Advances in Water Resources 77 (2015) 17–36

Contents lists available at ScienceDirect

Advances in Water Resources

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

# Stationarity is undead: Uncertainty dominates the distribution of extremes

Francesco Serinaldi *, Chris G. Kilsby

*School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK*
*Willis Research Network, 51 Lime St., London EC3M 7DQ, UK*

ARTICLE INFO

*Article history:*
Received 7 July 2014
Received in revised form 28 December 2014
Accepted 30 December 2014
Available online 8 January 2015

*Keywords:*
Nonstationary flood frequency analysis
Nonstationary return period
Risk of failure
Nonstationary confidence intervals
Generalized linear models
Generalized additive models

ABSTRACT

The increasing effort to develop and apply nonstationary models in hydrologic frequency analyses under changing environmental conditions can be frustrated when the additional uncertainty related to the model complexity is accounted for along with the sampling uncertainty. In order to show the practical implications and possible problems of using nonstationary models and provide critical guidelines, in this study we review the main tools developed in this field (such as nonstationary distribution functions, return periods, and risk of failure) highlighting advantages and disadvantages. The discussion is supported by three case studies that revise three illustrative examples reported in the scientific and technical literature referring to the Little Sugar Creek (at Charlotte, North Carolina), Red River of the North (North Dakota/Minnesota), and the Assunpink Creek (at Trenton, New Jersey). The uncertainty of the results is assessed by complementing point estimates with confidence intervals (CIs) and emphasizing critical aspects such as the subjectivity affecting the choice of the models' structure. Our results show that (1) nonstationary frequency analyses should not only be based on at-site time series but require additional information and detailed exploratory data analyses (EDA); (2) as nonstationary models imply that the time-varying model structure holds true for the entire future design life period, an appropriate modeling strategy requires that EDA identifies a well-defined deterministic mechanism leading the examined process; (3) when the model structure cannot be inferred in a deductive manner and nonstationary models are fitted by inductive inference, model structure introduces an additional source of uncertainty so that the resulting nonstationary models can provide no practical enhancement of the credibility and accuracy of the predicted extreme quantiles, whereas possible model misspecification can easily lead to physically inconsistent results; (4) when the model structure is uncertain, stationary models and a suitable assessment of the uncertainty accounting for possible temporal persistence should be retained as more theoretically coherent and reliable options for practical applications in real-world design and management problems; (5) a clear understanding of the actual probabilistic meaning of stationary and nonstationary return periods and risk of failure is required for a correct risk assessment and communication.


## 1. Introduction

Do nonstationary distributions and related concepts really provide more reliable predictions of hydrological variables to be used in design, planning and management? This question can synthesize in a simple but effective manner what practitioners ask more and more often of researchers and academics when they face for instance the increasing requirement to account for the effect of climate fluctuations, anthropogenic interventions in river basins, and other sources of so-called nonstationarity in their hydrological frequency analyses. This question also arises when looking at the increasing number of scientific papers dealing with nonstationary extreme value frequency analysis (e.g., [3,4,32,50,87,88,102,109,110], among others), and the requirements of the national adaptation programs/strategies to climate change that implicitly or explicitly claim an update of the current protocols of flood risk assessment to take nonstationarity into account (e.g., [2,18–20,114]).

Focusing on flood risk assessment (but the discussion holds true also for other environmental variables such as extreme rainfall and drought events), Madsen et al. [51] provided an up-to-date overview of the current practices adopted across the European countries to deal with the changing risk under possible future climate scenarios, concluding that "*The review of existing guidelines in Europe on design floods and design rainfalls… are based on climate*

* Corresponding author at: School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK.

*E-mail address:* francesco.serinaldi@ncl.ac.uk (F. Serinaldi).

http://dx.doi.org/10.1016/j.advwatres.2014.12.013
0309-1708/

*change adjustment factors to be applied to current design estimates and may depend on design return period and projection horizon...Most of the studies reported are based on frequency analysis assuming stationary conditions in a certain time window (typically 30 years) representing current and future climate. There is a need for developing more consistent non-stationary frequency analysis methods that can account for the transient nature of a changing climate*". As recognized by Stedinger and Griffis [93], extension of the models of flood probability to include changes in flood risk over time is relatively easy mathematically and available (along with well documented software packages) for the last years (e.g., [5,8,17,25,33–35,95,96]). However, in spite of the extensive literature on specific methods concerning EDA, model fitting and goodness-of-fit testing, only few works have recently tackled the practical problems of moving from stationary to nonstationary frequency analyses facing with real-world applications and uncertainty (e.g., [9,60,65,76,78,87,88,93,111]).

Provided that we have sufficient technology to perform nonstationary frequency analysis and extend the current practices available in standard guidelines into a nonstationary framework [65], can we really do that safely under the usual lack of data and information characterizing hydrological variables? Even though possible problems and warnings related to nonstationary frequency analysis have already been mentioned by some authors, the focus is usually on particular aspects such as the preliminary exploratory analyses or the choice of the model/distribution. Therefore, the aim of this paper is to summarize the whole inference procedure involved in nonstationarity modeling, the available tools and their relationship with stationary concepts, and explicitly highlight the possible operational difficulties faced by an analyst in each step of the data analysis. We also emphasize the role of uncertainty and the importance of performing a fair comparison between stationary and nonstationary results. It should be noted that this study tackles nonstationary frequency analysis from a "frequentist" perspective as this is the most applied approach both in the scientific and technical literature. However, based on the empirical results, the possible advantages of using a Bayesian approach are mentioned in the conclusions, where some key references are also listed.

This paper is organized as follows. In Section 2, a simple Gumbel distribution with linearly time-varying parameters is used to introduce the rationale of nonstationary models whose general mathematical structure is discussed in Section 3. Section 4 revises the approaches proposed in the literature to summarize the risk and obtain design values with a prescribed degree of rarity. In particular, we highlight the probabilistic meaning, the link between such approaches, and the actual information they convey. Section 5 reviews the methods to quantify the sampling uncertainty via confidence intervals (CIs) in the nonstationary framework highlighting its fundamental role for a fair comparison between models and a fair assessment of the output reliability. Three case studies already discussed in the literature are examined in Section 6, highlighting the critical point of each stage of the analysis. Each example has its own merit, emphasizing particular aspects. Conclusions are provided in Section 7 as critical guidelines to perform nonstationary frequency analyses in real-world applications.

## 2. Setting the framework

The difference between stationary and nonstationary frequency analyses can be summarized as follows: stationary analyses imply that the observed values of a variable under study (e.g., flood peak discharge) are assumed to be independent and identically distributed (*iid*) realizations of a random variable $Y$ with stationary distribution $F_Y(y;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters, whereas nonstationary analyses imply that the observations are independent but not necessarily identically distributed (*i/nid*) realizations drawn from a nonstationary distribution $F_Y(y;\boldsymbol{\theta}(\mathbf{X}))$ whose parameters $\boldsymbol{\theta}$ are not constant but change as a function of $\mathbf{X}$, which is a vector of so-called covariates. For example, $\mathbf{X}$ may be time $t$, meaning that the parameters change as a function of $t$. To clarify the concepts, let us consider the Gumbel (GUM) distribution, a well-known model in engineering practice. The cumulative distribution function (CDF) for the stationary model reads as:

$$F_Y(y;\mu,\delta) = \exp\left(-\exp\left(-\frac{y-\mu}{\delta}\right)\right), \tag{1}$$

where $-\infty < y < \infty$, $\mu$ is a location parameter and $\delta$ is a scale parameter. Assuming that both parameters are simple linear functions of time $t$, a possible version of the nonstationary GUM model is

$$F_Y(y;\mu(t),\delta(t)) = \exp\left(-\exp\left(-\frac{y-(\mu_0+\mu_1 t)}{(\delta_0+\delta_1 t)}\right)\right). \tag{2}$$

Eq. (2) is a simple example of the generic nonstationary representation $F_Y(y;\boldsymbol{\theta})$, where $F_Y$ specializes as a GUM CDF, $\mathbf{X}$ is $t$, and the vector of parameters $\boldsymbol{\theta}$ specializes as $(\mu(t)=\mu_0+\mu_1 t,\ \delta(t)=\delta_0+\delta_1 t)$. Salas and Obeysekera [78] recognized that actually $\mathbf{X}$ can include $t$ (e.g., [9,33,88,94,109]), stochastic models with shifting patterns (e.g., [98]), exogenous covariates such as large scale climate indices, population, and other hydrological variables (e.g., [26,84,86,108,110]).

In spite of its conceptual simplicity, Eq. (2) already helps highlight some critical points:

- In a stationary framework, an analytical model (e.g., Eq. (1)) is fitted to data in order to predict the probability corresponding to a given value of the hydrological variable or vice versa to estimate the value corresponding to a prescribed probability of exceedance. That is to say that the model is used to extrapolate the probability law beyond the range of the observed values and frequencies (e.g., [80,82]). In a nonstationary context the dependence of the parameters on $t$ or other covariates implies the additional extrapolation of the law linking parameters and covariates (e.g., [76]). For instance, if we want to use the model in Eq. (2) to compute design values (with a given probability of exceedance) over a future design life period, we need to assume that the linear variation of the GUM parameters holds true for the entire design life period. Therefore, nonstationary analysis introduces a further source of uncertainty that should carefully be taken into account if the relationships between model parameters and covariates are defined inductively (i.e. making direct use of observations) and covariates vary in time according to deterministic laws.
- The above remark highlights another aspect: the relationship between parameters and covariates and the temporal pattern of the covariates have to be deterministic in order to generate truly nonstationary models because purely random or stationary stochastic fluctuations of the parameters or parameters depending on stochastically varying covariates simply generate stationary compound or mixed distributions (e.g., [15,40,106]). Therefore, the law of variation of the parameters should reflect reasonable predictable physical mechanisms to guarantee that the patterns observed in the period of record is not just an effect of fluctuations of stationary processes whose dynamics evolve over longer time scales (e.g., [42,43]).

– The previous remark rises in turn a third aspect concerning the use itself of nonstationary models. Provided that nonstationarity is actually a property characterizing stochastic processes and models [41,43,61], inductive inference of nonstationary models for a hydrometeorological process from finite time series of observations might be theoretically inconsistent and practically not so easy owing to the interaction of multiple factors. Indeed, inferring ensemble statistics from temporal statistics implies the assumption that the process is ergodic; however, if the process is nonstationary, the ergodicity cannot hold (see e.g., [43], and references therein) thus making inductive inference from data theoretically impossible. The use of nonstationary models can therefore be considered as an option only if the evolution of the distribution function can be related to controllable factors which can be predicted in deterministic terms [43,61]. This implies that the relationships between model parameters and predictable covariates (i.e. the model structure) cannot result from an estimation procedure from the data, but need to be defined a priori based on the temporal evolution laws of the factors and physical mechanisms mentioned above. When covariates have a stochastic behavior (in time), the inference of the model structure is justified as the final model is still stationary (in time). In this context, diagnostics devised to check for monotonic trends, abrupt changes or more complex nonlinear temporal patterns and relationships between a variable $Y$ and covariates $\mathbf{X}$ should not be used to infer nonstationarity but as tools to identify possible deterministic predictable mechanisms whose temporal evolution needs to be deduced by meta-data (e.g., effects of water abstraction scheduling, dams' construction and operation, etc.).

These points will be further discussed and expanded in the context of real-world data analyses described in Section 6.

## 3. Mathematical structure of nonstationary distributions: an overview of available models

Eq. (2) is only a very simple case of more general models developed to describe nonstationary distributions with dynamically varying parameters. Consultancy experience tells us that practitioners who were asked to implement nonstationary analyses often are not very familiar with (or do not know at all) the wide range of available alternatives. We also noted that often even stationary analyses are performed in quite a superficial manner lacking for instance a suitable assessment of uncertainty (see e.g., [77,80,82,91], for discussions involving different perspectives). Therefore, it is worth mentioning which models could be used in nonstationary frequency analysis.

The simplest class of nonstationary distribution is the so-called Generalized Linear Models (GLM) (e.g., [12,55,63]). Such models imply a distribution function $F_Y$ belonging to the exponential family (which is a large class of distributions including normal, lognormal (LNO), gamma (GAM), among many others), a linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$ and a link function $g$ such that $\mathrm{E}[Y] = g^{-1}(\eta)$, where $\mathrm{E}[\cdot]$ denotes the expected value operator. An example can help clarify the GLM structure. Since the GUM distribution in Eq. (2) belongs to the exponential family when $\delta$ is known, the model

$$F_Y(y; \mu(t), \delta) = \exp\left(-\exp\left(-\frac{y - (\mu_0 + \mu_1 t)}{\delta}\right)\right) \tag{3}$$

is an example of GLM where $F_Y$ is GUM, $\eta = \mu_0 + \mu_1 t$ (i.e. $\mathbf{X} = t$ and $\beta = (\mu_0, \mu_1)^\top$), and $\mathrm{E}[Y] = g^{-1}(\eta) = \eta + 0.5772\delta$. In other words, GLM describe distribution functions whose mean (expected value $\mathrm{E}[Y]$) is a function $g^{-1}$ of a linear combination of covariates (here, $t$ for simplicity).

Even though GLM are powerful tools with extensive application in many disciplines (e.g. biometrics, environmetrics, and econometrics, among others), they have been further generalized in three main aspects: (1) the introduction of distributions not belonging to the exponential family; (2) the introduction of nonparametric and nonlinear relationships between the model parameters and the covariates; (3) the introduction of functional relationships between higher moments and covariates, thus allowing for dynamically varying variance, skewness, and kurtosis. When the covariates reduce to $t$, these models allow not only for time varying mean, but also for changes in the full shape of a CDF. Such generalizations include Generalized Additive Models (GAM) (e.g., [30]), Generalized Linear Mixed Models (GLMM) (e.g., [56,57]), Generalized Additive Mixed Models (GAMM), (e.g., [21]), Vector Generalized Linear/Additive Models (VGLM and VGAM) (e.g., [116,117]), and Generalized Additive Models for Location, Scale and Shape (GAMLSS) (e.g., [75]).

In particular, VGLM, VGAM, and GAMLSS can be seen as the most complete frameworks devised to account for nonstationarity in the whole set of parameters for a very large set of distributions, including those mostly used in hydrology (e.g. Generalized Extreme Value (GEV), Generalized Pareto, Log-Pearson III (LP3), LNO, GAM, etc.). In order to fulfill adaptation protocols, these models are expected to enter into the tool box of hydrologists in the future. It is therefore worth giving an overview in order to understand their structure. As the rationale of both VGLM/VGAM and GAMLSS is essentially the same (main differences being in the fitting algorithms), we briefly recall the mathematical theory following the GAMLSS notation used by Rigby and Stasinopoulos [75].

Denoting $Y$ the response variable (as for GLM notation), for the GAMLSS models it is assumed that independent observations $y_i$, for $i = 1, \ldots, n$, have distribution function $F_Y(y_i; \boldsymbol{\theta}^i)$ with $\boldsymbol{\theta}^i = (\theta_1^i, \ldots, \theta_\pi^i)$ a vector of $\pi$ distribution parameters accounting for location, scale, and shape. Commonly, $\pi$ is less than or equal to four, since 1- to 4-parameter families provide enough flexibility for most applications (e.g., GUM, GEV, GP, PE3, LNO,GAM, Logistic are all 2- or 3-parameter distributions). Given an $n$-length vector of observations of the response variable $\mathbf{y}^\top = (y_1, \ldots, y_n)$, let $g_k(\cdot)$, for $k = 1, \ldots, \pi$, be monotonic link functions relating the distribution parameters to explanatory variables and random effects through an additive model given by:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk}\gamma_{jk} \tag{4}$$

where $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are vectors of length $n$, e.g. $\boldsymbol{\theta}_k^\top = \{\theta_k^1, \ldots, \theta_k^n\}$, $\boldsymbol{\beta}_k^\top = \{\beta_{1k}, \ldots, \beta_{J_k}k\}$ is a parameter vector of length $J_k$, $\mathbf{X}_k$ is a known design matrix of order $n \times J_k$, $\mathbf{Z}_{jk}$ is a fixed known $n \times q_{jk}$ design matrix and $\gamma_{jk}$ is a $q_{jk}$-dimensional random variable.

