



**HAL**  
open science

## Internal variability versus multi-physics uncertainty in a regional climate model

Alvaro Lavin-gullon, Jesus Fernandez, Sophie Bastin, Rita M. Cardoso, Lluís Fita, Theodore M. Giannaros, Klaus Goergen, Jose Manuel Gutiérrez, Stergios Kartsios, Eleni Katragkou, et al.

### ► To cite this version:

Alvaro Lavin-gullon, Jesus Fernandez, Sophie Bastin, Rita M. Cardoso, Lluís Fita, et al.. Internal variability versus multi-physics uncertainty in a regional climate model. *International Journal of Climatology*, 2021, 41 (S1), pp.E656-E671. 10.1002/joc.6717 . insu-02885105

**HAL Id: insu-02885105**

**<https://insu.hal.science/insu-02885105v1>**

Submitted on 3 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3 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

4 June 15, 2020

5 A. Lavin-Gullon, J.M. Gutierrez

6 Grupo de Meteorología. Instituto de Física de Cantabria (IFCA), CSIC-  
7 Universidad de Cantabria. Santander. Spain

8 `alvaro.lavin@unican.es`

9 J. Fernandez, J. Milovac

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

12 T. Lorenz, S. Sobolowski

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

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

17 L. Fita

18 Centro de Investigaciones del Mar y la Atmósfera (CIMA), CONICET-UBA,

19 CNRS UMI-3351 IFAECI, C. A. Buenos Aires, Argentina  
20 K. Goergen  
21 Institute of Bio- and Geosciences (Agrosphere, IBG-3), Forschungszentrum  
22 Jülich (FZJ), Jülich, Germany  
23 E. Katragkou, S. Kartsios  
24 Department of Meteorology and Climatology, School of Geology, Aristotle  
25 University of Thessaloniki, Thessaloniki, Greece, R. M. Cardoso, P. M. M.  
26 Soares  
27 Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Lisbon,  
28 Portugal  
29 T. M. Giannaros  
30 National Observatory of Athens, Institute for Environmental Research and  
31 Sustainable Development, Athens, Greece  
32 K. Warrach-Sagi  
33 Institute of Physics and Meteorology (IPM), University of Hohenheim, Stuttgart,  
34 Germany

35 **Abstract**

36 In a recent study, Coppola et al (2020) assessed the ability of an en-  
37 semble of convection-permitting models (CPM) to simulate deep con-  
38 vection using three case studies. The ensemble exhibited strong dis-  
39 crepancies between models, which were attributed to various factors.  
40 In order to shed some light on the issue, we quantify in this paper  
41 the uncertainty associated to different physical parameterizations from  
42 that of using different initial conditions, often referred to as the inter-  
43 nal variability. For this purpose, we establish a framework to quantify  
44 both signals and we compare them for upper atmospheric circulation  
45 and near-surface variables. The analysis is carried out in the context of  
46 the CORDEX Flagship Pilot Study on Convective phenomena at high  
47 resolution over Europe and the Mediterranean, in which the interme-  
48 diate RCM WRF simulations that serve to drive the CPM are run  
49 several times with different parameterizations. For atmospheric circu-  
50 lation (geopotential height), the sensitivity induced by multi-physics  
51 and the internal variability show comparable magnitudes and a similar  
52 spatial distribution pattern. For 2-meter temperature and 10-meter  
53 wind, the simulations with different parameterizations show larger dif-  
54 ferences than those launched with different initial conditions. The  
55 systematic effect over one year shows distinct patterns for the multi-  
56 physics and the internal variability. Therefore, the general lesson of  
57 this study is that internal variability should be analyzed in order to  
58 properly distinguish the impact of other sources of uncertainty, espe-  
59 cially for short-term sensitivity simulations.

60 **Keywords:** Internal variability, Regional climate models, Uncer-  
61 tainty, Physical parameterizations, Ensemble

## 62 1 Introduction

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

78 Internal, unforced climate variability is one of the main sources of uncer-  
79 tainty in global climate simulations (Hawkins and Sutton, 2009). Due to the  
80 non-linear and chaotic nature of the climate system, small perturbations to a  
81 given state of the system grow and develop different trajectories in the state  
82 space (Palmer, 2005). In a relatively short period of time, two slightly per-  
83 turbed simulations in which initial conditions are modified can differ as much  
84 as two randomly chosen states of the climate system (Kalnay, 2003). When  
85 considering coupled systems that exhibit modes of low-frequency variability,  
86 even mean states over long periods of time can differ considerably. This  
87 internal or natural variability of the system is commonly explored using en-  
88 sembles of simulations started from perturbed initial conditions (Haughton

89 et al, 2014). The uncertainty arising from internal variability is not negli-  
90 gible compared to other sources of uncertainty, such as GCM modelling or  
91 GHG-scenario uncertainty (Hawkins and Sutton, 2009; Deser et al, 2012;  
92 van Pelt et al, 2015; Kumar and Ganguly, 2018).

93 In contrast, internal variability emerging in regional climate models  
94 (RCMs) is usually smaller than that in GCMs (Caya and Biner, 2004). This  
95 uncertainty is also commonly assessed by using a multi-initial-conditions  
96 ensemble (MICE) in order to separate RCM internal variability from the  
97 signal of forced variability (Giorgi and Bi, 2000; Christensen et al, 2001;  
98 Caya and Biner, 2004; Lucas-Picher et al, 2008b; Giorgi, 2019). Several  
99 studies concluded that at least 5-6 members should be considered to obtain  
100 robust estimates of internal variability (Lucas-Picher et al, 2008b; Laux et al,  
101 2017). The amplification of perturbations in the initial conditions is damped  
102 somewhat by the continuous flow of information through the boundaries of  
103 the limited area domain. Lucas-Picher et al (2008a) quantified the relation  
104 between the RCM internal variability and the lateral boundary forcing over  
105 the domain. In mid-latitudes, internal variability has a seasonal behaviour  
106 with higher (lower) values in summer (winter), when the boundary forcing  
107 (e.g. storm track intensity) is weaker (stronger) and the model is more (less)  
108 free to develop its own circulation (Caya and Biner, 2004; Lucas-Picher et al,  
109 2008b). According to the general atmospheric circulation, prevalent winds  
110 (e.g. westerlies in mid-latitudes) force a flow of information through the  
111 boundary. As a result, this forcing imposes a typical pattern that exhibits  
112 increasing internal variability as one travels downwind across the domain.  
113 Flow perturbations develop and grow as they travel through the RCM do-  
114 main, reaching a maximum near the downwind boundary where they are  
115 forced back to the flow of the GCM in the relaxation zone (Lucas-Picher

116 et al, 2008b).

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

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

138       Multi-RCM ensembles sample the dynamical downscaling methodolog-  
139 ical uncertainty. As such, it is challenging to discern the contributions to  
140 uncertainty from other sources (e.g. physical process parameterizations, in-  
141 ternal variability). This is because RCMs developed by different groups  
142 differ in so many aspects that the results from different models and mem-

143 bers cannot be used to understand the processes responsible for the spread.  
144 There have been different attempts to decompose multi-model uncertainty  
145 into other sources of uncertainty that can be more systematically explored.  
146 Perturbed-Physics Ensembles (PPE; Yang and Arritt, 2002; Bellprat et al,  
147 2012) consider a given RCM and explore the uncertainty associated to se-  
148 lected parameters, by sweeping a range of acceptable parameter values. This  
149 approach allows to link the resulting uncertainty to a specific parameter.  
150 Multi-physics ensembles (MPE; see e.g. García-Díez et al, 2015) provide a  
151 way to link modelling uncertainties to specific processes. These ensembles  
152 are generated using a single RCM by switching between different alternative  
153 physical parameterizations, which are the model components representing  
154 sub-grid-scale processes such as cloud microphysics, radiation, turbulence,  
155 etc. Physical parameterization are one of the key differences between dif-  
156 ferent RCMs and, therefore, MPEs mimic multi-model ensembles with the  
157 advantage of a fixed dynamical core and the rest of non-sampled physics  
158 schemes. Of course, these fixed components also limit model diversity and,  
159 therefore, MPEs cannot replace multi-model ensembles. Quite a few anal-  
160 yses tested the ability of different MPEs to encompass the regional climate  
161 in different areas (Fernández et al, 2007; Evans et al, 2012; Solman and  
162 Pessacg, 2012; Jerez et al, 2013; García-Díez et al, 2015; Katragkou et al,  
163 2015; Stegehuis et al, 2015; Devanand et al, 2018). Some of these analyses  
164 mentioned internal variability as potential source of background noise that  
165 impacts the sensitivity to the physical parameterization schemes (Tourpali  
166 and Zanis, 2013; Stegehuis et al, 2015), though internal variability was not  
167 formally investigated.

