Clim Dyn
DOI 10.1007/s00382-015-2529-x

An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation

Markel García-Díez · Jesús Fernández · Robert Vautard

Received: 2 April 2014 / Accepted: 13 February 2015

Abstract Regional Climate Models are widely used tools to add detail to the coarse resolution of global simulations. However, these are known to be affected by biases. Usually, published model evaluations use a reduced number of variables, frequently precipitation and temperature. Due to the complexity of the models, this may not be enough to assess their physical realism (e.g. to enable a fair comparison when weighting ensemble members). Furthermore, looking at only a few variables makes difficult to trace model errors. Thus, in many previous studies, these biases are described but their underlying causes and mechanisms are often left unknown. In this work the ability of a multi-physics ensemble in reproducing the observed climatologies of many variables over Europe is analysed. These are temperature, precipitation, cloud cover, radiative fluxes and total soil moisture content. It is found that, during winter, the model suffers a significant cold bias over snow covered regions. This is shown to be related with a poor representation of the snow-atmosphere interaction, and is amplified by an albedo feedback. It is shown how two members of the ensemble are able to alleviate this bias, but by generating a too large cloud cover. During summer, a large sensitivity to the cumulus parameterization is found, related to large differences in the cloud cover and short wave radiation flux. Results also show that small errors in one variable are sometimes a result of error compensation, so the high dimensionality of the model evaluation problem cannot be disregarded.

Keywords WRF · Multi-physics · Model evaluation · Radiation · Soil moisture · CERES · E-OBS · GLDAS · CORDEX · EURO-CORDEX

Electronic supplementary material The online version of this article (doi:10.1007/s00382-015-2529-x) contains supplementary material, which is available to authorized users.

M. García-Díez (✉)
Unitat de Dinàmica i Impacte Climàtic (UDIC), Institut Català de Ciències del Clima, 08005 Barcelona, Catalonia, Spain
e-mail: markel.garcia@ic3.cat

M. García-Díez · J. Fernández
Grupo de Meteorología, Dept. Matemática Aplicada y Ciencias de la Computación, Univ. de Cantabria, Avda. de los Castros, s/n, 39005 Santander, Spain

R. Vautard
Laboratoire des Sciences du Climat et de l'Environnement, Institute Pierre-Simon-Laplace, CEA, CNRS, UVSQ, Gif sur Yvette, France

1 Introduction

Over the last years, climate science has intensified focus on regional scales and impacts. In this context, dynamical downscaling arises as one of the main tools to get regional information. This methodology uses Regional Climate Models (RCMs) to produce high resolution climate variables from the coarser Global Climate Models (GCMs). The COordinated Regional climate Downscaling EXperiment (CORDEX) (Giorgi et al. 2009) is the first worldwide coordination framework for downscaling climate information. Embedded into it, the European regional climate modeling community has set up the Euro-CORDEX framework, coordinating the contributions to the European CORDEX domain (Jacob et al. 2013; Vautard et al. 2013; Kotlarski et al. 2014). The main goal of CORDEX is to assess regional climate change and the associated uncertainties by means of an ensemble of simulations for each region. It is, however, unclear how to weight the individual contributions in these ensembles. Some authors (see, among others Christensen et al. 2010; Herrera et al. 2010) propose to underweight or remove the models performing worse in the evaluation simulations nested into state-of-art reanalyses. When comparing these simulations with observations, a fundamental problem arises. Climate models are complex programs adjusted with many parameters, some of which are difficult or impossible to measure (Mauritsen et al. 2012). Most evaluation studies use only a few variables, being precipitation and temperature (P&T onwards) the most popular ones, followed by sea level pressure (SLP) or 500 hPa geopotential height. This can lead to reduce the bias by balancing out errors, instead of improving physical realism. For example, Samuelsson et al. (2011) performed a thorough evaluation of an RCM, and found that excessive incoming solar radiation was being compensated by a too large albedo in snow-free areas over Southern Europe. (Pessacg et al. 2013) studied the surface energy balance of seven RCMs over South America. They found that some models reached small temperature biases by compensating large errors in the radiative and heat fluxes. These were related with errors in the cloud fraction and albedo. To our knowledge, most RCM developers do evaluate them with many observations apart from P&T but, often, these results are kept unpublished and retained as know-how of the group. As P&T are key variables to assess the biophysical impacts of climate, it is reasonable to focus on them. However, model reliability requires transparency in the complete evaluation and adjustment process.

In this context, multi-physics ensembles (MPEs) appear as a methodology to improve the physical insight behind model biases. In these ensembles, the model is perturbed by changing the physical parameterizations used to represent unresolved phenomena (e.g. microphysics, cumulus, etc). Each parameterization combination leads to different simulated climates, so their spread is an estimate of the model uncertainty arising from the representation of the unresolved phenomena. Most previous multi-physics studies with RCMs focused in P&T (e.g. Mooney et al. 2013; Argüeso et al. 2011; Awan et al. 2011). Also, the majority of these studies were carried out using the Weather Research and Forecasting Model (WRF) (Skamarock et al. 2008), or its predecessor MM5 (Grell et al. 1995), because these models allow to easily choose among a large amount of state-of-the-art parameterizations.

Argüeso et al. (2011) compared eight parameterization combinations with observations over southern Spain. They found precipitation to be more sensitive to the choice of parameterizations (especially to cumulus and the planetary boundary layer—PBL) than temperature. Although they provide some recommendations, they conclude that there is no combination clearly better than the others. This conclusion is shared by most multi-physics studies (Fernández et al. 2007; García-Díez et al. 2012; Jerez et al. 2013).

Awan et al. (2011) analysed a large number of parameterization combinations over the Alps, including cumulus, microphysics and PBL. They found that it was possible to consistently improve the model results by choosing the most adequate combination. They also found that the schemes interact in a non-linear manner, so it is not possible to predict the result of a combination from the effects of changing each scheme alone from a control run. Mooney et al. (2013) compared a 12-member MPE with the observations over Europe. The MPE was constructed combining two land-surface schemes, two PBL schemes, two long-wave radiation schemes and two microphysics schemes. However, they did not use additional observations apart from P&T and SLP. The authors concluded that WRF reproduces temperature reasonably, but that it has problems with precipitation, which is largely overestimated. Although these studies find relevant results, they do not provide much information about the misrepresented physical processes or the model deficiencies behind the biases.

In this work, our goal is to show how the multi-variable analysis of an MPE can be used to improve the understanding of the physical realism of a model and to identify sources of error compensation. With this aim, a MPE is evaluated regarding not only P&T, but also radiation fluxes, cloud cover, soil moisture and albedo. It is well known that these additional variables play an essential role in representing climate and climate sensitivity (Jaeger and Seneviratne 2011; Samuelsson et al. 2011; Watanabe et al. 2012). Studying all these variables in an RCM MPE, which is unprecedented to our knowledge, shows how the uncertainty introduced by physical parameterizations behaves depending on the season and region. Note that multi-physics design does not account for all the uncertainty but, in contrast with the multi-model approach, in an MPE the differences between the members are traceable to the physical processes parametrized. Thus, another goal of this work is to analyse the main deficiencies of the model as well as their sensitivity to the parameterizations. The MPE approach allows us to discern whether these deficiencies are general or characteristic of one parameterization or parameterization set, and this helps to trace their origin. The large amount of dimensions involved (variable, physics, seasons) prevents a complete evaluation of the MPE. Thus, only the most relevant results are shown.

The domain used is the CORDEX-compliant domain for Europe at 0.44° horizontal resolution (Fig. 1), and the model used is the WRF model. Thus, the present results are directly relevant for the WRF community involved in CORDEX, but also for other RCMs, given that some of the problems detected in WRF are also present in other Euro-CORDEX RCMs (Kotlarski et al. 2014; Samuelsson et al. 2011).

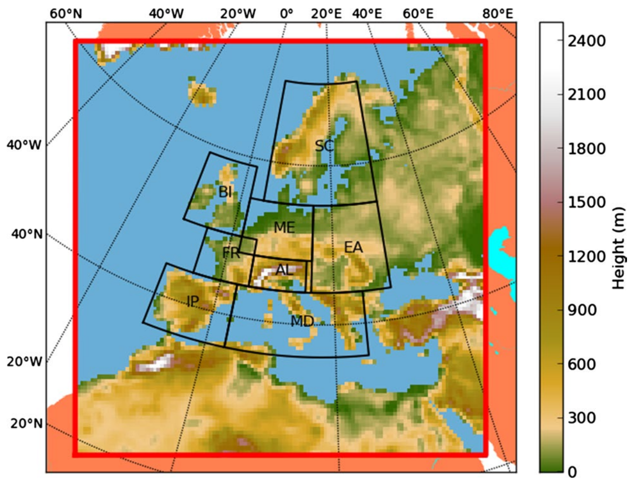

Fig. 1 Model domain and topography

Table 1 Summary of the parameterization combinations used in the WRF multi-physics ensemble

Label	Cumulus	Microphysics	Radiation
MPE-A	KF	WSM6	CAM
MPE-C	KF	WSM3	CAM
MPE-D	BM	WSM6	RRTMG
MPE-F	GD	WSM5	RRTMG
MPE-G	GD	WSM6	CAM
MPE-H	KF	M-2M	CAM
MPE-M	GD	M-2M	CAM
REFOR	GD	WSM6	CAM

Each scheme is individually described in Table 2

2 Methododology and data

2.1 Model configuration

A new 7-member MPE has been produced with the WRF model. The parameterization combinations are based in those used by the WRF contributions to the Euro-CORDEX evaluation simulations Vautard et al. (2013), and are summarized in Table 1. Table 2 shows the main features of each parameterization. As agreed in CORDEX, ERA-INTERIM reanalyses (Dee et al. 2011) have been used as a "perfect" GCM to downscale in the evaluation simulations. A 5-year period (2002–2006) was covered, leaving 1 year (2001) as spin up. The RCM used was WRF model version 3.3.1, which is an open source model described in detail in Skamarock et al. (2008). This model allows to choose among a large amount of state-of-the-art parameterizations. The WRF4G execution framework (Fernández-Quiruelas et al. 2010) was used to configure, execute and monitor the MPE experiment.

