

What is the role of the observational dataset in the evaluation and scoring of climate models?

J. J. Gómez-Navarro,¹ J. P. Montávez,² S. Jerez,³ P. Jiménez-Guerrero,² and E. Zorita¹

Received 18 October 2012; revised 7 November 2012; accepted 8 November 2012; published 20 December 2012.

[1] Climate models are usually assessed through their capacity to reproduce present climate conditions, which in turn are established by comparing the output of climate simulations with observational data sets including gridded products. However, due to the nature of the procedures to obtain observations and the statistical techniques employed to extrapolate this information onto reference gridded databases, they contain important uncertainties which may compromise the evaluation process. This paper examines to what extent the evaluation and ranking of an ensemble of regional climate models, according to their ability to reproduce the observed climatologies, is sensitive to the choice of the reference observational data set. Results show that even in areas covered by dense monitoring networks such as Spain, uncertainties in the observations are comparable to the uncertainties within state-of-the-art Regional Climate Models, at least when they are driven by nominally perfect boundary conditions like reanalysis. These findings point out that model evaluation should take into account the observational uncertainties. In particular, weighting models according to how well they perform with respect to a unique observation dataset, without acknowledging uncertainties in the observational dataset, might reduce the quality of the weighted ensemble average. **Citation:** Gómez-Navarro, J. J., J. P. Montávez, S. Jerez, P. Jiménez-Guerrero, and E. Zorita (2012), What is the role of the observational dataset in the evaluation and scoring of climate models?, *Geophys. Res. Lett.*, 39, L24701, doi:10.1029/2012GL054206.

1. Introduction

[2] Regional Climate Models (RCMs) are powerful tools that provide high-resolution information of a number of climatological variables. The simulations performed with these models can be used for a variety of purposes, such as to gain insight in the past and future evolution of the climate, climate change impact studies, air quality evaluation, or assessment of wind power resources among many others. However, the skill of the simulations must be established before the generated information can be safely employed, given that they are not free of uncertainties. These uncertainties arise from many factors, such as the unknown evolution of climate

forcings in the past, caveats in the input fields used to drive the regional simulation, or from physical processes not properly described in the regional model itself.

[3] A standard approach to address the model uncertainty is through ensembles of simulations. In particular, the uncertainties attributable to the use of regional models are especially noticeable when several RCMs are driven by the same boundary conditions. In this sense, large community projects such as PRUDENCE [Jacob *et al.*, 2007], ENSEMBLES [Christensen *et al.*, 2010] or more recently CORDEX [Nikulin *et al.*, 2012], have produced large ensembles of regional climate simulations with different models over the same area and period, with the aim of producing accurate climate change projections but also identifying the uncertainties associated with the use of regional models. Two such ensembles have been recently produced for the Iberian Peninsula. The ESCENA project [Jiménez-Guerrero *et al.*, 2012] was aimed at producing a multi-model ensemble of regional simulations over an area encompassing Spain up to the Canary Island. A second ensemble was produced by Jerez *et al.* [2012], in which all the members share the RCM but use different parametrization schemes, with the aim of identifying the uncertainties attributable to these parametrizations within the same model.

[4] In the context of model ensembles, a key question is how to combine the output of different models to improve the quality of the information from that available from any individual simulation. Experience has demonstrated that this is not a simple task, and so far there is no consensus on what is the best methodology to achieve optimum result. The easiest approach consists of assigning one vote to each model [Meehl *et al.*, 2007], while other authors propose using different weights for every model within the ensemble according to different criteria. A common criterion involves comparing the model results with observations of the present climate. Several techniques have been developed since the first attempts by Giorgi and Mearns [2002] based on Reliability Ensemble Average (REA) method, many of which have been tested within the ENSEMBLES project. Coppola *et al.* [2010] introduced a model weight based on mesoscale structures in precipitation and temperature. Kjellström *et al.* [2010] developed a metric based on the ability of the models to reproduce the Probability Distribution Function (PDF) of daily and monthly series of temperature and precipitation. Lenderink [2010] concentrated the scoring on how the models reproduced extreme precipitation events, and found that they generally tend to overestimate these events, whereas Lorenz and Jacob [2010] focused it on how models reproduced the seasonal and annual trends in the period 1960–2000 compared to ERA40. These studies found that although the ensemble mean weighted according to each technique performs better than the unweighted version in the simulation of the observed climate, no metric was able to identify a model that outperforms the

¹Department of Palaeoclimatology, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany.