In Eq. (4), the linear predictors $\boldsymbol{\eta}_k$ comprise a parametric component $\mathbf{X}_k\boldsymbol{\beta}_k$ (linear functions of explanatory variables), and additive components $\mathbf{Z}_{jk}\gamma_{jk}$ (linear functions of stochastic variables, also denoted as random effects). VGLM/VGAM and GAMLSS involve several important sub-models. For example, when the distribution function belongs to the exponential family and only the location parameter $\theta_1$ (related to the expected value) depends linearly upon the covariates via $\mathbf{X}_k\boldsymbol{\beta}_k$, the GAMLSS reduce to GLM. In spite of the apparently complex notation, Eq. (4) simply says that the parameters $\theta_k$ are functions of some covariates. Thus, it follows that the model in Eq. (2) is a simple example of VGLM where $\theta_1 = \mathrm{E}[Y]$ and $\theta_2 = \mathrm{Var}[Y]$, and both are linear functions of the explanatory variable $t$.

Whereas stationary analyses require only the choice of a suitable distribution of $Y$, the inference procedure for GLM, VGLM/

VGAM or GAMLSS also implies the selection of the explanatory variables, the link functions, and the structure of the systematic part (i.e., linear and/or nonlinear, parametric and/or additive nonparametric functions between parameters and covariates). When the estimation method is based on the maximum likelihood principle, the model selection can be carried out by checking the significance of the fitting improvement in terms of information criteria such as the Akaike Information Criterion (AIC) [1], the Schwarz Bayesian Criterion (SBC) [79], the generalized AIC (GAIC) [92] or deviance statistic and maximum likelihood ratio tests (e.g., [8, pp. 33–36]). Forward, backward, and step-wise procedures can be applied to select the meaningful explanatory variables. Diagnostic plots to check the fitting performance comprise diagrams of residuals, such as Q–Q plots or the more effective worm plots (e.g., [92,105]) are also available. In this study, all models are estimated by the maximum likelihood method.

## 4. Risk communication under nonstationarity: a farewell to return period?

In the conventional stationary analyses for designing hydraulic structures, the terms probability of exceedance, return period, and risk of failure during the design life are widely used. Referring to Chow et al. [6], Loaiciga and Mariño [49], Fernández and Salas [23,24], Douglas et al. [13], Cooley [9] and references therein for formal introductions, here, we would like to recall and slightly expand the critical discussion reported by Serinaldi [83].

As stated by Cooley [9], the concept of return period becomes ambiguous when we move from stationary to nonstationary conditions. However, it can still be defined for operational purposes at least in two ways. The first definition is the extension to nonstationary conditions of the concept of expected occurrence interval (expected waiting time until an exceedance occurs). This extension was first derived by Olsen et al. [66], and independently by Salas and Obeysekera [78] using a simpler procedure based on a non-homogeneous geometric distribution with time-varying parameters. Under nonstationarity, $p_i = \mathbb{P}[Y_i > y_T] = 1 - F_{Y,i}(y_T)$ is not longer constant but changes for each trial (time step) $i$ along the time series. Therefore, the geometric distribution describing the number of realizations (observed at fixed time steps) that one has to wait before observing an event whose magnitude exceeds a fixed value $y_T$ becomes [78]

$$\mathbb{P}[N = n] = p_n \prod_{i=1}^{n-1} (1 - p_i), \quad n = 1, 2, \ldots, \infty \tag{5}$$

and the return period, i.e. the corresponding expected value, is [9,78]

$$T = \mathrm{E}[N] = 1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} F_{Y,i}(y_T). \tag{6}$$

Parey et al. [70,71] extended to nonstationary conditions an alternative definition of return period such that $y_T$ is the value for which the expected number of exceedances in $T$ years (trials) is equal to one. Therefore, $y_T$ is the solution of the equation [9]

$$1 = \sum_{n=1}^{T} p_i = \sum_{n=1}^{T} (1 - F_{Y,i}(y_T)). \tag{7}$$

It should be noted that this definition is not influenced by temporal dependence if present. Both Eqs. (6) and (7) need to be solved numerically. From a practical point of view, it should be noted that the summation in Eq. (6) can be extended to a finite number of steps large enough to find the solution $y_T$ by zero-cross finding algorithms (see [78], for numerical details). On the other hand, the summation in Eq. (7) has to be extended only $T$ steps ahead, thus implying the extrapolation of the relationships that describe the evolution of $F_Y$ just until $T$ time steps from the reference time (usually the present). However, from a numerical point of view both methods require almost the same implementation and computational effort.

Under nonstationarity, the risk of failure, i.e. probability of observing at least one failure in the design life period $M$, specializes as follows [66,78,89]:

$$p_M = 1 - \prod_{i=1}^{M} (1 - p_i) = 1 - \prod_{i=1}^{M} F_{Y,i}(y_T), \tag{8}$$

where $p_i = 1 - F_{Y,i}(y_T)$ is the probability that $Y$ (e.g., the annual maximum) is greater than $y_T$ in any given trial (e.g., each year), under a given climate/environmental state $i$.

It should also be mentioned that some additional measures of risk have been proposed. For instance, Rootzén and Katz [76] introduced the so-called "design life level" and "minimax design life level". The first one is strictly related to the formulation of the risk of failure. Indeed, it is the value $y_M$ that solves the equation

$$p_M = 1 - \prod_{i=1}^{M} (1 - p_i) = 1 - \prod_{i=1}^{M} F_{Y,i}(y_M) \tag{9}$$

for a fixed design life $M$ and a prescribed probability $p_M$. The minimax design life level is the value of $Y$ such that the maximal probability of exceedance in any year in the design life period is at most $p$. For example, let us suppose that $F_Y$ changes along the next $M$ years so that we have $M$ different distributions $F_{Y,i}$, for $i = 1, \ldots, M$; the minimax design life level is obtained by computing the $p$ quantile for each $F_{Y,i}$ and taking the maximum. This guarantees that the selected value has *annual* probability of exceedance equal to or larger than $p$ along the design life period.

Rootzén and Katz [76] proposed also two diagrams devised to visualize the changing risk in the design life. The so-called "risk plot" shows how the annual probability of exceedance $p_i$ corresponding to a fixed value of $Y$ changes in the design life period, whereas the "constant risk plot" shows the values $y_i$ corresponding to a fixed probability of exceedance $p$. We notice that these diagrams (especially the "constant risk plot") are quite intuitive representations previously used by Hundecha et al. [32] and Villarini et al. [110] among others, even though no particular name was assigned to them. The "constant risk plot" is used in the next sections (we refer to [76] for further details).

Eqs. (6)–(9) deserve some further remarks. As highlighted by Obeysekera and Salas [65], the previous expressions can be used in two ways: (1) given a specified value of $Y$ we need to compute its probability of exceedance or risk of failure; (2) given a prescribed value of $p$, $T$ or $p_M$, one needs to compute the corresponding values of $y_T$ or $y_M$. The first case typically occurs when one needs to verify an existing project, and is the simplest case as the equations above allow explicit calculations when the quantile $y$ (i.e. $y_T$ or $y_M$) is known. The second case corresponds with the typical design problem and requires numerical solution of Eqs. (6), (7), and (9). Keeping in mind this difference, it is evident that Eqs. (8) and (9) are identical, being different only their use: Eq. (8) implies that $y_T$ is known and is used to compute $p_M$ in a verification setting, whereas Eq. (9) is used to compute $y_M$ for a given value of risk of failure $p_M$ in a design setting.

Regardless of its definition, the return periods and/or corresponding return levels $y_T$ yielded by Eqs. (6) and (7) simply summarize exactly or approximately the average annual probability of exceedance similar to the univariate case. This can easily be deduced for instance from Eq. (7) dividing both terms by $T$ and taking the reciprocal:

$$\frac{1}{T} = \frac{1}{T} \sum_{n=1}^{T} p_i = \bar{p} \quad \Rightarrow \quad T = \frac{1}{\bar{p}} = \frac{1}{1 - \overline{F_{Y,i}(y_T)}}. \tag{10}$$

Therefore, the difference between stationary and nonstationary return levels $y_T$ only depends on how much the nonstationary distribution diverges from the stationary. Indeed, if the pattern of the nonstationary distribution fluctuates around the stationary one so that $\bar{p} \approx p$, the stationary and nonstationary values of $y_T$ might be close to each other (examples are shown in Section 6). Moreover, similar to stationary $T$, Eq. (10) reveals that the nonstationary return period simply provides an alternative description of the average value of the probabilities of exceedance $p_i$. In other words, one can select a prescribed average *annual* probability of exceedance $\bar{p}$ to be met in the $T$ period and compute $y_T \equiv y_{\bar{p}}$ directly from Eq. (10) without introducing the concept of return period [83].

Furthermore, since also design values with large $p$ and $\bar{p}$ or small $T$ are often characterized by high risk of failure, as for the stationary case, the variables of true interest are $p_M$ and $y_M$. Indeed, unlike $\bar{p}$ (or $T$) and $y_T$, which describe the average annual risk (note the sum of probabilities in Eqs. (6) and (7)), $p_M$ and $y_M$ really measure the joint probability of exceedance in the design life period under *i*/*nid* conditions (note the product of probabilities in Eqs. (8) and (9)). In addition, the concept of risk of failure can easily be extended to non-independent non-identically distributed observations replacing the product of probabilities with the joint distribution $F_{\mathbf{Y}}$ (e.g., [10,11]). Using copula notation [64] to highlight the role of the nonstationary marginal distributions, we can write:

$$\begin{aligned} p_M &= 1 - F_{\mathbf{Y}}(Y(1) \leqslant y_M, Y(2) \leqslant y_M, \ldots, Y(M) \leqslant y_M) \\ &= 1 - C(F_{Y,1}(y_M), F_{Y,2}(y_M), \ldots, F_{Y,M}(y_M)), \end{aligned} \tag{11}$$

where $C$ denotes the copula describing the dependence structure, i.e. the temporal dependence of $Y(i)$ when the marginal distribution $F_Y$ varies dynamically with time $t$. Eq. (11) (1) describes the actual risk of failure in the design life period; (2) specializes as Eqs. (8) and (9) (according to its use for verification or design, respectively) under *i*/*nid* conditions, and as other forms under *iid* and non-independent *id* conditions [10,11,13,23,83]; (3) does not imply elaborated analytical derivations and/or reasoning, and extrapolations beyond the design life, and has an easy and straightforward interpretation. In this respect, the minimax design life level proposed by Rootzén and Katz [76] can be considered an alternative method (as many others) essentially based on the marginal distributions $F_{Y,i}$, $i, \ldots, M$, and particularly on that marginal distribution returning the most extreme quantile for a fixed *annual* probability $p$. Therefore, in some cases implying particular trends, it can be a cautionary option compared with the $\bar{p} = 1/T$ quantile, but does not summarize the collective risk to which the project is exposed in the design life period.

## 5. Quantifying the sampling and estimation uncertainty

Since hydrological frequency analyses are usually based on short time series such as a few tens of annual maxima or slightly longer sequences of peaks over threshold, the sampling and estimation uncertainties are commonly large. For instance the 95% CI of the probability of exceedance of the largest observation in a sample of size 50, which is estimated as $1/(50+1) = 0.0196$ according to the Weibull probability unbiased estimator, ranges from 0.0005 to 0.071 [38]. In other words, we can only say that there is a 95% probability that the interval between about fourteen and two thousand years contains the "true" return period of our "observed 50-year event". "*Anything beyond this kind of specification is speculation, notwithstanding any mathematical legerdemain by which it could have been obtained*" [38]. This result is not surprising and already provided by conventional statistical inference techniques for *iid* data when uncertainty is accounted for (e.g., [80,82,93]).

A fair comparison between stationary and nonstationary models also requires the assessment of the uncertainty of nonstationary probabilities, risk of failure, and design quantiles. Indeed, comparing point estimates is unfair and of little use if we do not evaluate if the differences are significant from an operational point view, based on the available knowledge (data and meta-data). In this respect, Obeysekera and Salas [65] provided a detailed overview of the available methods to compute CIs for quantiles corresponding to a given nonstationary return period. Such methods include (1) the so-called delta method (e.g., [8, pp. 31–33]), (2) the bootstrap resampling (e.g., [16]); and (3) the profile likelihood function (e.g., [8, pp. 34–36]). Referring to Obeysekera and Salas [65] for technical details, here we recall the rationale of these approaches and their suitability for practical applications. The delta method relies on the asymptotic properties of the maximum likelihood estimates of the model parameters and their covariance matrix. It yields the variance of the quantiles ($y_T$ or $y_M$) which is used to calculate approximate symmetric CIs under the hypothesis that the distribution of the quantiles is reasonably described by a Gaussian distribution. Thus, the $100(1-\alpha)\%$ confidence limits read as $\hat{y}_T \pm z_{1-\alpha/2}\hat{\sigma}$, where $z_\alpha$ denote the $\alpha$ standard Gaussian quantile, and $\hat{\sigma}$ is the standard deviation of the quantile estimator. The bootstrap method relies on the resampling of the observed series (nonparametric bootstrap) or the simulation of *iid* realizations drawn from suitable standardized distributions (referred to as parametric bootstrap or Monte Carlo simulation). This method is data-driven, is independent of the estimation method, does not rely on asymptotic assumptions, and gives an assessment of the sampling and parameter estimation uncertainties by realistic asymmetric CIs. Similar to delta method, CIs based on the profile likelihood function also rely on the asymptotic properties of maximum likelihood estimators, in particular on the asymptotic distribution of the deviance statistic (which is also used for model selection as mentioned above). Since the profile likelihood function is usually asymmetric, the resulting CIs are commonly asymmetric and deemed more accurate than those obtained by the delta method (e.g., [8, p. 35]). However, this method can be quite burdensome computationally [65].

Cooley [9] highlighted that any method for generating CIs has drawbacks, but most of them return a useful measure of the uncertainty of the quantile estimates. Even though Obeysekera and Salas [65] did not provide a ranking of the three methods, they suggest the profile likelihood as preferable although they caution of the computational burden involved. Provided that the performance of CI estimators should be assessed in terms of coverage probability by extensive Monte Carlo simulations for instance (see e.g., [44,80]), our experience with models involving different families of distributions (GEV, LP3 and others) as well as complex model structures (see e.g., [81]) suggests that the profile likelihood is not very practical not only computationally but also because it is strictly related to a single estimation method (maximum likelihood) and relies on large-sample properties which often are not met in real-world analyses (e.g., [40,68,69,85]). On the other hand, the delta method is easy to implement even for models with complex structure and is very effective to have a quick (even if approximated) look at the magnitude of the uncertainty without resorting to Monte Carlo simulations. Nonparametric or parametric bootstrap method is the most practical technique because it strictly relies on the available information without any asymptotic hypotheses, does not depend on a particular estimation method, and can easily be implemented regardless of the model complexity. Therefore, the delta method can be used for a quick preliminary check, and then bootstrap CIs calculated to refine the uncertainty estimates to be communicated. In this study, we use parametric bootstrap and the so-called "percentile" CIs (e.g., [16, pp. 170–174]). Moreover, as is shown in Section 6, the uncertainty of extreme quantiles is usually so large that the discrepancies between different types of CIs can be no much relevant. In this respect, a fair and suitable visualization of the results plays a key role in showing that the choice of the method used to compute CIs can be secondary and cannot remove the usual lack of knowledge and information characterizing extreme events (e.g., [36,38]).

## 6. Applications

### 6.1. Analysis of the Little Sugar Creek data

#### 6.1.1. Little Sugar Creek: data, preliminary remarks and EDA

The first example of nonstationary frequency analysis concerns the annual peak records of the Little Sugar Creek previously studied by Villarini et al. [110] and Salas and Obeysekera [78]. Referring to Villarini et al. [110] for a comprehensive presentation of the data, we recall that the Little Sugar Creek drains the urban core of Charlotte in Mecklenburg County (North Carolina, United States) with a drainage area of 110 km$^2$ at the United States Geological Survey (USGS) gaging station at Archdale. The gaging station was relocated in 1977 at a site upstream of its current location; the composite record from 1924 to 2007 comprises data from the dismissed station 02146500 (1924–1977) and the operating station 02146507 (1978–2012). A drainage area correction was applied to the record prior to 1977 to reduce the impact of relocation. USGS flagged the data from 1927 with the qualification code C ("All or part of the record affected by Urbanization, Mining, Agricultural changes, Channelization, or other"; data accessed on February 11, 2014).

The Little Sugar Creek at Archdale presents quite a complex time evolution of flood peaks, with an increase in magnitudes and variability of flood peaks beginning in the 1960s corresponding with a rapid urbanization processes in Charlotte. Villarini et al. [110] used GAMLSS to model the 83 years of flood peaks (from 1924 to 2007) selecting a GUM distribution (Eq. (1)), with identity and logarithmic link functions $g_k$, $k = 1, 2$, for location and scale parameters $\mu$ and $\delta$, and cubic spline smoothing curves to describe the relationship between the distribution parameters and covariates (namely, rainfall and population), i.e. the additive components $\mathbf{Z}_{jk}\gamma_{jk}$ in Eq. 4. Salas and Obeysekera [78] repeated the modeling exercise using a simpler GUM model with linearly time-varying location and scale parameters (Eq. (2)) assuming that the location parameter was constant before 1945.

