**Natural Hazards
and Earth
System Sciences**

Multiphysics superensemble forecast applied to Mediterranean heavy precipitation situations

M. Vich and R. Romero

Meteorology Group, Physics Dept., University of the Balearic Islands, Palma de Mallorca, Spain

Abstract. The high-impact precipitation events that regularly affect the western Mediterranean coastal regions are still difficult to predict with the current prediction systems. Bearing this in mind, this paper focuses on the superensemble technique applied to the precipitation field. Encouraged by the skill shown by a previous multiphysics ensemble prediction system applied to western Mediterranean precipitation events, the superensemble is fed with this ensemble.

The training phase of the superensemble contributes to the actual forecast with weights obtained by comparing the past performance of the ensemble members and the corresponding observed states. The non-hydrostatic MM5 mesoscale model is used to run the multiphysics ensemble. Simulations are performed with a 22.5 km resolution domain (Domain 1 in http://mm5forecasts.uib.es) nested in the ECMWF forecast fields. The period between September and December 2001 is used to train the superensemble and a collection of 19 MEDEX cyclones is used to test it.

The verification procedure involves testing the superensemble performance and comparing it with that of the poor-man and bias-corrected ensemble mean and the multiphysic EPS control member. The results emphasize the need of a well-behaved training phase to obtain good results with the superensemble technique. A strategy to obtain this improved training phase is already outlined.

1 Introduction

The superensemble forecast technique is a powerful post-processing method for the estimation of weather forecast parameters, like precipitation. In order to accomplish an improvement in the prediction skill of the heavy precipitation events that characterize the western Mediterranean coastal countries the superensemble technique is tested in the region. In previous studies where the goal was also to improve the prediction skill of these potentially dangerous events, similar approaches like ensemble prediction systems (EPS) were tested in the same region with good results. For example, Vich et al. (2010) showed an improvement of prediction skill if an EPS based on varying the model physical parameterization was used instead of a deterministic forecast.

The novelty of the superensemble method is that instead of poor man ensemble mean, where each EPS member weights equally, the method rewards or punishes the past performance of each member assigning weights accordingly. During the training phase, when these weights are calculated, the method computes a linear regression of the observed data against the performance of each ensemble member.

Afterwards, the actual forecast is derived using the data gathered through the training phase and the current ensemble members forecasts, this is called the forecast phase. The superensemble technique is described and applied successfully for medium range real-time global weather forecast in Krishnamurti et al. (1999, 2000a,b, 2001). As examples of more recent applications, Cane and Milelli (2005, 2006) apply the superensemble approach on variables like temperature and wind, while Yun et al. (2005) focuses on seasonal precipitation, Cane and Milelli (2010) on average precipitation over warning areas and Krishnamurti et al. (2008) on the diurnal cycle of the precipitation.

Correspondence to: M. Vich
(mar.vich@uib.es)

The superensemble weights are obtained for each observation location through a linear regression technique involving a minimization function that acts to limit the spread between the variables of the members and the observed state. This minimization function is described by

$$G = \sum_{t=1}^{T} (S_t - O_t)^2 \, , \tag{1}$$

where G is the minimization function, T is the length of the training period, S_t is the superensemble prediction, and O_t is the observed state. The length of a particular training set is important for achieving high skill forecasts (roughly 4 months of past daily forecast are vital for that purpose) as well as the quality of the observational database (Krishnamurti et al., 2001). The superensemble is derived in the forecast phase using the data gathered through the training phase and the current ensemble members forecasts by the following expression

$$S_t = \bar{O} + \sum_{i=1}^{N} a_i \left(F_{t_i} - \bar{F}_i \right) \, , \tag{2}$$

where $\bar{O}$ is the observed mean value of the forecasting variable in the training phase, N is the number of ensemble members, a_i represents the regression coefficient or weight, for member i. F_{t_i} is the variable forecast by member i, and $\bar{F}_i$ is the mean of this particular variable over all the forecasts in the entire training period.

