Climatic Change
DOI 10.1007/s10584-017-1940-1

When timing matters-considering changing temporal structures in runoff response surfaces

Klaus Vormoor[1] · **Ole Rössler**[2,3] · **Gerd Bürger**[1] · **Axel Bronstert**[1] · **Rolf Weingartner**[2,3]

Received: 26 January 2016 / Accepted: 6 March 2017

Abstract Scenario-neutral response surfaces illustrate the sensitivity of a simulated natural system, represented by a specific impact variable, to systematic perturbations of climatic parameters. This type of approach has recently been developed as an alternative to top-down approaches for the assessment of climate change impacts. A major limitation of this approach is the underrepresentation of changes in the temporal structure of the climate input data (i.e., the seasonal and day-to-day variability) since this is not altered by the perturbation. This paper presents a framework that aims to examine this limitation by perturbing both observed *and* projected climate data time series for a future period, which both serve as input into a hydrological model (the HBV model). The resulting multiple response surfaces are compared at a common domain, the *standardized* runoff response surface (SRRS). We apply this approach in a case study catchment in Norway to (i) analyze possible changes in mean and extreme runoff and (ii) quantify the influence of changes in the temporal structure represented by 17 different climate input sets using linear mixed-effect models. Results suggest that climate change induced increases in mean and peak flow runoff and only small changes in low flow. They further suggest that the effect of the different temporal structures of the climate input data considerably affects low flows and floods (at least 21% influence), while it is negligible for mean runoff.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-017-1940-1) contains supplementary material, which is available to authorized users.

✉ Klaus Vormoor
kvormoor@uni-potsdam.de

[1] Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

[2] Oeschger Centre for Climate Change Research, University of Bern, Falkenplatz 16, 3012 Bern, Switzerland

[3] Institute of Geography, University of Bern, Hallerstr. 12, 3012 Bern, Switzerland

Published online: 27 March 2017

1 Introduction

Climate change impact assessments provide a valuable basis for the development of adaptation measures. Typically, they follow the so-called top-down approach which utilizes general circulation model (GCM) and/or regional climate model (RCM) scenario data in impact models, as for example hydrological models. Usually, some kind of downscaling is needed that is either downscaling in sensu strictu bridging the gap in the spatial or even temporal resolution or bias correction of GCM or RCM data compared to climate observations (Fowler et al. 2007; Maraun et al. 2010).

As an alternative to the top-down approach, quite recently, the so-called scenario-neutral bottom-up approach has been introduced by Prudhomme et al. (2010) and Wehren (2010). This approach performs a sensitivity study of an environmental system (represented by a model) towards changes in the climate system and summarizes its response in a two-dimensional matrix, the so-called response surface. Each pixel within this domain shows the average change of a target variable (the impact) of a certain combination of systematically perturbed climate input data. Different projected changes of the climate system can afterwards be located in the response surface enabling a probabilistic assessment of the climate impact. Response surfaces have been used in a variety of research fields such as agriculture (Hirschi et al. 2012), ecology (Fronzek et al. 2011), and hydrology (Bastola et al. 2011; Wetterhall et al. 2011).

The major advantage of this approach is its flexibility towards new climate data sets that can easily be incorporated into the original surface to receive an updated version. Furthermore, it provides an intuitive illustration of the uncertainties of the impact response relative to the spread of climate projections within the response surface: the wider the spread the higher the uncertainties, and response values covered by the spread show the range of possible impacts.

A challenging step in the application of response surfaces is the adequate choice of the most sensitive climate parameters for the model response that are at the same time also sensitive under climate change conditions. The second challenge is the relative high computational cost, which is required to produce a response surface in the first place. Finally, yet importantly, a huge challenge is the choice of the appropriate perturbation method applied to generate the response surface by altering the climate parameters. Reviewing previous studies, at least three different approaches are followed: (i) linear scaling of observed time series to the level expected under climate change projections (e.g., Fronzek et al. 2011), (ii) linear scaling of observed time series with different seasonal change factor scenarios (e.g., Prudhomme et al. 2010; Wetterhall et al. 2011), and (iii) application of weather generators in combination with change factors (e.g., Wehren 2010; Bastola et al. 2011).

Critical to all approaches is that the temporal sequencing of the climate input data (meaning the annual, seasonal, and day-to-day climate variability) remains statistically unchanged. Changes in seasonal variability are partly covered by applying monthly change factors (e.g., Prudhomme et al. 2010), but statistical measures of the observed time series like wet and dry spells are not altered (in terms of sequencing) but only modulated as a result of the change factors. However, particularly in hydrology, extreme events like floods and droughts are highly influenced by the day-to-day variability in a certain season (e.g., rainy days in the snow melt season or dry days in summer). Thus, the effects of changes in climate variability are important to consider and they might have been considered inappropriately in the existing studies.

In this paper, we present a framework for incorporating changes in the temporal structure of the climate data in the construction of runoff response surfaces. This allows us to analyze the effect of changing temporal structures on key hydrological variables like mean flow, floods,

and droughts. Assuming that climate model future projections inhere good estimates of changing temporal structures, we apply linear scaling to both observed *and* projected climate data from eight GCM-RCM combinations, downscaled by two different methods. This results 17 different response surfaces (one generated with perturbed climate observation; 16 with perturbed future projections) which can be compared only if the corresponding perturbations pertain to the same base state. Therefore, the projected climate data need to be remapped to observations, which we do by removing the mean climate change signal, resulting in a *standardized* response surface. This way, we are able to assess the impact of changes in the temporal sequencing covered by the climate projections since the differences between the standardized response surfaces can be interpreted as the result of the differing temporal structures of the climate input data. Finally, linear mixed-effect models are applied to quantify the effects of considering (or disregarding) changes in the temporal structure in the construction of response surfaces.

