Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections

T. Bosshard,¹ M. Carambia,² K. Goergen,³ S. Kotlarski,¹ P. Krahe,² M. Zappa,⁴ and C. Schär¹

Received 19 October 2011; revised 9 December 2012; accepted 10 December 2012.

[1] The quantification of uncertainties in projections of climate impacts on river streamflow is highly important for climate adaptation purposes. In this study, we present a methodology to separate uncertainties arising from the climate model (CM), the statistical postprocessing (PP) scheme, and the hydrological model (HM). We analyzed ensemble projections of hydrological changes in the Alpine Rhine (Eastern Switzerland) for the near-term and farterm scenario periods 2024–2050 and 2073–2099 with respect to 1964–1990. For the latter scenario period, the model ensemble projects a decrease of daily mean runoff in summer (-32.2%, range [-45.5% to -8.1%]) and an increase in winter (+41.8%, range [+4.8% to -8.1%])+81.7%]). We applied an analysis of variance model combined with a subsampling procedure to assess the importance of different uncertainty sources. The CMs generally are the dominant source in summer and autumn, whereas, in winter and spring, the uncertainties due to the HMs and the statistical PP gain importance and even partly dominate. In addition, results show that the individual uncertainties from the three components are not additive. Rather, the associated interactions among the CM, the statistical PP scheme, and the HM account for about 5%–40% of the total ensemble uncertainty. The results indicate, in distinction to some previous studies, that none of the investigated uncertainty sources are negligible, and some of the uncertainty is not attributable to individual modeling chain components but rather depends upon interactions.

Citation: Bosshard, T., M. Carambia, K. Goergen, S. Kotlarski, P. Krahe, M. Zappa, and C. Schär (2013), Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, *Water Resour. Res.*, 49, doi:10.1029/2011WR011533.

1. Introduction

[2] The projected impacts of climate change on river streamflow are associated with large uncertainties. For water management, these projection uncertainties importantly contribute to the total uncertainty, in addition to factors such as natural variability and changes in water demand [Kundzewicz et al., 2008]. Although the use of climate-impact projections for water management planning purposes has been debated [Kundzewicz and Stakhiv, 2010], most published strategies actually make use of such information [Dessai and Hulme, 2007; Milly et al., 2008]. The quantification of uncertainties in climate-impact projections is therefore of particular interest [Pappenberger and Beven, 2006]. So far, the overall uncertainty of projected climate impacts is probably underestimated, which is due to an incomplete

sampling of the uncertainty sources [*Knutti*, 2008; *Wilby*, 2010]. Improving the knowledge about the importance of different uncertainty sources might thus help to design climate-impact studies with a more complete uncertainty assessment.

[3] Impact modeling systems that include a cascade of different models are commonly used to assess climate impacts (for some recent studies, see, e.g., Vicuna et al. [2010], Campbell et al. [2011], Quintana-Seguí et al. [2011], Köplin et al. [2012], and Teutschbein and Seibert [2012]) and to provide information for water management [e.g., Schaefli et al., 2007; Lopez et al., 2009]. Elements of this cascade are an emission scenario, a global circulation model (GCM), a dynamical downscaling step by means of a regional climate model (RCM), a statistical postprocessing (PP), and a hydrological model (HM). Alternatively, the dynamical downscaling and the PP steps can be replaced by a statistical downscaling. In the remainder of this article, we call this cascade of emission scenarios and models an impact modeling chain. Uncertainties in hydrological climate-impact projections arise due to different assumptions and model combinations in the whole impact modeling chain (e.g., CM or HM structure uncertainty [Masson and Knutti, 2011; Seiller et al., 2012]), CM or HM parameter uncertainty [Bellprat et al., 2012; Beven, 2006], instationarity of PP parameters [Buser et al., 2009; Boberg and Christensen, 2012], and natural variability [see, e.g., Lucas-Picher et al., 2008]. For a complete analysis of uncertainty in runoff projections, it is therefore important to investigate the contributions of all existing sources.

¹Institute for Atmospheric and Climate Science, ETH Zurich, Zürich, Switzerland.

²German Federal Institute of Hydrology, Koblenz, Germany.

³Département Environnement et Agro-Biotechnologies (EVA), Centre de Recherche Publi - Gabriel Lippmann, Belvaux, Luxembourg.

⁴Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland.

Corresponding author: T. Bosshard, Hydrological Research Unit, Swedish Meteorological and Hydrological Institute, Folkborgsvägen 1, SE-60176 Norrköping, Sweden. (thomas.bosshard@smhi.se)

^{©2012.} American Geophysical Union. All Rights Reserved. 0043-1397/13/2011WR011533

[4] Numerous previous studies have investigated hydrological climate-impact projections and their sensitivity to different uncertainty sources. Here, a nonexhaustive summary of studies conducted in the Rhine basin follows. Shabalova et al. [2003] compared two PP methods and found that for the end of the 21st century, both methods agree on a decrease of summer runoff and an increase of winter runoff, but the two methods lead to different increases of winter flood risk. Lenderink et al. [2007] also investigated uncertainties due to PP and confirmed the results of Shabalova et al. [2003] that the choice of the PP rather affects the changes in runoff extremes than in the mean runoff. Jasper et al. [2004] investigated the projected impact on runoff of an ensemble of 17 climate scenarios derived from seven GCMs and four emission scenarios in two Swiss catchments. They found that changes in the seasonality of runoff are robust, but the magnitude of the changes is strongly affected by the choice of the climate scenario. Combined uncertainties from emission scenarios, GCMs and RCMs, were investigated by Graham et al. [2007] who found that the choice of the GCM has a larger impact on projected hydrological changes than the choice of the RCM or emission scenario.

[5] More recently, studies that systematically investigate multiple uncertainty sources along the whole impact modeling chain have been published. Wilby and Harris [2006] assessed uncertainties from emission scenarios, GCMs, statistical downscaling, HM structure, and HM parameters. Using a probabilistic framework, they showed that GCMs and the downscaling step were the most important sources of uncertainty in simulating changes of low flows in the Thames River (UK). In a study of the impact of climate change on hydropower production in the Mauvoisin catchment (Switzerland), Schaefli et al. [2007] investigated the importance of uncertainties due to the global mean temperature projection, the regional scaling relationship, the glacier model, HM parameters, and the hydropower management model. They found the uncertainty in the global mean temperature and the regional scaling factors to be of comparable magnitude and both being more important than the other uncertainty sources. Prudhomme and Davies [2009] assessed uncertainties due to emission scenarios, GCMs, downscaling methods, and HMs in four mesoscale British catchments and concluded that the driving GCM is the dominant source of uncertainty. They further stated that uncertainty due to PP and the choice of the emission scenario are of comparable magnitude, whereas uncertainty due to the HMs is negligible in two out of four basins. Kay et al. [2009] investigated the same uncertainty sources as Wilby and Harris [2006] and also included the effect of internal variability in a case study that assessed changes of flood frequency in two British catchments. They found GCMs being the dominant source of uncertainty. However, after excluding one outlier from the GCM ensemble, other uncertainty sources such as RCMs and internal variability became more important than GCMs. For two catchments in Oregon (USA), Jung et al. [2011] found the natural variability and the driving GCM to be the major sources for uncertainty with respect to flood frequency changes.

[6] In our study, we perform an ensemble of hydrological climate-impact projections for an Alpine river catchment and the two scenario periods (SCE) 2024–2050 and 2073–2099 with respect to the control period (CTL) 1964– 1990. From this ensemble, we aim to infer several sources of uncertainty, and we use the variance as a measure for the uncertainty. The three uncertainty sources considered are (i) climate models (CMs) consisting of a GCM and an RCM, (ii) PP, and (iii) HMs. The Alpine study area is a challenging region for all three impact modeling chain elements. CMs, both GCMs and RCMs, for instance, cannot fully resolve the complex topography; PP methods have to correct for potentially large biases; and HMs are challenged by complicated and spatially highly variable hydrological processes such as accumulation and melt of snow.