²Department of Physics, UIOyN, Regional Campus of International Excellence “Campus Mare Nostrum”, University of Murcia, Spain.

³IDL, Universidade de Lisboa, Lisbon, Portugal.

Corresponding author: J. J. Gómez-Navarro, Department of Palaeoclimatology, Helmholtz-Zentrum Geesthacht, Max-Planck-Strasse 1, DE-21502 Geesthacht, Germany. (juan.gomez-navarro@hzg.de)

others in every area, season and/or variable. However there are also weighting schemes that depend on other factors than simply the similarity between model output and observations. This includes the similarities among the model simulations themselves [Tebaldi *et al.*, 2005] or a combination of both [Räisänen *et al.*, 2010].

[5] When using a weighting methodology according to how models reproduce the observed present, it should be kept in mind that the observational datasets used as a reference also contain non-negligible uncertainties. Although it has been suggested that errors in observations are relatively small compared with intermodel differences [Gleckler *et al.*, 2008], both kinds of uncertainties become comparable at local scale when regional models are driven by “perfect” boundary conditions. In this respect, Lenderink [2010] pointed out within the ENSEMBLES project context that the spread between models and observations was generally larger over areas with a low density of observations, which might indicate that the quality of the observational dataset should also be considered. In the same special issue, Christensen *et al.* [2010] also suggested that the quality of the observations should be assessed, but they postponed it for future studies.

[6] The present study tries to determine whether the observational dataset used to score climate models might play an important role in the ranking of climate simulations within an ensemble. To do so, the results of two independent ensembles are analysed and compared with three different observational databases over a region with a relatively good station coverage (Spain) during the last part of the 20th century, where and when the spread between observational databases is expected to be small.

2. Data Employed

[7] This study employs several datasets: three gridded observational databases (E-OBS, SPAIN02 and AEMET) and two ensembles of hindcast simulations (MULTI-MODEL and MULTI-PHYSICS), presented in greater detail below.

[8] The observational dataset E-OBS [Haylock *et al.*, 2008] covers Europe and was originally developed within the context of the ENSEMBLES project. It has been updated several times to improve its quality by increasing the number of meteorological stations in poorly-covered areas. This study employs version 5, which includes a better coverage of observations over the Iberian Peninsula compared to earlier versions. SPAIN02 [Herrera *et al.*, 2010] was developed in the University of Cantabria. It only considers Spain and compiles information from a large number of meteorological stations over the area. Finally, the AEMET database [Yolanda-Luna *et al.*, 2008] is an observational database developed by the Spanish Meteorological Office and is the one that includes the largest number of observations. All these observational datasets share a number of important features which make them suitable for this analysis. First, the overlap period covers more than fifty years (1950–2008), which is a fairly long period for validation purposes. Second, their spatial resolution is also similar, around 25 km, which allows interpolating all databases onto the same grid, thereby minimizing errors. Finally, all observational databases contain daily series of the same three variables: maximum temperature (TMAX), minimum temperature (TMIN) and accumulated precipitation (PRE).

