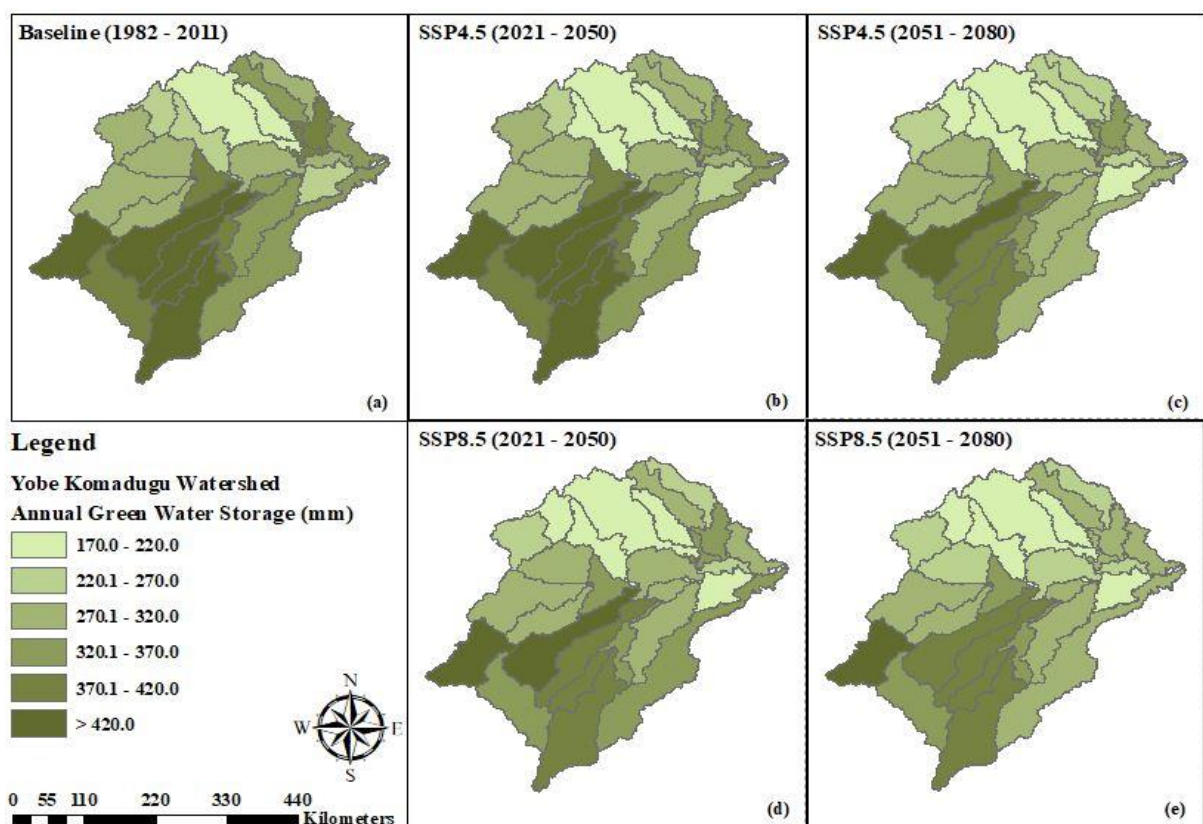
This decline may not be unconnected to the huge over-exploitation of groundwater resources for irrigation practices, by further lowering the water table level in the basin and possibly increasing surface air temperature could also affect soil water flow regimes thereby increasing the groundwater evaporative demands.

The variations and decline of soil moisture may be attributed to the likely increase in drought occurrences in regions that are currently drought prone areas and is projected to decrease soil moisture conditions in high emission scenarios as alluded by Cook et al., (2020).

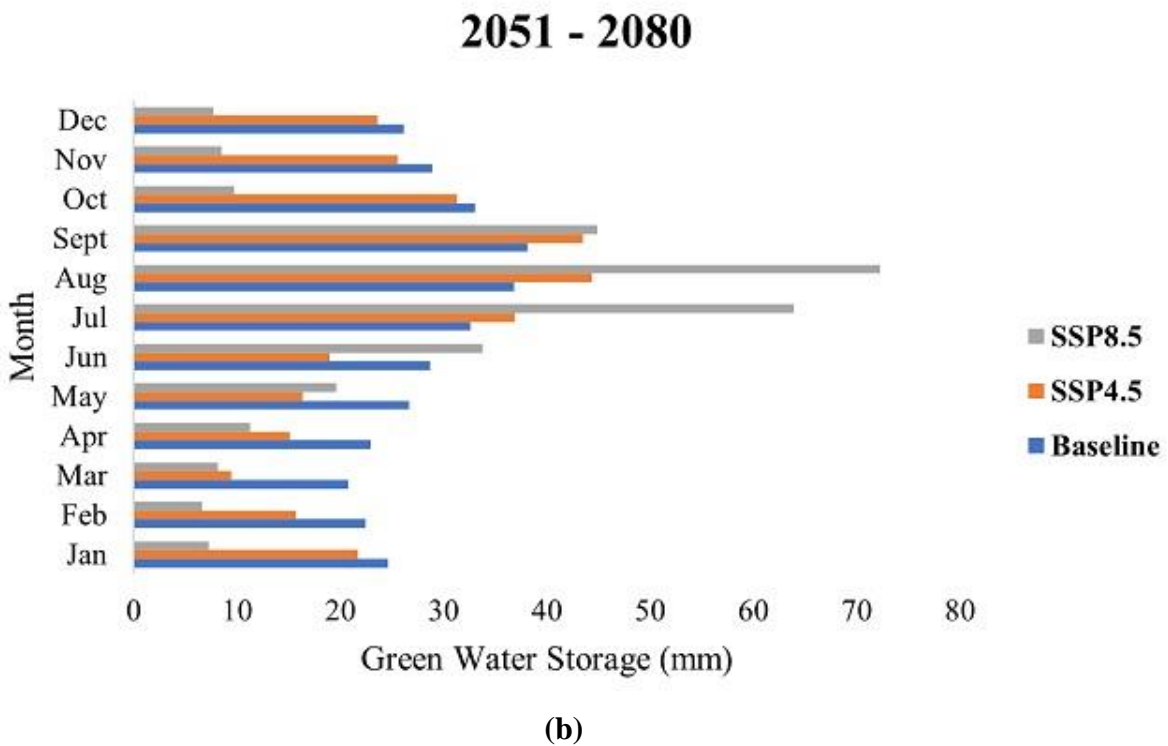
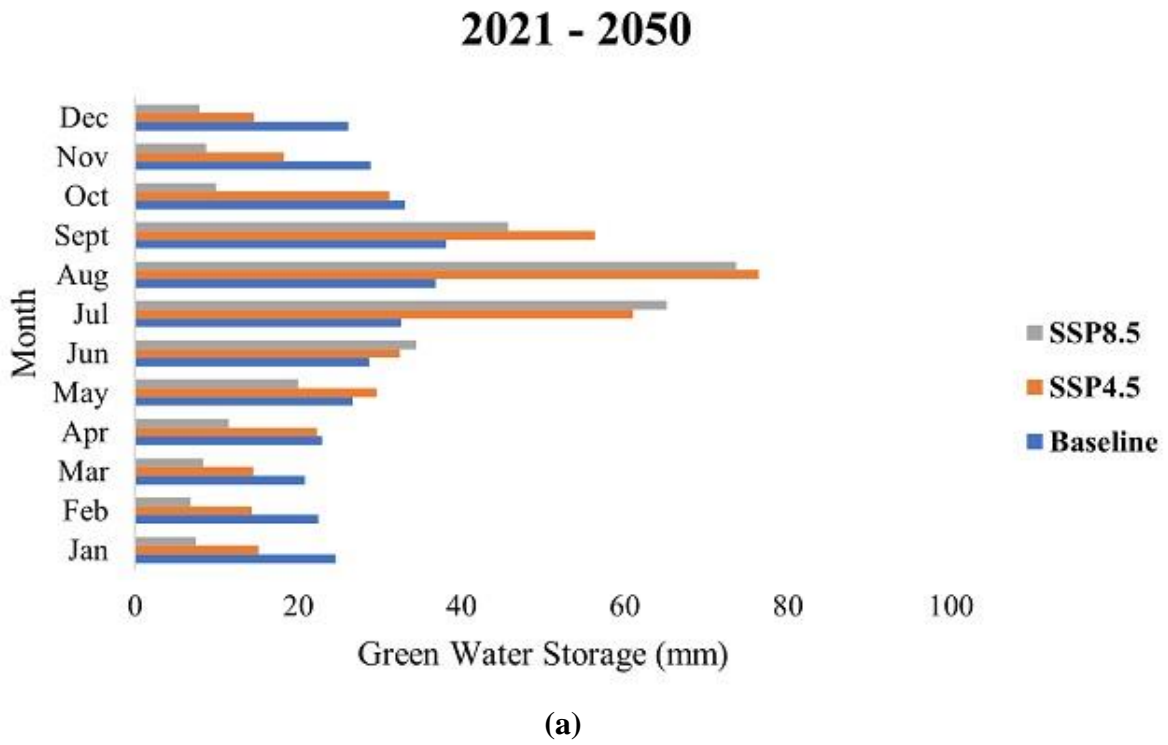
Analysis of the distribution and changes in the mean monthly variation of GWS of the near and far future relative to the baseline scenario, depicted in **Figure 6.9**, showed a consistent projected decline in most months in the range of 2.75 – 44.11% at a mean rate of 6.95 mm/month and 24.97 – 69.99% at a mean rate of 15.65 mm/month (**Figure 6.9a**), in 2021 – 2050 and 5.47 – 54.45% at a mean rate of 6.28 mm/month and 26.47 – 70.59% at a mean rate of 15.85 mm/month (**Figure 6.9b**), in 2051 – 2080 for SSP2(4.5) and SSP5(8.5) respectively.

However, there is an exception in the monsoon (May – September) season across the basin, with a projected increase of GWS of around 53.28% (18.58 mm/month) and 60.02% (20.71 mm/month) in 2021 – 2050 and (June - September) at 15.87% (5.72 mm/month) and 56.83% (23.91 mm/month) in 2051 – 2080 based on the two emission scenarios respectively.

The projected increase in the monsoon season is generally significant between the month of July – September which is associated with high rainfall intensities and interannual seasonal variability as corroborated by Almazroui et al., (2020).



**Figure 6.8:** Variations in (mm) of spatial distribution of green water storage in the Yobe-Komadugu Watershed.



**Figure 6.9:** Variations (mm) in the temporal distribution of mean monthly green water storage (a) 2021 – 2050 (b) 2051 – 2080 in Yobe-Komadugu Watershed.



### 6.5.2.3 Spatial and temporal variation of blue water flow under different climate change scenarios

The climate change impact on spatial and temporal variation of blue water flow was quantified at the sub-basin level in the watershed. BWF showed high variability, and the dynamics are quite distinct in the upstream and downstream parts of the watershed.

For example, BWF is projected to decline at a mean annual rate of 38.9 mm/year and 37.25 mm/year at the sub-basin (1 – 13, 15, 18, and 22) upstream, while an associated projected increase of 54.66 mm/year and 55.27 mm/year at sub-basin (14, 16, 17, 19 – 21, and 23 – 30) downstream, in 2021 – 2050 shown in **Figure 6.10b** and c, for SSP2(4.5) and SSP5(8.5) respectively, from the baseline period (1982 – 2011) depicted a mean annual BWF of 37.83 mm for the entire basin in **Figure 6.10a**.

Similarly, the dynamics are the same for the far future but with a reduced magnitude of decline from the baseline of 25.98 mm/year and 29.69 mm/year upstream and increase the magnitude of 77.23 mm/year and 98.97 mm/year downstream in 2051 – 2080 shown in **Figure 6.10d** and e, for SSP2(4.5) and SSP5(8.5) emission scenarios respectively.

However, analysis of changes in BWF in the entire watershed depicted a projected increase from the baseline period of 2.85 mm/year and 4.76 mm/year in 2021 – 2050 and 20.21 mm/year and 52.01 mm/year in 2051 – 2080 for the respective CO<sub>2</sub> emissions scenarios.

Analysis of the distribution and changes in the mean monthly variation of BWF of the near and far future relative to the baseline scenario, depicted in **Figure 6.11**, showed that the projected decline is prevalent between months in the winter and spring season shown in **Figure 6.11a** and b, where precipitation events are non-existent or sub-optimal (i.e. below long term basin average) in the tropical regions.

However, the summer and autumn months showed a projected increase in BWF relative to the baseline period which may be associated with increased monsoon rainfall events and intensities thereby intensification of wet extremes and dry spell lengths by shortening the Sahel rainy seasons as predicted in previous studies (Almazroui et al., 2020; Sarr, 2012; Sylla et al., 2016).

The projected decline of BWF is in the range of 0.17 – 4.88 mm/month and 0.17 – 6.42 mm/month in **Figure 6.11a**, between 2021 – 2050 and 0.13 – 6.0 mm/month and 0.07 – 5.49 mm/month in **Figure 6.11b**, between 2051 – 2080 for SSP2(4.5) and SSP5(8.5) respectively.

However, there is a generally sharp increase in BWF in monsoon season across the basin especially in August, with a projection of up to 4.76 and 4.96 mm/month between 2021 – 2050 and 11.66 and 23.8 mm/month between 2051 - 2080 based on the two emission scenarios.

These sharp changes in BWF across the months validate the changes seen in recent trends in significant increase in heavy rainfall events and changes in seasonality that exacerbated incidences of frequent weather extremes i.e., flooding and droughts in the Sahel (Boko et al., 2007; Niang et al., 2014).

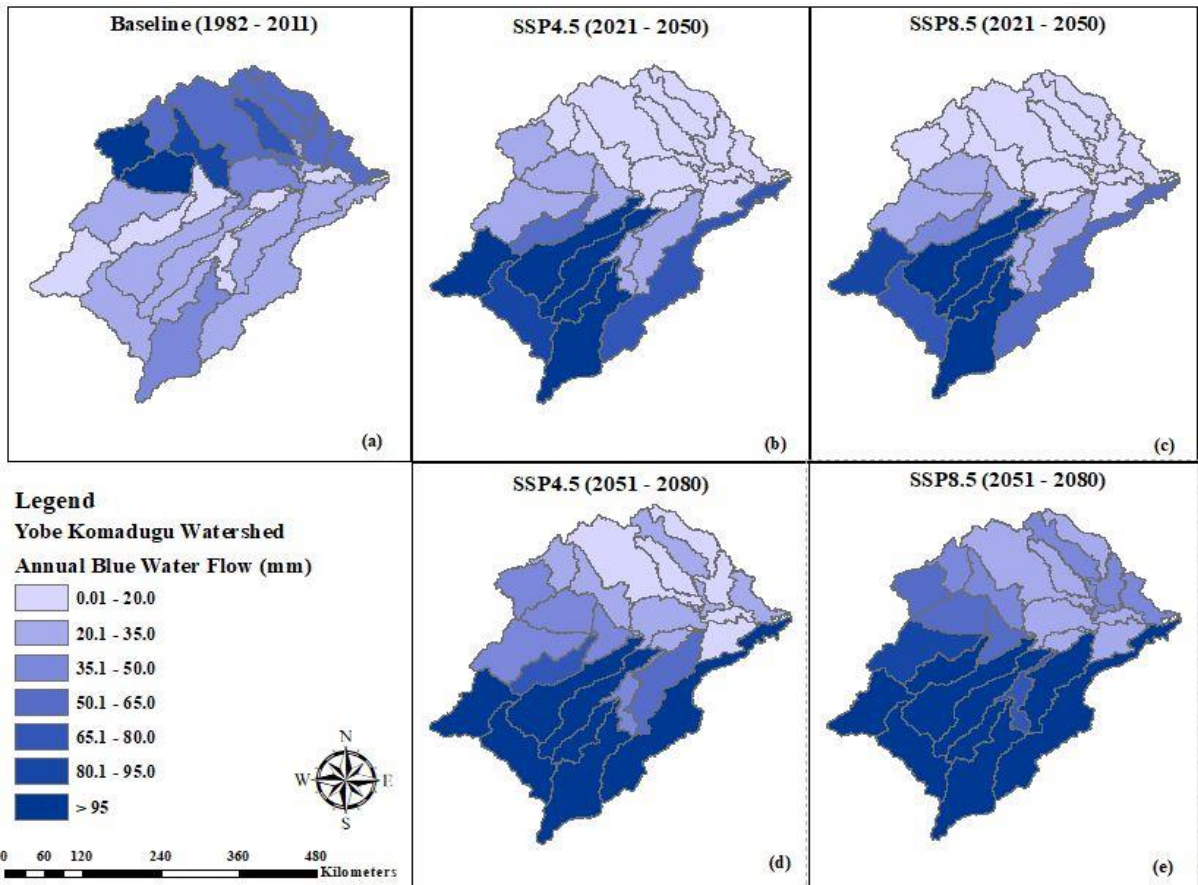
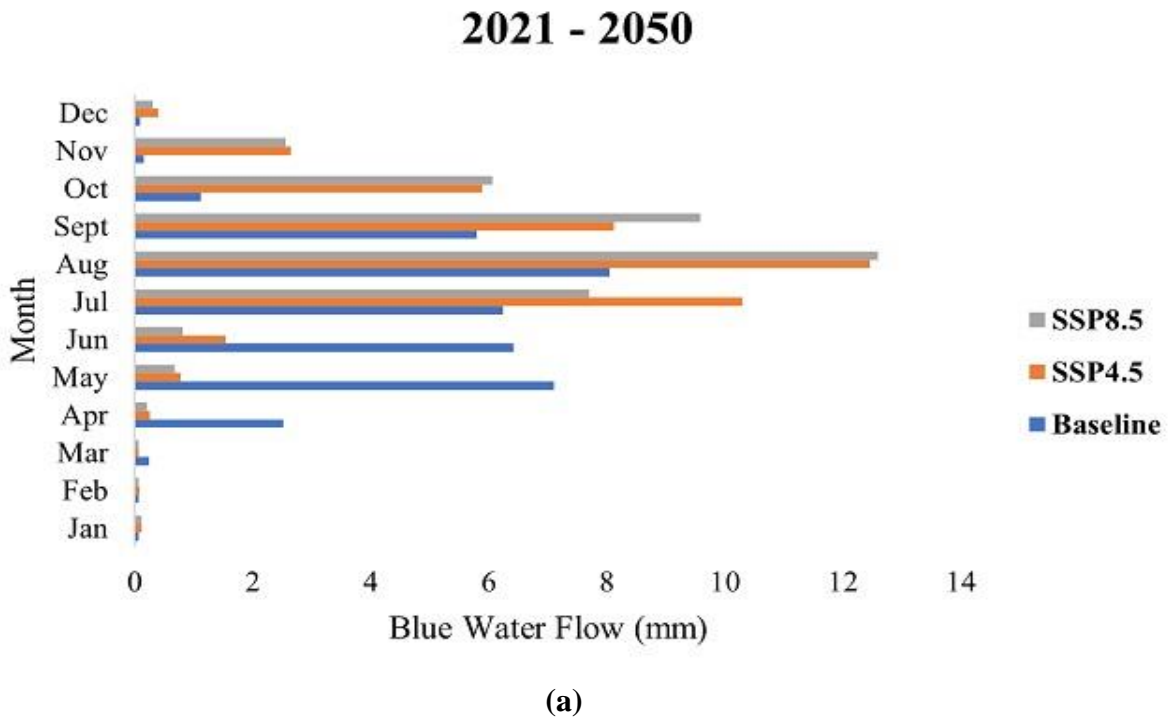
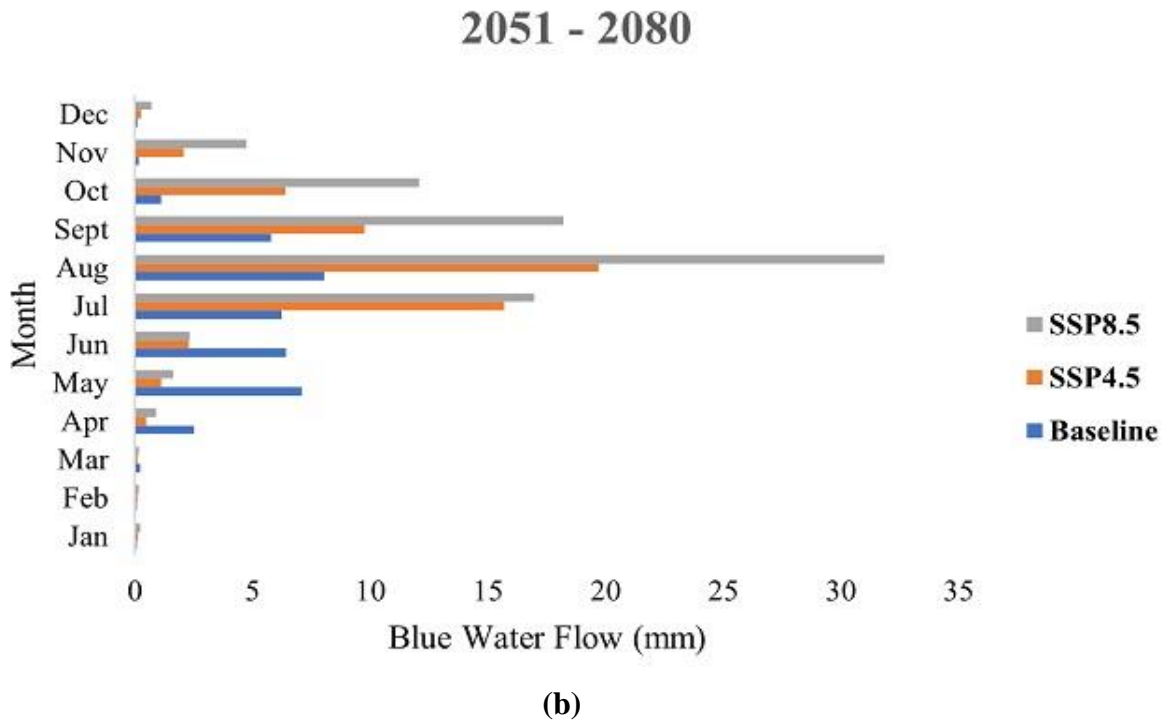


Figure 6.10: Variations (mm) in the spatial distribution of blue water flow in the Yobe-Komadugu watershed





**Figure 6.11:** Variations (mm) in the temporal distribution of mean monthly blue water flow (a) 2021 – 2050 (b) 2051 – 2080 in Yobe-Komadugu Watershed.

### 6.5.3 Climate change impact and socioeconomic drivers on spatial variation of projected green and blue water sustainability.

The green and blue water sustainability was determined at the sub-basin scale for baseline (1982 – 2011) and projected changes in the near (2021 – 2050) and far (2051 – 2080) future based on the two CO<sub>2</sub> emission scenarios using the sustainability indices as shown by the spatial maps in **Figure 6.12** and **Figure 6.13** respectively.

The baseline period showed that green water is moderate to extremely sustainable (ES), in 7 sub-basins accounting for 16.50% of the watershed area with sustainability indices (SI) ranging from 0.19 to 0.3, 0.5 – 0.71 and 0.81 – 1.0 in sub-basin (10 and 27), (7 and 26) and (1, 2, and 8) in **Figure 6.12a**. The remainder of the watershed was characterized with a low level of green water sustainability, except sub-basin 3, 20 and 24 which are high-risk geographic hotspots.