[7] The limited number of uncertainty sources and models included in the ensemble results in an underestimation of the overall uncertainty associated with the hydrological climate impacts (i.e., the uncertainty if all possible uncertainty sources were fully sampled). Throughout this paper, we call the spread in our ensemble the *total ensemble uncertainty* to clearly distinguish it from the overall (true) uncertainty.

[8] Our two research questions are as follows: (1) how large is the total ensemble uncertainty in the runoff projections and (2) how do different uncertainty sources contribute to the total ensemble uncertainty in the runoff projections for both SCEs. In particular, we are interested in how the contributions vary throughout the annual cycle and how they affect the uncertainty in changes of different runoff quantiles. We include fewer sources of uncertainty than some of the previously mentioned studies, but, instead of performing single-propagation runs, we conduct a multipropagation study; that is, we vary the different CMs, PP methods, and HMs in all possible combinations. This approach allows for an assessment of interactions between the uncertainty sources [Kay et al., 2009; Finger et al., 2012]. We quantify the contributions of the different uncertainty sources using the decomposition of the sum of squares as described within the analysis of variance (ANOVA) theory (see Déqué et al. [2007] or Yip et al. [2011] for a detailed description). Here, we refer to this method as the ANOVA approach. Following the idea of a multipropagation experiment, the ANOVA approach allows to consider interactions between the uncertainty sources. These interactions represent uncertainty contributions that do not behave linearly. For instance, a snowmelt bias of an HM may depend upon the temperature projection of the driving CM that could lead to a nonlinear response in river runoff. The ANOVA is complemented with a subsampling scheme to account for the different sample sizes of the three uncertainty sources.

[9] This paper is structured as follows. Section 2 briefly describes the study area and the data. Section 3 introduces the employed HMs, explains the PP methods, and presents the subsampling procedure in combination with the ANOVA. In section 4, we present the results of the hydrological climate-impact projections and the ANOVA. Section 5 summarizes our study and its main findings.

2. Study Region and Data

2.1. Study Region

[10] The study region consists of the Alpine Rhine catchment down to the gauge Diepoldsau in Eastern Switzerland (see Figure 1). It encompasses an area of 6119 km² and has



Figure 1. Map showing the catchment of the Alpine Rhine river down to Diepoldsau. The subbasin structures used in (left) HBV and (right) PREVAH are shown.

a mean elevation of about 1800 m above sea level. The runoff regime is nival, i.e., snow-dominated, but altered to some extent by hydropower production. The hydropower main effect is a seasonal redistribution of water from summer to winter [*Verbunt et al.*, 2005]. The buildup of the hydropower capacity in the period 1945–2009 is depicted in Figure 2 (solid black line). At the end of the period, the storage capacity amounts to about 10% of the annual runoff volume.



Figure 2. Validation of the two HMs HBV and PREVAH at the gauge Diepoldsau. Lines show the NSE (solid lines), 99% (dash-dotted lines), and 5% (dashed lines) runoff quantile levels (left-hand scale). The values are calculated for moving 3 year periods. The thick lines indicate the performance of the CTL run from 1961 to 1990 with the first 3 years being cutoff. The thin lines show the performance of the same model configurations in a longer simulation covering the period 1954–2006 (HBV) and 1951–2009 (PREVAH). The shaded areas depict the calibration period for each HM. The buildup of the total hydropower reservoir volume is shown as a solid thick black line (right-hand scale).

Table 1. List of Key Characteristics of the Two Employed F

	HBV	PREVAH	
Model type	Conceptual, semidistributed	Conceptual, semidistributed	
Meteorological input data	Precipitation (P), temperature (T), and global radiation (R) or sunshine duration (S)	Precipitation (P), temperature (T), sunshine duration (S), cloud cover (C), relative humidity (H), wind speed (V)	
Number of subbasins	2	6	
Spatial resolution of the underlying digital elevation model	$1000 \text{ m} \times 1000 \text{ m}$	$500 \text{ m} \times 500 \text{ m}$	
Internal time step	1 day	1 h	
Land use classes	4	29	
HRU definition	Elevation, land use	Elevation, land use, aspect, soil type	
Snow/glacier melt modeling approach	Degree-day factor	Degree-day factor with aspect and slope correction	
Evapotranspiration parameterization	Penman-Wendling	Penman-Monteith	
Calibration period	1970–1984	1985–1990	
References	Eberle et al. [2005], Lindström et al. [1997], Görgen et al. [2010]	Gurtz et al. [1999], Zappa and Gurtz [2003], Viviroli et al. [2009]	

2.2. Observational and CM Data

[11] Throughout this study, we use the CTL and the two SCEs, all three periods with an additional preceding 3 year spin-up period, as temporal subsets of the data series.

[12] The two HMs HBV and PREVAH (see section 3.1) require different kinds of input data with respect to the spatial resolution and the variables (see Table 1). HBV uses subbasin-averaged daily time series of precipitation, temperature, and global radiation or sunshine duration. Subbasin-averaged observational data have been provided by the International Commission for the Hydrology of the Rhine Basin (CHR; referred to as OBS_{CHR} in the remainder of this article) [*Görgen et al.* 2010]. PREVAH uses daily station data of the hydrometeorological variables precipitation (80 stations), temperature (36 stations), relative humidity (41 stations), sunshine duration (28 stations), and wind speed (43 stations) from the measurement network of MeteoSwiss (referred to as OBS_{ST} in the remainder of this article). Both observational data sets cover the whole CTL.

[13] For the calibration of the HMs, we used daily runoff data of the gauges depicted in Figure 1. The data were provided by the Swiss Federal Office for the Environment (see www.hydrodaten.admin.ch).

[14] As climate data, we used eight transient climate modeling chains of the ENSEMBLES project [van der Linden and Mitchell, 2009], as shown in Figure 3. Déqué et al. [2012] showed that, within the ENSEMBLES GCM-RCMs, the GCM is the largest contributor to the variance in the projections of seasonal mean temperature and precipitation, except for summer precipitation for which the RCMs are the largest source of variance. Thus, it is important to sample both the GCMs and the RCMs in an impact modeling chain. The ensemble used in this study encompasses three GCMs and seven RCMs, which allows to partly sample CM uncertainty.

[15] All modeling chains are driven by the A1B emission scenario [*Nakicenovic and Swart*, 2000]. This emission scenario belongs to the A1 emission scenario family that assumes rapid economic growth, population growth until mid-century, a homogenization of the global wealth across the different regions, and a rapid introduction of new efficient technologies. The B stands for a balanced (fossil and nonfossil) use of energy sources. The corresponding greenhouse gas emissions increase until 2060 and slightly decrease afterward, resulting in an atmospheric CO_2 concentration of about 700 ppm by 2100. The spatial resolution of the RCMs is about 25 km. We used subbasin-averaged CM time series (for HBV) and CM data interpolated to station locations (for PREVAH).

3. Methods

[16] Figure 3 depicts the modeling chain combination scheme employed in this study. In the following, we describe the HMs and the PP methods and explain the ANOVA approach in combination with a subsampling scheme.

3.1. Hydrological Models

[17] In our study, we use HMs that have already been set up in the catchment. Both the HBV and the PREVAH model are semidistributed conceptual rainfall-runoff models. Both models use the hydrological response unit (HRU) approach to cluster the spatial units according to their hydrological characteristics. Table 1 summarizes the characteristics of the two HMs in the Alpine Rhine catchment,



Figure 3. Modeling chain combination scheme. The three analyzed modeling chain elements are depicted from left to right. The naming of the CM chains provided by the ENSEMBLES project is based on the following pattern: institution that did run the RCM, employed GCM and RCM [*van der Linden and Mitchell*, 2009].

and Figure 1 shows the subbasin structures of the two HMs. For the HBV model, we use the HBV134 setup of the German Federal Institute of Hydrology. The HBV134 requires subbasin-averaged hydrometeorological time series as input data and applies a lapse rate correction to disaggregate basin-averaged temperature to zones of different elevations. For the PREVAH model, we use the setup by *Verbunt et al.* [2006], which we have recalibrated in the period 1985–1990 to use sunshine duration instead of global radiation as an input variable from which shortwave radiation is derived according to *Schulla* [1997]. PREVAH requires hydrometeorological station data as an input. The station data are interpolated using inverse distance weighting. For temperature, a height-dependent regression is applied (detrended inverse distance weighting).

