Analysis of high streamflow extremes in climate change studies: How do we calibrate hydrological models?

Bruno Majone[1], Diego Avesani[1], Patrick Zulian[1], Aldo Fiori[2], Alberto Bellin[1]

[1]Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, I-38123, Italy
[2] Department of Engineering, Roma Tre University, Roma, I- 00154, Italy

Correspondence to: Bruno Majone (bruno.majone@unitn.it)

Abstract. Climate change impact studies on hydrological extremes often rely on hydrological models with parameters inferred through calibration procedures using observed meteorological data as input forcing. We show that this procedure can lead to a biased evaluation of the probability distribution of high streamflow extremes when climate models are used. As an alternative approach, we introduce a methodology, coined Hydrological Calibration of eXtremes (HyCoX), in which the calibration of the hydrological model, as driven by climate models' outputs, is carried out by maximizing the probability that the modelled and observed high streamflow extremes belong to the same statistical population. The application to the Adige river catchment (southeastern Alps, Italy) by means of HYPERstreamHS, a distributed hydrological model, showed that this procedure preserves statistical coherence and produces reliable quantiles of the annual maximum streamflow to be used in assessment studies.

Key Points/Highlights:

- A methodology for devising reliable extreme high streamflow scenarios from climate change model simulations
- Accurate reproduction of observed ECDF of annual streamflow maximum
- Preservation of statistical coherence between observed and simulated ECDFs of annual streamflow maximum

Keywords: Goal-oriented calibration; high streamflow extremes, Climate change; statistical coherence; hydrological modelling

1 Introduction

The recognition that an altered climate may severely impact water availability and exacerbate floods and droughts, led in the past decades to a flourish of climate change impact assessment studies. Several studies investigated the likely impact of climate change on hydrology through hydrological modelling performed with meteorological forcing obtained from an ensemble of projections from multiple climate models under different greenhouse gas emissions scenarios [e.g., Kundzewicz et al., 2007; Todd et al., 2010, Wilby and Harris, 2006 for a comprehensive review]. A wealth of studies focused on long-term annual

and/or seasonal changes in hydrological variables such as runoff, streamflow, snow melt and soil moisture [e.g., Chiew et al. 2009; Majone et al., 2012; Buytaert and De Bièvre, 2012]. Much less studies addressed projected changes of hydrological extremes, i.e. floods and droughts, though they are expected to exert profound and dramatic impacts on agriculture, economy, human health, energy and many other water-related sectors [e.g., Arnell 2011; Taye et al. 2011; Bouwer, 2013; Thornton et al., 2014].

The role of hydrological calibration and the way to perform it in climate change impact studies has been much debated in the hydrological community [e.g. Peel and Blöschl, 2011; Muñoz et al., 2013; Montanari et al., 2013; Thirel et al., 2014]. According to the most used approach the hydrological model is first calibrated against the observed streamflow by using observed meteorological data as input. The calibrated hydrological model is then run with the climate models as input to assess the projected changes of selected indicators, including those related to extremes [e.g. flow quantiles, see Ngongondo et al., 2013; Aich et al., 2016; Pechlivanidis et al., 2017; Vetter et al., 2017; Hattermann et al. 2018]. The drawbacks of such an approach are, however, twofold: i) optimality in the reproduction of the time series of observed stream flow does not automatically imply optimality in the reproduction of extremes; and ii) because of epistemic uncertainty, a model calibrated with a given set of observations may respond differently when fed with projections obtained from climate change scenarios. Concerning this latter aspect, some studies evidenced that the calibrated model parameters depend on the climatic characteristics of the input forcing used for the calibration of the hydrological model [e.g., Vaze et al., 2010; Laiti et al., 2018]. Although recognized, this additional source of uncertainty is mostly ignored in climate change impact studies.

Several studies suggested that observed streamflow extremes provide valuable information about the hydrological behaviour of investigated catchments [Grubbs, 1969; Laio et al., 2010]. Similarly, Perrin et al. [2007] and Seibert and Beven [2009] concluded that a limited number of streamflow extremes encapsulate a significant amount of information that may be useful for hydrological model calibration. Beven and Westerberg [2011] suggested also that, when dealing with extremes, including the entire time series might not be informative. This occurs, for instance, when streamflow extremes belong to a different population than ordinary flows [e.g., Calenda et al, 2009], such that the latter does not provide useful information for inferring the former. Hence, quantifying the influence of such extreme events on model calibration is still a challenge in hydrological studies [Brigode et al., 2015], such as quantifying the uncertainty associated with these estimates [Honti et al., 2014].

To overcome the aforementioned limitations, we propose an innovative methodology in which the calibration of a hydrological model, as driven by climate models, is conducted by maximizing the probability that the modelled and observed streamflow extremes belong to the same population within the reference period. While the approach is exemplified in this work for high streamflows (given the broad interest in the topic), it can be applied to low flows as well (e.g., for droughts assessment). The methodology, coined here as Hydrological Calibration of eXtremes (HyCoX), targets specifically climate change impact assessment studies and relies on the use of the two-sample Kolmogorov-Smirnov statistic [Smirnov, 1939] as an efficiency metric during the calibration procedure. We emphasize that the suggested approach is by definition "*goal-oriented*", as recently discussed in Fiori et al. [2016], Guthke [2017] and Laiti et al., [2018].

Studies adopting the two-sample Kolmogorov-Smirnov test to evaluate whether simulated hydrological variables are distributed according to a given probability distribution [e.g., Kleinen and Petschel-Held, 2007] are relatively common in the literature. This statistical test was also used to detect changes in hydrological variables [e.g., Wang et al., 2008], and to verify if calibrated parameters of a hydrological model belong to a given probability distribution [e.g., Wu et al., 2017; Wang and Solomatine, 2019]. This notwithstanding, we are not aware of existing studies adopting this statistical test in the context of hydrological model calibration oriented to the reproduction of extremes.

The main objective of the present work is therefore twofold. From one side, we introduce the HyCoX framework and assess its capability to reproduce observed high streamflow extremes using climate models, applied to the same time frame of the observational data, as input. On the other, the strength of the proposed methodology is tested against the commonly adopted procedure of calibrating the model by using observational data.

The paper is organized as follows: Sect. 2 presents the hydrological modelling framework, the calibration metrics and the adopted statistical test; a description of the study area, the climate change projections, the observational hydro-meteorological datasets and the simulations set-up are summarized in Sect. 3. The main findings are presented and discussed in Sect. 4, whereas conclusions are finally drawn in Sect. 5.

2 Methods

2.1 Hydrological modelling

Hydrological simulations were performed at the daily time scale with the HYPERstreamHS model [Avesani et al., 2021; Laiti et al., 2018; Larsen et al., 2021] which couples the HYPERstream routing scheme, recently proposed by Piccolroaz et al., [2016], with a continuous module for surface and subsurface flow generation. HYPERstream routing scheme is specifically designed to facilitate coupling with climate models and, in general, with gridded climate datasets. HYPERstream can share the same computational grid as that of any overlaying product providing the meteorological forcing, still preserving geomorphological dispersion of the river network [Rinaldo et al., 1991] irrespective of the grid resolution. This "perfect upscaling" [cf. Piccolroaz et al., 2016] is obtained by the application of suitable transfer functions derived from a high-resolution Digital Elevation Model (DEM) of the study area. Separation between surface flow and infiltration was obtained by using the continuous SCS-CN model [Michel et al., 2005], which receives as input the total precipitation given by the sum of rainfall and snow melting, the latter being evaluated by the degree-day model coupled with mass balance, which includes snow accumulation [Rango and Martinec, 1995]. The infiltrating water enters into a non-linear bucket mimicking soil moisture dynamics [Majone et al., 2010] with evapotranspiration, which is computed by the Hargreaves and Samani [1982] model, and deep infiltration as output fluxes. Finally, deep infiltration enters a linear bucket used to represent return flow. The surface and subsurface flow generation module was already successfully applied in previous studies conducted in Alpine catchments [Piccolroaz et al., 2015; Bellin et al., 2016; Galletti et al., 2021]. The model requires a total of 12 parameters, which are assumed spatially uniform, but uncertain and to be determined through calibration. Spatial heterogeneity of evapotranspiration,

infiltration and runoff generation was accounted for by computing for each macrocell all relevant properties (e.g., maximum infiltration capacity, average elevation, soil type, crop coefficient etc.) based on available DEM and land-use/land-cover spatial maps. The list of the 12 parameters with their units together with a short description and range of variation is presented in Table 1. A detailed description of the hydrological model can be found in Laiti et al. [2018] and Avesani et al. [2021].

Table 1: List of model parameters with their units and parameters range.

Model Component	Parameters	Description	Unit	Parameter range
Snow model	T_{snow}	temperature threshold for snow precipitation	$°C$	-2 – 6
	T_{melt}	temperature threshold for snow melting	$°C$	-2 – 6
	c_{melt}	snow melting factor	$mm\ °C^{-1}d^{-1}$	0 – 10
Continuous soil-moisture accounting SCS-CN based model	c_s	parameter of the rainfall excess model	-	0.1 – 10
	c_a	parameter of the rainfall excess model	-	0.01 – 1
	q_{ref}	parameter of the nonlinear bucket	$mm\ s^{-1}$	10^{-7} – 10^{-3}
	μ	parameter of the nonlinear bucket	mm	0.5 – 300
	c_{fc}	coefficient for field capacity	-	0 – 1
	c_r	coefficient for residual soil moisture	-	0 – 0.25
Base-flow model	k	mean residence time for baseflow linear reservoir	day	200 – 1000
	α	partition coefficient for leakage flux	-	0 – 1
HYPERstream routing	v	stream velocity	$m\ s^{-1}$	0.2 – 4.0

2.2 Hydrological model calibration

The HYPERstreamHS hydrological model was calibrated against streamflow observations using as meteorological forcing both the observational dataset ADIGE (see Sect. 3.2) and the output of three climate models under two emission scenarios. A short description of these datasets is provided in Sect. 3.3. Parameters were inferred by optimizing three efficiency metrics using the Particle Swarming Optimization (PSO) algorithm [Kennedy and Eberhart, 1995]. PSO is an iterative algorithm belonging to the swarm intelligence category, which is based on the exploration of the space of parameters by a set of particles,

called bees. Particles' positions were first randomly initialized and then iteratively updated in the search for the optimal solution, with the location updating procedure considering the memory of all locations visited by the whole collection (swarm) of particles.

The first metric is the classic Nash-Sutcliffe model efficiency [Nash and Sutcliffe, 1970], which is widely used in hydrological applications:

$$NSE = 1 - \frac{\sum_{i=1}^{m}\left(Q_{s,i}(\boldsymbol{\theta}) - Q_{o,i}\right)^2}{\sum_{i=1}^{m}\left(Q_{o,i} - \bar{Q}_o\right)^2}, \quad (1)$$

where m is the total number of daily time steps, $Q_{s,i}(\boldsymbol{\theta})$ and $Q_{o,i}$ are the simulated (s) and observed (o) streamflow at time step i, respectively, $\bar{Q}_o$ is the mean of the observed values and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]$ are the $q=12$ model parameters. Since this metric considers the chronological time series of simulated and observed daily streamflow, it was applied only when the observational dataset ADIGE was used as meteorological input.

The second efficiency metric (R_{FDC}) is an adaptation of the objective function proposed in Westerberg et al. [2011] to obtain a good match between simulated, $\hat{Q}_{s,(i)}(\boldsymbol{\theta})$, and observed, $\hat{Q}_{o,(i)}$, flow duration curves (FDCs, i.e., the ranked streamflow values in descending order):

$$R_{FDC} = 1 - \frac{\sum_{i=1}^{n_{EP}}\left|\hat{Q}_{s,(i)}^{EP}(\boldsymbol{\theta}) - \hat{Q}_{o,(i)}^{EP}\right|}{\sum_{i=1}^{n_{EP}}\left|\hat{Q}_{o,(i)}^{EP} - \bar{Q}_o\right|}, \quad (2)$$

where $\hat{Q}_{s,(i)}^{EP}(\boldsymbol{\theta})$ and $\hat{Q}_{o,(i)}^{EP}$ are the simulated and observed streamflow values at the n_{EP} evaluation points (EPs) in which the flow duration curves are partitioned and $\bar{Q}_o$ is the mean of the observed time series. According to this metric, $R_{FDC} = 1$ when the two flow duration curves coincide (i.e., they are the same at all the EPs). Given that the flow duration curve is insensitive to chronologic sequence, R_{FDC} has been used as objective function for streamflow maxima obtained with both climate models and the observational dataset ADIGE. Furthermore, following Westerberg et al. [2011], the so-called volume method was employed in which EPs were identified as the upper boundary of the elements with the same ara V/n_{EP} below the FDC, where V is the total streamflow volume, i.e. the total area below the FDC. Given the same number of EPs, we remark that the procedure is performed independently for observed and simulated FDCs and it is indeed possible that the total volume V under the curves and the water volume V/n_{EP} of the n_{EP} intervals differ between observations and simulations. The water volume pertaining to each interval as well as the total water volume of the flow duration curve are computed by using the right Riemann sum procedure [Protter and Morrey, 1977]. In the computations we used $n_{EP} = 50$, which has been shown sufficient to obtain convergence of the statistic (2) irrespective of the integration scheme [Vogel and Fennessey, 1994].

