



Design of two ensemble prediction systems based on multiphysics and potential vorticity perturbations: Test for western Mediterranean precipitation and cyclones

Maria-del-Mar Vich *^a, Romualdo Romero^a and Harold Brooks^b

^a*University of the Balearic Islands, Dpt. Physics, Meteorology Group Palma de Mallorca, Spain*

^b*National Oceanic and Atmospheric Administration, National Severe Storms Laboratory Norman, Oklahoma, United States*

Abstract: The western Mediterranean is a very cyclogenetic area and many of the cyclones developed over the region are associated with high impact weather phenomena that affect the society of the coastal countries. With the aim of improving the short to mid-range numerical forecasts of cyclones and heavy rain events, ensemble prediction systems based on multiphysics and perturbed initial and boundary conditions were designed. All simulations are performed with the MM5 mesoscale model nested in the ECMWF forecast large-scale fields, which provides results at 22.5 km resolution for a two-day period over the western Mediterranean countries.

In the multiphysics ensemble, different combinations of model physical parameterizations are chosen. In the perturbed initial and boundary conditions ensemble, a potential vorticity (PV) inversion technique is used to perturb the initial state and boundary forcing of the mesoscale model. A PV error climatology (PVEC) derived from the large-scale fields allows to perturb the ECMWF PV fields using the appropriated error range.

The results show a good improvement of the prediction capability, evaluated through precipitation verification indices, for both ensembles over the traditional deterministic prediction, and a better performance of the perturbed initial and boundary conditions ensemble than the multiphysics. The verification procedure highlights the difficulties of evaluating the rainfall field forecast, since the discontinuities of this field and the insufficient samples of extreme events affects the verification scores negatively. Despite this, the results reinforce the idea that an EPS represents an improvement over a deterministic prediction. Copyright © 2009 Royal Meteorological Society

KEY WORDS Ensemble prediction system; Forecast verification; multiphysics; PV perturbations

Received 15 May 2009

1 Introduction

The western Mediterranean often suffers heavy rain events that have a high socio-economical impact. Several studies have established a connection between heavy rainfall and cyclones over the western Mediterranean

*Correspondence to: Universitat de les Illes Balears, Dpt. Física, Meteorology Group, Cra. Valldemossa, km. 7.5, 07122, Palma de Mallorca, Spain. E-mail: mar.vich@uib.es



region (e.g. [Jansà *et al.* 2001](#)), a very cyclogenetic area as a consequence of its latitude and complex topography ([Reiter, 1975](#); [Meteorological Office, 1962](#)). Different types of cyclones associated with heavy rain events have been described, such as: 1) shallow weak disturbances with a warm core over land masses of thermal origin ([Romero *et al.*, 2001](#)); 2) shallow weak lows with a warm core over the sea at the lee of important mountain ranges, linked to the orographic effect on the atmospheric flow ([Romero *et al.*, 2000](#)); and 3) baroclinic systems with great vertical amplitude developed along frontal zones under the intrusion in the Mediterranean region of an upper-level through ([Homar *et al.*, 2002](#)). Thus, the prediction of these events is very important to prevent or reduce the damages they cause.

The purpose of this study is to apply a probabilistic forecast approach to improve the prediction of these high impact weather events. A probabilistic forecast accounts for the uncertainty in the initial state ([Lorenz, 1963](#)) and the errors introduced by the models imperfections ([Frank, 1983](#)). It is derived using several numerical model runs. This technique is referred to as ensemble prediction system (EPS). The EPS pretends to produce a forecast probability density function (PDF) that captures the true PDF.

The construction of an EPS can be done using a wide range of techniques. One of them is perturbing the initial state. This technique was implemented operationally in the 1990s at the National Centers for Environmental Prediction (NCEP; [Toth and Kalnay 1997](#)) and at the European Center for Medium-Range Weather Forecasts (ECMWF; [Palmer *et al.* 1992](#); [Molteni *et al.* 1996](#)). Lately, other meteorological centers have also implemented it operationally, like the Meteorological Service of

Canada (MSC; [Pellerin *et al.* 2003](#)). Another EPS building technique takes advantage of several studies that show how model parameterization schemes can contribute to forecast errors ([Houtekamer *et al.*, 1996](#); [Stensrud *et al.*, 1999, 2000](#); [Wandishin *et al.*, 2001](#)). Thus, ensembles containing different models or different parameterization schemes are more skillful than ensembles that do not contain some aspect of model uncertainty ([Stensrud, 2007](#)). Hence, the use of multiphysics and multimodel ensembles is widely used to account for the model imperfections. These previous EPS constructing techniques deal with a Global Area Model (GAM) while our study requires a Local Area Model (LAM). The transition from GAM to LAM presents difficulties and is yet to be defined unequivocally. How one deals with this transition is part of the strategy one needs to choose when defining a mesoscale EPS.

In this paper a multiphysics ensemble that uses different physical parameterization for each ensemble member and an ensemble with perturbed initial and boundary conditions are studied. The latter ensemble introduces the perturbations on the PV field through a derived potential vorticity error climatology. We have chosen to deal with GAM-LAM transition by nesting a mesoscale model (MM5) in the ECMWF forecast large-scale fields, and when it corresponds, perturbing both the initial and the boundary conditions. There are other recent studies that compare the contributions of initial/lateral boundary conditions and mixed physics to the spread and skill of 120-h precipitations ensemble forecast ([Clark *et al.*, 2008](#)), that deal with the GAM-LAM transitions in a similar way.

To evaluate the performance of both ensembles a collection of 19 MEDEX[†] cyclones is used as a trial

[†]MEDEX is a Mediterranean Experiment on cyclones that produce high impact weather in the MEDiterranean, a project endorsed by WMO (<http://medex.aemet.uib.es>).

set. These MEDEX cyclones are associated with floods and strong winds over the western Mediterranean and represent the kind of events that this study aims. The verification process is applied to the precipitation field, a hard field to predict and verify due to its complex nature but with a direct impact on the society. The precipitation is a highly discontinuous variable, meaning that there is a large fraction of zones and times where the precipitation is zero. Also, it is difficult to match the forecast and the verifying data because the forecasts provide regularly spaced grid fields while the verifying observations usually are in irregular spaced networks of stations.

There is a wide range of verification methods that can be used to prove the quality of an ensemble prediction system. The quality of the forecasts involves measures of the relationship between a forecast, or set of forecasts, and the corresponding observations. As [Murphy and Winkler \(1987\)](#) and later [Murphy \(1993\)](#) point out, different quality attributes of the forecast such as reliability, resolution, uncertainty and sharpness need to be evaluated in the verification process to test the performance of such forecast. These quality attributes can be evaluated through different techniques such as the Brier score ([Brier, 1950](#)), reliability diagrams ([Wilks, 1995](#)), the relative operating characteristic, or ROC ([Mason, 1982](#)), and the rank histogram, also known as a Talagrand diagram.

This paper summarizes the design and test for the MEDEX cyclones of the two above mentioned ensemble prediction systems. A description of the different datasets used in the study is done in section 2. Section 3 contains the development of the methodology designed to build both ensembles. The ensemble verification procedure is described in section 4. Finally, some concluding remarks and future outlooks are presented in section 5.

2 Datasets

The ensemble dataset needs to account for our main goal, designing an EPS for the Mediterranean area able to predict the occurrence of extreme precipitation events. Ideally, the EPS needs to be a real time fully operational system. The UIB Meteorology Group has been running the MM5 model on a daily basis for some years ‡. Owing to computational limitations, these 48 h numerical forecasts, centered over the western Mediterranean region, are initialized with global coarse resolution 24 h forecast fields valid at 00 UTC and forced at the lateral boundaries with the subsequent data (i.e. with 30,36, ..., 72 h global forecasts). Thus, the mesoscale prediction system can only be considered as quasi-operational in the sense that forecasts, rather than analyses, are used to nest the MM5 model. With such a testbed in mind, the same type of data will be used to initialize and force the numerical weather model in the present study. This meteorological data is provided by the European Center for Medium-Range Weather Forecasts, ECMWF.

