

## Article

# Assessing the National Water Model's Streamflow Estimates Using a Multi-Decade Retrospective Dataset across the Contiguous United States

Mohamed Abdelkader <sup>1,\*</sup>, Marouane Temimi <sup>1</sup> and Taha B.M.J. Ouarda <sup>2</sup>

<sup>1</sup> Department of Civil, Environmental and Ocean Engineering (CEOE), Stevens Institute of Technology, Hoboken, NJ 07030, USA; mtemimi@stevens.edu

<sup>2</sup> Institut National de la Recherche Scientifique, Centre Eau Terre Environnement, INRS-ETE, 490 De la Couronne, Québec City, QC G1K 9A9, Canada; taha.ouarda@inrs.ca

\* Correspondence: mabdelka@stevens.edu

**Abstract:** The goal of this study is to evaluate the performance of the National Water Model (NWM) in time and space across the contiguous United States. Retrospective streamflow simulations were compared to records from 3260 USGS gauging stations, considering both regulated and natural flow conditions. Statistical metrics, including Kling–Gupta efficiency, Percent Bias, Pearson Correlation Coefficient, Root Mean Squared Error, and Normalized Root Mean Squared Error, were employed to assess the agreement between observed and simulated streamflow. A comparison of historical trends in daily flow data between the model and observed streamflow provided additional insight into the utility of retrospective NWM datasets. Our findings demonstrate a superior agreement between the simulated and observed streamflow for natural flow in comparison to regulated flow. The most favorable agreement between the NWM estimates and observed data was achieved in humid regions during the winter season, whereas a reduced degree of agreement was observed in the Great Plains region. Enhancements to model performance for regulated flow are necessary, and bias correction is crucial for utilizing the NWM retrospective streamflow dataset. The study concludes that the model-agnostic NextGen NWM framework, which accounts for regional performance of the utilized model, could be more suitable for continental-scale hydrologic prediction.

**Keywords:** continental-scale hydrological model; simulated streamflow; NWM; model performance; natural flow; regulated flow



**Citation:** Abdelkader, M.; Temimi, M.; Ouarda, T.B.M.J. Assessing the National Water Model's Streamflow Estimates Using a Multi-Decade Retrospective Dataset across the Contiguous United States. *Water* **2023**, *15*, 2319. <https://doi.org/10.3390/w15132319>

Academic Editors: Venkat Sridhar, Ankur Srivastava and Nikul Kumari

Received: 9 May 2023

Revised: 18 June 2023

Accepted: 18 June 2023

Published: 21 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The estimation of streamflow plays a pivotal role in water resource management. It facilitates the forecasting of forthcoming flood events and contributes to the optimization of water demand processes, thereby serving as an indispensable component of effective and sustainable water system management [1–4]. The ability to forecast streamflow accurately can assist water resource managers in making informed decisions regarding the release of water from reservoirs, irrigation, and other water uses [5–7]. The effectiveness of streamflow forecasting in water management has been demonstrated through the use of hydrological models to optimize the operation of reservoir systems to meet the demands of irrigation and hydropower generation [8]. Streamflow forecasts are also utilized to issue flood warnings and aid emergency managers in preparing for, and responding to, flooding events. Furthermore, the accuracy and quality of hydrological forecasts play a vital role improving the efficiency of hydropower generation by allowing power companies to better align their generation schedules with available water resources [9,10].

Simulating streamflow is complex due to the intricate nature of the hydrologic system [11]. A number of factors, encompassing precipitation, evaporation, groundwater recharge, and land use, changes impact streamflow estimation. The scarcity of precise and

reliable data further exacerbates the difficulty in accurately estimating streamflow [12]. Additionally, quantifying and modeling these factors is a challenging task [13,14]. Moreover, it is often challenging to obtain long-term streamflow data for model calibration, particularly in ungauged or inadequately gauged catchments. Therefore, developing accurate prediction models, particularly for ungauged catchments, is a daunting task. Furthermore, estimating streamflow entails increased uncertainty due to the combined impacts of land use development and climate change [15–17].

The estimation of streamflow is further complicated by the nonlinearity and non-stationarity of the hydrological system. Due to its nonlinearity, even small variations in input can lead to substantial changes in output. Previous studies have demonstrated how alterations in precipitation can have significant impacts on streamflow estimation [18–21]. Moreover, the hydrological modeling process uncertainty raises concerns from the simplified or incorrect representation of the hydrological system. Therefore, the lack of a full comprehension of the underlying mechanisms that govern streamflow exacerbates the challenge of creating accurate forecasting models [3,22]. A variety of modeling approaches are used to issue these forecasts, including empirical models and process-oriented models. Process-oriented models use mathematical equations to simulate the physical processes operating in the watershed and estimated streamflow. Several types of process-oriented hydrological models, such as conceptual models and physically based models, which rely on a high degree of spatial discretization, are employed for operational streamflow forecasting [22].

The estimation of streamflow can be issued using hydrological models at various scales, including watersheds, regions, and even the continental scale. Several hydrological models have been developed to estimate streamflow on large scales, including the continental and global scale. Some notable examples of large-scale hydrological models include the Variable Infiltration Capacity (VIC) model [23], the Land Dynamics (LaD) Model [24], and, more recently, the GEOGloWS ECMWF Streamflow service [25]. The models have been employed to simulate streamflow, furnishing valuable insights for water resource managers, thereby enabling them to make well-informed decisions.

In the United States, the National Weather Service (NWS) Advanced Hydrologic Prediction Service (AHPS) supplies streamflow forecasts and warnings for more than 3600 locations across the country [26,27]. However, those forecasts are specific to NWS River Forecast Centers (RFC). The system leverages streamflow forecasts, precipitation observations, and river-gauge measurements to issue flood and flash flood warnings. These alerts are communicated to the public through various channels, encompassing the NWS website, social media platforms, and emergency management agencies. Nonetheless, the National Research Council identified a discrepancy between the state-of-the-art modeling capabilities currently considered cutting edge and those utilized in the AHPS [27]. This observation underscores the necessity to integrate more advanced hydrologic models. In 2016, the National Water Model (NWM), a group of physics-based, hydrodynamic, and hydrologic models, was implemented by NOAA's Office of Water Prediction (OWP). The NWM is based on the Weather Research and Forecasting Model Hydrological modeling system (WRF-Hydro) framework [28] and generates streamflow forecasts for over 2.7-million locations across the CONUS [29].

Streamflow estimation is a key component of flood-risk mitigation. Thus, it is imperative to evaluate hydrological models' performance. It is essential to compare the model's predictions to observed data to validate the model and determine its capacity to reproduce in-land processes. Moreover, assessing the performance of hydrological models can help identify areas for improvement, including addressing model biases, improving the representation of key processes, and incorporating new data and observational capabilities. Additionally, continental-scale models are commonly used to support water management decisions and facilitate decision-making [30]. Thus, evaluating the performance of continental scale models can assist in communicating the model's capabilities and limitations to stakeholders, which can build trust in the model and its predictions.

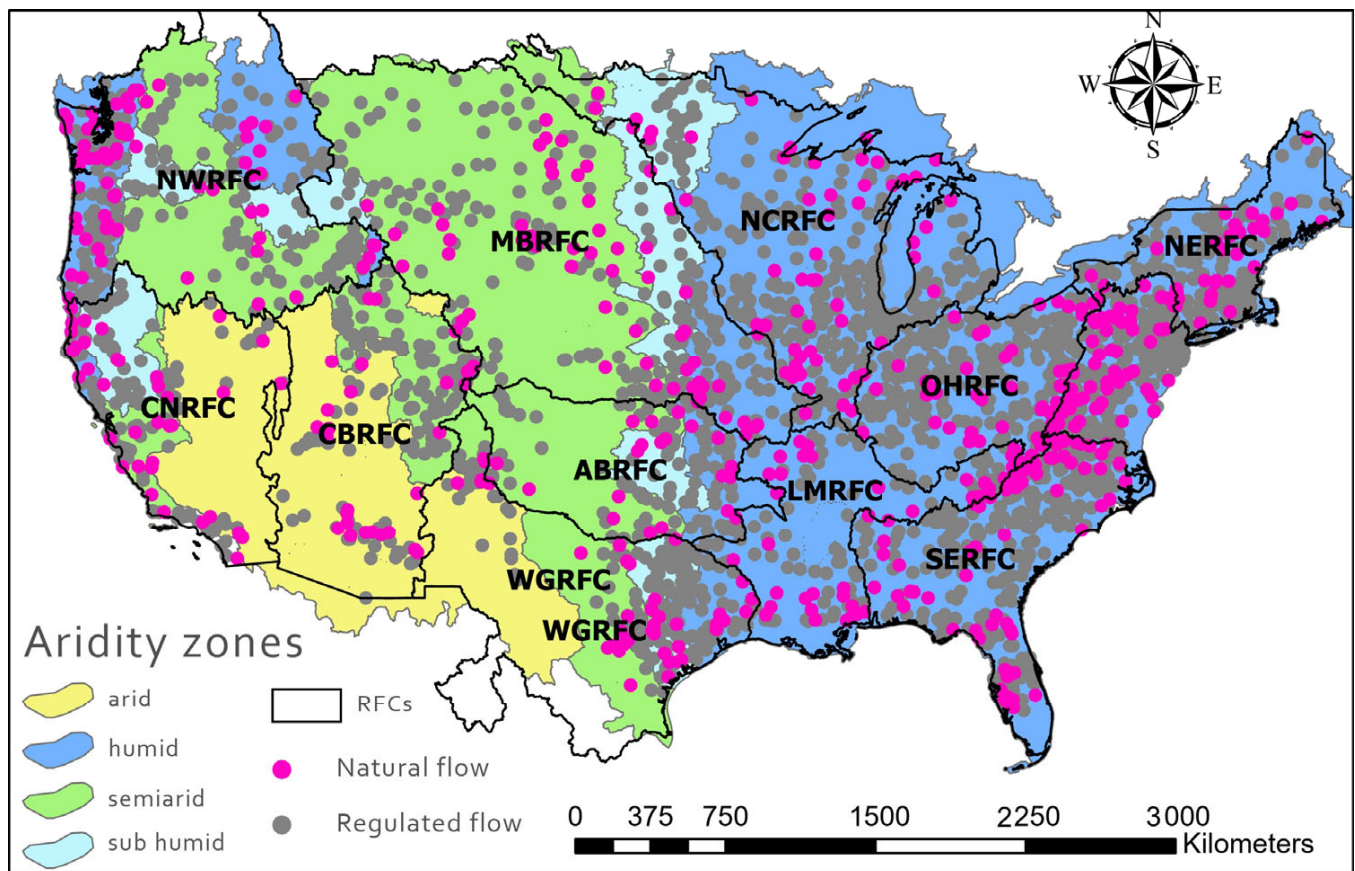
The evaluation of continental-scale hydrological models is critical in ensuring their accuracy and reliability for various applications, such as water-resource management and flood forecasting. A key component of model evaluation involves comparing model estimates to observed data and assessing the model's ability to reproduce historical streamflow patterns. Retrospective simulations are a valuable tool in assessing the performance of continental-scale hydrological models. This approach involves running the model using historical data and comparing its output with observed data from the same period. The retrospective analysis serves to evaluate the model's ability to accurately simulate past hydrological conditions, identify sources of bias or error, and improve its performance. This analysis enables the identification of areas for model improvement, thereby increasing confidence in its predictions. Thus, the evaluation of retrospective simulations plays a crucial role in the continual improvement and refinement of continental-scale hydrological models.

This study aims to evaluate the performance of the NWM V2.1 through the analysis of the results of multi-decadal retrospective simulations across the contiguous United States (CONUS). The study identifies the limitations of the NWM as reported by the retrospective runs, which could be indicative of the necessary areas of improvement to enhance the operational runs. The study investigates the spatial and temporal variability of streamflow bias, with the objective of identifying instances and locations where the model underestimates or overestimates streamflow. Understanding the factors underlying the variations in the model's performance is essential in this regard. Eventually, recommendations could be made to enhance the modeling of hydrologic processes across the US based on the determined biases.

## 2. Materials and Methods

### 2.1. NWM Retrospective Dataset

The National Water Model is a parallelized distributed hydrologic modeling framework based on the WRF–Hydro hydrologic model architecture [28]. In a one-way modeling framework, the model can be forced using precipitation and atmospheric surface data, or in a coupled framework by using the Weather Research and Forecasting (WRF–ARW) atmospheric model. With an hourly modeling cycle, the NWM simulates streamflow for 2.7 million river reaches across the CONUS (Figure 1). In the NWM, hydrologic processes and routing components have been sourced from the community WRF–Hydro modeling system developed at the National Center for Atmospheric Research (NCAR). In the NWM system, land surface processes are represented by the Noah Multi-Parameterization (Noah–MP) land surface model (LSM) [31], while water routing is represented by separate flow routing modules. Based on a 1-km grid, the LSM simulates the vertical exchange of water and energy between the Earth's surface and the atmosphere. The routing modules include diffusive wave surface routing and saturated subsurface routing, utilizing a 250-m grid, and Muskingum–Cunge channel routing utilizing vectorized NHDPlusV2 stream units. In order to improve the model's forecasting cycles' initial states, a nudging data assimilation (DA) scheme was implemented. However, it should be noted that streamflow observations are not incorporated into the retrospective simulations. Currently, there are three versions of the NWM reanalysis dataset available. The NWM versions 1.2 and 2.0 incorporate a 25- and 26-year retrospective simulation, respectively, while Version 2.1 includes an extended retrospective simulation spanning 42 years, from February 1979 to December 2020. A retrospective simulation of NWM V2.1 was used in this study in order to evaluate the performance of the model. All model output and forcing input fields are available in the NetCDF format. Furthermore, all model output fields, as well as the precipitation forcing field, are accessible in the Zarr format. Our study utilized the Zarr format, a cloud-friendly format, to obtain streamflow data. The data were downloaded based on the NWM forecast point reference number and processed locally [32]. The data retrieval process from Zarr files is well-documented in the Amazon Web Services (AWS) portal, where the data is stored [33].