The third efficiency metric (KS) is the two-sample Kolmogorov-Smirnov statistic (D_n):

$$KS = D_n = \max_{i\in[1,n]}\left|F_s\left(Q_{s,(i)}^{M}(\boldsymbol{\theta})\right) - F_o\left(Q_{o,(i)}^{M}\right)\right|, \quad (3)$$

where F_s and F_o are the Empirical Cumulative Distribution Functions (ECDFs) of the simulated, $Q_{s,(i)}^{M}(\boldsymbol{\theta})$, and observed, $Q_{o,(i)}^{M}$, samples of daily average annual streamflow maxima ranked in increasing order, respectively, and n is the number of

years considered in the simulation (29 in the present work, one for each year of the investigated period excluding the first two, see Sect. 3.4). Before ranking in increasing order, samples of annual streamflow maxima are extracted from the chronological daily time series of observed and simulated streamflow, respectively. Afterwards, ECDFs of the simulated and observed samples of annual maxima are computed according to the classic Weibull formulation [Weibull, 1939]:

$$F_j\left(Q^M_{j,(i)}\right) = \frac{i}{n+1}, \;\; j = o, s, \;\; i \in [1, n]. \tag{4}$$

This metric, which is at the core of the proposed approach, aims to maximize the probability that the modelled and observed samples of high streamflows extremes belong to the same population. In other words, among all possible sets of model parameters, we consider the one leading to the smallest maximum absolute distance D_n between simulated and observed ECDFs of daily annual streamflow maxima. Since KS is not sensitive to the temporal sequence of observed and simulated streamflows, similar to R_{FDC}, it has been applied to climate projections in addition to the simulations with the observational dataset ADIGE.

2.3 Evaluation of statistical coherence

After calibration, statistical coherence between the observed and simulated samples of high streamflow extremes was evaluated employing the two-sample Kolmogorov-Smirnov test [Smirnov, 1939], applied under the null hypothesis that the two samples are drawn from the same underlying distribution. In the two-tail application of interest here the test's statistic, D_n is given by Eq. (3). The closer D_n is to 0 the more likely it is that the two samples are drawn from the same population. In addition, the two-sample Kolmogorov-Smirnov test returns a p-value (p) corresponding to the computed D_n statistic [Conover, 1999]. The larger the p-value the stronger the evidence in favour of the null hypothesis, i.e., that the samples are drawn from the same distribution.

In this study, the p-value has been used as a measure of the statistical coherence between samples of simulated and observed high streamflow extremes. Furthermore, this evaluation step has been performed a-posteriori for each simulation experiment described in Sect. 3.4.

2.4 Probability distribution computation and confidence intervals

The theoretical probability distributions of simulated and observed annual streamflow maxima were obtained by fitting the Extreme Value Type I (Gumbel) [Gumbel, 1941] distribution, $P(Q \leq q) = exp\left[-exp[-\beta(q-u)]\right]$, with the Maximum Likelihood Method (MLE) [Hosking, 1985] to the respective samples. The Pearson's chi-squared test [Pearson, 1990] with a confidence level $\alpha_s = 0.05$ was then applied to validate the parameters β and u provided by the MLE. Extrapolation of high quantiles (i.e., estimation of quantiles for a return period larger than the available number of observation and simulation years) of observed and simulated annual streamflow maxima was then performed for all the simulation experiments described in Sect.3.4.

Confidence intervals of observed streamflow ECDF were computed using parametric bootstrap [Efron, 1982] under the assumption that the quantity of interest was distributed according to the above parametric Gumbel probability distribution. In particular, 90% confidence band was estimated by using 10000 uniform random samplings from the underlying inferred distribution.

3 Study area, hydro-climatic datasets and simulations set-up

3.1 Study area

To exemplify the application of the methodology the upper part of the Adige river basin (Italy), located in the south-eastern Alpine region (see Figure 1), at the gauging station of Trento (11° 06' 54.8" E, 46° 04' 13" N, drainage area of about 9850 km^2) was selected as a case study. The Adige river originates at the Resia Pass (close to the Alpine divide) and ends its course after 410 km in the northern Adriatic Sea. It is a typical Alpine river basin, with terrain elevations ranging from 185 m a.s.l. at Trento to 3500 m a.s.l. at the Italian-Austrian border. The morphology is characterized by deep valleys and high mountain crests.

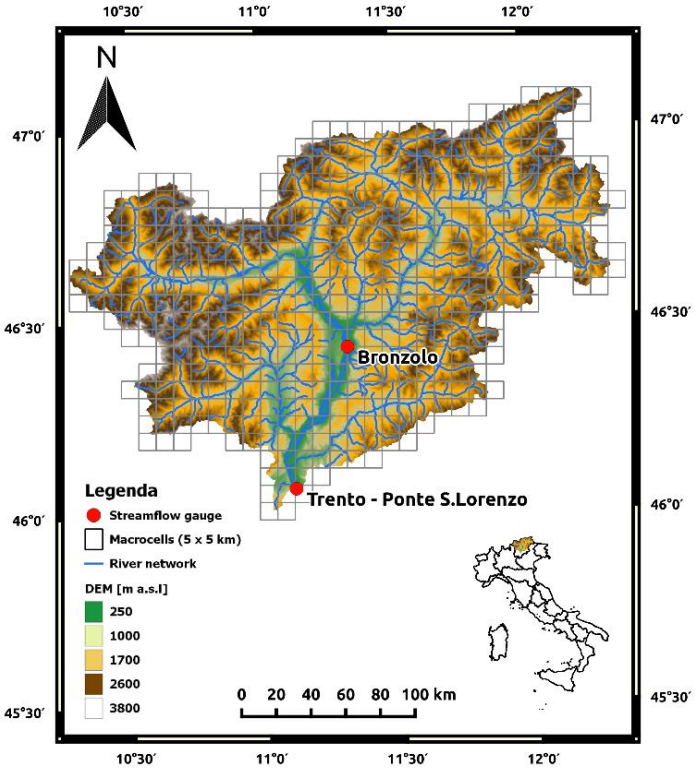


Figure 1: Map of the Adige river basin, with the computational grid cells ("macrocells") superimposed on the Digital Elevation Model (DEM) and the river network. The streamflow gauging stations used in the study are marked with red dots. The inset shows the location of the Adige river basin within the Italian territory.

The climate of the river basin is characterized by relatively dry and cold winters followed by humid summers and autumns. Streamflow is minimum in winter, when precipitation falls as snow over most of the river basin, and shows two maxima: one occurring early in summer, due to snowmelt, and the other in autumn, triggered by intense cyclonic storms. The average annual precipitation ranges from 500 mm in the North-West to 1600 mm in the southern part of the basin [Lutz et al., 2016; Diamantini et al., 2018; Laiti et al., 2018]. Projected decrease of snowfall in winter and anticipation of earlier snow-melting, essentially due to rising temperatures associated with global warming [Gobiet et al., 2014; Gampe et al., 2016], will likely affect the Adige streamflow regime by the second half of the 21st century [Bard et al., 2015; Majone et al., 2016]. This may have relevant consequences on water resources and hydropower production, which is particularly relevant in this region of the Alps [Zolezzi et al., 2009; Bellin et al., 2016; Majone et al., 2016; Avesani et al., 2022]. See also Chiogna et al. [2016] for a comprehensive review of the hydrological stressors acting in the Adige basin, as well as its ecological status.

3.2 Observational datasets

The regional dataset ADIGE developed by Mallucci et al., [2019] by using the meteorological stations within the catchment and in the nearby Austrian territory bounding the catchment from the north, was used as observational precipitation and temperature dataset within the time window 1950-2010. ADIGE was selected since it is the most accurate gridded meteorological dataset of the investigated river basin (as shown in the recent paper by Laiti et al., 2018). Meteorological data at the selected stations were provided by the Austrian Zentralanstalt für Meteorologie und Geodynamik (www.zamg.ac.at) and the meteorological offices of the Autonomous Provinces of Trento (www.meteotrentino.it) and Bolzano (www.provincia.bz.it/meteo). The time series were interpolated over a 1-km grid at a daily time step using the kriging with external drift algorithm [Goovaerts, 1997; Journel and Rossi, 1989], with an exponential semivariogram and by using the 16 closest neighbouring stations in the linear combination providing the estimate. The optimal spatial distribution model was selected by Mallucci et al. [2019] according to the leave-one-out cross-validation procedure, applied to both ordinary kriging and kriging with external drift algorithms. Several semi-variogram models (i.e., Gaussian, spherical and exponential models) and different numbers of neighbouring stations (namely 8, 16 and 32 stations) were tested and the model providing the minimum average absolute error of daily estimates was identified. As described in Mallucci et al. [2019] the optimal semivariogram model was the exponential one, providing an average absolute error of the daily estimates of about 1.32 mm for precipitation and 0.02°C for temperature, both comparable with the error estimates provided by widely used datasets available for the Alpine region such as APGD [Isotta et al., 2014]. Daily streamflow at the Ponte San Lorenzo in Trento and Bronzolo gauging stations (see Figure 1) were provided by the Hydrological Offices of the Autonomous Province of Trento (www.floods.it) and Bolzano (http://www.provincia.bz.it/hydro).

3.3 Climate change projections

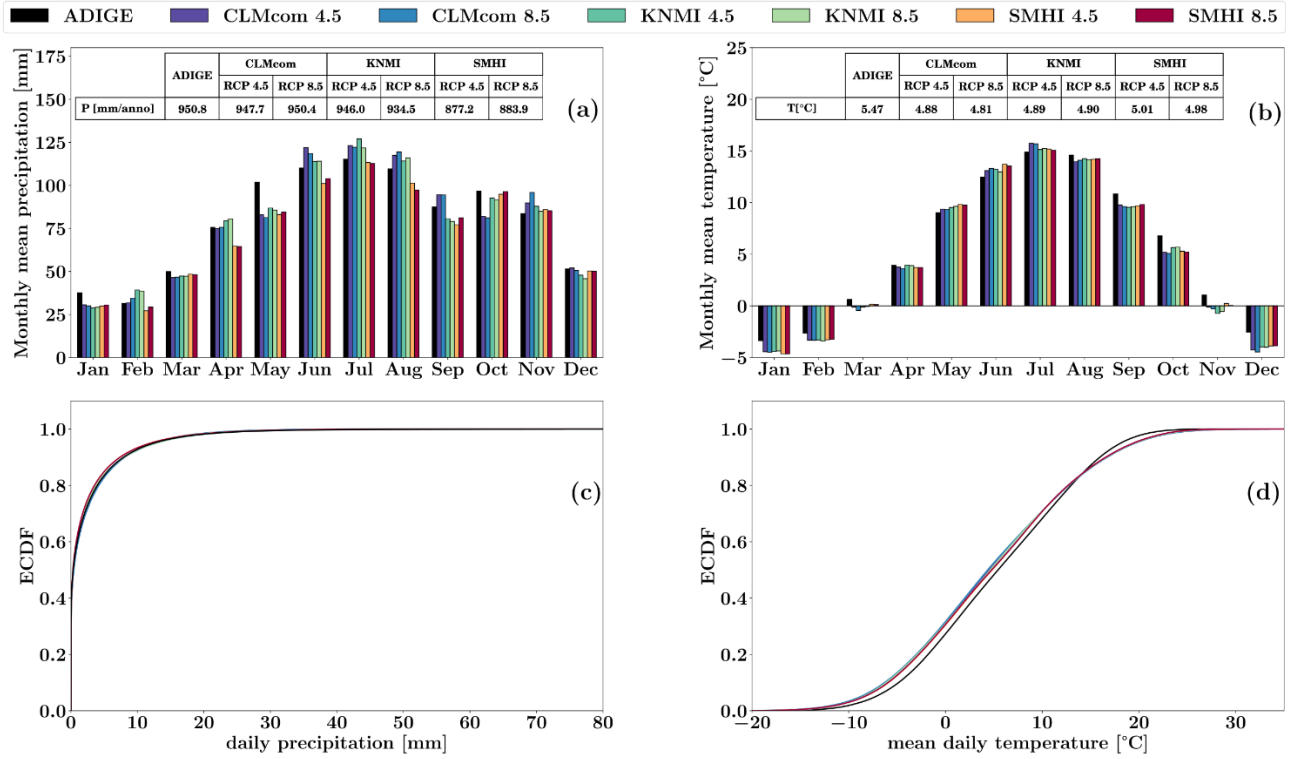
Climate projections used in the present work were derived from the combination of General Circulation Models (GCMs) and Regional Climate Models (RCMs) available from the EURO-CORDEX initiative under 4.5 and 8.5 Representative

Concentration Pathways (RCP4.5 and RCP8.5), at a spatial resolution of about 12 km [EUR-11, http://www.eurocordex.net/, Jacob et al., 2014]. To reduce the computational burden of the hydrological modelling experiments, we adopted the model sub-selection proposed by Vrzel et al. [2019] who applied a hierarchical clustering approach [Wilcke and Bärring, 2016] in selected European river basins (including the Adige) to reduce the number of available Climate Model (CM) simulations (i.e., GCM-RCM combinations) while preserving the variability of the ensemble of climate change signals. In particular, model reduction involved 5 steps: 1) identification of the meteorological variables; 2) transformation of variables into orthogonal and therewith uncorrelated variables using singular vector decomposition; 3) identification of the optimum number of clusters; 4) hierarchical clustering to group the simulations; and finally, 5) selection of the simulations closest to the group's mean as representative. This procedure led to the selection of the three GCM-RCM combinations (out of the 12 available), here referred to as CLMcom, KNMI and SMHI (see Table 2).

