



Experience with EPS verification with VERSUS at MCH

**COSMO GM Session VERSUS2-WG7 on EPS verification
9 September 2014, Eretria (GR)**

Francis Schubiger, André Walser, Pirmin Kaufmann
MeteoSwiss



Status of EPS verification with VERSUS 3.3

- Extensive tests of pre-release VERSUS 3.3: many bugs have been detected and corrected by VERSUS PL
 - Points to discuss
 - Cross Model Graphics and ROC area curve
- Problems of loading times, dimension of DB and execution times with COSMO-E experiments
- List of remaining bugs and extensions



Data Report

Availability

Forecast surface

From: 2014-07-17 **To: 2014-07-18**

Stratification: All European Stations

Model: COSMO-E **Run: 12**

Perturbation member :
0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20

Parameter: Precipitation

Method: 06) Mean of points Circle R=15 km

Count	Parameter	Method	Date
[129360] -LEV: 0	Precipitation	06) Mean of points Circle R=15 km (surface)	2014-07-17
[129360] -LEV: 0	Precipitation	06) Mean of points Circle R=15 km (surface)	2014-07-18

Delete Download Text Download Excel Back

Time spent in second = 1

One of the solved bug:

Now members from 0-20 (instead 1-21) is possible

It was not a bug but an undefined requirement

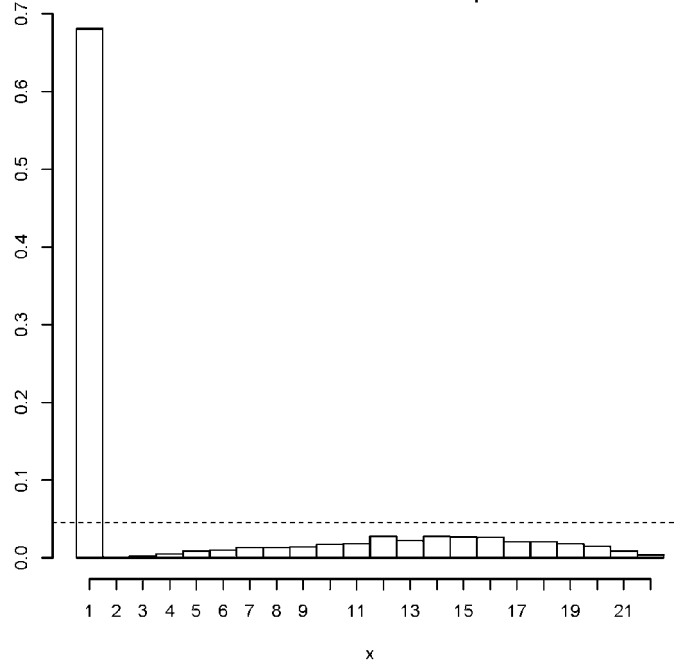


Discussion points

- Rank histogram: 'rank 1' – problem because the least rank will be assigned instead of the random one in case that several members give the same value, i.e. no rain observed

-> what are the advantages of an assignment to the least rank

Rank Histogram COSMO-E 12 PREC06 JJ14 EU – 12 Run
Stratification: All European Stations
From: 2014-07-01 To: 2014-07-31 Step: 24 Th: 2 -9999



instead of the random one ?

-> shall we move to assignment to random rank (instead least), or average rank ?

We followed the instruction described in the document (2011 by A. Boundel)

It can be done but definitely decided

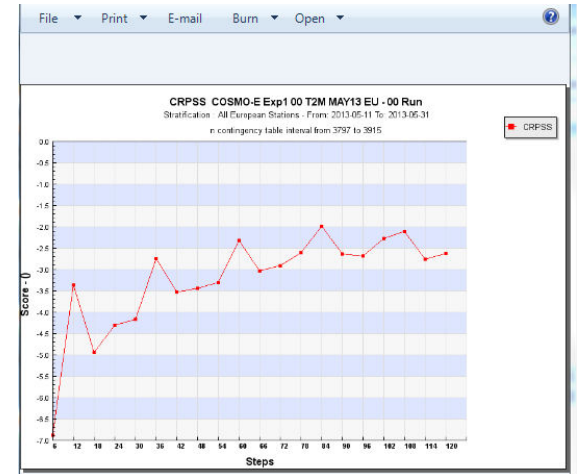


Discussion points (1)

- CRPSS scores seems to be wrong
- does VERSUS calculates a CRPS.clim as reference?
- How is it defined?
→ Anastasia provided input

You can find the details of the calculation in the Angela's document

- For EPS verification the choice of grid-point selection (nearest grid-point [3D-optimized] or radius of 15 km) will have a great influence on the results, namely for spread ?
-> any consensus what to do ?

