Climate Services 13 (2019) 1–13

Contents lists available at ScienceDirect

Climate Services

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

Original research article

Evaluating the added value of the new Swiss climate scenarios for hydrology: An example from the Thur catchment

Ole Rössler[a,b,*], Sven Kotlarski[c], Andreas M. Fischer[c], Denise Keller[c,d], Mark Liniger[c], Rolf Weingartner[a,b]

[a] *Hydrology Group, Department of Geography, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland*
[b] *Oeschger Centre for Climate Change Research, University of Bern, Switzerland*
[c] *Federal Office of Meteorology and Climatology MeteoSwiss, Zurich Airport, Switzerland*
[d] *Centre for Climate Systems Modeling, ETH Zurich, Switzerland*

ABSTRACT

The availability of new climate greenhouse gas scenario data often prompts the question in what respect the new data provide added value with respect to previous versions and whether or not impact models have to be rerun with the new climatic forcing. This question is the case not only for updated sets of underlying climate model ensembles but also for changes in the applied postprocessing method, such as in the upcoming new climate change projection suite CH2018. The new local projection data are no longer post-processed based on the delta change approach but using quantile mapping. Here, we evaluate the added value of new climate projections from a hydrological perspective. We propose an evaluation scheme that comprises both reference and greenhouse gas scenario periods, average values on different temporal aggregation levels, as well as extreme-related multiday indices. For a test catchment (Thur, pre-alpine, 1700 km^2) we show that the question about an added value, strongly depends on the variable and aspect (average and extreme) of interest. In many cases, basic hydrological characteristics are similarly represented when employing different climate model postprocessing techniques. However, we found differences in the climate change signal already for mean monthly runoff values and even more for several extreme-related indices. Some of them reveal very similar change signals, while the indices related to the intensity/volume of the extremes can strongly diverge. We argue that the comprehensive comparative analysis presented here is transferable and provides useful guidance for the assessment an added value, especially for climate data providers and impact modellers.

Practical implication

End of 2018, a Swiss scientific consortium will launch the new national climate change projections CH2018 to the public (www.climate-scenarios.ch). This new set of projections supersedes the previous projections of the year 2011 (CH2011). The latter was based on a selection of ENSEMBLES model chains for which change factors on multiple-scales were provided for Switzerland. This standardized dataset gained great attention and considerable application in the Swiss scientific, impact modeller and societal user communities. Many different impact assessment studies have been based on this dataset, including several hydrological studies. The new CH2018 climate projections will not only apply new climate model data from the EURO-CORDEX initiative but also downscale and bias correct the outputs employing quantile mapping (QM). Upfront, critical voices raised concern if these new projections will truly provide an added value that justifies the computational and labour effort to rerun the hydrological reference simulations. In the present study, we analyse the effect of changed post-processing methods when switching from CH2011 to CH2018: the use of QM instead of the delta change (DC) approach. This change is likely to influence the results profoundly, as the QM is regarded to overcome some of the known limitations of the DC approach like changes to extreme values and to the temporal variability. We aim to answer the question if an added value is detectable for hydrological impact assessment studies. We specify the question by asking for which variable and aspect of interest (average values and extremes, such as droughts and floods) this added value is found. This preview study was conducted in a prominent, pre-alpine catchment in northern Switzerland. For the sake of a consistent methodological comparison, the underlying climate model data will be that of the CH2011 dataset. In addition to QM and DC, a multi-site weather generator (WG) is investigated as a third climate model postprocessing method with the objective to assess the necessity to include WG-generated realizations in a possible later update of

* Corresponding author.
E-mail address: ole.roessler@giub.unibe.ch (O. Rössler).

https://doi.org/10.1016/j.cliser.2019.01.001
Received 26 June 2018; Received in revised form 4 January 2019; Accepted 21 January 2019
Available online 02 February 2019

CH2018.

As no established comparison scheme is known for this type of analysis, we here suggest a structure that starts with very general characteristics such as the annual water balance, specifies this water balance for mean monthly values, and finally, examines the related extreme indices in more detail. All aggregation levels are analysed for both a reference and a future greenhouse gas scenario period. The former allows for an evaluation of the performance quality of the climate realizations; the latter enables a comparison of the projections when employing different climate model postprocessing approaches. We find that already at the monthly aggregation level, smaller differences in the magnitude of the change signals occur between the three postprocessing methods. These differences are exacerbated for the extremes, and especially for the simulated intensity of both droughts and floods. Moreover, with respect to these extremes, we even find different signs of the change signal. These differences can be related to the formulation of each postprocessing method, with each one having its inherent limitations and advantages. Therefore, the generation of an ensemble median of hydrological simulations based on different postprocessing approaches is not recommended.

The upcoming CH2018 QM-based hydrological simulations will likely provide new and partly contradicting results to the previous CH2011-based experiments, already due to its new downscaling approach. Hence, we clearly found an added value of applying the new climate projections, especially for both the high and low extremes. At the same time, studies that purely focus on monthly to yearly average runoff changes, or on the total number of droughts and floods, do not need to rerun their simulations – at least not in order to account for the change in the postprocessing method. We believe that the presented comprehensive, index-based comparison will be of great help for many climate projections users—impact modellers, societal users, and especially for the climate service centres—as it provides a comprehensive, differentiated outlook on what to expect when updating hydrological climate change projections.

1. Introduction

Numerous climate change impact assessment (CCIA) studies in the field of hydrology have been conducted in the past two decades. Considerable efforts have been spent on issues such as how to best apply climate model data to hydrological models, which impacts are to be expected in different regions under different climate greenhouse gas scenarios, where the associated uncertainties originate, and how to best reduce them. In recent years, especially the climate model postprocessing strategy has been an issue: Which is the optimal downscaling or bias correction method and which are the inherent uncertainties, caveats and advances? An entire scientific community is working to optimally transfer a coarse-resolved and potentially biased climate model output to a point scale. Intercomparison projects have been initiated such as COST VALUE (Maraun et al., 2015) and BCIP (Nikulin et al., 2015), targeting meteorological variables, as well as studies focusing on hydrological features (Chen et al., 2013; Teutschbein and Seibert, 2012; Hundecha et al., 2016). The former, “meteorology-focused” approaches involve the validation of different aspects of downscaled time series, such as their temporal structure, spatial structure, variability, and intervariable consistency, to name a few (Maraun et al., 2015). “Hydrology-focused” comparison studies focus on the impact itself and relevant discharge features, such as low flow or flood indices.

These intercomparisons primarily found that the optimal choice of a climate model postprocessing technique[1] needs to be based on the target variable and the specific application considered (Chen et al., 2013), as different methods do not perform equally well for different variables, statistics (means and extremes) and regions (Quintana Seguí et al., 2010; Hundecha et al., 2016; Rössler et al., 2012; Diaz-Nieto and Wilby, 2005; Dobler et al., 2012). More recently, in hydrology, a strong focus was laid on bias correction methods and comparisons thereof (Graham et al., 2007; Hundecha et al., 2016; Lenderink et al., 2007; Fang et al., 2015; Teng et al., 2015; Bosshard et al., 2013), revealing again that strong differences stem from the choice of method (Teutschbein and Seibert, 2012).

Despite these scientific insights, the choice of a certain climate dataset to apply, a certain downscaling approach to select, and the application of a certain hydrological model is often controlled by practical issues, such as the availability of the data, the applicability of the downscaling approach or bias correction, and the hydrological model, rather than by scientific progress (Rössler et al., 2017). Hence, the generation and distribution of reference climate scenario products like, e.g. CH2018 or UKCP18, that make the best use of recent scientific advances are decisive to foster state-of-the-art climate change impact assessment studies.

In this respect, the choice for a certain reference projection product has far-reaching consequences. On the one hand, it allows many scientists and practitioners to apply the (latest) climate projections for their purposes in a straightforward manner. On the other hand, the provision of reference projections set standards and promotes impact studies with certain advantages, limitations, and possible structural uncertainties. The rather simple scientific question of which climate dataset and which downscaling approach to use is, therefore, loaded by an ethical dimension, as these projections have a great impact on the succedent impact assessment studies. The latter are challenged by the high demand from society for accurate information about future changes, and the ethical aspiration by the provider to provide credible projections (Rössler et al., 2017).

For the case of Switzerland the first national climate projections were released in 2011 by a comprehensive scientific consortium (CH2011, 2011). They made use of 20 GCM-RCM model chains of the ENSEMBLES project (van der Linden and Mitchell, 2009) and provided delta change signals for both larger regions and specific measurement sites in Switzerland for three future time periods and the greenhouse gas scenarios SRES A1B. At the site scale, spectrally smoothed daily change factors were provided following the work of Bosshard et al. (2011). These climate projections are regarded as very successful as they provide a common basis for CCIA studies from different disciplines and are comparatively easy to implement. After seven years, by the end of 2018, a new generation of climate projections for Switzerland is to be released by the same consortium. This time, the climate projections will be based on simulations of the EURO-CORDEX GCM-RCM ensemble (Jacob et al., 2014; Kotlarski et al., 2014), again downscaled and bias-corrected to provide local-scale information, comprising projections for the three RCP emission scenarios 2.6, 4.5, 8.5 (CH2018, 2018).

For this purpose, the quantile mapping (QM) method is applied. Despite certain drawbacks of this approach (e.g. Maraun et al., 2017), it comes along with a number of advantages compared to the delta change approach, and its choice was fostered by the stakeholders’ claim to improve the consideration of extreme values and changes in the day-to-day variability of the climate models (MeteoSwiss, 2016). These features were found in the intercomparison studies mentioned above to be especially improved by distribution-based correction methods, such as QM. In a second stage, the consortium might also provide results

[1] Note that we here define «postprocessing» in a general sense as any processing of raw climate model data with the aim to provide directly applicable climate scenario products. In particular, this can involve bias correction, statistical downscaling or the computation of delta change values.

obtained from the application of a recently developed multisite weather generator (Keller et al., 2015, 2016) to fulfil the stakeholder claims for numerous realizations of climate greenhouse gas scenarios.

