Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation

Hossein Tabari

KU Leuven, Department of Civil Engineering, Belgium (hossein.tabari@kuleuven.be)

Abstract

A reliable estimation of hydrological extremes with potentially severe socio-economic impacts is of crucial importance for efficient planning and design of hydraulic structures. Extreme value theory provides a firm theoretical foundation for the statistical modelling of extreme hydrological events. The dilemma in the modelling is on whether to use block maxima (BM) or peak-over-threshold (POT) method, each with its own cons and pros. It remains unexplored to what extent future projected changes in extreme hydrological events are influenced by the method choice, especially when some simplifications are made to lessen the computational burden for large-scale studies. This study addresses this research question by a comparative analysis between the BM and POT methods on future climate change impact on global flood and extreme precipitation. The extreme precipitation analysis is performed using 24 CMIP5 general circulation models (GCMs), while the flood analysis is based on a multi-model ensemble of 20 members including five global impact models forced by four CMIP5 GCMs. The results reveal that the BM and POT methods agree on the sign of changes in flood and extreme precipitation intensities, but disagree on the magnitude. The discrepancy between the BM and POT results increases with event extremity. The difference also varies with season, where the difference in the global land area with increasing signals peaks in winter at the rates of 11% for extreme precipitation and in summer at the rates of 3.5% for flood.

Keywords: Hydrological extremes, Block maxima, Peak-over-threshold, Climate change signals, Spatial distribution, Event extremity

1. Introduction

Hydrological extremes – floods and droughts – engross society as a formidable concern to public safety, welfare and the economy (Vörösmarty et al., 2013), justifying rigorous statistical analyses of the extreme events. There is, therefore, a pressing need for decision-makers and water managers to better understand the complexity of these events to mitigate the related human and economic losses. In this context, a precise approximation of hydrological extremes is one of the keys towards an efficient hydraulic infrastructure design (Hosseinzadehtalaei et al., 2020a), as fallacious analysis may lead to flawed estimates, resulting in either an underestimation of the design criteria and increased failure risk or an overestimation and increased infrastructure construction cost (Papalexiou & Koutsoyiannis, 2013; Yan et al., 2018).

Inference about extremes in a complex interacting hydro-climatology system is challenging due to the scarcity of events and thus their small sample size (Tabari, 2019). Extreme value theory (EVT; Coles, 2001) identifies extreme events, characterized by either very small or very large values. EVT consists of two fundamental methods that are well-adapted to the disciplines of water resources engineering and hydrology (Katz et al., 2002). The traditional method, known as block maxima (BM), considers the maximum of all recorded values in each year, making the sample size equal to the number of years of data. The major advantages of the BM method are its simplicity and the independence of extracted extremes. Its main shortcoming is that it samples only one event per year which may result in a loss of information as other events than the maximum of a year may exceed the annual maxima of other years. It is particularly problematic when the data record is short (i.e., small sample size). Another drawback of the BM method is the inclusion of some lower observations that are still the maximum value in the year (annual maxima in dry years). Annual maximum precipitation depths of coarse time-resolutions can also be potentially underestimated (Hershfield, 1961; Weiss, 1964; Young and McEnroe, 2003; Yoo et al., 2015; Morbidelli et al., 2017, 2018, 2020).

The second method for the extreme value analysis, which has become the method of choice in recent hydrological applications, is peak-over-threshold (POT) method. The POT method supplies a larger sample size for a more precise distribution parameter approximation compared to the BM method which chooses a single maximum event in each year (Lang et al., 1999). It also functions properly in the case of an asymmetric distribution tail (McNeil & Frey, 2000). However, the use of the POT method involves some analytical complexities, including choosing an appropriate threshold level for extreme events and assuring the independence of extremes. The choice of the threshold is based on a trade-off between variance and bias. A high threshold value satisfies trustworthy tail asymptotics underlying the distribution estimation and reduces the associated bias, however, increases the variance of the parameter approximation from a smaller sample size (Scarrott & MacDonald, 2012; Serinaldi & Kilsby, 2014; Nagy et al., 2017). Ensuring the independence of data series is also crucial as extreme hydrological events often appear in clusters, being strongly autocorrelated, e.g., occurrence of some flood peaks on the recession curves of the precedent peaks (Bezak et al., 2014; Mangini et al., 2018).

More recently, increasing attention has been paid to hydrological extremes because climate change is expected to intensify the key elements of the hydrological cycle such as precipitation and river flows and make them more extreme (Kharin et al. 2013; Tabari & Willems, 2018; van der Wiel et al., 2019; Allan et al., 2020; Tabari, 2020). The issue derives basically from the propensity of the atmosphere to retain more moisture as it warms as described by the Clausius–Clapeyron (C-C) relationship (about 7% per degree Celsius increase in temperature) (Held & Soden, 2006). Climate change makes the statistical analysis of hydrological extremes more challenging and further complicates resolving the dilemma of whether to use BM or POT method. Though, from a theoretical point of view, the two methods should yield the same results if applied properly for the same sample size (Wang and Holmes, 2020), some simplifications in large-scale climate change studies to decrease the computational burden may enlarge the numerical discrepancies between the results of the two methods. Among them is the selection of a single threshold for the POT method for the entire domain to make the statistical analyses more computationally feasible, while

the threshold is recommended to be selected based on regional (distribution of humid and arid regions) and seasonal characteristics of precipitation (Wang et al., 2019).

Practical drawbacks and advantages of the BM and POT methods should be considered when choosing between them for each case study based on its context and purpose. For instance, the selection of the optimum threshold for the POT method for global/continental studies is a time-consuming process and makes the total computational time for the POT method much longer compared to the BM method. Because of these difficulties, the POT method is still under-employed for large-scale climate change impact studies on hydrological extremes (Huang et al., 2017; Hosseinzadehtalaei et al., 2020a, 2020b), and the application of the BM is more widespread (Arnell and Gosling, 2016; Ward et al., 2020; Willner et al., 2018; Eilander et al., 2020).

Although several studies have compared the POT and BM methods for extreme value analyses, the comparison has been made on a limited historical time series in a small area (e.g., Jaruskova and Hanek, 2006; Li et al., 2016, Nagy et al., 2017; Vrban et al., 2018). It remains to determine the extent to which climate change results for extreme events on large scales are influenced by the choice of extreme value analysis method. The central question is how the difference between the two methods is propagated into climate change signals (narrowed or enlarged). Other open questions concern the extent to which the climate change signal difference varies spatially, and with season and event extremity. This study is the first global study to address these research questions by a comparative analysis between the BM and POT methods for future (2070–2099) climate change footprints on global flood and extreme precipitation. The climate change impact assessment for extreme precipitation is performed using 24 Coupled Model Intercomparison Project Phase 5 (CMIP5) general circulation models (GCMs) under the representative concentration pathway (RCP) 8.5. The flood analysis is based on five global impact models (GIMs) fed by four CMIP5 GCMs under the RCP8.5 scenario. Using this high-end scenario makes the results of this study comparable with those of previous global flood (Dankers et al., 2014; Giuntoli et al., 2015; Willner et al., 2018) and extreme precipitation (Tabari et al., 2019) studies which have used this scenario.

2. Materials and Methods

2.1. Data

Daily precipitation simulations from 24 CMIP5 (Taylor et al., 2012) GCMs for the historical period from 1971 to 2000 and the future period from 2070 to 2099 forced by RCP8.5 scenario are used: ACCESS1-0, ACCESS1-3, CanESM2, CMCC-CESM, CMCC-CM, CMCC-CMS, CNRM-CM5, CSIRO-Mk3-6-0, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, HadGEM2-AO, HadGEM2-CC, Had-GEM2-ES, INMCM4, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3. The data are taken only from the first run (r1) of each GCM. In this way, the models have an equal weight in the ensemble median. The historical and future (RCP8.5) monthly mean temperature simulations from the same 24 GCMs are also used to estimate global warming. The CMIP5 GCMs were selected based on the availability of the historical and RCP8.5 simulations of both daily precipitation and monthly temperature. The extreme precipitation analyses are limited to all land grid cells where the extremes will impact the society the most.

