Advances in Water Resources 143 (2020) 103672

Contents lists available at ScienceDirect

Advances in Water Resources

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

A comparison of estimators of the conditional mean under non-stationary conditions

Richard M. Vogel[a,*], Charles N. Kroll[b]

[a] Department of Civil and Environmental Engineering, Tufts University, Medford, MA, United States
[b] Environmental Resources Engineering, SUNY ESF, Syracuse, NY, United States

ARTICLE INFO

Keywords:
Floods
Droughts
Natural hazards
Hydrologic design
Urbanization
Climate change
Streamflow
Rivers
Infrastructure
Hydraulic
Efficiency
Robustness
Resistance
Parsimony
Subset
Update
Current mean

ABSTRACT

There is increasing attention to the development of a myriad of complex methods for nonstationary frequency analysis (NFA) of floods, droughts and other hydrologic processes. We assume that the need for NFA arises from well understood deterministic mechanisms of change. A common assumption in NFA, questioned here, is that more accurate estimators of hydrologic statistics result when more realistic, complex and sophisticated models are employed. By considering the mean annual flood (drought or other hydrologic event), general conditions are derived when the sample mean (*SM*) is a more efficient (lower mean square error, *MSE*) estimator than a regression estimate of the mean (*RM*). We introduce an optimal fractional mean estimator, FM^*, which is simply the *SM* of the most recent period of record nf^*, where f^* is the optimal fraction of the full sample n, which leads to minimum *MSE* among all possible values of f. Interestingly, FM^* is generally preferred over *RM* for attained significance levels associated with the fitted regression model in excess of about 0.05. Given the considerable attention and uncertainty surrounding potential nonstationary conditions, we demonstrate that a parsimonious estimator which exploits an optimal recent subset of the historical record may be more attractive than many of the more complex nonstationary approaches commonly advocated.

1. Introduction

One of the most common challenges facing hydrologists involves a determination of ‘design events’ associated with floods, droughts, rainfall, sediment and other constituent loads as well as a host of other important hydroclimatic variables. Such ‘design events’ are normally determined by developing the relationship between the magnitude and frequency of the hydroclimatic variable of interest. This study applies to nearly any type of hydrologic frequency analysis (HFA) however most of our discussions pertain to flood frequency analysis (FFA) and drought frequency analysis (DFA). Readers are encouraged to envision and apply the findings of this study to other areas of HFA, such as its most obvious extension to frequency analysis of rainfall, sediment and other constituent loads. Critical to nearly all traditional approaches to HFA in general, and FFA and DFA in particular, is the assumption of stationarity, loosely defined as conditions when key population statistics of the variable of interest, such as their moments or L-moments, and/or probability distribution function (PDF) parameters, do not systematically change over time. We distinguish between stationary and nonstationary FFA using the notation SFFA and NFFA, respectively.

Over the past decade, there has been a surge of literature on the topic of NFFA as evidenced from recent review articles and special journal issues (e.g. Khaliq et al., 2006; Petrow and Merz, 2009; Kiang et al., 2011; Salas et al., 2012; Madsen et al., 2013; Hall et al., 2014; Bayazit, 2015; Salas et al., 2018; Francois et al., 2019). In spite of this recent surge of literature on NFFA, Serago and Vogel (2018) describe the current situation: (1) there is no consensus on the need for NFFA, (2) considerable debate exists over whether one should use SFFA or NFFA, in practice, and (3) there is no consensus on an appropriate design event to employ under nonstationary conditions. Due to the tremendous uncertainty associated with the impacts of climate change on water resources, this lack of consensus is to be expected for studies which attempt to capture the impact of climate change on DFA and FFA (e.g. Koutsoyiannis et al., 2008; Hirsch and Ryberg, 2012). However, Serago and Vogel (2018), Blum et al. (2019) and Hecht and Vogel (2020) cite numerous reasons why the need for NFFA in urbanized or urbanizing watersheds is paramount. Those three studies argue that when historical trends in

* Corresponding author.
E-mail address: richard.vogel@tufts.edu (R.M. Vogel).

https://doi.org/10.1016/j.advwatres.2020.103672
Received 6 April 2020; Received in revised form 21 June 2020; Accepted 26 June 2020
Available online 27 June 2020

streamflow series are obvious due to documented changes in historical land use and/or water infrastructure, it is imperative to provide updated estimates of design events that reflect current conditions.

1.1. When to consider a nonstationary frequency analysis?

It is not our goal to provide a comprehensive analysis of the general conditions under which a stationary or nonstationary FFA is to be employed. Even when change is evident in a hydrologic record, such changes may not arise from nonstationary processes (Cohn and Lins, 2005; Koutsoyiannis and Montanari, 2015). Importantly, one is ill-advised to employ NFFA on the basis of information obtained only from available hydrologic records (Koutsoyiannis, 2016; Luke et al., 2017; Serinaldi and Kilsby, 2015; Serinaldi et al., 2018). Given these concerns, it is only when a very clear physical understanding of the deterministic causes of nonstationary behavior are present that a nonstationary analysis is warranted, and this is particularly true in any analysis involving extrapolation of historical change into the future. Koutsoyiannis and Montanari (2015) suggest a simple rule to decide if a nonstationary analysis is warranted by answering the question: can the change be predicted in deterministic terms? They argue that only if the answer is positive is it legitimate to invoke nonstationarity. Examples of such physical drivers of nonstationary behavior which could be predicted in deterministic terms include: changes in climate, numerous forms of land use change (between forest, agricultural, and urban), increased degree of artificial drainage (such as tile drainage or ditching), and changing agricultural practices (including conservation tillage or irrigation).

A primary assumption inherent in our analysis is that a clear physical basis and understanding exists concerning the mechanisms which result in the nonstationary process under consideration. Thus our study does not answer the question of whether or not to employ a stationary or nonstationary analysis, but rather, we assume that the flood, drought or other hydrologic series of interest is known to exhibit nonstationary behavior which can be represented in the form of a mathematical model. Our analysis then attempts to evaluate a number of different approaches for estimation of a relatively simple and important statistic, the "current mean" (i.e. the conditional mean) of that nonstationary hydrologic process. Here we consider a very simple nonstationary model (a linear regression model) and a very simple statistic (conditional mean), as opposed to estimating an extreme quantile under more complex nonstationary conditions.

Consider the very common problem in which a hydrologic record of annual maximum (or minimum) streamflow is available which has been subject to change due to known urbanization processes. In such situations the mean annual flood (MAF), along with all other associated hydrologic statistics, such as the T-year design event, will change over time. Thus an important and common question which forms the basis of this study is how to update the MAF or other hydrologic statistic such as the T-year design event to reflect current conditions (Serago and Vogel, 2018; Blum et al., 2019; Hecht and Vogel, 2020) when nonstationary processes are known, apriori, to govern hydrologic behavior. Our central goal is to examine the level of model and associated estimation complexity needed to update the mean of a hydrologic process to reflect current conditions. A natural tradeoff exists between the level of complexity and the reliability of the estimator to be employed. This tradeoff involves an understanding of the relationship between the sampling properties (i.e. bias and mean square error) associated with various estimators of the design statistic of interest and the sample size and significance level (or goodness-of-fit) associated with the fitted nonstationary model under consideration.

The central challenge addressed by this study involves the appropriate level of complexity associated with the estimator of a hydrologic design event when the hydrologic process under consideration is known to exhibit nonstationary behavior. This question is likely to remain an open question for some time to come, as evidenced from the recent findings of O'Brien and Burn (2014), Serinaldi and Kilsby (2015), Luke et al. (2017), Yu and Stedinger (2018) and Serinaldi et al. (2018). In a careful comparison of the precision (uncertainty) associated with various estimates of design flood events using both stationary and nonstationary methods, Luke et al. (2017) found that stationary methods were nearly always preferred over nonstationary approaches. Here we further explore this hypothesis.

1.2. On the need for parsimonious models

Serago and Vogel (section 2, 2018) review the fundamental virtues of most widely accepted methods of SFFA which relate to their proven properties of robustness, resistance and efficiency, all of which stem from the principle of parsimony, and all of which have been given little or no attention in the area of NFFA. It has long been known that efficient (low *MSE*) estimators also tend to be parsimonious (Box and Jenkins, 1976). A parsimonious estimator is one that accomplishes a desired level of prediction efficiency with as few model parameters as possible. Serago and Vogel (section 2, 2018) document clear evidence of the value of parsimonious models in both SFFA (Kuczera, 1982; Slack et al., 1975; Lu and Stedinger, 1992; Laio et al., 2009; and Di Baldassarre et al., 2009) and NFFA (Serinaldi and Kilsby, 2015; Luke et al., 2017; and Yu and Stedinger, 2018). Serago and Vogel (2018) and others provide extensive evidence of the value of parsimony in SFFA, and after the three recent studies by Serinaldi and Kilsby (2015), Luke et al. (2017), and Yu and Stedinger (2018), we expect the principle of parsimony to play an increasing and equally important role in NFFA. The primary goal of this study is to contrast the general performance (in terms of robustness and efficiency) of various estimators of the conditional mean with an emphasis on nonstationary conditions.

