Importance of hydrological uncertainty assessment methods in climate change impact studies

M. Honti\textsuperscript{1,2}, A. Scheidegger\textsuperscript{1}, and C. Stamm\textsuperscript{1}

\textsuperscript{1}Eawag: Swiss Federal Institute of Aquatic Sciences and Technology, Ueberlandstrasse 133, 8600 Dübendorf, Switzerland
\textsuperscript{2}Water Research Group of the Hungarian Academy of Sciences, Műegyetem rkp. 3, 1111 Budapest, Hungary

Received: 22 November 2013 – Accepted: 17 December 2013 – Published: 14 January 2014

Correspondence to: M. Honti (honti@vit.bme.hu)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Climate change impact assessments have become more and more popular in hydrology since the middle 1980’s with a recent boost after the publication of the IPCC AR4 report. During hundreds of impact studies a quasi-standard methodology emerged, which is mainly shaped by the growing public demand for predicting how water resources management or flood protection should change in the following decades. The “standard” workflow relies on a model cascade from global circulation model (GCM) predictions for selected IPCC scenarios to future catchment hydrology. Uncertainty is present at each level and propagates through the model cascade. There is an emerging consensus between many studies on the relative importance of the different uncertainty sources. The prevailing perception is that GCM uncertainty dominates hydrological impact studies. Our hypothesis was that the relative importance of climatic and hydrologic uncertainty is (among other factors) heavily influenced by the uncertainty assessment method. To test this we carried out a climate change impact assessment and estimated the relative importance of the uncertainty sources. The study was performed on two small catchments in the Swiss Plateau with a lumped conceptual rainfall runoff model. In the climatic part we applied the standard ensemble approach to quantify uncertainty but in hydrology we used formal Bayesian uncertainty assessment with two different likelihood functions. One was a time-series error model that was able to deal with the complicated statistical properties of hydrological model residuals. The second was an approximate likelihood function for the flow quantiles. The results showed that the expected climatic impact on flow quantiles was small compared to prediction uncertainty. The source, structure and composition of uncertainty depended strongly on the uncertainty assessment method. This demonstrated that one could arrive to rather different conclusions about predictive uncertainty for the same hydrological model and calibration data when considering different objective functions for calibration.
1 Introduction

Climate change impact assessments have become more and more popular in hydrology since the middle 1980’s (Gleick, 1986, 1989; Arnell, 1992) with the recognition that the global climate can be influenced by humankind and that the growing emission of greenhouse gases into the atmosphere has already started a global warming. This topic received another boost when the public perception of climate change transformed after the publication of the IPCC AR4 report (IPCC, 2007) and climate change became a politically and economically accepted boundary condition for the future. From that point onwards no responsible planning could omit the possible effects of an altered climate on water availability, flood levels or other hydrological resources or threats. Hundreds of studies were carried out on almost every significant catchment of the world (for a global summary see Kundzewicz et al., 2007; for a selection of more recent studies refer to Todd et al., 2011). During this bloom of impact studies a quasi standard methodology emerged (Blöschl and Montanari, 2010; Todd et al., 2011). The procedure is mostly shaped by the growing public demand for predicting how water resources management or flood protection should change in the close future. Impact studies need to accomplish an apparently impossible task: simulate future relevant hydrological events driven by local or extreme meteorological phenomena, which cannot be described by present climatic models. The common procedure based on a pragmatic approach contains the following steps (Blöschl and Montanari, 2010; Todd et al., 2011):

The future climate under a specific IPCC emission scenario is simulated with global circulation models (GCMs), whose output preferably feeds into dynamical downscaling: the GCM results serve as boundary conditions for a regional climate model (RCM) running on a much finer spatial resolution. The output from the climate models is typically corrected for the bias between the observed weather and the climate simulations for the reference period. When the daily resolution of GCMs and RCMs is too coarse compared to the requirements of the study, a frequent choice is to apply stochastic
...downscaling (frequently in the form of weather generators). Then a calibrated hydro-
logical model is run on the past and future meteorological data to analyse the impacts of climate change on the hydrological indicators of interest.

The impact predictions are as uncertain as any forecast that tries to describe the behaviour of an extremely complex system decades into the future. First of all, future climate predictions are uncertain due to the intrinsic uncertainty of their inputs: future emission scenarios are represented by a handful of representative story-lines managed by the IPCC, the translation of emissions and projected radiative forcing into actual weather is done with by GCMs that exhibit obvious deficiencies in simulating phenomena on finer resolution than continental scale (Xu, 1999; Blöschl and Montanari, 2010; Ehret et al., 2012) or according to some metrics do not work at all (Koutsoyiannis et al., 2008; Koutsoyiannis, 2010). Consequently, the GCM-based descriptions of the future climate are preferred to be called as “projections” instead of forecasts due to the immense amount of uncertainty caused by the above described factors (IPCC, 1995). Additionally, there is a non-quantified uncertainty that does not appear in ensembles of emission-scenario-GCM combinations (Jones, 2000). Like in any hierarchical model system, uncertainty propagates from the climate predictions through the descendant components to regional or local hydrological projections. Downscaling increases uncertainty with the deficiencies of RCMs and/or the imperfect stochastic description of the weather by a weather generator (Khan et al., 2006; Kay et al., 2009). Bias correction adds a strong deterministic shift to the input data (Ehret et al., 2012). Finally the predictive uncertainty of the hydrological model ends the cascade that leads to the total uncertainty of the hydrological impact assessment.

The high uncertainty of the impact of climate change on stream flow is usually admitted, but less often quantified properly. Some studies publish the impacts without any quantification of their uncertainty (Arnell, 2003; Gosain et al., 2006; Thodsen, 2007). Others mostly follow the semi-qualitative description of uncertainty throughout the entire model hierarchy by performing ensembles of simulations with different climate and hydrological model components and settings (Boorman and Sefton, 1997;...
Nijssen et al., 2001; Booij, 2005; Kingston and Taylor, 2010; Gosling et al., 2011; Chen et al., 2011) or focus only on climatic uncertainty and neglect hydrological uncertainty at all (Christensen et al., 2004; Maurer, 2007; Chiew et al., 2009) or even take a single climatic projection and assess only the hydrological uncertainty (Steele-Dunne et al., 2008). Despite the continuous development of quantitative uncertainty assessment methods such as formal Bayesian statistical approaches (Kuczera et al., 2006; Kavetski et al., 2006; Honti et al., 2013) or the GLUE methodology (Beven and Freer, 2001), these methods are relatively rarely preferred over taking a hydrological model ensemble. There are a few examples of applying GLUE for the estimation of hydrological predictive uncertainty in the context of climate change impact assessment (Cameron, 2006; Wilby, 2005; Wilby and Harris, 2006; Prudhomme and Davies, 2009a; Zambrano-Bigiarini, 2010), but Bayesian uncertainty assessment methods are missing (to our best knowledge).

Despite the diversity in the uncertainty assessment methodology applied in the context of hydrological climate change assessment, there is an emerging consensus between many studies on the relative importance of the different uncertainty sources. The prevailing perception is that GCM uncertainty dominates hydrological impact studies (Wilby and Harris, 2006; Graham et al., 2007; Prudhomme and Davies, 2009b; Kay et al., 2009; Kingston and Taylor, 2010; Arnell, 2011; Hughes et al., 2011; Gosling et al., 2011). There are only few studies, which found that the predictive uncertainty of hydrological models can be in the same range or even larger than climatic uncertainty. This special finding was typically coupled to unusual circumstances: poor hydrologic model performance already in the calibration period (Ludwig et al., 2009), application of an extremely error-tolerant equifinality criterium (Zambrano-Bigiarini, 2010) or very different spatial scales treated together during the hydrological modelling (Abbaspour et al., 2009).