As the choice of QM as postprocessing method involves a change from the previous delta change approach, it is of special interest for the climate projection providers, as well as for impact modellers, to know about the advantages and disadvantages of the new climate projection generation and its underlying methodologies. Moreover, a climate projection provision is not limited to the mere data handling, but rather includes guidance for impact modellers or societal users towards the optimal use of the climate projection products and potential limitations and pitfalls (see, for example, the recently published EURO-CORDEX user guidelines available from http://www.euro-cordex.net/imperia/md/content/csc/cordex/euro-cordex-guidelines-version1.0-2017.08.pdf).

Although numerous studies have been published that compare different climate projections and realizations of climate projections for hydrological impact studies, a transfer of these results to the Swiss situation beyond some very general remarks is hardly possible. That is, all comparison studies emphasized the very specific response signal of each downscaling method. According to Hundecha et al. (2016), projection differences could amount to up to 60% for hydrological extremes. Consequently, understanding the advantages and disadvantages of new climate projections generations, based on revised climate model postprocessing strategies, requires a consistent one-to-one comparison of the resulting projection products and their respective influence in subsequent applications (e.g., in hydrological modelling). This also holds true for updating the underlying climate model data, but for the sake of simplicity, we will here focus on the postprocessing methodology applied.

The overall objective of the present work is to assess the hydrological consequences of a revised climate model postprocessing procedure that is envisaged for the next generation of Swiss climate projections. We compare the hydro-climatic differences of the spectrally smoothed delta change approach applied in CH2011 with the QM approach of the forthcoming CH2018 projections. For the sake of completeness, we also include products based on a recently developed multisite weather generator (Keller et al., 2015, 2016) that will likely complement the QM-based CH2018 realizations for certain applications. We specifically ask the question to what extent the new climate projections give reason to expect new insights in hydrological projections.

The answer is not clear from scratch. Scanning the published intercomparison studies, one finds classical features, such as water balances and hydrograph comparison, both for the reference and the greenhouse gas scenario time period, as well as indices describing hydrological extremes. However, no common systematic analysis was used, further challenging the comparability. Addor and Seibert (2014) hence proposed “a more systematic quantification of the consequences of bias correction on impact simulations”. Furthermore, they emphasized the need to also consider multiday indices in the analytics of the consequences of bias–correction methods.

We here aim to address these claims and suggest a set of analyses that comprehensively evaluate the impacts of different downscaling or bias–correction approaches on hydrological projections.

For consistency, we use identical underlying climate model data for all three approaches: the ENSEMBLES dataset. For the sake of simplicity and to keep the computational time within reason, we solely focus on one mesoscale pre-alpine catchment in Switzerland, the Thur catchment. However, the general approach presented, as well as the applied set of hydro-climatic indicators to evaluate the effect of updated climate projections, are transferable to further Swiss catchments as the three climate post-processing products (DC, WG, and QM) are, in principle, available for the entire Swiss domain.

Table 1
List of the ENSEMBLES subset of GCM-RCM combinations used in this study.

GCM	RCM	Institution
BCM	RCA	SMHI
HadCM3Q0	CLM	ETH Zurich
HadCM3Q0	HadRM3Q0	MetOffice
HadCM3Q3	HadRMQ3	MetOffice
ECHAM5	REMO	MPI
ECHAM5	HIRHAM	DMI
ECHAM5	RACMO	KNMI
ECHAM5	RCA	SMHI
ECHAM5	REGCM3	ICTP
ARPEGE	ALADIN	CNRM

2. Methods and data

2.1. Climate model data and observations

As a common basis for all three postprocessing approaches, we chose climate model data comprising 10 GCM-RCM chains from the ENSEMBLES project (van der Linden and Mitchell 2009), assuming the SRES A1B emission scenario. Table 1 lists the climate model projections from which the reference period, 1980–2009, and the scenario period, 2070–2099, have been extracted. Fig. 1 provides an impression of the spatial resolution of the RCM grid cells (25 km) in comparison to the station locations and the catchment size.

As a reference for the hydrological modelling and the different downscaling methods, we used the observed daily mean runoff at the catchment outlet, Andelfingen, as well as temperature and precipitation records in the reference period, 1980–2009, at nine temperature and five precipitation stations that are well distributed across the catchment and that thoroughly cover the altitudinal spread (Fig. 1).

2.2. Hydrological modelling

The physically-based distributed hydrological model WASIM-ETH in the 8.5 version (Schulla 2015), was applied. As only downscaled temperature and precipitation data were available we limited the complexity of the model to algorithms that only rely on temperature and precipitation. For example, to determine evapotranspiration, we used the Hamon equation instead of the Penman-Monteith approach, and we calculated snowmelt by applying a degree-day-factor approach instead of an energy-balance model.

As the hydrological model requires spatial information about temperature and precipitation, we first conducted a spatial interpolation for each time step based on all nine temperature and five precipitation station data series. From the available interpolation techniques, we chose an inverse distance weighting (IDW) approach for precipitation and a LOESS regression against elevation for temperature. Although this choice was subjective, we tried to find the optimal combination of the most plausible spatial pattern and the least modifying technique. In doing so, we intended to retain as much climate model information and as many of the downscaling effects as possible in the hydrological model. However, the investigated indices are not based on station data, but rather on their derived spatial pattern.

We calibrated the hydrological model against the daily runoff for the time period 1990–1997, and validated it for the reference period 1982–2009. The model performance amounts to a Nash-Sutcliff-Efficiency (NSE) = 0.86, Kling-Gupta-Efficiency (KGE) = 0.91, RSME standard deviation ratio (RSR) = 0.37, and percent bias (PBIAS) = −5.9% (following Moriasi et al. 2007). In addition, a graphical comparison based on the hydrograph (a) and the Q-Q-plot (b) visualizes the overall good performance of the hydrological model. Some limitations are noticeable in spring and late summer (for long-term means, Fig. 2, left), and a slight underestimation of high runoff

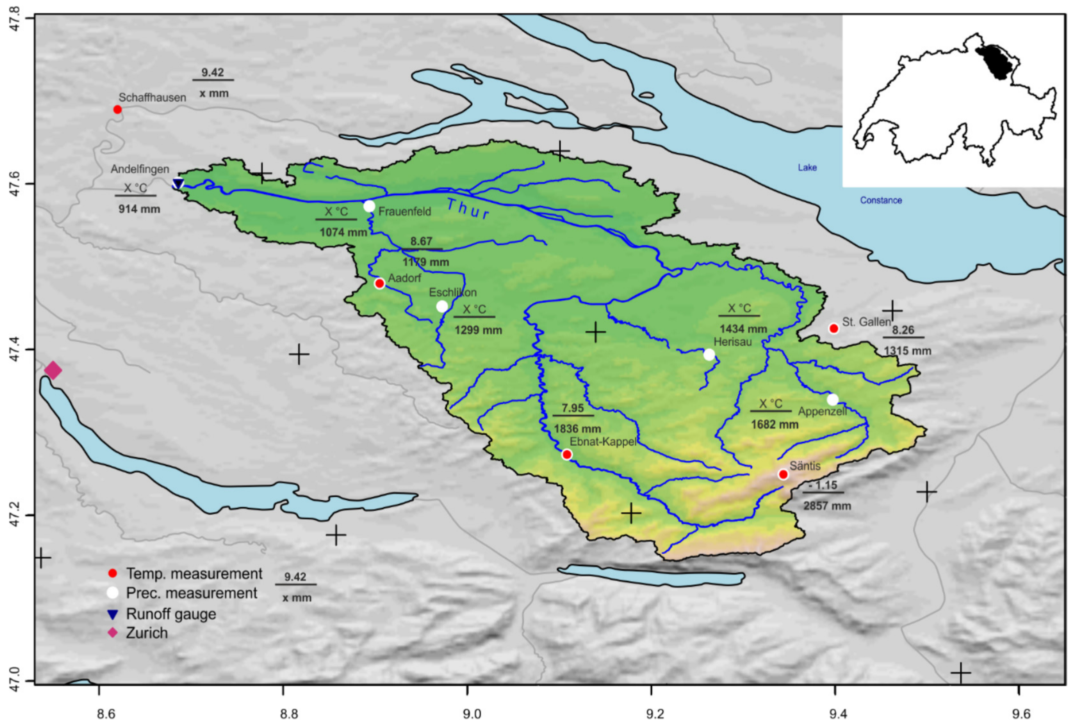

Fig. 1. Location of the Thur test catchment in Switzerland and the spatial distribution of meteorological stations recording temperature (red) and precipitation (white). The grid network (black crosses) refers to the corner points of the underlying RCM grid (25 km × 25 km). Values above the fraction line indicate the mean annual temperature [°C] at that site for the reference period, and values below the line indicate their annual precipitation [mm]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

values (over 600 m^3/s, Fig. 2, right).

