Physically based evaluation of climate models over the Iberian Peninsula

Ma. del Carmen Sánchez de Cos Escuín · Jose Ma. Sánchez-Laulhé · Carlos Jiménez Alonso · Juan Manuel Sancho Ávila · Ernesto Rodríguez-Camino

Received: 8 February 2012 / Accepted: 29 November 2012

Abstract A novel approach is proposed for evaluating regional climate models based on the comparison of empirical relationships among model outcome variables. The approach is actually a quantitative adaptation of the method for evaluating global climate models proposed by Betts (Bull Am Meteorol Soc 85:1673–1688, 2004). Three selected relationships among different magnitudes involved in water and energy land surface budgets are firstly established using daily re-analysis data. The selected relationships are obtained for an area encompassing two river basins in the southern Iberian Peninsula corresponding to 2 months, representative of dry and wet seasons. The same corresponding relations are also computed for each of the thirteen regional simulations of the ENSEMBLES project over the same area. The usage of a metric based on the Hellinger coefficient allows a quantitative estimation of how well models are performing in simulating the relations among surface magnitudes. Finally, a series of six rankings of the thirteen regional climate models participating in the ENSEMBLES project is obtained based on their ability to simulate such surface processes.

Keywords Climate models · Evaluation

Ma. del Carmen Sánchez de Cos Escuín ·
J. Ma. Sánchez-Laulhé · C. J. Alonso · J. M. S. Ávila
AEMET, Centro Meteorológico de Málaga, Demóstenes 4,
29010 Málaga, Spain

E. Rodríguez-Camino (✉)
AEMET, Servicios Centrales, Leonardo Prieto Castro 8,
28040 Madrid, Spain
e-mail: erodriguezc@aemet.es

1 Introduction

Climate models are numerical representations of the climate system based on the physical, chemical, and biological properties of its components, their interactions and feedback processes. Different climate models constitute multiple realizations of the climate system based on computer programs. Climate models differentiate among them by the approximations and discretizations used to solve the mathematical equations representing its physics, chemistry and biology. Although climate models continue to have significant limitations which lead to uncertainties in the magnitude and timing, as well as regional details, they have consistently provided a robust and unambiguous picture of the climate system. There is currently a considerable confidence in the simulations provided by climate models due to the fact that model principles are based on well established physical laws, such as conservation of mass, energy and momentum. An additional source of confidence is their ability to simulate important aspects of the current and past climates, as well as their changes (Randall et al. 2007).

The climate system includes a variety of physical processes, such as cloud processes, radiative processes and boundary-layer processes, which interact with each other on many temporal and spatial scales. Due to the limited resolutions of the models, many of these processes are not resolved adequately by the model grid and must therefore be parameterized. As confidence in global models decreases at smaller scales, higher resolution regional climate models (RCMs) provide quantitative value to climate simulations. With finer resolution, mesoscale phenomena, contributing e.g. to intense precipitation, and coupling between regional circulations and convection can be resolved. Higher resolution RCMs also include other types of scale-dependent variability such as extreme winds and locally extreme temperature that coarse-resolution global models will smooth. Regional-scale simulations also have phenomenological value, being able to represent processes that global models either cannot resolve or can resolve only poorly (CCSP 2008).

As climate models are very complex systems, they have different capabilities and limitations which can be evaluated using a variety of methods and approaches. Models can be tested either globally at the system-level or at component-level. Whereas system-level evaluation is focused on the outputs of the full model, component-level evaluation isolates particular components of the model (e.g. atmosphere, ocean, land surface, etc.) or even subcomponents (e.g., numerical methods, parameterizations of different physical processes, etc.,) to test them independently of the complete model. A hybrid approach consists of evaluating the whole system but putting the focus on some specific process or component. For example, we may be interested in exploring how well climate models are able to simulate surface processes or interaction between land and atmosphere (Randall et al. 2007).

A number of metrics have been designed to compare quantitatively climate model simulations against past or current observed climates. Although many different metrics of model reliability have been proposed (see, e.g., Gleckler et al. 2008) there is at present little consensus on a particular metric to discriminate ''good'' and ''bad'' models. In fact, the main issue is the virtually infinite number of metrics that can be defined, being each of them appropriate for different purposes (Knutti et al. 2010). Land-surface processes and interaction between land-surface and atmosphere are especially relevant for the evaluation of climate models simulations as they are very much responsible for precipitation and surface temperature, which traditionally have been used to define local climate. The performance of a climate model when simulating the interaction between land-surface and atmosphere depends critically on the correct coupling between land-surface fluxes and state variables (e.g., evapotranspiration, sensible heat flux, radiative fluxes, soil moisture, etc.). Some researchers (e.g., Betts 2004, 2007; Betts et al. 2006; Jaeger et al. 2009; Santanello et al. 2009; Seneviratne et al. 2010) have pointed out that an alternative way to identify coupling between related variables is to derive empirical relationships by displaying the investigated variables as a function of one another. These relationships can only be suggestive of coupling mechanisms at the land–atmosphere interface without pointing to any direction of causality. As these relationships can be derived for both observations and model data, they are also of strong relevance for model evaluation. We extend in this paper the method for evaluating global climate models proposed by Betts (2004) to RCMs including as main novelties, first, the quantification—by introducing the Hellinger distance—of how well different pairs of empirical relationships are represented by models and, second, the usage of such metric to evaluate and rank models according to accuracy of their simulation of atmosphere/land surface coupling.

In recent years a large number of RCM simulations have been produced for simulating the future European climate (e.g. Christensen and Christensen 2007; Déqué et al. 2005, 2007; van der Linden and Mitchell 2009). As indicated by Kjellström and Giorgi (2010), a relevant finding in these multi-model experiments is that climate change scenarios with different RCMs can differ significantly, even if the lateral boundary conditions are taken from the same global climate model. Therefore, an additional level of uncertainty to the total uncertainty is added by the downscaling process associated to regional climate change simulations. In order to explore such uncertainties, it is reasonable to make use of multi-model ensembles of RCMs for deriving detailed climate change information at the regional scale. It can even be envisaged the application of some kind of performance-based weighting schemes in the process of combining multi-model results, to increase the reliability of the projections (Giorgi and Mearns 2002). In the European project ENSEMBLES (van der Linden and Mitchell 2009), a work package was devoted to designing and testing a weighting system for a multi-model ensemble of RCMs. Kjellström and Giorgi (2010) have described the set of metrics derived in the framework of the ENSEMBLES project to combine RCMs simulations based on their performance and aiming at the production of probabilistic climate change projections (see also Climate Research, Special Issue No 23 2010 on 'Regional Climate Model evaluation and weighting'). Christensen et al. (2010) have explored six metrics designed to capture different aspects of RCM performance in reproducing large-scale circulation patterns, meso-scale signals, daily temperature and precipitation distributions and extremes, trends and the annual cycle. Most of their explored metrics were based on the performance of different aspects of temperature and precipitation fields but none of them relied on the correctness of physical processes simulations.

Within this frame our method proposes an evaluation of the interaction between land and atmosphere simulated by regional climate models as a complement to the above described methods to measure the performance of RCMs. The method here described characterizes the differences or distances of two 2D-scattered plots describing the empirical relationship linking pairs of land surface variables by making use of the Hellinger coefficient (Cramer 1946). The Hellinger coefficient—initially introduced in probability and statistics theories to measure the closeness of two probability distribution functions—will therefore allow us to quantify how close the same empirical relation obtained

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

from a climate model simulation and from observation are. In order to compare the here proposed method of evaluation based on the interaction between land and atmosphere with the six metrics proposed by Christensen et al. (2010), we have computed the Hellinger coefficient for the pair temperature and precipitation (T2m−PP) and also standard scores for temperature and precipitation.

ERA-Interim re-analysis (Dee et al. 2011) has been used as a proxy of actual observations for the selected surface magnitudes due to the lack of spatial coverage of observations for most of the fluxes and surface variables considered here. Direct measures of fluxes and surface/soil variables are frequently restricted to a few reference observatories or recent satellite measurements. Data assimilation algorithms provide a full and consistent 3D representation of the atmosphere constrained by the available observations and physical relationships among variables describing the state of the atmosphere, The four-dimensional variational data assimilation used in ERA-Interim includes, apart of the relationships of the forecast model, those of the complex statistical balance between the first guess error variables. We are fully aware that fluxes—and certain variables not directly observed-provided by a re-analysis are very much dependent on the constraints imposed by the data assimilation algorithm and the underlying model. Variables not directly observed are mainly produced by the underlying forecasting model. In fact, it may happen that fluxes and non-analysed soil/surface variables show bias attributable to the inaccuracies of the assimilation procedure. Therefore, before using re-analysis data as reference or ground-truth some efforts must be devoted to verify this assumption for the variables, region and seasons selected. Nevertheless, it should be stressed that this paper focuses on the proposed method to evaluate model outputs based on empirical relationship linking pairs of surface relevant magnitudes and not on a comprehensive validation of the reference.