The favourable sustainability indices of the sub-basins located upstream of the watershed may be due to land use and land cover features which is a mixture of scanty vegetation and

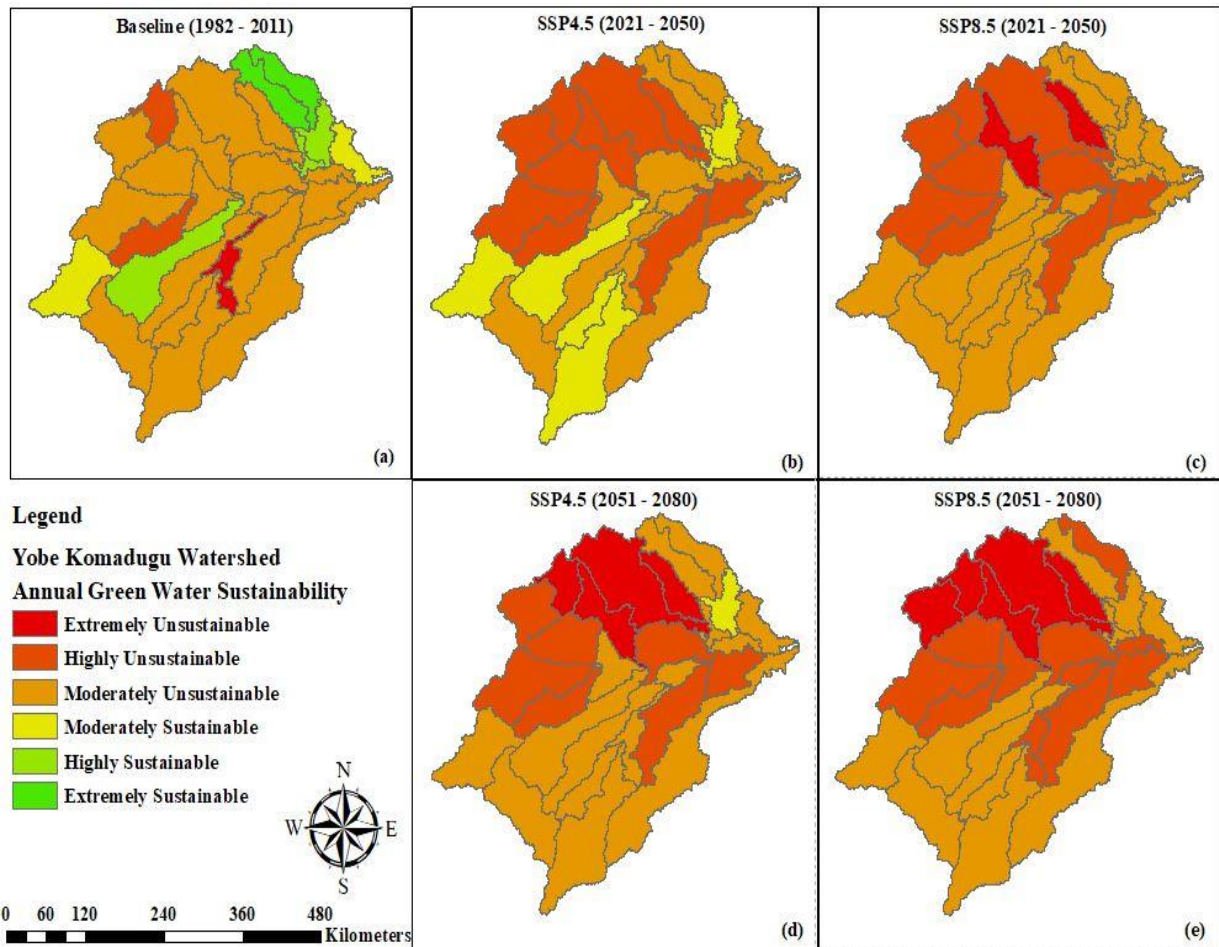
bare land with associated low green water footprint as a result of little to non-existent rainfed agricultural practices.

Similarly, plant water use efficiency due to climate change may be a significant influence in the differences of freshwater sustainability status between the upstream and downstream part of the watershed, which may have limit or counteract the rising projected evapotranspiration in high emission scenarios as corroborated in (Lemordant et al., 2018; Milly and Dunne, 2016).

Analysis of the projected green water sustainability indicated that there is a one-to-two-fold shift in sustainability threshold across the basin with a steady to sharp decline from a favourably basin green water sustainability status from the baseline of 16.50% to 15.9% for SSP2(4.5) (**Figure 6.12b**), and 0% for SSP5(8.5) in 2021 – 2050 (**Figure 6.12c**), and the far future also indicated a decline of watershed green water sustainability threshold of 1.86% for SSP2(4.5) (**Figure 6.12d**) and 0% for SSP5(8.5) (**Figure 6.12e**) in 2051 – 2080 of the watershed area respectively.

The geographic hotspots ( $SI < -0.5$ ) are generally situated upstream of the watershed in all scenarios and are an indication that climate change may have a more profound effect on the high to extremely unsustainable green water status which is evident in the continuous increase in green water flow and decreased green water storage.

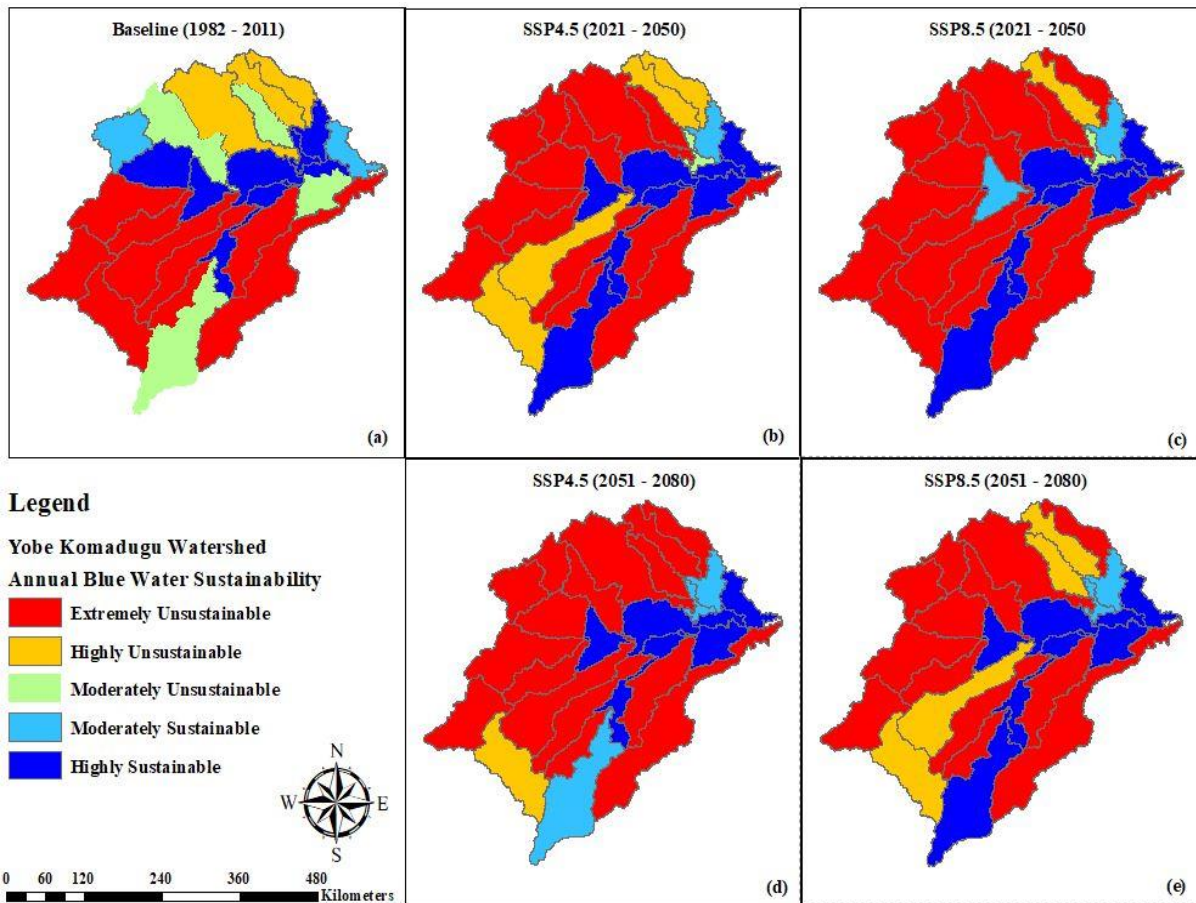
This phenomenon may have been the cause of increased humidity, affecting the timing, spatial pattern and intensity of rainfall in a basin as suggested by Du et al., (2018), and as CO<sub>2</sub> emissions rise, the efficiency of the photosynthetic process' utilisation of water increases, resulting in CO<sub>2</sub> fertilisation (Donohue et al., 2017).



*Figure 6.12: Spatial hazards map of changes of baseline and projected green water environmental sustainability in Yobe-Komadugu Watershed.*

The changes in blue water sustainability in the watershed for the baseline period in **Figure 6.13a** were assessed to be 15.61% HS, 5.4% MS, 15.65% MU and 63.34% HU – EU (potential blue water geographic hotspots) of the watershed area respectively.

The high level of blue water sustainability is predominant upstream of the watershed, however, sub-basins 1, 2, and 6 are shown to be highly unsustainable which may be related to the absence of viable stream channels and high rate of evaporative demands which characterized the basin as semi-arid with severe droughts events and high interannual rainfall variability due to the effect of Inter-Tropical Convergence Zone (ITCZ) migration (Thompson and Polet, 2000).



**Figure 6.13:** Spatial hazards map of changes of baseline and projected blue water environmental sustainability in Yobe-Komadugu Watershed.

Analysis of the influence of climate change and changes in socio-economic activities on projected blue water sustainability indicated a further increase in blue water geographic hotspots across the watershed area of 71.53% and 75.38% in **Figure 6.13b** and **c** between 2021 – 2050, and 73.51% and 76.35% in **Figure 6.13d** and **e** between 2051 – 2080 for SSP2(4.5) and SSP5(8.5) respectively.

Our model results showed that the blue water security hotspots regions have negative SI in the range of – 0.5 up to as high as – 16.58 for both SSP2(4.5) and SSP5(8.5) respectively. The blue water's continued unviability may be caused by major river systems drying up and reduced flows brought on by the overuse of groundwater and surface water resources as a result of intensive irrigation practices.

These consistent patterns could be scaled with the SSPs emissions scenarios, which have shown a strong correlation between anthropogenic GHG emissions and potential environmental impacts as corroborated by Adeyeri et al., (2019).

Some of the viable blue water sustainable sub-basins are characterised by interconnected large streams that form the Komadugu Yobe and Komadugu Gana river sub-systems that support different ecological processes and socioeconomic activities such as fish production, pastoralism, forest regeneration etc., with a population of over 20 million people depending on this activity in the basin.

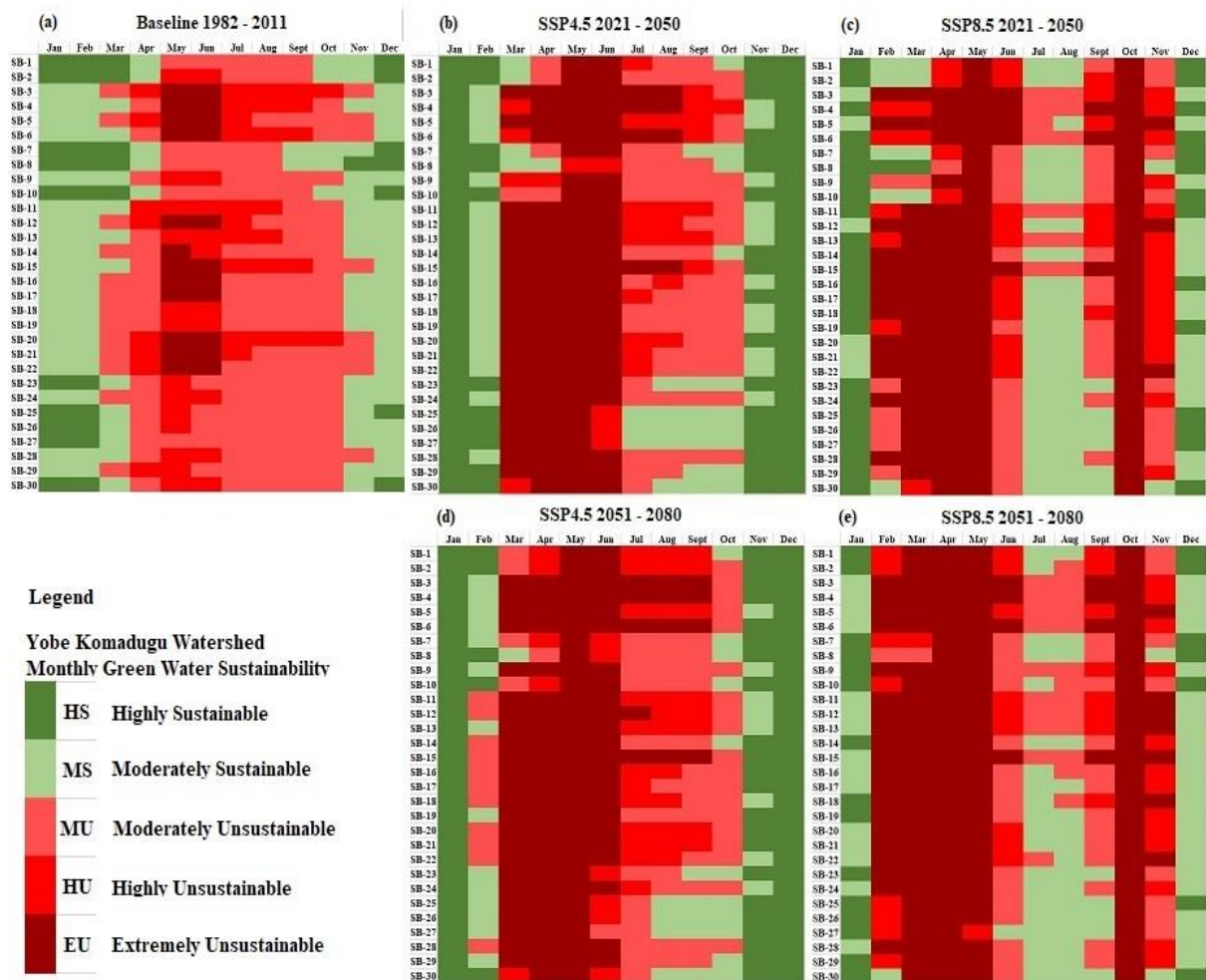
The continued decline in sustainable blue water may be worrisome to local and national strategic freshwater management plans and a threat to diplomatic relationships among countries that shared the basins.

#### **6.5.4 Climate change impact and socioeconomic drivers on the temporal variability of projected changes of green and blue water sustainability.**

The green and blue water sustainability assessment at the local basin scale will require an understanding of the temporal pattern of freshwater circulation at a monthly timescale to improve and stabilize the basin ecosystems. **Figure 6.14a – e**, showed a heat map of the severity of the baseline and projected monthly changes in green water sustainability across the 30 sub-basins of the watershed.

The results indicated that green water is more sustainable in the pre-monsoon and post-monsoon months with indices between 0.15 – 0.95 in **Figure 6.14a**, although there is a consistent projected change of sustainability status from moderately unsustainable to highly and extremely unsustainable green water in the monsoon months between April – June indicating a transition to potential geographic water sustainability hotspots across all the climate change scenario shown in **Figure 6.14b - e**.





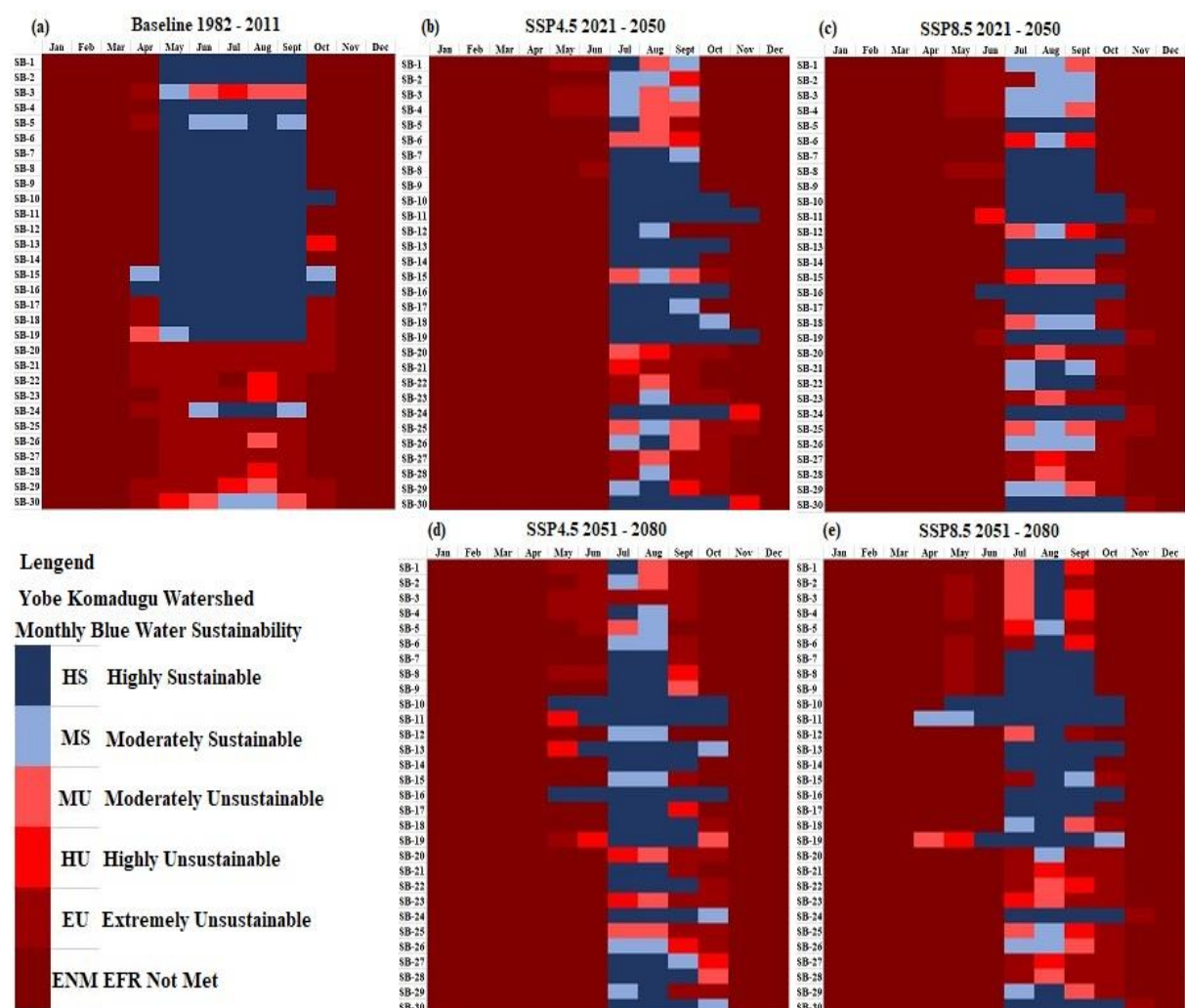
**Figure 6.14:** Heat Map showing temporal changes of mean monthly baseline and projected green water environmental sustainability in Yobe-Komadugu Watershed

However, a gradual change in favourable green water sustainability status is emerging in the month of July – August in 23% (**Figure 6.14b**) and 80% (**Figure 6.14c**) in the near future and 16.7% (**Figure 6.14d**) and 70% (**Figure 6.14e**) in far future for SSP2(4.5) and SSP5(8.5) emission scenarios respectively across the entire watershed, and this may be connected to the sudden projected increase in rainfall events and totals in the semi-arid climate.

Assessment of baseline blue water sustainability status in **Figure 6.15a** indicated that sub-basins 1 – 19 (upstream) showed a moderate to high blue water sustainability in the monsoon months of May – September with indices that ranged from 0.47 – 0.98, conversely, sub-basins 20 - 30 (downstream) are generally exhibiting the potential for geographic blue water

sustainability hotspots indicating highly to extremely unsustainable status in 92.2% of the time during the monsoon period with indices of  $-0.78$  to  $-4.2$ .

However, our analysis of monthly blue water availability indicated that environmental flow requirement cannot be met at 60.3% of the months in the baseline period to maintain a healthy aquatic ecosystem and have been generally identified during low flow periods between the months of November – March and should have been classed as “No abstraction period” and streams should be protected across the basin.



**Figure 6.15:** Heat Map showing temporal changes of mean monthly baseline and projected blue water environmental sustainability in Yobe-Komadugu Watershed.

The unsustainable blue water status may be closely related to the mass exploitation of groundwater and surface water for domestic and agricultural (irrigation) practices, high rate of surface water evaporation and plant transpiration due to increase surface air temperature

that triggered declining runoff contribution and shrinkage to the major Lake Chad which is consistent with the findings of Lemoalle et al., (2012); and Zhu et al., (2017).

The assessment of the model output for blue water sustainability status showed a projected increase of “No abstraction period” to 74.4% (**Figure 6.15b**) and 65% (**Figure 6.15c**) for the near future and 66.11% (**Figure 6.15d**) and 65.3% (**Figure 6.15e**) in the far future for CO<sub>2</sub> emissions scenarios SSP2(4.5) and SSP5(8.5) respectively.

The projections here indicate that the gradual increase in precipitations may have a direct impact on the sustainability of green water resources, where the monsoon months of July – August experienced a projected change in green water sustainability status from MU to MS as seen in **Figure 6.14c** and e.

However, blue water sustainability status tends to be degraded relative to the baseline in all emissions scenarios considered and a favourable blue water status may only be achieved through enforcing regulations to protect intense groundwater withdrawal, especially during low flow periods and exploring innovative river basin water conservation strategies.

According to the anticipated changes in the sustainability of green and blue water, more than half of the watershed will be ecologically fragile. Without prompt action by water authorities to improve ecological resilience and adaptation to reduce the shrinkage of wetlands and the larger Lake Chad in the face of changing climate and socioeconomic activities, some regions' freshwater geographic sustainability hotspots status may be beyond the tipping points, which will make restoration quite difficult.

Although improved data assimilation resulting from this framework that identifies the relevant pathway of the GCMs ensembles process which is capable of producing the best information, to strengthen monitoring and serve as an efficient information exchange can help to reduce uncertainties in climate change projections to a particular degree.

However, it does not imply that the projections are certain, and climate and hydrologic modellers or end-users may still consider the need for a worst-case output of future hydrological response in water policy decisions.

This is because significant uncertainties and risks will remain, and they are multifaceted due to the combination of climate change and other related socioeconomic drivers in water security assessment is expected to increase uncertainty (Martin et al., 2020).

Notwithstanding, the implementation of suitable measures, like integrating professional opinions, acquiring and refining novel data sources, and process comprehension of model structures and management strategies in water policy decisions, has the potential to augment and mitigate the effects of uncertainty in climate change impact modelling.

## **6.6 Conclusions and future work**

In this study, we developed a framework by integrating machine learning-based Boruta random feature selection as an input data refining process with process-based SWAT hydrologic models to optimize calibration process.

This framework was used to test whether models can simultaneously enhance baseline and future climate projections and accurately simulate water balance components by accepting or rejecting parameter solutions based on a defined error threshold when alternative satellite-based measurements of hydrologic fluxes are used in data-sparse watersheds where local observations are insufficient and are required for successful and reliable hydrologic modelling at the local scale.

In addition, the calibration and validation of the models with AET data is essential in basins where there is lack of or inadequate ground observation monitoring (streamflow) data in sufficient quantity and quality to drive hydrologic models. This is the main limitation of hydrologic modelling in data-sparse watersheds.

This also helps to estimate the components of infiltrated water such as aquifer recharge, soil moisture content and actual evapotranspiration with high degree of confidence, which are the main hydrologic variables and drivers essential for water footprint and sustainability assessment studies at basin scale.

Applying the framework to four sub-watersheds that form the larger Lake Chad basin defined by distinct morphological properties, we found that, the model simulates the hydrologic fluxes of ET with varying degree of acceptability. While ET can be simulated accurately, there are significant trade-offs in parameter sensitivity ranges in the calibration process across the sub-watersheds. Some of the key findings and conclusions in the research are summarized below.

The integrated hydrologic modelling process in this study can reliably represent the spatiotemporal distribution of the watershed hydrology irrespective of the different morphological characteristics of the four sub-watersheds, and reduce uncertainty from the input data (e.g., precipitation and temperature), which are the main drivers of water balance models.

The feature selection mechanism could reduce uncertainty propagation within acceptable thresholds in the data input process and provide ensembles whose projections can be relied upon and consistent with previous studies for water security assessment.

Green water is the dominant freshwater component across the basin relative to blue water, and climate change may be a significant factor in the spatial and temporal changes of projected green water sustainability status. However, the combination of socioeconomic drivers (i.e., only considered for blue water sustainability assessment) and climate change may have a significant impact on the projected blue water sustainability status across the basin.

High GWF, temperature and the flat terrain in the Yobe Komadugu watershed may affect the spatial distribution of projected natural runoff distribution and thus, the projected blue water footprint exceeds the blue water availability, and human water use can only be met by using up the environmental flows, resulting in degradation of rivers and groundwater potential.

Additionally, given the WF hotspots found in this study, new appropriate water abstraction targets should be quantified as part of future research, as well as its impact on blue water, which has a higher opportunity cost due to its potential as an input in many supply chains for emerging industries other than agriculture to help improve water management efforts at local river basin scale.

We should also point out that the results and conclusions reported in this study are based on certain configurations of the model parameters, input dataset, reference data, and hydrologic model. The established modelling framework, however, is independent of model and data type and may be used to assess the effectiveness of hydrologic state variables and fluxes at small-scale watershed levels.

Nevertheless, some obvious limitations are that the study does not consider the effects of some watershed management practices like irrigation withdrawals, reservoir regulations due to unavailability of data and future topographic changes in terrain and slope which will be significant driving factors governing the hydrologic response to land use and land cover changes.

This research focuses on blue and green water sustainability, however, efforts are required to extend the current work to grey water assessment by developing innovative ways and building observation datasets to further extend the model calibration and validation efforts to increase the confidence of hydrologic variable outputs required to reliably measure

and quantify grey water footprint and sustainability for managing wastewater discharge, consumption of fertilizer and pesticides for control of water pollution.