**Figure 1.** Geographical location of USGS streamflow gaging stations collocated with NWM forecast points.

## 2.2. In-Situ Streamflow Data

In this study, the Geospatial Attributes of Gages for Evaluating Streamflow, Version II (GAGES-2) database, maintained by the United States Geological Survey (USGS), was utilized to obtain stream gage locations. The GAGES-2 dataset comprises daily streamflow data for over 9000 stream gages located throughout the United States and its territories, as well as metadata describing the stream gages and the data, including both natural flow and regulated flow sites. Notably, the GAGES-2 dataset was employed solely for the purpose of identifying watersheds, which have natural flow conditions. In the GAGES-2 dataset, stream gages, which at some point in their history had periods which represented natural flow, are identified as USGS hydro-Climatic Data Network (HCDN) [34]. It should be noted that the original composition of the HCDN dataset incorporated a total of 743 distinct sites. The flow type (natural or regulated) was further investigated through the NHDPlus version 2.1 dataset [35]. Subsequently, streamflow data were obtained from the USGS National Water Information System (NWIS) portal [36].

In this study, streamflow data were obtained from the USGS stream gage network. These gages are instruments that measure the water level of a stream or river, and the data are subsequently quality controlled and processed by USGS scientists and technicians to ensure its accuracy and reliability. The USGS stream gage network offers a comprehensive dataset of streamflow records with hourly or sub-hourly intervals, which enables the development of a continental-scale performance evaluation of simulated streamflow. The streamflow records were obtained by converting measured water levels into discharge using established rating curves, which were developed by the USGS through periodic measurement of the stage–discharge relationship, particularly during low- and high-flow



events. It is important to note that the accuracy of the USGS rating curves was assumed in the analysis, and their uncertainty was not considered.

In this study, streamflow data from the USGS gauging stations were temporally and spatially matched with forecast points from the NWM to ensure alignment with the modeled domain. The integrity and continuity of the data are of utmost importance as we evaluate the temporal evolution of streamflow, comparing outputs from the NWM simulations with USGS station records. In line with widely accepted standards for hydroclimatic research, we uphold the need for a minimum of 30 years of records to ensure a reliable long-term analysis. Furthermore, our stringent data completeness criteria stipulate less than 10% missing data, with no consecutive gaps exceeding 24 months [37–40]. Thus, the used dataset was refined by retaining stations meeting the requisite recording period and excluding those with more than 10% missing data. This ensures the findings of this study are based on a robust and reliable data foundation.

Furthermore, we identified certain locations in the NWM’s output where the model was unable to accurately simulate streamflow, resulting in either a constant value or zero for the entirety of the simulation period. These situations, which we refer to as “simulation failures,” introduced potential inaccuracies into the analysis. As such, stations corresponding to these forecast points exhibiting simulation failures were judiciously excluded from our study. This ensured the reliability of our assessment by focusing solely on locations where the model effectively represented the streamflow temporal variability. The final list of streamflow gauging stations used in this study consisted of 3260 stations, with 548 stations located in catchments with natural flow and 2712 located in regulated catchments (Figure 1, Table 1).

**Table 1.** Description of gauging stations, drainage basins, and dominant climate in the studied RFCs.

River Forecast Center	No. of Streamflow Gauging Stations		Median Drainage Basin Area (km <sup>2</sup> )		Dominant Climate Class
	NF	RF	NF	RF	
Missouri Basin (MBRFC)	58	269	760.1	3467.9	Semi-arid
Colorado Basin (CBRFC)	26	165	203.3	1362.3	Arid
Arkansas–Red Basin (ABRFC)	22	125	485.6	2395.7	Semi-humid
California–Nevada (CNRFC)	37	147	189.1	924.6	Arid
Lower Mississippi (LMRFC)	38	111	786.1	1279.4	Humid
Middle Atlantic (MARFC)	60	282	264.2	323.7	Humid
North Central (NCRFC)	55	362	916.8	1414.1	Humid
Northeast (NERFC)	32	183	225.3	489.5	Humid
Northwest (NWRFC)	69	271	401.43	1491.8	Semi-humid
Ohio (OHRFC)	42	263	374.2	1090.4	Humid
Southeast (SERFC)	67	295	323.7	981.6	Humid
West Gulf (WGRFC)	42	236	463.6	1102	Arid & Semi-arid

Note: NF: Natural flow gauging stations; RF: Regulated flow gauging stations.

### 2.3. Ancillary Data

In this study, the Aridity Index (AI) was computed to investigate the NWM model performance for the different climate types. AI was calculated as a function of the Mean Annual Precipitation (MAP) and Mean Annual Potential Evapotranspiration (MAE) as suggested by United Nations Environment Program [41]:

$$AI = \frac{MAP}{MAE'} \quad (1)$$

Total precipitation and potential evapotranspiration data for the period spanning from January 1979 to December 2020 were obtained from the North American Land Data Assimilation System (NLDAS) and aggregated over the Hydrological Unit Code 8 (HUC8) basins. The annual precipitation and potential evapotranspiration time series were processed using the BASINS 4.5 software in accordance with prior investigations that appraised NLDAS products over the CONUS [42,43]. HUC-8 basins were classified into different climate types based on the calculated AI values, which were determined using the UNEP generalized climate classification [41]. The classification was performed based on the AI values, and the resulting categories were Arid ( $0.03 < AI < 0.2$ ), Semi-Arid ( $0.2 < AI < 0.5$ ), Sub-Humid ( $0.5 < AI < 0.65$ ), and Humid ( $AI > 0.65$ ) (Figure 1). It is important to mention that the data used for climate zones classification is subject to climate change signal and to the effects of teleconnections and climate variability (impact of different low frequency climate oscillation indices with different phases), and this topic is important.

#### 2.4. Model Evaluation Metrics

In this study, the accuracy of the NWM streamflow simulations was assessed by comparing time series from the model forecast point to in-situ streamflow observations from USGS stations network. Five statistics were used, including the modified Kling–Gupta efficiency metric (KGE), Percent Bias (PB), Pearson Correlation Coefficient (CC), the Root Mean Squared Error (RMSE), and the Normalized Root Mean Squared Error (NRMSE). It is important to mention that the acceptance threshold for the evaluation metrics used in our study was inferred from commonly used practices in hydrological studies [19,44–49].

The NWM performance at the hourly time scale was evaluated using the KGE metric. The Kling–Gupta efficiency (KGE) metric is used to evaluate the performance of hydrological models in terms of their ability to reproduce the temporal and spatial patterns of observed streamflow. It is a statistical measure that compares the correlation, bias, and variability of simulated streamflow to those of the observed streamflow. A KGE value of 1 indicates perfect agreement between the observed and simulated streamflow, while a value less than 1 indicates that the model is not performing as well. The KGE metric is widely used in the hydrological community to evaluate model performance and to identify areas where model improvement is needed [50]. The assessment of hydrological dynamics in the KGE metric is defined by three components: the temporal error through correlation (CC), bias errors ( $\beta$ ), and variability errors ( $\gamma$ ) expressed as follows:

$$KGE = 1 - \sqrt{(CC - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}, \quad (2)$$

$$CC = \frac{\sum_{i=1}^N (o_i - \mu_o)(s_i - \mu_s)}{\sqrt{\sum_{i=1}^N (o_i - \mu_o)^2 \sum_{i=1}^N (s_i - \mu_s)^2}}, \quad (3)$$

$$\beta = \frac{\mu_s}{\mu_o}, \quad (4)$$

$$\gamma = \frac{\sigma_s / \mu_s}{\sigma_o / \mu_o}, \quad (5)$$

where  $N$  is the total number of observations  $i$  at each station, the streamflow observations are denoted by ( $o$ ), and simulations obtained from the NWM are denoted by ( $s$ ). The mean and standard deviation of the streamflow are defined by  $\mu$  and  $\sigma$ , respectively.

The root mean square error (RMSE) constitutes a prevalent metric employed for assessing the performance of numerical models in simulating hydrological processes. As a measure of deviation between predicted and observed values, RMSE is computed by obtaining the square root of the mean of squared differences between these values. This particular metric proves advantageous in hydrological applications, as it furnishes a quantitative method for comparing model performance across distinct temporal and

spatial scales and delivers a singular, comprehensive metric encapsulating the overall model accuracy. In general, lower RMSE values signify superior model performance, with a value of zero denoting an impeccable correspondence between predicted and observed values. The RMSE for each forecast point was calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (s_i - o_i)^2}{N}}, \quad (6)$$

It is imperative to highlight that the root mean square error (RMSE) is inherently sensitive to the magnitude and frequency of deviations between simulated and observed values. In the realm of hydrological modeling, the primary objective is to accurately represent the temporal and spatial variability of hydrological processes, with a particular focus on peak streamflow values, as these significantly influence the overall model precision. Owing to the computation of RMSE as the square root of the mean of squared discrepancies, substantial deviations in peak values may disproportionately affect the composite RMSE value. For example, in cases where a model persistently underestimates peak flow values, the resultant RMSE will be considerably inflated, even if the model demonstrates adequate performance in simulating the majority of flow values. This inherent limitation of RMSE as a performance metric stems from its inability to discern between errors in simulating peak and low-flow events. Consequently, additional metrics, including normalized root mean square error, correlation coefficient, and percent bias, are incorporated into this study to furnish a comprehensive assessment of model performance.

In order to evaluate the performance of the NWM relative to in-situ observations, the NRMSE was computed. The utilization of NRMSE is particularly advantageous in scenarios where the range of observed values exhibits considerable variability or spans a wide spectrum, as it accounts for the inherent fluctuations within the observed data and furnishes a more robust and equitable comparison of model performance. Additionally, NRMSE was employed primarily to probe the spatial performance of the model across the continental United States. Given the broad spatial scale of our model's application, it is fundamental to understand how its relative performance varies spatially when compared to the observed dataset. The NRMSE was calculated as follows:

$$\text{NRMSE} = \frac{\text{RMSE}}{\mu_o}, \quad (7)$$

In conjunction with the aforementioned evaluation metrics, the percent bias metric was computed to estimate the error and uncertainty associated with the National Water Model (NWM) streamflow simulations. PB serves as a metric employed in the assessment of hydrological model performance by quantifying the disparity between predicted and observed values, expressed as a percentage of the observed values. PB is beneficial in hydrological model evaluation, as it offers a measure of the overall bias inherent in a model's predictions. A PB value of zero signifies an unbiased model, where the predicted values, on average, correspond to the observed values. Positive PB values denote an overestimation of observed values by the model, while negative PB values imply model underestimation of the observed values. In this study, the PB was calculated as follows:

$$\text{PB} = \frac{\sum_{i=1}^N (s_i - o_i)}{\sum_{i=1}^N o_i} \times 100, \quad (8)$$

Further analysis of streamflow time series for monotonic trends (one direction, either increasing or decreasing) was performed using the non-parametric Modified Mann–Kendall trend test.

### 2.5. Modified Mann–Kendall Trend Test

In the present study, the spatiotemporal evaluation of NWM streamflow estimates was partially investigated by performing streamflow trend analysis utilizing the Modified Mann–Kendall (MMK) trend test [51]. The non-parametric MMK test, which is widely employed [39,52–54], was selected for this investigation due to its lack of requirements for data to adhere to a specific distribution, reduced sensitivity to abrupt shifts resulting from non-homogeneity in the data, and its ability to account for autocorrelation in streamflow time series. Within this context, we analyze trend outcomes for observed and simulated streamflow series, comparing the spatial variation of the trend results. Consequently, the congruence between trends derived from the NWM streamflow series and observed series is scrutinized to evaluate the model's capacity to replicate historical streamflow patterns. It is crucial to note that the trend analysis test was applied exclusively to natural flow series.

## 3. Results

The accuracy of the NWM streamflow simulation was assessed by comparing with in-situ data using five evaluation statistics, namely KGE, PB, RMSE, NRMSE, and CC. The hourly streamflow data acquired from in-situ and retrospective simulations from February 1979 to December 2020 were used to analyze the Spatiotemporal Variability of the model accuracy. The basin drainage area and RFCs differences were considered.

