# Evaluation of Nonparametric and Parametric Statistical Procedures for Modeling and Prediction of Cluster-Correlated Hydroclimatic Data

**Sarmistha Singh[a], Ash Abebe[b], and Puneet Srivastava[c]**

[a]Department of Civil Engineering, Indian Institute of Technology, Palakkad, India. **Address:** Ahalia Integrated Campus, Palakkad, Kerala, Pin: 678557, India

[b]Department of Mathematics & Statistics, Auburn University, AL, USA. **Address:** 221 Parker Hall, Auburn University, Auburn, AL 36849

[c]Biosystems Engineering, Auburn University, AL, USA. **Address:** 350 Mell Street, Tom E. Corley Building, Auburn University, Auburn, AL 36849

**Corresponding author:** Sarmistha Singh, Indian Institute of Technology, Palakkad, Ahalia Integrated Campus, Kerala, Pin: 678557, India; e-mail: sarmistha.research@gmail.com

## Abstract

Climate and hydrologic variables such as temperature, precipitation, streamflow and baseflow generally do not follow Gaussian distribution due to the presence of outliers and heavy tails. Therefore, they are usually analyzed using the nonparametric Wilcoxon rank-sum (WRS) test rather than parametric methods like classical t-tests and analysis of variance. Furthermore, in addition to having a non-Gaussian distribution, these data exhibit monthly/seasonal variability which leads to within month/season cluster-correlation. In this study, a nonparametric procedure, called Joint Rank Fit (JRFit), for analyzing cluster-correlated data was implemented and compared against traditional methods such as restricted maximum likelihood (REML), least absolute deviations (LAD), and Rank-Based Fit (RFit, a model-

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2017WR022254

based extension of WRS) for studying the coupled effect of the phases of El Niño Southern Oscillation (ENSO) and Atlantic Multidecadal Oscillation (AMO) on baseflow levels. The results from a large Monte Carlo simulation experiment showed that JRFit was more efficient than the other three methods for data with (i) high variability, (ii) outliers due to contamination, or (iii) strong monthly/seasonal correlation. The efficiency gain of JRFit was up to 50% compared to REML for heavy tailed and highly correlated data. Predictive performance evaluated using the mean absolute prediction error and mean prediction standard error from an out-of-sample cross-validation study showed JRFit to be optimal for providing predictions of baseflow on the basis of the phases of ENSO and AMO. Thus, it is recommended that JRFit be implemented in hydroclimatic studies to provide powerful inference when there is evidence of clustering in the data.

*Keywords*: Joint Rank Fit, Wilcoxon rank-sum, Restricted maximum likelihood, Baseflow, El Niño Southern Oscillation

## 1. Introduction

Oceanic-atmospheric phenomena such as El Niño Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multidecadal Oscillation (AMO), and North Atlantic Oscillation (NAO) are natural, cyclical (recurring at interannual, decadal, and multi-decadal scales) phenomena that are caused by fluctuations in sea surface temperature (SST) and sea level pressure (SLP) (Ropelewski and Halpert, 1986; Kiladis and Diaz, 1989; MacDonald and Case, 2005). These oscillations have strong effect on components of hydrologic cycle across the world (Kahya and Dracup, 1993; Regonda et al., 2005; Tootle et al., 2005; Lee and Julien, 2016; Schulte et al., 2017; Steirou et al., 2017). Therefore, studies of interannual, decadal, and multi-decadal climate variability phenomena and their interactions with hydrologic processes can provide useful information towards strategies for mitigating their adverse effects on water resources (Climate Research Committee and National Research Council, 1995).

ENSO, a major mode of climate variability affecting the global climate system (Diaz and Markgraf, 1992), is the fluctuation (occurring with a periodicity of two to seven years) in SST in the east-central equatorial Pacific Ocean. ENSO has three phases, namely Neutral, El Niño and La Niña (Philander, 1990). The terms "El Niño" and "La Niña" refer to respective warming and cooling of SST at Eastern Tropical Pacific. Similar to ENSO, AMO is caused by the fluctuations in ocean-atmospheric temperature. However, it occurs in the North Atlantic Ocean. High and low SST anomalies are characterized by warm/positive and cold/negative phases, respectively, of AMO cycles that oscillate with a periodicity of 60-80 years (Tootle et al., 2005; Johnson et al., 2013). The importance of understanding the teleconnections between natural climate and hydrologic variability has increased since its near-future predictability helps in planning and formulating water resources management (Zorn and Waylen, 1997; Cayan et al., 1999; Poveda et al., 2001; Schmidt et al., 2001; Räsänen and Kummu, 2012). These teleconnections (coupled/interaction studies) have been widely exploited in long lead-time forecasts of streamflow (e.g. Gutierrez and Dracup 2001; Chiew et al., 2003; Tootle et al., 2005).

To test and quantify the teleconnections of ocean atmospheric phenomena on hydroclimatic variables such as temperature, precipitation, streamflow, and groundwater, the conventional nonparametric Wilcoxon rank-sum (WRS) test has been widely applied (Diaz and Markgraf, 1992; Chiew et al., 1998; Tootle et al., 2005; Roy, 2006; Keener et al., 2010; Johnson et al., 2013; Mitra et al., 2014). Since hydroclimatic data sets are typically skewed (not normally distributed and contain outliers), nonparametric procedures provide a viable approach to minimize the influence of outliers and non-normality in testing and estimation (Helsel and Hirsch, 2002; Tootle et al., 2005; Johnson et al., 2013). The nonparametric WRS test makes no distributional assumptions on data and is resistant to the adverse effects of outliers (Bradley, 1968; Hogg et al., 2005). Therefore, WRS is more suitable than parametric methods such as classical t-tests and analysis of variance (ANOVA) for testing hypotheses when non-normality is evident (Crawford et al., 1983; Rousseeuw and Leroy, 1987; Hogg et al., 2005). Although the two sample WRS test is ideal for dealing with data sources that are non-normal, it does not readily extend to testing the interactive effects of multiple climatic oscillation variables. WRS requires splitting the data into different phases of climatic cycles in order to detect significant differences and quantify the comparative effects of multiple climatic oscillations on hydroclimatic variables (Tootle et al., 2005; Johnson et al., 2013; Mitra et al., 2014). Therefore, previous studies have not performed direct interaction (coupled) tests between two ocean atmospheric phenomena, but instead made inferences about interaction by splitting the data into different phases of climatic oscillations (Tootle et al., 2005; Roy, 2006; Keener et al., 2010; Johnson et al., 2013; Mitra et al., 2014).

Furthermore, as Galbraith et al. (2010) demonstrated, despite its robustness, the performance of WRS is suboptimal when data exhibit high monthly or seasonal clustering. Monthly or seasonal clustering refers to grouping of data points resulting from monthly (seasonal) variation of data points that is higher than the variation within months (seasons). In addition to outliers and heavy tails, meteorological and hydrological variables such as temperature, precipitation, streamflow, baseflow and groundwater display monthly or seasonal clustering in that their values tend to be similar on a monthly or seasonal basis, irrespective of year (Singh et al., 2015). Thus, WRS is typically performed on standardized monthly or seasonal anomalies that are obtained from subtracting long-term monthly or seasonal medians and dividing by standard deviation (Johnson et al., 2013; Mitra et al., 2014). In addition, Rosner et al. (2003) and Datta and Satten (2005) have proposed modified WRS tests for cluster-correlated data. While these tests are appropriate for clustered non-

Gaussian data, they still do not allow direct testing for coupled effects, measuring effect sizes, including other explanatory variables, or making predictions.

Robust approaches for modeling include the least absolute deviations (LAD) estimator (Koenker and Basset, 1978) and rank-based regression (RFit) estimator (Adichie, 1967; Jureckova, 1971; Jaeckel, 1972; Hettmansperger and McKean, 2011). RFit is a direct extension of WRS to a modeling framework. Neither LAD nor RFit account for correlated hydrological responses. The common modeling approach employed for studying phenomena that include cluster-correlated (monthly or seasonal) responses, is via the linear mixed effect (LME) model (Milliken and Johnson, 2004). Typically, the fitting of LMEs involves the use of the parametric restricted maximum likelihood (REML) method under the assumption that the responses are derived from the Gaussian distribution (Milliken and Johnson, 2004; Bates et al., 2015). REML is appealing since it allows one to not only estimate effect sizes and test their significance, but also estimate intraclass correlation coefficient that measures cluster effects (Milliken and Johnson, 2004). Since the assumption that the responses follow the Gaussian distribution may not be appropriate for hydroclimatic data, the use of Joint Rank Fit (JRFit) procedure (Kloke et al., 2009) is proposed in this study that gives a genuine nonparametric alternative to REML for fitting LME models. JRFit, like RFit, formulates the WRS method as a linear model, but additionally estimates the effect of cluster correlation in the model without requiring any assumption on the distribution of the data.

Therefore, the goal of this study was to propose a modeling framework to (1) provide a robust mechanism for testing of main and interaction effects of climate variables on hydrological variables; (2) account for cluster correlation in hydrological data; and (3) give accurate estimates and out-of-sample predictions of hydrological variables using climate phenomena. The interaction or the coupled effect of interannual and multi-decadal ocean atmospheric phenomena such as ENSO and AMO on baseflow levels were modeled, tested, and compared using the LAD, RFit, REML, and JRFit procedures. In this study, the authors aim to demonstrate the application and examine the efficiency of the JRFit procedure against other parametric and nonparametric procedures (RFit, LAD, and REML) in evaluating the influence of an interannual cycle (e.g. ENSO) and a multidecadal cycle (e.g. AMO) on baseflow levels as well as providing out-of-sample predictions. The paper is organized as follows. Section two presents baseflow data and methods and section three presents the results of simulation experiments as well as results of an analysis using baseflow data. Finally, section four provides conclusions and recommendations of the study.