168 Few studies consider both physics sensitivity and internal variability. For  
169 instance, Laux et al (2017) explicitly aim to separate the effects of internal



170 variability from those of changes in land-use, suggesting that internal vari-  
171 ability has a significant impact on precipitation. Crétat and Pohl (2012)  
172 also studied the effect of physical parameterizations on internal variability  
173 and questioned the robustness of previous physics sensitivity studies which  
174 did not take into account internal variability.

175 The Flagship Pilot Study on Convective phenomena at high resolution  
176 over Europe and the Mediterranean (FPS-Convection) is an ongoing ini-  
177 tiative endorsed by CORDEX. This initiative aims at studying convective  
178 processes with CPM over the Alpine region (Coppola et al, 2020) by produc-  
179 ing both multi-model and multi-physics ensembles of RCM simulations. The  
180 initial results showed large discrepancies between individual ensemble mem-  
181 bers in their representation of selected heavy precipitation events. In this  
182 work, we take advantage of the ensembles produced in the FPS-Convection  
183 to follow up the study of Coppola et al (2020), in which the origin of these  
184 discrepancies was determined out of the scope. Since causation is difficult to  
185 address in a multi-model approach, we focus on the multi-physics ensemble  
186 within the FPS-Convection RCMs that serve to drive the CPM. We quan-  
187 titatively compare the signal arising from the use of different model compo-  
188 nents (physical parameterizations) against that associated to the background  
189 noise referred to internal variability at different time scales. The objective  
190 is twofold: (1) to assess whether modelling discrepancies in Coppola et al  
191 (2020) fall within the range of internal variability and (2) to quantify how  
192 much uncertainty in a multi-physics ensemble can be explained by internal  
193 variability.

194 The paper is structured as follows: The methodology and data used  
195 in this work are detailed in Section 2. Section 3 presents and discusses the  
196 results. First, applied to a case study presented in Coppola et al. (2020) and,

197 second, we extend the study to consider the role and relative magnitude of  
198 internal variability with respect to multi-physics uncertainty over an annual  
199 cycle. Finally, the conclusions are summarized in Section 4.

## 200 **2 Data & methods**

### 201 **2.1 Multi-physics ensemble**

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

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

221 All FPS-Convection WRF simulations consider a high-resolution ( $\sim 3\text{km}$ ),

222 convection-permitting domain centered over the Alpine region (ALP-3) nested  
223 into a coarser-resolution ( $\sim 12$  km), and much larger, pan-European domain.  
224 Except for the deep convection parameterization scheme, that is switched off  
225 in ALP-3, physical configuration does not differ between both domains. All  
226 WRF ensemble members used one-way nesting, so there is no communica-  
227 tion from the convection-permitting back to the coarser domain. Therefore,  
228 the convection-permitting inner domain did not alter in any way the results  
229 for the pan-European domain used in this work. Our analyses focus only  
230 on this pan-European domain, since we are interested in the uncertainty of  
231 the synoptic conditions over Europe, which drive the needed moisture that  
232 leads to unstable conditions over the Alpine area (see Section 3.1). The  
233 ALP-3 domain is not large enough to alter significantly the large-scale syn-  
234 optic conditions, so, in order to reproduce the case studies of Coppola et al  
235 (2020) in the ALP-3 domain, the right sequence of observed events should  
236 be preserved first in the pan-European domain forcing simulations.

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

240 **Experiment A** is described in Coppola et al (2020) and consisted of a  
241 preliminary test with all participating models, including WRF. Three heavy  
242 precipitation events in the Alpine region were simulated in two modes, iden-  
243 tified as “weather-like” and “climate mode”. Weather-like simulations were  
244 started one day before the onset of the events, aiming at simulating the event  
245 as closely as possible to the reality, aided by the predictability provided by  
246 the initial conditions. As the proximity of the initial conditions constrains  
247 the internal variability, we did not consider weather-like simulations in this

248 study. Climate-mode simulations were started one month before the event,  
249 so that initial conditions were not a source of predictability in this case and  
250 the models were mainly driven by the lateral boundary conditions, which  
251 is typical in regional climate modeling. We focus on a single event that  
252 occurred around the 23<sup>rd</sup> June, 2009, and was covered by climate-mode  
253 simulations running for the period from 1<sup>st</sup> June to 1<sup>st</sup> July, 2009 (see Sec-  
254 tion 3.1). WRF members of the ensemble showed the largest differences in  
255 terms of predictability of this particular event. WRF simulations for this ex-  
256 periment used a pan-European domain at  $0.11^\circ \times 0.11^\circ$  horizontal resolution  
257 (EUR-11), corresponding to the official EURO-CORDEX domain setup.

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

## 274 **2.2 Multi-initial-conditions ensemble**

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

286 We ran the 1-year MICE corresponding to experiment B (BI-r1 to BI-  
287 r5) only for the EUR-15 domain, without the inner ALP-3 nesting, so as to  
288 significantly reduce computational demands. Since no feedback from ALP-  
289 3 back to EUR-15 was allowed in the MPE, our EUR-15 MICE is fully  
290 comparable to EUR-15 MPE.

## 291 **2.3 Quantification of uncertainty**

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

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

295 where  $X(s, t, m)$  is the value of a given variable  $X$  at position  $s$  (summariz-  
296 ing, in this case, typical bi-dimensional position indices  $i, j$ ), at time step  $t$

297 and from ensemble member  $m$ .  $M$  is the total number of ensemble members.  
 298 The term  $\langle X \rangle(s, t)$  is the ensemble mean at a given position  $s$  and time  $t$ :

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

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

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

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

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

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

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

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

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

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

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

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

336 where  $\bar{X}^t(s, m)$  is the time average (i.e. the climatology) of each ensemble

337 member  $m$  and  $\langle \bar{X}^t \rangle(s)$  is the ensemble mean of the climatologies. Note  
338 that LTI measures the "uncertainty" of climate, while US measures the  
339 "climate" of the uncertainty. The latter is sensitive to the correspondence  
340 of meteorological events (e.g. heavy precipitation convective events) in time  
341 and space, while the former measures systematic deviations among members  
342 that lead to a different mean state (climate).

### 343 **3 Results & discussion**

#### 344 **3.1 Event reproducibility**

345 As an example, we focus first on a heavy precipitation case study ana-  
346 lyzed by Coppola et al (2020). The event was mostly driven by large-scale  
347 features, which consisted of a cut-off low over the Balkans inducing a persis-  
348 tent northeasterly flow over Austria. This unstable flow was warm and wet  
349 enough to trigger extreme precipitation by orographic lifting upon reaching  
350 the Alps. Observations reveal precipitation peaking on the 23<sup>rd</sup> June, 2009,  
351 over Austria. RCM simulations consistently reproduced this heavy precipi-  
352 tation event under weather-like initialization (see Section 2.1), but Coppola  
353 et al (2020) reported mixed results when considering the climate-mode ini-  
354 tialization. Some members of the multi-model/multi-physics ensemble com-  
355 pletely missed the precipitation event or represent highly damped versions  
356 of it (see Figure 4 of Coppola et al (2020)). They speculated on a poten-  
357 tially weak background synoptic forcing for this event, which we investigate  
358 in this work.

359 Notably, the WRF MPE alone also exhibited mixed results in reproduc-  
360 ing the event. For illustration, Figure 1 (left) shows the accumulated pre-  
361 cipitation on 23<sup>rd</sup> June for 4 WRF configurations. Only WRF configuration



362 AF is able to reproduce the event, with extended precipitation over Austria.  
363 Other WRF configurations (AB, AE, AD) miss the event and show some  
364 precipitation over southern Italy or very scarce precipitation (configurations  
365 AC, AG, AI, not shown in Figure 1).

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

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

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

395 An alternative hypothesis is that the different development of the event  
396 in the different MPE members is just the result of internal variability. To  
397 test this hypothesis, we considered a MICE based on configuration AI, which  
398 did not develop the event under the standard MPE initialization setup (start  
399 date: 00UTC, 1<sup>st</sup> June, 2009). Configuration AI (AI-r0) developed a low  
400 over southern Italy (Figure 2a), as many of the other configurations (Fig-  
401 ure 1). Many of the MICE members also developed a low over this area  
402 (see e.g. Figure 2), but member AI-r1 (start date: 00UTC, 31<sup>st</sup> May, 2009)  
403 presents a low in the right place, when compared against ERA-Interim.  
404 This was achieved by perturbing the initial conditions, starting the simula-  
405 tion one day earlier, and preserving exactly the same model configuration.  
406 Note that this is not a matter of improved initial conditions, since there are  
407 more than 20 days simulated from the geopotential height fields shown in  
408 Figures 1 (right) and 2, well beyond the limit of deterministic predictabil-  
409 ity of an atmospheric state. This is the result of internal variability. The  
410 slight perturbations in the initial conditions grew up by the non-linear dy-  
411 namical model. This process is in competition with the constraints imposed  
412 by the lateral boundary conditions, which bring the flow towards that of  
413 ERA-Interim close to border of the domain. This constraint can be seen in  
414 Figures 1 (right) and 2.

