



HORIZON 2020

THEME SC5-2017



European Climate Prediction system

(Grant Agreement 776613)

European Climate Prediction system (EUCP)

Deliverable D5.4

Merging methods based on, and added value of, the high-resolution regional climate simulations

Deliverable Title	<i>Merging methods based on, and added value of, the high-resolution regional climate simulations.</i>	
Brief Description	<i>Report on merging methods based on, and added value of, the high-resolution regional climate simulations</i>	
WP number	WP5	
Lead Beneficiary	<i>Geert Lenderink, KNMI</i>	
Contributors	<i>Samuel Somot, CNRS-CNRM</i> <i>Antoine Doury, CNRS-CNRM</i> <i>Cécile Caillaud, CNRS-CNRM</i> <i>Hylke de Vries, KNMI</i> <i>Ole Christensen, DMI</i> <i>Dominic Matte, DMI/NBI</i> <i>James Ciarlo, ICTP</i> <i>Marie-Estelle Demory, ETH</i> <i>Peter Lind, SMHI</i> <i>Jorge Baño-Medina, University of Cantabria, Guest (not paid by EUCP)</i>	
Creation Date	<i>12/01/2022</i>	
Version Number	<i>[1.0]</i>	
Version Date	<i>10/03/2022</i>	
Deliverable Due Date	<i>31/03/2022</i>	
Actual Delivery Date	<i>13/05/2022</i>	
Nature of the Deliverable	R	R – Report
		<i>P - Prototype</i>
		<i>D - Demonstrator</i>
		<i>O - Other</i>
Dissemination Level/ Audience	PU	PU - Public
		<i>PP - Restricted to other programme participants, including the Commission services</i>
		<i>RE - Restricted to a group specified by the consortium, including the Commission services</i>
		<i>CO - Confidential, only for members of the consortium, including the Commission services</i>

Version	Date	Modified by	Comments
0.1	12/01/2022	Lenderink & all	Merged all input from google documents and merged into deliverable template
0.2	01/02/2022	Lenderink & all	Derived from input period up febr. 1 in google deliverable document and references
0.3	07/02/2022	Lenderink	Version submitted for internal review
1.0	09/03/2022	Lenderink & all	Response to internal reviewer comments

Table of contents

1	Executive summary	5
2	Project objectives.....	7
3	Detailed report.....	8
3.1	Introduction	8
3.2	Added value	11
3.2.1	Activity AV1: Added value of precipitation distributions (ICTP)	11
3.2.2	Activity AV2: Added value of precipitation, snow, and winds over Fenno-Scandinavia (SMHI, DMI and UCPH)	16
3.2.3	Activity AV3: Added value of hourly precipitation extremes (KNMI).....	24
3.2.4	Activity AV4: (Sub)hourly precipitation extremes (ETHZ).....	28
3.3	Signal to noise and regional predictability (SNP).....	33
3.3.1	Activity SNP1, predictability and signal to noise of fine-scale information in a regional SMILE (KNMI)	33
3.3.2	Activity SNP2: Spatial pooling to improve signal-to-noise (KNMI).....	38
3.3.3	Activity SNP3: Early emergence of the forced signal (UCPH/DMI)	39
3.4	Spatial Merging and emulators (SME).....	44
3.4.1	SME1: Large-scale circulation change and spatial merging (KNMI).....	44
3.4.2	SME2: UNet-based statistical emulator of RCM (CNRS-CNRM).....	53
3.5	Common framework on emulators (COM).....	64
3.5.1	Introduction and motivations.....	64
3.5.2	Definition of the common framework.....	65
3.5.3	Short descriptions of the 5 spatial merging methods.....	66
3.5.4	Results.....	67
3.5.5	Lessons learnt and range of applicability of each method	71
3.5.6	Appendix A: Method ANOVA (DMI)	73
3.5.7	Appendix B: Multiple regression method (KNMI).....	78
4	Closing remarks, lessons learned and recommendations	83
5	Publications.....	87
6	References	89

1 Executive summary

This deliverable captures the progress on a range of research activities connected to the provision and usefulness of high-resolution climate information, mostly derived from the regional climate models – conventional regional climate models (RCMs) and convection-permitting regional climate models (CPMs) as applied in WP3 of EUCP. The first aim is to establish the added value of the high-resolution climate model streams, RCMs and CPMs, compared to the coarser resolution global climate models GCMs (activity AV; section 3.2). This part of the research is mainly incremental to the large international efforts (see also WP3) and long-term scientific discussion on added value. As a more innovative and explorative part of the research, this deliverable discusses the predictability of fine-scale climate features from a signal-to-noise perspective (activity SNP; section 3.3) and captures the development of *statistical* spatial merging methods (activity SME; section 3.4) to emulate high-resolution climate information. Finally, for the first time a common framework inter-comparison between the different spatial merged methods is described (activity COM; section 3.5). Main findings can be summarised as follows:

- Improvement of climate statistics in high-resolution models in comparison to observations is well established for many variables. Due to their higher resolution, RCMs provide precipitation statistics closer to observations than coarser GCMs, with more heavy rain and less drizzle precipitation. Due to their ability to explicitly resolve deep convection, CPMs provide more realistic short-term and heavy precipitation than conventional RCMs, a more realistic diurnal cycle, as well as an ability to simulate sub-hourly precipitation statistics. Also more realistic snow and wind statistics are obtained in CPMs (activities AV1-AV4; section 3.2.1 to 3.2.4). We also found evidence that CPMs are able to capture absolute and relative humidity dependencies of hourly extremes related to convective rain, providing physically-based confidence in their ability to project future changes in convective extremes (activity AV3; Section 3.2.3).
- High-resolution models predict different changes than simply spatially interpolated GCM results to the high-resolution grid, in particular in mountainous areas or near coastlines (activity AV in section 3.2, SNP1 in section 3.3.1). This result is usually interpreted as “added value”; that is, the downscaling model modifies the climate change response in a systematic sense.
- For many climate variables, internal climate variability – climate noise resulting from the non-linear dynamics of the weather/climate system – imposes substantial constraints on the straightforward applicability of the modelling results. Even for 30-year simulations, systematic responses due to greenhouse gas forcing may be masked by unpredictable internal climate variability. This signal-to-noise issue gets worse for fine-scale climate change features, and systematic changes at these fine-scales may not be easily derived from single model simulations. We note, however, that this strongly depends on the variable at stake; temperature is generally much more predictable than precipitation (activity SNP1; section 3.3.1). In general, large(*r*) ensembles of regional simulations are needed to obtain reliable estimates of the systematic changes at small scales. However, we also showed that the pseudo-global warming (PGW) simulation technique is (in certain cases) able to provide reliable estimates at (much) lower computation costs; e.g. one 30-year simulation is sufficient

to capture the small-scale response in winter precipitation from a much larger ensemble (activity SNP1-SME1; sections 3.3.1 and 3.4.1).

- Spatial pooling of results may also increase the signal-to-noise ratio, yet only for phenomena that are small-scale, but not location specific. Convective showers represent a case where pooling appears successful and reasonably robust results can be obtained from CPM simulations even with a length of only 10 years (activity SNP2; section 3.3.2).
- Ideally, to cover all uncertainties related to downscaling, all available GCM simulations have to be downscaled with all available regional models. Obviously, obtaining such a full downscaling matrix is far beyond what can be achieved realistically (computationally but also practically). Statistical emulators can be used to fill the gaps in the GCM downscaling matrix (activity SME1,2, COM; sections 3.4.1, 3.4.2, 3.5). Two developments on spatial merging are specifically mentioned here:
 - It is investigated whether large-scale atmospheric circulation patterns provide a valuable basis for spatial merging. Various (relatively simple) merging methods using simple large-scale circulation predictors have been explored (regression, circulation clusters and circulation analogues). The most advanced method based on circulation analogues shows that most of the response patterns in a 30-year climate cannot be realistically reproduced from changes in large-scale circulation only. Apparently, other factors (such as sea surface temperature, change in other aspects e.g., stability change or subtle circulation changes) play a crucial role as well (activity SME1; section 3.4.1).
 - Within EUCP an advanced statistical emulator based on neural network machine learning has been developed by CNRS-CNRM. Such an emulator is trained on a large variety of large-scale predictors. The strength of this method is its ability to produce time series of high-resolution data, which in principle makes the method applicable for emulating extremes as well (activity SME2; section 3.4.2).
- The diversity of statistical emulators is large; both in terms of what they need as input, as well as the emulated output they can produce. Here, we provide for the first time a common framework to understand their differences, investigate their strengths and weaknesses, and compare their performance (activity COM; section 3.5). Five spatial merging techniques ranging from very simple to very advanced — pattern scaling, analysis of variance (ANOVA), multiple regression, and 2 artificial intelligence (AI) based methods – have been compared in this common framework. It is found that the simpler methods usually produce the most reliable results when evaluating mean winter and summer precipitation and temperature response between the middle of the century and the present-day climate. Nevertheless, advanced methods based on AI have a wider range of applicability and are more targeted to display realistic day-to-day variability, which the simple methods cannot represent. It appears that the focus on the day-to-day weather variations gives a penalty in these methods in a climatological mean sense. More research and development into these advanced AI based methods is therefore needed.

2 Project objectives

These deliverables have contributed to the following EUCP objectives (Description of Action, Section 1.1):

No.	Objective	Yes	No
1	Develop an ensembles climate prediction system based on high-resolution climate models for the European region for the near-term (~1-40 years)	Y	
2	Use the climate prediction system to produce consistent, authoritative and actionable climate information	partly	
3	Demonstrate the value of this climate prediction system through high impact extreme weather events in the near past and near future	partly	
4	Develop, and publish, methodologies, good practice and guidance for producing and using EUCP's authoritative climate predictions for 1-40 year timescales	Y	

3 Detailed report

3.1 Introduction

In EUCP three different climate modelling systems are used to project future climate change: global climate models (GCMs), regional climate models (RCMs), and convection permitting (climate) models (CPMs). These systems differ in terms of horizontal resolution, ranging from a grid spacing of ~100 km in the present-day generation GCMs, ~10 km in the RCMs, and ~2 km in the CPMs. Moreover, while GCMs and RCMs tend to share the same physics and dynamics – yet usually optimised for different resolutions – CPMs are fundamentally different in terms of their small-scale dynamics, which allows them to run without a deep-convective parameterization. One of main target application areas of CPMs is therefore the projections of the lifetime and intensity of convective showers, often occurring in so-called mesoscale convective systems (MCS) and the study of the sub-daily time scale or of local climates such as coastal zones, islands, cities or mountains.

One may question the benefits of the high-resolution climate modelling systems, compared to the low-resolution systems, RCMs vs. GCMs, and CPMs vs. RCMs (see a recent review in Lucas-Picher et al. 2021). Such added value is expected from better resolving small-scale features related to the geography – topography, small lakes, cities, surface heterogeneities, land-sea borders – and also atmospheric phenomena at different scales, like frontal zones, convective showers and wind gusts. While it is relatively straightforward to show “added value” for the present-day climate by using available observational datasets (see the assessment done in Del 3.2, but also Fumière et al. 2020, Caillaud et al. 2021), it is more difficult to “prove” that high-resolution models are more trustworthy for future changes (Brisson et al. 2021). Various efforts in Task 5.4 aim to quantify the “added value” of high-resolution systems both in evaluation and scenario modes. Some aim at comparing RCMs to GCMs, others at the added value of CPMs as compared to RCMs.

Related to the issue of added value is the signal-to-noise ratio that determines the importance of the climate change signal relative to the natural climate variability. The climate noise due to natural internal climate variability tends to increase at smaller spatial and temporal scales. It is possible to evaluate the signal-to-noise ratio by conducting multi-member ensembles of climate simulations using the same climate model. Each member of such an ensemble differs by small perturbations in the initial state so that differences between each member are solely due to climatic noise that is mostly unpredictable. Change patterns from each member show large variability and poor signal-to-noise ratios. However, spatial pooling can enhance the signal-to-noise ratio. For example, changes in convective rain statistics at a specific location are dominated by large uncertainty due to random climate noise but it is possible to obtain robust changes by spatially pooling rain data over larger areas.

RCMs use boundary data provided by the GCMs, and typically can only be run for a (rather small) subset of the available GCM runs (see Figure 1 for a hypothetical example). This implies that not all possible large-scale future climate states (as simulated by the GCMs) are covered with the current set of RCMs. These holes in the GCM/RCM matrix may introduce a bias in the climate response of RCM

simulation ensembles with respect to the GCM ensembles (Evin et al. 2021), as well as “unexplored futures” that may be important from a certain user risk perspective. International coordinated actions, such as the EURO-CORDEX initiative, have generated a large GCM/RCM matrix using 3 greenhouse gas scenarios (RCP), 10 GCMs and 13 RCMs to mitigate these problems and provide regional climate projections as trustworthy as possible. Nevertheless, there are still inevitable holes in the matrix (Evin et al. 2021). Moreover, while the EURO-CORDEX set focusses on sampling the high-end emission scenario RCP8.5 well, other emission scenarios are still relatively poorly sampled.

The fullness of the multi-model simulation matrix is even more problematic with CPMs. CPMs are computationally very demanding and can therefore only simulate rather short time slices of typically 10-30 years depending on the size of the computational domain. Despite large efforts in EUCP WP3 to produce these simulations, they constitute a so-called “ensemble of opportunity” in which the choice of the GCMs has mostly been based on pragmatic grounds, such as the availability of GCM driving data, rather than on their uncertainty range. This limits the straightforward applicability of CPM results, and careful analysis is needed to extract the useful information.

Thus, the other main task of T5.4 is to investigate how to deal with the fact that we have only a relatively small set of downscaling experiments, given the large number of GCM simulations available in CMIP5/CMIP6. Methods are developed that aim to statistically fill in, or emulate, these gaps in the matrix (“orange” blocks in Figure 1). Most of this research is still in an early development stage: uncertainties in the appropriate large-scale predictors and how to combine them using statistical methods are still large. Methods range from simple interpolation methods, such as pattern scaling or ANOVA-based approach (DMI/NBI), to advanced methods emulating the daily scale using neural network machine learning techniques based on various large-scale predictors (CNRS-CNRM, activity SME2; section 3.4.2).

Given the diversity of these approaches – in methodology, in input, and what they can emulate – a simple common framework to test and compare them has been designed. In this framework a mid of century climate is emulated using information from the present-day and end of the century climate (RCP8.5), using CNRM-CM5 r1i1p1@150km (GCM), CNRM-ALADIN63@12km (RCM) and CNRM-AROME41t1@2.5km (CPM). Here, we compare the strengths and weaknesses of the different methods; and try to establish what are promising ways to go forward. Besides comparing different methods, another aim of the common framework is to develop simple benchmarks of the quality of the emulated products, measuring the improvement compared to very simple statistical methods.

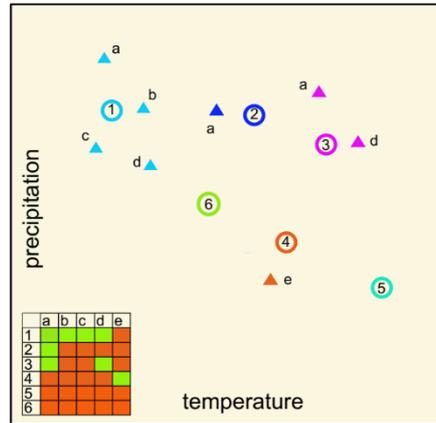


Figure 1. Schematic of a set of GCM/RCM simulations called “the matrix” (not based on real data) showing the response in temperature and precipitation in a system with 6 GCMs (circles, 1-6) and 5 RCMs (triangles, a-e). The matrix shows available simulations in green, whereas missing simulations are in orange. We note that this figure only shows the concept, and that there are many more GCM simulations, but also downscaled experiments available (e.g., ~130 in EURO-CORDEX@12km set). Also, the “emptiness” of the matrix varies considerably per region and per emission scenario.

In the next sections we discuss our activities. These are divided into three groups: activities on “added value” (AV), activities on signal to noise and predictability of the regional signals (SNP), and activities on spatial merging and emulators (SME). Since work on AV is a very broad research field – covering added value from RCMs compared to GCMs and CPMs compared to RCMs and GCMs, as well as many different climate variables at different spatial and temporal scales – it is not possible to cover the entire range here. The research in this part is therefore more incremental with respect to existing literature, and mostly focussed on rainfall extremes, where we expect substantial improvements from the high-resolution modelling systems. In addition to the AV question, we have put a considerable part of our research efforts on signal-to-noise and predictability (SNP) and the spatial merging (SME), which is the more innovative and explorative part of our research. As an example, the common framework to compare spatial merging methods is too our knowledge the first time such a coordinated action has been done. In addition, the hybrid AI based spatial merging technique are the first efforts to combine dynamical downscaling with advanced statistical methods.

3.2 Added value

3.2.1 Activity AV1: Added value of precipitation distributions (ICTP)

A new metric that quantifies added value (AV) was developed in ICTP. This method, adapted from past studies (Kanamitsu & DeHaan, 2011; Torma et al., 2015; Giorgi et al., 2016; Rummukainen, 2016; Fantini et al., 2018), compares the difference within the entire probability distribution functions (PDFs) of the global and regional models (GCM and RCM) with the observation at every grid point, to obtain a spatial distribution of AV (Ciarlo et al., 2021).

The data-sets required for the comparison (RCM, driving GCM, and high-resolution observation) must first be interpolated to a common grid, and the resultant PDFs must have the same characteristics (min, max, bin sizes). A distance-weighted interpolation was used, and the analysis was conducted at two grid scales (0.11°, 1.00°), showcasing consistent results of spatial AV, although with reduced spatial detail for the coarser resolution (seen in Figure 1). A small bin-size (analogous to a higher spatial-resolution; 1 mm/day for precipitation) was also necessary to maintain all the details of the PDF and obtain the most informative results possible.

The PDFs of a model (M) and the observation (O) are compared using the Relative Probability Difference, D (equation 1), where N is the number of events in a bin, v , and Δv is the bin size. This results in two spatial distributions that describe how much the models (RCM, GCM) differ individually from the observation, with larger values (unitless) describing a greater difference and zero describing a perfect match. These distributions are finally compared to obtain the AV (AV_i ; equation 2), which describes the improvement as positive values (or degradation as negative values) of the RCM over the GCM. The results of this process are showcased in Figure 2.

$$D_M = \frac{\sum_{v=1}^{v_t} |(N_M - N_O)\Delta v|}{\sum_{v=1}^{v_t} (N_O \Delta v)} \quad (1)$$

$$AV_i = D_{GCM} - D_{RCM} \quad (2)$$

The method can be adapted to focus on the tail-end of the distribution, as some studies (Torma *et al.*, 2015; Prein *et al.*, 2016; Fantini *et al.*, 2018) have revealed that GCMs struggle to resolve precipitation extremes. To achieve this, the threshold value of the percentile of interest (for example, the 95th percentile) is calculated for the observation. This spatially-varying threshold is then applied to the PDF data as a filter, after which the corresponding D and AV_i can be obtained.

Equation 1 can also be adapted to assess the Climate Change Downscaling Signal (CCDS) of climate projections, by comparing the modelled future projection (Mf) to the corresponding historical data-set (Mh) instead of an observation source (Eq. 3 and 4). The sign of the resulting quantity is not analogous to that of equation 2 however, but instead describes which model has a dominant climate change signal. The downscaling signal is present where values are further from a zero value.

$$D_{Mf} = \frac{\sum_{v=1}^{v_t} |(N_{Mf} - N_{Mh})\Delta v|}{\sum_{v=1}^{v_t} (N_{Mh} \Delta v)} \quad (3)$$

$$CCDS = D_{GCMf} - D_{RCMf} \quad (4)$$

The AV and CCDS metrics were tested on the daily precipitation of a 55-model EUR-11 CORDEX ensemble (Figure 3), giving attention to the larger AV for extreme precipitation (Ciarlo` *et al.*, 2021; Giorgi *et al.* 2021). Such AV is expected from better resolved small-scale geographical features such as topography, small lakes, cities, and many others, as well as atmospheric phenomena, like convective showers and wind gusts.

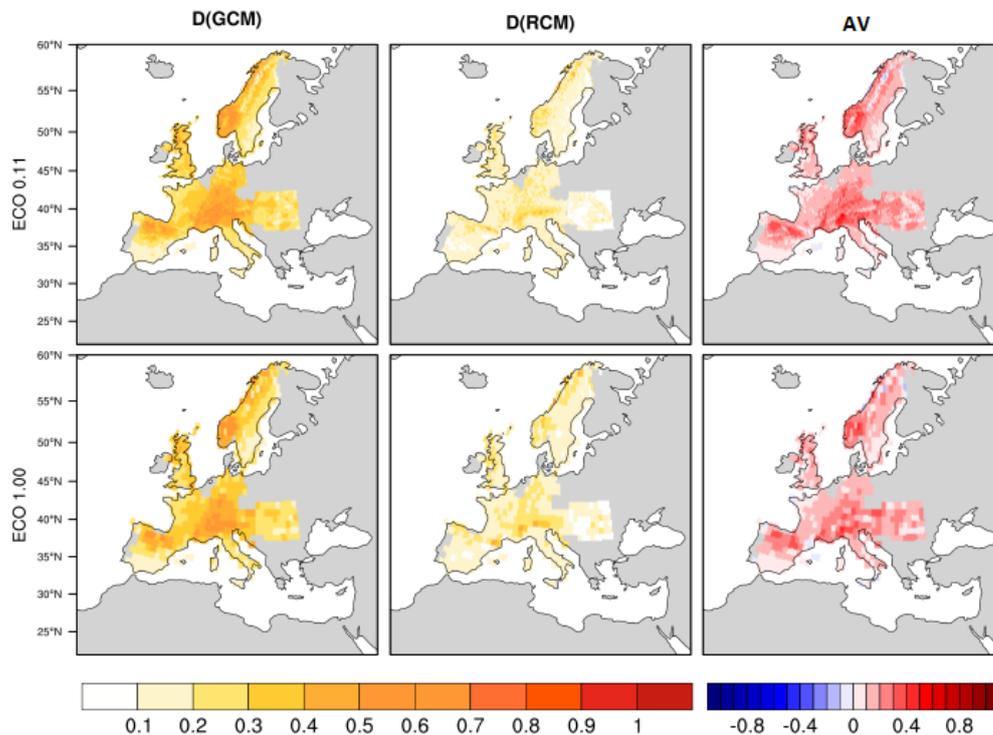


Figure 2. Relative Probability Difference (D) for the GCM and RCM 55-model ensemble, and the corresponding Added value (AVi) of daily precipitation, following an analysis on a composite of high-resolution observation sources (ECO) at two spatial resolutions (adapted from Ciarlo *et al.*, 2021).

This method is now being adapted to km-scale simulations using a multi-model approach, due to evidence of AV coming from such a scale which is worth objectively evaluating (Ban *et al.*, 2021, Pichelli *et al.*, 2021). The main challenges experienced at this stage involve the computational time and memory needed to process the added value in this manner. Although the final product is a small file, this method requires handling of large data-sets, and grid-by-grid comparisons, especially to calculate the PDF of each grid-point. The higher the resolution, and the larger the ensemble, the more challenging this metric becomes. The analysis at the hourly-scale is even more challenging because the data-set is 24-times larger.

The analysis is being focused on both daily and hourly data from a 14-model ensemble of the ALP-3 domain (shown in Table 1) using 5 high-resolution observation sources as reference (see Table 2). The primary objective is to assess the AV of the CPM with the driving RCM, but a comparison to the GCM is also included. A preliminary result projected on the 12 km RCM grid is shown in Figure 4 with the full PDFs, and Figure 5 using only the 99-100 percentile interval. One can note that the CPM run adds value over the RCM, possibly with emphasis in models/regions with lower RCM AV (requires

confirmation by comparing RCMs to the driving GCMs). This may explain the low AV seen beyond the 99th percentile, suggesting that the CPM may compensate for the ‘lacking’ of RCMs at lower percentile intervals.

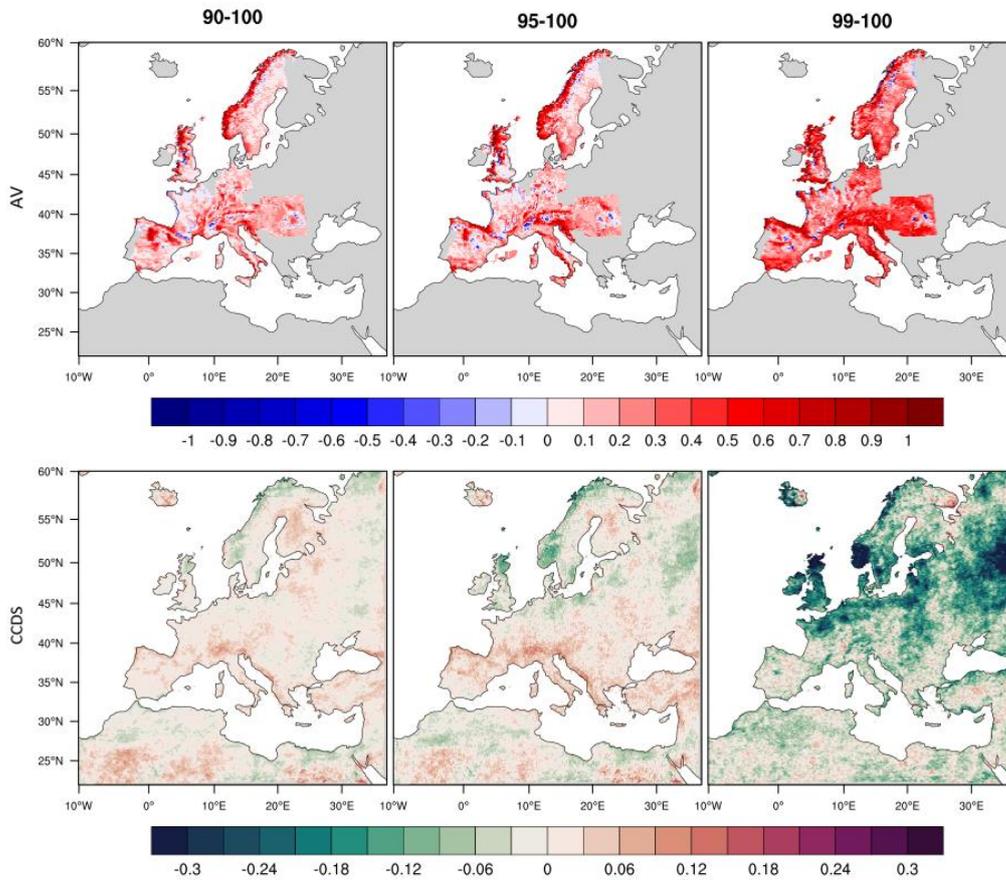


Figure 3. Added value (AV) and Climate Change Downscaling Signal (CCDS) of daily precipitation (at different percentile intervals: 90, 95, 99 to 100) for the 55-model EURO-CORDEX ensemble and the RCP8.5 far future at 0.11° (adapted from Ciarlo et al., 2021). The AV is compared to a composite of high-resolution observation sources, and the CCDS compares the 2080-2099 far future to the 1995-2014 reference period.

Table 1. A description of the 14-model ALP-3 ensemble members including the Convection Permitting Model (CPM), the intermediate RCM (where applicable), and driving GCM (together with its variant label). Note: The near/far-future simulations for BCCR and FZJ were prepared in collaboration with IDL and AUTH-MC (respectively); and will be included when the CCDS analysis is complete.

Institute	CPM	RCM	GCM	variant
CNRM	AROME41t1	ALADIN63	CNRM-CM5	r1i1p1
MOHC	HadREM3-RA-UM10-1	n/a	HadGEM3-GC3-1-N512	r1i1p1
CMCC	CCLM5-0-9	CCLM5-0-9	EC-EARTH	r12i1p1
HCLIM	HCLIM38-AROME	HCLIM38-ALADIN	EC-EARTH	r12i1p1
IPSL	WRF381CE	WRF381CE	CM5A-MR	r1i1p1
KNMI	HCLIM38h1-AROME	RACMO23E	EC-EARTH	r14i1p1
DWD	CCLM5-0-15	CCLM4-8-17	HadGEM2-ES	r1i1p1
ICTP	RegCM4-7	RegCM4-7	HadGEM2-ES	r1i1p1
KIT	CCLM5-0-15	CCLM4-8-17	MPI-ESM-LR	r1i1p1
ETHZ	COSMO-crCLIM	COSMO-crCLIM	MPI-ESM-LR	r1i1p1
JLU	CCLM5-0-15	n/a	MPI-ESM-LR	r1i1p1
WEGC	CCLM5-0-09	CCLM4-8-17	MPI-ESM-LR	r2i1p1
BCCR ¹	WRF381DA	WRF381DA	NorESM1-ME	r1i1p1
FZJ ²	IBG3-WRF381CA	IBG3-WRF381CA	EC-EARTH	r12i1p1

Table 2. Observation datasets used to assess the added value of the ALP-3 ensemble.

Region	Source	Type	Resolution	Period used	Reference
Italy	GRIPHO	gauge-based	12 km	2001-2016	Fantini (2019)
Alps	EURO4M	gauge-based	5 km	1971-2008	Isotta et al. (2014)
France	COMEPHORE	radar & gauge	1 km	1997-2017	Fumière et al. (2019); Tabary et al. (2012)
Germany	RADKLIM	radar & gauge	1 km	2001-2009	Winterrath et al. (2018)
Switzerland	RdisaggH	radar & gauge	2 km	2003-2010	Wuest et al. (2010)

The analysis is currently being replicated using hourly precipitation instead of daily, with a 0.1 mm/hr bin size. A preliminary result is shown in Figure 6, however, additional analysis is required in order to explain the significant differences observed with this data. Furthermore, the CPM data is currently being prepared for an analysis using the CCDS metric.

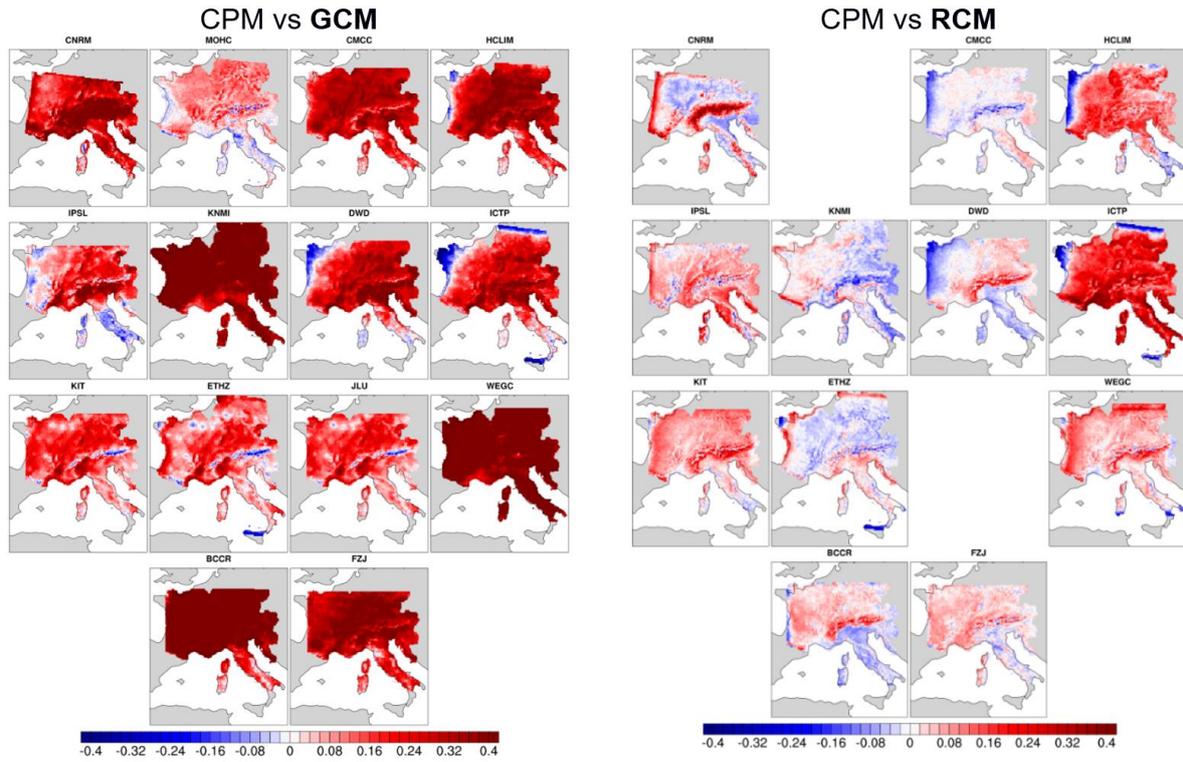


Figure 4. Added value for daily precipitation of the CPM ensemble comparing the CPM directly with the driving GCMs (left), and with the intermediate RCMs (right). Note that MOHC, and JLU members are excluded from the RCM comparison as these simulations were run directly from the GCM.

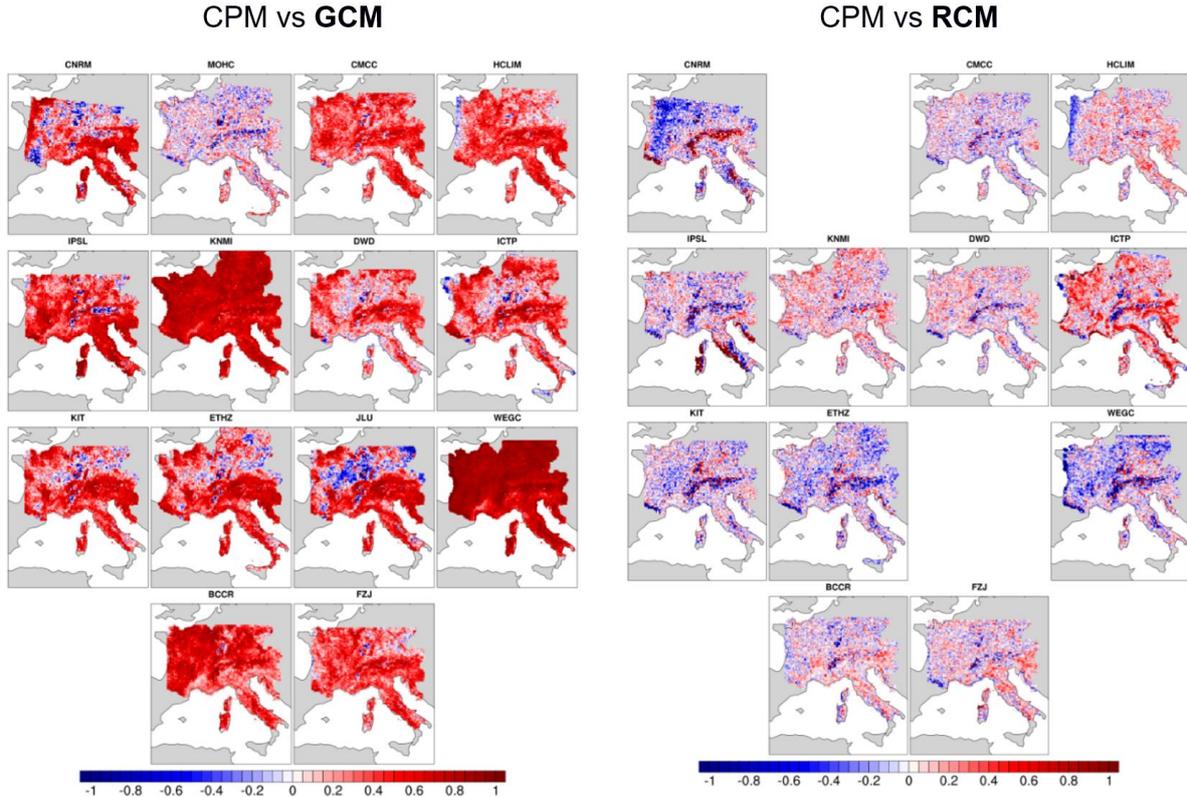


Figure 5. Added value for daily precipitation of the CPM ensemble as Figure 4, focusing on the 99-100 percentile interval.

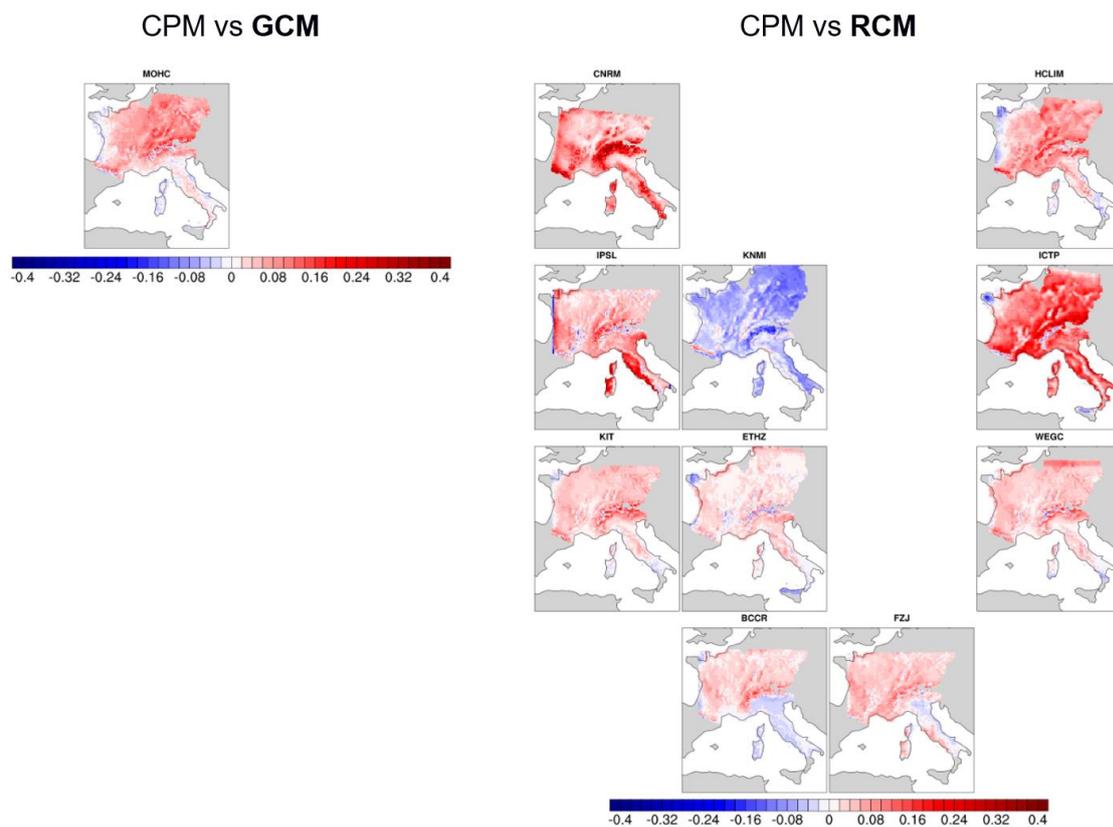


Figure 6. Added value for the CPM ensemble as Figure 4, focusing on hourly precipitation. Note that some panels are missing due to a lack of availability of hourly data (especially for the GCMs).