## 2. Data and Methods

This study was performed in the Apalachicola-Chattahoochee-Flint (ACF) River Basin which is located in southeastern United States (Figure 1) (Mitra et al., 2014; Singh et al., 2015). It covers approximately 50,800 km$^2$ where much of the basin lies in Georgia, smaller areas of the basin are contained in southeastern Alabama and northwestern Florida (Figure 1). Soil of the ACF river basin consists of different land-resource areas where 97% of this basin is covered by Southern Piedmont, Georgia Sand Hills, Southern Coastal Plain, and Eastern Gulf Coast Flatwoods land-resource areas (Couch et al., 1996). The physiography of this river basin contains parts of the Blue Ridge, Piedmont, and Coastal Plain. There are six aquifers: the surficial aquifer system, the Floridan aquifer system, the Claiborne aquifer, the Clayton aquifer, the Providence aquifer, and the crystalline rock aquifer underlie this basin (Couch et al., 1996). The climate of the ACF basin is humid subtropical with mild winters and long summers. The average annual precipitation and temperature of this basin are about 127 cm and 17 °C, respectively (Mitra et al., 2014; Singh et al., 2015). The ACF River Basin is predominantly affected by ENSO-induced droughts, and studies have shown that other climate variability cycles (such as AMO) also have considerable influence in the region (Kiladis and Diaz, 1989; Hansen and Maul, 1991; Enfield et al., 2001; Schmidt et al., 2001;

Johnson et al., 2013; Singh et al., 2015). Moreover, low baseflow due to municipal, industrial, and agricultural water withdrawals is often a concern in the humid Southeast US. Specifically, in the study basin, low baseflows threaten protected endangered mussel species and diminish US Army Corps of Engineers' ability to meet minimum flows requirements in the Apalachicola River and Bay during droughts. Therefore, in this basin, the relationship between baseflow and the coupled effects of interannual (ENSO) and multidecadal (AMO) climatic phenomena was studied using several parametric and nonparametric procedures (i.e., LAD, REML, RFit, and JRFit). Furthermore, the efficiency of these procedures was examined based on their estimation and prediction errors. A detailed description of the models used for the estimation and prediction of baseflow is provided in Section 2.3. It is to be noted that baseflow data were used for the demonstration purpose only. Other non-normal, cluster-correlated hydrological data with heavy tails and outliers can also be used for this purpose.

**2.1 Data Sets**

**2.1.1 Baseflow Data**

In order to study the effect of climate variability on baseflow, it is important to obtain unregulated streamflow datasets (those that are not affected by reservoirs and dams). In this study, the streamflow gauging stations on the Flint River (Figure 1) were selected since Flint River is relatively unaffected by water control structures as compared to the other parts of the ACF basin that are highly regulated. For example, the Chattahoochee River has 5 federal dams and 6 private river dams, while the Flint River has only 2 small, run-of-the-river dams (Johnson et al., 2013). Daily streamflow data (in cubic feet per second, $ft^3/s$) for approximately 59 years were collected from six United States Geological Survey (USGS) gauging stations (Table 1). Baseflow were separated from daily streamflow using Web-based

Hydrograph Analysis Tool (WHAT) that has two digital filter methods for baseflow separation (Lim et al., 2005; Singh et al., 2015), namely BFLOW and Eckhardt. In this study, Eckhardt filter method with baseflow index 0.9, which is used for perennial rivers (Lim et al., 2005), was used for baseflow separation. The equation used for the Eckhardt filter method is given below.

$$b_t = \frac{(1 - BFI_{max})\alpha + b_{t-1} + (1 - \alpha)BFI_{max}Q_t}{1 - \alpha BFI_{max}} \tag{1}$$

where, $b_t$ is the filtered baseflow at time step t; $BFI_{max}$ is the maximum value of long term ratio of base flow to total streamflow; $\alpha$ is the filter parameter; $b_{t-1}$ is the filtered base flow at time step t-1; and $Q_t$ is the total streamflow at time step t. Finally, the daily values were changed into monthly cubic meters per second ($m^3/s$) for further analysis. The time series plots of monthly baseflow levels for each station are presented in Figure 2. The quantile-quantile plots (Figure 3) indicate that baseflow has a non-Gaussian distribution with heavy-tails and potential outliers.

**2.1.2 Oceanic-atmospheric Climate Variability Indices**

In this study, the Niño 3.4 SST index (ERSST.v3b) was used to define ENSO phases and durations (Trenberth, 1997; Trenberth and Stepaniak, 2001). The Niño 3.4 index is based on the SST anomalies in the Niño 3.4 region (5$^o$N–5$^o$S, 120$^o$–170$^o$W) (Trenberth, 1997). The monthly Niño 3.4 index values were obtained from the National Oceanic and Atmospheric Administration (NOAA), Climate Prediction Center, United States of America (USA) (http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php). When Niño 3.4 index value is between -0.5$^o$C and +0.5$^o$C, ENSO is considered to be in Neutral phase, and indices above +0.5 $^o$C or below -0.5$^o$C values indicate that ENSO is in El Niño or La Niña phase, respectively (Kiladis and Diaz, 1989; Ropelewski and Halpert, 1986).

The AMO index is identified as the coherent pattern of SST variability in the North Atlantic Ocean (0°–70°N) (Schlesinger and Ramankutty, 1994; Enfield et al., 2001; Tootle et al., 2005) and is defined by the warming and cooling pattern of SST. The warm/positive and cold/negative phases of AMO were defined based on the positive and negative numerical values from 121-month smoothed index values, and each phase lasts for about 20-40 years. AMO index values were obtained from the Physical Sciences Division of the Earth Systems Research Laboratory, NOAA, USA (ESRL, 2012; Johnson et al., 2013) (https://www.esrl.noaa.gov/psd/data/timeseries/AMO/). The positive phase of AMO considered in this study spanned from 1950 to 1963 and 1995 to 2008, and the negative phase spanned from 1964 to 1994.

## 2.2 Statistical Methods

### 2.2.1 Wilcoxon Rank-Sum Procedure

The problem of testing for significance of the effect $\Delta$ on hydrological responses to the change from one phase to another of a climate variable is often represented as a two-population statistical testing problem. Given hydrologic data $U_1, \dots, U_m$ and $V_1, \dots, V_n$ from two climate phases, where m and n are the respective sample sizes, with expected effect of phase change $\Delta$ $(= \mathrm{V} - \mathrm{U})$ only, interest lies in testing the null hypothesis $H_0\colon \Delta = 0$ versus the alternative $H_A\colon \Delta \neq 0$, $H_A\colon \Delta > 0$, or $H_A\colon \Delta < 0$. The WRS test proceeds by ranking all the data ($U_1, \dots, U_m$ and $V_1, \dots, V_n$) together from the smallest (rank 1) to the largest (rank $m + n$) and then summing the ranks of one of the samples, say $V$, to get the WRS statistic $W = R(V_1) + \cdots + R(V_n)$ (Lehmann, 1975). The standardized WRS statistic follows an asymptotic standard Gaussian distribution (Lehmann, 1975). The estimator of the treatment effect associated with the WRS is the median of all pairwise differences (Hodges and Lehmann, 1963)

$$\tilde{\Delta} = median(V_i - U_j), \qquad 1 \leq i \leq n;\ 1 \leq j \leq m$$

(2)

and $\tilde{\Delta}$ is approximately Gaussian with mean $\Delta$ and standard deviation $\tau\left({}^{1}/_{m} + {}^{1}/_{n}\right)^{1/2}$, where $\tau$ is a scale parameter that needs to be estimated from the data (Koul et al., 1987). Asymptotic relative efficiency (ARE) comparisons of WRS test and the classical *t* test indicate that the *t* test is only 4.5% more powerful than the WRS test when the data distribution is Gaussian; however, the WRS is 10% and 24% more powerful than the *t* test, for the heavier-tailed logistic and *t* distribution with 5 degrees of freedom, respectively (Lehmann, 1975).

Hydrological and climatic data have several interacting variables and it is often relevant to understand the interaction effects. Moreover, a modeling framework that allows for accurate estimation and prediction of hydrological phenomena from climate variables, in addition to testing of hypotheses, is of interest. For this reason, a generalization of the WRS to the linear model, Rank-Based Fit (RFit), first proposed by Jaeckel (1972) is considered in this study.

**2.2.2 General Linear Models: RFit and LAD**

Consider the general linear model that relates a set of *p* predictors (*X*) collected on *n* subjects to their response (*Y*) using the plane

$$Y = \alpha 1_n + X\beta + \varepsilon,$$

(4)

where $Y$ is an $n \times 1$ vector of responses, $X$ is an $n \times p$ matrix of predictors, $\varepsilon$ is an $n \times 1$ vector of random errors, and $1_n$ is an $n \times 1$ vector of ones (Seber and Lee, 2003). Estimation of the $p \times 1$ vector of slope parameters $\beta = (\beta_1, \ldots, \beta_p)^T$ as well as test for the significance of the components of $\beta$ are objectives of interest.

