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 the use of hydrological models with parameters inferred by means of calibration procedures using ~~observational~~observed meteorological data ~~of daily streamflow.~~as input forcing. In this work we show that this ~~is an error prone~~ procedure ~~when the interest is~~can lead to ~~develop reliable Empirical Cumulative Distribution Function curves~~a biased evaluation of ~~annual~~the probability distribution of high streamflow ~~maximum~~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 ~~is carried out by directly targeting the probability distribution of high flow extremes. In particular, hydrological simulations conducted during a reference period~~, as driven by climate models' outputs, ~~are constrained to maximize~~is carried out by maximizing the probability that the ~~modeled~~modelled and observed high ~~flow~~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 produce 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 ~~modeling~~modelling

1 Introduction

The recognition that an altered climate may affect severely water availability, floods and droughts, ~~or other water related resources and sectors,~~ led in the past decades to a wealth of climate change impact assessment studies ~~in hydrological literature~~.

A number of studies investigated the likely impact of climate change on hydrology by combining ensemble of projections from multiple climate models under different greenhouse gas emissions scenarios and hydrological ~~modeling~~modelling [e.g., Kundzewicz et al., 2007; Todd et al., 2010 and Wilby and Harris, 2006 for a comprehensive review]. A wealth of studies ~~focus~~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 ~~address~~studies addressed projected changes in 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 peculiarity of hydrological calibration in climate change impact studies has been highly debated in the hydrological modelling community [e.g. Peel and Blöschl, 2011; Muñoz et al., 2013; Montanari et al., 2013; Thirel et al., 2014]. ~~The standard~~ According to the most used approach ~~is to calibrate~~ the ~~selected~~ hydrological model ~~using a chronological time series of~~ is first calibrated against the observed streamflow ~~observations and then feed the model with future climate projections to evaluate changes of~~ by using observed meteorological data as input. The calibrated hydrological model is then run with the output of climate models as input to assess the projected changes of selected indicators ~~[see~~, 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]. ~~However, several~~The drawbacks of such approach are, however, twofold: i) a model correctly reproducing the time series of observed streamflow does not guarantee the correct reproduction of the desired statistics for extremes; and ii) because of epistemic uncertainty, a model calibrated with a given set of observations may respond in a different way when fed with projections obtained from climate change scenarios. Concerning this latter aspect, a number of studies ~~showed~~evidenced that model parameters are highly dependent on the climatic characteristics of the input forcing used ~~during~~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.

~~Indeed, a number of~~Several studies ~~suggest~~suggested that observed streamflow extremes provide valuable information about the hydrological ~~behavior~~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 ~~in the context of~~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 do 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 to these estimates [Honti et al., 2014].

~~Despite the above evidences, the typical approach adopted in impact assessments of hydrological extremes is to estimate the projected changes of selected indicators (e.g. flow quantiles) using a hydrological model calibrated against chronological time series of streamflow observations during a reference period [e.g., Wilby and Harris, 2006; Hattermann et al. 2018]. The~~

~~drawbacks of such approach are however threefold: i) a model correctly reproducing the time series of observed streamflow does not guarantee the correct reproduction of the desired statistics for extremes, as discussed above; ii) due to epistemic uncertainty, a model calibrated with a given set of observed ground data may respond in a different way when fed with projections obtained from climate change scenarios; and iii) due to the impossibility of obtaining totally unbiased climate simulations there is no a-priori guarantee that simulations fed by climate models produce samples (e.g. time series of simulated annual streamflow maximum) that are statistically coherent with observations.~~

To overcome the aforementioned limitations, we propose an innovative methodology in which the calibration of a physically-based hydrological model, as driven by climate models, is conducted by ~~directly targeting~~maximizing the probability ~~distribution of high~~that the modelled and observed streamflow extremes ~~(i.e., the Empirical Cumulative Distribution Function, ECDF, of annual maxima).~~belong to the same population within the reference period. While the approach is exemplified in this work for high streamflows (also because of 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~~. In particular, hydrological simulations conducted during a reference period, as driven by climate models, are constrained to maximize the chances that the simulated and observed high flow extremes belong to the same population. Statistical coherence is obtained by using the two-sample Kolmogorov-Smirnov statistic [Smirnov, 1939] as the efficiency metric during the calibration procedure. The parameterization of the hydrological model obtained following this approach is then used in future climate change scenario runs to project changes in the distribution of high flows. The strengths of the methodology are highlighted by performing a comparison with experiments in which model parameterizations are obtained by calibrating the hydrological model, fed by observed ground data, with a suite of efficiency metrics customarily used in hydrological applications, i.e., Nash-Sutcliffe efficiency, [Nash and Sutcliffe,1970] and flow duration curve-related metric [Westerberg et al., 2011~~ and relies on the use of the two-sample Kolmogorov-Smirnov statistic [Smirnov, 1939] as 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 if simulated hydrological variables are distributed according to a given probability distribution [e.g., Kleinen and Petschel-Held, 2007], to detect changes in hydrological variables [e.g., Wang et al., 2008], or to understand if calibrated parameters of hydrological models belong to a given probability distribution [e.g., Wu et al., 2017; Wang and Solomatine, 2019~~We emphasize that the suggested approach is by definition "*goal-oriented*", according to the definition discussed in Fiori et al. [2016], Savoy et al. [2017], Guthke [2017], Laiti et al., [2018] and Li et al. [2018]. In other words, the selection of the hydrological model, its level of complexity and the metric to be employed for the evaluation of the statistical coherence of simulated and observed extremes depend on modeler's choice and the particular goal at hand. In the present work the main focus is on high streamflow extremes, and along the goal-oriented approach the proposed calibration procedure is tailored around it.~~

~~The ~~], are relatively common in the literature. This notwithstanding, we are not aware of any study adopting this statistical test in the context of hydrological model calibration on 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 as input meteorological forcing. On the other, the strength of the methodology is checked by performing a comparison with experiments in which model parameterizations are obtained calibrating the hydrological model by using observed streamflow and meteorological data (standard procedure) with a suite of objective functions customarily used in hydrological applications.

The paper is organized as follows: ~~Section~~Sect. 2 presents the hydrological modelling framework, the calibration ~~framework,~~metrics and the adopted statistical ~~coherence test and the calibration metrics~~test; a description of the study area, the climate change projections~~ available and~~, the observational hydro-meteorological datasets ~~used~~ and the simulations set-up are summarized in ~~Section~~Sect. 3. The main findings are presented and discussed in ~~Section~~Sect. 4, whereas conclusions are finally drawn in ~~Section~~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 ~~SCS-CN~~ module for surface and subsurface flow generation~~ [Michel et al., 2005].~~. HYPERstream routing scheme is specifically designed for being easily coupled 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] ~~can be achieved~~is obtained by the application of suitable transfer functions derived from a high-resolution Digital Elevation Model of the study area. ~~The~~ Surface flow is computed by using the continuous ~~soil-moisture accounting~~ SCS-CN model ~~for surface~~[Michel et al., 2005], which receives as input the total precipitation given by the sum of rainfall and ~~subsurface flow generation, based on the SCS-CN methodology, is~~ snow melting evaluated by the degree-day model coupled with mass balance for taking into account snow accumulation [Rango and Martinec, 1995]. The remaining flow enters into a non-linear bucket ~~model for~~mimicking soil moisture dynamics [Majone et al., 2010~~], a degree-day model for snow melting and accumulation [Rango and Martinec, 1995] and~~]. Evapotranspiration is computed by the Hargreaves and Samani [1982] model~~ for evapotranspiration.~~. Furthermore, deep infiltration enters a linear bucket used to represent return flow. ~~Notice that the~~The surface and subsurface flow generation module was already successfully applied in ~~two~~ 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 as spatially uniform but uncertain and all subject to calibration. 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 ~~modeling~~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	Units	Parameters 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 Evaluation of statistical coherence~~

~~Statistical coherence between the observed and simulated populations of extremes was evaluated by means of 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 is defined as the maximum absolute distance, D_n, between the simulated (F_s) and observed (F_o) ECDFs of annual daily streamflow maximum (Q^M):~~

$$D_n = \max_{i\in[1,n]} \left| F_s\left(Q^M_{s,(i)}\right) - F_o\left(Q^M_{o,(i)}\right) \right|, \qquad (1)$$

~~where i is the position of $Q^M_{s,(i)}$ and $Q^M_{o,(i)}$ in the ranked samples of the simulated (s) and observed (o) annual streamflow maxima, respectively, and n is the number of years considered in the simulation (29 values in this work, one for each year of the investigated period excluding the first two, see Sect. 2.4). As customary in statistics $Q^M_{s,(i)}, i = 1, \ldots, n$ indicates the ranked~~

~~time series of the annual maxima $Q^{M}_{s,i}$ of simulated streamflow. A similar definition has been introduced for observed streamflow.~~2.2 ~~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. The p-value is the probability of rejecting the null hypothesis when it is true. It can also be defined as the smallest significance level α_s at which the null hypothesis would be rejected [Conover, 1999]. In simpler terms, the larger the p-value the stronger is the evidence in favor of the null hypothesis, i.e., in our work that the samples are drawn from the same distribution.~~

~~In our framework the evaluation of statistical coherence of observed and simulated populations of extremes is performed a-posteriori for each simulation experiment described in Section 2.4. In particular, the p-value will be used as a measure of the statistical coherence between simulated and observed high streamflow extremes.~~

~~2.3 ECDF computation~~

~~The daily average annual streamflow maxima are extracted from the chronological daily time series of observed (o) and simulated (s) streamflow and their empirical probability, *F*, is computed separately according to the classic Weibull formulation [Weibull, 1939]:~~

$$\cancel{F_j\left(Q^{M}_{j,(i)}\right) = \frac{i}{n+1}, \quad j = o, s, \quad i \in [1, n].} \qquad \text{~~(2)~~}$$

~~Empirical probabilities provided by Eq. (2) are used in Eq. (1) to compute the Kolmogorov-Smirnov statistic D_n.~~

~~2.4~~ Hydrological model calibration

The HYPERstreamHS hydrological model was calibrated against streamflow observations using as meteorological forcing both an observational dataset (i.e. ADIGE, see Sect. 3.2) and three climate models each one under two emission scenarios. A short description of these datasets is provided in Sect. 3.3. ~~Calibrations were performed during the period 1980-2010, assumed as reference, with the first two years used as spin-up and therefore excluded from the computation of model performances.~~ Parameters were inferred by optimizing three efficiency metrics by means of the Particle Swarming Optimization ~~algorithm [Kennedy and Eberhart, 1995].~~(PSO) algorithm [Kennedy and Eberhart, 1995]. PSO is an iterative method 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 locations are first randomly initialized and then iteratively updated in the search of the optimal solution, with the location updating procedure considering the memory of all locations visited by the whole collection of particles.