[18] Both HMs correct the observed precipitation to account for undercatch and interpolation errors [*Sevruk and Nevenic*, 1998]. In the HBV model, a linear precipitation correction factor is applied to precipitation, and no distinction is made between rain and snow. In the PREVAH model, two correction factors for rain and snow are used. The precipitation correction factors are estimated by calibration using simulations driven by observed data (OBS_{CHR} or OBS_{ST}). The water-balance-corrected precipitation is then used as a reference in the PP. This provides a consistent way to process the CM data without artificially modifying the precipitation change signal by a nonlinear precipitation correction (i.e., different corrections for snow and rain). Consistent with the above, all figures in the remainder of this article show water-balance-corrected precipitation.

[19] For further details about the HMs, we refer to the references listed in Table 1.

3.2. Statistical PP

[20] We use a bias-correction (BC) and delta-change (DC) approach for the PP step in the impact modeling chain. The parameters of the PP methods were estimated based on the full 30 year periods 1961–1990 (CTL), 2021–2050 (SCE1), and 2070–2099 (SCE2), including a 3 year spin-up period.

[21] The two PP methods differ distinctively regarding the treatment of changes in the variability. In the BC, the time series used to drive the HM are based on CM data, which are corrected toward the climatological mean of the observations in the CTL. Thus, the variability in the CM data determines the variability in the bias-corrected forcing data, e.g., the succession of wet and dry days. The DC approach scales observed time series according to a climate change signal estimated from CM data. The scaled observational time series are used to force the HMs in the SCE. It is therefore the variability in the observed time series that determines the variability in the HM's forcing data. Also, the two PP methods differ in the number of variables they include. The BC approach corrects all required variables from the output of the CMs, whereas the DC method is applied to temperature and precipitation only, and unscaled observed time series are used for the other variables.

3.2.1. Bias Correction

[22] For the BC of the CM data, we use a linear scaling as employed by *Lenderink et al.* [2007] and *Görgen et al.* [2010].

[23] Let X be a meteorological input variable for the HM. As the observational reference, we use subbasin-averaged

time series and call it X_{avg}^{obs} . For OBS_{ST} used by PREVAH, we first spatially interpolate the station data and derive subbasin-averaged time series. Next, for each CM *j*, we calculate a subbasin mean time series as area weighted grid cell averages and denote it with X_{avg}^j . We estimate the correction parameters $a^X(m)$ between the subbasin mean time series X_{avg}^{obs} and X_{avg}^j for each month *m* in the annual cycle. We use the following linear correction models for the various meteorological variables (see Table 1 for an explanation of the abbreviations):

$$P_{\text{avg}}^{*j} = a^P P_{\text{avg}}^j, \quad a^P(m) = \frac{\overline{P_{\text{avg}}^{obs}(m)}}{\overline{P_{\text{avg}}^j(m)}}$$
(1)

$$T_{\text{avg}}^{*j} = T_{\text{avg}}^j + a^T, \quad a^T(m) = \overline{T_{\text{avg}}^{\text{obs}}}(m) - \overline{T_{\text{avg}}^j}(m)$$
(2)

$$S_{\text{avg}}^{*j} = \min\left[S_0, a^S S_{\text{avg}}^j\right], \quad a^S(m) = \frac{\overline{S_{\text{avg}}^{\text{obs}}(m)}}{\overline{S_{\text{avg}}^j(m)}}$$
(3)

$$R_{\text{avg}}^{*j} = a^R R_{\text{avg}}^j, \quad a^R(m) = \frac{\overline{R_{\text{avg}}^{\text{obs}}(m)}}{\overline{R_{\text{avg}}^j(m)}}$$
(4)

$$H_{\text{avg}}^{*i} = \max\left[0, 100 - a^{H}\left(100 - H_{\text{avg}}^{i}\right)\right], a^{H}(m) = \frac{100 - \overline{H_{\text{avg}}^{\text{obs}}(m)}}{100 - \overline{H_{\text{avg}}^{i}(m)}}$$
(5)

$$V_{\text{avg}}^{*j} = a^V V_{\text{avg}}^j, \quad a^V(m) = \frac{\overline{V_{\text{avg}}^{\text{bos}}(m)}}{\overline{V_{\text{avg}}^{j}(m)}},\tag{6}$$

where the overbar denotes climatological monthly means in the CTL, and the superscript * stands for the bias-corrected subbasin-averaged daily time series. The equations in the left column show the linear BC model, whereas the equations in the right column explain how the correction parameter is derived. Note that, although not explicitly indicated in the left column, the correction factor a^X varies according to the month. As indicated in Table 1, each HM uses only a selection of the meteorological variables listed above.

[24] Some CMs do not provide sunshine duration. Thus, we use cloud fraction (C) as a proxy and derive S according to

$$S_{\rm avg}^{j} = S_0 \left(1 - C_{\rm avg}^{j} \right) \tag{7}$$

with S_0 representing the maximum possible sunshine duration.

[25] After the BC on the basin level, a further spatial disaggregation step is necessary for the PREVAH model to account for the finer spatial input structure. In the CTL, we derive for every month *m* and every meteorological subarea *i* a disaggregation relation based on the observations according to

$$r_i(m) = \overline{X_i^{\text{obs}}}(m) - \overline{X_{\text{avg}}^{\text{obs}}}(m)$$
 for temperature, and (8)

$$r_i(m) = \frac{X_i^{\text{obs}}(m)}{\overline{X_{\text{avg}}^{\text{obs}}(m)}} \qquad \text{for the other variables}, \tag{9}$$

and scale the bias-corrected basin mean time series like

$$X_i^{*j} = X_{\text{avg}}^{*j} + r_i$$
 for temperature and (10)

$$X_i^{*j} = r_i X_{avg}^{*j}$$
 for the other variables. (11)

3.2.2. Delta Change

[26] We use the DC method as described by *Bosshard* et al. [2011] and *CH2011* [2011], which includes a spectral smoothing of the time series to damp the influence of natural variability on estimates of the climate change signal. Due to the spectral smoothing in the DC approach, the annual cycle is resolved continuously as opposed to the BC method where monthly steps are used. The DC scaling is applied to temperature and precipitation data only. For HBV, we derive the climate change signal from the subbasin-averaged time series X_{avg}^{j} and scale OBS_{CHR} accordingly. For PREVAH, we first interpolate the CM data to the station locations using inverse distance weighting, derive the climate change signal at the station sites, and scale OBS_{ST} data accordingly.

3.3. Variance Decomposition

[27] In the variance decomposition, we aim to decompose the total ensemble uncertainty into contributions from different elements of the impact modeling chain and interactions among them. In our study, the total ensemble uncertainty is the variance of the climate change signal in the mean annual cycle of runoff Q_{AC} (as estimated by a 31 day moving average; 31d MA) and in different runoff quantile levels (Q_Q). The model ensemble consists of 32 impact modeling chain combinations (Figure 3). For each of the 32 impact modeling chain combinations, we estimate Q_{AC} and Q_Q in the CTL and SCE (i.e., SCE1 or SCE2). Then, we calculate the climate change signal Y as

$$Y_{\rm AC} = Q_{\rm AC}^{\rm SCE} - Q_{\rm AC}^{\rm CTL} \tag{12}$$

$$Y_Q = Q_Q^{\text{SCE}} - Q_Q^{\text{CTL}}.$$
 (13)

In the following, we neglect the subscripts referring to the mean annual cycle or the runoff quantiles and describe the methodology for the general variable *Y*. To relate the target variable *Y* to the uncertainty sources, we use superscripts in $Y^{j,k,l}$ with *j*, *k*, and *l* representing the different samples of CMs, PP methods, and HMs, respectively.