The Mediterranean Experiment on Cyclones that produce High Impact Weather in the Mediterranean, MEDEX, is a project designed to contribute to a better understanding and short-range forecasting of high impact weather events in the Mediterranean, mainly heavy rain and strong winds, and is focused on Mediterranean cyclones that produce high impact weather. Thus MEDEX provides a suitable database for our study and the ensembles trial set will consist of a collection of 19 MEDEX cyclonic episodes comprising 56 different days between September 1996 and October 2002 (Table I).

Finally, a precipitation dataset is needed for the verification process of the ensembles forecasted precipitation. This dataset comes from the AEMET (Agencia Estatal

‡See <http://mm5forecasts.uib.es>

Table I. List of the MEDEX cyclonic episodes used for this study. A more detailed description can be found at <http://medex.aemet.uib.es>.

| | Date | Country affected |
|----|------------------------|----------------------------------|
| 01 | 11-12 Sep. 1996 | Spain |
| 02 | 06-09 Oct. 1996 | Italy, Spain |
| 03 | 14 Oct. 1996 | Spain, Italy |
| 04 | 04-06 Nov. 1997 | Portugal, Spain, France |
| 05 | 11-14 Nov. 1999 | Italy, Spain, France |
| 06 | 10 Jun. 2000 | Spain |
| 07 | 21-26 Oct. 2000 | Spain |
| 08 | 02-05 Nov. 2001 | Spain |
| 09 | 09-13 Nov. 2001 | Algeria, Spain, Croatia, Morocco |
| 10 | 14-16 Nov. 2001 | Spain |
| 11 | 14-15 Dec. 2001 | Spain |
| 12 | 11 Apr. 2002 | Spain |
| 13 | 06-08 May. 2002 | Spain |
| 14 | 12-15 Jul. 2002 | Spain, Croatia |
| 15 | 31 Jul. - 01 Aug. 2002 | Spain |
| 16 | 08-10 Sep. 2002 | France |
| 17 | 12-13 Sep. 2002 | Spain |
| 18 | 23-24 Sep. 2002 | Spain |
| 19 | 08-10 Oct. 2002 | Spain |

de Meteorología - Spanish Metoffice) climatological rain-gauge network, which provides 24 h accumulated precipitation from 06 UTC to 06 UTC the next day. Figure 1 shows the spatial distribution of this network over the Mediterranean influenced regions of Spain, with a total amount of 2300-2400 stations depending on the event.

3 Ensembles Construction

The ensemble consists of 13 members, 12 regular members plus a chosen control member. These members are run with the MM5 model. The MM5 non-hydrostatic mesoscale model is a high resolution short-range weather forecast model developed by the National Center for Atmospheric Research (NCAR) and the Pennsylvania State University (PSU) (Dudhia, 1993; Grell *et al.*, 1995). The simulation domain is defined as a 22.5 km resolution horizontal grid mesh with 120x120 modes, centered at

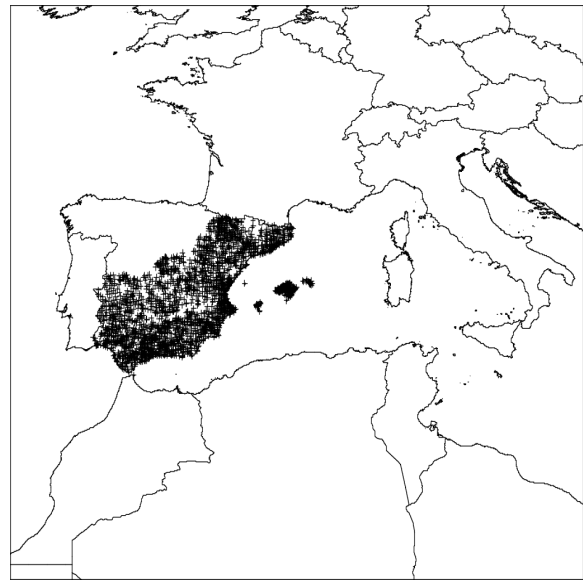


Figure 1. Geographical domain used for the MM5 numerical simulations. The spatial distribution of the AEMET rain-gauge network used for the ensembles verifications is plotted using crosses.

39.8° latitude and 2.4° longitude. The vertical grid mesh is defined by 30 sigma levels. This domain (Fig. 1) contains all the areas affected by the selected MEDEX cyclones and corresponds to the Domain 1 used in the deterministic quasi-operational model runs done by UIB Meteorology Group. To simplify the verification process, the forecasting period starts at 00 UTC but extends for 54 h instead of the regular 48 h, since the observed daily rainfall accumulations correspond to 06 UTC. The MEDEX trial set will be covered by all possible simulations comprising a MEDEX day within the 24-48 h forecast range. Thus, 56 different simulations have to be done per ensemble member.

3.1 Perturbed initial and boundary conditions ensemble

In this study the ensemble perturbations are defined for the wind and temperature fields through the PV field by applying the potential vorticity inversion technique developed by Davis and Emanuel (1991). This PV inversion technique uses the mass-wind balance condition derived by Charney (1955) to link the wind field to the temperature field and it has been used successfully by different

authors (e.g. [Romero 2001](#) and [Romero 2008](#)) to understand quantitatively the relevance of the PV-anomalies on the atmospheric behavior.

The perturbations are introduced in the initial and boundary PV fields randomly, along the zones with the most intense values and gradients of PV under the assumption that these are the most sensitive zones ([Garcies and Homar, 2009](#); [Romero et al., 2005, 2006](#)) of the subsequent atmospheric evolution, like the cyclogenesis process to occur over the western Mediterranean. If this hypothesis is true, the spread in the ensemble realizations would be benefited. The inverted mass and wind fields from the perturbed PV field will define, once compared with the unperturbed fields, the perturbations to be introduced in the MM5 model initial and boundary states. This is a neat approach of building physically-consistent perturbations of the primitive mass and wind fields. However, to be of value the EPS method requires the introduction of PV perturbations of realistic magnitude. To this aim, a PV error climatology or quantitative assessment of the PV uncertainty range has been derived.

The Potential Vorticity error climatology is obtained by comparing the ECMWF 24 h forecast and ECMWF analysis PV fields (the alternative approach of comparing two different analysis sources, like NCEP and ECMWF, seems to offer essentially the same results; [Horvath 2008](#)). It is assumed that the main error sources are the displacement between PV structures and differences in intensity, thus we will assess separately the displacement and intensity errors. This climatology is derived from the 19 MEDEX cyclonic episodes described above, sampled over a 200x200 grid points large domain at 6 h intervals at the standard pressure levels (100, 200, 300, 400, 500, 700, 850, 925 and 1000 hPa). The PV error climatology will tabulate, for each pressure level, the coefficients of

the functions fitted to the displacement and intensity error percentiles extracted from this large amount of data.

The displacement error (DE) represents the minimum of the forecast field displacement necessary to obtain maximum local correlation with the analysis field. DE is found after displacing a forecast PV field local matrix (21x21 grid points, or 450x450 km) in all directions, up to 10 grid points (i.e. 225 km). The minimum displacement showing local maximum correlation between the displaced field and the analysis field is assigned to the matrix central grid point as the corresponding displacement error. This process is repeated for each domain grid point along with its associated 21x21 grid points matrix. The choice of the matrix dimension and the maximum tested displacements are not entirely arbitrary, actually the 450x450 km matrix dimension ensures that, according to the PV field typical scales, subsynoptic structures will be effectively sampled. The possible displacements up to 225 km ensure that plausible lags between forecast and analysis PV fields can be identified. It should also be noted that, by definition, the displacement error is a discrete magnitude.

The intensity error (IE) represents the difference between the displaced forecast field and analysis PV. It is calculated for each domain grid point as the difference in the 21x21 grid points PV average between the aforementioned optimally displaced forecast field and the analysis field, and will be expressed relative to the analysis PV average (i.e. as an error percentage).

