Clim Dyn
DOI 10.1007/s00382-012-1539-1

A multi-physics ensemble of present-day climate regional simulations over the Iberian Peninsula

Sonia Jerez · Juan Pedro Montavez · Pedro Jimenez-Guerrero · Juan Jose Gomez-Navarro · Raquel Lorente-Plazas · Eduardo Zorita

Received: 15 March 2012 / Accepted: 20 September 2012

Abstract This work assesses the influence of the model physics in present-day regional climate simulations. It is based on a multi-phyiscs ensemble of 30-year long MM5 hindcasted simulations performed over a complex and climatically heterogeneous domain as the Iberian Peninsula. The ensemble consists of eight members that results from combining different parametrization schemes for modeling the Planetary Boundary Layer, the cumulus and the microphysics processes. The analysis is made at the seasonal time scale and focuses on mean values and interannual variability of temperature and precipitation. The objectives are (1) to evaluate and characterize differences among the simulations attributable to changes in the physical options of the regional model, and (2) to identify the most suitable parametrization schemes and understand the underlying mechanisms causing that some schemes perform better than others. The results confirm the paramount importance of the model physics, showing that the spread among the various simulations is of comparable magnitude to the spread obtained in similar multi-model ensembles. This suggests that most of the spread obtained in multi-model ensembles could be attributable to the different physical configurations employed in the various models. Second, we obtain that no single ensemble member outperforms the others in every situation. Nevertheless, some particular schemes display a better performance. On the one hand, the non-local MRF PBL scheme reduces the cold bias of the simulations throughout the year compared to the local Eta model. The reason is that the former simulates deeper mixing layers. On the other hand, the Grell parametrization scheme for cumulus produces smaller amount of precipitation in the summer season compared to the more complex Kain-Fritsch scheme by reducing the overestimation in the simulated frequency of the convective precipitation events. Consequently, the interannual variability of precipitation (temperature) diminishes (increases), which implies a better agreement with the observations in both cases. Although these features improve in general the accuracy of the simulations, controversial nuances are also highlighted.

Keywords Parameterization schemes · Multi-physics ensemble · Regional climate modeling · Iberian Peninsula

S. Jerez · J. P. Montavez (✉) · P. Jimenez-Guerrero · J. J. Gomez-Navarro · R. Lorente-Plazas
Departmento de Fisica, Universidad de Murcia, Murcia, Spain
e-mail: montavez@um.es

S. Jerez
e-mail: sonia.jerez@gmail.com

P. Jimenez-Guerrero
e-mail: pedro.jimenezguerrero@um.es

J. J. Gomez-Navarro
e-mail: jjgomeznavarro@um.es

R. Lorente-Plazas
e-mail: lorente.plazas@gmail.com

S. Jerez
IDL, Universidade de Lisboa, Lisbon, Portugal

E. Zorita
Institute for Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany
e-mail: eduardo.zorita@hzg.de

1 Introduction

Regional Climate Models (RCMs) have become an extensively used tool in climate research thanks to the fast

Published online: 16 October 2012

growth of computational power, free access to the source codes, and their multiple applications, which include climate change projections (Gao and Giorgi 2008; Koo et al. 2009; Gomez-Navarro et al. 2010), air quality studies (Leung and Gustafson 2005; Forkel and Knoche 2006; Jimenez-Guerrero et al. 2011) and the evaluation of the renewable energy resources (Frank and Landberg 1997; Pan et al. 2004; Pryor et al. 2005) among many others.

The main purpose of RCMs is to obtain climate information at a high resolution by dynamically downscaling coarser databases used as driving conditions at the boundaries of the domains, mainly from General Circulation Models (GCMs). This comes at the expense of limiting the model domain size, i.e. by focusing over a limited area. The higher spatial resolution of RCMs allows a more detailed representation of land-sea contrasts, vegetation cover and topography, as well as a more detailed simulation of 'small-scale' physical processes. Consequently, their use leads to a better representation of mesoscale atmospheric circulations while preserving the large scale features of the driving conditions (Rummukainen 2010).

The reliance placed on RCM simulations is rooted on two very different factors. On the one hand, the need of high resolution, quality and comprehensive climate information over long periods, which, unfortunately, can not be supplied by direct observations or field measurements. On the other hand, the robustness of the RCMs performance, their good skill to reproduce observed climatologies, and their added value in comparison to GCMs simulations, which has been widely reported (e.g. Castro et al. 2005; Gomez-Navarro et al. 2011). However, many aspects cause considerable uncertainty affecting regional simulations still exists. For instance, different RCMs produce different results even when driven by the same boundary conditions. Therefore, sensitivity studies and ensemble approaches are necessary (and commonly conducted) to further improve our understanding of the behavior of regional climate models and ultimately reduce this uncertainty.

Stensrud (2007) states that the most important components of any numerical weather prediction model are the parametrization schemes representing the subgrid-scale physical processes. Indeed, many studies deal with the sensitivity of meteorological forecasts to the physical parametrizations employed within RCMs (e.g. Grubii et al. 2005; Rakesh et al. 2007; Han et al. 2008). However, their role in regional simulations at climatic scales is still poorly assessed. While multi-model ensembles of regional climate simulations have been widely performed and investigated in an attempt to evaluate and overcome intermodel-related uncertainties or discrepancies (e.g. Jacob et al. 2007; Herrera et al. 2010), fewer studies deal with similar multi-physics ensembles aimed at elucidating associated intramodel discrepancies (e.g. Fernandez et al. 2007; Argüeso et al. 2011).

Therefore, the aim of this work is to conduct a comparative numerical modeling study of limited area climate simulations which have been performed with different physical set-ups within the same RCM. We use the mesoscale model MM5 (Grell et al. 1994), whose performance for reproducing local circulations has been profusely tested (Kanamitsu et al. 2002; Leung et al. 2003; Solman et al. 2008; Gomez-Navarro et al. 2011). In addition, MM5 has a large spectrum of physics options allowing a single-model multi-physics ensemble aimed at identifying the effects produced by changes in the physical set-up of the model, as pursued here. In particular, we focus on the simulated climatologies of temperature and precipitation, obtained by using different physical configurations of the same RCM, over the Iberian Peninsula (IP).

It should be stressed that the performance of a given parametrization scheme is usually dependent on the area, variable and application of interest. However, whereas parametrization schemes are developed and tested over (for) concrete areas (applications), they are later used for other areas or other applications on the basis of the evidences in the previous cases. In this sense, where and how we focus and drive our study is not a minus point and the results presented here may help in the design of new modeling systems focused on the IP climate.

The IP is located at the Northern Hemisphere mid-latitudes, surrounded by the Atlantic ocean to the west and by the Mediterranean Sea to the east. It has a complex orography. Though it is a small region, it exhibits an heterogeneous climate ranging from the Mediterranean climate, characterized by warm and dry summers with convective-predominant precipitation and cold and humid winters with large-scale induced precipitation, to milder winters and wetter summers toward the north and west (Font-Tullot 2000). These characteristics pose a strict test for climate models, making, therefore, the IP a suitable scenario for this kind of sensitivity studies, even beyond the particular interest as a single-case study.

This study is closely preceded by the work of Fernandez et al. (2007), who already dealt with a multi-physics ensemble of 5-year long MM5 regional simulations focusing on the reproduction of annual cycles of temperature and precipitation over the IP. The results showed a great seasonal and regional dependence of the most suitable physical set-ups of the model, with no single parametrization combination being optimal for all cases. Here we (1) extend the analysis to the evaluation of not only annual cycles but also interannual variability of the temperature and precipitation series, taking advantage of the length of our simulations that span three decades of the recent past, (2) propose an objective methodology to identify the most influential parametrization schemes and thereby the best (in case) and, from there, the best (or most advisable) physical

configuration of the model for reproducing the IP climate, and (3) have a closer look at the underlying mechanisms that make some of the parametrization schemes included in this study more suitable than others. This may provide useful information for improving the assumptions made in the formulation of different parametrization schemes (Kain 2004). Furthermore, we have employed a more sophisticated land-surface model than the one used by Fernandez et al. (2007) that largely reduces some of the biases detected previously (Jerez et al. 2010) which may masked interesting features.

This paper is structured as follows. Section 2 describes the simulations and the parametrization schemes included in the multi-physics ensemble. Section 3 describes the methodology and the observational database used as a reference for the validation of the simulated fields. Sections 4, 5, 6 and 7 present the results. Discussion and conclusions are provided in Sect. 8.

2 Experimental design

An ensemble of hindcasted simulations (HNDC) spanning the period 1970–1999 has been carried out using a climate version of the mesoscale model MM5 (Grell et al. 1994) which has been already used in other studies (e.g. Gomez-Navarro et al. 2010; Jerez et al. 2010, 2012; Gomez-Navarro et al. 2011; Rodrigo et al. 2012). The multi-physics ensemble consists of eight members that result from combining two planetary boundary layer (PBL) schemes, two cumulus (CML) schemes and two microphysics (MIC) schemes (Table 1). The processes represented by these schemes and how they work are described in detail below. For each case, the two schemes were chosen following two criteria: (1) both should have proven skill and should be, therefore, commonly used, and (2) both should follow, however, very different physical approaches. The remainder of the physical set-up of MM5 is common in all the experiments and includes the RRTM longwave radiation scheme (Mlawer et al. 1997) and the Noah Land-Surface Model (Chen and Dudhia 2001).

Table 1 Combination of the PBL, CML and MIC schemes employed in each simulation of the multi-physics ensemble

Sim.	PBL	CML	MIC
1	Eta	GR	SI
2	MRF	GR	SI
3	Eta	KF	SI
4	MRF	KF	SI
5	Eta	GR	MP
6	MRF	GR	MP
7	Eta	KF	MP
8	MRF	KF	MP

The spatial configuration of the model consists of two two-way nested domains with resolutions of 90 km in the outer domain (D1) and 30 km in the inner domain (D2) (Fig. 1). This latter domain covers the whole IP, even after removing the blending area (five cells from the borders). The outer domain is extended to the east in order to capture the strong influence that the Mediterranean Sea exerts over the IP (Font-Tullot 2000). Vertically, 24 sigma-levels up to 100 hPa, unevenly spaced (more closely spaced near the surface), were included.

The initial and boundary conditions (updated every 6 h) were obtained from the ERA40 reanalysis (Uppala et al. 2005). It should be acknowledged that, since RCMs are limited to the quality of the driving conditions, the evaluation of the ability of a RCM to reproduce observed climates must be carried out by providing reanalysis data, as a surrogate of reality, to the model. Otherwise, it would be too difficult to discern whether the obtained biases are mainly consequence of weaknesses in the RCM performance or in the boundary conditions. The simulations were performed by splitting the whole period (30 years) into subperiods of 5-years length that were then integrated by continuous runs with a spin-up period of 4 months in order to prevent noisy outputs during the model stabilization (Giorgi and Bi 2000) and errors from a possible poor initialization of, specially, soil variables (Christensen 1999).

Since all simulations were identically performed and their set-ups only differ in the physical configuration of the model, differences among them are directly attributable to the choice of the physical parametrizations. However, as we focus only on the inner domain, and given that the physical configuration of the model is the same for both domains and thus different climatologies are developed in the outer domain in the various simulations, it should be acknowledged that, actually, the boundary conditions for the inner domain do also differ among the various simulations. Differences in the inner domain are therefore due to both: (1) the different physical parametrization in the inner domain, and (2) the different physical parametrization in the outer domain leading to different conditions for the nesting with the inner domain. This latter may amplify the differences observed in the inner domain, while preserving the very nature of them. On the other hand, it should also be noted that the domains configuration could not be optimal for the driven conditions used as the ERA40 resolution ($1.125° \times 1.125°$) is not so far from the D1 resolution. However, for climate change experiments (driven by coarser GCM runs), as we will be assessing in a future work, this configuration is quite appropriate as it is coherent with the typical GCM resolutions and optimizes the computational cost of the simulations in comparison

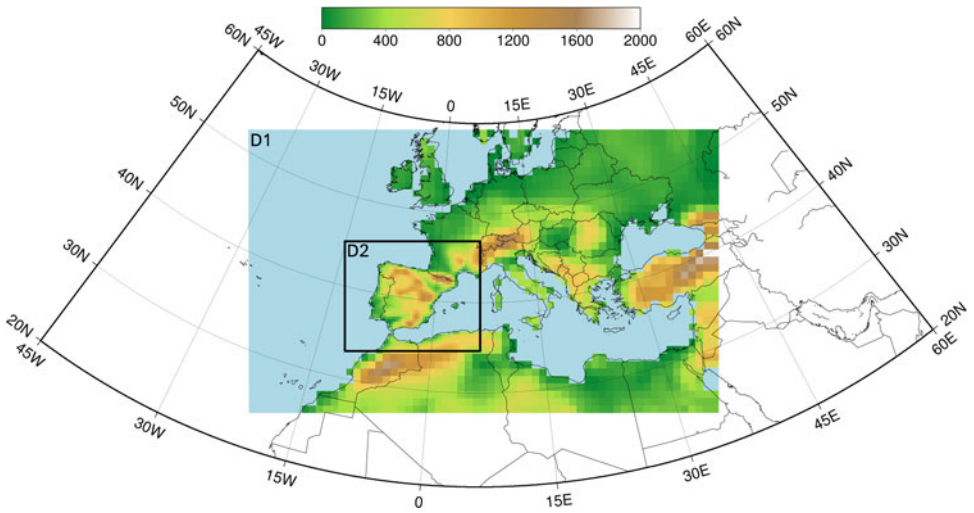

Fig. 1 Domains configuration used in the MM5 simulations. *Shaded colors* depict the orography (height above the sea level, in m) seen by the model at the corresponding spatial resolutions: 30 km in D2, and 90 km in D1. Water masses (sea and ocean) are plotted in *light blue*

with the use of only one but bigger domain at the pursued resolution (i.e. 30 km in our case). These experiments consist of present and future period simulations and, as a first step, it is a mandatory to validate the skill of the model under the GCM-simulated present conditions. For accomplishing this task, we will follow the common procedure of contrasting them with analogous hindcasted simulations (Galos et al. 2007; Gomez-Navarro et al. 2011), that are the ones presented here. This comparison would lack of fairness if the simulations compared were performed with different model set-ups. Therefore, we employed this domains configuration also in the hindcasted simulations presented here.

2.1 PBL schemes

The evolution of the PBL is primarily governed by surface heat, moisture and momentum fluxes. The PBL schemes handle the latent and sensible heat fluxes into the atmosphere, the frictional effects with the surface and the strong subgrid-scale mixing which takes place in the lower levels because of these processes. Once the wind, moisture and temperature are known at a point within the PBL, simple bulk PBL models (i.e. half-order closure models) diagnose their values at all heights without including effects of processes internal to the PBL, such as turbulent eddies. The more complex first-order closure schemes already include turbulent fluxes that flow downgradient. If these fluxes are proportional to the local vertical gradient of the quantity being transferred, the scheme is called "local"; if they are a function of predicted quantities at several heights through the depth of the PBL, the scheme is called "non-local". Non-local schemes treat turbulence as a superposition of eddies of various sizes. Local schemes have been extended to higher order closure schemes. These higher order closure schemes, in addition to prognostic equations for the mean quantities, retain prognostic equations for turbulent fluxes. For further details, the reader is referred to Stensrud (2007).