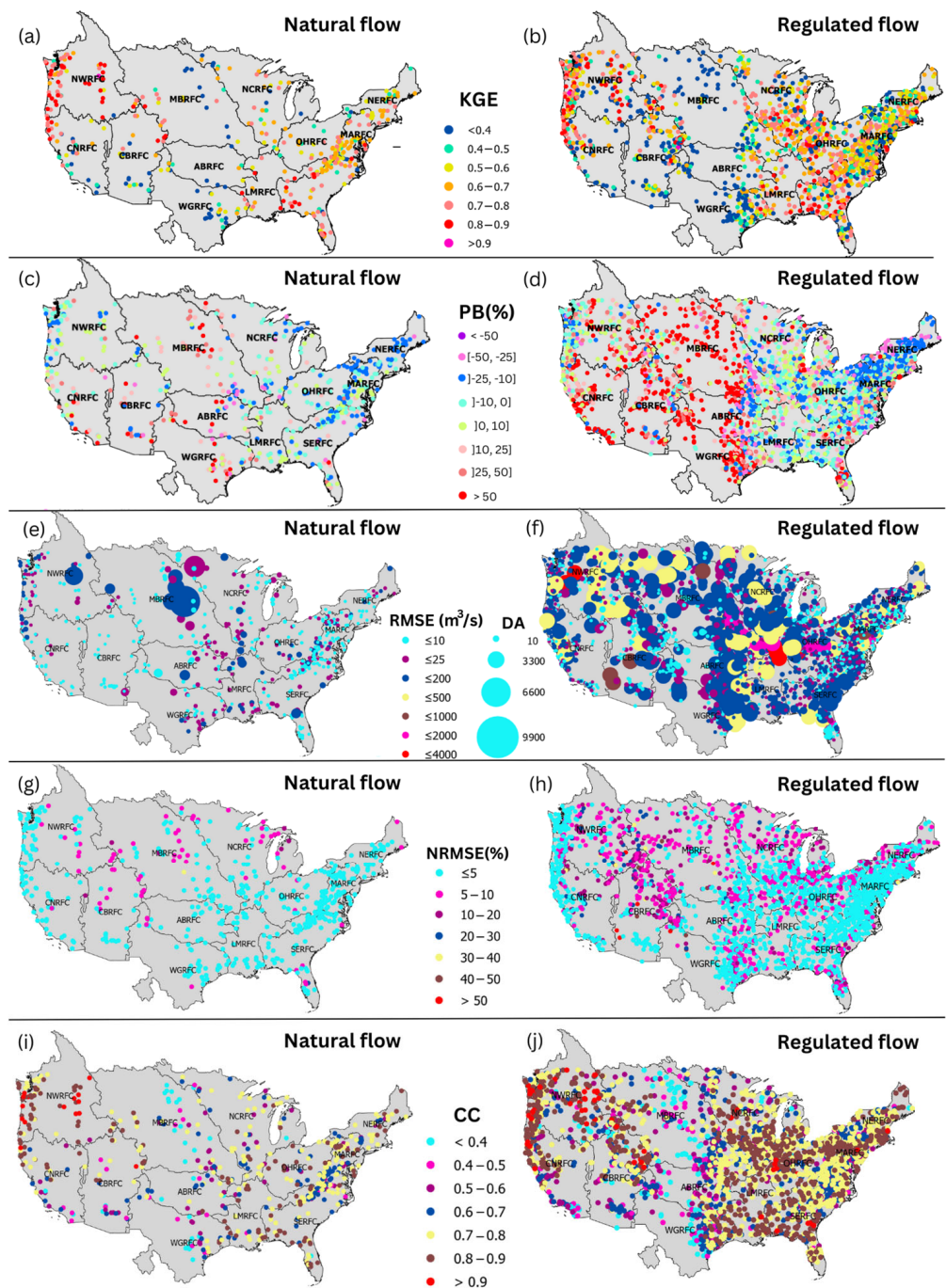
### 3.1. Spatial Analysis

The findings derived from the evaluation metrics assessing the NWM streamflow simulations' performance are illustrated in Figure 2. The acquired results demonstrate that 57% of natural flow forecast points exhibit KGE values exceeding 0.5 (Figure 2a), while 43% of forecast points associated with regulated flow present KGE values surpassing 0.5 (Figure 2b). The outcomes reveal suboptimal model performance for forecast points situated within the MBRFC, ABRFC, and WGRFC. Conversely, the model exhibits a commendable performance in estimating streamflow within the NERFC, MARFC, and NWRFC. It is crucial to note that KGE values represent an overall estimation of the concordance between observed and simulated streamflow. In essence, KGE values may be influenced by errors related to correlation, variability, bias, or a combination of these metric components.

The variation in bias error across the different River Forecast Centers (RFCs) was investigated utilizing the percent bias (PB) metric. For natural flow, a satisfactory agreement between observed and simulated streamflow was achieved for 50% of the forecast points, with PB values ranging between  $-10\%$  and  $10\%$ . PB values did not exhibit a discernible spatial distribution of bias error. However, a tendency for the model to underestimate streamflow values was observed in the eastern and northwestern portions of the contiguous United States, specifically within the NERFC, MARFC, SERFC, and NWRFC. Conversely, the model exhibited a propensity to overestimate streamflow in the central and western regions of the CONUS, predominantly over the MBRFC, ABRFC, and WGRFC (Figure 2c).

The findings are also applicable to regulated flow forecast points, where the model tends to underestimate streamflow values in eastern RFCs and overestimate streamflow in central and western RFCs. The results revealed that 36% of regulated flow estimates demonstrated low bias error with PB values between  $-10\%$  and  $10\%$ . Low PB values (close to zero) were primarily observed in the NERFC, MARFC, SERFC, NCRFC, and NWRFC (Figure 2d). It is vital to note that RFCs where the NWM displays low PB values are characterized by high annual precipitation rates (reference). In contrast, high PB values were observed in regions with low annual precipitation rates. The results suggest that the NWM configuration is well-suited for regions dominated by high precipitation. In other words, the model structure and calibration parameters accurately represent surface runoff processes.





**Figure 2.** Spatial variation of evaluation metrics results for the NWM hourly streamflow series against observed natural and regulated streamflow series. PB, RMSE, NRMSE, and CC results for natural flow sites are presented in subfigures (a,c,e,g,i), respectively. PB, RMSE, NRMSE, and CC results for regulated flow are presented in subfigures (b,d,f,h,j), respectively.

Natural streamflow estimates exhibited low RMSE values in the eastern and western RFCs of the CONUS, irrespective of drainage basin area (Figure 2e). High RMSE values were observed for natural streamflow estimates in the central and eastern regions of the United States. In terms of drainage area, the highest RMSE values were associated with drainage basins encompassing an area exceeding 3000 km<sup>2</sup>. Both RMSE and normalized NRMSE demonstrated optimal results for NWM streamflow estimates in the eastern and western RFCs. The NWM natural streamflow data displayed elevated NRMSE results in the central region, particularly in the MBRFC and portions of the CBRFC. Regulated flow

data exhibited inferior performance compared to natural flow, as indicated by higher RMSE and NRMSE values (Figure 2g,h) across the various RFCs within the CONUS. Regulated streamflow estimates yielded the most favorable outcomes in terms of RMSE and NRMSE values in the northeastern region, primarily for forecast points with a drainage area of less than 1000 km<sup>2</sup>. NRMSE values below 5% were obtained for 86% of natural forecast points and 63% of regulated forecast points, respectively.

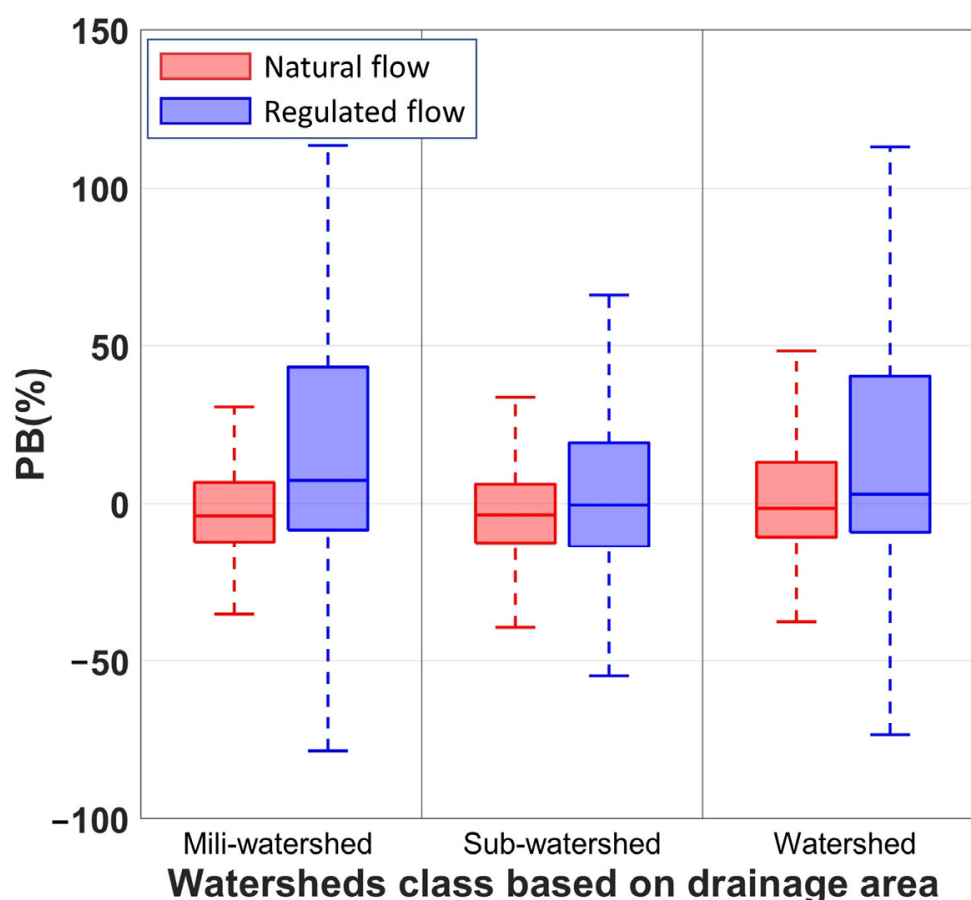
Overall, the model error outcomes in terms of RMSE and NRMSE indicate enhanced performance of the NWM in estimating natural streamflow. The obtained results can be attributed to the fact that regulated flow watersheds exhibit altered hydrological regimes due to the presence of water control structures, such as dams, weirs, and other hydraulic structures. These structures can significantly alter the timing and magnitude of downstream discharge. Consequently, it is challenging to capture the effects of these alterations on a continental-scale hydrological system. In essence, the NWM configuration may not be capable of simulating the intricate interactions between hydraulic structures and the hydrological system. Furthermore, regulated flow basins are often subject to human interventions, which can further complicate the hydrological regime, rendering it challenging for the NWM to accurately simulate streamflow.

Natural streamflow estimates demonstrated optimal results in terms of CC values ( $CC > 0.50$ ) within the eastern and western RFC. The natural streamflow estimates exhibited the weakest correlation with in situ streamflow observations in the central region, particularly for the MBRFC and WGRFC (Figure 2i). Overall, 89% of the natural flow forecast points had CC values exceeding 0.5. Conversely, 88% of the regulated flow forecast points had CC values greater than 0.5 (Figure 2j). Forecast points with CC values larger than 0.8 represented 30% and 31% of the natural flow and regulated flow forecast points, respectively. The observed results indicated a favorable agreement between observed and estimated streamflow values for regulated flow across the majority of the RFCs. However, the model performance deteriorated for regulated flow forecast points situated within the MBRFC, ABRFC, and WGRFC. The results revealed that the NWM is proficient in capturing the temporal evolution of streamflow for both natural and regulated flow. It is crucial to note that the inferior performance of the NWM in terms of CC is observed in regions with low annual precipitation rates. The findings suggest that the total error associated with streamflow estimates in terms of KGE predominantly originates from the bias error when compared to RMSE, NRMSE, and CC values. Consequently, further model accuracy analysis in the present study will be conducted based on percent bias results.

Overall, the spatial analysis demonstrated a superior agreement between the NWM streamflow estimates and natural flow observations compared to regulated flow observations. The NWM's performance was suboptimal in the Missouri Basin, Arkansas–Red Basin, and West Gulf River Forecast Centers. However, a more favorable agreement was attained in RFCs situated along the east and northwest coasts of the CONUS. Notably, a commendable model performance ( $KGE > 0.5$ ) was generally observed in humid regions characterized by substantial precipitation rates, suggesting that the NWM accurately represents the rainfall-runoff process. Conversely, the model's poor performance in drier regions may be attributed to its limitations in capturing the seasonal variability of the hydrological system in arid and semi-arid regions. The NWM may be incapable of capturing these types of variations in the natural system, resulting in a weak agreement between observed and estimated streamflow in regions where precipitation contribution to streamflow generation is limited.