We start the re-analysis from the time series diagrams. The sequence of the flood peaks are shown in Fig. 1 with up-to-date observations for the period 2008–2012. Villarini et al. [110], Cooley [9] and Salas and Obeysekera [78], used a symbol graph (showing only data points) as in Fig. 1(a), whereas we use a connected symbol graph (showing data connected with lines; Fig. 1(b)–(e)). Even though this may seem to be a trivial detail, actually it allows for a better portrayal of the temporal pattern in time series according to research on visual perception [7, pp. 180–192]. Indeed, in spite of its simplicity, this visual expedient sheds more light on the temporal pattern than more complex statistical tools. Fig. 1(b) shows that the flood peaks experienced periods with values persistently below and above the overall average. Fig. 1(c) and (d) illustrate the patterns chosen by Salas and Obeysekera [78] and Villarini et al. [110], respectively, to describe the time-varying location parameter (which is related to the time-varying average) of the GUM distribution. The functional patterns are reasonable (and were selected using information criteria, likelihood ratio tests and several diagnostic plots); however, the stepwise pattern in Fig. 1(e) is also credible even though it was simply drawn using the "eyeball" criterion. To confirm our guess (i.e., the possible presence of change points in the mean and variance of the flood peaks) we applied ex-post the Pettitt test [72,109] recognizing a significant change point in mean (variance) in 1970 (1968) and two non significant shifts in 1937 (1942) and 1992 (1994), which are in agreement with our guess estimate. It should be noted that we are not looking for the exact location of possible change points; we only want to show that a careful look at suitable diagrams can give reasonable clues about the behavior of a time series. On the other hand, if we are not able to identify a cause–effect mechanism for step changes or monotonic patterns, we cannot infer about their future evolution making hydrological predictions beyond the period of record and the range of values used in the fitting stage difficult and sometimes unrealistic [110] and also theoretically incorrect (see Section 2).

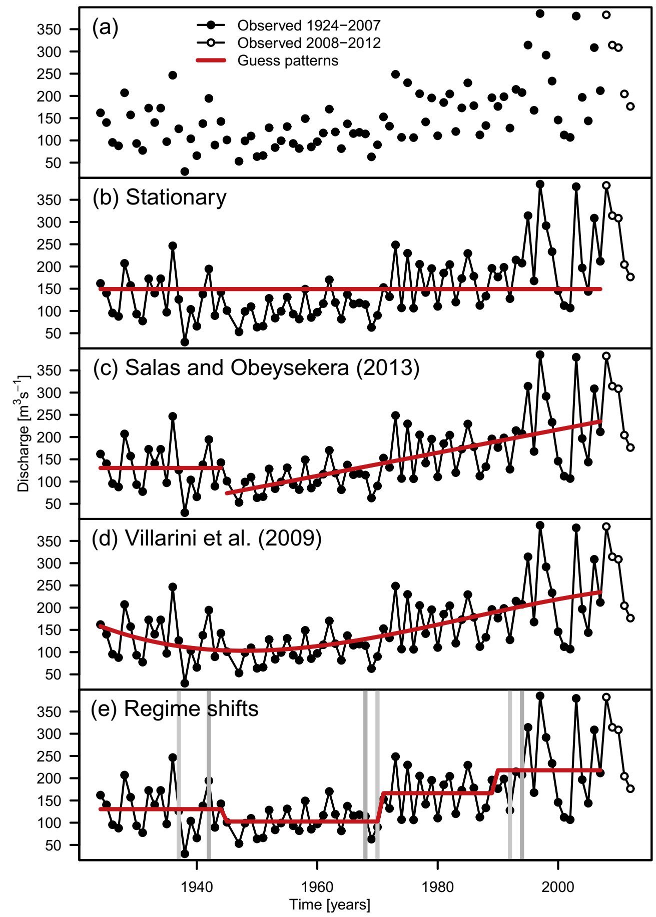


**Fig. 1.** Little Sugar Creek time series. (a) Disconnected symbol diagram of the observed peak flow time series. (b-d) Connected symbol diagrams and patterns adopted in different studies to model the evolution of the average values. (e) regime shift patterns. Vertical light-grey and grey lines denote the dates of change points in mean and variance, respectively, detected by Pettitt test.

These results highlight the practical implications of the remarks reported in Section 2. Adopting a nonstationary model requires that the relationships between parameters and predictable covariates are "deterministic" i.e. driven by clear physical mechanisms which allow for (reasonably) safe extrapolation. The patterns in Fig. 1 also highlight that statistical selection criteria cannot help in this context because the "best" model is just one among a usually limited set of competitors and can easily be outperformed by a new entry (e.g., a stepwise model or something else). Moreover, purely statistical criteria cannot reveal physical mechanisms, thus making the "optimal" model no more credible than other reasonable alternatives.

Compared with trend patterns, the regime shift pattern allows for some remarks: it is compatible with the interplay of different driving phenomena (e.g., large scale climate patterns, urbanization, mining, agricultural changes, channelization, etc.) that evolve and exhibit fluctuations over multiple time scales, better highlights the actual lack of knowledge of the cause–effect mechanisms when these mechanisms are not clearly identified/identifiable and cannot be described/quantified by well-defined "deterministic" laws, is compatible with stationary processes whose dynamics evolve over time scales larger than the observed period of record and are characterized by a variability (and thus, uncertainty) larger than *iid* or short memory processes. This allows for ascribing the magnitude of the fluctuation to the actual larger variability of the underlying process and thus maintaining the stationary option alive (see e.g., [43,46], for a discussion).

#### 6.1.2. Little Sugar Creek: modeling results

The implications of the above remarks are clearer when we compare the outcomes of the stationary and nonstationary frequency analyses. We compared four competitors: a GEV model with linearly time-varying parameters similar to the model of Salas and Obeysekera [78] (denoted as Model 1), a GEV model with parameters nonlinearly time-varying parameters similar to the model of Villarini et al. [110] (denoted as Model 2), a GEV model with parameters shifting according to the occurrence of the four regimes in the observed period (denoted as Model 3), and a benchmark stationary GEV model. The GEV model was chosen instead of Gumbel to check the impact of the shape parameter, and the model of Villarini et al. [110] was slightly modified using only the time $t$ as covariate and replacing the spline functions with suitable two- and three-order polynomials, which allow us to avoid the poor performance of nonparametric smoothing curves in the prediction stage (see the discussion in [110]). The comparison is made in terms of constant risk plots, stationary and nonstationary return levels and risk of failure. Nonstationary indices are calculated assuming that the design life spans 50 years from 2007 to 2056. All point estimates are complemented by CIs.

Fig. 2 shows the patterns of three different quantiles (corresponding to annual probabilities of exceedance 0.02, 0.01, 0.005) given by the four models both for record period (1924–2006) and design life period (2007–2056) (i.e. the so-called constant risk plots) along with the 95% CIs obtained by delta method. In all cases, the estimate of the shape parameter of the GEV model is close to zero confirming the results of Villarini et al. [110] and Salas and Obeysekera [78]. As expected, Model 1 (embedding a linear increasing trend) yields quantile values higher than the stationary model. As the probability of nonexceedance increases, also the uncertainty increases and the difference between stationary and nonstationary point estimates in the design life period becomes less and less statistically significant, with extensive overlap of CIs. Model 2 (with nonlinear time-varying parameters) yields 0.02 quantile values twice the maximum record at the end of the design life. Without questioning the physical meaning of these values, the widths of the CIs show that the model is not much informative. It should be noted that the symmetric CIs computed by delta method can give decreasing and also negative lower limits. This statistical artifact is a drawback of this method and highlights its physical inconsistency, being the approach based on over-approximating asymptotic properties. Therefore, the delta method may be used for quick preliminary checks rather than the communication of the actual uncertainty. Results of Model 3 are based on the hypothesis that the latter regime holds true for the entire

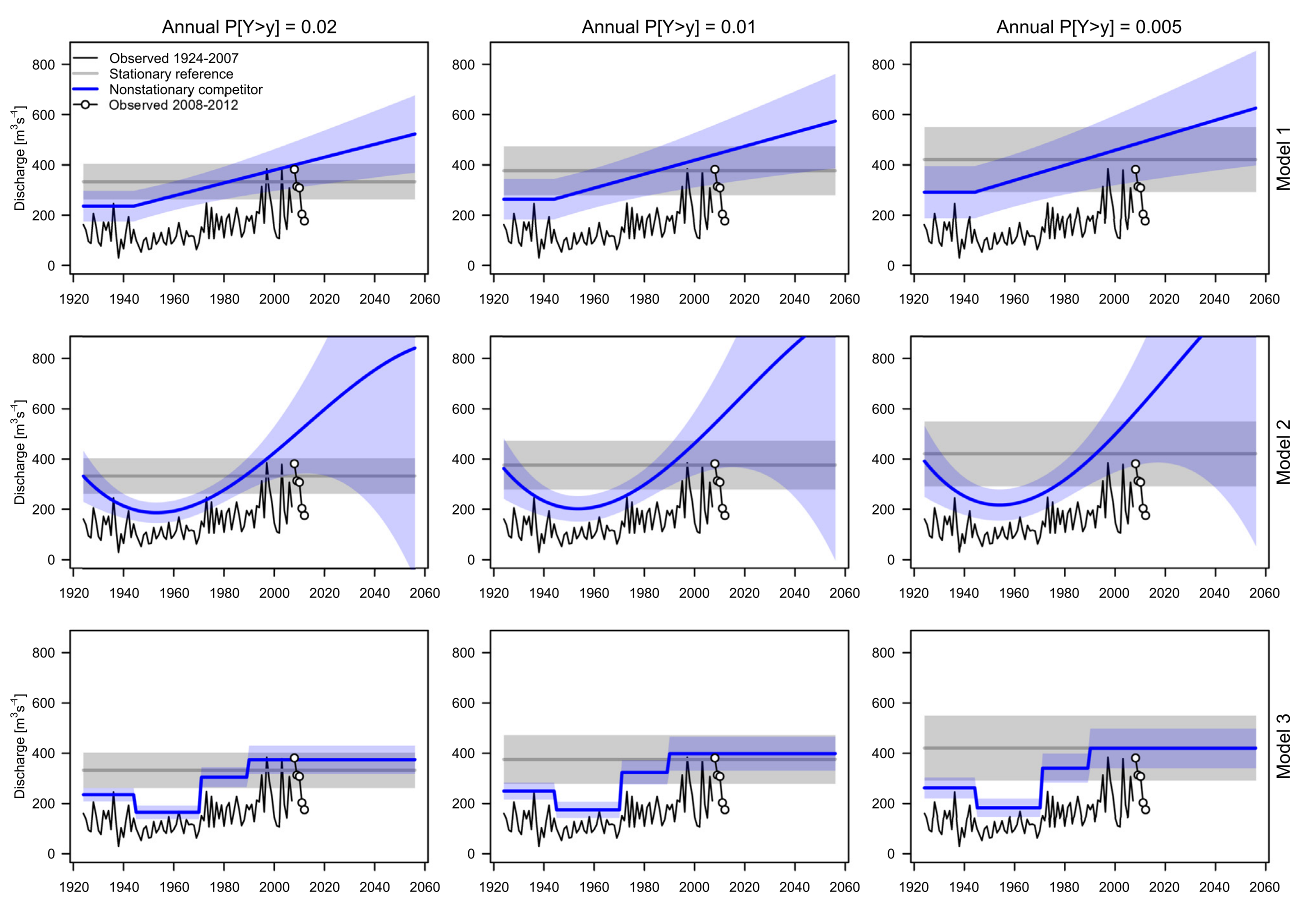


**Fig. 2.** Little Sugar Creek. Annual $p$ quantile plots for $p = \{0.02, 0.01, 0.005\}$ and the three models described in the text. Delta 95% CIs are shown.

design life. Even though this assumption is questionable, it is however as credible and defensible as Model 2 leading to 800 $m^3\ s^{-1}$ at the end of the next 50 years (which is about the double of the maximum value recorded in the past 83 years) or the linear increasing pattern of Model 1. Under this assumption, Model 3 and stationary GEV give similar results in the design life period. Model 3 shows CIs locally narrower than the stationary model. This behavior is also not surprising as the variance of the whole time series is globally higher than that of each sub-series within each hypothesized regime.

Fig. 3 shows the same point estimates reported in Fig. 2 complemented with bootstrap 95% CIs along with nonstationary 50-, 100- and 200-year return levels computed by Eqs. (6) and (7) (denoted as nsT1 and nsT2, respectively). Unlike delta CIs, bootstrap CIs are slightly asymmetric. Such asymmetry is not prominent as the upper tail of GEV models is approximately exponential (the GEV shape parameter is close to zero). The width of the CIs confirms that a large uncertainty characterizes the results obtained by Model 1 and Model 2. For instance, nonstationary 100-year return levels (nsT1 and nsT2 in the middle-top panel of Fig. 3) range approximately between 400 and 900 $m^3\ s^{-1}$ for Model 1, whereas the huge width of the CIs for every return level (nsT1 and nsT2 in the second row of panels in Fig. 3) highlights the lack of reliability and practical applicability of Model 2. This further stresses that the increase of model complexity is paid in terms of increase of uncertainty, and more complex models cannot replace information if this is not available. For an easier visualization, Fig. 4 summarizes stationary and nonstationary return levels for a range of $T$ values using return levels diagrams commonly used in stationary frequency analysis highlighting the closeness of the results given by nsT1 and nsT2 formulations. The comparison between stationary and nonstationary return levels in a unique diagram is possible as both are associated with (constant or average) *annual* probabilities of exceedance.

Even though all the four models are reasonable options and perform satisfactory in the fitting stage according to purely statistical criteria, the heterogeneity of the results in terms of design quantiles shows the difficulty to reliably assign a credible probability to such quantiles, thus making the choice between different models no much dissimilar from the choice of different safety factors, which risk-based approaches attempt to rationalize in a convenient way (e.g., [36]). Indeed, a nonstationary risk-based approach could not improve engineering design if the increased sources of uncertainty are not balanced by a suitable amount of additional information.

The above remarks are further supported by the values of risk of failure reported in Fig. 5. These diagrams look like those reported by Salas and Obeysekera [78] without CIs, highlighting the difference between stationary and nonstationary results. Is this difference really significant? CIs show that the uncertainty of the curves is so large that each curve falls almost completely within the CIs of the other.

