

# Assessment of Global Reanalysis Precipitation Datasets for Hydrological Applications in Kenya



Maureen Wanzala, Hannah Cloke, Andrea Ficchi, Liz Stephens, Heou Badjana, David Lavers

**Email:** [maureen.wanzala@welthungerhilfe.de](mailto:maureen.wanzala@welthungerhilfe.de)

Annual Seminar 2023

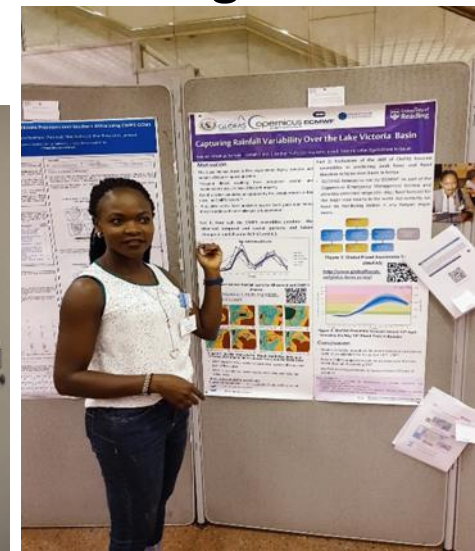
ECMWF – 07/09/2023

 @\_Maureen\_AW



# Introduction

- Meteorologist, hydro-meteorologist and climate scientist.
- Provide research, monitoring, and analysis of fast (floods and tropical storms) and slow (droughts) onset weather and climate hazards to support to country offices and RBs to enhance preparedness and Anticipatory Humanitarian Actions ahead of and during hazards.
- Science communications, working with broader scientific community and humanitarian agencies



**Email:** [maureen.wanzala@welthungerhilfe@de](mailto:maureen.wanzala@welthungerhilfe.de)

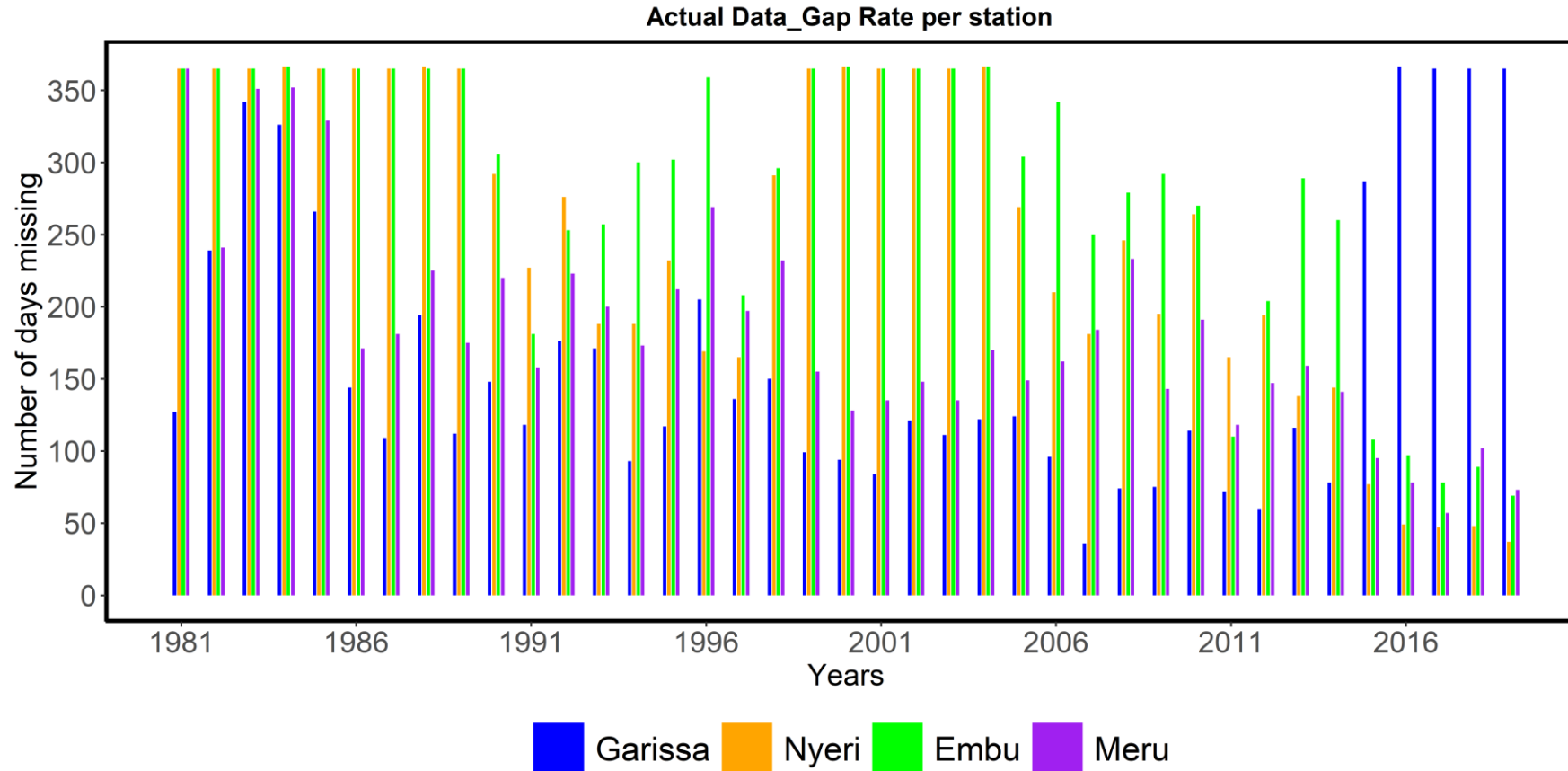


@\_Maureen\_AW

# Background

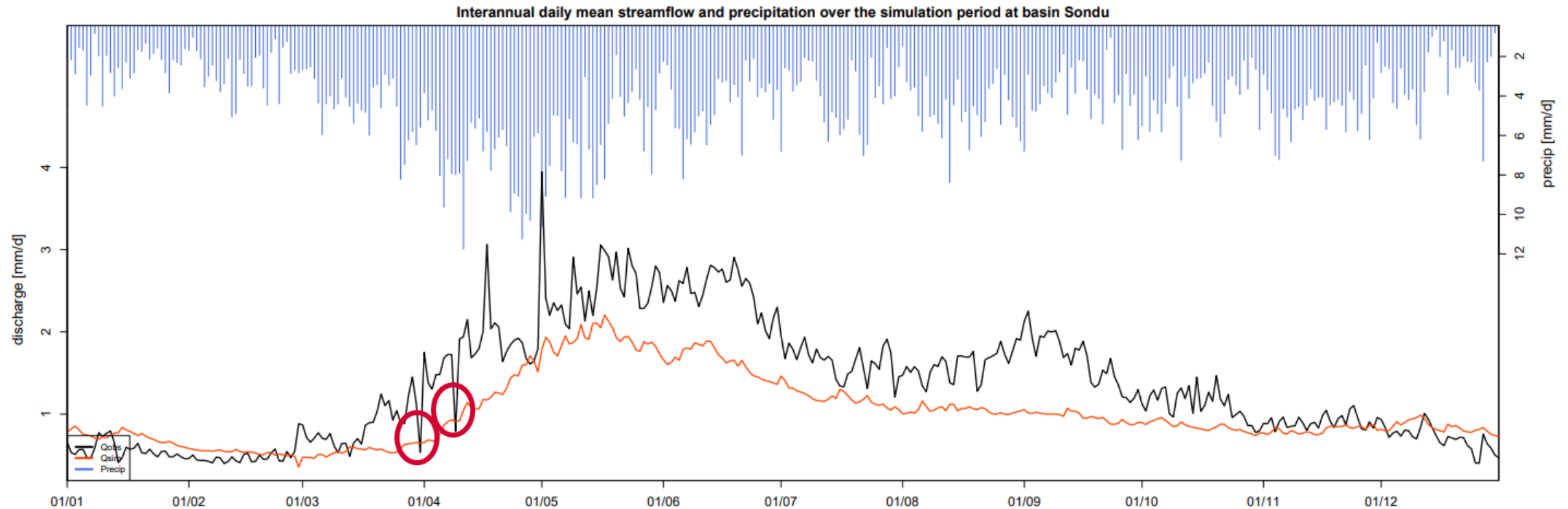
Reanalysis vary in **quality** and are used as substitutes to observations in data scarce regions:-