In the framework of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Warszawski et al., 2014), global impact models (GIMs) are driven by climate projections from four CMIP5 GCMs (GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and MIROC5). To this end, the raw outputs of the four GCMs were interpolated to 0.5° spatial resolution and bias-corrected to the EWEMBI (Earth2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP; Lange, 2019) dataset using a trend preserving method (Hempel et al., 2013). The EWEMBI dataset, compiled to support the bias correction of climate input data in phase 2b of ISIMIP (ISIMIP2b; Frieler et al., 2017), is based on ERA-Interim reanalysis data (ERA-Interim; Dee et al., 2011), WATCH forcing data methodology applied to ERA-Interim reanalysis data (WFDEI; Weedon et al., 2014), earth2Observe forcing data (E2OBS; Calton et al., 2016) and NASA/GEWEX Surface Radiation Budget data (SRB; Stackhouse Jr. et al., 2011) to bias-correct E2OBS shortwave and longwave radiation (Lange, 2019).

Global daily discharge simulations for the historical (1971-2000) and future (2070-2099) periods under RCP8.5 scenario from five GIMs including three global hydrological models of H08 (Hanasaki et al., 2008),

MATSIRO (Takata et al., 2003) and WaterGAP2 (Mueller Schmied et al., 2016), the dynamic global vegetation model of LPJmL (Bondeau et al., 2007) and the global land surface model of CLM4.5 (Thiery et al., 2017) are utilized (Table 1). Two types of human influence and land-use scenarios were used in the discharge simulations of the GIMs: time-varying (“histsoc” experiment) and fixed (“2005soc” experiment) scenarios. In the former, land use and other human influences (e.g., reservoirs & dams, water abstraction for domestic, industrial and agricultural uses) vary across the historical period, while in the latter, they are held fixed at 2005 (present-day) levels. The histsoc and 2005soc scenarios were used respectively for historical and future discharge simulations in all the GIMs except for CLM4.5 which made use of the 2005soc scenario for both historical and future simulations.

The GIMs also differ in their choice of evapotranspiration scheme; the Priestley-Taylor method is used in LPJmL and WaterGAP2, the bulk formula in H08 and MATSIRO and the Monin-Obukhov similarity theory in CLM4.5 (Table 1). There are differences in the snow schematization across the GIMs; CLIM4.5, H08 and MATSIRO use the physically-based energy balance method to simulate snow accumulation and melt, and LPJmL and WaterGAP2 use the degree-day method. The runoff scheme and input climatic variables also differ among the GIMs. All these differences in the model parameterization can potentially lead to different simulations of streamflow timing and magnitude (Do et al., 2020).

For the global analysis, model grid cells with historical annual flow maxima of $<1 \text{ m}^3 \cdot \text{s}^{-1}$ are excluded, because of less importance for flood assessment and inadequate data for probability distribution fitting. The Southern Hemisphere high-latitude land mass ($> 60^\circ \text{S}$) is also screened out which is almost uninhabited and unaffected by flooding.

2.2. Extreme value analysis

As for the first method considered here for extreme value analyses, the block maxima time series of precipitation and flow extremes for each grid cell are derived for both historical and future periods. Extreme value theory shows that BM extremes can be approximated most accurately using the generalized extreme value (GEV) distribution (Coles, 2001). The GEV is fitted to the flow and precipitation BM series. The

GEV distribution is characterized by three parameters of location (μ), scale (σ) and shape (ξ) which respectively describe the center of distribution, the deviation around the mean and the tail behavior of the distribution. According to the shape parameter, three extreme distributions are defined as Gumbel (type I), Fréchet (type II) and Weibull (type III) corresponding to $\xi=0$, $\xi>0$ and $\xi<0$, respectively. These three cases are called ‘light-tailed’, ‘bounded’, and ‘heavy-tailed’, respectively. The cumulative distribution function (CDF) of the GEV for a random variable x is formulated as follows:

$$F_{(\mu,\sigma,\xi)}(x) = \begin{cases} \exp\{-[1 + \xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}}\} & \text{for } \xi \neq 0 \\ \exp\{-\exp[-(\frac{x-\mu}{\sigma})]\} & \text{for } \xi = 0 \end{cases} \quad (1)$$

The return level of extreme precipitation and river flow for return periods ranging from 2 to 50 years at each grid cell is computed by:

$$I_T = \begin{cases} \mu + \frac{\sigma}{\xi} \left\{ 1 - \left[-\ln \left(1 - \frac{1}{T} \right) \right]^\xi \right\} & \text{if } \xi \neq 0 \\ \mu - \sigma \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] & \text{if } \xi = 0 \end{cases} \quad (2)$$

where I is the return level or event intensity associated with the return period T .

As the second approach for the extreme value analysis, the POT series (or partial duration series) are extracted from the whole time series by selecting events exceeding a pre-specified critical value (threshold). Extreme value theory indicates that POT extremes follow generalized Pareto distribution (GPD) (Katz et al., 2002; Lazoglou et al., 2019; Hosseinzadehtalaei et al., 2020a). Similar to the GEV, the GPD is characterized by μ , σ and ξ . Accordingly, $\xi=0$, $\xi>0$, and $\xi<0$ correspond to Exponential, Pareto and Beta distributions, respectively. The CDF of the GPD is defined by:

$$F_{(\mu,\sigma,\xi)}(x) = \begin{cases} 1 - [1 + \xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}} & \text{for } \xi \neq 0 \\ 1 - \exp[-\frac{x-\mu}{\sigma}] & \text{for } \xi = 0 \end{cases} \quad (3)$$

The return level of extreme precipitation and river flow in each grid cell for return periods ranging from 2 to 50 years for the chosen threshold u is calculated as:

$$I_T = \begin{cases} u + \frac{\sigma}{\xi} [1 - (\lambda T)^{-\xi}] & \text{if } \xi \neq 0 \\ u + \sigma \ln(\lambda T) & \text{if } \xi = 0 \end{cases} \quad (4)$$

where λ is equal to n/M , where n is the total number of POT and M is the number of years of the data (30 in this case).

Previous studies have shown that the POT method satisfies the Poisson process (greater statistical efficiency) and outperforms the BM method when the average number of the sampled peaks per year is greater than 1.65 (Cunnane, 1973; Nagy et al., 2017; Maposa and Cochran, 2017). This threshold is, therefore, selected for the POT method for the comparison with the BM results. The sensitivity of the results to the choice of the threshold for the POT method is also investigated. The sensitivity is characterized here by standard deviation across the climate change signals derived from different thresholds (1.65, 3, 4 and 5 peaks per year, on average; PPY). To estimate the bias of the inclusion of lower observations of dry years in the BM method, the lowest block maxima is used as the threshold for the POT method and its results are compared with the 1.65 PPY case. Also, to ensure the independence of the flow POT series, only one peak within each 15-day interval is selected following the previous studies (Mallakpour & Villarini, 2015; Dong et al., 2019). The precipitation POT series is assumed to be independent as the extremal behavior of precipitation process is not significantly influenced by dependence at the daily scale (Lombardo et al., 2019).

2.3. Climate change impact assessment

The changes in extreme precipitation and flood intensities are described in this study as the ratio between the intensities of the end twenty-first century (2070–2099) and the end twenty century (1971–2000). The changes are calculated for the whole year and per season, where seasons are defined as winter (December–February in the Northern Hemisphere and June–August in the Southern Hemisphere), spring (March–May in the Northern Hemisphere and September–November in the Southern Hemisphere), summer (June–August in the Northern Hemisphere and December–February in the Southern Hemisphere) and autumn (September–November in the Northern Hemisphere and March–May in the Southern Hemisphere). The signal-to-noise ratio (SNR) is used to quantify the agreement in the changes across ensemble members and the significance of the signals (Prudhomme et al., 2014; Marx et al., 2018). Here, the SNR is calculated as

the ensemble median divided by the inter-quantile range of the multi-model ensemble (i.e., 75th percentile minus 25th percentile). A significant change in extreme precipitation and flood intensities is expressed by a SNR of higher than 1 (i.e., signal larger than noise).

To estimate extreme precipitation-temperature scaling, extreme precipitation changes for each of the 24 GCMs are scaled by its change in global mean surface air temperature. The air temperature change is computed by comparing the 30-year global mean annual temperature between the end twenty century (1971–2000) and the end twenty-first century (2070–2099) under RCP8.5 scenario. The significance of the difference between the BM and POT results is tested using the Wilcoxon rank sum test.

