Journal of Hydrology: Regional Studies 28 (2020) 100675

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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

# Uncertainty introduced by flood frequency analysis in projections for changes in flood magnitudes under a future climate in Norway

Deborah Lawrence

*Hydrological Modelling Section, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway*

ARTICLE INFO

*Keywords:*
Climate change impacts
Nordic
Extreme value analysis
Ensemble uncertainty
HBV hydrological model

ABSTRACT

*Study region:* The study considers 115 unregulated catchments in Norway, with areas from 6 to 15,449 km$^2$ and flood generation regimes ranging from snowmelt-dominated to 'mixed' (snowmelt and rainfall) to fully rainfall-driven.

*Study focus:* Bias-corrected EURO-CORDEX RCM output for RCP 8.5 is used to generate an ensemble of 500 hydrological simulations for assessing changes in flood magnitudes under a future climate. Flood estimates are based on three extreme value distributions (EVDs), Gumbel, Generalised Extreme Value and Generalised Pareto, with confidence intervals calculated using parametric bootstrapping, and uncertainty introduced into the ensemble by the flood estimation is evaluated using variance decomposition. Changes in EVD parameters under future conditions are also assessed.

*New hydrologic insight for the region:* There are large differences in projected changes between catchments, with median estimates ranging from −48 % to +99 % for the 200 year flood. Flood magnitudes in all catchments with rainfall-dominant or mixed flood regimes are expected to increase. EVDs with a shape parameter (GEV and GPD) indicate larger increases in higher flood quantiles than the Gumbel distribution (*e.g.* by 5–8 percentage points for the 200 year flood). Flood frequency estimation contributes 30–52 % of the total ensemble range in individual catchments. Location and scale parameters generally increase in catchments with increasing flood magnitudes, and some catchments with mixed flood regimes also exhibit increases in the shape parameter under future conditions.

## 1. Introduction

There is considerable interest in and a pressing need for assessments of likely changes in extreme flows under a future climate due to the potential impact of extreme flooding on critical infrastructure and on society (*e.g.* Hirabayashi et al., 2013; Winsemius et al., 2016). As both housing and structures such as flood protection measures and hydropower-related dams and reservoirs can have expected lifetimes of up to 100 years and the consequences of inundation or failure can be catastrophic, estimates for high flows with long return periods are required for design purposes and for land use planning. This task is a significant challenge even under the current climate, as extrapolation beyond the length of a local streamflow record is almost always required. In addition, the extreme value analyses underlying this extrapolation entail fairly stringent assumptions as to the statistical properties of the observed data series, *i.e.* that extreme events are independent, identically distributed and stationary, that are often not satisfied in practise (Rossi et al., 1984; Franks and Kuczera, 2002; Read and Vogel, 2015). For estimates under a future climate, observed time series are not available and estimates are often generated from simulated hydrological data derived from climate projections. This leads to an even

*E-mail address:* dela@nve.no.

https://doi.org/10.1016/j.ejrh.2020.100675
Received 2 December 2019; Received in revised form 20 February 2020; Accepted 20 February 2020
Available online 03 March 2020

higher degree of uncertainty in flood estimates as several additional factors come into play, including the reliability of downscaled climate projections (Maraun et al., 2010), the need for and suitability of local bias adjustment of input climate data (*e.g.* Leander et al., 2008; Teutschbein et al., 2011), the suitability of the hydrological model (*e.g.* Poulin et al., 2011; Bosshard et al., 2013; Krysanova et al., 2018) and its parameterisation (*e.g.* Vormoor et al., 2018; Dakhlaoui et al., 2019) for modelling streamflows under changing conditions, and the most likely greenhouse gas concentration pathway. The general aim of the study presented here is, therefore, to investigate the uncertainty introduced by standard methods for statistical flood frequency estimation when applied to an ensemble of hydrological projections. A comparison is made between estimates based on extreme value distributions with and without a shape parameter for describing the skewness of the distribution. The study also presents projections for likely changes in flood magnitudes under a future climate in a region in which both snowmelt and rainfall contribute significantly to extreme high flows and in which the relative contributions of these factors are expected to change in a warmer climate.

A range of previous studies have analysed likely future changes in flooding using ensembles of hydrological simulations driven by precipitation and temperature time series from climate model outputs at catchment, regional and global scales (*e.g.* Hirabayashi et al., 2013; Köplin et al., 2014; Vormoor et al., 2015; Alfieri et al., 2015; Arheimer and Lindström, 2015; Arnell and Gosling, 2016; Osuch et al., 2016; Meresa and Romanowicz, 2017; Thober et al., 2018; and for earlier examples, see review in Madsen et al., 2014). Three issues considered or arising from these published studies are particularly relevant for the work presented here: 1) expected changes in regions where melt processes contribute to high flows; 2) the target variables used to assess changes in high flows and methods for estimating these; and 3) the components of the modelling chain considered in the ensemble analysis, and these aspects are reviewed in the following paragraphs.

Regions and catchments in which snowmelt clearly dominates flood generation in today's climate can, in many cases, expect both a decrease in flood magnitudes and an earlier spring flood (*e.g.* Veijalainen et al., 2010; Poulin et al., 2011). An earlier spring flood is, in fact, one of the few statistically significant changes in flood regimes that already is observed in the present climate (*e.g.* Wilson et al., 2010; Arheimer and Lindström, 2015; Vormoor et al., 2016; Blöschl et al., 2017). Expected future patterns of change are, however, more complex for catchments in which both snowmelt and rainfall contribute to flood generation, as the competing trends of a decreasing winter snowpack *vs.* increases in precipitation that falls as rain become important and can in some cases lead to a change in flood seasonality and in flood generating processes (Köplin et al., 2014; Vormoor et al., 2015). Projected changes in flood magnitudes and even the sign of the projected change can vary significantly with local factors, such as catchment size and median elevation, particularly in regions with complex topography (*e.g.* Musselman et al., 2018). There is, therefore, little direct correspondence between changes in precipitation and in flood magnitudes (Sharma et al., 2018), and in some cases, projected percentage changes in flood magnitudes under a future climate are higher than the expected increases in precipitation extremes.

Previous studies have used a variety of measures to assess changes in high flows, typically applying the measure to two or more time slices (*e.g.* 30 year periods) representing simulations for present and future conditions and comparing the results to estimate the degree of change (*e.g.* as a change factor or percentage change). Time slices of 30 years or less, rather than the full length of the simulated series (*e.g.* 1950–2100), are often used to avoid problems of non-stationarity (Madsen et al., 2014). Widely applied indices include both quantities that can be estimated without extrapolation, such as the mean annual flood (*e.g.* Lawrence and Haddeland, 2011; Köplin et al., 2014; Thober et al., 2018), average magnitude or number of over-threshold values (*e.g.* Cloke et al., 2013; Alfieri et al., 2015; Hundecha et al., 2016; Vormoor et al., 2016), or empirical return levels based on the time slice considered (*e.g.* Steele-Dunne et al., 2008; Hundecha et al., 2016) as well as estimates based on a fitted extreme value distribution (EVD) for flood magnitudes of a given return period. Although extrapolation with an EVD invariably leads to a more uncertain estimate, knowledge of likely changes in these higher flood quantiles are often required in practise for climate change adaptation. Several earlier studies of climate change impacts on flooding (*e.g.* Dankers and Feyen, 2008; Veijalainen et al., 2010; Lawrence and Hisdal, 2011) have applied a two-parameter Gumbel distribution to the simulated annual maximum series, and in some cases, have justified its use relative to a 3-parameter GEV (Generalised Extreme Value) distribution based on the likelihood-ratio (*e.g.* Dankers and Feyen, 2008). The use of a GEV distribution is more common in recent studies (*e.g.* Hattermann et al., 2011; Condon et al., 2015; Meresa and Romanowicz, 2017; Soriano et al., 2019), although the Gumbel distribution is still frequently used (*e.g.* Hirabayashi et al., 2013; Alfieri et al., 2015). The choice of a particular extreme value distribution relative to other alternatives is generally not discussed in these studies.

The selection of 'the best' methods for estimation of flood quantiles using extreme value analyses is a well-established and advanced research topic and covers a range of issues, including appropriate statistics for evaluating reliability and stability or robustness (*e.g.* Renard et al., 2013; Kochanek et al., 2014) and goodness-of-fit (*e.g.* Laio, 2004), as well as the most suitable methods for fitting the extreme value distributions. Some of these issues have recently been investigated for 526 Norwegian catchments for local (cf. regional) analyses using block maximum methods applied to observed data (Kobierska et al., 2018). Consistent with other studies, the results indicate that the 3-parameter GEV gives slightly more reliable, but less stable, estimates than the 2-parameter Gumbel distribution for 30 year data records. The results also favour the use of L-moments or Bayesian methods for fitting the EVD. For goodness of fit, both the Anderson-Darling test and the Cramer von Mises test are preferred for testing extreme value distributions (Laio, 2004), relative to traditional tests such as the Kolmogorov-Smirnov test, due to the weighting they place on higher quantiles.

Alternatives to the block maxima approach, such as the fitting of over-threshold values to a Generalised Pareto (GP) distribution, are less commonly used in studies of climate change impacts on flooding, although they also have been applied (*e.g.* Prudhomme et al., 2003, 2013; Collet et al., 2017). A recent study of flooding (Buchanan et al., 2018) due to sea level rise has also investigated this approach and demonstrates its advantages relative to the fitting of a Gumbel distribution. For streamflow flooding, a clear advantage of a peak over threshold series over an annual maximum series is that one potentially has a larger sample size and that the sample represents all of the highest flows during a given period, including subannual maxima (Madsen et al., 1997; Lang et al., 1999). This is relevant for Norwegian catchments, as many catchments exhibit 'mixed' flood regimes, including both snowmelt-generated

floods during spring and early summer and rainfall-generated floods during the autumn under the current climate (Vormoor et al., 2016). In some cases, the largest observed floods have occurred in response to rainfall in the autumn, although the annual maximum series is dominated by snowmelt floods during the spring and early summer (Roald, 2013). As the balance between various types of flood generating processes can change gradually under a warming climate, the use of a method that is not limited to sampling one event per year is more likely to detect insipient trends. A principal drawback of peak-over-threshold EVDs is that a suitable threshold must be used to select the events comprising the series and ensuring that they in fact represent an extreme value series. This is traditionally done using, for example, mean residual life plots (see discussion in Coles, 2001). This approach is, however, unfeasible for a large ensemble of simulations derived from climate data, such as is considered here. An alternative is to use an assumed threshold that is high enough to ensure that the sampled events are, in fact, extreme values (*e.g.* Madsen et al., 1997) such that the peak-over-threshold approach can be applied to a large number of data series without the need for an analysis of each individual data series.

