Subsampling Impact on the Climate Change Signal over Poland Based on Simulations from Statistical and Dynamical Downscaling

ABDELKADER MEZGHANI, ANDREAS DOBLER, RASMUS BENESTAD, JAN ERIK HAUGEN, AND
KAJSA M. PARDING

Norwegian Meteorological Institute, Oslo, Norway

MIKOLAJ PINIEWSKI

*Potsdam Institute for Climate Impact Research, Potsdam, Germany, and Department of Hydraulic Engineering,
Warsaw University of Life Sciences, Warsaw, Poland*

ZBIGNIEW W. KUNDZEWICZ

*Potsdam Institute for Climate Impact Research, Potsdam, Germany, and Institute for Agricultural and Forest Environment
of the Polish Academy of Sciences, Poznan, Poland*

ABSTRACT

Most impact studies using downscaled climate data as input assume that the selection of few global climate models (GCMs) representing the largest spread covers the likely range of future changes. This study shows that including more GCMs can result in a very different behavior. We tested the influence of selecting various subsets of GCMs on the climate change signal over Poland from simulations based on dynamical and empirical–statistical downscaling methods. When the climate variable is well simulated by the GCM, such as temperature, results showed that both downscaling methods agree on a warming over Poland by up to 2° or 5°C assuming intermediate or high emission scenarios, respectively, by 2071–2100. As a less robust simulated signal through GCMs, precipitation is expected to increase by up to 10% by 2071–2100 assuming the intermediate emission scenario. However, these changes are uncertain when the high emission scenario and the end of the twenty-first century are of interest. Further, an additional bootstrap test revealed an underestimation in the warming rate varying from 0.5° to more than 4°C over Poland that was found to be largely influenced by the selection of few driving GCMs instead of considering the full range of possible climate model outlooks. Furthermore, we found that differences between various combinations of small subsets from the GCM ensemble of opportunities can be as large as the climate change signal.

1. Introduction

The changing climate impacts society and ecosystems in a broad variety of ways. For example, climate change affects human health, can cause damage to property and infrastructure, affect the crop yield, and cause changes to forests (Field et al. 2012). Global climate models (GCMs) combined with projections of greenhouse gas concentrations form the basis to assess climate projections in the future. The GCMs are designed to simulate large-scale phenomena and processes; however, they are not capable of providing local details on a country level. Therefore, downscaling techniques continue to be used as unavoidable alternatives to provide local climate information more accurately. There are two commonly used strategies in downscaling: regional climate downscaling, also referred to as dynamical downscaling (DD), and empirical–statistical downscaling (ESD).

In most regional or local climate impact studies, there has been an increasing demand in selecting a small

---

ð Denotes content that is immediately available upon publication as open access.

Corresponding author: Abdelkader Mezghani, abdelkaderm@met.no

number of climate simulations in order to reduce computational costs (Vano et al. 2015). Ideally, models that are employed in the simulation framework should represent the observed climate well (Lutz et al. 2016) and cover the range of future climate scenarios appropriately (Lutz et al. 2016; Vano et al. 2015), at least for the (climate) variable(s) under investigation and the region of interest. Often, the multimodel ensemble provides an overall good estimation of future climate at continental scales; however, individual models' performance vary for different regions, variables under investigation, and evaluation metrics, causing the subselection process to be subjective and based on the priorities and background of users (Pierce et al. 2009; Overland et al. 2011).

There are few key climatic aspects for selecting a subset of climate model simulations that are generally required. For instance, Pierce et al. (2009) indicated the importance of including climate simulations with perturbed initial conditions to reduce biases caused by the effects of natural internal climate variability. They also demonstrated that the good performance of the multimodel ensemble mean is often caused by the cancellation of systematic errors in the individual global models. Overland et al. (2011) underlined 1) the large-scale climate physics and 2) the use of a multiple models instead of a single one as key elements in choosing a subset of climate models simulations. McSweeney et al. (2012) emphasized the need of a reliable sampling strategy to extract the most credible information on regional climate change, while minimizing the redundancy in climate model simulations and hence reducing the computational time and costs. Mendlik and Gobiet (2013, 2016) argued that a representative subset of simulations enhances the quality of the ensemble as it excludes redundant simulations that would lead to biased statistics. Similarly, Wilcke and Bärring (2016) suggested the use of more specific information around the requirements of the impact study in climate model selection so that the characteristics of the subset fit the purpose of the research more appropriately. By selecting models having small biases in the present-day seasonal cycle, Herger et al. (2017) were able to considerably reduce the dependence between the members while maintaining "acceptable" spread in their projections. Most of the studies mentioned above have assumed that selecting a few GCMs that represent the largest spread in the multimodel ensemble of simulations would systematically lead to covering the largest sensitivity of the driven regional climate model (RCM) or local climate model. To our knowledge, this statement has never been tested, and different combinations of GCM–RCM may also lead to unexpected climate responses, hence, affecting the spread of the climate signal.

There have also been various applications of ESD approaches to project climate information at local scale into the future as an alternative to the dynamical downscaling (Benestad 2008; Huth 2002; Mezghani and Hingray 2009; Wilby et al. 1998). In basic terms, ESD tries to establish a direct statistical link between the large-scale climate variables, for example, as simulated by the GCMs, and the target local climate variables, for example, as measured by the observational network. A more detailed discussion of ESD approaches can be found in Maraun et al. (2015), Takayabu et al. (2016), and in textbooks such as Benestad et al. (2008) and Maraun and Widmann (2017). An exhaustive list of studies is given below. Pioneering works were done by Wilby (1998), who implemented a statistical downscaling of daily precipitation using daily airflow and seasonal teleconnection indices. The use of ESD moved on with Huth (2002), who used statistical downscaling to estimate daily temperature in Central Europe. Later, Linderson et al. (2004) proposed using statistical downscaling to construct scenarios of precipitation in southern Sweden. It was also evident that ESD methods were useful tools for assessing uncertainties as they are able to account for the spread in large multimodel ensembles (e.g., Benestad 2002; Schmidli et al. 2007). For instance, Chen et al. (2006) used the method to quantify the uncertainty originating from using GCMs in regional precipitation changes over Sweden. Few attempts have also been made by comparing ESD results with those from dynamical downscaling in simulating climate variables (Hellström et al. 2001; Haylock et al. 2006; Chen et al. 2012). For instance, Busuioc et al. (2001) found a good agreement between the two methods in estimating winter precipitation changes over Romania. Similarly, Mezghani et al. (2017) found a good agreement in simulating temperature warming over Poland; however, disagreements in simulating precipitation changes are obtained, especially when a high emission scenario and the end of the century are considered.

Much of the downscaling within the climate science community is coordinated within a project of the World Climate Research Programme's (WCRP) project known as the Coordinated Regional Downscaling Experiment (CORDEX; Giorgi and Lionello 2008). Within this project, an ensemble of high-resolution regional climate projections for Europe (EURO-CORDEX, the European branch of the CORDEX initiative) have been made available for use in climate change impact, adaptation, and mitigation studies (Jacob et al. 2014). Even though the EURO-CORDEX ensemble size is considered to be large ($\approx$15) by the RCM community, the regional climate simulations are usually based on few ($<$10) GCMs. Similarly, an ESD branch of the CORDEX experiment has recently been implemented following recommendations from

the IPCC (2007) (http://cordex.org/domains/cordex-esd/) and aims at supporting joint DD and ESD activities. This, for instance, permits increasing the size of downscaled ensembles because ESD is less time consuming. Therefore, one could expand the projections to the large ensemble, and one might improve the projected outcome by taking into account for uncertainties associated with the ensemble spread.

The main objective of this paper is to assess the impact of selecting a subset of GCMs being used in regional or local impact studies over Poland. As most of the above studies have assumed that the selection "should" preserve the full spread of the climate change signal, an additional test using a bootstrapping technique is performed on raw GCM results by resampling all possible combinations of selected GCMs and multimodel ensemble sizes. A secondary objective is to evaluate how robust (i.e., whether the two downscaling strategies agree or not) is the simulated climate change signal by confronting results of local climate projections based on dynamical and empirical downscaling strategies. The climate impact is defined here in terms of mean and spread of the future climate, based on temperature (as a robust climate variable simulated by the models) and precipitation (as a less robust simulated signal through GCMs). We also made use of the statistically based ESD simulations to investigate the more physically based ones, as it can easily be extended to cover the full set of global climate simulations based on phase 5 of the Coupled Model Intercomparison Project (CMIP5). The study has been conducted over Poland because a collection of high-resolution gridded observational and climate-projection datasets is publicly available (http://data.4tu.nl/repository/collection:CHASE-PL).

2. Data

a. CMIP5 GCMs

CMIP5 (WCRP 2011) comprises a set of multimodel ensembles of climate projections that follow different representative concentration pathways (RCPs). We used climate runs following the intermediate RCP4.5 and the high RCP8.5 emission scenarios. There are 108 runs for temperature and 105 runs for precipitation assuming the RCP4.5 emission scenario, and there are 81 runs for temperature and 77 runs for precipitation assuming the RCP8.5 emission scenario. The above runs include those that have perturbed initial conditions and cover the period 1900–2100, as described in Taylor et al. (2012). Table B1 in appendix B shows the distribution of the different model runs by the corresponding institution and for each climate variable.

b. Bias-corrected EURO-CORDEX simulations

The RCM simulations consist of 18 bias-corrected EURO-CORDEX simulations for the period 1949–2100 produced within the CHASE-PL project: Climate change impact assessment for selected sectors in Poland (http://www.chase-pl.pl). The CHASE-PL climate-projection dataset (CPLCP-BCDPT5) was created by Mezghani et al. (2016) and contains adjusted historical and projected daily minimum and maximum air temperatures and precipitation totals of nine regional climate simulations following the two representative concentration pathways RCP4.5 and RCP8.5 (Table 2).

c. Definition of the various multimodel ensembles of projections

To assess the influence of selecting few simulations from the multimodel "ensemble of climate opportunities" such as CMIP5-based simulations, we constructed various multimodel ensembles of climate projections depending on the 1) driving GCMs and 2) downscaling method involved in the simulation framework. In total, we identified five (sub) ensembles of simulations for each climate variable (Table 1) that will be subject to various tests and analyses. These are the GCM-All ensemble of all CMIP5 raw GCM simulations (i.e., no downscaling is involved), the DD-All (also referred to as DD for simplicity) bias-corrected EURO-CORDEX simulations as described in Table 2, the ESD-All (also referred to as ESD for simplicity) ensemble of all empirical–statistical downscaled simulations, and the GCM-Com and ESD-Com as subsamples of GCM-All and ESD-All ensembles by keeping the set of simulations based on the common GCMs used in both DD and ESD only. Then, near- and far-future changes in temperature and precipitation are estimated for each of the predefined ensembles described above and for both the RCP4.5 and RCP8.5 emission scenarios.

