UNIVERSITY OF PADOVA

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

MASTER THESIS IN WATER AND GEOLOGICAL RISK ENGINEERING

COMPARATIVE ANALYSIS OF HYDROLOGICAL MODELING USING GROUND-OBSERVED AND GLOBAL REANALYSIS PRECIPITATION DATASETS IN THE SECCHIA RIVER BASIN

THESIS ADVISOR:

PROF. MARCO MARANI

THESIS CO-ADVISOR:

ENG. PAOLO MASTROCOLA

STUDENT:

FREHIWOT KASSA TEFERI

STUDENT ID: 2045471

ACADEMIC YEAR

2023/2024

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to Padova University for providing me with the opportunity and resources to pursue my master's degree and conduct this study. I am thankful to the entire staff of the Water and Geological Risk Engineering for their Continuous support and guidance throughout my academic journey.

I am deeply thankful to my supervisor, Professor Marco Marani, for invaluable guidance, unwavering encouragement and insightful feedback, it has been vital to the success of this study. Your expertise and dedication have significantly shaped both my research and academic growth.

I would also like to extend my sincere thanks to my co-adviser, Eng. Paolo Mastrocola from HydroNova, for his thoughtful advice, constructive feedback and constant support throughout this process. I am especially grateful to his staff, particularly Paolo Polo and Giacomo Cogo, for their invaluable assistance.

I would also like to extend my acknowledgement to the Agenzia Interregionale per il fiume PO and the Servizio Idrografia e Idrologia Regionale e Distretto PO Struttura IdroMeteoClima - Arpae Emilia-Romagna for providing the observed ground stations hydro meteorological data.

I also thank my colleagues for their collaboration, friendship and support, which made this journey both fulfilling and motivating.

Finally, I would like to thank my family for their unconditional love and support. Your encouragement has been my greatest source of strength and motivation throughout this journey.

Thank you all for your contributions to this achievement.

Abstract

The accurate prediction of the hydrological processes responsible for flood generation is critical for effective water resource management and flood risk mitigation. This thesis presents a comparative analysis of hydrological modeling results when using ground-observed or ERA5-Land reanalysis data, within the Secchia river basin in Italy. The study employs the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) to perform continuous simulations based on the Soil Moisture Accounting (SMA) approach. The HEC_HMS model used was originally calibrated using observed rainfall and stream flow values. Here the performance of the pre-calibrated model is assessed when ERA5-Land rainfall estimates are used in place of ground observations. Additionally, bias correction is applied to the ERA5-Land reanalysis data to address systematic discrepancies, and hydrological simulations are conducted both before and after this correction to evaluate its impact. Evaluation metrics, such as the Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) are used to quantify model performance. Ground-observed data generally provide higher accuracy in streamflow simulations, whereas ERA5-Land reanalysis data demonstrate better spatial and temporal coverage, crucial in the case of ungauged or sparsely gauged basins. Bias-corrected ERA5-Land reanalysis data significantly improve the model performance, aligning more closely with the results derived from ground observations. This comparative analysis underscores the potential of integrating reanalysis datasets into hydrological models to improve water resource management and contribute valuable data for ungauged basins. The findings of this study enhance our understanding of the use of reanalysis data, especially after bias correction, in hydrological modeling and provide practical insights for their application in water resource management and future hydrological research.

Keywords: Hydrological model, Global reanalysis, Flood simulations, Comparative analysis, Secchia river basin.

1 Introduction

Flood risks quantification is a critical component in support of decision making and risk reduction strategies [1,2]. This study focuses on the Secchia River Basin in Italy, a tributary to the Po River and a region where hydrological modeling can offer valuable insights into streamflow behavior under different precipitation scenarios.

By using both ground-observed and ERA5-Land reanalysis precipitation datasets within a comprehensive hydrological modeling framework, the study aims to evaluate the performance of these datasets for streamflow modeling and improve our understanding of the basin hydrologic response. The Soil Moisture Accounting (SMA) approach is utilized, capturing the interactions between precipitation, infiltration and runoff in hydrological model effectively. SMA approach is especially valuable in regions with variable soil characteristics and complex hydrological conditions [3].

The primary objective is to compare the streamflow modelling accuracy and reliability that may be achieved using ERA5-Land precipitation data instead of ground-observed precipitation. For this study, hourly precipitation data from ERA5-Land are utilized for the period from January 1, 1999, to December 31, 2019, which comprises several flood events in the Secchia river basin. Additionally, the study explores the effects of bias-correcting ERA5-Land data on the hydrologic model performance.

1.1 Problem Statement

The accuracy of hydrological models depends significantly on the quality and resolution of precipitation data used as input. While ground-observed precipitation data are typically reliable, they are often spatially limited and may not adequately represent precipitation variability across the entire Basin. This limitation can lead to incomplete assessments of hydrological processes and to a limitation in streamflow prediction accuracy. On the other hand, global reanalysis datasets, such as ERA5-Land, offer high resolution and continuous precipitation information that could enhance hydrological modeling in ungauged or sparsely gauged areas. However, the reliability of these reanalysis datasets remains uncertain.

This study aims to address these challenges by conducting a comparative analysis of ground-observed precipitation data and ERA5-Land reanalysis data in the Secchia river basin. By incorporating bias correction techniques, this study evaluates the effectiveness of these datasets in improving streamflow predictions. Ultimately, the findings will provide valuable insights into the applicability of reanalysis data for water resource management and flood mitigation strategies, particularly in regions where ground-based observations are limited.

1.2 Objectives

This study aims to evaluate the accuracy of ERA5-Land precipitation data in simulating streamflow within the Secchia river basin, at Rubiera station. By conducting a comparative analysis between ground-observed precipitation data and ERA5-Land reanalysis data, the study specifically assesses the reliability of ERA5-Land in capturing rainfall variability and streamflow dynamics. The primary objectives are listed below:

1. Assess the performance of hydrological simulations from a model calibrated with ground-observed data when ERA5-Land reanalysis precipitation data are used.
2. Evaluate the effect of bias correction on the ERA5-Land reanalysis precipitation data.
3. Investigate model performance when uncorrected and bias-corrected ERA5 data are used, using event-scale metrics (peak flow accuracy) and long-term metrics (mean streamflow rate, precipitation partitioning). This evaluation aims to determine the accuracy and reliability of each dataset in representing hydrological conditions.
4. Perform continuous model simulations from January 1, 2000, to December 31, 2019, to analyze the basin long-term hydrological response and understand annual variations in streamflow.
5. Compare ground-observed and ERA5-Land reanalysis precipitation data, to understand the strengths and limitations of each dataset in hydrological modeling.
6. Offer practical recommendations for integrating these datasets into hydrological models to enhance water resource management and flood mitigation strategies in the Secchia river basin and similar regions.
7. Provide information that can be used to inform hydrological analysis in ungauged or sparsely gauged regions, highlighting the potential of corrected reanalysis datasets to fill observational gaps.

1.3 Structure of the Thesis

This thesis is organized into the following chapters, each focusing on different aspects of the study to provide a comprehensive understanding of the comparative analysis of hydrological modeling using different precipitation datasets.

Chapter 1, Introduction: This chapter provides an overview of the problem and objectives the of the study. It introduces the Secchia River Basin, highlights the importance of hydrological modeling using precipitation datasets and outlines the structure of the thesis.

Chapter 2, Literature Review: Sets the foundation by reviewing existing literature on hydrological modeling techniques and the role of precipitation datasets in these models. It covers the characteristics of ground-observed and global reanalysis data and discusses bias correction methods.

Chapter 3, Study Area and Model Setup: Introduces the Secchia River Basin, detailing its geographical and hydrological characteristics. It also describes the setup, including the division into sub-basins and the configuration of the HEC-HMS model using the existing calibrated model developed by HydroNova. The Soil Moisture Accounting (SMA) approach and other key modelling components are explained.

Chapter 4, Data and Methods: Outlines the data sources and methods used in the study. It includes descriptions of the ground-observed precipitation data and ERA5-Land reanalysis precipitation data, as well as the process of bias correction applied to ERA5-Land reanalysis data. This chapter also details the setup of the HEC-HMS model, the simulation scenarios and the performance evaluation metrics such as the Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) are employed. The methods for partitioning precipitation, as well as calculating mean streamflow, are also discussed.

Chapter 5, Results and Discussion: Presents the results from the model simulations and analyzes their implications. It includes results from simulations using both ground-observed and ERA5-Land reanalysis precipitation data. It discusses the performance of the model in predicting streamflow, both before and after bias correction. The comparative analysis is presented using performance metrics, such as event-scale and long-term metrics and the impact of correction on model performance is evaluated. This chapter also discusses the analysis of precipitation partitioning, mean streamflow and other key hydrological responses within the basin.

Chapter 6, Conclusion: Summarizes the key findings of the study, highlights the contributions to the field of hydrological modeling and discusses the limitations of the study. It also offers recommendations for future research based on the study results.

The Bibliography section lists all references used throughout the thesis.

2 Background

2.1 Hydrological Modeling

Hydrological modeling plays a crucial role in understanding and predicting water flows within river basins. It is an essential tool for water resource management, flood control and environmental protection, as it enables the simulation of streamflow, water balance and various hydrological processes based on input data such as precipitation, temperature and land use patterns [5]. The accuracy of hydrological models is significantly influenced by the quality of the input data, with precipitation being one of the most critical drivers. Precipitation data directly influence the prediction of surface runoff, river discharge and soil moisture content, thus affecting the overall performance of hydrological simulations [2].

Given its importance, the accurate representation of precipitation in hydrological models is a major challenge. Precipitation data can be obtained from multiple sources, including ground-based observations, satellite-based remote sensing and global reanalysis datasets [2]. While ground-observed data are often considered the most accurate due to direct measurements, their spatial coverage is limited, especially in remote and topographically complex regions [6]. Therefore, researchers have increasingly turned to global reanalysis products, such as ERA5, which provide extensive spatial and temporal coverage of precipitation data for hydrological modeling. However, these datasets often require further correction to account for biases and uncertainties before being applied in hydrological studies [2].

In this study, the HEC-HMS model was employed for continuous hydrological simulations with runoff/infiltration described based on the Soil Moisture Accounting (SMA) approach. The objective of the present analysis is achieving a more accurate representation of hydrological processes in the Secchia River Basin, ensuring better predictions of streamflow.

2.2 Precipitation Datasets in Hydrological Modeling

Various global-scale precipitation datasets that exist, such as Climate Hazards group Infrared Precipitation with Stations (CHIRPS) (https://www.chc.ucsb.edu/data/chirps), ERA5 (https://cds.climate.copernicus.eu/) and Multi-Source Weighted-Ensemble Precipitation (MSWEP) (https://www.gloh2o.org/), have been evaluated for their performance in hydrological modeling applications [2]. Each dataset has its strengths and weaknesses depending on spatial resolution, temporal frequency, and underlying methodologies. Hence, the importance of choosing appropriate precipitation inputs based on regional hydrological characteristics to enhance model reliability [2]. Various large-scale climate data sets at different spatiotemporal scales have also been developed from station (in situ) observations [6]. However, gauge measurements have several drawbacks, such as limited spatial coverage, and deficiencies over most oceanic and sparsely populated areas [6]. Satellite-based datasets, on the other hand, offer broader coverage but may include random noise and systematic errors, which can lead to discrepancies in hydrological model outputs [2].

Most of the existing research has focused on comparisons between datasets or their applications in specific case studies using daily, monthly or annual data [2, 4, 6]. This study aims to assess the performance of hydrological simulations driven by both ground-observed and ERA5-Land reanalysis hourly precipitation data cover a 20-year period. By utilizing two decades of hourly data, the study aims to evaluate the datasets capability to replicate observed streamflow and capture hydrological responses under diverse climatic and hydrological scenarios. This long-term assessment will provide valuable insights into the suitability of using ERA5-Land hourly data for detailed hydrological modeling and water resource management, particularly in regions where long-term ground observations are sparse or unavailable.

2.3 Bias Correction Methods in Hydrology

Applying bias correction techniques can enhance the accuracy of hydrological models [2]. The spatial and temporal biases observed in datasets like ERA5 and MSWEP suggest that a single dataset may not always provide reliable inputs for hydrological simulations without appropriate corrections [2]. Bias correction methods can address these issues by aligning the simulated precipitation data with observed values, thereby improving the reliability of model outputs [2]. Systematic errors in precipitation estimates can lead to significant inaccuracies in hydrological simulations, especially in streamflow predictions. Various bias correction techniques have been developed to address these challenges.

Scaling (SCL) is one of the simplest methods employed for bias correction. This technique adjusts the mean of simulated hourly precipitation to align with observed values using a constant scaling factor. While SCL effectively addresses mean biases, it may not adequately capture variability or extreme events, which are crucial for accurate hydrological modeling. This limitation can significantly affect the representation of temporal variations in precipitation, especially in regions with substantial hourly fluctuations [7].

Local Intensity Scaling (LOCI) focuses on adjusting both the frequency of wet days and the intensity of precipitation by establishing a threshold. LOCI ensure that the number of simulated rainy days above this threshold matches observed data, effectively mitigating the overestimation of light rainfall events. LOCI's focus on hourly data makes it particularly relevant for hydrological applications where timing and intensity are critical [7].

Another widely used technique, Empirical Quantile Mapping (EQM), addresses the entire distribution of precipitation data. This non-parametric approach calibrates the cumulative distribution function (CDF) of simulated hourly precipitation to align with observed distributions. EQM effectively corrects mean, variance and quantiles, capturing both average and extreme conditions [8]. This method is particularly advantageous for correcting data with non-linear distributions and is commonly applied in climate and hydrological studies.

Global Homogeneous Bias Correction (GHBC) is another important method that aims to correct systematic biases across a uniform bias correction across the entire dataset. GHBC operates by applying a single correction factor to the entire dataset, making it particularly efficient for datasets with large spatial coverage. This method is beneficial in studies where maintaining spatial consistency is essential, such as in regional hydrological modeling [9]. While GHBC bias is assumed to be consistent throughout the dataset, meaning that the same correction factor is applied for the entire dataset, the SCL adjusts the data for each sub-basins based on the intensity or distribution of precipitation.

In this study, the Global Homogeneous Bias Correction method was employed for the ERA5-Land precipitation data. This technique assumes a consistent bias across the study area, applying a single correction factor to the entire dataset to address systematic over or underestimations. This approach simplifies the correction process while enhancing the accuracy of hydrological models.

Overall, bias correction is critical for improving reanalysis data performance in hydrological simulations. By applying Global Homogeneous Bias Correction to ERA5-Land reanalysis data, this study ensures that model outputs are reliable. The corrected hourly precipitation data were compared to original ERA5-Land data and ground observations in the Secchia River Basin, emphasizing the importance of bias correction in enhancing hydrological modeling accuracy.

3 Study Area and Model Setup

3.1 Secchia River Basin Overview

The Secchia River Basin, located in the Emilia-Romagna region of northern Italy, is an important affluent of the Po River, one of Italy's largest and most vital rivers. The Secchia river basin covers approximately 2,300 square kilometers, encompassing a diverse range of topographical features and land uses [10].

The Secchia River is 172 kilometers long, originating from the Alpe di Succiso at an elevation of 2,017 meters above sea level in the Tuscan-Emilian Apennines. From its mountainous source, the river flows southeastward through the region before merging with the Po River near Mantova. This river systems varied landscape and flow regime are crucial for understanding the hydrological dynamics of the basin [10].

This study specifically focuses on the 1,278 square kilometers of the basin upstream of Rubiera, the outlet point for the modeled area. The study area includes the rivers reach from its mountainous origins to Rubiera. Within this area, the basin is subdivided into eleven sub-basins, each with distinct hydrological characteristics. The upper reaches are defined by steep, forested slopes that contribute to rapid runoff and potential soil erosion, whereas the lower reaches are predominantly agricultural, making them more susceptible to flooding during heavy rainfall events [11].

The Rubiera gauge station, positioned at the outlet of the study area, provides measured discharge data used for validating the hydrological models applied in this study. By incorporating data from Rubiera and considering the diverse characteristics of the sub-basins, this study aims to evaluate the influence of different precipitation datasets on streamflow simulations.

Figure 1: Study Area of the Secchia River Basin

3.2 Sub-basins and Hydrological Characteristics

The study area consists of eleven sub-basins, each demonstrating distinct hydrological and geographical characteristics that influence their responses to precipitation events. Variations in size, elevation, slope and land use across these sub-basins directly impact processes such as streamflow generation, infiltration, surface runoff and baseflow contributions. Understanding these differences is critical for accurately modeling the basin's hydrological behavior and predicting streamflow under varying precipitation conditions.

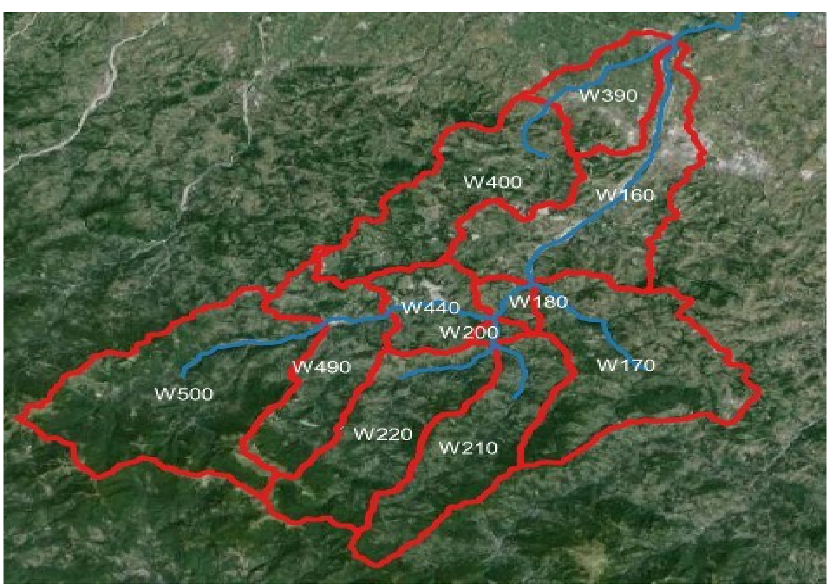

Figure 2: Map of sub-basins

Table 1. Characteristics of Sub-basins in the Study Area

Sub basin	W390	W160	W170	W180	W440	W200	W210	W220	W400	W490	W500
Area (km^2)	71.58	174.13	186.80	17.34	56.38	6.76	130.86	136.11	149.72	113.14	235.27
Basin Slope (°)	5.27	14.03	19.66	21.19	19.44	26.94	23.17	26.25	17.76	23.58	29.98

The sub-basins within the study area contribute significantly to the overall complexity of the basin's hydrological behavior. Their diverse sizes and slopes play a crucial role in determining how precipitation translates into runoff, thereby influencing the performance and accuracy of hydrological models. Variations in sub-basin area can affect the amount of water retained during precipitation events, while differences in slope can determine the speed at which runoff occurs. Understanding these factors is essential for developing effective models that can reliably simulate streamflow across the Secchia River Basin.