In this work we have used the MRF and Eta PBL schemes. MRF (Hong and Pan 1996) is a first-order non-local scheme, in which the countergradient transport of temperature and moisture under unstable conditions are added to local gradient transport. Under stable conditions, the local approach is employed for all prognostic variables. Eta (Janjic 1994) is a 1.5-order local closure scheme that computes eddy diffusivity based on local vertical wind shear, static stability, turbulence length scale, and turbulent kinetic energy, which is also predicted by a prognostic equation.

Based on MM5 simulations, some authors showed that the non-local first-order closure PBL schemes that take into account large eddies, such as the MRF scheme, lead to warmer, dryer, better mixed and higher boundary layers, which are closer to observations than those simulated by the more sophisticated 1.5-order local closure schemes available, such as the Eta model, which tend to be too cool, moist and shallow over mid-latitudes semi-arid regions (de Arellano et al. 2001; Bright and Mullen 2002; Zhang and Zheng 2004; Berg and Zhong 2005; Han et al. 2008).

2.2 CML schemes

Cumulus parametrization schemes determine how convection is triggered and how convection modifies moisture and temperature in the atmospheric column and interacts with grid-scale dynamics using the grid-scale information of the host model. This is, CML parametrizations account

for unresolved cloud formation. Therefore, they play an important role in influencing the dynamic and thermodynamic state of the atmosphere. Basic features that differentiate CML schemes are the closure assumptions upon which they are based, and the trigger mechanisms that activate them.

The Kain-Fritsch (KF) and the Grell (GR) schemes are used in this work. KF (Kain and Fritsch 1990) explicitly represents the effects of moist updrafts and downdrafts and the detraiment and subsequent evaporation/sublimation of cloud condensate into the downdraft. Convection is initiated when there is net column instability and sufficient grid-resolved upward vertical velocity to overcome any negative buoyancy in the lower atmospheric layers. The CAPE on the resolved scale governs the quantity of the convective mass flux required to consume the grid-resolved CAPE over a convective time step. The closure assumption is that convection is proportional to the available buoyant energy. GR (Kain and Fritsch 1990) is a simpler one-dimensional mass flux scheme of a single updraft-downdraft couplet. Unlike in the KF scheme, there is no direct mixing between the updraft and downdraft or with the surrounding atmosphere, except at the top and bottom of the clouds. Thus, the convective mass fluxes are constant with height. Closure is achieved by assuming that the rate at which instability is produced at the resolved grid scale is equal to the rate at which instability is removed at the subgrid cloud scale. The scheme removes all the available buoyant energy immediately, and converts all liquid water into precipitation, leaving no portion for cloud water. The GR trigger function depends on the rate of destabilization based on the change in the available buoyant energy due to large scale or subgrid effects. In contrast to KF, GR does not need a rising motion to be activated, and a weak advection of wet air, for instance, is able to trigger it. Nevertheless, while KF remains active until the complete CAPE is removed, GR checks for its activation at every time step. Thus, once KF is activated, it may lead to longer-lasting clouds and more moist convection.

Within MM5, Wang and Seaman (1997); Ferretti et al. (2000); Gochis et al. (2002); Yang and Tung (2003); Liang et al. (2004); Mapes et al. (2004); Ratnam and Kumar (2005) agree that schemes with closure assumption based on convective available potential energy (CAPE), such as the KF scheme, perform slightly more consistently, while simpler cumulus parametrizations, such as the GR scheme, tend to underestimate moisture, temperature and convective precipitation. Nevertheless, they also point out that the latter shows a higher accuracy in specific cases, mainly during light rainfall events. Moreover, other authors report the better performance of the GR scheme for most of the cases assessed (Kerkhoven et al. 2006; Rakesh et al. 2007). Therefore, it is still unclear if any single cumulus parametrization scheme consistently outperforms all others.

2.3 MIC schemes

Cloud formation is accomplished primarily by upward vertical air movement in cloud-free regions, leading to patches of air that have relative humidity in excess of 100 %. Once the relative humidity is above 100 %, cloud droplets can form, producing clouds. The MIC parametrization scheme accounts for the microphysical processes that govern the formation, growth and dissipation of cloud particles, i.e. water condensation, freezing, sublimation, evaporation, melting and deposition. Many aspects of these processes are still not completely understood. However, it is well known that they play an important role in how moist convection develops and evolves, as well as in the radiative energy budget of the earth-atmosphere system through both their albedo and greenhouse effects.

The schemes selected in this work are the Simple Ice model (SI) and the Mixed Phase scheme (MP). SI (Dudhia 1989) does not model mixed phase processes and does not allow for supercooled water. Snow and cloud ice are assumed to melt immediately at the melting point (above 0 °C), and vice versa, liquid water or vapor water condenses or sublimates immediately below that freezing threshold. The prognostic variables are the mass content of precipitation water and the cloud water. The more complex MP scheme (Reisner et al. 1998) does allow for supercooled water in liquid phase below 0 °C and ice does not immediately melt above 0 °C. It has mass contents for liquid water cloud, ice cloud, rain, snow, graupel and ice number concentrations as prognostic variables.

Previous works showed that the scheme for rainwater and cloud water with simple ice, with no mixed-phase processes, tends to simulate more snow at the expense of rainfall compared with more complex schemes, such as the MP scheme, which seem to perform slightly better (Kotroni and Lagouvardos 2001; Grubii et al. 2005). However, the inclusion of supercooled water by the latter reduces sometimes the amount of precipitation, leading to an overall worse performance (Colle and Mass 2000). In general, the representation of cloud microphysical processes in MM5 simulations is more controversial (Chiriaco et al. 2006) but less influential (McFarquhar et al. 2006) than other parametrized processes.

3 Methodology

The analysis focuses on seasonal averages of 2-m temperature (T) and precipitation (P). Both mean values and interannual variability are assessed, the latter defined as the standard deviation (sdev) of the detrended series. Three typical skill scores are used to quantify the ability of the simulations to reproduce the observed climatological

patterns of these magnitudes: the spatial correlation index (r), the Mean Absolute Error (*MAE*) and the Mean Bias Error (*MBE*). r allows to evaluate the spatial distribution of the simulated magnitudes independently of biases, *MAE* gives an unambiguous measure of the average error, and *MBE* provides information on the average error sign and, together with *MAE*, on the homogeneity of the error sign across the domain (Willmott and Matsuura 2005). In addition, the temporal correlation between the simulated and the observed seasonal series (ρ) is also explored.

The observational E-OBS database (Haylock et al. 2008) is used as a reference to evaluate the accuracy of the simulations in order to (1) elucidate deficiencies and (2) establish possible better configurations or more accurate schemes. E-OBS is a reconstruction of the evolution of the near surface air temperature and precipitation for the recent past (spanning from 1950 to 2006). It is the result of an interpolation of observational data to a high resolution regular grid (0.25×0.25) that homogeneously covers Europe over land grid points. E-OBS was initiated by the European Climate Support Network (ECSN) and supported by the Network of European Meteorological Services (EUMETNET). It was originally developed as part of the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com), and is now maintained and elaborated as part of the EURO4M project (EU-FP7). We have used the third version of this data set. Although some problems regarding precipitation have been reported (Hofstra et al. 2009), we use this database because it is commonly used for model validation purposes in large projects such as ENSEMBLES (van der Linden and Mitchell 2009), and the daily temporal resolution of this data base allows the study of extreme events, which will be assessed in future studies complementing this work.

For the comparison of the simulations and the E-OBS database, a spatial interpolation is performed from the MM5 grid onto the observational E-OBS grid (therefore, the evaluation is constrained to land points). Such interpolation consists of distance-weighted means involving those points of the MM5 grid that fall inside a circle centered at every point of the observational grid with a radius equal to twice the minimum distance between the corresponding E-OBS grid point and the closest MM5 grid point.

In order to characterize the mean model skill and the intramodel discrepancies, we use the following estimates:

- The ensemble mean (*EM*), defined as:

$$EM = \frac{1}{N}\sum_{i=1}^{N} m_i \tag{1}$$

where m_i denotes a given magnitude (Tmean, Tsdev, Pmean or Pdsdev in our case) as simulated by the i-th ensemble member, and N denotes the number of ensemble members (8 in our case).

- The ensemble spread (*ES*), defined as the maximum difference (in absolute value) in a given magnitude between any pair of simulations of the ensemble:

$$ES = max\{|m_i - m_j|\} \quad \forall i,j \quad i,j = 1,\ldots,N \tag{2}$$

In order to asses the relative importance of the *ES*, the signal-to-noise ratio, defined as the ratio between the *ES* and the standard deviation of the *EM* series, will be used along this work. Higher values of this ratio indicate that it is more unlikely that the internal variability could mask the influence of the physical configuration of the model. In the case of mean values, we impose a somehow subjective threshold of one (which means that the *ES* exceeds one standard deviation of the *EM* series). In the case of the standard deviation of the series (sdev), this signal-to-noise ratio actually depicts the *ES* in percentage with respect to the *EM* sdev values and, as such, it will be considered.

Finally, we propose a methodology aimed at isolating the effect of changing a particular parametrization scheme and, thereby, identifying the most accurate parametrization schemes for the case studies. This methodology is based on the analysis of subgroups of simulations within the ensemble (called subensembles), which are obtained by considering the experiments sharing the same PBL or CML or MIC scheme. Thus, these subensembles consist of four members. For example, the Eta-subensemble is composed of the four simulations performed with the Eta PBL scheme (i.e. the simulations 1, 3, 5 and 7 in Table 1). The subensemble mean is analogous to the ensemble mean but includes just these four members. Then, the mean ability of the model when using the Eta PBL scheme is assessed through the Eta-subensemble mean (note that, in this way, it is expected that the contribution of the interaction between the Eta PBL scheme and the various CML and MIC schemes was filtered out by averaging). The differences between the mean performances of the Eta-subensemble and the MRF-subensemble will display an amount that can be interpreted as the 'spread' attributable to the change of the PBL scheme. Hence, it is here denoted as the PBL_{spread} (Eq. 3). Identical procedures are applied for the CML and MIC schemes.

$$PBL_{spread} = \left| \frac{1}{4}\sum_{i=2,4,6,8} m_i - \frac{1}{4}\sum_{j=1,3,5,7} m_j \right| \tag{3}$$

We define the mean ensemble spread (*MES*) as the sum of the PBL_{spread}, the CML_{spread} and the MIC_{spread} (Eq. 4):

$$MES = PBL_{spread} + CML_{spread} + MIC_{spread} \tag{4}$$

This provides a framework for evaluating the relative importance of changing either the PBL, the CML or the

MIC scheme, since the contribution of the PBL_{spread}, following with the former example, to the *MES* can be expressed as a percentage (through Eq. 5) and compared to contributions of the CML_{spread} and the MIC_{spread} computed in the same way. In this context, we will say that the PBL is the leading parametrized process (LP) if the PBL_{spread} has the greatest contribution to the *MES*.

$$\frac{PBL_{spread}}{MES} \times 100 \quad (5)$$

4 Ensemble mean and ensemble spread

In this section we evaluate the mean ability of the ensemble to reproduce the observed climatology by focusing on the *EM*, as well as how this mean skill varies across the ensemble members by focusing on the *ES*. The analysis is based on Figs. 2, 4 depicting the *EM* values and its bias errors with respect to the E-OBS database, and Figs. 3, 5, which depict the *ES* for all the variables/statistics. Table 2 summarizes the skill scores of the ensemble mean (*EM*), as well as the magnitude of their variations among the ensemble members (i.e. the *ES* in these skill scores). The spread in *MBE* can be directly compared to the intermodel spread (called *MS*) obtained in a multi-model ensemble of similar hindcasts reported in Jacob et al. (2007).

4.1 Temperature

4.1.1 Mean values

The model is able to accurately reproduce the spatial distribution of the mean temperature (Tmean) patterns, being *r* around 0.95 without much variations between the ensemble members (Table 2). Nevertheless, large errors and spreads appear regarding other estimators. Tmean is overall underestimated (Fig. 2, first row), with the largest biases appearing in the warmest south-western areas in winter (about −2 °C) while in the north-east in summer (up to −3 °C); being intermediate in spring and autumn. In spatial average, *MAE* (*MBE*) is around or slightly above 1 (−1) (Table 2). In spite of the large spread in Tmean within the multi-physics ensemble (Fig. 3, first row), particularly large in winter and summer (up to 3 °C) and well above one standard deviation of the *EM* series (i.e. the signal-to-noise ratio is amply above the unit in most of the cases), the observations remain in most of the cases below the simulated range (see dots in Fig. 3, first row). Such a large spread involves variations over 2 °C in the *MBE*.

Beyond details, there are two features worth stressing. On the one hand, that the *ES* in the *MBE* in the multi-physics single-model ensemble is similar to the multi-model spread *MS* (Table 2). On the other hand, that it is appreciable some similarity in the spatial distribution of both the bias patterns and the *ES* patterns. These two facts would indicate that an appropriate physical configuration of the models could largely reduce biases and still improve the spatial distribution of the simulated patterns.

4.1.2 Interannual variability

The ability of the simulations to reproduce the spatial distribution of the Tsdev patterns is generally poor (*r*, between the *EM* patterns and the observations, is below 0.5 in winter and spring and just slightly above 0.6 in summer and autumn) and largely experiment-dependent, as the *ES* in *r* is over 0.3 (Table 2). Spatially, errors are as follows. Where the observations show the lowest (highest) values of Tsdev, which is in winter (spring), the *EM* mostly overestimates (underestimates) it, up to 0.4 (0.5) °C, which represents around the 50 % of the observed values. In summer and autumn errors are smaller but still important, being Tsdev underestimated up to 0.2–0.3 °C (up to 30 %). Again, a large spread (around 20 % of the *EM* values except for summer, when it grows above 60 %) appears over some of the areas showing the largest errors. Thus, although nothing similar to a linear relationship between bias and spread can be established, the particularly important role of the physical configuration of the model over the areas worst represented can be anew recognized. Again, variations in the *MBE* are similar to the *MS* values (Table 2).

In spite of the errors in the simulated magnitude of the variability of the temperature series, there is a fairly good agreement between the temporal evolution of the simulated and the observed temperature series, with ρ being above 0.8 almost everywhere (Fig. 2, third row), and above 0.9 when it is computed for the spatially averaged series of the *EM* (Table 2). Moreover, variations of ρ among the ensemble members are not important but for the summer season, when, however, the *ES* in ρ is up to 0.3–0.4 in the southern half of the IP (Fig. 3, third row).