1.3. Use of a subset of historical record to reflect current watershed conditions

Estimation of streamflow statistics which reflect current watershed conditions is challenging under nonstationary conditions due to the ever present and natural stochastic aspect of streamflow combined with the fact that historical streamflow records may not be representative of current watershed conditions if environmental conditions are changing. Under nonstationary conditions, multiple anthropogenic impacts often occur simultaneously, making it very difficult to attribute and model changes in streamflow as a function of climate, land use, water use and/or other watershed characteristics (Hirsch, 2011; Allaire et al., 2015). As an alternative to the development of a predictive model of streamflow, some authors have suggested using a recent subset of the streamflow record that better reflects current conditions at a site than the entire historical record (Riggs, 1972; Gebert et al., 2016; Blum et al., 2019). For example, in an evaluation of the relationship between magnitude and frequency of low flows for 15 basins in Wisconsin, Gebert et al. (2016) recommended use of the 1969–2008 period for determining current streamflow characteristics for design and compliance purposes, because that period was shown to represent current land use and climatic condition, and was generally free of trends and thus could be considered approximately stationary. Similarly, Riggs (1972) suggests that "[i]f the pattern of regulation of a stream has been consistent for several years and is expected to continue so, low-flow frequency curves based on the record for those years may be useful."

Blum et al. (2019) considered estimation of the 7-day 10-year low flow statistic (7Q10) under nonstationary conditions. Blum et al. (2019) performed Monte-Carlo experiments to evaluate the performance of a stationary nonparametric quantile estimator applied to a recent subset of the (nonstationary) annual minimum flow record, in comparison with using the complete historical flow record. They found that a nonparametric stationary quantile estimator fit using the most recent 30 or 50 years of a 75-year record led to improved accuracy and reduced bias of 7Q10 estimators when compared to using the entire flow record, under nonstationary conditions. This study takes an analogous approach, with the goal of obtaining more general conclusions concerning the value of using the sample mean of a subset of the recent observed record (which we term the fractional mean, *FM*) as compared to a more traditional nonstationary regression model estimator of the mean streamflow (which we term *RM*) or employing the entire historical record (which we term SM).

2. Methodology

2.1. Theoretical background and estimators

To enable us to develop general recommendation regarding the behavior of various hydrologic statistics under nonstationary conditions, in this initial study, we only consider estimation of the current mean, defined as the mean under current conditions. Under stationary conditions a true mean exists, or in other words, the current mean is always the true mean. However, the central focus of this work is on nonstationary conditions in which case we define the current mean as the conditional mean under 'current' watershed conditions.

Although estimation of the mean may seem trivial compared with other statistics, nearly all SFFA and NFFA methods begin with estimation of the mean. Consider the problem of fitting a nonstationary PDF to a series of annual maximum floods (AMF) denoted by x or its natural logarithm y. Any application of either SFFA or NFFA will typically involve estimation of the mean of either x or y (or both), depending on which PDF estimation method is used. Our results are general and apply to the mean μ_z, of any variable z regardless of its PDF. In this section we introduce several estimators of μ_z suitable for NFFA. An example of the importance of estimation of the mean would be in low flow frequency analysis when the annual minimum 7-day low flows follow a lognormal distribution. In that case the widely used 7-day 2-year low flow statistic is simply the mean of the logarithms of the annual minimum 7-day flows.

2.2. Conditional mean of a nonstationary hydrologic process

Most existing approaches to NFFA, summarized in review articles by Khaliq et al. (2006), Petrow and Merz (2009), Kiang et al. (2011), Salas et al. (2012), Madsen et al. (2013), Ehret et al. (2014), Hall et al. (2014), Bayazit (2015), Salas et al. (2018), and Francois et al. (2019), involve fitting PDFs whose parameters and/or moments are related to exogenous variables which are in turn are related to drivers of nonstationary behavior. Here we relate streamflow to an exogenous variable using a bivariate regression model:

$$z = \mu_z + \beta\left(w - \mu_w\right) + \varepsilon \tag{1}$$

where z and μ_z represent the streamflow variable of interest and its unconditional mean, respectively, w and μ_w represent an explanatory variable and its mean, respectively, β is the trend slope coefficient and ε is model error which is assumed to have zero mean and constant variance with $\sigma_\varepsilon^2 = \sigma_z^2(1-\rho^2)$ where ρ denotes the Pearson correlation coefficient between $z$and w.

The model in (1) is not limited to linear relationships because a wide range of monotonic non-linear functions can be linearized with transformations, such as Tukey's ladder-of-powers (Helsel et al., 2020), enabling the application of a variety of linear regression methods for fitting highly nonlinear relationships. In addition to the ability to model nonlinear trends, there are numerous other advantages of regression methods outlined by Serago and Vogel, 2018 including: rigorous graphical displays, parsimony, prediction intervals associated with trend extrapolations, accommodation of complex multivariate relationships, and an ability to account for missing observations, abrupt changes, and the impact of stochastic persistence.

The streamflow variable z could be a series of streamflows, x, their natural logarithms, y, or some other suitable transformation. Here we emphasize that the unconditional mean, μ_z, is of little value to NFFA because it does not generally exist for a nonstationary process, though it can be interpreted as the conditional mean of z when $w = \mu_w$

The model error term ε in (1) is often treated as an independent and identically distributed random variable, in spite of our now widespread knowledge that some hydrologic processes exhibit heteroscedasticity (Hecht and Vogel, 2020), stochastic persistence and possibly deterministic trends (Cohn and Lins, 2005). Ignoring stochastic persistence in (1), when it exists, would lead to incorrect statistical inference concerning the variance of the estimators of the model parameters β and μ_z, issues central to this study. Matalas and Sankarasubramanian (2003) provide approximations to the inflation in the variance associated with estimators of β in (1) when ε arises from Markov, ARMA(1,1) and Fractional Gaussian Noise stochastic processes. However, it is extremely difficult to identify such higher order stochastic persistence structures in typical hydrologic records (see examples in Vogel et al., 1998), thus in this initial study we assume that ε arises from a serially independent stochastic process and we encourage future investigators to employ the results of Matalas and Sankarasubramanian (2003) to extend the analysis which follows.

The regression coefficient β in (1) is defined by

$$\beta = \rho\frac{\sigma_z}{\sigma_w} \tag{2}$$

where σ_z and σ_w denote the standard deviation of z and w, respectively.

The conditional mean is the expected value of z conditioned upon w, denoted $\mu_{z|w}$, and is obtained by taking the expectation of (1) to obtain:

$$\mu_{z|w} = \mu_z + \beta\left(w - \mu_w\right) \tag{3}$$

because $E[\varepsilon] = 0$. See Serago and Vogel (2018) for the derivation of other conditional moments of regression in (1). Numerous authors have employed a regression model of the form given in (3) in NFFA. For example, Vogel et al. (2011), Prosdocimi et al. (2014) and Brady et al. (2019) found (1) (with z equal to the logarithm of the annual maximum flood (AMF) series and w equal to time) to be useful for modeling flood series at hundreds of rivers in the U.S. and the U.K., regardless of whether or not trends exist. We highlight the multilevel or panel version of (1) employed by Brady et al. (2019) which accounts for interactions among sites in a region in a Bayesian framework. Hirsch and Ryberg (2012) used (1) to relate the natural log of AMF series at 200 long term rivers in the US to the covariate w equal to carbon dioxide concentrations. We emphasize that the form of the regression model in (1) and (2) makes **no** assumption regarding the PDF of z.

2.3. Regression estimator of conditional mean, RM

A regression estimator of the conditional mean, $\mu_{z|w}$, termed RM, may be obtained using ordinary least squares estimators, in which case:

$$RM = \bar{z} + \hat{\beta}(w - \bar{w}) \tag{4}$$

where

$$\hat{\beta} = \hat{\rho}\frac{\hat{\sigma}_z}{\hat{\sigma}_w}, \quad \hat{\rho} = r = \frac{\sum_{i=1}^{n}\left(w_i - \overline{w}\right)\cdot\left(z_i - \overline{z}\right)}{\sqrt{\sum_{i=1}^{n}\left(w_i - \overline{w}\right)^2 \sum_{i=1}^{n}\left(z_i - \overline{z}\right)^2}}$$

$$\hat{\sigma}_z = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(z_i - \overline{z}\right)^2}, \quad \overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i,$$

$$\hat{\sigma}_w = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(w_i - \bar{w})^2}, \quad \bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i.$$

As is standard practice, hats over greek symbols are used to denote sample estimators. The correlation estimator $\hat{\rho} = r$ is the common Pearson correlation coefficient estimator (see Barber et al. 2019). We consider the special case in which the explanatory variable w is *time*, which is not a random variable. For this case, with no missing observations, Prosdocimi et al. (2014, Appendix A3) derive expressions for the sample mean and variance of *w=time* given by $\bar{w} = (n+1)/2$ and $\hat{\sigma}_w^2 = n(n+1)/12$ where $w = 1, 2, \ldots, n$. We are interested in estimating the current mean, defined as the conditional mean when $w = n$.