- Ground validation is key but very challenging, in sparse rainfall gauge networks regions



# Background

- Validation process uses ground proxies which fail to capture the real-time evolution and magnitude of events (e.g., floods and droughts)



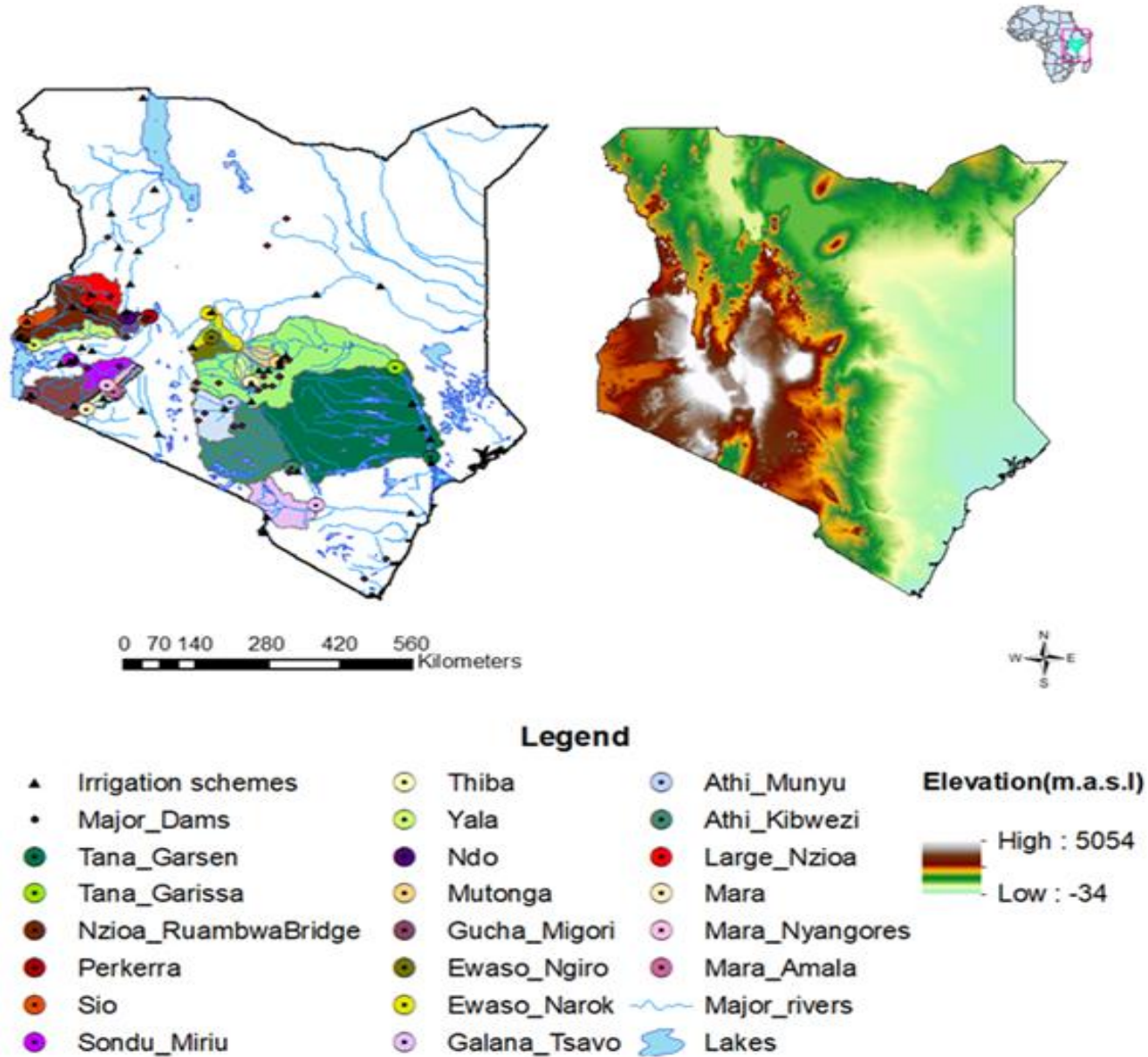
# Background

- Thus, the need to assess the impact on the river discharge on a catchment-by-catchment basis due to the stark differences between the river basins
- **Objective:** To assess different **reanalysis datasets** and identify which are the most suitable in **simulating streamflow** when coupling both the **performance statistics** and **sensitivity analysis**

## Research Questions

- How well do the precipitation datasets compare in terms of temporal dynamics at the basin scale? Which product is the most accurate compared to observations?
- How well do precipitation datasets compare in terms of spatial patterns? Which product shows consistency in spatial heterogeneity compared to observations?
- How does the general hydrological model performance vary with different datasets?
- How does the sensitivity of a rainfall runoff model (GR4J) vary with alternative rainfall forcing?

# Study Area



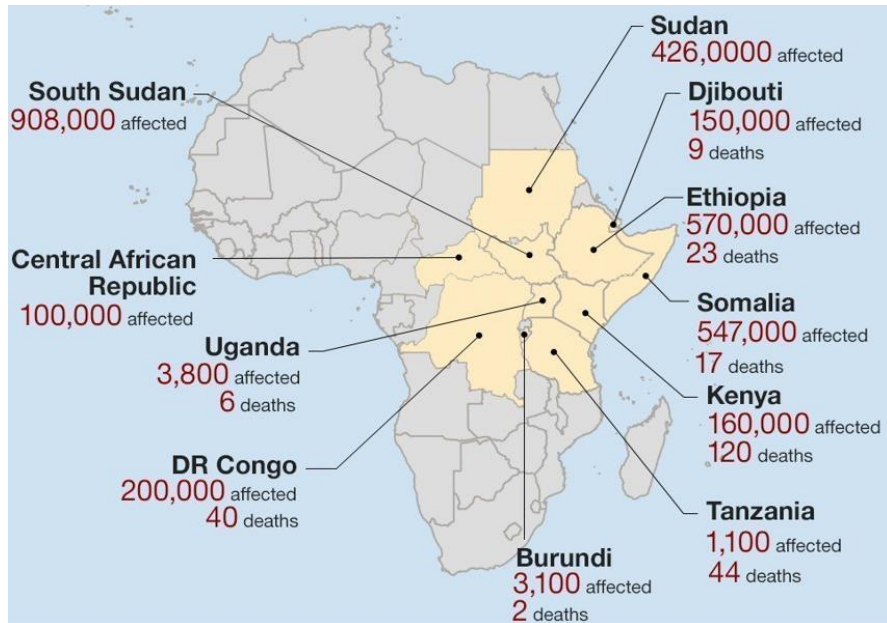
- Kenya lies astride of the equator at longitude 34°E–42°E and Latitude 5°S–5°N and along the quasi-meridional western edge of the Indian Ocean at the eastern side of Africa

- The study is undertaken in **19 Kenyan catchments** with good and satisfactory streamflow data records

*Study catchments, with the location of the outlet river gages used in this study and the main irrigation schemes and major dams across Kenya (left) and topography of Kenya (right)*

