Sistema de predicción por conjuntos a corto plazo sobre la Península Ibérica y Baleares

A short-range ensemble precipitation prediction system over the Iberian Peninsula and Balearics

Mª Yolanda LUNA[1], Ana MORATA[1] & Alvaro PASCUAL[3]

[1] Agencia Estatal de Meteorología (AEMET), Spain
dsantos@inm.es

[2] Departamento de Matemática Aplicada. Escuela Universitaria de Informática de Segovia.
Universidad de Valladolid, Spain.
mlmartin@eis.uva.es

[3] Departamento de Astrofísica y Ciencias de la Atmósfera
Universidad Complutense de Madrid, Spain
valero@fis.ucm.es

Received: 29 May 2009
Accepted: 16 September 2009

RESUMEN

Un sistema de predicción por conjuntos ha sido desarrollado para la generación de predicciones probabilistas a corto plazo del campo de precipitación sobre la Península Ibérica y las islas Baleares. El sistema está basado en la quinta generación del modelo desarrollado en colaboración entre la Universidad de *Pennsylvania State* y el *National Center for Atmospheric Research* (MM5). El sistema se fundamenta en la combinación de dos condiciones iniciales y de contorno provenientes de dos modelos globales, IFS-ECMWF y GFS-NCEP, con cinco configuraciones de parametrizaciones físicas de los fenómenos de sub-rejilla del modelo. De este modo se han obtenido diez miembros del sistema para el mes de Octubre del año 2006. La verificación mesoescalar del sistema se ha llevado a cabo mediante el contraste de las predicciones con los valores observados de precipitación de la Red Climática Española. El sistema de predicción por conjuntos muestra alta correlación dispersión-pericia para los valores de precipitación diaria. El histograma de rango o diagrama de Talagrand indica por su forma un comportamiento infradispersivo y afectado por sesgo del sistema. El área bajo las curvas R.O.C. indica una buena capacidad de discriminación del sistema. Los diagramas de fiabilidad revelan una buena fiabilidad del sistema, sugiriendo, en general, un buen grado de acuerdo entre las probabilidades predichas y las frecuencias promedio observadas. Por todo ello, el sistema se muestra como una herramienta útil para la predicción de la precipitación en el área de estudio.

Palabras clave: Predicción por conjuntos; prediccion a corto plazo; precipitación; Península Ibérica; Baleares.

ABSTRACT

A short-range ensemble precipitation forecast system has been constructed over the Iberian Peninsula and Balearics by means of the fifth-generation Pennsylvania State University-National Center for Atmospheric Research Model (MM5). The ensemble system consists of ten members, each run with a different combination of two different initial conditions from global models, IFS-ECMWF and GFS-NCEP, and five different subgrid-scale physics configurations for one month period, October 2006. The mesoscale verification is made by using observational precipitation data of the Spanish Climatic Network. The created short-range ensemble shows high spread-skill correlation values for daily precipitation. However, the asymmetric shape of the rank histogram indicates some underdispersion, suggesting a biased behaviour. The Talagrand shows as well the underdispersive effect because of its asymmetric distribution. The Relative Operating Characteristic curve shows a very outstanding area indicating the good discrimination capacity. The reliability diagrams are also indicative of the good reliability of the forecasting system, depicting in general good agreement between forecast probability and the mean observed frequency. Because of that, the verification proves the usefulness of the forecasting system over the study area.

Key words: Ensemble forecasting; short-range prediction; precipitation; Iberian Peninsula; Balearics.