2.4. The fractional sample mean, FM, and sample mean, SM

Suppose a record of streamflow is available over the past n years which is known to exhibit nonstationary behavior due to known deterministic effects, and guidance exists for selecting a subset of the entire record which best represents current watershed conditions. Consider as an alternative to the *RM*, a fractional mean estimator *FM*, which is simply the sample mean of the most recent $n \bullet f$ years of record where f denotes the fraction of the most recent record of n observations which represents current watershed conditions. A fractional mean estimator *FM* based on a sample $z_1, z_2, \ldots z_{n-nf}, z_{n-nf+1}, \ldots, z_n$ is defined as

$$FM = \frac{1}{nf} \sum_{i=1}^{nf} z_{n-nf+i} \tag{5}$$

Note that the sample mean of the entire historical record, $\bar{z}$ defined in (4) and referred to hereafter as *SM*, is a special case of the *FM*, when *f=1*.

We derive an optimal fractional mean estimator *FM**, which employs (5) with an optimal fraction f termed f^*, derived in the next section as that value of f which leads to a minimum *MSE* of the *FM* estimator. For example, suppose a sample of length $n = 50$ is available and one has knowledge that the last $n \bullet f = 30$ years are representative of current watershed conditions. Thus $f = 30/50 = 0.6$, and *FM* in (5) would be simply the sample mean of the last 30 values of z.

2.5. The attained significance level of a regression model

In addition to having solid physical evidence corresponding to the causative mechanisms for a trend in a hydrologic process, investigators often augment such physical knowledge with a hypothesis test associated with the estimated regression slope based on hydrologic measurements. This is in spite of the numerous concerns over null hypothesis significance testing summarized by Vogel et al. (2013), Serinaldi et al. (2018) and others cited therein. Normally a critical significance level α of say 0.05 is assumed, and then an attained significance level p is computed from the data, which indicates the smallest level of significance at which the null hypothesis of no trend ($\beta = 0$) is rejected. We assume that one has good physical evidence of the direction of the trend, if it exists, thus the sign of the alternative hypothesis would be either $\beta < 0$ or $\beta > 0$, in which case, if the attained significance level $p < \alpha$, one should have enough evidence to reject the null hypothesis and thus support the alternative nonstationary analysis. Assuming the regression model residuals are normally distributed, the attained significance level p for such a one-sided test can be computed from (see Helsel et al., 2020)

$$p = 1 - P\left[t_{n-2} \le \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}\right] \tag{6}$$

where $P[]$ denotes the cumulative probability operator and t_{n-2} represents a Students-t variate with n-2 degrees of freedom. We express the attained significance level p in (6) as a function of the true correlation ρ, rather than a sample estimator as is typical in hypothesis testing. Instead, we are only employing this relationship for the purpose of integrating the impact of both sample size and correlation on resulting attained significance levels.

Given concerns raised earlier in Section 1.1, as well as the fact that the hypothesis test in (6) does not account for the inflation in variance of the β estimator due to possible stochastic persistence associated with the model errors ε in (1), we do not advocate the use of this or other hypothesis tests to evaluate whether to perform a SFFA or NFFA. In addition to those concerns, the result of such tests do not provide critical information concerning the robustness or efficiency of the resulting estimators, *RM* and *FM* in (4) and (5), respectively. Instead we employ the attained significance level p in (6) in later comparisons as a metric which measures the impact of both sample size n, and correlation coefficient ρ, thus reducing the dimensionality of our comparative assessments.

2.6. Comparisons among estimators of the condional mean

If an estimator $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$, the estimator $\hat{\theta}_1$ will yield a "better" estimate of the statistic θ than $\hat{\theta}_2$. The notion of a "better" estimator of a statistic relies upon the choice of a loss function corresponding to the problem of interest, where a loss function quantifies the relative economic and other losses associated with estimation errors. Among statisticians, the most common choice of the form of a loss function is quadratic, resulting in the mean squared error (*MSE*) criterion of optimality (see Everitt, 2002, page 128). One can define the efficiency of the estimator $\hat{\theta}_1$ relative to the estimator $\hat{\theta}_2$ using

$$Eff[\hat{\theta}_1, \hat{\theta}_2] = \frac{E\left[\left(\hat{\theta}_2 - \theta\right)^2\right]}{E\left[\left(\hat{\theta}_1 - \theta\right)^2\right]} = \frac{MSE[\hat{\theta}_2]}{MSE[\hat{\theta}_1]} \tag{7}$$

where *MSE* denotes mean square error. In general, when $Eff[\hat{\theta}_1, \hat{\theta}_2] > 1$, the estimator $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ and is thus preferred over the estimator $\hat{\theta}_2$ for estimation of θ

It is well known that the *MSE*, variance and bias of an estimator are related via $MSE[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = (E[\hat{\theta}] - \theta)^2 + E[(\hat{\theta} - E[\hat{\theta}])^2] = Bias[\hat{\theta}]^2 + Var[\hat{\theta}]$, so that an unbiased estimator has *MSE* equal to its variance. Thus if both estimators are unbiased, then the efficiency in (7) reduces to the ratio of the variances of both estimators. To enable comparisons among estimators under both stationary and nonstationary conditions using the concept of efficiency in (7), we derive analytical expressions for the bias, variance and *MSE* of the estimators in the following sections.

2.7. MSE of estimators under stationary conditions

Under stationary conditions, the current mean is equal to μ_z which does not change over time, and both $\rho = 0$ and $\beta = 0$ in Eqs. (1)–(3). Under stationary conditions, *RM, SM* and *FM* are unbiased estimators of the true mean, regardless of the value of f, thus their *MSE* is made up entirely of their variance, so that for any independent and identically distributed series of (potentially transformed) streamflows z:

$$MSE_S[FM] = Var_S[FM] = \frac{\sigma_z^2}{n \cdot f} \tag{8}$$

with the subscript S denoting stationary conditions. Helsel et al. (2020) report the variance of a regression estimator of the mean in the current year n, which can be combined with the mean and standard deviation of w given in (4) to obtain

$$MSE_S[RM] = Var_S[RM] = \sigma_z^2\left(1-\rho^2\right)\left[\frac{1}{n} + \frac{\left(n-\bar{w}\right)^2}{n\hat{\sigma}_w^2}\right] = \sigma_z^2\left[\frac{1}{n} + \frac{\left(n - \frac{n+1}{2}\right)^2}{n\left(\frac{n(n+1)}{12}\right)}\right] \tag{9}$$

since under stationary conditions $\rho = 0$.

2.8. MSE of estimators under nonstationary conditions

We focus on the current mean in year n ($w = n$) which would reflect current conditions at the end of the streamflow record. Under nonstationary conditions, the current mean is equal to the conditional mean of z in year n denoted $\mu_{z|w=n}$, in which case *RM* is unbiased, but *FM* (and thus *SM*) will always be biased, regardless of the value of f (unless $n \bullet f = 1$). One can rewrite the expression for FM in (5) using the regression model for z given in (1) which leads to the bias of *FM* under nonstationary conditions:

$$Bias_{NS}[FM] = E[FM] - \mu_{z|w=n}$$
$$= E\left[\frac{1}{nf}\sum_{i=1}^{nf}\left[\mu_z + \beta\left[(n - nf + i) - \left(\frac{n+1}{2}\right)\right] + \varepsilon_i\right]\right] - \left[\mu_z + \beta\left[n - \left(\frac{n+1}{2}\right)\right]\right]$$
$$= \frac{\beta(1 - nf)}{2} \tag{10}$$

where the subscript NS denotes nonstationary conditions. Eq. (10) results from the fact that $E[\sum_{i=1}^{nf} i/(n \cdot f)] = (1 + nf)/2$ and $E[\varepsilon] = 0$. For a given sample size n, the bias in *FM* increases as both β and f increase. Under nonstationary conditions, each observation z is defined in (1) as $z = \mu_z + \beta(w - \mu_w) + \varepsilon$. Noting that the first two terms are not random variables, thus $Var(z) = Var(\varepsilon)$, so that $Var_{NS}[FM] = Var_S[FM]$ in Eq. (8), so that the *MSE* of the fractional mean is given by the sum of its bias squared and its variance:

$$MSE_{NS}[FM] = \left(\frac{\beta(1 - nf)}{2}\right)^2 + \frac{\sigma_z^2(1 - \rho^2)}{nf} \tag{11a}$$

which can be simplified using the fact that $\beta = \rho\sigma_z/\sigma_w = \rho\sigma_z\sqrt{12/(n(n+1))}$, resulting in

$$MSE_{NS}[FM] = \sigma_z^2\left[\frac{3\rho^2(1 - nf)^2}{n(n+1)} + \frac{1 - \rho^2}{nf}\right] \tag{11b}$$