415 In this particular flow state, there seem to be two preferred weather

416 regimes over the southern Mediterranean area or, at least, our model sim-  
417 ulations were only able to generate these two weather regimes: one with  
418 a low evolving over southern Italy and the other with the low positioned  
419 over the Balkans. The observed flow took the Balkan low path even though  
420 the model has difficulties to reproduce this path. Note that these weather  
421 regimes and their probability of occurrence are likely model dependent. In  
422 any case, this is just one particular event. Once we have shown that internal  
423 variability can trigger flow deviations similar to those from different physi-  
424 cal parameterizations, we focus on quantifying their relative uncertainty, i.e.  
425 the spread of MPE and MICE ensembles.

426       The evolution of inter-member variance in time for MPE and MICE (Fig-  
427 ure 3) can reach comparable values. MPE member simulations take exactly  
428 the same initial and lateral boundary conditions from ERA-Interim, hence  
429 the uncertainty (essentially the member-to-member variability) at the start  
430 is very small (close to zero during the first day), indicating that all members  
431 produce similar circulation patterns. As the different physical parameteri-  
432 zations have an effect on the model, each member simulated a different syn-  
433 optic situation and the uncertainty increases. Regarding the MICE, since  
434 its members were initialized before the MPE start date shown in Figure  
435 3, the spread among members is larger than in the MPE in the beginning  
436 of June. MICE uncertainty (i.e. internal variability) remains fairly stable  
437 along the 1-month time span of the simulation. After about 10 days, the  
438 magnitude of MPE and MICE inter-member variance are comparable, with  
439 internal variability (MICE spread) generally larger than MPE spread. This  
440 suggests that the different physical parameterizations used in the MPE in-  
441 troduce smaller differences among members than those arising from internal  
442 variability.

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

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

### 463 **3.2 Analysis over an annual cycle**

464 We extended the analysis to an one-year period taking advantage of FPS-  
465 Convection experiment B (Section 2.1). In particular, we extended Figure 3  
466 to one year using the year 1999 from the WRF MPE of experiment B and  
467 a MICE based on configuration BI. The resulting inter-member variance in  
468 time (Figure 4) shows a very similar behaviour of MPE spread and inter-

469 nal variability (MICE spread) along the whole year. MPE members started  
470 again from the same initial conditions. Therefore, they show very low dif-  
471 ferences on January 1st, which increases after about 10 days. After this  
472 10-day transient evolution affected by the initial conditions, both ensembles  
473 show comparable inter-member variance, exhibiting an annual cycle with  
474 increased uncertainty in summer. Moreover, even weekly to monthly vari-  
475 ability in these UT time series seems to match in both ensembles. Notably  
476 in the last months (Oct-Dec), and also in many other peaks along the year.  
477 This suggests that the differences introduced by the different physics formu-  
478 lations along the time are amplified by the model in a similar way than the  
479 perturbations of the initial conditions. No systematic effect is noticeable in  
480 the circulation. Put in another way, for this variable at least, multi-physics  
481 uncertainty can be fully explained by internal variability.

482 As in previous studies (Caya and Biner, 2004; Lucas-Picher et al, 2008b),  
483 we used transient-eddy variability (Equation 5) as a reference for uncer-  
484 tainty. This is the natural variability of a meteorological field associated to  
485 weather systems traveling along the storm track. TEV can be computed  
486 from any of the ensemble members. We used simulation BI (top line in  
487 Figure 4), which is the only member common to both MPE and MICE. To  
488 evaluate the uncertainty associated to the selection of this particular mem-  
489 ber, we computed the monthly TEV from each member, and its standard  
490 deviation for each ensemble and for each month is shown as error bars in  
491 Figure 4. TEV spread is very low and any member could have been used as  
492 the reference. As already found in previous studies in mid-latitudes, TEV is  
493 larger in winter than in summer, due to the more frequent passage of weather  
494 systems from the Atlantic. The faster atmospheric circulation in winter im-  
495 poses a strong boundary forcing, which may explain the lower spread among

496 ensemble members. TEV and the associated boundary forcing is lower dur-  
497 ing summer. As a result, the model has more freedom to develop its own  
498 circulation features, increasing the spread between the members. During  
499 summer, the spread reaches approximately half of the TEV, which would be  
500 the maximum attainable. This maximum is what one would expect from a  
501 GCM, which has no lateral boundary constraints. For such a model, MICE  
502 spread (i.e. internal variability) would increase during 1-2 weeks to reach  
503 the TEV line and remain around this limit along the year. In this sense,  
504 RCM internal variability is negligible compared to GCM internal variability  
505 during winter, but it represents an important fraction (approximately one  
506 half, in this example) during summer.

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

523 shown) are very similar to those in Figure 5. They show higher (lower)  
524 intensity in JJA (DJF), reaching 45 m (25 m) over the Baltic Sea.

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

537 The LTI of internal variability (Figure 6b) shows a distinct pattern, with  
538 the largest values in the northern half of the domain. The magnitude is  
539 comparable to that of the MPE, though. Therefore, even though the spatial  
540 patterns are different, the systematic differences among MPE members are  
541 still comparable to the internal variability. This would suggest that one-year  
542 simulations are not enough to distinguish the systematic effect of a particular  
543 parameterization configuration compared to the impact of different initial  
544 conditions on the circulation. Since the MICE is just composed of multiple  
545 realizations of the same model configuration, its LTI must tend to zero as the  
546 simulation length increases and the climatology of all members tends towards  
547 the “true” model climatology. Longer simulations, such as those currently  
548 under way in the FPS-Convection, should provide a better assessment of  
549 the LTI of the MPE. For example, for 10-year simulations, the values on

550 Figure 6b should be divided by a factor of  $\sqrt{10} \approx 3.2$  (Lucas-Picher et al,  
551 2008b). Up to this point, we have focused on the circulation (850 hPa  
552 geopotential height) and we have seen that multi-physics uncertainty is hard  
553 to distinguish from internal variability. The results for the circulation at 700  
554 hPa or 500 hPa (not shown) are qualitatively similar.

### 555 **3.3 Surface variables**

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

566 The evolution of inter-member variance for near-surface temperature,  
567 both for the MPE and MICE is different from the geopotential height shown  
568 in Figure 4. The annual cycle is clearer in the TEV than in the variance,  
569 which only shows a hint of a seasonal cycle during April through October.  
570 In summer, MPE and MICE spread evolution is uncorrelated, with some  
571 peak MPE uncertainty events (e.g. end of July) clearly standing out of  
572 internal variability. However, the strong winter variability seems coherent  
573 between MPE and MICE spread. Even if multi-physics spread is usually  
574 the greatest, internal variability seems to modulate it. This is in appar-  
575 ent contradiction with the results of Cr etat and Pohl (2012), who claimed



576 that physical parameterizations modulate IV. They show that two MICE  
577 under different physical parameterization configurations develop a different  
578 amount of IV on average. However, they also show (their Figure 4b) a co-  
579 herent evolution in time of the IV between model configurations. In our  
580 setup, physical parameterizations cannot modulate IV time evolution since  
581 the model configuration is fixed in the MICE. Still, Figure 7 shows that,  
582 despite the different spread amounts in MICE and MPE, both evolve coher-  
583 ently in time. It is likely that a third variable, such as the strength of the  
584 external forcing (i.e. boundary conditions), modulates the degree to which  
585 both physics and IV uncertainties can grow.

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

599       The spatial distribution of the inter-member variance for surface tem-  
600 perature (Figure 8) reveals, as before, a similar pattern of increasing spread  
601 towards the northeast in both ensembles. In this case, despite the similar  
602 pattern, MPE shows larger spread values in accordance with Figure 7. MPE

603 reaches a maximum value of about 3.5 K while MICE reaches about 2 K.

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

## 617 4 Conclusions

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

625 The analyses over the one-month period already shed light on the 2 main  
626 objectives of this work: (1) The failure of some WRF model configurations to  
627 reproduce the case study, as reported by Coppola et al (2020), is not related  
628 to physical parameterizations, but to the absence of a synoptic circulation

629 pattern that favoured the event. Some members of the MICE were able  
630 to reasonably reproduce the observed synoptic pattern without modifying  
631 the model parameterization setup. (2) From a quantitative perspective, the  
632 spread due to the parameterization differences has a magnitude comparable  
633 to that from internal variability. Therefore, in these one-month simulations,  
634 the effect of the different physical parameterizations on the circulation can-  
635 not be distinguished from internal variability.

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

644 The results, however, depend on the variable, with surface variables  
645 (known to be sensitive to parameterized processes) showing higher MPE  
646 spread. For example, for near-surface temperature the spread associated to  
647 parameterizations was above that due to the internal variability. This sug-  
648 gests that it is easier to discern both sources of uncertainties when analyzing  
649 variables more constrained by the model physics, which is typically the case  
650 in RCM parameterization sensitivity studies (Fernández et al, 2007; Evans  
651 et al, 2012; Solman and Pessacq, 2012; Jerez et al, 2013; García-Díez et al,  
652 2015; Katragkou et al, 2015; Stegehuis et al, 2015; Devanand et al, 2018).