The representation of the displacement error density function, expressed as function of number of grid lengths, shows a very clear symmetry along South-North and West-East directions (see [Fig. 2](#) for 300 hPa). This fact allows to express the error as absolute value in both directions. Likewise the intensity error presents a very high symmetry between positive and negative values (not

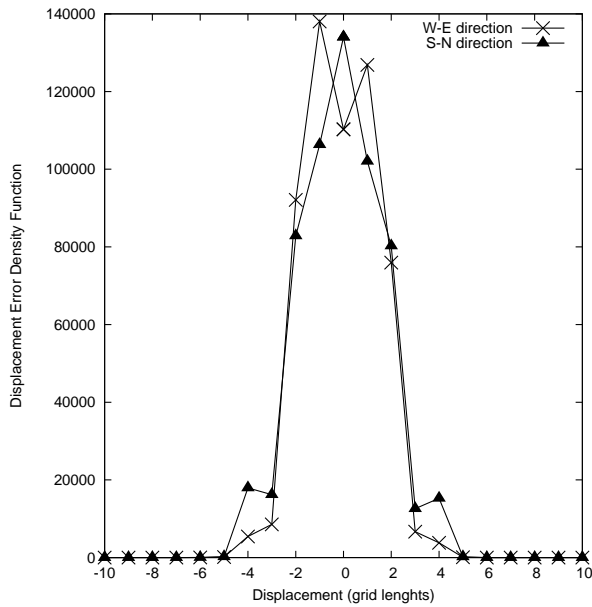


Figure 2. Displacement error density function at 300 hPa. It represents the number of occurrences (grid points) as function of the displacement error in grid lengths. Results for the South-North (solid line) and West-East (dotted line) directions are represented separately.

shown), so the absolute value is used too. With regard to the dependency on the PV value, DE appears to be insensitive to this value, while IE exhibits an appreciable dependency. It could be argued that this later case owes to the fact of expressing IE in relative terms, but a similar behavior is found when expressed as an absolute error. For the intensity error, the percentiles calculation requires, as a previous step, organizing the analysis PV values in 0.1 PVU -an arbitrary choice- classes for each pressure level. Then the IE percentiles as function of PV value and isobaric level are easily calculated from our population of error data. However the DE needs a different approach due to its discrete nature. The first step is to obtain the percentage of events with a displacement of X grid lengths (X ranges from 0 to 10) and then associate this percentage to $X.5$ grid lengths. Finally to obtain the standard percentile levels a linearly interpolation of the obtained percentages is done.

The representation of the calculated error percentiles for the whole range of PV values displays the lack of

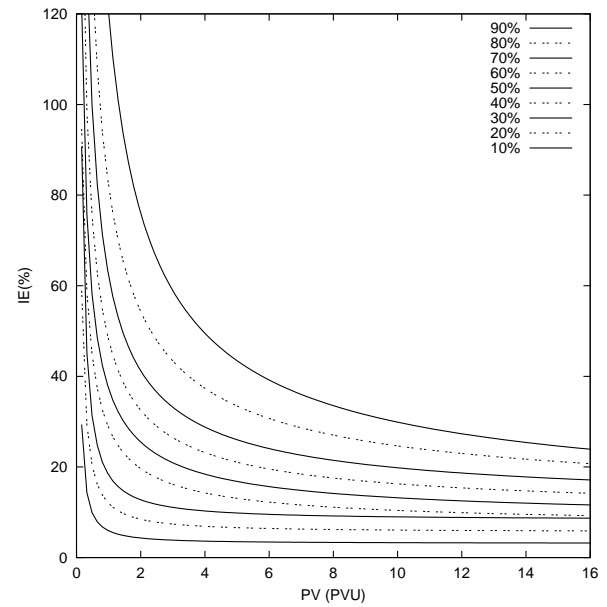


Figure 4. Intensity Error percentile levels at 300 hPa.

dependence on the PV value for the displacement error (Fig. 3) and the clear dependence on the PV value for the intensity error (Fig. 4). The analytical functions fitted to the percentile levels are linear-like for each geographical direction for the displacement error, and power-law for the intensity error. This PV error climatology (DE and IE percentiles) is used to introduce realistic perturbation along the zones with the most intense PV values and gradients that tend to characterize the most influential PV structures of mid-latitude weather systems. These sensitivity zones are delineated objectively by the following method. First, a pseudo-sensitivity field is defined as the difference between the PV field and a highly smoothed version of it. Then, for each pressure level, a threshold is defined as the domain averaged pseudo-sensitivity field in absolute value. The zones with PV higher than the threshold and lower than the negative of the threshold are then detected along the three dimensional domain. The total detection volume covers approximately half of the domain and each three-dimensional detected piece generally corresponds to an identified PV structure.

Finally, three different percentile levels are assigned

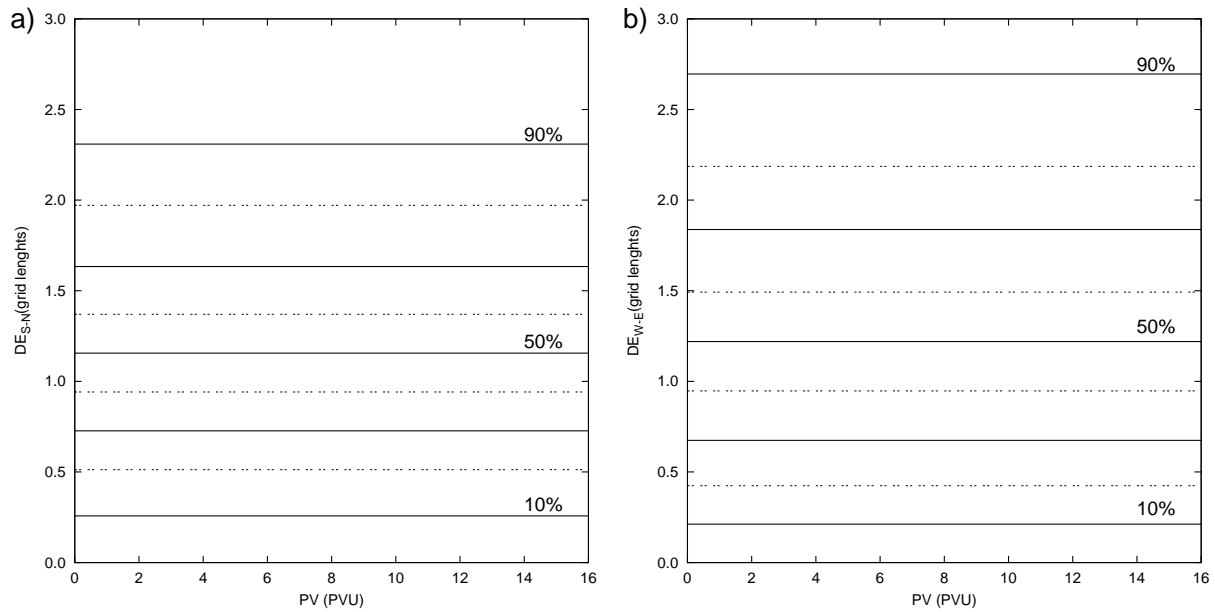


Figure 3. Displacement Error percentile levels along a) South-North and b) West-East directions at 300 hPa.

randomly to each detected sensitivity volume, one for intensity error and two for displacement error (one for each direction). The intensity and displacement error orientations (positive or negative) are generated randomly as well, globally for DE (i.e. common direction the whole domain) in order to avoid discontinuities in the perturbed PV field. These percentile levels are converted to PV perturbations through the previous PV error climatology. It is worth to notice that a superior limit is imposed to the IE ($IE_{\max} = 200\%$) since this relative error tends to infinity when PV tends to zero due to the power-law function properties (see Fig. 4).

To generate the ensemble members, the PV inversion technique is applied to the perturbed PV field. The difference between the original and perturbed balanced fields is added to the control, unperturbed mass and wind fields, to produce the perturbed initial and boundary conditions.

3.2 Multiphysics ensemble

This multiphysics ensemble is constructed using different MM5 physical parameterization sets. These sets are the

result of combining three explicit moisture schemes (Godard microphysics, Reisner graupel and Schultz microphysics), two cumulus parameterizations (Grell and Kain-Fritsch) and two PBL schemes (Eta and MRF), plus the set used in the operational model run by our group (the explicit moisture scheme Reisner graupel, the cumulus parameterization Kain-Fritsch 2 and the PBL scheme MRF). The parameterization set of the operational model is defined as the control member of the ensemble.

4 Verification procedure and results

A first evaluation of the EPSs is done by assessing the performance of each EPS individual member in order to assure that it is good enough to be included within the ensemble. Afterwards, each EPS is verified using probabilistic scores and indices, assuming each deterministic forecast (ensemble member) is an independent realization of the same underlying process. In this study the verification is performed for the 24 h accumulated precipitation period corresponding to the second day of the simulations.