2.3. Postprocessing approaches

1. Delta change (DC)
   The starting point for this comparison study was the delta change approach employed in the CH2011 framework according to the procedure of Bosshard et al. (2011). This method first involved estimating the mean annual cycles of the reference and scenario periods over 30 years in the raw climate model output. To ensure a smooth mean annual cycle with daily granularity, spectral smoothing (based on a superposition of harmonics) was applied. To spatially interpolate the coarse model data to the station locations, the four nearest grid points were used, applying an inverse distance weighting scheme. The resulting product were mean annual cycles of change factors (additive for temperature and multiplicative for precipitation) that could be applied onto observed series in order to generate future climatological forcing series for the hydrological model.
2. Multisite, multi-variate weather generator (WG)
   An advantage of weather generators is their flexibility and their ability to generate multiple realizations of weather based on the observed record. In a climate change context, a WG calibrated for today's climate is perturbed by changes in the WG parameters as derived from climate models. WGs have been widely used in CCIA, especially in the fields of agriculture and hydrology (Klein et al., 2013; Khalili et al., 2011). One common limitation of WGs is their lack of consideration of the spatial dependence of simulated time series. For hydrology, this particularly concerns precipitation and its areal sums over a hydrological catchment. The WG developed by

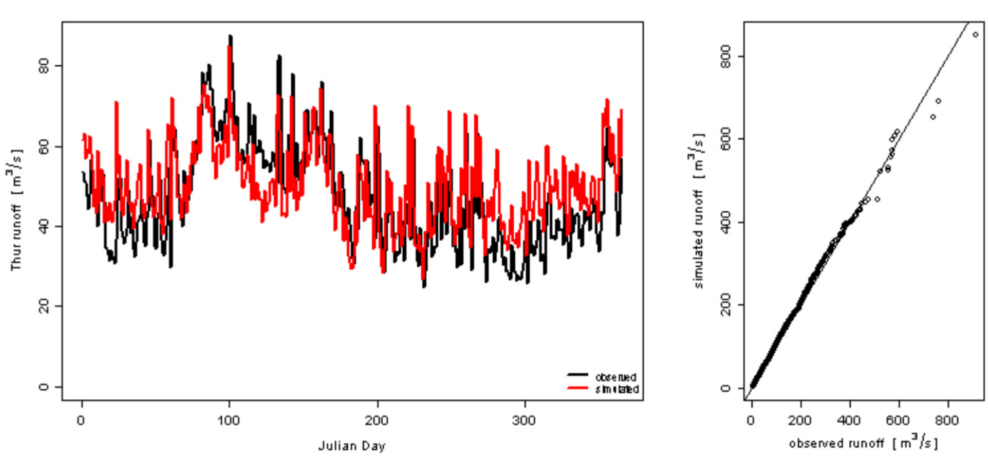

Fig. 2. Performance of the hydrological model in reproducing the daily observed runoff. Left panel: mean annual cycle over the period 1982–2009.Right panel observed versus simulated daily runoff quantiles for the same period.

Keller et al. (2015) overcomes this limitation by explicitly enforcing the spatial structure using spatially correlated random number streams. This multisite weather generator has been tested over the complex topography in Switzerland for current and future climate conditions (Keller et al., 2015, 2016).
Keller's approach is, in essence, a Richardson-type WG (Richardson 1981) that builds on a first-order two-state Markov chain model. Following Wilks (1998), the WG was extended to generate synthetic weather with the correct spatial dependence among several station locations. This was ensured by spatially correlated, but serially independent, random number streams. Conditioned on whether a dry or wet day is generated, the WG simulates the daily minimum and maximum temperature in a multisite mode. A threshold of 1 mm/day was assumed to separate wet from dry days. The daily mean temperature (used as input for WaSiM-ETH) was calculated by taking the mean of the generated minimum and maximum temperatures. For this study, we generated and processed 50 synthetic time series with daily granularity for the reference period, mimicking the observations. In the case of the projection period, 50 realizations were generated for each of the ten GCM-RCM simulations, resulting in 500 different realizations of future weather. As the WG approach inheres a stochasticity component reflecting meteorological variability, it is not directly comparable to the two other approaches. Yet we argue that a qualitative comparison keeping this additional information in mind is justified.
3. Quantile mapping (QM)
Quantile mapping is classified as a model output statistic (MOS, Maraun et al., 2010) that seeks to correct the distribution functions of climate model outputs (region or grid cell) against the distribution function of related observations. Transfer functions are derived for each defined quantile for a calibration period and are then applied to the projected time series under the assumption that the bias between the observations and climate model output for a given quantile remains constant. One must differentiate between a pure bias correction mode of QM, where the climate model output and observations have the same spatial scale, and a downscaling mode, where the observations are on a finer scale than the model and QM implicitly includes a downscaling component. Several studies showed that QM has equal or better performance than perfect prognosis downscaling approaches (Boé et al., 2007; Themeßl et al., 2011; Gutiérrez et al,. 2018). Concerns about intervariable consistency issues have been dismissed for single cases (Wilcke et al., 2013), and another study (Rajczak et al., 2016) showed that biases in the temporal structure could be removed by QM. D its popular usage, QM has limitations particularly in its downscaling mode. These include the following:
(a) its deterministic character and its inability to properly represent climatic variability on fine temporal and spatial scales (Maraun 2013),
(b) its altered climatic trends due to inflated variance (Maraun 2013), and
(c) the typical unsolved question of the extent to which a grid-cell based climate model output is informative for climatic conditions at single sites in complex terrain (Maraun and Widmann 2015).
(d) (its potential to distort the change signal due to the time dependency of the correction function Grillakis et al. (2017). Although, Grillakis et al. (2017) presented a fixture of this issue, it is not included in the CH2018 projections.

According to Gudmundsson et al. (2012) , three different types of quantile–quantile transformations exist: transformations based on theoretical distributions (e.g., Piani et al. 2010a), parametric transformations (Piani et al. 2010b), and non-parametric transformations that use the full empirical distribution (Déqué et al. 2007; Themeßl et al. 2012). Here, we use a non-parametric, linear transformation. This method has recently been proven to show the most robust results for Switzerland, especially in terms of extreme values (Ivanov and Kotlarski 2016). This QM implementation is based on Rajczak et al. (2016) and has been extended by a frequency adaptation for precipitation that ensures correction of biased wet-day frequencies. Quantile-quantile relationships were estimated for each day of the year in the reference period, with a 91-day moving window to account for seasonality in the correction function. For the treatment of extremes outside of the observed distribution, we used an additive correction according to the transformation function for the 1st and 99th percentile in the reference period.

2.4. Hydro-meteorological characteristics as evaluation criteria

Here, we suggest and apply a systematic scheme to analyse the impacts on the hydro-climate due to different postprocessing approaches.

As a primary principle, all analytics are performed for both the reference and the projection period. The former provides insights into the capabilities and limitations of the postprocessing approaches in reproducing the current hydro-meteorological conditions. In turn, the evaluation of the projection time period focuses on the comparison of climate change signals.

For each time period, the comparison is performed on different temporal aggregation levels, starting with the annual water balance, followed by monthly values to analyse the seasonality, and proceeds to an index-based analysis of specific hydro-climatologic characteristics and events. We thereby focus on the two major hydro-climatic variables: precipitation and runoff in terms of the behaviour of high and low flows as well as related precipitation characteristics. Table 2 lists the applied indices and their definitions.

Low flows are defined as days on which the runoff has fallen below the 5th percentile of 10 years (in Switzerland termed Q347) in the reference run for at least five consecutive days. The Q347 is the multiyear (at least 10 years) 5% quantile of natural runoff. Here, this threshold is

Table 2
List of meteorological indices used to evaluate the performance of the postprocessing methods.

Index	Description
Number of drought events	Number of events undershooting the median Q347 of the reference period model run for at least 5 consecutive days [-]
Low flow index Q347	Median of multiple ten years 5% quantile of runoff [m^3/s]
Drought intensity	Deficit volume under the Q347 threshold per drought [m^3/drought]
Drought duration	Length of a drought [days]
flood days (≥Q999)	Number of days consecutively exceeding the Q999 of the reference period [days]
Mean AMF	Mean annual maximum flood peak [m^3/s]
Flood volume	Mean volume of floods exceeding the mean AMF [m^3/event]
Flood duration	Mean duration of floods exceeding the mean AMF [days]
3daysRainSum	Median annual max of sums of rain in three consecutive days [mm]
5 days Rain Sum	Median annual max of sums of rain in five consecutive days [mm]
Average event precipitation	Mean precipitation sum during flood events [mm/event]
Snowmelt fraction	Fraction of snowmelt contributing to the discharge [%]

calculated multiple times applying a moving window of 10 years and recording the median. Based on this definition, drought-related indices, such as the number of droughts over the evaluation period, mean duration and mean intensity are calculated. We applied a 10-day moving average before defining the drought events to obtain "mutually dependent deficits and remove minor events" (WMO 2009). In contrast, we defined high flows as the 0.999 quantile (Q999) of the respective time period, representing an approximately 3-year return period high flow. As with droughts, we not only counted the high flows but also calculated flood-related indices, such as the mean flood peak, the mean flood volume and duration. Here, we empirically defined a minimum gap of 5 days between events to ensure independence. For both high and low extremes, we defined the threshold for each post-processing approach separately in the reference time period. These thresholds are accordingly applied in the simulation time period of each postprocessing method.

To trace possible causes of changes in flood occurrences back to the meteorological input, flood-related meteorological parameters, such as 3-day precipitation sums (3daysRainSum) and 5-day precipitation sums (5daysRainSum) are analysed. With respect to changes in droughts, we analysed the representations and projections of the dry spell lengths, where we defined a dry day as a day with less than 0.1 mm areal precipitation. Finally, to account for changes in the cryosphere, we took a closer look at the snowmelt contribution to runoff in this pre-alpine catchment by analysing the number of generated snowfall days (> 1 mm snowfall) among the downscaling methods.

Specifying the settings for the present application, the reference run is defined as the hydrological simulation, which is driven by observations for the period 1982–2009, and the projection period refers to 2072–2099. The first two years of each model run (i.e., 1980–1981 and 2070–2071) are omitted to account for model spin up. All analyses are based on the spatially interpolated meteorological parameters that are – only for the analyses - averaged over the entire catchment and the hydrological model output over 28 years.

To evaluate the performance of the downscaling methods, we compared the WG- and QM-driven hydrological model results with the reference runs for 1980–2009 (the DC approach in the reference period is the same as the reference run by definition). The period 1980–2009 also corresponds to the calibration period of WG and QM and implies that certain aspects of the downscaled climate are realistically represented by definition (such as the transition frequencies in the WG or the distributional quantities in QM at the scale of individual stations). The number of realizations from the different downscaling approaches varies between methods and between periods and is a product of the number of climate model runs multiplied by the number of stochastic realizations. The final number of realizations is given in Table 3.

