Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

Updating estimates of low-streamflow statistics to account for possible trends

Annalise G. Blum, Stacey A. Archfield, Robert M. Hirsch, Richard M. Vogel, Julie E. Kiang & Robert W. Dudley

HYDROLOGICAL SCIENCES JOURNAL
https://doi.org/10.1080/02626667.2019.1655148

Updating estimates of low-streamflow statistics to account for possible trends

Annalise G. Blum[a,b,c], Stacey A. Archfield[b], Robert M. Hirsch[b], Richard M. Vogel[a], Julie E. Kiang[b] and Robert W. Dudley[d]

[a]Civil and Environmental Engineering, Tufts University, Medford, Massachusetts, USA; [b]Water Mission Area, US Geological Survey, Reston, Virginia, USA; [c]Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland, USA; [d]New England Water Science Center, US Geological Survey, Augusta, Maine, USA

ABSTRACT
Accurate estimators of streamflow statistics are critical to the design, planning, and management of water resources. Given increasing evidence of trends in low-streamflow, new approaches to estimating low-streamflow statistics are needed. Here we investigate simple approaches to select a recent subset of the low-flow record to update the commonly used statistic of 7Q10, the annual minimum 7-day streamflow exceeded in 9 out of 10 years on average. Informed by low-streamflow records at 174 US Geological Survey streamgages, Monte Carlo simulation experiments evaluate competing approaches. We find that a strategy which estimates 7Q10 using the most recent 30 years of record when a trend is detected, reduces error and bias in 7Q10 estimators compared to use of the full record. This simple rule-based approach has potential as the basis for a framework for updating frequency-based statistics in the context of possible trends.

ARTICLE HISTORY
Received 24 August 2018
Accepted 16 July 2019

EDITOR
A. Castellarin

ASSOCIATE EDITOR
E. Volpi

KEYWORDS
low-streamflow; trends; Monte Carlo experiments; statistical hydrology; frequency statistics

1 Introduction

Ensuring sufficient streamflow during the driest part of each year is critical for maintaining water quality, energy production, and habitat, as well as for municipal, industrial, and agricultural water supply (Smakhtin 2001). Across the USA and around the world, many streams are experiencing wetting or drying trends in low-streamflow (Lins and Slack 1999, Douglas *et al.* 2000, Smakhtin 2001, Stahl *et al.* 2010, Du *et al.* 2015, Kam and Sheffield 2016, Kormos *et al.* 2016). Accounting for these trends in the estimation of low-streamflow frequency statistics is essential for appropriate design, planning, and management of water resources; however, few promising approaches to do so have been developed. For example, a commonly used low-streamflow statistic is 7Q10, the annual minimum 7-day streamflow which is exceeded in 9 out of 10 years on average. Estimates of 7Q10 are needed in water quality management and water supply planning, as well as for a wide range of activities relating to the determination of minimum downstream release requirements from hydropower, irrigation, water supply, cooling plant, recreation and other facilities. Water quality management applications of 7Q10 include the determination of wasteload allocations, discharge permits, and the siting of treatment plants and sanitary landfills.

Commonly, when a suitable streamflow record is available, estimation of flow frequency statistics (for low- and high-streamflow) consists of four steps: (1) selection of data; (2) selection of a probability distribution function; (3) estimation of parameters of the distribution; and (4) calculation of the desired quantile using estimated model parameters. The entire period of record available is generally selected in Step 1, based on the assumption that the frequency at which a particular statistic occurred in the past is representative of how often it will occur in the future (stationarity assumption). Under stationary conditions, use of the whole record is expected to yield the most accurate estimator; however, under nonstationary conditions, estimators which fail to account for changing conditions are likely to yield inaccurate and biased results (Yu 2017).

Given the growing evidence of trends in low streamflow in the USA, the assumption of stationarity may not always be appropriate when estimating low-flow statistics. For much of the USA, particularly in the east, recent increases in low flows (wetter conditions) have been documented (Lins and Slack 1999, Douglas *et al.* 2000, McCabe and Wolock 2002), which one study partially attributed to increases in fall precipitation (Small *et al.* 2006). Decreasing low streamflow (drier conditions) have been found in the Pacific northwest, northern California, and parts of the southeast (Lins and Slack 1999, Sawaske and Freyberg 2014). For the Pacific northwest, Kormos *et al.* (2016) linked these trends to changes in precipitation, whereas, in northern California, Asarian and Walker (2016) found that changes were likely due to human impacts or vegetation changes. In the southeast, Sadri *et al.* (2016) speculated that the cause of reduced low streamflow was the pumping of groundwater.

Although accounting for trends in flood frequency is an area of active research, relatively few studies have focused on low streamflow. For floods, a number of review papers have summarized proposed nonstationary approaches to frequency analysis, which generally includes fitting a probability density function with parameters dependent upon either time or

CONTACT Annalise G. Blum annaliseblum@gmail.com Department of Earth and Planetary Sciences, Johns Hopkins, 301 Olin Hall, 3400 N. Charles Street, Baltimore, MD 21218, USA

physical drivers of nonstationarity (Khaliq *et al.* 2006, Salas *et al.* 2012, 2018, Hall *et al.* 2014, Bayazit 2015). In contrast, the literature on frequency approaches for low flows under nonstationary conditions is more sparse. Much of the literature focuses on drought indices, proposing nonstationary approaches to modeling drought length (Tu *et al.* 2016, Cancelliere 2017), precipitation series (García Galiano *et al.* 2011, Giraldo Osorio and García Galiano 2012, Wang *et al.* 2015), or soil moisture (Burke *et al.* 2010).

A few studies have focused specifically on modeling nonstationarity in annual minimum streamflow or the impact of trends in flows on the estimation of 7Q10. Copulas with time-dependent parameters have been used to model low-flow series in the Connecticut River Basin, USA (Ahn and Palmer 2016) and on the Hanjiang River in China (Jiang *et al.* 2014). Compared to a stationary analysis, Liu *et al.* (2015) found that a nonstationary climate-informed model provided a better fit to streamflow observations from a gaging station downstream of the Three Gorges Dam in China. For two rivers in the northeastern USA, Steinschneider and Brown (2012) showed how a Bayesian approach informed by prior information on regional sea-surface temperature was preferable to the use of an uninformative prior for forecasting 7Q10. Often multiple anthropogenic impacts occur simultaneously, hindering attribution of changes in low streamflow to specific factors (Hirsch 2011, Allaire *et al.* 2015). When predicting the nonstationary distribution of low-flow series for the Wei River in China, Du *et al.* (2015) identified irrigation and urbanization as important to trends in low streamflow. However, the authors deemed these factors too uncertain and difficult to identify to include in their modeling. In fact, the overall impact of urbanization on low streamflow is uncertain. Some factors associated with urbanization are expected to cause increases, such as decreased evapotranspiration from loss of vegetation, wastewater and stormwater return flows and water-supply and stormwater leakage, while other factors are likely to cause decreases, such as increased surface runoff due to impervious area and losses due to groundwater pumping (Price 2011, Allaire *et al.* 2015).