SUMMARY: 1. Introduction. 2. Precipitation data. 3. Description of the short-range ensemble prediction system. 4. Results. 5. Summary and conclusions. 6. Acknowledgements. 7. References

1. INTRODUCTION

The occurrence and location of heavy precipitation events can be improved by using numerical weather forecasting in the short-range prediction. Although deterministic numerical models are able to reasonably accurately give forecasts, unknown error sources remain, reducing the confidence of forecasters in the simulated fields. Such sources can be atmospheric forcings imperfectly included in the model, scarce knowledge of the initial conditions including instrument error, sampling error, and initialization error, and limitations related with the non-linearity of the atmospheric system (Stenrud et al., 2000). Aditionally, small errors or perturbations in the model initial conditions are amplified as the forecast period grows, leading to noticeable differences in the forecast. Therefore, the forecast can be improved if the generation of several predictions based on slightly different initial conditions with the same probability is considered (Toth and Kalnay, 1997). An ensemble forecasting approach could provide an improvement of the skill when comparing with an individual deterministic one, providing as well a measure of the forecast uncertainty and of its reliability. The helpfulness and usefulness of an Ensemble Prediction System (EPS) resides generally in the diversity of possible solutions offered to a given meteorological forecast problem by the system. Differences characterized by each member of the EPS, related to the variance within the ensemble forecasts, entail study of the spread-skill relationship. If the forecasts are coherent, i.e., with small spread, the atmosphere is in a more predictable state than if the forecasts diverge or present large spread. So, if results show low ensemble variance, the system presents higher skill and consequently better confidence in the forecasts.

In this work, a short-range ensemble prediction system, based on a set of mesoscale models with different subgrid-scale physic schemes and two different initial conditions, is developed over the Iberian Peninsula and Balearics. A ten members short-range ensemble forecast system has been constructed as a result of combining two different initial conditions from global models and five different physics configurations of the non-hydrostatic Mesoscale Model (MM5, version 3). The ensemble simulations have been investigated for precipitation during October 2006. The quality and value of precipitation forecasts have been evaluated against observations of the Spanish Climatic Network. Comparisons between the ensemble system and observations provide an overview of the mean model performance and forecast variability besides of an evaluation of the ensemble accuracy. Additionally, ensemble probability distribution functions for precipitation will provide information on uncertainty of ensemble forecast. Both spatial and probabilistic approaches will be used to verify both each individual ensemble member and the ensemble mean evaluations.

The organization of the paper is as follows. Section 2 describes the precipitation dataset used to verify the ensemble precipitation forecasts. Section 3 presents a review of the constructed ensemble system, with a description of the model set-up perturbations used to generate the different ensemble members. The skill of the ensemble system in forecasting precipitation is examined in Section 4. Finally, main conclusions of results are drawn in Section 5.

2. PRECIPITATION DATA

The quality and value of precipitation forecasts will be evaluated against observations of the Spanish Climatic Network. Comparisons between the ensemble system and observations provide an overview of the mean model performance and forecast variability besides of an ensemble accuracy evaluation. In this paper, the data used for precipitation verification comes from a high-resolution daily precipitation data base derived from in-situ measurements coming from the station network of the Spanish Meteorological Service (Agencia Estatal de Meteorología, AEMET). The AEMET Climatological Area has elaborated an Iberia daily precipitation database by means of statistical spatial interpolation of more than 4000 in-situ measurements onto a regular grid. The purpose was to build a complete daily dataset necessary as input of spatially distributed models and for understanding the climate variability at daily scale. All the in-situ measurements of daily precipitation data from the Historical Database of AEMET were extracted for the period 1 January 1961 to 31 December 2008, regardless of their time coverage. Thus, the number of available stations depended on the date. These stations, irregularly distributed over the Iberian Peninsula and the Balearics, have provided good coverage over the domain. In order to complete the observations in Portugal, data from the

European Climate Assessment & Dataset project (ECA&D) were used (Haylock et al., 2008).