3.3 HEC-HMS Model Setup

The Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS), developed by the U.S. Army Corps of Engineers, is a flexible tool designed to model rainfall-runoff processes and assess various hydrological responses, including streamflow, infiltration and surface runoff [13]. The model's adaptability to various watershed configurations makes it ideal for simulating complex hydrological conditions.

3.3.1 Model Configuration

For this study, the HEC-HMS model configuration is based on an existing calibrated version developed by HydroNova for the Secchia River Basin. This pre-existing model has been previously adjusted to fit historical streamflow data, ensuring that its parameters are well-suited for the basin's specific conditions. The setup was used without changes to its original parameters to ensure a direct comparison of results under different precipitation scenarios.

The model configuration includes the division of the basin into sub-basins, each with its own set of hydrological parameters. The Soil Moisture Accounting (SMA) approach is used as the primary loss method, which involves continuous simulation to track soil moisture changes and partition precipitation. Additional model components include the Clark Unit Hydrograph for precipitation-runoff transformation, the simple canopy method for canopy interception, the linear reservoir method for baseflow and the Muskingum-Cunge method for routing.

The basin is divided into 11 sub-basins to capture spatial variability and key parameters are based on the model's predefined configuration. The model includes the following main components:

❖ Loss Method: Soil Moisture Accounting (SMA) approach

The Soil Moisture Accounting (SMA) method is utilized to partition of precipitation into various hydrological components, including infiltration, surface runoff, evapotranspiration and groundwater recharge. It is a continuous loss method that tracks changes in the soil moisture profile over time, making it suitable for simulating annual variations and antecedent moisture conditions. This model is designed to capture sort-term as well as long-term hydrological portioning.

Key parameters for the SMA method include:

- Maximum Soil Storage Capacity (mm): Represents the total water-holding capacity of the soil profile.
- Initial Soil Moisture Content (% of storage): The starting soil moisture level at the beginning of the simulation.
- Impervious Area (%): The proportion of the sub-basin that is covered by impervious surfaces, directly contributing to surface runoff.
- Maximum Percolation Rate (mm/hr.): The rate at which water percolates from the soil into the groundwater system.

The SMA method uses the following water balance equation:

$$\Delta S = P - (Qs + ET + I + G) \qquad (1)$$

Where ΔS is the change in soil moisture storage (mm), P is the precipitation (mm), Qs is the surface runoff (mm), ET is the evapotranspiration (mm), I is the infiltration (mm) and G is the groundwater recharge (mm).

❖ Transform Method: Clark Unit Hydrograph method

The Clark Unit Hydrograph method is used to transform excess precipitation into direct runoff at the outlet of each sub-basin. This method combines Time of Concentration (Tc) and Storage Coefficient (R) parameters to represent the timing and attenuation of peak flows.

Key parameters for this method include:

- Time of Concentration (Tc): The time required for water to travel from the furthest point in the sub-basin to the outlet.
- Storage Coefficient (R): Represents the delay and attenuation of runoff due to storage effects within the sub-basin.

The Clark method uses the following integral equation to compute the outflow hydrograph:

$$Q(t) = \frac{A}{Tc} \times \int_0^t P(\tau) e^{\frac{-(t-\tau)}{R}} d\tau \qquad (2)$$

Where Q(t) is the direct runoff at time t, A is the sub-basin area and P(τ) is the excess precipitation at time τ.

This equation allows for detailed representation of runoff timing and magnitude, making it ideal for simulating peak flow events.

❖ Muskingum-Cunge Method

The Muskingum Method is used for channel routing to simulate the downstream movement of water through the Secchia River and its tributaries. This method accounts for storage and attenuation effects, making it suitable for simulating riverine flow dynamics and peak discharge propagation.

Key parameters include:

- Muskingum Storage Coefficient (K): Reflects the travel time of water through the reach.
- Muskingum Weighting Factor (X): A factor ($0 \leq X \leq 0.5$) that determines the relative weight of inflow and outflow in the storage calculation.

The Muskingum method is represented by the following equation:

$$Q_{out}(t) = C_1 {Q_{in}}^{t} + C_2 {Q_{in}}^{t-1} + C_3 {Q_{out}}^{t-1} \qquad (3)$$

Where Q_{out} (t) is the outflow at time t, Q_{in}^{t} is the inflow at time t, Q_{in}^{t-1} is the inflow at previous time step and Q_{out}^{t-1} is the outflow at previous time step. The Coefficients C1, C2 and C3 are calculated using:

$$C_1 = \frac{\Delta t - 2KX}{2K(1-X) + \Delta t} \qquad (4)$$

$$C_2 = \frac{\Delta t + 2KX}{2K(1-X) + \Delta t} \qquad (5)$$

$$C_3 = \frac{2K(1-X) - \Delta t}{2K(1-X) + \Delta t} \qquad (6)$$

These coefficients determine how inflow and storage are converted into outflow over each time step.

❖ Meteorological Inputs: Temperature and Potential Evapotranspiration

The existing model configuration includes temperature and potential evapotranspiration (PET) data, which were directly used without modification. These values are sourced from the original HydroNova setup. The inclusion of PET data allows for an accurate representation of evapotranspiration processes, a critical component of the water balance within the basin.

In the existing model developed by HydroNova, the PET is represented using monthly average and the snowmelt process is modeled using the temperature index which relies on temperature data. By using the predefined meteorological inputs defined above from existing model configuration, the study ensures consistency with the previously validated model setup, enabling us to focus on analyzing the impact of varying precipitation inputs on hydrological responses.

4 Data and Methods

4.1 Data Sources

For this study, data were collected from two primary sources, ground-observed precipitation and discharge data, and ERA5-Land reanalysis data. Each dataset provides distinct advantages and plays a key role in evaluating the performance and accuracy of the hydrological models used in the Secchia River Basin.

4.1.1 Ground Observations: Precipitation and Discharge

The ground observations utilized in this study were sourced from the Emilia-Romagna Regional Environmental Protection Agency (ARPA), covering the period from January 1, 1999, to December 31, 2019. This dataset consists of hourly precipitation data for each sub-basins, and hourly discharge measurement at Rubiera station. The ground observation precipitation data used in the simulation of this study are interpolated values from the records obtained from ARPA for each sub-basin in the HMS model.

The interpolation was performed by HydroNova using the inverse distance weighting (IDW) method. The inverse distance weighting method is commonly used method in watershed precipitation interpolation. The horizontal distance between the location without the precipitation records, and its surrounding rainfall station, and the order of distance decide the weights given to each rainfall station [14]. The weight for precipitation at rainfall stations i in sub-watershed p, W_{pi} is given by:

$$W_{pi} = \frac{\frac{1}{{d_{pi}}^m}}{\sum_i^n \frac{1}{{d_{pi}}^m}} \qquad (7)$$

$$\mathrm{P}_p = \sum_{i=1}^{n} W_{pi} * P_i \qquad (8)$$

where m is the order of distances; d_{pi} is the distance between the center of sub-watershed p and rainfall station i; n is the total number of rainfall stations. After the weights are decided, unknown precipitation for each sub-watershed could be estimated. The precipitation of sub-watershed p, P_p, is defined as formula (8), in which Pi is the precipitation at rainfall station i.

Figure 3 below shows a map of the rain gauges (red dots) and flow gauge (the blue dot) stations locations, where the ground observed precipitation data and the discharge measurement are collected.

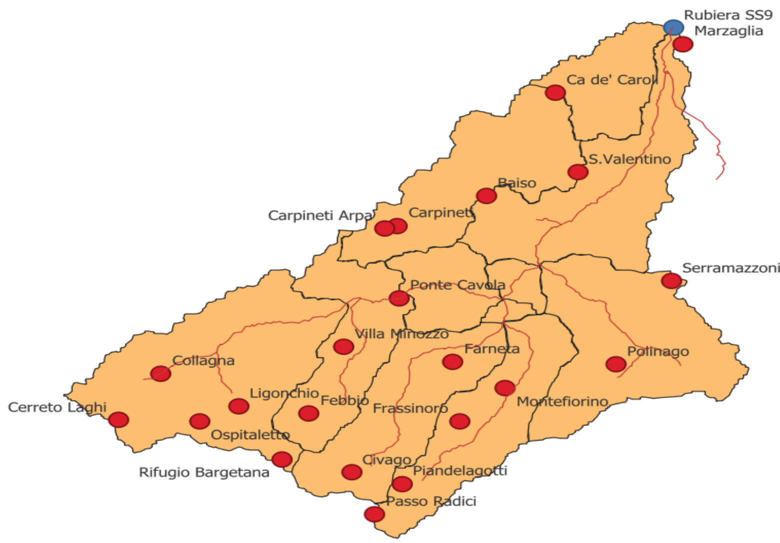

Figure 3: Rain gauge and flow gauge station

The precipitation data are complete for the entire study period (January 1, 1999, to December 31, 2019) and provide detailed insights into localized precipitation patterns across the basin.

Figure 4 illustrates the hourly precipitation for the sub-basin W160 based on ground-observed data from January 1, 1999, to December 31, 2019.

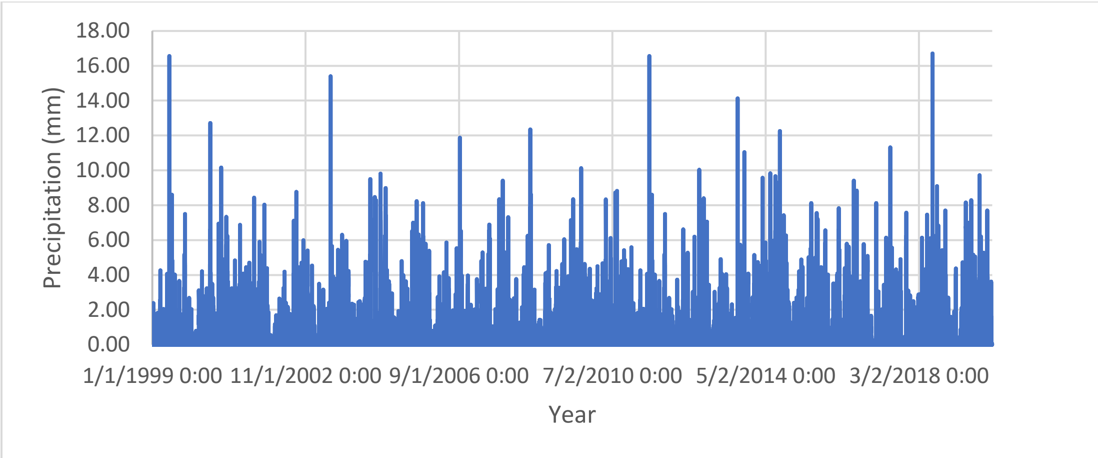

Figure 4: Ground-Observed Hourly Precipitation for the sub-basin W160 (1999–2019)

Similarly, the available discharge data, covering the period from January 1, 2000, to December 31, 2019, were collected at the Rubiera station, offering essential information on streamflow characteristics. However, there are some gaps in the discharge records, specifically missing data for the whole years in 2000, 2001, 2002, 2013 and 2014, except for one day record on December 31, 2013. There are also missing records in 2015 from June 22, 2015, to September 3, 2015, and in December 2015. To address these gaps, the analysis focused on the available discharge data to ensure a comprehensive evaluation of the hydrological models.

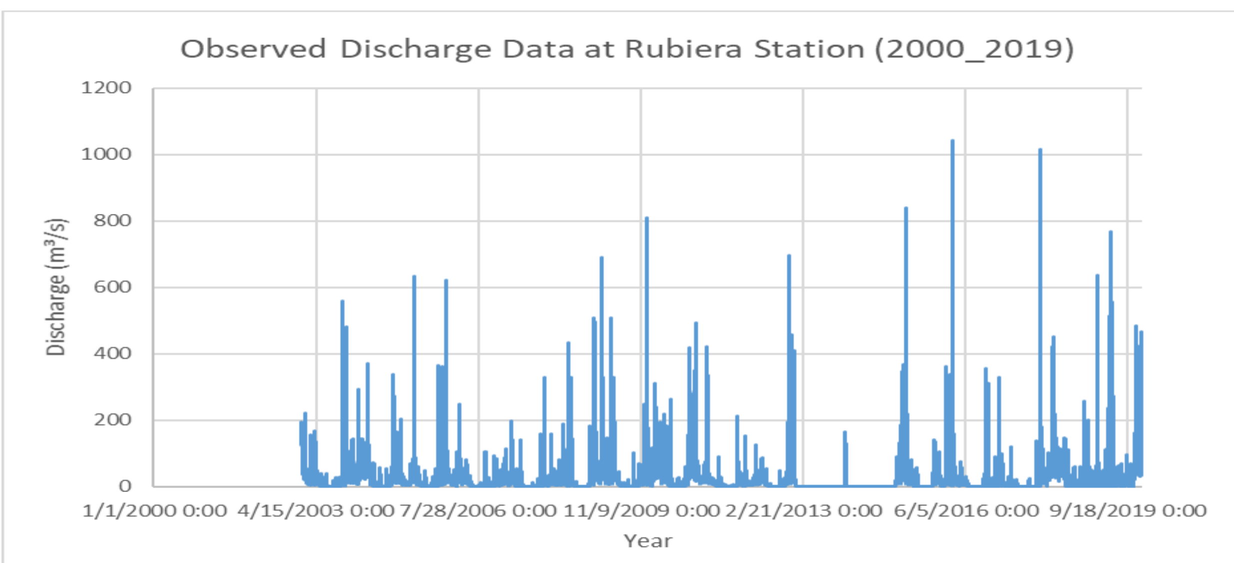

Figure 5: Hourly Observed Discharge Data at Rubiera Station from 2000 to 2019

Figure 5 illustrates key hydrological dynamics within the Secchia River Basin, highlighting significant peak discharge events. Notable peaks recorded in 2009, 2015, 2016 and 2017 are illustrated in table 2 below. Despite the gaps in data, the available records provide valuable insights into the hydrological patterns of the basin and enhance our understanding of the relationship between precipitation and discharge over time.

Table 2: Observed Peak Discharges at Rubiera Station

Date	Time	Observed Peak Discharge (m^3/s)
February 29, 2016	04:00 AM	1041.3
December 12, 2017	05:00 AM	1016.9
March 25, 2015	08:00 PM	839.4
December 25, 2009	03:00 AM	809

4.1.2 ERA5-Land Reanalysis Precipitation Data

For this study, hourly ERA5-Land reanalysis precipitation data is downloaded from Google Earth Engine (https://earthengine.google.com/), using the shapefile of the Secchia River sub-basins to extract relevant data specific to the study area. The data cover the period from January 1, 1999, to December 31, 2019, aligning with the ground-observed precipitation data.

Reanalysis is a process that combines historical weather observations with models to generate consistent time series of multiple climate variables. They provide a comprehensive description of the observed climate as it has evolved during recent decades [12]

ERA5-Land, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), is an advanced reanalysis product providing a spatial resolution of 9 km and a temporal resolution of 1 hour, offering detailed insights into precipitation dynamics. The dataset improves on the standard ERA5 product with enhanced spatial granularity, making it suitable for capturing variations in land surface variables. ERA5-Land, covering the period from January 1950 to the present, is a valuable resource for environmental management and climate adaptation, enhancing our understanding of land-atmosphere interactions [4, 12].

Total precipitation in ERA5-Land includes liquid and frozen water (rain and snow) falling to the surface, generated by both large-scale weather patterns (e.g., troughs and cold fronts) and convective systems. Precipitation measurements exclude moisture such as fog, dew, or water that evaporates before reaching the ground. The variable is accumulated over the forecast period and expressed in meters, representing the depth of water evenly distributed over the model grid [12].

Figure 6 presents the hourly precipitation for the sub-basin W160, obtained from ERA5-Land reanalysis, for the period from January 1, 1999, to December 31, 2019.

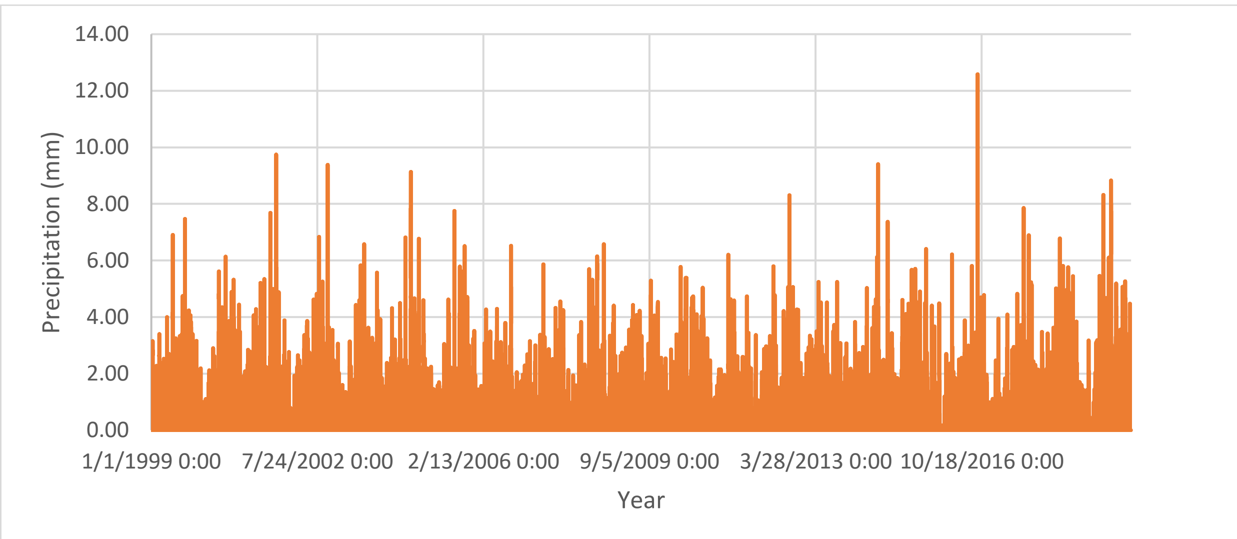

Figure 6: ERA5-Land hourly Reanalysis Precipitation Data for the sub-basin W160 (1999–2019)

4.1.3 Precipitation Data Comparison (Ground-Observed vs. ERA5-Land Reanalysis)

Figure 7 shows the discrepancies between ground observed precipitation data and the ERA5-Land precipitation data across the sub-basins study area. For instance, sub-basin W390 has a ratio of 0.84, indicating that the observed precipitation is lower than the ERA5-Land estimate. Similarly, the other basins indicate deviation of the ERA5-Land precipitation data from the ground observed precipitation. On the other hand, sub-basin W220 shows a ratio of 0.99, indicating close correlation between observed and ERA5-land precipitation. Notable exception is sub-basin W500 where the observed precipitation exceeds the ERA5-land precipitation data with a ratio of 1.15.

