Testing of
DWD products at the
ESSL Testbed 2012

ESSL Technical Report 2012-04

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

Contents

1 The ESSL Testbed 2012

1.1 Summary of activities

The ESSL Testbed is a project to enhance severe weather forecasting across Europe. The Testbed provides forecaster training, testing of forecasting tools, and is a platform for interaction between researchers and forecasters. At the first ESSL Testbed, occurrences of high-impact weather across Europe were investigated from 4 June to 6 July 2012. This first Edition of the ESSL Testbed took place at ESSL's Research and Training Centre in Wiener Neustadt, Austria, that was formally opened on 21 June 2012.

At the Testbed, 67 participants from 21 countries, including both researchers and forecasters, worked closely together on putting new forecast supporting products and methods to the test. Main activities were to prepare experimental forecasts for severe weather for day 1, 2, 3, 4 and 5 as well as "nowcasts" for the following 2 hours using the available Testbed tools and standard meteorological data. Subsequently, a verification of these forecasts was performed using the European Severe Weather Database, followed by an evaluation of forecasting tools and techniques.

Given the various backgrounds of the participants, an important goal of the Testbed is to acquaint its participants with severe weather forecasting methods and techniques that work universally. The tools that were evaluated included visualizations of high-resolution ensemble NWP models, the satellite-based cloud top cooling and overshooting top detection algorithms, lightning detection, ECMWF's Extreme Forecast Index, and cell-tracking algorithms.

Fig. 1. Testbed Participants preparing an experimental forecast.

In 15 daily "Expert Lectures", broadcast online to remote participants, researchers provided background information on their products and internationally renowned experts in forecasting presented their viewpoints on storm forecasting and its scientific roots.

The 2012 Testbed was organized in close collaboration with the Austrian Central Institute for Meteorology and Geodynamics (ZAMG) and supported by EUMETSAT, DWD, WMO Region VI, VAISALA, the German

Aerospace Center DLR, the City of Wiener Neustadt, the land of Lower Austria, EUMETCAL, EUCLID, ECMWF, Austro Control, and NOAA's GOES-R programme.

For a more detailed description of all Testbed activities, the lectures, and all partner organizations, the reader is referred to the **2012 Testbed Operations Plan** (ESSL Technical Report 2012-01, downloadable at: http://essl.org/media/publications/essl-tech-rep-2012-01.pdf). In the next section, an overview of the forecasting activities at the Testbed is given.

1.2 Experimental convective forecasts

On a daily basis, participants have prepared experimental forecasts for severe weather. These forecasts differ in their validity time period, domain, and predictands. They range from Nowcasts, that have validity time of two hours starting at the moment the forecasts issuance, to Day 5 forecasts, that deal with the weather occurring four days ahead. The forecasts have been issued using a programme with which draw lines are drawn to designate areas in which a particular probability of severe weather or lightning is expected. Table 3 lists all forecasts and areas that can be drawn.

Type	Deadlines (UTC)	Validity (UTC)	Predictands	Domain
Nowcast	1300 (1500 CEST) 1400 (1600 CEST) 1500 (1700 CEST)	1300 – 1500 1400 – 1600 1500 – 1700	watch with indication of expected severe weather type	selected sub-domain
Day 1	0855 (1055 CEST)	0900 – 0600 (next day)	thunder 15 % thunder 50 % level 1 (> 5% severe) level 2 (> 15% severe) level 3 (> 15 % extreme)	selected sub-domain
Day 2	0800 (1000 CEST)	0600 (next day) – 0600 (day + 2)	thunder 15 % thunder 50 % level 1 (> 5% severe) level 2 (> 15% severe) level 3 (> 15 % extreme)	selected sub-domain
Day 3	0855 (1055 CEST), on Mondays: 1300 (1500 CEST)	0600 (day + 2) – 0600 (day + 3)	5% severe	Europe
Day 4		0600 (day + 3) – 0600 (day + 4)	5% severe	Europe
Day 5		0600 (day + 4) – 0600 (day + 5)	5% severe	Europe

Table 1. Forecasts at the Testbed.

The forecasts are issued at fixed times and deal with a particular forecast domain. In the case of the Day 3, Day 4, and Day 5 forecasts, the domain is Europe in its entirety, whereas the Nowcasts and Day1 and Day 2 forecasts are issued for a sub-domain that is decided based on the pre-conceived likelihood of severe weather occurring within that sub-domain. The subdomains are indicated in Fig. 2.

Many of the predictands of the various forecasts relate to the probabilities of severe weather and extreme weather, which, for the ESSL Testbed are defined as in Table 4. A typical forecast map is depicted in Fig. 3, in this case augmented with the observational data to verify the forecasts (severe weather reports from the European Severe Weather Database and lightning strikes from the EUCLID network).

Severe weather	Extremely severe weather
• hail 2.0 cm or larger in diameter • wind gusts 25 m/s or higher • any tornado • rainfall causing significant damage	• hail with 5.0 cm diameter or larger • wind gusts 32 m/s or higher • tornado F2 or higher
The quantities to be forecast in the Day 1-5 forecasts are the probabilities that lightning / severe / extreme weather occurs within a radius of 40 km of any given point.	

Table 2. ESSL Testbed criteria for severe weather.

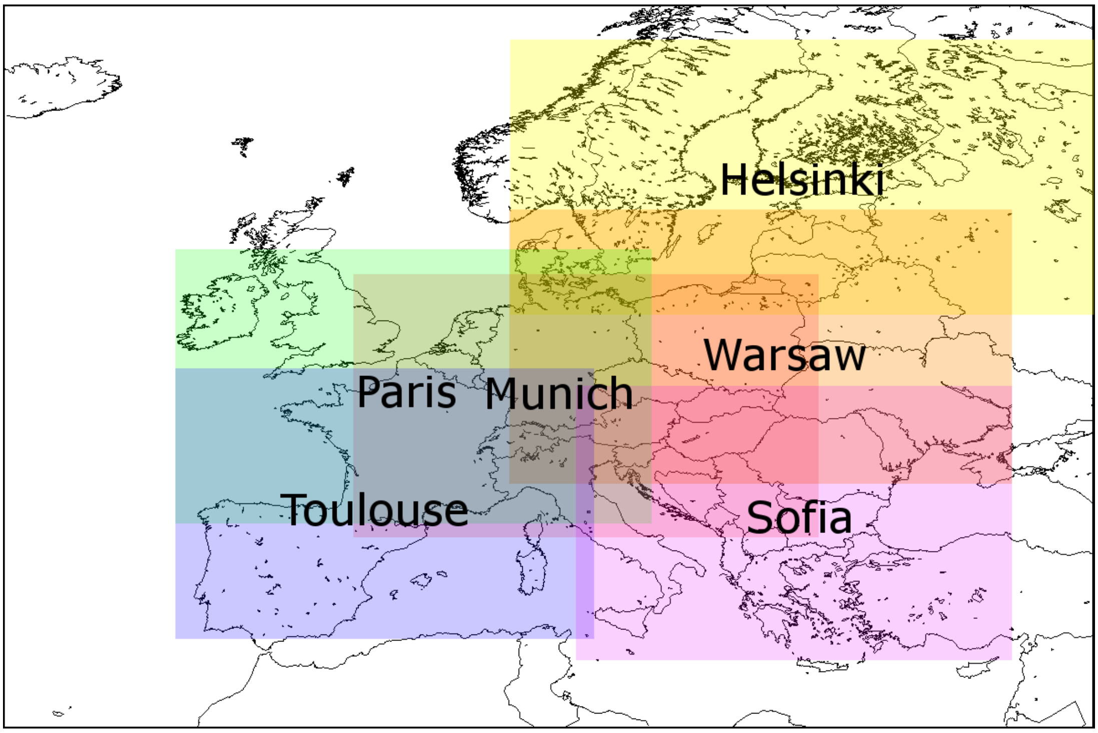


Fig. 2. The ESSL Testbed domain and the six subdomains for which forecasts are to be made.

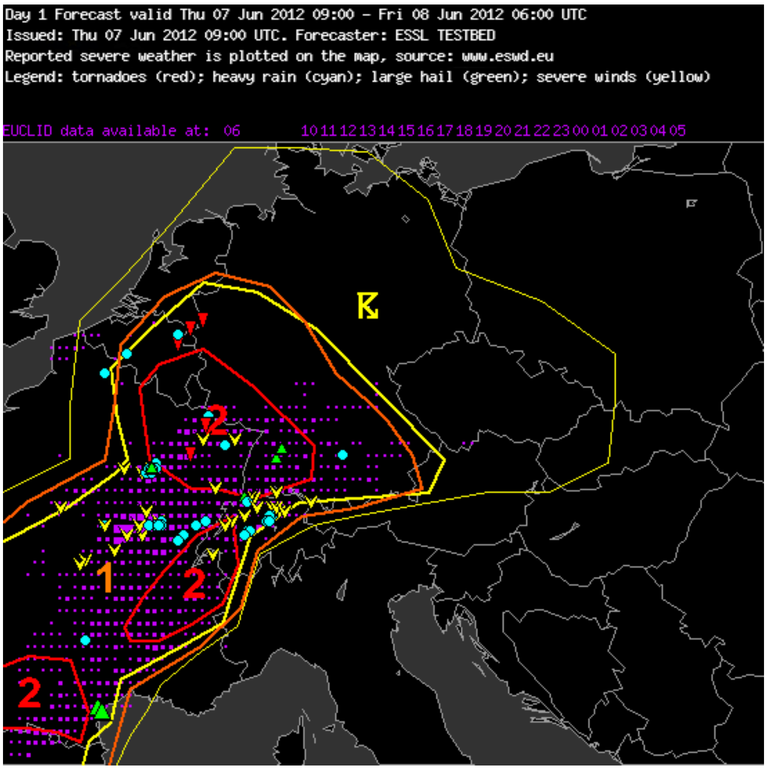


Fig. 3. Day 1 forecast map for the domain Munich showing the areas drawn to depict low (15 – 50 %) thunderstorm probability (inside the narrow yellow contour) and high (> 50%) thunderstorm probability (inside the thick yellow line), as well as a level 1 and 2 of severe weather threat (see Text and Tables 1 and 2 for definitions). The map also shows the data to verify the forecast, i.e. the observed severe weather reports from the European Severe Weather Database: heavy rain (blue circles), tornadoes (red triangles), yellow V's (severe wind > 25 m/s), and large (>= 2 cm) hail (green triangles), as well as lightning detected by the EUCLID network (purple).

2 The testing of DWD products

2.1 Offered products

The testing of DWD products at the 2012 ESSL Testbed has focused on a number of products. For some of these products, feedback was actively collected by means of questionnaires. The products were:

- Visualizations of COSMO-DE-EPS ensemble data, and the COSMO-DE deterministic model
- NowcastMIX

The underlying data of the COSMO data was provided by DWD in gridded binary (GRIB) format on a server, and was processed at ESSL to be visualized in various ways on maps. The NowcastMIX was provided as graphics files. Both products were available throughout the Testbed operations period in near real-time without any noteworthy interruptions.

In addition to these products, DWD kindly provided the Testbed with COSMO-EU model data, that because of its larger domain provided extremely useful information for forecasting outside the COSMO-DE domain.

2.2 Testing Procedure

The products were offered to the participants for use in experimental forecasting. The nature of the COSMO-DE(-EPS) and NowcastMIX implies that they were used for short-term forecasting and nowcasting. The participants were informed about the products and their interpretation through the Testbed Operations Plan, lectures from DWD scientists (Marcus Paulat and Paul James), and a brief introduction at the beginning of each week by Testbed staff. In addition Testbed staff were always present to answer questions arising throughout the week.

