

A practice-oriented framework for stationary and nonstationary flood frequency analysis

Cuauhtémoc Tonatiuh Vidrio-Sahagún, Jake Ruschkowski, Jianxun He^{*}, Alain Pietroniro

Civil Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, Canada, T2N 1N4



ARTICLE INFO

Keywords:

Flood frequency analysis framework
Stationarity
Nonstationarity
Uncertainty
Software.

ABSTRACT

In flood frequency analysis (FFA), choices of distribution and methods can hinder the reproducibility of results. Besides, changes in climate, land use/cover, and water management can induce nonstationarity. Frameworks to select between stationary FFA (S-FFA) and nonstationary FFA (NS-FFA) are lacking, and NS-FFA tools are limited. Therefore, this paper introduces a systematic and software-supported framework enabling repeatable workflows for both S-FFA and NS-FFA. The framework has three modules to a) process flood series for exploratory data analysis (EDA) and NS-FFA model determination (if needed), b) select the S-FFA or NS-FFA approach underpinned by the EDA, and c) perform FFA including model determination, parameter estimation, uncertainty quantification, and model performance assessment. The framework incorporates various distributions, methods, and metrics, and recent advancements in NS-FFA for model determination and uncertainty quantification and allows for the modeller's intervention while ensuring reproducibility. The software is freely available to the public.

1. Introduction

Floods are one of the most common, dangerous, and costliest natural hazards (Razavi et al., 2020) and have impacted 2.8 billion people across the world from 1980 to 2009, from which approximately 540,000 people lost their lives (Doocy et al., 2013). Flood hazards are exacerbated globally, as the expansion of urban areas in flood-prone regions has increased since 1985, particularly after 2000 (Andreadis et al., 2022). In this context, assessing the recurrence of floods is crucial for designing and evaluating hydraulic structures, determining flood hazard and associated risk, and managing water resources. Hence, flood frequency analysis (FFA), which evaluates the return period of extreme flows by fitting a probability distribution to extreme-event observations (e.g., annual maxima), is a fundamental practice in water resources engineering and management.

Broadly speaking, the FFA can be performed in two general approaches regarding the behavior of floods over time: stationary FFA (S-FFA) and nonstationary FFA (NS-FFA). In S-FFA, observed floods are assumed independent and identically distributed (i.i.d.), and thus, realizations from a stationary process. The stationary assumption implies that statistical characteristics of extremes remain constant over time. However, this assumption might not hold due to climate change,

changes in the land use/cover of the watershed, and/or changes in water management (e.g., dam regulation) (Blöschl et al., 2019; Blum et al., 2020; Milly et al., 2008; Slater et al., 2021b; Yin et al., 2018). All these human-induced changes may lead to temporal variations in the statistics of extremes and consequently to nonstationarity. Under such circumstances, nonstationarity has been handled in several typical ways in practice (Mishra et al., 2022), including (a) ignoring the nonstationarity and using S-FFA, (b) employing a recent subsample of hydrological records in S-FFA to reflect more up-to-date conditions and bypass explicitly modeling the nonstationarity, (c) adding a safety factor to the S-FFA frequency estimates, or (d) conducting NS-FFA. The NS-FFA explicitly addresses the nonstationarity by using a time-varying distribution to depict changes in the floods over time (e.g., Cunderlik and Burn, 2003; El Adlouni et al., 2007; Prosdocimi et al., 2015; Salas et al., 2018; Villarini et al., 2009b). However, nonstationarity is often omitted from the analysis in practice due to a general lack of guidelines on when and how to conduct NS-FFA (Salas et al., 2018; Slater et al., 2021a; Villarini et al., 2018). Neglecting the nonstationarity in FFA may lead to substantial over- or underestimation of quantiles (O'Brien and Burn, 2014) as well as increased uncertainty and/or degradation of the analysis accuracy (Vidrio-Sahagún et al., 2021).

Furthermore, in the context of S-FFA, a variety of probability

* Corresponding author.

E-mail address: jianhe@ucalgary.ca (J. He).

distributions and methods for estimating distribution parameters have been employed. However, the selections of the distribution and the parameter estimator are often subjective in practice. While selecting an appropriate distribution is necessary, the distribution parameter estimation is also crucial, as it involves approximating the true (yet unknown) distribution based on the available data. Moreover, the distribution parameter estimations rely on historical observations, which are subject to natural variability and consequently produce uncertainty. Thus, it is indispensable to complement the point estimates with an uncertainty assessment that accounts for sampling variability. The uncertainty is often estimated as confidence intervals. Despite being commonly overlooked by practitioners, uncertainty quantification is fundamental in evaluating analysis reliability and has implications in infrastructure design, hazard assessment, and risk mitigation. Similar challenges arise in NS-FFA as well. Therefore, a systematic framework that aids in the selection of either the S-FFA or NS-FFA approach and considers diverse metrics and methods for distribution selection, distribution parameter estimation, and uncertainty quantification, is indeed desired. Such a framework would be beneficial in avoiding inconsistent results and consequently enhancing the reproducibility of results.

2. Background

When conducting S-FFA, the probability distribution has been either arbitrarily chosen or selected from a pool of candidates through statistical tests (e.g., Kolmogorov-Smirnov, modified Anderson-Darling, and likelihood ratio tests), goodness-of-fit metrics (e.g., root mean squared error and Akaike and Bayesian Information Criteria (*AIC* and *BIC*)), and the statistical properties of the sample and the theoretical distributions (e.g., methods of moments and L-moments), among others. When selecting the distribution from several candidates, the choice of candidates themselves can vary substantially from one modeller to another, and personal biases or considerations, such as familiarity with certain distribution(s), may influence choices. Similarly, the distribution parameters can be estimated by different methods, including the methods of moments, L-moments, probability-weighted moments, maximum likelihood (ML), generalized maximum likelihood (GML), least squares, and Bayesian approaches with Markov Chain Monte Carlo (MCMC) samplers (Hosking, 1990; Iliopoulou et al., 2018; Kobierska et al., 2018; Martins and Stedinger, 2000; Reis and Stedinger, 2005). The L-moments method is often preferred in S-FFA because it is robust and more unbiased, especially when dealing with extreme observations and/or relatively short samples (Hosking, 1990; Sankarasubramanian and Srinivasan, 1999). In the context of uncertainty quantification, popular frequentist techniques reported in the literature include the bootstrap, delta, and profile likelihood (PL) methods (Gilleland and Katz, 2016; Kjeldsen et al., 2014b; Salas et al., 2018; Wang et al., 2017). Among these methods, the PL method is often deemed more reliable than and theoretically superior to other methods (Bolívar-Cimé et al., 2015; Coles, 2001; Obeysekera and Salas, 2014), although the bootstrap method is arguably the most common choice.

If nonstationarity is suspected, the stationary assumption can be assessed with the aid of exploratory data analysis (EDA) on the time series, which can detect the presence of change point(s) and temporal trend(s) and identify the trend types (deterministic or stochastic). A change point indicates an abrupt shift or a temporal pattern switch (e.g., the beginning of a trend) in the time series. In the context of FFA, change points in a flood series imply inhomogeneous periods that require piecewise analysis, as a single FFA model may not be representative for all time periods. Change points in hydrological time series are commonly detected using the Mann-Whitney-Pettitt test (commonly known as Pettitt) and the Mann-Kendall-Sneyers (MKS) test (also known as sequential Mann-Kendall) (François et al., 2019; Li et al., 2015; Slater et al., 2021a; Vidrio-Sahagún and He, 2022c; Villarini et al., 2009a; Zhang et al., 2014). A temporal trend suggests a continuous change over

time in the floods. The trend can be either deterministic or stochastic, although this distinction is often overlooked. A deterministic trend refers to a non-random and continuous change driven by external factors, whereas a stochastic trend is a spurious trend resulting from random variability and internal dynamics (Franzke et al., 2020). Therefore, the presence of a deterministic trend would support the need for NS-FFA. The non-parametric Mann-Kendall (MK) test is typically used to detect trends (e.g., François et al., 2019; Khalil et al., 2006; Mondal and Mujumdar, 2016; Slater et al., 2021a) due to its higher power compared to parametric tests (Yue and Pilon, 2004). The block bootstrap MK (BB-MK) test, which also accounts for the impact of serial correlation (Bayazit, 2015; Khalil et al., 2009; Sonali and Nagesh Kumar, 2013), is relatively less popular. Additionally, the parametric White test is used to assess the heteroscedasticity of flood series and investigate nonstationary signatures in their variability (Ragno et al., 2019). The non-parametric Mann-Kendall test coupled with the moving window method is also used to detect trends in the variability of flood series (e.g., in the standard deviation) (Cunderlik and Burn, 2003; Vidrio-Sahagún et al., 2021; Vidrio-Sahagún and He, 2022c), as it can avoid possible flaws of parametric tests due to model misspecification. To distinguish between the trend types, the Phillips-Perron (PP) and the Kwiatkowski-Phillips-Schmidt-Shim (KPSS) unit root tests can be employed in conjunction (Barbosa, 2011; Fatichi et al., 2009; Franzke et al., 2020; Li et al., 2021; Saidi et al., 2015), and have been adopted in a few FFA studies (Basu and Srinivas, 2013; Wang et al., 2006; Wu et al., 2019). Previous FFA studies generally show limited use of EDA for evaluating the stationary assumption with considerable variations from one study to another. For instance, several studies solely investigated the presence of temporal trends without identifying the trend types and/or the presence of change points (Cui et al., 2023; Hesarkazazi et al., 2021; Lu et al., 2020; Mondal and Mujumdar, 2016; Šraj et al., 2016; Zhang et al., 2019). Therefore, there is a need for a systematic EDA to assess the stationarity assumption and provide evidence for guiding the selection between the S-FFA or NS-FFA approaches.

To capture the nonstationarity, the statistical model in NS-FFA is composed of a probability distribution and a nonstationary structure that governs its evolution over time. The nonstationary structure is a deterministic function of the selected covariate(s) that controls the changes in the distribution parameters/moments. Typical nonstationary structures are linear functions of time or physical covariates modeling the changes in the distribution's location and/or scale parameters. The NS-FFA model is commonly selected by ranking a number of candidate models according to a performance metric, such as the *AIC* and *BIC*. However, this approach confronts practical and theoretical obstacles, including ergodicity violation, potential overfitting, and equifinality of multiple models (Luke et al., 2017; Serinaldi and Kilsby, 2015; Vidrio-Sahagún and He, 2022c). In this context, ergodicity is the property of stationary stochastic processes that ensures the convergency of sample statistics to population statistics as the sample size increases. This property does not hold for nonstationary processes, where population statistics change over time, making a single time series insufficient for reliable inference. Equifinality, on the other hand, denotes situations where multiple candidate models show nearly identical optimal performance according to an evaluation metric (although potentially producing different estimates), thereby introducing uncertainty into the model selection. In contrast, the decomposition-based approach (Cunderlik and Burn, 2003; Vidrio-Sahagún and He, 2022c) offers a means to mitigate these challenges by selecting the probability distribution and nonstationary structure of the NS-FFA model separately and explicitly using knowledge of the nonstationarity. In this NS-FFA model selection approach, applying an appropriate decomposition technique is relevant, as the distribution and nonstationary structure are chosen using the decomposed stationary stochastic and deterministic time-dependent components of flood series, respectively. Vidrio-Sahagún and He (2022c) developed a procedure based on the general decomposition of nonstationary stochastic processes that preserves higher-order moments

of the stochastic component (e.g., skewness) and prevents bias in the identification of the optimal distribution.

The introduction of the nonstationary structure in NS-FFA models makes the parameter estimation more complicated than in S-FFA. In NS-FFA, the parameter estimation is commonly conducted using the ML method due to its flexibility to incorporate the nonstationary structure into the distribution (Coles, 2001; Katz, 2013). Some other available techniques include GML (El Adlouni et al., 2007), Bayesian approaches with MCMC samplers (Luke et al., 2017; Ragno et al., 2019), and particle filters (Vidrio-Sahagún et al., 2021). For quantifying the uncertainty in NS-FFA, the bootstrap method is popular probably because it is compatible with any parameter estimation methods, easy to implement, and can yield relatively asymmetric confidence intervals (Gilleland and Katz, 2016; Ouarda et al., 2019; Serinaldi and Kilsby, 2015). Whereas the theoretical superiority of the PL method was overshadowed by its limited formulations under nonstationarity, computational burdensome, and potential numerical divergence when dealing with short datasets (Cooley, 2013; Obeysekera and Salas, 2014). Most recently, Vidrio-Sahagún and He (2022a) developed enhanced versions of the PL method for NS-FFA, namely the *regula-falsi* PL method (RF-PL method) and its generalized version (RF-GPL method). These modifications can reduce the computational demand by up to 96% and mitigate or eliminate the numerical instability. Since no consensus has been reached on which methods are preferable, especially for general practice, the use of different methods is still expected.

Furthermore, frameworks, guidelines, and software have been developed for conducting FFA in practice. Examples of frameworks and guidelines include the Bulletin 17C for the USA (England Jr. et al., 2019), the Flood Estimation Handbook and its updates in the UK (Institute of Hydrology, 1999; Kjeldsen et al., 2008; 2014a), and the Federal Hydrologic and Hydraulic Procedures for Flood Hazard Delinement for Canada (Natural Resources Canada; Public Safety Natural Resources Canada, 2019). While there is a growing effort worldwide to incorporate potential climate change impacts into flood estimation, the development of frameworks and guidelines considering the nonstationarity in floods is still in its infancy in general (Ball et al., 2016; Wasko et al., 2021). Several software programs for S-FFA that are readily accessible to the public include CFA (<https://www.trentu.ca/iws/research-projects/software>), HYFRAN (<https://www.wrpllc.com/index.html>), and RMC-BestFit (<https://www.rmc.usace.army.mil/Software/RMC-BestFit/>), among many others. However, comprehensive software packages that enable NS-FFA are scarce and often limited to certain distributions and/or methods (e.g., for estimating distribution parameters and uncertainty). These limitations restrict their applications. For example, the R package extRemes 2.0 (<https://cran.r-project.org/web/packages/extRemes/index.html>) (Gilleland and Katz, 2016) only considers the Generalized Extreme Value (GEV) and Generalized Pareto distributions and some estimation methods are limited to S-FFA, such as the profile likelihood method. The Matlab-based ProNEVA toolbox (https://github.com/elisaragno/ProNEVA_Toolbox) (Ragno et al., 2019) includes the log-Pearson Type 3 (LP3) distribution but is restricted to a Bayesian MCMC method for parameter and uncertainty estimation and only uses MK and White tests for EDA. Other recently released programs provide incremental improvements, such as mixed-type distributions and regionalization (de Luca and Napolitano, 2023).

Despite the above advancements in FFA, there is still a need for a practitioner-oriented and software-supported framework for conducting both the S-FFA and NS-FFA consistently and systematically. Such a framework needs to include procedures to guide the selection of either S-FFA or NS-FFA, determine the FFA model, estimate the FFA model parameters and frequency estimates, and quantify uncertainty in the frequency estimates. Therefore, the objective of this paper is to develop a framework that allows for repeatable workflows and includes state-of-art methods for S-FFA and NS-FFA. The framework provides a platform for enhancing the reproducibility of the FFA by covering from the

EDA to frequency estimates and their confidence intervals. To promote the use of the framework, a Graphical User Interface (GUI) is created to improve accessibility and simplify its application. The framework is demonstrated through several practical applications.

3. Proposed framework

Algorithm 1 outlines the end-user steps for implementing the proposed framework, while **Fig. 1** further presents the general workflow. As illustrated, the framework consists of three primary modules, namely (a) data processing, where a flood series is inputted and then processed for the EDA and/or NS-FFA model determination (if needed); (b) FFA approach selection (S-FFA or NS-FFA) aided by the EDA, which consists of a suite of statistical tests; and (c) flood frequency analysis, consisting of state-of-the-art methodologies to select the model, estimate the model parameters, quantify the uncertainty as confidence intervals, and assess the model performance. More details of the EDA sub-module in the FFA approach selection module are provided in **Fig. 2**. This framework sets a baseline of methods for the FFA approach to narrow wide-ranging practices while considering several alternative but commonly used methods to maintain flexibility. The framework allows for consistency and repeatability, as all decision points in the process highlighted in **Fig. 1** are documented. Although modellers can tailor the implementation at their own discretion, such as using a different FFA approach, model, metric, and/or method, all the decision points are guided by recommendations. In addition, the framework is constructed in a modular fashion to allow future expansion and developments, including the incorporation of nonstationary attribution and the use of more distributions and methods at any step of the FFA.

Algorithm 1. (main algorithm – general framework workflow).

Input: Hydrological annual maximum series (AMS).

Output: Flood frequency estimates with confidence intervals.