The Jaeckel (1972) rank-based estimator (RFit) of $\beta$, say $\tilde{\beta}$, minimizes the objective function

$$D(\beta) = \sum_{k=1}^{n} \varphi\left(\frac{R(e_k(\beta))}{n+1}\right) e_k(\beta)$$

(5)

where $e_k(\beta)$ is the $k$th entry of $Y - X\beta$, $R(e_k(\beta))$ is the rank of $e_k(\beta)$ among $e_1(\beta), \ldots, e_n(\beta)$, and $\varphi$ is a non-decreasing function defined on the interval $(0,1)$. Jaeckel (1972) established that $D(\beta)$ is a convex, continuous, and positive function of $\beta$. When $\varphi$ is odd about $1/2$, a natural estimator of the intercept is the median of the estimated residuals $e_1(\tilde{\beta}), \ldots, e_n(\tilde{\beta})$. Heiler and Willers (1988) have shown that the $\tilde{\beta}$ follows an asymptotic $p$ dimensional Gaussian distribution with mean $\beta$ and covariance matrix $\tau_\varphi^2 (X'X)^{-1}$, where $\tau_\varphi^2$ represents a scale parameter analogous to the error variance $\sigma^2$ in least squares estimation (Hettmansperger and McKean, 2011). A consistent estimator $\tilde{\tau}_\varphi^2$ of $\tau_\varphi^2$ is given in Koul et al. (1987). The estimator of $\tau_\varphi^2$ along with the asymptotic distribution can be used to construct test statistics for testing various types of hypotheses. Particularly, a Wald $t$ test for the significance of the $j$th individual slope, $1 \leq j \leq p$, uses the statistic

$$W_j = \frac{\tilde{\beta}_j}{\sqrt{\tilde{\tau}_\varphi^2 (X'X)_{jj}^{-1}}}$$

(7)

and the null hypothesis $H_0\colon \beta_j = 0$ is rejected in favor of $H_A\colon \beta_j \neq 0$ if $|W_j| > t_{n-p-1}(\gamma/2)$ where $t_{n-p-1}(\gamma/2)$ is the upper $\gamma/2$ percentile of the $t$ distribution with $n - p - 1$ degrees of freedom (Hogg et al., 2005). A Wald t-test uses a t-statistic formulated on the basis of the asymptotic Gaussian distribution of an estimator where a consistent estimator of the true variance is used in the calculation of the standard error of the estimator (Hogg et al., 2005). This RFit Wald test is equivalent to the WRS test in the case of a linear model with single binary (0 and 1) regressor indicating group membership (Hettmansperger and McKean, 2011). For example, if $Y$ is baseflow and $X$ is an indicator corresponding to the phases of ENSO, then the Wald test for significance of $\beta$ using the RFit estimator in the regression $Y = \alpha + \beta X + \varepsilon$ is identical to the WRS test comparing baseflow of two phases of ENSO.

A classical robust approach for estimating $\beta$ is the least absolute deviations (LAD) method (Koenker and Basset, 1978), where the 1-norm of the errors $\|\varepsilon\|_1 = \sum |\varepsilon_i|$ is minimized to obtain the estimator of $\beta$. If the errors $\varepsilon_1, \dots, \varepsilon_n$ are assumed independently drawn from a distribution that has probability density function $f$, then the LAD estimator of $\beta$ follows an approximate $p$-dimensional Gaussian distribution with mean $\beta$ and covariance matrix $\xi^2 (X^T X)^{-1}$, where $\xi = (2f(0))^{-1}$ (Hettmansperger and McKean, 2011).

**2.2.3 Linear Models with Cluster Correlation: JRFit and REML**

Assume that a total of $N = n_1 + \cdots + n_m$ observations in $m$ clusters are available, where cluster $k$ has $n_k$ observations. Within cluster $k$, let $Y_k$, $X_k$, and $\varepsilon_k$ denote the $n_k \times 1$ vector of responses, the $n_k \times p$ design matrix, and the $n_k \times 1$ vector of errors, respectively. Let $1_{n_k}$ denote a vector of $n_k$ ones. Then the linear model for $Y_k$, $k = 1, \dots, m$, is

$$Y_k = \alpha 1_{n_k} + X_k \beta + \varepsilon_k,$$

(8)

where $\alpha$ and $\beta$ represent the scalar intercept and the $p \times 1$ vector of slope parameters, respectively (Bates et al., 2015; Kloke et al., 2009). The errors in the same cluster are not assumed to be independent but errors in different clusters are assumed independent. The within cluster covariance matrix denoted by $Cov(\varepsilon_k) = \sigma^2\Omega_k$ is an $n_k \times n_k$ positive definite matrix. Model (8) reduces to the independent general linear model (4) if $\Omega_k = I_{n_k}$ for all $k$. In this study, $\Omega_k$ is assumed to be compound symmetric (Milliken and Johnson, 2004); that is, all the off-diagonal elements are equal, and all the diagonal elements are also equal. In this study the clusters are monthly or seasonal and there is no indication that the underlying correlations are different for different years. Monthly baseflow values fluctuate around the same value irrespective of year but the level tends to be different from month to month (Singh et al., 2015).

An extension of RFit to the clustered data case is given by Joint Ranking Fit (JRFit) estimation method that starts by stacking $Y_k$ into an $N \times 1$ response vector $Y$. The $N \times p$ predictor matrix $X$ is similarly defined by stacking $X_k$. The residuals for the stacked model are defined by the vector $e(\beta) = Y - X\beta$ with $i$th element $e_i(\beta)$, $i = 1, \ldots, N$. JRFit defines the dispersion function using $e(\beta)$ as $D_{JR}(\beta) = \sum_{i=1}^{N} \varphi\left(\frac{R(e_i(\beta))}{n+1}\right) e_i(\beta)$. Thus, JRFit is exactly the minimization of Jaeckel's dispersion for linear models with cluster-correlated errors with the resulting estimator denoted by $\tilde{\beta}_{JR}$. Kloke et al. (2009) showed that $\tilde{\beta}_{JR}$ follows an asymptotic Gaussian distribution with mean $\beta$ and covariance matrix given by $V_\varphi = \tau_\varphi^2 (X^T X)^{-1} \left(\sum_{k=1}^{m} X_k^T \Sigma_{\varphi,k} X_k\right) (X^T X)^{-1}$, where $\Sigma_{\varphi,k} = Cov(\varphi(F(\varepsilon_k)))$ is the $m \times m$ score intra-cluster covariance matrix. This asymptotic distribution is used to derive Wald tests of significance of the model parameters.

It is noted that both RFit and JRFit provide identical estimates of $\beta$ because they use the same formulation in their fixed-effects components. However, their random-effects components are different; thus, the standard errors of the estimators from RFit and JRFit are different (Kloke et al., 2009). JRFit calculates the within month variation in baseflow separately from the between month variation which is accounted for as random effects. For example, if we only want to study the effect of ENSO phases on baseflow, JRFit calculates baseflow differences between El Niño and La Niña phases within each month and the overall effect is compiled from the monthly effects. In RFit, since between month variations are not considered systematically as random effects, large month-to-month variations inflate the variance of the model random error. Thus, the standard errors of the JRFit estimators are generally smaller than those of RFit estimators and substantially smaller for data with high within month correlation.

The traditional approach of fitting model (8) involves using likelihood methods within the linear mixed effects model framework. Absent any distributional information on the population from which the data are drawn, the likelihood equation is constructed based on the assumption that $\varepsilon_k$ follow an $n_k$ dimensional Gaussian distribution with mean 0 and variance-covariance matrix $\Omega_k$. Estimation is performed using the REML method by first using regression to estimate the fixed effects residuals and using these residuals to estimate the variance components (Bates et al., 2015). The REML estimator of $\beta$, denoted by $\hat{\beta}_{REML}$, has an asymptotic Gaussian distribution with mean $\beta$ and variance-covariance matrix $V = \sigma^2\left(\sum_{k=1}^{m} X_k^T (I_{n_k} + X_k^T \Omega_k X_k)^{-1} X_k\right)^{-1}$.

For the linear score function, the ARE for comparing the JRFit estimator with the REML estimator is given by (Kloke et al., 2009)

$$ARE\left(\tilde{\beta}_{JR}, \hat{\beta}_{REML}\right) = \frac{1-\rho}{1-\rho_F} 12\sigma^2 \left(\int f^2(t)dt\right)^2 \tag{11}$$

where $\rho$ is the within cluster correlation and $\rho_F = Cor(F(\varepsilon_{11}), F(\varepsilon_{12}))$ is the within cluster rank correlation. The usual approach involves estimating the integral in equation (11) using kernel density estimates of the data distribution $f$ on the basis of a preliminary fit of the model (Koul et al., 1987). In this study, the effects of tail thickness, outliers, and correlation on the efficiency of JRFit (as well as LAD and RFit) versus REML were investigated using a Monte Carlo simulation experiment (Section 2.4) since analytical computation of AREs involving $\tilde{\beta}_{JR}$ is generally very complicated and often cannot be derived in closed form.

**2.3 Baseflow Models**

The effect of climate variability phenomena ENSO and AMO on baseflow ($BF$) were estimated using the linear model (Model 1)

$$BF = \beta_0 + \beta_1 ENSO + \beta_2 AMO + \varepsilon$$

(13)

where $\varepsilon$ represents random errors and $ENSO = 0$ and $ENSO = 1$ represent the La Niña and El Niño phases of ENSO and $AMO = 0$ and $AMO = 1$ represent the positive and negative phases of AMO, respectively. The baseline $BF$ value is $\beta_0$, which is the expected baseflow for the combination of La Niña and AMO positive phases. The value of $\beta_1$ measures the change in baseflow from baseline due to change from La Niña to El Niño for the same AMO phase, while $\beta_2$ measures the change in baseflow from baseline due to change from AMO positive to AMO negative for the same ENSO phase (Table 2). Since this is an additive model, if ENSO changes from La Niña to El Niño and AMO changes from positive to negative, the expected change in baseflow will be $\beta_1 + \beta_2$ (Table 2). This does not capture the modulation effects of the phases of one climate phenomenon by another one.

In this study, the coupled effect of climate variability phenomena on baseflow were studied using a linear model that allows us to test and estimate the interaction of the ENSO and AMO and their effect on baseflow. For that, we used the statistical model (Model 2)

$$BF = \beta_0 + \beta_1 ENSO + \beta_2 AMO + \beta_3 ENSO * AMO + \varepsilon$$

(14)

where, $\beta_3$ measures the interaction (coupled) effect of ENSO and AMO. The significance of the interaction effect $\beta_3$ indicates that the effect on baseflow of at least one of the phases of ENSO depends on the phases of AMO.

Under Model 1, the effect on expected baseflow of changing ENSO phase from La Niña to El Niño is $\beta_1$ regardless of whether AMO is in its positive phase or negative phase (Table 2). Under Model 2, however, the effect on expected baseflow of changing ENSO phase from La Niña to El Niño is $\beta_1$for the positive phase of AMO but $\beta_1$+ $\beta_3$ for the negative phase of AMO (Table 2). Hence $\beta_3$represents the effect on expected baseflow of the interaction of ENSO and AMO. Its significance indicates significant baseflow modulation of ENSO by the phases of AMO (Table 2). Model 2 is estimated as a linear mixed effect (LME) model where the errors $\boldsymbol{\varepsilon}$ are cluster correlated. In our study, JRFit, REML, LAD, and RFit were used to fit Model 2. The last two do not take cluster correlation into account. For the methods that account for cluster correlation, intraclass correlation coefficients were calculated as the proportion of total baseflow variance that is due to monthly variability (West et al., 2007). Moreover, the WRS test for clustered data (Rosner et al., 2003) was applied to baseflow data, where individual climate variability phases were compared separately since the method is not capable of including ENSO and AMO simultaneously as in Model 1 and Model 2.