3.3.1. Subsampling of the CMs

[28] The ANOVA approach (see section 3.3.2) is based on a biased variance estimator that underestimates the variance in small sample sizes (see also *Déqué et al.* [2007] for a short discussion of the problem). To diminish the effect of the biased variance estimator on the quantification of the variance contribution (see section 3.3.2), we subsample the eight different CMs. In each subsampling iteration *i*, we select two CMs out of the eight that results in a total of 28 possible CM pairs. For each of the 28 subsampling iterations, we end up with two CMs, two PP methods, and two HMs that define our model combination matrix for the variance decomposition. To differ between the full set of eight CMs and the subsampled CM pair, we replace the superscript *j* with g(h,i). The superscript *g* is a 2 × 28 matrix that contains the selected CMs for the particular subsampling iteration *i*:

$$g = \begin{pmatrix} 1 & 1 & \cdots & 1 & 2 & 2 & \cdots & 6 & 6 & 7 \\ 2 & 3 & \cdots & 8 & 3 & 4 & \cdots & 7 & 8 & 8 \end{pmatrix}.$$
 (14)

3.3.2. ANOVA Approach

[29] We use the statistical theory of the ANOVA to partition the variance (see, e.g., von Storch and Zwiers [1999] for an introduction and Déqué et al. [2007] and Yip et al. [2011] for an application in climate modeling). In our experimental setup, there is one data point Y for every possible combination of the three modeling chain elements CM, PP, and HM. In the terminology of the ANOVA, the modeling chain elements are "effects." An effect is a variable that one suspects could have an influence on the variability of the variable Y. In other words, we construct an ANOVA model based on the hypothesis that the CM, PP, and HM elements of the modeling chain have an influence on the variability of the variable Y, and we want to quantify the influence. We define the model as follows:

$$Y^{j,k,l} - Y^{\circ,\circ,\circ} = a^{j} + b^{k} + c^{l} + ab^{j,k} + ac^{j,l} + bc^{k,l} + abc^{j,k,l}, \quad (15)$$

where *a*, *b*, and *c* are the effects corresponding to CM, PP, and HM, respectively; the expressions *ab*, *ac*, *bc*, and *abc* are the interaction terms; and *j*, *k*, and *l* indicate samples of the different effects. The symbol $^{\circ}$ indicates averaging over the particular index. According to the ANOVA theory, the model allows us to split the total sum of the squares (SST) into sums of squares due to the individual effects (SSA, SSB, SSC) and their interactions (SSAB, SSAC, SSBC, SSABC) as

$$SST = SSA + SSB + SSC + SSI, \qquad (16)$$

$$SSI = SSAB + SSAC + SSBC + SSABC.$$
(17)

In this model, we summarize all interaction terms into the term SSI.

[30] We estimate the terms in equations (16) and (17) using the subsampling procedure introduced in section 3.3.1 as follows:

$$SST_{i} = \sum_{h=1}^{H} \sum_{k=1}^{K} \sum_{l=1}^{L} \left(Y^{g(h,i),k,l} - Y^{g(\circ \cdot i),\circ,\circ} \right)^{2},$$
(18)

$$SSA_{i} = K \cdot L \cdot \sum_{h=1}^{H} \left(Y^{g(h,i) \cdot \circ, \circ} - Y^{g(\circ,i) \cdot \circ, \circ} \right)^{2},$$
(19)

$$SSB_{i} = H \cdot L \cdot \sum_{k=1}^{K} \left(Y^{g(\circ,i)\cdot k\cdot\circ} - Y^{g(\circ,j)\cdot\circ,\circ} \right)^{2},$$
(20)

$$SSC_{i} = H \cdot K \cdot \sum_{l \neq 1}^{L} \left(Y^{g^{(\circ,i),\circ,l}} - Y^{g^{(\circ,i),\circ,\circ}} \right)^{2}, \quad \text{and}$$
(21)

$$SSI_{i} = \sum_{h=1}^{H} \sum_{k=1}^{K} \sum_{l=1}^{L} \left(Y^{g(h,i),k,l} - Y^{g(h,i),\circ,\circ} - Y^{g(\circ,i),k,\circ} - Y^{g(\circ,i),\circ,l} + 2Y^{g(\circ,i),\circ,\circ} \right)^{2}.$$
(22)

Then, for each effect, the variance fraction η^2 is derived as follows:

$$\eta_{\rm CM}^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{\rm SSA_i}{\rm SST_i},\tag{23}$$

$$\eta_{\rm PP}^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{\rm SSB}{\rm SST}_i,$$
(24)

$$\eta_{\rm HM}^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{\rm SSC}{\rm SST}_i, \text{ and }$$
(25)

$$\eta_{\text{Interactions}}^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{\text{SSI}_i}{\text{SST}_i}.$$
 (26)

Values of 0 and 1 for the variance fraction η^2 correspond to a contribution of an effect to the total ensemble variance (uncertainty) of 0% and 100%, respectively.

3.3.3. Effect of the Subsampling on the Estimated Biased Variance Contribution

[31] Without subsampling of the CMs, the estimated variance contributions using the ANOVA method would be biased, depending on the sample sizes of the different variance sources. For a random variable X of which N samples are drawn, the unbiased (σ_{UB}^2) and biased variance estimators (σ_{R}^2) are

$$\sigma_{\text{UB}}^2 = \frac{1}{N-1} \sum_{i}^{N} \left(X_i - \overline{X} \right)^2, \quad \text{and} \quad (27)$$

$$\sigma_B^2 = \frac{1}{N} \sum_{i}^{N} \left(X_i - \overline{X} \right)^2, \tag{28}$$

respectively. Thus, σ_B^2 underestimates the variance by a factor of

$$B\left(\frac{\sigma_B^2}{\sigma_{\rm UB}^2}\right) = \frac{N-1}{N},\tag{29}$$

where *B* denotes the expected bias of the argument. Thus, the bias $B\left(\frac{\sigma_B^2}{\sigma_{UB}^2}\right)$ depends on the sample size. The larger the sample size, the smaller the bias. In the ANOVA approach, we compare the variance of multiple variables using the biased variance estimator. Based on equation (29), the ratio *R* of two variances from two samples having sample sizes N_1 and N_2 is biased by the factor

$$B(R) = \frac{\frac{N_1 - 1}{N_1}}{\frac{N_2 - 1}{N_2}}.$$
(30)

For example, for the case of $N_1 = 8$ and $N_2 = 2$ representing the eight CMs and the two HMs, respectively, the bias in the variance ratio would be 1.75. Without subsampling, we would thus overestimate the uncertainty contribution of the CMs by a factor of 1.75. Figure 4 (left) shows B(R) for all combinations of two random variables having samples sizes between 2 and 10.

[32] We test the effect of the subsampling procedure on B(R) in a synthetic two-way ANOVA model without interactions and errors:

$$Z^{i,j} = a^{i} + b^{j}$$
(31)
$$a^{i} \sim \mathcal{N}(0, \sigma)$$

$$b^{j} \sim \mathcal{N}(0, \sigma).$$

By construction, the variances of the two effects *a* and *b* are equal. This simple model allows us to estimate the variance ratio *R* using unbiased and biased variance estimators. For different combinations of sample sizes for *a* and *b*, we constructed $Z^{i,j}$ according to equation (31) and subsequently estimated the variance ratio *R* using ANOVA and the unbiased variance estimator. We did this once without and once with subsampling (i.e., subsampling min(N_a, N_b) from the larger sample) and conducted 1000 bootstrap samples, each. Subsequently, we estimated B(R) as the mean of the 1000 bootstrap samples. The closer to one B(R) is, the smaller the bias.



Figure 4. Multiplicative bias of the variance ratio B(R) in the synthetic ANOVA experiment with two effects A and B represented by different sample sizes. A value of B(R) = 1 corresponds to no bias. (left) Expected B(R) based on the statistical theory (see equation (30)), (middle) estimated B(R) from 1000 bootstrap samples without using a subsampling approach, and (right) estimated B(R) from 1000 bootstrap samples with a subsampling scheme.

[33] Figure 4 shows the resulting B(R) from the experiment without subsampling (middle) and with subsampling (right). The results without subsampling match the expected bias from equation (30) perfectly, with negligible bias (i.e., B(R) = 1) on the diagonal, only. In contrast, the results with the subsampling procedure are much closer to B(R) = 1. Thus, the subsampling effectively diminishes the bias introduced by the biased variance estimator.