During the entire period of Testing This facility will be visited by approximately 12 external participants each week (67 in total) who worked with the products. The participants had various backgrounds, some being forecasters, others scientists, and coming from various European countries and the U.S.A.

Besides introducing the products to the participants, ESSL staff actively requested feedback and encouraged discussions among participants, some of which were plenary and took place in the Evaluation Session organized for this purpose every afternoon. In addition, DWD research and development staff actively sought feedback.

The participants' feedback was collected in written form through questionnaires which were filled in by the respective page of the Testbed intranet. This feedback is provided in the appendices to this report.

In the next section, the DWD products and their testing procedure are described.

2.3 COSMO-DE-EPS and COSMO-DE

Description:

COSMO-DE-EPS is an ensemble of the German Weather Service's high-resolution, convection permitting model COSMO-DE (COnsortium for Small-scale MOdelling - DEutschland). Different initial and boundary conditions of the ensemble members are obtained from four different COSMO runs at 7 km which are in turn driven by different global NWP models (ECMWF IFS, DWD GME, NCEP GFS and UKMO UM) and varying parameters during the assimilation phase. Additionally, the ensemble members use varying model physics parameters.

Testing:

Visualization of large volumes of ensemble model data is a challenge. For example, in plotting the average and standard deviation of forecast quantities much information is lost. For a field like radar reflectivity such a display would not make much sense. At the Testbed, the following visualizations of the following fields will be evaluated:

Field	**Visualizations**
Reflectivity	1. maximum value of any ensemble member 2. coloured areas where any ensemble member exceeds threshold of 40 dBZ
10 m wind gust	1. maximum value of any ensemble member 2. coloured areas where any ensemble member exceeds threshold of 25 m/s
Supercell Detection Index (Wicker et al., 2005)	maximum value of any ensemble member coloured areas where any ensemble member exceeds threshold of 0.001
3 hourly accumulated precipitation	maximum value of any ensemble member coloured areas where any ensemble member exceeds threshold of 20 mm (contours) or 40 mm (shaded)

An illustration of the Testbed display of these COSMO-DE-EPS fields is given in Fig. 4. Several other examples are provided in Section 3, that discussed the results of the testing.

In addition to the COSMO-DE-EPS displays, a standardized set of parameters will be displayed for the COSMO-DE and COSMO-EU models, together with model data delivered to ESSL by other Testbed partners, the global ECMWF IFS, and NCEP GFS, as well as the regional ALARO5 run by ZAMG in Vienna.

Those parameters are: temperature and wind at 300, 500, 700 and 850 hPa, Convective Available Potential Energy, Isentropic Potential Vorticity at the 320 K isentropic surface, mixing ratio and streamlines at 950 hPa, storm-relative helcity, 2 temperature and dew point temperature, vorticity at 500 hPa, convective (where applicable) and large scale precipitation, bulk wind shear in the 0-6000, 0-3000 and 0-1000 AGL layers.

In addition to compiling the questionnaires, ESSL will comment on, and relay comments from participants, on remarkable good or poor performance of DWD models in comparison with other NWP guidance, and on any noteworthy aspects that emerge while using them.

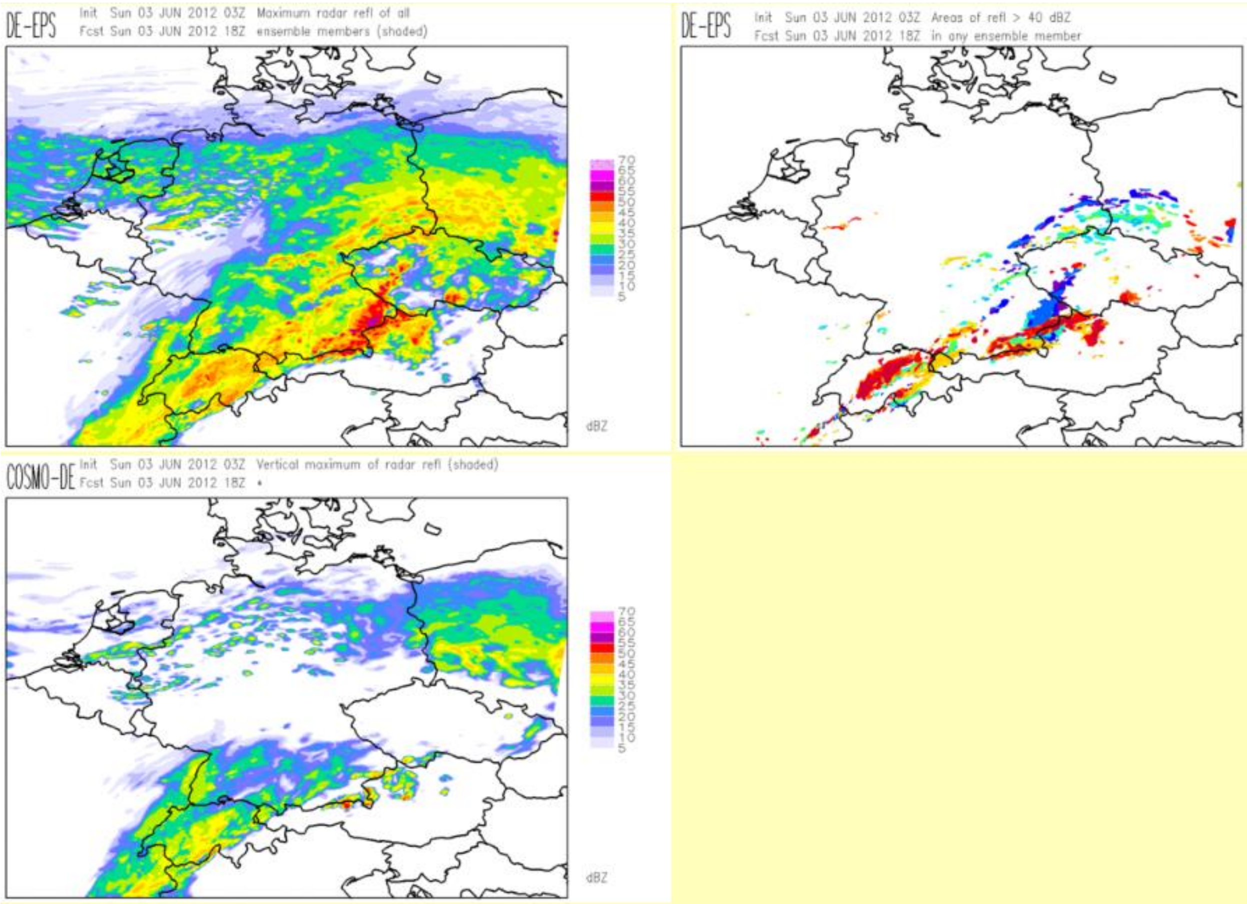


Fig. 4. Top left: Maximum of any member display of COSMO-DE-EPS, top right: Members exceeding 40 dBZ threshold display, bottom: deterministic COSMO-DE run for comparison.

Questionnaire

The questionnaire presented to participants specifically addresses the usability of the ensemble visualizations and the requests feedback on the added value of having convection-permitting ensemble data in comparison with determinstic data only. The questionnaire reads:

Question 1: Please characterize the weather situation in terms of forcing, CAPE, wind shear and other relevant characteristics in the COSMO-DE domain.

Question 2: How would you characterize the deterministic COSMO-DE's performance in indicating the location and time of convective initiation? very poor/poor/fair good/very good

Question 3: How would you characterize the guidance of the COSMO-DE-EPS in indicating the time and location of convective intiation?

Question 4: Please leave your comments on the performance of COSMO-DE and COSMO-DE-EPS regarding convective intiation (too early/too late; spatial displacement).

Question 5: How would you characterize the deterministic COSMO-DE's performance in indicating the intensity and mode of the strongest storms (looking at wind gusts, reflectivity, percipitation and SDI)?

Question 6: How would you characterize the guidance of the COSMO-DE-EPS in indicating the intensity and mode of the strongest storms (looking at wind gusts, reflectivity, percipitation and SDI)?

Question 7: Please leave your comments on the performance of COSMO-DE and COSMO-DE-EPS regarding the intensity and mode of the strongest storms.

Question 8: Reflectivity: Please comment on the usability of the two COSMO-DE-EPS visualizations of maximum reflectivity (maximum of any ensemble member [left], spaghetti plot [right])

Question 9: Wind gust: Please comment on the usability of the two COSMO-DE-EPS visualizations of maximum wind speed (maximum of any ensemble member [left], spaghetti plot [right])

Question 10: SDI: Please comment on the usability of the two COSMO-DE-EPS visualizations of SDI (maximum of any ensemble member [left], spaghetti plot [right])

Question 11: Precipitation: Please comment on the usability of the two COSMO-DE-EPS visualizations of 3-hourly precipitation accumulations (maximum of any ensemble member [left], spaghetti plot [right])

2.4 NowcastMix

Description:

The following description of NowcastMIX was provided by Paul James (DWD):

The German Weather Service's AutoWARN system integrates various meteorological data and products in a warning decision support process, generating real-time warning proposals for assessment and possible modification by the duty forecasters. These warnings finally issued by the forecaster are then exported to a system generating textual and graphical warning products for dissemination to customers. On very short, nowcasting timescales, several systems are continuously monitored. These include the radar-based storm-cell identification and tracking methods, KONRAD and CellMOS; 3D radar volume scans yielding vertically integrated liquid water (VIL) composites; precise lightning strike locations; the precipitation prediction system, RadVOR-OP as well as synoptic reports and the latest high resolution numerical analysis and forecast data. These systems provide a huge body of valuable data on rapidly developing mesoscale weather events. However, without some form of pre-processing, the forecasters could become overwhelmed with information, especially during major, widespread summer convective outbreaks. NowCastMIX thus processes all available nowcast data together in an integrated grid-based analysis, providing a generic, optimal warning solution with a 5-minute update cycle, combining inputs using a fuzzy logic approach. The method includes optimized estimates for the storm cell motion vectors by combining raw cell tracking inputs from the KONRAD and CellMOS systems with vector fields derived from comparing consecutive radar images. Finally, the resulting gridded warning fields are spatially filtered to provide regionally-optimized warning levels for differing thunderstorm severities which can be managed adequately by the duty forecasters. NowCastMIX thus delivers an on-going real-time synthesis of the various nowcasting and forecast model system inputs to provide consolidated sets of most-probable short-term forecasts.

Testing:

The NowCastMIX output is provided in graphical format by DWD on a 5 minute basis, and displayed on the Testbed intranet web site with less than 2 minutes delay (Fig. 5). The Testbed participants will be requested to monitor the product when performing the Nowcast activities for Testbed domains that include Germany, and comment on its performance and usability as well as any other aspects.