However, the universal dominance of climatic uncertainty can be challenged if we consider that the most popular formal and informal likelihood calculation methods (RMSE in GLUE, i.i.d. white noise in formal Bayesian calibration) tend to underestimate
hydrological predictive uncertainty due invalid statistical assumptions about the residuals (Schoups and Vrugt, 2010; Reichert and Schuwirth, 2012). Our hypothesis is that the relative importance of climatic and hydrologic uncertainty does not only depend on the hydrological and climate models and the application site, but are also conditional on the uncertainty assessment method.

Our objective is to test the above hypothesis with a climate change impact assessment including statistically sound estimates on the relative importance of the uncertainty sources. The study is performed on 2 small catchments in the Swiss Plateau with a lumped conceptual rainfall runoff model (CRRM). In the climatic part we apply the standard ensemble approach to quantify uncertainty but in hydrology we use formal Bayesian uncertainty assessment method with two different likelihood functions. One is a time-series error model that is able to deal with the complicated statistical properties of hydrological model residuals. The second is an approximate likelihood function directly for the flow quantiles. The use of this quantile approach roots in two observations: first, climate change impact assessment is mostly interested in magnitudes of flow of a given return periods. The exact timing and hence the time-series is not in the focus – the objective functions are statistics on the predicted time-series. The second reason for the quantile approach is the fact that uncertainty of these estimates when derived from the time-series are not straightforward to quantify properly in a statistical sense. Directed targeting the objective function of interest may therefore offer advantages that shall be explored in this article.

2 Methods

2.1 Study sites and discharge data

Our test catchments are the Mönchaltofer Aa (46 km²) and the Gürbe (137 km²), both lying on the Swiss Plateau on the Northern side of the Alps (Fig. 1). The dominant landuse types are intensive agriculture followed by forests in both sites (57 and 15 %
in the Mönchaltorfer Aa, 51 and 21 % in the Gürbe; swisstopo, 2008). Topography is rather different, the altitude difference between the uppermost point and the outflow is moderate for the Mönchaltorfer Aa (440–850 m a.s.l.), while highest point in the southern mountainous headwater catchment of the Gürbe is 1650 m above the river’s mouth near Belp (500–2150 m a.s.l.). Soil texture in the Mönchaltorfer Aa catchment is predominantly loamy (BLW, 2008) with cambisols on hillsides and gleysols on flat areas (Wittmer et al., 2010) as major soil types. The lowland area of the Gürbe has similar characteristics, while the alpine part is dominated by coarser, sandy soil material (BLW, 2008).

Discharge is monitored only at a single point along the Mönchaltorfer Aa, close to the outlet (Mönchaltorf) by the Office for Waste, Water, Energy and Air Quality of Kanton Zürich with 10 min frequency (AWEL, 2010). The Gürbe possesses 2 regular discharge gauges in the main channel operated by the Office of Waste and Water of Kanton Bern (AWA, 2010). One is slightly upstream from the river outlet (Belp), while the other is located about halfway to the headwaters (Burgistein).

2.2 Climatic input data

2.2.1 Observed climatic data

Regular high-resolution meteorological measurements were only available at one off-catchment location for each test catchment. The automatic measurement station of MeteoSchweiz at Wädenswil (10 km SW from Mönchaltorf) was used to drive the model of the Mönchaltorfer Aa, while the station at Bern Zollikofen (10 km N/NE from Belp) was the input for the lower subcatchments of the Gürbe. Additional daily rainfall data from the nearby Blumenstein gauge were used for the uppermost subcatchment of the Gürbe due the significant altitude and climatic difference compared to the lower parts (1300 vs. 700 m a.s.l. average elevation, 1260 vs. 1140 mm yr\(^{-1}\) in average precipitation).
Daily potential evapotranspiration was calculated from radiation and temperature with the simple Hargreaves–Samani method (Hargreaves and Samani, 1982), which was calibrated to match reference crop evapotranspiration given by the full FAO Penman–Monteith equation (Allen et al., 1998).

2.2.2 Stochastic weather generation

The EARWIG/UKCP09 statistical weather generator (Kilsby et al., 2007) was trained on the 1981–2010 weather series from Wädenswil and Bern-Zollikofen. This weather generator relies on the Neyman-Scott rectangular pulses (NSRP) model (Rodriguez-Iturbe et al., 1987; Cowpertwait et al., 1996) for the generation of hourly precipitation and simple autoregressive models for the daily values of other weather variables. The NSRP model was trained by optimising the formal statistical properties of the model to match those of the observed data following the procedure described in Fatichi et al. (2011a, b). The autoregressive coefficients were calibrated conditionally on the season (determined by half-monthly periods) and the transitions between wet (W) and dry (D) days. There were altogether 4 + 1 transition types: WW, WD, DW, DD (Kilsby et al., 2007) and additionally the DDD type for lasting droughts, which was introduced in the latest version of UKCP09 (Jones et al., 2011).

The difference in weather between the alpine and lowland parts of the Gürbe catchment should have been reflected in the generated weather data too, but the application of a spatial weather generation method like the STNSRP model (Burton et al., 2008) or the NSAR model (Burton et al., 2010) was impossible due to the lack of high-frequency observations for the Blumenstein gauge. To overcome this problem we generated the alpine precipitation conditionally on the lowland weather with a black box model (Appendix A).
2.2.3 Future climatic data

Climate change was represented by 10 GCM-RCM model chains from the ENSEMBLES project database (http://www.ensembles-eu.org) featuring 4 GCMs (including HadCM3 with 2 different sensitivities) and 8 RCMs, all ran transiently on the IPCC A1B emission scenario (Table 6). Despite the transient data, we chose to have two distinct stationary climates, one for the present and 1 for the future for a better statistical representation of climatic variability. The reference period was 1981–2010, while the forecast period was a 30-yr period centred around 2050 (2035–2064). The relatively close forecast time horizon meant that it was sufficient to take a single emission scenario as a representative for all, because the temperature effects of different emission scenarios are still quite similar in this period.

Direct RCM output was not usable for hydrological modeling because both test catchments are situated in the prealpine precipitation gradient zone, which is poorly captured with the coarse spatial resolution of RCMs. This caused that differences between annual precipitation sums from raw RCM data and observations were always significant, for some model chains reaching even 200%. According to the common practice, we applied bias correction to the statistics of precipitation and air temperature. The resulting factors of change were introduced to the weather generator following the procedure outlined in Kilsby et al. (2007).

2.3 Hydrological model

We used a modified version of the logSPM model (Kuczera et al., 2006; Honti et al., 2013). LogSPM belongs to the saturated path family of conceptual rainfall-runoff models, where the heart of the model is a non-linear function describing the saturated proportion of the catchment area as the function of the mean soil moisture content (Kavetski et al., 2003). Water is routed in the soil storage as either runoff, subsurface flow or recharge, proportionally to the saturated area.
The parameterisation of the saturation function relies on the catchment-scale analogies of characteristic soil moisture contents:

\[
    f_{\text{sat}}(h_s) = \frac{1}{1 + \exp \left( \frac{4 h_{\text{FS}} + h_{\text{FC}} - 2 h_s}{h_{\text{FS}} - h_{\text{FC}}} \right)} - \frac{1}{1 + \exp \left( \frac{4 h_{\text{FS}} + h_{\text{FC}}}{h_{\text{FS}} - h_{\text{FC}}} \right)},
\]  

(1)

where \( h_s \) is the average soil moisture content of the catchment, \( h_{\text{FS}} \) and \( h_{\text{FC}} \) are the catchment-scale storage level equivalents of full saturation and field capacity with 98 and 2 % of the catchment area saturated, respectively. Evapotranspiration from the soil moisture storage is controlled in a similar manner to \( f_{\text{sat}} \):

\[
    f_{\text{et}}(h_s) = \frac{1}{1 + \exp \left( \frac{8 \frac{3 h_{\text{WP}} + h_{\text{FC}} - h_s}{3 h_{\text{WP}} + h_{\text{FC}}}} \right)} - \frac{1}{1 + \exp (8)},
\]  

(2)

where \( h_{\text{WP}} \) is the catchment-scale moisture level equivalent of the wilting point (the actual evapotranspiration is only 5 % of the potential). The groundwater and stream storages are simple linear reservoirs (Table 1). To simulate hydrology under different topographic or landcover conditions, this basic conceptual model was combined with a snow module based on the degree-day method (Martinec and Rango, 1981), a canopy module based on the interception model of Vrugt et al. (2003) and a simple non-leaking threshold storage for paved areas (Table 1, Fig. 2).