Some authors have suggested a simple approach of estimating low-flow statistics under nonstationary conditions using a recent subset of the flow record (Riggs 1972, Gebert *et al.* 2016). While return periods of interest are often longer than the period of record for floods, for low flows, we often rely on shorter return periods (i.e. 10 years for 7Q10). The purpose of this study is to evaluate methods of updating 7Q10 estimates to reflect possible trends in the historical data. A simple modification to current practice for estimation of streamflow statistics, this approach does not rely on access to additional sources of data or attribution of changes. Use of a recent subset of the historical record, which does not require extrapolation of historical trends or claim to know future streamflows, is attractive given concerns over assumptions made about future nonstationary conditions (Montanari and Koutsoyiannis 2014, Serinaldi *et al.* 2018). Instead, the approach relies only on the assumption that the recent period of record more accurately reflects current conditions at a streamgage compared to the longer period of record which may be available. However, guidance on how to select an appropriate subset length is lacking. If a relatively long streamflow record is available and there are reasons to suspect trends in flows, how should a practitioner select a subset of that record to estimate 7Q10?

The goal of this study is to investigate and provide guidance on approaches to select a subset of a long low streamflow record when there is a good reason to suspect changes in the flow regime. We explore a variety of approaches, including selecting a recent subset from every flow record, as well as "adaptive" approaches in which we only select a more recent subset of the flow record at sites where statistically significant trends are detected. Our aim is to identify subset approaches for estimation of 7Q10 which (1) improve accuracy and (2) reduce bias when trends are present, as well as (3) maintain adequate performance when there is no trend detected. We use Monte Carlo simulation experiments in which the underlying behavior of the annual minimum flow series is known *a priori* and, therefore, the true value of 7Q10 is known. In contrast, the true value of 7Q10 is unknown for empirical data because only a limited sample of the low-flow record has been observed. To the extent possible, we design the experiment to mimic the real world: each simulation is based on the estimated trend at one of the 174 US Geological Survey (USGS) streamgages. In the following section, we explain the approaches to data selection and the design of the Monte Carlo experiments. We then present our results and conclude with a discussion of these results and directions for future work.

2 Methods

2.1 Streamgages used to inform experiment

To generate plausible low-streamflow series for the Monte Carlo experiments, we needed to identify a range of credible magnitudes for trends in low streamflow. To inform the simulations, we considered the characteristics of 174 USGS streamgages located in the Chesapeake Bay watershed, an ecologically and economically important region of the USA with water quality challenges (Fig. 1(a)). These streamgages were selected for having relatively complete and long records (56–75 years) and representing a diversity of low-flow characteristics and geology. Streamgages with any days of zero streamflow in the record were excluded from this analysis as study of ephemeral streams was beyond the scope of the experiment. We did not exclude any basins based on particular human interference, such as regulation or land-use change, as we wanted to include basins representing a wide range of upstream anthropogenic impacts. Three-quarters of these stream basins are classified as "non-reference" for having substantial human interference, whereas the remaining "reference" basins are minimally disturbed by humans (Falcone 2011). The drainage areas of the basins upstream of these streamgages range from 8 to 47 364 km^2. We obtained mean daily streamflow at these streamgages from the USGS National Water Information System (USGS 2017) and estimated rolling 7-day average streamflows. Annual minimum 7-day streamflow values (7Q) were determined based on climate

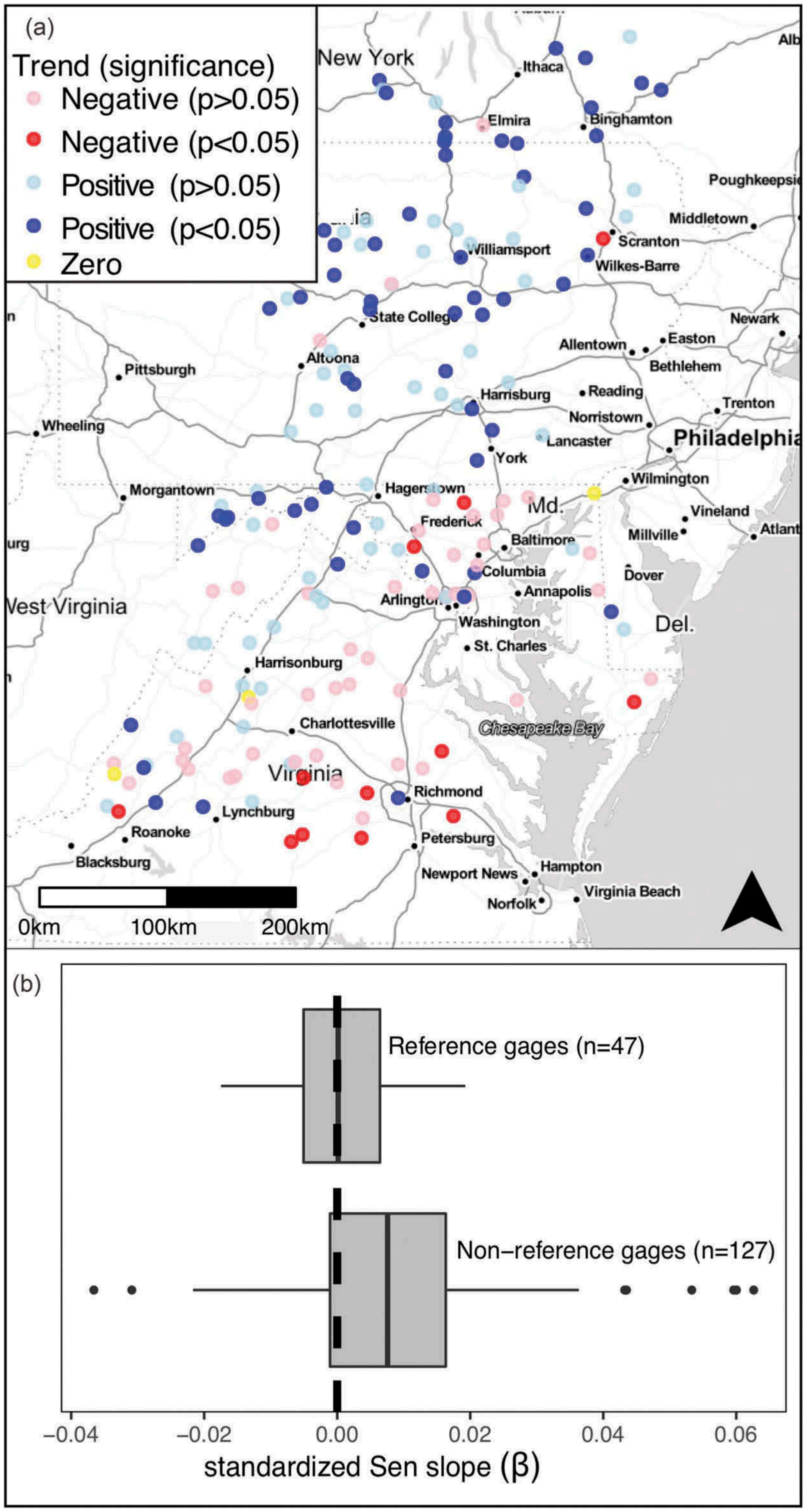

Figure 1. (a) Location of 174 USGS streamgages used to inform Monte Carlo experiment. The color of each marker indicates the color of the sign and statistical significance of a trend in annual 7-day minimum streamflows for the available record (ranging from 55–70 years). (b) Boxplots (25th–75th percentiles, with whiskers to 1.5 times this interquartile range) illustrating the range of standardized Sen slopes for reference and non-reference streamgages. Note that there are different sample sizes for each box.

years (1 April–31 March) to minimize the chance of beginning the year during a possible low-flow period. At these streamgages, 7*Q* ranged across many orders of magnitude, from 0.0002 to 252 m^3/s. We assumed that synthetic 7*Q* were independent between years as there appeared to be a weak serial correlation (or lag one autocorrelation) exhibited in 7*Q* at only 13% (22/174) of study sites, but future work should investigate the implications of this assumption.