Since linear models can be used for prediction, an out of sample cross validation was performed to evaluate the predictive performance of JRFit, REML, LAD, and RFit. The out of sample cross validation used a 10-fold cross-validation procedure where the data were randomly divided into 10 parts and 9 of the 10 parts were used as a training set while the remaining one part was used as a testing set. All methods were used to fit Model 2 using the training data, and resulting models were used for predicting baseflow values of the held-out sample. Prediction errors were computed by calculating the mean absolute prediction error (MAPE) between the predicted baseflow values and the true testing set baseflow values. Similarly, mean prediction standard errors (MPSE) were calculated from the testing set prediction variances. The MAPE and MPSE values of the different methods from the 10-fold cross-validation were compared using paired t-tests (Wong et al., 2014). These were corrected for multiple comparisons using the Bonferroni procedure (Bretz et al., 2011).

**2.4 Monte Carlo Evaluation of the Relative Efficiency of JRFit**

A Monte Carlo simulation was used to evaluate the relative efficiency of JRFit. In this simulation, 60 years hypothetical climate data, with 30 years assumed to be under climate phase $A$ and the remaining 30 assumed to be under climate phase $B$, were generated. The responses (baseflow values) were assumed to be measured seasonally; that is, there were four measured responses per year corresponding to each season. A compound-symmetric seasonal clustering structure was imposed where within cluster observations were correlated with correlation $\rho$. To simulate the heavy-tailed nature of climate data, the responses were generated using the $t$ distribution with various degrees of freedom. The tails of the $t$ distribution are heavy for small degrees of freedom and they approach the tails of the Gaussian distribution for degrees of freedom approaching infinity. Thus, a single set of responses under the two climate phases $A$ and $B$ was generated as

$$Y_A = T_{30}(d, \rho) + S \tag{15}$$

$$Y_B = T_{30}(d, \rho) + \Delta + S \tag{16}$$

where $T_{30}(d,\rho)$ is a random variate from the 30-dimensional $t$ distribution with degrees of freedom $d$ and correlation $\rho$, $\Delta$ is the effect of climate phase change, $S$ is the season effect. Therefore, the difference in expected responses of two different climate phases is $E(Y_B) - E(Y_A) = \Delta$. Different techniques were judged by how precisely and accurately they were able to recover the true value of $\Delta$ in this noisy setting. In our simulation, we considered various combinations of $(d,\rho,\Delta)$. The values of $d$ and $\rho$ considered were $df = (3, 4, 5, 10, 15, 20, 30, 60, 100)$ representing decreasing tail-thickness (decreasing variability in baseflow) and $\rho = (0, .05, .1, .2, .3, .4, .5)$ representing increasing degrees of seasonal correlation. Several values of $\Delta = (0, 1, 2, 3)$ were used to validate and check the sensitivity of estimation procedures.

The performance of JRFit versus traditional approaches was evaluated by studying the errors in the estimation of $\Delta$ using JRFit, REML, LAD, and RFit methods. The mean squared errors (MSE) of the estimates of $\Delta$ were calculated based on $M$ =10,000 iterations as in Kloke et al. (2009) where fresh data were generated in every iteration. For example, for JRFit we have $M$ estimates $\tilde{\Delta}_{JR,1}(df,\rho), \dots, \tilde{\Delta}_{JR,M}(df,\rho)$ corresponding to various degrees of freedom of the t distribution and various degrees of within season correlation. The Monte Carlo estimate of the MSE was calculated as

$$MSE_{JR}(df,\rho) = \frac{1}{M}\sum_{i=1}^{M}\left(\tilde{\Delta}_{JR,i}(df,\rho) - \Delta\right)^2. \quad (17)$$

$MSE_{REML}(df,\rho)$, $MSE_{LAD}(df,\rho)$, and $MSE_{WRS}(df,\rho)$ were calculated analogously. The finite sample relative efficiencies (RE) of JRFit, LAD, and RFit versus REML are the ratios $\frac{MSE_{REML}(df,\rho)}{MSE_{JR}(df,\rho)}$, $\frac{MSE_{REML}(df,\rho)}{MSE_{LAD}(df,\rho)}$, and $\frac{MSE_{REML}(df,\rho)}{MSE_{R}(df,\rho)}$, respectively. If RE = 1, then the two methods were judged to be equally efficient, whereas, RE > 1 indicated that the competitor was more efficient than REML in estimating phase effect.

The effect of outliers was evaluated by replicating the above simulation procedure using data from the contaminated normal distribution instead of the *t* distribution. Huber contamination (Huber, 1964; El-Shaarawi, 1989) of a base standard Gaussian with a Gaussian contaminant that has a higher variance was employed. This distribution is given as $(1-\delta)N(0,1) + \delta N(0,\sigma^2)$, where $\delta$ and $\sigma^2$ represent the proportion and the variance of contamination, respectively. In our simulation study, 0% to 35% contamination and contamination variance of $\sigma^2 = 9$ were used, where 0% represents no contamination and 35% represents heavy contamination (Abebe and Bindele, 2016; Bindele and Abebe, 2015).

## 3. Results and Discussion

### 3.1 ENSO and AMO

The estimates and standard errors for Model 1 (additive effect of ENSO and AMO on baseflow) as well as Model 2 (including the coupled effect of ENSO and AMO on baseflow) are provided in Table 3. The coefficients represent changes in baseflow ($m^3/s$) from the reference group baseflows where the reference group is taken to be the set of baseflow values of La Niña - AMO positive phase years (Table 2). ENSO coefficients indicate the change in baseflow as the phase changes from La Niña to El Niño. AMO coefficients indicate the change in baseflow as the phase changes from AMO Positive to AMO Negative. Considering the additive Model 1, ENSO and AMO coefficients were found to be positive and significant (at 5% level of significance) by REML and JRFit methods for all stations. However, LAD failed to find ENSO effects to be significant for all the stations. It also failed to find AMO effects to be significant for Stations C, D, and E. RFit failed to find ENSO to be significant for Stations A and F. Considering the interactive Model 2, the interaction term was found to be negative and significant by all the methods and for all stations except LAD and RFit that failed to find the interaction to be significant for Station F. Negative and significant interaction terms indicate that baseflow decreased overall when AMO changed phase from negative to positive primarily associated with drops in baseflow during La Niña phases while baseflow remained largely unchanged during El Niño phases. Comparing Model 1 and Model 2, it was found that removing the interaction term decreased the individual effects (ratio of coefficient estimates to standard errors) of ENSO and AMO on baseflow indicating the increased power of the interactive Model 2 in comparison to the additive Model 1. The Rosner et al. (2003) WRS test for clustered data using individual climate variability phases separately gave similar results to JRFit used on individual climate variability phases.

The results from the out of sample cross-validation study that calculated MAPE and MPSE values for the four different procedures (REML, LAD, RFit and JRFit) are presented in Table 4 and Figure 4. The within-month baseflow clustering effect found by calculating the intraclass correlations is also reported in Table 4. MAPE and MPSE values of different methods were compared using paired t-tests. Methods that differed significantly following a Bonferroni correction (Bretz et al., 2011) are indicated by different letter superscripts (a, b, c and d) in Table 4. The recommended optimal procedure is given in the last column of Table 4. The MAPE values for REML were found to be larger than those for LAD, RFit, and JRFit (Table 4 and Figure 4). MAPE values were similar for JRFit, LAD, and RFit where, as expected, JRFit and RFit gave equal MAPE values (Table 4 and Figure 4). The MPSE values for JRFit were found to be significantly smaller than all the other methods for Stations C, D, and E (Table 4 and Figure 4). The MPSE values for JRFit were not found to be significantly different from REML for Stations A and B, and from LAD and RFit for Station F. Table 4 shows that JRFit had either the lowest MAPE and/or the lowest MPSE in comparison to all the methods, except for Station F where it is tied with LAD and RFit (Figure 4). Thus, JRFit was found to be an optimal procedure for providing out-of-sample predictions of baseflow responses using climate variables. The MAPE and MPSE values (Figure 2 and Table 4) are increasing from upstream to downstream (from station A to F). This might be due to the large variation in baseflow levels at the downstream stations (Figure 2) where the range of the baseflow values are high for the stations E and F as compared to the upstream stations. It is possible the variation could be due to the size of the drainage basin (Table 1) and/or the existence of a dam upstream of station E thus affecting the free flow of water (Figure 2).

### 3.2 Evaluation of the Relative Efficiency of JRFit

The relative efficiency (RE) values reported for various combinations of $(df, \rho)$ are given in Figure 5 for the *t* distribution and in Figure 6 for the contaminated normal distribution. The value of $\Delta$ did not have much effect on measured relative efficiencies. So, only the results for $\Delta = 3$ are reported. Considering the effect of tail thickness and clustering strength on the efficiency of the various methods (Figure 5), it is noted that the methods that did not take clustering into account (i.e., RFit and LAD) had relative efficiency curves always below that of JRFit. Hence, RFit and LAD were inefficient compared to JRFit in all the cases evaluated; in some cases losing over 100% in efficiency. RFit and LAD were also inefficient compared to REML, especially as the tails of the distribution approach the tails of the Gaussian distribution (increasing $df$). However, they tended to perform better than REML for distributions that have tails substantially thicker than Gaussian tails, especially when the correlation is high (large $\rho$). For instance, for $\rho = 0.4$, RFit and LAD were 18-19% more efficient than REML for 3 degrees of freedom *t* distribution but this efficiency quickly dropped to a loss of 6-7% efficiency for 4 degrees of freedom. For clustered data, the competition is between REML and JRFit. For heavy tailed data ($df = 3$), JRFit was 24-51% more efficient than REML for the entire range of correlation scenarios. The efficiency of

JRFit versus REML also increased consistently as the clustering in data became stronger (increasing $\rho$). JRFit was found to be less efficient than REML for lighter tails (increasing $df$) and weak cluster correlation (decreasing $\rho$) with relative efficiency approaching the theoretical 95.5% value, which is the case for independent data, as $\rho$ approached 0. Generally, in all the scenarios evaluated, JRFit's efficiency loss relative to REML was never more than 8%, but efficiency gain was up to 50% for heavy tailed and highly correlated data.