Once the selected relationships have been determined for the ERA-Interim re-analysis data, the corresponding relationships are also determined for each of the thirteen regional simulations of the ENSEMBLES project (van der Linden and Mitchell 2009) using daily data over the same area. Finally, a measure of the closeness based on the Hellinger coefficient is applied to produce a ranking of the thirteen regional climate models participating in the ENSEMBLES project focused mainly on their ability to simulate surface processes.

The paper is organized as follows. Section 2 describes the data sets used in this study. The ground truth from ERA-Interim re-analysis is evaluated is Sect. 3. The principles, advantages and limitations of the method are described in Sect. 4. Main results are presented in Sect. 5. Finally, conclusions are summarized in Sect. 6.

2 Data

The ERA-Interim re-analysis data (Dee et al. 2011) has been used through the whole study as a reference to compare with RCMs outputs. Although it can be argued that some soil/surface variables and surface fluxes provided by a re-analysis are not the ideal reference to be used as an accurate representation of the observed atmosphere and/or land surface, it is however a practical approach which circumvents the problem of the insufficient spatial coverage of in situ data and of the inaccuracy of satellite data for certain surface variables. It must be always kept in mind that fluxes values correspond to 12 h forecasting and therefore they are very much dependent on the underlying model.

The following data have been used for this study:

(a) Daily analysis (0000, 0600, 1200, 1800 UTC) from 1989 to 2008 of Skin Temperature (SKT) and 2-meter Temperature (T2m) and daily averaged 12 h forecasts (0000, 1200 UTC) of Surface Net Thermal Radiation (LW_{net}), Surface Net Solar Radiation (SW_{net}), Surface Sensible Heat Flux (SSHF) and Total Precipitation (PP) from the European Centre for Medium-Range Weather Forecast (ECMWF) ERA-Interim reanalysis (Dee et al. 2011). The ERA-Interim atmospheric model is configured with 60 levels in the vertical; a T255 spherical-harmonic representation for the basic dynamical fields and a reduced Gaussian grid with approximately uniform 79 km spacing for surface and other grid-point fields.

(b) Daily fields from 1991 to 2000 of Maximum Soil Temperature (T_{smx}), Minimum Soil Temperature (T_{smn}) and 2-m Temperature (T2m), and daily averaged fields of Surface Net Thermal Radiation (LW_{net}), Surface Net Solar Radiation (SW_{net}), Surface Sensible Heat Flux (SSHF) and Precipitation (PP) from the thirteen RCMs participating in the Research Theme 3 (RT3) of the ENSEMBLES project (van der Linden and Mitchell 2009). All regional simulations for the period 1991–2000 were driven by ERA-40 reanalysis (Uppala et al. 2005). Table 1 provides information of the 13 models considered in this study: institution, model, number of vertical levels and key references. The fields were obtained from the ENSEMBLES RT3/RT2B data archive (http://ensemblesrt3.dmi.dk).

Only the months of July and November corresponding to ERA-Interim and RT3-ENSEMBLES data have been used. The election is justified by the fact that July is representative of the dry season, whereas November is representative of the wet season over Southern Spain. ERA-Interim and all 13 RT3-ENSEMBLES regional

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

Table 1 List of regional climate models participating in the EU-FP6 ENSEMBLES project

Institution	RCM	Vertical levels	Reference
CHMI	ALADIN	31	N/A
C4I	RCA3	31	Kjellström et al. (2005)
DMI	HIRHAM	31	Christensen et al. (2007)
ETHZ	CLM	32	Böhm et al. (2006)
HC	HadRM3Q0	19	Collins et al. (2006)
HC	HadRM3Q3	19	Collins et al. (2006)
HC	HadRM3Q16	19	Collins et al. (2006)
KNMI	RACMO	40	Van Meijgaard et al. (2008)
METNO	HIRHAM	31	Haugen and Haakensatd (2006)
MPI	REMO	27	Jacob (2001)
SHMI	RCA	24	Kjellström et al. (2005)
UCLM	PROMES	28	Sánchez et al. (2004)
OURANOS	CRCM	29	Plummer et al. (2006)

Fig. 1 Selected area for the study of ERA-Interim re-analysis and ENSEMBLES datasets

models datasets have been interpolated to a common grid (0.25° latitude × 0.25° longitude) defined by a rectangular area (from 40.5°N to 37.5°N, and from 7.0°W to 2.0°W) covering part of Tagus and Guadiana river basins in southern Iberian Peninsula (see Fig. 1).

3 Evaluation of ground-truth ERA-Interim data

Although the quality of ERA-Interim is not the subject of this paper, its selection as ground-truth requires of previous discussion and some validation against in situ and satellite observations. In particular, the quality of the ERA-Interim selected fluxes (LW_{net}, SW_{net} and SSHF) must be carefully validated—as these quantities are not analyzed—before accepting them as ground-truth reference to compare against the corresponding quantities from regional climate models. The validation of ERA-Interim fluxes implies a certain degree of difficulty as the corresponding observational satellite data, mainly from EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) products (see http://www.cmsaf.eu) are available only for recent years and these last data do not overlap in time with RT3-ENSEMBLES regional models simulations.

For the evaluation of LW_{net} and SW_{net}, we have made use of CM SAF products. The CM SAF data products are categorized in monitoring data sets obtained in near real time and data sets based on carefully inter-sensor calibrated radiances. The homogenous sets of high-quality data are derived from several instruments on-board meteorological operational satellites in geostationary and polar orbit as the Meteosat and EUMETSAT Polar System satellites, respectively. Surface radiation products are retrieved from SEVIRI/GERB instruments on MSG satellite and AVHRR instruments on METOP and NOAA satellites. They are available as gridded monthly and daily means data at 15 × 15 km resolution.

Figure 2 shows the comparison of daily SW_{net} obtained from ERA-Interim and from CM SAF averaged for the same area and for the months of July and November corresponding to years 2006, 2007 and 2008. The figure shows a remarkable coincidence between ERA-Interim and CM SAF values for clear sky days. Cloudy days show a tendency of ERA-Interim SW_{net} to have higher values than the corresponding CM SAF ones. The mean absolute difference (MAD) between both curves is 7.52 and 13.52 Wm^{-2} for July and November, respectively (see red lines in Fig. 4). The lower value for July is mainly due to the predominance of clear sky conditions. Computation of MAD between the ENSEMBLES regional models and ERA-Interim show clearly larger values (see box plots in Fig. 4) and therefore it can be reasonably assumed that ERA-Interim SW_{net} is a good approximation for the observed reference. As data available from ENSEMBLES RCMs do not cover the period 2006–2008, we have instead compared ERA-Interim against each of the ENSEMBLES regional models for the months of July and November of years 1998, 1999 and 2000 (see Fig. 4).

Unfortunately, there is no daily data available from CM SAF for LW_{net}. Therefore, the evaluation of ERA-Interim LW_{net} will be based on monthly averages. Figure 3 depicts monthly mean LW_{net} obtained from ERA-Interim and from CM SAF averaged for the same area and for years 2006–2010. The mean absolute difference between both

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

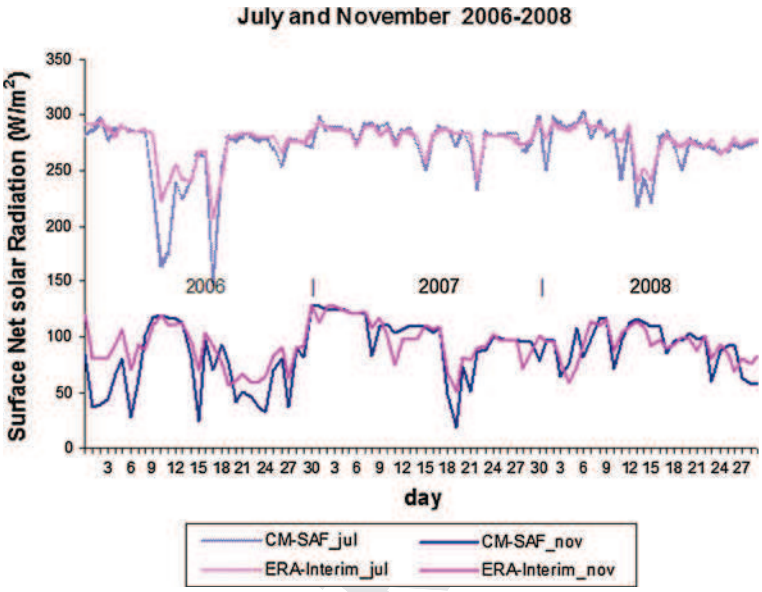


Fig. 2 Daily 12 h mean Surface Net Solar Radiation (SW_{net}) averaged over the selected area (see Fig. 1) from ERA-Interim and CM-SAF data for 3 months of July and November corresponding to years 2006–2008

curves is 4.67 Wm^{-2} for the whole period. Again, the corresponding computation of MAD between each of the 13 RT3-ENSEMBLES regional models and ERA-Interim show clearly larger values (see box plots in Fig. 4), but for the period 1996–2000, and therefore it can be reasonably assumed than ERA-Interim LW_{net} is a good approximation for the observed reference.