To detect trends, we used the nonparametric Mann-Kendall test for monotonic trends under the assumption of independence (Helsel and Hirsch 2002). Nonparametric approaches are often more powerful than parametric ones when the true distribution and/or trend model form is unknown, as is the case with observed data (Helsel and Hirsch 2002). The Mann-Kendall test identified statistically significant trends ($p < 0.05$) at 40% (69/174) of the streamgages. To characterize the magnitude of the trend at each site, we estimated the Sen slope, defined as the median slope of all slopes generated by joining every pair of points (Sen 1968). Of those sites with statistically significant trends, 83% (57/69) had positive Sen slopes, indicating trends towards increasing 7*Q* (wetter streamflow conditions). The range of estimated nonparametric Sen slopes was found to be very similar to the range of estimated parametric slope coefficients based on an ordinary least squares linear regression between the natural logarithm of 7*Q* and year.

To compare results across gages, we focus on nonparametric Sen slopes standardized by the standard deviation of the residual error, defined as the difference between observations and the Sen slope line. This is analogous to a standardized residual in parametric analysis, in which a residual is divided by its standard error (Helsel and Hirsch 2002). Representing a "signal-to-noise ratio", this strategy also reduced dimensionality of the experiment by combining two variables into one, which enabled the most straightforward design and clear presentation of findings. Standardized Sen slopes (also referred to as "standardized trends") estimated for each of the streamgages, along with statistical significance of trends, are illustrated in Fig. 1(a), while boxplots showing the range of standardized trends across reference and non-reference streamgages are given in Fig. 1(b). Compared to the reference streamgages, the range of standardized Sen slopes at the non-reference streamgages is larger and shifted toward a higher frequency of positive trends. It is important to note; however, that there are more non-reference streamgages than reference streamgages.

2.2 Subset approaches to estimate 7Q10

Time series of 7*Q* at three example USGS sites are plotted in Fig. 2. These three sites illustrate examples showing no apparent trend (Fig. 2(a)), a wetting trend (Fig. 2(b)), and a drying trend (Fig. 2(c)). Estimators of 7*Q*10 based on the entire full available (black line with 90% confidence intervals shown as grey lines) and most recent 30 years (dashed green line) are also shown. Note that for the site without a trend (Fig. 2(a)), the two estimators are almost identical, with the 7*Q*10 estimator based on the most recent 30 years falling within the 90% confidence intervals. For the site with a wetting trend (Fig. 2(b)), 7*Q*10 estimated using the last 30 years of the record better reflects 7*Q* flow conditions in recent years compared to 7*Q*10 estimated using the full record and is not contained within the 90% confidence interval. For Fig. 2 (c), the estimator of 7*Q*10 using the last 30 years is lower than the estimator based on the full record, but within the confidence interval.

In addition to evaluating 7*Q*10 estimated using the most recent 30 years of flow, we also evaluate use of the most recent 10 and 50 years. We term this type of approach "non-adaptive" because a fixed subset of the flow record is selected from every flow record. We compare this "non-adaptive" approach to an "adaptive" approach, in which a subset of the record is selected only when a statistically significant trend is detected; otherwise, the entire record available is used for estimation. For the adaptive approach, we compare three commonly used levels of statistical significance (0.01, 0.05, 0.1) to determine how and if statistical significance level impacts accuracy of 7*Q*10 estimators.

2.3 Experimental design

To generate plausible synthetic low-flow records for the experiment, we identified the simplest probability distribution which could approximate the probability distribution of 7*Q*. The World Meteorological Organization manual on low-flow estimation and prediction recommends using the Weibull distribution for 7*Q*10 estimation (WMO 2008). In the USA, the log-Pearson type 3 (LP3) is widely used to describe the probability distribution of 7*Q*. Based on over 1200 US streamgages, Kroll and Vogel (2002) identified the three-parameter lognormal distribution (LN3) as providing the best fit to perennial streams, as studied here. Because the two-parameter lognormal distribution (LN2) is a special case of both LN3 and LP3, we expected this parsimonious model to provide a good approximation of the distribution of 7*Q*. The LN2 model has previously been used to describe the

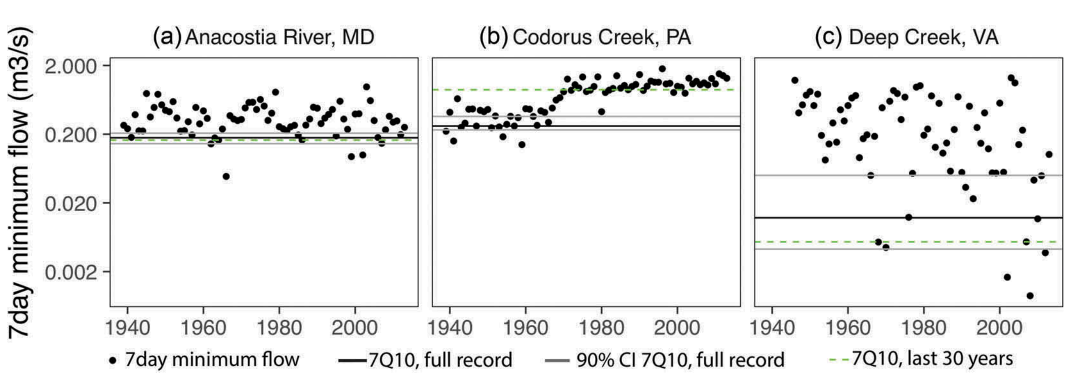

Figure 2. Time series of annual 7-day minimum streamflow and estimators of 7*Q*10 at three USGS streamgages. Confidence intervals around the full record estimator of 7*Q*10 were calculated using a bootstrap approach, as described in Ames (2006). (a) USGS gage 01649500 with standardized Sen slope 0.005 ($p = 0.4$); (b) USGS gage 01574500 with standardized Sen slope 0.06 ($p < 0.01$) and (c) USGS gage 02041000 with standardized Sen slope –0.02 ($p < 0.01$). Note the *y*-axis is on a natural logarithm scale.

distribution of low streamflow in Belgium (Grandry *et al.* 2013), China (Jiang *et al.* 2014), Iran (Modarres 2008), as well as in the USA in Massachusetts (Vogel and Kroll 1989, 1990), Vermont, and New Hampshire (Dingman and Lawlor 1995). For the streamgages used to inform the experiments, L-moment ratio diagrams confirmed that LN2 approximated the probability distribution of 7*Q* at the gaged sites (results not shown), which was also found in Blum (2017).