Figure 6 contains the relative efficiency values with respect to changing levels of data contamination and clustering strength. It was found that JRFit is more efficient than RFit and LAD as its relative efficiency curve lies above those for RFit and LAD. REML was also more efficient than RFit and LAD as the RFit and LAD relative efficiency curves were below 1. However, it was observed that the efficiency of REML deteriorated for data with strong clustering and with increasing percentage of contamination. For $\rho = 0.5$ and 35% contamination, RFit and LAD were 20% more efficient than REML. When there was no contamination in the data, JRFit's relative efficiency versus REML increased from 92% for $\rho = 0$ to 111% for $\rho = 0.5$. One of the most interesting results was that, for this finite sample analysis, JRFit outperformed REML for highly correlated Gaussian data. Moreover, JRFit's relative efficiency increased steadily as percentage of contamination increased. For data with 35% contamination, JRFit's relative efficiency versus REML increased from 122% for $\rho = 0$ to 134% for $\rho = 0.5$. In summary, while all methods lost efficiency with increasing contamination, REML lost efficiency at a higher rate. This gave an increasing relative efficiency (versus REML) curve for all the methods as contamination increased.

Similar observations have been made for JRFit versus REML in Kloke et al. (2009) who performed a simulation study limited to only two clusters where data were drawn from Gaussian and contaminated Gaussian with 20% contamination. The results from the Monte

Carlo simulation performed in this study demonstrate that the efficiency of JRFit holds for linear models with cluster-correlated errors in a larger setting. It was observed that rank methods tended to perform better than REML when the cluster structure exhibits strong correlation. While we simultaneously estimate and test the significance of effects, Galbraith et al. (2010) have also discussed WRS for clustered data (Rosner et al., 2003; Datta and Satten, 2005) from a testing perspective. They have also reported the perils of ignoring clustering from the perspective of inflated Type I error rates of tests.

**4. Conclusions and Recommendations**

The hydroclimatic variables such as temperature, precipitation, streamflow, baseflow and groundwater are typically not normally distributed and often contain outliers. Thus, the nonparametric WRS test has gained popularity for the analysis of such data due to its robustness to deviations from normality as well as the presence of outliers (Figure 3) (Diaz and Markgraf, 1992; Chiew et al., 1998; Tootle et al., 2005; Roy, 2006; Keener et al., 2010; Johnson et al., 2013; Mitra et al., 2014). However, these data display monthly or seasonal clustering and the WRS test does not properly account for the intra-cluster correlations. The purpose of this study was to evaluate the fidelity and efficiency of JRFit, an extension of WRS to modeling framework that accounts for cluster-correlation, against traditional statistical procedures. This was done via a Monte Carlo simulation experiment where datasets were generated under various scenarios. The efficiency of JRFit was compared to three traditional methods: restricted maximum likelihood, least absolute deviations and the RFit (a model-based equivalent of WRS) methods. The results confirmed that JRFit provides more efficient estimates of effects than the other three methods for clustered data with heavier tails (or data with outliers) or strong correlation. JRFit's efficiency gain was up to 50% as compared to REML for heavy tailed and highly correlated data. Researchers have extensively used the WRS method in past studies. However, our results conclusively show that using methods that fail to account for cluster correlations might lead to inefficiencies and possibly erroneous conclusions.

If interest lies in only testing, then the WRS methods of Rosner et al. (2003) or Datta and Satten (2005) that are specifically designed for clustered data may be used. However, if one is also interested in measuring effect sizes and prediction of future responses, then we recommend the use of JRFit that simultaneously provides estimation and testing. Moreover, the interaction of two climate phenomena can also be efficiently incorporated into the model and tested using the JRFit approach (Singh et al., 2015). The prediction errors of the JRFit provided the lowest mean absolute prediction error when intra-month correlations values were high. Thus, the nonparametric approach, JRFit, which was the focus of the this study, was not only found to be efficient for heavy tailed and contaminated datasets but also provided more consistent prediction of future values in the presence of cluster-correlation.

The results obtained from this study give credence to the importance of examining the coupled effect of interannual (e.g. ENSO) and multidecadal (e.g. AMO) climate variability phenomena. Incorporating decadal and multidecadal climatic cycles along with ENSO can help provide a clearer picture of climate impacts on baseflow. Moreover, the prediction standard errors of baseflow may further be reduced by incorporating other informative variables such precipitation, temperature, topographic elevations, etc. This can provide useful information to policymakers in devising water management policy and help in promoting drought severity-based water restrictions in this region. The linear mixed effects modeling framework used in JRFit is conducive for including more variables in the regression and performing model selection. For example, depending on the availability of data on several climatic/environmental variables, one may use the rank-based least absolute selection and shrinkage operator (LASSO) of Abebe and Bindele (2016) to simultaneously select the climatic/environmental variables that provide optimal prediction of hydrological processes as well as estimate their effect.

There are methods, such as generalized linear mixed models (Breslow and Clayton, 1993), that can be used to model clustered data from non-normal distributions, but they require the data distribution to be specified. JRFit is distribution-free; thus, investigators are not burdened with making a distributional choice. Despite its broad appeal JRFit has certain limitations. The computational burden of using JRFit for high dimensional data may be quite large. So, there is a need for improved algorithms to calculate JRFit estimates. A promising approach may be to extend the iterated reweighted least squares fitting approach (Sievers and Abebe, 2004; Miakonkana and Abebe, 2014) in combination with REML. Moreover, at the moment, JRFit can efficiently handle data for which the correlation structure is compound symmetric (Milliken and Johnson, 2004), but it is still of interest to develop a version of JRFit that can provide efficient estimation and prediction for hydroclimatic data whose correlation structure may not be compound symmetric. The current version of JRFit is freely available as an open source R package (R Core Team, 2017; Kloke, 2014).

**Acknowledgment**

The authors wish to acknowledge funding provided by the Alabama Agricultural Experiment Station, National Oceanic and Atmospheric Agency (NOAA) Sectoral Applications Research Program (SARP), and the National Integrated Drought Information System (NIDIS) Program for this study. The authors would like to thank the Editors and the three anonymous reviewers for constructive suggestions to improve the quality of the paper. Monthly streamflow data were obtained from the United States Geological Survey (USGS) Surface-Water Data for the Nation (https://waterdata.usgs.gov/nwis/sw). The monthly Niño 3.4 index values were obtained from the Climate Prediction Center, NOAA

(http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php). The Atlantic Multidecadal Oscillation (AMO) index values were obtained from the Physical Sciences Division of the Earth Systems Research Laboratory, NOAA (https://www.esrl.noaa.gov/psd/data/timeseries/AMO/). To create Figure 1, river basin boundary data was obtained from Hydrography, USGS (https://nhd.usgs.gov/index.html), the state boundary and stream network data from the National Map, USGS (https://viewer.nationalmap.gov/launch/), lake information from the Environmental Protection Division, Georgia Department of Natural Resources, (https://epd.georgia.gov/geographic-information-systems-gis-databases-and-documentation), and dam information from the National Map Small Scale, USGS (https://nationalmap.gov/small_scale/atlasftp.html).

**References**

Abebe, A., and Bindele, H. F., 2016. Robust Signed-Rank Variable Selection in Linear Regression. In Regina Liu and Joseph W. McKean, editors. Robust Rank-Based and Nonparametric Methods. Springer Proceedings in Mathematics & Statistics, 168, 25-45.

Adichie, J. N., 1967. Estimates of Regression Parameters Based on Rank Tests. Ann. Math. Statist., 38, 894-904.

Bates, D., Maechler, M., Bolker, B., and Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Beebee, R.A., and Manga M., 2004. Variation in the Relationship between Snowmelt Runoff in Oregon and ENSO and PDO. Journal of the American Water Resources Association, 40:1011–1024.

Bindele, H. F. and Abebe, A. (2015), Semi-parametric Rank Regression with Missing Responses. Journal of Multivariate Analysis, 142, 117-132.

Bradley, J. V., 1968. Distribution-Free Statistical Tests. PrenticeHall, Inc., Englewood Cliffs, New Jersey, 388 pp.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American statistical Association, 88(421), 9-25.

Bretz, F., Maurer, W., and Hommel, G., 2011. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. Statistics in Medicine, 30:1489–1501.

Cayan, D. R., Redmond, K. T., and Riddle, L. G., 1999. ENSO and Hydrologic Extremes in the Western United States. Journal of Climate, 12, 2881-2893.

Chen, W.Y., 1982. Assessment of Southern Oscillation sea-level pressure indices. Mon. Weather Rev. 110, 800–807.

Chiew, F. H. S., Zhou, S. L. and McMahon, T. A., 2003. Use of seasonal streamflow forecasts in water resources management. Journal of Hydrology, 270, 135-144.

Chiew, F.H.S., Piechota, T.C., Dracup, J.A., and McMahon, T.A., 1998. El Niño/Southern Oscillation and Australian rainfall, streamflow, and drought: links and potential for forecasting. Journal of Hydrology, 204, 138–149.

Climate Research Committee and National Research Council, 1995. Natural Climate Variability on Decade-to-Century Time Scales. Washington, DC: National Academy Press.

Couch, C. A., Hopkins, E. H., and Hardy, P. S., 1996. Influences of Environmental settings on aquatic ecosystems in the Apalachicola-Chattahoochee-Flint River basin: U.S. Geological Survey Water-Resources Investigations Report, 95-4278.

Crawford, C.G., Slack, J.R., and Hirsch, R.M., 1983. Nonparametric Tests for Trends in Water Quality Data using the Statistical Analysis System. USGS Open-File Report, 83-550.