3. Downscaling methods

Projections of local climate over Poland involved two widely used downscaling methods such as dynamical and empirical–statistical downscaling (Fig. 1). Details on how these methods are implemented and used in this study are described below along with a developed strategy on how one could achieve a fair comparison of the downscaled results.

a. Dynamical downscaling

The DD is based on bias-corrected EURO-CORDEX RCM simulations (section 2). The bias-corrected procedure follows Gudmundsson et al. (2012), which

TABLE 1. Definition of the multimodel ensembles based on the ''ensemble of opportunities'' of CMIP5 simulations. The first column gives the acronym of the defined ensemble or experiment. A short description of the simulations that belong to each ensemble or experiment being tested is given in the second column. The third and fourth columns show the number of runs and the corresponding GCMs for both temperature T and precipitation P assuming intermediate (RCP4.5) and high (RCP8.5) emission scenarios.

Expt	Description	No. of simulations RCP4.5: T—P	No. of simulations RCP8.5: T—P
GCM-All	Ensemble of all CMIP5 GCM simulations as described in Table B1	108—105 runs 39—36 GCMs	81—78 runs 35—35 GCMs
DD-All	Ensemble of all EURO-CORDEX bias-corrected simulations as defined in Table 2	9—9 runs 4+4—4+4 GCMs+RCMs	9—9 runs 4+4—4+4 GCMs+RCMs
ESD-All	Ensemble of all ESD simulations	108—105 runs 39—36 GCMs	81—78 runs 35—35 GCMs
GCM-Com	Ensemble of the subset of GCM-All based on common GCMs used in both DD and ESD	9—8 runs 4—4 GCMs	9—7 runs 4—4 GCMs
ESD-Com	Ensemble of the subset from ESD-All based on the set of GCMs common to both DD and ESD	9—8 runs 4—4 GCMs	9—7 runs 4—4 GCMs

attempts to adjust the distribution of simulated training data so that they fit closely to the observed climatology. The quantile-mapping approach was used to correct any systematic biases in the RCM simulations (Piani et al. 2010). The quantiles of the simulated time series were then mapped on to the observed ones at each grid cell and the number of quantiles was set to 1000. Further, a constant model bias similar to the one proposed by Boé et al. (2007) was used to enable simulating values lying outside the range of the training period including an adjustment of the wet-day frequencies between observations and simulations. Furthermore, the RCM outputs were mapped onto the high-resolution (5×5 km) gridded daily precipitation and temperature dataset (GDPT5; Berezowski et al. 2016) to produce high-resolution projections of gridded daily precipitation and temperature for the three time slices: 1971–2000 (historical), 2021–50 (near-future horizon), and 2071–2100 (far-future horizon), as described in section 2. Additional details about the bias correction procedure used in this study are presented in

Mezghani et al. (2017). A validation of both downscaling methods is presented in appendix A.

b. Empirical–statistical downscaling method

In this study, a similar approach as proposed by Benestad (2001) was adopted. First, common-empirical orthogonal functions (common-EOF) were applied to the reanalysis and modeled fields by the global models to extract common spatial patterns to define the predictors. Then, part of the corresponding principal components associated with the reanalysis (in this case, NCEP reanalyses) was used to calibrate the empirical–statistical model while the corresponding spatial pattern associated with the GCM was used for the downscaling (identical spatial patterns). Similar to EOFs, the predictand was represented by a set of principal components—as in Benestad et al. (2015)—computed on the set of available weather stations to represent local climate patterns. Further, large-scale fields of 2-m air temperature and precipitation were

TABLE 2. Summary of the CORDEX simulations. The simulations are based on a combination of RCMs in which the lateral boundary conditions are driven by GCMs. ICHEC is the Irish Centre for High-End Computing. Other GCM institute expansions can be found online (https://www.ametsoc.org/PubsAcronymList). The institute expansions for the RCMs can be found online (https://is-enes-data.github.io/CORDEX_RCMs_info.html).

| No.	GCM Institute	GCM Model	GCM Run	RCM Institute	RCM Model	Period
1	CNRM-CERFACS	CNRM-CM5	r1i1p1	CLMcom	CCLM4-8-17	1 Jan 1950–31 Dec 2100
2	CNRM-CERFACS	CNRM-CM5	r1i1p1	SMHI	RCA4	1 Jan 1970–31 Dec 2100
3	ICHEC	EC-EARTH	r12i1p1	CLMcom	CCLM4-8-17	1 Dec 1949–31 Dec 2100
4	ICHEC	EC-EARTH	r12i1p1	SMHI	RCA4	1 Jan 1970–31 Dec 2100
5	ICHEC	EC-EARTH	r1i1p1	KNMI	RACMO22E	1 Jan 1950–31 Dec 2100
6	ICHEC	EC-EARTH	r3i1p1	DMI	HIRHAM5	1 Jan 1951–31 Dec 2100
7	IPSL	IPSL-CM5A-MR	r1i1p1	SMHI	RCA4	1 Jan 1970–31 Dec 2100
8	MPI	M-MPI-ESM-LR	r1i1p1	CLMcom	CCLM4-8-17	1 Jan 1970–31 Dec 2100
9	MPI	M-MPI-ESM-LR	r1i1p1	SMHI	RCA4	1 Dec 1949–31 Dec 2100
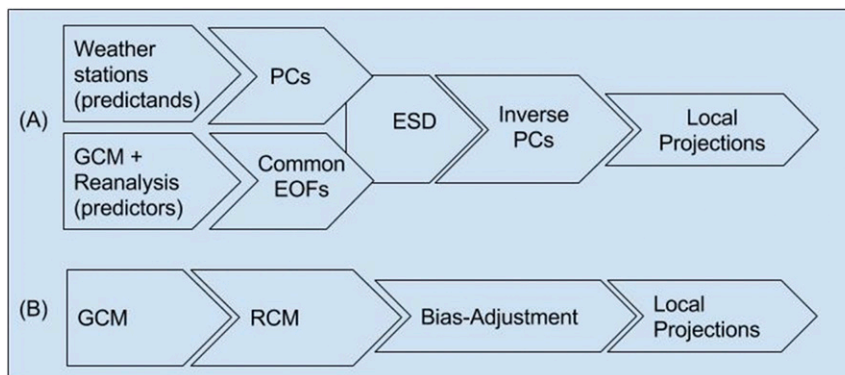
FIG. 1. Schematic diagram of the two downscaling strategies: (A) ESD and (B) DD.

used as predictors to project local mean temperature and precipitation, respectively.

c. Enabling a fair comparison of the downscaling methods

Both DD and ESD make use of independent sources of information to downscale climate information and are often designed for different research questions. This makes them more powerful when combined, as converging results would suggest a more robust signal. They have different strengths and weaknesses, and in some cases, one may outperform the other. For instance, ESD can only provide projections for variables for which long and good quality of observational datasets are available, but because of its low computational demands it is well suited for downscaling large multimodel ensembles (Benestad et al. 2016). DD, on the other hand, can provide a more complete picture of the processes involved and does not require a full set of observations, but it is more computationally expensive and is typically applied to a smaller ensemble of GCM runs. Observations are important in both DD and ESD for validation purposes that are often necessary before the downscaled results can be used in impact studies (cf. appendix A). Whereas the ESD method was applied to observations from selected weather stations and using the full set of CMIP5 GCMs, DD used gridded datasets and a small subset of GCMs. This makes the comparison between the two downscaling methods difficult to achieve. To do so, the DD projections were additionally bilinearly interpolated onto the set of weather stations used in ESD. Downscaled values from both strategies were then averaged over all common GCMs and locations. To enable a fair comparison between the two downscaling strategies, we selected the common set of GCMs employed in the two downscaling strategies. The selection led to a limited number of four GCMs, which are marked with stars in Table B1 and are summarized in Table 2. ESD results were

additionally weighted on the basis of how many times the driving GCMs have been used in regional downscaling (i.e., different RCMs driven by the same GCM).

4. Results

a. Projected future climate changes based on the small set of common GCMs

Seasonal and annual changes in projected temperatures and precipitation by both ESD and DD methods assuming the intermediate and high emission scenarios are presented in Table 3 and Figs. 2 and 3. Remember that here the changes are derived from the all bias-corrected DD dataset (DD-All, similar to DD) and the ESD-Com subensemble of common GCMs to both downscalings. Results are presented in Table 3 for projected seasonal and annual changes in mean temperature (absolute changes; °C) and monthly sums of precipitation (relative changes; %) by the near future (2021–50) and far future (2071–2100) and assuming both intermediate (RCP4.5) and high (RCP8.5) emission scenarios.

1) PROJECTED CHANGES BY THE DYNAMICAL DOWNSCALING

Projections made by the dynamical downscaling suggest increasing temperature in the future over Poland (Table 3). The annual mean temperature over Poland is expected to rise by 1.1° and 1.3°C for the period 2021–50 and by 2.0° and 3.6°C for the period 2071–2100, assuming the RCP4.5 and RCP8.5 emission scenarios, respectively. Results also exhibit low seasonal variations, except for winter RCP8.5 by 2071–2100 indicating an increase by up to 4.4°C. On average, there are small intermodel deviations varying approximately about 0.5° and 1.0°C on an annual and seasonal basis, respectively. Similarly to temperature, DD projected annual means of monthly sums of precipitation are

TABLE 3. Comparison between projected temperature and precipitation, as given by a summary of absolute changes in seasonal and annual means of daily temperature ΔT and relative changes in seasonal and annual means of monthly sums of precipitation ΔP (%) by the near (2021–50) and far (2071–2100) future time horizons and following the RCP4.5 and RCP8.5 emission scenarios, respectively. All values are estimated from the common subset of GCMs and averaged over the selected weather stations. Values in parentheses indicate the 5th and 95th percentiles, respectively, of the projected ensembles and, hence, represent the 90% confidence interval of the mean estimates.

