Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe

M. A. Sunyer[1], Y. Hundecha[2], D. Lawrence[3], H. Madsen[4], P. Willems[5,6], M. Martinkova[7], K. Vormoor[8], G. Bürger[8], M. Hanel[7,9], J. Kriaučiūnienė[10], A. Loukas[11], M. Osuch[12], and I. Yücel[13]

[1]{DTU Environment, Lyngby, Denmark}

[2]{Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum, Potsdam, Germany}

[3]{Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway}

[4]{DHI, Hørsholm, Denmark}

[5]{Hydraulics Laboratory, KU Leuven, Heverlee, Belgium}

[6]{Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, Belgium}

[7]{Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic}

[8]{Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany}

[9]{T. G. Masaryk Water Research Institute, Prague, Czech Republic}

[10]{Lithuanian Energy Institute, Kaunas, Lithuania}

[11]{Department of Civil Engineering, University of Thessaly, Volos, Greece}

[12]{Department of Hydrology and Hydrodynamics, Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland}

[13]{Middle East Technical University, Civil Engineering Department, Ankara, Turkey}

Correspondence to: M.A. Sunyer (masu@env.dtu.dk)

Abstract

Information on extreme precipitation for future climate is needed to assess the changes in the frequency and intensity of flooding. The primary source of information in climate change impact studies is climate model projections. However, due to the coarse resolution and biases of these models, they cannot be directly used in hydrological models. Hence, statistical downscaling is necessary to address climate change impacts at the catchment scale.

This study compares eight statistical downscaling methods often used in climate change impact studies. Four methods are based on change factors, three are bias correction methods, and one is a perfect prognosis method. The eight methods are used to downscale precipitation output from fifteen regional climate models (RCMs) from the ENSEMBLES project for eleven catchments in Europe. The overall results point to an increase in extreme precipitation in most catchments in both winter and summer. For individual catchments, the downscaled time series tend to agree on the direction of the change but differ in the magnitude. Differences between the statistical downscaling methods vary between the catchments and depend on the season analysed. Similarly, general conclusions cannot be drawn regarding the differences between change factor and bias correction methods. The performance of the bias correction methods during the control period also depends on the catchment, but in most cases they represent an improvement compared to RCM outputs. Analysis of the variance in the ensemble of RCMs and statistical downscaling methods indicates that at least 30% and up to approximately half of the total variance is derived from the statistical downscaling methods. This study illustrates the large variability in the expected changes in extreme precipitation and highlights the need of considering an ensemble of both statistical downscaling methods and climate models. Recommendations are provided on selection of the most suitable statistical downscaling methods to include in the analysis.

1. Introduction

Both the frequency and intensity of extreme precipitation are expected to increase under climate change conditions in Europe (Christensen and Christensen, 2003; IPCC, 2012). Several climate studies have focused on assessing these changes (e.g. Fowler and Ekström, 2009; Frei et al., 2006; Kendon et al., 2008) and their consequences in relation to the risk of flooding (Christensen and Christensen, 2003; IPCC, 2012; Leander et al., 2008; Vansteenkiste et al., 2013). The main steps often followed in these studies comprise the selection of one or several global climate models (GCM), regional climate models (RCM) and/or statistical downscaling methods (SDM). In climate change impact studies, hydrological models are then used to estimate changes in hydrological variables.

GCMs are the most comprehensive and widely used models for simulating the response of the global climate system to changes in greenhouse gas emissions. However, their spatial resolution (approximately 150 km) is often too coarse for addressing climate change impacts at the local scale, and variables such as precipitation are often biased. RCMs are climate models that cover a specific region (e.g. Europe) and use GCMs as boundary condition. RCMs have a higher spatial resolution (often approximately 25 km, but the new EURO-CORDEX simulations (Jacob et al., 2013) have a resolution of approximately 11 km) than GCMs, which makes them more adequate for assessing changes at the local scale. Nonetheless, RCMs often inherit the biases from the GCMs and their spatial resolution might still be too coarse for some impact studies (Maraun et al., 2010). Hence, further statistical downscaling is often needed to obtain bias-corrected projections at the local scale (Fowler et al., 2007). Statistical downscaling is based on defining a relationship between the large scale outputs of the RCMs (or GCMs) and the local scale variables required in impact studies (Fowler et al., 2007; Wilby et al., 2004).

In recent years, a relatively large number of RCM outputs have been made available, but there is no consensus on the best way to assess their performance (Knutti et al., 2010). There are several challenges in evaluating RCMs. For example, a RCM might perform well for some variables in some regions but not for other variables. Moreover, even if a climate model performs well under present climate conditions it might not perform equally well under future conditions (Knutti, 2010). For these reasons, it is generally recommended to use a multi-model ensemble of RCMs (or GCMs) instead of using a single model (Knutti et al., 2010; van der Linden and Mitchell, 2009; Tebaldi and Knutti, 2007).

Similarly, a large number of SDMs have been suggested in the literature, but there is no consensus on the best SDM. Fowler et al. (2007) and Maraun et al. (2010) provide comprehensive reviews of the methods suggested in the literature and their suitability for different applications. As in the case of climate models, the validation of SDMs is challenging. Only a few recent studies address this issue (e.g. Maraun et al., 2013; Räisänen and Räty, 2013; Teutschbein and Seibert, 2012; Vrac et al., 2007).

In order to account for the uncertainties in climate change impact studies and due to the lack of consensus on the best climate model and SDM, a number of studies consider multiple climate models and SDMs. For example regarding extreme events, Bürger et al. (2012 and 2013) used eight SDMs to downscale six GCMs forced with three emission scenarios, Sunyer et al. (2012) used five SDMs to downscale four RCMs driven by two GCMs, Hanel et al. (2013) used four SDMs and fifteen RCMs, and Kidmose et al. (2013) used two SDMs and nine RCMs. Bürger et al. (2012 and 2013) assessed the performance and variance arising from the SDMs and GCMs. They concluded that the main influence on the overall results for different extreme indices (including both precipitation and temperature indices) was the downscaling method used followed by the climate model selected. In their study, the main source of variance depended on the index considered, but overall the climate models had more influence on precipitation than on temperautre indices. Sunyer et al. (2012) and Hanel et al. (2013) showed that the variation in the results arising from the use of several statistical downscaling methods is larger in the case of extreme events (extreme precipitaiton in the case of Sunyer et al. (2012) and droughts in the case of Hanel et al. (2013)). Kidmose et al. (2013) found that in the case of extreme groundwater levels in Denmark the variance arising from the RCMs was larger than from the SDMs, but in this case only two SDMs were considered.

Some studies also consider hydrological models in the chain of uncertainties. For example, Wilby and Harris (2006) used two SDMs, four GCMs, and two emission scenarios combined with two hydrological model structures and two sets of hydrological model parameters. They concluded that the main sources of variation in the case of low flows are associated with the SDMs and GCMs used. Lawrence and Haddeland (2011) compared two SDMs, six RCMs driven by two GCMs, and two emission scenarios and used multiple parameter sets for the hydrological impact model. They found that for rainfall dominated catchments, the uncertainty arising from the hydrological parameters was more significant than other sources. In snowmelt dominated catchments, however, climate scenarios and SDMs were the main source of uncertainty. Wetterhall et al. (2012) assessed the variability in extreme discharge using three SDMs, sixteen RCMs, one hydrological model and a set of model parameters. The performance of the SDMs was evaluated and a best method was found, but it was not possible to reject the hypothesis that all SDMs perform equally well. Wetterhall et al. (2012) also concluded that more complex SDMs performed better than simple methods. A similar conclusion was reached by Räty et al. (2014) and Teutschbein and Seibert (2012). These two studies mainly focused on the validation of SDMs. Teutschbein and Seibert (2012) considered six SDMs and eleven RCMs for five Swedish catchments, while Räty et al. (2014) considered nine SDMs and six RCMs and considered two regions, Northern and Southern Europe.

The main focus of this study is to assess and compare the changes in extreme precipitation obtained using a range of SDMs and RCMs in eleven European catchments. For this purpose, precipitation outputs from fifteen RCMs driven by six GCMs from the ENSEMBLES project (van der Linden and Mitchell, 2009) are downscaled using eight SDMs based on different underlying assumptions. Four SDMs are change factor methods, three are bias correction methods and one is a perfect prognosis method. Some previous studies have compared the results from change factors and bias correction methods (e.g. Hanel et al., 2013; Ho et al., 2012; Räisänen and Räty, 2013) for mean temperature and mean precipitation for specific catchments. Here we focus on changes in extreme precipitation in a range of catchments over Europe with different climates. A key objective of this study is to assess whether it is possible to identify general advantages and deficiencies of the different SDMs when applied to the different catchments, and hence outline recommended use of SDMs. In addition, this study also focuses on whether there are common trends in projected changes in extreme precipitation over Europe and what the main sources of variation in the changes in extreme precipitation are.

The results presented here are based on a coordinated effort carried out as part of the COST Action FloodFreq (European Procedures for Flood Frequency Estimation, www.cost-floodfreq.eu). The outputs from this study have also been used as inputs to hydrological impact modelling in order to assess the changes in extreme discharge and flood frequency in the eleven catchments (Hundecha et al., submitted).

The next section describes the case study catchments and the data used, followed by the methodology section. Section 4 presents and discusses the results, and section 5 summarizes the findings and conclusions of the study.

2. Case study catchments and data

2.1. Observations

Figure 1 shows the location of the eleven catchments studied and the main properties of each catchment are summarized in Table 1. The two most northern catchments are the Norwegian catchments Nordelva at Krinsvatn (NO2) and Atna at Atnasjø (NO1), and the most southern catchment is Yermasoyia (CY) in Cyprus. The size of the catchments varies from the 6171 km^2 of Mulde (DE) in Germany to the 67 km^2 of Upper Metuje (CZ2) in the Czech Republic. Different precipitation patterns are represented in the catchments. The mean precipitation ranges between 2437 mm yr^{-1} in NO2 to 589 mm yr^{-1} in Nysa Kłodzka in Poland (PL). The season with more extreme precipitation events is summer for most of the catchments: NO1, DE, Aarhus in Denmark (DK), Merkys in Lithuania (LT), Grote Nete in Belgium (BE), and Jizera in the Czech Republic (CZ1). In NO2 and CY, winter is the season where most extremes occur, while in the Turkish catchment Omerli (TR) it is autumn. The season which is most subject to extremes is estimated from the extreme value series obtained considering the 1-yr threshold level and the whole time series (see section 3.2 for more details on how extreme precipitation is defined).

FIGURE 1

The observational data used is daily catchment precipitation, as the data were to be further used in catchment-based hydrological modelling in separate work (Hundecha et al., submitted). Different methods have been used to obtain areal precipitation time series. The catchments NO2, NO1, DK, and CZ2 use gridded data (derived from station data) to obtain areal average daily values for the catchment, while the remaining ones use station data to construct areal values. The cut-off value (threshold for dry days) for the observational data differs somewhat between the catchments. These catchment specific thresholds were not applied to the RCMs as they are not considered relevant for the analysis of extreme precipitation. Nonetheless some of the SDMs use thresholds to define dry and wet days (see section 3).

TABLE 1

2.2. Regional Climate Models

The climate model data used in this study is an ensemble of fifteen RCMs from the ENSEMBLES project (van der Linden and Mitchell, 2009). These fifteen simulations are based on eleven RCMs driven by six different GCMs. Table 2 shows the combinations of RCMs-GCMs used. The spatial resolution of all the models is 0.22° (approximately 25 km). For all the models, daily precipitation time series are available for the time period 1951–2100. In this study, we consider the time period 1961–1990 and 2071–2100 as the control and future time periods, respectively. It must be noted that six RCMs do not have data available for the year 2100. The future period used for these models is 2071-2099; this is not expected to have an influence on the results of this study. For each catchment, daily precipitation has been extracted from the 15 RCMs for the two periods using nearest neighbour interpolation to the catchment centroid. It must be noted that to simplify the calculations, the same control period is used for all the catchments. Therefore, in some catchments, the time period with observations (see Table 1) and the control period used from the RCMs do not fully overlap.