Datta, S., and Satten, G. A., 2005. Rank-Sum Tests for Clustered Data. Journal of the American Statistical Association, 100 (471), 908-915.

Diaz, H. F. and Markgraf, V., 1992. El Niño: historical and paleoclimatic aspects of the Southern Oscillation. Cambridge University Press.

Earth Systems Research Laboratory (ESRL), 2012. Climate Timeseries: AMO (Atlantic Multidecadal Oscillation) Index. <http://www.esrl.noaa.gov/psd/data/timeseries/AMO/>

El-Shaarawi, A. H., 1989. Inferences about the Mean from Censored Water Quality Data. Water Resources Research, 25 (4), 685-690.

Enfield, D.B., Mestas-Nuez, A.M., and Trimble, P.J., 2001. The Atlantic multi-decadal oscillation and its relation to rainfall and river flows in the continental US. Geophysical Research Letters, 28 (10), 2077-2080.

Galbraith, S., Daniel, J. A., and Vissel, B., 2010. A Study of Clustered Data and Approaches to Its Analysis. The Journal of Neuroscience, 30(32), 10601–10608.

Gutierrez, F., and Dracup, J. A., 2001. An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño–Southern Oscillation indicators, J. Hydrol. Amsterdam, 246, 181–196.

Hansen, D. V., and Maul, G. A., 1991. Anticyclonic current rings in the eastern tropical Pacific Ocean. Journal of Geophysical Research, 96 (C4), 6965-6979.

Heiler, S. and Willers, R., 1988. Asymptotic normality of R-estimates in the linear model. Statistics, 19, 173–184.

Helsel, D.R. and R.M. Hirsch, 2002. Statistical Methods in Water Resources. US Geological Survey TWRI 4A3.

Hettmansperger, T. P. and McKean, J. W., 2011. Robust nonparametric statistical methods, volume 119 of Monographs on Statistics and Applied Probability. CRC Press, Boca Raton, FL, second edition.

Hodges, J. L. and Lehmann, E. L., 1963. Estimates of location based on rank tests, Ann. Math. Statist., 34, 598-611.

Hogg, R. V., McKean, J. W., and Craig, A.T., 2005. Introduction to mathematical statistics. 6th Edition Pearson Prentice Hall; Upper Saddle River, N.J.

Huber, P. J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35, 73–101.

Jaeckel, L. A., 1972. Estimating regression coefficients by minimizing the dispersion of the residuals. Annals of Mathematical Statistics, 43, 1449–1458.

Johnson, N. T., Martinez, C. J., Kiker, G. A., and Leitman, S., 2013. Pacific and Atlantic sea surface temperature influences on streamflow in the Apalachicola–Chattahoochee–Flint river basin. Journal of Hydrology, 489, 160–179.

Jureckova, J., 1971. Nonparametric Estimate of Regression Coefficients. Ann. Math. Statist., 42, 1328-1338.

Kahya, E., and Dracup, J.A., 1993. US streamflow patterns in relation to the El Niño/Southern Oscillation. Water Resour. Res., 29 (8), 2491–2503.

Keener, V.W., Feyereisen, G.W., Lall, U., Jones, J.W., Bosch, D.D., and Lowrance, R., 2010. El-Niño/Southern Oscillation (ENSO) influences on monthly NO3 load and concentration, stream flow and precipitation in Little River Watershed, Tifton, Georgia (GA). J. Hyd. 381: 352-363.

Kiladis, G.N., and Diaz, H.F., 1989. Global climate anomalies associated with extremes in the Southern Oscillation. Journal of Climate, 2 (9), 1069–1090.

Kloke, J.D., 2014. Joint Ranking Estimates. GitHub repository, https://github.com/kloke/jrfit/.

Kloke, J.D., McKean, J.W., and Rashid, M., 2009. Rank-Based Estimation and Associated Inferences for Linear Models with Cluster Correlated Errors. Journal of the American Statistical Association, 104 (485), 384-390.

Koenker, R., and Basset, G., 1978, “Regression Quantiles” Econometrica, 46, p.33-50.

Koul, H. L., Sievers, G., McKean, J. W., 1987. An estimator of the scale parameter for the rank analysis of linear models under general score functions. Scandinavian Journal of Statistics, 14, 131–141.

Lee, J.H., and Julien, P.Y., 2016. Teleconnections of the ENSO and South Korean precipitation patterns, Journal of Hydrology 534, 237-250.

Lehmann, E., 1975. Nonparametrics: Statistical Methods based on Ranks. San Francisco, USA: Holden-Day.

Lim, K.J., Engel, B.A., Tang, Z., Choi, J., Kim, K., Muthukrishnan, S., and Tripathy, D., 2005. Web GIS-based Hydrograph Analysis Tool, WHAT. JAWRA, 41(6): 1407-1416.

MacDonald, G.M., and Case, R.A., 2005. Variations in the Pacific Decadal Oscillation over the past millennium. Geophysical Research Letters, 32(8): L08703.

Miakonkana, G., and Abebe, A., 2014. Iterative Rank Estimation for Generalized Linear Models. Journal of Statistical Planning and Inference, 151/152, 60-72.

Milliken, G. A., and Johnson, D. E., 2004. Analysis of Messy Data, Volume 1: Designed Experiments, Second Edition. Chapman and Hall/CRC, New York.

Mitra, S., Srivastava, P., Singh, S., and Yates, D., 2014. Effect of ENSO-Induced Climate Variability on Groundwater Levels in the Lower Apalachicola-Chattahoochee-Flint River Basin. Transactions of the ASABE, 57(5), 1393-1403.

Philander, S. G., 1990. El Niño, La Niña and the Southern Oscillation. Elsevier, New York.

Poveda, G., Jaramillo, A., Gil, M. M., Quiceno, N., and Mantilla, R., 2001. Seasonality in ENSO related precipitation, river discharges, soil moisture, and vegetation index (NDVI) in Colombia. Water Resour. Res., 37, 2169–2178.

R Core Team, 2017. R: A Language and Environment for Statistical. R Foundation for Statistical Computing. <https://www.R-project.org>

Räsänen, T.A., and Kummu, M., 2012. Spatiotemporal influences of El Niño Southern Oscillation (ENSO) on precipitation and flood pulse in the Mekong River Basin. Journal of Hydrology, 476(7):154-168.

Regonda, S.K., Rajagopalan, B., Clark, M., and Pitlick, J., 2005. Seasonal Cycle Shifts in Hydroclimatology over the Western United States. Journal of Climate 18 (2), 372–384.

Ropelewski, C.F., and Halpert, M.S., 1986. North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation. Monthly Weather Review, 114, 2352– 2362.

Rosner, B., Glynn, R. J., and Lee, M.L., 2003. Incorporation of clustering effects for the Wilcoxon rank sum test: A large-sample approach. Biometrics, 59:1089–1098.

Rousseeuw, P.J., and Leroy, A.M., 1987. Robust Regression and Outlier Detection, WileyInterscience, New York (Series in Applied Probability and Statistics), 329 pages. ISBN 0-471-85233-3.

Roy, S.S., 2006. The impacts of ENSO, PDO, and local SSTs on winter precipitation in India. Physical Geography, 27(5), 464-474.

Schlesinger, M. E., and Ramankutty, N., 1994. An oscillation in the global climate system of period 65-70 years. Nature, 367, 723-726.

Schmidt, N., Lipp, E. K., Rose, J. B., and Luther, M. E., 2001. ENSO Influences on Seasonal Rainfall and River Discharge in Florida. Journal of Climate 14: 615-628.

Schulte, J.A., Najjar, R.G., and Lee, S., 2017. Salinity and streamflow variability in the Mid-Atlantic region of the United States and its relationship with large-scale atmospheric circulation patterns, Journal of Hydrology 550, 65-79.

Seber, G. A. F., and Lee, A. J., 2003. Linear Regression Analysis. Wiley Series in Probability and Statistics, Hoboken, NJ, USA: Wiley.

Sievers, G., and Abebe, A., 2004. Rank estimation of regression coefficients using iterated reweighted least squares. Journal of Statistical Computation and Simulation, 74: 821-831.

Singh, S., Srivastava, P., Abebe, A. and Mitra, S., 2015. Baseflow response to climate variability induced droughts in the Apalachicola–Chattahoochee–Flint River Basin, U.S.A. Journal of Hydrology, 528, 550–561.

Steirou, E., Gerlitz, L., Apel, H., and Merz, B., 2017. Links between large-scale circulation patterns and streamflow in Central Europe: A review. Journal of Hydrology 549, 484-500.

Tootle, G.A., Piechota, T.C., and Singh, A., 2005. Coupled oceanic–atmospheric variability and US streamflow. Water Resour. Res., 41, 1–11.

Trenberth, K.E., 1997. The definition of El Niño. Bull. Am. Meteorol. Soc. 78, 2771–2777.

Trenberth, K.E., Stepaniak, D.P., 2001. Indices of El Niño evolution. J. Clim. 14, 1697–1701.

West, B. T., Welch, K. B., and Gałecki, A. T., 2007. Linear Mixed Models: A Practical Guide Using Statistical Software. Chapman & Hall/CRC.

Wong, R. K. W., Yao, F., and Lee, T. C. M., 2014. Robust Estimation for Generalized Additive Models. Journal of Computational and Graphical Statistics, 23, 270–289.

Zorn, M. R., and Waylen, P. R., 1997. Seasonal response of mean monthly streamflow to El Nin˜o/Southern Oscillation in north central Florida, Prof. Geogr., 49, 51– 62.

