SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Short-Range Ensemble Prediction System
Computer Project Account:	spitlekf
Start Year - End Year :	2015 - 2017
Principal Investigator(s)	Lucio Torrisi
Affiliation/Address:	COMET - Italian Air Force Operational Center for Meteorology Via di Pratica di Mare, 45 00071 Pomezia (Roma) ITALY
Other Researchers (Name/Affiliation):	Francesca Marcucci (COMET)

The following should cover the entire project duration.

Summary of project objectives

The goal of this project is to improve the existing short-range ensemble prediction system, based on the Ensemble Kalman Filter (EnKF) approach (COMET-LETKF [1] [2]) for the data assimilation component (estimation of the initial conditions) and the COSMO local model (www.cosmo-model.org) for the prognostic one, in the framework of a comprehensive plan for development of a new set of tools for the probabilistic forecast.

Summary of problems encountered

No real problem was encountered, neither technical nor conceptual.

Experience with the Special Project framework

We are familiar with the Special Project framework and we appreciated this way to proceed.

Summary of results

The atmospheric short-range ensemble prediction system (COSMO-ME EPS) based on the COMET-LETKF analysis [1] [2] and the COSMO model is running operationally at the Italian Air Force Meteorological Service (pre-operational since July 2013, operational since May 2014).

The relevant characteristics of the atmospheric COSMO-ME EPS are:

- Domain and resolution: COSMO model is integrated 40 times on COSMO-ME domain with a 7km (10km up to September 2017) grid spacing and 45 vertical levels
- IC and BC: initial conditions are derived from the COMET-LETKF system; lateral boundaries conditions are from the most recent IFS deterministic run perturbed using ECMWF-EPS.
- Model error: stochastics physics perturbation tendencies (SPPT).
- Forecast range: 72 hours at 00/12 UTC.

During the first year of the project, the almost two year (2013-2014) of forecast data have been analyzed, in order to evaluate the performances of the ensemble and the possibility of further improvements.
First of all a few statistical scores have been developed and applied in order to evaluate the usefulness of the COSMO-ME EPS forecast system in term of sharpness, bias and skill.
Rank histogram, spread-skill relationship and continuous rank probability scores have been chosen as best candidates for a first investigation. Scores are computed, with respect to synoptic observation and IFS analysis, for temperature, wind and relative humidity over pressure levels and for pressure, wind, temperature and precipitation at the surface.
The spread skill relationship has been computed following the method proposed in Whang and Bishop [3]. The results show that the ensemble has on average a good spread-skill behavior with respect to IFS analysis, but is slightly under-dispersive when compared to observations for some forecast steps. In Fig.1 the spread-skill results for temperature and wind are shown respectively at 500 hPa and 700 hPa. The under-dispersion of the COSMO-ME EPS after second day of forecast is also confirmed by rank histogram results.