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

$$NSE = \max_{\boldsymbol{\theta} \in P^q} \left(1 - \frac{\sum_{i=1}^{m} \left(Q_{s,i}(\boldsymbol{\theta}) - Q_{o,i}\right)^2}{\sum_{i=1}^{m} \left(Q_{o,i} - \bar{Q}_o\right)^2}\right), \qquad (\cancel{3}1)$$

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 ~~value~~ 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 forming the parameter space P^q. ~~A daily time step is used here.~~ Since this metric ~~is sensitive to~~considers the chronological time series of simulated and observed daily streamflow, it was applied only in the simulations calibrated by using the observational dataset ADIGE as meteorological input.

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

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

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 ~~(ranked from the larger to the smaller value)~~ 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], ~~we employed~~ the so-called volume method was employed in which EPs are identified as the upper boundary of the elements obtained by partitioning the area below the curve in n_{EP} elements such that each of them is characterized by the same water volume. 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 under the curves and the water volume of each interval differ between observations and simulations. The water volume pertaining to each interval as well 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 (~~4~~2) irrespective of the integration scheme [Vogel and Fennessey, 1994].

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

$$KS = \min_{\boldsymbol{\theta} \in P^q}(D_n) = \min_{\boldsymbol{\theta} \in P^q} \left(\max_{i \in [1,n]} \left| F_s\left(Q_{s,(i)}^M(\boldsymbol{\theta})\right) - F_o\left(Q_{o,(i)}^M\right) \right| \right), \qquad (\text{~~defined in Eq. (1):~~}$$

$$\cancel{KS = \min_{\boldsymbol{\theta} \in P^q}(D_n) = \min_{\boldsymbol{\theta} \in P^q} \left(\max_{i \in [1,n]} \left| F_s\left(Q_{s,(i)}^M(\boldsymbol{\theta})\right) - F_o\left(Q_{o,(i)}^M\right) \right| \right),} \qquad (\cancel{5}3)$$

where F_s and F_o are the simulated and observed ~~ECDF as introduced in Eq. (2), and~~Empirical Cumulative Distribution Functions (ECDFs) of the simulated, $Q_{s,(i)}^M(\boldsymbol{\theta})$ ~~$= \max(Q_{s,k}(\boldsymbol{\theta}),\ k \in i-th\ year)$ and~~ , and observed, $Q_{o,(i)}^M$ ~~$= \max(Q_{o,k},\ k \in i-th\ year)$ are the simulated and observed~~, samples of daily average annual streamflow maxima ~~of the *i-th* year~~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, their ECDFs are computed according to the classic Weibull formulation [Weibull, 1939]:

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

This metric, which is at the core of the proposed approach, aims to maximize the ~~chances~~probability that the ~~modeled~~modelled and observed samples of ~~extreme~~ high ~~flows~~streamflows extremes belong to the same population. In other words, among all possible sets of model parameters we consider the one leading to the ~~sample of simulated annual maxima with 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 the R_{FDC} case, 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 by means of 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~~(i~~ [Conover, 1999]. The larger the p-value the stronger is the evidence in favor of the null hypothesis, i.e., ~~KS metric). Since KS is not sensitive to the temporal sequence of observed and simulated streamflows, similar to the R_{FDC} case, it has been applied to climate projections in addition to the simulations with the observational dataset ADIGE.~~ that the samples are drawn from the same distribution.

In this study 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.~~5~~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, $P(Q \le q) = exp\left[-exp[-\beta(q-u)]\right]$, 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. ~~The adaptation~~Extrapolation of high quantiles (i.e., estimation of quantiles for a return period beyond the ~~Gumbel distribution to the annual maxima of the~~available number of simulation years) of observed and simulated ~~daily~~annual streamflow ~~was~~maxima were then performed ~~a-posteriori for comparison purposes in order to extrapolate high flow quantiles (i.e., high return periods)~~ for all the simulation experiments ~~presented~~described in Sect.3.4.

Confidence intervals of observed streamflow ECDF were computed by means of 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 ~~and~~, hydro-climatic datasets and simulations set-up

3.1 Study area

To exemplify the application of the methodology ~~we selected as a case study~~ the upper ~~portion~~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 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 watershed, 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.

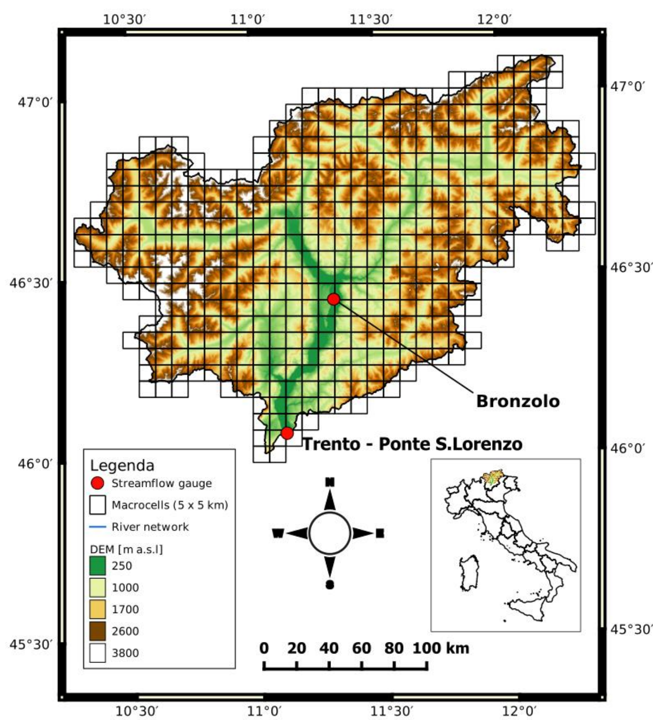


Figure 1: Map of the Adige river basin, with the computational grid cells ("macrocells") superimposed to the Digital Elevation Model (DEM) and the river network. The streamflow gauging stations used in the study are marked with red dots. The inset shows the location of the Adige river basin within the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

The climate of the watershed is characterized by relatively dry and cold winters followed by humid summers and autumns. Streamflow is minimum in winter, when precipitation ~~falls~~falling 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 annual average 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 Adige streamflow regime by the second half of 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 of 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 an observational precipitation and temperature dataset~~.~~ during 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 by means of the kriging with external drift algorithm [Goovaerts, 1997; Journel and Rossi, 1989], with an exponential semivariogram and by using the 16 closest neighbouring stations ~~to generate~~in the ~~estimates~~linear combination providing the estimate. The spatial distribution model was selected by Mallucci et al. [2019] according to the leave-one-out cross-validation procedure, applied to ordinary kriging and kriging with external drift algorithms, in association with multiple semi-variogram models (i.e., Gaussian, spherical and exponential models) and different numbers of neighbouring stations (namely 8, 16 and 32 stations). An average absolute error of the daily estimates of about 1.32 mm for precipitation and 0.02°C for temperature is reported in Mallucci et al. [2019], comparable with the error estimates provided by other ~~ground data-based~~ datasets available in the Alpine region such as APGD [Isotta et al., 2014]. ~~To comply with~~ Daily streamflow at the ~~computational grid adopted by HYPERstreamHS model, ADIGE was aggregated by areal averaging to the 5 km grid depicted in~~Trento Ponte San Lorenzo 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).

~~Daily streamflow data collected at the Trento Ponte san Lorenzo gauging station (see Figure 1) during the period 1980-2010 were provided by the Hydrological Office of the Autonomous Province of Trento (www.floods.it).~~

3.3 Climate change projections

Climate projections ~~applied~~used in ~~this study~~the present work were derived from the combination of General Circulation Models (GCMs) and Regional Climate Models (RCMs) available from 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 in the hydrological ~~modeling~~modelling experiments, we adopted the model sub-selection proposed by Vrzel et al. [2019] whom applied a hierarchical clustering approach [Wilcke and Bärring, 2016] in selected European river basins (including the Adige) in order to reduce the number of available Climate Model (CM) simulations (i.e., GCM-RCM combinations) while preserving the variability of the ~~overall~~ ensemble of climate change signals. In particular, model reduction involved 5 steps: 1) identification of meteorological variables; 2) transformation of variables into orthogonal and therewith uncorrelated variables by means of singular vector decomposition; 3) identification of the optimum number of clusters; 4) hierarchical clustering to group the simulations; and finally, 5) selection of the ~~simulation~~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 ~~1~~2).

These three ~~Climate Models (CMs)~~GCM-RCM combinations provide ~~an assessment~~projections of likely future climate changes for the mid-term horizon 2040-2070, with the time window 1980-2010 selected as a 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 ~~three adopted CMs~~combinations, thereby leading to a total of six CMs which are investigated in the present study (see Table ~~1~~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. ~~As customarily done in most of the climate change impacts studies,~~In this work we rely on ~~bias-corrected~~ products retrieved from EURO-CORDEX~~ initiative. For the reference period 1989-2010 EURO-CORDEX products~~, which are available bias-corrected by the distribution-based scaling approach [DBS, Yang et al., 2010] using as ~~a reference~~observations the MESAN gridded reanalysis datasets of daily precipitation and temperature [Landelius et al., 2016]. CMs forcing in the reference period 1980-2010~~ 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 observed values; and ii) the correction ~~used in the~~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. ~~Finally, before using them as input in hydrological calibration experiments, the gridded CMs outputs were resampled by means of~~This is needed because MESAN data are available only for the ~~nearest neighbour method to the computational 5 km spacing grid used in the discretization of the Adige river basin (see the grid shown in Figure 1).~~former period.