The assessment of a weather forecast is not an easy task due to the lack of a proper understanding of what

is a good forecast. To clarify the goodness of a forecast [Murphy \(1993\)](#) defines three distinct types of goodness: consistency (forecasters' judgments versus forecasts), quality (forecasts versus observations) and value (benefits). Traditionally, forecast verification practices consist of an evaluation of the correspondence between forecast and observations. Any procedure for verification of the quality of the forecasts involves measures of the relationship between a forecast or set of forecasts and the corresponding observations, hence a verification process involves a comparison between matched pairs of forecasts and the observations to which they pertain. In order to analyze this relationship, for dichotomous forecasts, contingency tables are used. A dichotomous forecast refers to the yes/no nature of the forecast at each point, for example a rainfall threshold defining the transition between a rain event and a nonrain event. The contingency table, shown in [Table II](#), counts for each of the four possible outcomes for the observed and forecast event: Hit (a), False alarm (b), Miss (c) and Correct rejection (d). Some scores derived from the contingency table are the *probability of detection* $POD = a/(a+c)$, the *probability of false detection* $POFD = b/(b+d)$, the *frequency bias* $Bias = (a+b)/(a+c)$ and *base rate* $s = (a+c)/(a+b+c+d)$; see [Jolliffe and Stephenson \(2003\)](#) and [Wilks \(1995\)](#) for more details. Another useful tool is the *Relative Operating Characteristics Curve*, ROC, a graph of probability of detection against probability of false detection. The ROC, introduced in meteorology by [Mason \(1982\)](#), comes from the signal detection theory and measures the ability of the forecast to discriminate between two alternative outcomes.

Since this study is not focused on verifying a single event but on evaluating the general performance of the ensembles, the definition of the observed event is not fixed. Therefore several rainfall amount thresholds (2,

Table II. Contingency table for observed event and forecast event in a 2x2 problem.

| Event Forecast | Observed | |
|-------------------|----------|-----------------------|
| | Yes | No |
| Yes | Hit (a) | False alarm (b) |
| No | Miss (c) | Correct rejection (d) |

30 and 100 mm) have been chosen to define a range of observed events and calculated its corresponding contingency table.

The ROC curves obtained for each EPS and threshold (not shown) indicate a very good performance of the individual members taking into account that a perfect ROC score is a curve that travels from bottom left to top left of the diagram and then across to top right, and that the diagonal line indicates no skill. The area under the ROC curve (ROC area) is thus a good indication of the performance of the system, in fact an area of 0.5 indicates no skill and of 1 a perfect skill. Taking this consideration in mind, [Figure 5](#) shows that the multiphysics EPS members have a better performance than the PV-perturbed ones for all thresholds, but nevertheless both EPSs present ROC areas above 0.76, a very satisfying result. The Bias indicates how the forecast event frequency compares to the observed event frequency. The obtained results for this index ([Fig. 6](#)) show that all members of both EPS overpredict ($Bias > 1$) rainfall amounts less than 5 mm while underpredict ($Bias < 1$) the larger rainfalls amounts. At approximately 5 mm threshold both EPS are nearest to the perfect score ($Bias = 1$). The Bias fast decay towards zero can be due to a sample problem, as the base rate shows a rapid decrease of the number of event samples for extreme precipitation values ([Fig. 6](#)). The Taylor diagrams plot several statistics related to the model performance in a single diagram ([Taylor, 2001](#)), yielding a graphical representation of the decomposition of the mean squared error. These statistics are the correlation coefficient and

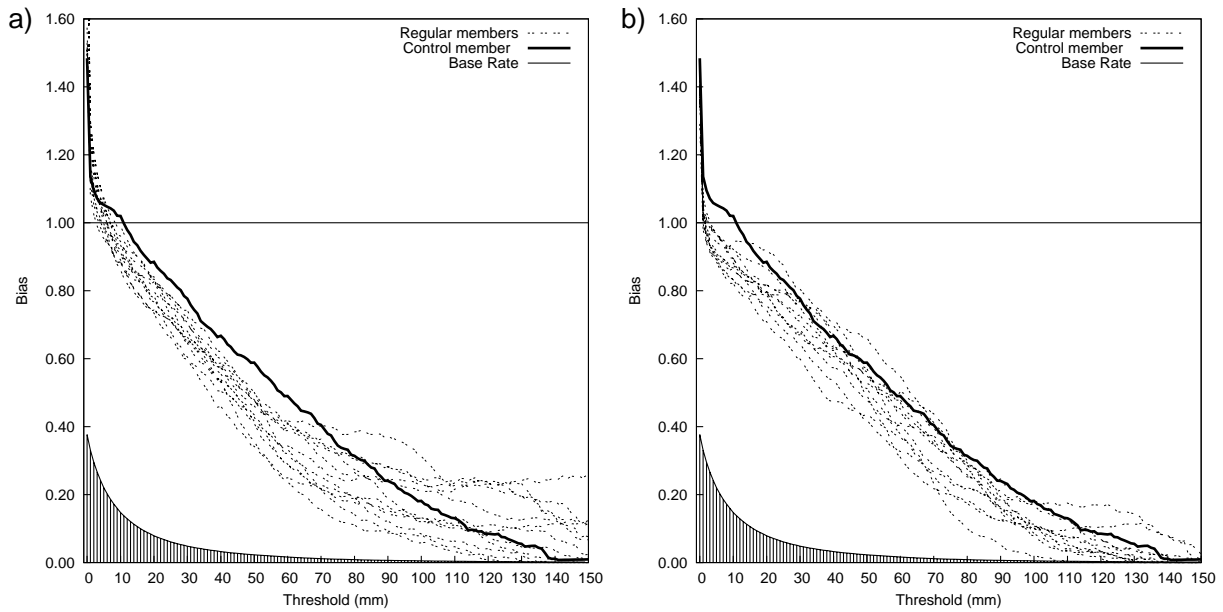


Figure 6. Bias and base rate of the events, for a) Multiphysics and b) PV-perturbed EPS.

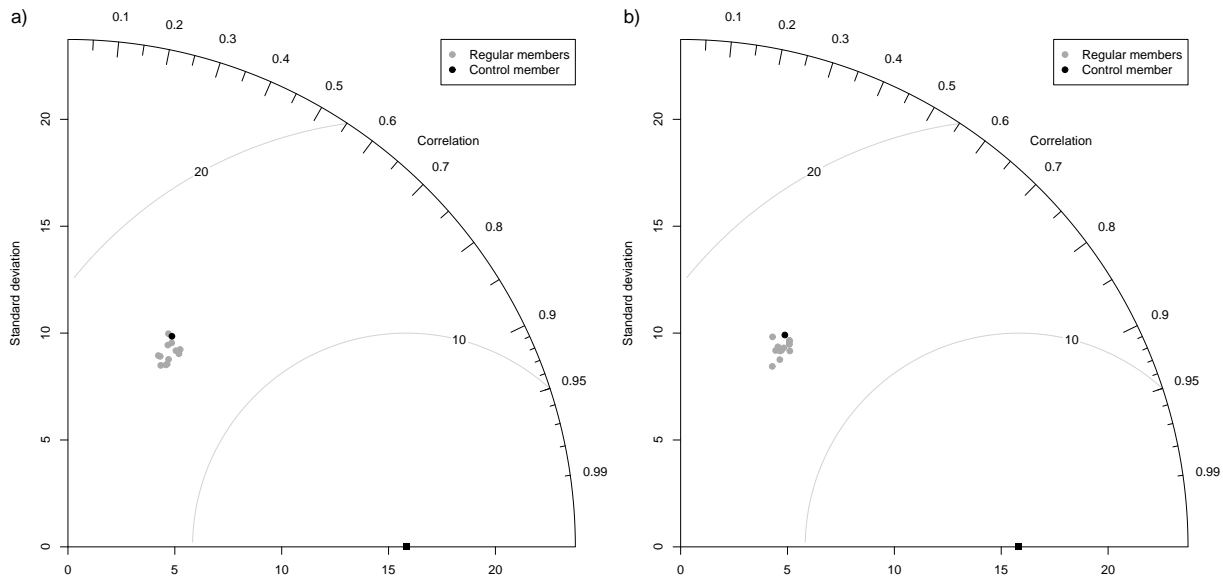


Figure 7. Taylor diagrams for a) Multiphysics and b) PV-perturbed EPS. The black square represents the observed field and the dots the forecasted fields. The radial distance from the origin is proportional to the standard deviation of a pattern. The centered RMS difference between the observed and forecast field is proportional to their distance apart. The correlation between the two fields is given by the azimuthal position of the forecast field. The standard deviation and centered RMS difference units are rainfall millimeters.