[9] In the simulations, two independent model simulation ensembles have been employed. The MULTI-MODEL

ensemble consists of a five-member group of high-resolution climate simulations performed with five Regional Climate Models (RCM) over the Iberian Peninsula produced within the context of the ESCENA project [Jiménez-Guerrero *et al.*, 2012]. All simulations were driven by the reanalysis ERA Interim from the ECMWF [Dee *et al.*, 2011]. The period considered is 1989–2008, and all the simulated domains, although slightly different in each model, cover the Iberian Peninsula with a resolution of 25 Km. The MULTI-PHYSICS ensemble consists of eight members, having in common the RCM used to conduct all the simulations, MM5 [Jerez *et al.*, 2012]. The driving conditions are ERA40 [Uppala *et al.*, 2005], and the simulated period is 1970–2000. In this case the domain is the same in all simulation, encompassing the Iberian Peninsula with a resolution of 30 Km. The ensemble has been generated by selecting two different options of physical schemes for each one of the following three parametrizations schemes: Planetary Boundary Layer, Microphysics and Cumulus. Thus, the major difference between the two ensembles is the driving data (ERA40 vs. ERA Interim, but in all cases updated every 6 hours), and the fact that the MULTI-MODEL is more heterogeneous in terms of grid and model setup. The use of two independent ensembles, with different nature and spanning different periods, solidify the findings analysed in this paper.

[10] To facilitate the comparison between different sources, all datasets have been interpolated onto the same grid using a bilinear method. Given that observations do not cover the ocean, a land mask has been applied also to the model fields, so that calculations herein do not consider ocean grid-cells. The common grid and land mask where all calculations are performed is the original SPAIN02 grid, consisting in a regular $0.2^\circ \times 0.2^\circ$ latitude-longitude grid.

3. Results

3.1. Scoring the Models

[11] As mentioned above, this study uses three different observational datasets to rank the models within two independent ensembles, to analyze the impact of the choice of a reference frame in the evaluation process. The ranking is established according to how well models reproduce the seasonal mean values of the three variables considered in a reference period (1970–2000 for MULTI-PHYSICS and 1989–2008 for MULTI-MODEL). For the sake of simplicity, a straightforward metric has been chosen, consisting of the spatial correlation of observed and simulated fields, although other metrics such as the Root Mean Squared Error, (RMSE) have been employed with similar results. Top half of Figure 1 shows the results of this scoring within the MULTI-PHYSICS ensemble. First, the figure illustrates that a “best model”, one that outperforms the others for all variables and seasons, does not really stand out. This is in good agreement with former analysis with ensembles of simulations [Coppola *et al.*, 2010; Jiménez-Guerrero *et al.*, 2012; Jerez *et al.*, 2012]. However, more interesting is to note that even focusing on one variable and one season, the model ranking is very sensitive to the choice of the observational database. For example in the case of TMAX in winter, the pink model is the best when SPAIN02 or AEMET databases are used as reference, but is fourth when E-OBS is considered. The yellow model is the best for reproducing maximum temperature in winter with respect to E-OBS, while it is the worst with respect to

