Advances in Geosciences

Verification of a short-range ensemble precipitation prediction system over Iberia

D. Santos-Muñoz[1], M. L. Martin[2], A. Morata[1], F. Valero[3], and A. Pascual[3]

[1]Agencia Estatal de Meteorología, Leonardo Prieto Castro, 8, 28040 Madrid, Spain
[2]Escuela Universitaria de Informática de Segovia, Universidad de Valladolid, Pza. Sta. Eulalia, 9–11, 40005 Segovia, Spain
[3]Dpto. Astrofísica y CC. de la Atmósfera, Facultad de CC. Físicas, Universidad Complutense de Madrid, Madrid, Spain

Abstract. The purpose of this paper is the verification of a short-range ensemble prediction system (SREPS) built with five different model physical process parameterization schemes and two different initial conditions from global models, allowing to construct several versions of the non-hydrostatic mesoscale MM5 model for a 1-month period of October 2006. From the SREPS, flow-dependent probabilistic forecasts are provided by means of predictive probability distributions over the Iberian Peninsula down to 10-km grid spacing. In order to carry out the verification, 25 km grid of observational precipitation records over Spain from the Spanish Climatic Network has been used to evaluate the ensemble accuracy together with the mean model performance and forecast variability by means of comparisons between such records and the ensemble forecasts. This verification has been carried out upscaling the 10 km probabilistic forecast to the observational data grid. Temporal evolution of precipitation forecasts for spatial averaged ensemble members and the ensemble mean is shown, illustrating the consistency of the SREPS. Such evolutions summarize the SREPS information, showing each of the members as well as the ensemble mean evolutions. The Talagrand diagram derived from the SREPS results shows underdispersion which indicates some bias behaviour. The Relative Operating Characteristic (ROC) curve shows a very outstanding area, indicating potential usefulness of the forecasting system. The forecast probability and the mean observed frequency present good agreement with the SREPS results close to the no-skill line. Because the probability has a good reliability and a positive contribution to the brier skill score, a positive value of this skill is obtained. Moreover, the probabilistic meteogram

of the spatial daily mean precipitation values shows the range of forecast values, providing discrete probability information in different quantile intervals. The epsgram shows different daily distributions, indicating the predictability of each day.

1 Introduction

Precipitation over the Iberian Peninsula has a strong seasonal character. Particularly in some regions, over the eastern flank of the Iberian Peninsula, the maximum seasonal precipitation occurs in autumn or spring. The sparse summer rainfall in most part of the Peninsula depends on local factors and is mainly caused by convective storms associated with ground heating, moisture content and upper instability (Sumner et al., 2001; Valero et al., 2004). The occurrence and location of heavy precipitation events can be improved by using numerical weather forecasting in the short-range prediction (Shuman, 1989). Although deterministic numerical models are able to reasonably accurately give forecasts, unknown error sources remain, reducing the confidence of forecasters in the simulated fields. Small errors or perturbations in the model initial condition are amplified as the forecast period grows, leading to noticeable differences in the forecast. So, the generation of several predictions, based on slightly different initial conditions with the same probability, can improve the forecast (Leith, 1974; Hoffman and Kalnay, 1983; Toth and Kalnay, 1997). An ensemble forecasting approach could provide an improvement of the skill when comparing with an individual deterministic one. At the same time, the ensemble forecasting provides a measure of the forecast uncertainty and of its reliability. The ensemble forecasting method aims to predict the probability of future events as accurately as possible (Epstein, 1969; Mullen and Baumhefner, 1988).

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. 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 PSU/NCAR Mesoscale Model (MM5, version 3). The ensemble simulations have been investigated for precipitation field during October 2006. The quality and value of precipitation forecasts have been evaluated against observations of the Spanish Climatic Network. 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 structure of the paper is as follows: Sects. 2 and 3 present a review of the constructed ensemble system with a description of the model set-up perturbations used to generate the different ensemble members and the precipitation dataset used to verify the ensemble precipitation forecasts, respectively. The results of the ensemble system verification in forecasting precipitation are examined in Sect. 4. Finally, main conclusions of results are drawn in Sect. 5

2 Short-range ensemble prediction system

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 in this study to generate a short-range EPS with five combinations of the model parameterizations and different initial and boundary conditions.