The water sustainability indices and status could be useful in the development of effective local river basin policies and regulations. Future work should involve addressing some of the limitations identified by extending the current study through the incorporation of more water balance components into the calibration process and analysing their effects on the overall trade-offs in the accuracy of modelling output.

## **6.7 Afterward**

This work has shown that by investigating and integrating transdisciplinary novel concepts to enhance current knowledge of climate change models and evaluation, effective predictive hydrologic modelling may be produced in data-sparse regions and give important water resource information at a local watershed scale.

The quality and resolution of gridded and GCM climate data and their interactions through downscaling and bias corrections scheme, which was a key aspect of the strategy explored in this research to improve on the theoretical underpinnings of process-based hydrologic models, regardless of their complexity and intended purpose, is largely dependent on the appropriate extreme (SPEI drought and flood hazard) event representation, improved understanding of key watershed climate processes (Lawal et al., 2023b) and synthesis of a watershed process of building a reliable model while acknowledging data limitations and uncertainty.

The goal is to assess and recommend an alternative hydrologic modelling concept that places a focus on process optimization and offers flexibility to define multiple representations of spatial variability and hydrologic connectivity through careful refinement of data structure, complexity, and reliability as demonstrated in chapters 4 and 5 without overextending model

parameter sensitivities to ensure robust predictions of watershed water resources dynamics under uncertainty, accurately on a wider spatial scale exhibited in this chapter.

The objectives of the research have been achieved and to encourage further inspection and adoption, this exploratory approach might be expanded to watersheds with abundant requisite modelling data. Future work activity recommendation is detailed in **CHAPTER 8:**



## CHAPTER 7: SUMMARY OF FINDINGS AND DISCUSSIONS

### 7.1 Assessment of quality-controlled observation data for consistency

The improvement in the development and availability of gridded climate datasets has led to the continual growth of the range of diverse users especially in watersheds where ground-based observations are inadequate and the quality requirements for their application in hydroclimatic study have evolved accordingly.

Basic understanding of the spatial and temporal variability or dynamics of this climate data is important in watershed modelling especially at daily timescale and this has become a challenge due to likely uncertainties associated with their development.

The efficacy and reliability of gridded data were validated relative to the observation station measurements. However, the missing data points of the station data were gap-filled by multivariate imputation by chained equations for both precipitation and temperature using MICE package in R software (**See 250**), due to its ability to impute effectively a continuous two-level data and maintain consistency (van Buuren and Groothuis-Oudshoorn, 2011).

The filled data points were checked for 100% completeness and subjected to a consistency and homogeneity test at individual station observations. The homogeneity test results using double mass curve, and absolute homogeneity tests by standard normal homogeneity (SNHT), Pettitt, and Von Neumann ratio test indicate that a straight-line plot without a breakpoint (**Figure 4.2**) for double mass curve and the estimated homogeneity test values were less than the critical value for precipitation and temperature respectively.

The application of several test here, was to provide a robust result preventing overestimation and correction of false inhomogeneities that leads to unreliable climate analysis. Therefore, the gauge measurement of the station records is reliable and can be used for validation.

## 7.2 Reliability of gridded climate data using multi-criteria approach

Some of the notable findings of the multi-criteria performance evaluation of the behaviour of the gridded climate data are summarized below:

The relative and absolute homogeneity test results are consistent and homogeneous across the stations. Some of the factors that may affect the true climatic conditions of the observation data series and consequently, the outcome of its application in further analysis due to the existence of bias such as gradual changes in instrumentation i.e., maintenance and calibration problems of the equipment, urbanisation resulting in changes of equipment locations and expertise of the personnel taking records are controlled.

This factors has been corroborated by Yozgatligil and Yazici, (2016), and is crucial to achieving an understanding of the climate system and its associated changes especially in data-sparse regions where data series quality is in question.

The result of the quality-controlled data series used to assess the capability of the gridded datasets to replicate the observation quite varied across the watershed and inconsistent across the different performance metrics. However, this may be attributed to the varied level of systematic errors in the gauged station measurements.

Additionally, the choice of interpolation technique, source and quality and quantity of the observation data used in the development of gridded data products may affect their ability to capture the spatiotemporal internal variability as acknowledged in (Faiz et al., 2018).

Analysis results of the agreement of gridded dataset relative to observation across the stations for precipitation and temperature in **Table 4.5 – Table 4.6** using statistical coefficients KGE, md, and NRMSE are quite satisfactory, except for PGF whose KGE values are quite low (0.33), while SC from symmetric uncertainty technique for precipitation are also low for all products shown in **Table 4.4**.

The result showed that temperature data series are less likely to be inconsistent and portrayed low NRMSE values indicating low level of uncertainty relative to the observation.

Further analyses were explored such as precipitation and temperature trend at annual and monsoon season. The result of the unidirectional trends of precipitation and temperature are adequate relative to the observation data. The analysis indicates a varying degree of mismatch between the gridded dataset relative to the observation data across the stations.

The agreement in precipitation data as displayed in **Table 4.7 – Table 4.10** based on Z – statistics values at annual, and the monsoon season for CPC, CRU, GPCC, PGF and UDel relative to the observation station data respectively. While the temperature data were consistent especially the CRU, PGF, and UDel with observed unidirectional trend agreement at annual and monsoon season respectively.

The annual precipitation and temperature exhibit a statistically declining trend in the Sahelo-Sudanian and Sudano-Guinean zone respectively, while increasing trend were noticed in all other stations and this consistent with notable findings from earlier studies (Conway et al., 2009; Nkiaka et al., 2017; Sarr, 2012).

It was also noted that some of the gridded data (**Table 4.7 – Table 4.10**) over and underestimate the unidirectional trends of precipitation and temperature by 1 to 2.5 order of magnitude relative to the observations in some stations. This mismatch has been acknowledged as one of the sources of uncertainty in hydrologic modelling process and caution should be applied while selecting alternative data for watershed representation.

Therefore, a single metric adopted for assessment may be misleading and a robust approach that minimize the variation of the means and standard deviation in **Figure 4.5a-f** is critical and multicriteria decision approach that combines a plethora of techniques and showed the strength and weaknesses across metrics and ranked the products by intercomparisons based on performance to reflect their order of suitability (**Table 4.11**) for

climate model assessment into the appropriate context will increase confidence and accurate watershed assessment. Therefore, the gridded dataset recommended are robust and can fairly represent basin climatologic feature for reliable application in impact study.

This is essential especially in data-sparse watershed where modelling output information and their projections are within acceptable uncertainty limits are required as a decision support tool for accurate and adequate water resource management.

### **7.3 Data pruning as an effective approach for GCM uncertainty reduction in modelling**

Appropriate watershed planning to deal with projected water resources sustainability can be carried out with GCMs taking into account environmental changes such as urbanization, population growth, climate change etc., which are some of the limitations of the gridded dataset. However, they are essential in downscaling of the GCMs from coarse to finer resolutions suitable for application at the regional and local basin scale.

Some of the gaps identified from the literature are that the time scale and resolutions of the GCMs datasets and these are often ignored in the assessment of their efficacy for hydrologic impact studies and conclusions drawn from previous studies may not be relied on for realistic projections at local basin scale.

Questions arising are that the approach of ensemble averaging of a lot of GCMs for application may be cumbersome and time-consuming, while other drawbacks are as stated by Cook, (2008) that the uncertainties are likely overstated, and models tend to simulate small changes in rainfall and the projected increase are likely influenced by a few outlier models and this flaws in the input rainfall data, if not limited might be amplified by the non-linearity of hydrologic models.

These concerns thought-provoked the second research objective to reassess GCMs critically and objectively by providing a robust pathway that captures accurately the

watershed's temporal and spatial variability such as trends and magnitude of climate extremes and generate projections of the climatology of a region.

These watershed climate and hydrologic variabilities should be consistent with best practices and recommendations of the intergovernmental panel on climate change (IPCC) by the selected appropriate GCM multi-model ensemble climate variable outputs that enhances and preserve the internal variability after reparameterization exerted by re-gridding, downscaling and bias correction which is an important factor that ensures the successes of hydrologic model performance especially in regions where data quality is essential.

Machine learning based algorithms Symmetric uncertainty (SU) and Boruta random forest (BRF) (See 254) were developed and applied for the evaluation of 16 GCMs that meets the essential requirements for hydrologic study. Multi-point downscaling and bias corrections were applied at 54 grid point in Lake Chad basin.

The downscaling and bias correction scheme indicated that the simulated output of daily precipitation distribution from the delta change method was far superior performance, and correlate well with the gridded data, while the empirical quantile mapping effectively captures the daily temperature distribution relative to the other methods across all stations respectively.

Whilst the metric of evaluation showed a promising result in temperature distribution, there are observed difficulties of the downscaled models to capture the peak values of minimum temperature (**Figure 5.3c**) across all stations, but notably more pronounced in the Sahelo-Saharan zone and this has been attributed to scale gap between the GCMs and gridded data and the variations are insignificant and cannot be accounted for by the output as corroborated by Sachindra et al., (2014) in a basin with similar climatology.

A more efficient performance was observed in Sudano-Guinean zone, with a cluster of grid points with almost perfect correlation relative to the gridded data, while Saharan zone

exhibit some inadequacies in simulating the climate features which may be related to the low precipitation events.

Assessment of the GCMs capability to capture the climatic features of the gridded data used as surrogate showed a varied level of association. However, the similarity coefficient (SC) from the assessment exhibits an improved performance relative to the latter phases of GCMs and the estimated range of SC (**Table 5.2**) for daily precipitation, maximum and minimum temperature respectively.

The varied level of association may be attributed to the effect of downscaling on initial resolutions, improvement in parameterization techniques and quality of observation data in model development and it has been acknowledge in studies by Ayugi et al., (2021); Grose et al., (2020); and Wang et al., (2021).

The evaluation using BRF revealed that the GCMs showed a varied level of performance to capture the significant antecedent feature of the observation of daily precipitation (**Figure 5.4a**) and this is common across the test grid points. Although the temperature data are quite well correlated with the observations as seen in **Figure 5.4b** and c for maximum and minimum temperature respectively. The important score was reaggregated across the basin and ranked in descending order of importance in **Table 5.3**.

The different results in simulating the precipitation and temperature observations in different parts of the basin (**Figure 5.5a – c**) implied that post-bias correction reassessment using filters like BRF as an important data pruning (i.e., process of GCM data elimination or isolating sub-optimal tuples that are non-critical or redundant to reduce noise and improve MME representation of basin climatic trend, magnitude and pattern of variability at spatial and temporal scale) methodology to limit the number of GCMs might reduce residual biases and uncertainties if present, that are likely to aggregate and yield errors in the further

application of the data in hydrologic modelling, particularly when dealing with the estimation of high and low flows conditions.

#### **7.4 Spatial and temporal assessment of GCM ensembles for hydrologic modelling**

Further reassessment of the machine learning approaches was conducted by developing multi-model ensemble of the four best performing GCMs after reaggregation by multi-criteria ranking across variables shown in **Table 5.4** and compared with the traditional approach of using all the GCMs identified here in referred as AME to evaluate the spatial and temporal correlation and mean annual precipitation and temperature biases across the entire grid points and validated by assessing the trend and magnitude of the return period of flood and drought hazards across the four climatic zones of the basin.

The result of the spatial correlation and pattern of the annual precipitation and temperature (**Figure 5.7a** and **b**) has shown that the BRF approach is consistent in simulating the observation and effectively correlate well with the observations (**Figure 5.6a** and **b**) across the basin respectively. However, the result of SU and AME ensemble approach are satisfactory, but poor correlation value were observed, especially in Sahelo-Saharan and Saharan zone for AME relative to the observations respectively.

Temporal assessment of the time series data across the basin indicates consistent and quite similar results for the ensemble mean temperature for all the approaches i.e., BRF, SU and AME respectively.

The assessed ensemble however, showed that multimodel ensemble formed using BRF data pruning approach effectively captured the annual precipitation spatially with reduced the annual mean precipitation bias (**Figure 5.8**) and correlates well with the observations relative to the tested approaches respectively. The biases or deviations are significant at grid points situated in the Saharan and Sahelo-Saharan zone of the basin.

The output from the three approaches was validated by examining the basin scale dynamics and relative skills in predicting the pattern, trend and magnitude of return period of drought and flood hazards across the four climatic zones. The result showed a satisfactory depiction of trend shift (**Figure 5.9a**), from wet extreme (1980 – 1998) to a gradual transition to a moderate to extreme droughts (1999 – 2012) in the Saharan zone.

While seemingly the performance of the three approaches in capturing the trends shift may be similar by visual graphical representation in **Figure 5.9b – d** across the climatic zone relative to the observation, the statistical trend based on the multi-year SPEI indices for the study period across all the climatic zones indicated that the BRF approach reliably captured the extreme event indices and direction quite accurately relative to the observation.

This is evident based on the trend indices displayed in **Table 5.5**, with z-statistic values of BRF approach within the same trend envelop with the observation and this is crucial in the accurate depiction of basin hydrology by limiting input uncertainties in modelling studies.

Further assessment of projected mean changes in precipitation and temperature in the Yobe-Komadugu watershed in the near and far future base on SSP2-4.5 and SSP5-8.5 scenarios relative to the baseline periods, has shown that the BRF methodology is robust and consistent with reported findings from studies within the watershed in Almazroui et al., (2020); Sylla et al., (2016); and Vizy et al., (2013).

### **7.5 Modelling evapotranspiration using integrated SWAT-BRF framework.**

This approach was integrated to the traditional modelling framework using SWAT model and tested on a typical case study area of four sub-watershed of Lake Chad basin with quite distinct morphological characteristic and climatological dynamics, where hydrologic long term streamflow data are sparse and unreliable.

Rather than resorting to regionalisation approach of modelling whose output is not convincing, we set out to investigation the application of the integrated framework to model



actual evapotranspiration across the four sub-watershed using remote sensing MODIS evapotranspiration data with a far-reaching spatial dimension to increase confidence in the modelling output at local basin scale.

The model calibration process results across the watersheds indicate variation in model parameter values and sensitivities (**Table 6.4**), however, all the parameter values are well within the acceptable absolute parameter ranges. The variations may be influenced by the combination of HRU regionalisation, soil, and land use conditions of each calibration point across the watershed which may aid lack of uniqueness in the calibration process.

The objective functions and algorithms used provided a range of values across the watershed that are satisfactory and well within acceptable uncertainty range (**Figure 6.4**) with a wider spatial areal extent as discussed in **6.5.1**.

The simulated results showed that the objective functions across the calibrated and validated points are in very good ranges ( $R^2 > 0.62$ ,  $NSE > 0.34$ ) and uncertainty ranges of P-factor and R-factor values of 0.68 – 0.93 and 0.73 – 1.31 respectively in 83%, 67%, 85.7% and 81.3% of the measured point across the sub-watershed.

While the optimized models are quite good, there are few simulated points with low objective function with correlation and Nash-Sutcliffe efficiency values as low as 0.25 and 0.14 respectively. However, they exhibited a good representation of data uncertainty range with satisfactory P-factor and R-factor values which are associated with difficulty in matching peak points especially in **Figure 6.4d**.

This anomaly in depiction may be related to limitations in modelling process such as simplification of the model structure like reaggregation of land use features, soil conditions, inadequate data that accounts for some of the essential basin scale water management processes for example sufficient information on reservoir and dam operations, water transfer and irrigation processes. This phenomena is discussed in Abbaspour, (2015); Schuol et al.,

(2008) and classed as technical modelling uncertainties and modifying the natural heterogeneity of the watershed conditions will generally affect model performance.

The modelling output has shown that the framework developed with the application of remote sensing data can objectively provide an opportunity to improve watershed representation of hydrologic process in data-sparse regions and differs from the previous studies using the regionalisation approach in Faramarzi et al., 2013; and Schuol et al., (2008) with associated poor watershed representation and large uncertainty range in the Lake Chad region.

## **7.6 Impact of climate change on projected green and blue water resources**

The model output variables from the modelling approach were applied to Yobe-Komadugu watershed to analyse the dynamics of projected green and blue water footprint, availability and sustainability in response to climate change based on SSP2-4.5 and SSP5-8.5 emission scenarios.

This study was necessary to understand was resources status at local basin scale. However, the lack of comprehensive and long-term sectoral water use information in these regions (Data-sparse) and previous research could not assessed green and blue water sustainability status and identify water resource hotspots and their projection.

This study attempts to address that by developing a theoretical framework to estimate conservative blue water footprint at annual and monthly scale using model-based parameters, limited water uses information available and gridded population data (See Eqn. (6.6). this was used to address research objective 4, 5, and 6 at local basin scale. The findings are summarized below.

A general marked increase in the spatial changes in projected mean annual GWF from the baseline period (**Figure 6.6a-e**) for all the emission scenarios across the time slices, while

analysis of monthly distribution and changes in **Figure 6.7a – b**, depicted a consistent projected increase between spring and summer months.

However, a sharp decline in autumn and winter months projected in all scenarios. This phenomenon may be related to the continuous warming caused by increase in average air temperature in the tropical regions especially between the month of April and September.

Analysis of basin changes in annual soil moisture conditions represented as GWS shown in **Figure 6.8a – e** has indicated a substantial projected decline across the watershed from baseline condition (1982 – 2011). While the decline is consistent in the monthly variation (**Figure 6.9a – b**) in all scenarios.

This decline may be related to the huge exploitation of groundwater resources for irrigation practices that further lowers the soil water table level and possibly the increase in surface air temperature affecting soil water flow regimes thereby increasing groundwater evaporative demand.

The general decline in soil moisture conditions experienced in the basin may be related to a general increase in evaporative demand that has been projected in over most part of the world in high emission scenarios mostly related to and as a consequence of increased vapour pressure deficit consistent with expert findings in (Scheff and Frierson, 2014; Vicente-Serrano et al., 2020)

However, exception is observed in the monsoon season with a projected increase in the near and far future based on the SSP2-4.5 and SSP5-8.5 emission scenarios. The projected increase in the GWS in the monsoon season is generally significant between the month of July – September and generally associated with high rainfall intensities and interannual seasonal variability as corroborated by Almazroui et al., (2020).

The impact on spatial and temporal variation of blue water flow dynamics was quite distinct in the upstream and downstream parts of the watershed. A general projected decline

in the upstream, while an associated increase was observed at the downstream of the watershed (**Figure 6.10b** and e) for SSP2-4.5 and SSP5-8.5 respectively. The results are also consistent with the far future time slice although at a different rate. However, mean annual changes across the entire watershed depicted a projected increase from the baseline period across all the emission scenarios.

The monthly variation of the BWF (**Figure 6.11a** and b) for near (2021 – 2050) and far future (2051 – 2080) scenarios showed a projected decline prevalent between the winter and spring season and this is quite expected because the months are characterized with sub-optimal precipitation (i.e. below long term basin average) totals in the tropics.

This is generally attributed to the likely increase in drought occurrences in this regions under the high emission scenarios (Cook et al., 2020), while the projected increase in the summer and autumn is associated with increase in monsoon rainfall events and intensities that characterizes the intensification of wet extremes that causes frequent flood events in the watershed supported by studies reported in Niang et al., (2014) and increased dry spell lengths especially in winter and autumn by shortening the Sahel rainy season as reported in numerous studies (Almazroui et al., 2020; Sylla et al., 2016).

Therefore, increased amount and intensity of precipitation as projected by the high emissions scenarios may not always correspond to increased green and blue water resource availability since temperature-induced increases in evapotranspiration can offset precipitation as alluded by Vicente-Serrano et al., (2020).

### **7.7 Impact of climate change on projected green and blue water sustainability**

The assessed annual green water sustainability status of the watershed in this study has shown that most of the sub-basins will experience a steady decline in annual green water sustainability status from the baseline period across the watershed area for SSP2-4.5 (**Figure**

6.12b and c) and for SSP5-8.5 (**Figure 6.12d** and e) in the near and far future time slices respectively.

These changes are usually from the moderately to extremely sustainable status to moderately and extremely unsustainable status based on the hazards map. Some of the sub-basin are already projected and classed green water geographic hotspots in all scenarios indicating that climate change may have a profound effects on the dynamic changes on the green water sustainability status which may have a ripple effect on the timing, spatial pattern and intensity of rainfall as suggested by Du et al., (2018).

The dynamic changes in annual blue water sustainability status are quite pronounced in the upstream where some of the sub-basins were projected to experience a rapid change moderately to highly sustainable to highly and extremely unsustainable status.

These dynamic shifts are related to the absence of viable streams channels and year-round freshwater infrastructure where the vast population depends on groundwater as an alternative water source. Even the sub-basins with major river systems will experience reduced flow and these consistent patterns will be scaled up with the SSPs emission scenario and these may be worrisome to local and national strategic freshwater management plans and could cause a threat to transboundary diplomatic ties on freshwater exploitation.

The sustainability heat map (**Figure 6.14 – Figure 6.15**) of the dynamic changes in monthly sustainability status implied that green water is projected to be sustainable in winter months while blue water is mostly during the summer months due to excessive rainfall events. However, the pattern has shown to be shrinking considering the two emission scenarios due to increase dry spell length triggered by global climate change.

The assessment has shown that some of the sub-basins water resource will be violated as available freshwater in the stream is significantly less than the environmental flow requirement and this is a threat to aquatic biodiversity.

According to the projected changes in the sustainability of green and blue water, more than half of the watershed will be ecologically fragile. Without prompt action by water authorities, some regions' freshwater geographic sustainability hotspots status may be past the recovery points, making restoration very challenging.

This is especially true if there is no improvement in ecological resilience and adaptation to lessen the shrinkage of wetlands and the larger Lake Chad in the face of changing climate and socioeconomic activities.

Therefore, the projected future population growth in conjunction with climate change may considerably raise the demand on the local basin-scale available blue water resources and in effects its sustainability. It is anticipated that climate change will have a major effect and worsen the watershed's projected green water resource and sustainability.

## **7.8 Overview of integrated framework for water security modelling**

The purpose of the research was to provide a viable alternative solution to hydrologic modelling in data-sparse regions where traditional framework required is not feasible and handy to assess water security status. The objective has been addressed and achieved by successfully testing the approach in a typical case study area of Lake Chad basin.

The combination of an integrated framework of SWAT and BRF GCM data pruning methodology produces a coherent and workable hydrologic modelling process that effectively simulate actual evapotranspiration to limit input data uncertainty spatially on a wider scale.

Finally, a theoretical blue water footprint accounting framework was developed using alternative data sources, model output information to cater for inadequate long term sectoral blue water use information, which is shown to be effective that furthers the understanding of green and blue water sustainability studies in response to climate change at the local basin scale.

## **CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

### **8.1 Conclusions**

This thesis highlights challenges faced in watershed modelling in data-sparse regions and some of the novel initiatives taken to address the literature gaps identified which ranges from data choice and quality, parameterization, model conceptualization and data assimilation approach to improve water security studies are explored.

The inferences and conclusion drawn from the studies are summarized below.

- Gridded climate data can fairly represent the chosen watershed climatology of interest both spatially and temporally, although their performance varies considerably and the choice of data to maintain a high degree of accuracy and low level of uncertainty especially in data-sparse regions is dependent on the method of valuation and specific application.
- A comprehensive assessment using a plethora of performance metrics and a robust multi-criteria decision that considers three broad aspects viz.: time scale, resolutions, trend and magnitude of data series can effectively showcase the strength and weaknesses across metrics and the suitability of a data product based on intercomparison will increase confidence in the assessment of watershed climatology.
- The choice of reference dataset and data pruning process using the BRF approach can strongly influence the accurate depiction of future projections of precipitation and temperature and effectively accounts for the inadequacies of bias correction schemes to constrain observational uncertainties of downscaled GCM ensemble simulations for hydrologic impact studies at the local basin scale.

- The data pruning approach using BRF algorithms of the GCM ensembles has been shown to maintain and captures the trend and magnitude of spatial and temporal pattern of the return period of flood and drought extremes across the different climatic zones of the basin accurately.
- GCMs evaluation using BRF data pruning algorithms is found to be extremely effective, although the knowledge and science of the underlying process are limited. However, the process when integrated into the hydrologic modelling framework is capable of reducing the propagation of uncertainties originating from the GCM ensembles and scenarios from SSPs input data in impact assessment modelling studies.
- The developed modelling framework can effectively simulate watershed hydrologic fluxes of actual evapotranspiration satisfactorily with a varied degree of model performance. However, there are significant trade-offs in parameter sensitivity due to the distinct watershed morphological features across the sub-watersheds.
- Assessment of green and blue water availability and sustainability suggested that green water is projected to be the dominant freshwater component relative to blue water resources and climate change may be a significant factor in the spatial and temporal changes of projected green water sustainability status. However, the combination of socioeconomic drivers and climate change may have a significant impact on the projected blue water sustainability status across the basin.
- The identified geographic water footprint hotspots based on sustainability indices can be improved by reviewing freshwater abstraction targets by easing up



environmental flows to prevent watershed degradation of rivers and groundwater potential to enhance river basin water management efforts.