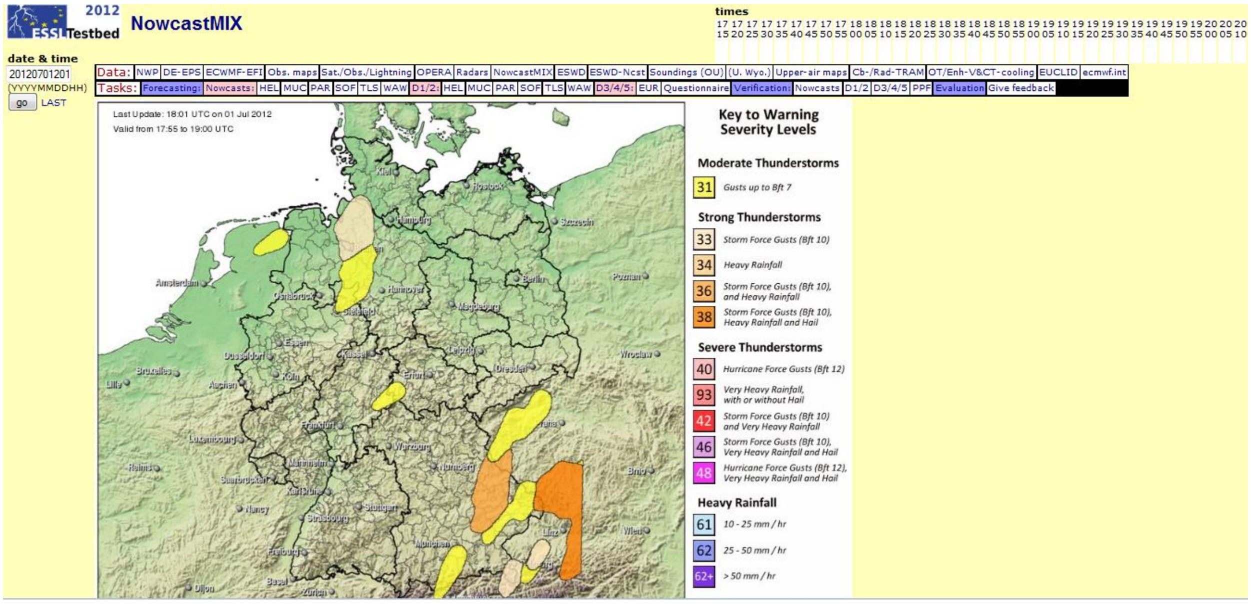


Fig 5. NowCastMIX display at the Testbed.

Questionnaire:

The questionnaire compiled in cooperation with Paul James of DWD, reads:

Question 1: Are the predicted storm cell motion vectors accurate and/or helpful?

Question 2: Do the storm intensity levels (with respect to the attributes: hail, heavy rain and severe gusts) shown in the warning areas give a realistic picture of the current and/or imminent storm characteristics?

Question 3: Does NowCastMIX appear to have specific strengths (weaknesses) in capturing specific storm attributes (hail, heavy rain and severe gusts) more (less) successfully than others?

Question 4: Is the temporal and/or spatial extent of the warnings appropriate (note NowCastMIX must extend the warnings up to the end of the next full hour - a maximum of 2 hours lead time)?

Question 5: Do you have any other comments about NowcastMIX?

3 Results

3.1 COSMO-DE-EPS and COSMO-DE

Before discussing the answers and compiling a summary, it is important to note that the answers to the questions addressed both the usability of the visualizations and the perceived quality of the ensemble simulations. These are two different quantities, but it is natural that anyone asked to evaluate the products will critique both aspects, and that the comments provided here will thus address both as well.

Visualizations

The visualizations of the COSMO-DE-EPS, as described above, for each of the quantities reflectivity, wind gust, SDI and 3-hourly accumulated precipitation use two fundamental ways to display ensemble information, namely one in which for each location the value of the most extreme member is plotted (VMEM) and on in which the threshold-value contours of exceeding members are shown (TCEM).

The VMEM display has the advantage that is uses the same colour scale as the display of the deterministic model. In the case of reflectivity, of isolated storms initiating only in some ensemble members it gives signals that storms may be developing in some ensemble members, then when some extreme value needs to be exceeded as in the TCEM display.

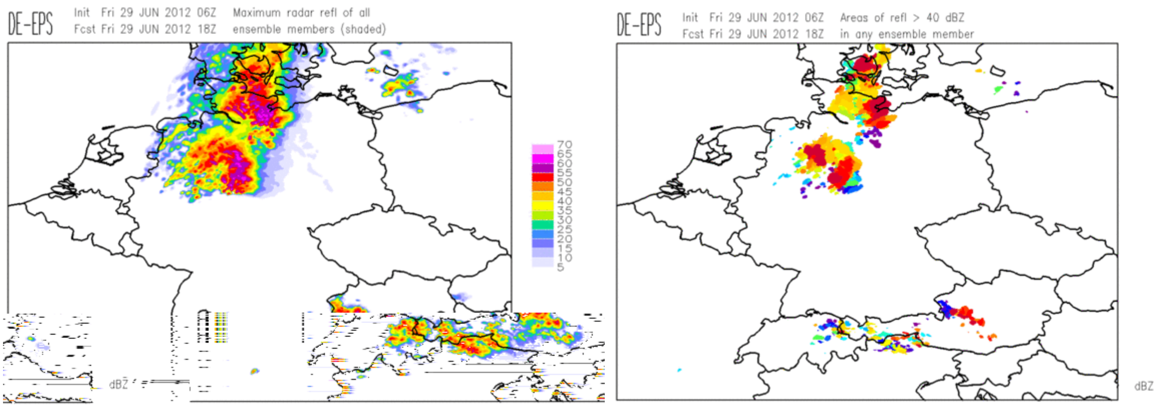


Fig. 6. Value of the Most Extreme Member (VMEM) display (left) and Threshold Contours of Exceeding Members display (right) for reflectivity forecast on 29 June (0600 UTC + 12 h).

The TCEM display has the advantage that it simultaneously indicates where at least one member exceeds a threshold, and gives an impression of how many of the members exceed it. However, since the displays mostly used shades rather than outline contours and because the colours of some ensemble members were necessarily quite similar, given that there were 20 members in total, it was not possible to know the number of exceeding members exactly. A participant remarked:

- *The two ensemble plots are complementary in that they contain different information. The suggestion was made to test plotting the number of ensembles members exceeding the threshold in an additional map.*

In other words, another alternative was suggested, which could not be implemented during the testbed, namely that of plotting the number (or fraction) of threshold-exceeding members for each location (NTEM). In the light of possible enlargements of the COSMO-DE-EPS ensemble to more than 20 members, that would aggravate the possibilities of the TCEM display providing a simple overview, this appears to be an interesting product to develop and test.

The aim of an ensemble is to reflect all possible outcomes given a particular initial state of the atmosphere and its inherent uncertainties, which in case of COSMO-DE-EPS is aimed for both by driving the members by different global models and perturbing constants in the physical parameterizations of the mode. It is clear that DWD developers will try to accomplish that each thus perturbed member represents an equally likely outcome as any other member. Therefore, it may be interesting to have feedback from Testbed participants regarding the handling of deep convection by particular members. In order to make this possible, it is necessary the Testbed be provided with a list that shows which member number is driven by which global model and run with which parameterization settings. In such a way, one could receive feedback like “members with physics perturbation X tends to produce too strong convective outflow wind, regardless of the model driving it”, or “physics perturbation Y tends to be too reluctant with convective initiation when the mid-troposphere is dry”.

Notes about the ensemble

On some occasions the deterministic run of the ensemble system was deemed to be relatively conservative in initiating convective storms. On such days, the signals that the ensemble maps gave by showing that, in contrast to the deterministic run, a small number of members did have storms, was deemed very valuable. An illustration is the case of 18 June (Fig. 7).

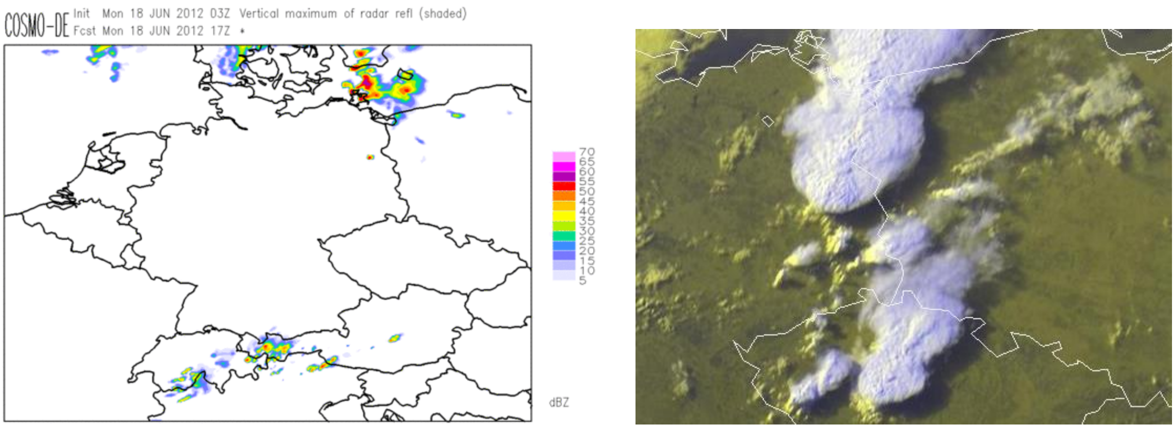


Fig. 7. Reflectivity forecast for 1700 UTC from the COSMO-DE run of 18 June 2012 0300 UTC (left), and E-View satellite imagery for part of the domain at the same time (right).
Satellite data: (c) EUMETSAT/ZAMG/ESSL.

The COSMO-DE-EPS ensemble did however show some subtle hints that strong storms were possible across NE Germany and the Czech Republic (see Fig. 8.).

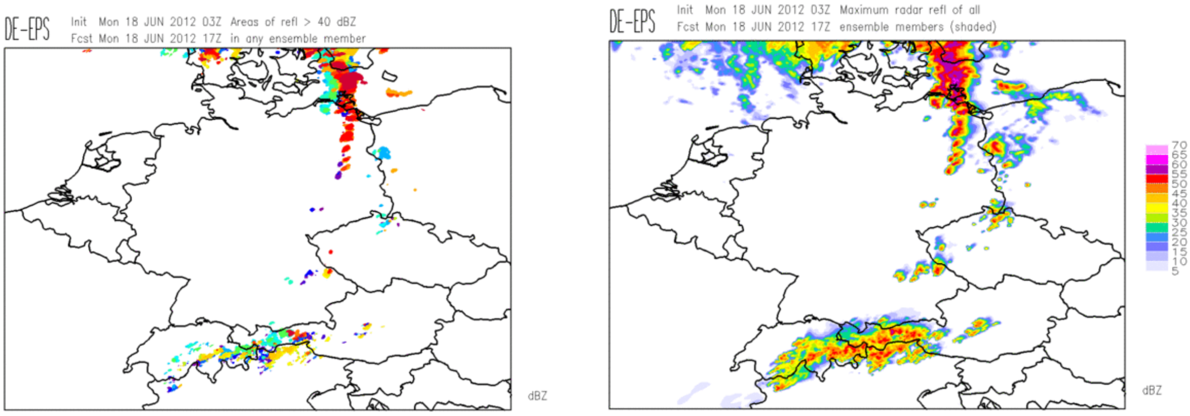


Fig. 8. TCEM (left) and VMEM display (right) of reflectivity for 1700 UTC from the 18 June 2012 0300 UTC COSMO-DE-EPS run.