In order to provide a congruous evaluation of extreme precipitation changes across the CMIP5 GCMs with different spatial resolutions, they are resampled from their native scale to a 0.5° (latitude) $\times 0.5^\circ$ (longitude) grid. Resampling can be done using two procedures: 1) last-step procedure where climate change signals are calculated directly on the native grids of climate models, and then resampled to the destination grid, and 2) first-step procedure where precipitation is first resampled to the destination grid, and then climate change signals are computed. Diaconescu et al. (2015) compared these two procedures and found that daily precipitation indices and standard deviation fields present more differences in the first-step procedure, regardless of the mathematical method used for resampling. Hence, in order to minimize the errors in the calculation of the extreme precipitation changes, the last-step procedure is used in this study.

3. Results

3.1. Extreme precipitation change difference between the BM and POT methods

A comparison between global median changes in extreme precipitation intensity with return periods ranging from 2 to 50 years derived from the BM and POT methods for the whole year shows almost same changes up to the return period of 15 years (Fig. 1). For the return periods of longer than 15 years, the ensemble multi-model medians for the BM and POT methods diverge, with slightly larger BM-based changes (up to 2.5% for a 50-year return period). The difference between the results of the two methods is,

however, significant for the return periods of longer than 36 years (Fig. 1). For both methods, the changes increase with precipitation extremity. The model uncertainty also increases with the event extremity, leading to a constant change significance (SNR) with the event extremity (Fig. S1).

The difference between the BM and POT results varies with season, reaching 6.8% for a 50-year return period in summer. For high return periods of whole year ($T > 13$), winter ($T > 21$) and spring ($T > 30$) and all return periods of summer and autumn, the uncertainty range in the BM-derived changes is larger (Fig. 1). The changes for both methods and for all cases are statistically significant; however, SNR for the POT-derived changes are generally larger compared with the BM-derived ones for whole year, summer and autumn, and smaller for winter and spring (Fig. S1). The difference between the changes computed by the BM and POT methods is larger for summer where the difference is significant for the return periods of longer than 15 years (Fig. 1). For spring, autumn and winter, the difference is significant for the return periods of higher than 24, 46 and 49 years, respectively.

< Fig. 1 here please >

A strong and statistically significant ($\alpha < 0.01$) relationship is found between changes in extreme precipitation intensities and the mean temperature response in the individual model simulations (Fig. 2). The whole-year extreme precipitation intensities with return periods of 5, 15, 25 and 50 years obtained from the BM method increase respectively by 4.7%, 4.8%, 4.9% and 5.1% per degree warming, which is close to the rate (5 % °C⁻¹) reported by Donat et al. (2016) for annual maximum daily precipitation (Rx1day). The extreme precipitation-temperature scaling rate is smaller for the POT method: 4.6, 4.7, 4.7 and 4.7 % °C⁻¹ for return periods of 5, 15, 25 and 50 years, respectively. Although the scaling rates from the BM method are higher, the significance and the correlation coefficient are larger for the POT method (Fig. 2). The difference between the scaling rates computed by the BM and POT methods is significant at the 0.05 level only for the 50-year return period (not shown).

< Fig. 2 here please >

The latitudinal distribution of changes in extreme precipitation intensity with a 25-year return period as a function of global warming is compared between the BM and POT methods (Fig. 3). There is a high consistency between the results of the two methods at the latitudes poleward of 28° in the Northern Hemisphere, where the whole-year extreme precipitation-temperature scaling rate gradually grows towards the North Pole and exceeds 10% °C⁻¹. The results of the BM and POT methods are also consistent at the latitudes poleward of 20° in the Southern Hemisphere. The difference between the scaling rates of the two methods is statistically significant at the 0.05 level around the Tropic of Cancer in the Northern Hemisphere (latitudes between 16° and 26°) and close to the Equator (south latitudes between 1° and 5°) (Fig. 3).

< Fig. 3 here please >

Fig. 4 shows that 25-year winter extreme precipitation increase with warming is close to what would be expected from the C-C relationship, while the increase is smallest for summer. The POT-derived 25-year extreme precipitation increases by 6.9%, 5.1%, 3.8% and 6.2% per degree warming for winter, spring, summer and autumn, respectively. The extreme precipitation-temperature scaling rate for the BM method is slightly smaller: 6.7%, 4.3%, 3.8% and 6.0% per degree warming for winter, spring, summer and autumn, respectively. The scaling relationships for both methods are statistically significant at the 0.01 level, although the significance and the correlation coefficients are larger for the POT method (Fig. 4). The difference between the BM- and POT-derived scaling rates is significant at the 0.01 level for summer and at the 0.05 level for autumn, while it is not significant for winter and spring (not shown).

Latitudinally, similar to the whole-year case, the highest agreement between the seasonal results of the BM and POT methods is seen at mid and high latitudes in the Northern Hemisphere (Fig. 3). In the Southern Hemisphere, the extreme precipitation-temperature scaling rates derived from the two methods are consistent for latitudes poleward of 31°, 33°, 32° and 9° for winter, spring, summer and autumn,

respectively. For most of the latitudes in tropical and subtropical regions (Fig. S2), the scaling rates calculated by the methods are inconsistent for all seasons, and significant differences are observed for latitudes between 10° and 35° in the Northern Hemisphere and between 3° and 34° in the Southern Hemisphere (Fig. 3).

< Fig. 4 here please >

The magnitude of the change in extreme precipitation is not homogenous over the globe (Figs. 5 and S3), albeit the sign of the change is almost uniform. Based on the BM results, about 97.9%, 99.7%, 99.8% and 99.9% of the global land area show an increasing signal of extreme precipitation with return periods of 5, 15, 25 and 50 years, respectively. Less than 2% difference in the share of the global land area with increasing signals is seen for the POT-derived results, with about 96%, 98%, 98.8% and 99.3% of the land area for the respective return periods. For extreme precipitation of a 25-year return period, a 20-40% increase is expected for about 70% of the global land area, while the percentage for the POT method for the same range of increase is 65%. As one can see, the difference in the spatial distribution of the changes between the two methods increases with return period (Figs. 5 and S3). The difference is also more pronounced in arid regions in the tropics, subtropics and Australia (see Fig. S2 for global geographical and climatological classifications). In fact, larger extreme precipitation changes are derived from the BM method for these regions compared to the POT method.

< Fig. 5 here please >

The spatial distribution of extreme precipitation changes varies seasonally (Fig. 6). For both methods, the smallest share of the global land area with increasing signals of extreme precipitation is found for winter, e.g., accounting for 94% and 83% for a 25-year return period for the BM and POT methods, respectively. The percentage of the global land area with an increasing signal of 25-year extreme precipitation reaches

99% and 97% for autumn in the BM and POT methods, respectively. For all seasons in both methods, 20-40% increase is the dominant change class for extreme precipitation. The discrepancy in the spatial distribution of changes between the BM and POT methods differs among seasons. The difference in the global land area with increasing extreme precipitation signals between the methods peaks for winter at the rate of 11%.

< Fig. 6 here please >

The sensitivity analysis of the POT results to the choice of the threshold shows that the sensitivity of the quantiles to the threshold is canceled out in relative changes. The standard deviation across the extreme precipitation changes obtained using the thresholds of 1.65 to 5 peaks per year is smaller than 1.5% except for the arid regions in the Middle East and North Africa (MENA) and parts of Australia (Figs. 7 and S4). A comparison of the POT-derived extreme precipitation changes for the whole year with the 1.65 PPY and lowest annual maxima (LAM) thresholds shows no difference for return periods of higher than 15 years and a negligible difference for lower return periods (Fig. S5). At the seasonal scale, no difference in the change magnitudes between the 1.65 PPY and LAM thresholds is observed for all seasons except winter when the changes using the LAM threshold is overestimated up to 1.1%.

< Fig. 7 here please >

3.2. Flood change difference between the BM and POT methods

The global median BM- and POT-derived changes in flood intensity are generally of comparable magnitude, pointing to an increase in flood intensity (Fig. 8). The SNR is larger than one for both methods, but slightly larger for the POT results especially for shorter return periods (Fig. S6). The discrepancy between the BM- and POT-derived results, however, differs among impact models. Whereas the BM and POT results are almost identical for the H08 and LPJmL models, the BM-derived changes are slightly larger

for the CLM4.5 and WaterGAP2 models and the opposite is found for the MATSIRO model. Similar to extreme precipitation changes, the difference between the BM and POT results for the mentioned models gets larger with return period. Taking the ensemble median of each impact model into account, changes in flood intensity vary considerably between impact models (e.g., 9-29% for 25-year return period) (Fig. 8).