4.2 Precipitation

4.2.1 Mean values

As for mean temperature, the agreement between simulated and observed patterns of the mean amount of precipitation (Pmean) is quite satisfactory, with *r* around 0.75–0.8 in all seasons. However, a larger *ES* in *r* appears in this case, up to 0.15–0.2 in summer and autumn (Table 2). The main biases in the simulated Pmean are up to 50 mm/month, representing up to 50 % of the observed values, and briefly consist of underestimation (overestimation) in the wettest (driest) areas/seasons, e.g. the western IP in winter, spring

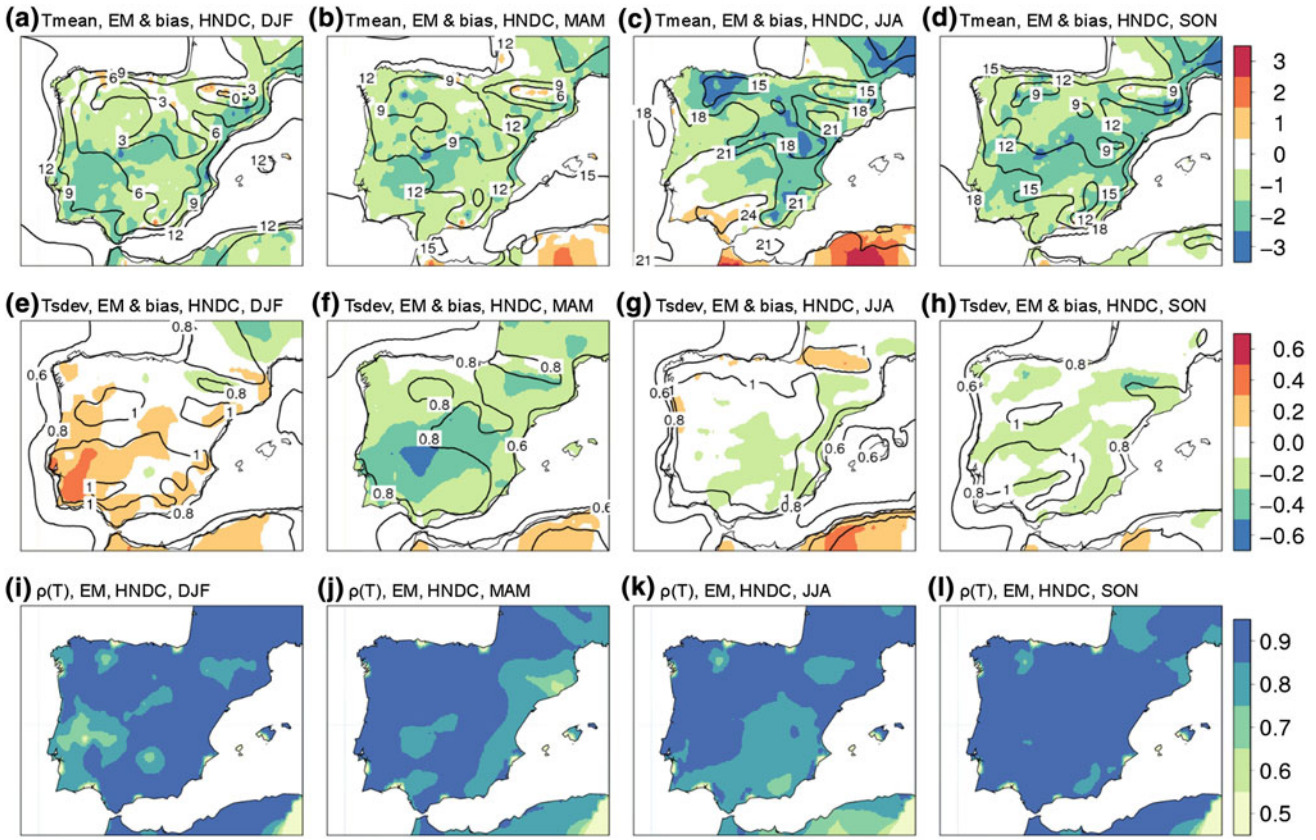

Fig. 2 Mean ability of the ensemble of hindcasts to reproduce the observed climatology: ensemble mean (*EM*, contours) and *EM*-bias (*shaded colors*) for Tmean (*first row*) and Tsdev (*second row*). Units in °C. *Third row* depicts the temporal correlation (ρ) between the simulated series (from the *EM*) and their analogs from the E-OBS database. Each column represents one season: winter (DJF), spring (MAM), summer (JJA) and autumn (SON)

and autumn (the whole IP in summer). It is worth stressing that the overall mixture of positive and negative biases throughout the entire domain leads to notable differences between the absolute values of the *MAE* and the *MBE* (Table 2), highlighting that the analysis of spatial averages of precipitation must be considered with great care in the case of the IP.

The *ES* in Pmean (Fig. 5, first row) is very localized in northern areas in winter, spring and autumn, where it represents about 30 % of the *EM* values, while it is likely negligible in the rest of the IP in comparison to the natural variability of the precipitation series (i.e. the signal-to-noise ratio is below one). However, in summer, the signal-to-noise ratio is clearly larger than one everywhere, with the *ES* depicting an orographic pattern with maxima over the main mountain systems ($\sim$40 mm/month, which is $\sim$100 % of the *EM* values) (Fig. 5c). This orographic distribution of the spread resembles the appearance of the corresponding bias pattern (Fig. 4c), although the observations still remain above the range of simulated values. Albeit the values of the spread in Pmean are large, the *ES* in the *MBE* is in this case about the half than the *MS* values even in summer (Table 2). This feature further highlights the strong dependence of precipitation on the large-scale advective phenomena and on the synoptic configuration (Koster and Suarez 1995), which is actually expected to show a wider spectrum in a multi-model ensemble of simulations than in this multi-physics single-model ensemble of simulations, since the latter are all performed with the same dynamic modeling system (Sanchez-Gomez et al. 2009).

Finally, the large spread in Pmean over the water mass areas (up to 100 % of the *EM* mean precipitation over the Mediterranean Sea) is noteworthy. This feature does not appear in the *ES* patterns of temperature, likely because the sea surface temperature, provided by ERA40, is the same in all the simulations. This boundary condition is not as determinant for precipitation as for temperature, which would explain this asymmetry.

4.2.2 *Interannual variability*

The model skill regarding the spatial distribution of the Psdev patterns shows a marked seasonality, with *r* being

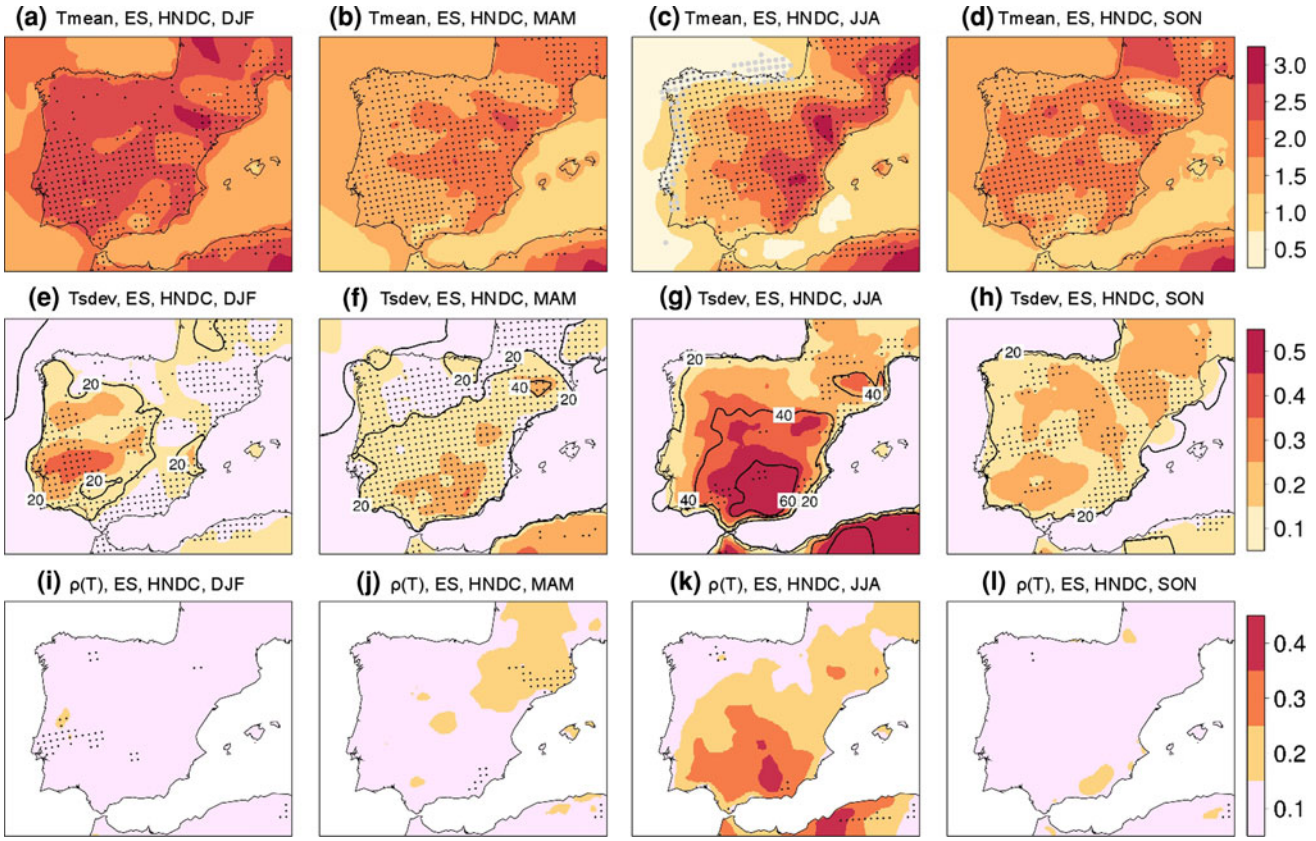

Fig. 3 Ensemble spread (*ES*) in the simulated Tmean (*first row*, units in °C), Tsdev (*second row*, units in °C by *shaded colors*; contours depict the *ES* in percentage with respect to the *EM* values) and in ρ (*third row*). *Black dots* mean that the observed values are out of the range of simulated values, except in the *third row* where they mean that ρ remains below 0.8 in all the simulations. *Gray dots* in the first row mean that the signal-to-noise ratio is below the unit. Each column represents one season: winter (DJF), spring (MAM), summer (JJA) and autumn (SON)

0.8 and showing a small spread in winter, while dropping below 0.5 with increased spread (up to 0.3 in summer) in the rest of seasons (Table 2). The *EM* patterns of errors, which are up to 40 mm/month or up to 50 % of the observed values (Fig. 4, second row), resemble the *EM* patterns of errors in Pmean. Therefore, the precipitation variability is underestimated in the same regions where the precipitation amount is underestimated, and vice versa. Indeed, the *EM* patterns for both mean precipitation and precipitation variability have similar distributions, i.e., the larger the amount of precipitation is, the larger the inter-annual variability of the precipitation series is. Moreover, the *ES* patterns for both magnitudes, Pmean and Psdev, exhibit also similar structures, although in percentage terms the *ES* is larger in Psdev than in Pmean, up to 60 % in every season and even above 100 % in summer. However, the *ES* for the *MBE* index is again just half of the *MS* (Table 2).

Regarding the temporal correlation of the simulated series with observations, lower skill and larger spread are found for precipitation than for temperature. Albeit ρ is over 0.8 if we compare the spatially averaged series of the *EM* with the corresponding observational series, except for the summer season when $\rho = 0.5$ (Table 2), there are wide areas where ρ drops below 0.5 in all seasons (Fig. 4, third row). The largest *ES* (over 0.5) appears in summer in the northern half of the IP, being also quite considerable in spring and autumn (Fig. 5, third row). In general, a large *ES* appears in region where the skill of the *EM* is poor, which again suggests that large improvements may be achieved with a proper physical configuration of the regional model.

5 A single best physical configuration?

The large spread obtained (often over the areas where simulations and observations do not show a good agreement) highlights the major role played by the physical configuration of the regional model for accurately simulating temperature and precipitation over the IP. The question immediately arising is whether a single ensemble member consistently outperforms the others. In this section

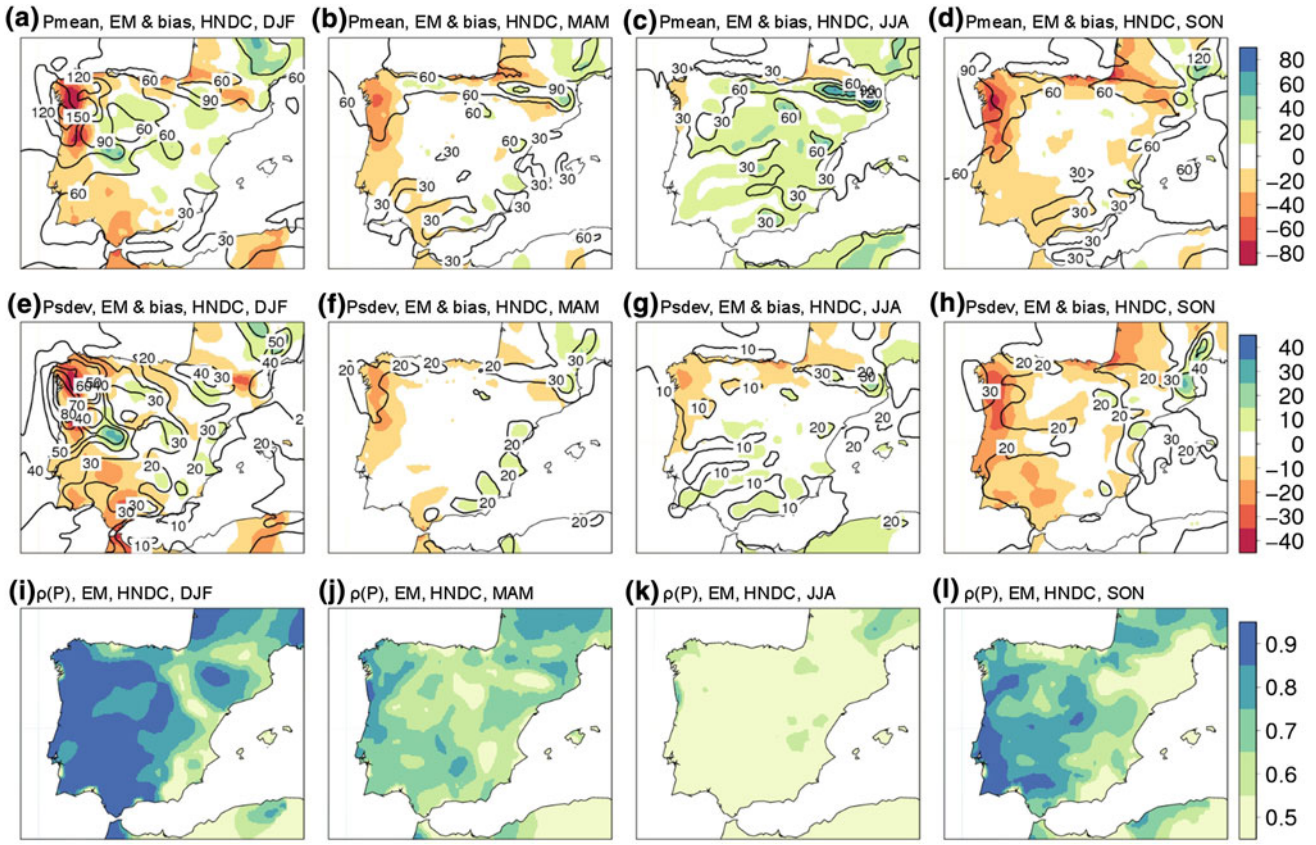

Fig. 4 As Fig. 2 for precipitation. Units in mm/month instead of °C

we perform a very simple analysis aimed at detecting the most accurate ensemble member for each site and season, and for each variable and statistic. Figures 6, 7 depict which ensemble member is closest to the observations in terms of the lowest error or highest ρ in each case (from now on we will call it "the best member" of the ensemble) through a color code assigning one color to each simulation of the ensemble (the reader should refer to Table 1 for identifying the simulations). The dots in these plots indicate that the *EM* reproduces the observation better than any single ensemble member.