Scenario by horizon	Method	Winter	Spring	Summer	Autumn	Annual
			ΔT (°C)			
RCP4.5 by 2021–50	ESD	+0.9 (+0.4, +1.6)	+1.2 (+1.0, +1.4)	+1.2 (+1.1, +1.8)	+1.2 (+1.0, +1.4)	+1.0 (+0.9, +1.5)
	DD	+1.2 (+0.3, +1.9)	+1.0 (+0.5, +1.8)	+1.0 (+0.6, +1.5)	+1.1 (+0.6, +1.6)	+1.1 (+0.7, +1.5)
RCP4.5 by 2071–2100	ESD	+2.2 (+1.2, +3.1)	+1.9 (+1.5, +2.2)	+2.3 (+1.8, +3.0)	+2.1 (+1.4, +2.5)	+2.2 (+1.5, +2.5)
	DD	+2.5 (+1.2, +3.4)	+2.0 (+1.1, +3.0)	+1.7 (+1.2, +2.5)	+1.9 (+1.4, +2.6)	+2.0 (+1.3, +2.6)
RCP8.5 by 2021–50	ESD	+1.7 (+0.8, +4.9)	+3.5 (+2.1, +4.6)	+1.5 (1.1, +2.1)	+1.6 (−0.1, +1.9)	+1.6 (+1.4, +3.3)
	DD	+1.4 (+0.3, +2.4)	+1.3 (+0.8, +2.1)	+1.1 (+0.6, +1.4)	+1.3 (+0.6, +1.8)	+1.3 (+0.8, +1.8)
RCP8.5 by 2071–2100	ESD	+5.1 (+3.6, +12.2)	+8.6 (+6.4, +11.1)	+4.0 (+2.9, +5.8)	+3.2 (+0.5, +4.4)	+4.7 (+4.2, +8.3)
	DD	+4.4 (+3.6, +5.5)	+3.2 (+2.4, +4.2)	+3.2 (+2.4, +4.1)	+3.5 (+2.7, +4.2)	+3.6 (+2.9, +4.1)
			ΔP (%)			
RCP4.5 by 2021–50	ESD	+ 5 (−7, +16)	+ 9 (+2, +16)	+ 6 (−5, +20)	+ 2 (−5, +11)	+ 5 (−1, +12)
	DD	+ 8 (−2, +17)	+ 8 (−1, +21)	+ 4 (−6, +16)	+ 6 (−7, +20)	+ 6 (+1, +12)
RCP4.5 by 2071–2100	ESD	+ 6 (−11, +22)	+ 4 (+13, +24)	+ 4 (−10, +21)	+ 0 (−8, +8)	+ 5 (−2, +14)
	DD	+16 (+4, +29)	+18 (+7, +29)	+ 5 (−9, +19)	+ 7 (−3, +17)	+10 (+4, +16)
RCP8.5 by 2021–50	ESD	+14 (−2, +32)	−1 (−12, +12)	−3 (−15, +13)	−4 (−16, +9)	+2 (−11, +17)
	DD	+10 (−1, +27)	+13 (+2, +26)	+ 5 (−5, +15)	+ 7 (−3, +20)	+ 8 (+3, +14)
RCP8.5 by 2071–2100	ESD	+36 (−7, +77)	+ 1 (−25, +28)	−15 (−49, +26)	−4 (−33, +30)	−4 (−17, +9)
	DD	+27 (+10, +44)	+27 (+12, +41)	+ 6 (−8, +22)	+14 (−5, +32)	+16 (+8, +26)

expected to increase for both future time horizons and emission scenarios. On average, the rate of precipitation increase is expected to be about 6% and 10% assuming the RCP4.5 emission scenario and about 8% and 16% assuming the RCP8.5 emission scenario. The strongest increases are expected to occur in winter and spring reaching up to approximately 30% by the far future horizon (2071–2100). However, summer changes exhibit a low and steady increase by about 5% regardless of the future horizon and the emission scenario of interest. Model differences in simulating precipitation changes were also found to be large and vary between −8% for spring and +44% for winter for the period 2071–2100 when assuming the RCP8.5 emission scenario.

2) PROJECTED CHANGES BY THE SMALL ESD ENSEMBLE

Like DD projections, ESD-Com based projections indicate a ubiquitous warming over Poland in the near and far futures and the annual warming rate is estimated to be about 1.0°C and 2.2°C, respectively, assuming the intermediate emission scenario RCP4.5. This warming accelerates and is expected to reach 4.7°C by the end of the twenty-first century assuming the high emission scenario RCP8.5. On a seasonal basis, results show low variations among seasons for the RCP4.5 emission scenario and enlarged for the RCP8.5 emission scenario. Seasonal changes are not consistent and depend on time horizon and the emission scenario of interest. For instance,

the highest increase is expected to occur in summer (1.2° ± 0.5°C) for the RCP4.5 emission scenario, whereas it is expected to occur in spring (8.6° ± 2.5°C) when considering the high emission scenario (RCP8.5). In contrast to temperature, projected precipitation show inconsistencies in simulating future changes. Assuming the RCP4.5 emission scenario, results show that Poland is expected to get more precipitation by about 5% in decades to come. Unlike RCP4.5, projections based on the RCP8.5 emission scenario reveal uncertain changes indicating wetter or drier climate conditions over Poland, especially by the end of the twenty-first century. This inconsistency remains on a seasonal level, except for winter during which there is a general agreement among ESD-Com simulations on an increase in monthly sums of precipitation, reaching up to ≈40% by 2071–2100 assuming the RCP8.5 emission scenario. However, it exhibits a large ensemble spread that varies between −7% and +77% (Table 3).

3) COMPARISON OF PROJECTIONS BASED ON THE COMMON SUBSET OF GCMs

The comparison between the two downscaling strategies suggests that when the climate variable is well simulated by the GCMs, which is the case for temperature, the two methods show consistent results. It also indicates that there is a close connection between the large-scale patterns and the local response of temperature, for example, between predictor and predictand as in ESD.
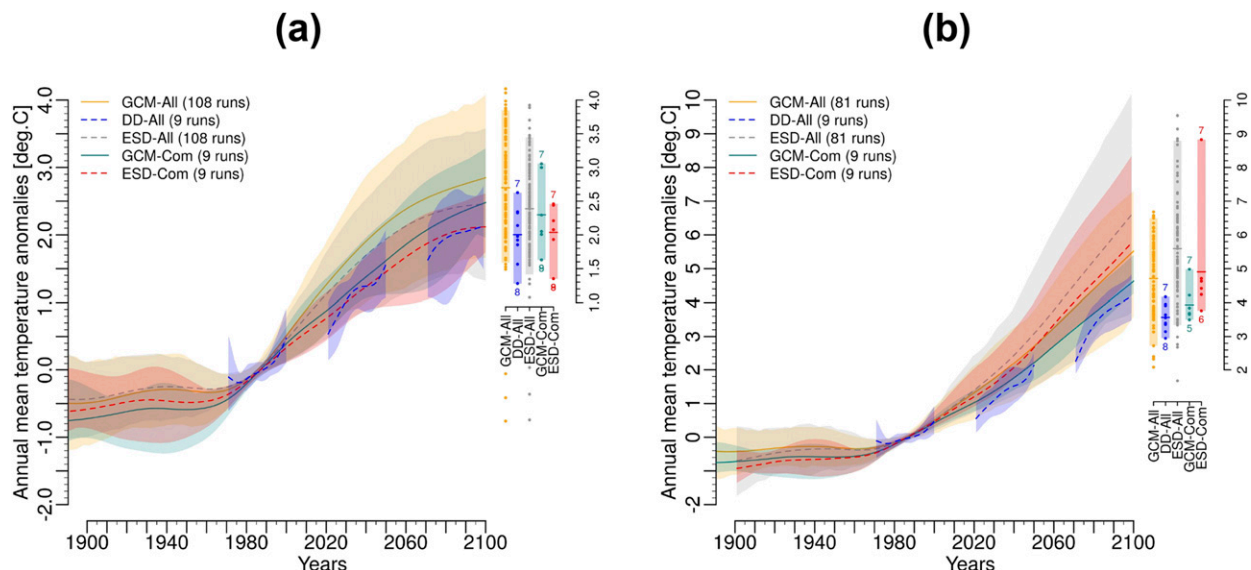
FIG. 2. Comparison between changes in raw (i.e., without downscaling; solid curves) and downscaled (dashed curves) global mean temperature (°C) anomalies assuming the (a) RCP4.5 and (b) RCP8.5 emission scenarios over Poland. These are derived from raw GCM-All ensemble (orange) and GCM-Com ensemble (dark cyan) and downscaled DD-All ensemble (blue), ESD-All ensemble (gray), and ESD-Com ensemble (red). Area averages are estimated from the spatial domain covering Poland. Envelopes indicate the 5th and 95th percentiles of the projected ensembles and, hence, represent the 90% confidence interval from the mean estimates. The legend displays the different (sub) ensembles as described in Table 1, and the values in parentheses indicate the number of runs included in each global and regional multimodel ensemble. Changes in annual mean temperature anomalies by the end of the twenty-first century (2071–2100) with regard to the reference period (1971–2000) for each scenario and each ensemble are drawn as boxes and inner dots. Inside each box, the mean (horizontal line) and the confidence interval (shading) are shown. The upper and lower limits of each box represent the minimum and maximum values for small ensembles and the 5th and 95th percentiles for large ensembles. The dots highlight individual climate model results. The numbers at the edges of each box refer to the EURO-CORDEX simulations described in Table 2.

When the climate variable is more difficult to predict by the climate models, such as precipitation, the results of the bias-corrected DD and ESD exhibit considerable differences. This suggests that it is necessary to include more than only precipitation to characterize the large-scale pattern (the predictor) in the ESD procedure because the large-scale precipitation may not be able to carry a climate change signal very well. The difference may also be related
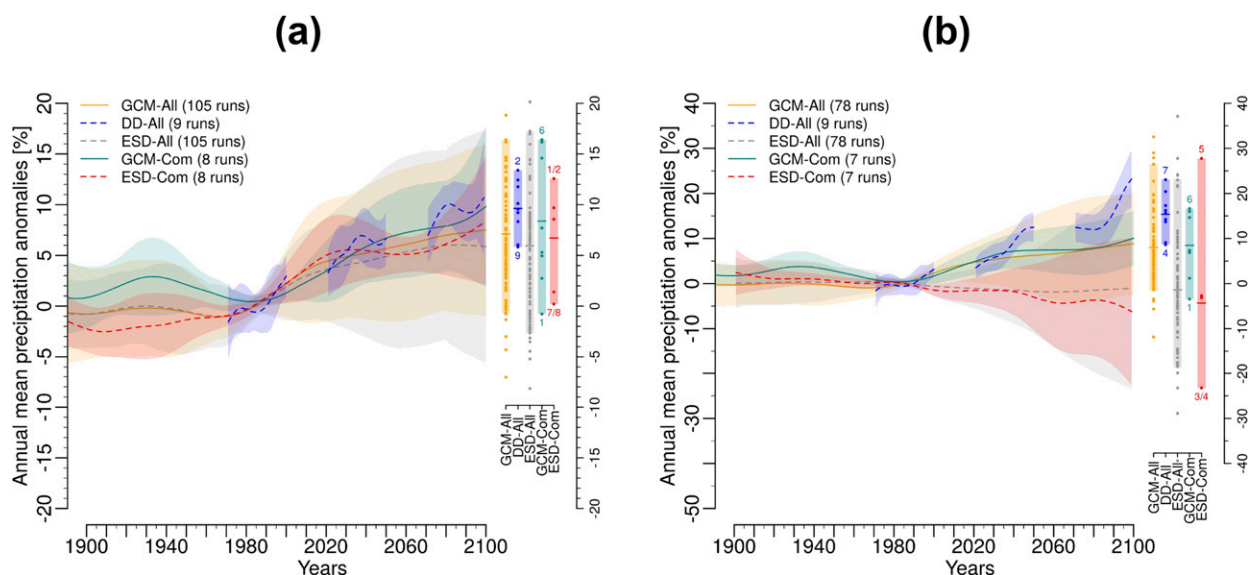

FIG. 3. As in Fig. 2, but for monthly sums of precipitation (relative changes; %) anomalies.

to the fact that temperature is less affected by local phenomena and is more homogeneous in space, unlike precipitation, which can be more variable in space and more affected by local conditions such as topography that are misrepresented in coarse-resolution GCMs (e.g., more than 1°).