TABLE 2

3. Methodology

3.1. Statistical downscaling methods

Eight SDMs are used to obtain downscaled RCM projections at the catchment scale. These methods are based on the idea that it is possible to define a relationship between the large scale variables (RCM outputs) and local scale variables (catchment precipitation). Wilby and Wigely (1997) and Fowler et al. (2007) classify SDMs based on the relationship used to link large and local scale. They consider three groups: regression methods, weather type approaches and stochastic weather generators. Rummukainen (1997) classifies SDMs based on the information used from the large scale variables and defines two groups: perfect prognosis (PP) and model output statistics (MOS). Maraun et al. (2010) integrate both Rummukainen (1997) and Wilby and Wigely (1997) classifications and consider three groups: PP, MOS, and weather generators. According to this last classification, seven of the eight methods used here are MOS methods, and one method is a PP method.

Here we further classify the seven MOS methods into change factor (CF) methods and bias correction (BC) methods. Four of the MOS methods considered are CF and three are BC methods. CF methods estimate the change from control to future period projected by the RCM in one or several statistics and apply this change to the observations. These methods are based on the idea that RCMs represent the change from the control to the future climate better than the absolute values of the variables. The BC methods define a transfer function for the RCM outputs for the control period to match certain statistical properties of the observations. This transfer function is then used to correct the RCM outputs for the future period. CF methods preserve the temporal structure in the observed time series while BC methods preserve the temporal structure in the RCM outputs. It must be noted that both approaches are based on the assumption that the bias for the future period is identical to the bias for the control period, which may not be the case. Sunyer et al. (2014) show that the precipitation bias of the RCMs depends on the precipitation intensity and might change in the future.

The following sub-sections briefly describe the eight SDMs. In the results section we refer to the SDMs as either CF or BC methods. For simplicity, the perfect prognosis method is grouped with the BC methods even though it does not strictly correct the RCMs. It is included with the BC methods because it defines a transfer function between the RCM for the control period and the observations and then applies this to the RCM output for the future period.

A common terminology is used for describing the methods: P^{Obs} and P^{Fut} refer to the observed precipitation and the downscaled precipitation for the future period, respectively; and P^{RCMCon} and P^{RCMFut} refer to the precipitation output from the RCMs for the control and future time period, respectively. Similarly, $\mathrm{ECDF}^{\mathrm{Obs}}$ and $\mathrm{ECDF}^{\mathrm{Fut}}$ refer to the empirical cumulative distribution function (ECDF) for the observed precipitation and for the downscaled precipitation for the future while $\mathrm{ECDF}^{\mathrm{RCMCon}}$ and $\mathrm{ECDF}^{\mathrm{RCMFut}}$ refer to the ECDF estimated from the RCMs for control and future time period, respectively. The methods used here have been implemented as suggested in the literature, i.e. no harmonisation has been applied to enable, for example, a common method for accounting for seasonality or the definition of wet days. This is due to this study focused on the intercomparison of approaches in the way they are applied by the partners of FloodFreq COST Action, which was designed for the exchange and compilation of ideas and knowledge across participating countries. Table 3 summarizes the main advantages and disadvantages of each method.

3.1.1. Bias correction of mean

The bias correction of mean, BCM, is a simple method based on removing systematic errors in mean daily precipitation. It has been used in several hydrological applications (e.g. Hanel et al., 2013; Leander and Buishand, 2007; Leander et al., 2008). Here the method proposed by Leander and Buishand (2007) is used. This is based on the transformation:

$$P_{y,j}^{\mathbf{Fut}} = a_j P_{y,j}^{\mathbf{RCMFut}} \tag{1}$$

where y is the year , j is the day of the year and a_j is the transformation parameter. a_j is estimated in two steps. First, for all the years a subset of 61 days centred on day j is created for $P_{.,j}^{\mathrm{Obs}}$ and $P_{.,j}^{\mathrm{RCMCon}}$. Then, a_j is estimated as the mean of $P_{.,j}^{\mathrm{Obs}}$ divided by the mean of $P_{.,j}^{\mathrm{RCMCon}}$.

3.1.2. Bias correction of mean and variance

The bias correction of mean and variance method, BCMV, is an extension of the BCM method. It corrects the RCM outputs considering systematic errors in both the mean and the variance. This method has been applied in several studies (e.g. Hanel et al., 2013; Leander and Buishand, 2007; Leander et al., 2008). The method suggested by Leander and Buishand (2007) is followed here, which is based on the transformation:

$$P_{y,j}^{\mathbf{Fut}} = a_j \left(P_{y,j}^{\mathbf{RCMFut}} \right)^{b_j} \tag{2}$$

where a_j is estimated as described above for BCM, and b_j is estimated by equating the coefficient of variation of $(a_j P_{.,j}^{\mathrm{RCMCon}})^{bj}$ and $P_{.,j}^{\mathrm{Obs}}$. b_j is found by iteration since it is not possible to solve this equation in closed form.

3.1.3. Bias correction quantile mapping

Bias correction based on quantile mapping, BCQM, has been widely used to correct RCM outputs over Europe (e.g. Dosio and Paruolo, 2011; Gudmundsson et al., 2012; Piani et al., 2010). The non-parametric empirical quantile method suggested in Gudmundsson et al. (2012) is followed here. It is based on the concept that there exists a transformation h, such that:

$$P^{\mathbf{Obs}} = h\left(P^{\mathbf{RCMCon}}\right) = ECDF^{\mathbf{Obs}\ -1}\left(ECDF^{\mathbf{RCMCon}}\left(P^{\mathbf{RCMCon}}\right)\right) \tag{3}$$

First, all the probabilities in $\mathrm{ECDF}^{\mathrm{Obs}}$ and $\mathrm{ECDF}^{\mathrm{RCMCon}}$ are estimated at a fixed interval of 0.01. Then, h is estimated as the relative difference between the two ECDFs in each interval. Interpolation between the fixed intervals is based on a monotonic tricubic spline interpolation. A threshold for the correction of the number of wet days is estimated from the empirical probability of non-zero values in P^{Obs}. All RCM values below this threshold are set to zero. The precipitation values for the full annual daily series are corrected without subsampling by season or month, as suggested by Piani et al., 2010. The method was implemented in R using the *qmap* package (Gudmundsson, 2014).

3.1.4. Expanded downscaling

Expanded Downscaling, XDS, is a perfect prognosis technique which maps large-scale atmospheric fields to local station data. XDS was originally introduced for weather forecasting purposes, but it has been recently used in climate change studies (e.g. Bürger and Chen, 2005; Bürger et al., 2013; Dobler et al., 2012). The XDS approach is based on defining a multivariate linear regression between predictors y (multivariate fields of atmospheric variables) and predictands x (local scale variables, i.e. catchment precipitation), extended by the side condition that the local co-variability between the variables (and stations) is preserved:

$$XDS = \arg\min_{Q} \|xQ - y\|, \text{ subject to } Q'x'xQ = y'y, \tag{4}$$

where XDS is the least square-solution of the matrix Q which is found among those that preserve the local covariance ($Q'x'xQ = y'y$). By this approach, the estimation of extremes is supposed to be improved compared to regular linear regression models. See Bürger et al. (2009) for a detailed description of this method.

The XDS model is first trained on RCM atmospheric fields driven by the ECMWF ERA-40 reanalysis (Uppala et al., 2005) and local scale observations with at least 10 yrs of data. Then, RCM outputs for the control and future periods are used to generate time series at the local scale. Generally XDS allows for exploring a range of large scale variables as predictors. Large-scale reanalyses, however, are generally in better agreement with local observations than an RCM simulation driven by those reanalyses, simply because that the simulation likely differs from the actual weather realization which is used for XDS calibration. This has the consequence that a perfect prognosis approach is no longer perfect. A second data assimilation based on the RCM-ERA-40 runs (in addition to the data assimilation which has already been done for the ERA-40 reanalysis) would overcome this problem to some degree. However, such runs are not available for the RCMs accessible from the ENSEMBLES archive. For this study, the predictors were therefore chosen rather 'conservatively', with predictor variables being limited to large-scale total and convective precipitation. The result is a set of predictors that is, moreover, unique across all catchments. The XDS source code and documentation can be downloaded from: http://xds.googlecode.com.

3.1.5. Change factor of mean

The change factor of mean, CFM, is a simple method which has been widely applied in hydrological applications (Hanel et al., 2013; Prudhomme et al., 2002; Sunyer et al., 2012). It is based on applying the change in mean precipitation projected by the RCMs to the observed data.

The method described in Sunyer et al. (2012) is followed here. Similarly to BCM, this method is based on the transformation:

$$P_{m,t}^{\mathbf{Fut}} = a_m P_{m,t}^{\mathbf{Obs}} \tag{5}$$

where m refers to the month and t to each time step in the observations; a_m is the relative change in the precipitation mean for month m. a_m is estimated as the mean of $P_{m,.}^{\text{RCMFut}}$ divided by the mean of $P_{m,.}^{\text{RCMCon}}$.

3.1.6. Change factor of mean and variance

The change factor of mean and variance, CFMV, is an extension of CFM. It has been applied in several studies (e.g. Hanel et al., 2013; Räisänen and Räty, 2013; Sunyer et al., 2012). CFMV modifies the observed time series using the change in both the mean and variance. The method described in Sunyer et al. (2012) is followed here. Similar to BCMV, the method is based on the transformation:

$$P_{m,t}^{\mathbf{Fut}} = a_m \left(P_{m,t}^{\mathbf{Obs}}\right)^{b_m} \tag{6}$$

where a_m is estimated as described for CFM. b_m is estimated by equating the coefficient of variation of the time series $(a_m P_{m,.}^{\text{Obs}})^{bm}$ and the coefficient of variation estimated for the future period. As in BCMV, this is solved by iteration. The coefficient of variation for the future period is calculated from the relative change in the mean and variance projected by the RCMs.

3.1.7. Change factor quantile mapping

The change factor quantile mapping, CFQM, is based on using the relative change in the ECDF projected by the RCMs to modify the observed data. It has been applied in several climate change studies (e.g. Boé et al., 2007; Olsson et al., 2009).

This method uses the ECDF of wet days estimated for each month m for the observations, and the RCM output for the control and future periods. The probability intervals considered are 0.001 for quantiles lower than 0.9 and 0.0005 for higher quantiles (linear interpolation between intensities is applied to obtain the precipitation intensity for all the quantiles). Wet days are defined as days with precipitation higher than 1 mm. The perturbation of the observed time series is carried out in three steps. First, for each wet day in each month m, $\text{ECDF}_m^{\text{Obs}}$ is used to estimate the probability of the precipitation intensity. Second, the relative change in the intensity for this probability is estimated from $ECDF_m^{RCMFut}$ and $ECDF_m^{RCMCon}$. This change is then multiplied to the observed precipitation intensity to obtain the intensity for the future period. Dry days in the observations are not modified.

3.1.8. Change factor quantile perturbation

The change factor quantile perturbation, CFQP, is similar to CFQM but it also accounts for changes in the number of wet days. Quantile perturbation methods can be performed either in a non-parametric way (Ntegeka et al., 2014; Vansteenkiste et al., 2014; Taye et al., 2011; Willems and Vrac, 2011) or in a parametric way based on distribution calibration (Willems, 2013; Rana et al., 2014). The version used here is the non-parametric one that was applied by Willems and Vrac (2011).

The observations are perturbed using a two-step approach. First, the number of wet days (days with precipitation higher than 0.1 mm day^{-1}) is changed for each month. The relative change in the frequency of wet days is estimated from the RCM output. If the frequency increases, dry days are randomly selected and replaced by random wet day intensities from the time series. Otherwise, wet days are randomly replaced by zero precipitation. In the second step, the wet day intensities are perturbed in a similar way as in the CFQM method. The empirical probability of each intensity is estimated, and the relative change in the intensity for each probability is then calculated (linear interpolation is applied when different probabilities are obtained for the control and future period) and used to perturb the observations.

These two steps are repeated 10 times. The repetition that leads to the results closest to the mean monthly precipitation value of all the repetitions is selected. See Willems and Vrac (2011) for more details on this method, including checks of the coefficient of variation, skewness and autocorrelation for the results.

It must be noted that in the case of BCQM, CFQM, and CFQP the use of empirical quantiles may lead to large fluctuations representing a lack of robustness in the values of the CF (or correction factors in the case of BCQM) for the highest quantiles. This is due to the fact that the highest quantiles are estimated using a limited number of values.