- Step 1: Perform the EDA in the mean and variability of the AMS.
- Step 2: Decide whether to perform S-FFA or NS-FFA considering the EDA results. If S-FFA is selected, proceed to Step 5.
- Step 3: If NS-FFA is selected, identify the deterministic time-dependent component behind the nonstationarity considering the EDA results, and in turn, determine the nonstationary structure of the NS-FFA model.
- Step 4: Obtain the stationary stochastic component of the AMS using the nonstationary decomposition (data processing module) and the deterministic component identified in Step 3.
- Step 5: Determine the best-fit distribution using the L-moment ratio diagram based on the AMS for S-FFA or the decomposed stationary stochastic component of the AMS for NS-FFA. The selected distribution alone and coupled with the nonstationary structure (Step 3) determine the S-FFA and NS-FFA models, respectively.
- Step 6: Choose the method and estimate the model parameters. The L-moments and ML methods are recommended defaults for S-FFA and NS-FFA, respectively.
- Step 7: Choose the method and quantify the uncertainty (confidence intervals). The parametric bootstrap and RF-PL methods are recommended defaults for S-FFA and NS-FFA, respectively.
- Step 8: Produce the frequency estimates along with confidence intervals, which are constant and time-varying in S-FFA and NS-FFA, respectively, and perform the model assessment.

3.1. Data processing module

The AMS dataset is fed into the framework through the data processing module. This module does not include the wide range of possible data pre-processing methods, such as “naturalization” or “de-regulation”, data infilling, and estimation of instantaneous peak flows from

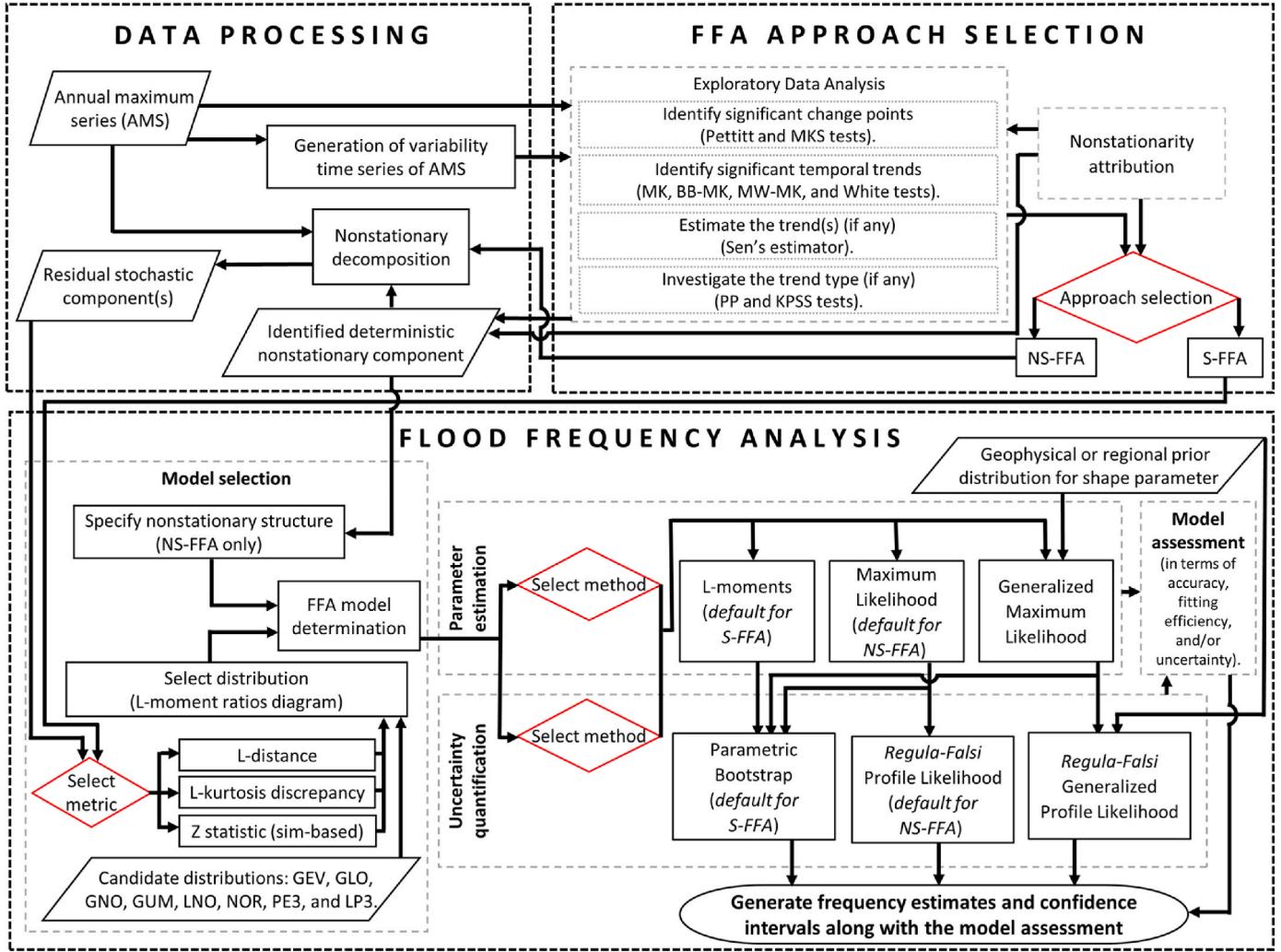


Fig. 1. General workflow of the proposed FFA framework. The red diamonds indicate the decision points and choices that can be made at the discretion of the modeller.

daily peaks, among others. In addition, the use of peaks over threshold (POT) series, which consists of all records above a sufficiently high threshold such that all events are considered independent extremes, was not included. The POT approach was excluded because it is typically constrained to the Generalized Pareto distribution and the threshold selection is often subjective (Coles, 2001; Nerantzaki and Papalexiou, 2021; Salas et al., 2018). Furthermore, this module generates the variability time series of the AMS for the EDA and conducts the nonstationary decomposition if NS-FFA is selected.

3.1.1. Generation of variability time series of AMS

Nonstationarity can present in the mean and/or high-order statistical moments of the AMS (e.g., variability and skewness). In the context of NS-FFA, the vast majority of the literature has primarily focused on the weak/second-order nonstationarity (Lindgren et al., 2013; von Storch and Zwiers, 2002), which encompasses up to the variability of the AMS. Thus, the current version of the framework only considers second-order nonstationarity. To explore potential evidence of nonstationarity in the variability of AMS, a time series of standard deviations is generated using the moving windows method. The generated time series of standard deviations is inputted to the EDA sub-module in the module of FFA approach selection. In the moving windows approach, the choice of the window hyperparameters is key, as their inadequate selection can lead to biased analysis results (Vidrio-Sahagún and He, 2022c). Following the simulation findings from Vidrio-Sahagún and He (2022c), the window

length and step are set to 10 and 5 years, respectively, as defaults. Yet, these can be tailored by modellers.

3.1.2. Dataset nonstationary decomposition

If NS-FFA is conducted, the AMS is decomposed to derive its stationary stochastic component, which is used to select the distribution of the NS-FFA model. The decomposition procedure follows that proposed by Vidrio-Sahagún and He (2022c), which relies on the statistical representation of nonstationary stochastic processes. The general decomposition is formulated as $G[Y_t] = G[X_t] + Z_t$, where Y_t is the observable nonstationary stochastic process, X_t is the stationary stochastic component, Z_t is the time-dependent deterministic (nonstationary) component that governs the changes of $G[Y_t]$ over time, and $G[\bullet]$ is a generic operation (e.g., $E[\bullet]$ or $\text{Var}[\bullet]$). The dataset decomposition consists of estimating X_t by removing Z_t from Y_t using an appropriate decomposition scheme according to the structural form of Z_t . Here, the decomposition is conducted on homogenous datasets (i.e., without change points) under three general scenarios of nonstationarity, namely, (a) scenario 1 (S1): temporal trend in the mean only (i.e., $E[Y_t] = E[X_t] + Z_t$); (b) scenario 2 (S2): temporal trend in the standard deviation only (i.e., $\text{Var}[Y_t]^{1/2} = \text{Var}[X_t]^{1/2} + Z_t$); and (c) scenario 3 (S3): temporal trends in both mean and standard deviation (i.e., $E[Y_t] = E[X_t] + Z_{1,t}$; $\text{Var}[Y_t]^{1/2} = \text{Var}[X_t]^{1/2} + Z_{2,t}$). In these scenarios, the stationary dataset $x_t = \{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$ is obtained by removing Z_t in the form of $f(t)$ and/or $g(t)$ from the

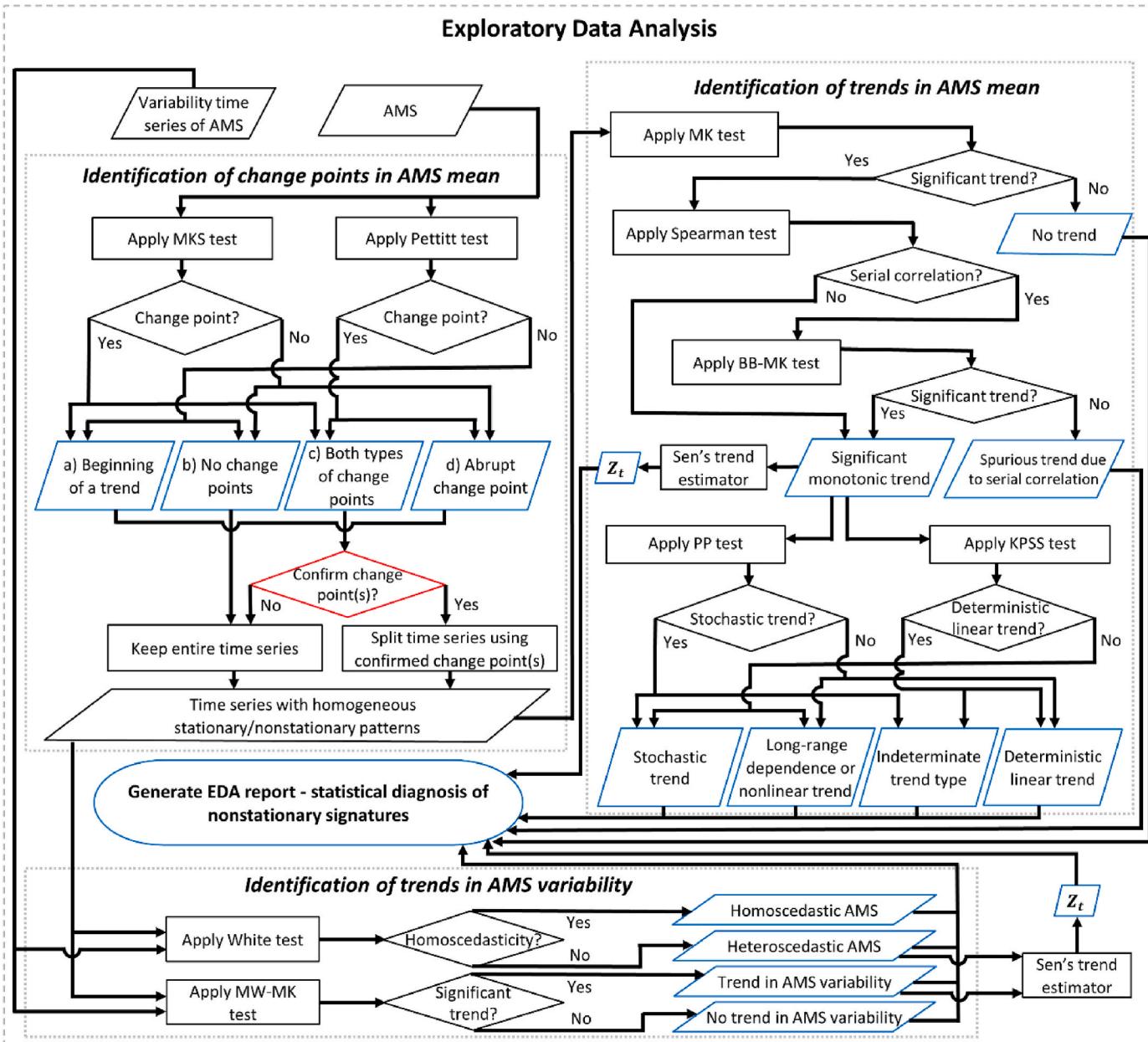


Fig. 2. Workflow of the exploratory data analysis (EDA) sub-module of the FFA approach selection module. The red diamond indicates the decision made at the discretion of the modeller, while the blue parallelograms indicate outputs for the EDA report containing the statistical diagnosis of significant nonstationary signatures.

AMS ($y_t = \{y_{t_1}, y_{t_2}, \dots, y_{t_n}\}$) using the corresponding equations from the set below:

$$\begin{aligned}
 S1: & \{x_t = y_t - f(t)\} \\
 S2: & \left\{ x_t = \hat{\mu}_{Y_t} + (y_t - \hat{\mu}_{Y_t}) \bullet h(t); \text{ where } h(t) = \frac{\hat{\sigma}_{X_t}}{\hat{\sigma}_{X_t} + g(t)} \right. \\
 S3: & \left. \left\{ x_t = \hat{\mu}_{Y_t-f(t)} + (\{y_t - f(t)\} - \hat{\mu}_{Y_t-f(t)}) \bullet h(t); \text{ where } h(t) = \frac{\hat{\sigma}_{X_t}}{\hat{\sigma}_{X_t} + g(t)} \right\} \right\}
 \end{aligned} \quad (1)$$

where $f(t)$ is the estimated Z_t in the nonstationary mean of Y_t ; $\hat{\mu}_{Y_t}$ is the estimated nonstationary mean of Y_t (which is time-independent in S2); $h(t)$ is a normalized multiplicative term depicting Z_t in the nonstationary standard deviation of Y_t ; $g(t)$ is the estimated Z_t in the nonstationary standard deviation of Y_t ; $\hat{\sigma}_{X_t}$ is the estimated stationary standard deviation of X_t , using the beginning of the observation period as the refer-

ence; and $\hat{\mu}_{Y_t-f(t)}$ is the estimated stationary mean of the partially decomposed dataset after $f(t)$ is removed but $g(t)$ is still present. This method can preserve the time-invariant statistical moments of the dataset after removing the Z_t , especially the higher-order moments (Vidrio-Sahagún and He, 2022c).

The dataset decomposition relies on the identified Z_t . Broadly speaking, Z_t can be modeled as a function of the selected physical covariate(s) or time. However, the use of physical covariates requires a detailed attribution of nonstationarity, which is often challenging due to the generally and often low dependence of the nonstationarity and floods on physical covariate(s) (e.g., climate indices, temperature, and rainfall) (Archfield et al., 2016; Burn and Whitfield, 2017, 2023; Ray and Goel, 2019). Instead, in the literature, Z_t is often modeled as a function of time to depict and surrogate any underlying time-varying physical processes behind nonstationarity (Abdelmoaty and Papalexiou, 2023; Obeysekera and Salas, 2016; Prosdocimi and Kjeldsen,

2021; Serago and Vogel, 2018; Sun et al., 2018; Vidrio-Sahagún and He, 2022c). For instance, temperature, which influences extreme rainfall and snowmelt and consequently can impact floods, has exhibited roughly monotonic increases over time (Trenberth et al., 2003; Wasko, 2021) and hence would be reasonably depicted using time. Therefore, time is used as the covariate in the framework and the Z_t corresponds to the temporal trends in the mean and/or standard deviation estimated in the EDA. However, it is worth noting a caveat when extrapolating temporal patterns into the future, as historical trends might not adequately capture the future evolution of the processes driving non-stationarity. In such cases, the use of physical covariate(s) might be more suitable (Schlef et al., 2018, 2023).

3.2. FFA approach selection

This module guides the selection of the FFA approach (S-FFA or NS-FFA) by leveraging the results from the EDA. In the EDA, the feasibility of the stationary assumption is assessed considering the presence of statistically significant nonstationary signatures, such as change points and deterministic temporal trends in the mean and/or the variability of the AMS. If any nonstationary signature is diagnosed, the stationarity assumption is not supported by the available evidence and NS-FFA is recommended. Otherwise, S-FFA is suggested. It is worth mentioning that performing the attribution of nonstationarity is fundamental to selecting the appropriate FFA approach with more confidence (Merz et al., 2012; Serinaldi et al., 2018; Slater et al., 2021a). Since the attribution of nonstationarity is an ongoing research topic, it is not currently included in the framework but is recommended for future development, as discussed in Section 6. Under such a circumstance, it is advised to adopt deductive reasoning for supporting the FFA approach selection (e.g., see Koutsoyiannis and Montanari, 2015; Serinaldi et al., 2018).

