

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

By Maria-del-Mar Vich^{1*}, Romualdo Romero¹ and Harold Brooks²,

¹*Universitat de les Illes Balears, Dpt. Física, Meteorology Group, Palma de Mallorca, Spain;*

²*National Oceanic and Atmospheric Administration, National Severe Storms Laboratory, Norman, Oklahoma, United States*

(Manuscript received 20 April 2010; in final form 25 November 2003)

ABSTRACT

The western Mediterranean is a very cyclogenetic area and many of the cyclones developed over this region are associated with high-impact weather phenomena that affect the society of the coastal countries. Two ensemble prediction systems based on multiphysics and perturbed initial and boundary conditions are designed in order to improve the forecast of these heavy rain episodes. The MM5 mesoscale model nested in the ECMWF forecast fields provides the simulations, run at 22.5 km resolution for a two-day period.

The multiphysics ensemble combines different model physical parameterizations while the other ensemble uses a potential vorticity (PV) inversion technique to perturb the initial state and boundary forcing of the model (PV-perturbed). A PV error climatology derived from the large-scale fields allows to perturb the ECMWF PV fields using the appropriated error range.

The results show an improvement of the prediction skill for both ensembles over the traditional deterministic prediction. Moreover, the PV-perturbed ensemble performs better than the multiphysics EPS. The verification procedure highlights the difficulties of evaluating the rainfall field, since its discontinuities and the limited samples of extreme events affect the scores negatively. Despite this, the results reinforce the idea that an EPS represents an improvement over a deterministic prediction.

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 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 trough (Homar et al., 2002). Thus, the prediction of these events is very important to prevent or reduce the damages they cause.

Since the purpose of this study is to improve the prediction of these high impact weather events, a probabilistic forecast is a suitable approach since it accounts for the uncertainty in the initial state (Lorenz, 1963) and the errors introduced by the models imperfections (Frank, 1983).

The construction of a probabilistic forecast or ensemble prediction system (EPS) can be done using a wide range of techniques. Perturbing the initial state is one of them. 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). Ensembles containing different models or different parameterization schemes have shown more skillfulness 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. The previous ensemble constructing techniques deal with a Global

* Corresponding author.

e-mail: mar.vich@uib.es

Area Model (GAM) while our study requires a Local Area Model (LAM). The transition from GAM to LAM ensembles 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 and an ensemble that perturbs the initial and boundary conditions are studied. The latter ensemble introduces the perturbations on the PV field using a potential vorticity error climatology to delimit them. The EPSs simulations are done by the mesoscale model MM5 nested into the ECMWF forecast large-scale fields. 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) also nest their mesoscale model into Global Area Model fields.

To evaluate the performance of both ensembles a collection of 19 MEDEX¹ cyclones is used as a trial 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 tough 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 EPS measuring 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 a verification process to test the performance of such forecast. These quality attributes can be obtained 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 (Talagrand et al., 1997).

¹ MEDEX is the Mediterranean Experiment on cyclones that produce high impact weather in the MEDiterranean, a project endorsed by WMO (<http://medex.aemet.uib.es>).

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 heavy 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 ². 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. Since MEDEX provides a suitable database for our study, 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 1).

Finally, a precipitation dataset is also needed for the verification process. This dataset comes from the AEMET (Agencia Estatal de Meteorología - Spanish Metoffice) climatological raingauge 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

² See <http://mm5forecasts.uib.es>

Mediterranean influenced regions of Spain, with more than 2300 stations depending on the event.

3 Ensembles Construction

Each ensemble consists of 13 MM5 runs, one of them corresponds to the control member. 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 nodes, centered at 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 our group. Since the observed daily rainfall accumulations are from the 06 to 06 UTC period, the forecasting period starts at 00 UTC and extends for 54 h, in order to simplify the verification process. In addition, forecast gridded fields are interpolated over the raingauge locations for the verification. The MEDEX trial set will be covered by all possible simulations comprising a MEDEX day within the 24-48 h forecast range (56 in total according to Table 1).

3.1 *Perturbed initial and boundary conditions ensemble*

The ensemble perturbations are introduced into the PV field using the zones of most intense values and gradients of this field as guidance. This choice assumes 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. In practical terms, this guidance field is defined as the difference of the PV field and a highly smoothed version of it. Then we define the perturbations volumes as the spatial volumes where the guidance field exceeds, in absolute value, a certain threshold. This threshold is defined by the domain averaged guidance field in absolute value.

The reason of perturbing the PV field is that thanks to the potential vorticity inversion technique developed by Davis and Emanuel (1991) these perturbations can also be extended to the wind and temperature field in a dynamically consistent manner. The 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 already 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. Then 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 quantitative assessment of the PV uncertainty range has been derived.