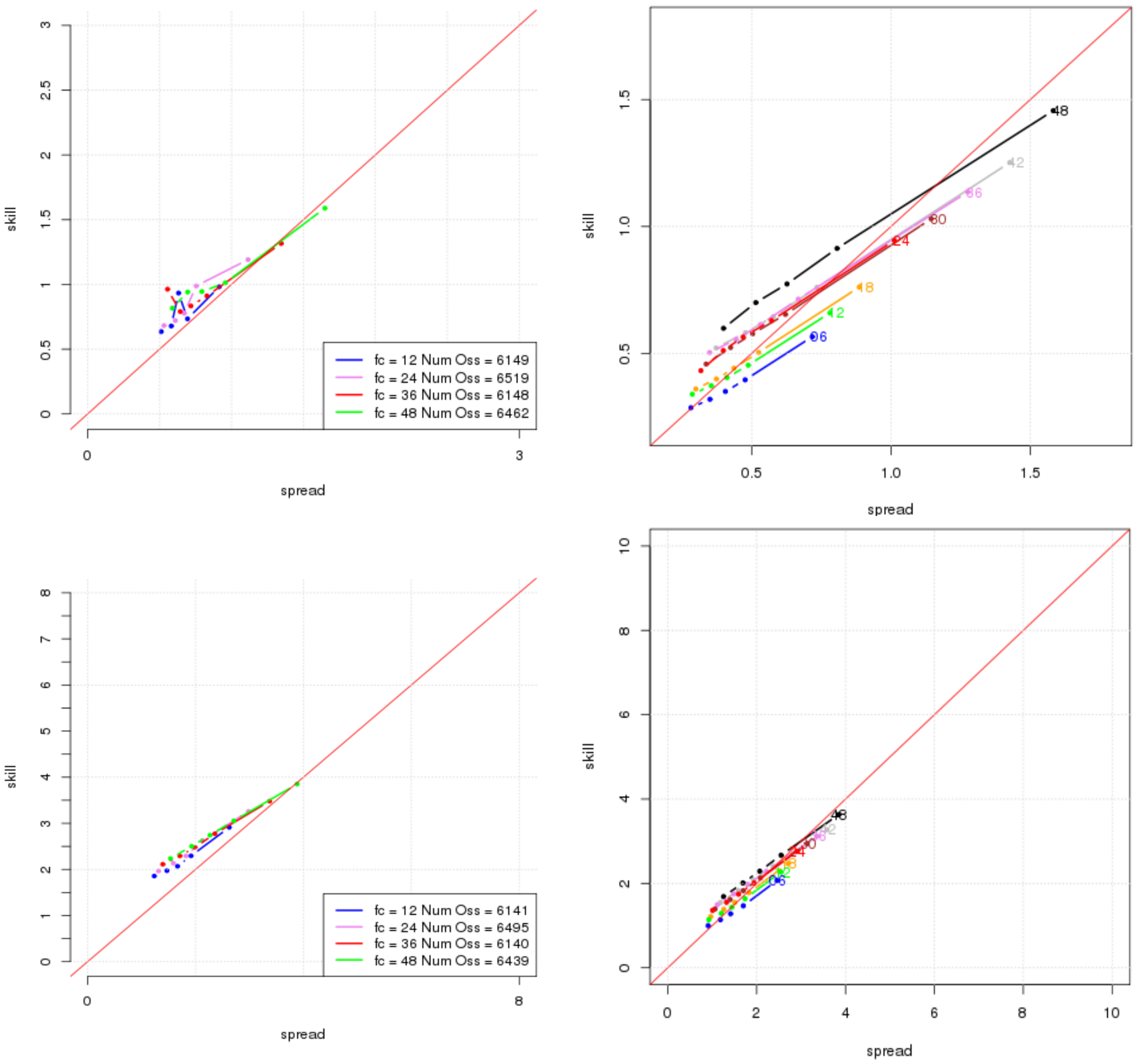

Fig.1 Left panels show temperature at 500 hPa (top) and u-wind component at 700 hPa (down) spread-skill binned relationship for all the forecast steps (in different colors) with respect to synop observation. Right panel shows same results with respect to IFS analysis.

During the second year of the project, one year (2014) forecast data have been analyzed, in order to evaluate COSMO-ME EPS precipitation forecast performances, using the precipitation observations from the high resolution Italian raingauges network. In particular, the usefulness of the COSMO-ME EPS forecast system in term of sharpness, bias and skill has been investigated using same scores as for the other continuous variables.
Looking at the results, it can be outlined that the rank histogram and spread-skill relationship confirm the under-dispersion of the COSMO-ME EPS for all forecast steps and the CRPS follows the behaviour of the spread-error as expected.

In addition resolution, reliability and accuracy of the ensemble, for different precipitation thresholds, have been evaluated by the implementation and computation of new scores: brier score and brier skill score (fig 2).