# Why Kenya in Particular

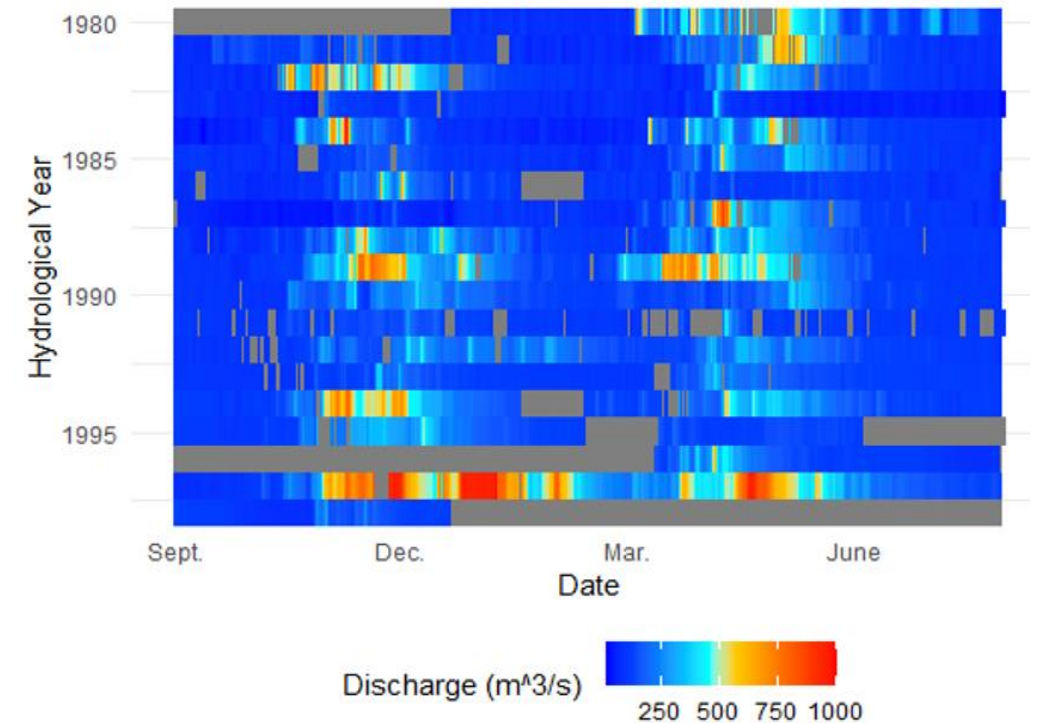
- Flood occurs annually during the rainy seasons (March-May “**long rains**” and Oct-Dec “**short rains**”)
- Flooding in EA is remarkable in terms of **duration, scale, and severity**. By late **May**, the **2021** floods had killed more people than COVID-19.



<https://medium.com/@icpac/recurring-floods-in-eastern-africa-amidst-projections-of-frequent-and-extreme-climatic-events-for-30d20d0d6f76>

# Flood Risk in Kenya

- Kenya is among the highest climate risk countries in the world (Global Climate Risk Index, 2019)
- The common climate and weather extremes are particularly droughts and floods (Eckstein et al., 2019)
- Major flood events occur roughly every two years on average (Emergency Events Database (EM-DAT))
- The typical population of people affected per event is approximated to be 70,000 (Parry et al., 2012)
- Between 1964 - 2019, Kenya recorded 18 major flood events, with 1961, 1997–1998, 2002, 2003, 2006, 2010, 2012, and 2018 recording particularly high impact flood events, and declared national disasters



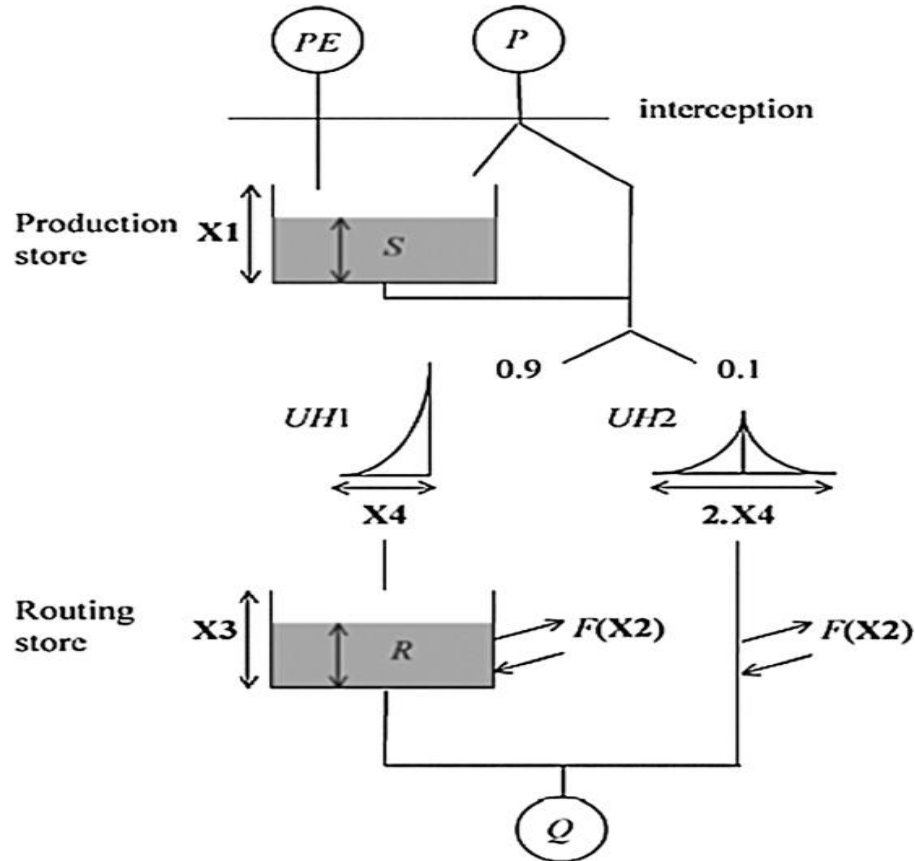
- High discharge occur during the wet seasons
- There is a large annual variation in river discharges



## Data and Methods

- Reanalysis datasets (ERA 5, ERA-Interim, NCEP/NCAR-CFSR, and JRA-55), and CHIRPS, as well as observed river discharge data for the period 1981-2016
- Using a lumped bucket-style hydrological model (**GR4J**, Perrin et al. 2003), we assess the **model performance** via the **KGE criterion and parameter uncertainty** via Sobol's Sensitivity Analysis
- Performance metrics: Correlation Coefficient (CC), Root Mean Square Error (RMSE),BIAS, Kling Gupta Efficiency (KGE), Sobol Total Sensitivity Index(TSI), Model Suitability Index(MSI) were computed from the calibration and validation using GR4J model

# GR4J Hydrological Model



## GR4J Parameters

Parameter	Description	Typical Value
X1	Capacity of the production store (mm)	(0, 1000)
X2	Groundwater exchange coefficient(mm)	(-5, 5)
X3	Capacity of the nonlinear routing store (mm)	(0, 300)
X4	Unit hydrograph time base (day)	(0.5, 5)

- **GR4J**: daily continuous lumped, conceptual model based on neutralisation for interception, a soil moisture accounting (SMA) store (production), an excess rainfall and routing store & unit hydrographs (Perrin et al., 2003).