For projected seasonal changes in precipitation, our results show merely good agreement in the magnitude of changes by 2021–50 assuming the RCP4.5 emission scenario. However, they exhibit large discrepancies when the end of the twenty-first century is considered; which shows weaker increases in ESD projections for all seasons than in bias-corrected DD projections, especially, when the RCP8.5 emission scenario is considered. For instance, winter and spring precipitation changes by ESD projections are expected to be about +6% and +4%, respectively, whereas the bias-corrected DD projections indicate higher increases of about +16% and +18%, respectively. An exception is made for winter, which shows similar order of magnitudes. Projected precipitation for autumn also differs depending on the downscaling method. The ESD projections show almost no changes (0% ± 8%) in the ensemble mean by 2071–2100 assuming the intermediate emission scenario, whereas the bias-corrected DD projections suggest an increase of about 7% in monthly sums of precipitation, with a range between −3% and +17% (Table 3).

Results also suggest a good agreement in both the magnitude and spread of projected seasonal changes in downscaled temperature, regardless of the representative concentration pathway of interest (Table 3). The only exception was found in spring in which there was a stronger increase in ESD than bias-corrected DD projections by approximately 2°C more in the near future (2021–50) and 5°C more in the far future (2071–2100), assuming the RCP4.5 and RCP8.5, respectively, and reaching up to approximately +9°C by the end of the twenty-first century when the high emission scenario is considered.

b. Impact of selecting few driving GCMs

To better understand the impact of the multimodel ensemble size on the climate change signal over Poland, a statistical test was performed using a bootstrapping technique applied to the set of GCM simulations, that is, using raw data without downscaling. The application of this test can be summarized in the following steps:

1) Select a random subset of n climate model simulations from the full "ensemble of opportunity," that is, CMIP5 ensemble. The subset represents then one subset of C combinations of n climate model simulations.

2) From this selection, estimate the mean value and range (maximum − minimum) of the mean climate change (temperature and precipitation) toward the end of the twenty-first century (2071–2100) with regard to the reference period (1971–2000).

3) Repeat steps 1 and 2 10 000 times to produce a bootstrap distribution of the mean and range of the climate change signal for ensembles of size n ($n = 1, \ldots, N$).

4) Last, repeat steps 1–3 for increments of multimodel ensemble sizes until the full ensemble is sampled (e.g., $N = 108$ for the temperature ensemble assuming the RCP4.5 emission scenario).

Because the CMIP5 ensemble includes multiple runs for some of the GCMs, this test was applied in two different ways by allowing or not allowing resampling of various realizations of the same GCM. This enables evaluating the uncertainty that may arise from internal climate variability (Deser et al. 2012). Although simulations from different GCMs may have shared code and developmental history and therefore are not guaranteed to be independent of each other, the exclusion of simulations from the same GCM reduces the interdependence within the ensemble. Results of the bootstrap test (Figs. 4 and 5) show that reducing the ensemble size can result in a large (positive or negative) bias in the ensemble mean, regardless of the climate variable and emission scenario of interest. This is evident from the large spread in the mean climate change associated with low ensemble sizes (Figs. 4a,b and 5a,b) and indicates that the ensemble mean can be heavily skewed if many simulations sharing strong climate change signal of the same sign are selected. A small ensemble size is also associated with a systematic underestimation of the range of possible climate change outcomes (Figs. 4c,d and 5c,d). The potentially misleading effect of a small ensemble size is enhanced when the climate change signal is strong, for example, when assessing the climate change signal assuming the high emission scenario RCP8.5 (Figs. 4b,d and 5b,d). For the high emission scenario (RCP8.5), the mean temperature change slightly increases (0.2°C) when limiting the sampling to only one realization from each GCM (Fig. 4b); however, for the intermediate RCP4.5 emission scenario (Fig. 4a) and for precipitation (Figs. 5a,b), the mean change signal remains unchanged while the uncertainty around the spread is impacted. This suggests that allowing resampling of multiple realizations from a single GCM can, but does not always, lead to a biased estimation of the ensemble mean and spread of the climate change signal.
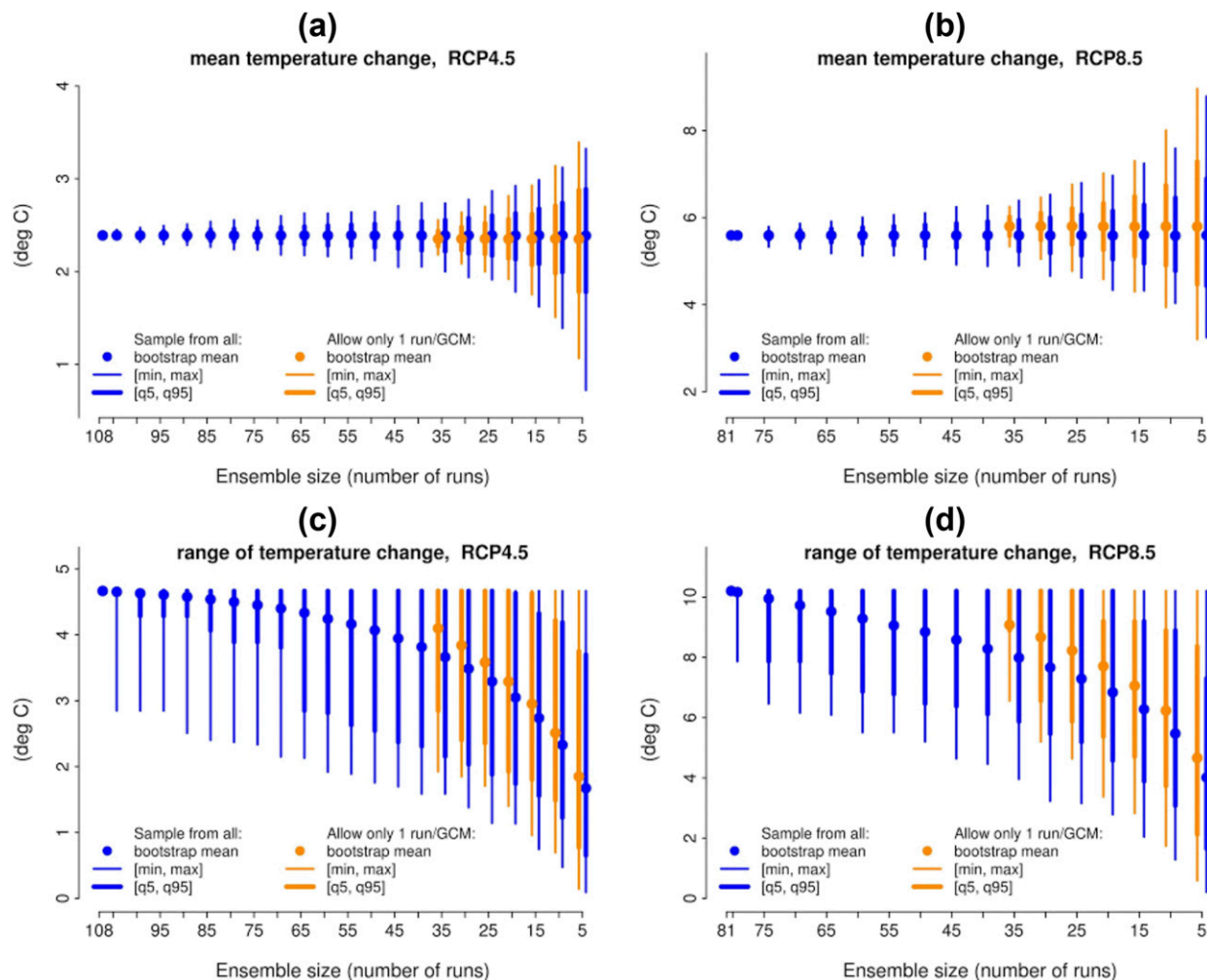
FIG. 4. Bootstrap test of ensemble size impact on the projected temperature change over Poland by 2071–2100 with regard to 1971–2000. The calculations were based on random sampling from the CMIP5 ensemble of raw GCM temperature projections. The figures show the (a),(b) mean and (c),(d) spread (maximum − minimum) of the temperature change as a function of the ensemble size assuming the (left) intermediate (RCP4.5) and (right) high (RCP8.5) emission scenarios, respectively. The plots show results from two experiments, allowing multiple simulations (blue) and only one GCM member (orange) in the selection. The dots represent the estimated mean value, and the thick and thin vertical segments represent the 90th confidence interval and the range from the bootstrap distribution, respectively. The range derived from the bootstrap distribution is displayed between the minimum and maximum estimates.

5. Discussion

Both DD and ESD have their drawbacks and strengths, and we believe that the combined use of the two downscaling methods would be beneficial to produce reliable projections in the future. Contrary to RCM output that can inform about both weather and climate, a limitation of an empirical–statistical downscaling method, such as the one implemented in this paper, is that it gives information about the climate only (i.e., fitting parameters of probability distribution functions) and not the weather (i.e., reproducing day-by-day patterns). Another limitation is the degree of applicability of the ESD approach to other regions and climate conditions that should also be tested, because it relies on the place-specific relationship between large-scale patterns and local climate response.

