Uncovering the shortcomings of a weather typing based statistical downscaling method

Els Van Uytven[1], Jan De Niel[1], Patrick Willems[1,2]

[1]Hydraulics Section, Department of Civil Engineering, KU Leuven, Kasteelpark Arenberg 40, 3001 Leuven, Belgium
5 [2]Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

Correspondence to: Els Van Uytven (els.vanuytven@kuleuven.be)

Abstract. In recent years many methods for statistical downscaling of the climate model outputs have been developed. Each statistical downscaling method (SDM) has strengths and limitations, but those are rarely evaluated. This paper proposes an approach to evaluate the skill of SDMs for the specific purpose of impact analysis in hydrology. The skill is evaluated by the
10 verification of the general statistical downscaling assumptions, and by the perfect predictor experiment that includes hydrological impact analysis. The approach has been tested for an advanced weather typing based SDM and for impact analysis on river peak flows in a Belgian river catchment. Significant shortcomings of the selected SDM were uncovered such as biases in the frequency of weather types and non-stationarities in the extreme precipitation distribution per weather type. Such evaluation of SDMs becomes of use for future tailoring of SDM ensembles to end user needs.

15 **1 Introduction**

For a 1.5°C temperature rise, the worldwide direct flood damage is estimated to increase in the range of 160% to 240% (Dottori et al., 2018). To minimize that potential impact, our society opts for two complementary strategies: climate mitigation and adaptation (IPCC, 2014). Consequently, vulnerability, impact and adaptation studies find ground in our society (Alfieri et al., 2016; Åström et al., 2016; Brekke et al., 2009; De Niel et al., 2018b; Termonia et al., 2018; Vansteenkiste et al., 2014a;
20 Vermuyten et al., 2018; Willems, 2013). These studies require projected hydro-meteorological time series, using the output of global climate models as primary information. However, the direct application of this output for impact modelling is challenged due to climate model biases (Kotlarski et al., 2014; Tabari et al., 2016) and, the mismatch in temporal and spatial resolutions between the climate model output and the time series required for impact modelling (Ochoa-Rodriguez et al., 2015; Salvadore et al., 2015). Therefore, statistical or dynamical downscaling is applied. The statistical downscaling approach bridges the
25 resolution gap through statistical relations between the predictors and predictand, whereas in the dynamical downscaling approach regional and limited area climate models (RCMs and LAMs) are developed. Despite the increased resolution of RCMs and LAMs, their climate model output remains biased and requires bias correction (Ehret et al., 2012; Maraun, 2016; Teutschbein and Seibert, 2012). Related to their underlying assumptions, both downscaling approaches have strengths and

limitations (Casanueva et al., 2016; Flaounas et al., 2013; Le Roux et al., 2018; Maraun et al., 2010; Vaittinada Ayar et al., 2016).

The statistical downscaling approach builds on four general assumptions (Benestad et al., 2008; Maraun and Widmann, 2018; Maraun et al., 2010; Schoof, 2013):

(1)   The informative assumption: a strong relation exists between the predictors and the predictand. It is of importance in the development of new statistical downscaling methods (SDMs), which requires the selection of predictors (Fu et al., 2018; Sachindra et al., 2018; Wilby and Wigley, 2000; Yang et al., 2017). Precipitation responds to large scale atmospheric circulation and thermo-dynamical laws (Emori and Brown, 2005; Kröner et al., 2017; Santos et al., 2016) and, hence, sea level pressure, geopotential height, relative humidity and/or (dew point) temperature are common predictors (Maraun and Widmann, 2018).

(2)   The perfect prognosis assumption: the predictors are adequately and accurately simulated by the climate models. The evaluation of this assumption is foremost performed under the name 'bias analysis'. Examples for a thorough bias analysis and the identification of underlying reasons can be found in the studies by (Anstey et al., 2013; Arakawa et al., 2014; Davini et al., 2017; Deser et al., 2012; Hartung et al., 2017; Hosseinzadehtalaei et al., 2018; Prein et al., 2015; Rybka and Tost, 2014; Tabari et al., 2016; Vanden Broucke et al., 2018; Watterson et al., 2014). These studies attribute the differences in the climate model skill to model resolution, parametrisation schemes and internal variability.

(3)   The greenhouse gas scenario sensitivity assumption: the predictors are sensitive to increasing greenhouse gas concentrations or, in other words, the predictors are greenhouse gas scenario sensitive. The greenhouse gas scenario sensitivity is measurable through trends in observed time series (Madsen et al., 2014). In climate model data, it is shown by the temporal evolution of the predictor changes (Maraun, 2013; Nguyen et al., 2018) and a comparison of the predictor changes for different representative concentration pathways (RCPs) (Jacob et al., 2014; KNMI, 2014; Van Uytven and Willems, 2018).

(4)   The stationarity assumption: the relation between the predictors and the predictand is time-invariant. Of all assumptions, this assumption is most difficult to validate (Hertig et al., 2017; Salvi et al., 2016; Wang et al., 2018). The difficulty arises from the internal variability of the climate system, which introduces anomalies in large scale circulation and precipitation (Merkenschlager et al., 2017; Ntegeka and Willems, 2008).

Alongside the main statistical downscaling assumptions, each SDM has method specific assumptions. They are encapsulated in the downscaling methodology, create the methods' strengths and limitations, and drive the statistical downscaling uncertainty contribution (Chen et al., 2011; Sunyer et al., 2015). An overview of commonly applied SDMs for precipitation downscaling and, their strengths and limitations is provided by (Hewitson et al., 2014; Maraun and Widmann, 2018; Maraun et al., 2010; Sunyer et al., 2015). In case of weather typing (WT) based SDMs, downscaling is performed in two steps. In the first step, by means of atmospheric classification systems (Philipp et al., 2016), the weather types (WTs) corresponding to the (mean sea level) pressure patterns are identified. In the second step, the relationship between the predictors (WTs) and

predictand (e.g. point precipitation) is established, using transfer functions, weather generators, re-sampling or analogues. Overall strengths emerge from the physically based WT characteristics, which represent the large scale atmospheric circulation and its changes. For analogues, limitations involve the analogues prerequisites, which divide the calibration data in smaller samples and therefore reduce the statistical reliability. Also, a historical analogue might be selected multiple times and, thus,

5 decrease the variance of the downscaled precipitation time series. Nevertheless, as climate projections indicate an intensification of extremes, the biggest limitation of analogues is that the produced precipitation amounts remain in the range of the historical data.

In this study, we evaluate the downscaling skill of a WT based SDM for climate change impact modelling on river peak flows

10 in a Belgian river catchment. The involved SDM is the WT based SDM "SD-B-7" developed by (Willems and Vrac, 2011). The downscaling skill is comprehensively evaluated through the individual verification of the general statistical downscaling assumptions and application of the perfect predictor experiment. The latter experiment has been extended by a hydrological impact analysis. In doing so, we respond to the recent focus of the statistical downscaling community on the validation of SDMs (Gutiérrez et al., 2018; Haberlandt et al., 2015; Hertig et al., 2018; Maraun et al., 2018, 2017, 2015; Pryor and Schoof,

15 2019; Soares et al., 2018; Werner and Cannon, 2016) and bring the validation into practice.

This study is organized as follows. Section 2 introduces the studied SDM, the selected river catchment and hydrological model applied for the rainfall-runoff simulation and, the observed and climate model simulated hydro-meteorological data. Section 3 outlines the evaluation and, corresponding results and discussions are included in Sect. 4. Section 5 summarizes the main

20 findings and highlights directions for future research.

2 Statistical downscaling method, case study and data

2.1 Statistical downscaling method

The SDM of study is the WT based SDM "SD-B-7" developed by (Willems and Vrac, 2011). The method downscales the daily gridded data to point time series with a sub-daily time step in three steps. In the first step, the Jenkinson-Collison

25 automated Lamb WT classification system is applied and the corresponding WTs are identified (Sect. 2.1.1). Next, by means of WT analogues, downscaled precipitation and potential evapotranspiration (ETo) time series are produced (Sect. 2.1.2). Finally, the precipitation amounts are scaled using the Clausius-Clapeyron relation (Sect. 2.1.3).

2.1.1 Jenkinson-Collison automated Lamb WT classification scheme

Based on the mean sea level pressure (MSLP) data for a 16 points grid centered around Uccle, the southerly, westerly and total

30 component of the geostrophic surface flow and shear velocity are calculated. Next, by inter-comparing these flow indices, 27 WTs are identified (Jenkinson and Collison, 1977; Jones et al., 1993; Philipp et al., 2016): 8 pure directional WTs (west 'W',

Hydrology and
Earth System
Sciences

Discussions

Open Access

northwest 'NW', north 'N', northeast 'NE', east 'E', southeast 'SE', south 'S' and southwest 'SW'), 2 non-directional WTs (cyclonic WT 'C' and anti-cyclonic WT 'A'), 16 hybrid WTs (8 hybrid cyclonic-directional WTs and 8 hybrid anti-cyclonic-directional WTs) and 1 undefined WT 'U'. The corresponding MSLP patterns can be found in (Brisson et al., 2011; Demuzere et al., 2009).

5

The 27 WTs are regrouped to 11 WTs by equally dividing the hybrid WTs over the corresponding non-directional WTs (C or A) and directional WTs (W, NW, N, NE, E, SE, S or SW). This might lead to information loss (Schiemann and Frei, 2010); however, it is in line with previous case studies for Belgium (Brisson et al., 2011; Demuzere et al., 2009; De Niel et al., 2017; Willems and Vrac, 2011).

10

A detailed method description (spacing and numbering of the grid, and WT classification scheme) is included in Supplementary Information – Sect. S1.

2.1.2 Statistical downscaling by analogues

Downscaled time series are produced by finding analogues for the climate model output. In the first step, the bias in the number

15 wet days is removed using a climate model and seasonal dependent wet day threshold. In the following step, the downscaled precipitation time series are constructed by WT analogues.

In the first place, an analogue wet day is based on the season and WT. Consider day 'd' of the climate model output for the scenario period, corresponding with season 's', WT 'wt' and a daily precipitation amount 'p'. Then, the search for an analogue

20 day is conducted among the historical wet days in season s for which the WT equals wt. The second criterion considers the exceedance probability of the daily precipitation amount p, which is calculated using the precipitation amounts of wet days occurring in season s and corresponding with the WT wt. As such, the analogue precipitation amount for day d equals the precipitation amount of the historical day with the closest exceedance probability. The sub-daily precipitation variation corresponds to the sub-daily variation of the analogue historical day. The downscaled ETo time series are produced by the

25 ETo amounts corresponding to the analogue days.

2.1.3 Precipitation scaling by the Clausius-Clapeyron relation

Precipitation responds to large scale circulation patterns and thermo-dynamical processes. In the SDM of this study, they are accounted for by the WT analogues and precipitation scaling by the Clausius-Clapeyron relation (CC relation), respectively. The CC relation describes the water holding capacity in air masses and, more specifically, increases by 7% per degree warming.

30 Application of this scaling rate to precipitation intensities is valid assuming that extreme precipitation amounts are controlled by local moisture availability and are not influenced by large scale atmospheric circulation patterns. However, in reality, physical processes interact and also higher scaling rates are found (Barbero et al., 2018; Blenkinsop et al., 2018; Manola et al.,

2018; Lenderink et al., 2017; Zhang et al., 2017). The CC relation is determined on annual time scale. The temperature rise, to be applied for the CC scaling, is computed using a seasonal quantile based approach.

2.2 Catchment characteristics and rainfall runoff modelling

The skill of the WT based SDM is evaluated for climate change impact analysis on river peak flows (Sect. 3). For this purpose,
5   the Grote Nete catchment, a catchment in the northeast of Belgium is selected. The catchment covers 402 km², of which 33% is cropland, 23% forest, 19% grass land, 19% urban and built-up area. Recently, (De Niel et al., 2018b) investigated the climate change impact on the catchments' hydrological extremes. They identified a minor uncertainty contribution by the hydrological models in the peak flow changes and therefore solely NAM, a lumped conceptual hydrological model for rainfall runoff simulation, is applied in this study (DHI, 2009; Nielsen and Hansen, 1973). The NAM model has been calibrated for the Grote
10 Nete catchment by (Vansteenkiste et al., 2014b).

