APPLICATION OF CLUSTER ANALYSIS TECHNIQUES TO THE VERIFICATION OF QUANTITATIVE PRECIPITATION FORECASTS

Stefano Serafin, Alessio Bertò, Dino Zardi

Department of Civil and Environmental Engineering, University of Trento
Via Mesiano 77, I-38050 Trento, Italy
E-mail: *stefano.serafin@ing.unitn.it*

Abstract: The results of the verification of precipitation forecasts are highly affected by the distribution of rain gauges, and depend mostly on the model performance in areas where the gauges network is denser. For model verification purposes, in the present work the whole set of available measurements is divided into a series of geographical subsets, each of them displaying a similar precipitation pattern. Cluster analysis is adopted as an objective method to create groups of rain gauges displaying interrelated measurements. Clusters display a well defined spatial structure, related to the interaction of air masses with orographic obstacles. Verification scores computed in each cluster help in detecting the spatial variation of a model's performance.

***Keywords** – precipitation forecast, verification, cluster analysis*

1 INTRODUCTION

Quantitative precipitation forecasts by numerical weather prediction models are usually verified by means of pointwise comparisons between local observations by rain gauges and interpolated forecast fields. The agreement between forecasts and observations is then evaluated by means of a contingency table. Finally, standard verification scores such as equitable threat score and bias are computed (Wilks, 1995). As a consequence, the distribution of rain gauges in the simulation domain plays a major role in verification, whose results depend mostly on the model performance in the areas where the rain gauge network is denser. Nevertheless, standard verification procedures do not take into account the spatial variability of the observed and forecast precipitation fields. The key points addressed here are: (1) do verification scores vary significantly within a simulation domain? (2) if so, are the variations related to geographical, meteorological, or other features?

The database collected during the MAP campaign in 1999 is particularly suited to discuss such issues, having very high resolution both in space and in time. Several events of the campaign (Intensive Observation Periods) have been selected, among those which displayed relevant precipitation amounts over Northern Italy: IOP2a, IOP2b, IOP3, IOP8, IOP15. In the present work, for each event the set of available observed data is split into a series of subsets, each of them having a similar precipitation pattern. Cluster analysis (Jain et al., 1999) is adopted as an objective method to create groups of rain gauges with interrelated measurements. The data used to perform clustering are precipitation observations and geographical information about rain gauges. Verification scores are computed in each of the subsets detected by clustering.

2 CASE STUDIES AND MODEL SIMULATIONS

Most of the events under consideration were characterized by similar synoptic conditions, with an upper level trough and a surface level low approaching northern Italy from the West. The flow associated with this configuration is in general southerly over this region. IOP2a is a convective rainfall event over the Lago Maggiore area. Heavy rainfall was widespread over Italy and also on the northern slope of the Alps during IOP2b, while it insisted only over Piedmont in IOP3. Rainfall was moderate and mainly concentrated in the Po Valley during IOP8. Finally, during IOP15 a deep cut-off low insisted over the Mediterranean sea, drawing north-westerly currents over Italy. Two models, namely PSU/NCAR MM5 (non-hydrostatic, Grell et al., 1994) and ISAC-CNR BOLAM (hydrostatic, Buzzi et al., 1994), have been used to simulate these events. The area of interest is approximately located between 43–49°N and 4–16°E, in order to include Northern Italy as a whole. The models were run with similar resolutions: a single domain with 12.5 km grid spacing for BOLAM and 2 two-way nested domains for MM5, with the inner one having 6 km resolution. All of the simulations were initialized with ECMWF analyses and were 48 hours long (startup times: IOP2a, 1200UTC 17 Sep; IOP2b, 1200UTC 19 Sep; IOP3, 0000UTC 25 Sep; IOP8, 1200UTC 20 Oct; IOP15, 1200UTC 5 Nov).

3 CLUSTERING TECHNIQUE

In order to identify the most significant precipitation patterns in the wide number of available observation points (about 1500), suitable criteria to gather them around representative modes have been borrowed from clustering analysis techniques. Several hierarchical agglomerative clustering techniques have been tested (e.g. Brankov et al., 1998) in various meteorological applications. The average-linkage clustering algorithm (Bertò et al., 2004) has been selected due to its good performances.