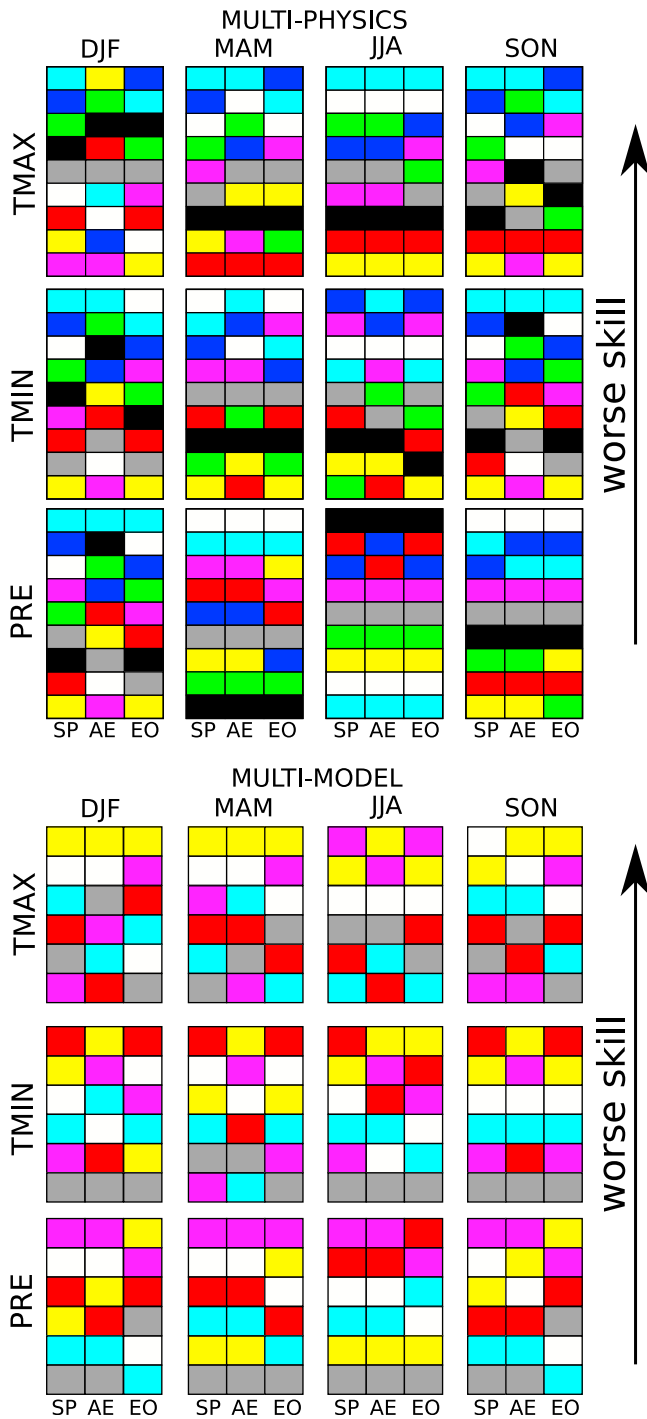


Figure 1. Scoring results within the (top) MULTI-PHYSICS ensemble and (bottom) MULTI-MODEL ensemble. In each case, there are 12 panels representing the result of the scoring separately for each season (by columns) and variable (by rows) considered here. Within a panel, each colour represents a model configuration. The lower is the square in the column, the better the model performs with respect to a given observational database. The three columns within a panel represent the scoring using the three observational datasets in this order (from left to right): SPAIN02, AEMET and E-OBS. The colours legend is not shown deliberately, and grey squares represent the ensemble average.

AEMET. These are not isolated cases, but rather the general picture that emerges from the figure. If the choice of the observational database played a weak role in the validation process, all colours would tend to line up. However, this is clearly not the case. Instead, Figure 1 shows a rather quasi-random distribution of skills, with the noticeable exception of summer TMAX and PRE, where the squares tend to line up. Results for the MULTI-MODEL ensemble, shown in the bottom half of the Figure 1 are very similar and reinforce the findings described for the MULTI-PHYSICS ensemble.

3.2. Spread in the Scores

[12] Figure 1 qualitatively illustrates how the order in the ranking of climate models is sensitive to the observational dataset employed as reference. In order to assess quantitatively the contribution of observations and models to the evaluation of uncertainties, two new statistical parameters have been defined. The spatial correlations of the climatological patterns in a reference period are computed between all combinations of models and observations. Thus, a set of correlations $\rho_{i,j}$ is generated, where i runs over the 3 observational datasets and j the ensemble (8 elements in the MULTI-PHYSICS ensemble and 5 in the MULTI-MODEL ensemble). Within this set, the model spread (Δ_{mod}) is defined as the mean (within the three observational databases) of the maximum distance in the correlations across the ensemble models. This is:

$$\Delta_{\text{mod}} = \frac{1}{3} \sum_i \left[\max_j \{ \rho_{i,j}, \forall j \} - \min_j \{ \rho_{i,j}, \forall j \} \right] \quad (1)$$

Similarly, the observation-derived spread (Δ_{obs}) is defined as the mean (within the ensemble) of the maximum distance in the correlations through the databases:

$$\Delta_{\text{obs}} = \frac{1}{8} \sum_j \left[\max_i \{ \rho_{i,j}, \forall i \} - \min_i \{ \rho_{i,j}, \forall i \} \right] \quad (2)$$

These statistics indicate the variability in the calculation of the spatial correlation due to the use of different models (Δ_{mod}) or observational datasets (Δ_{obs}). Obviously, these spreads can be calculated using metrics other than the spatial correlation, such as RMSE, and their comparison can determine whether the uncertainties related to climate models are comparable to those in the observations.