All the ensemble members use a terrain following the σ-coordinate system with 30 vertical levels, enhancing the vertical resolution in the lower troposphere in order to get a more accurately representation of boundary layer processes. The spatial coverage of the coarse 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. A 10-km grid covering the Iberian Peninsula for the fine domain (Fig. 1a) is used to enhance the model horizontal resolution. Fig. 1b shows the orography of the inner grid of simulation. Depending on the model member, time step has been chosen from 35 s till 240 s. The simulations were generated two-way nesting between models domains which allows flow between high-to-low resolution grid effects, permitting a more realistic terrain features and atmospheric energy fluxes. Thus, two different sets of analysis and forecasts data bases have been used as initial and boundary conditions to generate the ensemble members: the ECMWF-IFS data from the European Centre for Medium-
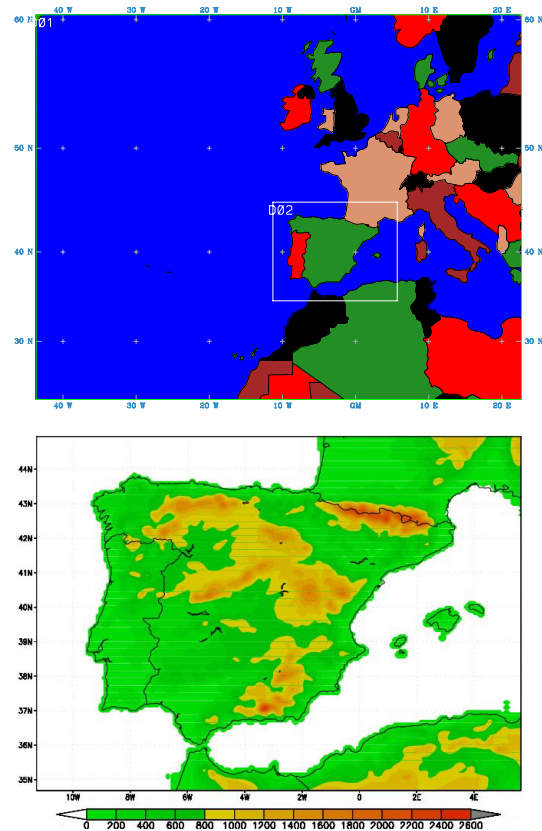

Fig. 1. (a) Geographic coverage of the two domains (coarse and fine) of the MM5 short-range ensemble simulations. **(b)** Fine domain corresponding with the Iberian Peninsula with the orography of the model detailed.

Range Weather Forecast (ECMWF) Improved Forecast System (IFS) ($0.5° \times 0.5°$ grid spacing, 21 isobaric vertical levels) and the NCEP data from National Center for Environmental Predictions (NCEP) Global Forecast System (GFS) ($1.0° \times 1.0°$ grid spacing, 26 isobaric vertical levels). The selected time period, October 2006, was simulated by means of a daily single run (starting at 00:00 UTC) of each ensemble member with a 36-h forecast horizon. On the other hand, the 24-h daily forecast precipitation has been estimated by using forecast data of 30-h minus forecast data of 6-h. Therefore, area and forecast length were selected in order to avoid boundary conditions effects, maintaining, at the same time, the capability of generating probabilities.

Physical uncertainties were incorporated into the ensemble by using different physical parameterization schemes. Details of the physics used to build up the two sets of 5 different ensemble members are listed in Table 1. The different model configurations chosen to create the physics ensemble are built changing both convection and planetary boundary layer parameterization schemes, generating plausible and realistic solutions to the predictability problem (Wang and Seaman, 1997). Two different convective parameterization schemes besides the default scheme are used in the ensemble

Table 1. Combinations of the MM5 parameterizations selected to generate the short-range ensemble prediction system.

NAME	CUMULUS	PBL	MICROPHYSICS	RADIATION	LAND SURFACE	SOIL LEVELS
DFE	Grell	MRF	SIMPLE ICE	CLOUD	5 LAYER	7,28,100,255
BEE	Betts-Miller	MYJ	REISNER	RRTM	NOAH	7,28,100,255
KEE	Kain-Fritsch	MYJ	REISNER	RRTM	NOAH	7,28,100,255
BGE	Betts-Miller	GAYNO-SEAMAN	GODDARD	RRTM	5 LAYER	7,28,100,255
KGE	Kain-Fritsch	GAYNO-SEAMAN	GODDARD	RRTM	5 LAYER	7,28,100,255
DFN	Grell	MRF	SIMPLE ICE	CLOUD	5 LAYER	10,40,100,200
BEN	Betts-Miller	MYJ	REISNER	RRTM	NOAH	10,40,100,200
KEN	Kain-Fritsch	MYJ	REISNER	RRTM	NOAH	10,40,100,200
BGN	Betts-Miller	GAYNO-SEAMAN	GODDARD	RRTM	5 LAYER	10,40,100,200
KGN	Kain-Fritsch	GAYNO-SEAMAN	GODDARD	RRTM	5 LAYER	10,40,100,200

to calculate moist convection effects on the domain: the Betts-Miller parameterization scheme (Betts, 1986; Betts and Miller, 1986) and the Kain-Fritsch scheme (Kain and Fritsch, 1990). To incorporate Planetary Boundary Layer (PBL) physical uncertainties, two PBL schemes in addition to the default scheme have been chosen: Mellor-Yamada Janjic (MYJ, Janjic and Zavisa, 1994) and Gayno-Seaman scheme (Ballard et al., 1991; 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 other than the default (Lin et al., 1983; Tao and Simpson, 1993; Reisner et al., 1998) and two land surface model schemes (Dudhia, 1996; Chen and Dudhia, 2001) have been combined with the cumulus and PBL schemes. All these parameterization combinations have been chosen after several tests trying to capture the model uncertainties on the processes related to generation, growing and deposition of water in the atmosphere, which finally drives to the model forecasted precipitation.