The AEMET choses a 25 km regular grid because of the space scale suitable needed for risk analysis models and for climatic variability studies, including evaluation of climate change impacts in Spain. The Kriging method was used to interpolate daily precipitation. This interpolation technique preserves more variance than other methods such as the inverse distance weighting method. Moreover, this spatial interpolation tool has widespread been used in software related to Geographical Information Systems, allowing comparisons with databases from different countries. All these considerations are outlined in the COST Action 719 (The Use of GIS in Meteorology and Climatology) in which the AEMET had been an active participant. The precipitation database has already successfully been used by the authors in several studies to validate other new precipitation datasets (Sotillo et al., 2006; Morata et al., 2008), giving as well as both characterization of the rainfall regime and evaluation of the potential improvement of such new precipitation database versus current datasets. Further information about this precipitation data set can be found in Luna and Almarza (2007).

3. DESCRIPTION OF THE SHORT-RANGE ENSEMBLE PREDICTION SYSTEM

In order to generate a short-range EPS, the non-hydrostatic Mesoscale Model (MM5, version 3), developed by the Pennsylvania State University-National Center of Atmospheric Research (Anthes and Warner, 1978; Grell et al., 1994) has been chosen to be used in this study with five combinations of the model parameterizations and two initial boundary conditions.

All the ensemble members use a terrain following σ-coordinate system with 30 vertical levels, being vertical resolution enhanced in the lower troposphere in order to get a more accurately representation of boundary layer processes. The spatial domain, which comprises the whole Iberian Peninsula as well as the most western side of the Mediterranean basin, is covered with a 30-km horizontal grid spacing in the coarse domain and 10-km for the fine domain (Fig. 1). Depending on the model member, time step has been chosen from 35 s till 240 s. Two different data have been used as initial and boundary conditions to generate the ensemble members: the IFS-ECMWF data (0.5° x 0.5° grid spacing, 21 isobaric vertical levels) and GFS-NCEP analysis and forecasts (1.0° x 1.0° grid spacing, 26 isobaric vertical levels). Run processes were obtained by bilinear interpolation to the MM5 grid, i.e. with two-way nesting which allows flow between high-to-low resolution grid effects, permitting realistic terrain features. The period was simulated by means of a daily single run (starting at 00 UTC) of each ensemble member with a 36-hour forecast horizon. Therefore, area, output runs and forecast length were selected in order to avoid boundary conditions effects, maintaining, at the same time, the capability of generating probabilities.

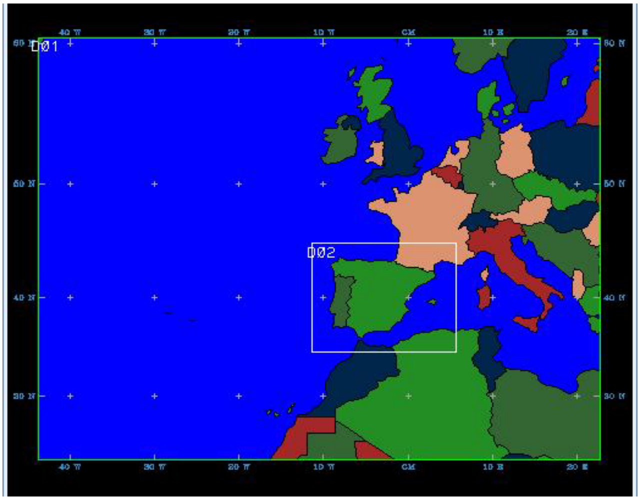
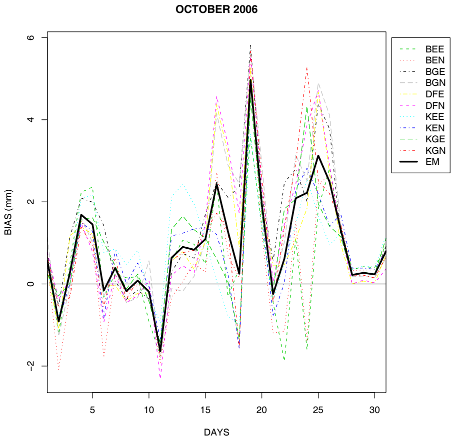


Figure 1. Geographic coverage of the MM5 short-range ensemble simulations.