These three GCM-RCM combinations provide projections of likely future climate changes for the mid-term horizon 2040-2070, with the time window 1980-2010 selected as the period of reference. The projected climate change meteorological signals in the Adige are discussed in Gampe et al. [2016]. Both RCP4.5 and RCP8.5 emission scenarios are available for all the combinations, thereby leading to a total of six CMs which are investigated in the present study (see Table 2). Since GCMs/RCMs combinations are prone to model biases, especially in complex terrain [Kotlarski et al., 2014], bias-correction is needed to accurately reproduce historical meteorological forcing during the reference period. In the present work, we rely on products retrieved from EURO-CORDEX, which are available bias-corrected by the distribution-based scaling approach [DBS, Yang et al., 2010] using as observations the MESAN gridded reanalysis datasets of daily precipitation and temperature [Landelius et al., 2016]. Basin-averaged monthly mean precipitation and temperature of the six CMs are presented in Figure 2 with reference to the period 1980-2010 together with those of the ADIGE dataset. Notice that CMs slightly differ between the two RCPs as a consequence of: i) the bias correction method adopted, which matches observed and simulated frequency distributions rather than the observations; and ii) the correction performed with reference to the period 1989-2010 is extended to the previous 9 years to obtain bias-corrected scenarios for the entire reference period 1980-2010. This is needed because MESAN data are available only for the former period. Figure 2a and 2b show that the six CMs basin-averaged monthly mean time series for both variables are in close agreement with ADIGE, with the largest deviations observed in May for precipitation (differences in the range of 15 - 21 mm), and in December for temperature (differences in the range of 1.3 - 1.9 °C), respectively. Accordingly, differences at the annual scale are rather small as highlighted in the insets of Figures 2a and 2b. ECDFs of basin-averaged daily precipitation and temperature for both ADIGE and the 6 CMs are presented in Figures 2c and 2d. For precipitation, no appreciable differences are observed between CMs and ADIGE throughout the entire range of variability. For the temperature (Figure 2d), small differences are observed which reduce progressively as temperature increases and become undetectable at high temperatures. Overall, these results indicate that CMs' outputs are in good agreement with the observations during the reference period, a statement which is also valid for the extremes of precipitation and temperatures which are indeed at the base of our approach.

Table 2: List of the EURO-CORDEX CMs used in this study. Acronyms adopted are listed in the last column.

RCM	GCM	Institute	RCP	Acronym
CLMcom-CCLM4-8-17	EC-EARTH-r1	Climate Limited-area Modelling Community (CLM-Community)	4.5	CMLcom
			8.5	
KNMI-RACMO22E	EC-EARTH-r12	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands	4.5	KNMI
			8.5	
SMHI-RCA4	HadGEM2-ES	Swedish Meteorological and Hydrological Institute, Rossby Centre	4.5	SMHI
			8.5	



	ADIGE	CLMcom RCP 4.5	CLMcom RCP 8.5	KNMI RCP 4.5	KNMI RCP 8.5	SMHI RCP 4.5	SMHI RCP 8.5
P [mm/anno]	950.8	947.7	950.4	946.0	934.5	877.2	883.9

	ADIGE	CLMcom RCP 4.5	CLMcom RCP 8.5	KNMI RCP 4.5	KNMI RCP 8.5	SMHI RCP 4.5	SMHI RCP 8.5
T[°C]	5.47	4.88	4.81	4.89	4.90	5.01	4.98

Figure 2: Annual cycle of basin-averaged monthly mean precipitation (a) and temperature (b) during the reference period 1980-2010 for both ADIGE and the 6 CMs used (different colour bars). The associated annual averages are also shown in the insets. ECDFs of basin-averaged daily precipitation and temperature for the same datasets are presented in subplots (c) and (d), respectively.

3.4 Simulations set-up

All the simulations were performed with the HYPERstreamHS hydrological model by using a daily time step and the 5 km computational grid depicted in Figure 1. Accordingly, precipitation and temperature provided by the ADIGE dataset and by

the six CM simulations presented in Sect. 3.3 were projected to this grid using the nearest neighbour method. Notice that the contributing area of the macrocells at the border of the domain was reduced by the amount belonging to the neighbouring basin, such as to preserve the overall contributing area of the investigated case study.

In a first set of simulations presented in Sect. 4.1, the HYPERstreamHS model was calibrated at the Trento gauging station by using the metrics NSE, KS and R_{FDC} as objective functions and the period 1980-2010 as reference. In order to ease the presentation of results, these three parameterizations are hereafter called NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE, respectively. Validation of the modelling framework was then performed, for these three parameterizations, by computing the efficiency metrics at the Bronzolo gauging station (drainage area of about 6000 km^2, see Figure 1) within the same time window, and at the Trento gauging station in the period 1950-1980, not used for calibration.

In a second set of simulations, presented in Sect. 4.2, we assessed whether the model calibrated with observational data and fed with precipitation and temperature obtained from climate models produces samples of annual streamflow maxima statistically coherent with the observations. Here we considered simulations performed in the period 1980-2010 by using precipitations and temperature from the three GCM-RCM combinations selected as described in Sect. 3.3 each one for both RCP4.5 and RCP8.5 emission scenarios, for a total of six CM combinations (see Table 2). The parameters of the hydrological model were those referring to NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE parameterizations.

In Sect. 4.3, we present the results of the calibration experiments performed by using in HYPERstreamHS the precipitations and temperature distributions provided by the six CMs for the period 1980-2010, and KS and R_{FDC} as objective functions. Following the procedure described in Sect. 2.4, extrapolations were then performed under the assumption that simulated and observed ECDFs were distributed according to the parametric Gumbel probability distribution. The Pearson's chi-squared test was then applied to verify the inferred model.

For all time windows and all simulations, the first two years were used as spin-up and therefore excluded from the computation of model performances. Furthermore, statistical coherence between simulated and observed samples of annual streamflow maxima was evaluated a-posteriori by using the p-values associated with the Kolmogorov-Smirnov two-sample test described in Sect. 2.3.

The effects on model parameters of calibrations conducted using different meteorological forcing (observational data as well CMs simulations) are investigated in Sect. 4.4 with reference to the KS metric. For each calibration experiment performed with the PSO algorithm, we considered 100 particles that, with a maximum number of 400 iterations, lead to a maximum of 40000 hydrological simulations for each external forcing. Parameters ranges considered during the search for the optimal solution were those presented in Table 1, and have been set by means of preliminary simulations such as to minimize the probability of excluding from the searching domain combinations of parameters leading to behavioural solutions [Beven and Binley, 1992]. In addition, we considered as a metric of uncertainty for the calibrated parameter the range, $\bar{d}$, between the maximum and minimum value of each parameter in the 200 simulations presenting the highest efficiency metric [see Piccolroaz et al., 2015]. We remark that the procedure adopted here aims at quantifying only the differences in the range of calibrated parameters and not to perform a full uncertainty analysis of predictions.

Finally, in Sect. 4.5 the projected changes of high flow extremes in the future period 2040-2070 are evaluated. For each CM we considered the following parameterizations obtained during calibration in the reference period: calibrations with KS and R_{FDC} as objective functions, and NSE-ADIGE as representative of a standard calibration procedure using the observational dataset ADIGE as input forcing.

4 Results and discussion

4.1 Simulations using the observational dataset ADIGE

Figure 3a shows the simulated ECDFs obtained by using the three metrics NSE, KS and R_{FDC} as objective functions and the observational ADIGE dataset as input forcing. Table 3 shows the associated p-values of the Kolmogorov-Smirnov test. From a statistical viewpoint, all three metrics provide simulated samples of annual streamflow maxima belonging to the same population as the observed ones, given that in all cases p>0.05, with a maximum for KS (p = 1.000) and a minimum for R_{FDC} (p = 0.372). However, calibration conducted by using KS as an objective function leads to NSE and R_{FDC} values (0.4 and 0.564, respectively, see Table 3) which are lower than those obtained when calibration is performed by optimizing (separately) these two metrics (NSE = 0.822 and R_{FDC} = 0.975, respectively, see Table 3). This is in accordance with several studies showing that the adoption of a given metric in calibration may lead to suboptimal results for other metrics since each one of them is sensitive to specific aspects of the time series with its limitations and trade-offs [see e.g., Schaefli and Gupta, 2007; Gupta et al., 2009; Mcmillan et al., 2017; Fenicia et al., 2018]. This latter limitation is, in our opinion, outweighed by the improvements in representing the ECDFs of observed high flow extremes when the model is calibrated considering explicitly such information, i.e. by minimizing the KS metric. Accordingly, in our analyses, the use of different efficiency metrics leads to different simulated ECDFs and hence to different p-values in the application of the statistical coherence test (see Table 3).

Validation of the hydrological modelling framework was performed by evaluating the model performance in the time frame 1952-1980, not used for calibration, at the gauging station of Ponte San Lorenzo in Trento. The validation was done by using the ADIGE dataset as input and the parameterizations obtained by calibrating the model in the time frame 1982-2010 (i.e., NSE-ADIGE, R_{FDC}-ADIGE and KS-ADIGE, as described above). NSE-ADIGE and R_{FDC}-ADIGE parameterizations led to NSE and R_{FDC} values (NSE = 0.803 and R_{FDC} = 0.804, see Table 3) which are only slightly lower than those obtained in calibration. KS-ADIGE parameterization leads to an increase of KS from 0.067 in calibration to 0.233 in validation, still rather small. The limited modifications of the efficiency metrics in validation is an encouraging result which shows that the HYPERstreamHS model provides a good representation of the hydrological system independently of the metric adopted in calibration. Simulated and observed ECDFs of annual streamflow maxima and the associated p-value of the Kolmogorov-Smirnov test are presented in Figure 3b. Reproduction of observed ECDF is satisfactorily for all the 3 parameterizations, particularly for high flow quantiles, with p-values in the range between 0.222 and 0.372 (see also Table 3). The three parameterizations provide simulated samples of annual streamflow maxima belonging to the same population of observations

also in the time window 1952-1980; the reduction of the p-value from calibration to validation is significant but rather common in hydrological models.

Spatial validation of the modelling framework was also performed by simulating streamflow at the Bronzolo gauging station (see Figure 1) in the same time window of the calibration conducted at the Trento Gauging station (1982-2010). Similarly, to the previous case, efficiency metrics in validation evidence a small reduction of performance with respect to those obtained in calibration (see Table 3). On the other hand, the results presented in Figure 3c highlight an excellent reproduction of the observed ECDF of annual streamflow maxima for all the 3 parameterizations, with the associated p-values in the range between 0.791 (NSE-ADIGE) and 0.951 (R_{FDC}-ADIGE and KS-ADIGE). The latter is a noteworthy result which indicates that the parameterization obtained using KS as an objective function is reliable, though relying on a limited number of observations, and does not introduce distortion in the spatial representation of the hydrological processes, particularly those controlling high streamflow events, i.e., runoff generation and streamflow concentration processes. This latter aspect will be further investigated in Sect. 4.4.

Table 3: Efficiency metrics for calibration and validation runs obtained by using the ADIGE dataset as input forcing. The terms NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE refer to the parameterizations described in Sect. 3.4. Grey shaded area indicates the metric optimized in calibration. p-values of the Kolmogorov-Smirnov test are also reported in the bottom line for the calibration experiments and the validation runs highlighted by bold numbers.