3. Results

3.1. Performance of the downscaling methods

In general, both WG and QM are able to reproduce the average conditions in the reference period (compare the reference to the light green and light blue boxes in Fig. 3). This result is especially true for temperature, while for precipitation larger deviations occur. For the QM approach, the deviations occur in certain months, especially in March, June, August, September, and December. Thereby, the deviations tend to the direction of the monthly levels of the precedent or following month: An underestimation occurs if preceding or following month (or both) comprise lower precipitation sums. This is an effect of the underlying 91 days used for calibration stretching across at least three months. The underestimations of QM precipitation are inherited by the runoff estimates and are exacerbated in some months. For the WG-generated precipitation, deviations from the reference data on a monthly scale are smaller, yet still present. WG also shows highest variability, both in the reference and the projection period as a result of this stochastic component and thereby hinting at the amount of climate variability.

Table 3
The number of model realizations per downscaling method and projection period.

	Reference Run	Quantile Mapping	Weather Generator	Delta Change
1980–2009	1	10	50	n.a.
2070–2099	n.a.	10	500	10

Thereby, WG underestimates the runoff in some months in particular (i.e., June, July, September) with unclear rationale; deficient temperature and precipitation values on individual days might be responsible (not shown). However, the underestimations are again smaller than for QM based simulations.

In hydrology, a model's performance should be assessed not only by evaluating the target variable against which the model was calibrated but also by considering its ability to correctly reproduce the water balance. The analysis of the water balance representation in the WG- and QM-driven hydrological model outputs in the reference period revealed very similar values for the mean annual water balance components of precipitation, evapotranspiration, and runoff (Table 4); deviations were ± ~5% at the most. This similarity of results also extends to the interannual variability, as indicated by the maximum and minimum annual values. The main difference between the methods relates to the separation of total precipitation into rain- and snowfall. WG and QM both tend to underestimate the snowfall amounts. This is likely related to some deficiencies in correctly reflecting the intervariable dependence between temperature and precipitation.

How do the differently post-processed meteorological input data influence extreme hydrological quantities? Fig. 4 shows their performance for several hydro-climatic indices under the current climate conditions (reference and first two boxplots per panel a – l). The notches of each boxplot help to interpret whether the differences found are significant (if the notches overlap, they are not). Clearly, the distinction between the methods becomes more pronounced than for the mean values. For those investigated hydro-climatic indices that are closely related to precipitation intensity and/or timing (3daySum, 5daySum (Fig. 4i, j), and drought duration(d)), as well as the number of floods (f) and droughts (b), the QM approach is able to reproduce the reference run, and widely outperforms WG. For other indices that are related to flood magnitudes and duration, as well as for drought intensity, WG outcompetes QM. The remaining indices calculated like low flow index Q347, flood causing precipitation, as well as snowmelt fraction are not well represented by both post-processing methods. In terms of the letter, QM performs slightly better than. The tendency to overestimate this fraction is in line with the higher snowfall fractions (Table 4).

The deficit in reproducing multiday indices is remarkable for both postprocessing methods. This limitation is further exacerbated in Fig. 5, showing the representation of dry spell length. A strong overestimation of the frequency, especially of short, dry spell lengths, is apparent for the WG approach (c). At the same time, the frequencies of longer dry spells (10 days and more) are underestimated (c). At first, the strong overestimation is surprising, as the WG is calibrated to correctly reproduce dry-wet sequences of adjacent days. However, as the frequency of longer dry spells is underestimated, the missing long spells are compensated by a higher number of shorter dry spells. QM, on the other hand, shows smaller deviations from the reference (e), with the strongest deviations at the one-day spell length and a tendency to underestimate the frequency of longer spell lengths (5-, 7-, and 8-day spell lengths). The reasons for this might be similar to those outlined for WG, although with a smaller magnitude. These results need to be interpreted in light of the spell-length distribution found in the reference run (a). Thus, deviations are clearly present, but in the case of QM, they are of

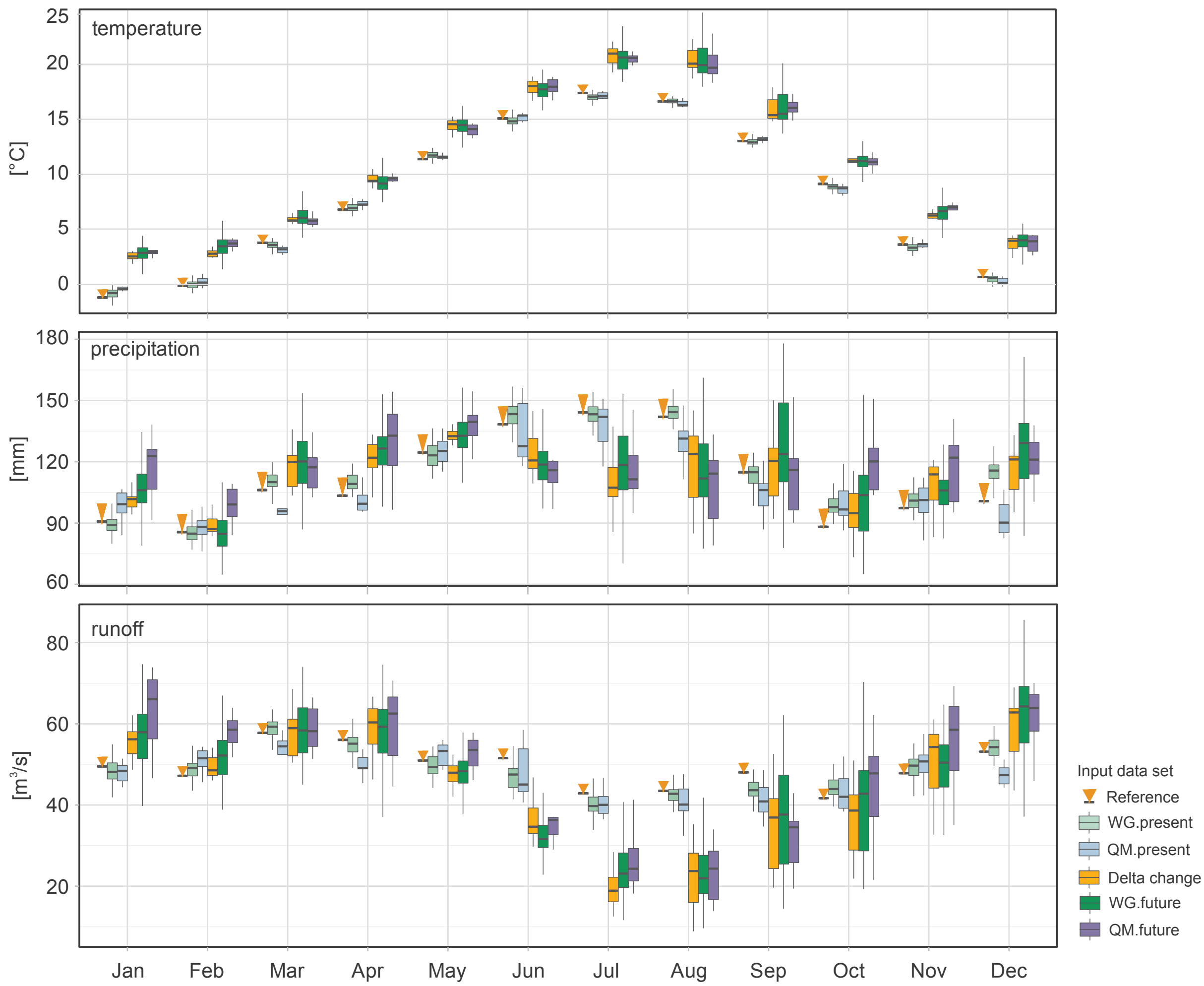

Fig. 3. Comparison of monthly mean values from the WG, QM, and DC methods for both the reference (first three boxplots per month) and projection periods (last three boxplots per month). Temperature and precipitation are depicted as catchment mean values after spatial interpolation; the runoff represents the Thur discharge at Andelfingen. Boxplots refer to the ensemble of realizations and depict the median (black horizontal line) and the interquartile range from 25 to 75% (box) as well as the total range of values (vertical line). Please note, WG of reference and projections period inheres a stochasticity component.

Table 4
Mean annual water balance components averaged over the Thur catchment for the reference run and for the three postprocessing approaches. Shown below are the arithmetic mean values across all realizations of the mean, maximum and minimum annual sums for the reference period (1980–2009) and the projection period (2070–2099). Maximum and minimum values indicate variability within the 28-year periods. Reference and delta change-derived values are presented in bold letters to enhance readability. The CS (climate change signal) columns indicate the relative climate change signal in percent.

		n = 1 Reference	n = 50 WG Present	n = 10 QM Present	n = 10 DC	CS [%]	n = 500 WG future	CS [%]	n = 10 QM future	CS [%]
	max	1728	1582	1714	1729	0	1582	0	1866	9
Precipitation [mm]	**mean**	**1383**	**1379**	**1350**	**1352**	**−2**	**1381**	**0**	**1408**	**4**
	min	1080	1187	986	1064	−1	1186	0	1055	7
	max	1505	1365	1513	1605	7	1462	7	1787	18
Rainfall [mm]	mean	1215	1155	1170	1280	5	1256	9	1318	13
	min	959	959	840	1013	6	1057	10	961	14
	max	299	353	317	158	−47	226	−36	181	−43
Snowfall [mm]	mean	168	223	180	71	−58	125	−44	89	−51
	min	57	121	78	19	−67	53	−56	26	−67
	max	490	485	475	544	11	562	16	535	13
Evapotranspiration [mm]	**mean**	**449**	**451**	**446**	**515**	**15**	**520**	**15**	**499**	**12**
	min	414	417	414	474	14	481	15	441	7
	max	1223	1093	1068	956	−22	1101	1	1128	6
Runoff [mm]	**mean**	**914**	**887**	**861**	**837**	**−8**	**823**	**−7**	**878**	**2**
	min	622	734	712	654	5	589	−20	647	−9