A key issue in this application of cluster analysis is the use of the appropriate input dataset to relate different precipitation regimes to specific geographical areas. Thus, a multidimensional phase space has been adopted where dimensions include normalized space coordinates (latitude, longitude, and height asl.) and 6-hours normalized accumulated precipitation (ie. 8 time lags spanning 48 hours), resulting in 11 degrees of freedom. Geographical variables have been weighted with small coefficients (0.7 for latitude and longitude, 0.2 for height) to emphasize the influence of the precipitation factor in the clustering algorithm. In order to minimize the variance between data within the same cluster and to maximize the variance between different clusters, the Root Mean Square Deviation (RMSD) has been analyzed for various possible number of clusters, in search of sudden breaks. RMSD is defined as the root of the average of the cluster variances. Cluster variance is in turn defined as the averaged square distance between each point representing an observation in the phase space and the center of mass of the cluster. Sudden breaks in RMSD identify the merging of significantly different precipitation patterns and suggest the optimal number of clusters to retain at the end of the agglomerative procedure.

4 VERIFICATION INDICES

Verification scores are defined in terms of the elements of a contingency table, in which all the forecast-observation pairs are distributed according to their relationship pertaining a threshold (Wilks, 1995). The bias score, ranging from zero to infinity, allows to assess the overestimation or underestimation of precipitation above a certain threshold, although it bears no information about the correspondence between forecast and observations. In general, bias greater than 1 indicates precipitation overforecasting, while bias less than 1 shows underforecasting. On the other hand, the ETS roughly estimates the percentage of correct forecasts that can be ascribed to the model skill (i.e. the percentage of non-random correct forecasts), with values ranging from slightly negative (worse than random forecast) to 1 (perfect forecast). Another relevant verification measure is the root mean square error (RMSE) of the forecast field; this index has been normalized with the average observed precipitation in each event.

Verification scores are supposed to provide significant pieces of information only when computed on a collection of many events, but in this particular context some useful indications can also be obtained from their small-scale analysis, because of the large availability of spatially and temporally dense rainfall measurements. Bias, ETS and RMSE have been computed in each event for the whole data set and for every single cluster. Scores are evaluated for cumulated precipitation over 6 hours; this unusually short time interval makes verification particularly demanding. The spatial distribution of verification indices has been reconstructed by using an extremely low threshold in contingency tables (0.1 mm). This choice aims at evaluating the models' ability in forecasting the occurrence or lack of rainfall at given space and time coordinates. Representative values of verification scores in each event and cluster have been estimated by selecting suitable thresholds (ie. the 33rd and 66th percentile of the frequency distribution of precipitation amounts in each data subset). Verification scores are indeed very sensitive to the choice of the threshold, and thus also scores vs. threshold charts need to be carefully evaluated.

5 RESULTS

5.1 Global verification scores

Table 1 lists the global verification scores for MM5 (M) and BOLAM (B) in all IOPs. The RMSE is generally lower for B. The precipitation fields forecast by B are more homogenous than those by M: as a result, deviations from observations are on average less relevant, although the ETS of the two models is comparable. M turns out to be drier than B in IOPs 2a, 2b and 3, and wetter in IOPs 8 and 15 (bias score). IOPs 2b and 8 are better predicted by both models, which both behave worse in IOPs 15 and 2a. The more it rains, the better models perform (they have trouble in forecasting low or sparse rainfall).

Table 1. RMSE, ETS and bias for M and B in five MAP events. ETS and bias are computed at the 66th percentile threshold (see Section 4).

	iop2a	**iop2b**	**iop3**	**iop8**	**iop15**
RMSE_M	7,39	2,68	6,27	2,94	4,74
RMSE_B	5,77	2,72	6,21	2,9	4,51
ETS_M	0,11	0,29	0,16	0,29	0,17
ETS_B	0,08	0,3	0,12	0,33	0,23
bias_M	1,24	0,9	1,03	1,68	2,22
bias_B	0,97	0,93	1,37	1,16	1,5

5.2 Scores distribution

The distribution of ETS and bias is not spatially homogenous (not shown). For instance, overestimation of rainfall by both B and M is often apparent on the northern slope of the Alps, whereas M produces dry forecasts in the Po Valley during IOPs 2a, 2b and 3. Typical distributions of verification scores show that models can produce misleading forecasts in orographically complex areas, while the forecast of purely stratiform rain is usually much better. Additionally, models seem to achieve better results in forecasting precipitation over the western part of the domain: one reason for this may be that in most events mesoscale systems move west to east. The model performance decreasing as runtime progresses affects the latest stage of the simulation, when precipitation is expected over the eastern part of the domain.

5.3 Cluster distribution

The clusters detected by the average-linkage algorithm are well separated when plotted in geographical coordinates, although rainfall variables account for more than 70% of the total information used. Moreover, clusters are persistent in all of the events, i.e. they have about the same position and shape (see Figure 1).