3   Verification data base and skill scores

The quality and value of precipitation forecasts have been evaluated against observations of the Spanish Climatic Network. Comparisons between the ensemble system and ob-

servations provide an overview of the mean model performance and forecast variability besides an evaluation of the ensemble accuracy. In this paper, the data used for precipitation verification comes from a daily precipitation data base derived from in-situ measurements coming from the high-resolution 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.

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 (Shen et al., 2001). This technique has the advantage of using the daily observed available data in spite of variations on the number of stations. Moreover, this spatial interpolation tool has been used extensively in software related to Geographical Information Systems, allowing comparisons with databases from different countries. All these considerations are outlined in the European Cooperation in the Field of Scientific and Technical Research, (COST Action 719, The Use of GIS in Meteorology and Climatology) in which the AEMET has been an active participant (Dyras et al., 2005). The precipitation database has already been used by the authors in several studies to validate other new precipitation datasets (Sotillo et al., 2006; Morata et al., 2008), giving both characterization of the rainfall regime as well as an 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).

In this paper, the time period to analyze the EPS performance has been selected taking into account the observed precipitation database of AEMET. Fig. 2a shows the evolution of the precipitation anomalies respect of the climatological mean, in three different zones in the Iberian Peninsula (Morata, 2004) from 1991 to 2008. In the EPS generation, forecast data from the NCEP have been used, being available from 2005. Therefore for verification purposes, the biggest rainfall anomaly in the three areas corresponds to October 2006 as it can be seen in Fig. 2b. During the first part of October 2006, several lows located at northwestern Iberia with associated fronts affected the Iberian weather. Thus, the nature of precipitation in the first part of the selected period, that affected the western part of Iberia, was mainly of frontal character. On the contrary, the rainfall of the last part of the time period showed some convective characteristics due to a through situated over North Algeria in juxtaposition with a ridge at upper level involving the western Mediterranean area.

In this paper, skill scores, both spatial and probabilistic, will be used to assess the accuracy and skill of precipitation forecasts. Standard verification methods based on a point by point comparison such as bias or root mean square error are used in this work over grid-point neighbourhoods of different areas, defining a scale associated with the verification: as scale increases, forecast and observational fields tend to be subjected to a filtering process. The bias, defined as the difference between a model-derived dataset and the observational data, and the time root mean square error have been derived at single points and spatially averaged values of 24-h precipitation over the grid have been subsequently computed. In order to provide rainfall data at a resolution compatible with the abovementioned observational data, a regular grid of 25 km×25 km has been selected.

Taking into account verification probabilistic tools, a more detailed way of analyzing the EPS spread is to construct a rank histogram or the so called Talagrand diagram
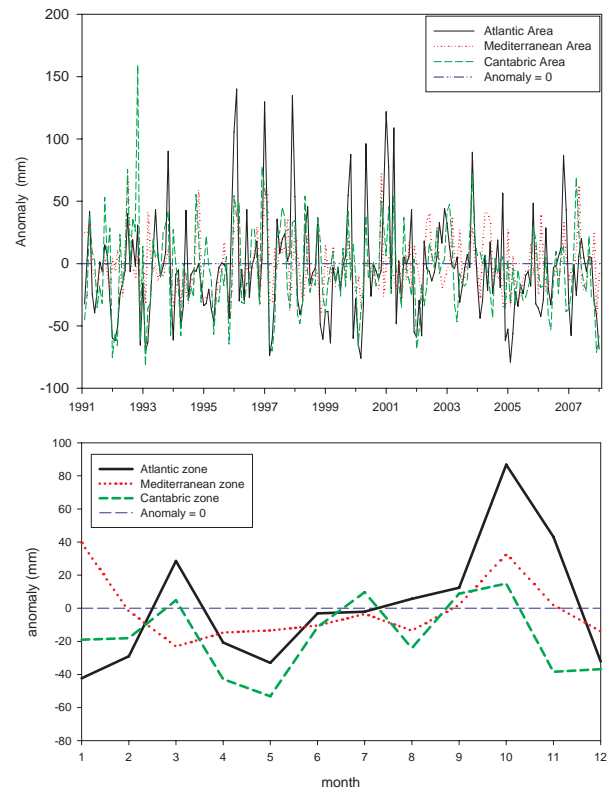

Fig. 2. Time evolution of precipitation anomalies in the Atlantic, Cantabric and Mediterranean zones of Iberia: **(a)** from 1991 to 2008; **(b)** the year 2006.

(Talagrand et al., 1997). In a perfect forecast system, the verifying analysis is equally likely positioned between any two ordered adjacent members, including cases when the analysis lies outside the ensemble range on either side of the distribution. Firstly, for each location, the forecast values of all ensemble members are sorted by increase order. Second, m+1 bins are made, m being the total number of ensemble members (in our case m=10), with each bin representing the range of nearby members. The first and last bins are representative of the values lower than the smallest member and greater than the largest member, respectively. The bins are examined, in order to see if the observed data fall into a given bin. Generally, in a good ensemble forecast system, all members should have equal ability to capture the observations, thus being the Talagrand diagram flat.