Based on the assumption of an LN2 distribution, the Monte Carlo experiment was carried out as follows:

Step 1. Calculate true 7Q10

For each standardized trend slope estimated from one of the 174 USGS streamgages, "true" 7*Q*10 values were simulated using a nonstationary LN2 quantile function (Vogel *et al.* 2011). The true value of 7Q10 in the last year of each record, assuming a log-linear trend in 7*Q*, is denoted $7Q10_{\text{true}}$ and was calculated:

$$7Q10_{\text{true}} = \exp(\mu_y + \beta(n - \bar{n}) + z_{0.1}\sqrt{\sigma_y^2 - \beta^2 s_n^2} \quad (1)$$

where μ_y is the mean and σ_y the standard deviation of y (with $y = \ln(7Q)$); $z_{0.1}$ is a standard normal variable with non-exceedance probability of 0.1; β is the magnitude of the standardized Sen slope; n is the record length; $\bar{n}$ is the mean year; and s_n^2 is the variance of record length, which has been derived for a non-random time variable so that $s_n^2 = \frac{n(n+1)}{12}$ (see Prosdocimi *et al.* 2014, Appendix A3). We assumed a record length n of 75 years, as this was both the median and mode of the record length of the USGS streamgages used to inform the experiment. We also assumed that the standardized Sen slope, β, estimated for each streamgage was the true slope. As an assumption of the true slope had to be made, we felt that this was a reasonable assumption for the purposes of this experiment; future work should explore different types of trends including abrupt, reversing, and multiple changes. Finally, we also set the mean of the flows in natural log space, μ_y, to zero and the standard deviation of y, σ_y, to one, which do not affect our findings. The standardized Sen slopes reflect a range of coefficients of variation of 7*Q* (ratio of standard deviation to mean) based on the experimental streamgages.

The nonstationary LN2 quantile function in Equation (1) assumes that an exponential trend (or, equivalently, a log-linear) model of 7*Q* is representative. This simple exponential trend model was found to fit the observed series of 7*Q* for this region relatively well, resulting in approximately normally distributed and constant-variance residuals for over two-thirds of the 174 gaged sites (p = 0.05). When the trend (here, standardized Sen slope) is zero, this model simplifies to the stationary quantile function for an LN2 variable:

$$7Q10_{\text{true}}(\beta = 0) = \exp(\mu_y + z_{0.1}\sigma_y) \quad (2)$$

Step 2. Simulate records and estimate 7Q10 using subset and full record

We generated 10 000 synthetic series of length 75 years for each of our 174 standardized trends, β, using the equation:

$$y_i = \mu_y + \beta(n_i - \bar{n}) + \varepsilon_i \quad (3)$$

where y_i = ln(7*Q*) for year i; n_i is the year i from 1 to 75; $\bar{n}$ is again the mean year (38 years); and ε_i is the residual error in year i where $\varepsilon_i \sim N(0, 1)$.

Given that the true distribution of annual low flows would be unknown in practice, we used a nonparametric estimator to estimate 7*Q*10 from each of the synthetic flow records. In addition, no extrapolation is necessary for records of at least 10 years when estimating a non-exceedance probability of 0.1 (and our shortest subset is 10 years). (In contrast, for flood frequency analysis, parametric frequency models are employed because usually the return period exceeds the available record length.)

The simplest way to estimate 7*Q*10 would be to select the 10th percentile flow from the empirical cumulative distribution of the time series of annual 7-day minimum streamflow, 7*Q*. For example, for a 10-year record, one could simply select the lowest 7*Q*. However, individual order statistics, particularly the lowest order statistic, can be highly variable, making this approach inefficient (Vogel and Fennessey 1994). We used a weighted average of adjacent order statistics, which increases the efficiency of nonparametric quantile estimators, particularly for small samples (Vogel and Fennessey 1994). One simple and attractive nonparametric quantile estimator based on the Weibull plotting position (Vogel and Fennessey 1994) is used here to estimate 7*Q*10 from the selected flow record:

$$7Q10_{\text{est}} = (1 - \theta)q_i + \theta q_{i+1} \quad (4)$$

where q_i is the annual minimum 7-day streamflow flow with ranking i based on sorting the entire record from smallest (i = 1) to largest (i = record length). In addition, $i = \text{floor}[(n + 1)p]$, where floor indicates rounding down to the next integer value, and $\theta = ((n + 1)p - i)$ represents the fraction that is rounded down in the calculation of i. Here, n is the record length (10, 30 or 50 years for the subset approaches, $7Q10_{\text{est\_sub}}$, and 75 years for the full record, $7Q10_{\text{est\_full}}$) and p is 0.1 because 7*Q*10 is the minimum streamflow which is exceeded in 9 out of 10 years, on average. For example, for estimators based on selecting a subset of 10 years of record, a weighted average of the first (q_1) and second (q_2) ranked annual minimum streamflows 7*Q* were used to estimate $7Q10_{\text{est\_sub10}}$:

$$7Q10_{\text{est\_sub10}} = 0.9q_1 + 0.1q_2 \quad (5)$$

Similarly, for estimators based on the full period of record, where n = 75, a weighted average of the seventh (q_7) and eighth (q_8) ranked annual minimum streamflows 7*Q* were used to estimate $7Q10_{\text{est\_full}}$.

Step 3. Calculate accuracy and bias of estimators relative to true 7Q10

We compared each subset approach ($7Q10_{\text{est\_sub}}$) to common practice of using the full available record ($7Q10_{\text{est\_full}}$) using what we term an "improvement factor", which is defined as the ratio of root mean squared error (RMSE) for $7Q10_{\text{est\_full}}$ to $7Q10_{\text{est\_sub}}$:

$$\text{improvement factor} = \frac{\sqrt{\sum_{i=1}^{17400}\left(7Q10_{\text{est\_full},i} - 7Q10_{\text{true},i}\right)^{2}}}{\sqrt{\sum_{i=1}^{17400}\left(7Q10_{\text{est\_sub},i} - 7Q10_{\text{true},i}\right)^{2}}} \quad (6)$$

Because we generated 10 000 simulations for each of 174 trend magnitudes, a total of 1 740 000 records were simulated. These experiments were divided equally into 100 bins for plotting, so each average in calculating RMSE was taken across 17 400 experiments (thus the summations in Equation (6) taken across 17 400 values of i). The RMSE represents the overall accuracy of the estimation method relative to the true value of 7Q10 ($7Q10_{\text{true}}$) and is composed of both bias and variance. An improvement factor of 2 thus reflects that the RMSE of the 7Q10 estimator using the full record is twice as large as the RMSE of the estimator based on the given subset approach. We also estimated mean bias for each of the estimators (subset and full):

$$\text{bias} = \frac{1}{17400}\sum_{i=1}^{17400}\left(7Q10_{\text{est},i} - 7Q10_{\text{true},i}\right) \quad (7)$$

To mimic real-world conditions, we attempted to use only information that would be available to a practitioner. We plotted both improvement factor and bias against standardized trend magnitude (average for each of 100 bins) representing standardized Sen slopes estimated from each simulated record.

In summary, the experimental design involved the following steps:

(1) Using the estimated standardized trend in low flow at the 174 USGS streamgages, calculate values of $7Q10_{\text{true}}$ using Equation (1)
(2) Simulate records (Equation (3)) and estimate $7Q10_{\text{est}}$ (Equation (4)) using full record and subset approaches for each standardized trend value:
- three non-adaptive: last 50 years, last 30 years, last 10 years; and
- three adaptive: only select a subset if a trend is detected based on the Mann-Kendall test at $p< 0.01$, 0.05, or 0.1

(3) Calculate accuracy and bias of estimators relative to true 7Q10 with Equations (6)–(7).