The performance of NWM streamflow estimates across different RFCs CONUS as a function of drainage areas was investigated. The watershed classification based on drainage area was derived from the classification scheme proposed by Singh (1994) [55]. Figure 3 depicts the PB variation for distinct watershed classes across the CONUS. The results indicated a superior performance of the model in estimating streamflow for natural flow compared to regulated flow. Overall, the model exhibited a propensity for underestimating streamflow for Milli-watersheds (1000–10,000 ha) and sub-watersheds

(10,000–50,000 ha) in natural flow estimates, with median PB values of  $-3.90\%$  and  $-3.60\%$ , respectively. Conversely, for the watershed class ( $>50,000$  ha), the results demonstrated a lower median PB value of  $-1.51\%$ , with the model tending to overestimate streamflow. In contrast, regulated flow outcomes displayed a suboptimal model performance with a broader spread in PB values. However, the model's behavior was more stable for sub-watersheds, exhibiting a median PB value of  $-0.49\%$  and a limited spread of obtained results ( $Q1 = -13.60$ , and  $Q3 = 19.06\%$ ). It is vital to note that the findings did not reveal any large-scale systematic bias in the model's performance as a function of watershed drainage area. Consequently, to better characterize the model bias on a regional scale, the PB values for each RFC were extracted and analyzed individually.



**Figure 3.** Boxplot showing PB variation as a function drainage area for natural and regulated flow. Boxes indicate the interquartile range (IQR) of the data. Within each box, horizontal lines indicate the median PB value; the first (Q1) and third (Q3) quantiles are marked by boxes, and whiskers extend to 1.5 interquartile ranges.

PB results for each RFC as a function of drainage area are delineated in Tables 2 and 3 for natural flow and regulated flow, respectively. The findings revealed that the model tends to underestimate natural flow in Milli-watersheds within the MARFC, NERFC, NWRFC, OHRFC, and SERFC. Conversely, the model exhibits a propensity to overestimate natural flow in Milli-watersheds situated in the MBRFC and CNRFC. For sub-watershed and watershed drainage basins, the model demonstrates a tendency to underestimate natural flow in the MARFC, NERFC, NCRFC, and OHRFC while overestimating natural flow in the MBRFC and CNRFC. The analysis suggests that NWM natural streamflow estimates are predominantly underestimated in eastern RFCs and overestimated in central RFCs, particularly for forecast points draining Milli-watersheds and sub-watersheds.

**Table 2.** Comparison of natural flow PB results for the different RFC as a function of watershed drainage area.

RFC	Milli-Watershed			Sub-Watershed			Watershed		
	Q1	Median	Q3	Q1	Median	Q3	Q1	Median	Q3
Missouri Basin (MBRFC)	2.66	4.46	12.53	−5.61	5.93	20.52	−9.24	11.02	32.78
Colorado Basin (CBRFC)	−6.90	−3.52	6.87	−16.02	2.91	27.74	−17.76	−13.58	65.53
Arkansas–Red Basin (ABRFC)	N/A	N/A	N/A	−27.75	1.55	15.35	−12.46	2.24	57.36
California–Nevada (CNRFC)	−6.53	12.98	28.84	4.42	10.78	31.22	5.82	13.44	18.52
Lower Mississippi (LMRFC)	N/A	N/A	N/A	1.54	5.20	6.33	−5.81	−1.75	3.66
Middle Atlantic (MARFC)	−12.91	−4.25	0.53	−10.48	−4.63	3.52	−13.36	−10.88	3.49
North Central (NCRFC)	N/A	N/A	N/A	−20.21	−10.15	2.19	−9.09	−3.08	6.62
Northeast (NERFC)	−12.81	−9.06	−5.57	−21.76	−18.64	−12.07	−23.71	−23.02	−19.25
Northwest (NWRFC)	−4.46	−2.10	1.18	−12.04	−4.77	0.60	−6.64	0.73	5.68
Ohio (OHRFC)	−6.75	−4.98	−0.10	−12.10	−6.41	−2.36	−11.33	−8.87	−4.07
Southeast (SERFC)	−15.87	−12.50	−9.35	−10.23	−4.60	1.95	−9.81	−1.83	10.35
West Gulf (WGRFC)	N/A	N/A	N/A	−20.08	7.73	22.74	6.52	21.74	58.78

Note: N/A: No forecast points representing this class, or limited number of forecast points.

**Table 3.** Comparison of regulated flow PB results for the different RFC as a function of watershed drainage area.

RFC	Milli-Watershed			Sub-Watershed			Watershed		
	Q1	Median	Q3	Q1	Median	Q3	Q1	Median	Q3
Missouri Basin (MBRFC)	−6.59	3.56	18.82	0.74	18.89	30.77	2.65	49.38	95.14
Colorado Basin (CBRFC)	−9.92	3.19	34.40	−15.22	−1.84	25.87	16.65	56.06	102.26
Arkansas–Red Basin (ABRFC)	N/A	N/A	N/A	−6.11	7.17	17.49	−10.69	11.03	83.97
California–Nevada (CNRFC)	−6.66	25.98	74.61	11.22	36.04	85.69	16.07	80.43	175.08
Lower Mississippi (LMRFC)	N/A	N/A	N/A	−10.29	−4.50	1.54	−9.99	−4.42	2.27
Middle Atlantic (MARFC)	−12.55	0.16	14.92	−12.52	−2.88	7.85	−13.78	−8.24	2.89
North Central (NCRFC)	15.08	47.14	78.24	−11.15	0.99	23.45	−6.86	2.21	13.65
Northeast (NERFC)	−11.07	2.47	35.84	−19.74	−12.67	−2.62	−25.40	−19.48	−12.88
Northwest (NWRFC)	−2.23	1.76	26.49	−11.34	−2.51	19.45	−3.73	6.75	39.54
Ohio (OHRFC)	−5.16	−2.33	14.04	−15.44	−5.20	6.67	−13.74	−6.35	−1.67
Southeast (SERFC)	−9.20	26.35	63.87	−10.09	−1.47	16.16	−8.47	−2.67	8.39
West Gulf (WGRFC)	15.42	59.55	84.37	−7.42	19.38	53.09	12.96	48.02	127.11

Note: N/A: No forecast points representing this class, or limited number of forecast points.

For regulated flow, the PB results exhibited less uniformity in bias distribution across various RFCs and drainage basin sizes. The outcomes indicated a propensity for the model to underestimate regulated flow over sub-watersheds and watersheds situated in the MARFC, NERFC, and OHRFC. Additionally, the results demonstrated a clear overestimation of streamflow in the MBRFC and CNRFC for different drainage basin sizes. The findings imply a random bias associated with streamflow estimates in regulated channels (Table 3). These results can be attributed to both natural and anthropogenic alterations in the drainage basin environment that may not be accurately captured by the model due to static inputs such as land use/cover and catchment topography. Consequently, such

changes may be more pronounced for smaller drainage basins where less consistency in the model behavior was observed.

A crucial aspect of the NWM's performance lies in streamflow estimates as a function of stream order. In the present study, PB values corresponding to forecast points for different stream orders were analyzed. As depicted in Figure 4, the PB values for natural flow estimates exhibit an approximately normal distribution, displaying a symmetrical distribution around the median and a relatively small interquartile range. It is also noteworthy that the PB values for natural flow series do not demonstrate significant skewness in any direction. Conversely, boxplots for the regulated flow outcomes indicate that PB values vary considerably as a function of stream order, exhibiting a substantial spread in IQR. The results also reveal high bias streamflow estimates obtained from first- and second-order regulated streams, suggesting a persistent bias associated with low-order regulated streams. The observed results can be attributed to the inherent assumptions related to channel properties in the NWM river network, including the adoption of trapezoidal channel geometry, uniform flow, and constant channel roughness that only varies depending on the stream order (up to 10 orders). These assumptions may influence flow routing and potentially introduce uncertainties in streamflow estimation that are order dependent. Additionally, the NWM employs the Noah-MP land surface model to simulate atmospheric exchanges with the surface and vertical fluxes within the soil moisture column at a 1-km grid resolution. Consequently, the model is anticipated to exhibit greater error generation in low order streams (small catchment), where the representation of channel infiltration may be insufficiently captured [56]. This limitation underscores the need for further refinements in the model's representation of channel properties and soil moisture distribution to improve streamflow predictions, particularly in small-scale drainage systems.

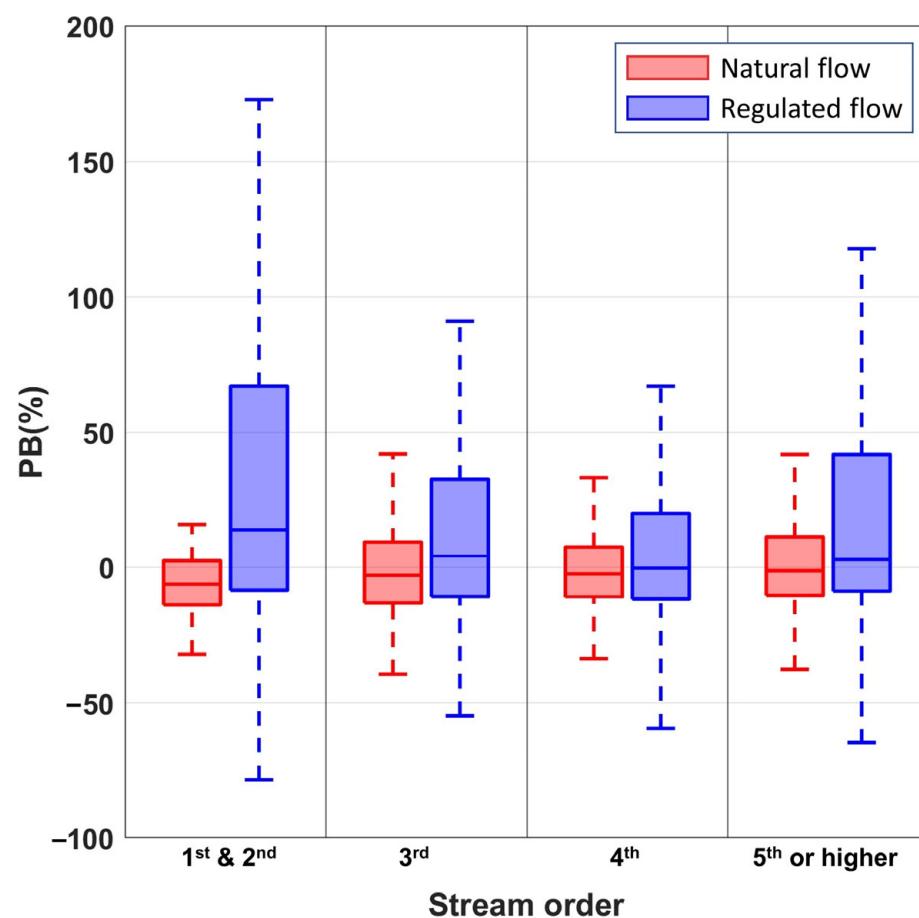


Figure 4. Boxplot showing PB variation as a function stream order for natural and regulated flow.



### 3.2. Temporal Analysis

The PB results were calculated for the NWM's natural and regulated flow estimates in comparison with in-situ streamflow data. As shown in Figure 5, a more favorable agreement between the estimated and observed streamflow data is achieved for natural flow across all 12 months of the year. For natural flow estimates, the results indicate that the model has a tendency to overestimate streamflow during the period between August and January, with median PB values greater than zero and a relatively positive skew in PB data. Natural flow estimates are predominantly underestimated during the period between March and June, with negative Q1 and median PB values, and Q3 values closer to 0. Higher negative PB values were obtained during March and April when seasonal transitions in streamflow regimes occur in most of the CONUS. The transition season coincides with snowmelt periods, which could be attributed to the NWM tendency to underestimate snow water equivalent (SWE). Conversely, the model's positive bias during the wet season could be attributed to errors in precipitation inputs that trigger the generation of excess runoff.

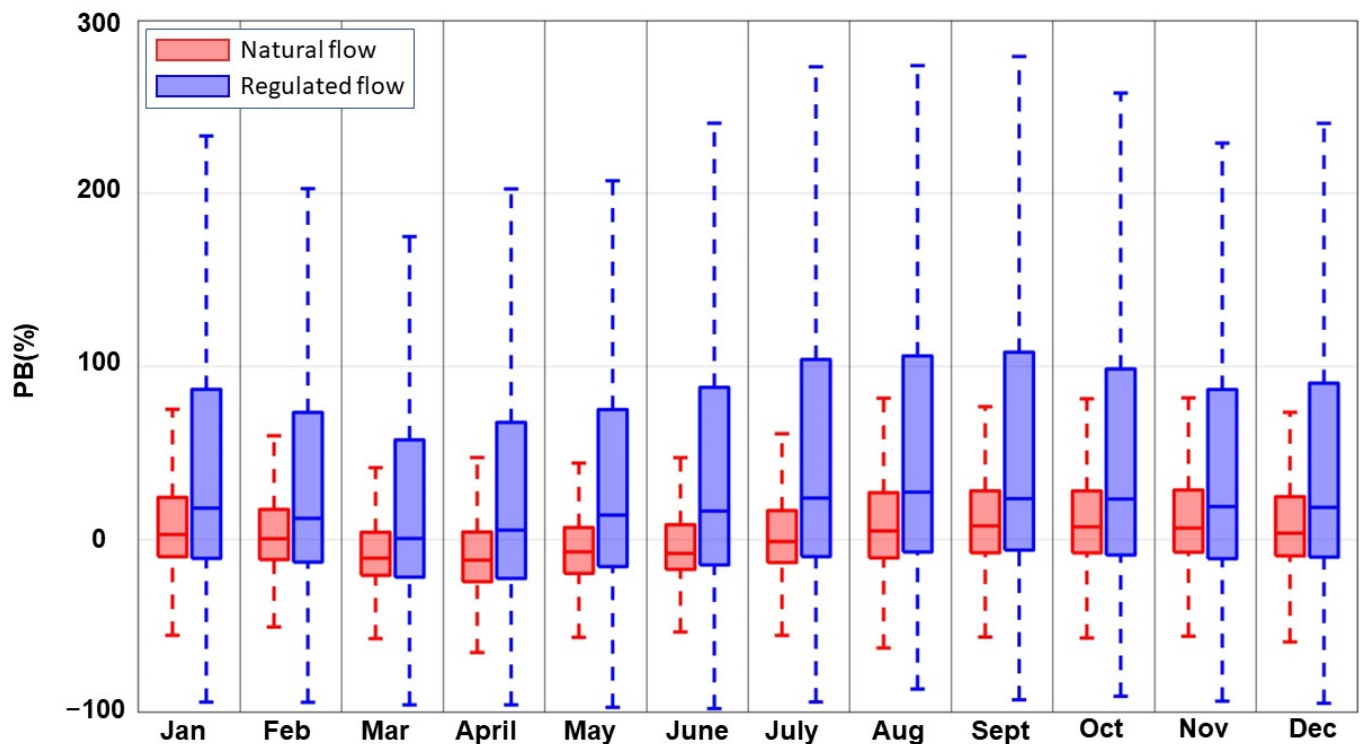


Figure 5. Boxplot showing monthly PB variation for natural flow and regulated flow hourly data.