For the evaluation of SSHF we have to rely on in situ observations from a number of flux tower networks (Král 2011). This evaluation made use of the 2006 data from the FLUXNET LaThuile Synthesis dataset which compiles flux tower eddy-covariance measurements from a number of regional flux tower networks across the globe (Baldocchi et al. 2001). Root mean square error of ERA-Interim SSHF compared against FLUXNET daily data for the whole 2006 show values ranging from 20 to 40 Wm^{-2} for most Western European towers, values are generally lower than the corresponding rmse of regional models computed with respect to ERA-Interim SSHF. This is an expected result, consequence of the land surface analysis combining synoptic observations over land with background estimates based on 6-hourly estimates of screen-level temperature and dew point from the latest atmospheric analysis (Douville et al. 1998). The analysis increments for screen-level temperature and humidity are subsequently used to update soil moisture and soil temperature estimates for each of the four layers of the land-surface model, by a simple empirical approach (Douville et al. 2000; Mahfouf et al. 2000). Therefore, surface sensible and latent fluxes are constrained in ERA-Interim by soil moisture and soil temperature which in turn are corrected by screen-level temperature and humidity observations.

4 Methodology

Atmosphere and land surface are strongly coupled subsystems of the climate system. Surface fluxes (of energy, water, momentum, carbon, etc.) enable the coupling of both sub-systems. In fact, climate variables, as e.g. surface equilibrium temperature, diurnal temperature range, near surface air temperature and humidity, are very dependent on surface fluxes. Moreover, the entire structure and features of the atmospheric boundary layer are in turn very influenced by land-surface and atmosphere coupling expressed in the form of surface fluxes (see, e.g., Stensrud 2007). Whenever we refer in this paper to coupling between two variables, we mean that one variable controls each other (following Seneviratne et al. (2010)) or even better that both are forced to change together in a way prescribed by the underlying processes. For example, for the particular case of the pair of variables $SW_{net} - LW_{net}$, Figure 6 shows that SW_{net} increases whenever LW_{net} increases (and vice versa) for November days, whereas this is only true when SW_{net} does not reach the maximum value (generally reduced by clouds) for July days. This coupling does not necessarily mean that the relationship between both variables is linear. In fact, in most of the cases, the relationship is linear only as a first approximation. The level of dispersion shown by 2D-scattered plots indicates—without any expression of causality—how tight the relationship between pairs of variables is.

Surface fluxes involved in the surface energy budget are especially relevant for land-surface and atmosphere coupling. The surface energy budget equation can be expressed in a simplified form as:

Springer

Journal : Large 382	Dispatch : 4-12-2012	Pages : 16
Article No. : 1619	□ LE	□ TYPESET
MS Code : CLIDY-D-12-00044	✔ CP	✔ DISK

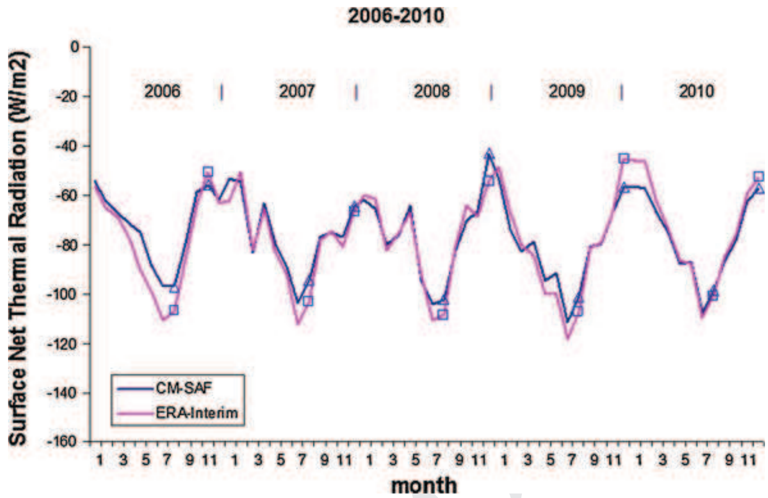


Fig. 3 Monthly mean Surface Net Thermal Radiation (LW_{net}) averaged over the selected area (see Fig. 1) from ERA-Interim and CM-SAF data for years 2006–2010. Months of July and November are additionally marked by *symbols*

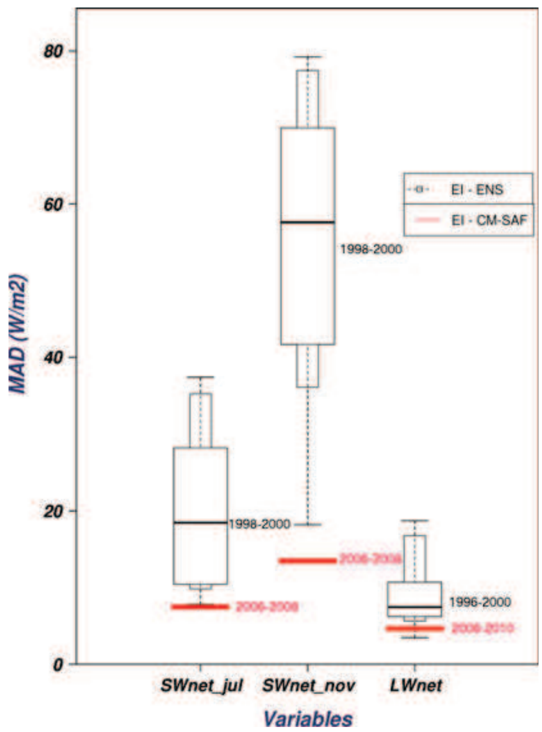


Fig. 4 Mean absolute difference of Net Solar Radiation fluxes averaged over the selected area from CM-SAF data (*red*) and thirteen ENSEMBLES RCMs (*box plot*) with respect to ERA-Interim. Daily Surface Net Solar Radiation (SW_{net}) for the months of July (*left*) and November (*centre*) and monthly Surface Net Thermal Radiation (LW_{net}) (*right*) are represented for the periods shown. *Box plots* represent the minimum, maximum, median and 10th, 25th, 75th and 90th percentiles

$$R_{net} = SW_{net} + LW_{net} = SSHF + SLHF + G \quad (1)$$

The net surface radiation, R_{net}, is the sum of net shortwave (SW_{net}) and longwave (LW_{net}) fluxes; R_{net} is balanced by the upward sensible heat flux (SSHF) the upward latent heat flux (SLHF) and the storage (G) (neglected on daily scales). Both heat fluxes are the important mechanisms to turn energy back into the atmosphere from land surface. Accuracy and minimal drift in the land-surface climate and the surface fluxes impact forecast skill on all timescales (Betts 2009; Stensrud 2007).

The surface LW_{net} plays a fundamental role in land–atmosphere coupling. Although upward and downward LW fluxes are strongly dependent functions of temperature, however, LW_{net} is largely determined by humidity and cloud cover on daily-mean timescales, due to the strong vertical coupling of the atmospheric temperature and moisture structure. For example, the depth of the daytime adiabatic mixed layer (ML) is a function of relative humidity (RH). Outgoing LW_{net} decreases as near-surface RH rises (and mean cloud-base falls), and decreases as cloud cover increases. LW_{net} plays in turn a fundamental role in the diurnal cycle over land. For example, a clear dry atmosphere gives place to an increased outgoing LW_{net} associated with surface cooling, lower minimum surface temperature at night and very stable nocturnal boundary layer, NBL. In terms of the daily climate, the strength of the NBL is closely related to the diurnal temperature range, DTR (defined as DTR = $T_{max} - T_{min}$, where T_{max}, T_{min} are the maximum and minimum values of 2-m Temperature). In the dry season, both atmospheric water vapour and cloud cover reach relatively low values and therefore the lifting condensation level (LCL) tends to reach relatively higher values, contributing all these factors to an increased outgoing LW_{net} (Betts 2009).

Surface water budget is also associated to energy budget, as latent heat flux, caused by evapotranspiration, plays an important role in both water and energy budgets. The surface water budget can be expressed as:

$$\delta S/\delta t = P - E - R \quad (2)$$

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

where S stand for terrestrial water storage, P for total amount of precipitation, E for evapotranspiration and R for total runoff.