653 As a reference for uncertainty, we computed transient-eddy variability,  
654 and quantified its spread due to the multi-physics and to internal variabil-  
655 ity. This type of uncertainty also depends on the variable. For the circu-

656 lation, transient-eddy variability of the different physical model configura-  
657 tions is similar to the internal variability range. However, for near-surface  
658 temperature, the different physics configurations exhibit a different level of  
659 transient-eddy variability. This requires further analysis on longer simula-  
660 tions to properly estimate the inter-annual contribution, but this is beyond  
661 the scope of the present work.

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

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

680 The inability of an RCM to reproduce the observed day-to-day circula-  
681 tion due to internal variability is not a matter of concern for mean cli-  
682 mate studies, given that long-term climate is preserved (Caya and Biner,

683 2004). However, with the arrival of convection-permitting simulations and  
684 the increasing interest in the climate of extremes, RCM internal variability  
685 re-emerges as a matter of concern for model evaluation. As an example, the  
686 FPS-Convection focuses on high-impact (low probability) convective phe-  
687 nomena that occur mainly during the summer season, when lateral bound-  
688 ary forcing is the weakest. The evaluation of models under these conditions  
689 poses a real challenge that can only be addressed by computationally expen-  
690 sive experiments including the simulation of long periods and/or the simula-  
691 tion of a corresponding MICE to disentangle the role of internal variability  
692 in the results. Other alternatives would be to constrain internal variability  
693 by using techniques such as spectral nudging, which has its own drawbacks  
694 (Alexandru et al, 2009), or frequently reinitializing the RCM (Lo et al, 2008;  
695 Lucas-Picher et al, 2013).

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

## 703 **Acknowledgements**

704 This work is partially funded by the Spanish government through grant  
705 BES-2016-078158 and MINECO/ FEDER co-funded projects INSIGNIA  
706 (CGL2016-79210-R) and MULTI-SDM (CGL2015-66583-R). Universidad de  
707 Cantabria simulations have been carried out on the Altamira Supercom-  
708 puter at the Instituto de Física de Cantabria (IFCA-CSIC), member of the

709 Spanish Supercomputing Network. EK and SK acknowledge the support of  
710 the Greek Research and Technology Network (GRNET) High Performance  
711 Computing (HPC) infrastructure for providing the computational resources  
712 of AUTH-simulations (under project ID pr003005) and the AUTH Scientific  
713 Computing Center for technical support. IPSL acknowledges the support  
714 from the EUCP project, funded by the European Union under H2020 Grant  
715 Agreement 776613, and from the IPSL mesocenter ESPRI facility which  
716 is supported by CNRS, UPMC, Labex L-IPSL, CNES and Ecole Polytech-  
717 nique. IPSL simulations were granted access to the HPC resources of TGCC  
718 under the allocation A0050106877 made by GENCI. The computational re-  
719 sources for NORCE/BCCR were provided by UNINETT Sigma2 (NN9280K,  
720 NS9001K), with funding from the Research Council of Norway's support for  
721 the strategic project on climate services. FZJ gratefully acknowledges the  
722 computing time granted by the John von Neumann Institute for Computing  
723 (NIC) and JARA-HPC provided on the supercomputer JURECA at Jülich  
724 Supercomputing Centre (JSC). We acknowledge the E-OBS dataset from  
725 the EU-FP6 project UERRA (<https://www.uerra.eu>) and the Copernicus  
726 Climate Change Service, and the data providers in the ECA&D project  
727 (<https://eca.knmi.nl>).

## 728 **References**

- 729 Alexandru A, de Elia R, Laprise R (2007) Internal variability in regional  
730 climate downscaling at the seasonal scale. *Monthly Weather Review*  
731 135(9):3221–3238, DOI 10.1175/MWR3456.1, URL [https://doi.org/  
732 10.1175/MWR3456.1](https://doi.org/10.1175/MWR3456.1)
- 733 Alexandru A, de Elia R, Laprise R, Separovic L, Biner S (2009) Sensitiv-

734 ity study of regional climate model simulations to large-scale nudging  
735 parameters. *Monthly Weather Review* 137(5):1666–1686, DOI 10.1175/  
736 2008MWR2620.1

737 Bellprat O, Kotlarski S, Lüthi D, Schär C (2012) Exploring perturbed  
738 physics ensembles in a regional climate model. *Journal of Climate*  
739 25(13):4582–4599, DOI 10.1175/JCLI-D-11-00275.1

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

743 Christensen OB, Gaertner MA, Prego JA, Polcher J (2001) Internal variabil-  
744 ity of regional climate models. *Climate Dynamics* 17(11):875–887, DOI  
745 10.1007/s003820100154

746 Coppola E, Sobolowski S, Pichelli E, Raffaele F, Ahrens B, Anders I, Ban  
747 N, Bastin S, Belda M, Belusic D, Caldas-Alvarez A, Cardoso RM, Davo-  
748 lio S, Dobler A, Fernandez J, Fita L, Fumiere Q, Giorgi F, Goergen K,  
749 Güttler I, Halenka T, Heinzeller D, Hodnebrog Ø, Jacob D, Kartsios S,  
750 Katragkou E, Kendon E, Khodayar S, Kunstmann H, Knist S, Lavín-  
751 Gullón A, Lind P, Lorenz T, Maraun D, Marelle L, van Meijgaard E,  
752 Milovac J, Myhre G, Panitz HJ, Piazza M, Raffa M, Raub T, Rockel  
753 B, Schär C, Sieck K, Soares PMM, Somot S, Srnec L, Stocchi P, Tölle  
754 MH, Truhetz H, Vautard R, de Vries H, Warrach-Sagi K (2020) A first-  
755 of-its-kind multi-model convection permitting ensemble for investigating  
756 convective phenomena over Europe and the Mediterranean. *Climate Dy-*  
757 *namics* DOI 10.1007/s00382-018-4521-8

758 Crétat J, Pohl B (2012) How physical parameterizations can modulate inter-

759     nal variability in a Regional Climate Model. *Journal of the Atmospheric*  
760     *Sciences* 69(2):714–724, DOI 10.1175/JAS-D-11-0109.1

761     Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, An-  
762     drae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM,  
763     van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M,  
764     Geer AJ, Haimberger L, Healy SB, Hersbach H, Hlm EV, Isaksen L, Kill-  
765     berg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette  
766     JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F  
767     (2011) The ERA-Interim reanalysis: configuration and performance of the  
768     data assimilation system. *Quarterly Journal of the Royal Meteorological*  
769     *Society* 137(656):553–597, DOI 10.1002/qj.828

770     Deser C, Phillips A, Bourdette V, Teng H (2012) Uncertainty in climate  
771     change projections: the role of internal variability. *Climate Dynamics*  
772     38(3):527–546, DOI 10.1007/s00382-010-0977-x

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

777     Evans JP, Ekström M, Ji F (2012) Evaluating the performance of a  
778     WRF physics ensemble over South-East Australia. *Climate Dynamics*  
779     39(6):1241–1258, DOI 10.1007/s00382-011-1244-5

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



- 784 Fernández J, Montávez J, Sáenz J, González-Rouco J, Zorita E (2007) Sen-  
785 sitivity of the MM5 mesoscale model to physical parameterizations for  
786 regional climate studies: Annual cycle. *Journal of Geophysical Research:*  
787 *Atmospheres* 112(D4)
- 788 Fernández J, Primo C, Cofiño AS, Gutiérrez JM, Rodríguez MA (2009) MVL  
789 spatiotemporal analysis for model intercomparison in EPS: application to  
790 the DEMETER multi-model ensemble. *Climate Dynamics* 33(2):233–243,  
791 DOI 10.1007/s00382-008-0456-9
- 792 García-Díez M, Fernández J, Vautard R (2015) An RCM multi-  
793 physics ensemble over Europe: multi-variable evaluation to avoid er-  
794 ror compensation. *Climate Dynamics* 45(11):3141–3156, DOI 10.1007/  
795 s00382-015-2529-x
- 796 Giorgi F (2019) Thirty years of regional climate modeling: Where are we and  
797 where are we going next? *Journal of Geophysical Research: Atmospheres*  
798 DOI 10.1029/2018JD030094
- 799 Giorgi F, Bi X (2000) A study of internal variability of a regional climate  
800 model. *Journal of Geophysical Research: Atmospheres* 105(D24):29,503–  
801 29,521, DOI 10.1029/2000JD900269
- 802 Giorgi F, Gutowski WJ (2015) Regional dynamical downscaling and the  
803 CORDEX initiative. *Annual Review of Environment and Resources*  
804 40(1):467–490, DOI 10.1146/annurev-environ-102014-021217
- 805 Gu H, Yu Z, Yang C, Ju Q, Yang T, Zhang D (2018) High-resolution en-  
806 semble projections and uncertainty assessment of regional climate change  
807 over China in CORDEX East Asia. *Hydrology and Earth System Sciences*  
808 22(5):3087–3103, DOI 10.5194/hess-22-3087-2018