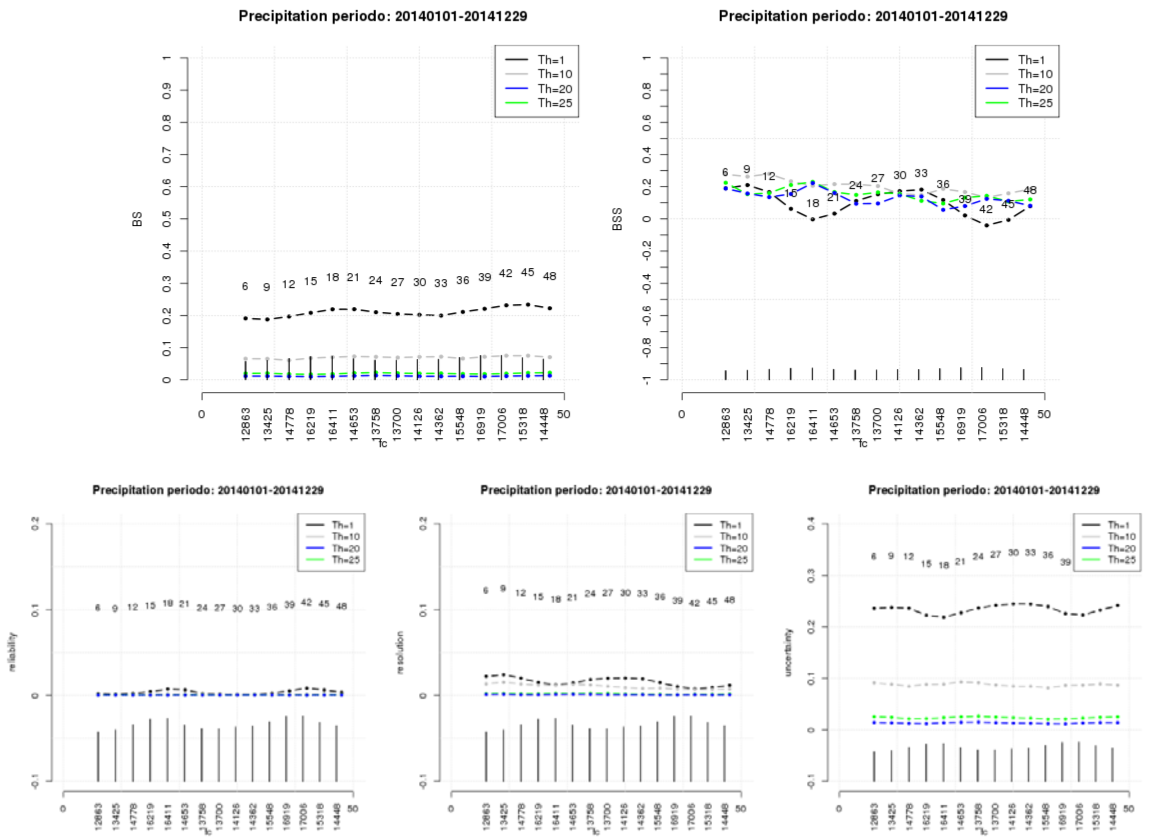

Fig.2 (top) Brie Score (left) and Brier Skill Score (right) for different thresholds and forecast steps. (bottom) From left to right reliability, resolution e uncertainty components of the Brier Score for different thresholds and forecast steps.

During the third year of the project, different calibration methods have been evaluated, in order to correct the bias of the existing ensemble and the derived uncertainty information of the precipitation field. The aim of the work has been to provide to our forecasters the best possible support for specific operational applications.

Because the small COSMO-ME EPS historical dataset, only two methods (quantile-to-quantile mapping QQM [4] and Local quantile-to-quantile transformation LQQT [5]) have been chosen as a starting point. Both methods need a smaller training period with respect to other algorithms proposed in literature. QQM and LQQT methods are used to calibrate the QPF (quantitative precipitation forecast) of each ensemble member, so that correspondent rain field maintains their spatial features.

In this project the two methods have been implemented using R language and tested over a 1 month period.

The LQQT method seems to correct the ensemble precipitation forecasts less than the QQM approach and, by contrast, it seems to correct more for heavy precipitation.

