Reducing uncertainty of design floods of two-component mixture distributions by utilizing flood timescale to classify flood types in seasonally snow covered region

Lei Yan[1,2], Lihua Xiong[1]*, Gusong Ruan[3], Chong-Yu Xu[1,4], Pengtao Yan[5], Pan Liu[1]

[1]*State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China*

[2]*College of Water Conservancy and Hydropower, Hebei University of Engineering, Handan 056021, China*

[3]*Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway*

[4]*Department of Geosciences, University of Oslo, P.O. Box 1022 Blindern, N-0315 Oslo, Norway*

[5]*School of Physics and Electronic Engineering, Xingtai University, Xingtai 054001, China*

E-mail addresses:
L. Yan (yanl@whu.edu.cn),
L. Xiong (xionglh@whu.edu.cn),
G. Ruan (guru@nve.no),
C-Y. Xu (c.y.xu@geo.uio.no),
P. Yan (mryanpt@163.com),
P. Liu (liupan@whu.edu.cn)

* *Corresponding author:*
Lihua Xiong, PhD, Professor
State Key Laboratory of Water Resources and Hydropower Engineering Science
Wuhan University, Wuhan 430072, P.R. China
E-mail: xionglh@whu.edu.cn
Telephone: +86-13871078660
Fax: +86-27-68773568

Abstract

The conventional flood frequency analysis typically assumes the annual maximum flood series (AMFS) result from a homogeneous flood population. However, actually AMFS are frequently generated by distinct flood generation mechanisms (FGMs), which are controlled by the interaction between different meteorological triggers (e.g., thunderstorms, typhoon, snowmelt) and properties of underlying surface (e.g., antecedent soil moisture and land-cover types). To consider the possibility of two FGMs in flood frequency analysis, researchers often use the two-component mixture distributions (TCMD) without explicitly linking each component distribution to a particular FGM. To improve the mixture distribution modeling in seasonally snow covered regions, an index called flood timescale (*FT*), defined as the ratio of the flood volume to peak value and chosen to reflect the relevent FGM, is employed to classify each flood into one of two types, i.e., the snowmelt-induced long-duration floods and the rainfall-induced short-duration floods, thus identifying the weighting coefficient of each component distribution beforehand. In applying the *FT*-based TCMD to model the AMFS of 34 watersheds in Norway, ten types of mixture distributions are considered. The design floods and associated confidence intervals are calculated using parametric bootstrap method. The results indicate that the *FT*-based TCMD model reduces the uncertainty in the estimation of design floods for high return periods by up to 40% with respect to the traditional TCMD. The improved predictive ability of the *FT*-based TCMD model is attributed to its explicit recognition of distinct

generation mechanisms of floods, thereby being able to identify the weighting coefficient and FGM of each component distribution without optimization.

Keywords: Flood frequency analysis; Two-component mixture distribution; Flood generation mechanisms; Flood types classification; Flood timescale; Norway

1. Introduction

The conventional flood frequency analysis is based on the assumption that the historical observations of an extreme hydrologic variable Z, denoted by z_t $(t = 1, ..., m)$ at time t, are independent and identically distributed (IID) realizations of a fixed single-type probability distribution $F_Z(z|\theta)$ whose moments and parameters are invariant. However, this IID assumption cannot be fulfilled for cases where hydrologic series exhibit more complex probabilistic structure (e.g., mixed populations and/or nonstationarity), and thus has been questioned by many researchers under either stationarity (Rulfová et al., 2016; Volpi et al., 2015; Baratti et al., 2012; Kochanek et al., 2012; Strupczewski et al., 2012; Singh et al., 2005; Klemeš, 2000; Waylen and Woo, 1982) or nonstationarity conditions (Xu, et al., 2018; Jiang et al., 2018; Schumann, 2017; Yan et al., 2017a, 2017b; Milly et al., 2015, 2008; Vogel et al., 2011; Villarini and Smith, 2010; Khaliq et al., 2006; Katz et al., 2002; Jain and Lall, 2001; Olsen et al., 1999).

Numerous studies have demonstrated the existence of flood records arising from distinct flood generation mechanisms (FGMs) due to combined actions of different meteorological conditions (e.g., thunderstorms, typhoon, cyclonic precipitation, convective precipitation and snowmelt) and basin properties (e.g., land-cover types, channel characteristics and soil moisture contents). Typically, different types of floods are mixed within a single annual maximum flood series (AMFS) with several particular FGMs dominating the flood regimes. Villarini and Smith (2010) and Smith

et al. (2011) reported that the flood events in the eastern United States are resulted from mixed populations which were dominated by landfalling tropical cyclones and extratropical systems. Barth et al. (2017) reported that the annual peak flow series in the western United States are generated from distinct FGMs, and particularly analyzed the contributions of atmospheric river to the peak flows based on 1375 stream gauge sites. Collins et al. (2014) analyzed distinct FGMs in New England and Atlantic Canada and found they were dominantly generated by storms from the Great Lakes and Coastal lows. Szolgay et al. (2016) analyzed 72 catchments in Northwest Austria and classified them into three different FGMs, i.e., rainfall-induced floods, flash floods and snowmelt-induced floods. Vormoor et al. (2015, 2016) found that two types of FGMs, i.e., rainfall-induced and snowmelt-induced floods, existed in most parts of Norway. Besides the rainfall-induced floods dominated western Norway and along the coast whereas snowmelt-induced floods dominated inland and northernmost Norway.

To address the issue of mixed flood populations generated by distinct FGMs in flood frequency analysis, researchers have developed two frequently used methods for mixture modeling (Alila et al., 2002). Provided that different FGMs are mutually independent and occur sequentially in each year, the first technique is the multiplicative model, often used for seasonal maxima, where the component distributions are combined multiplicatively. The other technique is the additive model or the finite mixture distribution (FMD) for annual maxima. In this method, the

probability distribution of AMFS is defined as weighted sum of several single-type probability distributions. Compared with the single-type distributions, the FMD is able to better model different types of skewness and tail behavior through an appropriate selection of their component distributions (Alila and Mtiraoui, 2002; McLachlan and Peel, 2000; Rossi et al., 1984). Since the increase in the number of mixture components of FMD requires larger number of observations and tends to make the parameter estimation method less robust and accurate, researchers usually assume that the AMFS are generated by two FGMs and recommend the use of two-component mixture distributions (TCMD) with different kinds of component distributions (e.g., lognormal, gamma, Weibull, Gumbel, generalized extreme value (GEV) and log Pearson type III) (Yan et al., 2017a; Zeng et al., 2014; Yoon et al., 2013; Evin et al., 2011; Villarini et al., 2011; Grego and Yates, 2010; Alila and Mtiraoui, 2002; Stedinger et al., 1993; Rossi et al., 1984; Singh and Sinclair, 1972). In the field of flood frequency analysis, often, TCMD is applied in cases where a prior identification of FGMs is not feasible due to the complexities of FGMs and the scanty long-term meteorological data needed for separating them (Evin et al., 2011; Grego and Yates, 2010; Alila and Mtiraoui, 2002; Rossi et al., 1984; Singh and Sinclair, 1972). Consequently, the distribution parameters of TCMD must be jointly estimated from the overall AMFS. However, this will probably result in some component distributions with a high probability of generating negative discharges (Bardsley, 2016) or lead to a larger standard error of the estimated quantiles (Strupczewski et al.,

120 2012). Otherwise, based on a prior separation of FGMs, the flood series are more
121 homogeneous within the flood samples belonging to each FGM. Thus, it is expected
122 to improve the physical justification of the mixture distributions and reduce the
123 standard error of the estimated design quantiles. However, to our knowledge in the
124 field of flood frequency analysis, a prior separation of annual maximum flood series
125 resulting from multi-source FGMs has not been incorporated into mixture distribution
126 modeling in estimating flood quantiles.

127 In fact, much attention has been paid to the identification and classification of
128 distinct FGMs (Brunner et al., 2017; Fischer et al., 2017; Alipour et al., 2016;
129 Antonetti et al., 2016; Berghuijs et al., 2016; Sikorska et al., 2015; Gaál et al., 2012;
130 Bárdossy and Filiz, 2005; Loukas et al., 2000). Among which, a measure named flood
131 timescale (*FT*), which is defined as the ratio of the flood volume to the flood peak was
132 proposed by Gaál et al. (2012) to improve our understanding of the interaction of
133 climate and basin processes. This event-based measure is closely related to the FGMs
134 of a basin because it integrates a series of meteorological information and basin
135 characteristics via a time parameter (Gaál et al., 2012). Fischer et al. (2016) first
136 employed the flood timescale to separate short and long flood events in summer.
137 Since the flood timescale is physically-based and does not require additional
138 meteorological information, this approach opens a way to identify distinct FGMs, and
139 determines the weighting coefficient and FGM of each component distribution
140 without optimization in mixture modeling. In the utilization of this approach, each

subsample is fitted to a single-type distribution, and then they are summed up via a weighting coefficient estimated by the proportion of each subsample in the overall AMFS.

In this study, we firstly try to identify and characterize distinct FGMs in a regional context. This is supported by the analyses of the flood seasonality and the relationship between flood volumes and peaks (flood timescale) based on 34 streamflow gauging stations throughout the entire Norway. In Norway, floods are primarily dominated by two major FGMs, i.e., the rainfall-induced floods and snowmelt-induced floods. Then we analyze the applicability and performance of the *FT*-based TCMD, denoted by TCMD-F. In the implementation of TCMD-F, we select four widely used flood probability distributions as the candidate component distributions, i.e., two-parameter lognormal (LN), Weibull (W), gamma (G), and generalized extreme value (GEV). As for parameter estimation of TCMD-F, we take the advantage of a prior separation of the observed AMFS into long-duration floods and short-duration floods based on a threshold of flood timescale. Finally, the design quantiles and associated confidence intervals (CIs), estimated by the parametric bootstrap method, of TCMD-F are compared with those yielded by a single-type distribution and the traditional TCMD, denoted by TCMD-T.

The rest of the paper is organized as follows. Firstly, we describe the study area and the data used in this study in Section 2. Secondly, the methodology used in the paper is presented in Section 3. Thirdly, the results along with several discussions of

TCMD-T and TCMD-F models are demonstrated in Section 4. Finally, the conclusions are drawn in Section 5.

2. Study area and data

Norway is located in the western part of Scandinavian Peninsula of Northern Europe, which has an approximate drainage area of 385,251 km^2 and lies between the geographical coordinates 57°-81°N and 4°-32°E (Fig. 1). The meteorological conditions in Norway exhibit large spatial variability due to its special geographical location, large latitudinal range and varied topography. The annual average air temperature ($\bar{t}_{emp}$) varies from more than 6℃ at the southern and south-western coastal regions to lower than -3℃ in the high-altitude regions in central Norway and the inland regions of northern Norway (Vormoor et al., 2016; Hanssen-Bauer et al., 2009). The annual average precipitation ($\bar{P}$) varies from approximately 300 mm in north-eastern and central-eastern Norway to more than 3500 mm in western Norway. With respect to the seasonal variation of the precipitation, the maximum precipitation volumes often occur during autumn and winter periods in western Norway, which is particularly influenced by the North Atlantic Oscillation (NAO) (Uvo, 2003), while cases are different for the inland areas of eastern Norway which experience cold dry winter and the maximum precipitation volumes concentrate on summer period (Vormoor et al., 2016; Støren and Paasche, 2014).

In most regions of Norway, both snowmelt and rainfall contribute to the runoff volume. However, due to the spatial variability of the temperature, the snowpack

183 volume and snow season vary considerably throughout the entire mainland of Norway,
184 resulting in varying levels of importance of snowmelt volumes in forming high flows.
185 Consequently, based on relative contributions of rainfall and snowmelt to floods, there
186 are three FGMs in Norway: (i) rainfall-induced floods particularly dominated in
187 western Norway and coastal regions during autumn and winter periods; (ii)
188 snowmelt-induced floods dominated in inland regions and northernmost Norway
189 particularly during spring and early summer; (iii) mixed rainfall and snowmelt driven
190 floods occurring in both autumn/winter and spring/summer (Vormoor et al., 2016,
191 2015).

192 In this study, we selected 34 watersheds throughout the entire Norway to represent
193 the three types of FGMs. It should be mentioned that only 8 of the 34 stations show
194 significant trends and only 9 of the 34 stations show significant abrupt changes at the
195 0.05 significance level, based on the results of Mann-Kendall trend test (Kendall,
196 1975) and Pettitt change point test (Pettitt, 1979). The main characteristics of these
197 watersheds, including the area, the annual mean runoff $\bar{Q}$, precipitation $\bar{P}_{rec}$ and
198 temperature $\bar{t}_{emp}$, are presented in Table 1. The daily average discharge and limited
199 peak discharge data are provided by the Norwegian Water Resources and Energy
200 Directorate's hydrometric observation network.

201 ## 3. Methodology

202 The methodologies used in the paper include: seasonality analysis method for
203 examining the existence of distinct FGMs, classification method of distinct FGMs

based on flood timescale, the method of two-component mixture distributions, the parameter estimation method of TCMD, goodness-of-fit tests and model selection criteria.

3.1. Identification of distinct FGMs by seasonality analysis

It is worth noting that before conducting mixture distribution modeling, the existence of distinct FGMs should be identified to improve the physical understanding of mixture nature of floods (Yan et al., 2017a; Villarini and Smith, 2010; Alila and Mtiraoui, 2002; Klemeš, 2000). Analyses of flood seasonality have been widely used in characterization of different FGMs within the AMFS since both the meteorological conditions and basin properties exhibit seasonal variability, and consequently some types of FGMs occur only in a specific season (Slater et al., 2017; Slater and Villarini, 2017; Yan et al., 2017a; Fischer et al., 2016; Beyene and Jain, 2015; Parajka et al., 2010; Sivapalan et al., 2005; Rossi et al., 1984). In this study, the seasonality analysis is based on the circular statistics or directional statistics (Mallakpour and Villarini, 2017; Villarini, 2016; Zhang et al., 2017; Dhakal et al., 2015; Chen et al., 2013; Burn, 1997). The flowchart of circular statistics is shown in Fig. 2.

In the circular statistics method, the date of occurrence of an annual maximum flood event z_t $(t=1,...,m)$, denoted by D_{z_t}, can be transformed to a polar coordinate Ω on a unit circle using:

$$\Omega_{z_t} = D_{z_t} \frac{2\pi}{L} \qquad 0 \le \Omega_{z_t} \le 2\pi \tag{1}$$

where L is the length of a year (L=365 or L=366 for a leap year); Ω_{z_t} is the angular

observation (in radians) of the flood event z_t. It should be noted that radian 0 represents January 1, and radian 2π represents December 31. For the AMFS with m flood events, Ω_{z_t} can be plotted on a unit circle to provide a visual representation of the flood seasonality. The direction representing the mean date of occurrence of m flood events, denoted by the polar coordinate $\bar{\Omega}$, can then be obtained by:

$$\bar{a} = \frac{1}{m}\sum_{t=1}^{m}\cos(\Omega_{z_t}) \tag{2}$$

$$\bar{b} = \frac{1}{m}\sum_{t=1}^{m}\sin(\Omega_{z_t}) \tag{3}$$

$$\bar{\Omega} = \arctan(\frac{\bar{b}}{\bar{a}}) \tag{4}$$

The variability of date of occurrence of m flood events can be characterized by the sample mean resultant length $\bar{r}$ (Burn, 1997):

$$\bar{r} = \sqrt{\bar{a}^2 + \bar{b}^2} \qquad 0 \le \bar{r} \le 1 \tag{5}$$

$\bar{r}$ is a measure of the spread of the data, ranging from 0 to 1. Values equaling to 0 indicate that the dates of occurrence of flood events are uniformly distributed throughout the year, while values equaling to 1 indicate that all the flood events occur on the same date.

$\bar{\Omega}$ and $\bar{r}$ are able to provide a preliminary and simplified summary of floods variability (Dhakal et al., 2015). In addition, we need more robust analyses to improve our understanding of the nature of model types for circular data using several well-designed statistical tests. Detailed statistical inference procedure can be found in Villarini (2016). Typically, there are three different model types of circular data, i.e.,

uniform model, reflective symmetric model and asymmetric model (Pewsey, 2013). It should be noted that the reflective symmetry model does not have much physical significance in hydrology, since we are particularly concerned with whether there is existence of seasonality (non-uniform model) or not (uniform model) when using circular statistics. So, it does not make more sense to distinguish asymmetric model or reflective symmetric model from non-uniform model. However, when we use seasonality to identify the existence of distinct FGMs, we are particularly interested in the existence of asymmetric models (multimodal) to characterize distinct FGMs. From the perspective of hypothesis test, if there is enough statistical evidence to reject the null hypothesis of uniform and reflective symmetry, the circular model is identified as asymmetry, including multimodal models, i.e., finite mixtures of unimodal symmetric and asymmetric models (Villarini, 2016). Therefore, in cases where the asymmetric model is recognized, the AMFS can be regarded as results of distinct FGMs.