3.2. Extreme precipitation Index

The outputs from all the statistical downscaling methods are analysed using an extreme precipitation index (EPI). This is defined as the average change in extreme precipitation higher than a defined return period. In this study, the return period is set equal to 1 and 5 yrs. EPI is estimated separately for each SDM, RCM, catchment, threshold return period, season and temporal aggregation. Four seasons are considered: winter (December to February), spring (March to May), summer (June to August), and autumn (September to November). Additionally, the index is estimated considering the whole time series, i.e. without dividing in seasons. The temporal aggregations considered are 1, 2, 5, 10, and 30 days. These are estimated using a moving average from the daily time series.

The first step in the calculation of EPI is the extraction of the extreme value series from the precipitation time series using a Peak Over Threshold (POT) approach. Peaks are extracted by using the 1- and 5-yr threshold return periods. For example, with a 30-yr record, the 30 and 6 most extreme events are included in the extreme series for the 1- and 5-yr threshold levels, respectively. An independence criterion based on the inter-event time is applied to make sure that extreme values are independent, i.e. only values separated by more than Δt days are considered. Δt is set equal to the temporal aggregation, i.e. for an aggregation time of 1 day, events must be separated by more than one day. EPI is then estimated as:

$$\mathrm{EPI} = \frac{\overline{\mathrm{POT}_2}}{\overline{\mathrm{POT}_1}} \tag{7}$$

where $\overline{\mathrm{POT}_1}$ and $\overline{\mathrm{POT}_2}$ are the averages of the selected POT values for reference and scenario, respectively. EPI takes the value of 1 if no change is estimated from reference to scenario and greater (less) than 1 if the average extreme precipitation is higher (lower) in the scenario time series.

In the results section, EPI is used to compare the changes in the downscaled time series from control to future. Additionally, three further comparisons are carried out. In total EPI is calculated for four different cases:

1. Comparison of the downscaled time series for the control and future periods.
2. Comparison of the RCM outputs for control and future periods. This allows us to compare the changes estimated from the downscaled precipitation, estimated in (1), to the changes projected by the RCMs.
3. For the four BC methods: comparison of the observations and the bias corrected RCMs for the control period. The value of the index for this comparison is a measure of the error of the BC methods in bias correcting the RCM outputs for extreme precipitation.

4. Comparison of the observations and RCM outputs for the control period. This comparison evaluates the performance of the RCMs in simulating extreme precipitation, and allows us to assess whether the error in the bias corrected time series, estimated in (3), is smaller than in the RCMs.

3.3. Variance decomposition

The variability in the EPI values found when comparing the downscaled time series for control and future arises mainly from three sources: GCMs, RCMs and SDMs. A variance decomposition approach is used to address the influence of each of these sources on the total variance for each catchment, return level, season and temporal aggregation. The approach described in Déqué et al. (2007, 2012) is followed here.

The total variance of EPI, V, can be split into the different contributions as:

$$V = R + G + S + \mathrm{RG} + \mathrm{RS} + \mathrm{GS} + \mathrm{RGS} \tag{8}$$

where R, G, and S are the individual parts of the variance explained by the RCMs, GCMs, and SDMs, respectively; RG, RS, and GS are the variance due to the interaction of RCM-GCM, RCM-SDM, and GCM-SDM, respectively; and RGS is the variance due to the interaction of all three sources. The part of the total variance explained by the RCMs, *V(R)* is:

$$V(R) = R + \mathrm{RG} + \mathrm{RS} + \mathrm{RGS} \tag{9}$$

The part of the total variance due to the GCMs, *V(G)*, and SDMs, *V(S)*, can be obtained in a similar way. The variances in Eq. (8) and Eq. (9) can be estimated as:

$$\begin{aligned}
R &= \frac{1}{11}\sum_{i=1}^{11}\left(\overline{\mathrm{EPI}}_{i..} - \overline{\mathrm{EPI}}_{...}\right)^2;\\
RG &= \frac{1}{11}\frac{1}{6}\sum_{i=1}^{11}\sum_{j=1}^{6}\left(\overline{\mathrm{EPI}}_{ij.} - \overline{\mathrm{EPI}}_{i..} - \overline{\mathrm{EPI}}_{.j.} + \overline{\mathrm{EPI}}_{...}\right)^2\\
RGS &= \frac{1}{11}\frac{1}{6}\frac{1}{8}\sum_{i=1}^{11}\sum_{j=1}^{6}\sum_{k=1}^{8}\left(\mathrm{EPI}_{ijk} - \overline{\mathrm{EPI}}_{ij.} - \overline{\mathrm{EPI}}_{i.k} - \overline{\mathrm{EPI}}_{.jk} + \overline{\mathrm{EPI}}_{i..} + \overline{\mathrm{EPI}}_{.j.} + \overline{\mathrm{EPI}}_{..k} - \overline{\mathrm{EPI}}_{...}\right)^2
\end{aligned} \tag{10}$$

where EPI_{ijk} is value of the index for RCM *i*, GCM *j* and SDM *k*, $\overline{\mathrm{EPI}}$ represents the average of EPI with respect to the subscripts that are replaced by a dot. The rest of the terms in Eq. (9) are estimated in a similar way as shown in Eq. (10). For more details see Déqué et al. (2007, 2012). Note that the observation errors in this approach are neglected in comparison with the other error sources.

As in Déqué et al. (2007), not all the terms in Eq. (10) can be estimated. This is because not all the combinations of RCM-GCMs are available (see Table 2). Déqué et al. (2007) suggested a simple method to reconstruct the missing data in the matrix of RCM-GCMs. This is based on minimizing the full interaction term RGS. However, this approach cannot be directly used here. This is because for the combinations of RCM *i* and GCM *j* that are not available there is no information on any of these SDM *k* values. Hence, in some cases it is not possible to estimate $\overline{\mathrm{EPI}}_{ij\cdot}$, which is needed to minimize the full interaction term *RGS*. For this reason, a slight modification is made to the approach suggested by Déqué et al. (2007). The approach followed here consists of two steps: (i) for all the combinations of *i* and *j* missing, $\overline{\mathrm{EPI}}_{ij\cdot}$ is estimated by minimizing RG; and (ii) the values of EPI_{ijk} missing are estimated by minimizing RGS.

A large number of gaps must be filled using this procedure. Two simple verifications have been carried out to check that the results are not largely affected by the matrix reconstruction approach. The first verification procedure is a simple comparison of the results from the variance decomposition described above with a variance decomposition approach, which considers only two sources of variance (climate models and SDMs). In the approach considering only these two sources, matrix reconstruction is not needed because all the elements in the matrix are known. The second verification procedure is similar to the verification carried out in Déqué et al. (2007). The two verification approaches and their results are described in Appendix A.

The results from the first verification procedure show that the conclusion as to which is the most important source of variance is nearly the same when considering two or three sources for all catchments. Conversely, the results from the second verification show that the reconstruction approach can influence the results. From the results of the first verification, we decide to analyse the variance explained by the GCMs and RCMs separately (i.e. considering three sources of variance) because, in our opinion, it adds value to separate the influence of the GCMs and RCMs. Nonetheless, we acknowledge that the results must be treated with caution due to the uncertainty added in the matrix reconstruction procedure.

TABLE 3

4. Results and discussion

This section is divided into two main parts. The first part analyses the results of all SDMs. The second part focuses on the performance of the three BC methods and perfect prognosis method. All the results are shown for winter and summer as these are the two seasons where most of the extremes occur under present conditions. However, it should be noted that in some catchments changes in other seasons might also be important due to their influence on floods, see examples in Hundecha et al. (submitted).

4.1. Comparison of the downscaled time series for the control and future periods

This subsection analyses the results of the eight SDMs driven by all RCMs. A summary of the results obtained for all the catchments is first presented followed by a more detailed analysis of the differences between the SDMs for three selected catchments.

4.1.1. Extreme precipitation index and variance decomposition for all catchments

Figure 2 summarizes the results of all the SDMs and RCMs for all the catchments for winter and summer for a temporal aggregation of 1 day. Additionally, it compares the results of the SDMs with the changes between the control and future periods projected by the RCMs. For the catchment CY for some SDMs, two special situations are encountered. For the methods BCM and BCMV for both winter and summer periods, due to the few rainy days in some of the RCM simulations, some of the parameters take unrealistic values which lead to unrealistic values of EPI. Similarly, it is not possible to estimate the CFs used in the case of CFM, CFMV and CFQM in the summer period. The results of these methods are, therefore, not included in the analysis for CY. For the other catchments such problems with the SDMs were not encountered and all results are included in the analysis.

For winter, extreme precipitation is expected to increase in all catchments (the median of EPI is greater than 1) except in CY. The median of EPI is similar for all catchments except for the two most northern catchments (NO1 and NO2) and the most southern catchment (CY). The EPI values range between 1.11 and 1.2 for the 1 yr threshold, and 1.14 and 1.22 for the 5 yr threshold. For this season, a similar variability is found for all catchments, except for CY, where the variability is slightly larger than in the other catchments. For summer, the median is also greater than 1 for all the catchments except for the two most southern catchments (CY and TR). These two catchments also have a larger variability. In general, there are larger differences between and within the catchments in summer than in winter.

In most catchments, and for both threshold (1 and 5 yrs), larger changes are expected for winter. Only in the case of NO2 the changes obtained for summer are larger than in winter. In the catchment in LT, CZ1 and CZ2, larger changes are obtained for winter for the 1 yr level and for summer for the 5 yr level. In both seasons and in most catchments, larger changes and variability are obtained for the 5 yr level.

Comparing the changes obtained from the SDMs with the mean changes projected by the RCMs (see Fig. 2), there is a general tendency that slightly smaller changes are estimated from the uncorrected RCM projections. However, there are some significant differences. For example, for NO2 in winter and the 5 yr level, the uncorrected RCM projections point to a decrease of extreme precipitation but the SDMs point to an increase. The opposite situation is obtained for CY for the same season and 1 yr level. For this catchment (CY) in summer, there is also a rather large difference between the changes estimated from the uncorrected RCM projections and the SDMs. The largest difference between the uncorrected RCMs and downscaled results is obtained in CY. The maximum difference is obtained in summer for the 5 yr level where the downscaled values lead to a change 20% higher than the uncorrected RCMs. Excluding CY, the average difference of the change between the downscaled and uncorrected series is small. For example, for the 1 yr level the average difference is 0.013 for winter and 0.022 for summer. The smallest difference in both seasons is obtained for the Danish catchment for which the difference is 0.003 in winter and 0.009 in summer. These overall results show that, in general, the SDMs do not modify the change projected by the uncorrected RCMs significantly. Nonetheless, in some cases the use of some downscaling methods might modify the magnitude of the change projected by the uncorrected RCMs. The influence of the SDM used with respect to the difference between the change projected by the uncorrected RCMs and the downscaled data is analysed in more detail in the next section.

FIGURE 2

Figure 2 does not differentiate between the variability due to the use of different SDMs and different RCM-GCM simulations. The variance decomposition approach is used to assess each of the sources of variance individually. Figure 3 shows the total variance decomposed in the variance arising from the GCMs, RCMs, SDMs and the interaction terms for all catchments for the 1 and 5 yr levels and temporal aggregation of 1 day. For CY the results for the summer are not shown and results for the winter do not include BCM and BCMV because EPI could not be calculated for a large number of cases (due to the few rainy days in some of the RCM simulations).

As shown in Fig. 2, the variance for the 5 yr level is higher for all catchments and seasons than the variance for the 1yr level. In summer, the variance tends to increase from north to south for the 5 yr level, and to some extent also for the 1 yr level. This trend is not observed in winter. The larger variance in the southern catchments for the 5 yr level may be partially caused by larger sampling variance (smaller number of extreme events). Figure 3 shows that in most cases the variance due to the RCM-GCM simulations is larger than the variance from the SDMs. However, the interaction term is in both seasons and in most catchments similar or larger than the individual sources of variance.

