ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Synthesis on the existing simulations at M43

2017/C3S_34b_Lot2_SMHI

Issued by: SMHI / Erik Kjellström

Date: 15/10/2021

Ref: C3S_34b_Lot2.4.3.4_202110_Synthesis_Report_v2

Official reference number service contract:

2017/C3S_34b_Lot2_SMHI/SC4

Contributors

CNRS
Robert Vautard (WP4 lead)

DMI
Fredrik Boberg
Ole B. Christensen

ETHZ
Silje L. Soerland
Marie-Estelle Demory
Jan Rajczac
Christoph Schär

KNMI
Emma Aalbers
Geert Lenderink
Erik van Meijgaard

GERICS
Katharina Bülow
Claas Teichmann

ICTP
James Ciarlo
Erika Coppola

MÉTEO-FRANCE
Lola Corre
Samuel Somot

MOHC
Erasmo Buonomo

SMHI
Erik Kjellström
Grigory Nikulin

Table of Contents

Introduction

The goals of the WP4 of C3S_34b_Lot210 are to (i) monitor the advancement by regular verifications of the progressive fulfillment of the spread requirements defined in C3S_34b_Lot210; (ii) perform, over a few key metrics the evaluation of existing and new simulations allowing a scientific “health check” of all simulations as they are produced, and (iii) provide regular syntheses and demonstrations.

In this report we review the state of the EURO-CORDEX ensemble as of May 2021 focusing on the RCP8.5 scenario as this is the one with most members (75), thereby spanning the largest spread of the three available scenarios (also RCP4.5 and RCP2.6). We first describe the status of the EURO-CORDEX GCM-RCM matrix and we perform an assessment of the simulations. The ensemble includes both simulations that were produced outside of the project and simulations produced in the project. The focus here is on reviewing the full EURO-CORDEX RCP8.5 GCM-RCM matrix available for C3S users. In addition, we also illustrate how the additional 48 RCP8.5-simulations produced within the project add value to the full 75-member GCM-RCM matrix. All simulations are now published on the Earth System Grid Federation.

For the assessment, a set of diagnostics and analyses is proposed. The set consists in a coordinated analysis of current biases and future trends over a number of indices in the form of maps and tables of numerical values for each PRUDENCE region. All these results are developed in two main articles (Vautard et al., 2021; Coppola et al., 2021), and we provide here only a short summary of these results. Other specific analyses are mentioned in a separate section. The Coppola et al study includes a first comparison between the EURO-CORDEX GCM-RCM matrix and a set of global models from the last generation global model intercomparison project CMIP6. For a more extensive set of CMIP6 models and how they compare to the EURO-CORDEX RCMs and their driving CMIP5 GCMs we refer to C3S_34b_Lot2.3.4.2 "Differences between CMIP5 GCMs downscaled by EURO-CORDEX RCMs and the full CMIP5 and CMIP6 ensembles, with focus on Europe".

After the key findings (Section 1), the coordinated evaluation design is described in Section 2 together with the model ensemble and the set of diagnostics. In Section 3 we present examples of the main results. Specific analyses are provided in Section 4.

1. Key findings

1.1 Concerning model biases

1. As of 1 May 2021, 75 EURO-CORDEX climate simulations, including those produced in this project were available (vs. 55 in November 2019). Eight global climate models (GCMs) downscaled by twelve regional climate models (RCMs), for the historical period and projections for the RCP8.5 scenario. Among the 75, 48 were carried out on C3S funding. The climate projections allow a better understanding of biases and drivers of climate change;

2. No GCM-RCM simulation exhibit outlying temperature biases, despite some systematic biases across regions for a few models; GCMs and RCMs generally have a cold bias in North and Western Europe and a warm bias in South-Eastern Europe; This trend is similar for means and hot temperatures; biases in frost days have varying patterns;

3. RCMs do not improve GCM temperature biases; Maximum and mean temperature bias patterns depend mostly on the GCM, while minimum temperature bias patterns are rather driven by RCMs;

4. RCMs exhibit a systematic wet bias as compared to E-OBS observations; the biases appear to originate from both RCMs and GCMs; RCMs generally overestimate heavy precipitations as observed in E-OBS, especially in winter;

5. Wet RCM biases are present in the water balance variables evaporation, runoff and soil moisture. Such variables show changes generally depending on the RCM to the largest extent;

6. Models generally reproduce well observed dynamical patterns and surface winds; A slight systematic underestimation of mean sea level pressure over North-Western Europe and the North-East Atlantic; and a systematic overestimation of ERA5 reanalysis surface winds. These biases could not be explained here but could also result partly from biases in ERA5;

7. RCM sea level pressure patterns are generally well correlated with driving GCM patterns, with pattern correlations of about 0.9 in winter and lower ones in summer (when the driving large-scale flow is weaker and other drivers such as soil moisture can be important);

8. Model simulated surface radiation shows biases ranging from about -50W/m² to about +50 W/m². Mean negative biases (10-20 W/m²) affect the Iberian Peninsula, and mean positive biases of the same amplitude affect Scandinavia. More analyses are needed to understand these biases affected by clouds and aerosols uncertainties;

9. The biases for a number of sectoral impact-indices have been analyzed, including extremes such as extreme heat or heat stress. Extreme heat (resp. cold) indices generally show negative (resp. neutral to positive) biases; extreme precipitation and wind biases are positive; Drought index biases are more balanced; most biases of extreme indices are driven by the RCMs, except for drought;

10. The addition of the C3S-funded simulations on top of already existing Euro-CORDEX simulations with the matrix-filling technique improves the ensemble mean for some variables and seasons, modifies projected trends, and clearly provides material for more robust statistics to be used in particular for analyzing climate extremes (e.g. for event attribution). The additional simulations produced in the project lead to an improvement compared to observations for some of the variables and a spread in the biases more centered around zero.

1.2 Concerning trends

1. A combined analysis of GCM and RCM projections was conducted for Europe; projections are available for two climate scenarios, the RCP2.6 and RCP8.5 from 11 RCMs driven by 8 GCMs. The RCM results are compared with the driving CMIP5 GCMs but also with a subset of GCMs from recent CMIP6 for the first time (Coppola et al., 2021). This analysis has allowed to feed the IPCC AR6 report on regional climate information;

2. Warming is largest for all 3 ensembles over Northern Europe in winter, associated with maximum precipitation increase, and maximum over Mediterranean and Southern European regions in summer, associated with maximum precipitation decrease;

3. CMIP6 projections indicate highest values of warming and highest precipitation changes while the EURO-CORDEX warms the least and has lower precipitation changes but with increased spatial details on complex topography, along the coasts and islands and a more accurate land-sea contrast;

4. An increase of heat extremes and decrease of cold extremes with analogous large-scale behavior for the 3 ensembles are found, with a similar order in change magnitude as for temperature;

5. RCMs indicate important change details relative to that in the GCMs such as an increase of the number of high heat stress thresholds (e.g. WBGT>35°C) in low-lying coastal areas of Southern Europe by the end of the century;

2. Model ensemble and diagnostics

Beyond the results of Vautard et al. (2020) and Coppola et al. (2020), the ensemble of (on ESGF) simulations available on 1 May 2021 is considered here, including different members of an ensemble using the same models (GCM, RCM), and different versions of the same model (for REMO). Two WRF versions were used but were considered as two different models due to a number of differences in parameterizations and implementation. In total 12 RCMs and 9 GCMs were used with 3 additional GCM members, and 75 GCM-RCM simulations are currently available. Table 1 lists the

simulations that were analyzed. Radiation is not available for all models or in some cases, for recently finalized simulations, not considered.

RCM-GCM	CNRM r1	ECEARTH r12, r1, r3	HADGEM r1	MPI r1, r2, r3	NORESM r1	IPSL r1	CANESM r1	MIROC r1
CCLM		r12	r1	r1			4	4
HIRHAM		r12,r1,r3		r1				
RACMO		r12,r1,r3		r1				
RCA		r12,r1,r3		r1,r2,r3				
REMO		r12		r1 2 ,r2 2 ,r3				
WRF361H		r12 4,5,6,7	4,5,6,7,8	r1 4,5,6,7				
WRF381P	9	r12	9	r1	9	9		
ALADIN53	3,4							
ALADIN63								
REGCM		r12	1	r1 1				
COSMO-crCLIM		r12,r1,r3		r1,r2,r3				
HADREM		r12		r1				

Table 1: Simulations analyzed in this study, which cover both historical and RCP8.5 periods (grey and purple). Grey cells indicate simulations that were funded by the C3S (48) while purple cells indicate simulations that were not funded by the C3S. The name of GCM simulations is given as ”GCM (rR)” where ”R” is the model ensemble realization. Footnotes: [1]: This simulation does not have sea-level pressure available; [2] This REMO version is REMO2009 while other versions are REMO2015; [3]: This simulation does not have daily surface wind available; [4] This simulation does not have daily maximum surface wind available; [5] This simulation does not have daily sea level pressure fields; [6] this simulation does not have surface solar downward radiation (rsds) field analyzed; [7] this simulation does not have WBGT index analyzed; [8] this simulation does not have total runoff; [9] this simulation does not have evaporation or total soil moisture content.

Diagnostics

In the evaluation presented in Vautard et al. (2021), and in the study presenting the projections (Coppola et al., 2021), a number of variables and indices are analyzed. The main variables (precipitation, temperature, wind, sea level pressure and

radiation) were analyzed, with seasonal decomposition for temperature and precipitation only. Other indices include extreme indices and impact-oriented indices, for a few sectors. For instance, the Heating Degree Day index and Cooling Degree Day indices are indicative for energy demand, while the Growing Degree Day and the Length of Frost-Free Period are indicative for agriculture. The simplified Wet Bulb Globe Temperature as well as the number of days above 35°C are indicative for health and labor productivity.

To keep the document short and the workload reasonable, we only use a selection of variables and indices in this synthesis report for the main analysis. We focus on 3 main variables (precipitation, temperature and radiation) and 8 indices with the full ensemble (Table 2). The ensemble results are, however, qualitatively similar to those of Vautard et al. (2021) and Coppola et al. (2021) despite the larger ensemble, and main results for the other variables and indices can be found there.

We present here briefly the main results for the biases of the ensemble and each model separately, and the changes between a reference period in the recent past and the mid-century. The choice to focus on the mid-century as a preferred period for analysis in the synthesis reports was made early in the project in dialogue with the ECMWF. This choice is partly also guided by the need for many stakeholders to have information for the mid-century or before. However, results can also easily be made available for the end of the century.

Time periods

For the bias evaluation analysis, we considered a single reference climate period which corresponds to the WMO period definition (1981-2010); since in CMIP5/CORDEX the historical period stops in 2005, we completed this period with RCP8.5 simulations until 2010. The choice of the RCP8.5 scenario for these five years (2006-2010) instead of any other scenario (RCP4.5 or 2.6) was pragmatic, as most simulations are available there and as these are the ones reported upon here. We note that a different scenario may give slightly different results. However, as differences between forcing in these scenarios in the first years are small, differences in the resulting climate would essentially result from natural chaotic variability and be small and are not taken into consideration here.

Time periods considered for the projection analysis are (i) the reference period as above, (ii) a mid-century period (2036-2065) in this report. A far-future period (2071-2100) was also considered in Coppola et al. (2021).

Changes are calculated as differences in the climate over the two periods.

Variable	ECV involved CF variable name	Observation data set	Season	Remark
daily mean temperature	Tas	E-OBS23 025	Year, DJF and JJA	Haylock et al. (2008)
daily max temperature	tasmax	E-OBS23 025	Year	
daily min temperature	tasmin	E-OBS23 025	Year	
daily precipitation amount	Pr	E-OBS23 025	Year, DJF and JJA	
surface solar radiation	rsds	HELIOSAT	Year	
Index	ECV involved CF variable name	Observation dataset used for assessment	Link to sector	Category of extreme or impact-oriented index
TXx	tasmax	E-OBS23 025	All	Warm
#days/year TX>35°C	tasmax	E-OBS23 025	agriculture, health	Warm
Length of frost-free period	tasmin	E-OBS23 025	agriculture, ecosystems	Warm
Growing degree days > 5°C	Tas	E-OBS23 025	agriculture, ecosystems	Warm
Cooling degree day > 22°C	tas, tasmin, tasmax	E-OBS23 025	energy	Warm
TNn (yearly min temperature)	tasmin	E-OBS23 025	All	Cold
Heating degree day < 15.5°C	tas, tasmin, tasmax	E-OBS23 025	energy	Cold
RX1day RX1d	Pr	E-OBS23 025	Flood	Wet

Table 2: Variables and indices used in this study with reference datasets, evaluation period and concerned economic sectors for the indices.

3. Results

We show here examples of our main results, and the reader is referred to the articles by Vautard et al. (2021) and Coppola et al. (2021) for detailed results, despite that they were made with a subset of 55 of the 75 models used here. Increasing the matrix size provides a more satisfactory coverage of the uncertainty in changes (more models, more balanced coverage of GCMs and RCMs and therefore of any possible climate change signal). It gives more confidence in results. However, despite added models slightly changing the numbers in biases and trends, we verified that these changes are small, which is expected unless a strong imbalance was present, and which is reassuring for the user in the fact that results are not very sensitive to model selection, and therefore robust.

In particular, the funded simulations increased the matrix by adding 5 more RCMs, and a full submatrix. The added value of the 48 C3S-funded RCP8.5-simulations is also discussed with a few example diagnostics in Section 3.4.