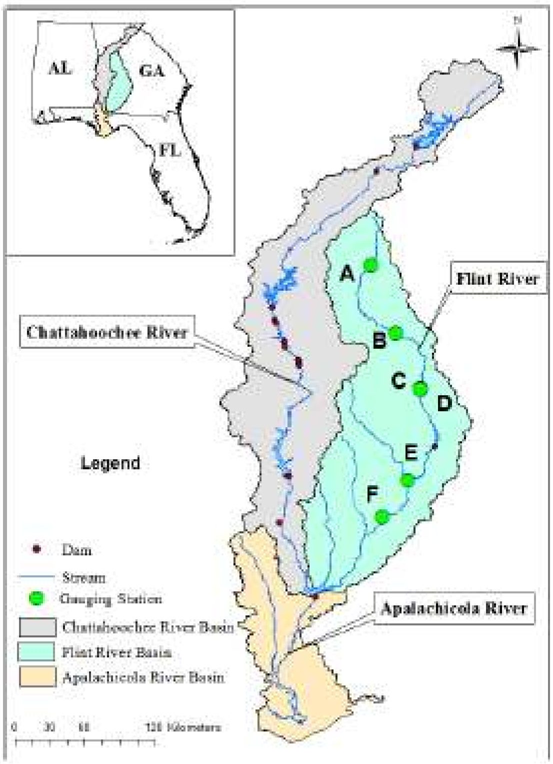


**Figure 1:** Map of Apalachicola-Chattahoochee-Flint (ACF) River basin showing the location of the gauging stations selected for this study. The ACF River basin is located in Alabama, Georgia, and Florida. The streamflow gauging stations are shown as green dots and it is to be noted that the flows in the Flint River are mostly unregulated.

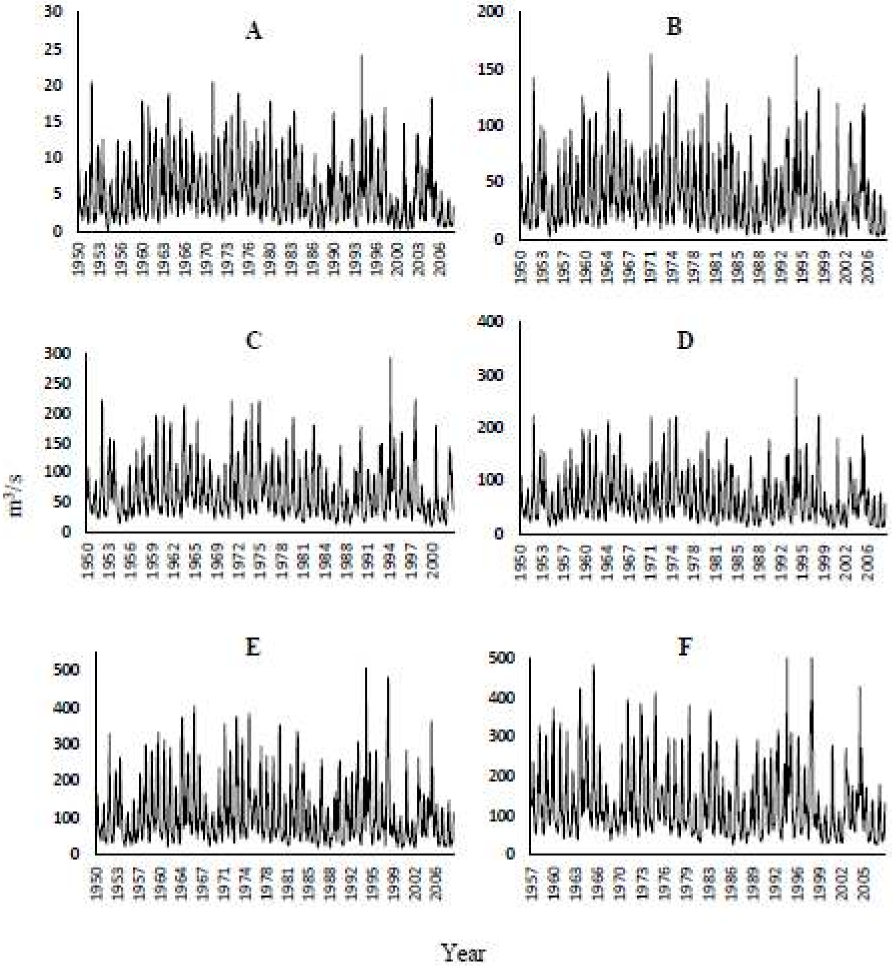


**Figure 2:** Monthly baseflow ($m^3/s$) time series plots for stations A, B, C, D, E, and F.

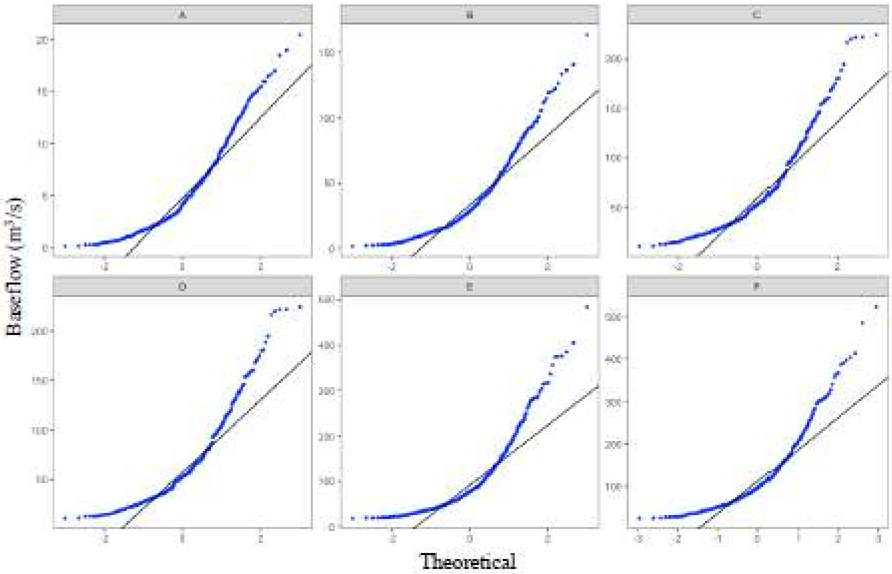


**Figure 3:** Quantile - Quantile plots for stations A, B, C, D, E, and F.

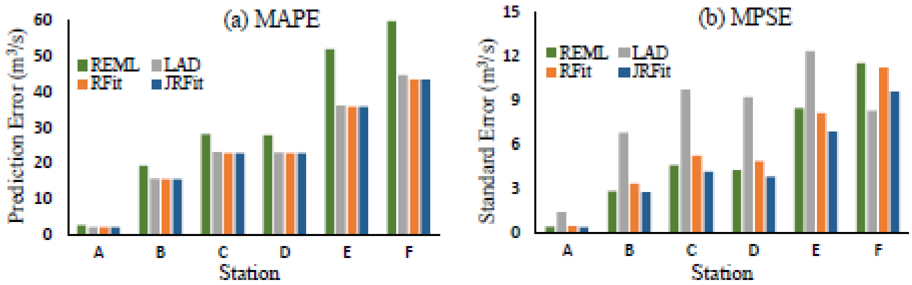


**Figure 4:** (a) Mean Absolute Prediction Error (MAPE) and (b) Mean Prediction Standard Error (MPSE) of Restricted Maximum Likelihood (REML), Least Absolute Deviations (LAD), Rank-Based Fit (RFit) and Joint Rank Fit (JRFit) across all the stations.

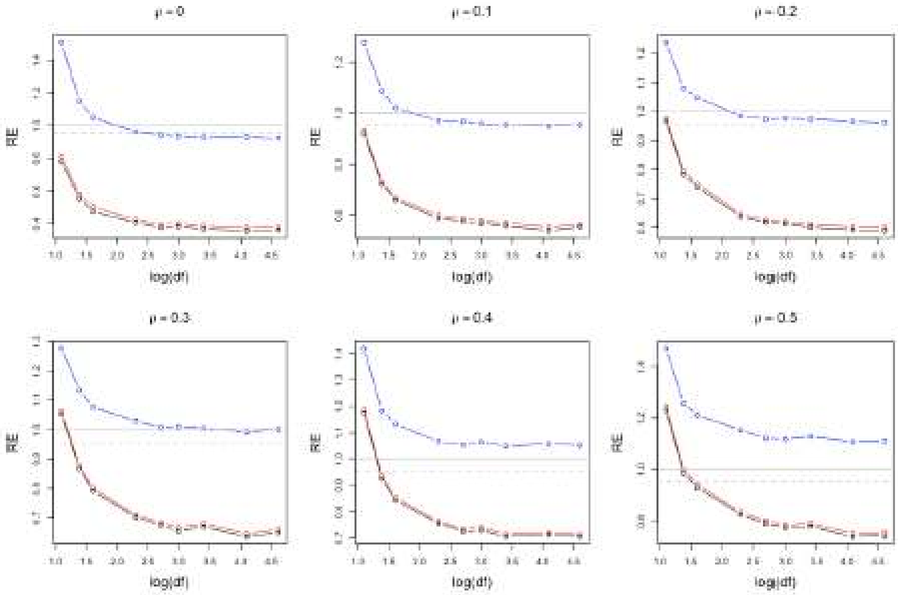


**Figure 5:** Estimated relative efficiencies (RE) versus Restricted Maximum Likelihood (REML): Joint Rank Fit (JRFit), Least Absolute Deviations (LAD), and Rank-Based Fit (RFit). Dashed line represents the theoretical ARE $3/\pi$ of JRFit vs REML for Gaussian ($df = \infty$) case when $\rho = 0$.