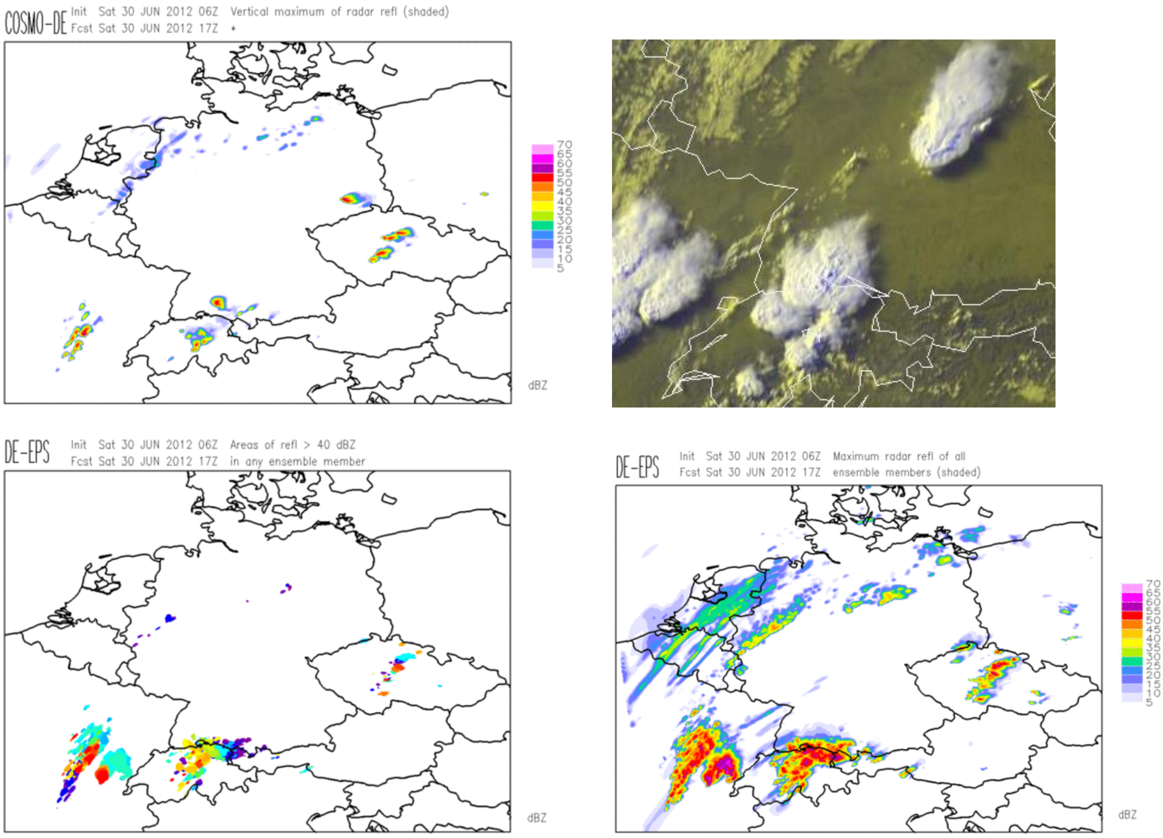


Fig. 9. Reflectivity forecast of COSMO-DE at 1700 UTC from the 30 June 2012 0600 UTC run (left top), and E-View satellite image at 1700 UTC (right top). TCEM (left bottom) and VMEM display (right bottom) of reflectivity valid at 1700 UTC from the 30 June 2012 0600 UTC COSMO-DE-EPS run.

Another example of this kind, a day with similar characteristics (very high CAPE, rather dry mid-levels) was 30 June 2012. The respective four graphics are given in Fig. 9. The COSMO-DE was slightly too reluctant with convective initiation. In this case, the storm located near Würzburg (or any convective activity in central Germany) was not reflected in any of the ensemble members. Fig. 10 shows soundings taken in the vicinity of storms in these two discussed cases with too little convective initiation in the COSMO model.

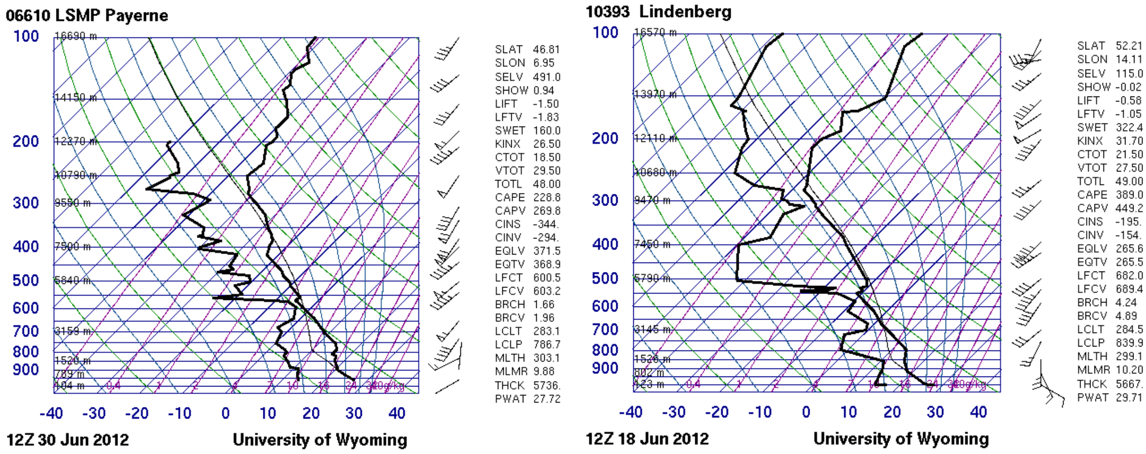


Fig. 10. Soundings on days with too little convective initiation by COSMO-DE.

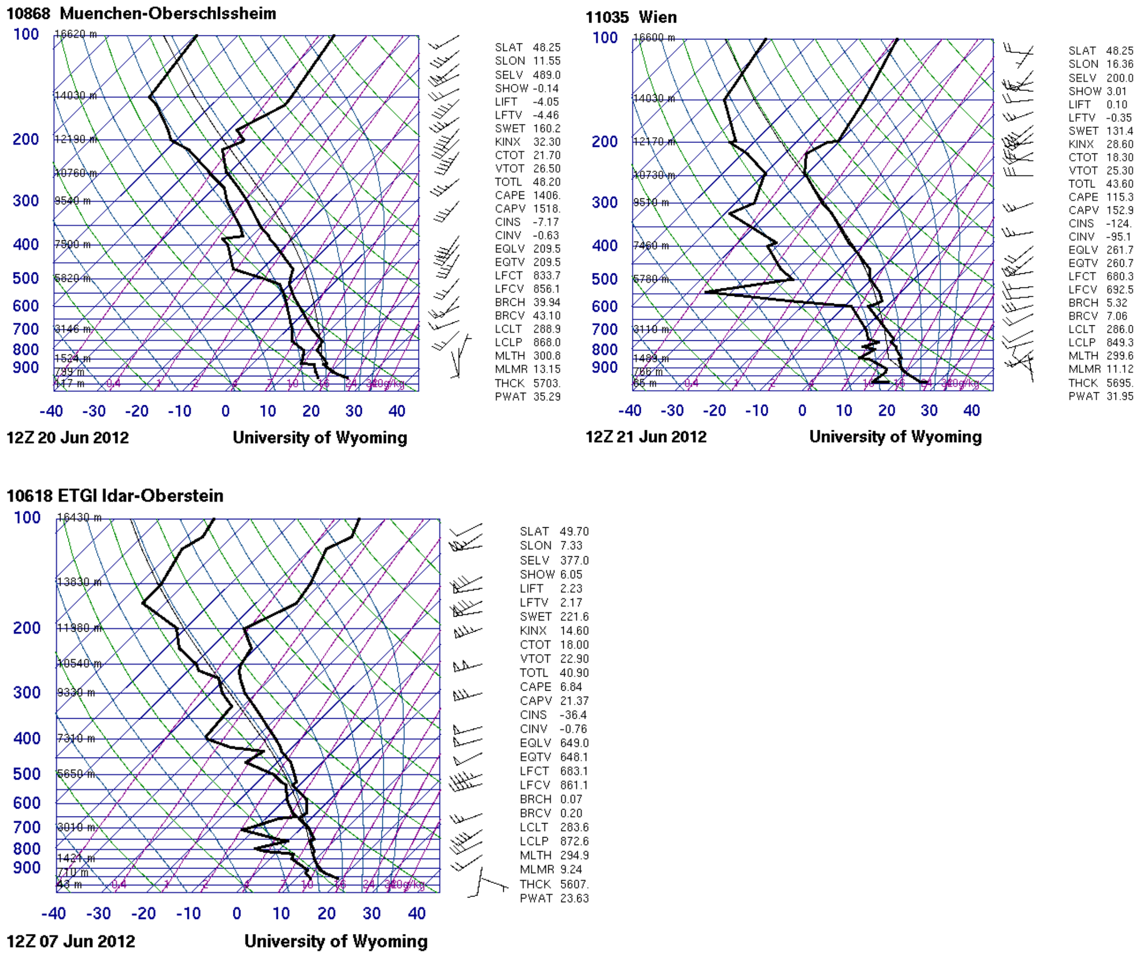


Fig. 11: Soundings on days with good performance of COSMO model regarding convective initiation.

On several other occasions, COSMO-DE performed well or even very well. These were, for example, 7 June, 20 June, and 21 June. A comparison of the soundings in the convecting air-masses on these days

(Fig. 11) with with those in Fig. 10 shows some clear differences. The soundings on the days where COSMO-DE performed well showed modest lapse rates and a relatively humid and shallow boundary layer. However, in the cases that COSMO was too reluctant with convective initiation, lapse rates were very high in a layer above the boundary layer (a so-called elevated mixed layer, EML), while the boundary layer was relatively dry and deep. The size of the presented sample of cases is naturally small, but it might be worthwhile to investigate this further, since it could potentially be an important clue to direct future model improvements. ESSL would be interested in contributing to such research.

Example forecast: 20 June 2012

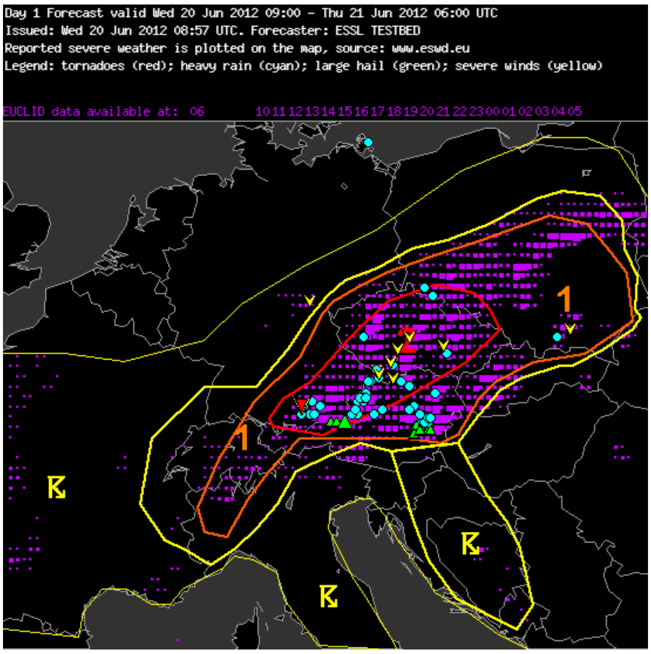


Fig. 12. Day 1 forecast map for the domain Munich on 20 June 2012. See Fig. 3.