Publication:

Ciarlo` JM, Coppola E, Fantini A, *et al.* (2021). A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles. *Climate Dynamics* 57, 1403–1424. <https://doi.org/10.1007/s00382-020-05400-5>

3.2.2 Activity AV2: Added value of precipitation, snow, and winds over Fenno-Scandinavia (SMHI, DMI and UCPH)

Climate models project future warming that is more pronounced in northern Europe, including Fenno-Scandinavia, than the global average (e.g., Christensen et al. 2018; Kjellström et al. 2018). Average precipitation amounts are also projected to increase, especially in winter, while in summer models typically show increases in northern and decreases in southern parts of Fenno-Scandinavia (Christensen et al., 2018; Coppola et al., 2021). Along with mean changes, heavy precipitation events and summer warm extremes are commonly projected to increase in frequency and intensity (Nikulin et al., 2011). Anticipated changes in temperature, precipitation amounts, snowpack and snow cover under global warming will considerably impact surface energy and water budgets including hydrology processes through, for example, changed surface runoff as well as timing and amplitude of the spring flood.

Here, we present the performance of a suite of climate simulations that are part of the Nordic Convection Permitting Climate Projections project (NorCP; Lind et al., 2020) and funded by EUCP. The

simulations apply the HARMONIE-Climate (HCLIM) model cycle 38 (HCLIM38; Belušić et al., 2020) at 3 km horizontal grid resolution (HCLIM38-AROME, from hereon HCLIM3) with an intermediate step at 12 km grid resolution (HCLIM38-ALADIN, from hereon HCLIM12). Simulations include perfect boundary experiments, forced by ERA-Interim reanalysis, over the 1998-2018 time period as well as down-scaling of two GCMs (EC-Earth and GFDL-CM3) simulating three 20-year time-slices (historic: 1985-2005; mid-century: 2041-2060; and end-of-century: 2081-2100). We seek to provide insight to what extent a state-of-the-art CPRCM, HCLIM3, provides added value in representing the present climate over the Fenno-Scandinavia region, particularly for precipitation, and if HCLIM3 has different and plausibly more realistic climate projections compared to the coarser-resolution HCLIM12 RCM. Note that parts of the results presented here (Figure 7, Figure 8, Figure 10) are from Lind et al. (2020) while other results are from, yet unpublished, ongoing work (Figure 9, Figure 11-Figure 14).

In the analysis we have used the Analyzing Scales of Precipitation (ASoP) method (Klingaman et al 2017). ASoP gives a distribution of the contributions of each precipitation intensity bin to the mean precipitation rate. The distributions are calculated for each model grid point, and then averaged over desired regions. In the first step, the method defines the precipitation intensity bins such that all bins have a similar number of events, except for the largest bins due to the small number of events there. We used the same number of bins (80) for all ASoP analysis where the bin edges were defined according to Eq. 1 in Klingaman et al 2017. In the second step, the frequency of events in each bin is multiplied by the mean precipitation rate of the bin to obtain the actual contribution of the bin to the mean precipitation rate. Note that the sum of all actual contributions gives the mean precipitation rate.

Average and diurnal variation of precipitation

HCLIM is able to reproduce the spatial and seasonal patterns of daily mean precipitation over Fenno-Scandinavia, although with clearly higher amounts compared to E-OBS gridded observations over the Scandes mountains, especially in winter (Figure 7). In summer, there are larger differences between HCLIM12 and HCLIM3 where the former in general exhibits wetter conditions with on average 25% more precipitation than E-OBS. A major part of the overestimation of summer precipitation in HCLIM12 is due to higher frequencies of wet days, primarily days with low-to-moderate rain intensities (see also next section and Figure 8).

On the sub-daily time scales there is more clear evidence of added value for precipitation in HCLIM3, especially in summer when smaller-scale processes like atmospheric convection and surface-atmospheric boundary layer interactions become more prominent. For example, the more pronounced diurnal variation of precipitation in summer is better represented in HCLIM3 compared to HCLIM12 with considerably improved timing and intensity of the afternoon peak (see Figure 11 in Lind et al., 2020).

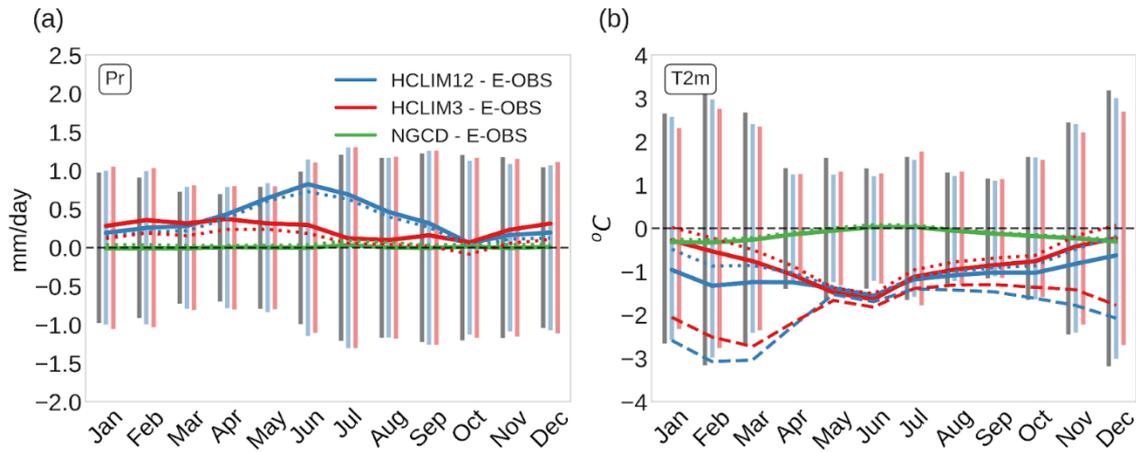


Figure 7. Annual cycles of precipitation (a) and near-surface temperature (b) anomalies with respect to E-OBS over Fenno-Scandinavia. Solid lines represent all grid points and dotted lines grid points below 500 m altitude based on E-OBS orography. Dashed lines are temperature from the open land tiles in HCLIM. Vertical bars represent \pm one standard deviation of monthly mean values in E-OBS (grey), HCLIM12 (blue) and HCLIM3 (red). Figure adapted from Lind et al. 2020.

Precipitation intensity distributions in present and future climate

Through the use of the ASoP method the full hourly precipitation intensity distributions are here evaluated for present climate and end-of-century climate scenarios over Fenno-Scandinavia. When compared to observations we note that the improved representation of precipitation in HCLIM3 is manifest over almost the entire spectrum of precipitation rates, both in winter (not shown) but especially in summer (Figure 8). This indicates a higher skill in representing the underlying processes, particularly through the explicit treatment of moist deep convection. HCLIM12, with parameterized convection, instead distinctly overestimates low-to-moderate precipitation rates (Figure 8), a major culprit for the wetter conditions seen in the seasonal averages (Figure 7).

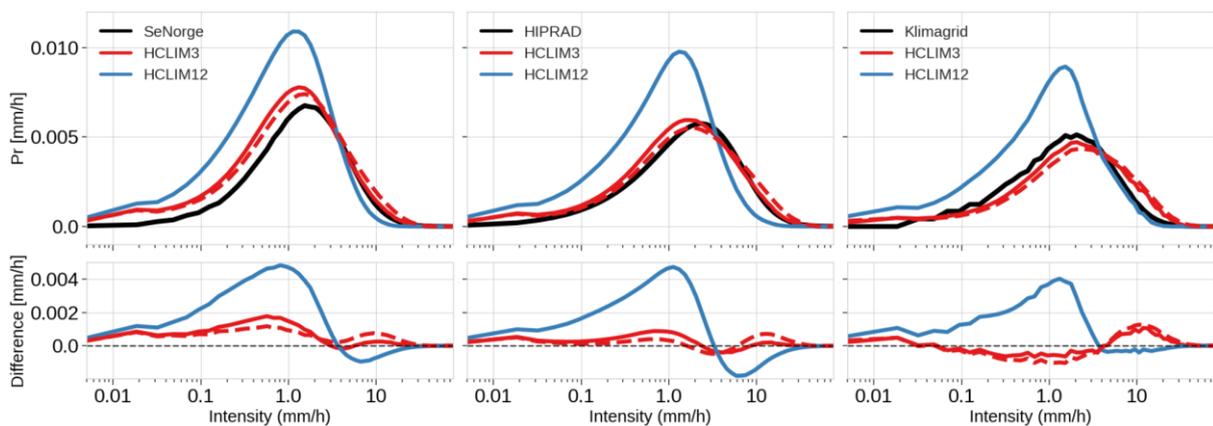


Figure 8. ASoP actual contributions per intensity bin to the total mean precipitation for JJA hourly precipitation over Norway (left), Sweden (middle) and Denmark (right). Model data is compared to national high-resolution data sets (in black). All data has been kept on native grids, except that for HCLIM3 the analysis was made both on the native-grid data (dashed red line) and the data remapped to the HCLIM12 grid (solid red line). Figure from Lind et al. 2020.

By the end-of-century (LC) in the RCP8.5 scenario, HCLIM12 and HCLIM3 show increases in DJF total mean precipitation (Figure 9a) compared to the historical period (CTRL), as depicted by the positive

change in contributions to the total mean at most precipitation rates (except for low intensities < 0.5 mm/h). In JJA, the bulk of the increase in contributions to the total mean occurs near intensities of around 8 mm/h in HCLIM3 while for HCLIM12 it occurs at lower intensities, around 3 mm/h. There is also a stronger decrease in contributions from lower precipitation rates in HCLIM3 (below ca 2 mm/h) (lower right panel in Figure 9a).

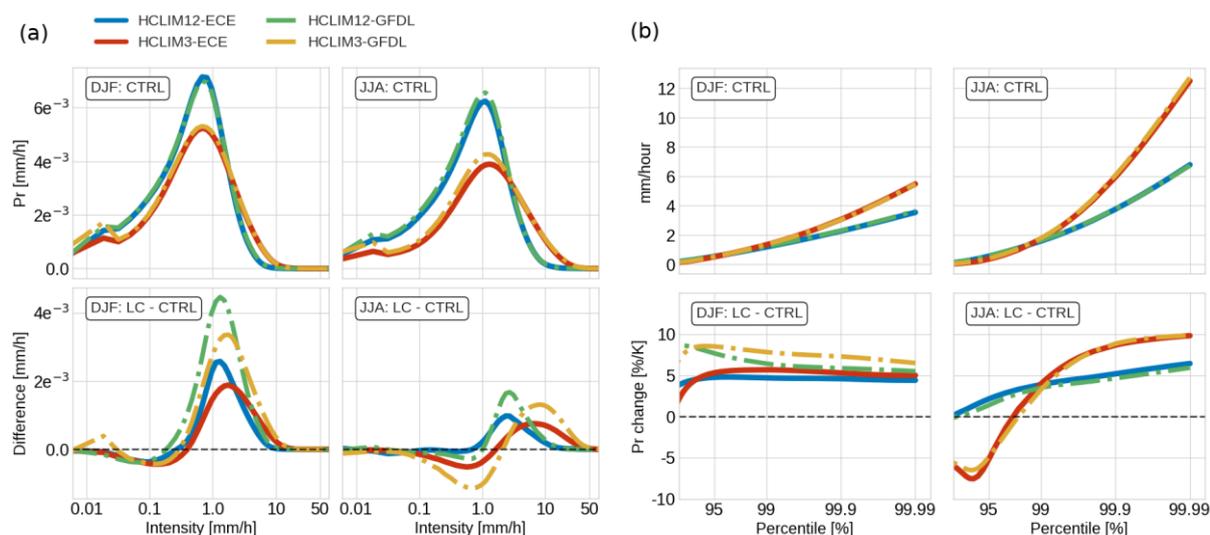


Figure 9. DJF and JJA hourly precipitation over Fenno-Scandinavia in historic (CTRL; top panels) and the change by end-of-century (LC-CTRL; bottom panels) in HCLIM3 and HCLIM12 forced by EC-Earth (ECE) and GFDL-CM3 (GFDL). (a): Actual contributions per intensity bin to the total mean precipitation; (b): High percentiles using all data, i.e. including dry hours. The differences (LC-CTRL) are normalised by the domain mean change in T2m. All data has been kept on native grids.

Changes in high intensity precipitation

The different responses in DJF and JJA precipitation between HCLIM12 and HCLIM3 is even further emphasised for the changes in high percentiles of hourly precipitation (normalised by change in near-surface temperature) (Figure 9b). In DJF the models tend to converge toward a value of between 5 and 7 %/K for the highest percentiles, and the HCLIM12 and HCLIM3 simulations forced by the same GCM follow relatively closely together. In JJA, in contrast, the two HCLIM12 and HCLIM3 runs respectively follow each other's change values more tightly, HCLIM3 having larger values than HCLIM12 for rankings above the 99th percentile. The different behaviour between winter and summer is an important indication of the impact of model grid resolution and physics (in particular the handling of cloud and precipitation processes) on these change patterns, as convection plays a more important role in JJA compared to DJF.

Snow climate in the Scandes mountains

Added value of very high-resolution models is also expected in areas of complex terrain as small-scale and local processes and interactions are potentially better represented, such as orographically forced precipitation (e.g., Gutowski et al., 2020). Evaluating simulated snowfall is challenging for several reasons, primarily related to the difficulties in measuring snowfall (e.g., Rasmussen et al. 2012). To partially overcome this issue, we investigated the annual fraction of occurrence of solid precipitation and compared it to station-based observations (Figure 10). It is calculated by dividing the number of

days with solid precipitation by the total number of wet days. The observations are provided by the Norwegian Meteorological Institute (Lussana et al. 2018). Figure 10b is showing a clear increase of the fraction of solid precipitation with higher elevation when compared to Figure 10a, mostly due to a more realistic representation of the topography in HCLIM3. There are, however, also considerable changes over low elevation areas in Sweden and Finland, indicating that the different physics and microphysical schemes in the two model versions are also impacting the amount of snowfall. When zooming in on a smaller area (Figure 10c,d), one can see how the topography has an important impact on solid precipitation; the smoother topography of HCLIM12 results in an underestimation. The observations overlaid on Figure 10c,d (dots) reveal that HCLIM3 overall has a better representation of simulated solid precipitation, as a combined result of the different physics schemes and higher spatial resolution. Figure 10e shows the fraction of solid precipitation as a function of the elevation in HCLIM12, HCLIM3 and the observations (blue, red and black, respectively). It appears that while HCLIM3 is still not reproducing the observed snowfall frequency at the station locations, there is a clear shift toward a larger snowfall fraction for the higher resolution model, making HCLIM3 more consistent with the observations.

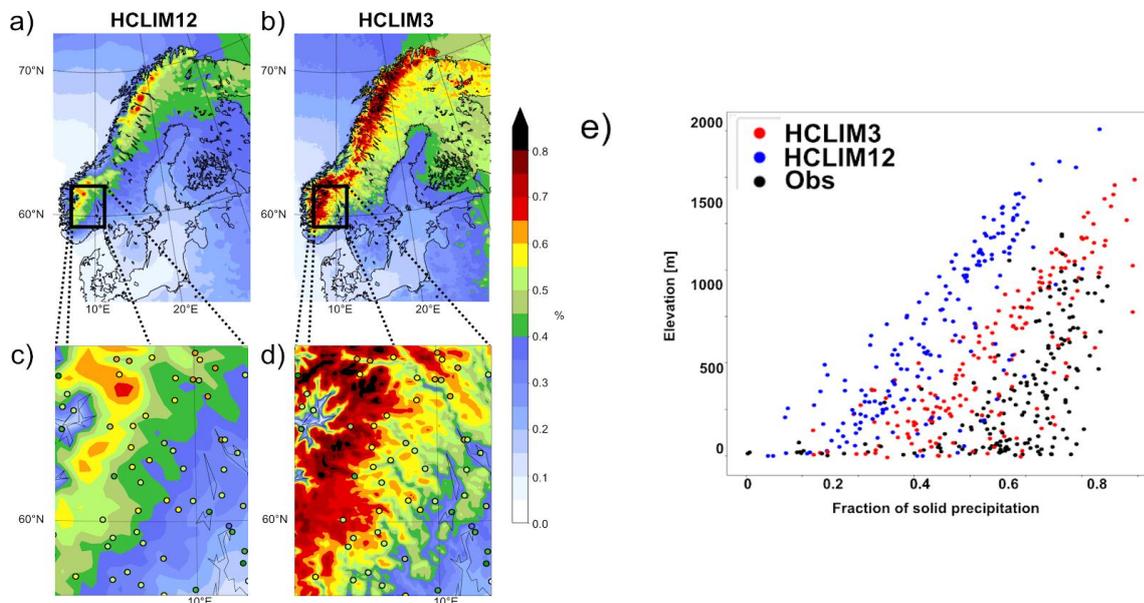


Figure 10. Annual mean fraction of solid precipitation compared to total precipitation as simulated by HCLIM12 (a) and HCLIM3 (b), while (c) and (d) are showing a regional zoom comparing the simulated results to station-based observations (black circles). (e) The fractions of solid precipitation as a function of elevation for observations (black dots) over Norway and the associated nearest grid point from HCLIM12 (blue) and HCLIM3 (red). Figure from Lind et al. 2020.

The warming seen in the scenarios, which is most pronounced in winter, has a strong impact on the snow climate over the Scandes with strong reductions in snow depth, cover and a shorter snow season (not shown). HCLIM12 and HCLIM3 exhibit similar daily mean precipitation in the cold season (October-March) between 200 and 800 metres in the Scandes (left panel in Figure 11). At higher altitudes HCLIM3 has a stronger increase with height in precipitation amounts, an indication of a stronger forcing from the orography to produce precipitation when air is moved over the mountains (as also seen in Figure 10). The middle and right panels of Figure 11 show the changes by end-of-century in daily mean precipitation and in the ratio of solid to total precipitation respectively. The change is positive in all simulations, but the magnitude is different between the two RCM-CPRCM combinations - with the largest increase in the HCLIM runs forced by GFDL GCM. Interestingly, the

reduction in the fraction of solid precipitation (right panel in Figure 11), due to the warming, is systematically less in HCLIM3 than in HCLIM12 by around 5-10 %.

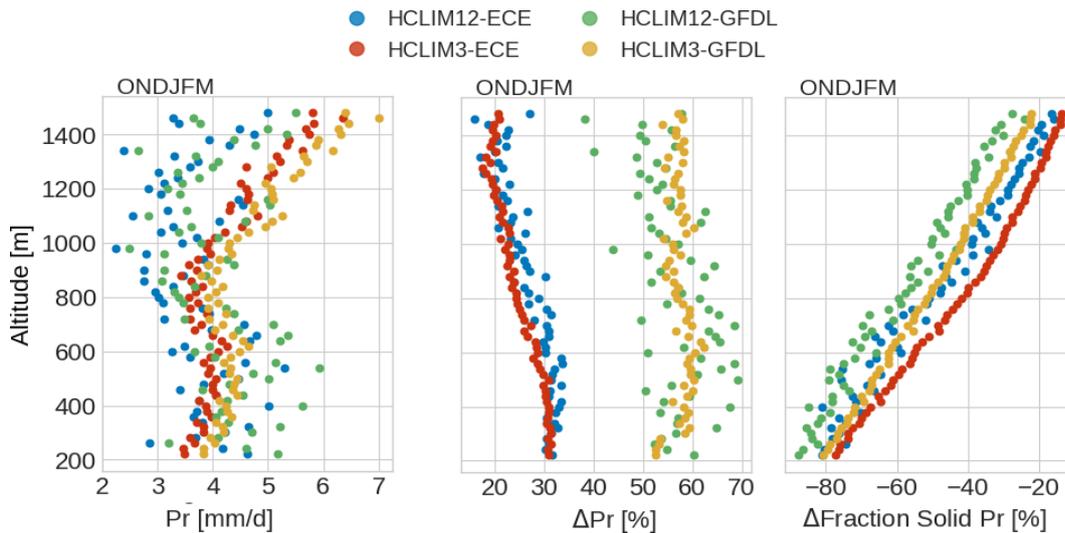


Figure 11. Daily mean precipitation (x -axis) in the cold season (October–March; ONDJFM) over southern part of Scandes mountains (slightly larger area than the black box in Figure 10, including also the coasts in southern Norway) as a function of altitude (y -axis). Left: Daily mean total precipitation in CTRL experiments; middle: Changes in total precipitation by end-of-century, in percent; and right: Changes in fraction of solid precipitation (to the total precipitation) by end-of-century, in percent. Model data has been binned into 20 m altitude (vertical) segments starting at 200 m and ending at 1500 m. Each dot represents the mean value in each of these segments. All model data have been kept at original grid resolution.

The shift from precipitation falling as snow to rain has been shown to cause a strengthening of rain shadows (increased drying ratio) (Pavelsky et al. 2012). Taking the average of grid points as a function of distance to the ridge-line of the Scandes shows that this is seen also in the HCLIM responses (less increase of total precipitation on leeward side). It further reveals that the smaller reduction in solid precipitation in HCLIM3 is mainly constrained to the mountain ridge and windward side (i.e., the west side due to predominantly westerly winds) of the mountains (Figure 12). Although Figure 12 only shows the EC-Earth forced simulations, the same response pattern is also seen in GFDL forced runs (however in HCLIM3-GFDL there is even a 10-15% increase in solid precipitation at and near the mountain ridge).

The reasons for the different responses in HCLIM12 and HCLIM3 are not entirely clear and still under investigation. A major part of the difference is likely due to an enhanced warming over topography in HCLIM12 (of around 1°C in DJF and to a lesser extent also in MAM and SON), somewhat more pronounced on the western slopes. Other factors could be the better resolved topography, especially the topographic variability that is pronounced on western coast of Norway, but also differences in model physics, e.g., microphysics. For example, even if HCLIM12 is colder than HCLIM3 in the present climate over the Scandes, HCLIM3 still has a larger fraction of solid precipitation, especially over the western slopes (not shown).

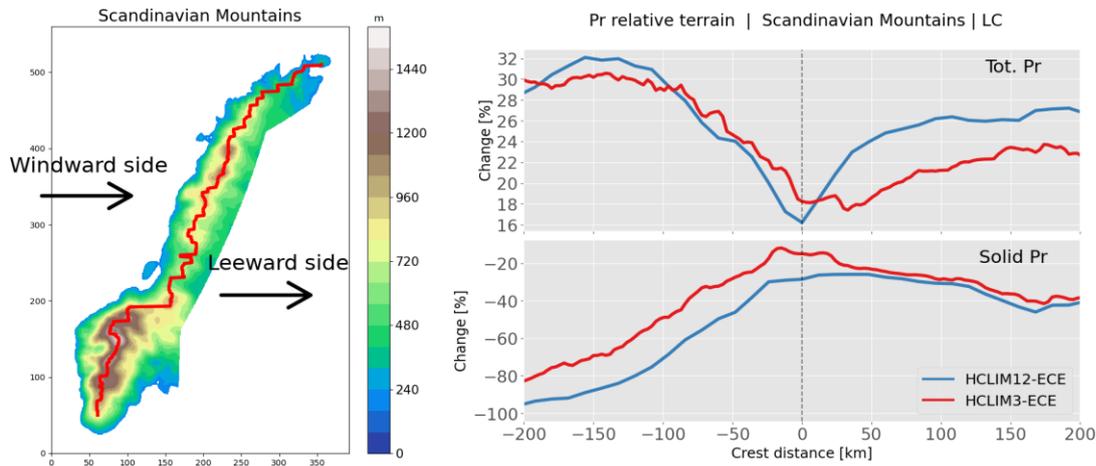


Figure 12. Left: map of Scandes mountains and its orography (in meters, smoothed field from HCLIM3). Right: Percentage change in daily mean total precipitation (top panel) and solid precipitation (bottom panel). For each latitude the highest point in smoothed HCLIM3 orography is identified. The red line in the map shows this "ridge" of highest points. Averages are then calculated for points on either side of the ridge ("windward" and "leeward" respectively). Results are only shown for HCLIM12 and HCLIM3 forced by EC-Earth. Note the different scales on the y-axes.

Winds

With more finely resolved land-surface heterogeneity stemming from complex topography and coast lines among other factors, high-resolution models are more suited to represent not only spatial variations in precipitation as discussed above but also improved near-surface winds and phenomena such as mountain-valley circulations or down-slope wind-storms (e.g., Jin et al., 2016; Junquas et al., 2018). In Figure 13 (top panel) we compare daily maximum 10-meter wind speed, over Fennoscandinavian land areas, between HCLIM12 and HCLIM3 as simulated in the GCM forced historical experiments, separating low-lying grid points (below 200 m) and grid points in the Scandes mountains at altitudes above 400 m. It is clear that outside of complex terrain the RCM and CPRCM models show similar wind speeds for all quantiles. However, at higher altitudes there is a shift to higher wind speeds in HCLIM3 and the differences become relatively larger for higher quantiles; and is also more pronounced in DJF than in JJA which is the season with the observed climatologically strongest winds. We further used measurements from stations located over Sweden (operated by SMHI) to evaluate the daily mean wind speed in the simulations forced by ERA-Interim, choosing the closest grid point in the model to the station locations (bottom panels in Figure 13). The probability distributions (PDFs) show differences between HCLIM and observations, for example an overestimation of days with 1-5 m/s, but also that HCLIM3 (labelled "AROME" in PDF figure legends, HCLIM12 is labelled "ALADIN") is much closer to observed probabilities for wind speeds above 5 m/s when comparing to stations located in the mountains. This provides some evidence for added value in HCLIM3 compared to HCLIM12 in simulating winds over complex terrain.

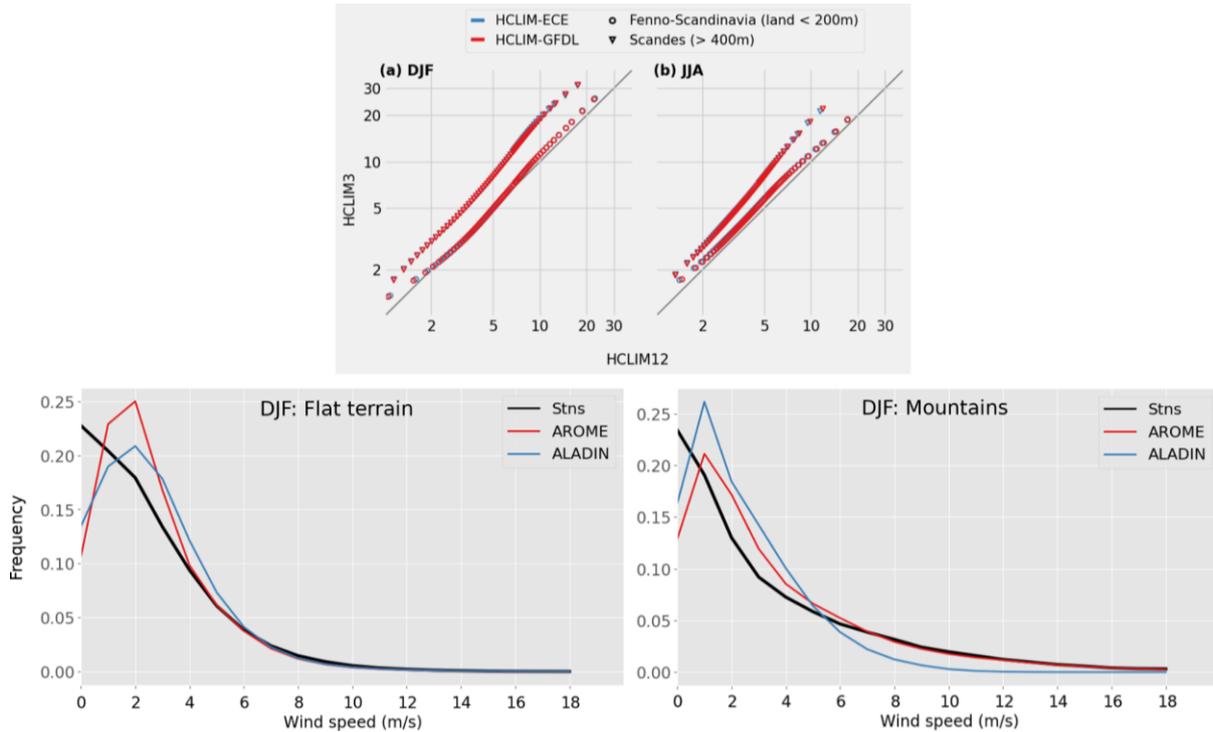


Figure 13. Top: Quantile-quantile plot of daily maximum 10m wind speed in HCLIM12 (x-axis) HCLIM3 (y-axis) forced by EC-Earth (blue markers) and GFDL-CM3 (red markers). Circles represent ocean and land grid points below 200 m, and triangles represent grid points above 400 m in the Scandinavian mountains. Units in m/s. Bottom: PDFs of DJF 10 meter wind speeds in HCLIM3 (red) and HCLIM12 (blue) compared to station data (black) over Sweden 1998-2018. Left panel considers stations (and corresponding model grid points) outside topography while the right panel is for locations in the Swedish mountains. HCLIM3 is here labelled “AROME” and HCLIM12 “ALADIN”.

The projected changes in seasonal average wind speeds over the Scandes are primarily governed by changes in the large-scale circulation and therefore tends, in the HCLIM scenarios shown here, to be largely influenced by the forcing GCM. Figure 14 shows the DJF and JJA 95th percentile of daily maximum 10m wind speed as a function of altitude over the southern part of Scandes mountains, for the historic period and the changes by end-of-century in RCP8.5. The larger wind speeds in HCLIM3 are evident for the present climate, increasingly larger with height. The magnitude of the climate change signal is less than 10% in all simulations, and even the sign of change is not consistent across simulations in DJF. At least from this analysis there are not any clear signs of potential added value in HCLIM3, i.e. systematic differences compared to HCLIM12, in the future response for wind speed in Fenno-Scandinavia. Other approaches, for example stratifying further using wind directions or selecting specific local or regional wind phenomena such as down-slope storms, could still potentially reveal different responses in HCLIM3 than in HCLIM12, a matter of future studies.

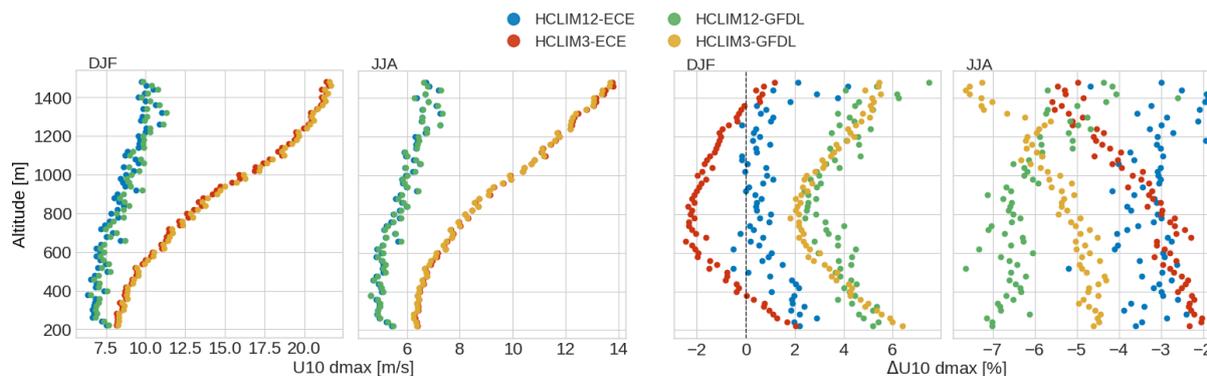


Figure 14. DJF and JJA 95th percentile of daily maximum wind speeds (x-axis) as a function of altitude (y-axis) in HCLIM3 and HCLIM12 RCP8.5 scenarios. Left two panels show wind speeds in the historical period, and the right two panels show the change, in percent, by 2080-2100. The 95th percentile was calculated per grid point, and the binning of data is the same as in Figure 11 with each dot representing the mean value in each of the vertical segments. All model data have been kept at original grid resolution.

Publication

Lind, P., Belušić, D., Christensen, O. B., et al, 2020: Benefits and added value of convection-permitting climate modelling over Fenno-Scandinavia. *Clim. Dyn.*, 55, 1893–1912, doi: 10.1007/s00382-020-05359-3.

3.2.3 Activity AV3: Added value of hourly precipitation extremes (KNMI)

Introduction

Convection permitting climate models have shown to vastly improve hourly rainfall statistics, for instance connected to extremes and the diurnal cycle of convective rain as investigated in WP3. Despite better present-day climate statistics, it is not so clear whether this translates to more reliable climate change projections as compared to RCM projections for two reasons. First, there still is limited understanding how the present-day observables link physically to future changes. Second, these CPM simulations are typically still rather short, with typically 10-30 years length. At these time scales the role of internal variability is large, which poses inherent problems with signal-to-noise (Aalbers et al. 2018).

Here, we tackle these issues using a scaling framework in which the dependency of hourly precipitation on absolute humidity and relative humidity is analysed. To improve signal-to-noise spatial pooling of results is used.

Scaling and added value

To investigate the added value of the CPMs compared to the RCMs, we analysed hourly rainfall in two areas: The Netherlands (~30 stations, almost complete data) and Southern France (~60 stations, approximately 50% data). Only CPM runs driven by reanalysis data are used, and runs that have data in both areas, giving in total 5 CPM simulations (two by HCLIM-AROME, UKMO, COSMO and

AROME41). We compared results of the CPMs to seven state of the art RCMs (CLMcom, HadRM3, RACMO2, RCA4, HIRHAM5, REMO, and ALADIN).

Scaling has been computed from pairing hourly precipitation to dew point temperatures from 4 hours earlier (Lenderink et al. 2011). The paired data is binned in dew point temperature bins, and different percentiles of the distribution conditioned on wet hours are computed. (We note that the same analysis has also been done for unconditional percentiles, giving similar results in most aspects.) Figure 15 shows the scaling based on dew point temperature, where we also subdivided into low and high relative humidity based on the dew point depression.

The CPMs are generally rather close to the observations; and capture the observed behaviour generally between one and two times the Clausius-Clapeyron (CC) relation in the observations. Most CPMs tend to slightly overestimate the 99th percentile, partly as the results of a tendency to underestimate the frequency of wet events. Most RCMs substantially underestimate extreme hourly rainfall. Also, the different models show more diverging behaviour (in comparison to the CPMs) and a number of RCMs strongly underestimate the dew point temperature dependency. The latter behaviour is most apparent for low relative humidity.

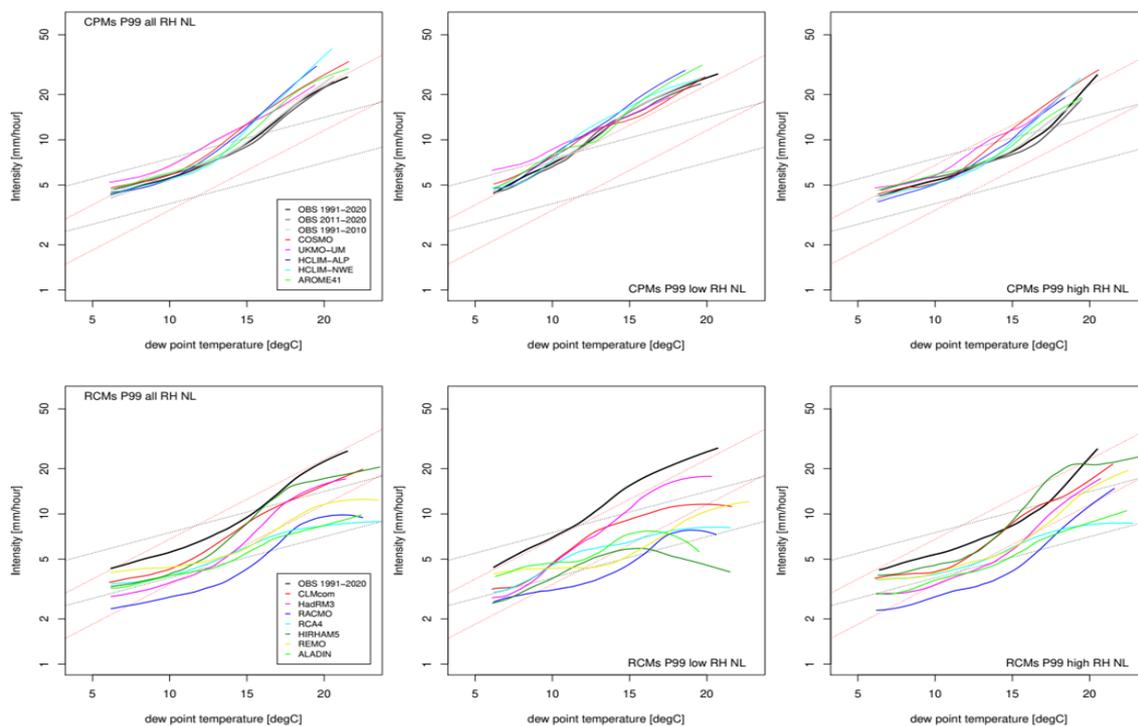


Figure 15. Scaling of the 99th percentile of hourly precipitation (conditional on wet hours) in NL based on, from left to right, dew point temperature, dew point temperature for hours with low relative humidity (dew point depression above 6 degrees), dew point temperature for hours with high relative humidity (below 3 degrees). Upper panels are CPMs, whereas lower panels are RCMs. Grey stippled lines are CC dependencies (6.5 % per degree) and red stippled lines 2CC dependencies (13 % per degree).

Further analysis shows that in particular the dependency on relative humidity, measured by the dew point depression, is rather different between most RCMs and the CPMs. Most RCMs (except HadRM3) display a tendency towards decreasing extremes with lower relative humidity. The CPMs, as well as

the observations, tend to display the opposite behaviour. Lower relative humidity tends to be associated with stronger precipitation extremes.

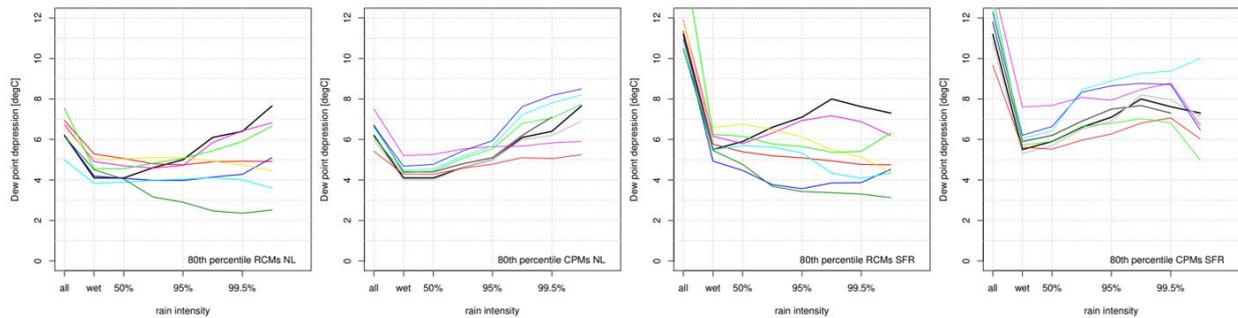


Figure 16. The 80th percentile of the distribution of dew point depression of the precipitation (dew point) temperature pairs, selected for (at the x-axis from left to right) all events, wet events (threshold) and exceeding different percentiles of the precipitation distributions (50, 90, 95, 99, 99.5% conditional). For instance, 80% of the rainfall events exceeding the 99.9th percentile have a dew point depression below ~8 degrees in The Netherlands.