Of course, the CIs' overlap does not mean that stationary and nonstationary curves (i.e. point estimates of return levels and risk of failure) are identical, but simply that the information used for calculations is not enough to reliably compute return levels and

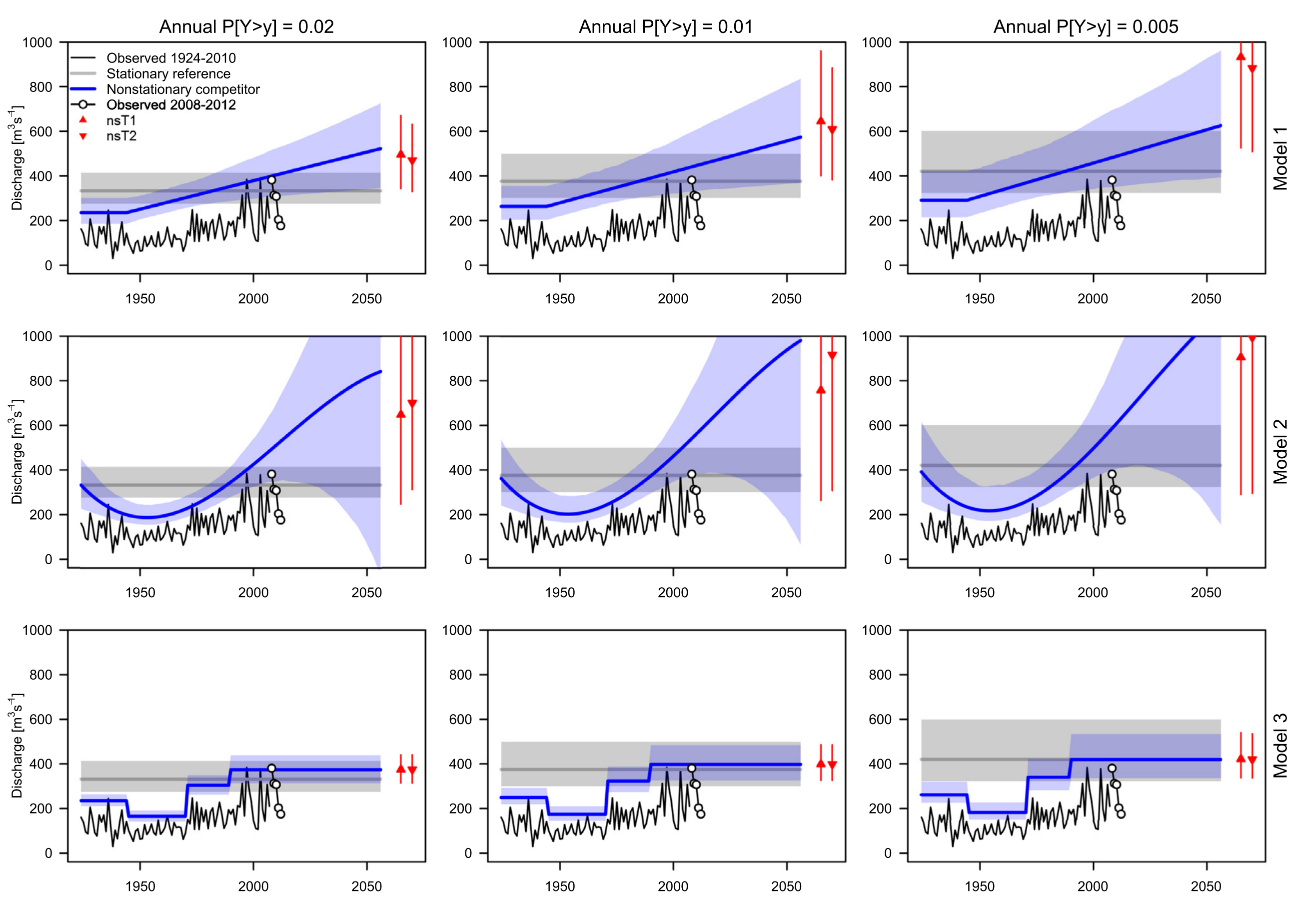


**Fig. 3.** As for Fig. 2, but showing bootstrap 95% CIs. nsT1 and nsT2 return levels with corresponding bootstrap 95% CIs are also shown. nsT1 and nsT2 return levels corresponding to $T = 50$ years are reported in the panels related to $p = 0.02$ (left column of plots). Similarly, nsT1 and nsT2 corresponding to $T = 100$ and $T = 200$ are shown in middle and right columns of plots related to $p = 0.01$ and $p = 0.005$.

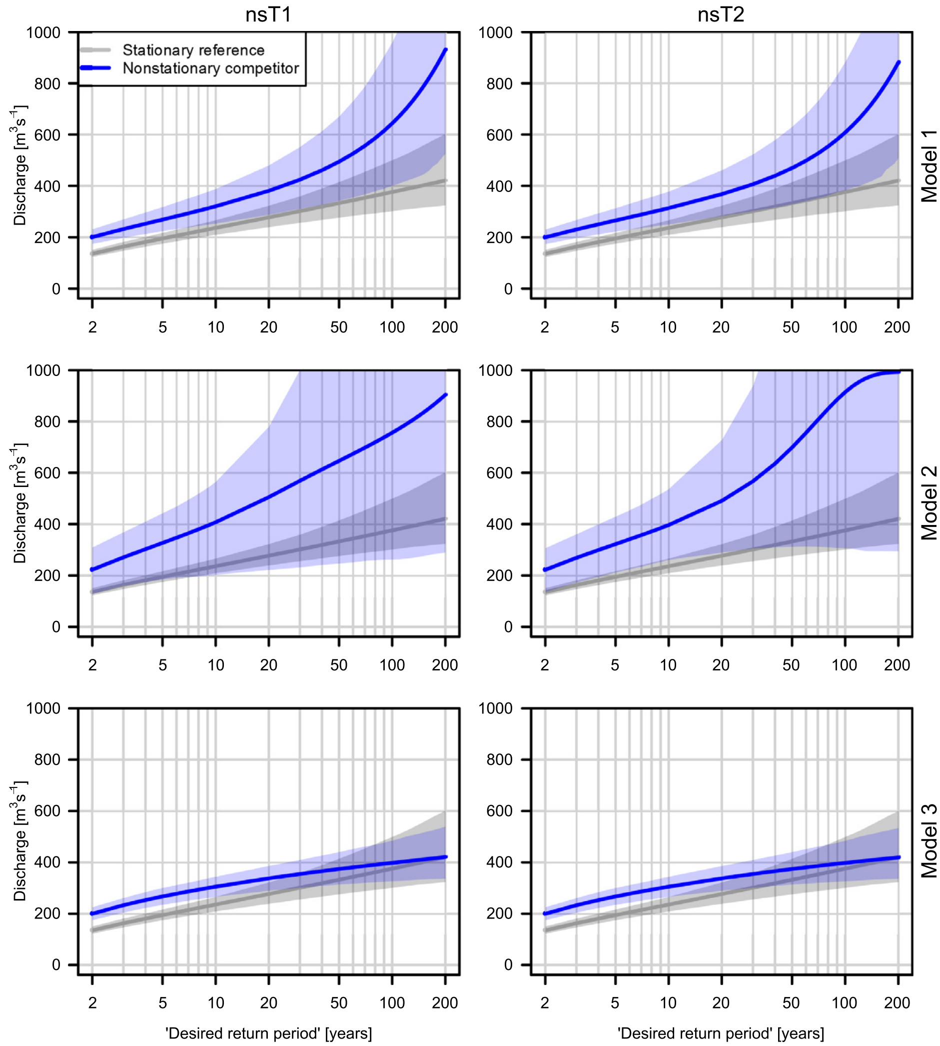


**Fig. 4.** Little Sugar Creek. Return level diagrams comparing stationary and nonstationary (nsT1 and nsT2) return levels complemented with bootstrap 95% CIs.

risk of failure and make conclusions about the real advantage of using a model more complex than the stationary. In other words, from a practical point of view, if a stationary model already provides sets (intervals) of critical values that largely comprise the values yielded by nonstationary models, and we bear in mind that these models are also characterized by a larger uncertainty (the model structure is unknown if it does not reflect physically-based predicable temporal patterns), there is no real justification to switch from simple but clear framework to something that is more complex and does not guarantee better future predictions.

In this respect, Villarini et al. [110] already recognized that even though simple characterizations of urbanization in terms of aggregate population and population growth are potentially useful for assessing changing flood frequency, they have limitations in describing the changing physical processes that affect flood frequency at a site. Fig. 12 in [110] clearly shows the sensitivity of nonstationary models and corresponding predictions to the unforeseen evolution of the temporal patterns of the variables when such patterns are not identified in a deductive manner and do not describe deterministic predictable dynamics.

Therefore, direct analysis of the impacts of urban development (such as alteration of infiltration and runoff production processes, expansions of the drainage network and changes of channel–floodplain interactions) is necessary for a comprehensive treatment of flood frequency in urbanizing watersheds. This point of view is coherent for instance with the recommendations of Klemeš [36] and is further emphasized accounting explicitly not only for the uncertainty of the covariates but also for that of the distribution family, functional relationships between model parameter and covariates as well as sampling uncertainty. In this respect, Wright et al. [115] showed how rainfall data driving a physics-based distributed hydrological model for a heavily urbanized watershed in Charlotte can provide a more reliable discharge frequency analysis.

Finally, it is worth mentioning that the benchmark stationary model describes *iid* data an does not account for the persistent-regime shift patterns shown in Fig. 1. The effect of such a persistence can be incorporated by hypothesizing the existence of an underlying (long-range) persistent process, thus resulting in stationary distributions with constant parameters generally characterized by a larger uncertainty (i.e. larger CIs). Alternatively, we can hypothesize the presence of a stochastic regime-switching process, which can be described by models such as hidden Markov models and Markov switching models, self-exciting threshold autoregressive models or similar (e.g., [39,99,100,104]), resulting in a mixture of distributions that alternate stochastically according to the transition probability from one regime (state) to another one. The latter kind of models exhibits locally persistent fluctua-

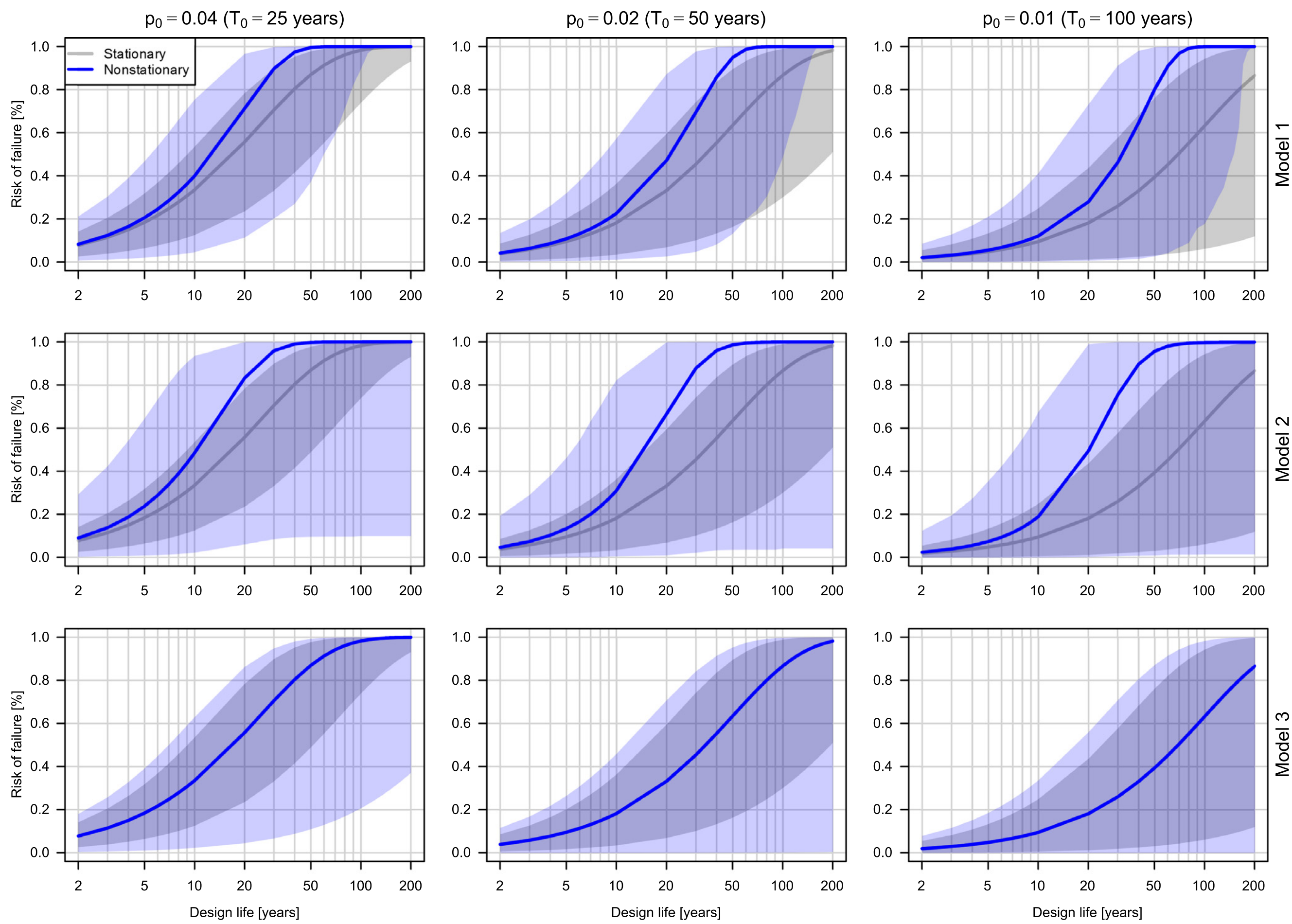


**Fig. 5.** Little Sugar Creek. Risk of failure diagrams for the design levels corresponding with $p_0 = \{0.04, 0.02, 0.01\}$ or more commonly $T_0 = \{25, 50, 100\}$ years, where $p_0$ and $T_0$ indicate the probability of exeedance and the "return period" ($T_0 = 1/p_0$) at the beginning of the design life period, respectively. Shaded areas define the pointwise bootstrap 95% CIs.

tions but is still globally stationary owing to the lack of a deterministic and predictable temporal evolution.

### *6.2. Analysis of the Red River of the North data*

#### *6.2.1. Red River of the North: data, preliminary remarks and EDA*

Cooley [9] used the annual peak flow measurements from 1942 to 2011 of the Red River of the North at Halstad (Minnesota, United States; USGS ID 05054500) to illustrate the application of the nsT1 and nsT2 return periods, return levels, and risk under nonstationarity by approaches that can be summarized as the risk plots and constant risk plots introduced later by Rootzén and Katz [76]. Cooley [9] selected the Red River of the North because of the recent flood activity and discussed in depth several aspects concerning the nature of the data and modeling strategy. In particular, Cooley [9] highlighted that (1) extrapolating the fitted trend into the future is always problematic as is taught in every introductory regression course; (2) information criteria (such as AIC) are not enough to select the "best" model and it is a good practice to include expert elicitation; (3) if the flow regime is driven by climate fluctuations, it is unlikely that it evolves linearly (a GEV with $\mu(t) = \mu_0 + \mu_1 t$ was assumed by Cooley [9] for the sake of illustration); and (4) it remains an open question if the recognized nonstationarity is due to selection bias or results from chance.

Addressing the above remarks in more depth was beyond the aim of the Cooley's methodological work; however, it can be done easily by collecting more information and performing some additional EDA. Additional information is provided for instance by the data recorded in other stations along the river, scientific literature and technical reports. A quick research in the USGS data base reveals that longer time series are available for the Red River of the North at Fargo (North Dakota; USGS ID 0554000; 1897–2012) and Grand Forks (North Dakota; USGS ID 05082500; 1882–2012) located upstream and downstream Halstad. Fig. 6 shows the three time series standardized by the drainage area for an easier comparison. It should be noted that data after 1940 are flagged with the qualification code 6 ("Discharge affected by regulation or diversion"; data accessed on February 19, 2014), and both upstream and downstream time series show historical records (occurred at the end of the 19th century) whose magnitude is comparable with the most recent events. It is therefore reasonable that similar events occurred also at Halstad, thus shedding a new light on the behavior of the peak flows at this site.

In light of the recent flood activity, the St. Paul District of the US Army Corps of Engineers prepared a feasibility study of alternative measures to reduce flood risk in the Fargo–Moorhead area [103]. The hydrological study is summarized in Appendix A of that report and by Mueller and Foley [62]. The most interesting aspects of that study are the involvement of a panel of experts and the homogenization of the data to account for the regulation operated after 1940. The outcome of the expert opinion elicitation (EOE) is a valuable example of how to proceed in such cases and how much discussion is needed to set up a reasonable modeling strategy. The EOE concluded that the peak flow records at Fargo exhibit nonstationarity

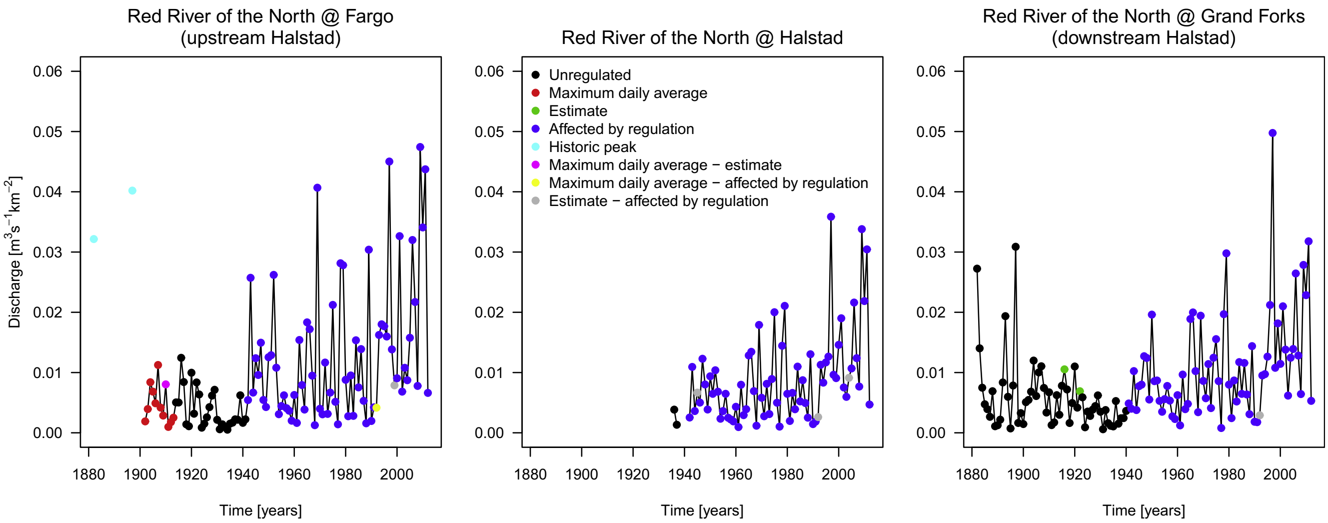


**Fig. 6.** Red River of the North at Fargo, Halstad, and Grand Forks. Connected symbol diagrams of the observed peak flow time series.