### 2.3.1 Spatial discretisation

The test catchments were spatially discretised using the HRU concept based on landuse and soil similarity. The subcatchments of discharge gauging stations were split into “forest”, “grassland” (including true grasslands, treeless agricultural areas, non-paved urban zones) and “paved” classes, each represented by a single HRU. Each HRU was assigned a separate soil and canopy unit. Similar HRUs shared a common parameter set. Soil types were assumed to be exclusively from the loamy category on
the entire Mönchaltorfer Aa catchment (Frey et al., 2011), while the Gürbe was divided into a lower and upper zone with loamy and sandy soil types, respectively. In the end there were 3 HRUs in the Mönchaltorfer Aa due to the 3 distinct landuse types and the lack of a topographic division, while for the Gürbe the separate treatment of the upper and lower zones above the Burgistein gauge and the additional lowland subcatchment between Burgistein and Belp forced the application of 9 HRUs in total.

2.3.2 Parameter priors

Due to the lack of previous conceptual modelling studies in the test catchments we collected prior knowledge about the parameters of the hydrological model by a literature review. Thanks to the reuse of simple and well-known modelling blocks for the snow, canopy and paved module we found several relevant parameter estimates (Tables 2 and 3). The prior values for the dripping rate from the canopy storage \( k_{\text{drip}} \) were so high compared to the daily resolution of the computation, that this parameter was fixed to \( 400 \text{ d}^{-1} \) and excluded from the calibration (Table 2). Priors for the characteristic average soil moisture contents were derived from the water retention curves of the dominant soil types with the van Genuchten model (van Genuchten, 1980) and the default parameters from the ROSETTA program (Schaap et al., 2001) and the assumption of a 1 m thick active surface layer (Table 4). Priors for the conceptual catchment parameters \( k_{\text{rge}}, k_{\text{bf}}, k_{q}, \text{etc.} \) were formulated with subjective estimation on their acceptable domain (Table 4).

2.3.3 Hydrological indicators

To facilitate the comparison of observed and predicted hydrological conditions, we rely on a small set of aggregated discharge statistics (namely flow quantiles) similarly to some previous studies (Arnell, 1992; Gosling et al., 2011). We use the 95, 50 and 5 % exceedance quantiles (Q95, Q50 and Q5, respectively) at the discharge gauge sites to represent low, medium and high flow. The selection of these less extreme occurrence
probabilities for the assessment ensures that the outcome does not depend heavily on truly extreme events. This is important for two technical reasons: first, the occurrence of extreme precipitation events and consequently extreme floods may be influenced more by internal climatic variability than by climate change itself (Fatichi et al., 2013). However, we describe this only to a very limited extent by taking 30 yr of data from each model chain, which is inappropriate to represent rare events. Second, extremes are generally poorly simulated by models of both climate and hydrology, so going for extreme flow indicators would further increase predictive uncertainty.

2.4 Impact and uncertainty assessment

The total uncertainty of our hydrological predictions was assessed with a hybrid approach. Similarly to the majority of climate impact studies we also assumed that the 10 GCM-RCM chains properly represent the uncertainty of the modelled future climate. However, contrary to others (Boorman and Sefton, 1997; Booij, 2005; Gosling et al., 2011) we did not apply the same approach to the hydrological side by representing the existing hydrological uncertainty with a set of different model structures or settings: we used a single conceptual model for hydrology with 3 versions of 2 Bayesian uncertainty assessment techniques.

2.4.1 Discharge time-series approach (TS)

The first variant for the hydrological uncertainty analysis relied on the predictive uncertainty of future discharge series. We considered an additive frequentist observation error together with a similarly additive Bayesian bias process that was designed to represent the effects of both model structural deficiencies and input uncertainty (Honti et al., 2013). The predicted future “true” discharge arose from the output of the deterministic CRRM plus the stochastic bias process reflecting epistemic uncertainty (Honti et al., 2013).
A sample of the posterior parameter distribution was generated with Markov Chain Monte Carlo sampling (for details see Honti et al., 2013). The posterior sample was used to produce model predictions of discharge and then the selected flow quantiles for each stage of the impact assessment workflow (Fig. 3, stage 0 being the observed discharge):

1. Based on observed weather data (1981–2010).

2. Based on generated weather data that reflect the reference climate (1981–2010).

3. Based on generated weather data that reflect a stationary future (2035–2064) climate projection by a selected GCM-RCM model chain.

Since all GCM-RCM model chains refer to the same future climate that realises under the IPCC A1B emission scenario, the predictions from the different model chains together represent the future climate. We did not differentiate between individual model chains based on their skill or performance as for example Gleckler et al. (2008) did, so the future climate in our assessment was represented with a model chain ensemble with uniform weights. Accordingly, the corresponding flow indices could be mixed together to get a sample of the future.

The uncertainty of weather generation and the stochastic downscaling of the future climate was assessed only implicitly. We assumed that the lengths of the baseline and prediction periods (30 yr both) are enough to produce a statistically well defined sample of the target flow indicators, so we used one realisation of the generated weather for the reference period and one for each model chain prediction. This is a reasonable assumption if we consider the probability of more extreme events than the selected flow quantiles: the expected number of days with flow outside the [Q95, Q5] interval is 1095 in a 30 yr period.
2.4.2 Discharge quantile approach (K)

Besides deriving the target flow indicators from the predicted time-series we also applied a direct approach. We kept the same CRRM, but performed the calibration and produced the parameter posterior sample based on the approximate likelihood of the quantiles themselves. Under some mild statistical assumptions the likelihood of the quantiles can be approximated by independent normal distributions. The details of the approximate quantile likelihood function are described in Appendix B.

Calibration without time-series fitting has already been quite common in hydrology. For example, Montanari and Toth (2007) used the spectral properties of the flow time-series as a measure of fit. Blazkova and Beven (2009) used certain flow quantiles among several other aggregated measures as acceptability criteria for their GLUE-based approach. Westerberg et al. (2011) performed model calibration based on fitting the flow-duration curves with a triangular informal likelihood function. In our case we used a formal statistical approach to the essentially same problem as Westerberg et al. (2011) addressed: using aggregated flow-statistics for reference offers interesting possibilities. Flow quantiles are independent of time. This means that timing errors, like slightly early or delayed flood peaks does not influence the model performance significantly.

We utilized this property in the estimation of climate change impacts. In the first variant (K1) of the quantile approach we went through the same workflow stages as described for the time-series approach. However, in the 2nd variant (K2) we merged stages 1–2 because we used the observed discharge data and the generated weather for the present together for calibration and the sampling of the parameter posterior (Fig. 3).

2.5 Comparison of different uncertainty effects

The relative importance of uncertainty entering different stages of the impact assessment and the effect of climate change itself was compared with a simple approach. The
change in the flow index distribution between the observed uncertainty of the flow quantiles and stage 1 corresponds to the effect of hydrologic modeling in TS and K1 and to the composite effect of hydrological modelling and weather generation in K2. Similarly, transition between stages 1 and 2 reflects the effect of using generated weather data instead of the observations in K1 and the effect of the internal variability of weather generation in K2.

