COSMO model validation using the Italian radar mosaic and the rain gauges estimated precipitation

N. Vela[1,2], R. Cremonini[1], E. Oberto[1], M. Giorcelli[1,3]

[1] *ARPA Piemonte, via Pio VII 9, 10135, Torino, Italy*

[2] *University of Torino, via Giuria 1, Torino, Italy*

[3] *Arianet S.r.l, via Gilino 9, 20218 Milano, Italy*

1 Introduction

The aim of the present work is to realise a new type of verification for the COSMO-I7 model. The verification is made against a precipitation field estimated by the Italian radar mosaic corrected with the data coming from the Italian rain gauges network. In order to perform the modification an ordinary kriging process of the differences between the radar data and the rain gauges measurement is needed. Once the modified precipitation field is obtained, firstly the relative error is calculated, then a fuzzy multi-scale verification is performed. The whole work is based upon a case study from the 24th to the 27th of October 2010. A second but not less important purpose of the work is to apply this method (already established in literature) to our territory, in order to analyse and elaborate data witch might be applicable to any future model verification. After completing this study, a second phase of the work will begin using the COSMO-I2 model.

2 Observed data: the rain gauges and the Italian radar mosaic

The rain gauges are unevenly distributed through the Italian territory with the exception of the Puglia and Sicily regions. The data delivered within the COSMO Project are used together with those observed by the rain gauges belonging to the different Regional Centres and made available through the Italian Civil Protection Department. The radar data come from 24 operative machines: 10 are installed and managed by Regions, 4 are owned by the Air Force, 2 are owned by ENAV (Air Traffic Control Agency) and 8 are installed by Civil Protection Departement (6 emitting in C band, 2 in X band). In Figure 1 the distribution of the rain gauges and the radar mosaic is shown.

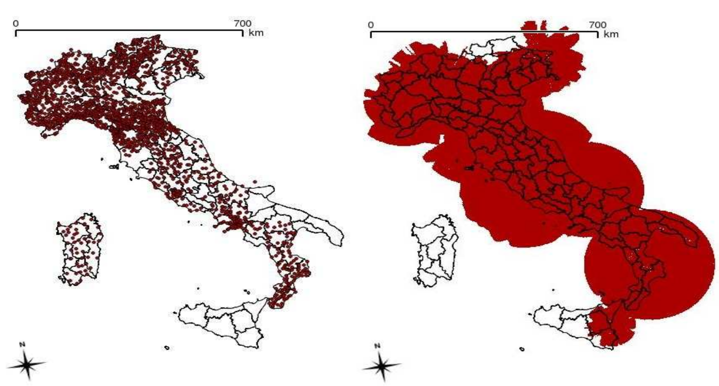

Figure 1: The Italian rain gauge network (left) and the Italian radar mosaic area (right).

3 Reconstruction of the precipitation field

3.1 Calculation of the difference (between radar and rain gauges)

The correspondence between the estimated 24 hour cumulative precipitation (radar) and the rain gauges measurements is calculated. The area associated to each rain gauge includes 9 radar grid points (witch has a 1040.9 m resolution). The median value among the 9 radar grid points is coupled with the rain gauge one. The difference (or deviation) between the two data is calculated as follows:

$$Difference = 10 * \log_{10}\left(\frac{r}{rg}\right) \tag{1}$$

where r is the precipitation estimated by the radar and rg is the rain gauge measurement.

3.2 Ordinary kriging

The ordinary kriging technique has been used to modify the estimated precipitation field recorded by the radars through the rain gauges network data. The "autoKrige" function of the R software has been used for this purpose. This function produces many outputs among witch two have been used: the kriging prediction and the kriging standard error.

4 Validation

4.1 Relative error

The relative error is calculated as follows:

$$Erel = \left[\frac{(F - O)}{O}\right] * 100 \tag{2}$$

where $Erel$ is the relative error, F is the forecast precipitation amount and O is the observed one (coming from the correction of the radar estimation). The relative error is calculated for the 24 hours cumulative precipitation (mm/24h). Concerning the model, the first day of forecast is used (00UTC run). The relative error is evaluated for the areas where the kriging standard error does not exceed the value of 4 dB.