We apply this approach in the field of hydrology and focus on different aspects of runoff response, covering mean runoff as well as extremes like low flows and floods in terms of their magnitude and seasonality in one Nordic catchment that is characterized by a mixed snowmelt/rainfall regime. The following two particular research questions are addressed by this paper: (i) What are the possible impacts of climate change on the previously mentioned aspects of runoff response derived by our proposed approach? (ii) What are the quantitative effects of the (un)changed temporal structure on mean runoff and extreme values in bottom-up approaches? This setting allows drawing conclusions about the validity of change factor approaches with unchanged temporal structures in general, and for bottom-up approaches in specific.

2 Material and methods

2.1 Study catchment

The analysis is carried out within the unglaciated mesoscale Kråkfoss catchment in southeast Norway (S1 in the supplement). The catchment is included in the benchmark dataset for climate change studies in Norway and tested for its suitability for daily analysis of mean-, flood-, and low flow discharge (Fleig et al. 2013). The catchment is 433 km^2 in size and its elevation ranges from 105 to 803 m a.s.l. with a median elevation of 445 m a.s.l. The mean annual precipitation sum is 910 mm and the mean annual runoff is about 600 mm (for the period 1961–1990). Runoff coefficients are rather high due to low evapotranspiration. Annual flood peaks either occur during spring and early summer, usually triggered by snowmelt or a combination of snowmelt and rainfall, or during autumn and early winter triggered by rainfall only. The magnitude of the autumn events tends to be slightly larger than those occurring during spring and early summer, whereas the frequency of spring and early summer events is slightly larger as compared with autumn events. Boreal forest is the dominant land cover (76%); lakes, mashes, and bogs account in total for 9% land cover. The influence of anthropogenic land use such as farming and urban settlements is rather small (11.2%).

2.2 Generating SRRS

Runoff response surfaces plot at a two-dimensional domain, the long-term means of a target variable (e.g., mean runoff) simulated by a hydrological model that has been driven with

systematically perturbed climate input data (e.g., Prudhomme et al. 2010). That is, observed daily time series of crucial meteorological variables such as temperature (T_0) and precipitation (P_0) are scaled within a given range of allowed change (the "user-defined" climate change signal). The simplest way of perturbing T- and P-data series is linear scaling:

$$T(i) \mapsto T_0(i) + X_t \tag{1}$$

$$P(i) \mapsto P_0(i) X_p \tag{2}$$

where X_t and X_p are the perturbation factors which incrementally scale the observed meteorological input data (additive for T, and multiplicative for P), and i indicates the daily time step. In this study, observed T is scaled with 0.5 °C increments so that the mean T of the perturbed time series ranges from 2.5 to 9.5 °C (15 perturbations). P is scaled with 100-mm increments to reflect the range of 800–1500 mm mean annual P sum (8 perturbations). The hydrological impact model is then run with all possible combinations of perturbed input data series and the result of each combination (i.e., the impact) is plotted as one single realization (i.e., one pixel) of possible climate change impacts over a two-dimensional domain (15 × 8 = 120 realizations). In other words, runoff response surfaces reflect the sensitivity of the hydrological impact model (or the catchment, respectively) to systematically changed climatological input data.

In the approach presented here, this concept is extended by not only scaling the observed climate time series but also the outputs of eight GCM-RCM combinations for a future period (2071–2099), downscaled in two different ways. To ensure the comparability between these response surfaces, the perturbation needs to pertain to the same base state so that all grids reflect the same dimensions (2.5–9.5 °C mean annual T on the x-axis; and 800–1500 mm mean annual P on the y-axis) irrespective of the climate input data used. The identical base state is obtained by removing the mean climate change signal from the corresponding climate projections, which is done additively for T and multiplicatively for P:

$$T(i) \mapsto T(i) - \left(\overline{T(i)} - T_0\right) + X_t \tag{3}$$

$$P(i) \mapsto P(i) \frac{P_0}{\overline{P(i)}} X_p, \tag{4}$$

with T_0 and P_0 being the observed mean states for T and P, respectively. That way, the scaling of the different input data series results in a standardized domain: the standardized runoff response surface (SRRS). If changes in the temporal structures of T and P, which are assumed to be represented by the downscaled RCMs, do not have an influence on the hydrological impact of interest, the SRRSs should be identical, irrespective of the data series used. In turn, every difference between the SRRSs can be interpreted as a result of the different temporal structure. In total, we generate 17 SRRSs, one based on observations and 16 based on eight GCM-RCM combinations that were downscaled and bias corrected, respectively, applying two different techniques (Section 2.4). That is, the SRRSs generated here are based on 17 × 15 (T) × 8 (P) = 2040 hydrological model simulations.

2.3 Temperature and precipitation data used for linear scaling

Gridded observation data for temperature and precipitation is available for the whole of Norway with a spatial resolution of 1 × 1 km^2 (seNorge data; e.g. Mohr and Tveito 2008). Daily mean temperature and precipitation for the domain of the study catchment and for the

period 1980–2008 (29 years) were extracted from the grids and applied as input into the HBV hydrological model (Section 2.5).