## **8.2 Recommendations to current practice**

The use of caution when addressing issues related to uncertainties associated with input data parameterization and choice to prevent error magnification is crucial for practitioners in the field of watershed modelling to understand, especially when applied to projected climate change impact and adaptation studies.

These errors and their sources, if not properly controlled, can have a negative impact on the range and distribution of water balance dynamics in a modelling scheme. The data-pruning framework presented in this study should be tested beyond data-sparse regions, because it can enable modellers to derive an accurate quantification of hydrologic model's output uncertainty and prevent its propagation without the hassle of compensating the effects of recalibrated parameters.

Assumptions and parameterizations of important climate features in model development, as well as creative and reliable ideas to manage and communicate uncertainties in impact studies, should be discussed in a collaborative forum open to practitioners, policymakers, and climate modelling centres.

In particular in developing economies where water management options are insufficient, this will be a significant step towards improving watershed representation by practitioners and providing accurate water resource information to be used as a decision support tool for adequate river basin water management and reducing transboundary water conflicts.

## **8.3 Recommendations for future research**

Some few recommendations are suggested to further the cause of research in data-sparse regions are itemized below.

- The effects of water management practices like irrigation withdrawal, reservoir regulations, when data are available, should be incorporated, analysed and compare the model performance findings with the results presented in this research.
- The effects of dynamic changes in grey water footprint and sustainability should be explored when data required is available to effectively calibrate models and to reliably simulate the grey water component to complete the cycle of water footprint assessment.
- The efficacy of the data pruning approach tested here can be extended to other hydrologic models with varying complexity to encourage scrutiny and widespread adoption of the framework to expand and improve scientific rigour. This should be expanded to examine the impact of the model structure and parameter sets on the data-pruned ensemble GCMs' behaviour on the hydrologic output and performance.
- Efforts in building hydrologic models for water security assessment need to move beyond traditional strategies to a multi-disciplinary framework required to improve the science of watershed representation that reduces various forms and sources of uncertainty and the general applicability in the hydrologic modelling process. This can be achieved by strengthening the linkages between theories, algorithms, data assimilation and observations to further improve model simplifications with little impact on fidelity of model simulations and increase confidence in model predictions of watershed resource dynamics.
- Risk and vulnerability assessments will serve as a basis for well-thought-out planning of climate resilience measures and strategies, and ultimately enhance future climate adaptation efforts in the water sector. These assessments entail

making appropriate use of current climate projections and carefully considering climate-related and socio-economic hazards. Therefore, access to basin water vulnerability information will provide a pathway for potential projected green and blue water vulnerability and risk assessment at basin scale.

## REFERENCES

- Abbasian, M., Moghim, S., Abrishamchi, A., 2019. Performance of the general circulation models in simulating temperature and precipitation over Iran. *Theoretical and Applied Climatology* 135, 1465–1483. <https://doi.org/10.1007/s00704-018-2456-y>
- Abbaspour, K.C., 2015. SWAT-CUP calibration and uncertainty programs. a user manual, department of systems analysis. Intergrated Assessment and Modelling (SIAM).
- Abbaspour, K.C., Faramarzi, M., Ghasemi, S.S., Yang, H., 2009. Assessing the impact of climate change on water resources in Iran. *Water Resources Research* 45, 1–16. <https://doi.org/10.1029/2008WR007615>
- Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology* 524, 733–752. <https://doi.org/10.1016/j.jhydrol.2015.03.027>
- Abbaspour, K.C., Vaghefi, S.A., Srinivasan, R., 2017. A guideline for successful calibration and uncertainty analysis for soil and water assessment: A review of papers from the 2016 international SWAT conference. *Water (Switzerland)* 10. <https://doi.org/10.3390/w10010006>
- Abbaspour, K.C., Vaghefi, S.A., Yang, H., Srinivasan, R., 2019. Global soil, landuse, evapotranspiration, historical and future weather databases for SWAT Applications. *Scientific Data* 6, 1–11. <https://doi.org/10.1038/s41597-019-0282-4>
- Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* 48, 1–14. <https://doi.org/10.1029/2010WR009945>
- Adeyeri, O.E., Lamptey, B.L., Lawin, A.E., Sanda, I.S., 2017. Spatio-Temporal Precipitation Trend and Homogeneity Analysis in Komadugu-Yobe Basin, Lake Chad Region. *Journal of Climatology & Weather Forecasting* 05. <https://doi.org/10.4172/2332-2594.1000214>
- Adeyeri, O.E., Lawin, A.E., Laux, P., Ishola, K.A., Ige, S.O., 2019. Analysis of climate extreme indices over the Komadugu-Yobe basin, Lake Chad region: Past and future occurrences. *Weather and Climate Extremes* 23, 1–21. <https://doi.org/10.1016/j.wace.2019.100194>
- Aghakhani Afshar, A., Hassanzadeh, Y., Pourreza-Bilondi, M., Ahmadi, A., 2018. Analyzing long-term spatial variability of blue and green water footprints in a semi-arid mountainous basin with MIROC-ESM model (case study: Kashafrood River Basin, Iran). *Theoretical and Applied Climatology* 134, 885–899. <https://doi.org/10.1007/s00704-017-2309-0>
- Aguilera, H., Guardiola-Albert, C., Serrano-Hidalgo, C., 2020. Estimating extremely large amounts of missing precipitation data. *Journal of Hydroinformatics* 22, 578–592. <https://doi.org/10.2166/hydro.2020.127>
- Ahmadalipour, A., Rana, A., Moradkhani, H., Sharma, A., 2017. Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis. *Theoretical and Applied Climatology* 128, 71–87. <https://doi.org/10.1007/s00704-015-1695-4>
- Ahmed, A.A.M., Deo, R.C., Ghahramani, A., Raj, N., Feng, Q., Yin, Z., Yang, L., 2021. LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios. *Stochastic Environmental Research and Risk Assessment* 3, 1–31. <https://doi.org/10.1007/s00477-021-01969-3>

- Ahmed, K., Sachindra, D.A., Shahid, S., Demirel, M.C., Chung, E., 2019a. Selection of multi-model ensemble of general circulation models for the simulation of precipitation and maximum and minimum temperature based on spatial assessment metrics. *Hydrology and Earth System Sciences* 23, 4803–4824. <https://doi.org/10.5194/hess-23-4803-2019>
- Ahmed, K., Shahid, S., Sachindra, D.A., Nawaz, N., Chung, E., 2019b. Fidelity assessment of general circulation model simulated precipitation and temperature over Pakistan using a feature selection method. *Journal of Hydrology* 573, 281–298. <https://doi.org/10.1016/j.jhydrol.2019.03.092>
- Ahmed, K., Shahid, S., Wang, X., Nawaz, N., Najeebullah, K., 2019c. Evaluation of gridded precipitation datasets over arid regions of Pakistan. *Water (Switzerland)* 11, 1–22. <https://doi.org/10.3390/w11020210>
- Aieb, A., Madani, K., Scarpa, M., Bonacorso, B., Lefsih, K., 2019. A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon* 5, 1–27. <https://doi.org/10.1016/j.heliyon.2019.e01247>
- Akhter, J., Das, L., Deb, A., 2017. CMIP5 ensemble-based spatial rainfall projection over homogeneous zones of India. *Climate Dynamics* 49, 1885–1916. <https://doi.org/10.1007/s00382-016-3409-8>
- Alamgir, M., Ahmed, K., Homsy, R., Dewan, A., Wang, J., Shahid, S., 2019. Downscaling and Projection of Spatiotemporal Changes in Temperature of Bangladesh. *Earth Systems and Environment* 3, 381–398. <https://doi.org/10.1007/s41748-019-00121-0>
- Alcamo, J., Flörke, M., Märker, M., 2007. Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrological Sciences Journal* 52, 247–275. <https://doi.org/10.1623/hysj.52.2.247>
- Ali, M., Deo, R.C., Xiang, Y., Li, Y., Yaseen, Z.M., 2020. Forecasting long-term precipitation for water resource management: a new multi-step data-intelligent modelling approach. *Hydrological Sciences Journal* 65, 2693–2708. <https://doi.org/10.1080/02626667.2020.1808219>
- Almazroui, M., Saeed, F., Saeed, S., Nazrul Islam, M., Ismail, M., Klutse, N.A.B., Siddiqui, M.H., 2020. Projected Change in Temperature and Precipitation Over Africa from CMIP6. *Earth Systems and Environment* 4, 455–475. <https://doi.org/10.1007/s41748-020-00161-x>
- Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *MARINE ECOLOGY PROGRESS SERIES* 271, 13–26.
- Arnell, N.W., 2004. Climate change and global water resources: SRES emissions and socio-economic scenarios. *Global Environmental Change* 14, 31–52. <https://doi.org/10.1016/j.gloenvcha.2003.10.006>
- Arnell, N.W., van Vuuren, D.P., Isaac, M., 2011. The implications of climate policy for the impacts of climate change on global water resources. *Global Environmental Change* 21, 592–603. <https://doi.org/10.1016/j.gloenvcha.2011.01.015>
- Arsenault, R., Brissette, F., Martel, J., 2018. The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology* 566, 346–362. <https://doi.org/10.1016/j.jhydrol.2018.09.027>
- 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, 69–83. <https://doi.org/10.1175/BAMS-D-13-00068.1>
- Assefa, Y.T., Babel, M.S., Sušnik, J., Shinde, V.R., 2019. Development of a generic domestic water security index, and its application in Addis Ababa, Ethiopia. *Water (Switzerland)*

- 11, 37. <https://doi.org/10.3390/w11010037>
- Autovino, D., Minacapilli, M., Provenzano, G., 2016. Modelling bulk surface resistance by MODIS data and assessment of MOD16A2 evapotranspiration product in an irrigation district of Southern Italy. *Agricultural Water Management* 167, 86–94. <https://doi.org/10.1016/j.agwat.2016.01.006>
- Awange, J.L., Ferreira, V.G., Forootan, E., Khandu, Andam-Akorful, S.A., Agutu, N.O., He, X.F., 2016. Uncertainties in remotely sensed precipitation data over Africa. *International Journal of Climatology* 36, 303–323. <https://doi.org/10.1002/joc.4346>
- Ayugi, B., Zhihong, J., Zhu, H., Ngoma, H., Babaousmail, H., Rizwan, K., Dike, V., 2021. Comparison of CMIP6 and CMIP5 models in simulating mean and extreme precipitation over East Africa. *International Journal of Climatology* 41, 6474–6496. <https://doi.org/10.1002/joc.7207>
- Azur, M.J., Stuart, E.A., Frangakis, C., Leaf, P.J., 2011. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research* 20, 40–49. <https://doi.org/10.1002/mpr.329>
- Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining Global Population Distribution: Methods, Applications and Data. *Advances in Parasitology* 62, 119–156. [https://doi.org/10.1016/S0065-308X\(05\)62004-0](https://doi.org/10.1016/S0065-308X(05)62004-0)
- Baroni, G., Schalge, B., Rakovec, O., Kumar, R., Schüler, L., Samaniego, L., Simmer, C., Attinger, S., 2019. A Comprehensive Distributed Hydrological Modeling Intercomparison to Support Process Representation and Data Collection Strategies. *Water Resources Research* 55, 990–1010. <https://doi.org/10.1029/2018WR023941>
- Barron, E.J., Moore, G.T., 1994. Climate Models and Their Application, in: *Climate Model Applications in Paleoenvironmental Analysis*. SEPM (Society for Sedimentary Geology), Virginia, USA, pp. 23–30. <https://doi.org/10.2110/scn.94.03.0023>
- Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I.J.M., 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, 207–224. <https://doi.org/10.5194/hess-23-207-2019>
- Beharry, S.L., Clarke, R.M., Kurmarsingh, K., 2014. Precipitation trends using in-situ and gridded datasets. *Theoretical and Applied Climatology* 115, 599–607. <https://doi.org/10.1007/s00704-013-0921-1>
- Benestad, R.E., Parding, K.M., Erlandsen, H.B., Mezghani, A., 2019. A simple equation to study changes in rainfall statistics. *Environmental Research Letters* 14, 084017. <https://doi.org/10.1088/1748-9326/ab2bb2>
- Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., 2013. Environmental Modelling & Software Characterising performance of environmental models. *Environmental Modelling and Software* 40, 1–20. <https://doi.org/10.1016/j.envsoft.2012.09.011>
- Berndt, C., Haberlandt, U., 2018. Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density. *Journal of Hydrology: Regional Studies* 15, 184–202. <https://doi.org/10.1016/j.ejrh.2018.02.002>
- Beven, K.J., Cloke, H.L., 2012. Comment on “Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water” by Eric F. Wood et al. *Water Resources Research* 48, 1–3. <https://doi.org/10.1029/2011wr010982>
- Biemans, H., Speelman, L.H., Ludwig, F., Moors, E.J., Wiltshire, A.J., Kumar, P., Gerten, D., Kabat, P., 2013. Future water resources for food production in five South Asian river basins and potential for adaptation - A modeling study. *Science of the Total Environment* 468–469, S117–S131. <https://doi.org/10.1016/j.scitotenv.2013.05.092>

- Bierkens, M.F.P., Bell, V.A., Burek, P., Chaney, N., Condon, L.E., David, C.H., de Roo, A., Döll, P., Drost, N., Famiglietti, J.S., Flörke, M., Gochis, D.J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R.M., Reager, J.T., Samaniego, L., Sudicky, E., Sutanudjaja, E.H., van de Giesen, N., Winsemius, H., Wood, E.F., 2015. Hyper-resolution global hydrological modelling: What is next?: “Everywhere and locally relevant” M. F. P. Bierkens et al. Invited Commentary. *Hydrological Processes* 29, 310–320. <https://doi.org/10.1002/hyp.10391>
- Boko, M., Niang, I., Nyong, A., Vogel, C., Githeko, A., Medany, M., Osman-Elasha, B., Tabo, R., Yanda, P., 2007. Africa, in: Frederick, S., Mohamed, S. (Eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution Of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom, pp. 433–467.
- Bontemps, S., Defourny, P., Bogaert, E. Van, Kalogirou, V., Perez, J.R., 2011. GLOBCOVER 2009 Products Description and Validation Report, ESA Bulletin.
- Bosilovich, M.G., Lucchesi, R., Suarez, M., 2016. MERRA-2: File Specification. *Earth* 9, 73.
- Breiman, L., 2001. Random forests. *Random Forests* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brunet, M., Jones, P., 2011. Data rescue initiatives: Bringing historical climate data into the 21st century. *Climate Research* 47, 29–40. <https://doi.org/10.3354/cr00960>
- Buma, W.G., Lee, S., Seo, J.Y., 2016. Hydrological evaluation of Lake Chad basin using space borne and hydrological model observations. *Water (Switzerland)* 8, 1–15. <https://doi.org/10.3390/w8050205>
- Byers, E., Gidden, M., Leclere, D., Balkovic, J., Burek, P., Ebi, K., Greve, P., Grey, D., Havlik, P., Hillers, A., Johnson, N., Kahil, T., Krey, V., Langan, S., Nakicenovic, N., Novak, R., Obersteiner, M., Pachauri, S., Palazzo, A., Parkinson, S., Rao, N.D., Rogelj, J., Satoh, Y., Wada, Y., Willaarts, B., Riahi, K., 2018. Global exposure and vulnerability to multi-sector development and climate change hotspots. *Environmental Research Letters* 13. <https://doi.org/10.1088/1748-9326/aabf45>
- Cai, X., Zeng, R., Kang, W.H., Song, J., Valocchi, A.J., 2015. Strategic Planning for Drought Mitigation under Climate Change. *Journal of Water Resources Planning and Management* 141, 1–10. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000510](https://doi.org/10.1061/(asce)wr.1943-5452.0000510)
- 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, 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Carter, T.R., Rovere, E.L.L., Jones, R.N., Leemans, R., Mearns, L.O., Nakicenovic, N., Pittock, A.B., Semenov, S.M., Skea, J., 2001. Developing and Applying Scenarios In *Climate Change 2001: In Climate Change 2001: Impacts, Adaptation and Vulnerability*, in: Gupta, S., Hulme, M. (Eds.), *Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom, pp. 145–190.
- Chandler, R.E., 2013. Exploiting strength, discounting weakness: Combining information from multiple climate simulators. *Philosophical Transactions of the Royal Society A* 371, 1–19. <https://doi.org/10.1098/rsta.2012.0388>
- Chase, T.N., Pielke, R.A., Kittel, T.G.F., Nemani, R.R., Running, S.W., 2000. Simulated impacts of historical land cover changes on global climate in northern winter. *Climate Dynamics* 16, 93–105. <https://doi.org/10.1007/s003820050007>
- Chellaney, B., 2016. Water, peace and war: Confronting the global water crisis. *Asia Pacific Peace Studies* 1, 71–86. <https://doi.org/10.1080/13623699.2014.890378>
- Chen, D., Rojas, M., Samset, B.H., Cobb, K., Niang, A.D., Edwards, P., Emori, S., Faria, S.H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S.K., Plattner,

- G.-K., Tréguier, A.-M., 2021. Framing, Context, and Methods, in: Chuersuwan, N., Hegerl, G., Yasunari, T. (Eds.), *N Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 147–286. <https://doi.org/10.1017/9781009157896.003>
- Chen, F.W., Liu, C.W., 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment* 10, 209–222. <https://doi.org/10.1007/s10333-012-0319-1>
- Chen, M., Xie, P., Janowiak, J.E., 2002. Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology* 3, 249–266. <https://doi.org/10.1175/1525-7541>
- Cherif, R., Bouteffeha, M., Gargouri-ellouze, E., Eslamian, S., 2023. Hydrologic models classification, calibration, and validation, *Handbook of HydroInformatics*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-821961-4.00023-3>
- Chouchane, H., Krol, M.S., Hoekstra, A.Y., 2018. Virtual water trade patterns in relation to environmental and socioeconomic factors: A case study for Tunisia. *Science of the Total Environment* 613, 287–297. <https://doi.org/10.1016/j.scitotenv.2017.09.032>
- Christ, M., Kempa-Liehr, A.W., Feindt, M., 2016. Distributed and parallel time series feature extraction for industrial big data applications, *Neurocomputing*. Karlsruhe, Germany.
- Chu, J.T., Xia, J., Xu, C.Y., Singh, V.P., 2010. Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theoretical and Applied Climatology* 99, 149–161. <https://doi.org/10.1007/s00704-009-0129-6>
- Clark, M.P., Fan, Y., Lawrence, D.M., Adam, J.C., Bolster, D., Gochis, D.J., Hooper, R.P., Kumar, M., Leung, L.R., Mackay, D.S., Maxwell, R.M., 2015. Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research* 51, 5929–5956. <https://doi.org/10.1002/2015WR017096>
- Clark, M.P., Schaefli, B., Schymanski, S.J., Samaniego, L., Luce, C.H., Jackson, B.M., Freer, J.E., Arnold, J.R., Moore, R.D., Instanbulluoglu, E., Ceola, S., 2016. Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research* 52, 2350–2365. <https://doi.org/10.1002/2015WR017910>
- Coe, M.T., Birkett, C.M., 2004. Calculation of river discharge and prediction of lake height from satellite radar altimetry: Example for the Lake Chad basin. *Water Resources Research* 40, 1–11. <https://doi.org/10.1029/2003WR002543>
- Coe, M.T., Foley, J.A., 2001. Human and natural impacts on the water resources of the Lake Chad basin. *Journal of Geophysical Research Atmospheres* 106, 3349–3356. <https://doi.org/10.1029/2000JD900587>
- Collins, M., 2017. Still weighting to break the model democracy. *Geophysical Research Letters* 44, 3328–3329. <https://doi.org/10.1002/2017GL073370>
- Collins, M., 2007. Ensembles and probabilities: A new era in the prediction of climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365, 1957–1970. <https://doi.org/10.1098/rsta.2007.2068>
- Conway, D., Pereschino, A., Ardoin-Bardin, S., Hamandawana, H., Dieulin, C., Mahé, G., 2009. Rainfall and water resources variability in sub-Saharan Africa during the twentieth century. *Journal of Hydrometeorology* 10, 41–59. <https://doi.org/10.1175/2008JHM1004.1>
- Cook, B.I., Mankin, J.S., Marvel, K., Williams, A.P., Smerdon, J.E., Anchukaitis, K.J., 2020. Twenty-First Century Drought Projections in the CMIP6 Forcing Scenarios. *Earth's Future* 8, 1–20. <https://doi.org/10.1029/2019EF001461>



- Cook, K.H., 2008. The mysteries of Sahel droughts. *Nature Geoscience* 1, 647–648. <https://doi.org/10.1038/ngeo320>
- Couasnon, A., Eilander, D., Muis, S., Veldkamp, T.I.E., Haigh, I.D., Wahl, T., Winsemius, H.C., Ward, P.J., 2020. Measuring compound flood potential from river discharge and storm surge extremes at the global scale. *Natural Hazards and Earth System Sciences* 20, 489–504. <https://doi.org/10.5194/nhess-20-489-2020>
- Coumou, D., Lehmann, J., Beckmann, J., 2015. The weakening summer circulation in the Northern Hemisphere mid-latitudes. *Science* 348, 324–327. <https://doi.org/10.1126/science.1261768>
- Coz, L.M., Delclaux, F., Genthon, P., Favreau, G., 2009. Assessment of Digital Elevation Model (DEM) aggregation methods for hydrological modeling: Lake Chad basin, Africa. *Computers and Geosciences* 35, 1661–1670. <https://doi.org/10.1016/j.cageo.2008.07.009>
- Creswell, J.W., 2009. *Research design: Qualitative, quantitative, and mixed methods approaches*, 3rd ed, University of Nebraska-Lincoln. SAGE Publications Ltd, London, UK. <https://doi.org/10.1128/microbe.4.485.1>
- Crouch, M.L., Jacobs, H.E., Speight, V.L., 2021. Defining domestic water consumption based on personal water use activities. *Aqua Water Infrastructure, Ecosystems and Society* 70, 1002–1011. <https://doi.org/10.2166/aqua.2021.056>
- Čuček, L., Klemeš, J.J., Varbanov, P.S., Kravanja, Z., 2015. Significance of environmental footprints for evaluating sustainability and security of development. *Clean Technologies and Environmental Policy* 17, 2125–2141. <https://doi.org/10.1007/s10098-015-0972-3>
- Cuthbert, M.O., Gleeson, T., Moosdorf, N., Befus, K.M., Schneider, A., Hartmann, J., Lehner, B., 2019. Global patterns and dynamics of climate – groundwater interactions. *Nature Climate Change* 9, 137–141. <https://doi.org/10.1038/s41558-018-0386-4>
- De Carvalho, J.R.P., Almeida Monteiro, J.E.B., Nakai, A.M., Assad, E.D., 2017. Model for Multiple Imputation to Estimate Daily Rainfall Data and Filling of Faults. *Revista Brasileira de Meteorologia* 32, 575–583. <https://doi.org/10.1590/0102-7786324006>
- de Castro-Pardo, M., Fernández Martínez, P., Pérez Zabaleta, A., 2022. An initial assessment of water security in Europe using a DEA approach. *Sustainable Technology and Entrepreneurship* 1, 100002. <https://doi.org/10.1016/j.stae.2022.100002>
- Déqué, M., 2007. Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change* 57, 16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>
- Dessai, S., Hulme, M., 2007. Assessing the robustness of adaptation decisions to climate change uncertainties: A case study on water resources management in the East of England. *Global Environmental Change* 17, 59–72. <https://doi.org/10.1016/j.gloenvcha.2006.11.005>
- Di Capua, G., Coumou, D., 2016. Changes in meandering of the Northern Hemisphere circulation. *Environmental Research Letters* 11, 094028. <https://doi.org/10.1088/1748-9326/11/9/094028>
- Diffenbaugh, N.S., Giorgi, F., 2012. Climate change hotspots in the CMIP5 global climate model ensemble. *Climatic Change* 114, 813–822. <https://doi.org/10.1007/s10584-012-0570-x>
- Diffenbaugh, N.S., Swain, D.L., Touma, D., Lubchenco, J., 2015. Anthropogenic warming has increased drought risk in California. *Proceedings of the National Academy of Sciences of the United States of America* 112, 3931–3936. <https://doi.org/10.1073/pnas.1422385112>
- Döll, P., Trautmann, T., Gerten, D., Schmied, H.M., Ostberg, S., Saaed, F., Schleussner, C.F.,