A Potential Vorticity error climatology is derived 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. This climatology is obtained from the 19 MEDEX cyclonic episodes described above, sampled over a 200x200 grid points large domain at 6 h intervals at the following pressure levels (100, 200, 300, 400, 500, 700, 850, 925 and 1000 hPa). The use of a domain bigger than the simulation one is done to avoid boundary issues when perturbing the model forcing. The PV error climatology will tabulate the coefficients of the functions fitted to the displacement and intensity error percentiles extracted from this large amount of data, for each pressure level.

The displacement error (DE) represents the minimum displacement of the forecast field necessary to obtain maximum local correlation with the analysis field. DE is found for each grid point displacing a centered forecast PV field area of 450x450 km in all directions, up to 225 km. The minimum displacement showing local maximum correlation between the displaced field and the analysis field is assigned as the corresponding displacement error. This process is repeated for each domain grid point. The choice of the area 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.

The intensity error (IE) represents the difference between the displaced forecast and analysis PV field. It is calculated for each domain grid point as the difference in PV field

between the aforementioned optimally displaced forecast field and the analysis field averaged over the corresponding area, 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 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. Analytical functions have been fitted to the percentile levels of the displacement error percentiles (Fig. 3) and the intensity error (Fig. 4) with a linear-like and power-law function, respectively.

These DE and IE percentiles (PV error climatology) are used to introduce realistic perturbation into the defined perturbation volumes. For each perturbation volume the intensity (magnitude and positive/negative orientation) and displacement magnitude perturbations are assigned randomly. The displacement direction perturbation is chosen randomly but will be the same for all the perturbation volumes in order to avoid discontinuities in the perturbed PV field. This assignment is done through the percentile level, so once the percentile level is randomly generated the perturbation can be obtained using the above mentioned tabulated coefficients which already consider the corresponding pressure level and PV value. It is worth to notice that a superior limit is imposed to the IE ($IE_{\max} = 200\%$) because this relative error tends to infinity when PV tends to zero due to the power-law function properties (see Fig. 4).

3.2 *Multiphysics ensemble*

The multiphysics ensemble is generated using a variety of physical parameterizations available in MM5 which, although computationally inexpensive, are adequate for mesoscale simulations of precipitation systems. Specifically, the multiphysics sets are the result of combining three explicit moisture schemes (Goddard 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 the ensemble members in order to assure that all ensemble members are good enough to be included within the ensemble. Afterwards, each EPS is verified using probabilistic scores and indices, assuming each ensemble member (deterministic forecast) as an independent realization of the same underlying process. The verification is focused on the 24 h accumulated precipitation period that corresponds to the second day of simulation.

Since this study is not focused on verifying a single observation threshold but on evaluating the general performance of the ensembles, the definition of the observed event is not fixed. Therefore several rainfall amount thresholds have been defined as observed events.

The area under the ROC curve (ROC area) is a good indicator of the ability of the forecast to discriminate between two alternative outcomes, in fact an area of 0.5 indicates no skill and of 1 a perfect skill (see Jolliffe and Stephenson (2003) and Wilks (1995) for more details on many verification scores). Figure 5 shows that the multiphysics EPS members perform better than the PV-perturbed ones for all thresholds, 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 results for this index (Fig. 6) show that all members of both ensembles 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, since 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 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 when the data point representing the forecast field matches up with the observed one. The diagrams (Fig. 7) show similar results for both EPSs, 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 perform similarly well is a positive feature since they make a good basis to build a valuable ensemble system. In order to verify the EPSs as a whole an estimate of the forecast probability of an event is calculated as the fraction of the EPS members predicting the event among all 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 also compared. The chosen approach to define the deterministic forecast is to consider a one-member ensemble made of the control member. The control member corresponds to the non-perturbed run using the physical parameterizations of our operational model, a member present on both ensembles by construction.

The ensembles ROC areas results (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 low ROC area values above 50 mm are influenced by a sample problem associated with the small number of observed events already noted for the individual ensemble members, and also by the difficulty of the EPS in predicting extreme rainfall values due to the 22.5 km grid length. The PV-perturbed ensemble shows better results than the multiphysics in almost all thresholds while both EPSs exhibit significantly better results than the control member.

The Bias results (Fig. 9) reveal a similar behavior to the EPS individual members, in the sense that both EPSs overpredict small rainfall amounts and underpredict large amounts. The higher score for both EPSs lies between 2 and 10 mm rainfall values. Like the ROC area, the Bias shows a slightly better performance of the PV-perturbed ensemble than the multiphysics system and a fast decay in skill for large rainfall values. Because of this from now on the verification procedure is focused on the thresholds ranging from 0 to 50 mm due to both the lack of statistical significance and to modeling limitations. The comparison between the EPSs and the control member is not easy since one can be better than the other depending on the threshold.