The climate projections are based on eight GCM-RCM combinations (Table S2 in the supplement) from the EU FP6 ENSEMBLES project (van der Linden and Mitchell 2009). The spatial resolution of the RCMs considered is 0.22° (approximately 25 km), and projections of daily values are available for the period 1950–2100. Within this study, we apply temperature and precipitation data for the 2071–2099 future period (29 years) for which the climate models are consistently driven by the SRES A1B scenario (IPCC 2000). The climate projections for the future period are assumed to represent changes in the annual, seasonal, and day-to-day variability of temperature and precipitation compared to current conditions (see Section 3.1).

2.4 Local adjustment of the RCM projections

Since the RCM outputs cannot be directly used for hydrological impact modeling due to limited process description and insufficient spatial resolution of the RCMs relative to the catchment scale, bias correction and downscaling is necessary to adjust large-scale RCM data to the catchment scale. A wide range of such methods are available and reviews on available approaches are given by, e.g., Maraun et al. (2010) and Teutschbein and Seibert (2010). In this study, we apply two methods for locally adjusting biased GCM-RCM data to the catchment scale: expanded downscaling (XDS) (Bürger 1996; Bürger et al. 2009) and empirical quantile mapping (QM) (Boé et al. 2007; Gudmundsson et al. 2012). XDS is a statistical downscaling method based on multiple linear regression but expanded by a side condition for preserving local co-variability between the variables so that the estimation of extremes is improved. QM represents a bias correction method that applies a transfer function to adjust the quantiles of biased RCM data to those of locally observed data. For more detailed descriptions of both methods, we refer to the supplement of this paper (S3 and S4).

2.5 The hydrological target variables

We generated SRRSs applying the "Nordic" version of the conceptually lumped HBV hydrological rainfall-runoff model (Bergström 1995; Saelthun 1996) using the linearly scaled T and P data mentioned above as inputs to the model. The model is applied on a daily time step and consists of three basic subroutines: (i) a semi-distributed snow routine representing 10 equal-area height zones, (ii) a lumped soil moisture routine, and (iii) a routine for runoff response. The advantage of the HBV model compared to more physically based hydrological models is that it only requires temperature and precipitation as input data. Evaporation is estimated by the model using the temperature index method. The model is calibrated using the PEST parameter estimation routines (Doherty 2004), and further details can be found in Lawrence et al. (2009). Four hydrological target variables are addressed within this study:

1. Mean daily discharge (MQ)
2. The mean magnitude of the annual maximum floods (AMFs)
3. The flood seasonality index S_D (Vormoor et al. 2015), which is a well-suited measure for characterizing the seasonality of AMFs in catchments that have two

distinctive flood seasons (spring/summer vs. autumn/winter), as it is the case in our study catchment:

$$S_D = \frac{\sum AMF_{\text{Sep-Feb}} - \sum AMF_{\text{Mar-Aug}}}{\sum AMF_{\text{all}}}, \tag{5}$$

which is the difference between the sum of the annual flood peaks occurring within September–February and those within March–August over the sum of all AMFs. The index ranges from −1 to +1, and negative numbers indicate dominant events during spring and summer, while positive numbers indicate dominant events during autumn and winter.

4. The low flow index LF7, which describes the mean of the annual 7-day minimum streamflow, derived for the period May–November to exclude winter low flows caused by water storage in snow and ice.

2.6 Quantifying the influence of perturbation and temporal structures on the overall variance

To analyze the relative influence of the perturbed parameters (T, P) and their temporal structure (V) on the runoff variables Q_i, we apply linear mixed-effect (LME) models that are an extension of linear regression models. LME models distinguish random and fixed effects. Random effects are predictors that are drawn out of a greater population, while fixed effects, i.e., predictors of the normal linear regression, are either the entire population or repeatable levels of experimental factors (Pinheiro and Bates 2004). A review including a user guide for LME models is published by Bolker et al. (2009) and an introducing tutorial is presented by Winter (2013).

The application of LME models is necessary since the temporal structure V is only a limited selection of many possible temporal structures. Moreover, the temporal structures of the future climate projections are non-independent since they are based on five GCMs, only. In this study, T and P are regarded as fixed effects and V represents random realizations (i.e., 17 perturbed time series) of many possible temporal structures. Since the realizations of V are not independent but rely on five GCMs (plus observations) and two downscaling methods, we itemized the variable V into the factors DS (downscaling) and G (originating GCM and observations, respectively). The mathematical structure of a LME model is very similar to a linear regression model, and Eq. 6 shows a well-known linear regression structure that is extended by our two random predictors denoted $(1|DS)$ and $(1|G)$.

$$Q^i_{j,k,l,m} = a + bT_j + cP_k + (1|DS_l) + (1|G_m) + \varepsilon, \tag{6}$$

with a, b, and c as the regression coefficients, Q_i as the runoff variable i = 1,…, 4, that is MQ, AMF, S_D, LF7, T_j as the temperature levels T_j of the SRRS, P_k as the precipitation levels P_k of the SRRS, DS_l as the factor respecting the temporal structure of each downscaled realization to be perturbed, with l = 1,…,17 (8 structures from XDS; 8 structures from QM; one structure from observations), G_m as the factor respecting the temporal structure of the underlying GCMs and observations, respectively (m: 5 GCMs and observations), and ε as the unexplained variance.