The monthly variation of PB for regulated flow was assessed in this analysis. The PB for NWM estimates exhibited positive median values for all months, indicating a general propensity for the model to overestimate regulated flow (Figure 5). The regulated flow estimates displayed enhanced performance during February and March, with median PB values closer to 0 and a relatively smaller IQR spread. Conversely, the results revealed the poorest PB values in July, August, and September, with median PB values of 23.90%, 27.37%, and 23.59%, respectively, and a large IQR spread, showcasing a tendency to overestimate streamflow. The larger PB values can be attributed to the fact that the NWM does not incorporate reservoir management practices. In other words, water retention in reservoirs is not represented in the model, which explains the substantial positive PB values.

### 3.3. Temporal–Spatial Analysis

The hourly streamflow data derived from the NWM retrospective simulation were contrasted with hourly in-situ streamflow data on a seasonal basis, taking into account the various climate zones encompassing the CONUS. It is crucial to acknowledge that

the data utilized for the classification of climate zones could be influenced by the signal of climate change, as well as the impacts of teleconnections and climate variability, including the implications of diverse low-frequency climate oscillation indices at varying phases. This topic warrants substantial attention. However, the primary objective of the present research is to offer a preliminary delineation of the spatial distribution of aridity zones.

Regardless, the methodologies derived for climate zone classification align well with both regional and global studies that have explored the spatial variation of the Aridity Index [57,58]. Additionally, the findings of the current study correspond closely with those of Heidari et al. (2020) [58]. Heidari et al. (2020) evaluated shifts in regional hydroclimatic conditions across the contiguous United States, in response to climate change throughout the 21st century. He generated Aridity Index maps for the period from 1989–2015, which exhibited patterns remarkably similar to those discerned in the present study [58].

Figure 6 illustrates the spatiotemporal variation in the accuracy of the NWM natural flow estimates in terms of PB. In arid regions, the results revealed that the model generally overestimates streamflow values across all seasons, with median PB values of 7.85%, 18.80%, 19.54%, and 11.89% for fall, winter, spring, and summer estimates, respectively. These findings imply that the model exhibits enhanced performance during the summer and fall seasons in arid regions, while the performance deteriorates during the winter and spring seasons.

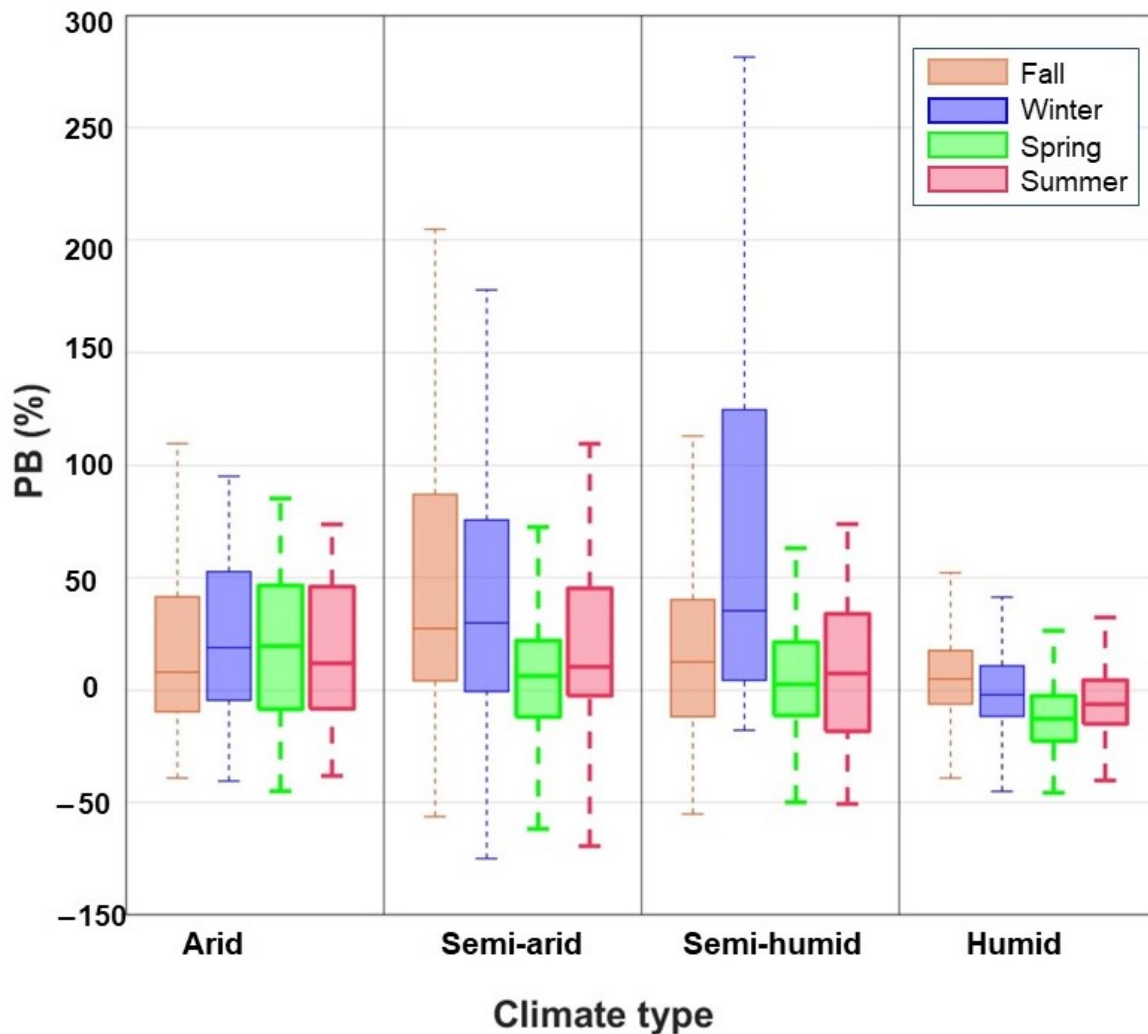


Figure 6. Box plot showing seasonal PB variation for natural flow as a function of climate type.

In semi-arid regions, the results indicated a favorable agreement between the model estimates and observed streamflow data during the spring season. In the remaining seasons, the model exhibited a tendency to overestimate streamflow, with median PB values exceeding zero and an IQR heavily skewed towards positive PB values. Streamflow estimates in semi-humid regions demonstrated a good correspondence with in-situ measurements, exhibiting a median PB value of 2.53%. The model displayed a clear inclination to overestimate streamflow during the winter season in the same region. Although the NWM streamflow estimates did not exhibit any systematic bias, the PB values were skewed in a positive direction.

The most favorable agreement between the NWM estimates and observed data was obtained for simulations conducted in humid regions. The results revealed low-bias error values with median PB values close to zero. The model's best performance was observed during the winter season, potentially reflecting the model's capability to generate accurate runoff during wet periods. It is worth noting that the model tends to underestimate streamflow during the spring season, which could be attributed to the model's limitations in snowmelt-dominated regions.

Based on the results, it can be concluded that the accuracy of the NWM streamflow estimates is significantly influenced by seasonal streamflow variation. In semi-arid and sub-humid climate zones, the NWM streamflow estimates displayed good agreement with in-situ observations for spring season results. In humid regions, streamflow estimates were more accurate during fall and winter. For arid regions, the model demonstrated better agreement with in-situ observations during the fall season.

Figure 7 shows that, compared to natural flow, the NWM streamflow estimates for regulated flow exhibit lower accuracy in terms of PB for the different climate regions except humid areas. For instance, the median PB values for all seasons over the arid, semi-arid, and semi-humid regions obtained for natural flow simulations (Figure 6) were remarkably lower than those obtained for the regulated flow simulations (Figure 7). The NWM-regulated flow estimates yielded the poorest PB results, predominantly during the fall season, across most of the studied climate zones. The NWM regulated flow estimates demonstrated the most favorable agreement with in-situ measurements in humid regions, exhibiting low PB values throughout the various seasons. Conversely, the model performance deteriorated in arid regions, with a tendency to overestimate streamflow, as evidenced by positive Q1 and median PB values across different seasons. It is crucial to emphasize that, in semi-arid and sub-humid regions, the model error in terms of PB during the spring season was comparatively lower than in other seasons. The categorization of model PB results as a function of climate types across different seasons revealed a more distinct pattern in the spatial and temporal distribution of model error. For both natural and regulated flow simulations, the model exhibited commendable performance in streamflow estimation for humid and sub-humid regions. However, a diminished agreement between the simulated and observed streamflow was observed in arid and semi-arid regions.

The obtained results shown in Figure 8 unveiled analogous streamflow trend patterns for the observed and simulated streamflow data from the years 1979 to 2020. The results showed a prevailing increasing trend in NERFC, MARFC, NCRFC, MBRFC, and MRRFC. Conversely, a dominant decreasing trend was observed in the western and southern portions of the CONUS, particularly for stations situated in the NWRFC, CNRFC, CBRFC, and WGRFC. However, this analysis primarily aims to compare results obtained from USGS streamflow records with NWM streamflow simulations. In the case of the NERFC, a significant discrepancy was observed between the MMK results for the in-situ observation series and the simulated series. The MMK outcomes for the NWM streamflow series exhibited a predominance of decreasing trends in the region, with 40% of the forecast point series revealing an opposing trend direction (decreasing trend) in comparison to the in-situ measurement trend direction (increasing trend). It is important to note that the NWM underestimated streamflow values for those simulations (negative PB values).

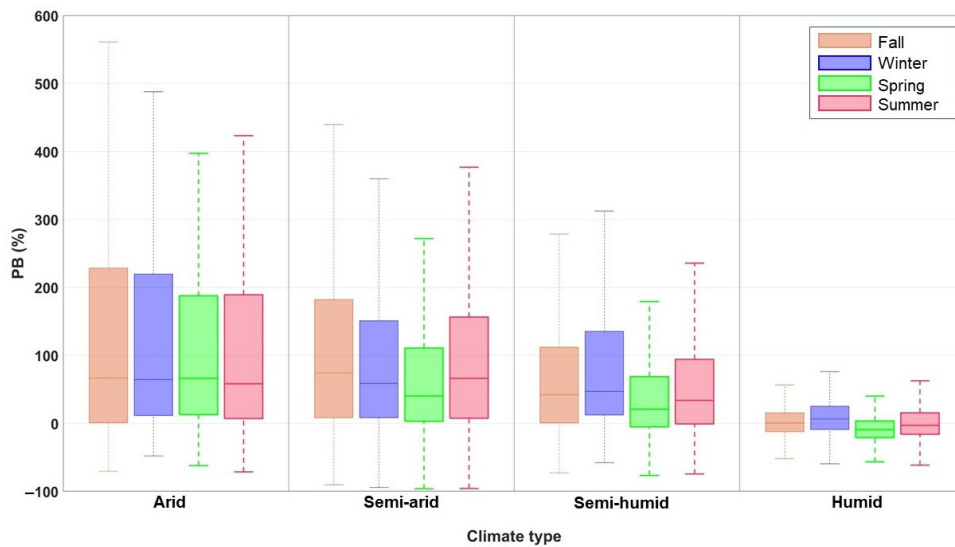


Figure 7. Box plot showing seasonal PB variation for regulated flow as a function of climate type.