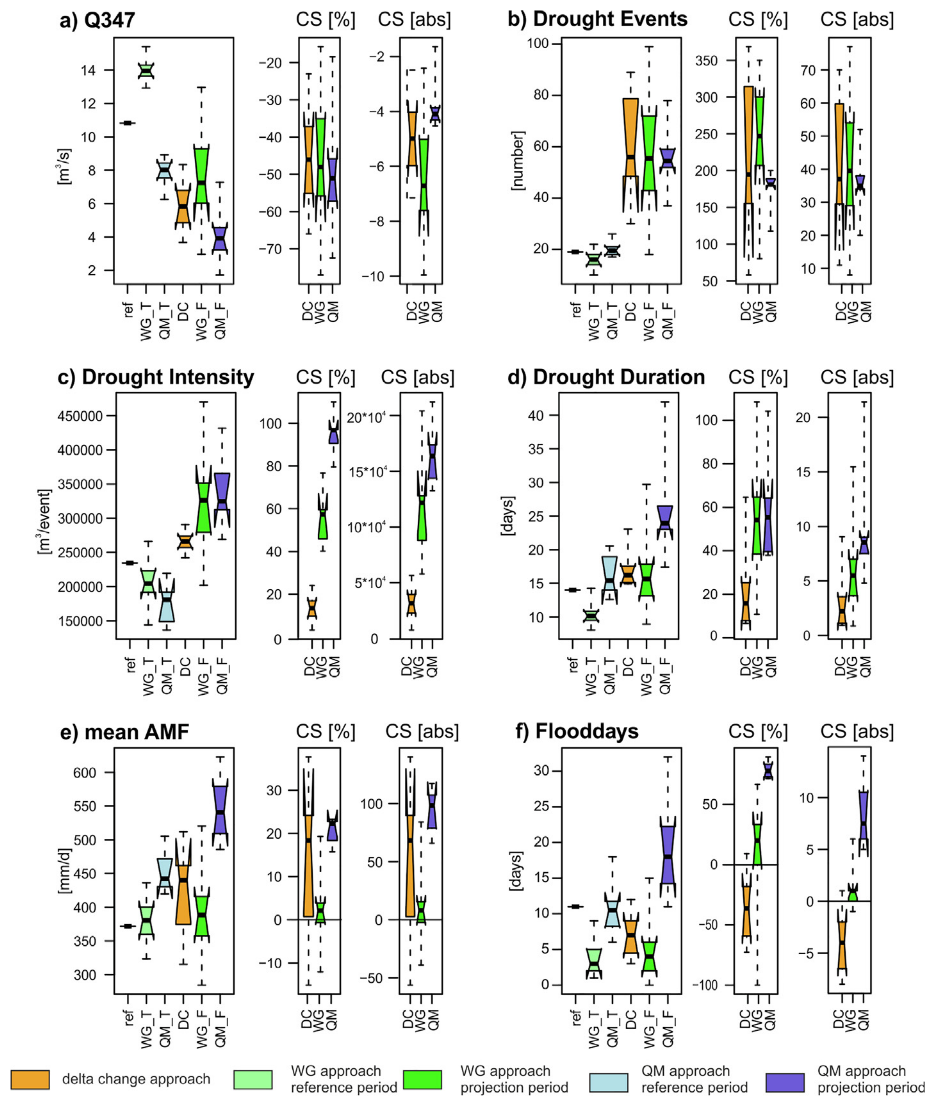

Fig. 4a. Summary figure – part one - of postprocessing effect on extreme-related hydro-climate indices in both the reference (first three boxplots per panel) and projection (last three boxplots per panel) periods. In addition, we show the change signals (CS) for each index expressed as absolute [abs] and relative change [%]. All meteorological indices (i–l) are based on catchment mean values. Boxplot margins display the median, IQR (25%–75%), as well as the range (min and max value). Please note, WG of reference and projections period inheres a stochasticity component.

rather small magnitude.

3.2. Comparison of hydro-meteorological projections from different downscaling methods

Finally, we compare the climate change signals from each postprocessing method, evaluate their plausibility, and thereby analyse the effect of updating climate projections in hydrological impact assessment studies. The effect of different postprocessing methods on the projected water balance (Table 4) summarizes the effects on an annual basis. In general, the mean, maximum and minimum annual values are very similar between the methods, but small differences still occur. That is, the precipitation projections using the QM approach show a slight increase, especially in the maximum and minimum annual values, while WG and DC show unchanged conditions. This difference in precipitation is further differentiated as the shift from snowfall to rainfall amounts found for DC and QM is less pronounced for the WG model runs. Since evapotranspiration has an increase of ~60 mm for all postprocessing methods, there is slightly lower annual runoff in the WG and DC approaches, whereas QM compensates for this loss with higher precipitation amounts.

In terms of mean monthly runoff (Fig. 3, and change signals in the Appendix Fig. S1), the typical response pattern of a central European pre-alpine catchment is projected by all methods: winter runoff increases, while summer runoff decreases. Both are a consequence of a rising snow line and increasing winter precipitation. At the same time, evapotranspiration increases, while precipitation decreases in summer. Despite very similar seasonal patterns, the magnitudes and uncertainties of the projected changes differ considerably among the methods. WG and especially QM show a slightly stronger increase in winter runoff (Jan-Apr, Fig. 3, and change signals in S1) than DC, which is mostly a result of higher precipitation amounts projected in these approaches at nearly equally increasing temperature levels projected by all approaches. In contrast, the summer runoff decrease (Jun - Sep, Fig. 3, and change signals in S1) is often stronger in the DC data. While QM shows the least drastic change signal in summer and the strongest change signal in winter, WG-based projections lie in between the two competitors. This pattern mainly mirrors the patterns found in the precipitation projections, as the differences in temperature projections are rather small.

The projections show a further differentiation regarding hydrological extremes (Fig. 4). While the change signal (smaller boxes beside

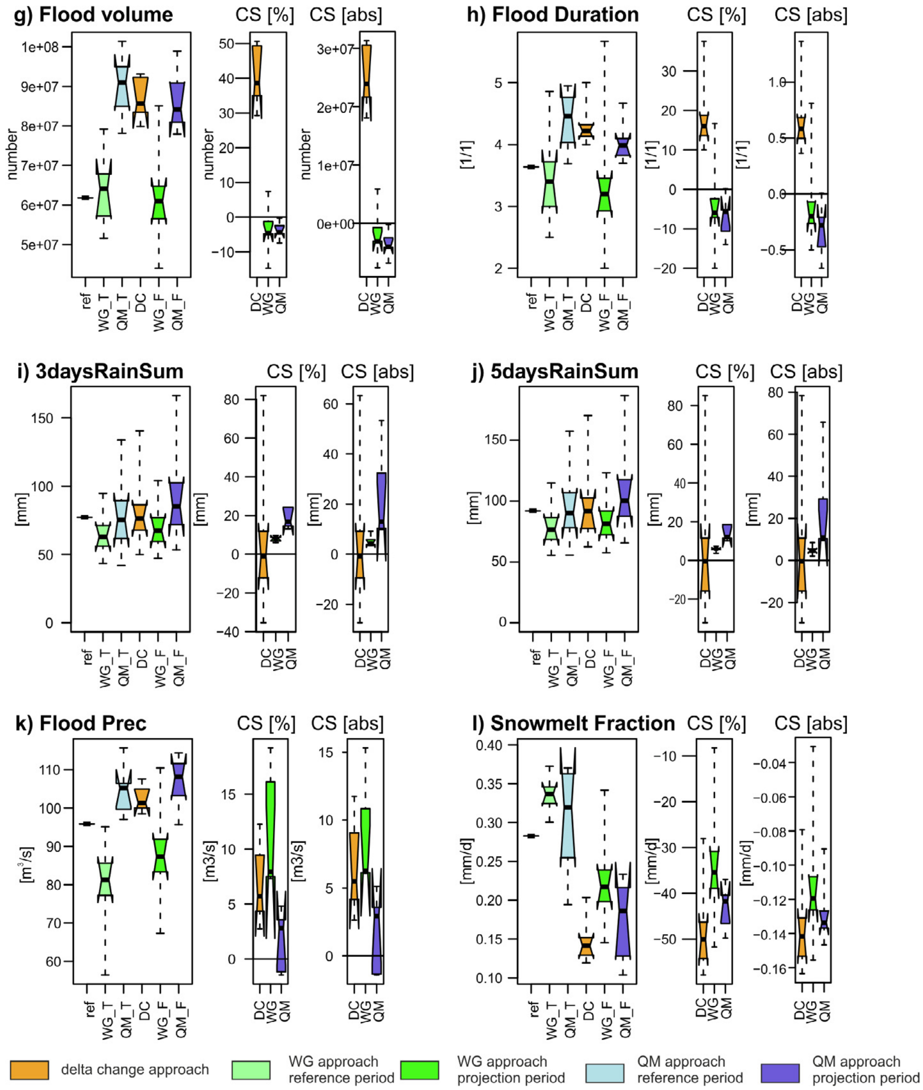

Fig. 4b. Summary figure – part 2 – of postprocessing effect on extreme-related hydro-climate indices in both the reference (first three boxplots per panel) and projection (last three boxplots per panel) periods. In addition, we show the change signals (CS) for each index expressed as absolute [abs] and relative change [%]. All meteorological indices (i - l) are based on catchment mean values. Boxplot margins display the median, IQR (25% −75%), as well as the range (min and max value). Please note, WG of reference and projections period inheres a stochasticity component.

each index) for the low flow index Q347 and the number of drought events are almost the same for the three postprocessing methods, drought intensity and drought duration are accentuated in the WG and QM-based realization, with drought intensity showing an even higher intensification for QM. While the direction of change is consistent between the methods for droughts, this is not the case for floods and related hydro-meteorological indices. We found that DC attenuates the number of flood events with widely unchanged mean AMF values. WG projects almost unchanged values with slightly more events, but the events are shorter and consequently of lower volume. QM showed an amplification (Fig. 4e, f) of mean AMF and number of events but is similar to WG of shorter duration and less volume. The mean event precipitation mirrors the flood volume projections, also indicating the reason for the partly strong deviation of the projection from the reference. The attenuation of the magnitude and number of floods in DC is the result of the linear scaling of each rainfall event in the reference period as a consequence of the decreasing mean monthly precipitation amounts in the RCMs in summer. Hence, in these months, each projected precipitation intensity is lowered, resulting in a lower probability of flood conditions. Although this change factor artefact impairs the WG as well, the WG is still able to generate daily precipitation amounts triggering flood conditions due to its inherent stochasticity. Finally, changes in the day-to-day variability, more specifically in the transition probability of wet and dry spells, which are considered in only the WG and QM projections, lead to a contrasting change pattern of flood duration and flood volume compared to DC.

