Fighting big data and ensemble fatigue in climate change impact studies: can we turn the ensemble cascade upside down?

Journal:	<i>International Journal of Climatology</i>
Manuscript ID	JOC-19-0577.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	04-Jan-2020
Complete List of Authors:	Van Uytven, Els; KU Leuven, Department of Civil Engineering De Niel, Jan; KU Leuven, Department of Civil Engineering Meert, Pieter; KU Leuven, Department of Civil Engineering Wolfs, Vincent; KU Leuven, Department of Civil Engineering Willems, Patrick; KU Leuven, Department of Civil Engineering
Keywords:	Ensembles < 1. Tools and methods, Climate < 2. Scale, Statistical methods < 1. Tools and methods, Rainfall < 3. Physical phenomenon, Hydrology < 4. Geophysical sphere
Country Keywords:	Belgium

SCHOLARONE™
Manuscripts

ORIGINAL ARTICLE

Fighting big data and ensemble fatigue in climate change impact studies: can we turn the ensemble cascade upside down?

E. Van Uytven¹ | J. De Niel¹ | P. Meert¹ | V. Wolfs¹
| P. Willems¹

¹KU Leuven, Department of Civil Engineering – Hydraulics Section, Kasteelpark 40 box 2448, 3001 Leuven, Belgium

Correspondence

E. Van Uytven, KU Leuven, Department of Civil Engineering – Hydraulics Section, Kasteelpark 40 box 2448, 3001 Leuven, Belgium
Email: van.uytven.els@gmail.com

Funding information

E. Van Uytven is funded by a doctoral grant from the Research Foundation – Flanders (FW.O., grant number 11ZY418N).

Climate change impact modellers consider the availability of large ensembles of climate model results more and more as problematic. They experience big data or ensemble fatigue and face computational limits. This study proposes an ensemble design approach based on clustering of the climate model skill, climate change signals and statistical downscaling skill, and investigates its potential for ensemble size reduction. The proposed approach is demonstrated for river and urban hydrological impact studies in Belgium, considering the average winter (summer) precipitation amount and extreme daily winter (summer) precipitation amount with a 10-year return period. The analysis starts from an original 240 membered multi-ensemble (48 climate models and five statistical downscaling methods) and is reduced to eight (12) members for the average seasonal winter (summer) precipitation amount and 18 (22) for the extreme daily winter (summer) precipitation amount. The range of the impact results by the original multi-ensemble is generally preserved. However, in some cases, the reduced ensemble shows biased impact results. The cluster analysis confirms the dependence between statistical downscaling methods and points to the interaction between climate models and statistical downscaling methods.

KEYWORDS

Clustering, Statistical downscaling, Validation, Bias, Skill, Perfect predictor experiment, Inter-dependence

1 | INTRODUCTION

Various comparative studies have analysed climate models and concluded that a single best climate model does not exist (Dai, 2006; Stocker et al., 2013; Jacob et al., 2014; Rajczak and Schär, 2017; Vanden Broucke et al., 2018). For a specific location, variable and season, the skill of different climate models and the climate change signals deviate among the models due to differences in spatial resolution (Anstey et al., 2013; Davini et al., 2017; Hartung et al., 2017; Mendoza et al., 2016; Watterson et al., 2014), initial climatic conditions representing the internal variability of the climate system (Deser et al., 2012; Fadhel et al., 2017; Hosseinzadehtalaei et al., 2017), parametrization schemes (Arakawa, 2004; Prein et al., 2015; Rybka and Tost, 2014) and model components (Alexander and Easterbrook, 2015; Lorenz et al., 2012; Pitman, 2003). Besides climate model differences, similarities arise from shared coding, the process of continuous model improvements and the fact that they all represent the same climate system (Knutti et al., 2013; Leduc et al., 2016; Masson and Knutti, 2011; Pennell and Reichler, 2011; Sunyer et al., 2013).

In addition to the large ensemble of climate models, a variety of statistical downscaling methods (SDMs) has been developed (Fowler et al., 2007; Gutmann et al., 2014; Li et al., 2017; Maraun et al., 2010; Maraun and Widmann, 2018; Smid and Costa, 2017; Sunyer et al., 2015; Wilby and Dawson, 2013). These can be classified based on their differences in the calibration data (perfect prognosis versus model output statistical methods), the statistical function (deterministic versus stochastic, and linear versus non-linear), and the nature of the predictor(s) and predictand (distributional versus discrete parameters). Within the same class, basic and advanced methods exist, with potentially large differences in their skill depending on the variable and impact application. Recently, within the COST Action VALUE (Action ES 1102, www.value-cost.eu), a validation framework for statistical downscaling methods has been set up (Maraun et al., 2014) and applied to assess the skill of 20 (22) precipitation (temperature) downscaling methods

across Europe (Gutiérrez et al., 2018; Hertig et al., 2018; Maraun et al., 2017, 2018; Soares et al., 2018; Widmann et al., 2019). To this end, historical observations and the output of re-analysis datasets are used to define pseudo climate model runs. Next, SDMs are applied to the pseudo climate model output and their skill was assessed based on the reproduction of time series statistics. Such so-called perfect predictor experiment is not original and has been applied in the past (Bürger et al., 2012; Werner and Cannon, 2016). As was the case for the precipitation projections, differences in downscaling skill are ascribed to the underlying methodology of the SDMs.

Differences between greenhouse gas scenarios, climate models, internal variability of the climate system and SDMs introduce uncertainties in climate change impact analyses. They are accounted for by applying a multi-ensemble approach, nowadays a standard practice in climate change impact studies (Karlsson et al., 2016; Marx et al., 2017; Potter et al., 2018; Seiller et al., 2017). Because more and more climate models and SDMs become available, the size of the multi-ensemble tends to increase when using this approach. It leads to the creation of big data and induces ensemble fatigue (Benestad et al., 2017). In particular, a combination of x greenhouse gas scenarios (GHSs), y global climate models (GCMs) and z SDMs, leads to a w -membered multi-ensemble, where w equals $x \times y \times z$. This means that w different impact simulations have to be conducted when applying the traditional multi-ensemble approach. This number would even be higher when considering stochastic statistical downscaling methods, for instance weather generators (Burton et al., 2008; Thorndahl et al., 2017) or stochastic re-sampling techniques (Willems and Vrac, 2011). Depending on the nature of the impact model, conceptual versus physically based models or lumped versus distributed models, this may pose a considerable computational burden to the impact modeller.

Big data in climate change impact studies has among others been addressed through the development of semi-automatic downscaling and processing tools (Benestad et al., 2015; Burton et al., 2008; Gudmundsson et al., 2012; Hanel et al., 2017; Hessami et al., 2008; Iturbide et al., 2019; Mehrotra et al., 2018; Wilby and Dawson, 2013), tailored climate change impact scenarios for urban and river catchments (Ntegeka et al., 2014; Vansteenkiste et al., 2014) and ensemble design approaches. Such approaches facilitate the big data processing (cf. end-of-pipe solution), but cannot prevent big data creation (cf. source control solution). The latter may be addressed by condensing the total ensemble

to a smaller sub-ensemble. Ensemble design approaches, capable of identifying primary information and removing redundant information, may prevent the creation of big data.