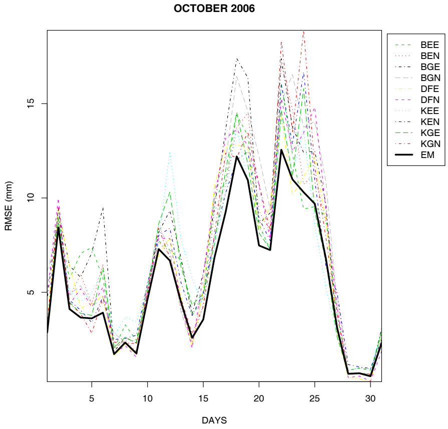
Physical uncertainties were incorporated into the ensemble by using different physical parameterization schemes, changing both convection and planetary boundary layer parameterization schemes and thus generating plausible and realistic solutions to the predictability problem. Two different convective parameterization schemes besides the default scheme are used in the ensemble to calculate moist convection effects on the domain: the Betts-Miller parameterization scheme (Betts and Miller, 1986) and the Kain-Fritsch scheme (Kain and Fritsch, 1990). To incorporate planetary boundary layer physical uncertainties, two PBL schemes in addition to the default scheme have been chosen: Mellor-Yamada Janjic (Janjic and Zavisa, 1994) and Gayno-Sheman scheme (Shafran et al., 2000). On the other hand, combinations of different physical model parameterizations (microphysic and land surface schemes) have produced more disperse probability density functions, even using the same model. Taking this into account and due to compatibility restrictions between the several parameterizations, two microphysical schemes, Reisner and Goddard, other than the default (Reisner et al., 1998) and two land surface model schemes (Chen and Dudhia, 2001) have been combined with the cumulus and PBL schemes.

4. RESULTS

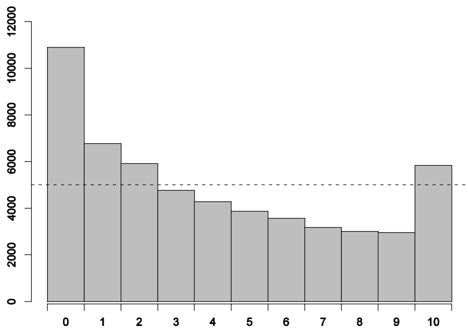
In this section the performance of the simulations obtained from the 10-members ensemble for October 2006 for the fine domain is illustrated, showing the results associated with both spatial and probabilistic verification methods. Figure 2 displays the mean bias and rmse of the 24-h precipitation over a regular grid (25 km x 25 km) in order to provide rainfall data at a resolution compatible with the observational data. Thus, an interpolation algorithm is used in which data located within a 5-km square, centered at each nearby grid point over the same period, were averaged to represent the rainfall amount at that grid point. Throughout the event simulation period, bias and rmse results (Figs. 2a and b) show similar evolution to the ensemble simulations. Time evolution of both bias and rmse show evidence of different behavior over the study period. In terms of the total daily precipitation, spatially averaged over the region, it can be noted wet bias throughout the event period except for particular days with negative bias results (Fig. 2a). The departures in the different stages of the period could indicate some rainfall over/underprediction. The ensemble mean provides a good forecast when it is compared with some ensemble members; in fact, when the time period is enough large, the ensemble mean constitutes the best forecast when is compared with any ensemble member. In this work, the ensemble mean offers the best forecasts when it is compared with any ensemble member used to build the EPS. The time evolution is similar for all ensemble members.



a)



b)