Peak flows are extracted using the Peak-Over-Threshold method of (Willems, 2009). A previous analysis (De Niel et al., 2018a) showed that, due to a high soil moisture saturation, the peak flows in the selected catchment mainly occur in the winter season. Therefore, the evaluation focusses on the winter season, which covers the months December, January and February.

15 ## 2.3 Hydro-meteorological data

With exception to the calibration and validation of the hydrological model, for which catchment observations are employed, this study makes use of the observations for the main station of the Royal Meteorological Institute in Uccle (central Belgium; latitude = 50.8° and longitude = 4.2°). They are assumed to be representative for the selected catchment. For Uccle, precipitation, temperature and evapotranspiration (ETo) time series are available for the period 1901-2000 with a 10 minutes
20 and daily time step, respectively.

The historical WTs are identified using the daily gridded MSLP output for the ERA40, NCEP/NCAR and EMULATE re-analysis datasets (Table 1). Hence, this study accounts for the recent findings of (Horton and Brönnimann, 2018) and (Stryhal and Huth, 2017): both studies point to the uncertainties introduced by the re-analysis datasets in WT classification and
25 statistical downscaling.

The climate model ensemble, presented in Table 2, includes 93 CMIP5 climate model runs of which 33 unique control runs (Taylor et al., 2012). For the climate change impact analysis, all four representative concentration pathways (RCPs) are considered (van Vuuren et al., 2011), where the RCP 2.6, 4.5, 6.0 and 8.5 sub-ensembles include 20, 28, 15 and 30 climate
30 model runs, respectively. RCP 8.5 represents the business as usual scenario and corresponds with a 2.6 to 4.8°C global mean surface temperature rise (IPCC, 2014), whereas RCP 2.6 is a strong mitigation scenario for which the temperature rise by the end of the 21st century is constrained within the range 0.3-1.7°C (IPCC, 2014). For each climate model run, daily MSLP,

precipitation and average temperature output are extracted for 1961-1990 (reference/control period) and 2071-2100 (scenario period). The precipitation and temperature data are extracted for the grid cell covering Uccle, whereas MSLP, required for the WT identification, is extracted for a larger area covering Uccle (Supplementary Information – Sect. S1; Jenkinson and Collison, 1977; Jones et al., 1993; Philipp et al., 2016).

5 **3 Evaluation of the statistical downscaling assumptions**

3.1 The informative assumption

The informative assumption defines the existence of an informative relationship between the predictors and predictand. In other words, the predictors should be representative for the predictand and a physically based correlation should build this relation. In the SDM of this study, the involved predictors are the daily average temperature and WTs. By means of statistical
10 techniques, they bridge the scale gap with the local predictand variable, this is precipitation.

The relation between WTs and precipitation across different stations in Belgium has been studied by (Brisson et al., 2011). Here we analyze the relationship between WTs and precipitation from a different point of view, i.e. the evaluation of the downscaling skill, and in this context we extend the analysis by considering multiple re-analysis datasets and focusing on the
15 extreme precipitation amounts. In order to examine the informative assumption, the precipitation statistics related to the individual WTs are determined. The studied precipitation statistics involve the precipitation accumulation, number of wet days and empirical distribution of independent extreme precipitation amounts. The independent precipitation amounts are determined using a peak over threshold method and an inter-dependency time of minimum 12 hours (Willems, 2000). The statistics are calculated for the reference period 1961-1990 using WT and precipitation time series with a daily and 10 minutes
20 time step, respectively. By using daily WTs, rapidly occurring changes in the large scale atmospheric circulation might be neglected (Åström et al., 2016). However, the winter season is of study and for this season no rapidly evolving circulation changes, i.e. within 1 day, are expected.

The relation between temperature and precipitation is described by the CC relation. For this study, the CC relation is verified
25 for the independent 10 minutes independent precipitation amounts. The independent precipitation amounts are determined for the 100 years long times series using a peak over threshold method and a minimum interval between consequent events (Willems, 2000). For each event, the corresponding daily average temperature is determined. Next, following the principle described in (Manola et al., 2018), the precipitation amounts and corresponding temperatures are classified in moving temperature bins and per bin sorted from low to high. In the end, the magnification of the 90th, 95th, 99th and 99.9[th] percentile
30 precipitation amount over increasing temperature bins is investigated.

3.2 The perfect prognosis assumption

The application of WT analogues follows the principle of Perfect Prognosis (PP) SDMs. For this method type, a statistical relation between observed predictors and predictand is defined and, thereafter, applied to the climate model output (Maraun et al., 2010). Consequently, the calibrated predictor-predictand relationship is only valid for a bias free simulated climate model output. The prefect prognosis assumption therefore implies an adequate and accurate simulation of both the predictors ánd the predictor-predictand relationship. More specifically, a comparison is made between the climate model simulated and observed WT occurrences and relationship between the WTs and precipitation. The other step in the SDM, the scaling of the precipitation amounts by the CC relation, is according to the principles of Model Output Statistical (MOS) SDMs. Different to PP methods, MOS methods apply predictor changes. To check whether the predictor simulation results are adequate and accurate, a comparison is made between the climate model simulated and observed daily average temperature statistics. The PP assumption moreover implies a credible predictor response to the greenhouse gas scenarios. In other words, if the climate model output would be biased, the bias should be time-invariant and not influenced by the increasing greenhouse gas concentrations. This part of the perfect prognosis assumption is not evaluated in this paper.

The climate model results considered for these evaluations are the outputs for 33 unique control runs (Table 2) and are compared against the observed results for the reference period 1961-1990. We note that the choice of the reference period (control period) and thus the internal climate variability influences the evaluation of the perfect prognosis assumption (Deser et al., 2012; Fadhel et al., 2017), in particular the WTs' associated extreme precipitation amounts.

3.3 The greenhouse gas scenario assumption

Referring to the greenhouse gas scenario sensitivity assumption, the predictor should respond to the greenhouse gas scenarios. Furthermore, the response may depend on the greenhouse gas scenario considered and may be magnified under increasing greenhouse gas scenarios. In this context, we analyze the change in the WT occurrences and average daily temperature in function of the four RCPs. The change is based on the model output statistics for 2071-2100 in comparison to those for 1961-1990 for an ensemble of 93 global climate models runs (Table 2). The changes for different greenhouse gas scenarios are visually inspected and focus is put on the uni-directionality and magnification of the changes for increasing greenhouse gas scenarios.

We also assess how the predictor changes contribute to the predictand changes. As such the strength of the greenhouse gas scenario sensitivity is measured, in particular for the change in the average daily precipitation amount for wet winter days. The total change accounts for the WT occurrence changes, i.e. the synoptic contribution '$\Delta P_{synoptic}$' (Eq. (1)), and remaining changes 'ΔP_{other}', including the thermo-dynamic and local/mesoscale feedback changes (Eq. (2)) (Souverijns et al., 2016). For both contributions, we make use of the absolute occurrence frequency of WT 'j' in the reference period $N_{j,contr}$, the absolute

occurrence frequency of WT 'j' in the scenario period $N_{j,scen}$, the average daily precipitation amount of the wet winter days associated with WT 'j' for the reference period $P_{j,contr}$ and the average daily precipitation amount of wet winter days associated with WT 'j' for the scenario period $P_{j,scen}$.

$$\Delta P_{synoptic} = \sum_{j=1}^{11}(N_{j,scen} - N_{j,contr})P_{j,contr} \,, \tag{1}$$

$$\Delta P_{other} = \sum_{j=1}^{11}(P_{j,scen} - P_{j,contr})N_{j,scen} \tag{2}$$

The evaluation of the greenhouse gas scenario sensitivity assumption is carried out assuming the bias in the climate model output is stationary and not influenced by the increasing greenhouse gas concentrations (see Sect. 3.2).

3.4 The stationarity assumption

The stationarity assumption requires the predictor-predictand relationship to be time-invariant or stationary. For this, we split the observed period in different sub-periods. The sub-periods are 10 years long moving windows, starting in 1901 and sliding up to 1991 by one year time step. For each of these sub-periods, the independent extreme precipitation amounts are determined and those associated with the wetter WTs are extracted. The wetter WTs were defined as part of the evaluation of the informative assumption (see Sect. 3.1). Finally, a comparison is made between the extreme precipitation amounts for different periods. This evaluation builds further on the findings of (Ntegeka and Willems, 2008), demonstrating the presence of positive and negative anomalies in the extreme precipitation amounts of the Uccle time series. Since the identification of precipitation anomalies requires long time series, the analysis is only performed for the EMULATE dataset. Once more, especially given the small length of the moving windows, we note the influence of the internal variability on precipitation extremes and its influence on the results (Deser et al., 2012; Fadhel et al., 2017).

3.5 The method specific assumptions

The method specific assumptions are associated with the downscaling methodology and define the SDM uncertainty. They involve the incorporation of the predictors in the statistical function and the extrapolation of the predictand values under increasing greenhouse gas scenarios For the studied SDM, the WTs and average daily temperature are included in the downscaling methodology using analogues and the CC relation, respectively. At the same time, the CC relation enables the extrapolation of the predictand variables. In the context of assessing the method specific assumptions, the SDM is applied for the perfect predictor experiment. For this experiment, the observations are divided in calibration and validation periods. The validation period is defined as 1991-2000. This period is selected over the other periods with positive precipitation anomalies because of the climate change amplification at the end of the 20th century (Ntegeka and Willems, 2008). This requires the application of the CC relation and therefore allows the assessment of the CC relation in practice. Different calibration periods are defined using a 10 years wide moving window, moving between 1962 and 1986 by an annual time step. The calibration period and control period are chosen equally. As such, there is no bias between the observed and climate model simulated

predictor-predictand relation. Finally, the SDM of this study is applied to the pseudo-climate model ensemble to obtain downscaled series. Ideally, the extreme precipitation amounts and peak flow discharges of the downscaled series approximate the observed ones for the period 1991-2000. However, due to the method specific assumptions, statistical downscaling adds uncertainties and deviations are expected.

4 Results and discussions

4.1 Weather type occurrences

The results on the identification of the WTs show that, generally, regardless the season and the re-analysis dataset, the A WTs occur most frequently and represent approximately 30%. For the winter season, the second highest WT occurrence is shared between the W, SW and C WTs, and are each approximately 12% (Fig. 1). Despite for some details, there are no differences in WT occurrences across the re-analysis datasets. For instance, the W occurrences for the ERA40 dataset are smaller compared to those for EMULATE and NCEP/NCAR. The difference measures approximately 3% and is at the expense of a higher occurrence of the N and NE WTs. A second difference is found for the C and U WTs, where higher occurrences are found for the EMULATE and ERA40 datasets. The WT persistency is shown in Fig. S2 in Supplementary Information – Sect. S2. Typically, frequently occurring WTs are associated with a stronger persistence. Among the frequently occurring WTs, the persistency for the A WTs varies between 10 and 15 days, whereas for the W, SW and C WTs it varies around 5 days. With exception to the extreme persistency values for the A WTs, the re-analysis datasets introduce no or limited uncertainties to the results. Regarding the A WTs, the EMULATE dataset shows deviating WT persistency for return periods higher than 3 years; higher than for the other datasets.

The WT occurrence patterns are in agreement with the recent findings of (Otero et al., 2018), apart from small differences that may be due to differences in the size and composition of the re-analysis ensemble, and the applied WT classification system (Stryhal and Huth, 2017).