3.2. Classification of distinct FGMs based on flood timescale

3.2.1. Flood timescale as an indicator for distinct FGMs

For catchments where multi-source FGMs mixed within the AMFS, a variety of process indicators (e.g., timing of the flood events, snowmelt, storm duration, rainfall depth, catchment characteristics) for classifying distinct FGMs have been suggested by Merz and Blöschl (2003). However, for practical applications, these process

indicators require meteorological or catchment-specific information which may not be available in many related studies, particular the data of snowmelt and antecedent soil moisture content. Following Bell and Kar (1969), Gaál et al. (2012) introduced the event-based measure termed flood timescale, denoted by *FT* (in hours), as a characteristic of the flood duration. *FT* is defined as the ratio of flood volume (denoted by *V*, in millimeter) and flood peak (denoted by Q_p, in millimeter/hour), which is given by:

$$FT = \frac{V}{Q_p} \tag{6}$$

The flood timescale was controlled by both meteorological conditions and basin-specific flood process (Gaál et al., 2012). In addition, Gaál et al. (2015) explored the causal factors controlling the relationship between flood peaks and volumes and argued that a weak dependence between flood peaks and volumes strongly indicates the existence of multiple FGMs. As schematically shown in Fig. 3, for cases in which the slim-type and fat-type hydrographs mixed with each other, the peak-volume relationship is not consistent and the FGM corresponds to the slope of peak-volume relationship. The slim-type hydrographs result in lower *FT* values (gentler slope), while the fat-type hydrographs result in larger ones (steeper slope). As discussed above, the flood timescale has sufficient explanatory power to distinguish multi-source flood events into groups. Fischer et al. (2016) first applied the flood timescale to specify FGMs by estimating the linear regression models between flood peaks and volumes.

It should be noted that drainage area would play a large part in the overall shape of the hydrograph in addition to the flood generation mechanism, especially for the rainfall-induced short-duration floods. It has been found that the difference in *FT* between long and short summer floods declines with the increase of drainage areas (Fischer et al., 2016). Fischer et al. (2016) also found that the timescales of long and short floods in different seasons (winter and summer) are relatively similar. It should be mentioned that we conduct the mixture modeling based on annual maxima rather than seasonal maxima, so the interaction of responses of drainage area and rainfall-generated/snowmelt-generated hydrographs through different seasons is not considered in the scope of this study.

3.2.2. Calculation of flood timescale based on disaggregated daily discharge

Following the mathematical definition of the flood timescale in Eq. (6), for an annual maximum flood event, we should determine the flood peak/maximum discharge and flood volume to estimate the flood timescale corresponding to this event. In this study, we have two kinds of discharge data, i.e., the annual peak flows (at most 30 years) and the daily average discharges (long sequence). Since the length of the observed peak flows are too short, the maximum value derived from the annual time series of daily average discharges is employed as the annual maximum discharge, and then it is used to calculate *FT* value. It is worth noting that if the annual maximum discharge is derived from the instantaneous discharge, then how sensitive is this *FT* ratio to the two types of annual maxima (e.g., annual maximum daily average

discharge and instantaneous discharge) should be further investigated, which may play a role in determining which regression line that flood event belongs to (slim or fat).

The practical calculation of the flood timescale closely depends on the estimation of flood volume associated with a flood event. To estimate the flood volume, the start and the end of a flood event should be identified. However, this is very difficult if only the daily average discharges are available, especially for small catchments whose runoff process is highly dynamic (Fischer et al., 2016; Wagner, 2012). There are several stochastic disaggregation methods for disaggregating discharge from daily scale to hourly scale (Koutsoyiannis, 2003), but in this study the case is simplified, since we focus on the disaggregation of just single flood event, not long-term disaggregation involving the simulation of wet and dry days. Therefore, in this study we use the semi-empirical approach proposed by Wagner (2012) to disaggregate daily discharges around the peak. Since the work of Wagner (2012) was written in German, here we would like to provide a brief description of this disaggregation procedure in Appendix A.

Having obtained the derived hourly hydrograph, in the next step, we should identify the beginning and end of a flood event to estimate its corresponding flood volume. In this study flood events are identified using a tool implemented in the R add-on package seriesdist (https://bitbucket.org/heisterm/seriesdist). This package enables the detection of flood peaks as well as their associated flood durations by specifying

beginning and end of the core flood event using a prescribed threshold (Vormoor et al., 2016, 2015). It should be noted that there exist other methods for specifying the beginning and end of a flood event (Longobardi et al., 2016). However, it must be mentioned that these automatic detection methods, including the seriesdist package used in this study, for the determination of beginning and end of a flood event contain an inherent level of subjectivity, and usually need manual inspection of their results. In this study, the baseflow component is also included in the estimation of event-specific flood volume to account for the role of soil moisture content in flood generation process as done by Fischer et al. (2016).

For a catchment of interest which has a sample of m annual maximum flood events, a sample of m flood timescales can be estimated based on the observed annual maximum discharge and their calculated flood volumes following Eq. (6).

3.2.3. Classification of distinct FGMs

The flood events in Norway are primarily dominated by the snowmelt-induced floods and rainfall-induced floods. The snowmelt-induced long-duration flood events (fat-type hydrographs) can typically be characterized by larger timescales than rainfall-induced short-duration flood events (slim-type hydrographs). In this study, a statistical procedure was employed to distinguish flood events from distinct FGMs into two groups according to different dependence structures between flood volumes and peaks, as proposed by Fischer et al. (2016). This classification method is based on a threshold FT_0 determined by the coefficient of determination of linear regressions

through the origin (RTO). For a sample of m flood timescales corresponding to m annual maximum flood events, denoted by FT_i $(i=1,...,m)$, if $FT_i \le FT_0$, it is assigned into the group of short-duration floods, otherwise it belongs to the group of long-duration floods. So, it is very important to calculate the threshold of the flood timescale FT_0 accurately. Firstly, FT_i are sorted in ascending order, i.e., $FT_{(1)} \le ... \le FT_{(m)}$, then we calculate the coefficient of determination $R^2(1,k)$ for the sample of first k order statistics and $R^2(k+1,m)$ for the rest of samples. FT_0 is the data point that maximizes $R^2(1,k)+R^2(k+1,m)$. See Fischer et al. (2016) for detailed information about this statistical procedure.

In order to check whether the estimated threshold FT_0 is influenced by outliers far from the center of other data, we recalculated the coefficient of determination for RTO and determined FT_0 after removing existing outliers, and found little difference as reported by Fischer et al. (2016). Of course, this issue can also be addressed by using other robust coefficient of determination for goodness-of-fit test of regression, such as the method introduced by Renaud and Victoria-Feser (2010).

3.3. The two-component mixture distributions

For cases where the existence of distinct FGMs is identified, it is appropriate and reasonable to turn to the mixture distribution modeling. In the hydrology community, the concept of finite mixture distributions was first introduced by Singh and Sinclair (1972) to address the issue of mixed flood populations in the flood frequency analysis. For a thorough discussion of this topic, see McLachlan and Peel (2000). Here, the

basic definitions and mathematical interpretations of the finite mixture distributions are briefly described as follows. For the observations of the AMFS z_t $(t=1,...,m)$, the corresponding probability density function (PDF), denoted by $f(z_t|\boldsymbol{\theta},\boldsymbol{w})$, is the weighted sum of a finite number of probability distributions, which is given by:

$$\begin{cases} f(z_t|\boldsymbol{\theta},\boldsymbol{w}) = \sum_{i=1}^{n} w_i f_i(z_t|\boldsymbol{\theta}_i) \\ \sum_{i=1}^{n} w_i = 1 \end{cases} \tag{7}$$

where $f_i(z_t|\boldsymbol{\theta}_i)$ is the *i*th density component of mixture distributions with the vector of parameters set $\boldsymbol{\theta}_i$. w_i is a weighting coefficient or mixing proportion $(0 \le w_i \le 1)$ representing the probability of z_t belonging to the *i*th density component. $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_n\}$ and $\boldsymbol{w} = (w_1,...,w_n)$. n is the number of mixture components.

In practical applications, Alila and Mtiraoui (2002) emphasizes the number of mixture components should be determined and kept to a minimum, for the reason that the increase in the number of mixture components requires larger number of observations and tends to make the parameter estimation method less robust and less accurate. In cases where a priori subdivision of the AMFS is not feasible, typically researchers assume that the AMFS are generated by two distinct FGMs and resort to the traditional two-component mixture distributions (TCMD-T), which are given by:

$$f_{TCMD-T}(z_t|\boldsymbol{\theta},w) = wf_1(z_t|\boldsymbol{\theta}_1) + (1-w)f_2(z_t|\boldsymbol{\theta}_2) \tag{8}$$

where w and $1-w$ are the probabilities of z_t belonging to an unknown flood population 1 and population 2, respectively. The vector of parameters set

$\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ represents the distribution parameters related to each component distribution. Correspondingly, the cumulative density function (CDF) of TCMD-T is given by:

$$F_{TCMD-T}(z_t \mid \boldsymbol{\theta}, w) = wF_1(z_t \mid \boldsymbol{\theta}_1) + (1-w)F_2(z_t \mid \boldsymbol{\theta}_2) \tag{9}$$

In practical applications, all the parameters of TCMD-T, namely $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, w, $1-w$, must be jointly estimated because no priori separation is done.

In this study, since we have classified the overall AMFS into two subsamples based on the flood timescale, it is reasonable to employ TCMD-F, whose PDF is given by:

$$\begin{cases} f_{TCMD-F}(z_t \mid \boldsymbol{\theta}, w) = w_L f_L(z_t \mid \boldsymbol{\theta}_L) + w_S f_S(z_t \mid \boldsymbol{\theta}_S) \\ w_L = m_L / (m_L + m_S) \\ w_S = m_S / (m_L + m_S) \end{cases} \tag{10}$$

where $f_L(\cdot)$ and $f_S(\cdot)$ refer to the PDFs for the long-duration floods component (*L*-component) and short-duration floods component (*S*-component), respectively. w_L and w_S denote the probabilities of z_t belonging to *L*-component and *S*-component, respectively. The vector of parameters set $\boldsymbol{\theta} = \{\boldsymbol{\theta}_L, \boldsymbol{\theta}_S\}$ represents the distribution parameters corresponding to $f_L(\cdot)$ and $f_S(\cdot)$, respectively. m_L is the sample size of the *L*-component and m_S is the length of the *S*-component. Correspondingly, the CDF of TCMD-F is given by:

$$F_{TCMD-F}(z_t \mid \boldsymbol{\theta}, w) = w_L F_L(z_t \mid \boldsymbol{\theta}_L) + w_S F_S(z_t \mid \boldsymbol{\theta}_S) \tag{11}$$

For the reason that the overall AMFS are classified into the *L*-component and *S*-component, the two parameter sets, i.e., $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_S$ can be separately estimated from the AMFS of each flood component. Besides, the weighting coefficients w_L

and w_S can be easily estimated by the proportion of each subsample.

It should be mentioned that TCMDs, including both TCMD-T and TCMD-F, are flexible tools which require neither the two component distributions belong to the same distribution family nor they have the same number of statistical parameters. The PDF of TCMD exists only if the component distributions are continuous (Egüen et al., 2016; Fischer et al., 2016; Shin et al., 2016; Ouarda et al., 2015). Thus, in the implementation of TCMDs, three two-parameter distributions, i.e., two-parameter lognormal distribution (LN), Weibull distribution (W), gamma distribution (G), and one three-parameter distribution, i.e., generalized extreme value distribution (GEV), are served as the candidate component distributions (Table 3) on the right-hand side of Eqs. (6)-(9). Thus, a total of 10 types of mixture distributions are considered in this study, including 4 homogeneous mixture distributions (e.g., a mixture of LN and LN) and 6 heterogeneous mixture distributions (e.g., a mixture of GEV and LN) (Table 4).

3.4. Parameter estimation of TCMD

Parameter estimation is an important procedure in the standard statistical inference. In this study, the maximum likelihood estimation method (MLE) was applied for parameter estimation of single-type probability distributions. However, if we go further and consider parameter estimation of TCMD-T, the MLE and other conventional parameter estimation methods tend to become less robust not only because of the doubled statistical parameters of TCMD-T but also because of the complexity of the estimation of weighting coefficients. To address this issue, in this

study, we use the meta-heuristic maximum likelihood estimation (MHML), which incorporates simulated annealing algorithm and MLE, to estimate parameters of TCMD-T (Yan et al. 2017a; Shin et al., 2015, 2014). MHML has advantages in estimating the weighting coefficients and finding global maximum with small samples. In addition, it can also be flexibly applied to various kinds of mixture distributions with different component distributions.

With regard to the parameter estimation of TCMD-F, we take the advantage of a prior classification of the observed AMFS into *L*-component, denoted by $z_L(i)$ $(i=1,...,m_L)$, and *S*-component, denoted by $z_S(i)$ $(i=1,...,m_S)$. Unlike the case of parameter estimation of TCMD-T, the weighting coefficients w_L and w_S were estimated by the proportions of *L*-component and *S*-component in the overall AMFS, respectively, without optimization (Eq. (10)), and the two parameter sets $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_S$ in Eq. (10) can be separately estimated from $z_L(i)$ and $z_S(i)$, respectively. Therefore, the issue of parameter estimation of mixture distributions becomes a simpler one similar to that of a single-type distribution. Just as in the case of single-type distribution, the MLE method was used to estimate the statistical parameters of each component distribution of TCMD-F.

3.5. Goodness-of-fit tests and model selection criteria

In this study, different types of extreme value distributions, including both the conventional single-type distributions and TCMD models with different component distributions were built to fit the AMFS. Therefore, to avoid model overfitting and

quantitatively evaluate the goodness-of-fit of these models, the Akaike Information Criterion (AIC) (Akaike, 1974), the adjusted coefficient of determination (R_a^2) (Shin et al., 2016), and the bootstrapped Kolmogorov-Smirnov test statistic (D_{ks}) (Sekhon, 2011) were employed.

3.5.1. Akaike Information Criterion

The Akaike Information Criterion (AIC) is used to measure the performance of a model with the level of complexity, whose expression is given by:

$$\mathrm{AIC} = -2l_{\max} + 2\rho \tag{12}$$

where $l_{\max}$ is the maximized value of the log-likelihood function for each candidate model and ρ is the total number of independently adjusted parameters of the model. The penalty term 2ρ is introduced to consider model parsimony of the distribution models. The lower the AIC score is, the better is the performance of the model.

3.5.2. Adjusted coefficient of determination

The conventional expression of the coefficient of determination R_0^2 is given by:

$$R_0^2 = 1 - \frac{\sum_{t=1}^{t=m}(F(z_t) - \hat{F}(z_t))^2}{\sum_{t=1}^{t=m}(F(z_t) - \bar{F})^2} \tag{13}$$

where $F(z_t)$ and $\hat{F}(z_t)$ are the empirical and theoretical cumulative probabilities of the tth observation z_t, respectively. $\bar{F}$ is the average empirical cumulative probability of observations. To take model parsimony into account, Shin et al. (2016) proposed the adjusted coefficient of determination R_a^2 by adding a penalty term for the number of parameters, which is given by:

$$R_a^2 = 1-(1-R_0^2)\frac{m-1}{m-\rho} \tag{14}$$

where m is the number of observations and ρ is the number of independently adjusted parameters of the model. The closer the R_a^2 is to 1, the better is the performance of the model.

3.5.3. Bootstrapped Kolmogorov-Smirnov test statistic

The conventional one-sample Kolmogorov-Smirnov test (K-S) is used to examine whether the sample is drawn from a specified distribution, and the K-S statistic is defined as:

$$D_{ks} = \max_{1\le t\le m}\left|F(z_t)-\hat{F}(z_t)\right| \tag{15}$$

where $F(z_t)$ and $\hat{F}(z_t)$ are the empirical and theoretical cumulative probabilities of the *t*th observation z_t, respectively. Note that researchers should always keep in mind the underlying distribution must be fully specified when using K-S test. That means, if location, scale, and shape parameters of the distribution are directly estimated from the observation data, the critical region of the K-S test becomes invalid, thus leading to accept the null hypothesis that the sample is generated from the prescribed distribution (Croarkin et al., 2006). To solve this problem, in this study the K-S test statistics are determined using bootstrap simulation method proposed by Sekhon (2011). The lower the D_{ks} value is, the better is the performance of the model.

3.5.4. Multi-criterion model selection measure

To comprehensively evaluate the overall performance of the employed model with respect to different goodness-of-fit measures and determine the optimal model, a multi-criterion measure is developed using the technique for order preference by similarity to ideal solution (TOPSIS) (Hwang and Yoon, 1981). TOPSIS is a widely used multi-criterion decision analysis approach which allows trade-off among different criteria, and is able to provide a ranking order for all alternative models. In this study, R_a^2 is a benefit criterion which means larger values are more appreciated, while AIC and D_{ks} are cost criteria which are more concerned about lower values. Note that each criterion is treated as equal importance when calculating the weighted normalized decision matrix. This procedure can be implemented using R package, i.e., topsis (Yazdi, 2013). See Hwang and Yoon (1981) for detailed description of TOPSIS.

4. Results and discussions

4.1. Seasonality analysis of AMFS in Norway

Robust analyses of seasonality were conducted following the flowchart in Fig. 2. The graphical representations of circular data are thought to be very helpful in visualization of clustering or seasonality. Fig. 4 summarized the preliminary analyses results of the floods seasonality via circular plots indicating the timing of AMFS. The seasonality of AMFS in Norway exhibited a spatial variability. For all stations in northern Norway or several stations in high-altitude area, flood events were

concentrated on the late spring or early summer (May and June) indicating that the flood regime was dominated by snowmelt-induced floods, while for most stations in western coastal Norway, flood events were usually concentrated on the summer and/or autumn seasons indicating that the flood regime was dominated by rainfall-induced floods. In the rest parts of Norway, AMFS were not concentrated on a particular season, especially in inland regions of southern and eastern Norway where AMFS can occur in both spring/summer and autumn/winter, indicating the existence of mixed FGMs.

The sample mean direction $\bar{\Omega}$ and mean resultant length $\bar{r}$ can give information about the time in which AMFS tend to occur and how strong the seasonality is, respectively (Table 2). As shown in Figs. 5 (a) and (b), the AMFS of several stations located in western coastal regions, high-altitude or northernmost regions exhibited seasonality and clustered in May or June, whereas the other stations showed strong variability of seasonality with $\bar{r} < 0.6$. In the latter cases there were no clusters in a particular season and mixed FGMs could exist.

