**Nonlinear Processes
in Geophysics**

Conditioning model output statistics of regional climate model precipitation on circulation patterns

F. Wetterhall[1,2]**, F. Pappenberger**[1,3]**, Y. He**[4]**, J. Freer**[5]**, and H. L. Cloke**[1,6]

[1]European Centre for Medium Range Weather Forecasts, Reading, UK
[2]Department of Geography, King's College London, London, UK
[3]Hohai University, Nanjing, China
[4]Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, UK
[5]School of Geographical Sciences, University of Bristol, UK
[6]Department of Geography & Environmental Science and Department of Meteorology, University of Reading, UK

Correspondence to: F. Wetterhall (fredrik.wetterhall@ecmwf.int)

Abstract. Dynamical downscaling of Global Climate Models (GCMs) through regional climate models (RCMs) potentially improves the usability of the output for hydrological impact studies. However, a further downscaling or interpolation of precipitation from RCMs is often needed to match the precipitation characteristics at the local scale. This study analysed three Model Output Statistics (MOS) techniques to adjust RCM precipitation; (1) a simple direct method (DM), (2) quantile-quantile mapping (QM) and (3) a distribution-based scaling (DBS) approach. The modelled precipitation was daily means from 16 RCMs driven by ERA40 reanalysis data over the 1961–2000 provided by the ENSEMBLES (ENSEMBLE-based Predictions of Climate Changes and their Impacts) project over a small catchment located in the Midlands, UK. All methods were conditioned on the entire time series, separate months and using an objective classification of Lamb's weather types. The performance of the MOS techniques were assessed regarding temporal and spatial characteristics of the precipitation fields, as well as modelled runoff using the HBV rainfall-runoff model. The results indicate that the DBS conditioned on classification patterns performed better than the other methods, however an ensemble approach in terms of both climate models and downscaling methods is recommended to account for uncertainties in the MOS methods.

1 Introduction

Global climate models (GCMs) are currently the best tools to model changes in the global climate caused by an increase in radiatively active gases (IPCC, 2007). However, despite constant improvements in model resolution and the description of the physical processes, modelling of precipitation in the current model versions is still inadequate for use in most, if not all, local impact studies (Leith and Chandler, 2010; Beven, 2011). There are many reasons for this, ranging from the difference in resolution between the GCM and the local hydrological scale, numerical implementation and the parameterisation of physical processes on the sub-grid scale (Schmidli et al., 2006).

Common approaches used in order to overcome this problem is the application of downscaling techniques, which are typically either dynamical or statistical (empirical). Statistical downscaling methods usually establish a statistical relationship between a local variable (predictand) and a larger-scale variable modelled by the global or regional model (predictor). There are a vast number of statistical downscaling techniques used in impact studies, and for recent overviews we refer to Fowler et al. (2007) and Maraun et al. (2010). Dynamical downscaling, where a regional climate model (RCM) is forced with boundary conditions from a GCM, includes many feedback processes which are important for the energy, radiation and water balances. These models are however very expensive to run both in terms of time and resources, and the ability to investigate the full

model uncertainty is problematic. Projects like PRUDENCE (Christensen and Christensen, 2007) and ENSEMBLES (van der Linden and Mitchell, 2009) have enabled ensemble output from RCMs to be easily available to the larger research community. However, even after dynamical downscaling, the RCM output is not directly suited for use in studies of climate impact on hydrology, ecology, and agriculture. This is especially the case for precipitation, and amongst the reasons for this is that the resolution in RCMs is currently too coarse to correctly resolve sub-grid processes on the scales that are hydrologically relevant and that relates to convective meteorological processes as well as orographic induced precipitation. In smaller catchments, on the scale of a few RCM grid points and with quick response times, this can potentially lead to an underestimation of high flow situations, regardless of whether the RCM has a wet or dry mean bias over the catchment. Other reasons are smoothing between grid points and inadequate representation of the physical processes in the model code (Cloke et al., 2012). Therefore at present, impact studies that rely on a correct representation of precipitation, such as those dealing with flood risk, cannot usefully use direct RCM output.

A solution to the problem of RCM misrepresentation of precipitation is to pre-process the RCM output through "bias correction" also known as a form of Model Output Statistics (MOS). MOS is the preferred term, because *bias* in statistics usually refers to a deviation in the expected value of the estimator from the true mean of the population (von Storch and Zwiers, 1999), whereas the deviations in the estimator are also found in the higher moments, such as the variance and skewness. The use of MOS can be seen as an additional downscaling step from the RCM, so the term downscaling will be used for this approach as well. MOS can also be applied directly to GCM output (e.g. Widmann et al., 2003; Eden et al., 2012). An example of a MOS technique is to use transfer functions to adjust the distribution of RCM precipitation to match the observed precipitation distribution (e.g. Déqué, 2007; Piani et al., 2009; Yang et al., 2010). However, the RCM may perform differently depending on season or governing atmospheric circulation. A typically wet weather regime (e.g. pattern or season) can have a different precipitation distribution in time and space than a dry regime.

Classification of weather patterns has been widely used in the statistical downscaling community to analyse and model the predictor-predictand relationship (e.g. Stehlik and Bardossy, 2002; Wetterhall et al., 2007). Recently a COST (European Cooperation in Science and Technology) action compiled a large collection of objective and subjective weather type classifications for Europe that are freely available (COST733, Philipp et al., 2011). Conditioning the MOS on months or on classified weather patterns rather than on the full distribution could potentially lead to better estimates of the parameters of the transfer function and therefore a more effective model error correction. However, there is a trade-off in terms of using less available information in the error
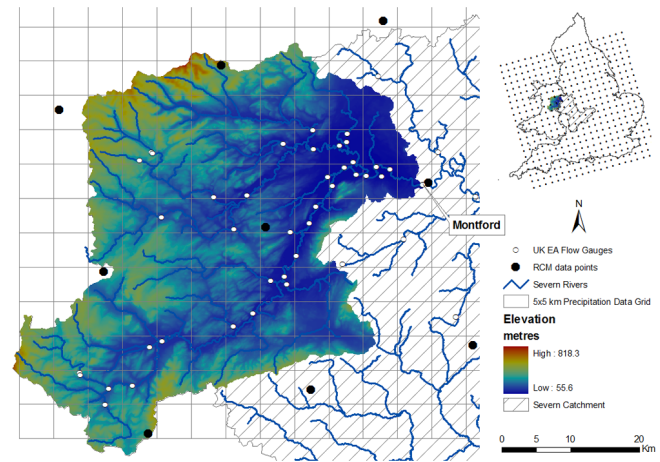
Fig. 1. The Montford catchment and its location in the UK. Large black dots the RCM data and the grid denotes the 5×5 km gridded dataset from the UK MetOffice.

correction, thereby suboptimising the estimation of the parameters.