< Fig. 8 here please >

The flood intensity at the seasonal scale is projected to increase for all seasons, consistently between the BM and POT methods (Fig. 9). Globally, the seasonal pattern of the changes in flood intensity resembles the pattern of extreme precipitation changes, with the largest changes in autumn and winter (in order of 40%), the smallest ones in spring and summer (in order of 12%) (Fig. 9). The ensemble median changes for winter and autumn and the high return periods of summer are larger than the multi-model uncertainty ($\text{SNR} > 1$; Fig. S6), while the changes for spring are smaller than the uncertainty ($\text{SNR} < 1$; Fig. S6). The largest differences between the BM and POT results are found respectively in autumn and summer, which enlarge with return period. The differences are not, however, statistically significant in all seasons.

< Fig. 9 here please >

The change in flood intensity is not uniformly distributed over the globe (Figs. 10 and S7). Globally, based on the BM results, about 73%, 77%, 78% and 79% of the analyzed land fraction will experience an increase in flood intensity with return periods of 5, 15, 25 and 50 years, respectively. The results of the POT method show similar percentages of the land with a flood intensity increase: about 73%, 75%, 76% and 77% for the respective return periods. About 21%, 24%, 26% and 28% of the analyzed land grid cells show a flood intensity increase of $>40\%$ respectively for return periods of 5, 15, 25 and 50 years based on the BM results, with similar percentages of 22%, 24%, 25% and 26% for the mentioned return periods for the POT method. Based on the results of both methods, the intensity of floods increases across large parts

of the world, including south and southeast Asia, northeast Eurasia, South America and Africa (Figs. 10 and S7). In contract, a decrease is expected mainly for central, northern and eastern Europe, Anatolia, southern part of Argentina and Chile and central parts of the USA and Canada.

< Fig. 10 here please >

There is a seasonality in the share of the land area with an increasing signal of flood intensity (Fig. 11). In contrast to extreme precipitation, the difference in the land area with increasing flood signals between the BM and POT method tops out for summer at the rate of 3.6%. The intensity of flood increases in 71-86% of unmasked land cells in autumn and winter for different return periods. The percentage of the land area with an increasing signal falls to the range of 52-69% for spring and summer. For all seasons in both methods, while the increase exceeding 60% is the dominant change class for winter and autumn, 0-20% increase is the dominant change range for spring and summer.

< Fig. 11 here please >

The sensitivity analysis of the POT results to the choice of the threshold for flow extremes shows a larger sensitivity in the regions with the largest projected increase in flood intensity (Figs. 12 and S8). The flood changes are generally more sensitive the choice of the threshold than extreme precipitation changes. A comparison of the global median changes in flood intensity using the POT method with the thresholds of the LAM and 1.65 PPY indicates a negligible difference for the whole year and per season except summer when an underestimation of up to 2.3% with a LAM threshold is found (Fig. S9).

< Fig. 12 here please >

4. Discussion

4.1. General pattern of changes

Both BM and POT methods show an intensification of extreme precipitation over most of the global land area. It is well in accordance with previous studies (Fischer et al., 2014; Lehtonen & Jylhä, 2019). The magnitude of extreme precipitation changes increases with event extremity. More extreme precipitation events are mainly influenced by thermodynamic factors (Emori et al., 2005), but this influence is weakened by dynamic factors for less extreme events (Norris et al., 2019). The increase of extreme precipitation change magnitudes with event extremity is supported by earlier studies (Pendergrass, 2018; Hosseinzadehtalaei et al., 2019; Myhre et al., 2019; Li et al., 2019).

By assuming that thermodynamic processes and hence the availability of atmospheric moisture are controlling factors for extreme precipitation changes (Zhang et al., 2017), the extreme precipitation-temperature scaling rate was also investigated. The scaling rate shows a clear latitudinal pattern. The pattern is governed by dynamic and thermodynamic effects. In regions such as extratropical latitudes where the thermodynamic effects (changes in atmospheric moisture) are dominant (Chen et al., 2019), the extreme precipitation-temperature scaling rate is close to the C-C rate. Wherever the dynamic factors (changes in atmospheric circulations) come into play such as tropical and subtropical latitudes, the scaling rate is deviated from the C-C rate (Weller et al., 2019; Norris et al., 2020). The overall latitudinal pattern of the extreme precipitation-temperature scaling rate is in line with a previous study based on an empirical distribution (Tabari et al., 2019).

The scaling rate of extreme precipitation with temperature varies with season, as a result of seasonally-varying thermodynamic and dynamic contributions to extreme precipitation changes (Pfahl et al., 2017; Bador et al., 2018). The largest scaling rate (close to the C-C relationship) is found for winter and the smallest one for summer. A similar seasonal behavior of extreme precipitation changes with global warming has been found using an empirical distribution (Tabari et al., 2019).

The results show a low sensitivity of the POT-derived extreme precipitation changes to the choice of the threshold except in arid regions. The high sensitivity in arid regions is due to a smaller sample size of precipitation in these regions where a lower threshold may result in inclusion of low (non-extreme)

precipitation values and thereby a larger climate change signal difference (higher sensitivity) between high and low thresholds. A higher threshold is therefore suggested for arid regions compared with humid regions (Wang et al., 2019). Compared to extreme precipitation changes, the sensitivity of the POT results to the choice of the threshold is larger for flood changes. It can be attributed to a smaller population size of flows (after applying the independence criterion) compared to precipitation from which the samples of extremes are taken. The population size of flows is 730, while the population sizes of precipitation are 10800, 10950 and 10958 for the climate models considering 360-day, 365-day and normal calendars, respectively.

When looking at climate change signals for flood, both BM and POT methods project an increase in the intensity of floods across large parts of the world. The global pattern of the change in flood intensity is in agreement with earlier studies using different hydrological and climate models and scenarios (Hirabayashi et al., 2013; Dankers et al. 2014; Donnelly et al., 2017; Thober et al., 2018). The spatially-varying footprint of climate change on flood intensity stems from spatially-varying flood generating mechanism: whether it is intense rainfall or antecedent conditions or snowmelt/snow accumulation (Arnell & Gosling, 2016). The decrease in flood intensity is projected in regions where extreme precipitation, which is expected to intensify over most of the globe, is not the dominant flood generating factor. Dankers et al. (2014) concluded that decreases in flood magnitude occur particularly in areas with a snowmelt-dominated hydrograph in spring. Warming temperatures are decreasing the magnitude of snowmelt-related floods as a result of a shift in snowmelt timing (Cunderlik & Ouarda, 2009; Burn & Whitfield, 2016). More precipitation is expected to drop in the form of rain rather than snow in a warmer climate, resulting in smaller rain-generated flood peaks compared to snowmelt-generated peaks (Arnell & Gosling, 2016). High resolution simulations from the Weather Research and Forecasting (WRF) model over western North America, however, show a climate and elevation dependence of these changes; the precipitation shift from snowfall to rain may lead to more frequent rain-on-snow (ROS) events at higher elevations, while ROS frequency is expected to decrease in lower elevated, warmer areas (Musselman et al., 2018). In the regions where antecedence soil moisture plays the dominant role, the sign of flood intensity change is not also clear, due to the complex interacting influences of seasonal and annual precipitation, evaporation and extreme

precipitation (Ivancic and Shaw, 2015; Bennett et al., 2018; Wasko & Nathan, 2019). For example, across Europe, soil moisture excess is the most important flood driver, followed by snowmelt or rain-on-snow (Berghuijs et al., 2019).

Seasonally, the BM and POT methods also project an increase in flood intensity in all seasons. The change magnitude, however, varies with season, with the largest magnitudes in autumn and winter and the smallest ones in spring and summer. The seasonality in the flood intensity change results from seasonally-varying flood generating mechanisms (Parajka et al., 2010; Villarini, 2016). Global warming may shift the flood regime from one generated by spring snowmelts to more frequent, smaller floods dominated by extreme precipitation in winter (Arnell & Gosling, 2016).