A first objective comparison of the two methods have been done using the continuous rank probability score (CRPS) computed over a longer period (00UTC run for november 2014) as a function of lead time. The results (fig.3) show that lower values of the CRPS are obtained with the LQQT method. An extension of the investigation period should be done, in order to have a more statistically robust result.

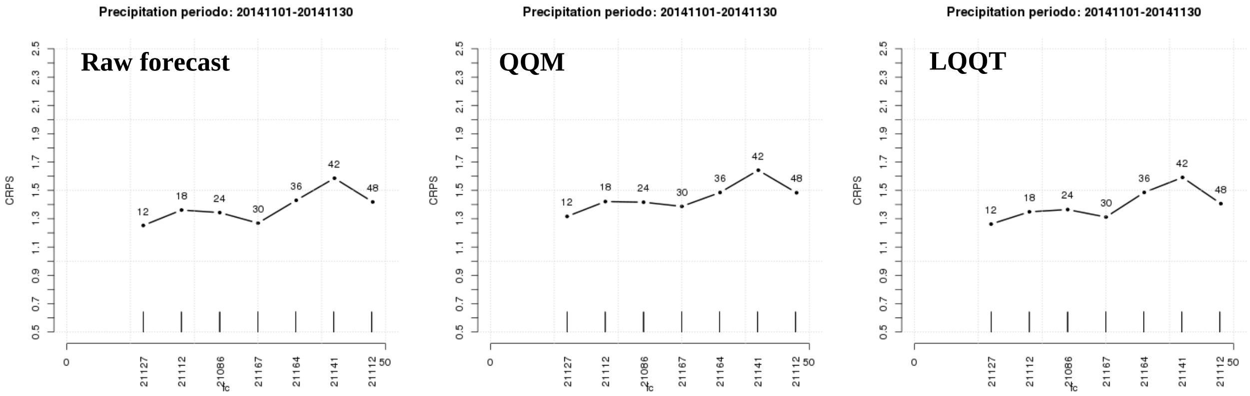

Fig 3. Comparison of CRPS score (00UTC run for november 2014) as a function of lead time of 6 hour cumulated precipitation forecast for raw ensemble (left), QQM calibrated ensemble (centre) and LQQT calibrated ensemble (right).

During last year a reliability-based calibration scheme [6] has been implemented and tested. This method was designed to target directly the reliability. The result should be reliable by construction, since each calibrated probability is based on a past observed event frequency.
Reliability calibration aim is to produce a reliable PDF conditioned on the forecast. The sum of all such conditional PDFs is the climatology, so reliability calibration can, in principle, simultaneously correct probabilities, spread, and climatology.

The scheme has been trained over one year of data cyclically preceding the target forecast. The training accumulates sample counts, forecast probabilities and observed frequencies for each grid-point, lead time, threshold, and forecast probability bin. The main implementation uses ten thresholds spaced in powers of two from 0.1 to 51.2 mm. Only five probability bins are used: three across the main probability range and one each for cases where zero or all members exceed the threshold.