According to the results of statistical inferences of AMFS in Norway (Table 2), the null hypothesis of uniformity can be rejected at 0.05 significance level except for Kråkfoss and Fustvatn stations with Rayleigh test. Especially for Fustvatn, the AMFS showed strong seasonality variability with the floods nearly evenly distributed throughout the year (Fig. 4). Then, the test for asymmetric models was conducted, and the results indicated that the null hypothesis of reflective symmetry was rejected at

0.05 significance level for half stations (Table 2). Besides, the majority of the asymmetric models were located in high-altitude regions where floods can be influenced by snow-melt during spring/summer seasons (Fig. 5c). The identification of asymmetry model can be used as indicator of mixed FGMs (Yan et al., 2017a; Villarini, 2016). It should be noted that sometimes AMFS following reflective symmetric model can occur in contrasting seasons and showing multimodal pattern (e.g., 20.2, 133.7 and 152.4 stations). Then, we selected 12 stations with seasonality (in bold in Table 2), containing 6 stations showing asymmetric type and 6 stations showing reflective symmetric type (multimodal), for subsequent classification of FGMs and mixture modeling.

4.2. Classification of FGMs

4.2.1. Disaggregation of daily discharge to hourly resolution

Fig. 6 provided a schematic diagram and illustrations of the disaggregation results using this semi-empirical approach. The results showed that the derived hourly hydrograph can reproduce the observed hourly hydrograph and preserve the daily volumes with satisfactory results. Besides, it is also obvious that the derived hourly discharge can improve the accuracy of volume estimation compared with using daily discharges. It is worth noting that this disaggregation approach does not guarantee the reproduction of flood peaks, and Fischer et al. (2016) recommended the use of observed peaks. However, for most stations in Norway we could only obtain observed

553 peaks with a length of at most 30 years, which is insufficient. Therefore in this study,
554 we use the use the observed annual maximum daily discharge for the calculation of
555 FT .

556 4.2.2. Classification

557 Based on the classification technique proposed by Fischer et al. (2016), a threshold
558 FT_0 can be determined taking account of the slopes of regression equations between
559 flood volumes and flood peaks. Therefore, the AMFS of the selected 12 stations were
560 grouped into two different FGMs, i.e., long-duration floods and short-duration floods
561 (Fig. 7). The coefficient of determination of the regression lines for short-duration
562 floods was larger than 0.9 for almost all stations, while for long-duration floods, there
563 were 4 stations whose coefficient of determination were smaller than 0.9, indicating
564 that there might exist more types of FGMs and more groups were required.
565 Nevertheless, the increase of the number of groups would reduce the sample size
566 within each group and lead to unreliable results in the subsequent statistical inference
567 procedure. Thus, AMFS were classified into only two groups in this study.

568 To further reveal the formation mechanisms associated with long-duration floods
569 and short-duration floods, we also analyzed the dates of occurrence of the two types
570 of floods using circular statistics in the Burn diagram (Burn, 1997). As shown in Fig.
571 8, for 9/12 of the stations, the proportion of long-duration flood events concentrated in
572 May and June is larger than 50%, to some extent, probably indicating the role of
573 snowmelt in the early flood events in spring/summer. On the contrary, the

574 short-duration floods were more dispersed within a year. More specifically, consistent
575 with the spatial and temporal pattern of rainfall in Norway, short-duration floods of
576 coastal areas occurred throughout the year, but those of middle and north regions did
577 not occur in winter season, probably indicating the role of rainfall in generating
578 short-duration floods.

579 *4.3. Flood frequency analysis using mixture distributions*

580 To model the heterogeneous flood series of the selected stations, both TCMD-T and
581 TCMD-F were applied on the basis of Eqs. (8) and (10). For each station, a total of 10
582 kinds of TCMD-T mixture models and 16 kinds of TCMD-F mixture models were
583 built considering different mixture types of component distributions.

584 As shown in Fig. 9, for all stations the lowest AIC values were obtained from
585 TCMD mixture models except for Krinsvatn station (ID: 133.7). In particular, the
586 GEV-GEV mixture distributions gave the best performance for almost 2/3 of stations.
587 Besides, compared with TCMD-F, TCMD-T led to better performance based on AIC
588 values for all stations except for Rygenetotal station (ID: 19.127). Fig. 10 presented
589 the boxplots of D_{ks} statistics, p-values of K-S test and R_a^2 statistics for all the
590 employed single-type and mixture distributions, respectively. Overall, all TCMD-T
591 models performed better than the single-type distributions with higher R_a^2, higher
592 p-values and lower D_{ks} values, and meanwhile nearly half of TCMD-F models
593 yielded comparable or higher R_a^2, p-values and lower D_{ks} values than the
594 single-type distribution. In addition, LN-G, GEV-LN, GEV-G, and GEV-GEV mixture

595 models had good fitting qualities for both TCMD-T and TCMD-F according to R_a^2,
596 D_{ks} and p-values measures in this analysis.

597 Given the above analysis, we have come to the conclusion that both TCMD-T and
598 TCMD-F mixture models perform better than the single-type distributions. However,
599 the performance of TCMD-F is not as good as that of TCMD-T. To further explore its
600 possible causes, we also analyzed the differences between the estimated distribution
601 parameters of TCMD-F and TCMD-T. Overall, the distribution parameters of
602 TCMD-F tended to be larger than those of TCMD-T (Fig. 11). In particular, the
603 largest over-estimations occurred in estimating the weighting coefficient w of
604 mixture distributions of LN and G for 12.228 station, while the largest
605 under-estimations occurred in estimating the shape parameter ε of mixture
606 distributions of GEV and GEV for 72.5 station. Furthermore, the ranges of estimated
607 distribution parameters of TCMD-T were generally lager than those of TCMD-F,
608 particularly for the weighting coefficient w which was fixed in TCMD-F (Fig. 12).
609 Priori classification of FGMs for TCMD-F is the reason for the difference between the
610 estimated distribution parameters of TCMD-F and TCMD-T. Therefore the imperfect
611 performance of TCMD-F is mainly due to the uncertainties, which are resulted from
612 the classification of FGMs and the parameter estimation procedure with reduced
613 sample size.

614 Fig. 13 presented the empirical frequencies and theoretical probability density
615 function of the optimal single-type distribution, the top two ranked TCMD-T models,

616 and the top two ranked TCMD-F models based on TOPSIS for each station. There
617 existed three different types of distributional characteristics of observed AMFS,
618 including the skewed unimodal (e.g., stations 41.1 and 133.7), the kurtotic unimodal
619 (e.g., station 72.5), and the asymmetric bimodal (e.g., stations 12.228, 16.23 and
620 19.127). See McLachlan and Peel (2000) for detailed description of different types of
621 distributional characteristics. Overall, TCMDs can describe different types of
622 distributional characteristics of AMFS, especially the kurtotic unimodal type and the
623 asymmetric bimodal type. Particularly, through an appropriate selection of its
624 component distributions, TCMD-T is able to better model complex types of skewness
625 and tail behavior, which seems difficult to be represented by single-type distribution.

626 *4.4. Comparison of design results*

627 To clearly demonstrate the differences in hydrologic design values between
628 TCMD-T and TCMD-F, the applications and comparisons of these two models for
629 estimating design floods and the associated CIs were demonstrated by using the
630 AMFS of Kirkevollbru station (1906-2015) (ID: 16.23). Kirkevollbru station was
631 selected as illustration for the reasons that it has a long discharge record of 108 years
632 (discarding 2 abnormal years), and the sample sizes of the two classified flood groups
633 were both larger than 45, i.e., 63 short-duration flood events and 45 long-duration
634 flood events.

635 Fig. 14a summarized the design floods for a range of return periods $T \in [2, 200]$
636 calculated by the optimal single-type (G), TCMD-T (LN-LN) and TCMD-F (LN-W)

637 models together with the return levels for the short-duration floods and long-duration
638 floods calculated by LN and W, respectively. LN-W yielded similar design flood with
639 G for $T \in [2,50]$, while LN-LN yielded smaller design floods. Besides, the estimated
640 return levels of the short-duration floods were the largest and the estimated return
641 levels of the long-duration floods were the smallest for $T \in [2,200]$, which was the
642 result of the larger flood magnitudes of the short-duration floods (Fig. 14b).

643 Quantifying uncertainty of design floods is an important procedure in conventional
644 statistical inference techniques for hydrologic designs (Obeysekera and Salas, 2014;
645 Coles, 2001). The delta method is a classical method to generate CIs. However, it
646 relies on the derivation of the covariance matrix of the estimated statistical parameters,
647 which would become further complicate and cumbersome for TCMD due to the
648 increase of model parameters. Because of the difficulties in driving analytical
649 solutions, CIs for design quantiles yielded by TCMD can be determined using the
650 parametric bootstrap method (Efron, 1979) in this study (see Appendix B).

651 Fig. 15 illustrated the return level diagram with 95% bootstrapped CIs for the
652 Kirkevollbru station. The results indicated that for both LN-LN mixture models, the
653 estimated deign floods of TCMD-F were larger than those of TCMD-T mainly
654 because of its larger weighting coefficient w and scale parameter σ of the first
655 component distribution (Table 5), while for LN-W mixture models, the estimated
656 deign floods of TCMD-F were larger than those of TCMD-T mainly because of its
657 lower w, higher μ and lower σ of the second component distribution. As for

uncertainty, the CIs of TCMD-F were always narrower than those of TCMD-T for $T \in [2, 200]$, with the largest reduction of 40%. The improved predictive ability of TCMD-F model is a result of its explicit recognition of distinct generation mechanisms of floods, thereby being able to identify the weighting coefficient and FGM of each component distribution without optimization.

4.5. Discussions

The above results confirmed the physical content of the motivation of mixture modeling and highlighted the advantages of TCMD-F models in reducing uncertainties of design floods during the statistical inference procedure. However, there are still three main comments should be made as follows:

Firstly, it should be emphasized that despite different flood generation mechanisms that can occur in flood series, the population of all flood events can more-or-less be described by a single-type distribution. The classical single distribution based frequency analysis method is convincing and cost-effective planning strategies, which is still the mainstream method in practical engineering. It's when the FGMs are markedly different from one another that we need to reevaluate our model assumptions and explore heterogeneous, mixture distribution approaches. Therefore, to strengthen the physical understanding of the mixture nature of floods, the existence of mixed populations must be identified before using mixture distributions. Conducting a statistical exercise without giving much attention to the physical process of floods is not recommended.

679 Secondly, it is worth noting that the reliability of statistical inference largely
680 depends on the sample size of flood record. However, in this study when the overall
681 sample of AMFS is divided into two subsamples based on flood timescale, the sample
682 size of each subsample will inevitably be smaller. Although we select Kirkevollbru
683 station which has a long flood record of 108 years as illustration in estimating flood
684 quantiles, uncertainty still exists. Moreover, the uncertainties would be larger if we
685 turn to nonstationary modeling, since nonstationary models are more complex and
686 possess more parameters to describe the trends of statistical parameters. Thus the
687 effects of nonstationarity is not considered in this study. In future, we plan to analyze
688 seasonal maximum flood series based on the multiplicative model, and extend it to
689 conduct nonstationary modeling. It has less model parameters (without weighting
690 coefficients) and does not reduce the sample size.
691 TCMD-T model, undoubtedly, provides accurate descriptions of AMFS with
692 mixture distributional characteristics based on the goodness-of-fit measures. However,
693 it fails to properly capture the underlying generation mechanisms and statistical
694 properties of the flood populations, compared with TCMD-F model. As discussed by
695 Bardsley (2016), TCMD-T pursues higher fitting qualities or some extra flexibility in
696 modeling mixed flood populations, however, at the expense of losing hydrological
697 mechanism. For instance, TCMD-T model is able to achieve better fit to the smallest
698 annual maximum floods by assigning a very small weighting coefficient to the
699 component distribution with smaller mean value, which may generate negative

discharges in some cases. Moreover, if so, it is inappropriate or even meaningless to utilize the mixture distributions since there actually exists only one dominant flood population. TCMD-F model, on the contrary, takes the advantage of a priori classification of flood types and is able to more effectively capture the underlying generation mechanisms and statistical properties of each flood population, and accordingly determine the weighting coefficient and FGM of each component distribution.

5. Conclusions

The main objectives of this study are to address the issue of mixed populations in the flood frequency analysis and further investigate the role of flood type classification on reducing uncertainty of design floods in the two-component mixture distributions modeling. For this purpose, ten types of mixture distributions are constructed to model the AMFS in Norway, which are classified into rainfall-induced short-duration floods and snowmelt-induced long-duration floods using flood timescale as indicator. Both the performance and design floods with 95% CIs of TCMD-F are compared with those of TCMD-T. The main conclusions of this study are drawn as follows:

(1) The seasonality of AMFS in Norway exhibits spatial variability. Mixed flood populations or distinct FGMs are identified particularly for stations located in southern and eastern inland regions of Norway based on the robust seasonality analysis of AMFS. However, there are also several stations in western coastal，

721 northernmost or high-altitude regions that exhibit strong clustering of seasonality,
722 indicating that the flood regime is dominated by rainfall-induced floods or
723 snowmelt-induced floods.

724 (2) Flood timescale is an effective tool to characterize and distinguish distinct FGMs.
725 The overall AMFS of the 12 selected stations in Norway are well classified into
726 snowmelt-induced long-duration floods and rainfall-induced short-duration floods
727 using the classification technique based on flood timescale. Overall, The
728 coefficient of determination of the regression lines for short-duration floods is
729 larger than 0.9, whereas there are 4 stations whose coefficient of determination
730 for long-duration floods are smaller than 0.9, indicating that more FGMs may
731 exist in long-duration floods.

732 (3) Mixture distributions are effective tools to capture and explain different kinds of
733 skewness and tail behavior. In general, both TCMD-T and TCMD-F mixture
734 models perform better than the single-type distributions. However, the
735 performance of TCMD-F is not as good as that of TCMD-T based on AIC, R_a^2,
736 D_{ks} and p-values, which is supposed to result from the uncertainties of
737 classification of FGMs and the parameter estimation procedure with reduced
738 sample size. Through an appropriate selection of its component distributions,
739 TCMD-T is able to better model complex types of skewness and tail behavior. In
740 addition, three kinds of heterogeneous mixture distributions (LN-G, GEV-LN and
741 GEV-G), and one kind of homogeneous mixture distributions (GEV-GEV)

perform well for both TCMD-T and TCMD-F.

(4) The estimated return levels of the short-duration floods are the largest and those of the long-duration floods are the smallest due to the larger flood magnitudes of the short-duration floods, whilst TCMDs can obtain return levels between them. Moreover, TCMD-F model is able to reduce the uncertainty in the estimation of design floods by up to 40% with respect to TCMD-T for high return periods. The improved predictive ability of TCMD-F model is attributed to its explicit recognition of distinct generation mechanisms of floods, thereby being able to identify the weighting coefficient and FGM of each component distribution without optimization.

Acknowledgements

The study is financially supported jointly by the National Natural Science Foundation of China (No. 51525902), the Research Council of Norway (FRINATEK Project 274310), and the "111 Project" Fund of China (B18037), all of which are greatly appreciated. Great thanks are due to the editor and reviewers for their professional and constructive comments and revision suggestions which are greatly helpful for the improvement of our manuscript.

Appendix A. Disaggregation of daily discharge into hourly discharge

Fig. 6a presents a schematic diagram of this disaggregation procedure. In this approach, the hourly discharge $Q(t_{h_i})$ at time step t_{h_i} (hourly time step within the current daily time step t_i) is represented using a third-order polynomial, which is given by:

$$Q(t_{h_i}) = a_{3_i} t_{h_i}^3 + a_{2_i} t_{h_i}^2 + a_{1_i} t_{h_i} + a_{0_i} \tag{A1}$$

where a_{j_i} $(j = 0,...,3)$ are the four parameters of the third-order polynomial at the current daily time step t_i. To estimate the four parameters, four conditions should be satisfied for each time step: the initial value (t_{i-1}), the volume balance of the current time step (t_i), and the volume balance of two subsequent time steps (t_{i+1} and t_{i+2}). The starting value Q_0 can be described by:

$$Q_0 = a_{3_i} t_{i-1}^3 + a_{2_i} t_{i-1}^2 + a_{1_i} t_{i-1} + a_{0_i} \tag{A2}$$

For the current time step t_i, the daily total volume can be represented by the definite integral of Eq. (A1), which is given by:

$$\begin{aligned} Q(t_i)\Delta t &= \int_{t_{i-1/2}}^{t_{i+1/2}} Q(t_{h_i}) \\ &= a_{3_i} \frac{t_{i+1/2}^4 - t_{i-1/2}^4}{4} + a_{2_i} \frac{t_{i+1/2}^3 - t_{i-1/2}^3}{3} + a_{1_i} \frac{t_{i+1/2}^2 - t_{i-1/2}^2}{2} + a_{0_i} (t_{i+1/2} - t_{i-1/2}) \end{aligned} \tag{A3}$$

where $t_{i-1/2}$ and $t_{i+1/2}$ are the beginning and end of current time step t_i, respectively. Δt is the length of the current time step. Similarly the total volume of other two time steps can be obtained. The four conditions can be characterized by a linear equation system with a general form of $K \cdot \vec{a} = \vec{c}$, as follows:

780
$$\begin{pmatrix} t_{i-1}^3 & t_{i-1}^2 & t_{i-1} & 1 \\ \frac{t_{i+1/2}^4 - t_{i-1/2}^4}{4} & \frac{t_{i+1/2}^3 - t_{i-1/2}^3}{4} & \frac{t_{i+1/2}^2 - t_{i-1/2}^2}{4} & t_{i+1/2} - t_{i-1/2} \\ \frac{t_{i+3/2}^4 - t_{i+1/2}^4}{4} & \frac{t_{i+3/2}^3 - t_{i+1/2}^3}{4} & \frac{t_{i+3/2}^2 - t_{i+1/2}^2}{4} & t_{i+3/2} - t_{i+1/2} \\ \frac{t_{i+5/2}^4 - t_{i+3/2}^4}{4} & \frac{t_{i+5/2}^3 - t_{i+3/2}^3}{4} & \frac{t_{i+5/2}^2 - t_{i+3/2}^2}{4} & t_{i+5/2} - t_{i+3/2} \end{pmatrix} \cdot \begin{pmatrix} a_{3_i} \\ a_{2_i} \\ a_{1_i} \\ a_{0_i} \end{pmatrix} = \begin{pmatrix} Q_0 \\ Q(t_i)\Delta t \\ Q(t_{i+1})\Delta t \\ Q(t_{i+2})\Delta t \end{pmatrix} \quad \text{(A4)}$$

781 This system of equations is established for each original daily time step and can be

782 solved by $\vec{a} = K^{-1} \cdot \vec{c}$.