In theory, the flow quantiles in stages 0–2 should not differ as these all represent the same present hydrology. Accordingly, any change in the distribution of flow quantiles during the transition between these stages can be attributed to existing hydrological and meteorological uncertainty. In contrast, the change in the distribution of flow quantiles between stage 2 and the joint predictions of the future by the 10 model chains should reflect the impact and uncertainty of climate change.

3 Results

We present the results for the different uncertainty assessment approaches by study site.

3.1 Mönchaltorfer Aa

3.1.1 Time-series approach

In accordance with our expectations, the CRRM performed well in simulating the observed discharge data with TS. The maximum likelihood solution had a Nash-Sutcliffe index (NS) of 0.8. Despite this good model performance, the selected flow quantiles (Q95, Q50 and Q5) showed significant uncertainty already in stages 1 and 2 without the effects of climate change (Fig. 4). Although these stages both should have corresponded to the observed reference meteorological and hydrological conditions, the simulated flow quantiles were biased in each stage and their variability was significantly larger than that of the observations (Fig. 5). Stage 1 introduced a relative
offset between −5 to −10 % for each quantile, while the 95 % uncertainty interval width was between 10 and 20 % of the observed flow indices. Weather generation (stage 2) caused a significant positive offset, which over-compensated for the underestimation in stage 1.

Since stages 1 and 2 did not produce flow quantiles identical to the observations, the operational definition of the climate change impact matters. Just comparing the observed quantiles with results from stage 3 would yield an increase for all three flow indices. However, quantifying the climate change effect as the difference between stages 2 and 3 shows that Q95 was predicted to increase slightly in the future, while Q50 and Q5 were likely to decrease. One had to notice that the expected climate change impact was always much smaller than the offsets caused by the previous stages. The procedure of analysing the difference between stages 2 and 3 essentially meant that we applied a bias-correction for the quantile offsets caused by the present uncertainty.

The variability of future flow indices was high for each of them (Table 7), compared to the expected climate change impact (Fig. 6). However, the source of this variability differed by flow index: for Q95 it was the modeling uncertainty (stage 1), while for Q50 and Q5 it was the future climate (stage 3) which contributed most to the final uncertainty.

3.1.2 Quantile approach

To our surprise, the calibration to quantiles with observed weather data (K1) also resulted in good agreement between the simulated and the observed flow time-series (NS = 0.75) although timing had not been considered in the calibration process. When calibrating with the generated weather series (K2), this performance decreased to NS = 0.65 but could still be considered as satisfactory. These good fits demonstrated that the temporal dynamics of discharge were clearly determined by precipitation, so one could predict discharge peaks quite well even without looking at the time-series during the calibration of CRRM parameters.
Flow quantiles could be calibrated well to the observed discharge in each version (K1,2); quantile offsets were between 1–6%. In K1, the quantiles still showed significant offsets upon calibration with the generated weather series (stage 2; Fig. 7). The weather generation again caused a positive bias in two flow quantiles (Q50, Q5). This could be completely avoided in K2. This meant that K2 resulted in different CRRM parameters to correct for any bias between the original stages 1 and 2 resulting from the interaction between uncertainty of the hydrological model and the weather generator.

Flow quantile variability was dominated by the future climate uncertainty in both versions (variance in stage 3 was much higher than in the previous stage(s)). In K1 weather generation was the most important source of bias (Fig. 7), while this was almost completely eliminated in K2 (Fig. 8).

The expected impact of climate change seemed to be a consistent decrease in all flow quantiles (Table 7). The decrease was between −1 and −8% in both versions. Variability was again large compared to the expected change.

3.2 Gürbe

3.2.1 Time-series approach

The performance of the CRRM was different in the Gürbe sites. In Belp, model performance was almost as good as in Mönchaltorf. However, the upper subcatchment above Burgistein had diverse problems. The complexity of alpine hydrology could not be completely captured by the simple CRRM despite the dedicated parameter set for the uppermost model unit. This caused a huge negative bias for Q95 at Burgistein, already in stage 1. Q50 was nicely reproduced, but Q5 was underestimated again.

Although the most complex weather generation procedure was applied for the Gürbe catchment, stage 2 dominated the quantile offsets. For variability the picture was different, the major source was the future weather uncertainty at both sites and for all quantiles.
The expected impact of climate change was a larger and a subtle decrease for Q95 and Q50, respectively, while Q5 was predicted to increase by 4–5 % at both gauging stations (Table 7). The relative uncertainty of these predictions varied between 2–4 % of the observed flow quantiles, which surprisingly suggests a stronger confidence despite the inferior CRRM performance (Table 7).

### 3.2.2 Quantile approach

With the K1 approach, the observed flow quantiles were almost perfectly matched for the Belp data, while Q50 and Q5 were overestimated by about 20 % in Burgistein. Weather generation (stage 2) meant a negative offset for each quantile at both gauging stations. The expected impact of climate change was quite similar to those from TS, Q95 and Q50 should decrease by 3 to 14 %, Q5 should increase by 2–5 % (Table 7). In contrast to TS, the variability of flow quantiles was much higher for Burgistein (Table 7). The poor performance of the CRRM for the alpine subcatchment resulted in high predictive uncertainty (29–82 % standard deviation relative to the observed quantities) for the flow quantiles already at stage 1. This was propagated through the entire workflow, which finally rendered the predictions for this site extremely unreliable (Table 7). As a result, future weather uncertainty could be considered to be responsible for most of the variability at Belp, but the offsets at Belp and the total uncertainty at Burgistein were dominated by the already existing uncertainty sources (meteorological and hydrological uncertainty, weather generation uncertainty).

In contrast to the Mönchaltorfer Aa case, the results for the Gürbe catchment with K2 yielded somewhat different climate change impacts compared to K1. Bias removal for stage 2 worked well again (Fig. 9), but the sign of expected change shifted for Q5 in both sites.

The performance difference between Belp and Burgistein and the variable relative importance of uncertainty components was seemingly contradicting with the fact that Burgistein covered a significant upstream subcatchment of the Belp gauge. Thus the two sites should have reflected comparable characteristics. The explanation for this
behaviour rooted in the different ability of the CRRM to simulate the observed flow time-series or quantiles. The inferior performance at Burgistein meant that already the uncertainty at stage 1 was much higher than for the Belp site. This elevated uncertainty was then propagated through the remaining stages of the workflow.

3.3 Comparison of uncertainty assessment approaches

The expected hydrological impacts of climate change were similar but not identical with the three uncertainty assessment approaches. The uncertainty of predicted change was always high compared to the mean predicted change, so few percent differences can be regarded as negligible. At the same time, the final uncertainty of the results and the accumulation of uncertainty at the separate workflow stages differed by between the approaches. The closest point to a full consensus between the uncertainty assessment approaches was at the single gauging site of the Mönchaltorfer Aa catchment, where the final uncertainty intervals were almost as similar as the expected impacts (Table 7). In the Gürbe catchment, where the performance of the CRRM was worse, the Belp site had less similar uncertainty intervals, but the standard deviations of the results were at least in the same order of magnitude. In Burgistein TS seemed to underestimate the variability of the climate change impact compared to K1 and K2. The poor performance of the CRRM suggests that the Burgistein predictions should have weaker confidence, yet this was only reflected in the results of the quantile approach, but not in TS (Table 7).