3.1 Model biases

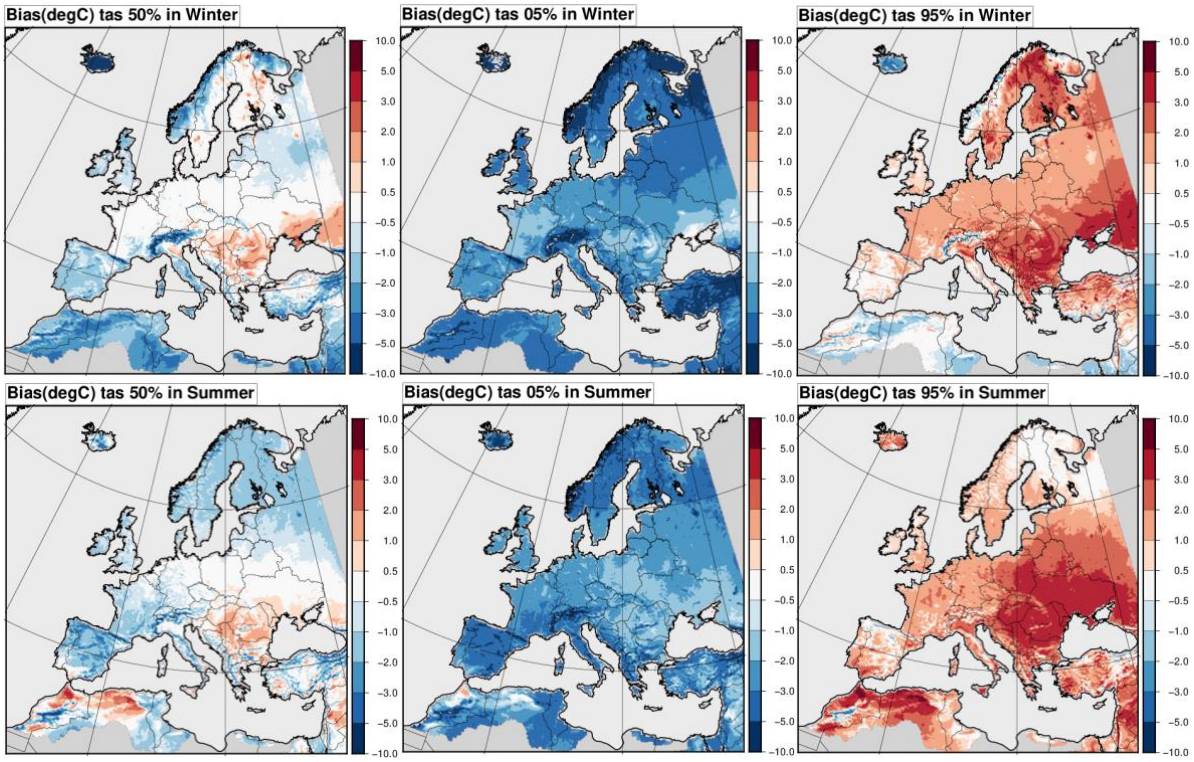


Figure 1: Median, 5% and 95% biases of the EURO-CORDEX ensemble for daily mean temperature (Tas). Top row: winter, bottom row: summer. Unit (°C).

Figure 1 shows ensemble median, lower 5% and upper 5% (that corresponds to the 95% percentile, here denoted as 95 % or p95) temperature biases in winter and in summer. Cold biases dominate in the median in large areas except in the Balkans where a slight positive bias is obtained as well as in parts of Fennoscandia in winter. However, a large band of small biases, smaller than 0.5°C in absolute value, can be seen over Central Europe, from Northern France to Ukraine. In Western Mediterranean areas, biases are negative and in the median reach about 1-2 degrees. In these areas, most models have a negative bias, as the p95 map shows near vanishing values. This suggests common issues in models (processes, resolution, forcing).

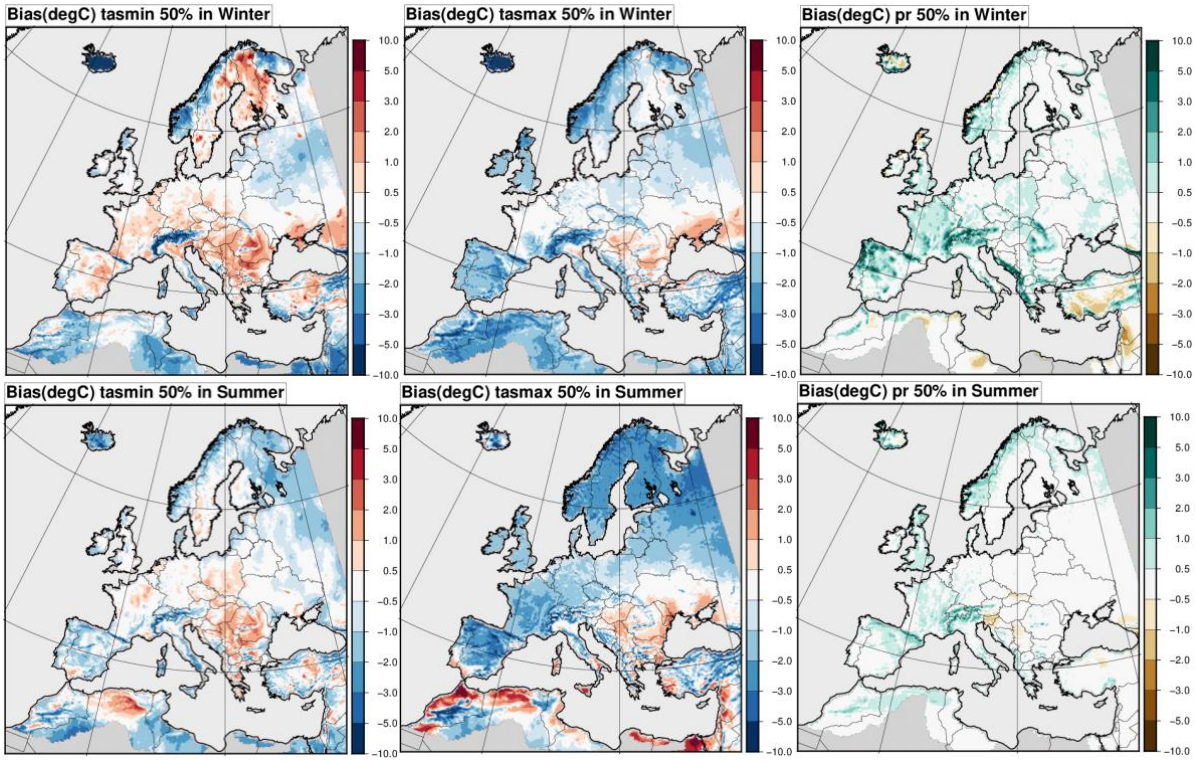


Figure 2: Median biases for daily minimum (tasmin), maximum temperatures (tasmax), left and middle columns, and mean precipitation amount (in mm/day), right column, for winter (top) and summer (bottom) seasons. Unit (°C).

Figure 2 shows the median bias for daily minimum and maximum temperatures as well as precipitation biases in winter and summer (without showing p05 and p95). Models underestimate maximum temperature, especially in summer where median biases exceed 1°C in most of western Europe. Models overestimate minimum temperature in many areas except in the Iberian Peninsula and Scandinavia in summer.

A positive bias is found for precipitation, especially in Southern Europe (Figure 2), as discussed in previous articles (e.g. Kotlarski et al., 2014). This bias is particularly high in the winter season, and more moderate in the summer season.

Area	West	East	South	North
1 (BI) British Isles	-10	2	50	59
2 (IP) Iberian Peninsula	-10	3	36	44
3 (FR) France	-5	5	44	50
4 (ME) Mid-Europe	2	16	48	55
5 (SC) Scandinavia	5	30	55	70
6 (AL) Alps	5	15	44	48
7 (MD) Mediterranean	3	25	36	44
8 (EA) Eastern Europe	16	30	44	55

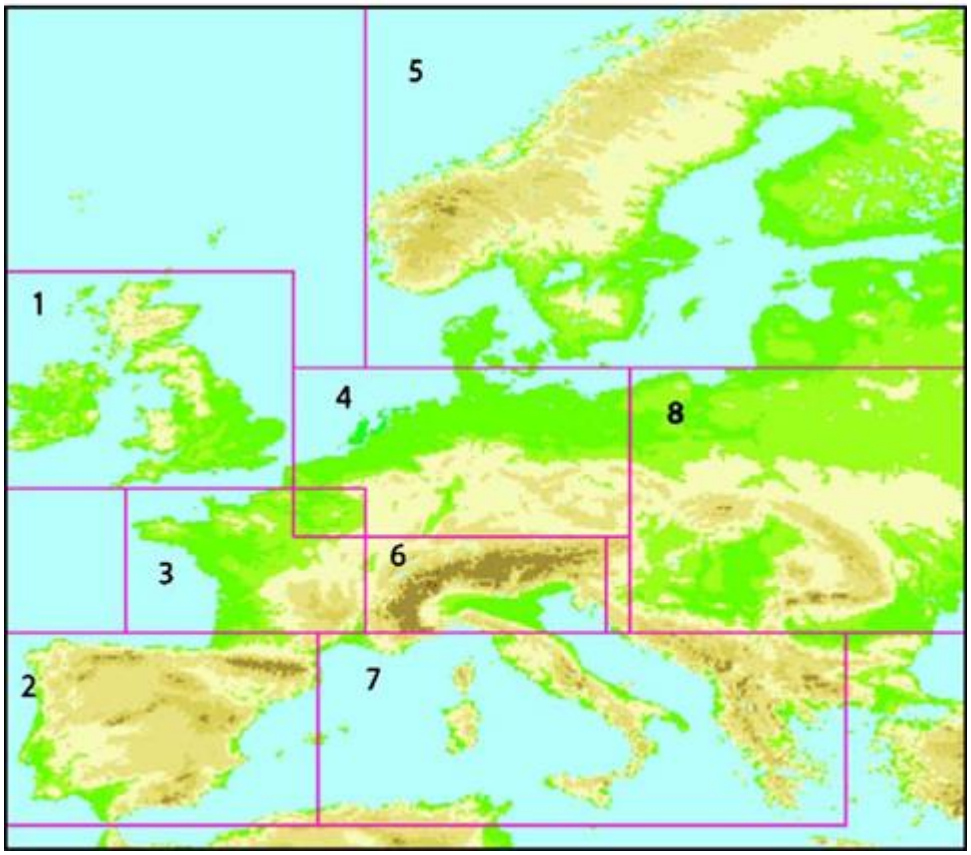


Figure 3: Prudence Regions and their extent in latitude and longitude.

To show individual model biases, we average grid-point biases over each of the PRUDENCE regions (see Figure 3 for definition). Each result is then plotted for mean precipitation, temperature (including min and max) and radiation in Figure 4. The influence of the GCM is clear in temperature and maximum temperature biases, while the RCM is the main driver for minimum temperatures (see also previous studies). For temperature and maximum temperature, one clearly distinguishes three

"cold GCMs": CNRM, EC-EARTH and IPSL. For minimum temperatures, "cold RCMs" are found (ALADIN, RACMO and RCA) with systematic negative biases while warm biases are not systematic.

For mean yearly precipitation, the distributions are rather centered around a 0.5 mm/day bias, but are skewed towards high biases which can reach 2 mm/day (about 700 mm/year). Such high biases are mostly found in the Alpine regions (AL) where wind can lead to severe undercatch of precipitation.

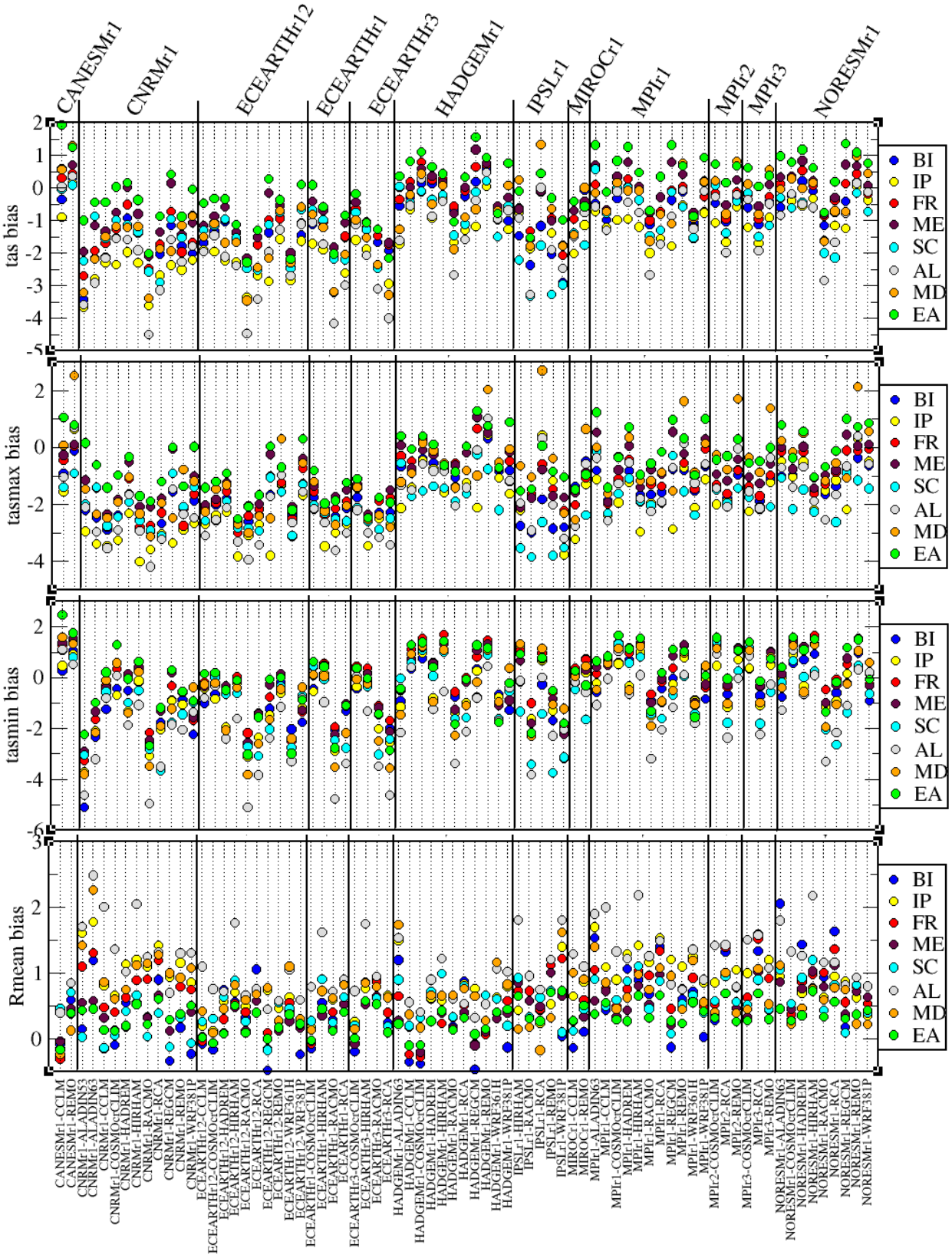


Figure 4: Biases of daily mean (Tas), maximum (tasmax), minimum (tasmin) temperature (°C) and precipitation (Pr, mm/day) (1981-2010) for each model and PRUDENCE region.