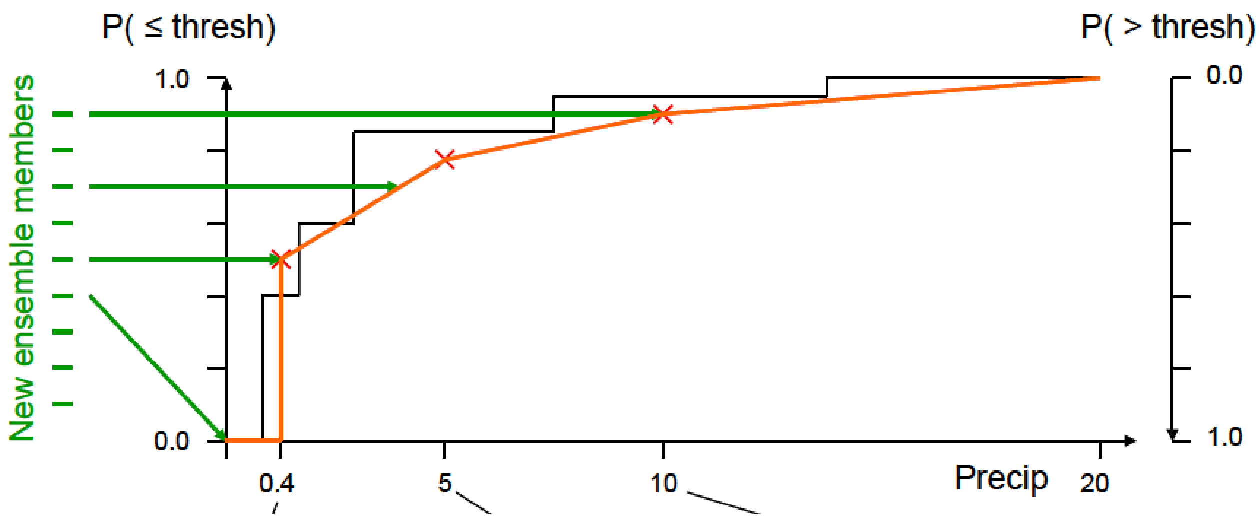

Fig.4: Illustration of the reliability-based calibration method

As shown in fig.4, the raw ensemble members imply a Cumulative Density Function (CDF). The training (reliability diagrams) implies a calibrated CDF. New members are assigned to equally divide the probability range, in the same order as the raw ensemble members.
Fig. 5 shows the maps of the difference between calibrated forecast probability and raw forecast probability related to events (single test case) with 6 h cumulated precipitation rain greater than 0.2 mm and 10 mm. For each grid point, the observed frequencies and the forecast probabilities, averaged on the probability bins, have been calculated starting from annual time series (training year 2014). The maps show in particular that there is almost no correction for the largest threshold

due to low frequency of events with rainfall above 10 mm. A comparison between seasonal and annual time series has been also performed, results are shown in fig. 6.
From a theoretical point of view, the reliability method seems to be the most advantageous, but it needs a continuous update of the calibrated CDF with the last available observations.
It needs therefore further investigation over a long period of test cases, in order to evaluate if the computational effort is worth it.