This paper investigates a number of MOS techniques on RCM precipitation from 16 RCM simulations from the EN-SEMBLES project over a small catchment ($\sim 2000 \, \text{km}^2$) in England. Further, it conditions the DBS and QM method on months and weather patterns to assess the potential added value. Section 2 describes the study area and data used in the study and a description of the methodology, the results and discussions are presented in Sect. 3 and the conclusions and future outlook in Sect. 4.

2 Material and methods

2.1 Study area and data

The study area was the upper part of the River Severn in the Midlands region of England, which flows through the Welsh mountains, passes the town of Shrewsbury and then continues southwards towards the Bristol Channel (Fig. 1). Flow levels are generally high in autumn and winter and low in summer. The area is prone to flooding and has recently been a study area for hydrological ensemble forecasting (He et al., 2009; Wetterhall et al., 2011; Cloke et al., 2012).

The observed precipitation and temperature data series used in this study consisted of daily 5×5 km grids interpolated from station data covering the period 1 January 1960–31 December 2006 provided by the UK Met Office. The interpolation of the temperature and precipitation was done using inverse-distance weighted interpolation (Perry et al., 2008). The observed discharge data at Montford were provided by the Environment Agency of England and Wales (EA) Midlands region and covered the period 1986–2006.

The RCM precipitation data were provided by the EN-SEMBLES project (van der Linden and Mitchell, 2009). The

RCM models used in this study were all driven by the ERA40 reanalysis over the time period 1 January 1961–31 December 2000 (Uppala et al., 2005). In total, 16 RCMs were available for this study, and one model was run with three different climate sensitivities, yielding 18 model projections available in total. The data is available publicly as daily values on a 0.22 degree grid (Fig. 1). The RCM precipitation was evaluated as an ensemble, equal weights are applied to all members since the aim of the study was to see how well the MOS worked regardless of the bias in the raw RCM. Furthermore, using equal weights enables the evaluation of the effects of MOS rather than ranking the performance of individual models.

2.2 Model output statistics

There are quite a few suggested methods of MOS in the literature (for an overview see Maraun et al., 2010). The type of MOS method most suited for a specific impact study will ultimately depend on the application objective/s and in this case, high river flow events were of most interest to quantify potential flood inducing events. Therefore, two methods that downscale the precipitation distribution were selected, (1) a quantile-quantile mapping (QM) correction as described by Boé et al. (2007) and (2) a distribution-based scaling (DBS) of the modelled and observed precipitation (Yang et al., 2010), but also a simpler factor change method, (3) the direct method (DM; Lenderink et al., 2007), was included as a benchmark. The rationale for choosing the first 2 methods are that they are sensitive to changes in the tails of the distribution and are therefore potentially more useful in hydrological impact studies. The direct method is merely correcting for the mean bias, and does not transfer any information on any modelled shift in the precipitation distribution from the RCM.

All methods were applied to the individual grid points (5×5 km) in the catchment. Since the RCM output resolution was approximately 25×25 km, the RCM modelled precipitation was first interpolated onto the same grid as the observed (5×5 km) using a simple nearest neighbour interpolation. 9 grid points from the RCM output covered the catchment area (Fig. 1).

All the identified MOS techniques were applied over the period where RCM data was available (1961–2000) through a split-sample validation on dry and wet conditions. The dry (wet) sample was constructed by selecting the driest (wettest) months over the training period. The MOS techniques were calibrated over each sample, and then validated over the other, respectively. By using this approach, the entire series can be used in the validation.

2.2.1 The direct method

The direct method (DM) scales the modelled precipitation by multiplying it with the ratio of the observed and

RCM-modelled precipitation:

$$\hat{z}_{\text{mod}}(t, n) = z_{\text{mod}}(t, n) \frac{\bar{Z}_{\text{obs}}(n)}{\bar{Z}_{\text{mod}}(n)} \tag{1}$$

where $\hat{z}_{\text{mod}}$ is the adjusted RCM-modelled precipitation z_{mod} at time t for point n in space, $\bar{Z}$ denotes the mean observed (obs) and modelled (mod) precipitation over calibration period. In this study a split-sample approach was used (see below). The direct method removes biases in the mean but not the coefficient of variance of the modelled precipitation (Lenderink et al., 2007). The direct method is sometimes referred to as the delta method in the scientific literature, since the method was originally developed for temperature, where the difference between modelled future and control scenarios was added to the observed temperature (Maraun et al., 2010).

2.2.2 Quantile-quantile mapping

Quantile-quantile mapping (QM) utilises the empirical cumulative distributions of the observed and modelled precipitation for the downscaling. Therefore, unlike the DMthis method improves any systematic biases throughout the statistical distribution of precipitation differences and hence improves the adjustments of extremes. The method modifies each RCM precipitation event by first calculating the probability of the event. The downscaled precipitation amount is then modelled as the observed precipitation with the same probability (Boé et al., 2007; following the denotation by Themeßl et al., 2010):

$$\hat{z}(t, n) = F_{t,n}^{\text{obs,cal-1}} \left(F_{t,n}^{\text{sim,cal}} \left(z^{\text{sim,val}}(t, n) \right) \right) \tag{2}$$

where $F_{t,n}^{\text{obs,cal-1}}$ denotes the inverse of the empirical cumulative distribution function (ecdf) $F_{t,n}^{\text{obs,cal}}$, $F_{t,n}^{\text{sim,cal}}$ is the inverse ecdf of the observed and simulated precipitation over the calibration period, and $z^{\text{sim,val}}$ is the simulated precipitation over the validation period. Quantile-quantile mapping is restricted to adjusting values that are in the observed range. If the method is to be used in climate impact studies it could be necessary to infer a change factor to values outside the observed range to account for changes in the upper tail of the distribution. However, since only historical records were used in this study, no extrapolation was made.