the centered pattern root-mean-square difference between the forecast and the observed field, and the standard deviation of both fields. The perfect score is obtained in the Taylor diagrams when the data point representing the forecast field matches up with the observed one. The diagrams (Fig. 7) show that both EPS obtain similar results, approximately all ensemble members present a RMS difference

of 15 mm and a correlation coefficient of 0.5, while the forecast and observation standard deviations are approximately of 10 and 16 mm, respectively. Taking into account that the RMS errors, correlation coefficients and standard deviations are sensitive to discontinuities, noise and outliers, the obtained results, for the rainfall field, are quite good. The fact that individual members from the EPSs

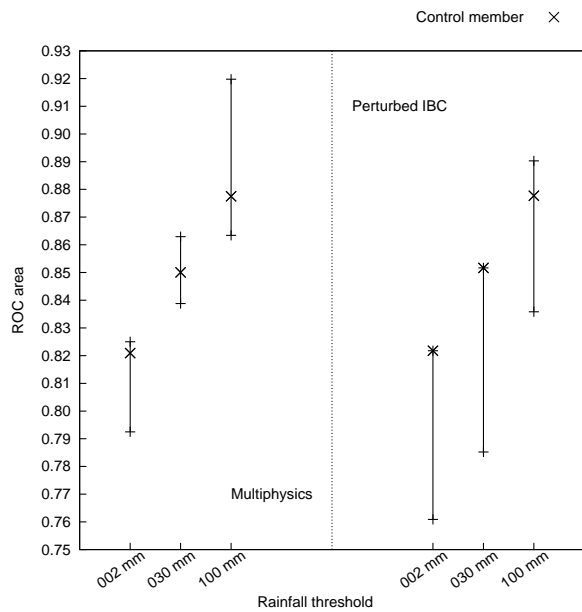


Figure 5. ROC area extremes for the Multiphysics and PV-perturbed ensemble members, as functions of different rainfall event thresholds..

perform similarly well is a positive feature. It is a desirable property that the individual members of an EPS present good performances on they own, so they can build a valuable ensemble system.

In other to verify each EPS as a whole an estimate of the forecast probability of an event needs to be obtained. This forecast probability is calculated as the fraction of the forecasts (or EPS members) predicting the event among all forecast considered (all EPS members). With the purpose of emphasizing the advantages of using a probabilistic forecast against a deterministic forecast, the performance of both types of forecasts is compared while evaluating the performance of the EPS. The chosen approach to define the deterministic forecast is to consider a one-member ensemble. That is, the forecast probability of the deterministic forecast can only be 0 or 100%. To satisfy our aims the deterministic forecast is the control member of the ensemble. In the multiphysics case, it is the member corresponding to the one using the parameterizations of our operational model run (explained above) and in the PV-perturbed ensemble is the one corresponding to

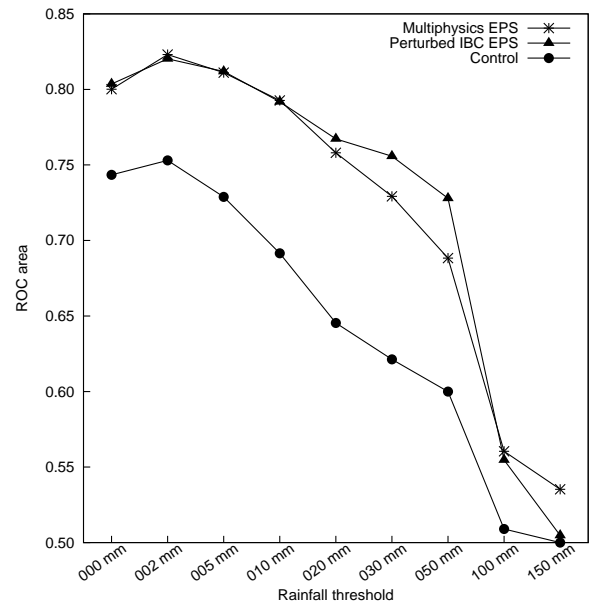


Figure 8. ROC area for the Multiphysics and PV-perturbed ensemble and the control one-member ensemble, as functions of different rainfall event thresholds.

the non-perturbed member using the same multiphysics parameterizations.

The ROC areas obtained for both ensembles (Fig. 8) are very encouraging since the area values lie over 0.8 for rainfall thresholds below 10 mm and over 0.7 below 50 mm. The results above 50 mm are influenced by the small sample problem associated with the small number of observed events. The low 0.5 - 0.6 ROC areas are due to this sample problem and to the difficulty of the EPS predicting the extreme rainfall values. The PV-perturbed ensemble shows better results than the multiphysics in almost all the rainfall thresholds, and both EPS exhibit significantly better results than the control member.

The Bias results (Fig. 9) reveal a similar behavior to the EPS individuals members, in the sense that both EPSs overpredict small rainfall amounts and underpredict large amounts. Between 2 and 10 mm rainfall values, both EPS are nearest to the perfect score. Again the Bias decays fast for large rainfall values due to the sample problem. Like the ROC area, the Bias shows a slightly better performance of the PV-perturbed ensemble than the

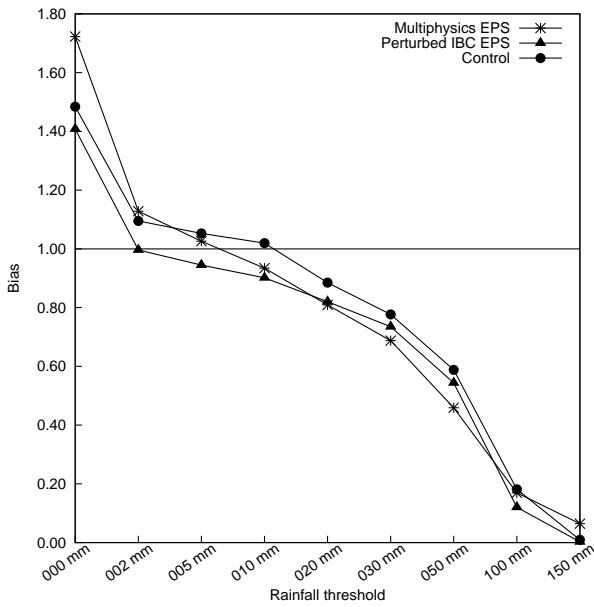


Figure 9. Bias for the Multiphysics and PV-perturbed ensemble and their respective control one-member ensemble, as functions of different rainfall event thresholds.

multiphysics system. The comparison between the EPSs and the control member is not easy since depending on the threshold one can be better than the other. The Brier Score (Brier, 1950) is a verification score for probabilistic forecasts. It is the mean squared error of probabilistic forecasts with events assigned a value of 1 and non-events zero. The Brier Score can be decomposed into the sum of three individual parts related to reliability, resolution and the underlying uncertainty of the observations (Murphy, 1973). Therefore, if the verification sample is partitioned into categories according to the forecast probabilities, the Brier Score can be defined as follows:

$$BS = \frac{1}{n} \sum_{i=1}^k (f_i - o_i)^2 = \frac{1}{n} \sum_{i=1}^k n_i (f_i - \bar{o}_i)^2 - \frac{1}{n} \sum_{i=1}^k n_i (\bar{o}_i - \bar{o})^2 + \frac{1}{n} \sum_{i=1}^k n_i (\bar{o}_i - \bar{o})^2$$

where k is the number of categories (often 11, corresponding to 0,10,...,100%), n the number of realizations of the forecast, f_i the forecast probability corresponding to the i category, n_i the number of times f_i is forecasted, o_i the observation probability corresponding to the i category that is 1 or 0 depending on whether the event occurred

or not, \bar{o}_i the mean occurrence of event for forecast category i and \bar{o} the base rate or climatological mean. The first term on the right-handed side is a measure of reliability, the second term refers to resolution, and the third term represents the uncertainty of the observations, so it is independent of the forecast system. A perfect reliability means a perfect agreement between the forecast values and the observed values, while a perfect resolution means that the forecast system can successfully separate one type of outcome from another. The Brier Score ranges from 0 to 1, and the perfect score is 0. The Brier Score can also be used to define a positively oriented index, the Brier Skill Score, that measures the difference between the Brier score for the forecast and the Brier score for the unskilled standard forecast normalized by the total possible improvement that can be achieved. The unskilled standard forecast is often the climatology, then the BSS can be written as

$$BSS = 1 - \frac{BS}{BS_{climatology}}$$

which ranges from $-\infty$ to 1; 0 indicates no skill compared to the climatology and the perfect score is 1.