783

Appendix B. Estimation of confidence intervals for design quantiles using parametric bootstrap method

The bootstrap method proposed by Efron (1979) is a feasible and convenient technique for generating CIs, which depends on computer simulations and resampling techniques to obtain CIs of statistical parameters and design quantiles, and has been recommended by many researchers for uncertainty analysis of hydrometeorological extremes (Rulfová et al., 2016; Serinaldi and Kilsby, 2015; Obeysekera and Salas, 2014; Serinaldi, 2009; Kyselý, 2008). The bootstrap method strictly depends on the observed data without any hypothesis and can be easily implemented despite the model complexity (Yan et al., 2017b; Serinaldi and Kilsby, 2015). Generally speaking, there exist two versions of bootstrap, namely the nonparametric bootstrap based on resampling with replacement from the original sample and the parametric bootstrap built on randomly generated samples from a specified probability distribution fitted to the original sample (Monte Carlo simulations) (Kottegoda and Rosso, 2008; Davison and Hinkley, 1997). Kyselý (2008) provided a comprehensive comparison concerning the performance of both parametric and nonparametric bootstrap methods in estimating uncertainties for extreme value distributions, and recommended the use of parametric bootstrap particularly in cases with small to moderate sample sizes. Based on the previous discussion, we employ the parametric bootstrap to generate CIs for TCMD. To generate the CIs for design quantile z_q corresponding to return period T of TCMD-T, based on Serinaldi (2009) and Kottegoda and Rosso (2008), the detailed

and general procedure of the parametric bootstrap method for TCMD-T is summarized as follows:

(1) Fit a TCMD-T model to the observed overall samples $\{z_t,\ t=1,...,m\}$ and calculate the design quantile z_q corresponding to return period T via $z_q = F^{-1}_{TCMD-T}(1-1/T \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, w)$, based on Eq. (9).

(2) Generate size-*m* bootstrap samples $\{z_t^b,\ t=1,...,m\}$ based on the fitted model at step (1). u_i $(i=1,...,m)$ are random realizations of a standard uniform distribution. If $u_i < w$, randomly generate a pseudo sample z_t^b by the inverse CDF of the unknown population 1, i.e., $F_1^{-1}(\cdot \mid \boldsymbol{\theta}_1)$ with statistical parameters $\boldsymbol{\theta}_1$, else if $u_i \geq w$, randomly generate a sample by the inverse CDF of the unknown population 2, i.e., $F_2^{-1}(\cdot \mid \boldsymbol{\theta}_2)$ with statistical parameters $\boldsymbol{\theta}_2$.

(3) Refit the bootstrapped data z_t^b using the same TCMD-T model established at step (1). Estimate new model parameters set $\boldsymbol{\theta}_1^b$, $\boldsymbol{\theta}_2^b$ and w^b, and compute the design quantile z_q for return period T via $z_q = F^{-1}_{TCMD-T}(1-1/T \mid \boldsymbol{\theta}_1^b, \boldsymbol{\theta}_2^b, w^b)$.

(4) Repeat steps (2) to (3) for a large number of times (e.g., 10000 in this study).

(5) Determine the empirical frequency distribution of z_q and calculate the corresponding confidence intervals as the $(\alpha/2)$ and $(1-\alpha/2)$ quantiles of the empirical frequency distribution of z_q.

As for generating the CIs for design quantile z_q corresponding to return period T of TCMD-F, we still take the advantage of a priori classification of the overall AMFS into *L*-component and *S*-component. The weighting coefficients w_L and w_S are

826 estimated directly from the observations and are assumed to be fixed during the
827 process of parametric bootstrap. Consequently, to preserve the mixture probabilities of
828 each component distribution, the above parametric bootstrap method should be
829 modified to allow component distributions independently and simultaneously generate
830 paired bootstrap samples $\{z_L^b(1),...,z_L^b(m_L),z_S^b(1),...,z_S^b(m_S)\}$, in which the first m_L
831 replicates are generated by the inverse CDF of the *L*-component $F_L^{-1}(\cdot|\boldsymbol{\theta}_L)$, and the
832 last m_S replicates are generated by the inverse CDF of the *S*-component $F_S^{-1}(\cdot|\boldsymbol{\theta}_S)$.
833 The above discussion is also known as the two-sample problem in the area of
834 bootstrap (Zieffler et al., 2011; Mudelsee and Alkio, 2007; Davison and Hinkley,
835 1997). Efron and Tibshirani (1986) considered a case where the data sets consist of
836 two independent random samples and modified the Monte Carlo simulations.
837 Following Efron and Tibshirani (1986), the general procedure of the parametric
838 bootstrap method for TCMD-F is summarized as follows:
839 (1) Fit a TCMD-F model to the classified samples $\{z_L(1),...,z_L(m_L),z_S(1),...,z_S(m_S)\}$,
840 and calculate the design quantile z_q corresponding to return period T via
841 $z_q = F_{TCMD-F}^{-1}(1-1/T|\boldsymbol{\theta}_L,\boldsymbol{\theta}_S,w)$, based on Eq. (11).
842 (2) Generate size-*m* bootstrap samples $\{z_L^b(1),...,z_L^b(m_L),z_S^b(1),...,z_S^b(m_S)\}$
843 $(m=m_L+m_S)$ independently and simultaneously by $F_L^{-1}(\cdot|\boldsymbol{\theta}_L)$ and $F_S^{-1}(\cdot|\boldsymbol{\theta}_S)$.
844 (3) Refit the bootstrapped data $\{z_L^b(1),...,z_L^b(m_L),z_S^b(1),...,z_S^b(m_S)\}$ using the same
845 TCMD-F model established at step (1). Estimate new statistical parameters set
846 $\boldsymbol{\theta}_L^b$ and $\boldsymbol{\theta}_S^b$ separately and compute the design quantile z_q corresponding to

847 return period T via $z_q = F^{-1}_{TCMD-F}(1-1/T \mid \boldsymbol{\theta}^b_L, \boldsymbol{\theta}^b_S, w)$.

848 (4) Repeat steps (2) to (3) for a large number of times (e.g., 10000 in this study).

849 (5) Determine the empirical frequency distribution of z_q and calculate the

850 corresponding confidence intervals as the $(\alpha/2)$ and $(1-\alpha/2)$ quantiles of

851 the empirical frequency distribution of z_q.

852

Reference

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19(6), 716-723.

Alila, Y., Mtiraoui, A., 2002. Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. Hydrol. Processes 16(5), 1065-1084.

Alipour, M.H., Rezakhani, A.T., Shamsai, A., 2016. Seasonal fractal-scaling of floods in two U.S. water resources regions. J. Hydrol. 540, 232-239.

Antonetti, M., Buss R., Scherrer S., Margreth M., Zappa M., 2016. Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations. Hydrol. Earth Syst. Sci. 20(7), 2929-2945.

Baratti, E., Montanari, A., Castellarin, A., Salinas, J.L., Viglione, A., Bezzi, A., 2012. Estimating the flood frequency distribution at seasonal and annual time scales. Hydrol. Earth Syst. Sci. 16(12), 4651-4660.

Bárdossy, A., Filiz F., 2005. Identification of flood producing atmospheric circulation patterns. J. Hydrol. 313(1-2), 48-57.

Bardsley, W.E., 2016. Cautionary note on multicomponent flood distributions for annual maxima. Hydrol. Processes 30(20), 3730-3732.

Barth, N.A., Villarini, G., Nayak, M.A., White, K., 2017. Mixed populations and annual flood frequency estimates in the western United States: the role of atmospheric rivers. Water Resour. Res. 53(1), 257-269.

Bell, F.C., Kar, S.O., 1969. Characteristic response times in design flood estimation. J. Hydrol. 8(2), 173-196.

Berghuijs, W.R., Woods, R.A., Hutton, C.J., Sivapalan, M., 2016. Dominant flood generating mechanisms across the United States. Geophys. Res. Lett. 43(9), 4382-4390.

Beyene, M. T., and S. Jain (2015), Wintertime weather-climate variability and its links to early spring ice-out in Maine lakes, Limnol. Oceanogr., 60(6), 1890–1905.

Brunner, M.I., Viviroli, D., Sikorska, A.E., Vannier, O., Favre, A., Seibert, J., 2017. Flood type specific construction of synthetic design hydrographs. Water Resour. Res. 53(2), 1390-1406.

881 Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality
882 measures. J. Hydrol. 202(1-4), 212-230.
883 Chen, L., Singh, V.P., Guo, S., Fang, B., Liu, P., 2013. A new method for identification of flood
884 seasons using directional statistics. Hydrol. Sci. J. 58(1), 28-40.
885 Coles, S., 2001. An introduction to statistical modeling of extreme values, Springer, London.
886 Collins, M.J., Kirk, J.P., Pettit, J., DeGaetano, A.T., McCown, M.S., Peterson, T.C., Means, T.N.,
887 Zhang, X., 2014. Annual floods in New England (USA) and Atlantic Canada: synoptic
888 climatology and generating mechanisms. Phys. Geogr. 35(3), 195-219.
889 Croarkin, C., Tobias, P., Filliben, J.J., 2003. NIST/Sematech e-handbook of statistical methods,
890 http://www.itl.nist.gov/div898/handbook/. (Date of access: 01/11/2017)
891 Davison, A.C., Hinkley, D.V., 1997. Bootstrap methods and their application, Cambridge University
892 Press, Cambridge, UK.
893 Dhakal, N., Jain, S., Gray, A., Dandy, M., Stancioff, E., 2015. Nonstationarity in seasonality of
894 extreme precipitation: a nonparametric circular statistical approach and its application. Water
895 Resour. Res. 51(6), 4499-4515.
896 Efron, B., 1979. Bootstrap methods: another Look at the Jackknife. Ann. Stat. 7(1), 1-26.
897 Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other
898 measures of statistical accuracy. Stat. Sci. 1(1), 54-75.
899 Egüen, M., Aguilar, C., Solari, S., Losada, M.A., 2016. Non-stationary rainfall and natural flows
900 modeling at the watershed scale. J. Hydrol. 538, 767-782.
901 Evin, G., Merleau, J., Perreault, L., 2011. Two-component mixtures of normal, gamma, and Gumbel
902 distributions for hydrological applications. Water Resour. Res. 47(8), W08525.
903 Fischer, S., Schumann, A., Schnurr, A., 2016. Ordinal pattern dependence between hydrological time
904 series. J. Hydrol., 548, 536-551.
905 Fischer, S., Schumann, A., Schulte, M., 2016. Characterisation of seasonal flood types according to
906 timescales in mixed probability distributions. J. Hydrol. 539, 38-56.
907 Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., Blöschl, G., 2012. Flood
908 timescales: understanding the interplay of climate and catchment processes through comparative

hydrology. Water Resour. Res. 48(4), W04511.

Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, K., Parajka, J., Viglione, A., Merz, R., Blöschl, G., 2015. Dependence between flood peaks and volumes: a case study on climate and hydrological controls. Hydrol. Sci. J. 60(6), 968-984.

Grego, J.M., Yates, P.A., 2010. Point and standard error estimation for quantiles of mixed flood distributions. J. Hydrol. 391(3), 289-301.

Hanssen-Bauer, I., Drange, H., Førland, E.J., Roald, L.A., Børsheim, K.Y., Hisdal, H., Lawrence, D., Nesje, A., Sandven, S., Sorteberg, A., 2009. Klima i Norge 2100-Bakgrunnsmateriale til NOU Klimatilpasning. Oslo (in Norwegian).

Hwang, C.L., Yoon, K., 1981. Multiple Attribute Decision Making: Methods and Applications. Springer-Verlag, New York.

Jain, S., Lall, U., 2001. Floods in a changing climate: does the past represent the future?. Water Resour. Res. 37(12), 3193-3205.

Jiang, C., Xiong, L., Yan, L., Dong, J., Xu, C-Y., 2018. Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/hess-2018-291

Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. in Water Resour. 25(8-12), 1287-1304.

Kendall, M.G., 1975. Rank Correlation Methods. Charles Griffin, 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.

Klemeš, V., 2000. Tall tales about tails of hydrological distributions. I. J. Hydrol. Eng. 5(3), 227-231.

Kochanek, K., Strupczewski, W.G., Bogdanowicz, E., 2012. On seasonal approach to flood frequency modelling. Part II: flood frequency analysis of Polish rivers. Hydrol. Processes 26(5), 717-730.

Kottegoda, N.T., Rosso, R., 2008. Applied Statistics for Civil and Environmental Engineers. Wiley-Blackwell, Oxford, UK.

Koutsoyiannis, D., 2003. Rainfall disaggregation methods: theory and applications, in Proceedings,

Workshop on Statistical and Mathematical Methods for Hydrological Analysis. available at https://www.itia.ntua.gr/en/docinfo/570/

Kyselý, J., 2008. A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. J. Appl. Meteorol. Clim. *47*(12), 3236-3251.

Longobardi, A., Villani, P., Guida, D., Cuomo, A., 2016. Hydro-geo-chemical streamflow analysis as a support for digital hydrograph filtering in a small, rainfall dominated, sandstone watershed. J. Hydrol. 539, 177-187.

Loukas, A., Vasiliades, L., Dalezios, N.R., 2000. Flood producing mechanisms identification in southern British Columbia, Canada. J. Hydrol. 227(1-4), 218-235.

Mallakpour, I., Villarini G., 2017. Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. Theor. Appl. Climatol. 130(1-2), 345-363.

McLachlan, G., Peel, D., 2000. Finite Mixture Model. John Wiley & Sons, New York, USA.

Merz, R., Blöschl, G., 2003. A process typology of regional floods. Water Resour. Res. 39(12), 1340.

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 Resour. Res. 51(9), 7785-7789.

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.

Mudelsee, M., Alkio, M., 2007. Quantifying effects in two-sample environmental experiments using bootstrap confidence intervals. Environ. Modell. Softw. 22(1), 84-96.

Obeysekera, J., Salas, J., 2014. Quantifying the uncertainty of design floods under nonstationary conditions. J. Hydrol. Eng. 19(7), 1438-1446.

Olsen, J.R., Stedinger, J.R., Matalas, N.C., Stakhiv, E.Z., 1999. Climate variability and flood frequency estimation for the upper Mississippi and lower Missouri rivers. J. Am. Water Resour. Assoc. 35(6), 1509-1523.

Ouarda, T.B.M.J., Charron, C., Shin, J., Marpu, P.R., Al-Mandoos, A.H., Al-Tamimi, M.H., Ghedira, H., Al Hosary, T.N., 2015. Probability distributions of wind speed in the UAE. Energy Convers.

965 Manage. 93, 414-434.
966 Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A.,
967 Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., Blöschl, G., 2010.
968 Seasonal characteristics of flood regimes across the Alpine-Carpathian range. J. Hydrol. 394(1-2),
969 78-89.
970 Pettitt, A.N., 1979. A non-parametric approach to the change-point detection. Appl. Statist. 28(2),
971 126-135.
972 Pewsey, A., Neuhäuser, M., Ruxton, G.D., 2013. Circular Statistics in R. Oxford University Press,
973 Oxford.
974 Renaud, O., Victoria-Feser, M.P., 2010. A robust coefficient of determination for regression. J. Stat.
975 Plann. Inference 140 (7), 1852–1862.
976 Rossi, F., Fiorentino, M., Versace, P., 1984. Two-component extreme value distribution for flood
977 frequency analysis. Water Resour. Res. 20(7), 847-856.
978 Rulfová, Z., Buishand, A., Roth, M., Kyselý, J., 2016. A two-component generalized extreme value
979 distribution for precipitation frequency analysis. J. Hydrol. 534, 659-668.
980 Schumann, A., 2017. Flood safety versus remaining risks-options and limitations of probabilistic
981 concepts in flood management. Water Resour. Manag. 31(10), 3131-3145.
982 Sekhon, J.S., 2011. Multivariate and propensity score matching software with automated balance
983 optimization: the matching package for R. J. Stat. Softw. 42(7), 1-52.
984 Serinaldi, F., 2009. Assessing the applicability of fractional order statistics for computing confidence
985 intervals for extreme quantiles. J. Hydrol. 376(3-4), 528-541.
986 Serinaldi, F., Kilsby, C.G., 2015. Stationarity is undead: uncertainty dominates the distribution of
987 extremes. Adv. Water Resour. 77, 17-36.
988 Shin, J., Heo, J., Jeong, C., Lee, T., 2014. Meta-heuristic maximum likelihood parameter estimation of
989 the mixture normal distribution for hydro-meteorological variables. Stoch. Env. Res. Risk A.
990 28(2), 347-358.
991 Shin, J., Lee, T., Ouarda, T.B.M.J., 2015. Heterogeneous mixture distributions for modeling
992 multisource extreme rainfalls. J. Hydrometeorol. 16(6), 2639-2657.