4.2. Differences between the BM- and POT-derived results

A comparison between the climate change signals of extreme precipitation derived from the BM and POT methods show a larger difference for longer return periods, in agreement with previous studies (e.g., Nagy et al., 2017). The reliability of quantile estimates decreases with enlarging the extrapolation beyond the length of the record. The error of the estimated quantiles would be larger when a smaller set of samples is used, which is the case for the BM method (Ding et al., 2008).

In terms of the scaling rate of extreme precipitation-temperature, for the whole year case, the rate obtained from the POT method is smaller than that from the BM method. By analyzing the scaling rate per season, the opposite pattern is found, where the scaling rate derived from the POT method is slightly larger for all seasons. The scaling rate difference between the BM and POT methods also varies with season. It implies that the quantified response of extreme precipitation to a warming atmosphere may vary with the method chosen for the extraction of extreme events, thus influencing our understanding of the response of the hydrological cycle to warming.

The largest difference between the BM and POT methods is found in arid regions. The tropics and subtropics are the regions where the largest inconsistency between the results of the two methods is found. The difference is also more pronounced in arid regions in the tropics, subtropics and Australia. It is expected

as the sensitivity to the sample size of extremes is larger in arid regions which are mainly located in tropical and subtropical latitudes. The results generally suggest that the choice of the method for the extreme value analysis is more important in arid regions.

When examining the difference between the BM and POT methods for flood intensity changes, similar to extreme precipitation changes, there is a clear seasonal and spatial dependence of the difference between the BM- and POT-derived results. In addition, the results of the two methods also differ among impact models. The difference between GIMs can be attributed to their different representation of snowmelt and evaporation processes (Haddeland, et al., 2011; Dankers et al., 2014). Snow accumulation and melt are simulated in CLM4.5, H08 and MATSIRO models using the physically based energy balance method, whereas LPJmL and WaterGAP2 models use the conceptual degree-day method (Table 1). The energy balance method generally leads to lower snow water equivalent values than the degree-day method (Haddeland et al., 2011). There are also differences in the way each snow scheme is implemented: in the number of snow layers, snow albedo values, and the considered liquid water-holding capacity of snowpack in the energy balance method and in threshold temperature, degree-day factor and the process of melted snow in the degree-day method. For example, models with a single snow layer (e.g., H08) may produce abundant snowmelt sooner following the accumulation of enough energy (Do et al., 2020) or models (e.g., H08) that consider a percolation of rainfall through the snowpack directly into the soil and runoff generation and no retention in the snowpack tend to simulate higher runoff values (Haddeland et al., 2011).

The evapotranspiration scheme and the details of implementation also substantially vary among the GIMs: the Monin-Obukhov similarity theory is used in CLM4.5, the Priestley-Taylor method with fixed and two alpha coefficients depending on the grid cell aridity respectively in LPJmL and WaterGAP2, and the bulk formula in H08 and MATSIRO. The differences in the evapotranspiration scheme are particularly important in arid and semi-arid regions where the runoff ratio is low, and a minor difference in simulated evaporation causes a big relative difference in runoff (Haddeland et al., 2011). Similarly, the runoff ratio in summer is lower than in winter, owing to a higher evapotranspiration in summer and a higher contribution of summer rainfall to recharge soil moisture (Merz and Blöschl, 2009; Rodríguez-Blanco et al., 2012;

Julander and Clayton, 2018). Based on the results of this study, the snow scheme appears to have a more important role than the evapotranspiration scheme in the differences among the selected GIMs, as the GIM uncertainty in winter is far larger than that in the other seasons (Fig. S10).

4.3. Caveats

There are some key caveats to note regarding the comparison between BM and POT methods performed in this study. The extreme precipitation changes were derived from coarse GCM simulations which might be biased and affect the comparative analyses. The spatial resolution of GCMs restricts their ability to represent many mesoscale atmospheric processes leading to extreme precipitation episodes (Lehtonen et al., 2013) and in turn, to adequately simulate the intensity, frequency and duration of precipitation extremes (Mishra et al., 2014; Xu et al., 2019; Wu et al., 2020). A good example is convection which is parameterized in GCMs, resulting in deficiencies in the diurnal cycle of convection, occurrence of precipitation events, and extremes (Kendon et al., 2017). This deficiency leads to a large uncertainty in extreme precipitation projections, especially where and when local scale processes to produce extreme precipitation are dominant (Tabari et al., 2019), since GCM uncertainty is one of the main uncertainty sources (e.g., Kay et al., 2009; Hawkins and Sutton, 2011; Tabari et al., 2019; Xu et al., 2019). The extreme precipitation simulations of GCM can be improved by dynamical (Meredith et al., 2018) and statistical (Tabari et al., 2020) downscaling, which needs further investigation in the future.

The comparison between the BM and POT methods for flood changes was based on the present state of accuracy that can be achieved on a global scale. The GIMs used to simulate streamflow have a global scope which assume globally-consistent within-cell routing parameters and reproduce flow regimes in typical catchments. They may not, therefore, be able to accurately describe climatological and hydrological systems at a specific location (Arnell and Gosling, 2016). Furthermore, to define the independence between two discharge peaks for the POT method, the criterion of independence used in the present work is a 15-day window between the events. This window, however, varies with catchment size and flow regime, being larger for larger catchments and snow-dominated regimes.

Out of the large ensemble of the CMIP5 GCMs, four GCMs were used to force GIMs to simulate daily flow. Although the four GCMs were selected in a way to cover the uncertainty of the full ensemble of the CMIP5 GCMs, the uncertainty in projected flood changes using the BM and POT methods is expected to be underestimated. A further avenue for research is forcing GIMs with more GCMs for a better uncertainty quantification.

5. Conclusions

This study addresses the question of how much the choice between the BM and POT methods matters for climate change impact assessment on global flood and extreme precipitation. The differences between the BM and POT results vary with variable, event extremity, season and space. The difference is generally larger for extreme precipitation changes compared to flood intensity changes. For the magnitude of extreme precipitation changes and the extreme precipitation-temperature scaling rates, the difference between the BM and POT methods is statistically significant for longer return periods. The method choice is therefore important for the applications that need the changes in more extreme events. The difference in the magnitude of projected extreme precipitation changes gets larger when sample size is small which is the case seasonally for summer and spatially for arid regions. Spatially, the significant inconsistency between the results of the two methods is found in the tropics and subtropics which cover most of the world's arid regions, while the results are highly consistent for the extratropics particularly in the Northern Hemisphere. The simplification in the extreme value analysis at large scales would thus be an issue for data limited conditions. In terms of the direction of extreme precipitation changes, the difference between the BM and POT results differs among seasons and peaks in winter. Similar to extreme precipitation, the difference in the magnitude of flood intensity changes is large for summer, but not statistically significant. The difference in the direction of flood intensity changes tops out for summer in contract to winter for extreme precipitation.

This study highlights the extra uncertainty which may be overlooked if the choice of the methods for the extreme value analysis is not taken into account for climate change analysis especially on a seasonal

scale. While the BM method may be considered a more straightforward method for large-scale climate change impact assessments, the choice of the method must be made based on the specific application and the availability of resources. This is because the POT method is computationally more expensive than the BM method to ensure the independence of the extremes and the selection of an appropriate threshold. There must be a trade-off between the accuracy of the results and available computational resources and neither should be sacrificed for another. The information needed from the extreme value analysis for a particular application defines how to appropriately tune between accuracy and efficiency.

Acknowledgments

The CMIP5 and ISIMIP data providers are acknowledged. The CMIP5 and ISIMIP data are publicly available for download from the Earth System Grid Federation (ESGF) server (<https://esgf-index1.ceda.ac.uk>). This research has been supported by the Research Foundation - Flanders (FWO grant no. 12P3219N). I also thank the editor and seven anonymous reviewers for their constructive feedback.

References

- Allan, R., et al. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, doi: 10.1111/nyas.14337.
- Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 134(3), 387-401.
- Bador, M., Donat, M. G., Geoffroy, O., & Alexander, L. V. (2018). Assessing the robustness of future extreme precipitation intensification in the CMIP5 ensemble. *Journal of Climate*, 31(16), 6505-6525.
- Bennett, B., Leonard, M., Deng, Y., & Westra, S. (2018). An empirical investigation into the effect of antecedent precipitation on flood volume. *Journal of Hydrology*, 567, 435-445.
- Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, 55(6), 4582-4593.