3.2.1. Exploratory data analysis

Algorithm 2 outlines the procedures of the EDA, and Fig. 2 further illustrates the general workflow including the statistical tests used. In this sub-module, the presence of change points is first used to identify homogenous periods with stationary or nonstationary monotonic patterns. Neglecting change points could lead to the inappropriate selection of FFA approach. For example, the presence of two opposite trends in a time series (e.g., upward and downward trends in two sub-periods) can counteract and negatively affect the trend investigation if analyzing the complete time series. The presence of abrupt change points is explored using the Pettitt test (Pettitt, 1979), whereas change points indicating the beginning of a significant trend are explored using the MKS test (Sneyers, 1975). The change point tests lead to four possible outcomes (Fig. 2), suggesting either retaining the entire AMS or dividing it into two or more sub-periods for the subsequent analyses. It is however worth mentioning that accounting for change points could undermine the reliability of both the EDA and FFA due to reduced sample sizes. Thus, in this framework, the modeller has the option to decide whether to split the time series or not, as the attribution of detected change point(s) is usually more demanding and requires more detailed justification. A significant change point can be assumed to emerge due to natural variability if there is no supporting physical evidence. In other cases, a change point could be introduced by a modeller when it is supported by a clear physical rationale (e.g., the construction of a dam or a change in dam operation).

Algorithm 2. (EDA).

Input: AMS and its variability time series (from the data processing module).

Output: EDA report - statistical diagnosis of nonstationary signatures.

Step 1: Apply the Pettitt and MKS tests to identify the presence of potential change point(s) in the AMS.

Step 2: Determine whether to use the entire AMS or split it into homogenous periods considering the Pettitt and MKS tests results. If the AMS is split and all periods are of interest, proceed with the AMS of the first homogeneous period in Steps 3 to 9, and repeat these steps for each homogeneous period. Otherwise, use the AMS of the homogeneous period of interest.

Step 3: Apply the MK test to the AMS to detect the presence of a temporal trend. If no significant trend is detected, proceed to Step 8.

Step 4: Apply the Spearman test to detect the presence of serial correlation. If there is no serial correlation, proceed to Step 6.

Step 5: Apply the BB-MK test to detect the presence of a temporal trend. If no significant trend is detected, proceed to Step 8.

Step 6: Estimate the trend slope and intercept using the Sen's trend estimator.

Step 7: Apply the PP and KPSS tests to identify the trend type.

Step 8: Apply the White and moving-window MK (MW-MK) tests to the AMS and the variability time series of the AMS, respectively. If there is no significant trend in the variability of the AMS and the AMS is homoscedastic, proceed to Step 10.

Step 9: Estimate the trend slope and intercept in the AMS variability using the Sen's trend estimator.

Step 10: Generate the EDA report - statistical diagnosis of nonstationary signatures.

The approaches to detect trends in the variability differ from those to detect trends in the mean due to the increased uncertainty when exploring higher-order moments. The presence of significant monotonic trends in the mean is investigated using the MK test (Kendall, 1975; Mann, 1945). The BB-MK (Khalil et al., 2009, Önöz and Bayazit, 2012) is adopted to prevent false detection of trends in the mean due to serial correlation (if present) and confirm the significance of a trend if identified by the MK test. Serial correlation is investigated by the Spearman's rho correlation test. The trend slope and intercept are estimated using the non-parametric Sen's trend estimator (Sen, 1968; Theil, 1950). Additionally, when a trend in the mean is detected, it is investigated whether it is deterministic or stochastic. Acknowledging typical sample size limitations, it is assumed that the deterministic trend (if any) is linear. Consequently, the PP (Phillips and Perron, 1988) and KPSS (Kwiatkowski et al., 1992) unit root tests are used in conjunction. The PP test evaluates the null hypothesis that the time series has a stochastic trend, whereas the KPSS test assesses the null hypothesis that the time series has a deterministic linear trend. If the PP rejects its null hypothesis and the KPSS accepts it, the statistical analysis supports the postulate of (linear) nonstationarity. If the PP accepts while the KPSS tests rejects their null hypothesis, the statistical analysis suggests that the detected trend is due to stochastic natural variability. If both tests fail to reject their hypothesis, there would be insufficient evidence to distinguish the type of trend, while the hypothesis rejections in both tests might suggest deterministic non-linear trend or long-range dependence (Franzke et al., 2020). Apart from the presence of a trend in the mean, the presence of a trend in the variability of the AMS and the homoscedasticity of the AMS are examined using the MW-MK and White (White (1980) tests, respectively. The MW-MK test evaluates the presence of temporal trends in the time series of standard deviations of the AMS (generated in the data processing module), whereas the White test explores the homoscedasticity of the AMS per se based upon its parametric assumptions. Hence, both tests complement each other's limitations. Similar to the trend in the mean, the trend in the standard deviation is also estimated using the Sen's estimator.

3.3. Flood frequency analysis

This module implements the selected FFA approach and includes sub-modules for model selection, parameter estimation, uncertainty quantification, and model performance assessment. Each sub-module offers various methods for its target task, with a recommended default

method. Note that if the time series contains a change point and the modeller decides to account for it, a piece-wise NS-FFA model structure is adopted. In such a case, this module is implemented independently for each homogenous period to determine the pieces of the model.

3.3.1. Model selection

In this sub-module, the statistical model is determined for S-FFA or NS-FFA. For S-FFA, a time-invariant probability distribution is selected from a pool of candidate distributions as the statistical model. For NS-FFA, which involves continuously time-varying parameters, a probability distribution is selected from the same pool of candidate distributions along with a nonstationary structure to depict its evolution over time. In the case of piece-wise NS-FFA, which captures multiple periods with different homogeneous patterns, the temporal dependence of the distribution is described in separate pieces that can be either time-variant or time-invariant distributions. The pool of candidate distributions includes the Generalized Extreme Value (GEV), Generalized Logistic (GLO), Generalized Normal (GNO), Gumbel (GUM), Log-Normal (LNO), Log-Pearson type III (LPE3), Normal (NOR), and Pearson type III (PE3). These distributions have been commonly used in FFA and cover a wide variety of statistical characteristics that AMSs may exhibit. Their cumulative distributions are listed in Appendix A.

The distribution is selected using the L-moment ratio diagram. The L-moment ratio diagram is a widely accepted tool to identify a probability distribution by comparing the sample L-moment ratios to theoretical L-moment ratios of candidate distributions (Hosking, 1990; Hosking and Wallis, 2009). Refer to Appendix B for the estimation of sample L-moment ratios. For the selection, three metrics can be employed to identify the best-fit distribution for the sample time series, namely, the L-distance, the L-kurtosis discrepancy, and the Z statistic. The L-distance is the shortest Euclidean distance from the sample's to the theoretical candidate distributions' L-skewness and L-kurtosis. The L-distance assesses the similarity of the L-moment ratios of the sample and the pool of candidate distributions (including both two- and three-parameter distributions). The L-kurtosis discrepancy measures the difference in the L-kurtosis between the sample and the candidate theoretical distributions. The L-kurtosis discrepancy is appropriate for selecting the distribution among three-parameter distributions only. This is because three-parameter distributions may considerably differ in the L-kurtosis even when they have similar L-skewness. The Z statistic builds upon the L-kurtosis discrepancy but takes into account its sampling variability. The Z statistic is computed as described by Hosking and Wallis (2009) with the simplification for a single site as outlined in Zhang et al. (2020) (see Appendix C). The Z-statistic calculation relies on a simulation procedure that involves random resampling. In this framework, the seed in the random number generator is set to ensure the reproducibility of results and avoid numerical variations in the Z-statistic calculation caused by its randomness component. As there is no consensus on which metric is preferred, this framework recommends that the modeller explore the three metrics and their corresponding selected best-fit distribution(s) to assess the overall agreement.

In NS-FFA, the identification of the nonstationary structure relies on the underlying deterministic nonstationary pattern Z_t , which is identified by the EDA and supported by deductive reasoning in the current framework (and/or nonstationarity attribution in future developments). The nonstationary structure is defined using the moment equations of the selected distribution (Stedinger, 2017; Stedinger et al., 1993) to identify the specific distribution parameter(s) that varies over time as per the Z_t . For instance, when the Z_t can be represented by a trend in the mean, the location parameter of the selected distribution is time dependent. When the Z_t can be represented by a trend in the standard deviation, both the location and scale parameters are time dependent. Since only linear trends are considered in the EDA as aforementioned, and consequently in the dataset decomposition for the distribution selection, the NS-FFA model also considers parsimonious linear nonstationary structures. Thus, the corresponding distribution parameter(s) of

NS-FFA models is expressed as a linear function of time.

In this paper, the NS-FFA models are denoted as $D_{u,w,0}$ and given by:

$$D_{u,w,0} \sim D(\xi_t, \alpha_t, \kappa); \quad \begin{cases} \xi_t = \xi_0 + \mathbb{I}(u)\xi_1 t \\ \alpha_t = \alpha_0 + \mathbb{I}(w)\alpha_1 t \\ \kappa = \text{constant} \end{cases} \quad (2)$$

where D denotes the selected probability distribution; u and w are sub-indices specifying the nonstationary structure and are set to "0" or "1" to indicate time-invariance (constant) or linear temporal dependence; ξ , α , and κ denote the location, scale, and/or shape parameters of all candidate distributions, respectively; $\mathbb{I}(\bullet)$ is the indicator function, which takes the value of 1 if its argument is "1" and 0 otherwise, indicating whether ξ and α are expressed as a function of time or constant; and ξ_0 and ξ_1 as well as α_0 and α_1 denote the regression coefficients of ξ and α , respectively. For instance, $GEV_{1,0,0}$ denotes the NS-FFA model constructed with the GEV distribution with time-varying ξ (expressed as $\xi_t = \xi_0 + \xi_1 t$) and constant α and shape parameter (κ); while $LPE3_{1,1,0}$ refers to the NS-FFA model constructed with the LPE3 distribution with time-varying ξ and α , expressed as $\xi_t = \xi_0 + \xi_1 t$ and $\alpha_t = \alpha_0 + \alpha_1 t$, respectively, and constant κ . Additionally, if a piece-wise NS-FFA is adopted, the pieces of the model are identified with superscripts. For example, $GNO_{0,0,0}^{\text{piece } 1}$ indicates the time-invariant GNO distribution in the first subperiod. Note that this sub-module employs the decomposition-based approach exclusively for determining the structural form of the NS-FFA model (i.e., the distribution and nonstationary structure) without estimating its parameters. The sub-module described in Section 3.3.2 parametrizes the NS-FFA model by jointly estimating the regression coefficients (ξ_0 , ξ_1 , and/or α_0 and α_1 in Eq. (2)) and the constant parameters (α and/or κ), as typically conducted in the NS-FFA literature.

3.3.2. Model parametrization

This sub-module estimates the distribution parameters of either the S-FFA or NS-FFA model. Since the NS-FFA model parameters are expressed as functions of the covariate, the parametrization of the model consists of estimating the regression coefficients of the time-varying parameters and the constant parameters. The available methods for parametrizing the models include the L-moments, ML, and GML methods. These methods are adopted due to their robustness, superior performance (particularly for limited sample sizes), and/or applicability for both S-FFA and NS-FFA. For S-FFA, the L-moments and ML methods are available for estimating the distribution parameters of any candidate distributions, while the GML method is currently limited to the GEV distribution. The L-moments method is recommended as the default for S-FFA due to its robustness and widespread use to date. For NS-FFA, the L-moments method is not available, whereas the ML and GML methods can be readily used due to their flexibility to incorporate the nonstationary structure into the probability distribution. The ML is the default method for NS-FFA, while the GLM method is only available for models based on the GEV distribution.

The L-moments method is based on the expectations of linear combinations of order statistics and is often used for parameter estimation because it is less sensitive to sampling variability and measurement errors. The L-moments and ratios are estimated from a sample as described by Hosking (1990) and Hosking and Wallis (2009) (see Appendix B).

The ML method relies on the principle of likelihood, which is defined as the joint probability of observations $y_t = \{y_1, y_2, \dots, y_n\}$ (here, the AMS) as a function of a given parameter set θ . Essentially, likelihood calculates the probability of observing y_t under certain values of θ . In both the S-FFA and NS-FFA, the likelihood ($L(\theta)$) and log-likelihood ($\ell(\theta)$) of y_t is expressed as follows (Coles, 2001):

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) \quad (3)$$

$$\ell(\theta) = \log[L(\theta)] = \sum_{i=1}^n \log[f(y_i; \theta)] \quad (4)$$

where $f(y_t; \theta)$ represents the probability density function of the AMS, which is constant and time-variant in S-FFA and NS-FFA, respectively. The ML estimate of θ , denoted as $\hat{\theta}$, is the set of parameters that maximizes either $L(\theta)$ or $\ell(\theta)$. This framework employs the $\ell(\bullet)$ function due to its computational efficiency in optimization procedures. However, the use of the ML method is challenged by its numerical instability particularly when dealing with short datasets. Thus, [Martins and Stedinger \(2000\)](#) proposed the GML for S-FFA when using the GEV distribution. The GML method introduces an informative geophysical prior distribution for the shape parameter of the GEV distribution to eliminate potentially invalid parameter values and thereby avoids numerical divergence. [El Adlouni et al. \(2007\)](#) followed the same notion and extended the GML method to NS-FFA for the GEV distribution using the same geophysical prior. The geophysical prior distribution constrains the GEV shape parameter to a range of statistically and physically reasonable values. The geophysical prior takes the form of a beta distribution $\pi(\kappa) \sim Beta(p=6, q=9)$ defined over the interval $(-0.5 < \kappa < 0.5)$, roughly 90% of its probability mass is concentrated in the range $-0.1 < \kappa < 0.3$ and has a mode of 0.12 and an expected value of 0.1. The generalized log-likelihood function is thus expressed as $G\ell(\theta) = \ell(\theta) + \log[\pi(\kappa)]$. In this framework, the same geophysical prior of [Martins and Stedinger \(2000\)](#) is adopted.

3.3.3. Uncertainty quantification

This sub-module generates confidence intervals to quantify the uncertainty in the estimated quantiles by the S-FFA or NS-FFA approach. Since the L-moments method is the default in the model parameterization for S-FFA, the bootstrap method ([Efron, 1992](#)) is recommended as the default for S-FFA due to its compatibility with any parameter estimation methods. Compared to the non-parametric bootstrap, the parametric bootstrap is deemed superior, as it often yields empirical coverage frequencies that more closely match the pre-selected nominal confidence level and is less prone to underestimating uncertainty ([Kyselý, 2008, 2010; Panagoulia et al., 2014](#)). Thus, the parametric bootstrap method is adopted in this framework. The parametric bootstrap method (refer to Appendix D) randomly resamples from the fitted FFA model, resulting in some variations in the confidence intervals from time to time. Hence, the seed in the random number generator is set to guarantee the reproducibility of results and avoid numerical variations in the confidence intervals caused by its randomness component. In contrast, the RF-PL method is the default method for NS-FFA, as it relies on the ML principle, which is used for parameter estimation in NS-FFA. When numerical instability occurs in the RF-PL method (especially when dealing with short datasets), the RF-GPL method is also available if using the GEV distribution.

The RF-PL method ([Vidrio-Sahagún and He, 2022a](#)) combines the *regula-falsi* and PL methods ([Coles, 2001](#)) to estimate the bounds of the confidence intervals, instead of computing the full profile. The RF-PL method derives the confidence interval of a parameter θ_j of the S-FFA or NS-FFA model by finding the intersections of its profile likelihood $\ell_p(\theta_j)$ with the threshold associated with the selected confidence level (e.g., 95%). In the RF-PL method, $\ell_p(\theta_j)$ is defined as the maximized log-likelihood function $\ell(\bullet)$ at fixed values of θ_j and optimized θ_{-j} (i.e., all other distribution parameters except θ_j) ([Coles, 2001](#)). In particular, the RF-PL method derives the confidence interval of the constant (stationary) quantiles (y_T) or time-variant (nonstationary) quantiles ($y_{T,t}$) using the re-parameterized log-likelihood function incorporating y_T or $y_{T,t}$ as a model parameter ([Coles, 2001; Vidrio-Sahagún and He, 2022a](#)). The required reparametrized log-likelihood functions for all the S-FFA models covered in the framework are provided in Appendix E, while those for NS-FFA can be found in [Vidrio-Sahagún and He \(2022a\)](#) and ([Vidrio-Sahagún et al. \(2023a\)](#).

Similar to the GML method, the RF-GPL method utilizes the geophysical prior $\pi(\kappa)$ to restrict the shape parameter of the GEV distribution to a reasonable range by applying the generalized maximum

likelihood principle. The RF-GPL method mainly differs from the RF-PL method in the use of the generalized log-likelihood function $G\ell(\bullet)$. The $G\ell(\bullet)$ includes the additional term of $\log[\pi(\kappa)]$ in the $\ell(\bullet)$ function to account for the geophysical prior distribution. The geophysical prior distribution used in the RF-GPL method is the same as the one used in the GML method. The Nelder-Mead simplex method ([Lagarias et al., 1998](#)) is used to maximize the $\ell(\bullet)$ and $G\ell(\bullet)$ in the RF-PL and RF-GPL methods, respectively.