Different climate models, downscaling methods and hydrological modelling techniques can produce widely varying estimates, and it is not possible to select the combination of models and approaches that give the most reliable prognoses for extreme flows under unknown conditions towards the end of the 21st century. A popular remedy for this is to use a so-called 'ensemble' method in which combinations of several models and methods are applied to generate multiple realizations of current and future conditions. Although the resulting 'ensembles of opportunity' (von Storch and Zwiers, 2013) lack the statistical rigor of 'classic' ensemble methods (in which, for example, multiple realisations are generated based on perturbed initial conditions), they are nevertheless useful for summarizing a range of possible future outcomes based on available methods and models. Numerous studies have investigated the 'uncertainty' in such ensemble estimates (corresponding in most cases to the ensemble range or 'variance') due to various ensemble components including climate models, post-processing methods and hydrological models and their parameterisations. The various analyses indicate that although uncertainty from differences between climate models makes a large contribution to the ensemble spread, other factors such as statistical post-processing methods (Hundecha et al., 2016; Osuch et al., 2016), hydrological model structure (Poulin et al., 2011; Bosshard et al., 2013; Steinschneider et al., 2015) and hydrological model parameterization (Lawrence and Haddeland, 2011) can also make significant contributions in some cases. Uncertainty introduced by the methods used to estimate extreme flood quantiles has received very little attention in previous work. Exceptions to this are the recent study by Meresa and Romanowicz (2017) in which the uncertainty introduced by flood frequency estimation based on a GEV distribution in estimates for future changes in both high flows and low flows is evaluated for a single catchment in Poland. They conclude that the flood frequency analysis contributes a considerable degree of uncertainty in the estimation of low flows, but not in high flows, in that catchment. In contrast with this finding, a recent study by Collet et al. (2017) found that the use of a GP or GEV distribution represents 40–60 % of the total ensemble uncertainty (relative to climate models) in estimates for changes in the 100 year flood in Great Britain.

To build on previous findings, the objectives of this study are, therefore, to compare estimates for future changes in hydrological flooding in Norway based on the application of 2-parameter (Gumbel) *vs*. 3-parameter (GEV and GP) extreme value distributions to a large ensemble of simulations. In particular, the study considers the following research questions:

1) Do EVDs with a shape parameter (GEV and GP) give significantly different estimates for projected changes in higher flood quantiles?
2) Do EVDs with a shape parameter give a better fit to the extreme value series?
3) What is the average contribution of flood frequency estimation to the range of ensemble estimates for changes in higher flood quantiles?
4) How do the average EVD parameters change under a future climate and can this change be related to flood generating processes?

The study also presents results for projected changes in four flood quantiles (with 10, 100, 200 and 1000 year return periods) for 115 catchments representing a range of flood generation regimes from fully pluvial to predominantly nival. The 200 year flood is used for flood hazard mapping in Norway in conjunction with land-use planning, and the 1000 year flood is required for design flood analysis for dam safety investigations. Estimates for projected changes in these flood quantiles are therefore used in practice for work with climate change adaptation.

## 2. Study area

Catchment-based hydrological modelling is applied in this study to develop hydrological projections for 115 near-natural catchments distributed across Norway, all without significant streamflow regulation affecting high flows (Fleig et al., 2013). The catchments represent a range of climate and hydrological regimes, topographic conditions and landscape types and can be grouped into six regions for summarising and presenting results (Fig. 1 and Table 1). The boundaries between the six regions are based primarily on established runoff regions (*e.g.* Pettersson, 2013) with a further grouping or subdivision of some regions to give a closer correspondence with administrative units (*i.e.* county boundaries). In all cases, however, the regions used here are based on catchment divides as boundaries between regions, rather than county boundaries. The adjustments to the runoff regions have been made to facilitate the presentation of summary results for regions that can be further used in planning related to climate change adaptation (*e.g.* Hanssen-Bauer et al., 2017).

The mean annual temperature for the whole of Norway during the reference period used in this study, 1971–2000, was +1.3°C (Hanssen-Bauer et al., 2015, 2017). There are, however, large regional variations, with annual averages of over +6°C in coastal

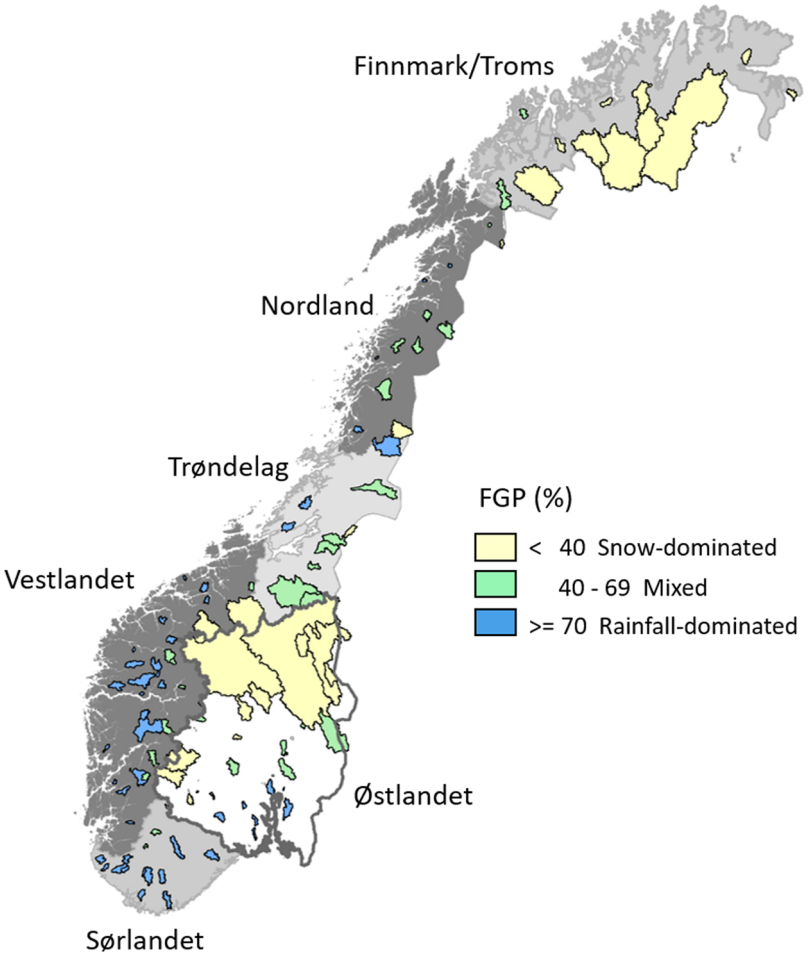


**Fig. 1.** Distribution and areal extent of the 115 catchments used in this study. The shaded areas indicate the 6 regions described in Table 1, where the region 'Østlandet' is without shading and its boundaries are indicated by the thick grey line. The fill colours indicate the flood generating process (FGP) under the current climate estimated as the percentage of the rainfall-dominated floods in the peak over threshold series for the catchment.

regions of Vestlandet, Sørlandet and Østlandet, and as low as −2.8°C in Finnmark and −4°C in the high mountain region of southern Norway (located near the boundary between Vestlandet and Østlandet in Fig. 1). Annual precipitation also shows significant spatial variation, with the driest regions in the inland zones of Østlandet and Finnmark receiving as little as 300 mm/yr and the wettest regions in Vestlandet and Nordland over 4000 mm/yr. Spatial patterns of annual runoff largely reflect regional temperature and precipitation patterns. The average annual runoff in the 115 catchments used in this study, together with the range of values for the catchments, are given for each region in Table 1. Catchments with the highest annual runoff are found in Vestlandet (avg. 2309 mm/yr) and in Nordland (avg. 1890 mm/yr) and the lowest annual runoff is associated with catchments in Finnmark (avg. 773 mm/yr.) and Østlandet (avg. 687 mm/yr.). There is, however, considerable local variation in each of these regions.

The seasonal pattern of discharge leading to flooding is strongly controlled by both temperature and precipitation during the winter period, as well as the catchment median elevation and hypsometry as they determine the potential for snow accumulation and melting. The number of days during the winter half year (1st October–31st March) with a temperature over 0°C, the precipitation during the winter half year, and the catchment median elevation and steepness are given as average values and ranges for each region in Table 1. In Finnmark/Troms and in inland Østlandet, winter conditions are favourable for significant snow storage, and snow accumulation is only limited by relatively low precipitation values. This leads to a seasonal pattern of runoff with high flows during the late spring to summer months and low flows during the winter. In contrast, Vestlandet and Sørlandet, and to some extent coastal regions of Nordland, Trøndelag and Østlandet, have relatively warmer winters with abundant precipitation. These areas, nevertheless, can also experience colder periods with snow accumulation. In addition, catchments in some of these areas can be quite steep, with snow storage and even glaciation, in the upper reaches. These factors produce a general pattern of high flows during the autumn and winter months with the heaviest precipitation, although peak melt periods can also lead to elevated discharge values during spring and summer months and during transient periods of snow accumulation and melting during autumn and winter.

Climate change projections for Norway (Hanssen-Bauer et al., 2015, 2017) indicate changes in both temperature and precipitation under a future climate. For the whole of Norway, temperature increases of 1.7–3.7°C under RCP 4.5 (moderate greenhouse gas

**Table 1**
Catchment characteristics by region (average and range).

| Region | No. of catchments | Area (km$^2$) | Average annual runoff (mm/yr) | Winter[a] days with T over 0° (of 182 days) | Avg. winter[a] precipitation (mm/6 mo.) | Median elevation (m.a.s.l.) | Catchment steepness[b] | FGP – Percentage rainfall floods |
|---|---|---|---|---|---|---|---|---|
| Finnmark/Troms | 12 | 2640 (135–14157) | 773 (365–1693) | 27 (15–69) | 546 (211–1131) | 506 (235–947) | 7.8 (0.53–17) | 32 (16–59) |
| Nordland | 12 | 205 (17–653) | 1890 (1044–3611) | 54 (14–96) | 1572 (850–2494) | 604 (157–1022) | 32.8 (7.9–115) | 60 (21–89) |
| Trøndelag | 11 | 737 (142–3079) | 1210 (675–1956) | 56 (20–89) | 906 (619–1542) | 630 (295–1347) | 6.6 (2.9–11) | 59 (33–79) |
| Vestlandet | 33 | 277 (7–2442) | 2309 (897–3758) | 72 (16–143) | 2157 (858–3215) | 942 (55–1546) | 28.2 (4.2–83) | 68 (23–97) |
| Sørlandet | 15 | 139 (19–276) | 1765 (894–2843) | 94 (31–143) | 1758 (884–2924) | 567 (178–1138) | 10.4 (3.6–23) | 84 (51–97) |
| Østlandet | 32 | 1338 (6–15449) | 687 (428–1554) | 42 (12–93) | 673 (394–1345) | 788 (153–1485) | 10.1 (1.4–24) | 50 (18–88) |

[a] Winter half-year, defined as the period 1st October–31st March.
[b] Defined as the difference between the 25th and 75th percentiles of the hypsometric curve divided by the catchment length.

concentration pathway) and 3.4–6.0°C under RCP 8.5 (high concentration pathway) are expected by 2071–2100 relative to a 1971–2000 reference period. Projected changes in annual precipitation are between 3 and 14 % (RCP 4.5) and 7 and 23 % (RCP 8.5). There are nevertheless large variations between regions and between seasons. The largest absolute increases (in mm) in precipitation are expected to occur during the autumn and winter seasons in the regions Vestlandet, Trøndelag and Nordland (see Fig. 1), while the largest relative increases (in %) are expected during the spring and winter periods in Østlandet and Finnmark. In addition, the number of days with heavy precipitation is expected to increase over the entire country during all seasons. In some cases, the average number of days with heavy precipitation will be more than double by the end of the 21st century under RCP 8.5.

