Internal variability vs multi-physics uncertainty in a regional climate model

A. Lavin-Gullon*, J. Fernandez, S. Bastin, R. M. Cardoso, L. Fita, T. M. Giannaros, K. Goergen, J. M. Gutierrez, S. Kartsios, E. Katragkou, T. Lorenz, J. Milovac, P. M. M. Soares, S. Sobolowski, K. Warrach-Sagi

June 18, 2020

*Corresponding author

A. Lavin-Gullon, J.M. Gutierrez

Grupo de Meteorología. Instituto de Fsica de Cantabria (IFCA), CSIC-Universidad de Cantabria. Santander. Spain

alvaro.lavin@unican.es

J. Fernandez, J. Milovac

Grupo de Meteorología y Computación. Dept. Matemática Aplicada y Ciencias de la Computación. Universidad de Cantabria. Santander. Spain

T. Lorenz, S. Sobolowski

NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway, S. Bastin

LATMOS/IPSL, UVSQ Université Paris-Saclay, UPMC Univ. Paris 06, CNRS, Guyancourt, France

L. Fita

Centro de Investigaciones del Mar y la Atmósfera (CIMA), CONICET-UBA, CNRS UMI-3351 IFAECI, C. A. Buenos Aires, Argentina

K. Goergen

Institute of Bio- and Geosciences (Agrosphere, IBG-3), Forschungszentrum Jülich (FZJ), Jülich, Germany

E. Katragkou, S. Kartsios

Department of Meteorology and Climatology, School of Geology, Aristotle University of Thessaloniki, Thessaloniki, Greece, R. M. Cardoso, P. M. M. Soares

Instituto Dom Luiz, Faculdade de Cincias, Universidade de Lisboa, Lisbon, Portugal

T. M. Giannaros

National Observatory of Athens, Institute for Environmental Research and Sustainable Development, Athens, Greece

K. Warrach-Sagi

Institute of Physics and Meteorology (IPM), University of Hohenheim, Stuttgart, Germany

Abstract

In a recent study, Coppola et al (2020) assessed the ability of an ensemble of convection-permitting models (CPM) to simulate deep convection using three case studies. The ensemble exhibited strong discrepancies between models, which were attributed to various factors. In order to shed some light on the issue, we quantify in this paper the uncertainty associated to different physical parameterizations from that of using different initial conditions, often referred to as the internal variability. For this purpose, we establish a framework to quantify both signals and we compare them for upper atmospheric circulation and near-surface variables. The analysis is carried out in the context of the CORDEX Flagship Pilot Study on Convective phenomena at high resolution over Europe and the Mediterranean, in which the intermediate RCM WRF simulations that serve to drive the CPM are run several times with different parameterizations. For atmospheric circulation (geopotential height), the sensitivity induced by multi-physics and the internal variability show comparable magnitudes and a similar spatial distribution pattern. For 2-meter temperature and 10-meter wind, the simulations with different parameterizations show larger differences than those launched with different initial conditions. The systematic effect over one year shows distinct patterns for the multi-physics and the internal variability. Therefore, the general lesson of this study is that internal variability should be analyzed in order to properly distinguish the impact of other sources of uncertainty, especially for short-term sensitivity simulations.

Keywords: Internal variability, Regional climate models, Uncertainty, Physical parameterizations, Ensemble

1 Introduction

The increasing resolution of Regional Climate Models (RCMs) has reached the so-called convection-permitting scale (Prein et al, 2015), by approaching resolutions of a few kilometers, typically used in Numerical Weather Prediction (NWP). A recent study by Coppola et al (2020) presented the largest multi-model ensemble of convection permitting RCMs to date, with an initial experiment exploring the ability of RCMs setup as NWP models and as regional climate modelling tools. Strong discrepancies between models were found in simulating three heavy precipitation events over the Alps. The explanation of these discrepancies was left open, and they speculated on three potential explanations: (1) the proximity of the event to the boundaries of the domain, (2) a failure in some RCMs to capture the response to the drivers of the event and (3) internal variability being responsible for the differences across models. This study is a follow up of Coppola et al (2020), where we investigate the role of internal variability in a selected event and we also further extend our analysis to a full annual cycle.

Internal, unforced climate variability is one of the main sources of uncertainty in global climate simulations (Hawkins and Sutton, 2009). Due to the non-linear and chaotic nature of the climate system, small perturbations to a given state of the system grow and develop different trajectories in the state space (Palmer, 2005). In a relatively short period of time, two slightly perturbed simulations in which initial conditions are modified can differ as much as two randomly chosen states of the climate system (Kalnay, 2003). When considering coupled systems that exhibit modes of low-frequency variability, even mean states over long periods of time can differ considerably. This internal or natural variability of the system is commonly explored using ensembles of simulations started from perturbed initial conditions (Haughton et al, 2014). The uncertainty arising from internal variability is not negligible compared to other sources of uncertainty, such as GCM modelling or GHG-scenario uncertainty (Hawkins and Sutton, 2009; Deser et al, 2012; van Pelt et al, 2015; Kumar and Ganguly, 2018).

In contrast, internal variability emerging in regional climate models (RCMs) is usually smaller than that in GCMs (Caya and Biner, 2004). This uncertainty is also commonly assessed by using a multi-initial-conditions ensemble (MICE) in order to separate RCM internal variability from the signal of forced variability (Giorgi and Bi, 2000; Christensen et al, 2001; Caya and Biner, 2004; Lucas-Picher et al, 2008b; Giorgi, 2019; Bassett et al, 2020). Several studies concluded that at least 5-6 members should be considered to obtain robust estimates of internal variability (Lucas-Picher et al, 2008b; Laux et al, 2017). Recent studies (Bassett et al, 2020) point to the need of even larger ensembles. The amplification of perturbations in the initial conditions is damped somewhat by the continuous flow of information through the boundaries of the limited area domain. Lucas-Picher et al (2008a) quantified the relation between the RCM internal variability and the lateral boundary forcing over the domain. In mid-latitudes, internal variability has a seasonal behaviour with higher (lower) values in summer (winter), when the boundary forcing (e.g. storm track intensity) is weaker (stronger) and the model is more (less) free to develop its own circulation (Caya and Biner, 2004; Lucas-Picher et al, 2008b). According to the general atmospheric circulation, prevalent winds (e.g. westerlies in mid-latitudes) force a flow of information through the boundary. As a result, this forcing imposes a typical pattern that exhibits increasing internal variability as one travels downwind across the domain. Flow perturbations develop and grow as they travel through the RCM domain, reaching a maximum near the downwind boundary where they are forced back to the flow of the GCM in the relaxation zone (Lucas-Picher et al, 2008b).

Despite its relevance, few studies have addressed other RCM uncertainties in the light of internal variability. Regarding multi-model uncertainty, Sanchez-Gomez et al (2009) explored the impact of internal variability for four different weather regimes, which showed different sensitivity depending on the lateral boundary conditions. The fraction of multi-model uncertainty in RCMs that can be explained by internal variability can be relatively large. For example, Gu et al (2018) suggest that it could be up to 70% of the total uncertainty for the precipitation in Asia. Also, Fathalli et al (2019) reported that internal variability was comparable to the inter-model precipitation spread in Tunisia during summertime, when the lateral forcing constraint is reduced. As for GCMs, the magnitude of RCM internal variability depends on the synoptic circulation, model configuration, region and season (Giorgi and Bi, 2000; Alexandru et al, 2007).

The relevance of RCM internal variability is also recognized by the Coordinated Regional climate Downscaling Experiment (CORDEX; Giorgi and Gutowski, 2015), an international ongoing initiative endorsed by the World Climate Research Program which coordinates the regional climate downscaling community. Under this framework, multiple institutions are producing and analysing the largest regional multi-model ensemble in history, covering all populated areas in the world with a standard set of continental-scale domains.

Multi-RCM ensembles sample the dynamical downscaling methodological uncertainty. As such, it is challenging to discern the contributions to uncertainty from other sources (e.g. physical process parameterizations, internal variability). This is because RCMs developed by different groups differ in so many aspects that the results from different models and members cannot be used to understand the processes responsible for the spread. There have been different attempts to decompose multi-model uncertainty into other sources of uncertainty that can be more systematically explored. Perturbed-Physics Ensembles (PPE; Yang and Arritt, 2002; Bellprat et al, 2012) consider a given RCM and explore the uncertainty associated to selected parameters, by sweeping a range of acceptable parameter values. This approach allows to link the resulting uncertainty to a specific parameter. Multi-physics ensembles (MPE; see e.g. García-Díez et al, 2015) provide a way to link modelling uncertainties to specific processes. These ensembles are generated using a single RCM by switching between different alternative physical parameterizations, which are the model components representing sub-grid-scale processes such as cloud microphysics, radiation, turbulence, etc. Physical parameterization are one of the key differences between different RCMs and, therefore, MPEs mimic multi-model ensembles with the advantage of a fixed dynamical core and the rest of non-sampled physics schemes. Of course, these fixed components also limit model diversity and, therefore, MPEs cannot replace multi-model ensembles. Quite a few analyses tested the ability of different MPEs to encompass the regional climate in different areas (Fernández et al, 2007; Evans et al, 2012; Solman and Pessacg, 2012; Jerez et al, 2013; García-Díez et al, 2015; Katragkou et al, 2015; Stegehuis et al, 2015; Devanand et al, 2018). Some of these analyses mentioned internal variability as potential source of background noise that impacts the sensitivity to the physical parameterization schemes (Tourpali and Zanis, 2013; Stegehuis et al, 2015), though internal variability was not formally investigated.

Few studies consider both physics sensitivity and internal variability. For instance, Laux et al (2017) explicitly aim to separate the effects of internal variability from those of changes in land-use, suggesting that internal variability has a significant impact on precipitation. Crétat and Pohl (2012) also studied the effect of physical parameterizations on internal variability and questioned the robustness of previous physics sensitivity studies which did not take into account internal variability.