4.2 Evaluating the informative assumption

In order to evaluate the informative assumption, the relationship between the observed WTs (predictors) and precipitation (predictand) is studied. The considered predictand statistics are the precipitation accumulation, number of wet days and independent extreme precipitation amounts. The relative winter precipitation accumulation associated per WT for the period 1961-1990 is plotted in Fig. 3. Although the A WTs occur most frequently (Fig. 1), they do not account for the highest precipitation accumulation. For the winter season, W, SW and C represent the highest share and their share to the total winter precipitation accumulation varies between 22 and 29%, 18 and 23%, and 13 and 24%, respectively. Combined, these WTs contribute on average 65% of the total winter precipitation accumulation. Different to the WT occurrences, the choice of re-analysis datasets is of importance and introduces uncertainties up to 11%. Compared to the remaining WTs, the higher

precipitation accumulation for the W, SW and C WTs is due to their higher occurrence (see Fig. 1) and higher number of wet days. As shown in Fig. 4, 89 to 93% of the W days are wet. For the SW and C days, the wet day frequency is estimated between 80 and 85% and, 73 and 87%, respectively. Additionally, on top of the higher wet day frequency, these WTs are associated with higher precipitation amounts (Fig. 5). For instance, for the NCEP/NCAR re-analysis dataset, the daily precipitation

5 amount with a 1 year return period associated with the W WTs measures 0.51 mm/h and is twice the size of the corresponding amount for the A WTs, which measures 0.19 mm/h. For the EMULATE dataset, a similar order of differences is found. For the ERA 40 dataset, the differences among the distributions are smaller, however still present. Also the NW WTs are characterized by higher extreme precipitation amounts and higher wet day frequency. However, compared to the W, SW and C WTs, their occurrence is rather low and therefore drives the lower precipitation accumulation.

10

Based on the WT occurrences (Fig. 1) and associated precipitation statistics (Fig. 3, Fig. 4 and Fig. 5), we have identified drier and wetter WTs. The W, SW and C WTs rank among the wetter WTs, while the remaining WTs are identified as the drier WTs. The distinction among the WTs is also built by MSLP composites and Belgium's geographical location. Because of the neighboring North Sea and nearby location of the Atlantic Ocean from the southwestern up to the northern border, the SW,

15 W, NW and N WTs transport maritime air. On the other hand, land encloses the northeastern to southern border and therefore NE, E, SE and S air flows are of continental nature and thus drier. These results are in agreement with the earlier study by (Brisson et al., 2011) investigating the spatial variation of precipitation statistics and their relation with atmospheric WTs across different locations in Belgium. Next sections mainly focus on the wetter WTs.

20 The relationship between precipitation and temperature is presented in Fig. 6. This figure demonstrates the intensification of the independent precipitation amounts for increasing temperatures. For instance, the 90th percentile precipitation amount increases by 7% per degree and this increase follows the CC relation. For temperatures higher than 10°C, the scaling rate increases up to 14% per degree. Similar scaling rates are obtained for the higher precipitation percentiles. We note that only extreme precipitation amounts follow these scaling rates. The relationship between precipitation and temperature found in this

25 study is in accordance with the earlier studies for Belgium by (De Troch, 2016) and neighbouring regions (Lenderink and Van Meijgaard, 2008).

4.3 Evaluating the perfect prognosis assumption

As discussed in Sect. 3.2, the perfect prognosis assumption firstly implies a bias free predictor simulation by the climate models. In order to investigate this, the inter-comparison is made between the WT occurrences for the climate model runs and

30 re-analysis datasets for the reference period 1961-1990 (Fig. 1). The W, SW, A and C WTs are hereby grouped among the frequently occurring WTs. However, the comparison with the re-analysis datasets reveals large biases, in particular for the W and A WTs. The climate models overestimate the occurrence of W WTs by approximately 11%, whereas the A WTs are underestimated by 14%. Moreover, different to the re-analysis datasets, the W WTs are the foremost occurring WTs in the

climate model outputs. These findings are in agreement with the recent study by (Stryhal and Huth, 2018). Using different atmospheric classification patterns, this study points to an overall overestimation of the westerly circulation, which is estimated at approximately 7% for the British isles and increases towards central Europe up to 21%. (Stryhal and Huth, 2018) also point to the poor climate model performance in reproducing the occurrence of A WTs; they found differences up to 40%. Since the

5   spatial scale of WTs extend the minimum skilful scale of the climate models (Benestad et al., 2008), a better agreement between the climate models and re-analysis dataset simulated WT occurrences would be expected, but internal variability plays a role as well. Although it would be possible to remove the bias in WT occurrences, the SDM in this study does not do that. Note that such bias correction would require a technique that simultaneously accounts for the bias in the WT persistence (Fig. S2 in Supplementary Information – Sect. S2) and relationship with precipitation and other hydro-meteorological variables.

10

The perfect prognosis assumption additionally implies a bias free simulation of the predictor-predictand relationship. In this study, the predictor-predictand relationship is firstly described by the precipitation accumulation per WT. Figure 2 shows the relative precipitation accumulation over the individual WTs for the re-analysis datasets and climate model runs for the reference period (1961-1990). The climate model runs are capable of reproducing the general features, attributing the highest

15   precipitation accumulation to the W, SW and C WTs. For the W WTs, the bias in the associated winter precipitation accumulation ranges between 3 and 10%, between 1 and 6% for the SW WTs % and between -2 and -13% for the C WTs. In order to further examine the bias and its drivers, the number of wet days and the empirical distribution of daily precipitation amounts for the climate model runs are compared against the results for the re-analysis datasets in Fig. 3 and Fig. 5. For the W WTs, the corresponding number of wet days and extreme precipitation amounts are underestimated. The overestimation of

20   winter precipitation accumulation for W WTs is, therefore, due to the large overestimation of the W WT occurrences. In contrast, the overestimation in the precipitation accumulation for the C WTs is mainly due to the overestimation of the daily precipitation amounts. Compared to the other WTs, the bias in the winter precipitation accumulation for SW WTs remains rather small. This is due the combined effect of the overestimation of SW WT occurrences and the underestimation of the number of wet days. Theoretically, the biases in the predictor-predictand relationship are also removable, but this requires a

25   simultaneous correction of different precipitation statistics: from the number of wet days to the precipitation amounts.

Figure 7 compares the observed and climate model simulated daily average winter temperatures in function of their empirical return period for the period 1961-1990. The climate model median is representative for the observed statistics. Hence, different to the precipitation statistics (Tabari et al., 2015), the WT occurrences and the relationship between WTs and precipitation, no

30   additional bias correction is required.

4.4 Evaluating the greenhouse gas scenario sensitivity assumption

In order to examine the greenhouse gas scenario sensitivity of the predictors, we determine the WT occurrences for 2071-2100 and compare them to 1961-1990 in function of the forcing RCP (Fig. 8 – left sub-figure). Under the total uncertainty range, i.e. all RCPs combined, the occurrence of W WTs is projected to increase by 7%, whereas for the RCP sub-ensembles the

5   increase is magnified from 5% for RCP 2.6, to 6% for RCP 4.5, 21% for RCP 6.0 and 11% for RCP 8.5. The changes are magnified under increasing greenhouse gas scenario or RCP, except for the changes from RCP 6.0 to RCP 8.5. Also the SW WTs will occur more often, up to 11%. Also these changes are uni-directional and increase under increasing greenhouse gas scenarios, except for the changes from RCP 6.0 to RCP 8.5. These discontinuities in the uni-directionality of the changes may be explained by the smaller ensemble size for RCP 6.0 compared to the other RCPs, and/or by the different RCP sub-ensemble

10   compositions (Table 2) and differences in the internal variability among the climate models. While the W and SW WTs will occur more often, the occurrence of the A and C WTs is projected to decrease by 11% and 3%, respectively. Also these changes are sensitive to the forcing greenhouse gas scenario. In particular, the C WTs are projected to increase by 5% for RCP 2.6, -3% for RCP 4.5, -6% for RCP 6.0 and -5% for RCP 8.5. The decrease in the A WTs is magnified from 9% for RCP 2.6, to 10% for RCP 4.5, 14% for RCP 6.0 and 12% for RCP 8.5. Overall, the frequency changes depend on the forcing RCP and,

15   hence, the predictor (WTs and their corresponding occurrence) is sensitive to increasing greenhouse gas scenarios. Furthermore, we note that larger changes in the A occurrences typically occur with larger biases, while this is not the case for the W occurrences (Supplementary Information – Sect. S3).

Compared to the W, SW, C and A WTs, larger differences between the RCP 4.5 and RCP 8.5 sub-ensemble are found for the

20   less frequently occurring WTs (N, NE, E, SE, S and U WTs). Because of their low occurrence rates, the relative occurrence changes are strong. The strengthened W and SW WT occurrence and decreasing C WT occurrence are in line with previous studies (Otero et al., 2018; Plavcová and Kyselý, 2013). However, the change in the A WT occurrences differs with the findings of (Otero et al., 2018). This may be due to the considered climate model ensemble, the location of the 16 points grid for the WT classification system and, the reference and scenario periods.

25

Also the greenhouse gas scenario sensitivity of the temperature changes is assessed. Fig. 8 (right sub-figure) presents the changes in average daily temperatures for different return periods in function of the forcing RCP. For all return periods, the temperature change increases for increasing RCP. Moreover, there are no discontinuities in the uni-directionality of the changes. For instance, the average daily temperature for a 10 years return period increases from 1.1°C for RCP 2.6, to 1.7°C

30   for RCP 4.5, 2.4°C for RCP 6.0 and 3.5°C for RCP 8.5. Similar changes are found for the other return periods. On the other hand, the spread of the changes is return period dependent, where it obviously increases for increasing return period.

To further investigate the greenhouse gas sensitivity of the predictors and the predictor-predictand relation, the total change in the average daily winter precipitation amount is decomposed in function of its drivers (Fig. 9). The total change (median value) is estimated at 0.43mm/day, for which the synoptic changes (WT occurrence changes) contributed 20% of the total change. The majority of the total change is driven by other processes, including the thermo-dynamical processes, which explain 80% of the total change. This is in agreement with the findings of (Kröner, 2016), indicating the importance of the thermo-dynamical processes in the winter season for Western Europe.

4.5 Evaluating the stationarity assumption

For the evaluation of the stationarity assumption, the predictor-predictand relationship is determined for different sub-periods in the observed period and an inter-comparison is made among these relationships. The analysis is based on the findings of (Ntegeka and Willems, 2008), which highlighted the presence of positive and negative anomalies in the independent extreme precipitation amounts. The independent extreme precipitation amounts associated with the wetter WTs (W, SW and C individually, and all combined) have been determined for 10 years long moving windows starting between 1901 and 1991. In Fig. 10, the extreme precipitation changes are shown for the more recent moving windows and for the moving windows starting between 1931 and 1941. The latter period is related with a negative anomaly, which means that the extreme precipitation amounts are smaller than those for moving windows starting between 1981 and 1991, which are associated with a positive anomaly (Ntegeka and Willems, 2008). Other periods among the sub-periods could be selected for the analysis (Supplementary Information – Sect. S4), but here we have chosen to highlight periods showing a strong time dependent relationship. Alike earlier findings of (Ntegeka and Willems, 2008), which were not based on a selection of WTs, the extreme precipitation amounts for moving windows starting between 1931 and 1941 are smaller compared to those for the moving windows between 1981 and 1991. The difference between the positive and negative precipitation anomaly is especially visible for the W WTs and this for all aggregation levels between 10 minutes to 1 day. More specifically, the 10 minutes precipitation amounts with a 1 year return period measures 6mm/h for the negative anomaly, whereas it is 14mm/h for the positive anomaly. Also at the daily time scale, for a 1 year return period, the precipitation amount for the positive anomaly is approximately twice the amount for the negative anomaly. For the SW WTs, the difference is only present at the sub-daily aggregation levels and return periods larger than 1 year. For the C WTs, no differences appear between the amounts for the positive and negative anomaly. Finally, we note that uncertainties are introduced in the results for the higher return periods. They arise due to the rather short moving window and the strong effect of internal variability on the extreme precipitation amounts.