- 4 parameters to calibrate: maximum capacity of the production store ( $X1$ , mm), groundwater exchange coefficient ( $X2$ , mm/day), maximum capacity of the non-linear routing store ( $X3$ , mm); and time base of the unit hydrograph ( $X4$ , days)

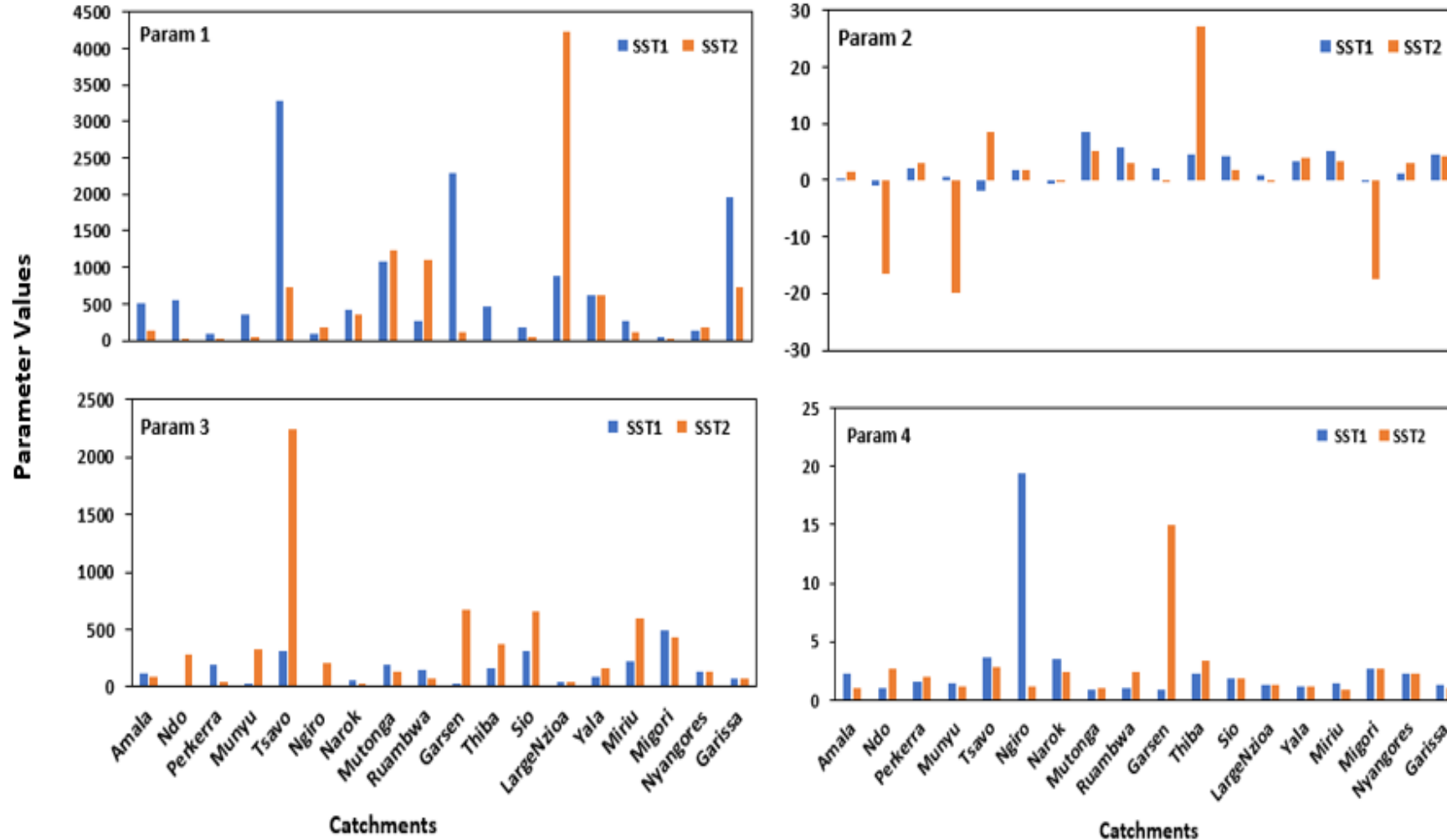
## GR4J Calibration Strategy

- GR4J conceptual rainfall-runoff model was calibrated to determine appropriate parameter values
- Parameters were calibrated to the Kling-Gupta Efficiency (KGE)((Gupta et al, 2009)) objective function
- 36 years of streamflow data for the catchments were used
- Split-sample validation testing (Klemes, 1986) was used to test model skill beyond the calibration period.

### Split Sample Test (SST)

<b>Y1 to Y4 (1981-1983)</b> <b>Model spin up</b>	<b>Y5 to Y15 (1984 – 1999)</b> <b>SST1</b>	<b>Y16 to Y36 (2001 – 2016)</b> <b>SST2</b>
---	---	--

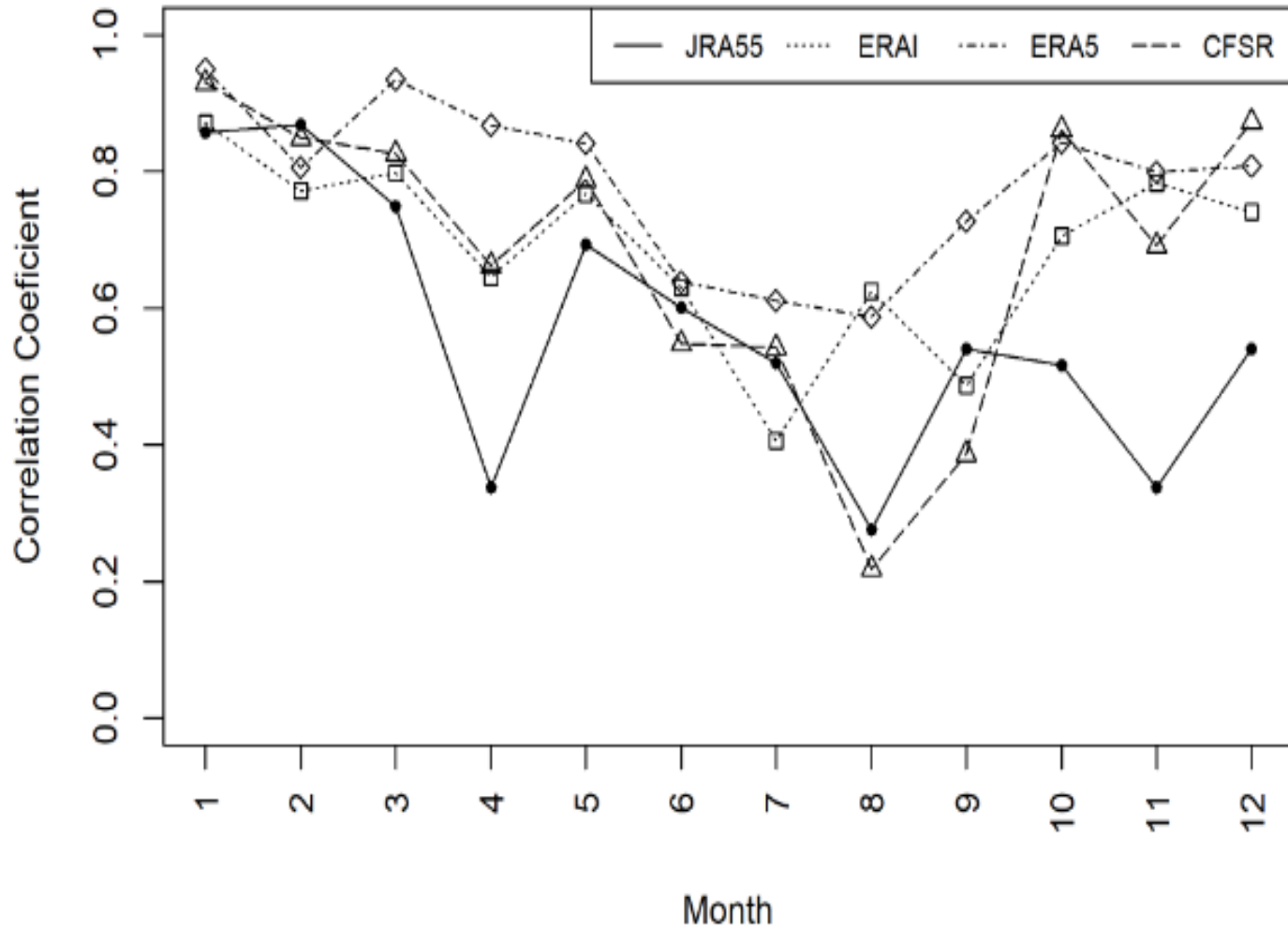
# GR4J Calibrated Parameters



Parameter	Description	Typical Value
X1	Capacity of the production store (mm)	(0, 1000)
X2	Groundwater exchange coefficient (mm/day)	(-5, 5)
X3	Capacity of the nonlinear routing store (mm)	(0, 300)
X4	Unit hydrograph time base (day)	(0.5, 5)