Table ~~1~~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	

~~4 Results and discussion~~

~~4.1 Direct calibrations forced by CMs~~

~~Here we consider~~3.4 Simulations set-up

All the ~~results of hydrological~~ simulations ~~performed~~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 by means of the nearest neighbour method.

In a first set of simulations, presented in Sect. 4.1, the HYPERstreamHS model was calibrated at the Trento gauging station by using NSE, KS and R_{FDC} as objective functions during the period 1980-2010, which is assumed as reference. In order to ease the presentation of results, these three parameterizations are hereafter termed as 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) during the same time window, and at the Trento gauging station during 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 during the period 1980-2010 by using precipitations and temperature from the three GCM-RCM combinations selected as described in ~~Section~~Sect. 3.3 each one ~~with~~for both RCP4.5 and RCP8.5 ~~pathways~~emission scenarios, for a total of six CM combinations~~, as shown in~~ (see Table ~~1.~~2). The parameters of the hydrological model ~~are obtained by optimizing~~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 during the period ~~1982~~1980-2010~~. The ECDFs of the daily average annual streamflow maximum are extracted a-posteriori from the six optimized models and then compared with~~, and KS and R_{FDC} as objective functions. Following the procedure described in Sect. 2.4, extrapolations were performed under the assumption that ~~of the historical observations in the same period. The comparison is performed by applying~~simulated and observed ECDFs were distributed according to the parametric Gumbel probability distribution. Verification of statistical inference procedure was performed in all cases with the successful application of the Pearson's chi-squared test.

For all time windows and for 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 to the Kolmogorov-Smirnov two-sample test ~~between simulated and observed ECDFs (see Sect. 2.2), with the former obtained by optimizing the two metrics KS (Eq. (5)) and R_{FDC} (Eq. (4)), respectively.~~ described in Sect. 2.3.

~~Results~~The effects on model parameters of calibrations conducted using different input forcing (observational data as well CMs simulations) is 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, leads 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]. ~~2 show that for both RCP4.5 and RCP8.5 emission scenarios simulations performed by optimizing KS (Eq., (5)) provided samples~~ We remark that ~~with~~ 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 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

~~probability (i.e. *p*~~4.1 Simulations using the observational dataset ADIGE

Figure 2a 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. In a strict statistical sense all the three metrics provide simulated samples of annual streamflow maxima belonging to the same population of 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} ($p = 0.372$). At the same time, calibration conducted by using KS as 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 more sensitive to specific aspects of the time series with its own 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 ECDF of observed high flow extremes when the model is calibrated considering explicitly such an information. 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 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 lead to a small 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 during the calibration procedure. Simulated and observed ECDFs of annual streamflow maxima and the associated p-value of the Kolmogorov-Smirnov test are presented in Figure 2b. 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). In a strict statistical sense, 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 p-value from calibration to validation is significant, but rather common in hydrological models.

A 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 are only slightly different from those obtained in calibration (see Table 3). Furthermore, results presented in Figure 3b 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 parameterization obtained using KS as objective function is reliable, though relying on the use of a limited number of observations, and does not introduce distortion in the spatial representation of the hydrological processes, particularly for those associated to 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 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 and bold numbers indicate the metric optimized in calibration. p-values of the Kolmogorov-Smirnov test are also reported in the bottom line for the three calibration experiments and for the validations runs.

	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			0.772		
FDC-ADIGE	0.488	**0.975**	0.233		0.804			0.830	
KS-ADIGE	0.400	0.564	**0.067**			0.233			0.137
p-value	0.951	0.372	1.000	0.222	0.372	0.372	0.791	0.951	0.951

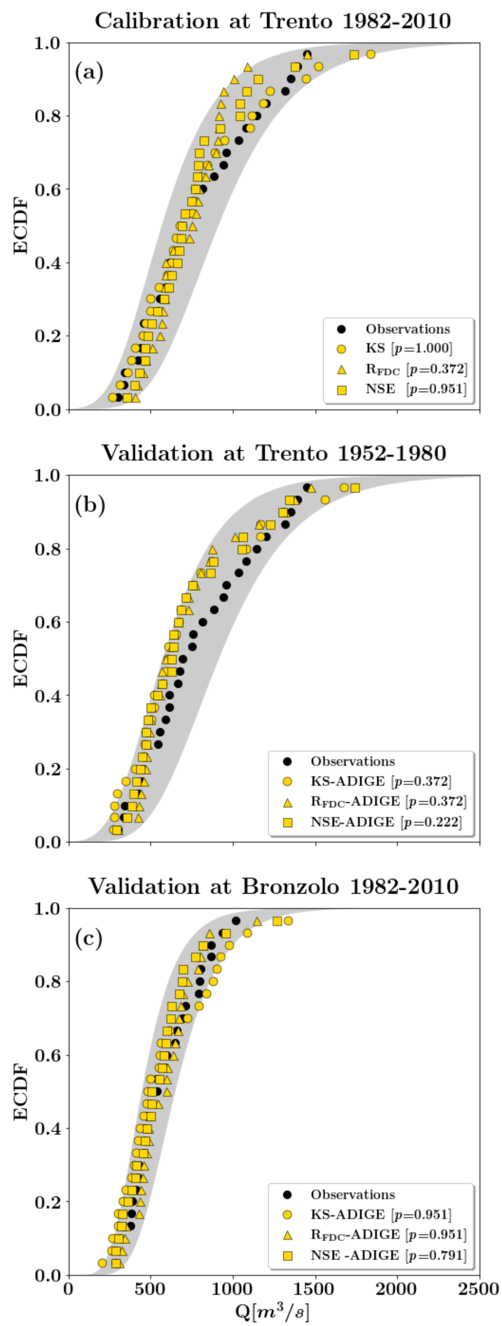


Figure 2: ECDFs of daily annual streamflow maximum obtained by using as input the observational dataset ADIGE and the parametrizations NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE at a) the Trento gauging station in the period 1982-2010; b) the Trento gauging station in the period 1952-1980, and c) the Bronzolo gauging station during the period 1982-2010. The experimental ECDF obtained from streamflow observations in the same time frames is 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 is run in the time frame 1982-2010 using as input the meteorological variables produced by the climate models with the three sets of parameters obtained by using ADIGE as input and NSE, R_{FDC} and KS as objective functions in calibration (i.e., NSE-ADIGE, R_{FDC}-ADIGE, KS-ADIGE, see Sect. 3.4). Visual inspection of Figures 3a, 3b and 3c 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 the Table 4. In particular, these 3 parameterizations lead 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 showing a slightly better performance: $p = 0.372$ for KNMI and SMHI under the RCP8.5 scenario (see Figures 3b and 3c and Table 4). Inspection of Table 4 also reveals that $p < 0.05$, and thereby the simulated ECDFs do not belong to the same population of the measured one, for the CLMcom model with 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 3a, 3b and 3c. These results are in agreement with previous studies evidencing that the hydrological models, calibrated against observed data, that performs 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]. ~~observed values for all the 3 CMs used.~~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 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) belong to the same population of the observed values. On the other hand, ~~the optimization of R_{FDC} leads to~~ when R-$_{FDC}$ is used as objective function, simulations lead to samples belonging to the same population with probability larger than p~~=~~ = 0.05 (i.e. the level of significance customarily adopted in the statistical literature to reject the null hypothesis), though significantly lower than for KS (~~fourth column~~sixth and seventh columns in Table ~~2~~4). The lowest p-value is obtained when the calibration is performed with R_{FDC} metric and by using the climate model CLMcom with RCP4.5 (p = *0.222*, see ~~Figure 2a and~~ Table ~~2~~4). Consistently, the absolute maximum ~~distance~~distances between the ECDF of observed and simulated ~~sample distributions (i.e., D_n)~~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 as objective function 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 fourth column in Table ~~2). The consequence in terms~~4). Since R_{FDC} employs the entire time series of ~~correspondence~~observational data, this result evidences that using KS metric during calibration avoids model's overparameterization, despite the limited number of observational data (i.e., 29 values of observed daily annual streamflow maxima).

The appreciable difference between observed and simulated ECDFs ~~is indeed appreciable.~~obtained in the calibration experiments conducted using KS and R_{FDC} metrics is highlighted in Figure 4. The ECDFs obtained from simulations ~~performed by adopting~~employing KS ~~as optimization criteria~~ are indeed in a better agreement with the observed ECDF than the ECDFs ~~obtained from simulations that use~~employing R_{FDC}, as ~~optimization criterion (see~~ showed in all subplots ~~in~~of Figure ~~2).~~4. These results also highlight that ~~adopting~~ the KS metric is preferable than ~~using~~ R_{FDC} when dealing with high flow extremes, thus strengthening the ~~idea of targeting~~approach envisaged here of addressing directly the desired statistics ~~for~~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 the different calibration experiments and for the validations conducted using NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE parameterizations.