4.2 Fuzzy verification

This kind of verification answers the question: "*Witch is the link between spatial forecast and a combination of the intensity of the precipitation and the scale of the event ?*" The scale decomposition methods allow us to diagnose the model errors and performances according to different scales. The scale-intensity approach links the traditional bi-dimensional verification categories: it returns the model skills according to different precipitation intensities and spatial scales. It verifies the model over the whole domain. It is useful in the spatial verification of discontinuous fields (like the precipitation). It supplies information both for single case studies and forecast systems evaluated over a longer time. Using a neighborhood verification method an exact correspondence between forecast and observation is not needed.

Re-sampling of the domain The domain is divided in boxes with a side of 10 km: each box contains the mean value of the precipitation records found in it (or a value which marks it as non valid if less than 75% of the included values are valid). Two different files are then created: one for the forecast data (COSMO-I7), the other one for the observations (radar corrected with the rain gauges).

Fraction Skill Score Answers the question: *What are the spatial scales at which the forecast resembles the observations ?* The Fraction Skill Score (FSS) directly links the portions of the grid which are covered by the forecast and by the observation (for example the rain exceeding a certain threshold) through spatial windows of increasing size. The FSS is calculated as follows:

$$FSS = 1 - \frac{\frac{1}{N}\sum_{N}(P_f - P_o)^2}{\frac{1}{N}\left[\sum_{N} P_f^2 + \sum_{N} P_o^2\right]} \tag{3}$$

where P_f is the portion of the box covered by the forecast, P_o is the portion of the box covered by the observation and N is the number of spatial boxes covering the entire domain. The Fractions Skill Score ranges from 0 (complete mismatch) to 1 (perfect match).

The value of FSS above which the forecasts are considered to have useful (better than random) skill is given by $FSS_{useful} = 0.5 + f_o/2$, where f_o is the domain average observed fraction. The smallest window size for which $FSS \geq FSS_{useful}$ can be considered the "skillful scale". As the size of the squares used to compute the fractions gets larger, the score will asymptote to a value that depends on the ratio between the forecast and observed frequencies of the event. The closer the asymptotic value to 1, the smaller the forecast bias. The score is most sensitive for rare events (small rain areas for example).

Equitable threat score (Gilbert skill score) Answers the question: *"How well did the forecast "yes" events correspond to the observed "yes" events (accounting for hits due to chance) ?"* It measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance (for example, it is easier to correctly forecast rain occurrence in a wet climate than in a dry climate). The ETS is often used in the verification of rainfall in NWP models because its "equitability" allows scores to be compared more fairly across different regimes. It is sensitive to hits. Since it penalises both misses and false alarms in the same way, it does not distinguish the source of forecast error. It is calculated as:

$$ETS = \frac{hits - hits_{random}}{hits + misses + falsealarms - hits_{random}} \tag{4}$$

where

$$hits_{random} = \frac{(hits + misses)\,(hits + falsealarms)}{total} \tag{5}$$

5 Case study: 2010/10/24-25-26-27

This case study has been chosen because of the preponderance of advective precipitation over the whole event, against a short convective phase at the beginning. The precipitation is well spread over the entire Italian territory on the 25^{th} and 26^{th} of October, while it is more concentrated over northern Italy on the 24^{th} and over the south on the 27^{th} of October.

