Summary of evaluation of DWD products at the ESSL Testbed 2013

ESSL Report 2013-01
Author: Pieter Groenemeijer

European Severe Storms Laboratory – Science and Training
Bräunlichgasse 6a / 6
2700 Wiener Neustadt, Austria

Tel: +49 151 59031839
Tel: +49 8153 281845
Fax: +49 8153 965999911
e-mail: testbed@essl.org

1. Overview

The ESSL Testbed 2013 took place 1 – 26 July 2013 at the ESSL Research and Training Centre in Wiener Neustadt. The Testbed was visited by two DWD R&D employees (Hengstebeck, Fritzsche), and three presentations were given online (Hengstebeck, Paulat/Gebhardt, James). Five DWD forecasters participated (Trippler, Tuschy, Manitta, Doerkopf, Herold). 46 persons in total participated on-site in addition to 8 online-only participants. The Testbed was visited by DWD's Head of Weather Forecasting Services, Hans-Joachim Koppert.

For the evaluation purposes of DWD, the weather could have been more favorable as only a few thunderstorms occurred across the domains covered by DWD products, much unlike the 2012 edition of the Testbed. Written feedback was collected primarily during the first and last weeks, and for a prior case (20 June 2013) for which most data was available. Despite the sub-optimal weather conditions, the Testbed was once again valued highly by its participants who graded the Testbed with an overall mark of 8.6 on a 1-to-10 scale (2012: 8.6).

Four DWD products were assessed at the Testbed: NowcastMIX, COSMO-DE-EPS, Mesocyclone Detection Algorithm, and Convective Initiation Product. For each of these products, questionnaires had been devised which were filled out by the participants. These answers are delivered to DWD in separate documents.

For the first time, the Testbed Data Interface has been made available online after the end of the Testbed at: http://www.essl.org/testbed/data

A Blog describing the daily activities at the Testbed can be found at: http://www.essl.org/testbed/blog

These are two general conclusions after the 2013 Testbed:

- We think that the best way for DWD to profit from the ESSL Testbed is through a direct person-to-person interaction of its research and development people with the participants. Therefore, we recommend that more R&D staff participate on site.
- We noted that, for some products, it is necessary to look at raw data in order to analyse their behaviour. This is particularly true for the mesocyclone detection algorithm, the convective initiation product and the NowcastMIX. We therefore welcome Hans-Joachim Koppert's suggestion to make a NinJo workstation available to the ESSL Testbed next year, which the DWD experts on-site can use to that aim.

2. COSMO-DE-EPS

Several COSMO-DE-EPS products, based on post-processed data by ESSL were offered to the forecasters. We thank the DWD partners who once again provided this data, helped design questions, and provided online training. These products are visualizations that intend to provide an efficient overview over the entire ensemble.

These are some conclusions:

- The various visualizations are complementary. The visualization that shows the shapes of the various areas in which members exceed the threshold is appreciated strongly, as well as the new probability of exceedence display. Perhaps the 'maximum of any member' display becomes slightly superfluous.
- We had a few cases (pre-testbed 20 June, and 3 July), in which the ensemble did not add useful information to the deterministic run. Generally, however, the ensemble provided important additional information.
- Interest was expressed to compare these visualizations with looking at probabilities of threshold exceedence in the neighborhood of a point.
- There were not enough cases during the Testbed period to allow definitive conclusions about cold pool properties. The following cases, just before and after the Testbed, respectively, exhibited a too warm and slightly too dry cold pool: 20 June, N Germany; 3 July, Austria/Czech Rep., 28 July, S Germany. This may have an inhibiting effect on secondary convective initiation, a tendency that was occasionally noted (storm systems propagating not enough 'into the warm air'). More cases can be viewed on the Testbed Data Interface: www.essl.org/testbed/data
- It was noted that on some occasions, as was the case in 2012, the COSMO model was too reluctant in initiating convective storms. In particular 20 June stands out. Although the ensemble provided good guidance for a severe squall-line across SW Germany during the evening, the widespread convection ahead of this system (including very severe storms!) over much of Germany was strongly underestimated. This case probably merits further study.
- Overall, it appears that frequency of convective initiation over mountains to relative to that over flat terrain, is biased a bit too much towards the mountains.

The ESSL Testbed team would like to further intensify cooperation with DWD's COSMO-developers. A few ideas from ESSL's side are the following:

- Evaluate visualizing 'probabilities of threshold exceedence in the neighborhood of a point'
- Test model-based predictors for large hail (i.e. graupel or hail fraction, updraught speed)
- Develop visualizations for ensemble CIN and CAPE, or even thermodynamic diagrams
- Use updraught helicity rather than (or in addition to) SDI to detect supercells

We are looking forward to discuss both these and other ideas.