The relative importance of latent and sensible heat fluxes depends strongly on surface features. In bare, dry soils, the absorbed radiative energy is mostly used to heat the surface, turning back energy to the atmosphere usually as a vigorous, turbulent sensible flux. On the other hand, densely vegetated surfaces with enough water available for evapotranspiration invest most of the radiative energy in extracting subsurface water through the root system. This process of transpiration is mainly controlled by leaves, opening and closing their stomata according to the environmental conditions and to the available soil wetness. Transpiration turns energy back to the atmosphere in form of latent heat flux. Over land the availability of water essentially determines evaporative fraction, EF, (being defined as SLHF/(SLHF + SSHF)). Soil water has a primary role in the surface energy partition between latent and sensible heat fluxes, and in turn in the diurnal cycle of 2-m Temperature and humidity. The latent and sensible heat fluxes play a different role for the atmosphere. Sensible heat at the bottom means energy immediately available to the atmosphere, and contributes to the heating and/or deepening of the planetary boundary layer. For an entire atmospheric column, the net radiative cooling is balanced by energy involved in phase changes inside the column (condensation of water vapour and evaporation of rain) and sensible heat flux at the surface (see, e.g., Garratt 1992; Stensrud 2007).

The three following relationships involving surface fluxes and temperatures were selected in order to evaluate the performance of the RT3-ENSEMBLE regional models when simulating atmosphere land-surface coupling:

- $SW_{net} - LW_{net}$,
- $SW_{net} - SSHF$,
- $LW_{net} - (T_{smx} - T_{smn})$.

The variables selected are readily available both from ERA-Interim and RT3-ENSEMBLE datasets and, as discussed above, are responsible and descriptive of different aspects related with energy and water budgets and with features of the atmospheric boundary layer.

The study area was selected inland of the Iberian Peninsula to avoid potential influences of the coast. The area encompassing two river basins—Tagus and Guadiana—also shows approximate homogeneity with respect to soil, vegetation and climate being predominantly flat. The selected area belongs to Mediterranean climate type with continental and Atlantic influences.

The three selected empirical relationships were derived from ERA-Interim, using daily data for July (representative of the dry season) and November (representative of the wet season), by displaying the three pairs of variables in 2D-scattered plots. The reason for the choice of these two months resides in the considerable differences appearing in the atmosphere-land surface coupling between dry and rainy seasons (Betts 2004). The 2D-scattered plots for each of the three relationships are represented in the upper left plots of Figs. 5, 6 and 7. They show some differences with the corresponding plots obtained by Betts (2004) for the Madeira (Brazil) river basin. These differences are justified by the fact that they are computed not only with different re-analysis but geographical location, period, terrain and weather conditions are also diverse. The largest differences between Madeira (tropical latitude, south of Equator) and the Iberian Peninsula (extratropical latitude) are mainly associated to minimum values of SW_{net}. Whereas the minimum value of SW_{net} in Madeira is approximately the same in dry and wet seasons, the corresponding minimum values show a difference of about 200 Wm^{-2} in the Iberian Peninsula. Also, the number of cloudless days is much higher in the Iberian Peninsula than in Madeira restricting considerably the SW_{net} range in the first case.

The corresponding relations for each of the RT3-ENSEMBLES regional simulations are then computed following the same procedure. Figures 5, 6 and 7 show 2D-scattered plots for the ERA-Interim and for the 13 regional models corresponding to each of the three relationships for dry (July) and wet (November) seasons.

Finally, in order to quantify differences or similarities in the empirical relationships between ERA-Interim and each one of the 13 regional models, the Hellinger coefficient (Hellinger 1909) has been used to measure distances of clouds of points in 2D-scattered plots. The Hellinger coefficient was originally designed to estimate the proximity of probability density functions (pdf's). The Hellinger coefficient is defined as:

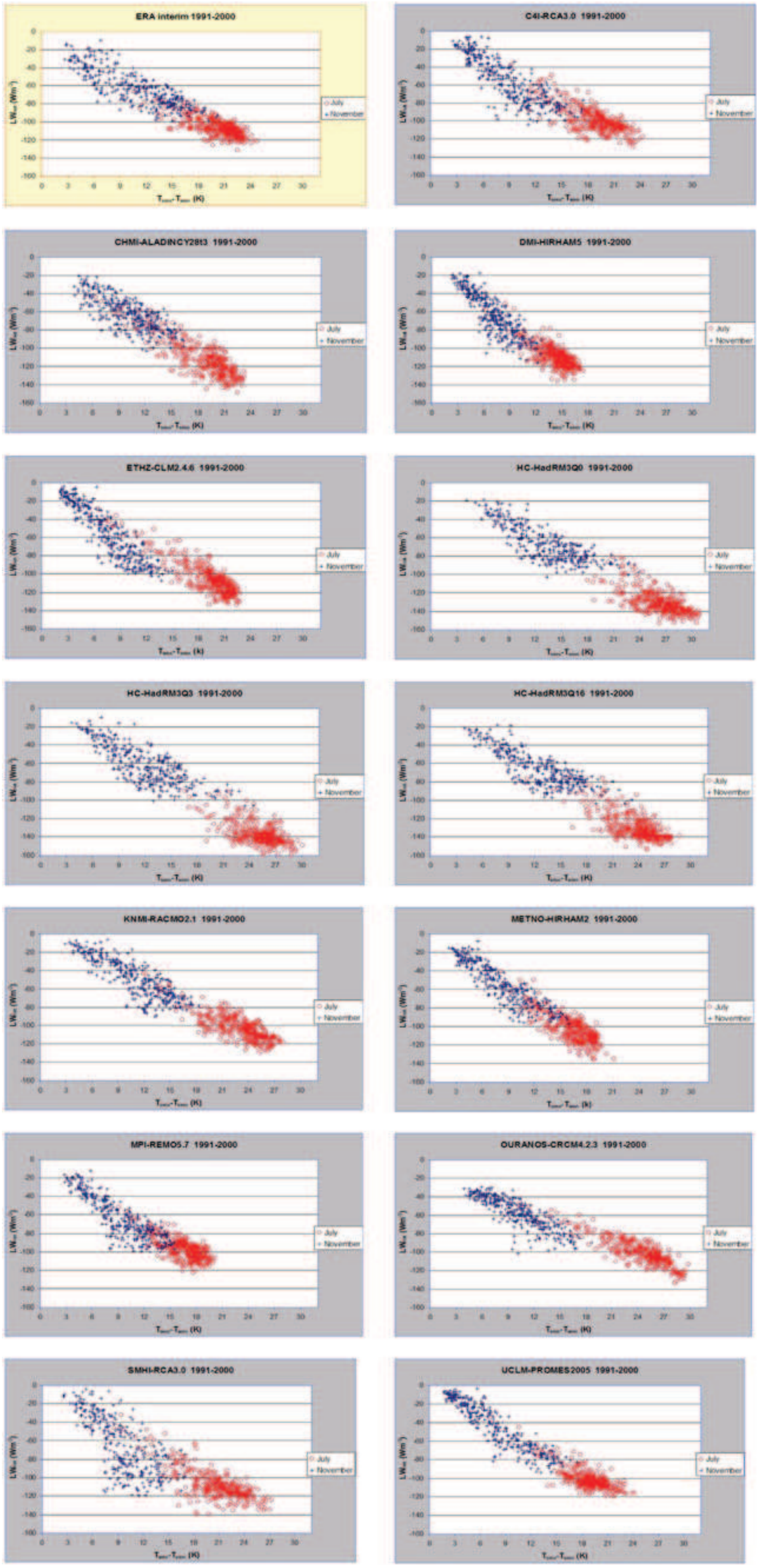
$$d_{Hell}^{(s)} = \int_R q(x)^s p(x)^{(1-s)} dx, \qquad (3)$$

where q(x) and p(x) are two pdf's to compare, and s is a parameter ($0 < s < 1$). The calculation was made choosing $s = 1/2$ which yields a symmetric measure with values between zero (p and q have disjoint supports) and one (p and q are identical). The Hellinger coefficient can be thought of as measure of the "overlap" between two distributions. Hellinger coefficient yields information about differences or similarities in relative position, shape and orientation of the pdf's. The definition given in Eq. (3) is in fact a measure of similarity.