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	RCP45	**0.943**	0.267	0.730	**0.067**	0.222	1.000	0.030	0.222	0.030
KNMI	RCP45	**0.940**	0.167	0.804	**0.067**	0.791	1.000	0.013	0.030	0.123
SMHI	RCP45	**0.972**	0.200	0.589	**0.067**	0.572	1.000	0.222	0.123	0.123

CLMcom	RCP85	0.980	0.200	0.449	0.067	0.572	1.000	0.123	0.372	0.222
KNMI	RCP85	0.961	0.167	0.456	0.067	0.791	1.000	0.123	0.222	0.372
SMHI	RCP85	0.932	0.167	0.484	0.067	0.791	1.000	0.123	0.372	0.123

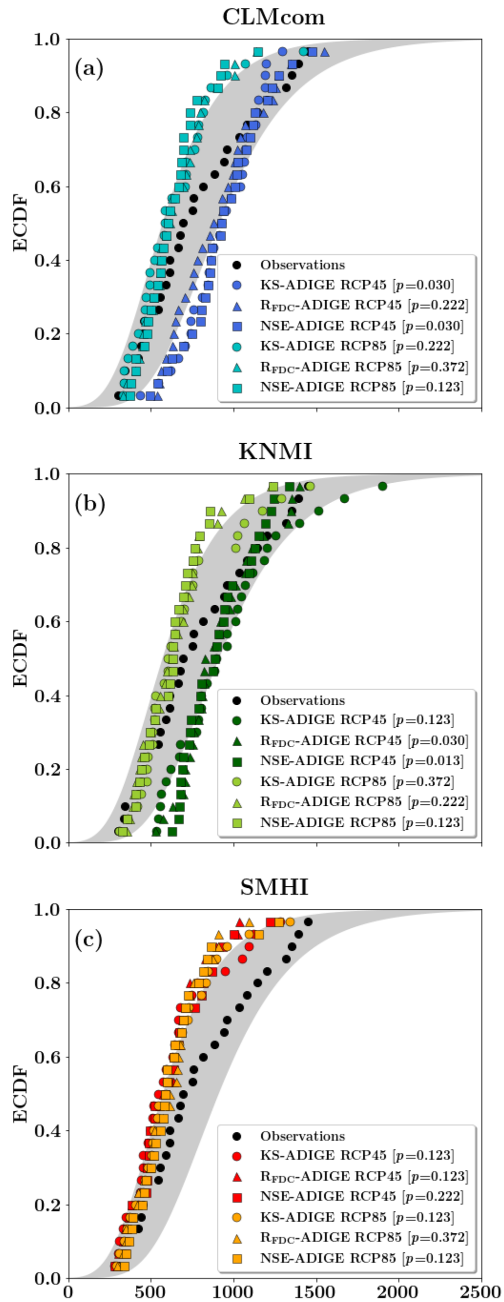


Figure ~~2~~3: ECDFs of annual maximum ~~daily~~ streamflow at Trento gauging station ~~during~~in the period 1982-2010 obtained by using NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE parameterizations and a) CLMcom ~~(first row),~~, b) KNMI ~~(second row),~~, and c) SMHI

~~(third row)~~ climate models as input forcing ~~under the RCP4.5 (left) and RCP8.5 (right) emission scenarios. Calibrations have been performed using~~ for both ~~KS~~RCP4.5 and ~~RFDC metrics~~RCP8.5 emission scenarios. The ECDF of observations is also shown with black dots together with the associated 90% confidence interval ~~obtained by bootstrapping as described in Section 2.5~~of the fitted Gumbel distribution (grey shaded area). p-values of the Kolmogorov-Smirnov two-sample test are also reported within brackets ~~in the legend.~~ for each simulation run.

~~The above findings can be cast and discussed in the context of the hydrological literature of hydrological extremes, and in the following we further expand the discussion presented in the Introduction in light of the results obtained so far. Climate change effects on extremes are typically assessed by using a hydrological-climatological modeling chain [Wilby and Harris, 2006] in which the adopted hydrological model is calibrated against chronological time series of observed streamflows during a reference period, and then used with GCM-RCM future projections as meteorological input to assess projected changes of the extremes [see e.g. Ngongondo et al., 2013; Aich et al., 2016; Pechlivanidis et al., 2017; Vetter et al., 2017; Hattermann et al. 2018]. Calibration of the hydrological model by targeting directly the statistics of extremes (e.g. high flow quantiles or KS metric as in our case) are indeed much less common in hydrological applications [see e.g. Honti et al., 2014]. This distinction may appear unnecessary since it seems reasonable to assume that a model correctly reproducing the chronological time series of streamflow will also reproduce the statistics of interest (or distribution). However, unavoidable epistemic and parametric errors, impairs this implicit assumption. A possible alternative, we explore here, is to calibrate on the distribution of the considered extremes, i.e. the annual streamflow maxima in our case.~~

Examples of hydrological models calibrated by using tailored information instead of the entire observed streamflow series are present in the hydrological literature~~.~~ [e.g., Montanari and Toth ~~[~~, 2007~~] used the spectral properties of the streamflow time series as a goodness of fit metric.~~; Blazkova and Beven ~~[~~, 2009~~] used certain flow quantiles as acceptability criteria within a GLUE framework.~~; Westerberg et al~~. [~~., 2011~~] used an informal triangular likelihood function for calibrating a hydrological model on the basis of observed flow duration curves. Furthermore,~~; Lindenschmidt ~~(~~, 2017~~) used water stage frequency curves as objective functions for reproducing ice jam formation in northern rivers.~~]. However, these approaches are typically adopted for reproducing watershed response to observed meteorological forcing ~~under observed conditions,~~ and have not been applied (to our best knowledge) ~~using directly the~~in combination with GCM-RCMs simulations ~~as input forcing in the calibration procedure~~. The only example somewhat similar to our approach we found in literature is that of Honti et al. [2014], which 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 ~~(i.e., timing errors do~~, for which time shift does not influence the ~~model performance)~~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].

~~Studies adopting the two-sample Kolmogorov-Smirnov test to evaluate if simulated hydrological variables are distributed according to a given probability distribution [e.g., Kleinen and Petschel-Held, 2007], to detect changes in hydrological~~

~~variables [e.g., Wang et al., 2008], or to understand if calibrated parameters of hydrological models belong to a given probability distribution [e.g., Wu et al., 2017; Wang and Solomatine, 2019] are relatively common in the literature. This notwithstanding, we are not aware of any study adopting directly the KS metric in the context of hydrological model calibration on extremes.~~

~~The proposed method, coined here as Hydrological Calibration on Extremes (HyCoX), hence consists in the calibration of a physically-based hydrological model on the probability distribution of extremes (in our case the ECDF of the annual maximum~~Quantiles of daily ~~streamflow), i.e. the possibility to constrain simulated (driven by GCM-RCM runs during the reference period) and observed data samples to belong to the same population. In this respect, HyCoX also provides a framework for excluding CMs forcing datasets not coherent with hydrological observations of high flow extremes (i.e., cases with~~ ~~$p < 0.05$~~ ~~in the Kolmogorov-Smirnov two-sample test). Finally, the parameterization obtained for the hydrological model guaranties statistical coherence between scenarios and observations during the reference period, and can then be used in future climate change scenario runs to project changes in extremes.~~

~~Table 2: NSE, RFDC and KS efficiency metrics of the period 1982-2010 with forcing provided by ADIGE dataset and 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 coherence test are also reported for each calibration experiment.~~

-	-	-	~~*NSE*~~	~~*R_{FDC}*~~	~~*KS*~~	~~*p-value*~~
-	~~NSE-Adige~~	-	**~~0.822~~**	~~0.874~~	~~0.133~~	~~0.951~~
	~~FDC-Adige~~		~~0.488~~	**~~0.975~~**	~~0.233~~	~~0.372~~
	~~FDC-CLMcom~~	~~RCP4.5~~	~~--~~	**~~0.943~~**	~~0.267~~	~~0.222~~
	~~FDC-KNMI~~	~~RCP4.5~~	~~--~~	**~~0.940~~**	~~0.167~~	~~0.791~~
	~~FDC-SMHI~~	~~RCP4.5~~	~~--~~	**~~0.972~~**	~~0.200~~	~~0.572~~
	~~FDC-CLMcom~~	~~RCP8.5~~	~~--~~	**~~0.980~~**	~~0.200~~	~~0.572~~
	~~FDC-KNMI~~	~~RCP8.5~~	~~--~~	**~~0.961~~**	~~0.167~~	~~0.791~~
	~~FDC-SMHI~~	~~RCP8.5~~	~~--~~	**~~0.932~~**	~~0.167~~	~~0.791~~
	~~KS-Adige~~		~~0.400~~	~~0.564~~	**~~0.067~~**	~~1.000~~
	~~KS-CLMcom~~	~~RCP4.5~~	~~--~~	~~0.730~~	**~~0.067~~**	~~1.000~~
	~~KS-KNMI~~	~~RCP4.5~~	~~--~~	~~0.804~~	**~~0.067~~**	~~1.000~~
	~~KS-SMHI~~	~~RCP4.5~~	~~--~~	~~0.589~~	**~~0.067~~**	~~1.000~~
	~~KS-CLMcom~~	~~RCP8.5~~	~~--~~	~~0.449~~	**~~0.067~~**	~~1.000~~
	~~KS-KNMI~~	~~RCP8.5~~	~~--~~	~~0.456~~	**~~0.067~~**	~~1.000~~
-	~~KS-SMHI~~	~~RCP8.5~~	~~--~~	~~0.484~~	**~~0.067~~**	~~1.000~~

~~4.2 Forward simulations using parameterizations derived from calibrations with observed ground data~~

~~The typical way to assess the impact of climate change on hydrology is to run a model, calibrated with observed meteorological and hydrological data, with the meteorological forcing provided by climate models. This approach is pursued here by using HYPERstreamHS calibrated against daily streamflow in the period 1982-2010, with precipitation and temperature extracted from the ADIGE dataset [Mallucci et al., 2019], and meteorological input provided by the six aforementioned CM simulations. The main objective is to assess if a model well calibrated with observational data responds coherently when applied with precipitation and temperature obtained from climate models. Again, the comparison is performed by applying the Kolmogorov-Smirnov two-sample test between simulated and observed ECDFs. HYPERstreamHS was calibrated at the Trento gauging station by using NSE, KS and R_{FDC} metrics. In order to ease the ensuing discussions, these three parameterizations are hereafter termed as NSE-ADIGE, KS-ADIGE and R_{FDC}-ADIGE, respectively (see also Table 2).~~