The temperature projections driven by the common subset of GCMs show systematic weaker changes in the mean temperature anomaly by almost 0.5°C by 2071–2100 and a systematically reduced ensemble spread in the future compared to the full ensemble of available simulations (Figs. 2 and 3). However, the bias-corrected DD and ESD projections of temperature (Figs. 2 and 6) agree well with each other both in terms of ensemble mean and spread when the same subset of models is taken into consideration. This lends support to both
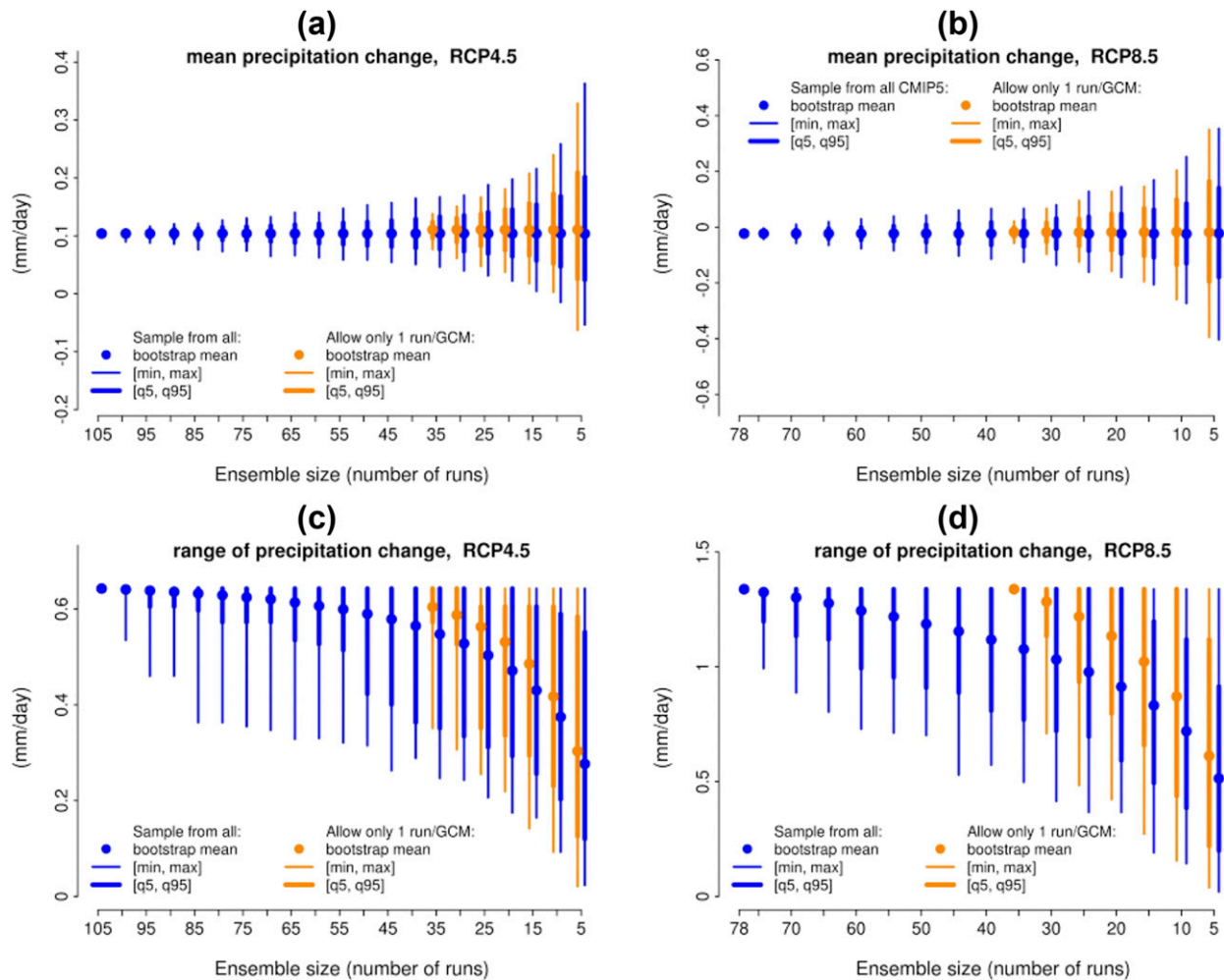
FIG. 5. As in Fig. 4, but for projected precipitation change.

downscaling methods, but suggests that the sampling of GCMs introduces a considerable bias (see section 4).

On average, downscaled projections indicate wetter future climate conditions over Poland in decades to come by assuming the intermediate emission scenario (RCP4.5), especially when the subset of GCMs as those used in the EURO-CORDEX simulations is considered. Figure 3 shows that, over all multimodel ensemble means, annual mean precipitation is expected to increase in the future regardless of the downscaling methods and the time horizons of interest and tends to closely follow raw GCM results (Figs. 6 and 7). The spread in simulated precipitation changes is largely affected by the selection of multimodel ensemble of projections. Ensembles based on the common set of GCMs (i.e., GCM-Com and ESD-Com) exhibit positive changes, whereas it is uncertain whether precipitation will increase or decrease in future when additional GCM simulations (i.e., large ensembles such as GCM-All and ESD-All) are included in the

multimodel ensemble. Nevertheless, bias-corrected DD tends to inflate the change signal in modeled precipitation by GCMs (i.e., raw GCM results), whereas downscaling based on ESD tends to weaken the climate change signal (see Figs. 6 and 7).

When assuming the RCP8.5 emission scenario, results of projected temperature changes (Fig. 2b) suggest that, if we discard the IPSL-CM5A GCM (simulation 7 in Table 2) from the ensembles, both the magnitude and the spread of temperature changes derived from the ESD-Com and DD-All datasets become very close and a better agreement between the two methods will be obtained. Results obtained with the IPSL-CM5A GCM exhibits also extreme changes in temperature when the DD results are considered. However, there are few discrepancies between the methods in simulating precipitation changes. For instance, the downscaled ESD ensemble mean is expected to decrease by 5% and is slightly increasing in the ESD-Com subset. ESD results
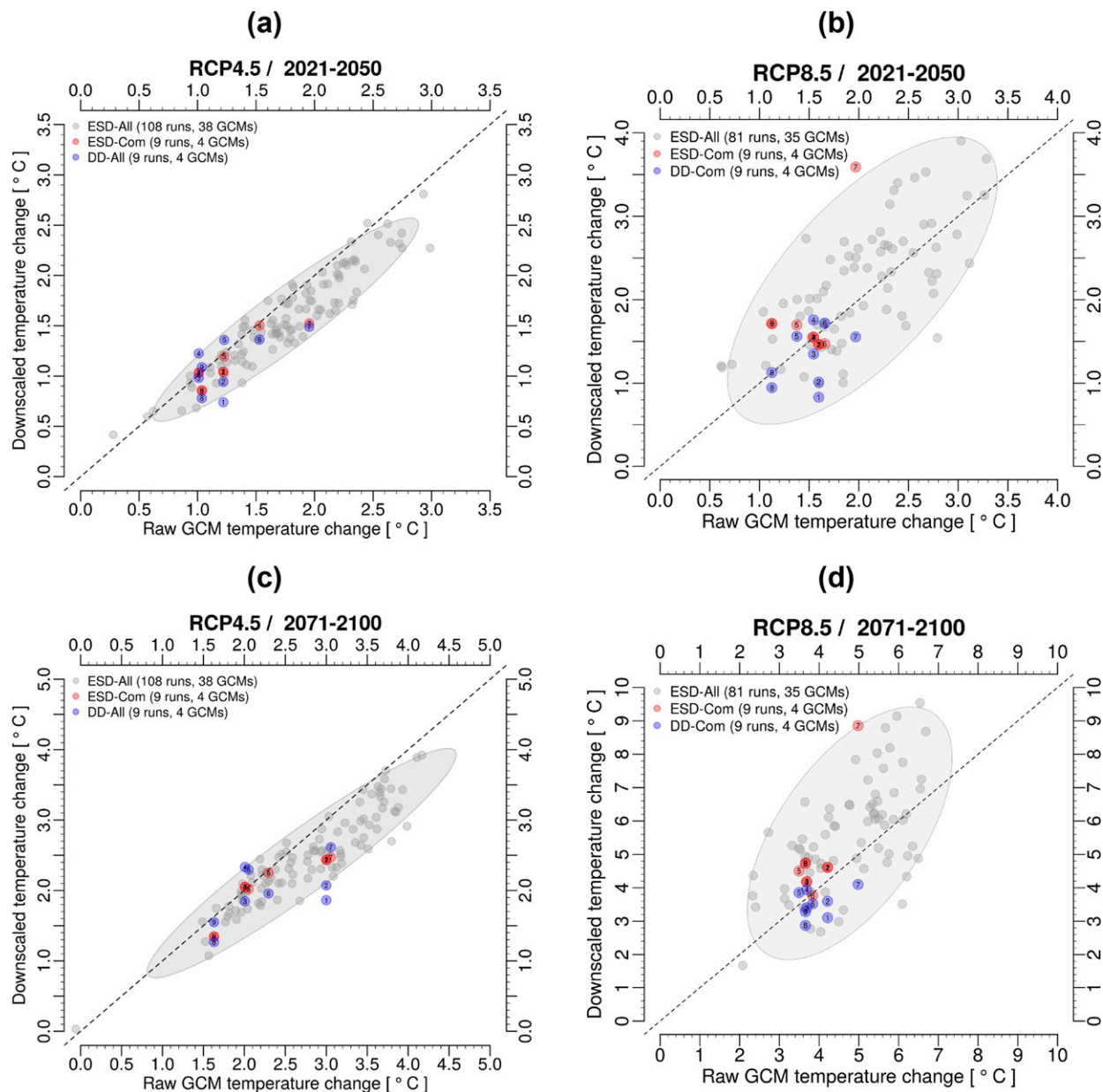
FIG. 6. Scatterplots of the changes in temperature as modeled by GCMs (*x* axis) and simulated by the DD (blue points) and the ESD (All: gray points and Com: red points) methods by the two horizons (a),(b) 2021–50 and (c),(d) 2071–2100 assuming the (left) intermediate (RCP4.5) and (right) high (RCP8.5) emission scenarios. The numbers inside the points refer to the simulation number as described in Table 2. The shaded area shows the 90% confidence zone derived from the ESD-All ensemble. Note that drawing a confidence zone for small ensembles is meaningless because the number of simulations in both DD-Com and ESD-Com is equal to or less than 9. The dashed line shows the line with slope 1 and intercept 0. Note also that the same coloring as in Fig. 2 has been used here.

also indicate inconsistencies between the subsets in simulating the sign and magnitude of precipitation changes toward the end of the twenty-first century. ESD-All ensemble mean of changes in precipitation is almost zero, whereas downscaling based on ESD indicates drier future conditions down to 8% when considering the common-GCM-based ensembles.