5.1 Preliminary operations

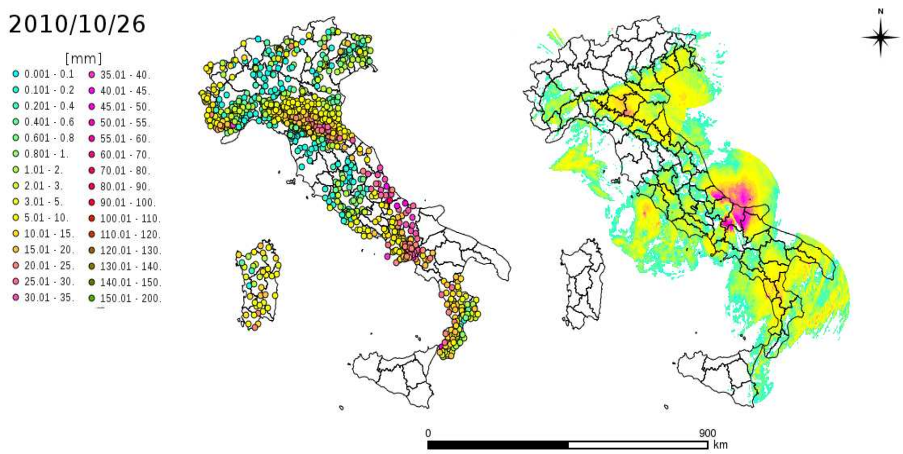

Figure 2: Rain gauges measurement (left) and pluviometric data from the Italian radar mosaic (right). (mm/24h, 26th of October)

In Figure 2 the rain gauge observations and the rain estimated by the radar are shown for the 26th of October (mm/24h). It is possible to notice that the the agreement between the two data is very good for what concerns the spatial dislocation, a little less good if we look at the intensity of the precipitation. The main differences are located over the Alps and the Apennines, and where the rain is very weak or very intense. The overall mean difference between the rain gauge measurements and the radar data falls between -5 and -2.5 dB: the radar seems to underestimate the precipitation. Figure 3 shows the spatial distribution of the difference between the two data expressed in dB (in red are shown the points where the radar underestimates, in blue those where it overestimates).

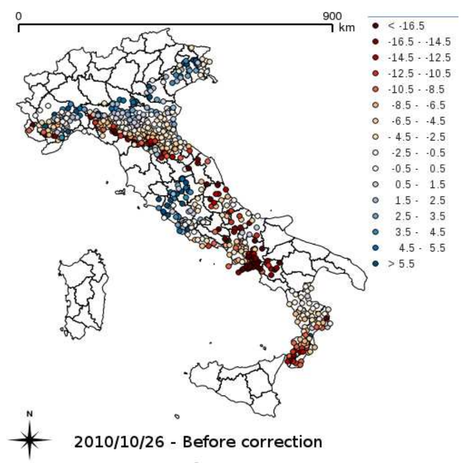

Figure 3: Difference (dB) between the pluviometric radar data and rain gauge measurements (spatial distribution). Blue: the radar data is higher than the associated rain gauge measurement. Red: the radar data is lower than the associated rain gauge measurement. In this figure it is possible to notice that the rain gauge measurements are higher than the pluviometric radar data over the more elevated points. (26th of October)

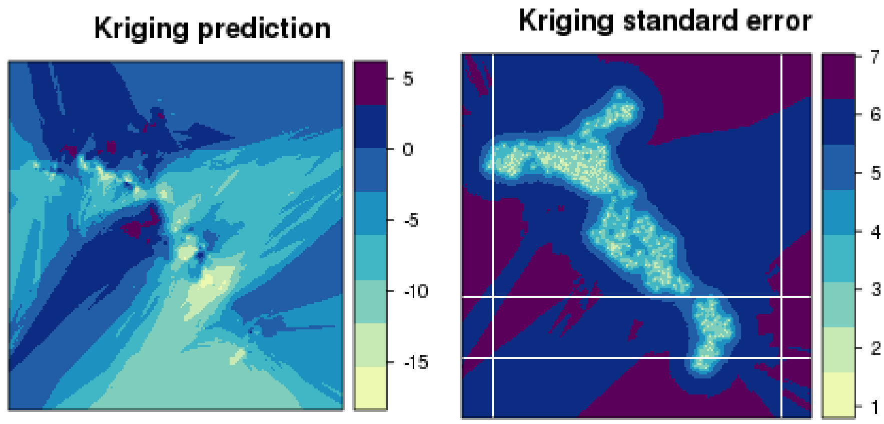

Figure 4: Two outputs among those coming from the autoKrige functionality in R. Left: kriging prediction for the difference between radar and the rain gauges network. Right: the associated kriging standard error. All the data are expressed in dB. (26th of October)

