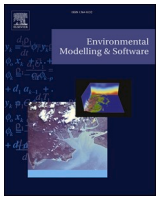
Environmental Modelling and Software 183 (2025) 106234

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Using national hydrologic models to obtain regional climate change impacts on streamflow basins with unrepresented processes

Patience Bosompemaa [a,*], Andrea Brookfield [b], Sam Zipper [a,c], Mary C. Hill [a]

[a] University of Kansas, Department of Geology, Lawrence, KS, 66045, USA
[b] University of Waterloo, Department of Earth and Environmental Sciences, Waterloo, ON, N2L 3G1, Canada
[c] Kansas Geological Survey, Lawrence, KS, 66047, USA

ARTICLE INFO

Handling Editor: Daniel P Ames

Keywords:
US Geological Survey national hydrologic model
NHM-PRMS
Bias correction
Flow duration curve (FDC)
Auto-regressive integrated moving average (ARIMA)
Arkansas river

ABSTRACT

Climate change is increasingly impacting water availability. National-scale hydrologic models simulate streamflow resulting from many important processes, but often without processes such as human water use and management activities. This work explores and tests methods to account for such omitted processes using one national-scale hydrologic model. Two bias correction methods, Flow Duration Curve (FDC) and Auto-Regressive Integrated Moving Average (ARIMA), are tested on streamflow simulated by the US Geological Survey National Hydrologic Model (NHM-PRMS), which omits irrigation pumping. A semi-arid agricultural case study is used. FDC and ARIMA perform better for correcting low and high flows, respectively. A hybrid method performs well at both low and high flows; typical Nash-Sutcliffe values increased from <-1.00 to about 0.75. Results suggest methods with which national-scale hydrologic models can be bias-corrected for omitted processes to improve regional streamflow estimates. Utility of these correction methods in simulation of future projections is discussed.

1. Introduction

Climate change is already impacting water availability for industrial and agricultural supplies (e.g., Foster et al., 2020; Ayers et al., 2022). Surface water and groundwater resources in heavily irrigated landscapes are facing the compound stressors of climate change and increased agricultural water use. In many watersheds worldwide, this has caused water demands to exceed water supply (Gleeson et al., 2012). Further supply depletion and more extreme high streamflows are projected in the future: at 4 °C global warming by the end of the 21st century, approximately 10% of the global land area is projected to face increasing high extreme streamflow and decreasing low extreme streamflow, affecting over 2.1 billion people (IPCC, 2023). This occurs as the warmer temperatures increase evapotranspiration (Condon et al., 2020) and change precipitation patterns (Patterson et al., 2022; Wigley and Jones, 1985; Zhang et al., 2011a). Resulting reduced surface-water storage and groundwater recharge would affect water availability. This work focuses on how to obtain insight about likely changes in water resources that can be used to manage water supplies and water dependent activities. Within this context, this work focuses on streamflow and how it is impacted by irrigation water use.

Prediction of water availability and responses to stressors such as weather, climate and human impacts like pumping have often been simulated using computer models (Acharki et al., 2023; Cao et al., 2021; Lewis et al., 2023; Maurya et al., 2023; Mohseni et al., 2023; Zhang et al., 2023a, 2023b). However, model dynamics are always simplified often by omitted selected factors, which results in poorly represented systems. Interactions between groundwater and surface water tend to be problematic. They are represented in substantial detail in, for example, the Scott Valley Integrated Hydrologic Model in Northern California (Foglia et al., 2013; Tolley et al., 2019) and the Ohio River Basin Agricultural Policy Environmental Extender (APEX) model (Santhi et al., 2014). However, developing detailed watershed-scale, site-specific models takes significant time, effort, and resources, and as a result many existing models lack such capabilities and many areas lack any hydrologic model (Zipper et al., 2022a). In addition, areas may be impacted by upstream effects that are omitted from the local models, such as climate impacts on snowpack in what may be distant upland areas.

One alternative for settings without reliable site-specific models and/or when larger-scale effects are important is using large-scale (regional, continental, or global) streamflow models. For instance, the Soil and Water Assessment Tool (SWAT) (Malik et al., 2022; White et al., 2022),

* Corresponding author.
E-mail address: pbosomp@ku.edu (P. Bosompemaa).

https://doi.org/10.1016/j.envsoft.2024.106234
Received 20 September 2024; Accepted 25 September 2024
Available online 27 September 2024

Variable Infiltration Capacity (VIC) (Bao et al., 2022; Srivastava et al., 2020; Treesa, 2017), Hydrologic Engineering Center - Hydrologic Modeling System (HEC-HMS) (Chu and Steinman, 2009; Dotson, 2001), Water Balance Model (WBM) (Arnell, 1999; Bock et al., 2018; Lohmann et al., 2004), ParFlow-CONUS (Maxwell et al., 2015; O'Neill et al., 2021), WRF-Hydro (Salas et al., 2018), Precipitation-Runoff Modeling System (PRMS; Markstrom et al., 2015) and many others have been successfully used for large-scale water resource assessment and reservoir management worldwide. Some of these models, such as WRF-Hydro, are designed for short timeframes and are more suitable for problems such as weather and flood predictions. Others, such as PRMS, are designed for long timeframes and are suitable for simulations involving years and even decades or centuries. In addition, some, such as WRF-Hydro, tend to have very large computational demands that make model calibration and testing more difficult. Others, such as PRMS, have smaller computational demands that make model calibration and testing easier.

Some constructed large-scale models allow for abstracting submodels to generate a complete local-scale model for enhancement after calibration (Kannan et al., 2019; Pagliero et al., 2019; Vansteenkiste et al., 2014; Yalew et al., 2013) but the transfer of these calibrated models directly to uncalibrated watersheds may not account for unique characteristics present in the uncalibrated watershed. For large-scale models to be used for local management-relevant investigations, they need to be evaluated for the relevant setting because they are most often calibrated over a large-scale sample of gages, the distribution of which is often biased towards reference and/or perennial gages (Krabbenhoft et al., 2022), which makes them inherently less well-suited for any specific local application (Blanc et al., 2014). Of interest in this work is how bias might be identified and, if significant, mitigated to allow the insights available from such large-scale models to be used to evaluate the impacts of climate change at regional scales.

In this work we use an application of PRMS developed by the United States Geological Survey (USGS) and published as the National Hydrologic Model (NHM-PRMS; Regan et al., 2019). This model was designed to provide nationally consistent estimates of total water availability, changes in the timing and source of streamflow, and measures of the uncertainty of these estimates (Regan et al., 2019). NHM-PRMS is used to compute internally-consistent simulation results of the temporal and spatial distribution of water availability and storage across the contiguous United States (CONUS) and computes daily streamflows (Regan et al., 2019; additional NHM-PRMS details in Section 2.1 of this article). PRMS was developed primarily to simulate hydrological processes in small to medium-sized gaged watersheds and has performed well in past evaluations. A study conducted by Hodgkins et al. (2020) used Daymet climate forcing version 3, version 1.0 of the USGS National Hydrography Geospatial Fabric (NHGF), and PRMS 5.0.0 to compare trends in simulated and measured streamflow. They found that the NHM-PRMS has a relatively good fit to observed data across minimally-developed basins spanning the continental US (Hodgkins et al., 2020). However, NHM-PRMS may underestimate baseflow in some settings (Foks et al., 2019). This may be due to a simplified representation of groundwater, including omission of groundwater extraction (Towler et al., 2023). In addition to simplified groundwater, because NHM-PRMS is based on NHGF, the resulting coarse spatial resolution and specific configuration of the PRMS code in NHM may limit its applicability in site-specific studies. In general, the NHM-PRMS performs better in minimally disturbed basins than in those impacted by human activities such as irrigation (Towler et al., 2023), and therefore the ability of NHM-PRMS to provide management-relevant streamflow estimates in heavily-stressed environments remains unknown. The difficulties these results suggest for NHM-PRMS are similar to difficulties likely in other large-scale models, so the goals of this work have wide applicability.

Bias correction of streamflow models can be accomplished using a variety of methods. Data-based bias correction methods using process-based models has been the subject of various studies using climate data, precipitation records, land use information, and other historical datasets to implement diverse bias correction methods aimed at enhancing simulated streamflow outcomes [e.g. Chen et al., 2013 (downscaling precipitation for hydrologic impact studies over North America); Lin et al., 2023 (data driven models for streamflow prediction); Yang et al., 2020 (ensemble precipitation forecasts in the improvement of summer streamflow prediction skill)]. Another way of reducing residual errors and biases in hydrologic models is the use of machine learning approaches such as artificial neural networks (Hunt et al., 2022; Meng et al., 2016) and random forest (Abbasi et al., 2021; Shen et al., 2022). However these methods require more input data and time in the computational analysis (Al-Jarrah et al., 2015; Kotsiantis et al., 2006; Najafabadi et al., 2015). This work investigates methods with smaller data and computational demands.

In this work two bias correction methods are evaluated: flow duration curve (FDC) (following Farmer et al., 2018) and Auto-Regressive Integrated Moving Average (ARIMA, following Kim et al., 2021). The FDC method was used because it is less complex and research suggests it can accurately represent high and low flow events of a hydrologic system (Farmer et al., 2018; Hrachowitz et al., 2013; Kim et al., 2021; Vogel and Fennessey, 1995). ARIMA reputedly captures the temporal dependencies and patterns in streamflow data (Abudu et al., 2010; Kim et al., 2021; Nigam et al., 2009). ARIMA and FDC in this context appear to offer the advantage of bias-correction across a range of flow variability to accommodate anthropogenic effects arising from water use and management practices. Here we seek to test the utility of these methods in the context of using national hydrologic models to obtain more accurate regional-scale streamflow estimates.

In this study, we first compare streamflows simulated by NHM-PRMS with historical data to determine the degree to which it can provide useful management insights. The evaluation is pursued using data from a semi-arid agricultural area characterized by heavy groundwater pumpage, the type of area where large-scale models are likely to face difficulties. The area chosen is the Central Arkansas River Basin (CARB), which includes parts of the USA in Kansas, Colorado, New Mexico, Texas, and Oklahoma. This area is underlain by the High Plains aquifer, including the Ogallala aquifer. Results shown in this work reveal that the simulation provided useful predictions of streamflow in some settings, but there were widespread discrepancies between measured and simulated NHM-PRMS streamflow that would impair decision making. This provides an ideal test case study for testing the bias correction methods identified. Therefore, our aim with this study is not to improve the NHM-PRMS itself, but rather to develop and test postprocessing methods that could improve the utility of NHM-PRMS output. The CARB is part of the depleted Ogallala aquifer system, it is important to global food supplies, it has a long history of groundwater extraction and depletion, and there is rich historical data available in much of the region. Exploration of this system provides important insights about potential limitations in the NHM-PRMS and whether it, and other large-scale models, can be used with bias correction to develop predictions given future water demands, including the effects of climate change.

2. Methods

This section presents NHM-PRMS and the data and methods used to evaluate NHM-PRMS simulated values, describes the bias correction methods, and presents a framework for evaluating how the bias correction methods might be tested for their ability to improve future projections.