In general, participants were positive about the COSMO-DE performance and in particular the DE-EPS ensemble, and at times the performance was even very impressive. A number of comments from the Testbed log, commenting on the model's behaviour on particular days, illustrate this

- *Very good performance with respect to all parameters (reflectivity, SDI, wind gusts, 3h precipitation totals) in both areas, namely the Eastern Alpine region and France! SDI signals seemed to be reliable this time. Region of another severe flash flood event in Central/Eastern Austria (again 50-100 mm of convective precipitation) was forecasted with surprising precision!*

- *The convective mode seems to be predicted pretty well, and it is great to have an ensemble showing supercell / linear-mode likelihood.*

A case illustrating good COSMO-DE-EPS performance and the visualizations was June 21. The forecast map with reports from the European Severe Weather Database (Fig. 12) shows that Bavaria, the Czech Republic, Upper and Lower Austria as well as Styria were affected by flash floods. In addition, large or even very large hail (>5 cm) was reported in northern Tyrol near the German/Austrian border and in central Styria. Severe wind gusts (25 m/s or stronger) were reported from northern Upper Austria into the Czech Republic.

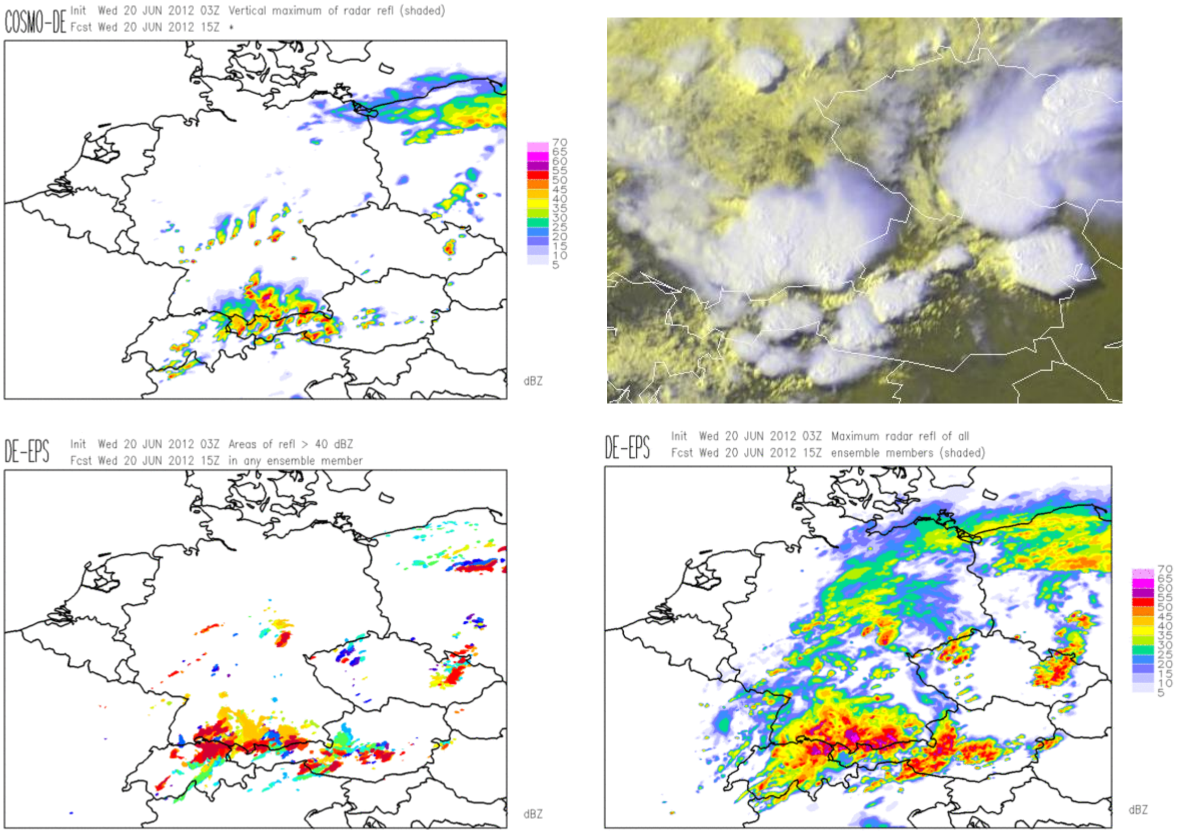


Fig. 13. Reflectivity forecast of COSMO-DE at 1500 UTC from the 20 June 2012 0300 UTC run (left top), and E-View satellite image at 1500 UTC (right top). TCEM (left bottom) and VMEM display (right bottom) of reflectivity valid at 1500 UTC from the 20 June 2012 0300 UTC COSMO-DE-EPS run.

The 03 UTC COSMO-DE and COSMO-DE-EPS simulated widespread convective initiation across southern Bavaria which indeed took place. In addition, the ensemble hinted at developments across much of the eastern Alps, much more so, by the way, than the COSMO-DE. The storms that eventually developed across the Austro-Czech border region were not simulated.

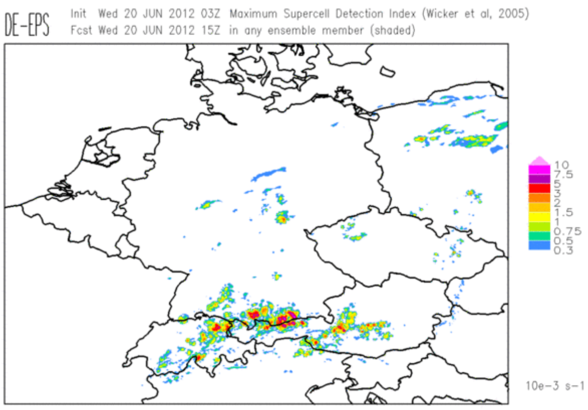


Fig. 14. VMEM display (right bottom) of SDI valid at 1500 UTC from the 20 June 2012 0300 UTC COSMO-DE-EPS run.

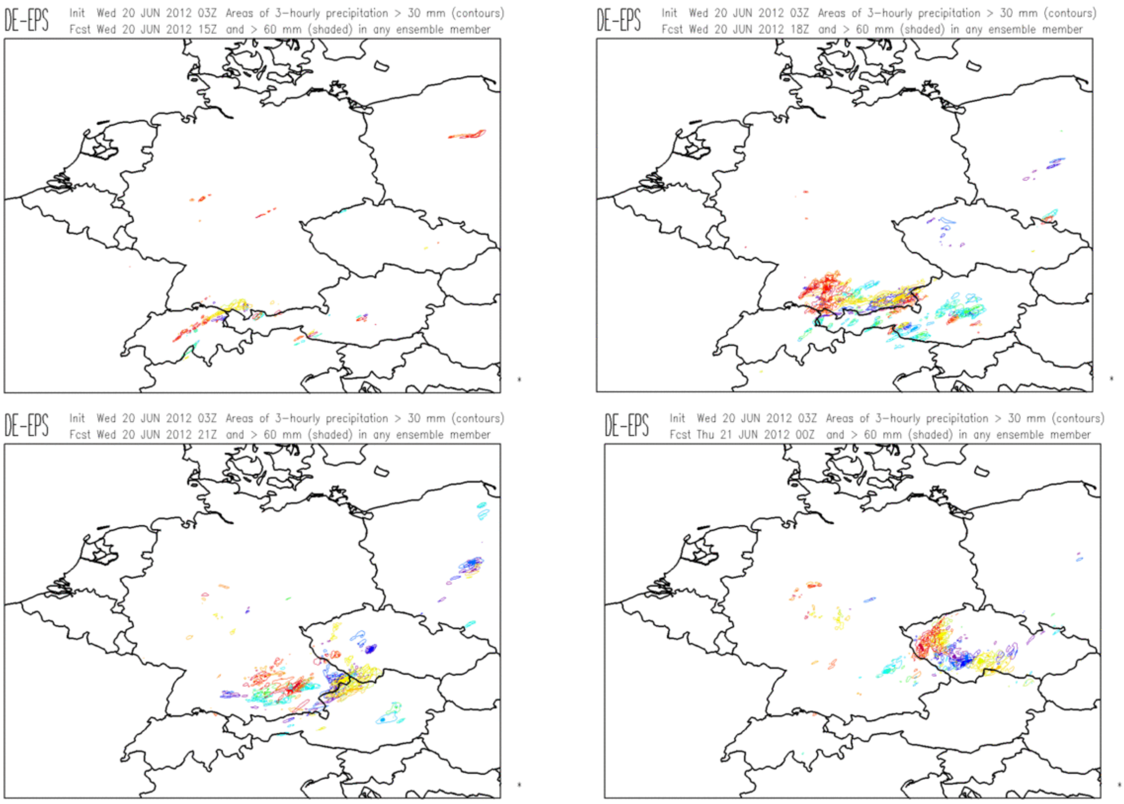


Fig. 15. TCEM display of 3-hourly accumulated precipitation valid at 1500 (left top), 1800 (right top), 2100 (left bottom), 0000 UTC (right bottom) from the 20 June 2012 0300 UTC COSMO-DE-EPS run.

The model suite also hinted at the occurrence of supercells to occur across extreme southern Germany and adjacent parts of Austria, the region in which very large hail was later reported, as very strong SDI signals were present (Fig. 14). Hail larger than 5 cm is typically a sign for supercells, so that this prediction can for most purposes considered to be verified.

For 15 UTC, some marginal signals for high 3-hour precipitation accumulations were present, but for 18 UTC, 21 UTC and 00UTC these signals were even stronger (Fig. 15). The correspondence of these areas with the ESWD heavy rain reports is very good. The one negative point to be noted is that the storms over southern and eastern Styria (area around Graz) that produced very large hail and flash floods (and resulting mud slides) were not represented well in the ensemble. Here, investigation of the thermodynamic profile suggest that, once more, exhibited a strong elevated mixed layer (EML).

Regarding wind gusts, the 03 UTC COSMO-DE-EPS ensemble of indicated a wind gust threat over SE Bavaria and parts of Styria at 18 UTC and across northern Austria and far southern Bohemia at 21 UTC (Fig. 18). A comparison with ESWD observations learns that most wind damage was reported from the Czech Republic, i.e. further to the north than the model's predictions.

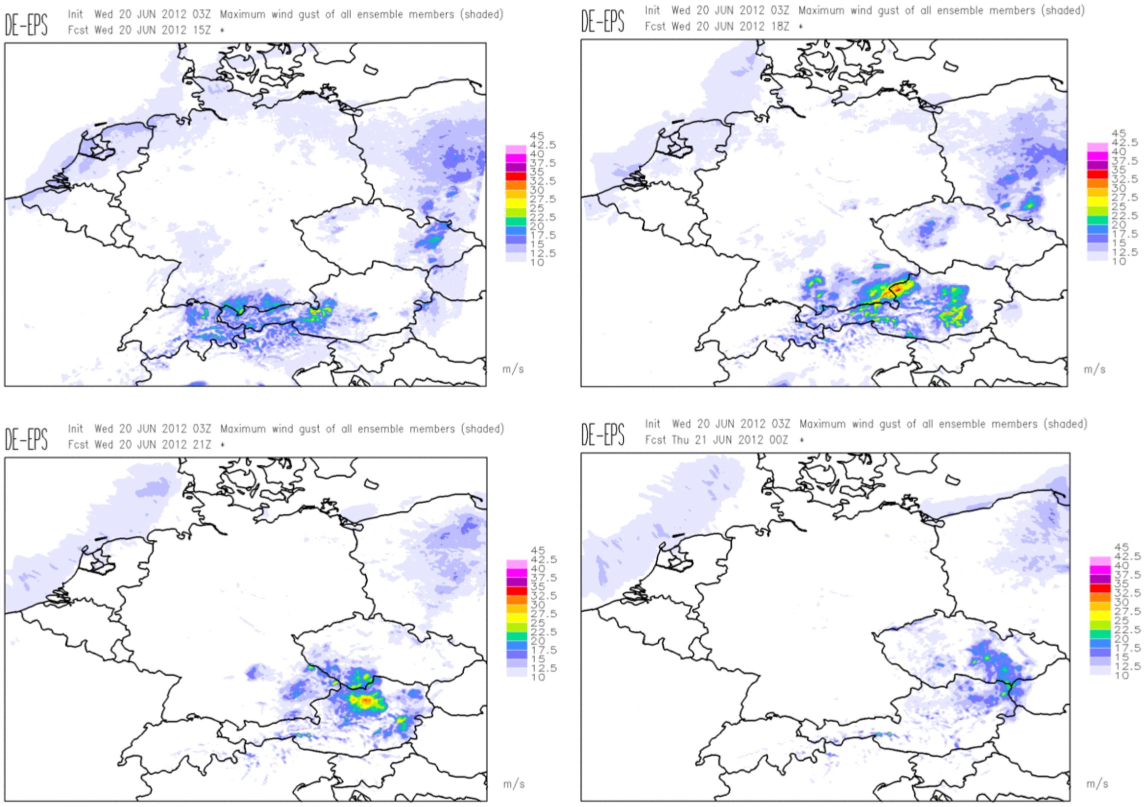


Fig. 18. VMEM display of wind gust valid at 1500 (left top), 1800 (right top), 2100 (left bottom), 0000 UTC (right bottom) from the 20 June 2012 0300 UTC COSMO-DE-EPS run.