2018. Risks for the global freshwater system at 1.5 °c and 2 °c global warming. *Environmental Research Letters* 13, 044038. <https://doi.org/10.1088/1748-9326/aab792>
- Donohue, R.J., Roderick, M.L., McVicar, T.R., Yang, Y., 2017. A simple hypothesis of how leaf and canopy-level transpiration and assimilation respond to elevated CO<sub>2</sub> reveals distinct response patterns between disturbed and undisturbed vegetation. *Journal of Geophysical Research: Biogeosciences* 122, 168–184. <https://doi.org/10.1002/2016JG003505>
- Douville, H., Raghavan, K., Renwick, J., Allan, R.P., Arias, P.A., Barlow, M., Cerezo-Mota, R., Cherchi, A., Gan, T.Y., Gergis, J., Jiang, D., Khan, A., Mba, W.P., Rosenfeld, D., Tierney, J., Zolina, O., 2021. Water Cycle Changes, in: Braconnot, P., Diedhiou, A. (Eds.), *In Climate Change 2021 – The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1055–1210. <https://doi.org/10.1017/9781009157896.010>
- Du, L., Rajib, A., Merwade, V., 2018. Large scale spatially explicit modeling of blue and green water dynamics in a temperate mid-latitude basin. *Journal of Hydrology* 562, 84–102. <https://doi.org/10.1016/j.jhydrol.2018.02.071>
- Duan, H., Zhang, G., Wang, S., Fan, Y., 2019. Robust climate change research: A review on multi-model analysis. *Environmental Research Letters* 14, 033001. <https://doi.org/10.1088/1748-9326/aaf8f9>
- 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. <https://doi.org/10.1016/j.jhydrol.2018.12.026>
- Elliott, J., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., Glotter, M., Flörke, M., Wada, Y., Best, N., Eisner, S., Fekete, B.M., Folberth, C., Foster, I., Gosling, S.N., Haddeland, I., Khabarov, N., Ludwig, F., Masaki, Y., Olin, S., Rosenzweig, C., Ruane, A.C., Satoh, Y., Schmid, E., Stacke, T., Tang, Q., Wisser, D., 2014. Constraints and potentials of future irrigation water availability on agricultural production under climate change. *Proceedings of the National Academy of Sciences of the United States of America* 111, 3239–3244. <https://doi.org/10.1073/pnas.1222474110>
- Eromo, S., Adane, C., Santosh, A., Pingale, M., 2016. Assessment of the impact of climate change on surface hydrological processes using SWAT : a case study of Omo-Gibe river basin , Ethiopia. *Modeling Earth Systems and Environment* 2, 1–15. <https://doi.org/10.1007/s40808-016-0257-9>
- Esmaili-Gisavandani, H., Lotfirad, M., Sofla, M.S.D., Ashrafzadeh, A., 2021. Improving the performance of rainfall-runoff models using the gene expression programming approach. *Journal of Water and Climate Change* 12, 3308–3329. <https://doi.org/10.2166/wcc.2021.064>
- Eum, H. Il, Dibike, Y., Prowse, T., Bonsal, B., 2014. Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada. *Hydrological Processes* 28, 4250–4271. <https://doi.org/10.1002/hyp.10236>
- Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* 9, 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Faiz, M.A., Liu, D., Fu, Q., Sun, Q., Li, M., Baig, F., Li, T., Cui, S., 2018. How accurate are the performances of gridded precipitation data products over Northeast China? *Atmospheric Research* 211, 12–20. <https://doi.org/10.1016/j.atmosres.2018.05.006>

- Faramarzi, M., Abbaspour, K.C., Ashraf Vaghefi, S., Farzaneh, M.R., Zehnder, A.J.B., Srinivasan, R., Yang, H., 2013. Modeling impacts of climate change on freshwater availability in Africa. *Journal of Hydrology* 480, 85–101. <https://doi.org/10.1016/j.jhydrol.2012.12.016>
- Fatahi, R., Pardis, N., Hamid, Y., Vanani, R., Ostad, K., Askari, A., Nouri, J., 2021. Eco-hydrologic stability zonation of dams and power plants using the combined models of SMCE and CEQUALW2. *Applied Water Science* 11, 1–7. <https://doi.org/10.1007/s13201-021-01427-z>
- Faticchi, S., Ivanov, V.Y., Paschalis, A., Peleg, N., Molnar, P., Rimkus, S., Kim, J., Burlando, P., Caporali, E., 2016. Uncertainty partition challenges the predictability of vital details of climate change. *Earth's Future* 4, 240–251. <https://doi.org/10.1002/2015EF000336>
- Fawzy, S., Osman, A.I., Doran, J., Rooney, D.W., 2020. Strategies for mitigation of climate change: a review. *Environmental Chemistry Letters* 18, 2069–2094. <https://doi.org/10.1007/s10311-020-01059-w>
- Fazeli Farsani, I., Farzaneh, M.R., Besalatpour, A.A., Salehi, M.H., Faramarzi, M., 2019. Assessment of the impact of climate change on spatiotemporal variability of blue and green water resources under CMIP3 and CMIP5 models in a highly mountainous watershed. *Theoretical and Applied Climatology* 136, 169–184. <https://doi.org/10.1007/s00704-018-2474-9>
- Feng, L., Zhou, T., Wu, B., Li, T., Luo, J.J., 2011. Projection of future precipitation change over China with a high-resolution global atmospheric model. *Advances in Atmospheric Sciences* 28, 464–476. <https://doi.org/10.1007/s00376-010-0016-1>
- Fischer, E.M., Knutti, R., 2016. Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change* 6, 986–991. <https://doi.org/10.1038/nclimate3110>
- Fischer, R., Nowicki, S., Kelley, M., Schmidt, G.A., 2014. A system of conservative regridding for ice-atmosphere coupling in a General Circulation Model (GCM). *Geoscientific Model Development* 7, 883–907. <https://doi.org/10.5194/gmd-7-883-2014>
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Rummukainen, M., 2013. Evaluation of climate models. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 741–866. <https://doi.org/10.1017/CBO9781107415324.020>
- Flörke, M., Schneider, C., McDonald, R.I., 2018. Water competition between cities and agriculture driven by climate change and urban growth. *Nature Sustainability* 1, 51–58. <https://doi.org/10.1038/s41893-017-0006-8>
- Folberth, C., Elliott, J., Müller, C., Balkovič, J., Chrýssanthacopoulos, J., Izaurralde, R.C., Jones, C.D., Khabarov, N., Liu, W., Reddy, A., Schmid, E., Skalský, R., Yang, H., Arneith, A., Ciais, P., Deryng, D., Lawrence, P.J., Olin, S., Pugh, T.A.M., Ruane, A.C., Wang, X., 2019. Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble, *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0221862>
- Fowler, H.J., Lenderink, G., Prein, A.F., Westra, S., Allan, R.P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H.X., Guerreiro, S., Haerter, J.O., Kendon, E.J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., Zhang, X., 2021. Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth and Environment* 2, 107–122. <https://doi.org/10.1038/s43017-020-00128-6>
- Fu, G., Liu, Z., Charles, S.P., Xu, Z., Yao, Z., 2013. A score-based method for assessing the performance of GCMs: A case study of southeastern Australia. *Journal of Geophysical*

- Research Atmospheres 118, 4154–4167. <https://doi.org/10.1002/jgrd.50269>
- Gabbert, S., van Ittersum, M., Kroeze, C., Stalpers, S., Ewert, F., Alkan Olsson, J.A., 2010. Uncertainty analysis in integrated assessment: The users' perspective. *Regional Environmental Change* 10, 131–143. <https://doi.org/10.1007/s10113-009-0100-1>
- Gal, G., Makler-pick, V., Shachar, N., 2014. Environmental Modelling & Software Dealing with uncertainty in ecosystem model scenarios: Application of the single-model ensemble approach \*. *Environmental Modelling and Software* 61, 360–370. <https://doi.org/10.1016/j.envsoft.2014.05.015>
- Galli, A., Wiedmann, T., Ercin, E., Knoblauch, D., Ewing, B., Giljum, S., 2012. Integrating Ecological, Carbon and Water footprint into a “footprint Family” of indicators: Definition and role in tracking human pressure on the planet. *Ecological Indicators* 16, 100–112. <https://doi.org/10.1016/j.ecolind.2011.06.017>
- Gampe, D., Ludwig, R., 2017. Evaluation of gridded precipitation data products for hydrological applications in complex topography. *Hydrology* 4, 1–21. <https://doi.org/10.3390/hydrology4040053>
- Gampe, D., Schmid, J., Ludwig, R., 2019. Impact of reference dataset selection on RCM evaluation, bias correction, and resulting climate change signals of precipitation. *Journal of Hydrometeorology* 20, 1813–1828. <https://doi.org/10.1175/JHM-D-18-0108.1>
- Gao, H., Bohn, T.J., Podest, E., McDonald, K.C., Lettenmaier, D.P., 2011. On the causes of the shrinking of Lake Chad. *Environmental Research Letters* 6, 1–7. <https://doi.org/10.1088/1748-9326/6/3/034021>
- Gay, L.R., Mills, G.E., Airasian, P.W., 2009. *Educational Research: Competencies for Analysis and Applications*, 9th ed. Pearson Education, Inc, Upper Saddle River: NJ. <https://doi.org/10.4236/oalib.1101538>
- Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2010. Variable selection using random forests. *Pattern Recognition Letters* 31, 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Gesualdo, G.C., Oliveira, P.T.S., Rodrigues, D.B.B., Gupta, H. V., 2019. Assessing Water Security in the Sao Paulo Metropolitan Region Under Projected Climate Change. *Hydrology and Earth System Sciences* 23, 4955–4968. <https://doi.org/10.5194/hess-23-2019>
- Ghimire, U., Srinivasan, G., Agarwal, A., 2019. Assessment of rainfall bias correction techniques for improved hydrological simulation. *International Journal of Climatology* 39, 2386–2399. <https://doi.org/10.1002/joc.5959>
- Gibson, S.J., Narendra, A., Dainotti, M.G., Bogdan, M., Pollo, A., Poliszczuk, A., Rinaldi, E., Liidakis, I., 2022. Using Multivariate Imputation by Chained Equations to Predict Redshifts of Active Galactic Nuclei. *Frontiers in Astronomy and Space Sciences* 9, 1–16. <https://doi.org/10.3389/fspas.2022.836215>
- Gidden, M.J., Riahi, K., Smith, S.J., Fujimori, S., Luderer, G., Kriegler, E., Van Vuuren, D.P., Van Den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J.C., Frank, S., Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., Takahashi, K., 2019. Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development* 12, 1443–1475. <https://doi.org/10.5194/gmd-12-1443-2019>
- Giorgi, F., Mearns, L.O., 2002. Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the “Reliability Ensemble Averaging” (REA) Method. *Journal of Climate* 15, 1141–1158. [https://doi.org/10.1175/1520-0442\(2002\)016](https://doi.org/10.1175/1520-0442(2002)016)
- Gleckler, P.J., Taylor, K.E., Doutriaux, C., 2008. Performance metrics for climate models. *Journal of Geophysical Research Atmospheres* 113, 1–20.

- <https://doi.org/10.1029/2007JD008972>
- Gong, G., An, X., Mahato, N.K., Sun, S., Chen, S., Wen, Y., 2019. Research on short-term load prediction based on Seq2seq model. *Energies* 12, 1–18. <https://doi.org/10.3390/en12163199>
- Gosling, S.N., Arnell, N.W., 2016. A global assessment of the impact of climate change on water scarcity. *Climatic Change* 134, 371–385. <https://doi.org/10.1007/s10584-013-0853-x>
- Goyal, M.K., Burn, D.H., Ojha, C.S.P., 2012. Evaluation of machine learning tools as a statistical downscaling tool: Temperatures projections for multi-stations for Thames River Basin, Canada. *Theoretical and Applied Climatology* 108, 519–534. <https://doi.org/10.1007/s00704-011-0546-1>
- Grose, M.R., Narsey, S., Delage, F.P., Dowdy, A.J., Bador, M., Boschat, G., Chung, C., Kajtar, J.B., Rauniyar, S., Freund, M.B., Lyu, K., Rashid, H., Zhang, X., Wales, S., Trenham, C., Holbrook, N.J., Cowan, T., Alexander, L., Arblaster, J.M., Power, S., 2020. Insights From CMIP6 for Australia’s Future Climate. *Earth’s Future* 8, 1–24. <https://doi.org/10.1029/2019EF001469>
- Gu, H., Yu, Z., Wang, J., Wang, G., Yang, T., Ju, Q., Yang, C., Xu, F., Fan, C., 2015. Assessing CMIP5 general circulation model simulations of precipitation and temperature over China. *International Journal of Climatology* 35, 2431–2440. <https://doi.org/10.1002/joc.4152>
- Guerreiro, S.B., Fowler, H.J., Barbero, R., Westra, S., Lenderink, G., Blenkinsop, S., Lewis, E., Li, X.F., 2018. Detection of continental-scale intensification of hourly rainfall extremes. *Nature Climate Change* 8, 803–807. <https://doi.org/10.1038/s41558-018-0245-3>
- 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, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gusain, A., Ghosh, S., Karmakar, S., 2020. Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research* 232, 104680. <https://doi.org/10.1016/j.atmosres.2019.104680>
- GWP, 2013. The Lake Chad Basin Aquifer System, Transboundary groundwater factsheet. Stockholm, Sweden.
- Gyau-Boakye, P., Schultz, G.A., 1994. Filling gaps in runoff time series in west africa. *Hydrological Sciences Journal* 39, 621–636. <https://doi.org/10.1080/02626669409492784>
- Hajnayeb, A., Ghasemloonia, A., Khadem, S.E., Moradi, M.H., 2011. Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis. *Expert Systems with Applications* 38, 10205–10209. <https://doi.org/10.1016/j.eswa.2011.02.065>
- Hammami, D., Lee, T.S., Ouarda, T.B.M.J., Le, J., 2012. Predictor selection for downscaling GCM data with LASSO. *Journal of Geophysical Research Atmospheres* 117, D17116. <https://doi.org/10.1029/2012JD017864>
- Harp, R.D., Horton, D.E., 2022. Observed Changes in Daily Precipitation Intensity in the United States. *Geophysical Research Letters* 49, e2022GL099955. <https://doi.org/10.1029/2022GL099955>
- Hassan, Ibrahim, Kalin, R.M., White, C.J., Aladejana, J.A., 2020. Evaluation of Daily Gridded Meteorological Datasets over the Niger Delta Region of Nigeria and Implication to Water Resources Management. *Atmospheric and Climate Sciences* 10, 21–39. <https://doi.org/10.4236/acs.2020.101002>

- Hassan, I, Kalin, R.M., White, C.J., Aladejana, J.A., 2020. Selection of CMIP5 GCM ensemble for the projection of spatio-temporal changes in precipitation and temperature over the Niger Delta, Nigeria. *Water (Switzerland)* 12, 1–19. <https://doi.org/10.3390/w12020385>
- Hattermann, F.F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Florke, F., Gosling, S.N., Hoffmann, P., Liersch, S., Masaki, Y., Motovilov, Y., Muller, C., Samaniego, L., Stacke, T., Wada, Y., Yang, T., Krysanova, V., 2018. Sources of uncertainty in hydrological climate impact assessment: a cross-scale study. *Environmental Research Letters* 13, 015006. <https://doi.org/10.1088/1748-9326/aa9938>
- Hawkins, E., Sutton, R., 2009. The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society* 90, 1095–1107. <https://doi.org/10.1175/2009BAMS2607.1>
- Hein, A., Condon, L., Maxwell, R., 2019. Evaluating the relative importance of precipitation , temperature and land-cover change in the hydrologic response to extreme meteorological drought conditions over the North American High Plains. *Hydrology and Earth System Sciences* 23, 1931–1950. <https://doi.org/10.5194/hess-23-1931-2019>
- Hejazi, M.I., Edmonds, J., Clarke, L., Kyle, P., Davies, E., Chaturvedi, V., Wise, M., Patel, P., Eom, J., Calvin, K., 2014. Integrated assessment of global water scarcity over the 21st century under multiple climate change mitigation policies. *Hydrology and Earth System Sciences* 18, 2859–2883. <https://doi.org/10.5194/hess-18-2859-2014>
- 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. <https://doi.org/10.1016/j.jhydrol.2017.03.008>
- Henry, B.M., 1945. Nonparametric Tests Against Trend. *Econometrica* 13, 245–259.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hirpa, F.A., Alfieri, L., Lees, T., Peng, J., Dyer, E., Dadson, S.J., 2019. Streamflow response to climate change in the Greater Horn of Africa. *Climatic Change* 156, 341–363. <https://doi.org/10.1007/s10584-019-02547-x>
- Hoekstra, A.Y., 2009. Human appropriation of natural capital: A comparison of ecological footprint and water footprint analysis. *Ecological Economics* 68, 1963–1974. <https://doi.org/10.1016/j.ecolecon.2008.06.021>
- Hoekstra, A.Y., Chapagain, A.K., Aldaya, M.M., Mekonnen, M.M., 2011. The Water Footprint Assessment Manual. Setting the Global Standard, Water Footprint Network. Earthscan, London, UK. <https://doi.org/10.1080/0969160x.2011.593864>
- Hoekstra, A.Y., Hung, P.Q., 2002. Virtual Water trade: a quantification of virtual water flows between nations in relation to crop trade. Value of Water Research Report Series. Delft, The Netherlands.
- Hoekstra, A.Y., Mekonnen, M.M., Chapagain, A.K., Mathews, R.E., Richter, B.D., 2012. Global monthly water scarcity: Blue water footprints versus blue water availability. *PLoS ONE* 7. <https://doi.org/10.1371/journal.pone.0032688>
- Holtanova, E., Kalvova, J., Pisoft, P., Miksovsky, J., 2014. Uncertainty in regional climate model outputs over the Czech Republic: The role of nested and driving models. *International Journal of Climatology* 34, 27–35. <https://doi.org/10.1002/joc.3663>
- Holthuijzen, M., Beckage, B., Clemens, P.J., Higdson, D., Winter, J.M., 2022. Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method: Advantages of a linear correction for extremes. *Theoretical and Applied Climatology* 149, 863–882. <https://doi.org/10.1007/s00704-022-04035-2>
- Holzkaemper, A., Klein, T., Seppelt, R., Fuhrer, J., 2015. Assessing the propagation of

- uncertainties in multi-objective optimization for agro-ecosystem adaptation to climate change. *Environmental Modelling and Software* 66, 27–35. <https://doi.org/10.1016/j.envsoft.2014.12.012>
- Hu, Z., Chen, Y., Yao, L., Wei, C., Li, C., 2016. Optimal allocation of regional water resources: From a perspective of equity-efficiency tradeoff. *Resources, Conservation and Recycling* 109, 102–113. <https://doi.org/10.1016/j.resconrec.2016.02.001>
- Huang, W., Wang, S., Chan, J.C.L., 2011. Discrepancies between global reanalyses and observations in the interdecadal variations of Southeast Asian cold surge. *International Journal of Climatology* 31, 2272–2280. <https://doi.org/10.1002/joc.2234>
- Hughes, D.A., Mantel, S., Mohobane, T., 2014. An assessment of the skill of downscaled GCM outputs in simulating historical patterns of rainfall variability in South Africa. *Hydrology Research* 45, 134–147. <https://doi.org/10.2166/nh.2013.027>
- Hur, J., Ihm, S., Park, Y., 2017. A variable impacts measurement in random forest for mobile cloud computing. *Wireless Communications and Mobile Computing* 2017, 1–13. <https://doi.org/10.1155/2017/6817627>
- IPCC, 2018. Global warming of 1.5°C, An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development., <https://doi.org/10.1002/9780470996621.ch50>
- Isotta, F.A., Frei, C., Weigluni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S.M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., Vertačnik, G., 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, 1657–1675. <https://doi.org/10.1002/joc.3794>
- Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software* 21, 602–614. <https://doi.org/10.1016/j.envsoft.2006.01.004>
- Jensen, O., Wu, H., 2018. Urban water security indicators: Development and pilot. *Environmental Science and Policy* 83, 33–45. <https://doi.org/10.1016/j.envsci.2018.02.003>
- Jiang, L., Wu, H., Tao, J., Kimball, J.S., Alfieri, L., Chen, X., 2020. Satellite-based evapotranspiration in hydrological model calibration. *Remote Sensing* 12, 428. <https://doi.org/10.3390/rs12030428>
- Jones, B., O’Neill, B.C., 2016. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environmental Research Letters* 11, 084003. <https://doi.org/10.1088/1748-9326/11/8/084003>
- 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, 487–503. <https://doi.org/10.1175/1525-7541>
- Kannan, S.S., Ramaraj, N., 2010. A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm. *Knowledge-Based Systems* 23, 580–585. <https://doi.org/10.1016/j.knosys.2010.03.016>
- Karner, K., Mitter, H., Schmid, E., 2019. The economic value of stochastic climate information for agricultural adaptation in a semi-arid region in Austria. *Journal of Environmental Management* 249, 109431. <https://doi.org/10.1016/j.jenvman.2019.109431>
- Kendall, M.G., 1948. *Rank Correlation Methods*, 4th ed. Griffin, London.
- Keys, P.W., Falkenmark, M., 2018. Green water and African sustainability. *Food Security* 10,

- 537–548. <https://doi.org/10.1007/s12571-018-0790-7>
- Keys, P.W., Wang-Erlandsson, L., Gordon, L.J., 2016. Revealing invisible Water: Moisture recycling as an ecosystem service. *PLoS ONE* 11, 1–16. <https://doi.org/10.1371/journal.pone.0151993>
- Khan, N., Shahid, S., Juneng, L., Ahmed, K., Ismail, T., Nawaz, N., 2019. Prediction of heat waves in Pakistan using quantile regression forests. *Atmospheric Research* 221, 1–11. <https://doi.org/10.1016/j.atmosres.2019.01.024>
- Kim, J., Ivanov, V.Y., Fatichi, S., 2016. Climate change and uncertainty assessment over a hydroclimatic transect of Michigan. *Stochastic Environmental Research and Risk Assessment* 30, 923–944. <https://doi.org/10.1007/s00477-015-1097-2>
- Kim, J.H., Chung, E.S., Song, J.Y., Shamsuddin, S., 2023. Quantifying the uncertainty in future runoff projection over SSP scenarios, GCMs and hydrological model parameters.
- Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* 42, 1–5. <https://doi.org/10.1029/2005WR004362>
- Kirchner, M., Mitter, H., Schneider, U.A., Sommer, M., Falkner, K., Schmid, E., 2021. Uncertainty concepts for integrated modeling - Review and application for identifying uncertainties and uncertainty propagation pathways. *Environmental Modelling and Software* 135, 104905. <https://doi.org/10.1016/j.envsoft.2020.104905>
- Knutti, R., Masson, D., Gettelman, A., 2013. Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters* 40, 1194–1199. <https://doi.org/10.1002/grl.50256>
- Knutti, R., Sedláček, J., Sanderson, B.M., Lorenz, R., Fischer, E.M., Eyring, V., 2017. A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters* 44, 1909–1918. <https://doi.org/10.1002/2016GL072012>
- Kohler, M.A., 1949. Double-mass analysis for testing the consistency of records and for making adjustments. *Bulletin of the American Meteorological Society* 30, 188–189.
- Konapala, G., Mishra, A., Leung, L.R., 2017. Changes in temporal variability of precipitation over land due to anthropogenic forcings. *Environmental Research Letters* 12, 024009. <https://doi.org/10.1088/1748-9326/aa568a>
- Konapala, G., Mishra, A.K., Wada, Y., Mann, M.E., 2020. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications* 11, 1–10. <https://doi.org/10.1038/s41467-020-16757-w>
- Koppa, A., Gebremichael, M., Yeh, W.W.G., 2019. Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach. *Advances in Water Resources* 130, 129–146. <https://doi.org/10.1016/j.advwatres.2019.06.005>
- Krabbenhoft, C.A., Allen, G.H., Lin, P., Godsey, S.E., Allen, D.C., Burrows, R.M., DelVecchia, A.G., Fritz, K.M., Shanafield, M., Burgin, A.J., Zimmer, M.A., Datry, T., Dodds, W.K., Jones, C.N., Mims, M.C., Franklin, C., Hammond, J.C., Zipper, S., Ward, A.S., Costigan, K.H., Beck, H.E., Olden, J.D., 2022. Assessing placement bias of the global river gauge network. *Nature Sustainability* 5, 586–592. <https://doi.org/10.1038/s41893-022-00873-0>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences* 22, 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., Simpson, J., 1998. The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology* 15, 809–817. [https://doi.org/10.1175/1520-0426\(1998\)015](https://doi.org/10.1175/1520-0426(1998)015)