4 Discussion

4.1 Climate change assessment based on quantiles of time-series

Although climate change effects on hydrology are generally assessed by some aggregate statistics, the most common approach is to use a hydrological model calibrated
to time-series of observed discharge for this purpose instead of calibrating models directly to the aggregated quantities of interest (e.g. flow quantiles). At first glance, this distinction seems unnecessary: if a model describes time-series of discharge properly, it is also expected to be a good descriptor of the derived flow quantiles. While this argument holds for a perfect model with no error whatsoever, the situation is more complex if one considers the predictive errors that are always present. The results shown before illustrated that point very clearly.

An additive stochastic time-series error – regardless whether it is an independent noise or an autoregressive process – automatically increases the variance of the CRRM model output, to which it is added to. Consequently, the simulated flow quantiles will spread outwards (Fig. 10), low flow quantiles will become lower, high flow quantiles will become higher. This has a profound effect in a time-series approach: if we account for the increasing non-observational uncertainty with an autoregressive bias term, it is guaranteed that the predictive flow quantiles will be more extreme there than in the calibration phase. This means that extreme events seem to be more likely due to our weaker knowledge about the future (compared to the past) without any change in the climate or hydrology.

However, while the increase in variance and the corresponding effect on the flow quantiles sounds obvious, it is more difficult to recognise the effect in the study outcome. Quantiles get biased due to the error addition, but at the same time their variance does not increase so much that their uncertainty interval would still encompass the original value. As a result the analyst must face some strongly biased but seemingly confident estimations on altered flow quantiles purely because of existing uncertainty.

The inevitability of biased quantiles in TS suggests that based on the change in flow quantiles alone one cannot unambiguously distinguish between the true impacts of climate change and uncertainty propagation in this approach – unless predictive uncertainty was negligible.
4.2 Interpretation of uncertain flow quantiles in K1 and K2

Quantile approaches could circumvent these inherent problems of the TS procedure. The elimination of the significant quantile bias of the additive time-series error models is a true improvement over TS. Furthermore, K2 corrects for the bias introduced by weather generation too. Boorman and Sefton (1997) list studies that used flow quantiles to derive the impact of climate change. Q95 was often used to describe low flow (Arnell, 1992; Wilby et al., 1994; Arnell and Reynard, 1996). Q5 was later also used to characterize flood levels (Gosling et al., 2011). We followed along these lines by implementing the two quantile approaches K1 and K2.

However, the use of a quantile approach also comes at some costs. The approximate likelihood applied in K1 and K2 does not explicitly make any assumptions about the sources and properties of uncertainty, so in this regard it is fundamentally different from the error model applied in TS. The use and definition of quantile uncertainty implies several limitations on the interpretation of results:

– Even if we used the same CRRM as for TS, in K1 and K2 the flow time-series are only intermediate products necessary to calculate the flow quantiles. Due to the possibly huge timing errors they cannot be used to derive any additional indicators that involve timing (for example the distribution of the length of baseflow periods). The quantile calculation procedure can be considered as an additional abstraction layer between the CRRM and the likelihood calculation, which renders the entire CRRM and its parameters somewhat more empirical.

– In K1 and K2 we do not know what mechanisms stand behind the uncertainty of flow quantiles. In contrast to this, TS defined observation, structural and input-related uncertainty and the posterior parameters of the error model could be used to specify the relative importance of these sources.

– With the simple definition of quantile uncertainty we assume that the uncertainty generation mechanisms are the same for the calibration and the predictive period.
This conflicts with our intention to calculate the quantiles of the true discharge for the prediction period without the observation error of the past. Nevertheless, the (random) observation uncertainty of (non-extreme) flow quantiles is very low for long discharge time-series so this theoretical limitation usually would not cause any practical problem.

Considering these pros and cons the different versions of the quantile approach provide a more empirical but viable alternative for uncertainty assessment in cases when flow quantiles are the only targets of the modelling exercise.

4.3 General aspects of impact assessment procedures

The climate change impact assessments procedure as used in this work relies on a complex procedure consisting of different steps as it is common in this field (Blöschl and Montanari, 2010; Todd et al., 2011). The current status of our prediction models does not allow for making hydrological predictions in a simple like feeding GCM output directly into a calibration-free hydrological model (Ehret et al., 2012). Today’s climate models are unable to simulate the present and thus the future hydrologic drivers without a significant bias (Xu, 1999) and conceptual rainfall-runoff models usually need a site-specific calibration (Blöschl and Montanari, 2010). All these required steps introduce uncertainty in the overall assessment procedure. In this article, we have tried to directly address some of these sources of uncertainty by either quantifying them explicitly in the TS or by avoiding some of them by directly calibrating the reference state model to the quantities of interest (i.e. flow statistics instead of time-series).

Despite this explicit treatment of sources of uncertainty one has to consider there remain several decisive pragmatic assumptions that could not be avoided:

– with the bias correction of GCM or RCM outputs we assume that the bias of the climatic model will stay invariant regardless the climatic change;

– the involvement of downscaling methods assumes that despite the inability of present climatic models to simulate small-scale and dynamic features of the
weather we trust that the relationships between local-scale phenomena and regional aggregated weather patterns will be the same in the future;

- the application of calibrated rainfall-runoff models relies on the temporal and climatic invariance of hydrologic model parameters, including their covariance structure.

Each of these assumptions have been refuted at least once based on scientific reasoning or evidence. Bias correction of climate model outputs ruins the physical consistency of climate models and can introduce arbitrary but significant changes into the meteorological forcing (Ehret et al., 2012). Downscaling is usually used to produce localised and often high resolution precipitation series that ought to drive the rainfall runoff models, but it can be simply considered as a rather speculative extrapolation that relies on the present extreme statistics and the biased, large-scale precipitation output of GCMs or RCMs (Blöschl and Montanari, 2010). Despite their definition, rainfall-runoff model parameters tend to vary in the very same catchment with time (Reichert and Mieleitner, 2009), season (Yang et al., 2007), climate (Merz et al., 2011) or just the internal state of the catchment (Romanowicz et al., 2006).

These problems together make the standard climate impact assessment error-prone and increase the uncertainty of results beyond what we have presented above. However, these errors typically result in a biased prediction instead of higher predictive variability, and thus are difficult to identify. Some of these pitfalls can be avoided by carrying out a step-by-step procedure as presented above (bias introduced by hydrological models and weather generation), but some major uncertainty sources will still remain outside the scope of hydrological impact assessment studies (bias of GCMs and RCMs).
5 Conclusions

Our study revealed that the naïve comparison of today’s observed flow quantiles to modelled flow quantiles under climate change with calibration to historic discharge time-series may lead to erroneous conclusions about the effects of climate change. The uncertainties that go with the different steps of the assessment procedure cause a divergence of the flow quantiles and may also introduce bias that is independent on any climate change effects. Hence, it is crucially important to make sure that effects on flow quantiles in a climate change assessment are actually due to the predicted change in climate and not caused by uncertainties related to other aspects of the assessment procedure including the structural uncertainty of the hydrological model itself. Interestingly, this important source of quantile bias was rarely mentioned in similar studies.

When only considering the effects of climate change by i.e. directly calibrating to flow quantiles with simulated weather data or by only considering the changes in the last step of the TS approach, our results delivered typical findings with regard to climate change impacts. The average impact signal was found to be very weak compared to the total uncertainty of future discharge predictions in both of our test catchments for all flow indices. A change of few percents was typically coupled with up to a few 10% of uncertainty, so for most sites and flow indices we could not even be sure about the sign of change. Irrespective of uncertainty assessment method and flow quantile, results suggest that in the future flow conditions may develop into quite different directions.