In this work, the performance of the superensemble during the forecast phase is evaluated using the data provided by the MEDEX[1] project, which consists of a collection of cyclonic events associated with floods and/or strong winds over the western Mediterranean, the kind of events that this study aims. The training phase period consists of a four-month period comprising September, October, November and December 2001. It is worth to notice that both phases are focused on the precipitation field, a field with direct impact on the society but difficult to predict due to its highly discontinuous nature.

This paper describes the application of the superensemble technique to western Mediterranean cyclones associated with heavy rainfall. A description of the construction of the superensemble is detailed in Sect. 2. Section 3 presents the results obtained by the superensemble in the verification procedure. Finally, some concluding remarks are found in Sect. 4.

2 Superensemble construction

The superensemble building requires different databases for each phase. In this study, the forecast phase consists of a collection of 19 MEDEX cyclones between September 1996 and October 2002 (see Vich et al., 2010 for a more detailed description) while the training phase consists of a very wet four-month period, September–December 2001, characteristic of the precipitation climatology of the region during Autumn.

The group of forecast members also needed to build the superensemble is provided by the multiphysics EPS develop in Vich et al. (2010). This multiphysics 13-member ensemble is generated using a variety of physical parameterizations available in mesoscale model MM5, specifically 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 simulation domain is defined as a 22.5 km resolution horizontal grid mesh with 120×120 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[2].

The meteorological fields used to initialize and force the model are provided by the ECMWF and the observational data by AEMET (Agencia Estatal de Meteorología – Spanish Weather Service) climatological raingauge network. The observational data consists of 24 h accumulated precipitation from 06:00 UTC to 06:00 UTC the next day, and the meteorological fields correspond to the ECMWF 24 h forecast fields. The use of 24 h forecasts instead of analyses, the best available guess of the atmospheric state, is due to computational limitations in our group, for a given future time, a previous forecast is available earlier than the analyses. The UIB Meteorology Group has been running the MM5 model on a daily basis for some years initializing it with global coarse resolution 24 h forecast fields valid at 00:00 UTC and forced at the lateral boundaries with the subsequent data (i.e. with 30, 36,…,72 h global forecasts) and we wish to test the superensemble technique exactly in the same quasi-operational framework. Figure 1 shows the raingauge network spatial distribution over the Mediterranean influenced regions of Spain, with more that 2300 stations for each event. In order to compare the observations with the regular gridded forecast fields, the fields are interpolated over the observational stations.

---

[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).

[2]See http://mm5forecasts.uib.es.

Nat. Hazards Earth Syst. Sci., 10, 2371–2377, 2010

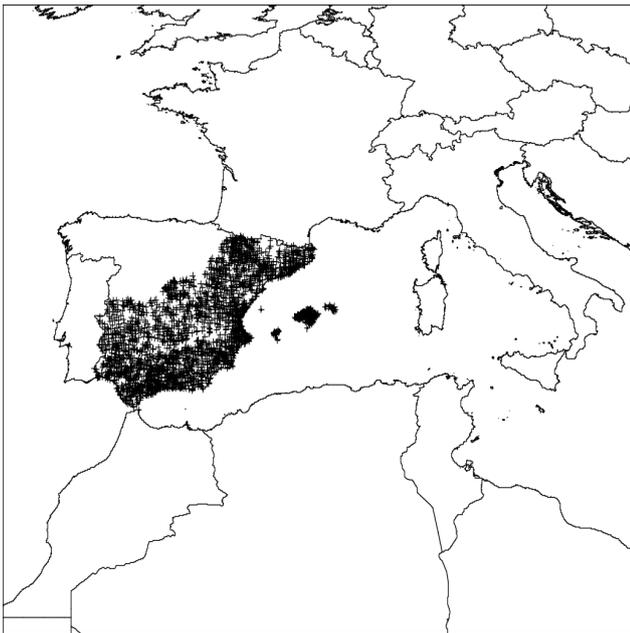www.nat-hazards-earth-syst-sci.net/10/2371/2010/

Fig. 1. Geographical domain used for the MM5 numerical simulations. The spatial distribution of the AEMET raingauge network used for the verification procedure is plotted using crosses.

Fig. 2. ROC area for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds.