Under nonstationary conditions *RM* in (4) is unbiased, thus its *MSE* is equal to its variance. Helsel et. al. (2020) report the variance of *RM*, which can be combined with the mean and standard deviation of w given in (4) and taking $w = n$ to obtain

$$MSE_{NS}[RM] = Var_{NS}[RM] = \sigma_z^2(1 - \rho^2)\left[\frac{1}{n} + \frac{(n - \bar{w})^2}{n\hat{\sigma}_w^2}\right]$$
$$= \sigma_z^2(1 - \rho^2)\left[\frac{1}{n} + \frac{\left(n - \frac{n+1}{2}\right)^2}{n\left(\frac{n(n+1)}{12}\right)}\right] \tag{12}$$

2.9. The optimal fractional sample mean, FM*

Here we introduce an optimal fractional mean denoted *FM**, as that value of *FM* which exhibits minimum *MSE* under nonstationary conditions. This is accomplished by deriving the optimal fraction f, denoted f^*, for which $MSE_{NS}[FM]$ is minimized. The function $MSE_{NS}[FM]$ given in (11) is minimized with respect to f by setting its first derivative equal to zero ($\partial MSE_{NS}[FM]/\partial f = 0$) and solving the resulting expression for ρ. The resulting optimal value of f, termed f^*, can be computed from

$$\rho = \sqrt{\frac{n+1}{6(f^*)^2 n(f^* n - 1) + n + 1}} \tag{13}$$

When the optimal fraction f^* is substituted into the estimator *FM* given in (5) we obtain the optimal fractional mean estimator *FM**, given by

$$FM^* = \frac{1}{nf^*}\sum_{i=1}^{nf^*} z_{n-nf^*+i} \tag{14}$$

*FM** has the unique property under nonstationary conditions of yielding an estimator of the sample mean with both minimum *MSE* as well as maximum efficiency relative to the regression estimator of the mean *RM*, as long as the correlation is known. Monte Carlo experiments are performed to evaluate the sampling properties of *FM** when the correlation is unknown.

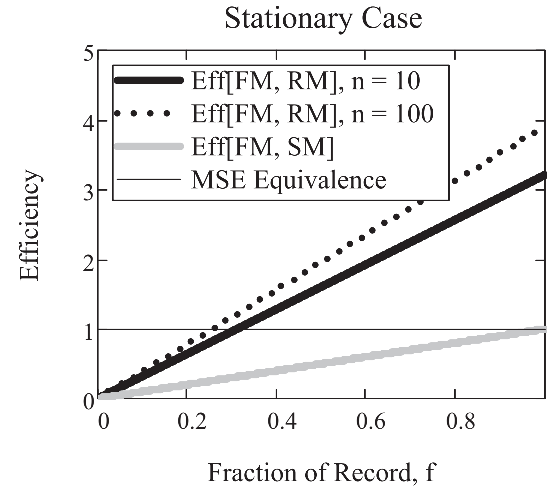

Fig. 1. Efficiency of the fractional mean *FM* relative to the regression mean RM under stationary conditions. Note that *SM* = *FM* when $f = 1$.

3. Results

We use the expressions for *MSE* and efficiency derived in the previous section to evaluate the general conditions under which one can expect *FM** to exhibit greater efficiency than either the sample mean (*SM* = *FM* with $f = 1$) or the regression estimator *RM*, under stationary and nonstationary conditions. All previous derived theoretical expressions for Bias, *MSE* and efficiency of the various estimators are a function of the sample size n and the population correlation ρ. All such analytical theoretical expressions for *RM* were derived assuming that in practice, an estimate of ρ is obtained from the Pearson correlation coefficient estimator r given in (4). Since an estimate of ρ is also needed to implement *FM** and we could not obtain analytical results for its Bias, MSE or efficiency when an estimate of ρ is used, Monte-Carlo experiments are performed to consider the additional uncertainty associated with *FM** resulting from the necessity to estimate ρ.

3.1. Results – stationary conditions

Unlike nonstationary conditions, under stationary conditions a true mean exists and a comparison among the estimators, *SM, FM, FM** and *RM*, is relatively simple because they are all unbiased, so their efficiency defined in (7) reduces to the ratio of their variances. For example, the efficiency of the sample mean *SM* relative to the fractional mean *FM* is equal to $1/f$, which we denote as $Eff[SM, FM] = 1/f$, so that under stationary conditions, SM is always more efficient than FM because if $f < 1$, then *Eff*[*SM, FM*] > 1. The efficiency of *FM* relative to *RM*, under stationary conditions, denoted as *Eff*[*FM, RM*], is obtained by substitution of Eqs. (8) and (9) into (7) which can be simplified to

$$Eff[FM, RM] = \frac{f(4n^2 - 5n + 3)}{n(n+1)} \tag{15}$$

because under stationary conditions, the correlation coefficient $\rho = 0$. Fig. 1 summarizes the efficiency of *FM* relative to *RM* under stationary conditions and also enables us to make a general efficiency comparison between *FM, FM**, *RM* and *SM*. Perhaps the most interesting result in Fig. 1 involves the efficiency of *SM* relative to *RM*, which corresponds to the value of *Eff*[*FM, RM*] when f=1. The range of efficiencies of *SM* over *RM* correspond approximately to $3 < Eff[SM, RM] < 4$ (for $10 \le n \le 100$), so under stationary conditions, *SM* will yield estimators of the mean with 3–4 times lower *MSE* than *RM*, even though both estimators are unbiased under those conditions. This result dramatizes the influence of the added uncertainty associated with *RM* over *SM*, resulting from having to estimate the regression slope β.

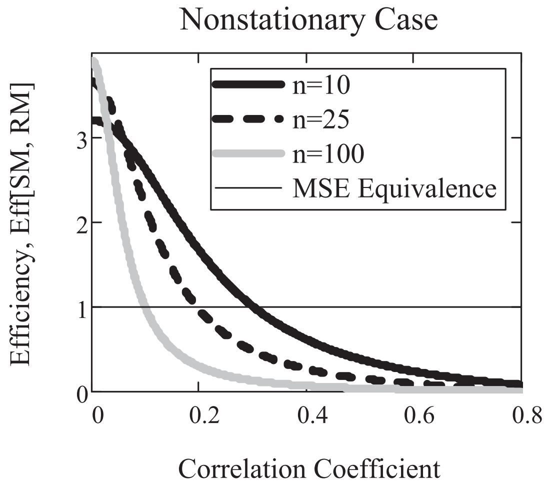

Fig. 2. Efficiency of the sample mean *SM* relative to the regression mean *RM* under nonstationary conditions as a function of sample size n and correlation coefficient ρ

As expected, Fig. 1 also illustrates that under stationary conditions, *SM* is always preferred over *FM* (for $f < 1$), regardless of the value of *n*, thus under those conditions, *SM* is also always preferred over *FM** (for $f < 1$). Fig. 1 also illustrates that *FM* is nearly always preferred over *RM* for values of *f* in excess of approximately 0.3. As is shown later on, the optimal fractional mean *FM** usually employs values of *f* significantly greater than 0.3, thus *FM** will usually be preferred over *RM*, under stationary conditions. These findings are analogous to some early flood frequency studies which documented the general conditions under which an 'at-site' estimator is to be preferred over one based on the use of a regional regression model (Hebson and Cunanne, 1987).

In summary, under stationary conditions and $f < 1$, SM is generally preferred to all other estimators considered, followed by FM (with $f < 1$), which is generally preferred to RM. Note that the gains in efficiency reported in Fig. 1 of *SM* over *FM* and of *SM* over *RM* are often rather considerable, and emphasize the need to employ stationary methods when stationary behavior cannot be rejected on the basis of a deterministic analysis as discussed in Section 1.1. Under nonstationary conditions, this ranking among the estimators is more complex and very different, highlighting the need to consider both stationary and nonstationary conditions in any hydrologic frequency analysis as emphasized by Vogel et al. (2013), Rosner et al. (2014), and Salas et al. (2018).

3.2. *Results - nonstationary conditions -* RM, FM *and* SM

Under nonstationary conditions, a true mean does not exist, and instead, the mean is changing over time. Unlike the stationary case when all the estimators, *SM, FM, FM** and *RM*, were unbiased; under nonstationary conditions, only *RM* is unbiased, and only when the true nonstationary model in (1) is known, as is assumed here. The bias associated with both *FM* and *SM* can be evaluated using (10), noting that when *f=1, FM* reduces to *SM*. In general, the sign of the bias associated with *FM* will depend on the sign of the slope coefficient β. For positive (negative) trends, *E[FM]* will generally be smaller (larger) than its true value, with that bias increasing as both *f* and *n* increase.