# Performance Statistics

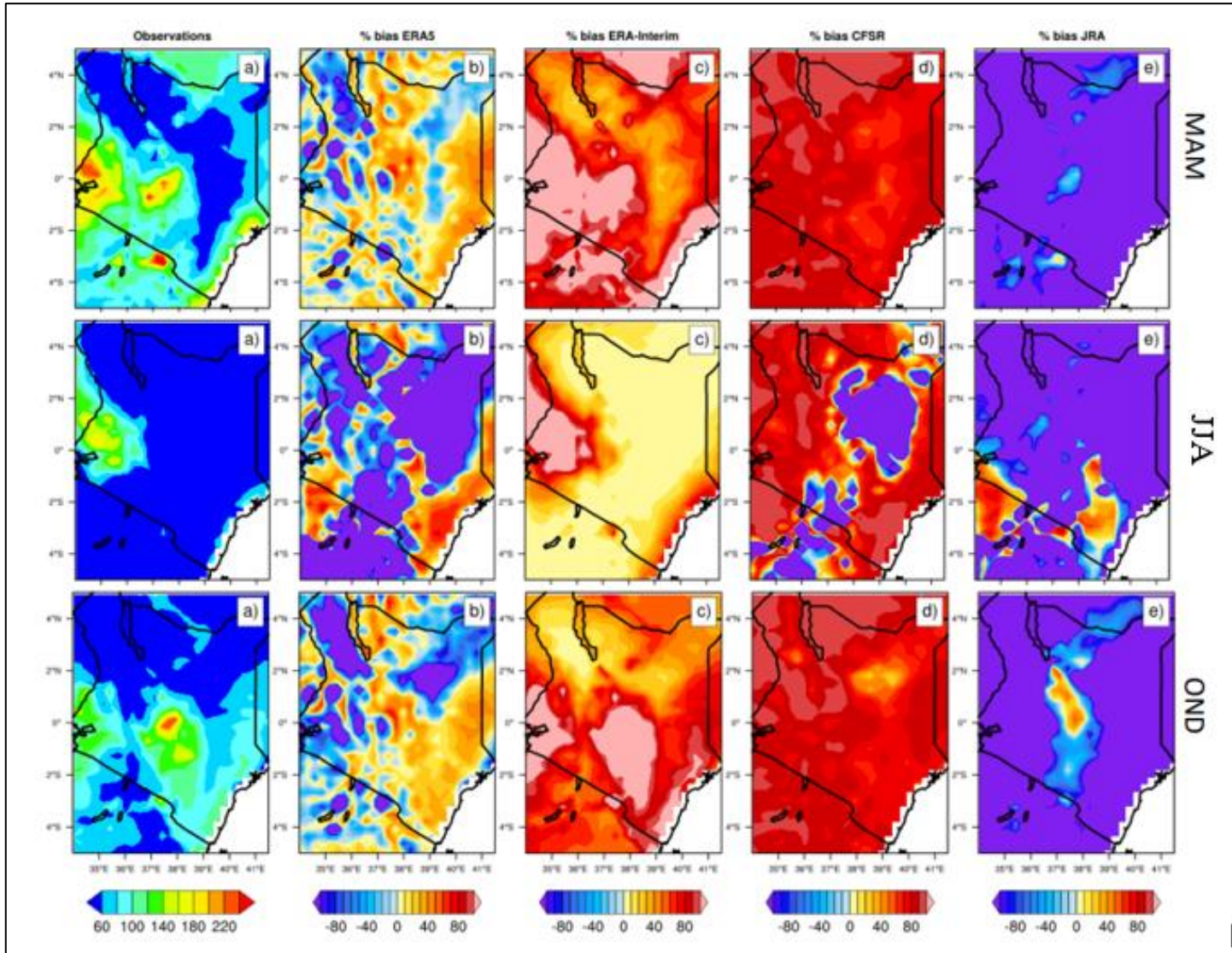
## Correlation Coefficient



- ERA5 has higher correlation coefficients (0.88) than JRA55, ERAI & CFSR in comparison to observations on monthly timesteps
- Same performance is observed on the seasonal scale (not shown)

# Performance Statistics

## Wet extreme rainy days



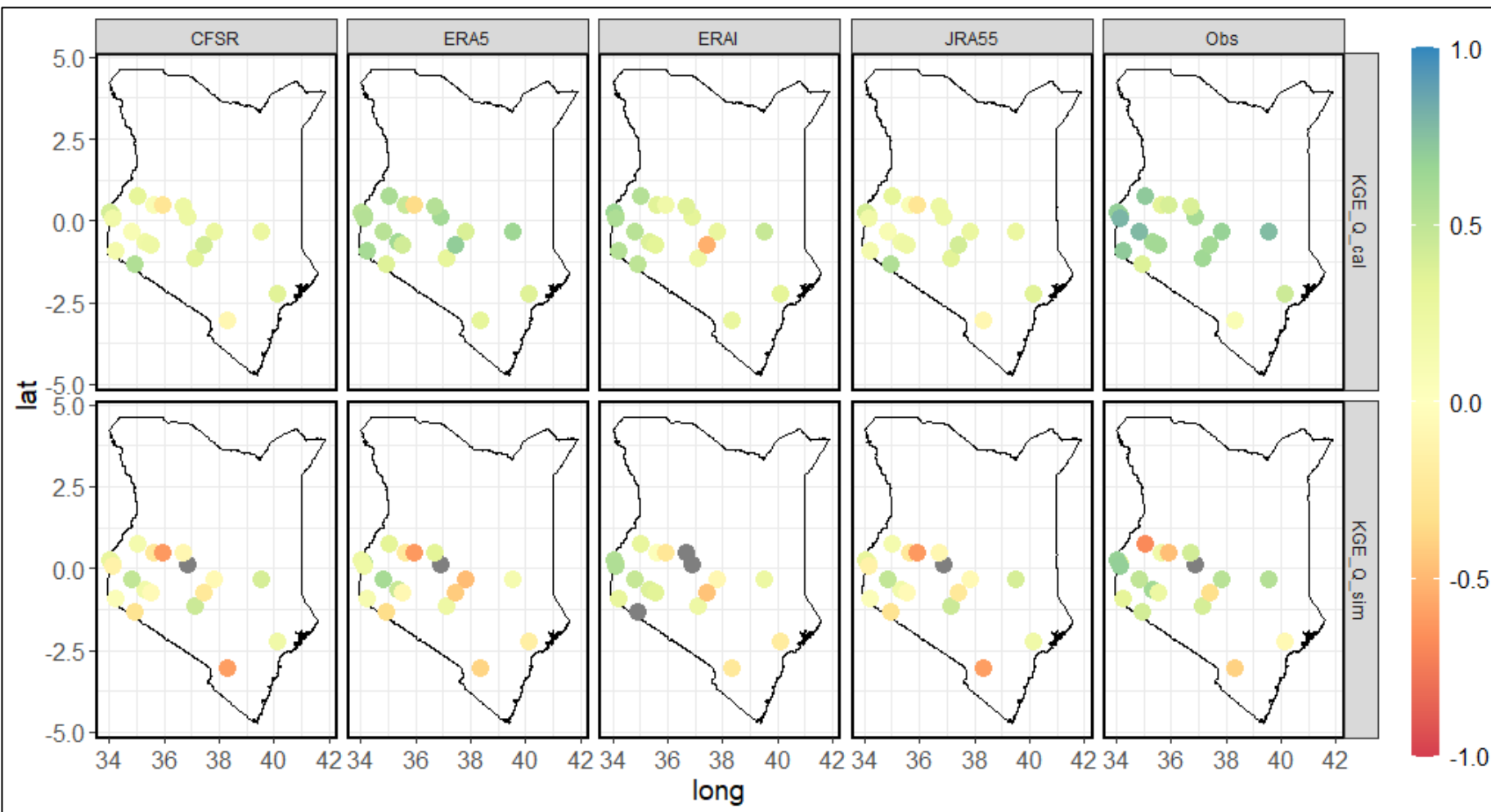
Seasonal observed precipitation (mm) and mean bias (%) of the extreme rainy days at 95th percentile in the four reanalysis products for 1981–2016 period

- The observed extreme precipitation varies between 60mm and 240mm across western, central highlands and coastal catchments for the rainy seasons (MAM and OND), whereas in the dry season (JJA) varied between 100mm and 160mm across the western catchments only, and less than 60mm in other regions.
- CFSR and ERAI show a positive bias for extreme precipitation across most parts of the country in all the three seasons, whereas JRA55 has an enhanced negative bias in most parts of the country.
- ERA5 has a positive bias in MAM and OND in most parts of the country with some patches of negative bias in the western and central highlands catchments.
- ERA5 outperforms other reanalysis products as it captures the wet extremes over the regions in which observations show enhanced precipitation in the respective seasons.