- Kundzewicz, Z.W., Schellnhuber, H.J., 2004. Floods in the IPCC TAR perspective. *Natural Hazards* 31, 111–128. <https://doi.org/10.1023/B:NHAZ.0000020257.09228.7b>
- Kunnath-Poovakka, A., Ryu, D., Renzullo, L.J., George, B., 2016. The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction. *Journal of Hydrology* 535, 509–524. <https://doi.org/10.1016/j.jhydrol.2016.02.018>
- Kursa, M.B., 2016. Embedded all relevant feature selection with random ferns. *ArXiv abs/1604.0*, 1–13. [https://doi.org/10.1007/978-3-319-60438-1\\_30](https://doi.org/10.1007/978-3-319-60438-1_30)
- Kursa, M.B., Jankowski, A., Rudnicki, W.R., 2010. Boruta - A system for feature selection. *Fundamenta Informaticae* 101, 271–285. <https://doi.org/10.3233/FI-2010-288>
- Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the boruta package. *Journal of Statistical Software* 36, 1–13. <https://doi.org/10.18637/jss.v036.i11>
- Kyselý, J., Plavcová, E., 2010. A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. *Journal of Geophysical Research Atmospheres* 115, 1–14. <https://doi.org/10.1029/2010JD014123>
- Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., Nikulin, G., Majone, B., 2018. Testing the Hydrological Coherence of High-Resolution Gridded Precipitation and Temperature Data Sets. *Water Resources Research* 54, 1999–2016. <https://doi.org/10.1002/2017WR021633>
- Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A., Olsson, E., 2016. A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Quarterly Journal of the Royal Meteorological Society* 142, 2132–2142. <https://doi.org/10.1002/qj.2813>
- Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated environmental modeling: A vision and roadmap for the future. *Environmental Modelling and Software* 39, 3–23. <https://doi.org/10.1016/j.envsoft.2012.09.006>
- Lanzante, J.R., Dixon, K.W., Adams-Smith, D., Nath, M.J., Whitlock, C.E., 2021. Evaluation of some distributional downscaling methods as applied to daily precipitation with an eye towards extremes. *International Journal of Climatology* 41, 3186–3202. <https://doi.org/10.1002/joc.7013>
- Launiainen, S., Futter, M.N., Ellison, D., Clarke, N., Finér, L., Högbom, L., Laurén, A., Ring, E., 2014. Is the water footprint an appropriate tool for forestry and forest products: The fennoscandian case. *Ambio* 43, 244–256. <https://doi.org/10.1007/s13280-013-0380-z>
- Lawal, I.M., Bertram, D., White, C.J., Jagaba, A.H., 2023a. Integrated framework for hydrologic modelling in data-sparse watersheds and climate change impact on projected green and blue water sustainability. *Frontiers in Environmental Science* 11, 1–26. <https://doi.org/10.3389/fenvs.2023.1233216>
- Lawal, I.M., Bertram, D., White, C.J., Jagaba, A.H., Hassan, I., Shuaibu, A., 2021. Multi-criteria performance evaluation of gridded precipitation and temperature products in data-sparse regions. *Atmosphere* 12, 1597. <https://doi.org/10.3390/atmos12121597>
- Lawal, I.M., Bertram, D., White, C.J., Kutty, S.R.M., Hassan, I., Jagaba, A.H., 2023b. Application of Boruta algorithms as a robust methodology for performance evaluation of CMIP6 general circulation models for hydro - climatic studies. *Theoretical and Applied Climatology* 153, 113–135. <https://doi.org/10.1007/s00704-023-04466-5>
- 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, 1–18. <https://doi.org/10.1029/2011JD016187>

- Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E.M., Brunner, L., Knutti, R., Hawkins, E., 2020. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics* 11, 491–508. <https://doi.org/10.5194/esd-11-491-2020>
- Lemoalle, J., 2014. Le développement du lac Tchad/Development of Lake Chad: Situation actuelle et futurs possibles/Current Situation and Possible Outcomes. Expertise collégiale. France Institut de Recherche pour le Développement (IRD), Marseille. <https://doi.org/http://doi.org/10.4000/books.irdeditions.11648>
- Lemoalle, J., Bader, J.C., Leblanc, M., Sedick, A., 2012. Recent changes in Lake Chad: Observations, simulations and management options (1973–2011). *Global and Planetary Change* 80, 247–254. <https://doi.org/10.1016/j.gloplacha.2011.07.004>
- Lemordant, L., Gentine, P., Swann, A.S., Cook, B.I., Scheff, J., 2018. Critical impact of vegetation physiology on the continental hydrologic cycle in response to increasing CO<sub>2</sub>. *Proceedings of the National Academy of Sciences of the United States of America* 115, 4093–4098. <https://doi.org/10.1073/pnas.1720712115>
- Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., Dech, S., Wegmann, M., 2012. Modelling forest  $\alpha$ -diversity and floristic composition - on the added value of LiDAR plus hyperspectral remote sensing. *Remote Sensing* 4, 2818–2845. <https://doi.org/10.3390/rs4092818>
- Li, C., Wang, L., Wanrui, W., Qi, J., Linshan, Y., Zhang, Y., Lei, W., Cui, X., Wang, P., 2018. An analytical approach to separate climate and human contributions to basin streamflow variability. *Journal of Hydrology* 559, 30–42. <https://doi.org/10.1016/j.jhydrol.2018.02.019>
- Li, W., Jiang, Z., Xu, J., Li, L., 2016. Extreme precipitation indices over China in CMIP5 models. Part II: Probabilistic projection. *Journal of Climate* 29, 8989–9004. <https://doi.org/10.1175/JCLI-D-16-0377.1>
- Liu, J., Yang, H., Gosling, S.N., Kummu, M., Flörke, M., Pfister, S., Hanasaki, N., Wada, Y., Zhang, X., Zheng, C., Alcamo, J., Oki, T., 2017. Water scarcity assessments in the past, present, and future. *Earth's Future* 5, 545–559. <https://doi.org/10.1002/2016EF000518>
- Liu, M., Zhang, P., Cai, Y., Chu, J., Li, Y., Wang, X., Li, C., Liu, Q., 2023. Spatial-temporal heterogeneity analysis of blue and green water resources for Poyang Lake basin, China. *Journal of Hydrology* 617, 128983. <https://doi.org/10.1016/j.jhydrol.2022.128983>
- López, P.L., Sutanudjaja, E.H., Schellekens, J., Sterk, G., Bierkens, M.F.P., 2017. Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences* 21, 3125–3144. <https://doi.org/10.5194/hess-21-3125-2017>
- Luken, K.J., Padhy, R., Wang, X.R., 2021. Missing Data Imputation for Galaxy Redshift Estimation 1–8. <https://doi.org/10.1101/2021.11.13806>
- Lyu, B., Zhang, Y., Hu, Y., 2017. Improving PM<sub>2.5</sub> air quality model forecasts in China using a bias-correction framework. *Atmosphere* 147, 1–15. <https://doi.org/10.3390/atmos8080147>
- Maggioni, V., Massari, C., 2018. On the performance of satellite precipitation products in riverine flood modeling: A review. *Journal of Hydrology* 558, 214–224. <https://doi.org/10.1016/j.jhydrol.2018.01.039>
- Magrin, G., 2016. The disappearance of Lake Chad: History of a myth. *Journal of Political Ecology* 23, 204–222. <https://doi.org/10.2458/v23i1.20191>
- Mahmood, R., Jia, S., 2019a. Observed and simulated hydro-climatic data for the lake Chad basin, Africa. *Data in Brief* 25, 1–5. <https://doi.org/10.1016/j.dib.2019.104043>
- Mahmood, R., Jia, S., 2019b. Assessment of hydro-climatic trends and causes of dramatically declining stream flow to Lake Chad, Africa, using a hydrological approach. *Science of*

- the Total Environment 675, 122–140. <https://doi.org/10.1016/j.scitotenv.2019.04.219>
- Mahmood, R., Jia, S., Zhu, W., 2019. Analysis of climate variability, trends, and prediction in the most active parts of the Lake Chad basin, Africa. *Scientific Reports* 9, 1–18. <https://doi.org/10.1038/s41598-019-42811-9>
- Maldonado, S., Weber, R., 2009. A wrapper method for feature selection using Support Vector Machines. *Information Sciences* 179, 2208–2217. <https://doi.org/10.1016/j.ins.2009.02.014>
- Manatsa, D., Chingombe, W., Matarira, C.H., 2008. The impact of the positive Indian Ocean dipole on Zimbabwe droughts Tropical climate is understood to be dominated by. *International Journal of Climatology* 28, 2011–2029. <https://doi.org/10.1002/joc.1695>
- Mao, G., Liu, J., Han, F., Meng, Y., Tian, Y., Zheng, Y., Zheng, C., 2020. Assessing the interlinkage of green and blue water in an arid catchment in Northwest China. *Environmental Geochemistry and Health* 42, 933–953. <https://doi.org/10.1007/s10653-019-00406-3>
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L.O., 2017. Towards process-informed bias correction of climate change simulations. *Nature Climate Change* 7, 764–773. <https://doi.org/10.1038/nclimate3418>
- Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themel, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* 48, 1–34. <https://doi.org/10.1029/2009RG000314>
- Marengo, J.A., Nobre, C.A., Tomasella, J., Oyama, M.D., de Oliveira, G.S., de Oliveira, R., Camargo, H., Alves, L.M., Brown, I.F., 2008. The drought of Amazonia in 2005. *Journal of Climate* 21, 495–516. <https://doi.org/10.1175/2007JCLI1600.1>
- Mariotti, L., Diallo, I., Coppola, E., Giorgi, F., 2014. Seasonal and intraseasonal changes of African monsoon climates in 21st century CORDEX projections. *Climatic Change* 125, 53–65. <https://doi.org/10.1007/s10584-014-1097-0>
- Martin, J.T., Pederson, G.T., Woodhouse, C.A., Cook, E.R., McCabe, G.J., Anchukaitis, K.J., Wise, E.K., Erger, P.J., Dolan, L., McGuire, M., Gangopadhyay, S., Chase, K.J., Littell, J.S., Gray, S.T., George, S.S., Friedman, J.M., Sauchyn, D.J., St-Jacques, J.M., King, J., 2020. Increased drought severity tracks warming in the United States’ largest river basin. *Proceedings of the National Academy of Sciences of the United States of America* 117. <https://doi.org/10.1073/pnas.1916208117>
- Martin, K., Marcel, S., Annika, K., Fred, H., Dennis, T., Tobias, P., André, M., 2020. Stop Floating, Start Swimming, Water and climate change - interlinkages and prospects for future action. Bonn and Eschborn, Germany.
- Martinez-García, F.P., Contreras-De-villar, A., Muñoz-Perez, J.J., 2021. Review of wind models at a local scale: Advantages and disadvantages. *Journal of Marine Science and Engineering* 9, 318. <https://doi.org/10.3390/jmse9030318>
- Martinsen, G., Liu, S., Mo, X., Bauer-Gottwein, P., 2019. Joint optimization of water allocation and water quality management in Haihe River basin. *Science of the Total Environment* 654, 72–84. <https://doi.org/10.1016/j.scitotenv.2018.11.036>
- Masud, M.B., McAllister, T., Cordeiro, M.R.C., Faramarzi, M., 2018. Modeling future water footprint of barley production in Alberta, Canada: Implications for water use and yields to 2064. *Science of the Total Environment* 616–617, 208–222. <https://doi.org/10.1016/j.scitotenv.2017.11.004>
- Maxino, C.C., McAvaney, B.J., Pitman, A.J., Perkins, S.E., 2008. Ranking the AR4 climate

- models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *International Journal of Climatology* 28, 1097–1112. <https://doi.org/10.1002/joc.1612>
- McMahon, T.A., Peel, M.C., Karoly, D.J., 2015. Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation. *Hydrology and Earth System Sciences* 19, 361–377. <https://doi.org/10.5194/hess-19-361-2015>
- McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C.D., Verdin, J.P., 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data* 4, 1–19. <https://doi.org/10.1038/sdata.2017.12>
- McSweeney, C.F., Jones, R.G., Lee, R.W., Rowell, D.P., 2015. Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics* 44, 3237–3260. <https://doi.org/10.1007/s00382-014-2418-8>
- Mekonnen, M.M., Hoekstra, A.Y., 2012. A Global Assessment of the Water Footprint of Farm Animal Products. *Ecosystems* 15, 401–415. <https://doi.org/10.1007/s10021-011-9517-8>
- Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* 29, 897–910. <https://doi.org/10.1175/JTECH-D-11-00103.1>
- Menne, M.J., Williams, C.N., Gleason, B.E., Jared Rennie, J., Lawrimore, J.H., 2018. The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *Journal of Climate* 31, 9835–9854. <https://doi.org/10.1175/JCLI-D-18-0094.1>
- Miao, C., Duan, Q., Yang, L., Borthwick, A.G.L., 2012. On the Applicability of Temperature and Precipitation Data from CMIP3 for China. *PLoS ONE* 7, 1–10. <https://doi.org/10.1371/journal.pone.0044659>
- Milly, P.C.D., Dunne, K.A., 2016. Potential evapotranspiration and continental drying. *Nature Climate Change* 6, 946–949. <https://doi.org/10.1038/nclimate3046>
- Min, S., Hense, A., 2006. A Bayesian assessment of climate change using multimodel ensembles. Part I: Global mean surface temperature. *Journal of Climate* 19, 3237–3256. <https://doi.org/10.1175/JCLI3784.1>
- Mishra, A., Liu, S.C., 2014. Changes in precipitation pattern and risk of drought over India in the context of global warming. *Journal of Geophysical Research: Atmospheres* 119, 7833–7841. <https://doi.org/doi:10.1002/2014JD021471>
- Mitter, H., Schmid, E., 2019. Computing the economic value of climate information for water stress management exemplified by crop production in Austria. *Agricultural Water Management* 221, 430–448. <https://doi.org/10.1016/j.agwat.2019.04.005>
- Mitter, H., Techen, A.K., Sinabell, F., Helming, K., Kok, K., Priess, J.A., Schmid, E., Bodirsky, B.L., Holman, I., Lehtonen, H., Leip, A., Le Mouél, C., Mathijs, E., Mehdi, B., Michetti, M., Mittenzwei, K., Mora, O., Øygarden, L., Reidsma, P., Schaldach, R., Schönhart, M., 2019. A protocol to develop Shared Socio-economic Pathways for European agriculture. *Journal of Environmental Management* 252, 109701. <https://doi.org/10.1016/j.jenvman.2019.109701>
- Mohsenipour, M., Shahid, S., Chung, E., Wang, X., 2018. Changing Pattern of Droughts during Cropping Seasons of Bangladesh. *Water Resources Management* 32, 1555–1568. <https://doi.org/10.1007/s11269-017-1890-4>
- Montaldo, N., Oren, R., 2018. Changing Seasonal Rainfall Distribution With Climate Directs Contrasting Impacts at Evapotranspiration and Water Yield in the Western Mediterranean Region. *Earth's Future* 6, 841–856. <https://doi.org/10.1029/2018EF000843>
- Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., Van Vuuren, D.P.,

- Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P., Wilbanks, T.J., 2010. The next generation of scenarios for climate change research and assessment. *Nature* 463, 747–756. <https://doi.org/10.1038/nature08823>
- Muratoglu, A., Iraz, E., Ercin, E., 2022. Water resources management of large hydrological basins in semi-arid regions: Spatial and temporal variability of water footprint of the Upper Euphrates River basin. *Science of the Total Environment* 846, 157396. <https://doi.org/10.1016/j.scitotenv.2022.157396>
- Naderi, M., Parsa, S., 2022. The difference of water footprint and availability as a physical metric for sustainable water use and management. *Ecohydrology* 15. <https://doi.org/10.1002/eco.2397>
- Nashwan, M.S., Shahid, S., 2019. Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmospheric Research* 230, 1–10. <https://doi.org/10.1016/j.atmosres.2019.104632>
- Nashwan, M.S., Shahid, S., Chung, E.S., 2019. Development of high-resolution daily gridded temperature datasets for the central north region of Egypt. *Scientific Data* 6, 1–13. <https://doi.org/10.1038/s41597-019-0144-0>
- Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A., Ramirez-Villegas, J., 2020. High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Scientific Data* 7, 1–14. <https://doi.org/10.1038/s41597-019-0343-8>
- Ndehedehe, C.E., Agutu, N.O., Okwuashi, O., Ferreira, V.G., 2016. Spatio-temporal variability of droughts and terrestrial water storage over Lake Chad Basin using independent component analysis. *Journal of Hydrology* 540, 106–128. <https://doi.org/10.1016/j.jhydrol.2016.05.068>
- Ndehedehe, C.E., Awange, J.L., Agutu, N.O., Okwuashi, O., 2018. Changes in hydro-meteorological conditions over tropical West Africa (1980–2015) and links to global climate. *Global and Planetary Change* 162, 321–341. <https://doi.org/10.1016/j.gloplacha.2018.01.020>
- Neelin, J.D., Sahany, S., Stechmann, S.N., Bernstein, D.N., 2017. Global warming precipitation accumulation increases above the current-climate cutoff scale, in: *Proceedings of the National Academy of Sciences of the United States of America*. Pnas, Washington DC, pp. 1258–1263. <https://doi.org/10.1073/pnas.1615333114>
- Neupane, R.P., Kumar, S., 2015. Estimating the effects of potential climate and land use changes on hydrologic processes of a large agriculture dominated watershed. *Journal of Hydrology* 529, 418–429. <https://doi.org/10.1016/j.jhydrol.2015.07.050>
- New, M., Hulme, M., Jones, P., 2000. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *Journal of Climate* 13, 2217–2238. [https://doi.org/10.1175/1520-0442\(2000\)013](https://doi.org/10.1175/1520-0442(2000)013)
- Niang, I., Ruppel, O.C., Abdrabo, M.A., Essel, A., Lennard, C., Padgham, J., Urquhart, P., 2014. Africa, in: Dube, P., Leary, N. (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199–1265.
- Nkiaka, E., Bryant, R.G., Ntajal, J., Biao, E.I., 2022. How useful are gridded water resources reanalysis and evapotranspiration 1 products for assessing water security in ungauged basins? *Hydrology and Earth System Discuss.* [preprint] 1–29.
- Nkiaka, E., Nawaz, N.R., Lovett, J.C., 2018a. Effect of single and multi-site calibration techniques on hydrological model performance, parameter estimation and predictive uncertainty: a case study in the Logone catchment, Lake Chad basin. *Stochastic*

- Environmental Research and Risk Assessment 32, 1665–1682. <https://doi.org/10.1007/s00477-017-1466-0>
- Nkiaka, E., Nawaz, N.R., Lovett, J.C., 2017. Analysis of rainfall variability in the Logone catchment, Lake Chad basin. *International Journal of Climatology* 37, 3553–3564. <https://doi.org/10.1002/joc.4936>
- Nkiaka, E., Nawaz, R., Lovett, J.C., 2018b. Assessing the reliability and uncertainties of projected changes in precipitation and temperature in Coupled Model Intercomparison Project phase 5 models over the Lake Chad basin. *International Journal of Climatology* 38, 5136–5152. <https://doi.org/10.1002/joc.5717>
- Novoa, V., Ahumada-Rudolph, R., Rojas, O., Munizaga, J., Sáez, K., Arumí, J.L., 2019. Sustainability assessment of the agricultural water footprint in the Cachapoal River basin, Chile. *Ecological Indicators* 98, 19–28. <https://doi.org/10.1016/j.ecolind.2018.10.048>
- O'Neill, B.C., Tebaldi, C., Van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K., Sanderson, B.M., 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development* 9, 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Ocio, D., Beskeen, T., Smart, K., 2019. Fully distributed hydrological modelling for catchmentwide hydrological data verification. *Hydrology Research* 50, 1520–1534. <https://doi.org/10.2166/nh.2019.006>
- Ogden, F.L., 2021. *Geohydrology: Hydrological Modeling*, 2nd ed, Encyclopedia of Geology, 2nd edition. Elsevier Ltd. <https://doi.org/10.1016/B978-0-08-102908-4.00115-6>
- Ogutu, B.O., D'Adamo, F., Dash, J., 2021. Impact of vegetation greening on carbon and water cycle in the African Sahel-Sudano-Guinean region. *Global and Planetary Change* 202, 103524. <https://doi.org/10.1016/j.gloplacha.2021.103524>
- Olden, J.D., Kennard, M.J., Pusey, B.J., 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* 5, 503–518. <https://doi.org/10.1002/eco.251>
- Ouallouche, F., Lazri, M., Ameer, S., 2018. Improvement of rainfall estimation from MSG data using Random Forests classification and regression. *Atmospheric Research* 211, 62–72. <https://doi.org/10.1016/j.atmosres.2018.05.001>
- Oudin, L., Andre, V., Perrin, C., Michel, C., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research* 44, W03413. <https://doi.org/10.1029/2007WR006240>
- Palazzi, E., Von Hardenberg, J., Provenzale, A., 2013. Precipitation in the hindu-kush karakoram himalaya: Observations and future scenarios. *Journal of Geophysical Research Atmospheres* 118, 85–100. <https://doi.org/10.1029/2012JD018697>
- Panda, K.C., Singh, R.M., Thakural, L.N., Sahoo, D.P., 2022. Representative grid location-multivariate adaptive regression spline (RGL-MARS) algorithm for downscaling dry and wet season rainfall. *Journal of Hydrology* 605, 127381. <https://doi.org/10.1016/j.jhydrol.2021.127381>
- Pang, J., Zhang, H., Xu, Q., Wang, Yujie, Wang, Yunqi, Zhang, O., Hao, J., 2020. Hydrological evaluation of open-access precipitation data using SWAT at multiple temporal and spatial scales. *Hydrology and Earth System Sciences* 24, 3603–3626. <https://doi.org/10.5194/hess-24-3603-2020>
- Parish, E.S., Kodra, E., Steinhaeuser, K., Ganguly, A.R., 2012. Estimating future global per capita water availability based on changes in climate and population. *Computers and*

- Geosciences 42, 79–86. <https://doi.org/10.1016/j.cageo.2012.01.019>
- Pattnayak, K.C., Abdel-Lathif, A.Y., Rathakrishnan, K. V., Singh, M., Dash, R., Maharana, P., 2019. Changing Climate Over Chad: Is the Rainfall Over the Major Cities Recovering? *Earth and Space Science* 6, 1149–1160. <https://doi.org/10.1029/2019EA000619>
- Pellicer-Martínez, F., Martínez-Paz, J.M., 2018. Probabilistic evaluation of the water footprint of a river basin: Accounting method and case study in the Segura River Basin, Spain. *Science of the Total Environment* 627, 28–38. <https://doi.org/10.1016/j.scitotenv.2018.01.223>
- Pellicer-Martínez, F., Martínez-Paz, J.M., 2016. The Water Footprint as an indicator of environmental sustainability in water use at the river basin level. *Science of the Total Environment* 571, 561–574. <https://doi.org/10.1016/j.scitotenv.2016.07.022>
- Pendergrass, A.G., Hartmann, D.L., 2014. Changes in the distribution of rain frequency and intensity in response to global warming. *Journal of Climate* 27, 8372–8383. <https://doi.org/10.1175/JCLI-D-14-00183.1>
- Perkins, S.E., Pitman, A.J., Holbrook, N.J., McAneney, J., 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *Journal of Climate* 20, 4356–4376. <https://doi.org/10.1175/JCLI4253.1>
- Pfister, S., Koehler, A., Hellweg, S., 2009. Assessing the environmental impacts of freshwater consumption in LCA. *Environmental Science and Technology* 43, 4098–4104. <https://doi.org/10.1021/es802423e>
- Pham-Duc, B., Sylvestre, F., Papa, F., Frappart, F., Bouchez, C., Crétaux, J.F., 2020. The Lake Chad hydrology under current climate change. *Scientific Reports* 10, 1–10. <https://doi.org/10.1038/s41598-020-62417-w>
- Poli, P., Hersbach, H., Dee, D.P., Berrisford, P., Simmons, A.J., Vitart, F., Laloyaux, P., Tan, D.G.H., Peubey, C., Thépaut, J.N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., Fisher, M., 2016. ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate* 29, 4083–4097. <https://doi.org/10.1175/JCLI-D-15-0556.1>
- Pour, S.H., Shahid, S., Chung, E., Wang, X., 2018. Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmospheric Research* 213, 149–162. <https://doi.org/10.1016/j.atmosres.2018.06.006>
- Prakash, S., Gairola, R.M., Mitra, A.K., 2015a. Comparison of large-scale global land precipitation from multisatellite and reanalysis products with gauge-based GPCP data sets. *Theoretical and Applied Climatology* 121, 303–317. <https://doi.org/10.1007/s00704-014-1245-5>
- Prakash, S., Mitra, A.K., Momin, I.M., Rajagopal, E.N., 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, 2326–2338. <https://doi.org/10.1002/joc.4129>
- Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2019. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena* 177, 149–166. <https://doi.org/10.1016/j.catena.2019.02.012>
- Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmospheric Research* 197, 42–63. <https://doi.org/10.1016/j.atmosres.2017.06.014>
- Prein, A.F., Gobiet, A., 2017. Impacts of uncertainties in European gridded precipitation