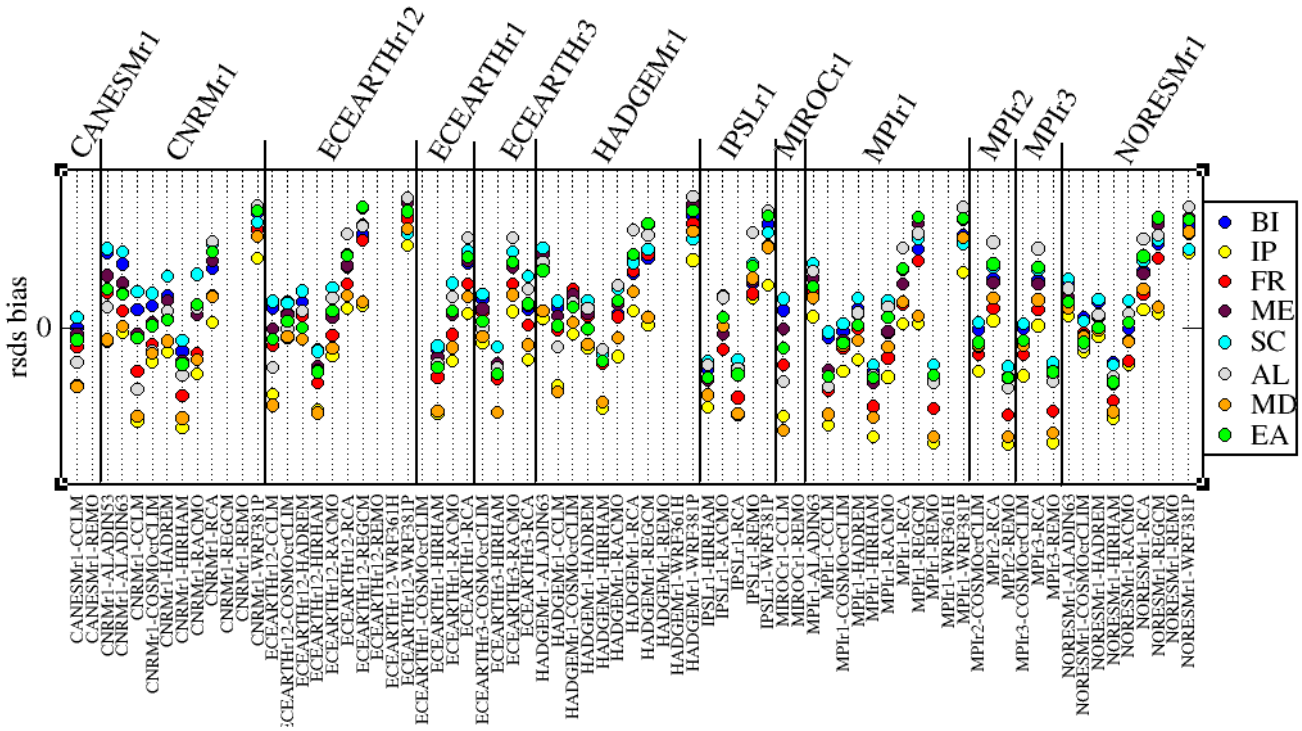


Figure 5: Biases of shortwave radiation at surface (rsds) (1981-2010) for each model and PRUDENCE region. Unit (W m^{-2}).

For radiation (Figure 5), the biases mostly depend on the RCM and its physical parameterizations. Some models have a systematically positive and high bias (e.g. WRF381P, REGCM, RCA) while others have a systematically negative bias (REMO, HIRHAM).

We then analysed the biases for the 8 indices (Figures 6 and 7). For the “warm indices” (eg. TXx, TX35, CDD, GDD) a clear dependence on GCM is found, with for instance warm biases for HADGEM and cold biases for EC-EARTH. Many models do not perform well for TXx in Scandinavia (too cold) and Eastern Europe (too hot), but the biases do not have systematic behavior in other regions. For very cold temperatures, biases can be very large (e.g. > 5°C), and either positive or negative. The effect of temperature biases can generate very different behavior for different cold indices. For instance, HDD has a general positive bias (too cold) while LFFP has a general neutral or positive bias (too warm). Hence the simple knowledge of temperature biases cannot allow conclusions as concerns impact indices. As for precipitation, RX1d has a systematic positive bias, with particularly high values in Mediterranean areas.

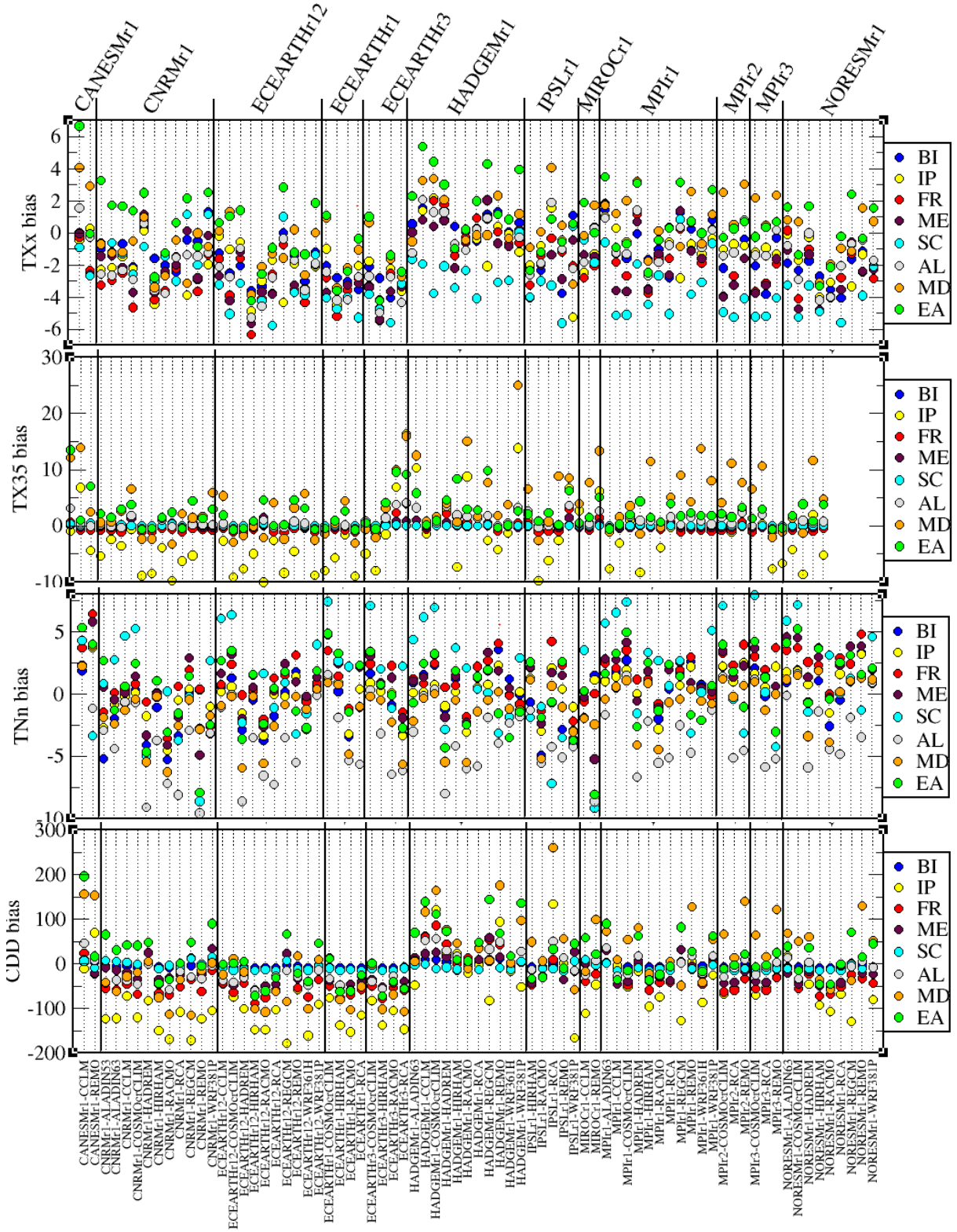


Figure 6: Biases for each model and PRUDENCE region for the 4 indices TXx (°C), TX35 (#day per year)), TNn (°C) and CDD (degree.day/year). Comparisons are made with calculations based on EOBS23 observations.

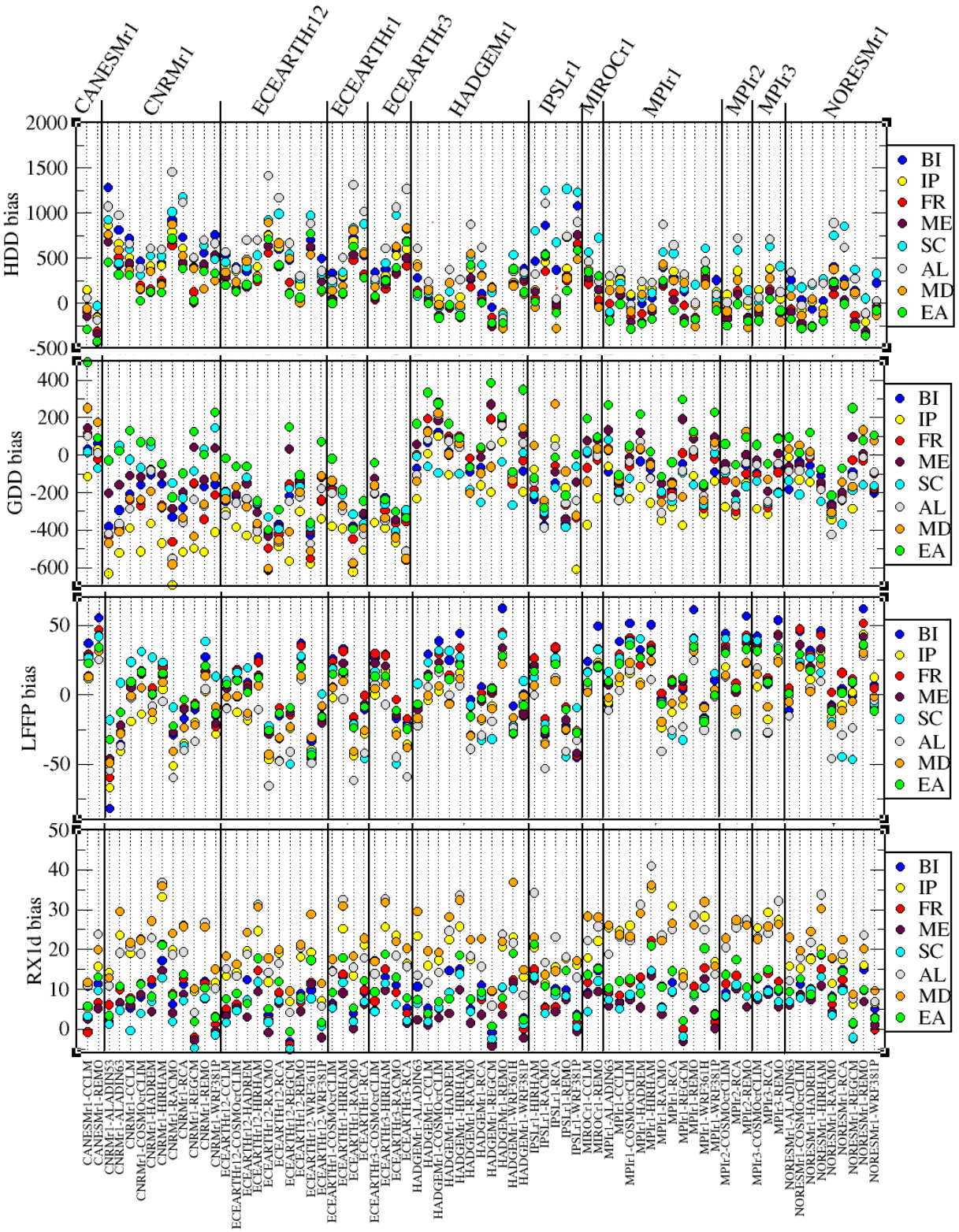


Figure 7: Same as Figure 6 for indices HDD, GDD (degree.day/year), LFFP (day/year) and RX1day (mm/day).

3.2 Future changes

Future changes were assessed in depth in the article of Coppola et al. (2021) using 55 models of the ensemble. They were compared with CMIP5 simulations as well as with the first CMIP6 simulations. This was the first model inter-comparison of this kind, which has been made possible thanks to the PRINCIPLES project. Here we show only a few examples of projected changes of the EURO-CORDEX ensemble. The comparison with GCMs is made in Coppola et al. (2021) and not repeated in any detail here. In section 3.4 we discuss the added value of the projections generated in the PRINCIPLES project and also give an example of the progression over time. A comparison of a larger set of CMIP6 GCMs with the EURO-CORDEX results including also its driving CMIP5 and the full CMIP5 ensemble is presented in C3S_34b_Lot2.3.4.2 "Differences between CMIP5 GCMs downscaled by EURO-CORDEX RCMs and the full CMIP5 and CMIP6 ensembles, with focus on Europe".

Temperature (and Tmin and Tmax) changes strongly depend on the GCM. Some GCMs show a particularly high sensitivity (e.g. CANESM, HADGEM, IPSL, MIROC), leading to changes generally between 1.5°C and 3°C in the different PRUDENCE regions. The other GCMs lead to changes mostly lying between 1°C and 2°C. Changes depend a lot on region, and models homogeneously catch this dependence with higher warming for Scandinavia and Mediterranean and lower warming for the British Isles. Highest changes are found over Scandinavia in daily minimum temperatures. In addition to the strong sensitivity to the GCMs, the results indicate that also RCMs can modify results strongly. For instance, there is about a 1°C difference in Tmin in Scandinavia between the two RCMs downscaling MIROC.

For precipitation, a less clear dependence on GCM is found. The dependence is regional, with more or less. clear decreases of precipitation in Mediterranean regions (IP and MD) and increases in Central and Northern Europe (SC, EA, ME).

Changes in indices are shown in Figures 9 and 10. For temperature extremes, changes are generally higher than for means. For TXx, warming until the mid-century can reach 4-5°C in some regions and for some models in FR, MD regions. TNn changes can also be higher, reaching 6-8°C for RCMs forced by some GCMs, especially in Scandinavia and Eastern Europe. Changes in the number of exceedances of 35°C have a strong model dependence and therefore large uncertainty, with changes ranging from 0 to about 20 days. The change of the length of frost-free period has a strong regional variation with about 20-40 days in Northern and Central Europe and weaker

values (less than 15 days) in the Mediterranean areas. Such regional dependence of change is also found in CDD, HDD and GDD.