The Flagship Pilot Study on Convective phenomena at high resolution over Europe and the Mediterranean (FPS-Convection) is an ongoing initiative endorsed by CORDEX. This initiative aims at studying convective processes with CPM over the Alpine region (Coppola et al, 2020) by producing both multi-model and multi-physics ensembles of RCM simulations. The initial results showed large discrepancies between individual ensemble members in their representation of selected heavy precipitation events. In this work, we take advantage of the ensembles produced in the FPS-Convection to follow up the study of Coppola et al (2020), in which the origin of these discrepancies was determined out of the scope. Since causation is difficult to address in a multi-model approach, we focus on the multi-physics ensemble within the FPS-Convection RCMs that serve to drive the CPM. We quantitatively compare the signal arising from the use of different model components (physical parameterizations) against that associated to the background noise referred to internal variability at different time scales. The objective is twofold: (1) to assess whether modelling discrepancies in Coppola et al (2020) fall within the range of internal variability and (2) to quantify how much uncertainty in a multi-physics ensemble can be explained by internal variability.

The paper is structured as follows: The methodology and data used in this work are detailed in Section 2. Section 3 presents and discusses the results. First, applied to a case study presented in Coppola et al. (2020) and, second, we extend the study to consider the role and relative magnitude of internal variability with respect to multi-physics uncertainty over an annual cycle. Finally, the conclusions are summarized in Section 4.

2 Data & methods

2.1 Multi-physics ensemble

In this work, we explore the uncertainty associated to physical parameterizations by using multi-physics ensembles (MPE, hereafter) generated in the context of the FPS-Convection. This initiative considers multiple RCMs, but here we will focus only on the sub-ensemble of simulations using the Weather Research and Forecasting (WRF) model (Skamarock et al, 2008). This modelling system provides the ability to switch among different physical parameterization schemes for a given sub-grid-scale process. Additionally, WRF allows for online telescopic nesting, running several nested domains simultaneously and exchanging information across domains at each time step. This approach gives rise to much smaller artifacts close to the borders of the inner domains, as compared to the standard procedure of running the model offline, nested into the output of a coarser resolution domain.

All institutions participating in FPS-Convection and using WRF have coordinated a MPE by setting different physical configurations so that at least one option differs among them (Table 1). The MPE considers different options varying the parameterization schemes for cloud micro-physics processes, surface and land processes, planetary boundary layer, and radiative processes. All other model configuration and experimental setup are fixed, including the model version (ARW-WRF v3.8.1).

All FPS-Convection WRF simulations consider a high-resolution (∼3km), convection-permitting domain centered over the Alpine region (ALP-3) nested into a coarser-resolution (∼12 km), and much larger, pan-European domain. Except for the deep convection parameterization scheme, that is switched off in ALP-3, physical configuration does not differ between both domains. All WRF ensemble members used one-way nesting, so there is no communication from the convection-permitting back to the coarser domain. Therefore, the convection-permitting inner domain did not alter in any way the results for the pan-European domain used in this work. Our analyses focus only on this pan-European domain, since we are interested in the uncertainty of the synoptic conditions over Europe, which drive the needed moisture that leads to unstable conditions over the Alpine area (see Section 3.1). The ALP-3 domain is not large enough to alter significantly the large-scale synoptic conditions, so, in order to reproduce the case studies of Coppola et al (2020) in the ALP-3 domain, the right sequence of observed events should be preserved first in the pan-European domain forcing simulations.

We use WRF data from two different FPS-Convection experiments driven by 6-hourly initial and lateral boundary conditions taken from the ERA-Interim Reanalysis (Dee et al, 2011):

Experiment A is described in Coppola et al (2020) and consisted of a preliminary test with all participating models, including WRF. Three heavy precipitation events in the Alpine region were simulated in two modes, identified as “weather-like” and “climate mode”. Weather-like simulations were started one day before the onset of the events, aiming at simulating the event as closely as possible to the reality, aided by the predictability provided by the initial conditions. As the proximity of the initial conditions constrains the internal variability, we did not consider weather-like simulations in this study. Climate-mode simulations were started one month before the event, so that initial conditions were not a source of predictability in this case and the models were mainly driven by the lateral boundary conditions, which is typical in regional climate modeling. We focus on a single event that occurred around the 23^{rd} June, 2009, and was covered by climate-mode simulations running for the period from 1^{st} June to 1^{st} July, 2009 (see Section 3.1). WRF members of the ensemble showed the largest differences in terms of predictability of this particular event. WRF simulations for this experiment used a pan-European domain at $0.11° \times 0.11°$ horizontal resolution (EUR-11), corresponding to the official EURO-CORDEX domain setup.

Experiment B consists of RCM evaluation simulations covering a 15-year period starting in 1999. All the WRF simulations started using the same initial conditions, with soil states generated by a 1-year spin-up run (1998). As in experiment A, the WRF model contributed with a MPE. However, the physical parameterizations for this experiment were slightly adjusted with respect to those used in experiment A (see Table 1) in order to consider more complex physics schemes and to avoid uncertainties from the interaction between distinct PBL and surface layer schemes. It should be noted that WRF simulations for this experiment used a slightly coarser $\sim$15 km horizontal resolution (EUR-15) than those in Experiment A, covering the same domain. This change was motivated to comply with the recommended odd nesting ratios for telescopic domains (5:1 in this case, from EUR-15 to ALP-3), which avoids interpolation between the staggered Arakawa-C grids used. In this way, fluxes across nested domains are more accurate and computationally efficient. In this study we used the first year (1999) of these simulations.

2.2 Multi-initial-conditions ensemble

A MICE was run to assess the role of internal variability in explaining the uncertainty developed by the MPE. We used WRF configurations AI and BI (see Table 1) to match the setup of experiments A and B, respectively, using a set of 6 different initial conditions. The set of perturbed initial conditions was generated using the lagged method (see e.g. Laux et al, 2017), i.e. by starting the simulations the day before (AI-r1), 2 days before (AI-r2), and so on, up to a 5-day lag (AI-r5). This is a simple way of perturbing the initial conditions while maintaining the physical consistency among variables. The extra simulated days are excluded, and we analyze only the period common to the MPE. The standard, no-lag runs AI and BI (say, AI-r0 and BI-r0) are part of both the 8-member MPE and this 6-member MICE.

We ran the 1-year MICE corresponding to experiment B (BI-r1 to BI-r5) only for the EUR-15 domain, without the inner ALP-3 nesting, so as to significantly reduce computational demands. Since no feedback from ALP-3 back to EUR-15 was allowed in the MPE, our EUR-15 MICE is fully comparable to EUR-15 MPE.

2.3 Quantification of uncertainty

In order to quantify the uncertainty (spread) in the two ensembles, we followed the approach of Lucas-Picher et al (2008b), who used an unbiased estimator of the inter-member variance:

$$\sigma_X^2(s,t) = \frac{1}{M-1}\sum_{m=1}^{M}\left(X(s,t,m) - \langle X\rangle(s,t)\right)^2 \tag{1}$$

where $X(s,t,m)$ is the value of a given variable X at position s (summarizing, in this case, typical bi-dimensional position indices i, j), at time step t and from ensemble member m. M is the total number of ensemble members. The term $\langle X\rangle(s,t)$ is the ensemble mean at a given position s and time t:

$$\langle X\rangle(s,t) = \frac{1}{M}\sum_{m=1}^{M} X(s,t,m). \tag{2}$$

To avoid confusion, we keep in this methodological summary the notation of Lucas-Picher et al (2008b) and earlier publications on internal variability, although the use of Greek letters (σ^2) to refer to a sample variance estimator is uncommon, and usually reserved for the population parameters to be estimated (Wilks, 2011). Note that even though this measure was proposed to quantify internal variability, it is just a measure of spread or uncertainty, that can be applied to any ensemble. This is typically employed to quantify internal variability on MICE. In this work, we apply it to both MPE and MICE.

The uncertainty, as represented by Eq. 1, is a spatio-temporal field. The evolution of uncertainty in time (UT) is calculated by considering the spatial average of the inter-member variance σ_X^2 as

$$UT^2 \equiv \overline{\sigma_X^2}^{s}(t) = \frac{1}{S}\sum_{s=1}^{S} \sigma_X^2(s,t) \tag{3}$$

where S is the total number of grid cells in the domain. UT^2 represents the domain average of the inter-member variance. To emphasize the quadratic nature of this uncertainty measure, we use the symbol UT^2 in Eq. 3 but, in the following, we consider always its square root UT, which has the units of the variable, and allows for an easier interpretation. In the same way, a spatial distribution of the uncertainty (US) is obtained by considering the time average of the inter-member variance σ_X^2 as

$$US^2 \equiv \overline{\sigma_X^2}^t(s) = \frac{1}{T}\sum_{t=1}^{T}\sigma_X^2(s,t) \tag{4}$$

where T is the total number of time steps in the period. This expression is an estimate of the expected value of the inter-member variance over a period of interest.

We consider transient eddy variability (TEV) as a reference for inter-member variability. Passing weather systems create a natural time variability in meteorological fields, which sets a limit to the maximum variability attainable at a given location. This variability is seasonally dependent, so Caya and Biner (2004) proposed to use a monthly estimator and compute a spatial average to make it comparable to UT:

$$TEV^2 \equiv \hat{\sigma}_X^2(\tau,m) = \frac{1}{S}\sum_{s=1}^{S}\overline{\left(X(s,t,m) - \overline{X}^\tau(s,m)\right)^2}^\tau \tag{5}$$

where the $\overline{}^\tau$ operator computes the monthly average, i.e. the mean for all time steps t corresponding to a given month τ. Again, the σ-notation is from previous literature but, in the following, we will simply refer to this monthly-averaged, transient-eddy variance as TEV. Note that TEV depends on the model and also suffers from sampling uncertainty, which will be quantified by computing it from different ensemble members.

Finally, the long-term impact (LTI) of the inter-member uncertainty on the climatology of a meteorological field is estimated by calculating the variance of the climate among ensemble members as

$$LTI^2 \equiv \sigma_{\overline{X}}^2(s) = \frac{1}{M-1}\sum_{m=1}^{M}\left(\overline{X}^t(s,m) - \left\langle\overline{X}^t\right\rangle(s)\right)^2 \tag{6}$$

where $\overline{X}^t(s,m)$ is the time average (i.e. the climatology) of each ensemble member m and $\left\langle \overline{X}^t \right\rangle(s)$ is the ensemble mean of the climatologies. Note that LTI measures the "uncertainty" of climate, while US measures the "climate" of the uncertainty. The latter is sensitive to the correspondence of meteorological events (e.g. heavy precipitation convective events) in time and space, while the former measures systematic deviations among members that lead to a different mean state (climate).