The Brier Score (Brier, 1950), defined by the mean squared error of probabilistic forecasts with events assigned a value of 1 and non-events zero, 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:

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

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 Brier Score ranges from 0 to 1 with a perfect score of 0. The first term on the right-hand 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. This score can 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

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

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

Figure 10 shows that the Brier Skill Score 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. The BS uncertainty term does not depend on the forecast but on the observations uncertainties so it is the same for both EPSs and the control run. 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 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 including the no-resolution and no-skill line. Figure 11 shows the results for rainfall thresholds of 2 and 30 mm which states 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 skill. Although both EPSs lie within in the skill zone, 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) as already stated by the Bias results (Fig. 9). 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. From moderate rainfall thresholds (Figure 11.b) the interpretation is more complex and highly dependent on the probability region.

The rank histograms (Talagrand et al., 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 represents the verification data in the combined data. 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 the ensemble spread is too small. An asymmetric curve means that the ensemble contains bias. More detailed information on rank histogram interpretation can be found in Hamill (2001). The rank histograms for rainfall thresholds of 2 and 30 mm (Fig. 12) present an U-shaped profile which 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 an U-shape combined with a strong right-asymmetry 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 designed multiphysics and perturbed initial and boundary conditions ensemble prediction systems have proved to be a good strategy to improve the short to mid-range numerical forecasts of cyclones and heavy rain events in the western Mediterranean. 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 (potential vorticity) 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 highlights 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 and difficulties of the model to forecast extreme precipitations. In spite of these difficulties the verification stresses the advantages of an EPS over a deterministic forecast and also that our 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 most intense PV values and gradients regions, the *semi-objective* sensitivity attribution criteria exploited in this paper.

6 Acknowledgments

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, G. W. 1950. Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.* **78**, 1–3.
- Charney, J. G. 1955. The use of primitive equation of motion in numerical prediction. *Tellus* **7**, 22–26.
- Clark, A. J., Gallus Jr., W. A. and 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, C. A. and Emanuel, K. A. 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, R. M. 1997. What is an adjoint model?. *Bull. Am. Meteorol. Soc.* **78**, 2577–2591.
- Frank, W. M. 1983. The cumulus parameterization problem. *Mon. Wea. Rev.* **111**, 1859–1871.
- Garcias, L. and Homar, V. 2009. Ensemble sensitivities of the real atmosphere: Application to mediterranean intense cyclones. *Tellus A* p. In Press.
- Grell, G., Dudhia, J. and Stauffer, D. R.: 1995, NCAR Tech. NCAR/TN-398+STR.
- Hamill, H. C. 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.* **129**, 550–560.
- Homar, V., Ramis, C. and 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, PhD thesis, Dep. Geophysics, Faculty of Science, University of Zagreb.
- Houtekamer, P. L., Lefaiivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.* **124**, 1225–1242.
- Jansà, A., Genovés, A., Picornell, M. A., Campins, J., Riosalido, R. and Carretero, O. 2001. Western mediterranean cyclones and heavy rain. part 2: Statistical approach.. *Meteorol. Appl.* **8**, 43–56.
- Jolliffe, I. T. and Stephenson, D. B. 2003. Forecast verification: A practitioner's guide in atmospheric science Jolliffe, I. T. and Stephenson, D. B. John Wiley and Sons.
- Lorenz, E. N. 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, Air Ministry. Vol. 1, 362 pp.
- Molteni, F., Buizza, R., Palmer, T. N. and 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. and 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. and Tribbia, J.: 1992, ECMWF Research Dept. Tech. Memo. 188, 45 pp.
- Pellerin, G., Lefaiivre, L., Houtekamer, P. and 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
Reiter, E. 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, C. A. and 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, C. A. and 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. and 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. and 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. and Rogers, E. 1999. Using ensembles for short-range forecasting. *Mon. Wea. Rev.* **127**, 433–446.
- Stensrud, D. J. 2007. Parameterization schemes: Keys to understanding numerical weather prediction models Stensrud, D. J. Cambridge University Press, 488 pp.
- Stensrud, D. J., Bao, J. and Warner, T. T. 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. and Strauss, B. 1997. Evaluation of probabilistic prediction systems. *Proceedings, ECMWF Workshop on Predictability*, ECMWF pp. 1–25.
- Taylor, K. E. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**, 7183–7192.
- Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at ncep and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.
- Wandishin, M., Mullen, S., Stensrud, D. and 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 Wilks, D. Academic Press.

Table 1. 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

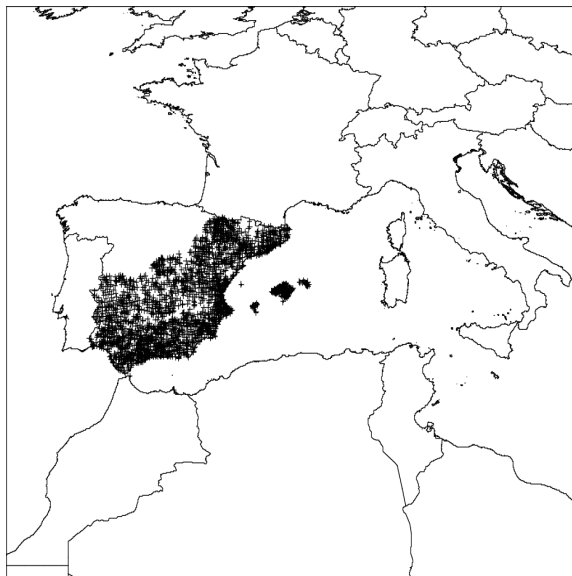


Figure 1. Geographical domain used for the MM5 numerical simulations. The spatial distribution of the AEMET rain gauge network used in the verification procedure is plotted using crosses.