	Calibration			Validation					
	Trento 1982-2010			Trento 1952-1980			Bronzolo 1982-2010		
	NSE	R_{FDC}	KS	NSE	R_{FDC}	KS	NSE	R_{FDC}	KS
NSE-ADIGE	**0.822**	0.875	0.133	**0.803**	0.760	0.260	**0.787**	0.705	0.166
R_{FDC}-ADIGE	0.488	**0.975**	0.233	0.552	**0.804**	0.233	0.506	**0.830**	0.133
KS-ADIGE	0.400	0.564	**0.067**	0.250	0.529	**0.233**	0.289	0.476	**0.137**
p-value	0.951	0.372	1.000	0.222	0.372	0.372	0.791	0.951	0.951

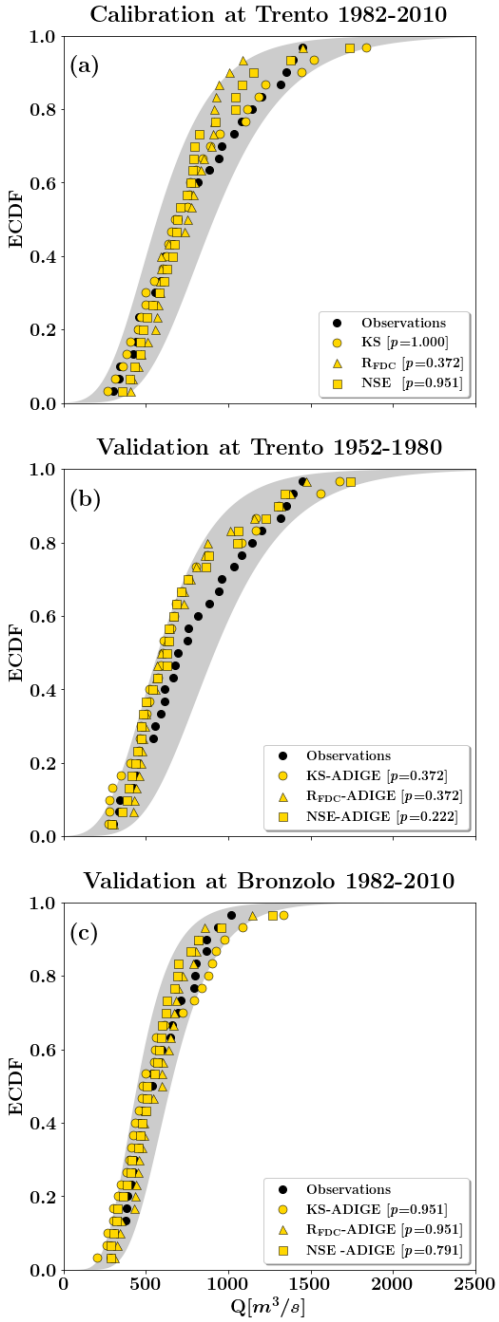


Figure 3: ECDFs of daily annual streamflow maximum obtained by using as input the observational dataset ADIGE and the parametrizations NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE at a) the Trento gauging station in the period 1982-2010; b) the Trento gauging station in the period 1952-1980, and c) the Bronzolo gauging station during the period 1982-2010. The experimental ECDFs obtained from streamflow observations in the same time frames are shown with black bullets with the grey shaded area indicating the associated 90% confidence interval of the fitted Gumbel distribution. p-values of the Kolmogorov-Smirnov two-sample test are also reported within brackets for each simulation run.

4.2 Simulations using parameterizations derived from calibrations with observed ground data

Here we analyse the case in which HYPERstreamHS was run in the time frame 1982-2010 using as input the meteorological variables produced by the climate models and the three parameterizations NSE-ADIGE, R_{FDC}-ADIGE, KS-ADIGE, described in Sect. 3.4. Visual inspection of Figures 4a, 4b and 4c evidence that for high quantiles the simulated ECDFs are often outside the 90% confidence interval of the Gumbel distribution fitted to observations for all the considered combinations of CMs and parameterizations. The p-values of these validation runs are shown in the last three columns of Table 4. In particular, these three parameterizations lead to p-values always lower than $p = 0.372$ for all the considered CMs and emission scenarios (see Table 4). NSE-ADIGE and R_{FDC}-ADIGE show on average the lowest p-values, with KS-ADIGE performing slightly better: $p = 0.372$ for KNMI and SMHI under the RCP8.5 scenario (see Figures 4b and 4c and Table 4). Inspection of Table 4 also reveals that values of $p < 0.05$, and thereby simulated ECDFs not belonging to the same population of the measured one, are obtained with the CLMcom model for both NSE-ADIGE and KS-ADIGE parameterizations under both emission scenarios, and for the KNMI model with NSE-ADIGE and R_{FDC}-ADIGE parameterizations under RCP4.5

The above results highlight how classical approaches based on feeding hydrological models, calibrated by using observed meteorological data and employing customary efficiency metrics (i.e., NSE and R_{FDC}), with meteorological forcing provided by Climate Models, produce results characterized by low statistical coherence with the observational data. Furthermore, our results indicate that the same drawback arises when employing parameterizations obtained with a calibration approach optimizing the desired statistic of extremes, but still using observational data as input, i.e., KS-ADIGE in Figures 4a, 4b and 4c. These results are in agreement with previous studies evidencing that the hydrological models, calibrated against observed data, that perform well within a baseline period may not be accurate nor consistent for simulating streamflow under future climate conditions [Brigode et al., 2013; Lespinas et al., 2014]. Indeed, it is recognized that the use of different datasets can lead to different optimized parameters that will partially account for their specific climate characteristics [Yapo et al. 1996; Vaze et al., 2010; Laiti et al., 2018]. Furthermore, it is acknowledged that climate change impact simulations are affected by uncertainty in climate modelling, but also the calibration strategy adopted during the reference period plays a role [Lespinas et al., 2014; Mizukami et al., 2019]. In this respect, we showed that the statistical coherence between climate scenarios and observations (i.e., high streamflow extremes in our case) should be preserved during hydrological calibration, at least in the reference period. This latter aspect will be further discussed in the ensuing Sect. 4.3.

4.3 Performance of the hydrological model calibrated using as input climate models' outputs

Table 4 summarizes the efficiency metrics and the p-values of the calibration experiments performed by using in HYPERstreamHS the precipitations and temperature distributions provided by the six selected CMs, and KS and R_{FDC} as objective functions. Simulations refer to the period 1982-2010. When KS is used in calibration, all the 6 simulations provided samples of annual streamflow maxima that with high probability (i.e. $p = 1.000$, column 8 of Table 4) belong to the same population of the observed values. A similar conclusion was reached for the objective function R_{FDC}, but with lower p values

(column 7 of Table 4), which are however larger than $p = 0.05$, the level of significance customarily adopted in the statistical literature to reject the null hypothesis. The lowest p-value was obtained with the climate model CLMcom under the RCP4.5 emission scenario and R_{FDC} as objective function ($p = 0.222$, see column 7 of Table 4). Consistently, the absolute maximum distances between the ECDF of observed and simulated samples obtained by using R_{FDC} as calibration metric are always larger than those obtained by using KS (see third and fifth columns in Table 4). When calibration is performed with KS the results are satisfactorily also with respect to the R_{FDC} metric, which is in the range between 0.449 and 0.804 for all the CMs (see the fourth column in Table 4). Since R_{FDC} employs the entire time series of observational data, this result evidences that using the KS metric in calibration does not introduce model's overparameterization, despite the reduced number of observational data used (i.e., 29 values of observed daily annual streamflow maxima).

The appreciable difference between observed and simulated ECDFs obtained in the calibration experiments conducted using KS and R_{FDC} metrics is highlighted in Figure 5. Figure 5 shows that the ECDFs obtained by extracting the annual maxima from the simulations calibrated with KS as objective function are in a better agreement with the observed ECDFs than those obtained by calibrating with R_{FDC}. This comparison highlights that the KS metric is preferable to R_{FDC} when dealing with high flow extremes, thus strengthening the approach envisaged here of addressing directly the desired statistics of extremes in calibration instead of calibrating the hydrological model on the entire streamflow record.

Table 4: R_{FDC} and KS efficiency metrics of the period 1982-2010 with forcing provided by CLMcom, KNMI, and SMHI climate models under the RCP4.5 and RCP8.5 emission scenarios. Grey shaded area and bold numbers indicate the metric optimized in calibration. p-values of the Kolmogorov-Smirnov test are also reported for all the calibration experiments and for the validations conducted using the parametrizations NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE.

Dataset		Efficiency metric				p-value				
						Direct calibration		Validations with ADIGE parameterizations		
		R_{FDC}	KS	R_{FDC}	KS	R_{FDC}	KS	NSE-ADIGE	R_{FDC}-ADIGE	KS-ADIGE
CLMcom	**RCP4.5**	**0.943**	0.267	0.730	**0.067**	0.222	1.000	0.030	0.222	0.030
KNMI	**RCP4.5**	**0.940**	0.167	0.804	**0.067**	0.791	1.000	0.013	0.030	0.123
SMHI	**RCP4.5**	**0.972**	0.200	0.589	**0.067**	0.572	1.000	0.222	0.123	0.123
CLMcom	**RCP8.5**	**0.980**	0.200	0.449	**0.067**	0.572	1.000	0.123	0.372	0.222
KNMI	**RCP8.5**	**0.961**	0.167	0.456	**0.067**	0.791	1.000	0.123	0.222	0.372
SMHI	**RCP8.5**	**0.932**	0.167	0.484	**0.067**	0.791	1.000	0.123	0.372	0.123

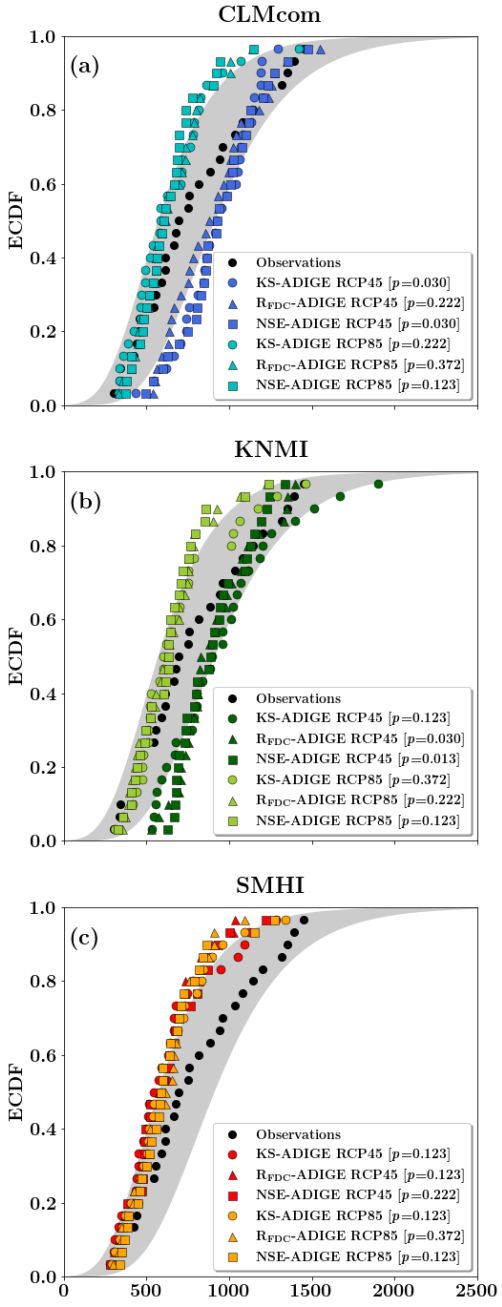


Figure 4: ECDFs of annual maximum streamflow at Trento gauging station in the period 1982-2010 obtained by using NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE parameterizations and a) CLMcom, b) KNMI, and c) SMHI climate models as input forcing for both RCP4.5 and RCP8.5 emission scenarios. The experimental ECDF is also shown with black dots together with the associated 90% confidence interval of the fitted Gumbel distribution (grey shaded area). p-values of the Kolmogorov-Smirnov two-sample test are also reported within brackets for each simulation run.

The literature reports a few examples of hydrological models calibrated by using tailored information instead of the entire observed streamflow time series [e.g., Montanari and Toth, 2007; Blazkova and Beven, 2009; Westerberg et al., 2011;

Lindenschmidt, 2017]. However, these approaches are typically adopted for reproducing basin response to observed meteorological forcing and have not been applied (to our best knowledge) in combination with GCM-RCMs simulations in climate change impact studies. The only example somewhat similar to our approach we found in the literature is that of Honti et al. [2014], who however used a stochastic weather generator trained by observed weather time series coupled with observed discharge data to sample the posterior distribution of model parameters. The adoption of a time-independent calibration, for which time shift does not influence the objective function, has the intrinsic advantage of allowing the use of GCM-RCM runs conducted without the assimilation of observational data, as in our case. In fact, these runs provide time-slice experiments representing a stationary climate for both reference and future periods [see e.g., Majone et al., 2012] and by definition cannot be used in the context of a classical day-by-day hydrological comparison experiment with observed historical data [see e.g., Eden et al., 2014].