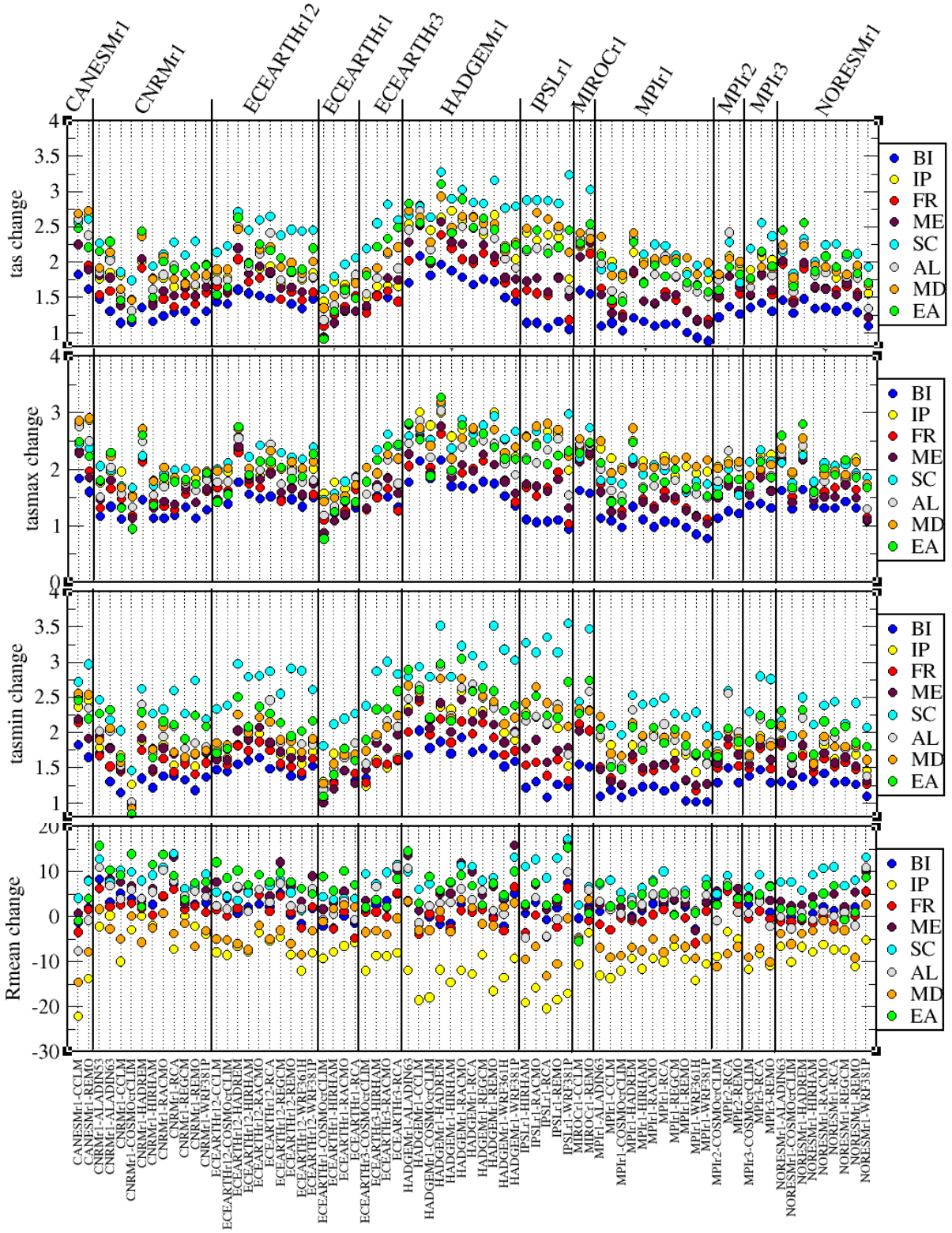


Figure 8: Changes in daily mean (tas), maximum (tasmax), minimum (tasmin) temperature (°C) and annual mean precipitation (Rmean, %) (2036-2065 – 1981-2010) in RCP8.5 for each model and PRUDENCE region.

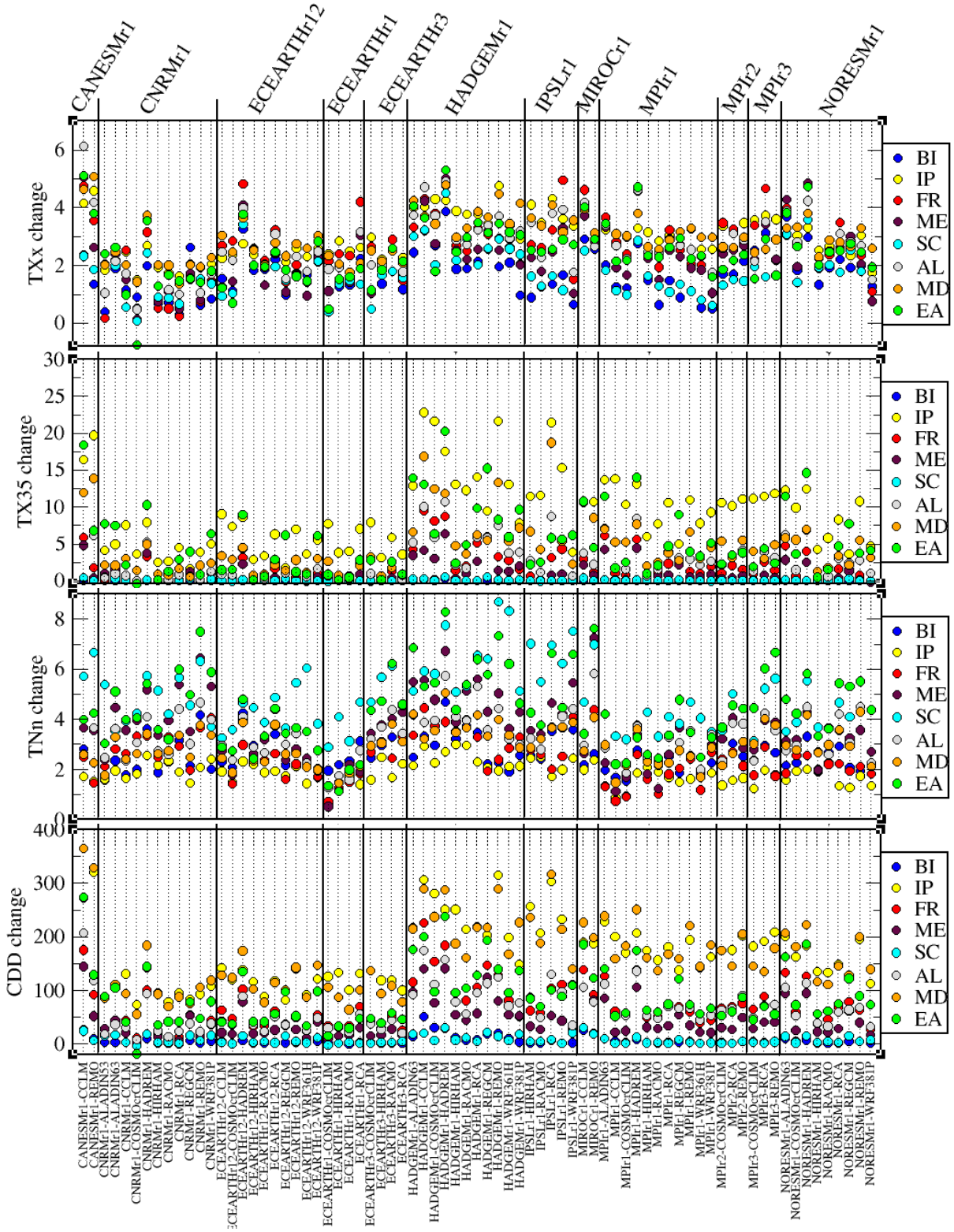


Figure 9: Changes expected by mid-century for RCP8.5 for TXx (°C), TX35 (#days/year), TNn (°C), and CDD (#degree.day/year).

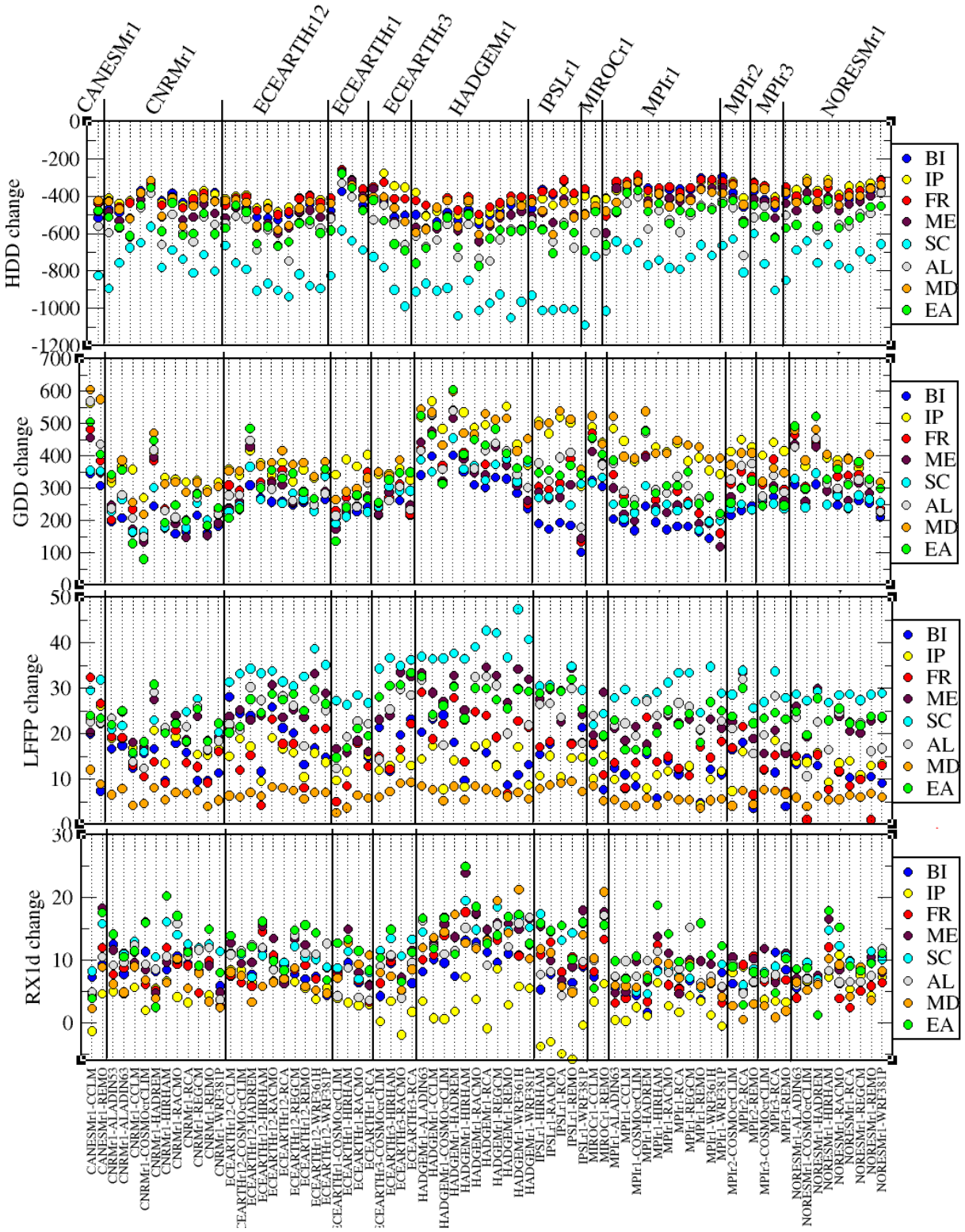
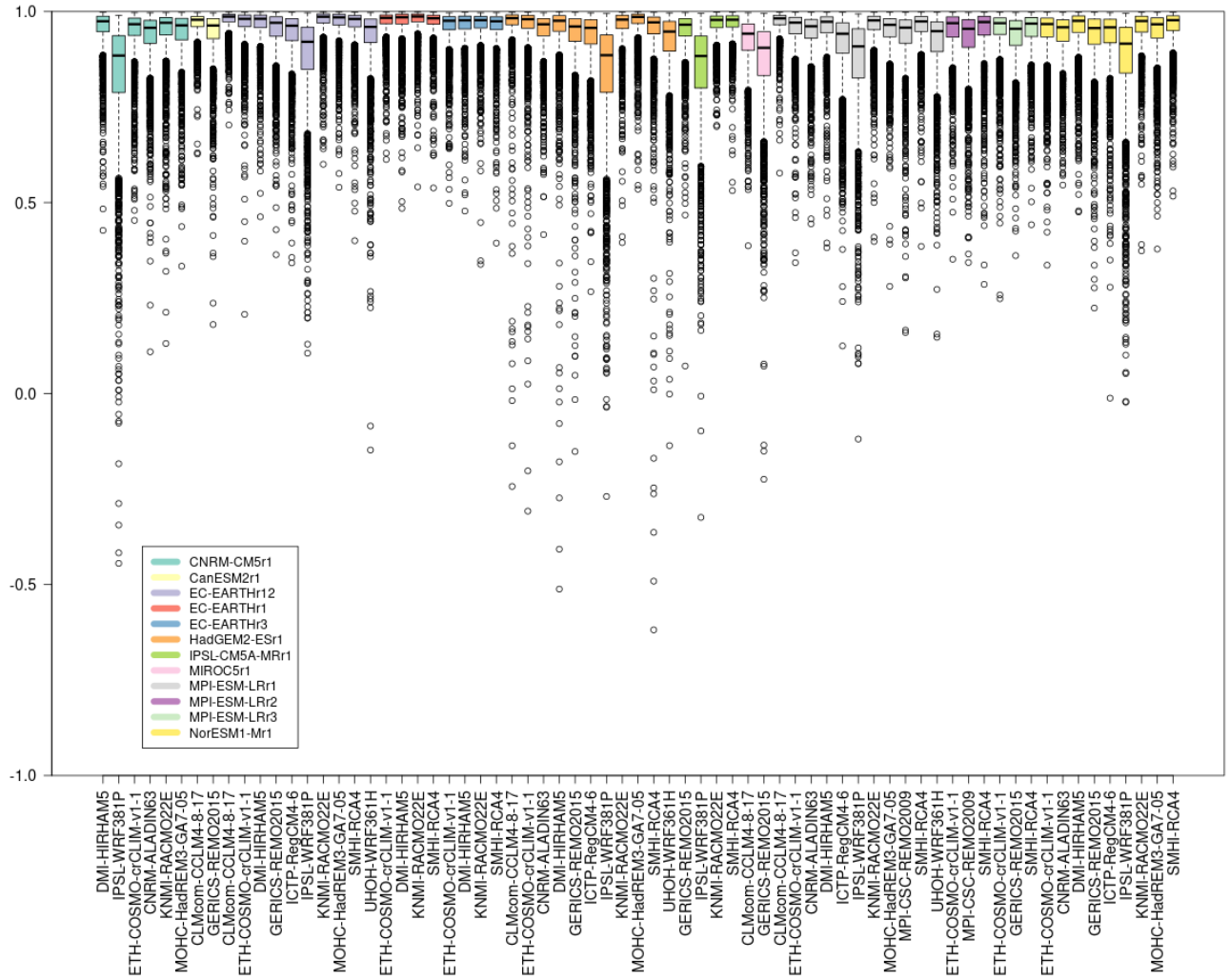


Figure 10: Changes expected by mid-century for RCP8.5 for HDD, GDD (#degree.day/year), LFFP (#day/year) and RX1d (%).

3.3 RCM-GCM consistency

We performed the same GCM-RCM consistency analysis as in previous analyses (e.g. C3S_D34b2.4.3.2) including a new regional climate model (HADREM) and all available simulations. The analysis is based on a spatial correlation analyses between sea level pressure fields of the RCM and it's driving GCM. Results are shown in Figures 11.