First, the mixture of colors within any of these patterns reveals that which of the ensemble members is closest to the observations depends on the region. Second, these patterns do not look similar for any of the magnitudes throughout the year, neither for any season throughout the various magnitudes. Therefore, there also exists a dependence on the season and on the magnitude of interest. Hence, this analysis prevents to pin down a best single ensemble member for all the cases. Moreover, the *EM* only stands out regarding the temporal evolution of the temperature series, but in the rest of the cases there is always one experiment that performs better than the *EM*.

This analysis has additionally revealed a worth mentioning feature. If we consider in each case (each site, season, variable and statistic) the value given by the best member and compare it with the observations (thus obtaining the patterns of the minimum errors within the ensemble; not shown), the resulting error patterns have the same spatial distribution as those corresponding to the *EM* shown in Figs. 2, 4, although they are obviously less intense. This indicates that the location of the areas showing the largest disagreement between the simulations and the observations (such as the eastern IP in summer regarding Tmean and the north-western IP in winter regarding Pmean, for instance) are hardly dependent of the physical configuration of the model. However, we found that these areas are largely sensitive to the choice of the parametrization schemes since they display a large model spread. The strong heterogeneity of these regions could be a plausible explanation for both features. On the one hand, a large *ES* over heterogeneous areas is actually expected. On the other hand, in heterogeneous regions, single measurements are not representative of the surrounding areas (i.e. of the area at the resolution employed in the regional model). Therefore, at the model resolution, it can not be expected that the simulations can capture a measured singularity. In our opinion, this does not imply a poor performance of the model but rather an inherent limitation of the simulations due to the spatial resolution employed.

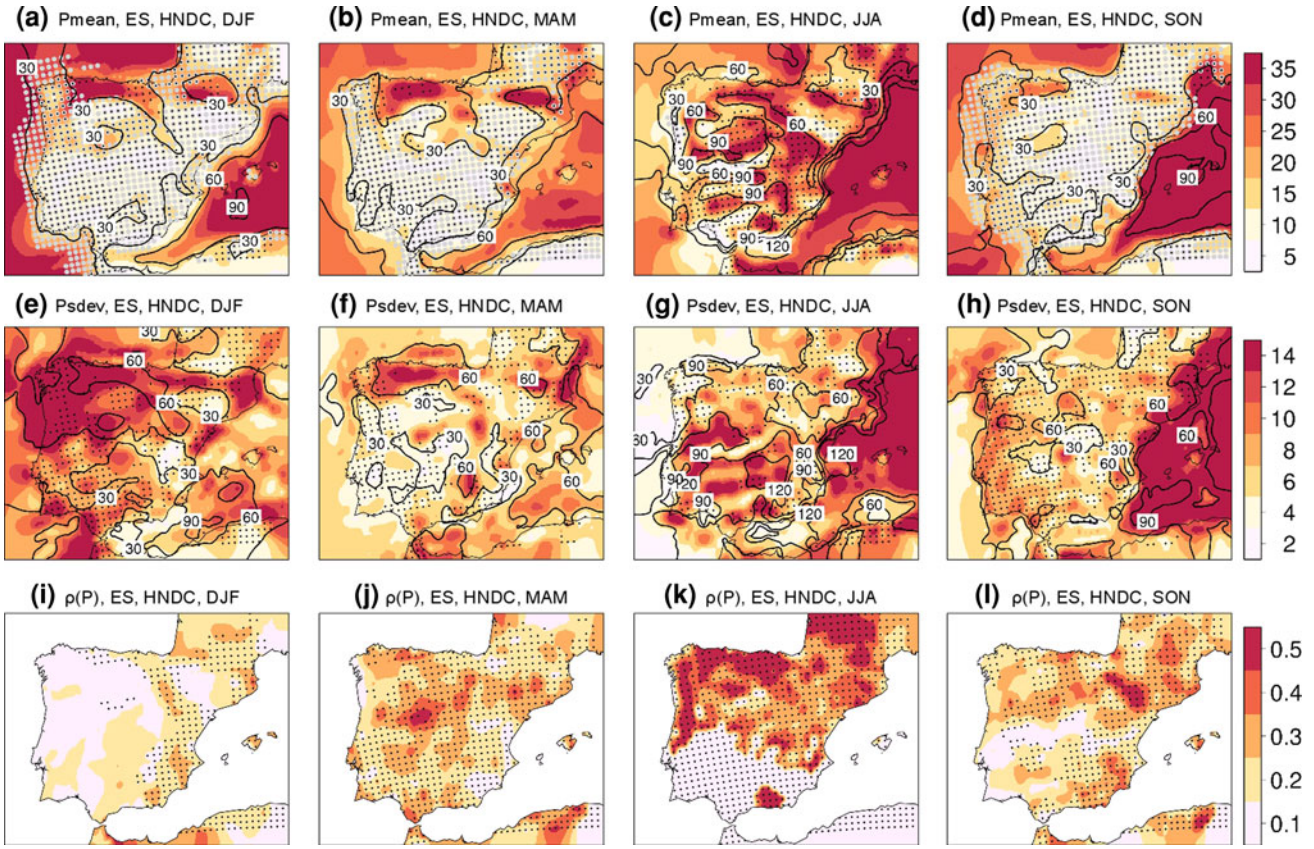

Fig. 5 As Fig. 3 for precipitation. Units in mm/month instead of °C

6 Insights from the subensemble methodology

Based on the simple approach presented before, one should either reject the existence of a single best physical configuration of the model for all the case studies or, better, propose alternative methodologies aimed at providing effective insights regarding the role of the various parametrization schemes. In this section we apply an alternative method focused not on the individual ensemble members but on the performance of the various parametrization schemes (see Sect. 3). First (Sect. 6.1), we identify the most influential parametrized processes. Second (Sect. 6.2), we identify the most accurate parametrization schemes, considering the IP as a whole and taking into account the relative importance of the physical configuration of the model across the domain.

6.1 Most influential parametrization schemes

Figures 8, 9 depict the mean ensemble spread (*MES*, with contours), the leading parametrized processes (LP, by the color) and the contribution to the *MES* of the scheme-induced spread associated to the identified LP (by the intensity of the color shading).

First, it is worth noting that the *MES* patterns resemble the *ES* patterns shown in Figs. 3, 5. This feature further supports our analysis by confirming that the spread obtained in the ensemble is not due to singularities arising from a single experiment, but to the systematic differences between the various experiments that remains even after averaging.

Regarding the influence of each parametrized process, some clear signals can be recognized in Figs. 8, 9, although most of these patterns still look quite heterogeneous. The PBL_{spread} greatly prevails over the CML_{spread} and the MIC_{spread} in the case of Tmean everywhere and every season. On the other hand, the CML_{spread} dominates in the case of Pmean (inland), Psdev and Tsdev in summer. The MIC_{spread} shows the greatest contribution to the *MES* in sparse occasions mainly related to the temperature variability. Do these signals indicate a better performance of one of the two schemes considered in each case?

6.2 Most accurate parametrization schemes

Motivated by the clear signals provided in the above evaluation regarding the 'leading parametrized processes', below we investigate the existence of most accurate/appropriate parametrization schemes for the cases studies. Figure 10 (left) illustrates the percentage of land points within the domain where each scheme performs better than its counterpart, i.e. MRF vs. Eta, GR vs. KF, and SI vs. MP.

Table 2 Skill scores (r, *MAE*, *MBE* and ρ) of the ensemble mean (*EM*) for each variable/statistic and each season

Var.	Stat.	What	DJF			MAM			JJA			SON		
			r	*MAE*	*MBE*	r	*MAE*	*MBE*	r	*MAE*	*MBE*	r	*MAE*	*MBE*
T	Mean	*EM*	0.95	1.11	−0.98	0.95	1.10	−0.95	0.93	1.44	−1.14	0.96	1.35	−1.29
		ES	0.01	1.62	2.23	0.00	1.24	1.64	0.02	0.96	1.23	0.00	1.59	1.84
		MS	NA	NA	2.50	NA	NA	NA	NA	NA	3.25	NA	NA	NA
	Sdev	*EM*	0.32	0.12	0.05	0.37	0.25	−0.23	0.61	0.09	−0.03	0.64	0.10	−0.09
		ES	0.37	0.14	0.12	0.24	0.13	0.16	0.30	0.05	0.26	0.15	0.28	0.17
		MS	NA	NA	0.15	NA	NA	NA	NA	NA	0.36	NA	NA	NA
	ρ	*EM*	0.93			0.91			0.91			0.93		
		ES	0.04			0.13			0.20			0.10		
		MS	NA			NA			NA			NA		
P	Mean	*EM*	0.81	18.8	−7.0	0.76	11.6	−5.6	0.78	15.5	13.0	0.74	15.8	−11.9
		ES	0.04	1.6	8.8	0.07	2.8	9.3	0.15	12.5	15.6	0.18	5.9	4.6
		MS	NA	NA	31.8	NA	NA	NA	NA	NA	31.8	NA	NA	NA
	Sdev	*EM*	0.81	9.5	−5.3	0.41	4.6	−2.3	0.45	4.9	0.3	0.39	8.7	−6.9
		ES	0.06	2.5	7.8	0.15	0.7	2.6	0.28	1.3	1.3	0.13	2.5	2.7
		MS	NA	NA	16.5	NA	NA	NA	NA	NA	2.4	NA	NA	NA
	ρ	*EM*	0.93			0.85			0.50			0.83		
		ES	0.13			0.15			0.43			0.10		
		MS	NA			NA			NA			NA		

r, *MAE* and *MBE* are computed only for the mean and sdev statistics (for ρ it makes no sense). ρ is computed for the spatially averaged *EM* series. *ES* denotes the ensemble spread computed as the difference between the highest and the lowest value of the corresponding skill score obtained across the ensemble members. *MS* denotes the multi-model spread (computed as the *ES* from data provided in Jacob et al. (2007) for the IP when available; *NA* non-available). Units for *MAE* and *MBE* are °C when referred to temperature and mm/month when referred to precipitation

The analysis is based on the subensemble approach, i.e. by comparing the corresponding subensemble means with the observations. Such a percentage does not remain always above the 50 %-line for any of the parametrization options. The darker bars (corresponding to the MRF PBL scheme, the GR CML scheme and the SI MIC scheme in each plot) seem to prevail over the gray ones, but it is difficult to discern any conclusive result. Instead, these plots further highlight that the accuracy of the schemes depends on the variable, statistic, site and season.

However, the aforementioned results do not take into account that there are areas with small spread and areas with large spread, where, thus, which parametrization scheme performs best should be prioritized. So far, we have just counted the number of grid points where each parametrization scheme performs better than the other option considered without pondering how much, i.e. regardless the value of the spread. However, from a holistic perspective, the use of a particular parametrization option could result advantageous, although eventually it is not (i.e. if it is not in the cases exhibiting small differences between the two options). Actually, Fig. 10 (right) reveals a clearer predominance of one schemes over the others when differences between the various subensemble means are considered only if they exceed some subjective thresholds. These are that the signal-to-noise ratio (as described in Sect. 3) computed for the particular scheme-induced spreads must be higher than 1 (0.1) in the case of mean (sdev) values, and that, in the case of the temporal correlations, the difference between the two subensemble means must be larger than 0.1. This procedure filters the cases where the role of the schemes is less relevant focusing on those where the schemes performances differ most. This way, Fig. 10b shows that the MRF PBL scheme overwhelmingly outperforms the Eta model, Fig. 10d shows that the GR CML scheme moderately outperforms the KF model, and Fig. 10f shows that the SI MIC scheme slightly outperforms the MP model.

In order to deepen into the identification of better performances of the individual schemes, we quantify the differences in the skill scores (r, *MAE* and ρ) of the corresponding subensemble means. Table 3 shows the values of the skill scores corresponding to each subensemble mean. Based on these values, we identify improvements of at least 10 % by imposing systematically the following thresholds (where the subscripts b and w refer to the best and the worst value respectively):

- The best spatial correlation (r_b) is over 0.5 and satisfies $r_b - r_w > r_w/10$.
- The lowest mean absolute error (MAE_b) holds that $\frac{MAE_w - MAE_b}{MAE_w} > MAE_w/10$ in the case of Tmean, and

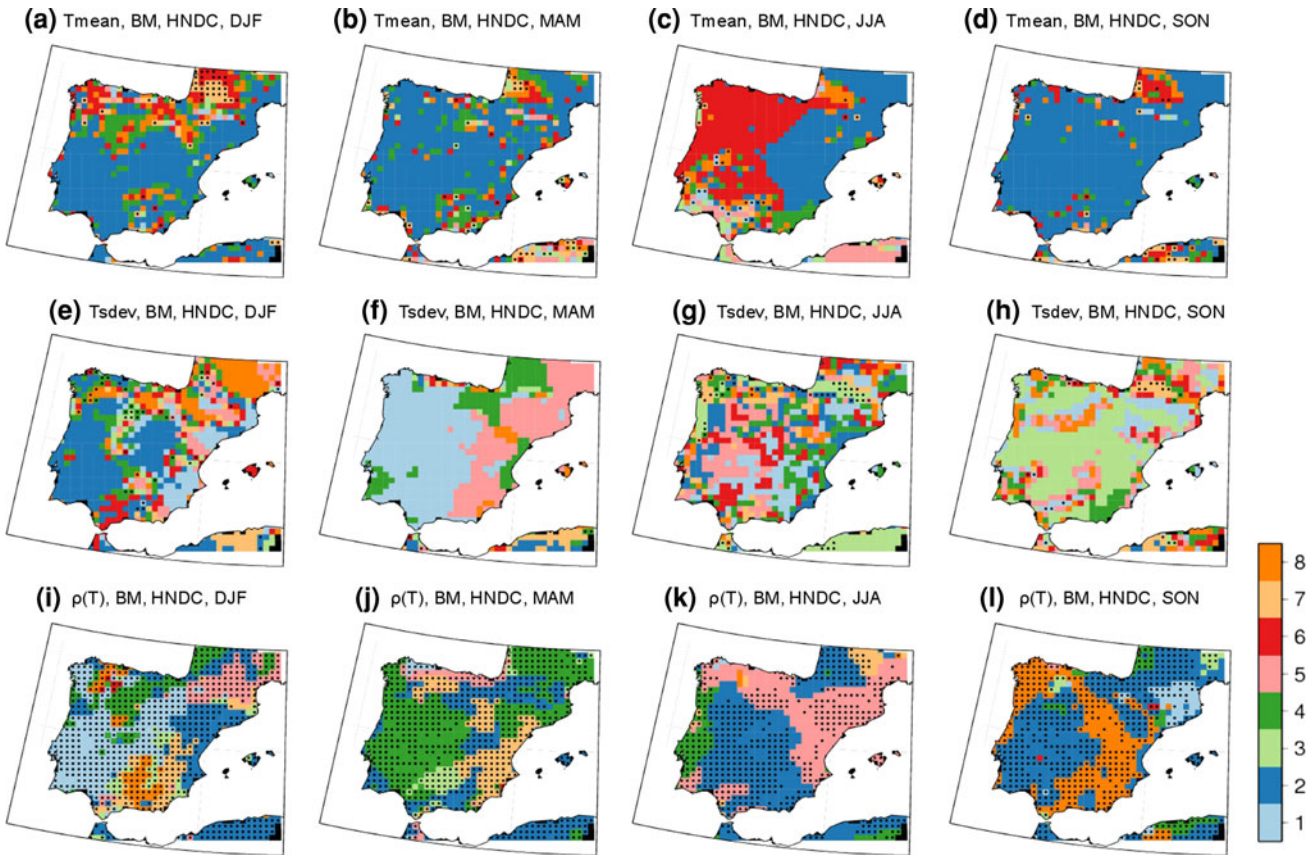

Fig. 6 Best ensemble member in reproducing the observed climatology. *Each color* represents one of the simulations composing the ensemble (see Table 1 for identifying them). Based on this colors code, these plots depict the ensemble member with the smallest error in reproducing Tmean (*first row*) and Tsdev (*second row*), and with the best performance in the analysis of temporal correlations (*third row*). *Dots* mean that the ensemble mean outperforms all the ensemble members

$MAE_w - MAE_b > 10$ if $MAE_b < 50$ %, or $MAE_w - MAE_b > 30$ if $MAE_b > 50$ % (expressing *MAE* in percentage with respect to the observed values at each grid point), in the cases of Tsdev, Pmean and Psdev.