3.3.4. Model performance assessment

This sub-module quantitatively assesses the model performance in terms of accuracy, fitting efficiency, and/or uncertainty. It is worth highlighting that the model assessment is not used to select models, especially in NS-FFA, because a model of superior performance may not necessarily describe the actual nonstationary stochastic process ([Luke et al., 2017; Serinaldi and Kilsby, 2015; Vidrio-Sahagún and He, 2022c](#)). In this framework, three evaluation metrics are included to assess the accuracy performance, namely, R^2 , root mean square error (*RMSE*), and *Bias*. These metrics are calculated by:

$$R^2 = \left[\frac{\sum(o_T - \bar{o}_T)(m_T - \bar{m}_T)}{\sqrt{\sum(o_T - \bar{o}_T)^2} \sqrt{\sum(m_T - \bar{m}_T)^2}} \right]^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (m_T - o_T)^2} \quad (6)$$

$$Bias = \frac{1}{n} \sum (m_T - o_T) \quad (7)$$

where o_T are the empirical quantiles associated with the return period T according to the unbiased Weibull plotting position formula $p_{1:n} = (r)/(n+1)$ (where r is the rank of the observation and n is the sample size), and m_T are the corresponding model quantiles. Since the calculations of these metrics require the use of the plotting position formula, they are only applicable to S-FFA.

In terms of fitting efficiency, the *AIC* and *BIC*, which deal with the trade-off between the goodness-of-fit offered by a model and its complexity based on information loss, are employed. These metrics can be estimated as a function of either likelihood or *RMSE* by:

$$\begin{aligned} AIC &= 2N_\theta + n \log(RMSE) \quad \text{or} \quad AIC = 2N_\theta - 2\ell(\hat{\theta}) \\ BIC &= N_\theta \log(n) + n \log(RMSE) \quad \text{or} \quad BIC = N_\theta \log(n) - 2\ell(\hat{\theta}) \end{aligned} \quad (8)$$

where N_θ is the number of parameters of the model, and $\ell(\hat{\theta})$ is the maximum log-likelihood of the fitted model. A model yielding a smaller *AIC* and *BIC* is superior in fitting efficiency.

The uncertainty is evaluated using the average bandwidth (*AW*), the percentage of observations coverage (*POC*), and the Coverage Width Index (*CWI*) ([Kasiviswanathan et al., 2019](#)). For S-FFA, these metrics are given as follows:

$$AW = \frac{1}{n} \sum (m_T^U - m_T^L) \quad (9)$$

$$POC = \frac{100}{n} \sum C_k, \text{ where } C_k = \begin{cases} 1 & \forall k \text{ s.t. } m_T^L \leq o_T \leq m_T^U \\ 0 & \text{elsewhere} \end{cases} \quad (10)$$

$$CWI = AW \bullet \exp \left([1 - \rho] - \frac{POC}{100} \right)^2 \quad (11)$$

where m_T^U and m_T^L are the upper and lower uncertainty bounds associated with the return period T , respectively; and ρ is the significance level. In this framework, the *AW* is averaged over a number of specified T s associated with the empirical exceedance probabilities determined by the unbiased Weibull plotting position. For NS-FFA, in the calculation of

the AW , the m_T^U and m_T^L are replaced by the time-variant $m_{t,T}^U$ and $m_{t,T}^L$, which are the m_T^U and m_T^L corresponding to the time slide of t . As the calculations of the POC and CWI require o_T , which cannot be estimated in NS-FFA, the POC and CWI are thus not computed. In this framework, the AW is derived in two ways, namely (a) as the average over both the temporal and frequency domains (using the T 's derived by the Weibull plotting position considering the available sample size), for comparison with the AW of S-FFA, and (b) as the average over the temporal domain only for a specific T of interest.

4. Framework deployment - Graphical User Interface

The framework code was written in Matlab and structured in a modular manner. To facilitate the application of the FFA framework by engineering practitioners, a Graphical User Interface (GUI) was developed (Fig. 3). The GUI is accompanied by a user manual (see supplementary material). The developed GUI is a standalone application that enables users to access its functionality without requiring a Matlab license. In the GUI, the data processing module runs in the background when executing the EDA and/or model selection sub-modules, and thus it is not directly shown in the main interface of the GUI. In addition, as

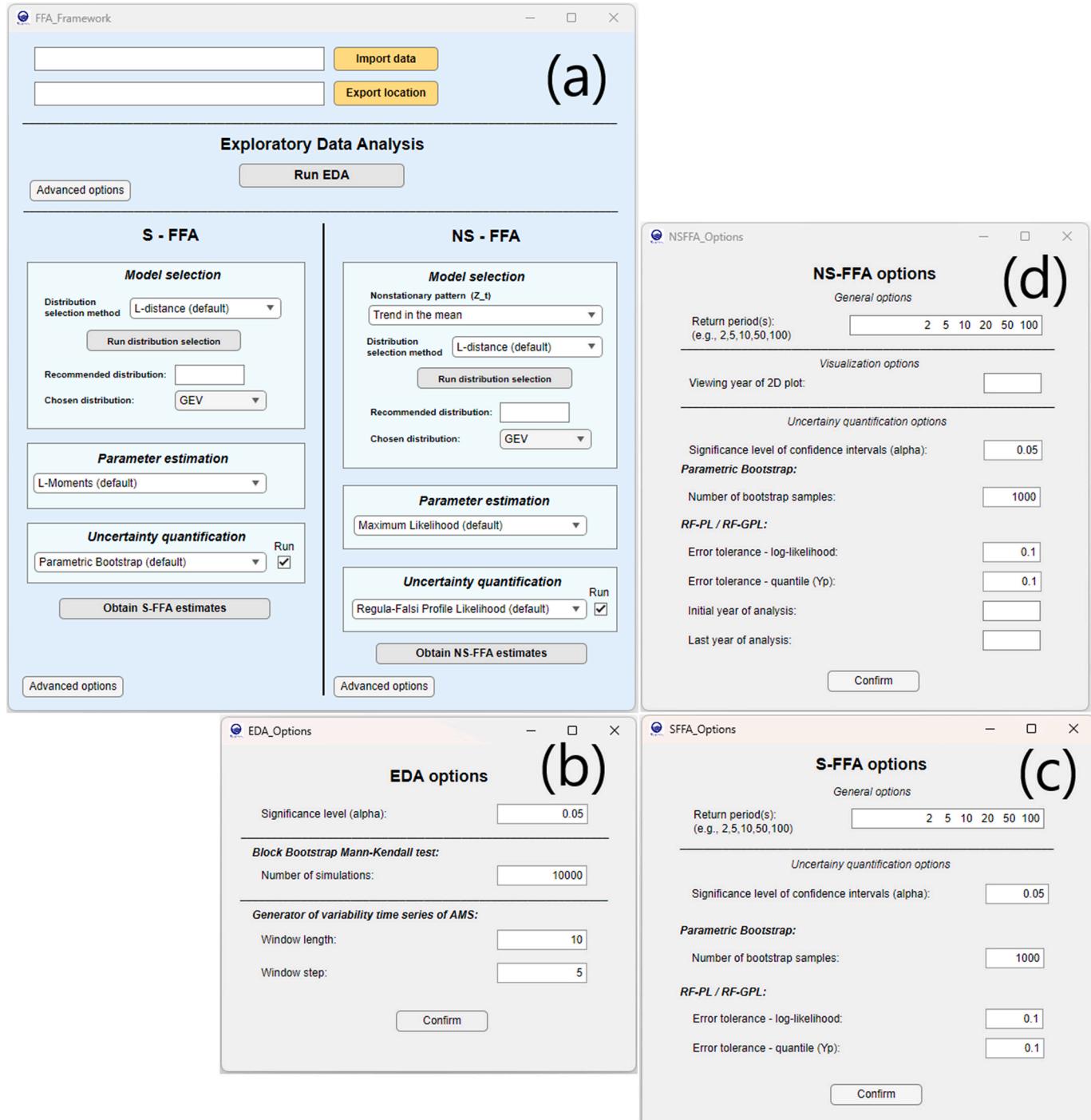


Fig. 3. FFA Framework GUI – (a) the main interface containing data input and result folder, the EDA sub-module, and FFA using S-FFA and NS-FFA; (b)–(d) the “Advanced options” for modellers to set the hyperparameters for the EDA, S-FFA, and NS-FFA, respectively.

shown in Figs. 1 and 2, modellers have the flexibility to intervene in various aspects, including the consideration of change point(s), the selection of the FFA approach, and the choice of metrics for distribution selection and the methods for model parametrization and uncertainty quantification. While the GUI provides default approaches and methods as recommendations, modellers can make their own selections from the readily available options. Moreover, setting hyperparameters, such as the significance level of confidence level, the number of simulations for the BB-MK test, and the number of samples for the bootstrap method, can be modified by modellers through the “Advanced options” as shown in Fig. 3 (b)–(d) that indicate their default values.

5. Application on real study cases

The application of the proposed framework was demonstrated using publicly available daily AMSs from Environment and Climate Change Canada (ECCC). The framework was implemented under both stationary (Fig. 4(a)) and nonstationary scenarios (Fig. 4(b)–(e)). In the nonstationary scenarios, various nonstationary signatures include a change point (Fig. 4(b)), a significant temporal trend in the mean (Fig. 4(c)), a significant temporal trend in the standard deviation (Fig. 4(d)), and significant temporal trends in both the mean and standard deviation (Fig. 4(e)). All the applications considered a significance level of 5%.

5.1. Application 1: S-FFA

This application employed the AMS of daily streamflow records at the hydrometric station ECCC-07BE001 Athabasca River at Athabasca (Fig. 4 (a)). The flood series displayed insignificant change points (according to the Pettitt and MKS tests), and insignificant temporal trends in the mean (using the MK test, in the absence of serial correlation) and variability (using the White and MW-MK tests). Thus, the EDA sub-module supported the stationarity assumption, and consequently, S-FFA was conducted on the entire AMS.

As shown in Fig. 5 (a) and (b), the distribution selection sub-module consistently supported that the GEV distribution was the best-fit distribution when using all three selection metrics. Thus, a time-invariant

GEV distribution was adopted to perform S-FFA in this application. Fig. 5 (c) presents the quantile estimates obtained using the default parameter estimation and uncertainty quantification methods for S-FFA. As expected, the uncertainty is higher when assessing more extreme floods, indicating lower reliability in the estimation of rarer flood events. Fig. 5 (d) presents the model performance.

5.2. Application 2: NS-FFA with change point

The flood series used in this application is the AMS of daily streamflow records measured at the hydrometric station ECCC-08NH021 Kootenai River at Porthill (Fig. 4(b)). The EDA detected two significant change points. The first is an abrupt change point that occurred in 1972 (p-value of <0.01 in the Pettitt test), coinciding with the completion of the Libby Dam. The second change point suggests the beginning of a trend in 1985 (p-value of 0.01 in the MKS test). The Libby Dam is located upstream of the Porthill station on the Kootenai River and is operated for hydropower generation and flood control since 1972 (Barton et al., 2004). The dam has a height of 130 m and creates a reservoir spanning 145 km along the Canada-USA border. Several studies have documented the significant impacts of the Libby Dam on hydrological regimes, maximum flows, and local ecosystems (Burke et al., 2009; McDonald et al., 2010; Paragamian et al., 2001). As a result, the detected change point in 1972 by the EDA is justified physically. To avoid small sample sizes and in turn high analysis uncertainty, only the change point in 1972 was considered. The first subperiod spanning 45 years (1928–1972) showed no significant temporal trends in the mean (MK test, without serial correlation) and variability (White and MW-MK tests). The second subperiod covering 46 years (1973–2018) showed a temporal trend in the mean (MK and BB-MK tests, the presence of lag-1 serial correlation), while no significant temporal trend in the variability was detected (White and MW-MK tests). However, the PP and KPSS tests failed to verify that the trend in the mean in the second subperiod is deterministic. As a result, piece-wise NS-FFA was performed using time-invariant distributions in each subperiod of the flood series. It is worth highlighting that when reservoir operations induce non-stationarity in the flood statistics of regulated rivers (e.g., by dams), the

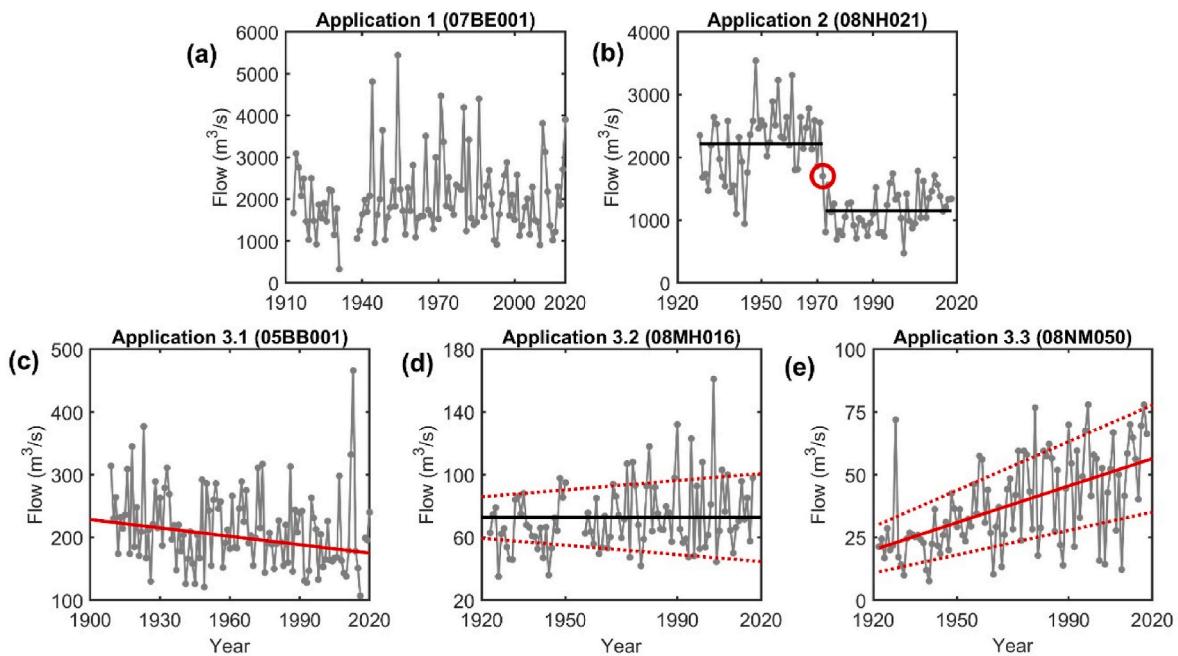


Fig. 4. Six flood series (daily AMSs) used to demonstrate the application of the FFA framework. The flood series exhibit (a) stationary behavior, as well as nonstationary signatures in the form of (b) a significant change point, (c) a significant temporal trend in the mean, (d) a significant temporal trend in the standard deviation, and (e) significant temporal trends in both the standard deviation and the mean. The solid and dotted red lines denote temporal trends in the mean and standard deviation, respectively, while the solid black lines indicate a time-invariant mean. The red circle indicates the location of the detected change point.