To complete the picture, we compared the change signals of dry spell lengths (Fig. 5a–f). As in the reference period, one must discriminate between shorter and longer spell lengths. In general, the changes are similar for all three postprocessing methods, but slightly deviating in the magnitude of the signals. We found DC and WG with rather unchanged shorter spells and more frequent longer spells of dry days. In contrast, QM projects more spell lengths of 5 to 8 days, at the expense of fewer long spells (10 or more) and fewer 1–day lengths.

4. Discussion

Do all postprocessing methods perform equally well in the present climate? One great advantage of the DC method lies in its observation-based reference climate (or reference run driven by observations), a fact that is often much appreciated by users. In contrast, WG and QM need to prove their capability of adequately reproducing the hydro-

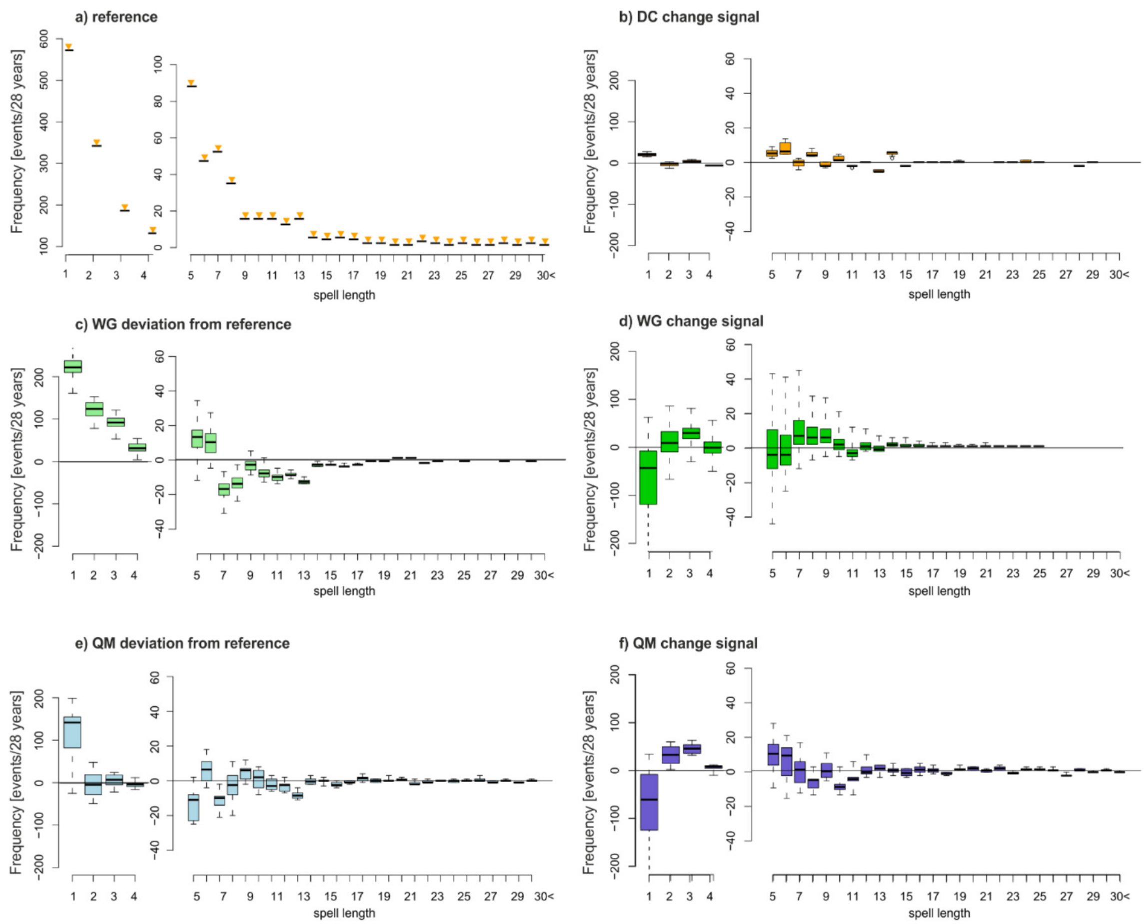

Fig. 5. Representations of dry spell length frequencies over the 28 years in the reference period (left panels) and changes in the projection period with respect to the reference. A dry spell is defined as a day with less than 0.1 mm of average precipitation over the catchment. Please note, WG of reference and projections period inheres a stochasticity component.

meteorological results and indices in the reference period prior to any interpretation of projections. For this Swiss example, we showed that both postprocessing approaches are widely capable of reproducing the hydro-climatic characteristics of the reference period, but reveal limitations for multiday extreme event magnitudes, such as Q347, drought intensity and event precipitation.

For QM, the amount of flood event rainfall, as well as the flood volume, are strongly overestimated. This could either indicate the (debated) presence of inflation (Maraun 2013; von Storch, 1999) or simply a less robust QM correction for the extreme values. The large variability of the 3 and 5daysRainSums for QM during the reference period (Fig. 4i, j) tends to support the latter rationale and is a likely explanation for this overestimation. However, Maraun (2013) demonstrated the effect of the omitted sub-grid variability that leads to increased spatial precipitation sums. Furthermore, the slight difference between the QM results and reference data in the occurrences of longer dry spells (1 days, and > 4 days) might also be ascribed to the omitted sub-grid variability, as smaller, local precipitation events can be spatially extended by the QM procedure. In a study evaluating QM in Switzerland for the same RCMs used here, Rajczak et al. (2016) demonstrated the overall very good performance of QM in reproducing dry and wet spells at individual stations. In their setting, however, stations and grid cells were always singular pairs, and they did not analyse spatial fields in their study.

Besides these extreme-related performance tests, QM-driven hydrological realizations also show deviations (greater than WG driven runs) in monthly mean values of runoff (Fig. 3) stemming from a misrepresentation of monthly mean precipitation. The most likely reason for this inaccuracy in the annual cycle of precipitation for QM in the calibration period is the fact that QM was calibrated using a 91-day moving window, which allows for slight deviations in the downscaled series with respect to the observed series in individual months. Furthermore, the deviations in areal precipitation can be explained by the fact that the QM approach underestimates the monthly precipitation at individual stations by approximately 10% (e.g., Säntis, Appenzell and Ebnat-Kappel, see Fig. 1). These stations exhibit the largest absolute precipitation sums over the catchment (see Fig. 1), and hence, their deviations have a stronger effect on the areal precipitation than the deviations of stations with lower precipitation sums. These limitations are clearly present and should be acknowledged in the communication with impact modellers and users who need to interpret this aspect with caution. However, the overall agreement of QM-based climate realizations with the reference is good.

Applying the WG data, duration-dependent indices (in particular indices that require or relate to weather sequences of > 2 days, see also Fig. 5) add to the list of underperforming aspects: drought duration, number of floods and multiday rainfall sums, and especially event precipitation. As stated earlier and shown by Keller et al. (2016), this failure can be traced back to the characteristic of first-order Markov chain as a core element of the WG. The better performance for flood duration likely relates back to the fact that flood durations defined by an exceedance of the Q999 value are seldom longer in the Thur catchment than one day (Schneeberger et al., 2018). More sophisticated, single-site weather generators, such as LARS_WG (Semenov and Stratonovitch, 2010), make use of a serial approach (Racsko et al., 1991) that has been shown to better capture the persistence of dry or wet events (Wilby et al., 2009). However, these serial WG approaches lack the ability to generate multisite rainfalls with the correct spatial structure. Aside from these limitations, mean monthly values, as well as the number of flood and drought events, are represented well.

4.1. What is the added value of using the updated climate projections?

Although the performance evaluation identifies the advantages and disadvantages of each postprocessing method, most users are likely more interested in the projected climate change signal. Thereby, the evaluation of the different hydrological projection is not as straightforward as the previous test in the reference period, as we lack a "correct" projection as a benchmark. Hence, we cannot judge which post-processing method is best performing but can only compare the climate change signals, check their plausibility, and compare these modelled signals with studies that analyse raw RCM output, such as those of Rajczak et al. (2013) and Fischer et al. (2015) for Switzerland, to ultimately conclude an added value.

First, it is notable that the differences in the change signals at the highest aggregation level, i.e., the water balance (Table 4), are rather negligible for all post-processing methods. However, a monthly perspective already reveals some larger differences. The good news is that all change signals point to the same pattern that has been found for many alpine and pre-alpine catchments (Köplin et al. 2012; Wagner et al. 2017): increasing winter runoff as a consequence of more rainfall and decreasing summer runoff as a result of less precipitation and enhanced evapotranspiration. Hence, the projection pattern is robust, but we found that the signal strength is influenced by the postprocessing method. Moreover, the signal strength is not only a function of the respective method but also varies for different methods in different months, affecting the projected change in hydrological regimes. Such seasonal differences in the change signal depending on the applied postprocessing method were also found by Hundecha et al. (2016) for small flood magnitudes. For the present study, we conclude that an added hydrological value using the new climate projections for Switzerland is already present at this rather high aggregation level.

Change signals for the extreme-related indices further accentuate the differences among the postprocessing approaches. For many indices, the established DC-realizations show the smallest change signals, while the signals are largest for QM. Moreover, the DC projections shows contradicting signals to the other two approaches. The likely rationale behind these different responses is two-fold: first, unintended effects of DC to reduce summer precipitation intensity due to lower summer mean precipitation sums, and second, the missing alteration of the temporal structure in the DC approach. This is in line with Vormoor *et al.* (2017) who proved the relevance of altered temporal sequences on the projections of different flow parameters and showed their major relevance for the extreme values in an alpine catchment. These caveats have always been known, but based on the present comparison with approaches that overcame these limitations, the downsides of DC can be quantified.