The Brier Skill Score results in Figure 10 show that the BSS is almost the same for both EPSs and the control run, while the BS terms show different behaviors depending on the EPS and the control. Both EPSs and control run present good skill for small rainfall thresholds that decreases as the rainfall threshold increases. As one should expect the BS uncertainty term is also the same for both EPSs and control run since it does not depend on the forecast uncertainties but on the observations. The results obtained on the BSS are explained through the BS reliability and BS resolution terms. These two terms are oppositely oriented and they almost compensate each other in the total score. Taking this into account the multiphysics ensemble shows better performance than the PV-perturbed

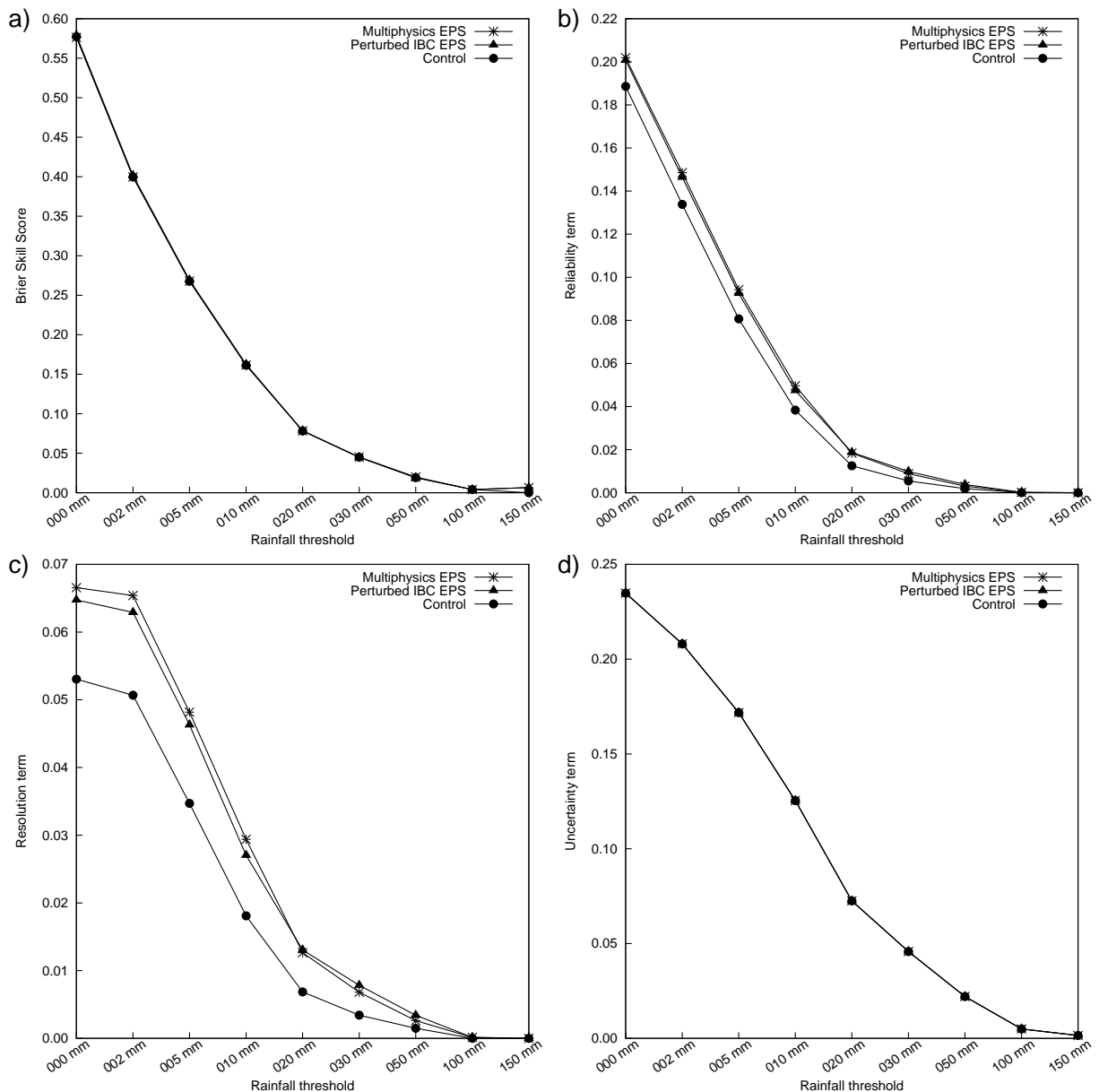


Figure 10. Brier Skill Score and the Brier Score three terms for the Multiphysics and PV-perturbed ensemble and the control one-member ensemble, as function of different rainfall event thresholds. a) Brier Skill Score, b) Brier Score Reliability term, c) Brier Score Resolution, and d) Brier Score Uncertainty.

EPS for both terms, while both ensembles show better performance over the control run for the BS resolution term, and vice versa for the BS reliability term. The attributes diagrams plot the observed frequency against the forecast probability, where the range of forecast probabilities is divided in bins, and also include the no-resolution line and the no-skill line. The perfect score is represented by a curve that matches the diagonal. The no-resolution line represents the climatological frequency of the event, while

the no-skill line presents a 0.5 slope and crosses the no-resolutions and perfect score lines. Figure 11 shows the results for rainfall thresholds of 2, 30 and 100 mm in order to represent the range of thresholds used in this work. The results state how the skill decreases as the threshold increases for both EPSs and the control run. Nevertheless the curves do not migrate out of the regions of no-skill except from the 100 mm threshold where the sample problem is already present as the graph reflects with its fluctuations. Although both EPSs lie within in the skill zone,