The results presented here showed that calibrating a CRRM to different quantities of interest (e.g. time-series of discharge versus flow quantiles) may result in slightly different parameterizations. Although a CRRM may predict reasonable discharge series even when only calibrated to flow quantiles where all timing information is lost, the differences in parameterization may induce relevant biases on the non-calibrated quantities. In a sense, this procedure degraded the hydrological model to a semi-empirical albeit rather complex mathematical function. There was no guarantee that the simulated discharge time-series or the model parameters had any connection with the true
physical quantities they originally referred to. This also demonstrated that one could arrive to rather different conclusions about the source, structure and composition of predictive uncertainty for the same hydrological model and calibration data when considering different objective functions for calibration.

On one hand this means that we can only make conditional statements about these internal details of uncertainty. On the other hand the robustness of total predictive uncertainty for the Mönchaltorf and Belp sites (where the hydrological model performance was good) indicates that the suitability of different uncertainty assessment procedures for different purposes (TS for timing-sensitive applications, K2 for flow quantiles) can be the major selection criterium between uncertainty assessment methods.

Appendix A

Conditional precipitation model

The task was to produce daily precipitation sums for the alpine subcatchment of the Gürbe based on the high frequency generated weather for the lowland part. We determined the joint probabilities of having wet/dry day in Blumenstein based on the weather in Belp. The analysis revealed that the elevated annual precipitation amount in the alpine zone comes with more wet days too (Table 5). The conditional prediction routine assumed that these probabilities are influenced mostly by the topographical differences and will remain constant regardless of the future climatic changes. During prediction we first decided about the wetness status of the alpine subcatchment considering the wetness status of the lowland and the given conditional probability.

For days which were wet in both parts of the catchment we applied a standard back-propagation neural network (R package neuralnet) with 5 nodes in each of its 2 hidden layers to predict the daily alpine precipitation from the daily values of precipitation, mean air temperature, mean global radiation, mean vapour pressure and mean wind speed of Belp. The residuals between the predicted and observed precipitation sums at
Blumenstein had 0 mean and were closely normally distributed after a Box-Cox power transform with $\lambda = 0.5$. So the final prediction method for homogeneously wet days was the application of the black-box model and the addition of a back-transformed normal noise term.

For days which were dry in the lowland but wet in the mountains we have drawn from a $\Gamma$ distribution, which was parameterized based on the historical precipitation data from such days. The choice of the $\Gamma$ distribution was motivated by the fact that this distribution type was used for drawing precipitation intensities in the Neyman–Scott rectangular pulses model which generated the rainfall for the UKCP09 weather generator.

Appendix B

An approximate likelihood function for flow quantiles

B1 Calculating the approximate likelihood

We denote set of observed flow quantiles with $q^o$, with the individual items indexed as $q^o_i$. The modelled flow quantiles ($q(\theta)$) depend on the model parameters ($\theta$) and are extracted from the modelled flow time-series. $F(y|\theta)$ is the distribution of the modelled discharge given the model parameters $\theta$, and $f(y|\theta)$ is the corresponding probability density function.

We assume that the quantiles are subject to an additive error $E$ so that

$$q^o = q(\theta) + E. \quad (B1)$$

If we consider samples from a distribution $E$ is approximately normally distributed with 0 mean and a variance depending on $\alpha$ and $f(y|\theta)$ (Sect. 10.10 in Stuart and Ord, 1994):
\[ \sigma^2_\alpha = \frac{1}{n} \alpha (1 - \alpha) \frac{1}{f(q_\alpha(\theta))^2} \]  

(B2)

where \( n \) is the sample size of \( y \). The error of this variance estimator is \( O\left(\frac{1}{n}\right) \).

This theorem applies to independent samples, \( \sigma^2_\alpha \) is underestimated for an autocorrelated sample. To make the theorem applicable for the quantile likelihood calculation of discharge series we make the following additional assumptions:

1. The correlation length of \( y \) is orders of magnitudes shorter than the length of the observation period (few days vs. many years), so \( q(\theta) \) can be regarded as a set of quantiles from an independent sample.

2. We replace \( n \) with \( n^* \) that refers to an effective sample size with \( 1 \leq n^* \leq n \). The new parameter characterizes the quality of fit to the observations. If \( n = n^* \) then the uncertainty of the modelled flow quantiles equals to the sampling uncertainty, so the uncertainty added by the model is negligible. If \( n^* \ll n \) then \( E \) is dominated by model-related uncertainty. The inverse of \( n^* \) can be considered as an uncertainty scaling parameter.

The resulting likelihood function is then

\[ p\left(q^0_\alpha|\theta\right) = N \left[ \mu = q_\alpha(\theta), \ \sigma^2_\alpha = \frac{1}{n^*} \alpha (1 - \alpha) \frac{1}{f(q_\alpha(\theta))^2} \right]. \]  

(B3)

We calculate \( f(q_\alpha(\theta)) \) from \( y \) with a kernel density estimator. The kernel distribution is normal, the standard deviation of the kernel is set to

\[ \sigma_k = 0.79 n^{-0.2} \left( q_{0.25}(y|\theta) - q_{0.75}(y|\theta) \right). \]  

(B4)
B2  Sampling from the predictive distribution

A predictive realization of the flow quantiles can be computed in two steps. First one needs to calculate \( q(\theta) \) with the deterministic rainfall-runoff model based on the model parameters and the input series for the prediction period. In the second step we need to draw a realization from the uncertainty distribution. To do this we need to calculate the probability density function for the modelled flow. Then we draw the predicted observation quantiles with Eq. (B3).

Acknowledgements. The iWaQa project is financed by the Swiss National Science Foundation (National Research Program #61 on Sustainable Water Management, Grant #406140-125866) and the Swiss federal Office for the Environment. The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged. We are grateful to Simone Fatichi and Pierluigi Calanca for their help with the stochastic downscaling of climatic data.

References


Arnell, N. W. and Reynard, N. S.: The effects of climate change due to global warming on river flows in Great Britain, J. Hydrol., 183, 397–424, 1996. 521


Bergström, S.: Parametervärden för HBV modellen i Sverige, SMHI Hydrologi, Norrköping, Sweden, 36 pp., 1990. 538


BLW: Soil suitability map (Bodeneignungskarte) of Switzerland, Bundesamt für Landwirtschaft, Bern, Switzerland, 2008. 507


Cameron, D.: An application of the UKCIP02 climate change scenarios to flood estimation by continuous simulation for a gauged catchment in the northeast of Scotland, UK (with uncertainty), J. Hydrol., 328, 212–226, 2006. 505


Khan, M., Coulibaly, P., and Dibike, Y.: Uncertainty analysis of statistical downscaling methods, J. Hydrol., 319, 357–382, 2006. 504


Maurer, E.: Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emission scenarios, Climatic Change, 82, 309–325, 2007. 505


Swisstopo: Digital land-use map of Switzerland, Bundesamt für Landestopographie, Wabern, Switzerland, 2008. 507

