

## Design flood estimation with varying record lengths in Norway under stationarity and nonstationarity scenarios

Lei Yan<sup>a,b</sup>, Lihua Xiong<sup>a,\*</sup>, Gusong Ruan<sup>c</sup>, Mengjie Zhang<sup>d</sup> and Chong-Yu Xu<sup>IWA<sup>e</sup></sup>

<sup>a</sup> State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

<sup>b</sup> College of Water Conservancy and Hydropower, Hebei University of Engineering, Handan 056021, China

<sup>c</sup> Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway

<sup>d</sup> Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA

<sup>e</sup> Department of Geosciences, University of Oslo, N-0315 Oslo, Norway

\*Corresponding author. E-mail: xiongh@whu.edu.cn

### ABSTRACT

In traditional flood frequency analysis, a minimum of 30 observations is required to guarantee the accuracy of design results with an allowable uncertainty, however, there has not been a recommendation for the requirement on the length of data in NFFA (nonstationary flood frequency analysis). Therefore, this study has been carried out with three aims: (i) to evaluate the predictive capabilities of nonstationary (NS) and stationary (ST) models with varying flood record lengths; (ii) to examine the impacts of flood record lengths on the NS and ST design floods and associated uncertainties; and (iii) to recommend the probable requirements of flood record length in NFFA. To achieve these objectives, 20 stations with record length longer than 100 years in Norway were selected and investigated by using both GEV (generalized extreme value)-ST and GEV-NS models with linearly varying location parameter (denoted by GEV-NS0). The results indicate that the fitting quality and predictive capabilities of GEV-NS0 outperform those of GEV-ST models when record length is approximately larger than 60 years for most stations, and the stability of the GEV-ST and GEV-NS0 is improved as record lengths increase. Therefore, a minimum of 60 years of flood observations is recommended for NFFA for the selected basins in Norway.

**Key words:** average design life level, flood record length, nonstationary flood frequency analysis, predictive capabilities, split-sample testing

### HIGHLIGHTS

- This study evaluates the predictive capabilities of nonstationary (NS) and stationary (ST) models with varying flood record lengths.
- This study examines the impacts of flood record lengths on the NS and ST design floods and associated uncertainties.
- A minimum of 60 years of flood observations is recommended for NFFA for the selected basins in Norway.

## 1. INTRODUCTION

Flooding is one of the major natural disasters and has caused casualties and huge economic losses worldwide in the 21st century (Hallegatte *et al.* 2013; Arnell & Gosling 2016; Kan *et al.* 2016; Slater & Villarini 2016; Kuang *et al.* 2018; Yin *et al.* 2018; Zhang *et al.* 2018; Yan *et al.* 2019a; Hou *et al.* 2020; Li *et al.* 2020; Venvik *et al.* 2020). In the hydrological communities, flood frequency analysis (FFA) is the most widely used approach to characterize flood risk and estimate design values for hydrological structures. The fundamental assumption of FFA is that the historical extreme flood events are independent and identically distributed, namely the so-called stationary (ST) assumption. However, it should be mentioned that the ST assumption is quite rigorous because of the dynamic nature of flood events caused by climate change and human activities (Vogel *et al.* 2011; Luke *et al.* 2017; Serago & Vogel 2018; Zhang *et al.* 2019). Thus, the ST assumption has been questioned by many researchers under changing environment (Milly *et al.* 2008; Villarini *et al.* 2009; Montanari *et al.* 2013; Milly *et al.* 2015; Yan *et al.* 2017a; Xu *et al.* 2018; Xiong *et al.* 2019; Yan *et al.* 2019b).

Nonstationary FFA (NFFA) is proposed to address the nonstationary (NS) issue of extreme flood events. As evidenced from recent review articles, there are several different methods for NFFA (Khaliq *et al.* 2006; Hall *et al.* 2013; Madsen *et al.* 2014; Bayazit 2015; Hao & Singh 2016; Yan *et al.* 2017b; Salas *et al.* 2018; Jiang *et al.* 2019; Qu *et al.* 2020; Xiong *et al.* 2020; Yan *et al.* 2021), among which the time-varying moments (TVM) method is the most popular approach. In the TVM method, the

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

statistical parameters of flood distribution are no longer invariant but change with time or other covariates to capture the trend of flood frequency (Rigby & Stasinopoulos 2005). However, despite the numerous studies in NFFA, there is neither a noncontroversial method for performing NFFA nor well accepted governmental guidelines for updating current design flood strategies (Serago & Vogel 2018). One of the concerns is the tremendous uncertainties in NFFA induced by model structure and parameter estimation of NS models (Hu *et al.* 2015; Serinaldi & Kilsby 2015; Luke *et al.* 2017; Serago & Vogel 2018; Sen *et al.* 2020). Typically, the model structure of the NS model is more complex than that of ST model, since additional parameters are introduced to capture the historical trend of extreme flood events. Thus, theoretically, more flood samples are required to provide reliable and robust parameter estimations for NS models. However, in many cases, the available flood record length is shorter than 50 years and is often limited to 20–30 years (McCuen & Galloway 2010; Kjeldsen *et al.* 2014; Kobierska *et al.* 2017; Ekeu-Wei 2018; Hu *et al.* 2020). So, for practical reasons, it is necessary to investigate the impacts of record lengths on NFFA considering the various availability of flood observations worldwide.

Another concern is the predictive capabilities of NS models. Numerous literatures have reported that NS models can improve the statistical representation of historical flood series compared with ST models (Sun *et al.* 2018; Xie *et al.* 2018; Lu *et al.* 2019; Su & Chen 2019), but the predictive capabilities of NS models for out-of-sample prediction are still questioned. Only very limited study has compared the predictive capabilities of NS and ST models. Luke *et al.* (2017) conducted split-sample testing and comprehensively evaluated predictive capabilities of NS log-Pearson Type III distributions compared with ST models in the USA, and recommended the use of updated ST model for flood risk analysis in basins that have been affected by human activities. But they did not consider the effects of different flood record lengths on the predictive capabilities of NS models. In Norway, Kobierska *et al.* (2017) evaluated the impacts of record lengths on the performance of ST models with different combinations of distributions and estimation methods and found that there is no clear threshold of record length about using a three-parameter distribution or a two-parameter distribution under ST condition.

The last concern is the engineering design problem based on the NS model. Theoretically, we can predict future annual flood distribution by extrapolating the statistical parameters of the established NS model. As a result, the predicted future flood distributions vary from one year to another year, which brings about new challenge for design flood estimation. Under the framework of NFFA, if we still follow the design method under stationarity, we can obtain annual design flood for a given return period, which is impractical for engineering design since the relationship between design flood and return period is no longer one-to-one. To address the problem of design flood estimation in NFFA, researchers have carried out many studies and developed several design flood methods in recent years (Olsen *et al.* 1998; Parey *et al.* 2007, 2010; Cooley 2013; Rootzén & Katz 2013; Salas & Obeyseker 2014; Acero *et al.* 2017, 2018; Hu *et al.* 2018; Wang *et al.* 2019; Byun & Hamlet 2020; Lu *et al.* 2020; Yan *et al.* 2020). Yan *et al.* (2019a) comprehensively compared different methods and recommended the use of average design life level (ADLL) and equivalent reliability (Hu *et al.* 2018). Byun & Hamlet (2020) developed an NS Monte Carlo method to address NS design problems by explicitly accounting for the projected changes of extreme hydrometeorological events. However, the effects of flood record length on the estimation of NS design flood have not been reported, which deserves more attention considering the larger uncertainties in the calculation of design flood with longer return period using relatively short flood series (Kjeldsen *et al.* 2014; Machado *et al.* 2015; Engeland *et al.* 2017; Strupczewski *et al.* 2017; Hu *et al.* 2020; Kochanek *et al.* 2020).

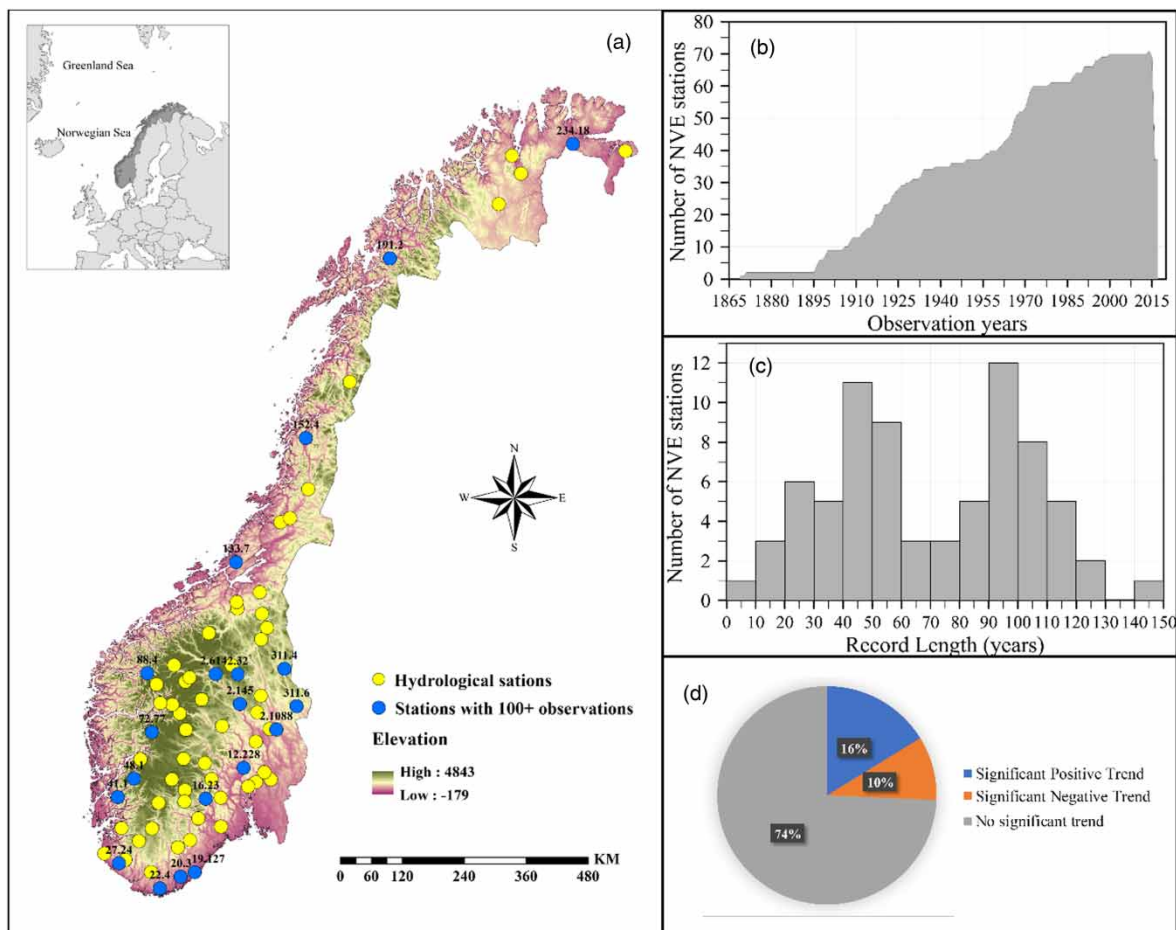
The aforementioned concerns are all related to the record length of flood series. There is a consensus that all the sampling schemes, parameter estimation and design flood estimation can be significantly influenced by the available flood record length. In Norway, dams should be able to withstand extreme floods with high return periods, such as 500 years or 1,000 years, based on the dam safety regulations of Norway (Lovdata 2010). Besides, infrastructures should withstand 20-, 200-, or 1,000-year return period flood events, resting with the impacts of flood events (TEK10 2016). Under ST condition, typically, a minimum of 30 years of flood samples is required for FFA (Stedinger *et al.* 1993; Midttømme *et al.* 2011; Kobierska *et al.* 2017). Luke *et al.* (2017) conducted NFFA using 1,250 annual maximum discharge records throughout the USA with record length ranging from 30 to 65 years, in which the record length of about 75% series is between 35 and 45 years. Yan *et al.* (2017a) conducted NFFA using annual maximum flood series (AMFS) with record lengths of 59, 62 and 66. Parey *et al.* (2010) constructed the NS model using 106 years of annual maximum temperature data. However, when we turn to nonstationarity, what are the record length requirements for NFFA with the expected accuracy or uncertainty? And how the record length might influence the predictive capabilities of NS models and the estimation of NS design flood? In this study, we attempted to answer these questions. Therefore, the objectives of this study are: (i) to evaluate the predictive capabilities of NS and ST models with varying flood record lengths, (ii) to examine the effects of varying flood record lengths on the

NS and ST design floods and associated uncertainties and (iii) to provide guidance for the probable requirements of record length for NFFA with the expected accuracy or uncertainty. To achieve these objectives, 20 AMFS with a record length of larger than 100 years in Norway were selected for illustration purposes.