The following parameterizations have been changed (see Table 2 for abbreviations and brief details): Cumulus (KD and GD), microphysics (WSM3, WSM5, WSM6 and M2M) and radiation (CAM and RRTMG). Some configurations differ from others in more than one parameterization (e.g. MPE-D). The choice is justified by our aim to reproduce the parameterization sets used in the Euro-CORDEX WRF ensemble (Vautard et al. 2013). We found in additional tests that results running with WSM5 and WSM6 are almost identical, so this leaves MPE-D as the only simulation with no possible one-step comparison. Additionally, a simulation in "reforecast mode" (REFOR) has been carried

Table 2 Summary of the WRF parameterizations used in this work

Label	Description
YSU	Yonsei University PBL scheme (Hong et al. 2006). Non-local diffusion scheme
KF	Kain–Frisch cumulus scheme (Kain 2004). Mass-flux scheme able to accumulate CAPE
BMJ	Betts–Miller–Janjic convection scheme (Janjic 2000). Deep layer control scheme unable to accumulate CAPE
GD	Grell–Devenyi cumulus scheme (Grell and Devenyi 2002). 144 member ensemble made with mass-flux schemes
WSM3	WRF single-moment microphysics parameterization (Hong et al. 2004) with three species (vapor, cloud water/ice and rain/snow)
WSM5	Similar to WSM3 with two more species (vapour, cloud water, cloud ice, rain and snow are treated independently)
WSM6	As WSM5 with one more species (graupel)
M2M	Morrison two-moment (Morrison et al. 2009). Complex parameterization with six species and two moments (density and mixing ratio)
CAM	Radiation parameterization of the NCAR Community Atmosphere Model (Collins et al. 2004). More complex than RRTMG
RRTMG	Long wave radiation parameterization. Improved version (Iacono et al. 2008) of the Rapid Radiative Transfer Model of Mlawer et al. (1997)
Noah	Noah Land-Surface Model (Chen and Dudhia 2001). Land-surface parameterization with four layers

out by restarting the model daily from ERA-INTERIM, and leaving 12 h of spin-up. This running scheme preserves the correlation with the driving reanalyses (Menéndez et al. 2014), and it was used to distinguish model errors that develop quickly from those that build up over a long period. The parameterization set used in REFOR and MPE-G is the same, enabling direct comparison.

2.2 Observational data

2.2.1 E-OBS dataset

E-OBS (Haylock et al. 2008) is an observation-based gridded product that covers Europe with a daily frequency. In the present work P&T data from the E-OBS v8.0 in the 0.5° grid have been used. The dataset was produced by interpolating station data from the European Climate Assessment and Data (ECA&D http://eca.knmi.nl). Recently, some studies have found problems and inaccuracies in E-OBS (Herrera et al. 2010; Kysely and Plavcova 2010; Hofstra et al. 2009). These affect especially precipitation extremes, in areas with complex orography and scarce stations. Over some of these areas the station coverage improved in latest versions, while it remained poor in a few areas (e.g. North Africa). In general, the mean climatologies derived from E-OBS can be considered of reasonable quality (Herrera et al. 2010), and our work will only use these.

2.2.2 Radiation data from CERES

The Cloud and Earth's Radiant Energy System (CERES) is an experiment from the National Aeronautics and Space Administration (NASA) devoted to process satellite observations, focusing in the earth radiation budget. In the present study, the radiation flux data from CERES labeled as energy balanced and filled (EBAF) have been used in its version 2.7. These data are provided as monthly averages with a resolution of 1°. In the CERES website[1] a complete description of the data elaboration process and its issues can be found. The raw data used by this dataset come from AQUA, TERRA and geostationary satellites. To produce the EBAF data, the energy balance is adjusted to that inferred by Loeb et al. (2012) from the measured warming of the oceans. Cloud cover observations are not available in EBAF, so we used those from SYN1deg (the processing step previous to EBAF, before adjusting the energy balance).

To check the robustness of the results, all the maps shown in the paper using CERES data have been reproduced with the independent GEWEX-SRB dataset, which does not use data from MODIS (Gupta et al. 2006). These are included in the supplementary material.

[1] http://ceres.larc.nasa.gov/order_data.php.

2.2.3 Global Land Data Assimilation System soil moisture content data

Surface observations of soil moisture are scarce, and the large spatial variability of this variable makes its use challenging (Greve et al. 2013). Furthermore, satellite products are also non-trivial to use, as they measure only the moisture content of the first centimeters of the soil (Dharssi et al. 2011), in contrast with the few meters that Land Surface Models (LSMs) use to represent the rooting zone. The Noah LSM (Chen and Dudhia 2001), which is the LSM used in all the simulations produced for the present study, is integrated in four layers up to 4 meters deep. Thus, to evaluate the total soil moisture content, data from the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004) has been used. Namely, we used GLDAS version 2 with 0.25° resolution. This dataset is a global soil reanalysis produced by running a LSM forced with data as realistic as possible. Actually, GLDAS data are available for four different LSMs, and one of them is the Noah LSM. Forcing Noah with observations allows soil variables to be comparable with those produced by WRF, avoiding the problems that arise when using direct observations. On the other hand, the use of a LSM reanalysis prevents the assessment of errors arising from the LSM. Thus, in the analysis of soil moisture (Sect. 3.3) we assume that the atmospheric forcing is the main source of error. Moreover, we did not change the LSM in our MPE, therefore, the uncertainty arising from the LSM was not addressed at all in this study (or in the EURO-CORDEX WRF simulations).

3 Results

This section is organized as follows. In Sect. 3.1, standard temperature and precipitation biases are shown, and their main features are discussed. Then, the comparison with CERES and GLDAS data is given in Sects. 3.2 and 3.3. The physical interpretation of the whole annual bias is difficult, since models usually show very different behaviour depending on the season (García-Díez et al. 2012). Thus, only seasonal biases are considered, focusing in summer and winter. In order to focus on the main results, and to avoid lengthy descriptions, some of the variables are only analysed for summer. Finally, to provide a more general picture, the whole annual cycles (regionally averaged) are compared in Sect. 3.4. The variables considered are summarized in Table 3 and, in the following, they are referred to by their short names.

3.1 Temperature and precipitation bias signatures

WRF shows a large cold bias in winter temperatures appearing in the NE quarter of the domain (Fig. 2), mainly

Table 3 Summary of the variables considered in the study

Short name	Long name	Units
TASMEAN	Daily mean surface air temperature	K
TASMAX	Daily maximum near-surface air temperature	K
TASMIN	Daily minimum near-surface air temperature	K
PR	Precipitation flux	$kg\,m^{-2}\,day^{-1}$
RSDS	Surface downwelling shortwave radiation flux	Wm^{-2}
RLDS	Surface downwelling longwave radiation flux	Wm^{-2}
RLUT	TOA outgoing longwave radiation flux	Wm^{-2}
CLT	Total cloud area fraction	1
ALB	Surface albedo	1
MRSO	Total soil moisture content	$kg\,m^{-2}$

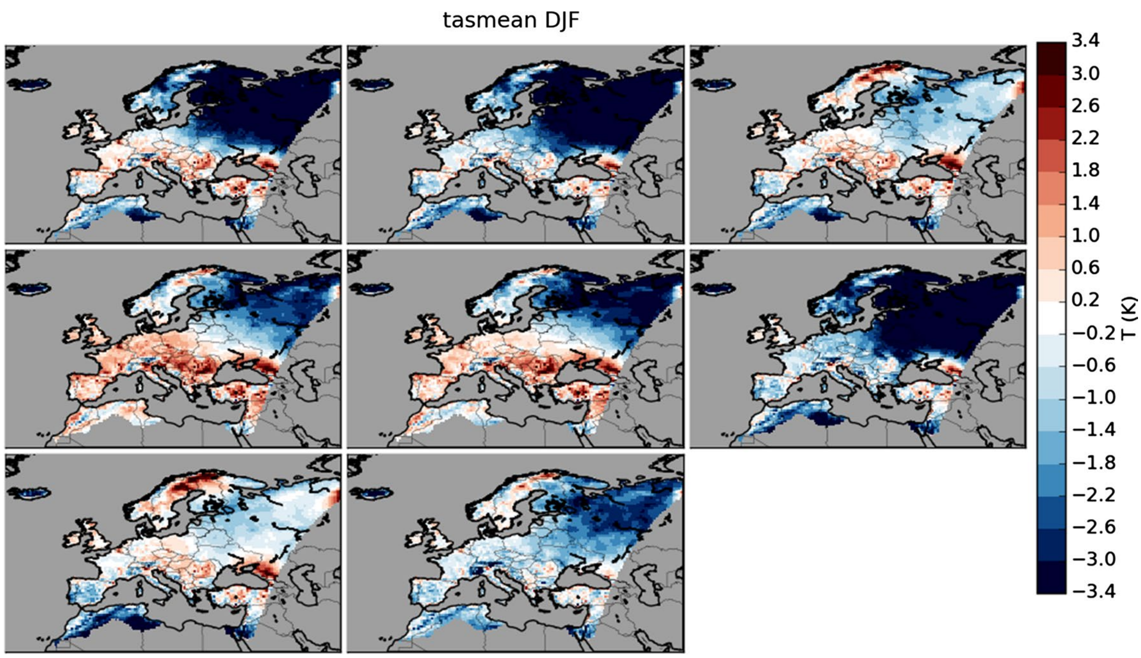

Fig. 2 Bias respect to E-OBS for the daily mean temperature in DJF. The areas with no observed data are painted *grey*

over Russia. This affects all simulations except those using M2M microphysics. It may seem that the larger complexity of M2M is able to overcome this problem, however, we will see below that it is balancing the temperature error with a large bias in cloud cover. This problem with the cloud cover is related to a bug in the code of M2M. A missing term in an equation, related to the cloud ice fall speed, causes the high cirrus clouds to be too persistent.[2] Unfortunately, this problem affects both MPE-H and MPE-M simulations. Arguably, these could be discarded as flawed, however, bugs are present in all computer code (McConnell 2004), and this one was present in the original release of WRF 3.3.1. These simulations have been used due to the interest of their results for the main point of the paper, which is a warning to avoid error compensation when evaluating models. REFOR (the simulation that has been restarted daily from ERA-INTERIM) also shows a cold bias in the NE corner, smaller than its continuous counterpart MPE-G. Therefore, part of the bias develops rapidly after starting the simulation. Other studies have found that the model is too cold over snow covered terrain (Mass 2013), and they have attributed it to a too simple representation of the snow, or to an error in the surface layer parameterization. Waliser et al. (2011) found that a scheme with a multi-layer snow pack (SSiB), recently added to WRF, is able to improve the results thanks to a better representation of the snow ageing and melting processes. Wang et al.