- The best temporal correlation (ρ_b) is over 0.5 and satisfies $\rho_b - \rho_w > \rho_w/10$. Here ρ is computed for the spatially averaged series.

Based on these subjective criteria, this procedure allows us to identify improvements of at least 10 % in some of the skill scores. These improvements are mainly related to the use of the MRF PBL scheme and the GR CML scheme. The former largely reduces biases in the Tmean patterns in every season (*MAE* is reduced around 1 °C ~50 %). The latter improves the representation of Tsdev (*r* rises from 0.54 to 0.64), Pmean (*r* increases from 0.68 to 0.82 and *MAE* is reduced to a half) and Psdev (*r* increases from 0.27 to 0.51 and *MAE* is reduced by 10 %) in the summer season (see bold numbers in Table 3). Regarding the MIC scheme, SI performs slightly better than MP as regards Tmean. Moreover, none of these schemes performs worse in any case (at least, not exceeding the above thresholds). Thus, we have clearly identified three schemes that provide higher confidence when simulating the climatology of the IP as a whole from a multi-variable/statistical point of view. It is worth mentioning that the combination of these three schemes is, in addition, the most efficient computationally.

7 Underlying mechanisms

The determination of the particularities of the schemes provoking the largest differences in the skill of the various simulations is beyond the scope of this study. Nonetheless, in this Section we try to physically explain the different responses. We focus on the performance of the PBL schemes when simulating Tmean, and on the performance of the CML schemes when simulating Tsdev, Pmean and Psdev (since these cases showed the largest signals in the previous assessment).

7.1 Comparison of the PBL schemes

The MRF PBL scheme always provides higher temperatures than the Eta model, which substantially reduces *MAE* and *MBE* over the entire domain and strongly drives the

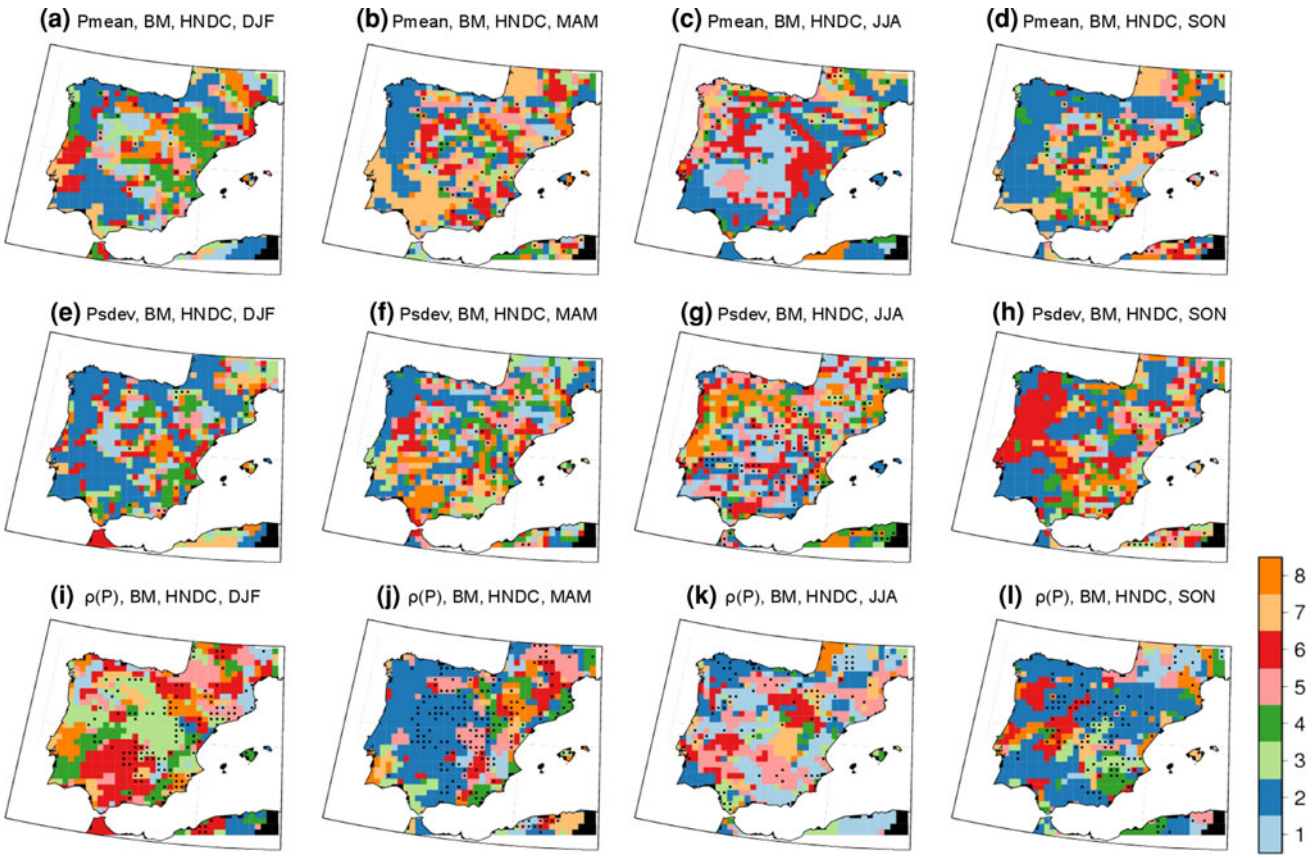

Fig. 7 As Fig. 6 for precipitation

mean ensemble spread patterns for Tmean. This feature must be linked to the fact that the non-local closure approach followed in the MRF scheme generates better mixed PBLs than the local Eta scheme. Figure 11 depicts the vertical profiles of the potential temperature in the lower eight atmospheric half-sigma levels obtained from the eight simulations, averaged over the whole IP. Note that biases at the first level are not included in this Figure in order to facilitate the comparison of the vertical gradients between the various experiments. The experiments performed with the MRF scheme show a sharper vertical gradient (i.e. with higher slope) than the experiments performed with the Eta scheme. Thus, the MRF scheme leads to develop deeper boundary layers. This well-known behavior (Wang and Seaman 1997; Ratnam and Kumar 2005) increases the potential temperature near the surface by reducing differences between the upper (warmer) and the lower (colder) levels.

Indeed, the use of the MRF PBL scheme enhances both maximum and minimum temperatures in comparison to the Eta PBL scheme (contours in Fig. 12, first and second rows). However, this is not an improvement in both cases. Meanwhile maximum temperature is systematically underestimated and, therefore, the use of the MRF PBL scheme improves the skill of the simulations reducing the *MBE* up to 50 % in comparison with the results for the Eta PBL scheme (Fig. 12a–d), minimum temperature is, on the contrary, systematically overestimated. In this latter case, the use of the Eta PBL scheme reduces *MBE* similarly up to 50 % in comparison with the results for the MRF PBL scheme (Fig. 12e–h). Therefore, while MRF is better at reproducing maximum temperatures, the local approach of the Eta PBL scheme is likely better for simulating the minima.

Nonetheless, it should be stressed that the above finding is mainly due to systematic differences among the two subensembles, i.e. one scheme is systematically warmer than the other. Note that the contours in Fig. 12a–d roughly resemble the contours in Fig. 12e–h. Actually, the analysis of the mean daily temperature range (DTR) depicted in Fig. 12i–l reveals no notable differences between both subensemble means (just eventually up to 1 °C). DTR is similarly underestimated throughout the year using either the MRF or the Eta PBL scheme (*MBE* is about 4−5 °C, negative, in all cases). This result suggests that both PBL schemes should be considered equally valid if the purpose is to evaluate deviations after removing the mean systematic warm or cold biases.

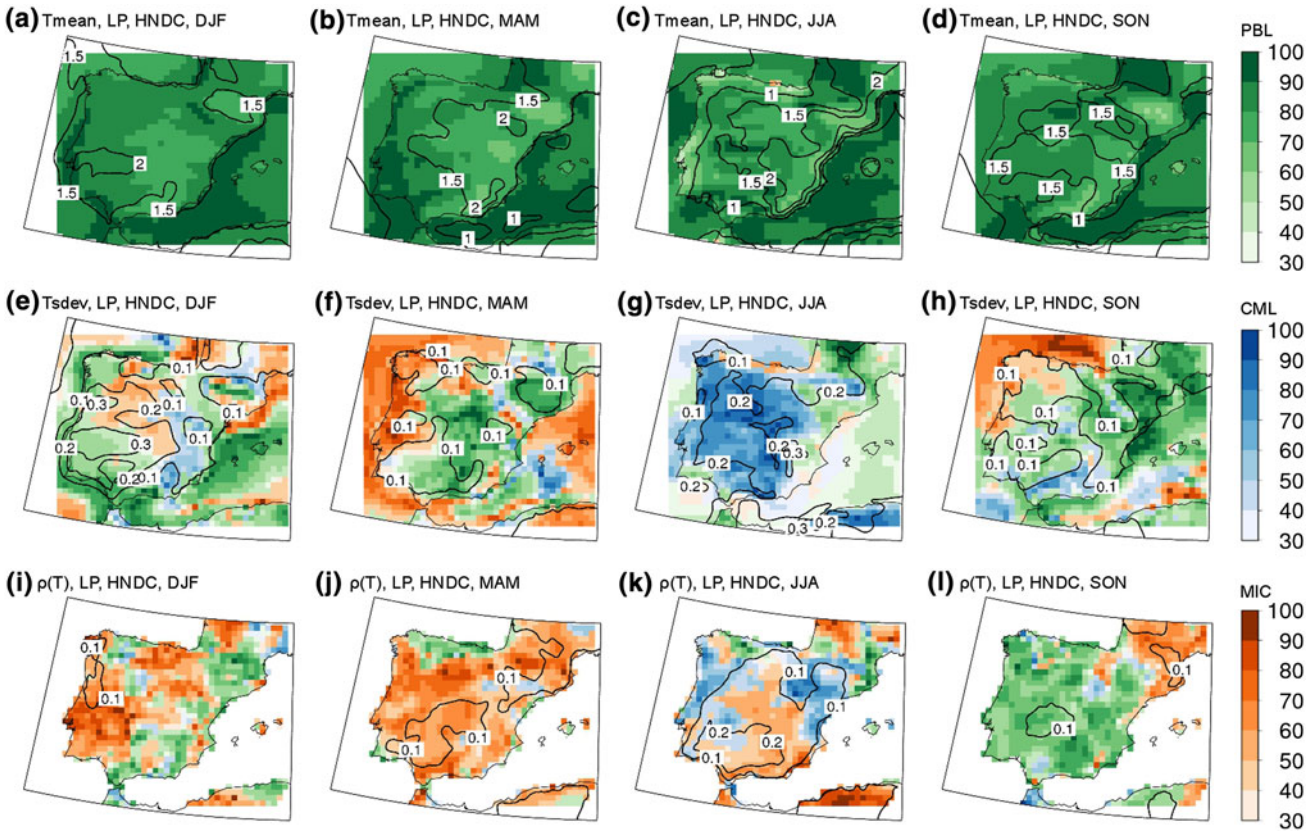

Fig. 8 Most influential parametrized processes for the simulation of Tmean (*first row*), Tsdev (*second row*) and the temporal evolution of the seasonal series of T, i.e. ρ (*third row*). The color (*green*, *blue* or *orange*) indicates which is the leading parametrized process (LP). The intensity of the color represents the percentage of contribution to the *MES* from either the PBL_{spread}, the CML_{spread} or the MIC_{spread} (the one being the LP). Contours depict the mean ensemble spread (*MES*), in °C in the *first* and *second rows*

7.2 Comparison of the CML schemes

The GR CML scheme simulates, in the summer season, less precipitation, lower precipitation variability and larger temperature variability than the KF scheme (Fig. 13a, d ,g). In all these cases, the differences between the performance of both CML schemes strongly contribute to the ensemble spread and render the GR scheme more suitable than the KF scheme, as shown above. In this section we delve into these differences.

Figure 13a–c confirms that the differences in mean precipitation mainly arise from the differences in the simulated convective precipitation. GR simulates lower convective precipitation, which improves the skill of the simulations with respect to KF since the model tends to overpredict the summertime precipitation over the IP. This finding is in agreement with Fernandez et al. (2007), although other authors have reported a poor performance of the GR scheme for other domains and situations (Wang and Seaman 1997; Ratnam and Kumar 2005). The reasons making the GR simulations drier than the KF simulations could be (1) the fact that the KF scheme strongly tends to handle precipitation at the subgrid scale for cases associated with raising motion (Ferretti et al. 2000), which are common during summertime over the IP, and (2) the fact that, once the KF scheme is triggered, it remains activated for longer than the GR scheme (Gochis et al. 2002; Ratnam and Kumar 2005). We explore both possibilities focusing on the frequency and intensity of the precipitation events simulated by both CML schemes. We define 'frequency' as the mean number of days with precipitation above 1 mm and 'intensity' as the mean precipitation amount in those rainy days. Figure 14 depicts the observed values of these diagnostics (upper panels), the errors of the GR-subensemble mean in reproducing them (shaded colors in bottom panels) and the difference between the GR and the KF-subensemble means (contours in bottom panels).