## 2. STUDY AREA AND DATA

In this study, a number of annual maximum flood records of hydrological stations in Norway was used for the purpose of evaluating the effects of varying record lengths on model performance, predictive capabilities and design flood estimation of GEV (generalized extreme value)-ST and GEV-NS0. Initially, we collected flood data of 74 hydrological stations throughout the Norwegian mainland from the Norwegian Water Resources and Energy Directorate (NVE). Figure 1 presents the geographical location and data information of the collected stations. Then, only the data sets with more than 100 observations were reserved to ensure the availability of at least 70 observations for the maximum record length in the fitting period and 30 observations in the evaluation period. Finally, the long flood records of 20 watersheds throughout Norway were chosen for illustration purposes (including 2 stations with 99 observations).

Hydrological variables are affected by both climate and land-use/land-cover changes to facilitate the study of climate change and its impact on hydrological variability, and a Hydrological Reference Dataset (HRD) consists of good-quality long-term data series not influenced by human activities causing non-climate-related variability or change was constructed by Norwegian Water Resources and Energy Directorate (NVE) (Fleig *et al.* 2013). HRD consists of 250 reference stations throughout Norway. One of the selection criteria is the degree of river basin development. Particularly, the proportion of



**Figure 1** | (a) Location information of the hydrological stations and station with record length of larger than 100 years. (b) The number of available stations with observations in given years. (c) Histogram of record lengths of available stations in this study. (d) Trend detection results of available stations.

urban area of the 250 reference stations is lower than 10%. The other criterion is the absence of significant regulations, diversions, or water use. The selected 250 potential reference stations are classified as ‘unregulated’ in the reports on NVE’s network of streamflow stations (Pettersson 2003; Kleivane 2006). In this study, two of the final selected nine stations are included in the HRD. For the remaining stations, similar criteria of no significant regulations and land-use/land-cover changes were used. For these reasons, the insensitive land-use and land-cover change as well as other minor physical changes, if any, are not explicitly considered in the study.

The main characteristics of the selected 20 watersheds, including the drainage areas, the geographical locations and starting and end years of the observations, are presented in Table 1. Previous studies have documented the existence of significant trend in the annual maxima and peaks over threshold in some regions of Norway (Vormoor *et al.* 2015, 2016; Debele *et al.* 2017). In this study, the trend-free pre-whitening Mann-Kendall test (TFPW-MK) (Yue *et al.* 2002) was employed to examine the trend of the selected AMFS by removing the effects of trend and serial correlation on the examination results (Table 1). The trend is statistically significant at 5% significant level if the *P*-value in Table 1 is smaller than 0.05. It should be mentioned that TFPW-MK suffered from type-1 error and Serinaldi & Kilsby (2016) proposed a corrected version called the TFPWcu method. See Serinaldi & Kilsby (2016) for the mathematical and numerical demonstration of TFPW flaws and the TFPWcu method, and see Serinaldi *et al.* (2020) for detailed discussion about the errors resulting from this superficial application of statistical analysis.

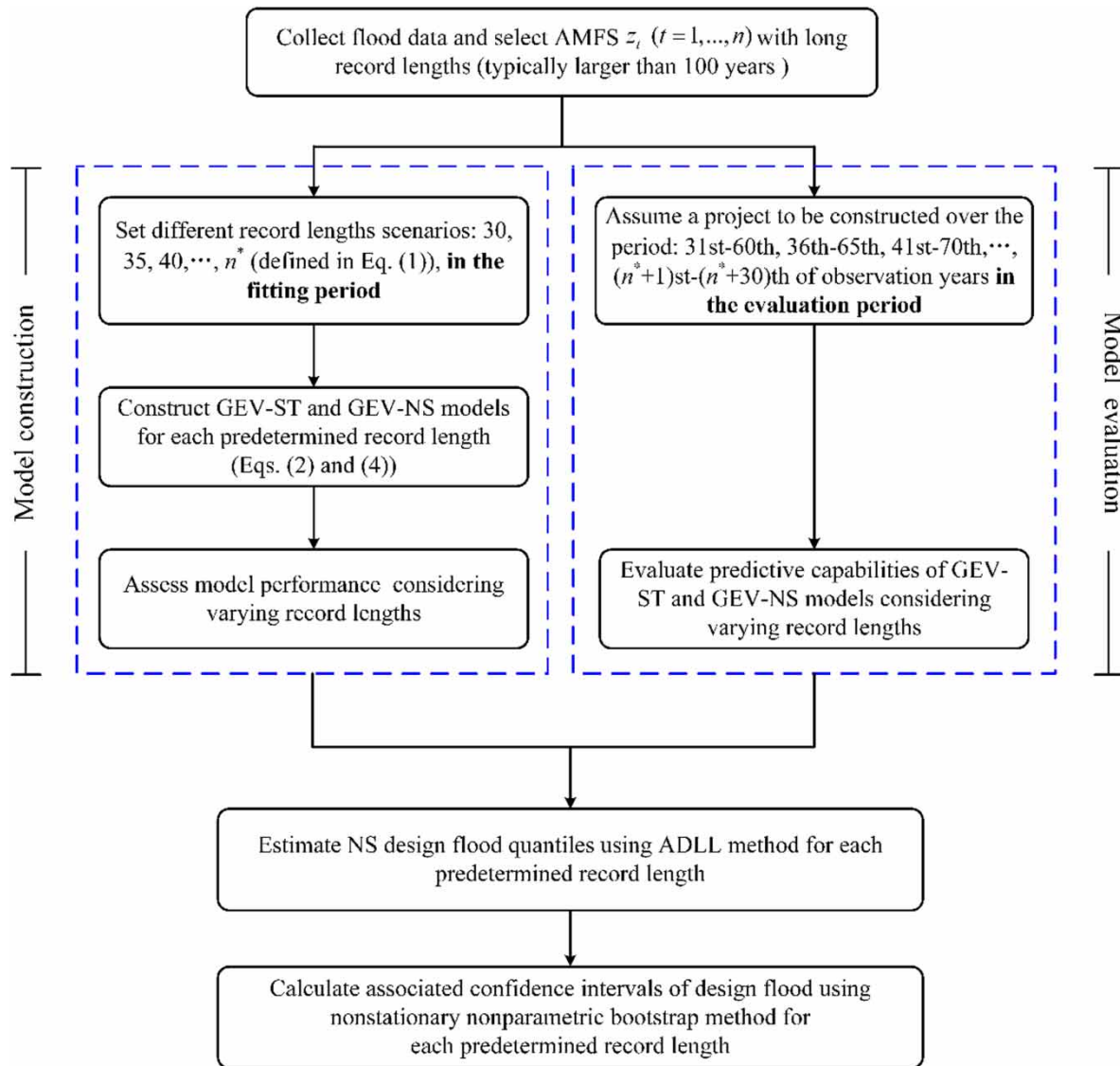
### 3. METHODOLOGY

In this section, we describe how the effects of varying record lengths are considered in NFFA from model construction to design flood estimation. The flowchart of this study is presented in Figure 2.

**Table 1** | Data Information of the selected 20 watersheds with long record lengths in Norway

Station_ID	Name	Area (km <sup>2</sup> )	Data period	Record length	Longitude	Latitude	<i>P</i> -value
2.32	Atnasjø	463.3	1916–2015	100	10.22	61.85	0.095
2.145	Losna	11,200	1896–2017	122	10.28	61.33	0.022*
2.614	Rosten	1,833	1917–2015	99	9.40	61.86	0.685
2.1088	Glomma ved Skogmuseet	15,500	1871–2017	147	11.56	60.87	0.000*
12.228	Kistefoss	3,703	1917–2015	99	10.36	60.22	0.050*
16.23	Kirkevollbru	3,845.4	1906–2015	110	9.04	59.69	0.001*
19.127	Rygenetotal	3,946.4	1900–2015	116	8.67	58.41	0.011*
20.3	Flaksvatn	1,780	1900–2017	118	8.20	58.33	0.719
22.4	Kjøemo	1,757.7	1897–2015	119	7.53	58.12	0.115
27.24	Helleland	184.7	1897–2015	119	6.15	58.53	0.009*
41.1	Stordalsvatn	130.7	1913–2015	103	6.01	59.68	0.340
48.1	Sandvenvatn	470	1908–2017	110	6.55	60.01	0.502
72.77	Flåm bru	263	1908–2017	110	7.12	60.84	0.020*
88.4	Lovatn	234.9	1900–2015	116	6.89	61.86	0.003*
133.7	Krinsvatn	206.6	1916–2015	100	10.23	63.80	0.187
152.4	Fustvatn	525.7	1909–2015	107	13.31	65.91	0.019*
191.2	Øvrevatn	526	1914–2015	102	17.94	68.86	0.679
234.18	Polmak	14,161.4	1912–2015	104	28.02	70.07	0.161
311.4	Femundsenden (Femunden)	1,790	1896–2017	122	11.94	61.92	0.081
311.6	Nybergsund	4,424.9	1909–2015	107	12.32	61.26	0.413

‘\*’ denotes the trend is significant at 5% significance level.



**Figure 2** | The flowchart of this study.

### 3.1. Setting of varying record length scenarios and the split-sample testing method

To assess the effects of record lengths on model performance, predictive capabilities and design flood estimation, different record length scenarios are generated. For an observed AMFS  $z_i(t = 1, \dots, n)$ , we assume there is a hydrological project to be constructed with a design life of 30 years within the observation years. Thus, the predetermined observations  $z_i^f(t = 1, \dots, n^*)$  with record length  $n^*$  in the fitting period and assumed future floods  $z_i^e(t = n^* + 1, \dots, n^* + 30)$  with a record length of 30 in the evaluation period are determined from the subset of the entire AMFS. The predetermined record lengths  $n^*$  are sampled every 5 years from 30 to  $n_{\max}^*$ , which is given by the following equation:

$$n_{\max}^* = 5 \times \text{floor}[(n - 30)/5] \quad (1)$$

where function floor returns the round down of  $(n - 30)/5$ , and 30 is the design life of a project. Herein the split-sample testing method is used. For example, if there are 106 observations,  $n_{\max}^*$  is equal to 75 and thus the record lengths are set to be 30, 35, 40, ..., 75. Subsequently, both ST GEV (GEV-ST) and NS GEV (GEV-NS) models are built for each predetermined record length in the fitting period, and the following 30 years' observations are assumed to be the design lifespan of a hydrological

project, which is retained for evaluating the predictive capabilities and estimating the design flood in the evaluation period (Figure 3).