- Bezak, N., Brilly, M. & Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Science Journal*, 59, 959–977.
- Bondeau, A. et al. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, 13, 679–706.
- Burn, D. H., & Whitfield, P. H. (2016). Changes in floods and flood regimes in Canada. *Canadian Water Resources Journal*, 41, 139–150.
- Calton, B., Schellekens, J., & Martinez-de la Torre, A. (2016). Water Resource Reanalysis v1: Data Access and Model Verification Results (Version v1.02). Zenodo (<https://doi.org/10.5281/zenodo.57760>).
- Chen, G., Norris, J., Neelin, J. D., Lu, J., Leung, L. R., & Sakaguchi, K. (2019). Thermodynamic and dynamic mechanisms for hydrological cycle intensification over the full probability distribution of precipitation events. *Journal of the Atmospheric Sciences*, 76(2), 497-516.
- Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. London: Springer.
- Cunderlik, J. M. & Ouarda, T. B. M. J. (2009). Trends in the timing and magnitude of floods in Canada. *Journal of Hydrology*, 375, 471–480.
- Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18, 257–271.
- Dankers, R., et al. (2014). First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences*, 111(9), 3257-3261.
- Dee, D. P. et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–97.
- Diaconescu, E. P., Gachon, P., & Laprise, R. (2015) On the remapping procedure of daily precipitation statistics and indices used in regional climate model evaluation. *Journal of Hydrometeorology*, 16(6), 2301-2310.

- Ding, Y., Cheng, B., & Jiang, Z. (2008). A newly-discovered GPD-GEV relationship together with comparing their models of extreme precipitation in summer. *Advances in Atmospheric Sciences*, 25(3), 507.
- Do, H. X., et al. (2020). Historical and future changes in global flood magnitude—evidence from a model-observation investigation. *Hydrology and Earth System Sciences*, 24(3), 1543-1564.
- Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A. & Maher, N. (2016). More extreme precipitation in the world’s dry and wet regions. *Nature Climate Change*, 6, 508–13.
- Dong, N., et al. (2019). Climate-induced hydrological impact mitigated by a high-density reservoir network in the Poyang Lake Basin. *Journal of Hydrology*, 579, 124148.
- Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P. and Ludwig, F. (2017). Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level. *Climatic Change*, 143, 13–26.
- Eilander, D., Couasnon, A., Ikeuchi, H., Muis, S., Yamazaki, D., Winsemius, H., & Ward, P. J. (2020). The effect of surge on riverine flood hazard and impact in deltas globally. *Environmental Research Letters*, doi: 10.1088/1748-9326/ab8ca6.
- Emori, S., & Brown, S. J. (2005). Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophysical Research Letters*, 32, L17706, doi:10.1029/2005GL023272.
- Fischer, E. M., & Knutti, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change*, 5(6), 560–564.
- Fischer, E. M., Sedláček, J., Hawkins, E., & Knutti, R. (2014). Models agree on forced response pattern of precipitation and temperature extremes. *Geophysical Research Letters*, 41(23), 8554-8562.
- Frieler, K., et al. (2017). Assessing the impacts of 1.5 degrees C global warming - simulation protocol of the inter-sectoral impact model intercomparison project (ISIMIP2b). *Geoscientific Model Development*, 10, 4321–45.

- Giuntoli, I., Vidal, J. P., Prudhomme, C., & Hannah, D. M. (2015). Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models. *Earth System Dynamics*, 6, 267–285.
- Haddeland I, et al. (2011). Multimodel estimate of the global terrestrial water balance: Setup and first results. *Journal of Hydrometeorology*, 12(5), 869–884.
- Hanasaki, N., et al. (2008). An integrated model for the assessment of global water resources—Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, 12, 1007–1025.
- Hawkins, E. & Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37, 407–18.
- Held, I. M. & Soden, B. J. (2006). Robust responses of the hydrologic cycle to global warming. *Journal of Climate*, 19, 5686–5699.
- Hempel, S., Frieler, K., Warszawski, L., Schewe, J. & Piontek, F. (2013). A trend-preserving bias correction—The ISI-MIP approach. *Earth System Dynamics*, 4, 219–236.
- Hershfield, D. M. (1961). Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years. US Weather Bureau Technical Paper N. 40, U.S. Dept. of Commerce, Washington, DC.
- Hirabayashi, Y., et al. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816–821.
- Hosseinzadehtalaei, P., Tabari, H., & Willems, P. (2019). Regionalization of anthropogenically forced changes in 3 hourly extreme precipitation over Europe. *Environmental Research Letters*, 14(12), 124031.
- Hosseinzadehtalaei, P., Tabari, H., & Willems, P. (2020a). Climate change impact on short-duration extreme precipitation and intensity–duration–frequency curves over Europe. *Journal of Hydrology*, 590, 125249.

- Hosseinzadehtalaei, P., Tabari, H., & Willems, P. (2020b). Satellite-based data driven quantification of pluvial floods over Europe under future climatic and socioeconomic changes. *Science of The Total Environment*, 137688.
- Huang, S. et al. (2017). Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. *Climatic Change*, 141(3), 381-397.
- Ivancic, T., & Shaw, S. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change*, 133, 681–693.
- Jaruskova, D., & Hanek, M. (2006). Peaks over threshold method in comparison with block-maxima method for estimating high return levels of several northern Moravia precipitation and discharge series. *Journal of Hydrology and Hydrometeorology*, 54(4), 309-319.
- Julander, R. P., & Clayton, J. A. (2018). Determining the proportion of streamflow that is generated by cold season processes versus summer rainfall in Utah, USA. *Journal of Hydrology: Regional Studies*, 17, 36-46.
- Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25, 1287–1304.
- Kay, A. L., Davies, H. N., Bell, V. A., & Jones, R. G. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic Change*, 92, 41-63.
- Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., ... & Wilkinson, J. M. (2017). Do convection-permitting regional climate models improve projections of future precipitation change?. *Bulletin of the American Meteorological Society*, 98(1), 79-93.
- Kharin, V. V., Zwiers, F. W., Zhang, X., & Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, 119, 345–357.
- Lang, M., Ouarda, T. B. M. J., & Bobee, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225, 103–117.
- Lange, S. (2019). Earth2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI). V. 1.1. GFZ Data Services. <https://doi.org/10.5880/pik.2019.004>.

- Lazoglou, G., Anagnostopoulou, C., Tolika, K., & Kolyva-Machera, F. (2019). A review of statistical methods to analyze extreme precipitation and temperature events in the Mediterranean region. *Theoretical and Applied Climatology*, 136, 99–117.
- Lehtonen, I., & Jylhä, K. (2019). Tendency towards a more extreme precipitation climate in the Coupled Model Intercomparison Project Phase 5 models. *Atmospheric Science Letters*, 20(5), e895.
- Lehtonen, I., Ruosteenoja, K., & Jylhä, K. (2014). Projected changes in European extreme precipitation indices on the basis of global and regional climate model ensembles. *International Journal of Climatology*, 34(4), 1208-1222.
- Li, C. et al. (2019). Larger increases in more extreme local precipitation events as climate warms. *Geophysical Research Letters*, 46(12), 6885-6891.
- Li, Z., Wang, Y., Zhao, W., Xu, Z., & Li, Z. (2016). Frequency analysis of high flow extremes in the Yingluoxia watershed in Northwest China. *Water*, 8(5), 215.
- Lombardo, F., Napolitano, F., Russo, F., & Koutsoyiannis, D. (2019). On the exact distribution of correlated extremes in hydrology. *Water Resources Research*, doi: 10.1029/2019WR025547.
- Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, 5(3), 250-254.
- Mangini, W., et al. (2018). Detection of trends in magnitude and frequency of flood peaks across Europe. *Hydrological Sciences Journal*, 63(4), 493-512.
- Maposa, D., & Cochran, J. J. (2017). Modelling extreme flood heights in the lower Limpopo River basin of Mozambique using a time-heterogeneous generalised Pareto distribution. *Statistics and Its Interface*, 10(1), 131-144.
- Marx, A., et al. (2018). Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 C. *Hydrology and Earth System Sciences*, 22(2), 1017-1032.
- McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4), 271–300.