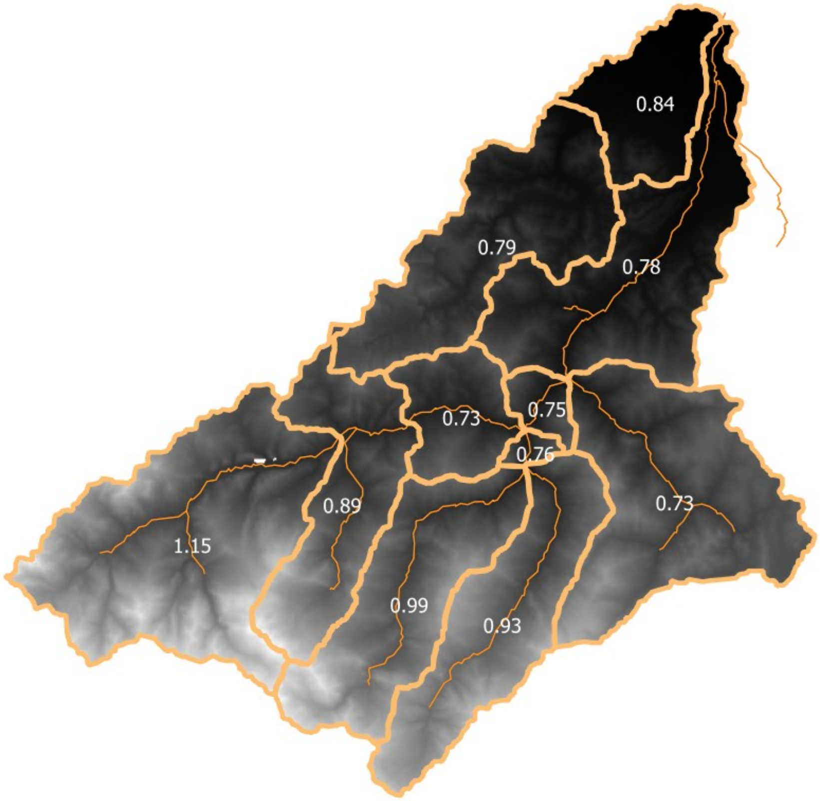

Figure 7: Precipitation Comparison Map

To evaluate the accuracy and reliability of different precipitation datasets, a comparative analysis is conducted using ground-observed precipitation data and ERA5-Land reanalysis data. This analysis focuses on two randomly selected sub-basins, one located in upstream (Sub-Basin W220), and another located closer to the downstream (Sub-basin W390), ensuring a representative comparison of both datasets across varying hydrological conditions. The sub-basin closer to each other display similar characteristics to each other compared sub-basins which are far from each other.

The comparison included analyzing annual cumulative precipitation for these sub-basins over the period from January 1, 1999, to December 31, 2019. The objective is to assess the degree of agreement between the datasets and to identify any significant differences.

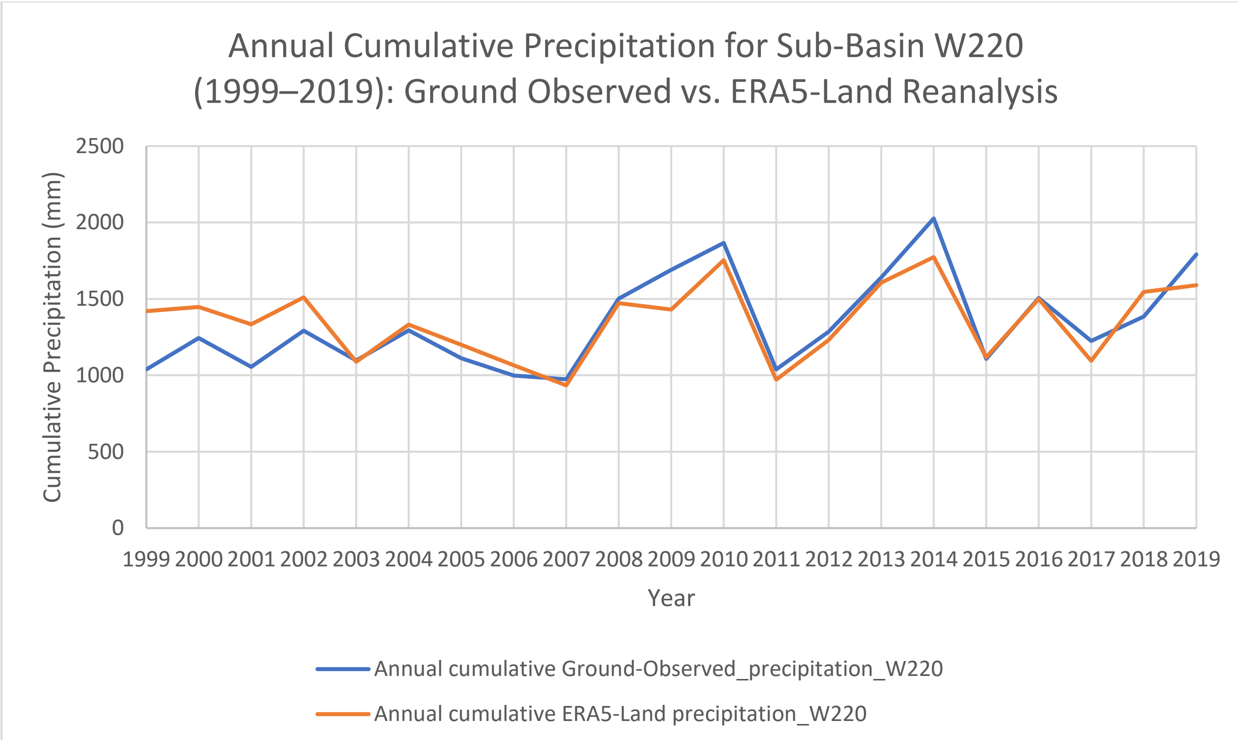

Figure 8a: Annual Cumulative Precipitation for Sub-Basin W220 Over time (1999–2019)

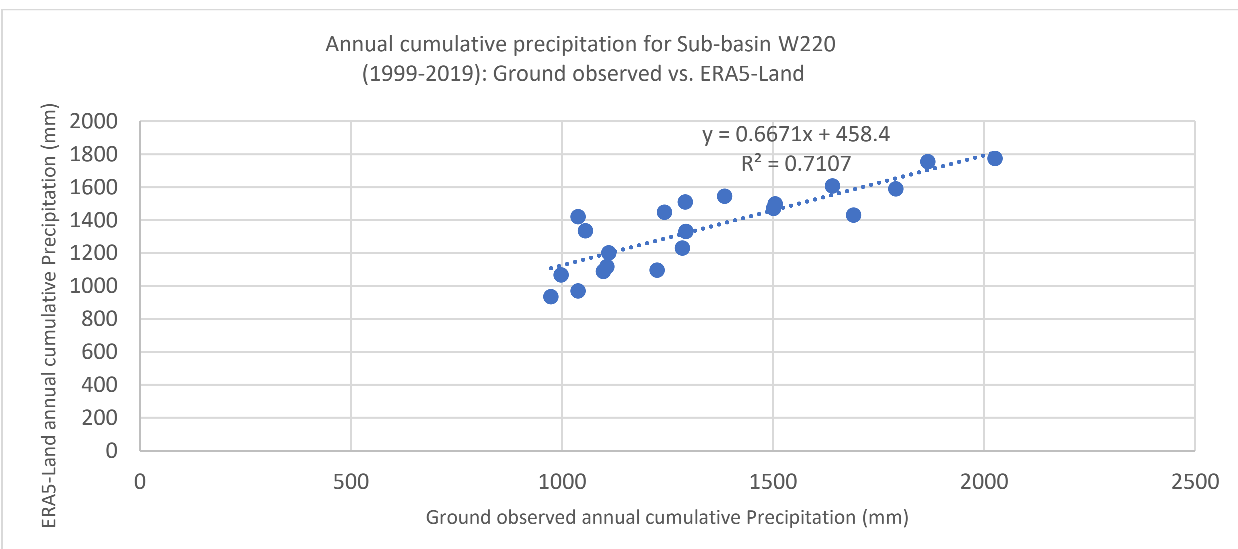

Figure 8b: Linear regression analysis for Sub-Basin W220

Figure 8a presents a comparative time-series plot of the annual cumulative precipitation for Sub-Basin W220, using both ground-observed data and ERA5-Land reanalysis data for the period from 1999 to 2019. The analysis shows ERA5-Land generally tending to overestimate precipitation.

The ground-observed precipitation data in Sub-Basin W220 showed notable fluctuations, ranging from a low of 973.68 mm in 2007 to a peak of 2026.27 mm in 2014. ERA5-Land typically estimated higher annual values. For example, in 1999, ERA5-Land recorded 1420.18 mm compared to 1038.35 mm observed at the ground, resulting in a difference of approximately 31%.

However, during extreme rainfall season the ERA5-Land precipitation underestimates the record. For example, in 2014, when ground observations recorded 2026.27 mm, ERA5-Land captured only 1773.39 mm, leading to a 252.88 mm underestimation. Conversely, in 2007, the datasets were more consistent, with a difference of just 39.46 mm.

The cumulative difference over the entire period for Sub-Basin W220 was 250.95 mm, reflecting a moderate overestimation by ERA5-Land. While the reanalysis data closely matched ground observations in some years, discrepancies during extreme events indicate the need for caution when using ERA5-Land for hydrological assessments in this region.

Figure 8b shows the linear regression analysis between annual cumulative ground-observed precipitation and ERA5-Land precipitation. The diagram illustrates ERA5-land underestimates the precipitation values compared to the ground observations.

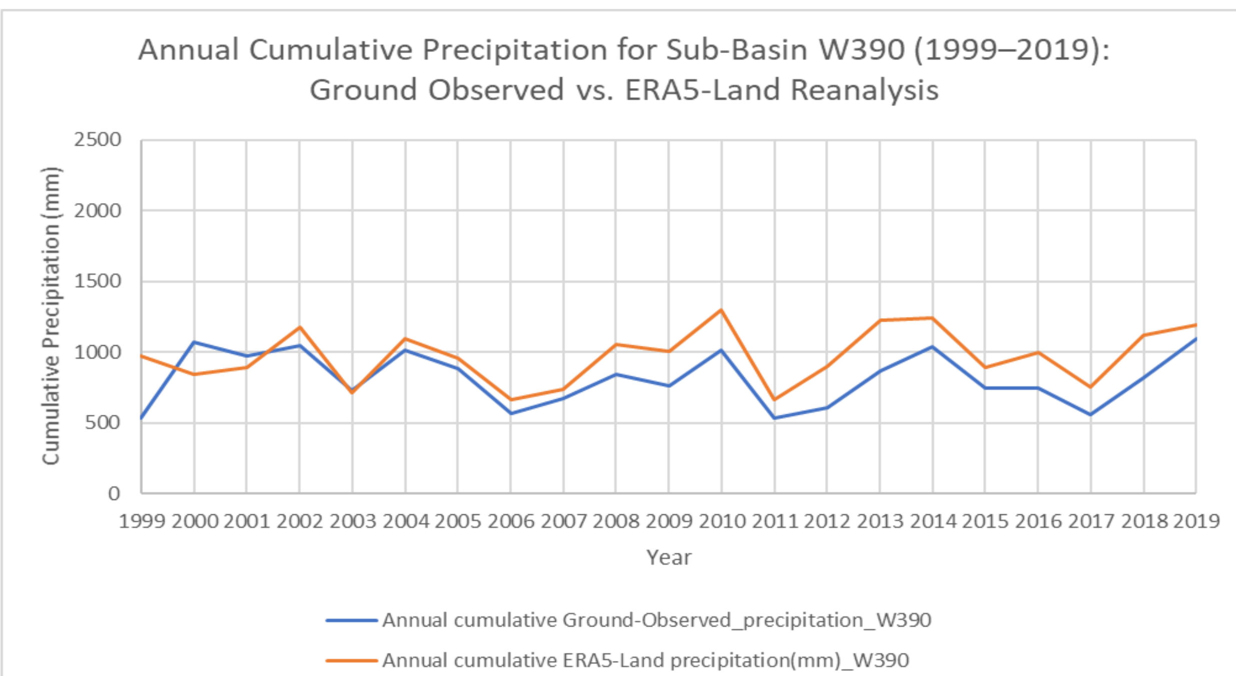

Figure 9a: Annual Cumulative Precipitation for Sub-Basin W390 (1999–2019)

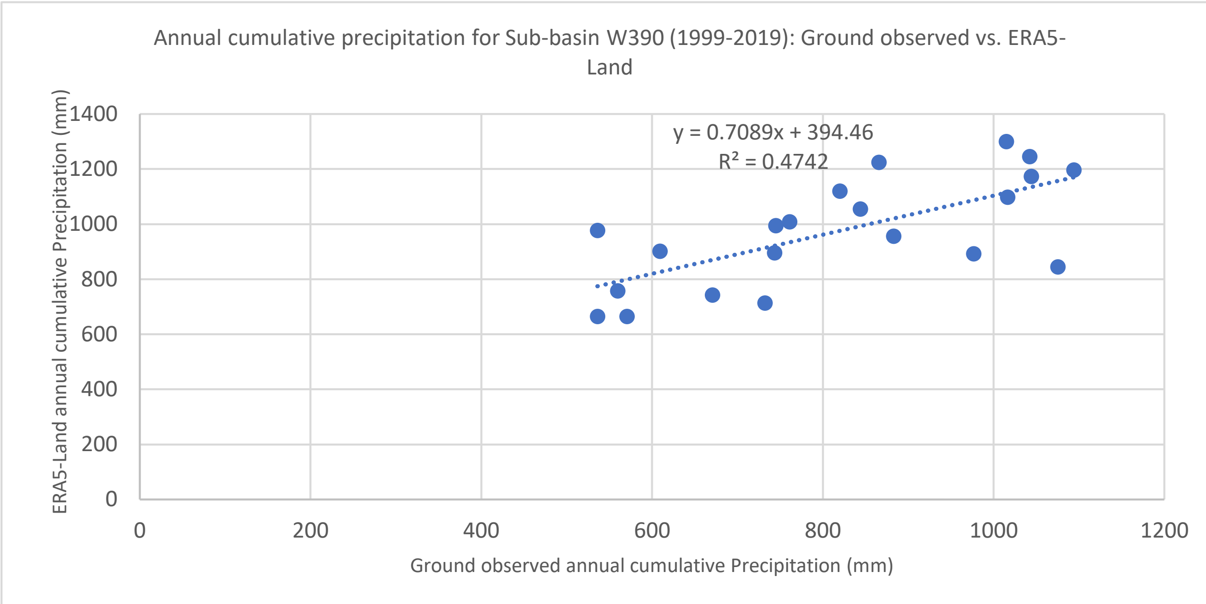

Figure 9b: Linear regression analysis for Sub-Basin W390 (1999–2019)

The analysis for Sub-basin W390 shows a notable pattern, with ground-observed precipitation values ranging from 536.01 mm in 1999 and 2011 to a maximum of 1094.07 mm in 2019. ERA5-Land consistently overestimated annual precipitation, for example, in 1999, ERA5-Land recorded 977.06 mm, leading to a notable discrepancy of -441.05 mm compared to ground observations. Similar overestimations were noted in 2010, where ERA5-Land indicated 1300.88 mm against an observed 1014.97 mm.

Figure 9b illustrates the linear regression analysis between ground-observed precipitation and ERA5-Land precipitation. The figure illustrates that there is a notable discrepancy between the two datasets.

The cumulative difference for Sub-Basin W390 over the study period amounted to 3294.52 mm, underscoring a consistent bias in the ERA5-Land dataset for this area. This significant overestimation suggests challenges in accurately representing localized climate effects and topography in this region.

Table 3: Summary of Annual Cumulative Precipitation: Sub-Basin W220 and W390

Datetime	Annual cumulative Ground-Observed _precipitation_W220 (mm)	Annual cumulative ERA5-Land precipitation _W220 (mm)	Precipitation Difference _W220 (mm)	Annual cumulative Ground-Observed_ precipitation _W390 (mm)	Annual cumulative ERA5-Land precipitation _W390 (mm)	Precipitation Difference _W390 (mm)
1999	1038.4	1420.2	-381.8	536.0	977.1	-441.1
2000	1242.7	1447.5	-204.8	1075.3	845.6	229.8
2001	1055.9	1334.6	-278.8	976.9	893.0	83.9
2002	1291.8	1510.4	-218.6	1044.4	1173.7	-129.2
2003	1097.6	1087.9	9.7	732.3	713.9	18.3
2004	1293.6	1331.7	-38.0	1016.4	1098.6	-82.2
2005	1110.9	1200.4	-89.6	883.0	957.1	-74.1
2006	998.0	1066.5	-68.5	570.5	664.8	-94.3
2007	973.7	934.2	39.5	671.0	742.4	-71.5
2008	1501.3	1471.5	29.8	844.1	1055.8	-211.7
2009	1691.1	1430.8	260.3	761.1	1009.5	-248.3
2010	1866.5	1754.0	112.5	1015	1300.9	-285.9
2011	1038.4	970.4	68.0	536.0	665.0	-129.0
2012	1285.2	1230.5	54.7	609.2	901.9	-292.7
2013	1640.7	1606.2	34.5	865.9	1225.2	-359.4
2014	2026.3	1773.4	252.9	1042.3	1245.3	-203.0
2015	1106.5	1117.3	-10.8	743.4	896.8	-153.5
2016	1505.4	1498.9	6.5	745.0	995.4	-250.4
2017	1225.6	1095.8	129.8	559.8	757.8	-198.1
2018	1385.1	1545.3	-160.2	820.12	1120.0	-299.8
2019	1791.0	1589.0	202.0	1094.1	1196.6	-102.6

Table 3 provides a detailed comparison of ground-observed and ERA5-Land precipitation data for both sub-basins. The findings suggest that while ERA5-Land captures overall precipitation trends, its accuracy varies significantly between sub-basins. The upstream sub-basin (W220) displayed relatively better alignment with ground observations, while the sub-basin near to the downstream area (W390) showed larger discrepancies, highlighting the limitations of reanalysis data in capturing localized hydrological conditions.

4.1.4 Bias Correction for ERA5-Land Data

To improve the accuracy of the ERA5-Land reanalysis data, a bias correction process is applied. This process involves comparing ERA5-Land precipitation data with ground-observed precipitation data to identify and adjust systematic biases. The bias correction adjusts the reanalysis data to better match the observed precipitation, ensuring that the ERA5-Land data provide a more accurate representation of precipitation for use in hydrological modeling. This corrected dataset is utilized in HEC-HMS simulations to evaluate its performance against the ground-observed data. This evaluation ensures that the simulation outputs reflect more realistic hydrological dynamics and can be reliably used for water resource management and planning.