2 http://www.mmm.ucar.edu/wrf/users/wrfv3.3/known-prob-3.3.1.html

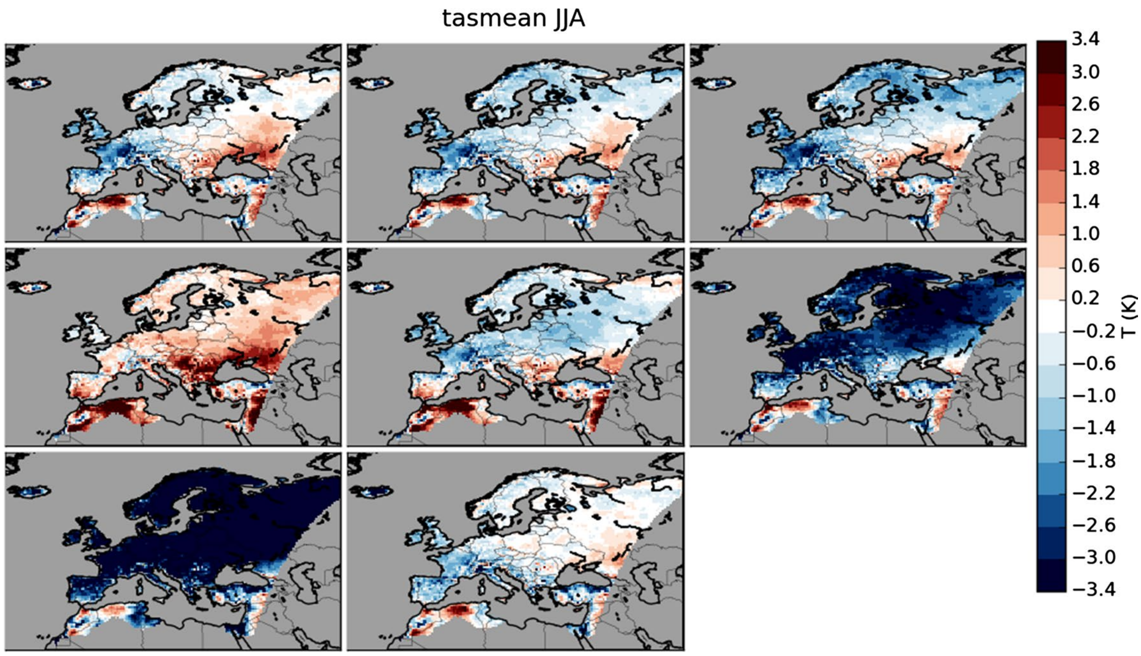

Fig. 3 Bias respect to E-OBS for the daily mean temperature in JJA. The areas with no observed data are painted *grey*

(2010) added some fixes and improvements to the way the Noah scheme represents snow, especially over woodland. These changes are implemented in WRF v3.5, however, a test was carried out with this version, and the cold bias over Russia persisted.

Despite the good behaviour of MPE-M in winter, this configuration is unrealistically cold in spring (not shown) and summer (Fig. 3), reaching 4 K cold biases. On the other hand, during summer, the temperatures of the simulations using KF (MPE-C, MPE-A, MPE-H) are very similar among each other, with small biases over large areas. In contrast, two of the simulations using GD are too cold (MPE-G and specially MPE-M), whereas MPE-F reproduces the observed temperatures well, with errors below ± 1.4 K.

Contrary to winter, summer cold bias in MPE-G does not appear in REFOR. Thus, it needs longer timescales to build up. Regarding the daily cycle, comparison with daily extremes (supplementary information), shows that most of the cold bias in MPE-G and MPE-M is confined to the maximum temperatures.

As we can consider WSM5 and WSM6 microphysics identical, the radiation schemes are the only difference between MPE-F (RRTMG) and MPE-G (CAM3). According to this, RRTMG produces consistently warmer temperatures than CAM. Previous experiments (not shown), showed that BMJ and KF have a similar warming effect in summer temperatures when compared with GD. Thus, despite non-linearities, we can conclude that the combined effect of BMJ and RRTMG made MPE-D the warmest simulation during summer.

WRF significantly overestimates precipitation in summer (Fig. 4) and winter (now shown), especially over the eastern half of the domain, and biases are in general larger than the differences between ensemble members. This overestimation of the observed precipitation is caused by a combination of different factors which are difficult to disentangle. These include: gauge undercatchment (Kotlarski et al. 2014; Frei et al. 2003), overestimation of the frequency of light rain events, biases in the atmospheric circulation, and issues with the parameterizations. For summer precipitation (Fig. 4), unlike for temperature, MPE-M is the configuration with the smallest error.

This is an example of how inappropriate it is to consider a single variable (e.g. precipitation), to evaluate a model, even if this is the variable of interest for a given study. On the other hand, simulations with correct temperatures, as MPE-F, or even too warm, as MPE-D, produce excessive precipitation. It is known that the Kain–Fritsch cumulus scheme overestimates convective precipitation because it does not represent the radiative effect of unresolved cumulus clouds (Herwehe et al. 2014; Alapaty et al. 2012). Other convection schemes, such as BMJ and GD, could be affected by similar problems, and these would be an important contribution to precipitation overestimation in summer.

3.2 Radiation fluxes and cloud cover

In the present section, CERES data are used to evaluate the model radiation balance at the surface and at the top of the atmosphere (TOA). This is also an indirect way to evaluate

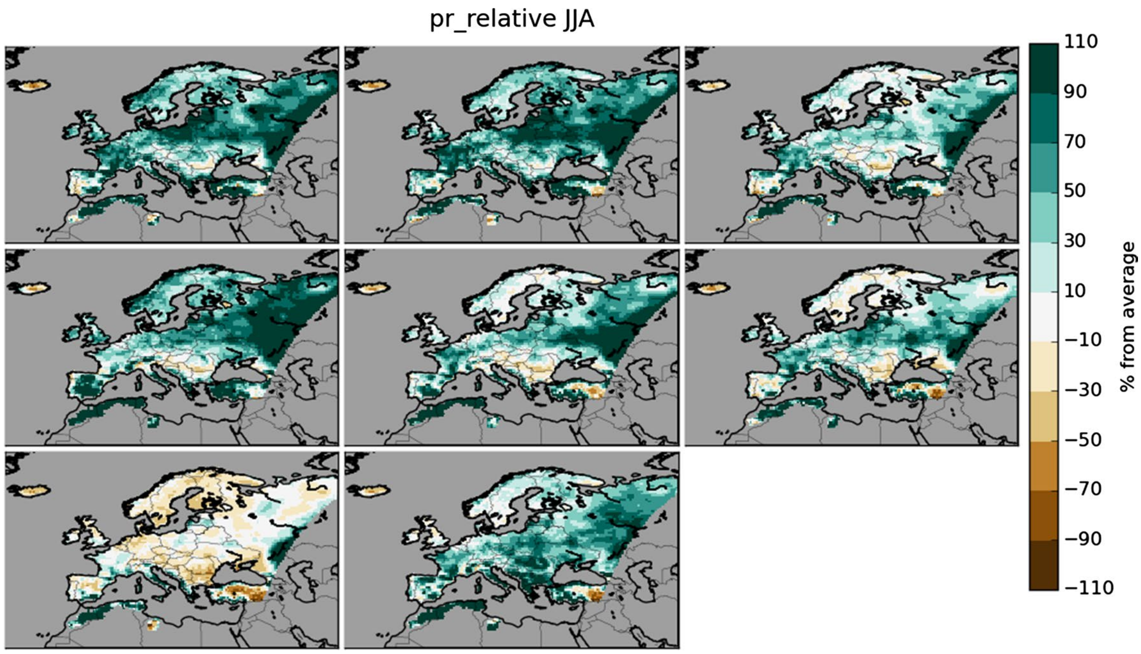

Fig. 4 Relative bias for the daily precipitation in JJA respect to E-OBS. The areas with no observed data are painted *grey*