- 809 Haughton N, Abramowitz G, Pitman A, Phipps SJ (2014) On the generation  
810 of climate model ensembles. *Climate Dynamics* 43(7):2297–2308, DOI  
811 10.1007/s00382-014-2054-3
- 812 Hawkins E, Sutton R (2009) The potential to narrow uncertainty in re-  
813 gional climate predictions. *Bulletin of the American Meteorological Soci-*  
814 *ety* 90(8):1095–1108, DOI 10.1175/2009BAMS2607.1
- 815 Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD,  
816 New M (2008) A European daily high-resolution gridded data  
817 set of surface temperature and precipitation for 1950–2006. *Jour-*  
818 *nal of Geophysical Research: Atmospheres* 113(D20), DOI 10.1029/  
819 2008JD010201, URL [https://agupubs.onlinelibrary.wiley.com/](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010201)  
820 [doi/abs/10.1029/2008JD010201](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010201), <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD010201>
- 822 Jerez S, Montavez JP, Gomez-Navarro JJ, Lorente-Plazas R, Garcia-Valero  
823 JA, Jimenez-Guerrero P (2013) A multi-physics ensemble of regional cli-  
824 mate change projections over the Iberian Peninsula. *Climate Dynamics*  
825 41(7):1749–1768, DOI 10.1007/s00382-012-1551-5
- 826 Kalnay E (2003) *Atmospheric Modeling, Data Assimilation and Predictabil-*  
827 *ity*. Cambridge University Press
- 828 Katragkou E, García-Díez M, Vautard R, Sobolowski S, Zanis P, Alexandri  
829 G, Cardoso RM, Colette A, Fernandez J, Gobiet A, Goergen K, Kara-  
830 costas T, Knist S, Mayer S, Soares PMM, Pytharoulis I, Tegoulis I, Tsik-  
831 erdekis A, Jacob D (2015) Regional climate hindcast simulations within  
832 EURO-CORDEX: evaluation of a WRF multi-physics ensemble. *Geosci-*  
833 *entific Model Development* 8(3):603–618, DOI 10.5194/gmd-8-603-2015

- 834 Kumar D, Ganguly AR (2018) Intercomparison of model response and  
835 internal variability across climate model ensembles. *Climate Dynamics*  
836 51(1):207–219, DOI 10.1007/s00382-017-3914-4
- 837 Laux P, Nguyen PNB, Cullmann J, Van TP, Kunstmann H (2017) How  
838 many RCM ensemble members provide confidence in the impact of land-  
839 use land cover change? *International Journal of Climatology* 37(4):2080–  
840 2100, DOI 10.1002/joc.4836
- 841 Lo JCF, Yang ZL, Pielke Sr RA (2008) Assessment of three dynamical cli-  
842 mate downscaling methods using the weather research and forecasting  
843 (WRF) model. *Journal of Geophysical Research: Atmospheres* 113(D9),  
844 DOI 10.1029/2007JD009216, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009216>
- 846 Lucas-Picher P, Caya D, Biner S, Laprise R (2008a) Quantification of  
847 the lateral boundary forcing of a Regional Climate Model using an ag-  
848 ing tracer. *Monthly Weather Review* 136(12):4980–4996, DOI 10.1175/  
849 2008MWR2448.1, URL <https://doi.org/10.1175/2008MWR2448.1>
- 850 Lucas-Picher P, Caya D, de Elía R, Laprise R (2008b) Investigation of re-  
851 gional climate models' internal variability with a ten-member ensemble  
852 of 10-year simulations over a large domain. *Climate Dynamics* 31(7):927–  
853 940, DOI 10.1007/s00382-008-0384-8, URL <https://doi.org/10.1007/s00382-008-0384-8>
- 855 Lucas-Picher P, Boberg F, Christensen JH, Berg P (2013) Dynamical  
856 downscaling with reinitializations: A method to generate finescale cli-  
857 mate datasets suitable for impact studies. *Journal of Hydrometeorol-*

858 ogy 14(4):1159–1174, DOI 10.1175/JHM-D-12-063.1, URL <https://doi.org/10.1175/JHM-D-12-063.1>

859

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

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

867 van Pelt SC, Beersma JJ, Buishand TA, van den Hurk BJJM, Schellekens J  
868 (2015) Uncertainty in the future change of extreme precipitation over the  
869 Rhine basin: the role of internal climate variability. *Climate Dynamics*  
870 44(7):1789–1800, DOI 10.1007/s00382-014-2312-4

871 Prein AF, Langhans W, Fosser G, Ferrone A, Ban N, Goergen K, Keller  
872 M, Tlle M, Gutjahr O, Feser F, Brisson E, Kollet S, Schmidli J, van  
873 Lipzig NPM, Leung R (2015) A review on regional convection-permitting  
874 climate modeling: Demonstrations, prospects, and challenges. *Reviews of*  
875 *Geophysics* 53(2):323–361, DOI 10.1002/2014RG000475, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000475>

876

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

880 Sanchez-Gomez E, Somot S, Déqué M (2009) Ability of an ensemble of re-  
881 gional climate models to reproduce weather regimes over Europe-Atlantic

882 during the period 1961–2000. *Climate Dynamics* 33(5):723–736, DOI  
883 10.1007/s00382-008-0502-7

884 Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker M, Duda KG,  
885 Y Huang X, Wang W, Powers JG (2008) A description of the ad-  
886 vanced research WRF Version 3. NCAR Technical Note pp 1–113, DOI  
887 10.5065/D68S4MVH

888 Solman SA, Pessacg NL (2012) Regional climate simulations over South  
889 America: sensitivity to model physics and to the treatment of lateral  
890 boundary conditions using the MM5 model. *Climate Dynamics* 38(1):281–  
891 300, DOI 10.1007/s00382-011-1049-6

892 Stegehuis AI, Vautard R, Ciais P, Teuling AJ, Miralles DG, Wild M (2015)  
893 An observation-constrained multi-physics WRF ensemble for simulating  
894 european mega heat waves. *Geoscientific Model Development* 8(7):2285–  
895 2298, DOI 10.5194/gmd-8-2285-2015

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

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

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

905 Yang Z, Arritt RW (2002) Tests of a Perturbed Physics Ensemble Approach  
906 for Regional Climate Modeling. *Journal of Climate* 15(20):2881–2896,

907 DOI 10.1175/1520-0442(2002)015<2881:TOAPPE>2.0.CO;2, URL [https:](https://doi.org/10.1175/1520-0442(2002)015<2881:TOAPPE>2.0.CO;2)  
908 [//doi.org/10.1175/1520-0442\(2002\)015<2881:TOAPPE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2881:TOAPPE>2.0.CO;2)

909 **List of Tables**

910	1	WRF multi-physics configurations considered in this study	
911		(see Section 2.1) for experiment A (one-month simulation,	
912		EUR-11 domain) and experiment B (one-year simulation, EUR-	
913		15). For each ensemble member, the table shows an Id. code,	
914		the institution performing the simulation and the physical	
915		parameterizations used. The ensembles explore the use of	
916		different schemes for micro-physics (MP), planetary bound-	
917		ary layer and surface layer (PBL), land surface (LSM), and	
918		shallow convection (ShC) processes. The PBL schemes de-	
919		noted with asterisk (*) used a different surface layer scheme	
920		despite sharing the MYNN2 PBL. See Table 2 for details of	
921		each parameterization scheme. . . . .	39
922	2	Physical schemes used in the multi-physics experiments shown	
923		in Table 1. . . . .	40

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
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.