2.2.3 Distribution-based scaling

The DBS is similar to the QM as it adjusts the precipitation amount by comparing the probability density functions of modelled and observed precipitation. The difference is that a theoretical density function is fitted for the observed and simulated series, respectively, and the probability of an event in the RCM precipitation is used together with the inverse of the cumulative density function from the observed data series to estimate new values. The gamma distribution was
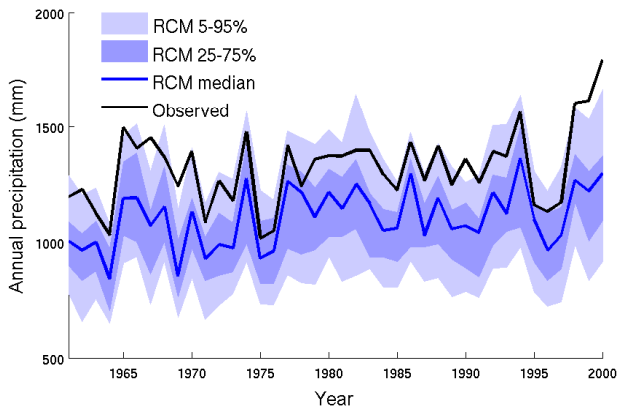
Fig. 2. Mean annual precipitation accumulated at the station in Montford catchment for the 18 RCMs over the study period 1961–2000. The 5–95 percentiles are shown in light blue, and the dark blue denotes the 25–75 percentiles. The solid black line shows the observed precipitation from the UK MetOffice.

selected for this study since it captures the asymmetrical and positively skewed distribution of daily precipitation intensities (Wilks, 1995; Haylock et al., 2006). The probability density function of a gamma-distributed variable x is defined as

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp(-x/\beta)}{\beta\Gamma(\alpha)} x, \alpha, \beta \prec 0 \qquad (3)$$

where α is the shape parameter, β is the scale parameter and $\Gamma(x)$ is the inverse gamma function. The distribution parameters were estimated using maximum likelihood estimation (MLE).

The method was applied in two steps. Firstly, the number of rainy days in the modelled and observed precipitation was compared, and a threshold value was defined in the RCM output to match the number of rainy days in both time series. This procedure is necessary since zero values are not defined by the gamma distribution. All days that had less precipitation than the threshold were considered dry days. In the second step, gamma distributions were fitted for the non-zero observed and modelled precipitation time series and the MOS function was then defined as

$$\hat{z}(t, n) =$$
$$F^{-1}\left(\alpha_{\text{obs,cal}}, \beta_{\text{obs,cal}}, F\left(z_{\text{sim,val}}(t, n), \alpha_{\text{sim,cal}}\beta_{\text{sim,cal}}\right)\right) \quad (4)$$

where obs denotes parameters estimated from observations and sim denotes parameters estimated from the RCM output and F^{-1} the inverse gamma distribution. MOS using the gamma distribution can yield values larger than in the records of observations, so care has to be taken in the upper tail of the distribution as to not obtain unreasonably large values. An upper threshold was explicitly implemented defined as the maximum observed precipitation for each point. Using
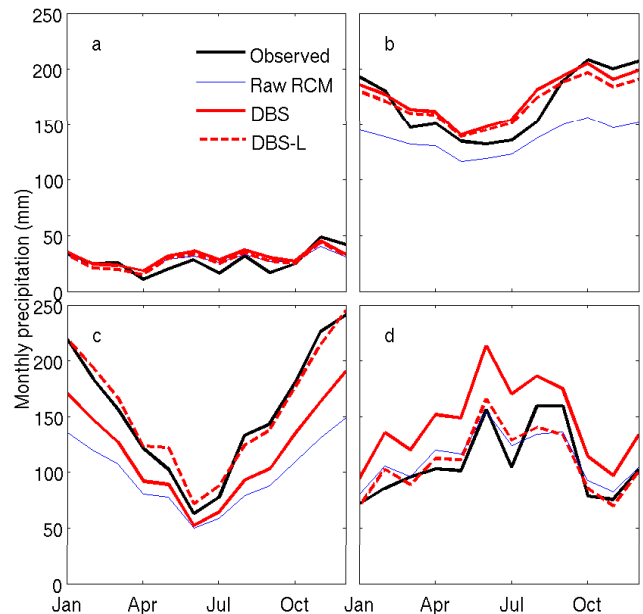
Fig. 3. Monthly precipitation for four classified weather patterns; **(a)** anticyclonic, **(b)** cyclonic, **(c)** westerly winds and **(d)** easterly/south-easterly winds. The figure shows the raw RCM precipitation and MOS using DBS and DBS classified on weather patterns.

a threshold is a rather crude method to avoid unreasonably large precipitation amounts but in this case study where we only were concerned with observed events it was not considered to have a large impact. However, if the method is applied to future simulations, the effect of an upper limit has to be studied in more detail.

The DBS and QM method were applied in three versions: (1) estimation of the distribution parameters was done using the entire time series (DBS/QM), (2) conditioned on specific months (DBS-M/QM-M) and (3) conditioned classified on circulation patterns (DBS-L/QM-L).

2.2.4 Classification of atmospheric circulation patterns

Since the precipitation pattern over the study area has a clear seasonality (Fig. 4), MOS conditioned on separate months implicitly accounts for this. However, this assumption may not be stationary when applied in future climate impact studies since the timing of seasons might shift. Classified weather patterns have been used in meteorological applications for a long time to distinguish between typical weather situations, for example the Grosswetterlagen for European mainland (Baur et al., 1944; Hess and Brezowsky, 1977) and weather types for the British Isles (Lamb, 1950, 1972). They have also been extensively used in precipitation modelling (e.g. Bardossy and Plate, 1992). The conditioning of the distribution parameters on discrete, classified weather patterns overcomes the problem of a potential shift of seasons in a future
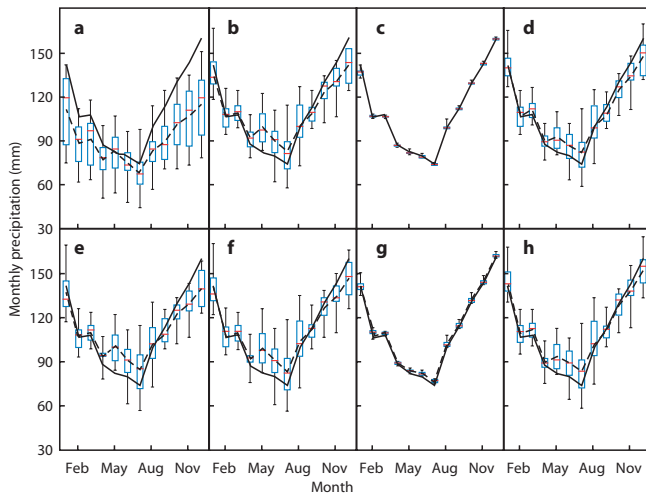
Fig. 4. Monthly mean precipitation over the Montford catchment evaluated over the time period 1961–2000. Boxplots denote the 25 and 50 percentiles where the whiskers are the upper and lower of the RCM values. The dashed line show the mean of the RCM data of **(a)** raw RCM output, and with MOS using **(b)** QM, **(c)** QM-M, **(d)** QM-L, **(e)** DM, **(f)** DBS, **(g)** DBS-M and **(h)** DBS-L. The black solid line shows the observed monthly mean precipitation.