2.1. NHM-PRMS

PRMS is a daily timestep, deterministic model which simulates the response of watersheds to combinations of climate and land use to improve understanding of hydrologic processes on the basis of historical and projected climate data (Markstrom et al., 2015; Regan et al., 2018, 2019). PRMS simulates diverse physical processes including plant

canopy, snow accumulation and melt, soil, surface depression, surface runoff, groundwater storage, climate, and stream segments to characterize and derive parameters required in simulation algorithms, spatial discretization, and topological connectivity (Farmer et al., 2019; Markstrom et al., 2015; Regan et al., 2018; Towler et al., 2023).

This study uses results from NHM-PRMS, consisting of PRMS version 5.0.0, Daymet gridded data set version 2 (Thornton et al., 2014) and version 1.0 of the NHGF. The PRMS hydrographic data such as hydrologic response units (HRUs), stream segments, and watershed parameters are extracted from NHGF database (Viger and Bock, 2014). The NHGF framework for CONUS-scale modeling derived from the National Hydrography Dataset Plus, version 1 (NHDPlus) (Bondelid et al., 2010). The NHGF divides the land surface into hydrologic response units (HRUs), totaling 109,951 across the CONUS, each with a median size of 33.2 km^2. The flow from these HRUs is routed through a stream network comprising 56,549 segments, utilizing Muskingum routing (Viger and Bock, 2014).

The NHM-PRMS was calibrated using reference streamflow sites from the Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAGES-II; Falcone, 2011), a database that spans CONUS. The calibration involved using measured streamflow and daily hydrographs from pooled ordinary kriging (Farmer, 2016) across 1410 gaged watersheds throughout the CONUS (Farmer et al., 2019). These reference sites represent minimally disturbed watersheds, and the sites are evenly distributed across the CONUS (Falcone, 2011). The NHM-PRMS was selected for this study because it simulates streamflow influenced by large-scale dynamics, and provides outputs for streamflow and components of flow (surface runoff, interflow, and groundwater flow) for each stream segment (Regan et al., 2018, 2019). In addition, the NHM-PRMS can be coupled with climate models or downscaled climate projections to simulate future hydrological responses under various climate change scenarios (Risley, 2019). The model does not perform well in non-reference gaged areas impacted by human activities (Towler et al., 2023) hence bias-correction methods might enhance usefulness in disturbed watersheds such as the CARB.

2.1.1. Demonstration site

The CARB is the middle third of the Arkansas River Basin, and it stretches from the foothills of the Colorado Rocky Mountains to northwestern Oklahoma, covering approximately 414,000 km^2 (161,000 mi^2) (USBR, 2016). The CARB includes areas of Colorado (CO), Kansas (KS), New Mexico (NM), Texas (TX), and Oklahoma (OK) (Fig. 1).

The CARB faces many of the problems common to agricultural regions in the Western US, including diminishing groundwater supplies, stream drying, increasing water and soil salinity, and increasing nitrate concentrations (Butler et al., 2018, 2023; Whittemore et al., 2016; Zipper et al., 2021, 2022b). Water sources in the region include precipitation, surface water from the Arkansas River and its tributaries, groundwater from the High Plains/Ogallala aquifer and some local alluvial aquifers.

The climate of the study area generally has high temperatures and low precipitation which affects surface-water availability. Changes to both precipitation and temperature will be a key to determining climate-change impacts on agricultural and other water resource intensive activities. Feddema et al. (2008) projected that there will be relatively little change in KS total annual precipitation, however, potential evaporation is projected to increase in the summer due to higher temperatures. As such, it is projected that irrigation water demand will increase significantly by 2–8 inches depending on the location, crop type, and agricultural practices (Feddema et al., 2008). Given the inherent challenge of predicting both human behavior and climate patterns accurately, a significant increase in water deficits would necessitate a reevaluation of anthropogenic impacts.

None of the gages in the CARB were used in NHM-PRMS calibration

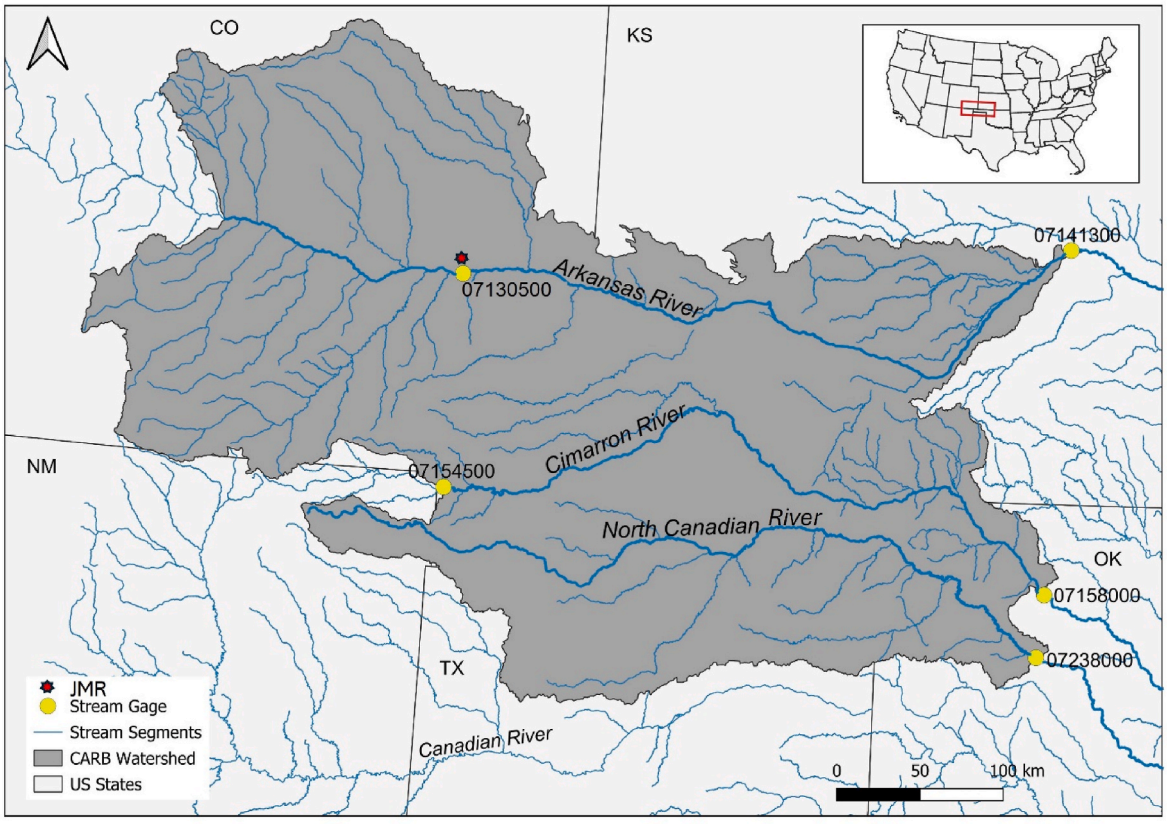


Fig. 1. The CARB region showing the Arkansas River, Cimarron River and the North Canadian River (dark blue), with selected stream gages for defining the study area and the John Martin Reservoir (JMR).

(Farmer et al., 2019). The NHM-PRMS does not simulate reservoir operations, surface or groundwater withdrawals, or stream releases (Towler et al., 2023). Therefore, the NHM-PRMS simulations for the CARB could approximate what streamflow would have been under reference conditions rather than current human-modified conditions, as the NHM is a natural flows model. In addition, the CARB region is heavily impacted by irrigation and most of the waters in the streams are allocated to meet crop demands. The John Martin Reservoir (JMR) located on the Arkansas River (Fig. 1) serves as a buffer against drier years (Bern et al., 2020) and has controlled outflow across the CO-KS border from year to year, though due to irrigation diversions the stream has been completely dry shortly downstream of the border for the past several decades (Zipper et al., 2022b). The NHM-PRMS does not account for any management of water resources or reservoir storage, but rather focuses on simulating natural flows in each stream segment and provides a crude estimate of stream segment storage based on Manning's equation, it is possible that the estimated storage used in the model is too high, which could result in biased timing of streamflow increases and decreases of the NHM-PRMS.

2.1.2. Data sources

NHM-PRMS version 5.0.0 daily streamflow results were downloaded from the ScienceBase (Hay, 2019) using R packages data.table (Dowle et al., 2019) from October 1, 1980 through December 31, 2016, approximately 36 years. Daily measured streamflow for stream gages was retrieved from the National Water Information System (NWIS) (USGS, 2022) using the dataRetrieval R package (DeCicco and Hirsch, 2021) for the same period. Stream gage data was filtered to determine gages in the CARB that had continuous daily streamflow data with no missing data. Out of the 58 gages found in the CARB, 27 had continuous data for the study period.

2.1.3. Model performance measures

NHM-PRMS simulated streamflow was compared to the USGS gage data to quantify how well the model represents the data. Fit metrics such as the Nash-Sutcliffe Efficiency (NSE), the coefficient of determination (R^2) and root mean square error (RMSE) were used to test the performance of the model as shown in Table 1, and were calculated using the "hydroGOF" (hydrological goodness-of-fit) package in R (Zambrano-Bigiarini, 2020).

2.2. Bias correction

We evaluated the performance of two bias correction methods, FDC and ARIMA, at all gages for their ability to reduce bias in the mean and variance of the simulated streamflow.

The FDC approach was used here to bias-correct the NHM-PRMS simulated time series of monthly streamflow. In this method we assumed that the probability distribution of the simulated streamflow should be shifted to approximately match the probability of the measured streamflow. This assumption was made to translate the model results that were a little too high to be low enough to reflect the measured and vice versa. Cumulative probability for the simulated and measured streamflow at a given time were generated from the "fdc" function (Vogel and Fennessey, 1994, 1995; Yilmaz et al., 2008) in R. The cumulative probability of the simulated streamflow was mapped to the cumulative probability of the measured streamflow for bias-correction (Farmer et al., 2018; Kim et al., 2021) and the matching streamflow was used as the bias-corrected simulated streamflow as shown in Fig. 2a and b.

Table 1
Metrics criteria for the model evaluation.

Perfect fit	Average fit	Worse fit
NSE = 1 (Schaefli and Gupta, 2007)	NSE = 0, simulated performs as good as the mean of the measured	NSE <0, the mean of the measured is a better predictor than the simulated
RMSE = 0 (Barnston, 1992)	The lower the RMSE value, the better the simulated fit.	Higher RMSE value means simulated does not fit measured perfectly
$R^2 = 1$ (Legates and McCabe Jr., 1999)	$R^2 < 1$	$R^2 \sim 0$

The stochastic approaches involved both nonseasonal models and seasonal models because the type of model that fits to a particular time series is problem dependent (Mishra and Desai, 2005) so there was a need to see how each gage would respond to the correction due to their difference in location and hydrological regimes. The nonseasonal model is Autoregressive Integrated Moving Average (ARIMA) and the seasonal model is Seasonal ARIMA (SARIMA), though since the SARIMA is a component of the ARIMA method, both are broadly within the category of ARIMA models. ARIMA and SARIMA are commonly applied in the hydrology field (Bazrafshan et al., 2015; Dwivedi et al., 2019; Kim et al., 2021; Mishra and Desai, 2005; Musarat et al., 2021; Phan and Nguyen, 2020; Valipour et al., 2013; Zhang et al., 2011b) to forecast reservoir inflows, water quality, drought, runoff, rainfall and other hydrological parameters. However, no study has used these stochastic approaches to bias-correct the NHM-PRMS.

