Science Plan status

COSMO GM Plenary session, 8-11 Sept 2014, Eretria, Greece

SP Outline: Chapter on Diagnostics and Validation

Proposed Actions in Validation and Diagnostics

- ***Tackling model performance improvement issues through the use of conditional verification (CV)***

- ***Statistical methods to identify the skill (added benefit) of convection permitting and near-convection-resolving model configurations***

- ***Development of tools for probabilistic and ensemble forecast verification***

- ***Exploitation of available observational dataset for operational and scientific purposes***

- ***Severe and High Impact Weather***

- ***User-oriented Verification products***

4.1 Science Plan: current situation
(from SMC videoconference 20/08)

- At the moment:

  → the revised Science Plan and answers to the reviewers' comments were prepared and distributed, also to the STC members, thank you!

  → I expected comments from Christoph Schaer, but there were none

  → currently, the comments from STC members are expected

  → already comments from MeteoSwiss are available; they call for <u>substantial additional discussion on the reviewers' comments and revision of the plan</u>

4.1 Science Plan: current state

- (continuation):

  → the STC discussion and feedback is expected in Eretria

  → I would propose that further work on Science Plan will take place after the STC feedback is known

  → I would ask you to still consider and discuss the Science Plan issues during parallel sessions in Eretria

SP chap9 – Reviewer Comments

<u>Jeanette Onvlee:</u>

Clear plan, and good to see so much attention devoted to diagnostics!

The conditional verification approach is a very interesting one. An element in the validation which is not really mentioned, but which I believe is quite relevant for the validation and testing of physics parametrizations and which would add to the conditional approach, **is the (routine) use of a set of well-defined, well-observed case studies (like ASTEX, RICO, GABLS, …) and associated forcing or observational information**. These can be used also for the diagnosis of timing errors/daily cycle behaviour etc., in a controlled way.

The way in which Conditional Verification is carried out in order to provide feedback to the modellers is discussed in detail in Section 11.2 where the use of properly chosen test cases is noted. It is true that for this purpose, the use of case studies from international experiments such as those mentioned could add value, but only in the case that the accompanying observational datasets contain the specialised data required for analysing interdependencies.

SP chap9 – Reviewer Comments

Jeanette Onvlee:

The feature-based, fuzzy, probabilistic and high impact weather verification techniques are sound, but they all depend very much on gridded data. What I find lacking a bit is a **strategy how to get the observations** you need for this, **other than radar**. It is important to also be able to assess non-precipitating conditions! **Observations on clouds from SAF's and ground-based systems like lidars, and surface observations** (both in-situ and remote sensing) deserve more consideration in my opinion, especially for validation of the boundary layer. Also: many data sources are mentioned on p.94 as possible data for verification. But for all these data the aspect of **quality control and bias correction** is one which should not be taken for granted.

It is true that most of verification techniques that require gridded data, heavily depend on precipitation data from radars that QC is an issue that has to be always faced. Even more tricky this can be with the use of satellite data. On the other hand the participation of COSMO to international initiatives like MesoVICT, can provide for selective test cases, gridded surface observations of very high resolution for a number of weather parameters suitable to be used for spatial methods.

SP chap9 – Reviewer Comments

State of the art, scientific developments

Marion Mittermaier:

"This can be one simple option, but it is clear that, in this new framework**, traditional "point-wise" verification measures can no longer reflect the real quality of forecasts information provided by such a model**":

F S: You are right: we should check your interesting approach of "A strategy for verifying near-convection-resolving forecasts at observing sites ". Could we say it is an extension of Susanne Theis approach some years ago?

FG: It is true that point verification approach cannot be adequate especially for precipitation forecasts but also for continuous parameters extracted from very high resolution model implementations, for this reason there is a dedicated action in 9.3 "Statistical methods to identify the skill of convection-permitting and near convection-resolving model configurations" which is based largely on the recent paper (2014) of M. Mittermaier on "A strategy for verifying near-convection-resolving forecasts at observing sites".

SP chap9 – Reviewer Comments

Strategy of COSMO and actions proposed

Barbara Früh:

What I always wonder about forecast verification is if the **forecasts are verified against observations which are assimilated**. I have not read anything about a **cross-validation** in the science plan.

It is true that there is no action that connects the QC of measurements that enter the DA cycle with verifications systems and limited international work on this (e.g. D.M. Barker, NCAR). There are major efforts undertaken in DA to accurately estimate observation and forecast error and some of the QC algorithms could potentially be used to "clean" data enter a verification system but this kind of approach is not part of the current focus of the verification group.

To detect severe weather events it might help to **compare the forecast with climatological probabilities of occurrences** of these kind of events from hindcasts to account for the model shortcomings.

This is partially performed for precipitation through the use of SEEPS index by using the climatological distribution at each location to define thresholds at which the score is calculated assessing the locally important aspects of the forecast.

SP chap9 – Reviewer Comments

Marion Mittermaier:

"Despite the fact that any long-term monitoring of forecast skill against gridded **observations can include biases not connected only with model evolution**…": Mittermaier et al 2013 make this point when using radar-rainfall. **It is okay for comparing models, but not so good when considering a single model**. Where should the errors be attributed?

The point of the reviewer is indeed taken into account in the "Exploitation of available observation datasets" action, that as the radar baseline is not as stable in time, the combination of radar precipitation and rain gauge data is used as the next stage in precipitation field estimation not as radar data correction but as a better representation of precipitation. In line with the reviewer comment, as it is written on page 92 "Despite the fact that any long-term monitoring of forecast skill against gridded observations can include biases not connected only with model evolution, these data can be valuable when comparing an existing model configuration against another".