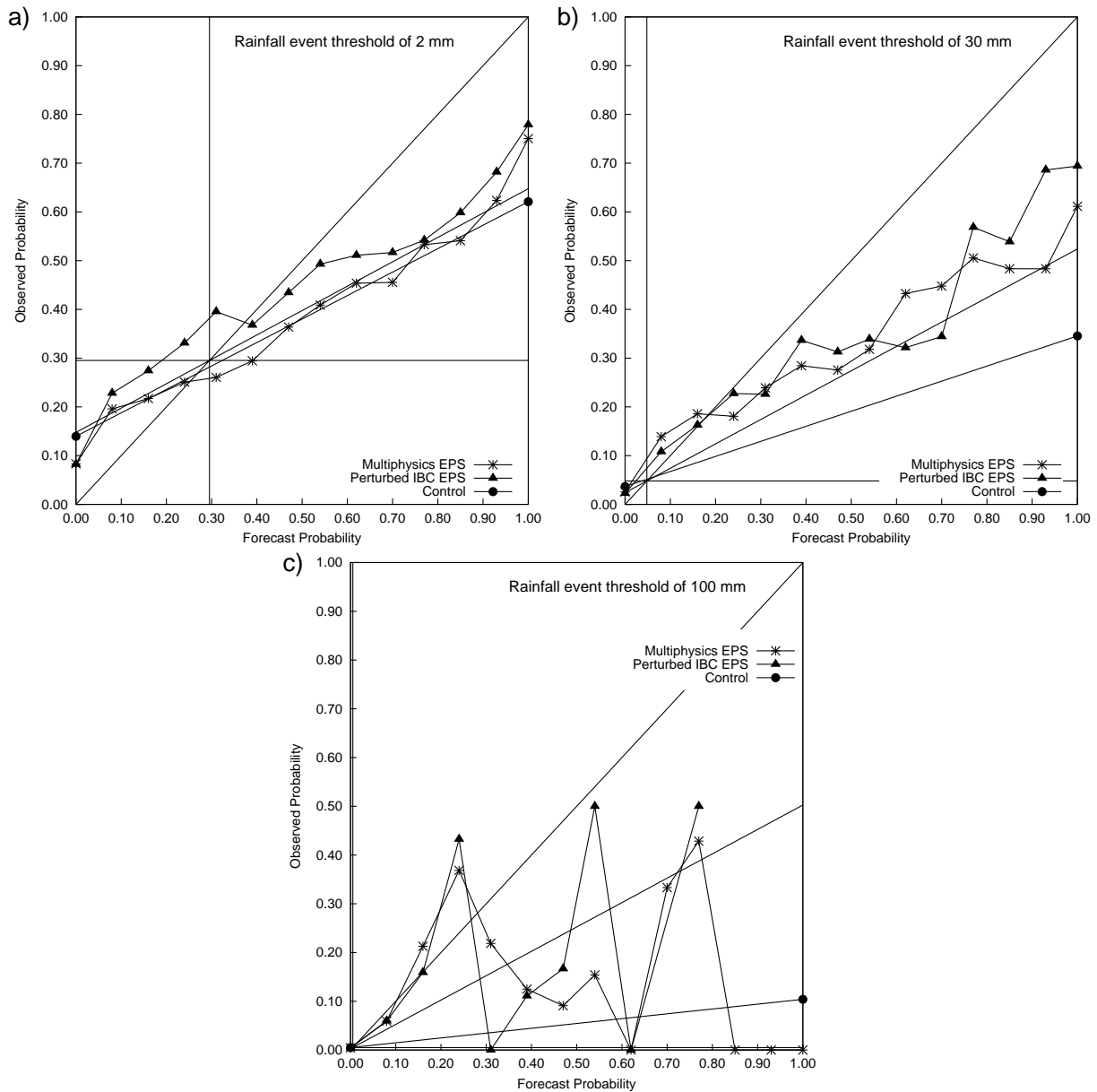


Figure 11. Attributes diagrams for the Multiphysics and PV-perturbed ensemble and the control one-member ensemble, for a) 2 mm, b) 30 mm and c) 100 mm rainfall event thresholds.

the PV-perturbed ensemble presents better skill than the multiphysics EPS in almost all regions. The control run is clearly presenting a bad skill. It is worth to note that the attributes diagrams give graphical information of the BS decomposition terms too. The reliability term corresponds to the weighted distance from the diagonal line to the attribute line, while the resolution term corresponds to the weighted distance from climatology line to the attribute line. Hence Figure 11.a shows a better reliability and resolution for high forecast probability (over 0.4) than for

low forecast probability (below 0.4) for both EPSs, meaning that they underforecast lower values and overforecast higher values (conditional bias). The fact that the PV-perturbed ensemble has higher observed probability than the multiphysics makes that conditional bias worse for low forecast probability and makes it better for high forecast probability. For moderate to extreme rainfall thresholds (Figure 11.b and 11.c) the interpretation is more complex and highly dependent on the probability region, especially for extreme values. The rank histograms (Talagrand *et al.*,

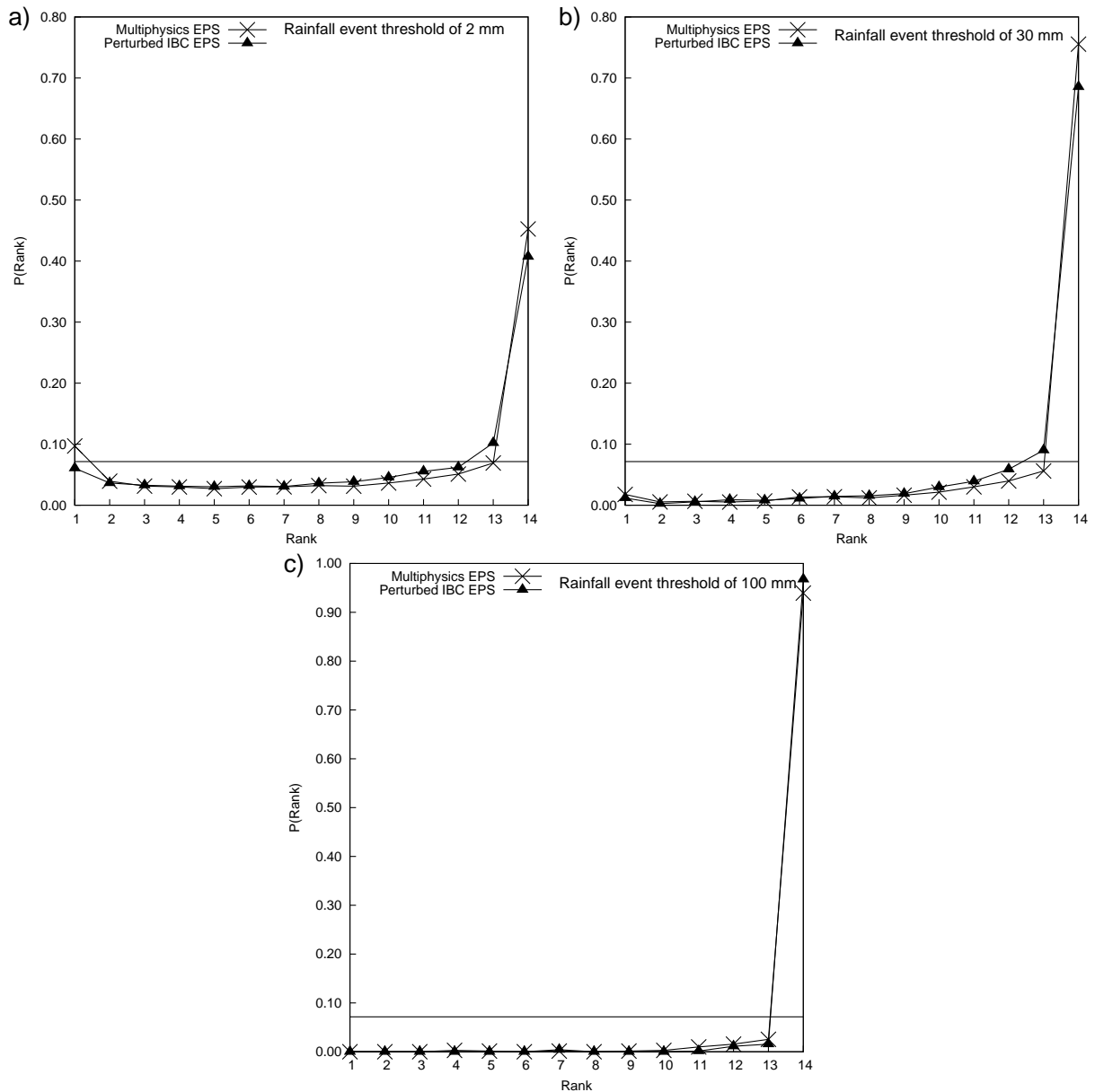


Figure 12. Rank Histograms for the Multiphysics and PV-perturbed ensemble, for a) 2 mm, b) 30 mm and c) 100 mm rainfall event thresholds..

1997) indicate how well the ensemble forecast spread represents the true variability or uncertainty of the observations. The rank histogram is constructed by combining the verification data with the ensemble members and determining what rank in the combined data the verification data represents. The rankings from the forecast samples are then represented in a histogram. A flat histogram means that verification is equally likely to be found anywhere within the ensemble, implying that the ensemble

has the right spread. A U-shape means that the verification is more likely to be found outside of the ensemble range, so that the ensemble spread is too small. A Dome-shape mean that the verification is found near the middle of the ensemble, implying that the spread is too large. An asymmetric curve means that the ensemble contains bias. More detailed information can be found in Hamill (2001). As for the attributes diagrams, the results (Fig. 12) are shown for rainfall thresholds of 2, 30 and 100 mm. All rank histograms present a U-shaped profile, except the 100

mm threshold that leads to an excessive population within the right extreme rank. That is, for the 100 mm threshold almost all the observation values exceed the forecasts provided by the ensemble members, partly due to the limited sample population and also to the well-known problems of mesoscale models to reproduce extreme rainfall values. The U-shape indicates that the spread in both EPSs is too small, as most of the observations fall outside the extremes of the ensemble. The 30 mm threshold presents a form between the U-shape and the right-asymmetric, meaning that the EPSs results fall outside the ensemble extreme with a negative bias. The PV-perturbed ensemble presents slightly better rank histograms than the multiphysics EPS. It should be noted that owing to the rank histogram construction method, it can not be calculated for the control run, a one-member ensemble.