- observations on regional climate analysis. *International Journal of Climatology* 37, 305–327. <https://doi.org/10.1002/joc.4706>
- Prudhomme, C., Giuntoli, I., Robinson, E.L., Clark, D.B., Arnell, N.W., Dankers, R., Fekete, B.M., Franssen, W., Gerten, D., Gosling, S.N., Hagemann, S., Hannah, D.M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., Wisser, D., 2014. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, in: *Proceedings of the National Academy of Sciences of the United States of America*. pp. 3262–3267. <https://doi.org/10.1073/pnas.1222473110>
- Pushpalatha, R., Perrin, C., Moine, N. Le, Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology* 420–421, 171–182. <https://doi.org/10.1016/j.jhydrol.2011.11.055>
- Quinteiro, P., Rafael, S., Villanueva-Rey, P., Ridoutt, B., Lopes, M., Arroja, L., Dias, A.C., 2018a. A characterisation model to address the environmental impact of green water flows for water scarcity footprints. *Science of the Total Environment* 626, 1210–1218. <https://doi.org/10.1016/j.scitotenv.2018.01.201>
- Quinteiro, P., Ridoutt, B.G., Arroja, L., Dias, A.C., 2018b. Identification of methodological challenges remaining in the assessment of a water scarcity footprint: a review. *International Journal of Life Cycle Assessment* 23, 164–180. <https://doi.org/10.1007/s11367-017-1304-0>
- Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology* 536, 192–207. <https://doi.org/10.1016/j.jhydrol.2016.02.037>
- Raju, K.S., Sonali, P., Kumar, N.D., 2017. Ranking of CMIP5-based global climate models for India using compromise programming. *Theoretical and Applied Climatology* 128, 563–574. <https://doi.org/10.1007/s00704-015-1721-6>
- Raju, S.K., Kumar, N.D., 2016. Selection of global climate models for India using cluster analysis. *Journal of Water and Climate Change* 7, 764–774. <https://doi.org/10.2166/wcc.2016.112>
- Ranzani, A., Bonato, M., Patro, E.R., Gaudard, L., Michele, C. De, 2018. Hydropower Future : Between Climate Change , Renewable Deployment , Carbon and Fuel Prices. *Water (Switzerland)* 10, 1197. <https://doi.org/10.3390/w10091197>
- Rashid, M.M., Beecham, S., Chowdhury, R.K., 2015. Statistical downscaling of CMIP5 outputs for projecting future changes in rainfall in the Onkaparinga catchment. *Science of the Total Environment* 530–531, 171–182. <https://doi.org/10.1016/j.scitotenv.2015.05.024>
- Razavi, T., Coulibaly, P., Asce, M., 2013. Streamflow Prediction in Ungauged Basins : Review of Regionalization Methods. *Journal of Hydrologic Engineering* 18, 958–975. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000690](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690)
- Refsgaard, J.C., Arnbjerg-Nielsen, K., Drews, M., Halsnæs, K., Jeppesen, E., Madsen, H., Markandya, A., Olesen, J.E., Porter, J.R., Christensen, J.H., 2013. The role of uncertainty in climate change adaptation strategies-A Danish water management example. *Mitigation and Adaptation Strategies for Global Change* 18, 337–359. <https://doi.org/10.1007/s11027-012-9366-6>
- Reichler, T., Kim, J., 2008. How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society* 89, 303–311. <https://doi.org/10.1175/BAMS-89-3-303>
- Reilly, M., Willenbockel, D., 2010. Managing uncertainty: A review of food system scenario analysis and modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 3049–3063. <https://doi.org/10.1098/rstb.2010.0141>
- Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N.,



- Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>
- Richter, B.D., Davis, M.M., Apse, C., Konrad, C., 2012. A presumptive standard for environmental flow protection. *River Research and Applications* 28, 1312–1321. <https://doi.org/10.1002/rra>
- Rockström, J., Falkenmark, M., Karlberg, L., Hoff, H., Rost, S., Gerten, D., 2009. Future water availability for global food production: The potential of green water for increasing resilience to global change. *Water Resources Research* 45, 1–16. <https://doi.org/10.1029/2007WR006767>
- Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoin, H.K., Landerer, F.W., Lo, M.H., 2018. Emerging trends in global freshwater availability. *Nature* 557, 651–659. <https://doi.org/10.1038/s41586-018-0123-1>
- Rodrigues, D.B.B., Gupta, H. V., Mendiondo, E.M., 2014. A blue/green water-based accounting framework for assessment of water security. *Water Resources Research* 50, 7187–7205. <https://doi.org/10.1002/2013WR014274>.
- Rodríguez, E., Sánchez, I., Duque, N., Arboleda, P., Vega, C., Zamora, D., López, P., Kaune, A., Werner, M., García, C., Burke, S., 2020. Combined Use of Local and Global Hydro Meteorological Data with Hydrological Models for Water Resources Management in the Magdalena - Cauca Macro Basin – Colombia. *Water Resources Management* 34, 2179–2199. <https://doi.org/10.1007/s11269-019-02236-5>
- Romanski, P., Lars, K., 2018. Selecting Attributes. Package ‘FSelector.’
- Roth, V., Lemann, T., 2016. Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands. *Hydrology and Earth System Sciences* 20, 921–934. <https://doi.org/10.5194/hess-20-921-2016>
- Sa’adi, Z., Shahid, S., Pour, S.H., Ahmed, K., Chung, E., Yaseen, Z.M., 2020. Multi-variable model output statistics downscaling for the projection of spatio-temporal changes in rainfall of Borneo Island. *Journal of Hydro-Environment Research* 31, 62–75. <https://doi.org/10.1016/j.jher.2020.05.002>
- Sachindra, D.A., Huang, F., Barton, A., Perera, B.J.C., 2014. Statistical downscaling of general circulation model outputs to precipitation-part 1: Calibration and validation. *International Journal of Climatology* 34, 3264–3281. <https://doi.org/10.1002/joc.3914>
- Salih, A.A.M., Elagib, N.A., Tjernström, M., Zhang, Q., 2018. Characterization of the Sahelian-Sudan rainfall based on observations and regional climate models. *Atmospheric Research* 202, 205–218. <https://doi.org/10.1016/j.atmosres.2017.12.001>
- Salman, S A, Shahid, S., Ismail, T., Ahmed, K., Wang, X., 2018. Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. *Atmospheric Research* 213, 509–522. <https://doi.org/10.1016/j.atmosres.2018.07.008>
- Salman, Saleem A., Shahid, S., Ismail, T., Ahmed, K., Wang, X.J., 2018. Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. *Atmospheric Research* 213, 509–522. <https://doi.org/10.1016/j.atmosres.2018.07.008>
- Salman, S.A., Shahid, S., Ismail, T., Al-Abadi, A.M., Wang, X. jun, Chung, E.S., 2019.

- Selection of gridded precipitation data for Iraq using compromise programming. *Measurement: Journal of the International Measurement Confederation* 132, 87–98. <https://doi.org/10.1016/j.measurement.2018.09.047>
- Santoso, H., Idinoba, M., Imbach, P., 2008. *Climate Scenarios: What we need to know and how to generate them*, Working Paper. Bogor, Indonesia.
- Sarch, M.T., Birkett, C., 2000. Fishing and farming at Lake Chad: Responses to lake-level fluctuations. *Geographical Journal* 166, 156–172. <https://doi.org/10.1111/j.1475-4959.2000.tb00015.x>
- Sarr, B., 2012. Present and future climate change in the semi-arid region of West Africa: A crucial input for practical adaptation in agriculture. *Atmospheric Science Letters* 13, 108–112. <https://doi.org/10.1002/asl.368>
- Savelsberg, J., Schillinger, M., Schlecht, I., Weigt, H., 2018. The Impact of Climate Change on Swiss Hydropower. *Sustainability (Switzerland)* 10, 2541. <https://doi.org/10.3390/su10072541>
- Savenije, H.H.G., 2000. Water scarcity indicators; the deception of the numbers. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* 25, 199–204. [https://doi.org/10.1016/S1464-1909\(00\)00004-6](https://doi.org/10.1016/S1464-1909(00)00004-6)
- Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M., Stender, P., 2014. Global gridded precipitation over land: A description of the new GPCP First Guess Daily product. *Earth System Science Data* 6, 49–60. <https://doi.org/10.5194/essd-6-49-2014>
- Scheff, J., Frierson, D.M.W., 2014. Scaling potential evapotranspiration with greenhouse warming. *Journal of Climate* 27, 1539–1558. <https://doi.org/10.1175/JCLI-D-13-00233.1>
- Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N.W., Clark, D.B., Dankers, R., Eisner, S., Fekete, B.M., Colón-González, F.J., Gosling, S.N., Kim, H., Liu, X., Masaki, Y., Portmann, F.T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., Kabat, P., 2014. Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences of the United States of America* 111, 3245–3250. <https://doi.org/10.1073/pnas.1222460110>
- Schneider, C., 2013. Three Shades of Water: Increasing Water Security with Blue, Green, and Gray Water, *CSA News*. <https://doi.org/10.2134/csa2013-58-10-1>
- Schoof, J.T., Pryor, S.C., 2003. Evaluation of the NCEP-NCAR reanalysis in terms of synoptic-scale phenomena: A case study from the Midwestern USA. *International Journal of Climatology* 23, 1725–1741. <https://doi.org/10.1002/joc.969>
- Schuol, J., Abbaspour, K.C., Yang, H., Srinivasan, R., Zehnder, A.J.B., 2008. Modeling blue and green water availability in Africa. *Water Resources Research* 44, 1–18. <https://doi.org/10.1029/2007WR006609>
- Sehad, M., Lazri, M., Ameer, S., 2017. Novel SVM-based technique to improve rainfall estimation over the Mediterranean region (north of Algeria) using the multispectral MSG SEVIRI imagery. *Advances in Space Research* 59, 1381–1394. <https://doi.org/10.1016/j.asr.2016.11.042>
- SEMA, 2022. Flood Updates: General overview of flood incidences across the state. Yobe, Nigeria.
- Sen, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association* 63, 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Senbel, M., McDaniels, T., Dowlatabadi, H., 2003. The ecological footprint: A non-monetary metric of human consumption applied to North America. *Global Environmental Change*

- 13, 83–100. [https://doi.org/10.1016/S0959-3780\(03\)00009-8](https://doi.org/10.1016/S0959-3780(03)00009-8)
- Seneviratne, S.I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A. Di, Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S.M., Wehner, M., Zhou, B., 2021. Weather and Climate Extreme Events in a Changing Climate. In: Chan, J., Sorteberg, A., Vera, C. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766. <https://doi.org/10.1017/9781009157896.013>
- Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate* 19, 3088–3111. <https://doi.org/10.1175/JCLI3790.1>
- Shekhar, A., Shapiro, C.A., 2019. What do meteorological indices tell us about a long-term tillage study? *Soil and Tillage Research* 193, 161–170. <https://doi.org/10.1016/j.still.2019.06.004>
- Shiferaw, B., Tesfaye, K., Kassie, M., Abate, T., Prasanna, B.M., Menkir, A., 2014. Managing vulnerability to drought and enhancing livelihood resilience in sub-Saharan Africa: Technological, institutional and policy options. *Weather and Climate Extremes* 3, 67–79. <https://doi.org/10.1016/j.wace.2014.04.004>
- Shiru, M.S., Chung, E., 2021. Performance evaluation of CMIP6 global climate models for selecting models for climate projection over Nigeria. *Theoretical and Applied Climatology* 146, 599–615. <https://doi.org/10.1007/s00704-021-03746-2>
- Shiru, M.S., Shahid, S., Chung, E.S., Alias, N., 2019a. Changing characteristics of meteorological droughts in Nigeria during 1901–2010. *Atmospheric Research* 223, 60–73. <https://doi.org/10.1016/j.atmosres.2019.03.010>
- Shiru, M.S., Shahid, S., Chung, E.S., Alias, N., Scherer, L., 2019b. A MCDM-based framework for selection of general circulation models and projection of spatio-temporal rainfall changes: A case study of Nigeria. *Atmospheric Research* 225, 1–16. <https://doi.org/10.1016/j.atmosres.2019.03.033>
- Shiru, M.S., Shahid, S., Shiru, S., Chung, E.S., Alias, N., Ahmed, K., Dioha, E.C., Sa’adi, Z., Salman, S., Noor, M., Nashwan, M.S., Idlan, M.K., Khan, N., Momade, M.H., Houmsi, M.R., Iqbal, Z., Ishanch, Q., Sediqi, M.N., 2019c. Challenges in water resources of Lagos mega city of Nigeria in the context of climate change. *Journal of Water and Climate Change* 4, 1067–1083. <https://doi.org/10.2166/wcc.2019.047>
- Shirvani, A., Landman, W.A., 2016. Seasonal precipitation forecast skill over Iran. *International Journal of Climatology* 36, 1887–1900. <https://doi.org/10.1002/joc.4467>
- Shreem, S.S., Abdullah, S., Nazri, M.Z.A., 2016. Hybrid feature selection algorithm using symmetrical uncertainty and a harmony search algorithm. *International Journal of Systems Science* 47, 1312–1329. <https://doi.org/10.1080/00207721.2014.924600>
- Shrestha, S., Aihara, Y., Bhattarai, A.P., Bista, N., Kondo, N., Futaba, K., Nishida, K., Shindo, J., 2018. Development of an objective water security index and assessment of its association with quality of life in urban areas of developing countries. *SSM - Population Health* 6, 276–285. <https://doi.org/10.1016/j.ssmph.2018.10.007>
- Shrestha, S., Chapagain, R., Babel, M.S., 2017. Quantifying the impact of climate change on crop yield and water footprint of rice in the Nam Oon Irrigation Project, Thailand. *Science of the Total Environment* 599, 689–699. <https://doi.org/10.1016/j.scitotenv.2017.05.028>
- Shukla, J., DelSole, T., Fennessy, M., Kinter, J., Paolino, D., 2006. Climate model fidelity and projections of climate change. *Geophysical Research Letters* 33, 3–6. <https://doi.org/10.1029/2005GL025579>

- Sian, K.T.C.L.K., Wang, J., Ayugi, B.O., Nooni, I.K., Ongoma, V., 2021. Multi-decadal variability and future changes in precipitation over Southern Africa. *Atmosphere* 12, 1–25. <https://doi.org/10.3390/atmos12060742>
- Simon, H.A., 1996. *The Sciences of the Artificial*, 3rd ed. MIT Press, Cambridge, MA, USA.
- Simonovic, S.P., Li, L., 2004. Sensitivity of the Red River basin flood protection system to climate variability and change. *Water Resources Management* 18, 89–110. <https://doi.org/10.1023/B:WARM.0000024702.40031.b2>
- Singh, V.P., 1995. *Computer models of watershed hydrology*. Water Resources Publications.
- Sivapalan, M., 2003. Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrological Processes* 17, 3163–3170. <https://doi.org/10.1002/hyp.5155>
- Smith, J.B., Hulme, M., 1998. Climate change scenarios, in: In: Feenstra, Jan F Burton, Ian Smith, Joel B Tol, Richard S J (Eds.) *UNEP Handbook on Methods for Climate Change Impact Assessment and Adaptation Strategies*. Version 2.0. Nairobi, Kenya and Institute for Environmental Studies, Amsterdam., pp. 18–19.
- Smith, K.A., Wilby, R.L., Broderick, C., Prudhomme, C., Matthews, T., Harrigan, S., Murphy, C., Analysis, I.C., 2018. Navigating Cascades of Uncertainty — As Easy as ABC? Not Quite . . . *Journal of Extreme Events* 5, 1–8. <https://doi.org/10.1142/S2345737618500070>
- Song, Y., Li, X., Bao, Y., Song, Z., Wei, M., Shu, Q., Yang, X., 2020. FIO-ESM v2.0 Outputs for the CMIP6 Global Monsoons Model Intercomparison Project Experiments. *Advances in Atmospheric Sciences* 37, 1045–1056. <https://doi.org/10.1007/s00376-020-9288-2>
- Stoppiglia, H., Dreyfus, G., Dubois, R., Oussar, Y., 2003. Ranking a Random Feature for Variable and Feature Selection. *Journal of Machine Learning Research* 3, 1399–1414. <https://doi.org/10.1162/153244303322753733>
- Sun, G., Peng, F., Mu, M., 2017. Uncertainty assessment and sensitivity analysis of soil moisture based on model parameter errors – Results from four regions in China. *Journal of Hydrology* 555, 347–360. <https://doi.org/10.1016/j.jhydrol.2017.09.059>
- Sun, J., Li, Y.P., Zhuang, X.W., Jin, S.W., Huang, G.H., Feng, R.F., 2018. Identifying water resources management strategies in adaptation to climate change under uncertainty. *Mitigation and Adaptation Strategies for Global Change* 23, 553–578. <https://doi.org/10.1007/s11027-017-9749-9>
- Swain, S.S., Mishra, A., Sahoo, B., Chatterjee, C., 2020. Water scarcity-risk assessment in data-scarce river basins under decadal climate change using a hydrological modelling approach. *Journal of Hydrology* 590, 125260. <https://doi.org/10.1016/j.jhydrol.2020.125260>
- Sylla, M.B., 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, 1805–1817. <https://doi.org/10.1002/joc.3551>
- Sylla, M.B., Giorgi, F., Pal, J.S., Gibba, P., Kebe, I., Nikiema, M., 2015. Projected changes in the annual cycle of high-intensity precipitation events over West Africa for the late twenty-first century. *Journal of Climate* 28, 6475–6488. <https://doi.org/10.1175/JCLI-D-14-00854.1>
- Sylla, M.B., Pinguinde, M.N., Peter, G., Ibourahima, K., Nana, A.B.K., 2016. Climate Change over West Africa: Recent Trends and Future Projections, in: *Adaptation to Climate Change and Variability in Rural West Africa*. Switzerland, pp. 25–40. <https://doi.org/10.1007/978-3-319-31499-0>
- Sýs, V., Fošumpaur, P., Kašpar, T., 2021. The impact of climate change on the reliability of water resources. *Climate* 9, 153. <https://doi.org/10.3390/cli9110153>

- Szczypta, C., Calvet, J.C., Albergel, C., Balsamo, G., Boussetta, S., Carrer, D., Lafont, S., Meurey, C., 2011. Verification of the new ECMWF ERA-Interim reanalysis over France. *Hydrology and Earth System Sciences* 15, 647–666. <https://doi.org/10.5194/hess-15-647-2011>
- Tabari, H., 2020. Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports* 10, 1–10. <https://doi.org/10.1038/s41598-020-70816-2>
- Tan, M.L., Duan, Z., 2017. Assessment of GPM and TRMM precipitation products over Singapore. *Remote Sensing* 9, 1–16. <https://doi.org/10.3390/rs9070720>
- Tanarhte, M., Hadjinicolaou, P., Lelieveld, J., 2012. Intercomparison of temperature and precipitation data sets based on observations in the Mediterranean and the Middle East. *Journal of Geophysical Research Atmospheres* 117, 1–24. <https://doi.org/10.1029/2011JD017293>
- Tani, S., Gobiet, A., 2019. Quantile mapping for improving precipitation extremes from regional climate models. *Journal of Agrometeorology* 21, 434–443.
- Tapiador, F.J., Turk, F.J., Petersen, W., Hou, A.Y., García-Ortega, E., Machado, L.A.T., Angelis, C.F., Salio, P., Kidd, C., Huffman, G.J., de Castro, M., 2012. Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research* 104, 70–97. <https://doi.org/10.1016/j.atmosres.2011.10.021>
- Taylor, J.R., 1997. Covariance and Correlation, in: *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, Sausalito, CA, p. 217.
- Taylor, K.E., 2005. Taylor Diagram Primer. Working Paper 1–4.
- Taylor, K.E., 2001. in a Single Diagram. *Journal of Geophysical Research* 106, 7183–7192.
- Taylor, R.G., Scanlon, B.R., Döll, P., Rodell, M., van Beek, R., Wada, Y., Bierkens, M.F.P., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.F., Holman, I., Treidel, H., 2013. Ground water and climate change. *Nature Climate Change* 3, 322–329. <https://doi.org/doi.org/10.1038/nclimate1744>
- Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., Moseley, C., Pfeifer, S., Rechid, D., Remedio, A.R., Ries, H., Petersen, J., Preuschmann, S., Raub, T., Saeed, F., Sieck, K., Weber, T., 2013. How does a regional climate model modify the projected climate change signal of the driving GCM: A study over different CORDEX regions using REMO. *Atmosphere* 4, 214–236. <https://doi.org/10.3390/atmos4020214>
- Themeßl, M.J., Gobiet, A., Heinrich, G., 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change* 112, 449–468. <https://doi.org/10.1007/s10584-011-0224-4>
- Thompson, J.R., Polet, G., 2000. Hydrology and land use in a sahelian floodplain Wetland. *Wetlands* 20, 639–659. [https://doi.org/10.1672/0277-5212\(2000\)020\[0639:HALUIA\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2000)020[0639:HALUIA]2.0.CO;2)
- Todzo, S., Bichet, A., Diedhiou, A., 2020. Intensification of the hydrological cycle expected in West Africa over the 21st century. *Earth System Dynamics* 11, 319–328. <https://doi.org/10.5194/esd-11-319-2020>
- Tokarska, K.B., Stolpe, M.B., Sippel, S., Fischer, E.M., Smith, C.J., Lehner, F., Knutti, R., 2020. Past warming trend constrains future warming in CMIP6 models. *Science Advances* 6, 1–14. <https://doi.org/10.1126/sciadv.aaz9549>
- Tortajada, C., González-Gómez, F., Biswas, A.K., Buurman, J., 2019. Water demand management strategies for water-scarce cities: The case of Spain. *Sustainable Cities and Society* 45, 649–656. <https://doi.org/10.1016/j.scs.2018.11.044>

- Trenberth, K.E., Fasullo, J.T., Mackaro, J., 2011. Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *Journal of Climate* 24, 4907–4924. <https://doi.org/10.1175/2011JCLI4171.1>
- Troost, C., Walter, T., Berger, T., 2015. Climate, energy and environmental policies in agriculture: Simulating likely farmer responses in Southwest Germany. *Land Use Policy* 46, 50–64. <https://doi.org/10.1016/j.landusepol.2015.01.028>
- Tzabiras, J., Vasiliades, L., Sidiropoulos, P., Loukas, A., Mylopoulos, N., 2016. Evaluation of Water Resources Management Strategies to Overturn Climate Change Impacts on Lake Karla Watershed. *Water Resources Management* 30, 5819–5844. <https://doi.org/10.1007/s11269-016-1536-y>
- UNEP, 2006. Environment for Development, Africa's Lakes: Atlas of our Changing Environment. Nairobi, Kenya.
- van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software* 45, 1–67. <https://doi.org/10.18637/jss.v045.i03>
- van de Giesen, N., Hut, R., Selker, J., 2014. The Trans-African Hydro-Meteorological Observatory (TAHMO). *WIREs Water* 1, 341–348. <https://doi.org/10.1002/wat2.1034>
- van Vuuren, D.P., Kriegler, E., O'Neill, B.C., Ebi, K.L., Riahi, K., Carter, T.R., Edmonds, J., Hallegatte, S., Kram, T., Mathur, R., Winkler, H., 2014. A new scenario framework for Climate Change Research: Scenario matrix architecture. *Climatic Change* 122, 373–386. <https://doi.org/10.1007/s10584-013-0906-1>
- van Vuuren, D.P., Riahi, K., Calvin, K., Dellink, R., Emmerling, J., Fujimori, S., KC, S., Kriegler, E., O'Neill, B., 2017. The Shared Socio-economic Pathways: Trajectories for human development and global environmental change. *Global Environmental Change* 42, 148–152. <https://doi.org/10.1016/j.gloenvcha.2016.10.009>
- Vanham, D., Bidoglio, G., 2013. A review on the indicator water footprint for the EU28. *Ecological Indicators* 26, 61–75. <https://doi.org/10.1016/j.ecolind.2012.10.021>
- Veettil, A.V., Mishra, A.K., 2018. Potential influence of climate and anthropogenic variables on water security using blue and green water scarcity, Falkenmark index, and freshwater provision indicator. *Journal of Environmental Management* 228, 346–362. <https://doi.org/10.1016/j.jenvman.2018.09.012>
- Veettil, A.V., Mishra, A.K., 2016. Water security assessment using blue and green water footprint concepts. *Journal of Hydrology* 542, 589–602. <https://doi.org/10.1016/j.jhydrol.2016.09.032>
- Velázquez, J.A., Schmid, J., Ricard, S., Muerth, M.J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R., Turcotte, R., 2013. An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrology and Earth System Sciences* 17, 565–578. <https://doi.org/10.5194/hess-17-565-2013>
- Vicente-Serrano, S.M., McVicar, T.R., Miralles, D.G., Yang, Y., Tomas-Burguera, M., 2020. Unraveling the influence of atmospheric evaporative demand on drought and its response to climate change. *Wiley Interdisciplinary Reviews: Climate Change* 11, 1–31. <https://doi.org/10.1002/wcc.632>
- Vicente-Serrano, S.M., Zabalza-Martínez, J., Borràs, G., López-Moreno, J.I., Pla, E., Pascual, D., Savé, R., Biel, C., Funes, I., Azorin-Molina, C., Sanchez-Lorenzo, A., Martín-Hernández, N., Peña-Gallardo, M., Alonso-González, E., Tomas-Burguera, M., El Kenawy, A., 2017. Extreme hydrological events and the influence of reservoirs in a highly regulated river basin of northeastern Spain. *Journal of Hydrology: Regional Studies* 12, 13–32. <https://doi.org/10.1016/j.ejrh.2017.01.004>
- Visessri, S., McIntyre, N., 2016. Regionalisation of hydrological responses under land-use