COSMO-DE and assessment of the convective environment

Forecasting severe thunderstorms is traditionally strongly based on a two-step process. The first step is an assessment of the environment of the potential storms. A forecaster will look at the potential of the environment of sustaining well-organized severe convection. The second step involves the detection of storms by remote sensing data. An example of this method is a situation in which a forecaster uses NWP output and find that both high CAPE and high wind shear are in place: two prerequisites for supercell and bow-echo storms. Subsequently, upon detecting the initiation of these storms by radar or satellite, the forecaster will be able to issue a warning for severe thunderstorms as he knows that the environment is in principle supportive of these storms. Without knowledge of the environment, the forecaster could not have issued the warning with confidence with the same lead time.

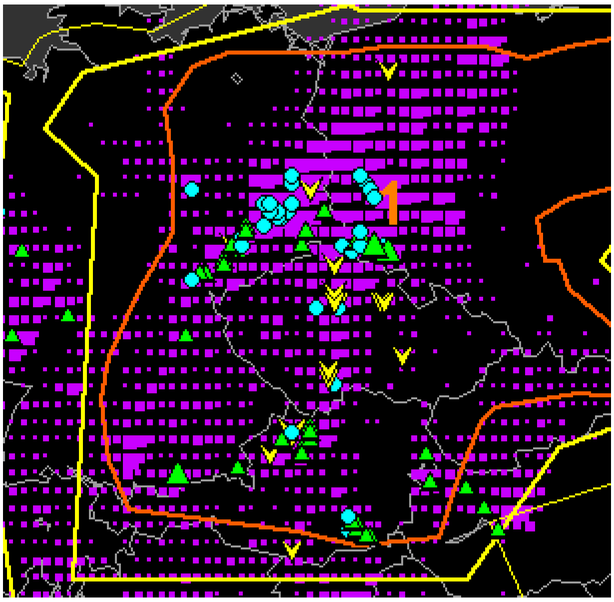


Fig. 19 Testbed forecast and ESWD reports of 5 July 2012. See Fig. 3 for legend.

As the resolution of convection-permitting NWP models has increased to the extent that storm rotation and squall-line structures are partly resolved, it may be tempting to believe that the assessment of the environment has lost importance. And indeed proxies such as the Supercell Detection Index (SDI) can provide guidance on the occurrence of, for example, storms that produce very large hail. There are however situations in which the method will fail. Most clearly this will be when the NWP model fails to produce storms, since the SDI or similar proxies will not show anything in that situation. Given difficulties that virtually all high-resolution models have with correctly simulating convective initiation, this is an important point. Moreover, the models do not explicitly simulate and resolve microbursts, large hail or tornadoes. Yet, the environment can give information about the probability of these phenomena occurring. In this light it is interesting to note that it appears that these environmental parameters may be forecast better in the COSMO-DE than in coarser models. The following example is an illustration.

On 5 June at 1500, convective storms were forming across Saxony and Southwestern Poland that produced widespread large hail. Hail exceeding 5 cm in diameter was reported from two locations in Lower Silesia, strongly hinting at cells having attained supercellular characteristics, i.e. a storm with a rotating updraught (Fig. 19). One of the parameters to assess the potential of storm rotation in Storm-Relative Helicity (SRH). With regard to this parameter it is therefore interesting to see that the COSMO-DE simulation featured a very pronounced maximum of SRH over Saxony and Silesia with values of 200-250 m^2/s^2, where the coarser models had values only near 150 m^2/s^2 (Fig. 20). We speculate that a better resolution of the mesoscale circulation around the Ore Mountains and Lusatian Mountains lead to stronger east-northeasterly surface winds to the north of these mountain ranges, which translated to higher SRH.

Similarly bulk wind shear across the 0-3 km layer, which if exceeding 15 m/s is an indicator of supercell potential, reached values around 20 m/s in COSMO-DE across Saxony and southwestern Poland, whereas all other models featured values of 10 – 15 m/s.

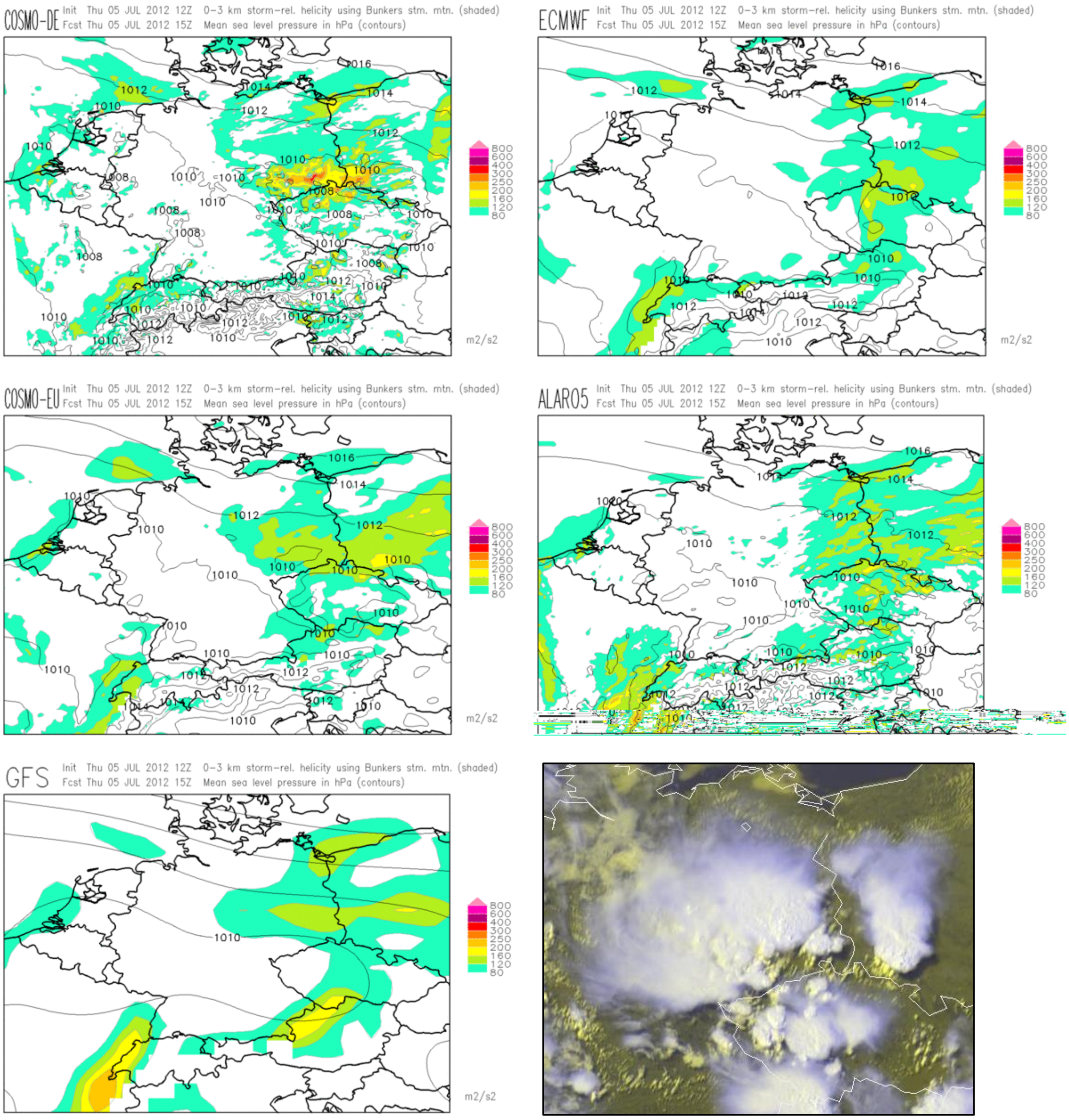


Fig. 19. Maps of 0-3 km storm-relative helicity and mean sea level pressure at 1500 from the 05 July 2012 1200 UTC runs of COSMO-DE (left-top), ECMWF (right top), COSMO-EU (left middle), ALARO5 (right middle) and GFS (left bottom), and the E-View satellite image at 1500 UTC.

During the Testbed more situations that involved similar enhancements of wind shear were noted. This happened at various occasions across the eastern and southeastern flanks of the Alps where the COSMO-DE showed (much) higher wind shear than the coarser models. In these cases, suspected supercells occurred which produced intense severe weather, most notably very large hail and flash floods.

Suggestions for future testing

Several suggestions were made for future testing of NWP at the ESSL Testbed.

- Firstly and most importantly, many participants expressed the wish to have vertical profiles of temperature, wind and humidity ("model soundings"). This is something ESSL would indeed like to offer, although the data volume to be transferred would potentially need to be increased drastically in order to obtain satisfying vertical resolution. A comparison between several numerical models (ECMWF, GFS, ZAMG-ALARO, COSMO-EU, COSMO-DE) could be particularly interesting.

- Another suggestion is to compare the SDI, which is a correlation coefficient between vorticity and vertical speed with an alternative that is used more frequently nowadays at the United States, the updraft helicity. This is the dot product of the vorticity and vertical speed.

- Since the COSMO model uses advanced microphysical parameterization schemes, it could be interesting to explore whether these fields can be used to signal the occurrence of large hail

- The output of the COSMO-DE-EPS does not only contain information on the convective storms that it resolves, but also about the mesoscale environments. It could be interesting to experiment with displays of fields that characterize these, e.g. of instability (CAPE) and wind shear (storm-relative helicity, bulk shear). As an example, the number/fraction of members exceeding a particular threshold value of such parameters, or the mean value could be displayed.

3.2 NowcastMIX

The NowcastMIX product was received with mixed opinions by the testbed participants. There are basically two aspects of the product which can be reported upon after the Testbed. The first aspect is its perceived skill in identifying and providing adequate warning suggestions. The second its the usability of the product in operational warning practice.