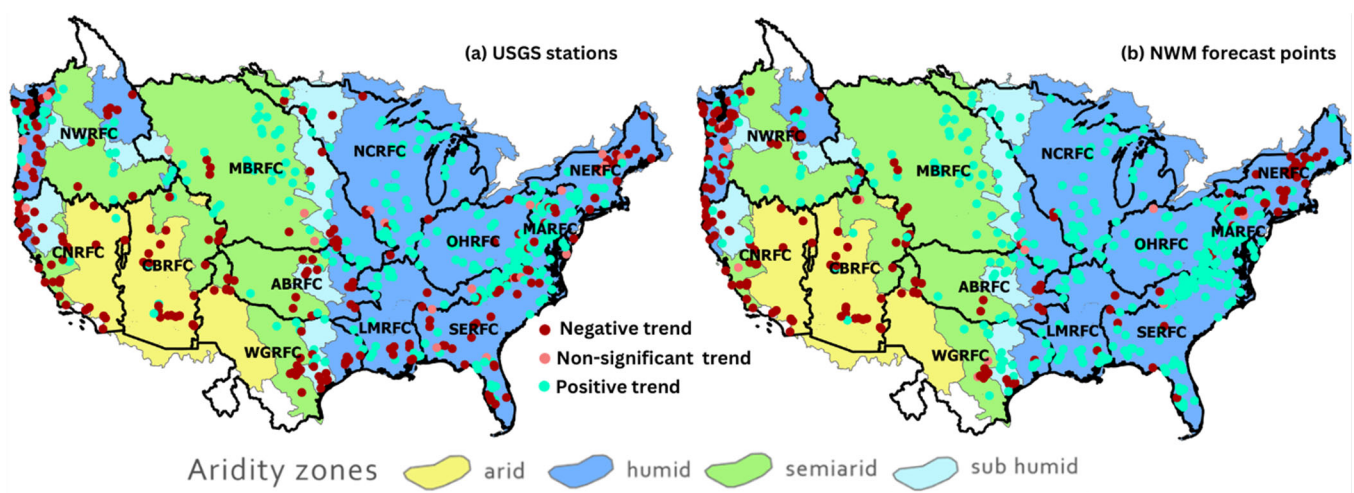


Figure 8. Spatial distribution of Z-MMK results for natural flow stations applied to in-situ observations and NWM retrospective data.

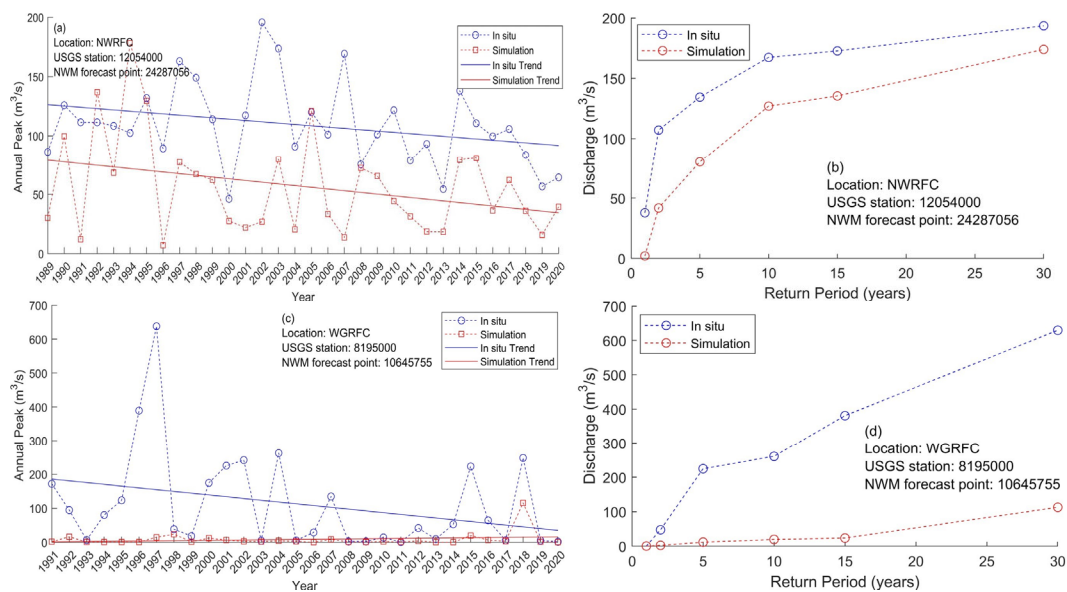
The opposite case was observed in parts of the SERFC, where in-situ observations disclosed a prevailing decreasing trend, while the NWM simulation displayed a dominant increasing trend. In those locations, PB results indicated a tendency of the model to overestimate streamflow. It is also pertinent to mention that distinct trend results for the NERFC and SERFC cases were discovered in forecast points within channels with a stream order of three or higher. The obtained results align with the findings presented in Figure 4, where forecast points with higher stream order exhibited larger PB results, leading to an inadequate representation of the temporal evolution of streamflow in high-order streams.

For other RFCs where streamflow values were overestimated, such as the MBRFC, ABRFC, and WGRFC, trend results obtained from the NWM series demonstrated a dominant increasing trend in those regions. However, in-situ measurement results revealed a prevailing decreasing trend in the ABRFC and WGRFC. The obtained results could be elucidated by the positive PB and low correlation for forecast points located in these regions (Figure 2c,i). Overall, the MMK test results disclosed a satisfactory agreement between the temporal evolution of the observed and simulated natural streamflow series for 74% of the studied dataset.



Regions exhibiting dominant negative streamflow trends and negative PB values pose significant concerns for products generated from the retrospective data, such as annual exceedance probability threshold values at forecast points, which are utilized to trigger flood inundation mapping in the operational NWM. To investigate this issue, we selected two representative stations with negative trends and negative PB values, characterized by close Z-MMK values, to retrieve the annual peak series and flood return periods. Initially, we extracted annual peak discharge from the daily streamflow time series and temporally matched these values with the corresponding values from the NWM simulations. It is crucial to note that the annual peak discharge values from the NWM simulation occurred within a 3-day time window of the USGS measurements, accounting for potential discrepancies in runoff generation lag time representation. In other words, within the same year, the NWM simulation could present a higher value than the selected value, which we consider a biased estimation of streamflow that could be attributed to biases in the forcing data.

As illustrated in Figure 9, the annual peak series and flood return period demonstrate that, although the NWM simulation captures the trend of annual peak discharge values, there is a significant concern regarding the model’s representation of flood return periods. The 15-year return period flood value, which is used to trigger flood mapping in the operational model, is underestimated by the NWM simulation, as shown in Figure 9b,d. This finding highlights the issue of the NWM underestimating peak flood values, necessitating bias correction of the NWM retrospective data. Several factors can contribute to the NWM model’s underestimation of peak discharge values. In addition to input data inaccuracies, model parameterization plays a substantial role in generating negative biases. Initial investigations into the source of the bias should primarily focus on land use and land cover data. The impact of updating the land use and land cover data while generating the retrospective dataset remains uncertain. Inaccurate land use and land cover data can lead to the misrepresentation of a catchment’s response to precipitation events, thereby causing an underestimation of peak discharge values. Furthermore, an incorrect representation of channel routing and floodplain storage can significantly impact the model’s ability to accurately simulate peak discharge values.



**Figure 9.** Annual peak series and flood return period series (up to 30 years) for selected stations from the NWRFC and WGRFC. Annual peak series (a,c) were used to estimate the 1, 2, 5, 10, 15, and 30-year return period discharge values, marked as dots in (b,d).



#### 4. Discussion

In this study, the evaluation of the National Water Model retrospective data concentrated on the model's streamflow estimation performance across the regional and temporal variations in the contiguous United States. The model bias and bias variation were scrutinized using multiple metrics, emphasizing percent bias (PB) variation for distinct climate classes. This approach facilitated a more standardized portrayal of the model's performance across diverse climate regions.

In this section, we delve into the sources of uncertainty associated with the NWM streamflow simulation, encompassing input and output data, model structure, and model parameters. In the present study, the spatial analysis facilitated the delineation of underperforming catchments, predominantly situated along the high plains and desert southwest. The analysis of the spatial dispersion of model bias shows that regions characterized by elevated non-seasonal streamflow variability and frequent precipitation occurrences, such as the Southeast River Forecast Center and Northwest River Forecast Center, exhibited low variability errors. In contrast, arid and semi-arid regions demonstrated heightened variability errors. Catchments dominated by snowpack, including those within the Missouri Basin River Forecast Center, and areas with a more distinct seasonal cycle, such as the California–Nevada River Forecast Center, also revealed substantial variability errors.

It is widely acknowledged that hydrological models are susceptible to errors in hydrometeorological input data [59]. Consequently, biases in forcing variables, particularly precipitation and near-surface air temperature, yield corresponding biases in model output. Lahmers et al. (2019) demonstrated that refining forcing data by integrating modeled atmospheric forcing with gauge-based precipitation measurements mitigated evapotranspiration biases and augmented the simulation of streamflow behavior [60]. In a separate study, Viterbo et al. (2020) conducted an event-based model assessment and deduced that the NWM streamflow bias is influenced by both model errors and input errors [61]. Furthermore, Garousi–Nejad and Tarboton (2022) identified a prevalent tendency for the NWM to underestimate Snow Water Equivalent (SWE) as a result of hydrologic process representation and hydrometeorological input errors [62]. Their investigation revealed that the incorporation of observed precipitation and bias-corrected air temperature data ameliorated the general downward bias in NWM–SWE estimations. The model is also certainly sensitive to potential biases in soil moisture especially in watersheds with natural flow. The advent of satellite missions like the NASA SMAP that hold a potential to enhance streamflow forecast through data assimilation. This requires an exhaustive calibration and validation of the sensor's estimates through various field campaigns [63,64].

For the NWM V2.1 retrospective simulations, the Analysis of Period of Record for Calibration (AORC) dataset was employed as forcing data. Although the blended product has high temporal and spatial resolutions and contains over a dozen individual rainfall datasets, there is need to evaluate the product performance over the CONUS while taking into account the seasonal variation of the product accuracy. For instance, Kim and Villarini (2022) assessed the AORC rainfall across Louisiana while focusing on the precipitation accuracy associated with Tropical Cyclone (TC) and non-TC conditions [65]. The study showed that AORC performs better for the TC period compared to the non-TC ones. Thus, for regions with dominant small precipitation amounts, it is expected that the bias in AORC data would affect the soil condition and infiltration rates in the model simulation, as well as impact the water storage and the baseflow estimations. Hong et al. (2022) analyzed the AORC performance over the Great Lakes basin, and the analysis showed compared to 632 gauge stations, the product tends to overestimate daily precipitation and underestimate heavy rains with a notable larger bias value in cold months [66]. The study also showed that the interannual change in AORC precipitation has divergent years while compared to other gridded precipitation products. To our knowledge, there are no continental scale assessments of AORC forcing hourly data, including precipitation, air temperature, specific humidity, surface pressure, radiation, and near-surface wind, which may have significant impact on accurately simulating evapotranspiration, snowmelt, and runoff generation

processes. A better understanding of the NWM bias inherited from the forcing data would include the assessment of the AORC forcing.

The NWM performance is contingent upon the quality and accuracy of the static input data, which fundamentally represent catchment characteristics and significantly influence runoff generation, flow direction, and the configuration of the stream network. For instance, land use and land-cover data are crucial for determining the spatial distribution of vegetation, urban areas, and other surface types within a catchment, thereby directly impacting parameters associated with infiltration, evapotranspiration, and surface runoff generation. In the case of the NWM retrospective simulation, it is unclear which land cover dataset was used. However, the usage of static datasets similar to the retrospective runs of Versions 2.0 and 2.1 may result in biased simulations for periods when the used land cover layer does not match with the land cover of the simulation period. Similarly, using static soil data for a long-term simulation, encompassing properties such as texture, depth, and hydraulic conductivity, play a pivotal role in shaping infiltration capacity, water retention, and groundwater recharge processes throughout the catchment. Lastly, the accurate representation of river network and geometry, including parameters such as river width, depth, and slope, is indispensable for modeling flow routing and overland flow processes. For example, Ghanghas et al. (2022) assessed the reliability of Synthetic Rating Curves (SRC) across the CONUS by comparing them with the rating curves from USGS gauges [67]. The identified errors in the SRC were attributed to topographical factors and inherent assumptions, such as employing reach-averaged channel properties, constant channel roughness, and uniform flow. These assumptions are consistent with those utilized in the NWM river network (NHDPlus). It is imperative to note that the NWM retrospective simulations serve as a basis for obtaining the 15-year return period flood discharge values across the CONUS. Subsequently, these discharge values are employed to initiate the flood inundation mapping process within the operational NWM framework. Consequently, inaccuracies in flood maps generated during forecasting may stem from errors in the estimated 15-year return period discharge values and/or the underlying SRC, emphasizing the importance of refining these components for improved flood prediction and mapping.

The uncertainty inherent in model streamflow simulations may also stem from errors in the in-situ measurements employed for model calibration and output assessment. While the gathered USGS streamflow data is verified through statistical and deterministic methods, such as flagging extreme discharge values, correlating independent observations, and conducting continuity tests, inherent errors in streamflow records may arise from alterations in stage–discharge relationships, backwater effects due to ice or debris, or equipment malfunctions. Thus, the correction of ice-affected streamflow is critical for evaluating models' simulations in cold regions [68]. Nevertheless, it is important to note that with advancements in estimating data uncertainty, the impact of such errors may be negligible. Conversely, it is imperative to emphasize that the NWM retrospective simulations scrutinized in this study do not incorporate streamflow observation nudging. As a result, the performance of the operational model version may be superior in localized areas where USGS observations are available. Despite this, the reported performance remains pertinent in identifying limitations in model parameterization and underlining the necessity for further refinement.