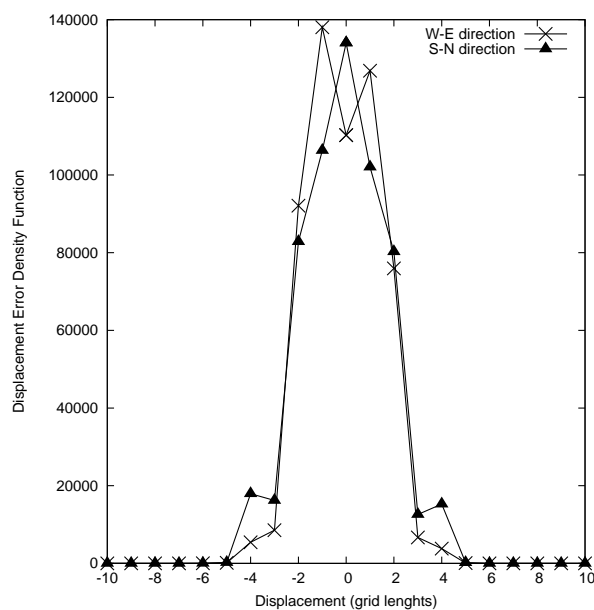


Figure 2. Displacement error density function at 300 hPa. It represents the number of occurrences (grid points) as function of the displacement error in grid lengths. Results for the South-North (solid line) and West-East (dotted line) directions are represented separately.

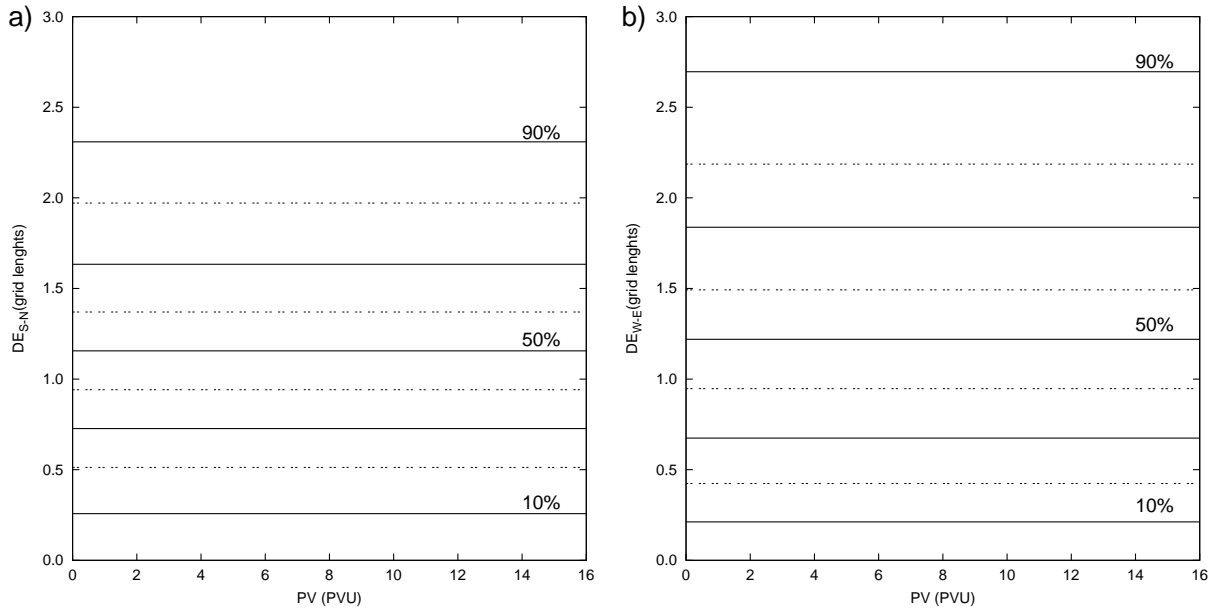


Figure 3. Displacement Error percentile levels along a) South-North and b) West-East directions at 300 hPa.