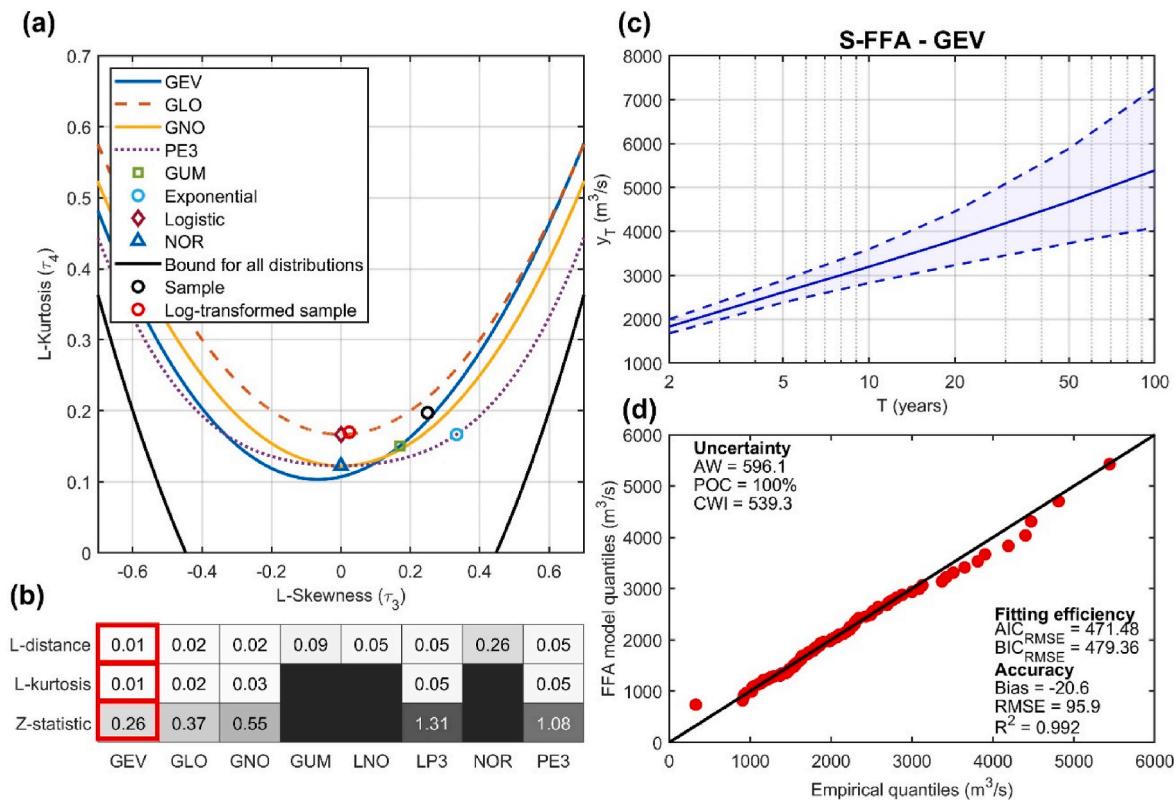


Fig. 5. Results when using S-FFA for the AMS at ECCC-07BE001 Athabasca River at Athabasca – (a) L-moment ratio diagram; (b) calculated L-distance, L-kurtosis, and Z-statistic for the distribution selection (red boxes highlight the best-fit distribution selected by each metric, while black cells indicate the non-applicable two-parameter distributions when adopting the L-kurtosis and Z-statistic metrics); (c) frequency estimates using the selected GEV distribution and default methods for the S-FFA (i.e., L-moments and parametric bootstrap) – point quantile estimates and confidence intervals are indicated by solid and dashed blue lines, respectively; and (d) model performance assessment.

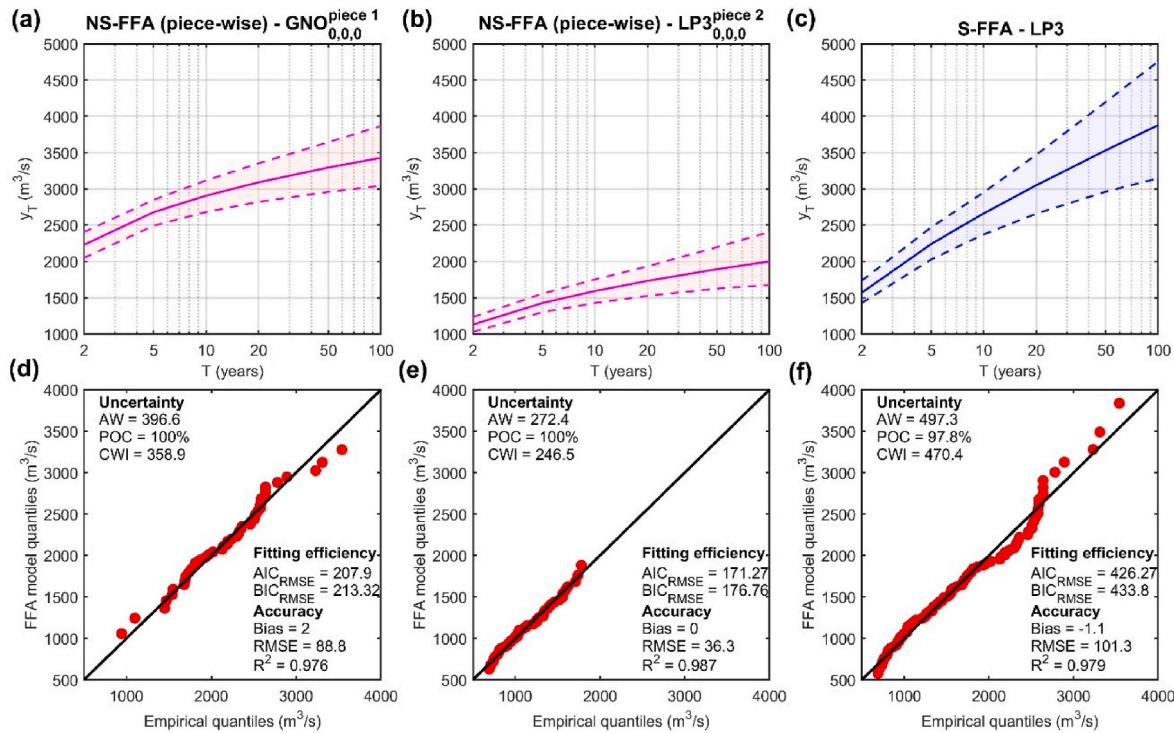


Fig. 6. Frequency estimates and model performance for the AMS at ECCC-08NH021 Kootenai River at Porthill when considering and ignoring an abrupt change point: (a) and (d) for the first subperiod (1928–1972), (d) and (e) second subperiod (1973–2018) in piece-wise NS-FFA, and (c) and (f) for the entire period in S-FFA (1928–2018).

reservoir indices can potentially serve as physical covariates in NS-FFA (López and Francés, 2013; Su and Chen, 2019). However, the sample size might impose constraints for a reliable NS-FFA, especially for recently constructed dams where the data available is limited. It is also worth noting that conducting the EDA again in each subperiod reveals change points in 1945 and 1994 in each subperiod. However, performing the EDA repeatedly for each split AMS (if change points are physically justified) would result in too small sample sizes for meaningful FFA. This would consequently compromise the reliability of the analysis. For example, a minimum of 30 years of data is typically required in S-FFA. Therefore, only the change point detected in the EDA conducted on the entire AMS was considered.

The GNO and LP3 were selected as the best-fit distributions for the first and second subperiods, respectively, according to all three selection metrics. These two distributions were used herein. For comparison purposes, S-FFA was also performed for the entire period ignoring the change point in 1972, for which the LP3 distribution was selected by the three metrics. The default parameter estimation and uncertainty quantification methods for S-FFA were employed in both piece-wise NS-FFA and S-FFA. Fig. 6 shows the frequency estimates and model performance of the piece-wise NS-FFA for each subperiod as well as those of the S-FFA. As shown in Fig. 6, the second subperiod of the piece-wise NS-FFA had lower quantiles compared to the first subperiod, indicating a reduction of flood hazards due to the operation of the Libby dam. In addition, the piece-wise NS-FFA yielded higher quantiles for lower T s and lower quantiles for higher T s for the first subperiod compared to the S-FFA. In contrast, the quantiles estimated in the second subperiod in the piece-wise NS-FFA were consistently lower than those in the S-FFA. These results reflected that ignoring the change point in the FFA could overlook the influence of the Libby Dam and consequently overestimate the frequency estimates. Furthermore, the piece-wise NS-FFA exhibited higher or approximately equivalent accuracy (i.e., lower RMSE and equivalent R^2) compared to the S-FFA. Notably, despite the smaller sample sizes of the two subperiods, the consideration of the change point in the piece-wise NS-FFA resulted in more reliable frequency estimates as it yielded a lower uncertainty (i.e., lower AW, CWI, and higher POC) than the S-FFA.

5.3. Application 3: NS-FFA with temporal trends

This application used the AMSs of daily streamflow records at the hydrometric stations ECCC-05BB001 Bow River at Banff (Fig. 4 (c), Application 3.1), ECCC-08MH016 Chilliwack River at the outlet of Chilliwack Lake (Fig. 4 (d), Application 3.2), and ECCC-08NM050 Okanagan River at Penticton (Fig. 4 (e), Application 3.3). The results of the EDA are shown in Table 1, demonstrating significant and deterministic nonstationary signatures. Specifically, in Application 3.1 (ECCC-05BB001 Bow River at Banff), the EDA revealed a significant and deterministic temporal trend in the mean but not in the variability. The EDA also detected a significant abrupt change point (Pettitt test) in 1974. The change point was disregarded due to the lack of a physical justification. The Bow River is unregulated upstream of this station. This station is part of the Canadian Reference Hydrometric Basin Network (RHBN), which consists of a set of stations with long records and minimal human impacts intended for climate change studies. Hence, the detected trend in the mean was considered a nonstationary signature of climate change. In Application 3.2 (ECCC-08MH016 Chilliwack River at the outlet of Chilliwack Lake), a significant temporal trend in the variability was identified, whereas no significant trend in the mean or change points were detected. This stream gauge station located at the Chilliwack River is also unregulated and a member of the RHBN. Thus, the observed nonstationary signature in this dataset was ascribed to climate change as well. In Application 3.3 (ECCC-08NM050 Okanagan River at Penticton), a significant deterministic temporal trend was detected in the mean (despite serial correlation), and a trend was also identified in the variability of the time series. Although the Pettitt test

Table 1

Summary of the significance (p-value) of statistical test results obtained from the EDA sub-module in Applications 3.1 (ECCC-05BB001 Bow River at Banff), 3.2 (ECCC-08MH016 Chilliwack River at the outlet of Chilliwack Lake), and 3.3 (ECCC-08NM050 Okanagan River at Penticton) along with their interpretation. Detected significant nonstationary signatures at the 95% confidence level are highlighted in boldface.

Application		3.1	3.2	3.3
Detection of change point(s)	Pettitt test	0.02	0.11	< 0.01
	MKS test	>0.99	0.16	>0.99
	Interpretation	Abrupt change point ^a	No change point	Abrupt change point ^a
Detection of trend in mean	MK test	0.01	0.07	< 0.01
	BB-MK test	N. A.	N. A.	< 0.01
	Interpretation	Significant trend	No trend	Significant trend
Estimation of trend in mean (if any)	Sen's intercept	228.13	N. A.	19.33
	Sen's slope	-0.49	N. A.	0.40
Identification of type of trend in mean (if any)	PP test	0.03	N. A.	0.02
	KPSS test	0.10	N. A.	0.10
	Interpretation	Deterministic trend	N. A.	Deterministic trend
Detection of trend in standard deviation	White test	0.09	0.12	0.13
	MW-MK test	0.98	< 0.01	< 0.01
Estimation of trend in standard deviation (if any)	Interpretation	No trend	Significant trend	Significant trend
	Sen's intercept	N. A.	9.89	8.49
	Sen's slope	N. A.	1.13	0.67

^a The detected abrupt change points were disregarded in these applications due to the lack of physical justification and/or the uncertain detection.

detected the presence of a potential change point, its statistic series exceeded the critical value and exhibited high values from 1945 to 1980, suggesting that any point within this period could be a significant change point. Hence, no change point was considered due to uncertain detection. The Okanagan River upstream of the station has been regulated since 1914 due to the construction of the first dam, followed by a second dam in 1920, and a regulation system in the early 1950s consisting of four dams and 38 km of engineered channel (Symonds, 2000). Rapid human settlement, development, and agricultural activity have occurred in the watershed. Apart from the trends detected in the AMS, nonstationary signatures (i.e., upward trends) in the average annual flow, as well as seasonal and monthly maximum flows, have been reported at this site (Rayne and Forest, 2010, 2011). Therefore, the detected nonstationary signatures (i.e., trends in the mean and variability) at this site could be resulting from the impacts of the anthropogenic activity and/or possible climate change. All these advocated the use of NS-FFA at these three sites.

These time series were decomposed according to their detected nonstationary patterns (i.e., significant temporal trends in the mean and/or standard deviation) and estimated Z_t in the EDA. The GNO distribution was consistently selected for Applications 3.1 and 3.2, while the GLO distribution was identified as the best-fit distribution for Application 3.3 according to all three distribution selection metrics. Thus, the NS-FFA models $GNO_{1,0,0}$, $GNO_{1,1,0}$, and $GLO_{1,1,0}$ were determined for Applications 3.1, 3.2, and 3.3, respectively, based on the nonstationary patterns detected in the EDA. These NS-FFA models reflect floods' statistics continually changing over time due to the effect of nonstationarity. In these applications, time was used as the covariate to depict the progressive response of the hydrological system to the time-varying causative physical drivers behind the nonstationarity, as discussed in Section 3.1.2. Fig. 7 (a) – (c) present the NS-FFA estimates over the record period in all three applications. As illustrated in these figures,

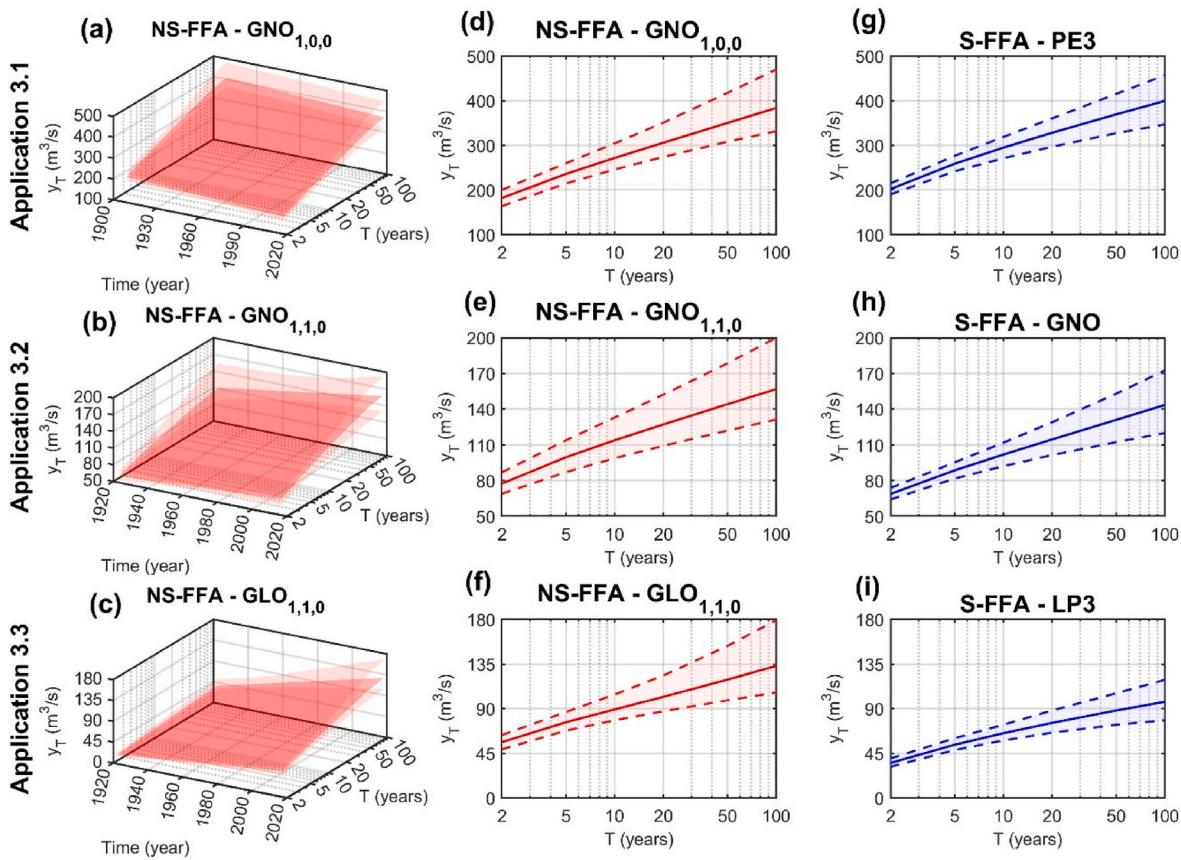


Fig. 7. Results of the framework application under the scenarios of the presence of temporal trends: (a)–(c) temporal variations of quantile estimates over the record periods in NS-FFA in Applications 3.1 (ECCC-05BB001 Bow River at Banff), 3.2 (ECCC-08MH016 Chilliwack River at the outlet of Chilliwack Lake), and 3.3 (ECCC-08NM050 Okanagan River at Penticton), respectively; (d)–(f) quantile estimates at the end of the observational period in NS-FFA in Applications 3.1, 3.2, and 3.3, respectively; and (g)–(i) quantiles estimated by S-FFA in Applications 3.1, 3.2, and 3.3, respectively. The point estimates are represented by dark solid lines, while the confidence intervals (upper and lower bounds) are shown by light or dashed lines. Note that in (b), the RF-PL method was numerically unstable at a few points, which were estimated through trial and error to create the continuous surface.

the quantiles decrease over time in Application 3.1 and increase over time in Applications 3.2 and 3.3, all of which well align with the nonstationary patterns identified. In addition, compared to Application 3.2, the quantiles in Application 3.3 increase over time at a more pronounced rate. The time series in Application 3.3 presents the more complex nonstationary pattern, i.e., temporal trends in both the mean and variability. Besides, the width of the confidence interval varies over time, indicating the need for assessing the uncertainty level at the time (s) of interest. In practice, the most up-to-date estimate is commonly of greater interest, for example, to assess flood hazard or existing infrastructure. Fig. 7 (d)–(f) present the NS-FFA quantile estimates at the end of the observational period in all three applications. For comparison purposes, the estimates by the S-FFA are also illustrated in Fig. 7 (e)–(g). Comparing the estimates of the NS-FFA at the end of the observational period and the S-FFA, the NS-FFA produced lower quantiles for Application 3.1 and higher quantiles for Applications 3.2 and 3.3. These results reveal that the use of S-FFA would lead to either overestimation (e.g., in Application 3.1) or underestimation (e.g., in Applications 3.2 and 3.3) due to ignoring the nonstationarity. All these argue the importance of considering the nonstationarity in FFA to more effectively assess the flood hazard in the context of a changing environment. Yet, the overall increased confidence intervals in the NS-FFA in the three applications could be ascribed to the use of more complex models due to the incorporation of the nonstationary structure compared to the S-FFA.