Here, we used the ARIMA and SARIMA methods to bias-correct the simulated NHM-PRMS streamflow results, both separately from and in combination with, the FDC approach. The ARIMA method has the ability to forecast future residuals based on past observations and provides valuable information about the trends and the presence of autocorrelation in time series data whereas the SARIMA requires few model parameters for the time series, which exhibit non-stationarity both within and across the seasons (Mishra and Desai, 2005). These methods require a univariate time series that assumes stationarity which means the dataset should have a constant mean and variance (Dwivedi et al., 2019; Mishra and Desai, 2005; Musarat et al., 2021; Valipour et al., 2013). ARIMA is usually specified with three parameters p, d and q described as ARIMA (p, d, q) where p indicates the order of the autoregressive (AR) component, d indicates the amount of differencing (I), and q indicates the order of the moving average (MA) component (Mishra and Desai, 2005). The SARIMA is a component of the ARIMA method which accounts for temporal dependencies and seasonality patterns that occur in the data at regular intervals. Since streamflow in the CARB exhibits strong seasonal patterns due to climate impacts on the water resource, for example higher streamflow in spring when snow melts than in summer, the SARIMA method was considered to adjust the patterns and forecast residuals accordingly.

SARIMA models are described as ARIMA (p, d, q) $(P, D, Q)_s$, where P is the order of seasonal autoregression (AR_s), D is the number of seasonal differencing (I_s), Q is the order of seasonal moving average (MA_s) and s is the length of season. The mathematical equation for the ARIMA and SARIMA (Eqs. (1) and (4) respectively) are:

$$\phi(\beta)\ (1 - \beta y_t) = \varphi(\beta)\varepsilon_t \tag{1}$$

where φ (β) and $\varphi(\beta)$ are polynomials of order p and q respectively, y_t is the observed series, and ε_t is the random error term.

$$\phi(\beta) = (1 - \phi_1\beta - \phi_2\beta^2 - \ldots - \phi_p\beta^p) \tag{2}$$

$$\varphi(\beta) = (1 - \varphi_1\beta - \varphi_2\beta^2 - \ldots - \varphi_q\beta^q) \tag{3}$$

$$\Psi_P\beta^S\phi_p\beta\ (1 - \beta^S)^D\ (1 - \beta)^d\ y_t = \lambda_Q\beta^S\varphi_q\beta\varepsilon_t \tag{4}$$

where ϕ is the parameter of AR, φ is the parameter of MA model, Ψ is the parameter of AR_s model, and λ is the MA_s model.

The ARIMA and SARIMA were constructed from a time series of the residuals calculated from the observed and NHM-PRMS simulated streamflow. The ARIMA models were carried out in three steps (Fig. 2c): (1) Identification (Box and Jenkins, 1976), (2) Estimation (Bras and Rodríguez-Iturbe, 1993), and (3) Diagnostic Checking/Validation

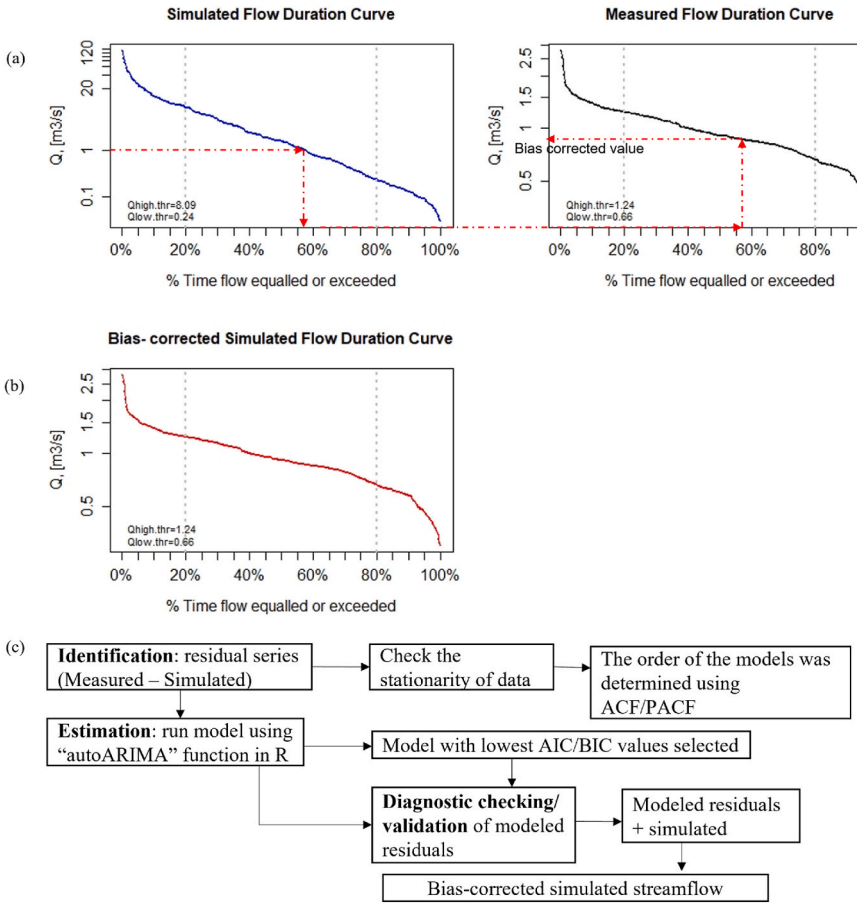


Fig. 2. Illustration of the bias correction methods. (a) The volume of NHM-PRMS simulated streamflow and probability at a point was mapped to the same probability on the measured streamflow and the corresponding streamflow value was used as the bias corrected FDC simulated streamflow at that point; (b) shows the bias corrected simulated FDC streamflow for all NHM-PRMS simulated streamflow values for gage 07156900 (Appendix, Table A.2); (c) a flowchart for the ARIMA and SARIMA residual-based bias-corrected model.

(Smith, 1985).

The identification step (Fig. 2c) involved data filtering, obtaining residual series by subtracting the simulated streamflow from that of the measured (Kim et al., 2021), and checking for stationarity of the dataset. If the residual series is not stationary, which we found in this work, the ARIMA method can be introduced to achieve stationarity and normality. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) approaches were used to check the stationarity of the dataset (Box and Jenkins, 1976). The temporal correlation structure of the transformed data was identified by examining its ACF and PACF functions (Mishra and Desai, 2005). The ACF is a useful statistical tool that measures if earlier values in the series have some relation to later values (Ömer Faruk, 2010). PACF is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all low order lags (Ömer Faruk, 2010). PACF is defined for positive lag only and their value also lies between −1 and +1 (Guha and Bandyopadhyay, 2016).

The estimation step (Fig. 2c) involved estimating the parameters of the ARIMA and SARIMA models for each gage using the "autoARIMA" function in R (Box et al., 2015; Box and Jenkins, 1976). For the diagnostic checking/validation step the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select the best model. Performance of these criteria was evaluated by (Foglia et al., 2013) and indicated that BIC tends to favor more model parameters. The AIC (Akaike, 1987) and BIC (Wit et al., 2012) are mathematically (Eqs. (5) and (6) respectively) represented as:

$$AIC = 2k - 2\ln(L) \quad (5)$$

$$BIC = k\ln(n) - 2\ln(L) \quad (6)$$

Where k is the number of parameters needed to captures the complexity of a model based on the criteria used, ln(L) is the log-likelihood of the model on the data which captures the goodness of fit and n is the number of data points (Akaike, 1987; Wit et al., 2012). The ARIMA model with the lowest AIC and BIC value was considered the best model and the autoregressive model of the residuals was generated. In this work AIC and BIC produced similar results. The "autoARIMA" function (Hyndman and Khandakar, 2008; Wang et al., 2006) automates the model processes and selects ARIMA or SARIMA as the best model for each of the gages. The seasonal and nonseasonal parameters for each gage were generated after running the model. The diagnostic checking/validation (Fig. 2c) step involved carrying out a residual check to see if the modeled residuals are normally distributed or not (Mishra and Desai, 2005). The bias-corrected flow was obtained by adding the modeled residuals from ARIMA to the simulated flow (Kim et al., 2021). The metrics of the bias-corrected flow versus the observed flow were obtained to see how the residual-based bias-correction methods performed. Forecasting (Peiris and Perera, 1988) was carried out on the modeled residuals for each gage to identify residual values that could be used to bias-correct the NHM-PRMS results in the absence of measured streamflow data, as needed to make future projections.

2.3. Applying the bias correction to future projections

The FDC and ARIMA methods rely on comparison between simulated and observed streamflow and therefore it is unknown how the methods would perform for future scenarios when there is no observational data available. To test the characteristics of projection, the bias-correction models were trained on a subset of the data (1980–2000) and the remainder of the data (2001–2016) was used to test model performance. The metrics for the trained and test methods were compared to see if the performance improved throughout the CARB. The method that gives better performance metrics would be the preferred bias correction method that could be applied for future climate scenarios to estimate future streamflow in the region.

3. Results and discussion

This section first presents streamflow measurements and output from the NHM-PRMS model. Then, bias corrected simulated values are compared with the measured streamflow to quantify error reduction. Finally, future bias corrections are calculated, method accuracy is evaluated, and recommendations for projection bias corrections are made.

3.1. Comparison between observed streamflow and NHM-PRMS output

NHM-PRMS simulated streamflows were compared to measurements for the 27 gages in the CARB with continuous data graphically and using the NSE. There are large differences between NHM-PRMS simulated and measured streamflow at most of the gage sites in the CARB (Figs. 3a and 4a). The mean and median NSE values for the monthly NHM-PRMS streamflow and measured streamflow across all the gages were −97.51 and −3.19 respectively, indicating poor performance (Table 1). Only five gages in the CARB had NSE >0 (Table A.2, Fig. 3a), and all of these were found in the westernmost portion of the domain closest to the headwaters in CO. The model predictions were likely best in western CO (Fig. 3a) because watersheds are smaller and groundwater withdrawal in that area is less (Dieter, 2018) than the other portions of the CARB. Thus, the area more closely matches the conditions of the reference gages used for NHM-PRMS calibration. This inference is supported by other studies that found that the NHM-PRMS did not represent groundwater withdrawals and stream releases (Towler et al., 2023), which suggests that areas that have less human activities would have better matches between model results and observed data. In addition, the NHM-PRMS model includes snowmelt parameters (Towler et al., 2023) and the model may have captured and predicted better patterns of snowmelt runoff from the headwaters. The NHM-PRMS performance gets worse moving downstream in the basin, as shown in Fig. 3a and Table A.2, likely because agricultural activities intensify downstream and possibly due to high storage estimates used in the model. Results from four stream gages A, B, C and D (Fig. 3a) were selected to demonstrate how the bias correction method performed at the upstream and downstream ends of the watershed. Gage A was selected because it is located at the inlet of the watershed and B, C and D are located at the outlet of different streams/rivers in the basin. The results for all gages can be found in Appendix A (Table A.2).

The performance of the model at gage locations below the JMR is poor compared to measured streamflow (Fig. 3a) as storage and releases from the reservoir are not simulated and they significantly impact downstream daily flows (Towler et al., 2023; Zipper et al., 2022b). Since the model did not accurately simulate streamflow in the CARB, further investigation was done to bias-correct the NHM-PRMS output.