924 **List of Figures**

925 1 Left: Accumulated precipitation (mm) on June, 23rd 2009  
926 according to E-OBS (Haylock et al (2008); top) and as simu-  
927 lated in the ALP-3 domain by experiment A for WRF MPE  
928 members AF, AD, AB and AE. Right: 850hPa geopotential  
929 height (m) according to ERA-Interim (top) and the corre-  
930 sponding MPE ensemble members in the EUR-11 domain in  
931 pink. An ERA-Interim 1500m-isoline (the same in all panels)  
932 is represented for reference in black. . . . . 42

933 2 As Figure 1 (right), but for 4 MICE members: AI-r0 to AI-r3. 43

934 3 Inter-member variance in time (Equation 3) for 850hPa geopo-  
935 tential height (m) in EUR-11 domain of experiment A (June  
936 2009). The spread is computed separately for MPE (blue)  
937 and MICE (red). The latter was computed both at  $0.11^\circ$  and  
938  $0.44^\circ$  horizontal resolution with similar number of ensemble  
939 members. . . . . 44

940 4 Inter-member variance in time (UT) for 850hPa geopotential  
941 height (m) in EUR-15 domain of experiment B (year 1999).  
942 The uncertainty is computed separately for MPE (blue) and  
943 MICE (red). Transient-eddy variability (Equation 5, black  
944 line) was computed from BI configuration and error bars show  
945 its standard deviation for MPE and MICE. . . . . 45

946 5 Spatial distribution of the inter-member variance (US) for  
947 the 850 hPa geopotential height (m) in EUR-15 domain of  
948 experiment B (year 1999). a) multi-physics ensemble. b)  
949 multi-initial-conditions ensemble. . . . . 46

950 6 Long-term impact of multi-physics (a) and multi-initial-conditions  
951 (b) on 850hPa geopotential height (m). . . . . 47

952 7 As Fig. 4 but for surface temperature over land. . . . . 48

953 8 Spatial distribution of the inter-member variance for surface  
954 temperature (K) in EUR-15 domain of experiment B (year  
955 1999). a) multi-physics ensemble. b) multi-initial-conditions  
956 ensemble. . . . . 49

957 9 Long-term impact of multi-physics (a) and multi-initial-conditions  
958 (b) on surface temperature (K). . . . . 50

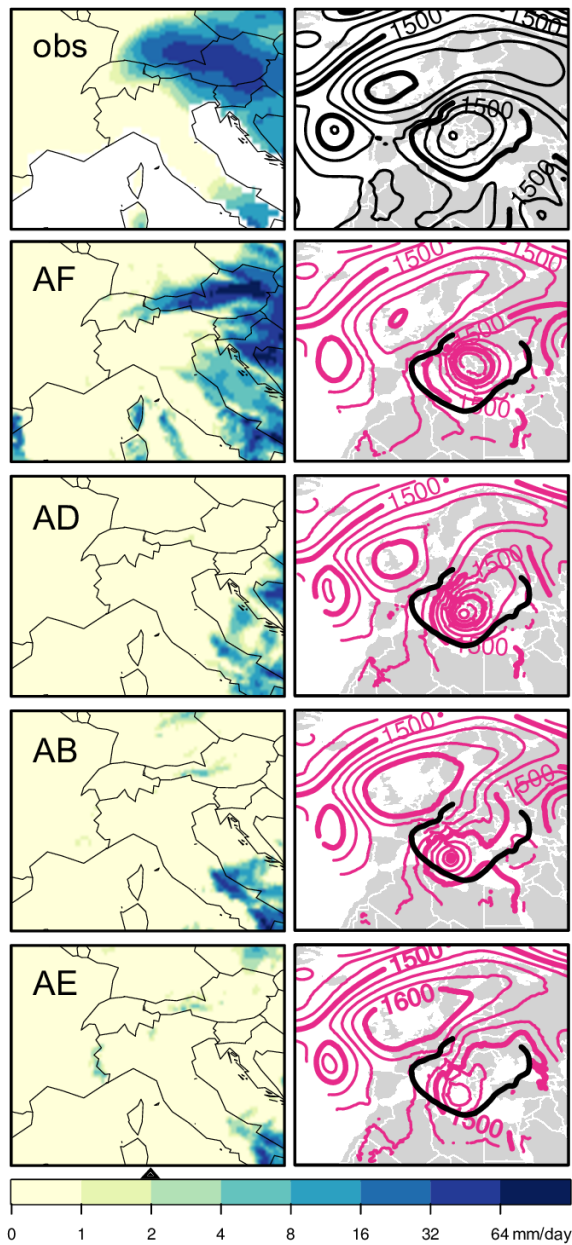


Figure 1: Left: Accumulated precipitation (mm) on June, 23rd 2009 according to E-OBS (Haylock et al (2008); top) and as simulated in the ALP-3 domain by experiment A for WRF MPE members AF, AD, AB and AE. Right: 850hPa geopotential height (m) according to ERA-Interim (top) and the corresponding MPE ensemble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

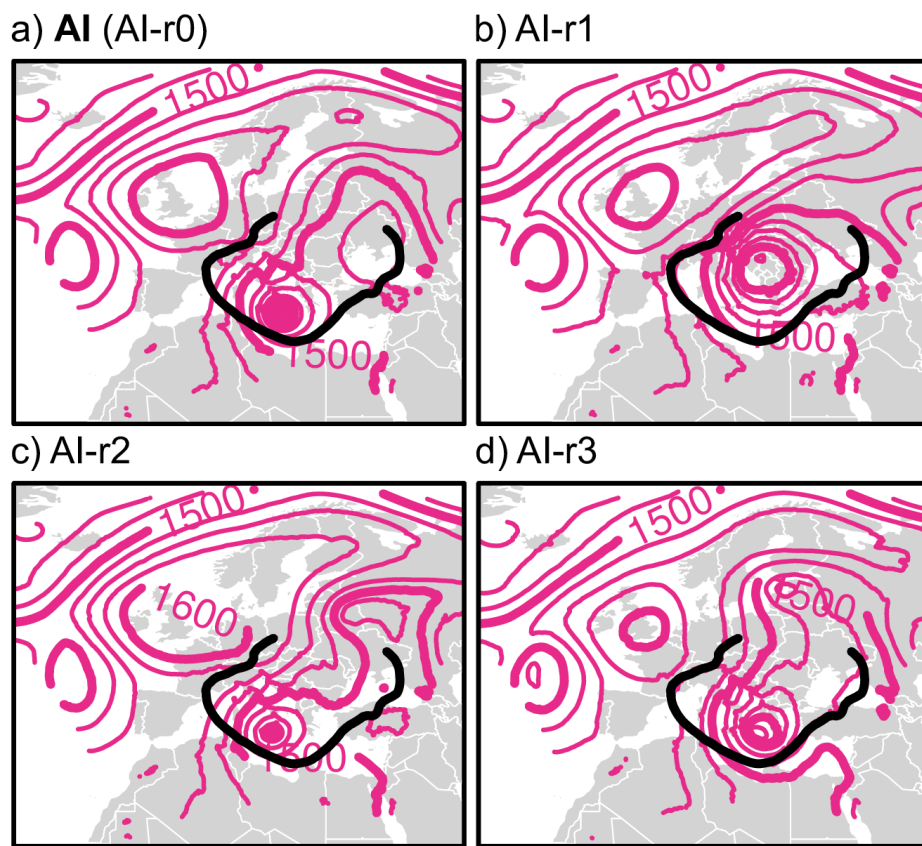


Figure 2: As Figure 1 (right), but for 4 MICE members: AI-r0 to AI-r3.

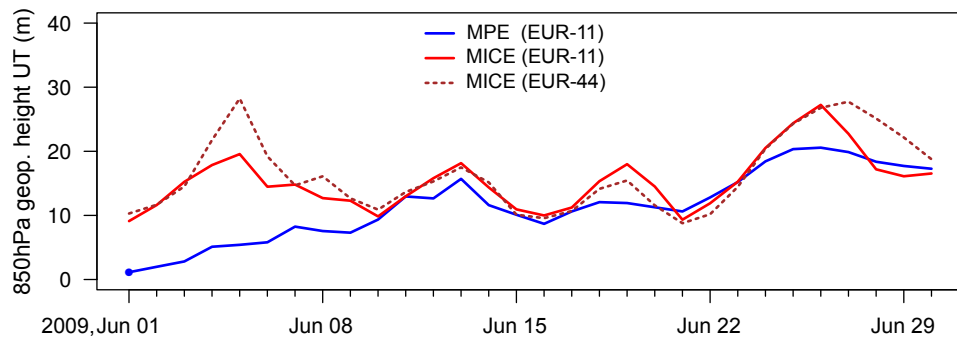


Figure 3: Inter-member variance in time (Equation 3) for 850hPa geopotential height (m) in EUR-11 domain of experiment A (June 2009). The spread is computed separately for MPE (blue) and MICE (red). The latter was computed both at  $0.11^\circ$  and  $0.44^\circ$  horizontal resolution with similar number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

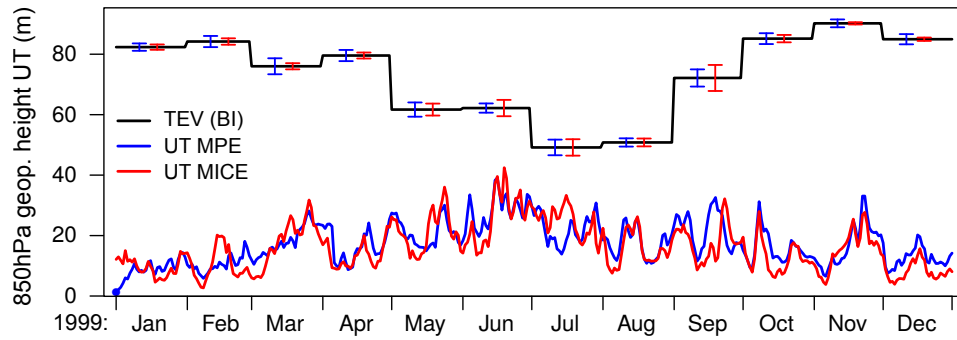


Figure 4: Inter-member variance in time (UT) for 850hPa geopotential height (m) in EUR-15 domain of experiment B (year 1999). The uncertainty is computed separately for MPE (blue) and MICE (red). Transient-eddy variability (Equation 5, black line) was computed from BI configuration and error bars show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