3 Experiments and results

The evaluation of the superensemble performance for the rainfall field is done thanks to a wide range of verification indices and also comparing the superensemble results with that of the ensemble mean (a simple average of all the members), the bias-corrected ensemble mean (i.e. expression 2 with $a_i = 1/N$) and the multiphysics EPS control member, the operational model run of our group. It is worth to notice that for both phases, training and forecast, all computations have been done for the 24 h accumulated precipitation field considering both the 6–30 h and 30–54 h accumulated periods indistinctly, this implies that the number of days for both phases are doubled. Eventhough Krishnamurti et al. (2000b) separates both time windows we have checked that for this study merging them does not affect negatively the superensemble performace and we gain statistical significance. 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. For example, if a basement gets flooded when it rains more than 50 mm day^{-1} this would be the observed event, since such threshold separates safety from disaster. Here nine rainfall amount thresholds (0, 2, 5, 10, 20, 30, 50, 100 and 150 mm) have been defined as observed events.
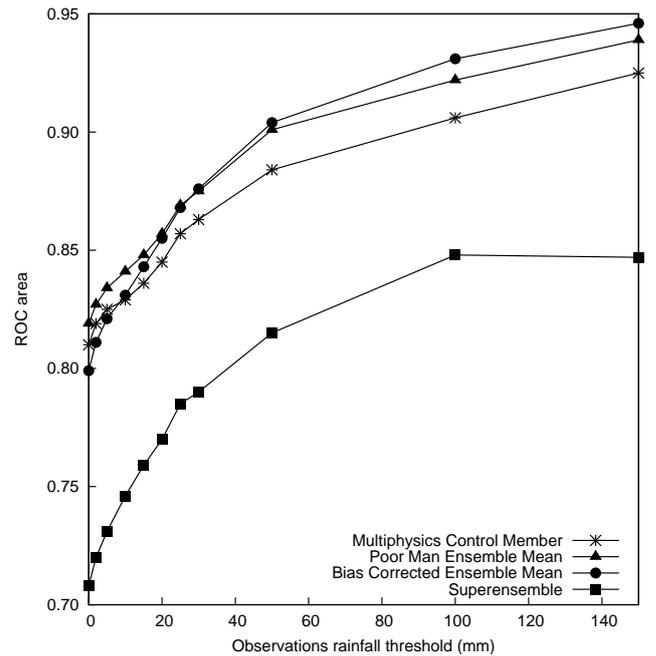
The Relative Operating Curve or ROC measures the ability of the forecast to discriminate between two alternative outcomes, thus measuring resolution. The ROC is obtained plotting probability of detection (fraction of the observed events that were forecast) against the probability of false detection (fraction of the non-observed events that were forecast). The area under the ROC curve (ROC area) is frequently used as score, 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). The ROC area results plotted on Fig. 2 show that the bias corrected ensemble mean performs better than the other forecasts followed by the ensemble mean, the control member and finally by the superensemble, nevertheless all forecasts present ROC areas above 0.7, a very satisfying result according to Stensrud and Yussouf (2007) who establish that forecasting systems with ROC area greater than the mentioned threshold are useful.

The Bias indicates how the forecast event frequency compares to the observed event frequency. The results for this index (Fig. 3) show that both ensembles mean, poor man and bias corrected, overpredict (Bias > 1) rainfall amounts less than 40 mm and underpredict (Bias < 1) the larger rainfalls amounts, while the control member presents the same behavior at a transitional threshold of 70 mm. On the other hand the superensemble overpredict rainfall amounts less than 80 mm and keeps steady around the perfect score (Bias = 1) for larger rainfalls amounts.
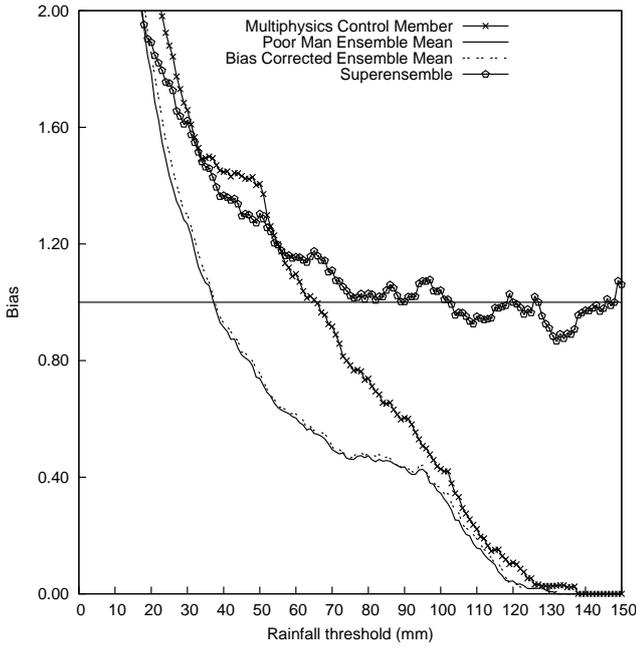
Fig. 3. Bias for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds.