# Model Calibration Results

## Kling Gupta Efficiency (KGE)

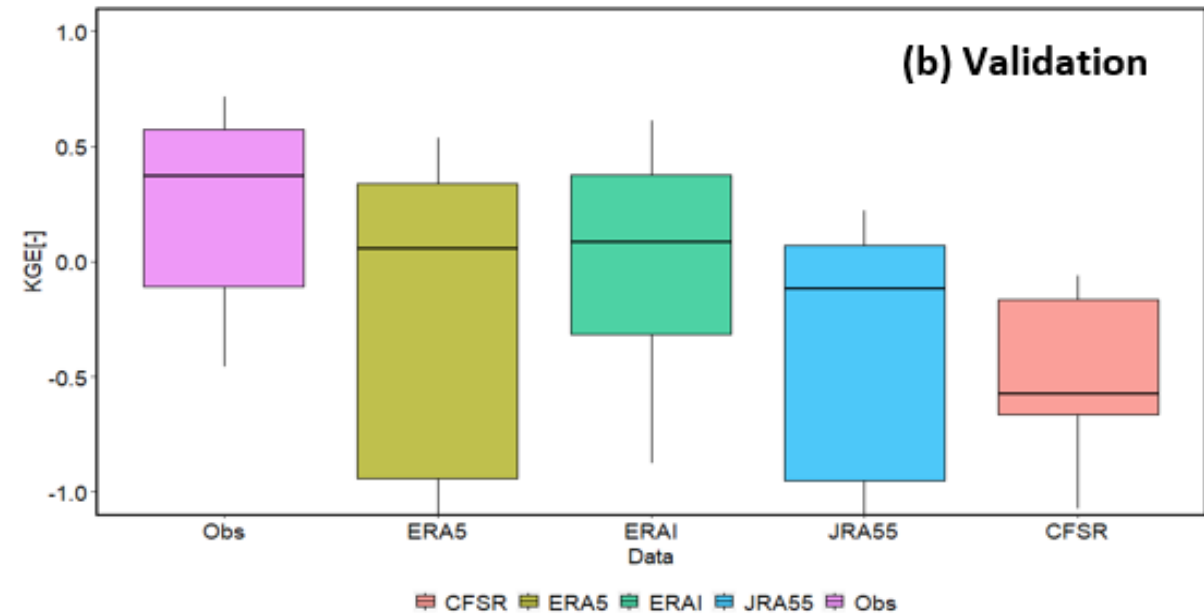
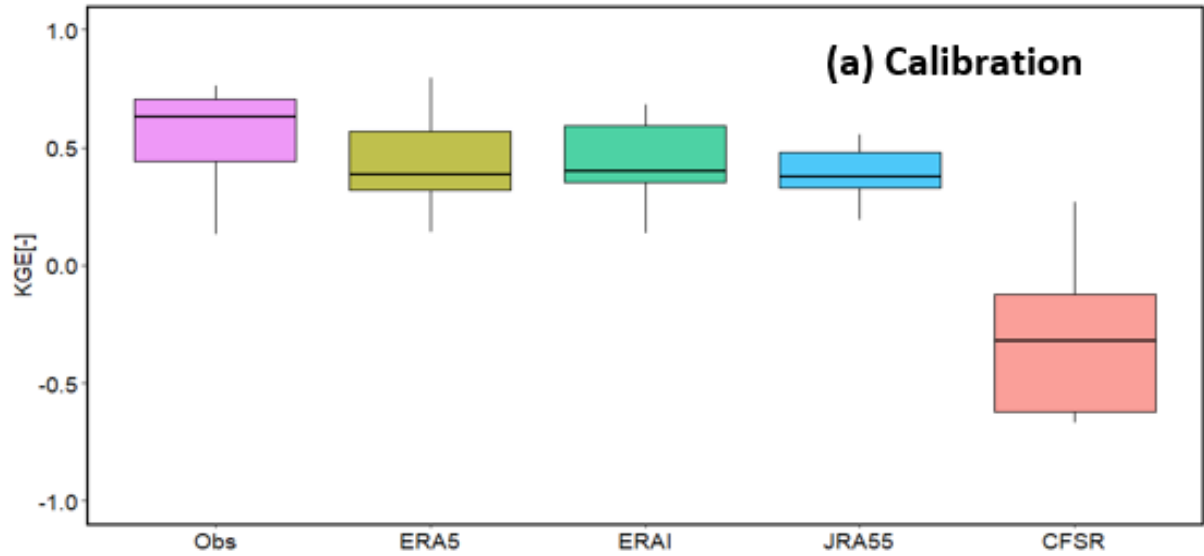
The KGE in calibration (top panel) and validation (bottom panel) scores obtained using different datasets



Source:- Wanzala et al., 2022

- Wetland catchments in the western and highlands of Kenya obtained relatively better calibration scores than those in the semi-arid regions.
- For each of the catchments, ERA5 showed better calibrated KGE scores compared to observations, while CFSR and JRA55 obtained poorer KGE scores.
- Caution- Strong influence by human activities and uncertainty in the input data.

# Model Calibration and Validation



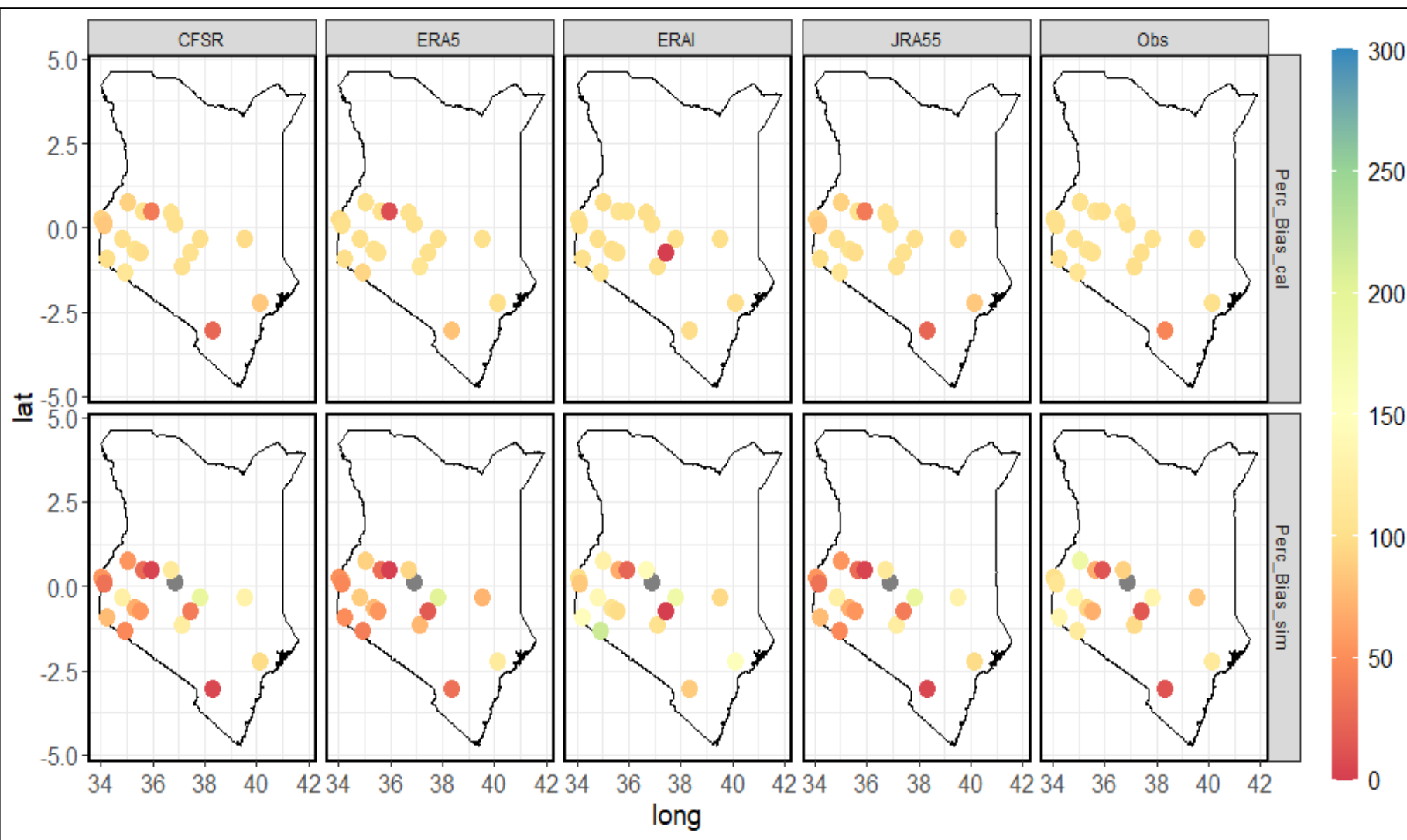
## Kling Gupta Efficiency (KGE)