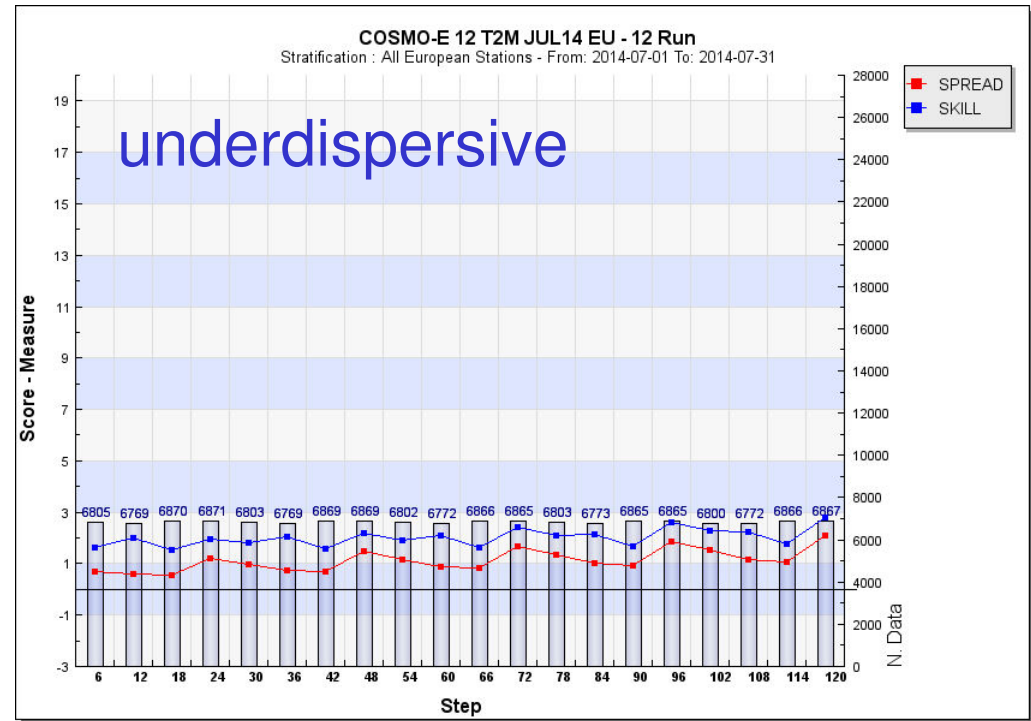
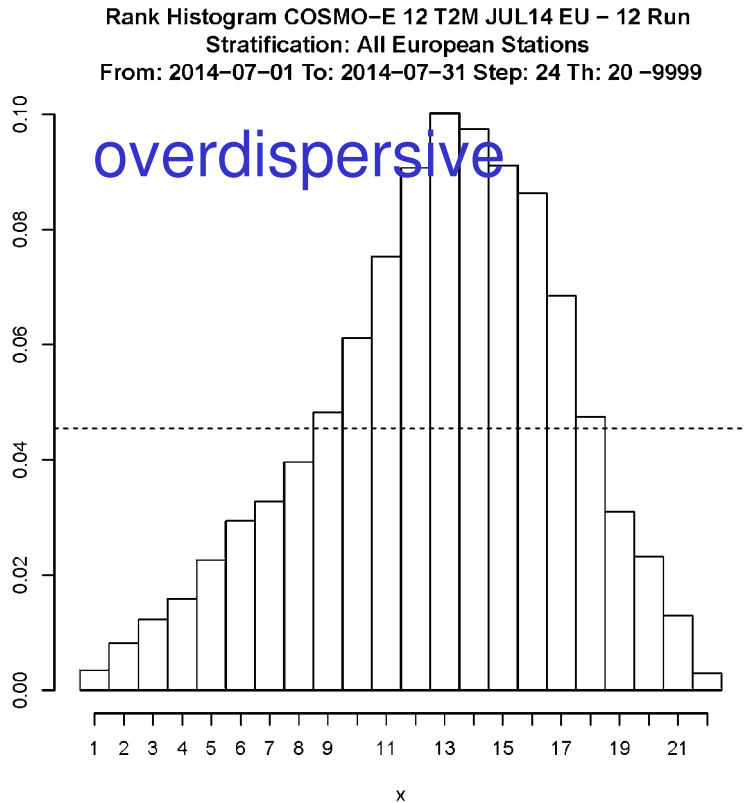




Discussion points (2)

inconsistent results for T2M:

- we rather believe t2m is underdispersive
- how is 'skill' defined, rather the error (RMSE or STDE?)?