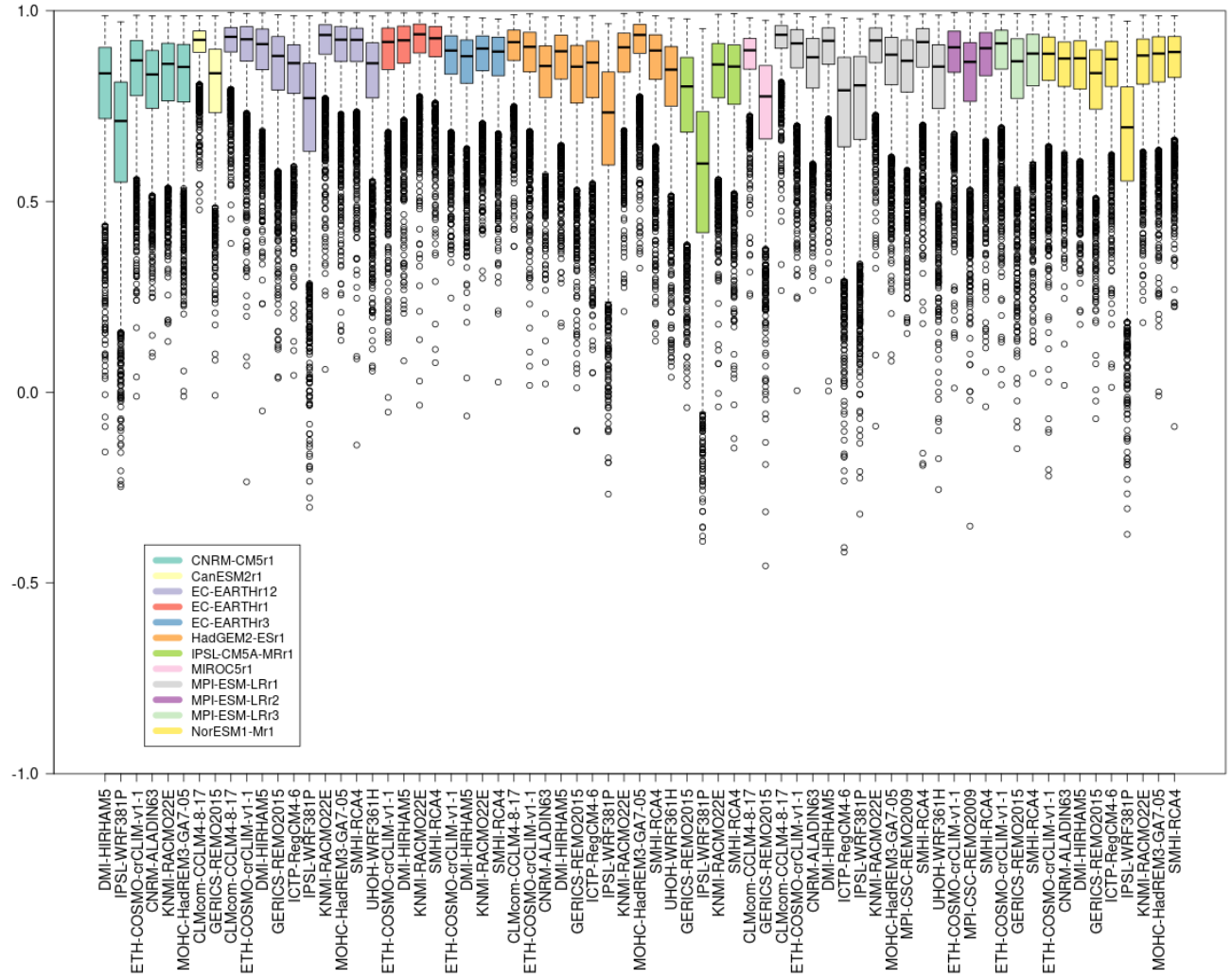


Figure 11: Distribution of mean sea level pressure spatial correlation between RCMs and driving GCMs, for simulation in the historical period (here taken as 1976-2005). Top panel, winter (DJF); bottom panel, summer (JJA). The X-axis lists the RCM name; GCMs are identified by different colors, reported in the legend. The boxes are constructed from the 25th and the 75th percentiles, with the median shown as a line inside it, while the whiskers extend to data up to a distance of 1.5 times the interquartile range. All the correlations outside the whiskers are shown as points. Note that the WRF361H simulations are included here using a simplified calculation of the sea-level pressure from the surface pressure and temperature.

In general, models show large RCM-GCM correlation, in particular in winter, as the large-scale flow is dominant during this season. In summer correlations are weaker. These results are similar to those presented for the smaller ensemble available at an earlier stage of the project (C3S_D34b2.4.3.2) confirming that the RCMs to a strong degree are consistent with their driving GCMs in how they represent the large-scale circulation.

In the earlier work it was noted that WRF381P showed a lower degree of correlation than the other RCMs. Here, a sensitivity experiment with WRF381P have been

conducted to try to explain why. The buffer zone used in all standard WRF381P experiments has 5 rows and the boundary of the RCM is relatively far away from the EURO-CORDEX domain (40 rows). By testing a new buffer zone with 30 rows implying that the inner domain is more similar to the EURO-CORDEX domain the pattern correlation increases, in particular in summer, from about 0.75 to about 0.85.

3.4 Added value of C3S funded simulations

The matrix-filling strategy has been defined and evaluated in WP1. However, the ensemble was built on a pre-existing ensemble and the C3S added a number of simulations to complete it. For the RCP8.5 scenario, 48 simulations were added (see Table 1) to the 27 already existing. Clearly, the full ensemble of 75 simulations now represents an improved balance across RCMs and GCMs and allows dedicated studies about the projection uncertainty. It also provides a large ensemble which can be used for robust statistics. Such are needed e.g. for extreme event attribution, and the full ensemble has already been used in a few recent attribution analysis cases (e.g. Luu et al., 2018; Vautard et al., 2019; see also https://www.worldweatherattribution.org).

In practice, ensemble averages from the 27-member (non C3S-funded) were consistent with those from the full ensemble, which is reassuring. In several cases biases are reduced with improved ensemble mean by the full ensemble, and the range of biases is more centered toward zero, which is also expected if simulations are independent.

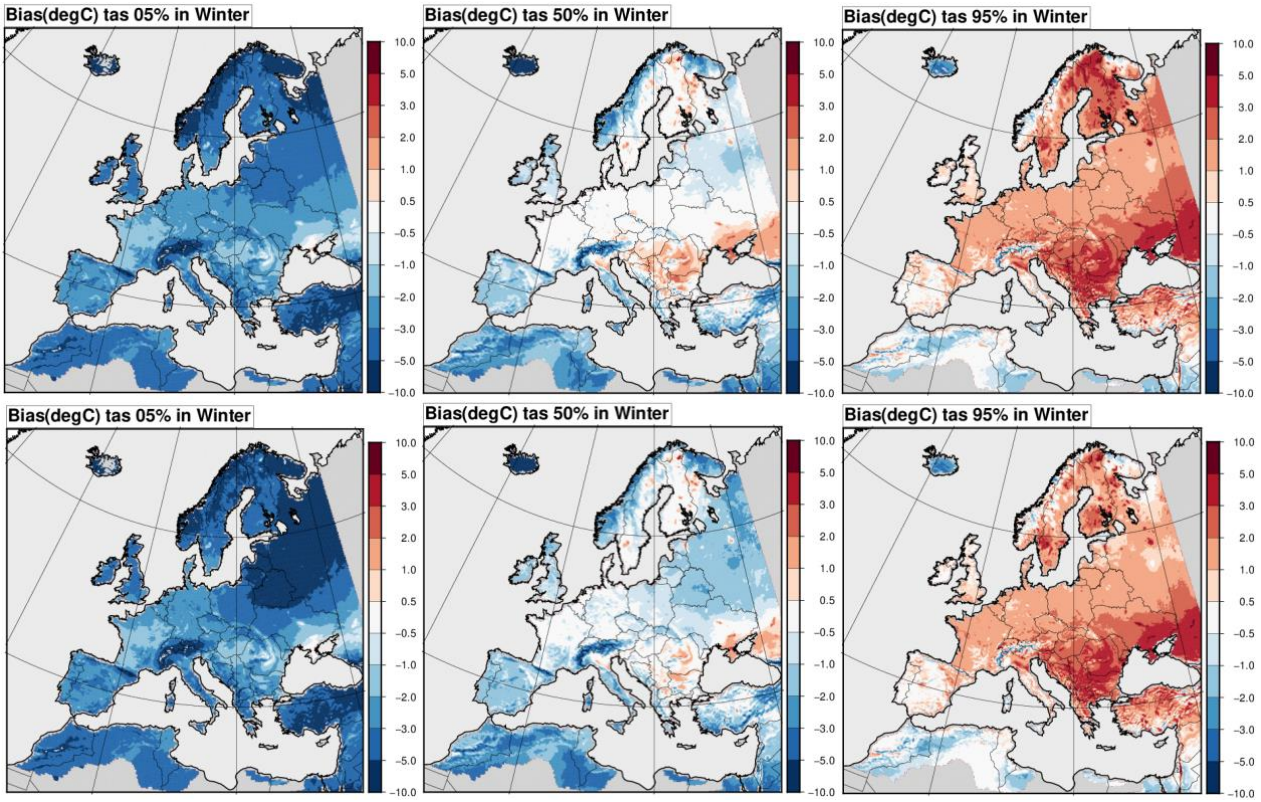


Figure 12: As in Figure 1 for temperature biases, but for daily mean temperature (Tas) in winter from the 75-member ensemble (top panels) and the 27-member ensemble (bottom panels).

Figure 12 shows such an example for winter temperatures. Over Eastern Europe and Russia, the 27-member ensemble (bottom panels) is more negatively biased than the 75-member panels, and gives more weight to cold-biased simulations in winter. We show this as an example, but there are also regional differences in summer and for other variables.

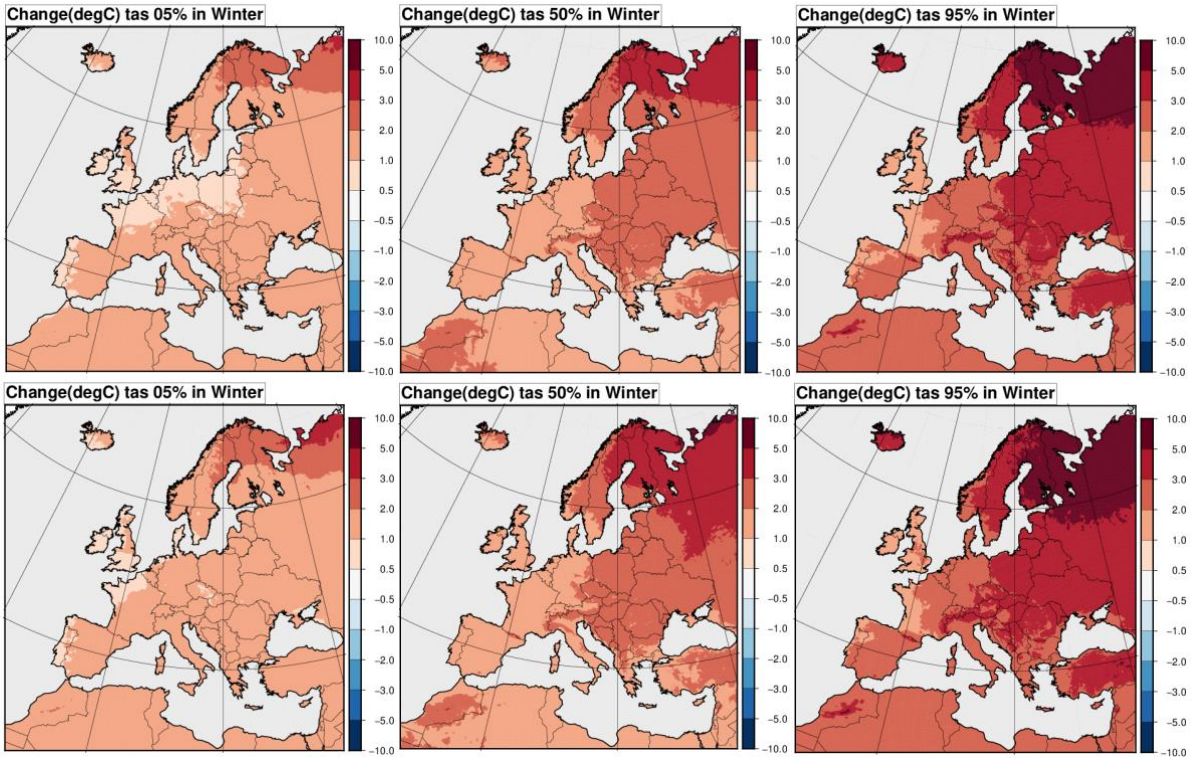


Figure 13: Same as Figure 12 for changes between 2036-2055 and 1981-2010 instead of biases.

The differences are also important for future climate change. Here, we note that for instance the 5%-95% range of uncertainty is modified by the full ensemble with regional differences. As an example, Figure 13 shows the changes in mean winter temperatures and the ensemble 5%, 50% and 95% centiles of the two ensembles. The patterns are very consistent, but the quantitative values change locally. For instance, the 75-member ensemble shows an extended range toward lower change values in Central Europe as compared to the 27-member ensemble. In fact, the whole distribution is shifted to lower values of change.

Figure 14 shows an example of the evolution of the ensemble over time along the PRINCIPLES project for one index, TXx, averaged over each region. The figure shows that median remains relatively stable while the bulk of the distribution can change significantly by increasing the number of ensemble members. It is hard to say if 75 is sufficient to reach convergence from these graphs, even though the difference between the 2020 (55 members) and the final ensemble appears minor in most cases.

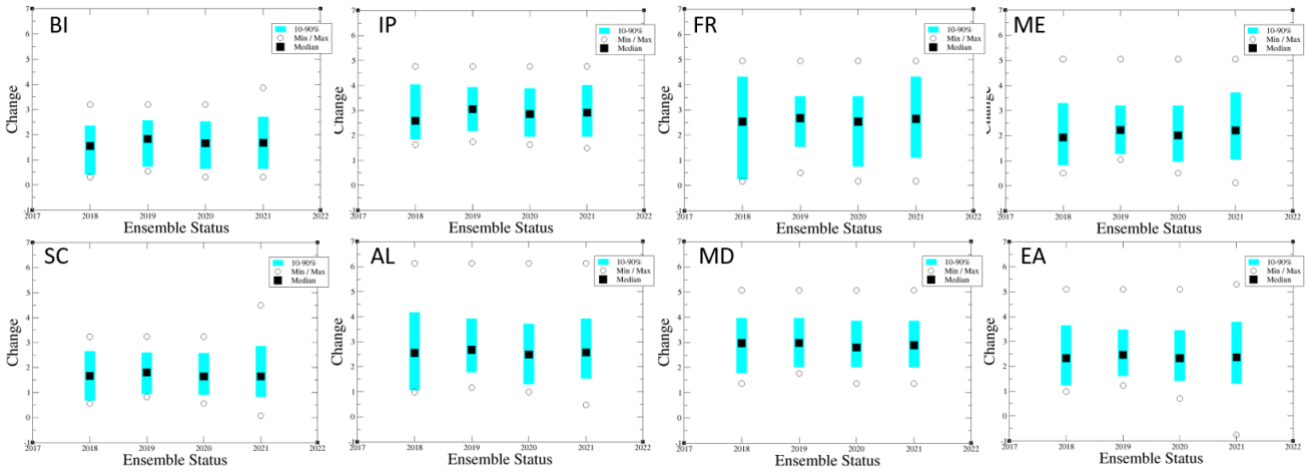


Figure 14: Range, 10-90% interval and median of the changes in TXx for each PRUDENCE region using the Euro-CORDEX ensemble made at different stages of the project (each year, in abscissa). The number of models increases with time to reach 75 in 2021.

