EMS Annual Meeting Abstracts Vol. 15, EMS2018-816, 2018 © Author(s) 2018. CC Attribution 4.0 License.



Quantifying observation uncertainty on verification measures – A MesoVICT example

Simon Kloiber and Manfred Dorninger

University of Vienna, Department of Meteorology and Geophysics, Wien, Austria (manfred.dorninger@univie.ac.at)

Observation errors are usually neglected when it comes to verification arguing that they are small compared to forecast errors. By increasing the quality of the forecasts in the last decades this assumption has to be revisited if it still holds. In this MesoVICT (Mesoscale Verification Inter-comparison over Complex Terrain) study we will show that observation errors may influence verification results considerable.

The observation error is quantified by the sophisticated quality control scheme of VERA (Vienna Enhanced Resolution Analysis). By adding the errors in a random procedure to the original observations and re-run of the analysis an analysis ensemble is produced. Several tests have been executed (Kolmogorov-Smirnov-Test, Finkelstein-Schafer Test, Chi-Square Test etc.) to investigate its distribution. This analysis ensemble is used to verify deterministic as well as ensemble forecasts. For the whole study short term ensemble forecasts (up to 48 hours) from COSMO-LEPS provided by Arpae-SIMC Emilia-Romagna and COSMO-E provided by MeteoSwiss are used. The time period covers the MesoVICT core case from 20-22 June 2007.

The spread of the verification results caused by the analysis error is investigated for a set of verification measures for single grid points as well as the whole area of investigation (larger Alpine area). This set covers Bias, RMSE, MAE and ETS for deterministic forecasts as well as Rank Histograms, CRPS, Reliability diagrams and ROC curves for ensemble forecasts.

Results will be presented for 10m-wind speed. They indicate that observation/analysis errors can remarkable influence verification results. The spread resulting from the analysis ensemble varies strongly in the 3-hourly time series and can reach up to 50% of the original verification score.