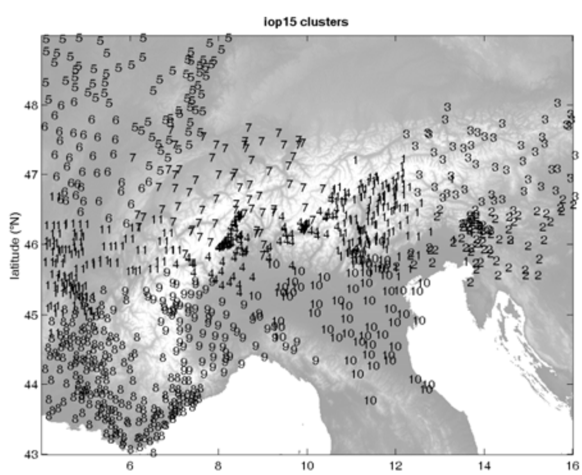

Figure 1. The clusters detected by the average-linkage algorithm from the data of IOP15.

Rainfall patterns are indeed similar in the different IOPs, but other factors seem to be particularly relevant. For instance, the Alpine ridge is also a separation between adjacent clusters. This is related to the physics of precipitation, the southern slope of the Alps being upstream and the northern downstream. Also, somewhere clusters seem to mirror the political borders between different countries or regions. This feature raises the suspect of systematical biases in the various rain gauge networks.

5.4 Clustered verification scores

Table 2 contains a sample of the clustered verification scores for IOP15. Scores change significantly from cluster to cluster. Among the three factors of variability that determine the values of verification indices, the most effective is the spatial variation between clusters. Less relevant factors are the event considered and, last, the model used (M or B). Models produce very similar forecasts (they were initialized with the same IC's and BC's), but variations in their performance are more relevant between different areas within the same event, than between different events.

Table 2. Clustered verification scores for IOP15. Boldface: M; italics: B. ETS and bias were computed at the 66th percentile threshold (see Section 4).

Cluster	1	2	3	4	5	6	7	8	9	10	11
RMSE	**2,2** *2,6*	**1,2** *1,5*	**1,8** *1,5*	**1,5** *1,8*	**1,8** *1,5*	**5,3** *3,2*	**2,9** *2,6*	**2,1** *2,1*	**1,5** *1,5*	**3,9** *3,9*	**3,1** *2,9*
ETS	**0,38** *0,24*	**0,46** *0,28*	**0,44** *0,42*	**0,48** *0,44*	**0,38** *0,50*	**0,17** *0,22*	**0,25** *0,25*	**0,21** *0,35*	**0,20** *0,23*	**0,15** *0,16*	**0,26** *0,26*
Bias	**0,96** *0,32*	**1,08** *0,47*	**1,37** *0,89*	**1,31** *0,49*	**1,42** *1,29*	**3,62** *3,02*	**0,70** *0,41*	**1,13** *1,10*	**0,94** *0,57*	**1,90** *0,67*	**1,41** *1,61*

6 CONCLUSIONS

Cluster analysis techniques have been applied to the aim of detecting any significant spatial variation in the skill of numerical weather prediction models. The subsets detected by the clustering algorithm were remarkably persistent in the events in exam, and forecast quality appears to depend on the spatial variability represented by clusters. Such variability seems to be related both to shortcomings in the models' physics and to biases in the observational data.

REFERENCES

Bertò, A., A. Buzzi and D. Zardi, 2004: Back-tracking Water Vapour Contributing to a Precipitation Event over Trentino: a case study. *Meteorol. Z.*, **13**, 189-200.

Buzzi, A., M. Fantini, P. Malguzzi, and F. Nerozzi, 1994: Validation of a limited area model in cases of Mediterranean cyclogenesis: surface fields and precipitation scores. *Meteor. Atmos. Phys.*, **53**, 137-154.

Dorling, S.R., T.D. Davies and C.E. Pierce, 1992: Cluster analysis: A Technique for Estimating The Synoptic Meteorological Controls on Air and Precipitation Chemistry-Method and Applications. *Atmos. Environ.* **26A**, 2575-2581.

Grell, G.A., J. Dudhia and D.R. Stauffer, 1994: A description of the Fifth-Generation Penn State/NCAR Mesoscale Model MM5. NCAR/TN-389+IA.

Jain, A.K., N.M. Murty, and P.J. Flynn, 1999: Data Clustering: A Review. *ACM Computing Surveys*, **31**, 264-323.

Wilks, D.S., 1996: Statistical Methods in the Atmospheric Sciences. Academic Press.