3.2. Bias correction results

To better match the measured streamflow at long-term gages in the CARB, bias correction methods were applied to the model output at these locations. Two bias correction methods, stochastic (ARIMA and SARIMA) and FDC, were applied independently and together (FDC-

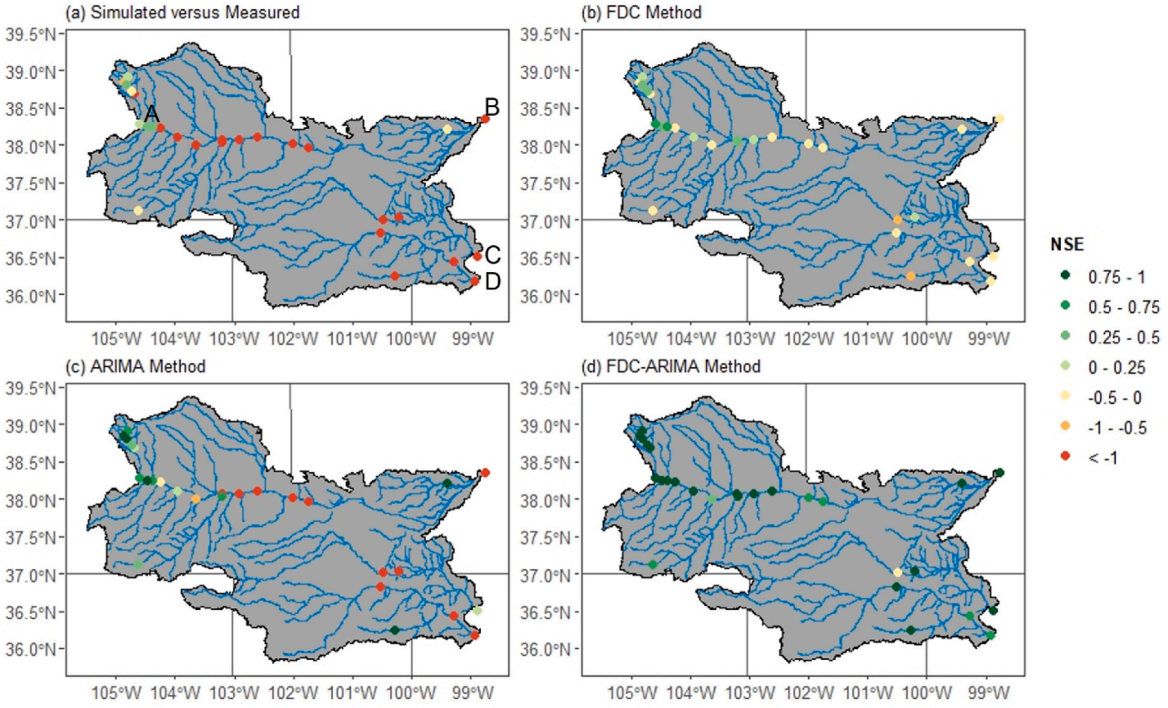


Fig. 3. Distribution map of performance metrics of gages in the CARB. (a) NHM-PRMS simulated streamflow versus USGS measured streamflow; (b) the FDC approach; (c) the ARIMA approach and (d) the FDC and ARIMA approach. Gage A represents 07108900 at St. Charles River at Vineland in CO, Gage B represents 07141300 at Arkansas River at Great Bend in KS, Gage C represents 07158000 at Cimarron River near Waynoka in OK and Gage D represents gage 07238000 at North Canadian River near Seiling in OK. Gray lines in the background show state boundaries (as in Fig. 1).

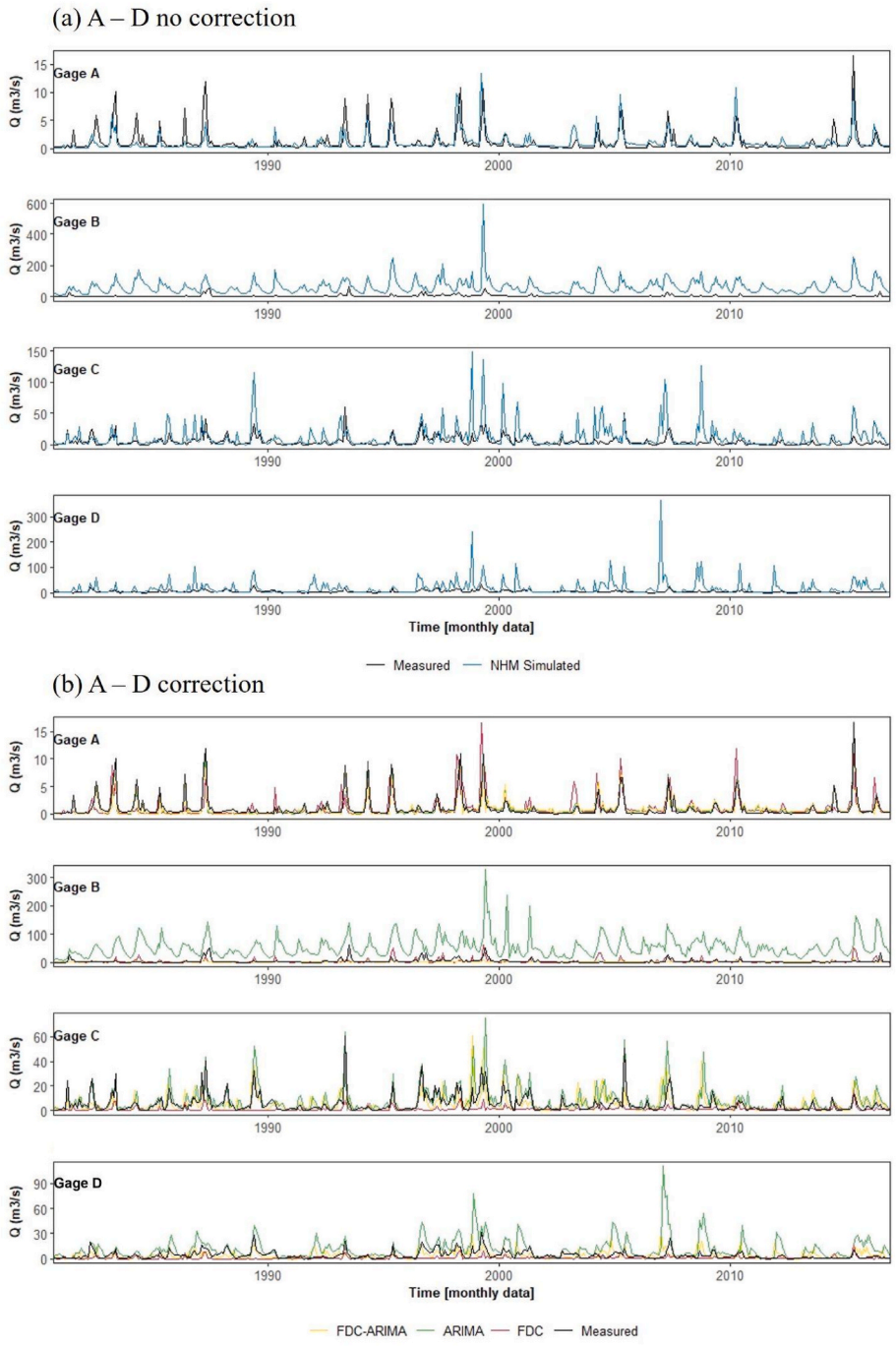


Fig. 4. (a) Time series plot of the NHM-PRMS simulated (blue) and the USGS measured (black) streamflow of selected gages in CARB; (b)Time series plot of measured (black) and bias-corrected (colors) streamflow at gaged sites. Gage A represents 07108900 at St. Charles River at Vineland in CO, gage B represents 07141300 at Arkansas River at Great Bend in KS, gage C represents 07158000 at Cimarron River near Waynoka in OK and gage D represents gage 07238000 at North Canadian River near Seiling in OK.

ARIMA) to try to best match observed data.

3.2.1. *ARIMA and SARIMA*

For the identification step of ARIMA and SARIMA, ACF and PACF were generated for residuals of all the gages in the CARB before and after the bias correction. To represent the ACF and PACF results for all the gages in this paper, gage A (Fig. 5) ACF and PACF results are shown. Before bias correction, the ACF and PACF graphs (Fig. 5) showed significant spikes at lag 1 and 12 with a few lags deviating from the confidence limits (CL) which indicates that the dataset was non-stationary and therefore a first order differencing was needed to make the data stationary. After bias correction, the ACF graph (Fig. 5) showed significant spikes at lag 1 and 12 and the PACF at lag 12, however most of the vertical lines were within the CL and the lags were near to zero, suggesting that the bias-correct model residuals are independently distributed and the model has accurately forecasted the time series (Musarat et al., 2021). The ACF and PACF graphs generated for all the gages had vertical lines within the CL after the bias-correction.

For the estimation step, the parameters for ARIMA and SARIMA, which included the p-values, standard errors and t-ratios, were estimated for all the gages (see Table A.3). The estimated parameters were statistically significant (p-values are <0.05) which reflects that the

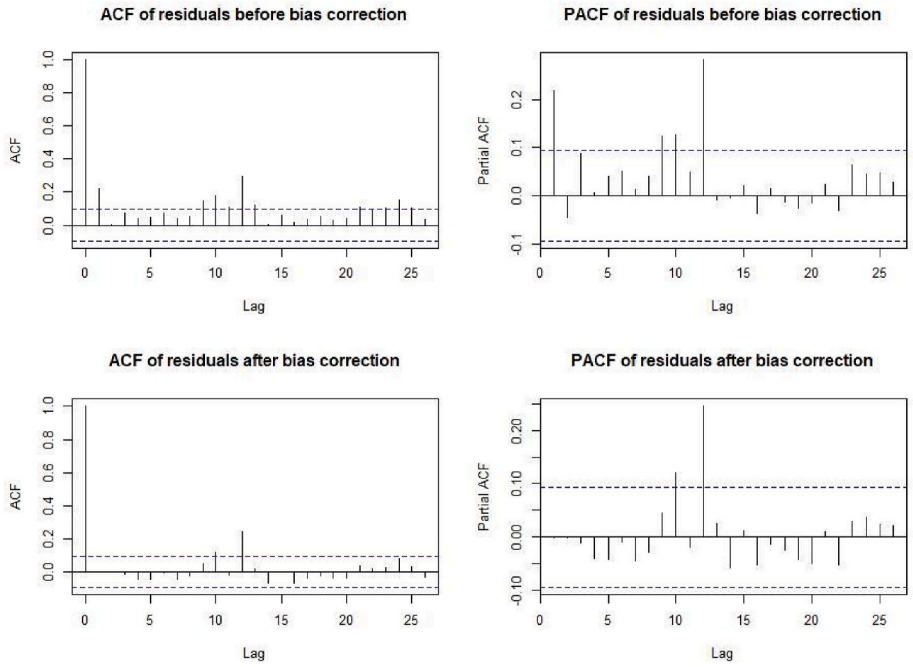


Fig. 5. ACF and PACF correlograms for gage A before (top graphs) and after (bottom graphs) bias-correction with 95% confidence limits (blue). The lag units are in months.

standard errors were generally small compared to the parameter values (Mishra and Desai, 2005). Thus, all parameters are statistically different than zero and are retained in the model. For the diagnostic step, the histogram generated for gages A, B, C, and D for the residuals (Fig. B.1) showed the modeled residuals are normally distributed and fits the dataset well.