- Meredith, E. P., Rust, H. W., & Ulbrich, U. (2018). A classification algorithm for selective dynamical downscaling of precipitation extremes. *Hydrology and Earth System Sciences*, 22, 4183-4200.
- Merz, R., & Blöschl, G. (2009). A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resources Research*, 45, W01405, doi:10.1029/2008WR007163.
- Mishra, V., Kumar, D., Ganguly, A. R., Sanjay, J., Mujumdar, M., Krishnan, R., & Shah, R. D. (2014). Reliability of regional and global climate models to simulate precipitation extremes over India. *Journal of Geophysical Research: Atmospheres*, 119(15), 9301-9323.
- Morbidelli, R. et al. (2020). The history of rainfall data time-resolution in a wide variety of geographical areas. *Journal of Hydrology*, 590, 125258.
- Morbidelli, R., et al. (2017). Effect of temporal aggregation on the estimate of annual maximum rainfall depths for the design of hydraulic infrastructure systems. *Journal of Hydrology*, 554, 710–720.
- Morbidelli, R., Saltalippi, C., Flammini, A., Corradini, C., Wilkinson, S. M., Fowler, H. J., (2018). Influence of temporal data aggregation on trend estimation for intense rainfall. *Advances in Water Resources*, 122, 304–316.
- Mueller Schmied, H., et al. (2016). Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrology and Earth System Sciences*, 20(7), 2877-2898.
- Musselman, K. N. et al. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change*, 8(9), 808-812.
- Myhre, G., et al. (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific Reports*, 9(1), 1-10.
- Nagy, B. K., Mohssen, M., & Hughey, K. F. D. (2017). Flood frequency analysis for a braided river catchment in New Zealand: Comparing annual maximum and partial duration series with varying record lengths. *Journal of Hydrology*, 547, 365-374.

- Nagy, B. K., Mohssen, M., & Hughey, K. F. D. (2017). Flood frequency analysis for a braided river catchment in New Zealand: Comparing annual maximum and partial duration series with varying record lengths. *Journal of Hydrology*, 547, 365-374.
- Norris, J., Chen, G., & Li, C. (2020). Dynamic amplification of subtropical extreme precipitation in a warming climate. *Geophysical Research Letters*, e2020GL087200.
- Norris, J., Chen, G., & Neelin, J. D. (2019). Thermodynamic versus dynamic controls on extreme precipitation in a warming climate from the Community Earth System Model Large Ensemble. *Journal of Climate*, 32(4), 1025-1045.
- Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, 49(1), 187-201.
- Parajka, J. et al. (2010). Seasonal characteristics of flood regimes across the Alpine–Carpathian range. *Journal of Hydrology*, 394(1-2), 78-89.
- Pendergrass, A. G. (2018). What precipitation is extreme? *Science*, 360(6393), 1072– 1073.
- Pfahl, S., O’Gorman, P. A., and Fischer, E. M. (2017). Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, 7, 423–427.
- Prudhomme, C. et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3262–3267.
- Rodríguez-Blanco, M. L., Taboada-Castro, M. M., & Taboada-Castro, M. T. (2012). Rainfall–runoff response and event-based runoff coefficients in a humid area (northwest Spain). *Hydrological Sciences Journal*, 57(3), 445-459.
- Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal*, 10(1), 33–60.
- Serinaldi, F., & Kilsby, C. G. (2014). Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resources Research*, 50(1), 336-352.

- Stackhouse, P. W. et al. (2011). 24.5 year SRB data set released GEWEX News 21 10–2
(http://www.gewex.org/gewex-content/files_mf/1432209318Feb2011.pdf).
- Tabari, H. (2019). Statistical analysis and stochastic modelling of hydrological extremes. *Water* 2019,
11(9), 1861.
- Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water
availability. *Scientific Reports*, 10, 13768.
- Tabari, H., & Willems, P. (2018). Seasonally varying footprint of climate change on precipitation in the
Middle East. *Scientific Reports*, 8(1), 1-10.
- Tabari, H., Buekenhout, D., & Willems, P. (2020). Comparison of statistical downscaling methods for
climate change impact analysis on drought. *Hydrology and Earth System Sciences Discussions*,
doi:10.5194/hess-2020-506.
- Tabari, H., Hosseinzadehtalaei, P., AghaKouchak, A., & Willems, P. (2019). Latitudinal heterogeneity and
hotspots of uncertainty in projected extreme precipitation. *Environmental Research Letters*, 14(12),
124032.
- Takata, K., Emori, S. & Watanabe, T. (2003). Development of the minimal advanced treatments of surface
interaction and runoff. *Global and Planetary Change*, 38, 209–222.
- Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design.
Bulletin of the American Meteorological Society, 93, 485–498.
- Thiery, W., Davin, E. L., Lawrence, D. M., Hirsch, A. L., Hauser, M. & Seneviratne, S. I. (2017). Present-
day irrigation mitigates heat extremes. *Journal of Geophysical Research: Atmospheres*, 122(3), 1403-
1422.
- Thober, S., et al. (2018). Multi-model ensemble projections of European river floods and high flows at 1.5,
2, and 3 degrees global warming. *Environmental Research Letters*, 13(1), 014003.
- Van der Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added value of large ensemble
simulations for assessing extreme river discharge in a 2 C warmer world. *Geophysical Research Letters*,
46(4), 2093-2102.

- Villarini, G. (2016). On the seasonality of flooding across the continental United States. *Advances in Water Resources*, 87, 80-91.
- Vörösmarty, C. J. et al. (2013). Extreme rainfall, vulnerability and risk: a continental-scale assessment for South America. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371, 20120408.
- Vrban, S., Wang, Y., McBean, E. A., Binns, A., & Gharabaghi, B. (2018). Evaluation of stormwater infrastructure design storms developed using partial duration and annual maximum series models. *Journal of Hydrologic Engineering*, 23(12), 04018051.
- Wang, C. H., & Holmes, J. D. (2020). Exceedance rate, exceedance probability, and the duality of GEV and GPD for extreme hazard analysis. *Natural Hazards*, 102, 1305–1321.
- Wang, Z., Wen, X., Lei, X., Tan, Q., Fang, G., & Zhang, X. (2020). Effects of different statistical distribution and threshold criteria in extreme precipitation modelling over global land areas. *International Journal of Climatology*, 40(3), 1838-1850.
- Ward, P. J., et al. (2017). A global framework for future costs and benefits of river-flood protection in urban areas. *Nature Climate Change*, 7(9), 642-646.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3228–3232.
- Wasko, C., & Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in flooding, *Journal of Hydrology*, 575, 432–441.
- Weedon, G. P. et al. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50, 7505–14.
- Weiss, L. L. (1964). Ratio of true to fixed-interval maximum rainfall. *Journal of the Hydraulics Division - Civil Engineering Database*, 90 (1), 77–82.

- Weller, E., Jakob, C., & Reeder, M. J. (2019). Understanding the dynamic contribution to future changes in tropical precipitation from low-level convergence lines. *Geophysical Research Letters*, 46(4), 2196-2203.
- Willner, S. N., Levermann, A., Zhao, F., & Frieler, K. (2018). Adaptation required to preserve future high-end river flood risk at present levels. *Science Advances*, 4(1), eaao1914.
- Wu, C. H., Yeh, P. J.-F., Chen, Y. Y., Hu, B. X., & Huang, G. R. (2020). Future precipitation-driven meteorological drought changes in the CMIP5 multi-model ensembles under 1.5 °C and 2 °C global warming. *Journal of Hydrometeorology*, 21(9), 2177–2196.
- Xu, K., Xu, B. B., Ju, J. L., Wu, C. H., Dai, H., & Hu, B. X. (2019). Projection and uncertainty of precipitation extremes in the CMIP5 multimodel ensembles over nine major basins in China. *Atmospheric Research*, 226: 122-137.
- Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., & Leung, R. (2018). Next-generation intensity-duration-frequency curves for hydrologic design in snow-dominated environments. *Water Resources Research*, 54(2), 1093-1108.
- Yoo, C., Jun, C., & Park, C. (2015). Effect of rainfall temporal distribution on the conversion factor to convert the fixed-interval into true-interval rainfall. *Journal of Hydrologic Engineering*, 20 (10), 04015018.
- Young, C. B., & McEnroe, B. M., (2003). Sampling adjustment factors for rainfall recorded at fixed time intervals. *Journal of Hydrologic Engineering*, 8 (5), 294–296.
- Zhang, X., Zwiers, F. W., Li, G., Wan, H., & Cannon, A. J. (2017). Complexity in estimating past and future extreme short-duration rainfall. *Nature Geoscience*, 10(4), 255-259.