To the present day, the IPCC has not recommended an approach for the design of ensembles in climate change impact studies (Stocker et al., 2013). Hence, different strategies exist for designing ensembles and confining the ensemble size. Impact modellers might arbitrarily sub-select climate models and in this way an *ensemble of opportunity* is designed. The arbitrariness of such an ensemble lies in the fact that researchers often select their properly developed models, the models they are familiar with, models which are commonly applied and/or models with specific strengths (Addor and Melsen, 2019; Rössler et al., 2017). On the contrary, ensemble design approaches are more transparent and frequently consider the reproduction of historical statistics. For instance, the total ensemble size could be condensed to a sub-ensemble by weighting ensemble members with zero and one. Weights equal to one are assigned to the better performing models, whereas models not capturing the dynamical and the thermo-dynamical precipitation drivers, and thus not accurately reproducing historical statistics, are excluded by a weight equal to zero. Some weighting schemes are more advanced and simultaneously address the dependence between ensemble members (Sanderson et al., 2017). Dalelane et al. (2018) tested an advanced weighting technique on a 15 membered EURO-CORDEX ensemble. They were able to reduce the original climate model ensemble size by half (15 to seven climate model runs), while maintaining 94% of the original information content. Besides member weighting, clustering techniques are frequently applied. For instance, Mendlik and Gobiet (2016) and Wilcke and Bärring (2016) applied clustering techniques to multi-variate climate change signals. In that way, Mendlik and Gobiet (2016) succeeded in condensing the original climate model ensemble size from 25 to five climate model runs, while preserving the main statistics of the projected changes. Using concepts of information theory and entropy, Pechlivanidis et al. (2018) also succeeded in defining a representative sub-ensemble. Their results indicate that approximately 33% of the original ensemble size contains over 80% of the total information.

In this study, a multi-ensemble design approach is presented. The approach considers the individual skill of and dependence between ensemble members and accounts for both through clustering. Rather than applying the full ensemble for impact modelling, the ensemble size is reduced by means of the proposed technique as schematically shown in Figure 1. Note that the application of clustering techniques and/or considering the skill of and dependence between ensemble members is not new (Chen et al., 2017; Crawford et al., 2019; Dalelane et al., 2018; Knutti et al., 2017; Mendlik and Gobiet, 2016; Sanderson et al., 2017). However, they have not yet been applied in the design of multi-ensembles, including the statistical downscaling step.

FIGURE 1 a) Traditionally, the total ensemble includes different greenhouse gas scenarios (GHSs), several global climate models (GCMs) with different initial climatic conditions (representing the internal variability of the climate system) and different statistical downscaling methods (SDMs). Doing so, the total ensemble size builds up and this build-up is referred to as the *ensemble cascade*. b) By combining validation and inter-dependence analyses, as proposed in this study, the *ensemble cascade* is turned upside down and the total ensemble size is reduced.

The structure of this study is as follows. The design approach is presented in Section 2 and demonstrated for case studies in Belgium (Section 3). More precisely, the approach is demonstrated for the average winter (summer) precipitation accumulation and extreme daily winter (summer) precipitation amounts with a 10-year return period. The condensed ensembles are validated for two hydrological climate change impact studies, one on extreme river peak flow discharges and one on overflow volumes in a stormwater retention basin. The results are presented in Section 4, followed by the discussion and conclusions in Section 5.

2 | METHODOLOGY

2.1 | The multi-ensemble design approach

The proposed multi-ensemble design approach turns the ensemble cascade upside down in two steps (Figure 1). In the first step, clusters are defined between the climate model skill and climate change signal for each SDM. Through clustering, the dependence between the climate model greenhouse gas scenario sensitivity (i.e. the response of the climate model to a specific greenhouse gas scenario) and climate model skill are simultaneously accounted for. Next, a representative climate model run is identified for each cluster. Doing so, the application of clustering allows the identification and removal of redundant information, but does not reduce the climate model uncertainties. The latter requires the improvement of the climate model skill.

In the second step, information on the representative cluster members is collected for all SDMs and clusters are determined between the climate change signals for varying SDMs and the corresponding statistical downscaling skill. By doing so, the dependence between SDMs is addressed and the interaction between the climate models and SDMs is accounted for. The latter does not reduce the uncertainties introduced by the statistical downscaling step.

The clustering itself is performed using k -medoids (partitioning around medoids *pam*) and the number of clusters is identified through the average silhouette method. For both methods, the R packages *cluster* (Maechler et al., 2016) and *factoextra* (Kassambara and Mundt, 2016) are used. More specifically, the function *fviz_nbclust()* is first applied to identify the optimal number of clusters. Inputs for this function are the clustering method and the method for estimating the optimal number of clusters, respectively *pam* and the average silhouette method in this case study. Next, the optimal number of clusters is used in the function *pam()* in order to define the k -medoid clusters. The k -medoid clusters are finally illustrated using the function *fviz_cluster()*. Prior to each clustering step, the data are re-scaled by subtracting the mean and subsequently dividing by the standard deviation. The representative cluster member is the cluster member closest to the medoid.

110

111 An optimal number of clusters is advised by the applied R packages, but in some cases this optimal number is
112 actually sub-optimal. More specifically, the suggested number of clusters does not account for the physical reason-
113 ing behind the climate change signals and the evaluation of the climate model and statistical downscaling skill. For
114 instance, it would not be logical to combine climate models projecting a strong intensification with climate models
115 projecting drying. Similarly, one would not cluster climate models overestimating the observations-based statistics
116 with models underestimating these statistics. Such cases typically occur when the suggested number of clusters is
117 small and disappear by increasing the number of clusters.

118 **2.2 | Climate model skill**

119 The climate model skill is evaluated through a comparison of statistical properties obtained from the available obser-
120 vations and the climate model output over a historical period. The climate model bias ($BIAS_{CM}$) is defined as follows:

$$BIAS_{CM} = \left(\frac{STAT_{CM}}{STAT_{OBS}} - 1 \right) \times 100\% \tag{1}$$

121 with $STAT_{OBS}$ the observed statistics and $STAT_{CM}$ the climate model produced statistics.

122 **2.3 | Statistical downscaling skill**

123 The skill of the SDMs is assessed through the perfect predictor (PP) experiment. In the first step of this experiment,
124 daily pseudo climate model data are defined using the observations. To fulfil the PP assumption of this experiment, the
125 observed time series equals the pseudo climate model time series for the control period. Next, the statistical down-
126 scaling methods are applied to the pseudo climate model data and downscaled time series are produced. Provided that

the statistical downscaling step would not involve assumptions, the statistics for the downscaled time series would match the statistics for the observations of the pseudo scenario period. In reality, however, the statistical downscaling step involves assumptions, adding uncertainties and inducing differences between the statistics for the downscaled results and the observed statistics.

The bias in the PP experiment ($BIAS_{PP}$), is defined as follows:

$$BIAS_{PP} = \left(\frac{STAT_{DOWNSC,PP}}{STAT_{OBS,PP}} - 1 \right) \times 100\% \quad (2)$$

with $STAT_{DOWNSC,PP}$ the statistics of the downscaled time series and $STAT_{OBS,PP}$ the observed statistics for the scenario period.

2.4 | Climate change signal

The climate change signal CCS is defined as follows:

$$CCS = \left(\frac{STAT_{DOWNSC}}{STAT_{OBS}} - 1 \right) \times 100\% \quad (3)$$

with $STAT_{DOWNSC}$ the statistics of the downscaled time series and $STAT_{OBS}$ the observed statistics.