Similar to the results based on RCP4.5 emission scenario, the bias-corrected DD results are consistent with the raw GCM results and confirm increases in precipitation. This increasing rate is enhanced and can reach up to 25% by the end of the twenty-first century (highest values are simulated by simulation 7) when assuming the high emission scenario (RCP8.5). Although raw GCM
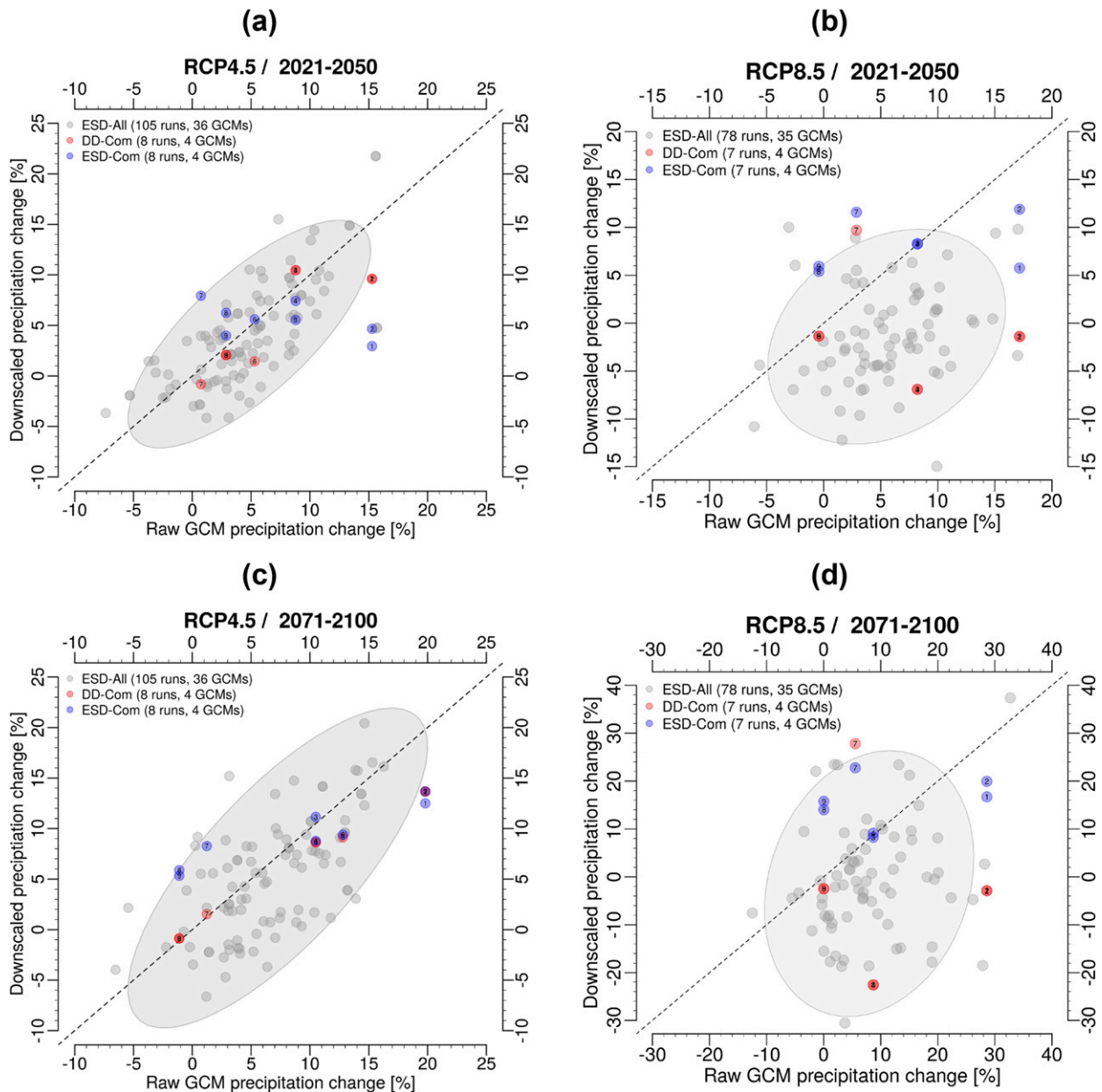
FIG. 7. As in Fig. 6, but for changes in precipitation.

results exhibit increases in annual mean of precipitation amount by about 10%, ESD results suggest altered precipitation change signal, showing in some cases a decreasing rate of precipitation by almost 25% (simulations 3 and 4). One explanation could be that modeled precipitation by the GCM, which is used as a predictor in the ESD framework, does not seem to carry enough information about the climate change signal and including other predictors would be necessary.

Of interest is that the subset of GCMs preserves both the mean and spread of modeled precipitation (i.e., raw

GCM without involving downscaling) reasonably well (Fig. 3) but that the two downscaling strategies show considerable disagreement, especially for the high emission scenario. The intermodel variations in the ESD projections tend to follow the large-scale climate signal; hence, they are more sensitive to the driving GCM (see Figs. 6 and 7). The bias-corrected DD projections, on the other hand, produce a smaller range and stronger increase in precipitation than the corresponding ESD-Com results. This suggests that the bias-corrected RCMs are less influenced by the driving GCM, possibly related to their

tendency to develop their own internal climate system (Evans and Westra 2012). Although the bias-corrected DD projections tend to slightly overlap with the ESD projections for the near future (Fig. 3a), they show a stronger increase in precipitation by the end of the twenty-first century. The discrepancy between the two indicates that one or both of the methods are flawed, revealing large uncertainties in projecting precipitation over Poland. However, both the downscaled results (bias-corrected DD and ESD; Fig. 3) suggest that precipitation is likely to increase in future, although this trend is uncertain. This further supports the findings of Piniewski et al. (2017) who argued that changes in annual precipitation were not robust for Poland, even though the climate models agreed on an increase. The disagreement in precipitation change between the two downscaling methods becomes more evident for the high emission scenario RCP8.5 in both the sign and magnitude (Table 3).

The uncertainty originating from subsampling a selection of GCM simulations (i.e., range of the estimated values without including downscaling) tends to produce large differences when compared to the full CMIP5 GCM ensemble of climate simulations. For instance, given an ensemble size of $n = 5$ when sampling from different GCMs (Figs. 4 and 5), has led to differences in the 90% confidence interval by up to $\pm 1°C$ by the end of the twenty-first century when assuming the intermediate emission scenario RCP4.5, which is doubled ($\pm 2°C$) when the high emission scenario (RCP8.5) is considered. This suggests that the difference between various combinations of small subsets from the CMIP5 ensemble of opportunities (including or not including the internal variability) can be as large as the climate change signal, which is in line with Dwyer et al. (2012) and Yettella and England (2018).

The impact of the subsampling is also clearly depicted in Figs. 2 and 3, which compare projected annual mean temperature (Fig. 2) and precipitation (Fig. 3) over Poland derived from various datasets as defined in Table 1. The selection of a small number of "representative" GCMs seems to underestimate the projected temperature change by DD and ESD over Poland. Both the magnitude and the spread of the change are affected by the subsampling, suggesting that the subsample in the CORDEX simulations does not cover the full range of possibilities in the CMIP5 ensemble. Even without involving any of the downscaling methods, the projected temperature tends to be systematically underestimated by about 0.5°C toward the end of the twenty-first century and covers approximately one-half of the full range spread of all available GCMs.

Although the RCM community is partly aware of the need for increasing the number of GCMs used in regional downscaling, as stressed by Zubler et al. (2016), more work remains to be done. For instance, in the Producing Regional Climate Projections Leading to European Services (PRINCIPLES) project (C3S_34b Lot2), although the existing ensemble of regional climate simulations for Europe will be extended to 10 GCMs, these efforts are still limited to a very small proportion of CMIP5 (10 of 40 GCMs) or CMIP6 global climate simulations and hence lead to a misrepresentation of the spread of the "effective" climate change signal on regional and local scales.

6. Conclusions

We made use of empirical–statistical methods and demonstrated that selecting a subset of GCMs to run RCMs, as in the CORDEX experiment, may lead to under/over estimation of the magnitude and/or spread of the climate change signal over Poland. For instance, the effect of the subselection is clear for temperature (Figs. 2 and 4) as indicated by an underestimation by about 0.5°C in the mean and by 1°C in the spread of temperature change assuming the intermediate RCP4.5 emission scenario, a result that is doubled toward the end of the twenty-first century and assuming the high RCP8.5 emission scenario.

Model selection can be done in a number of ways (e.g., (Herger et al. 2017; McSweeney et al. 2015; Knutti et al. 2017). The subset of models is often influenced by subjective choices such as the metrics and methods of evaluation as well as regions and variables of interest. We demonstrated that a small ensemble size, as is typically the case for most of the dynamically downscaled ensembles (e.g., CORDEX simulations) as compared with ESD, may not fully represent the uncertainty in the CMIP5 ensemble of possible climate change outlooks over Poland. There are substantial differences in the climate change signal among climate model simulations, and a random or even more sophisticated selection method of few GCMs may lead to a biased warming rate and strong biases in precipitation changes. This was demonstrated by the bootstrapping tests, which highlight the need for using large multimodel ensembles of climate model simulations.

For a better adaptation to climate change over Poland, we recommend using additional but not only representative GCM simulations such as EURO-CORDEX simulations when driving the boundary lateral conditions in the RCMs to ensure a better representation of the spread of the climate variability on regional and/or local scales.
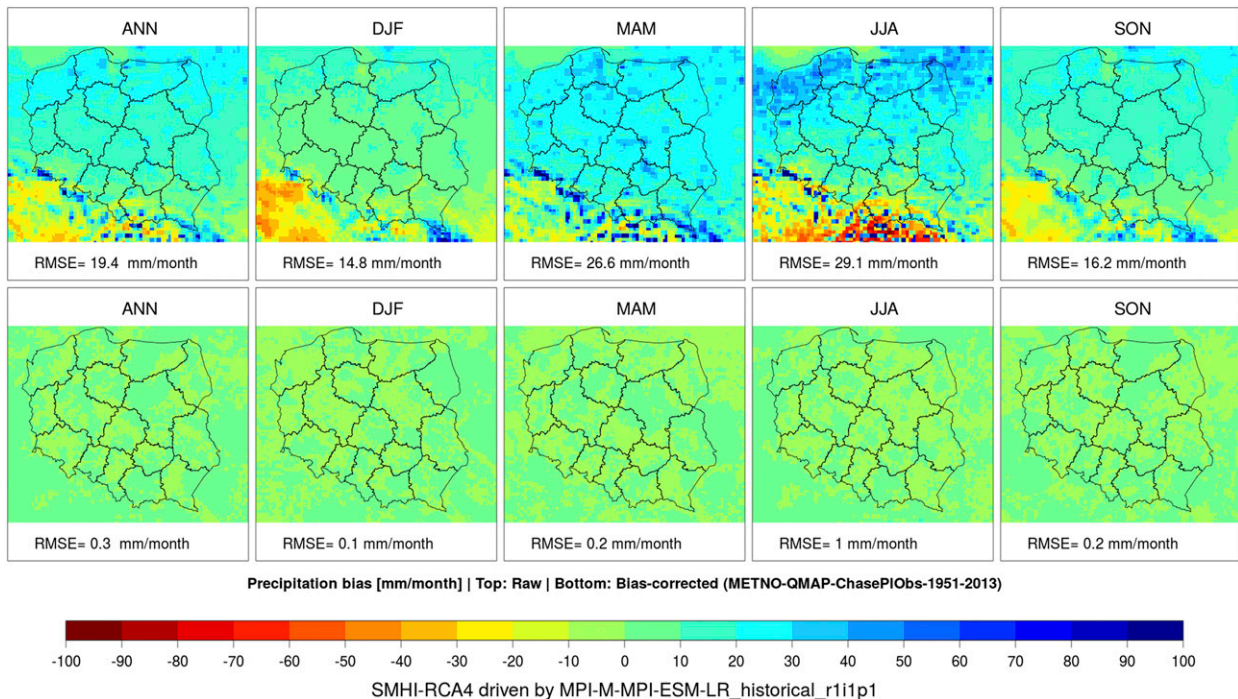
FIG. A1. Example of an evaluation of the DD method showing simulated precipitation by the SMHI-RCA4 RCM driven by the MPI-ESM-LR GCM. The maps show the root-mean-square errors (RMSE; mm month^{-1}) derived from (top) raw and (bottom) corrected precipitation with respect to the gridded observational dataset GDPT5 for annual and all seasons. RMSE values displayed in the bottom of each map show the area average over all individual RMSEs (i.e., at grid cells). This figure was adopted from Mezghani et al. (2017).

and the World Climate Research Programme's Working Group on Regional Climate as well as the Working Group on Coupled Modelling, the former coordinating body of CORDEX and the responsible panel for CMIP5. We also thank the climate modelling groups (listed in Table 2 of this paper) for producing and making available their model output. We also acknowledge the Earth System Grid Federation infrastructure—an international effort led by the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, the European Network for Earth System Modelling, and other partners in the Global Organization for Earth System Science Portals (GO-ESSP). Coauthor M. Piniewski is additionally grateful for support from the Alexander von Humboldt Foundation and the Ministry of Science and Higher Education of the Republic of Poland. We also thank the three anonymous reviewers for their constructive comments that helped in improving the analyses and the quality of the paper.