6. Framework limitations and future research and development

While the proposed framework incorporates state-of-the-art methods for NS-FFA, there are still opportunities for improvement. One of the challenges posed by nonstationarity is its impact on statistical inference. The ergodic statistical property, which is necessary for reliable statistical inference from a time series, does not hold under nonstationary conditions (Koutsoyiannis and Montanari, 2015; Serinaldi et al., 2018). This framework is currently formulated to perform either S-FFA or NS-FFA by assessing the stationarity assumption through the EDA. The reliability of the EDA results heavily depends on the sample size of the AMS. Moreover, the EDA of the current framework cannot distinguish the presence of non-linear deterministic nonstationary signatures in the mean of the AMS or deterministic nonstationary signatures in higher-order moments of the AMS. It has been well acknowledged that the attribution of nonstationarity is critical to confirm the violation of the stationarity condition and thus support the use of NS-FFA. Also, the nonstationarity attribution could facilitate the incorporation of physical covariates in NS-FFA, which would be the physical driver(s) behind the nonstationarity. However, methodologies for nonstationarity attribution are still in their early stages of development, and further research is still needed to refine this aspect. Consequently, the use of deductive reasoning is recommended for diagnosing nonstationarity in the underlying system (Koutsoyiannis and Montanari, 2015; Serinaldi et al., 2018) and was employed in the current framework. The addition of nonstationarity attribution into the framework is recommended when suitable methodologies are available.

Regarding the methods adopted in the framework, the ML and RF-PL methods are advantageous for modeling the nonstationarity through the principle of likelihood. However, these methods might be subject to numerical instability when estimating distribution parameters and quantifying uncertainty, respectively. These issues are particularly more likely to occur when dealing with time series of short sample sizes and heavy tails, and/or when exploring large return periods (Obeysekera and Salas, 2014; Vidrio-Sahagún and He, 2022a). To overcome these potential issues, the framework provides alternative methods, such as L-moments and GML methods as well as the bootstrap and RF-GPL methods. However, the L-moments method is only applicable to S-FFA, while the GML and RF-GPL methods are only devised for the GEV distribution. Thus, advancements in these methods for other commonly used probability distribution families in NS-FFA are recommended for future studies. These advancements would resolve or mitigate the issue of numerical instability. Furthermore, the FFA framework can be extended to include approaches that harness data beyond the AMS. For example, the Generalized Pareto distribution can be included to analyze POT data, for which suitable guidelines for selecting the threshold would be indispensable. Another potential approach for POT data is the two-step Kappa approach (O'Shea et al., 2023), which is however not readily available for NS-FFA yet. In addition, the Metastatistical approach (Marani and Ignaccolo, 2015) and its simplified version (Marra et al., 2019) are attractive alternatives, as they leverage information from ordinary independent events and recently have been applied and extended for explicit NS-FFA (Vidrio-Sahagún et al., 2023b; Vidrio-Sahagún and He, 2022b). However, standard practices for separating ordinary streamflow events and identifying a suitable ordinary-event distribution are needed to incorporate these approaches into the framework.

Furthermore, the current framework is devised to analyze historical flood series. In the context of NS-FFA, several strategies can be followed to generate future (out-of-sample) predictions (François et al., 2019; Vidrio-Sahagún and He, 2022c; Wasko et al., 2021), including (a) directly fitting the NS-FFA model to projections of floods on the future horizon (e.g., produced by hydrological models feed by downscaled climate model outputs); (b) fitting the NS-FFA model to historical flood series with selected physical covariate(s), and then applying it for the future using the projections of the physical covariate(s); and (c) fitting the NS-FFA model to historical flood series with the temporal covariate and then extrapolating over time. No consensus on which strategy is preferable has been reached yet in the research community. The implementation of these strategies would rely on factors such as (a) the availability and reliability of flood projections, (b) the availability and reliability of projections of the selected physical covariate(s) and the validity of both the selected covariate(s) and its link function with the model parameters in the future, and/or (c) the assumption that temporal trends observed in the past will remain in the future. Thus, to extend the framework to enable future FFA projections, further research is desired to identify a practically preferable strategy.

Lastly, for NS-FFA, the framework provides the time-dependent quantiles associated with selected T 's, also known as effective return levels (ERL) (Katz et al., 2002). However, there are other nonstationary metrics used for design and hazard assessment, such as expected waiting time (EWT) (Olsen et al., 1998; Wigley, 2009), expected number of exceedances (ENE) (Parey et al., 2007, 2010), effective return period (ERP) (Katz et al., 2002), risk of failure (R) or reliability (R_f) (Read and Vogel, 2015; Salas and Obeysekera, 2014; Serinaldi and Kilsby, 2015), design life level (DLL) (Rootzén and Katz, 2013), and average design life level (ADLL) (Yan et al., 2017), among others. All metrics have their particularities. Some metrics (e.g., EWT) require extrapolation of the NS-FFA model to achieve convergence, which is not guaranteed and depends on the distribution tail and trend direction. DLL and ADLL take

into consideration the design life of the project. A module for estimating the metrics for design and hazard assessment under nonstationarity can be integrated into the framework as future development. For this addition, the selection of metrics should consider their specific requirements and the feasibility of meeting those requirements (e.g., the availability of future NS-FFA projections, if needed).

7. Conclusions

This paper developed a practice-oriented framework to guide users in conducting the FFA while accounting for potential nonstationarity. The framework consists of three main modules: a) data processing for the EDA and/or NS-FFA model determination (if required), b) selection of a suitable FFA approach (i.e., either the S-FFA or NS-FFA) aided by the EDA, and c) performing FFA, including determining the FFA model, estimating distribution parameters, quantifying uncertainty, and assessing model performance. Compared to existing frameworks and tools for the FFA, the developed framework offers several key features that address practical gaps and enable consistency. Firstly, the framework encompasses all required steps for the FFA and offers flexible options for methods and metrics, allowing for systematic and repeatable workflows. Secondly, the framework includes documented decision points that allow for modellers' intervention (e.g., the selections of the FFA approach, distributions, methods, etc.) while maintaining reproducibility. Lastly, it employs state-of-the-art methods to deal with nonstationarity, addressing the growing demand from practitioners and government agencies in the face of changes in climate, land use/cover, and water management practices. The source codes of the framework and a user-friendly standalone GUI are publicly available. For future developments, it is recommended to advance and extend the framework by including nonstationarity attribution, FFA future projections, and improved methods (especially for the NS-FFA) as they become available.

Software availability

Name of software: FFA Framework 1.0.0.

Developers: Cuauhtémoc Tonatiuh Vidrio-Sahagún, Jake Ruschkowski, Jianxun He, and Alain Pietroniro.

Contact information: cuauhitemoc.vidriosah@ucalgary.ca; Year first available: 2024.

Software required: Windows 10 or later versions as Operating System (OS) for standalone GUI; Matlab 2021a or higher for source codes.

Availability and cost: Freely available at <https://zenodo.org/records/8012096>.

Program language: Matlab.

Program size: 2.39 GB.

Declaration of competing interest

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

Data availability

Data used are publicly available.

Acknowledgments

This work was funded by the Flood Hazard Identification and Mapping Program of Environment and Climate Change Canada as well as the Canada Research Chair (Tier 1) awarded to Dr. Pietroniro.

Appendix F. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envsoft.2024.105940>.

Appendix A. Candidate distributions

Table A. Cumulative distribution functions and quantile functions of the eight candidate distributions considered in the FFA framework.

Distribution	Cumulative distribution (F_Y) and quantile (y_T) functions
GEV	$F(y; \theta) = \exp\left(-\left[1 + \kappa\left(\frac{y-\xi}{\alpha}\right)\right]^{\frac{1}{-\kappa}}\right), \kappa \neq 0$ $y_T = F_Y^{-1}(p; \theta) = \xi - \frac{\alpha}{\kappa}\left[1 - \left\{-\ln(1-p)\right\}^{-\kappa}\right], \kappa \neq 0$
GLO	$F(y; \theta) = \frac{1}{1+e^{-z}}$; where $z = -\kappa^{-1}\log\left(1 - \frac{\kappa(y-\xi)}{\alpha}\right)$, $\kappa \neq 0$ $y_T = F_Y^{-1}(p; \theta) = \xi + \frac{\alpha}{\kappa}\left[1 - \left\{\frac{p}{1-p}\right\}^{\kappa}\right], \kappa \neq 0$
GNO	$F(y; \theta) = \Phi\left(-\kappa^{-1}\log\left(1 - \frac{\kappa(y-\xi)}{\alpha}\right)\right), \kappa \neq 0$ $y_T = F_Y^{-1}(p; \theta) = \xi + \frac{\alpha}{\kappa}[1 - \exp\{-\kappa\Phi^{-1}(1-p)\}], \kappa \neq 0$
GUM	$F_Y(y; \theta) = \exp\left(-\exp\left(-\left(\frac{y-\xi}{\alpha}\right)\right)\right)$ $y_T = F_Y^{-1}(p; \theta) = \xi - \alpha \log[-\log(1-p)]$
NOR (LNO)	$F(y; \theta) = \Phi\left(\frac{y-\xi}{\alpha}\right)$ $y_T = \exp(\xi + \alpha\Phi^{-1}(1-p))$
PE3 (LP3)	Note that in LNO, the log-transformed $y^* = \log(y)$ is used instead of y . $F(y; \theta) = \begin{cases} \frac{G\left(d, \frac{y-c}{b}\right)}{\Gamma(d)} & \kappa > 0 \\ 1 - \frac{G\left(d, \frac{c-y}{b}\right)}{\Gamma(d)} & \kappa < 0 \end{cases}$ $y_{p_e} = \xi + \alpha K_p(\kappa), K_p(\kappa) = \frac{2}{\kappa} \left\{ \left[1 - \left(\frac{\kappa}{6}\right)^2 + \frac{\kappa z_p}{6}\right]^3 - 1 \right\}$ where $d = 4/\kappa^2$, $b = \frac{\alpha \kappa }{2}$, and $c = \xi - 2\alpha/\kappa$; Note that in LP3, the log-transformed $y^* = \log(y)$ is used instead of y .

Where p is the exceedance probability, i.e., $p = 1 - F_Y(y; \theta)$; $\Phi(\bullet)$ is the cumulative standard Normal distribution; $\Phi^{-1}(\bullet)$ is the inverse standard Normal distribution; $\Gamma(\bullet)$ is the gamma function; $G(\bullet)$ is the incomplete gamma function; $K_p(\kappa)$ is the frequency factor given by the Wilson-Hilferty transformation; and z_p is the p th quantile of the standard normal distribution.

Appendix B. Estimation of sample L-moments and ratios

The sample L-moments and ratios used for both the model selection and parametrizations sub-modules of the Flood Frequency Analysis module are calculated as follows. Let $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ be the ordered sample in ascending order of magnitude. The sample L-moments are calculated by the following steps:

1. Calculate the unbiased estimators of the probability weighted moments (b_r , for $r = 0, 1, 2$ and 3):

$$b_r = n^{-1} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_{1:n} \quad (\text{B.1})$$

2. Estimate the first four sample L-moments by

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 6b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned} \quad (\text{B.2})$$

3. Compute the sample L-moment ratios (t_r) by

$$t_r = \frac{l_r}{l_2} \quad (\text{B.3})$$

where t_3 and t_4 correspond to the L-skewness and L-kurtosis, respectively.

4. If estimating the parameters of a chosen distribution, equate the sample L-moments and L-moment ratios to their population counterparts ($\lambda_1, \lambda_2, \dots, \tau_3$, and τ_4) and use their relationships with the theoretical distribution parameters (Hosking, 1990; Hosking and Wallis, 2009).

Appendix C. Z-statistic

The Z statistic used for the model selection sub-module in the Flood Frequency Analysis module is calculated as follows.

1. Fit the four-parameter kappa distribution to the sample AMS with sample size n , considering its sample L-moment ratios (denoted herein as t_3^s) and $l_1 = 1$.
2. Generate N_{sim} synthetic series from the fitted kappa distribution with the same n .
3. Calculate the L-skewness $t_3^{[i]}$ and L-kurtosis $t_4^{[i]}$ of each synthetic dataset $i = 1, 2, \dots, N_{sim}$.
4. Calculate the bias and standard deviation of t_4^s , respectively, by:

$$B_4 = N_{sim}^{-1} \sum_{i=1}^{N_{sim}} (t_4^{[i]} - t_4^s) \quad (C.1)$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{ \sum_{i=1}^{N_{sim}} (t_4^{[i]} - t_4^s)^2 - N_{sim} B_4^2 \right\} \right]^{\frac{1}{2}} \quad (C.2)$$

5. Determine the theoretical L-kurtosis (t_4^{Dist}) of all the candidate distributions fitted to the sample AMS, i.e., considering that $t_3^{Dist} = t_3^s$.
6. Calculate the Z statistic for each candidate distribution (Z^{Dist}) by

$$Z^{Dist} = \frac{(t_4^{Dist} - t_4^{[i]} + B_4)}{\sigma_4} \quad (C.3)$$

7. The candidate distribution with smallest $|Z^{Dist}|$ is the best-fit distribution. Also, if $|Z^{Dist}| \leq 1.96$, the distribution is deemed to have an acceptable goodness-of-fit at the 5% significance level.

Appendix D. Parametric bootstrap

The parametric bootstrap resampling method generates samples from a parametric model fitted to the data. This method is implemented by the following steps.

1. Obtain a parametric estimate \hat{F} of the population F using the available sample (AMS).
2. Draw N_{samp} synthetic samples from the parametric estimate \hat{F} .
3. Estimate the model parameters $\hat{\theta}_i, i = 1, 2, \dots, N_{samples}$ for each synthetic sample.
4. Estimate the quantiles $\hat{y}_i = F^{-1}(p; \hat{\theta}_i)$ for each synthetic sample.
5. Derive the confidence intervals at a $(1 - \rho)$ 100% significance level using the $\rho/2$ and $1 - \rho/2$ percentiles from the set of N_{samp} estimated \hat{y}_i s.

Appendix E. Reparametrized log-likelihood functions for the RF-PL method

Table E

Reparametrized log-likelihood functions ($\ell(y_T, \theta_{-j})$) for all candidate distributions included in the framework. The θ_{-j} denotes all parameters of $\ell(\bullet)$ except y_T .