Quantiles of daily annual streamflow maxima as a function of the return period at the Trento gauging station are shown in Figure 6, where results obtained by calibrating the hydrological model with the meteorological input provided by the Climate Models (for both KS and R_{FDC} metrics as objective functions) are compared with those obtained using the same meteorological input but employing NSE-ADIGE, R_{FDC}-ADIGE, and KS-ADIGE parameterizations. Visual inspection of Figure 6 reveals that for all return periods parametrizations obtained by calibrating with the observed precipitations and temperatures as provided by the ADIGE dataset significantly underestimate the quantiles of the observations and fall outside the confidence interval of the fitted Gumbel distribution (i.e., outside the grey area). The only exceptions are the quantiles derived from simulations conducted with KNMI (KS-ADIGE, dotted line in Figure 6c) and CLMcom (all the 3 metrics, Figure 6a) climate models under RCP4.5. We note, however, how these curves are obtained with forward simulations providing low p-values of the Kolmogorov-Smirnov test with respect to the other cases (always lower than $p = 0.222$). Instead, quantiles obtained from simulations optimized directly on Climate Models and by using KS as metric are in a very good agreement with the experimental data, while those obtained by using R_{FDC} are outside or at the lower bound of the interval of confidence, though they are generally in a better agreement with the quantiles of the experimental data than those obtained with the aforementioned NSE-ADIGE, R_{FDC}-ADIGE, and KS-ADIGE parametrizations. Exceptions are the quantiles obtained with CLMcom and KNMI under RCP4.5 emission scenario and R_{FDC} as metric which are characterized by the largest deviations from observations (see Figures 6a and 6c, respectively). We attribute this occurrence to the additional source of uncertainty arising from the extrapolation procedure (i.e., the selection of the probability distribution and of the statistical inference method for the parameters, MLE in our case). The interval of confidence of the fitted Gumbel distribution to the observational data (grey area) widens as the return period increases and this is in line with the recent findings of Meresa and Romanowicz [2017], which showed that errors in fitting theoretical distribution models to annual maxima streamflow series might contribute significantly to the overall uncertainty associated to projections of future hydrological extremes.

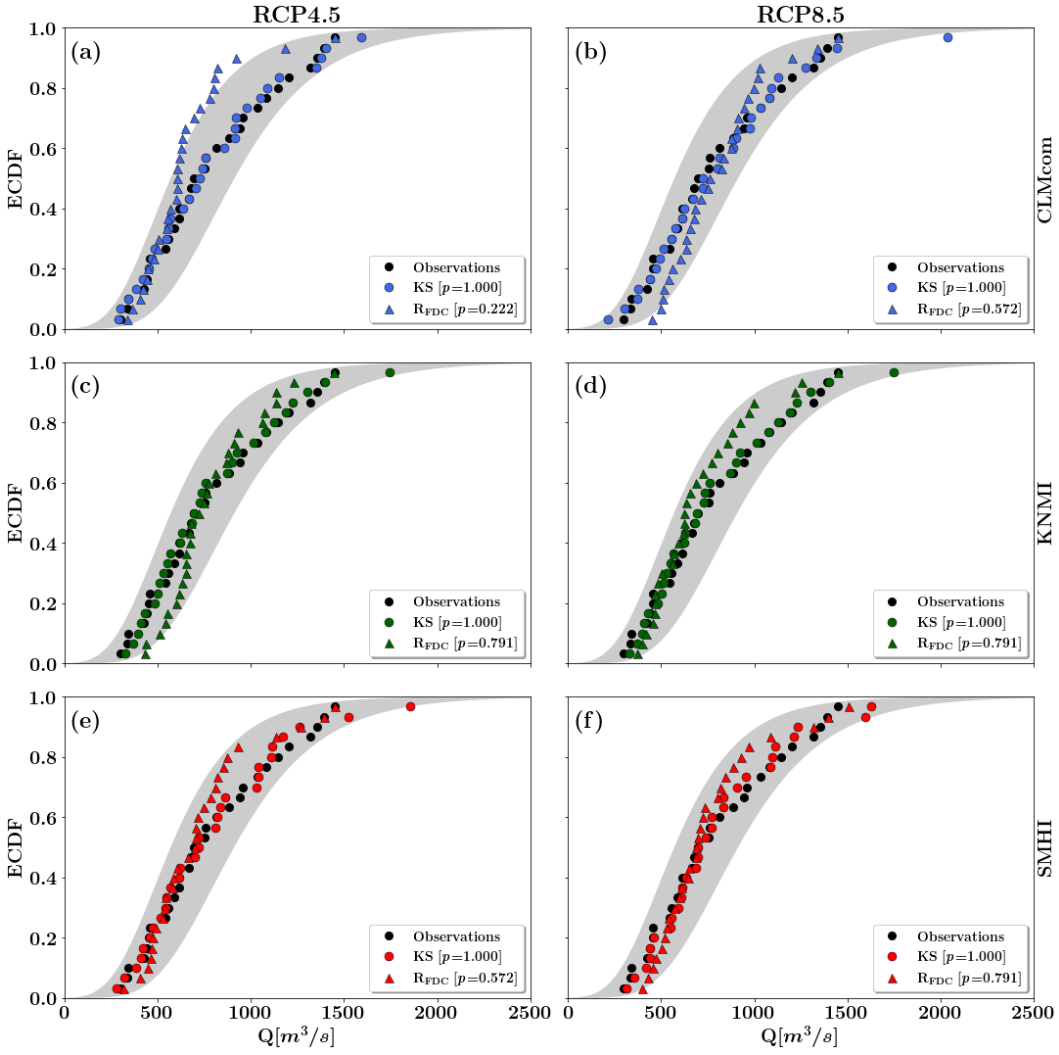


Figure 5: Simulated ECDFs of daily annual maximum streamflow at the gauging station of Trento in the period 1982-2010 with precipitation and air temperature provided by CLMcom (first row), KNMI (second row), and SMHI (third row) climate models under the RCP4.5 (left) and RCP8.5 (right) emission scenarios. Calibration of HYPERstreamHS was performed using both KS and R_{FDC} metrics as objective functions. The ECDF of observations is also shown with black dots together with the associated 90% confidence interval of the fitted Gumbel distribution (grey shaded area). p-values of the Kolmogorov-Smirnov two-sample test are also reported within brackets for each simulation run.

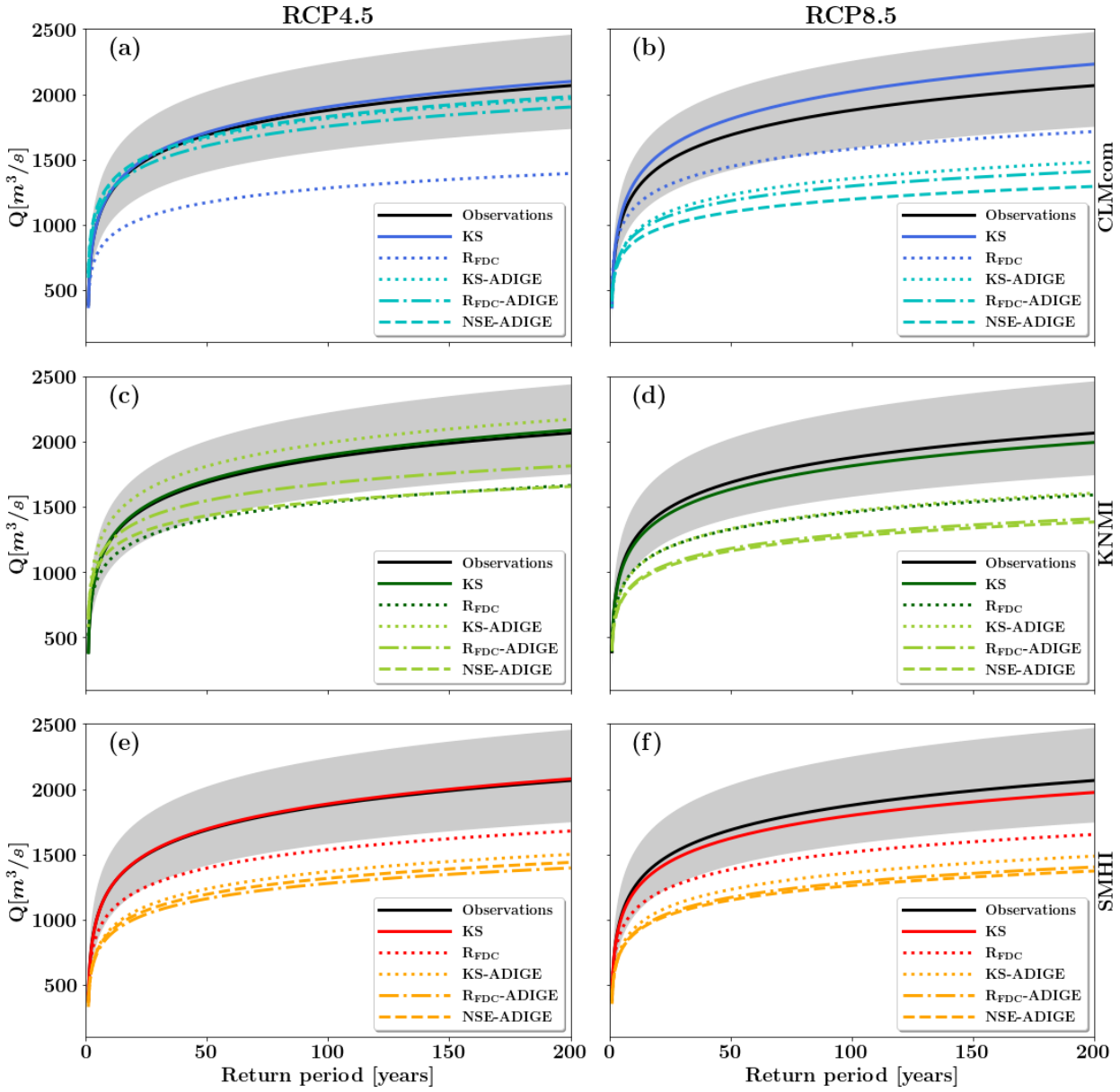


Figure 6: Quantiles of daily annual streamflow maxima as a function of return period at the Trento gauging station. Extrapolations are based on simulations conducted during the period 1982-2010 using as input forcing the CLMcom (first row), KNMI (second row), and SMHI (third row) climate models under the RCP4.5 (left) and RCP8.5 (right) emission scenarios, respectively. Each curve represents a combination of CM, emission scenario and parameterization obtained with the calibration. Simulations conducted using the parameterizations obtained by using the observational dataset ADIGE in calibration are labelled as NSE-ADIGE, R_{FDC}-ADIGE and KS-ADIGE. Extrapolation from observed streamflow maxima is also shown (continuous black line) together with the associated 90% confidence interval of the fitted Gumbel distribution (grey shaded area).

4.4 Model parameters

The results presented in the previous Sections highlight that the better statistical coherence between observations and simulations (performed with CMs simulations as input) was achieved by optimizing the desired statistics of extremes, in our case KS (see the curves labelled KS in Figures 5 and 6), in the calibration of the hydrological model. Starting from this evidence, we investigated the effect on model parameters of performing the calibration by using either observed or derived from CMs meteorological data and KS as objective function. Figure 7 shows the range, $\bar{d}$, between the maximum and minimum values, here represented by the length of the vertical bar, of each parameter among the 200 accepted values corresponding to the behavioural models (see Sect. 3.4), together with the corresponding optimal parameter set, which is represented with a horizontal segment. The values of the parameters are normalized with respect to their range (see Table 1) such that they are directly comparable. In all simulations the normalized parameters range $\bar{d}$ is well distributed between 0 and 1, indicating a proper choice of the parameters range in the PSO algorithm, although for a few of them the optimal value was located close to the boundary of the searching domain. As shown in Figure 7 the majority of the parameters obtained by using the proposed approach span a range $\bar{d}$ that is similar in terms of amplitude (or slightly larger) to that obtained for KS-ADIGE, thus supporting the conclusion that calibration using CMs simulations does not lead, for both RCPs, to bias parameterizations. Figure 7 also shows that for most of the parameters, simulations performed with CMs lead to generally overlapping ranges for $\bar{d}$ with respect to the case in which the observational dataset ADIGE was used. The largest deviations in terms of $\bar{d}$ are observed for KS-KNMI, particularly under the RCP8.5 emission scenario. Notably, the parameters shaping the continuous soil-moisture accounting module result in values of the optimum which are very similar for all cases (see q_{ref}, μ, and c_{fc} in Figures 7a and 7b). Visual inspection of Figure 7 also highlights that the parameters controlling runoff generation and streamflow concentration (in particular, v, c_s, q_{ref}, and c_{fc}) present very good identifiability (i.e., a small range $\bar{d}$). This is not the case for parameters controlling snowmelting and groundwater contribution, the latter being relevant only for low flow conditions (see k in Figures 7a and 7b). These results, together with the good performances obtained in the validation runs presented in Sect. 4.1, suggest that, although the model is calibrated considering a limited number of observations, in the continuous simulations the maxima are well reproduced but this is achieved only if the interaction between the precipitation and streamflow relevant during high flow extremes is correctly reproduced. We cannot exclude that additional analyses could be envisioned for improving the identifiability of some parameters (e.g., by reducing the number of model parameters, introducing constraints in the parameters range, etc.) in applications dealing with different hydrological models and different data availabilities (e.g. lower number of streamflow extremes). However, the analysis presented here provides clear evidences that the parameterizations derived from the use of KS metric are reliable.