3 Results & discussion

3.1 Event reproducibility

As an example, we focus first on a heavy precipitation case study analyzed by Coppola et al (2020). The event was mostly driven by large-scale features, which consisted of a cut-off low over the Balkans inducing a persistent northeasterly flow over Austria. This unstable flow was warm and wet enough to trigger extreme precipitation by orographic lifting upon reaching the Alps. Observations reveal precipitation peaking on the 23^{rd} June, 2009, over Austria. RCM simulations consistently reproduced this heavy precipitation event under weather-like initialization (see Section 2.1), but Coppola et al (2020) reported mixed results when considering the climate-mode initialization. Some members of the multi-model/multi-physics ensemble completely missed the precipitation event or represent highly damped versions of it (see Figure 4 of Coppola et al (2020)). They speculated on a potentially weak background synoptic forcing for this event, which we investigate in this work.

Notably, the WRF MPE alone also exhibited mixed results in reproducing the event. For illustration, Figure 1 (left) shows the accumulated precipitation on 23^{rd} June for 4 WRF configurations. Only WRF configuration AF is able to reproduce the event, with extended precipitation over Austria. Other WRF configurations (AB, AE, AD) miss the event and show some precipitation over southern Italy or very scarce precipitation (configurations AC, AG, AI, not shown in Figure 1).

The synoptic situation, as represented by the 850hPa geopotential height (Figure 1, right), shows the cut-off low located as observed (ERA-Interim) over the Balkans for the AF configuration. For the rest of the MPE members, a low-pressure system is simulated in southern Italy, which alters the circulation so that the warm-moist airflow over the Alps is strongly reduced and precipitation is eventually not occurring or occurring over other areas (southern Italy).

Given that MPE members differ only in their physical parameterization schemes, one might be tempted to assume that configuration AF outperforms the rest. That would imply e.g. that the use of the YSU non-local boundary layer scheme somehow helps in developing the cut-off low at the right location, as opposite to the MYNN2 local mixing scheme. This is the only difference between configurations AF and AD. Moreover, YSU alone cannot explain the ability of AF to represent the event, because configuration AB also used this PBL scheme. The only difference between configurations AF and AB is the land surface model (LSM). AF used Noah-MP, a much extended version (Niu et al, 2011) of the Noah LSM (used in AB), considering a multi-layer snow model with more realistic snow physics, canopy shadows, snow on canopy, an aquifer layer, and many other improvements. Other configurations used Noah-MP (AD, AE or AI), though, and the low pressure system and precipitation still did not occur on the right place. Therefore, either the exact parameterization combination of configuration AF is the key or there must be a different explanation for the discrepancies.

Note that WRF was run using one-way, online telescopic nesting and, therefore, we can also rule out the proximity of the high precipitation event to the ALP-3 domain boundaries as potential cause for the different model results in Coppola et al (2020). Boundary artifacts close to the inner boundaries are greatly reduced in this setup and still some WRF members reproduced the event while others missed it.

An alternative hypothesis is that the different development of the event in the different MPE members is just the result of internal variability. To test this hypothesis, we considered a MICE based on configuration AI, which did not develop the event under the standard MPE initialization setup (start date: 00UTC, 1^{st} June, 2009). Configuration AI (AI-r0) developed a low over southern Italy (Figure 2a), as many of the other configurations (Figure 1). Many of the MICE members also developed a low over this area (see e.g. Figure 2), but member AI-r1 (start date: 00UTC, 31^{st} May, 2009) presents a low in the right place, when compared against ERA-Interim. This was achieved by perturbing the initial conditions, starting the simulation one day earlier, and preserving exactly the same model configuration. Note that this is not a matter of improved initial conditions, since there are more than 20 days simulated from the geopotential height fields shown in Figures 1 (right) and 2, well beyond the limit of deterministic predictability of an atmospheric state. This is the result of internal variability. The slight perturbations in the initial conditions grew up by the non-linear dynamical model. This process is in competition with the constraints imposed by the lateral boundary conditions, which bring the flow towards that of ERA-Interim close to border of the domain. This constraint can be seen in Figures 1 (right) and 2.

In this particular flow state, there seem to be two preferred weather regimes over the southern Mediterranean area or, at least, our model simulations were only able to generate these two weather regimes: one with a low evolving over southern Italy and the other with the low positioned over the Balkans. The observed flow took the Balkan low path even though the model has difficulties to reproduce this path. Note that these weather regimes and their probability of occurrence are likely model dependent. In any case, this is just one particular event. Once we have shown that internal variability can trigger flow deviations similar to those from different physical parameterizations, we focus on quantifying their relative uncertainty, i.e. the spread of MPE and MICE ensembles.

The evolution of inter-member variance in time for MPE and MICE (Figure 3) can reach comparable values. MPE member simulations take exactly the same initial and lateral boundary conditions from ERA-Interim, hence the uncertainty (essentially the member-to-member variability) at the start is very small (close to zero during the first day), indicating that all members produce similar circulation patterns. As the different physical parameterizations have an effect on the model, each member simulated a different synoptic situation and the uncertainty increases. Regarding the MICE, since its members were initialized before the MPE start date shown in Figure 3, the spread among members is larger than in the MPE in the beginning of June. MICE uncertainty (i.e. internal variability) remains fairly stable along the 1-month time span of the simulation. After about 10 days, the magnitude of MPE and MICE inter-member variance are comparable, with internal variability (MICE spread) generally larger than MPE spread. This suggests that the different physical parameterizations used in the MPE introduce smaller differences among members than those arising from internal variability.

A qualitative look at the UT evolution (Figure 3) shows that, even if uncertainty remains quite stable, there are periods of increased uncertainty that seem to be synchronous in both ensembles. These must be periods of either weaker lateral boundary forcing (the only external forcing) or increased internal variability due to a particular situation of the internal dynamics. Notably, the period 22-26 June, when the heavy precipitation event occurred over Austria, is a period of increased uncertainty, where internal variability surpasses MPE spread. Also, MPE spread seems to develop a linear trend along the 1-month period. If sustained, this trend would overcome internal variability in longer periods. Unfortunately, FPS-Convection experiment A only considered 1-month-long simulations. In order to explore MPE vs. MICE uncertainty over a longer period, we use the output from FPS-Convection experiment B in the next section.

Experiment B produced a MPE with slightly different model configurations (Table 1) and also on a slightly coarser domain (EUR-15). In order to discard a sensitivity to this coarser resolution, we simulated a new MICE using AI configuration but on a much coarser $0.44° \times 0.44°$ horizontal resolution (EUR-44). Its spread (dashed line on Figure 3) is very similar to that of EUR-11, which suggests that a major part of the uncertainty is due to the large-scale synoptic pattern and not to smaller scale variability.

3.2 Analysis over an annual cycle

We extended the analysis to an one-year period taking advantage of FPS-Convection experiment B (Section 2.1). In particular, we extended Figure 3 to one year using the year 1999 from the WRF MPE of experiment B and a MICE based on configuration BI. The resulting inter-member variance in time (Figure 4) shows a very similar behaviour of MPE spread and internal variability (MICE spread) along the whole year. MPE members started again from the same initial conditions. Therefore, they show very low differences on January 1st, which increases after about 10 days. After this 10-day transient evolution affected by the initial conditions, both ensembles show comparable inter-member variance, exhibiting an annual cycle with increased uncertainty in summer. Moreover, even weekly to monthly variability in these UT time series seems to match in both ensembles. Notably in the last months (Oct-Dec), and also in many other peaks along the year. This suggests that the differences introduced by the different physics formulations along the time are amplified by the model in a similar way than the perturbations of the initial conditions. No systematic effect is noticeable in the circulation. Put in another way, for this variable at least, multi-physics uncertainty can be fully explained by internal variability.

As in previous studies (Caya and Biner, 2004; Lucas-Picher et al, 2008b), we used transient-eddy variability (Equation 5) as a reference for uncertainty. This is the natural variability of a meteorological field associated to weather systems traveling along the storm track. TEV can be computed from any of the ensemble members. We used simulation BI (top line in Figure 4), which is the only member common to both MPE and MICE. To evaluate the uncertainty associated to the selection of this particular member, we computed the monthly TEV from each member, and its standard deviation for each ensemble and for each month is shown as error bars in Figure 4. TEV spread is very low and any member could have been used as the reference. As already found in previous studies in mid-latitudes, TEV is larger in winter than in summer, due to the more frequent passage of weather systems from the Atlantic. The faster atmospheric circulation in winter imposes a strong boundary forcing, which may explain the lower spread among ensemble members. TEV and the associated boundary forcing is lower during summer. As a result, the model has more freedom to develop its own circulation features, increasing the spread between the members. During summer, the spread reaches approximately half of the TEV, which would be the maximum attainable. This maximum is what one would expect from a GCM, which has no lateral boundary constraints. For such a model, MICE spread (i.e. internal variability) would increase during 1-2 weeks to reach the TEV line and remain around this limit along the year. In this sense, RCM internal variability is negligible compared to GCM internal variability during winter, but it represents an important fraction (approximately one half, in this example) during summer.

The similarity between MPE and MICE uncertainty is not restricted to domain averages. In Figure 5, we show the spread in space, by averaging inter-member variance in time for each model grid point (Equation 4). Both maps show a typical spatial distribution of internal variability in mid-latitudes, with increasing variability from the southwestern to the north-eastern part of the domain. The patterns are remarkably similar, with MPE inter-member variance (Figure 5a) only slightly larger than internal variability (Figure 5b). Both reach about 35 m over the Baltic Sea and a steeper gradient towards the outflow (eastern) boundary than in the inflow (western) one. The westerly input flow is slowly modified by the RCM as it travels along the domain, but it is suddenly modified at the outflow boundary to match again the ERA-Interim flow at the eastern border. Christensen et al (2001) suggested that, for a domain over Europe, the lower uncertainty in south-western Europe is also due to the fact that the area is mainly sea, and not only due to the distance to the boundaries. Seasonal winter (DJF) and summer (JJA) patterns of MPE and MICE inter-member variance (not shown) are very similar to those in Figure 5. They show higher (lower) intensity in JJA (DJF), reaching 45 m (25 m) over the Baltic Sea.