Fig. 1. Comparison of changes in extreme precipitation intensity with return periods ranging between 2 and 50 years derived from the BM and POT methods for the whole year and per season. Lines and uncertainty bounds represent ensemble median and standard deviation, respectively. The difference between the results of the two methods is significant at the 0.05 level based on the Wilcoxon rank sum test for the return periods on the right side of the vertical red dotted line.

Fig. 2. Comparison of changes in extreme precipitation intensity with 5-, 15-, 25- and 50-year return periods as a function of global warming derived from the BM and POT methods for the whole year. The numbers on top of the plots show changes per degree global warming (slope of the fitted linear regression line), and the related coefficient of determination (R^2) and p-value (α). Markers represent the scaling rate in individual GCMs, and solid lines represent the linear regression fit through all points of an extreme value analysis method.

Fig. 3. Comparison of the latitudinal distribution of changes in extreme precipitation intensity with a 25-year return period as a function of global warming derived from the BM and POT methods for the whole year and per season. The scaling rate based on the slope of the linear regression line between the extreme precipitation change and the warming rate in individual GCMs is shown. The significant difference between the BM- and POT-derived scaling rates at the 0.05 level using the Wilcoxon rank sum test is highlighted in green color.

Fig. 4. Comparison of changes in extreme precipitation intensity with a 25-year return period as a function of global warming derived from the BM and POT methods per season. The numbers on top of the plots show changes per degree global warming (slope of the fitted linear regression line), and the related coefficient of determination (R^2) and p-value (α). Markers represent the scaling rate in individual GCMs, and solid lines represent the linear regression fit through all points of an extreme value analysis method.

Fig. 5. Spatial distribution of the ensemble median changes in extreme precipitation intensity with a 25-year return period derived from the BM and POT methods.

Fig. 6. Percentage of the global land area with different classes of extreme precipitation changes for the return periods ranging between 2 and 50 years derived from the BM and POT methods per season. Each circle in the concentric circles shows the results for a return period, with the return period increasing from the inner to the outer circle. The change is based on the ensemble median of 24 CMIP5 GCMs.

Fig. 7. Sensitivity analysis of the POT-derived changes in extreme precipitation intensity with a 25-year return period to the choice of the threshold. The sensitivity is characterized by standard deviation across the changes derived from different thresholds (1.65, 3, 4 and 5 PPY).

Fig. 8. Comparison of changes in flood intensity with return periods ranging between 2 and 50 years derived from the BM and POT methods using CLM4.5, H08, LPJmL, MATSIRO and WaterGAP2 impact models for the whole year. Lines and uncertainty bounds represent ensemble median and standard deviation for each impact model, respectively.

Fig. 9. Comparison of changes in flood intensity with return periods ranging between 2 and 50 years derived from the BM and POT methods per season. Lines and uncertainty bounds represent ensemble median and standard deviation across all impact models (20 experiments), respectively.

Fig. 10. Spatial distribution of the ensemble median (20 experiments) changes in flood intensity with a 25-year return period derived from the BM and POT methods. Grid cells with annual maxima of $<1 \text{ m}^3 \text{ s}^{-1}$ of the historical model period are screened out.

Fig. 11. Percentage of the unmasked global land area with different change classes for flood intensity for the return periods ranging between 2 and 50 years derived from the BM and POT methods per season. Each circle in the concentric circles shows the results for a return period, with the return period increasing from the inner to the outer circle. The change is based on the ensemble median of 20 experiments.

Fig. 12. Sensitivity analysis of the POT-derived changes in flood intensity with a 25-year return period to the choice of the threshold. The sensitivity is characterized by standard deviation across the changes derived from different thresholds (1.65, 3, 4 and 5 PPY). Grid cells with annual maxima of $<1 \text{ m}^3 \text{ s}^{-1}$ of the historical model period are screened out.

Table 1. Hydrological processes represented in the GIMs used in this study.

Model name	Model type	Climate variables	ET scheme	Snow scheme	Runoff scheme
CLM4.5	Global land surface model	ta, huss, sfcWind, rlds, rsds, pr	Monin-Obukhov similarity theory	Energy balance/multi-layer model	Saturation excess/fractional area
H08	Global hydrological models	tas, rlds, rsds, prsn, ps, pr	Bulk formula	Energy balance	Saturation excess/non-linear
LPJmL	Dynamic global vegetation model	tas, rsds, pr	Priestley-Taylor	Degree-day	Saturation excess
MATSIRO	Global hydrological models	ta, huss, sfcWind, tasmax, tas, tasmin, rlds, rsds, prsn, ps, pr	Bulk formula	Energy balance	Infiltration and saturation excess/groundwater
WaterGAP2	Global hydrological models	tas, rlds, rsds, pr	Priestley Taylor with two alpha coefficients contingent on grid cell aridity	Degree-day	Saturation excess/beta function

ET: evapotranspiration; ta: air temperature; tas: daily mean temperature; tasmax: daily maximum temperature; tasmin: daily minimum temperature; huss: near-surface specific humidity; sfcWind: near-surface wind speed; rlds: longwave downwelling radiation; rsds: shortwave downwelling radiation; pr: total precipitation; prsn: snowfall; ps: surface air pressure.

Supporting Information for

**Extreme value analysis dilemma for climate change impact
assessment on global flood and extreme precipitation**

Hossein Tabari

KU Leuven, Department of Civil Engineering, Belgium (hossein.tabari@kuleuven.be)

Contents of this file

Figs. S1 to S10

Fig. S1. Comparison of signal-to-noise ratio (SNR) for the change in extreme precipitation intensity for the whole year and per season derived from the BM and POT methods for return periods ranging from 2 to 50 years.

Fig. S2. Global climate classification based on the aridity index (AI). AI is calculated as the ratio between potential evapotranspiration (PET) and precipitation (P): $AI = PET/P$. The grid cells with $AI < 1$ are classified as humid, $1 \leq AI < 2$ as semi-humid, $2 \leq AI < 5$ as semi-arid and $AI \geq 5$ as arid. PET and P data were obtained from the Climate Research Unit of the University of East Anglia (CRUTSv4.01). Geographical zones are shown on the map. The Southern Hemisphere high-latitude land mass ($> 60^\circ S$) which includes the southern polar region is screened out in this study.

Fig. S3. Spatial distribution of the ensemble median changes in extreme precipitation intensity with 5-, 15- and 50-year return periods derived from the BM and POT methods.

Fig. S4. Sensitivity analysis of the POT-derived changes in extreme precipitation intensity with 5-, 15- and 50-year return periods to the choice of the threshold. The sensitivity is characterized by standard deviation across the changes derived from different thresholds (1.65, 3, 4 and 5 PPY).

Fig. S5. Comparison of the changes in extreme precipitation intensity for the whole year and per season using the POT method with the thresholds of lowest annual maxima (LAM) and 1.65 PPY for return periods ranging from 2 to 50 years.

Fig. S6. Comparison of signal-to-noise ratio (SNR) for the change in flood intensity for the whole year and per season derived from the BM and POT methods for return periods ranging from 2 to 50 years.

Fig. S7. Spatial distribution of the ensemble median changes in flood intensity with 5-, 15- and 50-year return periods derived from the BM and POT methods. Grid cells with annual maxima of $<1 \text{ m}^3 \text{ s}^{-1}$ of the historical model period are screened out.

Fig. S8. Sensitivity analysis of the POT-derived changes in flood intensity with 5-, 15- and 50-year return periods to the choice of the threshold. The sensitivity is characterized by standard deviation across the changes derived from different thresholds (1.65, 3, 4 and 5 PPY). Grid cells with annual maxima of $<1 \text{ m}^3 \text{ s}^{-1}$ of the historical model period are screened out.

Fig. S9. Comparison of the changes in flood intensity for the whole year and per season using the POT method with the thresholds of lowest annual maxima (LAM) and 1.65 PPY, for return periods ranging from 2 to 50 years.

Fig. S10. GIM uncertainty of changes in flood intensity for different seasons. The GIM uncertainty is computed as the standard deviation of the changes across GIMs after taking the median of changes over GCMs and extreme value analysis methods.