The differences observed in the optimal values of model parameters are due to the use of datasets for the meteorological forcing with different capabilities to reproduce the present climate. Along the concepts brought forward here, this source of uncertainty can be addressed effectively by calibration of the hydrological model to the quantities of interest (i.e. the observed streamflow

statistics of extremes) using as input the forcing provided by a specific CM. This approach can be seen as a "*hydrologic-based bias-correction*" and is rooted in the adoption of a "goal-oriented" calibration framework [see e.g., Laiti et al., 2018] along the lines stated in the Introduction.

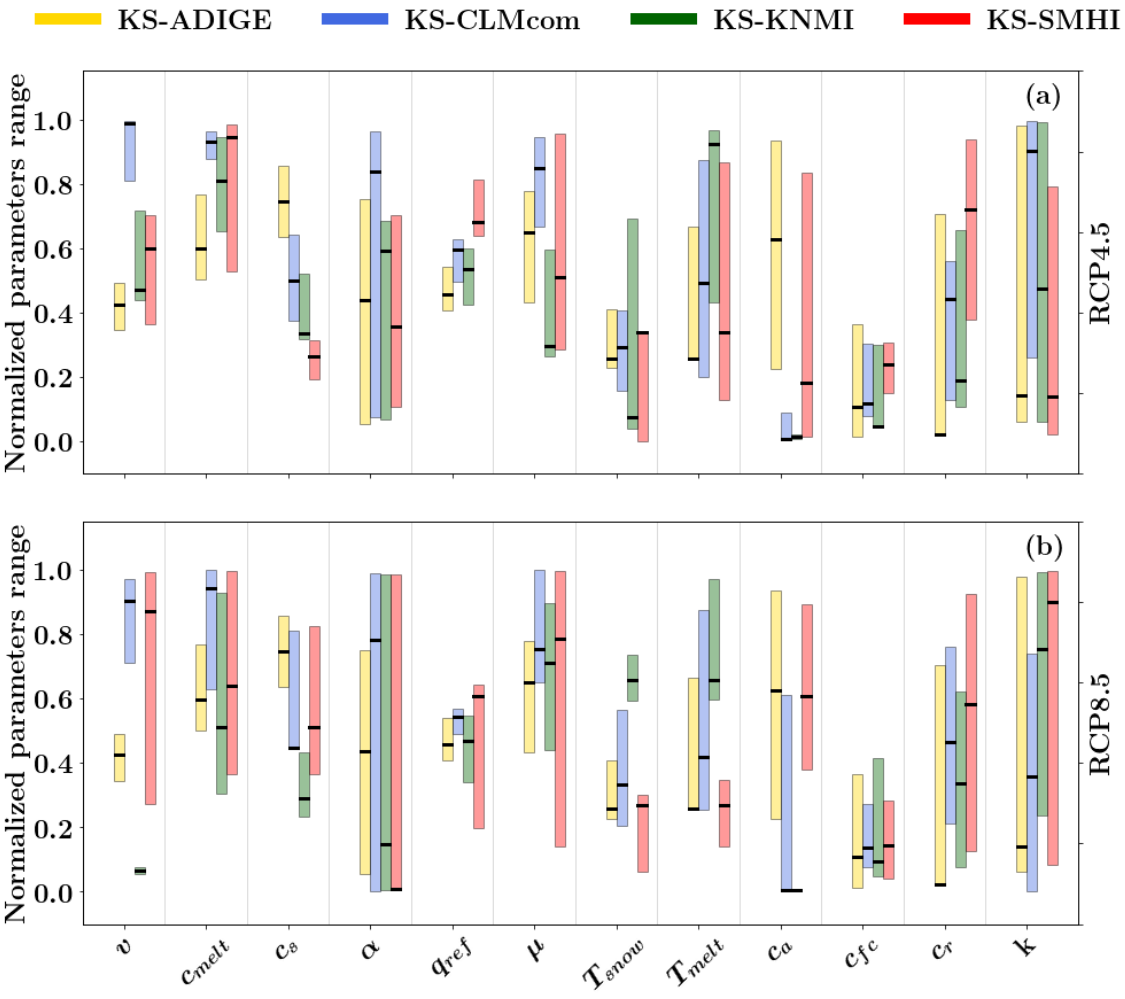


Figure 7: Range, $\bar{d}$, between the maximum and minimum value of each parameter associated with the 200 simulations presenting the highest efficiency plotted as a normalized range with respect to the parameter range presented in Table 1. Calibrations are conducted for the 3 different CMs under (a) RCP4.5 and (b) RCP8.5 emission scenarios with reference to the KS metric. Bold horizontal segments indicate the optimal parameter sets for all experiments.

4.5 Projected changes of streamflow quantiles

Figure 8 shows the annual maximum streamflow at the Trento gauging station as a function of the return period in the future time window 2040-2070 and for the 6 selected CMs. Visual inspection of Figure 8 confirms that in all cases using the standard calibration (i.e., NSE-ADIGE) of the hydrological model leads to a significant underestimation of all quantiles with respect to

using KS and R_{FDC}. This is in agreement with the results obtained for the reference period (see Figure 6), where simulations using NSE-ADIGE parameterization provided streamflow quantiles systematically lower than that obtained with the CMs. In addition, KS-based calibrations always provide larger quantiles with respect to the cases in which the R_{FDC} metric is adopted (considering the same RCP emission scenario). We remark how the adoption of the KS metric is preferable since it provided an almost perfect match with observed streamflow quantiles in the calibration period (see Figure 6).

Moreover, Figure 8 shows that projected changes of high flows extremes depend on the selected CM and emission scenario. Projected streamflow quantiles under RCP8.5 are larger than those under RCP4.5 for all the CMs. In general, the projected streamflow quantiles do not exceed those obtained by fitting the Gumbel distribution to the observational data of the period 1982-2010 (continuous black lines in Figures 6 and 8), with the exceptions of CLMcom and SMHI models under RCP8.5 and SMHI under RCP4.5 when KS metric is adopted. These results are in line with other recent contributions which concluded that the sign and magnitude of projected changes of high flow extremes vary significantly with the location of the investigated river basin, the climate models used, the emission scenario as well as the selection of the time window [Ngongondo et al., 2013; Aich et al., 2016; Pechlivanidis et al., 2017; Vetter at al., 2017]. Our results are in line with the analysis of Brunner et al. [2019] who implemented a stochastic framework to simulate future streamflow time series in 19 regions of Switzerland and concluded that future maximum streamflow will increase and decrease in rainfall-dominated and melt-dominated regions, respectively. Similarly, Di Sante et al. [2019] showed that a moderate increase in high flow magnitude (return time of 100 years) is projected for large river basins (drained area >10.000 km^2) in the Central Europe region under RCP8.5 and considering a mid-century time slice.

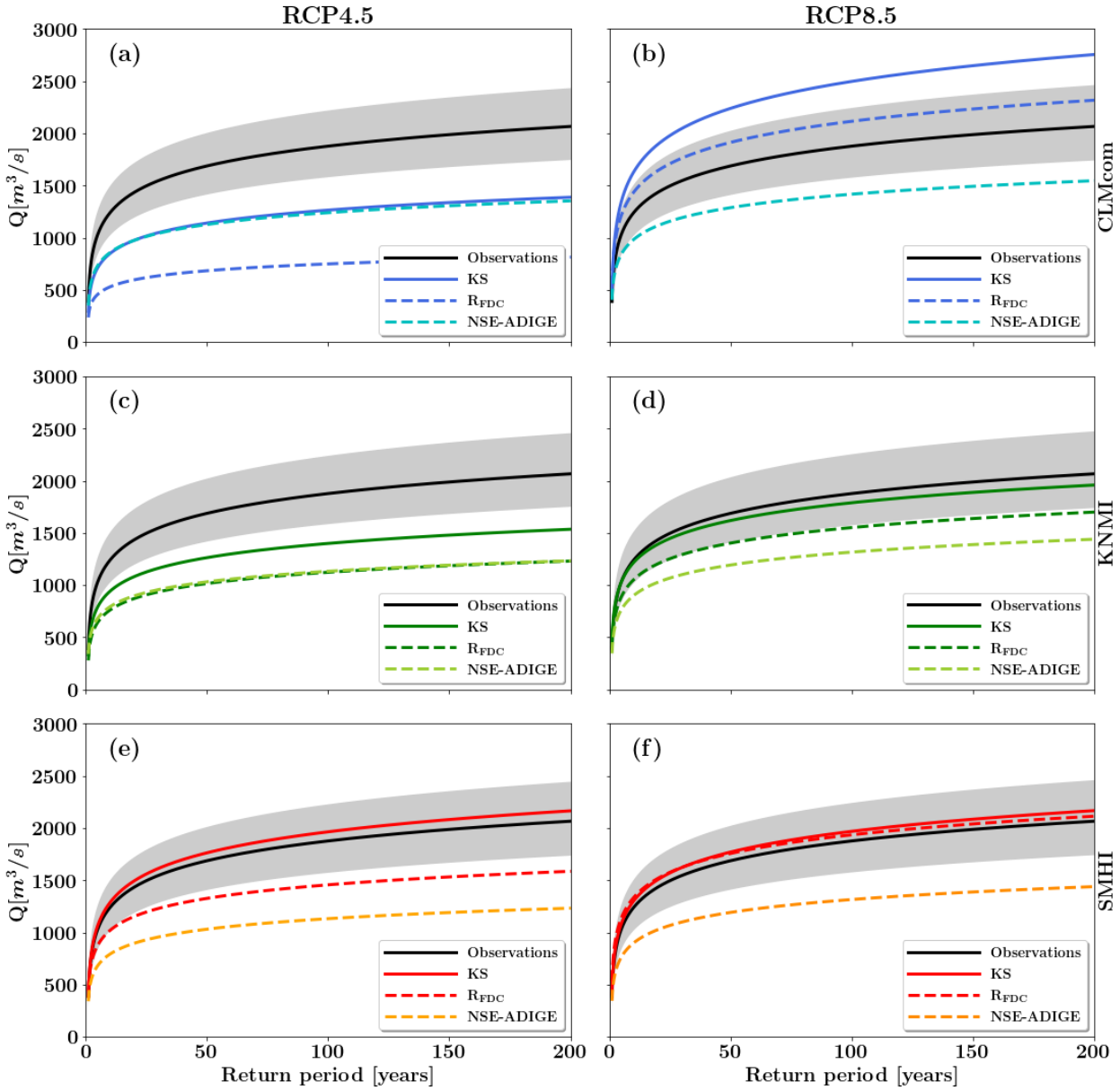


Figure 8: Quantiles of annual maximum daily streamflow as a function of return period at the Trento gauging station. Projections are based on simulations of the future time period 2042-2070 using as input the CLMcom (first row), KNMI (second row), and SMHI (third row) climate models under the RCP4.5 (left) and RCP8.5 (right) emission scenarios, respectively. The continuous black line shows the quantile distribution of high flow extremes evaluated with the observational data of the period 1982-2010 together with the associated 90% confidence interval of the fitted Gumbel distribution (grey shaded area).

5 Conclusions

In this work, we proposed the methodological framework HyCoX in which the calibration of the hydrological model is carried out by maximizing the probability that the modelled and observed high streamflow extremes belong to the same statistical population. The proposed framework is "goal-oriented" and aims at improving the estimation of streamflow extremes by directly calibrating the selected hydrological model to the quantities of interest (i.e. flow statistics instead of time series) using as input directly the meteorological data provided by Climate Models. In particular, the framework relies on the use of the two-sample Kolmogorov-Smirnov statistic (KS) as an objective function during the calibration procedure. This approach ensures statistical coherence between scenarios and observations in the reference period, and, likely, preserves it in the future climate change scenario runs performed to project changes in streamflow extremes. The goal-oriented approach envisaged in this work can be applied to a variety of hydrological scenarios and modelling approaches. Furthermore, we remark that the HyCoX methodology is not metric dependent, and any type of metric assessing the statistical coherence between observed and simulated streamflow extremes can be employed without any loss of generality.