The projected changes in temperature and precipitation regimes are expected to lead to changes in seasonal runoff that also vary significantly between regions. Due to a decrease in snow storage and an increase in precipitation that falls as rain in the winter under a future climate, winter runoff is projected to increase in all regions under RCP 8.5, with the largest relative increases in Nordland and Finnmark and in inland and mountainous regions in the mid- and southern Norway (Hanssen-Bauer et al., 2015). Summer runoff is projected to decrease considerably in all regions as a result of increased evapotranspiration and an earlier snowmelt season due to warmer temperatures.

## 3. Data and methods

### 3.1. Classification of flood regimes

Previous analyses of observed and projected future high flows in Norway indicate that the dominant flood generation process (FGP) in a given catchment exerts a strong control on the direction and magnitude of changes (Vormoor et al., 2015, 2016). In this study, we therefore use the simple categorisation of FGPs proposed by Vormoor et al. (2015, 2016) to describe the contribution of precipitation *vs*. snowmelt to the generation of high flows in individual catchments. The FGP represents the percentage of the independent events in the over threshold series (based on a threshold corresponding to the 98.5th percentile of the flow duration curve) comprised predominantly of rainfall (see Vormoor et al., 2016 for further details) and is quantified here as a percentage between 0 and 100 %. Although more detailed classifications could be considered (*e.g.* Merz and Blöschl, 2003), we use the simple 3-member classification proposed by Vormoor to broadly distinguish flood regimes and to characterise the spatial pattern of flood generation. Using this classification, the categories snow-dominated *vs.* mixed *vs.* rainfall-dominated flood regimes shown in Fig. 1 highlight snowmelt as the main FGP in the inland regions of southern Norway and in Finnmark/Troms; and rainfall as the dominant driver of high flows in Vestlandet and Sørlandet and in coastal regions of Østlandet, Trøndelag and Nordland. Mixed regimes are prominent in transitional areas between the two other regimes. Although actual changes in these flood regimes under a future climate are not analysed in this study, the catchment FGP under the current climate is used to identify the types of catchments that are most susceptible to increases *vs*. decreases in flood levels under changing climatic conditions.

### 3.2. Ensemble of hydrological projections

#### 3.2.1. Climate model data and bias adjustment

The hydrological simulations analysed in this study are driven by data from 10 RCM (Regional Climate Model) runs generated by the EUROCORDEX initiative (Jacob et al., 2014) for RCP 8.5 using the EUR-11 grid with a spatial resolution of approximately 12 km (Table 2). RCP 8.5 is used here because the current national guidance for adaptation related to climate change in Norway recommends that estimates for the likely effects of global warming are based on concentration pathways representing a high level of emissions (Miljøverndepartement, 2012). The selection of the GCM-RCM combinations shown in Table 2 represents EUROCORDEX runs available at the time of the most recent Norwegian climate assessment report (*e.g.* summarized in Hanssen-Bauer et al., 2015, 2017) which were deemed suitable for analyses in the Nordic region by the Norwegian Meteorological Institute. Precipitation and temperature series were extracted from the EUROCORDEX runs for 115 catchments distributed across Norway for a reference period (1971–2000) and a future period (2071–2100). Data from the RCM grid cells covering each catchment were used to develop area-averaged values (weighted according to the proportion of the catchment area covered by individual grid cells) for bias adjustment.

**Table 2**
GCM/RCM combinations used in the ensemble analysis.

| Global Climate Model (GCM) | Regional Climate Model (RCM) | Institute |
|---|---|---|
| CNRM-CER-FACS-CM5 | CCLM4-8-17 | CLM-Community |
| CNRM-CER-FACS-CM5 | RCA4 | SMHI |
| ICHEC-EC-EARTH | CCLM4-8-17 | CLM-Community |
| ICHEC-EC-EARTH | HIRHAM5 | DMI |
| ICHEC-EC-EARTH | RACMO22E | KNMI |
| ICHEC-EC-EARTH | RCA4 | SMHI |
| IPSL-CM5A-MR | RCA4 | SMHI |
| MOHC-HADGEM2-ES | RCA4 | SMHI |
| MPI-ESM-LR | CCLM4-8-17 | CLM-Community |
| MPI-ESM-LR | RCA4 | SMHI |

Precipitation time series from the RCMs were bias adjusted using two alternative techniques: a) empirical quantile mapping (Gudmundsson et al., 2012); and b) a distribution-based mapping using a double gamma function (Yang et al., 2010). These two particular techniques were selected as they can give different corrections for the highest precipitation quantiles. Empirical quantile mapping extrapolates values beyond the highest observed values using a tri-cubic spline and is therefore sensitive to the highest values. This can, however, also undermine the robustness of the correction. Double gamma correction employs separate gamma functions for values above and below the 95th percentile of the cumulative distribution and assumes that the extreme values follow a thgamma distribution in each of the two segments. In practice, the use of a theoretical distribution, such as a gamma function, leads to more robust estimates of the highest quantiles, although there can be deviations from the highest observed values. Temperature time series were adjusted using a normal distribution, as this has previously been shown to be suitable for temperature distributions (Piani et al., 2010).

The data were bias adjusted relative to 'observed' area-averaged values for each catchment derived from 'seNorge' gridded precipitation and temperature data available at a $1 \times 1$ km scale from the Norwegian Meteorological Institute (see discussion in Lussana et al., 2019). Daily values were adjusted using corrections for individual months, developed by applying a 3-month moving window centered on the month of interest (see Wong et al., 2016 for further details). Corrections were developed from and applied to residual values after trend removal following the procedures recommended by Hempel et al. (2013). This approach is used to ensure that the climate change signal is not altered significantly by the bias adjustment process.

#### 3.2.2. Hydrological modelling

The bias adjusted time series were used as input to the HBV hydrological model (Sælthun, 1996), previously calibrated using observed precipitation and temperature series for each of the 115 catchments. The observed time series used in the calibration are the same as those used for the bias correction of climate data discussed in the previous section. Calibration was performed using PEST optimization routines (Lawrence et al., 2009) to produce 150 possible parameter sets for each catchment with a Nash-Sutcliffe efficiency (NSE) criterion of at least 0.50. The 25 best-fit parameter sets were then selected for each catchment such that the NSE was within 2 percentage points of the maximum for the catchment. Both the NSE criterion and the volumetric bias were used as objective functions in the model calibration, and all daily values were given equal weight (*i.e.* no weighting of peak flow values was used). Simulations used a daily time step, and the average validation NSE value for all 115 catchments was 0.76, and ranged from 0.54 to 0.93. In general, better model fits are achieved in catchments with seasonal flow regimes dominated by snowmelt (*i.e.* Østlandet and Finnmark/Troms regions in Fig. 1) relative to those obtained in areas dominated by rainfall. This difference in model performance in catchments with differing flood regimes appears to primarily reflect the higher variance in the streamflow series in rainfall-dominated catchments as compared with the relatively smooth annual pattern of flow found in snowmelt-dominated catchments. The 25 best-fit parameter sets selected from the model calibration and validation are used to assess some of the uncertainty introduced by the parameterization of the HBV hydrological model to the total ensemble considered in this study. A full assessment of the uncertainty introduced by model parameterization would, however, require a significantly larger number of parameter sets (*e.g.* Lawrence and Haddeland, 2011) but is not the focus of the work presented here.

### 3.3. Flood frequency analysis methods

Previously published catchment-based hydrological projections for future changes in flood frequency in Norway (*e.g.* Lawrence and Hisdal, 2011; Lawrence, 2016) apply a 2-parameter Gumbel distribution for estimating return levels for the 200 year flood in a reference and a future period. The Gumbel distribution is used primarily due to its relative robustness for extreme value series of limited length. This ensures that small differences in subsequent analyses based on newer climate projections do not lead to unduly large changes in projected estimates. Changes in the tail of the distribution are, however, also of interest in climate change studies, and such changes are better captured using an extreme value function which also accounts for the skewness of the distribution. Estimates for the projected changes in flood quantiles developed using a 2-parameter Gumbel distribution and a 3-parameter GEV distribution based on block maxima (*i.e.* the annual maximum series) are, therefore, compared in this study. In addition, a GP distribution is applied to the over threshold series representing independent high flow events. This approach has the advantage of a potentially larger sample of events for fitting the distribution. In that case, a fixed threshold corresponding to the 98.5 percentile of the partial duration series is used to select events. Events were assumed to be independent if they were separated by at least 6 days, a value which has previously been shown to be suitable for the 115 catchments considered here (with the exception of two large catchments with an area $> 10{,}000$ km$^2$ and having a predominantly snowmelt flood regime). This procedure gave on average 1.5–2 events per year for most of the catchments considered.

In order to compare the goodness-of-fit of the three alternative EVDs considered, the Anderson-Darling (Laio, 2004) test has been used. This test was implemented using the gnFit R package (Saeb, 2018) that includes extensions of the Anderson-Darling (A–D) test (originally developed for testing of normality) to extreme value distributions. The Cramer von Mises test was also applied, but as this gave very similar results to the A–D test, those results are not reported here. In principle, both the reliability and robustness of the flood estimates based on the three EVDs could also be evaluated for the cases considered here. This has not been undertaken, as it is deemed unlikely that the results would differ from those previously reported, *i.e.* that the 2-parameter Gumbel distribution is more stable, but slightly less reliable than the 3-parameter GEV for record lengths of 30 years (Kobierska et al., 2018).

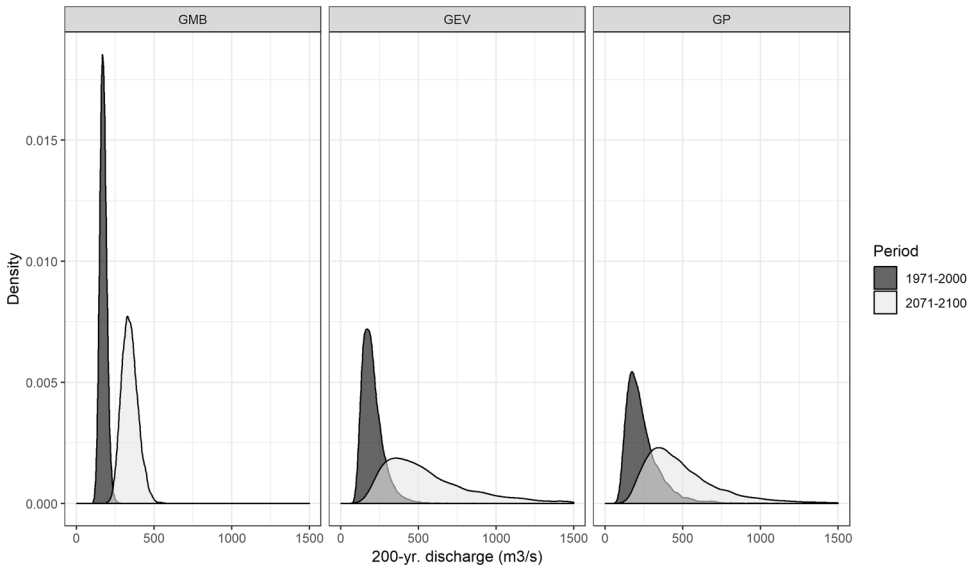


**Fig. 2.** Distribution of return level estimates for the 200 year discharge in a catchment in eastern Norway (Atnasjø, 463 km$^2$) for the reference (1971–2000) and the future (2071–2100) periods based on the Gumbel (GMB), the GEV and the GP extreme value distributions for one (of 500) ensemble member.