4.6 Evaluating the method specific assumptions

In order to evaluate the method specific assumptions, and thus the downscaling methodology, the SDM is applied for the perfect predictor experiment. For this experiment, sub-periods in the observed period are defined as pseudo climate model runs. Next, the SDM is applied to the pseudo climate model ensemble and downscaled time series are produced. Finally, a comparison is made between the observed and downscaled series, and the downscaling methodology is evaluated based on the

reproduction of the extreme precipitation amounts (Fig. 11) and peak flow discharges (Fig. 12). We verify the presence of the common limitations of WT analogues, involving the extrapolation of predictand values and the underestimation of the predictand variance.

5 The median results of the climate model ensemble are close to the observed extreme precipitation amounts for all re-analysis datasets and aggregation levels (Fig. 11). Since the downscaling methodology assumes independency among consequent days and WT analogues have limitations regarding the predictand extrapolation, these results are remarkable. In particular, because of the assumed independency, an underestimation of the extreme precipitation amounts, aggregated over at least 1 day, is expected. Moreover, because of analogues, one would expect an overall underestimation of the extreme precipitation amounts.

10 Both types of underestimation are cancelled by precipitation scaling following the CC relation (Supplementary Information – Sect. S5), where the scaling is especially visible for the extreme 10 minutes precipitation amounts (Fig. 11). For that aggregation level, the highest observed extreme precipitation amounts are systematically lower than the downscaled ones. This might be due to randomness, which is induced by the short length of the time series. Furthermore, the CC relation removes the underestimation of the variance of the daily precipitation amounts, which is induced by the limited sampling variability among

15 analogue days (Supplementary Information – Sect. S5).

Different to the extreme precipitation amounts, the peak flow probability distribution is not accurately reproduced for any of the re-analysis datasets (Fig. 12). In particular, for empirical return periods smaller than 1 year, the downscaled time series underestimate the observed peak flow discharges. This underestimation reaches a maximum deviation of $2 m^3/s$ and takes place

20 at a return period equal to 0.4 year. An unique reason for this underestimation could not be identified. We expect contributions from the seasonal calibration of the SDM, the rather short data range for the SDM calibration and the sequencing of precipitation events and corresponding precipitation amounts. Besides precipitation, the dry periods are also of importance. They influence the soil moisture balance of the hydrological model and therefore might contribute to the peak flow underestimation. For higher return periods, the median of the pseudo climate model ensemble reproduces the observed peak

25 flow discharges. Furthermore, alike the reproduced precipitation amounts (Fig. 11), the spread for the pseudo climate model ensemble increases for increasing return period. The largest spread is found for the NCEP/NCAR re-analysis dataset.

Nor for the reproduced extreme precipitation amounts, neither for the peak flow discharges, systematic differences are found across the re-analysis datasets.

30 **5 Conclusions and outlook**

The first statistical downscaling assumption, i.e. the informative assumption, implies a strong relation between the predictor and predictand. In the SDM of this study, the predictand is precipitation and the predictors are the Jenkinson-Collison Lamb

WTs and temperature. Among all WTs, the W, C, SW and A WTs are identified as frequently occurring. More specifically, 30% of all winter days are accompanied by A WTs, whereas the other WTs occur each approximately 12%. Apart from the A WTs, these WTs are associated with a high precipitation accumulation and together explain up to 71% of the total winter precipitation amount. Respectively 91%, 83% and 79% of W, SW and C occurrences are wet days; they moreover relate to

5   higher precipitation extremes. The WTs are found representative and informative for the target predictand variable, precipitation. The choice of the re-analysis dataset does not influence the WT occurrence. However, when examining the predictor-predictand relationship, the re-analysis datasets show differences, for which no systematic grouping is found. Also the relationship between temperature and precipitation is found informative. More specifically, extreme precipitation amounts are magnified under increasing temperatures following the CC relation (+7%/°C). For higher temperatures, the scaling rate

10   evolves into the super CC relation rate (+14%/°C).

The second assumption, i.e. the perfect prognosis assumption, firstly presupposes a bias free and adequate predictor simulation. The climate model runs generally reproduce the WT occurrence pattern. However, large biases are found for the W and A occurrences. The occurrence of W WTs is overestimated by 11% at the expense of an underestimation of the A WTs. Due to

15   the magnitude of the bias, the W WTs become the foremost occurring WT in the climate model outputs, while in the re-analysis datasets this is the A WT. The perfect prognosis assumption also implies an adequate and accurate simulation of the predictor-predictand relation. The strong biases in the WT occurrences influence the biases in the predictor-predictand relation. As such, the overestimation of the precipitation accumulation for the W WTs is due to the overestimation of the W occurrences. Also the underestimation of the precipitation accumulation for the A WTs is driven by a bias in the occurrence frequency. This bias

20   is thereafter amplified by the underestimation of the wet day frequency. In contrast, for the C WTs, for which no bias in the WT occurrence nor wet day frequency is found, the bias in the associated precipitation accumulation is due to an overestimation of the daily (extreme) precipitation amounts. Since biases are identifiable, the perfect prognosis assumption is not valid, but bias correction may be possible. On the contrary, for temperature, no significant biases are identified, implying a valid perfect prognosis assumption.

25

The third statistical downscaling assumption concerns the predictor response to increasing RCP greenhouse gas scenarios. The WT occurrences and average daily temperature statistics are determined for 2071-2100 and are compared to the results for 1961-1990. The W WTs are projected to increase by 6% for RCP 4.5 and 11% for RCP 8.5, the SW WTs by 10% for RCP 4.5 and 14% for RCP 8.5, the A WTs by -10% for RCP 4.5 and -12% for RCP 8.5 and, the C WTs by -3% for RCP 4.5 and -5%

30   for RCP 8.5. The changes in the WT occurrences are generally uni-directional: they are magnified under increasing greenhouse gas scenarios. They are moreover in line with the 'wet becomes wetter' winter projections for Western Europe. Also for temperature, the change increases for increasing RCP. More specifically, for RCP 2.6 the change measures 1.3°C and increases up to 1.9°C for RCP 4.5, 2.1°C for RCP 6.0 and 3.3°C for RCP 8.5. Consequently, both predictors meet the greenhouse gas scenario sensitivity assumption. Due to greenhouse gas scenario sensitivity behavior, the changes in WT occurrences explain

20% of the change in the average precipitation amount of wet winter days, while the temperature and local/mesoscale scale changes explain 80%. These unbalanced contributions demonstrate the importance of especially temperature as predictor for precipitation downscaling.

5   Finally, the fourth assumption requires the predictor-predictand relationship to be time-invariant or stationary. However, we found the extreme precipitation amounts, related to the wetter WTs, to be time dependent. The end of the 20th century correlates with a positive anomaly of extreme precipitation amounts, meaning that the extreme precipitation amounts for this period are higher than the overall average of extreme precipitation amounts. Moreover, this positive anomaly is expected to be strengthened by climatic changes. For other periods, such as the period 1931-1950, a negative anomaly of extreme
10   precipitation amounts is detected. The stationarity assumption is known as the main caveat of the statistical downscaling approach. Therefore, the failure against the stationarity assumption is not surprising.

In addition, the method specific assumptions were evaluated in the perfect predictor experiment. The downscaled time series can accurately reproduce the extreme precipitation amounts. Hence, the application of the CC relation and the prerequisites
15   for an analogue day are found effective. However, for the peak flows, only the tail of the peak flow discharges, ranging from 1 year to 10 years, is accurately reproduced. For the smaller return periods, the downscaled time series underestimate the observed peak flows. This underestimation might be induced by the sequences of dry periods, which influence the moisture balance of the hydrological model and therefore the peak flows.

20   The studied SDM does not meet all assumptions. It is shown that the SDM has limitations and its skill should be improved. Improvements should especially address the failing perfect prognosis assumption. Since the large scale atmospheric circulation in RCMs remains biased (Addor et al., 2016; Jury et al., 2018), we suggest bias correcting WT occurrences before statistical downscaling. However, the number of WT based bias correction methods remains limited (Photiadou et al., 2016). Moreover, unless taken into account, bias correction techniques may disturb the WT persistency and correlation among variables.
25   Uncovering the limitations of the applied WT based SDM does not mean that we discourage its' use; one should not forget that other SDMs may have other types of shortcomings and biases (Chen et al., 2011; Sunyer et al., 2015; Haberlandt et al., 2015). By considering an ensemble of SDMs, the uncertainties introduced by these limitations can be taken into account.

It is also important to mention that the above findings depend on the case study and thus the choice of the SDM and river
30   catchment. However, the evaluation is general and therefore applicable to other case studies. Ideally and assuming the skill is time-invariant, the SDM skill is evaluated for the current climate and thereafter accounted for in the SDM ensemble. When SDM ensembles would be considered, ensemble members could be weighted based on their skill in the perfect predictor experiment. We note the resemblance with existing climate model weighing techniques (Sanderson et al., 2017). A first step towards a weighted SDM ensemble is still to be made by the statistical downscaling community.

Acknowledgments

5   **Author contributions**

EVU conceptualised and developed the approach to evaluate the skill of statistical downscaling methods. Under the supervision of PW, EVU and JDN performed the formal analysis and investigation. EVU prepared the visualisation and wrote the initial draft, which was critically reviewed and revised by PW and JD.

Data availability

10   We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in Table 2) for producing and making available their model output. The climate model data is available through the website of the Earth System Grid Federation, https://esgf.llnl.gov/. The ERA40 re-analysis data set is available through the ECMWF Public Datasets Web Interface http://apps.ecmwf.int/datasets/, the EMULATE data set through https://www.metoffice.gov.uk/hadobs/emslp/ and the NCEP/NCAR data set through

15   https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.re-analayis.html. The observed precipitation time series for the Uccle station were made available by Royal Meteorological Institute.

References

Addor, N., Rohrer, M., Furrer, R., and Seibert, J.: Propagation of biases in climate models from the synoptic to the regional scale: Implications for bias adjustment, J. Geophys. Res. Atmos., 121, 2075–2089, doi: 10.1002/2015JD024040, 2016.

Alfieri, L., Feyen, L., and Di Baldassarre, G.: Increasing flood risk under climate change: a pan-European assessment of the benefits of four adaptation strategies, Clim. Change, 136(3-4), 507-521, https://doi.org/10.1007/s10584-016-1641-1, 2016.

Ansell, T.J., Jones, P.D., Allan, R.J., Lister, D., Parker, D.E., Brunet, M., Moberg, A., Jacobeit, J., Brohan, P., Rayner, N.A., Aguilar, E., Alexandersson, H., Barriendos, M., Brandsma, T., Cox, N.J., Della-Marta, P.M., Drebs, A., Founda, D., Gerstengarbe, F., Hickey, K., Jónsson, T., Luterbacher, J., Nordli, Ø., Oesterle, H., Petrakis, M., Philipp, A., Rodwell, M.J., Saladie, O., Sigro, J., Slonosky, V., Srnec, L., Swail, V., García-Suárez, A.M., Tuomenvirta, H., Wang, X., Wanner, H., Werner, P., Wheeler, D., and Xoplaki, E.: Daily Mean Sea Level Pressure Reconstructions for the European–North Atlantic Region for the Period 1850–2003, J. Climate, 19, 2717–2742, https://doi.org/10.1175/JCLI3775.1, 2006.

Anstey, J.A., Davini, P., Gray, L.J., Woollings, T.J., Butchart, N., Cagnazzo, C., Christiansen, B., Hardiman, S.C., Osprey, S.M., and Yang, S.: Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution, J. Geophys. Res. Atmos., 118, 3956–3971, https://doi.org/10.1002/jgrd.50231, 2013.

Arakawa, A.: The Cumulus Parameterization Problem: Past, Present, and Future, J. Climate, 17, 2493–2525, https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2, 2004.