Figure 3 also shows the fractional percentage explained by *V(G)*, *V(R)*, and *V(S)*, such that the three terms sum to 100%. The scaling of the percentages to obtain a total of 100% is needed because some interaction terms are included in several factors. As already mentioned, the percentage explained by the RCM-GCM simulations is in most cases larger than the percentage explained by the SDMs. The only exception is TR for summer and PL for winter for the 1 yr level. However, in all cases, the percentage explained by the SDMs is at least 30% of the total variance, which is considerable. Similar results are obtained for winter and summer for the 1 and 5 yr levels. For both seasons and return levels, there are no clear spatial patterns in the percentages. These results are in agreement with the results obtained by Räty et al. (2014). They carried out a similar variance decomposition to study the variance arising from climate models and statistical downscaling methods over northern and southern Europe. For northern Europe, they found that for the 70th and higher precipitation percentiles, the climate models are the main source of variance and the variance arising from the SDMs is at least 20% and the interaction term accounts for approximately 20%. For southern Europe, the contribution of the SDMs is also at least 20%, but the variance arising from the interaction term is higher (it ranges between 20 and 50% for all percentiles). In addition, and also in agreement with the results shown here, Kidmose et al. (2013) found that for extreme groundwater levels in a Danish catchment the variance arising from the ensemble of climate models is higher than the variance arising from the SDMs, although only two downscaling methods were considered. They also highlighted the importance of natural variability, which in their case was higher than the variability related to climate models and downscaling methods. The results for Norway (NO2 and NO1) are also in agreement with the results found by Lawrence and Haddeland

(2011). The influence of the SDMs is larger in the snow dominated catchment, NO1, than in the rainfall dominated catchment, NO2.

In all cases the percentage of the variance explained by the RCMs is larger than the percentage explained by the GCMs. For both return levels, in winter the average percentage explained by the GCMs is approximately 20% while in summer it is approximately 15%. The smaller percentage for the GCMs in the summer is due to the larger relative influence of both the RCMs and SDMs. This is likely due to the fact that in Europe, extreme precipitation from convective storms occurs more frequently during summer (e.g. Lenderink, 2010; Hofstra et al., 2009), and this has a larger influence on the outputs from the RCMs and SDMs due to their higher spatial resolution. Several studies have shown that the errors of the RCMs are larger in the representation of daily extreme precipitation in summer over Europe (e.g. Frei et al., 2006; Fowler and Ekström, 2009).

The results of the variance decomposition obtained for aggregation levels larger than 1 day (not shown) point towards a smaller total variance. For these temporal aggregations, the main source of variation is also the RCM-GCMs, although the percentage explained by SDMs is slightly larger than for the 1 day aggregation. The decrease in total variance and in the percentage explained by RCM-GCMs mainly reflects that the model outputs being more similar for larger temporal aggregations. The results from the variance decomposition highlight the need for considering both a range of SDMs and an ensemble of RCMs driven by different GCMs for assessing the uncertainty in the projection of changes in extreme precipitation.

FIGURE 3

4.1.2. Extreme precipitation index for three selected catchments

The previous section summarizes the main results regarding the expected changes in extreme precipitation when considering all the RCMs and SDMs. This section focuses on the differences between the statistical downscaling methods. For this purpose, three catchments have been selected: NO2, DE, and TR (distributed north to south and with different precipitation patterns). Figure 4 shows the median, 25th, and 75th quantile of EPI for each SDM for the three catchments for the 1 yr level and a temporal aggregation of 1 day.

In NO2, for both seasons, the SDMs based on BC show a lower EPI than the methods based on CFs. In winter, all the CF methods point towards an increase in extreme precipitation, although some of the BC methods show a decrease for some RCMs. In summer, all methods point to an increase except XDS, which produces a small EPI and a large variability. There are several factors which may contribute to these differences. As this region is projected to generally have an increase in winter precipitation, use of change factor methods that do not correct for changes in the number of wet days will automatically produce higher values for extreme precipitation in winter. If this precipitation increase is, however, also associated with a change in storm patterns, such that the increase simply reflects an increase in wet days rather than wet day extremes, then this difference would be reflected in the results for the BC methods.

In DE, all the SDMs lead to similar median values except the BCMV in winter and CFM in summer. The differences between BCMV and the other two BC methods are due to some RCMs leading to very large changes when they are downscaled with BCMV, e.g. for RCA-ECHAM5, the values of EPI are 1.18 for BCM, 1.16 for BCQM and 1.63 for BCMV. This large value of EPI is caused by unexpectedly large precipitation intensities obtained from the non-linear transformation in BCMV, which is one of the disadvantages of this method (see Table 3). For the BCMV method two events of 55 and 60 mm/day are obtained while the largest events for the two other BC methods are below 40 mm/day (for the control period all the events are lower than 30 mm/day).

CFM leads to the lowest value of EPI obtained in summer. This is also the case for all the other catchments considered in this study except NO2 and Yermasoyia in Cyprus (results not shown). It indicates that mean precipitation is likely to increase less than the more extreme precipitation intensities. In addition, it illustrates that the CFM method is not suitable for regions where the expected changes in extreme precipitation are different than the changes in mean precipitation.

In TR, the results of the SDMs vary more than in DE and NO2. For this catchment, CFM leads to the lowest EPI in both seasons, which indicates a lower increase in mean precipitation than in extreme precipitation, as in DE. In summer, all SDMs point to a decrease of extreme precipitation except BCM and BCMV, which do not show a change in extreme precipitation. These two methods show the largest variability for both winter and summer. The high variability for these two methods is due to the same issue identified in CY, i.e. only a few rainy days in the RCM simulations, the annual percentage of rainy days ranges between 12% and 28% rainy days.

For all catchments and both seasons, very similar results are obtained for CFQM and CFQP. This is expected since the main difference between the two methods is the treatment of wet day frequency. This is expected to have a minor impact, except for TR in the summer, where there are only very few rainy days during the summer period. This implies that in some cases all the rainy days are included in the selection of extreme events. Hence, the change in the number of wet days may have an effect on the changes in extreme precipitation. Similar results to those illustrated in Fig. 4 were also obtained for the 5 yr level (results not shown).

The results for the three catchments show that there is not a clear tendency in the differences between CF and BC methods. In addition, there is no evidence that methods that are based on the same statistics for the correction (e.g. BCM and CFM or BCMV and CFMV) will lead to similar results. Hence, it is not possible to generalize the results with respect to the use of SDM. This result contrasts with the findings in Hanel et al. (2013) for low flows in the Czech Republic. They found that, in general, the SDMs which account for changes in variance (such as BCMV and CFMV) led to larger changes in runoff. In addition, they also found larger changes in runoff for BC than for CF methods.

The EPI estimated using the uncorrected RCMs can be used as a reference to assess whether the downscaled data preserves the changes projected by the RCMs and the differences depending on the SDM. In the case of NO2, the EPI estimated using the uncorrected RCMs lie in between the values from the BC and CF methods. The downscaling method that shows the closest agreement with the changes projected by the RCMs is BCQM. Overall for the three catchments and both seasons this method is the one that shows values of EPI closest to the ones estimated from the uncorrected RCMs. This points towards the suitability of this method to downscale extreme precipitation as it corrects the properties of interest for representing extreme precipitation. On the other hand, EPI obtained from CFM tend to produce the largest deviations from the EPI of the uncorrected RCMs (except in the case of TR in summer), which again shows that this method is not suitable for projecting changes in extreme precipitation. In addition, problems of producing unrealistic extreme precipitation values with some of the methods, such as BCM and BCMV in TR in summer, XDS in TR in winter and NO2 in summer are clearly seen when comparing their EPI values with those obtained from the uncorrected RCMs. The above examples illustrate that some SDMs are better suited for downscaling extreme precipitation and some SDMs are less robust with respect to downscaling various precipitation patterns.

FIGURE 4

Figure 5 analyses the eight SDMs for the three catchments for two temporal aggregations: 1 and 30 days. In general, the variability in EPI in the RCM ensemble decreases with increasing temporal aggregation, except for a few cases, e.g. XDS in NO2 and BCM for DE in summer. There is no general indication that EPI either increases or decreases with increasing temporal aggregation.

In NO2, EPI is larger for a temporal aggregation of 30 days for BCM, BCMV and BCQM, and it is lower for the CF methods and XDS for summer. In winter, EPI for BCM, BCMV and BCQM is also slightly larger for a temporal aggregation of 30 days (in the case of BCM and BCMV, this means a smaller reduction of extreme precipitation). In DE, most methods show a lower EPI for 30 days except CFM in summer and CFM, CFMV and XDS in winter. Similarly, in TR all the methods show lower EPI for 30 days except for CFM, XDS and CFQM in summer. For all catchments, the results of the SDMs at 30 days temporal aggregation are more similar than for 1 day aggregation.

In most cases, EPI at 1 and 30 days are not considerably different and show the same signal (except in the case of TR for BCM and BCMV for both seasons and BCQM in winter). As for the 1 day aggregation, the results with temporal aggregation of 30 days do not allow general conclusions with respect to the use of SDM.

FIGURE 5

4.2. Comparison of observations and bias corrected RCMs for the control period

The previous section focuses on the analysis of the expected changes in extreme precipitation. This section uses EPI to compare the results from the BC methods for the control period and the observations. This allows us to evaluate how well the different BC methods correct extreme precipitation from the RCMs. As in the previous section, a summary of the results found for all the catchments is first presented, followed by a more detailed analysis of the results found for each BC method for three of the catchments. It must be noted that this comparison of the results for the control period does not provide a validation of the downscaling methods. The data used to downscale the RCMs for the control period is the same as the data used for the calibration of these methods. Nonetheless, it should be noted that the validation of downscaling methods is crucial and relevant for assessing how well we can estimate changes in extreme precipitation. However, the validation of SDMs is challenging as it requires either observational data that have different properties to enable assessing whether the downscaling methods can be used to project climate changes (e.g. Refsgaard et al. 2014; Teutschbein and Seibert, 2012) or, alternatively, the use of pseudo-realities (e.g. Räisänen and Räty 2013; Vrac et al. 2007; Maraun et al., 2015). If the observational data do not show pronounced changes in extremes, then the results of the validation analyses are questionable with respect to the suitability of the methods for use in climate change analyses. There is, thus, a clear need for further research on validation methods for SDMs. It will not be addressed in this paper.

For BE, CY, CZ2, DK, and PL, the control period considered for the RCMs does not fully overlap with the observation period. In the case of DK, for example, there is only an overlap of 2 yrs. The use of different periods assumes that the statistics are stationary between the periods. However, some of the disagreements between the observations and bias corrected results may well be due to non-stationary statistics between the two periods.

4.2.1. Extreme precipitation index for all catchments

Figure 6 shows EPI estimated using the observations and the bias corrected RCM. In this figure (and the rest of the figures in this section), a value of 1 indicates that there is no difference between the extreme value statistics from the observations and the bias corrected RCM. A value greater (less) than one indicates that the bias corrected RCM overestimates (underestimates) extreme precipitation. It must be noted that for the catchments LT and TR there is a perfect overlap between the time period of the observations and RCMs, while for the other catchments the observation period includes the RCM period or there is only a partly overlap between the time period of the observations and RCMs (see Table 1 for details).

FIGURE 6

For extreme winter precipitation there is no clear tendency across catchments for under- or overestimation with the bias corrected data. The catchments that have the largest underestimation are for the most northern and southern catchments (NO2, NO1, DK and CY), whereas LT, BE and PL have the largest overestimation. For extreme summer precipitation, there is a pronounced underestimation for a number of catchments. The three most northern catchments (NO2, NO1, and DK) show the lowest mean bias based on the median values for all downscaled projections. The most southern catchment (CY) has the largest underestimation of extreme summer precipitation. Both the median and variance of EPI depend on the catchment and the season. For example, the bias corrected data for LT, BE and PL tend to overestimate extreme precipitation in winter, but underestimate this in summer. CZ1 in winter and NO2 in summer are the catchments that lead to the median closest to 1. The largest variability is found for PL in winter and TR and CY in summer.

The comparison of the error in the RCMs before and after bias correction shows that, in general, the error after bias correction is smaller than before bias correction. This shows that the BC methods improve the representation of extremes. However, in a few cases the error of the RCMs before bias correction is smaller than after bias correction. This is because some of the RCMs result in large errors after bias correction. For example for BE in winter with the HadRM3Q3-HadCM3Q3 model, values of 1.18 for BCM, 1.37 for BCMV, 1.24 for BCQM, and 1.23 for XDS are obtained, while a value of 0.98 is obtained from the uncorrected data. In fact, the average over all the RCMs shows that none of the downscaling methods improves the results of the uncorrected RCMs for this catchment. A similar result is obtained for the DE catchment. In the summer period, the results after bias correction for all the downscaling methods in the LT catchment show larger differences compared to the observations than the uncorrected RCMs. In both seasons, these results (error of the RCMs before bias correction is smaller than after bias correction) are obtained for catchments where the RCMs have the lowest error in representing observed extreme precipitation (i.e. EPI closer to 1). This indicates that if the agreement between the observations and RCMs is high, the downscaling methods considered in this study are not able to improve it. The next section describes in more detail the difference between EPI of the uncorrected RCMs and the downscaled series for each bias correction method.