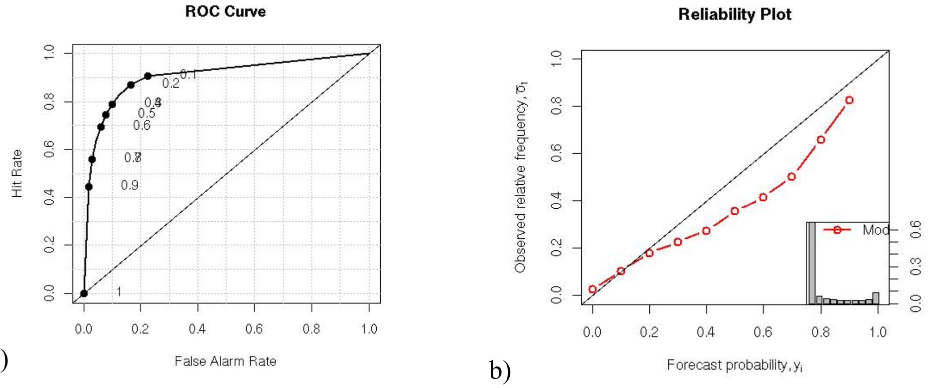
c)

Figure 2. Temporal evolution of bias (a) and rmse (b) for each ensemble member and the ensemble mean of the total daily precipitation, spatially averaged over the region. Units are in mm/m2.; (c) Talagrand diagram showing the frequencies at which the verifying analysis falls in each category, defined by the 10 ordered ensemble members at each grid points. Dashed line represents the equiprobability line.

The relationship between the rmse and the spread of the ensemble mean is quite close to the ideal diagonal with a correlation value of 0.86. Here, the whole period shows a spread value shorter than the ensemble mean error, indicating an under-dispersive behaviour.

The Verification Rank Histogram or Talagrand distribution is the histogram of frequencies of the rank of the observed data within the forecast ensemble. That is, Talagrand counts where the verifying observation falls with respect to the ensemble forecast data, which is arranged in increasing order at each grid point. In an ensemble with perfect spread, each member represents an equally likely scenario, so the observation is equally likely to fall between any two members. Although a rank histogram measures whether the observed probability distribution is well represented by the ensemble, a flat Talagrand does not necessarily indicate a skilled forecast. Nevertheless a good EPS should have a uniform distribution, representing correctly the forecast uncertainty. Thus, the observed data set should uniformly be distributed among the ensemble members. Then, in an ideal EPS, the Talagrand should be flat with many verifications regularly in each interval, meaning exchangeability between deterministic predictions and observed data. A distribution can present a U inversed-shaped, indicating a too large ensemble spread in which many observations falling near the centre of the ensemble; On the contrary, if the distribution presents slightly U–shaped, some cases can be over-represented, falling the verification outside the ensemble and, other cases can be under-represented when verification is located in the ensemble centre. That is, a U-shaped would indicate ensemble spread too small with many observations falling outside the extremes of the ensemble. For some parameters such U-shape degenerates into an asymmetric shape, indicating the presence of bias in the system for such parameter. Taking into account the different abovementioned shapes that a Talagrand can show, Fig. 2c depicts an asymmetric distribution in which the first bin is the most frequent. Therefore, the ensemble contains bias. If the null values, on the other hand so frequently in precipitation, are not considered, it can be observed some underdispersion by the EPS, i.e., the verifying observation falls outside the envelope forecasts generated by the ensemble members. Without bias correction, an overpopulation of the extreme ranks of the histogram can be noted, suggesting this asymmetric shape a bias behaviour. This situation is nowadays accurate by applying statistical postprocesses with a bias correction and further generation of a flat histogram or members equally probability.