The kind of evaluation here described is in the same spirit as those proposed by several authors (Perkins et al. 2007; Perkins and Pitman 2009; Casado and Pastor 2012)

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

Fig. 5 Scattered plots of LW_{net} as a function of $(T_{smx} - T_{smn})$ for ERA-Interim and thirteen ENSEMBLES RCMs over the selected area. *Red circles* and *blue crosses* correspond to dry (July) and wet (November) seasons, respectively



Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

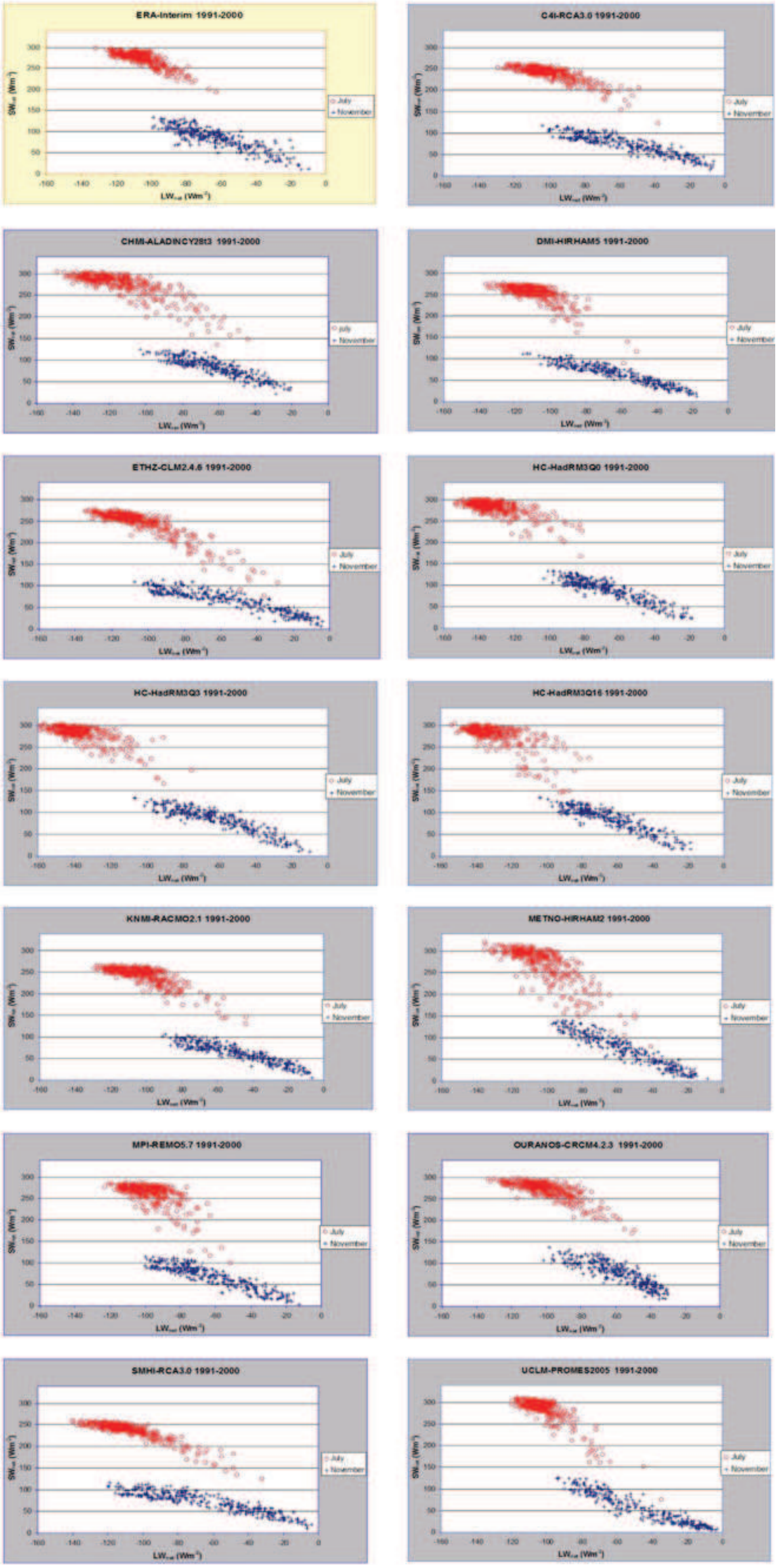
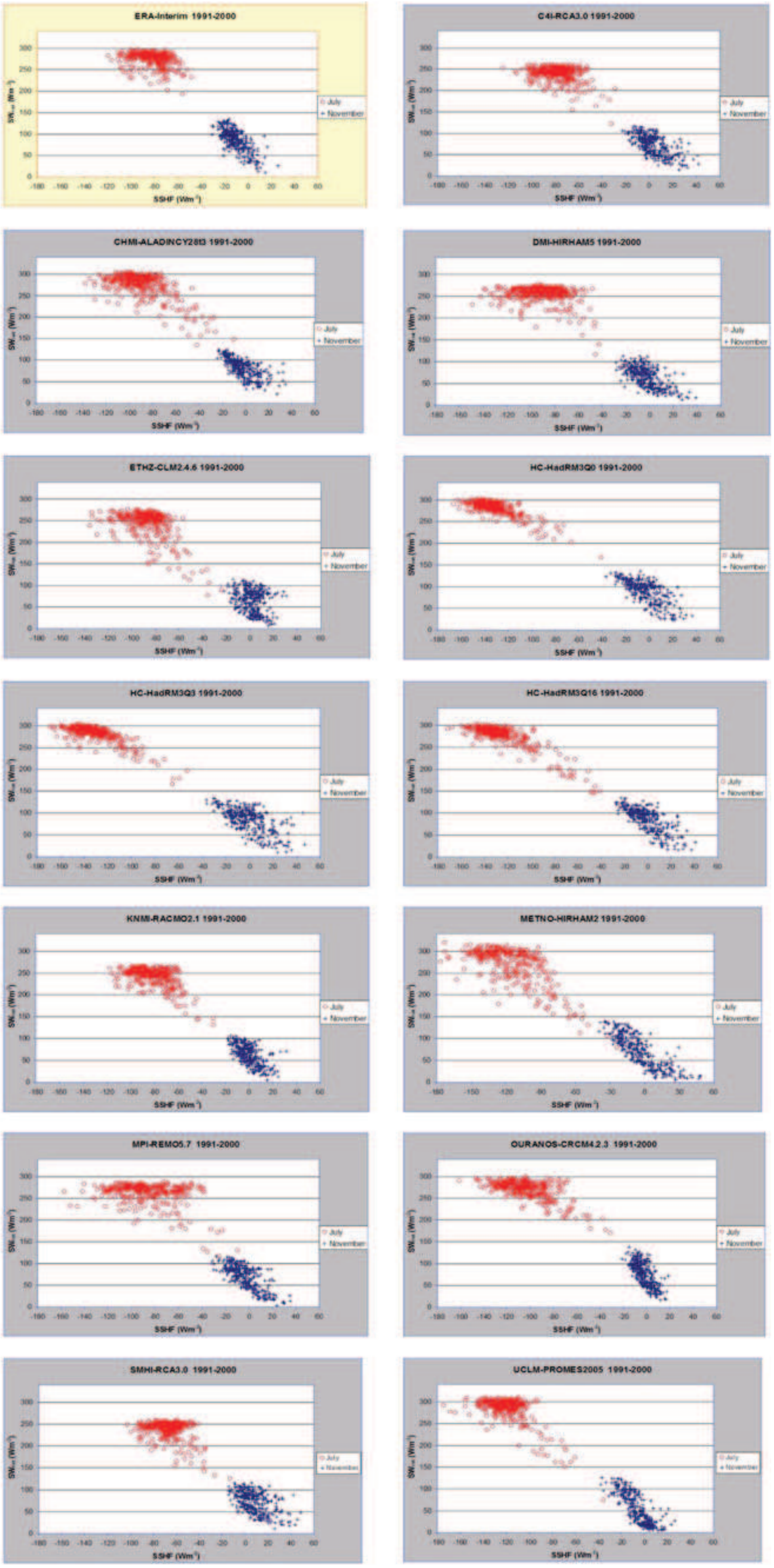


Fig. 6 The same as Fig. 5, but for SW_{net} as a function of LW_{net}

Journal : **Large 382** Dispatch : **4-12-2012** Pages : **16**
Article No. : **1619** ☐ LE ☐ TYPESET
MS Code : **CLIDY-D-12-00044** ☑ CP ☑ DISK

Fig. 7 The same as Fig. 5, but for SW_{net} as a function SSHF



Journal : **Large 382**	Dispatch : **4-12-2012**	Pages : **16**
Article No. : **1619**	☐ LE	☐ TYPESET
MS Code : **CLIDY-D-12-00044**	☑ CP	☑ DISK

who considered the great advantage of assessing climate models using metrics derived from pdf's estimated from daily data.

5 Results

Figure 5 shows the scattered plot of LW_{net} as a function of the diurnal range of soil temperature (DTR) for ERA-Interim and for each of the thirteen RT3-ENSEMBLES regional models. Points corresponding to July and November merge in a single quasi-linear distribution for most models. Other months (not shown here) fall in between filling in the same distribution. This behaviour was explained by Betts (2009) that showed that for any latitude DTR $\approx -LW_{net}$ $(1/(4\sigma T3))$, being σ the Stefan-Boltzmann constant ($\sigma = 5.67 \times 10^{-8}$ $Wm^{-2}K^{-4}$). A clear dry atmosphere above causes high values of LW_{net} and therefore cooling at the surface, leading to lower minimum surface temperature at night, and a 'stronger' nocturnal boundary layer (NBL). In terms of daily climate, this strength of the NBL is closely related to the diurnal temperature range DTR $= T_{max} - T_{min}$. Most of the plots show that the range of DTR is roughly double for November (wet season) as compared to July (dry season). LW_{net} also shows higher values for the wet season as compared to dry season. The reasons for such higher values of LW_{net} during the wet season reside principally in the usually greater cloud cover and higher lifting condensation level (LCL). From a daily climate perspective, day-time and night-time boundary layers are a fully coupled system, frequently being a deep residual mixed layer from the previous day. LW_{net} is usually correlated with the strength of NBL and the thickness of the diurnal boundary layer.