A very useful and powerful way for assessing the ability of probabilistic forecasts to discriminate dichotomous events is the Relative Operating Characteristic or ROC (Swets, 1973; Mason, 1982) based on contingency tables for different thresholds. Using different thresholds as criteria, a set of different contingency tables can be constructed for a probabilistic forecast that can range from 0%, representing an always forecast event, to 100%, indicating a never forecast event. Moreover, the ensemble forecast probabilistic performance can be also evaluated following the difference between a

forecast probability distribution and the observed probability distribution, using the Brier score (Brier, 1950). The Brier score can be decomposed on three sums (Murphy, 1993) called reliability, resolution and uncertainty.

If the sample climatology is used, the Brier skill score can be defined as:

$$BSS = \frac{BS_{Resolution} - BS_{Reliability}}{BS_{Uncertainty}} \qquad (1)$$

The different Brier skill score contributions can be represented on the attribute diagram. In this diagram, the curve is the named reliability curve which shows the relationship between the forecast probability and the observed frequency. Actually, the reliability is the area between the diagonal line (y=x) and the reliability curve, being a small area at initial time and becoming larger later on. The resolution could be defined as the ability of the forecast system in creating different probability distributions for different forecast values. If the sample climatology is used as a forecasting system, the forecast would not have resolution because it does not discriminate at all between events and non-events. Thus, the frequency of the forecasted event, obtained from the sample climatology, is represented over the diagram as horizontal and vertical lines labelled as "no resolution" lines. The bisectrix line between the diagonal and "no resolution" line defines the "no skill" line. Points between the "no skill" line and the diagonal contribute positively to the Brier skill score. The uncertainty is defined as the variability of the observations and it is measured by the variance of the sample climatology. The greater the uncertainty, the more difficult the forecast will tend to be. The uncertainty is related with the observations and it is impossible to improve by the forecast system. Another measurement related to the marginal distribution of the forecasts is the sharpness which quantifies the ability of the forecasts in deviating from the mean climatological value for deterministic forecast or from climatological mean probabilities for probabilistic forecasts. In a calibrated forecast, sharpness is related to resolution, or ability of the forecast to be sorted into subsamples where the verifying event is different (Wilks, 1995).

Finally, another way to represent the ensemble variance is used in this paper: the probabilistic meteogram or epsgram. The ensemble system information can also be exposed on box-and-whiskers plots, displaying the ensemble distribution of the forecast. These plots are represented by rectangles or boxes with upper and lower values corresponding to 75% and 25% percentiles, i.e., the corresponding third and first quartiles, respectively; the forecast median is also depicted and finally, the forecast extremes (95% and 5%) are represented by the whiskers. Thus, the rectangle symbol with two additional vertical lines represent the dispersion of a variable for the all members of the ensemble at a given time.
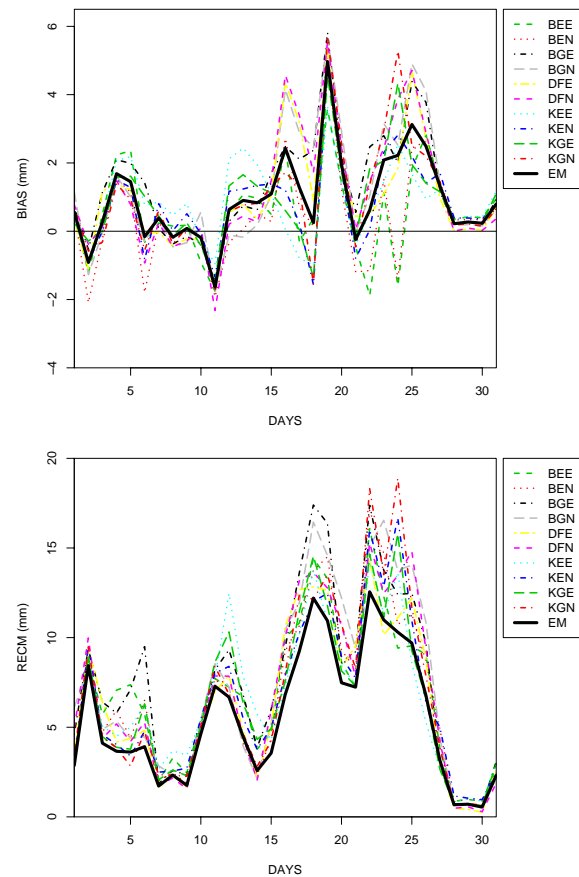

Fig. 3. Temporal evolution of bias **(a)** and RMSE **(b)** for each ensemble member and the ensemble mean of the total daily precipitation, spatially averaged over the region. Units are in mm/m^2. The x-axis indicates the days of the period time.

4 Results

In this section the results associated with both spatial and probabilistic verification methods are shown, assessing accuracy and skill of precipitation forecasts for the fine domain of the simulations obtained from the 10-members ensemble for October 2006. Figure 3 displays the mean bias and RMSE of the 24-h precipitation over a regular grid (25 km×25 km) in order to provide rainfall data at a resolution compatible with the observational data.