The bias correction process includes calculating a Global Homogeneous Correction Factor, which adjusts for systematic biases across the entire Secchia River Basin. This factor ensures consistency between the ERA5-Land precipitation data and the observed measurements across various sub-basins. The correction factor is computed using the following equation:

$$G_{correction} = \frac{\sum_{i=1}^{n} \mathrm{A}i * P_{observedi}}{\sum_{i=1}^{n} \mathrm{A}i * P_{ERA5Landi}} \qquad (9)$$

where:

- $G_{correction}$ is the global homogeneous correction factor.
- n is the number of sub basins in the Secchia River Basin.
- $P_{observed,i}$ is the observed precipitation for sub-basin *i*
- $P_{ERA5\text{-}Land,\, i}$ is the ERA5-Land precipitation for sub-basin *i*
- A_i is Area of sub-basin *i*

The correction factor computed using the above equation is equal to 0.87. This correction factor is applied uniformly across all sub-basins to ensure that the reanalysis precipitation data are consistently adjusted. The corrected precipitation value is calculated as follows.

$$P_{corrected} = P_{ERA5Land} * G_{Correcttion} \qquad (10)$$

where: $P_{corrected}$ is the adjusted precipitation value and $P_{ERA5\text{-}Land}$ is the original ERA5-Land precipitation value.

By using the Global Homogeneous Correction Factor, the ERA5-Land precipitation data are aligned more closely with observed precipitation data, effectively correcting for any overall bias. This adjustment improves the accuracy of the precipitation input used in hydrological simulations, leading to more reliable model outputs and better representation of hydrological conditions in the Secchia River Basin.

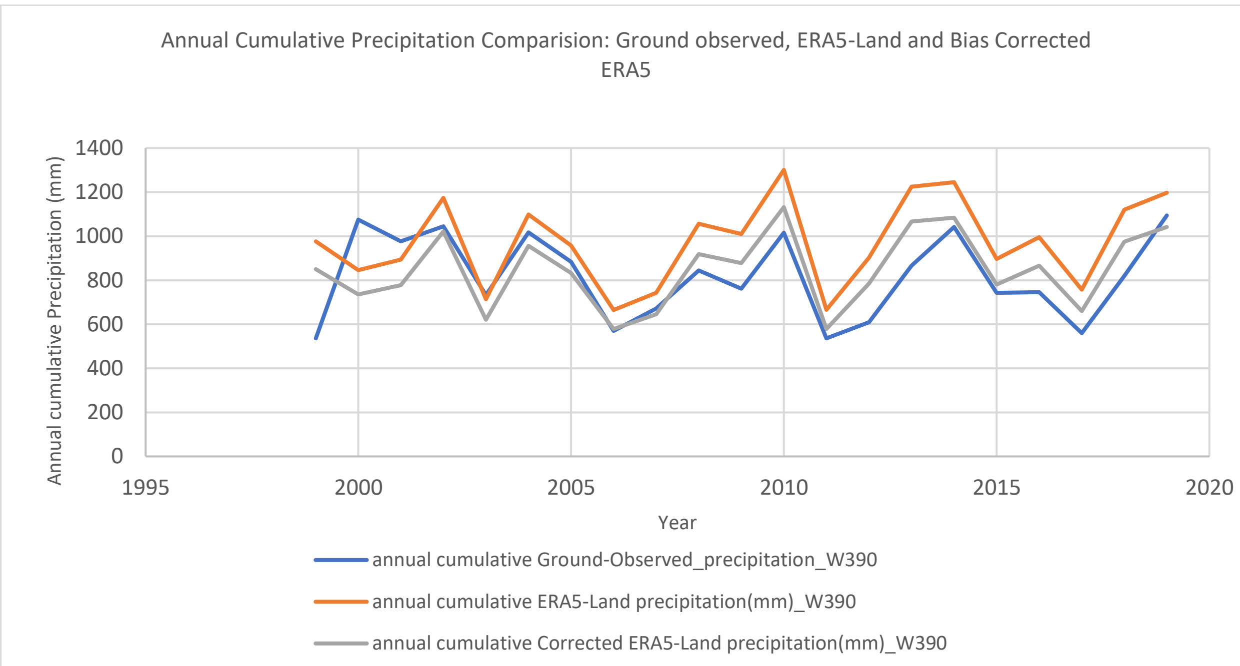

Figure 10a: Annual Cumulative Precipitation for Sub-Basin W390 (1999–2019)

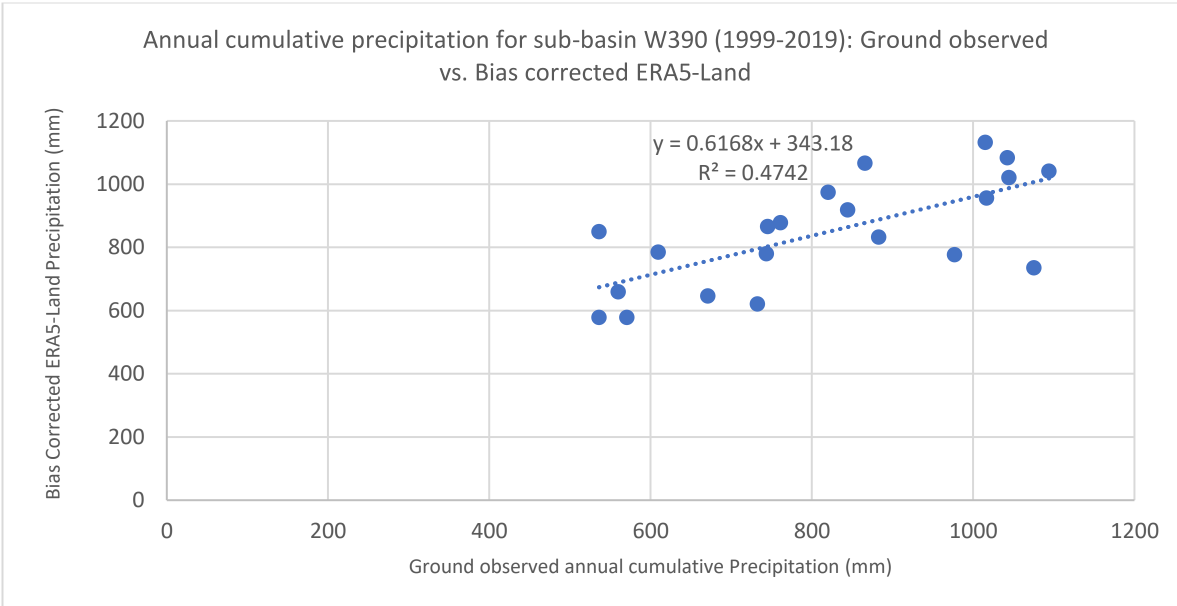

Figure 10b: Linear regression analysis for Sub-Basin W390 (1999–2019)

Figure 10a illustrates the relationship between annual cumulative precipitation data from 1999 to 2019 for three different sources, ERA5-land, ground-observed and corrected ERA5-Land precipitation. In most of the years, ERA5-Land precipitation estimates are higher than ground-observed values. The corrected ERA5-Land values provide a closer match to the ground-observations, reducing the overestimation illustrated in the original ERA5-Land data.

Figure 10b shows the linear regression analysis between ground-observed precipitation and corrected ERA5-Land precipitation. The regression equation shows the tendency of corrected ERA5-land underestimate higher precipitation values in comparison to the ground observations.

4.2 Hydrological Model Setup

The Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) was used to simulate the hydrological responses within the Secchia River Basin. HEC-HMS, its robust capabilities for representing different meteorological conditions and simulating hydrological processes make it an ideal choice for basin-scale water resource assessments, flood management and long-term hydrological analysis.

4.2.1 HEC-HMS Simulation

The HEC-HMS model was used to simulate continuous hydrological processes in the Secchia River Basin over a 20-year period, from January 1, 2000, to December 31, 2019. However, since the model employs the Soil Moisture Accounting (SMA) approach, it requires an initialization period to accurately set soil moisture and other hydrological storage conditions. To achieve this, precipitation data from January 1, 1999, to December 31, 1999, was used as input to ensure proper initialization of the model state variables before the main simulation period began.

This initialization period allows the model to account for essential processes such as soil infiltration, evapotranspiration and runoff, which influence hydrological conditions at the start of 2000. Without this initialization phase, the simulation for 2000 and subsequent years could be affected by arbitrary initial conditions. Including data from 1999 ensures that the results reflect realistic hydrological dynamics, providing a solid foundation for accurate simulations starting in 2000.

4.2.1.1 Simulation Scenarios and Timeframes

This study conducted multiple simulation scenarios to assess the impact of different precipitation datasets on hydrological modeling in the Secchia River Basin. The key scenarios included running the HEC-HMS model using two different precipitation inputs: ground-observed data and ERA5-Land reanalysis data. The simulation period covered from January 1, 2000, to December 31, 2019, with 1999 serving as an initialization year to stabilize the model state variables using the Soil Moisture Accounting (SMA) method.

The scenarios examined the hydrological response under varying meteorological conditions, enabling a comparison between the two datasets in terms of discharge, infiltration, runoff and other hydrological components. Both event-based and long-term hydrological performance, including annual trends, were analyzed. This comparison provided insights into the dataset's effectiveness in simulating hydrological behavior across different timeframes.

4.3 Calibration and Validation

The calibration and validation process for the HEC-HMS model is crucial for ensuring the accuracy and reliability of hydrological simulations. In this study, the focus is on evaluating the model performance with different precipitation datasets rather than recalibrating the model parameters.

4.3.1 Existing Calibrated Model

The hydrological model used in this study is based on an existing calibrated HEC-HMS model developed for the Secchia River Basin by HydroNova. Calibration refers to the process of adjusting model parameters to optimize the fit between observed and simulated streamflow data. The model was previously calibrated using historical ground-observed precipitation and discharge data.

For this study, no additional recalibration is performed. The existing calibrated model serves as a baseline, allowing for a direct comparison of model performance under different precipitation scenarios without altering the calibrated parameters. This approach ensures consistency in evaluating how different precipitation datasets affect the model ability to simulate streamflow.

4.3.2 Validation Process

Validation was carried out to evaluate the model performance in replicating observed streamflow at the Rubiera station, which serves as the outlet for the study area. The validation process involved comparing simulated streamflow results with actual discharge measurements from this station. This validation ensures that the model is capable of accurately simulating streamflow based on the existing calibration and evaluating the impact of different precipitation datasets on model performance.

4.4 Performance Evaluation Metrics

In this study, a combination of event-scale and long-term evaluation metrics is used to assess the performance of the HEC-HMS model in simulating hydrological processes under different precipitation datasets. These metrics enable a comprehensive assessment of the model’s ability to replicate observed hydrological behaviors accurately. The study employs two primary metrics: Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE), which are applied in different contexts to capture both event-specific and long-term model performance.

4.4.1 Nash-Sutcliffe Efficiency (NSE)

The Nash-Sutcliffe Efficiency (NSE) is a widely used metric that measures the degree of agreement between simulated and observed discharge values, providing an overall indication of model performance. It is defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{obs,i} - Q_{sim,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{obs,i} - \bar{Q}_{obs,i}\right)^2} \qquad (11)$$

Where $Q_{obs,i}$ is the observed discharge at time step i, $Q_{sim,i}$ is the simulated discharge at time step i, $\bar{Q}_{obs}$ is the mean observed discharge and n is the total number of observations.

An NSE value of 1.0 indicates a perfect match between the observed and simulated values, while values below zero suggest that the model predictions are less accurate than using the mean of the observed data as a predictor. In this study, the NSE is applied in event-scale evaluation and long-term and annual evaluation.

In the event-scale evaluation, the focus is on assessing the model's ability to accurately simulate specific hydrological events. In this study, the accuracy of peak flows and the timing of peak flows are evaluated to determine how well the model replicates the dynamics of individual, high-intensity events. Peak flow accuracy is assessed to measure how closely the simulated peak flows match the observed magnitudes during significant hydrological events. This metric is crucial for understanding the model's performance in capturing extreme precipitation events. Additionally, the time to peak metric evaluates how precisely the model predicts the timing of peak flows, which is essential for assessing the model's capability to simulate the temporal dynamics of hydrological responses accurately.

In the long-term and annual evaluation, NSE is also employed to evaluate the overall model performance across the entire simulation period. This context helps gauge the model's effectiveness in representing long-term water balance and streamflow variability within the basin. By capturing the cumulative hydrological response over diverse conditions, the NSE helps validate the model's suitability for long-term water resource assessments.

4.4.2 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) measures the average magnitude of deviations between observed and simulated values, reflecting the model's predictive accuracy. It is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Q_{obs,i} - Q_{sim,i}\right)^2}{n}} \qquad (12)$$

where $Q_{obs,i}$ is the observed discharge, $Q_{sim,i}$ is the simulated discharge and n is the total number of observations.

Lower RMSE values indicate a closer fit between the observed and simulated data, with zero representing a perfect match. In this study, RMSE is primarily used for long-term evaluations to quantify the overall error in model predictions over the entire simulation period. By capturing the magnitude of prediction errors, RMSE serves as a reliable indicator of the model's accuracy in representing the overall hydrological dynamics within the basin.

4.5 Precipitation Partitioning

4.5.1 Method for Partitioning Precipitation

The ability of the model to partition precipitation into surface runoff, infiltration and baseflow is essential for understanding water movement and distribution within the Secchia River Basin. This analysis provides valuable insights into the hydrological dynamics of the study area.

In this section, precipitation partitioning is performed using output from the HEC-HMS model, specifically utilized the Soil Moisture Accounting (SMA) method. The SMA approach continuously tracks the movement of water through various hydrological processes, including infiltration, runoff and baseflow over a 20-year period (2000-2019). This method facilitates a comprehensive analysis of how precipitation is allocated within each sub-basin of the Secchia River Basin.

The analysis was carried out for both ground-observed data and ERA5-Land reanalysis data, allowing for a direct comparison of how each dataset influenced the partitioning of precipitation. This analysis used to assess the accuracy of ERA5-Land data in capturing key hydrological processes in comparison to ground-observed data. This comparison also evaluated the reliability of ERA5-Land reanalysis under varying meteorological conditions, particularly where ground data was sparse or unavailable.

By incorporating the water balance equation and examining the partitioning of precipitation across different datasets, this analysis provided critical insights into the hydrological behavior of the Secchia River Basin and the effectiveness of reanalysis data in representing these processes.

4.6 Mean Streamflow Calculation

The mean streamflow is a fundamental metric used to quantify the average discharge rate of a river over a specified period. It serves as a key indicator for understanding overall hydrological responses, evaluating water availability and informing water resource management decisions. In this study, mean streamflow calculations were performed to assess the hydrological performance of the HEC-HMS model and to determine the accuracy of various precipitation datasets for simulating streamflow within the Secchia River Basin.

For this analysis, three precipitation datasets were utilized: ground-observed precipitation, ERA5-Land reanalysis data and bias-corrected ERA5-Land reanalysis data. The objective was to compare the ability of these datasets to replicate the observed hydrological dynamics in the basin. The mean streamflow (Q_{mean}) was calculated using hourly discharge data over the main simulation period from January 1, 2000, to December 31, 2019. The following formula was used for the calculation:

$$Q_{mean} = \frac{1}{N}\sum_{i=1}^{N} Qi \qquad (13)$$

where Q_{mean} represents the mean streamflow (m^3/s), N is the total number of hourly time steps within the simulation period and Qi denotes the discharge value at each time step i (m^3/s).

The analysis included both model-simulated discharge values and observed discharge measurements to provide a comparative assessment of the model's accuracy. Although the model setup required an initialization period from January 1, 1999, to December 31, 1999, to establish initial soil moisture and other hydrological conditions, this initialization period was excluded from the mean streamflow calculation. Focusing only on the main simulation period from 2000 to 2019 ensured that the results accurately reflect the hydrological conditions.

The mean streamflow values derived from different precipitation datasets, such as ground-observed precipitation, ERA5-Land reanalysis data and corrected ERA5-Land reanalysis data, were compared to evaluate the model’s performance under varying input conditions. This comparison is crucial for understanding the model's sensitivity to different precipitation inputs and for identifying the most reliable dataset for simulating hydrological processes within the basin. Accurate mean streamflow representation is essential for effective water resource planning, flood management and sustainable development within the Secchia river basin.

5 Results and Discussion

Here I compare with observations discharge simulations obtained by using ground-observed precipitation data, ERA5-Land reanalysis precipitation data, and bias-corrected ERA5-Land reanalysis precipitation data. Key aspects of the analysis include continuous simulation results, performance metrics computed on flood event simulations and long-term metrics, the impact of bias correction on reanalysis data and the evaluation of mean streamflow and precipitation partitioning.

5.1 Model Simulation Results using Different Input Data

In this section the results of the simulations are compared with Rubiera station serving as the outlet for comparison and validation.

5.1.1 Discharge Simulations Results Using Ground-Observed Precipitation Data

Figure 11a shows the observed discharge hydrograph compared with simulations between 2000 and 2019. It is important to note the absence of observed discharge data for some years as specified in section 4.1.1. These data gaps limit the validation of the model's performance during these periods and were excluded from the Nash-Sutcliffe Efficiency (NSE) calculations to ensure an accurate assessment of the model's fit.

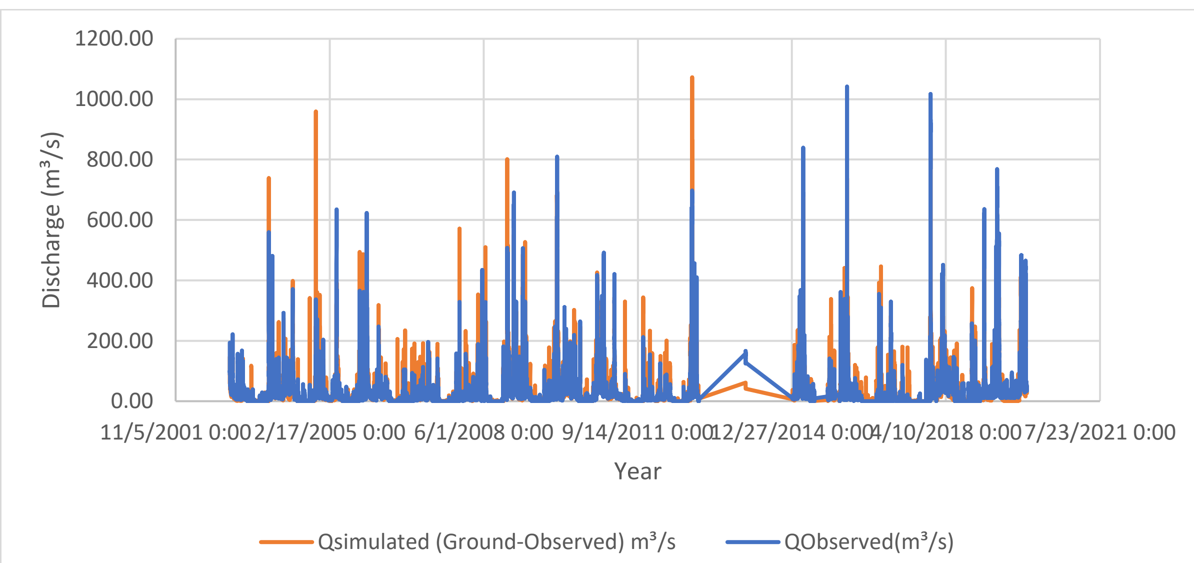

Figure 11a: Observed Discharge and Simulated Discharge Hydrographs Using Ground-Observed Precipitation Data at Rubiera Station

The scatter plot in figure 11b indicates a good fit between observed discharge and ground observed simulated discharges, with the model successfully capturing major trends and hydrological events. As demonstrated by R^2 value of 0.75. This suggests that 75% of the variability in the observed discharge can be explained by the model.