3 | METEOROLOGICAL DATA, SDMS AND CASE STUDIES

3.1 | Observed and projected meteorological data

Uccle, situated near Brussels in Central Belgium (4.36° longitude, 50.8° latitude), is selected as study case. It is the location of the main station of the Royal Meteorological Institute of Belgium, where precipitation amounts have been measured every ten minutes since 1898. Daily potential evapotranspiration data are available for the same location since 1901. For both meteorological variables, the data series for 1901-2000 are considered for this study.

The output of 48 CMIP5 global climate model runs is employed in this study. The ensemble composition is specified in Table 1 and includes climate model runs for representative concentration pathway 8.5 (RCP8.5). The control and scenario periods are respectively taken as 1961-1990 and 2071-2100. The potential evapotranspiration data for the climate model outputs were computed from the data on average, maximum and minimum temperature, solar radiation, relative humidity and wind speed by the Bultot method, which is a Penman-Monteith method with parameters calibrated for Belgium (Bultot et al., 1993; Tabari et al., 2015).

The PP experiment is limited to one set of pseudo climate model runs. The control and scenario periods of the pseudo climate model data are defined as 1901-1930 and 1971-2000. The statistics of the target period are the statistics for 1971-2000.

TABLE 1 Composition of the climate model ensemble employed in this study

Id	Model	Research institution	Runs
1	ACCESS1-0	Commonwealth Scientific and Industrial Research Organisation, Bureau of Meteorology (BOM, Australia)	1
1	ACCESS1-3	Commonwealth Scientific and Industrial Research Organisation, Bureau of Meteorology (BOM, Australia)	1
2	BCC-CSM1.1(m)	Beijing Climate Center, Beijing Normal University	1
3	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	1
4	CanESM2	Canadian Centre for Climate Modelling and Analysis	5
5	CCSM4	National Center for Atmospheric Research	2
6	CESM1-CAM5	National Science Foundation, Department of Energy, National Center for Atmospheric Research	1
7	CMCC-CESM	Centro Euro-Mediterraneo per i Cambiamenti Climatici	1
7	CMCC-CM	Centro Euro-Mediterraneo per i Cambiamenti Climatici	1
7	CMCC-CMS	Centro Euro-Mediterraneo per i Cambiamenti Climatici	1
8	CNRM-CM5	Centre National de Recherches Meteorologiques, Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	1
9	CSIRO-Mk3.6	Commonwealth Scientific and Industrial Research Organisation, Queensland Climate Change Centre of Excellence	10
10	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Tsinghua University	1
11	GFDL-CM3	Geophysical Fluid Dynamics Laboratory	1
11	GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory	1
11	GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory	1
12	HadGEM2-AO	National Institute of Meteorological Research, Korea Meteorological Administration	1
13	HadGEM2-ES	Met Office Hadley Centre	1
14	INM-CM4	Institute for Numerical Mathematics	1
15	IPSL-CM5A-LR	Institut Pierre-Simon Laplace	4
15	IPSL-CM5A-MR	Institut Pierre-Simon Laplace	1
15	IPSL-CM5B-LR	Institut Pierre-Simon Laplace	1
16	MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies	1
16	MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies	1
17	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology	3
18	MP-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	1
18	MP-ESM-MR	Max Planck Institute for Meteorology (MPI-M)	1
19	MRI-CGCM3	Meteorological Research Institute	1
20	NorESM1-M	Norwegian Climate Centre	1
Climate model ensemble size:			48

3.2 | SDM ensemble

The SDM ensemble includes the following members:

- **SDM 1:** the simple delta change method based on change factors for the mean of the meteorological variable considered (Sunyer et al., 2015). For daily precipitation time series, this means that relative change factors are obtained for the average seasonal precipitation accumulation and applied to all values of the daily precipitation time series of that season. This method does not consider changes in the duration of wet or dry spells. Moreover, it assumes that the average and extreme precipitation amounts change by the same amount.
- **SDM 2:** the quantile perturbation method developed by Ntegeka et al. (2014). This change factor method first modifies the number of wet and dry days in the observed time series. Next, relative changes are applied to the daily precipitation amounts, considering the quantile of the observed daily precipitation amount. More precisely, the observed daily precipitation amount with quantile q is modified by the relative change in the daily precipitation amount with quantile q based on the climate model output. In other words, this method assumes that the average and extreme precipitation amounts change by different amounts and that the highest change is applicable to the day with the highest precipitation amount. Note that modifying the number of wet and dry days and applying the precipitation amount changes at a daily time scale may disturb the timeseries' autocorrelation. Also remark that the changes for the extreme precipitation amounts are based on a few values.
- **SDM 3:** the quantile perturbation method referred to as SD-A-5 by Willems and Vrac (2011). This method only modifies the precipitation amounts. More specifically, the applied changes are based on the quantile of the observed precipitation amount, which is defined by the highest sub-daily precipitation amount rather than the total daily precipitation amount. In other words, the highest change is applied to the day with the highest sub-daily precipitation amount. Hence, focus is put on the intensification of the extreme short-duration precipitation amounts. Similar to SDM 2, this method assumes that the average and extreme precipitation amounts change by different amounts, that the application of the changes at daily time scale may disturb the timeseries' autocorrelation and that the changes for the extreme precipitation amounts are based on a few values.

• **SDM 4:** the event-based perturbation method referred to as *SD-B* by Sørup et al. (2017). This perturbation method first identifies precipitation events and next applies event dependent, relative precipitation amount changes. The changes are defined after fitting and comparing exponential probability distributions for the daily precipitation amounts in the control and scenario periods. Hence, in comparison with SDMs 2 and 3, the changes for the extreme precipitation amounts are more robust and the autocorrelation is not disturbed. Note that the majority of the precipitation events is identified as non-extreme. For these non-extreme precipitation events, the performed downscaling methodology is similar to SDM 1.

• **SDM 5:** a non-linear transformation method developed by van Pelt et al. (2012). Projected precipitation time series are produced through a power relation. The coefficients of the power relation are derived from the climate model output for 5-days precipitation amounts. Consequently, this method has shortcomings with respect to the extreme short duration precipitation amounts. Furthermore, the power-relation may produce large precipitation amounts and may disturb the autocorrelation. Note that the method does not alter the duration of dry and wet spells.

For a detailed description of all methods the reader is referred to the indicated references.

3.3 | Case studies

The potential of the presented approach for ensemble size reduction is tested for the average winter (summer) precipitation accumulation and the extreme daily winter (summer) precipitation amount for a 10-years return period. The winter season includes January, February and December. The summer season includes June, July and August.

The clusters for the average winter precipitation accumulation are validated for climate change impact analysis on extreme peak flow discharges in the Grote Nete catchment. The rainfall runoff processes for that catchment are simulated by a lumped conceptual NAM model, which has been calibrated in a previous study (Vansteenkiste et al., 2014). The daily precipitation and potential evapotranspiration time series are applied as input series of the

hydrological model, simulating a daily discharge time series. Next, a peak-over-threshold method is applied to the simulated discharge time series to identify independent peak flow discharges (Willems, 2009). Eventually, a comparison is made between the projections of the largest peak flow discharges (empirical return periods varying between 10 and 100 years) for the original and condensed sub-ensemble. The downscaled precipitation input series are obtained by each of the five SDMs presented in Section 3.2. The downscaled potential evapotranspiration time series are produced by the change factor of mean method. This means that the median climate change signal for the RCP8.5 sub-ensemble is applied to the observed potential evapotranspiration time series. In summary, the clustering approach is applied to precipitation and not to potential evapotranspiration.