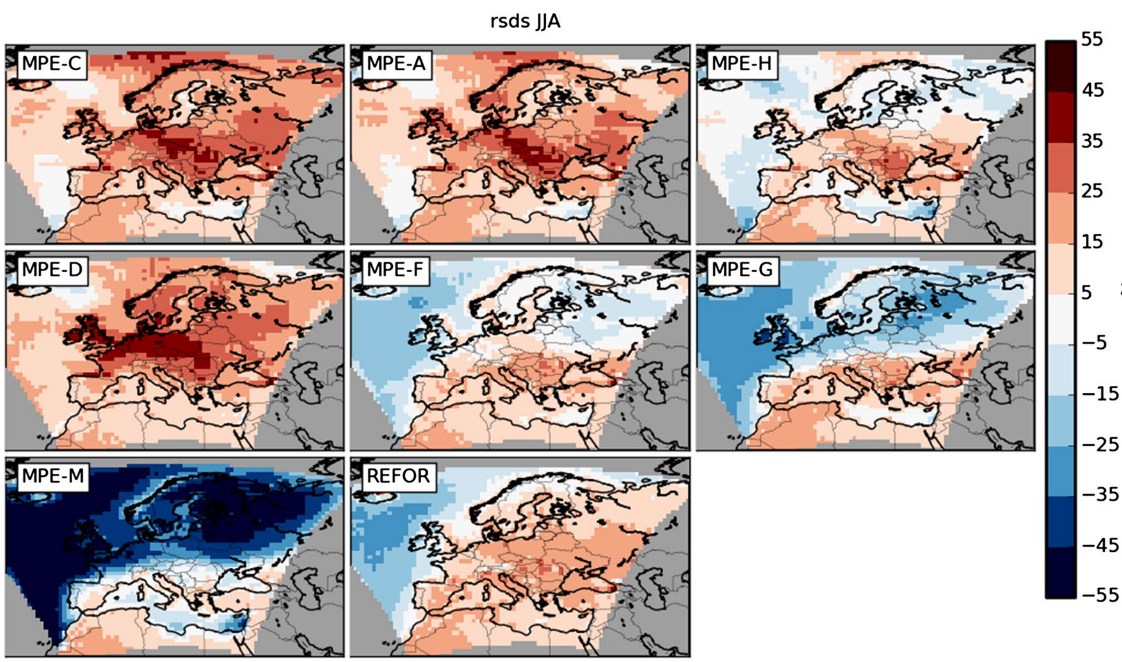

Fig. 5 Relative bias for RSDS in JJA respect to CERES EBAF. The areas with no observed data are painted *grey*

the simulated cloud cover, which is not straightforward to compare with observations (Díaz et al. 2015).

Figure 5 shows the bias for the downward shortwave radiation flux at the surface (RSDS) during the summer. Cold biases in this season for MPE-G and MPE-M (Fig. 3) are correlated with an underestimation of RSDS. However, cold biases persist in areas with no RSDS bias, such as southern France or northern Spain. On the other hand, MPE-C and MPE-A, which produced realistic temperatures over central Europe, overestimate RSDS during summer. Thus, it seems that the model is still too cold with a correct RSDS. In the case of MPE-F, RSDS is close to the observation in most of the domain. Interestingly, the REFOR simulation is fairly different from MPE-G. Thus, again we see

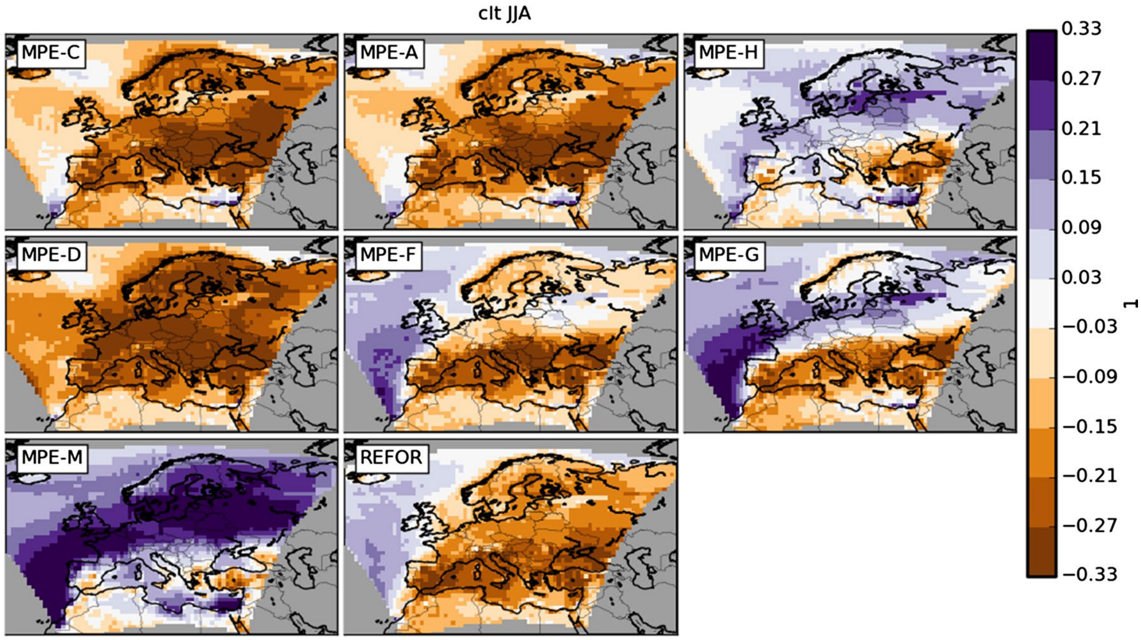

Fig. 6 Bias for total cloud cover in JJA respect to CERES-SYNdef1. The areas with no observed data are painted *grey*

that the bias pattern for RSDS and temperature that appears in MPE-G needs more than 12–24 h to build.

A simple comparison with SYNdef1 total cloud cover (CLT) data has been carried out to check if the results are consistent to those found for the radiation fluxes. To compute CLT in WRF we followed (Sundqvist et al. 1989), which assumes maximum overlapping in each cloud layer and random overlapping between layers. CLT biases in summer (Fig. 6) correlate well with those found for RSDS. The spatial patterns of the summer temperature biases in MPE-G and MPE-M are, thus, related to the cloud cover. Clouds can either warm or cool the surface depending on their altitude and the phase of the diurnal cycle. Also, locally, advection can be as important for temperature as the point energy balance and, in fact, both processes can feed back. This complicates the rigorous attribution of temperature biases to cloud cover. Here, we try to get the most complete possible picture of what is happening into the model by looking at many variables, but analyses to rigorously address causal relationships would require a more systematic approach. Causality in non linear systems is not straightforward to define (Sugihara et al. 2012).

A negative CLT bias appears over the Mediterranean Sea and its surrounding countries in all simulations except MPE-H and MPE-M. This bias is also found when comparing with data from an independent dataset (GEWEX-SRB, see supplementary material). On the other hand, ERA-INTERIM cloud cover is very similar to WRF over this region (not shown). A more detailed analysis would be needed to address whether the problem is in the observations or in the models.

The long wave downward radiation flux (RLDS), depends on the emissivity and temperature of the troposphere and, if clouds are present, on the temperature of the cloud base. Thus, the presence of low-base clouds tends to increase RLDS. During summer, MPE-C, MPE-A and MPE-H underestimate RLDS over most of the domain (Fig. 7). This is consistent with the overestimation of RSDS found, and its relationship to a lack of cloudiness. Interestingly, simulations using the Grell–Devenyi cumulus scheme (MPE-F, MPE-G and MPE-M) show very small biases over the northern half of continental Europe, despite the differences found for RSDS in MPE-G and -M. This apparent inconsistency can be either related to the cooler temperatures of MPE-G and MPE-M and/or to the presence of clouds with different longwave/shortwave transmissivities, such as cirrus clouds. Over the almost cloud-free region in the southern part of the domain (Mediterranean sea and northern Africa), simulations using the RRTMG LW radiation parameterization (MPE-F and MPE-D), display a larger RLDS, which is closer to observation than the rest, which use CAM. MPE-F and MPE-D also tend to be warmer over this area (Fig. 3). As MPE-D and MPE-C display a very similar cloud cover over the whole domain, Fig. 7 shows that, to equal cloudiness, RRTMG produces a larger RLDS and warmer temperatures than CAM.

Finally, the TOA radiation fluxes can also provide information about the clouds. The long wave upward flux at the TOA (RLUT) is a good indicator of the height of cloud

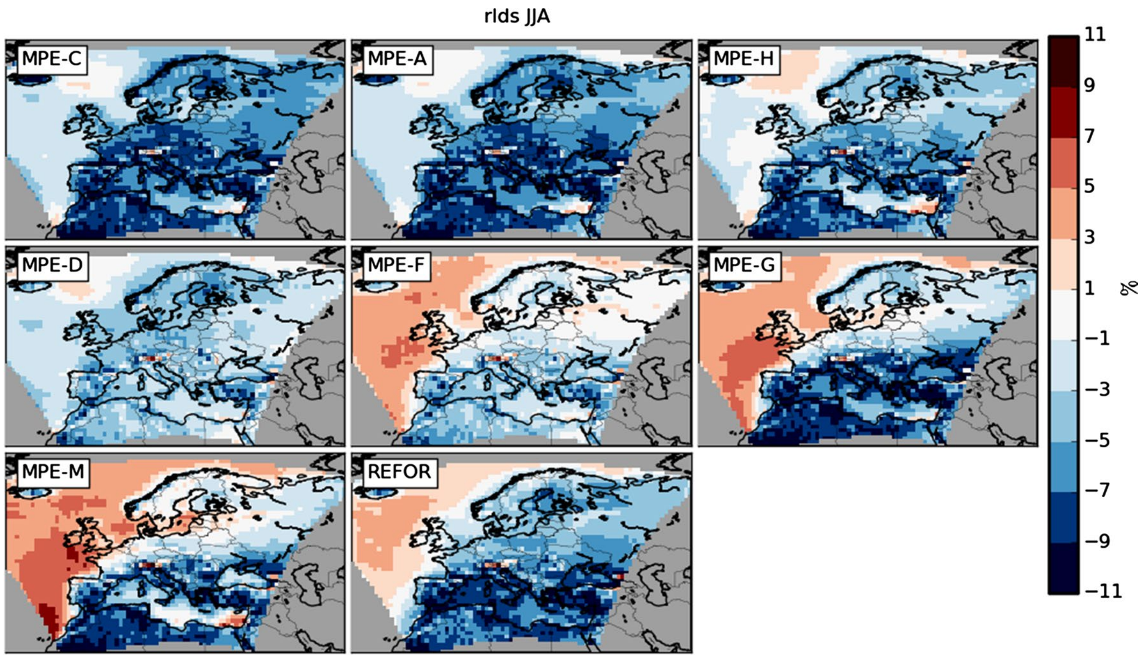

Fig. 7 Relative bias for RLDS in JJA respect to CERES EBAF. The areas with no observed data are painted *grey*