In this study, several key findings and recommendations emerge from the analysis of the model's performance. First, it is evident that incorporating soil moisture data assimilation in humid regions, where frequent precipitation events are common, is essential to improve the model's accuracy, as soil moisture dynamics play a crucial role in partitioning precipitation into runoff and infiltration, thereby directly impacting streamflow estimates. We would like to emphasize that the NWM's soil moisture estimates are grid-based with a spatial resolution of 1 km. Consequently, a faithful representation of soil moisture within these grid cells would contribute substantially to a more accurate estimation of streamflow in such regions. In addition, it is also important to note that the operational

version of the NWM actively nudging streamflow data from USGS stations. This nudging process is critical in updating the streamflow estimates within the model, but it does not affect the other hydrological components. Consequently, continuous and accurate soil moisture data assimilation would not only maintain the up-to-date status of soil moisture but also significantly enhance the model's streamflow estimates, especially in predicting and managing flood events. These collective measures, in effect, would amplify the overall reliability and accuracy of the NWM's streamflow estimates. Second, in cold regions, the assimilation of river ice data is necessary to better capture the effects of ice processes, such as freeze-up and break-up events, on streamflow dynamics, ultimately leading to more accurate predictions in these areas. Third, the uncertainty associated with the lack of knowledge regarding reservoir operation rules has been identified as a potential source of error in the model's performance.

The challenges posed by river systems in the context of hydrological modeling necessitate a more sophisticated and adaptive approach. To tackle these complexities, we advocate for the exploration and integration of advanced computational methods, such as physics-informed machine learning and data mining, into the operational procedures of the NWM. Physics-informed machine-learning algorithms, which couple the power of data-driven models with the underlying physical principles governing the system, can provide insightful and reliable representations of reservoir operation rules. In this context, these techniques can help leverage the existing wealth of observational and operational data, enabling the extraction of complex relationships and patterns that traditional methods may struggle to discern. Moreover, enhancing the NWM's capabilities through data-assimilation techniques could also significantly improve the simulation of streamflow estimates in regulated systems. Data assimilation, which blends model predictions with observed data to improve forecast accuracy and reduce uncertainty, has emerged as a powerful tool in hydrological modeling. Incorporating these inferred operation rules into the NWM would allow for a more faithful representation of the regulated systems, thereby significantly improving the accuracy of the model's streamflow estimates. In this way, these advanced techniques could serve as catalysts in boosting the model's overall performance, while simultaneously offering a more nuanced understanding of the hydrological processes at play in these regulated systems. These proposed recommendations, while demanding in terms of computational resources and methodological implementation, are projected to considerably enhance the capacity of the NWM to provide more reliable and accurate hydrological forecasts, ultimately benefiting a range of stakeholders in water resources management, flood prediction, and environmental conservation.

On the other hand, as we move towards the implementation of the Next Generation Modeling System (NextGen) for the National Water Model, opportunities for enhanced streamflow estimation, particularly in regions exhibiting poor performance with the current NWM, are anticipated [69]. The NextGen framework, inherently model agnostic, offers immense flexibility, permitting the application of disparate methods or models to simulate specific hydrological fluxes. These simulations can be adjusted for diverse temporal and spatial scales, thereby accommodating unique hydrological characteristics across regions. The collaborative ethos embedded in the open-source philosophy of the NextGen framework augments methodological development within the scientific community. This atmosphere of collective engagement accelerates advancements in hydrological modeling techniques, promoting knowledge exchange and bolstering scientific growth. Moreover, this open-source approach empowers agencies to integrate proven models effectively within the NextGen framework, a feature that is notably advantageous for regions wherein certain models have demonstrated exceptional performance [69]. Overall, the evolution towards the NextGen framework heralds a promising era for the NWM, offering pathways to surmount current limitations in streamflow estimation and substantially enhance the accuracy and reliability of hydrological predictions across the Continental United States. This transition signifies a noteworthy advancement in our ongoing journey towards comprehensive and adaptable hydrological modeling.

## 5. Conclusions

This study aimed to assess the NWM streamflow estimates using a multi-decade retrospective dataset across the contiguous United States. The performance of the models was found to vary regionally, with the primary factors influencing this variation being aridity and precipitation alternation and chaotic dynamics, snowmelt contribution, and runoff seasonality. Model static input data, such as river network and land cover data, were also identified as potential sources of bias. Our findings revealed that 57% of natural flow forecast points exhibited KGE values greater than 0.5, while 43% of regulated flow forecast points surpassed a KGE value of 0.5. The model performance was notably poor in the Great Plains region, while it demonstrated better performance in estimating streamflow in the Northeastern and Northwestern regions. Ultimately, the model bias was found to be governed by both the representation of hydrologic processes and the presence of forcing errors. Our findings revealed that the accuracy of the NWM streamflow estimates is significantly influenced by seasonal streamflow variation. In semi-arid and sub-humid climate zones, the NWM streamflow estimates displayed good agreement with in-situ observations for spring season results. In humid regions, streamflow estimates were more accurate during fall and winter, while for arid regions, the model demonstrated better agreement with in-situ observations during the fall season. These insights can inform future model improvements and contribute to more accurate and reliable hydrologic predictions across the contiguous United States. Future work will focus on leveraging the outcomes of the assessment of the performance of the NWM in this study to guide the evaluation of the streamflow of the operational NWM. In addition, future work will involve the bias correction of streamflow from the retrospective simulation and its frequency analysis to better estimate the annual exceedance probability threshold values at the forecast points, which are used to trigger flood inundation mapping, anticipated in late 2023, using the NWM V3.0. On the other hand, the transition towards the NextGen modeling system holds promising prospects for improved streamflow estimations, particularly in areas of suboptimal current performance of the NWM. Its model-agnostic structure and open-source philosophy foster innovation and integration of proven models. This transformative shift signifies a critical advancement towards a more robust and adaptable hydrological modeling system.

**Author Contributions:** Conceptualization, M.A., M.T. and T.B.M.J.O.; Methodology, M.A., M.T. and T.B.M.J.O.; Software, M.A.; Formal analysis, M.A.; Investigation, M.A., M.T. and T.B.M.J.O.; Resources, M.A. and M.T.; Data curation, M.A.; Writing—original draft, M.A. and M.T.; Supervision, M.T. and T.B.M.J.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data for the National Water Model (NWM) streamflow retrospective simulations are openly available in the following website: <https://registry.opendata.aws/nwm-archive/> (accessed on 17 June 2023). The in-situ streamflow data from the United States Geological Survey (USGS) are openly available in the following website: <https://waterdata.usgs.gov/nwis/sw> (accessed on 17 June 2023). Codes developed for this study are publicly available in the following HydroShare repository: Abdelkader, M., J. H. Bravo Mendez (2023). NWM version 2.1 model output data retrieval, HydroShare, <https://doi.org/10.4211/hs.c4c9f0950c7a42d298ca25e4f6ba5542> (accessed on 17 June 2023).

**Acknowledgments:** The authors would like to thank the editor and the anonymous reviewers for their comments that improved the quality of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Makwana, J.J.; Tiwari, M.K. Intermittent Streamflow Forecasting and Extreme Event Modelling Using Wavelet Based Artificial Neural Networks. *Water Resour. Manag.* **2014**, *28*, 4857–4873. [CrossRef]
2. Bai, T.; Chang, J.; Chang, F.J.; Huang, Q.; Wang, Y.; Chen, G. Synergistic Gains from the Multi-Objective Optimal Operation of Cascade Reservoirs in the Upper Yellow River Basin. *J. Hydrol.* **2015**, *523*, 758–767. [CrossRef]



3. Pagano, T.C.; Wood, A.W.; Ramos, M.-H.; Cloke, H.L.; Pappenberger, F.; Clark, M.P.; Cranston, M.; Kavetski, D.; Mathevet, T.; Sorooshian, S.; et al. Challenges of Operational River Forecasting. *J. Hydrometeorol.* **2014**, *15*, 1692–1707. [[CrossRef](#)]
4. Li, J.; Wang, Z.; Wu, X.; Xu, C.Y.; Guo, S.; Chen, X. Toward Monitoring Short-Term Droughts using a Novel Daily Scale, Standardized Antecedent Precipitation Evapotranspiration Index. *J. Hydrometeorol.* **2020**, *21*, 891–908. [[CrossRef](#)]
5. Chiew, F.H.S.; Zhou, S.L.; McMahon, T.A. Use of Seasonal Streamflow Forecasts in Water Resources Management. *J. Hydrol.* **2003**, *270*, 135–144. [[CrossRef](#)]
6. Li, X.; Rankin, C.; Gangrade, S.; Zhao, G.; Lander, K.; Voisin, N.; Shao, M.; Morales-Hernández, M.; Kao, S.C.; Gao, H. Evaluating Precipitation, Streamflow, and Inundation Forecasting Skills during Extreme Weather Events: A Case Study for an Urban Watershed. *J. Hydrol.* **2021**, *603*, 127126. [[CrossRef](#)]
7. Sushanth, K.; Mishra, A.; Mukhopadhyay, P.; Singh, R. Real-Time Streamflow Forecasting in a Reservoir-Regulated River Basin using Explainable Machine Learning and Conceptual Reservoir Module. *Sci. Total Environ.* **2023**, *861*, 160680. [[CrossRef](#)]
8. Anghileri, D.; Voisin, N.; Castelletti, A.; Pianosi, F.; Nijssen, B.; Lettenmaier, D.P. Value of Long-Term Streamflow Forecasts to Reservoir Operations for Water Supply in Snow-Dominated River Catchments. *Water Resour. Res.* **2016**, *52*, 4209–4225. [[CrossRef](#)]
9. Cassagnole, M.; Ramos, M.H.; Zalachori, I.; Thirel, G.; Garçon, R.; Gailhard, J.; Ouillon, T. Impact of the Quality of Hydrological Forecasts on the Management and Revenue of Hydroelectric Reservoirs—A Conceptual Approach. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1033–1052. [[CrossRef](#)]
10. Kao, S.C.; Sale, M.J.; Ashfaq, M.; Uria Martinez, R.; Kaiser, D.P.; Wei, Y.; Diffenbaugh, N.S. Projecting Changes in Annual Hydropower Generation using Regional Runoff Data: An Assessment of the United States Federal Hydropower Plants. *Energy* **2015**, *80*, 239–250. [[CrossRef](#)]
11. Sivakumar, B. Nonlinear Dynamics and Chaos in Hydrologic Systems: Latest Developments and a Look Forward. *Stoch. Environ. Res. Risk Assess.* **2009**, *23*, 1027–1036. [[CrossRef](#)]
12. Perrin, C.; Oudin, L.; Andreassian, V.; Rojas-Serna, C.; Michel, C.; Mathevet, T. Impact of Limited Streamflow Data on the Efficiency and the Parameters of Rainfall-Runoff Models. *Hydrol. Sci. J.* **2007**, *52*, 131–151. [[CrossRef](#)]
13. Butts, M.B.; Payne, J.T.; Kristensen, M.; Madsen, H. An Evaluation of the Impact of Model Structure on Hydrological Modelling Uncertainty for Streamflow Simulation. *J. Hydrol.* **2004**, *298*, 242–266. [[CrossRef](#)]
14. Orth, R.; Staudinger, M.; Seneviratne, S.I.; Seibert, J.; Zappa, M. Does Model Performance Improve with Complexity? A Case Study with Three Hydrological Models. *J. Hydrol.* **2015**, *523*, 147–159. [[CrossRef](#)]
15. Hung, C.L.J.; James, L.A.; Carbone, G.J.; Williams, J.M. Impacts of Combined Land-Use and Climate Change on Streamflow in Two Nested Catchments in the Southeastern United States. *Ecol. Eng.* **2020**, *143*, 105665. [[CrossRef](#)]
16. Sunde, M.G.; He, H.S.; Hubbard, J.A.; Urban, M.A. An Integrated Modeling Approach for Estimating Hydrologic Responses to Future Urbanization and Climate Changes in a Mixed-Use Midwestern Watershed. *J. Environ. Manag.* **2018**, *220*, 149–162. [[CrossRef](#)] [[PubMed](#)]
17. Zhou, Q.; Leng, G.; Su, J.; Ren, Y. Comparison of Urbanization and Climate Change Impacts on Urban Flood Volumes: Importance of Urban Planning and Drainage Adaptation. *Sci. Total Environ.* **2019**, *658*, 24–33. [[CrossRef](#)] [[PubMed](#)]
18. Abbas, S.A.; Xuan, Y. Impact of Precipitation Pre-Processing Methods on Hydrological Model Performance using High-Resolution Gridded Dataset. *Water* **2020**, *12*, 840. [[CrossRef](#)]
19. Shafqat Mehboob, M.; Kim, Y.; Lee, J.; Eidhammer, T. Quantifying the Sources of Uncertainty for Hydrological Predictions with WRF-Hydro over the Snow-Covered Region in the Upper Indus Basin, Pakistan. *J. Hydrol.* **2022**, *614*, 128500. [[CrossRef](#)]
20. Segond, M.L.; Wheeler, H.S.; Onof, C. The Significance of Spatial Rainfall Representation for Flood Runoff Estimation: A Numerical Evaluation Based on the Lee Catchment, UK. *J. Hydrol.* **2007**, *347*, 116–131. [[CrossRef](#)]
21. Gu, P.; Wang, G.; Liu, G.; Wu, Y.; Liu, H.; Jiang, X.; Liu, T. Evaluation of Multisource Precipitation Input for Hydrological Modeling in an Alpine Basin: A Case Study from the Yellow River Source Region (China). *Hydrol. Res.* **2022**, *53*, 314–335. [[CrossRef](#)]
22. Bourdin, D.R.; Fleming, S.W.; Stull, R.B. Streamflow Modelling: A Primer on Applications, Approaches and Challenges. *Atmos. Ocean* **2012**, *50*, 507–536. [[CrossRef](#)]
23. Nijssen, B.; Schnur, R.; Lettenmaier, D.P. Global Retrospective Estimation of Soil Moisture using the Variable Infiltration Capacity Land Surface Model, 1980–93. *J. Clim.* **2001**, *14*, 1790–1808. [[CrossRef](#)]
24. Milly, P.C.D.; Shmakin, A.B. Global Modeling of Land Water and Energy Balances. Part II: Land-Characteristic Contributions to Spatial Variability. *J. Hydrometeorol.* **2002**, *3*, 301–310. [[CrossRef](#)]
25. Hales, R.C.; Nelson, E.J.; Souffront, M.; Gutierrez, A.L.; Prudhomme, C.; Kopp, S.; Ames, D.P.; Williams, G.P.; Jones, N.L. Advancing Global Hydrologic Modeling with the GEOGloWS ECMWF Streamflow Service. *J. Flood Risk Manag.* **2022**, *16*, 12859. [[CrossRef](#)]
26. McEnery, J.; Ingram, J.; Duan, Q.; Adams, T.; Anderson, L. NOAA’S advanced hydrologic prediction service: Building pathways for better science in water forecasting. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 375–386. [[CrossRef](#)]
27. National Research Council. *Toward a New Advanced Hydrologic Prediction Service (AHPS)*; The National Academies Press: Washington, DC, USA, 2006; ISBN 0309101441. [[CrossRef](#)]
28. Gochis, D.J.; Barlage, M.; Dugger, A.; Fitzgerald, K.; Karsten, L.; Mcallister, M.; Mccreight, J.; Mills, J.; Rafieeiniasab, A.; Read, L.; et al. *The WRF-Hydro Modeling System Technical Description, Version 5.0*; NCAR Technical Note; UCAR: Boulder, CO, USA, 2018.
29. Office of Water Prediction. Available online: <https://water.noaa.gov/about/nwm> (accessed on 6 May 2023).