5.2 Ordinary kriging

The difference (expressed in dB) between the pluviometric radar data and the rain gauge measurements is then used to correct the radar itself. The ordinary kriging is used for this operation. In Figure 4 it is possible to see the output given by the autoKrige function of the R software (the kriging prediction on the left, the kriging standard error on the right, this is the result for the 26^{th} of October). It is possible to notice that the standard error is small where the rain gauge network is thicker, while it gets bigger where there are fewer of them (beyond Italy borders, over the sea, in Puglia and Sicily regions). The ordinary kriging procedure increases the precipitation recorded by the radar over most of the grid points (Figure 5).

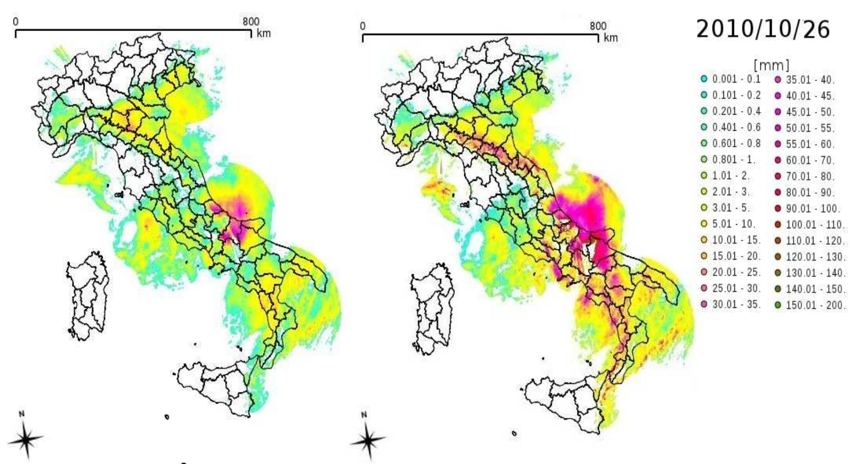

Figure 5: Pluviometric radar field (left) and pluviometric radar field after the rain gauge correction (right). Data expressed in mm. (26^{th} of October)

6 Verification

6.1 Preliminary analysis

The first step in the verification is an eyeball comparison (Figure 6, given as an example) between the modified radar field and the COSMO-I7 forecast. It is possible to notice a good agreement between the two fields. The agreement is good for what concerns the dislocation of the precipitation patterns, a little worst if we look at the intensity of the precipitation.

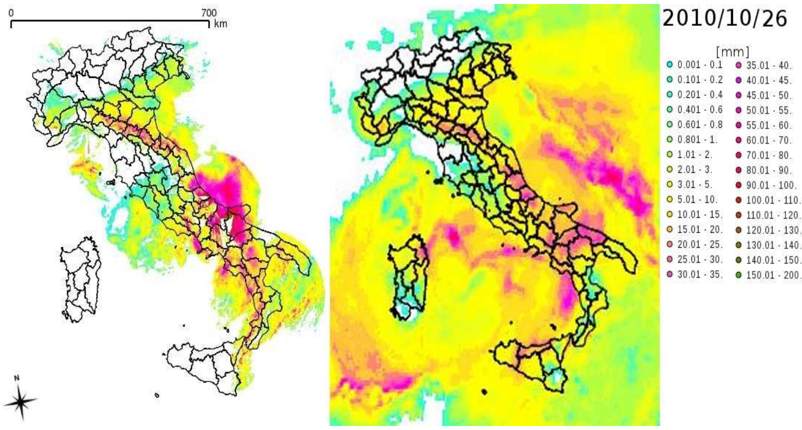

Figure 6: Pluviometric radar field corrected with the rain gauge measurements (left) and COSMO-I7 forecast cumulated (24h) precipitation (right). (26^{th} of October)