### 3.2. NFFA with varying record lengths (fitting period)

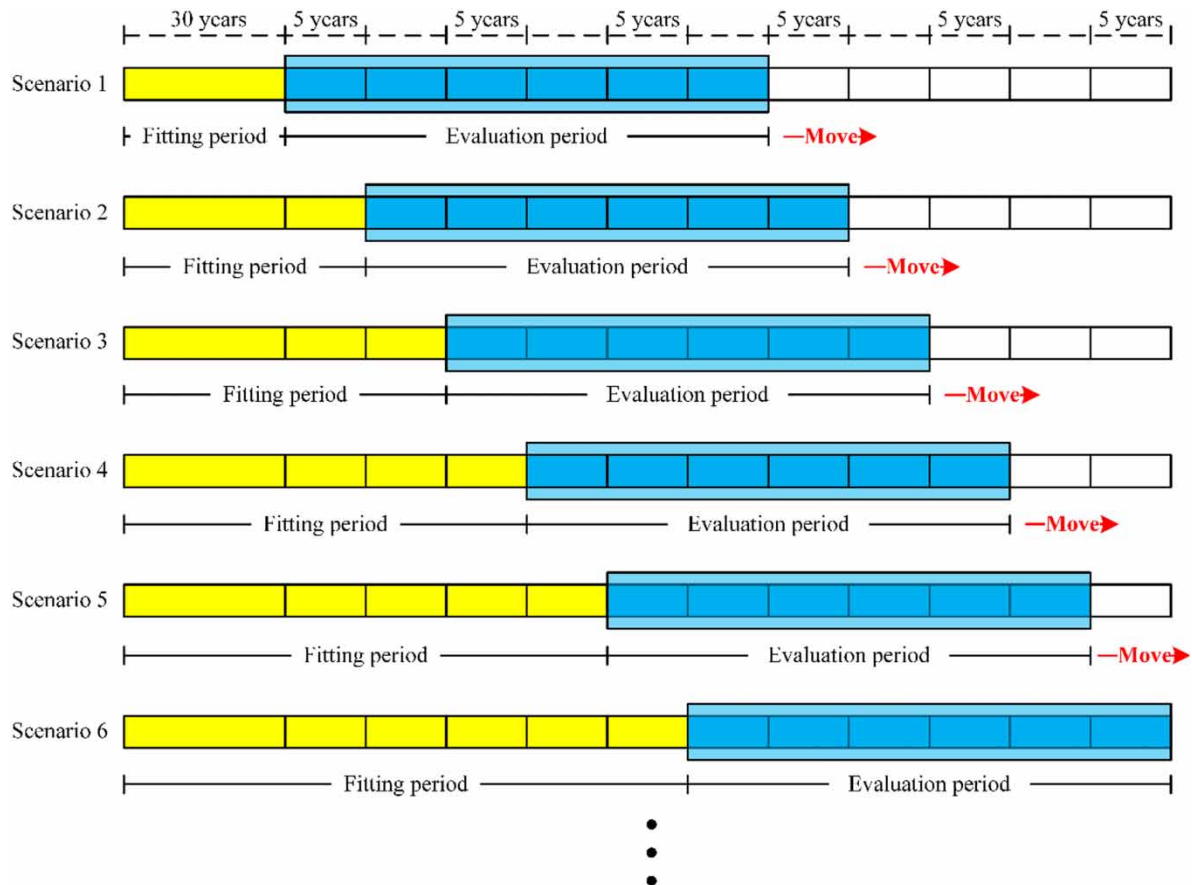
#### 3.2.1. NS GEV model

In this study, we focus on the NFFA using single distribution rather than the nonstationarity caused by abrupt change or mixed flood generation mechanisms. GEV distribution is often used to model the annual maxima, such as AMFS, under the ST context. In the fitting period, for the AMFS  $z_t^f(t = 1, \dots, n^*)$ , the cumulative distribution function (CDF) of GEV-ST model is given by the following equation:

$$\begin{cases} G(z_t^f | \mu, \sigma, \varepsilon) = \exp\{-[1 + \varepsilon(z_t^f - \mu)/\sigma]^{-1/\varepsilon}\}, t = 1, \dots, n^* \\ 1 + \varepsilon(z_t^f - \mu)/\sigma > 0 \end{cases} \quad (2)$$

where  $\mu, \sigma > 0$  and  $\varepsilon$  are the location, scale and shape parameters of GEV-ST model, respectively. However, under NS condition, the statistical parameters of GEV model can be modeled as a function of covariates, such as time, to capture the trend of AMFS. The CDF of GEV-NS model is as follows:

$$\begin{cases} G_t(z_t^f | \mu_t, \sigma_t, \varepsilon_t) = \exp\{-[1 + \varepsilon_t(z_t^f - \mu_t)/\sigma_t]^{-1/\varepsilon_t}\}, t = 1, \dots, n^* \\ 1 + \varepsilon_t(z_t^f - \mu_t)/\sigma_t > 0 \end{cases} \quad (3)$$



**Figure 3** | Schematic diagram of varying record length scenarios using moving windows. The yellow rectangular boxes indicate the predetermined record lengths in the fitting period, increased by 5 years in each scenario. The light blue boxes represent the assumed 30-year lifespan of a project, moving with the increase of predetermined record lengths. It should be noted that there are more scenarios in our study than the six scenarios displayed above. Please refer to the online version of this paper to see this figure in color: <http://dx.doi.org/10.2166/nh.2021.026>.

where  $G_t(\cdot)$  denotes the time-varying CDF of GEV-NS model.  $\mu_t, \sigma_t > 0$  and  $\varepsilon_t$  are the time-varying location, scale and shape parameters of GEV model, respectively, and  $t$  is the time scale. The time-varying statistical parameters, varying from one year to another year, are estimated based on the whole trend of the entire AMFS. Theoretically, although all three statistical parameters can be time-varying, the shape parameter in Equation (3) is sensitive and difficult to estimate (Cheng & Aghakouchak 2014; Du *et al.* 2015). Besides, the aim of this study is not to select the optimal model for a single basin, but to investigate the impacts of record lengths in NS frequency analysis. Thus, the structure of GEV-NS models should be simplified and fixed for each record length. The GEV-NS models only considering the time-varying location parameter become one of the most widely used GEV-NS models in practical applications. This kind of simplified GEV-NS models can generate realistic design quantiles consistent with the probabilistic behavior of extreme events (Cheng & Aghakouchak 2014; Sarhadi & Soulis 2017), which is described as follows:

$$G_t^{-1}(1-p|\mu_t, \sigma, \varepsilon) = \mu_0 + \mu_1 \times t + \frac{\sigma}{\varepsilon} \times [(-\ln(1-p))^{-\varepsilon} - 1], t = 1, \dots, n^* \quad (4)$$

where  $G_t^{-1}(\cdot)$  is the time-varying inverse CDF of GEV-NS model.  $p$  is the annual exceedance probability.  $\mu_0$  and  $\mu_1$  are the model parameters to model the change of location parameter  $\mu_t$ , that is  $\mu_t = \mu_0 + \mu_1 \times t$ . The GEV-NS model in Equation (4) is simplified and convenient, considering that the shape parameter  $\varepsilon$  of GEV is quite sensitive and difficult to be estimated. Therefore, in this study, the GEV-NS model is constructed based on Equation (4) as done by previous researchers (Zwiers *et al.* 2010; Westra *et al.* 2013; Luke *et al.* 2017; Um *et al.* 2017).

### 3.2.2. Assessment of model performance for the fitting period

In this study, both GEV-ST and GEV-NS models with linearly varying location parameter, denoted by GEV-NS0, are built for each predetermined record length. Thus, to assess the fitting qualities and compare the difference between GEV-ST and GEV-NS0, the Akaike information criterion (AIC) (Akaike 1974) is employed, which is described as follows:

$$\text{AIC} = -2 \ln(\zeta) + 2\rho \quad (5)$$

where  $\rho$  is the number of independently adjusted parameters of the model.  $\zeta$  is the maximum likelihood value of the established model. It should be noted that both the statistical parameters of GEV-ST and GEV-NS0 are estimated by the maximum likelihood method. Thus,  $\zeta$  can be estimated by calculating the maximum value of likelihood function via deriving the partial derivatives of the log-likelihood function, which is, for instance,  $\ln L(\theta|z_t^f) = \ln \prod_{t=1}^{n^*} g(z_t^f|\mu, \sigma, \varepsilon)$  for the GEV-ST model.

### 3.3. Evaluation of predictive capabilities with varying record lengths (evaluation period)

For each predetermined record length, the likelihood function is used to evaluate the predictive capabilities of both GEV-ST and GEV-NS0. Theoretically, the likelihood function can provide information about how plausible the estimated statistical model is with the observations  $z_t^e(t = n^* + 1, \dots, n^* + 30)$  during the evaluation period. For the GEV-ST model, the likelihood function  $L(\theta|z_t^e)$  for the evaluation period is described as follows:

$$L(\theta|z_t^e) = \prod_{t=n^*+1}^{n^*+30} g(z_t^e|\mu, \sigma, \varepsilon) \quad (6)$$

where  $g(\cdot)$  is the probability density function (PDF) of GEV, and  $n^*$  is the record length of predetermined observations. However, the likelihood function  $L(\theta_t|z_t^e)$  of GEV-NS model for the evaluation period is defined as follows:

$$L(\theta_t|z_t^e) = \prod_{t=n^*+1}^{n^*+30} g(z_t^e|\mu_t, \sigma, \varepsilon) \quad (7)$$

where  $\theta_t$  represents the time-varying statistical parameters. It can be seen from Equations (6) and (7) that under ST condition, the likelihood function is evaluated based on a set of invariant statistical parameters, whereas under NS condition, the likelihood function is evaluated based on variant statistical parameters.

Besides, the probabilistic coverages were also used to assess the predictive capabilities of GEV-ST and GEV-NS0. Theoretically, the percentage of flood observations below the prescribed percentiles, namely 5th, 25th, 50th, 75th and 95th

percentiles, should be 5, 25, 50, 75 and 95%. The absolute difference between theoretical and actual probabilistic coverages, denoted by  $D_{pc}$ , can be calculated by the following equation:

$$\begin{cases} D_{pc} = |PC_T - PC_A| \\ PC_A = \text{countif}(z_t < \text{percentiles})/n \end{cases} \quad (8)$$

where  $PC_T$  and  $PC_A$  are the theoretical and actual probability coverages, respectively, and  $PC_T = 5, 25, 50, 75$  and  $95\%$ . percentiles denote the 5th, 25th, 50th, 75th and 95th percentiles.  $n$  represents the number of observations retained in the evaluation period and equals 30 in this study.

### 3.4. NS design floods and uncertainties with varying record lengths

#### 3.4.1. Average design life level

In this study, the ADLL method is used to solve the NS hydrological design problem using GEV-NS0. The ADLL method was proposed by Yan *et al.* (2017a) based on the concept of annual average reliability, denoted by  $RE_{T_1-T_2}^{ave}$ , under NS context.  $RE_{T_1-T_2}^{ave}$  is defined as follows (Read & Vogel 2015):

$$RE_{T_1-T_2}^{ave} = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} (1 - p_t) = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} G_t(z_q | \theta_t) \quad (9)$$

where  $p_t$  is the time-varying nonexceedance probability,  $z_q$  is the design quantile over the design life of a hydrological project to be constructed from  $T_1$  to  $T_2$ . The ADLL method assumes that the annual average reliability corresponding to  $z_q$  should be equal to the yearly reliability  $1 - 1/m$  for the return period  $m$ . Thus, the  $m$ -year design quantile  $z_{T_1-T_2}^{ADLL}(m)$  estimated by the ADLL method can be determined by the following equation:

$$\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} G_t(z_{T_1-T_2}^{ADLL}(m) | \theta_t) = 1 - 1/m \quad (10)$$

ADLL is recommended for practical use because it has associated design floods with the design life period of projects and is able to yield reasonable design quantiles and confidence intervals (CIs). See Yan *et al.* (2017a), which compared different NS design methods and recommended the ADLL method, for details of the advantages of ADLL. It should be noted that the maximum record length is equal to the full record length when analyzing the effects of record lengths on design flood quantiles and associated CIs. In addition, the coefficient of variation (CV) of design flood is also estimated for each record length using the nonstationary nonparametric bootstrap (NNB) method to assess the stability of GEV-ST and GEV-NS0.