The clusters for the extreme daily summer precipitation amount with a 10-years return period are validated by a climate change impact analysis on the overflow volumes for a stormwater retention basin (Willems, 2013). The overflow events are simulated by means of a conceptual reservoir model (Wolfs and Willems, 2017), with a capacity of 250 m³ per hectare of effective connected impervious area. The reservoir empties linearly and has a maximum outflow of 20 l/s. The 10-minutes precipitation time series are applied to the model, simulating a continuous time series of 10-minutes overflow discharges. Next, independent overflow events are identified using a peak-over-threshold method. The overflow volumes are obtained by aggregating the overflow discharges over time.

4 | RESULTS

4.1 | Step 1: Clustering the climate change signals and climate model skill

In the first step, clusters are defined based on the climate model skill and climate change signal per SDM and precipitation variable. Figure 2 presents the clusters for the average winter precipitation accumulation. For sake of clarity, the results of only one SDM are shown, namely SDM 2. Note that the model skill and climate change signals have been individually standardised (Section 2.1), which means that a negative climate change signal does not directly imply a decrease in the average winter precipitation accumulation.

224

225 As a first observation, climate model runs with different initial conditions are clustered together. This is in partic-
226 ular the case for the different realisations of CanESM2, CSIRO, IPSL-CM5A-LR and MIROC5. Clustering of different
227 model realisations is a general phenomenon and appears also in the results for other SDMs and precipitation variables.
228 Besides clustering of different realisations of the same model, clustering by the institution that developed the model
229 is observed. For instance, in the results for the average winter precipitation accumulation (Figure 2), MIROC-ESM and
230 MIROC-ESM-CHEM are often clustered together, as well the GFDL and IPSL models. Apart from inter-dependence
231 between climate models, originating from the same research institution, inter-dependence is also observed between
232 climate models originating from different research institutions. This indicates that inter-dependence can not solely be
233 explained by the fact that climate models would have been developed by the same research institution (Leduc et al.,
234 2016). For the extreme daily summer precipitation amount, none of the MIROC-ESM, GFDL and IPSL models are
235 clustered together, only the ACCESS models do.

236

FIGURE 2 Clusters between the standardised climate model skill and climate change signal for the average winter precipitation accumulation. The change signals are produced by SDM 2. Each cluster is depicted with a different colour and symbol. Cluster members are connected with or bounded by lines. Two clusters are situated outside the presented ranges and are therefore not shown.

For each precipitation variable, the number of identified clusters in function of the SDM is indicated in Table 2. The number of clusters for extreme precipitation amounts is higher than the number of clusters for the average seasonal precipitation accumulation. More specifically, the results for the average summer precipitation accumulation show only 11 to 14 clusters, while for the extreme daily summer precipitation amount there are between 13 and 19 clusters. This may be related to the type of rainfall storms. In summer, precipitation events in Belgium may be both stratiform and convective. While the stratiform precipitation events are more or less adequately simulated by the climate models, the convective events are not due to their local and short-term nature (Tabari et al., 2016, 2015). More specifically, the longitudinal resolutions of the global climate, varying between 1° and 4° , do not capture the local convective precipitation events. Furthermore, the mesoscale processes producing extreme precipitation amounts are not resolved at the global climate model resolutions and are for this reason parameterised (Prein et al., 2015). The parametrisation schemes introduce additional random uncertainties in the climate model outputs for summer. Hence, the skill and projections for extreme precipitation amounts vary a lot (Figure 3). This might explain why a larger number of clusters is required to preserve the spread on the skill and projected changes. The application of parametric changes as done in SDM 4 is not able to remove all random uncertainties and also for this SDM a larger number of clusters is required.

TABLE 2 Number of clusters and thus size of the condensed ensemble after clustering the standardised climate model skill and climate change signal. The original GCM ensemble includes 48 climate model runs.

SDM	1	2	3	4	5
Average winter precipitation accumulation	12	11	10	10	11
Average summer precipitation accumulation	12	12	14	12	11
Extreme daily winter precipitation amount	20	21	15	15	18
Extreme daily summer precipitation amount	16	17	19	19	13

FIGURE 3 Clusters based on the standardised climate model skill and climate change signal for the extreme daily summer precipitation amount with a 10-year recurrence period. The change signals are produced by SDM 2. Each cluster is depicted with a different symbol and members of one cluster are connected with or bounded by lines.

For winter season, more clusters have been identified for the extreme precipitation amount than for the seasonal precipitation accumulation. This is likely due to the higher importance of the internal variability of the climate system in simulating precipitation extremes, as is confirmed for the CSIRO realisations (Figure 4). More specifically, the total range of the climate change signals divided by the median climate change signal measures 14% for the average winter precipitation accumulation, whereas it equals 27% for the extreme daily winter precipitation amount. For summer, the range equals 30% for the average seasonal precipitation accumulation and 50% for the extreme daily precipitation amount.

FIGURE 4 Ratio of the difference between the individual climate change signals and median of all climate change signals to the median of all climate change signals. The change signals for the considered CSIRO realisations are produced by SDM 2.

The differences in the number of clusters for different SDMs remain limited. One would expect a smaller number of clusters for the more simple SDMs and a higher number for the more advanced ones. However, no systematic dependence is found between the number of clusters and the complexity of the method.

4.2 | Step 2: Clustering the prior selected climate change signals and statistical downscaling skill

The second clustering step builds on the results of the prior clustering step and additionally considers the statistical downscaling skill. The size of the final sub-ensembles is presented in Table 3. Generally, the size of the condensed ensembles for the average seasonal precipitation accumulation is smaller than for the extreme precipitation amounts. As indicated in previous section, this can be firstly explained by the scattered pattern between the climate change signals and the skill for extreme precipitation amounts. Additionally, the dependence between SDMs is higher for average seasonal precipitation accumulation than for extreme daily amounts. This is demonstrated in the PP experiment, where the spread on the reproduced statistics versus the corresponding median is smaller for average seasonal precipitation accumulation in comparison to extreme daily amounts (Figure 5). Also, the SDM ensemble includes methods with varying complexity, which is especially of importance for extreme precipitation amounts. More specifically, SDM 1 is a simple downscaling method and assumes that all precipitation amounts change by the same factor. The other methods apply more advanced changes, such as quantile or event-based changes. The complexity of the SDM is especially important for the summer season and less for winter. For summer, drying is projected for the seasonal precipitation accumulation, while extreme precipitation amounts are projected to increase (Van Uytven and Willems, 2018). For winter, on the other hand, an increase is projected for all precipitation amounts. Therefore, the dependence between SDMs is stronger for winter than for summer.

TABLE 3 Size of the final sub-ensemble. The original ensemble includes 240 members (five SDMs × 48 GCMs)

Average winter precipitation accumulation	8
Average summer precipitation accumulation	12
Extreme daily winter precipitation amount	18
Extreme daily summer precipitation amount	22

FIGURE 5 Ratio of the difference between the individual SDM skill and median skill for all SDMs to the median skill for all SDMs.

Due to the scaling against the mean and the potentially unequally sized SDM sub-ensembles (Table 2), misinterpretations regarding the SDM skill might arise. Based on Figure 6a, which presents the clusters for the average winter precipitation accumulation, it appears that SDM 3 is the best performing SDM in the PP experiment. This is, however, not the case. Upon inspecting Figure 7, SDM 2 in fact has the best performance in reproducing the statistics of the PP experiment. This is due to the changes in the number of wet (dry) days and the quantile, daily and empirical nature of the applied precipitation amount changes.