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. It is worth to note that the means of the fields are subtracted, so the diagram does not provide information about overall biases, but solely characterizes the centered pattern error. The perfect score is obtained when the data point representing the forecast field matches up with the observed one. 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 diagram (Fig. 4) shows similar results for both ensemble means, poor man and bias corrected, approximately both present a RMS difference of 12 mm and a correlation coefficient of 0.5, while the standard deviations of the forecast are between 11 and 13 mm and the observed standard deviation is approximately 13 mm. The control member and the superensemble show a higher RMS difference, lower correlation and higher forecast and observation standard deviation. It is worth to mention that the statistics used on the Taylor diagram are negatively affected owing to the discontinuities, noise and outliers characteristic of the rainfall field.
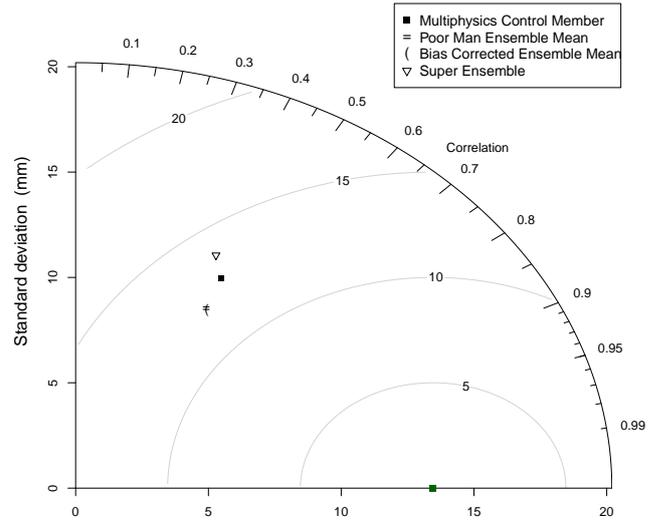

Fig. 4. Taylor diagram for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds.
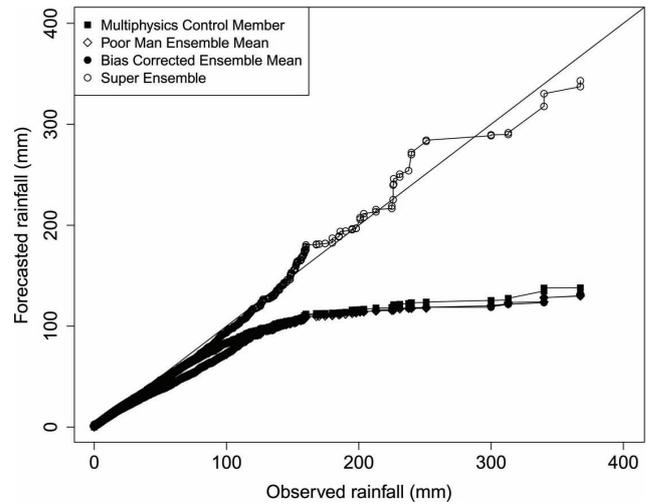

Fig. 5. Q-Q plot for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds.