#### 3.4.2. Uncertainty analysis using the NNB method

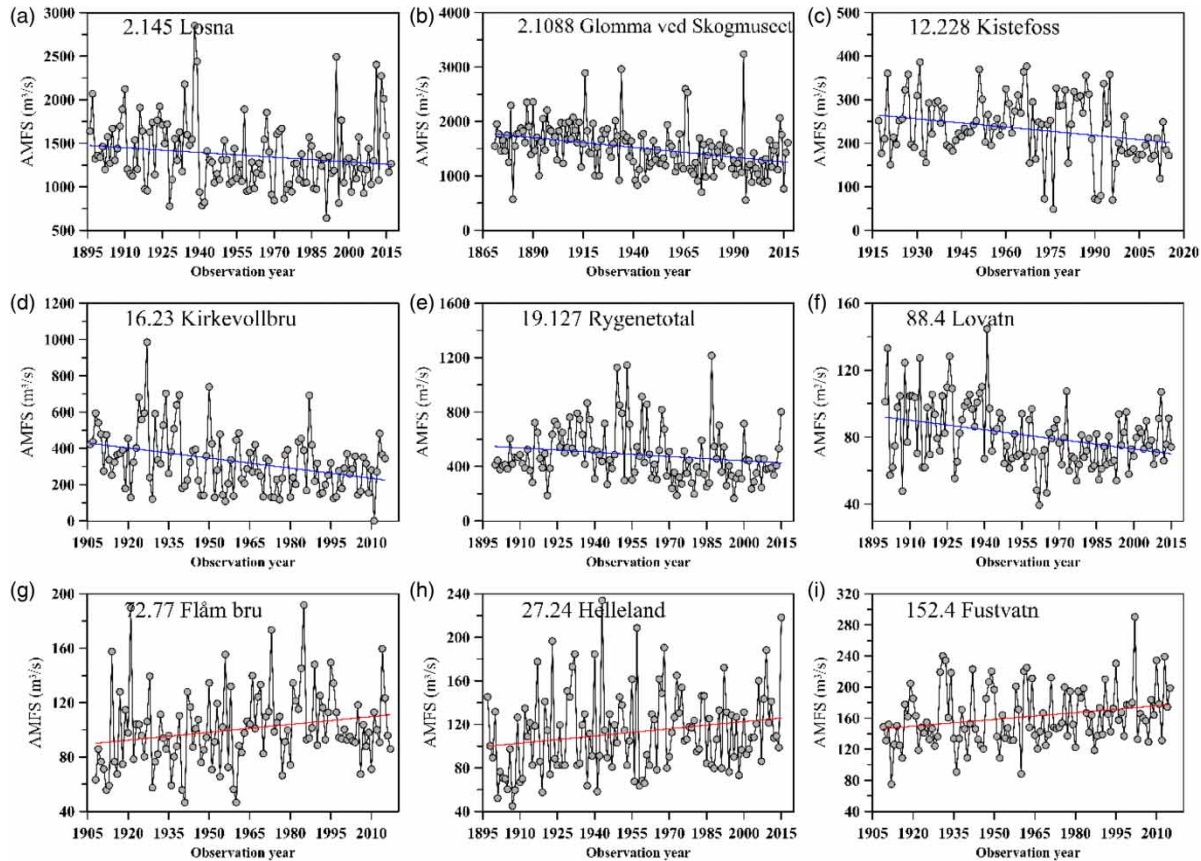
In this study, to comprehensively evaluate the effects of record lengths on the ST and NS design results, the uncertainties or CIs of design floods estimated by the ADLL method are also calculated based on the NNB method. The advantage of NNB model is that it depends strictly on the available data without any hypotheses and can be implemented regardless of model complexity. In the NNB method, first, an NS model is used to fit the original observations. Then, the original observations are transformed to identically distributed residuals, and the transformed residuals are resampled with replacement and back-transformed to obtain the new samples. Next, the same NS model constructed in the first step is employed to fit the bootstrapped new samples and estimate corresponding design quantiles. Repeat the above steps for a large number of times and finally determine the CIs of design quantiles. See Obeysekera & Salas (2014) and Yan *et al.* (2017a) for detailed information and operating steps about the NNB method.

## 4. RESULTS AND DISCUSSION

### 4.1. NFFA with varying flood record lengths

According to the results of the TFPW-MK test, 9 of the 20 selected AMFS exhibited a significant increasing or decreasing trend at 5% significant level, as displayed in Figure 4. Then, NS frequency analysis of the AMFS of the nine stations was





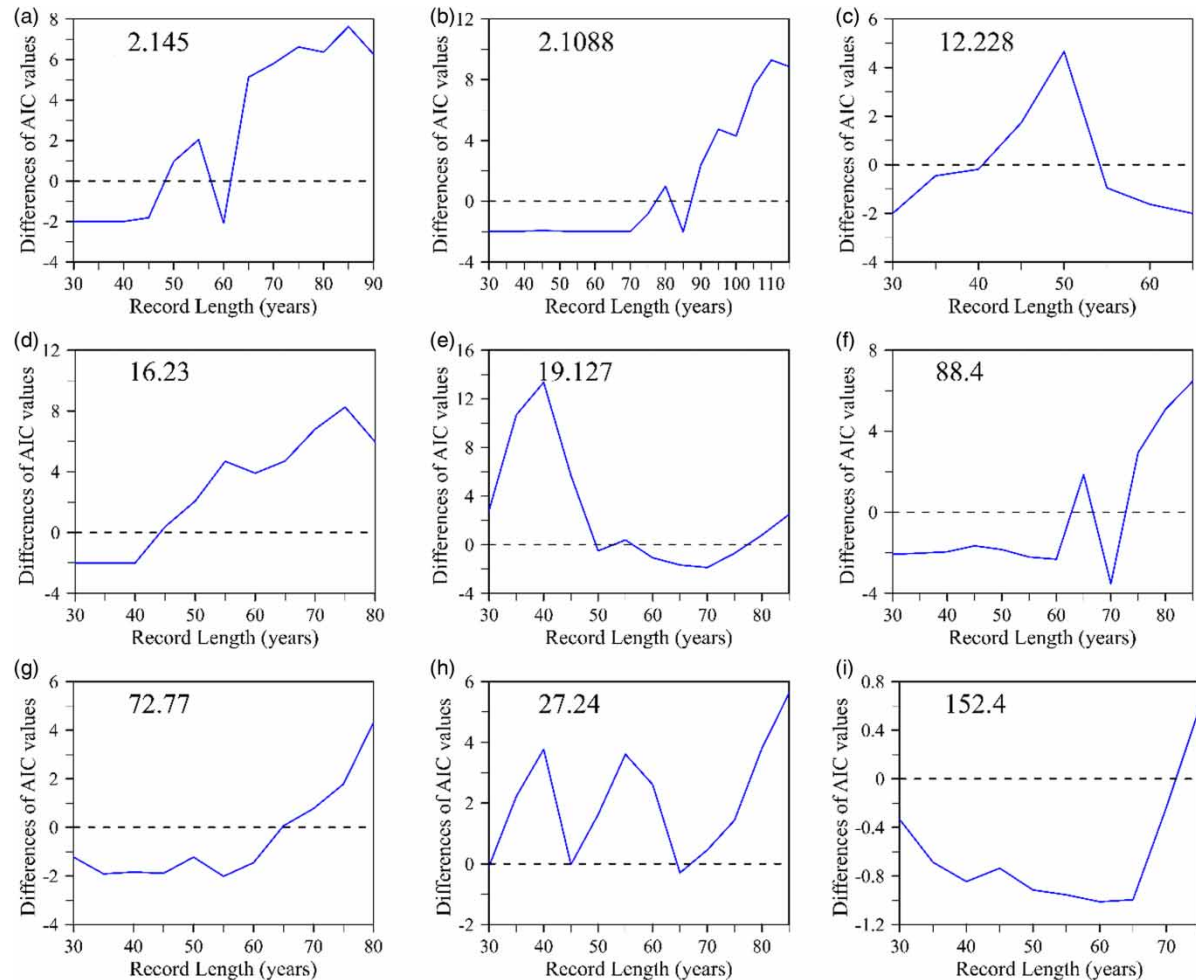
**Figure 4** | AMFS of the selected nine stations with increasing or decreasing trends.

carried out using the generalized additive models for location, scale and shape (GAMLSS) package available in the R language platform.

Both GEV-ST and GEV-NS0 were constructed for different record length scenarios defined in Section 3.1. We evaluated and compared the model performance of GEV-ST and GEV-NS0 using the AIC values. As shown in Figure 5, for eight out of the nine stations, the AIC values of GEV-ST were eventually larger than those of GEV-NS0 with the increase of record lengths, indicating the better performance of GEV-NS0 particularly for samples with large record lengths. Furthermore, the differences between AIC values of GEV-ST and GEV-NS became larger with the increase of record lengths except for stations 12.228 and 19.127. Quantitatively, for 8 out of the 9 stations, the biggest AIC differences in Figure 5 are larger than 5, and for 2 out of the 9 stations, the biggest AIC differences are around or larger than 10, which confirms the advantage of the GEV-NS0. Besides, we also found that the superiority of GEV-NS0 starts to emerge roughly for record lengths of larger than 60 years, indicating the necessity of collecting more flood samples in NFFA rather than the minimum requirement of 30 samples in conventional FFA.

#### 4.2. Predictive capabilities of GEV-ST and GEV-NS models with varying record lengths

Likelihood values were employed to evaluate the predictive capabilities of GEV-ST and GEV-NS0. As shown in Figure 6, for majority stations (six out of nine stations), the likelihood values of GEV-NS0 are eventually higher than those of the GEV-ST models for larger record length, particularly when the record length is larger than 60 years. It should be mentioned that the likelihood values in Figure 6 may be affected by high uncertainty for small samples. Besides, we also examined the difference between theoretical probability coverage and actual probability coverages which are the percentage of flood observations under 5th, 25th, 50th, 75th and 95th percentiles, denoted by  $D_5$ ,  $D_{25}$ ,  $D_{50}$ ,  $D_{75}$  and  $D_{95}$ , respectively. As shown in Figures 7 and 8, the average value of  $D_5$  is smaller than 10% for both GEV-ST and GEV-NS0 when record length is larger than 60, while the average value of  $D_{95}$  is smaller than 5% when record length is larger than 40, indicating the good predictive



**Figure 5** | Absolute differences of the AIC values calculated using the AIC values of GEV-ST models minus those of GEV-NS models. The values larger than zero indicate the better performance of GEV-NS models, and vice versa.