### *3.4. Quantification of uncertainty in return level estimates*

The 10 climate projections (Table 2), the 2 bias correction procedures (Sec. 3.2.1), and the 25 HBV parameter sets (Sec. 3.2.2) yield 500 simulations for each of the two time periods considered (1971–2000 and 2071–2100) for each of the 115 catchments. For each simulation, the 10, 100, 200 and 1000 year return levels were estimated based on EVDs fitted with the L-moments method, implemented in R using the 'lmomco' package (Asquith, 2017). Parametric bootstrapping following the procedure described in Kuczera and Franks (2006) was then used to quantify the 5 and 95 % confidence intervals for each return level using 2500 resamples. An example of the results for one ensemble member after bootstrapping is shown in Fig. 2 for the estimated return levels for the 200 year flood for the reference and future periods for the Gumbel, the GEV and the GP distributions. This example illustrates the large spread in the estimates for a given simulation period, particularly for the GEV and GP distributions, and the significant overlap between the reference and future periods for those distributions. To generate estimates for changes in flood levels between the two periods, 100 random samples with replacement were drawn from the distributions for each period and used to estimate the percentage change between the two periods.

### *3.5. Ensemble modelling and variance decomposition*

An ensemble consisting of 50,000 estimates (*i.e.* 500 ensemble members, each with 100 samples) for the percentage change in each of the flood quantiles considered was generated for each of the three extreme value functions for each of the 115 catchments. The contributions of each of three factors (differences between climate model data, hydrological model parameterization, and uncertainty in the extreme flood estimate) to the total spread in the ensemble of results for a given catchment were analysed using variance decomposition (see Déqué et al., 2007 or Sunyer et al., 2015 for full details). The decomposition procedure uses an ANOVA linear model for the ensemble variance and estimates the variance introduced by individual factors and the so-called 'interaction' terms between the factors. The interaction terms arise when the variance introduced by two or more factors cannot be explained using a simple linear combination of the individual factors. In order to produce an ensemble for the analysis of variance in which all components are equally represented (Bosshard et al., 2013), the results for the hydrological parameters sets and for the flood frequency analysis were resampled using a uniform distribution on the 5–95 % range to produce 20 subsamples for each component. Results for the 10 climate models with the 2 bias correction techniques were also combined to produce 20 'samples' representing climate data. The variance decomposition was then performed on the resulting ensemble comprised of 60 members for each catchment, in which the three components 'climate data', 'hydrological model parameterisation', and 'flood frequency analysis' are equally represented, and in which the decomposition is effectively performed on the ensemble range. This is the only decomposition approach that can be justified, given that the full distributions of 'climate data' and 'hydrological parameterisation' are unknown.

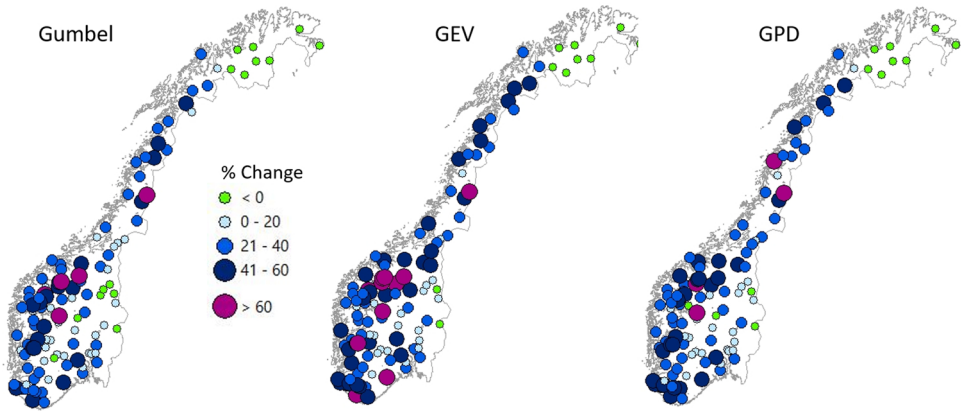


**Fig. 3.** Median value of estimated percentage change in the 200 year flood based on an ensemble of 500 simulations for each of 115 catchments for RCP 8.5. Estimated change is based on a comparison of estimates for the 1971-2000 reference period and the 2071–2100 future period. The flood quantiles for the two periods are estimated using the three extreme value distributions indicated: Gumbel, Generalised Extreme Value (GEV) and Generalised Pareto (GP).

## 4. Results

### 4.1. Projected changes in the flood magnitudes

The median values of the 500-member ensemble of estimated changes in the 200 year flood magnitude are illustrated in Fig. 3 for each of the 115 catchments for the Gumbel, GEV and GP distributions. The estimated median change based on the 2-parameter Gumbel distribution ranges from −48 to +99 %, and the spatial distribution of projected changes indicate a regional pattern of decreases or small increases (*i.e.* < 20 %) in northernmost Norway (Finnmark), in mid-Norway (Trøndelag) and in inland regions of southern Norway. Larger increases are projected for western Norway, Nordland in northern Norway and in some catchments in coastal regions in southern and eastern Norway. Although a somewhat similar regional pattern can be seen in the results based on the GEV and GP distributions, estimated increases are generally larger in most regions and especially along the southwestern coast of Norway. It is only in the northernmost region, *i.e.* Finnmark, that projected decreases or small increases continue to dominate the regional pattern when the GEV and GP distributions are applied. The results also show a high degree of correspondence between the results for the GEV and GP distributions, in that areas with higher estimates for the GEV, such as in coastal southwestern Norway, also indicate higher estimates for the GP distribution.

The projected percentage changes in the 200 year flood magnitude estimated using the three EVDs are shown as a function of the flood generating process (FGP) in Fig. 4. The figure confirms the general tendency for the GEV and GP distributions to give slightly higher estimates for the estimated change seen in Fig. 3, both for catchments with projected increases and decreases. The results also

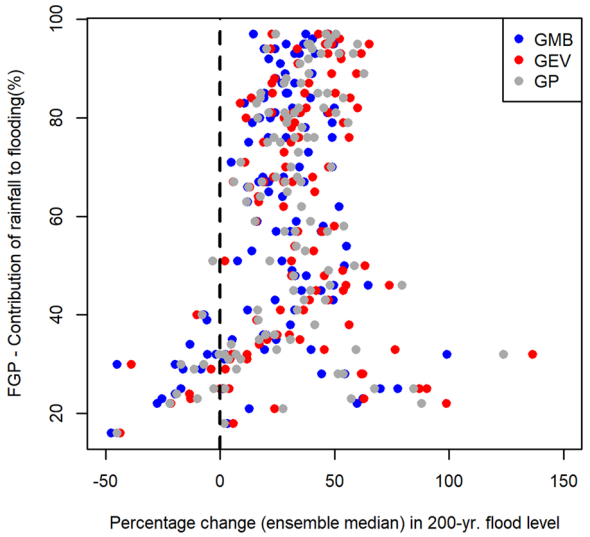


**Fig. 4.** Projected percentage change in the 200 year flood levels relative to the FGP for that catchment for the three EVDs. Each point represents the median value of a 500 member ensemble.

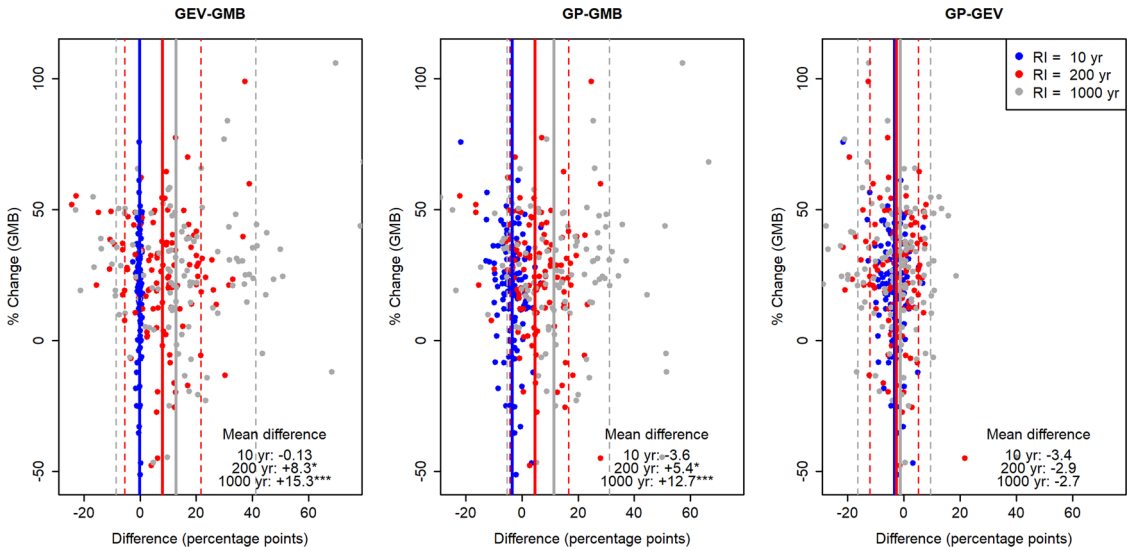


**Fig. 5.** Difference (in percentage points) between the projected changes in the 10, 200, and 1000 year return levels based on a) a GEV *vs*. a Gumbel distribution (GEV-GMB); b) a GP *vs*. a Gumbel distribution (GPD-GMB); and c) a GP *vs*. a Gumbel distribution (GPD-GEV). The solid lines indicate the mean difference for each return level (blue for the 10 year, red for the 200 year and gray for the 1000 year flood level). The stippled red and gray lines indicate the $10^{th}$ and $90^{th}$ percentiles of the distribution of mean differences for the 200 and 1000 year return levels, respectively. The differences in the mean values for each return level are indicated at the bottom of each figure, and the significance level of the Tukey HSD means test is also given (* - 0.05; ** - 0.01; *** - 0.001). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

indicate that virtually all median ensemble estimates are positive for catchments with an FGP of 40 % or higher. It is only in catchments with flood regimes dominated by snowmelt under the current climate that decreases in the 200 year flood are projected by all three EVDs, and in that case, less than half of the catchments are projected to decrease. Fig. 4 also indicates that there are larger differences between catchments with similar FGPs when the FGP is 50 % or less.