The systematic effects of the physical parameterizations on the circulation can be seen in the long-term impact (Figure 6a). LTI summarizes the variability of the climatology for the different ensemble members (Equation 6). Note that this variability is about one order of magnitude smaller than the uncertainty measures shown previously (cf. the scales of Figures 5 and 6). Nevertheless, LTI has an impact on the simulated climate, while the (time) mean inter-member variance explored previously is mainly due to a lack of correlation (Caya and Biner, 2004). The largest differences among the simulations using different parameterizations occur in the center of the domain, between Germany and Poland, and extend towards the Alpine region. Remarkably, systematic differences develop also on the northwestern boundary.

The LTI of internal variability (Figure 6b) shows a distinct pattern, with the largest values in the northern half of the domain. The magnitude is comparable to that of the MPE, though. Therefore, even though the spatial patterns are different, the systematic differences among MPE members are still comparable to the internal variability. This would suggest that one-year simulations are not enough to distinguish the systematic effect of a particular parameterization configuration compared to the impact of different initial conditions on the circulation. Since the MICE is just composed of multiple realizations of the same model configuration, its LTI must tend to zero as the simulation length increases and the climatology of all members tends towards the "true" model climatology. Longer simulations, such as those currently under way in the FPS-Convection, should provide a better assessment of the LTI of the MPE. For example, for 10-year simulations, the values on Figure 6b should be divided by a factor of $\sqrt{10} \approx 3.2$ (Lucas-Picher et al, 2008b). Up to this point, we have focused on the circulation (850 hPa geopotential height) and we have seen that multi-physics uncertainty is hard to distinguish from internal variability. The results for the circulation at 700 hPa or 500 hPa (not shown) are qualitatively similar.

3.3 Surface variables

Since circulation is only indirectly affected by physical parameterizations, in this section we focus on near-surface (2-meter) temperature. This is just one example of a variable affected by surface radiative and heat flux balances, which are parameterized in RCMs. In particular, the set of parameterizations tested in the FPS-Convection WRF ensemble (Table 1) directly affects cloud cover, surface energy (and mass) exchange and transport. As a result, this MPE shows a spread in surface temperature that substantially exceeds internal variability (Figure 7). Other near-surface variables, such as 10-meter wind, were also checked (not shown) and showed qualitatively similar results as near-surface temperature.

The evolution of inter-member variance for near-surface temperature, both for the MPE and MICE is different from the geopotential height shown in Figure 4. The annual cycle is clearer in the TEV than in the variance, which only shows a hint of a seasonal cycle during April through October. In summer, MPE and MICE spread evolution is uncorrelated, with some peak MPE uncertainty events (e.g. end of July) clearly standing out of internal variability. However, the strong winter variability seems coherent between MPE and MICE spread. Even if multi-physics spread is usually the greatest, internal variability seems to modulate it. This is in apparent contradiction with the results of Crétat and Pohl (2012), who claimed that physical parameterizations modulate IV. They show that two MICE under different physical parameterization configurations develop a different amount of IV on average. However, they also show (their Figure 4b) a coherent evolution in time of the IV between model configurations. In our setup, physical parameterizations cannot modulate IV time evolution since the model configuration is fixed in the MICE. Still, Figure 7 shows that, despite the different spread amounts in MICE and MPE, both evolve coherently in time. It is likely that a third variable, such as the strength of the external forcing (i.e. boundary conditions), modulates the degree to which both physics and IV uncertainties can grow.

Transient-eddy variability for surface temperature (monthly step line in Figure 7) shows again the mid-latitude maximum during winter. A key difference compared to the geopotential height is the large variability of TEV within MPE members, as compared to the MICE members. In fact, uncertainty in MPE nearly doubles internal variability during some months. Notably, a peak uncertainty event by the end of July reaches the TEV line (especially, when considering its uncertainty), indicating that surface temperature patterns for the different physics differ as much as two random temperature patterns in this month. Note, however, that TEV was computed using a single month and, therefore, this estimate does not consider interannual variability. This might explain the reversal of the TEV cycle during November and December. The strong uncertainty in the November UT estimate is likely pushing up the TEV value for this month.

The spatial distribution of the inter-member variance for surface temperature (Figure 8) reveals, as before, a similar pattern of increasing spread towards the northeast in both ensembles. In this case, despite the similar pattern, MPE shows larger spread values in accordance with Figure 7. MPE reaches a maximum value of about 3.5 K while MICE reaches about 2 K.

Finally, apart from the higher day-to-day uncertainty of the MPE for surface temperature, a systematic, long-term impact is clearly developed for this variable (Figure 9a). Unlike the circulation variable, the long-term impact of MPE for temperature is of comparable magnitude to its uncertainty. Also, it falls well above the long-term impact of internal variability (Figure 9b), suggesting that for variables directly influenced by physical parameterizations (such as surface temperature), one-year simulations suffice to discern the systematic effect of a given parameterization with respect to another. Not only the magnitude, but also the spatial pattern of LTI differs between that of internal variability and the effect of parameterizations. The latter shows three main maxima over Africa, central Europe and Russia. As expected, impact is negligible over the sea, where surface temperatures are prescribed.

4 Conclusions

In this study we quantified the uncertainty arising from WRF model MPEs, on two different time scales, developed within the FPS-Convection international initiative. Additionally, for each MPE, new MICEs were performed to assess the role of internal variability in explaining the different ability of MPE members to reproduce specific convective events. The study was carried out for a one-month period focusing on a particular case study of heavy precipitation over Austria, and extended to one-year timescale.

The analyses over the one-month period already shed light on the 2 main objectives of this work: (1) The failure of some WRF model configurations to reproduce the case study, as reported by Coppola et al (2020), is not related to physical parameterizations, but to the absence of a synoptic circulation pattern that favoured the event. Some members of the MICE were able to reasonably reproduce the observed synoptic pattern without modifying the model parameterization setup. (2) From a quantitative perspective, the spread due to the parameterization differences has a magnitude comparable to that from internal variability. Therefore, in these one-month simulations, the effect of the different physical parameterizations on the circulation cannot be distinguished from internal variability.

The extended study over a one-year period showed similar results for circulation variables (geopotential height). Multi-physics spread is comparable to internal variability both in its time evolution along the year and its spatial pattern. In this regard, we found multi-physics circulation uncertainty to behave according to previous RCM internal variability studies (Lucas-Picher et al, 2008b), with an annual cycle exhibiting increased uncertainty during summer and a spatial pattern of increased uncertainty towards the outflow boundaries of the regional domain.

The results, however, depend on the variable, with surface variables (known to be sensitive to parameterized processes) showing higher MPE spread. For example, for near-surface temperature the spread associated to parameterizations was above that due to the internal variability. This suggests that it is easier to discern both sources of uncertainties when analyzing variables more constrained by the model physics, which is typically the case in RCM parameterization sensitivity studies (Fernández et al, 2007; Evans et al, 2012; Solman and Pessacg, 2012; Jerez et al, 2013; García-Díez et al, 2015; Katragkou et al, 2015; Stegehuis et al, 2015; Devanand et al, 2018).

As a reference for uncertainty, we computed transient-eddy variability, and quantified its spread due to the multi-physics and to internal variability. This type of uncertainty also depends on the variable. For the circulation, transient-eddy variability of the different physical model configurations is similar to the internal variability range. However, for near-surface temperature, the different physics configurations exhibit a different level of transient-eddy variability. This requires further analysis on longer simulations to properly estimate the inter-annual contribution, but this is beyond the scope of the present work.

The long-term impact of the internal variability has been found to be of comparable magnitude to that of multi-physics for atmospheric circulation variables on year-long simulations. For surface temperature, however, the long-term impact of the multi-physics is larger, standing out of internal variability. For both variables, the spatial patterns of MPE and MICE differ, and this calls for a detailed study of each physical parameterization considered.

The techniques for quantification of internal variability (Lucas-Picher et al, 2008b) were applied here to explore also multi-physics spread, which proved to be a useful method for comparing both sources of uncertainty. They revealed that uncertainty arising from perturbations of the model physics (full replacement of a physics scheme) are seen from the circulation point of view as perturbations of initial conditions, i.e. as internal variability "noise". Both types of perturbations seem amplified in a similar way by the dynamical system and synchronously constrained by the lateral boundary conditions. This view of a structured near-surface perturbation as a random upper air circulation noise was also found, in a completely different context, by Fernández et al (2009).

The inability of an RCM to reproduce the observed day-to-day circulation due to internal variability is not a matter of concern for mean climate studies, given that long-term climate is preserved (Caya and Biner, 2004). However, with the arrival of convection-permitting simulations and the increasing interest in the climate of extremes, RCM internal variability re-emerges as a matter of concern for model evaluation. As an example, the FPS-Convection focuses on high-impact (low probability) convective phenomena that occur mainly during the summer season, when lateral boundary forcing is the weakest. The evaluation of models under these conditions poses a real challenge that can only be addressed by computationally expensive experiments including the simulation of long periods and/or the simulation of a corresponding MICE to disentangle the role of internal variability in the results. Other alternatives would be to constrain internal variability by using techniques such as spectral nudging, which has its own drawbacks (Alexandru et al, 2009), or frequently reinitializing the RCM (Lo et al, 2008; Lucas-Picher et al, 2013).

Finally, the magnitude of internal variability in an RCM has been shown to depend on the domain size and location (Giorgi and Bi, 2000; Rinke and Dethloff, 2000; Alexandru et al, 2007). Given that, for circulation variables, MPE variability behaves as internal variability, we could argue that a similar dependence on domain size and location might affect MPE variability. The generalization of these results for other domain sizes and for regions with a weaker lateral boundary forcing is left for a forthcoming study.