Åström, H.L.A., Sunyer, M., Madsen, H., Rosbjerg, D., and Arnbjerg-Nielsen, K.: Explanatory analysis of the relationship between atmospheric circulation and occurrence of flood-generating events in a coastal city, Hydrol. Process., 30, 2773–2788, https://doi.org/10.1002/hyp.10767, 2016.

Barbero, R., Westra, S., Lenderink, G., and Fowler, H.J.: Temperature-extreme precipitation scaling: a two-way causality?, Int. J. Climatol., 38, e1274–e1279, https://doi.org/10.1002/joc.5370, 2018.

Benestad, R.E., Hanssen-bauer, I., and Chen, D.: Empirical statistical downscaling, https://doi.org/10.1142/6908, 2008.

Blenkinsop, S., Fowler, H.J., Barbero, R., Chan, S.C., Guerreiro, S.B., Kendon, E., Lenderink, G., Lewis, E., Li, X.-F., Westra, S., Alexander, L., Allan, R.P., Berg, P., Dunn, R.J.H., Ekström, M., Evans, J.P., Holland, G., Jones, R., Kjellström, E., Klein-Tank, A., Lettenmaier, D., Mishra, V., Prein, A.F., Sheffield, J., and Tye, M.R.: The INTENSE project: using observations and models to understand the past, present and future of sub-daily rainfall extremes, Adv. Sci. Res., 15, 117–126, https://doi.org/10.5194/asr-15-117-2018, 2018.

Brekke, L.D., Maurer, E.P., Anderson, J.D., Dettinger, M.D., Townsley, E.S., Harrison, A., and Pruitt, T.: Assessing reservoir operations risk under climate change, Water Resour. Res. 45, 1–16, https://doi.org/10.1029/2008WR006941, 2009.

Brisson, E., Demuzere, M., Kwakernaak, B., and Van Lipzig, N.P.M.: Relations between atmospheric circulation and precipitation in Belgium, Meteorol. Atmos. Phys., 111, 27–39, https://doi.org/10.1007/s00703-010-0103-y, 2011.

Casanueva, A., Herrera, S., Fernández, J., and Gutiérrez, J.M.: Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative, Clim. Change, 137, 411–426, https://doi.org/10.1007/s10584-016-1683-4, 2016.

Chen, J., Brissette, F.P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of climate change on
5   hydrology, J. Hydrol., 401, 190–202, https://doi.org/10.1016/j.jhydrol.2011.02.020, 2011.

Davini, P., Corti, S., D'Andrea, F., Rivière, G., and von Hardenberg, J.: Improved Winter European Atmospheric Blocking Frequencies in High-Resolution Global Climate Simulations, J. Adv. Model. Earth Syst., 9, 2615–2634, https://doi.org/10.1002/2017MS001082, 2017.

Demuzere, M. , Werner, M. , van Lipzig, N. P., and Roeckner, E.: An analysis of present and future ECHAM5 pressure fields
10   using a classification of circulation patterns, Int. J. Climatol., 29, 1796-1810, https://doi.org/10.1002/joc.1821, 2009.

De Niel, J., Van Uytven, E., and Willems, P.: On the correlation between precipitation and potential evapotranspiration climate change signals for hydrological impact analyses, under review, 2018a.

De Niel, J., Van Uytven, E., and Willems, P.: Uncertainty analysis of climate change impact on river flow extremes based on a large multi model ensemble, under review, 2018b.

15   De Niel, J., Demarée, G., and Willems, P.: Weather Typing-Based Flood Frequency Analysis Verified for Exceptional Historical Events of Past 500 Years Along the Meuse River, Water Resour. Res., 53, 8459–8474, https://doi.org/10.1002/2017WR020803, 2017.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: The role of internal variability, Clim. Dyn., 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, 2012.

20   De Troch, R.: The application of the ALARO-0 model for regional climate modeling in Belgium, Ugent & KMI, 2016.

DHI: Mike 11 software manual, 2009.

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R.A., and Feyen, L.: Increased human and economic losses from river flooding with anthropogenic warming., Nat. Clim. Chang., https://doi.org/10.1038/s41558-018-0257-z, 2018.

25   Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391-3404, https://doi.org/10.5194/hess-16-3391-2012, 2012.

Emori, S. and Brown, S.J.: Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate, Geophys. Res. Lett., 32, 1–5, https://doi.org/10.1029/2005GL023272, 2005.

30   Fadhel, S., Rico-Ramirez, M.A., Han, D.: Uncertainty of Intensity–Duration–Frequency (IDF) curves due to varied climate baseline periods, J. Hydrol., 547, 600–612, https://doi.org/10.1016/j.jhydrol.2017.02.013, 2017.

Flaounas, E., Drobinski, P., Vrac, M., Bastin, S., Lebeaupin-Brossier, C., Stéfanon, M., Borga, M., and Calvet, J.C.: Precipitation and temperature space-time variability and extremes in the Mediterranean region: Evaluation of dynamical and statistical downscaling methods, Clim. Dyn., 40, 2687–2705, https://doi.org/10.1007/s00382-012-1558-y, 2013.

Fu, G., Charles, S.P., Chiew, F.H.S., Ekström, M., and Potter, N.J.: Uncertainties of statistical downscaling from predictor selection: Equifinality and transferability, Atmos. Res., 203, 130–140, https://doi.org/10.1016/j.atmosres.2017.12.008, 2018.

Gutiérrez, J.M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua,

5   J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M.J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D.E., Fischer, A.M., Cardoso, R.M., Soares, P.M.M., Czernecki, B., and Pagé, C.: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment, Int. J. Climatol., 1–36, https://doi.org/10.1002/joc.5462, 2018.

Haberlandt, U., Belli, A., and Bárdossy, A.: Statistical downscaling of precipitation using a stochastic rainfall model

10   conditioned on circulation patterns - an evaluation of assumptions, Int. J. Climatol., 35, 417–432, https://doi.org/10.1002/joc.3989, 2015.

Hartung, K., Svensson, G., and Kjellström, E.: Resolution, physics and atmosphere–ocean interaction – How do they influence climate model representation of Euro-Atlantic atmospheric blocking?, Tellus A Dyn. Meteorol. Oceanogr., https://doi.org/10.1080/16000870.2017.1406252, 2017.

15   Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J.M., Wibig, J., Casanueva, A., and Soares, P.M.M.: Comparison of statistical downscaling methods with respect to extreme events over Europe: Validation results from the perfect predictor experiment of the COST Action VALUE, Int. J. Climatol., 1–22, https://doi.org/10.1002/joc.5469, 2018.

Hertig, E., Merkenschlager, C., and Jacobeit, J.: Change points in predictors–predictand relationships within the scope of statistical downscaling, Int. J. Climatol., 37, 1619–1633, https://doi.org/10.1002/joc.4801, 2017.

20   Hewitson, B.C., Daron, J., Crane, R.G., Zermoglio, M.F., and Jack, C.: Interrogating empirical-statistical downscaling, Clim. Change, 122, 539–554, https://doi.org/10.1007/s10584-013-1021-z, 2014.

Horton, P. and Brönnimann, S.: Impact of global atmospheric reanalyses on statistical precipitation downscaling, Clim. Dyn., https://doi.org/10.1007/s00382-018-4442-6, 2018.

Hosseinzadehtalaei, P., Tabari, H., and Willems, P.: Precipitation intensity–duration–frequency curves for central Belgium

25   with an ensemble of EURO-CORDEX simulations, and associated uncertainties, Atmos. Res., 200, 1–12, https://doi.org/10.1016/j.atmosres.2017.09.015, 2018.

IPCC, Climate Change 2014 Synthesis Report Summary Chapter for Policymakers, IPCC, https://doi.org/10.1017/CBO9781107415324, 2014.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski,

30   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., van 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, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Jenkinson, A.F. and Collison, F.P.: An initial climatology of gales over the North Sea, Synoptic Climatology Branch Memorandum 62, Meteorological Office, London. 1977.

Jones, P.D., Hulme, M., and Briffa, K.R.: A comparison of Lamb circulation types with an objective classification scheme, International Journal of Climatology, 13, 655–663, 1993.

5 Jury, M.W., Herrera, S., Gutiérrez, J.M., and Barriopedro, D.: Blocking representation in the ERA-Interim driven EURO-CORDEX RCMs, Clim. Dyn., https://doi.org/10.1007/s00382-018-4335-8, 2018.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Amer. Meteor. Soc., 77, 437–472,

10 https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

KNMI: KNMI'14 climate scenarios for the Netherlands, KNMI, 2014.

Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., Van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7,

15 1297–1333, https://doi.org/10.5194/gmd-7-1297-2014, 2014.

Kröner, N., Kotlarski, S., Fischer, E., Lüthi, D., Zubler, E., and Schär, C.: Separating climate change signals into thermodynamic, lapse-rate and circulation effects: theory and application to the European summer climate, Clim. Dyn., 48, 3425–3440, https://doi.org/10.1007/s00382-016-3276-3, 2017.

Kröner, N.: Identifying and Quantifying Large-Scale Drivers of European Climate Change, ETH Zürich,

20 https://doi.org/10.3929/ethz-a-010793497, 2016.

Le Roux, R., Katurji, M., Zawar-Reza, P., Quénol, H., and Sturman, A.,: Comparison of statistical and dynamical downscaling results from the WRF model, Environ. Model. Softw., 100, 67–73, https://doi.org/10.1016/j.envsoft.2017.11.002, 2018.

Lenderink, G., Barbero, R., Loriaux, J.M., and Fowler, H.J.: 2017. Super-Clausius-Clapeyron scaling of extreme hourly convective precipitation and its relation to large-scale atmospheric conditions, J. Clim., 30, 6037–6052,

25 https://doi.org/10.1175/JCLI-D-16-0808.1, 2017.

Lenderlink, G. and van Meijgaard, E.: Increase in hourly precipitation extremes beyond expectations from temperature changes, Nat. Geosci., 1, 511–514, https://doi.org/10.1038/ngeo262, 2008.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T.R.: Review of trend analysis and climate change projections of extreme precipitation and floods in Europe, J. Hydrol., 519, 1–48, https://doi.org/10.1016/j.jhydrol.2014.11.003,

30 2014.

Manola, I., van den Hurk, B., De Moel, H., and Aerts, J. C. J. H.: Future extreme precipitation intensities based on a historic event, Hydrol. Earth Syst. Sci., 22, 3777-3788, https://doi.org/10.5194/hess-22-3777-2018, 2018.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research, https://doi.org/10.1017/9781107588783, 2018.

Maraun, D. , Widmann, M. and Gutierrez, J. M.: Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment, Int J Climatol., https://doi.org/10.1002/joc.5877, 2018.

Maraun, D., Huth, R., Gutiérrez, J.M., Martín, D.S., Dubrovsky, M., Fischer, A., Hertig, E., Soares, P.M.M., Bartholy, J., Pongrácz, R., Widmann, M., Casado, M.J., Ramos, P., and Bedia, J.: The VALUE perfect predictor experiment: evaluation of temporal variability, Int. J. Climatol., https://doi.org/10.1002/joc.5222, 2017.

Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Curr. Clim. Chang. Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Maraun, D., Widmann, M. , Gutiérrez, J. M., Kotlarski, S. , Chandler, R. E., Hertig, E. , Wibig, J. , Huth, R. and Wilcke, R. A., VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 1-14, https://doi.org/10.1002/2014EF000259, 2015.

Maraun, D.: When will trends in European mean and heavy daily precipitation emerge?, Environmental Research Letters, https://doi:10.1088/1748-9326/8/1/014004, 2013.

Maraun, D., Wetterhall, F., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, 1–38, https://doi.org/10.1029/2009RG000314, 2010.

Merkenschlager, C., Hertig, E., and Jacobeit, J.: Non-stationarities in the relationships of heavy precipitation events in the Mediterranean area and the large-scale circulation in the second half of the 20th century, Glob. Planet. Change, 151, 108–121, https://doi.org/10.1016/j.gloplacha.2016.10.009, 2017.