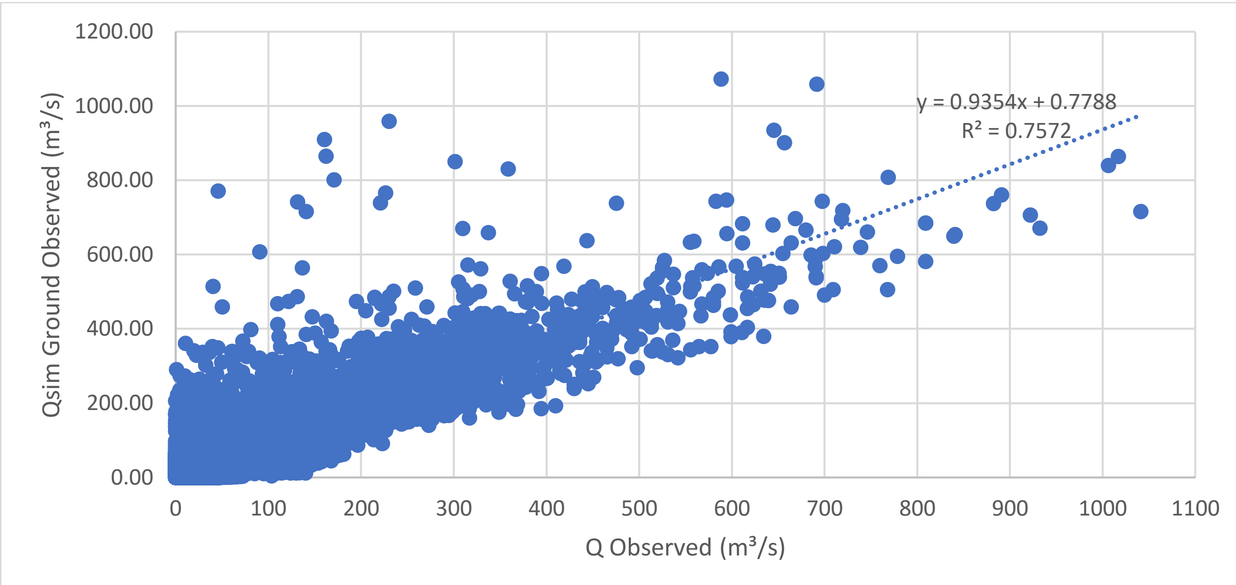

Figure 11b: Scatter Plot of Observed Discharge and Simulation Discharge Using Ground-Observed Precipitation Data at Rubiera Station

To evaluate the performance, Nash-Sutcliffe Efficiency (NSE) was calculated using only the available observed discharge data, yielding an NSE of 0.71. This value indicates a good fit between simulated and observed discharge where data was present.

5.1.2 Discharge Simulations Results Based on ERA5-Land Reanalysis Precipitation Data

In addition to ground-observed precipitation, the hydrological model was run using ERA5-Land reanalysis precipitation data for the same period (2000–2019).

Figure 12a illustrates the discharge hydrographs at the Rubiera station, comparing simulated discharge using ERA5-Land precipitation data with observed discharge from 2000 to 2019.

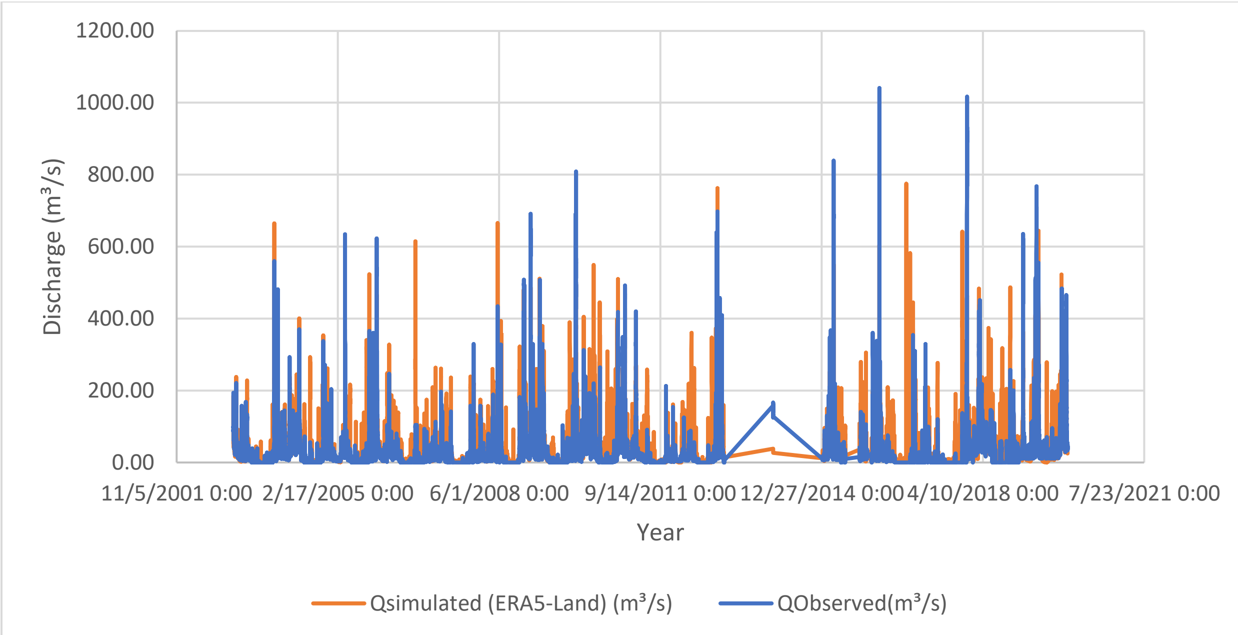

Figure 12a: Observed Discharge and Simulated Discharge Hydrographs Using ERA5 Precipitation Data at Rubiera Station

The hydrographs show that while the model captures some general trends in discharge, there are notable discrepancies during significant hydrological events. Specifically, the model tends to underestimate peak discharges, indicating challenges in accurately simulating high-flow conditions. These discrepancies between the observed discharge and simulated discharge using ERA5-Land Precipitation data is evident in the scatter plot in Figure 12b.

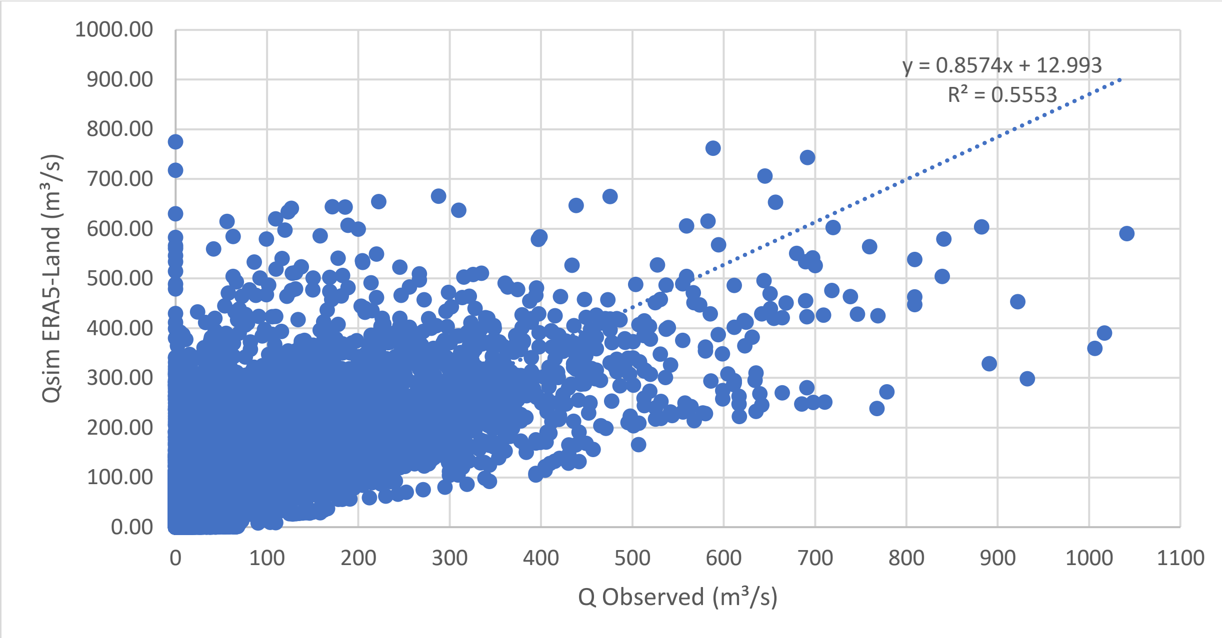

Figure 12b: Scatter Plot of Observed Discharge and Simulated Discharge Using ERA5-Land Precipitation Data at Rubiera Station

To assess the model performance, the Nash-Sutcliffe Efficiency (NSE) was calculated. The NSE value obtained for the ERA5-Land-based simulations was 0.34, indicating a moderate fit between the simulated and observed discharge data. This suggests that while the reanalysis data captures some variability in discharge, its performance is weaker compared to that achieved with ground-observed precipitation data.

The moderate NSE highlights the limitations of using reanalysis data for hydrological modeling. This finding emphasizes the importance of validating model outputs with reliable ground-observed data to enhance the accuracy of hydrological simulations. While the ERA5-Land reanalysis precipitation data provide valuable insights into the hydrological dynamics of the Secchia River Basin, its ability to simulate discharge accurately is limited.

5.1.3 Discharge Simulations Results Based on Bias-Corrected ERA5-Land Reanalysis Precipitation Data

Figure 13a displays the discharge hydrographs at the Rubiera station, comparing simulated discharge derived from bias-corrected ERA5-Land reanalysis precipitation data obtained using GHCF described in section 4.14 with observed discharge over the same period.

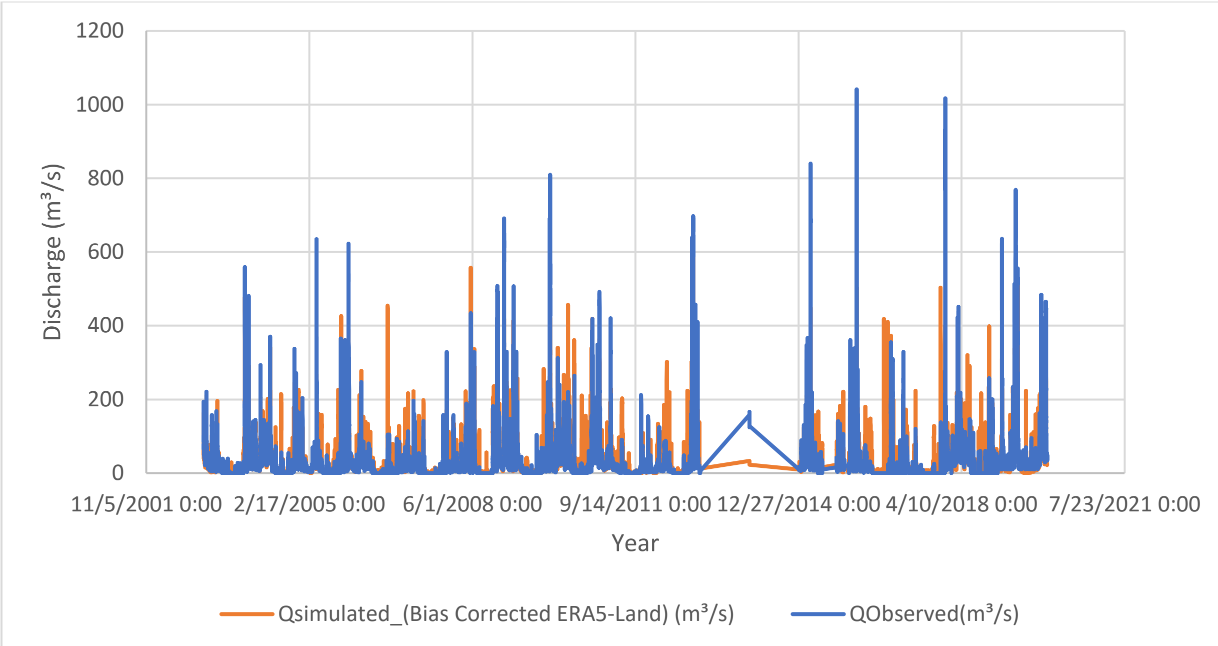

Figure 13a: Observed Discharge and Simulated Discharge Hydrographs Using Bias-Corrected ERA5-Land Reanalysis Precipitation Data at Rubiera Station

In the scatter plot in Figure 13b, the linear regression relationship (R^2) value of 0.59 indicates that approximately 59% of the variance in observed discharge is explained by the bias-corrected simulation. In comparison, R^2 value from the uncorrected ERA5-Land simulation was 0.55, indicating that only 55% of the variance in observed discharge was explained by the model. The increase in R^2 from 0.55 to 0.59 demonstrates that the application of GHCF bias correction has improved the accuracy of the simulated discharge. However, despite the improvements, some differences remain during extreme high-flow events, indicating that more advanced bias correction methods or additional refinement may be required to capture peak discharges more accurately.

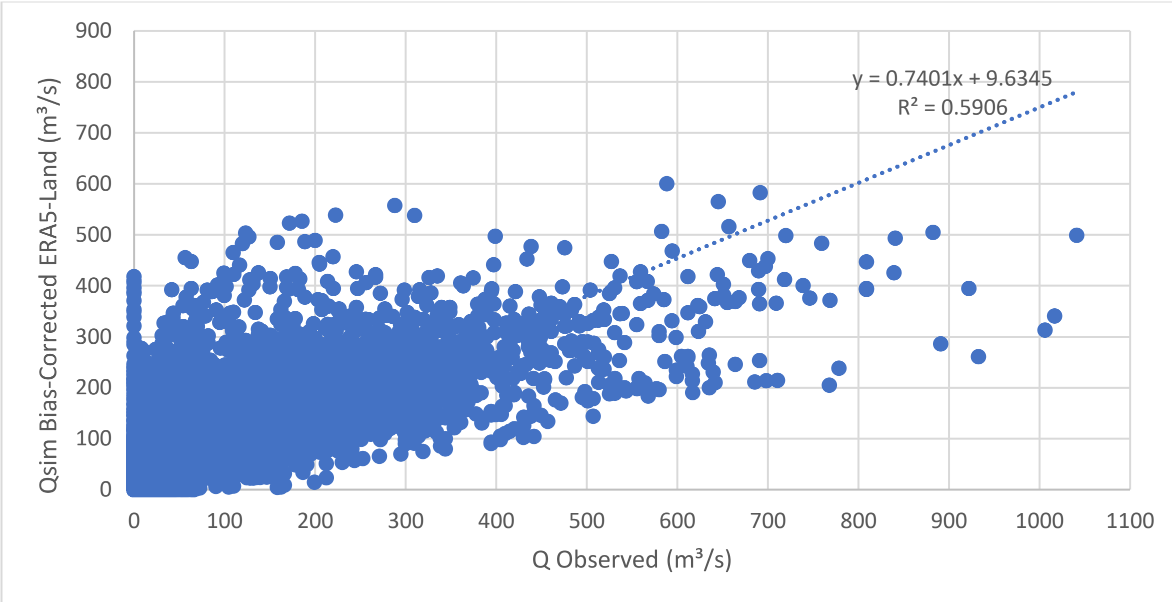

Figure 13b: Scatter Plot of Observed Discharge and Simulated Discharge Using Bias-Corrected ERA5-Land Reanalysis Precipitation Data at Rubiera Station

Before applying bias correction, the discharge simulations based on ERA5-Land reanalysis data resulted in a Nash-Sutcliffe Efficiency (NSE) of 0.34. Following bias correction, the NSE increased to 0.55, indicating a moderate fit between simulated and observed discharge data. These results highlight the critical role of precipitation data quality in hydrological modeling; ground-observed precipitation data achieved a higher NSE of 0.71, demonstrating its reliability for discharge simulations.

While the initial NSE for ERA5-Land data was lower due to limitations in capturing localized precipitation events, the application of the GHCF technique significantly improved its performance. This underscores the potential of global reanalysis data as a valuable resource when corrected for systematic errors.

Overall, the findings suggest that while ERA5-Land reanalysis data can be used for discharge simulations, bias correction is essential to enhance its accuracy. Ground-observed precipitation remains the preferred input for hydrological models when available. However, bias-corrected reanalysis data serves as a possible alternative in data sparse regions or for long-term studies where observed data may be limited.

5.2 Performance Metrics

This section provides a comprehensive analysis of the performance metrics used to evaluate the discharge simulations for the Secchia River Basin. The focus is on assessing model performance at different temporal scales, including event-scale performance and annual or long-term metrics. This multi-scale evaluation helps to capture both short-term hydrological responses during extreme events and the model's ability to reproduce seasonal or interannual discharge patterns.

5.2.1 Event-Scale Performance

Event-scale performance metrics are critical for evaluating the model's ability to simulate rapid hydrological responses, such as during floods or high-flow events, which are crucial for effective flood management and early warning systems. For this study, event-scale performance was assessed by comparing the simulated discharge to observed discharge during notable flood events and extreme precipitation periods. This analysis focuses on four notable extreme events within the Secchia River Basin to evaluate how well the hydrological model replicates observed discharge characteristics under these conditions. The selected events are December 2009, March 2015, February 2016 and December 2017. These events were chosen for their significant hydrological impact, providing a robust benchmark to test the model's capability in simulating both the magnitude and timing of extreme discharges. To assess the accuracy of the simulations, discharge data from these events are compared against observed discharge measurements, with a focus on peak discharge values and timing.

The subsequent figures and tables provide a comparative analysis of the model's performance using ground-observed precipitation and ERA5-Land reanalysis datasets to capture variations in event-scale discharge simulations.

December 2009 Extreme Event

The December 2009 extreme hydrological event presented multiple peak flows, making it a challenging case for evaluating the model's performance. This event was characterized by two distinct peaks within a short period, providing an opportunity to assess how well the model captures both timing and magnitude under complex hydrological conditions. The discharge simulations were conducted using ground-observed precipitation, ERA5-Land reanalysis precipitation and bias-corrected ERA5-Land reanalysis precipitation data.

The figure 14 shows the discharge hydrographs for the extreme hydrological event of December 2009 at the Rubiera station. The figure highlights two prominent peaks occurring on December 23 and December 25, providing a detailed comparison of how each dataset captures these key hydrological features.