perturbed climate. Any MOS conditioned on months is inherently stationary, but if the MOS is conditioned on weather patterns it instead becomes more sensitive to a possible shift in the large-scale circulation. Furthermore, weather pattern might be more reasonable to use in areas where the large-scale circulation rather than time of year depicts the precipitation pattern, such as the British Isles.

The COST733 action recently provided a catalogue of objective and subjective classifications of atmospheric circulation patterns over Europe using large-scale parameters such as mean sea level pressure, atmospheric thickness, vorticity and winds (Philipp et al., 2010). If the classifications are to be applied to condition the parameters of DBS-L, it is important that the classified patterns can discriminate between wet and dry precipitation patterns, since the distribution of precipitation is very different depending on the large-scale pattern. The circulation patterns from the COST733 catalogue were therefore evaluated according to their ability to discriminate between wet and dry patterns using two indices, one for precipitation occurrence I_1 and precipitation intensity I_2:

$$I_1 = \frac{1}{T} \sum_{t-1}^{T} \sqrt{(p(CP(t)) - \bar{p})^2} \tag{5}$$

$$I_2 = \frac{1}{T} \sum_{t=1}^{T} \left| \ln \left(\frac{z(CP(t))}{\bar{z}} \right) \right| \tag{6}$$

where T is the total number of classified events t, $p(CP(t))$ is the probability and $z(CP(t))$ is the amount of precipitation for event t with classification CP. I_1 distinguishes between wet and dry patterns regarding the frequency of wet

and dry days. A classification that separates rainy days from dry days yields a high score. As a complement, I_2 distinguishes between wet and dry patterns in terms of total precipitation The indices have been used in earlier studies of downscaling as measures of how well a classification can distinguish wet and dry conditions, and were used to select the classification which was mostly suited to downscale precipitation (Wetterhall et al., 2007). To narrow down the possible options, the classifications with maximum 12 patterns were selected from the COST733 database. The Lamb objective classification with 11 patterns centred over the British Isles had the highest score on both I_1 and I_2 and was accordingly selected for this study.

2.3 Rainfall-runoff modelling with the HBV model

The MOS-adjusted precipitation series were finally used as input to the hydrological HBV rainfall-runoff model (Lindström et al., 1997). The HBV model was run in a lumped configuration upstream of Montford river gauging station (Fig. 1). The HBV model requires daily temperature, precipitation and potential evapotranspiration as driving variables. Potential evapotranspiration (PE) was calculated through a simple method using radiation and temperature (McGuinness and Bordne, 1972) modified by Oudin et al. (2005)

$$PE = \frac{R_e}{\lambda \rho} \frac{T_a + K_2}{K_1}. \quad \text{if } T_a + K_2 > 0$$
$$PE = 0 \qquad \text{otherwise} \tag{7}$$

where R_e is extra-terrestrial radiation (depending only on latitude and ordinal date), T_a is the mean daily temperature in C°, λ is the latent heat flux, ρ is density of water and K_1 and K_2 are constants in C° that can be calibrated. This simple formulation of evapotranspiration has been found to be robust when applied in climate impact studies (Oudin et al., 2005).

2.3.1 Sensitivity analysis of parameter values

The HBV model has a large number of free parameters that have to be estimated when applied to catchment modelling (Seibert, 1997). In this study, a simplified version of the HBV-model was used, with 17 free parameters. Of these parameters, 6 describe snow and snowmelt, and were not included in the sensitivity analysis since snow melt processes were found to not have a huge influence on the studied catchment. The remaining 12 free parameters were tested using a base sample set generated using a quasi-random Lp Tau method, which is an efficient method to generate a quasi-random sequence for Monte Carlo experiments (Sobol, 1979; Pappenberger et al., 2006). The base sample was 200 000 simulation runs, this was deemed to be a large enough number of simulations to explore the parameter space. From these simulations, all parameter sets that generated a Nash-Sutcliffe coefficient > 0.85 were selected as behavioural sets.

This follows the principle of the Generalised Likelihood Uncertainty Estimation method. (GLUE; Xingguo et al., 2006; Pappenberger et al., 2008; Cloke et al., 2008). The threshold was subjectively selected to obtain a large enough set of parameter values to estimate the uncertainty in the HBV model. The number of behaviour models obtained was 2146, roughly the top percent of the tested parameter space. The Nash-Sutcliffe coefficient is defined as (Nash and Sutcliffe, 1970)

$$R = 1 - \frac{\sum\limits_{t=1}^{T} (q_{\text{obs}}(t) - q_{\text{sim}}(t))^2}{\sum\limits_{t=1}^{T} (q_{obs}(t) - \bar{q}_{\text{obs}})^2} \qquad (8)$$

where t is the time step, q the observed and simulated discharge at time step t, respectively, and $\bar{q}$ the mean observed discharge. See Table 1 for an overview of the values of the free parameters that was retained from the calibration. The RCM precipitation was evaluated as ensemble input to the HBV model using observed temperature and potential evapotranspiration.

2.4 Validation measures of MOS performance

The MOS methods were validated using precipitation statistics on mean daily precipitation and the mean annual maximum 5-day accumulation of precipitation (max5day) as well as a number of performance scores for precipitation and the modelled runoff, respectively. The max5day was calculated by filtering the daily precipitation series with a running 5-day mean filter and taking the mean of the annual maximum values. This measure identifies precipitation events that are prone to cause fluvial flooding in the catchment. The performance regarding precipitation intensities of the different methods was further validated through intensity-duration curves, calculated as the maximum intensity over accumulation times ranging from 1 to 30 days. The maximum values for different accumulation times shows biases against observations on time scales that are hydrologically interesting, both regarding fast-responding floods and longer-duration wet periods. The objective function for the temporal variation was selected as the mean rank of the observed intensities for each accumulation time in comparison to modelled intensities:

$$R = 1 - \frac{1}{T} \sum_{t=1}^{T} \left| \frac{2R(t)}{N} - 1 \right| \qquad (9)$$

where N is the number of RCMs, T is the number of accumulation times T over which the rank R was calculated (1 to 30 days). The rank score is constructed to yield a value of 1 for a perfect score and 0 for no skill to be easily comparable to the other scores.