6.2 Relative error calculation

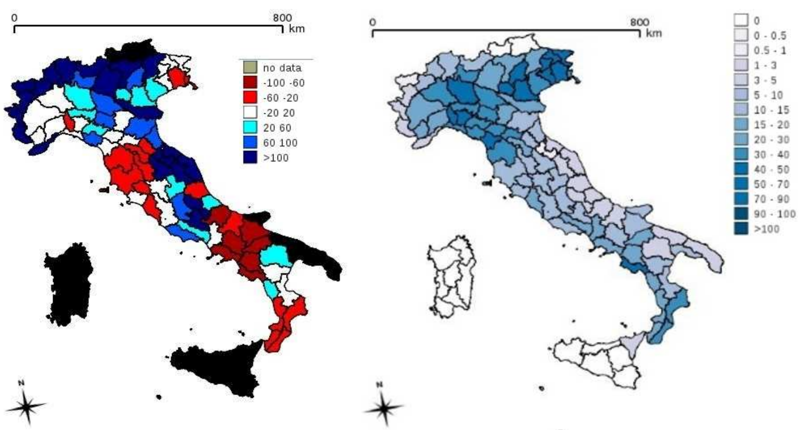

Figure 7: Relative error (left) and cumulative precipitation (mean over alert areas) (right). (25^{th} of October)

In Figure 7 the calculation of the relative error between the forecast and the observed precipitation for the 25^{th} of October is reported. The relative error is calculated for each of the 102 alert areas in which the Italian territory is subdivided. The ones coloured in black are those where the kriging standard error exceeds the value of 4 dB (no rain gauges or no rain). The red ones are those where the forecast underestimates the precipitation, the blue ones are those where COSMO-I7 overestimates it. It is possible to notice a general overestimation of the model over northern Italy, more marked in the alpine region. For what concerns the peninsula, the model underestimates almost everywhere, with the exception of the Marche and part of the Lazio regions where there is overestimation.

6.3 FSS calculation

As written before, the two fields (forecast and observed) must be brought to a common grid to calculate the FSS. This common grid is made of boxes with a side of 10 km. From these two new grids the files for each exceeding threshold are written (0.2, 1, 2, 5, 10, 20, 40, 50, 75 mm) (Figure 8, example: 25^{th} of October, 20 mm). The portions coloured in blue are those where the precipitation exceeds the threshold, the red ones are those where the threshold is not reached, while the white ones are those where the precipitation data are not valid. The

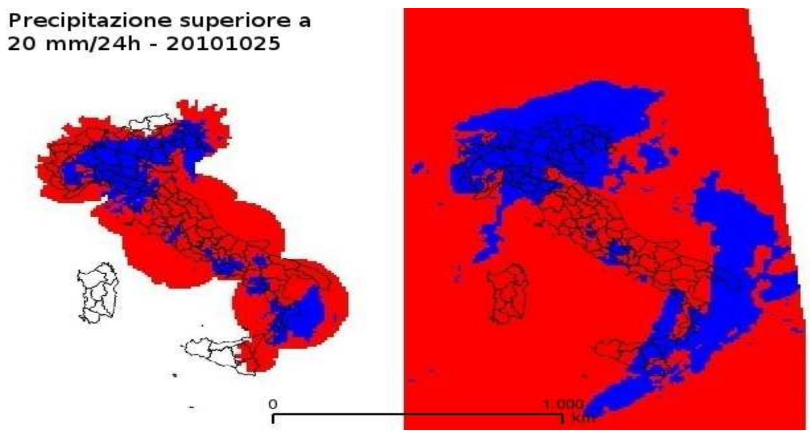

Figure 8: Observed (corrected radar field) (left) and forecast (right) precipitation exceeding the 20mm/24h threshold for the 25^{th} of October. White: no data. Red: precipitation not exceeding the threshold. Blue: precipitation exceeding the threshold.