3. Mesocyclone Detection Algorithm

ESSL thanks Thomas Hengstebeck for his strong engagement in the Testbed 2013. The ESSL team thinks the MDA is a product with great potential. We intensively discussed the MDA product with Thomas in the time leading up to, and during his presence at the Testbed. Despite the great potential, it became clear that the MDA presently still suffers from a few problems. These include:

1. detections of meteorologically insignificant mesocyclones
2. a complicated, and not optimal indication of mesocyclone intensity

3. spurious detections of what appear to be very large mesocyclones, caused by merging of individual regions of rotation (this problem was already known in advance)

Our recommendations include

- to base the intensity indication on the intensity of the detected rotation, and no other parameters such as reflectivity or derived quantities (except where these are needed to perform a basic quality control on the data, e.g. to eliminate spurious detections in the absence of strong precipitation). This makes the product conceptually easier to use in addition to other products. For example, the NowcastMIX is already product designed to assess storm intensity, so that attempts for the MDA to do the same would be superfluous.
- A suggestion would be to aim to calculate rotation in the same units as the cosmo-de model can forecast (vorticity?)
- to require all detected mesocyclones to have a total depth of at least 3 kilometres (to mitigate problem 1).
- to make sure that the base height of the rotation does not have an overly high weight in the intensity level, since it was noted that hail-producing supercells with strong mid-level mesocyclones where given a low intensity number (1, 2 out of 5) , despite producing 8 cm diameter hail. A suggestion could be to leave out the base of height of the rotation altogether and additionally introduce a parameter to characterize concentrated low-level rotation suggestion a probability of tornadoes.
- to introduce detection of meso-anticyclones, since a fair share of supercells in Germany are anticyclonic. Examples of left-moving cells likely to possess significant mesoanticyclones from this year are 3 July (Bavaria), 26 July (Black Forest). For the 3 July case, a mesocyclone was detected, on the flank of where one would expect a strong meso-anticyclone.
- to address the problem of the undesired merging several mesocyclones into one detection of a very large mesocyclone

As was mentioned above, we found for the MDA product that the lack of access to the raw data impeded the analysis of the product's behaviour. This is why in 2014, we would like to use a NinJo workstation to access this data. ESSL would hope that in 2014 a new version of the MDA can be evaluated at the ESSL Testbed.

4. NowCastMIX

The NowCastMIX was presented to the Testbed participants by Paul James, who programmed a visually pleasing display of the product. He was, unfortunately, not able to attend on site this year. The following points constitute a major part of the feedback from the Testbed:

- The NowCastMIX provided a fairly good guidance of the motions of the storms, even for storms not moving with the mean wind.

- Opinions varied among participants whether the NowcastMIX severity levels were appropriate. It would be fair to say that the mean to verify this at the Testbed were limited, and that the availability of the input fields to the algorithm through NinJo would probably help.
- In general, people had trouble with the multitude of severity levels that DWD uses. There are combinations of storm properties that, following the rules of formal logic, have no place in the warning scheme. For other combinations, it is ambiguous which warning level would be appropriate. A simplification of the scheme would be recommendable from a forecaster's point of view. As an unfortunate example, a storm that produced extreme hail damage across northern Germany on 27 July (one day after the Testbed officially closed) was assigned warning level 95, one that does not explicitly mention hail.

We hope to include the NowCastMIX again in the 2014 Testbed programme. It would be good if the NowCastMIX can be delivered in a format that can be integrated into the Testbed display. ESSL strongly encourages personal attendance on site.

Convective Initiation Product

ESSL appreciates the work done by Pierre Fritzsche that enabled us to integrate the Convective Initation algorithm into the Testbed data interface, which made it possible to display it in combination with various other types of data.

It was fortunate that two other persons who co-developed the product were present at the Testbed, John Mecikalski (Univ. Alabama) and Zsofia Kocsis (OMSZ/ EUMETSAT).

The evaluation of the CI product showed that it needs some significant improvements before it can successfully be used by forecasters, as the product exhibited many false detections in what appeared to be stratiform clouds, and missed the majority of initiating convective storms.

Discussions with Pierre, John and Zsofia showed that these were most likely caused by:

1. The usage of a cloud mask produced by EUMETSAT's NowcastSAF which is not optimized for the use of a CI product. It is recommended that a better cloud mask be used.

2. The use of standard 15-minute METEOSAT data, instead of 5-minute RapidScan data. The other experts on-site said that usage of RapidScan data has the potential to greatly improve the algorithm's performance.

3. Problems with the large number of criteria (10) the algorithm used to identify a pixel as being related to convective initiation. It is recommended to study in detail the role that each of these criteria play in the algorithm, and evaluate whether some of them might be redundant.

4. A review of the thresholds to better capture vertical cloud growth into levels which are more closely related to CI, as it was noted that 'bubbling' shallow cumulus clouds are detected more often than deep clouds undergoing the cumulus to cumulonimbus transition.