~~Figure 3a shows the simulated ECDFs and the associated p-values of the Kolmogorov-Smirnov test for calibrations conducted with the ADIGE dataset as input and the three aforementioned metrics (NSE, KS and R_{FDC}). In a strict statistical sense all the three metrics provide simulated samples of~~ annual streamflow maxima ~~belonging to the same population of the observed ones, given that in all cases p>0.05, but with an evidence in favor of this conclusion that is maximum for KSE (p = 1.000) and minimum for R_{FDC} (p = 0.372). At the same time, calibration conducted by using KS as efficiency metric leads to NSE and R_{FDC} values (0.4 and 0.564, respectively, see Table 2) which are lower than those obtained optimizing the two metrics in the calibration (NSE = 0.822 and R_{FDC} = 0.975, see Table 2). 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 more sensitive to specific aspects of the time series with its own 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 ECDF of observed high flow extremes during the calibration of the hydrological model. 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 2).~~

~~When the 3 aforementioned parameterizations (i.e., NSE-ADIGE, R_{FDC}-ADIGE, KS-ADIGE) are used in the forward simulations of the reference period 1982-2010 using as input the meteorological variables produced by the climate models, the conclusions are significantly different (see Figures 3b, 3c, 3d). In particular, the adopted parameterizations lead to simulations that show low p-values (always lower than p = 0.372) for all the considered CMs and pathways. In particular, NSE-ADIGE and R_{FDC}-ADIGE show on average the lowest p-values, with KS-ADIGE showing a slightly better performance: p = 0.372 for KNMI and SMHI under the RCP8.5 scenario (see Figures 3c and 3d).~~

~~The above results highlight how the classical approaches based on feeding the hydrological model calibrated to observed streamflow data (i.e., by using the NSE metric) or to flow duration curve (i.e., by using the R_{FDC} metric), 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 3b, 3c and 3d. These results are in agreement with previous studies evidencing that the use of parameterizations of hydrological models (obtained with observational input forcing) providing a good reproduction during a baseline period is questionable for simulating streamflow under future climate conditions [Brigode et al., 2013; Lespinas et al., 2014]. Indeed, it is well recognized that the use of different datasets can lead to different optimized parameter sets 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 certainly affected by uncertainties in climate modeling, but also the calibration strategy adopted during the reference period plays a, often dominant, role [Lespinas et al., 2014; Mizukami et al., 2019]. In this respect, we believe that the preservation of statistical coherence between climate scenarios and observations (i.e., high flow extremes in our case) should be taken into account directly during hydrological calibration, at least in the reference period.~~

~~Figures 3b), 3c) and 3d) evidence that for high quantiles the simulated ECDFs are often outside the 90% confidence interval of the observed ECDF for all the considered forward simulations. This is further highlighted in Figure 4, which shows the quantiles of annual maximum daily streamflow~~ as a function of return period at the Trento gauging station~~. Following the procedure described in Section 2.5 extrapolations are performed under the assumption that the simulated ECDFs are distributed according to the parametric Gumbel probability distribution. All the adaptations to the simulated ECDFs passed the Pearson's chi-squared test. Confidence intervals of observed streamflow are derived from the parametric bootstrap procedure outlined in Sect. 2.5.~~ are shown in Figure 5, 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 ~~4~~5 shows that for all ~~investigated~~ return periods ~~parameterizations~~parametrizations obtained by ~~fitting~~ calibrating with the ~~model to~~ observed ~~streamflow with meteorological data~~precipitations and temperatures as provided by the ADIGE dataset ~~(i.e., NSE-ADIGE, R_{FDC}-ADIGE, KS-ADIGE)~~ significantly underestimate the ~~ECDF~~quantiles of ~~annual maximum~~the observations and fall outside the confidence interval of ~~observations~~the fitted Gumbel distribution (i.e., outside the grey area). The only exceptions are ~~curves~~ the quantiles derived from simulations conducted with KNMI (KS-ADIGE, dotted line in Figure ~~4c~~5c) and CLMcom (all the 3 metrics, Figure ~~4a~~5a) climate models under RCP4.5~~ pathway~~. 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). Quantiles obtained by calibrating the hydrological model with the meteorological input provided by the Climate Models and 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 generally are 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 represented by CLMcom and KNMI under RCP4.5 emission scenario and R_{FDC} as metric that present the largest deviations from observations (see Figures ~~4a~~5a and ~~4c~~5c, 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 ~~(gray~~of the fitted Gumbel distribution (grey area) widens as the return period increases and this is line with the recent findings of Meresa and Romanowicz [2017], which showed that errors in fitting theoretical distribution models to annual ~~extreme~~maxima streamflow series might contribute significantly to the overall uncertainty associated to projections of future hydrological extremes.

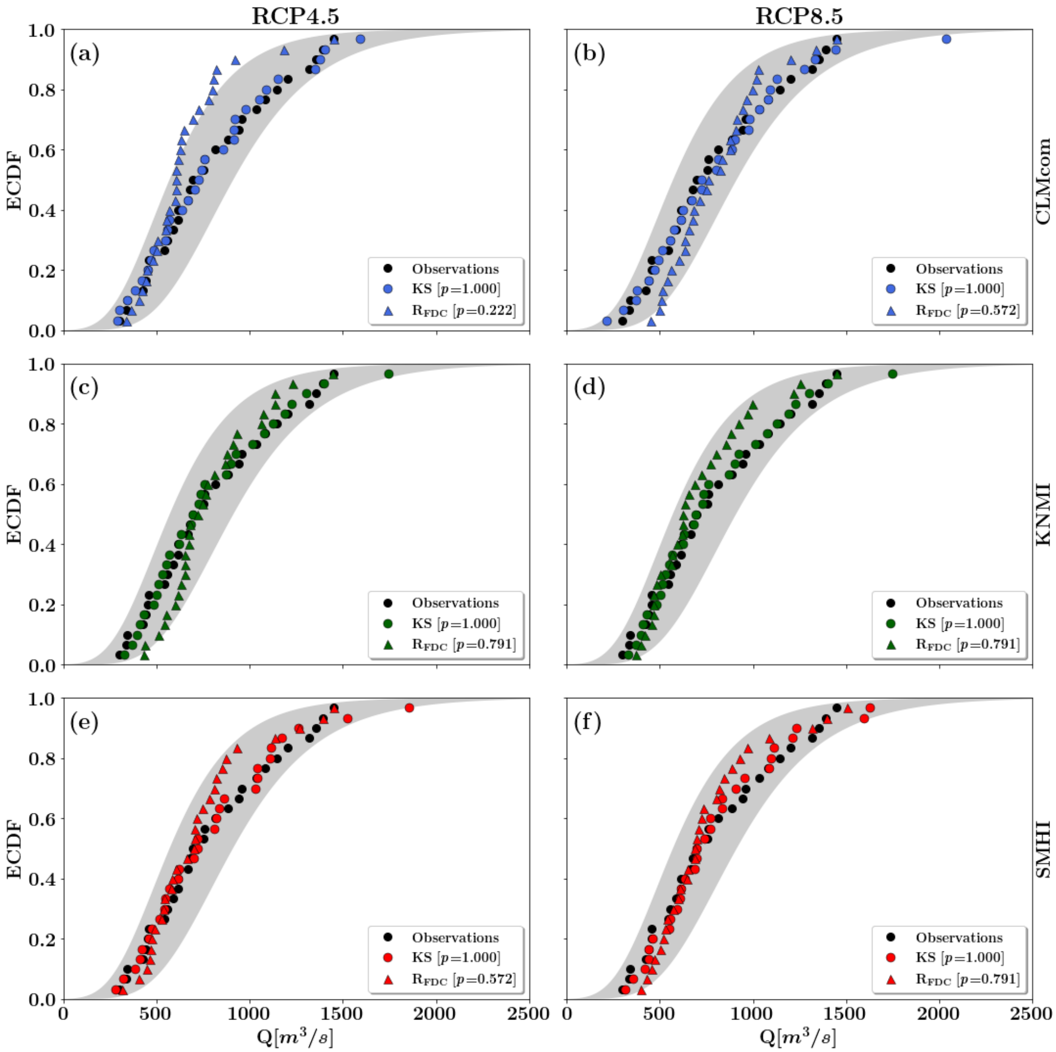


Figure 4: Simulated ECDFs of daily annual maximum streamflow at Trento gauging station in the period 1982-2010 with precipitation and air temperature provided by CLMcom (first row), KNMI (second row), and SMHI (third row) climate models under the RCP4.5 (left) and RCP8.5 (right) emission scenarios. Calibration 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.