The influence of each predictor is expressed by the fraction of explained- over the total variance. To elucidate each factor's influence, we conduct a series of models with increasing complexity: First, we assess the variance explained by each fixed parameter applying two univariate regression models (models I and II, Table 1). Second, we add the second fixed predictor and evaluate the parameter's interaction (models III and IV, Table 1). Third, we add one of the two random predictors (models V and VI, Table 1), before we finally conduct the full model with and without interactions (models VII and VIII, Table 1). Each model is tested for its disparity to the next simpler model on the 5% significance level using the likelihood ratio approach. As *DS* is widely, apart from observations, a specification of the information included in *G* (i.e., each downscaling method is applied to every GCM-RCM combination), we assess the influence of *G* and *DS* separately (models V and VI, Table 1). Differences of these three parameter models to the full interaction model (model VIII) are related to interactions between all factors. Based on this series of models, the influences of each predictor can be derived from the differences in explained variances of models.

The procedure is applied to all 17 SRRS data points, with each 8 (P) × 15 (T) = 120 samples. Only significant predictors (at the 99% level) are used in the models. To ensure homoscedasticity, we have transformed the target variables with a log-transformation, and for S_D, we additionally have scaled the values with +1.

3 Results

3.1 Characteristic values of the climate input data sets

We have calculated six climate indices to explore the differences between the climate data sets that are used to generate the SRRSs (observed data; downscaled RCM data) and the RCM raw data. Detailed results can be found in the supplement to this paper (S5). The differences in the estimated indices compared to observations can be seen as a result of changes in the temporal structure of the (downscaled) climate projections. We have analyzed the number of wet days per year (rain >1 mm day^{-1}) (S5a), the wet day intensity (mm per rainy day) (S5b), the maximum wet spell length in the time period (days) (S5c), the number of ice days (S5d), the

Table 1 Series of LME models applied and the corresponding estimates of explained variance for the four hydrological target variables mean discharge (MQ), mean annual maximum floods (AMF), the flood seasonality index SD, and the low flow index LF7.

		Explained variance			
Model ID	Model structure	MQ	AMF	S_D	LF7
I	T	0.031	0.062	0.52	0.18
II	P	0.935	0.686	0.02	0.5
III	$T + P$	0.967	0.74	0.55	0.68
IV	$T \times P$	0.967	0.75	0.56	0.68
V	$T + P + G$	0.979[a]	0.89	0.59[a]	0.917
VI	$T + P + DS$	0.989	0.955	0.801	0.983
VII	$T + P + G + DS$	0.989	0.955	0.801	0.983
VIII	$T \times P \times (1\|G) \times (1\|DS)$	0.99	0.958	0.816	0.983

[a] Not significantly different from model III

Markov Chain stagnation probability (S4e) that is the probability of a dry (wet) day to be followed by a dry (wet) day, and the sum of precipitation over 5 days (mm) (S5f).

In general, both downscaling techniques reduce both the spread and the magnitude of the climate change signal given by the RCM raw data. Only regarding wet day intensity, the downscaled RCM data show larger values than the RCM raw data. The analysis illustrates not only the different temporal structures of the data set used to construct the different SRRSs but also the different effects of the two different downscaling methods. That is, the methods used for downscaling the climate projections also introduce modifications in the temporal structure of the data sets. While QM is designed to preserve the temporal structure of the climate projections, XDS modulates the temporal structure as a result of linear mapping of multiple atmospheric fields (predictors) to the local scale.

3.2 SRRSs for the four hydrological target variables

For each of the four hydrological target variables, we have generated 17 SRRSs (one using observed climate data; 16 downscaled climate projections). They are shown in the supplement to this paper (S6–S9). To aggregate the information of these plots, for each variable, we have estimated SRRSs showing the mean of the ensemble; the differences between the ensemble members are summarized in SRRSs showing the coefficient of variation (CV) for MQ, AMF, and LF7, and the standard deviation for S_D (Fig. 1a–h). The CV and standard deviation indicate the influence of the different climate input data sets and particularly their temporal structures on the SRRSs. Note that due to the dependency of some climate projections (i.e., same underlying GCM (S2)), we have iteratively estimated the mean and the CV for every independent subsample of downscaled climate projections. The ensemble mean and CV shown here represent the mean for each variable over the multiple subsamples.

3.2.1 Mean daily discharge (MQ)

Figure 1a demonstrates that the ensemble mean for MQ ranges from ~5 to about ~13.4 $m^3 s^{-1}$. As expected, the lowest values are found for the combinations with the highest mean annual *T* and the lowest mean annual *P* sums, while the largest values are found for the lowest *T* and highest *P* combinations. Changes in MQ are dominantly driven by scaling *P*. Only some smaller changes occur due the scaling of *T* which affects the relative role of snowmelt. Today's MQ of ~7 $m^3 s^{-1}$ (black 1) will likely increase under climate change conditions, although the spread of projections (red and green numbers) is large. There is a high level of agreement between the individual ensemble members (S6) as expressed by the low CV values shown in Fig. 1b (2.9–6.2%). This indicates a small influence of the different temporal structures on mean discharge.