993 Shin, J., Ouarda, T.B.M.J., Lee, T., 2016. Heterogeneous mixture distributions for modeling wind
994 speed, application to the UAE. Renew. Energ. 91, 40-52.
995 Sikorska, A.E., Viviroli, D., Seibert, J., 2015. Flood-type classification in mountainous catchments
996 using crisp and fuzzy decision trees. Water Resour. Res. 51(10), 7959-7976.
997 Singh, K.P., Sinclair, R.A., 1972. Two-distribution method for flood-frequency analysis. J. Hydraul.
998 Div. Amer. Soc. Civil Eng. 98 (HYl), 29-44.
999 Singh, V.P., Wang, S.X., Zhang, L., 2005. Frequency analysis of nonidentically distributed hydrologic
1000 flood data. J. Hydrol. 307(1-4), 175-195.
1001 Sivapalan, M., Blöschl, G., Merz, R., Gutknecht, D., 2005. Linking flood frequency to long-term water
1002 balance: incorporating effects of seasonality. Water Resour. Res. 41(6), W06012.
1003 Slater, L.J., Villarini, G., 2017. Evaluating the drivers of seasonal streamflow in the U.S. Midwest.
1004 Water 9(9), 695.
1005 Slater, L.J., Villarini, G., Bradley, A.A., Vecchi, G.A., 2017. A dynamical statistical framework for
1006 seasonal streamflow forecasting in an agricultural watershed. Clim. Dyn.
1007 https://doi.org/10.1007/s00382-017-3794-7. (in press)
1008 Smith, J.A., Villarini, G., Baeck, M.L., 2011. Mixture distributions and the hydroclimatology of
1009 extreme rainfall and flooding in the eastern United States. J. Hydrometeorol. 12(2), 294-309.
1010 Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In
1011 Maidment, D.R. (Ed), Handbook of Hydrology, McGraw-Hill, New York.
1012 Støren, E.N., Paasche, Ø., 2014. Scandinavian floods: from past observations to future trends. Global
1013 Planet. Change 113, 34-43.
1014 Strupczewski, W.G., Kochanek, K., Bogdanowicz, E., Markiewicz, I., 2012. On seasonal approach to
1015 flood frequency modelling. Part I: Two-component distribution revisited. Hydrol. Processes 26(5),
1016 705-716.
1017 Szolgay, J., Gaál, L., Bacigál, T., Kohnová, S., Hlavčová, K., Výleta, R., Parajka, J., Blöschl, G., 2016.
1018 A regional comparative analysis of empirical and theoretical flood peak-volume relationships. J.
1019 Hydrol. Hydromech. 64, 367.
1020 Uvo, C.B., 2003. Analysis and regionalization of northern European winter precipitation based on its

relationship with the North Atlantic oscillation. Int. J. Climatol. 23(10), 1185-1194.

Villarini, G., 2016. On the seasonality of flooding across the continental United States. Adv. Water Resour. 87, 80-91.

Villarini, G., Smith, J.A., 2010. Flood peak distributions for the eastern United States. Water Resour. Res. 46(6), W06504.

Villarini, G., Smith, J.A., Baeck, M.L., Krajewski, W.F., 2011. Examining flood frequency distributions in the midwest U.S.. J. Am. Water Resour. As. 47(3), 447-463.

Vogel, R.M., Yaindl, C., Walter, M., 2011. Nonstationarity: flood magnification and recurrence reduction factors in the United States1. J. Am. Water Resour. As. 47(3), 464-474.

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., Koutsoyiannis, D., 2015. One hundred years of return period: strengths and limitations. Water Resour. Res. 51(10), 8570-8585.

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. Hydrol. Earth Syst. Sci. 19(2), 913-931.

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. J. Hydrol. 538, 33-48.

Wagner, M., 2012. Regionalisierung von Hochwasserscheiteln auf Basis einergekoppelten Niederschlag–Abfluss-Statistik mit besonderer Beachtung von Extremereignissen Dissertation. Inst. für Hydrologie und Meteorologie Lehrstuhl für Hydrologie. (in German)

Waylen, P., Woo, M.K., 1982. Prediction of annual floods generated by mixed processes. Water Resour. Res. 18(4), 1283-1286.

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 Resour. Manag. 32(4), 1343-1366.

Yan, L., Xiong, L., Liu, D., Hu, T., Xu, C-Y., 2017a. Frequency analysis of nonstationary annual maximum flood series using the time-varying two-component mixture distributions. Hydrol. Processes 31(1), 69-89.

1049 Yan, L., Xiong, L., Guo, S., Xu, C-Y., Xia, J., Du, T., 2017b. Comparison of four nonstationary
1050 hydrologic design methods for changing environment. J. Hydrol. 551, 132-150.
1051 Yazdi, M.M., 2013. topsis: TOPSIS method for multiple-criteria decision making (MCDM).
1052 https://CRAN.R-project.org/package=topsis. (Date of access: 01/11/2017)
1053 Zhang, Q., Gu, X., Singh, V.P., Shi, P., Luo, M., 2017. Timing of floods in southeastern China:
1054 seasonal properties and potential causes. J. Hydrol. 552, 732-744.
1055 Zeng, H., Feng, P., Li, X., 2014. Reservoir flood routing considering the non-stationarity of flood series
1056 in north China. Water Resour. Manag. 28(12), 4273-4287.
1057 Zieffler, A.S., Harring, J.R., Long, J.D., 2011. Comparing groups: randomization and bootstrap
1058 methods using R. John Wiley & Sons, New Jersey.

Table

Click here to download Table: Table1.docx

Table 1. Data Information of the 34 watersheds in Norway

Station ID	Name	Area (km^2)	Longitude	Latitude	Data period	$\bar{Q}$ (mm/yr)	$\bar{P}_{rec}$ (mm/yr)	$\bar{t}_{emp}$ (℃)
2.268	Akslen	789.3	8.447	61.800	1934-2015	992.7	1195.6	-3.18
2.279	Kråkfoss	435.2	11.080	60.133	1966-2015	613.0	1030.7	2.69
2.291	Tora	262.1	7.866	62.008	1967-2015	1511.1	1542.5	-2.30
2.32	Atnasjø	463.3	10.222	61.852	1917-2015	705.4	859.0	-2.10
2.614	Rosten	1833	9.405	61.859	1917-2015	558.6	884.3	-1.31
12.228	Kistefoss	3703	10.362	60.222	1917-2015	502.3	1035.5	1.11
12.7	Etna	570.3	9.626	60.952	1920-2015	541.6	1177.0	-0.58
15.21	Jondalselv	126	9.555	59.707	1920-2015	750.5	1212.8	2.26
16.23	Kirkevollbru	3845.4	9.038	59.690	1906-2015	755.2	1475.4	-0.66
19.127	Rygenetotal	3946.4	8.670	58.411	1900-2015	930.8	1512.7	3.43
20.2	Austenå	276.4	8.101	58.840	1925-2015	1224.8	1872.1	2.42
22.4	Kjæøemo	1757.7	7.528	58.120	1897-2015	1490.2	2266.3	3.62
24.9	Tingvatn	272.2	7.223	58.401	1923-2015	1755.2	2628.5	3.56
27.24	Helleland	184.7	6.149	58.534	1897-2015	2338.0	3430.2	4.69
28.7	Haugland	139.4	5.648	58.693	1919-2015	1520.7	2082.9	6.31
41.1	Stordalsvatn	130.7	6.010	59.683	1913-2015	3093.8	4029.7	3.93
50.1	Hølen	232.7	6.746	60.357	1923-2015	1596.8	2671.5	0.33
72.5	Brekkebru	268.2	7.114	60.850	1944-2014	1940.4	2383.8	-0.36
75.23	Krokenelv	45.9	7.398	61.347	1965-2015	1537.7	1976.3	0.70
76.5	Nigardsbrevatn	65.3	7.242	61.667	1963-2015	3082.0	3221.6	-1.34
88.4	Lovatn	234.9	6.890	61.859	1900-2015	2148.7	2872.3	0.36
122.11	Eggafoss	655.2	11.184	62.890	1941-2015	833.5	1160.1	-0.03
122.17	Hugdalbru	545.9	10.246	62.994	1973-2015	750.2	1136.6	1.45
122.9	Gaulfoss	3085.9	10.229	63.108	1958-2015	849.0	1182.3	0.78
123.31	Kjeldstad	143	11.131	63.266	1930-2015	1608.0	1441.7	2.21
133.7	Krinsvatn	206.6	10.232	63.804	1916-2015	1903.3	2337.0	3.80
152.4	Fustvatn	525.7	13.308	65.905	1909-2015	1933.0	2365.0	1.60
163.5	Junkerdalselv	422	15.411	66.815	1938-2015	1079.6	1294.2	-1.44
191.2	Øvrevatn	526	17.941	68.858	1914-2015	1294.4	1642.6	-0.70
223.1	Stabburselv	1067.3	24.883	70.176	1924-2015	641.1	697.7	-1.82
224.1	Skoganvarre	940.7	25.085	69.837	1922-2014	504.0	598.2	-2.33
234.18	Polmak	14161.4	28.016	70.070	1912-2015	379.1	527.9	-3.01
247.3	Karpelva	128.9	30.384	69.660	1928-2015	556.9	668.5	-0.76
311.6	Nybergsund	4424.9	12.322	61.259	1909-2015	493.2	894.3	-0.90

Table

Click here to download Table: Table2.docx

Table 2. Results of seasonality analyses of AMFS based on the circular statistical analysis. The zero direction is at radian $\pi/2$ from the mathematical origin corresponding to the positive horizontal axis. The stations in bold are selected for subsequent analyses. It should be noted that ** denotes p value < 0.05 and * denotes 0.05 < p value < 0.1.

Station ID	Basic circular statistics		Tests for uniformity				Tests for symmetry
	$\bar{\Omega}$ (radian)	$\bar{r}$	Rayleigh	Kuiper	Watson	Rao spacing	Asymptotic theory based test
2.268	3.07 (26 Jun)	0.85	0.85**	6.08**	3.61**	245.27**	2.20**
2.279	3.03 (24 Jun)	0.14	0.14	2.72**	0.47**	215.68**	2.09**
2.291	3.02 (24 Jun)	0.94	0.94**	5.91**	2.91**	288.64**	0.75
2.32	2.71 (6 Jun)	0.88	0.88**	7.38**	4.97**	265.44**	3.19**
2.614	2.75 (8 Jun)	0.95	0.95**	7.93**	5.83**	284.94**	2.24**
12.228	3.04 (25 Jun)	0.57	0.57**	4.62**	2.19**	202.64**	5.65**
12.7	2.50 (25 May)	0.85	0.85**	7.67**	4.92**	280.94**	3.43**
15.21	3.15 (1 Jul)	0.42	0.42**	4.03**	1.21**	190.65**	1.53
16.23	3.16 (2 Jul)	0.53	0.53**	4.96**	1.99**	201.40**	5.08**
19.127	4.48 (17 Sep)	0.27	0.27**	3.23**	0.80**	166.58**	6.82**
20.2	3.58 (26 Jul)	0.29	0.29**	3.70**	0.82**	203.34**	0.40
22.4	5.08 (22 Oct)	0.33	0.33**	3.14**	0.93**	164.95**	2.07**
24.9	5.42 (11 Nov)	0.45	0.45**	4.20**	1.27**	179.34**	0.69
27.24	5.56 (19 Nov)	0.59	0.59**	4.69**	2.22**	171.85**	0.38
28.7	5.69 (27 Nov)	0.59	0.59**	4.48**	1.77**	180.61**	0.47
41.1	5.12 (25 Oct)	0.51	0.51**	3.87**	1.52**	167.64**	1.31
50.1	3.00 (22 Jun)	0.76	0.76**	6.44**	3.65**	245.15**	3.63**
72.5	3.35 (13 Jul)	0.69	0.69**	4.63**	1.99**	222.02**	5.05**
75.23	2.89 (16 Jun)	0.76	0.76**	4.87**	1.98**	253.73**	3.06**
76.5	3.75 (6 Aug)	0.94	0.94**	5.65**	2.94**	266.01**	0.94
88.4	3.61 (29 Jul)	0.92	0.92**	7.95**	6.05**	265.93**	2.58**
122.11	2.61 (31 May)	0.92	0.92**	6.71**	4.21**	270.84**	2.51**
122.17	2.54 (27 May)	0.86	0.86**	4.27**	1.65**	244.88**	2.20**
122.9	2.77 (9 Jun)	0.81	0.81**	4.37**	1.84**	240.22**	3.59**
123.31	3.33 (11 Jul)	0.35	0.35**	2.66**	0.62**	159.80**	2.31**
133.7	0.26 (15 Jan)	0.34	0.34**	2.65**	0.64**	146.88**	0.94
152.4	3.64 (30 Jul)	0.07	0.07	1.77**	0.23**	154.88**	0.51
163.5	3.00 (23 Jun)	0.85	0.85**	5.72**	3.24**	234.26**	0.80
191.2	3.07 (26 Jun)	0.63	0.63**	4.78**	2.39**	197.23**	1.00
223.1	2.75 (8 Jun)	0.96	0.96**	7.98**	5.75**	293.52**	1.26
224.1	2.72 (6 Jun)	0.96	0.96**	8.07**	5.77**	291.82**	1.54
234.18	2.52 (26 May)	0.97	0.97**	8.78**	6.82**	304.14**	1.01
247.3	2.49 (24 May)	0.93	0.93**	7.85**	5.29**	294.33**	1.33

311.6	2.52 (26 May)	0.78	0.78**	6.34**	4.05**	231.06**	0.35

Table

Click here to download Table: Table3.docx

Table 3. Summary of the employed single-type distributions

Distributions	Probability density function (PDF)	Number of parameters
Lognormal (LN)	$f_{LN}(z \mid \mu_{LN}, \sigma_{LN}) = \frac{1}{\sqrt{2\pi}\sigma_{LN}} \frac{1}{z} \exp\left[-\frac{(\log(z) - \mu_{LN})^2}{2\sigma_{LN}^2} \right]$ $z > 0, \mu_{LN} > 0, \sigma_{LN} > 0$	2
Gamma (G)	$f_G(z \mid \mu_G, \sigma_G) = \frac{1}{(\mu_G \sigma_G^2)^{1/\sigma_G^2}} \frac{z^{(1/\sigma_G^2 - 1)} e^{-z/(\mu_G \sigma_G^2)}}{\Gamma(1/\sigma_G^2)}$ $z > 0, \mu_G > 0, \sigma_G > 0$	2
Weibull (W)	$f_W(z \mid \mu_W, \sigma_W) = \frac{\sigma_W z^{\sigma_W - 1}}{\mu_W^{\sigma_W}} \exp\left[-\left(\frac{z}{\mu_W} \right)^{\sigma_W} \right]$ $z > 0, \mu_W > 0, \sigma_W > 0$	2
GEV	$f_Z(z \mid \mu_{GEV}, \sigma_{GEV}, \varepsilon_{GEV}) = \frac{1}{\sigma_{GEV}} \left[1 + \varepsilon_{GEV} \left(\frac{z - \mu_{GEV}}{\sigma_{GEV}} \right) \right]^{(-1/\varepsilon_{GEV}) - 1} \exp\left\{ -\left[1 + \varepsilon_{GEV} \left(\frac{z - \mu_{GEV}}{\sigma_{GEV}} \right) \right]^{-1/\varepsilon_{GEV}} \right\}$ $-\infty < z < \infty, -\infty < \mu_{GEV} < \infty, \sigma_{GEV} > 0, -\infty < \varepsilon_{GEV} < \infty$	3

Table

Click here to download Table: Table4.docx

Table 4. TCMDs used to model the AMFS in the study

Distributions	Probability density function (PDF)	Number of parameters
LN-LN	$f_{LN-LN}(z \mid w,\mu_{LN1},\sigma_{LN1},\mu_{LN2},\sigma_{LN2}) = wf_{LN}(z \mid \mu_{LN1},\sigma_{LN1}) + (1-w) f_{LN}(z \mid \mu_{LN2},\sigma_{LN2})$ $z > 0$	5
G-G	$f_{G-G}(z \mid w,\mu_{G1},\sigma_{G1},\mu_{G2},\sigma_{G2}) = wf_{G}(z \mid \mu_{G1},\sigma_{G1}) + (1-w) f_{G}(z \mid \mu_{G2},\sigma_{G2})$ $z > 0$	5
W-W	$f_{W-W}(z \mid w,\mu_{W1},\sigma_{W1},\mu_{W2},\sigma_{W2}) = wf_{G}(z \mid \mu_{W1},\sigma_{W1}) + (1-w) f_{G}(z \mid \mu_{W2},\sigma_{W2})$ $z > 0$	5
GEV-GEV	$f_{GEV-GEV}(z \mid w,\mu_{GEV1},\sigma_{GEV1},\varepsilon_{GEV1},\mu_{GEV2},\sigma_{GEV2},\varepsilon_{GEV2}) =$ $wf_{GEV}(z \mid \mu_{GEV1},\sigma_{GEV1},\varepsilon_{GEV1}) + (1-w) f_{GEV}(z \mid \mu_{GEV2},\sigma_{GEV2},\varepsilon_{GEV2})$ $-\infty < z < \infty$	7
LN-G	$f_{LN-G}(z \mid w,\mu_{LN},\sigma_{LN},\mu_{G},\sigma_{G}) = wf_{LN}(z \mid \mu_{LN},\sigma_{LN}) + (1-w) f_{G}(z \mid \mu_{G},\sigma_{G})$ $z > 0$	5
LN-W	$f_{LN-W}(z \mid w,\mu_{LN},\sigma_{LN},\mu_{W},\sigma_{W}) = wf_{LN}(z \mid \mu_{LN},\sigma_{LN}) + (1-w) f_{W}(z \mid \mu_{W},\sigma_{W})$ $z > 0$	5
G-W	$f_{G-W}(z \mid w,\mu_{G},\sigma_{G},\mu_{W},\sigma_{W}) = wf_{G}(z \mid \mu_{G},\sigma_{G}) + (1-w) f_{W}(z \mid \mu_{W},\sigma_{W})$ $z > 0$	5
GEV-L	$f_{GEV-L}(z \mid w,\mu_{GEV},\sigma_{GEV},\varepsilon_{GEV},\mu_{L},\sigma_{L}) = wf_{GEV}(z \mid \mu_{GEV},\sigma_{GEV},\varepsilon_{GEV}) + (1-w) f_{L}(z \mid \mu_{L},\sigma_{L})$ $-\infty < z < \infty$	6
GEV-G	$f_{GEV-G}(z \mid w,\mu_{GEV},\sigma_{GEV},\varepsilon_{GEV},\mu_{G},\sigma_{G}) = wf_{GEV}(z \mid \mu_{GEV},\sigma_{GEV},\varepsilon_{GEV}) + (1-w) f_{G}(z \mid \mu_{G},\sigma_{G})$ $-\infty < z < \infty$	6
GEV-W	$f_{GEV-W}(z \mid w,\mu_{GEV},\sigma_{GEV},\varepsilon_{GEV},\mu_{W},\sigma_{W}) = wf_{GEV}(z \mid \mu_{GEV},\sigma_{GEV},\varepsilon_{GEV}) + (1-w) f_{W}(z \mid \mu_{W},\sigma_{W})$ $-\infty < z < \infty$	6

Table

Click here to download Table: Table5.docx

Table 5. Summary of the estimated parameters of the optimal TCMD-T and TCMD-F models (in bold) and their comparison models fitted to the AMFS of Kirkevollbru station (ID: 16.23). μ_1 and σ_1 are the distribution parameters belonging to the first component distribution, while μ_2 and σ_2 are the distribution parameters belonging to the second component distribution.