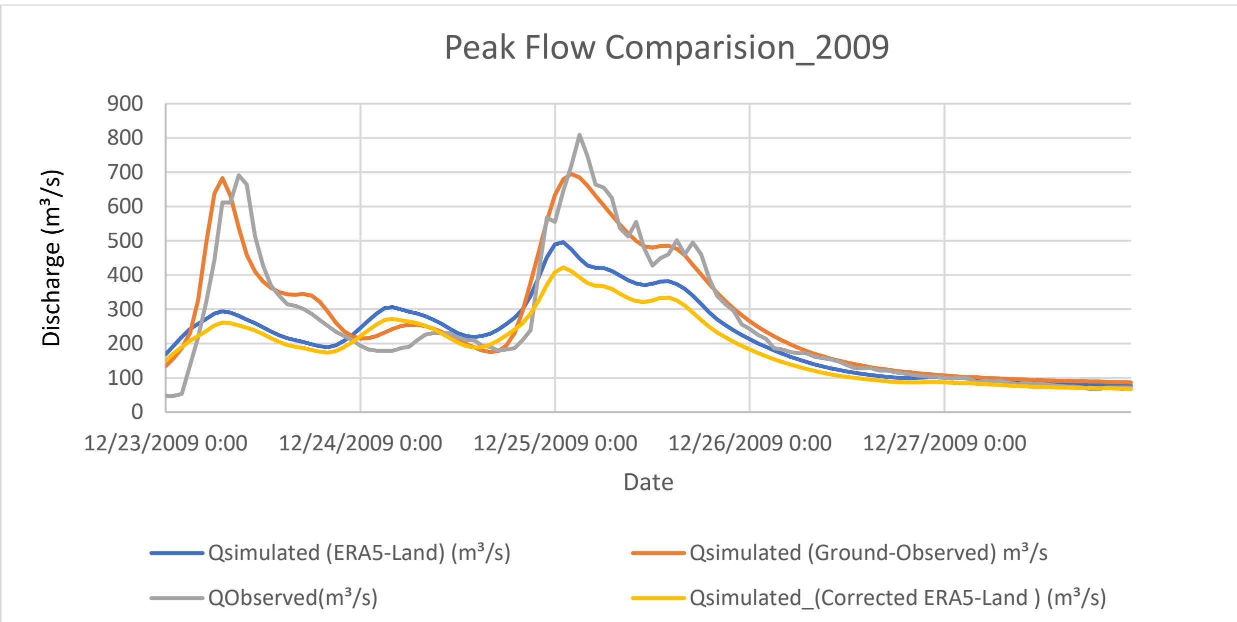

Figure 14: Discharge Simulation Hydrographs for the December 2009 Extreme Event

Table 4 summarizes the peak discharge values and timing for both the observed and simulated flows during this event. The first peak occurred on December 23, 2009, at 10:00, with an observed discharge of 664.1 m^3/s. The simulation using ground-observed precipitation data closely captured this peak, predicting a discharge of 682.7 m^3/s at 7:00, resulting in a small relative error of 2.8%. However, the ERA5-Land reanalysis simulation significantly underestimated the first peak, yielding a discharge of 294.2 m^3/s and a higher relative error of 55.7%, also predicting the peak earlier, at 7:00. The Bias corrected ERA5-Land simulation further reduced the peak discharge to 261.3 m^3/s, and higher relative error of 60.1%.

Table 4: Summary of Peak Discharge and Peak Time During the December 2009 Extreme Event

Metric	Observed Flow	Simulated Flow (Observed Precipitation)	Simulated Flow (ERA5-Land Reanalysis)	Simulated Flow (ERA5-Land Corrected)
Peak 1 Discharge (m^3/s)	664.1	682.7	294.2	261.3
Peak 1 Time	23-Dec-2009,10:00	23-Dec-2009,7:00	23-Dec-2009,7:00	23-Dec-2009, 7:00
Peak 2 Discharge (m3/s)	809	694.6	495.5	421.5
Peak 2 Time	25-Dec-2009, 03:00	25-Dec-2009, 02:00	25-Dec-2009, 01:00	25-Dec-2009 1:00
Nash-Sutcliffe Efficiency (NSE)	-	0.92	0.62	0.49

The second peak, which occurred on December 25, 2009, at 03:00, had an observed discharge of 809 m^3/s. The ground-observed precipitation data simulation predicted a peak of 694.6 m^3/s, with a relative error of 14.14%, occurring at 2:00, one hour earlier than observed. The ERA5-Land reanalysis simulation further underestimated this second peak, producing a discharge of 495.5 m^3/s, leading to a relative error of 38.7%, and predicted the peak two hours earlier, at 1:00. The Bias corrected ERA5-Land simulation lowest peak 412.5 m^3/s, peak at 1:00.

In terms of peak timing, the observed peak occurred at 03:00 on December 25, 2009. The simulation using ground-observed precipitation data predicted the peak at 02:00, just one hour earlier, while the ERA5-Land reanalysis simulation and Bias corrected ERA5-Land predicted the peak at 01:00, two hours earlier. These timing discrepancies reflect the differences in precipitation intensity and distribution captured by each dataset.

The Nash-Sutcliffe Efficiency (NSE) values further reflect the model's ability to replicate observed discharge patterns during this event. For the second peak, the simulation using ground-observed precipitation achieved an NSE of 0.92, indicating a very strong fit between the simulated and observed discharge values. This suggests that the model effectively captures the hydrological dynamics of the event using ground-based measurements. while the ERA5-Land-based simulation obtained an NSE of 0.62, reflecting a moderate fit and highlighting the limitations of using reanalysis data to simulate peak discharges. The Bias corrected showed a moderate fit, with an NSE of 0.49.

March 2015 Extreme Event:

The extreme hydrological event of March 2015 is significant due to its impact on the Secchia River basin, making it an essential case for evaluating the performance of discharge simulations.

Figure 15 illustrates the discharge hydrographs for the March 2015 extreme event at Rubiera station, the outlet for our study area. The figure highlights how closely the simulations align with the observed discharge, particularly during critical periods.

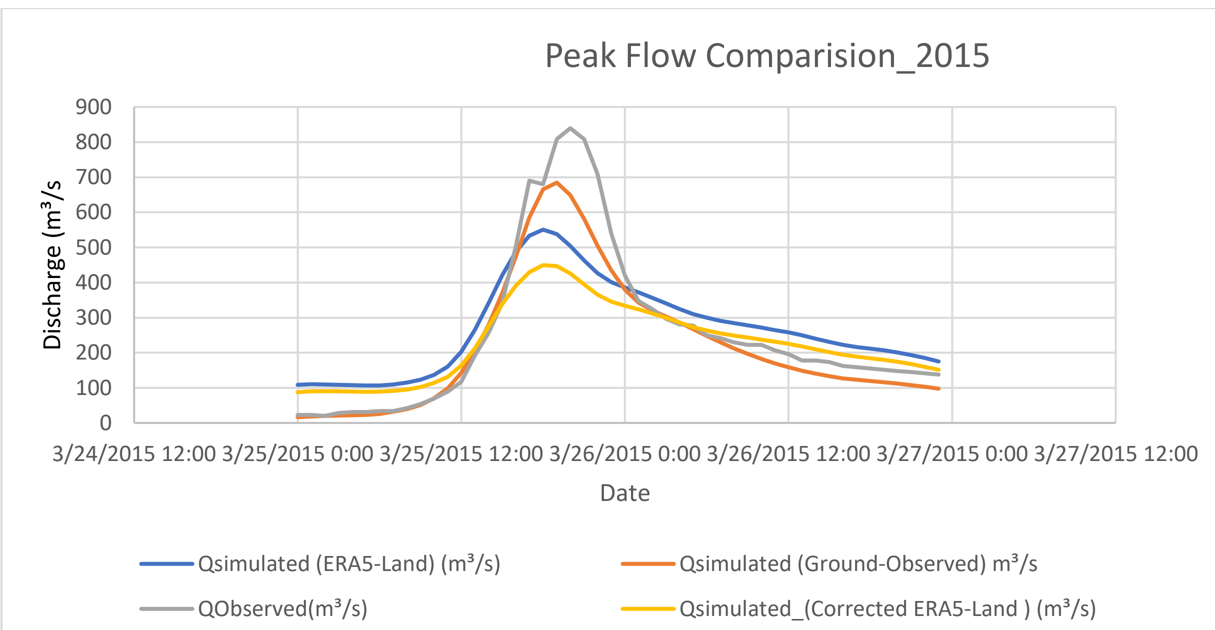

Figure 15: Discharge Hydrographs for the Extreme Event of March 2015

Table 5 presents a comparison of peak discharge and peak timing for the extreme event that occurred in March 2015 at Rubiera station. The observed peak discharge was 839.4 m³/s. The simulation based on ground-observed precipitation data resulted in a peak discharge of 684.8 m³/s, while the simulation using ERA5-Land reanalysis data produced a peak discharge of 550.7 m³/s.

In terms of peak timing, the observed peak occurred at 20:00 on March 25, 2015. The peak times predicted by the simulations were earlier, with the ground-observed data forecasting the peak at 19:00 and the ERA5-Land data forecasting it at 18:00. The Peak discharge for the Bias corrected ERA5-Land is 449.2 m^3/s, which is lower among all the simulated flows and significantly lower than observed flow of 839.4 m^3/s.

Table 5: Summary of Peak Discharge and Peak Time During the March 2015 Extreme Event

Metric	Observed Flow	Simulated Flow (Observed Precipitation)	Simulated Flow (ERA5-Land Reanalysis)	Simulated Flow (ERA5-Land Corrected)
Peak Discharge (m3/s)	839.4	684.8	550.7	449.2
Peak Time	25-Mar-2015, 20:00	25-Mar-2015, 19:00	25-Mar-2015, 18:00	25-Mar-2015, 18:00
Nash-Sutcliffe Efficiency (NSE)	-	0.92	0.75	0.66

The Nash-Sutcliffe Efficiency (NSE) for the event was 0.92 for the simulation based on ground-observed precipitation data, indicating a strong fit between simulated and observed discharge. In contrast, the ERA5-Land reanalysis data yielded an NSE of 0.75, suggesting a moderate fit. These NSE values indicate that the model performed significantly better with ground-observed precipitation data, closely aligning with both the peak discharge and timing of the observed data. The Nash-Sutcliffe Efficiency (NSE) for the Bias corrected ERA5-Land is 0.66, indicating a moderate fit between the simulated and observed flows.

February 2016 Extreme Event:

The February 2016 extreme hydrological event is notable for its significant impact on the Secchia river basin, marking it as an important case for evaluating model performance.

Figure 16 displays the discharge hydrographs for the February 2016 extreme event at the Rubiera station. The figure reveals the performance of the model in capturing discharge dynamics during this critical period.

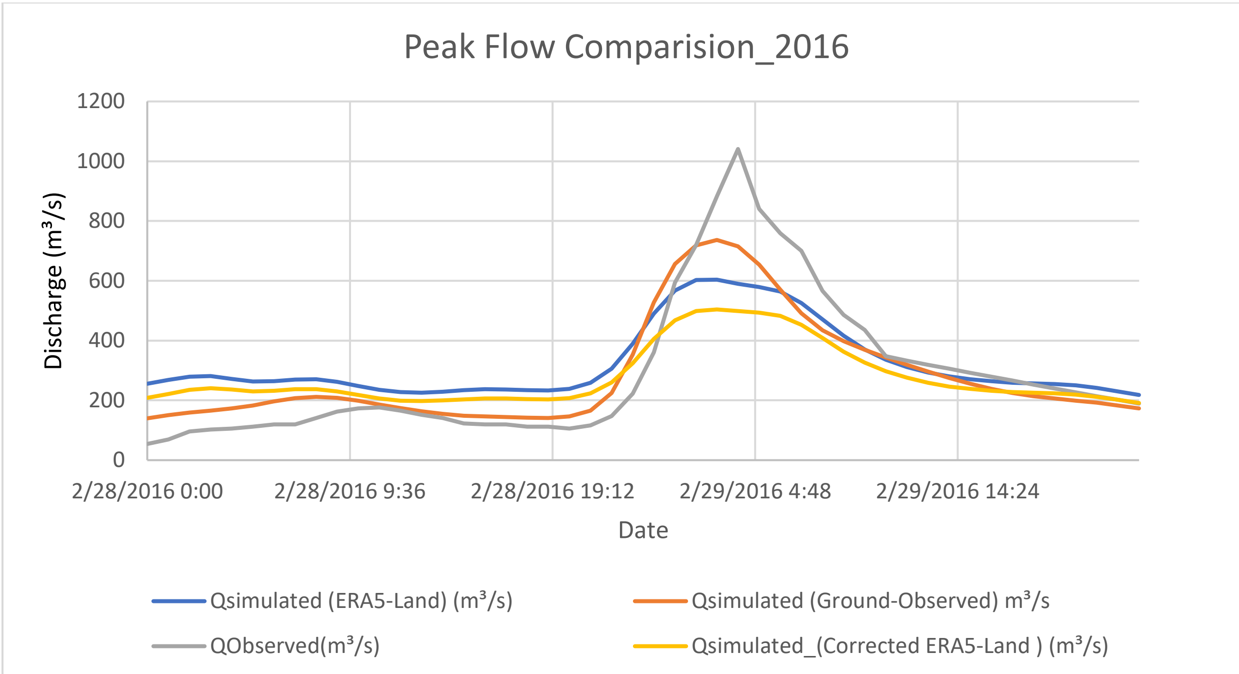

Figure 16: Discharge Hydrographs for the Extreme Event of February 2016

During the February 2016 extreme hydrological event, the observed peak discharge reached 1041.3 m³/s. The simulation based on ground-observed precipitation estimated a peak discharge of 737.1 m³/s, while the ERA5-Land reanalysis data predicted a peak of 603.8 m³/s. Both simulations underestimated the observed peak discharge, with the ERA5-Land data showing a larger discrepancy. The Bias corrected ERA5-Land provided the lowest discharge estimate of 504.3 m³/s, with the peak time at 3:00.

Table 6: Overview of Peak Discharge and Timing for the February 2016 Extreme Event

Metric	Observed Flow	Simulated Flow (Observed Precipitation)	Simulated Flow (ERA5-Land Reanalysis)	Simulated Flow (ERA5-Land Reanalysis)
Peak Discharge (m3/s)	1041.3	737.1	603.8	504.3
Peak Time	29-Feb-2016, 04:00	29-Feb-2016, 03:00	29-Feb-2016, 03:00	29-Feb-2016, 3:00
Nash-Sutcliffe Efficiency (NSE)	-	0.85	0.66	0.6

The observed peak occurred at 4:00 AM on February 29, 2016, while both models predicted the peak to occur an hour earlier at 3:00 AM. The Nash-Sutcliffe Efficiency (NSE) for the ground-observed precipitation simulation was 0.85, indicating strong performance in simulating the event. In contrast, the ERA5-Land simulation produced an NSE of 0.66, suggesting a moderate level of accuracy in replicating the observed discharge. The bias corrected ERA5-land simulation had an NSE of 0.6, indicating a moderate fit between observed discharge and the simulated discharge.

December 2017 Extreme Event:

The December 2017 extreme hydrological event stands out as the highest peak discharge recorded during the entire study period, making it a crucial case for evaluating the model's performance under severe conditions. The implications of accurately modeling such high-flow events are significant for flood management and water resource planning in the Secchia River basin.

Figure 17 illustrates the discharge hydrographs for the December 2017 event at Rubiera station. The figure highlights the discrepancies between observed and simulated values, particularly during peak flows.

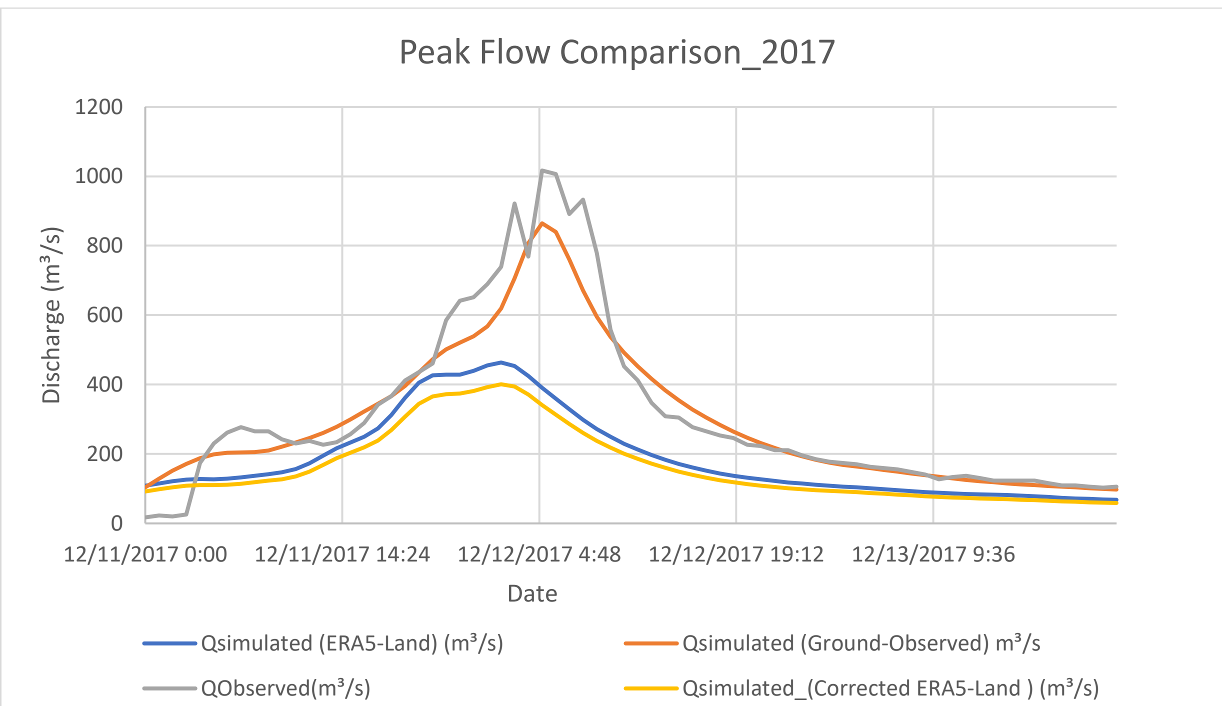

Figure 17: Hydrograph of December 2017 Extreme Event at Rubiera Station

Table 7 presents the peak discharge and peak time data for the extreme hydrological event that occurred on December 12, 2017. The observed peak discharge was 1016.9 m^3/s. The model simulations yielded different results: the simulation based on ground-observed precipitation data estimated a peak discharge of 864.1 m^3/s, whereas the simulation using ERA5-Land reanalysis data significantly underestimated the peak discharge at 463.4 m^3/s. Both simulations based on ground-observed precipitation and the actual observations identified the peak time as 5:00 AM on December 12, 2017. In contrast, the ERA5-Land simulation predicted the peak time as 2:00 AM on the same day. The bias corrected ERA5-Land showed the lowest peak discharge estimate of 400.5 m^3/s, the peak time 2:00 AM on December 12, 2017.