5 Concluding remarks and future outlook

The design of ensemble prediction systems based on multiphysics and perturbed initial and boundary conditions with the aim of improving the short to mid-range numerical forecasts of cyclones and heavy rain events in the western Mediterranean is proved to be a good strategy. The methodology developed for building the PV-perturbed ensemble appears to be a promising tool. On the one hand, the use of a single variable (PV) on which to define perturbations, combined with the PV Inversion Technique, keeps the method simple while ensures modifications of all the meteorological fields without compromising the mass-wind balance. On the other hand, the results show that a skillful PV-perturbed ensemble is obtained with this method.

The verification procedure shows the difficulties of evaluating the rainfall field, highly discontinuous in space and time and observed over irregularly spaced networks,

as well as the sample problems associated with extreme rainfall values. In spite of these difficulties the verification highlights the advantages of EPSs over deterministic forecasts and also that the PV-perturbed ensemble, in general, performs better than the multiphysics.

Encouraged by these results and the better performance of the PV-perturbed EPS over the more traditional multiphysics approach, a new ensemble is currently under development. This ensemble consists of using the PV sensitivity regions calculated by the MM5 adjoint model (Errico, 1997), an *objective* choice, instead of the zones with the most intense PV values and gradients, the *subjective* sensitivity attribution criteria exploited in this paper. Preliminary results are very encouraging, since they show an increase in the ensemble skill and spread.

Acknowledgement

Support from PRECIOSO/CGL2005-03918/CLI and MEDICANES/CGL2008-01271/CLI projects and PhD grant BES-2006-14044, all from the Spanish *Ministerio de Ciencia e Innovación*, is acknowledged.

References

- Brier GW. 1950. Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.* **78**: 1–3.
- Charney JG. 1955. The use of primitive equation of motion in numerical prediction. *Tellus* **7**: 22–26.
- Clark AJ, Gallus Jr WA, Chen T. 2008. Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon. Wea. Rev.* **136**: 2140–2156.
- Davis CA, Emanuel KA. 1991. Potential vorticity diagnostics of cyclogenesis. *Mon. Wea. Rev.* **119**: 1928–1953.
- Dudhia J. 1993. A non-hydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.* **121**: 1493–1513.
- Errico RM. 1997. What is an adjoint model? *Bull. Am. Meteorol. Soc.* **78**: 2577–2591.
- Frank WM. 1983. The cumulus parameterization problem. *Mon. Wea. Rev.* **111**: 1859–1871.

- Garcies L, Homar V. 2009. Ensemble sensitivities of the real atmosphere: Application to mediterranean intense cyclones. *Tellus A* : In Press.
- Grell G, Dudhia J, Stauffer DR. 1995. A description of the fifth-generation of the Penn State/NCAR mesoscale model (MM5). NCAR/TN-398+STR.
- Hamill HC. 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.* **129**: 550–560.
- Homar V, Ramis C, Alonso S. 2002. Numerical study of the october 2000 torrential precipitation event over eastern spain: Analysis of the synoptic-scale stationarity. *Ann. Geophys.* **20**: 2047–2066.
- Horvath K. 2008. Dynamical processes in the upper-troposphere and lee cyclogenesis in the western mediterranean. PhD thesis, Dep. Geophysics, Faculty of Science, University of Zagreb.
- Houtekamer PL, Lefavre L, Derome J, Ritchie H, Mitchell HL. 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.* **124**: 1225–1242.
- Jansà A, Genovés A, Picornell MA, Campins J, Riosalido R, Carretero O. 2001. Western mediterranean cyclones and heavy rain. part 2: Statistical approach. *Meteorol. Appl.* **8**: 43–56.
- Jolliffe IT, Stephenson DB. 2003. *Forecast verification: A practitioner's guide in atmospheric science*. John Wiley and Sons.
- Lorenz EN. 1963. Deterministic nonperiodic flow. *Atmos. Sci.* **20**: 130–141.
- Mason I. 1982. A model for assessment of weather forecasts. *Austr. Meteor. Mag.* **30**: 291–203.
- Meteorological Office. 1962. Weather in the mediterranean. Vol. 1, 362 pp.
- Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. The ecmwf ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.* **122**: 73–119.
- Murphy A. 1973. A new vector partition of the probability score. *J. Appl. Met.* **12**: 534–537.
- Murphy A. 1993. What is a good forecast? an essay on the nature of goodness in weather forecasting. *Wea. Forecasting* **8**: 281–293.
- Murphy A, Winkler R. 1987. A general framework for forecast verification. *Mon. Wea. Rev.* **115**: 1330–1338.
- Palmer F, Molteni F, Mureau R, Buizza P, Chapelet P, Tribbia J. 1992. Ensemble prediction. 188, 45 pp.
- Pellerin G, Lefavre L, Houtekamer P, Girard C. 2003. Increasing the horizontal resolution of ensemble forecasts at cmc. *Nonlinear Proc. Geophys.* **10**: 463–468.
- Reiter E. 1975. *Handbook for forecasters in the mediterranean. part 1: General description of the meteorological processes*. Naval Environmental Research Facility: Monterey, California.
- Romero R. 2001. Sensitivity of a heavy rain producing western mediterranean cyclone to embedded potential vorticity anomalies. *Quart. J. Roy. Meteor. Soc.* **127**: 2559–2597.
- Romero R. 2008. A method for quantifying the impacts and interactions of potential vorticity anomalies in extratropical cyclones. *Quart. J. Roy. Meteor. Soc.* **134**: 385–402.
- Romero R, Doswel III CA, Riosalido R. 2001. Observations and fine-grid simulations of a convective outbreak in northeastern spain: Importance of diurnal forcing and convective cool pools. *Mon. Wea. Rev.* **129**: 2157–2182.
- Romero R, Doswell III CA, Ramis C. 2000. Mesoscale numerical study of two cases of long-lived quasistationary convective systems over eastern spain. *Mon. Wea. Rev.* **128**: 3731–3751.
- Romero R, Martín A, Homar V, Alonso S, Ramis C. 2005. Predictability of prototype flash flood events in the western mediterranean under uncertainties of the precursor upper-level disturbance: The hydroptimet case studies. *Nat. Haz. and Earth. Syst. Sci.* **5**: 505–525.
- Romero R, Martín A, Homar V, Alonso S, Ramis C. 2006. Predictability of prototype flash flood events in the western mediterranean under uncertainties of the precursor upper-level disturbance: the hydroptimet case studies. *Adv. Geosciences* **7**: 55–63.
- Stensrud D, Brooks H, Du J, Tracton M, Rogers E. 1999. Using ensembles for short-range forecasting. *Mon. Wea. Rev.* **127**: 433–446.
- Stensrud DJ. 2007. *Parameterization schemes: Keys to understanding numerical weather prediction models*. Cambridge University Press, 488 pp.
- Stensrud DJ, Bao J, Warner TT. 2000. Using initial conditions and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.* **128**: 2077–2107.
- Talagrand O, Vautard R, Strauss B. 1997. Evaluation of probabilistic prediction systems. *Proceedings, ECMWF Workshop on Predictability*, ECMWF : 1–25.
- Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**: 7183–7192.
- Toth Z, Kalnay E. 1997. Ensemble forecasting at ncep and the breeding method. *Mon. Wea. Rev.* **125**: 3297–3319.
- Wandishin M, Mullen S, Stensrud D, Brooks H. 2001. Evaluation of a short-range multi-model ensemble system. *Mon. Wea. Rev.* **129**: 729–747.
- Wilks D. 1995. *Statistical methods in the atmospheric sciences: An introduction*. Academic Press.