Acknowledgements

This work is partially funded by the Spanish government through grant BES-2016-078158 and MINECO/ FEDER co-funded projects INSIGNIA (CGL2016-79210-R) and MULTI-SDM (CGL2015-66583-R). Universidad de Cantabria simulations have been carried out on the Altamira Supercomputer at the Instituto de Física de Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network. EK and SK acknowledge the support of the Greek Research and Technology Network (GRNET) High Performance Computing (HPC) infrastructure for providing the computational resources of AUTH-simulations (under project ID pr003005) and the AUTH Scientific Computing Center for technical support. IPSL acknowledges the support from the EUCP project, funded by the European Union under H2020 Grant Agreement 776613, and from the IPSL mesocenter ESPRI facility which is supported by CNRS, UPMC, Labex L-IPSL, CNES and Ecole Polytechnique. IPSL simulations were granted access to the HPC resources of TGCC under the allocation A0050106877 made by GENCI. The computational resources for NORCE/BCCR were provided by UNINETT Sigma2 (NN9280K, NS9001K), with funding from the Research Council of Norway's support for the strategic project on climate services. FZJ gratefully acknowledges the computing time granted by the John von Neumann Institute for Computing (NIC) and JARA-HPC provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC). We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (https://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://eca.knmi.nl).

References

Alexandru A, de Elia R, Laprise R (2007) Internal variability in regional climate downscaling at the seasonal scale. Monthly Weather Review 135(9):3221–3238, DOI 10.1175/MWR3456.1,URL `https://doi.org/10.1175/MWR3456.1`

Alexandru A, de Elia R, Laprise R, Separovic L, Biner S (2009) Sensitivity study of regional climate model simulations to large-scale nudging parameters. Monthly Weather Review 137(5):1666–1686, DOI 10.1175/2008MWR2620.1

Bassett R, Young P, Blair G, Samreen F, Simm W (2020) A Large Ensemble Approach to Quantifying Internal Model Variability Within the WRF Numerical Model. Journal of Geophysical Research Atmospheres 125(7):e2019JD031,286, DOI 10.1029/2019JD031286

Bellprat O, Kotlarski S, Lüthi D, Schär C (2012) Exploring perturbed physics ensembles in a regional climate model. Journal of Climate 25(13):4582–4599, DOI 10.1175/JCLI-D-11-00275.1

Caya D, Biner S (2004) Internal variability of RCM simulations over an annual cycle. Climate Dynamics 22(1):33–46, DOI 10.1007/s00382-003-0360-2

Christensen OB, Gaertner MA, Prego JA, Polcher J (2001) Internal variability of regional climate models. Climate Dynamics 17(11):875–887, DOI 10.1007/s003820100154

Coppola E, Sobolowski S, Pichelli E, Raffaele F, Ahrens B, Anders I, Ban N, Bastin S, Belda M, Belusic D, Caldas-Alvarez A, Cardoso RM, Davolio S, Dobler A, Fernandez J, Fita L, Fumiere Q, Giorgi F, Goergen K, Güttler I, Halenka T, Heinzeller D, Hodnebrog Ø, Jacob D, Kartsios S, Katragkou E, Kendon E, Khodayar S, Kunstmann H, Knist S, Lavín-Gullón A, Lind P, Lorenz T, Maraun D, Marelle L, van Meijgaard E, Milovac J, Myhre G, Panitz HJ, Piazza M, Raffa M, Raub T, Rockel B, Schär C, Sieck K, Soares PMM, Somot S, Srnec L, Stocchi P, Tölle

MH, Truhetz H, Vautard R, de Vries H, Warrach-Sagi K (2020) 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

Crétat J, Pohl B (2012) How physical parameterizations can modulate internal variability in a Regional Climate Model. Journal of the Atmospheric Sciences 69(2):714–724, DOI 10.1175/JAS-D-11-0109.1

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hlm EV, Isaksen L, Kllberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137(656):553–597, DOI 10.1002/qj.828

Deser C, Phillips A, Bourdette V, Teng H (2012) Uncertainty in climate change projections: the role of internal variability. Climate Dynamics 38(3):527–546, DOI 10.1007/s00382-010-0977-x

Devanand A, Ghosh S, Paul S, Karmakar S, Niyogi D (2018) Multi-ensemble regional simulation of Indian monsoon during contrasting rainfall years: role of convective schemes and nested domain. Climate Dynamics 50(11):4127–4147, DOI 10.1007/s00382-017-3864-x

Evans JP, Ekström M, Ji F (2012) Evaluating the performance of a

WRF physics ensemble over South-East Australia. Climate Dynamics 39(6):1241–1258, DOI 10.1007/s00382-011-1244-5

Fathalli B, Pohl B, Castel T, Safi MJ (2019) Errors and uncertainties in regional climate simulations of rainfall variability over Tunisia: a multi-model and multi-member approach. Climate Dynamics 52(1):335–361, DOI 10.1007/s00382-018-4150-2

Fernández J, Montávez J, Sáenz J, González-Rouco J, Zorita E (2007) Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle. Journal of Geophysical Research: Atmospheres 112(D4)

Fernández J, Primo C, Cofiño AS, Gutiérrez JM, Rodríguez MA (2009) MVL spatiotemporal analysis for model intercomparison in EPS: application to the DEMETER multi-model ensemble. Climate Dynamics 33(2):233–243, DOI 10.1007/s00382-008-0456-9

García-Díez M, Fernández J, Vautard R (2015) An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. Climate Dynamics 45(11):3141–3156,DOI 10.1007/s00382-015-2529-x

Giorgi F (2019) Thirty years of regional climate modeling: Where are we and where are we going next? Journal of Geophysical Research: Atmospheres DOI 10.1029/2018JD030094

Giorgi F, Bi X (2000) A study of internal variability of a regional climate model. Journal of Geophysical Research: Atmospheres 105(D24):29,503–29,521, DOI 10.1029/2000JD900269

Giorgi F, Gutowski WJ (2015) Regional dynamical downscaling and the CORDEX initiative. Annual Review of Environment and Resources 40(1):467–490, DOI 10.1146/annurev-environ-102014-021217

Gu H, Yu Z, Yang C, Ju Q, Yang T, Zhang D (2018) High-resolution ensemble projections and uncertainty assessment of regional climate change over China in CORDEX East Asia. Hydrology and Earth System Sciences 22(5):3087–3103, DOI 10.5194/hess-22-3087-2018

Haughton N, Abramowitz G, Pitman A, Phipps SJ (2014) On the generation of climate model ensembles. Climate Dynamics 43(7):2297–2308, DOI 10.1007/s00382-014-2054-3

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society 90(8):1095–1108, DOI 10.1175/2009BAMS2607.1

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. Journal of Geophysical Research: Atmospheres 113(D20), DOI 10.1029/2008JD010201, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010201, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD010201

Jerez S, Montavez JP, Gomez-Navarro JJ, Lorente-Plazas R, Garcia-Valero JA, Jimenez-Guerrero P (2013) A multi-physics ensemble of regional climate change projections over the Iberian Peninsula. Climate Dynamics 41(7):1749–1768, DOI 10.1007/s00382-012-1551-5

Kalnay E (2003) Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press

Katragkou E, García-Díez M, Vautard R, Sobolowski S, Zanis P, Alexandri G, Cardoso RM, Colette A, Fernandez J, Gobiet A, Goergen K, Karacostas T, Knist S, Mayer S, Soares PMM, Pytharoulis I, Tegoulias I, Tsikerdekis A, Jacob D (2015) Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. Geoscientific Model Development 8(3):603–618, DOI 10.5194/gmd-8-603-2015

Kumar D, Ganguly AR (2018) Intercomparison of model response and internal variability across climate model ensembles. Climate Dynamics 51(1):207–219, DOI 10.1007/s00382-017-3914-4

Laux P, Nguyen PNB, Cullmann J, Van TP, Kunstmann H (2017) How many RCM ensemble members provide confidence in the impact of land-use land cover change? International Journal of Climatology 37(4):2080–2100, DOI 10.1002/joc.4836

Lo JCF, Yang ZL, Pielke Sr RA (2008) Assessment of three dynamical climate downscaling methods using the weather research and forecasting (WRF) model. Journal of Geophysical Research: Atmospheres 113(D9), DOI 10.1029/2007JD009216, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009216

Lucas-Picher P, Caya D, Biner S, Laprise R (2008a) Quantification of the lateral boundary forcing of a Regional Climate Model using an aging tracer. Monthly Weather Review 136(12):4980–4996, DOI 10.1175/2008MWR2448.1, URL https://doi.org/10.1175/2008MWR2448.1

Lucas-Picher P, Caya D, de Elía R, Laprise R (2008b) Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain. Climate Dynamics 31(7):927–940, DOI 10.1007/s00382-008-0384-8, URL https://doi.org/10.1007/s00382-008-0384-8

Lucas-Picher P, Boberg F, Christensen JH, Berg P (2013) Dynamical downscaling with reinitializations: A method to generate finescale climate datasets suitable for impact studies. Journal of Hydrometeorology 14(4):1159–1174, DOI 10.1175/JHM-D-12-063.1, URL https://doi.org/10.1175/JHM-D-12-063.1

Niu GY, Yang ZL, Mitchell KE, Chen F, Ek MB, Barlage M, Kumar A, Manning K, Niyogi D, Rosero E, et al (2011) The community Noah land surface model with multiparameterization options (Noah-MP): 1. model description and evaluation with local-scale measurements. Journal of Geophysical Research: Atmospheres 116(D12), DOI 10.1029/2010JD015139

Palmer T (2005) Global warming in a nonlinear climate. Can we be sure? Europhysics news 36(2):42–46

van Pelt SC, Beersma JJ, Buishand TA, van den Hurk BJJM, Schellekens J (2015) Uncertainty in the future change of extreme precipitation over the Rhine basin: the role of internal climate variability. Climate Dynamics 44(7):1789–1800, DOI 10.1007/s00382-014-2312-4

Prein AF, Langhans W, Fosser G, Ferrone A, Ban N, Goergen K, Keller M, Tlle M, Gutjahr O, Feser F, Brisson E, Kollet S, Schmidli J, van Lipzig NPM, Leung R (2015) A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Reviews of

Geophysics 53(2):323–361, DOI 10.1002/2014RG000475, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000475

Rinke A, Dethloff K (2000) On the sensitivity of a regional Arctic climate model to initial and boundary conditions. Climate Research 14:101–113, DOI 10.3354/cr014101

Sanchez-Gomez E, Somot S, Déqué M (2009) Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961–2000. Climate Dynamics 33(5):723–736, DOI 10.1007/s00382-008-0502-7