Table 7: Peak Discharge and Peak Time Data for the December 2017 Extreme Hydrological Event

Metric	Observed Flow	Simulated Flow (Observed Precipitation)	Simulated Flow (ERA5-Land Reanalysis)	Simulated Flow (ERA5-Land Reanalysis)
Peak Discharge (m3/s)	1016.9	864.1	463.4	400.5
Peak Time	12-Dec-2017, 05:00	12-Dec-2017, 05:00	12-Dec-2017, 02:00	Dec-12-2017, 2:00
Nash-Sutcliffe Efficiency (NSE)	-	0.91	0.35	0.19

The Nash-Sutcliffe Efficiency (NSE) values for this event reflect the performance of each simulation: the simulation based on ground-observed precipitation achieved a high NSE of 0.91, indicating a strong fit with the observed data. In comparison, the ERA5-Land simulation had an NSE of 0.35, demonstrating a lower level of accuracy in replicating the observed discharge during this event. The Bias corrected ERA5-Land showing the lowest NSE of 0.19. This value indicates poor fit between bias corrected ERA5-Land simulated peak and observed peak discharge.

5.2.2 Annual Performance Metrics

In this section, annual performance metrics are evaluated to assess the effectiveness of different precipitation datasets in simulating streamflow within the Secchia river basin. The purpose of this analysis is to compare the accuracy of various datasets, specifically Ground-Observed, ERA5-Land reanalysis and Bias-Corrected ERA5-Land over a 20-year period, from 2000 to 2019. This comparison aimed to identify the strengths and limitations of each dataset, thereby informing future hydrological modeling efforts.

The Nash-Sutcliffe Efficiency (NSE) metric is employed to evaluate annual performance. NSE values closer to 1 indicate strong agreement between observed and simulated data, while values closer to zero or negative reflect poor performance. Table 7 presents the results for each dataset over the study period, providing a comprehensive overview of their reliability in replicating observed streamflow patterns.

Table 8: Annual Performance Metrics for the Study Period

Year	NSE based on Ground-Observed Simulation	NSE based on ERA5-Land Simulation	NSE based Bias-Corrected Simulation
(2000-2002)	-	-	-
2003	0.68	0.67	0.79
2004	0.3	0.34	0.62
2005	0.76	0.34	0.54
2006	0.37	-1.15	-0.13
2007	0.41	-0.62	0.11
2008	0.65	0.18	0.53
2009	0.89	0.61	0.71
2010	0.70	0.31	0.56
2011	0.45	-0.14	0.31
2012	0.64	0.15	0.40
2013-2014	-	-	-
2015	0.57	0.42	0.56
2016	0.68	-0.35	0.17
2017	0.80	0.20	0.42
2018	0.60	0.12	0.40
2019	0.76	0.51	0.52

For the years 2000, 2001, 2002, 2013 and 2014, the NSE values are missing because observed streamflow measurements were unavailable for these periods. For the remaining years, the analysis shows significant variation in model performance across the datasets and years. The Ground-Observed Simulation consistently performed with NSE values exceeding 0.5 in most years, indicating a fairly good fit between simulated and observed streamflow. The highest performance values are recorded in 2009 (NSE = 0.89), 2017 (0.80) and 2005 (0.76), demonstrating the robustness of using ground-based observations for streamflow simulation.

The ERA5-Land Simulation, on the other hand, showed considerable variability in its performance, with negative NSE values in certain years, such as 2006 (-1.15) and 2007 (-0.62), indicating that the ERA5-Land dataset performed poorly in replicating observed streamflow for these years. While positive NSE values in some years, such as 2003 (0.67) and 2009 (0.61), suggest that ERA5-Land was able to capture streamflow patterns, the overall lower values relative to the Ground-Observed Simulation indicate that it is less reliability when used without further correction.

The Bias-Corrected ERA5-Land Simulation generally performed better than the uncorrected ERA5-Land Simulation, demonstrating that the applied corrections substantially improved its alignment with observed data. In most years, the Bias-Corrected dataset showed positive NSE values, even in years where the original ERA5-Land data exhibited negative performance (e.g., 2006 and 2016). Notably, in certain years such as 2003 (0.79), 2004 (0.62) and 2009 (0.71), the Bias-Corrected Simulation even outperformed the Ground-Observed Simulation, highlighting the effectiveness of bias-correction techniques in enhancing the reliability of reanalysis data for streamflow simulation.

Overall, the annual performance metrics indicate that the Ground-Observed Simulation consistently delivered the best performance, demonstrating that ground-based precipitation data is highly reliable for hydrological modeling. On the other hand, the ERA5-Land Simulation showed significant weaknesses in some years. The Bias-Corrected ERA5-Land Simulation showed marked improvements, highlighting its potential as a viable alternative when ground-based data is unavailable.

5.3 Mean Streamflow Analysis

In this section, the mean streamflow values are assessed to evaluate the effectiveness of different precipitation datasets in simulating average streamflow conditions, which is crucial for understanding the hydrological dynamics within the Secchia river basin. This section focuses on the average streamflow rates derived from different precipitation datasets over the study period from 2000 to 2019: Ground-observed data, ERA5-Land reanalysis data and bias-corrected ERA5-Land data.

To assess the effectiveness of these datasets, mean streamflow values were calculated for each year, providing insights into the model's ability to simulate hydrological responses over time and highlighting annual variations and long-term trends. The mean streamflow values for the study period (2000-2019) are summarized in Table 8, which compares observed values alongside simulated outputs from the different datasets. Notably, observed streamflow values are not available for the years 2000, 2001, 2002, 2013, and 2014, while simulated values from the Ground-observed, ERA5-Land and Bias-corrected ERA5-Land datasets were present for all years. The observed streamflow ranged from a minimum of 10.9 m^3/s in 2007 to a peak of 64.8 m^3/s in 2019, showcasing substantial variability across the years.

Table 9: Mean Streamflow value for the study period (2000-2019)

Year	Mean Streamflow _ Observed (m^3/s)	Simulated Mean Streamflow based on Ground-Observed (m3/s)	Simulated Mean Streamflow based on ERA5-Land (m^3/s)	Simulated Mean Streamflow based on Bias-Corrected ERA5-Land (m^3/s)
2000	-	25.3	33.1	26.9
2001	-	24.1	33.0	26.8
2002	-	21.5	43.5	36.2
2003	21.4	17.1	22.8	18.1
2004	28.3	24.1	36.1	29.6
2005	20.3	19.5	28.8	23.2
2006	13.8	14.1	24.0	19.0
2007	10.9	12.0	18.8	14.6
2008	21.6	25.9	37.9	31.2
2009	29.1	31.1	37.3	30.7
2010	36.1	36.5	51.7	43.4
2011	13.3	15.4	20.9	16.5
2012	17.0	19.9	30.1	24.3
2013	-	33.1	45.9	38.3
2014	-	40.4	51.5	43.1
2015	32.3	17.9	27.6	22.1
2016	16.0	23.4	38.6	31.7
2017	11.9	16.2	21.9	17.4
2018	35.4	24.9	45.1	37.6
2019	64.8	34.5	43.1	35.5
Overall Mean	24.8	23.8	34.6	28.3

The analysis reveals that the ERA5-Land simulations consistently overestimated streamflow, with an overall mean value of 34.6 m^3/s compared to the observed mean of 24.8 m^3/s and the simulated ground-observed mean of 23.8 m^3/s. For example, in 2009, the ERA5-Land streamflow was 37.3 m^3/s, while the observed streamflow was 29.1 m^3/s and the Ground-Observed simulation showed 31.1 m^3/s. This indicates that the ERA5-Land data overestimated flow relative to both observed and simulated values. However, the Bias-Corrected ERA5-Land streamflow is adjusted to 30.7 m^3/s, which is much closer to the observed value, highlighting the effectiveness of the bias correction. A similar pattern is visible in 2010, where the ERA5-Land simulation produced a high mean streamflow of 51.7m^3/s, while the observed value is 36.1 m^3/s and the bias-corrected reduced the ERA5-Land simulated mean streamflow to 43.4 m^3/s, resulting in a better approximation.

To further evaluate the performance of these simulations, the Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) are calculated. The NSE values are computed based on the hourly discharge model results, indicating varying degrees of model performance. The Ground-Observed dataset achieved an NSE of 0.71, indicating a good fit between observed and simulated values. In contrast, the ERA5-Land reanalysis data resulted an NSE of 0.34, reflecting a low level of accuracy and suggesting that the model struggled to replicate observed patterns using this data source. However, Era5-Land bias-corrected data improved the NSE value to 0.55, showing a moderate level of accuracy and notable enhancement compared to the original data. This variation in model performance underscores the importance of dataset selection and bias correction in achieving reliable predictions.

The simplified RMSE values, calculated based on the overall mean streamflow values, highlighted discrepancies in simulation accuracy across the datasets. The RMSE for the Ground-Observed dataset is relatively low at 0.98 m^3/s, highlighting a strong alignment with observed streamflow values. The ERA5-Land dataset showed a higher RMSE of 9.77 m^3/s, emphasizing significant overestimations in streamflow. The Bias-Corrected ERA5-Land dataset showed an RMSE of 3.49 m^3/s, reflecting a marked improvement in accuracy due to bias correction.

5.4 Precipitation Partitioning Analysis

This section provides a comprehensive analysis of precipitation partitioning in the Secchia river basin from 2000 to 2019, utilizing both ground-observed precipitation data and ERA5-Land reanalysis precipitation data. The focus is on understanding how total precipitation is distributed among direct runoff, soil storage, and loss volumes across various sub-basins. The insights gained from this analysis are crucial for effective water resource management and hydrological modeling.

5.4.1 Precipitation Partitioning Using Ground-Observed Data

The analysis of precipitation partitioning using ground-observed data from the years 2000 to 2019 shows significant insights into the hydrological behavior of the Secchia river basin. Table 10 presents the precipitation partitioning results based on ground-observed data for the Secchia River sub-basins. The table provides a comprehensive breakdown of various hydrological components, including total precipitation volume, direct runoff, baseflow, total discharge, total loss volume and soil storage for each sub-basin. The analysis of these helps to illustrate the water balance and hydrological dynamics within the basin, offering valuable insights into the distribution and movement of water across different sub-regions.

Table 10: Precipitation Partitioning Using Ground-Observed Data (2000-2019)

Sub-basin	Total Precipitation Volume (mm)	Total Direct Runoff Volume (mm)	Total Baseflow Volume (mm)	Total Discharge Volume (mm)	Total Loss Volume (mm)	Soil Storage (mm)
W160	17148.8	793.1	5696.7	6489.7	14587.9	1767.8
W170	18498.3	563.9	7857.2	8421.0	13635.6	4298.9
W180	18611.5	1.7	7418.8	7420.6	14283.4	4326.3
W200	19206.1	12.5	7865.6	7878.1	14832.8	4360.9
W210	25117.8	87.0	14915.2	15002.2	20675.8	4355.1
W220	27127.1	116.2	14771.8	14888.0	22624.5	4386.4
W390	16605.8	1879.1	4510.7	6389.8	11488.2	3238.5
W400	17672.9	1820.3	4581.5	6401.8	14097.0	1755.6
W440	18517.2	2.6	7627.7	7630.3	14226.9	4287.7
W490	23791.7	47.1	13569.2	13616.3	19429.9	4314.6
W500	30736.2	237.7	20093.1	20330.8	26100.4	4398.0
Total sum	233033.3	5561.1	108907.5	114468.6	185982.4	41489.8

The total precipitation volume over the 20-year period varied significantly across the sub-basins, ranging from 16,605.8 mm in sub-basin W390 to 30,736.2 mm in sub-basin W500. Sub-basin

W500, with the highest total precipitation, also recorded the highest total discharge volume (20,330.8 mm) and the largest baseflow volume (20,093.1 mm).

Direct runoff, representing the immediate surface response to precipitation, indicates a notable difference among sub-basins. Sub-basins such as W400 and W390 showed the highest direct runoff volumes (1,820.3mm and 1,879.1 mm, respectively), contributing 10.3% and 11.3% of their total precipitation, indicating that these areas are more prone to generating surface runoff. Conversely, sub-basins like W180 and W440 recorded very low direct runoff volumes (1.7 mm, 2.6 mm, respectively), contributing less than 0.01%, indicating that these areas experience limited surface runoff. This precipitation is translated into direct runoff, which totaled 5,561.1 mm across the sub-basins.

Baseflow volumes, a substantial contributor to streamflow, totaling 108,907.5mm**,** varied widely across the sub-basins. Sub-basin W500 had the highest baseflow volume (20,093.1 mm), followed by W210 (14,915.2 mm). In contrast, W390 and W400 had the lowest baseflow volumes, with 4510.7 mm and 4581.5mm, respectively.

The total discharge volume, which is the sum of direct runoff and baseflow, reached 114,468.6 mm. Sub-basin W500 had the highest total discharge (20,330.8 mm), while W390 had the lowest total discharge (6,389.8 mm).

The total loss volume, which includes water lost through evapotranspiration, soil retention and other non-discharge processes, amounted to 185,982.4 mm, representing approximately 79.8% of the total precipitation volume. The cumulative soil storage over the analyzed period was 41,489.8 mm.

Figure 18a illustrates the precipitation partitioning across basins providing a clear picture of the hydrological behavior of each sub-basin, emphasizing variations in surface runoff generation, soil storage and loss. These insights are crucial for effective water resources management and the development of strategies to enhance water availability and sustainability in the region.

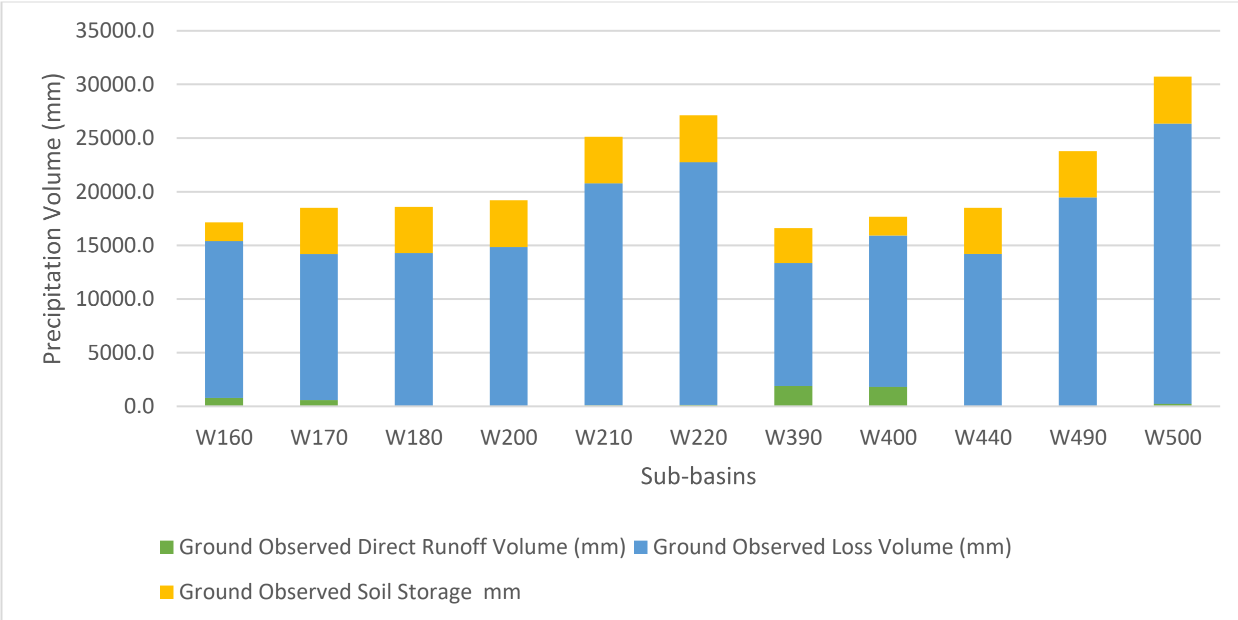

Figure 18a: Precipitation Partitioning Using Ground-Observed Data Per sub-basins (2000 – 2019)

Overall, the cumulative precipitation volume across all sub-basins over the 20-year period is 233,033.3 mm. The total direct runoff is 5,561.1mm, representing 2.4% of total precipitation. Additionally, total losses amounted to 185,982.4 mm, which corresponds to 79.8% of total precipitation) and total soil storage is 41,489.8 mm, equating 17.8% of total precipitation this is illustrated in Figure 18b. This result indicates that a large proportion of precipitation in the Secchia River basin is lost.

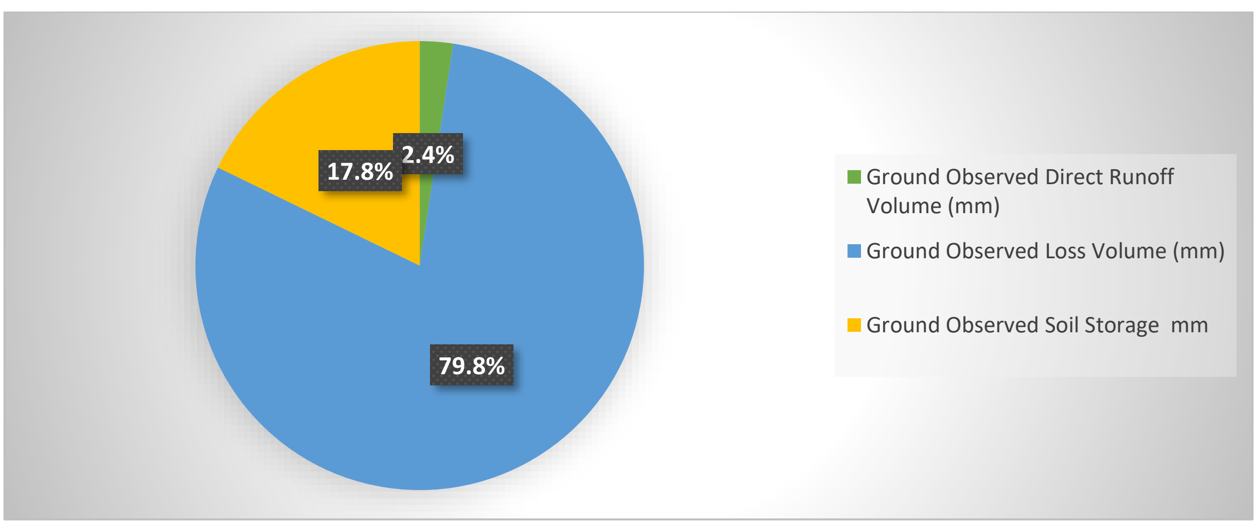

Figure 18b: Total Precipitation Partitioning Using Ground-Observed Data in the study area

5.4.2 Precipitation Partitioning Using ERA5-Land Reanalysis Data

The partitioning analysis based on ERA5-Land reanalysis data for the Secchia River basin during the period from 2000 to 2019 is detailed in Table 11. This analysis highlights the hydrological dynamics and responses of the basin to precipitation inputs, providing insights into the performance of the ERA5-Land dataset in simulating water movement.