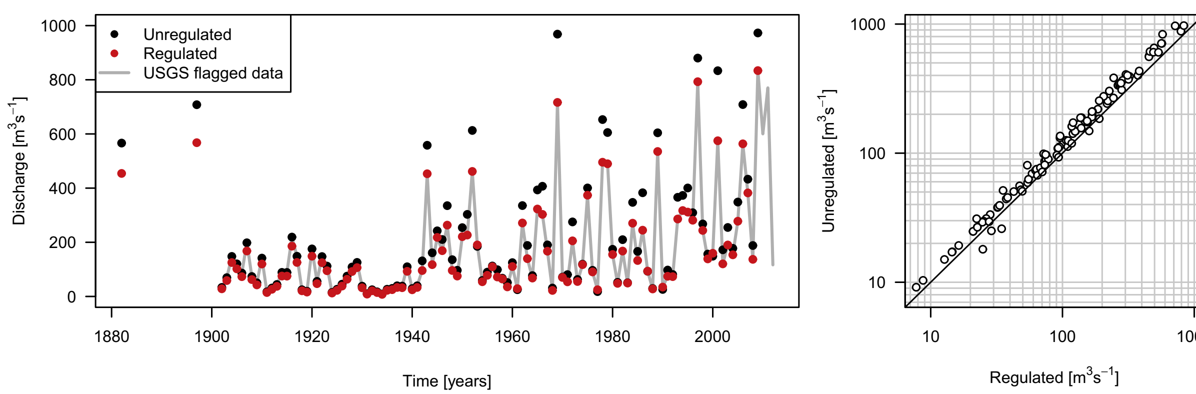


**Fig. 7.** Red River of the North at Fargo. Regulated and unregulated peak flow time series. The scaling effect of regulation is highlighted in the right side.

and this result had to be incorporated in the development of the flow frequency analysis. On the other hand, the procedure devised to rebuild the natural discharge values showed that the change point identified by Villarini et al. [109] in 1942 cannot be ascribed to regularization activities. Regulated and unregulated time series are shown in Fig. 7. It should be noted that the effect of the naturalization procedure is a simple rescaling of all values (see the constant shift in log–log scale reported in the right plot of Fig. 7). Thus, the final model adopted was a mixture of two distributions describing two different regimes defined as dry (prior 1941) and wet (after 1941) [62]. Such a distribution has form $F_Y = p_w F_{Y,w} + p_d F_{Y,d}$, where $p_w$ and $p_d$ are the probabilities to be in one of the two regimes, and $F_{Y,w}$ and $F_{Y,w}$ are the peak flow distributions in the two regimes. For the projection into the future 50 years (i.e. the design life period) it was assumed $p_w = 0.65$ and $p_d = 0.35$ (computed as the percentages of dry and wet years in the period of record), whereas $p_w = 0.80$ and $p_d = 0.20$ were used for the 25-year projection. The USACE study assessed the uncertainty of the design peak flow values as the differences between the quantiles yielded by $F_{Y,w}$, $F_{Y,d}$, and $F_Y$ with the different values of $p_w$ and $p_d$ and accounting or not for regulation; however, as is shown in the next section, this sensitivity analysis does not account for the most important source of uncertainty, i.e. the sampling uncertainty.

Todhunter [101] studied the peak flow data for the Grand Forks station highlighting the necessity of examining the assumptions required for the application of the conventional stationary frequency analysis based on LP3 distribution recommended by the US federal guidelines detailed in *Bulletin 17B* [112]. In particular, Todhunter [101] checked the presence of climatic trends, the temporal independence of the records, watershed changes and the flood generating mechanism. However, in spite of the author's suggestion of quantifying an effectively communicate the various uncertainties in flood risk assessment, the sampling uncertainty of some of the methods used in that study is not assessed, thus affecting the conclusions. For instance, Todhunter [101] used the values of the first three sample moments (mean, standard deviation, and coefficient of skewness) calculated incrementally for the period of record and their increments to assess the "stationarity of the time series" (see remarks in Section 2). The rationale of this method is that the sample moments should converge to population moments as the sample size increases. However, the convergence can be slower than expected. This is shown in Fig. 8 where the curves reported by Todhunter [101] are complemented by 1000 possible alternatives obtained by bootstrapping the observed time series (top plots) and simulating by an autoregressive AR(1) model (bottom plots). Bootstrap yields patterns corresponding to *iid* observations, whereas AR(1) allows us to account for the significant autocorrelation detected by Todhunter [101]. It is evident that the observed patterns (for mean and standard deviation) are not in the middle of the simulated bundle of curves but are compatible with the range of fluctuations corresponding with

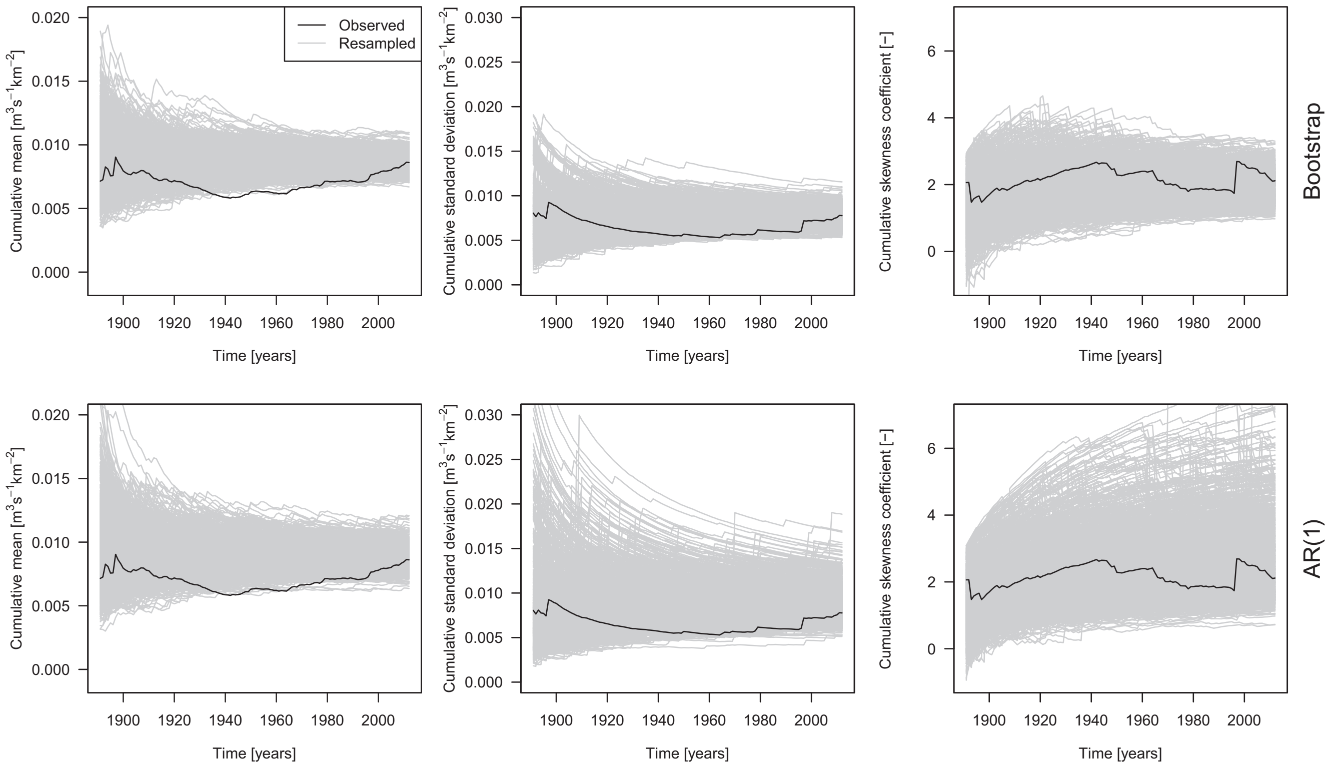


**Fig. 8.** Red River of the North at Grand Fork. Evolution of the first three sample moments of the peak flow records. Alternative patterns corresponding to 1000 bootstrapped series (top) and 1000 series simulated by a autoregressive AR(1) model (bottom) are also shown to highlight the slow convergence of the moments under temporal independence (bootstrap) and AR(1) dependence.

stationary (memoryless or short-range persistent) conditions. As stated by Todhunter [101], this diagnostic method is simple and widely used, but like any statistical method, it is also affected by the sampling uncertainty and often widely misused (if this uncertainty is not accounted for), leading to conclusions which are much less evident than expected and reported.

Referring to Todhunter [101] for the details of other analyses, we are interested in summarizing the lesson learned by the above EDA before presenting the modeling results:

- In real world applications, both stationary and nonstationary analyses cannot be reduced to fitting a small set of models implemented in some ready-to-use software as is often done in the literature and technical reports.
- Focusing on flood frequency analysis (but the discussion holds true for every type of frequency analysis, mutatis mutandis), a thorough EDA is always required and should include all the possible information from data recorded in nearby stations, other variables (such as rainfall, temperature and large scale climate indices), and expert elicitation. A thorough review of the scientific and technical literature is a must.
- EDA should rely on statistical methods well-suited for the specific scope and requires familiarity not only with the analysis methods (and their limits) but also with effective methods for uncertainty quantification, the analyzed variables, underlying physical processes, and engineering/management problems.
- In spite of the level of accuracy, a detailed EDA could yield no definite answers, but is always informative as it increases our awareness of our lack of knowledge, thus resulting in a more correct understanding of the actual reliability of the results.
- As far as the stationarity of the peak flows of the Red River of the North is concerned, no definite conclusions can be drawn. Indeed, the presence of memory recognized by Todhunter [101] reduces the effective size of the time series, increasing the uncertainty of every exploratory statistical index/test (e.g., [42]). Moreover, trends and change points can be purely stochastic and related to the persistence of the signal. On the other hand, Todhunter [101] highlighted the possible effects of the extensive and complex changes in land use/land cover occurred in the northern Great Plains (going beyond the regularization effects considered by the USACE study) along with the recognition of two distinct hydroclimatic regimes characterizing the period of record, and the different flood generating mechanisms (snowmelt and warm season rainfall). However, no analysis allows a definite conclusion on the stationarity. Indeed, being difficult to identify actual dynamics of the watershed hydrology and deterministic predictable evolution laws (based on the available information reviewed above), it is also difficult to conclude if the underlying process is stationary or nonstationary, and how it will evolve in the design life period (indeed, Cooley [9] and US Army Corps of Engineers [103] tested different alternative model structures, respectively).

*6.2.2. Red River of the North at Fargo: modeling results*

Given the lack of definite conclusions about a predictable evolution of peak flow dynamics of the Red River of the North, we adopt a pragmatic approach and model the peak flow data at Fargo by four different models, comparing their performance in terms of design quantiles and risk of failure. We analyze Fargo data as this allows for a comparison with the results reported by USACE [103, Appendix A-1c].

The four models are: LP3 distribution with linearly time-varying location parameter (denoted as Model 1) fitted on the time series of the unregulated peak flows provided by USACE [103, Appendix A-1c] starting from 1942 (i.e., for the wet regime); LP3 with location and scale parameters characterized by a regime shift in 1924, thus mimicking the two possible dry and wet regimes recognized by USACE EOE (see [103, Appendix A-1b]); this model splits in two sub-models in the prediction stage: (1) a sub-model assuming that the wet regime persists for the whole design life period (denoted as Model 2), and (2) a sub-model assuming that wet and dry states alternate with probabilities $p_w = 0.65$ and $p_d = 0.35$ (denoted as Model 3). The latter model does not account for the persistence of the two possible regimes which seems to occur in the period of record (see e.g., [47], for a discussion); however, it provides results corresponding to the mixed distribution used by USACE (see [103, Appendix A-1b]).

The constant risk plots in Fig. 9 show that the three nonstationary models yield quantiles (with annual probability of exceedance 0.02, 0.01, and 0.005) whose differences are always smaller than the width on the 95% CIs. Focusing on the design life period, which is the time window of actual interest for design purposes, the largest differences are between the stationary model and Model 1, the latter being however the less realistic among the competitors. Indeed, as mentioned in the previous section, there is no physical justification to assume such a pattern. Moreover, even though Model 1 would be a realistic option, confidence intervals of annual quantiles, nsT1 and nsT2 return levels are very large, encompassing values which are physically questionable. The width of the CIs highlights the actual lack of information as well as the marginal importance of focusing on the research of the "best" family (GEV, LP3, etc.) when the differences (in terms of point estimates of extreme quantiles) are often negligible compared with their uncertainty (see e.g., [38,80,82,93], among others, for a discussion).

Model 2 and 3 (with regime shifts) yield annual quantiles and nsT1 and nsT2 return levels not significantly different from the stationary LP3 fitted on the whole time series. Also in these cases the magnitude of the uncertainty dominates the differences of the point estimates. Stationary and nonstationary return levels for a range of $T \in [2, 200]$ years displayed in Fig. 10 confirm the above remarks. As the desired return period increases, Model 1 provides point estimates that diverge toward physically unrealistic values (more that ten times larger than the maximum record) complemented by CIs that highlight the unreliability of these estimates. On the other hand, the differences between Model 2 and 3 and stationary LP3 are negligible in terms of point estimate, being the difference only in the CIs of the most extreme return levels. The risk of failure diagrams (Fig. 11) convey the same message from a different point of view.

To further highlight the implication of performing a fair comparison accounting for uncertainty, we compared the nsT1 return levels given by Model 3, stationary LP3 return levels, and return levels calculated by USACE under several model settings [103, Appendix A-1c] (Fig. 12). Such settings were used to obtain a picture of the uncertainty; however, they describe only the model uncertainty and overlook the sampling uncertainty, which is the most important. Apart from the case corresponding to LP3 fitted to the dry period, all USACE scenarios fall within the confidence intervals of the stationary LP3 and/or Model 3. Therefore, none of

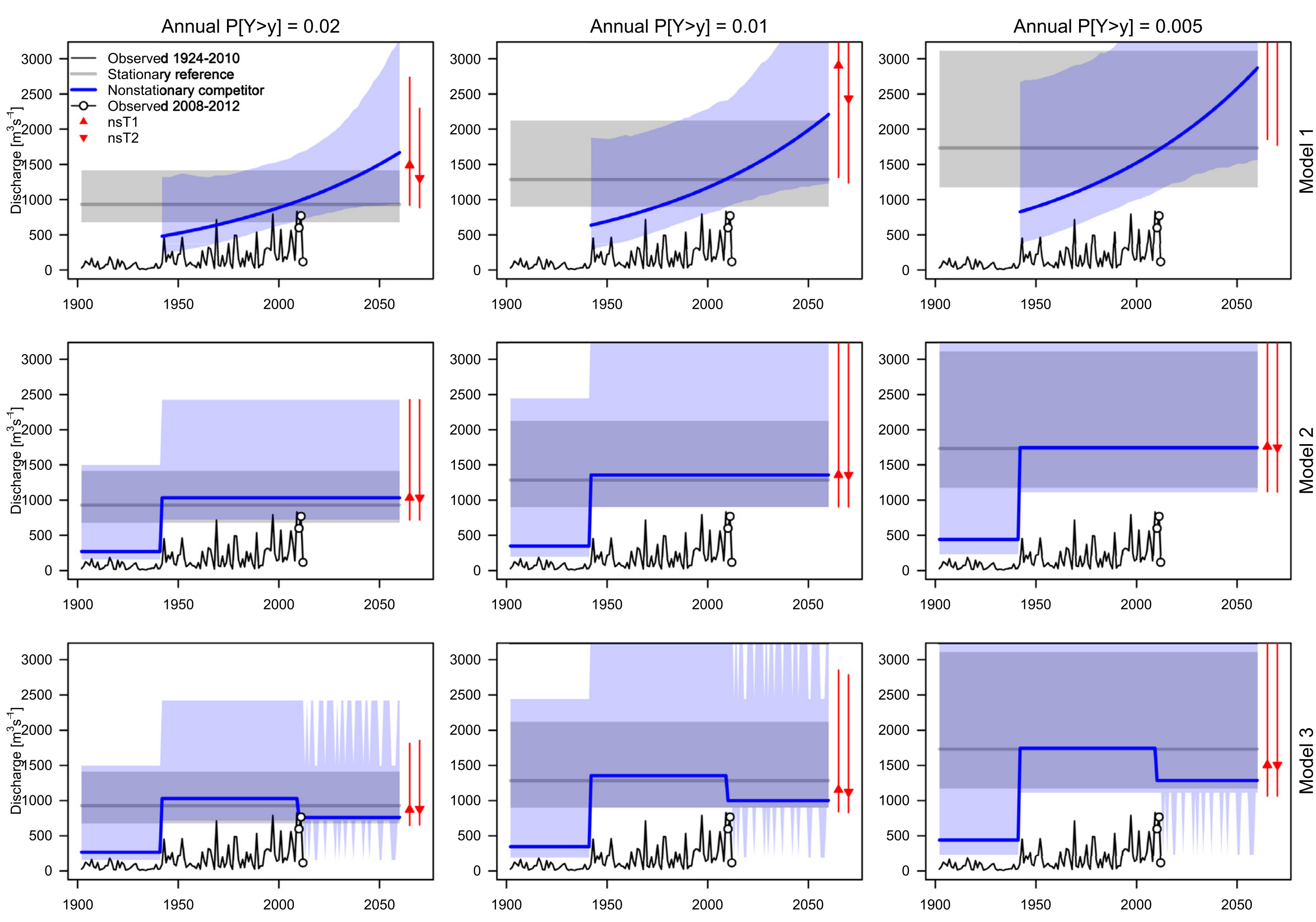


**Fig. 9.** Red River of the North at Fargo. Annual $p$ quantile plots with bootstrap 95% CIs for $p = \{0.02, 0.01, 0.005\}$ and the three models described in the text.