ESSL would be happy to include an improved version of the CI product in the Testbed in 2014.

ESSL Testbed 2013 - COSMO-DE-EPS

#1

COMPLETE

Collector: Web Link (Web Link)
Started: Thursday, July 04, 2013 5:32:33 AM
Last Modified: Thursday, July 04, 2013 5:48:36 AM
Time Spent: 00:16:02
IP Address: 80.109.154.58

PAGE 1

Q1: Which of the three visualizations do you find most useful for the parameters Reflectivity, Wind gust speed, Supercell Detection Index, and 3-hourly precipitation accumulations ?

About the reflectivity display:

If the ensemble would cover all possible options, the VMEM visualization would help to rule out that some areas would experience convection: in practice the ensemble does not cover all possibilities.

The TCEM is nice, because you can track individual model forecasts.

The PET has the useful property that it is the most objective way to know the probabilty of an event occurring.

Q2: How do the temperature, humidity and wind within convectively generated cold pools compare to observations? Is their expansion realistic?

Respondent skipped this question

Q3: How much and what information did the COSMO-DE-EPS ensemble add to the deterministic run?

On 3 July: it appears that the determinstic COSMO run at 06 UTC was a bit better than most ensemble members in locating the areas of convective activity. The ensemble did not add any information.

Q4: Do your answers above apply to the algorithm's performance on one particular day (e.g. today or yesterday), or on multiple days?

One day,

If one day, please enter the date here: 3 July

Q5: Do you have any other comments regarding the COSMO-DE-EPS?

No recommendation possible based on this one case.

Looking at updraught speeds could be interesting in addition to looking at reflectivity only.
Are there microphysical parameters (graupel hydrometeor class?) that can give guidance about hail occurrence?

#2

COMPLETE

Collector: Web Link (Web Link)
Started: Thursday, July 04, 2013 6:03:40 AM
Last Modified: Thursday, July 04, 2013 6:35:29 AM
Time Spent: 00:31:48
IP Address: 80.109.154.58

PAGE 1

Q1: Which of the three visualizations do you find most useful for the parameters Reflectivity, Wind gust speed, Supercell Detection Index, and 3-hourly precipitation accumulations ?

Respondent skipped this question

Q2: How do the temperature, humidity and wind within convectively generated cold pools compare to observations? Is their expansion realistic?

Background on 3 July 2013: weak cold front moving in from the west, widespread convection along its wind shift line over N Austria and Czech Republic. Weak vertical wind shear, only poor storm organization. Two stations were directly hit by storms with cold pools (hourly sequence of temperature/dew point):

Kramolin-Kosetice (CZ):
13z: 25/13
14z: 18/12
15z: 17/15 (and 16 m/s gust)

Litschau (AT):
14z: 24/16
15z: 15/14 (and 18 m/s gust)

Timing and overage of these storms were very well forecasted by COSMO-DE deterministic run (not so well by the other ensemble members), which allowed a direct comparison with observations. At the two mentioned sites, forecast temperature dropped from 24 to 20C, forecast dew point remained constant around 15C. Highest gusts in close vicinity of these sites were forecasted to be in the 17.5 to 20 m/s range.

Conclusions (for this particular situation):
Temperature drop was underestimated. Gust forecast was in good range of the station measurements, but possibly also too low, given that there was a damage report which suggested maximum gusts at least slightly above 25 m/s.

Interpretation (of this particular situation):
It seems like cold pools were stronger in reality than forecasted by COSMO-DE, but also stronger than the Testbed forecasters expected it in such a weakly sheared and rather moist situation. The 12z soundings (Prague, Postejov, Vienna) showed very dry layers of air around 850 hPa, probably originating from Foehn-like subsidence in the wake of the Alps, given that there was some 10 m/s southerly flow at 850 hPa. Maybe the underestimation of cold pool formation can be traced back to an underestimation of these dry layers, which enhanced evaporative cooling with the thunderstorms. This, along with the small sample of storms, does not allow a conclusion about possible systematic forecast errors yet, but it is logged for further reference.

Q3: How much and what information did the COSMO-DE-EPS ensemble add to the deterministic run?

Respondent skipped this question

Q4: Do your answers above apply to the algorithm's performance on one particular day (e.g. today or yesterday), or on multiple days?

One day,

If one day, please enter the date here: 3 july 2013

Q5: Do you have any other comments regarding the COSMO-DE-EPS?

(this is an amendment to the other evaluation entry about 3 july 2013)

#3

COMPLETE

Collector: Web Link (Web Link)
Started: Thursday, July 04, 2013 8:03:02 AM
Last Modified: Thursday, July 04, 2013 8:09:01 AM
Time Spent: 00:05:59
IP Address: 80.109.154.58

PAGE 1

Q1: Which of the three visualizations do you find most useful for the parameters Reflectivity, Wind gust speed, Supercell Detection Index, and 3-hourly precipitation accumulations ?