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker M, Duda KG, Y Huang X, Wang W, Powers JG (2008) A description of the advanced research WRF Version 3. NCAR Technical Note pp 1–113, DOI 10.5065/D68S4MVH

Solman SA, Pessacg NL (2012) Regional climate simulations over South America: sensitivity to model physics and to the treatment of lateral boundary conditions using the MM5 model. Climate Dynamics 38(1):281–300, DOI 10.1007/s00382-011-1049-6

Stegehuis AI, Vautard R, Ciais P, Teuling AJ, Miralles DG, Wild M (2015) An observation-constrained multi-physics WRF ensemble for simulating european mega heat waves. Geoscientific Model Development 8(7):2285–2298, DOI 10.5194/gmd-8-2285-2015

Thompson G, Field PR, Rasmussen RM, Hall WD (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. Monthly Weather Review 136(12):5095–5115, DOI 10.1175/2008MWR2387.1

Tourpali K, Zanis P (2013) Winter anticyclonic blocking effects over Europe during 1960-2000 from an ensemble of regional climate models. Climate Research 57:81–91, DOI 10.3354/cr01169

Wilks DS (2011) Statistical methods in the atmospheric sciences, vol 100. Academic press

Yang Z, Arritt RW (2002) Tests of a Perturbed Physics Ensemble Approach for Regional Climate Modeling. Journal of Climate 15(20):2881–2896, DOI 10.1175/1520-0442(2002)015⟨2881:TOAPPE⟩2.0.CO;2,URL https://doi.org/10.1175/1520-0442(2002)015<2881:TOAPPE>2.0.CO;2

List of Tables

Exp.	Id.	Institution	MP	PBL	LSM	ShC
A	AB	Forschungszentrum Jülich (FZJ-IBG3), Germany	Thomp.	YSU	NOAH	GRIMS
	AC	National Observatory of Athens (NOA), Greece	Thomp.	MYNN2	NOAH	GRIMS
	AD	University of Hohenheim (UHOH), Germany	Thomp.	MYNN2*	NOAH-MP	GRIMS
	AE	Intitute Pierre Simon Laplace (IPSL), France	Thomp.	MYNN2	NOAH-MP	UW
	AF	Bjerknes Centre for Climate Res. (BCCR), Norway	Thomp.	YSU	NOAH-MP	GRIMS
	AG	Aristotle University of Thessaloniki (AUTH), Greece	WDM6	YSU	NOAH	GRIMS
	AH	Instituto Dom Luiz (IDL), Portugal	WDM6	MYNN2	NOAH	GRIMS
	AI	Universidad de Cantabria (UCAN), Spain	WDM6	MYNN2*	NOAH-MP	GRIMS
B	BB	Forschungszentrum Jülich (FZJ-IBG3), Germany	Th-AA	YSU	NOAH	GRIMS
	BC	National Observatory of Athens (NOA), Greece	Thomp.	MYNN2	NOAH	GRIMS
	BD	University of Hohenheim (UHOH), Germany	Th-AA	MYNN2	NOAH-MP	GRIMS
	BE	Intitute Pierre Simon Laplace (IPSL), France	Th-AA	MYNN2	NOAH-MP	UW
	BF	Bjerknes Centre for Climate Res. (BCCR), Norway	Thomp.	YSU	NOAH-MP	GRIMS
	BG	Aristotle University of Thessaloniki (AUTH), Greece	WDM6	YSU	NOAH-MP	GRIMS
	BH	Instituto Dom Luiz (IDL), Portugal	WDM6	MYNN2	NOAH	GRIMS
	BI	Universidad de Cantabria (UCAN), Spain	WDM6	MYNN2	NOAH-MP	GRIMS

Table 1: WRF multi-physics configurations considered in this study (see Section 2.1) for experiment A (one-month simulation, EUR-11 domain) and experiment B (one-year simulation, EUR-15). For each ensemble member, the table shows an Id. code, the institution performing the simulation and the physical parameterizations used. The ensembles explore the use of different schemes for micro-physics (MP), planetary boundary layer and surface layer (PBL), land surface (LSM), and shallow convection (ShC) processes. The PBL schemes denoted with asterisk (*) used a different surface layer scheme despite sharing the MYNN2 PBL. See Table 2 for details of each parameterization scheme.

Acronym	Physical scheme
Thomp.	Thompson et al (2008) scheme with ice, snow and graupel processes suitable for high-resolution simulations
Th-AA	New Thompson aerosol-aware scheme considering water- and ice-friendly aerosols
WDM6	WRF Double-Moment 6-class microphysics scheme with cloud condensation nuclei for warm processes
YSU	Yonsei University non-local closure PBL scheme with revised MM5 Monin-Obukhov surface layer
MYNN2	Mellor-Yamada Nakanishi and Niino Level 2.5 (*combined with revised MM5 Monin-Obukhov surface layer)
NOAH	Noah LSM with multilayer soil temperature and moisture, snow cover and frozen soil physics
NOAH-MP	Noah LSM-Multi Physics. NOAH with multiple options for land-atmosphere processes
GRIMS	Shallow cumulus scheme from the Global/Regional Integrated Modeling System
UW	University of Washington shallow cumulus scheme from the Community Earth System Model

Table 2: Physical schemes used in the multi-physics experiments shown in Table 1.

List of Figures

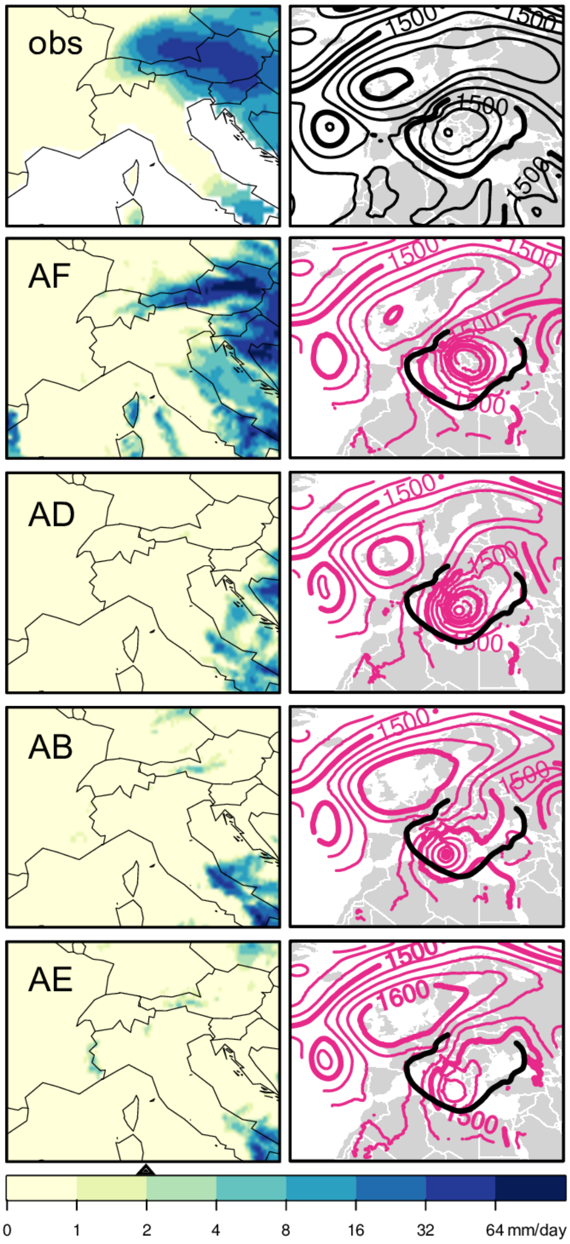

Figure 1: Left: Accumulated precipitation (mm) on June, 23rd 2009 according to E-OBS (Haylock et al (2008); top) and as simulated in the ALP-3 domain by experiment A for WRF MPE members AF, AD, AB and AE. Right: 850hPa geopotential height (m) according to ERA-Interim (top) and the corresponding MPE ensemble members in the EUR-11 domain in pink. An ERA-Interim 1500m-isoline (the same in all panels) is represented for reference in black.

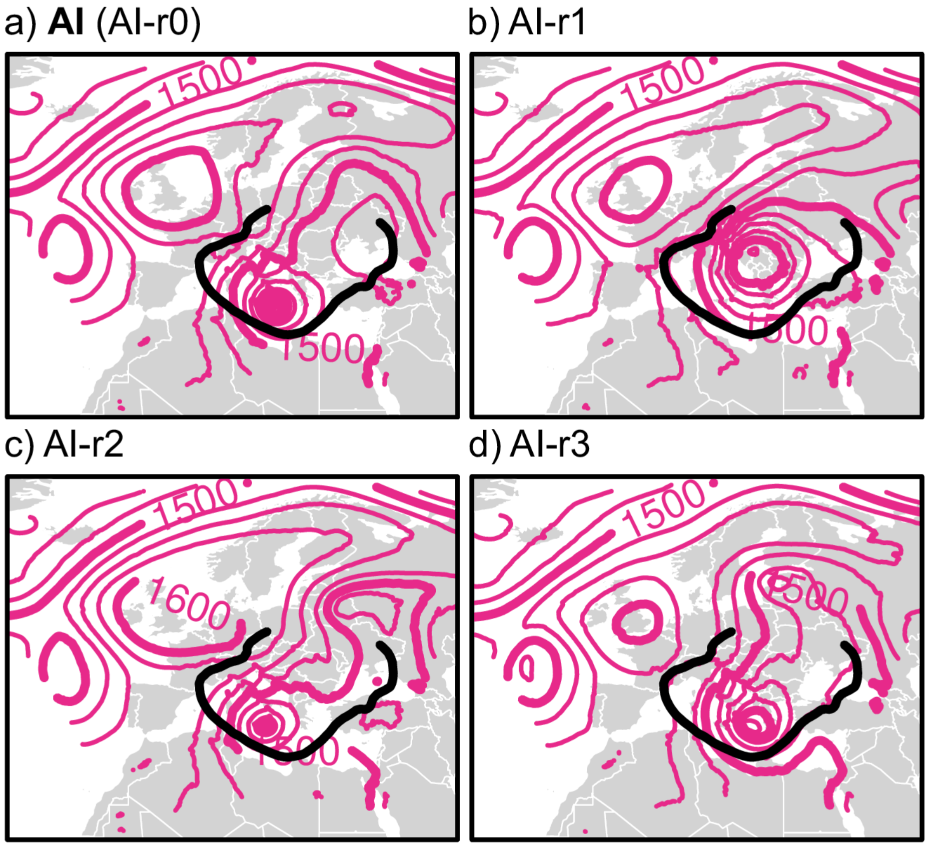

Figure 2: As Figure 1 (right), but for 4 MICE members: AI-r0 to AI-r3.