The time evolution of both bias and RMSE is similar for all ensemble members (Figs. 3a and b, respectively), showing evidence of different behaviour over the study period. In terms of the total daily precipitation, spatially averaged over the region, positive bias can be noted throughout the event period except for particular days with negative bias results (Fig. 3a). The ensemble mean provides a good forecast when it is compared with some ensemble members and the time period is enough large (Grimit and Mass, 2002). In fact, in this paper, the ensemble mean offers the best forecasts when it is compared with any ensemble member used to build the EPS.
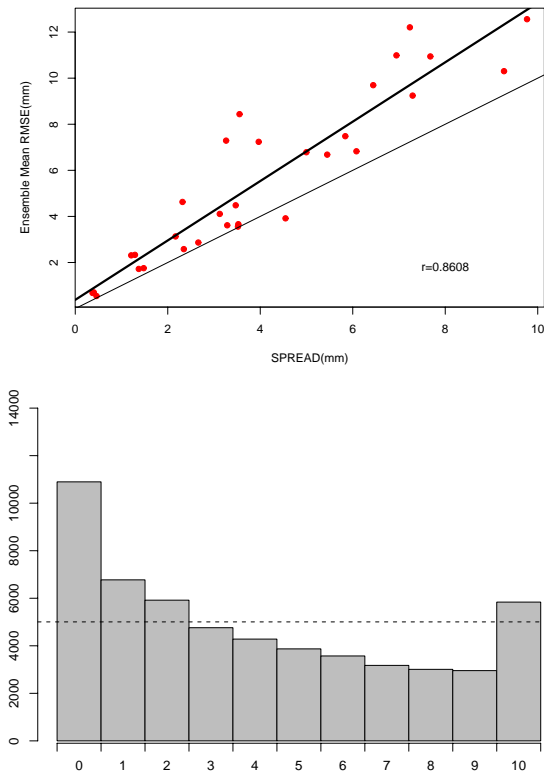
Fig. 4. (a) Scatter diagram of the ensemble RMSE (mm) and the ensemble spread (mm) for daily precipitation, showing in thin line the ideal diagonal; **(b)** Talagrand diagram showing the frequencies at which the verifying analysis falls in each category, defined by the 10 ordered ensemble members at each grid points. Dashed line represents the equiprobability line.

The departures in the different stages of the period could indicate some rainfall over/underprediction. Large ranges in the bias can suggest the existence of large variability in the accumulated precipitation. Concerning the RMSE (Fig. 3b), similar explanation can be stated, also highlighting great departures in the final stages of the period and low RMSE values in the initial part of the study period.

As it is abovementioned in the previous section, in this paper several skill scores for ensemble forecasts have been used. The probabilities were calculated by equally weighting each member of the system. Thus, each forecast model represents 1/10 of the total probability. In Fig. 4a the relationship between the RMSE and the spread of the ensemble mean is shown. Such relationship is quite close to the ideal diagonal with a correlation value of 0.86 (corresponding with the line regression displayed by the thick line). The ideal behaviour corresponds with a coincidence with the y=x line (thin line), indicating that forecast spread is appropriate to represent the uncertainty by means of member spreading.

The Talagrand distribution is the histogram of frequencies of the rank of the observed data within the forecast ensemble. Although a rank histogram measures whether the observed probability distribution is well represented by the ensemble,
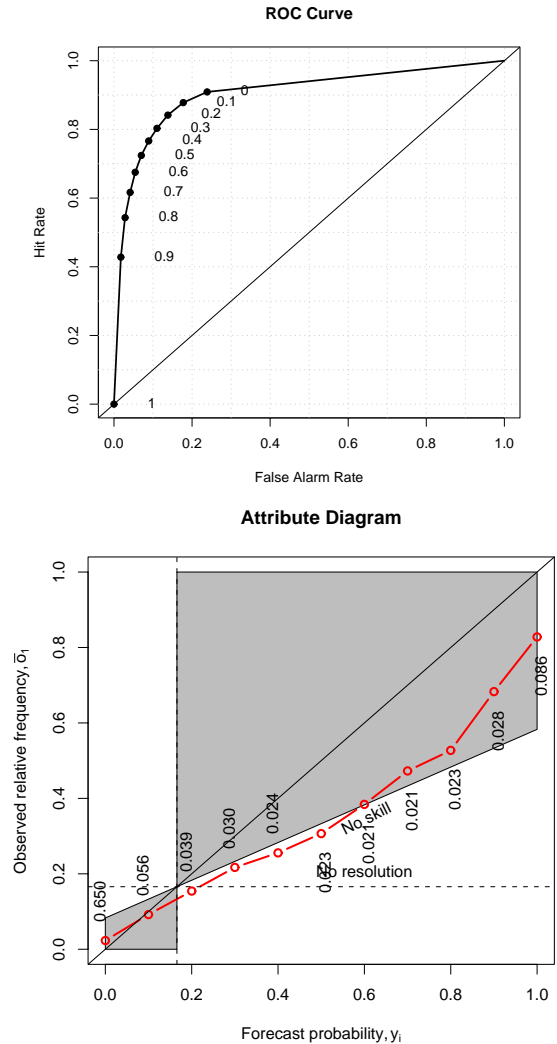