Overall, this analysis shows that the addition of all C3S-funded simulation quantitatively modifies the distributions of biases and changes, but not the general spatial patterns.

4. Specific analyses

In this report we provide two specific sets of analyses apart from those reported upon above: one related to the water balance (section 4.1) and one to scaling methods for hourly precipitation extremes (4.2). In previous synthesis reports a number of specific analyses have been presented based on the simulations available at the time of writing. A few have been repeated while others have only been performed once.

The results of these can be seen as examples of different aspects of RCM performance and of GCM-RCM response to changing forcing conditions. Depending on application it is suggested that these may be repeated for any ensemble in focus for further work including impact studies.

As a repetition, here we briefly summarize which analyses that have been included in previous synthesis reports together with some main findings:

- Performance of EURO-CORDEX RCMs at 12 and 50 km resolution in comparison with a set of high-resolution (25-50 km) global climate model simulations from the European PRIMAVERA project. In general, it is found that these high-resolution GCMs show similar strengths and weaknesses as do the EURO-CORDEX RCMs. Both sets of models show improvements compared to coarse-scale CMIP5 GCMs. (see deliverable C3S_D34b_Lot2.4.3.3)
- Estimation of added value of RCMs relative to GCMs. A new method for calculating the spatial added value considering the full differences between PDF distributions was applied on 34 EURO-CORDEX simulations for precipitation, maximum winter and maximum summer temperatures. Results showed added value for all variables in most models. For precipitation, the added value was shown to be retained also when upscaled to a lower resolution of 2 degrees representative of a typical GCM resolution. (see deliverable C2S_D34b_Lot2.4.3.2)
- A comparison of the spread in the bivariate climate change signals for temperature and precipitation comparing GCMs and RCMs for 34 EURO-CORDEX simulations. It was shown that the RCMs substantially, and in some cases systematically, modulate the projected temperature and precipitation changes for summer compared to the GCMs. For winter, the bivariate changes in the RCMs are much more similar to those in the GCMs. (see deliverable C3S_D34b_Lot2.4.3.2)
- A simple scaling analysis relating dew point temperature with hourly precipitation extremes for two EURO-CORDEX RCMs at 0.11° resolution and for a convection-permitting regional climate model downscaling ERA-Interim reanalysis data to 2.5 km resolution. Results indicate that the two EURO-CORDEX RCMs show some weaknesses compared to the high-resolution convection-permitting model. The work presented in section 4.2 is an extension of this initial study. (see deliverable C3S_D34b_Lot2.4.3.2)
- A study of the water balance involving precipitation, evaporation, runoff and soil moisture, for 23 EURO-CORDEX RCM simulations was performed in deliverable C3S_D34b_Lot2.4.3.1 and later repeated in deliverable C3S_D34b_Lot2.4.3.2 for 30 simulations. Biases w.r.t. the observed climate was found to be determined to a large extent by the RCMs. For climate change signals it was less evident if the GCMs or the RCMs determined the level of change. A study of the water balance was also included in Vautard et al. (2021) and is extended with more simulations in section 4.1 below.
- A study of GCM-RCM spatial correlation was performed in deliverable C3S_D34b_Lot2.4.3.1 for 18 RCM simulations. This has later been repeated in

deliverable C3S_D34b_Lot2.4.3.2 for 31 RCM simulations and here in this report in section 3.3. The analysis shows a very high degree of correlation between RCMs and GCMs for wintertime conditions while correlations are lower in summer. It has also helped in identifying errors and inconsistencies in some RCMs or RCM-GCM combinations (e.g. see section 3.3).

- A study on past trends in near-surface temperature and surface shortwave downwelling radiation was performed based on 25 simulations. It was found that only one of the RCMs could reproduce the observed trend in shortwave radiation. The particular RCM applied a transient aerosol forcing instead of a constant one that is used in most other RCMs. For the temperature trends, different RCMs instead show different results depending on which GCM they are driven with. (see deliverable C3S_D34b_Lot2.4.3.1)

4.1 Water balance

The water balance analysis presented in Vautard et al. (2021) has been updated with all available new models (here, 71 models) and is represented in Figure 15. This figure is an update of that in Vautard et al. (2021) to which the reader is referred for more detail.

Figure 15 compares models for which the full water balance could be assessed. It shows the average over the 8 sub-regions for precipitation (pr), evaporation (evap), runoff (mrro) and soil moisture (mrso). For precipitation, regions IP and MD are dry in JJA while BI and AL are wet in both seasons. Evaporation in dry regions is limited by low soil moisture amounts and therefore lower than in other regions in summer.

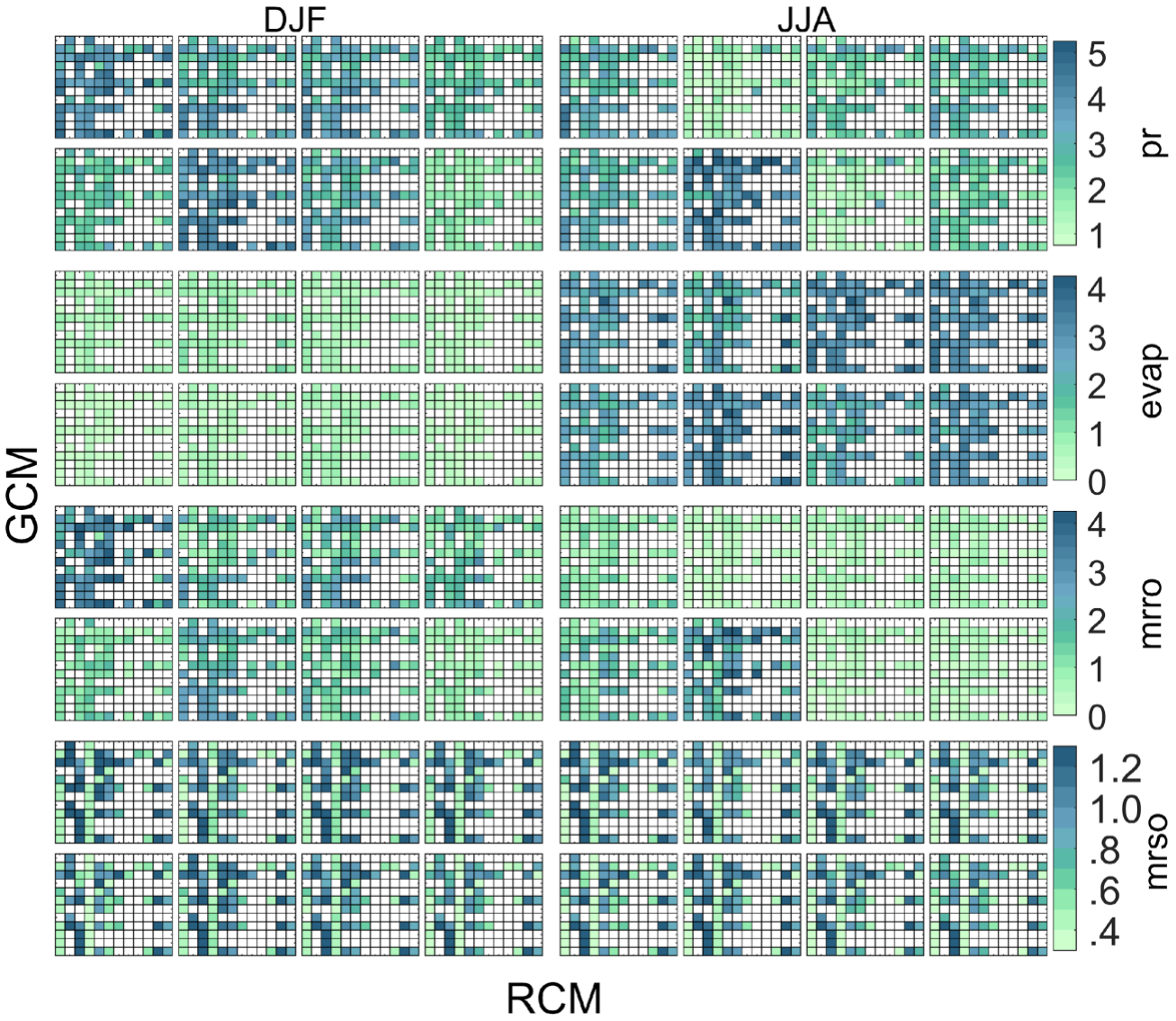


Figure 15: 1981-2010 means for 71 RCM simulations divided into a GCM/RCM matrix where the rows (GCMs) and columns (RCMs) match those in Table 1. Left column of panels is for DJF and the right column for JJA. The variables shown are precipitation (pr, first 2 rows, unit mm/day), evaporation (evap, second 2 rows, unit mm/day), runoff (mrro, third 2 rows, unit mm/day) and soil moisture (mrso, fourth 2 rows, unit m). For each variable and season, the data are divided into 8 PRUDENCE regions: British Isles, Iberian Peninsula, France and Mid-Europe in the first row and Scandinavia, Alps, the Mediterranean and Eastern Europe in the second row.

High run-off values over BI in DJF and AL in JJA are found as expected and models generally agree. By contrast, for total soil moisture content, a large degree of model disagreement is found. This may stem from the very different formulation of soil layers and soil modelling.

The qualitative conclusions are very similar using this larger ensemble than using the published one in Vautard et al. (2021).

4.2 Scaling of hourly precipitation extremes with near surface temperature and humidity in EURO-CORDEX evaluation experiments and observations.

In the preceding sections we have focused on assessing model performance and climate change at daily time scales considering daily data from the RCMs. As there is a strong interest from users of climate data to also utilize models for sub-daily data, such as that associated with heavy rain showers, we present here an investigation of how hourly precipitation extremes are represented by the models.

Scaling methods – utilizing dependencies of rainfall intensity on temperature, dewpoint temperature, and relative humidity derived from present-day climate variability – can be useful to analyze historical data and model behavior and to help understand potential future changes. As a fundamental background to all these methods is the Clausius-Clapeyron relation that states that air can contain more water vapor the warmer it gets with about a 7% increase per degree warming. Dependencies on surface dew point temperature have shown surprising regular behavior, in particular for hourly and sub-hourly rainfall (Lenderink et al. 2011; Westra et al. 2014; Fowler et al. 2021). Although these scaling methods cannot fully explain changes to (sub)hourly extremes, when used with care they can provide useful information on future changes to rainfall extremes (Fowler et al. 2021; Lenderink et al. 2021; Lenderink and Attema 2015). Here, we therefore provide a number of scaling results for the hourly rainfall derived from seven RCM simulations forced with reanalysis boundary conditions (evaluation experiments), in comparison with observations from the Netherlands.

Methods: All statistics are derived for rainfall events in the Netherlands. Data from 32-36 automatic weather stations relatively evenly covering the Netherlands have been pooled in space and time to avoid double-counting of events; details are given in (Lenderink et al. 2017). The period considered is 1991-2010 for the model simulations as well as the observations. In order to compare with more recent observations, the period 2011-2020 is also shown. In case of the observational data, some stations have missing data, but the records from around 32 stations are almost complete.

For the model data we took the grid point nearest to the location of the observational data. This sometimes led to the same grid point being chosen (one or two stations). In that case, the grid point is only included once in the pooled data set. Also, we did not

ignore model data from grid points at which observational data were missing, but just analyzed all available pooled data. In general, such mismatches in the data has hardly no impact on the results in this type of analysis, as shown e.g. in (Lenderink et al. 2021).

Pairing with two meter temperature and dewpoint has been done by taking (on average) the temperature measurement four hours before the rainfall measurement. Since the temperature data in the RCMs is stored on a three-hourly basis, this time period varies, but taking four hours at least guarantees that the rainfall occurred after the temperature/dewpoint measurement.

We did not consider the scale difference between surface observations (locally at a station) and modelled data (12x12 km2 grid box). Furthermore, we did not analyse any radar data but instead favored the longer time series of the station data as time series of radar data are shorter. Also, the rain radar data is known to be somewhat biased in the high intensity range. Finally, we note that, although absolute values between local rain measurements and 12x12 km^2 average rain radar data differ, the shown relations in the report are only marginally affected by this.

We performed two types of analysis here. First, we looked at the distributions of temperature, dew point temperature and dew point temperature depression (the difference between temperature and dew point temperature, as a measure of relative humidity). These cumulative distributions are plotted for selected events, based on rainfall amounts in the paired data: all hours, hours with rain, and hours with rain exceeding the 90^{th}, 95^{th} and 99^{th} percentiles of rainfall (a percentile is unconditionally based on all hours). This type of analysis was introduced in (Lenderink et al. 2011) as a simple means to evaluate temperature and moisture dependencies.

The second analysis is a so-called scaling analysis. To this purpose the paired (dew point) temperature-precipitation time series is divided into temperature bins. Typically, these are two degrees wide, and in each bin different percentiles of the distribution of rainfall are computed. Here, we used so-called wet conditioned (threshold 0.1 mm) percentiles: the 90^{th}, 99^{th}, and 99.9^{th} percentile (see e.g. Lenderink and van Meijgaard, 2008). Scaling diagrams based on temperature as well as dew point temperature are presented. We note, that, although scaling based on dew point temperature makes physically more sense (Lenderink et al. 2018), scaling diagrams based on temperature also give useful insights into the sensitivity of the model to temperature and humidity perturbations.

We considered the evaluation experiments with the following RCMs:

Institute	Forcing	RCM	version
DMI	ERAINT, 1991-2010	HIRHAM5	v1
KNMI	ERAINT, 1991-2010	RACMO22E	v1
SMHI	ERAINT, 1991-2010	RCA4	v1
GERICS	ERAINT, 1991-2010	REMO2015	v2
MOHC	ERAINT, 1991-2010	HadREM3-GA7-05	v1
ETH	ERAINT, 1991-2010	COSMO-crCLIM	v1
CNRM	ERAINT, 1991-2010	ALADIN63	v1

4.3 Results and Discussion

First, we consider the cumulative distribution of dew point temperature depression (that is, a measure of relative humidity), for which the results are shown in Fig. 16. The observations show a remarkable difference with most of the RCM model simulations. In the observations, heavier rain (from all wet events, events exceeding the 90th percentile to events exceeding the 99th percentiles) display higher dew point temperature depressions, equivalent to lower relative humidity. In most of the models, this behavior is not reproduced: for example, a reverse behavior is observed in HIRHAM5 and high intensity events exceeding the 99th percentile (5.8 mm hour^{-1}) occur predominantly (>80%) at dew point depression below 2 degrees (approximately relative humidity exceeding 90%). In the observations, the latter number is approximately 6 degrees. Most other models display similar behavior, yet less clear. HadREM3 is the only model displaying a behavior rather close to the observed behavior, although still not entirely matching the observed dependency.