Model	$\ell(y_T, \theta_{-j})$ function
GEV	$\ell(y_T, \alpha, \kappa) = -n \log \alpha - \left(1 + \frac{1}{\kappa}\right) \sum \log \left[\frac{\kappa}{\alpha} (y_t - y_T) + \{-\log(1-p)\}^{-\kappa} \right] - \sum \left[\frac{\kappa}{\alpha} (y_t - y_T) + \{-\log(1-p)\}^{-\kappa} \right]^{-1/\kappa}$
GLO	$\ell(y_T, \alpha, \kappa) = -n \log \alpha + \left(\frac{1}{\kappa} - 1\right) \sum \log \left[\left\{ \frac{p}{1-p} \right\}^\kappa - \frac{\kappa}{\alpha} (y_t - y_T) \right] - 2 \sum \log \left[1 + \left(\left\{ \frac{p}{1-p} \right\}^\kappa - \frac{\kappa}{\alpha} (y_t - y_T) \right)^{\frac{1}{\kappa}} \right]$
GNO	$\ell(y_T, \alpha, \kappa) = -n \log \alpha - \frac{n}{2} \log(2\pi) - \sum \left\{ \log \left\{ \exp[-\kappa \Phi^{-1}(1-p)] - \frac{\kappa}{\alpha} (y_t - y_T) \right\} + \frac{\left(\log \left\{ \exp[-\kappa \Phi^{-1}(1-p)] - \frac{\kappa}{\alpha} (y_t - y_T) \right\} \right)^2}{2\kappa^2} \right\}$
GUM	$\ell(y_T, \alpha) = -n \log \alpha - \sum \left(\frac{y_t - y_T - \alpha \log\{-\log(1-p)\}}{\alpha} \right) - \sum \exp \left(- \left[\frac{y_t - y_T - \alpha \log\{-\log(1-p)\}}{\alpha} \right] \right)$
LNO	$\ell(y_T, \alpha) = -n \log \alpha - \frac{n}{2} \log(2\pi) - \sum \log(y_t) - \frac{1}{2\alpha^2} \sum \{ (\log(y_t) - \log(y_T) + \alpha \Phi^{-1}(1-p))^2 \}$
LP3	$\ell(y_T, \alpha, \kappa) = -\frac{4n}{\kappa^2} \log \left(\frac{\alpha \kappa }{2} \right) - n \log \left(\Gamma \left[\frac{4}{\kappa^2} \right] \right) + \sum \left\{ \left(\frac{4}{\kappa^2} - 1 \right) \log \left(\log(y_t) - \log(y_T) + \alpha \left\{ K_p(\kappa) + \frac{2}{\kappa} \right\} \right) - 2 \left(\frac{\log(y_t) - \log(y_T) + \alpha \left\{ K_p(\kappa) + \frac{2}{\kappa} \right\}}{\alpha \kappa } \right) - \log(y_t) \right\}, \text{ for } \kappa > 0;$
	$\ell(y_T, \alpha, \kappa) = -\frac{4n}{\kappa^2} \log \left(\frac{\alpha \kappa }{2} \right) - n \log \left(\Gamma \left[\frac{4}{\kappa^2} \right] \right) + \sum \left\{ \left(\frac{4}{\kappa^2} - 1 \right) \log \left(\log(y_T) - \log(y_t) - \alpha \left\{ K_p(\kappa) + \frac{2}{\kappa} \right\} \right) - 2 \left(\frac{\log(y_T) - \log(y_t) - \alpha \left\{ K_p(\kappa) + \frac{2}{\kappa} \right\}}{\alpha \kappa } \right) - \log(y_t) \right\}, \text{ for } \kappa < 0.$
NOR	$\ell(y_T, \alpha) = -n \log \alpha - \frac{n}{2} \log(2\pi) - \frac{1}{2\alpha^2} \sum \{ (y_t - y_T + \alpha \Phi^{-1}(1-p))^2 \}$

(continued on next page)

Table E (continued)

Model	$\ell(y_T, \theta_{-j})$ function
PE3	$\ell(y_T, \alpha, \kappa) = -\frac{4n}{\kappa^2} \log\left(\frac{\alpha \kappa }{2}\right) - n \log\left(\Gamma\left[\frac{4}{\kappa^2}\right]\right) + \sum \left\{ \left(\frac{4}{\kappa^2} - 1 \right) \log\left(y_t - y_T + \alpha\left\{K_p(\kappa) + \frac{2}{\kappa}\right\}\right) - 2\left(\frac{y_t - y_T + \alpha\left\{K_p(\kappa) + \frac{2}{\kappa}\right\}}{\alpha \kappa }\right) \right\}, \text{ for } \kappa > 0;$ $\ell(y_T, \alpha, \kappa) = -\frac{4n}{\kappa^2} \log\left(\frac{\alpha \kappa }{2}\right) - n \log\left(\Gamma\left[\frac{4}{\kappa^2}\right]\right) \sum \left\{ \left(\frac{4}{\kappa^2} - 1 \right) \log\left(y_T - y_t - \alpha\left\{K_p(\kappa) + \frac{2}{\kappa}\right\}\right) - 2\left(\frac{y_T - y_t - \alpha\left\{K_p(\kappa) + \frac{2}{\kappa}\right\}}{\alpha \kappa }\right) \right\}, \text{ for } \kappa < 0.$

References

- Abdelmoaty, H.M., Papalexiou, S.M., 2023. Changes of extreme precipitation in CMIP6 projections: should we use stationary or nonstationary models? *J. Clim.* 1–40. <https://doi.org/10.1175/JCLI-D-22-0467.1>.
- Andreadis, K.M., Wing, O.E.J., Colven, E., Gleason, C.J., Bates, P.D., Brown, C.M., 2022. Urbanizing the floodplain: global changes of imperviousness in flood-prone areas. *Environ. Res. Lett.* 17 (10), 104024 <https://doi.org/10.1088/1748-9326/ac9197>.
- Archfield, S.A., Hirsch, R.M., Viglione, A., Blöschl, G., 2016. Fragmented patterns of flood change across the United States. *Geophys. Res. Lett.* 43 (19) <https://doi.org/10.1002/2016GL070590>, 10,232-10,239.
- Ball, J.E., Babister, M.K., Nathan, R., Weinmann, P.E., Weeks, W., Retallick, M., Testoni, I., 2016. Book 1 - scope and philosophy. In: *Australian Rainfall and Runoff-A Guide to Flood Estimation*. Commonwealth of Australia.
- Barbosa, S.M., 2011. Testing for deterministic trends in global sea surface temperature. *J. Clim.* 24 (10), 2516–2522. <https://doi.org/10.1175/2010JCLI3877.1>.
- Barton, G.J., Moran, E.H., Berembrock, C., 2004. Surveying Cross Sections of the Kootenai River between Libby Dam, Montana, and Kootenay Lake, British Columbia, Canada. US Geological Survey.
- Basu, B., Srinivas, V.V., 2013. Formulation of a mathematical approach to regional frequency analysis. *Water Resour. Res.* 49 (10), 6810–6833. <https://doi.org/10.1002/wrcr.20540>.
- Bayazit, M., 2015. Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. *Environ. Processes* 2 (3), 527–542. <https://doi.org/10.1007/s40710-015-0081-7>.
- Blöschl, G., Hall, J., Viglione, A., Perdigão, R.A.P., Parajka, J., Merz, B., et al., 2019. Changing climate both increases and decreases European river floods. *Nature* 573 (7772), 108–111. <https://doi.org/10.1038/s41586-019-1495-6>.
- Blum, A.G., Ferraro, P.J., Archfield, S.A., Ryberg, K.R., 2020. Causal effect of impervious cover on annual flood magnitude for the United States. *Geophys. Res. Lett.* 47 (5) <https://doi.org/10.1029/2019GL086480>.
- Bolívar-Cimé, A., Díaz-Francés, E., Ortega, J., 2015. Optimality of profile likelihood intervals for quantiles of extreme value distributions: application to environmental disasters. *Hydrol. Sci. J.* 60 (4), 651–670. <https://doi.org/10.1080/02626667.2014.897405>.
- Burke, M., Jorde, K., Buffington, J.M., 2009. Application of a hierarchical framework for assessing environmental impacts of dam operation: changes in streamflow, bed mobility and recruitment of riparian trees in a western North American river. *J. Environ. Manag.* 90, S224–S236.
- Burn, D.H., Whitfield, P.H., 2017. Changes in cold region flood regimes inferred from long-record reference gauging stations. *Water Resour. Res.* 53 (4), 2643–2658. <https://doi.org/10.1002/2016WR020108>.
- Burn, D.H., Whitfield, P.H., 2023. Climate related changes to flood regimes show an increasing rainfall influence. *J. Hydrol.* 617, 129075 <https://doi.org/10.1016/j.jhydrol.2023.129075>.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag London, London.
- Cooley, D., 2013. Return periods and return levels under climate change. In: *Extremes in a Changing Climate*. Springer, pp. 97–114.
- Cui, H., Jiang, S., Gao, B., Ren, L., Xiao, W., Wang, M., et al., 2023. On method of regional non-stationary flood frequency analysis under the influence of large reservoir group and climate change. *J. Hydrol.* 618 <https://doi.org/10.1016/j.jhydrol.2023.129255>.
- Cunderlik, J.M., Burn, D.H., 2003. Non-stationary pooled flood frequency analysis. *J. Hydrol.* 276 (1–4), 210–223. [https://doi.org/10.1016/S0022-1694\(03\)00062-3](https://doi.org/10.1016/S0022-1694(03)00062-3).
- de Luca, D.L., Napolitano, F., 2023. *A User-Friendly Software for Modelling Extreme Values: EXTRASTAR (Extremes Abacus for Statistical Regionalization)*. Environmental Modelling & Software, 105622.
- Doocy, S., Daniels, A., Murray, S., Kirsch, T.D., 2013. The human impact of floods: a historical review of events 1980–2009 and systematic literature review. *PLoS Curr.* 5.
- Efron, B., 1992. Bootstrap methods: another look at the jackknife. In: *Breakthroughs in Statistics*. Springer S), New York, NY. Springer.
- El Adlouni, S., Ouarda, T.B.M.J., Zhang, X., Roy, R., Bobée, B., 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resour. Res.* 43 (3), 1–13. <https://doi.org/10.1029/2005WR004545>.
- England Jr., J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A. G., et al., 2019. Guidelines for determining flood flow frequency—Bulletin 17C (Version 1. In: Techniques and Methods. <https://doi.org/10.3133/tm4B5>. Reston, vol. A.
- Faticchi, S., Barbosa, S.M., Caporali, E., Silva, M.E., 2009. Deterministic versus stochastic trends: detection and challenges. *J. Geophys. Res. Atmos.* 114 (18), 1–11. <https://doi.org/10.1029/2009JD011960>.
- François, B., Schlef, K.E., Wi, S., Brown, C.M., 2019. Design considerations for riverine floods in a changing climate – a review. *J. Hydrol.* 574 (September 2018), 557–573. <https://doi.org/10.1016/j.jhydrol.2019.04.068>.
- Franzke, C.L.E., Barbosa, S., Blender, R., Fredriksen, H., Laepple, T., Lambert, F., et al., 2020. The structure of climate variability across scales. *Rev. Geophys.* 58 (2) <https://doi.org/10.1029/2019rg000657>.
- Gilleland, E., Katz, R.W., 2016. extRemes 2.0: an extreme value analysis package in R. *J. Stat. Software* 72 (8). <https://doi.org/10.18637/jss.v072.i08>.
- Hesiodi, S., Arabzadeh, R., Hajibabaei, M., Rauch, W., Kjeldsen, T.R., Prosdocimi, I., et al., 2021. Stationary vs non-stationary modelling of flood frequency distribution across northwest England. *Hydrol. Sci. J.* 00 (00), 1–16. <https://doi.org/10.1080/02626667.2021.1884685>.
- Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc. Ser. C* 52 (1), 105–124. Retrieved from. <http://www.jstor.org/stable/2345653>.
- Hosking, J.R.M., Wallis, J.R., 2009. Regional frequency analysis. *Regional Frequency Anal.* 1–13 <https://doi.org/10.1017/cbo9780511529443.003>.
- Iliopoulos, T., Koutsogiannis, D., Montanari, A., 2018. Characterizing and modeling seasonality in extreme rainfall. *Water Resour. Res.* 54 (9), 6242–6258. <https://doi.org/10.1029/2018WR023360>.
- Kasivananthan, K.S., He, J., Tay, J., Sudheer, K.P., 2019. Enhancement of Model Reliability by Integrating Prediction Interval Optimization into Hydrogeological Modeling. *Water Resources Management*, pp. 229–243. <https://doi.org/10.1007/s11269-018-2099-x>.
- Katz, R.W., 2013. Statistical methods for nonstationary extremes. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., Sorooshian, S. (Eds.), *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*. Springer Netherlands, pp. 15–37.
- Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. *Adv. Water Resour.* 25 (8–12), 1287–1304. [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8).
- Kendall, M.G., 1975. *Rank Correlation Methods, Book Series*. Charles Griffin. Oxford University Press, USA, London.
- Khaliq, M.N., Ouarda, T.B.M.J., Ondo, J.C., Gachon, P., Bobée, B., 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. *J. Hydrol.* 329 (3–4), 534–552. <https://doi.org/10.1016/j.jhydrol.2006.03.004>.
- Khaliq, M.N., Ouarda, T.B.M.J., Gachon, P., Sushama, L., St-Hilaire, A., 2009. Identification of hydrological trends in the presence of serial and cross correlations: a review of selected methods and their application to annual flow regimes of Canadian rivers. *J. Hydrol.* 368 (1–4), 117–130. <https://doi.org/10.1016/j.jhydrol.2009.01.035>.
- Kjeldsen, T.R., Jones, D.A., Bayliss, A.C., 2008. *Improving the FEH Statistical Procedures for Flood Frequency Estimation*. Environment Agency.
- Kjeldsen, T.R., Jones, D.A., Morris, D.G., 2014a. Using multiple donor sites for enhanced flood estimation in ungauged catchments. *Water Resour. Res.* 50 (8), 6646–6657. <https://doi.org/10.1002/2013WR015203>.
- Kjeldsen, T.R., Lamb, R., Blazkova, S.D., 2014b. Uncertainty in flood frequency analysis. In: *Applied Uncertainty Analysis for Flood Risk Management*, pp. 153–197. https://doi.org/10.1142/9781848162716_0008.
- Kobierska, F., Engeland, K., Thorarinsdottir, T., 2018. Evaluation of design flood estimates - a case study for Norway. *Nord. Hydrol.* 49 (2), 450–465. <https://doi.org/10.2166/nh.2017.068>.
- Koutsogiannis, D., Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrol. Sci. J.* 60 (7–8), 1174–1183. <https://doi.org/10.1080/02626667.2014.959959>.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom.* 54 (1–3), 159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-y](https://doi.org/10.1016/0304-4076(92)90104-y).
- Kysely, J., 2008. A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *J. Appl. Meteorol. Climatol.* 47 (12), 3236–3251. <https://doi.org/10.1175/2008JAMC1763.1>.
- Kysely, J., 2010. Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. *Theor. Appl. Climatol.* 101 (3), 345–361. <https://doi.org/10.1007/s00704-009-0190-1>.
- Lagaras, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.* 9 (1), 112–147. <https://doi.org/10.1137/S1064926596303470>.
- Li, J., Liu, X., Chen, F., 2015. Evaluation of nonstationarity in annual maximum flood series and the associations with large-scale climate patterns and human activities.