FIGURE 6 a) Clusters based on the standardised SDM skill and the prior selected climate change signals for the average winter precipitation accumulation. Each cluster is depicted with a different symbol and colour. Members of one cluster are connected with or bounded by lines. One cluster is situated outside the presented ranges of the standardised skill and change and is therefore not shown in this figure. b) Close-up of the two clusters combining results for SDM 1, SDM 2 and SDM 4. Results for the same climate models are encircled in black.

As shown in Figures 5 and 6a, SDM 5 does not cluster with other SDMs for any of the considered cases. Depending on the variable, clusters include SDM 1, SDM 2, SDM 3 and/or SDM 4. Explaining why specific methods cluster together is not always possible. One possible reason could be that both SDM 2 and SDM 3 are quantile based perturbation methods and have a comparable skill (Figure 5). Also SDM 1 and SDM 4 have a comparable skill, although SDM 4 is more advanced due to the application of event dependent changes. Moreover, the majority of the events are classified as non-extreme events and for these events average precipitation amount changes are applied, which makes SDM 4 similar to SDM 1. Furthermore, SDMs developed by the same research institution also tend to cluster together. This is the case for the different versions of the presented event based perturbation method, i.e. SDM 4 developed by Sørup et al. (2017). In that work, the authors suggest the application of one of the four presented SDMs, such that the dependence between SDMs is indirectly addressed.

FIGURE 7 Projected and observed extreme daily summer precipitation amounts in the PP experiment.

Additionally, clusters combining different SDMs often include similar representative cluster members as indicated in Figure 6b. More precisely, the climate change signals for the quantile perturbation method, the change factor of mean method and the event-based change factor method are close when the change in the number of wet days is negligible and the extreme precipitation amounts change approximately by the same amount as the moderate precipitation amounts.

4.3 | Precipitation projections for the original and condensed sub-ensemble

The changes in the average winter precipitation accumulation and extreme daily summer precipitation amount for the original and final condensed ensemble are presented in Figure 8. Additionally, for each SDM, the variation of the median climate change signal for 100 random ensembles is presented. The size of these random ensembles equals the size of the condensed ensemble after the first clustering step (Table 2).

FIGURE 8 a) Projections for the average winter precipitation accumulation for the original (dark grey filled boxplot) and condensed ensemble (light grey filled boxplot). For each SDM, also the variation of the median climate change signal for 100 random ensembles, equally sized to the corresponding condensed SDM sub-ensemble (Table 2), is shown by the third boxplot (non filled boxplot). b) Projections for the extreme summer precipitation amount for the original, condensed and random ensemble.

As indicated in Figure 8, the overall good agreement between the spread of the original and condensed ensembles for both variables and each of the SDMs indicates that the spread is preserved and the ensemble cascade is successfully turned upside down. In some cases, however, the main statistics are not accurately reproduced. For instance, for SDM 2, the 25th and 50th percentile changes in the average winter precipitation accumulation are accurately reproduced, whereas the 75th percentile change is underestimated by the condensed ensemble (Figure 8a). The underestimation is explained by the presence of large clusters with a neutral to positive standardised change signal (Figure 2). Replacing these large clusters by only one representative member leads to an underestimation of the 75th percentile change. Also the statistics of the extreme daily summer precipitation amount changes are not preserved (Figure 8b). More specifically, for SDM 2, larger clusters are defined for the negative standardised climate change signals (Figure 3). Hence, replacing these clusters by one representative member yields an overestimation of the statistics. Over- or underestimation of the main statistics is also observed for the remaining SDMs and second clustering step. To overcome such over- or underestimation, the representative cluster members could be weighted based on the total number of cluster members and the relative position of the representative member with respect to the remaining cluster members. However, one might question whether weighting to better reproduce the statistics of the original ensemble would make much sense. The statistics of the original ensemble might be biased since the dependence between and the skill of the ensemble members is not accounted for.

Based on a comparison with the variation of the median signal for 100 random ensembles, the condensed ensembles do not under perform. On the contrary, based on the median change in the extreme summer precipitation amounts by SDM 2, for instance, the condensed ensemble outperforms the random ensemble and original ensemble. Due to the Clausius-Clapeyron relation (Manola et al., 2018), indicating an increase of the precipitation amounts by 7% per degree temperature increase, an intensification of the extreme daily summer precipitation amount is actually expected. The latter has not been observed in the original nor random ensemble, but by coincidence appears in the condensed ensemble.

4.4 | Urban and river impact projections for the original and condensed sub-ensemble

The projected peak flow discharges are presented in Figure 9, with the simulated results for a return period between ten and 100 years in Figure 9a. For some return periods, a detailed comparison between the original and condensed ensemble is made in Figure 9b. For the discharges with a return period up to ten years, there are no differences between the median of the original and condensed ensemble. However, differences appear in the results for the higher return periods and, moreover, increase for increasing return period. Finally, for the discharges with a 100-year return period, the difference reaches $1.1 \text{ m}^3/\text{s}$, corresponding with 5% difference in the associated climate change impact factor. Also the difference in the spread between the original and condensed ensemble increases for increasing return period. In this way, for a return period of 100 years, interquartile range for the condensed ensemble matches the 25th-50th percentile range for the original ensemble.

Figure 10 presents the simulated overflow volumes for the stormwater retention basin for the original and condensed ensemble. When studying the overflow results for a 20-year return period, which is typically used for the design of urban stormwater systems in Flanders (Coördinatiecommissie Integraal Waterbeleid (CIW), 2018), the statistics for the condensed ensemble overestimate the statistics for the original ensemble. More specifically, the difference measures 18 m^3 for the median, 42 m^3 for the 75th percentile and 34 m^3 for the 25th percentile. While the deviation with respect to the median is negligible, the deviation with respect to the 75th percentile is not and corresponds to approximately 17% of the reservoir volume.

FIGURE 9 a) Simulated peak flow discharges in function of their empirical return period for the projected time series of the original and condensed ensembles. The grey area indicates the total spread for the original ensemble and the black lines depict the results for the members of the condensed ensemble. b) Simulated peak flow discharges with a 10-, 20-, 50- and 100-year return period. The results for the original and condensed ensemble are shown in the boxplots, outliers are not shown.

354 The differences in impact statistics between the original and condensed ensemble are primarily introduced by
 355 the clustering step. This is for instance observed in the similarity in the overestimation of the 75th percentile change
 356 between the extreme daily summer precipitation amount and the overflow volume (Figure 6b versus Figure 10b).
 357 Moreover, the differences in impact results might indicate that the chosen variable is not representative and does not
 358 drive the impact model. More specifically, there is no reason to assume that the impact for the urban stormwater
 359 retention basin would be mainly explained by the extreme daily summer precipitation amount with a 10-year return
 360 period and not by a different return period or by extreme sub-daily precipitation amounts. Depending on the type
 361 of stormwater system, e.g. infiltration reservoirs, the reservoir is generally filled to a certain level for most of the
 362 time. Hence, the occurrence of an overflow event and the corresponding overflow volume depend on the alternating
 363 occurrence of wet and dry periods and the precipitation amounts during the wet periods. The same applies to the
 364 extreme peak flow discharges. More specifically, peak flow discharges are driven by the soil moisture balance, which
 365 in turn is influenced by the cumulative precipitation volumes, the intensity of precipitation events and the dry periods
 366 between successive precipitation events.