Differences in the projected median changes given by the three distributions for each catchment are shown for the 10, 200 and 1000 year return periods in Fig. 5. The differences are plotted relative to the projected median change for the catchment, based on a Gumbel distribution. For each return period, the mean difference of all the values for each case are shown as solid lines and the mean value for that case is given in text at the bottom of the diagram. The 10th and 90th percentiles of the distribution of the mean values are also shown for the 200 year and 1000 year return levels. The values at the bottom of each diagram give the mean difference between each case and its significance level. The results indicate that for a return level of 10 years, there is, on average, no significant difference between the estimates based on the three different EVDs. There is, however, a significance difference between the distribution of values for the 200 year and 1000 year return levels estimated by the EVDs with a shape parameter (GEV and GP) and those estimated using a Gumbel EVD. The EVDs with a shape parameter give, on average, higher estimates. The figure also confirms that the GEV and GP distribution give, on average, similar estimates in that there is only a small difference between the two and this is difference is not statistically significant.

To further compare the three EVD distributions, the Anderson-Darling (A–D) statistic was calculated for each case (*i.e.* ensemble member) for the fits to the series representing the reference and future periods, and the P value indicating the significance level was determined for each case. An average P value was then calculated based on the average of all cases for a given catchment for each of the 3 EVDs. The results indicate that it is only for the GP distribution that one can reject the null hypothesis (*i.e.* that the time series do not come for the distribution indicated) at the 0.05 significance level in most catchments, based on the average P value for that catchment. For the case of the GEV and Gumbel distributions, there were no catchments with average P values indicating that the null hypothesis can be rejected at the 0.05 level, although the average P values based on the GEV distribution tend to be slightly lower (*i.e.* slightly 'better' and closer to the 0.05 level) than those based on the Gumbel distribution.

### *4.2. Uncertainty in ensemble projections for changes in flood magnitudes*

The total ensemble variance decomposed into four components (climate data, flood frequency analysis, hydrological model parameterization, and interaction terms) is shown in Fig. 6a as an average value for all 115 catchments for three return periods (10, 100, and 1000 year) for each of the three EVDs. The component 'climate models' represents the variance introduced by the use of 10 different climate models and 2 different bias adjustment methods to process the climate data before it is used in hydrological modelling. The component 'FFA', *i.e.* flood frequency analysis, is the ensemble variance derived from the fitting of a given EVD. The component 'Hydro model' represents the variance associated with the use of differing parameterisations of the HBV model for individual catchments. The component 'interaction' is the sum of several interaction terms identified in the variance decomposition,

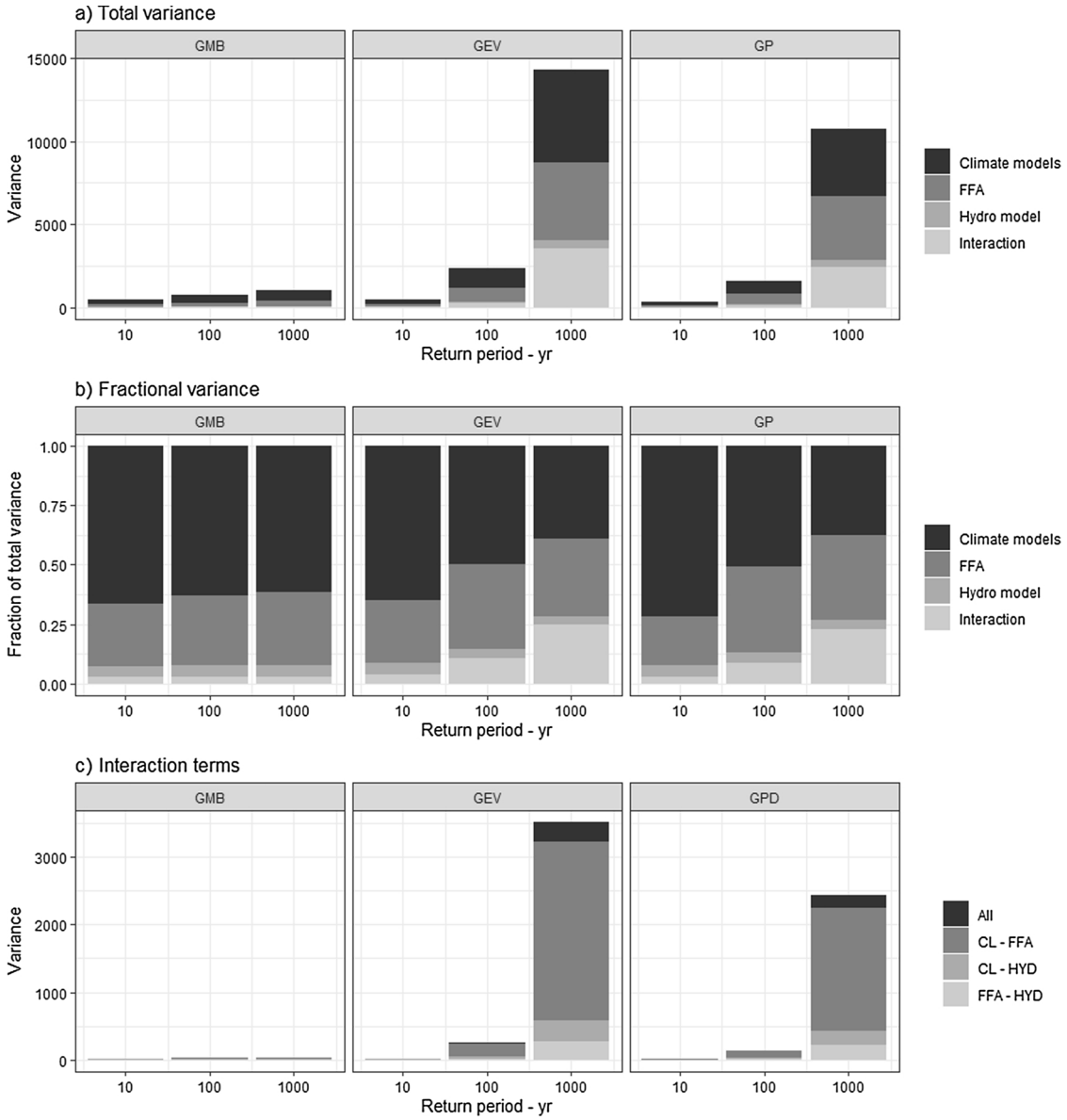


**Fig. 6.** Summary results of variance decomposition: a) average total ensemble variance decomposed into four components (climate model data, flood frequency analysis, hydrological model parameterisation, and interactions) for each EVD (Gumbel, GEV and GP) and return period (10 year, 100 year and 1000 year) indicated; b) average fractional ensemble variance for the four components for each EVD and return period; and c) variance attributed to the interaction term decomposed into its components: All – interaction between all three terms (climate data, flood frequency analysis, and hydrological model parameterisation); CL-FFA – interaction between climate data and flood frequency analysis; CL-HYD – interaction between climate data and hydrological model parameterisation; FFA-HYD – interaction between flood frequency analysis and hydrological model parameterisation.

including both interactions between two individual components (*e.g.* climate models and FFA) and between all components.

The results illustrated in Fig. 6a indicate that for the 10 year return period, the three EVDs give a similar average total variance, although the Gumbel distribution has a slightly higher average value. For the 100 year return level, the total variance is 2–3 times larger for the GEV and GP distributions in comparison with the Gumbel distribution, and for the 1000 year return level, they are an order of magnitude larger. The average total variance is, however, in all cases greater for the GEV than the GP distribution. The contribution of the four components to the total variance indicates that for the case of the GEV and the GP distributions, the contribution of the climate models and the flood frequency analysis is of a similar magnitude and much larger than the contribution of the hydrological model parameterization for all EVDs and return periods. This is further illustrated in Fig. 6b which shows the fraction of the total variance contributed by the four ensemble components. For the case of the Gumbel distribution (GMB), climate data contributes on average approximately two-thirds of the ensemble variance for all return periods. This increases to over 70 % for

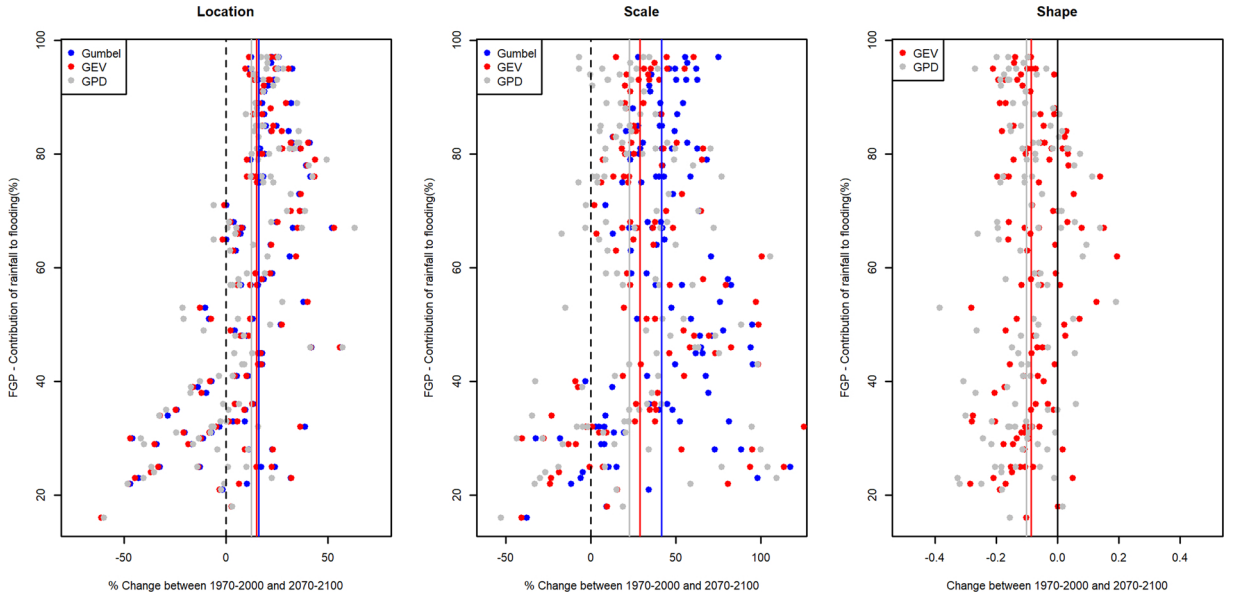


**Fig. 7.** Average change in the location, scale and shape parameters between the reference (1971-2000) and the future (2071-2100) periods for the EVDs indicate. Each point represents the average of 50,000 parameter estimates for a given catchment.

the GP distribution for the 10 year return period. In general, the contribution of the FFA component increases slightly with return period for all EVDs. The largest change, however, is in the increase in the interaction terms for the 100 year and 1000 year return periods for the GEV and GP distributions.