Fig. 5. (a) ROC Curve precipitation; **(b)** Attribute diagram, over 5 mm/24 h threshold.

a flat Talagrand does not necessarily indicate a skilled forecast. Nevertheless a good EPS should have a uniform distribution, representing correctly the forecast uncertainty and the observed data set uniformly be distributed among the ensemble members. A distribution can present a ∩-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 itself slightly ∪–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 ∪-shaped would indicate ensemble spread too small with many observations falling outside the extremes of the ensemble. For some parameters such ∪-shape degenerates into an asymmetric shape, indicating the presence of bias in the system for such parameter. Thus, Fig. 4b shows an asymmetric distribution in which the first bin is the most frequent. Therefore, the ensemble contains bias. Without bias correction,

an overpopulation of the extreme ranks of the histogram can be noted, suggesting with this asymmetric shape a bias behaviour. This situation can be made accurate by applying statistical postprocesses with a bias correction and further generation of a flatter histogram.

The ROC is considered as a measure of potential utility, measuring ability of the forecast in discriminating between two alternative outcomes: events and non-events (resolution). 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 resolution. Although the Talagrand diagram (Fig. 4b) showed bias as it has abovementioned, the good ROC curve (Fig. 5a) 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 resolution of the EPS. On the other hand, the attribute 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, indicating good reliability a line close to the diagonal. However, deviation from the diagonal points out conditional bias. Thus, deviation below the diagonal represents too high forecast probabilities; on the contrary, points above diagonal indicate too low forecast probabilities. Points between the diagonal and "no skill" line have a positive contribution to the BSS. The flatter the curve in the diagram, the less resolution it has. In this work, a good agreement between forecast probability and the mean observed frequency can be noted in the Fig. 5b in which the circle-line is close to the no-skill line. Although some forecast probabilities, ranging from 20 to 50%, contribute negatively to the BSS due to them being located below the no-skill line, the remainder probabilities are in between the no-skill and the diagonal lines. On the other hand, the sharpness values are shown in the Fig. 5b as numbers over/below the diagram points, displaying the highest value in the 0% forecast probability. Since this probability has a good reliability and a positive contribution to the BSS, a positive value of 0.4054 of this skill can be obtained (see Table 2). To avoid undersampling effects, works in extending this study to other larger samples are in progress.

Finally, the probabilistic meteogram, also called epsgram, shows the ensemble distribution of the forecast, displaying the information on box-and-whiskers plots. Usually, the ensemble information exposed by the epsgram is relative to an individual grid–point location, indicating the time evolution of a parameter for all ensemble members and also pointing to the spread by the range of forecast values. In this work, epsgrams show the time evolution of the spatially mean daily precipitation for the ensemble system (Fig. 6a) as well as the evolution of the total daily precipitation, spatially aver-

Table 2. Brier Score, BS component values and Brier Skill Score of precipitation over 5mm/24h threshold.

BS	Reliability	Resolution	Uncertainty	BSS
0.08048	0.009457	0.0643	0.1354	0.4054

aged over the region, for each ensemble member (Fig. 6b), rather than the time evolution for all ensemble members at an individual grid–point location. Figure 6a shows departures between the box plots, representing the dispersion of the precipitation forecasts for the all members of the EPS at the date of the time period. Great departures in the final part of the time period and minimum ones in the initial part can be noted, thus exhibiting a similar behaviour that the RMSE evolution as it can be observed in Fig. 3b. The limit of the upper (lower) vertical line represents the value of the highest (lowest) member. Except for the 10th, 13th, 15th an 24th days, presenting outliers, there are not excessive variation in the forecast values, suggesting confidence in the ensemble system. Moreover, the evolution of the spatially averaged observed precipitation (continuous line) is quite similar to the time evolution of the spatially mean daily precipitation for the ensemble system. On the other hand, the epsgram evolution of each ensemble member (Fig. 6b) shows in general a good agreement among them. Their median, first and third quartile values are quite similar among the members, indicating that most of the member forecasts present similar behaviour throughout all the time period. Most of the forecast higher extremes, represented by whiskers, present values between 10–15 mm, except for the KEE and KGE members, showing outlier values. Moreover, values of the mean and the median of the spatially averaged observed precipitation (dashed lines) maintain in the box plot values in all ensemble members. In fact, the observed median is embraced between the lower and the upper quartiles.

5 Conclusions

The purpose of this study is the verification of a short-range ensemble prediction system (SREPS) built with five different model physical process parameterization schemes and two different initial conditions from global models, allowing to construct several versions of the non-hydrostatic mesoscale MM5 model for a 1-month period of October 2006. 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 mean model performance and forecast variability, together with the evaluation of the ensemble accuracy have shown a good EPS performance.