3 Results

First, we focus on the three non-adaptive approaches applied to our simulated records: subset the most recent 10, 30, or 50 years of the 75 years of each synthetic record (referred to as the Subset 10, 30, and 50 strategies; Fig. 3). The roughly parabolic shape of the points centered at zero illustrates how the largest increases in accuracy (largest improvement factors) are associated with the largest magnitude standardized trends. An improvement factor of 1 represents that subset estimators of 7Q10 have the same RMSE as estimators using the full record, >1 indicates improved accuracy (and reduced RMSE) associated with the subset approach, and <1 indicates reduced accuracy (increased RMSE). For Figs. 3–5, each point illustrates the mean value of 17 400 simulated records which fall into that bin. These points are more diffuse at larger trend magnitudes because each point represents the average of a wider range of points. Additionally, the range of x-values (standardized trends) is not symmetric around zero because the simulated trend magnitudes reflect the range of standardized trends from the streamgage sites, which were not symmetrically distributed.

As expected, for the non-adaptive "no-trend" scenarios (Fig. 3), use of the full record provides more accurate estimators of 7Q10 relative to the subset approaches. However, for the Subset 30 and 50 strategies, this reduced accuracy is relatively small compared to the large improvements in accuracy in the presence of trends. For the most extreme trends, Subset 10 shows the largest improvements in accuracy; however, this approach shows reduced accuracy relative to the full record for most of the standardized trends studied. Subset 30 improvement factors are mostly above one, except for no-trend and for very small trend scenarios. Additionally, Subset 30 improvement factors are generally higher than the Subset 10 (except for very extreme trends) or the Subset 50 (except for very small trends) strategies. Given our goal to improve accuracy and reduce bias of 7Q10 estimators in the face of trends while maintaining performance in the absence of trends, we find that the Subset 30 strategy appears to be a reasonable approach.

Turning to the adaptive approaches, we apply an adaptive Subset 30 approach to each simulated record: the full 75-year record is used to estimate 7Q10 unless a trend is detected by the Mann-Kendall test (for p values of 0.01, 0.05, and 0.1), in which case 7Q10 is estimated using the last 30 years of the record. For comparison, Fig. 4 includes the non-adaptive Subset 30 as well as the adaptive Subset 30 approaches.

For no-trend and small-trend scenarios, the adaptive approaches have an improvement factor of one. This is because the full record was used when no statistically significant trend was detected. As a result, these adaptive approaches eliminate the loss in accuracy for no trend scenarios associated with the non-adaptive Subset 30 strategy.

In terms of statistical significance, we find very little difference between the three significance levels considered here (Fig. 4). For the case of $p < 0.1$ (which is inclusive of the most trends), the points rise most steeply, which suggests that this approach is marginally more accurate than the other levels of statistical significance, but this difference is small.

Finally, Fig. 5 illustrates bias in 7Q10 estimators associated with all of the estimation approaches (full record, as well as non-adaptive and adaptive). We plot mean bias for each approach directly, rather than a ratio relative to the full record (as done in Figs. 3 and 4) because bias takes on both positive and negative values. Not surprisingly, we find positive bias for negative trends and negative bias for positive trends. Overall, we find that the largest magnitude biases associated with use of the full record, as these points appear farthest from zero. Other than the Subset 10 strategy, all approaches show zero mean bias for no trend scenarios (recall that each point represents the average of 17 400 simulations). Generally, the Subset 30 strategy, and the three adaptive Subset 30 strategies appear to have the smallest magnitude

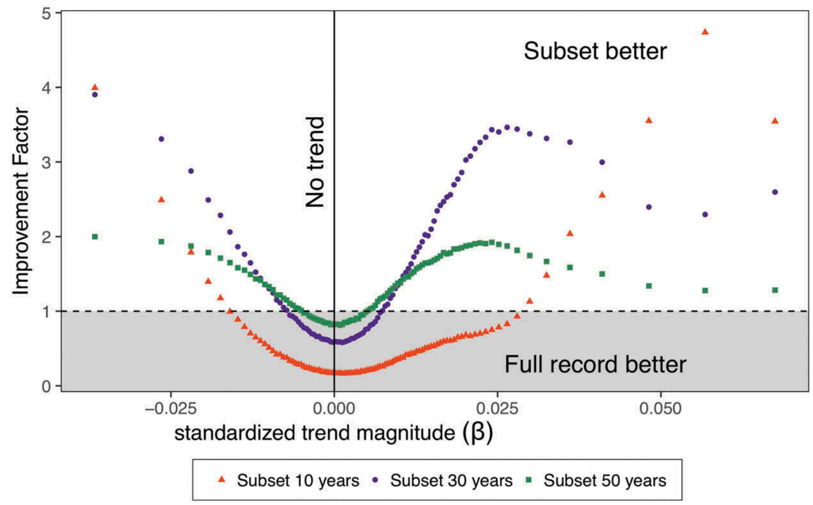

Figure 3. Accuracy of 7*Q*10 estimators based on Subset 10, 30, and 50 strategies, relative to 7*Q*10 estimators calculated using a full 75-year record. Improvement factor is defined as the ratio of RMSE for 7*Q*10 estimated using the full 75-year record relative to RMSE for 7*Q*10 estimated using the specified Subset. Standardized trend magnitude refers to the nonparametric Sen slope standardized by residual errors relative to the Sen slope line.

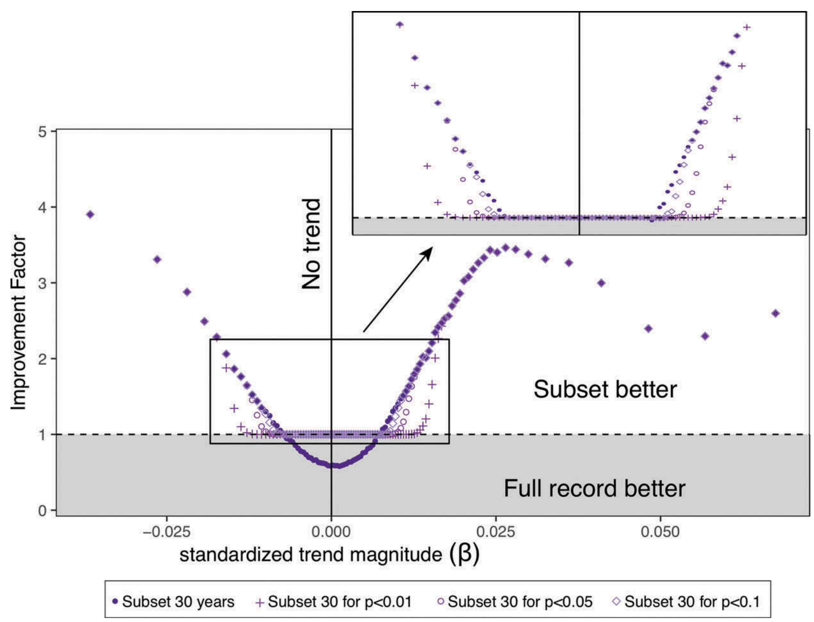

Figure 4. Accuracy of 7*Q*10 estimators based on adaptive approaches, where the Subset 30 strategy is only applied when a statistically significant trend is detected at the $p < 0.01$, 0.05, and 0.1 significance levels (identical to Fig. 3), relative to 7*Q*10 estimators calculated using a full 75-year record. Improvement factor is defined as the ratio of RMSE for 7*Q*10 estimated using the full 75-year record relative to RMSE for 7*Q*10 estimated using the specified Subset. Standardized trend magnitude refers to the nonparametric Sen slope standardized by residual errors relative to the Sen slope line.

mean bias. For all approaches, absolute bias increases as standardized trend magnitude increases; however, bias associated with use of the full record is consistently larger than any of the subset approaches. While there is large bias for the most extreme trends, most points are concentrated between standardized Sen slopes of ±0.025 and have a relatively low bias.