Fig. 6c illustrates the contributions of the individual terms comprising the lumped 'interaction' term shown in Fig. 6b, as a total (rather than fractional) variance. The figure confirms that the interaction terms are negligible when the Gumbel distribution is used relative to the two other EVDS. For the case of the GEV and GPD, the 'climate model – FFA' term comprises over 70 % of the portion of the variance due to the interaction of two or more factors, for both the 100 year and the 1000 year return periods. Thus, the results suggest that for the GEV and GP distributions, the relative contribution of FFA to the ensemble variance is indeed higher for these distributions at the higher quantiles. If one assumes that the interaction between the climate model and FFA terms can be divided equally between the two components, then Fig. 6b and c also indicate that the relative contribution of FFA to the ensemble variance is similar to that of climate models for the GEV and GP distributions for the higher quantiles.

### *4.3. Changes in EVD parameters in a future climate*

The projected changes illustrated in Figs. 3 and 4 suggest systematic changes in the EVD parameters between the reference and future periods. Average changes in the three EVD parameters (location, scale and shape) are therefore shown for each EVD in Fig. 7. As in Figs. 3–5, values are shown for the 115 catchments, and in this case each point is an average of 50,000 parameter estimates for the catchment (derived from 500 ensemble members, each with 100 samples from the distribution of boot-strapped parameter values). The average parameter values are shown relative to the dominant FGP in the current climate in a particular catchment, as in Fig. 4. Changes in the location parameter correspond to a shift in the distribution (*e.g.* to a higher or lower mean value), and changes in the scale parameter correspond to an expansion or contraction of the distribution (*i.e.* to a higher or lower gradient of values as a function of return period). Changes in the shape parameter represent a change in the skewness of the distribution, *i.e.* in the 'heaviness' in the tails of the distribution. It should also be noted that although a fixed threshold representing the 98.5$^{th}$ quantile is used to set the location parameter for the GP distribution, the value of the location parameter will nevertheless vary between periods as the cumulative distribution function of discharge varies.

The results shown in Fig. 7 indicate that in most catchments the location and scale parameters are expected to increase for all three EVDs, with the exception of some catchments with lower values of FGP (*e.g.* 55 % or less) representing catchments in which snowmelt plays a significant role in flood generation under the current climate. For these snow-dominated catchments, the majority of catchments show a decrease in the location parameter and about half are associated with a decrease in the scale parameter. The three EVDs indicate similar changes, although the average change in the scale parameter is slightly larger for the Gumbel distribution and it is slightly less for the GP distribution. The results also indicate that the average shape parameter decreases in most catchments, for both the GEV and the GP distribution. An exception, however, to this general decrease is seen for a number of catchments with FGPs between 42 and 84 %. These are catchments with a mixed flood generation regime under the current climate, and the increase is consistent with a likely transition to a more rainfall dominated flood regime in the future period. The spatial distribution of the direction of change in the shape parameter is shown in Fig. 8 and indicates a high degree of spatial coherency among the catchments exhibiting an increase in the shape parameter between the reference and future periods. Most of these catchments are found in

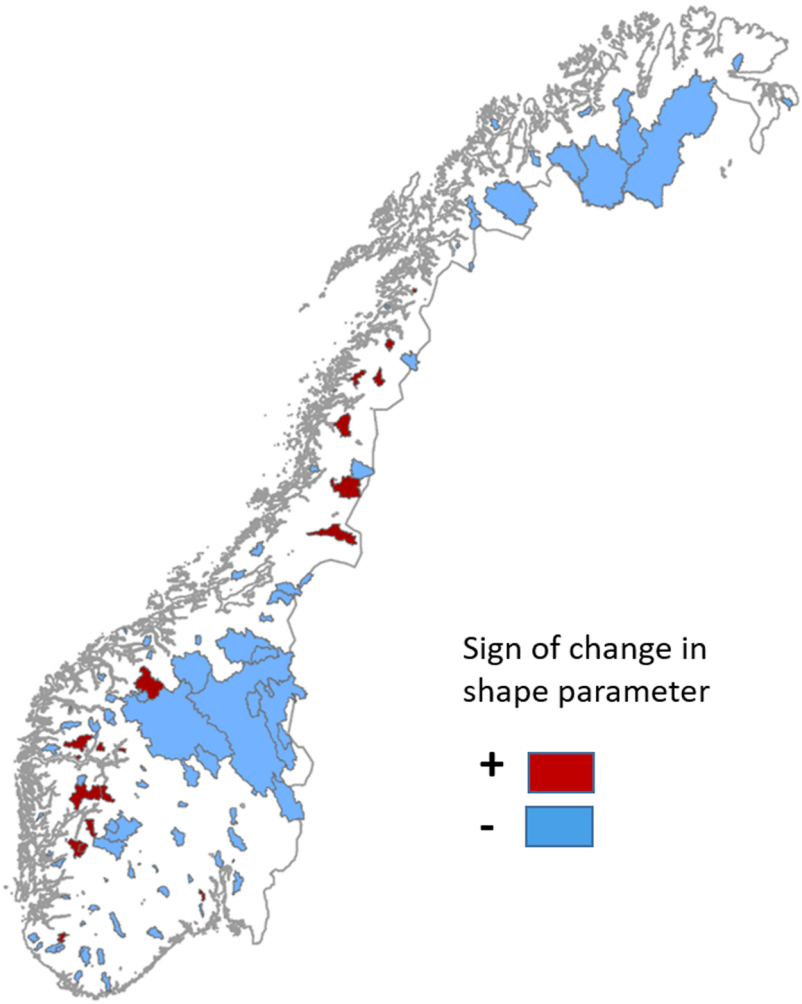


**Fig. 8.** Location of catchments with positive (red) *vs*. negative (light blue) changes in the average shape parameter between the reference (1971-2000) and the future periods (2071-2100). The results are shown for the GEV distribution, and as suggested in Fig. 7, the pattern is very similar for the GP distribution (not shown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

western Norway (Vestlandet) and in Nordland, and several of them correspond with the catchments exhibiting the highest percentage change in the 200 year flood shown in Fig. 3.

## 5. Discussion

Estimated changes in the 200 year flood based on the ensemble median indicate both increases and decreases (Fig. 3) in the catchments considered. The direction of change is related to the dominant flood generation process (Fig. 4) in that only catchments with an FGP of 40 % or less are associated with projected decreases. Simultaneously, Fig. 4 also illustrates that the majority of snow-dominated catchments exhibit small to large increases in the projected median value of the 200 year flood level. The largest median increases are actually seen in catchments with FGPs of 50 % or less. Above this FGP value, the results are less variable between individual catchments, indicating a more consistent response to the effect of projected increases in precipitation. The variable response of flood magnitudes in snow-dominated catchments most likely reflects the competing effects of precipitation and temperature, both of which are projected to increase during the winter season throughout Norway (Hanssen-Bauer et al., 2017). These increases can lead to an increase in winter snow storage, and a larger, albeit earlierspring snowmelt, or to an increase in rain-on-snow events throughout the winter half year, depending on the latitude and the elevation of the catchment. Such variable responses have also been reported for catchments in the Alpine region of Europe (Köplin et al., 2014) and in the western region of North America (Musselman et al., 2018)

The three EVDs give similar projected increases for the 10 year flood magnitude, but for the 200 and 1000 year return periods, the EVDs with a shape parameter (GEV and GP) give larger increases than the Gumbel distribution (Fig. 5). The mean difference between

the distributions of results for EVDs with a shape parameter and those generated by the Gumbel distribution is statistically significant at the (at least) 0.05 significance level for the two higher flood quantiles (*i.e.* 200 and 1000 year return periods). This is not entirely surprising and suggests that distributions that also model the skewness of the extreme values indicate larger changes in the most extreme flood levels. The GEV and GP EVDs give similar (but not identical) estimates for the higher return periods, although they represent analyses on somewhat different time series (annual maxima *vs.* partial duration series). This result concurs with a recent study comparing the use of a GEV *vs*. a GP distribution to assess future changes in the 100 year flood for catchments in the UK (Collet et al., 2017). This suggests that despite the potential problems of stability associated with fitting an EVD with a shape parameter to a time series of limited length (Renard et al., 2013; Kobierska et al., 2018), the average values obtained from the ensemble analyses used here give consistent results for the EVDs with a shape parameter. The goodness-of-fit statistic indicates, however, that it is only in the case of the GP that the null hypothesis can be rejected for most catchments. That the GP distribution has much lower P values than the GEV is due, at least in part, to the larger number of 'observations' available for each time period (*i.e.* typically 50–60) as compared with the 30 annual maximum values used for fitting the GEV and Gumbel distributions. In addition, the threshold value (*i.e.* that determines the location parameter) is fixed to correspond to the 98.5 % level of the cumulative distribution of daily runoff values, leading to only 2 parameters (scale and location) fitted from the partial duration series. Although the selection of this value is somewhat arbitrary, in that thresholds should be tailored to each individual partial duration series (Coles, 2001), the similarity in the results for the GEV and GP EVDs shown in Fig. 5, support the use of this approach, at least for the case when average ensemble values (rather than individual cases) are of interest.

The analysis of the ensemble variance (Fig. 6) indicates that flood frequency analysis introduces a high degree of 'uncertainty' into the range of estimated changes, particularly for the highest flood quantiles. For the case of the GEV and GP distributions, the total variance is, for example, 8–9 times larger for the 1000 year flood estimated with a GEV distribution relative to a Gumbel distribution (Fig. 6a). At first glance, it might be concluded that the use of a Gumbel distribution is preferable, as it does not entail the fitting of a shape parameter and the additional uncertainty introduced by this estimation. However, given that the A–D statistic suggests a better fit for the EVDs with a shape parameter, it is more likely that for the higher flood quantiles the much larger spread in the ensemble of estimates is a realistic depiction of the high degree of uncertainty associated with projections for changes in these flood quantiles. It is also noteworthy that for the case of the 10 year flood, the GP distribution gives, on average, the lowest total variance of all 3 EVDs, thus supporting the use of this distribution for the full range of flood quantiles.

The fractional contribution of FFA to the total ensemble variance is higher for the GEV and GP distributions and is, on average, equal to the contribution from the differences between climate models. The role of hydrological model parameterisation is clearly secondary to these other factors, and in most cases the uncertainty introduced by this factor is (on average for all catchments) almost an order of magnitude less than the combined effects of climate models and flood frequency analyses. It should be emphasised, however, that the decomposition analysis considers estimates for the flood frequency analysis based on an exhaustive parametric bootstrapping procedure and a resampling to arrive at 20 subsamples representing the full range of estimates. For the case of the climate models, the full distribution of models is unknown and the 10 GCM/RCM combinations considered are not necessarily 'independent' models (*i.e.* they represent in many cases the same GCM run with a different RCM; Table 2). Similarly, a complete investigation of the uncertainty introduced by hydrological model parameterisation would need to include a much larger sample of parameter sets. If it were possible to identify the range of estimates from a complete distribution of climate models and hydrological model parameterisations, then these two factors would presumably make a larger overall contribution to the ensemble range than is shown here. In addition, only one hydrological model has been considered, and previous studies indicate that hydrological model structure is also a considerable source of uncertainty in hydrological projections for extreme flows (Poulin et al., 2011; Bosshard et al., 2013). However, the ensemble analysed here represents that which is available in practise for climate change adaptation in Norway (*e.g.* Lawrence, 2016; Hanssen-Bauer et al., 2017). Given this ensemble, it can be concluded from Fig. 6 that statistical flood frequency analysis makes a considerable contribution to the total ensemble variance if uncertainty associated with fitting an EVD is taken into account. This result agrees with the findings of Collet et al. (2017) for estimates of the 100 year flood in Great Britain.