Concerning the spatial verification scores, the temporal evolution of the mean bias and rmse of the 24-h precipitation
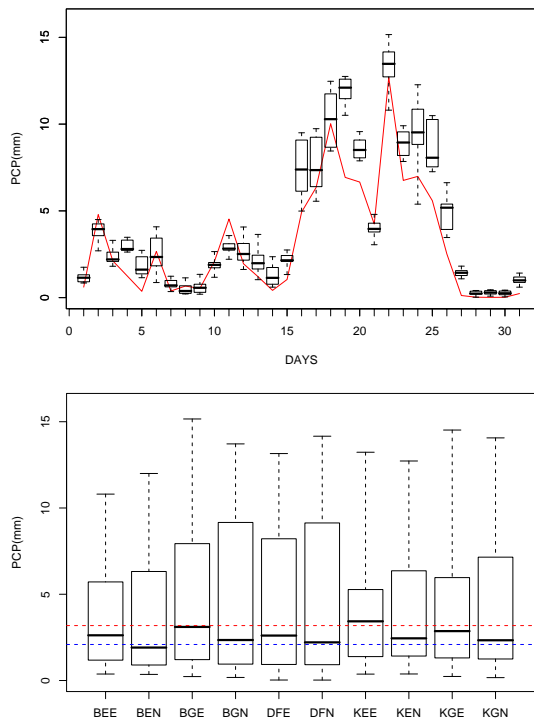
Fig. 6. Epsgram of the total daily precipitation, spatially averaged over the region, for **(a)** the ensemble system for each day of the period time with juxtaposition in solid line of the spatially averaged observed precipitation; **(b)** each ensemble member cited in the x-axis; dashed red and blue lines represent the mean and the median of the spatially averaged observed precipitation. Units are in mm.

over a regular grid have shown similar evolution of the ensemble simulations, excepting particular ensemble members that have shown large departures. Different behaviour over the study period has been also noticed with large departures in the first and last stages of the period, while a quasi-zero evolution during the central part of the time period is noticed. Therefore, large variability in the accumulated precipitation is exhibited when these large ranges are considered.

High spread-skill correlation values (> 0.86) for daily precipitation are noted, but with small spread values, indicating that the EPS is affected by underdispersion. Moreover, such underdispersive effect is noted in the Talagrand diagram distribution, indicating that the generated envelope forecasts are not including the verifying observation. The Talagrand diagram has shown an asymmetric distribution, pointing out the presence of bias in the ensemble system.

Concerning verification probabilistic methods, the ROC curve suggests a greater forecast quality of the EPS, with the curve showing an evolution tending to the upper left corner, i.e., pointing where there are no false alarms and only hits. Such situation is indicative of the usefulness of the ensemble system.

The attribute diagram shows some forecast probabilities contribute negatively to the BSS due to the fact that they are

located below the no-skill line; the remaining probabilities are in between the no-skill and the diagonal lines. The result of all these contributions lead to a positive value of BSS, pointing out the skill of the system. To avoid undersampling effects, works in extending this study to other larger samples are in progress.

Epsgrams of each ensemble member have shown a good agreement among them, with quartile values quite similar among all members. Moreover, epsgrams inform about the spread of the ensemble system. In this paper, departures between them throughout the time period are shown. This dispersion is mainly noted in the early and final stages of the period, being coincident with the RMSE evolution. Thus, those days presenting large errors will be coincident with those days exhibiting big dispersion. The evolution of the spatially averaged observed precipitation presents quite similarity to the time evolution of the spatially mean daily precipitation for the ensemble system.

Edited by: S. C. Michaelides
Reviewed by: two anonymous referees

References

Anthes, R. A. and Warner, T. T.: Development of hydrodynamic models suitable for air pollution and other mesometeorological studies, Mon. Weather Rev., 106, 1054–1078, 1978.

Ballard, S. P., Golding, B. W., and Smith, R. N.: Mesoscale model experimental Forecasts of the Haar of Northeast Scotland, Mon. Weather Rev., 119, 2107–2123, 1991.

Betts, A. K.: A new convective adjustment scheme, Part I: Observational and theoretical basis, Q. J. Roy. Meteor. Soc., 112, 677–692, 1986.

Betts, A. K. and Miller, M. J.: A new convective adjustment scheme, Part II: Single column tests using GATE wave, BOMEX, ATEX and Artic air-mass data sets, Q. J. Roy. Meteor. Soc., 112, 693–709, 1986.

Brier, G. W.: Verification of forecasts expressed in terms of probabilities, Mon. Weather Rev., 78, 1–3, 1950.

Chen, F. and Dudhia, J.: Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system, Part I: Model implementation and sensitivity, Mon. Weather Rev., 129, 569–585, 2001.

Dudhia, J.: A multi-layer soil temperature model for MM5. Preprints, The Sixth PSU/NCAR Mesoscale Model Users' Workshop, 22–24 July 1996, Boulder, Colorado, 49–50, 1996.

Epstein, E. S.: Schocastic dynamic prediction, Tellus, 21, 739–759, 1969.