### Table 1. Process Matrix of the CRRM

<table>
<thead>
<tr>
<th>Process</th>
<th>Symbol</th>
<th>Rate</th>
<th>Affected storage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$h_{\text{snow}}$</td>
</tr>
<tr>
<td>snowfall</td>
<td>$P_{\text{snow}}$</td>
<td>( \begin{cases} P &amp; \text{if } T &lt; T_{\text{crit}} \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td>+</td>
</tr>
<tr>
<td>snowmelt</td>
<td>$P_{\text{melt}}$</td>
<td>( \begin{cases} k_{\text{melt}} (T - T_{\text{melt}}) &amp; \text{if } T &gt; T_{\text{melt}} \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td>-</td>
</tr>
<tr>
<td>rainfall</td>
<td>$P_{\text{rain}}$</td>
<td>( \begin{cases} P &amp; \text{if } T \geq T_{\text{crit}} \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td>+</td>
</tr>
<tr>
<td>throughfall</td>
<td>$P_{\text{through}}$</td>
<td>( (1 - k_{\text{capt}}) (P_{\text{melt}} + P_{\text{rain}}) )</td>
<td>-</td>
</tr>
<tr>
<td>canopy evapotr.</td>
<td>$E_{\text{c}}$</td>
<td>$k_{\text{ce}} E_{\text{pot}} \begin{cases} \frac{h_{c}}{h_{\text{c,sat}}} &amp; \text{if } h_{c} \leq h_{\text{c,sat}} \ 1 &amp; \text{otherwise} \end{cases}$</td>
<td>-</td>
</tr>
<tr>
<td>dripping</td>
<td>$P_{\text{drip}}$</td>
<td>$k_{\text{drip}} \begin{cases} h_{c} - h_{\text{c,sat}} &amp; \text{if } h_{c} \geq h_{\text{c,sat}} \ 0 &amp; \text{otherwise} \end{cases}$</td>
<td>-</td>
</tr>
<tr>
<td>ground evapotr.</td>
<td>$E$</td>
<td>$E_{\text{pot}} f_{\text{et}}$</td>
<td>-</td>
</tr>
<tr>
<td>sat. excess runoff</td>
<td>$Q_{r}$</td>
<td>$f_{\text{sat}} (P_{\text{drip}} + P_{\text{through}})$</td>
<td>-</td>
</tr>
<tr>
<td>groundwater rech.</td>
<td>$Q_{\text{rge}}$</td>
<td>$k_{\text{rge}} f_{\text{sat}} q_{\text{seep}}$</td>
<td>-</td>
</tr>
<tr>
<td>subsurface flow</td>
<td>$Q_{\text{ssf}}$</td>
<td>( (1 - k_{\text{rge}}) f_{\text{sat}} q_{\text{seep}} )</td>
<td>-</td>
</tr>
<tr>
<td>baseflow</td>
<td>$Q_{\text{bf}}$</td>
<td>$k_{\text{bf}} h_{gw}$</td>
<td>-</td>
</tr>
<tr>
<td>paved evapor.</td>
<td>$E_{p}$</td>
<td>$k_{\text{p,e}} E_{\text{pot}} \frac{h_{p}}{h_{\text{p,e}} + h_{\text{p,seep}}}$</td>
<td>-</td>
</tr>
<tr>
<td>paved runoff</td>
<td>$Q_{p}$</td>
<td>( k_{p,r} (h_{p} - h_{p,\text{sat}}) ) if $h_{p} \geq h_{p,\text{sat}} )</td>
<td>-</td>
</tr>
<tr>
<td>stream discharge</td>
<td>$Q_{q}$</td>
<td>$k_{q} h_{q}$</td>
<td>-</td>
</tr>
</tbody>
</table>

* All storages are in [mm] while processes are in [mm d$^{-1}$]. $f_{\text{sat}}$ and $f_{\text{et}}$ are defined in Eqs. (1) and (2), respectively.
Table 2. Prior distributions of snow, canopy and paved area parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Reference</th>
<th>Applied distribution*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{crit}}$ [°C]</td>
<td>Critical temperature for snowfall</td>
<td>Kokkonen et al. (2006)</td>
<td>N(1, 0.5)</td>
</tr>
<tr>
<td>$T_{\text{melt}}$ [°C]</td>
<td>Threshold temperature for snowmelt</td>
<td>Kokkonen et al. (2006)</td>
<td>N(0, 1)</td>
</tr>
<tr>
<td>$k_{\text{melt}}$ [mm°C]</td>
<td>Temperature-specific snowmelt rate constant</td>
<td>Bergström (1990)</td>
<td>LN(3, 1.2)</td>
</tr>
<tr>
<td>$k_{\text{capt}}$ [-]</td>
<td>Precipitation capturing efficiency of fully developed canopy</td>
<td>Vrugt et al. (2003)</td>
<td>B(0.7, 0.15)</td>
</tr>
<tr>
<td>$k_{\text{e,c}}$ [-]</td>
<td>Evaporation multiplier of canopy</td>
<td>Vrugt et al. (2003)</td>
<td>LN(1, 0.1) grass</td>
</tr>
<tr>
<td>$k_{\text{drip}}$ [d$^{-1}$]</td>
<td>Dripping rate from canopy storage</td>
<td>Vrugt et al. (2003)</td>
<td>δ(400)</td>
</tr>
<tr>
<td>$h_{\text{c,sat}}$ [mm]</td>
<td>Storage in fully wetted canopy</td>
<td>Rutter et al. (1971)</td>
<td>LN(1.2, 0.2) forest</td>
</tr>
<tr>
<td>$k_{\text{LAI}}_{\text{min}}$ [-]</td>
<td>Relative winter leaf area index</td>
<td>estimation</td>
<td>B(0.1, 0.05) grass</td>
</tr>
</tbody>
</table>

*a Values are represented by range (min–max) or mean±standard deviation. b Distribution types: N(µ, σ): normal, LN(µ, σ): lognormal, B(µ, σ): beta, δ(µ): Dirac-delta. µ and σ are the mean and the standard deviation of the distributions, respectively.

538
Table 3. Prior distributions of paved area parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Reference</th>
<th>Applied distribution&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_{p,\text{sat}} ) [mm]</td>
<td>Paved area storage</td>
<td>Falk and Niemczynowicz (1979)</td>
<td>LN(1, 0.3)</td>
</tr>
<tr>
<td>0.1–1.1</td>
<td>Kidd (1978)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1–1.5</td>
<td>Heaney et al. (1976)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>Arnell (1982)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4–0.7</td>
<td>estimation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( k_{p,r} \) [d<sup>−1</sup>] Paved area runoff rate LN(20, 1)