Figure 14a, b shows that the observed intensity of the summertime rainy days grows towards the coast (up to 8 mm per rainy day on average), being softer in central areas (2–4 mm per rainy day), while the pattern of the observed frequency of the precipitation events displays a meridional gradient with the smallest values in the south (precipitation occurring 5–10 % of the summer days) and

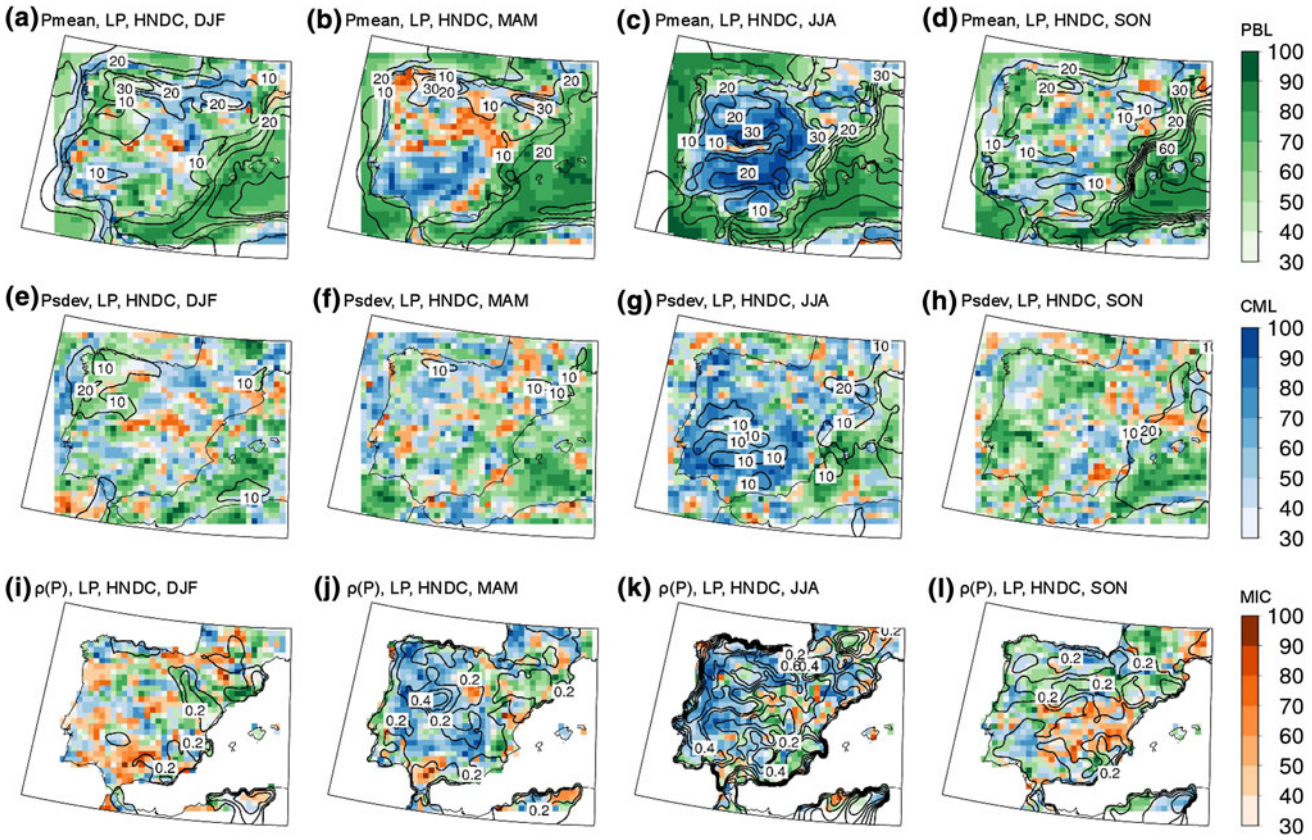

Fig. 9 As Fig. 8 for precipitation. Units in mm/month instead of °C regarding the *MES*

the highest in the north (precipitation in the 20–30 % of the days). Regarding the simulations (Fig. 14c, d), although the intensity is overall underestimated, with the error pattern (Fig. 14c) resembling the observed pattern (Fig. 14a), i.e the errors grow towards the coast up to 3 mm (about 30 % of the observed values), the frequency is overestimated, up to 30 % in north-eastern mountain systems. These numbers correspond to the GR-subensemble mean, but both intensity underestimation and frequency overestimation persist in the KF simulations (see the *MBE* values in Fig. 14c, d). These errors indicate that the observations show less frequent but more intense precipitation events than the simulations. This kind of errors is not unique in our simulations, and may related to the nature of the threshold criteria for triggering convection, originally developed for use in coarser resolutions (Gianotti et al. 2012). Therefore, the overestimation of the mean precipitation amount indicated above should be due to the overestimation of the frequency of the precipitation events (i.e. convective precipitation is initiated too frequently) since their intensity is actually underestimated. Indeed, the worst performance of KF is mainly due to the higher frequency of precipitation events simulated, while slightly reduces the dry bias because the higher precipitation simulated in the rainy days.

Some valuable insights can be drawn from this analysis. First, since the frequency of the rainy days is overestimated but their intensity is underestimated, the strongest events may be underestimated by the simulations. Second, the main reason making the GR scheme more suitable is that this scheme checks for its activation at every time step of the simulation. This reduces the overestimation of the frequency of the rainy days causing the reported overestimation of mean precipitation. However, KF is better able than GR to reproduce the intensity of the rainy events, while it overestimates more their frequency. This latter feature of the KF scheme is likely due to the fact that it remains activated for longer than GR, as commented before.

The GR scheme outperforms the KF scheme regarding also Psdev by simulating smaller interannual variability of the precipitation series. According to the general behavior of the model described in previous sections (related lower precipitation and lower interannual variability), the fact that GR provides a lower precipitation amount would also explain these differences in the precipitation variability.

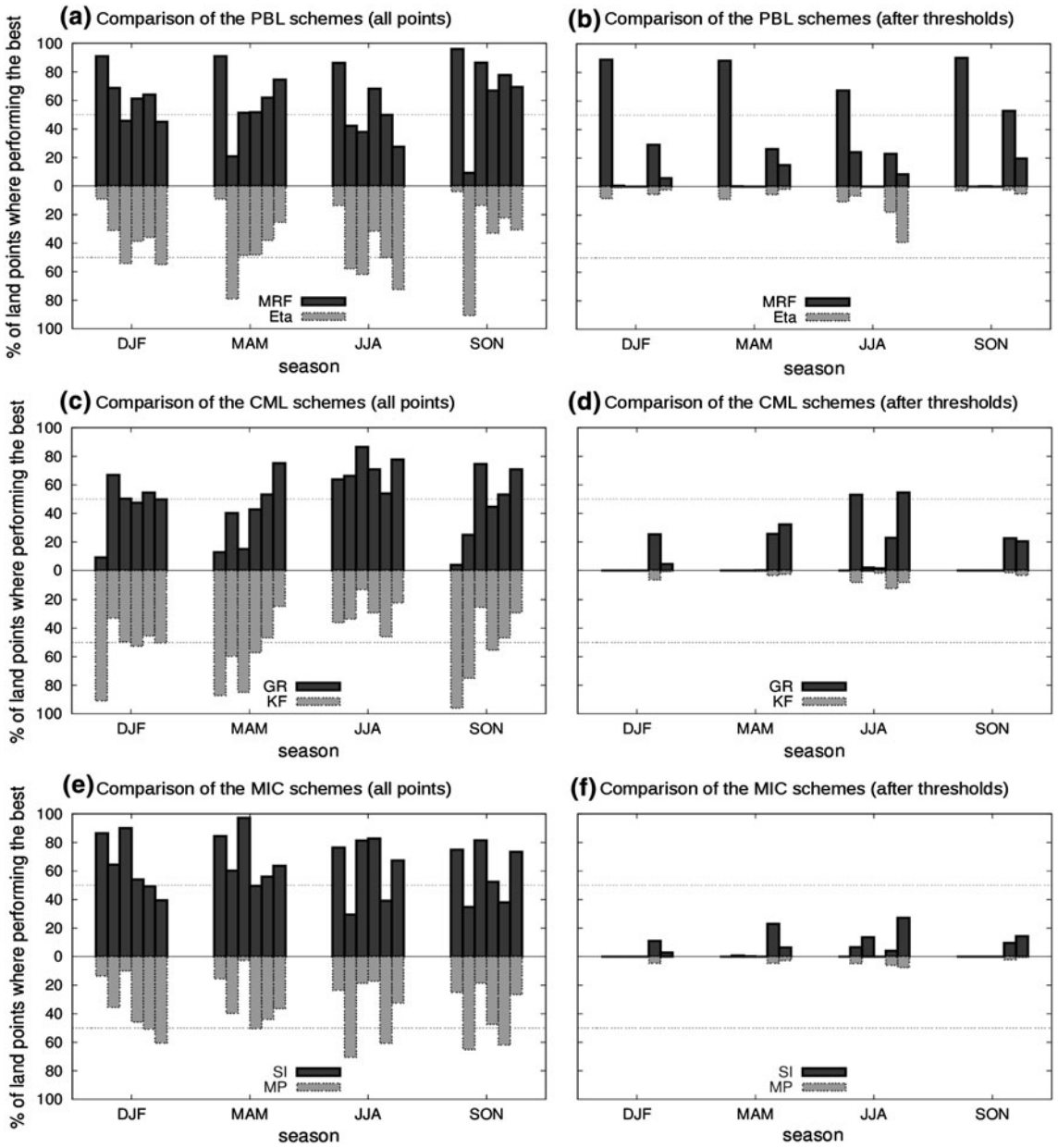

Fig. 10 Percentage of land points where each PBL, CML or MIC parametrization scheme outperforms the other option considered: **a** and **b** MRF versus Eta, **c** and **d** GR versus KF, and **e** and **f** SI versus MP. The analysis is made for each variable/statistic and every season, based on the subsensembles approach, i.e. by comparing each subensemble mean with the observations. Within each season, each box corresponds (from *left* to *right*) to the assessment of Tmean, Tsdev, ρ (between temperature series), Pmean, Psdev and ρ (between precipitation series). *Left graphs* (**a**, **c** and **e**) are obtained considering all the land grid points. *Right graphs* (**b**, **d** and **f**) are obtained considering only the land grid points where the signal-to-noise ratio computed for the particular scheme-induced spreads is above 1 (0.1) for mean (sdev) values and, in the case of ρ, where the differences between both subensemble means are above 0.1, as exposed in the text

This relationship is stronger for convective precipitation, since it is less dependent on the synoptic forcing than non-convective precipitation. In fact, Fig. 13d–f show that the largest differences between both subensemble means in Psdev appear effectively regarding the convective precipitation. Furthermore, the lower precipitation amount simulated by GR may be closely linked to the larger temperature variability also simulated by this scheme. The absence of precipitation could eventually imply totally dry soils provoking uneven positive anomalies of temperature due to the intensification of the positive soil moisture-temperature feedback (Jerez et al. 2010, 2012). Indeed,

Table 3 Skill scores (r, MAE and ρ) of the subensemble means given by fixing either the PBL scheme, the CML scheme or the MIC scheme to one of the two options considered, for each variable/statistic and every season

Var.	Stat.	Fixed scheme	DJF		MAM		JJA		SON	
			r	MAE	r	MAE	r	MAE	r	MAE
T	Mean	MRF	0.95	**0.70**	0.95	**0.71**	0.94	**1.06**	0.96	**0.87**
		Eta	0.95	**1.75**	0.95	**1.61**	0.92	**1.91**	0.96	**1.93**
		GR	0.95	1.16	0.95	1.13	0.93	1.42	0.96	1.43
		KF	0.95	1.07	0.95	1.06	0.93	1.47	0.96	1.31
		SI	0.95	**1.06**	0.95	**1.02**	0.94	1.39	0.96	1.32
		MP	0.95	**1.17**	0.95	**1.17**	0.93	1.50	0.96	1.40
	Sdev	MRF	0.48	11.5	0.29	24.8	0.57	8.9	**0.60**	11.7
		Eta	0.13	15.8	0.43	22.3	0.62	10.7	**0.66**	7.8
		GR	0.29	12.4	0.36	23.7	**0.65**	8.9	0.64	10.3
		KF	0.34	14.7	0.37	23.4	**0.54**	11.9	0.64	8.7
		SI	0.33	12.0	0.41	22.8	0.61	8.8	0.65	9.1
		MP	0.30	15.2	0.31	24.4	0.59	10.0	0.62	9.9
	ρ	MRF	0.93		0.91		0.88		0.94	
		Eta	0.94		0.90		0.90		0.91	
		GR	0.93		0.90		0.92		0.94	
		KF	0.93		0.92		0.87		0.92	
		SI	0.94		0.93		0.91		0.94	
		MP	0.92		0.88		0.86		0.92	
P	Mean	MRF	0.82	25.7	0.77	19.6	0.76	**88.0**	0.79	21.9
		Eta	0.79	27.0	0.73	20.6	0.78	**119.8**	0.67	25.0
		GR	0.81	26.5	0.75	20.6	**0.82**	**74.6**	0.74	23.8
		KF	0.81	26.0	0.76	19.2	**0.68**	**133.3**	0.73	22.8
		SI	0.81	25.9	0.76	19.9	0.77	95.8	0.75	22.9
		MP	0.81	26.4	0.76	19.8	0.78	110.6	0.73	23.3
	Sdev	MRF	0.82	25.6	0.44	20.6	0.45	45.2	0.44	26.6
		Eta	0.78	27.6	0.37	23.2	0.43	46.1	0.33	32.3
		GR	0.80	26.3	0.42	21.3	**0.51**	**41.8**	0.43	28.5
		KF	0.81	26.9	0.39	22.2	**0.27**	**55.2**	0.34	30.3
		SI	0.80	26.2	0.42	21.1	0.44	44.9	0.39	29.7
		MP	0.81	26.8	0.39	22.2	0.46	45.5	0.39	28.7
	ρ	MRF	0.93		0.86		0.31		0.83	
		Eta	0.93		0.84		0.48		0.82	
		GR	0.93		0.88		**0.55**		0.83	
		KF	0.93		0.81		**0.24**		0.82	
		SI	0.92		0.85		0.42		0.84	
		MP	0.94		0.85		0.38		0.82	

r and MAE are computed only for the mean and sdev statistics (for ρ it makes no sense). ρ is computed for the spatially averaged subensemble mean series. In the cases of Tsdev, Pmean and Psdev, MAE is computed from the values of error expressed in percentage with respect to the observations at each grid point. For Tmean it is in °C. Bold characters indicate that values fit the criteria described in the text

Fig. 13g–i shows larger values of the CML_{spread} for the variability of maximum temperatures than for the variability of minima. Since soil moisture-temperature feedback is stronger at daytime (Jerez et al. 2012), this would support the former explanation.

It should be stressed that these results can be dependent of the spatial resolution considered. For instance, it has been shown that changes in resolution produce different responses in precipitation depending on the selected convective parametrization scheme, being the triggering criteria the main contributor to the obtained differences (Giorgi and Marinucci 1996; Gallus 1999). In addition, finer resolutions could lead to more frequent, local and more intense precipitation events (Soares et al. 2012a).