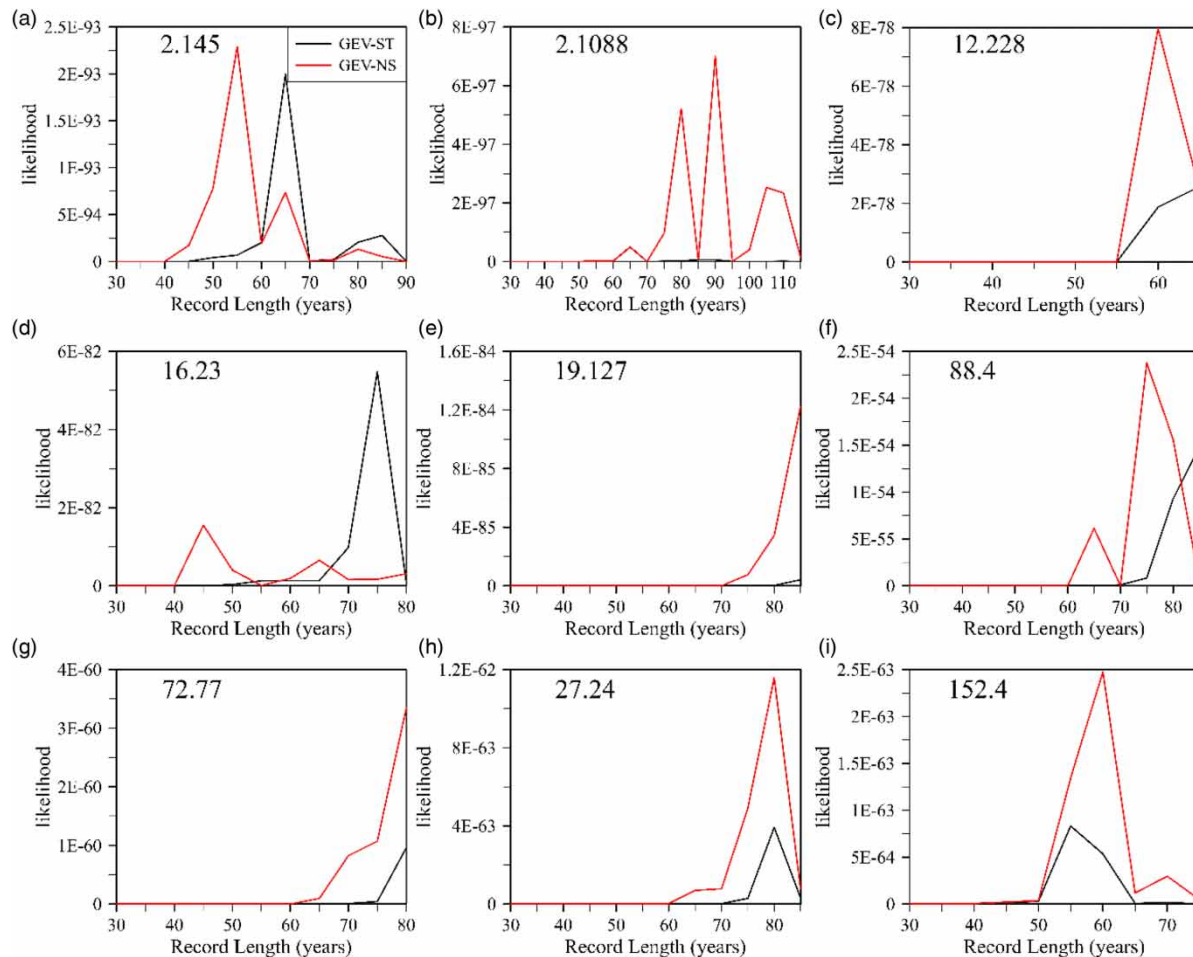
capabilities of extreme events for both GEV-ST and GEV-NS0. Furthermore, the results indicate that the difference in nearly all five plots tends to be smaller with the increase of record lengths for GEV-NS0, particularly when the record length is roughly longer than 60 years. While for the GEV-ST models, there is no significant decrease of the five plots with the increase of record lengths, and  $D_{25}$  and  $D_{50}$  even exhibit upward trends. It should be mentioned that the difference of probability coverage in Figure 7 may be affected by less stations for larger record lengths. Particularly, for  $D_{25}$  and  $D_{50}$ , the decreasing trends become more significant for the two largest sample sizes of 90 and 95. The '95-size' value relies on station 2.1088, while the '90-size' average is taken over stations 2.145 and 2.1088. In fact, decreasing trend can still be identified regardless of these two record lengths, particularly compared with the increasing trends of  $D_{25}$  and  $D_{50}$  for the GEV-ST model.

The above results confirm the better predictive capabilities of GEV-NS0 compared with GEV-ST models for most stations. The enhanced model performance and predictive capabilities of GEV-NS0 can be attributed to their better capture of the variability of extreme flood events when new flood samples are added into the flood population.

#### 4.3. Influence of varying record lengths on NS design floods and uncertainties

After the evaluation of model performance and predictive capabilities of both GEV-ST and GEV-NS0, we further examine the influence of record length on design flood quantiles and associated CIs. Assuming the design life of hydrological structures is 30 years, both GEV-ST and GEV-NS0 were employed to estimate design floods using the ADLL method for each record length.

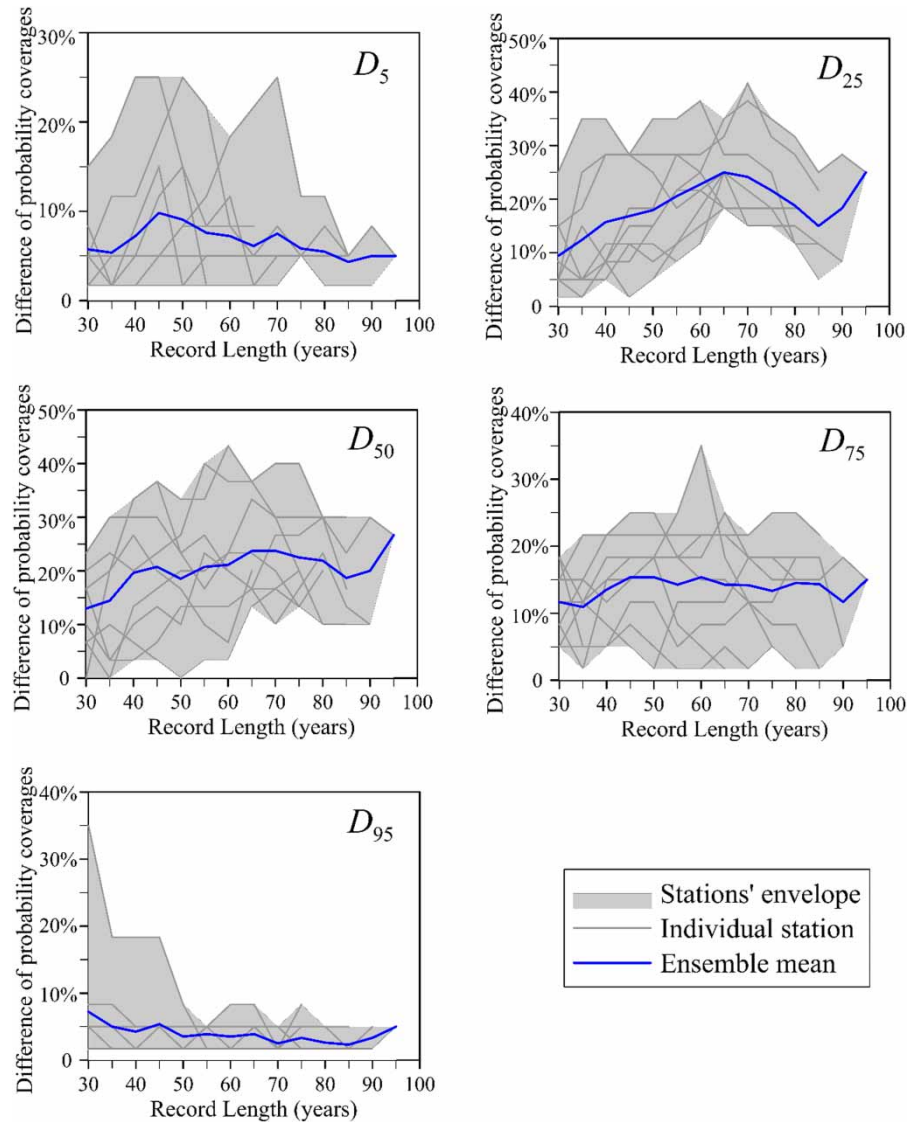
For each station, we first evaluated the stability of the established GEV-ST and GEV-NS0 by calculating the CVs of the 20-, 50- and 100-year design flood quantiles for each record length. Figure 9 shows the evolution of CV as a function of record



**Figure 6** | Likelihood values of GEV-ST (black lines) and GEV-NS (red lines) models. Please refer to the online version of this paper to see this figure in color: <http://dx.doi.org/10.2166/nh.2021.026>.

length for all the selected nine stations. It is shown that for both ST and NS conditions, the CVs of 20-, 50- and 100-year design floods tend to be smaller with the increase of record length for nearly all stations, showing the increasing stability of the established statistical models as record lengths increase. Besides, when the record length is approximately larger than 50–60 years, the fluctuation of CV becomes gentle for most stations. It should be mentioned that for stations 19.127, 27.24 and 152.4, the stabilization of both models occurs for a record length of smaller than 60. As for the comparison of the stability of GEV-ST and GEV-NS0, the CVs of GEV-ST are smaller than those of GEV-NS0 for all the three return periods, which is due to the more complex model structure of GEV-NS0.

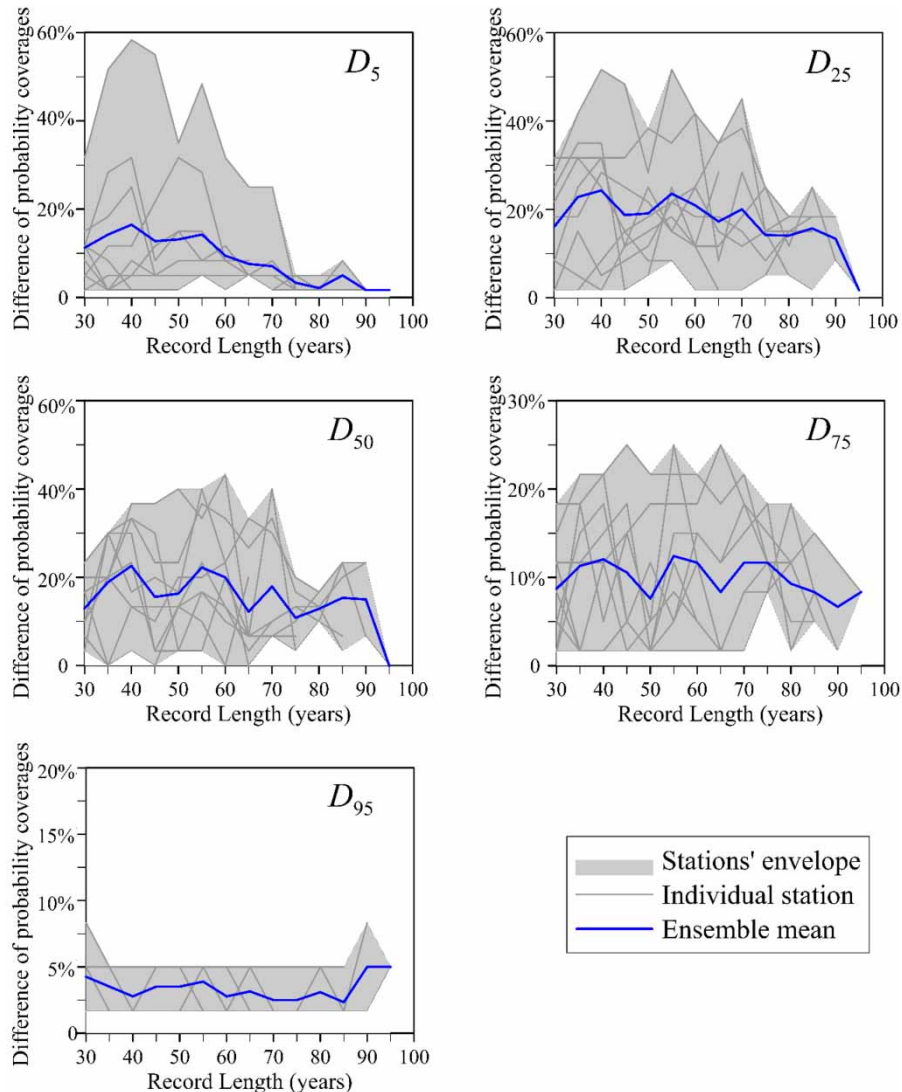
Figure 10 summarizes the evolution of 100-year design floods and associated 95% CIs with the increase of record length under both ST and NS contexts. The design values yielded by the GEV-ST model and the simplified GEV-NS model based on the ADLL method are similar, particularly with the increase of record length. For six out of the nine stations, the difference between 100-year design flood of ST and NS models is relatively small with the increase of record length or even becomes smaller when record length is larger than 50–60 years, see stations 12.228 and 19.127. Thus, it is difficult to discriminate between these two design strategies for a larger record length. For the rest of the stations, a representative example is station 2.1088, for which the difference between ST and NS design floods becomes larger when the record length is larger than 70 years and the largest difference is obtained when record length is 145 years. In terms of CIs, it is found that the associated bootstrapped 95% CIs of the 100-year design flood become smaller as the record lengths increase for both GEV-ST and GEV-NS0, showing the advantage of collecting more flood samples under both ST and NS conditions. In addition, there is no significant difference between the CIs of ST and NS design floods for most stations.



**Figure 7** | The difference between theoretical probability coverage and actual probability coverages and the percentage of flood observations under 5th, 25th, 50th, 75th and 95th percentiles, denoted by  $D_5$ ,  $D_{25}$ ,  $D_{50}$ ,  $D_{75}$  and  $D_{95}$ , respectively, using GEV-ST models. The difference was calculated using Equation (8), and ensemble mean was the average difference of nine stations.