We begin by exploring *Eff*[*SM, RM*] as a function of ρ and *n* in Fig. 2 for nonstationary conditions, which is obtained by combining (7), with (11, assuming *f=1*), and (12). Also shown in Fig. 2 are conditions of *MSE* equivalence among the two estimators, which occurs when *Eff*[*SM, RM*] = 1 (i.e. *MSE* of *SM* is equal to the *MSE* of *RM*). Fig. 2 illustrates that for very small values of ρ, *SM* is generally preferred over *RM*. In contrast, we note that for larger values of ρ, *RM* exhibits much lower *MSE* than *SM*. Clearly, the values of *n* and ρ are of considerable importance in understanding the advantages of *RM* over *SM*, and vice versa.

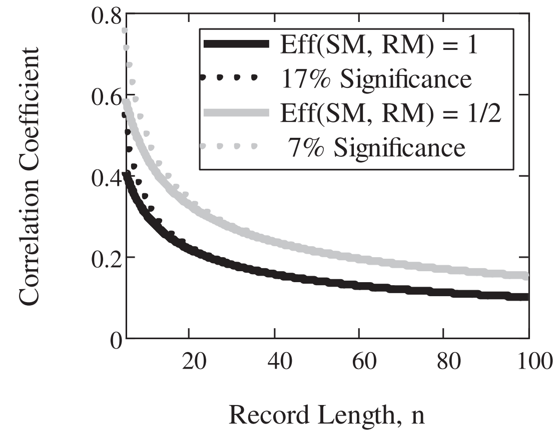

Fig. 3. Relationship between correlation coefficient ρ, record length *n* and attained significance level *p* when the efficiency of *RM* relative to *SM* is equal to 1 and 1/2 under nonstationary conditions.

To understand the behavior of *Eff*[*SM, RM*] more fully, Fig. 3 superimposes the relationship between *n*, ρ, and the attained significance level *p* (given in (6)) over the relationship between *Eff*[*SM, RM*], *n* and ρ given by substitution of (11, with *f*=1) and (12) into (7) for the cases when *Eff*[*SM, RM*] = 1 and when *Eff*[*SM, RM*] = 1/2. Note that the condition *Eff*[*SM, RM*] = 1/2 in Fig. 3 implies that *RM* is preferred over *SM* because under those conditions *RM* has twice the information content (or half the variance) of *SM*. Values of *p* were adjusted, until both of the superimposed relationships in Fig. 3, roughly line up with each other. The important result in Fig. 3 is that under nonstationary conditions, MSE equivalence between RM and SM (*Eff*[*SM, RM*] = 1) occurs when the attained significance level associated with the fitted regression is approximately equal to 17%. This result emphasizes that the estimator RM will have lower MSE than SM when the attained significance level is below about 17%. It is common practice to choose the regression estimator RM over the SM when significance levels are below 5% or 10%. Fig. 3 illustrates that use of a 5% or 10% level test as is common practice, corresponds to conditions under which RM will have considerably more information than SM thus perhaps a higher significance level (roughly 17%) may be a more appropriate significance threshold, in practice.

3.3. *Results – nonstationary conditions – optimal fractional mean, FM**

The critical question addressed in this section is whether or not a sample mean of an optimal subset of the historical record can be used to improve upon the performance of *RM* under nonstationary conditions. We begin by illustrating the peaked nature of the relationship between *Eff*[*FM, RM*] and *f* in Fig. 4(a), which was constructed by combining (7) with (11) and (12) for the cases when $\rho = 0.2$ and $n = 10$, 25 and 100, values close to those typically encountered in practice. We note that for a given information content associated with the historical record, as evidenced by the combined values of *n* and ρ, there always exists a maximum value of *Eff*[*FM, RM*] corresponding to a single optimal value of *f* which we term f^*. For values of *f* either greater or less than f^*, there is a decrease in *Eff*[*FM, RM*]. Fig. 4a shows that for $\rho = 0.2$ and $n = 25$, any *FM* estimator with $0.3 < f < 0.95$ is more efficient than *RM*. The general relationship between the optimal fraction f^* and values of *n* and ρ, given in (13), is illustrated in Fig. 4b.

Fig. 5a displays the efficiency of *FM** relative to *RM* as a function of ρ and *n*, under nonstationary conditions, computed by combining (7) with (11) and (12) with the optimal value f^* computed from (13). Fig. 5a shows that for the cases when $n = 10$ and $n = 100$, correlations in excess of about 0.5 and 0.2 are needed, respectively, for RM to have a

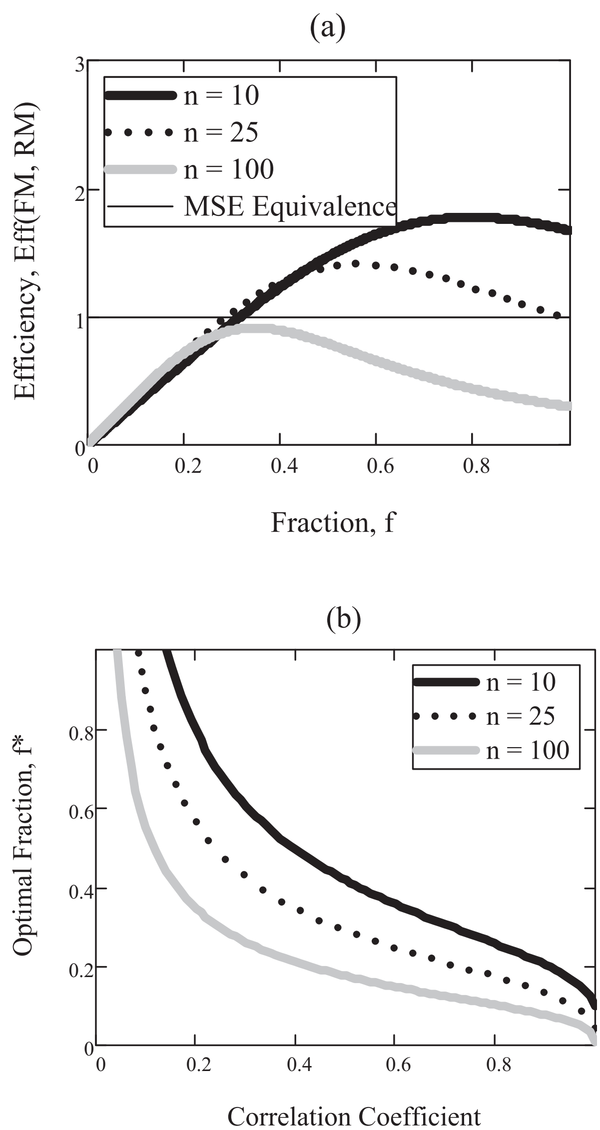

Fig. 4. Under nonstationary conditions: (a) illustration of the peaked shape of *Eff[FM, RM]* as a function of *f* and *n* when ρ=0.2; and (b) illustration of the relationship between the optimal fraction f^* and sample size n and correlation ρ.

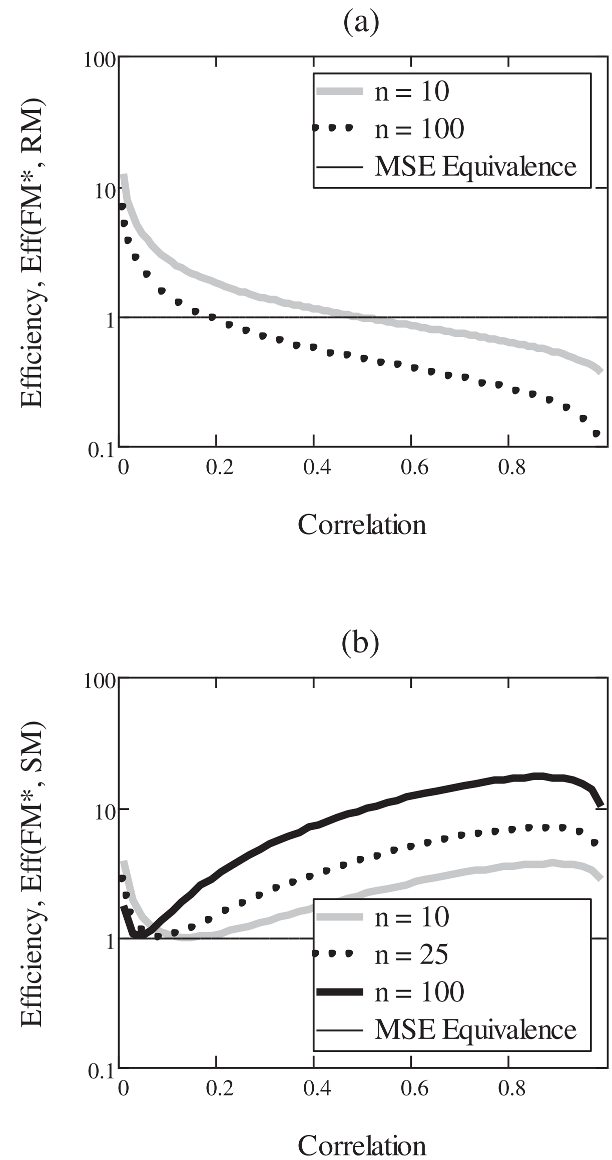

Fig. 5. Efficiency of (a) *FM** relative to *RM* and (b) *FM** relative to *SM* as a function of sample size *n* and correlation ρ under nonstationary conditions.