While the CPM ensemble shows much better correspondence to the observations than the RCM ensemble, the CPMs are not without faults either. Most CPMs tend to be too dry and display too low relative humidity. This likely affects the rainfall distribution too. For the Netherlands low relative humidity tends to enhance rainfall extremes (given same absolute humidity; see 2nd and 3rd panel in Figure 15). However, for the climatological dryer climate in southern France, the relation between relative humidity and extreme rainfall is likely broken for the most extreme events (compare 2nd and 4th panel in Figure 16). So, in that case a dry bias of the model tends to suppress extremes. This also appears to be reflected in a too small difference between the intensity of precipitation extremes in SFRA and NL in all CPMs; in the observations ~50% more intense in SFRA compared to NL whereas in the CPMs this is only 10-30%.

Scaling and future response

While increases in precipitation extremes are generally explained by the humidity increases, it is strongly debated whether the above scaling rates (usually called apparent scaling rates) between 6.5 and 13 % per degree can be used as a simple predictor of future changes to precipitation extremes. It is often argued that apparent scaling rates lead to an overprediction (Ban et al. 2015, 2020). We tested this in a small ensemble of three model experiment using the CPM HCLIM-AROME: a surrogate warming experiment (E_{SW}) using 2 degrees uniform warming and unchanged relative humidity, a pseudo global warming (PGW) approach based on a full 3D perturbed climate derived from ECEARTH at 1.5 degree global warming (E_{PGW}), and a full GCM driven experiment 2089-2099 compared to 1995-2005 with approximately 3 degrees global warming (E_{GCM}). We note here that a similar analysis based on COSMO is performed in the next section.

Comparison between apparent scaling rates and the climate change response rate (change in hourly rainfall extremes scaled by the increase in dew point temperature representative for moderate to extreme events) shows that apparent scaling rates indeed overestimate the climate response rate (Figure 17). A main cause of this overestimation in E_{PGW} and E_{GCM} is the decrease in number of rain events, which is tied to the increased influence of high-pressure systems in the future in the experiments (see activity SME1; section 3.4.1). However, for the tail behaviour this is less the case,

and climatic response of the rainfall extremes appears to converge to values close to or slightly below the apparent scaling rates. We note that the climate change response is derived from unconditional percentiles to avoid statistical artefacts (Schär et al. 2016).

In the above results we pooled data from a large area, mostly covering the full CPM domain and low altitude area (below 400 m). How does this translate to subregions of intermediate size, such as the size of the Benelux? Results are broadly similar, yet with substantial variability. A strong overestimation of changes in moderate extreme events is obtained from apparent scaling (left panel in Figure 18). But for the most extreme events apparent scaling and climatic scaling are similar in magnitude and both scatter round the 2CC rate. Comparing different subregions there is generally no correlation between apparent and climatic scaling, but we note also that in the simple surrogate warming experiment (ESW) there is some correlation (open dots in left and middle figure).

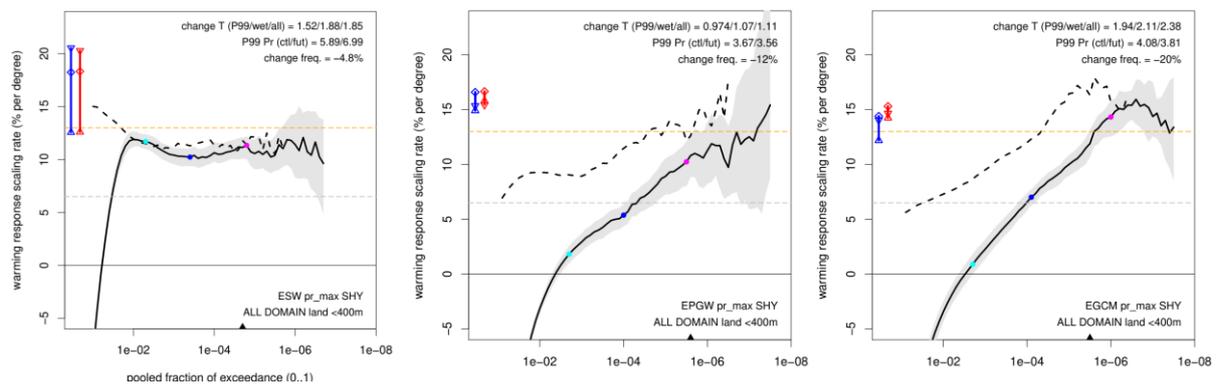


Figure 17. Climate change response (scaled by the dew point temperature change for heavy rain events) as a function of the “pooled fraction of exceedance” for simulation domain low altitude area (land points below 400 m). Solid lines are the unconditioned change rates, whereas the dashed lines are conditioned on the occurrence of rain (wet threshold is 0.1 mm hour⁻¹). The change in the frequency of rain is shown at the top right. Grey and orange lines denote 1 and 2 times the CC rate respectively. The blue (control) and red (future) lines and symbols on the left denote the apparent scaling rates with different symbols corresponding to a different dew point range of the fit. The colour dots indicate the positions where rainfall amount for the present-day simulation are 10 (cyan), 30 (blue) and 60 (magenta) mm hour⁻¹, respectively, and the triangle at the bottom marks the top 1000 rainfall hours.

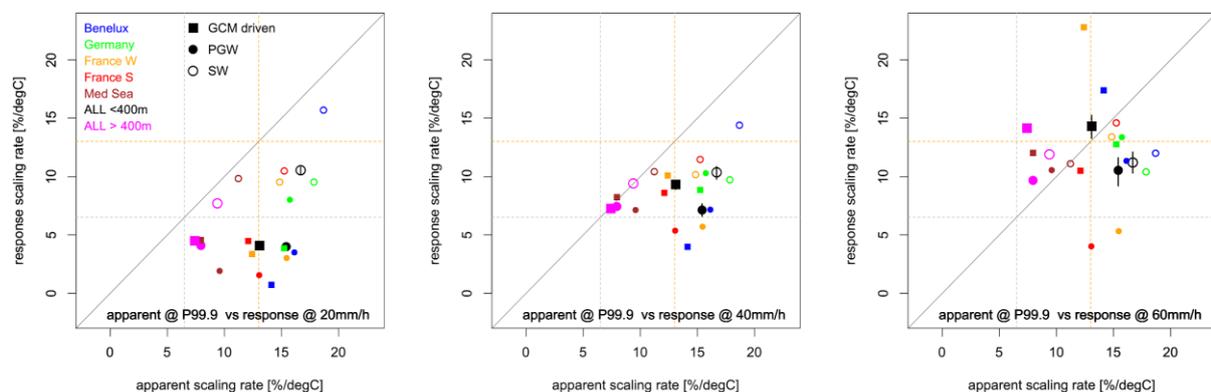


Figure 18. Climate response rate compared to apparent scaling rates of hourly precipitation extremes for typical intensities of 20 (left), 40 (centre), and 60 (right) mm hour⁻¹ in the present-day simulation for a number of sub-regions and the three experiments: E_{SW} , E_{PGW} and E_{GCM} . Regions are low altitude land areas only, and “ALL” is the full model domain covering all sub-areas (see Lenderink et al. 2021 for definition areas).

Perspective

CPMs display a much better (than RCMs) consistency between humidity variables and precipitation with much better correspondence to observations. However, the general climatology still needs to be improved to fully make use of their potential. This may for instance imply that improvements in the soil hydrology are needed.

Although apparent scaling rates are no straightforward predictor of climate change response of hourly extremes, we provide here evidence that apparent scaling rates could inform us on potential response of the most extreme events to climate change (Lenderink et al. 2018, 2021). These are also the events that are most important in terms of impacts, such as the devastating floods over Germany and Benelux summer 2021 showed. Flash floods occurring in several hours' time caused extreme damage and loss of lives in the Ahr and Erft catchments. Thus, scaling does provide indirect support for the trustworthiness of models, such as the CPMs, as well as a potential indicator of the tail behaviour of the extremes.

Obviously, the short simulation periods of the CPMs impose challenges in terms of signal-to-noise. By extensive pooling it appears that still reasonable estimates of the tail behaviour can be obtained. However, longer simulations and more analysis is needed here. The user however should be aware that the plain statistics even over reasonably sized areas (as large as The Benelux) are quite uncertain, in particular for the most extreme events (beyond e.g., 50 mm hour⁻¹). In addition, in order to assess the robustness of these results, we need to further explore the other CPMs simulations. This is currently underway. In the next section an analysis of COSMO results is presented.

Publications

Lenderink, G., H. de Vries, H. J. Fowler, R. Barbero, B. van Uft, 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 Math. Phys. Eng. Sci.*, **379**, 20190544, <https://doi.org/10.1098/rsta.2019.0544>.

In draft:

Lenderink et al. [Evaluation of hourly precipitation in convection permitting models using scaling: are they better than parameterized models?](#)

3.2.4 Activity AV4: (Sub)hourly precipitation extremes (ETHZ)

Introduction

Recent studies have shown the inadequacy of low-resolution models using convection parameterizations to simulate short-term mean and heavy precipitation. They strongly overestimate summer mean hourly precipitation frequency together with a compensating underestimation of mean and heavy hourly precipitation intensity (e.g., Ban et al. 2021). During the last decade, a tremendous effort has become evident to overcome this challenge by refining the horizontal grid spacing of global and regional climate models to less than 4 km (Stevens et al. 2019, Schär et al. 2020). This development enables the explicit representation of deep convective clouds, without semi-empirical

parameterization schemes. It yields a significantly improved simulation of the diurnal cycle and allows a more adequate representation of clouds, precipitation systems and winds (Ban et al. 2015, Belušić et al. 2018, Hentgen et al. 2019, Ban et al. 2020, Ban et al. 2021, Pichelli et al. 2021).

This study aims to assess the performance of convection-resolving models in simulating even shorter precipitation events at sub-hourly timescales. Such extreme events have extremely large temporal and spatial variations and can cause huge damage at the local scale, such as flooding, landslides, erosion and water damage. For example, measurements indicated that rain amounted to 41.3 mm in 10 min during the June 11th 2018 flash flood that occurred in Lausanne (Switzerland). This amount corresponds to more than a third of the climatological monthly mean. More recently, 31.1 mm of rain fell in 10 min over the village of Waldegg (Zürich, Switzerland) on July 13th 2021. It is therefore essential to simulate such sub-hourly precipitation events accurately to assess how they might evolve in response to global warming. Due to the inadequacy of climate models that rely on convection parameterization to simulate such events, that effort was until recently impossible.

Studies based on observations (e.g. Lenderink and van Meijgaard 2008, Westra et al. 2014) and convection-resolving simulations (e.g. Kendon et al. 2014, Ban et al. 2015, Lenderink et al. 2019) have highlighted that changes in hourly precipitation events tend to follow or even exceed the Clausius-Clapeyron relationship (see also previous section; section 3.2.3), which corresponds to an increase of 6-7% per degree warming. In this study, we consider 6-min precipitation to assess their changes in response to global warming. Such events have been shown to be well represented in km-scale simulations over the Netherlands (Loriaux et al. 2013), the UK (Chan et al. 2016) and Catalonia (Meredith et al. 2020). 6-min precipitation data is extracted from the COSMO-crCLIM regional climate model (Schär et al. 2020). The simulations were conducted at 2 km resolution over a greater Alpine region, using an intermediate nest at 12 km resolution over Europe. The simulations were driven by the Max-Planck Institute global climate model under present (1996-2006) and RCP8.5 future (2090-2100) climate conditions (see Table 1).

Scaling of sub-hourly precipitation extremes

Compared to the Swiss automatic meteorological surface network (SwissMetNet, formerly ANETZ), as used by Ban et al. (2014), results show that the model is able to capture the statistical values of extreme precipitation events for aggregation times as short as 10 minutes over the Alps (Figure 19). The model credibly simulates all percentiles of extreme precipitation for 10-min and 30-min accumulations, particularly the very extreme percentiles (99.99% and 99.999%), while the 99% and 99.9% are systematically overestimated. The hourly accumulated precipitation events are underestimated compared to observations. This evaluation increases trustworthiness in climate projections of extreme precipitation events as short as 10-min.

Figure 20 shows projected changes in extreme precipitation intensity at sub-hourly to hourly timescales over several European regions under the RCP8.5 greenhouse-gas warming. Extreme precipitation events are projected to increase with warming at a scaling rate that increases with the intensity of the events and that is close to the Clausius-Clapeyron rate for the most extreme events. The scaling rates never exceed the Clausius-Clapeyron rate for any of the regions considered. This result is similar to previous studies based on hourly extreme precipitation events (Ban et al. 2015). Over all regions considered, there is a clear decrease in the scaling rate from hourly to 6-min precipitation accumulation. This indicates that very short precipitation events will increase but at a

rate that is smaller than the Clausius-Clapeyron rate. This is particularly visible in the analysis of the most extreme percentile (99.999%). It indicates that dynamical changes in such extreme short-term events may not play an important role in scaling changes in precipitation intensity. Over orographic regions such as the Alps, the scaling does not depend on the precipitation accumulation periods.

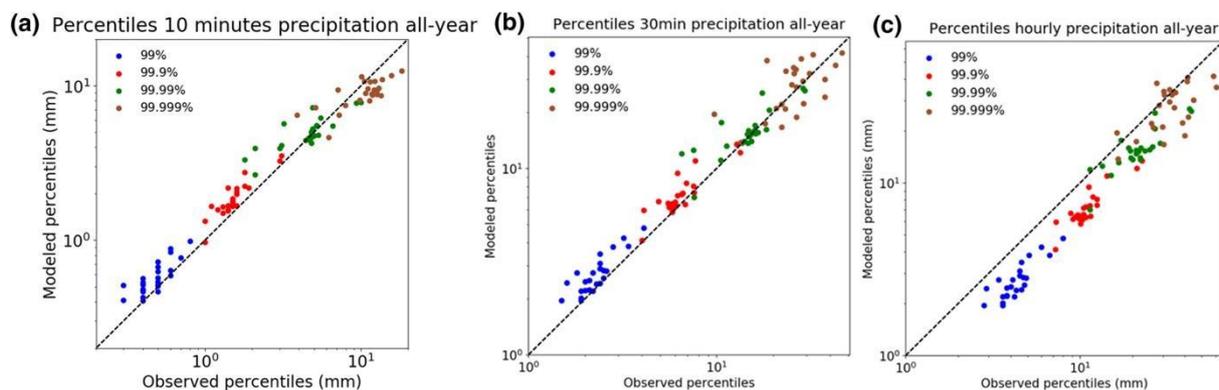


Figure 19. Evaluation of modelled percentiles of precipitation amount (mm) in the present climate simulation against station observations for the all-year percentiles. Each dot corresponds to one Alpine station. The different colours represent different percentiles of intensities (including dry periods). Panels (a–c) correspond to precipitation amounts over 10-min, 30-min, and 1-h periods, respectively. The units are the absolute rain amount for each of the presented accumulation periods measured in mm.

The projected increase in extreme precipitation occurs with an overall decrease in summer mean precipitation over most of central and southern Europe (Figure 21a,b). Convective available potential energy (CAPE) increases largely (Figure 21c). CAPE is associated with the maximum velocity of rising air in a convective cloud. An increase in CAPE therefore signifies a more favourable environment for the development of severe convective events. Convective inhibition (CIN) also increases with temperature, but to a smaller extent (Figure 21). As CIN represent the amount of energy that is needed to start convection; and increase in CIN tends to delay/suppress convection and is often associated with a stronger built up of convective instability (CAPE) and therefore results in more severe convection. This result is in line with less frequent and more intense precipitation events projected in the future scenario (Figure 21e,g).

Perspective

This study represents one of the first model-based decadal projections of sub-hourly precipitation extremes. It shows the adequacy of convection-resolving models to analyse and project extreme precipitation events at the sub-hourly timescale. By using the temperature increase in the lower part of the cloud (700 hPa) to normalise the climate change response, the results suggest that the intensity of sub-hourly precipitation extremes will increase largely with temperature increase, at a rate that is close to the Clausius-Clapeyron rate. This is in line with earlier results on hourly timescales over the Alps (e.g., Ban et al. 2015, 2020). However, the rate of increase does not exceed the Clausius-Clapeyron rate as found by previous studies performed over other regions (e.g. Lenderink and Meijgaard 2008, Loriaux et al. 2013, Kendon et al. 2014), Lenderink et al. 2019, 2021. This result may depend on the choice of the temperature variable used for the scaling (e.g., 700 hPa, 2 m temperature or surface dew point temperature). It also indicates that the increase in short-term precipitation intensity by dynamical or microphysical feedbacks may be negligible in these simulations for the regions considered, and highlights that the results might be model dependent, as also suggested by

Lenderink et al. 2022 (section 3.2.3). In addition, the relative temperature large lapse rate changes in these runs may also dampen dynamical feedbacks leading to super CC behaviour (Loriaux et al. 2013). Repeating such study with the multi-model ensemble provided through EUCP would give valuable information regarding future changes in extreme precipitation events as short as 10-min.

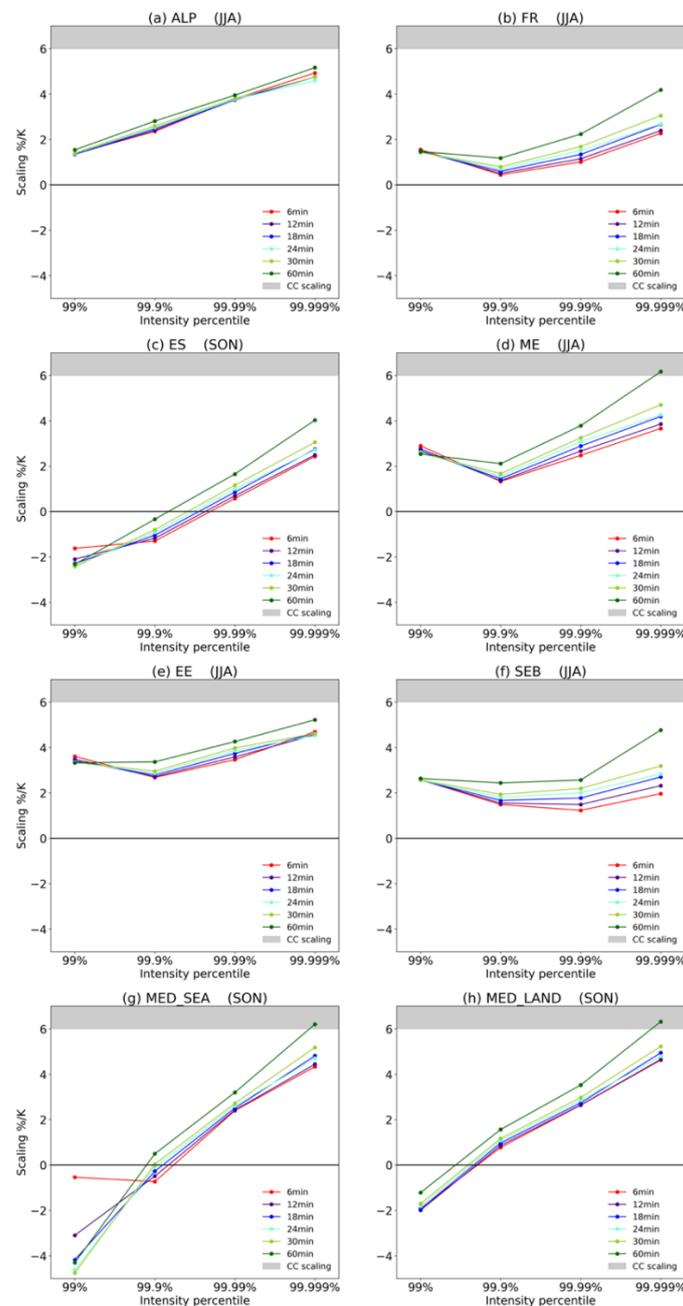


Figure 20. Climate change projections of short-term heavy precipitation events: scaling of extreme precipitation percentiles for different accumulation periods (different colours) across European regions: (a) Alps, (b) France, (c) Eastern Spain, (d) Mid-Europe, (e) Eastern Europe, (f) South-East Britain, (g) Mediterranean Sea, (h) Mediterranean Land. The scaling expresses the difference between future and current climate simulations in % per K of warming. The grey band highlights the range of the Clausius- Clapeyron scaling (6 to 7 %/K). The percentiles are computed for all-event percentiles (wet and dry events) and normalised with the temperature increase (at 700 hPa) of the season with the most intense precipitation (JJA for all regions except SON for Eastern Spain and the Mediterranean).

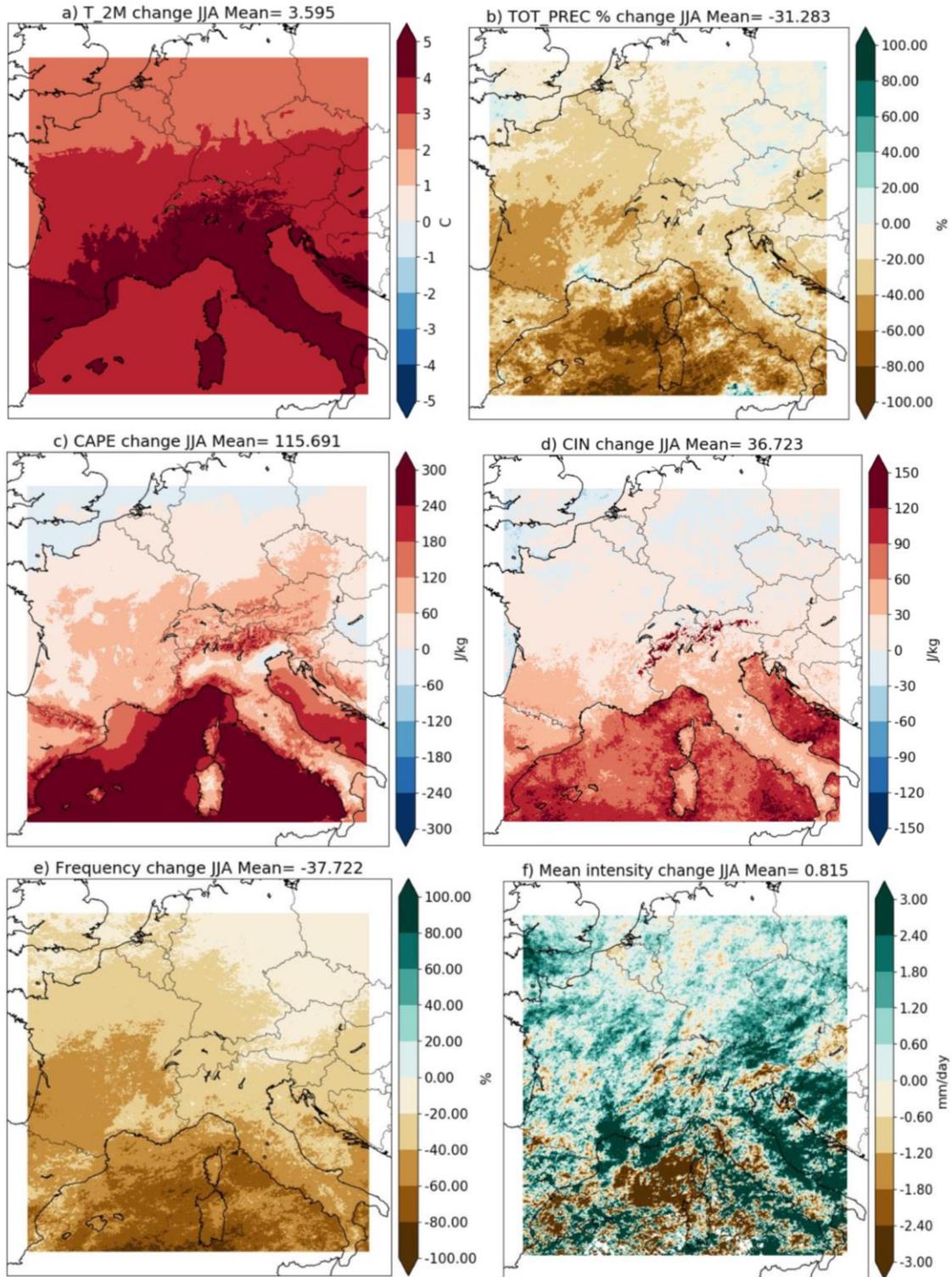


Figure 21. Projected summer changes in (a) surface temperature (°C), (b) total precipitation (%), (c) convective available potential energy (J/kg, at 12 UTC), (d) convective inhibition (J/kg, at 12 UTC), (e) daily precipitation frequency (%), and (f) daily precipitation intensity (mm/day). Mean intensity refers to the average precipitation across all days where total precipitation exceeds more than 1 mm/day. The values on top of each panel correspond to the regional mean changes.

Publication:

Vergara-Temprado, J., N. Ban and C. Schär (2021). Extreme sub-hourly precipitation intensities scale close to the Clausius-Clapeyron rate over Europe. *Geophysical Research Letters*, 48, e2020GL089506. <https://doi.org/10.1029/2020GL089506>

3.3 Signal to noise and regional predictability (SNP)

3.3.1 Activity SNP1, predictability and signal to noise of fine-scale information in a regional SMILE (KNMI)

Introduction

The added value (AV) of high-resolution models in producing more reliable *climate change* information and/or better projections of the future climate is a complex issue (Feser et al. 2011; Di Luca et al. 2015; Giorgi 2019; Lloyd et al. 2021). Obviously, the fact that local processes are better resolved in high-resolution climate models increases confidence in their ability to make more reliable projections of future changes. Not disputing this view, one may, however, also question whether this added physical realism (and usually proven better present-day climatology) is a sufficient reason to spent computational resources on downscaling experiments as compared to, for example, producing larger ensembles of global climate model simulations, or trying to improve those models (Nishant and Sherwood 2021). We mention two primary reasons for this point of view. First, a regional model will inherit most of the large-scale atmospheric flow features from the driving global climate model and the reliability of a downscaled projection is therefore limited by the realism of, or uncertainty in, in these large-scale conditions (Maher et al. 2021). Second, internal climate variability tends to increase at smaller scales, so that the enhanced spatial detail may easily get lost in the small-scale unpredictable noise (Aalbers et al. 2018).

To study these internal climatic variations in relation to the systematic greenhouse gas induced forced climate change signal, so-called Single Model Initial Large Ensembles (SMILEs) are used. Apart from this signal-to-noise problem, SMILEs can also be used to study very rare extremes or changes in variability (Wood et al. 2021; van der Wiel et al. 2021; see also WP2 deliverable D2.4). A small number of downscaled regional SMILEs exists (von Trentini et al. 2020).

Here we tackle the question of AV in the context of a perfect model approach using a regional SMILE (RACMO2 embedded in ECEARTH; Aalbers et al. 2018). Our main goal is to learn in this environment what are good strategies to cope with the signal-to-noise problem, and what kind of information could be reliable in this respect, and how much AV we can expect from the fine-scale information. As part of coping with the signal-to-noise problem, we also present results using an ensemble produced within EUCP using a pseudo global warming (PGW) approach (Schär et al. 1996; Brogli et al. 2019). In this approach, the weather in control period is repeated under warmer/moisture conditions, using mean perturbations in temperature/humidity and large-scale flow derived from global climate simulations. This filters out a considerable part of the climate variations related to variability in the large-scale atmospheric flow as quantified by our analysis.

Analysis

Two analyses have been performed:

- First, we take a signal-to-noise standpoint, and investigate the internal variability (spread in the ensemble) and the systematic forced signal (mean of the ensemble response). A spatial filtering technique – a Gaussian filter with width of 10 RCM grid-points – is used to separate out large-scale and fine-scale features. Then, we evaluate the performance of individual members to predict the forced signal; and compare results from the standard ensemble (GCM driven) with the PGW ensemble.
- Second, we approach the problem in a prediction framework. A perfect model approach is used to evaluate how well a 30-year future climate state can be predicted, taking one ensemble member as the truth, and trying to predict the future state using the change information from the other members. Various options to produce high-resolution future states are explored to investigate the added value of the high-resolution modelling systems.

Our analysis is focused on the mean winter temperature and precipitation change between the period 1981-2010 (present-day) and 2071-2100 (future climate). Here, we document mainly the results for winter precipitation, which is characterized by high internal variability and can be considered as a worst case scenario in terms of signal to noise. Results for temperature can be found in the Appendix.

Forced signal and internal variability

The forced response shows the expected increase in mean winter precipitation over central parts of Europe of 10-20%, mostly as a result of the increase in water vapor of the atmosphere in a warming climate. In addition to this large-scale pattern, substantial fine-scale deviations are obtained, typically amounting up to 10-15% over regions with high topography. For instance, south of the Austrian Alps precipitation increases are up to 30%, as well as east of the Norwegian mountains. Although these fine-scale features in the systematic response pattern are substantial, they are generally small compared to the internal variability (compare the left panel in bottom row of Figure 22 to the other panels in bottom row). Internal variability at large scales, but also at fine scales is much larger than the systematic fine-scale change pattern. Results for temperature are more favourable, and although the internal variability is still considerable, the systematic fine-scale change patterns in temperature are relatively large (compared to internal variability) in most topographic areas.

The PGW approach is often employed with the motivation (but no actual proof) to obtain more reliable information for future changes in this signal-to-noise context. Indeed, the PGW ensemble has much better signal-to-noise characteristics. At large scales internal variability is now very small, and even at fine scales it is much smaller than in the original ensemble. At large scales, however, the PGW approach produces a small bias (2-5% higher) in the response in mean winter precipitation as compared to the standard ensemble. Nevertheless, it also gives a very reliable estimate of the fine-scale pattern in the response, with a pattern correlation of 0.92 with the original pattern from the standard ensemble (Figure 23). A similar result is obtained for temperature. The mean response is

somewhat biased at large scales, but pattern correlations of the fine-scale response between PGW and standard ensemble are very high.

One may question how much information on the systematic change can be obtained from a single member of the ensemble, and whether in that case having the fine-scale information adds value. With the forced signal from the standard ensemble as a reference, all PGW members are better at predicting this change pattern compared to the standard ensemble members (Figure 23, left top). This also holds for the fine-scale response pattern (middle pattern); and is obviously more true taking the PGW forced response as a reference (right panel). Comparing the large-scale response with the full response pattern (containing large-scale as well as fine-scale) in the standard ensemble no improvement is found by adding the fine-scale information (bottom left, panel), but some improvements are obtained by using the ensemble mean large-scale response (cyan dot). In the PGW ensemble, adding fine-scale information slightly improves the results, at least with somewhat higher spatial correlation (from 0.9 to 0.92). Improvements are again (obviously) larger taking the PGW forced response as a reference.

Temperature results are more positive; and adding fine-scale information generally improves the prediction of the forced change pattern. In this case, the PGW approach has about the same (or slightly better skill) than the standard approach. Adding the fine-scale information in the PGW again improves the skill more than in the standard ensemble.

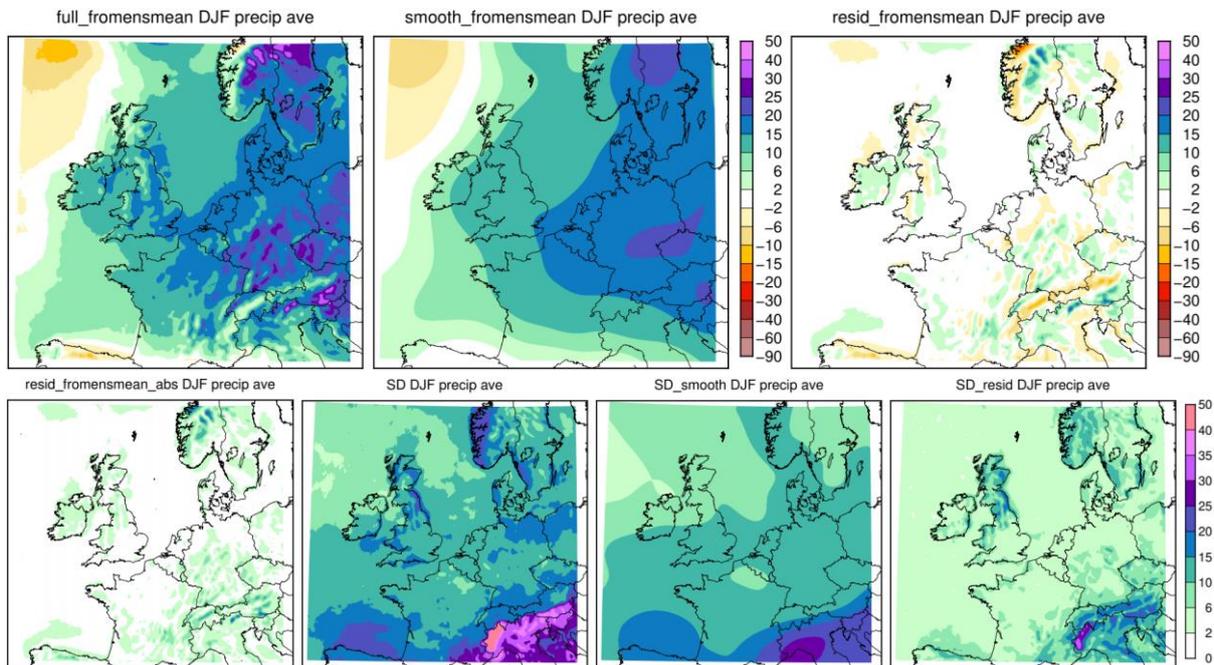


Figure 22. Change (%) in mean winter precipitation from the SMILE (16 member RACMO ensemble). Upper panels: forced response in precipitation (mean of the 16 members) with the from left to right, full field, filtered field and fine-scale field. Lower panels: absolute value of fine-scale forced response (left), in comparison with internal variability (2 times standard deviation between the members, %) in full, filtered and fine-scale field (2nd to 4th panel).

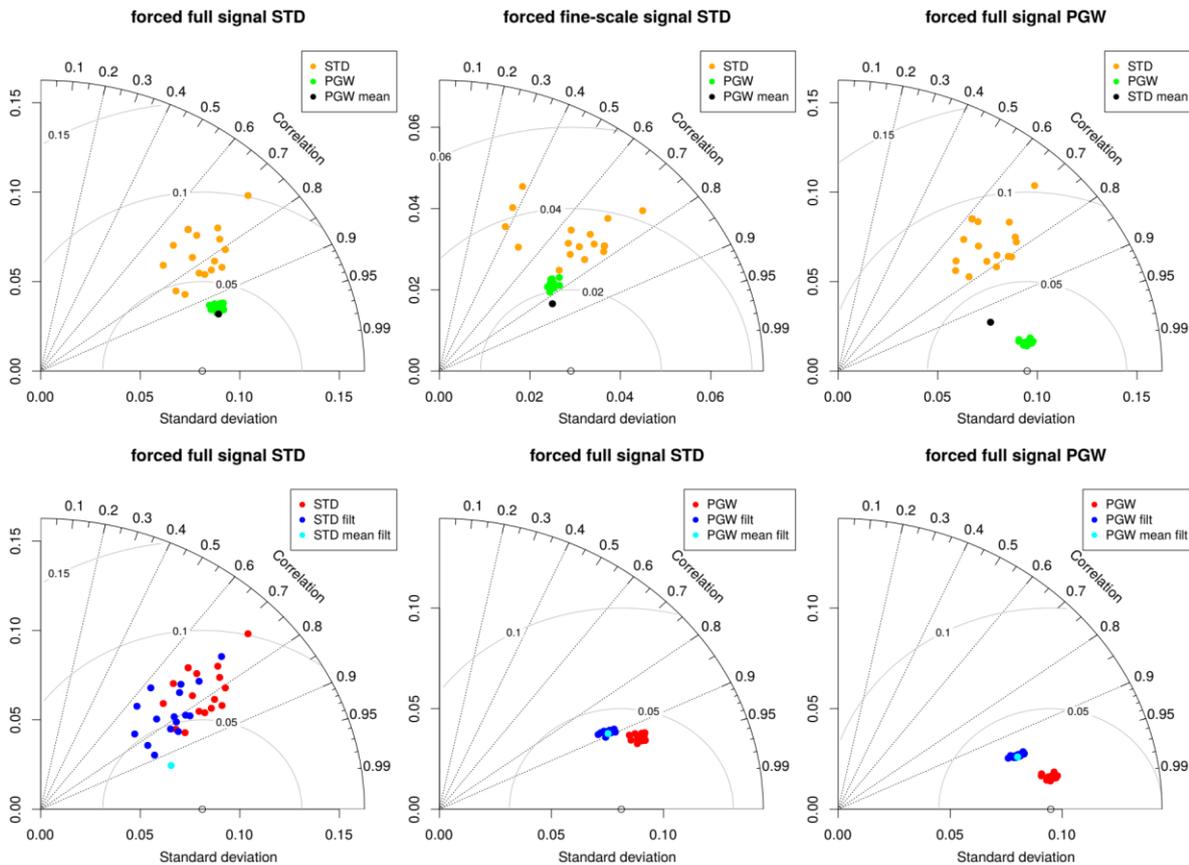


Figure 23. Taylor diagrams showing correlation and standard error of mean winter precipitation change between the forced response (reference) and different projections (e.g., individual members of the ensemble). The reference is given on top; e.g., in the left top panel each member of the STD (orange) and PGW (green) ensemble is compared to the full forced response of the STD ensemble (note the comparison to the PGW forced response in the right-hand plots). Upper plots compare individual members of the PGW (green) and STD (orange) members to the forced response; lower plots compare filtered (blue) and full resolution, unfiltered (red) results.

Prediction of the future climate state

We further investigate the potential to predict an actual future climate state; a climate state is defined as a realisation of a 30-year climate, in our case a member of the ensemble. Here, we used a perfect model approach in which one member i is considered to be the truth, and we aim to predict the future state of member i using information from the other members j . We also assume that the present-day climatic state of member i is known. Three prediction pairs are used, in which the added value of fine-scale change information (derived from member j) is assessed. The first pair starts from the climate state of member i , and adds the change – large-scale only as well as the full response including the fine scale – from member j . The second pair starts from the ensemble mean control state, adding the same changes. This avoids an error in the prediction because the bias in the control state (that is, the deviation in the control climate state of i as compared to the ensemble mean) will carry over to the predictions of the future state. The third, prediction pair uses information from the large-scale change from member i ; that is, assumes a perfect knowledge of large-scale change. Added value is established by comparing this prediction, with a prediction in which the fine-scale information of member j is added on top of the perfect large-scale change. Finally, we also compared with a direct prediction, taking the future climate state of member j to be the truth. As a measure of skill we used here the

spatial mean of the relative precipitation error (in fraction) and absolute temperature difference between prediction and true state averaged across the full simulation domain.

For precipitation shown in Figure 24, no added value of the fine-scale information is found, comparing neighbouring red (including fine-scale) and blue bars (excluding fine-scale information). This is true for the full domain (shown), but it also for a subdomain around the Alpine region where we might expect more added value. In general, the PGW based predictions are better than the standard ensemble based predictions. For temperature this is different. Adding the fine-scale information generally improves the predictions, for the full domain () and even more pronounced for the Alpine region. In this case, the PGW approach is worse than the standard ensemble for the full domain (caused by a mean bias in the response over the domain), but is still better focusing on the Alpine region.

Finally, the direct prediction from the standard ensemble is the best prediction apart from the predictions using perfect knowledge of the large-scale change. This may look surprising, but actually this is well understandable. Since we assumed a perfect model, each member is the best prediction of another member acknowledging that the internal variability component cannot be predicted. The two other approaches (first and second prediction set) can be rewritten as the direct approach with a correction based on the difference in two climate states for the present-day climate (in the first prediction set of members i and j). In an error perspective this correction term adds two error terms due to internal variability as compared to the direct prediction and therefore these predictions are worse than the direct prediction. We note however that the advantage of the direct prediction is only academic. In practice, the model will be biased and a bias correction term will therefore be applied. This bias correction term will generally involve comparing the control climate state of the model to the observed climate state. This will again introduce two error terms due to internal variability, as with the first prediction set.