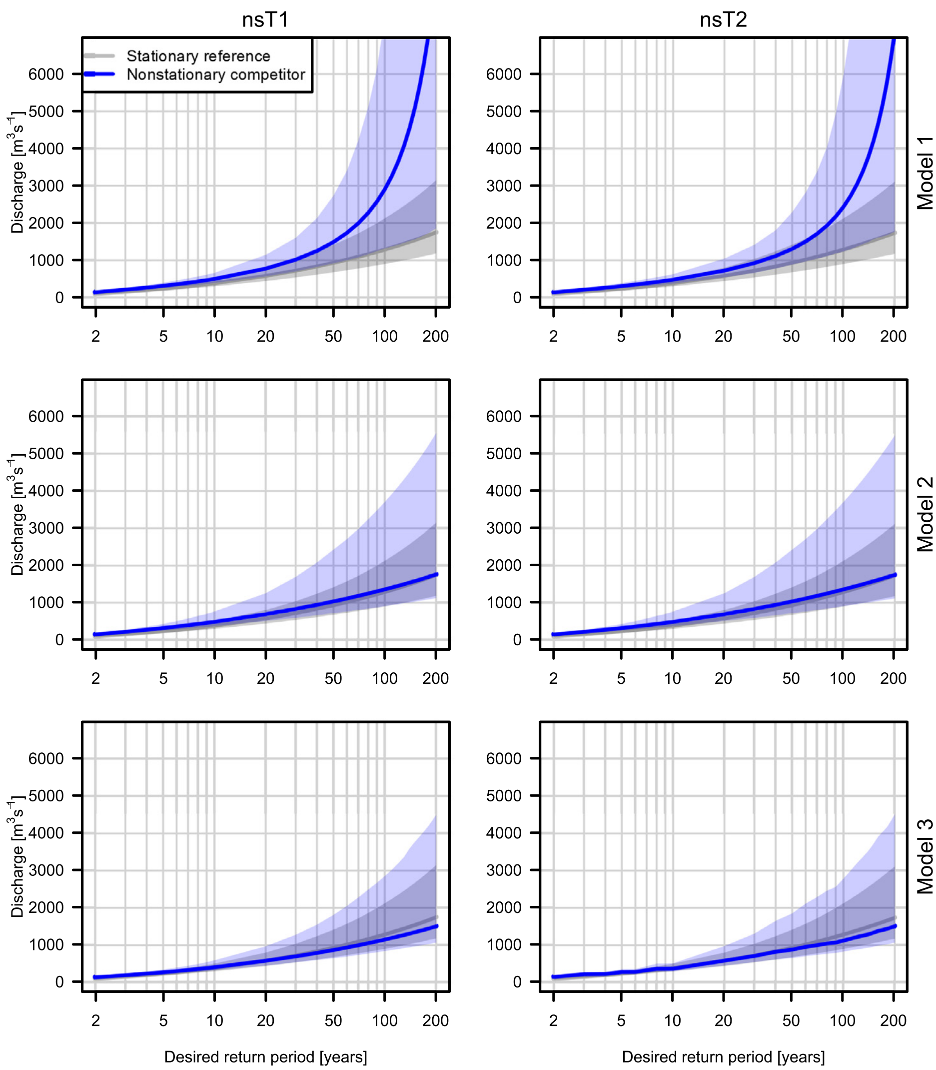


**Fig. 10.** Red River of the North at Fargo. Return level diagrams comparing stationary and nonstationary (nsT1 and nsT2) return levels complemented with bootstrap 95% CIs.

the point estimates is better than the other given the small differences of the less extreme return levels (with $T \leqslant 20$ years) and the large uncertainty of the most extreme (with $T \geqslant 50$ years).

Thus, complex nonstationary models do not provide better, more cautionary or reliable results than the stationary models because what really matters are not the point estimates but the interval estimates. The latter communicate the actual level of knowledge we have about the process and a suitable set of values to be used in effective sensitivity analyses involving additional criteria such as physical/design constraints, socio-economic variables and cost-benefit balance. Of course, the usefulness of nonstationary models cannot be assessed a priori; however, if the actual dynamics and corresponding predictable evolution laws generating the nonstationarity are not identified (i.e. if we are not sure that the process is really nonstationary), nonstationary models are unlikely to provide any improvement. In these cases (which are the most common in real-world applications), such models can be used at most for the sake of comparison, but cannot be considered definitely better than other competitors.

### *6.3. Analysis of the Assunpink Creek data*

The third case study relies on the peak flow measurements from 1924 to 2010 (here updated to 2011) of the Assunpink Creek at Trenton (New Jersey, United States; USGS ID 01464000) used by Obeysekera and Salas [65] to illustrate the applicability of nonstationary concepts (models and return periods) and different methods to build CIs. For the sake of space, we present only the main aspects of the analysis, referring to the Supplementary material for a detailed discussion.

For these data, Obeysekera and Salas [65] selected a GEV distribution with linearly time-varying location parameter and constant scale and shape parameters. However, a regime shift seems to occur in the first half of 1960s, when USGS started to flag data as affected by human activities. The presence of a possible abrupt change is confirmed by the ex post application of the Pettitt test. As for the Little Sugar Creek, a regime shift behavior seems to be more reasonable than slowly varying linear or monotonic patterns, which in turn cannot easily be justified. However, since these findings are still not sufficient to set up a nonstationary model if we are not able to reliably associate the regime shift with predictable mechanisms (anthropogenic or non anthropogenic), we slightly deepened the literature review, concluding that abrupt shift behavior seems to be a rule rather an exception for rivers in the conterminous United States (e.g., [45,48,52,113]). Moreover, the attribution of these changes is generally difficult, as no obvious hydraulic modifications were identified in many rivers. For instance, focusing on the Delaware River basin, Smith et al. [90]

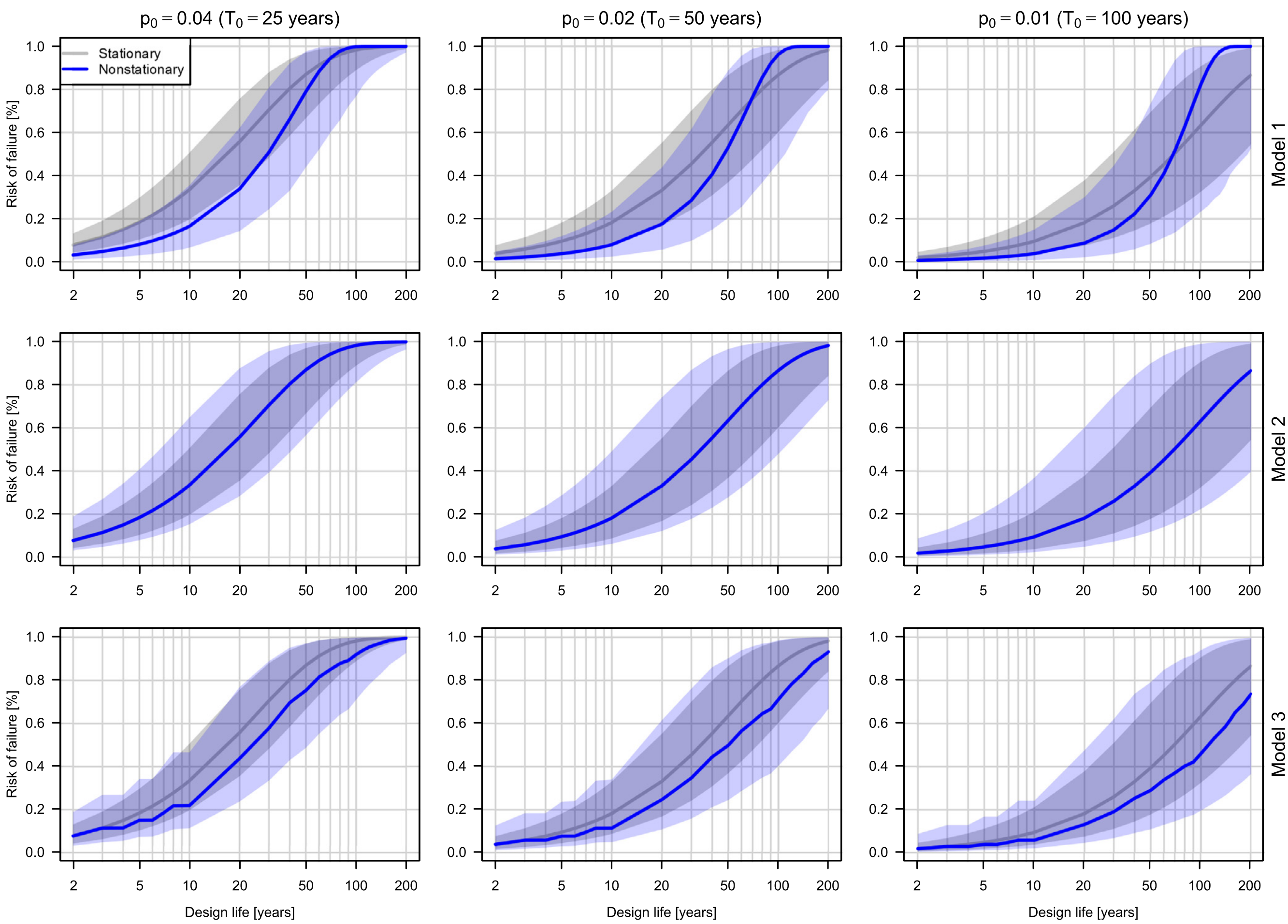


**Fig. 11.** Red River of the North at Fargo. Risk of failure diagrams for the design levels corresponding with $p_0 = \{0.04, 0.02, 0.01\}$ ($T_0 = 1/p_0 = \{25, 50, 100\}$ years), where $p_0$ and $T_0$ indicate the probability of exceedance and the "return period" at the beginning of the design life period, respectively. Shaded areas define the pointwise bootstrap 95% CIs.

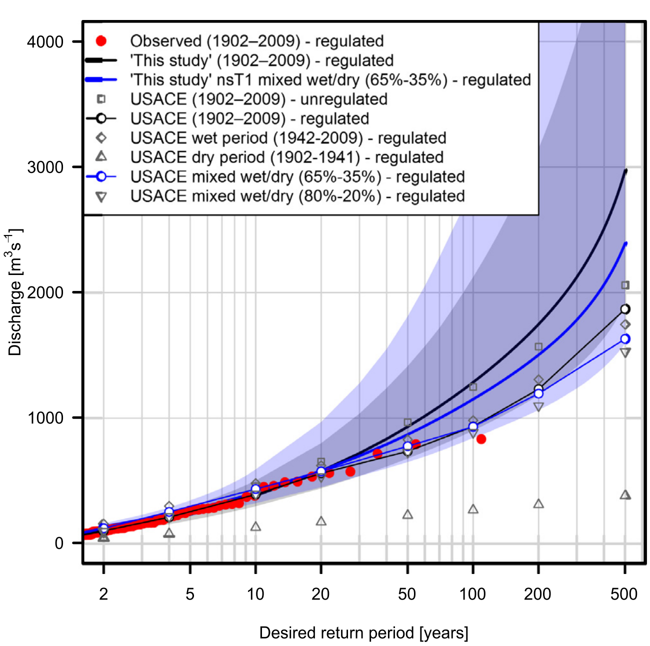


**Fig. 12.** Red River of the North at Fargo. Return level plots comparing stationary and nsT1 return levels (complemented with bootstrap 95% CIs) with several alternative scenarios reported by USACE ([103, Appendix A-1c]).

found regulation by dams and reservoirs plays an important role in determining change points, but the downstream effects of reservoirs on flood distributions are limited, whereas McCabe and Wolock [53,54] found that temporal patterns of flow summary statistics in the conterminous United States are only weakly associated with well-known climate indices, thus concluding that "*most of the temporal variability in flow is unpredictable in terms of relations to climate indices and infer that, for the most part, future changes in flow characteristics cannot be predicted by these indices*".

Therefore, the preliminary analyses clarify that studying a single time series is not enough to set up models and make inference in a nonstationary context. Techniques like the Mann–Kendall test can be used for preliminary examinations, but are not sufficient to make conclusions, whereas the collection of meta-data (existing reports and planning studies, historical information about land use, engineering interventions, infrastructures, etc.) is a fundamental (and often overlooked) step, especially in nonstationary frequency analyses, to identify the possible causes of changes and make an appropriate attribution. In this respect, as the above analysis does not explicitly identify predictable mechanisms causing the detected abrupt changes, we can only conclude in agreement with Lins [45] that "*Both sudden and gradual changes can be found in historical climatic and hydrologic records, and each type of change has distinct implications. A slow, gradual trend implies a pattern that is likely to continue into the future. A rapid step change typically indi-*

*cates a regime shift from one set of conditions to another, with the new conditions likely to persist until the next sudden shift occurs. What this may mean for future variations and changes in U.S. streamflow will only be revealed with time but, based on nearly a century of observations, we should expect our rivers and streams to continue to be characterized by both short- and long-term variations*".

Based on the EDA, four different models were fitted to the Assunpink Creek data, obtaining results similar to those of the other case studies (see Supplementary material for details), allowing us to conclude that, moving from illustrative applications to real-world applications, the principle of parsimony (or Occam's razor) should always be the first criterion to follow because "*a simple model with well-understood flaws may be preferable to a sophisticated model whose correspondence to reality is uncertain*" [46].

## 7. Conclusions

In this paper we have provided a critical overview of the concepts and methods used in nonstationary frequency analysis. We have explored all the aspects of the analysis to be considered in real-world applications going from the EDA to communication of the uncertainty of design quantiles. In view of the likely inclusion of nonstationary analyses in standard guidelines for practitioners, our aim was to highlight the critical points that the analyst should carefully keep in mind before approaching these tools and the related software in order to avoid misuses easily leading to incorrect conclusions. Therefore, we think that before adopting policies that make nonstationary analyses mandatory, it should be realized that advanced technology needs advanced training to be correctly applied. In this respect, the data analyses in Section 6 showed how large the uncertainty can be, given the increased model complexity introduced in the nonstationary frequency analyses, and thus how careful we should be in handling these techniques and their outputs.