APPENDIX A

Evaluation of the Downscaling Methods

To ensure that downscaled estimates are representative of the local climate response, it is important to evaluate whether the ESD and DD methods are performing as intended. One common quality control method is cross validation, in which downscaled values for the historical period are compared with independent subsamples taken from observations.

a. Bias-corrected DD method

For DD, the quality control of the RCMs typically focuses on the evaluation of near-surface temperature, precipitation and mean sea level pressure on the basis of gridded observational datasets, that is, reanalysis products. A detailed analysis of the performance of the RCMs involved in this study was presented in Kotlarski et al. (2014). They found that the RCMs are able to capture the basic features of the climate system over Europe, including Poland, with regional and seasonal mean biases generally below 1.5°C for temperature and ±40% for precipitation. Similarly, Mezghani et al. (2017) assessed the raw EURO-CORDEX regional climate simulations over Poland and found large biases in the multimodel ensemble mean of mean annual precipitation ($\approx \pm 30\%$) and temperature ($\approx -2°C$). The authors indicated also the existence of an overall cold bias for all seasons except summer, during which a positive bias was found in the eastern part of the region and the mountains in the south. They argued that
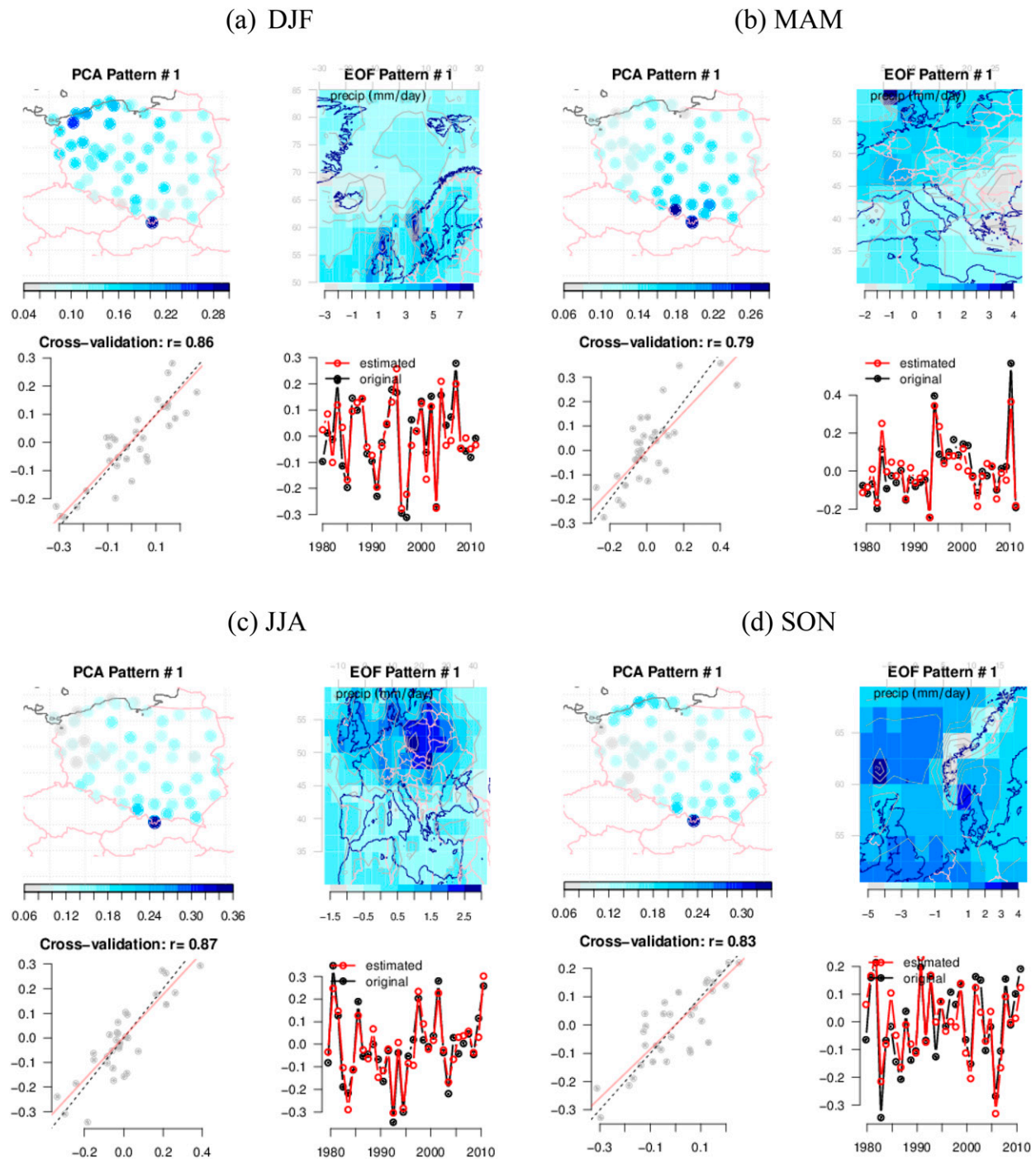
FIG. A2. Evaluation (by season) of the ESD method in reproducing precipitation characteristics found in the reference datasets, i.e., recorded precipitation at weather stations (top-left plot of each panel) and NCEP reanalysis (top-right plot of each panel). Also shown in each panel are the results of the cross-validation procedure (bottom-left plot) and how well the estimated values follow the reference data (bottom-right plot). In this case, the reference data are the first principal component.

all biases found in the raw data were considerably reduced (close to zero) in the bias-corrected datasets for both annual and seasonal values (Fig. A1). However, they underlined that there remained small biases in summer precipitation as an effect of the inclusion of the wet-day frequency in the adjustment procedure (less than 5%) and in temperature (less than 0.2°C) in the south enhanced by the Carpathian Mountains.

TABLE B1. Summary of CMIP5 GCM runs driven by the RCP4.5 emission scenario. The values in parentheses show the horizontal grid resolution of the climate model and the total number of individual runs provided by each institute. The last two columns show the total number of available simulations used for temperature T and precipitation P, respectively. Lines marked with an asterisk indicate the names of the GCMs (also marked with an asterisk) that have been used to produce regional climate simulations that are described in Table 2.

Institute	Model	No. runs	
		T	P
SF-DOE-NCAR	CESM1-CAM5 (0.9, 3), CESM1-BGC (0.9, 1)	4	4
CMCC	CMCC-CM (1.125, 1), CMCC-CMS (1.875, 1)	2	2
*CNRM-CERFACS	CNRM-CM5* (0.95)	1	1
CSIRO-QCCCE	CSIRO-Mk3.6.0 (1.875, 10)	10	10
*ICHEC	EC-EARTH* (1.125, 7)	7	6
LASG-CESS	FGOALS-g2 (2.8125, 1)	1	1
FIO	FIO-ESM (2.8, 3)	3	3
NOAA-GFDL	GFDL-CM3 (1.8, 1), GFDL-ESM2G (2.5, 1), GFDL-ESM2M (2.5, 1)	3	3
NASA-GISS	GISS-E2-H-CC (1, 1), GISS-E2-H (2, 15), GISS-E2-R (2, 17), GISS-E2-R-CC (1, 1)	34	18
MOHC	HadGEM2-CC (1.875, 1), HadGEM2-ES (1.875, 4)	5	5
NIMR-KMA	HadGEM2-AO (1.875, 1)	1	1
INM	INM-CM4 (2, 1)	1	1
*IPSL	IPSL-CM5A-LR (1.9, 4), IPSL-CM5A-MR* (1.25, 1), IPSL-CM5B-LR3 (1.25, 1)	6	7
MIROC	MIROC5 (1.4, 3), MIROC-ESM (2.8, 1), MIROC-ESM-CHEM (2.8, 1)	5	5
*MPI	MPI-ESM-LR* (1.875, 3), MPI-ESM-MR (1.875, 3)	6	12
MRI	MRI-CGCM3 (1.125, 1)	1	2
NCC	NorESM1-M (1.9, 1), NorESM1-ME (1.9, 1)	2	2
Total		108	105

b. Empirical–statistical downscaling method

For ESD, this was done by excluding a subsample (one fifth) of the observational data when the statistical model was calibrated and then using these data points to test the method. The cross validation was repeated (five times) until each part of the observational time series had been withheld from calibration and used for independent comparison. This procedure ensures a close relationship between the predictand and the predictor, that is, that selected large-scale climate patterns can be used to predict the local response. Before applying the ESD model to future projections, one should also make sure that the predictor is well represented by the GCMs. In our case this was achieved by using a common-EOF analysis to ensure similar spatial patterns in the NCEP reanalysis products and the CMIP5 GCM data as described in section 2. By examining the statistical quantities of the common-EOF, one can verify how similar the structures of the GCM and reanalysis data are (Benestad 1999). Further, we used data from the same observational sites in Poland as in Szwed et al. (2017), which consists of 43 weather stations recording daily precipitation and temperature and covering the period 1952–2013. The data were provided by the Institute of Meteorology and Water Management–National Research Institute (IMGW-PIB) and were used to train and evaluate the empirical–statistical downscaling regression models. The location of stations is displayed in

the top-left plot of each panel in Fig. A2. Furthermore, when the ESD is successful, the statistical properties of the downscaled ensemble, for example, the mean and standard deviation during the historical period, tend to match those from observations. If the ESD ensemble has a much smaller variability than observations or the multimodel ensemble mean is far off, this can inform us about a systematic failure of the procedure. An example of ESD evaluation is presented in Fig. A2, which shows the PCA pattern (top left of each panel), the EOF pattern (top right), the cross validation (bottom left), and the simulated-versus-original time series (bottom right) of precipitation characteristics across Poland. In this case, the cross-validation procedure indicated good performance for all seasons, with correlation coefficient values r that are higher than 0.79. This good performance was reflected by the good agreement between the estimated and original datasets (transformations based on the observations and/ or reanalysis datasets). For temperature, the ESD performance was even better, with cross-validation correlation values of higher than 0.9 for all seasons (not shown).

APPENDIX B

CMIP5 Ensemble of Climate Model Simulations

Details about the distribution of the CMIP5 simulations for each running institution and climate variable of

interest are presented in Table B1. This table also gives the total number of available simulations used for temperature T and precipitation P assuming the intermediate (RCP4.5) and high (RCP8.5) emission scenarios.