We also note that the intensities projected by the model are too low, an outcome that has been obtained many times (see e.g. Berg et al. 2018). Most RCMs predict the 99th percentile to be between 3.4 and 5.2 mm hour^{-1}, whereas in the observations it is over 8 mm hour^{-1}. Partly, this may be explained by the difference in spatial scale. However, the area reduction factor at hourly time resolution estimated from radar data over Germany by Eggert et al. (2015) as approximately 20%, agrees well with the value obtained in convection permitting models from the grid-scale to the 12x12 km^2 scale. Therefore, the difference in scale is not the major factor here. Two models, COSMO (6.6 mm hour^{-1}) and HIRHAM (5.8 mm hour^{-1}) are reasonably close to the observed values. The distributions of dry-bulb temperature, shown in Fig. 17, and dew point temperature, shown in Fig. 18, partly reflect the same behavior. In the

observations, “extremes” occur at higher temperatures, for the exceedances of the 99th percentile approximately 4 degrees warmer than the average wet hour. In most models it is 1-2 degrees less, although COSMO and HadREM are close to the observations. The distribution of dew point temperature themselves is generally rather close to the observations.

Considering scaling based on 2m dew point temperature, as shown in Fig. 19, all models capture the positive dependency. Some models, e.g. ALADIN and RCA, remain close to the Clausius-Clapeyron (CC) rate of 7% per degree. The other models capture a ~2 times CC behavior in particular for the high dew point temperature regime. A number of models (HIRHAM, REMO, RACMO) show a rather clear change from ~CC scaling to ~2CC scaling at a dewpoint of 12-14 °C. HIRHAM also shows a leveling off at dew point temperatures above 17 °C, and we also note that this model has too many high dew point temperature events thereby extending the scaling curve beyond 20 °C (see black line at the bottom of the graph extending to 21 °C versus 19 in the observations, and also Figure 18). COSMO and REMO also appear to have too many of these high moisture events, yet to a lesser degree. COSMO and HadREM show a more even super CC behavior, which is closest to the observations across the whole dew point temperature range.

Scaling on (dry-bulb) temperature, as shown in Fig. 20, reveals more diversity. In the observations, a 2CC behavior is observed up to 20-22 °C. For higher temperatures the dependency falls back to the CC rate. A large number of RCMs (COSMO, HIRHAM, REMO, RACMO, RCA) do not capture this behavior very well, and show a levelling off or even decreasing rainfall intensities for temperatures beyond 20-24 °C. ALADIN and RCA do not capture a super CC behavior at all. HadREM3 is the only model that captures the observed dependencies rather well.

Concluding, the results of the RCMs show quite varying behavior. Some of the models do not capture a 2CC dependency (ALADIN and RCA), all models (except HadREM) are too sensitive to decreases in relative humidity, and as a consequence do not capture the temperature dependency in the high temperature regime. HadREM is the only RCM that appears to have a realistic balance representing both the dependencies on absolute as well as relative humidity, although also this model is too sensitive to decreases in relative humidity.

What do we learn from this in a climate change context? It looks that most RCMs studied here are unreliable to predict future changes in hourly extremes. With

reliable, we refer here to the ability of the model to predict *deviations from* the simple CC behavior. Taking the appropriate dew point temperature change, a CC behavior is directly linked to an increase in near-surface atmospheric moisture. Looking at the dependency on absolute humidity, two models fail to reproduce the observed 2CC dependency; the other models do reproduce this behavior to a reasonable degree (although some tend to overestimate the dependency for intermediate high dew points, and tend to underestimate the dependency for the highest dew points).

Yet, climate change is not only characterized by increases in absolute humidity, in particular in summer reductions of relative humidity over land are generally projected to occur. From the five models that reproduce a ~2CC behavior on dew point, only HadREM has a realistic dependency on relative humidity. In reality, decreases in relative humidity are associated with increases in intensity, at least for the moderate climate of the Netherlands. So, at the same absolute humidity as measured by the dew point temperature, lower relative humidity leads to stronger extremes (not directly shown here, but consistent with the dependencies on dew point depression, Figure 16). In the majority of models this dependency is not well captured, or even reversed (like HIRHAM). Convective permitting models appear to be better at representing these relations, with in general very good representation of the dew point dependencies as well as adequate representations of the dew point depression dependency (relative humidity). Considering that climate change is associated with change to both absolute and relative humidity, and that most RCMs do not realistically capture the observed intensity dependencies on these, we think that RCMs studied here (with possible exception of HadREM), are not a reliable tool to predict future change in hourly rainfall extremes.

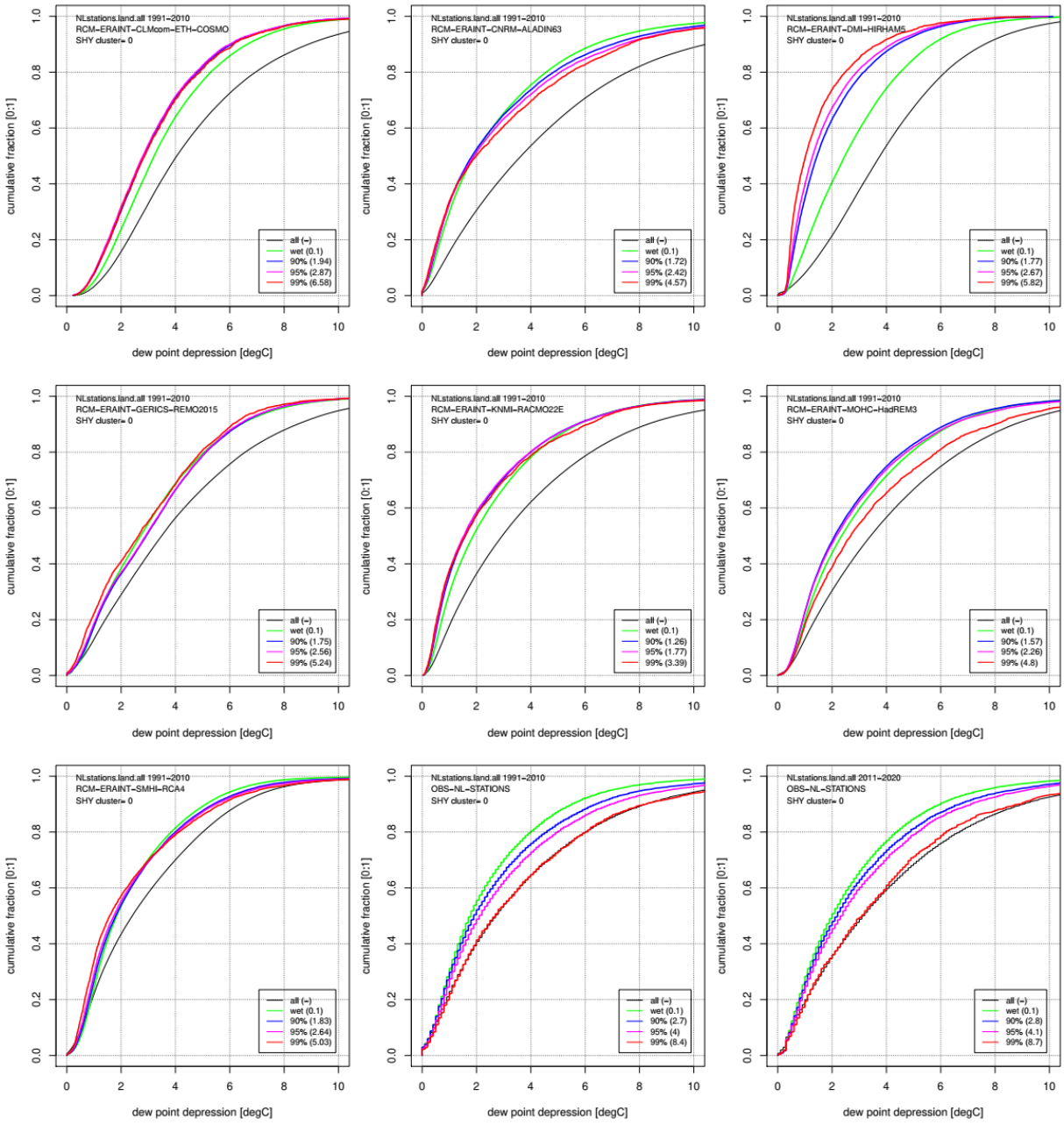


Figure 16. Cumulative distributions of dew point depression (the difference between the actual dry-bulb temperature and the dew point, which is a measure of relative humidity with approximately a change in relative humidity of 5% per degree) for all measurements (black), and a selection of rain events. Temperature and dew point are taken "4" hours before the rainfall. From green to red are the distributions for increasing amounts of rainfall, with the lower bound of the selection defined as indicated by the percentiles in the legend. The numbers between brackets in the legend indicate the corresponding precipitation threshold. The bottom-center and bottom-right panels show the observations from The Netherlands for different periods, while the other panels show model results derived for the period 1991-2010.

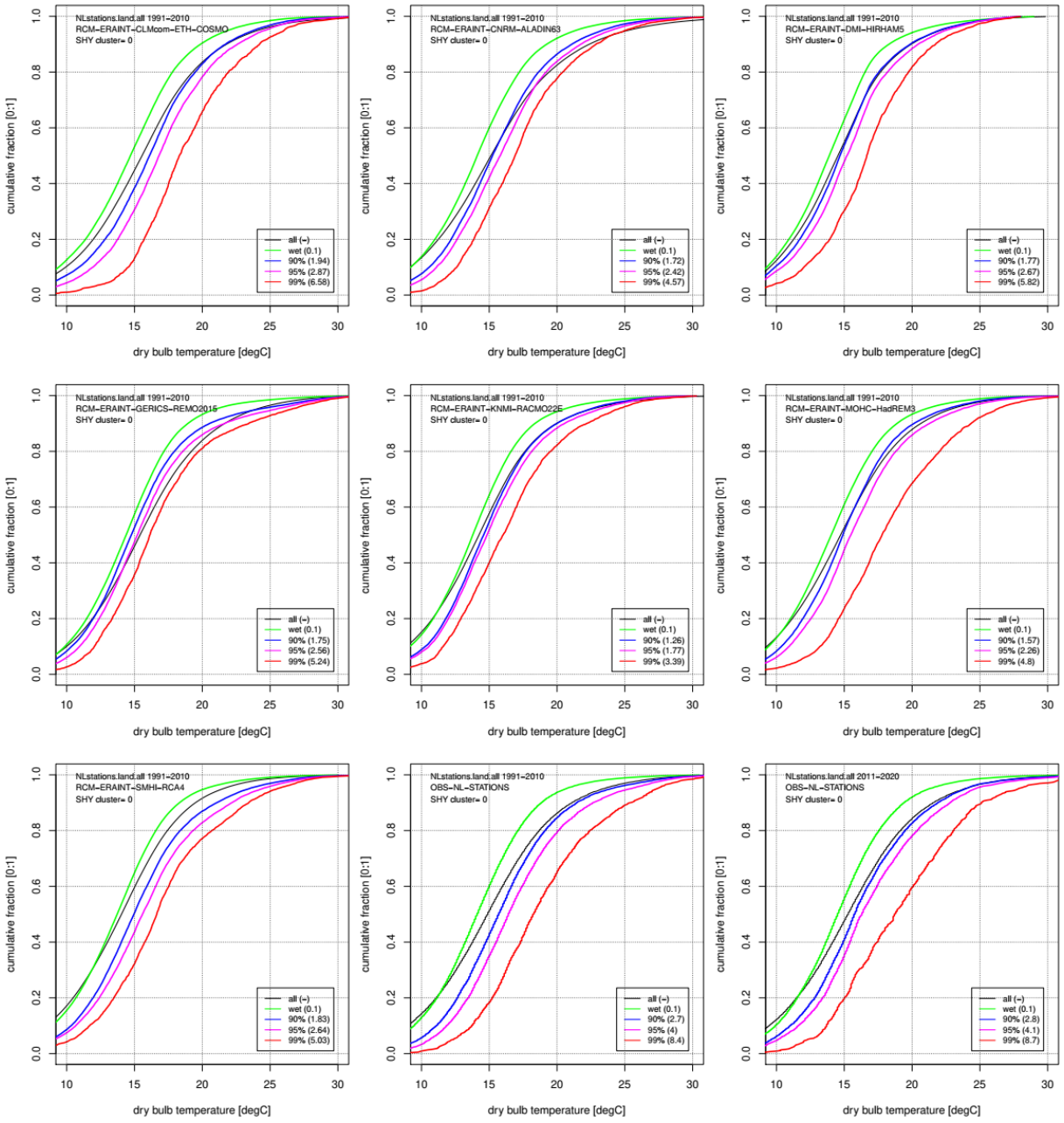


Figure 17. As Figure 16, but now for the distribution of dry-bulb temperature.