Despite these theory-based rationales the stronger and partly contrasting change signals obtained using QM-based realizations also need to be justified. These realizations show longer droughts and shorter flood duration under climate change projections. Furthermore, it projects more flood days with higher mean annual flood levels, but lower flood volumes and less precipitation per event—thus, an intensification of shorter flood events. These projections are in line with the analysis of RCM signals by Fischer et al. (2015). They found changes in the temporal precipitation structure and precipitation intensity and demonstrated a shift to longer dry spells leading to shorter but more intense wet spells. Furthermore, Rajczak et al. (2013) found that there are similar change signals for the mean and higher precipitation intensities (10–50%) shown for the study region. Therefore, we considered most of the projected values as valid, although the flood volume and the event precipitation derived from QM-based climate scenarios need to be interpreted with caution as they show strong deviations during the reference period and different change signals in comparison to DC. Our findings are in contrast to those of Snell et al. (2018) who compared DC and QM based projections for forest ecosystem services in Switzerland. In their case, climate model uncertainty was much more important than the postprocessing approach. We relate this disagreement to the different time scales and processes relevant for analysed impacts.

The main advantage of applying a WG—its ability to account for the variability and changes in temporal sequences and the generation of multiple ensembles—is partly contradicted by the rather simple first-order Markov chain of this multisite WG. The discussed limitations for multiday indices during the reference period also hold true for the projection periods with respect to changes in mean annual flood, changes in the number of flood days, and, to some degree, for changes in the multiday precipitation sums. Thus, the use of the WG data is primarily recommended for mean runoff values when multiple realizations are needed. However, several change signals are at the same order of magnitude as the QM-based projections; namely, flood and drought duration, as well as the change of Q347, and the number of drought and flood events. Hence, the WG is not as limited as expected given the stressed similarity of WG and DC results by Keller et al. (2016).

Finally, climate change signals for the dry spell distribution demonstrate the large impacts of the three different downscaling concepts. The analysis of Fischer et al. (2015) regarding changes in the temporal precipitation structure and its intensity demonstrate the shift to longer dry spells at the expense of shorter spells. This typical pattern is detected for the WG change signals in our study. This signal change is consistent with the findings of Keller et al. (2015), who identified an increase in the mean lengths of dry spells. Thus, although WG underestimates long dry spells, it simulates an increased number due to the influence of the climate models. The longer dry spell change signal of QM is presumably affected by the same spatiotemporal consistency problem that we discussed with regard to the QM performance compared to the observational data (see above).

4.2. General limitations of the present study

Apart from the limitations elaborated so far, some further limitations of the entire model setup must be discussed. First, the hydrological model itself is imperfect and is unable to capture the entire range of runoff characteristics measured; runoff is especially underestimated for the highest quantiles. Furthermore, the model overestimates the mean runoff in late summer and is, therefore, less sensitive to low flow conditions during the months when most drought events occur. Although we only compared model results with other model results to ensure relatively consistent results, this possible lack of sensitivity cannot be ignored.

A further source of uncertainty is the interpolation approach used to obtain meteorological input data for the entire catchment. Out of the possible interpolation approaches, we chose those that promised to introduce the fewest artefacts into the station input dataset. However, we cannot assess the extent of this limitation. This source of uncertainty is widely neglected in hydrological modelling and impact assessment studies and might be of high interest for further studies.

5. Conclusion

The starting points for this study were the following questions: Is there a potential added value of using the newly derived Swiss climate projections in hydrology compared to their DC-based precursor, and if so, in which respect? Not surprisingly, the answer is related to the variable of interest and the aspect of interest (averages or extremes). The new climate projections often, but not always, provide better or more trustworthy results. However, in cases in which the old scenarios are clearly limited by construction, the new projections clearly provide an added value, more so as comparisons with the old projections enable quantification of the shortcomings. For climate service providers, this information is presumably of high interest as it allows for a differentiated answer about the validity of past studies and the necessity to provide updates.

A first essential step in the evaluation of new climate projections is the comparison with the reference run. Here, we proved its value, as it enables the quantification of the quality of the postprocessing methods and the ability of the underlying climate model data to represent today's conditions. The comparison with the observation-driven reference run proved the overall suitability of the new CH2018 QM-based climate projections to represent the average and extreme hydro-climatic conditions. Only the water volume during flood events is overestimated, which is likely an effect of the omitted sub-grid variability and the less robust correction functions for precipitation extremes. However, minor deviations, especially in the mean monthly runoff values, occur, emphasizing again the need to strictly separate model and observation-based results in the reference period. This is especially emphasized in settings such as the present one when changing from the DC approach to other postprocessing methods.

Regarding the projected climate change signals (here exemplarily 2070–2099 vs. 1981–2010, SRES A1B), deviations between DC- and QM-based climate realizations occur already for mean monthly precipitation and runoff values yet are still within the excepted and known change signal pattern of pre-Alpine catchments. From a scientific perspective, multi-ensemble simulations and quantification of postprocessing uncertainty are surely an interesting option. From a user's perspective, it is questionable whether the found differences truly require new hydrological simulations.

This statement certainly does not extend to the projection of extremes. Indeed, we showed that DC can still be a valid option as the number of droughts as well as the days with runoff above the mean annual maximum flood showed similar projections. However, the strong added value of the new climate projections comes from the magnitude of extreme events, for both floods and droughts. Here, the commonly known limitations of DC are clearly overcome as the change signals of QM-based simulations are much closer to changes found in the raw RCM signals. The WG change signals are for several indices in line with the QM-based signals but have a lower magnitude. Due to their possibility to generate numerous realizations, the WG, therefore, likely remains an interesting option for some hydrological simulations in Switzerland.

Addor and Seibert (2014) pointed to a weakness of current postprocessing comparison studies: A structured analysis was and is still missing, and multiday-based statistics are required, especially for hydrological extremes. Here, we suggest an index-based comparison that covers different temporal aggregation levels, reference, and projection periods, as well as both extremes (high and low). The latter is partly based on multiday statistics and comprises different aspects of these extremes (intensity/volume and duration). While this suggestion works very well in the presented case by pointing out the weak and well-performing variables and aspects, it requires further applications in other comparison studies.

Based on the presented comparison results, one might also argue for the need for multi-postprocessing ensembles. We only partly agree with this opinion, as many differences found could be ascribed to a certain limitation of a certain postprocessing method. Hence, these methods do not generate equally valid datasets such as is the case of different climate models. In those cases, a multi-approach ensemble is invalid.

As the major differences found are closely related to the formulation of each postprocessing method and not to catchment-specific characteristics, we are confident that the general results derived from this example will also hold for applications in other catchments, at least in Switzerland. Clearly, the focus on one catchment is a limitation of this study, but necessary due to the computational effort a distributed, more physical-based model requires (especially when incorporating WG realizations).

We hope that the presented comprehensive analysis of the hydrological added value of updated climate projections will a) lead to a discussion of the value and content of structured comparison analyses, and b) help climate data providers, impact modellers and users to decide on their variable and aspect of interest if an update of hydrological projections is truly necessary, and what to expect when applying the new CH2018 climate projections.

Declaration of interests

None.

Acknowledgements

We would like to thank MeteoSwiss and FOEN for kindly supplying the meteorological and hydrological observations. Furthermore, we acknowledge the ENSEMBLES dataset from the EU-FP6 project ENSEMBLES. Finally, we thank Jan Rajczak (ETH Zurich) for sharing the QM code, and Luise Keller (GIUB) for the R-code on flood property extraction.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cliser.2019.01.001.

References

Addor, N., Seibert, J., 2014. Bias correction for hydrological impact studies – beyond the daily perspective. Hydrol. Process. 28, 4823–4828. https://doi.org/10.1002/hyp.10238.

Boé, J., Terray, L., Habets, F., Martin, E., 2007. Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. Int. J. Climatol. 27, 1643–1655. https://doi.org/10.1002/joc.1602.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resour. Res. 49, 1523–1536. https://doi.org/10.1029/2011WR011533.

Bosshard, T., Kotlarski, S., Ewen, T., Schär, C., 2011. Spectral representation of the annual cycle in the climate change signal. Hydrol. Earth Syst. Sci. 15, 2777–2788. https://doi.org/10.5194/hess-15-2777-2011.

CH2011. Swiss Climate Change Scenarios CH2011 2011 Zurich.

CH2018. (2018). CH2018 – Climate Scenarios for Switzerland: Technical Report. Zurich: National Centre for Climate Services.

Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resour. Res. 49, 4187–4205. https://doi.org/10.1002/wrcr.20331.

Déqué, M., Rowell, D.P., Lüthi, D., Giorgi, F., Christensen, J.H., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., van den Hurk, B., 2007. An intercomparison of regional climate simulations for Europe. Assessing uncertainties in model projections. Clim. Change 81, 53–70. https://doi.org/10.1007/s10584-006-9228-x.

Diaz-Nieto, J., Wilby, R.L., 2005. A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. Clim. Change 69, 245–268. https://doi.org/10.1007/s10584-005-1157-6.

Dobler, C., Hagemann, S., Wilby, R.L., Stötter, J., 2012. Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed. Hydrol. Earth Syst. Sci. 16, 4343–4360. https://doi.org/10.5194/hess-16-4343-2012.

Fang, G.H., Yang, J., Chen, Y.N., Zammit, C., 2015. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. Hydrol. Earth Syst. Sci. 19, 2547–2559. https://doi.org/10.5194/hess-19-2547-2015.

Fischer, A.M., Keller, D.E., Liniger, M.A., Rajczak, J., Schär, C., Appenzeller, C., 2015. Projected changes in precipitation intensity and frequency in Switzerland: a multi-model perspective. Int. J. Climatol. 35, 3204–3219. https://doi.org/10.1002/joc.4162.