The proposed procedure is exemplified through the application of six Climate Models and observational data to the analysis of the annual maximum streamflow of the Adige river basin (Italy) using the distributed hydrological model HYPERstreamHS. While the approach is exemplified here for high flows, it can be applied to low flows as well (e.g. for drought assessment). The results highlight that adopting KS is preferable to other popular metrics (e.g. NSE or fit to flow duration curve, R_{FDC}) when dealing with high streamflow extremes. This validates our hypothesis that addressing directly the statistics of extremes under consideration during the calibration exercise leads to coherent and reliable hydrological models for assessing the impact of climate change. We warn that such an approach may lead to a suboptimal performance if the target is different from the one employed in this study, in line with the goal-oriented framework here pursued. Alternatively, a multi-objective approach could be envisioned to investigate the trade-off in model performance emerging from the use of multiple metrics, including the one proposed here. This latter aspect is indeed beyond the objective of the present contribution, though it is worthy of further analysis. Furthermore, investigation of optimal values highlighted that direct calibration using CMs outputs and KS as objective function leads to unbiased identification of model parameters.

Overall, we showed that the way the hydrological model is calibrated against observations assumes paramount importance in climate change impact assessments on streamflow extremes. In particular, we highlighted how the classical approach of calibrating on daily streamflow observations by using observed meteorological data can lead to a biased probability distribution of streamflow extremes when climate models are used as input forcing during the reference period, with high streamflow quantiles being dramatically underestimated with respect to the fitted distribution of the observed extremes. Extrapolations performed by using the proposed calibration procedure, with input provided by CMs, are instead more reliable and they provide a good match with observed quantiles.

Author contribution

Bruno Majone: Writing - original draft preparation, Writing - review & editing, Investigation, Software, Conceptualization, Methodology, Supervision, Funding acquisition; **Diego Avesani**: Writing - review & editing, Investigation, Software, Visualization, Data Curation; **Patrick Zulian**: Software, Data curation; **Aldo Fiori**: Writing - review & editing, Conceptualization, Methodology, Supervision; **Alberto Bellin**: Writing - review & editing, Conceptualization, Methodology, Supervision, Funding acquisition.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research received financial support from the Italian Ministry of Education, University and Research (MIUR) under the Departments of Excellence, grant L.232/2016, and by the Energy oriented Centre of Excellence (EoCoE-II), grant agreement number 824158, funded within the Horizon2020 framework of the European Union. Bruno Majone also acknowledges support by the project "Seasonal Hydrological-Econometric forecasting for hydropower optimization (SHE)" funded within the Call for projects “Research Südtirol/Alto Adige” 2019 Autonomous Province of Bozen/Bolzano – South Tyrol. Diego Avesani acknowledges financing from the European Union - FSE-REACT-EU, PON Research and Innovation 2014-2020 DM1062 / 2021. The authors acknowledge the climate modelling groups listed in Table 2 of this paper, for producing and making available their model output within the EURO-CORDEX initiative (https://www.euro-cordex.net/index.php.en). Streamflow data were kindly provided by the Service for Hydraulic Works of the Autonomous Province of Trento (www.floods.it) and Hydrological Office of the Autonomous Province of Bolzano (https://meteo.provincia.bz.it/default.asp). We also thank the two anonymous Referees whose comments and suggestions helped improve and clarify this manuscript.

References

Aich, V., Liersch, S., Vetter, T., Fournet, S., Andersson, J.C.M., Calmanti, S., Van Weert, F.H.A., Hattermann, F.F., and Paton, E.N.: Flood projections within the Niger River Basin under future land use and climate change, Sci. Total Environ., 562, 666-677, doi:10.1016/j.scitotenv.2016.04.021, 2016.

Arnell, N.W.: Uncertainty in the relationship between climate forcing and hydrological response in UK catchments. Hydrol Earth Syst Sci, 15, 897-912. doi:10.5194/hess-15-897-2011, 2011.

Avesani, D., Galletti A., Piccolroaz S., Bellin A. and Majone B.: A dual layer MPI continuous large-scale hydrological model including Human Systems, Environ. Model. Softw., 139, 105003. doi:10.1016/j.envsoft.2021.105003, 2021.

Avesani D., Zanfei A., Di Marco N., Galletti A., Ravazzolo F., Righetti M. and Majone B.: Short-term hydropower optimization driven by innovative time-adapting econometric model, 310, 118510, Appl. Energy, 2022.

Bard, A., Renard, B., Lang, M., Giuntoli, I., Korck, J., Koboltschnig, G., Janža, M., D'Amico, M., Volken D.: Trends in the hydrologic regime of Alpine rivers, J. Hydrol., 529, 1823-1837, doi:10.1016/j.jhydrol.2015.07.052, 2015.

Bellin, A., Majone, B., Cainelli, O., Alberici, D., and Villa F.: A continuous coupled hydrological and water resources management model, Environ. Model. Softw., 75, 176-192, doi:10.1016/j.envsoft.2015.10.013, 2016.

Beven, K. J., and Binley A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Processes, 6(3), 279-298, doi:10.1002/hyp.3360060305, 1992.

Beven, K. and Westerberg I.: On red herrings and real herrings: disinformation and information in hydrological inference, Hydrol. Process., 25(10), 1676-1680, doi:10.1002/hyp.7963, 2011

Blazkova, S. and Beven K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resour. Res., 45, W00B16, doi:10.1029/2007WR006726, 2009.

Bouwer, L. M., Projections of Future Extreme Weather Losses Under Changes in Climate and Exposure, Risk Analysis, Vol. 33 (5), doi:10.1111/j.1539-6924.2012.01880.x, 2013.

Brigode, P., Oudin, L., and Perrin, C.: Hydrological model parameter instability: a source of additional uncertainty in estimating the hydrological impacts of climate change? J. Hydrol., 476, 410-425, doi:10.1016/j.jhydrol.2012.11.012, 2013.

Brigode, P., Paquet, E., Bernardara, P., Gailhard, J., Garavaglia, F., Ribstein, P., Bourgin, F., Perrin C., and Andréassian, V.: Dependence of model-based extreme flood estimation on the calibration period: the case study of the Kamp River (Austria), Hydrological Sciences Journal, 60 (7-8), 1424-1437, doi.org/10.1080/02626667.2015.1006632, 2015.

Brunner, M. I., Farinotti, D., Zekollari, H., Huss, M., and Zappa M.: Future shifts in extreme flow regimes in Alpine regions, Hydrol. Earth Syst. Sci., 23 (11), 4471-4489, doi:10.5194/hess-23-4471-2019, 2019.

Buytaert, W., De Bièvre B.: Water for cities: the impact of climate change and demographic growth in the tropical Andes, Water Resour. Res., 48, W08503, doi:10.1029/2011WR011755, 2012.

Calenda, G., Mancini, C.P., and E. Volpi, E.: Selection of the probabilistic model of extreme floods: The case of the River Tiber in Rome, J. Hydrol., 371(1-4), 1-11, doi:10.1016/j.jhydrol.2009.03.010, 2009.

Chiew, F., Teng, J., Vaze, J., Post, D., Perraud, J., Kirono, D., and Viney, N.: Estimating climate change impact on runoff across southeast Australia, Method, results, and implications of the modeling method, Water Resour. Res., 45, W10414, doi:10.1029/2008WR007338, 2009.

Chiogna, G., Majone, B., Cano Paoli, K., Diamantini, E., Stella, E., Mallucci, S., Lencioni, V., Zandonai, F. and Bellin, A.: A review of hydrological and chemical stressors in the Adige basin and its ecological status, Sci. Tot. Env., 540, 429-443, doi:10.1016/j.scitotenv.2015.06.149, 2016.

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R. and Brekke, L. D.: Characterizing Uncertainty of the Hydrologic Impacts of Climate Change, Curr. Clim. Change Rep., 2, 55–64, doi:10.1007/s40641-016-0034-x, 2016.

Conover, W.J.: Practical Nonparametric Statistics, Third edition, Wiley Series in Probability and Statistics: Applied Probability and Statistics Section, John Wiley & Sons. INC., New York, 1999.

Diamantini, E., Lutz, S.R., Mallucci, S., Majone, B., Merz, R., and Bellin, A.: Driver detection of water quality trends in three large European river basins, Sci. Total Environ., 612, 49-62, doi.org/10.1016/j.scitotenv.2017.08.172, 2018.

Di Sante, F., Coppola, E., and Giorgi F.: Projections of river floods in Europe using EURO-CORDEX, CMIP5 and CMIP6 simulations, International Journal of Climatology, 41(5), doi:10.1002/joc.7014, 2019.

Eden, J. M., Widmann, M., Maraun D., and Vrac M.: Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing, J. Geophys. Res. Atmos., 119, 11,040–11,053, doi:10.1002/2014JD021732, 2014.

Efron, B.: The jackknife, the bootstrap, and other resampling plans, 38, Society of Industrial and Applied Mathematics CBMS-NSF Monographs, ISBN 0-89871-179-7, 1982.

Fenicia, F., Kavetski, D., Reichert, P., and Albert, C.: Signature-domain calibration of hydrological models using approximate Bayesian computation: Empirical analysis of fundamental properties. Water Resources Research, 54, 3958-3987, https://doi.org/10.1002/2017WR021616, 2018.

Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., et al.: Debates-stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for?, Water Resour. Res., 52, 9228-9234, https://doi.org/10.1002/2015WR017525, 2016.

Galletti, A., Avesani, D., Bellin A, and Majone, B.: Detailed simulation of storage hydropower systems in large Alpine watersheds. J. Hydrol., 603, 127125, http://dx.doi.org/10.1016/j.jhydrol.2021.127125, 2021.

Gampe, D., Nikulin, G. and Ludwig, R.: Using an ensemble of regional climate models to assess climate change impacts on water scarcity in European river basins, Sci. Total Environ., 573, 1503-1518, doi: 10.1016/j.scitotenv.2016.08.053, 2016.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J. and Stoffel, M.: 21st century climate change in the European Alps. A review, Sci. Total Environ., 493, 1138-1151, doi: 10.1016/j.scitotenv.2013.07.050, 2014.

Goovaerts, P.: Geostatistics for natural resources evaluation, Oxford University Press, 483 pp, 1997.

Grubbs, F. E.: Procedures for Detecting Outlying Observations in Samples, Technometrics 11(1), 1-21, doi:10.1080/00401706.1969.10490657, 1969.

Gumbel, E. J.: The return period of flood flows, Ann. Math Stat., 12(2), 163-190, 1941.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, 2009.

Guthke, A.: Defensible model complexity: A call for data-based and goal-oriented model choice. Groundwater, 55(5), 646-650, https://doi.org/10.1111/gwat.12554, 2017.

Hargreaves, G.H. and Samani, Z.A.: Estimating potential evapotranspiration, J. Irrig. Drain. Eng., 108, 225-230, 1989.

Harris, I., P.D. Jones, T.J. Osborn, and D.H. Lister (), Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset, Int. J. Climatol., 34, 623-642, doi:10.1002/joc.3711, 2014.

Hattermann, F.F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Florke F., Gosling, S.N., Hoffmann, P., Liersch, S., Masaki, Y., Motovilov, Y., Muller, C., Samaniego, L., Stacke, T., Wada, Y., Yang, T., and Krysnaova, V.: Environ. Res. Lett., 015006, doi.org/10.1088/1748-9326/aa9938, 2018.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New., M.: A European daily high-resolution gridded dataset of surface temperature and precipitation, J. Geophys. Res., 113, D20119, doi:10.1029/2008JD10201, 2008.

Heistermann, M., and Kneis, D.: Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling, Water Resour. Res., 47, W06514, doi:10.1029/2010WR009153, 2011.

Hock, R.: Temperature index melt modelling in mountain areas, J. Hydrol., 282, 104-115, doi: 10.1016/S0022-1694(03)00257-9, 2003.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky C. T.: Bayesian model averaging: A tutorial, Stat. Sci., 14(4), 382-417, 1999.

Hofstra, N., Haylock, M., New, M. and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, J. Geophys. Res., 114, D21101, doi:10.1029/2009JD011799, 2009.

Hofstra, N., New, M. and McSweeney, C.: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, Clim. Dyn. 35, 841–858, doi: 10.1007/s00382-009-0698-1, 2010.

Honti M., A. Scheidegger, A. and Stamm, C.: The importance of hydrological uncertainty assessment methods in climate change impact studies, Hydrol. Earth Syst. Sci., 18, 3301–3317, doi:10.5194/hess-18-3301-2014, 2014.

Hosking, J.R.: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution, Appl. Stat., 34, pp. 301-310, doi.org/10.2307/2347483, 1985.