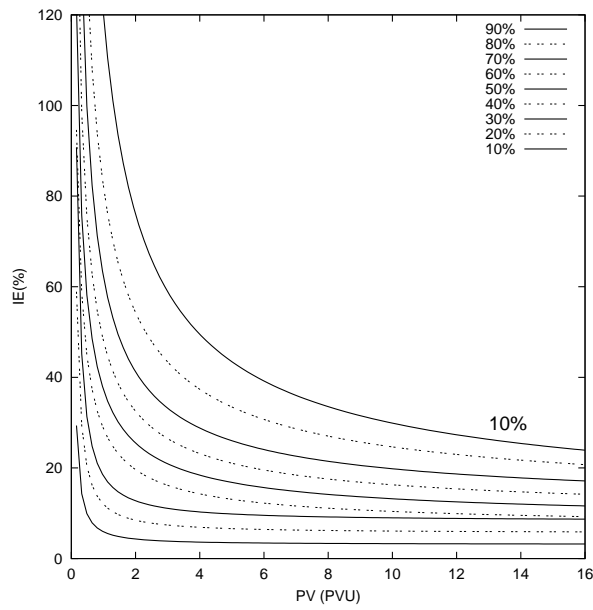


Figure 4. Intensity Error percentile levels at 300 hPa.

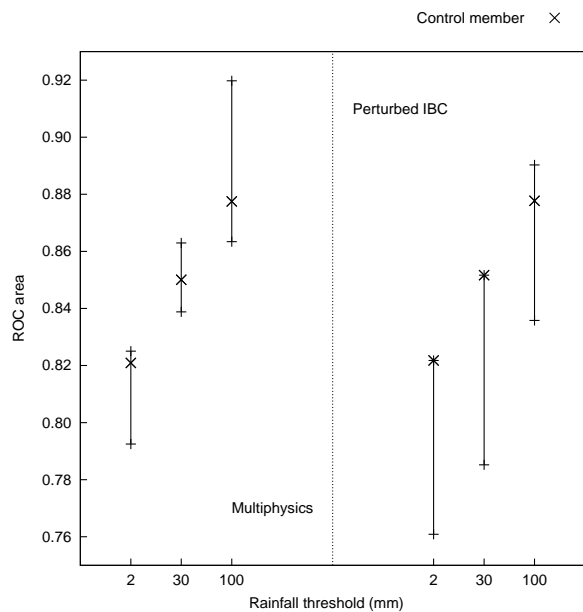


Figure 5. ROC area extremes for the multiphysics and PV-perturbed ensemble members, as functions of different rainfall event thresholds.

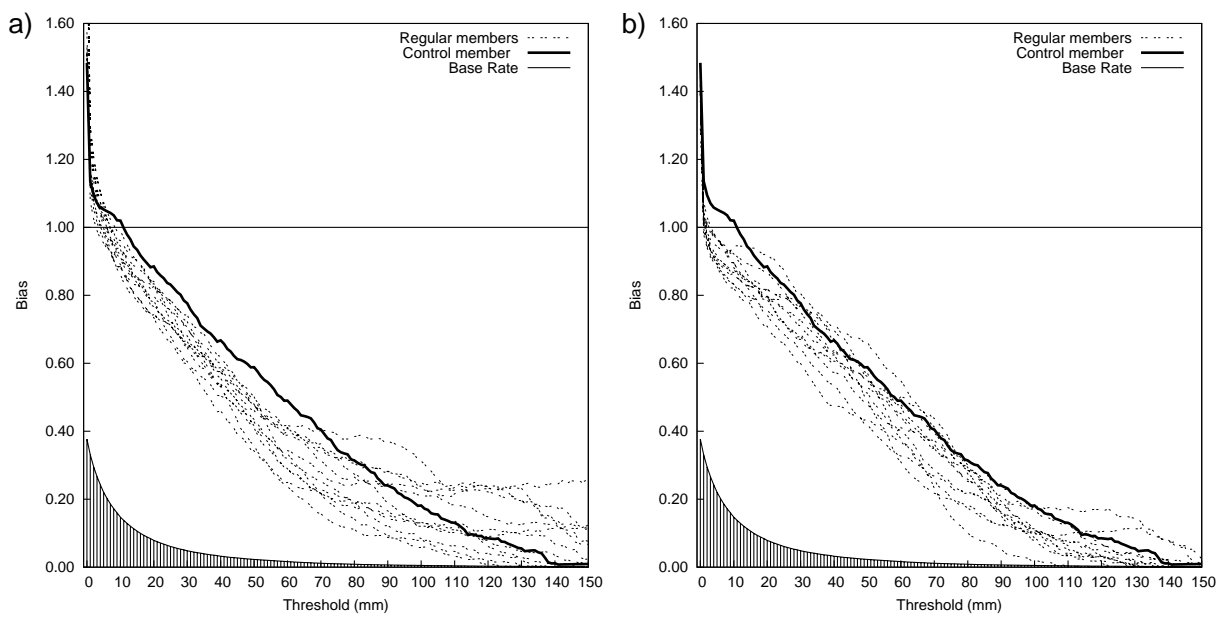


Figure 6. Bias and base rate of the events, for a) multiphysics and b) PV-perturbed EPS.