Respondent skipped this question

Q2: How do the temperature, humidity and wind within convectively generated cold pools compare to observations? Is their expansion realistic?

Q3: How much and what information did the COSMO-DE-EPS ensemble add to the deterministic run?

unfortunately on jul02 and 04, the eps did not hit the thunderstorm areas...

Q4: Do your answers above apply to the algorithm's performance on one particular day (e.g. today or yesterday), or on multiple days?

Multiple days

Q5: Do you have any other comments regarding the COSMO-DE-EPS?

some other parameters could make sense:
- likely location of covergence zones in the 10m wind field
- CAPE values over threshold

#4

COMPLETE

Collector: Web Link (Web Link)
Started: Thursday, July 18, 2013 1:48:23 PM
Last Modified: Thursday, July 18, 2013 2:48:58 PM
Time Spent: 01:00:34
IP Address: 84.115.17.181

PAGE 1

Q1: **Which of the three visualizations do you find most useful for the parameters Reflectivity, Wind gust speed, Supercell Detection Index, and 3-hourly precipitation accumulations ?**

After looking at several past cases:

The visualizations are complementary. Certainly the visualizations in which individual ensemble members are plotted are necessary to help interpret what the other visualizations that combine data from several members mean. Smoothed probabilities (à la Theis) may be useful in addition to the current visualizations.

Q2: **How do the temperature, humidity and wind within convectively generated cold pools compare to observations? Is their expansion realistic?**

There were no good cases this week.

Q3: **How much and what information did the COSMO-DE-EPS ensemble add to the deterministic run?**

There were no good cases this week.

Q4: **Do your answers above apply to the algorithm's performance on one particular day (e.g. today or yesterday), or on multiple days?**

Multiple days

Q5: **Do you have any other comments regarding the COSMO-DE-EPS?**

We looked at the case of 20 June 2013. On that day, EPS performance (e.g. the 06 runs) was poor, missing most widespread (severe) convective activity except for a severe strongly forced squall line during the mid and late evening across Switzerland and western Germany. The entire ensemble produced solutions very different from the real evolution. This is probably a important case to simulate for model improvement.

Although performance of the ensemble is probably going to increase during the coming years because of various efforts (introducing more ensemble spread through ensemble data assimilation, stochastic physics, improving parameterizations), such cases will, with a reducing frequency, occur for a long time to come.

In such cases, the forecaster must become active in his role of someone "guarding the process". Therefore, it is important that the means are provided to the forecaster to optimally monitor the behaviour of the ensemble that enable him to see if the "ensemble is still on track".

For example, at some point on 20 June (around 12-14 UTC), it became clear that at least the deterministic run had too high temperatures over SW Germany. The significance of that observation was not clear since this field was not easily accessible for the 20 ensemble members.

Later on, it was observed that more storms developed in central and southern Germany than in in any ensemble member. An important question is whether it could have been detected earlier that there was a problem with the ensemble.

A way to do this could be to offer the forecaster better possibilities to check the presence of the fundamental ingredients of convection (steep lapse rates, moisture and lift vs. convective inhibition) before storm formation (and wind shear too).

Several participants recommended that ways must be sought to monitor the evolution of these ingredients within the ensemble. This is, of course, very challenging. However, it is a necessary step to be prepared for those (rare, but often extreme) cases in which the ensemble does not give good guidance.

#5

COMPLETE

Collector: Web Link (Web Link)
Started: Thursday, July 25, 2013 6:03:54 AM
Last Modified: Thursday, July 25, 2013 8:13:44 AM
Time Spent: 02:09:49
IP Address: 80.109.154.58

PAGE 1

Q1: Which of the three visualizations do you find most useful for the parameters Reflectivity, Wind gust speed, Supercell Detection Index, and 3-hourly precipitation accumulations ?

When describing the areas and member fraction over a fixed threshold for instance 20mm/h precipitation in 3hours, the information about large values like 19.8mm/h will be lost. Therefore, we will suggest to use the 75 percentil of ensemble members.
The maximum values could lead to high false alarm rate.

Q2: How do the temperature, humidity and wind within convectively generated cold pools compare to observations? Is their expansion realistic?

Respondent skipped this question

Q3: How much and what information did the COSMO-DE-EPS ensemble add to the deterministic run?

The COSMO-EPS is able to depict areas of convection not detected by the deterministic run.

Q4: Do your answers above apply to the algorithm's performance on one particular day (e.g. today or yesterday), or on multiple days?

One day,

If one day, please enter the date here: July 25 2013

Q5: Do you have any other comments regarding the COSMO-DE-EPS?

The color scale should be ajusted to the OPERA reflectivity colour scale.