4. Results

4.1. Validation of the HMs

[34] Figure 2 shows the results of the validation of the HMs against observed runoff at the gauge Diepoldsau, using three validation metrics. Additional evaluation is presented in sections 4.2.3 and 4.2.4, where the results of the runoff projections for the CTL are discussed. The validation is done for the whole CTL that partly overlaps with the two different calibration periods of the HMs. In addition to the CTL simulations, we conducted two long simulations covering the period 1954–2006 (HBV) and 1951–2009 (PREVAH) with the same parameter setup of the HMs. These long simulations are meant to give some insight into the model performance over a longer period.

[35] Within the CTL, the Nash-Sutcliffe efficiency (NSE) [Nash and Sutcliffe, 1970] values of 3 year periods shifted in yearly intervals vary in the range of 0.81–0.91. With regard to the 99% runoff quantile values, the model HBV follows closely the observations. PREVAH tends to overestimate the high runoff values. The 5% runoff quantile values are simulated quite well by HBV, whereas PREVAH has a systematic underestimation. It should be noted though that, in the Alpine Rhine catchment, the low flows are influenced by hydropower operation [Margot et al., 1992] that leads to an increase of the low flows during winter. This effect is also shown in Figure 2 in the time series of the observed 5% runoff quantiles. They are lower in the early period, before the major part of the storage volume has been installed. Thus, the performance of the HMs in the low-flow range has to be interpreted with caution, as they both do not include a hydropower module.

[36] The long simulation runs in Figure 2 further show that there is a high degree of variation in the model performance. Generally speaking, variations in the model performance are due to model simplifications, which lead to an imperfect representation of reality. It is not clear to what extent calibrated parameters for the CTL are transferable to a different climate in the SCE. Indeed, the NSE series show lower values outside the CTL. Recent research has shown that stationary model parameters may lead to drifts in model performance over a longer period of time [*Merz et al.*, 2011] and for changes in climatic conditions [*Coron et al.*, 2012].

[37] Overall, the validation results indicate that the HMs perform reasonably well in the CTL. In our ensemble of opportunity, we use both models equivalently. Noting that a good performance during the past does not imply a good performance in a future climate [e.g., *Blöschl and Montanari*, 2010], we try to answer the question how much of the uncertainty is due to the choice of the HM system. The effect of the model imperfections on the climate change

signal clearly needs further systematic investigations. In particular, a potential relationship between model biases and the uncertainty in the climate projections needs to be addressed.

4.2. Hydrological Climate-Impact Projections

[38] We first discuss the two main input variables temperature and precipitation, followed by the two target variables that are the mean annual cycles of runoff and the runoff quantiles.

4.2.1. Temperature

[39] Figure 5 shows the results for the temperature projections, spatially averaged over the whole basin down to Diepoldsau. In the CTL, there is a clear annual cycle with a maximum in summer of about 11°C and a minimum in winter of about -6° C. The differences between the two input data sets OBS_{CHR} and OBS_{ST} are smaller than 1°C (see CTL DC in Figure 5). The spread of the BC runs in the CTL is smaller than 1°C throughout the whole annual cycle. In the two SCEs, the pattern of the annual cycle remains the same but is shifted to higher temperatures. The HadCM3Q0-driven CMs are at the higher end of the ensemble range in both SCEs, whereas SMHI-BCM-RCA is at the lower end. The temperature increase is in the range of 0.3°C-3.4°C for SCE1 and in the range of 1.7°C-5.9°C for SCE2. In the SCE1, there is no clear pattern in the annual cycle of the temperature change signal except for the HadCM3Q0-driven CMs that show a peak increase in winter and summer. In the SCE2, all CMs except DMI-ECHAM-HIRHAM and SMHI-BCM-RCA show a summer peak of the temperature increase.

4.2.2. Precipitation

[40] Figure 6 shows the annual cycle of the precipitation projections, averaged over the whole basin area down to Diepoldsau. The annual cycles for the DC runs in the CTL do not fully agree. This is because the two HMs use slightly different precipitation data sets as input (see section 2.2) and correct the observed precipitation using model-dependent calibrated precipitation correction factors (see section 3.1). Also, the precipitation of HBV-BC and PREVAH-BC simulations in the CTL does not fully agree, since the BC is done using the different precipitation reference data shown in Figure 6 (top left). The BC is able to correct the CM precipitation to match the observed mean annual cycle (see BC for the CTL in Figure 6). Note that the BC is carried out on a monthly basis. Thus, natural variability causes the 31d MA precipitation of the individual CMs to diverge between the monthly correction steps. Also, the BC was made for the full period 1961-1990, whereas we here only analyze 1964-1990 that causes some further deviations between the CMs as can be seen in June.

[41] In the SCEs, the ensembles of the BC runs show a larger spread than the DC ensembles. The bias-corrected DMI-ECHAM-HIRHAM runs project considerably higher precipitation peaks in summer than the rest of the ensemble. Also, the BC runs driven by HadCM3Q0 show a clear peak around May and a minimum around July. In the climate change signal as well, the mentioned CMs show some deviations from the rest of the ensemble. DMI-ECHAM-HIRHAM projects the strongest increase in summer precipitation, whereas the HadCM3Q0-driven CMs show the strongest increase in late spring of the SCE2. The DC runs generally



Figure 5. Mean annual cycle of basin-averaged temperature in (top) the CTL period, (middle) the two SCE periods, and (bottom) climate change signal as the differences SCE-CTL. The coloring of the CMs is the same for all PP-HM combination panels. The PP-HM combination is indicated in the title of each panel. For the DC runs, there is only one CTL run per HM, and, thus, the HBV and PREVAH runs are combined in the top left. The annual cycle has been low-pass filtered using a 31d MA.

show smaller variations in the annual cycle that is partly due to the spectral smoothing employed in the DC as opposed to the monthly based correction used in the BC. The pattern of the precipitation change signal is clearer in SCE2 than in SCE1 and shows decreasing precipitation in summer and slightly increasing precipitation in the rest of the year.

4.2.3. Annual Cycle of Runoff

[42] Figure 7 shows the annual cycles of simulated runoff at the gauge Diepoldsau in the CTL and the SCEs as well as the climate change signal. Annual and seasonal values are listed in Table 2. [43] The analysis of the mean annual cycle (see CTL DC in Figure 7) shows that HBV underestimates the mean runoff during the summer months as well as in February, March, and April. This is partly related to a lower waterbalance-corrected precipitation in OBS_{CHR} compared with OBS_{ST} (see CTL DC in Figure 6). The results for the BC runs in the CTL show similar characteristics as the CTL DC results. For example, HBV-BC runs tend to simulate a delayed onset of the snowmelt period in May instead of in April and an underestimation of the mean runoff during summer, whereas all the PREVAH-BC runs simulate a too low mean runoff in winter.



Figure 6. Same as Figure 5 but for water-balance-corrected precipitation.



Figure 7. Same as Figure 5 but for mean runoff at the gauge Diepoldsau. The observed mean runoff is added to the CTL period panels.

[44] In the SCE1, the peak in the annual cycle occurs about half a month earlier compared to the CTL for all modeling chains except for the runs using the CM DMI-ECHAM-HIRHAM, which projects a later peak and higher runoff. In the SCE2, the runoff peak occurs about 1 month earlier and has a smaller amplitude in comparison to the CTL. In the climate change signal, all modeling chains agree on an increase of winter runoff. All impact modeling chain combinations (except those driven by DMI-ECHAM-HIR-HAM for SCE1) show a decrease of summer runoff. These signals are stronger for SCE2 than for SCE1. Most BC runs project a peak in the increase of runoff in the snowmelt season around May. The climate change signals of the DC runs are smoother than the ones of the BC runs, which is due to spectral smoothing in the DC as opposed to monthly correction intervals in the BC. The model spread is highest in summer.

[45] The general pattern of seasonal runoff changes (a decrease and increase of summer and winter runoff, respectively, and an earlier runoff peak in the annual cycle) is consistent with the results of numerous previous studies in snow-dominated Alpine catchments [e.g., *Horton et al.*, 2006; *Graham et al.*, 2007; *Köplin et al.*, 2012].