Optimal models	w	μ_1	σ_1	μ_2	σ_2
LN-LN (TCMD-T)	0.152	4.893	0.089	5.818	0.404
LN-LN (TCMD-F)	0.583	5.873	0.428	5.423	0.461
LN-W (TCMD-T)	0.868	5.799	0.418	135.5	16.50
LN-W (TCMD-F)	0.583	5.873	0.428	288.4	1.891
G (overall AMFS)	--	330.4	0.484	--	--

Figure

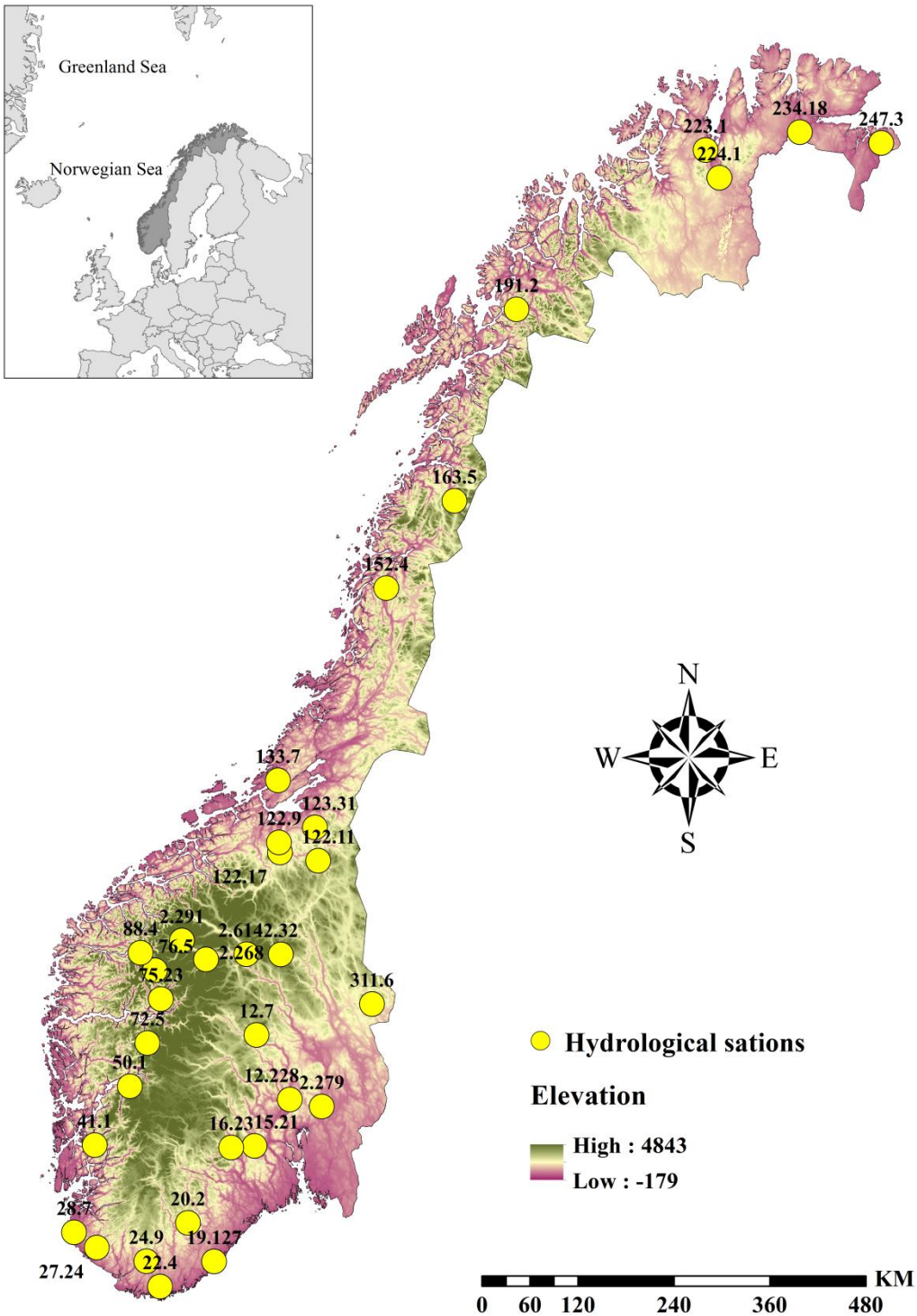

Fig. 1. Location of the selected 34 hydrological stations throughout the entire Norway mainland. The inserted frame in the top left corner depicts the geographical location of Norway in the map of Nordic region.

Figure

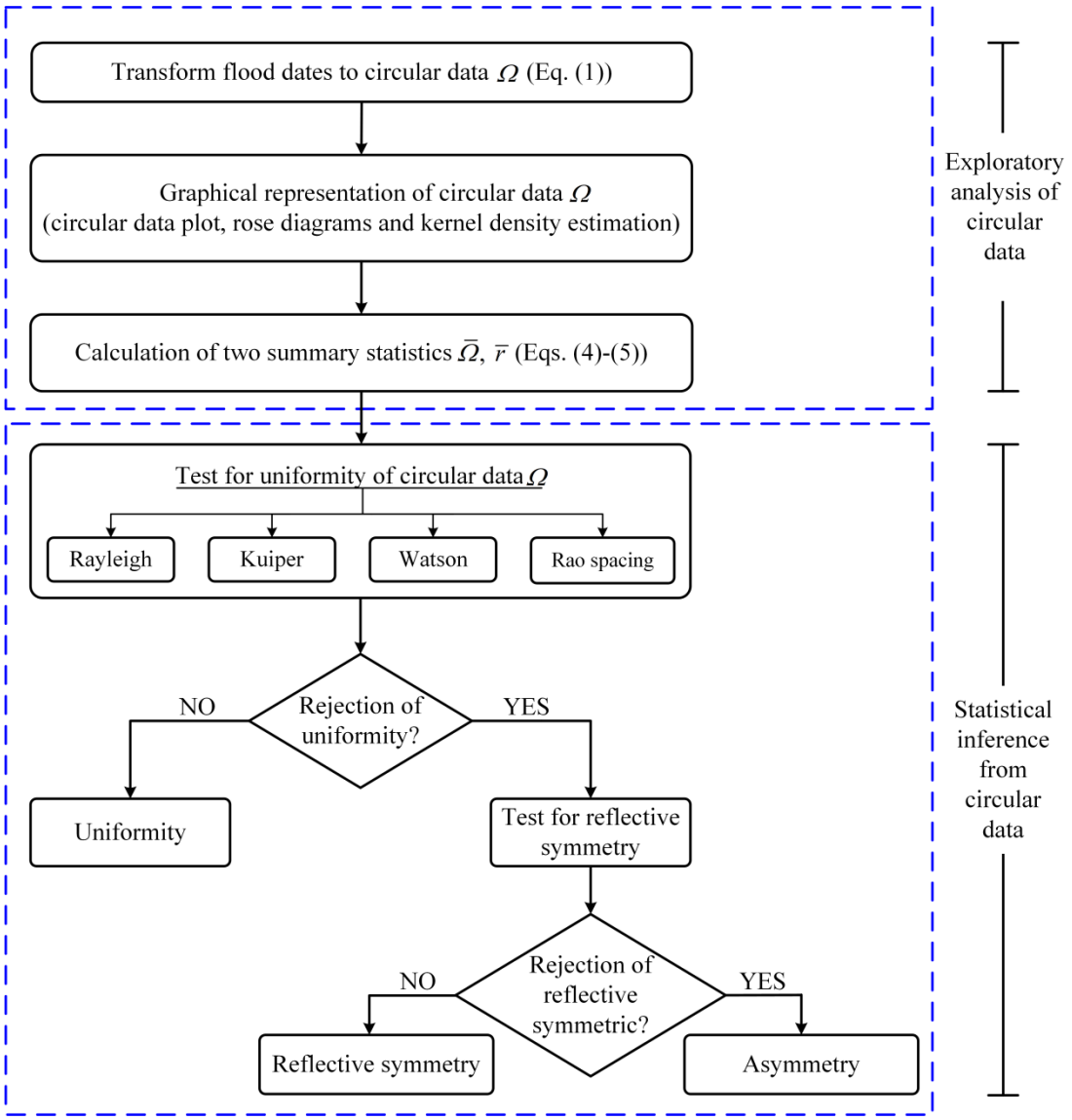

Fig. 2. Flow chart of exploratory analysis of circular data and statistical inference from circular model (adapted from Yan et al., 2017a).

Figure

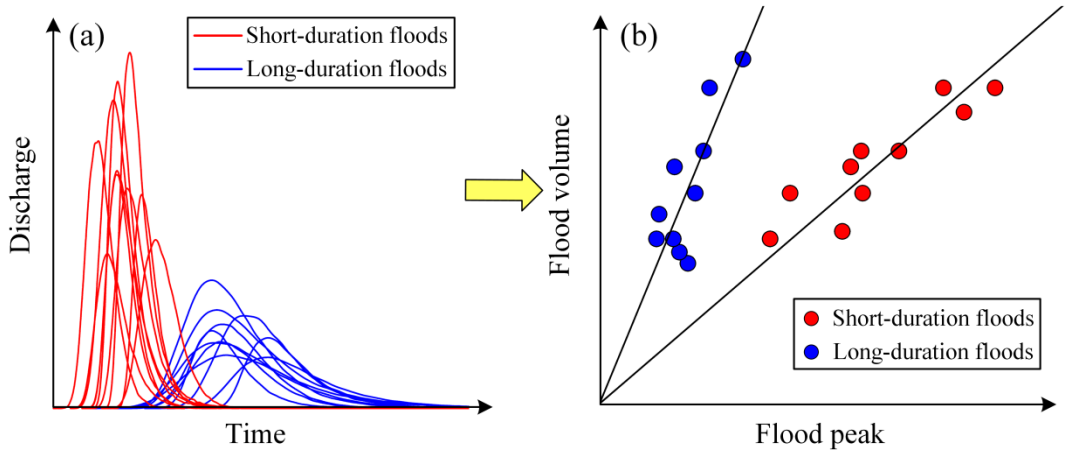

Fig. 3. Typical hydrographs (a) and the associated peak-volume relationships (b) of two types of flood events.

Figure

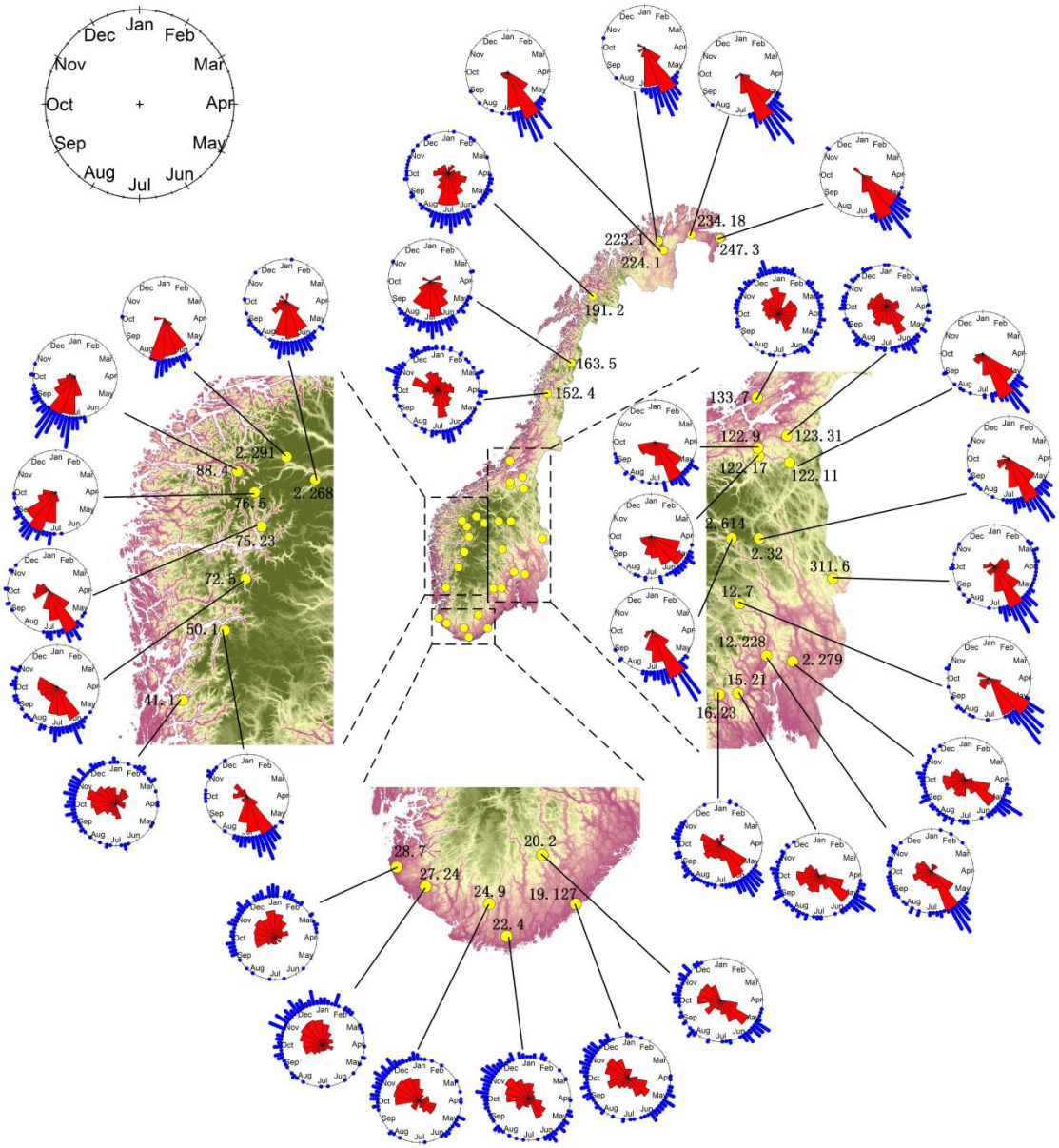

Fig. 4. Circular data for the selected stations in Norway. The blue points around the circle represent the timing of observed annual maximum flood events and the red wedges in the circles are the rose diagrams.

Figure

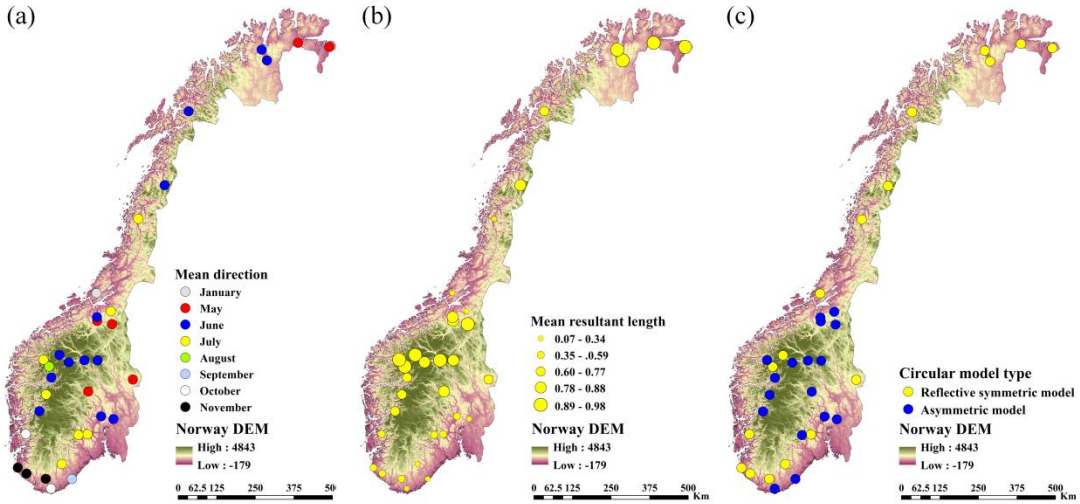

Fig. 5. Results of exploratory analysis and statistical inference of circular data for Norway. Maps of the sample mean direction (a), sample mean resultant length (b) and identified circular model types (c).