The maximum upward LW_{net} for ERA-Interim in July reaches a value of about -130 Wm^{-2}. The corresponding RCMs values for these maxima are highly variable, reaching values up to -160 Wm^{-2} (for HadRM3 model). In the month of November, maximum values of LW_{net} are of about -100 Wm^{-2} for all models (including ERA-Interim) except for SMHI-RCA and DMI-HIRHAM where maximum values rise up to -120 Wm^{-2} (see Fig. 5). These maxima correspond to clear days with low atmospheric humidity.

Figure 6 depicts the scattered plot of SW_{net} as a function of LW_{net}, showing two well differentiated distributions for July and November. The scattered plot corresponding to ERA-Interim suggests that SW_{net} and LW_{net} are coupled only in the few cloudy days of the month of July. However, no coupling seems to exist in clear days which are majority in July. None of the RCM seems to properly simulate this behaviour. Differences in the upper limits of SW_{net} of up to 30 Wm^{-2} between ERA-Interim and some RCMs might be due to different surface albedo. In November where clear days are infrequent, coupling between SW_{net} and LW_{net} is not so tight possibly caused by advection of atmospheric water vapour. Differences between RCMs and ERA-Interim are smaller in November than in July, showing several RCMs stronger $SW_{net} - LW_{net}$ coupling than for ERA-Interim.

The scattered plot of SW_{net} as a function of SSHF based on ERA-Interim (see Fig. 7) shows almost no coupling between SW_{net} and SSHF for the month of July. The surface energy budget equation (see Eq. 1) can be consequently simplified as $R_{net} = SW_{net} + LW_{net} = SSHF$ due to the lack of available water for evapotranspiration during dry season. Therefore, most of the net surface radiation, R_{net}, will turn back as SSHF to the atmosphere, favouring the coupling SSHF $- LW_{net}$ and preventing the coupling SSHF $- SW_{net}$. On the other hand, the month of November (wet season) shows a clear SW_{net} $-$ SSHF coupling. Some RCMs show greater coupling than ERA-Interim in cloudy July days. The behaviour of RCMs in November is highly variable as compared with ERA-Interim.

Table 2 summarizes Hellinger distances between ERA-Interim and each one of the ENSEMBLES RCMs and for each of the three selected relations describing the atmosphere-land surface coupling for July and November. The T2m $-$ PP relationship has also been added for the sake of comparison with previous studies (e.g., Christensen et al. 2010). Hellinger coefficients for July tend to be smaller than the corresponding values for November, meaning that coupling in dry season is worse simulated than in wet season. This effect is particularly clear for the relation SW_{net} $-$ SSHF. Tables 3 and 4 summarize for July and November standard skill scores between ERA-Interim and each one of the ENSEMBLES RCMs for 2-m Temperature and Daily Total Precipitation, respectively.

There is an overall agreement of temperature skill scores—including Hellinger coefficient for T2m – PP—discriminating consistently best and worst models (see Table 3). For example, KNMI-RACMO model in July is ranked respectively as second, first, first, fourth and first best model when using the following performance metrics: bias, mean absolute error, RMSE, correlation coefficient and Hellinger coefficient for T2m – PP. Also, HadRM3Q3 model in July is ranked as the worst model when using bias, mean absolute error and RMSE and the second and third worst when using correlation coefficient and Hellinger coefficient for T2m – PP, respectively.

Tables 3 and 4 clearly show that models performing well in 1 month and for one variable not necessarily they do in other months and variables. This fact is well known and it is a direct consequence of the predominance of certain processes in one or another season affecting more to one or another variable. For example, temperature in

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

Table 2 Values of Hellinger coefficient for the relations $LW_{net} - (T_{smx} - T_{smn})$, $SW_{net} - LW_{net}$, SW_{net} – SSHF and T2m – PP for the months of July and November

Institution-model	Hellinger coefficient July				Hellinger coefficient November			
	$LW_{net} - (T_{smx} - T_{smn})$	$SW_{net} - LW_{net}$	SW_{net} – SSHF	T2m – PP	$LW_{net} - (T_{smx} - T_{smn})$	$SW_{net} - LW_{net}$	SW_{net} – SSHF	T2m – PP
CHMI-ALADIN	0.86	0.83	**0.85**	0.84	0.96	0.99	**0.93**	0.98
C4I-RCA3	0.91	0.58	0.61	0.94	0.94	0.94	0.78	0.96
DMI-HIRHAM	0.39	0.85	0.79	0.93	<u>0.78</u>	0.91	0.85	**1.00**
ETHZ-CLM	0.88	0.70	0.74	0.86	0.85	0.90	<u>0.76</u>	0.99
METO-HC_HadRM3Q0	0.30	0.59	<u>0.25</u>	0.96	0.95	0.99	**0.93**	0.98
METO-HC_HadRM3Q3	<u>0.28</u>	<u>0.43</u>	0.28	0.84	**0.99**	**1.00**	0.88	0.98
METO-HC_HadRM3Q16	0.55	0.62	0.47	0.92	0.98	**1.00**	0.89	0.99
KNMI-RACMO	0.86	0.70	0.72	**0.99**	0.94	0.84	0.77	0.94
METNO-HIRHAM	0.71	0.79	0.51	0.92	0.91	0.93	0.84	0.96
MPI-M-REMO	0.69	0.84	0.81	0.95	0.96	0.95	0.89	0.98
SMHI-RCA	0.92	0.59	0.54	0.89	0.92	0.92	0.78	0.94
OURANOS-CRCM	0.75	**0.93**	0.77	<u>0.71</u>	0.96	0.97	0.87	<u>0.90</u>
UCLM-PROMES	**0.94**	0.89	0.40	–	0.86	<u>0.80</u>	0.85	–

The RCM acquiring the **highest** and the <u>lowest</u> respective value for each relation is indicated

Table 3 Bias, mean absolute error, root mean square error and correlation coefficient for 2-m Temperature

Institution-model	2-m Temperature July				2-m Temperature November			
	Bias	MAE	RMSE	Corr. Coeff.	Bias	MAE	RMSE	Corr. Coeff.
CHMI-ALADIN	1.23	1.29	1.63	0.92	<u>2.51</u>	<u>2.59</u>	<u>2.78</u>	0.91
C4I-RCA3	1.15	1.50	1.82	0.87	1.70	1.92	2.28	0.86
DMI-HIRHAM	−1.01	1.15	1.38	**0.94**	**0.11**	**0.73**	**0.94**	**0.94**
ETHZ-CLM	−1.07	1.33	1.52	**0.94**	0.81	1.14	1.38	0.93
METO-HC_HadRM3Q0	−1.61	2.02	2.51	0.76	1.02	1.60	2.06	<u>0.79</u>
METO-HC_HadRM3Q3	<u>−3.16</u>	<u>3.24</u>	<u>3.96</u>	0.66	0.82	1.43	1.88	0.81
METO-HC_HadRM3Q16	−2.15	2.42	3.08	0.70	0.81	1.47	1.85	0.82
KNMI-RACMO	0.70	**0.95**	**1.26**	0.93	1.84	1.95	2.24	0.90
METNO-HIRHAM	−1.34	1.73	2.17	0.84	0.25	1.04	1.30	0.89
MPI-M-REMO	−1.38	1.53	1.79	0.92	−0.48	0.91	1.21	0.92
SMHI-RCA	1.74	1.77	1.98	**0.94**	2.08	2.18	2.54	0.88
OURANOS-CRCM	2.47	2.45	2.87	0.87	2.36	2.48	2.73	0.89
UCLM-PROMES	**−0.25**	1.83	2.38	<u>0.64</u>	1.63	1.38	2.38	0.80

summertime is very much related with the correct partition of sensible and latent heat fluxes, which in turn depends on a reasonable simulation of soil water content. This is not the case in wintertime. Finally, Table 5 displays eight different rankings of the 13 ENSEMBLES RCMs according to the value of the Hellinger coefficient for each of the four considered relationships computed for the months of July and November. It is noticeable that for November there is a high consistency among rankings based on the here considered relationships. This consistency implies that one could use fewer relationships to select the models better simulating atmosphere-land surface coupling. However, discrepancy among different models rankings—depending on the chosen relation—is higher for July, possibly due to the different quality of radiation fluxes and heat fluxes. It is also noticeable the large differences appearing between dry and wet seasons in the rankings. It is very significant that some models highly scored for the wet season only get poor scores for the dry season and vice versa.