The modelled runoff was evaluated against observations by calculating the annual maximum runoff, and then

calculating a fuzzy membership function over the time period 1987–2000 to allow for a spin-up period of 1 yr. The fuzzy membership function accounts for uncertainties in observations and was defined as 1 if the simulated flow is ±10 % of observed flow, and a linear decreasing to 0 if simulated flow is ±30 % that of the observation (Cloke et al., 2008).

$$O = \frac{1}{T} \sum_{t=1}^{T} \begin{cases} 0 & q(t) < a, q(t) > d \\ \frac{q(t)-a}{b-a} & a \leq q(t) \leq b \\ 1 & b < q(t) < c \\ \frac{d-q(t)}{d-c} & c \leq (t) \leq d \end{cases} \quad \left. \begin{array}{l} a = 0.7 q_{\text{obs}}(t) \\ b = 0.9 q_{\text{obs}}(t) \\ c = 1.1 q_{\text{obs}}(t) \\ d = 1.3 q_{\text{obs}}(t) \end{array} \right\} \quad (10)$$

where q is the simulated flow at time t and, q_{obs} is the observed flow and T is the total number of time steps. The fuzzy membership function has a perfect score of 1, and was used as a measure of performance (O).

A contingency table was also constructed for the annual maximum runoff values where the threshold for a hit was set to $300 \, \text{m}^3 \, \text{s}^{-1}$. This threshold corresponds to a warning level of intermediate floods in the Montford station. From this contingency table, the Relative Operating Characteristic (ROC) curve can be plotted as the probability of detection (POD) against the probability of false alarm rates (FAR), and the area under this curve is denoted as the ROC score (Demargne et al., 2010). A ROC score of 1 denotes a perfect forecast and a value of 0.5 means that the forecast is performing in par with climatology.

3 Results and discussion

3.1 Precipitation scores

The RCMs generally underestimate the precipitation over the upper Severn catchment (Fig. 2). The inter-annual variability was however well captured, indicating that the RCMs overall perform well when driven by reanalysis data, although the intense precipitation events were underestimated.

The precipitation statistics for daily means and max5day show that all MOS techniques can remove the mean bias in the daily precipitation (Table 2). However, the bias in the max5day is underestimated using the DM method, whereas the DBS overestimates the 5-day maximum. This indicates the difficulty in estimating the parameters of the gamma distribution for the upper tail. Using weather patterns somewhat dampens the overestimation.

Figure 3 exemplifies how the conditioning of weather patterns improves the MOS for the DBS method (similar results were obtained with QM, not shown). For the dry anticyclonic circulation (a) and the wet cyclonic circulation (b), the influence of weather patterns is negligible; the MOS in general adjusts precipitation values to better match observations, especially for the wet cyclonic circulation. However, for the patterns dominated by westerly (c) and easterly/southeasterly (d) winds, weather pattern classification substantially improves the MOS.

Table 1. Range of parameter values derived during the calibration of the HBV model. Parameter definitions follow those of Seibert (1997).
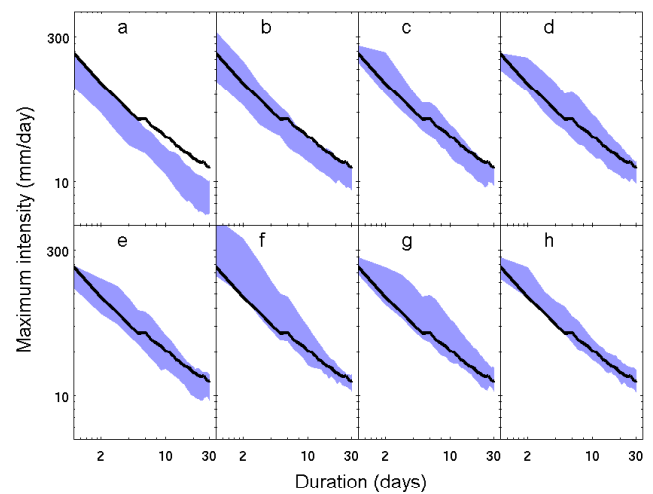
Parameter name	Description	Mean value	Standard deviation	Unit
fc	Maximum field capacity	296	147	mm
lp	Decides the threshold for reduction in ET	0.59	0.28	NA
beta	Shape coefficient for the soil moisture	4.2	2.6	NA
perc	Maximum ground water percolation from upper to lower storage	249	143	$mm\,d^{-1}$
uzl	Limit of upper ground water storage	506	287	mm
k0	Recession coefficient	0.50	0.28	d^{-1}
k1	Recession coefficient	0.41	0.23	d^{-1}
k2	Recession coefficient	0.21	0.15	d^{-1}
maxbas	Routing parameter	5.56	1.20	d
K1	Evapotranspiration constant	180	74	$C°$
K2	Evapotranspiration constant	20	7.8	$C°$

Table 2. Precipitation statistics for the MOS techniques calculated for the area upstream Montford.

MOS method	Mean precipitation $(mm\,day^{-1})$	Mean annual max 5-day precipitation (mm)	R	O	ROC
Observed	3.62 ± 0.46	87 ± 17		0.41	
Raw output	2.99 ± 0.37	76 ± 13	0	0.12	0.5
DM	3.62 ± 0.45	81 ± 16	0.23	0.24	0.71
QQ	3.62 ± 0.49	88 ± 18	0.27	0.26	0.78
QM-M	3.62 ± 0.49	87 ± 18	0.47	0.26	0.78
QM-L	3.60 ± 0.50	87 ± 17	0.37	0.26	0.79
DBS	3.62 ± 0.51	98 ± 23	0.62	0.25	0.75
DBS-M	3.67 ± 0.52	98 ± 22	0.63	0.24	0.8
DBS-L	3.66 ± 0.51	94 ± 21	0.76	0.25	0.83

Fig. 5. Intensity-duration curves for the RCM precipitation using **(a)** raw output and with MOS using **(b)** DBS, **(c)** DBS-M, **(d)** DBS-L, **(e)** DM, **(f)** QM, **(g)** QM-M and **(h)** QM-L. The black line is observed precipitation.