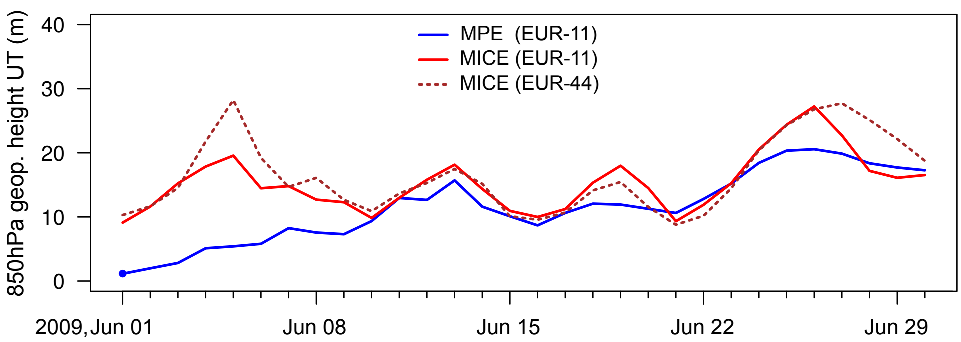

Figure 3: Inter-member variance in time (Equation 3) for 850hPa geopotential height (m) in EUR-11 domain of experiment A (June 2009). The spread is computed separately for MPE (blue) and MICE (red). The latter was computed both at 0.11° and 0.44° horizontal resolution with similar number of ensemble members.

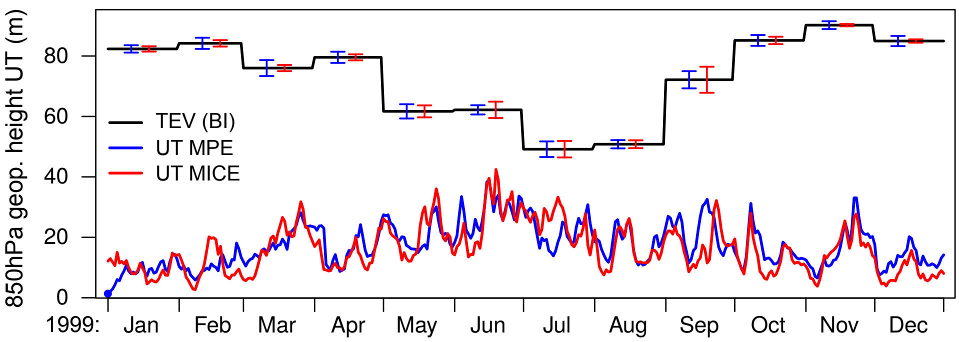

Figure 4: Inter-member variance in time (UT) for 850hPa geopotential height (m) in EUR-15 domain of experiment B (year 1999). The uncertainty is computed separately for MPE (blue) and MICE (red). Transient-eddy variability (Equation 5, black line) was computed from BI configuration and error bars show its standard deviation for MPE and MICE.

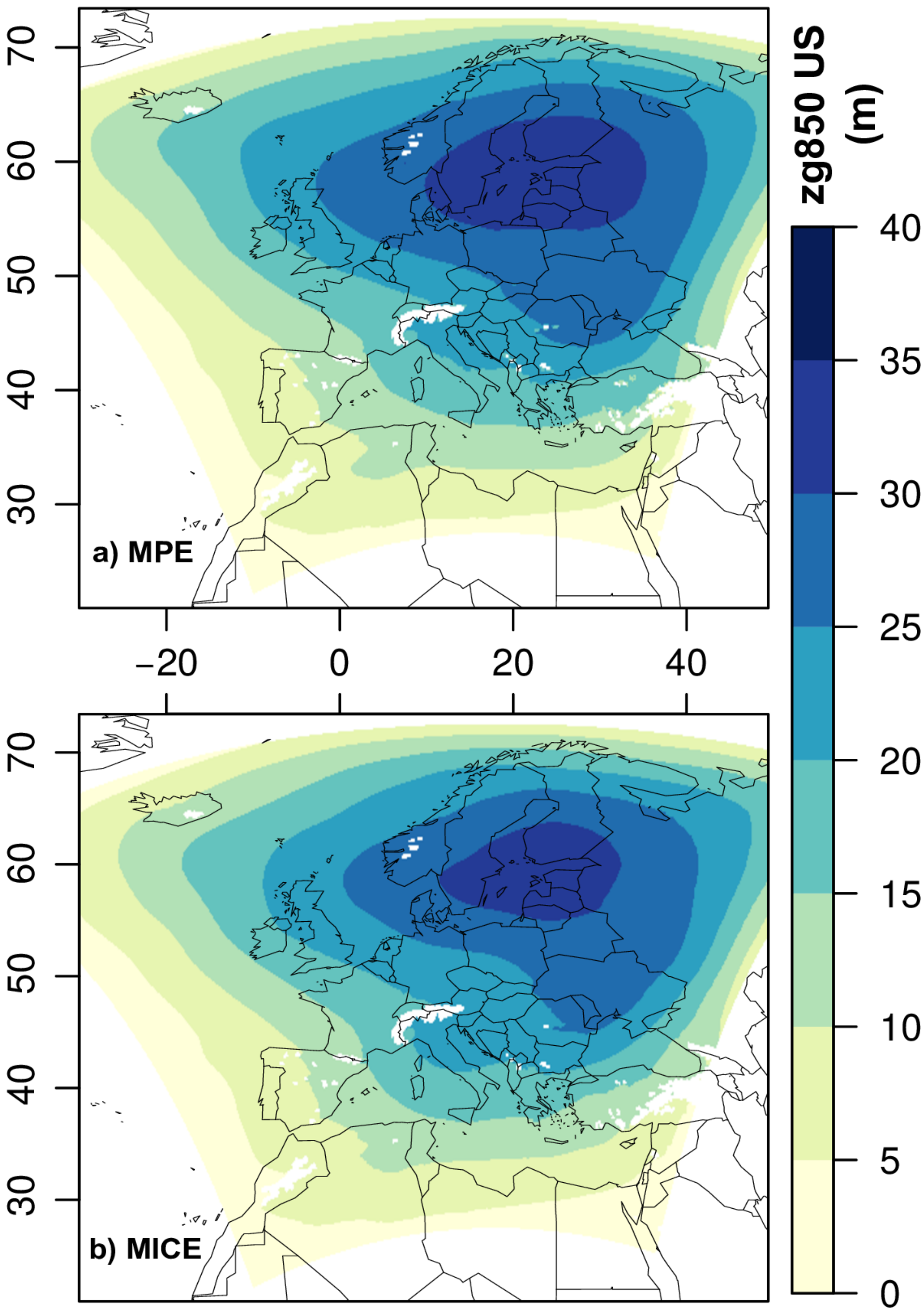

Figure 5: Spatial distribution of the inter-member variance (US) for the 850 hPa geopotential height (m) in EUR-15 domain of experiment B (year 1999). a) multi-physics ensemble. b) multi-initial-conditions ensemble.

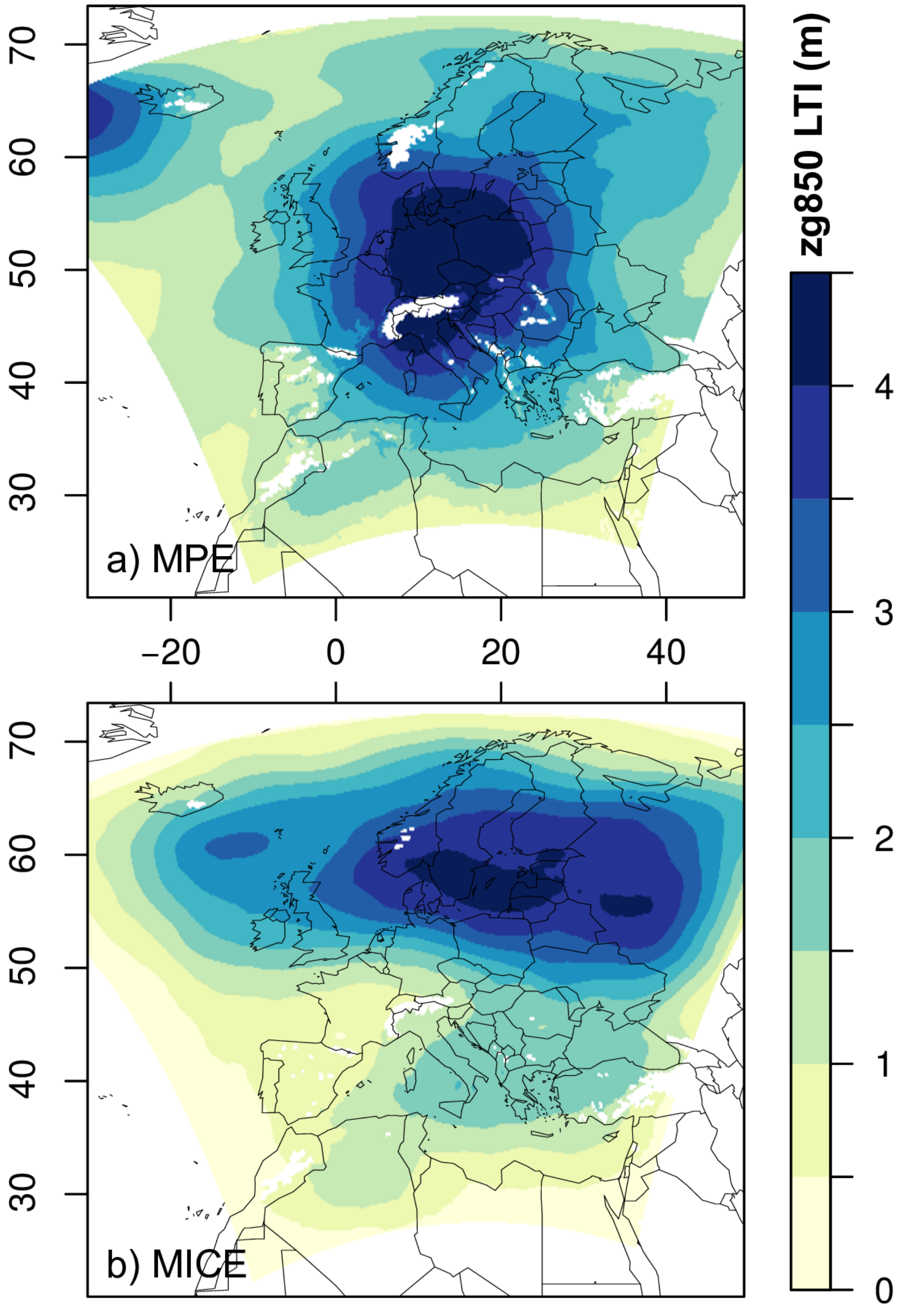

Figure 6: Long-term impact of multi-physics (a) and multi-initial-conditions (b) on 850hPa geopotential height (m).