"**SEDI and SEEPS together applied can provide complementary assessments of forecast performance**": The SEDS family of scores needs to be recalibrated if interested in extremes, which for me is a weakness. The **SEEPS was not designed for extremes but does provide a very useful, visual way of displaying forecast issues** to local biases because it utilises a climatology. The biggest challenge here is that everyone uses the **same climatology**, and updating/maintaining such a climatology is a non-trivial task that needs to be taken on by one institution. So far ECMWF has done this.

FS (answer): Thank you for these clarifications.

SP chap9 – Reviewer Comments

Marion Mittermaier:

"As the information content of **satellite observations increases**, effort must be given for further exploitation of the data as **radiation (shortwave, longwave),** cloudiness, vertically integrated water vapour content, for verification purposes.": e.g. Crocker and Mittermaier (2013), which also highlight the issues with analyses.

FS (answer): Yes, interesting with SAL.

"The VERSUS system is able to provide all these types of verification functionalities and the possibilities can be further expanded in the future due to the flexibility of its architecture": Engaging with the end-user is important, but challenging. However basic metrics, such **as tracking the percentage of forecasts with temperature errors within 2C is one such metric which could be meaningful to both modellers and users**.

This is indeed a basic useful information extracted directly from verification to both modellers and forecasters and such metrics should be summarised and provided within the "Application oriented verification information".

SP chap9 – Reviewer Comments

Processing verification feedback on model development

Authors: Matthias Raschendorfer (DWD), Flora Gofa (HNMS), Christoph Schraff (DWD)

Jeanette Onvlee:

The statistical hyperparametrization and automated parameter optimization approach is an interesting idea, but how to do this safely in a model in which parametrizations are tuned with respect to each other, and how to relate this optimally to EPS??

Matthias Raschendorfer

Well, the principal concept is like that: trying a process orientated optimization of individual parameterizations by 'component testing' (which is aiming just to minimize the feedbacks of parameterizations). Parameter calibration of the complete coupled system (s. PP CALMO)

Adding missing relationships (expressing remaining dependencies of model parameters on the model state by (the very preliminary idea of) a "statistical hyper-parameterization". Remaining "stochastic" errors are contributed to the part "stochastic physics" (Ch 11.3).

Statistical postprocessing is useful but difficult to remain up-to-date with regular model changes. In Hirlam, we plan to do several years of reforecasting runs about every other major release (2-3 years) to provide input for statistical postprocessing. Are you considering something like that? The analysis of such reforecasting data is usually also very helpful to study model biases (combine with your conditional verification?)

IDEA?

Work organization

@ Provide amended comments to reviewers if necessary

@ Define a work plan to implement proposed activities within WG5:

- ***Tackling model performance improvement issues through the use of conditional verification (CV)***

Initiate the regular exchange of common plots conditional reports to WG3, with main findings once it is proved (in various seasons, domains) the interdependency of variables in error. Feedback from modelers in definition of conditional tests is a prerequisite.

- ***Statistical methods to identify the skill (added benefit) of convection permitting and near-convection-resolving model configurations***

Idea of a new PP to be discussed within WG5

- ***Development of tools for probabilistic and ensemble forecast verification***

Part of new PP too, VERSUS extension of capabilities. Coordination with WG7

Work organization

• ***Exploitation of available observational dataset for operational and scientific purposes***

Use of data available from case studies of projects
MesoVICT (surface parameters, precipitation)
SRNWP datapool (radiation and other PBL properties, VERSUS capability)
EURO4M, European reanalysis and observations for monitoring
http://www.euro4m.eu/datasets.html
EU funded project to develop the capacity for and deliver the best possible and most complete gridded climate change time series and monitoring services covering all of Europe.

Radar composite data from OPERA (initate contact with DA community)

Satellite data : cloudiness, surface wind, vertically integrated water vapour content (NWP-SAF)

Work Group 5 Task List (changes are needed to reflect SP actions)

Common Verification Framework

1.1 Operational Verification *Responsible: ALL*

1.2 Responsibility for Common Plots Reports *Responsible: J.Linkowska,IMGW*

1.3 Verification of vertical profiles using TEMP observations, aircraft data (AMDAR) and wind-profiler data ***(add FeedBack Files)*** *Responsible: ALL*

1.4 Dissemination of daily Grib model output Files *Responsible: De Morsier, MCH*

2. Exploitation of observational dataset for operational and scientific purposes

2.1 High density verification of precipitation over Italy *Responsible: E.Oberto, ARPA-PT*

2.2 Exchange of a common data set of non-GTS data DWD *Responsible:U.Damrath* ***(STATUS?)***

2.3 Evaluation of COSMO models in the lower PBL Responsible: Raspanti, Gofa, Kaufmann

2.4 Satellite synthetic products …

3. Evaluation of convection permitting models performance

3.1 Long Term Trend Verification *Responsible: ALL*

3.2 Conditional Verification **(contact with WG3 for design of tests)** *Responsible: ALL*

3.3 Weather Dependant Verification (WDV) *Responsible: ALL*

3.4 Severe and High Impact Weather **(use of SEEPS can be added too)**

Work Group 5 Task List

4. Neighborhood method techniques

4.1 Verification of COSMO-7 **(to be changed)** precipitation forecast using Radar composite network

Responsible: MCH

4.2 Precipitation verification using radar composite network with neighborhood methods *Responsible: N. Vela, ARPA-PT*

5. Verification of EPS products (Cooperation with WG7)

6. Other

6.1 Annual Workshop/Tutorial on VERSUS2 & WG5