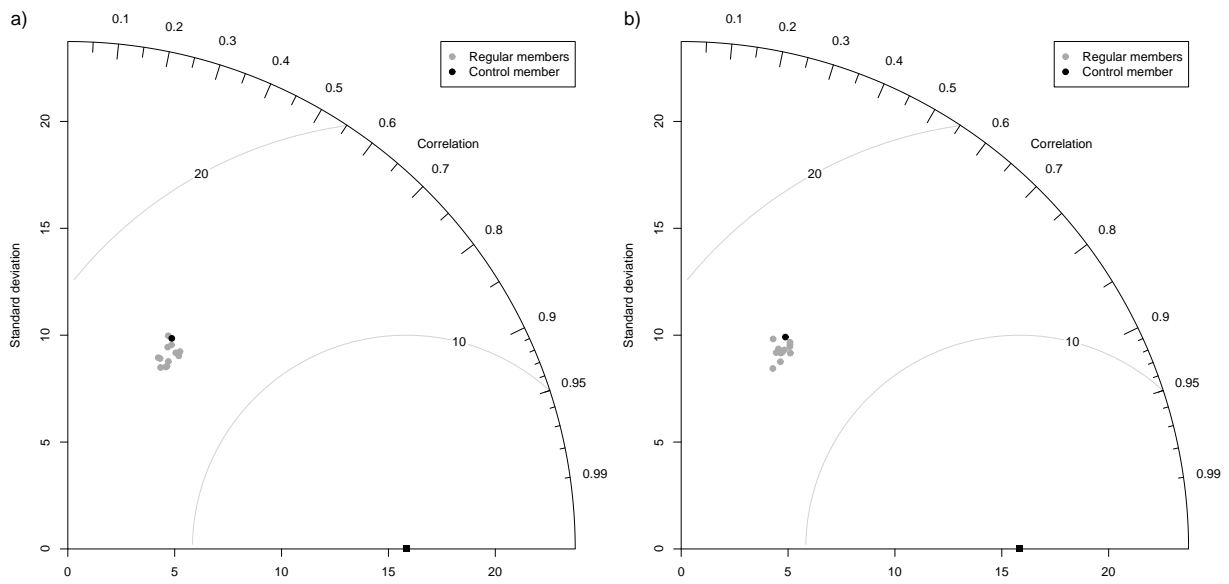


Figure 7. Taylor diagrams for a) multiphysics and b) PV-perturbed EPS. The black square represents the observed field and the dots the forecasted fields. The radial distance from the origin is proportional to the standard deviation of a pattern. The centered RMS difference between the observed and forecast 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.

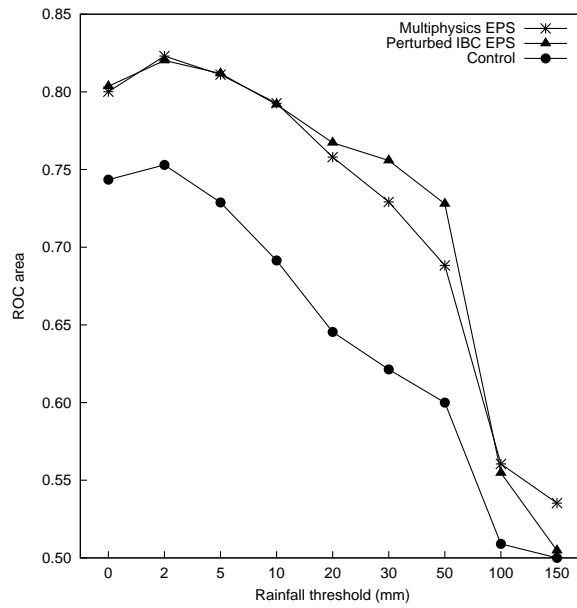


Figure 8. ROC area for the multiphysics and PV-perturbed ensemble and the control one-member ensemble, as functions of different rainfall event thresholds.

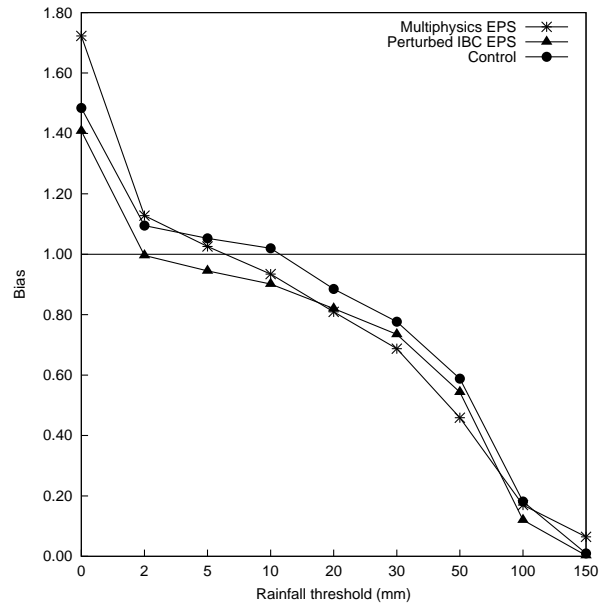


Figure 9. Bias for the multiphysics and PV-perturbed ensemble and their respective control one-member ensemble, as functions of different rainfall event thresholds.