[13] The results for the spreads associated with the calculation of spatial correlation and RMSE for the MULTI-PHYSICS ensemble are shown in the first four columns of Table 1. In precipitation, the spread among the models is generally larger than in observations, regardless of the metric employed to validate the models (with the noticeable exception of spring precipitation, where differences among observations are more prominent than those among models). This indicates that the physical processes involved in precipitation are still not fully understood (note that although the model is the same in all the ensemble members, each one employs different parametrization schemes), with underlying uncertainties that exceed those associated to observations. However, the situation is different for temperature, where the spread in the correlation is generally larger for observations than for models. This indicates that uncertainties in observations are at least as relevant as those within this ensemble of simulations when evaluating how models

Table 1. Spread in the Calculation of Spatial Correlation ($\times 100$) and RMSE (in Degrees Celsius for TMAX and TMIN and mm/season for PRE) Within the MULTI-PHYSICS Ensemble and MULTI-MODEL Ensemble for All Variables and Seasons

Var.	Seas.	MULTI-PHYSICS				MULTI-MODEL			
		COR ($\times 100$)		RMSE		COR ($\times 100$)		RMSE	
		Δ_{mod}	Δ_{obs}	Δ_{mod}	Δ_{obs}	Δ_{mod}	Δ_{obs}	Δ_{mod}	Δ_{obs}
TMAX	DJF	1.53	3.45	1.88	0.66	3.19	5.16	1.84	0.57
	MAM	1.42	3.73	1.49	0.95	3.52	3.27	1.85	0.58
	JJA	3.92	4.23	1.56	1.20	2.51	2.57	1.70	0.62
	SON	1.77	2.99	1.60	0.90	2.02	3.08	1.39	0.57
TMIN	DJF	1.86	4.71	1.51	0.46	4.02	4.89	1.12	0.41
	MAM	2.50	4.07	0.76	0.44	3.91	5.14	1.17	0.38
	JJA	1.87	4.50	0.46	0.56	4.54	5.55	2.28	0.64
	SON	1.32	4.49	0.94	0.53	3.56	4.77	1.84	0.45
PRE	DJF	3.83	1.71	10.73	9.97	7.74	4.95	38.69	10.84
	MAM	6.94	8.21	9.39	17.05	11.50	4.23	60.76	10.66
	JJA	16.6	3.64	51.94	11.81	8.74	4.46	53.03	9.85
	SON	11.09	6.99	17.06	15.87	10.55	6.79	46.78	16.30

reproduce maximum and minimum temperature patterns. In the case of RMSE, the spread between model configurations is larger due to the systematic biases that different parametrizations introduce, and in particular *Jerez et al.* [2012] pointed out that the largest differences among parametrizations schemes were due to the PBL scheme. An exception is TMIN in summer, where differences in the observations are even greater than among models (the spatial-average of TMIN in SPAIN02 is half a degree colder than in the other two databases, which has an important impact in the calculation of RMSE, and hence Δ_{obs}).

[14] The same calculations have been applied to the MULTI-MODEL ensemble, and are shown in the last four columns in Table 1. Performing the same analysis in two ensembles allows for an assessment of the robustness of these findings, given their different nature. The results are consistent with the former analysis: for precipitation, the spread in models is larger than for observations, while for temperature it depends on which metrics has been employed. As before, the observations exhibit larger uncertainties for correlation in temperature variables, but lower when RMSE is employed. It is noteworthy that although the size of the MULTI-MODEL ensemble is smaller, the spread among the models is generally larger than in the MULTI-PHYSICS ensemble. This is due to the fact that the former includes not only different parametrization schemes, but also completely different RCMs and even slightly different domains although with the same spatial resolution.

