# PP- C2I: Task 6.5 - Verification

*Flora Gofa*

## Introduction

The purpose of the document is to provide verification guidelines that can be followed by the partners for priority project C2I purposes. The overall goal of the PP-C2I is to ensure a smooth transition from the COSMO model to ICON-LAM. At the end of the PP C2I, each participating institution is free to choose when ICON-LAM replaces the COSMO model in their operational forecasting system and a major role in this will play the relative performance between the two systems.

## Tactic

COSMO has implemented for similar purposes an environment to perform carefully-controlled and rigorous testing, including the calculation of verification statistics, for any new COSMO model version. This testing phase offers the necessary information on the model forecasting performance, in order to determine whether the upgrade of a model test-version to a new release version is possible. Similarly for C2I project, a testing platform approach needs to be followed that will provide for each service a way to monitor and evaluate the forecasts of the COSMO and ICON modelling systems and determine in an objective way the relative advantage of each one.

According to the recently produced document by the WMO Joint Working Group for Forecast Verification Research, process-oriented or modelling oriented verification has its overall objective of performance monitoring, identifying the role of various model processes, understanding modelling errors. Within such framework, process-oriented verification contributes to a more efficient research-to-operations and operations-to-research cycle in NWP, both within and between NWP centres. In the context of verification, comparing forecasts from different modelling systems is an attempt to relate differences in error characteristics between forecasting systems to differences in their design. It can provide additional insights when extended further than just a ranking exercise, to include in-depth analyses of score differences and statistics. Finally, for physic parameterization related modelling systems such as this of COSMO and ICON, such verification can also contribute to the understanding of systematic errors, especially for surface parameters. In order to draw conclusions however about processes, a deep knowledge about the different parameterizations used in the models is required.

Previous efforts of NWP systems intercomparison experiments, provided us with the knowledge:

- to make more use of test case or idealized experiments to isolate processes,
- to put effort to link model output to observations with the creation of observation based datasets whose structure and metadata mirror that of the model-based datasets,
- to coordinate efforts to apply common evaluation packages and methods.

## Proposed Verification Software

COSMO consortium since 2006 has adopted the approach of Common Verification Software that would enable the production of homogeneous and comparable statistical results in order to perform activities such as the monitoring of the performance of the operation COSMO versions for the extraction of systematic errors and this approach is preferable also for C2I verification Task.

**VERSUS:** is a software was developed with this intention and is readily available to be used for point verification approaches. ICON offers the possibility to provide output on a regular longitude-latitude grid as COSMO model, hence it can be relatively easy to adapt to ICON results. It has to be pointed out that ICON does not offer the possibility of GRIB1 output as the available output formats are NetCDF and GRIB2. However, a conversion from GRIB2 to GRIB1 is supported by the official COSMO post-processing software Fieldextra, even though VERSUS has a developed module to ingest GRIB2 outputs.

**MEC/Rfdbk:** The Model Equivalent Calculator (MEC) software for the production of Feedback Files, and verification scripts based on the R package Rfdbk, are tools that were developed and are currently used operationally at DWD for the operational verification of both COSMO and ICON model chains. PP-CARMA is a project that is currently running in COSMO consortium with a goal to replace VERSUS verification software environment as a CVS in order to perform part of the verification activities in the consortium (http://www.cosmo-model.org/content/tasks/priority Projects/carma/default.htm . The software is now available to all partners for installation and use, while support will be provided through the project to accomplish these goals. Following the project timeline, the software will be partially used in all services until the end of 2019 so it can facilitate verification objectives of C2I on a later stage and definitely as a tool to accomplish all point surface and upper air verification needs of ICON model in the future.

**VAST/DIST/SpatialVx:** As for very high resolution (convection permitting) modeling systems, is necessary to apply spatial verification approaches, such tools are also recommended. VAST is COSMO software that is based on Beth Ebert fuzzy verification IDL code. It includes the following methods: Upscaling, anywhere in window, minimum coverage, fuzzy logic, joint probability and multi-event contingency table, brier skill score, fraction skill score and practically perfect hindcast belong to the verification methods employed in this package. VAST main code utilizes txt gridded files for each weather parameter, but also a preprocessing of input files is available with the help of LIBSIM software. As these tools are based on grib1 format as input, a preprocessing of ICON files needs to be performed beforehand.
DIST is verification code that is used to apply the DIST (Distribution) method that was developed by ARPA-SIMC. It is a spatial verification method (a form of upscaling) based on catchment area or boxes of fixed size comparing different indices (max, mean, median, percentiles). DIST is linked with LIBSIM preprocessing of observation and forecast file library.
SpatialVx (International Community) is an R package for performing spatial forecast verification. Most of the state-of-the-art verification methods are included. This R package comprises several libraries that one has to apply both for input file adaptations and verification purposes.