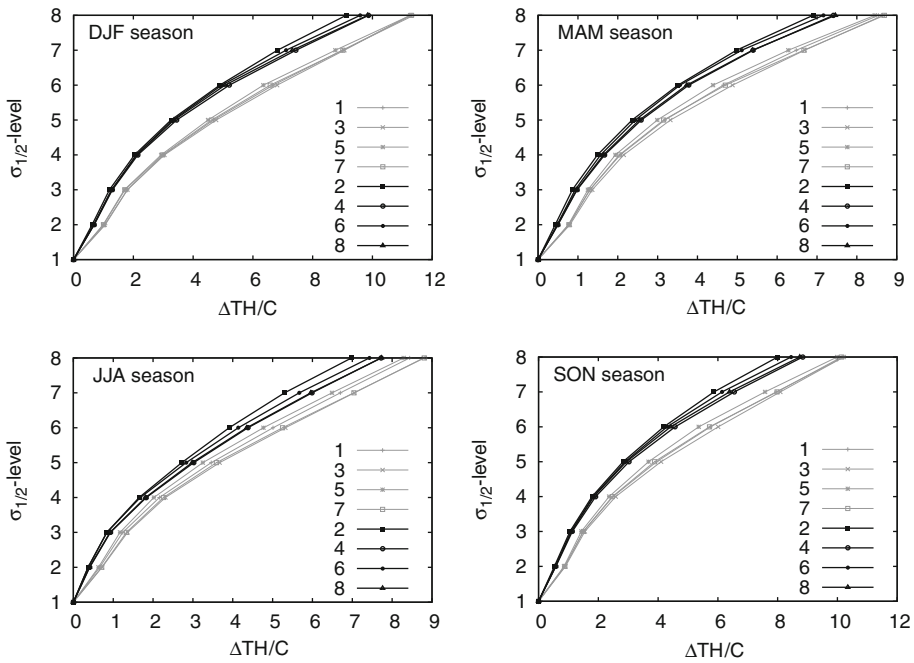

Fig. 11 Vertical profiles of potential temperature averaged over the whole IP from each experiment (see legend and Table 1) and for every season. For each experiment, the value at the first vertical level has been removed from the rest of the values at the upper levels in order to facilitate the comparison between the slope of the various curves. *Black lines* correspond to the experiments performed with the MRF PBL scheme, and *gray lines* correspond to experiments performed with the Eta PBL scheme. The $\sigma_{1/2}$-levels are (from *bottom* to *top*) 0.995, 0.985, 0.97, 0.945, 0.91, 0.87, 0.825 and 0.775, with approximate heights of 35, 110, 220, 400, 670, 980, 1350 and 1760 m above the land surface)

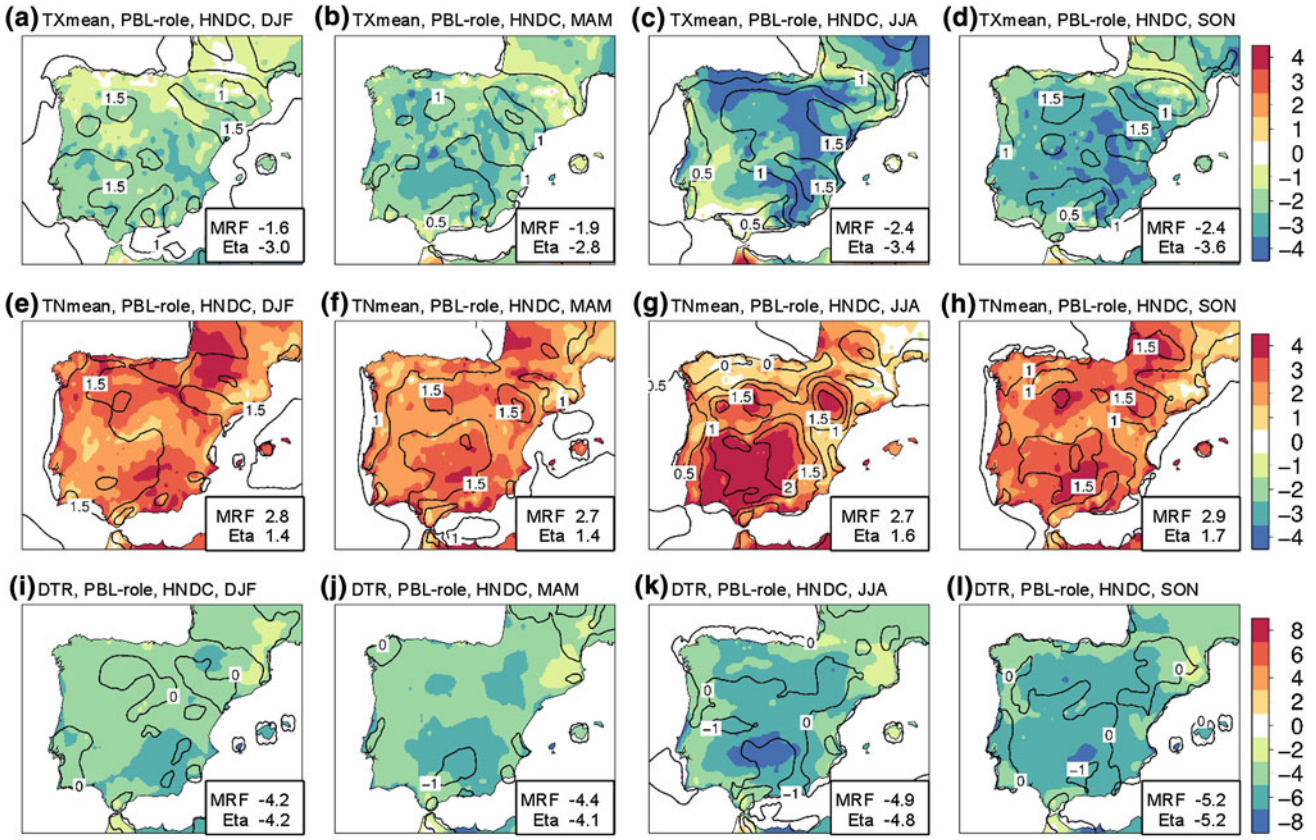

Fig. 12 *Shaded colors* represent the bias errors of the MRF-subensemble mean in reproducing mean values of daily maximum (TX; *first row*) and minimum (TN; *second row*) temperature, and the daily temperature range (DTR; *third row*). Contours depict the difference between the MRF-subensemble mean and the Eta-subensemble mean. All units in °C. In each panel the *MBE* corresponding to both subensemble means is provided (*right bottom*). Each column represents one season: winter (DJF), spring (MAM), summer (JJA) and autumn (SON)

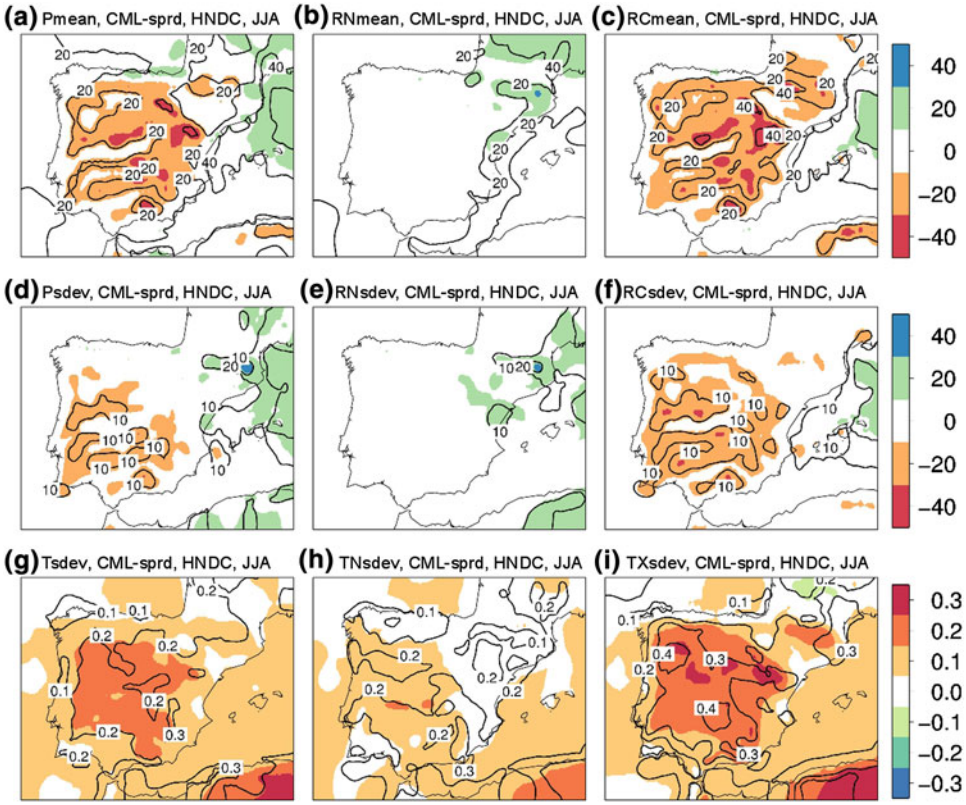

Fig. 13 Mean ensemble spread (contours) and the difference between the GR-subensemble mean and the KF-subensemble mean (*shaded colors*) in the summertime (JJA) patterns of Pmean, RNmean (RN is non-convective precipitation), RCmean (RC is convective precipitation), Psdev, RNsdev, RCsdev, Tsdev, TNsdev (TN is averaged daily minimum temperature) and TXsdev (Tx is averaged daily maXimum temperature). Units: mm/month for precipitation and °C for temperature

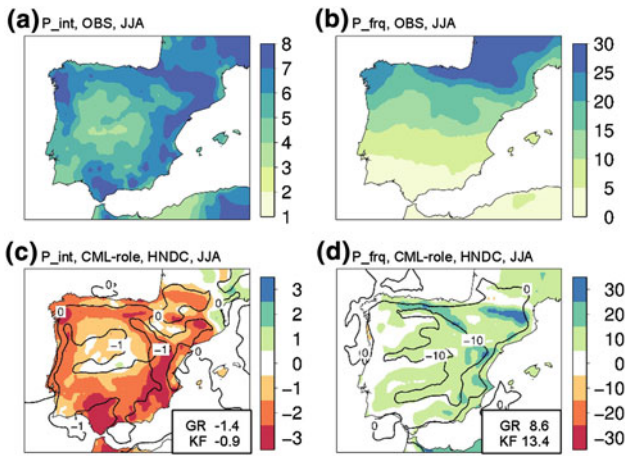

Fig. 14 *Upper panels* observed values of **a** the intensity and **b** the frequency of the rainy days in the summer (JJA) season (units in mm and percentage of rainy days respectively). *Bottom panels shaded colors* represent the bias errors of the GR-subensemble mean in reproducing those diagnostics and contours depict the difference between the GR-subensemble mean and the KF-subensemble mean. The *MBE* corresponding to both subensemble means is provided (*right bottom*)

However, we still believe that our results are likely non-dependent of the resolution from a qualitative point of view (although we acknowledge that it would have to be specifically checked), as these qualitative responses of convective precipitation to the model resolution are to be expected in all cases, i.e. regardless of the parametrization scheme considered.

8 Conclusions and remarks

This work assesses a multi-physics ensemble of present-climate simulations over the Iberian Peninsula. The ensemble consists of eight members that result from combining two schemes for the PBL, two for CML and two for the MIC processes within the same host regional model, MM5. The assessment is aimed at elucidating the role of the physical parametrized processes when reproducing the observed climatology in order to deepening the knowledge about intramodel discrepancies related to the model physics, as done (more frequently) with intermodel differences

(e.g. Jacob et al. 2007). The advantage of our experimental design is that the identical set-up of all the ensemble members, except for the physical configuration of the regional model, allows us to easily identify the leading underlying mechanisms.

The analysis focuses on seasonal averages of temperature and precipitation (mean values and interannual variability). The main characteristics of the ensemble are quantified through the ensemble mean and the ensemble spread. Furthermore, in order to better isolate the impact of each physical option (i.e. PBL, CML and MIC), we also deal with subensemble means (these subensembles are obtained by fixing one of the parametrization schemes at stake) and the so-called schemes-induced spreads. The E-OBS database is used for the validation purposes and several skill scores are considered: spatial and temporal correlation (r and ρ), mean absolute error (*MAE*) and mean bias error (*MBE*).

The purpose of this study is not to favor one parametrizations over others, but rather to characterize the responses and systematic errors for climate mesoscale applications of these parametrization schemes. This characterization is performed over a domain (the Iberian Peninsula) different from those used by modelers to test their developed schemes, and thus where their applicability is not obvious. This information will be useful for the design of future numerical prediction systems, for a further understanding of the parametrized processes and, hence, for future improvements of the parametrization schemes. It must be stressed, however, that some errors may be related to deficiencies in the internal model components or in the driving boundary conditions used (Herrera et al. 2010). Moreover, errors arising from the observational database could have masked some interesting signals (Fernandez et al. 2007). Thus, the reported behavior of the parametrization schemes must be considered within the host mesoscale model and may be dependent on both the reliability of the lateral boundary forcings and the confidence of the observational data. Despite these limitations and assumptions, much useful information can be drawn:

- The ensemble mean acceptably reproduces the climatology of the Iberian Peninsula, as characterized by mean values and interannual variability of temperature and precipitation, and thus the mean model skill can be considered satisfactory bearing in mind the following weaknesses. Mean temperature (Tmean) is systematically underestimated (Fernandez et al. 2007), but its spatial and interannual variations are well captured. The patterns of temperature variability (Tsdev) are, however, quite unrealistic, and Tsdev is strongly underestimated in the spring season. The spatial structure of the mean precipitation (Pmean) patterns is again quite reliable, although the simulations tend to overpredict the areas/seasons of light rainfall and to underpredict the heavy amounts, as found by Wang and Seaman (1997) over the United States and Rakesh et al. (2007) over India. Given that mean precipitation and precipitation variability (Psdev) are closely related, the bias patterns of both magnitudes are quite similar and precipitation variability is underestimated (overestimated) in the places where the amount of precipitation is underestimated (overestimated). The interannual variability of precipitation is well captured in the winter season, but this ability decreases dramatically in summer.
- Large variations in the model skill are found among the ensemble members. Spreads in Tmean are up to 3 °C, and spread in Tsdev, Pmean and Psedv amount to 50 % of the ensemble mean values (eventually even more). Regarding the temporal correlation of the simulated series with the observations, the spread may be up to 0.3. Some of these spreads are of similar magnitude as those obtained in a multi-model ensemble (Jacob et al. 2007). This suggests that a large part of the intermodel spread could be attributed to differences in the physical setup of the various models.
- Although a proper choice of parametrization schemes may considerably reduce the magnitude of biases since there are areas showing both large bias and large spread, the spatial distribution of the largest biases does not change much when considering each member of the ensemble. This indicates that these areas are particularly complex. This complexity, on the one hand, would promote a wider variety in the response to the various schemes, and thus a larger spread. On the other hand, such complexity would prevent that single measurements are representative of surrounding areas. Thus, since the model is not able to reproduce singularities, large biases appear.
- No single experiment always outperforms the others (as Fernandez et al. (2007) pointed out already). Nonetheless, some schemes have been found to perform better than others. We first identified that the largest spreads are related to changes of the PBL scheme and the CML scheme, while the change of the MIC scheme seems less relevant. The PBL mainly affects the mean temperature patterns throughout the year, while the CML mainly affects the patterns of mean precipitation and of interannual variability of temperature and precipitation in the summer season. Bearing in mind the regional variations regarding the most appropriate parametrization schemes, we have found in these cases that the use of the MRF and GR schemes (for the PBL and cumulus modeling, respectively) provides important benefits when considering the entire Iberian

Peninsula. The underlying responsible mechanisms have been investigated, which further yields some caveats. Although these findings do not provide an universal framework for the parametrization schemes, they could be a helpful reference for future decisions. They are:

- MRF produces warmer temperatures than the Eta PBL scheme, which reduces the cold bias over the entire domain. Such improvement may be related to the higher and better mixed PBLs simulated by the MRF scheme (de Arellano et al. 2001; Berg and Zhong 2005). Actually, MRF produces warmer values than Eta for both maximum and minimum temperature. However, although maximum temperature is underestimated, minimum temperature is overestimated by the model, and thus the Eta model would outperform the MRF scheme for simulating cold extremes. Nonetheless, the PBL_{spread} disappears regarding the daily temperature range, which indicates that both schemes should be considered equally valid if the purpose is to evaluate deviations after removing the systematic cold or warm biases.
- GR is drier than the KF CML scheme in the summer season (as reported by Ferretti et al. (2000); Gochis et al. (2002); Liang et al. (2004); Ratnam and Kumar (2005) in other situations), which produces larger temperature variability and smaller precipitation variability. Differences, especially in the simulated convective precipitation, help to reduce the overprediction of the summer precipitation and its interannual variability over the entire domain, and to improve the representation of the temperature variability due to the larger variability of the maximum temperatures with drier soils (Jerez et al. 2010, 2012). The key is that, contrary to the KF scheme, the GR scheme checks at every time step whether or not the triggering conditions are fulfilled, and this reduces the overestimation in the number of rainy days detected in the simulations. However, GR is less able than KF to adequately reproduce the intensity of the precipitation events, which is systematically underestimated.