- change and variable data quality. *Hydrological Sciences Journal* 61, 302–320. <https://doi.org/10.1080/02626667.2015.1006226>
- Vizy, E.K., Cook, K.H., Crétat, J., Neupane, N., 2013. Projections of a wetter sahel in the twenty-first century from global and regional models. *Journal of Climate* 26, 4664–4687. <https://doi.org/10.1175/JCLI-D-12-00533.1>
- Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. *Nature* 467, 555–561. <https://doi.org/10.1038/nature09440>
- Wagena, M.B., Collick, A.S., Ross, A.C., Najjar, R.G., Rau, B., Sommerlot, A.R., Fuka, D.R., Kleinman, P.J.A., Easton, Z.M., 2018. Science of the Total Environment Impact of climate change and climate anomalies on hydrologic and biogeochemical processes in an agricultural catchment of the Chesapeake Bay watershed , USA. *Science of the Total Environment* 637–638, 1443–1454. <https://doi.org/10.1016/j.scitotenv.2018.05.116>
- Wang, D., Liu, J., Shao, W., Mei, C., Su, X., Wang, H., 2021. Comparison of CMIP5 and CMIP6 Multi-Model Ensemble for Precipitation Downscaling Results and Observational Data: The Case of Hanjiang River Basin. *Atmosphere* 12, 1–20. <https://doi.org/https://doi.org/10.3390/atmos12070867>
- Wang, L., Ranasinghe, R., Maskey, S., van Gelder, P.H.A.J.M., Vrijling, J.K., 2016. Comparison of empirical statistical methods for downscaling daily climate projections from CMIP5 GCMs: A case study of the Huai River Basin, China. *International Journal of Climatology* 36, 145–164. <https://doi.org/10.1002/joc.4334>
- Wang, X., Piao, S., Ciais, P., Friedlingstein, P., Myneni, R.B., Cox, P., Heimann, M., Miller, J., Peng, S., Wang, T., Yang, H., Chen, A., 2014. A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature* 506, 212–215. <https://doi.org/10.1038/nature12915>
- Wang, X., Zhang, J., Shahid, S., Guan, E., Wu, Y., Gao, J., He, R., 2016. Adaptation to climate change impacts on water demand. *Mitigation and Adaptation Strategies for Global Change* 21, 81–99. <https://doi.org/10.1007/s11027-014-9571-6>
- Wang, X.J., Zhang, J.Y., Shamsuddin, S., Bi, S.H., He, R.M., Xu, Z., 2015. Assessing water security and adaptation measures in a changing environment, in: *IAHS-AISH Proceedings and Reports*. pp. 129–130. <https://doi.org/10.5194/piahs-366-129-2015>
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., Schewe, J., 2014. The inter-sectoral impact model intercomparison project (ISI-MIP): Project framework, in: *Proceedings of the National Academy of Sciences of the United States of America*. pp. 3228–3232. <https://doi.org/10.1073/pnas.1312330110>
- Washington, R., Harrison, M., Conway, D., Black, E., Challinor, A., Grimes, D., Jones, R., Morse, A., Kay, G., Todd, M., 2006. African climate change: Taking the shorter route. *Bulletin of the American Meteorological Society* 87, 1355–1366. <https://doi.org/10.1175/BAMS-87-10-1355>
- Weigel, A.P., Knutti, R., Liniger, M.A., Appenzeller, C., 2010. Risks of model weighting in multimodel climate projections. *Journal of Climate* 23, 4175–4191. <https://doi.org/10.1175/2010JCLI3594.1>
- Wijngaard, J.B., Klein Tank, A.M.G., Können, G.P., 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* 23, 679–692. <https://doi.org/10.1002/joc.906>
- Wilby, R.L., Dessai, S., 2010. Robust adaptation to climate change. *Weather* 65, 180–185. <https://doi.org/10.1002/wea.504>
- Willems, P., Vrac, M., 2011. Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology* 402, 193–

205. <https://doi.org/10.1016/j.jhydrol.2011.02.030>
- William, H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1996. *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing*. Cambridge University Press, London, UK.
- Willmott, C.J., 1982. Some comments on the evaluation of model performance. *Bulletin - American Meteorological Society* 63, 1309–1313. <https://doi.org/10.1175/1520-0477>
- Willmott, C.J., 1981. On the validation of models. *Physical Geography* 2, 184–194. <https://doi.org/10.1080/02723646.1981.10642213>
- Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, 79–82. <https://doi.org/10.3354/cr030079>
- Witten, I.H., Hall, M.A., Frank, E., 2005. Ensemble Learning, in: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Amsterdam, pp. 351–372.
- Wójcik, R., Pilarski, M., Miętus, M., 2014. Statistical downscaling of probability density function of daily precipitation on the Polish coast. *Meteorology Hydrology and Water Management* 2, 27–36. <https://doi.org/10.26491/mhwm/21590>
- Wright, D.B., Knutson, T.R., Smith, J.A., 2015. Regional climate model projections of rainfall from U.S. landfalling tropical cyclones. *Climate Dynamics* 45, 3365–3379. <https://doi.org/10.1007/s00382-015-2544-y>
- Wu, Y., Zhang, A., 2004. Feature selection for classifying high-dimensional numerical data, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 251–258. <https://doi.org/http://dx.doi.org/10.1109/CVPR.2004.1315171>
- Wu, Z., Chen, X., Lu, G., Xiao, H., He, H., Zhang, J., 2017. Regional response of runoff in CMIP5 multi-model climate projections of Jiangsu Province, China. *Stochastic Environmental Research and Risk Assessment* 31, 2627–2643. <https://doi.org/10.1007/s00477-016-1349-9>
- Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang, S., 2007. A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology* 8, 607–626. <https://doi.org/10.1175/JHM583.1>
- Xie, P., Zhuo, L., Yang, X., Huang, H., Gao, X., Wu, P., 2020. Spatial-temporal variations in blue and green water resources, water footprints and water scarcities in a large river basin: A case for the Yellow River basin. *Journal of Hydrology* 590, 125222. <https://doi.org/10.1016/j.jhydrol.2020.125222>
- Xu, H., Xu, C.Y., Sælthun, N.R., Zhou, B., Xu, Y., 2015. Evaluation of reanalysis and satellite-based precipitation datasets in driving hydrological models in a humid region of Southern China. *Stochastic Environmental Research and Risk Assessment* 29, 2003–2020. <https://doi.org/10.1007/s00477-014-1007-z>
- Xuan, W., Ma, C., Kang, L., Gu, H., Pan, S., Xu, Y.P., 2017. Evaluating historical simulations of CMIP5 GCMs for key climatic variables in Zhejiang Province, China. *Theoretical and Applied Climatology* 128, 207–222. <https://doi.org/10.1007/s00704-015-1704-7>
- Yang, X., Delsole, T., 2012. Systematic comparison of enso teleconnection patterns between models and observations. *Journal of Climate* 25, 425–446. <https://doi.org/10.1175/JCLI-D-11-00175.1>
- Yapo, O.P., Gupta, V.H., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* 204, 83–97.
- Yozgatligil, C., Yazici, C., 2016. Comparison of homogeneity tests for temperature using a simulation study. *International Journal of Climatology* 36, 62–81.



<https://doi.org/10.1002/joc.4329>

- Yu, L., Liu, H., 2003. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. *Proceedings, Twentieth International Conference on Machine Learning* 2, 856–863.
- Yuan, Z., Xu, J., Wang, Y., 2019. Historical and future changes of blue water and green water resources in the Yangtze River source region, China. *Theoretical and Applied Climatology* 138, 1035–1047. <https://doi.org/10.1007/s00704-019-02883-z>
- Yue, S., Pilon, P., Phinney, B., Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* 16, 1807–1829. <https://doi.org/10.1002/hyp.1095>
- Zandler, H., Haag, I., Samimi, C., 2019. Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions. *Scientific Reports* 9, 1–15. <https://doi.org/10.1038/s41598-019-51666-z>
- Zang, C., Liu, J., 2013. Trend analysis for the flows of green and blue water in the Heihe River basin, northwestern China. *Journal of Hydrology* 502, 27–36. <https://doi.org/10.1016/j.jhydrol.2013.08.022>
- Zeng, Z., Piao, S., Li, L.Z.X., Wang, T., Ciais, P., Lian, X., Yang, Y., Mao, J., Shi, X., Myneni, R.B., 2018. Impact of Earth greening on the terrestrial water cycle. *Journal of Climate* 31, 2633–2650. <https://doi.org/10.1175/JCLI-D-17-0236.1>
- Zhang, Q., Körnich, H., Holmgren, K., 2013. How well do reanalyses represent the southern African precipitation? *Climate Dynamics* 40, 951–962. <https://doi.org/10.1007/s00382-012-1423-z>
- Zhang, X., Jiang, C., Huang, J., Ni, Z., Sun, J., Li, Z., Wen, T., 2022. Spatiotemporal Evaluation of Blue and Green Water in Xinjiang River Basin Based on SWAT Model. *Water (Switzerland)* 14. <https://doi.org/10.3390/w14152429>
- Zhang, Y., Chiew, F.H.S., 2009. Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research* 45, W07412. <https://doi.org/10.1029/2008WR007504>
- Zhang, Y., Peña-Arancibia, J.L., McVicar, T.R., Chiew, F.H.S., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y.Y., Miralles, D.G., Pan, M., 2016. Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific Reports* 6, 1–12. <https://doi.org/10.1038/srep19124>
- Zhou, F., Zhang, W., Su, W., Peng, H., Zhou, S., 2021. Spatial differentiation and driving mechanism of rural water security in typical “engineering water depletion” of karst mountainous area—A lesson of Guizhou, China. *Science of the Total Environment* 793, 148387. <https://doi.org/10.1016/j.scitotenv.2021.148387>
- Zhu, K., Qiu, X., Luo, Y., Dai, M., Lu, X., Zang, C., Zhang, W., Gan, X., Zhula, W., 2022. Spatial and temporal dynamics of water resources in typical ecosystems of the Dongjiang River Basin, China. *Journal of Hydrology* 614, 128617. <https://doi.org/10.1016/j.jhydrol.2022.128617>
- Zhu, K., Xie, Z., Zhao, Y., Lu, F., Song, Xinyi, Li, L., Song, Xiaomeng, 2018. The assessment of greenwater based on the SWAT model: A case study in the Hai River Basin, China. *Water (Switzerland)* 10. <https://doi.org/10.3390/w10060798>
- Zhu, W., Yan, J., Jia, S., 2017. Monitoring recent fluctuations of the southern pool of lake chad using multiple remote sensing data: Implications for water balance analysis. *Remote Sensing* 9, 1032. <https://doi.org/10.3390/rs9101032>
- Zhuang, X.W., Li, Y.P., Huang, G.H., Liu, J., 2016. Assessment of climate change impacts on watershed in cold-arid region: an integrated multi-GCM-based stochastic weather generator and stepwise cluster analysis method. *Climate Dynamics* 47, 191–209. <https://doi.org/10.1007/s00382-015-2831-7>

## APPENDICES

### Appendix 1: Multivariate Imputation and Chained Equation (MICE) missing data script.

```

1 # Libraries
2 library(mice)
3 library(VIM)
4
5
6 setwd("H:/LAKE CHAD OBSERVED DATASET/TEMP")
7 # Data
8 data <- read.csv(file.choose(), header = T)
9 str(data)
10 summary(data)
11
12 # Missing data
13 p <- function(x) {sum(is.na(x))/length(x)*100}
14 apply(data, 2, p)
15 md.pattern(data)
16 md.pairs(data)
17 marginplot(data[,c("Jan", "Feb")])
18
19 # Impute
20 impute <- mice(data[,2:18], m=3, seed = 123)
21 print(impute)
22 impute$imp$Jan
23 data[31,]
24 summary(data$Jan)
25 data[31,]
26 data[32,]
27
28 # Complete data
29 complete(impute, 3)
30
31 # Distribution of observed/imputed values
32 stripplot(impute, pch = 20, cex = 1.2)
33
34 # Convert matrix to dataframe
35 newdata <- as.data.frame(complete(impute, 3))
36
37 # Write csv file
38 write.csv(newdata,"ZINDER 3.csv")
39

```

### Appendix 2: Homogeneity test results of station data

#### Appendix 2a: absolute homogeneity test results of precipitation station data

Station	Pettitt's Test		SNHT Test		Von Neumann's Test	
	Critical	T-Stat	Critical	T-Stat	Critical	T-Stat
Abeche	0.05	0.032	0.05	0.036	0.05	0.016
Banda	0.05	0.044	0.05	0.047	0.05	0.046
Bongor	0.05	0.045	0.05	0.013	0.05	0.028
Bossangoa	0.05	0.034	0.05	0.035	0.05	0.015
Doba	0.05	0.064	0.05	0.045	0.05	0.035
Maiduguri	0.05	0.050	0.05	0.011	0.05	0.034
Moundou	0.05	0.013	0.05	0.019	0.05	0.011
N'Djamena	0.05	0.016	0.05	0.002	0.05	0.035
Nguigni	0.05	0.036	0.05	0.034	0.05	0.067
Potiskum	0.05	0.046	0.05	0.065	0.05	0.035
Sahr	0.05	0.029	0.05	0.029	0.05	0.039
Samry-I	0.05	0.039	0.05	0.036	0.05	0.037
Sategui D	0.05	0.027	0.05	0.028	0.05	0.029
Tsanaga	0.05	0.029	0.05	0.029	0.05	0.027
Zinder	0.05	0.026	0.05	0.028	0.05	0.029

## Appendix 2b: absolute homogeneity test results of temperature station data

Station	Pettitt's Test		SNHT Test		Von Neumann's Test	
	Critical	T-Stat	Critical	T-Stat	Critical	T-Stat
Bilma	0.05	0.001	0.05	0.009	0.05	0.008
Bossangoa	0.05	0.01	0.05	0.006	0.05	0.02
Bouar	0.05	0.02	0.05	0.02	0.05	0.019
Geneina	0.05	< 0.0001	0.05	< 0.0001	0.05	< 0.0001
Maiduguri	0.05	0.021	0.05	0.025	0.05	0.027
Maina S.	0.05	0.017	0.05	0.004	0.05	0.014
Moundou	0.05	0.025	0.05	0.027	0.05	0.034
N'Djamena	0.05	0.029	0.05	0.036	0.05	0.008
Ngaoundere	0.05	0.018	0.05	0.026	0.05	0.021
Nguigni	0.05	0.001	0.05	0.001	0.05	0.006
Sahr	0.05	0.036	0.05	0.023	0.05	0.022
Zinder	0.05	0.020	0.05	0.014	0.05	0.036

## Appendix 3: Climate data extraction R-script

### Appendix 3a: precipitation data extraction R-script

```

1 # Load libraries
2 library(raster)
3 library(tidyverse)
4 library(data.table)
5 library(ncdf4)
6 library(sp)
7
8 # Set working Directory
9 setwd("H:/GRIDDED DATASETS/UDEL")
10
11 # Read coordinates file
12 coords <- read_csv("Prec Chad-Basin.txt")
13
14 # Read netcdf file as rasterbrick
15 file <- brick("precip.mon.total.v501.nc")
16
17 # Get coordinates from coordinates file
18 lon <- coords$LONG
19 lat <- coords$LAT
20
21 # Extract values at coordinates
22 pr <- raster::extract(file, cbind(lon,lat))
23
24 # Convert matrix to dataframe
25 pr <- as.data.frame(pr)
26 pr <- pr*10
27 pr <- round(pr, digits = 2)
28 pr <- data.frame(coords$NAME,pr)
29
30 # Transpose dataframe
31 maxtemp <- data.table::transpose(pr)
32
33 # Assign column and row names
34 colnames(maxtemp) <- rownames(pr)
35 rownames(maxtemp) <- colnames(pr)
36
37 # write csv file
38 write_csv(maxtemp,"precip.mon.total.v501.csv")
39

```

## Appendix 3b: Temperature data extraction R-script

The screenshot displays the RStudio interface. The main editor shows an R script with the following code:

```
1 # Load libraries
2 library(raster)
3 library(tidyverse)
4 library(data.table)
5 library(sp)
6 library(ncdf4)
7
8 # Set working Directory
9 setwd("H:/GRIDDED DATASETS/PGF Gridded Data/Tmin")
10
11 # Read coordinates file
12 coords <- read_csv("1.0 Stations and Coordinates.txt")
13
14 # Read netcdf file as rasterbrick
15 file <- brick("tmin_daily_2012-2012.nc")
16
17 # Get coordinates from coordinates file
18 lon <- coords$LONG
19 lat <- coords$LAT
20
21 # Extract values at coordinates
22 tasmin <- raster::extract(file, cbind(lon,lat))
23
24 # Convert matrix to dataframe
25 tasmin <- as.data.frame(tasmin)
26 tasmin <- tasmin - 273
27 tasmin <- data.frame(coords$NAME,tasmin)
28
29 # Transpose dataframe
30 mintemp <- data.table::transpose(tasmin)
31
32 # Assign column and row names
33 colnames(mintemp) <- rownames(tasmin)
34 rownames(mintemp) <- colnames(tasmin)
35
36 # write csv file
37 write_csv(mintemp,"tmin_daily_2012-2012.csv")
38
```

The Environment pane on the right shows the following data objects:

Object	Details
coords	210 obs. of 5 variables
file	Formal class RasterBrick
mintemp	367 obs. of 210 variables
tasmin	210 obs. of 367 variables

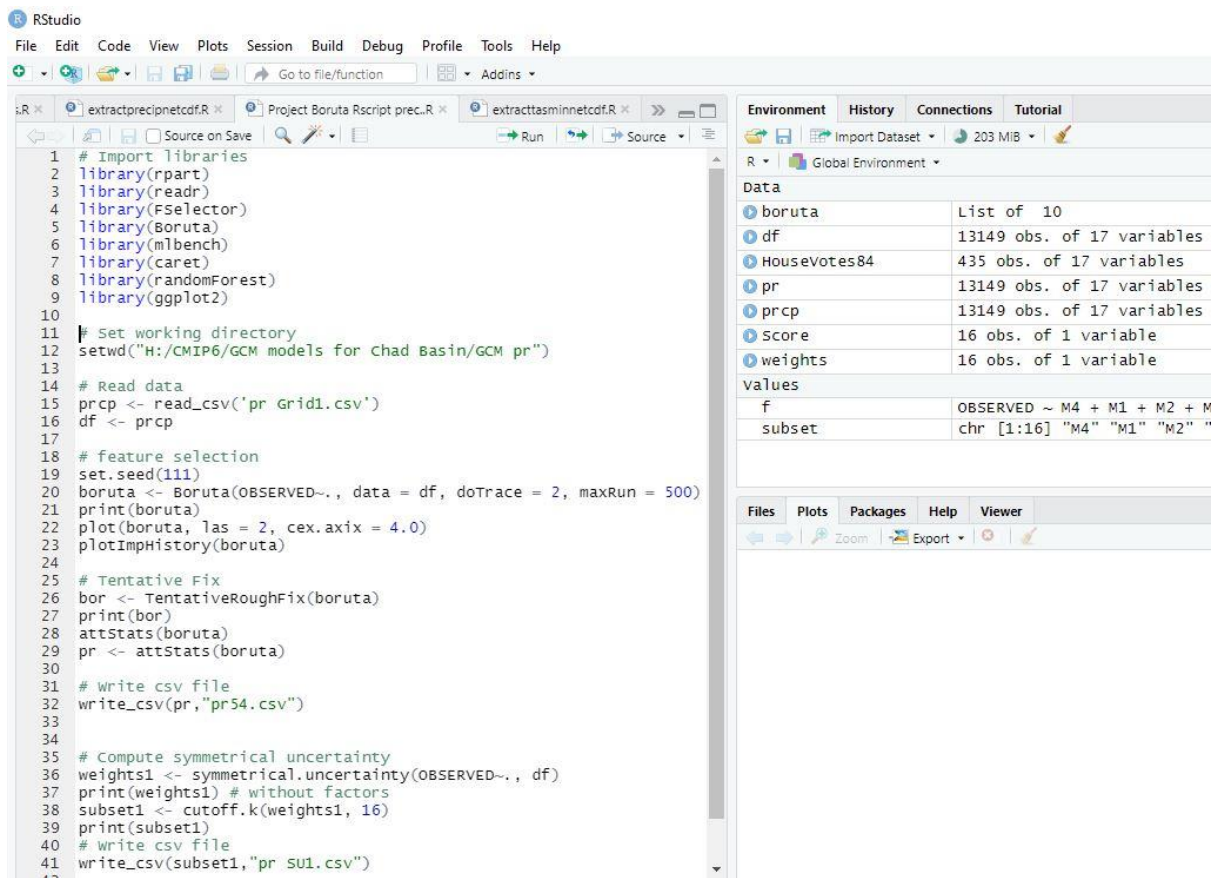
The Environment pane also shows the following values:

Variable	Class	Value
lat	num [1:210]	5.81 6.81 6.81
lon	num [1:210]	18.4 14.4 15.4

## Appendix 4: Mann-Kendall trend analysis R-script

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
metric Homogeneity Test Prec... x R data plots.R x Mann-Kendall Trend Test.R x Statistics hydroGOF.R x
Source on Save Run
1 library(modifiedmk)
2
3 setwd("C:/Users/drbb19224/Desktop/TREND TEST/PREC")
4 # Data
5 data <- read.csv(file.choose(), header = T)
6
7 # Autocorrelation check for choice of trend test
8 acf(data$Annual, lag.max = 1)
9
10 acf(data$Annual, lag.max = 1)$acf
11
12 acf(data$Annual, lag.max = 1, plot = T)
13
14 lag.plot(acf(data$Annual), lags = 20, do.lines = F)
15
16 # Trend Analysis for time series data (with significant autocorrelation)
17 bcpw(data$Annual)
18
19 # Trend Analysis for time series data (without significant autocorrelation)
20 mkttest(data$Annual)
21
22
23 # Autocorrelation check for choice of trend test
24 acf(data$Premonsoon, lag.max = 1)
25
26 acf(data$Premonsoon, lag.max = 1)$acf
27
28 acf(data$Premonsoon, lag.max = 1, plot = T)
29
30 lag.plot(acf(data$Premonsoon), lags = 20, do.lines = F)
31
32 # Trend Analysis for time series data (with significant autocorrelation)
33 bcpw(data$Premonsoon)
34
35 # Trend Analysis for time series data (without significant autocorrelation)
36 mkttest(data$Premonsoon)
37
38 # Autocorrelation check for choice of trend test
39 acf(data$Monsoon, lag.max = 1)
40
41 acf(data$Monsoon, lag.max = 1)$acf
42
43 acf(data$Monsoon, lag.max = 1, plot = T)
44
45 lag.plot(acf(data$Monsoon), lags = 20, do.lines = F)
46
47 # Trend Analysis for time series data (with significant autocorrelation)
48 bcpw(data$Monsoon)
49
50 # Trend Analysis for time series data (without significant autocorrelation)
51 mkttest(data$Monsoon)
52
```

## Appendix 5: Random forest algorithms for data pruning R-script



The screenshot displays the RStudio interface. The main editor shows an R script with the following code:

```
1 # Import libraries
2 library(rpart)
3 library(readr)
4 library(FSelector)
5 library(Boruta)
6 library(mlbench)
7 library(caret)
8 library(randomForest)
9 library(ggplot2)
10
11 # Set working directory
12 setwd("H:/CMIP6/GCM models for Chad Basin/GCM pr")
13
14 # Read data
15 prcp <- read_csv('pr Grid1.csv')
16 df <- prcp
17
18 # feature selection
19 set.seed(111)
20 boruta <- Boruta(OBSERVED~., data = df, doTrace = 2, maxRun = 500)
21 print(boruta)
22 plot(boruta, las = 2, cex.axis = 4.0)
23 plotImpHistory(boruta)
24
25 # Tentative Fix
26 bor <- TentativeRoughFix(boruta)
27 print(bor)
28 attStats(boruta)
29 pr <- attStats(boruta)
30
31 # write csv file
32 write_csv(pr, "pr54.csv")
33
34
35 # Compute symmetrical uncertainty
36 weights1 <- symmetrical.uncertainty(OBSERVED~., df)
37 print(weights1) # without factors
38 subset1 <- cutoff.k(weights1, 16)
39 print(subset1)
40 # write csv file
41 write_csv(subset1, "pr su1.csv")
```

The Environment pane on the right shows the following data objects:

Object	Description
boruta	List of 10
df	13149 obs. of 17 variables
Housevotes84	435 obs. of 17 variables
pr	13149 obs. of 17 variables
prcp	13149 obs. of 17 variables
Score	16 obs. of 1 variable
weights	16 obs. of 1 variable

Below the data objects, the 'values' section shows:

Variable	Value
f	OBSERVED ~ M4 + M1 + M2 + M
subset	chr [1:16] "M4" "M1" "M2" "