Even though we used examples referring to flood frequency analysis, the discussion was kept as general as possible and is applicable to whatever hydrometeorological data. The results can therefore be summarized as a list of suggestions and warnings that should be followed when performing these types of studies:

1. A thorough EDA is fundamental in both stationary and nonstationary frequency analysis:
   - *Information should always be integrated*. The common length of time series of annual maxima (or slightly longer peak-over-threshold series) is usually insufficient to obtain reliable estimates of design quantiles. This concept is well known in stationary analysis (e.g., [37]), but is much more important in nonstationary analysis, where we have a double extrapolation: toward frequencies more extreme than the observed and into the future [76]. In this respect, the predictions given by nonstationary models become a sort of probabilistic forecast with the unavoidable additional uncertainty implied by the unknown evolution of the process dynamics. Therefore, we need additional information in form of similar data from nearby sites (i.e., "*borrowing strength across space*" [76]; e.g., flow data from upstream and downstream stations), exogenous hydroclimatic and socio-economic variables describing possible natural and anthropogenic forcings (e.g., rainfall, temperature, large scale climate indices, population, degree of urbanization, etc.) and meta-data (expert elicitation, historical documents, technical reports describing human activities/interventions, scientific literature describing the same data and/or the same area, etc.). This information should be integrated and carefully analyzed to have further insight on the dynamics of the process under study and draw credible conclusions on its future evolution. A further source of information can be the projections from climate models [76]; however, this type of information should be used with great care as it does not convey historic (observed) information, but future possible scenarios resulting from models; thus, we believe that such projections should be used for sensitivity analysis rather than as an exploratory tool. As discussed throughout the paper, the exploratory analysis can sometimes be inconclusive, but always informative about our actual degree of knowledge.
   - *Spatio-temporal dependence matters*. Both stationary and nonstationary frequency analyses as well as diagnostics and tests devised to check statistical properties, such as monotonic trends and abrupt changes, rely on the hypothesis that the data are spatially and temporally independent. When this hypothesis is not fulfilled, information is redundant and the effective sample size is smaller than the record length, thus increasing the uncertainty of summary statistics (e.g., [41]). The slow convergence of the moments discussed in Section 6.2.1 (see also Fig. 8) is an example. Therefore, spatio-temporal dependence should always be checked before any other type of analysis. If present, it should be accounted for. Several methods (with their own advantages and drawbacks) are available in the literature and should be consulted before proceeding to subsequent steps (see e.g., [14,22,27–29,118,119]).
   - *Uncertainty is the rule not the exception*. Dealing with short times series, uncertainty affects not only model outputs but also exploratory diagnostics. Provided that nonstationarity is actually a property characterizing stochastic processes [41], inferring the "nonstationarity" (as usually intended in the literature) of a hydrometeorological process from finite time series of observations might be not so easy because of the multiple interacting factors. This makes the recognition of a clear law of evolution very difficult if not impossible, and "*without such physical understanding, we have little basis for expecting a trend to continue, stop, or to reverse*" [93].

2. Modeling strategy.
   - *Occam's razor "Entia non sunt multiplicanda preater necessitatem"*. Modeling strategy should always start from the simplest informative approach according to the principle of parsimony. Examples often reported in the literature apply nonstationary distributions based on superficial EDA, assuming that climate fluctuations produce simple linear (or low-order polynomial) patterns or even more simply assuming linear evolutions for sake of simplicity. Moreover, the stationary distributions are usually discarded on the basis of some purely statistical criteria. This approach is surely effective for the sake of illustration, but cannot be recommended and should be avoided in real-world applications. As mentioned in the points above and throughout the paper, without a clear understanding of physical dynamics there is no sufficient justification to move from simpler to more complex models and conclude that nonstationary competitors perform better than stationary. Moreover, in real-world cases, the comparison should be done in term of the difference of the final output of interest (e.g., design values) accounting for the different sources

of uncertainty (model and sampling uncertainty, at least) and the physical meaning of such outputs. As is shown in the case studies, some models seemingly providing a good fit in the period of records can easily yield unrealistic predictions whose large uncertainty reflects the lack of reliability of the model structure and its inapplicability for design purposes. In this respect, cautionary statements are provided by some authors (e.g., [9,46,76,78,93]); however, they are overlooked in most of the literature, which appears more focused on the presentation of the models than on their utilization in practical applications.

– *Stationarity should remain the default assumption* [46]. Even if we have evidence for nonstationarity, stationary models should be used as benchmark for every more complex competitor. Moreover, the comparison with the stationary benchmark cannot be limited to the differences between point estimates of design values because point estimates returned by different models will be unavoidably different. A fair comparison should be based on the interval estimates accounting for the sampling uncertainty and highlighting the reliability of the estimates, and the physical meaning of point and interval estimates. If necessary, such an assessment should be complemented by other criteria such as technical/legislation constraints and socio-economic considerations (implemented in cost-benefit or multi-criteria analyses).
– *Uncertainty still rules!*. From the previous point, it follows that the assessment of the uncertainty is fundamental in the modeling stage as well as in the EDA. The simplest method for summarizing the uncertainty is the calculation of CIs of flood quantiles. As mentioned above, several methods are available with their own advantages and shortcomings. We suggest the nonparametric and parametric bootstrap (also known as Monte Carlo simulation) as the most practical option to obtain CIs reflecting the properties of models and estimators. The method is data-driven, does not resort to asymptotic assumptions, can be applied for every model and estimator, and can easily be tailored to meet specific requirements of the analysis if needed. Other methods should not be excluded and could be used for the sake of comparison (as shown in Section 6.1.2). They generally require a bit more advanced theoretical knowledge, which however is a necessary prerequisite to properly apply nonstationary models. As mentioned in the introduction, this discussion is limited to "frequentist" inference methods (e.g., method of the moments, maximum likelihood, L-moments), which are the most commonly used in everyday standard hydrological frequency analyses; however, Bayesian approaches are very promising especially to reduce the uncertainty by incorporating exogenous information. Examples of Bayesian nonstationary frequency analyses are provided for instance by Hundecha et al. [32], Ouarda and El-Adlouni [67] and Sun et al. [97]. An overview is given by Renard et al. [74], whereas a Bayesian strategy to incorporate relevant information in flood frequency analysis is discussed by Merz and Blöschl [58,59] and Viglione et al. [107]. Finally, freely available software for Bayesian nonstationary extreme value analyses is described by Cheng et al. [5], whereas other options are discussed by Gilleland et al. [25].

3. Design quantities: what and how should be communicated.

– *Results' visualization*. The diagrams of the annual quantiles corresponding to a given probability (i.e., the annual $p$ quantile diagrams or constant risk plots) are valuable tools to visualize the evolution of nonstationary distributions. In this respect, the diagram reported in this study attempt to refine the plots displayed by Cooley [9], Rootzén and Katz [76] or Silva et al. [88]. Such diagrams should show the data observed in the period of record, the stationary point estimates complemented by CIs, and nonstationary point and interval estimates. Visualizing this information allows for a fair comparison and effective illustration of the coherence between predictions (under stationary and nonstationary conditions) and observations as well as of the credibility of the extrapolations into the future design life period.
– *Return periods and levels*. In nonstationary models, since the annual $p$ quantiles change for each time step, they cannot be used as design quantities without additional constraints. In this respect, the minimax design life levels are an attempt to define such constraints focusing on the annual distribution yielding the most extreme $p$ quantile. The concept of return period defined in terms of number of exceedances in $T$ years is a simple alternative way of communicating the average annual probability $\bar{p}$, and the corresponding return levels are simply the quantiles associated with this average probability. On the other hand, the definition of expected waiting time yields similar results in several circumstances (see the case studies discussed above) and its calculation relies on the evolution of the nonstationary distributions also after the design life period [76]. Moreover, similar to the stationary definitions, both formulations tend to conceal the underlying time-varying probabilities under the symbol "$T$" and measurement units whose actual meaning are often misunderstood, leading to incorrect conclusions on the degree of rarity of an event. In respect, a clear understanding of the derivation of $T$ both under stationary and nonstationary conditions and the meaning of the underlying probabilities $p$ is fundamental for appropriate use and communication of the risk.
– *Risk of failure and design life levels*. The risk of failure (to be used in a verification setting) and the (formally identical) design life level (to be used in a design setting) are the most coherent choices as they describe the joint probability of exceedance $p_M$ in the design life period and the corresponding $p_M$ quantile. Such concepts can be generalized under temporal dependence, are probabilistically sound, provide a fair assessment of the actual probability to observe a critical event during the design life without requiring further information about the evolution of the nonstationary distribution beyond the design life period, and do not conceal the role of the underlying time-varying distributions. Moreover, such concepts are already known in stationary conditions, even though the compact formulation resulting from *iid* assumption [6, p. 383] does not reveal their true nature and meaning. Obviously, alternative summary statistics can be proposed; however such ad hoc design values should fulfill some requirements such as a clear link with the problem at hand (e.g. they should be tailored to meet operation rules, management policies, etc.), and a clear communication of the associated probabilities (univariate, joint or conditional) and how they are handled to obtain the final risk measure.

## Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/K013513/1 "Flood MEMORY: Multi–Event Modelling Of Risk & recoverY", and Willis Research Network. The authors wish to thank the two eponymous reviewers Dr. Richard W. Katz (National Center for Atmospheric Research, Colorado, USA) and Prof. Jose D. Salas (Colorado State University, Colorado, USA) for their detailed and insightful remarks and constructive criticisms that helped improve the quality of the original manuscript. The analyses were performed in `R` [73] by using the contributed package `ismev` [31]. The authors and maintainers of this software are gratefully acknowledged.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.advwatres.2014.12.013.

## References

[1] Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19(6):716–23. http://dx.doi.org/10.1109/TAC.1974.1100705.
[2] Bauer A, Feichtinger J, Steurer R. The governance of climate change adaptation in 10 OECD countries: challenges and approaches. J Environ Policy Plann 2012;14(3):279–304. http://dx.doi.org/10.1080/1523908X.2012.707406.
[3] Beguería S, Angulo-Martínez M, Vicente-Serrano SM, López-Moreno JI, El-Kenawy A. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006. Int J Climatol 2011;31(14):2102–14. http://dx.doi.org/10.1002/joc.2218.
[4] Beguería S, Vicente-Serrano SM. Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. J Appl Meteorol Climatol 2006;45(1):108–24. http://dx.doi.org/10.1175/JAM2324.1.
[5] Cheng L, AghaKouchak A, Gilleland E, Katz RW. Non-stationary extreme value analysis in a changing climate. Clim Change 2014:1–17. http://dx.doi.org/10.1007/s10584-014-1254-5.
[6] Chow VT, Maidment DR, Mays LW. Applied hydrology. New York: McGraw-Hill; 1988.
[7] Cleveland WS. The elements of graphing data. Murray Hill (NJ): AT&T Bell Laboratories; 1994.
[8] Coles S. An introduction to statistical modeling of extreme values. Springer series in statistics. London: Springer-Verlag; 2001.
[9] Cooley D. Return periods and return levels under climate change. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S, editors. Extremes in a changing climate. Water science and technology library, vol. 65. Netherlands: Springer; 2013. p. 97–114. http://dx.doi.org/10.1007/978-94-007-4479-0_4.
[10] Şen Z. Simple risk calculations in dependent hydrological series. Hydrol Sci J 1999;44(6):871–8. http://dx.doi.org/10.1080/02626669909492286.
[11] Şen Z, Altunkaynak A, Özger M. Autorun persistence of hydrologic design. J Hydrol Eng 2003;8(6):329–38. http://dx.doi.org/10.1061/(ASCE)1084-0699(2003)8:6(329).
[12] Dobson AJ, Barnett A. An introduction to generalized linear models. 3rd ed. Chapman & Hall/CRC; 2008.
[13] Douglas EM, Vogel RM, Kroll CN. Impact of streamflow persistence on hydrologic design. J Hydrol Eng 2002;7(3):220–7. http://dx.doi.org/10.1061/40569(2001)75.
[14] Douglas EM, Vogel RM, Kroll CN. Trends in floods and low flows in the United States: impact of spatial correlation. J Hydrol 2000;240(1–2):90–105. http://dx.doi.org/10.1016/S0022-1694(00)00336-X.
[15] Dubey SD. Compound gamma, beta and F distributions. Metrika 1970;16(1):27–31. http://dx.doi.org/10.1007/BF02613934.
[16] Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman and Hall; 1993.
[17] El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobée B. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resour Res 2007;43(3):WR004545. http://dx.doi.org/10.1029/2005WR004545.
[18] European Commission. Directive 2007/60/EC on the assessment and management of flood risks. Tech. rep., Brussels; 2007.
[19] European Commission. Adapting to climate change: towards a European framework for action, COM(2009) 147 final. White Paper. Tech. rep., Brussels; 2009.
[20] European Environment Agency. Climate change and water adaptation issues. Tech. rep., Kongens Nytorv 6, 1050 Copenhagen K, Denmark; 2013.
[21] Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. J R Stat Soc: Ser C (Appl Stat) 2001;50(2):201–20. http://dx.doi.org/10.1111/1467-9876.00229.
[22] Fatichi S, Barbosa SM, Caporali E, Silva ME. Deterministic versus stochastic trends: detection and challenges. J Geophys Res Atmos 2009;114(D18). http://dx.doi.org/10.1029/2009JD011960.
[23] Fernández B, Salas JD. Return period and risk of hydrologic events. I: Mathematical formulation. J Hydrol Eng 1999;4(4):297–307. http://dx.doi.org/10.1061/(ASCE)1084-0699(1999)4:4(297).
[24] Fernández B, Salas JD. Return period and risk of hydrologic events. II: Applications. J Hydrol Eng 1999;4(4):308–16. http://dx.doi.org/10.1061/(ASCE)1084-0699(1999)4:4(308).
[25] Gilleland E, Ribatet M, Stephenson AG. A software review for extreme value analysis. Extremes 2013;16(1):103–19. http://dx.doi.org/10.1007/s10687-012-0155-0.
[26] Griffis V, Stedinger JR. Incorporating climate change and variability into bulletin 17B LP3 model; 2007. p. 1–8 [chapter 68]. http://dx.doi.org/10.1061/40927(243)69.
[27] Hamed KH. Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. J Hydrol 2008;349(3–4):350–63. http://dx.doi.org/10.1016/j.jhydrol.2007.11.009.
[28] Hamed KH. Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. J Hydrol 2009;368(1–4):143–55. http://dx.doi.org/10.1016/j.jhydrol.2009.01.040.
[29] Hamed KH. The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data. Hydrol Sci J 2011;56(5):841–53. http://dx.doi.org/10.1080/02626667.2011.586948.
[30] Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman and Hall; 1990.
[31] Heffernan JE, Stephenson AG, Gilleland E. ismev: an introduction to statistical modeling of extreme values. R package version 1.39; 2012. <http://CRAN.R-project.org/package=ismev>.
[32] Hundecha Y, St-Hilaire A, Ouarda TBMJ, El Adlouni S, Gachon P. A nonstationary extreme value analysis for the assessment of changes in extreme annual wind speed over the gulf of St. Lawrence, Canada. J Appl Meteorol Climatol 2008;47(11):2745–59. http://dx.doi.org/10.1175/2008JAMC1665.1.
[33] Katz RW. Statistical methods for nonstationary extremes. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S, editors. Extremes in a changing climate. Water science and technology library, vol. 65. Netherlands: Springer; 2013. p. 15–37. http://dx.doi.org/10.1007/978-94-007-4479-0_2.
[34] Katz RW, Parlange MB, Naveau P. Statistics of extremes in hydrology. Adv Water Resour 2002;25(8–12):1287–304. http://dx.doi.org/10.1016/S0309-1708(02)00056-8.
[35] Khaliq MN, Ouarda TBMJ, Ondo J, Gachon P, Bobée B. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. J Hydrol 2006;329(3–4):534–52. http://dx.doi.org/10.1016/j.jhydrol.2006.03.004.
[36] Klemeš V. Dilettantism in hydrology: transition or destiny? Water Resour Res 1986;22(9S):177S–88S. http://dx.doi.org/10.1029/WR022i09Sp0177S.
[37] Klemeš V. Tall tales about tails of hydrological distributions. J Hydrol Eng 2000;5(3):227–31. http://dx.doi.org/10.1061/(ASCE)1084-0699(2000)5:3(227).
[38] Klemeš V. Risk analysis: the unbearable cleverness of bluffing. In: Bogardi JJ, Kundzewicz ZW, editors. Risk, reliability, uncertainty, and robustness of water resource systems. International hydrology series. Cambridge University Press; 2002. p. 22–9. http://dx.doi.org/10.1017/CBO9780511546006.003.
[39] Komorník J, Komorníková M, Mesiar R, Szökeová D, Szolgay J. Comparison of forecasting performance of nonlinear models of hydrological time series. Phys Chem Earth Parts A/B/C 2006;31(18):1127–45. http://dx.doi.org/10.1016/j.pce.2006.05.006.
[40] Koutsoyiannis D. Statistics of extremes and estimation of extreme rainfall. I: Theoretical investigation. Hydrol Sci J 2004;49(4):575–90. http://dx.doi.org/10.1623/hysj.49.4.575.54430.
[41] Koutsoyiannis D. Nonstationarity versus scaling in hydrology. J Hydrol 2006;324(1–4):239–54. http://dx.doi.org/10.1016/j.jhydrol.2005.09.022.
[42] Koutsoyiannis D. Hurst–Kolmogorov dynamics and uncertainty. J Am Water Resour Assoc 2011;47(3):481–95. http://dx.doi.org/10.1111/j.1752-1688.2011.00543.x.
[43] Koutsoyiannis D, Montanari A. Negligent killing of scientific concepts: the stationarity case. Hydrol Sci J 2014. http://dx.doi.org/10.1080/02626667.2014.959959.
[44] Kyselý J. A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. J Appl Meteorol Climatol 2008;47:3236–51. http://dx.doi.org/10.1175/2008JAMC1763.1.
[45] Lins HF. Streamflow trends in the United States: from the national streamflow information program. Fact sheet. US Department of the Interior, US Geological Survey; 2005. <http://pubs.usgs.gov/fs/2005/3017/pdf/FS2005_3017.pdf>.
[46] Lins HF, Cohn TA. Stationarity: wanted dead or alive? J Am Water Resour Assoc 2011;47(3):475–80. http://dx.doi.org/10.1111/j.1752-1688.2011.00542.x.
[47] Lins HF, Hirsch RM, Kiang J. Water-the nations fundamental climate issue: a white paper on the US Geological Survey role and capabilities. US Geological Survey Circular 1347. Tech. rep.; 2010. http://pubs.usgs.gov/circ/1347.