Cross model graphics

- Cross model graphics:
 - **Selection for 2m-temperature and 10-windspeed** gives a window with “no verification found”, although these parameters are available for both experiments
 - **Selection for precipitation dichotomic** gives a window with error messages, although precipitation is available for both experiments with 00 UTC runs for 6h-precipitation sums and 12h-precipitation sums
 - **Selection for precipitation continuous** (instead of correctly dichotomic!) gives a window where the two 12h-precipitation verification can be selected (but not also the 6h-precipitation verification) and then as next window the “Graphic Registration”. No possibility to select separately the different scores and thresholds for the plots, i.e. no possibility to produce different plots for the different scores nor to choose the scale min/max.

The screenshot shows a 'Graphic Registration' window with the following configuration:

- Graphic Registration**
- Cross Graphic Registration**
- Criteria: COSMO-E Exp3 00 PREC12 MAY13 EU
COSMO-E Exp1 00 PREC12 MAY13 EU
From: 2013-05-11 To: 2013-05-31
- General**
- Title: TOTAL PRECIPITATION PAST 6 HOURS
- Scale Min: -0.1, Scale Max: 1, File extension: PNG
- X title: Step, Y title: Score - Measure
- Legend**
- Method: (06) Mean of points Circle R=15 km, Strati: All European Stations
- BRIER COSMO-E-TST-12, Line Style: Square, Color: red, Opacity: 1
- Legend**
- Method: (06) Mean of points Circle R=15 km, Strati: All European Stations
- BRIER COSMO-E-215-98, Line Style: Square, Color: blue, Opacity: 1
- Legend**
- Method: (06) Mean of points Circle R=15 km, Strati: All European Stations