The Q-Q plots compare the observed and forecasted distributions in terms of quantiles. A diagonal line indicates a perfect skill, while below the diagonal the forecast underestimates the observation and overestimates it over the diagonal (a more detailed description can be found at Wilks, 1995). The plot (Fig. 5) shows that over the 100 mm rainfall threshold all forecasts except the superensemble underpredict the observed precipitation distribution, while the superensemble captures the observed precipitation distribution (perfect score).
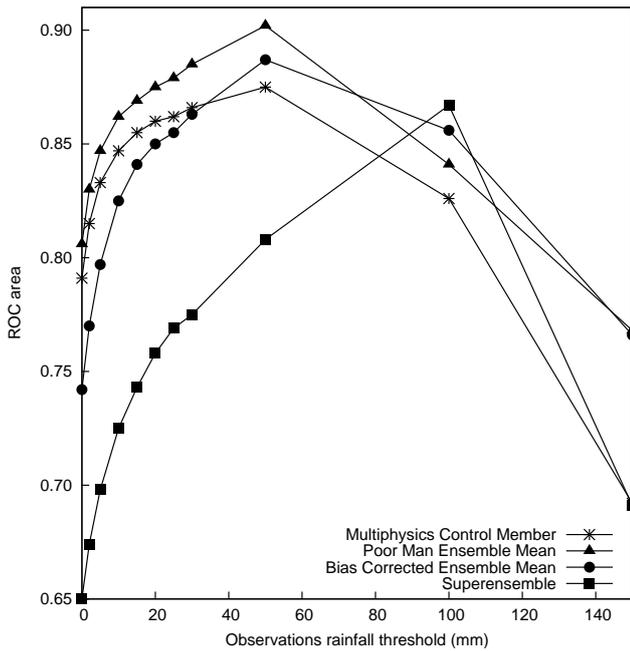
Fig. 6. ROC area for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds. Note: training period 19 MEDEX cases and forecast period September–December 2001.

Fig. 7. Bias for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds. Note: training period 19 MEDEX cases and forecast period September–December 2001.
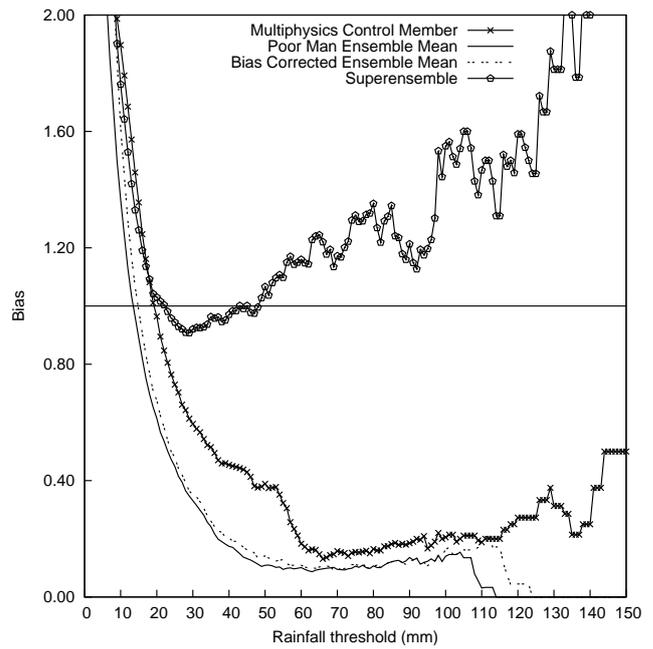
The obtained superensemble scores at the ROC area and Taylor diagrams are lower than expected. The cause of these low scores could be related to the superensemble dependency on the assumption that the performance of the members past forecasts accurately represent the performance of those members in the forecast period. Since this study deals with extreme and rare events this assumption might not be achieved. Also it is worth to notice that ECMWF forecasts, our initial and boundary conditions, have undergone severals updates during the period this study is focused on (from 1996 to 2002). Although these changes could also affect the superensemble skill, the possible effects will be neglected at this stage of the study under the assumption that the model physical parameterizations are the dominant source of variability in heavy precipitation simulations. Bearing this in mind a new test in done exchanging the training and forecast datasets in order to examine the stability of the results. In this new experiment the superensemble is trained for the MEDEX cyclones collection and tested for the 4-month period in the forecast phase.