The best model selected for gage A was ARIMA $(0, 1, 3)(0, 0, 1)_{12}$ (Table A.3) which means the ARIMA part of the model does not contain AR term ($p = 0$) but has I term ($d = 1$) which suggests that the original time series data was differentiated once for stationarity. The MA term ($q = 3$) suggests three lagged values of the forecasted errors were used to predict the residual values in the model. The SARIMA part does not contain AR_s or I_s terms ($P = 0$, $D = 0$), which suggests the seasonal dataset was stationary and did not require differentiation. The MA_s term ($Q = 1$) suggests one lagged value was used for forecasting the residuals for seasonal pattern recurring every 12 time units (months). The models selected for the other gages are shown in Table A.3. In addition, the estimated parameters from Eqs. (1) and (4) (Table A.3) with positive values suggest stronger positive correlation and the negative values suggest weaker negative correlation between consecutive values in the time series data. Overall, the estimated parameters for the ARIMA and SARIMA for all the gages were <1 (Table A.3) indicating a nonrandom mode model (stationarity) which suggests the modeled residuals for each gage fit the datasets.

Following ARIMA bias-correction, the mean and median NSE values for all the gages were −19.26 and 0.04 respectively. The NSE values for gage A improved from 0.42 to 0.92, gage B from −103.19 to −66.38, gage C from −4.99 to 0.09 and gage D from −46.07 to −7.19 as shown in Table A.2. The NHM-PRMS simulated streamflow improved slightly with a lot of negative NSE values and a few close to +1 (Table A.2) across the basin. In general, 14 gages in the region had NSE >0 (Table A.2) and the improvement was most frequently observed upstream in CO exemplified by Fig. 3c. There were slight improvements in the NHM-PRMS performance after the bias correction at all gages in the region as shown in Table A.2. However, the performance of the model after bias correction still degrades moving downstream (Fig. 3c). Fig. 4a and b shows that the bias-corrected simulated streamflow is still closer to the simulated streamflow than the measured. Based on Fig. 4a and b it could be observed that the ARIMA (seasonal and nonseasonal) approaches captured more high flows in the NHM-PRMS simulations and similar observation was seen across all the gages in the CARB. This suggests the ARIMA performed better in the high flows than in the low flows.

3.2.2. FDC

The mean and median NSE values for all the gages when bias-corrected using FDC were −0.04 and −0.09 respectively and the NSE values for gages A, B, C and D were 0.33, −0.37, −0.05 and −0.17 respectively (Table A.2). Fig. 3b shows that the bias corrected NHM-PRMS performed better across all the gages in the CARB, with 11 gages in the region having NSE >0 (Table A.2). The FDC may perform better than ARIMA because ARIMA assumes stationarity and linearity in the data for the modeled residuals, which may not always hold true for low flows. Since this study involves bias correcting real-world non-stationarity processes and anthropogenic effects, ARIMA is less effective for highly non-stationary hydrologic response to climate change and anthropogenic changes, such as changing agricultural practices (Milly et al., 2008).

Low flows are often climate controlled (Dallaire et al., 2021; Dierauer et al., 2018; Marx et al., 2018), thus during average and high temperatures precipitation decreases and evapotranspiration increases reducing the quantity of water in the streams. The time series plot (Fig. 6) shows that although the bias-corrected simulated streamflow exhibits a small degree of residual bias (Table A.2) it still matches the measured data better than the uncorrected NHM-PRMS results. This suggests that the FDC method reduced bias for low flows more effectively than for high flows.

3.2.3. FDC-ARIMA

The approach presented here involved combining the method that corrected for low flows and high flows to determine how they combine to bias-correct the simulated results. The bias-corrected simulated streamflow output from the FDC was used as the input simulated streamflow for the ARIMA methods. The FDC-ARIMA approach improved the performance of the NHM-PRMS model across all gages (Table A.2, Fig. 6) in the basin with a mean and median NSE value of

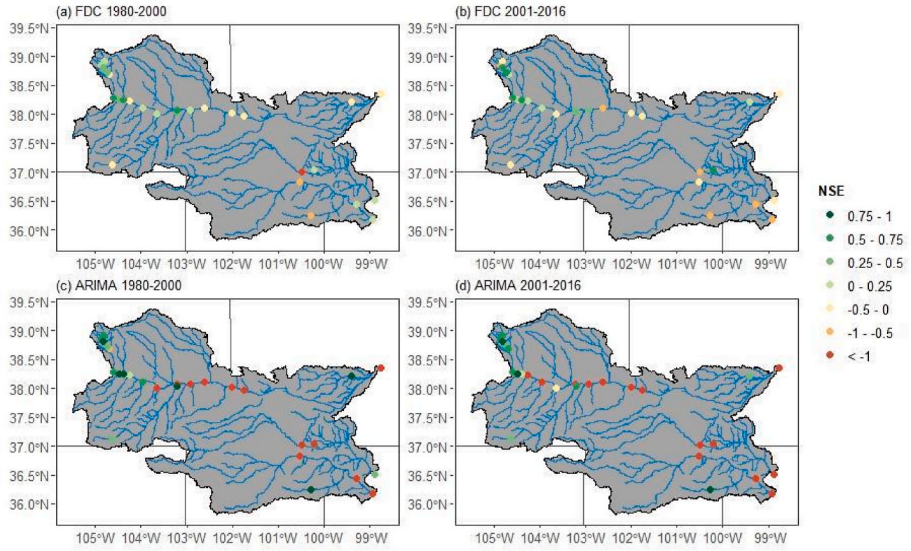


Fig. 6. Distribution map of FDC and ARIMA performance metrics of bias-corrected simulated streamflow versus USGS measured streamflow for gages in the CARB. (a) Training data for FDC (1980–2000); (b) Testing data for FDC (2001–2016); (c) Training data for ARIMA (1980–2000); (d) Testing data for ARIMA (2001–2016).

0.78 and 0.87 respectively. Gages A, B, C and D improved to an NSE value of 0.92, 0.80, 0.92 and 0.67 respectively (Table A.2). Figs. 3d and 6 show that the bias-corrected streamflow matches the measured, suggesting that the FDC-ARIMA method most effectively reduces bias in the NHM-PRMS in the CARB.

Comparing the three bias corrected approaches for the three outlet gages on the far eastern side of the CARB (gage B, C and D as shown in Fig. 3a), gage C below the Arkansas River responds to ARIMA bias correction differently than the other gages along the Arkansas River. The NSE for gage C improved from −4.99 to 0.09 and gage D from −46.07 to −7.19 as shown in Table A.2. In contrast, using the FDC method, the three outlets all behaved similarly with NSE between −0.5 and 0 (Fig. 3b). The improvement using the ARIMA method at the outlets may be due to drainage accumulation, as water from the Arkansas River and other tributaries in the CARB flows downstream, providing a potential source for streamflow even during drought. Applying the FDC-ARIMA method all the three outlets improved the NSE >0.5 (Fig. 3d) whereas for the ARIMA method only gage C improved (Fig. 3c).

3.3. *Lessons for future scenario simulations such as for climate-change evaluation*

The predictive ability of ARIMA and FDC were tested using a subset of data to train the model, and then that model was applied to the remaining data. Comparing the results to the observed testing data provides a way to evaluate how effective and robust the model is when bias-correcting for periods of time when no observed streamflow data are available.

The training data for ARIMA performed similarly as the testing (Fig. 6), where the model performed well upstream and worse moving downstream. From Table A.4 it was observed that the testing model performed worse than the training for some gages, indicating overfitting. This means the ARIMA methods at those sites learned to fit the training data too closely, including capturing the patterns leading to noise in the data, which caused the model to perform poorly on testing data in disturbed areas (Fig. 6c). For the forecasting analysis (Table A.4 and Fig. B.2), we observed that the forecasted residuals (Fig. B.2) start from the beginning of the time series (2017) and later flattened after a period of time indicating no trend in the data or weak signal. This suggests that forecasting of residuals for each gage is time bound and most of the gages in the region could be forecasted up to 3 years as shown in Table A.4. The forecasted residuals could be used to bias correct extreme high flows in simulated models to aid in predicting flood events accurately for more reliable early warnings and evacuation plans, effective reservoir management for flood mitigation and for proper water resource management decisions. These forecasted residuals could contribute to correcting biases in future streamflow simulations, although their applicability is limited to a three-year timeframe and therefore may be most useful for seasonal to annual streamflow forecasting. The ARIMA methods would not be useful for long-term water supply management, as it does not predict far enough out, or capture low flows well. Additionally, the ARIMA training data was parameterized based on static historical information and this could cause inaccuracies because the NHM parameterization data sets were calibrated at different spatial resolutions.

From Fig. 6 it can be observed that the training FDC method performed similarly as the testing FDC (Table A.4) and the fits were reasonable across the whole basin. Some streamflow peaks were not captured during the testing period (Fig. 7), consistent with our previous results (Section 3.2) which found that FDC corrects low flows better than high. Since the performance of some of the gages were not consistent throughout the testing period (Fig. 6b), the performance for projections further into the future may get worse at some sites, for example if a change within the watershed led to a shift in the flow distribution. This means that longer-term scenario simulation should be site specific and with appropriate context regarding limitations related to changes in management. Additionally, the FDC method for bias-correcting future scenarios assumes stationarity and homogeneity (Müller and Thompson, 2016) meaning the FDC approach may not accurately correct for bias if future climate scenarios include significant changes in precipitation patterns and when there are significant changes in water use, land use and water management.

Although FDCs constructed from historical observations at gaged sites may not represent current flow conditions well because flow regimes could be impacted by climate change and anthropogenic alterations of the catchments (Castellarin et al., 2013; Mu et al., 2007), the bias-corrected FDC approach could nevertheless be an important tool for reducing residuals in the NHM-PRMS model. This is particularly important for making future projections in areas that have been impacted by human activities because the NHM-PRMS does not account for processes like water withdrawals and reservoir operations, and so inherently is not representative of streamflow under these conditions (Towler et al., 2023).