Perspective

It is clearly established that in the systematic, greenhouse gas forced response patterns fine-scale information is present, both for temperature as well as precipitation. But, these systematic patterns may also be overwhelmed by internal variability looking at single model runs.

It is user dependent whether their decisions rely mostly on the forced systematic changes (therefore including the fine-scale information is relevant; and filtering out the internal variability) or whether they are mostly interested in a prediction perspective (therefore including the internal variability and comparing this to the systematic fine-scale changes). For mean winter precipitation with large-internal variability results showed that there is little added value from the fine-scale information in the latter prediction sense. However, for temperature, which is less affected by internal variability, there still is added value of the fine-scale response.

In general, the PGW approach appears promising to predict the systematic fine-scale response patterns, and these patterns are generally already contained in a single member. The amount of internal variability in these runs is (much) lower than in the free standard ensemble. However, this advantage should be compared to potential biases in the response introduced by the method (mostly at larger scales). For precipitation the balance is clearly dominated by the improved signal-to-noise ratio. For temperature, the advantage is still present but less outspoken. However, we also note that

these results do not necessarily carry over to other seasons; for instance, we noticed that the bias in the PGW approach in summer appeared to be too large to give improvements and that this bias was related to the rather small model domain we used for these simulations. Thus, we think that the PGW approach is promising, but more research on the method is definitely needed.

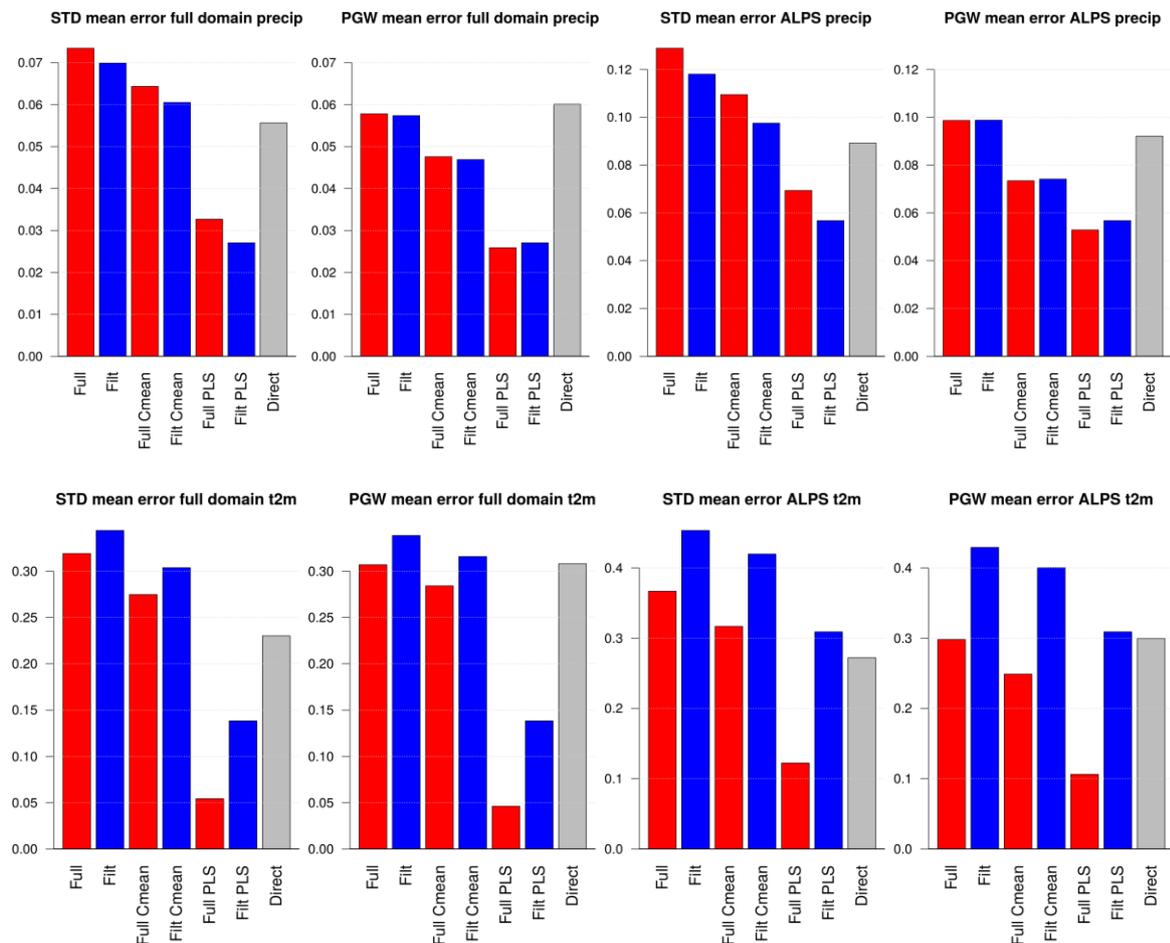


Figure 24. Absolute spatial mean of the relative error in winter-mean precipitation (upper panels, in fraction) and absolute error in temperature (lower panels, in degrees) across the full prediction matrix and the domain/Alpine region. Left two panels, mean over the domain, STD and PGW based predictions; right two panels, mean over the Alpine region. In red are predictions based on full changes, in blue predictions based on only large-scale changes.

In draft:

Lenderink et al. [A perfect model study on the feasibility of adding small-scale information in climate change projections](#)

3.3.2 Activity SNP2: Spatial pooling to improve signal-to-noise (KNMI)

Finally, we briefly mention work done by KNMI on spatial pooling. When changes in statistics are reasonably homogeneous over large areas improvements of signal-to-noise can be obtained from pooling of large areas. This can be advantageous for atmospheric small-scale phenomena, such as those related to convective showers, which are small-scale but not necessarily connected to specific locations. Results on spatial pooling of hourly extremes are discussed under activity AV3 in section 3.2.3.

3.3.3 Activity SNP3: Early emergence of the forced signal (UCPH/DMI)

Here, we discuss the emergence of the forced signal using the pattern scaling approach based on the results published in Matte et al. (2019). Climate models continue to exhibit large inter-model differences due to, among other things, differences in cloud parameterization schemes, resolution, physics, land-surface, water cycle representation, and sea ice treatment. Furthermore, climate models have systematic biases, which further complicates the extraction of useful climate change information. To provide such information, Santer et al. (1990) proposed using a pattern scaling approach. This approach assumes a linear relationship between patterns of regional climate change and the average global temperature change; and scales the grid-point change by the spatial average of the 2-meter temperature global change, e.g., the global forcing level. Scaled pattern as defined could be seen as the first emerging fingerprints of climate change pattern e.g., it shows an overall quite stable evolution of climate for some simple variable such as the 2-meter temperature and to some extent the mean precipitation (Matte et al., 2019). The approach has the considerable advantage of providing climate change information for time periods or emission scenarios for which no simulation is available (Lustenberger et al., 2014). Since Santer et al. (1990), pattern scaling has been widely used (Huntingford and Cox, 2000; Mitchell, 2003; Sanderson et al., 2011; Lustenberger et al., 2014; Tebaldi and Arblaster, 2014; Christensen et al., 2015, 2019). It is worth noting that one of the major conclusions of Mitchell (2003) is the necessity to use a large ensemble to achieve a sufficiently large change signal when compared with the inter-model spread [also called the signal-to-noise ratio (S/N)] to identify a robust signal.

Many coordinated experiments such as CMIP3, CMIP5, ENSEMBLES and CORDEX have offered the opportunity to deepen our understanding of pattern scaling by using model ensembles (Lustenberger et al., 2014; Tebaldi and Arblaster, 2014; Christensen et al., 2015, 2019). For the particular case of temperature and precipitation, Tebaldi and Arblaster (2014) analysed the robustness of pattern scaling across time, Representative Concentration Pathways (RCP), and models using CMIP3 and CMIP5. They concluded that pattern variability is explained by the inter-member variability rather than the RCP variability. Their results showed that the pattern scaling is insensitive to the choice of emission scenarios (RCP4.5 or RCP8.5). Overall, only small differences were noted, suggesting that pattern scaling might provide a robust type of information across RCPs. They have also shown a greater variability of the signal for precipitation compared to temperature, which is likely due to differences in parameterization of cumulus convection together with cloud formation (Santer et al., 1990). To better understand high-end scenarios, Christensen et al. (2015) investigated the European response to a global mean warming of 6°C. They showed that the response was largely linear in global temperature change, compared to the scaled patterns produced from previous experiments (ENSEMBLES and PRUDENCE), with extreme precipitation as a notable exception. Also, Barring and Strandberg (2018) looked at scenarios/pathways and how scenarios for transient climate change taken at some window of global warming could (or could not) be used as proxies for stabilisation scenarios at the same warming levels. They found that this was often the case but not necessarily for all aspects of climate change – notably in relation to those involving some extremes.

Recently, Christensen et al. (2019) applied and compared pattern scaling from several coordinated experiments (PRUDENCE, ENSEMBLES and CORDEX). Their results show comparable patterns and ranges between these projects, suggesting that pattern scaling is robust across modelling initiatives over time. They also compared the scaled patterns of an observational dataset, here using CRU (Harris

et al., 2014), and also here show a high correspondence with scaled patterns originating from the coordinated experiments. This result supports the idea that the linearity of pattern scaling is also seen in observations and can be extended to, at least, the end of the twenty-first century. Interpretation of these results allowed us to extrapolate and hypothesise that such a pattern should be identified also in early simulated periods despite a weak signal-to-noise ratio.

Using the same approach as in Christensen et al. (2019), Matte et al. (2019) have shown that indeed such patterns emerge earlier within the simulation than was previously thought. The next section is first highlighting the end-century scaled patterns as the one calculated from 2080 to 2099, the latest 30-year period available for all models, divided by the time averaged global mean temperature change for the period. The following section discussed the emergence's pattern. The temperature and precipitation fields from the EURO-CORDEX experiment at 0.11° (EUR-11) and 0.44° (EUR-44) are used for RCP 4.5 and 8.5. See methodology in Matte et al. (2019), for more details.

The end century scaled pattern.

The end-of-century scaled patterns for temperature in DJF (first two rows of Figure 25) are showing strong warming in the north-east of the domain with a smaller value over the Atlantic Ocean as also observed by other studies (Christensen et al., 2015, 2019). However, although some larger differences can be observed in the signal-to-noise ratio (S/N), the percentage of grid points where $S/N < 1$ (shown in the upper-left corner of each figure) is quite small if not zero. Only the area over the Atlantic Ocean shows $S/N < 1$ due to a moderate climate change signal. Note that EUR-44 is slightly warmer than the EUR-11. This is caused by the slightly different model ensembles available. Selecting identical model ensembles for the two resolutions, the scaled patterns between EUR-11 and EUR-44 do not show differences in temperature (not shown).

The scaled patterns of the precipitation fields for DJF (last two rows of Figure 25) are showing, overall, a future with a wetter climate over northern Europe and drier conditions over the southern and north-western parts of the domain; the S/N ratio is quite similar. However, the percentage of grid points with $S/N < 1$ is considerably higher for precipitation (53, 63% for EUR-11 and EUR-44, respectively) indicating a higher disparity between ensemble members. In general, Northern European land areas have S/N larger than one. This is consistent with the global tendency of increased intensity of the hydrological cycle with global warming. The results show relatively low-level noise over southernmost Europe and a larger signal over central and northern Europe. The low-level noise over the European continent is likely due to the large-scale circulation constraint dominating this mid-latitude region in this season (stratiform precipitation from large low-pressure systems) (Sørland et al., 2018; Ozturk et al., 2021).

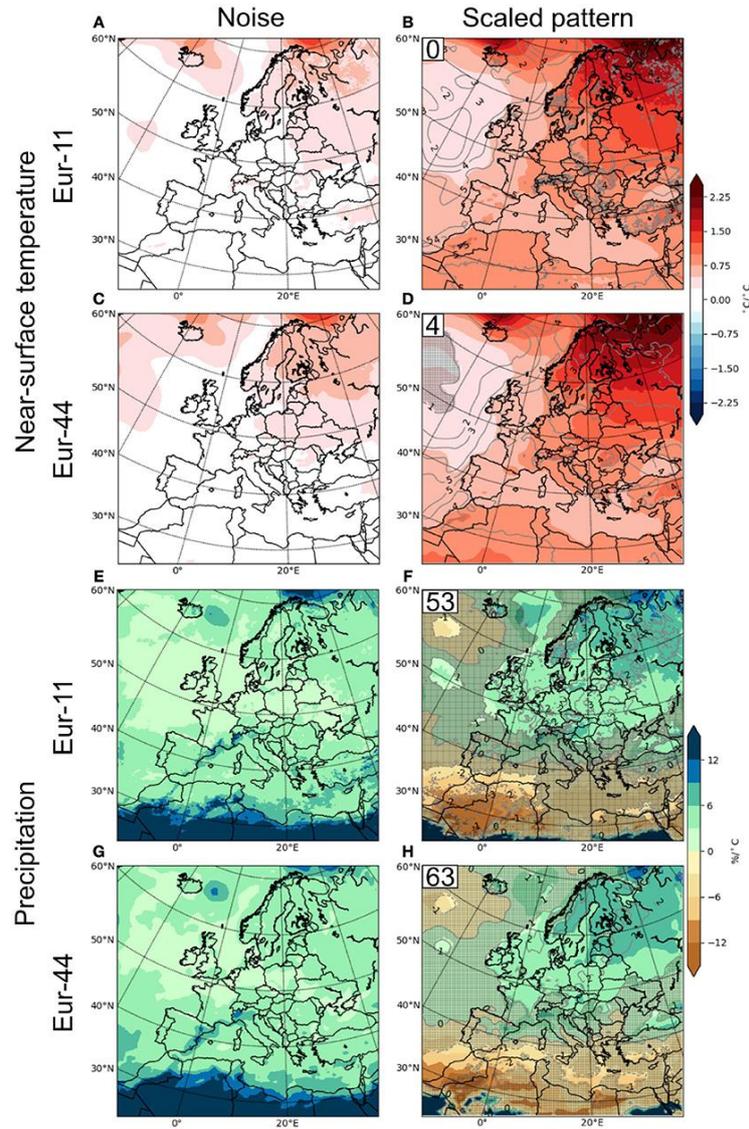


Figure 25. Scaled 2080–2099 DJF patterns for combinations of temperature (A–D)/precipitation (E–H) and EUR-11 (A,B,E,F)/EUR-44 (C,D,G,H). The left column shows the inter-member noise and the right column shows the 2080–2099 scaled patterns. The contour lines shown in the right column show the S/N ratio and the grey shading depicts areas of S/N < 1. The numbers in the upper left corner of the right column shows the percentage of grid points where S/N < 1. Note that both columns have the same unit (figure from Matte et al., 2019).

The scaled patterns of JJA temperature (Figure 26) is showing a more homogeneous warming over the domain than DJF with the highest warming rates in the southern and north-eastern parts of the domain as also observed in Christensen et al. (2019). In general, the JJA scaled precipitation patterns have a smaller S/N than those for DJF. Due to large inter-member disparities as seen here for JJA, as likely affected by the reproduction of convective precipitation, the percentage of grid points with S/N < 1 is higher (78, 83% for EUR-11 and EUR-44, respectively) than for DJF. The area where S/N > 1 over the Iberian Peninsula is due to a stronger signal. It is expected that noise levels are higher in summer than in winter, as weather is more locally generated, which also means that the role of the regional model for noise is higher than in winter; this was originally described by Déqué et al. (2007). Further discussions on the robustness in relation to S/N is discussed later. The large levels of noise for JJA and DJF in the southern parts of the domain are related to the use of relative rather than absolute changes.

This is supported by both a small absolute signal and a small absolute noise for this region (see Figure 12 of Matte et al., 2019).

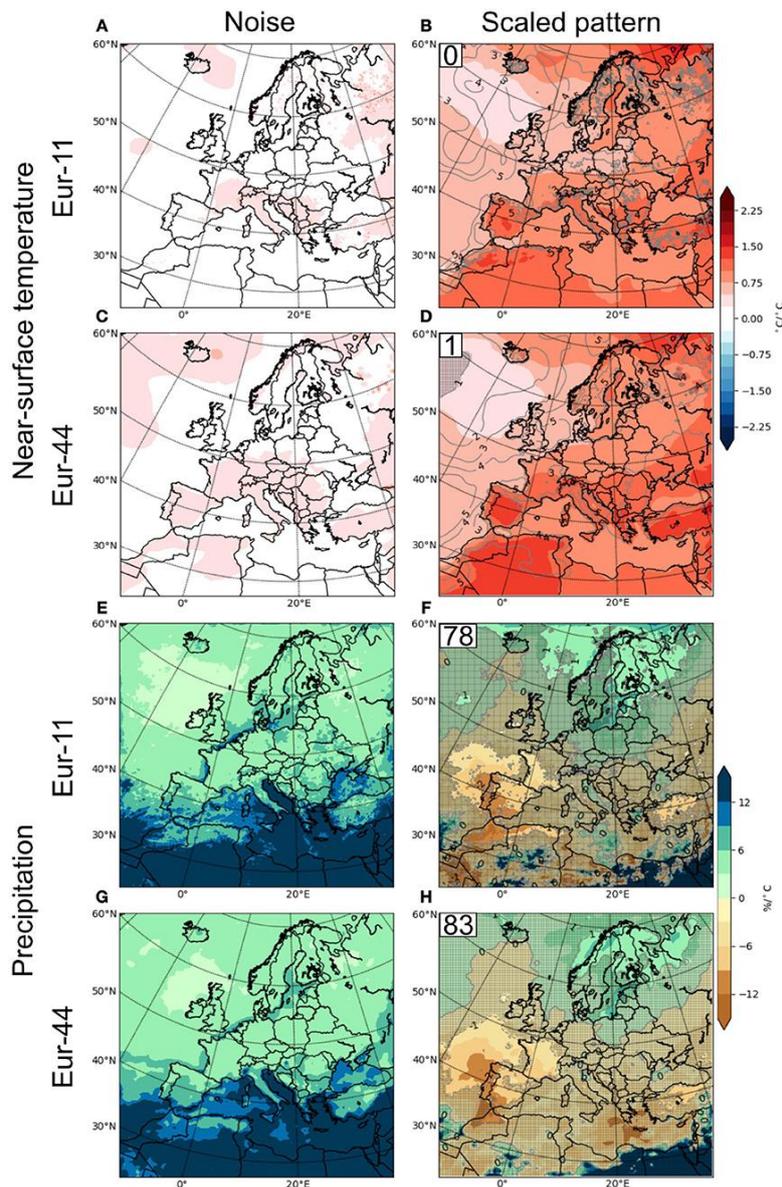


Figure 26. As for Figure 25 but for JJA (figure from Matte et al., 2019).

The emergence of end-century scaled pattern signal

This section is focusing on the emergence of the scaled patterns. To depict the emergence of the scaled patterns, movies are available (in the Supplementary Material of Matte et al., 2019) showing the temporal evolution of the temperature and the scaled precipitation patterns for JJA and DJF. Figure 27 shows the main statistics of the annual evolution of the scaled patterns from 2005 to 2090 (i.e., the 2005 level is calculated from the 1995–2014 period and so forth). The main purpose is to show at which time the scaled patterns, shown in the previous section, emerge at various geographical locations and scales.

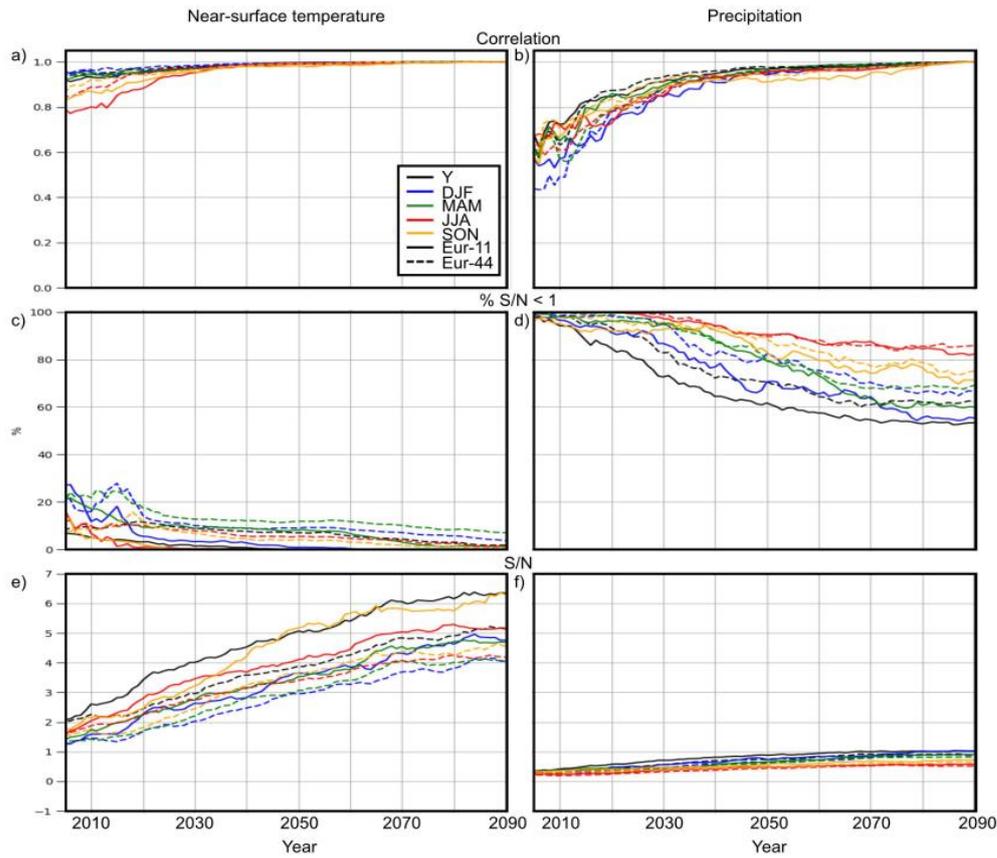


Figure 27. Pattern correlation between a running mean of 20-years scaled patterns from 2005 to 2090 (central year shown) against corresponding end-period levels (2080–2099) (A,B). Percentage of grid points where $S/N < 1$ (C,D). Spatial average of the S/N levels (using the absolute signal) of the scaled patterns (E,F). The results are shown for both variables (temperature and absolute precipitation, left and right respectively), and across resolutions and seasons as well as annually (adapted from Matte et al., 2019).

The first row of Figure 27 shows the spatial correlation between the scaled patterns calculated each year against the 2080–2099 scaled patterns (as shown in Figure 25 and Figure 26). The spatial correlation for temperature (Figure 27a) reaches the asymptotic unit value for all seasons and for both resolutions early in the century (around 2035). The latest alignment to the asymptotic value is seen for the JJA season, which again is likely due to smaller-scale convective weather systems. For absolute precipitation change, the evolution of correlations (Figure 27b) is more divergent than for temperature, reaching unity at a later stage (around 2080). However, the correlation levels seem to stabilise around 2050. For precipitation, DJF is the last season to reach its asymptotic value.

The second and the third rows of Figure 27 show the percentage of grid points where $S/N < 1$ and the value of the spatial average of S/N , respectively. For temperature (Figure 27c), the percentage of grid points where $S/N < 1$ is relatively low early in the period. For EUR-11, a level of 0% is reached around 2040 whereas EUR-44 decreases to around 2% at the end of the century. The scaled precipitation patterns differ substantially from those of temperature (Figure 27d), starting at approximately 100% of grid point where $S/N < 1$ for all seasons and resolutions, decreasing steadily to levels between 50% and 85% at the end of the century. Accordingly, the spatial average of S/N is increasing for both variables (Figure 27e,f). At the start of the period, spatially averaged temperature S/N levels are

already >1 reaching values between 4 and 6.4 at the end of the century (across seasons and resolutions) for temperature. However, although S/N levels also increase in precipitation, the spatial average is comparatively lower and shows a slower increase toward the end of the century, suggesting a much noisier field for precipitation. Despite some variations in the evolution of the statistics and relatively low S/N, the movies (see Supplementary Video S1 from Matte et al., 2019) suggest that the underlying emerging patterns are already recognizable from around 2020.

Perspective

It was shown that the scaled temperature patterns emerge faster than the corresponding patterns for precipitation. For temperature, the areas of credibility (grid points where $S/N > 1$) increase to 100% toward the end of the century whereas a stabilisation around 50–60% is seen for precipitation around 2050. The large noise throughout the century related to the scaled precipitation pattern suggests that the precipitation pattern is not scaling as linear as is the case with temperature. Although the signal is not formally discernible from the inter-member noise (the ratio of which decreases through the century), the consistency in the scaled precipitation patterns already from 2020 suggests that the precipitation field is actually scalable at longer timescales.

In view of the known model- and GCM/RCM configuration deficits, Christensen et al. (2019) note in their analyses of the PRUDENCE–ENSEMBLES–CORDEX sequence, that the pattern-scaled projections of temperature and precipitation change across 20 years of model evolution appear quite similar. Likewise, the patterns of change of annual mean temperature and precipitation can be found to compare well with the results from global models (e.g., Collins et al. 2013). This apparent robustness of the model sequence represents an important contribution to credibility of the projections provided by the Euro-CORDEX data set representing the current state of the art.

Publication

Matte, D., Larsen, M. A. D., Christensen, O. B., & Christensen, J. H. (2019). Robustness and scalability of regional climate projections over Europe. *Frontiers in Environmental Science*, 6, 163.

3.4 Spatial Merging and emulators (SME)

3.4.1 SME1: Large-scale circulation change and spatial merging (KNMI)

KNMI investigated the possibility to use measures of large-scale atmospheric circulation (change) as a basis to perform spatial merging between RCM and GCM simulations. Three different methods have been explored. First, we used a local regression method based on different components of the near-surface geostrophic wind. This method is employed in the common framework (and described in Appendix B). Second, we used an atmospheric clustering technique, assigning each day to an atmospheric cluster. Third, we used a circulation analogue technique, whereby each day in the future climate period is matched to the closest matching circulation in the control climate (and vice versa) allowing a separation of precipitation changes into circulation and thermodynamic contributions (*de Vries et al., 2022, accepted*). In all methods information from the large scale (GCM) is used to understand the signal at the local scale (RCM). The last two methods are discussed below.

The two methods discussed here were not applied in the common framework. Given time constraints we could only apply one method. Also, the methods discussed here rely on the existence of a very large data set (480 years of data for each 30-year time period analysed from the 16-member ensemble as discussed in section 3.3.1), which far exceeds the 10 years of data available in the common framework. They mostly provide meaningful results for the systematic changes within the 16-member ensemble and not for individual members (see also Figure 31). Finally, further analysis showed that already the simpler methods captured most of the circulation dependent components, and therefore we did not expect better results from these methods in the common framework than the simple regression method used there.

3.4.1.1 Tailored weather patterns and climate change

Introduction

One of the main goals in spatial merging is to quantify how the large scale influences the local scale and how this information can be used to understand changes occurring at these smaller scales. That this interaction exists is undebated: from daily experience we know that circulation patterns can be associated with radically different weather (Figure 28).

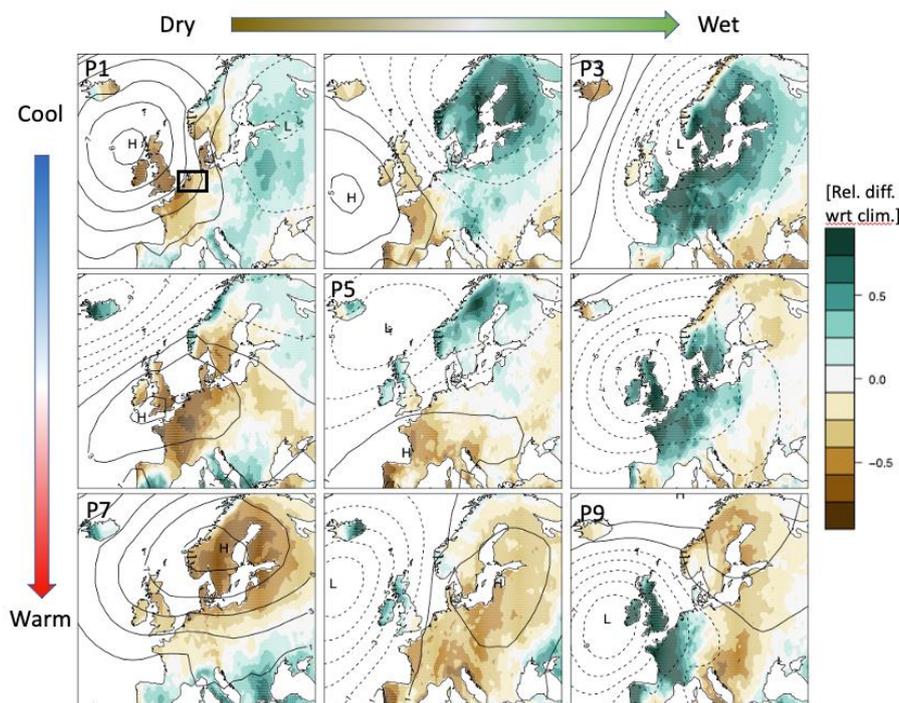


Figure 28. Example of “tailored” weather patterns. First the (2m-temperature) x (precipitation) PDF of domain NLXL (0-10E, 50-55N, see box) is binned in tertiles, yielding 9 composite patterns ranging from “cool and dry” (top-left) to “warm and wet” (bottom-right). Shading shows anomalous precipitation (relative difference with respect to climatology), contours mean sea level pressure anomalies (hPa).

These figures show a composite of mean sea-level pressure, mslp (taken from ERA5 reanalysis) and E-OBS gridded precipitation and temperature. The composite is created for the period 1981-2010, by first partitioning the two-dimensional daily precipitation-temperature summer PDF of a domain NLXL (0-10E, 50-55N, see box) into tertiles. The 9 resulting patterns are *tailored* to different parts of the temperature-precipitation distribution. The mslp field is then re-projected onto the 9 patterns such

that response variables can be computed. The patterns have a large spatial scale and are easily connected to the typical weather they produce.

Future changes

How climate change affects the frequency of weather patterns and their response fields is not immediately obvious, mainly because of the large natural variability in weather patterns. This necessitates the use of an ensemble approach. Here the single-model climate ensemble is used, that was obtained with the GCM EC-Earth (Hazeleger et al, 2012) and the RCM RACMO2 (van Meijgaard et al, 2012; Aalbers et al, 2018). The ensemble consists of 16 simulations (period 1850-2100 for the GCM, dynamically downscaled with RACMO between 1950-2100, and forced by RCP8.5 after 2005).

To investigate the frequency and response changes over time, each of the model summer days is projected on each of the composite mslp fields, using Euclidean distance as a metric. Figure 29 shows these changes over the period 1950 to 2100. A 30-year running average is used. The ensemble exhibits a clear drying of around 15% towards the end of the 21st century (black line in the right panel, labelled as “sum”). Part of these changes can be explained from frequency changes (left panel). Especially P9 and P6 and to lesser extent also P3 decrease in frequency. These three patterns were derived from the wettest tertiles. In contrast, the frequencies of driest patterns (P1, P4 and P7) increase. They are associated with high pressure over the Atlantic, central Europe and Scandinavia. Thus, from frequency changes alone we expect a precipitation decrease over NLXL. The pattern-specific precipitation, however, also changes (right panel): The *wet* patterns tend to get wetter (at least up to mid 21st century), whereas the drier patterns show almost no change or tend to get dryer toward the end of the century.

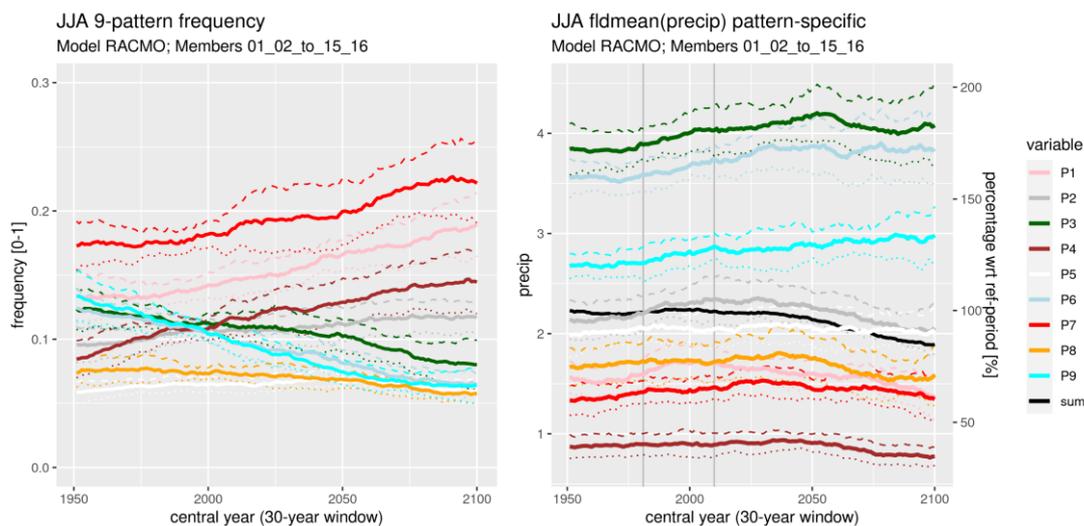


Figure 29. (left) Showing pattern-frequency obtained from EC-Earth and (right) the pattern-specific precipitation derived from RACMO [mm/day]. Running averages are used (30-years, smaller-window near 1950 and 2100). Dashed and dotted lines indicate +/- one ensemble standard deviation. Colour coding: Green/blue \Leftrightarrow wet patterns (P3,P6,P9); Red/brown \Leftrightarrow dry patterns (P1, P4, P7). The black line is the total relative change.

Mechanisms

Basic mechanisms can be defined through which the precipitation changes: (1) “between-cluster” circulation changes (bec_cc) following from frequency shifts (cf. Figure 29a); (2) “within-cluster” changes (wic) in which the pattern-specific precipitation changes (cf. Figure 29b) and (3) “non-linear”

changes (nonlin) in which both terms change. Finally, the “within-cluster” change term can further be subdivided in a (thermodynamic) response term (wic_rc, “response-change”) and a ‘within-cluster’ circulation-change term (wic_cc) that measures small circulation shifts within a cluster (de Vries et al., 2013). Figure 30 shows these contributions.

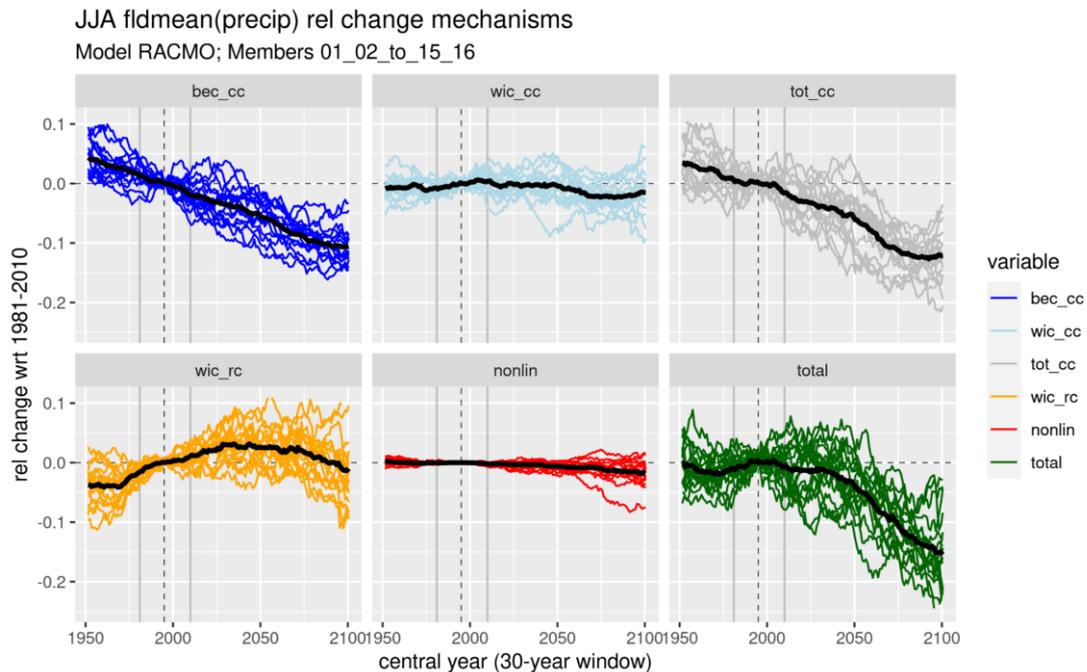


Figure 30. Contributions to the summer precipitation change in NLXL. Individual ensemble realisations are shown in thin coloured lines, the ensemble-average in black. The top-right panel shows the sum of the top-left and top-middle panels. The total change is shown in the bottom-right panel.

Frequency changes alone (top-left) account for much of the ensemble-mean decrease in mean precipitation over NLXL. Within-cluster circulation changes do not add much to this in the ensemble-mean yet show large ensemble spread. The “within-cluster” (thermodynamic) response changes (bottom-left) increase towards the mid-21st century, and may decrease thereafter. This temporal response pattern may be related to the competing influences of (absolute) humidity increases with warming and progressive soil drying towards the end of the century. Nonlinear effects (due e.g., to a reducing frequency of a pattern that gets wetter over time) are small. Note the large spread between members; individual ensemble realisations (thin lines) are noisy despite the 30-year running averages.

Discussion

Using a *tailored* weather-pattern approach we showed that systematic frequency shifts (towards more frequent patterns involving high pressure) can explain a large fraction of the ensemble-mean precipitation trend in the NLXL region in summer in the EC-Earth/RACMO ensemble. Up to the first half of the 21st century the resulting precipitation decrease is compensated for by a wetting of the wet-patterns.

These results, however, need to be interpreted with care as single realisations of the ensemble evolve quite differently than the ensemble mean (Figure 30). It was hoped, and to a certain extent expected, that a considerable part of the variability between ensemble members in the total signal could be explained and perhaps even predicted from changes of the large-scale weather-pattern variability. However, this is not really the case and other factors (such as the response changes and even the within-cluster circulation contributions) are also required.

We illustrate this limitation of the approach with one example. Figure 31 below shows a Taylor diagram constructed from predictions made using various combinations of weather-pattern frequencies and their response fields on a 5-year time scale. At shorter time scales we expect the influence of circulation to be more important.