Now, at this point, question arises whether a ranking of models based on standard skill scores for 2-m Temperature

Journal : Large 382 Dispatch : 4-12-2012 Pages : 16
Article No. : 1619 ☐ LE ☐ TYPESET
MS Code : CLIDY-D-12-00044 ☑ CP ☑ DISK

Table 4 The same as Table 3, but for Daily Total Precipitation

Institution-model	Daily total precipitation July				Daily total precipitation November			
	Bias	MAE	RMSE	Corr. Coeff.	Bias	MAE	RMSE	Corr. Coeff.
CHMI-ALADIN	−0.34	0.38	1.10	0.79	−0.51	0.83	1.84	**0.94**
C4I-RCA3	−0.20	0.31	0.81	0.62	−0.45	1.06	2.19	0.87
DMI-HIRHAM	**0.00**	**0.19**	0.74	0.78	−0.07	0.83	1.99	0.90
ETHZ-CLM	−0.13	0.26	1.11	0.66	−0.15	**0.74**	**1.66**	0.92
METO-HC_HadRM3Q0	−0.08	0.31	0.76	0.37	**−0.05**	0.92	2.31	0.86
METO-HC_HadRM3Q3	−0.01	0.26	0.73	0.30	−0.25	0.96	2.38	0.88
METO-HC_HadRM3Q16	−0.07	0.32	0.89	0.23	**−0.05**	0.89	2.24	0.87
KNMI-RACMO	0.05	**0.19**	0.72	0.50	−0.40	0.84	2.09	0.90
METNO-HIRHAM	−0.08	0.22	**0.69**	**0.81**	−0.89	1.18	3.17	0.89
MPI-M-REMO	−0.17	0.28	0.75	0.57	−0.24	0.82	2.39	0.89
SMHI-RCA	−0.14	0.26	0.82	0.76	−0.35	0.85	1.68	0.92
OURANOS-CRCM	−0.99	0.98	1.70	0.63	−0.07	1.00	1.87	0.90

Table 5 Rankings of 13 ENSEMBLES RCMs (in numbers) according to Hellinger coefficient based on the proximity of the relationships: $LW_{net} - (T_{smx} - T_{smn})$, $SW_{net} - LW_{net}$, SW_{net} − SSHF, and T2m – PP for the months of July and November

Institution-model	July				November			
	$LW_{net} - (T_{smx} - T_{smn})$	$SW_{net} - LW_{net}$	SW_{net} − SSHF	T2m – PP	$LW_{net} - (T_{smx} - T_{smn})$	$SW_{net} - LW_{net}$	SW_{net} − SSHF	T2m – PP
CHMI-ALADIN	6	5	**1**	10	4	4	**1**	7
C4I-RCA3	3	12	7	4	7	7	10	9
DMI-HIRHAM	11	3	3	5	13	10	7	**1**
ETHZ-CLM	4	7	5	9	12	11	13	2
HC-HadRM3Q0	12	10	13	2	6	3	2	5
HC-HadRM3Q3	13	13	12	11	**1**	**1**	5	6
HC-HadRM3Q16	10	9	10	6	2	2	3	3
KNMI-RACMO	5	8	6	**1**	8	12	12	10
METNO-HIRHAM	8	6	9	7	10	8	9	8
MPI-REMO	9	4	2	3	5	6	4	4
SMHI-RCA	2	11	8	8	9	9	11	11
OURANOS-CRCM	7	**1**	4	12	3	5	6	12
UCLM-PROMES	**1**	2	11	–	11	13	8	–

and Daily Total Precipitation would be consistent with a ranking based on Hellinger coefficients as it is here proposed. And provided that consistency of results holds, what would an evaluation based on Hellinger coefficients add to the more traditional approach based on skill scores for temperature and precipitation? Results summarized in Tables 2, 3, 4 and 5 allow us to conclude that not always models best/worst performing in terms of standard scores for temperature and precipitation show consistent performance in terms of Hellinger coefficients for the pairs of quantities here selected. As an example, the outstanding performance of KNMI-RACMO model in July for temperature (see Table 3) has not counterpart in terms of Hellinger coefficients (see Table 5). This can be explained by the fact that the overall surface energy budget is reasonably well captured although individual fluxes might not be properly simulated. On the other hand, the deficient performance of HadRM3Q3 model in July for temperature is also confirmed in terms of Hellinger coefficients. In November consistency among standard scores for temperature and Hellinger coefficients is less clear. This may be justified by the fact that local wintertime (heat and radiation) fluxes are not so strong and consequently 2-m Temperature is also affected by other non-local factors.

The comparison of our results with those of Christensen et al. (2010) is not straightforward for a number of reasons.

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | LE | TYPESET
MS Code : CLIDY-D-12-00044 | CP | DISK

First, their work was aiming to merge a collection of 6 performance metrics into an aggregated model weight with the purpose of combining climate change information from the range of RCMs. They proposed 3 different ways of combining the 6 performance metrics showing a relatively high degree of coincidence for the final weight. Second, the purpose of their work was to get a single valued model weight describing the overall performance of each RCM for the whole domain, for all seasons and for all considered variables. Contrary, our work does not intend to generate an overall performance score. We have instead attempted to propose some scores based on the Hellinger coefficient determining how well atmosphere-land surface coupling is simulated by models. Furthermore, this evaluation scores may help to detect problems which may be behind a poor model performance in terms of temperature and precipitation. Nevertheless, some coincidences appear in the results based on both approaches.

Therefore, we have preferred not to merge the obtained eight rankings into just one ranking in order to highlight how differences among rankings depend strongly on season and to a lesser extent on the particular relationship expressing the atmosphere-land surface coupling. We confirm with our results that model rankings are highly dependent on region, variables, seasons and metrics selected for the evaluation in full agreement with other authors (e.g., Knutti et al. 2010; Casado and Pastor 2012).

6 Conclusions

An original approach has been proposed for evaluating regional climate models based on the comparison of empirical relationships among model outcome variables. The proposed method provides tools to identify which processes related to the atmosphere-land surface coupling are not properly simulated by models. Contrary to more classical methods essentially focused on traditional climate variables—like air temperature and precipitation—here the focus is put on fluxes which are in the end terms appearing in the budget equations determining temperature and soil moisture. Soil moisture is responsible for the right partition of surface energy between latent and sensible heat fluxes, and in turn of the structure of boundary layer in terms of temperature and humidity. The approach provides a quantitative evaluation of models and therefore allows the establishment of model rankings focusing on the ability to properly simulate the interaction between atmosphere and land surface. Thirteen RCMs participating in the ENSEMBLES project were selected by the availability of daily data for the period 1991–2000 of the variables LW_{net}, SW_{net}, SSHF, Tsmax and Tsmin. Three pairs of relations among surface energy variables and fluxes relevant to the energy and water budget were obtained for an area covering part of two river basins within southern Iberian Peninsula and for 2 months representative of the dry and wet seasons, respectively. The truth to compare with model simulations was ERA-Interim re-analysis. As it was already mentioned in Sect. 1, the comparison of RCMs against ERA-Interim may have certain flaws mainly when comparing variables not directly observed, as it is the case for the fluxes. However, comparison of ERA-Interim fluxes against satellite estimations allow us to conclude that ERA-Interim fluxes have a reasonable quality to be used as ground truth reference. Our main aim, however, was to illustrate the value of comparing magnitudes representative of certain processes in order to quantify how well models are capturing them. Besides, significant deviation of some models for certain magnitudes and seasons can help to identify problems when simulating processes as complex as those responsible for the atmosphere-land surface coupling. The Hellinger coefficient was the metric selected to quantify the distance between each of the regional models and the reference represented by ERA-Interim.

The comparison of the relationships here obtained for southern Iberian Peninsula with those obtained by Betts (2004) for the Madeira basin (Brazil) confirms that such comparison is highly dependent on season, region and climate conditions. In that sense, this approach is very adequate to quantify the regional performance of climate models.

The proximity of modelled and reference scattered plots depends very much on the season. The generally higher value of Hellinger coefficient (lower distance) for the wet season is indicative of difficulties associated with the simulation of atmosphere-land surface coupling during the dry season. Moreover, the high coincidence of the four rankings for the wet season suggests that only one relation may be enough to discriminate the "best" and "worst" models at that time of the year. This is not the case for the dry season, where more relations seem to be needed to quantify the radiative and water aspects of modelled surface coupling. The range of Hellinger coefficient values tends to be narrower in the wet season showing a high degree of agreement among different model simulations in coincidence with results by Betts et al. (2006).

We would like to point out that most methods for evaluating climate models frequently put the focus on outcome variables (usually precipitation and temperature) disregarding important aspects related to the coupling between subsystems of the climate system. We are convinced of the importance of evaluation studies focusing on physical processes, and in particular on the features of interface between subsystems. In this line, our approach aims directly at the performance of models in connection with the atmosphere-land surface interaction which is in the end highly responsible for a realistic simulation of

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | ☐ LE | ☐ TYPESET
MS Code : CLIDY-D-12-00044 | ☑ CP | ☑ DISK

variables more commonly described in climate studies, such as precipitation and temperature.