For the purposes of predicting freshwater availability, the ability of

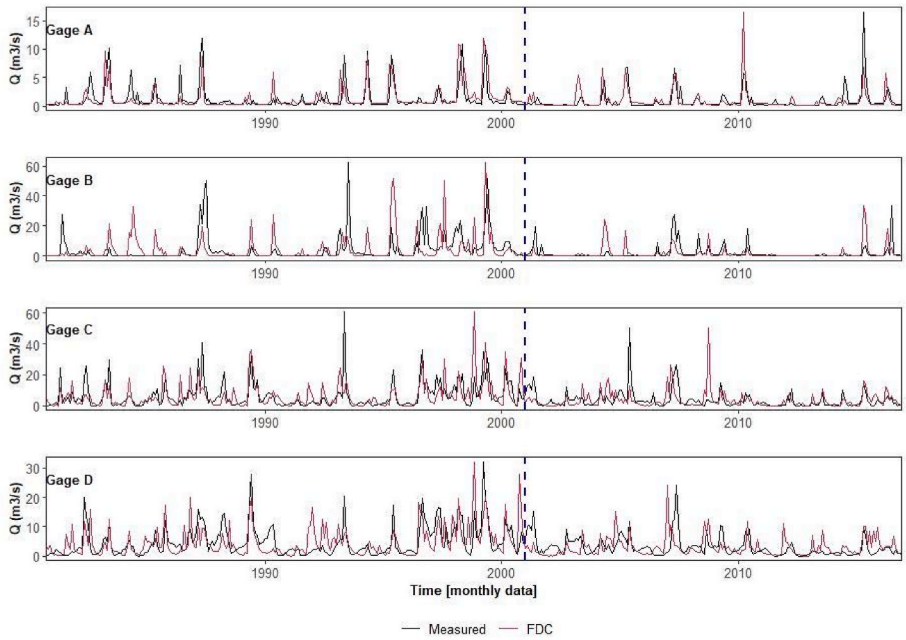


Fig. 7. Time series plot of FDC bias-corrected streamflow versus USGS measured streamflow at gaged sites with the blue line separating training (1980–2000) and testing results (2001–2016). Gage A represents 07108900 at St. Charles River at Vineland in CO, gage B represents 07141300 at Arkansas River at Great Bend in KS, Gage C represents 07158000 at Cimarron River near Waynoka in OK and Gage D represents gage 07238000 at North Canadian River near Seiling in OK.

the NHM-PRMS results to estimate low flows is limited but these flows are critical to understanding water availability during periods of drought or reduced precipitation (Van Loon et al., 2016). Furthermore, since the FDC approach presented here better captures low flows and the study is focused on making long term projections, the FDC would be a better approach for future scenarios compared to the ARIMA method for proper management and sustainability of streamflow.

A limitation of our work is that we only investigate bias-correction methods in settings with existing gaging data. While we do not demonstrate the application of bias correction methods in ungauged basins, other research works have utilized approaches such as spatial interpolation algorithms for FDC estimates (Farmer et al., 2019; Shu and Ouarda, 2012) to perform bias correction on national-scale streamflow models including ungauged settings. One limitation of bias correction methods lies in their assumption of a certain level of stationarity and linearity in hydrological processes. This assumption, coupled with evolving climate and land-use patterns, along with unreliable model performance (Vaittinada Ayar et al., 2021) could significantly restrict the effectiveness of these methods in predicting future conditions, particularly in heavily disturbed basins. In situations such as this, ensemble bias correction methodologies such as those proposed by Vaittinada Ayar et al. (2021) offer a viable solution to preserve internal variability, even amidst shifting climatic conditions.

4. Conclusions

In this work, bias correction was found to be necessary to make results from a national scale hydrologic model useful for estimating streamflow in a heavily stressed basin, as needed for future projection under climate change. Without bias correction, for a historic period (27 gages with continuous data between 1980 and 2016 from the Central Arkansas River basin (CARB)), simulated low flows were far greater than measured values. Indeed, in some areas zero streamflow was measured for long historic periods due to anthropogenic streamflow depletion, while simulated streamflow was consistently at non-zero values. Simulated high flows were less than measured values in the test case. Thus, the measured data had a wider range of values than the simulated values, including lower lows and higher highs. The CARB provided a useful test case due to its similarity to other highly stressed aquifers worldwide in that it is large (414,000 km^2) and has been heavily irrigated over past decades. Impacts on streams and groundwater systems in the region are well documented. The analysis was conducted using the USGS National Hydrologic Model (NHM-PRMS) because of its utility for multi-decadal simulations. The difficulties addressed in the CARB are expected to apply to other national scale models as well, because large-scale models tend to poorly represent selected aspects important to regional-scale resource management.

Results that closely matched streamflows for historical periods were obtained using a combination of the FDC and ARIMA bias correction methods. FDC (Flow Duration Curve) maps the cumulative probability of the NHM simulated streamflow onto the probability of the measured streamflow. FDC proved to be good at correcting low flows. ARIMA (Auto-Regressive Integrated Moving Average) involves subtracting the simulated from the measured streamflow and modeling the residual time series. ARIMA proved to be good at correcting high flows. A method of combining the two methods was devised that produced good results for both low and high flows during the historic period.

Tests conducted by withholding selected historical data suggests that ARIMA has poor long-term projection capabilities for the problem considered. Using FDC alone was found to be better for future simulations such as would be required for analysis of the effects of climate change. This means that the future projects are likely to have the strengths and deficiencies typical of FDC– low flows that are corrected well, and high flows that are more problematic.

Overall, this work identifies serious difficulties with using national-scale model results directly to project streamflows over climate-change scale time periods in locations with strong anthropogenic impacts on water resources, and that bias correction has some ability to improve results. This suggests ways to obtain national-scale simulated streamflows and provide results more useful for management and allocation of surface water and groundwater under climate change. Though the method development and tests in this work focused on agricultural systems, application to other water management settings is promising.

Software and data availability

Name of software: Streamflow-bias-correction-BP
Developers: Patience Bosompemaa, Andrea Brookfield, Sam Zipper, Mary C. Hill
Version: 1.0
Year first available: 2024
Program language: R
Cost: free
Availability https://cran.r-project.org/package=hydroTSMbosompee/Streamflow-Bias-Correction-BP: Codes and streamflow data (github.com)
Size of archive: 13.3 GB

CRediT authorship contribution statement

Patience Bosompemaa: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing, Investigation. **Andrea Brookfield:** Conceptualization, Data curation, Formal analysis, Resources, Software, Supervision, Validation, Writing – review & editing. **Sam Zipper:** Conceptualization, Data curation, Methodology, Software, Supervision, Validation, Visualization, Writing – review & editing. **Mary C. Hill:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Software, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known financial and personal relationships with other people or organizations that could inappropriately influence this work.

Data availability

Data are available on bosompee/Streamflow-Bias-Correction-BP: Codes and streamflow data (github.com).

Acknowledgements

This research was financed and supported by the National Science Foundation (NSF award number 1856084) and INFEWS/T2 FEWtures project: Innovation Analysis Framework for Resilient Futures, with Application to the Central Arkansas River Basin. We appreciate feedback from William Farmer and Jacob LaFontaine at the U.S. Geological Survey regarding the NHM.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2024.106234.

References

Abudu, S., Cui, C., King, J.P., Abudukadeer, K., 2010. Comparison of performance of statistical models in forecasting monthly streamflow of Kizil River, China. Water Sci. Eng. 3 (3), 269–281. https://doi.org/10.3882/j.issn.1674-2370.2010.03.003.

Acharki, S., Taia, S., Arjdal, Y., Hack, J., 2023. Hydrological modeling of spatial and temporal variations in streamflow due to multiple climate change scenarios in northwestern Morocco. Climate Services 30, 100388. https://doi.org/10.1016/j.cliser.2023.100388.

Akaike, H., 1987. Factor analysis and AIC. Psychometrika 52 (3), 317–332. https://doi.org/10.1007/BF02294359.

Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., Taha, K., 2015. Efficient machine learning for big data: a review. Big Data Research 2 (3), 87–93. https://doi.org/10.1016/j.bdr.2015.04.001.

Arnell, N.W., 1999. A simple water balance model for the simulation of streamflow over a large geographic domain. J. Hydrol. 217 (3), 314–335. https://doi.org/10.1016/S0022-1694(99)00023-2.

Ayers, J.R., Villarini, G., Schilling, K., Jones, C., Brookfield, A., Zipper, S.C., Farmer, W. H., 2022. The role of climate in monthly baseflow changes across the continental United States. J. Hydrol. Eng. 27 (5), 04022006. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002170.

Bao, Q., Ding, J., Han, L., 2022. Quantifying the effects of human activities and climate variability on runoff changes using variable infiltration capacity model. PLoS One 17 (9), e0272576. https://doi.org/10.1371/journal.pone.0272576.

Barnston, A.G., 1992. Correspondence among the correlation, RMSE, and heidke forecast verification measures; refinement of the heidke score. Weather Forecast. 7 (4), 699–709. https://doi.org/10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2.

Bazrafshan, O., Salajegheh, A., Bazrafshan, J., Mahdavi, M., Fatehi Maraj, A., 2015. Hydrological drought forecasting using ARIMA models (case study: karkheh basin). Ecopersia 3 (3), 1099–1117.

Bern, C.R., Holmberg, M.J., Kisfalusi, Z.D., 2020. Effects of John Martin Reservoir, Colorado on water quality and quantity: assessment by chemical, isotopic, and mass-balance methods. J. Hydrol. X 7, 100051. https://doi.org/10.1016/j.hydroa.2020.100051.

Blanc, E., Strzepek, K., Schlosser, A., Jacoby, H., Gueneau, A., Fant, C., Rausch, S., Reilly, J., 2014. Modeling U.S. water resources under climate change. Earth's Future 2 (4), 197–224. https://doi.org/10.1002/2013EF000214.

Bock, A.R., Farmer, W.H., Hay, L.E., 2018. Quantifying uncertainty in simulated streamflow and runoff from a continental-scale monthly water balance model. Adv. Water Resour. 122, 166–175. https://doi.org/10.1016/j.advwatres.2018.10.005.

Bondelid, T., C, J., C, M., R, M., A, R., 2010. NHD plus—NHDPlus version 1 (archive). https://nhdplus.com/NHDPlus/NHDPlusV1_home.php.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Bras, R.L., Rodríguez-Iturbe, I., 1993. Random functions and hydrology (Rev. ed., pp. 359-378). Courier Corporation. https://books.google.com/books?hl=en&lr=&id=AHPco48ZI1gC&oi=fnd&pg=PA1&dq=Bras+%26+Rodriguez-Iturbe,+1985+ARIMA&ots=ZpzpWtgNAQ&sig=7ovT5lW2dn9zGm_6rCJIcPXtbdk#v=onepage&q&f=false.

Butler, J.J., Whittemore, D.O., Wilson, B.B., Bohling, G.C., 2018. Sustainability of aquifers supporting irrigated agriculture: a case study of the High Plains aquifer in Kansas. Water Int. 43 (6), 815–828. https://doi.org/10.1080/02508060.2018.1515566.

Butler, Jr.J.j., Bohling, G.c., Perkins, S.p., Whittemore, D.o., Liu, G., Wilson, B.b., 2023. Net inflow: an important target on the path to aquifer sustainability. Groundwater 61 (1), 56–65. https://doi.org/10.1111/gwat.13233.

Cao, C., Yan, B., Guo, J., Jiang, H., Li, Z., Liu, Y., 2021. A framework for projecting future streamflow of the Yalong River basin to climate change. Stoch. Environ. Res. Risk Assess. 35 (8), 1549–1562. https://doi.org/10.1007/s00477-021-02009-w.

Castellarin, A., Botter, G., Hughes, D.A., Liu, S., Ouarda, T.B.M.J., Parajka, J., Post, D.A., Sivapalan, M., Spence, C., Viglione, A., Vogel, R.M., 2013. Prediction of flow duration curves in ungauged basins. In: Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H. (Eds.), Runoff Prediction in Ungauged Basins, first ed. Cambridge University Press, pp. 135–162. https://doi.org/10.1017/CBO9781139235761.010.

Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resour. Res. 49 (7), 4187–4205. https://doi.org/10.1002/wrcr.20331.

Chu, X., Steinman, A., 2009. Event and continuous hydrologic modeling with HEC-HMS. J. Irrigat. Drain. Eng. 135 (1), 119–124. https://doi.org/10.1061/(ASCE)0733-9437(2009)135:1(119.