The ROC area (Fig. 6) shows that the superensemble is tuned for the 100 m rainfall threshold being the forecast with the highest score. The Bias (Fig. 7) also shows that the superensemble is the nearest to the perfect score for a wider range of rainfall thresholds than the other forecasts, and while the others underpredict for higher thresholds
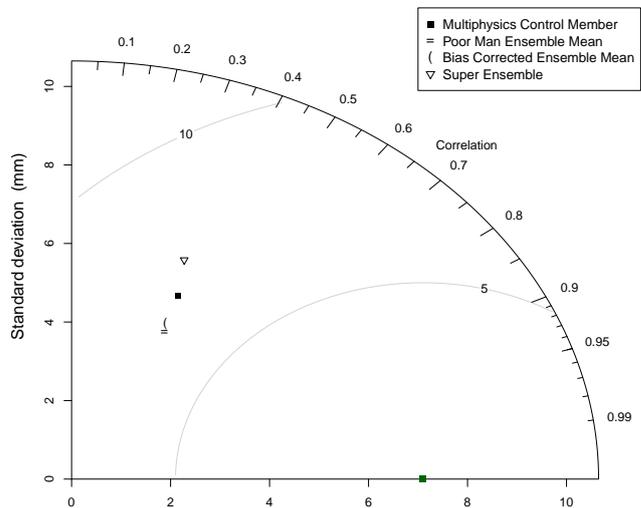

Fig. 8. Taylor diagram for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds. Note: training period 19 MEDEX cases and forecast period September–December 2001.

the superensemble slightly overpredicts them. The Taylor diagram (Fig. 8) behaves as in the previous experiment: both ensemble means (poor man and bias corrected) are the nearest to the perfect score, followed by the control forecast
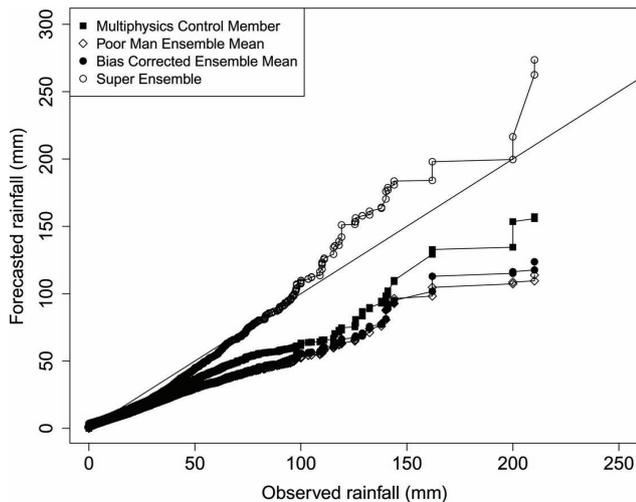
www.nat-hazards-earth-syst-sci.net/10/2371/2010/

Nat. Hazards Earth Syst. Sci., 10, 2371–2377, 2010

Fig. 9. Q-Q plot for the multiphysics ensemble control member, the multiphysics ensemble mean, the bias corrected ensemble mean and the superensemble, as functions of different rainfall event thresholds. Note: training period 19 MEDEX cases and forecast period September–December 2001.

and the superensemble. The Q-Q plot (Fig. 9) also shows that the superensemble is the nearest to the perfect score as in the previous test. These results seem to indicate that exchanging the superensemble datesets makes the superensemble more attuned to the higher precipitation thresholds.

4   Concluding remarks

The superensemble based on a multiphysics EPS instead of a multimodel ensemble and applied to rare and extreme events has not performed as expected, eventhough the superensemble has proved its value in previous studies dealing with ordinary situations. The fact that we are dealing with a multiphysics ensemble may lead to more correlation between ensemble members and therefore affect the multi-linear regression technique used to calculated the superensemble weights. Another fact that is worth to note is that the superensemble technique assumes that the past behavior of each ensemble member is representative of the present behavior, and this assumption may not be accurate for the kind of events tested in this study, cyclone-induced heavy rain events, rare and extreme by definition.