3.2.2 Mean magnitude of the AMF

Changes in the mean magnitude of AMFs follow a similar pattern as the changes in MQ, indicating that they are mainly driven by scaling *P* (Fig. 1c). The lowest magnitudes of AMFs (~43 $m^3 s^{-1}$) occur for combinations with the highest *T* and lowest *P*; the highest magnitudes of AMFs (~109 $m^3 s^{-1}$) are found for the opposite combinations. Likewise, the projections (red and green numbers) indicate likely increases in the magnitudes of AMFs, though with a

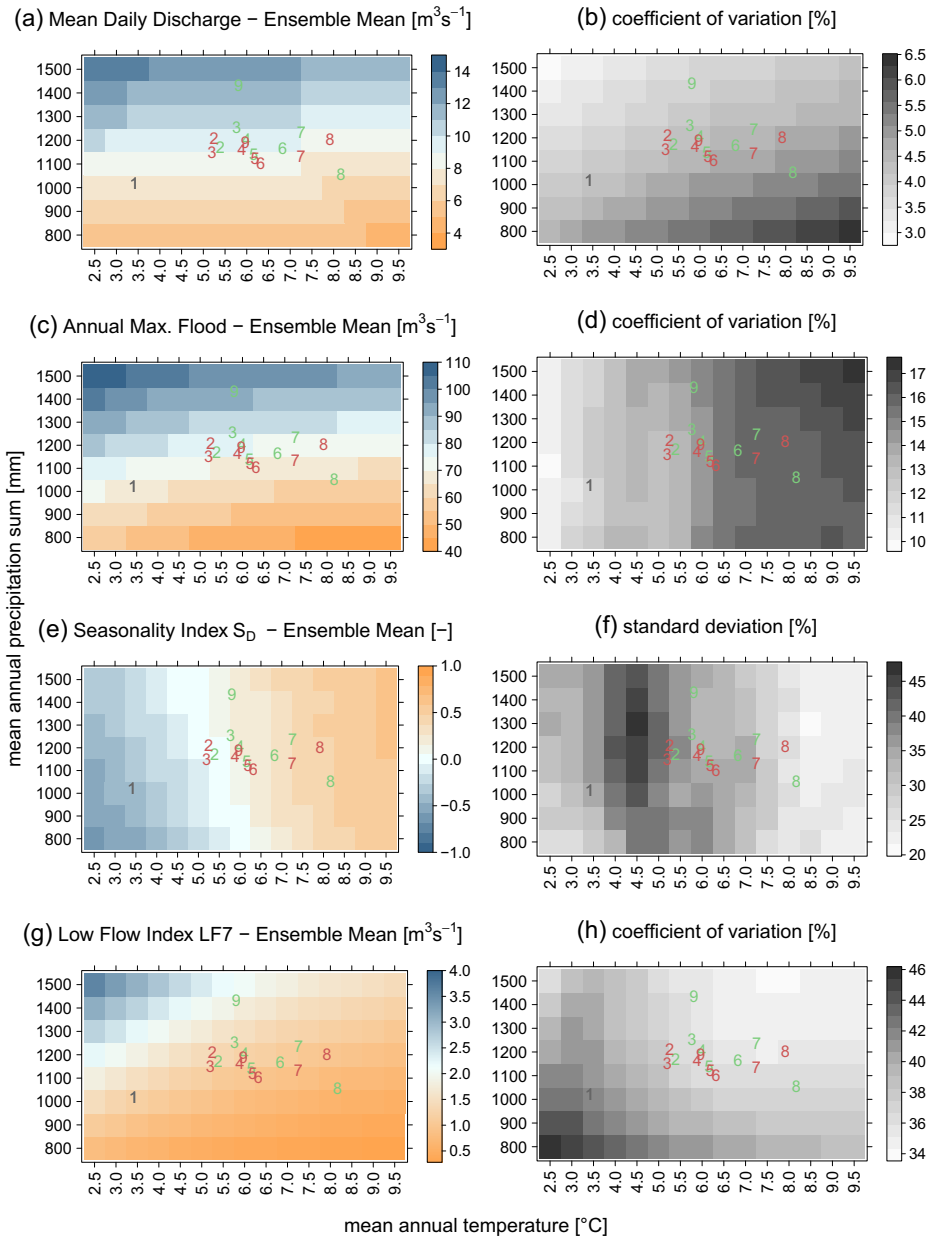

Fig. 1 **a**–**h** Standardized runoff response surfaces (SRRSs) showing the mean (*left column*) and coefficient of variation (standard deviation for S_D) (*right column*) for the four hydrological target variables obtained from multiple subsampling of the entire ensemble of linearly scaled input data. The numbers in each plot indicate where the particular data set lies prior to the perturbation. Observed time series (*1*), BCA_HIRHAM (*2*), BCA_RCA (*3*), ECHAM5_RACMO (*4*), ECHAM5_REGCM (*5*), ECHMA5_REMO (*6*), HadCM3Q0_HadRM3Q0 (*7*), HadCM3Q16_HadRM3Q16 (*8*), and HadCM3Q0_HadRM3Q0 (*9*). *Colors* represent the method applied for the local adjustment of the RCM projections. *XDS* expanded downscaling (*green*), *QM* empirical quantile mapping (*red*). The acronyms for the climate model combinations are given in the supplement to this paper (S2)

large spread. A *T*-driven change in the pattern of AMFs emerges for *T* below 6.0 °C. Here, the mean magnitude of the AMF tends to rise with constant *P*. This pattern is reflected by most of the individual SRRSs (S7) and might be explained by an increased storage of snow during winter due to lower *T* which potentially leads to higher runoff during the snowmelt season.

The SRRS for the AMF is much more affected by the different temporal structures than those for mean discharge as illustrated by the high CV values (10.0–17.2%; and thus, 7–11 percentage points larger than for MQ; Fig. 1d). Note the location of several projections (red and green 6–9) in areas of the response surface which show high CV values indicating comparable high uncertainties in the simulations of high flows. Generally, the differences between the patterns of the individual SRRSs for the mean magnitude of the AMFs (S7) are much more distinctive as compared to the patterns of MQ (S6).