4.2.4. Runoff Quantiles

[46] The runoff quantiles (see Figure 8) in the CTL DC simulations show the performance of the HMs. PREVAH

underestimates low runoff quantiles but overestimates high runoff quantiles. HBV matches the observed low and high runoff quantiles better but slightly underestimates the intermediate quantiles. In the BC results, larger deviations from the observations are apparent in the high quantile range. This is due to the BC that corrects for biases in the mean but not for biases in the whole quantile distribution. In the climate change signal for SCE1, all the DC runs project slight decreases of runoff in the medium and moderate high quantile range, and most DC model combinations agree on an increase of the 99.9% runoff quantile level. For SCE2, the decrease in the moderate to high quantile ranges is larger, and some modeling chains simulate a decrease of the 99.9% runoff quantile level. The BC runs also show slight decreases of runoff in the medium-to-moderate high quantile range, but, for quantiles above 90%, the BC runs disagree on the sign of the change. Furthermore, the spread of projected changes in the high quantile range is similar in both SCEs.

4.3. Variance Decomposition

[47] As described in section 3.3, we analyze the contribution of the three sources CM, PP, and HM to the total ensemble uncertainty of the runoff projections. In principle, such a variance decomposition is feasible for any variable of interest. We conducted the analysis for differences in the

Table 2. Annual and Seasonal Changes in Mean Runoff Expressed in Differences and Percentages^a

	SCE1 2024–2050		SCE2 2073–2099	
	SCE-CTL (mm/d)	(%)	SCE-CTL (mm/d)	(%)
Annual	-0.1 [-0.2-0.2]	-2.0 [-6.0-6.1]	-0.2 [-0.4-0.2]	-4.8 [-12.7-7.1]
DJF	0.2 [0.1–0.5]	15.7 [5.5-40.7]	0.6 [0.1–1.0]	41.8 [4.8-81.7]
MAM	0.2 [0.0-0.6]	8.6 [-0.4-23.9]	0.6 [0.3–0.9]	21.0 [10.1–37.7]
JJA	-0.8[-1.5-0.2]	-14.6[-26.0-3.9]	-1.7[-2.4-0.4]	-32.2[-45.5-8.1]
SON	0.1 [-0.2-0.3]	1.7 [-7.2-11.2]	-0.1 [-0.6-0.3]	-3.0 [-20.0-11.4]

^aThe ensemble mean is shown, followed in brackets by the range of the 32 hydrological climate runoff scenarios. The first 3 years of each SCE were cut off to exclude spin-up effects from the analysis.



Figure 8. Same as Figure 7 but for different runoff quantiles.

annual cycle of mean runoff (see bottom row in Figure 7)and for runoff quantiles (see bottom row in Figure 8).4.3.1. Changes in the Mean Annual Cycle of Runoff

[48] Figure 9 shows the variance decomposition of the climate change signal in the mean annual cycle of runoff. The CM is the dominant source during summer and autumn. In May, the variance explained by the PP shows a peak that is clearer in SCE2 than in SCE1. The high contribution of PP is apparent neither in the variance decomposition of precipitation nor of temperature (analysis not shown). Thus, we suspect that the differences between the PP methods build up during the snow-accumulation period (e.g., through differences in the covariance between temperature and precipitation) and take effect only during the snowmelt period. The HM is the least important individual source of uncertainty in SCE1, but it becomes the dominant source of uncertainty in winter and early spring of SCE2. This indicates that the HMs are indifferent to the projected temperature and precipitation changes for SCE1 but react differently to the projected climate for SCE2 that strongly deviates from the CTL climate. PREVAH responds more sensitively to the large temperature changes, as projected precipitation changes are mostly below +0.5mm/d during winter for SCE2, but PREVAH simulates a runoff increase of nearly +1mm/d, whereas HBV projects runoff changes just slightly above +0.5 mm/d. This pronounced difference between the HMs could be related to the difference in the validation performance during winter (see section 4.1), but a systematic study of the relationship between model performance in the CTL and the uncertainty in the projections requires further research. Note that although the relative contribution of the HMs to the total ensemble uncertainty is high during winter and spring, the absolute level of the total ensemble uncertainty is lower than in summer.

^[49] The contribution of the interaction term to the total ensemble uncertainty varies throughout the annual cycle between 0.1 and 0.4. The high contribution around April, particularly in SCE2, might be due to strong nonlinearities in the snowmelt process.

4.3.2. Changes in Runoff Quantiles

[50] Figure 10 depicts the variance decomposition of the changes in different runoff quantiles. For the discussion, we divide the quantile bins in a low (5%-35%), an intermediate (35%-80%), and a high range (80%-99.9%). The low range is closely connected to the runoff in the winter months. The high runoff quantile range is related to the runoff in the summer months. Runoff values in the intermediate quantile range occur most often in the snowmelt season as well as in late summer and autumn. This quantile range is though not as clearly linked to a particular season in the year as the other two quantile ranges.



Figure 9. Variance decomposition of the uncertainty in mean runoff changes at the gauge Diepoldsau in the course of the annual cycle. The uncertainty sources are CM, PP, and HM.



Figure 10. Same as Figure 9 but for variance decomposition of changes in different runoff quantiles.

[51] In the low quantile range, the contributions of the different uncertainty sources to the total ensemble uncertainty change considerably from SCE1 to SCE2. While, in SCE1, the CMs and the interactions are the dominant sources of uncertainty, it is the HMs that explain about 50% of the total ensemble uncertainty in SCE2. Please see the discussion of the winter period in section 4.3.1 for an explanation. In the high quantile range, the CMs and PP methods are the most important uncertainty sources in both SCEs. Weather events leading to such high runoff values are linked to the precipitation frequency and intensity, both of which are determined by the CM and PP. The interpretation of the results for the intermediate quantile range is not straightforward. As this quantile range is not clearly determined by a particular process or season in the year, the results reflect the time-varying contribution of the different uncertainty sources that we see in Figure 9 as well.

[52] The contribution of the interaction term varies between 0.1 and 0.4 in SCE1 and 0.1 and 0.3 in SCE2, but interactions are more important for low and high quantile in SCE1, whereas, in SCE2, they contribute more to the uncertainty in the intermediate and high quantile range.

4.3.3. Discussion of the Variance Decomposition

[53] The ANOVA results identified nonnegligible contributions of interactions between the different uncertainty sources. This indicates that, without considering interactions, the importance of individual uncertainty sources is overestimated, and it demonstrates the importance of testing all possible combinations of impact modeling chain elements. Note, though, that our setup does not allow for a precise quantification of the interactions, as we cannot disentangle them from errors in the ANOVA model. A more thorough quantification of interactions would require multiple realizations of each impact modeling chain combination by using different realizations of GCMs [von Storch and Zwiers, 1999; Yip et al., 2011].

[54] The ANOVA approach assumes independence between the different samples (e.g., between the different CMs). This assumption is not fulfilled, as several RCMs use the same driving GCM, CMs in general also share model structures [*Masson and Knutti*, 2011], the PP methods both are based on the statistics for the mean variable, and the HMs are conceptually similar.

[55] We have evaluated the influence of the unbalanced representation of CMs sharing the same driving GCM. In our ensemble of opportunity, there are five CMs using ECHAM5, two CMs using HadCM3Q0, and one CM using BCM as the driving GCM. The climate change signals for temperature and precipitation (Figures 5 and 6) show a grouping of the CMs that share the same driving GCM. We tested the influence of the unbalanced representation of the driving GCMs by analyzing subsets of CMs with only one CM per driving GCM. There are 10 such subsets, and averaging over the 10 ANOVA results leads to qualitatively similar results (i.e., in terms of uncertainty source rank) as presented in Figures 9 and 10 (not shown). This is most probably related to the fact that although there are GCMgrouping effects visible in the temperature and precipitation change signals, these grouping effects are diminished in the runoff change signal (Figure 7). As this might be different in other basins (e.g., more pluvial basins), we suggest to conduct similar tests in every impact study. Further research is required to quantify the correlations between the different members of the same impact modeling chain element and to develop methods to account for them in the quantification of the uncertainty sources.