4.2.2. Extreme precipitation index for each bias correction method for three selected catchments

Figure 7 shows the results of the three BC methods and XDS for NO2, DE, and TR for the 1 yr level and 1 day temporal aggregation. The performance of each method varies depending on the season and catchment. For example, BCM overestimates extremes in NO2 in winter and TR in summer and underestimates them in NO2 in summer and TR in winter. In DE, BCM performs equally well as BCMV. This illustrates that simple BC methods can, in some cases perform similarly or better than more advanced methods. In the catchments considered in this study, there is no clear relationship between the performance of the BC methods and the precipitation regime for the catchments.

In winter, the errors obtained for DE are smaller than in the other two catchments. EPI ranges from an underestimation of 4% (EPI equal to 0.96) for BCM and BCMV, to an overestimation of approximately 6% for BCQM and XDS. For this catchment and both seasons, BCM and BCMV lead to better results than BCQM and XDS. In summer, the errors in NO2 are smaller than in the other two catchments. For this catchment and this season, XDS is the method that leads to the smallest error and variability.

The largest errors and variability in the results are found for the TR catchment in both seasons. For this catchment and in the winter period, the median of all methods underestimate extremes except XDS, while in summer BCM and BCMV overestimate extremes and the other two methods underestimate. A very large variability is obtained for BCM and BCMV in summer (the 25th and 75th percentiles range from 0.4 to 1.5).

Comparison of the results of the SDMs with EPI obtained from the uncorrected RCMs shows that in the case of NO2 all the SDMs clearly agree better with the observations. But for the other two catchments, the results depend on the downscaling method. In DE, BCM and BCMV lead to better results than the other two methods for both seasons. In the TR catchment, BCQM leads to the best result in winter but not in summer, where BCMV produces the best result. Even though the results depend on the catchment analysed, the BCM is the method leads to the least improvements in most cases compared to the results of the uncorrected RCM. This is in agreement with the main conclusion from the validation study carried out by Teutschbein and Seibert (2012). They concluded that the linear bias correction (equivalent to the BCM method used here) together with the delta-change method (equivalent to the CFM used here) are less reliable than other more complex methods. Similarly, the cross-validation study carried out by Räty et al. (2014) showed that the linear bias correction method tends to perform more poorly than the other more complex bias correction methods, especially for high percentiles (between 75th and 97th percentile) in southern Europe and between the 50th and 70th percentile in northern Europe. Nonetheless, it should be noted that even if in some cases it is possible to identify a method that performs better than others it might not be possible to reject the hypothesis that all SDMs perform equally well (Wetterhall et al., 2012). This points towards the advantage of using an ensemble of SDMs to represent the uncertainty related to the statistical downscaling.

FIGURE 7

The results from Figure 7 indicate that the bias correction methods do not in all cases improve the time series from the RCMs. This must be tested for each application. Figure 8 shows the error of each BC method for two temporal aggregations, 1 and 30 days, for the 1 yr level. In general, the performance of the BC methods for the winter period improves for large temporal aggregation (except for XDS in TR). However, in summer this is not the case. For this season, the difference between the results for 1 and 30 day aggregations depends on the catchment and the method. In NO2, the results for 1 day are better than for 30 days for BCQM and XDS, although the reverse is true for TR. In DE, the results for 1 day are better than for 30 days for all the methods except XDS.

As shown in Fig. 7, TR has the largest variability for 30 days followed by NO2 for both seasons. The results for DE appear to be the least dependent on the temporal aggregation. This may be the result of spatially averaging the observations from 43 stations to derive the catchment precipitation. For such a large basin (6171 km², see Table 1), this may simultaneously lead to temporally-averaged precipitation values from the gauged nested sub-catchments. In all cases, the variability for 30 days is smaller than for 1 day, indicating that the RCMs lead to more similar results for large temporal aggregations.

FIGURE 8

5. Summary and conclusions

This study analyses the expected changes in extreme precipitation in eleven European catchments. It focuses on the variability in the changes arising from the use of different statistical downscaling methods as well as different RCM-GCM simulations. Fifteen RCMs driven by six GCMs are downscaled using eight statistical downscaling methods. The statistical downscaling methods rely on different assumptions and different RCM outputs. The outputs from all the statistical downscaling methods are analysed using an extreme precipitation index.

Extreme precipitation is expected to increase in most catchments in both winter and summer. A decrease in extreme precipitation is only expected for both winter and summer in CY and for summer in TR. In most catchments, larger changes are expected in winter than in summer. Additionally, in all cases, larger increases and larger variability in the results are obtained for the higher return level, 5 years.

In most catchments and for both winter and summer, the RCM-GCM projections are the main source of variability in the results when compared to the differences between SDMs, although variability due to the SDMs explains at least 30% of the total variance in all cases. Additionally, in all cases, the RCMs represent a larger percentage of the total variability than the GCMs, especially in summer. For this season, the total variance tends to be higher for the most southern catchments.

In general, the eight statistical downscaling methods agree on the direction of the change but not the magnitude of the change. It is not possible to draw general conclusions regarding differences between the downscaling methods, as the differences depend on the physical geographical characteristics of the catchment and the season analysed. For example, for NO2 the bias correction methods lead to lower changes than the change factor methods, but this is not the case for the other catchments. A common result for all catchments except NO2 and CY is that the CFM method leads to the smallest increase of extreme precipitation in summer. This indicates that this method is not suitable for regions where the expected changes in extreme precipitation differ from the changes in mean precipitation. The changes obtained for different temporal aggregations also depend on the physical geographical characteristics of the catchment and season analysed, i.e. there is no general tendency for an increase or decrease in the index with increasing temporal aggregation.

Overall, the bias correction methods improve the representation of extreme precipitation, as compared with the uncorrected RCM outputs. However, the bias corrected time series tend to underestimate extreme precipitation. The magnitude of the errors depends on the catchment and season analysed. For example, the results of the bias correction of mean are worse than the other methods for the NO2 but not for the other catchments. There is no clear relationship between the performance of the bias correction methods and the precipitation regime of the catchment. There is also no clear indication of an increase or decrease in the error with increasing temporal aggregation.

The results from the statistical downscaling methods have been compared with the extreme precipitation obtained from the uncorrected RCMs. Although the results depend on the catchment and season as in the other comparisons discussed before, some overall conclusions can be extracted from this comparison. Regarding the comparison of the change in extreme precipitation projected by the uncorrected RCMs and the downscaled series, the SDM that showed the smallest differences relative to the RCM projections is the BCQM method, while the method that led to the largest differences is the CFM method. These differences between the methods are more pronounced for the summer period. From the comparison of the SDMs and the uncorrected RCMs in representing the current period it was found that in general the BCM method fails in more cases than the other SDMs in improving the representation of extreme precipitation from the uncorrected RCMs.

From the results of all these comparisons, it is possible to draw some general recommendations when selecting SDMs from the ones considered here for downscaling extreme precipitation. Downscaling methods that do not explicitly correct or take into account changes in extreme precipitation may lead to different climate change signals than the ones projected by the RCMs and should not be used. In this study, this occurs mainly with CFM. In addition, some methods fail to correct the errors in the RCMs in representing extreme precipitation. In this study, this occurred in more cases when using BCM than with the other methods. Finally, in catchments with long dry periods the BCM, BCMV, CFM, CFMV, and CFQM methods produce unrealistic results and should not be used (or should be configured differently than done in this study with respect to describing the seasonal patterns). BCMV may also lead to unrealistic results in other catchments as seen in the case of DE. The ability of the downscaling methods to improve the representation of extreme precipitation from the RCMs and to preserve the climate change signal should be assessed for each case study in order to select the most suitable SDMs.

This study illustrates that there is a large variability in the changes estimated from different statistical downscaling methods and RCMs. It also shows that the differences between the methods and the performance of the bias correction methods depend on the catchment studied. Hence, for a specific case study, the selection of a suitable statistical downscaling method may depend on the physical geographical characteristics of the catchment. However, we recommend the use of a set of statistical downscaling methods as well as an ensemble of climate model projections. The selection of statistical downscaling methods should include: methods that are able to project changes in extreme precipitation if they are expected to be different from other precipitation properties; methods based on different underlying assumptions, for example BC and CF methods; and methods that use different outputs from the RCMs as, for example, XDS, CF or BC methods including mean and variance of precipitation, and methods including a range of quantiles.

Acknowledgements

This work was carried out as part of the COST Action ES0901 "European Procedures for Flood Frequency Estimation (FloodFreq)" (http://www.cost-floodfreq.eu/). Maria A. Sunyer and Henrik Madsen were supported by the Danish Council for Strategic Research as part of the project RiskChange. Deborah Lawrence acknowledges support from NVE for the internal research project 'Climate change and future floods'. Patrick Willems was supported by research projects for the Fund for Scientific Research (F.W.O.) - Flanders, Flanders Hydraulics Research and the Flemish Environment Agency. Klaus Vormoor acknowledges funding from the Helmoltz graduate research school GeoSim. Marta Martínková was financially supported by Internal Grant Agency of the Faculty of Environmental Sciences, CULS Prague (00000869/2013). Gerd Bürger was funded by the Austrian Climate and Energy Fund as part of the Austrian Climate Research Programme. Ismail Yucel was supported by TÜBİTAK project number 110Y036. The data from the RCMs used in this work was funded by the EU FP6 Integrated Project ENSEMBLES contract number 05539 (http://ensembles-eu.metoffice.com), whose support is gratefully acknowledged.

References

Boé, J., Terray, L., Habets, F. and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int. J. Climatol., 27(12), 1643–1655, doi:10.1002/joc.1602, 2007.

Bürger, G. and Chen, Y.: Regression-based downscaling of spatial variability for hydrologic applications, J. Hydrol., 311(1-4), 299–317, doi:10.1016/j.jhydrol.2005.01.025, 2005.

Bürger, G., Reusser, D. and Kneis, D.: Early flood warnings from empirical (expanded) downscaling of the full ECMWF Ensemble Prediction System, Water Resour. Res., 45(10), n/a–n/a, doi:10.1029/2009WR007779, 2009.

Bürger, G., Murdock, T.Q., Werner, A.T., Sobie, S.R. and Cannon, A.J.: Downscaling Extremes—An Intercomparison of Multiple Statistical Methods for Present Climate, J. Clim., 25, 4366–4388, doi: 10.1175/JCLI-D-11-00408.1, 2012.

Bürger, G., Sobie, S. R., Cannon, A. J., Werner, A. T. and Murdock, T. Q.: Downscaling Extremes: An Intercomparison of Multiple Methods for Future Climate, J. Clim., 26(10), 3429–3449, doi:10.1175/JCLI-D-12-00249.1, 2013.

Christensen, J. H. and Christensen, O. B.: Climate modelling: Severe summertime flooding in Europe., Nature, 421(6925), 805–6, doi:10.1038/421805a, 2003.

Danish Meteorological Institute (DMI): Climate Grid Denmark. Dataset for use in research and education. Daily and monthly values 1989-2010 10x10 km observed precipitation 20x20 km temperature, potential evaporation (Makkink), wind speed, global radiation, Technical Report 12-10, report available at http://beta.dmi.dk/fileadmin/Rapporter/TR/tr12-10.pdf (last access: 3 June 2014), 2012.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E., Castro, M. and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Clim. Change, 81(S1), 53–70, doi:10.1007/s10584-006-9228-x, 2007.

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C. M., Jacob, D., Lenderink, G. and Christensen, O. B.: The spread amongst ENSEMBLES regional scenarios: regional climate models, driving general circulation models and interannual variability, Clim. Dyn., 38, 951–964, doi:10.1007/s00382-011-1053-x, 2012.