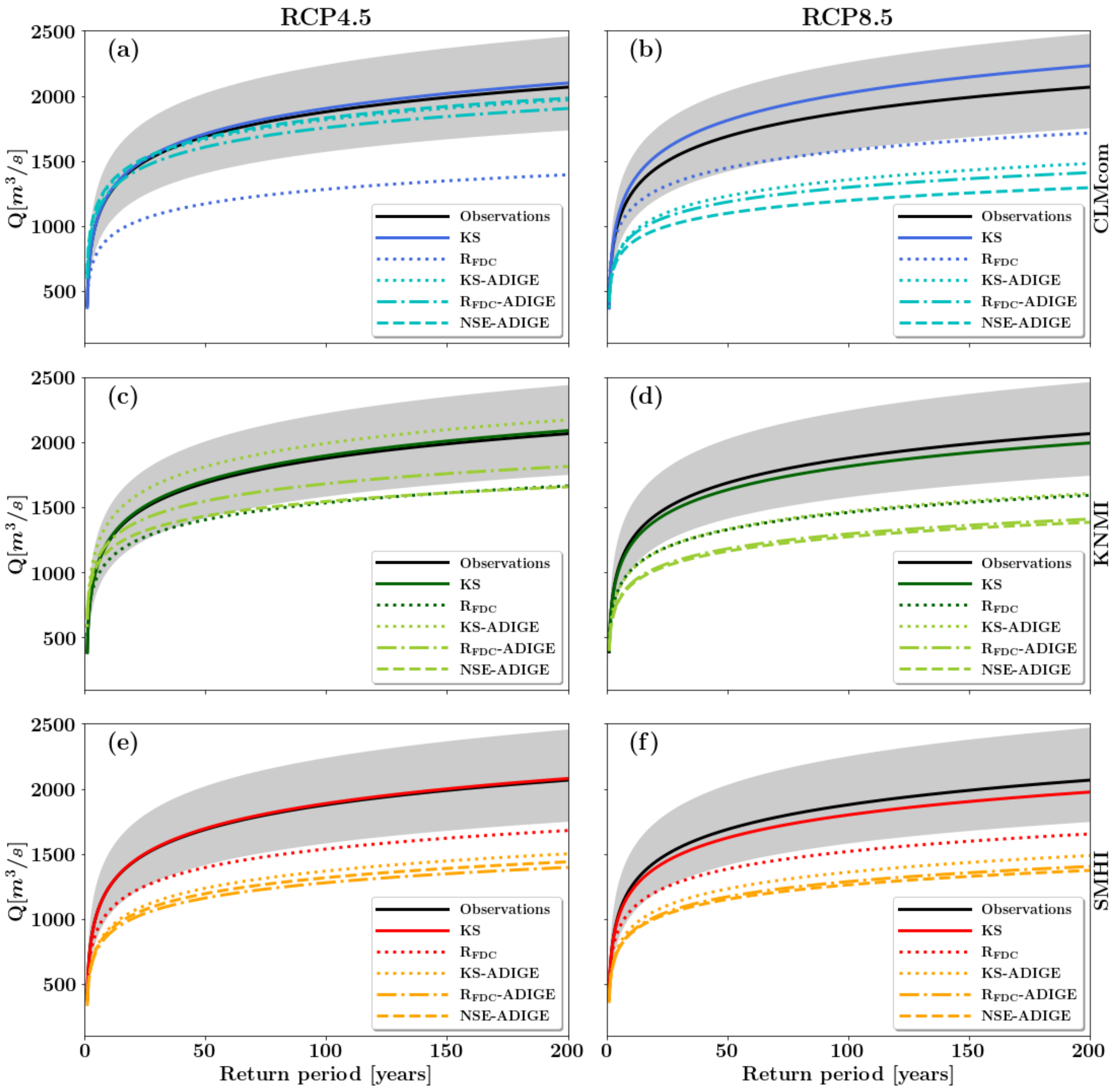


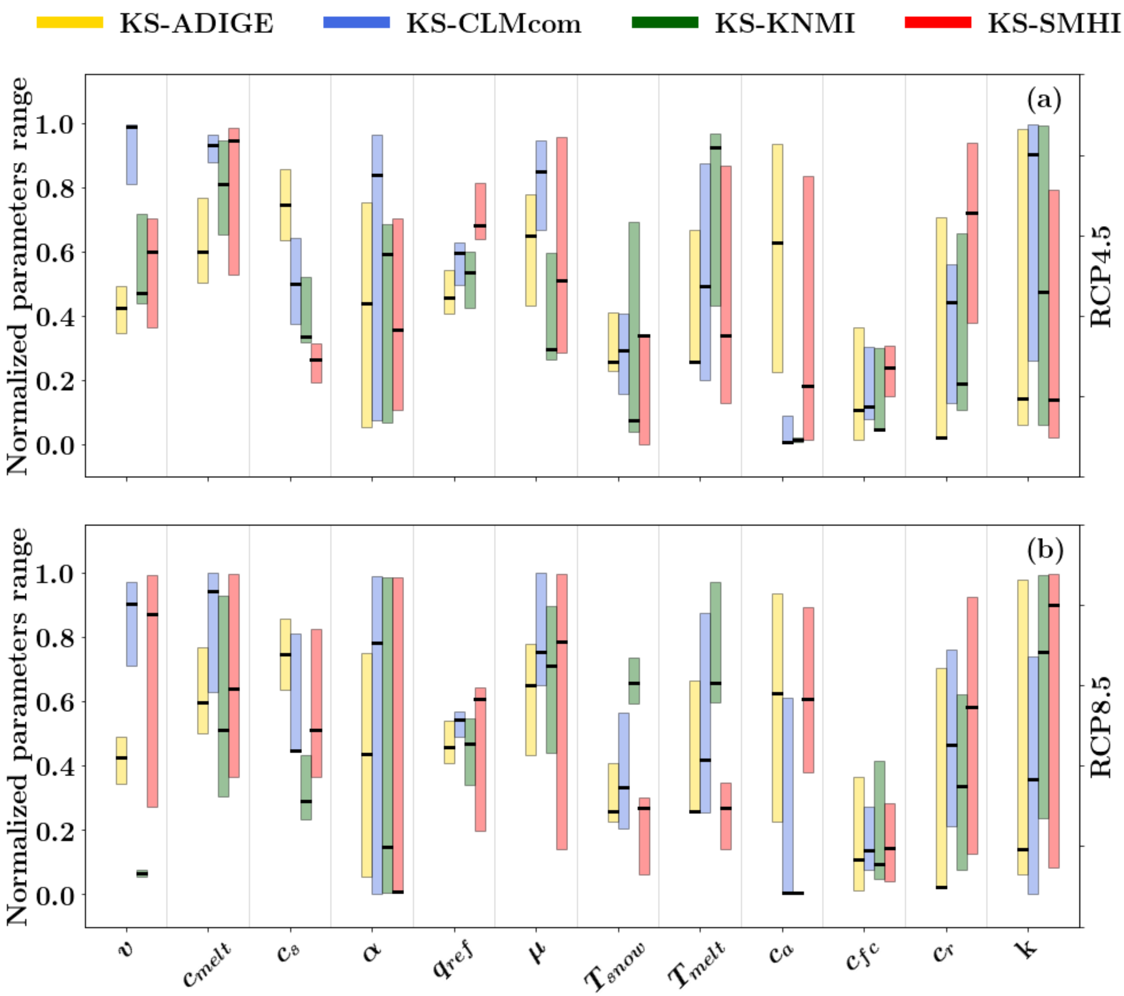
Figure 5: Quantiles of daily annual streamflow maxima as a function of return period at the Trento gauging station. Extrapolations are based on simulations conducted during the period 1982-2010 using as input forcing the CLMcom (first row), KNMI (second row), and SMHI (third row) climate models under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

4.~~3~~4 Model parameters

The results presented in the previous Sections highlight how the largest level of statistical coherence between observations and simulations (performed with CMs simulations as input) ~~can be~~is achieved ~~only~~ by optimizing the desired statistics of extremes (i.e., see the curves ~~labeled~~labelled KS in Figures ~~3~~4 and ~~4~~5) in the calibration of the hydrological model. Starting from this evidence, we investigated what is the effect on model parameters of performing the calibration by using ~~different input data (~~either observational data ~~as well~~or CMs simulations~~). The list of the 12 parameters of the model with their units together with a short description~~ as input data and ~~range of variation is presented in Table 3. Notice that all the 12 parameters of HYPERStreamHS are calibrated. To simplify the comparison, we considered for the 3 CMs and for both emission scenarios only the calibrations performed by using the~~ KS ~~metric. In particular, for each calibration run we considered the 200 simulations (out of 40000 performed in total) presenting the highest efficiency. In addition, we considered the 100% confidence bands resulting from the retained solutions, and used the distance between the upper and lower limits of the confidence interval, $\bar{d}$,~~ as ~~a metric of uncertainty for the calibrated parameter~~ objective function. ~~[see Piccolroaz et al., 2015].~~ ~~We remark that the procedure adopted here aims at quantifying only the differences in the calibrated parameters and not to perform a full uncertainty analysis of predictions.~~

Figure ~~5~~6 shows the range ~~of variability of the parameters~~, $\bar{d}$, between the maximum and minimum value of each parameter associated with the ~~retained solutions~~200 accepted sets of parameters (see Sect. 3.4), together with the corresponding optimal parameter ~~sets~~set. The values of the parameters are normalized with respect to their range (see Table ~~3). The boundary of the parameters space has been set by means of preliminary simulations~~1) such ~~as to minimize the probability of excluding from the search domain combinations of parameters leading to behavioral solutions [Beven and Binley, 1992]. Furthermore, we verified a-posteriori that the optimal parameters are inside the range of variation. In all cases, confidence bands are~~that they are directly comparable. In all simulations the range $\bar{d}$ is generally well distributed between 0 and 1, indicating a proper choice of ~~their~~the parameters range ~~of variation~~, although for a few parameters ~~optimality~~the optimal value was located close to the boundary of the search domain. As shown in Figure ~~5~~6 the majority of the parameters ~~have~~span a ~~confidence band~~range $\bar{d}$ that is similar in terms of amplitude (or slightly larger) to that obtained in the case of KS-ADIGE, thus supporting the conclusion that calibration using CMs simulations does not lead, for both ~~RCP~~RCPs, to ~~anomalous identification of model parameters.~~bias parameterizations. Figure ~~5~~6 also shows that for most of the parameters, simulations performed with CMs lead to generally overlapping ~~confidence bands~~ranges for $\bar{d}$ with respect to the case in which the observational dataset ADIGE is used. The largest deviations in terms of ~~confidence bands~~$\bar{d}$ 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 (see q_{ref}, μ, and c_{fc} in Figures ~~5a and 5b). Notwithstanding, this analysis evidences the possibility to identify to what degree each parameter is sensitive to the adoption of a different input in the calibration procedure.~~6a and 6b). Visual inspection of Figure 6 also highlights that the parameters controlling runoff generation and streamflow concentration (in particular, v, c_s, q_{ref}, and c_{fc}) present a very good identifiability (i.e., small range $\bar{d}$). This is not the case for parameters controlling snowmelting and groundwater contribution, the latter being relevant only for low flows conditions (see k in Figures 6a and 6b). 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 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., reduced number of model parameters, introduction of 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, results presented in this Section are in our view enough to consider the parameterizations derived from the use of KS metric as reliable.