## Proposed evaluation approach

**Long(er)-term verification: Monthly/Seasonal basis**
WG5 is already providing verification guidelines that can be followed by the partners for model intercomparison based on station based data. This basic approach can be easily implemented by all partners as it is already part of the work that is produced for Common Verification Plots in COSMO. The only additional work will be to adjust ICON model output to verification software

requirements. Guidelines can be found at: http://www.cosmo-model.org/view/repository/wg5/commonPlots/reports/.

**Conditional Verification: Monthly/Seasonal basis**

The stratification of verification results according to a control variable is a method that can facilitate the deeper study of the performance of COSMO and ICON models. This variable can be the verified quantity itself (e.g. cloud cover forecast error as a function of observed and/or forecast cloud cover). In the more general case, it is one or several variables which quantify the effect of specific processes on the verified quantity, allowing in this way a distinction between different 'regimes'. A typical example would be the evaluation of night-time 2m temperature biases as a function of cloud cover and wind speed.

Conditional verification can be used to quantify relationships between errors in different variables and can help to identify their sources. WG5 in the past has proposed several conditions that could provide indicative results linked to various model processes. Some of them that can be applied in COSMO-ICON forecast evaluation are:

| | |
|---|---|
| **2m Temperature** | |
| 1st condition:<br>2nd condition: | Total cloud cover >= 75% (overcast condition)<br>a. THICK using TQC (Total column cloud water)<br>b. THIN using TQC - Reference value TQC<5 g/m2 |
| 1st condition:<br>2nd condition: | Total cloud cover <= 25% (clear sky condition)<br>a. THICK using TQC - Reference value TQC>5 g/m2<br>b. THIN using TQC - Reference value TQC<5 g/m2 |
| 1st condition: | 2m Temp for various thresholds<br>2mT with wind in selected stations<br>2mT with snow cover<br>**2mT/Td with soil moisture** |
| 1st condition:<br>2nd condition: | **Total cloud cover <= 25% (overcast condition)**<br>Wind speed <= 2,5 m/s |
| **Precipitation** | |
| 1st condition: | Convective precipitation (unstable atmosphere)<br>Reference value of CAPE 50 J/Kg<br>Precipitation for various weather classes<br>Check pressure tendency availability |
| 1st condition: | Large scale precipitation (LSP)<br>using non convective CAPE values |
| **Cloud cover with stability index** | |
| **Wind Speed** | |
| **WS with roughness length** | |
| **Wind gust** | |
| 1st condition: | Convective (unstable atmosphere)<br>Wind gust for convective precipitation cases |
| 1st condition: | non convective atmosphere,<br>using non convective CAPE |

**Test case based verification- Spatial methods**

As it is mentioned before, it is recommended for modeling system intercomparison to work an a test case base, even idealized ones, in order to deeper analyze the differences of the systems with respect to various physical processes. WG4 has already gathered a number of cases with various characteristics in the weather phenomena that were reported during them, so this can be a good starting point. On a national level, special cases of interest can be recommended after collaboration with the forecasters. The recently completed PP-INSPECT, provided to the consortum an insight of the pros and cons of various spatial verification approaches that one can get advise from (http://www.cosmo-model.org/content/model/documentation/techReports/docs /techReport37.pdf). For this task of comparing the two models (ICON/COSMO), the most easy to imprement and to interpret methods are suggested to be used.

Neighborhood methods provide a tool to compare modeling systems of various resolutions and are particularly valuable in the case of high resolution forecasts. However, before deciding on the methodology or score that is more suitable for a model evaluation, the first step in this approach involves carefully defining the attributes of a good forecast and subsequently identifying the specific methods and their associated decision models best suited to the particular application. Neighbourhood verification is particularly valuable in the case of high resolution forecasts, providing useful feedback on the scale and intensity for which each model configuration is advantageous. Precipitation events on different spatial scales are produced by different physical processes (e.g. large-scale frontal systems or small-scale convective events). Verification at different spatial scales provides greater insight into model performance to simulate these different processes.

The fraction skill score (Roberts N.M., 2008) is one of most commonly used index. The basic idea of this score is to define an area or scale and to count events in this region (fraction area) both in the model and in the observation. One then obtains a score for different scales and a so called “useful scale”, is also defined above a certain score. It is suggested to reduce the available information by showing the most relevant part for standard verification on a one dimension plot which resembles usual station based verification or in a table. For the considered score and variable it is proposed to graphically represent cross sections of the data. One can select a single meaningful scale and to show for selected threshold the score as a function of leadtime. This meaningful scale can roughly correspond to the size of a warning region. Another proposed way of displaying the data is to focus on the useful scale information and graphically represent them, as a function of lead time for different thresholds.