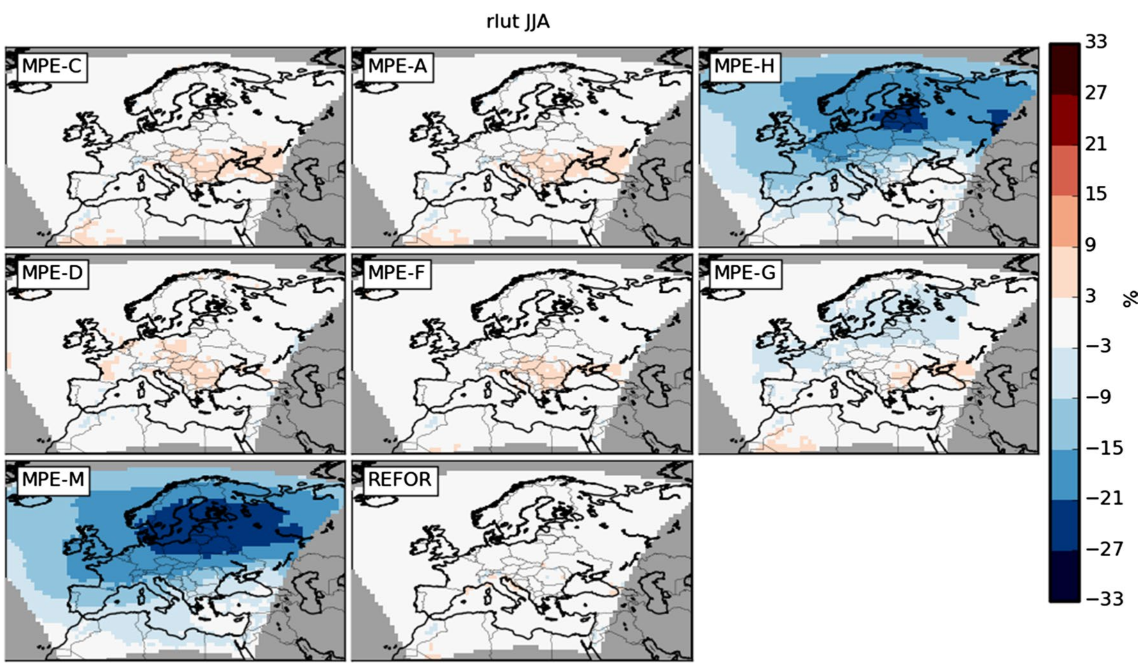

Fig. 8 Relative bias for RLUT in JJA respect to CERES EBAF. The areas with no observed data are painted *grey*

tops. Higher cloud tops are cooler and emit less LW radiation. When the sky is clear, RLUT provides information about surface temperature. RLUT biases is summer (Fig. 8) are small (±3 %) in all simulations, except those using the M2M microphysics, which significantly underestimate it. Thus, this scheme tends to produce too much high cloud cover. As mentioned before (Sect. 3.1), the cause is a bug related to the cloud ice fall speed, which causes high cirrus clouds to be too persistent.

The overestimation of RLUT by the simulations using the M2M microphysics occurs also in winter (not shown). Thus, the persistence of high clouds seems to be the factor

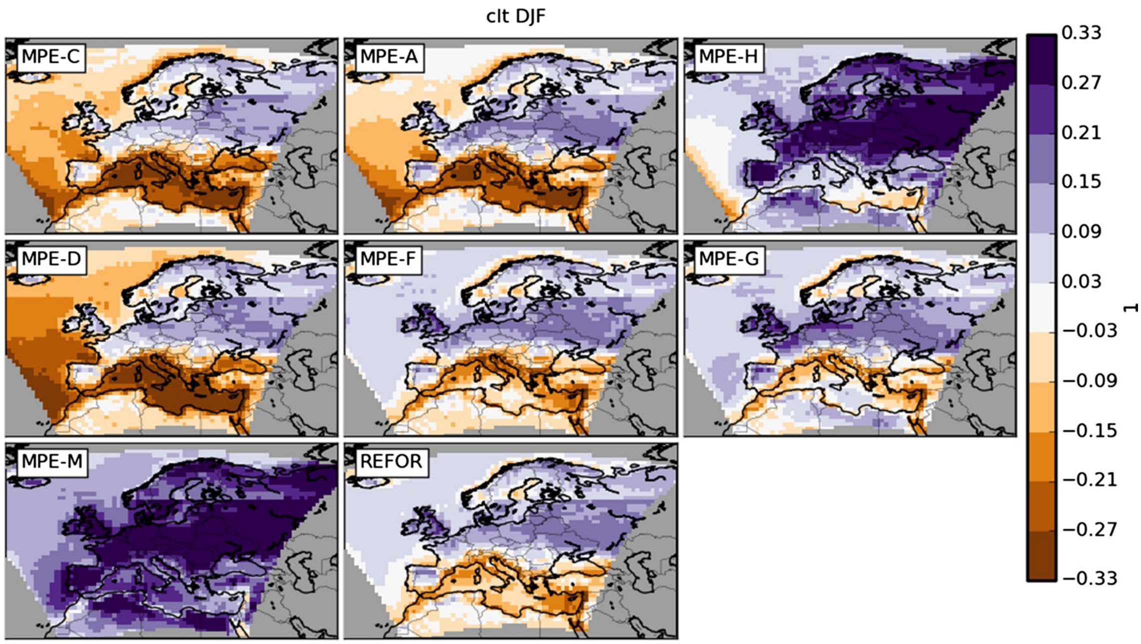

Fig. 9 Relative bias for CLT in DJF respect to CERES SYndef1. The areas with no observed data are painted *grey*. Note that the data are problematic above 60° latitude

compensating the winter cold bias in MPE-H and MPE-M (Fig. 2). This is confirmed by Fig. 9, where the CLT winter bias is shown. MPE-H and MPE-M overestimate winter CLT, which has a net warming effect by blocking the nighttime radiative cooling. Furthermore, the CLT overestimation by MPE-H and MPE-M is known to be affecting especially the high clouds, which are more transparent to SW radiation than to LW, increasing the warming effect. These simulations also overestimate winter RLDS (not shown), confirming this picture. This is a clear case of error compensation, where MPE-H and MPE-M produce seemingly realistic temperatures by fixing the cold bias with an unrealistic high cloud cover.

Furthermore, the relationship between the winter cold bias found in Fig. 2, and the snow cover suggests that the albedo can be playing a role in that problem. Surface albedo is not directly available in CERES data, but it can be estimated by simply dividing RSUS by RSDS. This is not the best estimation, because albedo can be directly observed, but this approach guaranties the consistency with the rest of the analyses. WRF overestimates the albedo (Fig. 10) in the snowy regions (Alps, Eastern Europe and Russia), which correlates with the winter cold bias (Fig. 2). Most of the overestimation is bound to high latitudes, where the observations are uncertain and show an unrealistic discontinuity. However, comparison with GEWEX-SRB data (supplementary material) yields similar results, with an even larger overestimation of the albedo, so this feature is robust. A similar result was found by Xu and Yang (2012) over North America. However, according to Mass (2013), albedo does not seem to be the main cause of the winter cold bias, which is unknown, but likely related to the treatment of the snow pack by the land surface and/or surface layer schemes. Thus, the albedo bias would be a feedback reinforcing too cold temperatures. The presence of the cold bias in REFOR is a measure of the part of this bias which appears immediately after starting from ERA-INTERIM.

3.3 Soil moisture

Total soil moisture content (MRSO) plays a major role in the surface flux partition (Jaeger and Seneviratne 2011). In this section, WRF MRSO is compared with data from the GLDAS reanalysis (see Sect. 2.2). For the sake of brevity, only results for summer will be shown. WRF overestimates MRSO (Fig. 11) in most places, except in the southern part of the domain. The bias is larger in the coldest (driest) simulations (MPE-G and MPE-M), and smaller in the warmest (wettest) (MPE-D). A comparison of the sensible and latent heat fluxes of WRF with GLDAS (not shown) reveals that the latter is too large in the simulations that overestimate RSDS the most (MPE-D, MPE-C, and MPE-A), while the former is correct. Thus, in these simulations the excessive soil moisture is influencing the energy partitioning (Bowen ratio), shifting it to a too large evaporation, and making difficult the occurrence of the "dry regime" where evapotranspiration is limited by soil moisture and not by incoming energy. This is consistent