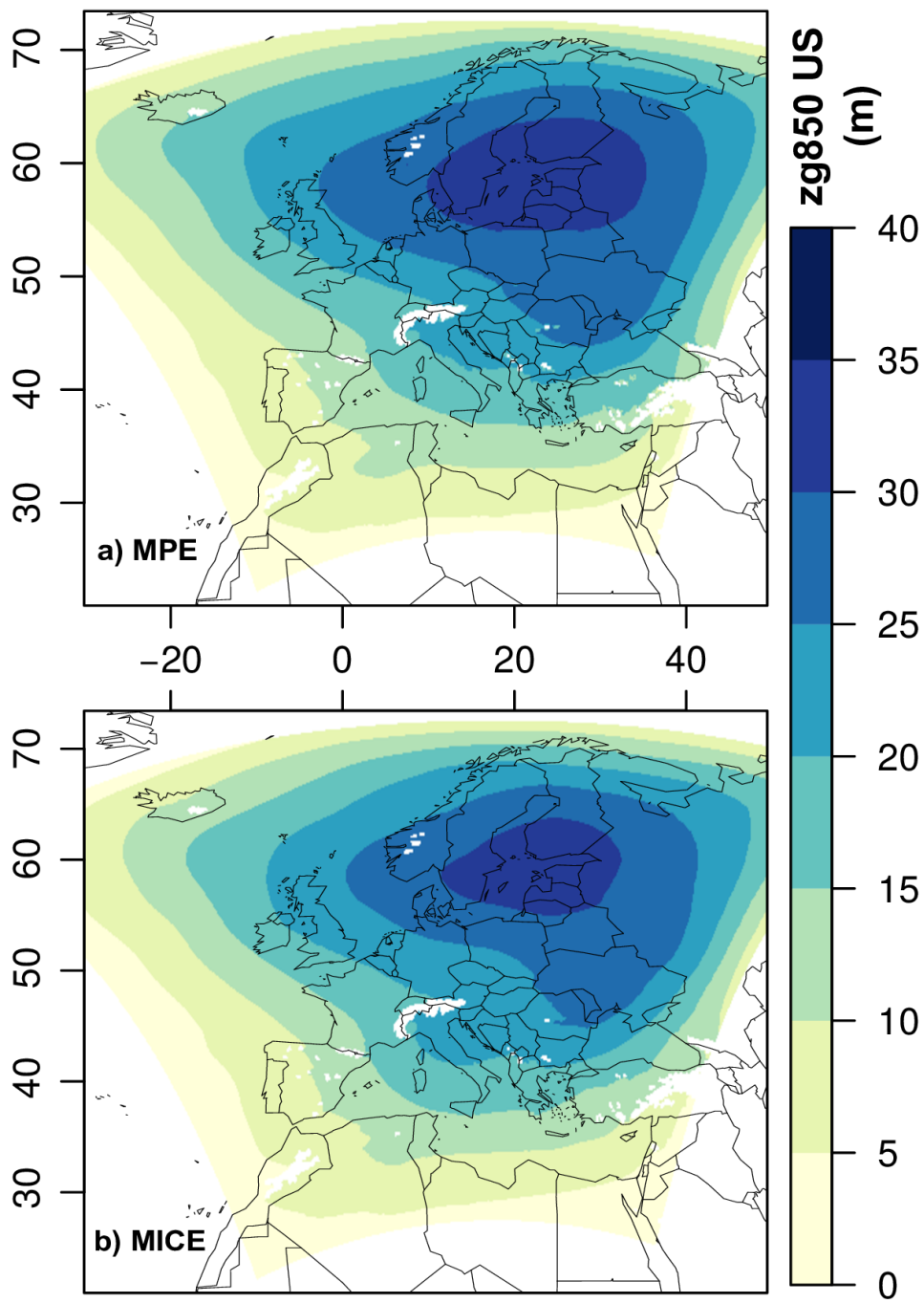


Figure 5: Spatial distribution of the inter-member variance (US) for the 850 hPa geopotential height (m) in EUR-15 domain of experiment B (year 1999). a) multi-physics ensemble. b) multi-initial-conditions ensemble.

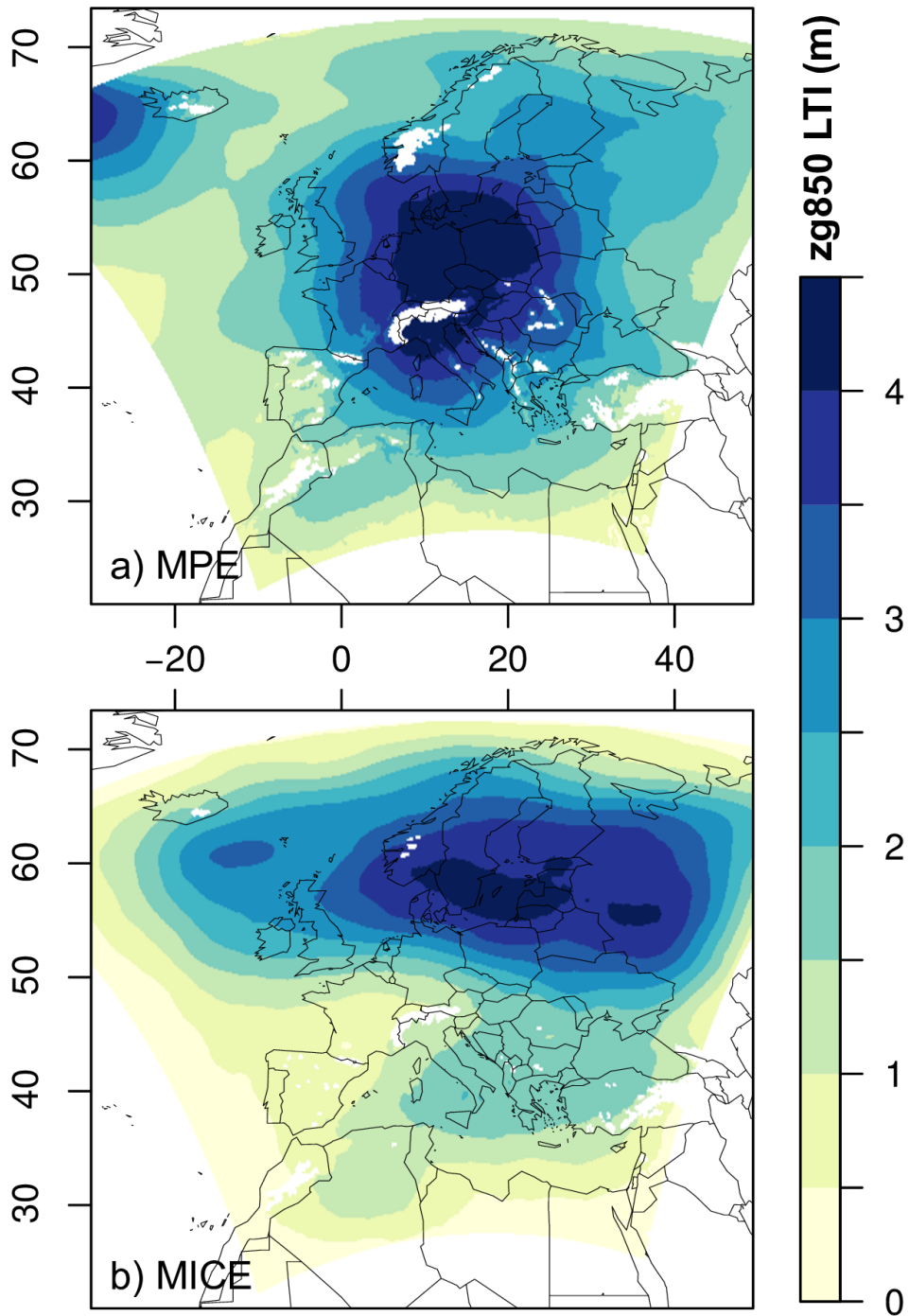


Figure 6: Long-term impact of multi-physics (a) and multi-initial-conditions (b) on 850hPa geopotential height (m).