[56] We have also analyzed the robustness of the results with respect to the outlier DMI-ECHAM5-HIRHAM, which is the only GCM-RCM that does not project a large decrease of runoff in summer. Excluding this CM from the analysis does not change the results substantially (not shown), and, thus, the results are robust with respect to this outlier.

5. Summary and Conclusions

[57] We investigated the hydrological climate-impact projections based on the eight different CMs, two statistical PP methods, and two HMs. An Alpine study region was chosen as a challenging test area. For the two scenario periods 2024–2050 and 2073–2099 with respect to 1964–1990, we estimated the total ensemble uncertainty of the hydrological projections and presented a methodological framework to quantify the contributions of different uncertainty sources.

[58] For both scenario periods, all modeling chain combinations show an increase in winter runoff, and most model combinations project a decrease in summer runoff. The resulting spread is highest in summer and for high (99% and 99.9%) runoff quantiles. For such high quantiles, the hydrological projections disagree on the sign of the changes.

[59] We used a method based on the theory of the ANOVA to decompose the total ensemble uncertainty into contributions from individual modeling chain elements. We complemented it with a subsampling scheme to minimize the biases related to the different sample sizes of the different uncertainty sources.

[60] The variance decomposition has identified the CMs to be the dominant source of uncertainty during summer and autumn for both scenario periods. This fully agrees with the results of previous studies [*Wilby and Harris*, 2006; Jasper et al., 2004; Graham et al., 2007; Schaefli

et al., 2007; Kay et al., 2009; Prudhomme and Davies, 2009; Jung et al., 2011]. For the far-term SCE period in winter and spring, however, the dominance of the CMs diminishes, and the PP methods and HMs become more important. The analysis of changes in different runoff quantiles reveals that the HMs become the dominant uncertainty source for changes in the low runoff quantiles for the farterm scenario period. For high runoff quantiles, both the PP methods and the CMs contribute between 30% and 60% to the total ensemble uncertainty. Interactions between the uncertainty sources are important as well, with contributions between 10% and 40%.

[61] The results shown are subject to a few limitations. First, they are based on an ensemble of opportunity with a fairly small number of ensemble members. The ensemble allowed us to assess the importance of three different uncertainty sources. Other potentially important sources such as natural climate variability [Jung et al., 2011], the emission scenario [Prein et al., 2011], and the HM parameters [Beven and Binley, 1992] have been neglected. Thus, the estimated uncertainty ranges are likely smaller than the overall uncertainty associated with hydrological climate impacts. Second, the variance decomposition assumes independence between the different samples of each impact modeling chain element. However, none of the modeling chain elements fulfill this assumption completely. Some CMs share the driving GCM, the PPs share the same approach in correcting for the mean, and the HMs share similar conceptual structures. Third, our setup does not allow to discriminate between interactions and potential model errors in the ANOVA model, as multiple realizations of every modeling chain combination would be necessary for that purpose.

[62] The present study was embedded in the AdaptAlp project (Adaptation to Climate Change in the Alpine Space [Korck et al., 2011], see www.adaptalp.org), which aimed to develop adaptation strategies based on scientific knowledge and risk management assessment. Knowledge about the contribution of different uncertainty sources may help to design future impact modeling studies. Our results indicate that none of the investigated sources are negligible and that some fraction of the uncertainty cannot be attributed to individual elements of the modeling chain but rather depends on interactions between these components. The potential for interactions furthermore requires future impact modeling studies to conduct multipropagation simulations [Kay et al., 2009], i.e., simulations in which the modeling chain elements are varied in all possible ways.

[63] Although our results as such are not transferable to catchments with different hydrological properties, the presented methods for quantifying different uncertainty sources are versatile and adjustable to other experiment designs and regions. The ANOVA approach in combination with the subsampling scheme presents a consistent framework for the quantification of different uncertainty sources in hydrological climate-impact modeling. of Meteorology and Climatology, the Swiss Federal Office for the Environment and the CHR. We would also like to thank the Seminar for Statistics of the ETH Zurich for its assistance. The Center for Climate Systems Modeling (C2SM) at ETH Zurich is acknowledged for providing technical and scientific support.

References

- Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär (2012), Exploring perturbed physics ensembles in a regional climate model, *J. Clim.*, 25(13), 4582–4599, doi:10.1175/JCLI-D-11-00275.1.
- Beven, K. (2006), A manifesto for the equifinality thesis, J. Hydrol., 320(1-2), 18-36, doi:10.1016/j.hydrol.2005.07.007.
- Beven, K. J., and A. M. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Proc*esses, 6, 279–298.
- Blöschl, G., and A. Montanari (2010), Climate change impacts—Throwing the dice?, *Hydrol. Processes*, 24, 374–381, doi:10.1002/hyp.7574.
- Boberg, F., and J. H. Christensen (2012), Overestimation of Mediterranean summer temperature projections due to model deficiencies, *Nat. Clim. Change*, 2, 433–436, doi:10.1038/nclimate1454.
- Bosshard, T., S. Kotlarski, T. Ewen, and C. Schär (2011), Spectral representation of the annual cycle in the climate change signal, *Hydrol. Earth Syst. Sci.*, 15, 2777–2788, doi:10.5194/hess-15-2777-2011.
- Buser, C. M., H. R. Künsch, D. Lüthi, M. Wild, and C. Schär (2009), Bayesian multi-model projection of climate: Bias assumptions and interannual variability, *Clim. Dyn.*, 33(6), 849–868, doi:10.1007/s00382-009-0588-6.
- Campbell, J. L., C. T. Driscoll, A. Pourmokhtarian, and K. Hayhoe (2011), Streamflow responses to past and projected future changes in climate at the Hubbard Brook Experimental Forest, New Hampshire, United States, *Water Resour. Res.*, 47, W02514, doi:10.1029/2010WR009438.
- CH2011 (2011), *Swiss Climate Change Scenarios CH2011*, C2SM, MeteoSwiss, ETH, NCCR Clim., and OcCC, Zurich, Switzerland, 88 pp.
- Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, 48, W05552, doi:10.1029/2011WR011721.
- Déqué, M., et al. (2007), An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections, *Clim. Change*, 81, 53–70.
- Déqué, M., S. Somot, E. Sanchez-Gomez, C. M. Goodess, D. Jacob, G. Lenderink, and O. B. Christensen (2012), The spread amongst ensembles regional scenarios: Regional climate models, driving general circulation models and interannual variability, *Clim. Dyn.*, 38, 951–964, doi:10.1007/s00382-011-1053-x.
- Dessai, S., and M. Hulme (2007), Assessing the robustness of adaptation decisions to climate change uncertainties: A case study on water resources management in the east of England, *Global Environ. Change*, 17(1), 59–72, doi:10.1016/j.gloenvcha.2006.11.005.
- Eberle, M., H. Buiteveld, K. Wilke, and P. Krahe (2005), Hydrological modelling in the River Rhine Basin. Part III—Daily HBV model for the Rhine Basin, *BfG-JAP 2555*, BfG, Koblenz, Germany.
- Finger, D., G. Heinrich, A. Gobiet, and A. Bauder (2012), Projections of future water resources and their uncertainty in a glacierized catchment in the Swiss alps and the subsequent effects on hydropower production during the 21st century, *Water Resour. Res.*, 48, W02521, doi:10.1029/ 2011WR010733.
- Görgen, K., et al. (2010), Assessment of climate change impacts on discharge in the Rhine River Basin: Results of the Rheinblick2050 project, *CHR Rep. I-23*, 229 pp., Lelystad, Netherlands.
- Graham, L. P., S. Hagemann, S. Jaun, and M. Beniston (2007), On interpreting hydrological change from regional climate models, *Clim. Change*, 81, 97–122, doi:10.1007/s10584-006-9217-0.
- Gurtz, J., A. Baltenswiler, and H. Lang (1999), Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins, *Hydrol. Processes*, 13, 2751–2768.
- Horton, P., B. Schäfli, A. Mezghani, B. Hingray, and A. Musy (2006), Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty, *Hydrol. Processes*, 20, 2091–2109, doi:10.1002/hyp.6197.
- Jasper, K., P. Calanca, D. Gyalistras, and J. Fuhrer (2004), Differential impacts of climate change on the hydrology of two alpine river basins, *Climate Res.*, 26, 113–129.