- ERA5, ERAI, & JRA55 obtained better KGE scores whereas CFSR had lower scores in both calibration and validation
- The 3 best performing datasets give similar scores to observations in calibration, but more variable and less robust results in validation (especially ERA5 and JRA55)



# Percentage Bias

- The bias in all the four reanalysis is higher in calibration, whereas in validation most catchments exhibit lower biases, except for Perkerra.

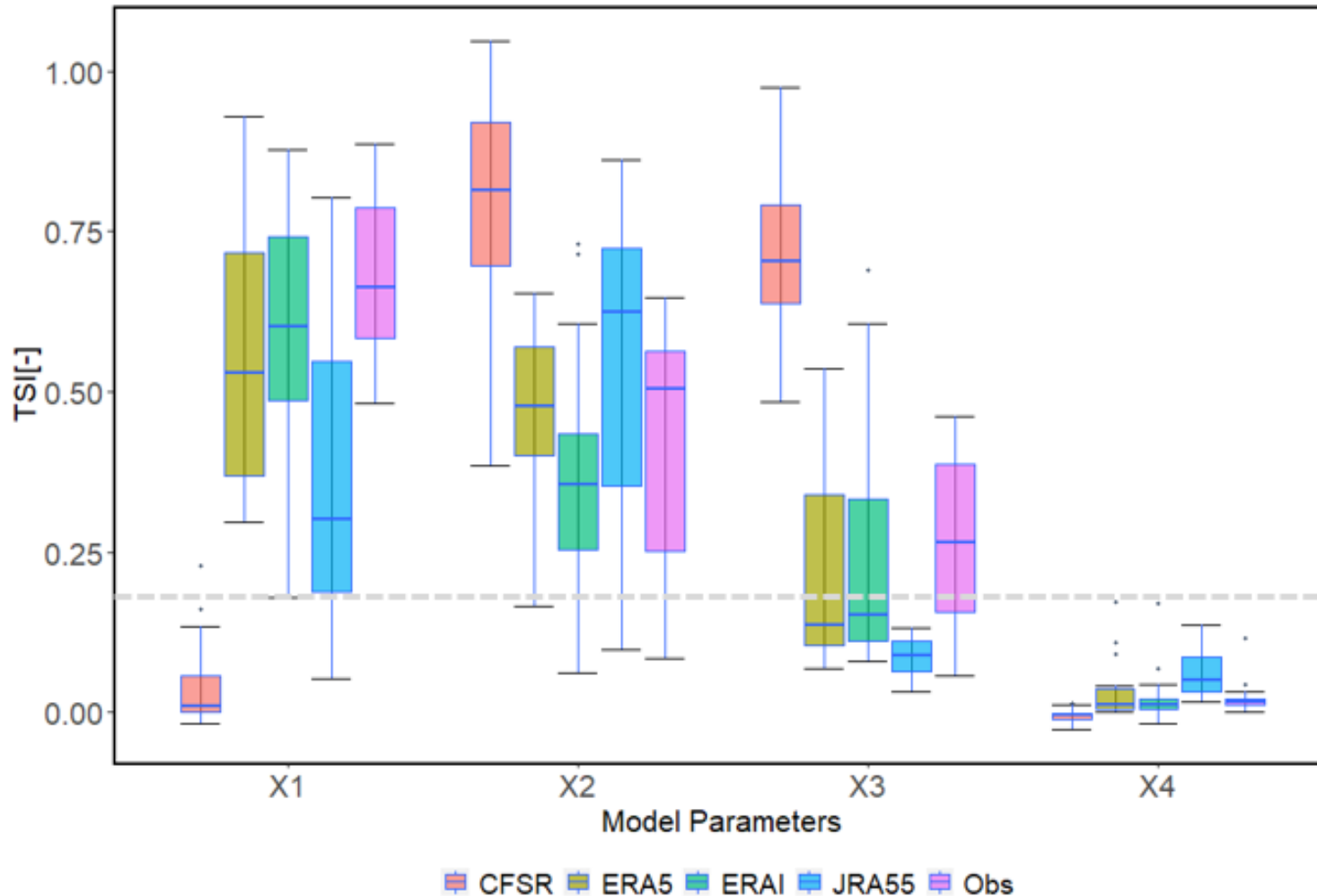


Source:- Wanzala et al., 2022

The percentage KGE - Bias in calibration (top panel) and validation (bottom panel) scores obtained using different datasets