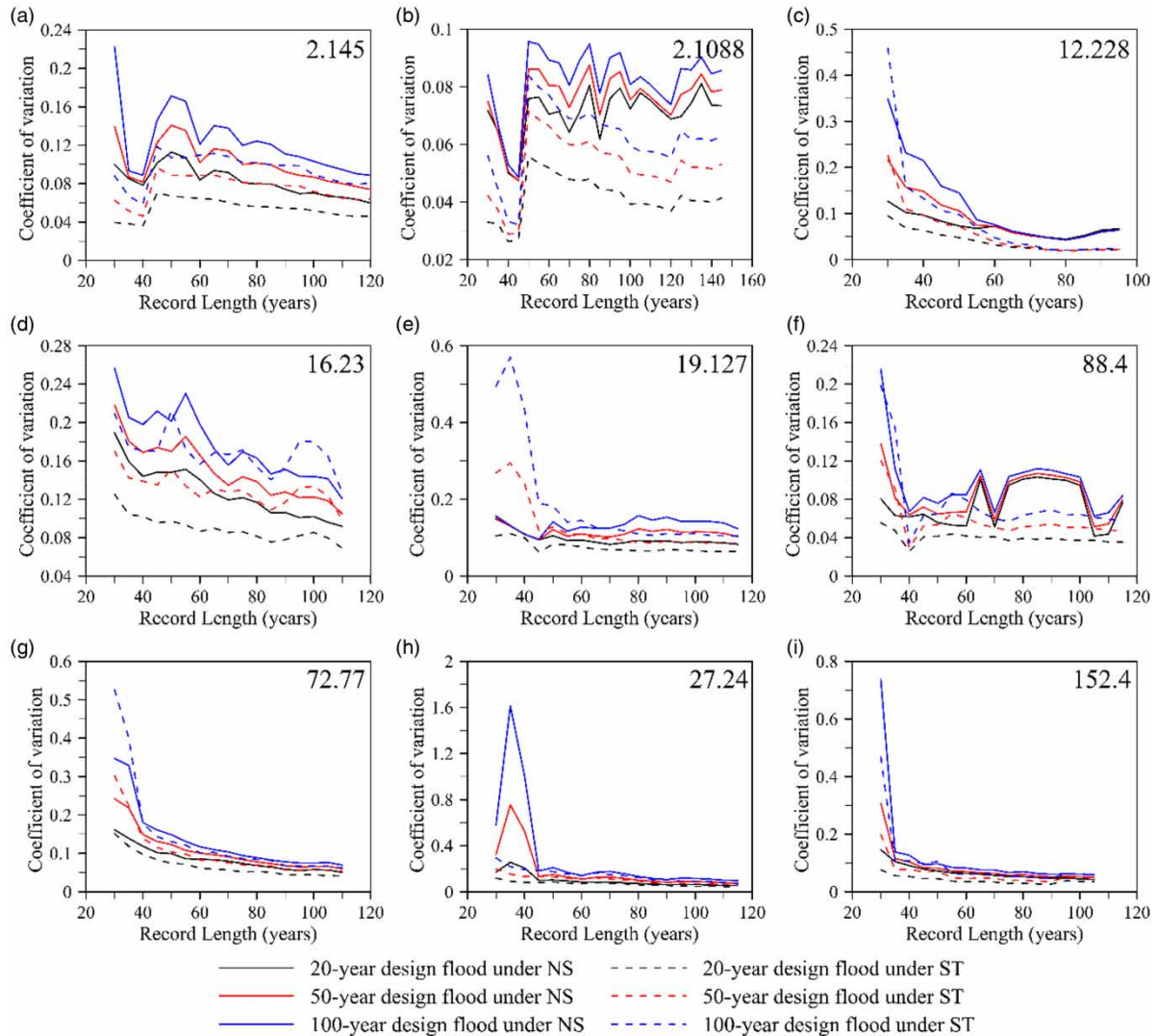
#### 4.4. Discussions

There is still a hot academic debate on whether we should switch to NS frequency analysis, considering the complex model structure, prediction of future distributions and the uncertainties associated with NS hydrological design. However, there is no doubt that we are living in a changing world. Thus, this study focuses on what we could do if the world is NS. Again, it should be emphasized that nonstationarity cannot be inferred only by trend tests but should be confirmed by attribution analysis, statistical analysis and empirical analysis (Milly *et al.* 2015; Yan *et al.* 2017a). As discussed by Serinaldi *et al.* (2018), trend analysis can at most be used as statistical tools for preliminary screening whose outcome should be carefully checked and complemented with exogenous information. If a clear physical mechanism related to a predictable evolution of the properties of the process at hand is not identified, we cannot make conclusions about the reason of rejection or lack of rejection, since multiple factors not included in the null and alternative hypotheses can actually play a role. Furthermore, because of the different availability of flood observations all over the world, it is crucial to assess the impacts of varying record lengths on NFFA for practical reasons, and our research can provide technical guidance for engineers and policymakers.



**Figure 8** | The difference between theoretical probability coverage and actual probability coverages and the percentage of flood observations under 5th, 25th, 50th, 75th and 95th percentiles, denoted by  $D_5$ ,  $D_{25}$ ,  $D_{50}$ ,  $D_{75}$  and  $D_{95}$ , respectively, using GEV-NS models. The difference was calculated using Equation (8), and ensemble mean was the average difference of nine stations.

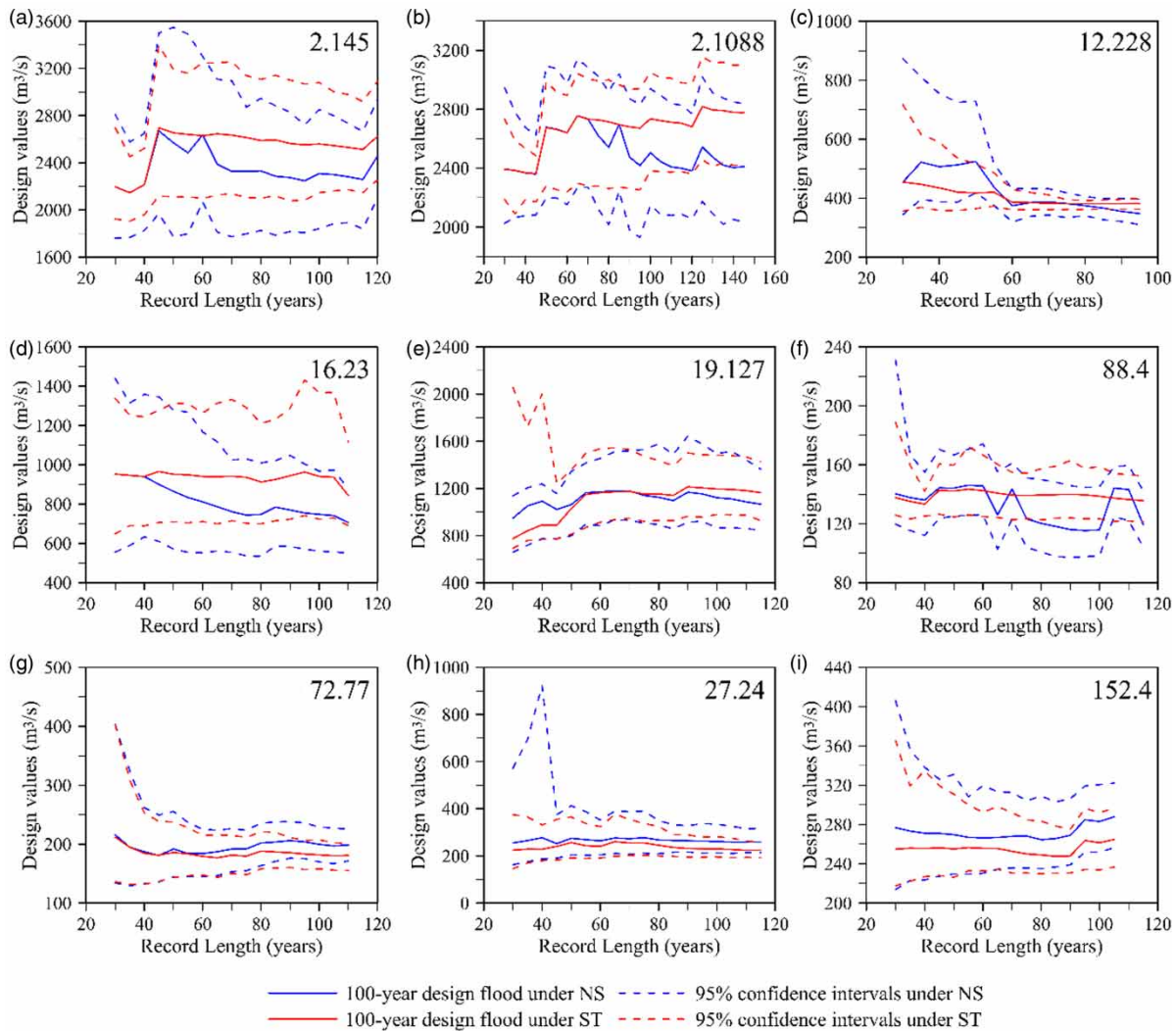
The results of this study may be affected by the employed NS model. NS models make sense if we know *a priori* mechanism of change, and therefore we can set up parameters varying accordingly. Otherwise, we need to consider the model uncertainty including a variety of possible options and combinations in terms of covariates and link functions. Theoretically, all the three statistical parameters can be time-varying, but the shape parameter is sensitive and difficult to estimate. In practice, either location or scale parameters are likely to change under a changing climate, but in this study, only the location parameter of GEV-NS0 is time-varying to facilitate the model construction and investigation procedures. Although this kind of simplified GEV-NS0 model may reduce the model set and neglect an infinite set of possible alternative NS models, the aim of this study is not to select the optimal model for a selected basin but to investigate the impacts of record lengths on NFFA. It is a trade-off choice between the complexity of model structure of the NS model and the practical applications of NS model in this study. Besides, to strengthen the physical mechanism of NS models, researchers have recommended the use of covariates more relevant to flood process in the construction of NS models to the regions where human activity is an important contributor. In this study, we used time as a covariate for two reasons: (1) to focus the study on the main objectives, which are to examine the impacts of flood record lengths on the NS and ST design floods and associated uncertainties, and to explore the probable requirements of flood record length in NFFA and (2) human activities on the selected Norwegian catchments are not



**Figure 9** | Evolution of the CVs of 20-, 50- and 100-year design flood with the increase of record lengths under both ST (dashed lines) and NS (solid lines) conditions.

significant and the used maximum extreme values are normally less affected by non-climate factors compared with daily values or low-flow values. From the perspective of feasibility, in this study, all the record lengths of selected stations are larger than 100 years. So, it is very difficult to obtain physical covariates with such a long record length. It should be strengthened that time covariate is just a simple reflection of the gradual change of time series, and it lacks physical meaning. Thus, in future, when more observations data are available, it is feasible to incorporate physical covariates which are relevant to the flood process in the NS modeling.

Theoretically, the required sample size for a given analysis depends on the target statistic of interest (e.g., 50-, 100- and 1,000-year design flood) and required accuracy. In practical flood design problems, hydrological engineers work with the available flood observations they have and account for the corresponding uncertainties to a targeted return period. The recommended minimum sample size in flood design in the handbook of hydrology is to guarantee the reliability of design results for cases with no specific degrees of uncertainty in advance, and urge the hydrological engineers to collect more samples or turn to 'peaks over threshold' sampling scheme to reduce the huge uncertainty when estimating 1,000-year design flood using limited observations. Based on the results of this study, a minimum of 60 years of observations is recommended for NFFA to guarantee the superiority of the NS models. It should be mentioned that the threshold of 60-year observations switching from FFA to NFFA is not a 'magic' number. It is inferred based on fitting quality, predictive capabilities and the stability of GEV-



**Figure 10** | Evolution of the 100-year design flood (solid lines) and associated CIs (dashed lines) with the increase of record lengths under both ST (red lines) and NS (blue lines) conditions. Please refer to the online version of this paper to see this figure in color: <http://dx.doi.org/10.2166/nh.2021.026>.