^[64] Acknowledgments. The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (contract 505539) whose support is gratefully acknowledged. We are also thankful for the RheinBlick2050, CLIMPACT (FNR C09/SR/16), the KLIWAS, the AdaptAlp, and the CCHydro projects that fostered collaboration. The observational data were provided by MeteoSwiss, the Swiss Federal Office

- Jung, I.-W., H. Change, and H. Moradkhani (2011), Quantifying uncertainty in urban flooding analysis considering hydro-climatic projection and urban development effects, *Hydrol. Earth Syst. Sci.*, 15, 617–633, doi:10.5194/hess-15-617-2011.
- Kay, A. L., H. N. Davies, V. A. Bell, and R. G. Jones (2009), Comparison of uncertainty sources for climate change impacts: Flood frequency in England, *Clim. Change*, 92, 41–63, doi:10.1007/s10584-008-9471-4.
- Knutti, R. (2008), Should we believe model predictions of future climate change?, *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.*, 366(1885), 4647–4664, doi:10.1098/rsta.2008.0169.
- Köplin, N., B. Schädler, D. Viviroli, and R. Weingartner (2012), Relating climate change signals and physiographic catchment properties to clustered hydrological response types, *Hydrol. Earth Syst. Sci.*, 16, 2267– 2283, doi:10.5194/hess-16-2267-2012.
- Korck, J., B. Mayer, F. Rudolf-Miklau, P. Greminger, A. Zischg, and M. Damm (2011), Climate Adaptation and Natural Hazard Management in the Alpine Space: Final report, AdaptAlp.
- Kundzewicz, Z. W., and E. Z. Stakhiv (2010), Are climate models "ready for prime time" in water resources management applications, or is more research needed?, *Hydrol. Sci. J.*, 55(7), 1085–1089, doi:10.1080/ 02626667.2010.513211.
- Kundzewicz, Z. W., L. J. Mata, N. W. Arnell, P. Döll, B. Jimenez, K. Miller, T. Oki, Z. Şen, and I. Shiklomanov (2008), The implications of projected climate change for freshwater resources and their management, *Hydrol. Sci. J.*, 53(1), 3–10, doi:10.1623/hysj.53.1.3.
- Lenderink, G., A. Buishand, and W. van Deursen (2007), Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach, *Hydrol. Earth Syst. Sci.*, 11(3), 1145–1159, doi:10.1023/B:CLIM.0000013692.50640.55.
- Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, 201, 272–288.
- Lopez, A., F. Fung, M. New, G. Watts, A. Weston, and R. L. Wilby (2009), From climate model ensembles to climate change impacts and adaptation: A case study of water resource management in the southwest of England, *Water Resour. Res.*, 45, W08419, doi:10.1029/ 2008WR007499.
- Lucas-Picher, P., D. Caya, R. de Elía, and R. Laprise (2008), Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain, *Clim. Dyn.*, 31(7–8), 927–940, doi:10.1007/s00382-008-0384-8.
- Margot, A., R. Sigg, B. Schädler, and R. Weingartner (1992), Beeinflussung der Fliessgewässer durch Kraftwerke (≥300 kW) und Seeregulierung, in *Hydrologischer Atlas der Schweiz*, vol. Plate 5.3, edited by D. Bérod, M. Spreafico, and R. Weingartner, Fed. Off. for the Environ., Bern, Switz.
- Masson, D., and R. Knutti (2011), Climate model genealogy, *Geophys. Res. Lett.*, 38, L08703, doi:10.1029/2011GL046864.
- Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, 47, W02531, doi:10.1029/2010WR009505.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity is dead: Whither water management?, *Science*, 319(5863), 573–574, doi:10.1126/science.1151915.
- Nakicenovic, N., and R. Swart (2000), *IPCC Special Report on Emission Scenarios*, 570 pp., Intergov. Panel on Clim. Change, Cambridge Univ. Press, Cambridge, U. K.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models. Part I—A discussion of principles, J. Hydrol., 27(3), 282–290.
- Pappenberger, F., and K. J. Beven (2006), Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resour. Res.*, 42, W05302, 1–8, doi:10.1029/2005WR004820.

- Prein, A. F., A. Gobiet, and H. Truhetz (2011), Analysis of uncertainty in large scale climate change projections over Europe, *Meteorol. Z.*, 20(4), 383–395, doi:10.1127/0941-2948/2011/0286.
- Prudhomme, C., and H. Davies (2009), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: Future climate, *Clim. Change*, 93(1–2), 197–222, doi:10.1007/s10584-008-9461-6.
- Quintana-Seguí, P., F. Habets, and E. Martin (2011), Comparison of past and future Mediterranean high and low extremes of precipitation and river flow projected using different statistical downscaling methods, *Nat. Hazards Earth Syst. Sci.*, 11(5), 1411–1432, doi:10.5194/nhess-11-1411-2011.
- Schaefli, B., B. Hingray, and A. Musy (2007), Climate change and hydropower production in the Swiss Alps: Quantification of potential impacts and related modelling uncertainties, *Hydrol. Earth Syst. Sci.*, 11(3), 1191–1205.
- Schulla, J. (1997), Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaveränderungen, Ph.D. dissertation, ETH Zurich, Zurich, Switz.
- Seiller, G., F. Anctil, and C. Perrin (2012), Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, *Hydrol. Earth Syst. Sci.*, 16(4), 1171–1189, doi:10.5194/hess-16-1171-2012.
- Sevruk, B., and M. Nevenic (1998), The geography and topography effects on the areal pattern of precipitation in a small prealpine basin, *Water Sci. Technol.*, *37*(11), 163–170.
- Shabalova, M. V., W. P. A. van Deursen, and T. A. Buishand (2003), Assessing future discharge of the river Rhine using regional climate model integrations and a hydrological model. *Climate Res.*, 23, 233–246.
- Teutschbein, C., and J. Seibert (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, doi:10.1016/j.jhydrol.2012.05.052.
- van der Linden, P., and J. Mitchell (Eds.) (2009), *ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project*, Met Office Hadley Cent., Exeter, U. K.
- Verbunt, M., M. G. Zwaaftink, and J. Gurtz (2005), The hydrologic impact of land cover changes and hydropower stations in the Alpine Rhine basin, *Ecol. Model.*, 187, 71–84.
- Verbunt, M., M. Zappa, J. Gurtz, and P. Kaufmann (2006), Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, J. Hydrol., 324, 224–238, doi:10.1016/ j.hydrol.2005.09.036.
- Vicuna, S., J. A. Dracup, J. R. Lund, L. L. Dale, and E. P. Maurer (2010), Basin-scale water system operations with uncertain future climate conditions: Methodology and case studies, *Water Resour. Res.*, 46, W04505, 1–19, doi:10.1029/2009WR007838.
- Viviroli, D., M. Zappa, J. Gurtz, and R. Weingartner (2009), An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, *Environ. Model. Software*, 24(10), 1209–1222, doi:10.1016/j.envsoft.2009.04.001.
- von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, U. K.
- Wilby, R. L. (2010), Evaluating climate model outputs for hydrological applications, *Hydrol. Sci. J.*, 55(7), 1090–1093, doi:10.1080/ 02626667.2010.513212.
- Wilby, R. L., and I. Harris (2006), A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK., *Water Resour. Res.*, 42, W02419, doi:10.1029/2005WR004065.
- Yip, S., C. A. T. Ferro, D. B. Stephenson, and E. Hawkins (2011), A simple, coherent framework for partitioning uncertainty in climate predictions, *J. Clim.*, 24, 4634–4643, doi:10.1175/2011JCLI4085.1.
- Zappa, M., and J. Gurtz (2003), Simulation of soil moisture and evapotranspiration in a soil profile during the 1999 MAP-Riviera campaign, *Hydrol. Earth Syst. Sci.*, 7(6), 903–919.