Nguyen, TH., Min, SK., Paik, S., and Lee, D.: Time of emergence in regional precipitation changes: an updated assessment using the CMIP5 multi-model ensemble, Clim Dyn., https://doi.org/10.1007/s00382-018-4073-y, 2018.

Nielsen, S.A. and Hansen, E.: Numerical simulation of the rainfall-runoff process on a daily 3 basis, Nord Hydrol, 4, 171–190, 1973.

Ntegeka, V. and Willems, P.: Trends and multidecadal oscillations in rainfall extremes, based on a more than 100-year time series of 10 min rainfall intensities at Uccle, Belgium, Water Resour. Res., 44, 1–15, https://doi.org/10.1029/2007WR006471, 2008.

Ochoa-Rodriguez, S., Wang, L.P., Gires, A., Pina, R.D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., Van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., and Ten Veldhuis, M.C.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, J. Hydrol., 531, 389–407, https://doi.org/10.1016/j.jhydrol.2015.05.035, 2015.

Otero, N., Sillmann, J., and Butler, T.: Assessment of an extended version of the Jenkinson–Collison classification on CMIP5 models over Europe, Clim. Dyn., 50, 1559–1579, https://doi.org/10.1007/s00382-017-3705-y, 2018.

Philipp, A., Beck, C., Huth, R., and Jacobeit, J.: Development and comparison of circulation type classifications using the COST 733 dataset and software, Int. J. Climatol., 36, 2673–2691, https://doi.org/10.1002/joc.3920, 2016.

Photiadou, C., van den Hurk, B., van Delden, A., and Weerts, A.: Incorporating circulation statistics in bias correction of GCM ensembles: hydrological application for the Rhine basin, Clim. Dyn., 46, 187–203, https://doi.org/10.1007/s00382-015-2578-1, 2016.

Plavcová, E. and Kyselý, J.: Projected evolution of circulation types and their temperatures over Central Europe in climate
5   models, Theor. Appl. Climatol., 114, 625–634, https://doi.org/10.1007/s00704-013-0874-4, 2013.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Van Lipzig, N.P.M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev. Geophys., 53, 323–361, https://doi.org/10.1002/2014RG000475, 2015.

Pryor, S., and Schoof, J.: A hierarchical analysis of the impact of methodological decisions on statistical downscaling of daily
10 precipitation and air temperatures, Int J Climatol., https://doi.org/10.1002/joc.5990, 2019.

Rybka, H., and Tost, H.: Uncertainties in future climate predictions due to convection parameterisations, Atmos. Chem. Phys., 14, 5561–5576, https://doi.org/10.5194/acp-14-5561-2014, 2014.

Sachindra, D.A., Ahmed, K., Shahid, S., and Perera, B.J.C.: Cautionary note on the use of genetic programming in statistical downscaling, Int. J. Climatol., 1–17, https://doi.org/10.1002/joc.5508, 2018.

15 Salvadore, E., Bronders, J., and Batelaan, O.: Hydrological modelling of urbanized catchments: A review and future directions, J. Hydrol., 529, 62–81, https://doi.org/10.1016/j.jhydrol.2015.06.028, 2015.

Salvi, K., Ghosh, S., and Ganguly, A.R.: Credibility of statistical downscaling under nonstationary climate, Clim. Dyn., https://doi.org/10.1007/s00382-015-2688-9, 2016.

Sanderson, B.M., Wehner, M., and Knutti, R.: Skill and independence weighting for multi-model assessments, Geosci. Model
20 Dev., 10, 2379–2395, https://doi.org/10.5194/gmd-10-2379-2017, 2017.

Santos, J.A., Belo-pereira, M., Fraga, H., and Pinto, J.G.: Understanding climate change projections for precipitation over western Europe with a weather typing approach, J. Geophys. Res. Atmos., 121, 1170–1189, https://doi.org/10.1002/2015JD024399, 2016.

Schiemann, R. and Frei, C.: How to quantify the resolution of surface climate by circulation types: An example for Alpine
25 precipitation, Phys. Chem. Earth, 35, 403–410, https://doi.org/10.1016/j.pce.2009.09.005, 2010.

Schoof, J.: Statistical downscaling in Climatology, Geography Compass, 7(4), 249-265, https://doi.org/10.1111/gec3.12036, 2013.

Soares, P.M.M., Maraun, D., Brands, S., Jury, M.W., Gutiérrez, J.M., San-Martín, D., Hertig, E., Huth, R., Belušić Vozila, A., Cardoso, R. M., Kotlarski, S., Drobinski, P., and Obermann-Hellhund, A.: Process-based evaluation of the VALUE perfect
30 predictor experiment of statistical downscaling methods, Int J Climatol., https://doi.org/10.1002/joc.5911, 2018.

Souverijns, N., Thiery, W., Demuzere, M., and Van Lipzig, N.P.M.: Drivers of future changes in East African precipitation Drivers of future changes in East African precipitation, Environ. Res. Lett., 11, 1–9, https://doi.org/10.1088/1748-9326/11/11/114011, 2016.

Stryhal, J. and Huth, R.: Classifications of winter Euro-Atlantic circulation patterns: An intercomparison of five atmospheric reanalyses., J. Clim., 30, 7847–7861, https://doi.org/10.1175/JCLI-D-17-0059.1, 2017.

Stryhal, J. and Huth, R.: Classifications of winter atmospheric circulation patterns: validation of CMIP5 GCMs over Europe and the North Atlantic, Clim. Dyn., https://doi.org/10.1007/s00382-018-4344-7, 2018.

5 Sunyer, M. A., Hundecha, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., Vormoor, K., Bürger, G., Hanel, M., Kriaučiūnienė, J., Loukas, A., Osuch, M., and Yücel, I.: Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe, Hydrol. Earth Syst. Sci., 19, 1827-1847, https://doi.org/10.5194/hess-19-1827-2015, 2015.

Tabari, H., De Troch, R., Giot, O., Hamdi, R., Termonia, P., Saeed, S., Brisson, E., Van Lipzig, N., and Willems, P.: Local impact analysis of climate change on precipitation extremes: Are high-resolution climate models needed for realistic

10 simulations?, Hydrol. Earth Syst. Sci., 20, 3843–3857, https://doi.org/10.5194/hess-20-3843-2016, 2016.

Tabari, H., Taye, M., and Willems, P.: Water availability change in central Belgium for the later 21st century, Global and Planetary Change, 131, 115-123, https://doi.org/10.1016/j.gloplacha.2015.05.012, 2013.

Taylor, K.E., Stouffer, R.J., and Meehl, G.A.: An overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc., 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

15 Termonia, P., Van Schaeybroeck, B., De Cruz, L., De Troch, R., Caluwaerts, S., Giot, O., Hamdi, R., Vannitsem, S., Duchêne, F., Willems, P., Tabari, H., Van Uytven, E., Hosseinzadehtalaei, P., Van Lipzig, N., Wouters, H., Vanden Broucke, S., van Ypersele, J.P., Marbaix, P., Villanueva-Birriel, C., Fettweis, X., Wyard, C., Scholzen, C., Doutreloup, S., De Ridder, K., Gobin, A., Lauwaet, D., Stavrakou, T., Bauwens, M., Müller, J.F., Luyten, P., Ponsar, S., Van den Eynde, D., and Pottiaux, E.: The CORDEX.be initiative as a foundation for climate services in Belgium, Clim. Serv.,

20 https://doi.org/10.1016/j.cliser.2018.05.001, 2018.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, J. Hydrol., 456–457, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J.,

25 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. R. Meteorol. Soc., 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

30 Vaittinada Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M., and Gallardo, C.: Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations, Clim. Dyn., 46, 1301–1329, https://doi.org/10.1007/s00382-015-2647-5, 2016.

Vanden Broucke, S., Wouters, H., Demuzere, M., and Van Lipzig, N.: The influence of convection-permitting regional climate modeling on future projections of extreme precipitation: dependency on topography and timescale, Clim. Dyn., https://doi.org/10.1007/s00382-018-4454-2, 2018.

Van Uytven, E. and Willems, P.: Greenhouse gas scenario sensitivity and uncertainties in precipitation projections for central
5   Belgium, J. Hydrol., 558, 9–19, https://doi.org/10.1016/j.jhydrol.2018.01.018, 2018.

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.J., and Rose, S.K.: The representative concentration pathways: An overview, Clim. Change, 109, 5–31, https://doi.org/10.1007/s10584-011-0148-z, 2011.

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

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F., and Willems, P.: Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation, J. Hydrol., 511, 335–349, https://doi.org/10.1016/j.jhydrol.2014.01.050, 2014b.

15 Vermuyten, E., Van Uytven, E., Meert, P., Wolfs, V., and Willems, P: Long-term optimization of river flood control under historical and future climatic conditions, under review, 2019.

Wang, Y., Sivandran, G., and Bielicki, J.M.: The stationarity of two statistical downscaling methods for precipitation under different choices of cross-validation periods, Int. J. Climatol., 38, e330–e348, https://doi.org/10.1002/joc.5375, 2018.

Watterson, I.G., Bathols, J., and Heady, C.: What influences the skill of climate models over the continents?, Bull. Am.
20 Meteorol. Soc., 95, 689–700, https://doi.org/10.1175/BAMS-D-12-00136.1, 2014.

Werner, A. T. and Cannon, A. J.: Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods, Hydrol. Earth Syst. Sci., 20, 1483-1508, https://doi.org/10.5194/hess-20-1483-2016, 2016.

Wilby, R.L., and Wigley, T.M.L.: Precipitation predictors for downscaling: Oberserved and General Circulation Model relationships, Int. J. Clim., 20, 641–661, https://doi.org/10.1002/(Sici)1097-0088(200005)20:6<641::Aid-Joc501>3.0.Co;2-1,
25 2000.

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, https://doi.org/10.1016/j.jhydrol.2013.05.037, 2013.

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

30 Willems, P.: A time series tool to support the multi-criteria performance evaluation of 8 rainfall-runoff models, Environ Model Softw, 24, 311–321, https://doi.org/10.1016/j.envsoft.2008.09.005, 2009.

Willems, P.: Compound intensity / duration / frequency-relationships of extreme precipitation for two seasons and two storm types, J. Hydrol., 233, 189–205, https://doi.org/10.1016/S0022-1694(00)00233-X, 2000.

Yang, C., Wang, N., and Wang, S.: A comparison of three predictor selection methods for statistical downscaling, Int. J. Climatol., 37, 1238–1249, https://doi.org/10.1002/joc.4772, 2017.

Zhang, X., Zwiers, F.W., Li, G., Wan, H., and Cannon, A.J.: Complexity in estimating past and future extreme short-duration rainfall, Nat. Geosci., 10, 255–259, https://doi.org/10.1038/ngeo2911, 2017.
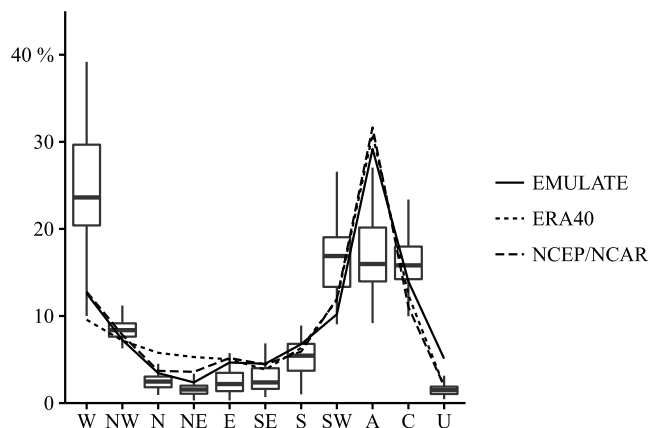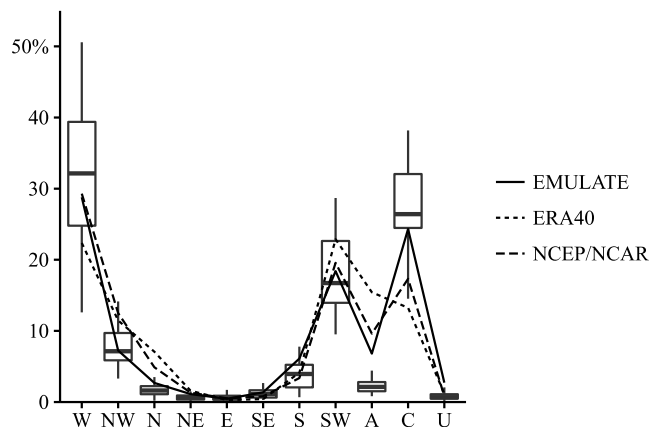
5

Figure 1 Relative WT occurrence in the winter season for different re-analysis datasets (lines) and climate model runs (boxplots). The results are obtained for the reference period 1961-1990.