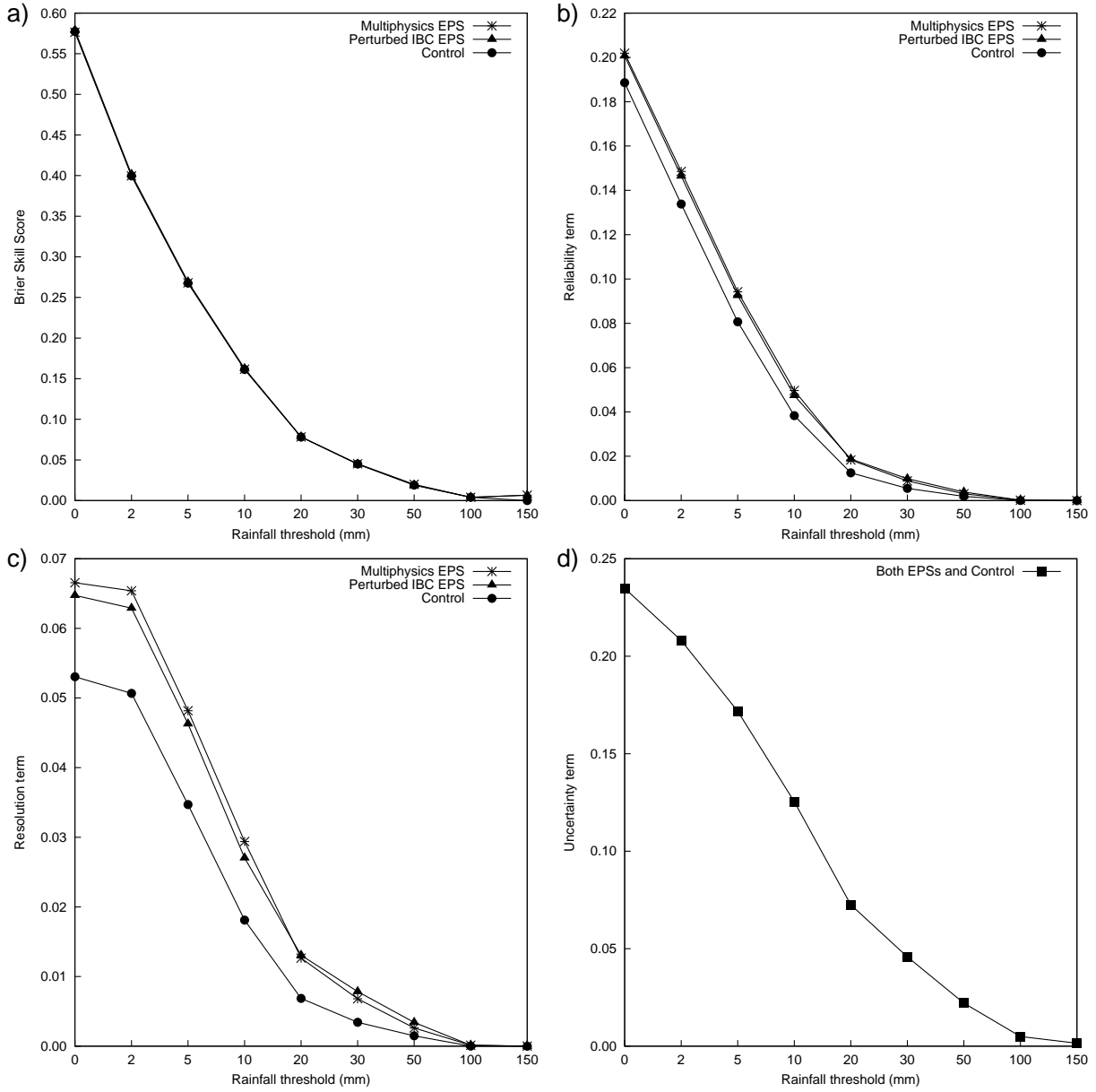


Figure 10. Brier Skill Score and the Brier Score three terms for the multiphysics and PV-perturbed ensemble and the control one-member ensemble, as function of different rainfall event thresholds. a) Brier Skill Score, b) Brier Score Reliability term, c) Brier Score Resolution, and d) Brier Score Uncertainty.