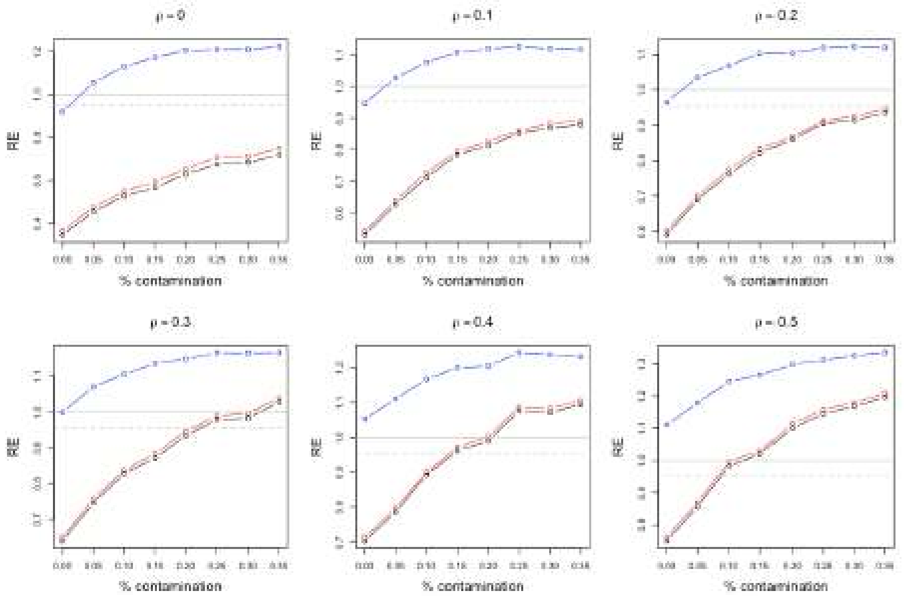


**Figure 6:** Estimated relative efficiencies (RE) versus Restricted Maximum Likelihood (REML): Joint Rank Fit (JRFit), Least Absolute Deviations (LAD), and Rank-Based Fit (RFit). Dashed line represents the theoretical ARE $3/\pi$ of JRFit vs REML for Gaussian $(\delta = 0)$ case when $\rho = 0$.

**Table 1:** Streamflow gauging stations used in this study showing the USGS station ID, location, their assigned names used in the manuscript, and their respective date ranges.

| Station ID | Location | Latitude | Longitude | Given Name | Drainage Area ($km^2$) | Data Range (Year) |
|---|---|---|---|---|---|---|
| 02344500 | Flint River near Griffin, GA | 33.244 | -84.429 | **A** | 704 | 1950-2008 |
| 02347500 | Flint River near Carsonville, GA | 32.721 | -84.232 | **B** | 4791 | 1950-2008 |
| 02349500 | Flint River at Montezuma, GA | 32.298 | -84.044 | **C** | 7511 | 1950-2003 |
| 02349605 | Flint River near Montezuma, GA | 32.293 | -84.043 | **D** | 7563 | 1950-2008 |
| 02352500 | Flint River, Albany, GA | 31.594 | -84.144 | **E** | 13753 | 1950-2008 |
| 02353000 | Flint River, Newton, GA | 31.307 | -84.339 | **F** | 14867 | 1957-2008 |

**Table 2:** Expected baseflow values under Model 1 and Model 2. ENSO = 0 and ENSO = 1 represent the La Niña and El Niño phases of ENSO, and AMO = 0 and AMO = 1 represent the positive and negative phases of AMO, respectively.

| ENSO | AMO | Expected baseflow under Model 1 | Expected baseflow under Model 2 |
|---|---|---|---|
| 0 | 0 | $\beta_0$ | $\beta_0$ |
| 1 | 0 | $\beta_0 + \beta_1$ | $\beta_0 + \beta_1$ |
| 0 | 1 | $\beta_0 + \beta_2$ | $\beta_0 + \beta_2$ |
| 1 | 1 | $\beta_0 + \beta_1 + \beta_2$ | $\beta_0 + \beta_1 + \beta_2 + \beta_3$ |

**Table 3:** Coefficient of estimation and standard error values (in the bracket) for Model 1 and Model 2 for Restricted Maximum Likelihood (REML), Least Absolute Deviations (LAD), Rank-Based Fit (RFit), and Joint Rank Fit (JRFit). The values represent changes in baseflow in $m^3/s$ per changes in climate variable phase. Numbers in bold are not significant.

| Station | Effect | REML | | LAD | | Rfit | | JRFit | |
|---|---|---|---|---|---|---|---|---|---|
| | | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 |
| A | ENSO | 0.816 | 2.445 | **0.656** | 2.329 | **0.570** | 1.541 | 0.571 | 1.540 |
| | | (0.290) | (0.399) | **(0.490)** | (0.727) | **(0.299)** | (0.438) | (0.235) | (0.347) |
| | AMO | 1.544 | 3.136 | 2.035 | 2.570 | 1.275 | 2.212 | 1.274 | 2.212 |
| | | (0.290) | (0.394) | (0.490) | (0.719) | (0.299) | (0.433) | (0.219) | (0.361) |
| | ENSO*AMO | | -3.183 | | -2.492 | | -1.946 | | -1.946 |
| | | | (0.558) | | (1.016) | | (0.612) | | (0.331) |
| B | ENSO | 7.773 | 19.155 | **7.812** | 15.063 | 5.199 | 11.380 | 5.198 | 11.384 |
| | | (1.970) | (2.700) | **(3.995)** | (5.251) | (2.181) | (3.189) | (1.680) | (2.552) |
| | AMO | 9.324 | 20.443 | 8.142 | 12.549 | 6.733 | 12.775 | 6.734 | 12.773 |
| | | (1.969) | (2.668) | (3.996) | (5.191) | (2.182) | (3.153) | (1.645) | (2.636) |
| | ENSO*AMO | | -22.232 | | -15.939 | | -12.446 | | -12.447 |
| | | | (3.775) | | (7.342) | | (4.460) | | (2.447) |
| C | ENSO | 9.951 | 29.246 | **9.295** | 22.366 | 7.207 | 19.705 | 7.207 | 19.698 |
| | | (3.126) | (4.451) | **(5.179)** | (5.747) | (3.418) | (5.045) | (2.683) | (3.968) |
| | AMO | 10.452 | 28.025 | **7.571** | 20.058 | 7.111 | 17.800 | 7.110 | 17.799 |
| | | (3.143) | (4.248) | **(5.206)** | (5.482) | (3.436) | (4.812) | (2.179) | (3.477) |
| | ENSO*AMO | | -35.076 | | -25.102 | | -22.540 | | -22.531 |
| | | | (6.000) | | (7.744) | | (6.798) | | (3.673) |
| D | ENSO | 11.482 | 29.651 | **9.347** | 22.366 | 7.221 | 17.826 | 7.230 | 17.823 |
| | | (2.973) | (4.055) | **(5.032)** | (5.538) | (3.282) | (4.577) | (2.419) | (3.365) |
| | AMO | 12.239 | 29.988 | **7.607** | 20.058 | 8.530 | 18.473 | 8.533 | 18.475 |
| | | (2.973) | (4.007) | **(5.033)** | (5.474) | (3.283) | (4.524) | (2.095) | (3.587) |
| | ENSO*AMO | | -35.484 | | -25.102 | | -20.724 | | -20.727 |
| | | | (5.669) | | (7.744) | | (6.400) | | (3.675) |
| E | ENSO | 28.668 | 57.529 | **13.314** | 32.930 | 13.443 | 27.804 | 13.434 | 27.800 |
| | | (5.768) | (7.997) | **(8.004)** | (9.538) | (5.194) | (7.662) | (4.680) | (6.329) |
| | AMO | 22.849 | 51.041 | **15.178** | 28.262 | 14.157 | 27.649 | 14.166 | 27.650 |
| | | (5.765) | (7.903) | **(8.005)** | (9.429) | (5.195) | (7.574) | (3.887) | (6.238) |
| | ENSO*AMO | | -56.368 | | -29.785 | | -28.882 | | -28.876 |
| | | | (11.179) | | (13.337) | | (10.713) | | (6.524) |
| F | ENSO | 24.820 | 59.828 | **7.271** | **15.914** | **9.866** | 24.047 | 9.886 | 24.053 |
| | | (6.997) | (11.058) | **(10.340)** | **(17.726)** | **(6.662)** | (10.762) | (4.440) | (6.447) |
| | AMO | 23.288 | 54.108 | 21.437 | **26.039** | 18.689 | 30.929 | 18.687 | 30.958 |
| | | (7.173) | (10.375) | (10.599) | **(16.615)** | (6.829) | (10.088) | (5.347) | (8.948) |
| | ENSO*AMO | | -56.651 | | **-16.673** | | **-23.787** | | -23.806 |
| | | | (14.082) | | **(22.560)** | | **(13.698)** | | (8.890) |

**Table 4:** Mean Absolute Prediction Error (MAPE) and Mean Prediction Standard Error (MPSE) in $m^3/s$ of Restricted Maximum Likelihood (REML), Least Absolute Deviations (LAD), Rank-Based Fit (RFit), and Joint Rank Fit (JRFit).

| Station Name | Mean values | REML | LAD | RFit | JRFit | Intra-month Correlation | Selected Optimal Procedure |
|---|---|---|---|---|---|---|---|
| **A** | MAPE | 2.733[a] | 2.245[b] | 2.225[b] | 2.225[b] | 54.50% | JRFit |
| | MPSE | 0.421[a] | 1.414[b] | 0.465[c] | 0.384[a] | | |
| **B** | MAPE | 19.446[a] | 15.922[b] | 15.622[b] | 15.621[b] | 61.04% | JRFit |
| | MPSE | 2.849[a] | 6.794[b] | 3.346[c] | 2.772[a] | | |
| **C** | MAPE | 28.246[a] | 23.322[b] | 23.017[b] | 23.016[b] | 59.20% | JRFit |
| | MPSE | 4.609[a] | 9.714[b] | 5.249[c] | 4.145[d] | | |
| **D** | MAPE | 27.959[a] | 23.120[b] | 23.051[b] | 23.048[b] | 59.00% | JRFit |
| | MPSE | 4.278[a] | 9.188[b] | 4.884[c] | 3.811[d] | | |
| **E** | MAPE | 52.105[a] | 36.254[b] | 35.942[b] | 35.942[b] | 55.00% | JRFit |
| | MPSE | 8.439[a] | 12.320[b] | 8.123[c] | 6.852[d] | | |
| **F** | MAPE | 59.951[a] | 44.731[b] | 43.554[b] | 43.553[b] | 55.30% | LAD/RFit/ JRFit |
| | MPSE | 11.503[a] | 8.280[ab] | 11.201[b] | 9.573[b] | | |

For each station, MAPE and MPSE results with different superscripts (a, b, c and d) indicate significant difference between procedures according to paired t-test comparison followed by a Bonferroni correction.