The bias corrected ensemble mean and the poor-man ensemble mean show a clear improvement over the control member, as expected. Furthermore, the superensemble is the best in the bias and Q-Q plots scores but is not good enough in the ROC area and Taylor diagrams scores. The second experiment points out that inverting the superensemble datasets attunes the superensemble better for higher rainfall thresholds. It is worth to note that the verification procedure focuses on the rainfall field, which

is highly discontinuous in space and time and observed over irregularly spaced networks and therefore difficult to be evaluated. In spite of these difficulties the verification stresses the good performance of the forecasts, specially how the superensemble captures the quantile distribution of the precipitation.

To deepen in the reasons behind the superensemble behavior found in these experiments, a new set of experiments needs to be done. The relation between past and present performance of each ensemble member has to be evaluated in order to determine the actual representativeness of the past/present forecasts. For example, the updates undergone by the ECMWF forecasts between the training period and earlier MEDEX cases were considered irrelevant. In order to evaluate this assumption and further study the relation between past and present performance, all the experiments could be redone using the ERA-INTERIM fields as initial and boundary conditions. Also, an experiment that uses a classification of Mediterranean intense cyclones derived by Garcies and Homar (2010) as training dataset, would allow the superensemble to be trained with the same event typology of the 19 MEDEX cyclones used in the forecast phase. The superensemble performance for the kind of extreme events of interest is expected to improve thanks to the information provided by these two experiments.

Edited by: G. Boni
Reviewed by: two anonymous referees

References

Cane, D. and Milelli, M.: Use of Multimodel SuperEnsemble Technique for Mountain-area weather forecast in the Olympic Area of Torino, Croat. Met. J., 40, 215–219, 2005.

Cane, D. and Milelli, M.: Weather forecast with Multimodel SuperEnsemble Techinique in a complex orography region, Meteorol. Z., 15(2), 1–8, 2006.

Cane, D. and Milelli, M.: Multimodel SuperEnsemble technique for quantitative precipitation forecasts in Piemonte region, Nat. Hazards Earth Syst. Sci., 10, 265–273, doi:10.5194/nhess-10-265-2010, 2010.

Garcies, L. and Homar, V.: An optimal ensemble sensitivity climatology of Mediterranean intense cyclones, Nat. Hazards Earth Syst. Sci., accepted, 2010.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: A Practioner's Guide in Atmospheric Science, John Wiley and Sons, 2003.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: Improved weather and seasonal climate forecasts from multimodel superensemble, Science, 285, 1548–1550, 1999.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: Multi-model superensemble forecasts for weather seasonal climate, J. Climate, 13, 4196–4216, 2000a.

Krishnamurti, T. N., Kishtawal, C. M., Shin, D. W., and Williford, C. E.: Improving tropical precipitation forecasts from multianalysis superensemble, J. Climate, 13, 4217–4227, 2000b.

Krishnamurti, T. N., Surendran, S., Shin, D. W., Correa-Torres, R., Vijaya Kumar, T. S. V., Williford, C. E., Kummerow, C., Adler, R. F., Simpson, J., Kakar, R., Olson, W., and Turk, F. J.: Real time Multianalysis/Multimodel Superensemble forecasts of precipitation using TRMM and SSM/I products, Mon. Weather Rev., 129, 2861–2883, 2001.

Krishnamurti, T. N., Gnanaseelan, C., Mishra, A. K., and Chakraborty, A.: Improved Forecasts of the Diurnal Cycle in the Tropics Using Multiple Global Models. Part I: Precipitation, J. Climate, 21(16), 4029–4043, 2008.

Stensrud, D. and Yussouf, N.: Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system, Weather Forecast., 22, 3–17, 2007.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.

Vich, M., Romero, R., and Brooks, H.: Ensemble prediction of Mediterranean high-impact events using potential vorticity perturbations. Part I: Comparison against the multiphysics approach, Atmos. Res., submited, 2010.

Wilks, D.: Statistical Methods in the Atmospheric Sciences: An Introduction, Academic Press, 1995.

Yun, W. T., Stefanova, L., Mitra, A. K., Kumar, T. S. V. V., Dewar, W., and Krishnamurti, T. N.: A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts, Tellus A, 57, 280–289, doi:10.1111/j.1600–0870.2005.00 131.x, 2005.

www.nat-hazards-earth-syst-sci.net/10/2371/2010/

Nat. Hazards Earth Syst. Sci., 10, 2371–2377, 2010