FIGURE 10 a) Simulated overflow volumes in function of their empirical return period for the projected time series of the original and condensed ensembles. The grey area shows the total spread for the original ensemble and the black lines depict the results for the members of the condensed ensemble. b) Close-up of the results for an overflow event with a 20-year return period. The results for the original and condensed ensemble are shown in the boxplots, outliers are not shown.

5 | DISCUSSION AND CONCLUSIONS

The approach is successfully demonstrated using a multi-ensemble with 240 members (48 GCMs $\times$ five SDMs/GCM). The ensemble members have not been chosen to favour nor discriminate methods or models. Using the presented approach, the original 240 membered multi-ensemble is condensed to eight (12) members for average seasonal winter (summer) precipitation accumulation and 18 (22) for extreme daily winter (summer) precipitation amount with a 10-year return period. In the first clustering step, clusters are defined between the climate model skill and associated climate change signal. In here, different realisations of the same climate model and climate models developed by the same research institution often become members of the same cluster. No correlation is observed between the number of clusters and the complexity of the SDM. The ensemble size reduction rates for this first step are comparable to the reduction rates found in the studies by Dalelane et al. (2018) and Mendlik and Gobiet (2016), both applying a different technique. In the second clustering step, combining the prior selected climate change signals and the downscaling skill, the existence of SDM inter-dependence is confirmed. In particular, SDMs with similarities in the downscaling methodology tend to cluster together. The dependence between SDMs generally remains unnoticed

in the climate change impact studies. However, it has significant implications and produces biased projections and uncertainty estimates. This requires further research by the statistical downscaling community. Furthermore, in case of clustering across SDMs, the clusters frequently include projections for similar models. The latter points to the interaction between SDMs and GCMs and has also been observed in the studies by Anandhi et al. (2011) and Mezghani et al. (2019). Overall, the proposed design approach generally preserves the spread and main statistics of the original ensemble. In some cases, particularly when large clusters are defined, the main statistics are over- or underestimated. Theoretically, this could be resolved by an additional weighting step. However, the relevance of such weighting step is questionable as the statistics for the original ensemble are based on dependent ensemble members.

Advanced knowledge on the impact model and underlying processes is required to thoroughly exploit the potential for ensemble size reduction. One could easily select a secondary variable for clustering and thus not obtain a representative sub-ensemble of impact results. Moreover, some models respond to a range of hydro-meteorological variables and their response times might extend over several seasons. As the number of clusters and cluster characteristics depend on the variable and season, considering different variables and seasons would make the presented approach lose its potential for ensemble size reduction. Further research should therefore involve the identification of a sub-ensemble which is representative for different variables, as for instance done by Cannon (2015) and Farjad et al. (2019). For the presented methodology, the identification of such an ensemble might be based on the similarities between clusters for different variables. Challenges also arise when applying the presented approach for impact applications on fast responding water systems, i.e. urban hydrology. Nor global, regional, or limited area climate models produce precipitation output series with a 1- to 10-minutes time step, which is typically prescribed for urban hydrology (Ochoa-Rodriguez et al., 2015). Instead, daily precipitation projections are most often considered. Clusters based on such precipitation projections may, however, differ from those one would obtain based on sub-daily extreme precipitation amounts.

Since the IPCC currently does not recommend any approach for the design of multi-ensembles, the battle continues between the proponents for member weighting and selection, and the opponents, who prefer a more conservative

approach and equiprobable weighting of all available ensemble members. Moreover, the results reported in literature on ensemble member weighting are twofold. Some studies point to negligible and moderate differences, whereas others find significant differences. Therefore, a remaining key question is whether a best ensemble design approach and best ensemble size reduction technique does actually exist. Also, the tasks of evaluating the skill of the climate models and SDMs, of becoming familiar with the ensemble design approach and of, eventually, applying such an approach, remain for now time-demanding. It is therefore questioned whether the total effort involved in the design of a condensed ensemble outweighs the effort required to post-process the original ensemble of time series and/or apply these for impact modelling.

acknowledgements

E. Van Uytven is funded by a doctoral grant from the Research Foundation – Flanders (F.W.O., grant number 11ZY418N). This financial support is gratefully acknowledged. The authors would like to thank colleague S. Moustakas from KU Leuven for his help with the rainfall run-off simulations.

references

- Addor, N. and Melsen, L. A. (2019) Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resources Research*, **55**, 378–390.
- Alexander, K. and Easterbrook, S. M. (2015) The software architecture of climate models: a graphical comparison of cmip5 and emicar5 configurations. *Geoscientific Model Development*, **8**, 1221–1232.
- Anandhi, A., Frei, A., Pierson, D. C., Schneiderman, E. M., Zion, M. S., Lounsbury, D. and Matonse, A. H. (2011) Examination of change factor methodologies for climate change impact assessment. *Water Resources Research*, **47**.
- 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. (2013) Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution. *Journal of Geophysical Research Atmospheres*, **118**, 3956–3971.
- Arakawa, A. (2004) The cumulus parameterization problem: Past, present, and future. *Journal of Climate*, **17**, 2493–2525.

- Benestad, R., Mezghani, A. and Parding, K. M. (2015) esd- the empirical-statistical downscaling tool & its visualisation capabilities.
- Benestad, R., Sillmann, J., Thorarinsdottir, T. L., Guttorp, P., Mesquita, M. D., Tye, M. R., Uotila, P., Maule, C. F., Thejll, P., Drews, M. and Parding, K. M. (2017) New vigour involving statisticians to overcome ensemble fatigue. *Nature Climate Change*, **7**, 697–703.
- Bultot, F., Coppens, A. and Dupriez, G. (1993) Estimation de l'évapotranspiration potentielle en Belgique. *Météorologique de Belgique - Koninklijk Meteorologisch Instituut van België*, 28.
- Bürger, G., Murdock, T. Q., Werner, A. T., Sobie, S. R. and Cannon, A. J. (2012) Downscaling extremes—an intercomparison of multiple statistical methods for present climate. *Journal of Climate*, **25**, 4366–4388.
- Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. and O'Connell, P. E. (2008) RainSim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling and Software*, **23**, 1356–1369.
- Cannon, A. J. (2015) Selecting gcm scenarios that span the range of changes in a multimodel ensemble: Application to cmip5 climate extremes indices. *Journal of Climate*, **28**, 1260–1267.
- Chen, J., Brissette, F. P., Lucas-Picher, P. and Caya, D. (2017) Impacts of weighting climate models for hydro-meteorological climate change studies. *Journal of Hydrology*, **549**, 534–546.
- Coördinatiecommissie Integraal Waterbeleid (CIW) (2018) Code van goede praktijk voor rioleringsystemen. URL: <http://www.integraalwaterbeleid.be/nl/publicaties/code-goede-praktijk-rioleringsystemen>.
- Crawford, J., Venkataraman, K. and Booth, J. (2019) Developing climate model ensembles: A comparative case study. *Journal of Hydrology*, **568**, 160 – 173.
- Dai, A. (2006) Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, **19**, 4605–4630.
- Dalelane, C., Früh, B., Steger, C. and Walter, A. (2018) A Pragmatic Approach to Build a Reduced Regional Climate Projection Ensemble for Germany Using the EURO-CORDEX 8.5 Ensemble. *Journal of Applied Meteorology and Climatology*, **57**, 477–491.
- Davini, P., Corti, S., D'Andrea, F., Rivière, G. and von Hardenberg, J. (2017) Improved Winter European Atmospheric Blocking Frequencies in High-Resolution Global Climate Simulations. *Journal of Advances in Modeling Earth Systems*, **9**, 2615–2634.
- Deser, C., Phillips, A., Bourdette, V. and Teng, H. (2012) Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, **38**, 527–546.