<sup>a</sup> Values are represented by range (min–max). <sup>b</sup> Distribution types: LN(\( \mu \), \( \sigma \)): lognormal. \( \mu \) and \( \sigma \) are the mean and the standard deviation of the distribution, respectively.
Table 4. Prior distributions of catchment and stream parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Reference</th>
<th>Applied distribution$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{FS}$</td>
<td>Catchment-scale equivalent of full saturation ($f_{sat} = 98%$)</td>
<td>Schaar et al. (2001)$^c$</td>
<td>LN(430, 20) for loamy soils</td>
</tr>
<tr>
<td>[mm]</td>
<td>387–440 for clay-loam/sandy loam</td>
<td>Schaar et al. (2001)$^c$</td>
<td>LN(382, 3) for sandy soils</td>
</tr>
<tr>
<td></td>
<td>390–430 for clay-loam/sandy loam</td>
<td>Carsel and Parrish (1988)$^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>375–390 for sand/loamy sand</td>
<td>Schaar et al. (2001)$^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>440–490 for silt/silt-loam/silt-clay-loam</td>
<td>Schaar et al. (2001)$^c$</td>
<td></td>
</tr>
<tr>
<td>$h_{FC}$</td>
<td>Catchment-scale equivalent of field capacity ($f_{sat} = 2%$)</td>
<td>Schaar et al. (2001)$^c$</td>
<td>LN(220, 25) for loamy soils</td>
</tr>
<tr>
<td>[mm]</td>
<td>168–255 for clay-loam/sandy loam</td>
<td>Schaar et al. (2001)$^c$</td>
<td>LN(75, 9) for sandy soils</td>
</tr>
<tr>
<td></td>
<td>85–270 for clay-loam/sandy loam</td>
<td>Carsel and Parrish (1988)$^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55–105 for sand/loamy sand</td>
<td>Schaar et al. (2001)$^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>280–305 for silt/silt-loam/silt-clay-loam</td>
<td>Schaar et al. (2001)$^c$</td>
<td></td>
</tr>
<tr>
<td>$h_{WP}$</td>
<td>Catchment-scale equivalent of wilting point ($E_{wilting} = 5%$ of $E_{pot}$)</td>
<td>Schaar et al. (2001)$^c$</td>
<td>LN(90, 10) for loamy soils</td>
</tr>
<tr>
<td>[mm]</td>
<td>60–150 for clay-loam/sandy loam</td>
<td>Schaar et al. (2001)$^c$</td>
<td>LN(52.5, 1) for sandy soils</td>
</tr>
<tr>
<td></td>
<td>65–150 for clay-loam/sandy loam</td>
<td>Carsel and Parrish (1988)$^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52–53 for sand/loamy sand</td>
<td>Schaar et al. (2001)$^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70–120 for silt/silt-loam/silt-clay-loam</td>
<td>Schaar et al. (2001)$^c$</td>
<td></td>
</tr>
<tr>
<td>$k_{rge}$</td>
<td>Proportion of groundwater recharge from seepage</td>
<td>B(0.7, 0.1)</td>
<td></td>
</tr>
<tr>
<td>[–]</td>
<td>40–90 %</td>
<td>estimation</td>
<td></td>
</tr>
<tr>
<td>$q_{seep}^\text{max}$</td>
<td>Maximal seepage rate</td>
<td>LN(100, 50)</td>
<td></td>
</tr>
<tr>
<td>[mm d$^{-1}$]</td>
<td>50–200</td>
<td>estimation</td>
<td></td>
</tr>
<tr>
<td>$k_{bf}$</td>
<td>Baseflow constant</td>
<td>LN(0.0005, 0.0005)</td>
<td></td>
</tr>
<tr>
<td>[d$^{-1}$]</td>
<td>$10^{-5}$–$10^{-3}$</td>
<td>estimation$^d$</td>
<td></td>
</tr>
<tr>
<td>$k_q$</td>
<td>Stream constant</td>
<td>LN(10, 5)</td>
<td></td>
</tr>
<tr>
<td>[d$^{-1}$]</td>
<td>3–30</td>
<td>estimation</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Values are represented by range (min–max). $^b$ Distribution types: LN($\mu$, $\sigma$): lognormal, B($\mu$, $\sigma$): beta. $\mu$ and $\sigma$ are the mean and the standard deviation of the distribution, respectively. $^c$ Full saturation (FS), field capacity (FC) or wilting point (WP) moisture content of homogenous soils of the given type with 1 m thickness. $^d$ Groundwater residence time is estimated to be between 180 days and about 30 yr.
**Table 5.** Probability of wet ($W$) and dry ($D$) days in the upper and lower parts of the Gürbe catchment.

<table>
<thead>
<tr>
<th></th>
<th>$P_{Belp} = W$</th>
<th>$P_{Belp} = D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{Blumenstein} = W$</td>
<td>43.2 %</td>
<td>14.5 %</td>
</tr>
<tr>
<td>$P_{Blumenstein} = D$</td>
<td>4.2 %</td>
<td>38.0 %</td>
</tr>
<tr>
<td>Institution</td>
<td>Code</td>
<td>GCM</td>
</tr>
<tr>
<td>---------------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>CNRM</td>
<td>cnrm</td>
<td>Arpege</td>
</tr>
<tr>
<td>DMI</td>
<td>dmi</td>
<td>ECHAM5</td>
</tr>
<tr>
<td>ETHZ</td>
<td>ethz</td>
<td>HadCM3Q0(^a)</td>
</tr>
<tr>
<td>ICTP</td>
<td>ictp</td>
<td>ECHAM5</td>
</tr>
<tr>
<td>KNMI</td>
<td>knmi</td>
<td>ECHAM5</td>
</tr>
<tr>
<td>MetOffice-HC</td>
<td>hadley</td>
<td>HadCM3Q0(^a)</td>
</tr>
<tr>
<td>MPI</td>
<td>mpi</td>
<td>ECHAM5</td>
</tr>
<tr>
<td>SMHI</td>
<td>smhibcm</td>
<td>BCM</td>
</tr>
<tr>
<td></td>
<td>smhi_echam</td>
<td>ECHAM5</td>
</tr>
<tr>
<td></td>
<td>smhi_had</td>
<td>HadCM3Q3(^b)</td>
</tr>
</tbody>
</table>

\(^a\) Normal sensitivity, \(^b\) low sensitivity
Table 7. Relative changes in flow indices* due to climate change with different uncertainty assessment approaches.

<table>
<thead>
<tr>
<th>Flow index</th>
<th>TS</th>
<th>K1</th>
<th>K2</th>
<th>Sign consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mönchaltorf</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q95</td>
<td>+2 ± 9</td>
<td>−1 ± 11</td>
<td>−7 ± 12</td>
<td></td>
</tr>
<tr>
<td>Q50</td>
<td>−4 ± 8</td>
<td>−7 ± 10</td>
<td>−7 ± 9</td>
<td>✓</td>
</tr>
<tr>
<td>Q5</td>
<td>−4 ± 11</td>
<td>−8 ± 9</td>
<td>−8 ± 9</td>
<td>✓</td>
</tr>
<tr>
<td>Belp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q95</td>
<td>−15 ± 4</td>
<td>−14 ± 6</td>
<td>−4 ± 7</td>
<td>✓</td>
</tr>
<tr>
<td>Q50</td>
<td>−1 ± 4</td>
<td>−3 ± 6</td>
<td>−4 ± 5</td>
<td>✓</td>
</tr>
<tr>
<td>Q5</td>
<td>+5 ± 3</td>
<td>+5 ± 5</td>
<td>−5 ± 3</td>
<td></td>
</tr>
<tr>
<td>Burgistein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q95</td>
<td>−15 ± 4</td>
<td>−9 ± 38</td>
<td>−4 ± 45</td>
<td>✓</td>
</tr>
<tr>
<td>Q50</td>
<td>−1 ± 3</td>
<td>−3 ± 29</td>
<td>−4 ± 26</td>
<td>✓</td>
</tr>
<tr>
<td>Q5</td>
<td>+4 ± 2</td>
<td>+2 ± 82</td>
<td>−4 ± 80</td>
<td></td>
</tr>
</tbody>
</table>

* Changes are in [%] of the observed flow index ± standard deviation.
Fig. 1. The catchments and gauging sites (triangles) of this study and their locations in Switzerland.
Fig. 2. Schematic structure of the extended LogSPM model.
**Fig. 3.** Scheme of workflow stages for different uncertainty assessment approaches. White stars indicate stages of calibration.
Fig. 4. Modelled flow plotted against the observations in different workflow stages (Mönchaltorf, TS approach).
Fig. 5. Flow index predictions for Mönchaltorf with the TS approach. Predictions for the present (left side) were made using generated precipitation (stage 2). SLS: simple least squares calibration (for reference), BIAS: the Bayesian error model of TS. The right panel shows future predictions for the individual model chains. Future uncertainty is the joint prediction from all 10 model chains.
Fig. 6. Absolute (top panel) and relative (bottom panel) changes in flow indices during the workflow stages (Mönchaltorf, TS approach).
Fig. 7. Absolute (top panel) and relative (bottom panel) changes in flow indices during the workflow stages (Mönchaltorf, K1 approach).
Fig. 8. Absolute (top panel) and relative (bottom panel) changes in flow indices during the workflow stages (Mönchaltorf, K2 approach).
Fig. 9. Absolute (top panel) and relative (bottom panel) changes in flow indices during the workflow stages (Gürbe at Belp, K2 approach).
Fig. 10. A synthetic example for the different effect of increasing time-series and quantile uncertainty on the CDF of $Y + E$. $Y$ is a standard normal distribution, $E_{ts}$ is an i.i.d. normal time-series error and $E_q$ is a normal quantile error. The uncertainty intervals were derived from 2000 realisations, the time-series length was also 2000.