Condon, L.E., Atchley, A.L., Maxwell, R.M., 2020. Evapotranspiration depletes groundwater under warming over the contiguous United States. Nat. Commun. 11 (1), 873. https://doi.org/10.1038/s41467-020-14688-0.

Dallaire, G., Poulin, A., Arsenault, R., Brissette, F., 2021. Uncertainty of potential evapotranspiration modelling in climate change impact studies on low flows in North America. Hydrol. Sci. J. 66 (4), 689–702. https://doi.org/10.1080/02626667.2021.1888955.

DeCicco, L., Hirsch, R., 2021. Introduction to the dataRetrieval package. https://github.com/DOI-USGS/dataRetrieval/blob/HEAD/vignettes/dataRetrieval.Rmd.

Dierauer, J.R., Whitfield, P.H., Allen, D.M., 2018. Climate controls on runoff and low flows in mountain catchments of western North America. Water Resour. Res. 54 (10), 7495–7510. https://doi.org/10.1029/2018WR023087.

Dieter, C.A., 2018. Water Availability and Use Science Program: Estimated Use of Water in the United States in 2015. Government Printing Office.

Dotson, H.W., 2001. Watershed modeling with HEC-HMS (hydrologic engineering centers-hydrologic modeling system) using spatially distributed rainfall. In: Gruntfest, E., Handmer, J. (Eds.), Coping with Flash Floods. Springer, Netherlands, pp. 219–230. https://doi.org/10.1007/978-94-010-0918-8_21.

Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., Ritchie, S., 2019. Package data.table; Extension of 'data.frame'. https://rdatatable.gitlab.io/data.table/.

Dwivedi, D.K., Kelaiya, J.H., Sharma, G.R., 2019. Forecasting monthly rainfall using autoregressive integrated moving average model (ARIMA) and artificial neural network (ANN) model: a case study of Junagadh, Gujarat, India. Journal of Applied and Natural Science 11 (1), 1. https://doi.org/10.31018/jans.v11i1.1951.

Falcone, J.A., 2011. GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow. U.S. Geological Survey. https://doi.org/10.3133/70046617.

Farmer, W.H., 2016. Ordinary kriging as a tool to estimate historical daily streamflow records. Hydrol. Earth Syst. Sci. 20 (7), 2721–2735. https://doi.org/10.5194/hess-20-2721-2016.

Farmer, W.H., LaFontaine, J.H., Hay, L.E., 2019. Calibration of the US geological Survey national hydrologic model in ungauged basins using statistical at-site streamflow simulations. J. Hydrol. Eng. 24 (11), 04019049. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001854.

Farmer, W.H., Over, T.M., Kiang, J.E., 2018. Bias correction of simulated historical daily streamflow at ungauged locations by using independently estimated flow duration curves. Hydrol. Earth Syst. Sci. 22 (11), 5741–5758. https://doi.org/10.5194/hess-22-5741-2018.

Feddema, J., Brunsell, N., Jackson, T., Jones, A., 2008. Climate Change in Kansas. Prepared for the Climate and Energy Project of the Land Institute. Department of Geography, University of Kansas, pp. 1–10.

Foglia, L., McNally, A., Harter, T., 2013. Coupling a spatiotemporally distributed soil water budget with stream-depletion functions to inform stakeholder-driven management of groundwater-dependent ecosystems. Water Resour. Res. 49 (11), 7292–7310. https://doi.org/10.1002/wrcr.20555.

Foks, S.S., Raffensperger, J.P., Penn, C.A., Driscoll, J.M., 2019. Estimation of base flow by optimal hydrograph separation for the conterminous United States and implications for national-extent hydrologic models. Water 11 (8), 8. https://doi.org/10.3390/w11081629.

Foster, L.M., Williams, K.H., Maxwell, R.M., 2020. Resolution matters when modeling climate change in headwaters of the Colorado River. Environ. Res. Lett. 15 (10), 104031. https://doi.org/10.1088/1748-9326/aba77f.

Gleeson, T., Wada, Y., Bierkens, M.F., Van Beek, L.P., 2012. Water balance of global aquifers revealed by groundwater footprint. Nature 488 (7410), 197–200.

Guha, B., Bandyopadhyay, G., 2016. Gold price forecasting using ARIMA model. Journal of Advance Management Journal. https://doi.org/10.12720/joams.4.2.117-121.

Hay, L.E., 2019. Application of the national hydrologic model infrastructure with the precipitation-runoff modeling system (NHM-PRMS), by HRU calibrated version. https://www.sciencebase.gov/catalog/item/5a4ea3bee4b0d05ee8c6647b.

Hodgkins, G.A., Dudley, R.W., Russell, A.M., LaFontaine, J.H., 2020. Comparing trends in modeled and observed streamflows at minimally altered basins in the United States. Water 12 (6), 6. https://doi.org/10.3390/w12061728.

Hrachowitz, M., Savenije, H.h. g., Blöschl, G., McDonnell, J.j., Sivapalan, M., Pomeroy, J. w., Arheimer, B., Blume, T., Clark, M.p., Ehret, U., Fenicia, F., Freer, J.e., Gelfan, A., Gupta, H.v., Hughes, D.a., Hut, R.w., Montanari, A., Pande, S., Tetzlaff, D., et al., 2013. A decade of predictions in ungauged basins (PUB)—a review. Hydrol. Sci. J. 58 (6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183.

Hunt, K.M.R., Matthews, G.R., Pappenberger, F., Prudhomme, C., 2022. Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. Hydrol. Earth Syst. Sci. 26 (21), 5449–5472. https://doi.org/10.5194/hess-26-5449-2022.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. J. Stat. Software 27, 1–22. https://doi.org/10.18637/jss.v027.i03.

IPCC, 2023. Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, first ed. Cambridge University Press. https://doi.org/10.1017/9781009325844.

Kannan, N., Santhi, C., White, M.J., Mehan, S., Arnold, J.G., Gassman, P.W., 2019. Some challenges in hydrologic model calibration for large-scale studies: a case study of SWAT model application to Mississippi-atchafalaya River Basin. Hydrology 6 (1), 1. https://doi.org/10.3390/hydrology6010017.

Kim, K.B., Kwon, H.-H., Han, D., 2021. Bias-correction schemes for calibrated flow in a conceptual hydrological model. Nord. Hydrol 52 (1), 196–211. https://doi.org/10.2166/nh.2021.043.

Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E., 2006. Machine learning: a review of classification and combining techniques. Artif. Intell. Rev. 26 (3), 159–190. https://doi.org/10.1007/s10462-007-9052-3.

Krabbenhoft, C.A., Allen, G.H., Lin, P., Godsey, S.E., Allen, D.C., Burrows, R.M., DelVecchia, A.G., Fritz, K.M., Shanafield, M., Burgin, A.J., Zimmer, M.A., Datry, T., Dodds, W.K., Jones, C.N., Mims, M.C., Franklin, C., Hammond, J.C., Zipper, S., Ward, A.S., et al., 2022. Assessing placement bias of the global river gauge network. Nat. Sustain. 5 (7), 586–592. https://doi.org/10.1038/s41893-022-00873-0.

Legates, D.R., McCabe, Jr.G.J., 1999. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241. https://doi.org/10.1029/1998WR900018.

Lewis, J.W., Lytle, S.E., Tavakoly, A.A., 2023. Climate change projections of continental-scale streamflow across the Mississippi River Basin. Theor. Appl. Climatol. 151 (3), 1013–1034. https://doi.org/10.1007/s00704-022-04243-w.

Lin, Y., Wang, D., Meng, Y., Sun, W., Qiu, J., Shangguan, W., Cai, J., Kim, Y., Dai, Y., 2023. Bias learning improves data driven models for streamflow prediction. J. Hydrol.: Reg. Stud. 50, 101557. https://doi.org/10.1016/j.ejrh.2023.101557.

Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., 2004. Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. J. Geophys. Res. Atmos. 109 (D7). https://doi.org/10.1029/2003JD003517.

Malik, M.A., Dar, A.Q., Jain, M.K., 2022. Modelling streamflow using the SWAT model and multi-site calibration utilizing SUFI-2 of SWAT-CUP model for high altitude catchments, NW Himalaya's. Modeling Earth Systems and Environment 8 (1), 1203–1213. https://doi.org/10.1007/s40808-021-01145-0.

Markstrom, S.L., Regan, R.S., Hay, L.E., Viger, R.J., Webb, R.M.T., Payn, R.A., LaFontaine, J.H., 2015. *PRMS-IV, the Precipitation-Runoff Modeling System, Version 4* (Techniques and Methods) [Techniques and Methods].

Marx, A., Kumar, R., Thober, S., Rakovec, O., Wanders, N., Zink, M., Wood, E.F., Pan, M., Sheffield, J., Samaniego, L., 2018. Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 °C. Hydrol. Earth Syst. Sci. 22 (2), 1017–1032. https://doi.org/10.5194/hess-22-1017-2018.

Maurya, S., Srivastava, P.K., Zhuo, L., Yaduvanshi, A., Mall, R.K., 2023. Future climate change impact on the streamflow of mahi River Basin under different general circulation model scenarios. Water Resour. Manag. 37 (6), 2675–2696. https://doi.org/10.1007/s11269-022-03372-1.

Maxwell, R.M., Condon, L.E., Kollet, S.J., 2015. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. Geosci. Model Dev. (GMD) 8 (3), 923–937. https://doi.org/10.5194/gmd-8-923-2015.

Meng, C., Zhou, J., Tayyab, M., Zhu, S., Zhang, H., 2016. Integrating artificial neural networks into the VIC model for rainfall-runoff modeling. Water 8 (9), 9. https://doi.org/10.3390/w8090407.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: whither water management? Science 319 (5863), 573–574. https://doi.org/10.1126/science.1151915.

Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. Stoch. Environ. Res. Risk Assess. 19 (5), 326–339. https://doi.org/10.1007/s00477-005-0238-4.

Mohseni, U., Agnihotri, P.G., Pande, C.B., Durin, B., 2023. Understanding the climate change and land use impact on streamflow in the present and future under CMIP6 climate scenarios for the parvara mula basin, India. Water 15 (9), 9. https://doi.org/10.3390/w15091753.

Mu, X., Zhang, L., McVicar, T.R., Chille, B., Gau, P., 2007. Analysis of the impact of conservation measures on stream flow regime in catchments of the Loess Plateau, China. Hydrol. Process. 21 (16), 2124–2134. https://doi.org/10.1002/hyp.6391.

Müller, M.F., Thompson, S.E., 2016. Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes. Hydrol. Earth Syst. Sci. 20 (2), 669–683. https://doi.org/10.5194/hess-20-669-2016.

Musarat, M.A., Alaloul, W.S., Rabbani, M.B.A., Ali, M., Altaf, M., Fediuk, R., Vatin, N., Klyuev, S., Bukhari, H., Sadiq, A., Rafiq, W., Farooq, W., 2021. Kabul River flow prediction using automated ARIMA forecasting: a machine learning approach. Sustainability 13 (19), 19. https://doi.org/10.3390/su131910720.

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. Journal of Big Data 2 (1), 1. https://doi.org/10.1186/s40537-014-0007-7.

Nigam, R., Bux, S., Nigam, S., Pardasani, K.R., Mittal, S.K., Haque, R., 2009. Time Series Modeling and Forecast of River Flow, vol. 4.