5

Figure 2 Relative winter precipitation accumulation per WT for different re-analysis datasets (lines) and climate model runs (boxplots). The results are obtained for the reference period 1961-1990.
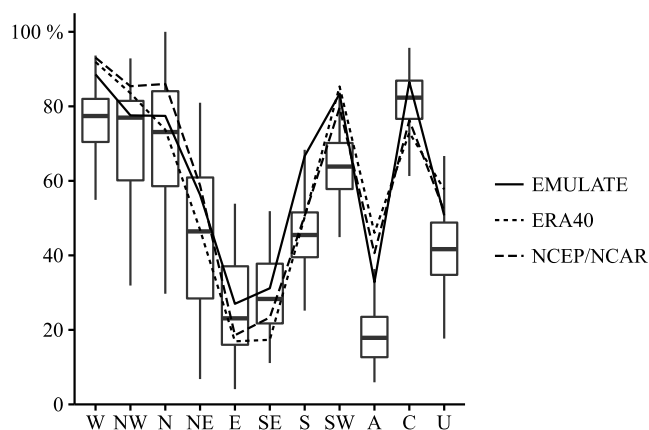
Figure 3 Percentage wet days per WT in the winter season for different re-analysis datasets (lines) and climate model runs (boxplots). The results are obtained for the reference period 1961-1990.
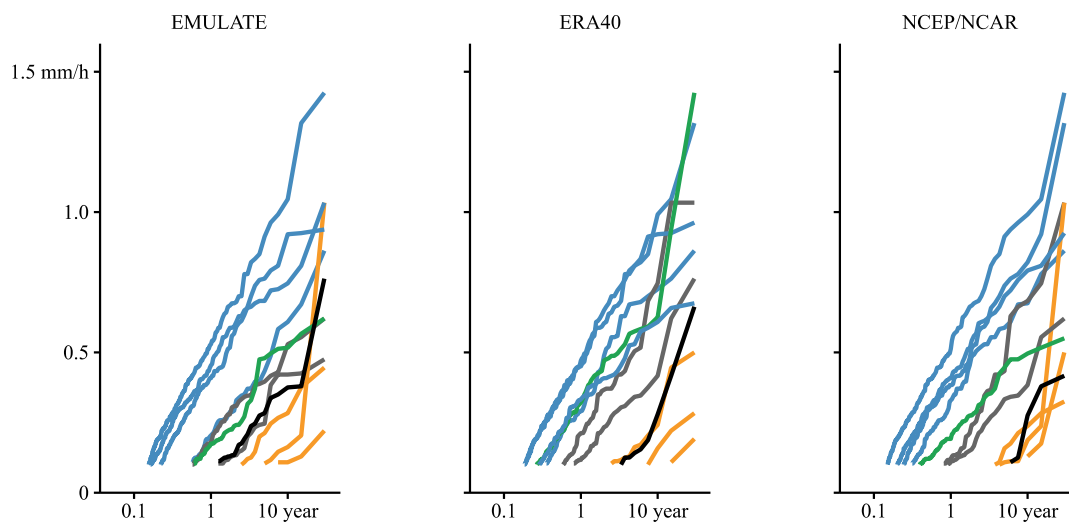
Figure 4 Daily winter precipitation amounts per weather type for different re-analysis datasets. The blue lines indicate the results for the W, NW, SW and C WTs, the green line for the A WT, the grey lines for the N and S WTs, the orange lines for the NE, E, SE WTs and the black line for the U WTs. The results are obtained for the reference period 1961-1990.
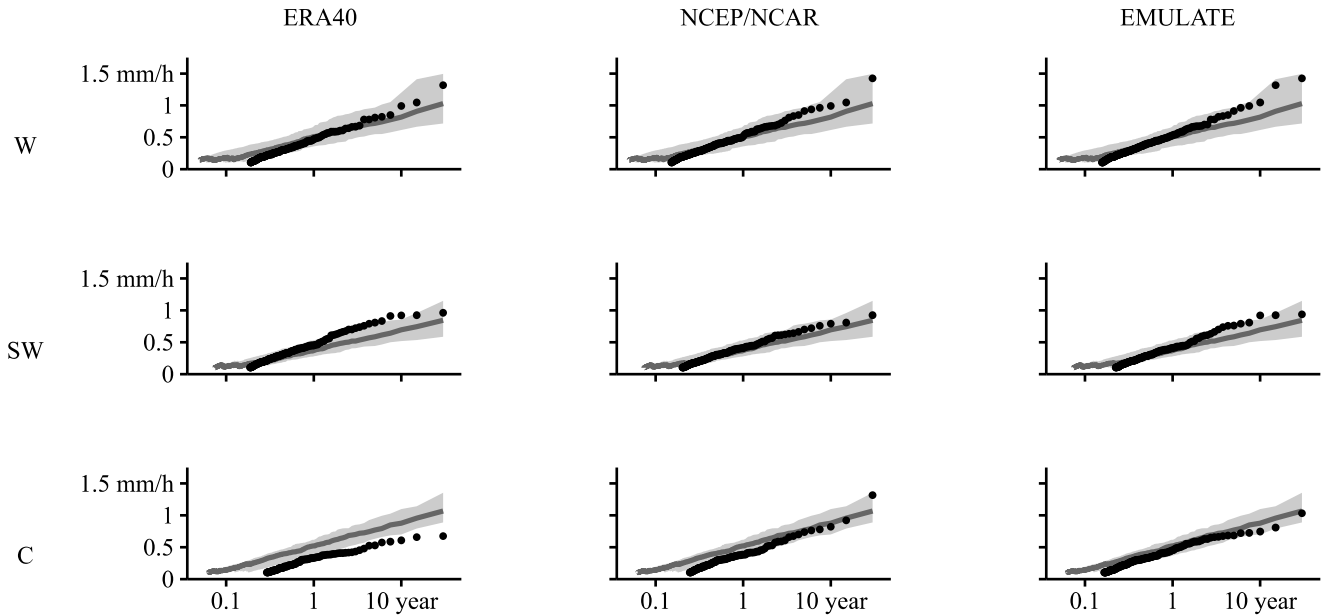
Figure 5 Daily winter precipitation amounts in function of their empirical return period for the wettest weather types (W, SW and C WTs) and different re-analysis datasets. The black dots represent the observed relationship. The relationship simulated by the climate model runs is indicated by the grey line (median result) and area (5th – 95th percentile range). The results are obtained for the period 1961-1990.
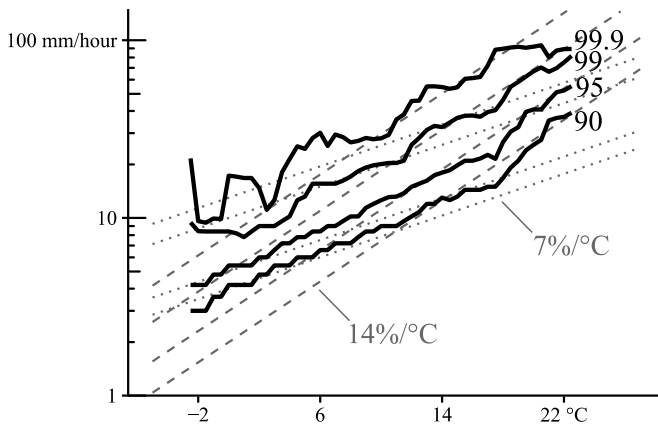

Figure 6 The relationship between daily average temperature and independent 10 minutes precipitation amounts. The relationship is defined on annual time scale and this by using the entire Uccle time series (1901-2000). The CC relation (+7%/°C) is indicated by the grey dotted lines, whereas the 2x CC relation (+14%/°C) by the grey striped lines. The black lines show magnification of the 90th, 95th, 99th and 99.9th percentile precipitation amount for increasing temperatures.
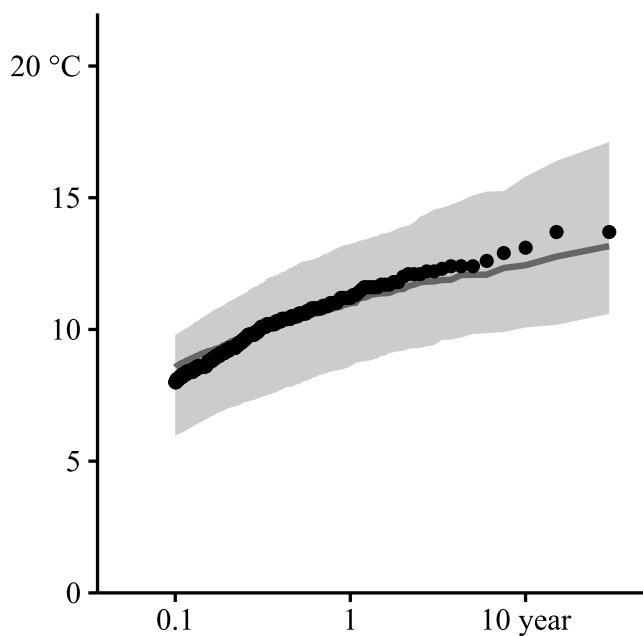
Figure 7 Daily average temperatures in function of their empirical return period. The black dots represent the observed relationship. The relationship simulated by the climate model runs is indicated by the grey line (median result) and area (5th – 95th percentile range). The results are obtained for the reference period 1961-1990.
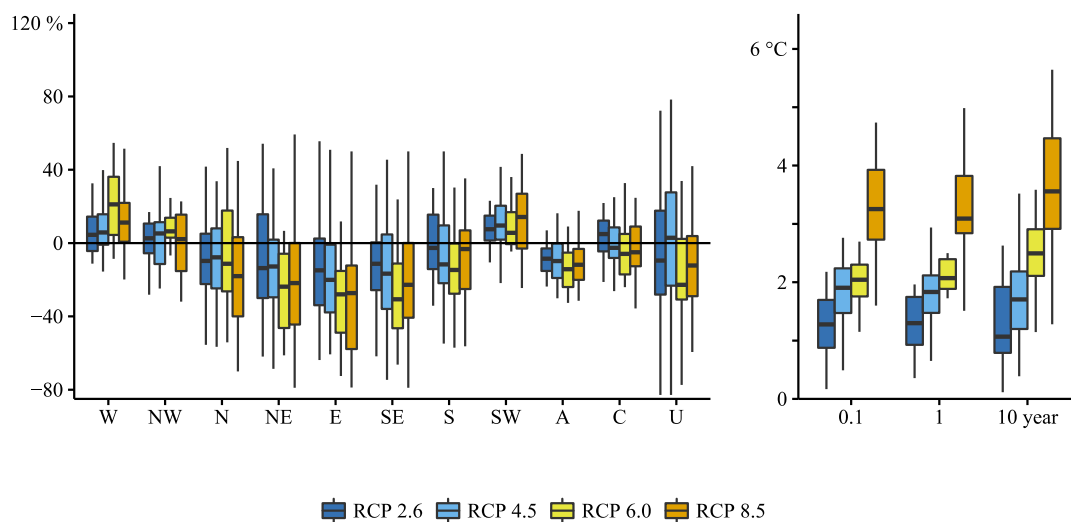
5