The ability of the forecast in discriminating between two alternative outcomes (events and non-events (discrimination)) is measured by the use of the Relative Operating Characteristic or ROC. Plotting hit rate versus false alarm rate in a set of varying probability thresholds helps to make the yes or no decision. The closer the point is to the upper left corner in the plot indicates a measure of the higher of the skill, giving the area below the ROC curve a forecast skill measure. Therefore, the ROC provides more information that can be summarized in a single data, the area. The greater is the area means the good system discrimination. On the contrary, the lower left corner, where both hit and false alarms rate are zero, represents a system which never warns of an event and the upper right corner, represents a system where the event never occurs. An imperfect system will give results with values on a long convex curve tending to the upper-left corner on the ROC curve, enabling comparisons between a probabilistic and a deterministic forecast system. Therefore, if the curve is close to upper left corner, the system presents good discrimination. Here, results from the ensemble indicate that the ensemble system is reasonably accurate using various threshold values for precipitation. Although the Talagrand (Fig. 2c) showed some bias as it has already been mentioned, the good ROC curve (Fig. 3a) tends to the upper left corner, indicating the greater forecast quality of the system, that is, the usefulness of the forecasting system. It is worth to note the very outstanding area presented in the ROC curve, pointing out again the good discrimination of the EPS.



a)

b)

Figure 3. (a) ROC Curve precipitation; (b) Reliability diagram, over 5mm/24h.

On the other hand, the reliability diagram measures how well the predicted probabilities of an event correspond to their observed frequencies (reliability). In this diagram, the observed frequency is plotted against forecast probability for all probability categories. Thus, a line close to diagonal represents good reliability; on the contrary, deviation from diagonal indicates conditional bias: below diagonal represents forecast probabilities too high; above diagonal, forecast probabilities too low. In this work, Figure 3b displays, in general, a good agreement between forecast probability and the mean observed frequency, except for certain forecast probability values ranging about 50%. The sharpness diagram shows the frequencies, having greater contribution to the resolution component (the 0.0 and 1.0 probability) and little contribution the middle ones; that is the reason of the good reliability value of the system.

5. SUMMARY AND CONCLUSIONS

In this paper an overview of a short-range EPS in the Mediterranean area for a 1-month period of 2006 is shown. The ensemble system consists of ten members at 10-km grid spacing from the MM5, with different combination of two different initial conditions from global models and five different subgrid-scale physics configurations, changing both convection and planetary boundary layer parameterization schemes and thus generating plausible and realistic solutions to the predictability problem. In order to evaluate the ensemble accuracy, comparisons between ensemble system forecasts and observations have been made, providing an overview of the mean model performance and forecast variability. The observational dataset consists of a high-resolution daily precipitation data base derived from in-situ measurements coming from the station network of the Spanish Meteorological Service.

The mean model performance and forecast variability, together with the evaluation of the ensemble accuracy have shown a good EPS performance. The main conclusions that can be extracted form this work are:

- Temporal evolution of the mean bias and rmse of the 24-h precipitation over a regular grid have shown similar evolution to the ensemble simulations.
- The ensemble mean offers the best forecasts when it is compared with any ensemble member used to build the EPS.
- It is noticeable the high spread-skill correlation values (> 0.80) for daily precipitation, but with small spread values, indicating that the EPS is affected by underdispersion.
- The Talagrand have shown an asymmetric distribution, pointing out the presence of bias in the ensemble system and showing againg the underdispersive effect noted in the spread-skill.
- The ROC curve tends to the upper left corner, pointing the existence of no false alarms and only hits. It is highlited the very great area shown by the

ROC curve, being suggestive of a greater forecast quality of the EPS, indicating that the ensemble system is useful.

- Reliability has shown a good agreement between forecast probability and the mean observed frequency, except for certain forecast probability values ranging about 50%. Frequencies with high contribution to the Brier score reliability component and those ones with low contribution, i.e., the middle ones, are shown in the sharpness.

All verification scores used in this work have generally exhibited hopeful results, encouraging the extension of this preliminary research to other verification periods and studying cases and the possible inclusion of new members using WRF-NMM and WRF-ARW models. In that aim, works in extending this study to other larger samples are in progress.