The raw output from the RCMs did not capture the intra-annual pattern of precipitation, especially for the wetter months September–January (Fig. 4a). Most floods occur during this period, and the deficiencies in modelling that period means that the RCMs cannot be used directly in impact studies to simulate winter floods. MOS generally improved the intra-annual variation (Fig. 4b–f). The DBS-M and QM-M reduced almost all bias in the RCM, and this was to be expected as the MOS was done on a monthly scale. The MOS methods that were not conditioned on months overestimated the precipitation in May–July, but also this was dampened when using weather patterns (Fig. 4).

The intensity-duration curves pointed to another issue regarding the RCM-modelled precipitation series (Fig. 5a). Even if the short-intensity precipitation events (1–10 days) were slightly underestimated, the discrepancy was more evident on longer time scales (10–30 days). Since the onset of a flood event is very dependent on the antecedent soil moisture and ground water conditions this can have the effect that a model run by the direct output from the RCM can be too dry, and therefore not be able to reproduce extreme flood events. Also here all MOS methods removed most of this bias (Fig. 5b–h, Table 2). The DBS methods removed most of the error on the long duration scale, but also shifted the bias for the high intensity events to the positive side. This was improved using months, and especially weather patterns to condition the MOS. The ROC from the precipitation-intensity show that the DBS performs best and again that using weather patterns improves the downscaling further, however not for the QM (Table 2).

This study was mainly focussing on how MOS techniques perform for the RCM ensemble as a whole, but it is also quite revealing to look at the performance of the individual RCMs, for example the correlation coefficients of the annual max5day in Table 3. This measure disregards any improvement in the daily means and focusses on the inter-annual variation of large events. The numbers in black denote that the MOS improves the score in comparison with the raw output, and stars mark that using months or weather patterns
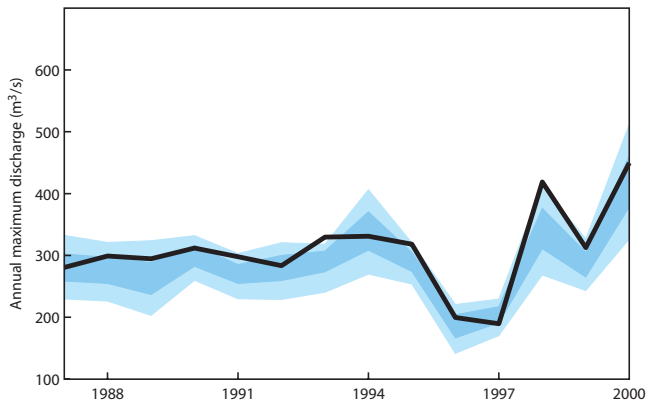
Fig. 6. Annual maximum runoff modelled with the HBV model at the station Montford with observed precipitation. The lighter blue areas show the 5th to 95th percentile, the darker blue show the 25th to 75th percentile of the parameter uncertainty for all models with NS>0.85. The black line is observed runoff.

Fig. 7. Annual maximum runoff modelled by the HBV model at the station Montford driven by (**a**) raw output and with MOS using (**b**) QM, (**c**) QM-M, (**d**) QM-L, (**e**) DM, (**f**) DBS, (**g**) DBS-M and (**h**) DBS-L. The lighter blue areas show the 5th to 95th percentile, the darker blue show the 25th to 75th percentile. The black line is observed runoff.

improved the performance in comparison with the original method. MOS do not always improve the RCM output, and in some cases the results are worse than the raw output. However, in most cases the results are better. Furthermore, the improvement with MOS is not correlated with the performance of the raw output, meaning that MOS cannot fully compensate for all errors in a bad performing RCM. Finally, the conditioning on weather patterns does improve the MOS, both for the QM and DBS methods.

3.2 Runoff modelling

The HBV model simulated runoff well over the calibration period and the best parameter set had a Nash-Sutcliffe efficiency of $R>0.90$. The annual maximum runoff was captured both in terms of magnitude and timing (Fig. 6). The RCM-driven runoff is not able to capture the pattern of either the maximum runoff (Fig. 7) or the daily runoff as simulated by the HBV model (Table 2) but all MOS methods improved the simulated runoff (Fig. 7). The event of 1994 is over-predicted by RCMs in comparison to observations, whereas the floods in 2000 are still very much under-predicted. The O score, show clearly that the MOS adds skill in comparison to raw output (Table 2). The more sophisticated methods performed better than the DM, but there is no clear signal as to which of the MOS techniques performs better.

3.3 Selecting the best MOS method

The overall best performance in this study was achieved with the DBS-L method, using gamma distribution conditioned on objective Lamb weather types (Tables 2 and 3). However, the improvements in comparison to the other techniques are not statistically significant, therefore the hypothesis that they all perform equally well cannot be rejected. Although the results are only strictly valid for this specific catchment, this work
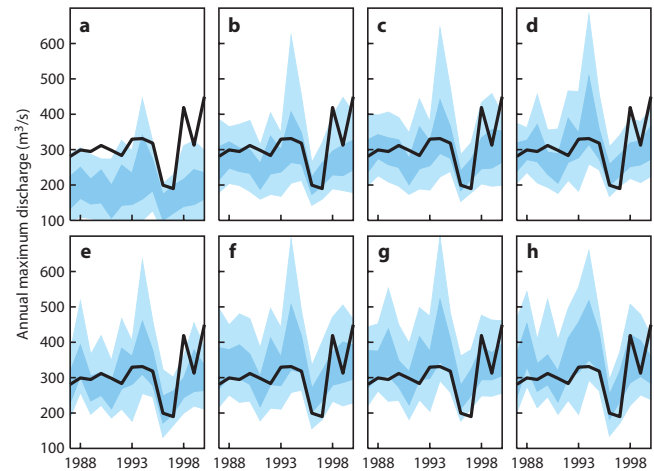
indicates the importance of a thorough test of MOS methods prior to their use in impact studies. In this case the focus was on flood events, but if the goal was to model the impact on water resources, a simpler method like the direct method could prove sufficient.