Dyras, I., Dobesch, H., Grueter, E., Perdigao, A., and Tveito, O.: The use of Geographic Information Systems in climatology and meteorology: COST 719. Meteorol. Appl., 12, 1–5, 2005.

Grell, G. A., Dudhia, and Stauffer, D. R.: Diagnosing coupled jet streak circulations for a northern plains snow band from the operational nested-grid model, Weather Forecast., 7, 26–48, 1994.

Grimit, E. P. and Mass, C. F.: Initial reults of a mesoscale short-range ensemble forecasting asystem over the Pacific Northwest, Weather Forecast., 17, 192–205, 2002.

Hoffman, R. N. and Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting. Tellus A, 35, 100–118, 1983.

Janjic, J. and Zavisa, I.: The step-mountain eta coordinate model: Further development of the convection, viscous sublayer, and turbulent closure schemes, Mon. Weather Rev., 122, 927–945, 1994.

Kain, J. S. and Fritsch, J. M.: A one-dimensional entraining/detraining plume model and its application in convective parameterization. J. Atmos. Sci., 47, 2784–2802, 1990.

Leith, C. E.: Theoretical skill of Monte Carlo forecast, Mon. Wea. Rev., 102, 409–418, 1974.

Lin, Y. L., Farley, R. D. and Orville, H.D.: Bulk parameterization of the snow field in a cloud model, J. Appl. Meteorol. Clim., 22, 1065–1092, 1983.

Luna, M. Y. and Almarza, C.: Interpolation of 1961–2002 daily climatic data in Spain, Spatial Interpolation in Climatology and Meteorology (ISBN 92-898-0033-X), edited by: Szalai, S., Bihari, Z., Szentimrey, T., and Lakatos, M., COST Office, 189–196, 2007.

Mason, I.: A model for assessment of weather forecasts. Aust. Meteorol. Mag., 30, 291–303, 1982.

Morata, A.: Influencia de la circulación atmosférica sobre la precipitación en la cuenca Mediterránea Occidental, Tesis Doctoral, Universidad Complutense de Madrid, Madrid, España, 193 pp., 2004.

Morata, A., Luna, M. Y., Martín, M. L., Sotillo, M. G., and Valero, F.: Iberian autumnal precipitation characterization through observed, simulated and reanalysed data, Adv. Geosci., 16, 49–54, 2008,
http://www.adv-geosci.net/16/49/2008/.

Mullen, S. L. and Baumhefner, D. P.: Sensitivity to numerical simulations of explosive oceanic cyclogenesis to changes in physical parameterizations, Mon. Weather Rev., 116, 2289–2329, 1988.

Murphy, A. H.: What is a good forecast? An essay on the nature of goodness in weather forecasting, Weather Forecast., 8, 281–293, 1993.

Reisner, J., Rasmussen, R. M., and Bruintjes, R. T.: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. Q. J. Roy. Meteor. Soc., 124, 1071–1107, 1998.

Shafran, P. C., Seaman, N. L., and Gayno, G. A.: Evaluation of numerical predictions of boundary layer structure during the Lake-Michigan Ozone Study, J. Appl. Meteorol., 39, 412–426, 2000.

Shen, S. S. P., Dzikowski, P., Li, G., and Griffith, D.: Interpolation of 1961–97 Daily Temperature and Precipitation Data onto Alberta Polygons of Ecodistrict and Soil Landscapes of Canada, J. Appl. Meteorol., 40, 2162–2177, 2001.

Shuman, F. G.: History of Numerical Weather Prediction at the National Meteorological Center, Weather Forecast., 4, 286–296, 1989.

Sotillo, M. G., Martin, M. L., Valero, F., and Luna M. Y.: Validation of an homogeneous 41-year (1961–2001) winter precipitation hindcasted dataset over the Iberian Peninsula: assessment of the regional improvement of global reanalysis, Clim. Dynam., 27, 627–645, 2006.

Sumner, G., Homar, V., and Ramis, C.: Precipitation seasonality in eastern and southern coastal Spain, Int. J. Climatol., 21, 219–247, 2001.

Swets, J. A.: The relative operating characteristic in psychology, Science, 182, 990–1000, 1973.

Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of probabilistic prediction systems, Proc. Workshop on Predictability, 1–25, 1997.

Tao, W.-K., and Simpson, J.: The Goddard cumulus ensemble model, Part I: Model description. Terr., Atmos. Oceanic Sci., 4, 35–72, 1993.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP: The breeding method, Mon. Weather Rev., 125, 3297–3318, 1997.

Valero, F., Luna, M. Y., Martín, M. L., Morata, A., and González-Rouco, F.: Coupled modes of large-scale climatic variables and regional precipitation in the Western Mediterranean in autumn, Clim. Dynam., 22, 307–323, 2004.

Wang, W. and Seaman, N. L.: A comparison study of convective parameterization schemes in a mesoscale model, Mon. Weather Rev., 125, 252–278, 1997.

Wilks, D. S.: Statistical methods in the atmospheric sciences: an introduction, Academic Press Inc., San Diego, California, USA, 467 pp., 1995.