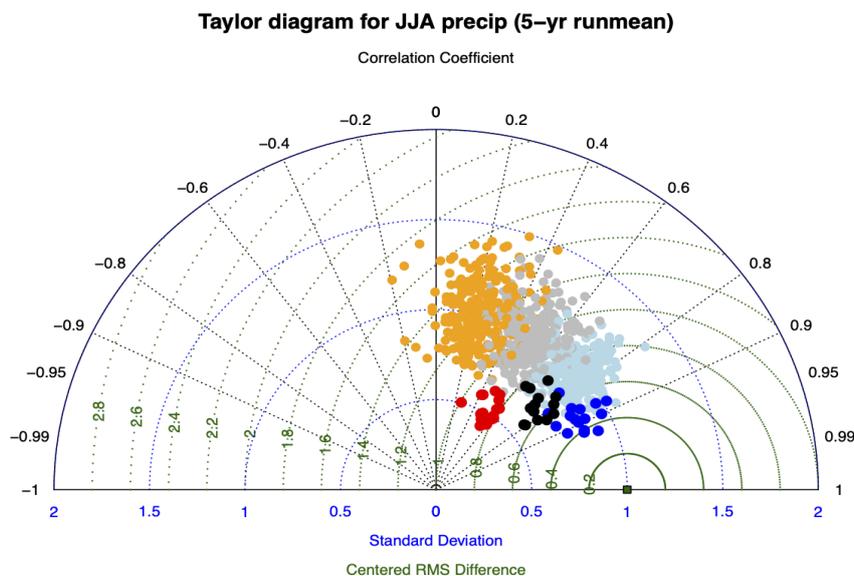


Figure 31. Taylor diagram of various precipitation predictions based on the circulation clusters, averaged across the domain NLXL and for the period 1950-2100. The 5-year running mean precipitation of each member is predicted in 6 different ways: (i, orange) using the 15 other members as predictions (giving 16x15 dots); (ii, red) using the ens-mean (16 dots), (iii, grey) using its frequency but the response from the 15 other members; (iv, light blue) using its response but the frequency of the 15 other members; (v, black) using its frequency but the ens-mean response; (vi, dark blue) using its response but the ens-mean frequencies.

The benchmark is to beat the (trivial) predictions: replacing a member by another member (orange, 16x15 dots) or by the ensemble mean itself (red, 16 dots). These still mostly yield a positive correlation due to the trend. Much better performance is obtained if one combines the frequencies of a member with the ensemble-mean response (black, 16 dots). Correlations are substantially higher, and better than when using individual responses (grey, 16x15 dots). Yet the highest correlations are found if the ensemble-mean frequencies are combined with the individual responses (dark blue, 16 dots). Thus even on shorter time scales it seems more beneficial to know the individual response field than the correct weather-pattern frequencies.

Thus, it appears that if we want to emulate a future 30-year climate, the information contained in the frequency changes of large-scale circulation clusters is limited. This could be due to the coarseness of the clustering technique assigning days to rather broadband circulation clusters. In the next section, we will investigate whether a more careful treatment of the large-scale circulation will improve this situation.

3.4.1.2 Using circulation analogues to explain future summer precipitation

Introduction

Regional climate projections show that European summer precipitation may change considerably in future (Meehl et al, 2007; Polade et al, 2014; Coumou et al, 2018). Southern Europe is likely to

experience drying while northern regions could become wetter. Future projections come with large uncertainties because many factors influence precipitation formation (e.g., radiation, soil-moisture, large-scale circulation, relative humidity, vertical stability, cloud microphysics), most of which may change (O’Gorman and Schneider, 2009; Fowler et al, 2021). To understand the pattern of future precipitation change, we need to find the underlying mechanisms in a background dominated by natural variability (Rowell and Jones, 2006; Polade et al, 2014; Pfahl et al, 2017; Norris et al, 2019). Generally, two main categories are distinguished: thermodynamic and dynamic mechanisms (e.g. Rowell and Jones, 2006; Norris et al, 2019). The latter includes effects from circulation change.

Here we focus on the role played by the large-scale circulation. The aim is to quantify to what extent large-scale circulation changes contribute to the regional precipitation-change pattern over Europe. The circulation-induced effects are identified using circulation analogues (Yiou, 2014; Clemins et al, 2019). With a good quantification of the large-scale circulation effects we can potentially estimate the precipitation response to a different circulation response as well, proving a potential basis for spatial merging techniques. The methods developed and used here could therefore prove valuable in coordinated regional climate-change experiments such as CORDEX (Giorgi et al, 2006; Coppola et al, 2018; Pichelli et al, 2021).

Data and methods.

We use the same single-model climate ensemble as in the previous section (RACMO2 driven by EC-Earth). We briefly summarise how the analogue method is used and refer to De Vries et al. (submitted) for more details. We consider the difference between two periods: 1991-2020 and 2071-2100.

Analogues. A (circulation) analogue is defined as a reconstruction of a climate whereby days are sought that match with a target mslp time series; for instance, a reconstruction of the future climate with the target mslp timeseries of the control period. These analogues are found for each day by minimising the Euclidean distance of the mslp anomaly difference field with all other days from all ensemble members within the same season. Other fields are taken from this reconstruction.

Decomposition. Let P and F be the present-day and future state of a climate variable X (such as precipitation). In this study the future change $\Delta X = F - P$ is written as a sum of three terms and a residual. The first term records changes occurring without circulation change (referred to as the *thermodynamic* contribution). The second and third term measure the contribution from changes in the large-scale *mean circulation* and its *variability*. Each term can be expressed as a specific combination of analogues. Finally, a δ -term collects a left-over signal that is unexplained by the analogues. While developing the separation into the different terms, we discovered that the method of analogue-selection easily creates a biased result; to correct for this we introduced analogues for the same time period (De Vries et al., submitted).

Mean response

Figure 32 shows the relative change of the summer mean precipitation scaled by the mean increase in global temperature (3.1 K). A clear meridional wet-dry gradient is seen, from modest increases in the North to substantial drying of locally more than 15% per degree warming further south. The east-west modification of the pattern is related to the gradual build-up of high pressures above Ireland. Compared to the amplitude of a daily weather pattern, its amplitude is modest, yet its impact on precipitation is seen nevertheless in a large region downstream. The change patterns are structurally

but also quantitatively similar in RCM and GCM. The signal-to-noise ratio (S2N, bottom panels) is determined as the ensemble mean change divided by the ensemble standard deviation. The south-west part of the domain displays the highest S2N values. Further to the east and north S2N is smaller. Note that the S2N is obtained by considering two periods far apart in time, and for the high RCP8.5 emission scenario. For the lower RCP4.5 scenario or for more nearby time periods, S2N will be reduced.

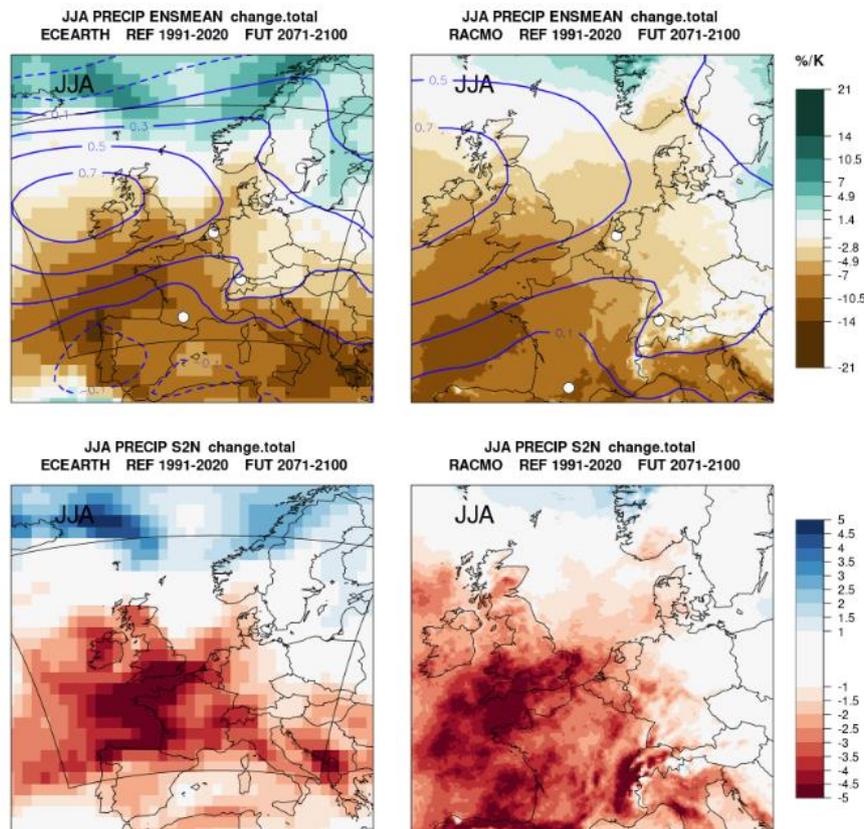


Figure 32. (top row) Change of summer precipitation per degree warming [%/K]; (bottom row) Signal-to-noise (the change of mean divided by ensemble standard deviation) for EC-Earth GCM (left) and RACMO RCM (right). Contours show mslp trend per degree warming (interval 0.2 hPa/K). [from De Vries et al. (submitted)].

Thermodynamic versus circulation-change contributions

Using the method of circulation analogues one can separate the total precipitation change pattern into contributions from circulation-change and thermodynamics. The circulation is further decomposed into a term related to the mean-circulation change and one that may pick up signals from changes in e.g., weather-pattern variability or altered persistence of weather patterns.

The panels in Figure 33 show the decomposition for the RCM (GCM results are similar). The thermodynamic pattern is more zonally oriented than the total change pattern (Figure 32), with robust drying in the south and a gradual change towards wetting further north. For a broad latitudinal band the ensemble-mean thermodynamic contribution is almost zero. The circulation contribution exhibits a drying pattern in the area centred and downstream of the region where the pressure increase is strongest. For a broad zonal band from Ireland towards the Netherlands and central Germany, almost the entire drying signal seen in the model can be explained from the change of the mean circulation.

The drying extends well into Scandinavia. There it competes with the wetting coming from the thermodynamic component (and reducing S2N).

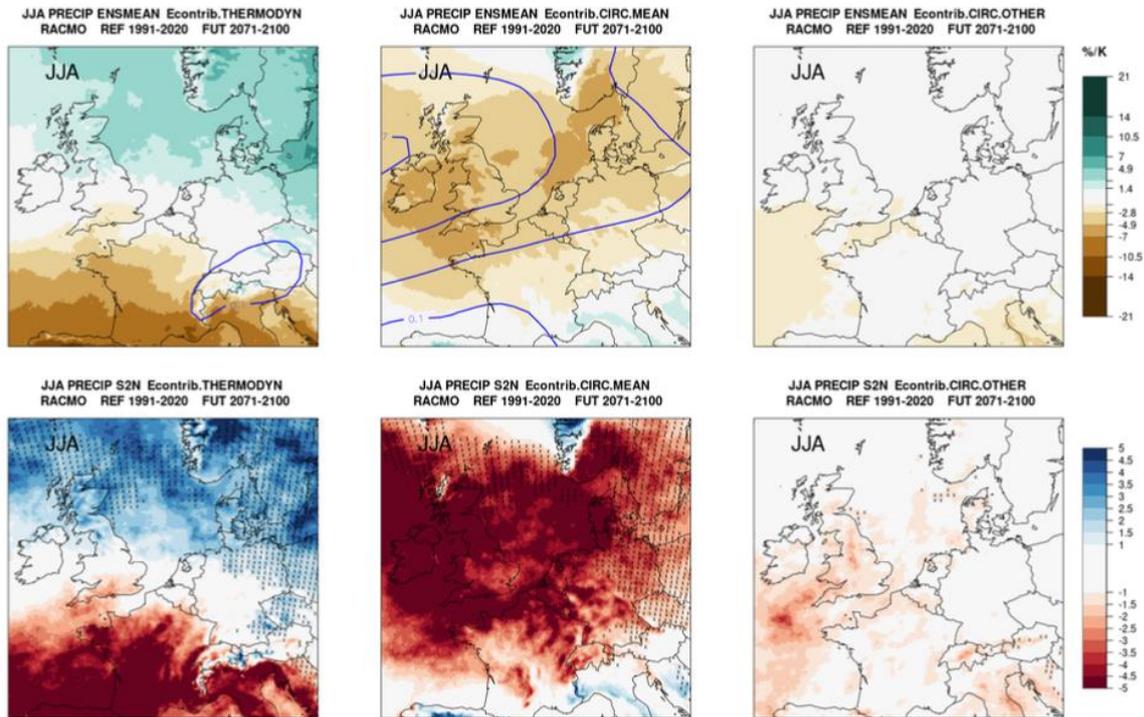


Figure 33. Contributions to RACMO (RCM) summer precipitation change per degree warming (top) and their signal-to-noise ratios (bottom). Contours show change of mslp [hPa per degree warming]. Crosses indicate the region where the absolute value of S2N of the total change is smaller than one while that of the component exceeds unity [from De Vries et al. (submitted)].

Separating the circulation terms into mean and variance contributions (middle and right panel), shows a dominance of the former. Off the coast of France a small drying contribution is found from the circulation variability term. However, there is no clear signal that altered circulation variability, persistence or more subtle circulation effects strongly impact the mean precipitation change in summer over land areas in central and western Europe. The ensemble spread of this term is however quite large (not shown), suggesting that it can still be important in individual climate realisations.

Connection to Pseudo-Global-Warming (PGW) experiments.

In PGW experiments a regional climate model is forced with present-day circulation variability augmented with a delta change signal (Schär et al, 1996). Analogues differ from PGW simulations in at least two aspects. First, by adding a delta-change field, each *futurized* PGW-member will still be (approximately) in the same state of natural variability as its original for all other precipitation drivers. For instance, if the original member had anomalously high SST, this is still also the case after applying the PGW perturbation. Future analogues differ because only the circulation is matched, and other drivers regress to the mean. Furthermore, PGW approaches include by construction only the *thermodynamic* and *mean-circulation* changes, not those of modified large-scale circulation variability (which are highly variable but have only a small ensemble-mean contribution). From the analysis

shown here we conclude that the error made in such PGW-experiments (that is, by neglecting the circulation-variability change component) is not particularly large for summer precipitation.

Summary and concluding remarks.

Our main findings are as follows. First, thermodynamic and mean circulation changes (as measured by mslp) are shown to be sufficient to explain the ensemble-mean precipitation change in summer over most of western and central Europe in the EC-Earth/RACMO ensemble. The thermodynamic pattern is characterised by a north to south wetting to drying gradient. The mean-circulation contribution is tied to high pressure developing west of Ireland. Downstream of this high-pressure region almost the entire (drying) signal is explained by this term (including region NLXL of the previous section). In some areas the contributions reinforce while in others they counteract. Second, changes in circulation-variability only play a marginal role in explaining the ensemble-mean precipitation signal. In contrast, however, this term has a disproportionately large ensemble spread.

However, the study also shows that large-scale circulation variability alone is not sufficient to explain the ensemble-spread in precipitation trends. A large fraction of the ensemble uncertainty is left unexplained by circulation. An example of this is given in Figure 34 where we show the different terms – thermodynamic, circulation and residual delta – for four locations (dots in Figure 32).

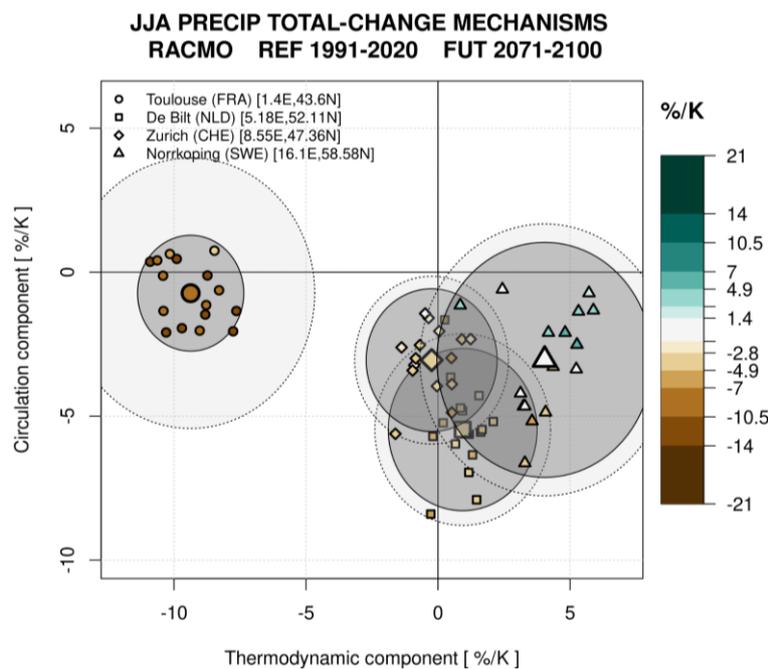


Figure 34. Total fractional change of summer precipitation in RACMO per degree warming for selected locations as a function of the thermodynamic (x-axis) and circulation (y-axis) contributions. Ensemble mean values are indicated by big symbols and colour indicates the total change, small symbols show the contributions for each member. The dark-grey circles are centred around the ensemble means and approximate the 95% ensemble spread. The light-grey circles approximate the 95% spread of the δ -term measuring the change unexplained by the circulation [from De Vries et al. 2022, accepted].

The unexplained ensemble spread is a consequence of the fact that circulation is not the only driver of precipitation, and that the natural variability in the other precipitation drivers is also important.

Publication

de Vries, H., Lenderink, G., Wiel, K. van der & van Meijgaard, E. (2022) Quantifying the role of the large-scale circulation on European summer precipitation change. *Climate Dynamics*. Accepted March 2022.

3.4.2 SME2: UNet-based statistical emulator of RCM (CNRS-CNRM)

Introduction

Delivering reliable regional or local climate change projections for the next decades that are both at fine scale and taking into account all sources of uncertainty is currently an unsolvable problem with dynamical models. Indeed, it requires at the same time large ensembles to capture the various sources of uncertainty (scenario choice, model choice, natural variability) and computationally expensive fine-scale climate model runs. To tackle this typical “spatial merging” issue, we propose in EUCP/T5.4 a completely new approach called “RCM-emulators”. The RCM emulators belong to the family of the hybrid downscaling approaches (Déqué et al. 2012, Walton et al. 2015, Erlandsen et al. 2020, Wang et al. 2021, Doury et al. 2022). It aims at combining the physical basis of the dynamical downscaling approach (i.e., RCM or CPRCM) with the low computational cost of the empirical statistical downscaling (ESD) using recent machine learning techniques. Maraun and Widmann (2018) first introduced the statistical emulators' concept, also mentioned in chapter 10 of IPCC-AR6 (2021). We are developing and applying it below.

RCM-emulator concept

RCM-Emulators are a novel hybrid downscaling approach that emulates the downscaling function of a given RCM. That is to say learning the transformation of the large-scale climate information into a local climate information performed by a regional climate model. The idea is to combine Dynamical Downscaling and Empirical Statistical Downscaling (ESD) in order to tackle their respective limitations. On one hand, it uses existing RCM simulations to estimate the downscaling function, which enables calibration of the emulator on longer series, in future periods, at different time scales, for not-observed variables or regions and over complete domains. On the other hand, the emulator relies on Machine Learning algorithms which grants good computational efficiency regarding the huge computational costs of RCM. The final aim of such a tool is then to enlarge the number of high-resolution simulations for a given variable and a given area, knowing the large-scale conditions from any input GCM simulation. Indeed, RCM-emulators allow downscaling more GCM simulations, including longer periods, multiple members, different scenarios and different models, at a much lower cost. Thanks to these bigger ensembles, it will be possible to better estimate the uncertainty associated with the high-resolution projections of the emulated variables. However, RCM-Emulators are a very novel hybrid downscaling method and it is necessary to first study their feasibility and robustness, which has been the main focus of the work performed by CNRS-CNRM within EUCP Task 5.4.

The conceptual framework used to build the emulator has been described in Doury et al. (2022). RCM-Emulator uses a neural network architecture to learn the relationship between large-scale fields and local-scale variables inside regional climate simulations. RCMs include a downscaling function F which transforms large scale information given through a couple X, Z (X for 2D variables and Z for 1D variables; more details in the Predictor section below) into high-resolution surface variables Y . The statistical RCM-emulator aims to solve a common Machine Learning problem

$$Y = F(X, Z)$$

which is to find a statistical estimator of F in order to apply it to new GCM simulations. The following paragraphs describe the list of predictors (X, Z) used as inputs and their domain, the predictand (or target, Y) and its domain, the neural network architecture, the framework used to train the emulator and the metrics used to evaluate its performances.

UNet-Emulator

The emulator used in this document relies on a specific convolutional neural network (CNN) architecture called UNet. The strong advantage of CNN is their very good ability to deal with images and extract features from them. In particular UNet has been introduced for its good results in image segmentation. This last point is particularly interesting in our case as we hope that the emulator identifies meteorological objects, to derive the high-resolution temperature. Moreover, as explained in Doury et al (2022), the “U” shape of the UNet was specifically interesting for the emulator we developed.

Training in perfect model framework

The RCM-Emulator is trained in a perfect model framework, meaning that both predictands and predictors are coming from the same RCM simulation. This framework guarantees a perfect correspondence between large scale predictors and a local scale predictand, avoiding any biases or temporal mismatches that can exist between a GCM and its driven RCM. The complete protocol used to train the emulator is schematized in Figure 35. The predictors are upscaled from the RCM resolution to the GCM one, and then smoothed using a 3 by 3 grid points moving average filter. There are 2 reasons behind this smoothing:

- It allows sticking to the effective resolution of the GCM that is at least 3 times the grid mesh size (Klaver et al. 2020).
- It targets to delete any local scale information, which might persist through the upscaling step.

Once the predictor dataset (see paragraph about the list of predictors) is built, the UNet-Emulator can be trained, by showing it multiple examples (here multiple days) of corresponding input and targets. Once the emulator is trained, we evaluate its performance in the perfect model world first with a GCM-RCM simulation not used for the training and apply it then to GCM simulation.

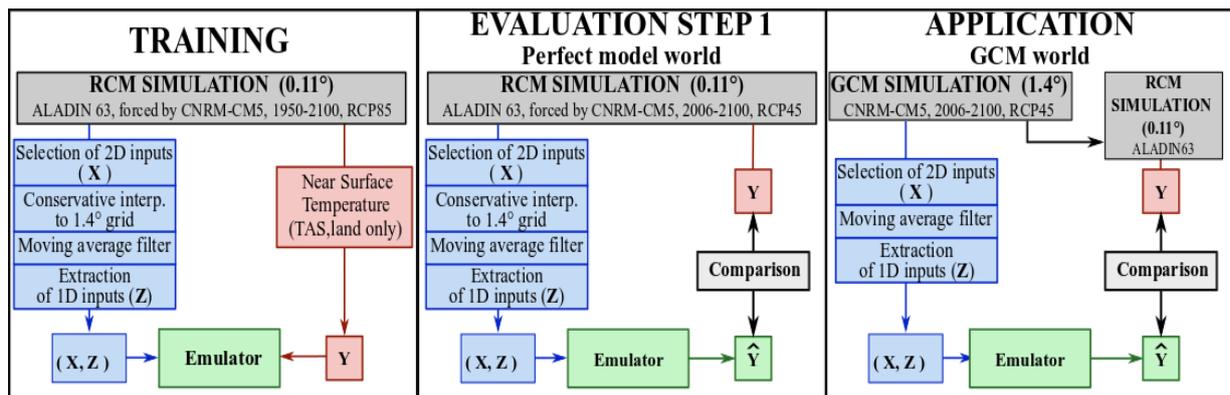


Figure 35. Scheme of the protocols for the training (left), the evaluation in perfect model (centre) and the application to real GCM data (right figure)

List of predictors

The good ability and efficiency of UNet to deal with images allows us to give multiple raw 2D predictors with minimal prior selection (which could introduce some bias) or statistical pre-work (which might delete some of the information). The 2D variables (X) used as predictors are the geopotential, the temperature, the humidity and the wind eastern and northern component at 850, 700, 500 hPa, plus the near-surface wind components and the sea level pressure. The total aerosol optical depth completes this list. On top of these 2D variables we also add some 1D variables (Z): the total greenhouse gases concentration, a seasonal (sinus, cosinus) vector which gives information about the position of the day in the year, and the solar and ozone forcings. Moreover, it is necessary to give standardised input to the network, such that each variable has comparable variances. We standardise the 2D variables spatially, with respect to the daily spatial mean and standard deviation of each variable. To keep the temporal information contained in these means and standard deviations, we include them in the list of 1D variables. The 1D variables are standardised “temporarily”, with respect to their means and standard deviations over a reference period.

UNet-Emulator for temperature over South-West Europe

Doury et al (2022) present the concept of RCM emulator as a new hybrid downscaling technique, propose a training-evaluation-application protocol and test the UNet-emulator for the daily temperature for multi-decadal scenario simulations and over a South-West European domain, including notably the Pyrenees, the French Alps, the French and part of the Spanish Mediterranean coast and the Balearic Islands. This domain was chosen for its diversity and complexity of the climate to be emulated. We only give a brief summary of the main conclusions of the publication here. The historical and RCP85 scenario simulations of the EURO-CORDEX CNRM-CM5 / CNRM-ALADIN63 pair (total period 1951-2100) were used for the training phase whereas the RCP45 scenario simulation (2006-2100) with the same model pair was used for the evaluation and application. The main conclusions are:

- The emulator generally fulfils its mission by capturing very well the transformation from low resolution information to the high-resolution near surface temperature.

- The emulator succeeds very well in reproducing the high-resolution spatial structure and daily variability of the RCM.
- Training the emulator in both past and future periods improves consequently the performance of the emulator.
- The emulator showed some limitations in accurately simulating extreme events and of the complete climate change magnitude.
- This study highlighted that the perfect model framework is the best to properly evaluate and compare emulators.
- The UNet-based emulator has been compared with emulators based on more standard ESD methods such as CDF-t and Multi-linear regression. It generally outperforms those more standard tools.

UNet-Emulator for precipitation over Alpine domain: Impact of the loss function

We present here some preliminary results on the emulation of the high-resolution precipitation field in CNRM-ALADIN63, over an Alpine domain visible on Figure 36, Figure 37, and Figure 38. The domain of 128 by 128 grid points is centred around the Alps. The statistical nature of high-resolution daily precipitation makes the task more complex than for near-surface temperature. During its training, the neural network calibrates a set of parameters according to a loss function which defines the cost of an error in the prediction. As explained in Doury et al (2022), the high-resolution daily temperature field has a gaussian behaviour, so the mean squared error was a natural choice for the loss function. We propose here to compare 3 different losses functions for the emulation of precipitation:

- The mean squared error (MSE)
- The mean absolute error (MAE)
- A custom made mean absolute error (CMAE), which penalises the network stronger when it underestimates the prediction in the case of extremes.

The three emulators were trained following the perfect model framework described above (Figure 35), using the historical and RCP85 simulations (1950-2100) from the same GCM/RCM pair as above. They are evaluated here in a perfect model approach over the RCP45 simulation and the results are commented below.

First of all, Figure 36 shows some illustrations of the results obtained with the emulators with respect to the RCM truth. The top row shows the high-resolution precipitation map of a given day for the three emulators and the reference. All emulators seem to have the good ability to locate the precipitations at the right places and seem to reproduce the good range of intensity. However, all three emulators seem to produce a precipitation field slightly smoother than the RCM truth. The second row of Figure 36 shows randomly chosen time series for the three emulators versus the reference, and all three emulators show a good temporal correlation. The bottom row shows the probability density functions (PDF) for the emulators and the RCM over the entire domain, a small box around the western Alps and one centred on the Paris region. Note that data from the various grid points of the chosen domains have been pooled to obtain the PDFs. These PDFs plots are normalised following the ASoP method (“Analysing Scales of Precipitation”) presented in Klingaman et al. (2017) and explained in Berthou et al (2020). We also use the Fractional Contribution (FC) index, introduced in Berthou et al

(2020), to quantify the absolute difference between the normalised PDF of the emulators and the RCM truth. The Emul_MSE reproduces the mean correctly but misses both tails, and specially the small precipitations. On the other hand Emul_MAE appears to fit extremely well the left tail of the PDF but misses the rest. For this reason we tried to build a loss function adapted from the MAE in order to reproduce better the higher intensities. The Emul_CMAE fits well the entire distribution for all three cases with very low FC indices with respect to the two other emulators.

Figure 37 confirms this last result. Both MSE and CMAE emulators are correctly reproducing (with less than 10% error) the RCM present climatology, while the MAE version is on average 20% drier. Regarding the percentage of dry days ($pr < 1\text{mm}$) over the 2005-2025 period the MAE performs very well, together with the Emul_CMAE, while the MSE emulator produces about 20% more days with more than 1mm rain. The maps showing the 99th quantile, highlights that both MAE and MSE emulators are performing poorly regarding the representation of extreme values while the CMAE emulator still performs accurately.

Finally, Figure 38 points out an interesting feature of the three emulators. Regarding the climate change signal for the climatology, the percentage of dry days and the 99th quantile, the three emulators appear to perform well. This result tends to tell us that the errors mentioned above stay coherent all along the series, and each emulator stays coherent. The climate change pattern is driven by large scale evolution, which is seen in the same way by all emulators.

To conclude, we transposed here the emulator for the near-surface temperature developed in Doury et al. (2022), to the daily amount of precipitation. Given the non-gaussian statistical nature of the field we proposed different loss functions to train the neural network. The three versions proposed here showed very interesting results, but the emulator trained with the Custom MAE seemed to reproduce the RCM simulation most accurately.

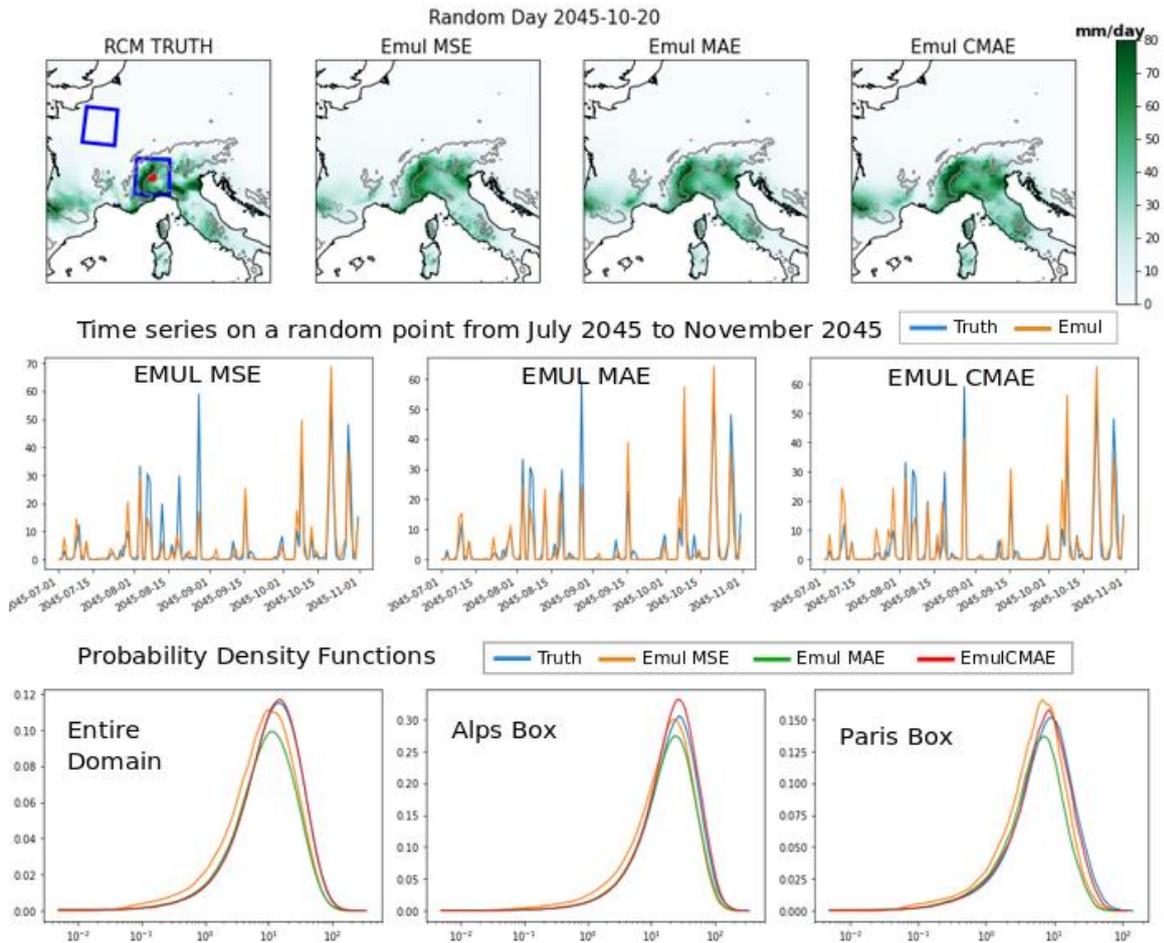
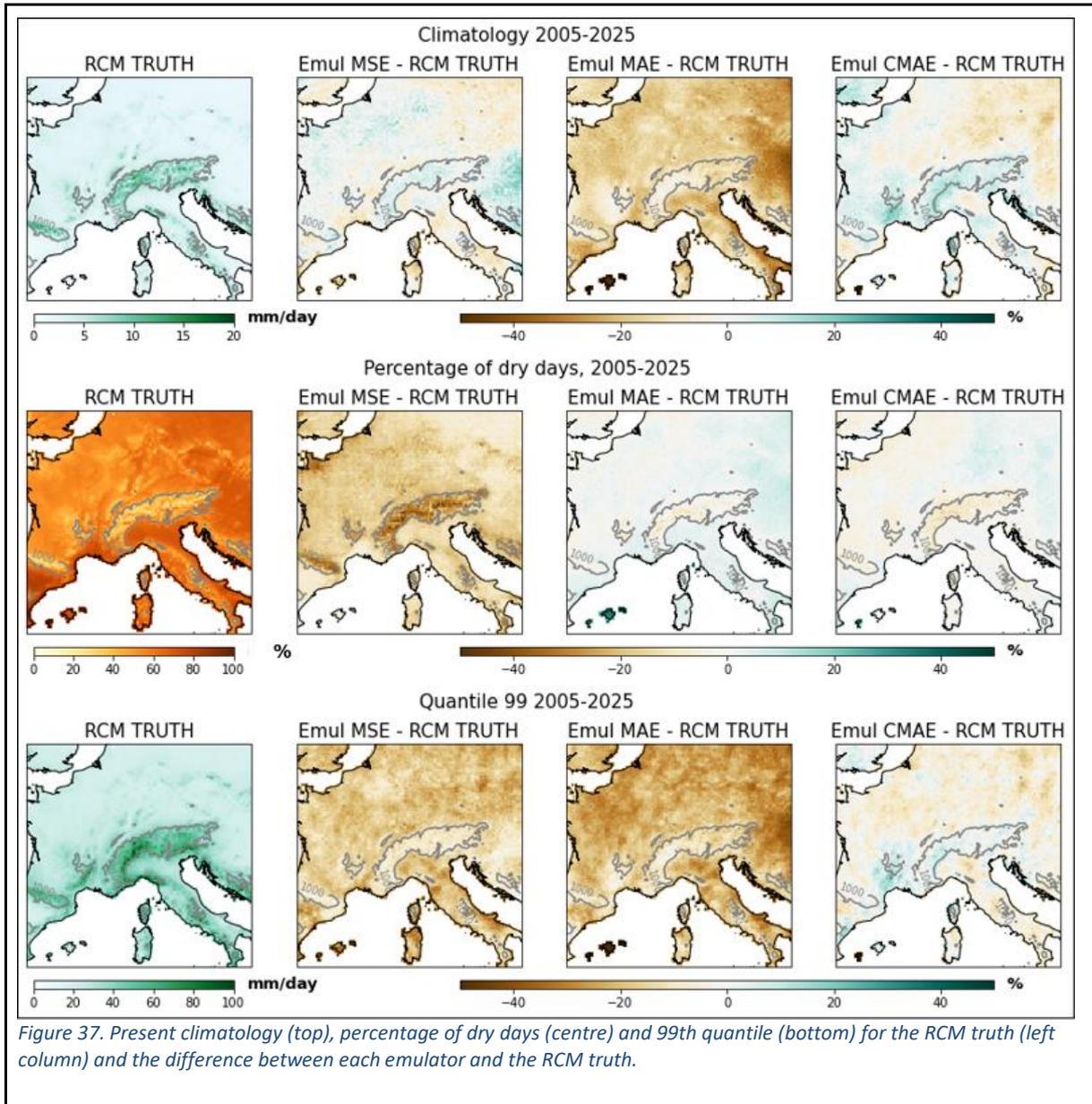
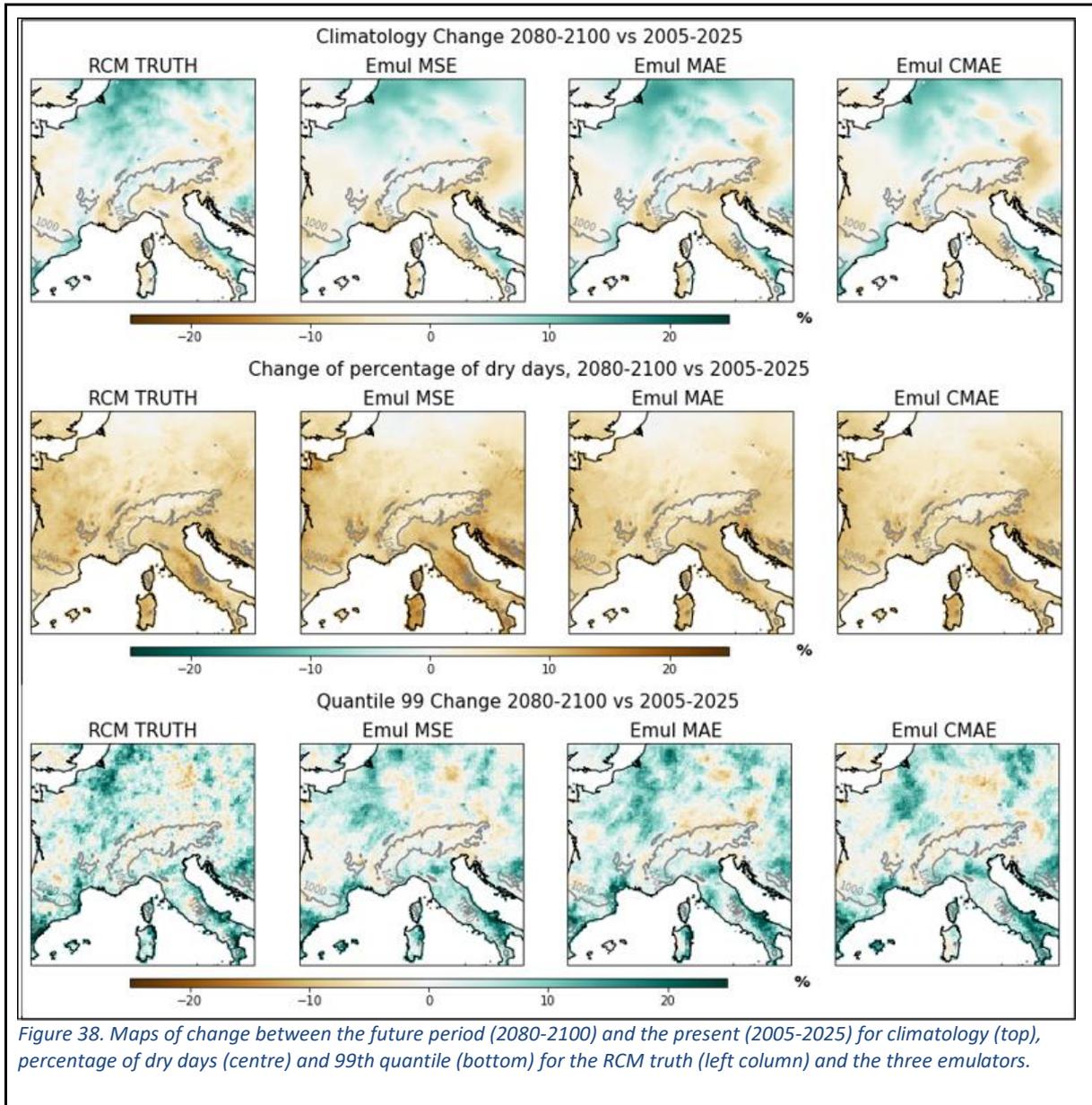


Figure 36. Precipitation map of a random day (Top) and time series of a random period (Center) at a random point in the north of Italy (in red in the top left map) for the RCM truth, and the three versions of the Emulator. Domain-pooled and normalised (Berthou et al 2020) Probability Density Function over the entire domain, a Alps and Paris box (shown on the top right map) and the Fractional Contribution Index (Berthou et al 2020) for the three Emulators with respect to the RCM truth.