Figure 8 Changes in the winter WT occurrences (left) and daily average temperature (right) between 2071-2100 and 1961-1990 in function of the RCP.
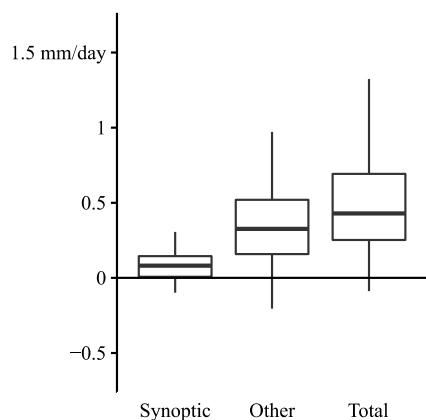
Figure 9 Change in the average daily precipitation amount of wet winter days due to synoptic effects, i.e. changes in the WT occurrence frequencies, and other effects, for instance due to thermo-dynamical changes. The results are based on the climate model output for the scenario (2071-2100) and control period (1961-1990).
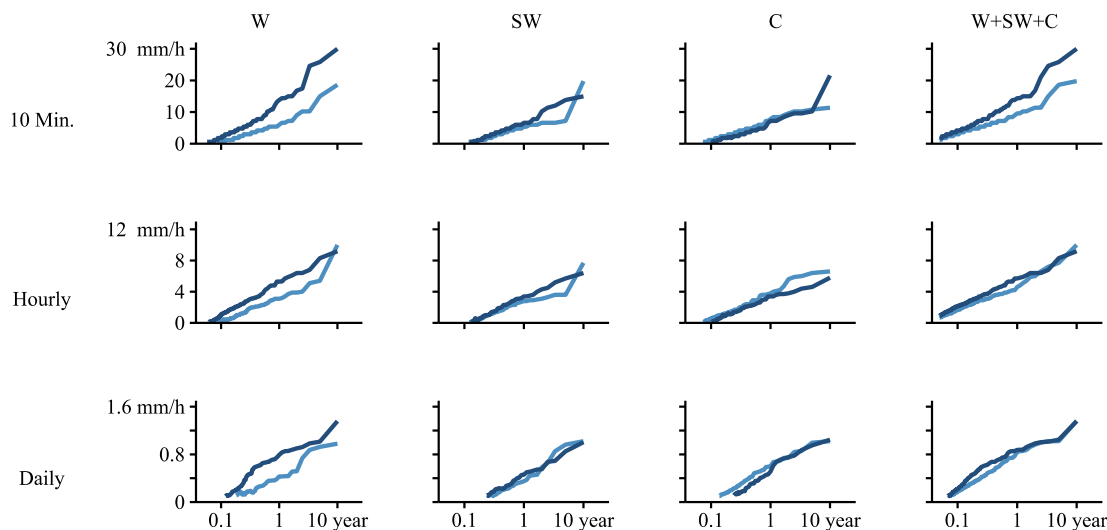
5

Figure 10 Independent 10 minutes, hourly and daily aggregated winter precipitation amounts associated with the wetter WTs. The precipitation amounts for moving windows starting between 1981 and 1991 are indicated by dark blue, whereas the light blue represent the amounts for moving windows starting between 1931 and 1941. The lines show the median result.
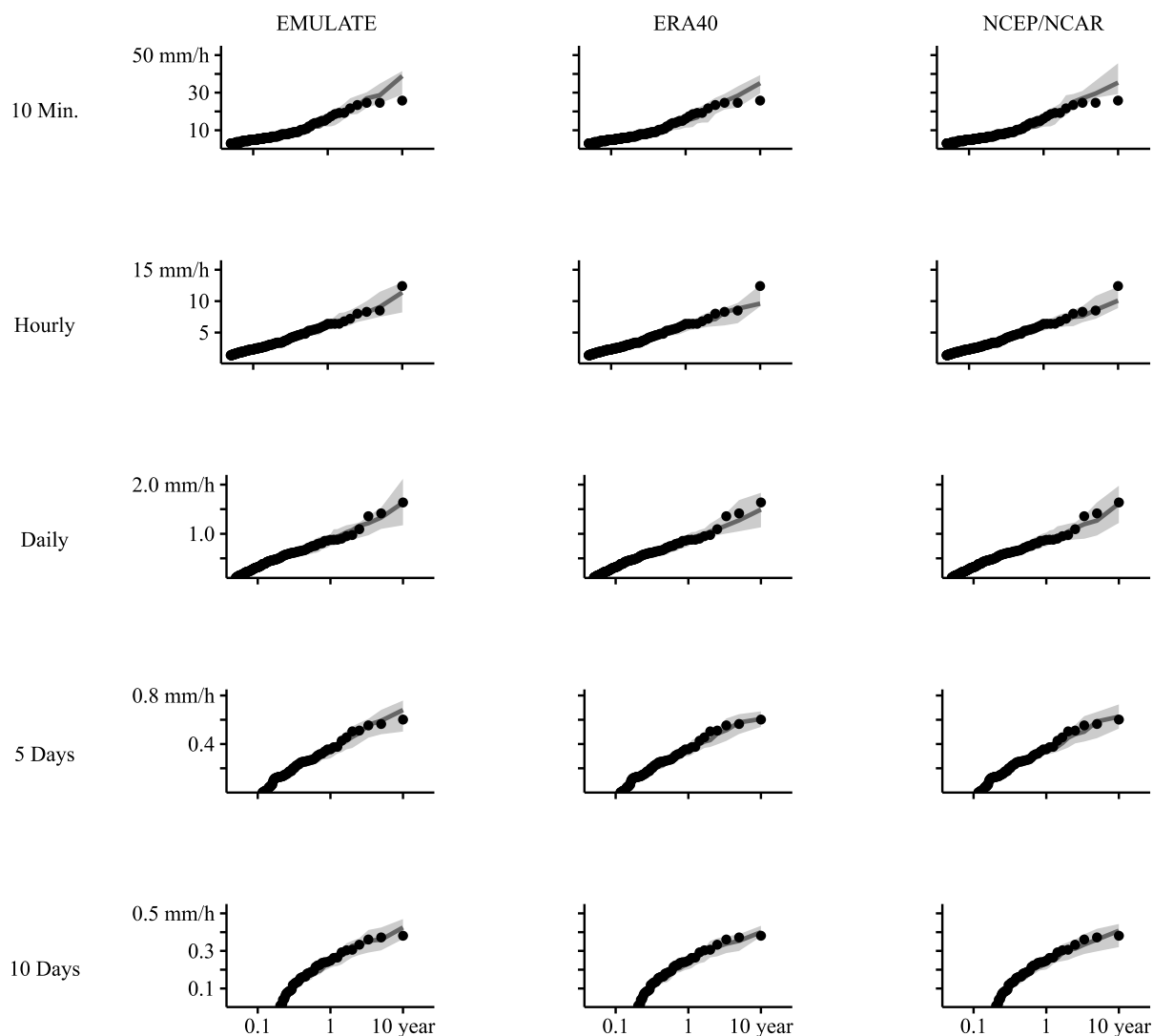
Figure 11 Independent winter precipitation amounts for the observed time series (1991-2000, black dots) and downscaled time series, which are obtained in the perfect predictor experiment (grey line: median result, grey area: 5th-95th percentile range). The inter-comparison is done for different aggregation levels (rows) and re-analysis datasets (columns).
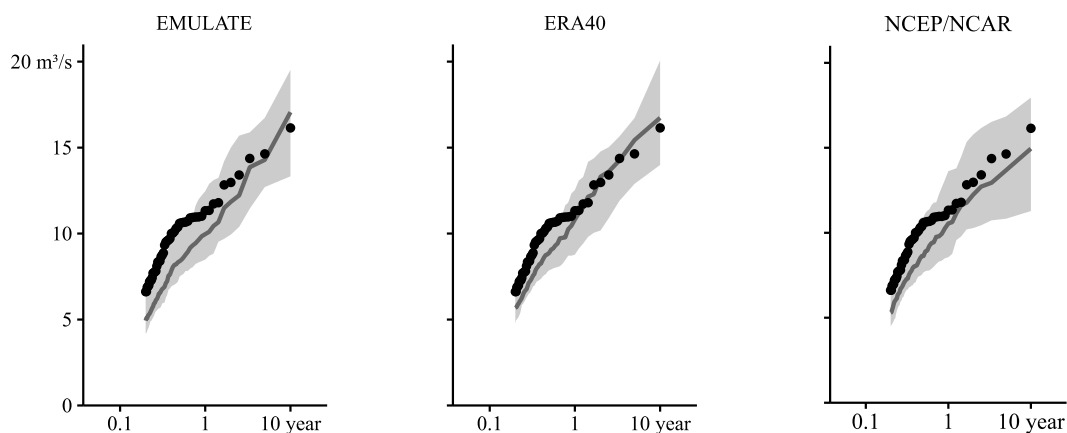
Figure 12 Simulated, independent daily winter peak flow discharges for the observed time series (1991-2000, black dots) and downscaled time series, which are obtained in the perfect predictor experiment (grey line: median result grey area: 5th-95th percentile range). The inter-comparison is done for different re-analysis datasets (columns).

5

Table 1 Overview of the re-analysis dataset ensemble employed in this study

Re-analysis dataset	Resolution (Lon x Lat)	Time range	Reference
EMULATE	5° x 5°	1881-2000	(Ansell et al., 2006)
ERA40	2° x 2°	1948-present	(Uppala et al., 2005)
NCEP/NCAR	2.5° x 2.5°	1957-2002	(Kalnay et al., 1996)

Table 2 Overview of the climate model ensemble employed in this study

| Modeling Center or Group | Model name | RCP | | | | Model resolution |
		2.6	4.5	6.0	8.5	(Lon x Lat)
Beijing Climate Center, China Meteorological Administration	BCC-CSM1.1(m)		1	1	1	1.1° x 1.1°
CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	ACCESS 1.0				1	1.9° x 1.3°
Canadian Centre for Climate Modelling and Analysis	CanESM2	5	5		5	2.8° x 2.8°
Centro Euro-Mediterraneo per I Cambiamenti Climatici	CMCC-CESM				1	3.8° x 3.7°
	CMCC-CMS		1		1	1.9° x 1.9°
Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	CNRM-CM5	1	1		1	1.4° x 1.4°
Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence	CSIRO-Mk3.6.0		1	1	1	1.9° x 1.9°
College of Global Change and Earth System Science, Beijing Normal University	BNU-ESM	1	1			2.8° x 2.8°
Institute for Numerical Mathematics	INM-CM4		1		1	2.0° x 1.5°
Institut Pierre-Simon Laplace	IPSL-CM5A-LR	1	3	1	3	3.8° x 1.9°
	IPSL-CM5A-MR	1	1	1	1	2.5° x 1.3°
	IPSL-CM5B.LR		1		1	3.8° x 1.9°
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC-ESM-CHEM	1	1	1	1	2.8° x 2.8°
	MIROC-ESM	1	1	1	1	2.8° x 2.8°
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC5	3	2	3	3	1.4° x 1.4°
Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)	HadGEM2-AO	1	1	1	1	1.9° x 1.3°
Max Planck Institute for Meteorology (MPI-M)	MPI-ESM-LR	1	1		1	1.9° x 1.9°
	MPI-ESM-MR	1	1		1	1.9° x 1.9°
Meteorological Research Institute	MRI-CGCM3	1	1	1	1	1.1° x 1.1°
NASA Goddard Institute for Space Studies	GISS-E2-R		2			2.5° x 2.0°
Norwegian Climate Centre	NorESM1-M	1		1	1	2.5° x 1.9°
Geophysical Fluid Dynamics Laboratory	GFDL-CM3			1	1	2.5° x 2.0°
	GFDL-ESM2G	1	1	1	1	2.5° x 2.0°
	GFDL-ESM2M		1	1	1	2.5° x 2.0°