We may conclude by saying that the here proposed method of evaluating RCMs does not only intend to present an additional set of performance-based metrics aiming to rank models or to weight them within an ensemble of RCMs as it was proposed by other authors (e.g., Christensen et al. 2010). Our proposal goes mainly in the direction of exploring and quantifying how well coupling between atmosphere-land surface is simulated by different RCMs. As we mentioned in the introduction, climate models are based on sound and well established physical laws and their success in simulating the climate system depends on an accurate representation of the climate relevant processes. Consequently, our proposal of evaluation heavily relies on physical processes—and in this particular case on interaction between subsystems—instead of the more traditional methods which are more focused on the behaviour of climate variables such as temperature and precipitation. Additionally, the analysis of the simulated coupling between subsystems could help to diagnose modelling deficiencies which may be behind a poor performance in terms of climate variables.

Acknowledgments The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged by the authors of the paper, without these data it would have been impossible to write this article. We also thank DWD and ECMWF for providing Climate Monitoring SAF (CM SAF) and ERA-Interim re-analysis datasets, respectively. Special thanks are also due to our colleagues M.J. Casado, M.A. Pastor, J.A. López, J.A. García-Moya and B. Navascués for their fruitful suggestions. Finally, we thank to two anonymous referees who have substantially contributed with their constructive comments to the final form of this work.

References

Baldocchi D, Falge E, Gu L, Olson R, Hollinger D, Running S, Anthoni P, Bernhofer Ch, Davis K, Evans R et al (2001) FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide. Bull Am Meteorol Soc 82:2415–2434

Betts AK (2004) Understanding hydrometeorology using global models. Bull Am Meteorol Soc 85:1673–1688. doi:10.1175 BAMS-85-11-1673

Betts AK (2007) Coupling of water vapor convergence, clouds, precipitation, and land-surface processes. J Geophys Res 112: D10108. doi:10.1029/2006JD008191

Betts AK (2009) Land surface-atmosphere coupling in observations and models. J Adv Model Earth Syst 1, Art.#4, 18 pp, doi: 10.3894/JAMES.2009.1.4

Betts AK, Ball J, Barr A, Black TA, McCaughey JH, Viterbo P (2006) Assessing land-surface-atmosphere coupling in the ERA-40 reanalysis with boreal forest data. Agric For Meteorol 140:355–382. doi:10.1016/j.agrformet.2006.08.009

Böhm U, Küchen M, Ahrens W, Block A, Hauffe D, Keuler K, Rockel B, Will A (2006). CLM-the climate version of LM: brief description and long-term applications. COSMO Newsletter No 6

Casado MJ, Pastor MA (2012) Use of variability modes to evaluate AR4 climate models over the Euro-Atlantic region. Clim Dyn 38:225–237. doi:10.1007/s00382-011-1077-2

CCSP (2008) Climate models: an assessment of strengths and limitations. A report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research [Bader D.C., C. Covey, W.J. Gutowski Jr., I.M. Held, K.E. Kunkel, R.L. Miller, R.T. Tokmakian and M.H. Zhang (Authors)]. Department of Energy, Office of Biological and Environmental Research, Washington, D.C., USA, 124 pp

Christensen JH, Christensen OB (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. Clim Change 81(Suppl 1):7–30

Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. Clim Res 44:179–194

Collins M, Booth BBB, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2006) Towards quantifying uncertainty in transient climate change. Clim Dyn 27:127–147. doi:10.1007/s00382-006-0121-0

Cramer H (1946) Mathematical methods of statistics. Princeton University Press, Princeton, p 354

Dee DP et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597. doi:10.1002/qj.828

Déqué M, Jones RG, Wild M, Giorgi F et al (2005) Global high resolution versus limited area model climate change projections over Europe: quantifying confidence level from PRUDENCE results. Clim Dyn 25:653–670

Déqué M, Rowell DP, Lüthi D, Giorgi F et al (2007) An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Clim Change 81(Suppl 1):53–70

Douville H, Mahfouf J-F, Saarinen S, Viterbo P (1998) The ECMWF surface analysis: diagnostics and prospects. Tech. Memo. No. 258. ECMWF, Reading, UK

Douville H, Viterbo P, Mahfouf J-F, Beljaars ACM (2000) Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data. Mon Weather Rev 128:1733–1756

Garratt JR (1992) The atmospheric boundary layer. Cambridge University Press, Cambridge, p 316

Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the Reliability Ensemble Averaging (REA) method. J Clim 15:1141–1158

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113:D06104. doi:10.1029/2007JD008972

Haugen JE, Haakensatd H (2006) Validation of HIRHAM version with 50 km and 25 km resolution. RegClim General Technical Report, No. 9, pp 159–173

Hellinger E (1909) Neue Begründung der Theorie quadratischer Formen von unendlich vielen Veränderlichen. J Reine Angew Math 136:210–271

Jacob D (2001) A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. Meteorol Atmos Phys 77:61–73

Jaeger EB, Stöckli R, Seneviratne SI (2009) Analysis of planetary boundary fluxes and land-atmosphere coupling in the regional climate model CLM. J Geophys Res 114:D17106

Kjellström E, Giorgi F (2010) Regional climate model evaluation and weighting, introduction. Clim Res 44:117–119. doi:10.3354/cr00976

Kjellström E, Bärring L, Gollvik S, Hansson U, Jones C, Samuelsson P, Rummukainen M, Ullersig A, Willen U, Wyser K (2005) A

140-year simulation of European climate with the new version of the Rossby Centre regional atmospheric climate model (RCA3). Reports Meteorology and Climatology, 108, SMHI, SE-60176 Norrköping, Sweden, 54 pp

Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010) Challenges in combining projections from multiple climate models. J Clim 23:2739–2758. doi:10.1175/2009JCLI3361.1

Král T (2011) Flux tower observations for the evaluation of land surface schemes: application to ERA-Interim. ERA report series No 11. ECMWF, Shinfield Park, Reading, RG2 9AX, England

Mahfouf J-F, Viterbo P, Douville H, Beljaars ACM, Saarinen S (2000) A revised land-surface analysis scheme in the integrated forecasting system. ECMWF Newsl 88:8–13

Mearns LO, Giorgi F, Whetton P, Pabon D, Hulme M, and Lal M (2003) Guidelines for use of climate scenarios developed from regional climate model experiments. Data distribution centre of the international panel of climate change, 38 pp. (Available for download from www.ipcc-data.org/guidelines/dgm_no1_v1_10-2003.pdf)

Perkins SE, Pitman AJ (2009) Do weak AR4 models bias projections of future climate changes over Australia?. Clim Change 93:527–558

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376

Plummer D, Caya D, Coté H, Frigon A, Biner S, Giguère M, Paquin D, Harvey R, de Elia R (2006) Climate and climate change over North America as simulated by the canadian regional climate model. J Clim 19:3112–3132

Randall DA, Wood RA, Bony S, Coleman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer RJ, Sumi A, Taylor KE (2007) Climate models and their evaluation, chapter of the book climate change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 589–662

Sánchez E, Gallardo C, Gaertner MA, Arribas A, Castro M (2004) Future climate extreme events in the Mediterranean simulated by a regional climate model: a first approach. Global Planet Change 44:163–180

Santanello JA Jr, Peters-Lidard CD, Kumar SV, Alonge C, Tao WK (2009) A modeling and observational framework for diagnosing local land-atmosphere coupling on diurnal time scales. J Hydrometeorol 10(3):577–599

Schmidli J, Goodess CM, Frei C, Haylock MR, Hundecha Y, Ribalaygua J, Schmith T (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. J Geophys Res 112:D04105. doi:10.1029/2005JD007026

Seneviratne SI, Corti T, Davin E, 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.2010.02.04

Stensrud DJ (2007) Parameterization schemes: keys to understanding numerical weather prediction models. Cambridge University Press. ISBN: 9780521865401, 459 p

Uppala SM, Källberg PW, Simmons AJ, Andrae U, da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, UnCTh A, Vasiljevic D, Viterbo P, Woollen J (2005) The ERA-40 re-analysis. Q J R Meteorol Soc 131:2961–3012. doi:10.1256/qj.04.176

van der Linden P, Mitchell JFB (eds) (2009) ENSEMBLES: climate change and its impacts. Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Exeter

Van Meijgaard et al (2008) Simulation of present day climate in RACMO2: first results and model developments. Technical report No 252, KNMI, 24 pp

Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004). Guidelines for use of climate scenarios developed from statistical downscaling method. Data Distribution centre of the international panel of climate change, 27 pp. (Available for download from www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf)

Journal : Large 382 | Dispatch : 4-12-2012 | Pages : 16
Article No. : 1619 | LE | TYPESET
MS Code : CLIDY-D-12-00044 | CP | DISK