Unet-Emulator for temperature over an Alpine domain: Impact of the learning period length

We present here some preliminary results on the impact on the length of the training set used to calibrate the emulator. The emulator used here is exactly the one introduced in Doury et al (2022), with the MSE loss function choice and targets to reproduce the near surface temperature over the same Alpine domain as in the previous section. The aim here is to give a first hint about the performances of this emulator according to the length of the data used for calibration. This sensitivity study is particularly relevant for the future development of CPRCM-emulators as only relatively short simulations (~10 years) can be performed currently with such models (see for example EUCP-WP3 runs). To do so we trained four emulators, using each time data from the historical run and the RCP85 projections and increasing the length of the calibration set:

- 2 times 10 years, 1996-2005 and 2090-2099

- 2 times 20 years, 1986-2005 and 2080-2099
- 2 times 40 years, 1976-2015 and 2060-2099
- Full period: the entire 1951-2100 period.

Then we compared the prediction of these four emulators for the 2006-2100 RCP45 projection in perfect model world, following the evaluation protocol of Figure 35. The results are illustrated in Figure 39 and Figure 40.

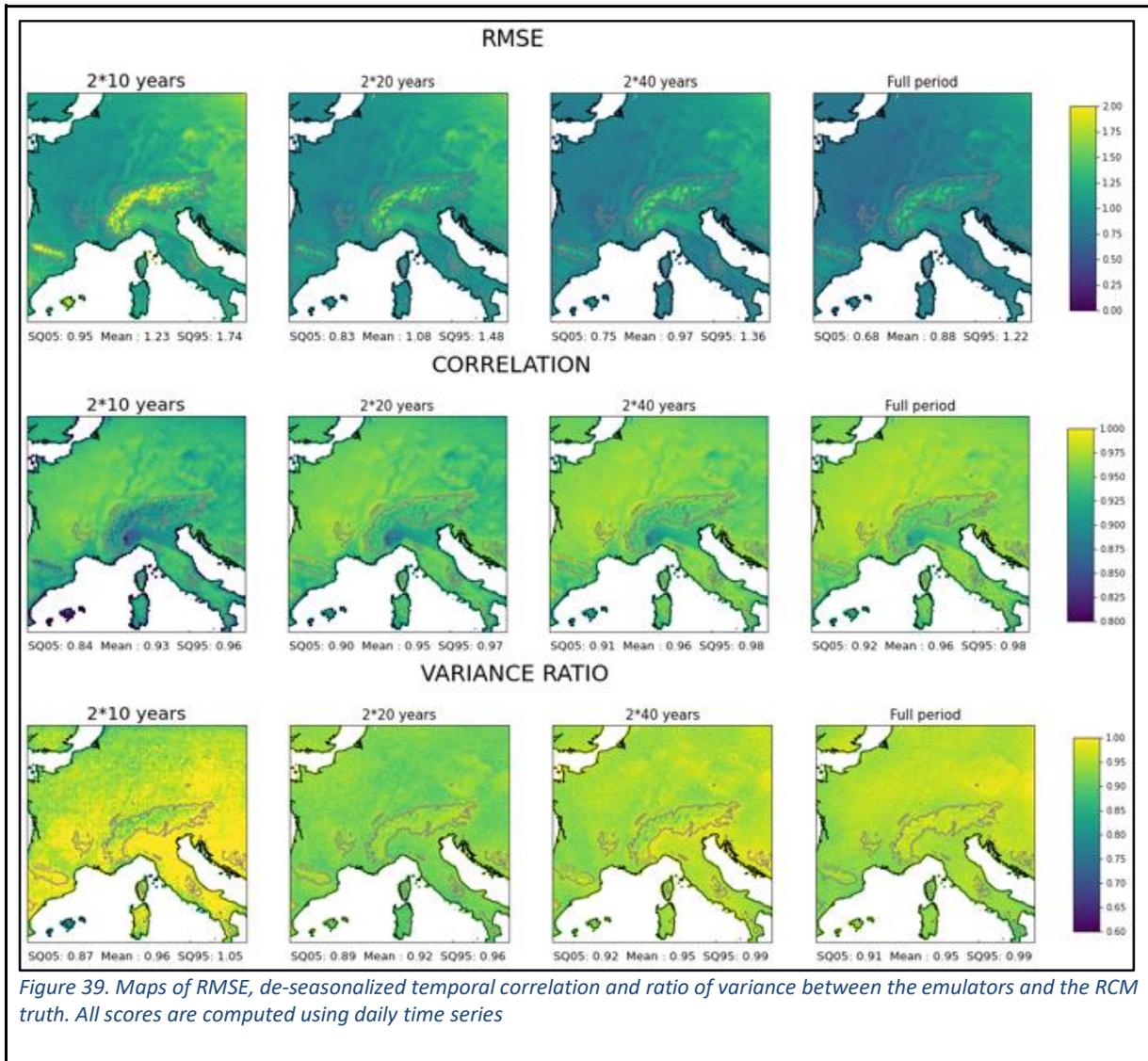
First of all, Figure 39 shows 3 point-wise scores: the RMSE, the temporal correlation and the ratio of variance between the emulated series and RCM reference. Some summary statistics are given under each of these maps: the mean and the super-quantile 05 and 95, which is the average of all values outside of these quantiles.

The four emulators present reasonable results, in each case they seem to have a good temporal correlation and RMSEs. However, it seems clear here that the longer the training set, the better the emulator. The RMSE is probably the best example, it constantly decreases as the length increases. The variance ratio for the emulator trained on 2 times 10 years appears to be a lot noisier than the others. It seems that 10 years in the past and in the future were not enough for the emulator to learn a stable relationship between large scale predictors and the high-resolution temperature. This hypothesis seems to be confirmed by Figure 40 where the climatology of the first emulator seems much noisier than the others, as well as the quantile map. The three other emulators seem to present more comparable results. They all three have a good representation of the future climatology and increasing the length of the period seem to slightly improve the results. Regarding the reproduction of the future 99th quantile, the emulator using the full training period has better results even if it also underestimates the warmest extremes. Finally, it is worth noting that all emulators are reproducing a coherent warming in winter with respect to the reference. They show a good spatial pattern with a greater warming over the reliefs and an important hot spot in the North-East of the domain. However, all emulators slightly underestimate the intensity of the warming even if a longer training period seems to correct this aspect.

It is worthwhile to mention that we trained emulators with different 10 and 20 years periods in order to compare the impact of the RCM internal variability and the one of the length of the training period. The results (not shown) indicate that the length of the training set has a stronger impact on the quality of this emulator than the choice of the periods, as long as they belong to a similar climate.

To conclude, the results here tend to show that using two periods of 10 years to train the emulator are not enough, as the neural network probably did not see enough diversity in the example shown. Two periods of 20 years seem to lead to more stable and coherent results, but the emulators trained with longer periods are better. As expected, the length of the training period has a strong impact on the emulator and the longer it is the better the emulator is. Based on these sensitivity tests as well as on additional tests performed in Doury et al. (2022), we conclude that emulating the near-surface temperature over this domain requires period of at least 2 times 20 years for the calibration covering relatively cold and warm climates with respect to the targeted emulated period. This can at least partly explain the relatively bad results of the UNet-based emulator in the common framework discussed in the next section for which only 10-year long runs were allowed to match with the EUCP WP3

simulation protocol. Developing CPRCM-emulators based on neural networks will therefore require longer training datasets.



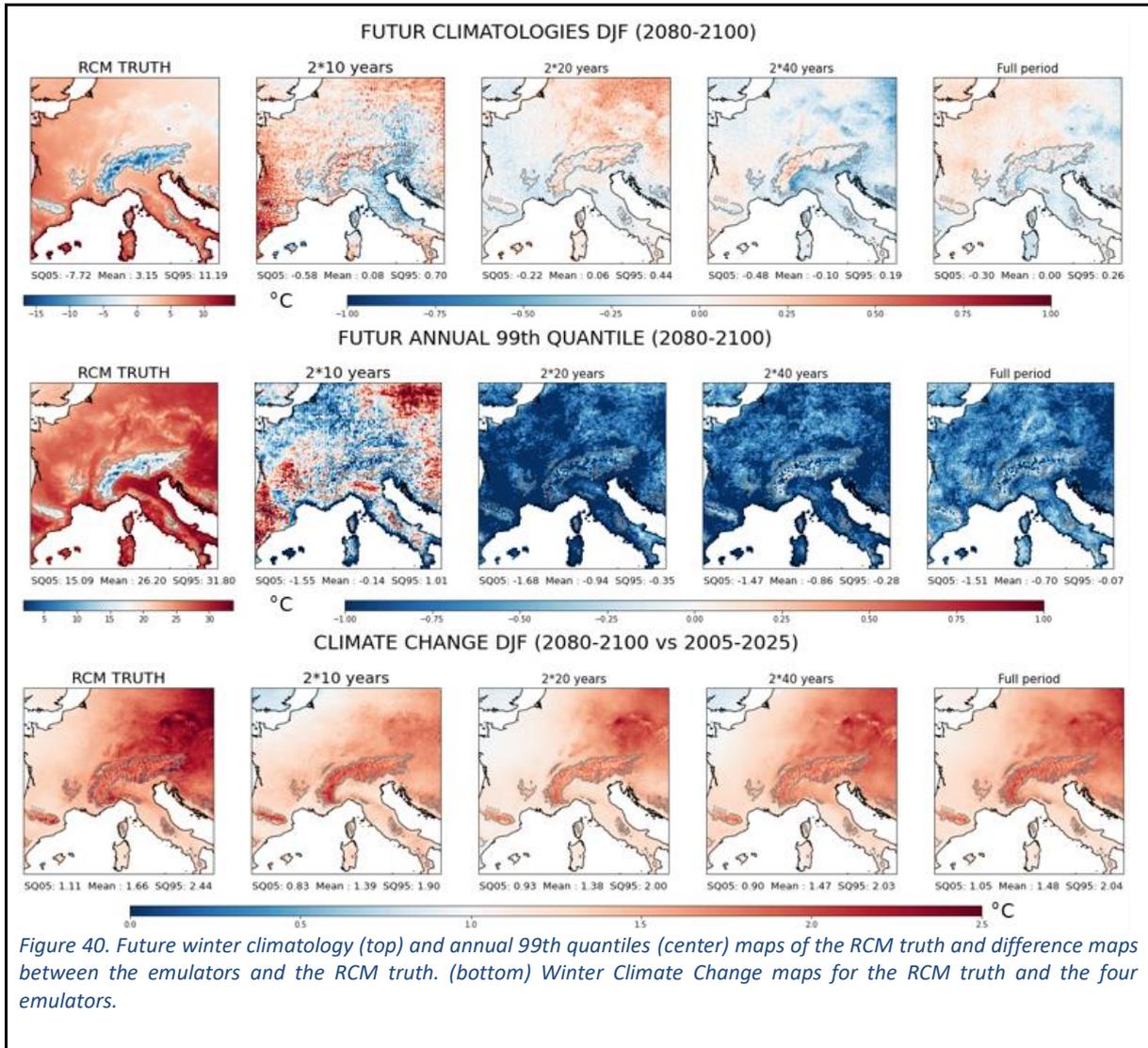


Figure 40. Future winter climatology (top) and annual 99th quantiles (center) maps of the RCM truth and difference maps between the emulators and the RCM truth. (bottom) Winter Climate Change maps for the RCM truth and the four emulators.

Perspectives

The emulator introduced in Doury et al (2022) showed a good ability to reproduce the high-resolution temperature field from the RCM in perfect model conditions which validates the concept of statistical RCM-emulator. We showed here that the emulator could be easily adapted to precipitation with satisfying results, and that a judicious choice of the loss function used to train the network could improve these results. However more work is necessary to deeper understand the impacts of the choice for the loss function, and find maybe a better one. The emulators for precipitation were here only evaluated in perfect model conditions and it is necessary to study its performance when using GCM inputs.

Doury et al (2022) evaluates the emulator for the near surface temperature also in GCM conditions, but using another simulation of the same GCM which drove the RCM simulation used for training the emulator. The emulator had good results in this exercise, but it is necessary to study its stability with other GCMs. This transferability will be an important axis of work for the coming months in order to

address the robustness of the tool and therefore hope to be able to produce very large ensembles of emulated RCM fields.

Another important challenge will be to build an emulator for CPRCM models. In order to build a new downscaling tool as the RCM-Emulator it is necessary to have a large amount of data available in order to explore its strengths and limits, and to fully understand it. For example, it is a key feature to understand how the size of the training sample impacts the performances of the emulator. The EURO-CORDEX matrix is probably the best playground for this exercise whereas the existing CPRCM-based ensembles (EUCP-WP3, CORDEX FPS-convection) remains so far limited in particular due to too short runs. At the same time, CPRCMs are much more expensive than RCMs, so the emulator will probably be a necessary tool to correctly explore all the uncertainties associated with these kilometre scale projections. Finally, emulating sub-daily frequencies, such as hourly precipitation also stays as an open challenge so far.

Publications

Doury, A., Somot, S., Gadat, S., Ribes, A., Corre, L. Regional Climate Model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics* (2022), in review.

3.5 Common framework on emulators (COM)

3.5.1 Introduction and motivations

Within EUCP Task 5.4, a number of different approaches are employed to add in fine-scale information coming from RCMs or CPRCMs on top of large-scale information coming from GCMs. This CPRCMs/RCMs/GCMs information blending has been referred to as “spatial merging” within EUCP. These methods range from very simple scaling approaches to more elaborate statistical approaches based on AI. A number of those blending methods are reported in detail in the other sections of Del 5.4. However, they rely on various training and application frameworks, making it impossible to compare their skills. We therefore decided during the course of the project to propose a *common training/evaluation framework* to test and compare the different methods. This common protocol is based on a perfect-model approach where the hidden truth is known and allows comparison of the various approaches to underline their strengths and weaknesses.

Four institutes from EUCP plus one institute not participating to the project agreed to participate to a common protocol that has been kept on purpose simple and affordable for all the proposed methods: pattern scaling (UCPH, Dominic Matte), ANOVA (DMI, Ole Christensen), regression (KNMI, Hylke de Vries), Emulator UNET (CNRS-CNRM, Antoine Doury), Emulator CNN10 (Uni of Cantabria, J. Bano-Medina). Note that the EUCP common framework protocol has also been proposed to be used in the CORDEX FPS-convection initiative.

3.5.2 Definition of the common framework

The following basic principles guided the proposed common protocol:

- a simple framework allowing all the methods to participate whatever their level of complexity and input variables, and where the truth (that is, the climate to be emulated) is known
- proximity with the EUCP-WP3 framework that focuses on 10-year long CPRCM simulations and the mid-21st century (2041-2050) in order to show the usability of the spatial merging methods in the EUCP context
- an existing Scenario-GCM-RCM-CPRCM quadruplet as virtual truth with enough years for allowing to separate a training period and an independent evaluation period, and with all necessary output available to apply the emulators
- a Scenario-GCM-RCM-CPRCM quadruplet used in WP3 and labelled by the WP2 GCM selection subtasks (constrained future range, GCM performance, ...).

Those basic principles lead to proposing the following *T5.4 common framework to compare Spatial Merging techniques*. The chosen quadruplet is:

- SCEN: CMIP5 RCP85 scenario
- GCM: CMIP5 CNRM-CM5 r1i1p1
- RCM: EURO-CORDEX CNRM-ALADIN63 at a 12km-resolution
- CPRCM: ALP-3 CNRM-AROME41t1 at a 2.5km-resolution

This complete running chain is available and CMORized for the periods 1986-2005, 2041-2060 and 2080-2099, and 3D altitude and 2D surface fields are available for every block of the chain needed for the AI based emulators (the complete list of required input variables can be found in Appendix below). GCM and RCM runs are official CMIP5 and EURO-CORDEX runs available on ESGF and the CPRCM runs have been produced within WP3 of EUCP.

As part of the common framework design, we have the following rules concerning the Training phase:

- The 2041-2050 period should not be used in the training phase
- The methods should be trained on the 10-year long historical run (1996-2005) and on the 10-year long end of the century (2090-2099, RCP85) as this is what will be available for the other CPRCMs within EUCP (optional 20 years of training could be used).

Outputs to be provided by each method:

- High-resolution climatology or climate change maps on the relevant regular grids (ALP-12i grid for the ALADIN RCM, ALP-3i grid for the AROME CPRCM) for the 2041-2050 period, for both the Winter (DJF) and Summer (JJA) seasons.

3.5.3 Short descriptions of the 5 spatial merging methods

- **Pattern scaling (UCPH):** Emulates climate change patterns from existing transient simulation using a linear relationship between patterns of regional climate change and the average global temperature change (see activity SNP3; section 3.3.3).
- **ANOVA (DMI):** Emulates the missing field as an ANOVA-based linear combination of the same field for other GCM/RCM/resolution/scenario/periods (Christensen and Kjellström 2020, 2021). Several setups have been examined: Calculating RCM from GCM; calculating CPRCM from RCM alone and from both (see Appendix A). We note that the other methods emulate the CPRCM pattern directly from the GCM large-scale information.
- **Regression (KNMI):** Emulator based on a simple linear regression using a number of flow indices derived from the GCM fields of mslp, global mean and regional temperature (see Appendix B)
- **Emulator UNET (CNRS-CNRM):** the method is based on Neural Network machine learning technique, in order to produce fine-scale details on top of the GCM simulations. We target RCM or CPRCM spatial scale and daily scale. This model emulator is considered as a hybrid downscaling technique. Contrary to the above-mentioned method, this method is a completely new development performed during the project and is able to produce local daily time series or maps of extreme indices (see activity SME.2; section 3.4.2), and more information in Doury et al. (in rev.).
- **Emulator CNN10 (University of Cantabria, voluntary contribution):** similar method as for CNRS-CNRM but based on another neural network architecture. The CNN1 and CNN10 models were introduced in a recent study dealing with the use of convolutional neural networks for “perfect-prognosis” statistical downscaling (Bano-Medina et al., 2020). These topologies are trained to establish an empirical mapping between a set of predictor variables and the parameters of conditional distributions - Gaussian for temperature and Bernoulli-Gamma for precipitation - by minimising the negative log-likelihood. The CNN1 (CNN10) model consists of an input layer with the predictor fields stacked as channels, feeding a 3-convolutional-layer stage of 50, 25 and 1 (10) filter maps with a kernel size of 3 by 3, respectively. Lastly, for temperature (precipitation) the last hidden layer is fully-connected to the parameters of the Gaussian (Bernoulli-Gamma) distribution, resulting into estimates of the mean and variance (the probability of rain, and the shape and scale factors of the Gamma density function) at each predictand site per sample. We can then either compute the expectance (deterministic) or sample (stochastic) from these estimated conditional distributions. For the CNN1 and CNN10 we use five variables (geopotential height, the specific humidity, the air temperature and the zonal and wind air velocities) at three pressure levels (500, 700 and 850 hPa) plus the sea level pressure. Regardless of the predictor dataset - either the upscaled RCM or the GCM, - we standardise the predictors at a grid-box level with respect to the baseline period 1996-2005. Nevertheless, following recommendations in previous works (Bano-Medina et al 2021), we apply a harmonisation process to the CNRM-CM5 variables prior to the standardisation stage, which bias adjusts the monthly mean towards the training predictor fields. This method will be called CNN10 in the following.

The methods differ widely in the used information for the training and application phases and also in what they predict (see a synthesis in the Table 3 below). The simpler methods (Pattern Scaling, ANOVA) basically predict a change signal; in the case of pattern scaling the change signal is derived from the change at the end of the century and scaled back with the ratio of global rise at the middle of the century compared to end of the century. In contrast, Regression and Emulators predict the actual state of the 2041-2050 or even the full daily time series of the period. We are aware that this may lead to a not-completely fair comparison.

Table 3. Synthesis of the information used by each method with the following notations: G: GCM run information, R: RCM run information, 0: present climate, 1: mid-21st century, 2: end-21st century. <> global temperature average.

Method	Used Information for Training	Used information for the Emulation	Emulated field
Emulator UNET Emulator CNN10	R0,R2	G1	R1, daily time series
Pattern Scaling	<G0>,<G2>,R0,R2	<G1>	R1-R0, climatology
ANOVA	G0,G2,R0,R2	G1	R1-R0, climatology
Regression	G0,G2,R0,R2 <G0>,<G2>	G1	R1, daily time series

3.5.4 Results

Based on the common framework developed above, we illustrate here some results obtained for each method for a given variable, a given season and a given model to be emulated. As showing the 2 seasons for 2 variables, 2 models and 5 methods plus the benchmark and the virtual truth represents too many figures for this deliverable, we first propose two contrasted illustrations (Figure 41 and Figure 42) before summarising all the cases in Taylor diagrams in Figure 43. The emulated results are compared to the RCM or CPRCM virtual truth for the same period and to a very simple benchmark, the GCM climate change map, linearly-interpolated on the RCM or CPRCM high-resolution grid. On the figures and for each method, we show the seasonal-mean climate change signal between the emulated 2041-2050 period and the RCM (or CPRCM) 1996-2005 period. We recall that the RCM or CPRCM data for the period 2041-2050 have not been used for the training.

Figure 41 shows the climate change signal in winter (DJF) precipitation for the mid-21st century for the RCM CNRM-ALADIN63. For this near-term projection, the RCM run shows a weak increase over the domain (mean value = +0.39 mm/d) with local stronger increases such as over the Alps (SQ95=1.33 mm/d) and with decreases in few places (SQ05<0). The interpolated GCM field allows to underline the existence of small-scale patterns in the RCM that are not simulated by the GCM, often related to the mountain ranges (see also activity SNP1; section 3.3.1). Note in particular that the GCM map is poorly correlated with the RCM map and that the driest or wettest values of change are underestimated by the GCM. The general RCM pattern is reproduced by all methods with specific strengths and weaknesses and with none of them showing a very good skill. For example, ANOVA simulates well the

mean value, the enhanced wettening over the Alps and shows a relatively good RMSE but fails to reproduce the drying areas ($SQ_{05} > 0$). The Pattern Scaling approach shows the best spatial correlation and a RMSE as good as ANOVA but underestimates the enhanced wettening over the Alps. Regression shows overall good mean and super-quantiles values but very low spatial correlation, whereas the Emulator UNET has the opposite behaviour with a good correlation but an underestimated mean change and a strong overestimation of the spatial variability with too strong drying over the Mediterranean area and too strong wettening over the Alps. The Emulator CNN10 is generally too wet in the mid-century period with respect to the RCM and has a too strong spatial variability as Emulator UNET.

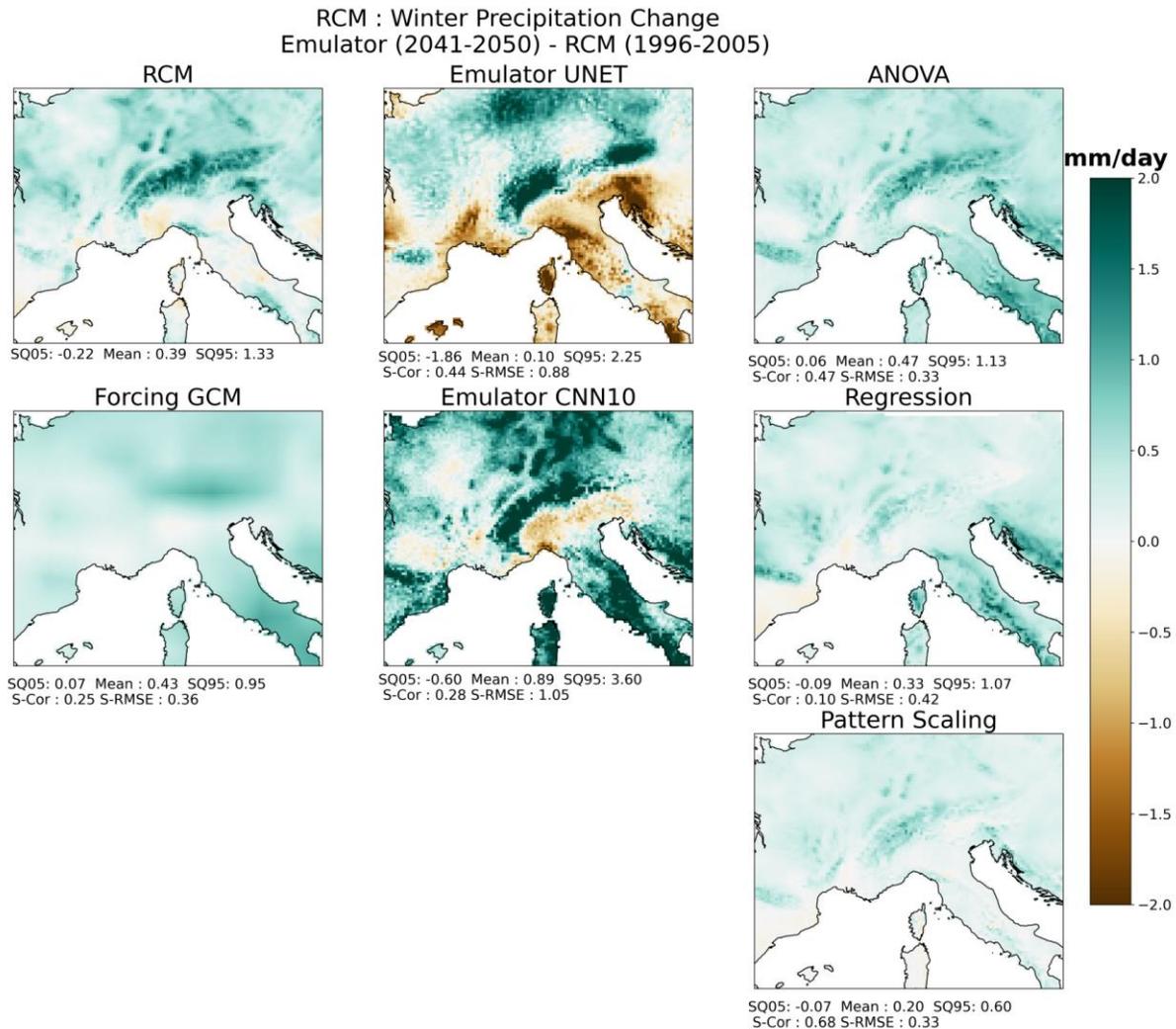


Figure 41. Winter precipitation climate change maps with respect to the historical RCM truth for the RCM (truth) and the 5 methods. The GCM climate change maps, linearly-interpolated on the RCM grid, is shown as a simple benchmark. For each map the mean, the 5th and the 95th super-quantiles (SQ) are given. For the 5 methods and the interpolated GCM, we also give the spatial correlation and RMSE with respect to the RCM map.

Figure 42 shows the climate change signal in summer (JJA) near-surface temperature for the mid-21st century for the CPRCM CNRM-AROME41t1. For this near-term projection, the CPRCM run shows a $+0.7^{\circ}\text{C}$ mean warming over the domain with a noticeable spatial variability, leading to local stronger increases mostly over the mountain tops ($SQ_{95}=1.44^{\circ}\text{C}$) and with lower increases over the Northern part of the domain and in the valleys ($SQ_{05}=0.22^{\circ}\text{C}$). Comparing the CPRCM map and the

interpolated GCM map is another way to underline the existence of small-scale patterns in the CPRCM, not simulated by the GCM. In particular, the interpolated GCM is poorly correlated with the RCM signal. The CPRCM pattern is well reproduced by the ANOVA method with good mean and extreme values, high spatial correlation, and low RMSE. The Pattern Scaling and Regression approaches show relatively good skills in reproducing the CPRCM pattern but overestimate the mean and extreme values of the warming. The Emulator UNET shows generally lower skills with an overestimated warming and a very noisy map, leading to very strong local warming (SQ95=3.28°C). This noisy pattern is likely attributable to a too short training period for the Emulator UNET. Despite its methodological proximity, the Emulator CNN10 does not show this noise issue and has a similar range of warming over the domain as the CPRCM, but presents a different spatial pattern than the model, in particular with a low spatial correlation.

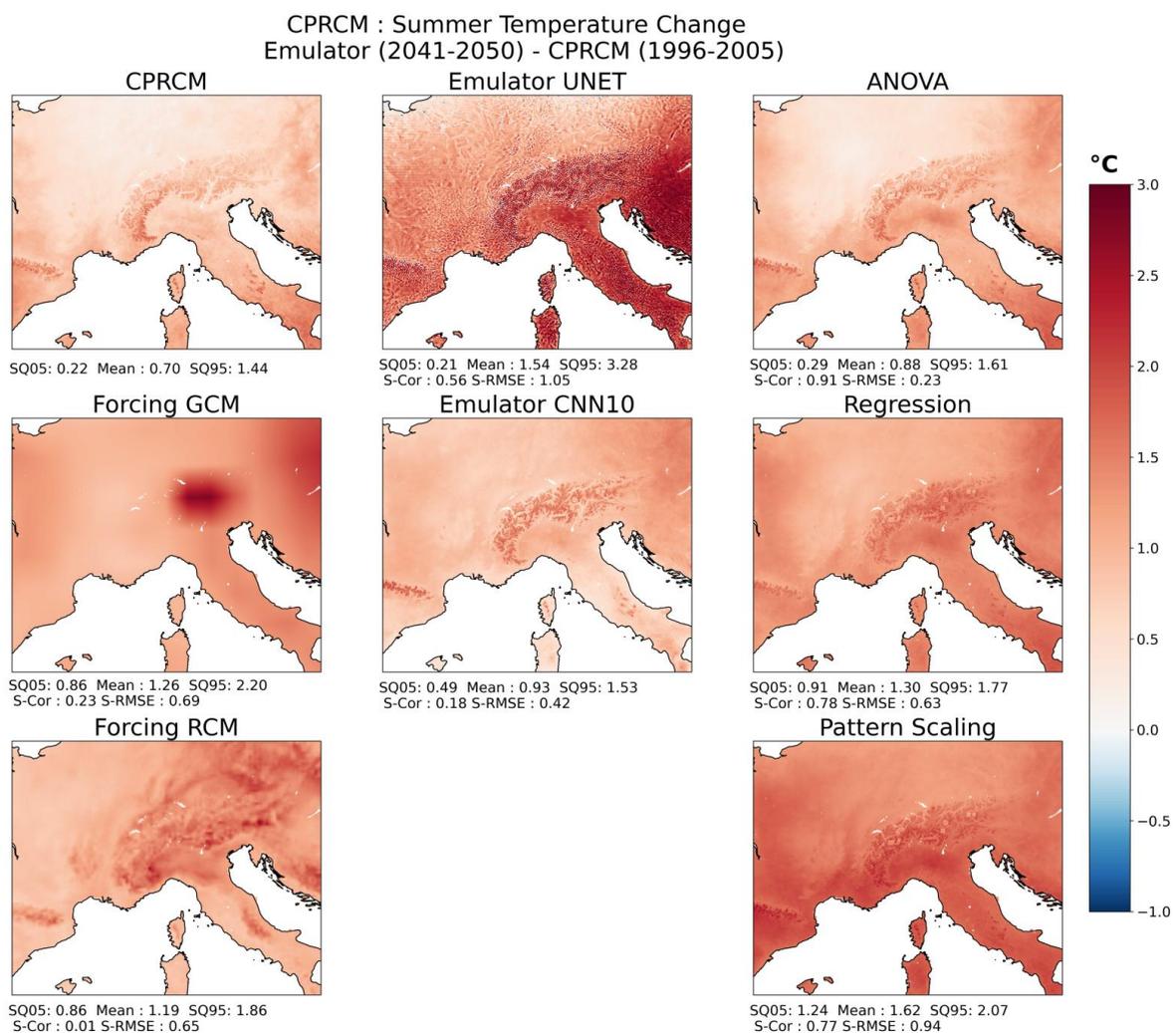


Figure 42. Same as Figure 41 but for Summer temperature and the CPRCM. The RCM climate change map, linearly-interpolated on the CPRCM grid is also shown.

Figure 43 summarises the results of the 5 methods in spatial Taylor diagrams for both variables (near-surface temperature, precipitation), both seasons (DJF, JJA), both models (RCM ALADIN and CPRCM AROME). This figure shows that none of the methods is really close to the model truth, but that Pattern Scaling and ANOVA seem to be the best approaches to emulate the high-resolution mean climate change maps and the Emulator UNET is the worst. Generally speaking, the methods designed to produce climatology maps (Pattern Scaling) get better results than the methods emulating times series (Regression and Emulators). The Emulator UNET shows its best results for the ALADIN temperature which corresponds to the variable and model used in Doury et al. (2022). This new method is probably not yet mature enough to emulate CPRCM or precipitation fields and further work is required. The Emulator CNN10 performs quite similarly which is expected as the methods are close, even if it generally shows lower bias than the Emulator UNET which is likely due to the different input pre-processing steps. It is worth noting that for the Pattern Scaling and ANOVA methods and contrary to what is expected, emulating the CPRCM or emulating the precipitation field does not seem to be more complicated than emulating the RCM or the temperature.

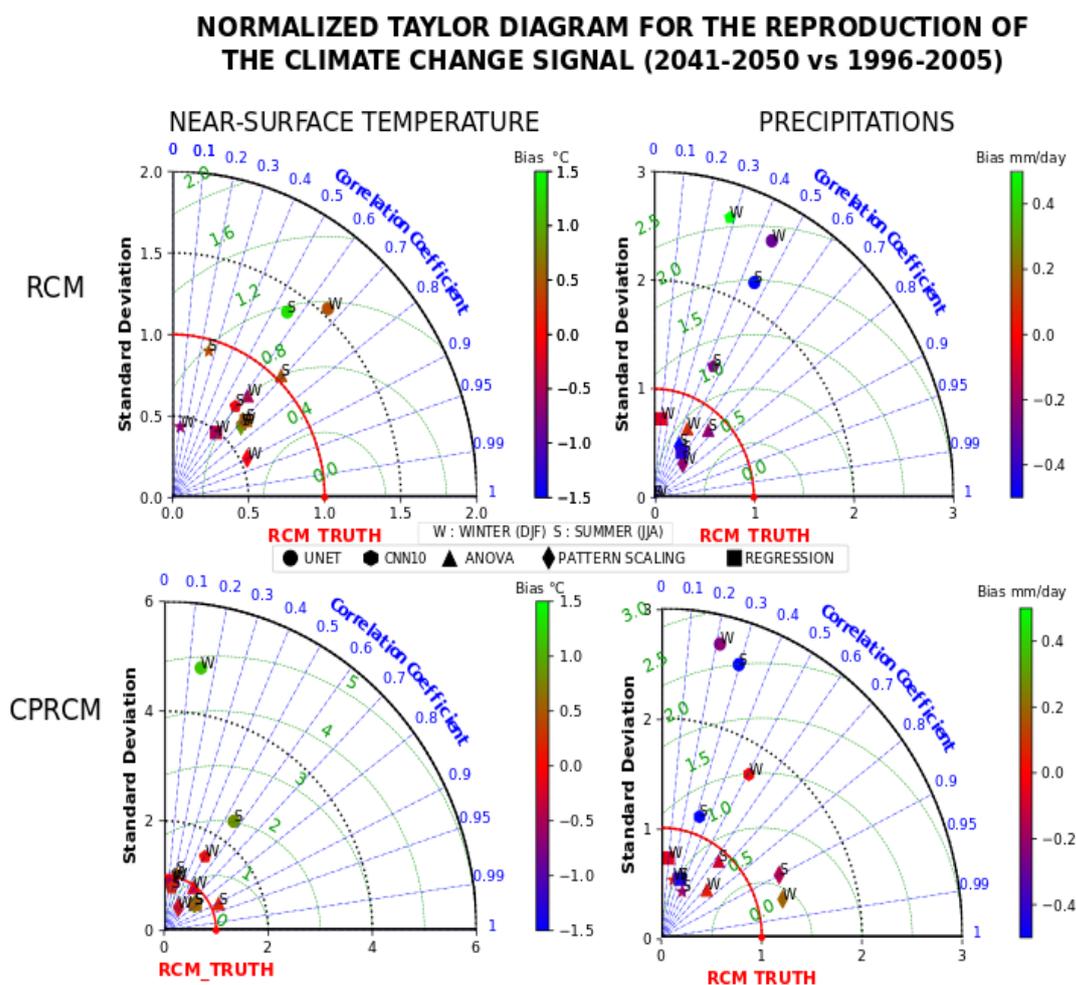


Figure 43. Normalised Taylor diagrams for the reproduction of the RCM (first line) and CPRCM (second line) climate change maps by the 5 methods for the summer (JJA) and winter (DJF) temperature (first column) and precipitation (second column). The Taylor diagrams show the ratio of the spatial standard deviations, the normalised spatial RMSD and the spatial correlation. The methods are labelled by different marker shapes, the colour scale represents the spatial-mean bias value and the seasons are mentioned by "W" (winter) or "S" (summer) next to the markers.

3.5.5 Lessons learnt and range of applicability of each method

To our knowledge, this is the first time that a common intercomparison framework is proposed to compare spatial merging methods or in other words, statistical methods targeting to complete missing periods of time in RCM simulations. This is also the first time that such methods are applied to CPRCM simulations. This means that the common framework proposed is possibly not the optimal one and may be refined in future studies.

The results show that the simple approaches (ANOVA, pattern scaling) are able to correctly reproduce the mean value and the spatial pattern of climate change for the emulated time period. Figure 43 shows that it is difficult to determine which is the best method between ANOVA or pattern scaling in our framework as results depend on the variable, model and season. Note that ANOVA uses the information from the RCM in addition to the information from the GCM when it emulates the CPRCM results. This is not the case for the other methods that emulate the CPRCM pattern directly from the GCM large-scale information.