3.2.3 The seasonality index S_D

In contrast to the variables before, the seasonal occurrence of floods (S_D) and their response to a changing climate is mainly *T*-driven (Fig. 1e). With increasing *T*, the mean S_D values increase up to 0.61 indicating dominant flood events during autumn and winter, while decreasing *T* reduces the mean S_D values down to −0.64 which indicates dominant events during spring and summer. The S_D value for the current climate is slightly below zero and reflects the mixed flood seasonality as outlined in Section 2.1. This is likely to change towards more dominant pluvial floods during autumn and winter. Scaling *T* reveals the changing role of snow storage and snowmelt for the generation of floods. Scaling *P* seems to have only a minor influence on S_D.

The disagreement (on median 32%) between the individual SRRSs (Fig. 1f) displays the high influence of the underlying temporal structures and is also controlled by the scaling of *T*. The lowest standard deviations (21%) are found for the impacts of *T* larger than 7.5 °C. The largest standard deviations (up to 46%) appear for *T* between 4.0 and 5.0 °C where S_D is simulated to be slightly below zero. Below 3.5 °C and above 5.5 °C, the standard deviations are decreasing again. The differences between the single SRRSs generated with the time series of XDS- and QM-adjusted climate projection data are comparatively large (S8), which indicates the crucial role of the methods used to locally adjust large-scale RCM data to the catchment scale.

3.2.4 The low flow index LF7

Changes in LF7 over the entire ensemble are influenced by scaling both *T* and *P*. This leads to the pronounced diagonal pattern of changes in LF7 as illustrated in Fig. 1g. As expected, the highest mean LF7 value (3.6 $m^3 s^{-1}$) emerges for low *T* and high *P*; the lowest mean LF7 value (0.3 $m^3 s^{-1}$) emerges for the opposite combination of *T* and *P*.

Compared to the three target variables before, the influence of the temporal structure is highest for LF7 as indicated by the high CV values (median 37%). The highest CV values (45%) are found for the lowest *T* and *P* combinations, while the lowest values (34%) are found for high *T* and *P* combinations (Fig. 1h). There are only small differences between the SRRSs generated by the XDS-adjusted and QM-adjusted time series. However, some remarkable differences occur among the climate models irrespective of the post-processing method applied (S8): the SRRSs generated by the post-processed ECHAM5_REMO data show comparable patterns to the SRRS generated by the observed input data. The SRRSs generated by post-

processed BCA_HIRHAM, BCA_RCA, and ECHAM_REGCM data show the highest LF7 values for all combinations of linear scaling.

3.3 The influence of perturbation and temporal structures on the SRRSs

So far, it has become obvious that differing temporal structures influence the long-term means of the estimated hydrological target variables displayed by SRRSs. In this section, we present the quantified effect of the different temporal structures *V* and its specifications *G* and *DS* compared to the influence of perturbing *T* and *P* on the different patterns of the SRRSs as it has been quantified by the LME models (Fig. 2 and Table 1).

In general, the effects of *T*, *P*, and *V* (comprises *DS* and *G*) strongly vary with the hydrological target variables. The influence of scaling *P* dominates for MQ, AMF, and LF7 (94, 69, and 50% of the explained variance), while it is negligible for S_D (2%). This result makes sense since major liquid water storages are missing in the catchment so that increasing/decreasing *P* results in increasing/decreasing runoff. Furthermore, Fig. 2 indicates that the flood seasonality is mainly controlled by scaling *T* (52% of the explained variance) since *T* determines how much water is stored as snow during winter in this Nordic catchment which is remobilized during the snowmelt season. A large influence of scaling *T* is also found for LF7 (18% of explained variance) since high *T* strongly controls low flows during the snow-free months of the year. Note also the small contribution of *T* to the total variance regarding MQ and AMF that refers to snowmelt during spring and to evapotranspiration which particularly

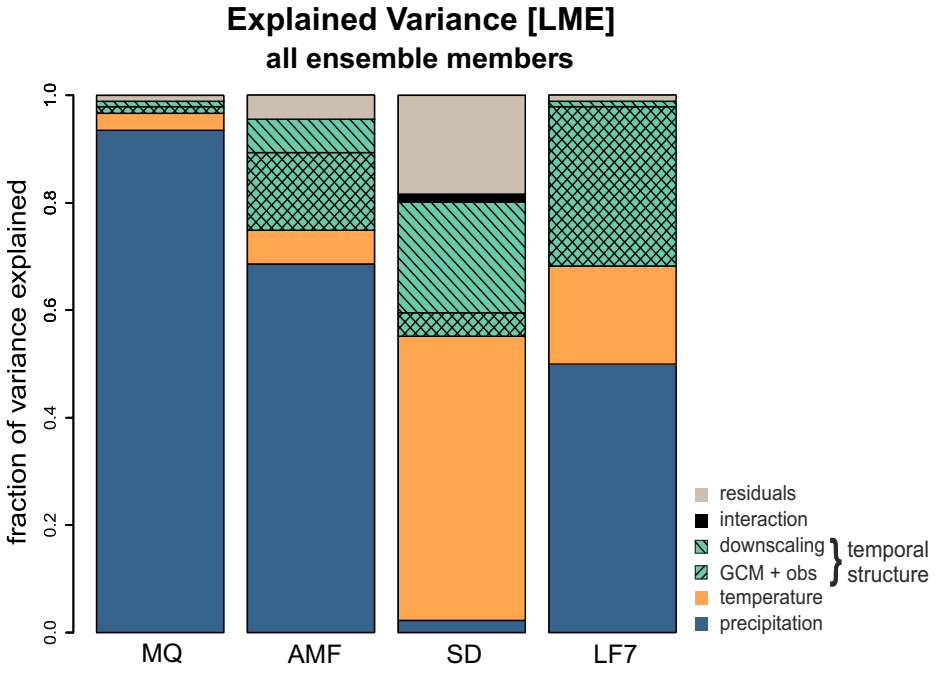

Fig. 2 Fraction of explained variance of the perturbed variables (temperature and precipitation) and the different temporal structures (by its specifications GCMs + observations and the statistical downscaling) to the total variance as described by the linear mixed-effect models for the four runoff target variables mean discharge (*MQ*), mean annual maximum flood (*AMF*), the flood seasonality index S_D, and the low flow index LF7. Proportion missing to 1 is unexplained variance (residuals)