[48] Lins HF, Slack JR. Streamflow trends in the United States. Geophys Res Lett 1999;26(2):227–30. http://dx.doi.org/10.1029/1998GL900291.
[49] Loaiciga H, Mariño M. Recurrence interval of geophysical events. J Water Resour Plann. Manage. 1991;117(3):367–82. http://dx.doi.org/10.1061/(ASCE)0733-9496(1991)117:3(367).
[50] López J, Francés F. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrol Earth Syst Sci 2013;17(8):3189–203. http://dx.doi.org/10.5194/hess-17-3189-2013.
[51] Madsen H, Lawrence D, Lang M, Martinkova M, Kjeldsen TR. A review of applied methods in Europe for flood-frequency analysis in a changing environment. Tech. rep., the report has been prepared by Working Group 4 (Flood frequency estimation methods and environmental change). Freely available online – Official URL link provides full text; 2013. <http://nora.nerc.ac.uk/501751/>.
[52] McCabe GJ, Wolock DM. A step increase in streamflow in the conterminous United States. Geophys Res Lett 2002;29(24):38-1–4. http://dx.doi.org/10.1029/2002GL015999.
[53] McCabe GJ, Wolock DM. Variability common to global sea surface temperatures and runoff in the conterminous United States. J Hydrometeorol 2014;15(2):714–25. http://dx.doi.org/10.1175/JHM-D-13-097.1.
[54] McCabe GJ, Wolock DM. Spatial and temporal patterns in conterminous United States streamflow characteristics. Geophys Res Lett 2014;41(19):6889–97. http://dx.doi.org/10.1002/2014GL061980.
[55] McCullagh P, Nelder JA. Generalized linear model. 2nd ed. CRC Press; 1989.
[56] McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. J Am Stat Assoc 1997;92(437):162–70. http://dx.doi.org/10.1080/01621459.1997.10473613.
[57] McCulloch CE. Generalized linear mixed models. IMS; 2003.
[58] Merz R, Blöschl G. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. Water Resour Res 2008;44(8):W08432. http://dx.doi.org/10.1029/2007WR006744.
[59] Merz R, Blöschl G. Flood frequency hydrology: 2. Combining data evidence. Water Resour Res 2008;44(8):W08433. http://dx.doi.org/10.1029/2007WR006745.
[60] Mondal A, Mujumdar P. Return levels of hydrologic droughts under climate change. Adv Water Resour 2015;75:67–79. http://dx.doi.org/10.1016/j.advwatres.2014.11.005.
[61] Montanari A, Koutsoyiannis D. Modeling and mitigating natural hazards: Stationarity is immortal! Water Resour Res 2014;50(12):9748–56. http://dx.doi.org/10.1002/2014WR016092.
[62] Mueller C, Foley P. Discussion of uncertainty of the assumptions required for estimating the regulatory flood: Red River of the North by Paul E. Todhunter. J Hydrol Eng 2014;19(1):275–6. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000808).
[63] Nelder JA, Wedderburn RWM. Generalized linear models. J R Stat Soc Ser A 1972;135:370–84. http://dx.doi.org/10.2307/2344614.
[64] Nelsen RB. An introduction to copulas. 2nd ed. New York: Springer-Verlag; 2006.
[65] Obeysekera J, Salas JD. Quantifying the uncertainty of design floods under nonstationary conditions. J Hydrol Eng 2014;19(7):1438–46. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000931.
[66] Olsen JR, Lambert JH, Haimes YY. Risk of extreme events under nonstationary conditions. Risk Anal 1998;18(4):497–510. http://dx.doi.org/10.1111/j.1539-6924.1998.tb00364.x.
[67] Ouarda TBMJ, El-Adlouni S. Bayesian nonstationary frequency analysis of hydrological variables. J Am Water Resour Assoc 2011;47(3):496–505. http://dx.doi.org/10.1111/j.1752-1688.2011.00544.x.
[68] Papalexiou SM, Koutsoyiannis D. Battle of extreme value distributions: a global survey on extreme daily rainfall. Water Resour Res 2013;49(1):187–201. http://dx.doi.org/10.1029/2012WR012557.
[69] Papalexiou SM, Koutsoyiannis D, Makropoulos C. How extreme is extreme? An assessment of daily rainfall distribution tails. Hydrol Earth Syst Sci 2013;17(2):851–62. http://dx.doi.org/10.5194/hess-17-851-2013.
[70] Parey S, Hoang TTH, Dacunha-Castelle D. Different ways to compute temperature return levels in the climate change context. Environmetrics 2010;21(7–8):698–718. http://dx.doi.org/10.1002/env.1060.
[71] Parey S, Malek F, Laurent C, Dacunha-Castelle D. Trends and climate evolution: statistical approach for very high temperatures in France. Clim Change 2007;81(3–4):331–52. http://dx.doi.org/10.1007/s10584-006-9116-4.
[72] Pettitt AN. A non-parametric approach to the change-point problem. J R Stat Soc. Ser C (Appl Stat) 1979;28(2):126–35. http://dx.doi.org/10.2307/2346729.
[73] R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0; 2013. http://www.R-project.org/.
[74] Renard B, Sun X, Lang M. Bayesian methods for non-stationary extreme value analysis. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S, editors, Extremes in a changing climate, Water science and technology library, vol. 65; 2013. p. 39–95. http://dx.doi.org/10.1007/978-94-007-4479-0_3.
[75] Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. Appl Stat 2005;54(3):507–54. http://dx.doi.org/10.1111/j.1467-9876.2005.00510.x.
[76] Rootzén H, Katz RW. Design life level: quantifying risk in a changing climate. Water Resour Res 2013;49(9):5964–72. http://dx.doi.org/10.1002/wrcr.20425.
[77] Salas JD. Discussion – pragmatic approaches for water management under climate change uncertainty by Eugene Z. Stakhiv. J Am Water Resour Assoc 2013;49(2):475–8. http://dx.doi.org/10.1111/jawr.12026.
[78] Salas JD, Obeysekera J. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. J Hydrol Eng 2014;19(3):554–68. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000820.
[79] Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6(2):461–4. http://dx.doi.org/10.1214/aos/1176344136.
[80] Serinaldi F. Assessing the applicability of fractional order statistics for computing confidence intervals for extreme quantiles. J Hydrol 2009;376(3–4):528–41. http://dx.doi.org/10.1016/j.jhydrol.2009.07.065.
[81] Serinaldi F. Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. Energy Econ 2011;33(6):1216–26. http://dx.doi.org/10.1016/j.eneco.2011.05.001.
[82] Serinaldi F. An uncertain journey around the tails of multivariate hydrological distributions. Water Resour Res 2013;49(10):6527–47. http://dx.doi.org/10.1002/wrcr.20531.
[83] Serinaldi F. Dismissing return periods! Stochastic environmental research and risk assessment; 2014. http://dx.doi.org/10.1007/s00477-014-0916-1.
[84] Serinaldi F, Kilsby CG. A modular class of multisite monthly rainfall generators for water resource management and impact studies. J Hydrol 2012;464–465(0):528–40. http://dx.doi.org/10.1016/j.jhydrol.2012.07.043.
[85] Serinaldi F, Kilsby CG. Rainfall extremes: Toward reconciliation after the battle of distributions. Water Resour Res 2014;50(1):336–52. http://dx.doi.org/10.1002/2013WR014211.
[86] Serinaldi F, Kilsby CG. Simulating daily rainfall fields over large areas for collective risk estimation. J Hydrol 2014;512:285–302. http://dx.doi.org/10.1016/j.jhydrol.2014.02.043.
[87] Silva AT, Naghettini M, Portela MM. On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates. Stochastic Environ Res Risk Assess, submitted for publication.
[88] Silva AT, Portela MM, Naghettini M. On peaks-over-threshold modeling of floods with zero-inflated poisson arrivals under stationarity and nonstationarity. Stochastic Environ Res Risk Assess 2014;28(6):1587–99. http://dx.doi.org/10.1007/s00477-013-0813-z.
[89] Sivapalan M, Samuel JM. Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment. Hydrol Process 2009;23(11):1671–5. http://dx.doi.org/10.1002/hyp.7292.
[90] Smith JA, Baeck ML, Villarini G, Krajewski WF. The hydrology and hydrometeorology of flooding in the Delaware River basin. J Hydrometeorol 2010;11(4):841–59. http://dx.doi.org/10.1175/2010JHM1236.1.
[91] Stakhiv EZ. Pragmatic approaches for water management under climate change uncertainty. J Am Water Resour Assoc 2011;47(6):1183–96. http://dx.doi.org/10.1111/j.1752-1688.2011.00589.x.
[92] Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS) in R. J Stat Software 2007;23(7):1–46. http://www.jstatsoft.org/v23/i07.
[93] Stedinger JR, Griffis VW. Getting from here to where? Flood frequency analysis and climate. J Am Water Resour Assoc 2011;47(3):506–13. http://dx.doi.org/10.1111/j.1752-1688.2011.00545.x.
[94] Strupczewski W, Singh V, Mitosek H. Non-stationary approach to at-site flood frequency modelling III. Flood analysis of Polish rivers. J Hydrol 2001;248(1–4):152–67. http://dx.doi.org/10.1016/S0022-1694(01)00399-7.
[95] Strupczewski WG, Kaczmarek Z. Non-stationary approach to at-site flood frequency modelling II. Weighted least squares estimation. J Hydrol 2001;248(1–4):143–51. http://dx.doi.org/10.1016/S0022-1694(01)00398-5.
[96] Strupczewski WG, Singh VP, Feluch W. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. J Hydrol 2001;248(1–4):123–42. http://dx.doi.org/10.1016/S0022-1694(01)00397-3.
[97] Sun X, Thyer M, Renard B, Lang M. A general regional frequency analysis framework for quantifying local-scale climate effects: a case study of ENSO effects on Southeast Queensland rainfall. J Hydrol 2014;512:53–68. http://dx.doi.org/10.1016/j.jhydrol.2014.02.025.
[98] Sveinsson O, Salas JD, Boes D. Prediction of extreme events in hydrologic processes that exhibit abrupt shifting patterns. J Hydrol Eng 2005;10(4):315–26. http://dx.doi.org/10.1061/(ASCE)1084-0699(2005)10:4(315).
[99] Thyer M, Kuczera G. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach. J Hydrol 2003;275(1–2):12–26. http://dx.doi.org/10.1016/S0022-1694(02)00412-2.
[100] Thyer M, Kuczera G. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2: Real data analysis. J Hydrol 2003;275(1–2):27–48. http://dx.doi.org/10.1016/S0022-1694(02)00411-0.
[101] Todhunter P. Uncertainty of the assumptions required for estimating the regulatory flood: Red River of the North. J Hydrol Eng 2012;17(9):1011–20. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000560.
[102] Tramblay Y, Neppel L, Carreau J, Najib K. Non-stationary frequency analysis of heavy rainfall events in southern France. Hydrol Sci J 2013;58(2):280–94. http://dx.doi.org/10.1080/02626667.2012.754988.
[103] US Army Corps of Engineers (USACE). Final feasibility report and environmental impact statement: Fargo–Moorhead metropolitan area flood

risk management. Tech. rep., St. Paul District, St. Paul, MN; 2011. <http://www.fmdiversion.com/eisconsultantsreport.php>.
[104] Valent P, Howden N, Szolgay J, Komorníková M. Analysis of nitrate concentrations using nonlinear time series models. J Hydrol Hydromech 2011;59(3):157–70. http://dx.doi.org/10.2478/v10098-011-0013-9.
[105] van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modeling growth reference curves. Stat Med 2001;20:1259–77. http://dx.doi.org/10.1002/sim.746.
[106] van Montfort MAJ, van Putten B. A comment on modelling extremes: links between multi-component extreme value and general extreme value distributions. J Hydrol New Zealand 2002;41(2):197–202. http://www.hydrologynz.co.nz/journal.php?article_id=69.
[107] Viglione A, Merz R, Salinas JL, Blöschl G. Flood frequency hydrology: 3. A Bayesian analysis. Water Resour Res 2013;49(2):675–92. http://dx.doi.org/10.1029/2011WR010782.
[108] Villarini G, Serinaldi F. Development of statistical models for at-site probabilistic seasonal rainfall forecast. Int J Climatol 2012;32(14):2197–212. http://dx.doi.org/10.1002/joc.3393.
[109] Villarini G, Serinaldi F, Smith JA, Krajewski WF. On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resour Res 2009;45(8):WR007645. http://dx.doi.org/10.1029/2008WR007645.
[110] Villarini G, Smith JA, Serinaldi F, Bales J, Bates PD, Krajewski WF. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Adv Water Resour 2009;32(8):1255–66. http://dx.doi.org/10.1016/j.advwatres.2009.05.003.
[111] Vogel RM, Yaindl C, Walter M. Nonstationarity: flood magnification and recurrence reduction factors in the United States. J Am Water Resour Assoc 2011;47(3):464–74. http://dx.doi.org/10.1111/j.1752-1688.2011.00541.x.
[112] Water Resources Council. "Guidelines for determining flood flow frequency" Bulletin 17B. Tech. rep., Washington, DC; 1982. http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.
[113] Watson KM, Schopp RD. Methodology for estimation of flood magnitude and frequency for New Jersey streams: US Geological Survey Scientific Investigations Report 2009v5167. US Department of the Interior, US Geological Survey, Reston, Virginia; 2009. http://pubs.usgs.gov/sir/2009/5167/pdf/sir2009-5167.pdf.
[114] Wilby RL, Keenan R. Adapting to flood risk under climate change. Prog Phys Geogr 2012;36(3):348–78. http://dx.doi.org/10.1177/0309133312438908.
[115] Wright DB, Smith JA, Baeck ML. Flood frequency analysis using radar rainfall fields and stochastic storm transposition. Water Resour Res 2014;50(2):1592–615. http://dx.doi.org/10.1002/2013WR014224.
[116] Yee TW, Stephenson AG. Vector generalized linear and additive extreme value models. Extremes 2007;10:1–19. http://dx.doi.org/10.1007/s10687-007-0032-4.
[117] Yee TW, Wild CJ. Vector generalized additive models. J R Stat Soc Ser B 1996;58(3):481–93. http://www.jstor.org/stable/2345888.
[118] Yue S, Kundzewicz ZW, Wang L. Detection of changes. In: Kundzewicz ZW, editor. Changes in flood risk in Europe. IAHS special Publication 10. Balkema, Wallingford, UK: IAHS Press and CRC Press; 2012. p. 22–9.
[119] Yue S, Pilon P, Phinney B. Canadian streamflow trend detection: impacts of serial and cross–correlation. Hydrol Sci J 2003;48(1):51–63. http://dx.doi.org/10.1623/hysj.48.1.51.43478.