NS0 for most selected stations in Norway for estimating 20-, 50- and 100-year design flood, using GEV-NS0, which is an estimation that may depend on the nature of the time series, its memory, its distribution, the models used and the level of accuracy required. With that said, we still think that the recommendation of a plausible standard to guide the estimation of design flood for engineers, particularly for cases with limited observations, is beneficial for pushing forward the practical application of NS frequency analysis. Besides, it should be noted that in the practical design procedure, no matter whether the available observations satisfy the minimum requirement of record length, performing a cost-benefit (multicriteria) analysis considering the range of uncertainty corresponding to a predefined design value, thus getting the solution that optimizes the required criteria (i.e., structural and economic) is important. Since a number of other variables that are not specified in this study may affect the required sample size. For example, if we improve the level of required accuracy of design values in a specific problem, a series with 60 observations is likely to be not enough, or if we use the NS model with a different model structure whose location parameter and scale parameter are both time-varying, the required sample size may be different. Besides, in this study, only two stations' record lengths are larger than 120 years, influencing the analysis of probability coverage of GEV-ST and GEV-NS0. With the increase of observations in future, to promote the development of NFFA, further studies concerning the record length required for NFFA in other basins worldwide should be carried out.

## 5. CONCLUSIONS

The aim of this study was to evaluate the influence of record length on the ST and NS FFA. This aim was achieved using the split-sample testing for the stations with long record lengths (larger than 100 years), and several goodness-of-fit measures were used to evaluate the model performance and predictive capabilities of statistical models. Subsequently, the ADLL design method was employed to predict NS design floods, and the NNB was used to calculate CIs by resampling the original flood observations for record lengths varying from 30 to 145 years.

Based on the aforementioned results, the main conclusions of this study are as follows:

1. For frequency analysis of NS flood extremes, the performance of GEV-NS0 becomes better with the increase of record length and particularly they outperform GEV-ST models when record length is larger than 60 years in the selected basins in Norway, indicating the necessity of collecting more flood samples in NFFA rather than the minimum requirement of 30 samples in traditional FFA.
2. The predictive capabilities of extreme events (95th and 5th percentiles) are good for both GEV-ST and GEV-NS0 in the selected basins in Norway. However, compared with GEV-ST models, the difference between theoretical and actual probabilistic coverages becomes smaller, as record lengths increase for GEV-NS0. More importantly, the likelihood values of GEV-NS0 are larger than those of GEV-ST models when the record length is larger than 60 years for most stations, indicating the better predictive capabilities of GEV-NS0. The improved model performance and predictive capabilities of GEV-NS0 are thought to be attributed to their better capture of the variability of extreme flood events when more flood samples are added into the flood population.
3. The stability of the established GEV-ST and GEV-NS0 is improved, as record lengths increase since the CVs of 20-, 50- and 100-year design floods become smaller with the increase of record length. Particularly, the fluctuation of CV becomes gentle when the record length is roughly larger than 50–60 years for most stations in the selected basins. Besides, the CVs of GEV-ST are smaller than those of GEV-NS0.
4. The difference of 100-year design flood estimated using the ADLL method between GEV-ST and GEV-NS0 is relatively small or even becomes smaller when record length is larger than 50–60 years. In addition, the bootstrapped 95% CIs of the 100-year design flood become smaller with the increase of record length for both GEV-ST and GEV-NS0 models in the selected basins in Norway.

## ACKNOWLEDGEMENTS

This study is financially supported jointly by the National Natural Science Foundation of China (Nos. 51525902 and 51909053), the Research Council of Norway (FRINATEK Project No. 274310 and KLIMAFORSK Project No. 302457), the Ministry of Education ‘Plan 111’ Fund of China (No. B18037), the Natural Science Foundation of Hebei Province (No. E2019402076), the Open Research Fund Program of State Key Laboratory of Water Resources and Hydropower Engineering Science (No. 2019SWG02) and the Youth Foundation of Education Department of Hebei Province (No. QN2019132), all of which are greatly appreciated.

## CONFLICT OF INTEREST

The authors declare there is no conflict of interest regarding the publication of this paper.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Aceró, F. J., Parey, S., Hoang, T. T. H., Dacunha-Castelle, D., García, J. A. & Gallego, M. C. 2017 [Non-stationary future return levels for extreme rainfall over Extremadura \(southwestern Iberian Peninsula\)](#). *Hydrological Sciences Journal* **62** (9), 1394–1411. doi:10.1080/02626667.2017.1328559.
- Aceró, F. J., Parey, S., García, J. A. & Dacunha-Castelle, D. 2018 [Return level estimation of extreme rainfall over the Iberian Peninsula: comparison of methods](#). *Water* **10** (2), 179. doi:10.3390/w10020179.
- Akaike, H. 1974 [A new look at the statistical model identification](#). *IEEE Transactions on Automatic Control* **19** (6), 716–723. doi:10.1109/TAC.1974.1100705.
- Arnell, N. W. & Gosling, S. N. 2016 [The impacts of climate change on river flood risk at the global scale](#). *Climatic Change* **134** (3), 387–401. doi:10.1007/s10584-014-1084-5.



- Bayazit, M. 2015 Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. *Environmental Processes* 2 (3), 527–542. doi:10.1007/s40710-015-0081-7.
- Byun, K. & Hamlet, A. F. 2020 A risk-based analytical framework for quantifying non-stationary flood risks and establishing infrastructure design standards in a changing environment. *Journal of Hydrology* 584, 124575. doi:10.1016/j.jhydrol.2020.124575.
- Cheng, L. & Aghakouchak, A. 2014 Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports* 4 (1), 378–381. doi:10.1038/srep07093.
- Cooley, D. 2013 Return periods and return levels under climate change. In: *Extremes in a Changing Climate* (AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S. & Sorooshian, S., eds). Springer, Dordrecht, Netherlands, pp. 97–114. doi:10.1007/978-94-007-4479-0\_4.
- Debele, S. E., Strupczewski, W. G. & Bogdanowicz, E. 2017 A comparison of three approaches to non-stationary flood frequency analysis. *Acta Geophysica* 65 (4), 863–883. doi:10.1007/s11600-017-0071-4.
- Du, T., Xiong, L., Xu, C.-Y., Gippel, C. J., Guo, S. & Liu, P. 2015 Return period and risk analysis of nonstationary low-flow series under climate change. *Journal of Hydrology* 527, 234–250. doi:10.1016/j.jhydrol.2015.04.041.
- Ekeu-Wei, I. T. 2018 Evaluation of hydrological data collection challenges and flood estimation uncertainties in Nigeria. *Environment and Natural Resources Research* 8, 44–54. doi:10.5539/enrr.v8n2p44.
- Engeland, K., Wilson, D., Borsányi, P., Roald, L. & Holmqvist, E. 2017 Use of historical data in flood frequency analysis: a case study for four catchments in Norway. *Hydrology Research* 49 (2), 466–486. doi:10.2166/nh.2017.069.
- Fleig, A. K., Andreassen, L. M., Barfod, E., Haga, J., Haugen, L. E., Hisdal, H., Melvold, K. & Saloranta, T. 2013 *Norwegian Hydrological Reference Dataset for Climate Change*. Norwegian Water Resources and Energy Directorate (NVE). Rapport 2.
- Hall, J., Arheimer, B., Borga, M., Brazdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kriauciūnienė, J., Kundzewicz, Z. W. & Lang, M. 2013 Understanding flood regime changes in Europe: a state-of-the-art assessment. *Hydrology and Earth System Sciences* 18 (7), 2735–2772. doi:10.5194/hess-18-2735-2014.
- Hallegatte, S., Green, C., Nicholls, R. J. & Corfee-Morlot, J. 2013 Future flood losses in major coastal cities. *Nature Climate Change* 3 (9), 802–806. doi:10.1038/nclimate1979.
- Hao, Z. & Singh, V. P. 2016 Review of dependence modeling in hydrology and water resources. *Progress in Physical Geography* 40 (4), 549–578. doi:10.1177/0309133316632460.
- Hou, J., Li, B., Tong, Y., Ma, L., Ball, J., Luo, H., Liang, Q. & Xia, J. 2020 Cause analysis for a new type of devastating flash flood. *Hydrology Research* 51 (1), 1–16. doi:10.2166/nh.2019.091.
- Hu, Y., Liang, Z., Liu, Y., Zeng, X. & Wang, D. 2015 Uncertainty assessment of estimation of hydrological design values. *Stochastic Environmental Research and Risk Assessment* 29 (2), 501–511. doi:10.1007/s00477-014-0979-z.
- Hu, Y., Liang, Z., Singh, V. P., Zhang, X., Wang, J., Li, B. & Wang, H. 2018 Concept of equivalent reliability for estimating the design flood under non-stationary conditions. *Water Resources Management* 32 (3), 997–1011. doi:10.1007/s11269-017-1851-y.
- Hu, L., Nikolopoulos, E. I., Marra, F. & Anagnostou, E. N. 2020 Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: an evaluation over the contiguous United States. *Journal of Flood Risk Management* 13 (1), e12580. doi:10.1111/jfr3.12580.
- Jiang, C., Xiong, L., Yan, L., Dong, J. & Xu, C.-Y. 2019 Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice. *Hydrology and Earth System Sciences* 23 (3), 1683–1704. doi:10.5194/hess-23-1683-2019.
- Kan, G., Zhang, M., Liang, K., Wang, H., Jiang, Y., Li, J., Ding, L., He, X., Hong, Y., Zuo, D., Bao, Z. & Li, C. 2016 Improving water quantity simulation & forecasting to solve the energy-water-food nexus issue by using heterogeneous computing accelerated global optimization method. *Applied Energy* 210, 420–433. doi:10.1016/j.apenergy.2016.08.017.
- Khalik, 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. *Journal of Hydrology* 329 (3–4), 534–552. doi:10.1016/j.jhydrol.2006.03.004.
- Kjeldsen, T. R., Macdonald, N., Lang, M., Mediero, L., Albuquerque, T., Bogdanowicz, E., Brázdil, R., Castellari, A., David, V., Fleig, A., Gül, G. O., Kriauciūnienė, J., Kohnová, S., Merz, B., Nicholson, O., Roald, L. A., Salinas, J. L., Sarauskiene, D., Šraj, M., Strupczewski, W., Szolgay, J., Toumazis, A., Vanneville, W., Veijalainen, N. & Wilson, D. 2014 Documentary evidence of past floods in Europe and their utility in flood frequency estimation. *Journal of Hydrology* 517, 963–973. doi:10.1016/j.jhydrol.2014.06.038.
- Kleivane, I. 2006 *Vassføringsstasjonar på Vestlandet*. Norwegian Water Resources and Energy Directorate (NVE). Rapport 17.
- Kobierska, F., Engeland, K. & Thorarinsdottir, T. L. 2017 Evaluation of design flood estimates – a case study for Norway. *Hydrology Research* 49 (2), 450–465. doi:10.2166/nh.2017.068.
- Kochanek, K., Strupczewski, W. G., Bogdanowicz, E. & Markiewicz, I. 2020 The bias of the maximum likelihood estimates of flood quantiles based solely on the largest historical records. *Journal of Hydrology* 584, 124740. doi:10.1016/j.jhydrol.2020.124740.
- Kuang, Y., Xiong, L., Yu, K., Liu, P., Xu, C.-Y. & Yan, L. 2018 Comparison of first-order and second-order derived moment approaches in estimating annual runoff distribution. *Journal of Hydrologic Engineering* 23 (8), 04018034. doi:10.1061/(ASCE)HE.1943-5584.0001683.
- Li, H., Gao, H., Zhou, Y., Xu, C.-Y., Ortega, M., & Sælthun, R. Z. & R, N. 2020 Usage of SIMWE model to model urban overland flood: a case study in Oslo. *Hydrology Research* 51 (2), 366–380. doi:10.2166/nh.2020.068.
- Lovdata 2010 *Dam Safety Regulation (Damsikkerhetsforskriften)*. Norway. Available from: <https://lovdata.no/dokument/SF/forskrift/2009-12-18-1600>.
- Lu, M., Xu, Y., Shan, N., Wang, Q., Yuan, J. & Wang, J. 2019 Effect of urbanisation on extreme precipitation based on nonstationary models in the Yangtze River Delta metropolitan region. *Science of the Total Environment* 673, 64–73. doi:10.1016/j.scitotenv.2019.03.413.