lower MSE than FM*. This is an important result because correlations of trend models for common hydrologic series typically exhibit relatively low correlations, often in the range in which *FM** is likely to have lower *MSE* than *RM*. Rather than citing examples of typical reported trend correlations, this issue is explored further below, using attained significance levels, which combine the impact of both sample size and the magnitude of the correlation coefficient. In Section 4.0 we report results from a study by Douglas et al. (2000) which documents the range of attained significance levels observed in practice for both flood and low streamflows.

Fig. 5a also indicates that the efficiency of *FM** relative to *RM* becomes extremely large as ρ approaches zero because under those conditions *FM** approaches *SM* (and $f^* = 1$ as shown in Fig. 4b) in which case *SM* has a much lower *MSE* than *RM* as was shown earlier in Fig. 2. Fig. 5a is important, because it shows that even though RM is an unbiased estimator under nonstationary conditions, FM* (and even SM) both have much lower MSE than RM when ρ approaches zero, due to added uncertainty associated with having to estimate the slope coefficient β. We conclude that the values of both ρ and n are central to determining how much better *FM** and *SM* are, compared to *RM* under nonstationary conditions. Fig. 5b illustrates the efficiency of *FM** relative to *SM* as a function of both ρ and *n*, which documents the considerable gains in efficiency of *FM** over *SM* as both n and ρ increase. We conclude from Fig. 5b that FM* is always an improvement over SM under nonstationary conditions.

Perhaps our most important practical result is illustrated in Fig. 6, which is analogous to Fig. 3, because it was constructed by superimposing the relationship between *n, ρ* and *p* (given in (6)), over the relationship between *Eff*[*FM**, *RM*], *n* and ρ given by substitution of (11 with f^* given in 13) and (12) into (7). The important result in Fig. 6 is that MSE equivalence between FM* and RM occurs approximately when the attained significance level associated with the fitted regression is equal to approximately 5%.

Fig. 6 also reports conditions when *Eff*[*FM**, *RM*] = 1/2 which implies that *RM* is preferred over *FM* * because under those conditions *RM* has twice the information content (or half the variance) of *FM**. Interestingly even though both *RM* and *FM** are designed to perform well under nonstationary conditions, *RM* is generally dominated by *FM** unless significance levels are below 0.05. This result emphasizes that for

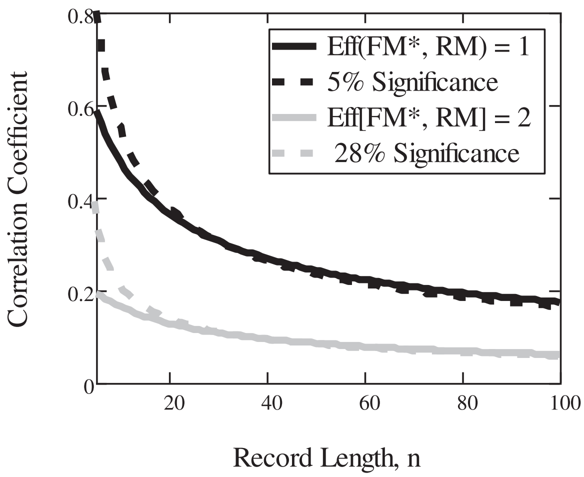

Fig. 6. Relationship between correlation coefficient ρ, record length n and attained significance level p when the efficiency of FM^* relative to *RM*, $Eff[FM^*, RM]$, is equal to 1 and 2 under nonstationary conditions.

RM to have clear advantages over FM^*, would require attained significance levels considerably lower than 5%.

3.4. *Impact of uncertainty in estimation of* ρ *on* $MSE_{NS}[FM^*]$

Until now, we have assumed knowledge of the true correlation ρ. In this section, we consider the ramifications of having to estimate ρ using the common Pearson correlation coefficient estimator, denoted r. Barber et al. (2019) provide a recent review of literature on estimation of ρ for skewed hydrologic processes. Our theoretical derivation of $MSE_{NS}[FM^*]$ (computed by combining (7) with (11) and (12) with the optimal value f^* computed from (13)), ignores uncertainty associated with the estimator of ρ needed to estimate f^* in FM^*. Monte Carlo experiments are performed to evaluate the impact of uncertainty in the ρ estimator, which in turn is needed to compute f^*, FM^* and finally $MSE_{NS}[FM^*]$. Simulation of uncertainty associated with an estimate of ρ (denoted *r*) is implemented using an approximation to what is termed the non-null sampling distribution of *r,* which describes the distribution of *r* when the regression model in (1) governs. Fisher (1915, 1921) first introduced exact expressions for the non-null sampling distribution of *r* for large and small samples. Due to their numerical complexity, numerous approximations have been introduced over the years. We employ Kraemers (1973) approximation who showed that if *r* is an estimator of the correlation coefficient based on a sample of *n* observations drawn from a bivariate normal distribution with correlation coefficient ρ, then the statistic

$$T = \frac{(r-\rho)\sqrt{n-2}}{\sqrt{(1-r^2)(1-\rho^2)}} \sim t_{n-2} \tag{16}$$

is approximately distributed as Students-t with *n*-2 degrees of freedom. Using (16) we generate 10,000 values of r corresponding to a wide range of values of ρ for the two cases $n = 25$ and $n = 100$. For each true value of ρ considered, the 10,000 generated values of *r* are used to estimate 10,000 values of f^*, which are then used to compute 10,000 values of $Eff[FM^*, RM]$. Fig. 7 compares the mean (or steady state) value of $Eff[FM^*, RM]$ from those Monte-Carlo experiments with the theoretical value of $Eff[FM^*, RM]$ which ignores uncertainty in our estimator of f^* (plotted earlier in Fig. 5a). Interestingly, for large samples (i.e. $n = 100$) the (approximate) analytical and (exact) Monte-Carlo results are in very good agreement, especially for values of ρ in excess of about 0.2. For smaller samples, accounting for uncertainty in the estimator of f^* leads to slightly lower efficiencies, $Eff[FM^*, RM]$, than expected on the basis of our analytical results. Additional Monte-Carlo experiments

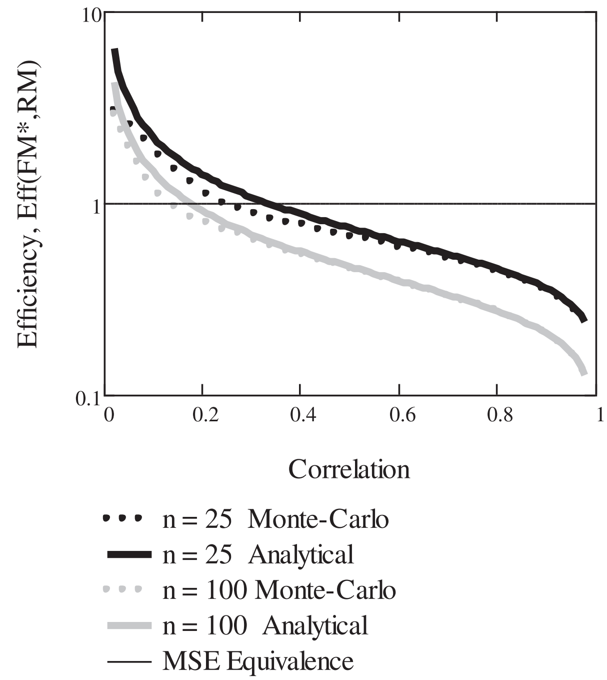

Fig. 7. Efficiency of FM^* relative to *RM*, $Eff[FM^*, RM]$, for analytical and Monte-Carlo results as a function of correlation ρ under nonstationary conditions.

are needed to formulate more general conclusions for a wider range of conditions.