During last week this page has been completely restructured. Need to be tested



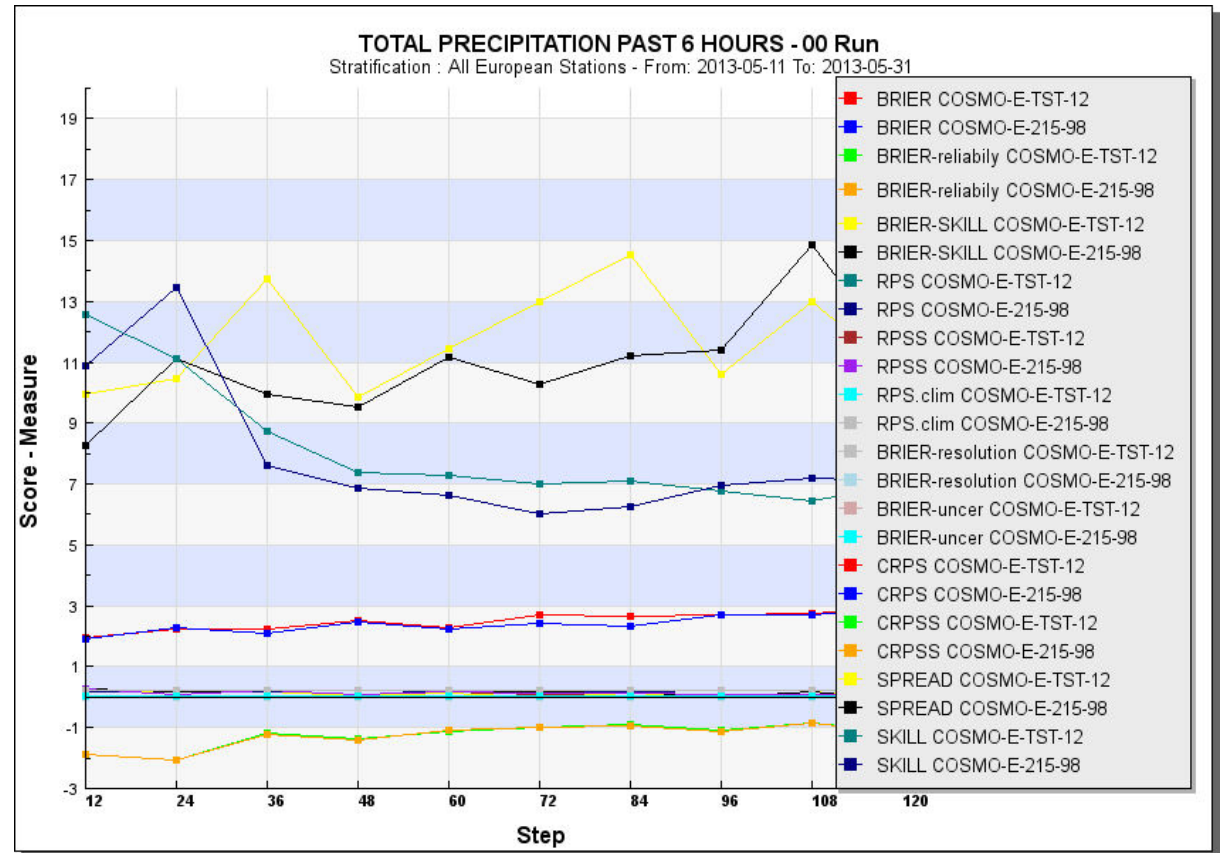
Cross model graphics

This page has been restructured. To be tested again

All scores on one graphic!

The cross model graphic must produce the same number of plots as for individual verification, for both the scores (png-plots) and the EPS results (pdf-files).

Only one value for each score; for the different Brier scores it is not clear which threshold has been taken.



The cross model graphic must produce the same number of plots as for individual verification, for both the scores (png-plots) and the EPS results (pdf-files).



ROC curve area

- **Summary graphic of the ROC area:**

The ROC curve can be plotted and the value of the ROC area is displayed (ROC area = 0.858). A plot of the ROC area vs lead time and vs threshold is needed to get a summary view of the ROC scores. Right figure shows such a graphic done with the software of ARPA-SIMC (for two experiments on the same plot).

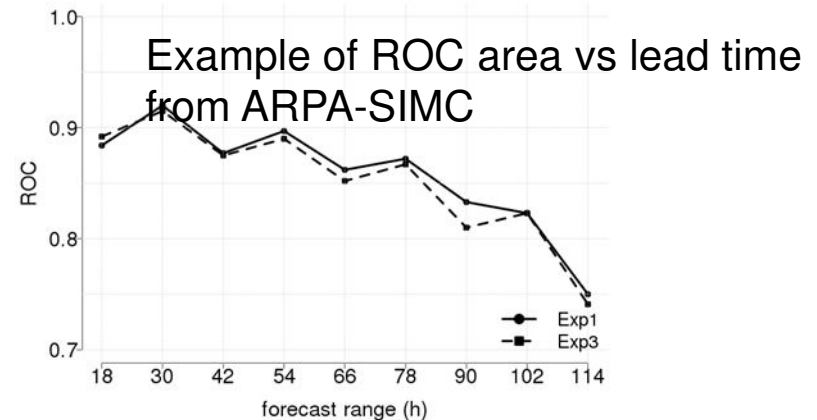
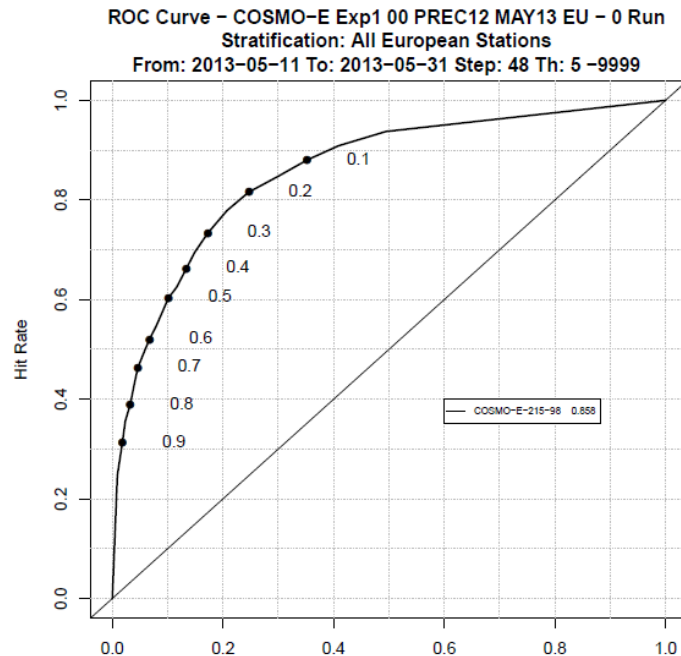


Figure 11. ROC area for 12h precipitation sums > 10 mm as a function of lead-time for exp1 (solid) and exp3 (dashed).

As defined during the CUS with the WG7 agreement, the ROC AREA values are saved in a separated file (txt) in order to be able to perform a ROC COURVE cross model out of VERSUS by yourself.