The differences observed in the optimal ~~value~~values of model parameters are due to ~~structural errors in~~ the ~~GCMs and RCMs (i.e., their~~use of datasets presenting different capabilities to simulate the present climate~~), which are a substantial source of uncertainty in the impact assessment modeling chain [Honti et al., 2014; Tian et al, 2016].~~. 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 ~~GCM-RCM combination.~~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. ~~Furthermore, our approach provides an answer to the need of reducing uncertainty in climate change impact assessments recently highlighted in the review by Clark et al. [2016].~~



~~Figure 5: Normalized parameters range obtained by retaining~~ Figure 6: 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. ~~Marked~~Bold horizontal dashes indicate the optimal parameter sets for all experiments. ~~See Table 3 for the description of parameters and their ranges.~~

4.~~4~~5 Projected changes of streamflow quantiles

~~Here we evaluate projected changes of high flow extremes in the future period 2040-2070. Notice that, similar to the reference period, the first two years of simulations (i.e., 2040 and 2041) have been used as a spin-up period. Figure 6~~Figure 7 presents the annual maximum streamflow as a function of return period at Trento gauging station in the future time window 2040-2070 for the 3 ~~considered~~selected CMs under both RCP4.5 and RCP8.5 emission scenarios. ~~As customary, extrapolations are performed under the assumption that the simulated ECDFs are distributed according to a parametric Gumbel probability distribution and the comparison is presented only for the optimal solution obtained during calibration. For each CM we~~

~~considered the following parameterizations referred to the reference period: direct calibration with KS and R_{FDC} metrics, and NSE-ADIGE as representative of a standard calibration procedure.~~

Visual inspection of Figure ~~6~~7 confirms that using the standard calibration (i.e., NSE-ADIGE) of the hydrological model leads to a significant underestimation of all quantiles with respect ~~to~~of using KS and R_{FDC} for all the 3 ~~considered~~ CMs under both RCPs. This is in agreement with the results obtained for the reference period (see Figure ~~4~~5), where simulations using NSE-ADIGE parameterization provided streamflow quantiles systematically lower than with the CMs. In addition, KS-based calibrations always provide larger streamflow quantiles with respect to the cases in which the R_{FDC} metric is considered (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 ~~4~~5).

Figure ~~6~~7 shows that projected changes of high flows extremes depend on the selected CM and emission scenario. Projected streamflow quantiles under RCP8.5 emission scenario are larger than those under RCP4.5 for all the CMs. In general, the projected streamflow quantiles do not exceed the extrapolations from observations in the period 1982-2010 (black lines in Figure ~~6~~7), 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 investigated time window [Ngongondo et al., 2013; Aich et al., 2016; Pechlivanidis et al., 2017; Vetter at al., 2017]. Similar results were also found by Brunner et al. [2019] who implemented a stochastic framework to simulate future streamflow time series in 19 regions of Switzerland, and concluded that future shifts in maximum streamflow will increase and decrease in rainfall-dominated and melt-dominated regions, respectively. ~~In agreement with~~Similarly to our results, Di Sante et al. [2019] ~~also~~ showed that a moderate increase in high flow magnitude (return time of 100 years) is projected for large river basin (drained area >10.000 km^2) in the Central Europe region under RCP8.5 and considering a mid-century time slice.

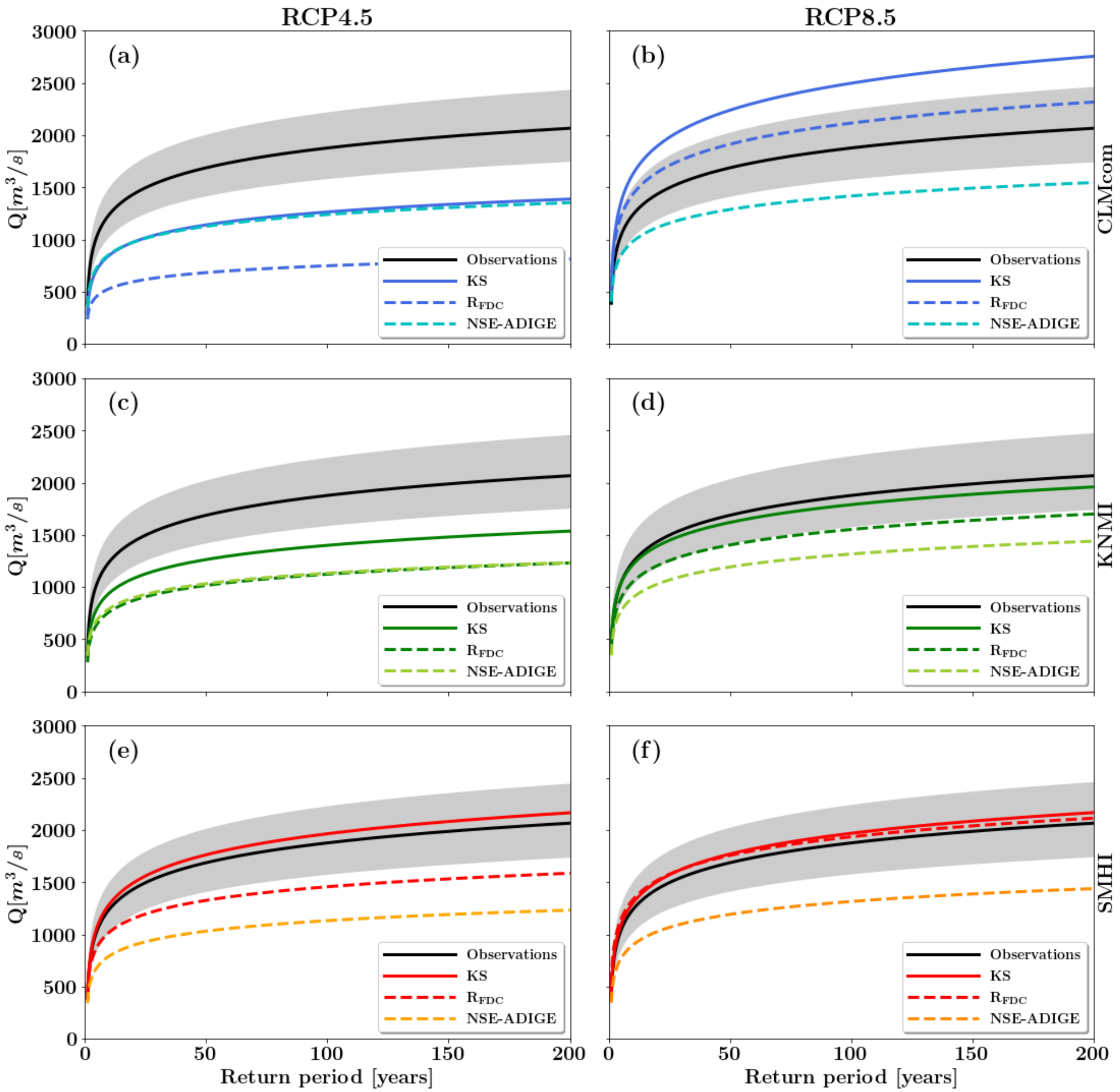


Figure ~~6~~7: Quantiles of annual maximum daily streamflow as a function of return period at Trento gauging station. Projections are based on simulations conducted during the future time period 2042-2070 using as input the CLMcom (first row), KNMI (second row), and SMHI (third row) climate models under the RCP4.5 (left) and RCP8.5 (right) emission scenarios, respectively. Black line denotes the extrapolation of observational data in the period 1982-2010~~.~~ together with the associated 90% confidence interval of the fitted Gumbel distribution (grey shaded area).

5 Conclusions

~~We investigated different strategies for the calibration of hydrological models to be used in assessing projections of hydrological high flows in light of climate change.~~ In ~~particular~~this work, we proposed ~~a~~the methodological framework HyCoX in which the calibration of the hydrological model is carried out by ~~optimizing the reproduction of~~maximizing the probability ~~distribution of high flows extremes. The methodology, coined here as Hydrological Calibration on eXtremes (HyCoX), is such~~ that the ~~hydrological simulations conducted during a reference period, as driven by climate model outputs, are constrained to maximize the probability that the modeled~~modelled and observed ~~extreme streamflows~~high streamflow extremes belong to the same statistical population. The proposed framework is "goal-oriented" and aims at ~~reducing the uncertainty associated to~~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 ~~(precipitations and temperature in the case at hand)~~ provided by ~~the~~ Climate ~~Model (CM).~~Models. In particular, the framework relies on the use of the two-sample Kolmogorov-Smirnov statistic (KS) as objective function during the calibration procedure. This approach ensures statistical coherence between scenarios and observations ~~during~~in the reference period, and, likely, preserves it in the future climate change scenario runs performed with the aim of projecting changes in streamflow extremes. The goal-oriented approach envisaged in this work can be also applied to a variety of hydrological scenario and modelling approaches. While the approach is exemplified here for high flows, it can be applied to low flows as well (e.g. for drought assessment).

~~The proposed procedure is exemplified through application of a few climate models and observational data to the analysis of annual maximum streamflow of the Adige river at the Ponte San Lorenzo gauging station in Trento (Italy). Though the nature of the present work is inherently methodological, it is worth mentioning that the application of the framework over a statistically-sized number of watersheds is currently ongoing in order to demonstrate method's generality. The hydrological model employed is HYPERstreamHS, a continuous simulation distributed model. Three performance metrics were adopted, including the proposed one, for which the Kolmogorov-Smirnov two-sample statistical test was employed (KS for brevity).~~ ~~We~~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 ~~could have been~~can be employed without any loss of generality.