4 Discussion and conclusions

Given evidence of trends in low-flow series in many streams (Lins and Slack, 1999, Stahl *et al.* 2010, Du *et al.* 2015, Kam and Sheffield 2016, Kormos *et al.* 2016), updated methods for estimation of low-flow statistics are needed. Using controlled Monte Carlo simulation experiments, we evaluated simple approaches to improve estimation of a common low-flow statistic, 7Q10, defined as the annual minimum 7-day streamflow which is exceeded in 9 out of 10 years, on average. Relative to the standard practice of using the entire available streamflow record, we found that selecting a more recent subset of a long record can improve accuracy and reduce bias of 7Q10 estimators, particularly for records which exhibit large magnitude trends.

Among the estimators considered here, 7Q10 estimators based on the most recent 30 years (Subset 30 strategy) provided a reasonable approach to improve the accuracy and reduce the bias of estimators when trends in the flow record were present. However, an adaptive approach using a trend

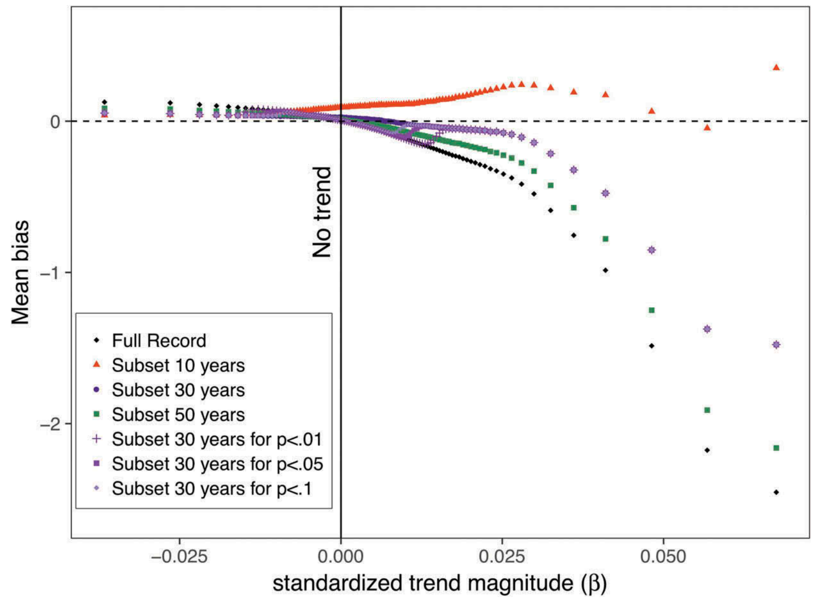

Figure 5. Mean bias of 7Q10 estimators calculated using a full 75-year record, non-adaptive approaches (Subset 10, 30, or 50 years), and adaptive approaches for Subset 30 ($p < 0.01$, 0.05, or 0.1). Standardized trend magnitude refers to the nonparametric Sen slope standardized by residual errors relative to the Sen slope line. While the points (purple) for the non-adaptive Subset 30 and the three Subset 30 adaptive approaches mostly coincide with one another, differences are distinguishable between standardized trend magnitudes of approximate 0 to 0.02.

test showed greater accuracy regardless of whether a trend was detected or not. By applying the Subset 30 strategy only to records with statistically significant trends, this adaptive approach eliminated the loss of accuracy associated with selecting a subset of a record without a detectable trend. We found that the choice of a statistical significance level (p values of 0.01, 0.05, and 0.1) yielded similar 7Q10 estimators with comparable accuracy and bias. The largest p value ($p = 0.1$) appeared to be marginally more accurate compared to the other p values tested, but the difference was small. Based on the experiment carried out here, Fig. 6 illustrates

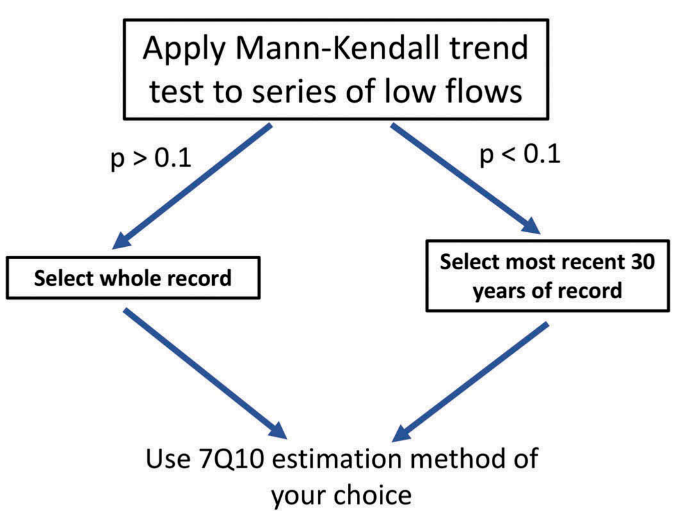

Figure 6. The adaptive Subset 30 strategy in which 7Q10 is estimated using the full record when no trend is detected (Mann-Kendall significance $p > 0.1$) and otherwise estimating 7Q10 using the most recent 30 years of record. This is the recommended strategy based on the simulation experiments for 174 USGS streamgages in the mid-Atlantic USA.

the use of the best-performing approach, the adaptive Subset 30 strategy.

This study provides a preliminary exploration of simple approaches to selecting a subset of a long record of annual minimum 7-day flows for use in low-flow frequency analysis. Many assumptions had to be made which should be more fully explored in future work. First, experimental trend magnitudes were based on 174 perennial streamgages in the mid-Atlantic USA and may not be representative of standardized trends in other locations. Multiple trends, abrupt shifts in trends, nonlinear trends, and reversing trends at some streamgages are probable but were beyond the scope of work considered here. We also did not consider a change-point in the variance of the flow in our simulation experiments.

Our assumptions that annual minimum 7-day flows could be approximated by a two-parameter lognormal distribution and that these flows were temporally independent helped to make the experiments of reasonable scope, but these are only approximations. Year-to-year correlations have been found previously in annual minimum low-flow series (Douglas *et al.* 2002). Future work could explore the impact of relaxing this independence assumption or of techniques such as trend-free pre-whitening processes, such as those introduced by Yue *et al.* (2002) and others, to remove possible lag-1 autocorrelation in annual and other time series. An assumption of independence generally results in a higher number of reported statistically significant trends; thus, pre-whitening would likely result in fewer identified trends, and, under application of the adaptive approach studied here, more frequent use of the full available streamflow record. Finally, we focused on a range of standardized trends because we were interested in the signal-to-noise ratio of trend-to-variance in the flow record. Blum (2017) found comparable results to those presented here with a similar experiment studying non-standardized trends across a range of coefficients of variance and including step change trends. Greater understanding of the combined role of variance of the flow record and trend magnitude is needed before operationalizing these methods. Additionally, study of how these approaches perform across a variety of basins, such as regulated, urban, or agricultural, could inform their application.