4. Discussion

[15] Most methodologies developed so far for weighting simulations of climate change do not take into account explicitly the uncertainty contained in the observational database used as reference. This approach is based in the assumption that the observational uncertainty is small compared with inter-model variability [*Gleckler et al.*, 2008]. However, *Räisänen et al.* [2010] explored the sensitivity to the observational database in the weighting of global circulation models, and concluded that uncertainties in observations are not negligible in all circumstances. This fact has also been pointed out by other authors in a regional model context [*Christensen et al.*, 2010; *Lenderink*, 2010]. In general terms, it is expected that when the intermodel variability decreases, as is the

case when an ensemble is driven by the same “perfect” boundary conditions, the observation uncertainties becomes relevant.

[16] In this analysis, three different gridded observational databases over Spain are employed, which contain daily series of maximum and minimum temperature as well as precipitation. The aim is to rank the models in two independent ensembles of high-resolution climate simulations driven by “perfect” boundary conditions. In good agreement with previous attempts [e.g., *Räisänen et al.*, 2010; *Kjellström et al.*, 2010; *Lenderink*, 2010; *Lorenz and Jacob*, 2010; *Jerez et al.*, 2012], the analysis shows that it is not possible to identify a “best model”, given that the performance of the model depends on which season and variable is considered more relevant. More importantly, the ranking of the models is also sensitive to the choice of the observational database used to evaluate the simulations. Even considering the same variable and season, a model can be scored as one of the best/worst in the ensemble depending on what database is used as reference. Two spreads that allow the comparison of quantitative uncertainties within an ensemble with those embedded in the observations are defined. For precipitation, the spread among the models is larger than that among the observations in the two ensembles. However, for maximum and minimum temperatures, it turns out that uncertainties among observations are at least as relevant, when ranking models according to how well they reproduce the climatological values, as uncertainties among the models within an ensemble. The fact that the same conclusions can be drawn from both ensembles reinforces our findings. Indeed, the major difference between them is that the spreads between models are generally larger in the MULTI-MODEL ensemble, although it is smaller. This is due to the fact that their members not only consider different parametrization schemes, but complete RCMs and even slightly different domains.

[17] Despite its successful implementation in seasonal forecasting, the development of an objective, general and widely accepted methodology to weight climate change projections has turned out to be difficult. *Weigel et al.* [2010] explored the risks involved with a conceptual model, and demonstrated that although errors in equally weighted multimodels can be reduced by model weighting, if the real uncertainties are not properly taken into account more information may actually be lost than gained by optimum average. The comparison exercise

developed in this paper demonstrates that the uncertainties in the observational databases are not negligible in some circumstances, and that non uniform weighting should only be used with great care. In particular, sticking to a unique observational database to weight an ensemble of climate models according to how well they reproduce these observations is somewhat inappropriate, as this procedure propagates the uncertainties in the observations to the weighted result. The authors consider that new techniques should be implemented to overcome this problem, although finding a proper solution is beyond the scope of the present paper. For example the observational uncertainty arising from the existence of different observational data sets can be naturally included in probabilistic projections of regional climate change within a Bayesian framework [Tebaldi et al., 2005], although to our knowledge it has not been done yet.

[18] **Acknowledgments.** This work was supported by the projects SPEQTRES, CORWES (Spanish Ministry of Economy and Competitiveness and EU FEDER funds) and ESCENA (Spanish Ministry of the Environment). The authors acknowledge the E-OBS data set from the EU-FP6 project ENSEMBLES and the data providers in the ECA & D project, the University of Cantabria and AEMET for providing the observational datasets employed in this study, as well as the members of the ESCENA project for providing the model data. J. J. Gómez-Navarro thanks the funding from the PRIME2 project (priority program INTERDYNAMIK, German Research Foundation). S. Jerez thanks the Portuguese Science Foundation (FCT) for her financial support through the project ENAC (PTDC/AAC-CLI/103567/2008). P. Jiménez-Guerrero also acknowledges the Ramón y Cajal Programme by the Spanish Government.