Dobler, C., Bürger, G. and Stötter, J.: Assessment of climate change impacts on flood hazard potential in the Alpine Lech watershed, J. Hydrol., 460-461, 29–39, doi:10.1016/j.jhydrol.2012.06.027, 2012.

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, J. Geophys. Res., 116(D16), D16106, doi:10.1029/2011JD015934, 2011.

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, doi:10.1002/joc1556, 2007.

Fowler, H. J. and Ekström, M.: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, Int. J. Climatol., 29(3), 385–416, doi:10.1002/joc.1827, 2009.

Frei, C., Schöll, R., Fukutome, S., Schmidli, J. and Vidale, P. L.: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, J. Geophys. Res., 111, D06105, doi:10.1029/2005JD005965, 2006.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16(9), 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Gudmundsson, L.: qmap: Statistical transformations for post-processing climate model output. R package version 1.0-2, technical note available at: http://cran.r-project.org/web/packages/qmap/qmap.pdf (last access: 3 June 2014), 2014.

Hanel, M., Mrkvičková, M., Máca, P., Vizina, A. and Pech, P.: Evaluation of Simple Statistical Downscaling Methods for Monthly Regional Climate Model Simulations with Respect to the Estimated Changes in Runoff in the Czech Republic, Water Resour. Manag., 27, 5261–5279, doi:10.1007/s11269-013-0466-1, 2013.

Ho, C. K., Stephenson, D. B., Collins, M., Ferro, C. a. T. and Brown, S. J.: Calibration Strategies: A Source of Additional Uncertainty in Climate Change Projections, Bull. Am. Meteorol. Soc., 93(1), 21–26, doi:10.1175/2011BAMS3110.1, 2012.

Hofstra, N., Haylock, M., New, M. and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, J. Geophys. Res., 114, D21101, doi:10.1029/2009JD011799, 2009.

Hundecha, Y., Sunyer, M.A., Lawrence, D., Willems, P., Bürger, G., Kriaučiūnienė, J., Loukas, A., Martinkova, M., Osuch, M., Vormoor, K., Yücel, I., Madsen, H.: Effect of downscaling climate data on indices of extreme river flow: A comparative study across Europe, in preparation.

IPCC: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, edited by C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley, Cambridge University Press, Cambridge, UK, and New York, NY, USA., 2012.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B. and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Chang., 14(2), 563–578, doi:10.1007/s10113-013-0499-2, 2013.

Kidmose, J., Refsgaard, J. C., Troldborg, L., Seaby, L. P. and Escrivà M. M.: Climate change impact on groundwater levels: ensemble modelling of extreme values, Hydrol. Earth Syst. Sci., 17, 1619–1634, doi:10.5194/hess-17-1619-2013, 2013.

Kendon, E. J., Rowell, D. P., Jones, R. G. and Buonomo, E.: Robustness of Future Changes in Local Precipitation Extremes, J. Clim., 21(17), 4280–4297, doi:10.1175/2008JCLI2082.1, 2008.

Knutti, R.: The end of model democracy?, Clim. Change, 102(3-4), 395–404, doi:10.1007/s10584-010-9800-2, 2010.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. and Meehl, G. a.: Challenges in Combining Projections from Multiple Climate Models, J. Clim., 23, 2739–2758, doi:10.1175/2009JCLI3361.1, 2010.

Lawrence, D. and Haddeland, I.: Uncertainty in hydrological modelling of climate change impacts in four Norwegian catchments, Hydrol. Res., 42(6), 457, doi:10.2166/nh.2011.010, 2011.

Leander, R. and Buishand, T. A.: Resampling of regional climate model output for the simulation of extreme river flows, J. Hydrol., 332(3-4), 487–496, doi:10.1016/j.jhydrol.2006.08.006, 2007.

Leander, R., Buishand, T. A., van den Hurk, B. J. J. M. and de Wit, M. J. M.: Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, J. Hydrol., 351(3-4), 331–343, doi:10.1016/j.jhydrol.2007.12.020, 2008.

Lenderink, G.: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations, Clim. Res., 44, 151–166, doi:10.3354/cr00946, 2010.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., 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(3), RG3003, 2010.

Maraun, D., Widmann, M., Benestad, R., Kotlarski, S., Huth, R., Hertig, E., Wibig, J. and Gutierrez, J.: VALUE - Validating and Integrating Downscaling Methods for Climate Change Research, EGU Gen. Assem. Conf. Abstr., 15, 12041, 2013.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A.I.: VALUE: A framework to validate downscaling approaches for climate change studies. Earth's Future. doi: 10.1002/2014EF000259, 2015.

Ntegeka, V., Baguis, P., Roulin, E. and Willems, P.: Developing tailored climate change scenarios for hydrological impact assessments, J. Hydrol., 508, 307–321, doi:10.1016/j.jhydrol.2013.11.001, 2014.

Olsson, J., Berggren, K., Olofsson, M. and Viklander, M.: Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden, Atmos. Res., 92, 364–375, doi:10.1016/j.atmosres.2009.01.015, 2009.

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, doi:10.1007/s00704-009-0134-9, 2010.

Prudhomme, C., Reynard, N. and Crooks, S.: Downscaling of global climate models for flood frequency analysis: where are we now?, Hydrol. Process., 16(6), 1137–1150, doi:10.1002/hyp.1054, 2002.

Räisänen, J. and Räty, O.: Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, Clim. Dyn., 41, 1553–1568, doi:10.1007/s00382-012-1515-9, 2013.

Rana, A., Foster, K., Bosshard, T., Olsson, J. and Bengtsson, L.: Impact of climate change on rainfall over Mumbai using Distribution-based Scaling of Global Climate Model projections, J. Hydrol.: Reg. Stud., 1, 107–128, doi: 10.1016/j.ejrh.2014.06.005, 2014.

Räty, O., Räisänen, J. and Ylhäisi, J.S.: Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations, Clim. Dyn., 42, 2287–2303, doi: 10.1007/s00382-014-2130-8, 2014.

Rummukainen, M.: Methods of statistical downscaling of GCM simulations, Reports Meteorology and Climatology 80, Tech. rep., Swedish Meteorological and Hydrological Institute, SE-601 76 Norrkping, Sweden, 1997.

Šercl, P: Assessment of methods for area precipitation estimates, Meteorological Bulletin, 61(2), 2008.

Sunyer, M. A., Madsen, H. and Ang, P. H.: A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change, Atmos. Res., 103, 119–128, doi:10.1016/j.atmosres.2011.06.011, 2012.

Sunyer, M.A., Madsen, H., Rosbjerg, D., Arnbjerg-Nielsen, K.: A Bayesian approach for uncertainty quantification of extreme precipitation projections including climate model interdependency and non-stationary bias, J. Climate, 27, 7113–7132, doi:10.1175/JCLI-D-13-00589.1, 2014.

Taye, M. T., Ntegeka, V., Ogiramoi, N. P. and Willems, P.: Assessment of climate change impact on hydrological extremes in two source regions of the Nile River Basin, Hydrol. Earth Syst. Sci., 15, 209–222, doi:10.5194/hess-15-209-2011, 2011.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections., Philos. Trans. A. Math. Phys. Eng. Sci., 365(1857), 2053–2075, doi:10.1098/rsta.2007.2076, 2007.

Teutschbein, C. and Seibert, J.: Is bias correction of Regional Climate Model (RCM) simulations possible for non-stationary conditions?, Hydrol. Earth Syst. Sci. Discuss., 9(11), 12765–12795, doi:10.5194/hessd-9-12765-2012, 2012.

Tveito, O. E., Bjørdal, I., Skjelvåg, A. O. and Aune, B.: A GIS-based agro-ecological decision system based on gridded climatology, Meteorol. Appl., 12(1), 57–68, doi:10.1017/S1350482705001490, 2005.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., 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, doi:10.1256/qj.04.176, 2005.

van der Linden P., and Mitchell J.F.B. (eds.): ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK. 160pp, 2009. Technical report available at: http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf (last access: 3 June 2014)

Vansteenkiste, Th., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F. and Willems, P.: Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. J. Hydrol., 519, 743–755, doi: 10.1016/j.jhydrol.2014.07.062, 2014.

Vrac, M., Stein, M. L., Hayhoe, K. and Liang, X.-Z.: A general method for validating statistical downscaling methods under future climate change, Geophys. Res. Lett., 34(18), L18701, doi:10.1029/2007GL030295, 2007.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P. and Mearns, L. O.: Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods., 2004.

Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resour. Res., 42(2), n/a–n/a, doi:10.1029/2005WR004065, 2006.

Willems, P. and Vrac, M.: Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change, J. Hydrol., 402, 193–205, doi:10.1016/j.jhydrol.2011.02.030, 2011.

Willems, P.: Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium, J. Hydrol., 496, 166–177, doi: 10.1016/j.jhydrol.2013.05.037, 2013.

1

Appendix A - Verification of matrix reconstruction approach

A.1 Comparison of results using 2 and 3 sources of variance

This verification approach assesses the influence of the matrix reconstruction procedure on the percentage of the total variance explained by climate models (influence of GCM-RCM simulations) and SDMs. For this purpose, the variance decomposition approach has been applied considering two sources of uncertainty: SDMs and climate models (the 15 RCM-GCM simulations). In the case of two sources of variance, there is no need to reconstruct the matrix.

Table A1 shows the percentage explained by the climate models and SDMs estimated considering two and three sources of variance. The percentages for CY are not shown for summer because EPI could not be calculated for a large number of cases, and the percentages for winter do not include the results from BCM and BCMV. The percentage explained by the GCM-RCM simulations and the SDMs is similar when considering two or three sources of variances. Additionally, the conclusion on which is the most important source of variance is the same for all catchments except for DE and PL in winter. For these two catchments, the percentage explained by the GCM-RCM simulations is approximately 50%.

TABLE A1

A.2 Comparison of reconstructed and original values

A similar verification approach as the one carried out in Déqué et al. (2007) has also been used. It consists in removing the data for one combination of RCM-GCM and using the matrix reconstruction approach to estimate its values for all SDMs. The reconstructed values are then compared with the original values and also with two other combinations of RCM-GCMs (one using the same RCM and one using the same GCM). This test is applied to two RCM-GCM simulations: RCA-ECHAM5 and HIRHAM-BCM.

The reconstructed vector for these combinations is referred to as $\mathbf{EPI_{RG}}$. In the case of RCA-ECHAM5, $\mathbf{EPI_{RG}}$ is compared with the vectors found for: (i) the original EPI values found for RCA-ECHAM5; (ii) the combination RCA-BCM ($\mathbf{EPI_{R}}$ in Table A2); (iii) and the combination REMO-ECHAM5 ($\mathbf{EPI_{G}}$ in Table A2). In the case of HIRHAM-ARPEGE, $\mathbf{EPI_{RG}}$ is compared with the original values, with HIRHAM-ARPEGE ($\mathbf{EPI_{R}}$), and RCA-BCM ($\mathbf{EPI_{G}}$). Table A2 shows the average of the RMSE obtained for all the catchments, T-yr levels, seasons, and temporal aggregations.

Table A2 shows that in the case of RCA-ECHAM5, the difference between the reconstructed and the original values is smaller than the difference between the reconstructed values and the other two RCM-GCM combinations. However, in the case of HIRHAM-BCM, the difference between the reconstructed and the original values is higher than the difference between the reconstructed and the other two RCM-GCM combinations.

TABLE A2

This results show that in some cases the reconstructed values can differ more from the original values than they differ from other models. Hence, the variances estimated in the variance decomposition approach are likely to be affected by the reconstructed values.

Table 1 - Summary of the main characteristics of the catchments. The column with label “extremes” indicates the season where most precipitation extremes occur. The catchments are sorted from north to south, with the most northern catchment in the top row.