affects MQ during summer. Still, the overall effect of scaling *T* on the results is lowest for all variables except for S_D.

Being negligible for MQ, the influence of *V* has a strong effect on all other target variables (21% for AMFs, 25% for S_D, and 30% for LF7). Although *V* does not show the largest contribution to the explained variance for any of the target variables, it strongly modifies all results except those for MQ. A closer look at the parameters *G* and *DS* reveals that for MQ, AMF, and LF7, the GCMs can already explain the majority of variance assigned to the temporal structure. *DS* comprises all information covered by the GCM and is able to explain an additional fraction of explained variance. This is expected since *DS* is a specification of temporal structures predetermined by the GCM. In turn, the target variable S_D is influenced merely by the downscaling method, indicating an excessive modifying effect of the methods applied.

The analysis underlines the critical role of the temporal sequencing of the climate data for the simulation of hydrological extremes. Most likely, this is because the temporal structure alters floods and low flows by extending or shortening the wet and dry periods on the day-to-day scale. On a seasonal scale, *DS* in combination with scaling *T* mainly explains the overall variance of the SRRSs for S_D. However, in contrast to the well-explained target variables MQ, AMF, and LF7, the large fraction of unexplained variance suggests that the LME models are not as good in representing S_D which may indicate a missing influencing factor.

4 Discussion and conclusions

The utility of scenario-neutral response surface approaches has lately been proven by different studies and for different target variables like flood magnitude (Prudhomme et al. 2010; Bastola et al. 2011; Wilby et al. 2014) and lake levels (Wetterhall et al. 2011). One major limitation of response surfaces is the underrepresentation of changes in the temporal structure of the climate input data sets used in the impact models. In this study, we have examined this limitation by scaling both observed *and* projected climate data time series, which both serve as input into a hydrological model. The resulting multiple response surfaces are compared at a common domain, the *standardized* runoff response surface (SRRS). This way, we were able (i) to study possible changes in mean and extreme runoff response due to climate change in a Norwegian case study catchment and (ii) to quantify the influence of changes in the temporal structure of the climate input data used within our bottom-up approach.

Results for this Nordic catchment suggest higher mean and peak flow discharge under future climate condition but only small changes for low flows. The results regarding increasing flood magnitudes and shifts towards more pronounced autumn and winter floods correspond to results which have been achieved by top-down approaches in Norway (Lawrence and Hisdal 2011; Vormoor et al. 2015). We generally agree that response surfaces provide a valuable option for the assessment of climate change impacts since they present an illustrative overview of many possible combinations of impacts in one plot. Furthermore, climate model-related uncertainties can be illustrated very intuitively since the spread of multiple GCMs/RCMs can be integrated into the response surface (as shown by, e.g., Prudhomme et al. 2010) and they can easily be updated with future generations of climate models.

However, our results also indicate the crucial role of perturbation and changes in the temporal structure of climate time series in the bottom-up approach. We have shown that changes in the temporal structure of the climate input data are of little importance for mean runoff but have a strong effect on the estimation of changes in peak flows and low flows (at

least 21% of the explained variability in this case study). This result confirms that modifications in the day-to-day as well as seasonal variability of the climate parameters will, most likely, highly influence the impacts of future climate conditions on hydrological extremes (Gobiet et al. 2014). This is in line with Steinschneider et al. (2015) who have found that the effects of internal climate variability on flood damage are at the same order of magnitude as the effects of uncertainty in future precipitation projections. It further points to a central aspect of the limitation associated with the generation of response surfaces in general, as stated by Prudhomme et al. (2010) in their initializing paper: the sensitivity of response surfaces towards the perturbation technique applied. Since it is hard to visualize multiple treatments at a two-dimensional domain (Wilby et al. 2014), response surfaces have focused so far on perturbing observed climate data in a two-dimensional way, only. Here, we have presented a framework that allows for comparing multiple response surfaces at a common domain. Although the computational costs are a multiple larger than for regular response surfaces, the aggregation of the full range of information to averages and coefficients of variations is a promising way to include more dimensions into the response surface approach without being too complex.

In conclusion, bottom-up approaches dealing with extreme values need to be aware that a substantial part of variability is missing when disregarding likely changes in the temporal structure of the climate data. At the same time, regular response surfaces prove to be a valuable tool for systematically analyzing climate change-driven impacts on mean values. This is in line with the discussion on the suitability of change-factor-based statistical downscaling methods for extreme value projections under climate change (Graham et al. 2007; Räisänen and Räty 2013; Räty et al. 2014; Sunyer et al. 2015). Delta change or simple perturbation techniques are appropriate for mean values, but more sophisticated methods are required for the simulation of extremes. The framework of SRRSs might be a good starting point to include such issues in the bottom-up approach.