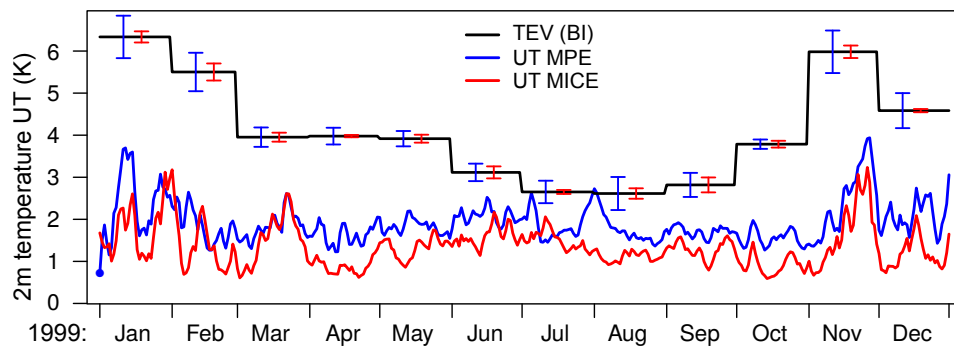


Figure 7: As Fig. 4 but for surface temperature over land.

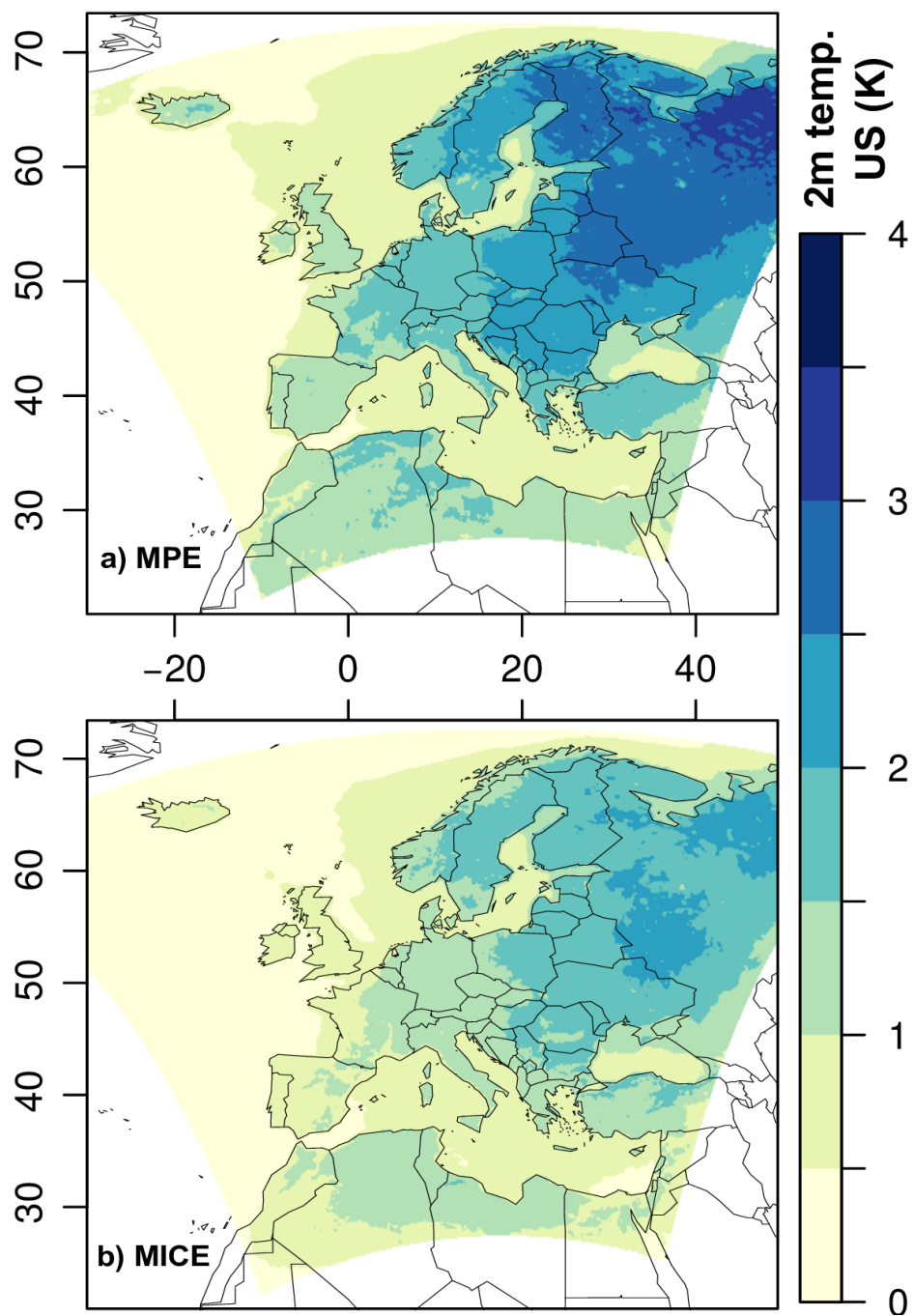


Figure 8: Spatial distribution of the inter-member variance for surface temperature (K) in EUR-15 domain of experiment B (year 1999). a) multi-physics ensemble. b) multi-initial-conditions ensemble.

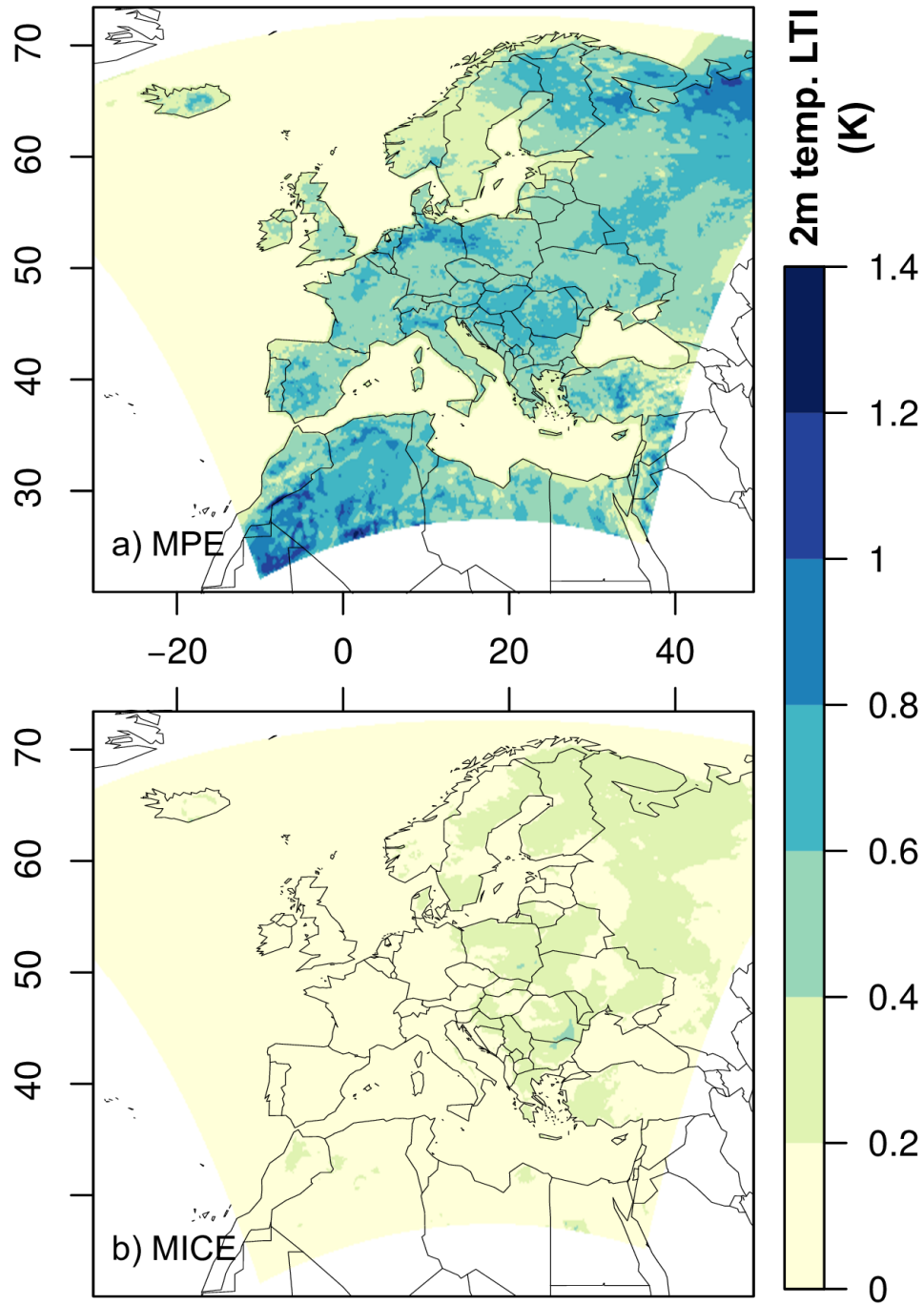


Figure 9: Long-term impact of multi-physics (a) and multi-initial-conditions (b) on surface temperature (K).