Name	River, Country	Area [km^2]	Median altitude [m]	Data used for calculation of catchment precipitation	Mean annual precipitation [$mm\ yr^{-1}$]	Extremes	Observation period
NO2	Nordelva, Norway	207	349	1x1 km grid (Tveito et al., 2005)	2437	Winter	1957 – 2010
NO1	Atna, Norway	463	1204	1x1 km grid (Tveito et al., 2005)	852	Summer	1957 – 2010
DK	Aarhus Å, Denmark	119	65	10x10 km grid (DMI, 2012)	868	Summer	1989 – 2010
LT	Merkys, Lithuania	4416	109	1 station	658	Summer	1961 – 1990
BE	Grote Nete, Belgium	383	32	6 stations	828	Summer	1986 – 2003
DE	Mulde, Germany	6171	414	43 stations	937	Summer	1951 – 2003

CZ2	Upper Metuje, Czech Republic	67	588	1x1 km grid (Šercl, 2008)	788	Summer	1980 – 2007
CZ1	Jizera, Czech Republic	2180	365	10 stations	860	Summer	1951 – 2003
PL	Nysa Kłodzka, Poland	1083	316	2 stations	589	Summer	1965 – 2000
TR	Gocbeylidere, Turkey	609	153	1 station	850	Autumn	1960 – 1990
CY	Yermasoyia, Cyprus	157	575	2 stations	640	Winter	1986 – 1997

Table 2 – Matrix of RCM-GCM combinations used in this study and source of the RCMs.

RCM\GCM	ECHAM5	BCM	HadCM3-Q3	HadCM3-Q16	HadCM3-Q0	ARPEGE	Institute
RM5.1						X	National Centre for Meteorological Research in France
RACMO2	X						Royal Netherlands Meteorological Institute
RCA	X	X	X				Swedish Meteorological and Hydrological Institute
REMO	X						Max Planck Institute for Meteorology
RCA3				X			Community Climate Change Consortium for Ireland
CLM					X		Swiss Federal Institute of Technology
HadRM3Q0					X		UK Met Office
HadRM3Q3			X				UK Met Office
HadRM3Q16				X			UK Met Office
HIRHAM5	X	X				X	Danish Meteorological Institute
RegCM3	X						International Centre for Theoretical Physics

Table 3 – Summary of the advantages and disadvantages of each statistical downscaling method. The name of the institution that undertook the downscaling work in this study is included in the first column. The advantages and/or disadvantages which are specific to the way the methods are applied in this application are stated.

SD method	Advantages	Disadvantages
Bias correction of mean, BCM (T. G. Masaryk Water Research Institute, Faculty of Environmental Sciences)	Easy to apply and little computer time required. Preserves the sequences of dry/wet days from the RCM. It accounts for different corrections in different time windows.	It only corrects the mean precipitation of the RCM.
Bias correction of mean and variance, BCMV (T. G. Masaryk Water Research Institute, Faculty of Environmental Sciences)	(same as bias correction of mean) It allows for distinct corrections between mean and variance.	The non-linear transformation may lead to unexpectedly large precipitation amounts. The autocorrelation from the RCM is not corrected, but it is affected by the bias correction approach.
Bias correction	Easy to apply and little computer time required.	The correction of the upper tail is based on relatively

quantile mapping, BCQM (NVE)	Preserves the sequences of dry/wet days from the RCM. Distinction between corrections in mean and extreme precipitation. The frequency of precipitation is corrected. No theoretical distribution is assumed.	few values (empirical distribution based). In this application, the same correction is applied for all seasons. The autocorrelation from the RCM is not corrected, but it is affected by the bias correction approach.
Expanded downscaling, XDS (U. Potsdam)	Generates realistic weather consistent with large-scale atmospheric patterns. Able to employ full range of predictor variables. It preserves co-variability between the predictands.	High demand for climate model accuracy; systematic biases can cause large errors. Requires large computation time and data preparation. No fully objective way of selecting the predictors.
Change factor of mean, CFM (DHI, DTU)	Easy to apply and little computer time required. It accounts for different changes in different months.	It only accounts for changes in mean precipitation. Does not account for changes in the length of dry/wet spells.
Change factor of mean and variance, CFMV (DHI, DTU)	(same as change factor of mean) Distinction between changes in mean and variance.	Does not account for changes in the length of dry/wet spells. The autocorrelation of precipitation may be disturbed. The non-linear transformation may lead to unexpectedly large precipitation amounts.

Change factor quantile mapping, CFQM (DTU)	(same as change factor of mean) Distinction between changes in mean and extreme precipitation. No theoretical distribution is assumed.	Does not account for changes in the length of dry/wet spells. The changes in the tails are based on relatively few values. The autocorrelation of precipitation may be disturbed.
Change factor quantile perturbation, CFQP (KU Leuven)	(same as change factor quantile mapping) Changes in the frequency of precipitation are accounted for.	The changes in the tails are based on relatively few values. The autocorrelation of precipitation may be disturbed (in this application, this is checked).

Table A1 – Percentage of the total variance explained by the GCM-RCM simulations (G+R) and SDMs (S) considering 2 and 3 sources of variance. The contribution of the GCMs and RCMs is shown in brackets.

		Winter		Summer	
	Nr. sources	G+R	S	G+R	S
NO2	2	68	32	51	49
	3	69 (29+40)	31	52 (14+38)	48
NO1	2	51	49	60	40
	3	51 (13+38)	49	61 (13+48)	39
DK	2	60	40	65	35
	3	62 (22+40)	38	67 (26+41)	33
LT	2	59	41	60	40
	3	57 (20+37)	43	57 (10+47)	43
BE	2	69	31	51	49
	3	71 (30+41)	29	52 (15+37)	48
DE	2	49	51	62	38
	3	51 (18+33)	49	61 (16+45)	39
CZ2	2	54	46	61	39
	3	55 (15+41)	45	57 (14+43)	43
CZ1	2	60	40	64	36
	3	58 (24+34)	42	59 (19+40)	41
PL	2	51	49	55	45
	3	48 (21+28)	52	50 (19+30)	50
TR	2	57	43	46	54
	3	55 (19+35)	45	42 (19+23)	58
CY	2	55	45		
	3	55 (21+34)	45		

Table A2 – Average RMSE from the comparison of the reconstructed and original values and the comparison with other combinations of GCM-RCM

RCM\GCM	Original	EPI_R	EPI_G
RCA-ECHAM5	0.47	0.60	0.61
HIRHAM-BCM	2.49	1.46	2.45

Figure 1 - Location of the eleven catchments studied.

Figure 2 – EPI estimated from the comparison of the downscaled time series for control and future period for 1 yr (light grey boxes) and 5 yr levels (dark grey boxes). The boxes indicate the 25, 50 and 75th percentiles and the whiskers the 5 and 95th percentiles. The circles show the median of all the values of EPI estimated from the comparison of the RCM outputs for the control and future periods. All the results are for a temporal aggregation of 1 day.

Figure 3 – In the top row, total variance decomposed in variance from GCMs, RCMs, SDMs and all the interaction terms (darkest to lighter grey colours). In the bottom row, percentage of the total variance explained by GCMs, RCMs, and SDMs (darkest to lighter grey colours). All the results are shown for 1 and 5 yr levels in the left and right column of each catchment, respectively. All the results are for a temporal aggregation of 1 day.

Figure 4 – EPI for each SDM for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. All results are for the 1 yr level and temporal aggregation of 1 day. Note the different scales used in the y-axis for winter and for summer.

Figure 5 - EPI for each SDM for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal aggregation. The same symbols are used for the different downscaling methods as in Fig. 4. Note the different scales used in the y-axis for winter and for summer.

Figure 6 – EPI estimated from the comparison of the observations and the downscaled time series by all BC methods for the control period for 1 yr (light grey boxes) and 5 yr levels (dark grey boxes). The boxes indicate the 25, 50 and 75th percentiles and the whiskers the 5 and 95th percentiles. The circles show the median of all the values of EPI estimated from the comparison of the observations and the uncorrected RCM outputs for the control period. All the results are for a temporal aggregation of 1 day.

Figure 7 - EPI for each BC method for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. All the results are for the 1 yr level and temporal aggregation of 1 day. Note the different scales used in the y-axis for winter and for summer.

Figure 8 - EPI for each BC method for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal aggregation. All the results are for 1 yr threshold. The same symbols are used for the different downscaling methods as in Fig. 7. Note the different scales used in the y-axis for winter and for summer.

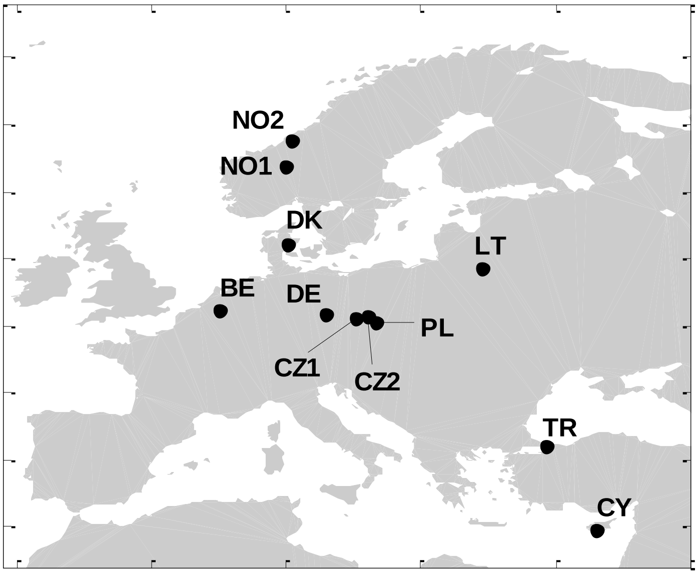

Figure 1 - Location of the eleven catchments studied.

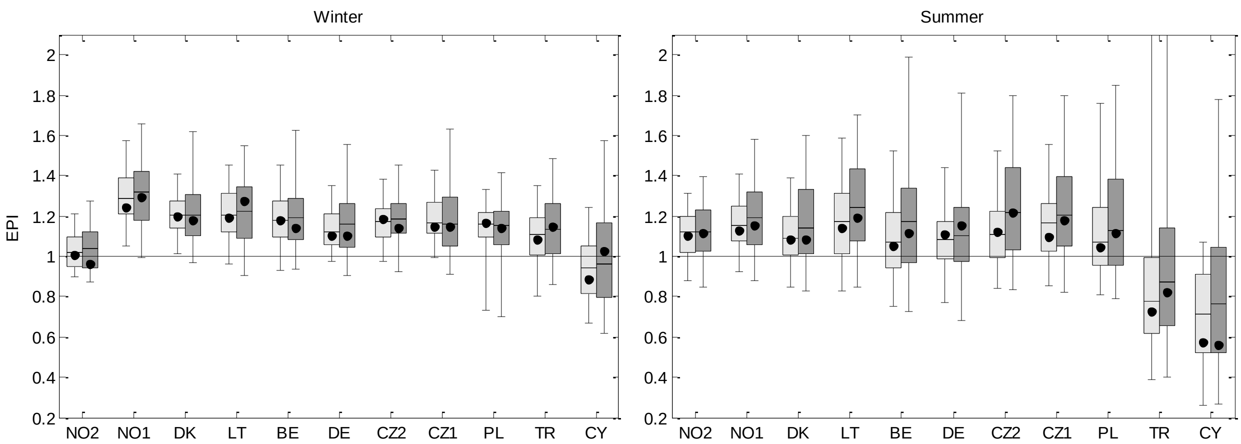

Figure 2 – EPI estimated from the comparison of the downscaled time series for control and future period for 1 yr (light grey boxes) and 5 yr levels (dark grey boxes). The boxes indicate the 25, 50 and 75th percentiles and the whiskers the 5 and 95th percentiles. The circles show the median of all the values of EPI estimated from the comparison of the RCM outputs for the control and future periods. All the results are for a temporal aggregation of 1 day.