Skill

The majority view of the participants was that the algorithm performed well in most cases. The good performance was reflected in participant comments such as

- *it was correctly predicted that the threat of severe wind gusts (25-30 m/s) was highest in the northern part of the squall line*

- *The skills seemed to be evenly distributed for the present case. All in all, the performance of the NowcastMix can be considered very good and reliable.*

- *NowCastMIX was a good guidance regarding the expected strength of thunderstorm activity.*

A recurring observation, however, was that storms with hurricane-force winds were warned for, and that such gusts were then subsequently not reported.

There was also recurring frustration with the large number of warning categories and the associated colour scheme which many participants found not intuitive.

Usability

The participants could be divided basically into two groups, one of them being willing to accept the NowcastMIX as a first guess and expressing that to use it may save time in time-critical situations, even though an forecaster should evaluate it carefully and would indeed typically be able to add quality to the forecast. The other group had the opinion that, if the mental process of analysing the situation must be made anyway to arrive at the best possible warning, the use of having an automated suggestion is of little value.

Either group expressed the desire to understand the NowcastMIX algorithm better than could be provided through the training. For example, on participant commented that...

- *it is very important to train forecasters with this product in such a way they gain knowledge what goes into this product. They have to know what can cause shifts in colours in the boxes and in edges.*

Others commented that...

- *it is important for other users as well to have a good overview what goes into this product.*
- *a learning process is necessary before being able to get the maximum value out of the product.*
- *it seems that forecasters would need to know the specific thresholds used in the algorithm in order to judge how the algorithm is performing or at least decide whether what the algorithm is showing is a reasonable assessment of the current convective activity and what made the algorithm decide to weaken or strengthen specific segments of the convection it is evaluating.*

The fact that these participants made such comments could be interpreted in such a way that they apparently at the Testbed did not feel that they knew enough background information to use the product optimally. An important question is whether NowCastMIX can reasonably be understood, or if the algorithms are inherently too complex for its warning suggestions to be reconstructed mentally in any practical forecasting situation. This is probably not the case, and indeed participants commented that...

- *it is a black box to us, would like to know how it works.*
- *Forecaster cannot understand the product (black box). Jumping behaviour frustrating when not knowing the basis for it.*

The "jumping behaviour" referred to are the changes from one warning category to another for almost the same warning area.

NowcastMIX adds two important components to the underlying cell-tracking algorithms, namely the designation of regions for suggested warnings, and the suggestion of the warning type. It may be that one of these components is better handled in an automated way than the other. It appears that the designation of warnings is generally well-handled by the algorithm. However, the rapid changes in suggested warning type with time and, especially, the lack of possibilities to understand what causes them leads to frustration with its users.

Paul James has provided information on the concept of fuzzy logic that NowcastMIX uses, and explained that the predictors for thunderstorm gusts, for example, are high lower-tropospheric wind speeds in COSMO-DE, rapid cell motion and the radar-determined vertically integrated liquid. It is thinkable that additional or other parameters have predictive value, such as the mode of convection (bowing line-shaped system versus small quasi-circular system), and others. In order to be able to engage in a circle of positive feedback between forecasters and the developers it is necessary that the values of these parameters are provided to the forecasters. For each cell, a pop-up list could be given containing information such as “VIL: 10mm, speed = 12 m/s, max COSMO wind = 14 m/s, max. surface gust obs.: 20 m/s”. In this way, forecasters would be able to reject the suggestion in case any of these input data would be suspicious, be able to provide feedback to developers, and gain more confidence in the product per se.

4 Conclusions and recommendations

Based on the evaluation at the Testbed, we formulate the following conclusions and recommendations. For a longer motivation for these recommendations, please refer to Chapter 3.

Regarding the COSMO-DE and COSMO-DE-EPS and their visualizations:

1. Both the "value of the most extreme member" and the "threshold-contours of exceeding members" displays were found useful by the forecasters with no preference for either of them. Instead the recommendation is that these visualizations be used jointly.

2. It is suggested that a plot will be tested that, for any location shows the number of ensemble members exceeding a particular threshold value.

3. It was suggested that the COSMO-DE-EPS be provided to the Testbed 2013 with a key of which member is perturbed in what way and driven by which global model, this to allow the collection of feedback on the characteristic behaviour of particular members, which can be used as an input to the developers of the ensemble system.

4. A tendency for the COSMO-DE and –EPS to be too reluctant with convective storm initiation was observed in cases featuring a distinct capping (elevated mixed) layer and a relatively deep and somewhat dry boundary layer. In cases where the aforementioned problem did not occur, the accuracy of the location and timing were generally good or very good. A more systematic study of this effect by model developers may be worthwhile.

5. The supercell detection index (SDI) in most cases indicated the location of rotating storms accurately, as far as could be judged from reports of tornadoes and very large (>5 cm) hail. For a better evaluation the Mesocyclone Detection Algorithm should be made available at the Testbed, and good radar imagery should be provided. Another suggestion regarding supercell detection is to test updraught helicity in addition to the SDI.

6. There are strong signals that the COSMO-DE is capable of better representing orographically-induced mesoscale flow systems better than coarser models. This leads to a significant improvement of the resolution of local maxima of vertical wind shear, which has important implications for the assessment of the risk of organized severe storms.

7. Since many participants expressed the wish to use model-based prognostic soundings, something that could not be offered at the first Testbed edition, DWD and ESSL should aim at making this possible for upcoming Testbed editions for the DWD NWP models.

8. Since COSMO-DE uses advanced microphysical parameterization schemes, it could be interesting to explore whether fields like graupel content are useful for forecasting large hail.

Regarding NowcastMIX:

1. No severe systematic problems were noted with the problem in the small number of cases studied at the Testbed, although it may be that the product warns for hurricane-force gusts relatively soon. The location, shape and size of the projected warning areas were thought to be adequate.

2. Many thought that DWD uses an unnecessarily high number of warnign levels for thunderstorms, and thought the associated colour codes were not intuitive.

3. The product was received differently by different forecasters. Some thought it will help in busy warning situations, but many were very reluctant to use it operationally as they felt that the algorithm’s assignment of warning levels was too intransparent.

4. Making the algorithm more transparent by providing forecasters the basic data used for the warning level assignment may help to overcome this reluctance and start a cycle of feedback between forecasters and the developers that will ultimately improve the product.

Appendix A: Unedited answers to questions presented to the participants

In this appendix, the raw feedback from the Testbed questionnaire can be found. Please note that the answers are from groups of Testbed participants. They do not necessarily always reflect the consensus opinion or that of ESSL.

Answers to questions about the COSMO-DE(-EPS) visualizations

On several occasions, participants and/or Testbed staff answered all or some of the questions. These were days after intense convection had occurred across the COSMO-DE domain. These Questions are listed in Section 2.3.

Feedback about 7 June 2012:

Question 4 (CI performance comments):

- CI occurred mostly outside the COSMO-DE domain over France and a mature convective system moved into COSMO-DE domain. COSMO initiated storms over France at the approximate location of this system.

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Good (3/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Good (3/4)

Question 7 (Comments regarding intensity/mode):

- COSMO-DE correctly indicated the location and timing of supercell occurrence (confirmed by observed tornadoes) over Belgium and the area south of Luxembourg. The initiation of storms over France, rather than advection of the more mature MCS into the domain, probably affected the quality of the precipitation forecasts. The COSMO-DE-EPS gave clues for the heavy rain over Switzerland during the evening. The wind gusts reported in the ESWD, with wind speeds likely exceeding 25 m/s, were not captured by the wind gust field of the ensemble model.

Note by ESSL staff*: in each of the COSMO-DE-EPS runs form 00 UTC through 12 UTC, there have been 1-5 members simulating 25 m/s gusts over very small areas (typically ~5 x 5 km), with the signal actually weakening from the 00 UTC 12 UTC. This indeed may be considered a signal to be (much) weaker than required, but it would not be true to say no signal was present at alt.*

Question 8 (Comments on maximum and spaghetti plots for precipitation):

- The two ensemble plots are complementary in that they contain different information. The suggestion was made to test plotting the number of ensembles members exceeding the threshold in an additional map.

Feedback about 12 June:

Question 1 (Characterize weather situation):

- Low pressure system centred over Austria, moving ENE-ward until 1800 UTC, dividing the domain into two areas of completely different characteristics:

  (1) Strong SW-erly upper level flow at its Southern and Eastern flank; numerous thunderstorms developing over Northern Italy, Slovenia, and Croatia in an environment of moderate CAPE and strong shear. Strong dynamic support (short-wave trough).

  (2) Heavy precipitation (supported by upslope flow) in the range of the back-bent occlusion at the Northern and Western flank of the low pressure core over Southern Germany and Western Austria. Some embedded convection occurred close to the low pressure center, but remained unorganized except for two probable supercells in the greater Vienna region

Question 2 (Performance COSMO-DE regarding CI):

- Fair (2/4)

Question 3 (Performance COSMO-DE-EPS regarding CI):

- Poor (1/4)

Question 4 (CI performance comments):

- We separate into the same areas as in question 1 again:

  (1) Convection in the strongly sheared environment over Northern Italy, Slovenia and Croatia was forecasted very well. Over Southeastern Austria, a lot of convection was forecasted as well, but in reality the anvil shading of the Slovenian thunderstorms and later Foehn effects suppressed almost all convection. The deterministic run of COSMO-DE performed better than most of the ensembles, as it at least captured the drying associated with the Foehn effects in the afternoon. The poor forecast of convection went along with an equally poor gust forecast as a "downstream error" (gusts around 20 m/s forecasted but no gusts at all occurred).

  (2) Northern side of the Alps: A large cluster of embedded convection (which caused significant flooding over Northern Austria) was forecasted several hours too early, but the location was forecasted quite well. However, the SDI showed signals even in the region of frontal cloudiness and mostly stratiform precipitation, which we could not understand (weak vertical wind shear and hardly any CAPE!) - especially a maximum of SDI signals in the Munich area, i.e. far away from any actual convective activity.

***Note by ESSL staff**: report of a funnel clouds from Sooß and Böheimkirchen, Lower Austria were later entered into the ESWD, suggesting at least shallow updraught rotation took place with some of the storms further east.*

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Poor (1/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Poor (1/4)

Question 7 (Comments regarding intensity/mode):

- Outside of the Alpine region (e.g. Slovenia, Croatia, or many parts of Germany), the forecast was quite good. However, either meso- and small-scale effects in the Alpine area or in the vicinity of the low pressure core (or both?) seemed to have a negative impact on the forecast quality.

  It is noteworthy, though, that it was a very difficult synoptic situation in general. Other models and even human forecasters had big problems as well to capture and correctly predict all the processes which were going on.

Feedback about 18 June:

Question 1 (Characterize weather situation):

- Synoptic regime dominated by a strongly marked and fast moving upper-level trough through the benelux countries and northern Germany/denmark and Baltic region. At the surface, a cold front extending southwestward from a surface low located over northern Germany and a warm front extending east-south-eastward from the surface low. MLCAPE values between 1500-2500 J/kg were located in the warm sector with deep layer shear values averaging between 15 and 30 m/s over this broad region. Several weaker surface boundaries were also present in satellite imagery within the warm sector over Germany.

Question 2 (Performance COSMO-DE regarding CI):

- Poor (1/4)

Question 3 (Performance COSMO-DE-EPS regarding CI):

- Fair (2/4)

Question 4 (CI performance comments):

- While both the deterministic and EPS runs underestimated the strength and southern extension of the convection, the EPS product better handled the potential for the convection to extend southward as visible via certain EPS member solutions. Perhaps one of the reasons for this was the presence of a rather nicely visible convergence line in satellite imagery that was most likely not well resolved by the COSMO model in general.