It has previously been shown that there can be large differences between downscaling techniques in terms of precipitation patterns and volumes over a certain catchment (Dibike and Coulibaly, 2005; Cloke et al., 2012). Also, even if the downscaled precipitation patterns are similar, the catchment may have a threshold response in which exactly this small difference may induce severe flooding or not. Beyond this threshold the difference in the techniques may be irrelevant. The threshold used in this study was selected to represent a value where flooding does occur, but the ROC score is sensitive to this threshold, and changing this might lead to slightly different results.

The selected classification was in this case a readily available classification, and it was not optimised on precipitation. An objective pattern classification optimised on precipitation or runoff events could further improve the DBS-L. Bardossy and Pegram (2011) applied an objectively derived classification to adjust precipitation values using quantile-matching technique and found that the RCMs were able to reproduce the patterns of current situations well.

The individual RCMs could be weighted based on their performance in the control period for use in studies of future climate change impact. The issue of weighting RCMs regarding their performance was recently discussed in Christensen et al. (2010). Their conclusion was that weighting schemes were very sensitive to the aggregation process of the modelled output and the selected metric, and that an extra

Table 3. Correlation coefficients for the annual maximum 5-day precipitation for the individual RCMs. Bold marks that the method improved the correlation compared with the raw output. Asterisks (*) denote that the conditioning on month or weather pattern improved results compared with the unconditional method.

RCM	Raw	DM	QM	QMM	QML	DBS	DBSM	DBSL
1	0.11	**0.12**	0.11	**0.24***	**0.15***	0.10	**0.14***	**0.16***
2	0.39	**0.40**	**0.44**	0.37	0.33	0.38	**0.44***	**0.39***
3	0.23	**0.24**	0.19	**0.37***	**0.23***	**0.25**	**0.34***	0.22
4	0.37	**0.40**	**0.39**	0.37	0.27	**0.38**	0.35	**0.38**
5	0.42	0.42	**0.43**	**0.47***	**0.43**	0.42	**0.51***	0.40
6	0.45	0.43	0.44	**0.45***	**0.45***	0.43	**0.46***	0.42
7	0.56	0.56	0.56	0.50	0.53	**0.57**	0.50	**0.60***
8	0.04	**0.07**	**0.08**	0.04	**0.06**	**0.05**	0.02	**0.08***
9	0.39	0.39	0.37	**0.48***	**0.40***	0.37	**0.46***	**0.39***
10	0.43	0.43	**0.44**	0.35	**0.44**	**0.44**	0.42	0.38
11	0.07	**0.08**	0.05	−0.03	**0.08***	0.07	0.03	**0.16***
12	0.12	**0.13**	**0.13**	**0.25***	**0.38***	**0.14**	**0.24***	**0.41***
13	0.11	0.11	**0.16**	0.02	**0.20***	0.07	0.05	**0.12***
14	0.36	**0.40**	0.30	0.28	0.31*	0.32	0.33*	0.35*
15	0.15	**0.16**	0.14	0.09	0.11	0.15	0.10	**0.19***
16	0.38	0.38	0.36	0.38*	**0.39***	0.38	**0.42***	**0.47***
17	0.41	0.41	0.41	**0.42***	0.40	0.41	**0.43***	**0.47***
18	0.48	0.48	0.47	0.45	0.40	**0.49**	**0.52***	**0.52***
Mean	**0.30**	**0.31**	**0.30**	**0.31**	**0.31**	**0.30**	**0.32**	**0.34**
Improvements		9	7	6 (8)	10 (9)	7	10 (10)	11 (13)

uncertainty was added in the procedure. MOS methods compensate for some of the errors in the RCMs, thus in this sense potentially lead to a reduction of the total uncertainty. However, their estimation requires extra parameters in the calibration, and therefore adds to the layer of uncertainty. On the other hand, the inclusion of local processes, especially a better representation of orography and description of land surface interactions, could increase the sharpness, for example higher resolution in space and time. This might mean that the total uncertainty in the impact model result could be reduced. This is especially true when the large-scale model is performing well (Diez et al., 2005).

Another issue is that different methods do perform well for different objective functions (Table 2). The type of MOS to be used in impact studies should therefore be carefully selected depending on the aim of the study in question. However, a range of MOS techniques as well as an ensemble of climate models is advisable in any impact study, since just using one technique might correct for a certain aspect of the precipitation bias, for example the number of rainy days or high-intensities. Using an ensemble of techniques expands the range of possible outcomes. It is also paramount to perform a MOS with utmost transparency to fully show the benefits as well as the possible pitfalls with MOS techniques.

4 Conclusions

This study evaluated three different methods to further downscale RCM precipitation for use in impact studies, the direct method, quantile-quantile mapping and distributions-based scaling. The QM and DM were employed in three variants, conditioned on (1) the entire time series, (2) individual months and (3) on classified weather patterns. The RCMs in the study were forced with ERA40 data and their performance could therefore be evaluated against observations directly. The main conclusions of this study were

- Using a range of performance measures to look at different aspects of the modelled precipitation shows the benefits of using more sophisticated MOS techniques in comparison with simpler schemes.

- Raw RCM output could not capture flood events when run through a hydrological model. MOS techniques improved the usability of the RCM output.

- Distribution scaling using a fitted gamma distribution conditioned on objectively classified Lamb weather types proved to be the better method in comparison with the direct method and the quantile-quantile mapping, however the results are not significant.

The development of RCMs moves constantly towards higher resolution and improvement in the parameterisation of physical parameters, and it is possible that these improvements will lead to the reduction in model errors. The recently

launched CORDEX aims to run RCMs with higher resolution, on the order of 10 km, and over many more regions of the world than what are presently available. The impact community will have the ability to perform studies over regions where to date few studies have been done, for example in arid and mountainous regions. The challenge in these regions will not only be to develop models that can describe the necessary processes, but also to tackle the problem of data availability.

Edited by: E. Coppola
Reviewed by: two anonymous referees

References

Bardossy, A. and Pegram, G.: Downscaling precipitation using regional climate models and circulation patterns toward hydrology, Water Resour. Res., 47, doi:10.1029/2010WR009689, 2011.

Bardossy, A. and Plate, E.: Space-time model for daily rainfall using atmospheric circulation patterns, Water Resour. Res., 28, 1247–1259, 1992.

Baur, F., Hess, P., and Nagel, H.: Kalender der Grosswetterlagen Europas 1881–1939, Bad Homburg, Germany, 35 pp., 1944.