Table 11: Precipitation Partitioning Using ERA5-Land Reanalysis Data (2000-2019)

Sub-basin	Total Precipitation Volume (mm)	Total Direct Runoff Volume (mm)	Total Baseflow Volume (mm)	Total Discharge Volume (mm)	Total Loss Volume (mm)	Soil storage (mm)
W160	21593.3	1007.6	12563.2	13570.7	18838.1	1747.7
W170	24766.5	377.9	17116.5	17494.4	20438.7	3950.0
W180	24432.3	2.5	15804.8	15807.2	20098.9	4331.0
W200	24722.2	1.8	16396.7	16398.5	20435.5	4284.8
W210	26577.8	11.8	20106.8	20118.6	22817.7	3748.3
W220	26996.2	9.0	17990.2	17999.2	23210.7	3776.6
W390	19459.3	2176.3	9304.4	11480.6	14621.5	2661.5
W400	22105.5	2286.3	11646.0	13932.3	18065.7	1753.6
W440	24907.7	0.5	16602.5	16603.0	20619.6	4287.6
W490	26394.4	0.9	19741.8	19742.7	22600.2	3793.3
W500	26406.9	5.4	19870.8	19876.2	22725.9	3675.6
Total sum	268362.3	5879.9	177143.7	183023.5	224472.5	38009.9

The hydrological assessment of the Secchia sub-basin, utilizing ERA5-Land reanalysis data over a 20-year period, shows significant insights into the partitioning of precipitation across various sub-basins as illustrated in Figure 19a. The total precipitation volume across all sub-basins amounted to 268,362.3 mm. Among the individual sub-basins, W220 recorded the highest total precipitation volume at 26,996.2 mm, while W390 showed the lowest at 19,459.3 mm.

Total direct runoff recorded across all sub-basins is 5,879.8 mm, accounting for approximately 2.2% of the total precipitation volume. Among the sub-basins, W390 and W400 indicated the highest direct runoff values, with 2,176.3 mm contributing (11.2% of its total precipitation) and 2,286.3 mm (10.3% of its total precipitation), respectively. Notably, these figures are close to those reported in the Ground-Observed data, suggesting that the ERA5-Land dataset provides a reliable estimate of runoff potential in these areas.

The total baseflow volume recorded across all sub-basins is 177,143.7 mm. The Sub-basin W210 showed the highest baseflow volume (20,106.8 mm), while W390 recorded lowest baseflow of 9,304.4 mm.

The total discharge volume, encompassing both direct runoff and baseflow, amounted to 183,023.5 mm. W210 once again exhibited the highest total discharge at 20,118.6 mm, while W390 had the lowest total discharge at 11,480.6 mm.

Total losses, which encompass evaporation, infiltration and other processes, are substantial at 224,472.5 mm. representing approximately 83.6% of the total precipitation volume. The sub-basin W500 experiencing the highest loss volume of 22,725.9 mm. This indicates a significant outflow of water through various pathways, even in areas receiving higher precipitation. The cumulative soil storage over the analyzed period is 38,009.9 mm.

Figure 19a: Precipitation Partitioning Using ERA5-Land Reanalysis Data per sub-basins (2000-2019)

Figure 19b shows the total precipitation partitioning across the entire basin from 2000 to 2019. It illustrates that the soil storage accounted for 14.2% of total precipitation, reflecting notable water retention. In contrast, total losses represented a significant 83.6%, while direct runoff contributed only 2.2%.

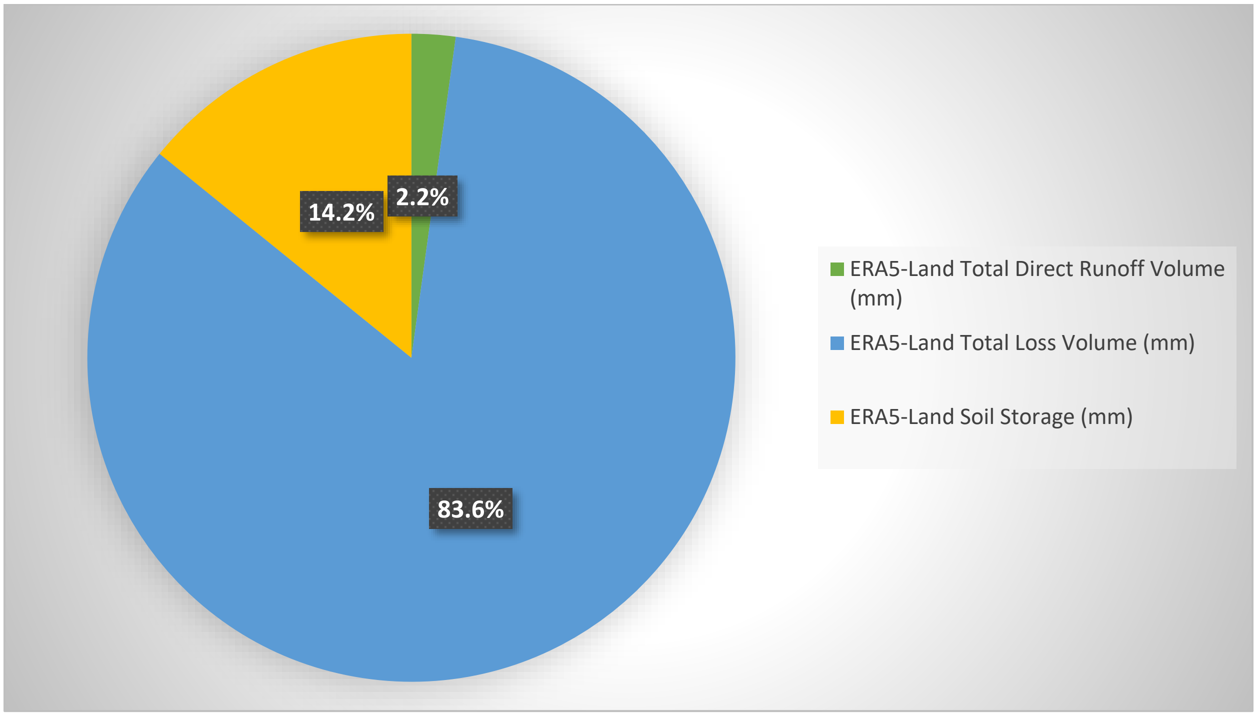

Figure 19b: Precipitation Partitioning Using ERA5-Land Reanalysis Data Over in the study area (2000-2019)

5.4.3 Precipitation Partitioning Using Bias Corrected ERA5-Land Reanalysis Data

The partitioning analysis based on bias corrected ERA5-Land reanalysis data for the Secchia River basin during the period from 2000 to 2019 is detailed in Table 12. The table illustrates that the precipitation volume across the entire basin over the 20 years period is 233459.9 mm, which closely correlates with the ground observed total precipitation value of 233033.3, indicating that the applied technique corrected the overestimation by the original ERA5-Land precipitation data.

Table 12: Precipitation Partitioning Using Bias corrected ERA5-Land Reanalysis Data (2000-2019)

Sub-basin	Total Precipitation Volume (mm)	Total Direct Runoff Volume (mm)	Total Baseflow Volume (mm)	Total Discharge Volume (mm)	Total Loss Volume (mm)	Soil storage (mm)
W160	18786.3	857.5	10003.0	10860.5	18786.3	857.5
W170	21546.7	215.1	13979.1	14194.2	21546.7	215.1
W180	21257.3	0.0	12729.8	12729.8	21257.3	0.0
W200	21508.2	0.0	13331.3	13331.3	21508.2	0.0
W210	23122.8	5.2	16864.9	16870.1	23122.8	5.2
W220	23514.0	1.9	14910.0	14911.9	23514.0	1.9
W390	16929.3	1946.2	6644.6	8590.7	16929.3	1946.2
W400	19232.0	1499.6	9528.1	11027.8	19232.0	1499.6
W440	21670.9	0.0	13438.3	13438.3	21670.9	0.0
W490	22918.2	0.0	15992.6	15992.6	22918.2	0.0
W500	22974.3	41.3	16139.4	16180.7	22974.3	41.3
Total sum	233459.9	4566.9	143561.1	148127.9	233459.9	4566.9

The precipitation partitioning varies across the sub-basins, with W390 receiving the lowest amount of 16,929.3 mm and W220 receiving the highest with 23,514 mm. The direct runoff volume varies significantly, with some basins recording no direct runoff at all, while others showing high runoff volumes. The largest baseflow is observed in W210 with 16,864.9 mm, similarly, the highest total discharge is recorded in the same sub-basin with the value of 16,870.1 mm.

Figure 20a shows the precipitation partitioning for each sub-basins for bias corrected ERA5-Land precipitation data, providing information about surface runoff, soil storage and water loss. The figure illustrates that the applied correction model improved the overestimation shown in the original ERA5-Land precipitation data.

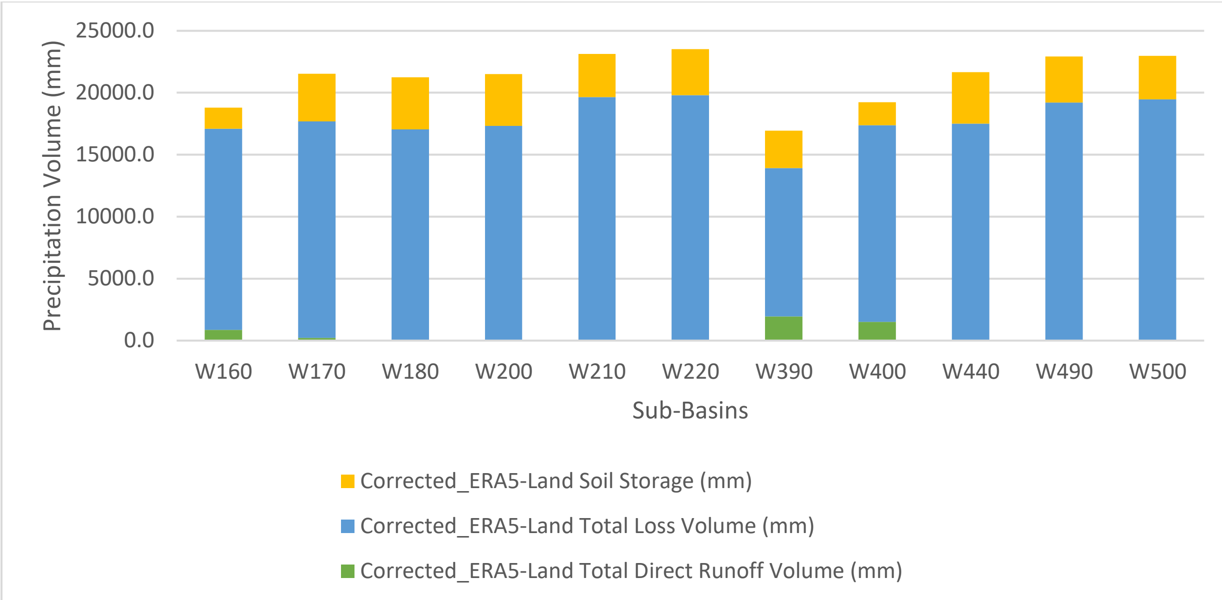

Figure 20a: Precipitation Partitioning Using Bias-Corrected ERA5-Land Reanalysis Data per sub-basins (2000-2019)

Figure 20b shows the cumulative precipitation percentage across the entire study area over the 20 years period. The chart shows that the largest proportion of the precipitation is the loss, which accounts 82% of the precipitation, while the soil storage takes 16% of the precipitation and the direct runoff volume percentage is 2%.

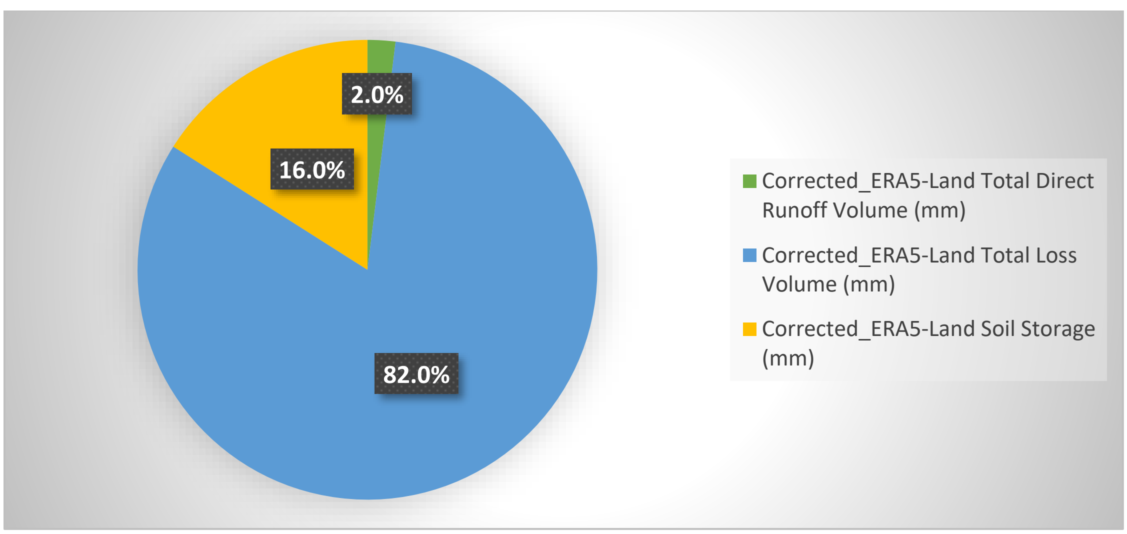

Figure 20b: Precipitation Partitioning Using Bias-Corrected ERA5-Land Reanalysis Data Over in the study area (2000-2019)

6 Conclusions

This thesis conducted a comparative analysis of hydrological modeling using two distinct hourly precipitation datasets, ground-observed data and ERA5-Land reanalysis data, aiming to evaluate their effectiveness in simulating streamflow within the Secchia river basin over a 20-year period from 2000 to 2019. Through a systematic approach, including bias correction of ERA5-Land reanalysis data, precipitation partitioning analysis, and detailed hydrological modeling using the HEC-HMS framework, the study determines strengths and weaknesses of using reanalysis information for hydrological purposes, with specific attention to peak discharge comparisons, mean streamflow assessments, and hydrologic partitioning.

The thesis used results from ground-observed rainfall data to provide a benchmark of model performance. HEC-HMS runs using ground rainfall observations provided estimates of streamflow with good reliability, as demonstrated by a high Nash-Sutcliffe Efficiency (NSE). Hydrological model runs using ERA5-Land reanalysis data exhibited significant discrepancies with respect to observed streamflow, particularly during extreme hydrological events, where peak discharge values were significantly underestimated. On the other hand, the overall rainfall volume provided by ERA5-Land overestimate observations, suggesting the need for a bias correction. In this thesis a simple correction, based on a multiplication factor, was applied to generate a bias-corrected ERA5-Land dataset. Evaluations based on bias-corrected ERA5-Land data improved the long-term analysis, aligning more closely with the observed conditions. But this bias correction significantly deteriorated the ability to reproduce significant events, such as that of February 2016 and December 2017.

The analysis of mean streamflow rates indicate that ERA5-Land simulations overestimated the mean streamflow compared to observed values, underscoring the importance of accurate precipitation data in hydrological modeling. The bias correction applied to the ERA5-Land data effectively minimized these discrepancies, highlighting its potential as a reliable alternative in the absence of ground observations for long-term water balance.

Furthermore, the precipitation partitioning analysis results indicate that while only a small fraction of precipitation contributed to direct runoff, a substantial volume contributed to baseflow, emphasizing the critical role of groundwater in maintaining streamflow in the study area.

The findings of this study have significant implications for water resource management and hydrological modeling practices. This thesis, in particular shows that reanalysis datasets need to be calibrated with the ground observations, and that these adjustments may differ depending on the objectives of the work, such as whether the interest is in reproducing flood events or the long-term hydrologic partitioning. This notion may help using reanalysis datasets to inform in the sustainable management of water resources in the Secchia River Basin and similar regions facing similar hydrological challenges.

While ground-observed data remains the most reliable source for streamflow simulation, the use of bias-corrected ERA5-Land data provides a viable alternative when ground observations are limited. Future research should focus on refining correction techniques, integrating diverse datasets and exploring advanced modeling approaches to further enhance the accuracy and reliability of hydrological predictions.

Bibliography

[1] Kundzewicz, Z. W., et al. (2014). Flood risk and climate change: Global and regional perspectives. Hydrological Sciences Journal.

[2] Ashworth, P. J., Gebrechorkos, S. H., Leyland, J., Dadson, S. J., Cohen, S., Slater, L., Wortmann, M., Hardy, R., Bennett, G. L., Boothroyd, R., Cloke, H., Delorme, P., Griffith, H., Vahidi, E., Hawker, L., McLelland, S., Neal, J., Nicholas, A., Tatem, A. J., Liu, Y., Sheffield, J., Parsons, D. R., & Darby, S. E. (2024). Global-scale evaluation of precipitation datasets for hydrological modelling. Hydrology and Earth System Sciences.

[3] Sahu, M.K., Shwetha, H.R. & Dwarakish, G.S (2023). State-of-the-art hydrological models and application of the HEC-HMS model: a review. Model. Earth Syst. Environ.

[4] Bredesen, A. and Brown, C.J. (2019) Comparison of Hydrologic Model Performance Statistics Using Rain Gauge and NEXRAD Precipitation Input at Different Watershed Spatial Scales and Rainfall Return Frequencies for the Upper St. Johns River, Florida USA.

[5] Johnston, R., Smakhtin, V. (2014). Hydrological Modeling of Large River Basins: How Much is Enough? Water Resour Manage

[6] Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2017). A review of global precipitation data sets: Data sources, estimation and intercomparisons.

[7] Huth, R., Kysely, J., & Pokorny, J. (2016). Evaluation of bias correction methods for regional climate models.

[8] Cannon, A. J., & B. J. H. (2015). Bias Correction of GCM Precipitation by Quantile Mapping.

[9] Bahram Choubin, Shahram Khalighi-Sigaroodi, Ashok Mishra, Massoud Goodarzi, Shahaboddin Shamshirband, Esmatullah Ghaljaee, Fan Zhang, (2019). A novel bias correction framework of TMPA 3B42 daily precipitation data using similarity matrix/homogeneous conditions, Science of The Total Environment.

[10] Vezzoli, R., & Mercogliano, P. (2013). Impacts of land cover and climate changes on peak floods probability distribution function. CMCC.

[11] Bianchi, F., & Montanari, A. (2013). Impacts of land use and climate change on peak flood probabilities in the Secchia River. Hydrology and Earth System Sciences.

[12] Copernicus Climate Change Service. (2023). ERA5-Land reanalysis dataset.

[13] U.S. Army corps of Engineers. HEC-HMS Hydrologic Modeling System. U.S. Army of Crops of Engineers.

[14] Chang, CL., Lo, SL. & Yu, SL. (2006). The Parameter Optimization in the Inverse Distance Method by Genetic Algorithm for Estimating Precipitation. *Environ Monit Assess*