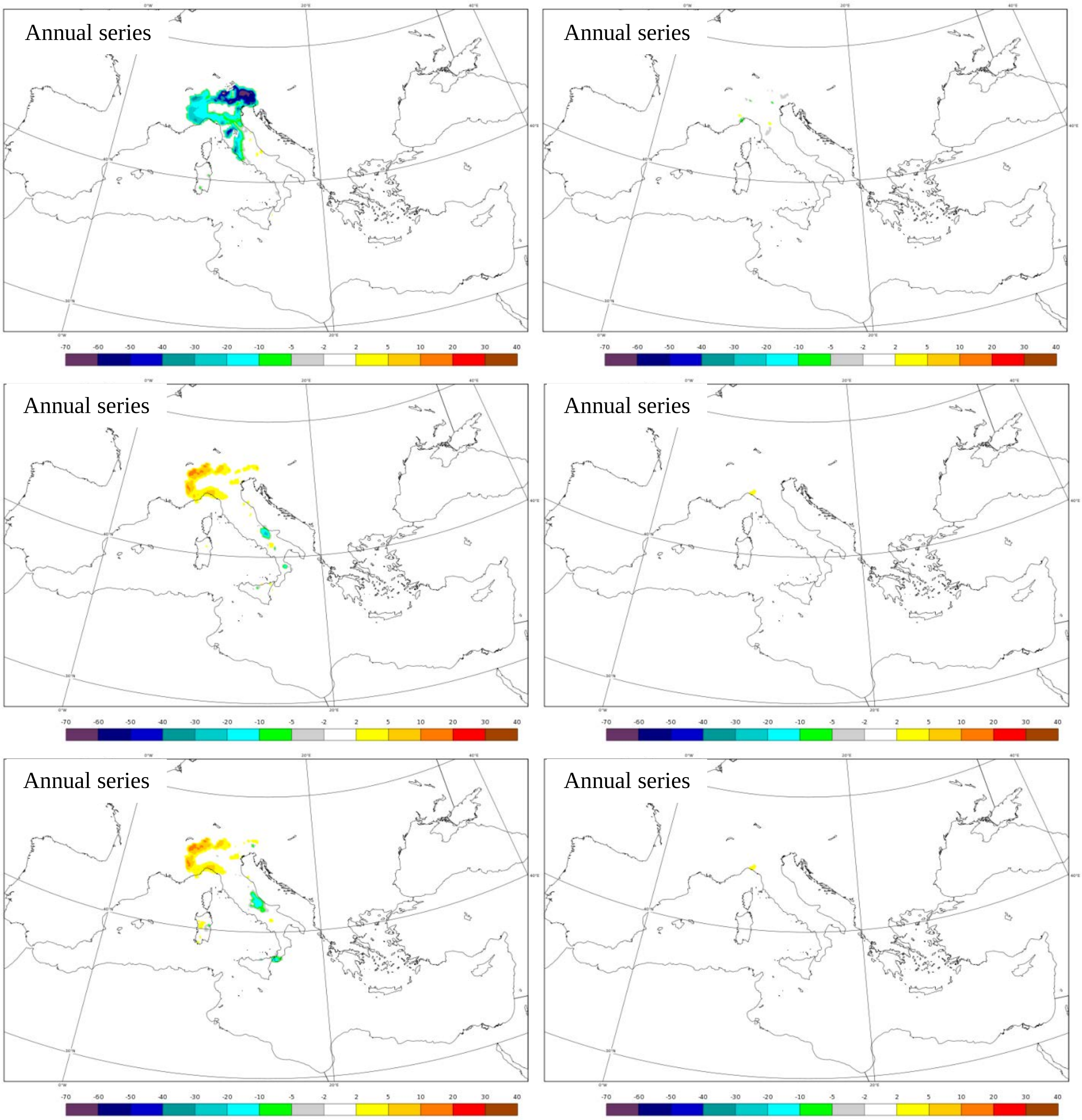

Figure 5 – Single test case. Maps of the difference between calibrated forecast probability and raw forecast probability related to rain events greater than 0.2 mm and rain greater than 10 mm (on the left and on the right respectively). For each grid point the observation frequencies and the forecast probabilities averaged on the probability bins have been calculated starting from the annual time series (2014).