30. Wagener, T.; Sivapalan, M.; Troch, P.A.; McGlynn, B.L.; Harman, C.J.; Gupta, H.V.; Kumar, P.; Rao, P.S.C.; Basu, N.B.; Wilson, J.S. The Future of Hydrology: An Evolving Science for a Changing World. *Water Resour. Res.* **2010**, *46*, W05301. [CrossRef]
31. Niu, G.Y.; Yang, Z.L.; Mitchell, K.E.; Chen, F.; Ek, M.B.; Barlage, M.; Kumar, A.; Manning, K.; Niyogi, D.; Rosero, E.; et al. The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 1. Model Description and Evaluation with Local-Scale Measurements. *J. Geophys. Res. Atmos.* **2011**, *116*, 1–19. [CrossRef]
32. Abdelkader, M.; Bravo Mendez, J.H. NWM Version 2.1 Model Output Data Retrieval. *HydroShare*. 2023. Available online: <https://www.hydroshare.org/resource/c4c9f0950c7a42d298ca25e4f6ba5542/> (accessed on 17 June 2023).
33. NOAA National Water Model CONUS Retrospective Dataset—Registry of Open Data on AWS. Available online: <https://registry.opendata.aws/nwm-archive/> (accessed on 6 May 2023).
34. USGS Water Mission Area NSDI Node. Available online: [https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII\\_Sept2011.xml](https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml) (accessed on 6 May 2023).
35. Database of Modified Routing for NHDPlus, Version 2.1; Flowlines: ENHDPlusV2\_us—ScienceBase-Catalog. Available online: <https://www.sciencebase.gov/catalog/item/5b92790be4b0702d0e809fe5> (accessed on 6 May 2023).
36. USGS Surface—Water Data for the Nation. Available online: <https://waterdata.usgs.gov/nwis/sw> (accessed on 6 May 2023).
37. Hamilton, A.S.; Moore, R.D. Quantifying Uncertainty in Streamflow Records. *Can. Water Resour. J.* **2012**, *37*, 3–21. [CrossRef]
38. Giuntoli, I.; Renard, B.; Vidal, J.P.; Bard, A. Low Flows in France and Their Relationship to Large-Scale Climate Indices. *J. Hydrol.* **2013**, *482*, 105–118. [CrossRef]
39. Yerdelen, C.; Abdelkader, M. Hydrological Data Trend Analysis with Wavelet Transform. *Comptes Rendus L'Academie Bulg. Sci.* **2021**, *74*, 1194–1202. [CrossRef]
40. Abdelkader, M.; Yerdelen, C. Hydrological Drought Variability and Its Teleconnections with Climate Indices. *J. Hydrol.* **2022**, *605*, 127290. [CrossRef]
41. World Atlas of Desertification: Second Edition. Available online: <https://wedocs.unep.org/20.500.11822/30300> (accessed on 6 May 2023).
42. Xu, T.; Guo, Z.; Xia, Y.; Ferreira, V.G.; Liu, S.; Wang, K.; Yao, Y.; Zhang, X.; Zhao, C. Evaluation of Twelve Evapotranspiration Products from Machine Learning, Remote Sensing and Land Surface Models over Conterminous United States. *J. Hydrol.* **2019**, *578*, 124105. [CrossRef]
43. Zhang, B.; Xia, Y.; Long, B.; Hobbins, M.; Zhao, X.; Hain, C.; Li, Y.; Anderson, M.C. Evaluation and Comparison of Multiple Evapotranspiration Data Models over the Contiguous United States: Implications for the next Phase of NLDAS (NLDAS-Testbed) Development. *Agric. For. Meteorol.* **2020**, *280*, 107810. [CrossRef]
44. Knoben, W.J.M.; Freer, J.E.; Woods, R.A. Technical Note: Inherent Benchmark or Not? Comparing Nash-Sutcliffe and Kling-Gupta Efficiency Scores. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4323–4331. [CrossRef]
45. Waseem, M.; Mani, N.; Andiego, G.; Usman, M. A Review of Criteria of Fit for Hydrological Models. *Int. Res. J. Eng. Technol.* **2008**, *9001*, 1765.
46. Liu, D. A Rational Performance Criterion for Hydrological Model. *J. Hydrol.* **2020**, *590*, 125488. [CrossRef]
47. Lamontagne, J.R.; Barber, C.A.; Vogel, R.M. Improved Estimators of Model Performance Efficiency for Skewed Hydrologic Data. *Water Resour. Res.* **2020**, *56*, e2020WR027101. [CrossRef]
48. Yuemei, H.; Xiaoqin, Z.; Jianguo, S.; Jina, N. Conduction between Left Superior Pulmonary Vein and Left Atria and Atria Fibrillation under Cervical Vagal Trunk Stimulation. *Colomb. Med.* **2008**, *39*, 227–234.
49. de Salis, H.H.C.; da Costa, A.M.; Vianna, J.H.M.; Schuler, M.A.; Künne, A.; Fernandes, L.F.S.; Pacheco, F.A.L. Hydrologic Modeling for Sustainable Water Resources Management in Urbanized Karst Areas. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2542. [CrossRef]
50. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
51. Hamed, K.H.; Ramachandra Rao, A. A Modified Mann-Kendall Trend Test for Autocorrelated Data. *J. Hydrol.* **1998**, *204*, 182–196. [CrossRef]
52. Naizghi, M.S.; Ouarda, T.B.M.J. Teleconnections and Analysis of Long-Term Wind Speed Variability in the UAE. *Int. J. Clim.* **2017**, *37*, 230–248. [CrossRef]
53. Vazifekhhah, S.; Kahya, E. Hydrological and Agricultural Droughts Assessment in a Semi-Arid Basin: Inspecting the Teleconnections of Climate Indices on a Catchment Scale. *Agric. Water Manag.* **2019**, *217*, 413–425. [CrossRef]
54. Yerdelen, C.; Tastan, M.; Abdelkader, M. Assessment of Trend Analysis Methods for Annual Streamflow. *Environ. Eng. Manag. J.* **2022**, *21*, 569–577.
55. Singh, V.P. *Elementary Hydrology*; Pearson: London, UK, 1992; 973p.
56. Lahmers, T.M.; Hazenberg, P.; Gupta, H.; Castro, C.; Gochis, D.; Dugger, A.; Yates, D.; Read, L.; Karsten, L.; Wang, Y.H. Evaluation of NOAA National Water Model Parameter Calibration in Semiarid Environments Prone to Channel Infiltration. *J. Hydrometeorol.* **2021**, *22*, 2939–2969. [CrossRef]
57. Srivastava, A.; Rodriguez, J.F.; Saco, P.M.; Kumari, N.; Yetemen, O. Global Analysis of Atmospheric Transmissivity using Cloud Cover, Aridity and Flux Network Datasets. *Remote Sens.* **2021**, *13*, 1716. [CrossRef]
58. Heidari, H.; Arabi, M.; Warziniack, T.; Kao, S.C. Assessing Shifts in Regional Hydroclimatic Conditions of U.S. River Basins in Response to Climate Change over the 21st Century. *Earth's Futur.* **2020**, *8*, e2020EF001657. [CrossRef]

59. Wu, X.; Guo, S.; Qian, S.; Wang, Z.; Lai, C.; Li, J.; Liu, P. Long-Range Precipitation Forecast Based on Multipole and Preceding Fluctuations of Sea Surface Temperature. *Int. J. Clim.* **2022**, *42*, 8024–8039. [[CrossRef](#)]
60. Lahmers, T.M.; Gupta, H.; Castro, C.L.; Gochis, D.J.; Yates, D.; Dugger, A.; Goodrich, D.; Hazenberg, P. Enhancing the Structure of the WRF-Hydro Hydrologic Model for Semiarid Environments. *J. Hydrometeorol.* **2019**, *20*, 691–714. [[CrossRef](#)]
61. Viterbo, F.; Mahoney, K.; Read, L.; Salas, F.; Bates, B.; Elliott, J.; Cosgrove, B.; Dugger, A.; Gochis, D.; Cifelli, R. A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood. *J. Hydrometeorol.* **2020**, *21*, 475–499. [[CrossRef](#)]
62. Garousi-Nejad, I.; Tarboton, D.G. A Comparison of National Water Model Retrospective Analysis Snow Outputs at Snow Telemetry Sites across the Western United States. *Hydrol. Process.* **2022**, *36*, e14469. [[CrossRef](#)]
63. Karamouz, M.; Alipour, R.S.; Roohinia, M.; Fereshtehpour, M. A Remote Sensing Driven Soil Moisture Estimator: Uncertain Downscaling with Geostatistically Based Use of Ancillary Data. *Water Resour. Res.* **2022**, *58*, e2022WR031946. [[CrossRef](#)]
64. Abdelkader, M.; Temimi, M.; Colliander, A.; Cosh, M.H.; Kelly, V.R.; Lakhankar, T.; Fares, A. Assessing the Spatiotemporal Variability of SMAP Soil Moisture Accuracy in a Deciduous Forest Region. *Remote Sens.* **2022**, *14*, 3329. [[CrossRef](#)]
65. Kim, H.; Villarini, G. Evaluation of the Analysis of Record for Calibration (AORC) Rainfall across Louisiana. *Remote Sens.* **2022**, *14*, 3284. [[CrossRef](#)]
66. Hong, Y.; Xuan Do, H.; Kessler, J.; Fry, L.; Read, L.; Rafieei Nasab, A.; Gronewold, A.D.; Mason, L.; Anderson, E.J. Evaluation of Gridded Precipitation Datasets over International Basins and Large Lakes. *J. Hydrol.* **2022**, *607*, 127507. [[CrossRef](#)]
67. Ghanghas, A.; Dey, S.; Merwade, V. Evaluating the Reliability of Synthetic Rating Curves for Continental Scale Flood Mapping. *J. Hydrol.* **2022**, *606*, 127470. [[CrossRef](#)]
68. Chaouch, N.; Temimi, M.; Romanov, P.; Cabrera, R.; Mckillop, G.; Khanbilvardi, R. An Automated Algorithm for River Ice Monitoring over the Susquehanna River using the MODIS Data. *Hydrol. Process.* **2014**, *28*, 62–73. [[CrossRef](#)]
69. Next Gen Water Modeling Framework Prototype. Available online: <https://github.com/NOAA-OWP/ngen> (accessed on 17 June 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.