The subset methods described here can only update 7Q10 to more recent conditions, which may or may not reflect streamflow behavior in the future. Thus, these procedures should not be used to predict future 7Q10, but rather can provide a more accurate estimator of present value 7Q10 compared to use of the full record, especially if future conditions are expected to be similar to current conditions. We recommend that 7Q10 be reported with confidence intervals, as described in Ames (2006) and shown in Fig. 2. Comparing the estimators introduced in this study to such confidence intervals can help inform practitioners whether changes in streamflow have resulted in significantly distinct values of 7Q10.

These sorts of approaches – providing a "snapshot" of the most recent low-streamflow conditions at a streamgage – would ideally be updated annually based on newly available flow records. Future research could compare subset approaches to other methods to account for nonstationarity, such as estimation of 7Q10 from a de-trended record with the trend added back into the estimator, which has the advantage of using all of the data and thus capturing the variability of the full record available. The disadvantage of such an approach would be the additional complexity of detangling the distinct deterministic and stochastic components of the time series such that the deterministic component can be added back.

Given the popularity of developing nonstationary models for floods, such approaches for low flows will likely increase as well. However, challenges with modeling nonstationarity in minimum streamflows identified by previous studies will remain a challenge, including simultaneous changes in anthropogenic impacts (Allaire *et al.* 2015, Du *et al.* 2015) and uncertainty about future conditions (Montanari and Koutsoyiannis 2014, Serinaldi *et al.* 2018). Determining the causes of observed trends, including flow regulation, land-use change, water withdrawals, and climate change, can help inform understanding for the physical drivers of trends (Steinschneider and Brown 2012, Luke *et al.* 2017) aiding more accurate prediction of 7Q10 under changing conditions. However, attributing causal effects of physical drivers on low flows is very challenging, as reflected by a range of inconsistent findings in the literature (Price 2011). As such, subset approaches such as those studied here provide a promising way forward in providing practitioners with a simple update to their current estimation approach. Additionally, it may be worthwhile to consider using low-flow metrics other than 7Q10 which can better reflect trends and changing streamflow conditions.

We hope that the findings from this study will promote and inform future work on nonstationary low-flow frequency analysis. As hydrologists, we now are presented with a good problem to have – so much data that we must thoughtfully consider which parts are most useful or appropriate for a particular application! Determining which data to use for different applications is thus a growing challenge. Approaches studied here are not limited to the estimation of 7Q10; similar Monte Carlo experiments could be useful in assessing methods of updating other low-flow statistics, such as the median annual 7-day minimum streamflow. Changing streamflow conditions makes accurate estimation of flow statistics difficult. Given uncertainty about future conditions, starting with small adjustments to existing and well-understood methods of frequency analysis presents a promising way forward. With more research on a broader range of trend types and differing experimental assumptions, we can have greater confidence in the applicability of simple subset approaches for record selection for estimation of low-flow statistics.

Acknowledgements

Many thanks to three anonymous reviewers, as well as to Associate editor, Elena Volpi, Glenn Hodgkins, Chuck Kroll, and Jory Hecht for their helpful comments on this paper.

Disclaimer

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This material is based upon work supported by the United States National Science Foundation [grant numbers DGE-1144081 and EEC-1444926] and EPA Interagency Agreement Number DW-14-92433201. AGB also would like to acknowledge the Glenadore and Howard L. Pim Postdoctoral Fellowship, which supported her while completing this work.

ORCID

Annalise G. Blum http://orcid.org/0000-0003-4618-6181
Stacey A. Archfield http://orcid.org/0000-0002-9011-3871
Robert M. Hirsch http://orcid.org/0000-0002-4534-075X
Richard M. Vogel http://orcid.org/0000-0001-9759-0024
Julie E. Kiang http://orcid.org/0000-0003-0653-4225
Robert W. Dudley http://orcid.org/0000-0002-0934-0568

References

Ahn, K. and Palmer, R.N., 2016. Use of a nonstationary copula to predict future bivariate low flow frequency in the Connecticut river basin. *Hydrological Processes*, 3532 (June), 3518–3532. doi:10.1002/hyp.10876

Allaire, M.C., Vogel, R.M., and Kroll, C.N., 2015. The hydromorphology of an urbanizing watershed using multivariate elasticity. *Advances in Water Resources*, 86, 147–154. doi:10.1016/j.advwatres.2015.09.022

Ames, D.P., 2006. Estimating 7Q10 confidence limits from data: a bootstrap approach. *Journal of Water Resources Planning and Management*, 132 (3), 204–208. doi:10.1061/(ASCE)0733-9496(2006)132:3(204)

Asarian, J.E. and Walker, J.D., 2016. Long-term trends in streamflow and precipitation in northwest California and southwest Oregon, 1953–2012. *JAWRA:Journal of the American Water Resources Association*, 52 (1), 241–261. doi:10.1111/1752-1688.12381

Bayazit, M., 2015. Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. *Environmental Processes*, 2 (3), 527–542. doi:10.1007/s40710-015-0081-7

Blum, A.G., 2017. *Characterizing streamflow variability: distributions, trends, and ecological impacts*. Dissertation (PhD). Medford, MA: Tufts University.

Burke, E.J., Perry, R.H.J., and Brown, S.J., 2010. An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology*, 388 (1–2), 131–143. doi:10.1016/j.jhydrol.2010.04.035

Cancelliere, A., 2017. Non stationary analysis of extreme events. *Water Resources Management*, 31 (10), 3097–3110. doi:10.1007/s11269-017-1724-4

Dingman, S.L. and Lawlor, S.C., 1995. Estimating low-flow quantiles from drainage-basin characteristics in New Hampshire and Vermont. *JAWRA: Journal of the American Water Resources Association*, 31 (2), 243–256. doi:10.1111/jawr.1995.31.issue-2

Douglas, E.M., Vogel, R.M., and Kroll, C.N., 2000. Trends in floods and low fows in the United States: impact of spatial correlation. *Journal of Hydrology*, 240 (1–2), 90–105. doi:10.1016/S0022-1694(00)00336-X

Douglas, E.M., Vogel, R.M., and Kroll, C.N., 2002. Impact of streamflow persistence on hydrologic design. *Journal of Hydrologic Engineering*, 7 (3), 220–227. doi:10.1061/(ASCE)1084-0699(2002)7:3(220)

Du, T., *et al.*, 2015. Return period and risk analysis of nonstationary low-flow series under climate change. *Journal of Hydrology*, 1–63.

Falcone, J.A., 2011. GAGES-II: geospatial attributes of gages for evaluating streamflow (Digital dataset). *US Geological Survey*, 527, 234–250. doi:10.3133/70046617

García Galiano, S.G., *et al.*, 2011. Assessing drought hazard under non-stationary conditions on southeast of Spain. *In*: G. Blöschl, *et al.*, ed. *Risk in water resources management*. Wallingford, UK: International Association of Hydrological Sciences, IAHS Publication no.347, 85–92.

Gebert, W.A., Garn, H.S., and Rose, W.J., 2016. Changes in streamflow characteristics in Wisconsin as related to precipitation and land use. Reston, VA: US Geological Survey, Scientific Investigations Report 2015 – 5140.