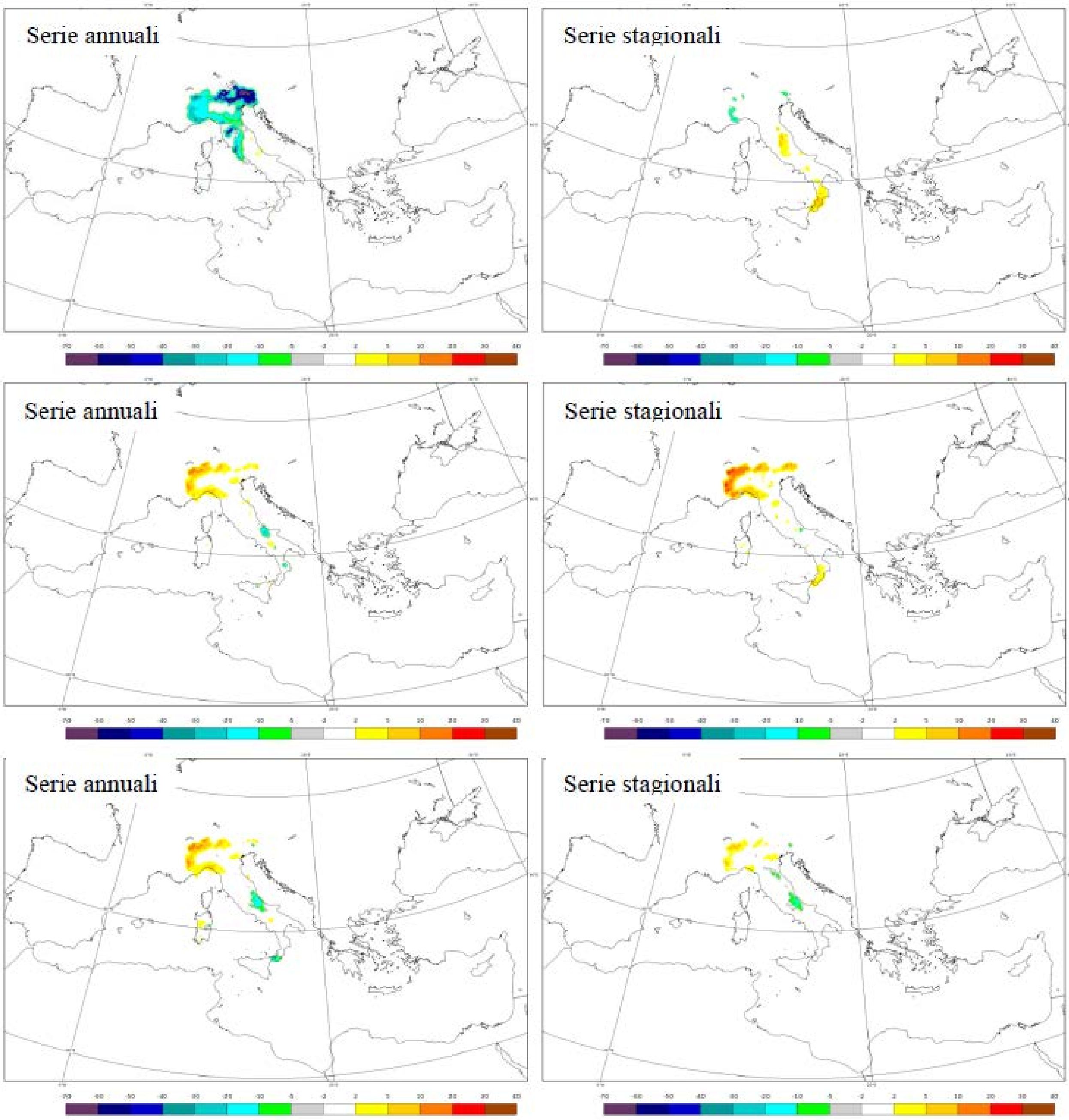

Figure 6- Single test case. Maps of the difference between calibrated forecast probability and raw forecast probability related to the rain event greater than 0.2 mm. For each grid point the observation frequencies and the forecast probabilities averaged on the probability bins have been calculated starting from annual time series (with training period equal to the year 2014, on the left) and seasonal time series (with training period equal to the season in which the target date falls, on the right).

List of publications/reports from the project with complete references

[1] Bonavita M, Torrisi L, Marcucci F. 2008. The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. *Q. J. R. Meteorol. Soc.* 134: 1733-1744.

[2] Bonavita M, Torrisi L, Marcucci F. 2010. Ensemble data assimilation with the CNMCA regional forecasting system. *Q. J. R. Meteorol. Soc.* 136: 132-145.

[3] Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. J. Atmos. Sci., 60, 1140-1158.

[4] B. Thrasher, E. P. Maurer, C. McKellar, and P. B. Duffy: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrol. Earth Syst. Sci., 16, 3309–3314, doi:10.5194/hess-16-3309-2012, 2012.

[5] Bremnes, J. B.: Improved calibration of precipitation forecasts using ensemble techniques. met.no report 04/02007, 2007.

Future plans

We planned to stop the work on tuning and calibration of a regional ensemble prediction system at 7 km resolution. This is mainly, because we decided to invest a much more effort on the convective permitting scales, also in view of the next future increase of resolution of the ECMWF forecasting system (deterministic and probabilistic). We planned to use computational resources for the improvement of a convection-permitting Ensemble Prediction System over Italy and for the operational production of a new set of tools for probabilistic forecast to support aviation. This will be the main goal of a new special project that will be proposed for the period 2019-2021.