Note, however, that our conclusions are based on the presented case study with 17 climate realizations in one single catchment in Norway applying the HBV model. Further studies using wider climate realizations and other study areas need to elaborate the effect of temporal structures on response surfaces in more detail. Moreover, we need to emphasize that other sources of variability/uncertainty, as for example hydrological model structure and parameter uncertainty, have not yet been analyzed within the framework of bottom-up approaches. This can also be a future application of the framework presented here.

Acknowledgements The first author acknowledges the Helmholtz graduate school GeoSim for funding the PhD studentship. The Norwegian Water Resources and Energy Directorate (NVE) is thanked for providing the hydrological model code and the runoff and climate observation data. The regional climate model simulations stem from the EU FP6 project ENSEMBLES, whose support is gratefully acknowledged. All statistical analyses and visualizations (except S1) were performed within the R software environment (www-r-project.org). Critical comments by three anonymous reviewers on an earlier version of this manuscript helped to improve this paper.

References

Bastola S, Murphy C, Sweeney J (2011) The sensitivity of fluvial flood risk in Irish catchments to the range of IPCC AR4 climate change scenarios. Sci Total Environ 409(24):5403–5415

Bergström S (1995) The HBV model. In: Singh VP (ed) Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, Colo., pp 443–476

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(12):1643–1655

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends in ecology and evolution 24:3,127–3,135
Bürger G (1996) Expanded downscaling for generating local weather scenarios. Clim. Res. (7):111–128
Bürger G, Reusser D, Kneis D (2009) Early flood warnings from empirical (expanded) downscaling of the full ECMWF Ensemble Prediction System. Water Resour Res 45(10):W10443
Doherty J (2004) PEST: Model independent parameter estimation. Fifth edition of user manual. Watermark Numerical Computing, Brisbane
Fleig A, Andreassen LM, Barfod E, Haga J, Haugen LE, Hisdal H, Melvold K (2013) Norwegian Hydrological Reference Dataset for climate change studies. NVE raport, 2, Oslo, Norway
Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol 27(12):1547–1578
Fronzek S, Carter TR, Luoto M (2011) Evaluating sources of uncertainty in modelling the impact of probabilistic climate change on sub-arctic palsa mires. Nat Hazards Earth Syst Sci 11(11):2981–2995
Gobiet A, Kotlarski S, Beniston M, Heinrich G, Rajczak J, Stoffel M (2014) 21st century climate change in the European Alps—a review. Sci Total Environ 493:1138–1151
Graham LP, 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 Chang 81(S1):293–307
Gudmundsson L, Bremnes JB, Haugen JE, 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(9):3383–3390
Hirschi M, Stoeckli S, Dubrovsky M, Spirig C, Calanca P, Rotach MW, Fischer AM, Duffy B, Samietz J (2012) Downscaling climate change scenarios for apple pest and disease modeling in Switzerland. Earth Syst Dynam 3(1):33–47
IPCC (2000) Emissions scenarios. IPCC special report, Edinburgh
Lawrence D, Hisdal H (2011) Hydrological projections for floods in Norway under a future climate. NVE raport, 5
Lawrence D, Haddeland I, Langsholt E (2009) Calibration of HBV hydrological models using PEST parameter estimation. Report 1/2009. Norwegian Water Resources and Energy Administration (NVE), Oslo
van der Linden P, Mitchell JFB (2009) ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project. FritzRoz Road, Exeter, UK
Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, 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(3)
Mohr M, Tveito OE (2008) Daily temperature and precipitation maps with 1 km resolution derived from Norwegian weather observations. 17th Conference on Applied Climatology, Whistler, BC, Canada, Amer Meteor Soc, 6.3. Available online at https://ams.confex.com/ams/pdfpapers/141069.pdf
Pinheiro JC, Bates DM (2004) Mixed-effects models in S and S-PLUS. Statistics and computing. Springer, New York
Prudhomme C, Wilby RL, Crooks S, Kay AL, Reynard NS (2010) Scenario-neutral approach to climate change impact studies: application to flood risk. J Hydrol 390(3–4):198–209
Räisänen J, Räty O (2013) Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations. Clim Dyn 41(5–6):1553–1568
Räty O, Räisänen J, Ylhäisi JS (2014) Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations. Clim Dyn 42(9–10):2287–2303
Saelthun NR (1996) The 'Nordic' HBV Model.: description and documentation of the model version developed for the project climate change and energy production. Norges Vassdrags-og Energiverk Report, 7
Steinschneider S, Wi S, Brown C (2015) The integrated effects of climate and hydrologic uncertainty on future flood risk assessments. Hydrol Process 29(12):2823–2839
Sunyer MA, Hundecha Y, Lawrence D, Madsen H, Willems P, Martinkova M, Vormoor K, Bürger G, Hanel M, Kriaučiūnienė J, Loukas A, Osuch M, Yücel I (2015) Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. Hydrol Earth Syst Sci 19(4):1827–1847
Teutschbein C, Seibert J (2010) Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies. Geography Compass 4(7):834–860
Vormoor K, Lawrence D, Heistermann M, Bronstert A (2015) Climate change impacts on the seasonality and generation processes of floods—projections and uncertainties for catchments with mixed snowmelt/rainfall regimes. Hydrol Earth Syst Sci 19(2):913–931
Wehren B (2010) Die Hydrologie der Kander - gestern, heute, morgen: Analyse und Modellierung der Hochwasser und ihrer raumzeitlichen Dynamik. Dissertationsschrift, Bern