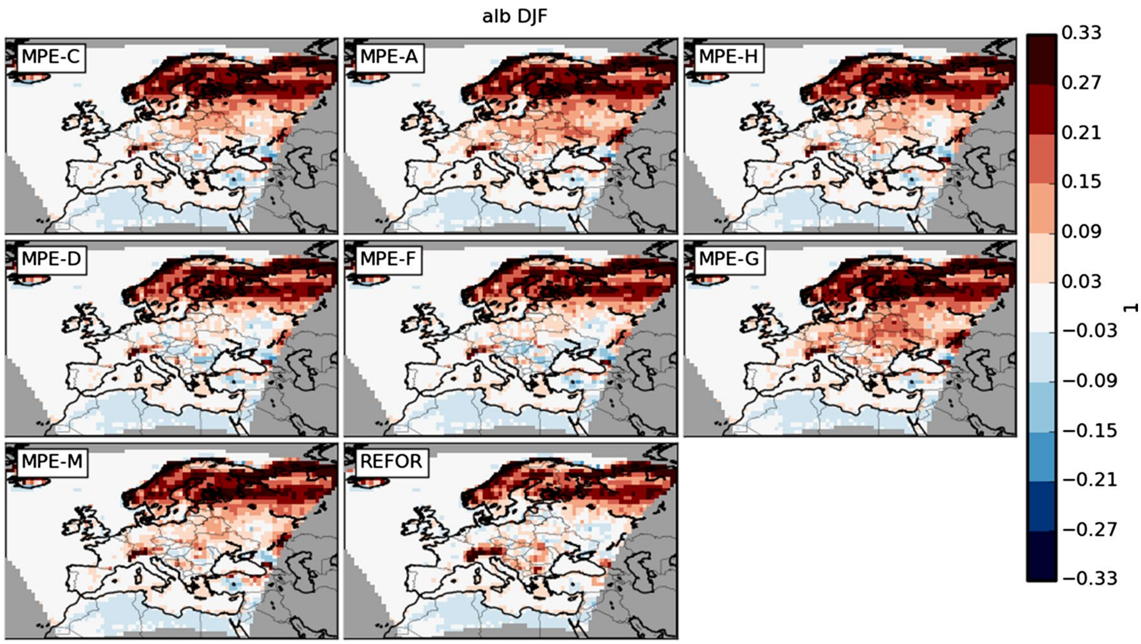

Fig. 10 Surface albedo bias in DJF respect to CERES EBAF. The areas with no observed data are painted *grey*. Note that the data are problematic above 60° latitude

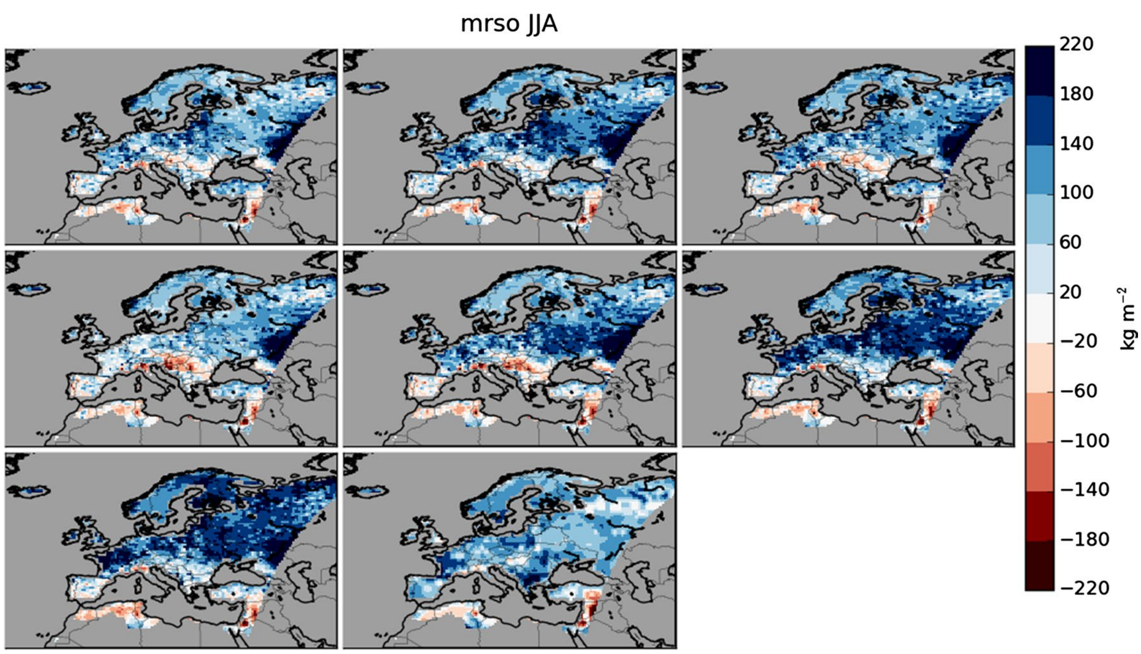

Fig. 11 Total soil moisture content bias in JJA respect to GLDAS2. The areas with no observed data are painted *grey*

with the general difficulty of WRF to simulate extreme heat waves using the NOAH soil scheme (Stegehuis et al. 2014). With a correct precipitation, soil moisture and Bowen ratio, the temperature would likely be higher, due to the too large RSDS. Therefore, this is another example of error compensation.

3.4 Annual cycles

Previous sections have mainly focused in summer and winter. The maps shown show that the spatial autocorrelation is generally large. Thus, spatial averages make sense in most regions. In this section we use spatially-averaged

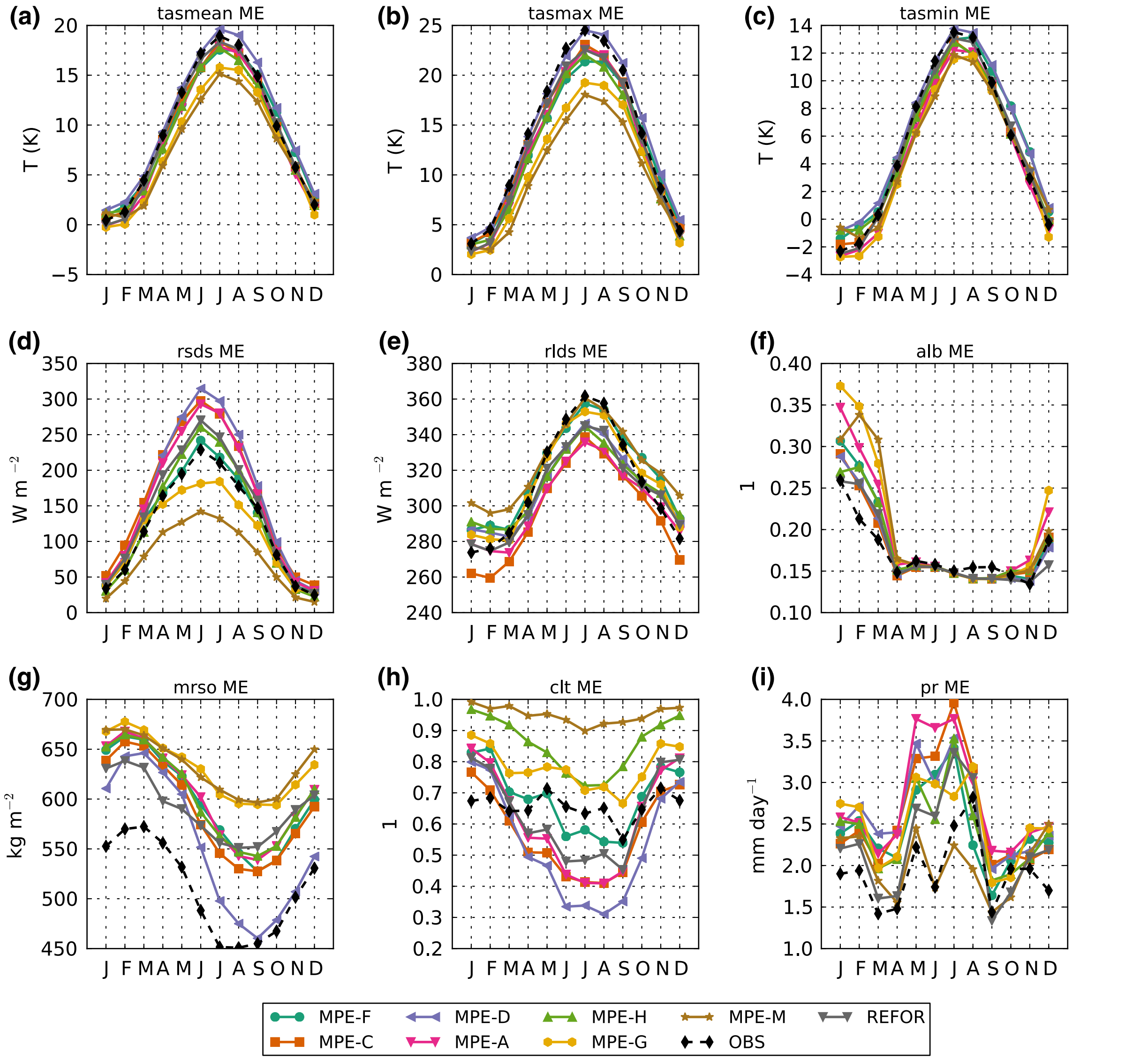

Fig. 12 Mean annual cycles of areal averages over the Middle Europe PRUDENCE region. The *variables plotted* are TASMEAN, TASMIN, RSDS, RLDS, MRSO, ALB, CLT and PR

biases to gain better perspective of the temporal structure of the model error. The regions chosen are the so-called PRUDENCE regions, shown in Fig. 1. We considered monthly time series (available as supplementary material), which show that the relative differences among the MPE members are preserved every year with very few exceptions. Thus, spatially-averaged monthly annual cycles have been computed, following these steps: First, the data from WRF, CERES and GLDAS have been bilinearly interpolated to the E-OBS grid and its land-sea mask has been applied. Second, monthly averages have been computed and, then, spatial averages have been applied. Finally, the annual cycles were obtained by averaging the five values available for each month. As mentioned, the regions were assumed to be homogeneous enough for the areal averages to be meaningful. This assumption holds in most regions for the mean biases (except ME and AL) and differences between ensemble members, but not for the absolute values of the variables. However, the absolute values have been plotted so the relationship between bias, sensitivity to physical parameterizations and the order of magnitude of the variable can be appreciated. These annual cycles summarize many variables in one panel [see Fig. 12 for Middle Europe (ME)]. The variables chosen (see Table 3) are TASMEAN, TASMAX, TASMIN, RSDS, RLDS, MRSO, ALB, CLT and PR.