- 456 Fadhel, S., Rico-Ramirez, M. A. and Han, D. (2017) Uncertainty of Intensity–Duration–Frequency (IDF) curves due to varied
457 climate baseline periods. *Journal of Hydrology*, **547**, 600–612.
- 458 Farjad, B., Gupta, A., Sartipizadeh, H. and Cannon, A. (2019) A novel approach for selecting extreme climate change scenarios
459 for climate change impact studies. *Science of The Total Environment*, **678**, 476 – 485.
- 460 Fowler, H. J., Blenkinsop, S. and Tebaldi, C. (2007) Linking climate change modelling to impacts studies: recent advances in
461 downscaling techniques for hydrological modelling. *International Journal of Climatology*, **27**, 1547–1578.
- 462 Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T. (2012) Technical Note: Downscaling RCM precipita-
463 tion to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System*
464 *Sciences*, **16**, 3383–3390.
- 465 Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S.,
466 San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanar, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J.,
467 Pórtolés, J., Rätty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M. J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T.,
468 Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A. M., Cardoso, R. M., Soares, P. M., Czernecki, B. and Pagé, C.
469 (2018) An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE
470 perfect predictor cross-validation experiment. *International Journal of Climatology*, 1–36.
- 471 Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A. and Rasmussen, R. M. (2014) An intercomparison of
472 statistical downscaling methods used for water resource assessments in the united states. *Water Resources Research*, **50**,
473 7167–7186.
- 474 Hanel, M., Kožín, R., Heřmanovský, M. and Roub, R. (2017) An R package for assessment of statistical downscaling methods
475 for hydrological climate change impact studies. *Environmental Modelling and Software*, **95**, 22–28.
- 476 Hartung, K., Svensson, G. and Kjellström, E. (2017) Resolution, physics and atmosphere–ocean interaction – How do they in-
477 fluence climate model representation of Euro-Atlantic atmospheric blocking? *Tellus A: Dynamic Meteorology and Oceanog-*
478 *raphy*, **69**, 1406252.
- 479 Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J. M., Wibig, J., Casanueva, A. and Soares,
480 P. M. M. (2018) Comparison of statistical downscaling methods with respect to extreme events over Europe: Validation
481 results from the perfect predictor experiment of the COST Action VALUE. *International Journal of Climatology*, 1–22.
- 482 Hessami, M., Gachon, P., Ouarda, T. B. and St-Hilaire, A. (2008) Automated regression-based statistical downscaling tool.
483 *Environmental Modelling and Software*, **23**, 813–834.

- Hosseinzadehtalaei, P., Tabari, H. and Willems, P. (2017) Uncertainty assessment for climate change impact on intense precipitation: how many model runs do we need? *International Journal of Climatology*, **37**, 1105–1117.
- Iturbide, M., Bedia, J., Herrera, S., no Medina, J. B., Fernández, J., Frías, M., Manzananas, R., San-Martin, D., Cimadevilla, E., no, A. C. and Gutiérrez, J. (2019) The r-based climate4r open framework for reproducible climate data access and post-processing. *Environmental Modelling and Software*, **111**, 42 – 54.
- 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., 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. (2014) EURO-CORDEX: New high-resolution climate change projections for European impact research. *Regional Environmental Change*, **14**, 563–578.
- Karlsson, I. B., Sonnenborg, T. O., Christian, J., Trolle, D., Duus, C., Olesen, J. E., Jeppesen, E. and Jensen, K. H. (2016) Combined effects of climate models , hydrological model structures and land use scenarios on hydrological impacts of climate change. *Journal of Hydrology*, **535**, 301–317.
- Kassambara, A. and Mundt, F. (2016) *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. URL: <https://CRAN.R-project.org/package=factoextra>.
- Knutti, R., Masson, D. and Gettelman, A. (2013) Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, **40**, 1194–1199.
- Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M. and Eyring, V. (2017) A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, **44**, 1909–1918.
- Leduc, M., Laprise, R., de Elía, R. and Šeparović, L. (2016) Is institutional democracy a good proxy for model independence? *Journal of Climate*, **29**, 8301–8316.
- Li, J., Johnson, F., Evans, J. and Sharma, A. (2017) A comparison of methods to estimate future sub-daily design rainfall. *Advances in Water Resources*, **110**, 215–227.
- Lorenz, R., Davin, E. L. and Seneviratne, S. I. (2012) Modeling land-climate coupling in Europe: Impact of land surface representation on climate variability and extremes. *Journal of Geophysical Research Atmospheres*, **117**, 1–16.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2016) *cluster: Cluster Analysis Basics and Extensions*. URL: <https://CRAN.R-project.org/package=cluster>.

- Manola, I., van den Hurk, B., De Moel, H. and Aerts, J. C. J. H. (2018) Future extreme precipitation intensities based on a historic event. *Hydrology and Earth System Sciences*, **22**, 3777–3788.
- 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. (2017) The VALUE perfect predictor experiment: evaluation of temporal variability. *International Journal of Climatology*.
- Maraun, D., Wetterhall, F., Chandler, R. E., Kendon, E. J., Widmann, M., Brieren, 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. (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, **48**, 1–38.
- Maraun, D. and Widmann, M. (2018) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press.
- Maraun, D., Widmann, M. and Gutiérrez, J. M. (2018) Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology*, 1–12.
- Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R. A. I. (2014) VALUE : A framework to validate downscaling approaches for climate change studies. *Earth 's Future*, 1–14.
- Marx, A., Kumar, R., Thober, S., Zink, M., Wanders, N., Wood, E. F., Ming, P., Sheffield, J. and Samaniego, L. (2017) Climate change alters low flows in Europe under a 1.5, 2, and 3 degree global warming. *Hydrology and Earth System Sciences Discussions*, 1–24.
- Masson, D. and Knutti, R. (2011) Climate model genealogy. *Geophysical Research Letters*, **38**, 1–4.
- Mehrotra, R., Johnson, F. and Sharma, A. (2018) A software toolkit for correcting systematic biases in climate model simulations. *Environmental Modelling and Software*, **104**, 130 – 152.
- Mendlik, T. and Gobiet, A. (2016) Selecting climate simulations for impact studies based on multivariate patterns of climate change. *Climatic Change*, **135**, 381–393.
- Mendoza, P. A., Mizukami, N., Ikeda, K., Clark, M. P., Gutmann, E. D., Arnold, J. R., Brekke, L. D. and Rajagopalan, B. (2016) Effects of different regional climate model resolution and forcing scales on projected hydrologic changes. *Journal of Hydrology*, **541**, 1003–1019.