Figure

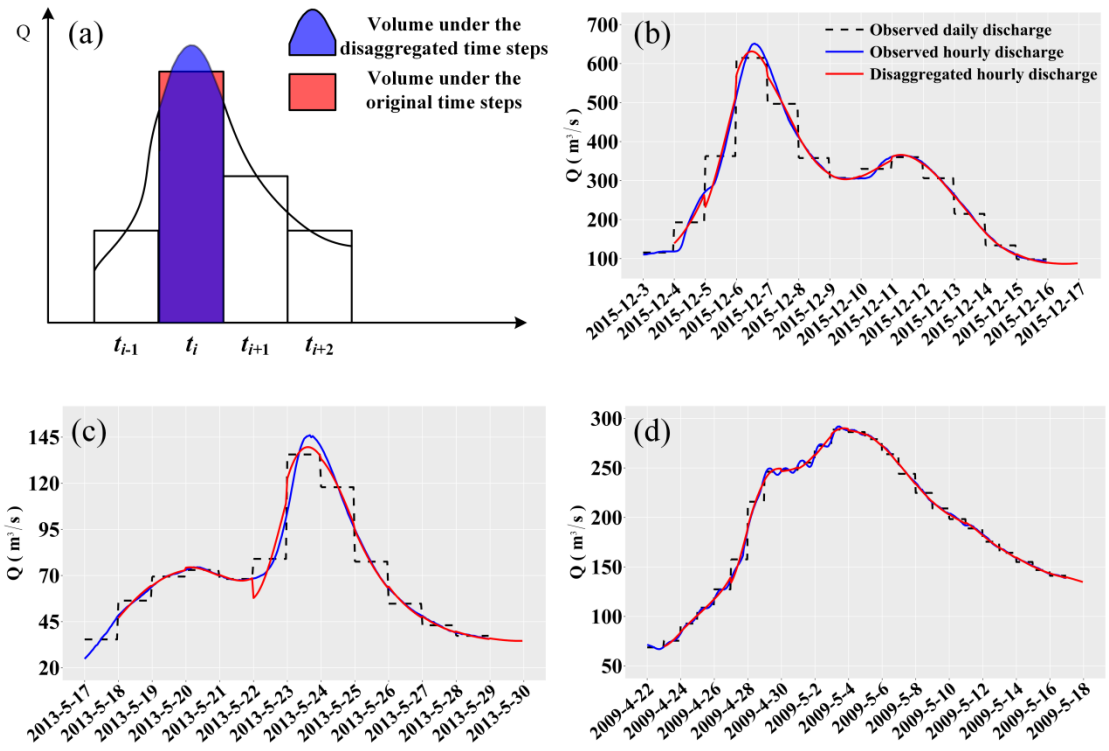

Fig. 6. Schematic diagram of the water volume at different original and disaggregated time steps (a), and illustrations of the disaggregation results of three flood events at the Kjæøemo gauge station (station ID: 22.4) (b), the Atnasjø gauge station (station ID: 2.32) (c), and the Nybergsund gauge station (station ID: 311.6) (d).

Figure

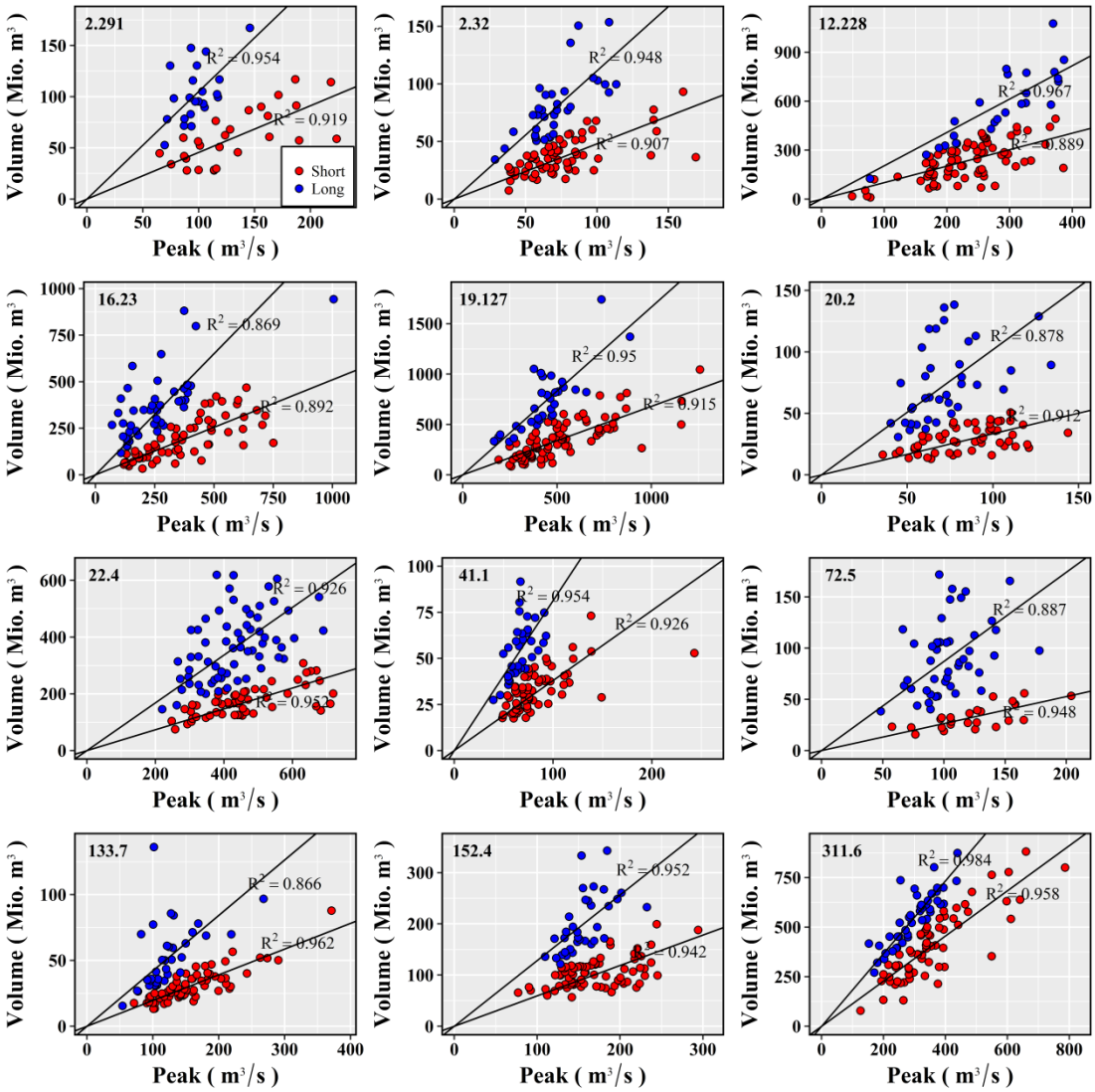

Fig. 7. Classification of the annual maximum flood events for the selected stations in Norway into either long-duration floods (blue points) or short-duration floods (red points) based on flood timescale.

Figure

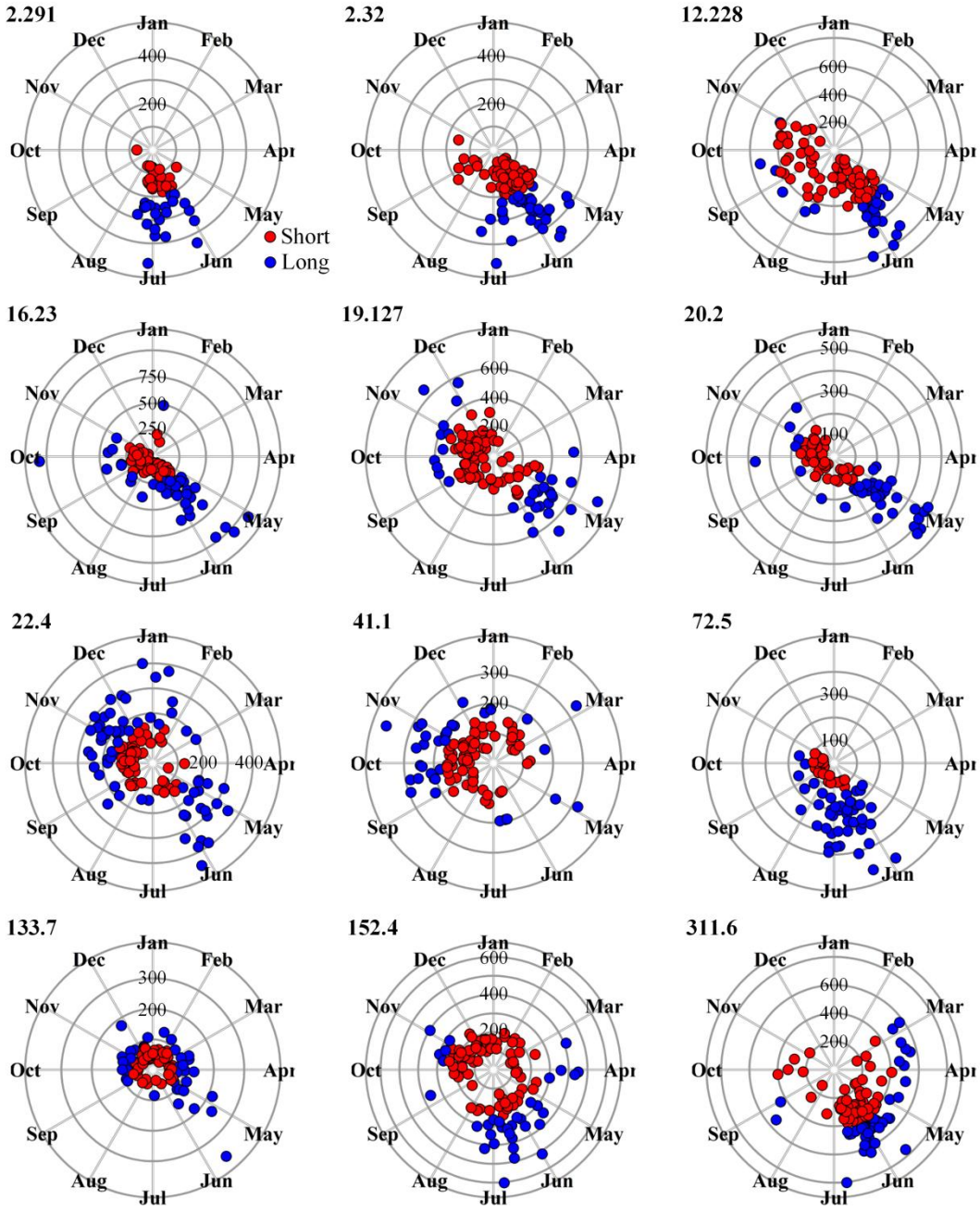

Fig. 8. Burn diagram of FT values for the selected 12 stations. The red filled circles represent the short-duration floods; the blue filled circles represent the long-duration floods; the radial distance from the origin indicates FT value.

Figure

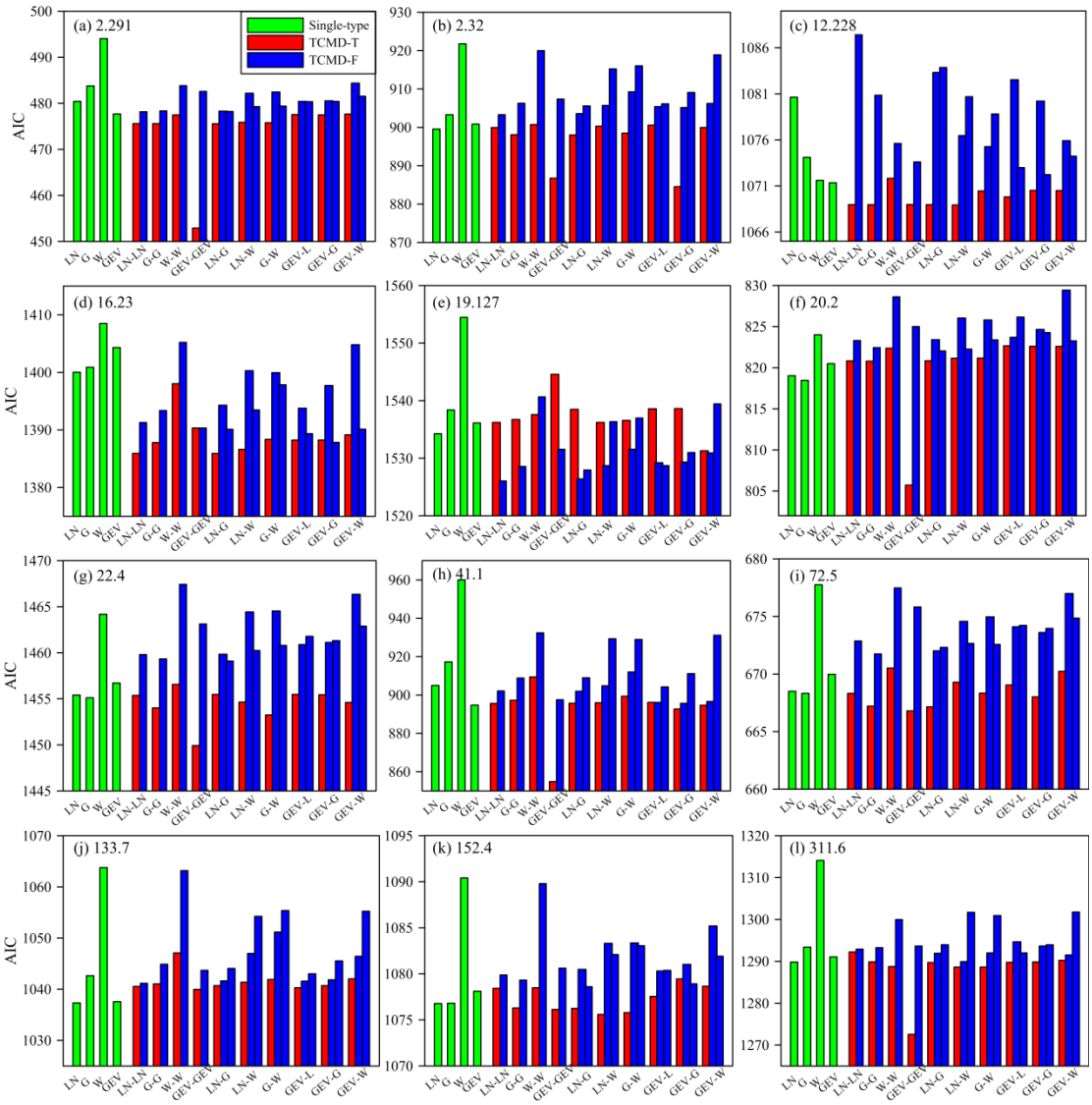

Fig. 9. AIC values of the employed single-type distributions, TCMD-T and TCMD-F for flood series of the selected 12 stations. Note that there are two different heterogeneous mixture distributions when using TCMD-F.