Giraldo Osorio, J.D. and García Galiano, S.G., 2012. Non-stationary analysis of dry spells in monsoon season of Senegal River Basin using data from regional climate models. *Journal of Hydrology*, 450–451 (March 2015), 82–92. doi:10.1016/j.jhydrol.2012.05.029

Grandry, M., *et al.*, 2013. A method for low-flow estimation at ungauged sites: a case study in Wallonia (Belgium). *Hydrology and Earth System Sciences*, 17 (4), 1319–1330. doi:10.5194/hess-17-1319-2013

Hall, J., *et al.*, 2014. Understanding flood regime changes in Europe: a state-of-the-art assessment. *Hydrology and Earth System Sciences*, 18, 2735–2772. doi:10.5194/hess-18-2735-2014

Helsel, D.R. and Hirsch, R.M.M., 2002. *Statistical methods in water resources.* Reston, VA: Techniques of Water-Resources Investigations of the United States Geological Survey Book 4, Hydrologic Analysis and Interpretation.

Hirsch, R.M., 2011. A perspective on nonstationarity and water management. *JAWRA: Journal of the American Water Resources Association*, 47 (3), 436–446. doi:10.1111/j.1752-1688.2011.00539.x

Jiang, C., *et al.*, 2014. Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula. *Hydrological Processes*. doi:10.1002/hyp.10288

Kam, J. and Sheffield, J., 2016. Changes in the low flow regime over the eastern United States (1962 – 2011): variability, trends, and attributions. *Climatic Change*, 135, 639–653. doi:10.1007/s10584-015-1574-0

Khaliq, M.N., *et al.*, 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. *Journal of Hydrology*, 329 (3–4), 534–552. doi:10.1016/j.jhydrol.2006.03.004

Kormos, P.R., *et al.*, 2016. Trends and sensitivities of low streamflow extremes to discharge timing and magnitude in Pacific Northwest mountain streams. *Water Resources Research*, 52, 4990–5007. doi:10.1002/2015WR018125

Kroll, C.N. and Vogel, R.M., 2002. Probability distribution of low streamflow series in the United States. *Journal of Hydrologic Engineering*, 7 (2), 137–146. doi:10.1061/(ASCE)1084-0699(2002)7:2(137)

Lins, H.F. and Slack, J.R., 1999. Streamflow trends in the United States. *Geophysical Research Letters*, 26 (2), 227–230. doi:10.1029/1998GL900291

Liu, D., *et al.*, 2015. Climate-informed low-flow frequency analysis using nonstationary modelling. *Hydrological Processes*, 29 (9), 2112–2124. doi:10.1002/hyp.v29.9

Luke, A., *et al.*, 2017. Predicting nonstationary flood frequencies: evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, 53, 2742–2759. doi:10.1002/2016WR019676

McCabe, G.J. and Wolock, D.M., 2002. A step increase in streamflow in the conterminous United States. *Geophysical Research Letters*, 29 (24), 38-1-38–4. doi:10.1029/2002GL015999

Modarres, R., 2008. Regional frequency distribution type of low flow in North of Iran by L-moments. *Water Resources Management*, 22 (7), 823–841. doi:10.1007/s11269-007-9194-8

Montanari, A. and Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: stationarity is immortal! *Water Resources Research*, 50 (12), 9748–9756. doi:10.1002/2014WR016092

Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review. *Progress in Physical Geography*, 35 (4), 465–492. doi:10.1177/0309133311402714

Prosdocimi, I., Kjeldsen, T.R., and Svensson, C., 2014. Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Natural Hazards and Earth System Sciences*, 14, 1125–1144. doi:10.5194/nhess-14-1125-2014

Riggs, H.C., 1972. Low-Flow investigation. *US Geological Survey Techniques of Water-Resources Investigations*. Washington, DC: United States Geological Survey.

Sadri, S., Kam, J., and Sheffield, J., 2016. Nonstationarity of low flows and their timing in the eastern United States. *Hydrology and Earth System Sciences*, 20, 633–649. doi:10.5194/hess-20-633-2016

Salas, J.D., *et al.*, 2012. Special section on climate change and water resources: climate nonstationarity and water resources management. *Journal of Water Resources Planning and Management*, 138 (October), 385–388. doi:10.1061/(ASCE)WR.1943-5452.0000279

Salas, J.D., Obeysekera, J., and Vogel, R.M., 2018. Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal*, 63 (3), 325–352. doi:10.1080/02626667.2018.1426858

Sawaske, S.R. and Freyberg, D.L., 2014. An analysis of trends in baseflow recession and low-flows in rain-dominated coastal streams of the pacific coast. *Journal of Hydrology*, 519, 599–610. doi:10.1016/j.jhydrol.2014.07.046

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's Tau. *Journal of the American Statistical Association*, 63 (324), 1379–1389. doi:10.1080/01621459.1968.10480934

Serinaldi, F., Kilsby, C.G., and Lombardo, F., 2018. Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources*, 111, 132–155. doi:10.1016/j.advwatres.2017.10.015

Smakhtin, V.U., 2001. Low flow hydrology: a review. *Journal of Hydrology*, 240, 147–186. doi:10.1016/S0022-1694(00)00340-1

Small, D., Islam, S., and Vogel, R.M., 2006. Trends in precipitation and streamflow in the eastern U.S.: paradox or perception? *Geophysical Research Letters*, 3 (33). doi:10.1029/2005GL024995

Stahl, K., *et al.*, 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences*, 14, 2367–2382. doi:10.5194/hess-14-2367-2010

Steinschneider, S. and Brown, C., 2012. Forecast-informed low-flow frequency analysis in a Bayesian framework for the northeastern

United States. *Water Resources Research*, 48 (10), n/a–n/a. doi:10.1029/2012WR011860

Tu, X., *et al*., 2016. Uncertainty and variability in bivariate modeling of hydrological droughts. *Stochastic Environmental Research and Risk Assessment*, 30 (5), 1317–1334. doi:10.1007/s00477-015-1185-3

US Geological Survey, 2017. *National Water Information System* [online]. Reston, VA. Available from: http://waterdata.usgs.gov/nwis/ [Accessed 15 November 2016].

Vogel, R.M. and Fennessey, N.M., 1994. Flow-duration curves. I: new interpretation and confidence intervals. *Journal of Water Resources Planning and Management*, 120 (4), 485–504. doi:10.1061/(ASCE)0733-9496(1994)120:4(485)

Vogel, R.M. and Kroll, C.N., 1989. Low-flow frequency analysis using probability-plot correlation coefficients. *Journal of Water Resources Planning and Management*, 115 (3), 338–357. doi:10.1061/(ASCE)0733-9496(1989)115:3(338)

Vogel, R.M. and Kroll, C.N., 1990. Generalized low-flow frequency relationships for ungaged sites in Massachusetts. *JAWRA: Journal of the American Water Resources Association*, 26 (2), 241–253. doi:10.1111/jawr.1990.26.issue-2

Vogel, R.M., Yaindl, C., and Walter, M., 2011. Nonstationarity: flood magnification and recurrence reduction factors in the United States. *JAWRA: Journal of the American Water Resources Association*, 47 (3), 464–474. doi:10.1111/j.1752-1688.2011.00541.x

Wang, Y., *et al*., 2015. A time-dependent drought index for non-stationary precipitation series. *Water Resources Management*, 29 (15), 5631–5647.

WMO (World Meteorological Organization), 2008. *Manual on low-flow estimation and prediction*. Geneva, Switzerland: World Meteorological Organization, Operational Hydrology Report no. 50.

Yu, X., 2017. *Flood frequency analysis in the context of climate change or with mixed populations*. Dissertation (PhD). Ithaca, NY: Cornell University.

Yue, S., *et al*., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, 16 (9), 1807–1829. doi:10.1002/(ISSN)1099-1085