Changes in the average EVD location parameter (reflecting changes in the mean discharge level in the flood series) between the reference and future periods (Fig. 7) have a high degree of correspondence with projected changes in the 200 year flood (Fig. 4). An exception are the several snow-dominated catchments which show large increases in the 200 year flood level, but apparently have decreases or only small increases in the location parameter. It appears that in these catchments, the increase in flood level is associated with an increased value of the scale parameter, such that flood levels increase as a function of return period although the mean flood discharge level is actually lower in the future period. It is also the change in the scale parameter that varies most between EVDs and between catchments. The projected changes in this parameter are, on average, largest for the Gumbel distribution, and this is most likely due to the lack of a shape parameter, *i.e.* a higher gradient is required to accommodate changes in the tail of the distribution that otherwise would be described by the shape parameter.

Most catchments show a decrease in the shape parameter, and this is initially somewhat surprising, particularly given the results shown in Fig. 5 which indicate larger increases for the GP and GEV distributions for the higher return periods. One might assume, for example, that the larger increases in the estimated changes at the higher return periods reflect increases in the shape parameter. Fig. 7, however, indicates that this may only be the case in a subset of catchments, *i.e.* several catchments with a 'mixed' snow regime in the current climate. An increase in rainfall-dominated events in this catchments may well contribute to an increase in the positive skewness of the distribution, *i.e.* reflecting an increase in the probability of the most extreme events. Catchments with a decrease in the shape parameter under a future climate, on the other hand, will most likely have an increase in the events that are extreme in today's climate, but have lower return intervals under a future climate, such that resulting distribution is actually less skewed (although the average flood levels are higher). Similar patterns are seen in the results for changes in flood regimes associated with sea

level rise presented by Buchanan et al. (2018). That the catchments which exhibit a positive increase in the shape parameter are only found in two well-defined geographic areas (Fig. 8), rather than in all catchments with mixed flood regimes, suggests possible links to changes in climatological drivers. This is beyond the scope of the current study, but such linkages between changes in climatic variables, flood generating processes, and the flood quantiles that are affected deserve investigation in further work. An additional factor that can influence the changes in the EVD parameters, but is beyond the scope of the work presented here, is the method used for fitting the extreme value function. In this work, we have applied the L-moments methods although other methods, such as maximum likelihood estimation or Bayesian estimation, could be used and may lead to somewhat different conclusions, particularly regarding changes in the scale *vs*. the shape parameter.

The results presented in this study indicate that EVDs with a shape parameter project somewhat larger increases in higher flood quantiles, particularly for the 1000 year flood, than does the 2-parameter Gumbel distribution. In practice, adaptation to likely future changes flood hazard in Norway is based on projections developed using a 2-parameter Gumbel distribution. Following these projections, a climate change allowance of 0, 20 or 40 % is used (Lawrence, 2016) depending on local conditions (*i.e*. types of catchments in which flood levels are projected to decrease *vs*. expected to have moderate or large increases). For the case of the 200 year flood level, the average difference between the Gumbel and GEV/GP distributions is 5–8 %. This is not a large difference, but does suggest that in some cases a higher category (*e.g*. 40 instead of 20 %) might be chosen for an individual catchment if the GEV or GP distribution were used to develop the estimates. The results also indicate that the use of 40 % as the highest standard allowance is not excessively conservative, in that many catchments actually have a higher projected increase for all three EVDs. For the case of the 1000 year flood level, the differences between the Gumbel and GEV/GP distributions is much larger (13–15 %). In addition, there is a high degree of uncertainty in these estimates, due in part to the fitting of the shape parameter to a data series of limited length. The comparison of the GEV and GP distributions point to a preference for the GP distribution, due to a better fit and a lower ensemble uncertainty.

## 6. Conclusions and recommendations for further work

This study has analysed projections for changes in flood magnitudes under a future climate using a large ensemble of hydrological simulations representing differing GCM/RCM combinations, bias correction methods, hydrological model parameterisations and EVDs for estimating flood return levels. Consistent with earlier work, the median ensemble values indicate an increase in flood hazard under a future climate in all catchments having rainfall-dominated or mixed flood regimes under the current climate. Snowmelt-dominated catchments show a wider range of responses, with some catchments projected to have a decrease in flood magnitudes and others to have an increase. The results also indicate that uncertainty introduced by flood frequency analyses in these estimates is considerable and is of the same magnitude as differences between GCM/RCM combinations. This is a source of uncertainty that has largely been neglected in previous analyses of ensemble uncertainty. There are also differences in estimated changes as a function of the extreme value distribution used for the statistical flood estimation, in addition to the uncertainty underlying estimates based on a given distribution. In particular, the application of EVDs with a shape parameter (GEV and GP) generally leads to larger projected increases in the 200 and 1000 year flood in comparison with the 2-parameter Gumbel distribution.

There are several relevant sources of uncertainty that have not been considered in this study, but which are suitable for further work. The analysis presented here, for example, has assumed stationarity during each of the 30 year time slices representing the current *vs*. the future period. Although this is consistent with previous work and can be partly justified due to the short period considered, a depiction of how flood probabilities change as a continuous function over time would be useful both in climate change adaptation planning and in enhancing our understanding of how flood processes respond to changes in climatological variables. Non-stationary flood frequency methods, however, require the fitting of additional parameters to account for changes in the EVD parameters over time. The comparison of average EVD parameter values between the two time slices presented here indicate both clear patterns (*e.g*. changes in the location parameter as related to projected changes in flooding), and more complex issues (*e.g*. changes in the shape parameter as a function of changes in flood generation processes). Applications of non-stationary methods should therefore seek to interrogate the physical factors behind this complexity (*e.g*. Condon et al., 2015), rather than simply fitting a non-stationary EVD with additional parameters. In addition, this study has evaluated sources of ensemble uncertainty in hydrological projections for 115 gauged catchments where it is possible to calibrate a hydrological model. In most cases, however, flood estimates based on a calibrated hydrological model are not available for the sites where they are needed in practical applications, as these sites often represent ungauged catchments. In current practice in Norway, climate change allowances are selected for ungauged catchments based on a qualitative assessment and comparison of the hydrological and climatological characteristics of the catchment of interest relative to published results for the 115 catchments. This informal transfer of information between gauged and ungauged catchment also introduces an element of subjectivity and inherent uncertainty. Methods for regionalising catchment-based hydrological projections for likely changes in flood levels and flood regimes to ungauged catchments are, therefore, also required if one is to improve the reliability of estimates available for climate change adaptation.

## CRediT authorship contribution statement

**Deborah Lawrence:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Funding acquisition, Project administration, Writing - original draft, Writing - review & editing.

## Declaration of Competing Interest

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

## Acknowledgements

This work has been partly supported by the Norwegian Research Council through the Klimaforsk research project 'ExPrecFlood' Project no. 244175 and by internal research funding from the Norwegian Water Resources and Energy Directorate (NVE). The author thanks NVE colleagues Wai Kwok Wong and Ingjerd Haddeland for extracting the EUROCORDEX data for the catchments used in this study and the EUROCORDEX initiative for making these climate model runs available.

Declaration of interest: None

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrh.2020.100675.

## References

Alfieri, L., Burek, P., Feyen, L., Forzieri, G., 2015. Global warming increases the frequency of river floods in Europe. Hydrol. Earth Syst. Sci. 19, 2247–2260.
Arheimer, B., Lindström, G., 2015. Climate impact on floods: changes in high flows in Sweden in the past and the future (1911-2100). Hydrol. Earth Syst. Sci. 19, 771–784.
Arnell, N., Gosling, S., 2016. The impacts of climate change on river flood risk at the global scale. Clim. Change 134, 387–401.
Asquith, W.H., 2017. Lmomco: L-Moments, Censored L-Moments, Trimmed L-Moments, L-Comoments, and Many Distributions. R Package Version 2.2.9. Texas Tech University, Lubbock, Texas.
Blöschl, G., Hall, J., Parajka, J., Perdigao, R.A.P., Merz, B., et al., 2017. Changing climate shifts timing of European floods. Science 357, 588–590.
Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resour. Res. 49, 1523–1536.
Buchanan, M.K., Oppenheimer, M., Kopp, R.E., 2018. Amplification of flood frequencies with local sea level rise and emerging flood regimes. Environ. Res. Lett. 12. https://doi.org/10.1088/1748-9326/aa6cb3.
Cloke, H., Wetterhall, F., He, Y., Freer, J.E., Pappenberger, F., 2013. Modelling climate impact on floods with ensemble projections. Q. J. R. Meteorol. Soc. 139, 282–297.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag London Limited.
Collet, L., Beevers, L., Prudhomme, C., 2017. Assessing the impact of climate change and extreme value uncertainty to extreme flows across Great Britain. Water 9, 103. https://doi.org/10.3390/w9020103.
Condon, L.E., Gangopadhyay, S., Pruitt, T., 2015. Climate change and non-stationary flood risk for the upper Truckee River basin. Hydrol. Earth Syst. Sci. 19, 159–175.
Dakhlaoui, H., Ruellland, D., Tramblay, Y., 2019. A bootstrap-based differential split-sample test to assess the transferability of conceptual rainfall-runoff models under past and future climate variability. J. Hydrol. 575, 470–486.
Dankers, R., Feyen, L., 2008. Climate change impact on flood hazard in Europe: an assessment based on high-resolution climate simulations. J. Geophys. Res. 113, D19105. https://doi.org/10.1029/2007/JD009719.
Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C.M., Jacob, D., Lenderink, G., Christensen, O.B., 2007. An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Clim. Change 81, 53–70.
Fleig, A., Andreassen, L.M., Barfod, E., Haga, J., Haugen, L.E., Hisdal, H., Melvold, K., Saloranto, T., 2013. Norwegian Hydrological Reference Dataset for Climate Change Studies. NVE Report 2/2013. (Accessed 18 February 2020). http://publikasjoner.nve.no/rapport/2013/rapport2013_02.pdf.
Franks, S.W., Kuczera, G., 2002. Flood frequency analysis: evidence and implications of secular climate variability, New South Wales. Water Resour. Res. 38. https://doi.org/10.1029/2001WR000232.
Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note: downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390.
Hanssen-Bauer, I., Førland, E. J., Haddeland, I., Hisdal, H., Mayer, S., Nesje, A., Sandven, J. E. Ø, Sandø, A. B., Sorteberg, A., Ådlandsvik, B. (Eds.), 2015. Klima i Norge 2100: Kunnskapsgrunnlag for klimatilpasning oppdatert i 2015, (Climate in Norway 2100: Knowledge basis for climate change adaptation updated in 2015). Norwegian Climate Services Centre Report 2/2015. (In Norwegian).
Hanssen-Bauer, I., Førland, E. J., Haddeland, I., Hisdal, H., Lawrence, D., Mayer, S., Nesje, A., Sandven, J. E. Ø, Sandø, A. B., Sorteberg, A., Ådlandsvik, B. (Eds.), 2017. Climate in Norway 2100: a knowledge based for climate adaptation. Norwegian Climate Services Centre Report 1/2017.
Hattermann, F.F., Weiland, M., Huang, S., Krysanova, V., Kundzewicz, Z.W., 2011. Model-supported impact assessment for the water sector in central Germany under climate change – a case study. Water Resour. Manage. 25, 3113.
Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction – the ISI-MIP approach. Earth Syst. Dyn. 4, 219–236.
Hirabayashi, Y., Mahendran, R., Loirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Chang. 3, 816–821.
Hundecha, Y., Sunyer, M., Lawrence, D., Madsen, H., Willems, P., Bürger, G., Kriaučiūnienė, J., Loukas, A., Martinkova, M., Osuch, M., Vasiliades, L., von Christierson, B., Vormoor, K., Yücel, I., 2016. Inter-comparison of statistical downscaling methods for projection of extreme flow indices across Europe. J. Hydro. 541, 1273–1286.
Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Deque, M., Georgievski, G., Georgopoulou, E., Gobiat, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., Yiou, P., 2014. EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg. Environ. Change 14, 563–578.
Kobierska, F., Engeland, K., Thorarinsdottir, Th., 2018. Evaluation of design flood estimates – a case study for Norway. Hydrology Research 49 (2), 450–465.
Kochanek, K., Renard, B., Arnaud, P., Lang, M., Cipriani, T., Sauquet, E., 2014. A data-based comparison of flood frequency analysis methods used in France. Nat. Hazards Earth Syst. Sci. 14, 295–308.
Köplin, N., Schädler, B., Viviroli, D., Weingartner, R., 2014. Seasonality and magnitude of floods in Switzerland under future climate change. Hydrol. Process 28, 2567–2578.
Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., Kundzewicz, Z.W., 2018. How the performance of hydrological models relates to credibility of projections under climate change. Hydrol. Sci. J. 63, 696–720.
Kuczera, G., Franks, S., 2006. Australian Rainfall and Runoff Book IV, Estimation of Peak-Discharge. Revision draft (online), Engineers Australia. Available from www.arr.org.au. (Accessed 28 November 2019). .