Isotta, F.A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S.M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G. and Vertačnik G.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, Int. J. Climatol., 34, 1657–1675, doi:10.1002/joc.3794, 2014.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Georgopoulou, E., et al.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Chang., 14, 563-578, 2014.

Journel, A. G. and Rossi, M. E.: When do we need a trend model in kriging?, Math. Geol., 21, 715–739, doi: 10.1007/BF00893318, 1989.

Kennedy, J. and Eberhart R.: Particle swarm optimization, in Proceedings of IEEE International Conference on Neural Networks, Institute of Electrical & Electronics Engineering, University of Western Australia, Perth, Western Australia, 1942–1948, doi:10.1109/ICNN.1995.488968, 1995.

Kleinen, T. and Petschel-Held, G.: Integrated assessment of changes in flooding probabilities due to climate change, Climatic Change, 81:283–312, doi:10.1007/s10584-006-9159-6, 2007.

Kotlarski, S., Keuler, K., Christensen, O.B, Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V.: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297-1333, doi:10.5194/gmd-7-1297-2014, 2014.

Kundzewicz, Z., Mata, L., Arnell, N., Döll, P., Kabat, P., Jiménez, B., Miller, K., Oki, T., Shen, Z., and Shiklomanov, I.: Freshwater resources and their management, in: Climate change: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel of Climate Change, edited by: Parry, M., Canziani, O., Palutikof, J., van der Linden, P., and Hanson, C., 173–210, Cambridge University Press, Cambridge, UK, 2007.

Laio, F., Allamano, P. and Claps, P.: Exploiting the information content of hydrological outliers for goodness-of-fit testing. Hydrol. Earth Syst. Sci. 14(10), 1909–1917, doi:10.5194/hess-14-1909-2010, 2010.

Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., Nikulin, G., and Majone, B.: Testing the hydrological coherence of high-resolution gridded precipitation and temperature datasets. Water Resources Research, 54, 1999–2016. https://doi.org/10.1002/2017WR021633, 2018.

Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A., and Olsson, E.: A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind, Q.J.R. Meteorol. Soc., doi: 10.1002/qj.2813, 2016.

Larsen, S., Majone, B., Zulian, P., Stella, E., Bellin, A., Bruno, M. C., and Zolezzi, G.: Combining hydrologic simulations and stream-network models to reveal flow-ecology relationships in a large Alpine catchment. Water Resources Research, 57, e2020WR028496. https://doi.org/10.1029/2020WR028496, 2021.

Lespinas, F., Ludwig, W., and Heussner, S.: Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers, J. Hydrol., 511, 403–422, https://doi.org/10.1016/j.jhydrol.2014.01.033, 2014.

K.E. Lindenschmidt: Using stage frequency distributions as objective functions for model calibration and global sensitivity analyses, Environ. Model. Softw., 92, 169-175. http://dx.doi.org/10.1016/j.envsoft.2017.02.027, 2017.

Lutz, S.R., Mallucci, S., Diamantini, E., Majone, B., Bellin, A. and Merz, R.: Hydroclimatic and water quality trends across three Mediterranean river basins, Sci. Tot. Env., 571, 1392-1406, doi:10.1016/j.scitotenv.2016.07.102, 2016.

Majone, B., Bertagnoli, A, and A. Bellin A.: A non-linear runoff generation model in small Alpine catchments, J. Hydrol., 385, 300-312, doi: 10.1016/j.jhydrol.2010.02.033, 2010.

Majone, B., Bovolo, C.I., Bellin, A., Blenkinsop, S. and Fowler, J.: Modeling the impacts of future climate change on water resources for the Ga´llego river basin (Spain), Water Resour. Res., 48, W01512, doi:10.1029/2011WR010985, 2012.

Majone, B., F. Villa, R. Deidda, and A. Bellin: Impact of climate change and water use policies on hydropower potential in the south-eastern Alpine region, Sci. Tot. Env., 543(B), 965–980, doi:10.1016/j.scitotenv.2015.05.009, 2016.

Mallucci, S., B. Majone, B. and Bellin A.: Detection and attribution of hydrological changes in a large Alpine river basin. Journal of Hydrology, 575:1214-1229, doi:10.1016/j.jhydrol.2019.06.020, 2019.

Mcmillan, H., I. Westerberg, I. and Branger, F.: Five guidelines for selecting hydrological signatures. Hydrological Processes, 31 (26), 4757-4761. https://doi.org/10.1002/hyp.11300, 2017.

Meresa, H.K., and Romanowicz, R.J.: The critical role of uncertainty in projections of hydrological extremes, Hydrol. Earth Syst. Sci., 21, 4245–4258, https://doi.org/10.5194/hess-21-4245-2017, 2017.

Michel, C., Andreassian, V. and Perrin, C.: Soil Conservation Service Curve Number method: How to mend a wrong soil moisture accounting procedure?, Water Resour. Res., 41, W02,011, doi:10.1029/2004WR003191, 2005.

Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, V.H. and Kumar, R.: On the choice of calibration metrics for "high-flow" estimation using hydrologic models, Hydrol. Earth Syst. Sci., 23, 2601-2614. https://doi.org/10.5194/hess-23-2601-2019, 2019.

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resour. Res., 43, W05434, doi:10.1029/2006WR005184, 2007.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., et al.: 'Panta Rhei-Everything Flows': Change in hydrology and society - The IAHS Scientific Decade 2013-2022, Hydrological 711 Sciences Journal 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

Muñoz, E., Arumí, J. L. and Rivera, D.: Watersheds are not static: Implications of climate variability and hydrologic dynamics in modelling, Bosque (Valdivia), 714 34(1), 7-11. doi:10.4067/S0717-92002013000100002, 2013.

Nash, J.E. and Sutcliffe, J.V.: River flow forecasting through conceptual models part I. A discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Ngongondo, C., Li, L., Gong, L., Xu, C. and Alemawm, B.F: Flood frequency under changing climate in the upper Kafue River basin, southern Africa: a large scale hydrological model application, Stoch. Environ. Res. Risk. Assess., 27:1883–1898, doi:10.1007/s00477-013-0724-z, 2013.

Pearson, K.: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philosophical Magazine Series, volume 5, 302, 157-175, 1900.

Pechlivanidis, I.G., Arheimer, B., Donnelly, C., Hundecha, Y., Huang, S., Aich, V., Samaniego, L., Eisner, S. and Shi, P.: Analysis of hydrological extremes at different hydro-climatic regimes under present and future conditions, Climatic Change, 141:467-481, doi:10.1007/s10584-016-1723-0, 2017.

Peel, M. C. and Blöschl, G. (2011), Hydrological modelling in a changing world. Progress in Physical Geography 35(2), 249–261, doi.org/10.1177/0309133311402550.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C. and Mathevet, T.: Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, Hydrological Sciences Journal 52(1), 131-151, doi:10.1623/hysj.52.1.131, 2007.

Piccolroaz, S., Majone, B., Palmieri, F., Cassiani, G., and Bellin, A.: On the use of spatially distributed, time-lapse microgravity surveys to inform hydrological modeling, Water Resour. Res., 51, 7270-7288, doi:10.1002/2015WR016994, 2015.

Piccolroaz, S., Di Lazzaro, M., Zarlenga, A., Majone, B., Bellin, A. and Fiori, A.: HYPERstream: a multi-scale framework for streamflow routing in large-scale hydrological model, Hydrol. Earth Syst. Sci., 20, 2047–2061, doi:10.5194/hess-20-2047-2016, 2016.

Protter, M.H., and Morrey, C.B.: College Calculus with Analytic Geometry, Addison-Wesley VLSI Systems Series, Addison-Wesley Publishing Company, 1977.

Rango, A., and Martinec, J.: Revisiting the degree-day method for snowmelt computations. J. Am. Water Resour. Assoc., 31 (4), 657–669, doi.org/10.1111/j.1752-1688.1995.tb03392.x, 1995.

Rinaldo, A., Marani, A., and Rigon, R.: Geomorphological dispersion, Water Resour. Res., 27, 513–525, doi:10.1029/90WR02501, 1991.

Schaefli, B., and H. V. Gupta: Do Nash values have value?, Hydrol. Processes, 21(15), 2075-2080, doi.org/10.1002/hyp.6825, 2007.

Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed? Hydrology and Earth System Sciences 13(6), 883–892, doi:10.5194/hess-13-883-2009, 2009.

Smirnov, N.V.: Estimate of deviation between empirical distribution functions in two independent samples. (Russian). Bull. Moscow Univ. 2(2), 3–16 (6.1, 6.2), 1939.

Taye M.T, Ntegeka V., Ogiramoi N.P., Willems P.: Assessment of climate change impact on hydrological extremes in two source regions of the Nile River basin. Hydrol Earth Syst Sci 15:209–222, doi:10.5194/hess-15-209-2011, 2011.

Thirel, G., Andréassian, V., Perrin, C., Audouy, J.-N., Berthet, L., Edwards, P., Folton, N., et al.: Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. Hydrol. Sci. J., 60:7-8, 1184-1199, doi:10.1080/02626667.2014.967248, 2014.

Thornton, P.K., Ericksen P.J., Herrero M., and Challinor A.J.: Climate variability and vulnerability to climate change: a review, Global Change Biology 20, 3313–3328, doi:10.1111/gcb.12581, 2014.

Todd, M. C., Taylor, R. G., Osborn, T. J., Kingston, D. G., Arnell, N. W., and Gosling, S. N.: Uncertainty in climate change impacts on basin-scale freshwater resources - preface to the special issue: the QUEST-GSI methodology and synthesis of results, Hydrol. Earth Syst. Sci., 15, 1035-1046, doi:10.5194/hess-15-1035-2011, 2011.

Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J.: Climate non-stationarity - validity of calibrated rainfall-runoff models for use in climate change studies, J. Hydrol. 394 (3-4), 447-457, 16/j.jhydrol.2010.09.018, 2010.

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I.G., Plötner, S., Seidou, O., Su, B., Vervoort, R.W. and Krysanova, V.: Evaluation of sources of uncertainty in projected hydrological changes under climate change in large-scale river basins, Climatic Change, 141:419–433, DOI:10.1007/s10584-016-1794-y, 2017.

Vogel, R. M., and Fennessey, N. M.: Flow-Duration Curves. 1: New Interpretation and Confidence-Intervals, J. Water Res. Planning and Management, 120(4), 485-504, doi:10.1061/(ASCE)0733-9496(1994)120:4(485), 1994.

Vrzel, J., Ludwig, R., Gampe, D., and Ogrinc, N.: Hydrological system behavior of an alluvial aquifer under climate change, Sci. Total Environ., 649, 1179-1188, https://doi.org/10.1016/j.scitotenv.2018.08.396, 2019.

Wang, W., Chen, X., Shi, P., and van Gelder, P.H.A.J.M.: Detecting changes in extreme precipitation and extreme streamflow in the Dongjiang River Basin in southern China, Hydrol. Earth Syst. Sci., 12, 207–221, doi.org/10.5194/hess-12-207-2008, 2008.

Wang, A., and Solomatine, D.P.: Practical Experience of Sensitivity Analysis: Comparing Six Methods, on Three Hydrological Models, with Three Performance Criteria, Water, 11, 1062; doi:10.3390/w11051062, 2019.

Weibull, W.: A statistical theory of strength of materials., Ing. Vetensk. Akad. Handl., 151, 1-45, 1939.

Westerberg I.K., Guerrero, J.L., Younger, P.M., Beven, K.J., Seibert, J., Halldin, S., Freer, J.E. and Xu, C.Y.: Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205–2227, doi:10.5194/hess-15-2205-2011, 2011.

Wilby, R. L. and Harris I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resour. Res., 42, W02419, https://doi.org/10.1029/2005WR004065, 2006.

Wilcke, R.A.I. and L. Bärring: Selecting regional climate scenarios for impact modelling studies, Environ. Model. Softw., 78, pp. 191-201, 10.1016/j.envsoft.2016.01.002, 2016.

Wu, Q., Liu, S., Cai, Y., Li, X. and Jiang, Y.: Improvement of hydrological model calibration by selecting multiple parameter ranges, Hydrol. Earth Syst. Sci., 21, 393–407, doi:10.5194/hess-21-393-2017, 2017.

Yang, W., Andréasson, J., Graham, L.P., Olsson, J., Rosberg, J. and Wetterhall, F.: Distribution based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, Hydrol. Res., 41, pp.211-229, 10.2166/nh.2010.004, 2010.

Yapo, P.O., Gupta, H.V., Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. J. Hydrol. 181, 23–48. http://dx.doi.org/10.1016/0022-1694(95)02918-4, 1996.

Zolezzi, G., Bellin, A., Bruno, M.C., Maiolini, B. and Siviglia, A.: Assessing hydrological alterations at multiple temporal scales: Adige River, Italy, Water Resour. Res., 45, W12421, doi:10.1029/2008WR007266, 2009.