FSS results for the four days are presented in Figure 9. The black line surrounds the values witch are higher than the FSS_{useful} (different for each threshold). Higher values of FSS can be found for large areas and very low thresholds (upper left corner in each panel). The best FSS values are those of the 25^{th} of October, where the precipitation is more extensive. It is in such a case that it is important to look at the FSS_{useful}: the value of FSS above which the forecasts are considered to have useful skill (better than random).

6.4 ETS calculation

Figure 10 shows the results for the ETS calculation over the four days of the event. Also in this case there is a better forecast for wide areas (with the exception of the 26^{th} of October for the medium-high thresholds). The graphs of the ETS are not monotone: the function contains a factor witch takes into account the frequency of the forecast event (i.e. precipitation exceeding higher thresholds are less common and for this reason more difficult to forecast). A relative maximum in the results of the ETS means a better performance of the model for those thresholds. As the FSS, the ETS shows a very good performance of the model for wider areas at low thresholds. ETS shows a good performance also for the mean thresholds.

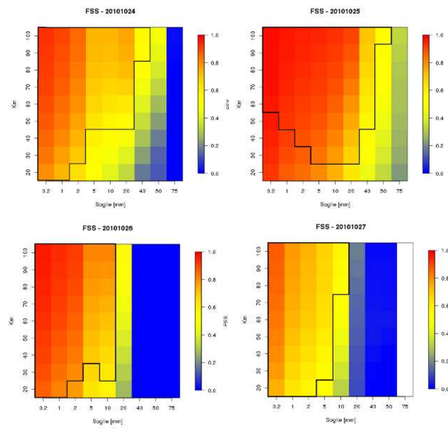

Figure 9: The graphs show the value of the FSS and the FSS_{useful} (black line) for each of the four days of the event. X-axis: thresholds (mm/24h). Y-axis: box side (km).

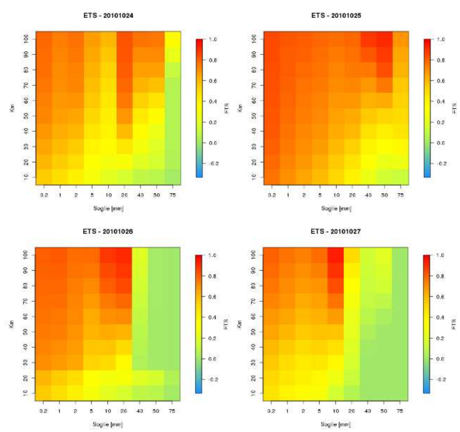

Figure 10: The graphs show the value of the ETS for each of the four days of the event. X-axis: thresholds (mm/24h). Y-axis: box side (km).

7 Conclusions

7.1 Ordinary kriging

It is not possible to integrate the radar field with the data coming from the rain gauges by simply applying a bias to the first one. On the one hand the rain estimated by the radar is affected by errors coming from the characteristics of the precipitation, the orography and the geometry of the beam itself. On the other hand, the rain gauges show some problems when displaced at higher altitude and do not supply a regular field. For the above reasons we decided to use the R functionality "autoKrige" to perform an ordinary kriging of the differences between the radar precipitation field and the rain gauge network measurements, and then to use the latter to correct the first.

7.2 Relative error

The field resulting from the correction of the pluviometric radar data has then been used as an observation to calculate the relative error of the COSMO-I7 precipitation forecast (cumulative over the 24 hours, first day forecast, 00UTC run). The evaluation of the relative error has been done by dividing the Italian territory into the 102 alert areas used by the Civil Protection Department. The mean forecast and observed precipitation has been calculated for each area (with the exception of those where the kriging standard error was too high). The results are concordant with those coming from a more classic verification (rain gauges only): COSMO-I7 tends to overestimate the precipitation over the Alpine area and underestimates (or overestimates less) over the plains.