Beven, K.: I believe in climate change but how precautionary do we need to be in planning for the future?, Hydrol. Process., 25, 1517–1520, 2011.

Boé, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydrometeorological studies. Int. J. Climatol., 27, 1643–1655, 2007.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7–30, 2007.

Christensen, J. H., Kjellström, E., Giorgi,, F., Lenderink, G., and Rummukainen, M.: Weight assignment in regional climate models. Clim. Res., 44, 179–194, 2010.

Cloke, H. L., Pappenberger, F., and Renaud, J.-P.: Multi-Method Global Sensitivity Analysis (MMGSA) for modelling floodplain hydrological processes, Hydrol. Process., 22, 1660–1674, 2008.

Cloke, H. L., Wetterhall, F., He, Y., Freer, J., and Pappenberger, F.: Modelling climate impact on floods with ensemble climate projections, Q. J. Roy. Meteor. Soc., online first, 2012.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global Planet. Chang., 57, 16–26, 2007.

Demargne, J., Brown,J., Liu, Y., Seo, D.-J., Wu, L., Toth, Z., and Zhu, Y.: Diagnostic verification of hydrometeorological and hydrologic ensembles, Atmos. Sci. Lett., 11, 114–122, 2010.

Dibike, Y. B. and Coulibaly, P.: Hydrologic Impact of Climate Change in the Saguenay Watershed: Comparison of Downscaling Methods and Hydrologic Models, J. Hydrol., 307, 145–163, 2005.

Díez, E., Primo, C., García-Moya, J. A., Gutiérrez, J. M., and Orfila, B.: Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts, Tellus A, 57, 409–423, 2005.

Eden, J. M., Widmann, M., Grawe, D., and Rast, S.: Skill, Correction, and Downscaling of GCM-Simulated Precipitation, J. Climate, 25, 3970–3984, 2012.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27, 1547–1578, 2007.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L., and Goodess, C.M.: Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios, Int. J. Climatol., 26, 1397–1415, 2006.

He, Y., Wetterhall, F., Cloke, H. L., Pappenberger, F., Wilson, M., Freer,, J., and McGregor, G.: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, Meteorol. Appl., 16, 91–101, 2009.

Hess, P. and Brezowsky, H.: Katalog der Grobwetterlagen Europas 1881–1976, 3 verbesserte und ergäntze, Auflage. Ber Dt. Wetterd., 15, 54 pp., 1977.

IPCC Pachauri, R. K. and Reisinger, A. (Eds): Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, Geneva, Switzerland, p. 104, 2007.

Lamb, H. H.: Types and spells of weather around the year in the British Isles : Annual trends, seasonal structure of the year, singularities, Q. J. Roy. Meteor. Soc., 76, 393–429, 1950.

Lamb, H. H.: British Isles weather types and a register of the daily sequence of circulation patterns 1861–1971, H. M. Stationery Off. (London), p. 85, 1972.

Leith, N. A. and Chandler, R. E.: A framework for interpreting climate model outputs, J. Roy. Stat. Soc. C-App., 59, 279–296, 2010.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, Hydrol. Earth Syst. Sci., 11, 1145–1159, doi:10.5194/hess-11-1145-2007, 2007.

Lindström G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, 1997.

McGuinness, J. L. and Bordne, E. F.: A comparison of lysimeter-derived potential evapotranspiration with computed values, Technical Bulletin 1452, Agricultural Research Service, US Department of Agriculture, Washington, DC, 1972.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themessl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change. Recent developments to

bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, doi:10.1029/2009RG000314, 2010.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I. A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2– Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, J. Hydrol., 303, 290–306, 2005.

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J. B., Pfister, L., and Fraipont de, P.: Influence of uncertain boundary conditions and model structure on flood inundation predictions, Adv. Water Resour., 29, 1430–1449, 2006.

Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P.: Multimethod global sensitivity analysis of flood inundation models, Adv. Water Resour., 31, 1–14, 2008.

Perry, M., Hollis, D., and Elms, M.: The generation of daily gridded datasets for temperature and precipitation, covering the UK for the period 1960 to 2006, Internal report, 2008.

Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Jourdain, S., Kreienkamp, F., Krennert, T., Lykoudis, S., Michalides, S. C., Pianko-Kluczynska, K., Post, P., Álvarez, D. R., Schiemann, R., Spekat, A., and Tymvios, F. S.: Cost733cat – A database of weather and circulation type classifications, Phys. Chem. Earth, Pt. A/B/C, 35, 360–373, 2010.

Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theor. Appl. Climatol., 99, 187–192, 2009.

Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods, Int. J. Climatol., 26, 679–689, 2006.

Seibert, J.: Estimation of parameter uncertainty in the HBV model, Nord. Hydrol., 28, 247–262, 1997.

Sobol, I. M.: On the Systematic Search in a Hypercube, Siam J. Numer. Anal., 16, 790–793, 1979.

Stehlík, J. and Bárdossy, A.: Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, J. Hydrol, 256, 120–141, 2002.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530–1544, doi:10.1002/joc.2168, 2010.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Da Costa, V., Bechtold, P., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.

van der Linden, P. and Mitchell, J. F. B.: ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project, technical report, Met Off. Hadley Cent., Exeter, UK 716, Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, 717 FitzRoy Road, Exeter EX1 3PB, UK, 2009.

von Storch, H. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, United Kingdom, 1999.

Wetterhall, F., Halldin, S., and Xu, C.-Y.: Seasonality properties of four statistical-downscaling methods in central Sweden, Theor. Appl. Climatol., 87, 123–137, 2007.

Wetterhall, F., He, Y., Cloke, H., and Pappenberger, F.: Effects of temporal resolution of input precipitation on the performance of hydrological forecasting, Adv. Geosci., 29, 21–25, doi:10.5194/adgeo-29-21-2011, 2011.

Widmann, M., Bretherton, C. S., and Salathé Jr., E. P.: Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor, J. Climate, 16, 799–816, 2003.

Wilks, D. S.: Statistical methods in the atmospheric sciences: an introduction, Academic Press, 501 UK, 1995.

Xingguo, M., Pappenberger, F., Beven, K. J., de Roo, A., and Suxia, L.: Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model, Hydrol. Sci. J., 51, 45–65, 2006.

Yang, W., Andréasson, J., Graham, L. P., Olsson, J., Rosberg, J., and Wetterhall, F.: Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. Hydrol. Res., 41, 211–229, 2010.