Graham, L.P., Andréasson, J., Carlsson, B., 2007. Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods – a case study on the Lule River basin. Clim. Change 81, 293–307. https://doi.org/10.1007/s10584-006-9215-2.

Grillakis, M.G., Koutroulis, A.G., Daliakopoulos, I.N., Tsanis, I.K., 2017. A method to preserve trends in quantile mapping bias correction of climate modeled temperature. Earth Syst. Dynam. 8, 889–900. https://doi.org/10.5194/esd-8-889-2017.

Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note. Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390. https://doi.org/10.

5194/hess-16-3383-2012.
Gutiérrez, J.M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M.J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D.E., Fischer, A.M., Cardoso, R.M., Soares, P.M.M., Czernecki, B., Pagé, C., 2018. An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. Int. J. Climatol. 28, 183. https://doi.org/10.1002/joc.5462.
Hundecha, Y., Sunyer, M.A., Lawrence, D., Madsen, H., Willems, P., Bürger, G., Kriaučiūnienė, J., Loukas, A., Martinkova, M., Osuch, M., Vasiliades, L., von Christierson, B., Vormoor, K., Yücel, I., 2016. Inter-comparison of statistical downscaling methods for projection of extreme flow indices across Europe. J. Hydrol. 541, 1273–1286. https://doi.org/10.1016/j.jhydrol.2016.08.033.
Ivanov, M.A., Kotlarski, S., 2016. Assessing distribution-based climate model bias correction methods over an alpine domain. Added value and limitations. Int. J. Climatol. https://doi.org/10.1002/joc.4870.
Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., Yiou, P., 2014. EURO-CORDEX. New high-resolution climate change projections for European impact research. Reg. Environ. Change 14, 563–578. https://doi.org/10.1007/s10113-013-0499-2.
Keller, D.E., Fischer, A.M., Frei, C., Liniger, M.A., Appenzeller, C., Knutti, R., 2015. Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment. Hydrol. Earth Syst. Sci. 19, 2163–2177. https://doi.org/10.5194/hess-19-2163-2015.
Keller, D.E., Fischer, A.M., Liniger, M.A., Appenzeller, C., Knutti, R., 2016. Testing a weather generator for downscaling climate change projections over Switzerland. Int. J. Climatol. https://doi.org/10.1002/joc.4750.
Khalili, M., Brissette, F., Leconte, R., 2011. Effectiveness of Multi-Site Weather Generator for Hydrological Modeling1. JAWRA J. Am. Water Resour. Assoc. 47, 303–314. https://doi.org/10.1111/j.1752-1688.2010.00514.x.
Klein, T., Holzkämper, A., Calanca, P., Seppelt, R., Fuhrer, J., 2013. Adapting agricultural land management to climate change. A regional multi-objective optimization approach. Landscape Ecol. 28, 2029–2047. https://doi.org/10.1007/s10980-013-9939-0.
Köplin, N., Schädler, B., Viviroli, D., Weingartner, R., 2012. Relating climate change signals and physiographic catchment properties to clustered hydrological response types. Hydrol. Earth Syst. Sci. 16, 2267–2283. https://doi.org/10.5194/hess-16-2267-2012.
Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., Wulfmeyer, V., 2014. Regional climate modeling on European scales. A joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci. Model Dev. 7, 1297–1333. https://doi.org/10.5194/gmd-7-1297-2014.
Lenderink, G., van Ulden, A., van den Hurk, B., van Meijgaard, E., 2007. Summertime inter-annual temperature variability in an ensemble of regional model simulations: analysis of the surface energy budget. Clim. Change 81, 233–247. https://doi.org/10.1007/s10584-006-9229-9.
Maraun, D., 2013. Bias correction, quantile mapping, and downscaling. Revisiting the inflation issue. J. Climate 26, 2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1.
Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L.O., 2017. Towards process-informed bias correction of climate change simulations. Nature Clim. Change 7, 664–773. https://doi.org/10.1038/NCLIMATE3418.
Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48. https://doi.org/10.1029/2009RG000314.
Maraun, D., Widmann, M., 2015. The representation of location by a regional climate model in complex terrain. Hydrol. Earth Syst. Sci. 19, 3449–3456. https://doi.org/10.5194/hess-19-3449-2015.
Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R., Wilcke, R.A.I., 2015. VALUE. A framework to validate downscaling approaches for climate change studies. Earth's Future 3, 1–14. https://doi.org/10.1002/2014EF000259.
MeteoSwiss (Ed). 2016. Analyse der Nutzerbedürfnisse zu nationalen Klimaszenarien.
Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900. https://doi.org/10.13031/2013.23153.
Nikulin, G, Bosshard, T, Yang, W, Bärring, L, Wilcke, R, Vrac, M, Vautard, R, Noel, T, Gutiérrez, JM, Herrera, S, Fernández, J, Haugen, JE, Benestad, R, Landgren, OA, Grillakis, M, Ioannis, T, Koutroulis, A, Dosio, A, Ferrone, A, Switanek, M. 2015. Bias Correction Intercomparison Project (BCIP): an introduction and the first results 17.
Piani, C., Haerter, J.O., Coppola, E., 2010a. Statistical bias correction for daily precipitation in regional climate models over Europe. Theor. Appl. Climatol. 99, 187–192. https://doi.org/10.1007/s00704-009-0134-9.
Piani, C., Weedon, G.P., Best, M., Gomes, S.M., Viterbo, P., Hagemann, S., Haerter, J.O., 2010b. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J. Hydrol. 395, 199–215. https://doi.org/10.1016/j.jhydrol.2010.10.024.
Quintana Seguí, P., Ribes, A., Martin, E., Habets, F., Boé, J., 2010. Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins. J. Hydrol. 383, 111–124. https://doi.org/10.1016/j.jhydrol.2009.09.050.
Racsko, P., Szeidl, L., Semenov, M., 1991. A serial approach to local stochastic weather models. Ecol. Model. 57, 27–41. https://doi.org/10.1016/0304-3800(91)90053-4.
Rajczak, J., Kotlarski, S., Schär, C., 2016. Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell lengths? J. Climate 29, 1605–1615. https://doi.org/10.1175/JCLI-D-15-0162.1.
Rajczak, J., Pall, P., Schär, C., 2013. Projections of extreme precipitation events in regional climate simulations for Europe and the Alpine Region. J. Geophys. Res.: Atmos. 118. https://doi.org/10.1002/jgrd.50297.
Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17, 182–190. https://doi.org/10.1029/WR017i001p00182.
Rössler, O., Diekkrüger, B., Löffler, J., 2012. Potential drought stress in a Swiss mountain catchment-Ensemble forecasting of high mountain soil moisture reveals a drastic decrease, despite major uncertainties. Water Resour. Res. 48, W04521. https://doi.org/10.1029/2011WR011188.
Rössler, O., Fischer, A.M., Huebener, H., Maraun, D., Benestad, R.E., Christodoulides, P., Soares, P.M.M., Cardoso, R.M., Pagé, C., Kanamaru, H., Kreienkamp, F., Vlachogiannis, D., 2017. Challenges to link climate change data provision and user needs - perspective from the COST-action VALUE. Int. J. Climatol. 50, 7541. https://doi.org/10.1002/joc.5060.
Schneeberger, K., Rössler, O., Weingartner, R., 2018. Spatial patterns of frequent floods in Switzerland, patial patterns of frequent floods in Switzerland. Hydrol. Sci. J./J. Sci. Hydrol. https://doi.org/10.1080/02626667.2018.1450505.
Schulla, J. 2015. Model Description WaSiM. http://www.wasim.ch/ (accessed 30 July 2017).
Semenov, M.A., Stratonovitch, P., 2010. Use of multi-model ensembles from global climate models for assessment of climate change impacts. Clim. Res. 41, 1–14. https://doi.org/10.3354/cr00836.
Snell, R.S., Elkin, C., Kotlarski, S., Bugmann, H., 2018. Importance of climate uncertainty for projections of forest ecosystem services. Reg. Environ. Change 22, 42. https://doi.org/10.1007/s10113-018-1337-3.
von Storch, H., 1999. On the use of "inflation" in statistical downscaling. J. Climate 12, 3505–3506. https://doi.org/10.1175/1520-0442(1999) 012<3505:OTUOII>2.0.CO;2.
Teng, J., Potter, N.J., Chiew, F.H.S., Zhang, L., Wang, B., Vaze, J., Evans, J.P., 2015. How does bias correction of regional climate model precipitation affect modelled runoff? Hydrol. Earth Syst. Sci. 19, 711–728. https://doi.org/10.5194/hess-19-711-2015.
Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J. Hydrol. 456–457, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052.
Themeßl, J., Gobiet, A., Leuprecht, A., 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int. J. Climatol. 31, 1530–1544. https://doi.org/10.1002/joc.2168.
Themeßl, M.J., Gobiet, A., Heinrich, G., 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Clim. Change 112, 449–468. https://doi.org/10.1007/s10584-011-0224-4.
van der Linden, P, Mitchell, JFB. 2009. ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK, Exeter (accessed 9 July 2015).
Wagner, T., Themeßl, M., Schüppel, A., Gobiet, A., Stigler, H., Birk, S., 2017. Impacts of climate change on stream flow and hydro power generation in the Alpine region. Environ Earth Sci 76. https://doi.org/10.1007/s12665-016-6318-6.
Wilby, R.L., Troni, J., Biot, Y., Tedd, L., Hewitson, B.C., Smith, D.M., Sutton, R.T., 2009. A review of climate risk information for adaptation and development planning. Int. J. Climatol. 29, 1193–1215. https://doi.org/10.1002/joc.1839.
Wilcke, R.A.I., Mendlik, T., Gobiet, A., 2013. Multi-variable error correction of regional climate models. Clim. Change 120, 871–887. https://doi.org/10.1007/s10584-013-0845-x.
Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol. 210, 178–191. https://doi.org/10.1016/S0022-1694(98)00186-3.
WMO, 2009. Manual on Low-Flow Estimation and Prediction. World Meteorological Organization, Geneva, Switzerland.