The proposed procedure is exemplified through application of a few climate models and observational data to the analysis of annual maximum streamflow of the Adige river basin (Italy) by means of the distributed hydrological model HYPERstreamHS. The results highlight that adopting KS is preferable to other popular metrics (e.g. ~~Nash-Sutcliffe~~NSE or fit to ~~Flow Duration Curve~~flow duration curve, R_{FDC}) when dealing with high ~~flow~~streamflow extremes. This ~~result~~ validates our hypothesis that targeting directly the statistics of extreme values under consideration during the calibration exercise leads to coherent and ~~consistent~~reliable hydrological models for addressing the impact of climate change. We remark that such approach may lead to a suboptimal performance if the target is different from the one employed in this study, limitation that is outweighed by the improvements in representing high flow extremes in line with the goal-oriented framework pursued in this work. Furthermore, investigation of optimal values highlighted that direct calibration using CMs outputs and KS as objective function lead to unbiased identification of model parameters.

~~In addition, our analysis revealed that using CMs simulations as input of the hydrological model, and adopting the parameterizations derived from calibration against historical time series, is an error prone procedure. Nowadays, it is generally acknowledged that the uncertainties arising at the different steps of the hydrological-climatological modeling chain can cause a significant divergence in the statistics of extremes. However, it appears crucial that the simulated effects on projected extremes in a climate change impact assessment can be safely attributed to the change in climate alone, and not to uncertainties arising from the selection of the efficiency metric in the calibration process. Finally, investigation of optimal values highlighted that direct calibration using CMs outputs lead to unbiased identification of model parameters.~~

~~In climate change impact assessments on streamflow extremes~~ In the present work we also showed that the way the hydrological model is calibrated against observations assumes paramount importance~~. In the present work we showed that~~ 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 ~~is an error prone procedure when the objective is~~can lead to ~~project~~a biased evaluation of the probability distribution of streamflow extremes ~~by using~~when climate models~~. Streamflow~~ are used as input forcing during the reference period, with high streamflow quantiles ~~are~~being dramatically underestimated ~~and fall outside the confidence interval of the quantiles of observed annual maxima when applied to the observation period~~with respect to the fitted extreme value distribution of observations. Extrapolations performed by using the proposed calibration procedure, with input provided by CMs, are instead more consistent and provide a good match with observed quantiles.

~~The goal-oriented approach envisaged in this work can be applied to a variety of hydrological scenario and modeling approaches. While the approach is exemplified here for high flows, it can be applied to low flows as well (e.g. for drought assessment). In any case, it is advisable to calculate the uncertainty band for both the simulation model and the ECDF from observations.~~

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.

Acknowledgments

This research received financial support by 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. B.M. and D.A. also acknowledge 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. The authors acknowledge the climate modelling groups listed in Table ~~1~~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). We also thank the two anonymous Referees whose comments and suggestions helped improve and clarify this manuscript.

References

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

Arnell, N.W. (2011), 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.

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

Avesani D., Zanfei A., Di Marco N., Galletti A., Ravazzolo F., Righetti M. and Majone B. (2022). Short-term hydropower optimization driven by innovative time-adapting econometric model. Applied Energy, in press.

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

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

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

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

Blazkova, S. and K. Beven (2009), 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. (2013), 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.

Brigode, P., Oudin, L., and Perrin, C. (2013), 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.

Brigode, P., E. Paquet, P. Bernardara, J. Gailhard, F. Garavaglia, P. Ribstein, F. Bourgin, C. Perrin and V. Andréassian (2015), 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, https://doi.org/10.1080/02626667.2015.1006632.

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

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

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

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

Chiogna, G., B. Majone, K. Cano Paoli, E. Diamantini, E. Stella, S. Mallucci, V. Lencioni, F. Zandonai, and A. Bellin (2016), 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.

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

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

Diamantini, E., Lutz, S.R., Mallucci, S., Majone, B., Merz, R., and Bellin, A. (2018), 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.

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

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

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

Fenicia, F., Kavetski, D., Reichert, P., and Albert, C. (2018), 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.

Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., et al. (2016), 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 Resources Research, 52, 9228–9234. https://doi.org/10.1002/2015WR017525~~https://doi.org/10.1002/2015WR017525.~~.

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

Gampe, D., G. Nikulin, and R. Ludwig (2016), 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.

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

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

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

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

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

Guthke, A. (2017), 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.

Hargreaves, G.H. and Z.A. Samani (1982), Estimating potential evapotranspiration, J. Irrig. Drain. Eng., 108, 225-230.

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

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. (2018), Environ. Res. Lett. 015006, https://doi.org/10.1088/1748-9326/aa9938.

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

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

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

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

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

Hofstra, N., M. New, and C. McSweeney (2010), 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.

Honti M., A. Scheidegger and C. Stamm (2014), 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.

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

Isotta, F.A., C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S.M. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo, and G. Vertačnik (2014), 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.

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

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

Kennedy, J., and R. Eberhart (1995), 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.

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

Kotlarski, S., K. Keuler, O.B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard, K. Warrach-Sagi, and V. Wulfmeyer, (2014), 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.

Kundzewicz, Z., Mata, L., Arnell, N., Döll, P., Kabat, P., Jiménez, B., Miller, K., Oki, T., Shen, Z., and Shiklomanov, I. (2007), Freshwater resources and their management, in: Climate change 2007: 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.

Laio, F., Allamano, P. and Claps, P. (2010), 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.

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

Landelius, T., P. Dahlgren, S. Gollvik, A. Jansson, and E. Olsson (2016), 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.

Larsen, S., Majone, B., Zulian, P., Stella, E., Bellin, A., Bruno, M. C., and Zolezzi, G. (2021). 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.

Lespinas, F., Ludwig, W., and Heussner, S. (2014), 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.

~~Li, X., Y. Li, C. Chang, B. Tan, Z. Chen, J. Sege, C. Wang, and Y. Rubin (2018), Stochastic, goal-oriented rapid impact modeling of uncertainty and environmental impacts in poorly-sampled sites using ex-situ priors, Advances in Water Resources, 111, 174-191, .~~

K.E. Lindenschmidt (2017), Using stage frequency distributions as objective functions for model calibration and global sensitivity analyses, Environmental Modelling & Software, 92, 169-175. http://dx.doi.org/10.1016/j.envsoft.2017.02.027.

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

Majone, B., A. Bertagnoli, A. Bellin (2010), A non-linear runoff generation model in small Alpine catchments, Journal of Hydrology, 385, 300–312, doi: 10.1016/j.jhydrol.2010.02.033.

Majone, B., C. I. Bovolo, A. Bellin, S. Blenkinsop, and H. J. Fowler (2012), 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.

Majone, B., F. Villa, R. Deidda, and A. Bellin (2016), 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.

Mallucci, S., B. Majone and A. Bellin (2019), 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.

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

Meresa, H.K., and R.J. Romanowicz (2017), 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.

Michel, C., V. Andreassian, and C. Perrin (2005), 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.

Mizukami, N., O. Rakovec, A. J. Newman, M. P. Clark, A. W. Wood, H. V. Gupta, and R. Kumar (2019), 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.

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

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., et al. (2013), ‘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.

Muñoz, E., Arumí, J. L. and Rivera, D. (2013), 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.

Nash, J.E. and J.V. Sutcliffe (1970), 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.

Ngongondo, C., L. Li, L. Gong, C. Xu and B.F. Alemaw (2013), 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~~doi:10.1007/s00477-013-0724-z.

Pearson, K. (1900), 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, issue 302, p. 157 – 175.

Pechlivanidis, I.G., B. Arheimer, C. Donnelly, Y. Hundecha, S. Huang, V. Aich, L. Samaniego, S. Eisner and P. Shi (2017), 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.

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. (2007), 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.

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

Piccolroaz, S., M. Di Lazzaro, A. Zarlenga, B. Majone, A. Bellin, and A. Fiori (2016), 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.

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

Rango, A., and Martinec, J. (1995), Revisiting the degree-day method for snowmelt computations. JAWRA Journal of the American Water Resources Association 31 (4), 657–669, doi.org/10.1111/j.1752-1688.1995.tb03392.x.

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

~~Savoy, H., Kalbacher, T., Dietrich, P., & Rubin, Y. (2017). Geological heterogeneity: Goal-oriented simplification of structure and characterization needs, Advances in Water Resources, 109, 1–13. .~~

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

Seibert, J. and Beven, K. J. (2009), 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.

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

Taye M.T, Ntegeka V., Ogiramoi N.P., Willems P. (2011), 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.

Thirel, G., Andréassian, V., Perrin, C., Audouy, J.-N., Berthet, L., Edwards, P., Folton, N., et al. (2014), 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.

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

~~Tian, Y., Xu Y.P., Booij, M.J., and Cao L. (2016), Impact assessment of multiple uncertainty sources on high flows under climate change, Hydrol. Res., 47.1, 61–74, https://doi.org/10.2166/nh.2015.008.~~

Todd, M. C., Taylor, R. G., Osborn, T. J., Kingston, D. G., Arnell, N. W., and Gosling, S. N. (2011), 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.

Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J. (2010), 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.

Vetter, T., J. Reinhardt, M. Flörke, A. van Griensven, F. Hattermann, S. Huang, H. Koch, I.G. Pechlivanidis, S. Plötner, O. Seidou, B. Su, R.W. Vervoort and V. Krysanova (2017), 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.

Vogel, R. M., and Fennessey, N. M. (1994), 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).

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

Wang, W., X. Chen, P. Shi, and P.H.A.J.M. van Gelder (2008), 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.

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

Weibull, W. (1939), A statistical theory of strength of materials., Ing. Vetensk. Akad. Handl., 151, 1–45.

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

Wilby, R. L. and Harris I. (2006), 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.

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

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

Yang, W., J. Andréasson, L.P. Graham, J. Olsson, J. Rosberg and F. Wetterhall (2010), 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.

Yapo, P.O., Gupta, H.V., Sorooshian, S. (1996), 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.

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