4. Discussion

We have shown that the decision as to which estimator of the conditional mean to use can be expressed as a function of the attained significance level p, which in turn is a complex function of the explanatory power of the trend and sample size as summarized by the values of ρ and *n,* respectively. If the Monte-Carlo and analytical results in Fig. 7 were in exact agreement, our analytical results would yield both exact and general decision rules regarding which estimator of the conditional mean to apply for a given observed value of attained significance level, when the true nonstationary model is known. However, given the results in Fig. 7, the resulting decision rules are only approximate, yet still could be quite useful in practice. Fig. 7 documents that our decision rules are only approximate for small values of n and ρ, yet are nearly exact for larger values. The approximate decision rules concerning the application of the three estimators of the conditional mean, *SM*, FM^*, and *RM,* can be summarized as follows:

1. Under stationary conditions all three estimators are unbiased, but *SM* is always preferred because it exhibits the lowest variance and thus the lowest *MSE*, followed by *FM* (with $f < 1$), which is generally preferred to *RM*.
2. Under nonstationary conditions *RM* is the only unbiased estimator considered, and the attained significance level *p* approximates which estimator is preferred. When $p < 0.05$ *RM* is generally preferred, and when $p > 0.05$ FM^* is generally preferred over *RM*. When $p < 0.17$, *RM* is generally preferred to *SM*.

Our results indicate that the decision as to whether to employ *RM,* FM^* or *SM* under nonstationary conditions depends critically upon the attained significance levels associated with the trend in the range of $p \leq 0.17$. Such values of attained significance levels are commonplace in trend assessments concerning hydrologic series. For example, in a national study of trends in low flow and flood flow discharges at unregulated watersheds across the U.S. over a 50 year period, Douglas et al. (2000, Table 2) found attained significance levels for low flow trends having a mean of 0.132 and a range from 0.007 to 0.429, and for flood flow trends a mean of 0.264 and a range from 0.035 to 0.386.

It is important to note that our results assume full knowledge of the true underlying structure of the nonstationary model. Naturally in practice one would not know a priori the form of the nonstationary trend model nor the form of any stochastic persistence present in the streamflow observations. Such knowledge may be critical in practice, thus future experiments are needed to evaluate the robustness of our findings under (1) different nonstationary model structures, (2) possible heteroscedasticity of the model residuals (see Hecht and Vogel, 2020), and (3) different forms of stochastic persistence.

5. Conclusions

There is a constant evolution of the field of hydrologic frequency analysis (HFA). Serago and Vogel (2018, Section 2) reviewed early developments in HFA in the 1970's and 1980's which led to the widespread appreciation of the superiority of, and need for parsimonious estimators for use in SFFA, because they generally yield more efficient estimators of design event quantiles than more complex models. In spite of the tremendous increase in the literature on NFFA, we could only find a few studies which have documented the value of parsimonious estimators such as the FM^* estimator introduced here. The findings of this study, along with recent work by Serinaldi and Kilsby (2015), Luke et al. (2017), and Yu and Stedinger (2018), provide substantial support for understanding and promoting the value of parsimonious estimators in *NFFA*.

This study is the first effort to develop a general decision rule concerning which estimator of a conditional mean to employ under nonstationary conditions. We have compared the performance of three estimators of the conditional mean, FM^*, *SM* and *RM*. The estimator *SM* is the most parsimonious estimator of the mean possible, because it is nonparametric and requires no parameters. In contrast FM^* and *RM* are both parametric estimators because FM^* requires an estimate of f^*, the optimal fraction of the most recent record to employ, and *RM* requires an estimate of both μ_z and β, regression model parameters. We emphasize that we have only considered relatively parsimonious estimators of a conditional mean, in contrast to the much more complex and much less parsimonious estimators normally recommended (see reviews by Khaliq et al., 2006; Petrow and Merz, 2009; Kiang et al., 2011; Salas et al., 2012; Madsen et al., 2013; Hall et al., 2014; Bayazit, 2015; Salas et al., 2018; and Francois et al., 2019). We outlined numerous advantages of FM^* over both *RM* and *SM*, under nonstationary conditions, and if we had considered less parsimonious models of the type often recommended, we anticipate the advantages of FM^* would appear even greater than those outlined here.

This study takes a different approach then most previous studies which only use the attained significance level p to decide whether or not to perform a nonstationary analysis. The result of a hypothesis test (i.e. attained significance level) does not provide the critical information needed, concerning the robustness or efficiency of the estimators under consideration. Instead, our approach was to relate the attained significance level to the performance of the various estimators in terms of their bias, variance and mean square error (*MSE*). We have shown that the decision whether to employ *RM*, FM^* or *SM* under nonstationary conditions depends critically upon the attained significance levels associated with the trend in the range of $p \leq 0.17$. We have derived analytical expressions for the *MSE* of *SM,* as well as the nonstationary estimators of the conditional mean, FM^* and *RM,* and used those expressions to arrive at the following general conclusions:

- Under stationary conditions a true fixed mean exists and *SM* is always preferred to all other estimators considered, followed by *FM* (with $f < 1$), which is generally preferred to *RM*. The gains in efficiency of *SM* over *FM* and of *SM* over *RM* are often rather considerable, and emphasize the need to employ stationary methods when conditions are known to be stationary as determined from a deterministic analysis (see Section 1.1).
- Under nonstationary conditions no fixed or true mean exists, thus we evaluate estimators of the current or conditional mean at the end of the historical record. Here the ranking among the estimators is more complex and very different than under stationary conditions, highlighting the need to consider both stationary and nonstationary conditions in any hydrologic frequency analysis as emphasized by Vogel et al. (2013), Rosner et al. (2014), and Salas et al. (2018).
- Under nonstationary conditions and very small values of the correlation coefficient, ρ, *SM* is generally preferred over *RM,* whereas in contrast for larger values of ρ, *RM* exhibits much lower *MSE* than *SM*. In addition to apriori having a solid physical basis for performing a nonstationary analysis, the values of sample size, n, and correlation coefficient, ρ, are of considerable importance in deciding which estimator to implement. This is because estimator *RM* will only have a lower *MSE* than *SM* when the attained significance level is below about 17%, thus use of a 5% level test as is often done in practice may be misleading.
- Interestingly, over the range of typical sample sizes $n = 10$ to $n = 100$, correlations in excess of about 0.5 and 0.2 are needed, respectively, for *RM* to have lower *MSE* than FM^*. This is an important result, because correlations of trend models for common hydrologic series exhibit relatively low correlations, often in the range in which FM^* is likely to have lower *MSE* than *RM.*
- Even though *RM* is an unbiased estimator under nonstationary conditions when the true model is known, both *SM* and FM^* have lower *MSE* than *RM* when ρ approaches zero due to added uncertainty associated with having to estimate the slope parameter β. In general, FM^* is always an improvement over *SM* under nonstationary conditions and is generally an improvement over *RM* as well, when the attained significance level is above about 5%.
- The above conclusions contrasting the behavior of *RM* and *SM* are quite rigorous, general and complete, because our analytical derivations consider the sampling variability of the estimators of β. To account for sampling variability in estimators of ρ needed to estimate f^* in FM^*, it was necessary to resort to a Monte Carlo simulation to more fully evaluate $Eff[FM^*, RM]$. Those experiments indicated only slight reductions in $Eff[FM^*, RM]$ compared to our analytical results. Those reductions in efficiency mostly occur for smaller sample sizes and are due to the need to estimate ρ. Additional Monte-Carlo experiments are needed to formulate more general conclusions under a wider range of conditions.
- Our theoretical analysis assumes full knowledge of the structure of the nonstationary model. Monte Carlo robustness experiments are needed to evaluate the performance of the various estimators considered, when the true nonstationary model is unknown.
- Our results are all based on the assumption that the nonstationary regression model in (1) holds with independent, homoscedastic and normally distributed residuals. While this assumption has been shown to be plausible on the basis of continental analyses of hundreds of annual maximum flood series (Vogel et al., 2011; and Prosdocimi et al., 2014), there is also good evidence that such series may also exhibit heteroscedasticity (Hecht and Vogel, 2020) as well as various forms of both short and long term stochastic persistence which can easily be confused or interpreted as deterministic trends (Cohn and Lins, 2005; and Koutsoyiannis and Montanari, 2015). A natural extension to this study would consider the impact of short and long term stochastic persistence within the context of the regression in (1), by using the results from Matalas and Sankarasubramanian (2003) for AR(1), ARMA(1,1) and fractional Gaussian noise stochastic processes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Richard M. Vogel: Conceptualization, Methodology, Formal analysis, Software, Data curation, Writing - original draft, Writing - review & editing, Visualization, Investigation, Validation. **Charles N. Kroll:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Investigation, Validation.

Acknowledgements

The authors are indebted to both Demetris Koutsoyiannis and Francesco Serinaldi, whose comments on an early version of this manuscript led to considerable improvements. The authors are also indebted to the associate editor and three anonymous reviewers for their highly constructive comments which led to additional improvements in our manuscript.