Figure

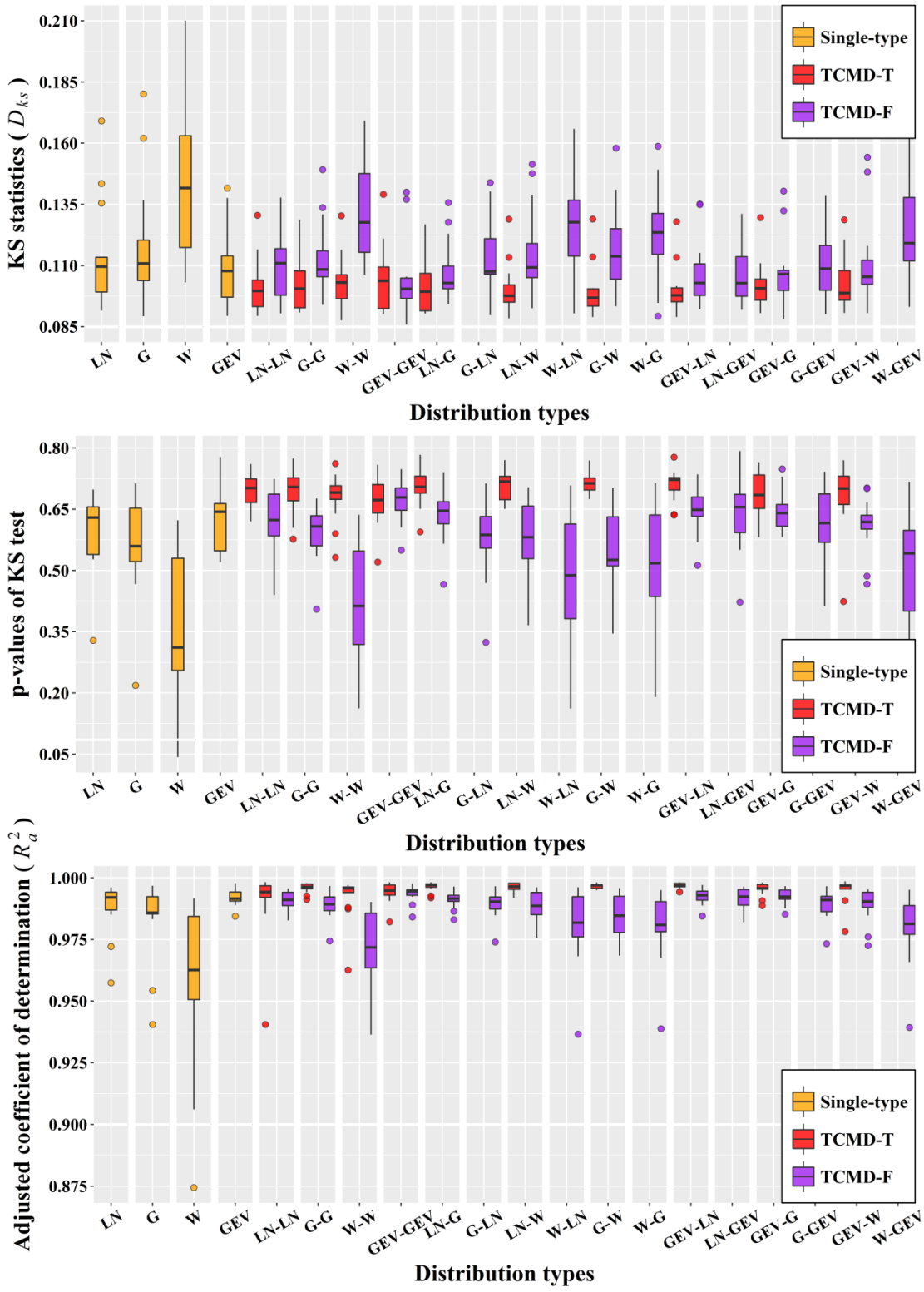

Fig. 10. Boxplots of the D_{ks} statistics (top panel), associated p-values (middle panel) and R_a^2 statistics (bottom panel) for all the single-type and mixture distributions.

Figure

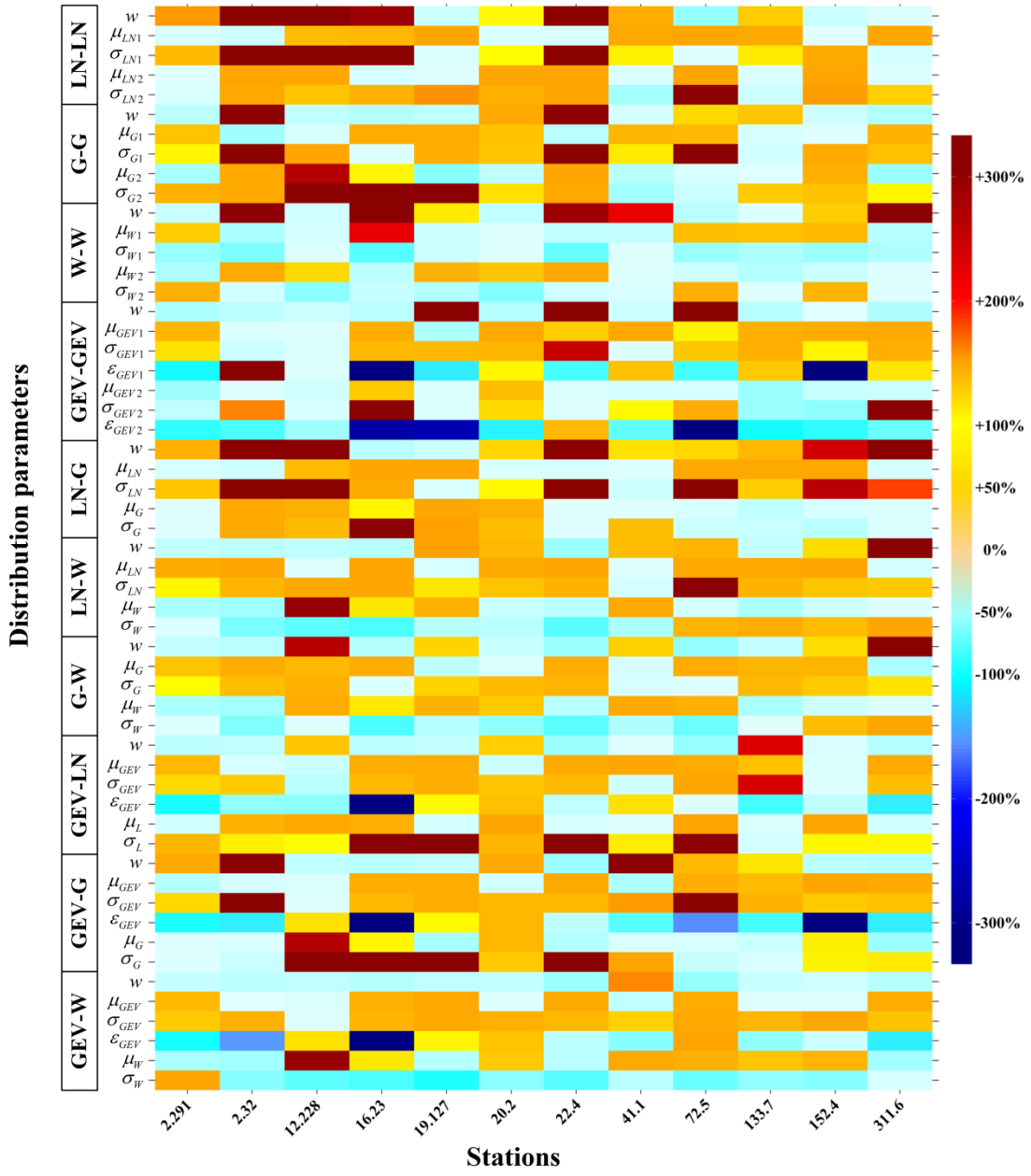

Fig. 11. Heatmap of the relative change between each distribution parameter of TCMD-F and TCMD-T calculated by using parameter value of TCMD-F minus that of TCMD-T, and then divided by that of TCMD-T.

Figure

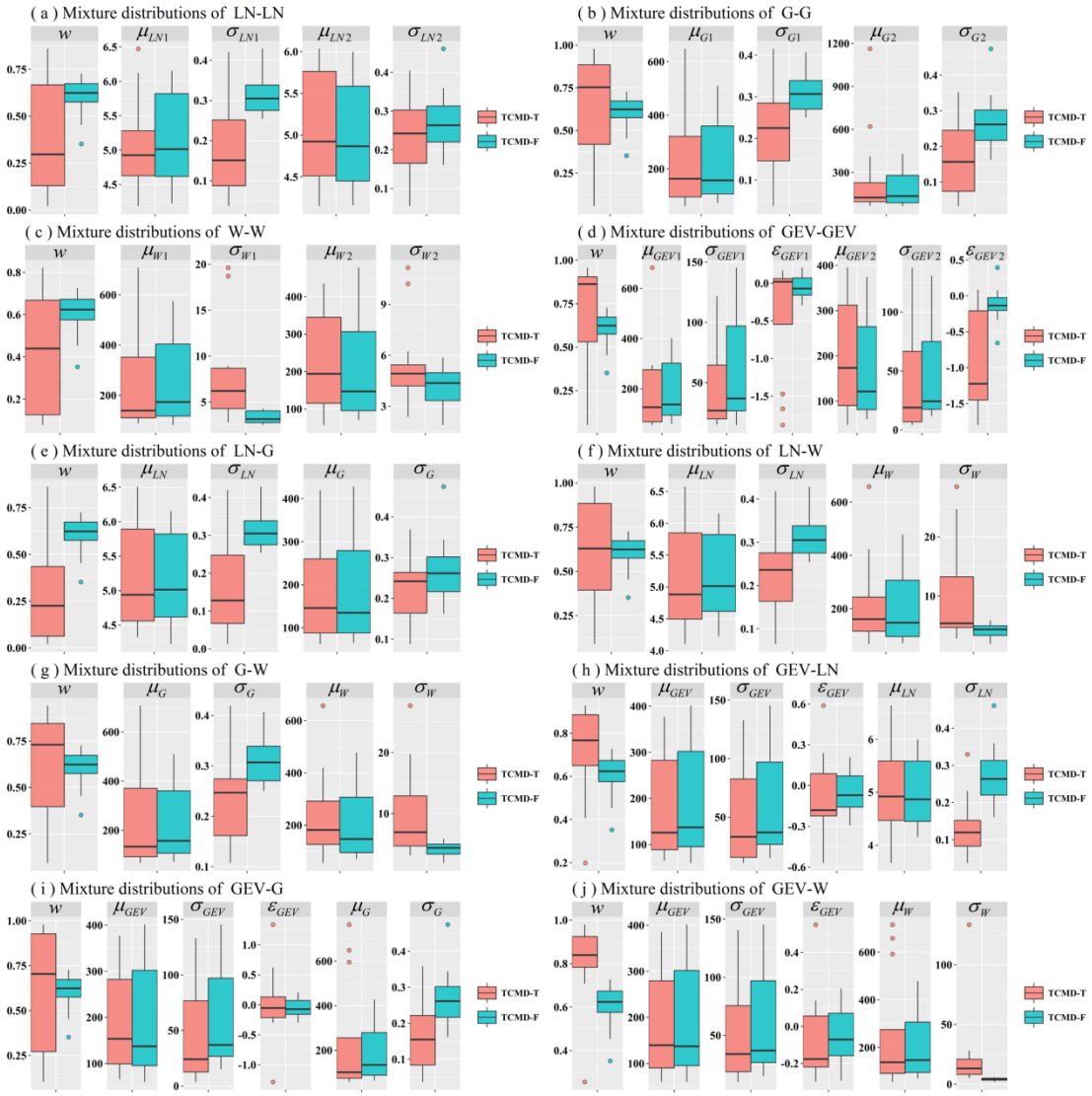

Fig. 12. Boxplots of the parameters for all the mixture distributions.

Figure

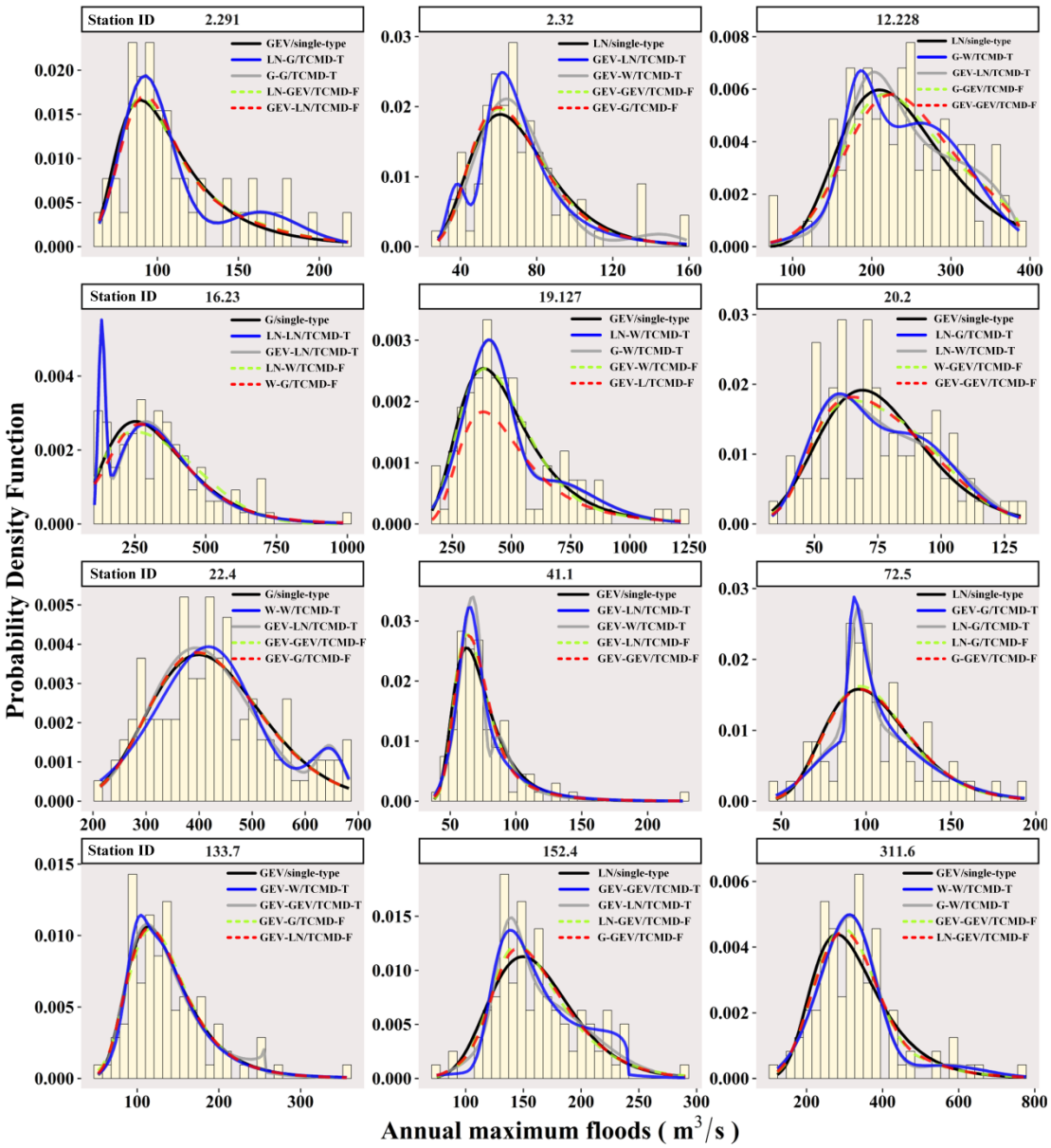

Fig. 13. Empirical frequencies and theoretical probability density function of the optimal single-type distribution, the top two ranked TCMD-T models, and the top two ranked TCMD-F models based on TOPSIS with respect to the three goodness-of-fit measures for each station.

Figure

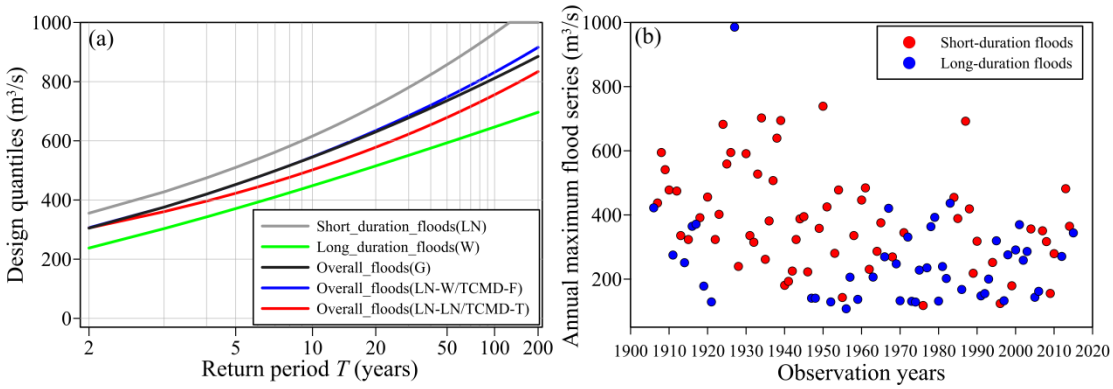

Fig. 14. Return level diagrams for the overall AMFS using optimal single-type, TCMD-T and TCMD-F models together with the return levels for the short-duration floods and long-duration floods at Kirkevollbru station (ID: 16.23) (a); The classified short-duration floods (red circle points) and long-duration floods (blue circle points) over the period of 1906-2015 at the Kirkevollbru station (b).

Figure

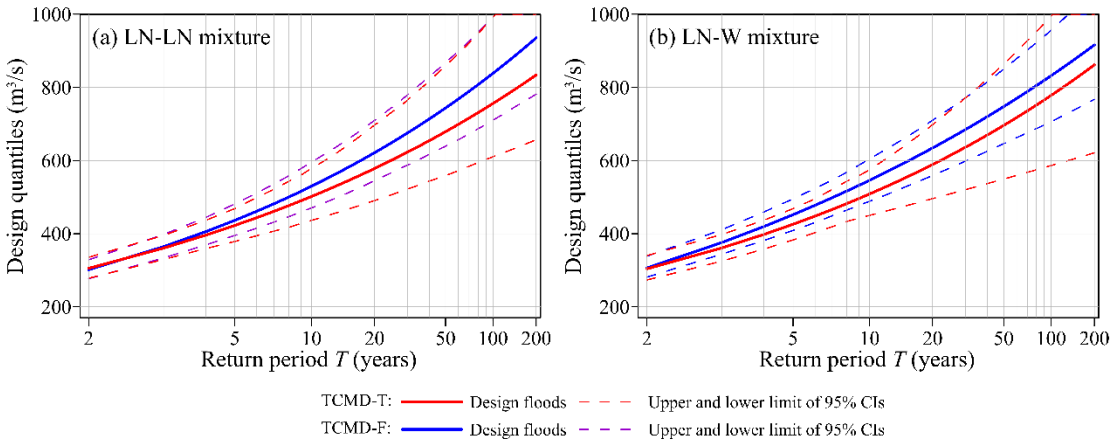

Fig. 15. Return level diagrams for the AMFS of Kirkevollbru station (ID: 16.23) estimated by (a) LN-LN mixture (the optimal TCMD-T model) and (b) LN-W (the optimal TCMD-F model) with 95% bootstrapped confidence intervals. The solid lines are the design floods while the dashed lines are the upper and lower limits of the 95% confidence intervals.