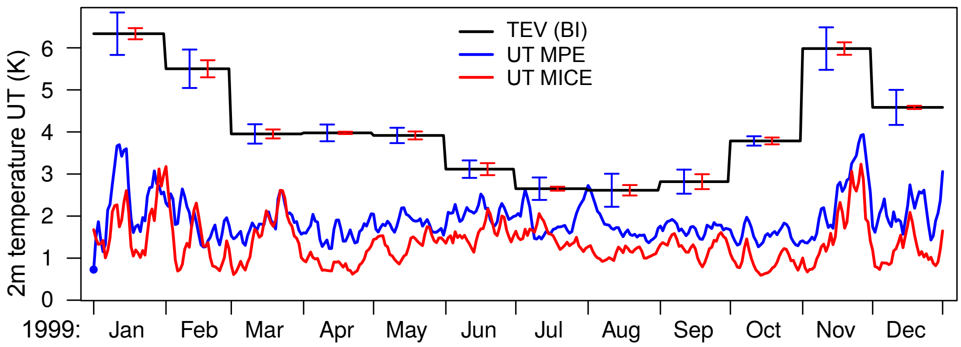

Figure 7: As Fig. 4 but for surface temperature over land.

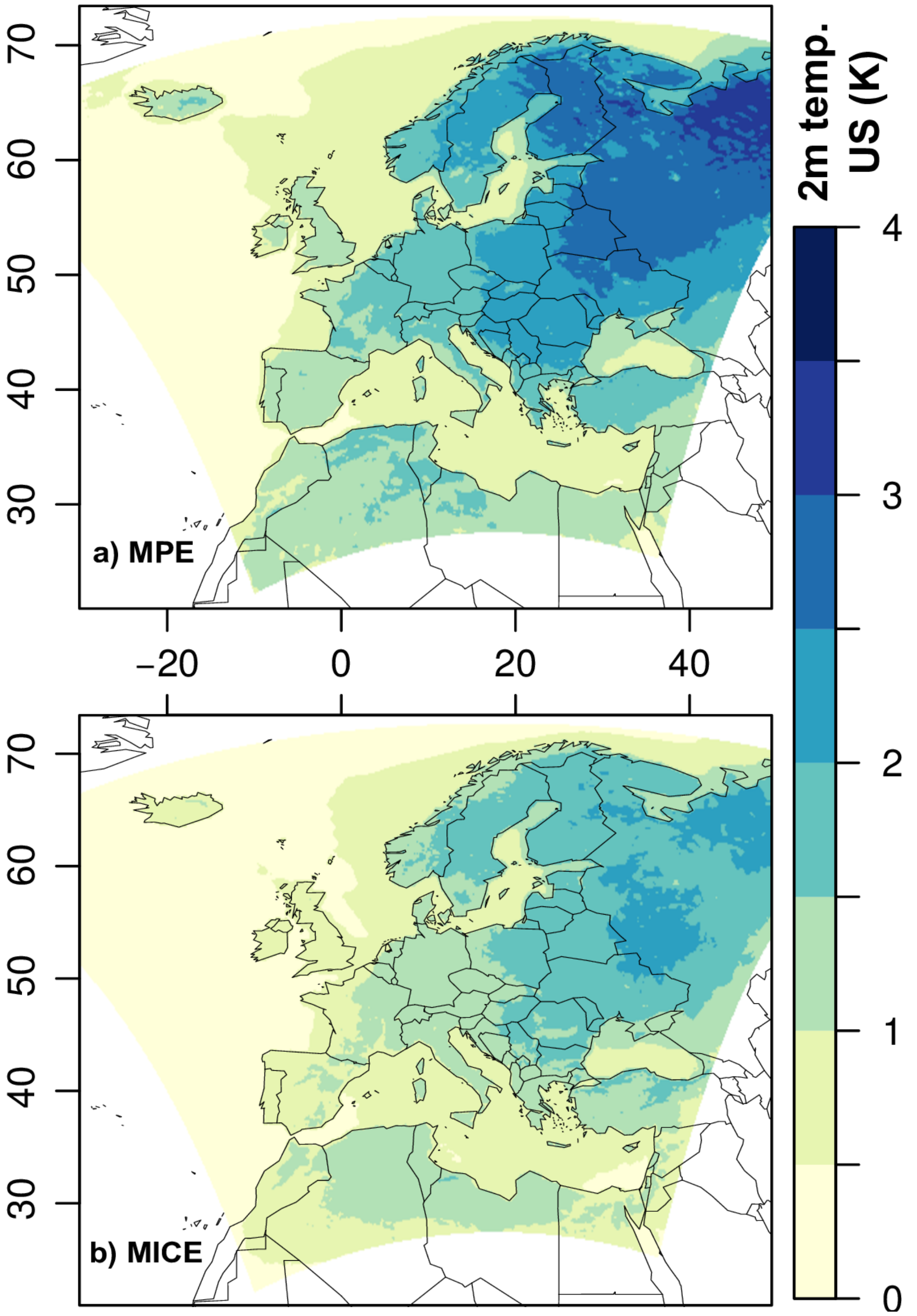

Figure 8: Spatial distribution of the inter-member variance for surface temperature (K) in EUR-15 domain of experiment B (year 1999). a) multi-physics ensemble. b) multi-initial-conditions ensemble.

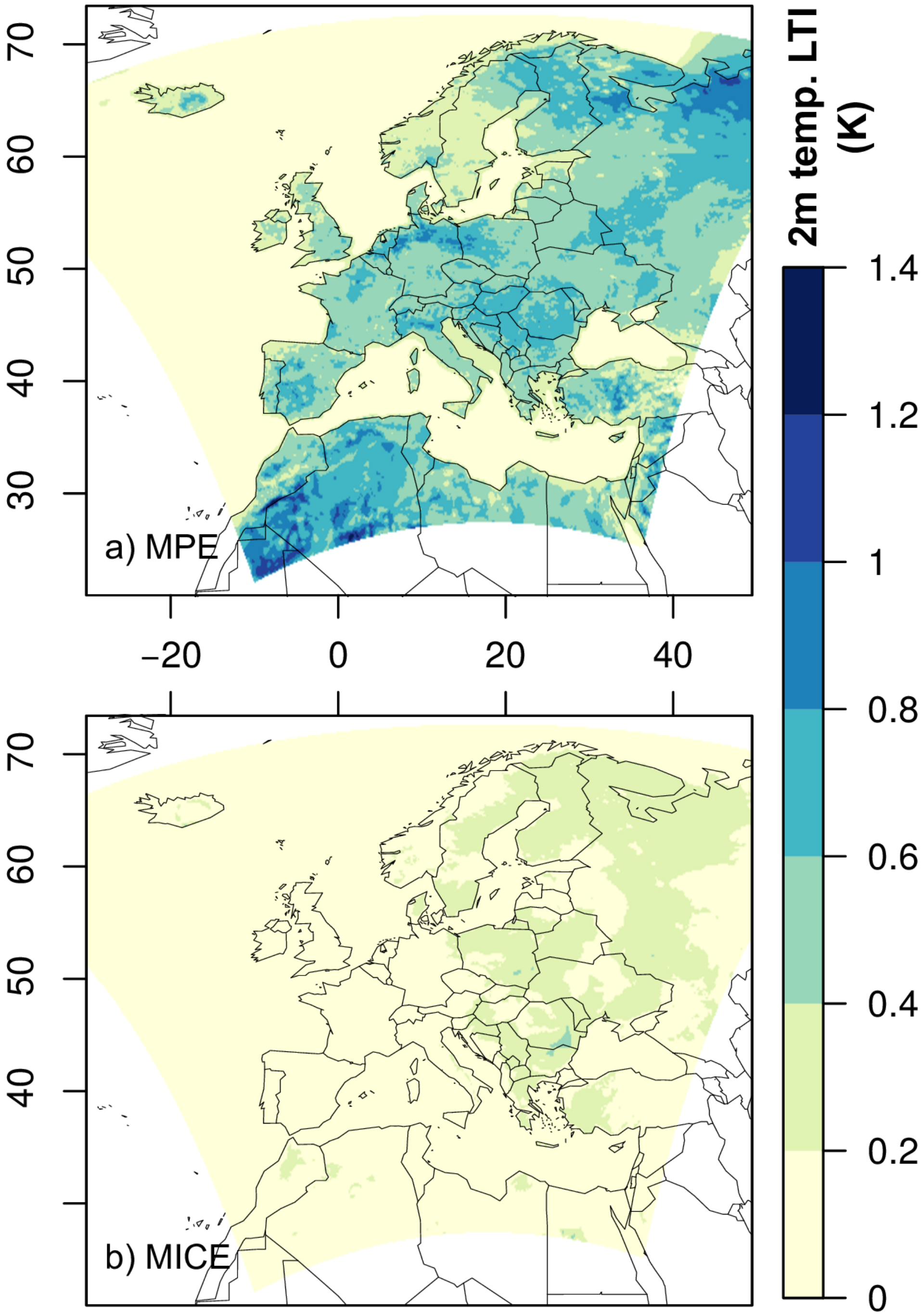

Figure 9: Long-term impact of multi-physics (a) and multi-initial-conditions (b) on surface temperature (K).