Laio, F., 2004. Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. Water Resour. Res. 40, W09308. https://doi.org/10.1029/2004WR003204.
Lang, M., Ouarda, T.B.M.J., Bobée, B., 1999. Towards operational guidelines for over-threshold modelling. J. Hydrol. 225, 103–117.
Lawrence, D., 2016. Klimaendring og framtidige flommer i Norge (Climate change and future floods in Norway). NVE Rapport 81/2016 (in Norwegian). (Accessed 18 February 2020). http://publikasjoner.nve.no/rapport/2016/rapport2016_81.pdf.
Lawrence, D., Haddeland, I., 2011. Uncertainty in hydrological modelling of climate change impacts in four Norwegian catchments. Hydrol. Res. 42, 457–471.
Lawrence, D., Hisdal, H., 2011. Hydrological Projections for Floods in Norway under a Future Climate. NVE Report 5/2011. (Accessed 18 February 2020). http://publikasjoner.nve.no/report/2011/report2011_05.pdf.
Lawrence, D., Haddeland, I., Langsholt, E., 2009. Calibration of HBV Hydrological Models Using PEST Parameter Estimation. NVE Report 1/2009. (Accessed 18 February 2020). http://publikasjoner.nve.no/report/2009/report2009_01.pdf.
Leander, R., Buishand, T.A., van den Hurk, B.J.J.M., deWit, M.J.M., 2008. Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output. J. Hydrol. 351, 331–343.
Lussana, C., Tveito, O.E., Dobler, A., Tunheim, K., 2019. seNorge_2018, daily precipitation and temperature datasets over Norway. Earth Syst. Sci. Data 11, 1531–1551.
Madsen, H., Rasmussen, P.F., Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At-site modeling. Water Resour. Res. 33, 747–757.
Madsen, H., Lawrence, D., Lang, M., Martinkova, M., Kjeldsen, T., 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. J. Hydrol. 519, 3634–3650.
Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeβl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.J., Onof, C., Vrac, M., Theile-Eich, I., 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48, RG3003. https://doi.org/10.1029/2009RG000314.
Meresa, H.K., Romanowicz, R.J., 2017. The critical role of uncertainty in projections of hydrological extremes. Hydrol. Earth Syst. Sci. 21, 4245–4258.
Merz, R., Blöschl, G., 2003. A process typology of regional floods. Water Resour. Res. 39, 1340.
Miljøverndepartement, 2012. Klimatilpasning i Norge (Climate change adaptation in Norway) Meld. St. nr. 33 (2012–2013). Miljøverndepartement, Oslo (Accessed 18 February 2020). https://www.regjeringen.no/contentassets/e5e7872303544ae38bdbdc82aa0446d8/no/pdfs/stm201220130033000dddpdfs.pdf.
Musselman, K.N., Lehner, F., Ikeda, K., Clark, M.D., Prein, A.F., Liu, C., Barlage, M., Rasmussen, R., 2018. Projected increases in rain-on-snow flood risk over western North America. Nat. Clim. Change 8, 808–812.
Osuch, M., Lawrence, D., Meresa, H.K., Napiorkowski, J.J., Romanowicz, R.J., 2016. Projected changes in flood indices in selected catchments in Poland in the 21st century. Stoch. Environ. Res. Risk Assess. 31, 2435–2457.
Pettersson, L.-E., 2013. Totalavløpet fra Norges vassdrag 1900-2010. (Runoff from Norwegian basins 1900-2010). NVE Rapport 39/2012. In Norwegian. http://publikasjoner.nve.no/rapport/2012/rapport2012_39.pdf. (Accessed 18 February 2020). .
Piani, C., Weedon, G.P., Best, M., Gomes, S.M., Viterbo, P., Hagemann, S., Haerter, J.O., 2010. Statistical bias correction of global simulation daily precipitation and temperature for the application of hydrological models. J. Hydrol. 395, 199–215.
Poulin, A., Brisette, F., Leconte, R., Arsenault, R., Malo, J.-S., 2011. Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. J. Hydrol. 409, 626–636.
Prudhomme, C., Jakob, D., Svensson, C., 2003. Uncertainty and climate change impact on the flood regime of small UK catchments. J. Hydrol. 277, 1–23.
Prudhomme, C., Crooks, S., Kay, A.L., Reynard, N., 2013. Climate change and river flooding: part 1. Classifying the sensitivity of British catchments. Clim. Change 119, 933–948.
Read, L.K., Vogel, R.M., 2015. Reliability, return periods and risk under nonstationarity. Water Resour. Res. 51 (8), 6381–6398.
Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., Najib, K., Carreau, J., Arnaud, P., Aubert, Y., Borchi, F., Soubeyroux, J.-M., Jourdain, S., Veysseire, J.-M., Sauquet, E., Cipriani, T., Auffray, A., 2013. Data-based comparison of frequency analysis methods: a general framework. Water Resour. Res. 49, 825–843.
Roald, L.A., 2013. Flom i Norge (Floods in Norway). Forlaget Tom & Tom, Vestfossen (In Norwegian).
Rossi, F., Fiorentino, M., Versace, P., 1984. Two-component extreme value distribution for flood frequency analysis. Water Resour. Res. 20, 847–856.
Saeb, A., 2018. Goodness of Fit Test for Continuous Distribution Functions. R Package 'gnFit' Version 0.2.0. Available at: https://cran.r-project.org/web/packages/gnFit/gnFit.pdf. (Accessed 18 February 2020). .
Sælthun, N.R., 1996. The "Nordic" HBV model, NVE Publication 7/1996. Available at http://publikasjoner.nve.no/publication/1996/publication1996_07.pdf. (Accessed 18. February 2020). .
Sharma, A., Wasko, C., Lettenmaier, D.P., 2018. If precipitation extremes are increasing, why aren't floods? Water Resour. Res. 54, 8545–8551.
Soriano, E., Mediero, L., Garijo, C., 2019. Selection of bias correction methods to assess the impact of climate change on flood frequency curves. Water 11, 2266. https://doi.org/10.3390/w11112266.
Steele-Dunne, S., Lynch, P., McGrath, R., Semmler, T., Wang, S., Hanafin, J., Nolan, P., 2008. The impacts of climate change on hydrology in Ireland. J. Hydrol. 356, 28–45.
Steinschneider, S., Wi, S., Brown, C., 2015. The integrated effects of climate and hydrologic uncertainty on future flood risk assessments. Hydrol. Process 29, 2823–2839.
Sunyer, M., Hundecha, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., Vormoor, K., Bürger, G., Hanel, M., Kriaučiūnienė, Loukas, A., Osuch, M., Yücel, I., 2015. Intercomparison of statistical downscaling methods for projection of extreme precipitation in Europe. Hydrol. Earth Syst. Sci. 19, 1827–1847.
Teutschbein, C., Wetterhal, F., Seibert, J., 2011. Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. Clim. Dyn. 37, 2087–2105.
Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rokovec, O., Samaniego, L., Sheffield, J., Wood, E.F., Zink, M., 2018. Multi-model ensemble projections of European river floods and high flows at 1.5, 2 and 3 degrees global warming. Environ. Res. Lett. 13 (1) 014003.
Veijalainen, N., Lotsari, E., Alho, P., Vehviläinen, B., Käyhkö, J., 2010. National scale assessment of climate change impacts on flooding in Finland. J. Hydrol. 391, 333–350.
von Storch, H., Zwiers, F., 2013. Testing ensembles of climate change scenarios for "statistical significance". Clim. Change 117, 1–9.
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, 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.
Vormoor, K., Heistermann, M., Bronstert, A., Lawrence, D., 2018. Hydrological model parameter (in)stability – "crash testing" the HBV model under contrasting flood seasonality conditions. Hydrol. Sci. J. 63, 991–1007.
Wilson, D., Hisdal, H., Lawrence, D., 2010. Has streamflow changed in the Nordic countries? Recent trends and comparisons to hydrological projections. J. Hydrol. 394, 334–346.
Winsemius, H.C., Aerts, J.C.J.H., van Beek, L.P.H., Bierkens, M.F.P., Bouwman, A., Jongman, B., Kwadijk, J.C.J., Liftvoet, W., Lucas, P.L., van Vuuren, D.P., Ward, P.J., 2016. Global drivers of future river flood risk. Nat. Clim. Change 6, 381–385.
Wong, W.K., Haddeland, I., Lawrence, D., Belding, S., 2016. Gridded 1 X 1 Km Climate and Hydrological Projections for Norway. NVE Rapport 59/2016. Available at http://publikasjoner.nve.no/rapport/2016/rapport2016_59.pdf. (Accessed 18 February 2020). .
Yang, W., Andreasson, J., Graham, L.P., Olsson, J., Rosberg, J., Wetterhall, F., 2010. Distribution-based scaling to improve the usability of regional climate model projections for climate change impact studies. Hydrol. Res. 41, 211–222.