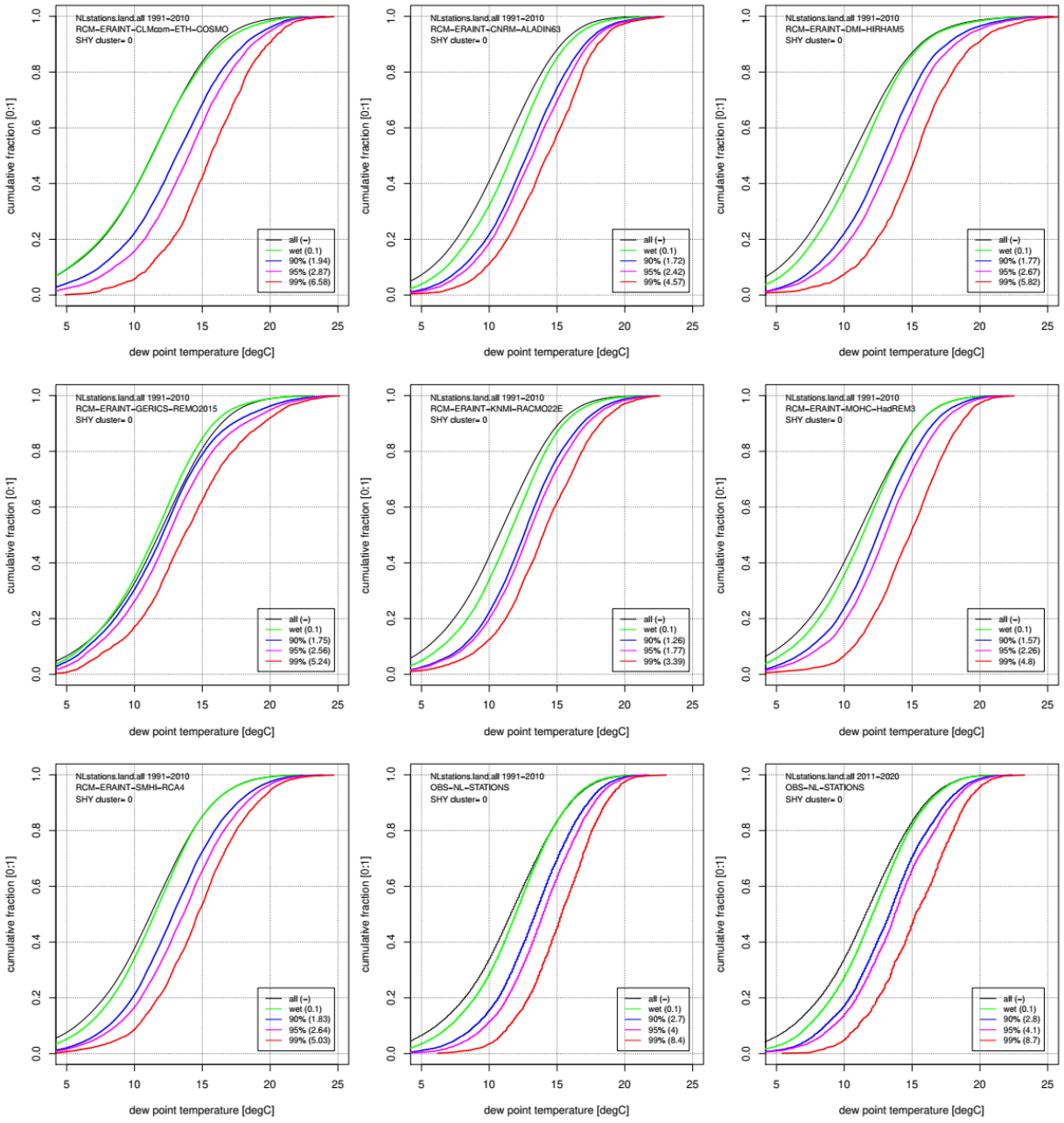


Figure 18. As Figure 16, but now for the distribution of dew point temperature.

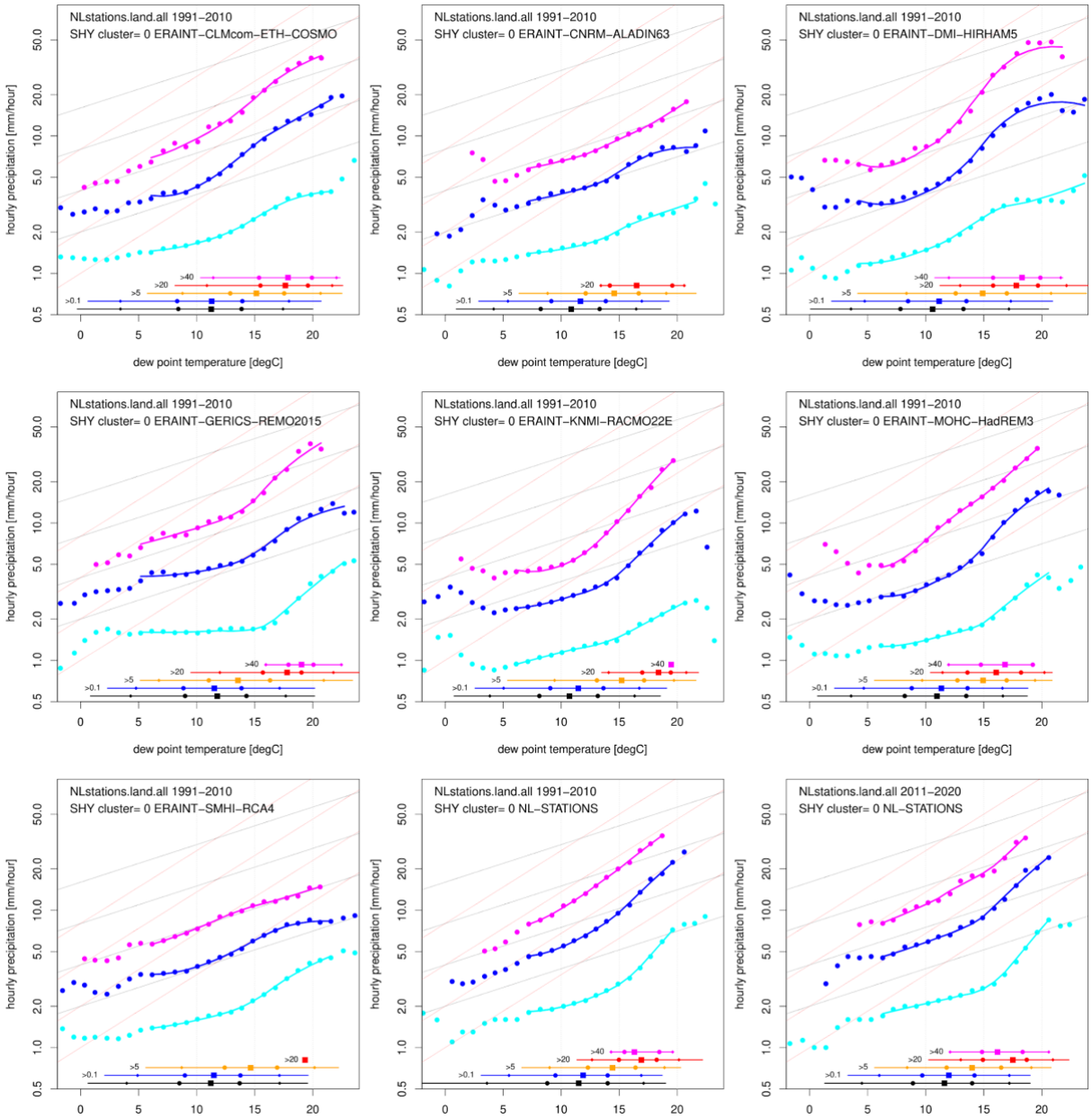


Figure 19. Scaling of hourly precipitation extremes as a function of dew point temperature, with the different percentiles shown in magenta (99.9th), blue (99th) and cyan (90th), conditional on the occurrence of rain. On the bottom the dew point temperature ranges for all data (black) and precipitation exceeding 0.1, 5, 20 and 40 mm are shown, with the 1 to 99 percentile range indicated by the line, and the 5,25,50,75,95th percentiles by the different markers. Straight stippled lines are CC (black) and 2CC (red) dependencies.

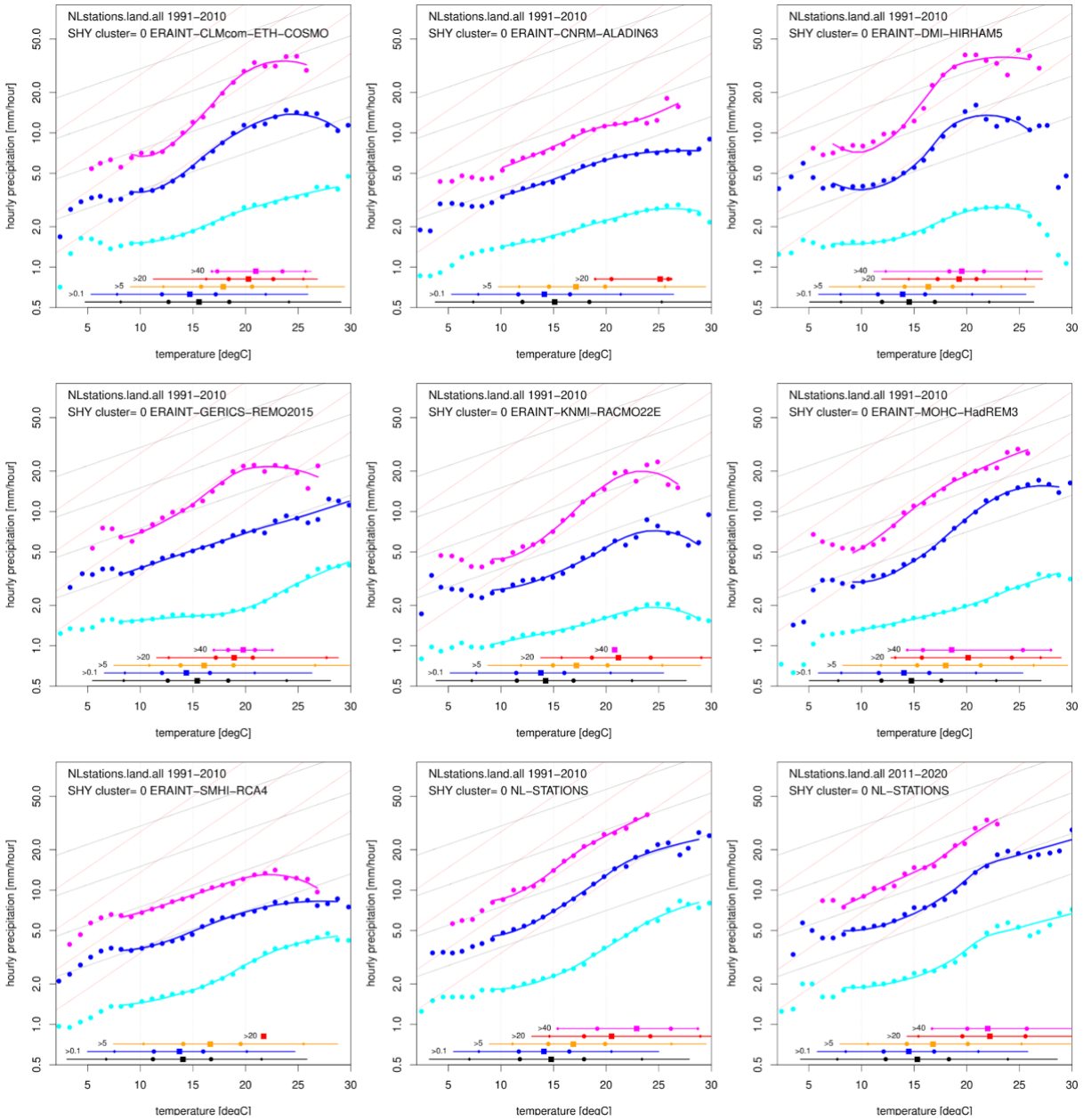


Figure 20. As Figure 19, but now for the (dry-bulb) temperature.

5. Acknowledgements

E-OBS temperature and precipitation v17 were used: We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu)" (Haylock et al., 2008)

For the WRF381P simulations the TGCC supercomputer centre graciously provided the computer time required for the simulations of the PRINCIPLES project.

6. References

Coppola, E., R. Nogherotto, J. M. Ciarlò, F. Giorgi, E. van Meijgaard, C. Iles, N. Kadygrov, L. Corre, M. S., S. Somot, P. Nabat, R. Vautard, G. Levavasseur, C. Schwingshackl, J. Sillmann, E. Kjellström, G. Nikulin, E. Aalbers, G. Lenderink, O. B. Christensen, F. Boberg, S. L. Sørland, M.-E. Demory, K. Bülow, C. Teichmann, 2021, Assessment of the European climate projections as simulated by the large EURO-CORDEX regional climate model ensemble, Journal of Geophysical Research: Atmospheres, 126(4), e2019JD032356.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. J. Geophys. Res. Atmos. 113(20). https://doi.org/10.1029/2008JD010201

Eggert, B., P. Berg, J. O. Haerter, D. Jacob, and C. Moseley, 2015: Temporal and spatial scaling impacts on extreme precipitation. Atmos. Chem. Phys., 15, 5957–5971, https://doi.org/10.5194/acp-15-5957-2015.

Fowler, H. J., and Coauthors, 2021: Anthropogenic intensification of short-duration rainfall extremes. Nat. Rev. Earth Environ., https://doi.org/10.1038/s43017-020-00128-6.

Lenderink, G., and J. Attema, 2015: A simple scaling approach to produce climate scenarios of local precipitation extremes for the Netherlands. Environ. Res. Lett., 10, 085001, https://doi.org/10.1088/1748-9326/10/8/085001.

Lenderink, G., H. Y. Mok, T. C. Lee, and G. J. van Oldenborgh, 2011: Scaling and trends of hourly precipitation extremes in two different climate zones – Hong Kong and the Netherlands. Hydrol. Earth Syst. Sci., 15, 3033–3041, https://doi.org/10.5194/hess-15-3033-2011.

Lenderink, G., R. Barbero, J. M. Loriaux, and H. J. Fowler, 2017: Super-Clausius–Clapeyron Scaling of Extreme Hourly Convective Precipitation and Its Relation to Large-Scale Atmospheric Conditions. J. Clim., 30, 6037–6052, https://doi.org/10.1175/JCLI-D-16-0808.1.

Lenderink, G., R. Barbero, S. Westra, and H. J. Fowler, 2018: Reply to comments on "Temperature-extreme precipitation scaling: a two-way causality?" Int. J. Climatol., 8–10, https://doi.org/10.1002/joc.5799.

Lenderink, G., H. de Vries, H. J. Fowler, R. Barbero, B. Van Ulft, and E. Van Meijgaard, 2021: Scaling and responses of extreme hourly precipitation in three climate experiments with a convection-permitting model. Philos. Trans. R. Soc. A, 1–33.

Luu, L. N., Vautard, R., Yiou, P., van Oldenborgh, G. J., & Lenderink, G. (2018). Attribution of extreme rainfall events in the South of France using EURO-CORDEX simulations. Geophysical Research Letters, 45, 6242– 6250. https://doi.org/10.1029/2018GL077807

Vautard, R., van Oldenborgh, G. J., Otto, F. E. L., Yiou, P., de Vries, H., van Meijgaard, E., Stepek, A., Soubeyroux, J.-M., Philip, S., Kew, S. F., Costella, C., Singh, R., and Tebaldi, C.: Human influence on European winter wind storms such as those of January 2018, Earth Syst. Dynam., 10, 271–286, https://doi.org/10.5194/esd-10-271-2019, 2019.

Vautard, R., N. Kadygrov, C. Iles, F. Boberg, E. Buonomo, K. Bülow, E. Coppola, L. Corre, E. van Meijgaard, R. Nogherotto, M. Sandstad, C. Schwingshackl, S. Somot, E. Aalbers, O. B. Christensen, James M. Ciarlo`, M.-E. Demory, F. Giorgi, D. Jacob, R. G. Jones, K. Keuler, E. Kjellström, G. Lenderink, G. Levavasseur, G. Nikulin, J. Sillmann, S. Lund Sørland, C. Steger, C. Teichmann, K. Warrach-Sagi, V. Wulfmeyer, 2021: Evaluation of the large EURO-CORDEX regional climate model ensemble, J. Geophys. Res., e2019JD032344.

Westra, S., and Coauthors, 2014: Future changes to the intensity and frequency of short-duration extreme rainfall. Rev. Geophys., 52, 522–555, https://doi.org/10.1002/2014RG000464.

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: info@copernicus-climate.eu