6. ACKNOWLEDGEMENTS

This work has been partially supported by the research projects CGL2007-61328/CLI and UE *Safewind* G.A. No. 213740. The authors wish to thank the Spanish Meteorological Agency (AEMET: Agencia Estatal de Meteorología) for providing the precipitation dataset; the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA from their Web site at http://www.cdc.noaa.gov for providing the GFS-NCEP analysis and forecast data and the European Centre for Weather Medium Forecast (ECWMF) for providing the IFS-ECMWF analysis and forecast data. Authors also want to acknowledge the information related to MM5 model that UCAR provides online at http://www.mmm.ucar.edu/mm5.

We do not want to finalize the wording of this article without a memory to the memory of Elvira Zurita, whose absence, although still short on time, becomes ever deeper into our hearts. Elvira, how long we stayed together in this home, and we can recall no time when our endearing friendship would have been minimally disturbed despite some vicissitudes lived jointly! You always will stay in our hearts.

No queremos finalizar la redacción de este artículo sin un recuerdo emocionado a la memoria de la profesora Elvira Zurita, cuya ausencia, aunque breve todavía en el tiempo, se torna cada vez más profunda en nuestro corazón. ¡Elvira, cuanto tiempo hemos permanecido juntos en esta casa sin que podamos recordar ningún momento en que nuestra entrañable amistad se hubiera visto minimamente perturbada a pesar de los múltiples avatares vividos conjuntamente!. Siempre permanecerás en nuestros corazones.

7. REFERENCES

ANTHES, R.A. & TT. WARNER (1978). Development of hydrodynamic models suitable for air pollution and other mesometeorological studies. *Mon. Wea. Rev.*, 106, 1054-1078.

BETTS, A.K. & M.J. MILLER (1986). A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX,ATEX and Artic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, 112, 693-709.

CHEN, F. & J. DUDHIA (2001). Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, 129, 569-585.

GRELL, G.A., J. DUDHIA & D.R. STAUFFER (1994). Diagnosing coupled jet streak circulations for a northern plains snow band from the operational nested-grid model. *Wea. Forecast.,* 7, 26-48.

HAYLOCK, M.R., N. HOFSTRA, A.M.G. KLEIN TANK, E. J. KLOK, P.D. JONES & 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.

JANJIC, J. & I. ZAVISA (1994). The step-mountain eta coordinate model: Further development of the convection, viscous sublayer, and turbulent closure schemes. *Mon. Wea. Rev.*, 122, 927-945.

KAIN, J.S. & J.M. FRITSCH (1990). A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.,* 47, 2784-2802.

LUNA, M.Y. & C. ALMARZA (2007). Interpolation of 1961-2002 daily climatic data in Spain. In *Spatial Interpolation in Climatology and Meteorology.* Ed. by S. Szalai, Z. Bihari, T. Szentimrey and M. Lakatos. COST Office (ISBN 92-898-0033-X), 189-196.

MORATA, A., M.L. MARTÍN, M.G. SOTILLO, F. VALERO & M.Y. LUNA (2008). Iberian autumn precipitation characterization through observed, simulated and reanalysed data. *Advances in Geosciences*, 16, 49-54.

REISNER, J., R.M. RASMUSSEN & R.T. BRUINTJES (1998). Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.,* 124, 1071-1107, 1998.

SHAFRAN, P.C., N.L. SEAMAN & G.A. GAYNO (2000). Evaluation of numerical predictions of boundary layer structure during the Lake-Michigan Ozone Study. *J. Appl. Meteor.*, 39, 412-426.

SOTILLO, M.G., M.L. MARTIN, F. VALERO & M.Y. LUNA (2006). Validation of an homogeneous 41-year (1961-2001) winter precipitation hindcasted dataset over the Iberian Peninsula: assessment of the regional improvement of global reanalysis. *Climate Dynamics*, 27, 627–645.

STENSRUD, D.J., J.W. BAO & T.T WARNER (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, 128, 2077-2107.

TOTH, Z. & E. KALNAY (1997). Ensemble forecasting at NCEP: The breeding method. *Mon. Wea. Rev.*, 125, 3297-3318.