- 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* **34** (1), 169–182. doi: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 Resources Research* **53** (7), 5469–5494. doi:10.1002/2016WR019676.
- Machado, M. J., Botero, B. A., López, J., Francés, F., Díez-Herrero, A. & Benito, G. 2015 Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrology and Earth System Sciences* **19** (6), 2561–2576. doi:10.5194/hess-19-2561-2015.
- Madsen, H., Lawrence, D., Lang, M., Martinkova, M. & Kjeldsen, T. R. 2014 Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology* **519**, 3634–3650. doi:10.1016/j.jhydrol.2014.11.003.
- McCuen, R. H. & Galloway, K. E. 2010 Record length requirements for annual maximum flood series. *Journal of Hydrologic Engineering* **15** (9), 704–707. doi:10.1061/(ASCE)HE.1943-5584.0000223.
- Middtømme, G., Pettersson, L. E., Holmqvist, E., Nøtsund, Ø., Hisdal, H. & Sivertsgård, R. 2011 *Guidelines for Flood Estimation*. NVE retningslinjer, Oslo, Norway.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. & Stouffer, R. J. 2008 Stationarity is dead: whither water management? *Science* **319** (5863), 573–574.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., Stouffer, R. J., Dettinger, M. D. & Krysanova, V. 2015 On critiques of ‘stationarity is dead: whither water management?’. *Water Resources Research* **51** (9), 7785–7789. doi:10.1002/2015WR017408.
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E. & Grimaldi, S. 2013 ‘Panta Rhei – Everything Flows’: change in hydrology and society – the IAHS Scientific Decade 2013–2022. *Hydrological Sciences Journal* **58** (6), 1256–1275. doi:10.1080/02626667.2013.809088.
- Obeyskera, J. & Salas, J. 2014 Quantifying the uncertainty of design floods under nonstationary conditions. *Journal of Hydrologic Engineering* **19** (7), 1438–1446. doi:10.1061/(ASCE)HE.1943-5584.0000931.
- Olsen, R., Lambert, J. H. & Haines, Y. Y. 1998 Risk of extreme events under nonstationary conditions. *Risk Analysis* **18**, 497–510. doi:10.1111/j.1539-6924.1998.tb00364.x.
- 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. doi: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. doi:10.1002/env.1060.
- Pettersson, L. E. 2003 *Norges Hydrologiske Stasjonsnett*. Rapport 7. Norwegian Water Resources and Energy Directorate (NVE).
- Qu, C., Li, J., Yan, L., Yan, P. & Lu, D. 2020 Non-stationary flood frequency analysis using cubic B-spline-based GAMLSS model. *Water* **12** (7), 1867. doi:10.3390/w12071867.
- Read, L. K. & Vogel, R. M. 2015 Reliability, return periods, and risk under nonstationarity. *Water Resources Research* **51** (8), 6381–6398. doi:10.1002/2015WR017089.
- Rigby, R. A. & Stasinopoulos, D. 2005 Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54** (3), 507–554. doi:10.1111/j.1467-9876.2005.00510.x.
- Rootzén, H. & Katz, R. W. 2013 Design life level: quantifying risk in a changing climate. *Water Resources Research* **49** (9), 5964–5972. doi:10.1002/wrcr.20425.
- Salas, J. D. & Obeyskera, J. 2014 Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering* **19** (3), 554–568. doi:10.1061/(ASCE)HE.1943-5584.0000820.
- Salas, J. D., Obeyskera, J. & Vogel, R. M. 2018 Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal* **63** (3), 325–352. doi:10.1080/02626667.2018.1426858.
- Sarhadi, A. & Soulis, E. D. 2017 Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate. *Geophysical Research Letters* **44** (5), 2454–2463. https://doi.org/10.1002/2016GL072201.
- Sen, S., He, J. & Kasiviswanathan, K. S. 2020 Uncertainty quantification using the particle filter for non-stationary hydrological frequency analysis. *Journal of Hydrology* **584**, 124666. doi:10.1016/j.jhydrol.2020.124666.
- Serago, J. M. & Vogel, R. M. 2018 Parsimonious nonstationary flood frequency analysis. *Advances in Water Resources* **112**, 1–16. doi:10.1016/j.advwatres.2017.11.026.
- Serinaldi, F. & Kilsby, C. G. 2015 Stationarity is undead: uncertainty dominates the distribution of extremes. *Advances in Water Resources* **77**, 17–36. doi:10.1016/j.advwatres.2014.12.013.
- Serinaldi, F. & Kilsby, C. G. 2016 The importance of prewhitening in change point analysis under persistence. *Stochastic Environmental Research and Risk Assessment* **30**, 763–777. doi:10.1007/s00477-015-1041-5.
- Serinaldi, F., Kilsby, C. G. & Lombardo, F. 2018 Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources* **111**, 132–155. doi:10.1016/j.advwatres.2017.10.015.
- Serinaldi, F., Chebana, F. & Kilsby, C. G. 2020 Dissecting innovative trend analysis. *Stochastic Environmental Research and Risk Assessment* **34** (5), 733–754. doi:10.1007/s00477-020-01797-x.
- Slater, L. J. & Villarini, G. 2016 Recent trends in US flood risk. *Geophysical Research Letters* **43** (24), 12–428. doi:10.1002/2016GL071199.
- Stedinger, J. R., Vogel, R. M. & Foufoula-Georgiou, E. 1993 Frequency analysis of extreme events. In: *Handbook of Hydrology* (Maidment, D. R., ed.). McGraw-Hill, New York.

- Strupczewski, W. G., Kochanek, K. & Bogdanowicz, E. 2017 Historical floods in flood frequency analysis: is this game worth the candle? *Journal of Hydrology* **554**, 800–816. doi:10.1016/j.jhydrol.2017.09.034.
- 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. *Advances in Water Resources* **124**, 29–40. doi:10.1016/j.advwatres.2018.12.004.
- 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. *Journal of Hydrology* **567**, 393–404. doi:10.1016/j.jhydrol.2018.10.031.
- TEK10 2016 *Building Regulations (Byggtekninsk forskrift)*. Available from: <https://dibk.no/byggeregler/tek/>.
- Um, M., Kim, Y., Markus, M. & Wuebbles, D. J. 2017 Modeling nonstationary extreme value distributions with nonlinear functions: an application using multiple precipitation projections for U.S. cities. *Journal of Hydrology* **552**, 396–406. doi:10.1016/j.jhydrol.2017.07.007.
- Venvik, G., Bang-Kittilsen, A. & Boogaard, F. C. 2020 Risk assessment for areas prone to flooding and subsidence – a concept model with case study from Bergen, Western Norway. *Hydrology Research* **51** (2), 322–338. doi:10.2166/nh.2019.030.
- Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D. & Krajewski, W. F. 2009 Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources* **32** (8), 1255–1266. doi:10.1016/j.advwatres.2009.05.003.
- Vogel, R. M., Yaindl, C. & Walter, M. 2011 Nonstationarity: flood magnification and recurrence reduction factors in the United States1. *Journal of the American Water Resources Association* **47** (3), 464–474. doi:10.1111/j.1752-1688.2011.00541.x.
- Vormoor, K., Lawrence, D., Heistermann, M. & Bronstert, A. 2015 Climate change impacts on the seasonality and generation processes of floods – projections and uncertainties for catchments with mixed snowmelt/rainfall regimes. *Hydrology and Earth System Sciences* **19** (2), 913–931. doi:10.5194/hess-19-913-2015.
- Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D. & Wong, W. K. 2016 Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway. *Journal of Hydrology* **538**, 33–48. doi:10.1016/j.jhydrol.2016.03.066.
- Wang, J., Xu, Y., Wang, Y., Yuan, J., Wang, Q. & Xiang, J. 2019 Non-stationarity analysis of extreme water level in response to climate change and urbanization in the Taihu Basin, China. *Stochastic Environmental Research and Risk Assessment* **33** (3), 891–904. doi:10.1007/s00477-019-01666-2.
- Westra, S., Alexander, L. V. & Zwiers, F. W. 2013 Global increasing trends in annual maximum daily precipitation. *Journal of Climate* **26** (11), 3904–3918. doi:10.1175/JCLI-D-12-00502.1.
- Xie, P., Wu, Z., Sang, Y., Gu, H., Zhao, Y. & Singh, V. P. 2018 Evaluation of the significance of abrupt changes in precipitation and runoff process in China. *Journal of Hydrology* **560**, 451–460. doi:10.1016/j.jhydrol.2018.02.036.
- Xiong, L., Yan, L., Du, T., Yan, P., Li, L. & Xu, W. 2019 Impacts of climate change on urban extreme rainfall and drainage infrastructure performance: a case study in Wuhan City, China. *Irrigation and Drainage* **68** (2), 152–164. doi:10.1002/ird.2316.
- Xiong, B., Xiong, L., Guo, S., Xu, C.-Y., Xia, J., Zhong, Y. & Yang, H. 2020 Nonstationary frequency analysis of censored data: a case study of the floods in the Yangtze River from 1470 to 2017. *Water Resources Research* **56**, e2020WR027112. doi:10.1029/2020WR027112.
- Xu, W., Jiang, C., Yan, L., Li, L. & Liu, S. 2018 An adaptive metropolis-hastings optimization algorithm of Bayesian estimation in non-stationary flood frequency analysis. *Water Resources Management* **32** (4), 1343–1366. doi:10.1007/s11269-017-1873-5.
- Yan, L., Xiong, L., Guo, S., Xu, C.-Y., Xia, J. & Du, T. 2017a Comparison of four nonstationary hydrologic design methods for changing environment. *Journal of Hydrology* **551**, 132–150. doi:10.1016/j.jhydrol.2017.06.001.
- Yan, L., Xiong, L., Liu, D., Hu, T. & Xu, C.-Y. 2017b Frequency analysis of nonstationary annual maximum flood series using the time-varying two-component mixture distributions. *Hydrological Processes* **31** (1), 69–89. doi:10.1002/hyp.10965.
- Yan, L., Li, L., Yan, P., He, H., Li, J. & Lu, D. 2019a Nonstationary flood hazard analysis in response to climate change and population growth. *Water* **11** (9), 1811. doi:10.3390/w11091811.
- Yan, L., Xiong, L., Ruan, G., Xu, C.-Y., Yan, P. & Liu, P. 2019b Reducing uncertainty of design floods of two-component mixture distributions by utilizing flood timescale to classify flood types in seasonally snow covered region. *Journal of Hydrology* **574**, 588–608. doi:10.1016/j.jhydrol.2019.04.056.
- Yan, L., Xiong, L., Luan, Q., Jiang, C., Yu, K. & Xu, C.-Y. 2020 On the applicability of the expected waiting time method in nonstationary flood design. *Water Resources Management* **34** (8), 2585–2601. doi:10.1007/s11269-020-02581-w.
- Yan, L., Xiong, L., Jiang, C., Zhang, M., Wang, D. & Xu, C.-Y. 2021 Updating intensity–duration–frequency curves for urban infrastructure design under a changing environment. *WIREs Water* e1519. doi:10.1002/wat2.1519.
- Yin, J., Gentine, 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. *Nature Communications* **9** (1), 4389. doi:10.1038/s41467-018-06765-2.
- Yue, S., Pilon, P., Phinney, B. & Cavadias, G. 2002 The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* **16** (9), 1807–1829. doi:10.1002/hyp.1095.
- Zhang, W., Villarini, G., Vecchi, G. A. & Smith, J. A. 2018 Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature* **563** (7731), 384–388. doi:10.1038/s41586-018-0676-z.
- Zhang, X., Dong, Q., Costa, V. & Wang, X. 2019 A hierarchical Bayesian model for decomposing the impacts of human activities and climate change on water resources in China. *Science of the Total Environment* **665**, 836–847. doi:10.1016/j.scitotenv.2019.02.189.
- Zwiers, F. W., Zhang, X. & Feng, Y. 2010 Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate* **24** (3), 881–892. doi:10.1175/2010JCLI3908.1.

First received 19 March 2021; accepted in revised form 5 July 2021. Available online 14 July 2021