# Sensitivity Analysis

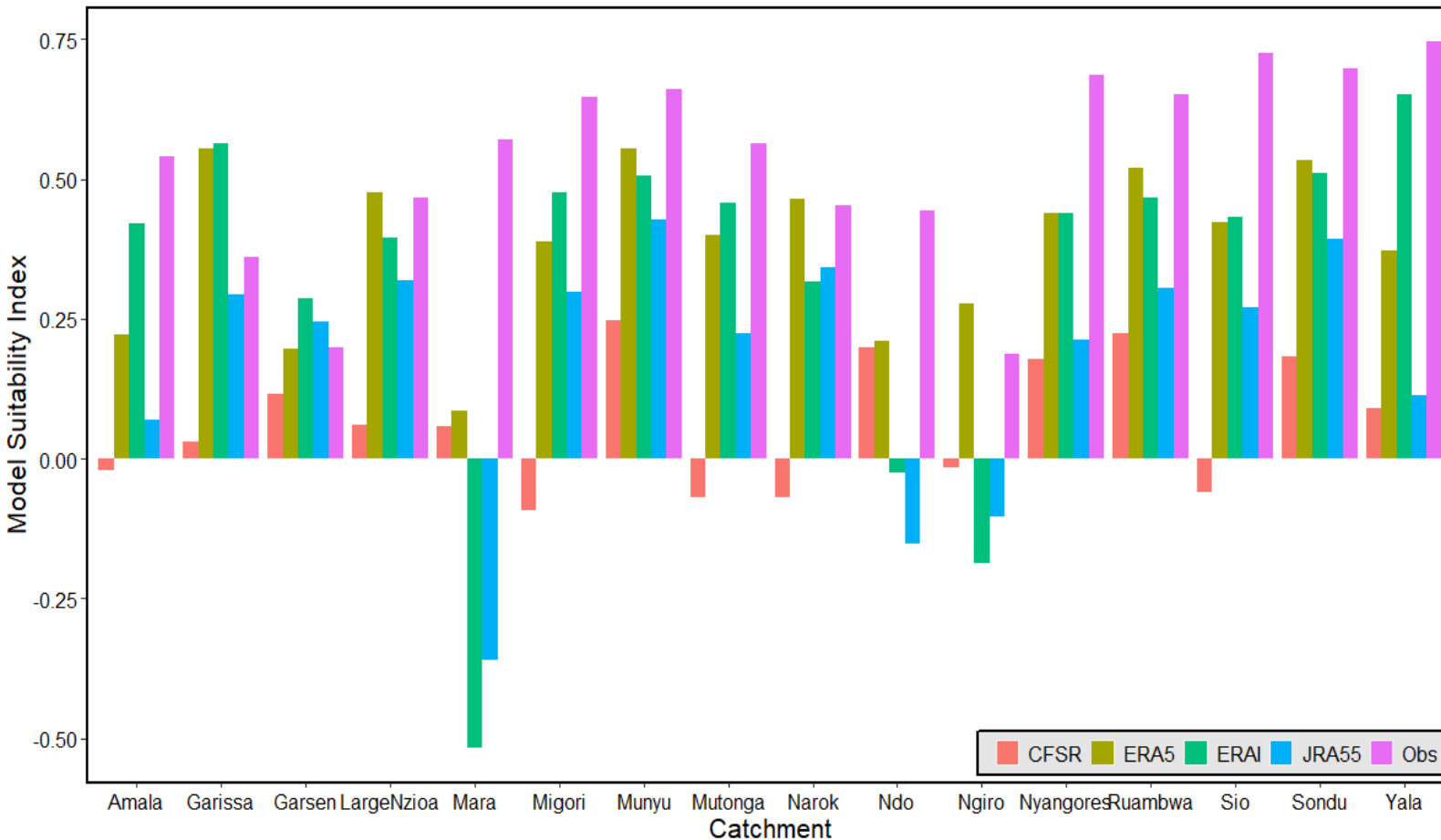
## Total Sensitivity Index (TSI)



- In all datasets, the production store capacity (X1) has the highest total effects (Total Sensitivity Index) whereas the time base of the unit hydrograph (X4) is the least influential parameter
- ERA5 show similar sensitivity of each of the model parameters compared to observations, whereas CFSR has highest **uncertainty**

# Model Suitability Index

## Model Suitability Index (MSI)



Bar chart comparing model suitability in terms of performance and parameter sensitivity across different reanalysis in the 19 catchments.

- MSI considers both sensitivity indices and performance statistics, on average, ERA5, ERAI & JRA have better scores across most of the catchments
- **MSI** has the strength of ease and clearness to judge the **superiority and inferiority** of the reanalysis when compared to observations at catchment scale

- The ERA5 has the highest MSI compared to observations across the nineteen catchments, followed by the ERAI reanalysis.
- Overall, the four reanalysis datasets obtained relatively lower MSI values particularly in catchments that are mainly in arid and semi-arid areas of Kenya.

## Key Take Aways

- **Hydrological model performance:** ERA5, ERAI, & JRA55 obtained better KGE scores whereas CFSR had lower scores in both calibration and validation
- The 3 best performing datasets give similar scores to observations in calibration, but more variable and less robust results in validation (especially ERA5 and JRA55)
- **Sensitivity analysis:** ERA5 show similar Sobol's **sensitivity indices** for all model parameters compared to observations, whereas CFSR has highest **uncertainty**
- The MSI aggregates both **sensitivity indices & performance statistics**, providing a clear index to judge the **superiority** (or inferiority) of a reanalysis with respect to observations
- On average ERA5, ERAI (& JRA55) have better MSI scores across most of the Kenyan catchments: ERAI & ERA5 perform better than JRA55 & CFSR, and lead to more robust model parameters

**THANK YOU FOR YOUR ATTENTION**



Happy to take Your Questions & Comments

**Email:** [maureen.wanzala@welthungerhilfe.de](mailto:maureen.wanzala@welthungerhilfe.de)

or

[manyango03@gmail.com](mailto:manyango03@gmail.com)

Journal of Hydrology: Regional Studies 41 (2022) 101105

Contents lists available at ScienceDirect

**Journal of Hydrology: Regional Studies**

journal homepage: [www.elsevier.com/locate/ejrh](http://www.elsevier.com/locate/ejrh)

**Assessment of global reanalysis precipitation for hydrological modelling in data-scarce regions: A case study of Kenya**

Maureen A. Wanzala<sup>a,\*</sup>, Andrea Ficchi<sup>b,c</sup>, Hannah L. Cloke<sup>a,d,e,f</sup>, Elisabeth M. Stephens<sup>g,h</sup>, Heza M. Badjana<sup>i</sup>, David A. Lavers<sup>a,k</sup>

<sup>a</sup> Department of Geography & Environmental Science, University of Reading, Reading, UK  
<sup>b</sup> Department of Electronics, Information, and Biogeography, Politecnico di Milano, Milano, Italy  
<sup>c</sup> Department of Meteorology, University of Reading, Reading, UK  
<sup>d</sup> Department of Earth System, Uppsala University, Uppsala, Sweden  
<sup>e</sup> European Centre for Medium-Range Weather Forecasts, Reading, UK  
<sup>f</sup> Centre of National Research and Research Institute, CNRS, Uppsala, Sweden  
<sup>g</sup> Real-time and Climate Change Centre, The Hague, The Netherlands  
<sup>h</sup> School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK

**ARTICLE INFO**

**ABSTRACT**

**Key words:** Reanalysis data; Kenya catchments; Model calibration; Model performance; Sensitivity analysis; Model reliability index

**Study region:** 29 flood-prone catchments in Kenya, Eastern Africa

**Study focus:** Flooding is a major natural hazard especially in developing countries, and the need for timely, reliable, and actionable hydrological forecasts is prominent. Hydrological modelling is essential to produce forecasts but is a challenging task, especially in poorly gauged catchments, because of the inadequate temporal and spatial coverage of hydro-meteorological observations. Open access global meteorological reanalysis datasets can fill in this gap, because they have significant reach. This study assesses the performance of four reanalysis datasets (ERA5, ERA-Interim, CNRS, and JRA55) over Kenya for the period 1981–2020 on daily, monthly, seasonal, and annual timescales. We firstly evaluate the reanalysis datasets by comparing them against observations from the Climate Hazards group infrared Precipitation with Station. Secondly, we evaluate the ability of these reanalysis datasets to simulate streamflow using GRU2 model considering both model performance and parameter sensitivity and identifiability.

**New hydrological insight for the region:** While ERA5 is the best performing dataset overall, performance varies by season, and catchment and therefore there are marked differences in the suitability of reanalysis products for forcing hydrological models. Overall, wetland catchments in the western region and highlands of Kenya obtained relatively better scores compared to those in the arid and regions, this can inform future applications of reanalysis products for setting up hydrological models that can be used for flood forecasting, early warning, and early action in data-scarce regions, such as Kenya.

**1. Introduction**

Precipitation is arguably the most important driver of catchment hydrological response (e.g. [McLeod et al., 2021](#)), but it is

\* Corresponding author.  
E-mail address: [m.a.wanzala@pps.reading.ac.uk](mailto:m.a.wanzala@pps.reading.ac.uk) (M.A. Wanzala).

<https://doi.org/10.1016/j.ejrh.2022.101105>  
Received 29 December 2021; Received in revised form 6 May 2022; Accepted 7 May 2022  
Available online 18 May 2022  
2204-5818/Content Copyright © 2022 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Wanzala et al., 2022

<https://doi.org/10.1016/j.ejrh.2022.101105>

7 September 2023



[@\\_Maureen\\_AW](https://twitter.com/Maureen_AW)