- Water Resour. Manag. 29 (5), 1653–1668. <https://doi.org/10.1007/s11269-014-0900-z>.
- Li, X., Sang, Y.F., Sivakumar, B., Gil-Alana, L.A., 2021. Detection of type of trends in surface air temperature in China. J. Hydrol. 596 (December 2020), 126061 <https://doi.org/10.1016/j.jhydrol.2021.126061>.
- Lindgren, G., Rootzén, H., Sandsten, Maria, 2013. *Stationary Stochastic Processes for Scientists and Engineers*. Chapman and Hall/CRC.
- López, J., Francés, F., 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrol. Earth Syst. Sci. 17 (8), 3189–3203. <https://doi.org/10.5194/hess-17-3189-2013>.
- Lu, F., Song, X., Xiao, W., Zhu, K., Xie, Z., 2020. Detecting the impact of climate and reservoirs on extreme floods using nonstationary frequency models. *Stochastic Environmental Research and Risk Assessment*. Springer. <https://doi.org/10.1007/s00477-019-01747-2>.
- Luke, A., Vrugt, J.A., AghaKouchak, A., Matthew, R., Sanders, B.F., 2017. Predicting nonstationary flood frequencies: evidence supports an updated stationarity thesis in the United States. Water Resour. Res. 53 (7), 5469–5494. <https://doi.org/10.1002/2016WR019676>.
- Mann, H.B., 1945. Nonparametric tests against trend. *Econometrica: J. Econom. Soc.* 245–259.
- Marani, M., Ignaccolo, M., 2015. A metastatistical approach to rainfall extremes. Adv. Water Resour. 79, 121–126. <https://doi.org/10.1016/j.advwatres.2015.03.001>.
- Marra, F., Zoccatelli, D., Armon, M., Morin, E., 2019. A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. Adv. Water Resour. 127, 280–290. <https://doi.org/10.1016/j.advwatres.2019.04.002>.
- Martins, E.S., Stedinger, J.R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resour. Res. 36 (3), 737–744. <https://doi.org/10.1029/1999WR900330>.
- McDonald, R., Nelson, J., Paragamian, V., Barton, G., 2010. Modeling the effect of flow and sediment transport on white sturgeon spawning habitat in the Kootenai River, Idaho. J. Hydraul. Eng. 136 (12), 1077–1092.
- Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., Hundecha, Y., 2012. HESS Opinions: “More efforts and scientific rigour are needed to attribute trends in flood time series.”. Hydrol. Earth Syst. Sci. 16 (5), 1379–1387. <https://doi.org/10.5194/hess-16-1379-2012>.
- Milly, A.P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Zbigniew, W., Lettenmaier, D.P., et al., 2008. Stationarity is dead: whither water management. Science 319 (5863), 573–574. <https://doi.org/10.1126/science.1151915>.
- Mishra, A., Mukherjee, S., Merz, B., Singh, V.P., Wright, D.B., Villarini, G., et al., 2022. An overview of flood concepts, challenges, and future directions. J. Hydrol. Eng. 27 (6), 1–30. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002164](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002164).
- Mondal, A., Mujumdar, P.P., 2016. Detection of change in flood return levels under global warming. J. Hydrol. Eng. 21 (8), 1–12. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001326](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001326).
- Natural Resources Canada, 2019. Public safety Canada. In: *Federal Hydrologic and Hydraulic Procedures for Floodplain Delineation* (Version 1.). Natural Resources Canada. <https://doi.org/10.4095/299808>.
- Nerantzaki, S., Papalexiou, S.M., 2021. Assessing extremes in hydroclimatology: a review on probabilistic methods. J. Hydrol. 605 (December 2021), 127302 <https://doi.org/10.1016/j.jhydrol.2021.127302>.
- Obeysekera, J., Salas, J.D., 2014. Quantifying the uncertainty of design floods under nonstationary conditions. J. Hydrol. Eng. 19 (7), 1438–1446. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000931](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000931).
- Obeysekera, J., Salas, J.D., 2016. Frequency of recurrent extremes under nonstationarity. J. Hydrol. Eng. 21 (5), 04016005 [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001339](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001339).
- Olsen, J.R., Lambert, J.H., Haimes, Y.Y., 1998. Risk of extreme events under nonstationary conditions. Risk Anal. 18 (4), 497–510. <https://doi.org/10.1111/j.1539-6924.1998.tb00364.x>.
- Önöz, B., Bayazit, M., 2012. Block bootstrap for Mann-Kendall trend test of serially dependent data. Hydrol. Process. 26 (23), 3552–3560. <https://doi.org/10.1002/hyp.8438>.
- Ouarda, T.B.M.J., Charron, C., St-Hilaire, A., 2019. Uncertainty of stationary and nonstationary models for rainfall frequency analysis. Int. J. Climatol. (September), 1–20. <https://doi.org/10.1002/joc.6339>.
- O’Brien, N.L., Burn, D.H., 2014. A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. J. Hydrol. 519 (PB), 2040–2048. <https://doi.org/10.1016/j.jhydrol.2014.09.041>.
- O’Shea, D., Nathan, R., Sharma, A., Wasko, C., 2023. Improved extreme rainfall frequency analysis using a two-step Kappa approach. Water Resour. Res. 59 (4) <https://doi.org/10.1029/2021wr031854>.
- Panagoulia, D., Economou, P., Caroni, C., 2014. Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. Environmetrics 25 (1), 29–43. <https://doi.org/10.1002/env.2252>.
- Paragamian, V.L., Kruse, G., Wakkinen, V., 2001. Spawning habitat of Kootenai River white sturgeon, post-Libby Dam. N. Am. J. Fish. Manag. 21 (1), 22–33.
- Parey, S., Malek, F., Laurent, C., Dacunha-Castelle, D., 2007. Trends and climate evolution: statistical approach for very high temperatures in France. Climatic Change 81 (3–4), 331–352. <https://doi.org/10.1007/s10584-006-9116-4>.
- Parey, S., Hoang, T.T.H., Dacunha-Castelle, D., 2010. Different ways to compute temperature return levels in the climate change context. Environmetrics 21 (7–8), 698–718. <https://doi.org/10.1002/env.1060>.
- Pettitt, A.N., 1979. A non-parametric approach to the change-point problem published by: wiley for the royal statistical society. A Non-parametric Approach to the Change-point Problem. J. Royal Statist. Soc. 28 (2), 126–135. Stable URL : <http://www.jstor.org/stable/2346729> <http://www.jstor.org/stable/2346729>.
- Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346. <https://doi.org/10.1093/biomet/75.2.335>. Retrieved from.
- Proscocimi, I., Kjeldsen, T., 2021. Parametrisation of change-permitting extreme value models and its impact on the description of change. Stoch. Environ. Res. Risk Assess. 35 (2), 307–324. <https://doi.org/10.1007/s00477-020-01940-8>.
- Proscocimi, I., Kjeldsen, T.R., Miller, J.D., 2015. Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. Water Resour. Res. 51 (6), 4244–4262. <https://doi.org/10.1002/2015WR017065>.
- Ragno, E., AghaKouchak, A., Cheng, L., Sadegh, M., 2019. A generalized framework for process-informed nonstationary extreme value analysis. Adv. Water Resour. 130, 270–282. <https://doi.org/10.1016/j.advwatres.2019.06.007>.
- Ray, L.K., Goel, N.K., 2019. Flood frequency analysis of narmada river basin in India under nonstationary condition. J. Hydrol. Eng. 24 (8), 1–15. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001808](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001808).
- Rayne, S., Forest, K., 2010. Historical trends in annual water yields for the Okanagan Basin, British Columbia, Canada. Nat. Precedings 1.
- Rayne, S., Forest, K., 2011. Historical temporal trends in monthly, seasonal, and annual mean, minimum, and maximum streamflows from the Okanagan River watershed in south-central British Columbia, Canada. Nat. Precedings 1.
- Razavi, S., Gober, P., Maier, H.R., Brouwer, R., Wheater, H., 2020. Anthropocene flooding: challenges for science and society. Hydrol. Process. 34 (8), 1996–2000. <https://doi.org/10.1002/hyp.13723>.
- Read, L.K., Vogel, R.M., 2015. Reliability, return periods, and risk under nonstationarity. Water Resour. Res. 51, 6381–6398. <https://doi.org/10.1002/2015WR017089>.
- Reis, D.S., Stedinger, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. J. Hydrol. 313 (1–2), 97–116. <https://doi.org/10.1016/j.jhydrol.2005.02.028>.
- Rootzén, H., Katz, R.W., 2013. Design life level: quantifying risk in a changing climate. Water Resour. Res. 49 (9), 5964–5972. <https://doi.org/10.1002/wrcr.20425>.
- Saidi, H., Ciampittiello, M., Dresti, C., Ghiglieri, G., 2015. Assessment of trends in extreme precipitation events: a case study in piedmont (North-West Italy). Water Resour. Manag. 29 (1), 63–80. <https://doi.org/10.1007/s11269-014-0826-5>.
- Salas, J.D., Obeysekera, J., 2014. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. J. Hydrol. Eng. 19 (3), 554–568. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000820](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820).
- Salas, J.D., Obeysekera, J., Vogel, R.M., 2018. Techniques for assessing water infrastructure for nonstationary extreme events: a review. Hydrol. Sci. J. 63 (3), 325–352. <https://doi.org/10.1080/02626667.2018.1426858>.
- Sankarasubramanian, A., Srinivasan, K., 1999. Investigation and comparison of sampling properties of L-moments and conventional moments. J. Hydrol. 218 (1–2), 13–34. [https://doi.org/10.1016/S0022-1694\(99\)00018-9](https://doi.org/10.1016/S0022-1694(99)00018-9).
- Schleff, K.E., François, B., Robertson, A.W., Brown, C., 2018. A general methodology for climate-informed approaches to long-term flood projection—illustrated with the Ohio river basin. Water Resour. Res. 54 (11), 9321–9341. <https://doi.org/10.1029/2018WR023209>.
- Schleff, K.E., Kunkel, K.E., Brown, C., Demissie, Y., Lettenmaier, D.P., Wagner, A., Wigmosta, M.S., Karl, T.R., Easterling, D.R., Wang, K.J., François, B., Yan, E., 2023. Incorporating non-stationarity from climate change into rainfall frequency and intensity-duration-frequency (IDF) curves. J. Hydrol. 616, 128757 <https://doi.org/10.1016/j.jhydrol.2022.128757>.
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall’s tau. J. Am. Stat. Assoc. 63 (324), 1379–1389.
- Serago, J.M., Vogel, R.M., 2018. Parsimonious nonstationary flood frequency analysis. Adv. Water Resour. 112 (May), 1–16. <https://doi.org/10.1016/j.advwatres.2017.11.026>.
- Serinaldi, F., Kilby, C.G., 2015. Stationarity is undead: uncertainty dominates the distribution of extremes. Adv. Water Resour. 77, 17–36. <https://doi.org/10.1016/j.advwatres.2014.12.013>.
- Serinaldi, F., Kilby, C.G., Lombardo, F., 2018. Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. Adv. Water Resour. 111, 132–155. <https://doi.org/10.1016/j.advwatres.2017.10.015>. June 2017.
- Slater, L.J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., et al., 2021a. Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. Hydrol. Earth Syst. Sci. 25 (7), 3897–3935. <https://doi.org/10.5194/hess-25-3897-2021>.
- Slater, L., Villarini, G., Archfield, S., Faulkner, D., Lamb, R., Khouakhi, A., Yin, J., 2021b. Global changes in 20-year, 50-year, and 100-year river floods. Geophys. Res. Lett. 48 (6), 1–10. <https://doi.org/10.1029/2020GL091824>.
- Sneyers, R., 1975. Sur l’analyse statistique des séries d’observations. WMO Tech Note.
- Sonali, P., Nagesh Kumar, D., 2013. Review of trend detection methods and their application to detect temperature changes in India. J. Hydrol. 476, 212–227. <https://doi.org/10.1016/j.jhydrol.2012.10.034>.
- Šraj, M., Viglione, A., Parajka, J., Blöschl, G., 2016. The influence of non-stationarity in extreme hydrological events on flood frequency estimation. J. Hydrol. Hydromechanics 64 (4), 426–437. <https://doi.org/10.1515/johh-2016-0032>.
- Institute of Hydrology, 1999. In: *Flood Estimation Handbook* (Vol. 3: Statist. Institute of Hydrology Wallingford, Wallingford, U. K.
- Stedinger, J.R., 2017. Chapter 76: flood frequency analysis. In: Singh, V.P. (Ed.), *Handbook of Applied Hydrology*, second ed. McGraw-Hill Professional.
- Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Maidment, D.R. (Ed.), *Handbook of Hydrology*. McGraw-Hill, New York.
- Su, C., Chen, X., 2019. Assessing the effects of reservoirs on extreme flows using nonstationary flood frequency models with the modified reservoir index as a

- covariate. *Adv. Water Resour.* 124, 29–40. <https://doi.org/10.1016/j.advwatres.2018.12.004>. June 2018.
- Sun, P., Wen, Q., Zhang, Q., Singh, V.P., Sun, Y., Li, J., 2018. Nonstationarity-based evaluation of flood frequency and flood risk in the Huai River basin, China. *J. Hydrol.* 567 (March), 393–404. <https://doi.org/10.1016/j.jhydrol.2018.10.031>.
- Symonds, B.J., 2000. *Background and History of Water Management of Okanagan Lake and River*, vol. 25. Prepared by Ministry of Environment, Lands, and Parks (Water Management), Penticton, BC.
- Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis (Parts 1-3). *Ned. Akad. Wetensch. Proc. Ser. A* 53, 1397–1412.
- Trenberth, K.E., Dai, A., Rasmusson, R.M., Parsons, D.B., 2003. The changing character of precipitation. *Bull. Am. Meteorol. Soc.* 84 (9), 1205–1218. <https://doi.org/10.1175/BAMS-84-9-1205>.
- Vidrio-Sahagún, C.T., He, J., 2022a. Enhanced profile likelihood method for the nonstationary hydrological frequency analysis. *Adv. Water Resour.* 161 (February), 104151 <https://doi.org/10.1016/j.advwatres.2022.104151>.
- Vidrio-Sahagún, C.T., He, J., 2022b. Hydrological frequency analysis under nonstationarity using the Metastatistical approach and its simplified version. *Adv. Water Resour.* 166 <https://doi.org/10.1016/j.advwatres.2022.104244>.
- Vidrio-Sahagún, C.T., He, J., 2022c. The decomposition-based nonstationary flood frequency analysis. *J. Hydrol.* 612 (September 2022), 128186 <https://doi.org/10.1016/j.jhydrol.2022.128186>.
- Vidrio-Sahagún, C.T., He, J., Kasiviswanathan, K.S., Sen, S., 2021. Stationary hydrological frequency analysis coupled with uncertainty assessment under nonstationary scenarios. *J. Hydrol.* 598 (July), 125725 <https://doi.org/10.1016/j.jhydrol.2020.125725>.
- Vidrio-Sahagún, C.T., He, J., Pietroniro, A., 2023a. Multi-distribution regula-falsi profile likelihood method for nonstationary hydrological frequency analysis. *Stoch. Environ. Res. Risk Assess.* <https://doi.org/10.1007/s00477-023-02603-0>.
- Vidrio-Sahagún, C.T., He, J., Pietroniro, A., 2023b. Nonstationary hydrological frequency analysis using the Metastatistical extreme value distribution. *Adv. Water Resour.* 176 <https://doi.org/10.1016/j.advwatres.2023.104460>.
- Villarini, G., Serinaldi, F., Smith, J.A., Krajewski, W.F., 2009a. On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resour. Res.* 45 (8), 1–17. <https://doi.org/10.1029/2008WR007645>.
- Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D., Krajewski, W.F., 2009b. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Adv. Water Resour.* 32 (8), 1255–1266. <https://doi.org/10.1016/j.advwatres.2009.05.003>.
- Villarini, G., Taylor, S., Wobus, C., Vogel, R., Hecht, J., White, K.D., et al., 2018. *Floods and Nonstationarity: A Review*, CWTS 2018-01. US Army Corps of Engineers, Washington, DC.
- von Storch, H., Zwiers, F.W., 2002. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge. Retrieved from. <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=77515&site=ehost-live>.
- Wang, W., Vrijling, J.K., Van Gelder, P.H.A.J.M., Ma, J., 2006. Testing for nonlinearity of streamflow processes at different timescales. *J. Hydrol.* 322 (1–4), 247–268. <https://doi.org/10.1016/j.jhydrol.2005.02.045>.
- Wang, J., Lu, F., Lin, K., Xiao, W., Song, X., He, Y., 2017. Comparison and evaluation of uncertainties in extreme flood estimations of the upper Yangtze River by the Delta and profile likelihood function methods. *Stoch. Environ. Res. Risk Assess.* 31 (9), 2281–2296. <https://doi.org/10.1007/s00477-016-1370-z>.
- Wasko, C., 2021. Review: can temperature be used to inform changes to flood extremes with global warming? *Phil. Trans. Math. Phys. Eng. Sci.* 379, 20190551 <https://doi.org/10.1098/rsta.2019.0551>, 2195.
- Wasko, C., Westra, S., Nathan, R., Orr, H.G., Villarini, G., Villalobos Herrera, R., Fowler, H.J., 2021. Incorporating climate change in flood estimation guidance. *Phil. Trans. Math. Phys. Eng. Sci.* 379 <https://doi.org/10.1098/rsta.2019.0548>, 2195.
- White, H., 1980. *A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*, vol. 48.
- Wigley, T.M.L., 2009. The effect of changing climate on the frequency of absolute extreme events. *Climatic Change* 97 (1), 67–76. <https://doi.org/10.1007/s10584-009-9654-7>.
- Wu, Y., Xue, L., Liu, Y., Ren, L., 2019. Uncertainty assessment of extreme flood estimation in the Dongting Lake basin, China. *Nord. Hydrol.* 50 (4), 1162–1176. <https://doi.org/10.2166/nh.2019.088>.
- Yan, L., Xiong, L., Guo, S., Xu, C.Y., Xia, J., Du, T., 2017. Comparison of four nonstationary hydrologic design methods for changing environment. *J. Hydrol.* 551, 132–150. <https://doi.org/10.1016/j.jhydrol.2017.06.001>.
- Yin, J., Gentile, P., Zhou, S., Sullivan, S.C., Wang, R., Zhang, Y., Guo, S., 2018. Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nat. Commun.* 9 (1) <https://doi.org/10.1038/s41467-018-06765-2>.
- Yue, S., Pilon, P., 2004. A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrolog. Sci. J.* 49 (1), 21–38. <https://doi.org/10.1623/hysj.49.1.21.53996>.
- Zhang, C., Zhang, B., Li, W., Liu, M., 2014. Response of streamflow to climate change and human activity in Xitaoxi river basin in China. *Hydrolog. Process.* 28 (1), 43–50. <https://doi.org/10.1002/hyp.9539>.
- Zhang, X., Duan, K., Dong, Q., 2019. Comparison of nonstationary models in analyzing bivariate flood frequency at the Three Gorges Dam. *J. Hydrol.* 579 (August), 124208 <https://doi.org/10.1016/j.jhydrol.2019.124208>.
- Zhang, Z., Stadnyk, T.A., Burn, D.H., 2020. Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. *Can. Water Resour. J./Revue Canadienne Des Ressources Hydrauliques* 45 (1), 43–58. <https://doi.org/10.1080/07011784.2019.1691942>.