As analysing all regions would be too extensive and redundant, the analysis is limited to the ME PRUDENCE region. Figure 12 shows many features common to other regions, and summarizes well many of the results found in previous sections. Namely:

- Summer cold bias is larger in TASMAX than in TASMIN (Fig. 12b, c), and is very large for MPE-G and MPE-M, simulations also underestimating RSDS. However, simulations overestimating RSDS still underestimate maximum temperatures during summer (MPE-C, MPE-A). MPE-D is the only simulation with correct TASMAX during summer, and also the only one that removes enough soil moisture to reach GLDAS during that season, although it overestimates RSDS by more than 50 Wm^{-2}.
- As previously seen for summer, WRF tends to generate a too large soil moisture content. This result extends to the whole annual cycle, except for MPE-D during autumn. Colder simulations, MPE-G and MPE-M, have wetter soils during summer.
- Ensemble spread is generally much smaller during the cold season. WRF shows an excessive surface albedo from January to March, larger in MPE-G, MPE-M and MPE-C. Although this region is not affected by the large cold bias found in winter in the NE quarter of the domain, most simulations show cold-biased maximum temperatures during winter and early spring, especially MPE-G and MPE-M. These are also the simulations that overestimate the albedo the most. The effect of an increased albedo in winter and/or cloud cover in summer, related to a moister soil, builds up over time, and illustrates how feedbacks drive these two simulations to different climates. The REFOR simulation, without spin-up, does not show many of the features found in MPE-G.
- Despite the large overestimation, WRF reproduces the shape of the observed precipitation annual cycle. The wet bias is more pronounced in the maxima on May and July, reaching large values of more than 1 $\mathrm{kg\,m^{-2}\,day^{-1}}$.
- Regarding total cloud cover, the ensemble spread is very large. As in the radiation fluxes, MPE-F is the configuration producing the most realistic cloud cover. Despite that in general it is not possible to find a better configuration in this kind of experiments, in this case, MPE-F is clearly outperforming the other configurations. Probably this would not be the case with a larger ensemble size.

4 Discussion and conclusions

With the aim of improving the understanding of the physical realism of a RCM, a new MPE over Europe has been produced and evaluated. The evaluation has been carried out using many variables in addition to the standard P&T. Namely, radiation fluxes, total cloud cover, surface albedo, and soil moisture have been compared with CERES observations and with GLDAS soil reanalysis, respectively. This enabled us to see how the errors of the different variables can sometimes compensate each other, so focusing on one or few variables can be misleading. The approach followed also revealed correlations between the biases of different variables, helping to identify the processes that are behind them. However, as temperature, precipitation, clouds and soil moisture are non-linearly coupled (Seneviratne et al. 2010), drawing strong conclusions about causal relationships behind the biases remains as a challenge, and caution must be taken with the interpretation of the results. This paper does not provide a systematic method to deal with non-linearities and identify the causes of the biases, but addresses the need of publishing more thorough evaluations of the models. This is needed to boost the reliability of the models and to enable fair ensemble weighting methodologies.

One of the goals of the paper is to analyse the sensitivity to physical parameterizations schemes. With this aim, we used three cumulus schemes (KF, BMJ and GD), two radiation schemes (RRTMG and CAM) and four microphysics schemes (WSM3, WSM5, WSM6 and M2M, though WSM5 and WSM6 are very similar). We found that RRTMG is generally warmer than CAM thanks to an enhanced downward long wave radiation flux. The slightly different CO_2 concentration prescribed in these schemes cannot explain the difference, as in fact it is larger in CAM (370 ppm in RRTMG versus yearly IPCC AR4 A2 scenario, in the range of 374–383 ppm, in CAM). Thus, there must be another cause, which remains unknown. During summer, we found that the sensitivity to the cumulus parameterization is large. When changing the cumulus scheme in MPE-G from GD to KF, most of the cold bias is removed. However, we found that this is due to a overestimated downward short wave radiation flux. The excess of energy received by the surface in these simulations is compensated by a too low downward long wave flux and a too large evaporation. Regarding the microphysics, differences between WSM-3, WSM5 and WSM-6 are generally small. Thanks to its increased cloud cover, M2M is much cooler in summer and warmer in winter over snow-covered areas, compensating the pronounced cold bias found over these. This cannot be attributed to the formulation of M2M itself but to the presence of a bug in M2M code which makes cirrus clouds too persistent. Multi-physics spread is large, comparable to that of a multi-model, which is consistent with other results in the literature (Jerez et al. 2013). Spread is found to be remarkably large for both RSDS and CLT, especially during summer. These two variables are key to explain the spread, which is reasonable given that the schemes changed are those more closely related with clouds and radiation. Other studies evaluating RCMs radiative fluxes (Kothe and Ahrens 2010; Kothe et al. 2011; Pessacg et al. 2013; Samuelsson et al. 2011) found similar relationships among the biases of the different variables, and similar error magnitudes (or even larger, in the case of tropical regions in (Pessacg et al. 2013)). They also found cases in which biases compensate each other to result in small errors in some variables for the wrong reasons.

Another goal was to analyse the main model deficiencies. Two main biases have been identified in the model. One is the overestimation of precipitation, which occurs in almost all seasons and ensemble members, except MPE-M, and especially in the eastern half of the domain. This problem is currently affecting most of the RCMs (Kotlarski et al. 2014). During summer, it is partly related to problems in adapting some aspects of the models to the resolution, namely the cumulus parameterization, which are being investigated (Alapaty et al. 2012; Tripathi and Dominguez 2013). The other causes behind this bias are difficult to measure, but probably include errors in the atmospheric circulation, gauge undercatchment, and other factors. The comparison with GLDAS data revealed also that WRF overestimates soil moisture content in most regions.

The second bias is the pronounced cold bias appearing in the NE quarter of the domain during winter and spring. We found that, during those seasons, the differences in albedo correlated well with the temperature bias. Thus, the cold bias is partially related to this albedo overestimation. This was also found by Xu and Yang (2012) in other region (U.S.-Canada). Other authors (Mass 2013) found that the cold bias appears systematically over snow-covered regions, regardless of the albedo. Evidence suggests that the problem is a too crude representation of the snow pack or either some problem in the computation of the skin temperature or ground heat flux. The albedo would be acting as a feedback and not as the main cause of the bias.

In general, WRF shows results comparable to other models (Mearns et al. 2012; Kotlarski et al. 2014), although the winter cold bias causes great deviations from observed temperatures not found in other models. The simulation labelled as MPE-M is also too unrealistic. Some of the biases found can also be spotted in the WRF members used in the EURO-CORDEX evaluation work (Kotlarski et al. 2014) covering a 20-year period and including other configuration differences apart from the physics options. Namely the winter cold bias in the NE quarter is present in the IPSL-INERIS and CRP-GL simulations, which use the physics of MPE-F and MPE-A. In the UHOH simulation, equivalent to MPE-H, winter temperatures are realistic in this region but, as shown in the present work, this is due to compensation of errors by a too large cloud cover.

The most important conclusion of this work are not the particular results for the WRF configurations tested (although these are valuable for the WRF community), but showing how, when abandoning the limited perspective of P&T, a rich and complex picture emerges, where the good model performance in some variables is sometimes related to compensation of errors and not to improved realism. The simulation labelled as MPE-M is an extreme example of this behaviour. It was found to produce realistic winter temperatures thanks to a wrong cloud cover, and a realistic summer precipitation thanks to a very large cold bias. Thus, we strongly encourage the regional climate modeling community to use as many variables as possible in model evaluation, and in the weighting process of ensembles.

As a final remark, results suggest that surface temperature is not well suited to assess the overall realism of a simulation. Some studies (Giorgi and Coppola 2010) found no clear relationship between temperature bias on evaluation runs and climate change amplitude or sign. The present work shows how cloud cover, radiation fluxes and soil moisture are key variables to show that the model is producing a realistic present time simulation.

Acknowledgments This work was partially supported by Projects EXTREMBLES (CGL2010-21869) and CORWES (CGL2010-22158-C02), funded by the Spanish R&D Programme. WRF4G (CGL2011-28864) provided the framework to run the model; this Spanish R&D project is co-funded by the European Regional Development Fund (ERDF). Partial support from the 7th European Framework Programme (FP7) through Grant 308291 (EUPORIAS) is also acknowledged. The E-OBS dataset was produced within the EU-FP6 project ENSEMBLES http://www.ensembles-eu.org and the data was provided through the ECA&D Project http://eca.knmi.nl. Radiation data from CERES were obtained from the NASA Langley Research Center Atmospheric Science Data Center. GLDAS data were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

References

Alapaty K, Herwehe JA, Otte TL, Nolte CG, Bullock OR, Mallard MS, Kain JS, Dudhia J (2012) Introducing subgrid-scale cloud feedbacks to radiation for regional meteorological and climate modeling. Geophys Res Lett. doi:10.1029/2012GL054031

Argüeso D, Hidalgo-Muñoz J, Gámiz-Fortis S, Esteban-Parra M, Dudhia J, Castro-Díez Y (2011) Evaluation of WRF parameterizations for climate studies over Southern Spain using a multi-step regionalization. J Clim. doi:10.1175/JCLI-D-11-00073.1

Awan N, Truhetz H, Gobiet A (2011) Parameterization induced error-characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble based analysis. J Clim 24(12):3107–3123. doi:10.1175/2011JCLI3674.1

Chen F, Dudhia J (2001) Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon Weather Rev

129(4):569–585. doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Christensen J, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. Clim Res 44(2–3):179–194. doi:10.3354/cr00916

Collins WD, Rasch PJ, Boville BA, Hack JJ, McCaa JR, Williamson DL, Kiehl JT, Briegleb B, Bitz C, Lin S et al. (2004) Description of the NCAR community atmosphere model (CAM 3.0). NCAR Tech Note NCAR/TN-464+ STR

Dee D, Uppala S, Simmons A, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M, Balsamo G, Bauer P et al (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

Dharssi I, Bovis KJ, Macpherson B, Jones CP (2011) Operational assimilation of ASCAT surface soil wetness at the Met Office. Hydrol Earth Syst Sci 15(8):2729–2746. doi:10.5194/hess-15-2729-2011