- Mezghani, A., Dobler, A., Benestad, R., Haugen, J. E., Parding, K. M., Piniewski, M. and Kundzewicz, Z. W. (2019) Subsampling impact on the climate change signal over poland based on simulations from statistical and dynamical downscaling. *Journal of Applied Meteorology and Climatology*, **58**, 1061–1078.
- Ntegeka, V., Baguis, P., Roulin, E. and Willems, P. (2014) Developing tailored climate change scenarios for hydrological impact assessments. *Journal of Hydrology*, **508**, 307–321.
- 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-Tuyts, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P. and Ten Veldhuis, M. C. (2015) Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal of Hydrology*, **531**, 389–407.
- Pechlivanidis, I., Gupta, H. and Bosshard, T. (2018) An information theory approach to identifying a representative subset of hydro-climatic simulations for impact modeling studies. *Water Resources Research*, **54**, 5422–5435.
- van Pelt, S. C., Beersma, J. J., Buishand, T. A., van den Hurk, B. J. J. M. and Kabat, P. (2012) Future changes in extreme precipitation in the rhine basin based on global and regional climate model simulations. *Hydrology and Earth System Sciences*, **16**, 4517–4530.
- Pennell, C. and Reichler, T. (2011) On the effective number of climate models. *Journal of Climate*, **24**, 2358–2367.
- Pitman, A. J. (2003) The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, **23**, 479–510.
- Potter, N., Ekström, M., Chiew, F., Zhang, L. and Fu, G. (2018) Change-signal impacts in downscaled data and its influence on hydroclimate projections. *Journal of Hydrology*, **564**, 12 – 25.
- 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. and Leung, R. (2015) A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, **53**, 323–361.
- Rajczak, J. and Schär, C. (2017) Projections of Future Precipitation Extremes Over Europe: A Multimodel Assessment of Climate Simulations. *Journal of Geophysical Research: Atmospheres*, **122**, 10,773–10,800.
- Rössler, O., Fischer, A. M., Huebener, H., Maraun, D., Benestad, R. E., Christodoulides, P., Soares, P. M., Cardoso, R. M., Pagé, C., Kanamaru, H., Kreienkamp, F. and Vlachogiannis, D. (2017) Challenges to link climate change data provision and user needs - perspective from the COST-action VALUE. *International Journal of Climatology*.

- Rybka, H. and Tost, H. (2014) Uncertainties in future climate predictions due to convection parameterisations. *Atmospheric Chemistry and Physics*, **14**, 5561–5576.
- Sanderson, B. M., Wehner, M. and Knutti, R. (2017) Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, **10**, 2379–2395.
- Seiller, G., Roy, R. and Anctil, F. (2017) Influence of three common calibration metrics on the diagnosis of climate change impacts on water resources. *Journal of Hydrology*, **547**, 280 – 295. URL: <http://www.sciencedirect.com/science/article/pii/S0022169417300744>.
- Smid, M. and Costa, A. C. (2017) Climate projections and downscaling techniques: a discussion for impact studies in urban systems. *International Journal of Urban Sciences*, 1–31.
- 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. (2018) Process-based evaluation of the value perfect predictor experiment of statistical downscaling methods. *International Journal of Climatology*, 1–28.
- Sørup, H. J. D., Georgiadis, S., Gregersen, I. B. and Arnbjerg-Nielsen, K. (2017) Formulating and testing a method for perturbing precipitation time series to reflect anticipated climatic changes. *Hydrology and Earth System Sciences*, **21**, 345–355.
- Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. (2013) *Summary for Policymakers*, book section SPM, 1–30. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. URL: www.climatechange2013.org.
- Sunyer, M. A., Hundeicha, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., Vormoor, K., Bürger, G., Hanel, M., Kriaučiūnienė, J., Loukas, A., Osuch, M. and Yücel, I. (2015) Inter-comparison of statistical downscaling methods for projection of extreme precipitation in europe. *Hydrology and Earth System Sciences*, **19**, 1827–1847.
- Sunyer, M. A., Madsen, H., Rosbjerg, D. and Arnbjerg-Nielsen, K. (2013) Regional interdependency of precipitation indices across denmark in two ensembles of high-resolution rcms. *Journal of Climate*, **26**, 7912–7928.
- Tabari, H., De Troch, R., Giot, O., Hamdi, R., Termonia, P., Saeed, S., Brisson, E., Van Lipzig, N. and Willems, P. (2016) Local impact analysis of climate change on precipitation extremes: are high-resolution climate models needed for realistic simulations? *Hydrology and Earth System Sciences*, **20**, 3843–3857.
- Tabari, H., Taye, M. T. and Willems, P. (2015) Water availability change in central Belgium for the late 21st century. *Global and Planetary Change*, **131**, 115–123.

- Thorndahl, S., Korup Andersen, A. and Larsen, A. B. (2017) Event-based stochastic point rainfall resampling for statistical replication and climate projection of historical rainfall series. *Hydrology and Earth System Sciences*, **21**, 4433–4448.
- Van Uytven, E. and Willems, P. (2018) Greenhouse gas scenario sensitivity and uncertainties in precipitation projections for central Belgium. *Journal of Hydrology*, **558**, 9–19.
- Vanden Broucke, S., Wouters, H., Demuzere, M. and van Lipzig, N. P. M. (2018) The influence of convection-permitting regional climate modeling on future projections of extreme precipitation: dependency on topography and timescale. *Climate Dynamics*.
- Vansteenkiste, T., Tavakoli, M., Ntegeka, V., Smedt, F. D., Batelaan, O., Pereira, F. and Willems, P. (2014) Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. *Journal of Hydrology*, **519**, 743 – 755.
- Watterson, I. G., Bathols, J. and Heady, C. (2014) What influences the skill of climate models over the continents? *Bulletin of the American Meteorological Society*, **95**, 689–700.
- Werner, A. T. and Cannon, A. J. (2016) Hydrologic extremes - an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, **20**, 1483–1508.
- Widmann, M., Bedia, J., Gutiérrez, J. M., Bosshard, T., Hertig, E., Maraun, D., Casado, M. J., Ramos, P., Cardoso, R. M., Soares, P. M. M., Ribalaygua, J., Pagé, C., Fischer, A. M., Herrera, S. and Huth, R. (2019) Validation of spatial variability in downscaling results from the value perfect predictor experiment. *International Journal of Climatology*, 1–27.
- Wilby, R. L. and Dawson, C. W. (2013) The statistical downscaling model: Insights from one decade of application. *International Journal of Climatology*, **33**, 1707–1719.
- Wilcke, R. A. and Bärring, L. (2016) Selecting regional climate scenarios for impact modelling studies. *Environmental Modelling and Software*, **78**, 191 – 201.
- Willems, P. (2009) A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling and Software*, **24**, 311 – 321.
- (2013) Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at uccle, belgium. *Journal of Hydrology*, **496**, 166–177.
- Willems, P. and Vrac, M. (2011) Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology*, **402**, 193–205.

- 619 Wolfs, V. and Willems, P. (2017) Modular conceptual modelling approach and software for sewer hydraulic computations.
620 *Water Resources Management*, **31**, 283–298.

Peer Review Only

Fighting big data and ensemble fatigue in climate change impact studies: can we turn the ensemble cascade upside down?

E. Van Uytven*, J. De Niel, P. Meert, V. Wolfs and P. Willems

a) The ensemble cascade

b) Turning the ensemble cascade upside down

a) Traditionally, the total ensemble includes GHS, several GCMs and different SDMs. Doing so, the total ensemble size builds up and this build-up is referred to as the ensemble cascade. b) By combining validation and inter-dependence analyses, as proposed in this study, the ensemble cascade is turned upside down and the total ensemble size is reduced.