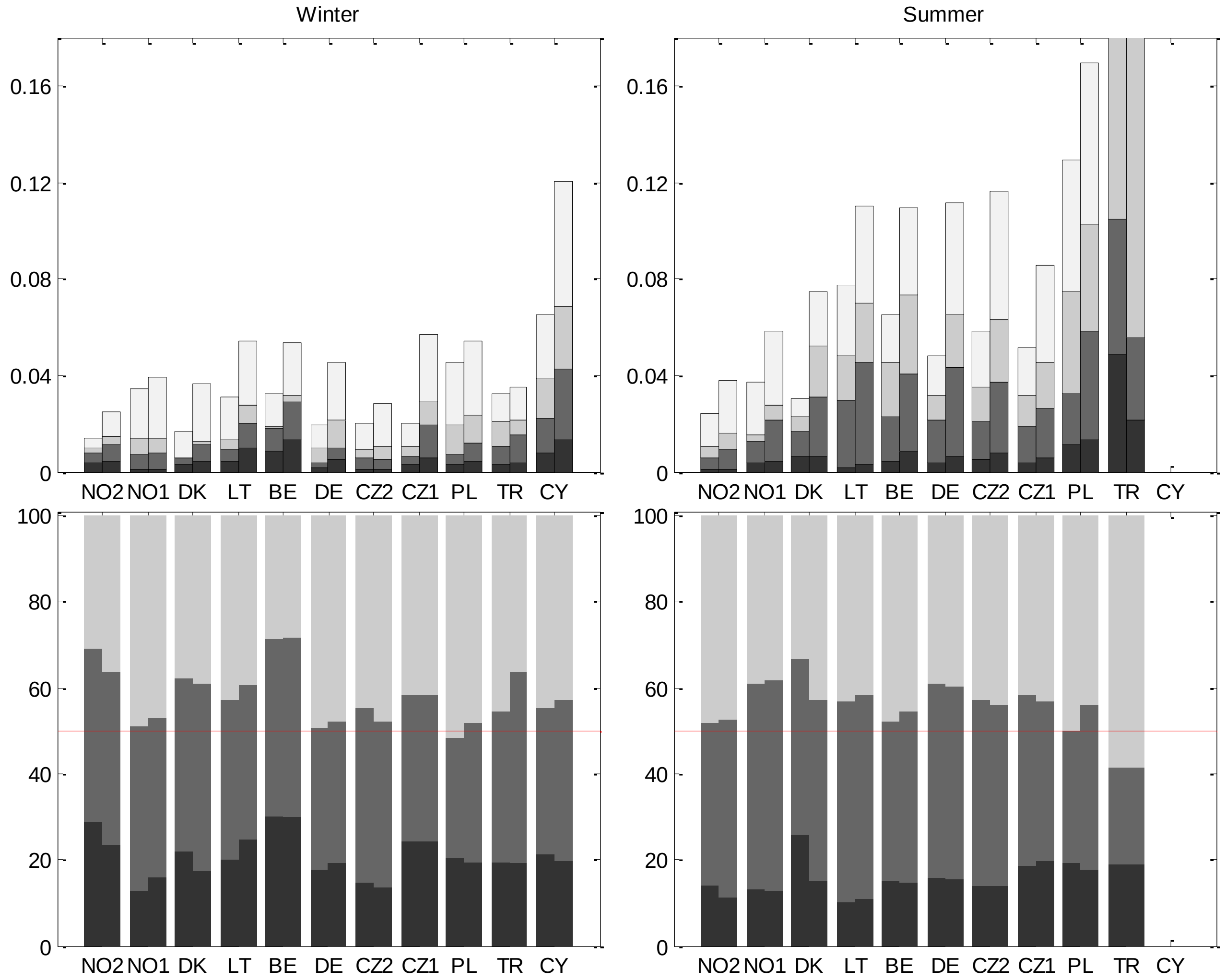

Figure 3 – In the top row, total variance decomposed in variance from GCMs, RCMs, SDMs and all the interaction terms (darkest to lighter grey colours). In the bottom row, percentage of the total variance explained by GCMs, RCMs, and SDMs (darkest to lighter grey colours). All the results are shown for 1 and 5 yr levels in the left and right column of each catchment, respectively. All the results are for a temporal aggregation of 1 day.

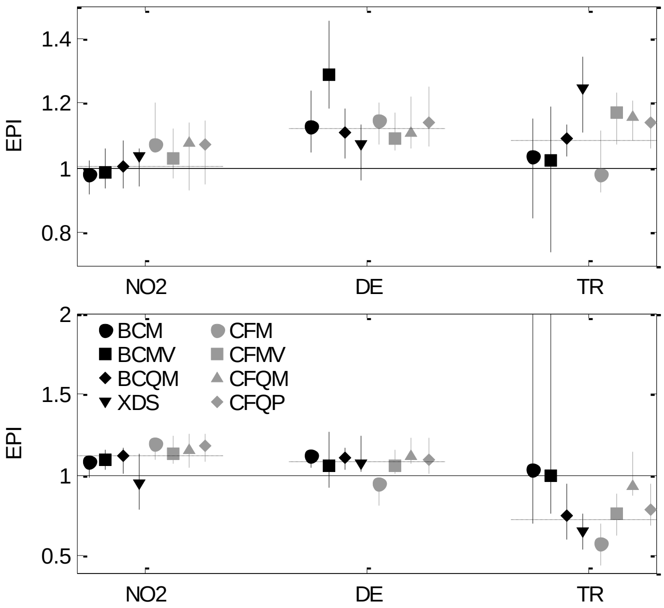

Figure 4 – EPI for each SDM for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. All results are for the 1 yr level and temporal aggregation of 1 day. Note the different scales used in the y-axis for winter and for summer.

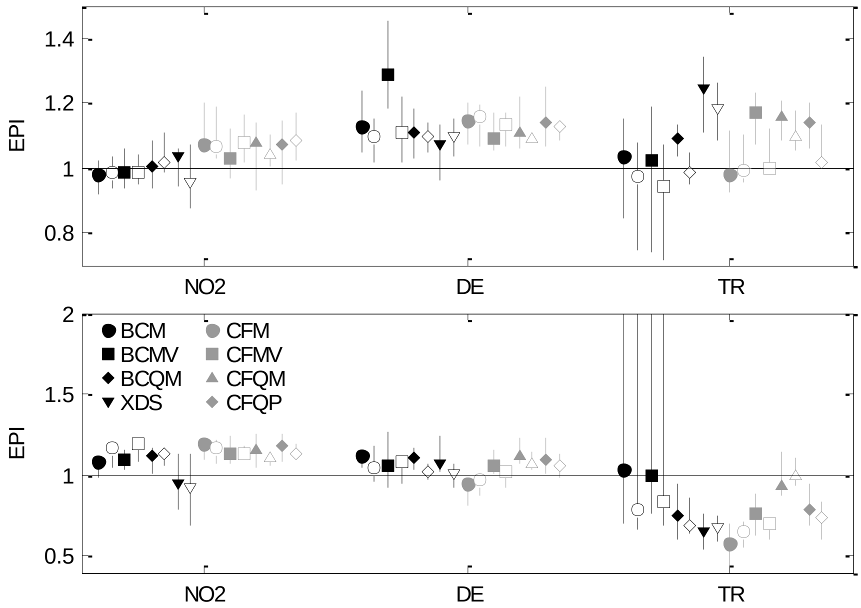

Figure 5 - EPI for each SDM for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal aggregation. The same symbols are used for the different downscaling methods as in Fig. 4. Note the different scales used in the y-axis for winter and for summer.

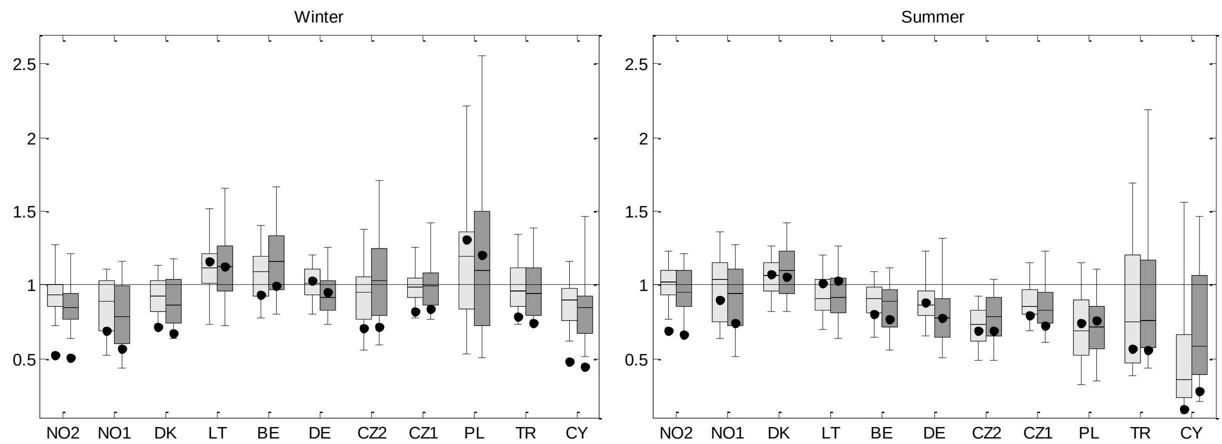

Figure 6 – EPI estimated from the comparison of the observations and the downscaled time series by all BC methods for the control period for 1 yr (light grey boxes) and 5 yr levels (dark grey boxes). The boxes indicate the 25, 50 and 75th percentiles and the whiskers the 5 and 95th percentiles. The circles show the median of all the values of EPI estimated from the comparison of the observations and the uncorrected RCM outputs for the control period. All the results are for a temporal aggregation of 1 day.

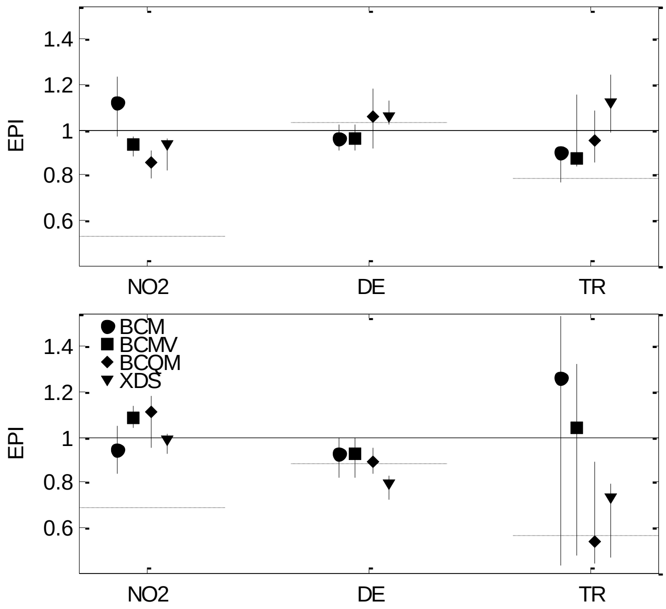

Figure 7 - EPI for each BC method for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. All the results are for the 1 yr level and temporal aggregation of 1 day. Note the different scales used in the y-axis for winter and for summer.

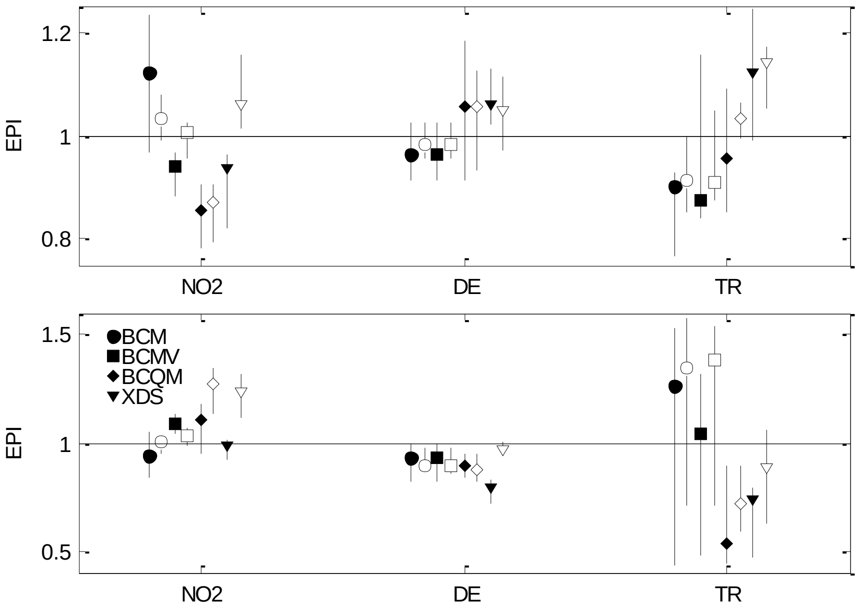

Figure 8 - EPI for each BC method for NO2, DE, and TR for winter (top) and summer (bottom). The markers indicate the median and the lines represent the range covered by the 25th and 75th percentiles. The results are shown for 1 day (filled markers) and 30 days (hollow markers) temporal aggregation. All the results are for 1 yr threshold. The same symbols are used for the different downscaling methods as in Fig. 7. Note the different scales used in the y-axis for winter and for summer.