Díaz JP, González A, Expósito FJ, Pérez JC, Fernández J, García-Díez M, Taima D (2015) WRF multi-physics simulation of clouds in the African region. Q J R Meteorol Soc (submitted)

Fernández J, Montávez J, Sáenz J, González-Rouco J, Zorita E (2007) Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: annual cycle. J Geophys Res 112(D4):101. doi:10.1029/2005JD006649

Fernández-Quiruelas V, Fita L, Fernández J, Cofiño A (2010) WRF workflow on the Grid with WRF4G. In: 11th WRF users' workshop. Boulder (CO), USA

Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL (2003) Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. J Geophys Res Atmos 108(D3):4124. doi:10.1029/2002JD002287

García-Díez M, Fernández J, Fita L, Yagüe C (2012) Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. Q J R Meteorol Soc. doi:10.1002/qj.1976

Giorgi F, Coppola E (2010) Does the model regional bias affect the projected regional climate change? An analysis of global model projections. Clim Change 100(3–4):787–795. doi:10.1007/s10584-010-9864-z

Giorgi F, Jones C, Asrar G et al (2009) Addressing climate information needs at the regional level: the CORDEX framework. World Meteorol Organ (WMO) Bull 58(3):175

Grell G, Devenyi D (2002) A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophys Res Lett 29(14):38-1. doi:10.1029/2002GL015311

Grell GA, Dudhia J, Stauffer DR (1995) A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Technical Note

Greve P, Warrach-Sagi K, Wulfmeyer V (2013) Evaluating soil water content in a WRF-Noah downscaling experiment. J Appl Meteorol Climatol 52(10):2312–2327. doi:10.1175/JAMC-D-12-0239.1

Gupta Stackhouse PW, Mikovitz JC, Cox SJ, Zhang T (2006) Surface radiation budget project completes 22-year data set. GEWEX WCRP News 14(4):12–13

Haylock M, Hofstra N, Tank A, Klok E, Jones P, 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, Fernández J, Gutiérrez J (2010) Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain. J Geophys Res 115(D21):117. doi:10.1029/2010JD013936

Herwehe JA, Alapaty K, Spero TL, Nolte CG (2014) Increasing the credibility of regional climate simulations by introducing subgrid-scale cloud-radiation interactions: RCM sims with Cu-radiation interactions. J Geophys Res Atmos 119(9):5317–5330. doi:10.1002/2014JD021504

Hofstra N, Haylock M, New M, Jones PD (2009) Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. J Geophys Res. doi:10.1029/2009JD011799

Hong S, Dudhia J, Chen S (2004) A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon Weather Rev 132(1):103–120

Hong S, Noh Y, Dudhia J (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev 134(9):2318–2341

Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. J Geophys Res Atmos. doi:10.1029/2008JD009944

Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, Braun A, Colette A, Déqué M, Georgievski G, others, (2013) EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg Environ Change. doi:10.1007/s10113-013-0499-2

Jaeger EB, Seneviratne SI (2011) Impact of soil moisture-atmosphere coupling on European climate extremes and trends in a regional climate model. Clim Dyn 36(9–10):1919–1939. doi:10.1007/s00382-010-0780-8

Janjic ZI (2000) Comments on "Development and evaluation of a convection scheme for use in climate models". J Atmos Sci 57(21):3686–3686. doi:10.1175/1520-0469(2000)057<3686:CODAEO>2.0.CO;2

Jerez S, Montavez JP, Jimenez-Guerrero P, Gomez-Navarro JJ, Lorente-Plazas R, Zorita E (2013) A multi-physics ensemble of present-day climate regional simulations over the Iberian Peninsula. Clim Dyn 40(11–12):3023–3046. doi:10.1007/s00382-012-1539-1

Kain JS (2004) The Kain–Fritsch convective parameterization: an update. J Appl Meteorol 43(1):170–181. doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2

Kothe S, Ahrens B (2010) On the radiation budget in regional climate simulations for West Africa. J Geophys Res Atmos. doi:10.1029/2010JD014331

Kothe S, Dobler A, Beck A, Ahrens B (2011) The radiation budget in a regional climate model. Clim Dyn 36(5–6):1023–1036. doi:10.1007/s00382-009-0733-2

Kotlarski S, Keuler K, Christensen OB, Colette A, Déqué M, Gobiet A, Goergen K, Jacob D, Lüthi D, van Meijgaard E, Nikulin G, Schär C, Teichmann C, Vautard R, Warrach-Sagi K, Wulfmeyer V (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7(4):1297–1333. doi:10.5194/gmd-7-1297-2014

Kysely J, Plavcova E (2010) A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. J Geophys Res Atmos. doi:10.1029/2010JD014123

Loeb NG, Lyman JM, Johnson GC, Allan RP, Doelling DR, Wong T, Soden BJ, Stephens GL (2012) Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. Nat Geosci 5(2):110–113. doi:10.1038/ngeo1375

Mass C (2013) Strange linear features in WRF clouds and precipitation: diagnosis and correction. In: 14th Annual WRF users' workshop, Boulder, Colorado

Mauritsen T, Stevens B, Roeckner E, Crueger T, Esch M, Giorgetta M, Haak H, Jungclaus J, Klocke D, Matei D, Mikolajewicz U, Notz D, Pincus R, Schmidt H, Tomassini L (2012) Tuning the climate of a global model. J Adv Model Earth Syst. doi:10.1029/2012MS000154

McConnell S (2004) Code complete. O'Reilly Media Inc., Sebastopol

Mearns LO, Arritt R, Biner S, Bukovsky MS, McGinnis S, Sain S, Caya D, Correia J Jr, Flory D, Gutowski W, Gutowski W et al

(2012) The North American regional climate change assessment program: overview of phase I results. Bull Am Meteorol Soc 93(9):1337–1362. doi:10.1175/BAMS-D-11-00223.1

Menéndez M, García-Díez M, Fita L, Fernández J, Méndez FJ, Gutiérrez JM (2014) High-resolution sea wind hindcasts over the Mediterranean area. Clim Dyn 42(7–8):1857–1872. doi:10.1007/s00382-013-1912-8

Mlawer E, Taubman S, Brown P, Iacono M, Clough S (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res 102(D14):16,663–16,716. doi:10.1029/97JD00237

Mooney P, Mulligan F, Fealy R (2013) Evaluation of the sensitivity of the weather research and forecasting model to parameterization schemes for regional climates of Europe over the period 1990–1995. J Clim 26(3):1002–1017. doi:10.1175/JCLI-D-11-00676.1

Morrison H, Thompson G, Tatarskii V (2009) Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one-and two-moment schemes. Mon Weather Rev 137(3):991–1007. doi:10.1175/2008MWR2556.1

Pessacg NL, Solman SA, Samuelsson P, Sanchez E, Marengo J, Li L, Remedio ARC, Rocha RPd, Mourão C, Jacob D (2013) The surface radiation budget over South America in a set of regional climate models from the CLARIS-LPB project. Clim Dyn. doi:10.1007/s00382-013-1916-4

Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng CJ, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. Bull Am Meteorol Soc 85(3):381–394. doi:10.1175/BAMS-85-3-381

Samuelsson P, Jones CG, Willén U, Ullerstig A, Gollvik S, Hansson U, Jansson C, Kjellström E, Nikulin G, Wyser K (2011) The Rossby Centre regional climate model RCA3: model description and performance. Tellus A 63(1):4–23

Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ (2010) Investigating soil moisture-climate interactions in a changing climate: a review. Earth Sci Rev 99(3–4):125–161. doi:10.1016/j.earscirev.2010.02.004

Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, Wang W, Powers J (2008) A description of the advanced research WRF version 3. Tech. rep, NCAR

Stegehuis AI, Vautard R, Ciais P, Teuling AJ, Miralles DG, Wild M (2014) An observation-constrained multi-physics RCM ensemble for simulating European mega-heatwaves. Geosci Model Dev Discuss 7(6):7861–7886. doi:10.5194/gmdd-7-7861-2014

Sugihara G, May R, Ye H, Hsieh Ch, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338(6106):496–500. doi:10.1126/science.1227079

Sundqvist H, Berge E, Kristjánsson JE (1989) Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. Mon Weather Rev 117(8):1641–1657. doi:10.1175/1520-0493(1989)117<1641:CACPSW>2.0.CO;2

Tripathi OP, Dominguez F (2013) Effects of spatial resolution in the simulation of daily and subdaily precipitation in the southwestern US. J Geophys Res Atmos 118(14):7591–7605. doi:10.1002/jgrd.50590

Vautard R, Gobiet A, Jacob D, Belda M, Colette A, Déqué M, Fernández J, García-Díez M, Goergen K, Güttler I, others (2013) The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. Clim Dyn. doi:10.1007/s00382-013-1714-z

Waliser D, Kim J, Xue Y, Chao Y, Eldering A, Fovell R, Hall A, Li Q, Liou KN, McWilliams J, Kapnick S, Vasic R, Sale FD, Yu Y (2011) Simulating cold season snowpack: impacts of snow albedo and multi-layer snow physics. Clim Change 109(1):95–117. doi:10.1007/s10584-011-0312-5

Wang Z, Zeng X, Decker M (2010) Improving snow processes in the Noah land model. J Geophys Res Atmos. doi:10.1029/2009JD013761

Watanabe M, Shiogama H, Yokohata T, Kamae Y, Yoshimori M, Ogura T, Annan JD, Hargreaves JC, Emori S, Kimoto M (2012) Using a multiphysics ensemble for exploring diversity in cloud-shortwave feedback in GCMs. J Clim 25(15):5416–5431. doi:10.1175/JCLI-D-11-00564.1

Xu Z, Yang ZL (2012) An improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate simulations. J Clim 25(18):6271–6286. doi:10.1175/JCLI-D-12-00005.1