The more advanced methods (regression, IA-based emulator) capture part of the climate change signal but generally underperform, showing mean biases and sometimes strange behaviours. Fully explaining the difficulties of those approaches to emulate the climatological state of the missing period is out of reach with the current protocol. We can however propose some hypotheses:

- (1) The model emulators target to emulate the daily time series in their full complexity. So, the climatology state is not their main target and is a by-product of these methods. Small errors in the reconstruction of the daily time series can lead to substantial errors in the climatology. This could be the case since correlations in day-to-day weather variability may not correspond well to the relations seen on longer time scales, such as in a climate change context. Therefore, methods optimised for day-to-day variability may well suffer in climatological sense.
- (2) In addition, the model emulator learns the downscaling function within the RCM or CPRCM and does not learn the way the RCM or CPRCM modifies the large-scale pattern of the GCM. So, if this large-scale transformation is strong within the training period or the emulated period, the emulator will add small scales on the top of the GCM large-scale but will not reproduce the RCM or CPRCM full field. See Doury et al. (2022) for a detailed discussion of this point.
- (3) The emulator targets to reconstruct the daily time scale of the high-resolution model in a varying climate (daily variability, seasonal cycle, climate change). To do so, it needs to “see” a large number of days with varying large-scale / small-scale relationships during the training. The current protocol that relies on the EUCP WP3 framework (10-year long time slices with CPRCMs) is therefore probably not adapted to train emulators with too short available training periods. Longer training dataset as in Doury et al. (2022) are probably required for better scores with this method. See also activity SME2 in section 3.4.2 of the current deliverable for sensitivity tests to the training period length.
- (4) Methods that target to predict a climate state (such as the Emulators or the Regression) have a disadvantage as compared to methods that primarily predict the change signal. This is because the climate change signal for these methods is derived from the emulated future state compared to the true (not emulated) control climate. It might have been better to compare to the emulated control period, but this option was not available for all methods.

The common framework proposed here targets to compare the spatial merging methods for only one of their possible applications, that is to say, to reconstruct mean climate change maps for a missing time period of a given Scenario-GCM-RCM triplet or Scenario-GCM-RCM-CPRCM quadruplet. However, some of the methods have a wider range of applications (see Table 4). Examples of additional applications are the emulation of small-scale information for:

- daily variability and extreme indices for missing time period for the same model chain
- full daily time series for missing time period for the same model chain
- other members for the same GCM
- other emission scenarios for the same GCM
- other GCMs, allowing to fill Scenario-GCM-RCM matrices
- Specific chosen storylines such as in EUCP tasks T2.4 or T5.5.

Note that even for the most advanced methods, the emulated product is not generic but depends on the RCM or CPRCM of the training set as the large-scale/small-scale relationship is model-dependent. For example, the Emulator UNET shown in Figure 41 is a model emulator of the CNRM-ALADIN63 RCM. Developing an emulator for another RCM would require another dedicated training set.

Table 4. : Summary of the potential applicability range for each method

Potential Application	Methods and associated Institute				
	Emulator UNET	Emulator CNN10	ANOVA	Pattern Scaling	Regression
	CNRM	U.Santander	DMI	UCPH	KNMI
climatology maps for missing time periods, same GCM-RCM-CPRCM as in training	Yes	Yes	Yes	Yes	Yes
Daily variance or extreme indice maps for missing time periods, same GCM-RCM-CPRCM as in training	Yes	Yes	Yes but to be evaluated	No	Yes but poor results
Full daily time series maps for missing time periods, same GCM-RCM-CPRCM as in training	Yes	Yes	No	No	Yes
other GCM members for the same GCM-RCM-CPRCM as in training	Yes	Yes			
other emission scenarios for the same GCM-RCM-CPRCM as in training	Yes	Yes			
other RCMs for the same GCM as in training	Need another training set	Need another training set			
other GCMs for the same RCM or CPRCM as in training	Yes but to be evaluated	Yes but to be evaluated			
emulate EUCP-selected storylines	Yes	Yes			

3.5.6 Appendix A: Method ANOVA (DMI)

The common framework challenge is to emulate a high-resolution field in the middle of the century, based on various input fields from higher-resolution simulations for three periods: present day, middle of the century and end of century, as well as fields from the high-resolution simulation for the first and the third period. The ANOVA-based filling method is very simple, and only uses the target fields themselves, i.e., 5 or 8 mean fields to find the 6th or 9th in this setting. In this study we have looked at surface air temperature (T) and mean precipitation (pr) in three settings:

1. Calculating RCM from GCM and RCM
2. Calculating CPRCM from both GCM and RCM
3. Calculating CPRCM from only RCM

This has been done for both winter (DJF) and summer (JJA).

Method

As a first step, all fields are put on the CPM grid with bilinear interpolation. Each point, field and season are treated independently, in the same way, in this method. We have fields $Y(i,j)$ where Y is the seasonal field in each point. $i=1,3$ is the period, 1996-2005; 2041-2050; 2090-2099. $j=1,3$ is the model resolution GCM; RCM; CPM.

- Setting 1 (Tier 0): Find the target RCM field from all GCM and the remaining RCM fields, i.e., $Y(2,2)$ based on the other 5 fields from $Y(1:3,1:2)$
- Setting 2 (Tier 1): Find CPRCM fields from both GCM and RCM and the remaining CPRCM fields, i.e., $Y(2,3)$ based on the other 8.
- Setting 3 (modified Tier 1): Find CPRCM from only RCM and CPRCM data, i.e., $Y(2,3)$ from the remaining 5 fields in $Y(1:3,2:3)$.

According to standard Analysis-of-Variance (ANOVA) procedure, we split each field into

$$Y(i,j) = M + T(i) + R(j) + TR(i,j)$$

where all indexed variables sum to 0 over each index. There is exactly one way to do this in the case where all values of Y are known. M is the mean over the ensemble; T depends on time and is the mean anomaly from M across resolution for the given time. Correspondingly, R is a resolution term signifying the time-mean deviation of the given resolution from M . Finally, TR is a cross term containing the individual simulation's deviation from the linear expectation and includes any noise.

The "hole filling", calculation of the missing data from the known data, follow Christensen and Kjellström (2021): Since the cross terms are unknown for the missing simulation, we set the TR term to zero, i.e., we let the missing simulation be a sum of linear terms. To find the missing simulation $Y(2,2)$ we solve $TR(2,2) = 0$ keeping all summing rules.

For settings 1 and 3, finding RCM (CPRCM) from GCM (RCM) this reduces to an equation like

$$Y(2,2) = Y(2,1) + (Y(1,2) - Y(1,1) + Y(3,2) - Y(3,1))/2 \quad (1)$$

In other words, we find the GCM -> RCM transfer function from the two known pairs and use it on the middle part of the century.

For setting 2, finding the missing simulation $Y(2,3)$ from both GCM and RCM, we solve $TR(2,3) = 0$. This reduces to:

$$Y(2,3) = (Y(1,3) + Y(3,3) + Y(2,1) + Y(2,2))/2 - (Y(1,1) + Y(1,2) + Y(3,1) + Y(3,2))/4 \quad (2)$$

$$= (Y(2,1) + (Y(1,3) - Y(1,1) + Y(3,3) - Y(3,1))/2)/2 + (Y(2,2) + (Y(1,3) - Y(1,2) + Y(3,3) - Y(3,2))/2)/2$$

In words, find the GCM->CPM transfer function from the two known pairs of GCM-CPM; do the same for RCM->CPM, and average the two.

Results

We will now look at T and pr, DJF and JJA. In order to have a point of comparison we also show simple interpolation of the two known high-resolution fields with the regional average of temperature with the area mean temperature from the GCM simulation, i.e., a regional pattern scaling not employing ANOVA.

First, in Figs COM.A1-4 we compare the simple regional GCM temperature-based interpolation between the present-day (now) and the end-of-century (end) for the target in the middle of the century (mid) with the emulation from GCM and RCM to the unknown RCM time slice. We show the deviation from the true simulated field in both cases. The quantities being subtracted between interpolation and truth are absolute climate change in temperature, and relative climate change in precipitation (percentage).

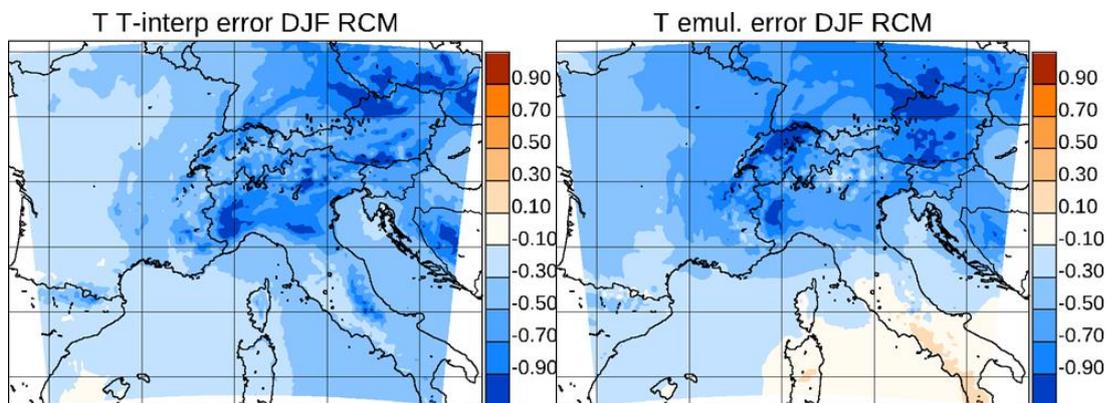


Figure A1. Winter RCM temperature emulation error (degrees K) for simple interpolation (left) and for setting 1 (right).

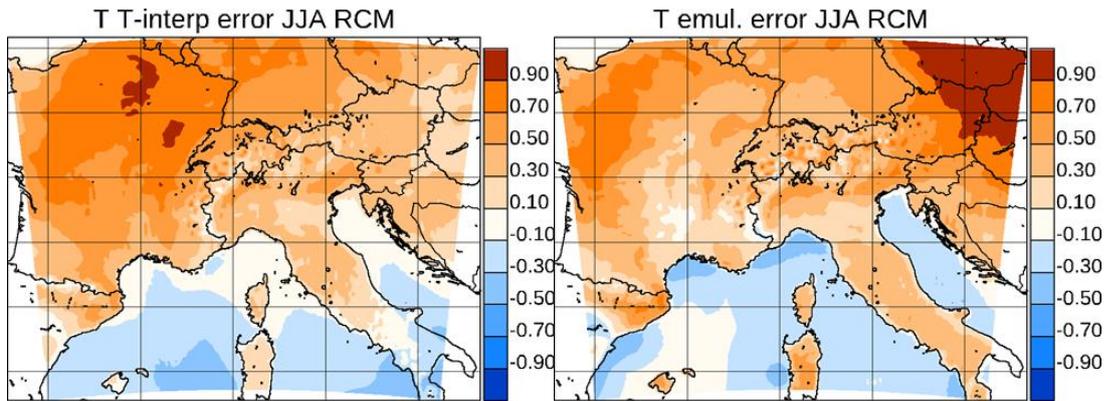


Figure A2. Summer RCM temperature emulation error (degrees K) for simple interpolation (left) and for setting 1 (right).

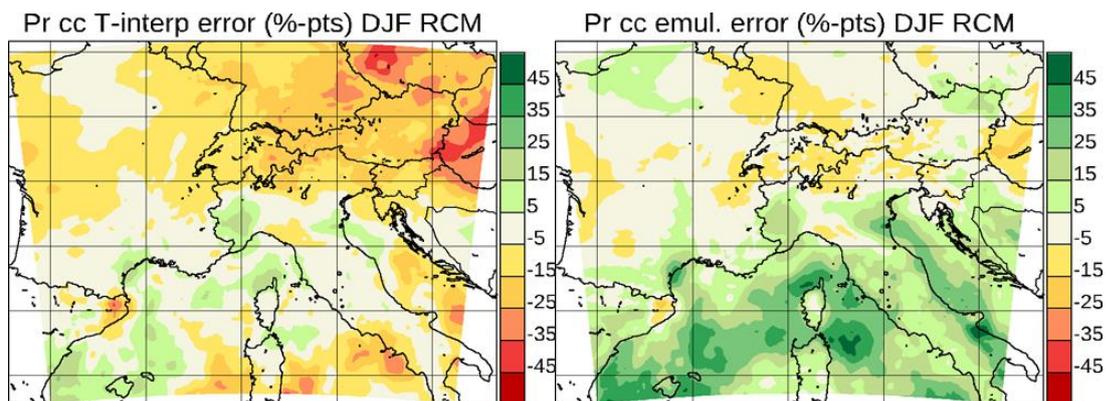


Figure A3. Winter RCM precipitation emulation error (percentage) for simple interpolation (left) and for setting 1 (right).

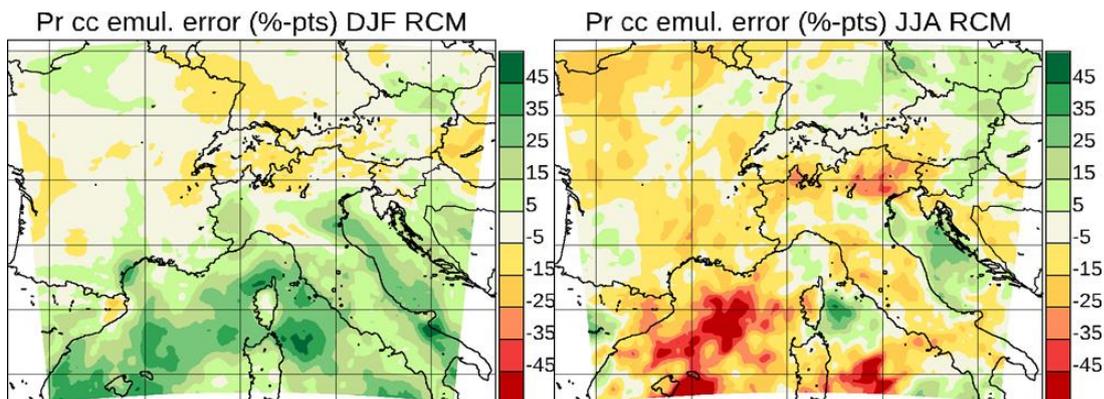


Figure A4. Summer RCM precipitation emulation error (percentage) for simple interpolation (left) and for setting 1 (right).

From Figs COM.A1-2 it can be seen that the actual mid-century winter temperature of the RCM simulation is warmer than would be expected from the corresponding GCM mean temperature evolution, and conversely colder in winter. This fact cannot be resolved by the emulation method, and the two methods show quite comparable quality. One exception is the Mediterranean Sea, where the emulation is closer to the actual simulation than the interpolation. This is due to the fact that RCM sea-surface temperature is directly interpolated from the GCM; therefore, the GCM-RCM transfer function for sea points will be close to the identity function, and that the mid-century GCM temperature will be very close to both the emulated values and the actual RCM simulated values.

For precipitation, the emulation tends to have more precipitation than the interpolation. For winter, the errors are of comparable magnitude between the methods, whereas there is some improvement for summer.

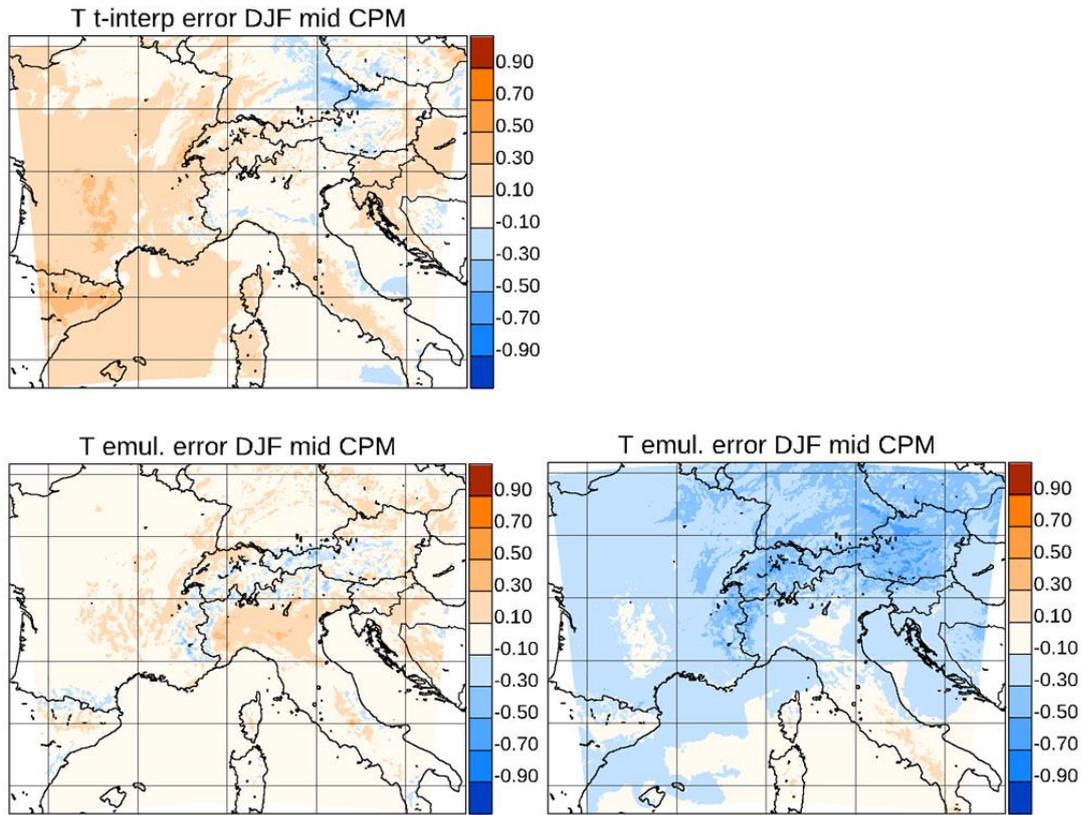
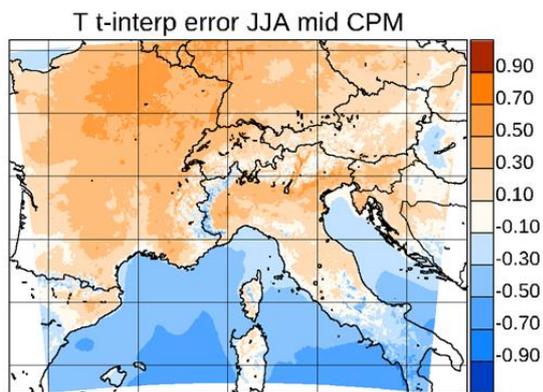


Figure A5. Winter temperature emulation error (degrees K) for simple interpolation (top), for setting 3 (bottom left), and for setting 2 (bottom right).



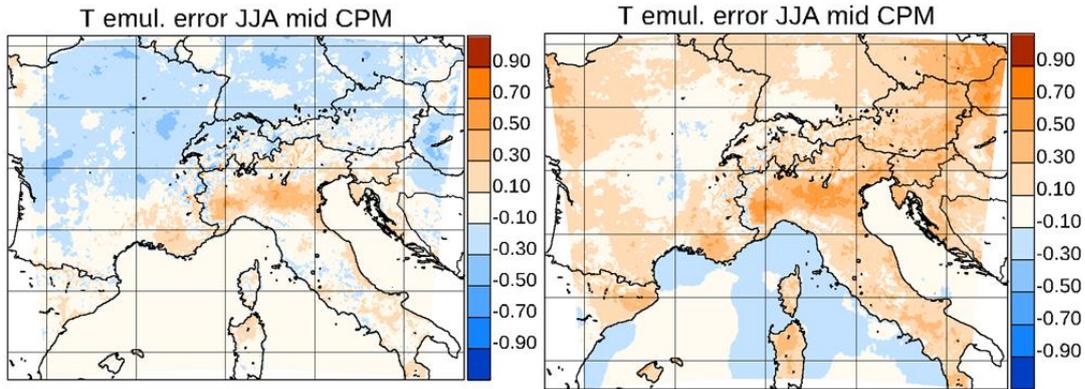
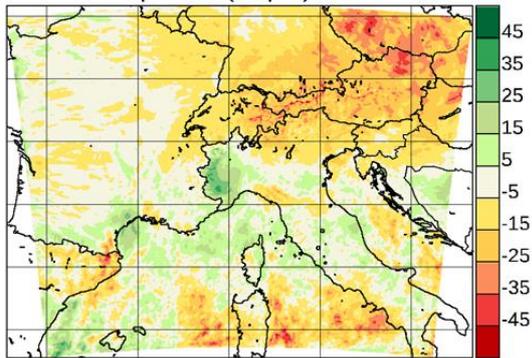
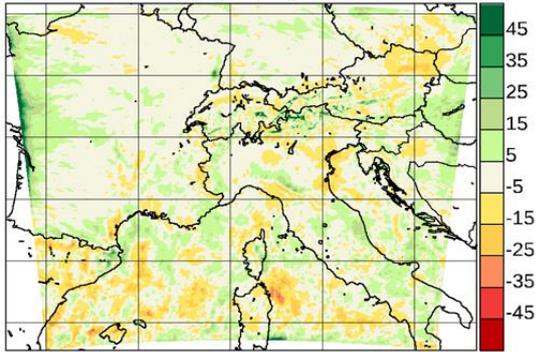


Figure A6. Summer temperature emulation error (degrees K) for simple interpolation (top), for setting 3 (bottom left), and for setting 2 (bottom right).

Pr cc T-interp error (%-pts) DJF mid CPM



Pr cc emul. error (%-pts) DJF mid CPM



Pr cc emul. error (%-pts) DJF mid CPM

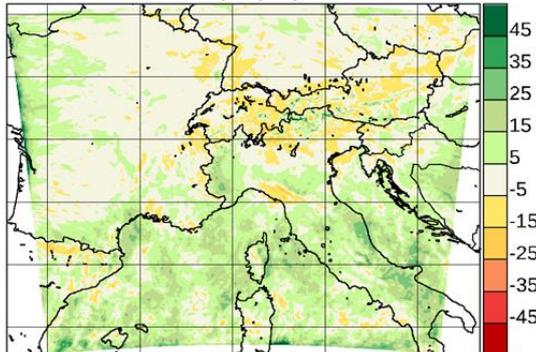
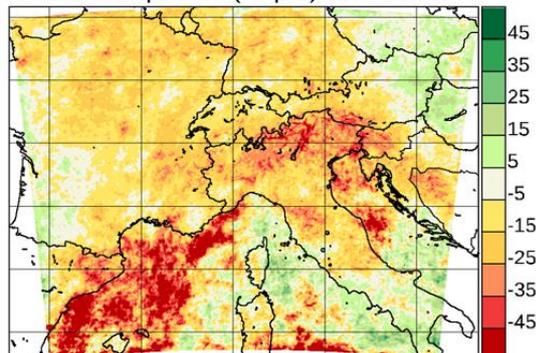


Figure A7. Winter error in relative precipitation change emulation (percentage points) for simple interpolation (top), for setting 3 (bottom left), and for setting 2 (bottom right).

Pr cc T-interp error (%-pts) JJA mid CPM



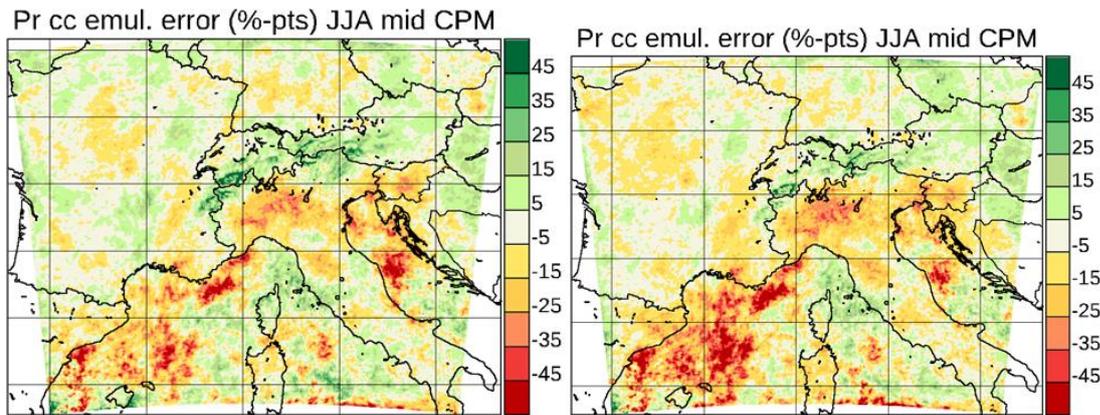


Figure A8. Summer error in relative precipitation change emulation (percentage points) for simple interpolation (top), for setting 3 (bottom left), and for setting 2 (bottom right).

In Figs A5-8 we show deviations from the true CPM simulation according to three different emulation methods: In the top panels we show GCM regional temperature-based interpolation of the now and end periods of the CPM simulation. In the bottom left we show ANOVA-based emulation results only using RCM and CPM results (setting 3), i.e., using exactly the same methods as above, just going from RCM and CPM to CPM (mid) instead of going from GCM and RCM to RCM (mid); this follows Eq. 1. In the bottom right panels we use all GCM and RCM results as well as CPM (now and end; setting 2) to obtain CPM (mid) according to Eq. 2.

In this case, emulation of CPM results, there is a clear ranking of methods with only winter temperature being a bit unclear: Both ANOVA-based methods outperform temperature interpolation, and the RCM-only setting 3 outperforms the full setting-2 emulation using all information. In other words, the GCM information is so coarse that it actually deteriorates the emulation of CPM information to include it! This result seems intuitive, as the size of the area under investigation corresponds to spatial scales where we do expect RCMs to perform well, and we expect GCM resolution to be too coarse to simulate spatial structure at this scale.

In conclusion, the simpler setting 3 should provide a good benchmark reference point for process-based and much more advanced methods as discussed elsewhere in this report.

Publication

Christensen, O.B., Kjellström, E., Filling the matrix: an ANOVA-based method to emulate regional climate model simulations for equally-weighted properties of ensembles of opportunity. *Clim Dyn* (2021). <https://doi.org/10.1007/s00382-021-06010-5>

3.5.7 Appendix B: Multiple regression method (KNMI)

Method

The method is based on multiple linear regression using a number of key predictors, following the work of van Oldenborgh et al (2008). We exploit the role of the large-scale circulation on local temperature and precipitation. A local regression model for RCM-temperature $T_{rcm}(x,y,t)$ is constructed as follows (van Oldenborgh et al, 2008; Lenderink et al, 2009):

$$T^{rcm}(x, y, t) = \alpha_0(x, y) + \underbrace{\alpha(x, y)T_g(t) + \sum_i \beta_i(x, y)C_i^{gcm}(x, y, t)}_{T_{pred}^{rcm}(x, y, t)} + \epsilon(x, y, t) \quad (1)$$

where T_g indicates (yearly-average) GCM global temperature, C_i^{gcm} are daily GCM-based fields (predictors), and α and β the spatial maps with regression coefficients. The daily GCM fields included here are GCM local temperature, and geostrophic westerly and southerly wind components and geostrophic vorticity derived from the mslp field (using a 20 x 10 degree longitude x latitude box and afterward bilinearly interpolated to the target RCM or CPM grid). The method has been used for identifying circulation-change induced temperature trends (van Oldenborgh et al, 2008). Training is carried out for each season independently and based on a regression over both the control period as well as the future period.

Predictions take the form of a gridded daily time-series. Multi-year seasonal averages are constructed from these and compared to the *truth*, i.e. the multi-year seasonal average as simulated by the RCM or CPM. Note that since the regressions are computed locally, there is no spatial coherence in the resulting daily predictions. Due to the spatial coherence of the GCM circulation fields, the resulting patterns are in practice nevertheless reasonably coherent.

Regression coefficients

The figure shows the RCM regression coefficients of the geostrophic circulation components for summer precipitation (CPM results are similar). They are physically recognizable patterns. Southerly wind is an important contributor (increased/decreased precipitation upstream/downstream of the main Alpine ridge with increasing large-scale winds from the south). A similar analysis carried out for the Scandinavian mountains shows a marked sensitivity to zonal wind.

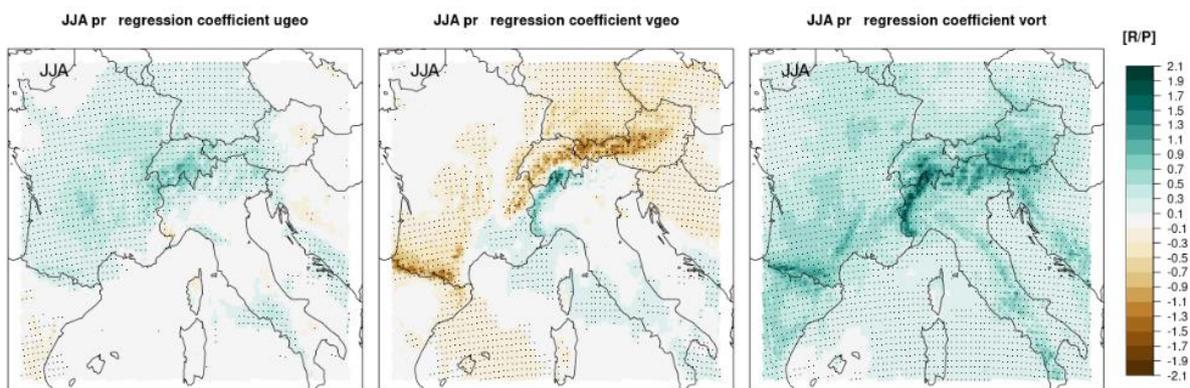


Figure B1. RCM summer precipitation; regression patterns of the geostrophic circulation components (left zonal circulation, middle: meridional circulation, right; large-scale vorticity). Green/brown colours occur when increases in the driver produce an increase/decrease in precipitation.

Best predictor

The most important predictor for RCM-temperature is (obviously) GCM-temperature. Without this component, time-correlations (see section on Performance below) between *truth* and *predicted* time series are rarely above 0.5. However, because local GCM temperature and the geostrophic wind

components are correlated, including GCM-temperature modifies the patterns of the regression coefficients of the circulation components. Typically their effect is reduced when local GCM temperature is included as a predictor. For precipitation this is not the case.

Performance: Time correlation

The KNMI regression method yields a daily time-series prediction from which seasonal means and variability are determined. Figure B2 and B3 show a few key results for the target domains of the selected RCM and CPM. They are compared to similar results based on pattern scaling using global temperature only (left column). Although it is not part of the common output, it is instructive to examine how good these predictions are for metrics other than the mean climatology. One such metric is the time-correlation between truth and prediction. The pattern scaling approach (using Tglob only) does not emulate day-to-day variability which makes this measure meaningless. The time-correlations are better in the winter than in the summer, and (much) better for temperature than for precipitation. In winter-time the weather patterns are generally of large spatial scale, making it more easy to pick up its systematic effect on local climate variables. Nevertheless, the approach clearly shows that there is a limit to the extent the GCM information is beneficial for predicting the RCM.

Performance: Variability and seasonal extremes

Finally, we discuss daily variability and the seasonal minimum and maximum temperature). In Tier 3 of the common approach aspects of daily variability are considered. For the RCM we computed the ratio of standard deviations (prediction over truth) as well as the mean yearly minimum and maximum values. Regression-based approaches are known to reduce variability. This is obtained here. With respect to winter temperature the regression model has a variability of up to 80-90% of the true daily variability. In summer the results are lower, especially in the eastern part of the domain. The substantial underestimation of the variability impacts the predictions of seasonal minimum and maximum values accordingly: They are over and underestimated by several degrees. For precipitation reduction in daily variability is even stronger.

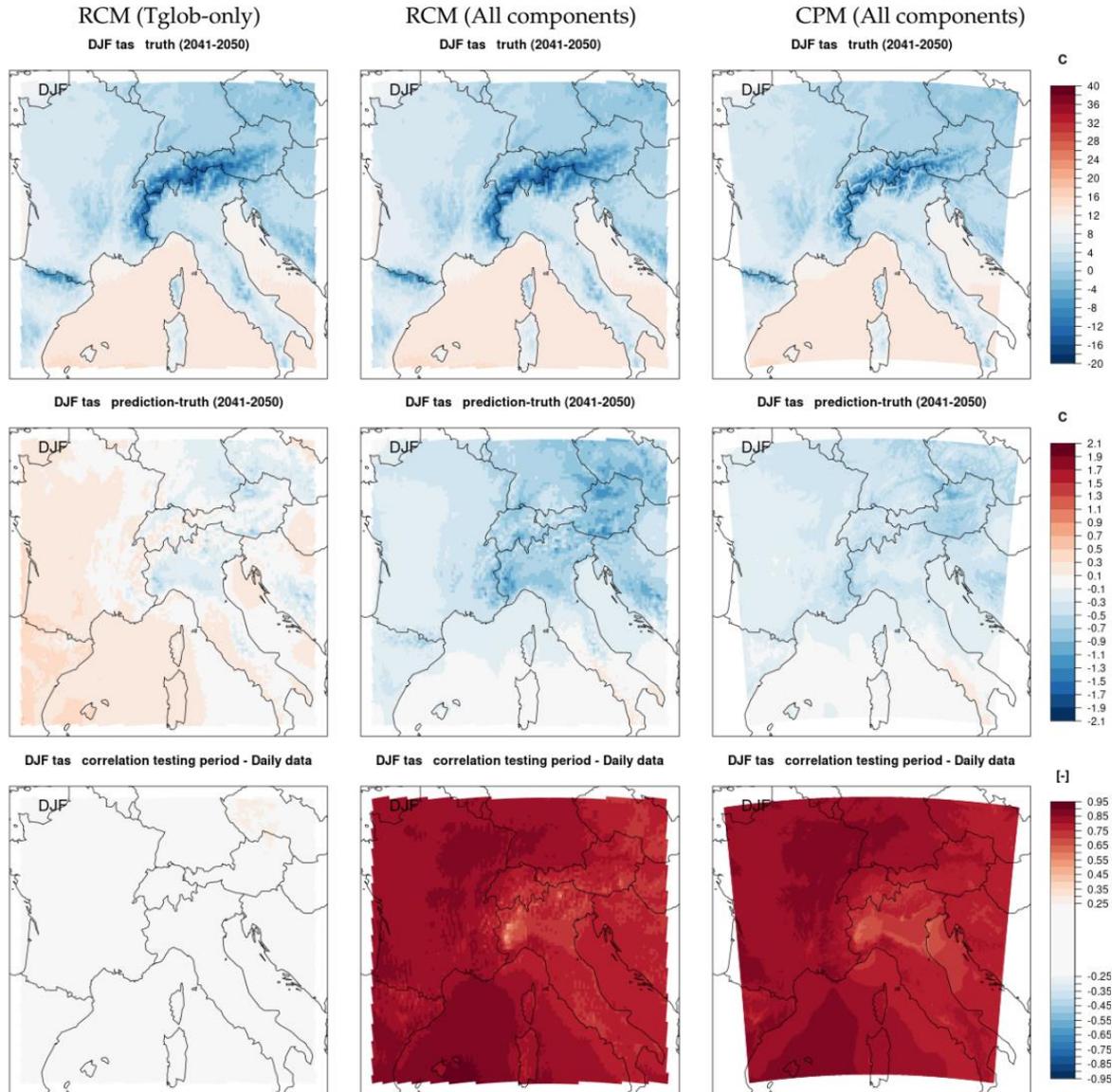


Figure B2. 2041-2050 Mean winter temperature in the RCM (left and middle) and CPM (right). Top-row shows the truth, second row the difference between prediction and truth, third row shows the daily correlation. Left column: pattern scaling with global temperature, middle and right columns: regression method.

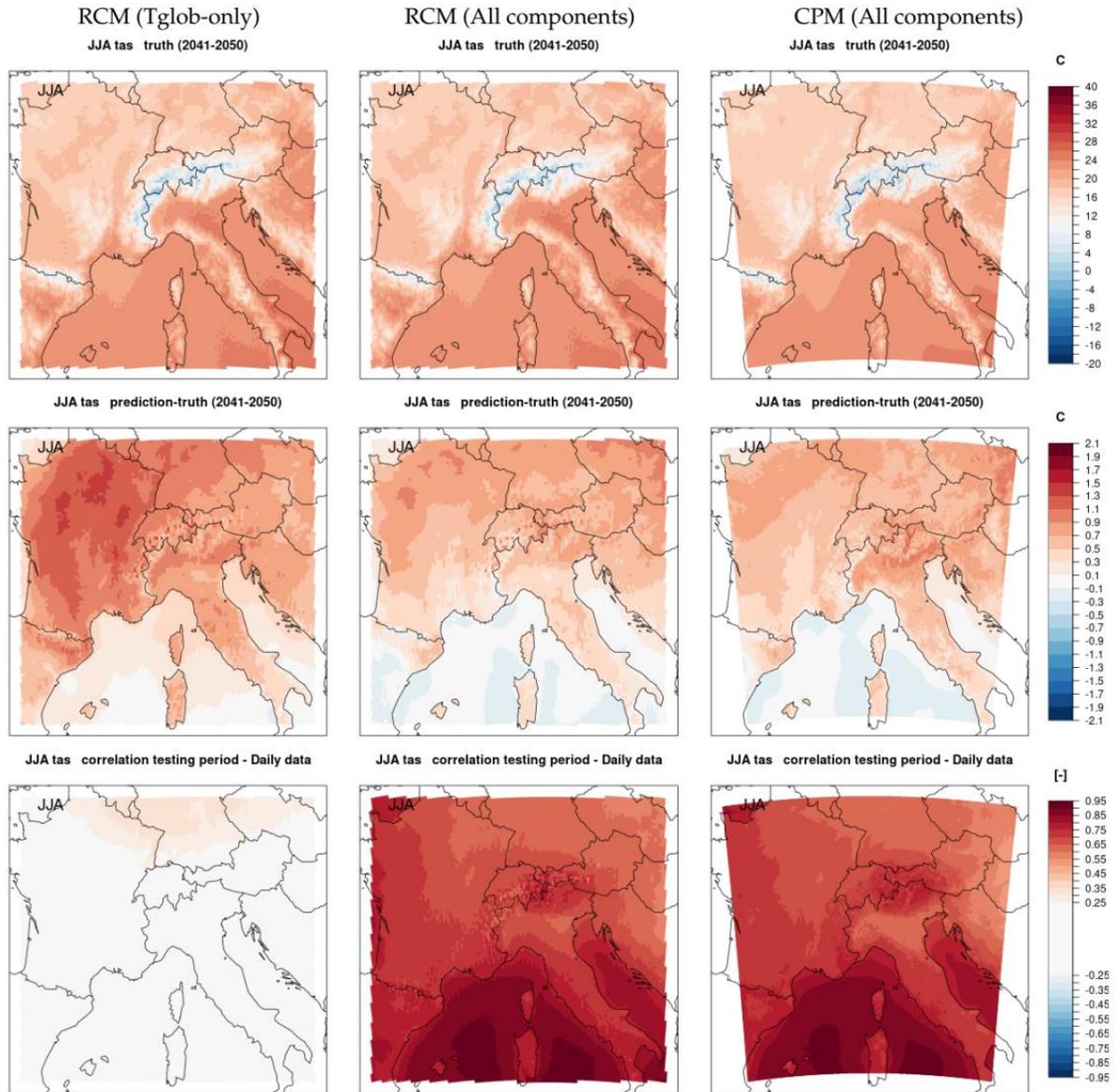


Figure B3: as previous figure but now for summer (JJA).

4 Closing remarks, lessons learned and recommendations

Summarising from our results it appears that, with the current state of the research, there is no single optimal strategy to provide user relevant data at high-resolution in an authoritative European climate prediction system setting. Sometimes simple techniques based on delta change (modifying observed data by large-scale changes such as studied in WP2) provide the best information, sometimes there is (clear) added value from the smaller-scale changes (such as the CPMs developed in WP3). For climate means simple emulators of future climate state appear to perform best, but additional information on day-to-day variability can only be provided by the advanced emulators based AI (this task). In some cases, post processing of data by statistical pooling can enhance the signal to noise, and provide more reliable estimates of the systematic changes due to climate change (this task).