Ömer Faruk, D., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23 (4), 586–594. https://doi.org/10.1016/j.engappai.2009.09.015.

O'Neill, M.M.F., Tijerina, D.T., Condon, L.E., Maxwell, R.M., 2021. Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States. Geosci. Model Dev. (GMD) 14 (12), 7223–7254. https://doi.org/10.5194/gmd-14-7223-2021.

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., McIntyre, N., 2019. Investigating regionalization techniques for large-scale hydrological modelling. J. Hydrol. 570, 220–235. https://doi.org/10.1016/j.jhydrol.2018.12.071.

Patterson, N.K., Lane, B.A., Sandoval-Solis, S., Persad, G.G., Ortiz-Partida, J.P., 2022. Projected effects of temperature and precipitation variability change on streamflow patterns using a functional flows approach. Earth's Future 10 (7), e2021EF002631. https://doi.org/10.1029/2021EF002631.

Peiris, M.S., Perera, B.J.C., 1988. On prediction with fractionally differenced arima models. J. Time Anal. 9 (3), 215–220. https://doi.org/10.1111/j.1467-9892.1988.tb00465.x.

Phan, T.-T.-H., Nguyen, X.H., 2020. Combining statistical machine learning models with ARIMA for water level forecasting: the case of the Red river. Adv. Water Resour. 142, 103656. https://doi.org/10.1016/j.advwatres.2020.103656.

Regan, R.S., Juracek, K.E., Hay, L.E., Markstrom, S.L., Viger, R.J., Driscoll, J.M., LaFontaine, J.H., Norton, P.A., 2019. The U. S. Geological Survey National Hydrologic Model infrastructure: rationale, description, and application of a watershed-scale model for the conterminous United States. Environ. Model. Software 111, 192–203. https://doi.org/10.1016/j.envsoft.2018.09.023.

Regan, R.S., Markstrom, S.L., Hay, Lauren E., Viger, R.J., Norton, P.A., Driscoll, J.M., LaFontaine, J.H., 2018. Description of the National Hydrologic Model for Use with the Precipitation-Runoff Modeling System (PRMS) (Techniques and Methods) [Techniques and Methods].

Risley, J.C., 2019. Using the precipitation-runoff modeling system to predict seasonal water availability in the upper Klamath River basin, Oregon and California. In: Scientific Investigations Report. U.S. Geological Survey, pp. 2019–5044. https://doi.org/10.3133/sir20195044.

Salas, F.R., Somos-Valenzuela, M.A., Dugger, A., Maidment, D.R., Gochis, D.J., David, C. H., Yu, W., Ding, D., Clark, E.P., Noman, N., 2018. Towards real-time continental scale streamflow simulation in continuous and discrete space. JAWRA Journal of the American Water Resources Association 54 (1), 7–27. https://doi.org/10.1111/1752-1688.12586.

Santhi, C., Kannan, N., White, M., Di Luzio, M., Arnold, J.G., Wang, X., Williams, J.r., 2014. An integrated modeling approach for estimating the water quality benefits of conservation practices at the River Basin scale. J. Environ. Qual. 43 (1), 177–198. https://doi.org/10.2134/jeq2011.0460.

Schaefli, B., Gupta, H.V. (Eds.), 2007. Do Nash Values Have Value? *Hydrological Processes*. https://doi.org/10.1002/hyp.6825.

Shu, C., Ouarda, T.B.M.J., 2012. Improved methods for daily streamflow estimates at ungauged sites. Water Resour. Res. 48 (2). https://doi.org/10.1029/2011WR011501.

Smith, J.Q., 1985. Diagnostic checks of non-standard time series models. J. Forecast. 4 (3), 283–291. https://doi.org/10.1002/for.3980040305.

Srivastava, A., Kumari, N., Maza, M., 2020. Hydrological response to agricultural land use heterogeneity using variable infiltration capacity model. Water Resour. Manag. 34 (12), 3779–3794. https://doi.org/10.1007/s11269-020-02630-4.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., Cook, R.B., 2014. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States). https://www.osti.gov/biblio/1148868.

Tolley, D., Foglia, L., Harter, T., 2019. Sensitivity analysis and calibration of an integrated hydrologic model in an irrigated agricultural basin with a groundwater-dependent ecosystem. Water Resour. Res. 55 (9), 7876–7901. https://doi.org/10.1029/2018WR024209.

Towler, E., Foks, S.S., Dugger, A.L., Dickinson, J.E., Essaid, H.I., Gochis, D., Viger, R.J., Zhang, Y., 2023. Benchmarking high-resolution hydrologic model performance of long-term retrospective streamflow simulations in the contiguous United States. Hydrol. Earth Syst. Sci. 27 (9), 1809–1825. https://doi.org/10.5194/hess-27-1809-2023.

Treesa, A., 2017. Assessment of Impact of Climate Change on Streamflows Using VIC Model.

USBR, 2016. Reclamation—Managing Water in the Wet" SECURE Water Act Section 9503(c)-Reclamation Climate Change and Water, p. 24, 2016.

USGS, 2022. Water resources of the United States—national water information system (NWIS) mapper (usgs.gov). https://maps.waterdata.usgs.gov/mapper/index.html.

Vaittinada Ayar, P., Vrac, M., Mailhot, A., 2021. Ensemble bias correction of climate simulations: preserving internal variability. Sci. Rep. 11 (1), 3098. https://doi.org/10.1038/s41598-021-82715-1.

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J. Hydrol. 476, 433–441. https://doi.org/10.1016/j.jhydrol.2012.11.017.

Van Loon, A.F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., Van Dijk, A.I.J.M., Tallaksen, L.M., Hannaford, J., Uijlenhoet, R., Teuling, A.J., Hannah, D.M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., Van Lanen, H.A.J., 2016. Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches. Hydrol. Earth Syst. Sci. 20 (9), 3631–3650. https://doi.org/10.5194/hess-20-3631-2016.

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F., Willems, P., 2014. Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation. J. Hydrol. 511, 335–349. https://doi.org/10.1016/j.jhydrol.2014.01.050.

Viger, R., Bock, A., 2014. A Geospatial Fabric (GF) for National Hydrological Modeling, vol. 2014, pp. H43I–H1078.

Vogel, R.M., Fennessey, N.M., 1994. Flow-duration curves. I: New interpretation and confidence intervals. J. Water Resour. Plann. Manag. 120 (4), 485–504. https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(485.

Vogel, R.M., Fennessey, N.M., 1995. Flow duration curves ii: a review of applications in water resources Planning1. JAWRA Journal of the American Water Resources Association 31 (6), 1029–1039. https://doi.org/10.1111/j.1752-1688.1995.tb03419.x.

Wang, X., Smith, K., Hyndman, R., 2006. Characteristic-based clustering for time series data. Data Min. Knowl. Discov. 13 (3), 335–364. https://doi.org/10.1007/s10618-005-0039-x.

White, M.J., Arnold, J.G., Bieger, K., Allen, P.M., Gao, J., Čerkasova, N., Gambone, M., Park, S., Bosch, D.D., Yen, H., Osorio, J.M., 2022. Development of a field scale SWAT + modeling framework for the contiguous U.S. JAWRA Journal of the American Water Resources Association 58 (6), 1545–1560. https://doi.org/10.1111/1752-1688.13056.

Whittemore, D.O., Butler, J.J., Wilson, B.B., 2016. Assessing the major drivers of water-level declines: New insights into the future of heavily stressed aquifers. Hydrol. Sci. J. 61 (1), 134–145. https://doi.org/10.1080/02626667.2014.959958.

Wigley, T.M.L., Jones, P.D., 1985. Influences of precipitation changes and direct CO2 effects on streamflow. Nature 314 (6007), 149–152. https://doi.org/10.1038/314149a0.

Wit, E., Heuvel, E.V.D., Romeijn, J.W., 2012. 'All models are wrong…': an introduction to model uncertainty. Stat. Neerl. 66 (3), 217–236. https://doi.org/10.1111/j.1467-9574.2012.00530.x.

Yalew, S., van Griensven, A., Ray, N., Kokoszkiewicz, L., Betrie, G.D., 2013. Distributed computation of large scale SWAT models on the Grid. Environ. Model. Software 41, 223–230. https://doi.org/10.1016/j.envsoft.2012.08.002.

Yang, C., Yuan, H., Su, X., 2020. Bias correction of ensemble precipitation forecasts in the improvement of summer streamflow prediction skill. J. Hydrol. 588, 124955. https://doi.org/10.1016/j.jhydrol.2020.124955.

Yilmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model. Water Resour. Res. 44 (9). https://doi.org/10.1029/2007WR006716.

Zambrano-Bigiarini, M., 2020. Package 'hydroGOF'. Goodness-Of-Fit Functions for Comparison of Simulated and Observed.

Zhang, Q., Singh, V.P., Sun, P., Chen, X., Zhang, Z., Li, J., 2011a. Precipitation and streamflow changes in China: changing patterns, causes and implications. J. Hydrol. 410 (3), 204–216. https://doi.org/10.1016/j.jhydrol.2011.09.017.

Zhang, Q., Wang, B.-D., He, B., Peng, Y., Ren, M.-L., 2011b. Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. Water Resour. Manag. 25 (11), 2683–2703. https://doi.org/10.1007/s11269-011-9833-y.

Zhang, Y., Hou, J., Tong, Y., Gong, J., Zhou, N., Zhang, Z., Wang, T., Ju, Q., 2023a. Impact of climate change on streamflow in the middle–upper reaches of the weihe River Basin, China. J. Hydrol. Eng. 28 (6), 05023007. https://doi.org/10.1061/JHYEFF.HEENG-5825.

Zhang, Y., Zheng, H., Zhang, X., Leung, L.R., Liu, C., Zheng, C., Guo, Y., Chiew, F.H.S., Post, D., Kong, D., Beck, H.E., Li, C., Blöschl, G., 2023b. Future global streamflow declines are probably more severe than previously estimated. Nature Water 1 (3), 261–271. https://doi.org/10.1038/s44221-023-00030-7.

Zipper, S.C., Hammond, J.C., Shanafield, M., Zimmer, M., Datry, T., Jones, C.N., Kaiser, K.E., Godsey, S.E., Burrows, R.M., Blaszczak, J.R., Busch, M.H., Price, A.N., Boersma, K.S., Ward, A.S., Costigan, K., Allen, G.H., Krabbenhoft, C.A., Dodds, W.K., Mims, M.C., et al., 2021. Pervasive changes in stream intermittency across the United States. Environ. Res. Lett. 16 (8), 084033. https://doi.org/10.1088/1748-9326/ac14ec.

Zipper, S.C., Farmer, W.H., Brookfield, A., Ajami, H., Reeves, H.W., Wardropper, C., Hammond, J.C., Gleeson, T., Deines, J.M., 2022a. Quantifying streamflow depletion from groundwater pumping: a practical review of past and emerging approaches for water management. JAWRA Journal of the American Water Resources Association 58 (2), 289–312. https://doi.org/10.1111/1752-1688.12998.

Zipper, S., Popescu, I., Compare, K., Zhang, C., Seybold, E.C., 2022b. Alternative stable states and hydrological regime shifts in a large intermittent river. Environ. Res. Lett. 17 (7), 074005. https://doi.org/10.1088/1748-9326/ac7539.