References

Allaire, M.C., Vogel, R.M., Kroll, C.N., 2015. The hydromorphology of an urbanizing watershed using multivariate elasticity. Adv. Water Res. 86, 147–154.

Barber, C., Lamontagne, J., Vogel, R.M., 2019. Improved estimators of correlation and R^2 for skewed hydrologic data. Hydrol. Sci. J.. https://doi.org/10.1080/02626667.2019.1686639.

Bayazit, M., 2015. Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. Environ. Process. 2, 527–542.

Box, G.E., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, Revised edition Holden-Day, San Francisco, Calif..

Blum, A.G., Archfield, S.A., Hirsch, R.M., Vogel, R.M., Kiang, J.E., Dudley, R.W., 2019. Updating estimates of low streamflow statistics to account for possible trends. Hydrol. Sci. J. https://doi.org/10.1080/02626667.2019.1655148.

Brady, A., Faraway, J., Prosdocimi, I., 2019. Attribution of long-term changes in peak river flows in Great Britain. Hydrol. Sci. J. 64 (10), 1159–1170. https://doi.org/10.1080/02626667.2019.1628964.

Cohn, T.A., Lins, H.F., 2005. Nature's style: naturally trendy. Geophys. Res. Lett. 32 (23), L23402. https://doi.org/10.1029/2005GL024476.

Di Baldassarre, G., Laio, F., Montanari, A., 2009. Design flood estimation using model selection criteria. Phys. Chem. Earth 34, 606–611.

Douglas, E.M., Vogel, R.M., Kroll, C.N., 2000. Trends in flood and low flows in the United States. J. Hydrol. 1–2, 90–105 (240).

Ehret, U., Gupta, H.V., Sivapalan, M., Weijs, S., Schymanski, S., Blöschl, G., Gelfan, A., Harman, C., Kleidon, A., Bogaard, T.others, 2014. Advancing catchment hydrology to deal with predictions under change. Hydrol. Earth Syst. Sci. 18, 649–671. https://doi.org/10.5194/hess-18-649-2014.

Everitt, B.S., 2002. The Cambridge Dictionary of Statistics, 2nd ed. Cambridge University Press, New York ISBN 0-521-81099-X.

Fisher, R.A., 1915. Frequency distribution of the values of correlation coefficient in samples from an indefinitely large population. Biometrika 10, 507–521.

Fisher, R.A., 1921. On the 'probable error' of a coefficient of correlation deduced from a small sample. Metron 1, 1–32.

Francois, B., Schlef, K.E., Wi, S., Brown, C.M., 2019. Design considerations for riverine floods in a changing climate – a review. J. Hydrol. 574, 557–573.

Gebert, W.A., Garn, H.S., Rose, W.J., 2016. Changes in streamflow characteristics in Wisconsin as related to precipitation and land use. USGS Sci. Investig. Rep. 2015–5140.

Hall, J., et al., 2014. Understanding flood regime changes in Europe: a state-of-the-art assessment. Hydrol. Earth Syst. Sci. 18, 2735–2772. https://doi.org/10.5194/hess-18-2735-2014.

Hebson, C.S., Cunnane, C., 1987. Assessment of use of at-site and regional flood data for flood frequency estimation. In: Singh, V.P. (Ed.), Hydrologic Frequency Modelling. D. Reidel Norwell, Massachusetts, USA, pp. 433–448.

Hecht, J., Vogel, R.M., 2020. Updating urban design floods for changes in central tendency and variability using regression. Adv. Water Res.. https://doi.org/10.1016/j.advwatres.2019.103484.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2020. Statistical Methods in Water Resources. U.S. Geological Survey Techniques and Methods, p. 458 book 4, chapter A3.

Hirsch, R.M., Ryberg, K., 2012. Has the magnitude of floods across the USA changed with global CO_2 levels. Hydrol. Sci. J. 57, 1–9.

Hirsch, R.M., 2011. A perspective on nonstationarity and water management. J. Am. Water Resour. Assoc. 47 (3), 436–446. https://doi.org/10.1111/j.1752-1688.2011.00539.x.

Khaliq, M., Ouarda, T., Ondo, J.-C., Gachon, P., Bobée, B., 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. J. Hydrol. 329, 534–552.

Kiang, J.E., Olsen, R.J., Reagan, R.M., 2011. Introduction to the featured collection on "nonstationarity, hydrologic frequency analysis, and water management". J. Am. Water Resour. Assoc. 47 (3), 433–435.

Koutsoyiannis, D., Efstratiadis, A., Mamassis, N., Christofides, A., 2008. On the credibility of climate predictions. Hydrol. Sci. J. 53, 671–684.

Koutsoyiannis, D., Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. Hydrol. Sci. J. 60 (7–8), 2–22. https://doi.org/10.1080/02626667.2014.959959.

Koutsoyiannis, D., 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrol. Sci. J. 61 (2), 225–244. https://doi.org/10.1080/02626667.2015.1016950.

Kraemer, H.C., 1973. Improved approximation to the non-null distribution of the correlation coefficient. J. Am. Statist. Assoc. 68 (344), 1004–1008.

Kuczera, G., 1982. Robust flood frequency models. Water Resour. Res. 18 (2), 315–324.

Laio, F., Di Baldassarre, G., Montanari, A., 2009. Model selection techniques for the frequency analysis of hydrological extremes. Water Resour. Res. 45, W07416. https://doi.org/10.1029/2007WR006666.

Luke, A., Vrugt, J.A., AghaKouchak, A., Matthew, R., Sanders, B.F., 2017. Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. Water Resour. Res. 53, 5469–5494. https://doi.org/10.1002/2016WR019676.

Lu, L-H, Stedinger, J.R., 1992. Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. J. Hydrol. 138, 247–267.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., Kjeldsen, T., 2013. A Review Of Applied Methods in Europe for Flood-Frequency Analysis in a Changing Environment. NERC/Centre for Ecology & Hydrology, p. 180.

Matalas, N.C., Sankarasubramanian, A., 2003. Effect of persistence on trend detection via regression. Water Resour. Res. 9 (12), 1342. https://doi.org/10.1029/2003WR002292.

O'Brien, N.L., Burn, D.H., 2014. A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. J. Hydrol. 519B, 2040–2048. https://doi.org/10.1016/j.jhydrol.2014.09.041.

Petrow, T., Merz, B., 2009. Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002. J. Hydrol. 371, 129–141.

Prosdocimi, I., Kjeldsen, T., Svensson, C., 2014. Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. Nat. Hazards Earth Syst. Sci. 14, 1125–1144.

Riggs, H.C., 1972. Low-flow investigations, in: Techniques of Water-Resources Investigations 04-B1. U.S. Government Printing Office, p. 18.

Rosner, A., Vogel, R.M., Kirshen, P.H., 2014. A risk-based approach to flood management decisions in a nonstationary world. Water Resour. Res. 50. https://doi.org/10.1002/2013WR014561.

Salas, J.D., Rajagopalan, B., Saito, L., Brown, C., 2012. Special section on climate change and water resources: climate nonstationarity and water resources management. J. Water Resour. Plann. Manage. 138 (5), 385–388.

Salas, J.D., Obeysekera, J., Vogel, R.M., 2018. Techniques for assessing water infrastructure for nonstationary extreme events: a review. Hydrol. Sci. J.. https://doi.org/10.1080/02626667.2018.1426858.

Serago, J., Vogel, R.M., 2018. Parsimonious nonstationary flood frequency analysis. Adv. Water Res. 112, 1–16.

Serinaldi, F., Kilsby, C.G., 2015. Stationarity is undead: Uncertainty dominates the distribution of extremes. Adv. Water Res. 77, 17–36.

Serinaldi, R., Kilsby, C.G., Lombardo, F., 2018. Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. Adv. Water Res. 111, 132–155.

Slack, J., Wallis, J., Matalas, N., 1975. On the value of information to flood frequency analysis. Water Resour. Res. 11, 629–647.

Vogel, R., Rosner, A., Kirshen, P., 2013. Brief Communication: likelihood of societal preparedness for global change: trend detection. Nat. Hazards Earth Syst. Sci. 13, 1773–1778.

Vogel, R.M., Yaindl, C., Walter, M., 2011. Flood magnification and recurrence reduction factors in the United States. J. Am. Water Resour. Assoc. 47 (3), 464–474. https://doi.org/10.1111/j.1752-1688.2011.00541.x, 2011.

Vogel, R.M., Tsai, Y., Limbrunner, J.F., 1998. The regional persistence and variability of annual streamflow in the United States. Water Resour. Res. 34 (12), 3445–3459.

Yu, X., Stedinger, J.R., 2018. LP3 Flood Frequency Analysis Including Climate Change. World Environmental and Water Resources Congress 2018 , June 3–7, 2018, Minneapolis, Minnesota https://doi.org/10.1061/9780784481400.043 .