To improve on this situation, we discuss a number of recommendations for future developments grouped in four categories: i) improvements on the experimental setups of high-resolution simulations in order to improve their usefulness and sampling of uncertainty, ii) improve methods to perform spatial merging, iii) statistical methods to improve the (reliability of) fine scale information and/or to emulate fine-scales learnt in a high-resolution model (instead of actually performing the simulation), and iv) recommendations on how to interact with users and guidance on the use of high-resolution climate information.

Improve experimental setups:

- **More careful selection of GCM boundaries.** Considering the computational expenses of CPM simulations, it is advisable to carefully design the matrix of simulations in order to sample internal variability as well as the uncertainty in global climate models well. Most of the present CPM ensembles are very much an ensemble of opportunity, which depends on the availability of resources, GCM forcing data and other pragmatic considerations of the institutes running the experiments. In hindsight it is checked how these simulations sit within the broader uncertainty space (see e.g., work in WP2). However, we also mention two ongoing international research activities which are targeted to improve on this situation. First, the CPM ensemble produced by Met Office downscales a global/regional perturbed physics ensemble to partly sample model uncertainty as well as internal variability (Kendon et al. 2021). Second, selection of “middle of the road¹” ensemble members within a single model initial perturbed global/regional climate model ensemble to reduce the influence of random internal climate variations (Lenderink et al. 2021; vanden Broucke et al. 2018). We note that the same arguments also apply to RCM simulations, although in general due to the much larger ensembles available this is generally considered to be a smaller issue.
- **Develop larger ensembles of CPMs.** CPMs are still relatively new and expensive models to run. However, with the development of the models and of the computer power, we may soon be able to design larger ensembles, with longer CPM simulations and by better exploring the various sources of uncertainties. Careful selection of GCMs and of GCM members will still be an important criterium, however, to design efficient and reliable CPM ensembles.

¹ that is, with average behaviour compared to the full ensemble in a number of climate statistics

- **Investigate the potential of PGW experiments.** Another way to filter out random climate variations is to consider pseudo global warming (PGW) experiments. A PGW experiment generally improves the signal-to-noise ratio by strongly reducing the internal variability of the large-scale flow conditions. This potentially leads to more reliable estimates of the forced change signal (research activity SNP1 and SME1; sections 3.3.1 and 3.4.1). However, the advantage of improved signal-to-noise should be balanced with potential biases introduced by the approach as well, such as the inability of the approach to represent the effect of changes in variability of the circulation and other drivers of change (e.g., SSTs). An additional advantage of the PGW approach is that the control period simulation is usually based on reanalysis boundary conditions, which reduces biases in the control period simulation as compared to GCM driven simulations.
- **Investigate physical storyline approaches.** Statistical approaches of a matrix design are usually designed on basic statistics; e.g., seasonal mean temperature and precipitation change. However, this may not select the spread in boundary conditions that are most relevant for the variables where we expect most added value in the high-resolution model simulations. For instance, the large-scale circulation change strongly determines the mean precipitation change, yet changes in extreme summer precipitation are also strongly influenced by atmospheric stability changes. If the matrix is designed on the basis of seasonal means (as simulated by the GCMs) we may therefore not capture all relevant large-scale conditions. Therefore, we may target experiments around known physical drivers of changes. GCM boundaries can be selected to cover uncertainty in these drivers by creating physically-consistent storyline scenarios that can be seen as plausible future options without associated probabilities. More drastic storylines such as worst-case scenarios for a phenomenon of interest for a given user is another useful option to limit the run cost with CPMs. Storyline approaches are developed in D5.5.

Improve spatial merging:

- **Cleverly combining high and low resolution information depending on usage.** The fact that most users need high-resolution climate information does not necessarily imply the use of high-resolution climate models. High-resolution future data can also be obtained statistically by merging a high-resolution observational dataset with the change information from a coarse resolution climate model. For some (or even many) applications or variables, this information may give satisfying results (see research activity SNP1; section 3.3.1). The more certain large-scale changes from large global climate model ensembles – more reliable in terms of the systematic changes or changes weighted based on observational constraints such as produced in WP2 – may outweigh the advantage of adding the fine-scale changes. In other cases, and for other variables the high-resolution model adds information that cannot be reliably estimated from the coarse resolution model. Hybrid downscaling approach (a.k.a model emulators) may in that case be a good solution. In such an approach (see 3.4), the fine-scale structure that can change over time is statistically learnt by machine learning techniques in high-resolution model runs before being added to the coarse resolution model simulations. However, training datasets longer than those available from EUCP WP3 for CPMs and covering diverse past and future climates are required for an optimal training of the model emulators.
- **Lesson from the common framework.** Although more complex emulators, such as those based on machine learning techniques, have a much wider range of applicability, they do not generally improve on the simpler methods concerning the mean climatology studied here.

The reason appears to be that those emulators are trained on weather variations and therefore do not necessarily capture the longer-term variations very well. Further improvements of the training dataset and of the evaluation framework are still needed to gain confidence in these new hybrid techniques and therefore to be able to generally apply these AI based methods to produce reliable fine-scale climate information.

Improving reliability of the information:

- **Spatial pooling.** Extracting the climate change signal from the noise is a major challenge for the relatively short CPM simulations. For some variables spatial pooling may strongly enhance signal-to-noise. This makes sense when we are dealing with small-scale atmospheric phenomena that are not necessarily location specific, such as small-scale convective rain events or wind gusts. Here, we used pooling for hourly extremes, which can be done as hourly extremes are usually rather homogeneous over large areas (research activity SNP2; section 3.3.2). However, for daily extremes this is definitely not true, and large differences occur related to large-scale circulation as well as how the circulation interacts with the orography. Improvement in the statistical methods of spatial pooling – which variables, which area, and possible ways to improve on the treatment of spatial inhomogeneities – are still needed.
- **Develop emergent/observational constraints and improvement of physical understanding.** High-resolution models often produce better present-day statistics (activities AV1-4, section 3.2 and also Lucas-Picher et al. 2021, Caillaud et al. 2021, Fumière et al. 2019), and also produce different climate change responses (see activities AV1, AV2, sections 3.2.1 and 3.2.2; Brisson et al. 2021). However, understanding of how the improved physics lead to more reliable climate change estimates is still limited. Observational constraints could help to fill this gap. Linking present-day observables to climate change by looking for correlation in model ensembles or across regions could be explored more (activity AV3; section 3.2.3). This could help to gain confidence in the added value of high-resolution modelling systems. Better understanding of the large-scale drivers of changes at small scales also helps in designing a matrix of model experiments (see point above).
- **Coordinated actions on analysis of CPM data.** There is a large variety of analysis of CPM data, with different analysis areas, different models and also different analysis methodologies. This makes differences often hard to interpret. For example, the response of hourly extremes to warming is complicated by the different definitions of the temperature rise which is used to derive the warming dependent sensitivity of rainfall extremes. It is difficult to assess whether apparently conflicting sensitivities (e.g. comparing Figure 9 and Section 3.2.3 with 3.2.4, first two supporting changes well beyond the Clausius-Clapeyron (CC) rate and the other not supporting this) are due to different choice for the representative *temperature*, specifics of the simulation (e.g. lapse-rate changes) or due to random climate variability, or perhaps other factors. A large coordinated action to analyse model in a more coherent framework was not possible in this task (too limited resources and late data availability and focus on the merging activities, yet we acknowledge the large efforts in WP3 into this direction) but should definitely be further pursued. Coordinated simple experiments in a more controlled PGW experimental environment context could also help to understand and interpret CPM differences.

Improve guidance and usability

- **Guidance on the difference between extracting the forced signal and exploring the full future climate variability.** Improve understanding of the signal-to-noise problem, and the difference between a change in systematic risk perspective and a prediction perspective. Acknowledge that many features of a future climate state (the manifestation of a 30-year climate, so what will actually happen) are hard to predict/project, and even more so their fine-scale features (Fatichi et al. 2016; DelSole and Tippett 2018). On the other hand, how climate statistics systematically change due to global warming, is potentially (much) more predictable and can in principle be improved by model development and usage of high-resolution models. Whether this is useful information is dependent on the user (risk) perspective. More effort should be on mutual understanding between user and climate data provider. The user concerning the inherent limitations of the data and the role in *unpredictable* internal variability, the climate data provider concerning the decision process.
- **Guidance on appropriate modelling streams for a given application.** Provide better guidance on appropriate modelling systems for specific users. This is obviously strongly connected to the first point here. For many users the high-resolution data by high-resolution modelling systems is very attractive, and gives a false impression of the predictability of the regional climate. This leads to a bias with respect to the use of high-resolution systems. For some applications this may bias the users' decisions as well. For instance, the signal-to-noise in the occurrence of large-scale winter storms in a 10-year simulation is very low. So, even though the CPMs will produce very realistic local wind patterns, it will do so in a large-scale environment dominated by internal variability. On the other hand, CPMs do provide essential information on how size, lifetime and intensity of convective systems could evolve in future (see EUCP D3.4), in particular in a spatially aggregated (pooled) sense; information that cannot be reliably estimated from the other modelling systems (RCMs and GCMs). For many applications, RCMs may still be the tool to use with relatively high-resolution and reasonable coverage of internal variability as well as large-scale driving conditions.
- **Guidance on the representativeness of high-resolution model simulations within the envelope of uncertainty.** As part of the guidance for users, simple information on where the high-resolution simulations sit within the uncertainty in large-scale driving conditions – and how these biases could affect the user decisions – should be more explored (see EUCP D2.3). In particular, being aware that a given CPM simulation is representative of the central part of the plausible distribution or is illustrating a high-risk-low-probability storyline has to be clarified for the CPM data users.

The challenge of providing actionable climate change information at regional and local scales is still enormous. Part of this is related to the *distillation* concept: how can we best make use of, and merge the results of the different modelling streams, from GCM to RCMs and CPMs? Can we use statistical methods to improve the distillation, and how do we account for physical/process understanding, as well as multiple lines of evidence? This part is mainly covered in this deliverable. The other part is the large diversity of user needs, entailing different variables, different time and spatial scales, and different risk perspectives. This deliverable provides a framework to further proceed on this situation. A key factor in further progress is a better integration of the different climate modelling communities (see Demory et al. 2020 for an example of such an activity). In addition, a stronger communication

between the climate modelling community and end users would be very valuable to better understand users' needs as well as what can be provided.

Finally, the situation due to Covid-19 imposed practical limitations on collaborative research. Many of the activities — e.g. work on regional signal-to-noise, (advanced) emulators, and common framework — reported here are still in an explorative/early development stage, and the lack of opportunities to meet live had an impact on collaborative efforts. In this task, we focussed collaboration on the common framework study as we felt this is an area which is completely unexplored and important as well as practically achievable with available resources. Consequently, work on added value is therefore mostly incremental to existing literature (and less coordinated) and focusses mainly on precipitation extremes, with a strong focus on the added value of the CPMs. We note however also that research on the (added value of) CPMs is relatively new/explorative, caused by relatively recent availability of modelling data (only becoming available during the course of this project) but also the relatively short simulation lengths, the difficulty to get good evaluation observational data as well as differences in (and different opinions on) evaluation strategies.

5 Publications

Published

Caillaud C., Somot S., Alias A., Bernard-Bouissières I., Fumière Q., Laurantin O., Seity Y., Ducrocq V. (2021) : Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. *Clim Dyn* **56**, 1717–1752 (2021). <https://doi.org/10.1007/s00382-020-05558-y>

Caillaud, C., Somot, S., Alias, A. *et al.* Correction to: Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. *Clim Dyn* (2021). <https://doi.org/10.1007/s00382-021-05887-6>

Christensen, O.B., Kjellström, E., Filling the matrix: an ANOVA-based method to emulate regional climate model simulations for equally-weighted properties of ensembles of opportunity. *Clim Dyn* (2021). <https://doi.org/10.1007/s00382-021-06010-5>

Ciarlo` JM, Coppola E, Fantini A, et al. (2021). A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles. *Climate Dynamics* 57, 1403–1424. <https://doi.org/10.1007/s00382-020-05400-5>

Brisson, E., Blahak, U., Lucas-Picher, P., Purr, C., & Ahrens, B. (2021). Contrasting lightning projection using the lightning potential index adapted in a convection-permitting regional climate model. *Climate Dynamics*, 1-15.

Evin G., Somot S., Hingray B. (2021) Balanced estimate and uncertainty assessment of European climate change using the large EURO-CORDEX regional climate model ensemble. *Earth Syst. Dynam.*, 12, 1543–1569, 2021, <https://doi.org/10.5194/esd-12-1543-2021>

Fumière Q., Déqué M., Nuissier O., Somot S., Alias A., Caillaud C., Laurantin O., Seity Y. (2020) Extreme rainfall in Mediterranean France during the fall: added-value of the CNRM-AROME

Convection-Permitting Regional Climate Model. *Clim. Dyn.*, 55:77-91, doi:10.1007/s00382-019-04898-8

Lind, P., Belušić, D., Christensen, O. B., et al, 2020: Benefits and added value of convection-permitting climate modelling over Fenno-Scandinavia. *Clim. Dyn.*, 55, 1893–1912, doi: 10.1007/s00382-020-05359-3.

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 Math. Phys. Eng. Sci.*, **379**, 20190544, <https://doi.org/10.1098/rsta.2019.0544>.

Lucas-Picher, P., Argüeso, D., Brisson, E., Trambly, Y., Berg, P., Lemonsu, A., ... & Caillaud, C. (2021). Convection-permitting modeling with regional climate models: Latest developments and next steps. *Wiley Interdisciplinary Reviews: Climate Change*, 12(6), e731.

Matte, D., Larsen, M. A. D., Christensen, O. B., & Christensen, J. H. (2019). Robustness and scalability of regional climate projections over Europe. *Frontiers in Environmental Science*, 6, 163.

Vergara-Temprado, J., N. Ban and C. Schär (2021). Extreme sub-hourly precipitation intensities scale close to the Clausius-Clapeyron rate over Europe. *Geophysical Research Letters*, 48, e2020GL089506. <https://doi.org/10.1029/2020GL089506>

In review

de Vries, H., Lenderink, G., Wiel, K. van der & van Meijgaard, E. (2022) Quantifying the role of the large-scale circulation on European summer precipitation change. *Climate Dynamics*. accepted.

Doury, A., Somot, S., Gadat, S., Ribes, A., Corre, L. (2022) Regional Climate Model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics*. in review

In draft

Lenderink et al. [Evaluation of hourly precipitation in convection permitting models using scaling: are they better than parameterized models?](#)

Lenderink et al. [A perfect model study on the feasibility of adding small-scale information in climate change projections](#)

6 References

- Aalbers EE, Lenderink G, van Meijgaard E, van den Hurk BJM (2018) Local-scale changes in mean and heavy precipitation in western Europe, climate change or internal variability? *Climate Dynamics* 50(11):4745–4766, DOI 10.1007/s00382-017-3901-9
- Ban N, Schmidli J, Schär C (2014) Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres* 119:7889–7907. <https://doi.org/10.1002/2014JD021478>
- Ban N., J. Schmidli, C. Schär (2015). Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophys. Res. Lett.*, 42, 1165–1172, <https://doi.org/10.1002/2014GL062588>
- Ban, N., J. Rajczak, J. Schmidli, C. Schär (2020). Analysis of precipitation extremes using generalized extreme value theory in convection-resolving climate simulations. *Clim. Dyn.*, 55, 61–75, <https://doi.org/10.1007/s00382-018-4339-4>
- Ban N, Caillaud C, Coppola E, *et al.* (2021). The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Climate Dynamics* 57, 275–302. <https://doi.org/10.1007/s00382-021-05708-w>
- Baño-Medina, J., Manzanar, R., & Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4), 2109–2124.
- Baño-Medina, J., Manzanar, R. & Gutiérrez, J.M. (2021) On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections. *Clim Dyn* 57, 2941–2951. <https://doi.org.libproxy.viko.it/10.1007/s00382-021-05847-0>
- Barring L, Strandberg G (2018) Does the projected pathway to global warming targets matter? *Environ Res Lett* 13:024029. <https://doi.org/10.1088/1748-9326/aa9f72>
- Belušić, A., M. Telišman Prtenjak, I. Güttler, N. Ban, D. Leutwyler, C. Schär (2018). Near-surface wind variability over the broader Adriatic region: insights from an ensemble of regional climate models, *Clim. Dyn.*, 50, 4455–4480, <https://doi.org/10.1007/s00382-017-3885-5>
- Belušić, D. *et al.* (2020). HCLIM38: a flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales, *Geosci. Model Dev.*, 13, 1311–1333, <https://doi.org/10.5194/gmd-13-1311-2020>.
- Berthou, S., Kendon, E.J., Chan, S.C. *et al.* (2020). Pan-European climate at convection-permitting scale: a model intercomparison study. *Clim Dyn* 55, 35–59. <https://doi.org/10.1007/s00382-018-4114-6>
- Brogli R, Kröner N, Sørland SL, *et al* (2019). The role of hadley circulation and lapse-rate changes for the future European summer climate. *Journal of Climate* 32:385–404. <https://doi.org/10.1175/JCLI-D-18-0431.1>
- Vanden Broucke S, Wouters H, Demuzere M, van Lipzig NPM (2019) The influence of convection-permitting regional climate modeling on future projections of extreme precipitation: dependency on topography and timescale. *Climate Dynamics* 52:5303–5324. <https://doi.org/10.1007/s00382-018-4454-2>
- Caillaud C., Somot S., Alias A., Bernard-Bouissières I., Fumière Q., Laurantin O., Seity Y., Ducrocq V. (2021) : Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. *Clim Dyn* 56, 1717–1752 (2021). <https://doi.org/10.1007/s00382-020-05558-y>
- Caillaud, C., Somot, S., Alias, A. *et al.* Correction to: Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. *Clim Dyn* (2021). <https://doi.org/10.1007/s00382-021-05887-6>
- Chan, S. C., E. J. Kendon, N. M. Roberts, H. J. Fowler, S. Blenkinsop (2016). The characteristics of summer sub-hourly rainfall over the southern UK in a high-resolution convective permitting model. *Environmental Research Letters*, 11, 094024, <https://doi.org/10.1088/1748-9326/11/9/094024>

- Christensen, J. H., Larsen, M. A. D., Christensen, O. B., Drews, M., and Stendel, M. (2019). Robustness of European climate projections from dynamical downscaling. *Climate Dyn.*
- Christensen, O. B., Yang, S., Boberg, F., Maule, C. F., Thejll, P., Olesen, M., et al. (2015). Scalability of regional climate change in Europe for high-end scenarios. *Clim. Res.* 64, 25–38. doi: 10.3354/cr01286
- Christensen, O. B., Kjellström, E. (2018): Projections for Temperature Precipitation, Wind, and Snow in the Baltic Sea Region until 2100. Oxford University Press. <http://dx.doi.org/10.1093/acrefore/9780190228620.013.695>
- Christensen, O. B., & Kjellström, E. (2020). Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Climate Dynamics*, 54(9), 4293-4308. <https://doi.org/10.1007/s00382-020-05229-y>
- Christensen, O. B., & Kjellström, E. (2021). Filling the Matrix: An ANOVA-Based Method to Emulate Regional Climate Model Simulations for Equally-Weighted Properties of Ensembles of Opportunity. *Clim Dyn.* <https://doi.org/10.1007/s00382-021-06010-5>
- Ciarlo` JM, Coppola E, Fantini A, et al. (2021). A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles. *Climate Dynamics* 57, 1403–1424. <https://doi.org/10.1007/s00382-020-05400-5>
- Clemins PJ, Bucini G, Winter JM, Beckage B, Towler E, Betts A, Cummings R, Queiroz HC (2019). An analog approach for weather estimation using climate projections and reanalysis data. *Journal of Applied Meteorology and Climatology* 58(8):1763–1777, DOI 10.1175/JAMC-D-18-0255.1
- Coppola E, et al. (2018). A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dynamics* DOI 10.1007/s00382-018-4521-8, URL <https://doi.org/10.1007/s00382-018-4521-8>
- Coppola, E., Nogherotto, R., Ciarlò, J. M., Giorgi, F., van Meijgaard, E., Kadyrov, N., et al. (2021). Assessment of the European Climate Projections as Simulated by the Large EURO-CORDEX Regional and Global Climate Model Ensemble. *Journal of Geophysical Research: Atmospheres*, 126, e2019JD032356. <https://doi.org/10.1029/2019JD032356>
- Coumou D, Di Capua G, Vavrus S, Wang L, Wang S (2018). The influence of Arctic amplification on mid-latitude summer circulation. *Nature Communications* DOI 10.1038/s41467-018-05256-8, URL <http://link.springer.com/10.1007/s40641-018-0105-2>
- DelSole T, Tippett MK (2018) Predictability in a changing climate. *Clim Dyn* 51:531–545. <https://doi.org/10.1007/s00382-017-3939-8>
- Demory M-E, Berthou S, Fernández J, et al (2020) European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP). *Geosci Model Dev* 13:5485–5506. <https://doi.org/10.5194/gmd-13-5485-2020>
- Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., et al. (2007). An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Clim. Change* 81(Suppl. 1), 53–70. doi: 10.1007/s10584-006-9228-x
- Déqué M, Somot S, Sanchez-Gomez E, Goodess CM, Jacob D, Lenderink G, Christensen OB (2012). The spread amongst ENSEMBLES regional scenarios: Regional climate models, driving general circulation models and interannual variability. *Climate Dynamics* 38(5-6):951–964, DOI 10.1007/s00382-011-1053-x
- Di Luca A, de Elía R, Laprise R (2015). Challenges in the Quest for Added Value of Regional Climate Dynamical Downscaling. *Current Climate Change Reports* 1:10–21. <https://doi.org/10.1007/s40641-015-0003-9>
- Doury A., Somot S., Gadat S., Ribes A., Corre L. (2022). Regional Climate Model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach. *Climate Dyn.*, available on <https://www.researchsquare.com/article/rs-725819/latest.pdf> (in review).

- Erlandsen HB, Parding KM, Benestad R, Mezghani A, Pontoppidan M (2020). A hybrid downscaling approach for future temperature and precipitation change. *Journal of Applied Meteorology and Climatology* 59(11):1–46, DOI 10.1175/jamc-d-20-0013.1
- Fantini A, *et al.* (2018). Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high-resolution observations. *Climate Dynamics*.
- Fantini A (2019). Ph. D. Thesis: climate change impact on flood hazard over Italy. <http://hdl.handle.net/11368/2940009>
- Feser F, Rockel B, von Storch H, et al (2011). Regional Climate Models Add Value to Global Model Data: A Review and Selected Examples. *Bulletin of the American Meteorological Society* 92:1181–1192. <https://doi.org/10.1175/2011BAMS3061.1>
- Fowler HJ, Lenderink G, Prein AF, Westra S, et al. (2021). Anthropogenic intensification of short- duration rainfall extremes. *Nature Reviews Earth & Environment* 2(2):107–122, DOI 10.1038/s43017-020-00128-6, URL <http://dx.doi.org/10.1038/s43017-020-00128-6>
- Fumière Q, Déqué M, Nuissier O, Somot S, Alias A, Caillaud C, Laurantin O, Seity Y (2019). Extreme rainfall in Mediterranean France during the fall: added-value of the CNRM-AROME Convection-Permitting Regional Climate Model. *Climate Dynamics*. <https://doi.org/10.1007/s00382-019-04898-8>
- Giorgi F, Jones C, Asrar GR (2006). Addressing climate information needs at the regional level: the CORDEX framework. *Bull Amer Meteor Soc* 58:175–183
- Giorgi F, Torma C, Coppola E, et al (2016). Enhanced summer convective rainfall at Alpine high elevations in response to climate warming. *Nature Geosci* 9:584–589. <https://doi.org/10.1038/ngeo2761>
- Giorgi F (2019). Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next? *Journal of Geophysical Research: Atmospheres* 124:5696–5723. <https://doi.org/10.1029/2018JD030094>
- Giorgi F, Coppola E, Jacob, *et al.* (2021). The CORDEX-CORE EXP-I initiative: Description and highlight results from the initial analysis. *Bulletin of the American Meteorological Society*.
- Gutowski, W. J., et al, C. (2020). The Ongoing Need for High-Resolution Regional Climate Models: Process Understanding and Stakeholder Information, *Bulletin of the American Meteorological Society*, 101(5), E664-E683. Retrieved Jan 11, 2022, from <https://journals.ametsoc.org/view/journals/bams/101/5/bams-d-19-0113.1.xml>
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset. *Int. J. Climatol.* 34, 623–642. doi: 10.1002/joc.3711
- Hazeleger W, Wang X, Severijns C, Stefanescu S, Bintanja R, Sterl A, Wyser K, Semmler T, Yang S, van den Hurk BJM, van Noije T, van der Linden E, van der Wiel K (2012). EC-EARTH V2.2: description and validation of a new seamless Earth system prediction model. *Climate Dynamics* 39:2611–2629, DOI 10.1007/s00382-011-1228-5
- Hentgen, L., N. Ban, N. Kröner, D. Leutwyler, C. Schär (2019). Clouds in convection-resolving climate simulations over Europe. *J. Geophys. Res. – Atmos.*, 124, 3849–3870, <https://doi.org/10.1029/2018JD030150>
- Huntingford, C., and Cox, P. M. (2000). An analogue model to derive additional climate change scenarios from existing GCM simulations. *Clim. Dyn.* 16, 575–586. doi: 10.1007/s003820000067
- Isotta FA, Frei C, Weigluni V, Perčec Tadić M, Lassègues P, *et al.* (2014). The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, 34: 1657–1675. <https://doi.org/10.1002/joc.3794>
- Jin, L., Li, Z., He, Q. *et al.* (2016). Observation and simulation of near-surface wind and its variation with topography in Urumqi, West China. *J Meteorol Res* 30, 961–982. <https://doi.org/10.1007/s13351-016-6012-3>

Junquas, C., Takahashi, K., Condom, T. *et al.* (2018). Understanding the influence of orography on the precipitation diurnal cycle and the associated atmospheric processes in the central Andes. *Clim Dyn* **50**, 3995–4017.

<https://doi.org/10.1007/s00382-017-3858-8>

Kanamitsu M, & DeHaan L (2001). The Added Value Index: A new metric to quantify the added value of regional models. *Journal of Geophysical Research*, 116 (D11106): 1-10.

Klaver R, Haarsma R, Vidale PL, Hazeleger W (2020). Effective resolution in high-resolution global atmospheric models for climate studies. *Atmospheric Science Letters* 21(4):1–8, DOI 10.1002/asl.952

Kendon, E. J., N. M. Roberts, H. J. Fowler, M. J. Roberts, S. C. Chan, C. A. Senior (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4, 570–576, <https://doi.org/10.1038/nclimate2258>

Kendon, Elizabeth, Chris Short, James Pope, Steven Chan, Jonathan Wilkinson, Simon Tucker, Phil Bett, and Glen Harris (2021). Update to UKCP Local (2.2km) projections, July 2021

https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/ukcp18_local_update_report_2021.pdf

Kjellström, E., Nikulin, G., Strandberg, G., Christensen, O. B., Jacob, D., Keuler, K., Lenderink, G., van Meijgaard, E., Schär, C., Somot, S., Sørland, S. L., Teichmann, C., and Vautard, R. (2018). European climate change at global mean temperature increases of 1.5 and 2 °C above pre-industrial conditions as simulated by the EURO-CORDEX regional climate models, *Earth Syst. Dynam.*, 9, 459–478, <https://doi.org/10.5194/esd-9-459-2018>

Klingaman N P, Martin G M and Moise A (2017). Asop (v1.0): a set of methods for analyzing scales of precipitation in general circulation models. *Geosci. Model Dev.* 10 57–83.

Lenderink, G. and E. Van Meijgaard (2008). Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, 1, 511–514, <https://doi.org/10.1038/ngeo262>

Lenderink G, Meijgaard E, Selten F (2009). Intense coastal rainfall in the Netherlands in response to high sea surface temperatures: Analysis of the event of August 2006 from the perspective of a changing climate. *Climate Dynamics* 32(1):19–33, DOI 10.1007/s00382-008-0366-x

Lenderink G, Mok HY, Lee TC, van Oldenborgh GJ (2011). Scaling and trends of hourly precipitation extremes in two different climate zones – Hong Kong and the Netherlands. *Hydrology and Earth System Sciences* 15:3033–3041. <https://doi.org/10.5194/hess-15-3033-2011>

Lenderink G, Barbero R, Westra S, Fowler HJ (2018). Reply to comments on “Temperature-extreme precipitation scaling: a two-way causality?” *International Journal of Climatology* 8–10. <https://doi.org/10.1002/joc.5799>

Lenderink, G., D. Belu, H. J. Fowler, E. Kjellström, P. Lind (2019). Systematic increases in the thermodynamic response of hourly precipitation extremes in an idealized warming experiment with a convection-permitting climate model. *Environmental Research Letters*, 14, 074012.

Lenderink G, de Vries H, Fowler HJ, et al (2021). Scaling and responses of extreme hourly precipitation in three climate experiments with a convection-permitting model. *Phil Trans R Soc A* 379:20190544. <https://doi.org/10.1098/rsta.2019.0544>

Loriaux, J. M., G. Lenderink, S. R. De Roode, A. P. Siebesma (2013). Understanding convective extreme precipitation scaling using observations and an entraining plume model. *Journal of the Atmospheric Sciences*, 70, 3641–3655, <https://doi.org/10.1175/JAS-D-12-0317.1>

Lloyd EA, Bukovsky M, Mearns LO (2021). An analysis of the disagreement about added value by regional climate models. *Synthese* 198:11645–11672. <https://doi.org/10.1007/s11229-020-02821-x>

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J. (2018). seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, *Earth Syst. Sci. Data*, 10, 235–249, <https://doi.org/10.5194/essd-10-235-2018>.

- Lustenberger, A., Knutti, R., and Fischer, E. M. (2014). The potential of pattern scaling for projecting temperature-related extreme indices. *Int. J. Climatol.* 34, 18–26. doi: 10.1002/joc.3659
- Maher N, Milinski S, Ludwig R (2021). Large ensemble climate model simulations: Introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics* 12:401–418. <https://doi.org/10.5194/esd-12-401-2021>
- Matte, D., Larsen, M. A. D., Christensen, O. B., & Christensen, J. H. (2019). Robustness and scalability of regional climate projections over Europe. *Frontiers in Environmental Science*, 6, 163.
- Meehl G, Stocker T, Collins W, Friedlingstein P, Gaye A, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper S, Watterson I, Weaver A, Zhao ZC (2007). Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- van Meijgaard E, van Ulft LH, Lenderink G, de Roode SR, Wipfler L, Boers R, Timmermans RMA (2012). Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe. KvR 054/12, ISBN/EAN 978-90-8815-046-3, *Climate changes Spatial Planning*, <https://library.wur.nl/WebQuery/wurpubs/fulltext/312258>
- Meinshausen M, Smith S, Calvin K, Daniel J, Kainuma M, Lamarque JF, Matsumoto K, Montzka S, Raper S, Riahi K, Thomson A, Velders G, van Vuuren D (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* 109:213–241, DOI 10.1007/s10584-011-0156-z
- Meredith, E. P., U. Ulbrich, H. W. Rust, (2020). Sub-hourly rainfall in a convection-permitting model. *Environmental Research Letters*, 15, 034031, <https://doi.org/10.1088/1748-9326/ab6787>
- Mitchell, T. D. (2003). Pattern scaling: an examination of the accuracy of the technique for describing future climates. *Clim. Change* 60, 217–242. doi: 10.1023/A:1026035305597
- Nikulin, G., Kjellström, E., Hansson, U., Strandberg, G., Ullerstig, A. (2011). Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations, *Tellus A: Dynamic Meteorology and Oceanography*, 63:1, 41-55, doi: [10.1111/j.1600-0870.2010.00466.x](https://doi.org/10.1111/j.1600-0870.2010.00466.x)
- Nishant N, Sherwood SC (2021). How Strongly Are Mean and Extreme Precipitation Coupled? *Geophysical Research Letters* 48:1–13. <https://doi.org/10.1029/2020GL092075>
- Norris J, Chen G, David Neelin J (2019). Thermodynamic versus dynamic controls on extreme precipitation in a warming climate from the Community Earth System Model Large Ensemble. *Journal of Climate* DOI 10.1175/JCLI-D-18-0302.1
- O’Gorman PA, Schneider T (2009). The physical basis for increases in precipitation extremes in simulations of 21st century climate change. *Proc Natl Acad Sci USA* 106:14,773–14,777, DOI 10.1073/pnas.0907610106
- van Oldenborgh GJ, Drijfhout S, van Ulden A, Haarsma R, Sterl A, Severijns C, Hazeleger W, Dijkstra H (2008). Western Europe is warming much faster than expected. *Climate of the Past* 5:1– 12, DOI 10.5194/cpd-4-897-2008, URL <http://arxiv.org/abs/0806.0715>, 0806.0715
- Ozturk, T., D. Matte, and J.H. Christensen (2021). Robustness of future atmospheric circulation changes over the EURO-CORDEX domain, *Clim. Dyn.* doi: 10.1007/s00382-021-06069-0
- Pavelsky, T. M., Sobolowski, S., Kapnick, S. B., and Barnes, J. B. (2012), Changes in orographic precipitation patterns caused by a shift from snow to rain, *Geophys. Res. Lett.*, 39, L18706, doi:10.1029/2012GL052741
- Pfahl S, O’Gorman PA, Fischer EM (2017). Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change* 7(6), DOI 10.1038/nclimate3287

- Pichelli E, Coppola E, Sobolowski S, *et al.* (2021). The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation. *Climate Dynamics* 56, 3581–3602. <https://doi.org/10.1007/s00382-021-05657-4>
- Polade SD, Pierce DW, Cayan DR, Gershunov A, Dettinger MD (2014). The key role of dry days in changing regional climate and precipitation regimes. *Scientific Reports* 4:1–8, DOI 10.1038/srep04364
- Prein AF, Gobiet A, Truhetz H, *et al.* (2016). Precipitation in the EUROCORDEX 0.11° and 0.44° simulations: high-resolution, high benefits? *Climate Dynamics* 46:383–412. <https://doi.org/10.1007/s00382-015-2589-y>
- Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., Gutmann, E. (2012). How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, *Bulletin of the American Meteorological Society*, 93(6), 811-829. Retrieved Jan 11, 2022, from <https://journals.ametsoc.org/view/journals/bams/93/6/bams-d-11-00052.1.xml>
- Rowell DP, Jones RG (2006). Causes and uncertainty of future summer drying over Europe. *Climate Dynamics* 27(2-3):281–299, DOI 10.1007/s00382-006-0125-9
- Rummukainen M. (2016). Added value in regional climate modelling. *Climatic Change*, 7: 145-159.
- Sanderson, M. G., Hemming, D. L., and Betts, R. A. (2011). Regional temperature and precipitation changes under high-end (≥ 4 °C) global warming. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* 369, 85–98. doi: 10.1098/rsta.2010.0283
- Santer, B. D., Wigley, T. M. L., Schlesinger, M. E., and Mitchell, J. F. B. (1990). Developing Climate Scenarios From Equilibrium GCM Results. Report Max Planck Institut für Meteorologie, Max-Planck-Institut für Meteorologie.
- Schär C, Frei C, Lüthi D, Davies HC (1996) Surrogate climate-change scenarios for regional climate models. *Geophys Res Lett* 23:669–672. <https://doi.org/10.1029/96GL00265>
- Schär C, Ban N, Fischer EM, *et al* (2016). Percentile indices for assessing changes in heavy precipitation events. *Climatic Change* 137:201–216. <https://doi.org/10.1007/s10584-016-1669-2>
- Schär, C., O. Fuhrer, A. Arteaga, N. Ban, C. Charpiloz, S. Di Girolamo, *et al.*, (2020). Kilometer-scale climate models: Prospects and challenges. *Bull. American Meteorol. Soc.*, 101, E567-E587, <https://doi.org/10.1175/BAMS-D-18-0167.1>
- Sørland, S. L., Schär, C., Lüthi, D., and Kjellström, E. (2018). Bias patterns and climate change signals in gcm-rcm model chains. *Environ. Res. Lett.* 13:074017. doi: 10.1088/1748-9326/aacc77
- Stevens, B., M. Satoh, *et al.*, (2019). DYAMOND: The DYNAMics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progr. Earth Planetary Sci.*, 6, Art. No. 61.
- Tabary P, Dupuy P, L’Henaff G, *et al.* (2012). A 10-year (1997–2006) reanalysis of quantitative precipitation estimation over France: methodology and first results. *IAHS Publ.* 351:255–260
- Tebaldi, C., and Arblaster, J. M. (2014). Pattern scaling: its strengths and limitations, and an update on the latest model simulations. *Clim. Change* 122, 459–471. doi: 10.1007/s10584-013-1032-9
- Torma C, *et al.* (2015). Added value of regional climate modelling over areas characterized by complex terrain – Precipitation over the Alps. *Journal of Geophysical Research: Atmospheres*, 3957-3972.
- von Trentini F, Aalbers EE, Fischer EM, Ludwig R (2020). Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe. *Earth System Dynamics* 11:1013–1031. <https://doi.org/10.5194/esd-11-1013-2020>
- Vrac, M., & Vaittinada Ayar, P. (2017). Influence of Bias Correcting Predictors on Statistical Downscaling Models, *Journal of Applied Meteorology and Climatology*, 56(1), 5-26. Retrieved Jan 25, 2022, from <https://journals.ametsoc.org/view/journals/apme/56/1/jamc-d-16-0079.1.xml>

- de Vries, H., Haarsma, R. J. & Hazeleger, W. (2013). On the future reduction of snowfall in western and central Europe. *Clim. Dyn.* **41**, 2319–2330.
- de Vries, H., Lenderink, G., Wiel, K. van der & van Meijgaard, E. (2022) Quantifying the role of circulation on European summer precipitation change (in review).
- Walton DB, Sun F, Hall A, Capps S (2015). A hybrid dynamical-statistical downscaling technique. Part I: Development and validation of the technique. *Journal of Climate* 28(12):4597–4617, DOI 10.1175/JCLI-D-14-00196.1
- Wang, J., Liu, Z., Foster, I., Chang, W., Kettimuthu, R., and Kotamarthi, V. R. (2021) Fast and accurate learned multiresolution dynamical downscaling for precipitation, *Geosci. Model Dev.*, 14, 6355–6372, <https://doi.org/10.5194/gmd-14-6355-2021>.
- Westra, S., H. J. Fowler, J. P. Evans, L. V. Alexander, P. Berg, F. Johnson, et al., (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, 52, 522–555, <https://doi.org/10.1002/2014RG000464>
- van der Wiel K, Lenderink G, de Vries H (2021). Physical storylines of future European drought events like 2018 based on ensemble climate modelling. *Weather and Climate Extremes* 33:100350. <https://doi.org/10.1016/j.wace.2021.100350>
- Winterrath T, *et al.* (2018). RADKLIM Version 2017.002: Reprocessed gauge-adjusted radar data, one-hour precipitation sums (RW) DOI: 10.5676/DWD/RADKLIM_RW_V2017.002
- Wood RR, Lehner F, Pendergrass AG, Schlunegger S (2021). Changes in precipitation variability across time scales in multiple global climate model large ensembles. *Environmental Research Letters* 16:. <https://doi.org/10.1088/1748-9326/ac10dd>
- Wuest M, Frei C, Altenho A, Hagen M, Litschi M, Schär C (2010). A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *International Journal of Climatology* 30:1764–1775
- Yiou P (2014). AnaWEGE: A weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development* 7(2):531–543, DOI 10.5194/gmd-7-531-2014