This study further highlights the strong dependence of the regional model's skill on its physical configuration when reproducing observed climatologies. When projecting future climate changes, such influence could be amplified, displaced or even canceled by compensation. How discrepancies in present period simulations could propagate and affect future climate change projections is still uncertain (Liang et al. 2008) and, given these results, the topic is worthy of further investigation. A accompanying second work is devoted to evaluate the climate change patterns obtained from analogous ensembles of control and scenario simulations, assessing how the physics-derived discrepancies behave and affect them.

Finally, it is in the very nature of the models that while the averaged climate is well represented, extreme conditions are not (Soares et al. 2012b). We also found this irrespective of the model physics configuration: as MRF reduces the underestimation of maximum temperature but enlarges the overestimation of minimum temperature, the daily temperature range is underestimated regardless of the PBL scheme; as GR reduces the overestimation of the frequency of the precipitation events but enlarges the underestimation of their intensity, the strong precipitation events are poorly reproduced regardless of the CML scheme. This raises a big question for future: if the occurrence of extreme events increases (Beniston et al. 2007) modifying the averaged climate, will models be able to catch that? To achieve that goal, improved parametrization schemes, adapted to the increasingly used high resolutions, would be desired.

Acknowledgments This work was funded by the Spanish Ministry of the Environment (project ESCENA, Ref. 200800050084265) and Project CORWES (CGL2010-22158-C02-02). The authors also gratefully acknowledge funding from the Euro-Mediterranean Institute of Water (IEA). Pedro Jiménez-Guerrero acknowledges the Ramón y Cajal Programme. Sonia Jerez thanks the Portuguese Science Foundation (FCT) for her current financial support through the project ENAC (PTDC/AAC-CLI/103567/2008) and Ricardo M. Trigo for his personal scientific support.

References

Argüeso D, Hidalgo-Muñoz JM, Gámiz-Fortis SR, Esteban-Parra MJ, Dudhia J, Castro-Díez Y (2011) Evaluation of WRF parameterizations for climate studies over Southern Spain using a multistep regionalization. J Clim 24:5633–5651

Beniston M, Stephenson DB, Christensen OB, Ferro CAT, Frei C, Goyette S, Halsnaes K, Holt T, Jylhä K, Koffi B, et al (2007) Future extreme events in European climate: an exploration of regional climate model projections. Clim Change 81:71–95

Berg LK, Zhong S (2005) Sensitivity of MM5-Simulated Boundary Layer characteristics to turbulence parameterizations. J Appl Meteorol 44:1467–1483

Bright DR, Mullen SL (2002) The Sensitivity of the Numerical Simulation of the Southwest Monsoon Boundary Layer to the choice of PBL turbulence parameterization in MM5. Wea Forecast 17(1):99–114

Castro CL, Pielke RA, Leocini G (2005) Dynamical downscaling: assessment of value retained and added using the regional atmospheric modeling system (RAMS). J Geophys Res 110:D5

Chen F, Dudhia J (2001) Coupling an advanced land surfacehydrology model with the penn state NCAR MM5 Modeling System. Part I: model implementation and sensitivity. Mon Wea Rev 129(4):569–585

Chiriaco M, Vautard R, Chepfer H, Haeffelin M, Dudhia J, Wanherdrick Y, Morille Y, Protat A (2006) The ability of MM5 to simulate ice clouds: systematic comparison between

simulated and measured fluxes and lidar/radar profiles at the SIRTA atmospheric observatory. Mon Wea Rev 134(3):897–918
Christensen OB (1999) Relaxation of soil variables in a regional climate model. Tellus A 51(5):674–685
Colle BA, Mass CF (2000) The 59 February 1996 flooding event over the Pacific Northwest: sensitivity studies and evaluation of the MM5 precipitation forecasts. Mon Wea Rev 128(3):593–617
de Arellano JVG, Vellinga OS, Holtslag AAM, Bosveld FC, Baltink HK (2001) Observational evaluation of PBL parameterizations modeled by MM5. Preprints 11th PSU/NCAR MM5 Users Workshop, Boulder, CO, NCAR
Dudhia J (1989) Numerical study of convection observed during the Winter Monsoon experiment using a mesoscale two-dimensional model. J Atmos Sci 46(20):3077–3107
Fernandez J, Montavez JP, Saenz J, Gonzalez-Rouco JF, Zorita E (2007) Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: annual cycle. J Geophys Res 112:D04,101
Ferretti R, Paolucci T, Zheng W, Visconti G, Bonelli P (2000) Analyses of the precipitation pattern on the alpine region using different cumulus convection parameterizations. J Appl Meteorol 39(2):182–200
Font-Tullot I (2000) Climatologa de Espaa y Portugal. Ed Universidad de Salamanca, Salamanca
Forkel R, Knoche R (2006) Regional climate change and its impact on photooxidant concentrations in southern Germany: simulations with a coupled regional chemistry-climate model. J Geophys Res 111:D12,302
Frank H, Landberg L (1997) Modelling the wind climate of Ireland. Boundary Layer Meteorol 85:359–378
Gallus WA (1999) Eta simulations of three extreme precipitation events: sensitivity to resolution and convective parameterization. Wea Forecast 14:405–426
Galos B, Lorenz P, Jacob D (2007) Will dry events occur more often in Hungary in the future?. Environ Res Lett 2(3):034,006
Gao X, Giorgi F (2008) Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model. Global Planet Change 62:195–209
Gianotti RL, Zhang D, Eltahir EAB (2012) Assessment of the regional climate model version 3 over the maritime continent using different cumulus parameterization and land surface schemes. J Clim 25(2):638–656
Giorgi F, Bi X (2000) A study of internal variability of a regional climate model. J Geophys Res 105(D24):29503–29529
Giorgi F, Marinucci MR (1996) An investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies. Mon Wea Rev 124(1):148–166
Gochis DJ, Shuttleworth W, Yang ZL (2002) Sensitivity of the modeled North American Monsoon regional climate to convective parameterization. Mon Wea Rev 130(5):1282–1298
Gomez-Navarro JJ, Montavez JP, Jimenez-Guerrero P, Jerez S, Garcia-Valero JA, Gonzalez-Rouco JF (2010) Warming patterns in regional climate change projections over the Iberian Peninsula. Meteorol Zeitschrift 19(3):275–285
Gomez-Navarro JJ, Montavez JP, Jerez S, Jimenez-Guerrero P, Lorente-Plazas R, Gonzalez-Rouco JF, Zorita E (2011) A regional simulation over the Iberian Peninsula for the last millenium. Clim Past 7(2):451–472
Grell GA, Dudhia J, Stauffer DR (1994) A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech Note 398+STR, Natl Cent for Atmos Res, Boulder, CO
Grubii V, Vellore RK, Huggins AW (2005) Quantitative precipitation forecasting of wintertime storms in the Sierra Nevada: sensitivity to the microphysical parameterization and horizontal resolution. Mon Wea Rev 133(10):2834–2859
Han Z, Hiromasa U, An J (2008) Evaluation and intercomparison of meteorological predictions by five MM5-PBL parameterizations in combination with three land-surface models. Atmos Environ 42(2):233–249
Haylock MR, Hofstra N, Tank AMGK, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res 113:D20,119
Herrera S, Fita L, Fernandez J, Gutierrez JM (2010) Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain. J Geophys Res 115(D21117):148–227
Hofstra N, Haylock M, New M, Jones P (2009) European high-resolution gridded data set of daily precipitation and surface temperature. J Geophys Res 114:D20,119
Hong SY, Pan HL (1996) Nonlocal Boundary Layer vertical diffusion in a medium-range forecast model. Mon Wea Rev 124:2322–2339
Jacob D, Barring L, Christensen OB, Christensen JH, de Castro M, Deque M, Giorgi F, Hagemann S, Lenderink G, Rockel B, Sanchez E, Schaer C, Seneviratne SI, Somot S, van Ulden A, van denHurk B (2007) An inter-comparison of regional climate models for Europe: model performance in present-day climate. Clim Change 81:31–52
Janjic ZI (1994) The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon Wea Rev 122(5):927–945
Jerez S, Montavez JP, Gomez-Navarro JJ, Jimenez-Guerrero P, Jimenez J, Gonzalez-Rouco JF (2010) Temperature sensitivity to the land-surface model in MM5 climate simulations over the Iberian Peninsula. Meteorol Zeitschrift 19(4):363–374
Jerez S, Montavez JP, Gomez-Navarro JJ, Jimenez PA, Jimenez-Guerrero P, Lorente-Plazas R, Gonzalez-Rouco JF (2012) The role of the land-surface model for climate change projections over the Iberian Peninsula. J Geophys Res 117:D01,109
Jimenez-Guerrero P, Gomez-Navarro JJ, Jerez S, Lorente R, Garcia-Valero JA, Montavez JP (2011) Variation of secondary inorganic aerosols (SIA) in Europe for the 21st century (1991-2100). Atmos Environ 45:1059–1063
Kain JS (2004) The KainFritsch convective parameterization: an update. J Appl Meteorol 43(1):170–181
Kain JS, Fritsch JM (1990) A one-dimensional entraining/detraining Plume model and its application in convective parameterization. J Atmos Sci 47(23):2784–2802
Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL (2002) NCEP-DOE AMIP-II reanalysis (R-2). Bull Am Meteorol Soc 83:1631–1643
Kerkhoven E, Gan TY, Shiiba M, Reuter G, Tanaka K (2006) A comparison of cumulus parameterization schemes in a numerical weather prediction model for a monsoon rainfall event. Hydrol Process 20:1961–1978
Koo GS, Boo KO, Kwon WT (2009) Projection of temperature over Korea using an MM5 regional climate simulation. Clim Res 40(2–3):241–248
Koster RD, Suarez MJ (1995) Relative contributions of land and ocean processes to precipitation variability. J Geophys Res 100:D7
Kotroni V, Lagouvardos K (2001) Precipitation forecast skill of different convective parameterization and microphysical schemes: application for the cold season over Greece. Geophys Res Lett 28(10):1977–1980
Leung LR, Gustafson WI (2005) Potential regional climate change and implications to U.S. air quality. Geophys Res Lett 32:L16,711
Leung LR, Qian Y, Bian X (2003) Hydroclimate of the Western United States based on observations and regional climate

simulation of 19812000. Part I: seasonal statistics. J Clim 16(12):1892–1911
Liang XZ, Li L, Dai A, Kunkel KE (2004) Regional climate model simulation of summer precipitation diurnal cycle over the United States. Geophys Res Lett 31:L24,208
Liang XZ, Kunkel KE, Meehl GA, Jones RG, Wang JXL (2008) Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections. Geophys Res Lett 35(8):L08,709
Mapes BE, Warner TT, Xu M, Gochis DJ (2004) Comparison of cumulus parameterizations and entrainment using domain-mean wind divergence in a regional model. J Atmos Sci 61(11):1284–1295
McFarquhar GM, Zhang H, Heymsfield G, Halverson JB, Hood R, Dudhia J, Marks F (2006) Factors affecting the evolution of hurricane Erin (2001) and the distributions of hydrometeors: role of microphysical processes. J Atmos Sci 63(1):127–150
Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res 102:16,663–16,682
Pan Z, Segal M, Arrit R, Takle E (2004) On the potential change in solar radiation over the US due to increases of atmospheric greenhouse gases. Renow Energy 29:19231928
Pryor S, Barthelmie R, Kjellstrom E (2005) Potential climate change impact on wind energy resources in northern Europe: analyses using a regional climate model. Clim Dyn 25:815–835
Rakesh V, Singh R, Pal PK, Joshi PC (2007) Sensitivity of mesoscale model forecast during a satellite launch to different cumulus parameterization schemes in MM5. Pure Appl Geophys 164:1617–1637
Ratnam JV, Kumar KK (2005) Sensitivity of the simulated monsoons of 1987 and 1988 to convective parameterization schemes in MM5. J Clim 18(14):2724–2743
Reisner J, Rasmussen RM, Bruintjes RT (1998) Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. Quart J R Meteorol Soc 124(548):1071–1107
Rodrigo FS, Gomez-Navarro JJ, Montavez JP (2012) Climate variability in Andalusia (southern Spain) during the period 1701–1850 based on documentary sources: evaluation and comparison with climate model simulations. Clim Past 8:117–133
Rummukainen R (2010) State-of-the-art with regional climate models. WIREs Clim Change 1:82–96
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. Clim Dyn 33(5):723–736
Soares PMM, Cardoso RM, Miranda PMA, de Medeiros J, Belo-Pereira M, Espirito-Santo F (2012a) WRF high resolution dynamical downscaling of ERA-Interim for Portugal. Clim Dyn. doi:10.1007/s00382-012-1315-2
Soares PMM, Cardoso RM, Miranda PMA, Viterbo P, Belo-Pereira M (2012b) Assessment of the ENSEMBLES regional climate models in the representation of precipitation variability and extremes over Portugal. J Geophys Res 117(D7):D07,114
Solman SA, Nunez MN, Cabré MF (2008) Regional climate change experiments over southern South America. I: present climate. Clim Dyn 30(5):533–552
Stensrud D (2007) Parameterization schemes: keys to understanding numerical weather prediction models. Cambridge University Press, Cambridge
van der Linden P, Mitchell JFB (2009) ENSEMBLES: climate change and its impacts. Summary of research and results from the ENSEMBLES projec. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK
Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, et al (2005) The ERA-40 re-analysis. Quart J R Meteorol Soc 131:2961–3012
Wang W, Seaman NL (1997) A comparison study of convective parameterization schemes in a mesoscale model. Mon Wea Rev 125(2):252–278
Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30(1):79
Yang MJ, Tung QC (2003) Evaluation of rainfall forecasts over Taiwan by four cumulus parameterization schemes. J Meteorol Soc Jpn 81(5):1163–1183
Zhang DL, Zheng WZ (2004) Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. J Appl Meteorol 43(1):157–169