7.3 Fuzzy verification

Fuzzy verification methods are called scientific and diagnostic and they analyse the nature of the error itself. Both the FSS and the ETS show how the COSMO-I7 model has always very good skills in forecasting the precipitation for low thresholds over wide areas. The ETS also show good skills for the middle thresholds (also on large areas). The quality of the forecast reduces if we look at higher thresholds: this might be because they are more spatially localized. This kind of verification lets us know what are the conditions in witch COSMO-I7 can be trusted for different types of forecasts (from the local to the large scale ones).

7.4 Future developments

This work, although it refers to a single case study, shows some potentiality and some promising result. The idea is to extend the approach to other cases using the model COSMO-I2 which probably is more suitable for this kind of analysis due to its higher horizontal resolution.

References

[1] Cherubini, T., Ghelli, A. and Lalaurette, F., 2002: Verification of precipitation forecasts over the Alpine region using a high-density observing network, *Wea. Forecasting*, 17, **238-249**.

[2] Ciach, G. J., W. F. Krajewski, E. N. Anagnostou, M. L. Baeck, J. A. Smith, J. R. McCollum, A. Kruger, 1997: Radar Rainfall Estimation for Ground Validation Studies of the Tropical Rainfall Measuring Mission., *J. Appl. Meteor.*, 36, **735747**.

[3] Cremonini R., R. Bechini, V. Campana, D. Rabuffetti, and I. Giudici, 2004: An implementation of Kalman filter in north-western Italy for radar calibration using 3-hourly gauges measurements and polarimetric c-band radar estimates, *Proceedings of ERAD* (2004): **352355**.

[4] Ebert, E.E. and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors, *J. Hydrology*, 239, **179-202**.

[5] Giuli D., E. Palmisano, G. Scarchilli, J. Tan, M. Thurai, 1996: Rain assessment through integration of radar and rain-gauge data in severe storms, *Remote Sensing Reviews*, Vol. 14, Iss. 1-3.

[6] E. Goudenhoofdt, L. Delobbe, 2009: Evaluation of radar-gauge merging methods for quantitative precipitation estimates, *Hydrol. Earth Syst. Sci.*, 13, **195-203**.

[7] Holleman, I., 2006: Bias adjustment of radar-based 3-hour precipitation accumulations KNMI publication, TR-290, 11/10/2006, pp**56**.

[8] Kaufmann, P., 2007: Post-processing of the aLMo Precipitation with the Neighborhood Method, *COSMO Newsletter* No. 7.

[9] Krajewski, W. F., 1987: Co-kriging radar-rainfall and rain gage data, *J. Geophys. Res.*, 92(D8), **95719580**.

[10] Jolliffe, I.T., and Stephenson, D.B., 2003: Forecast Verification: A Practitioner's Guide in atmospheric Science, *John Wiley and Sons*.

[11] Joss, J., R. Lee, 1995: The application of Radar-Gauge comparisons to operational precipitation profile corrections. *J. Appl. Meteorol.* 34, **2612-2630**.

[12] Oberto E., Milelli M. 2010: QPF verification for 2008/2009 of several COSMO-Model version, *COSMO NEWSLETTER* No.10.

[13] Pappenberger, F., Ghelli, A., Buizza, R., and Bodis, K., 2009: The skill of probabilistic forecasts under observational uncertainties within the generalized likehood uncertainty estimation framework for hydrological applications, *J. Hydrometerol.*, 10, **807819**.

[14] Sinclair, S. and Pegram, G. 2005: Combining radar and rain gauge rainfall estimates using conditional merging, *Atmos. Sci. Lett.*, 6, **1922**.

[15] Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. An Introduction, *Academic Press*, San Diego, xvii+**627** pp.

[16] Zawadzki, I., 1984: Factors affecting the precision of radar measurements of rain. In Proceedings of Preprints of the 22nd Conference on Radar Meteorology; *American Meteorological Society*: Boston, MA, USA.