Loading times for COSMO-E forecasts

- new VERSUS system (since June 2014) runs with 120 GB RAM
- time needed to load one COSMO-E forecast (21 members, all European stations on domain up to +120h, but only for each 3h and without loading cloudiness):
4.1h for an increase of 1.3 GB in the DB with ~3600 GRIB-files

Parameter	Time per Forecast
Precipitation	40'
Surface Parameter	90'
Gusts	35'
Observations (24 h)	8'
Wind Parameters	77'
Total	4h10'



Loading times for COSMO-E forecasts

- time needed to load one COSMO-E forecast
(21 members, all European stations on domain up to +120h,
but only for each 3h and without loading cloudiness):
4.1h for an increase of 1.3 GB in the DB with ~3600 GRIB-files
- If we extrapolate for 1 season 2 daily forecasts, also with total
cloudiness (that takes ~75% of total time) -> 2900h, i.e. 120 days
- If we load hourly data -> factor 3 more: **360 days for one season**
- time needed to load the data for one season COSMO-7 00 UTC
for only areas CH+common area: 9.3h,
i.e. for a season for the 3 daily runs: 28h
- Loading of COSMO-1 & COSMO-E & COSMO-2 /-7 operationally is
not possible if we want also make verifications of our testchains
- But our machine is fast for loading, see results on VERSUS feedback:
TEST6: 0.556" (for 20 days)
TEST5: (Standard): 39", [RHM: 133", RU: 26"]



Dimension of DB and execution times for COSMO-E forecasts

- data amount is ~1.2 GB for one COSMO-E forecast, i.e. $1.2 \times 2 \times 90 = 216$ GB for one season – but we have also COSMO-1 and COSMO-2/-7.
Angela wrote in the forum: “I’ve noted that a DB greater than 150 GB start to be slow”. But we have far more than 150 GB.
- actual dimension of DB: 360 GB
- amount of logfiles (for loading): 3.5 GB/forecast !, that gives 640 GB/season. What is the interest in these huge files?
- execution time: Standard EPS verification does not run for 2 months (1 month 1 parameter all 6h up to 120h) takes ~1.5h
 - in mysql: max_execution_time variable has been raised to 6000
 - still open issue



Analysis OF the DB

	MySQL Version	DB size	InnoDB tables	innodb_buffer_pool_size	innodb_file_per_table
DWD	?	251 GB	NO	4 GB	NO
HNMS	community-5.1.46-2.18 (OS)	322 GB	NO	2 GB	NO
IT	community-5.1.67-1	145 GB	YES	4 GB	YES
IT-EPS	community-5.1.73-1	12 GB	YES	4 GB	YES
IT-ARPA	5.5.33-0.11.1	33 GB	YES	2 GB	YES
MCH	community-5.1.46-2.18 (OS)	322 GB	YES	?	YES
POLAND	5.0.94-0.2.4.1 (OS)	8 GB	YES	8MB	NO
RHM	5.5.37	118 GB	YES	1GB	YES
ROMANIA	?	38 GB	YES	8 GB	NO

MCH :Obsolete MySQL and too big DB you need Update mysql, lower DB

N record = 515442323!!!! Too many records



List of remaining bugs and extensions

- **It is not yet possible to load data and run a verification:** the jobs run sequentially
- **No possibility for multi-users** to work independently on VERSUS

The limitation, derived from R softwarr, has been solved. Task 0.f PP 2014/2015 is a deeper Test phase

- **Implement observation-based verification:** needed in the same way as implemented for deterministic forecast verification; is even more needed than for deterministic forecasts, as ensemble forecasts have a longer lead time (COSMO-E up to 120h). **Requirements for all the partners? No MWS performs that**
- **Implement the suspect observation check:**
 $|\text{median of all members} - \text{OBS}| > \text{threshold}$ **Requirements for all the partners?**
- Produce **all plots in png (the conversion could be made externally)**
- **Produce also txt-files for the EPS-diagrams** (not only for ROC): is needed for cross model graphics and compute number of outliers
(mail of Chiara 25.08.14) **OK**
- **Treatment of log-files unclear:** for batch jobs they are archived; for interactive jobs they are always overwritten. Where are the logfiles of mysql ? **Contact the**



List of remaining bugs and extensions

- Remove the threshold value in the title of the EPS-diagrams (pdf-files), as these results are identical for all thresholds
- Title of the experiment for the plot “Reliability Diagram and Sharpness” is still missing: add it as it has been added for the other EPS diagrams
- Name of the “spread/skill” files is D_IDnumber.png instead of D_SPREAD-SKILL_IDnumber.png

Even if it could seem a easily solvable request, it straight acts on a generic php module which makes it not so easy that specific index to be recognised.