REFERENCES

Benestad, R. E., 1999: Evaluation of the common EOF approach in linear empirical downscaling of future ECHAM4/OPYC3 GSDIO climate scenarios. DNMI/KLIMA Rep. 35/99, 23 pp.

——, 2001: A comparison between two empirical downscaling strategies. *Int. J. Climatol.*, **21**, 1645–1668, https://doi.org/10.1002/joc.703.

——, 2002: Empirically downscaled multimodel ensemble temperature and precipitation scenarios for Norway. *J. Climate*, **15**, 3008–3027, https://doi.org/10.1175/1520-0442(2002)015<3008:EDMETA>2.0.CO;2.

——, 2008: Downscaled regional Norwegian temperature and precipitation series: Analysis for Statnett and CES. Norwegian Meteorological Institute Rep. 7, 67 pp., https://www.met.no/publikasjoner/met-report/met-report-2008/_/attachment/download/32c25aa1-a96b-4569-bf31-ae3806b20cbf:e78eb51010a9d8ac81ea8a2d489d4765b7b67400/MET-report-07-2008.pdf.

——, I. Hanssen-Bauer, and D. Chen, 2008: *Empirical–Statistical Downscaling*. World Scientific, 228 pp., https://doi.org/10.1142/6908.

——, D. Chen, A. Mezghani, L. Fan, and K. Parding, 2015: On using principal components to represent stations in empirical–statistical downscaling. *Tellus*, **67A**, 28326, https://doi.org/10.3402/tellusa.v67.28326.

——, K. M. Parding, K. Isaksen, and A. Mezghani, 2016: Climate change and projections for the Barents region: What is expected to change and what will stay the same? *Environ. Res. Lett.*, **11**, 054017, https://doi.org/10.1088/1748-9326/11/5/054017.

Berezowski, T., M. Szcześniak, I. Kardel, R. Michalowski, T. Okruszko, A. Mezghani, and M. Piniewski, 2016: CPLFD-GDPT5: High-resolution gridded daily precipitation and temperature data set for two largest Polish river basins. *Earth Syst. Sci. Data*, **8**, 127–139, https://doi.org/10.5194/essd-8-127-2016.

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

Busuioc, A., D. L. Chen, and C. Hellstrom, 2001: Performance of statistical downscaling models in GCM validation and regional climate change estimates: Application for Swedish precipitation. *Int. J. Climatol.*, **21**, 557–578, https://doi.org/10.1002/joc.624.

Chen, D., C. Achberger, J. Räisänen, and C. Hellström, 2006: Using statistical downscaling to quantify the GCM-related uncertainty in regional climate change scenarios: A case study of Swedish precipitation. *Adv. Atmos. Sci.*, **23**, 54–60, https://doi.org/10.1007/s00376-006-0006-5.

Chen, H., C.-Y. Xu, and S. Guo, 2012: Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *J. Hydrol.*, **434**, 36–45, https://doi.org/10.1016/j.jhydrol.2012.02.040.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, https://doi.org/10.1038/nclimate1562.

Dwyer, J. G., M. Biasutti, and A. H. Sobel, 2012: Projected changes in the seasonal cycle of surface temperature. *J. Climate*, **25**, 6359–6374, https://doi.org/10.1175/JCLI-D-11-00741.1.

Evans, J. P., and S. Westra, 2012: Investigating the mechanisms of diurnal rainfall variability using a regional climate model. *J. Climate*, **25**, 7232–7247, https://doi.org/10.1175/JCLI-D-11-00616.1.

Field, C. B., and Coauthors, Eds., 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge University Press, 582 pp.

Giorgi, F., and P. Lionello, 2008: Climate change projections for the Mediterranean region. *Global Planet. Change*, **63**, 90–104, https://doi.org/10.1016/j.gloplacha.2007.09.005.

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

Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess, 2006: Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios. *Int. J. Climatol.*, **26**, 1397–1415, https://doi.org/10.1002/joc.1318.

Hellström, C., D. Chen, C. Achberger, and J. Räisänen, 2001: Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Res.*, **19**, 45–55, https://doi.org/10.3354/cr019045.

Herger, N., G. Abramowitz, R. Knutti, O. Angélil, K. Lehmann, and B. M. Sanderson, 2017: Selecting a climate model subset ptimizeise key ensemble properties. *Earth Syst. Dyn.*, **9**, 135–151, https://doi.org/10.5194/esd-9-135-2018.

Huth, R., 2002: Statistical downscaling of daily temperature in central Europe. *J. Climate*, **15**, 1731–1742, https://doi.org/10.1175/1520-0442(2002)015<1731:SDODTI>2.0.CO;2.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

Jacob, D., and Coauthors, 2014: EURO-CORDEX: New high-resolution climate change projections for European impact research. *Reg. Environ. Change*, **14**, 563–578, https://doi.org/10.1007/s10113-013-0499-2.

Knutti, R., J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1918, https://doi.org/10.1002/2016GL072012.

Kotlarski, S., and Coauthors, 2014: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev.*, **7**, 1297–1333, https://doi.org/10.5194/gmd-7-1297-2014.

Linderson, M.-L., C. Achberger, and D. Chen, 2004: Statistical downscaling and scenario construction of precipitation in Scania, southern Sweden. *Nord. Hydrol.*, **35**, 261–278, https://doi.org/10.2166/nh.2004.0019.

Lutz, A. F., H. W. ter Maat, H. Biemans, A. B. Shrestha, P. Wester, and W. W. Immerzeel, 2016: Selecting representative climate models for climate change impact studies: An advanced envelope-based selection approach. *Int. J. Climatol.*, **36**, 3988–4005, https://doi.org/10.1002/joc.4608.

Maraun, D., and M. Widmann, 2017: *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 360 pp., https://doi.org/10.1017/9781107588783.

——, and Coauthors, 2015: VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, **3**, 1–14, https://doi.org/10.1002/2014EF000259.

McSweeney, C. F., R. G. Jones, and B. B. B. Booth, 2012: Selecting ensemble members to provide regional climate change information. *J. Climate*, **25**, 7100–7121, https://doi.org/10.1175/JCLI-D-11-00526.1.

——, ——, R. W. Lee, and D. P. Rowell, 2015: Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dyn.*, **44**, 3237–3260, https://doi.org/10.1007/s00382-014-2418-8.

Mendlik, T., and A. Gobiet, 2013: Selecting representative climate simulations for impact studies using cluster analysis. *Geophysical Research Abstracts*, Vol. 15, Abstract EGU2013-8460, https://meetingorganizer.copernicus.org/EGU2013/EGU2013-8460.pdf.

——, and ——, 2016: Selecting climate simulations for impact studies based on multivariate patterns of climate change. *Climatic Change*, **135**, 381–393, https://doi.org/10.1007/s10584-015-1582-0.

Mezghani, A., and B. Hingray, 2009: A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin. *J. Hydrol.*, **377**, 245–260, https://doi.org/10.1016/j.jhydrol.2009.08.033.

——, A. Dobler, and J. E. Haugen, 2016: CHASE-PL Climate Projections: 5-km Gridded Daily Precipitation and Temperature Dataset (CPLCP-GDPT5). Norwegian Meteorological Institute, accessed April 2018, https://doi.org/10.4121/uuid:e940ec1a-71a0-449e-bbe3-29217f2ba31d.

——, ——, ——, R. E. Benestad, K. M. Parding, M. Piniewski, I. Kardel, and Z. W. Kundzewicz, 2017: CHASE-PL Climate Projection dataset over Poland—Bias adjustment of EURO-CORDEX simulations. *Earth Syst. Sci. Data*, **9**, 905–925, https://doi.org/10.5194/essd-9-905-2017.

Overland, J. E., M. Wang, N. A. Bond, J. E. Walsh, V. M. Kattsov, and W. L. Chapman, 2011: Considerations in the selection of GCMs for regional climate projections: The Arctic as a case study. *J. Climate*, **24**, 1583–1597, https://doi.org/10.1175/2010JCLI3462.1.

Piani, C., G. P. Weedon, M. Best, S. M. Gomes, P. Viterbo, S. Hagemann, and J. O. Haerter, 2010: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, **395**, 199–215, https://doi.org/10.1016/j.jhydrol.2010.10.024.

Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting GCMs for regional climate change studies. *Proc. Natl. Acad. Sci. USA*, **106**, 8441–8446, https://doi.org/10.1073/pnas.0900094106.

Piniewski, M., A. Mezghani, M. Szcześniak, and Z. W. Kundzewicz, 2017: Regional projections of temperature and precipitation changes: Robustness and uncertainty aspects. *Meteor. Z.*, **26**, 223–234, https://doi.org/10.1127/metz/2017/0813.

Schmidli, J., C. M. Goodess, C. Frei, M. R. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schith, 2007: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *J. Geophys. Res.*, **112**, D04105, https://doi.org/10.1029/2005JD007026.

Szwed, M., I. Pińskwar, Z. W. Kundzewicz, D. Graczyk, and A. Mezghani, 2017: Changes of snow cover in Poland. *Acta Geophys.*, **65**, 65–76, https://doi.org/10.1007/s11600-017-0007-z.

Takayabu, I., H. Kanamaru, K. Dairaku, R. Benestad, H. von Storch, and J. H. Christensen, 2016: Reconsidering the quality and utility of downscaling. *J. Meteor. Soc. Japan*, **94A**, 31–45, https://doi.org/10.2151/jmsj.2015-042.

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

Vano, J. A., J. B. Kim, D. E. Rupp, and P. W. Mote, 2015: Selecting climate change scenarios using impact-relevant sensitivities. *Geophys. Res. Lett.*, **42**, 5516–5525, https://doi.org/10.1002/2015GL063208.

WCRP, 2011: Coupled Model Intercomparison Project—Phase 5. *CLIVAR Exchanges*, No. 56, Vol. 16 (2), International CLIVAR Project Office, Southampton, United Kingdom, 52 pp., http://www.clivar.org/sites/default/files/documents/Exchanges56.pdf.

Wilby, R. L., 1998: Statistical downscaling of daily precipitation using daily airflow and seasonal teleconnection indices. *Climate Res.*, **10**, 163–178.

——, T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.*, **34**, 2995–3008, https://doi.org/10.1029/98WR02577.

Wilcke, R. A. I., and L. Bärring, 2016: Selecting regional climate scenarios for impact modelling studies. *Environ. Modell. Software*, **78**, 191–201, https://doi.org/10.1016/j.envsoft.2016.01.002.

Yettella, V., and M. R. England, 2018: The role of internal variability in twenty-first-century projections of the seasonal cycle of Northern Hemisphere surface temperature. *J. Geophys. Res.*, **123**, 13 149–13 167, https://doi.org/10.1029/2018JD029066.

Zubler, E. M., A. M. Fischer, F. Fröb, and M. A. Liniger, 2016: Climate change signals of CMIP5 general circulation models over the Alps—Impact of model selection. *Int. J. Climatol.*, **36**, 3088–3104, https://doi.org/10.1002/joc.4538.