[19] The Editor thanks the anonymous reviewer for his/her assistance in evaluating this paper.

References

- Christensen, J. H., E. Kjellström, F. Giorgi, G. Lenderink, and M. Rummukainen (2010), Weight assignment in regional climate models, *Clim. Res.*, *44*, 117–194, doi:10.3354/cr00916.
- Coppola, E., F. Giorgi, S. A. Rauscher, and C. Piani (2010), Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models, *Clim. Res.*, *44*, 121–134, doi:10.3354/cr00940.
- Dee, D. P., 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.
- Giorgi, F., and L. Mearns (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, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *J. Geophys. Res.*, *113*, D06104, doi:10.1029/2007JD008972.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.
- Herrera, S., J. Gutiérrez, R. Ancell, M. Pons, M. Frías, and J. Fernández (2010), Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02), *Int. J. Climatol.*, *32*, 74–85, doi:10.1002/joc.2256.
- Jacob, D., et al. (2007), An inter-comparison of regional climate models for Europe: Model performance in present-day climate, *Clim. Change*, *81*, Suppl. 1, 31–52, doi:10.1007/s10584-006-9213-4.
- Jerez, S., J. P. Montávez, P. Jiménez-Guerrero, J. J. Gómez-Navarro, R. Lorente, and E. Zorita (2012), A multi-physics ensemble of present-day climate regional simulations over the Iberian Peninsula, *Clim. Dyn.*, *39*(7–8), 1–24, doi:10.1007/s00382-012-1539-1.
- Jiménez-Guerrero, P., J. Montávez, M. Domínguez, R. Romera, L. Fita, J. Fernández, W. Cabos, G. Liguori, and M. Gartner (2012), Description of mean fields and interannual variability in an ensemble of RCM evaluation simulations over Spain: Results from the ESCENA project, *Clim. Res.*, in press.
- Kjellström, E., F. Boberg, M. Castro, J. Christensen, and G. N. E. Sánchez (2010), Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models, *Clim. Res.*, *44*, 135–150, doi:10.3354/cr00932.
- Lenderink, G. (2010), Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations, *Clim. Res.*, *44*, 151–166, doi:10.3354/cr00946.
- Lorenz, P., and D. Jacob (2010), Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40, *Clim. Res.*, *44*, 167–177, doi:10.3354/cr00973.
- Meehl, G. A., et al. (2007), Global climate projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 747–845, Cambridge Univ. Press, Cambridge, U. K.
- Nikulin, G., et al. (2012), Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations, *J. Clim.*, *25*, 6057–6078, doi:10.1175/JCLI-D-11-00375.1.
- Räisänen, J., L. Ruokolainen, and J. Ylhäisi (2010), Weighting of model results for improving best estimates of climate change, *Clim. Dyn.*, *35*, 407–422, doi:10.1007/s00382-009-0659-8.
- Tebaldi, C., R. Smith, D. Nychka, and L. Mearns (2005), Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, *J. Clim.*, *18*, 1524–1540.
- Uppala, S., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*(612), 2961–3012, doi:10.1256/qj.04.176.
- Weigel, A. P., R. Knutti, M. A. Liniger, and C. Appenzeller (2010), Risks of model weighting in multimodel climate projections, *J. Clim.*, *23*(15), 4175–4191, doi:10.1175/2010JCLI3594.1.
- Yolanda-Luna, M., A. Morata, M. Martín, S.-M. D., and J. dela Cruz (2008), Validación de la base de datos reticular de la AEMet: Temperatura diaria máxima y mínima, in *Cambio Climático Regional y Sus Impactos*, edited by J. Sigró Rodríguez, M. Brunet India, and E. Aguilar Anfrons, pp. 189–196, Artyplan, Tarragona, Spain.