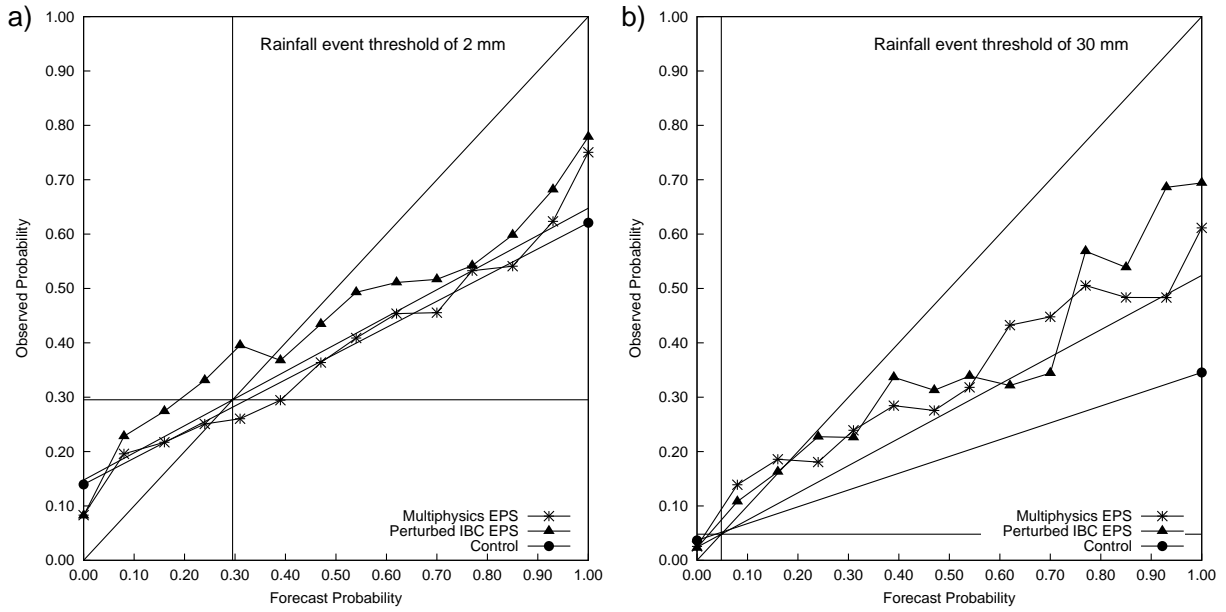


Figure 11. Attributes diagrams for the multiphysics and PV-perturbed ensemble and the control one-member ensemble, for a) 2 mm and b) 30 mm rainfall event thresholds. The perfect score is represented by a curve that matches the diagonal and the no-resolution line (both horizontal and vertical lines) 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.

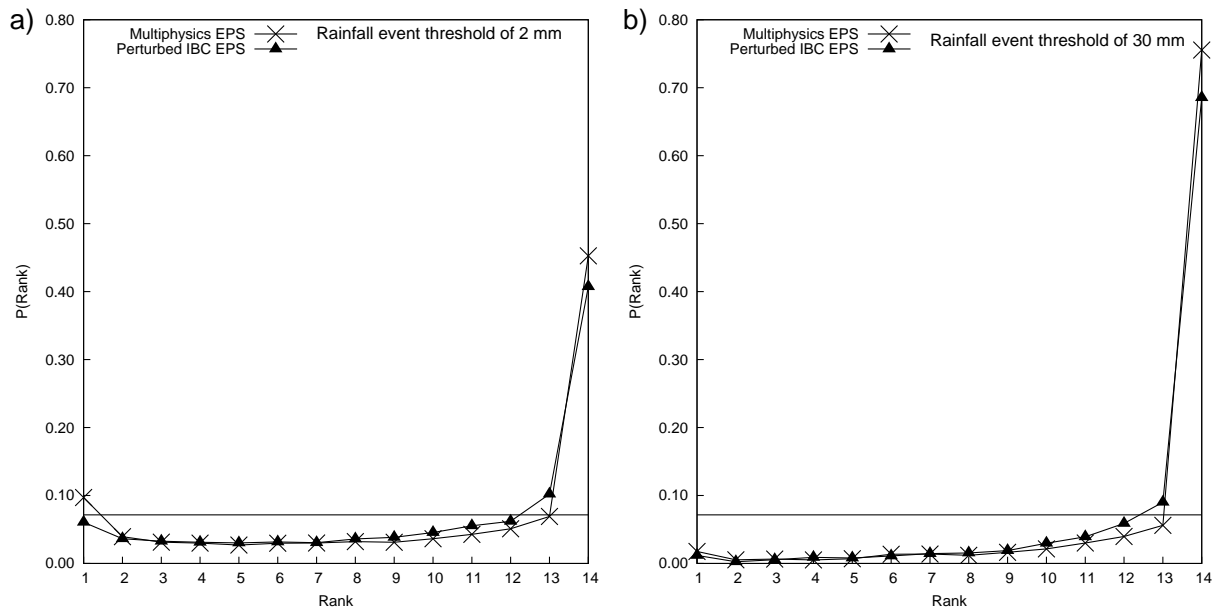


Figure 12. Rank Histograms for the multiphysics and PV-perturbed ensemble, for a) 2 mm and b) 30 mm rainfall event thresholds..