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Poor (1/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Fair (2/4)

Question 7 (Comments regarding intensity/mode):

- While the EPS was rather good at resolving the strength of the northern part of the convective line, the southern portion was greatly underestimated. However, the EPS was much better in resolving the correct intensity of the northern part of the line compared to the deterministic run.

Question 8 (Comments on maximum and spaghetti plots for maximum reflectivity):

- The maximum radar reflectivity ensemble product and the ensemble member product were very useful in showing the potential for stronger convection occurring compared to the single deterministic run. At the time of the forecast, it wasn't clear to the forecasters which solution (EPS or deterministic) was going to be the more accurate one.

Question 9 (Comments on wind gust maximum and spaghetti plots):

- It appears that the COSMO-DE-EPS rather nicely identified the region where the highest wind gusts were possible (extreme northern Germany) even if it perhaps exaggerated the intensity of the highest winds compared to the ESWD reports...

Question 10 (Comments on SDI maximum and spaghetti plots):

- The SDI signal appears consistent with the synoptic setting and dynamics present but lack of comparison with a mesocyclone detection algorithm makes a more in depth comparison difficult. No official tornado reports were observable in the ESWD database for this day.

Question 11 (Comments on 3-hourly precipitation maximum and spaghetti plots):

- Same comparison as for the maximum reflectivity products. However, to forecasters, maximum reflectivity allows direct comparison to radar data while 3hr precip totals do not. As a side note, the Alpine overnight MCS was not well handled by the COSMO-EPS (underestimated the precip amounts except for the runs where the initial convection was correctly assimilated).

Feedback about 20 June:

Question 1 (Characterize weather situation):

Eastern Alpine region: Weak SW-erly flow, very high CAPE, pronounced short wave trough provided support for lifting; widespread thunderstorms forming in the afternoon over orography

Question 2 (Performance COSMO-DE regarding CI):

- Very good (4/4)

Question 3 (Performance COSMO-DE-EPS regarding CI):

- Very good (4/4)

Question 4 (CI performance comments):

Timing and location of convective initiation was handled very well for both the orographic convection over Central Austria and the storms forming over Bavaria.

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Fair (2/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Fair (2/4)

Question 7 (Comments regarding intensity/mode):

- Storms were forecast to turn into a large cluster towards the evening hours in the DE/CZ/AT bordering region, whereas such a cluster moved further East in reality. Precipitation amounts were slightly underestimated (rather widespread 50-100 mm reported but not forecast). Nonetheless, the model provided a good hint for a severe flash flood threat in the Eastern Alpine region, and was right with it. Strong SDI signals were present in the forecast, but no hints for actual supercell development to be found in radar data and eye observations.

Feedback about 21 June:

Question 1 (Characterize weather situation):

- Eastern Alpine region: Rapid moisture recovery after last days excessive thunderstorms activity, high CAPE, vertical wind shear increasing to moderate values (15 m/s), again widespread thunderstorms forming over orographic features and later propagating to the SE

  France: Widespread well-organized supercells forming ahead of a powerful short-wave trough, later turning into a squall line. Strong vertical wind shear, favourably veering wind profiles, CAPE higher than expected.

Question 2 (Performance COSMO-DE regarding CI):

- Very good (4/4)

Question 3 (Performance COSMO-DE-EPS regarding CI):

- Very good (4/4)

Question 4 (CI performance comments):

- Convective initiation over the Eastern Alpine region and subsequent backbuilding / propagating was very well forecasted! Initiation over France was very close to the Western domain boundary, therefore some difficulties were visible; in general, initiation seemed to happen ~2h earlier and further to the West than predicted... resulting later in a good forecast, Lagrangian-wise.

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Very good (4/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Very good (4/4)

Question 7 (Comments regarding intensity/mode):

- Very good performance with respect to all parameters (reflectivity, SDI, wind gusts, 3h precipitation totals) in both areas, namely the Eastern Alpine region and France! SDI signals seemed to be reliable this time. Region of another severe flash flood event in Central/Eastern Austria (again 50-100 mm of convective precipitation) was forecasted with surprising precision!

Feedback about 29 June:

Question 1 (Characterize weather situation):

- large hail and rather widespread damaging winds. Ample CAPE and some 20 m/s 0-6km shear. Large-scale forcing for upward motion was somewhat displaced to the NW, but I would clearly categorize it as a strongly-forced pattern.

Question 2 (Performance COSMO-DE regarding CI):

- Good (3/4)

Question 3 (Performance COSMO-DE-EPS regarding CI):

- Good (3/4)

Question 4 (CI performance comments):

- COSMO-DE EPS:

  CI was predicted pretty well, as well as the path of the convective system. The small spread among the ensemble members resulted in rather high confidence in our forecast. (Day one forecast, issued on 29 June 2012)

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Good (3/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Fair (2/4)

Question 7 (Comments regarding intensity/mode) :

- Our forecast was based on the gust probability of COSMO-DE EPS, which showed a clear signal for gusts greater than 32 m/s. Since this was consistent with the environment and synoptic situation, which were deemed conducive to a widespread severe-wind event, our forecast was rather aggressive. However, It seemed that COSMO-DE was too optimistic regarding the occurrence of severe wind gusts. I suspect this could be related to the single-moment microphysics parameterization (fixed intercept parameter, leading to too many small drops and hence excessive evaporation and outflow). Usually, simulations with such microphysics schemes produce too much outflow, so I would not be surprised if this was a systematic problem.

The convective mode seems to be predicted pretty well, and it is great to have an ensemble showing supercell / linear-mode likelihood.

I think the product and the display are really helpful, but perhaps there is a high-gusts bias that the user should be aware of (I am aware that this comment is based on just one case; but perhaps others have observed a similar bias).

Overall, I like to compliment the developers for this product. I think it is quite a powerful tool.

Feedback about 4 July:

Question 1 (Characterize weather situation):

- 12Z: Comparing DE EPS to the other models (GFS, ECMWF,COSMO-EU showing 0-6km SHEAR and MUCAPE) it also doesnt shows high values but realtime Satelite and Radars datas shows that there is rich development of the isolated convective cells over Eastern Germany along the border with Czech Republic. There was a big miscalculation in the rest of the model as well which shows bigger CAPE values in westward and eastward directions from our area of interest.

Question 2 (Performance COSMO-DE regarding CI):

- Fair (2/4)

Question 3 (Performance COSMO-DE-EPS regarding CI):

- Poor (1/4)

Question 4 (CI performance comments):

- 9.00-12.00Z: There are some under- and over estimations over area of central Europe made by DE-EPS: over Nothern Czech Republic shows decaing of convective cells which is happening realtime as well but: over eastern Germany, western Germany, nothern part of Holland it underestimate cenvective initiation and it overestimates over eastern Belgium. Miscalculations seems to be significant for forecasting purposes due to quick development of that cells producing lightnings before 12Z.

Question 5 (Performance COSMO-DE regarding intensity/mode):

- Fair (2/4)

Question 6 (Performance COSMO-DE-EPS regarding intensity/mode):

- Good (3/4)

Question 7 (Comments regarding intensity/mode):

- COSMO -DE: comparing to the GFS, ECMWF and obviously ALAR05 models, COSMO-DE is the most conservative model. However during the early houres(6-12) of the day CAPE and MUCAPE is completly mislocated compering to on-going convection systems at real -time, however it shows some corect signals regarding moisture/dew point values. Going further into a day (after 12Z) its more trustworthly.

Question 8 (Comments on maximum and spaghetti plots for maximum reflectivity):

- Considering only revlectivity values over the corectly forecasted areas usabilty is rather high but there is big miscalculations in forecasting locations where significant reflectivities apeared (12Z overestimations or mislocations over Belgium-Germany border and Czech Republic).

Question 9 (Comments on wind gust maximum and spaghetti plots):

- There was not wind gusts obserwation available at the time of making feedback so making worthful comment is not possible.

Question 10 (Comments on SDI maximum and spaghetti plots):

- Considering first part of the day analize (6-14Z) SDI is high useful product to recognize the biggest cells which have apeared during early afternoon. Indicates areas where bigest development is going to take place. Although in some cases those areas are little bit mislocated (50-100km from real time obs.), indication of the most dangerous area is easier. Model may have some problems with mountainy areas (especially Slowenia also N Czech Rep. and SW Poland) where it does not predict any Supercells which existed in a real time.

Question 11 (Comments on 3-hourly precipitation maximum and spaghetti plots):

- In early time of the day (6Z-12Z) it underestimates possible convective precipitation over whole region (including some heavy rain cases) while in the next period of the day it shows corect locations and intensity of a precipitation. Although there is a few mislocated areas over Poland and Czech Rep. those area are in the minority.

Feedback about 5 July:

Comment:

- Regarding deterministic COSMO DE runs: It picked up well on the local enhancement of shear (example 5th July over southern Saxony with up to 20 m/s 0-6 km shear), which non of the other models predicted. Although on the scale of individual cells one should not rely on the exact time and location forecasted cells, the mode of convection and the pattern provide strong signals into the right direction.

Answers to questions about NowcastMIX

Feedback about 7 June:

Question 1 (Accuracy and usability of cell motion vectors):

- 7 June: looked fine but storms died before they reached into the predicted SW Germany area from France.

Question 2 (Accuracy of representation of storm characteristics):

- Hurricane force winds were not observed but the storm died instead of intensified. Hail was correctly indicated when high tops started to appear near the FR-DE border.

Question 3 (Strengths and weaknesses of capturing storm attributes):

- not enough opportunities yet to judge this, only one day this week.

Question 4 (Appropriateness of temporal and spatial extent)

- looked fine.

Question 5 (Any other comments)

- it is a black box to us, would like to know how it works.

Feedback entered on 13 June:

Question 1 (Accuracy and usability of cell motion vectors):

- Case 13 June: Slight differences between storm / system motion and cell motion possible (case 8 June, S Bavaria), not so obvious with single cells or single storms (EX N Germany same case)

Question 2 (Accuracy of representation of storm characteristics):

- case of 8 June: for the majority of all events the product is reliable. a longer obs period is needed for this question

Question 4 (Appropriateness of temporal and spatial extent)

- case 8 June, N Germany: formation and decay of short-life-cycle- cells affects the product, as for larger systems detail is lost due to time integration (compare northern to southern Germany)

Feedback about 13 June:

Question 1 (Accuracy and usability of cell motion vectors):

- Accuracy of detection overall good. Helpfulness: depending on the task. Not easy to use without long term experience with this product. Forecaster cannot understand the product (blackbox).

Question 2 (Accuracy of representation of storm characteristics):

- Learning process necessary before being able to get the maximum value out of the product. Difficult to distinguish between the different colors.

Question 3 (Strengths and weaknesses of capturing storm attributes):

- No relevant cases.

Question 5 (Any other comments)

- Distinction should be made between convective and non convective features in the terms of the warning duration. 2 hours might be appropriate for more stratiform and larger scale (MCS, frontal structures) features.

Feedback entered on 19 June about 18 June:

Question 1 (Accuracy and usability of cell motion vectors):

- Yes

Question 2 (Accuracy of representation of storm characteristics):

- Yes, it was correctly predicted that the threat of severe wind gusts (25-30 m/s) was highest in the northern part of the squall line, whereas there was a possibility of hail all along the line. The maximum wind intensities were slightly overestimated (hurricane-force gusts predicted but no evidence, taking into account that some reports in the ESWD may still be missing at the time of writing).

Question 3 (Strengths and weaknesses of capturing storm attributes):

- The skills seemed to be evenly distributed for the present case. All in all, the performance of the NowcastMix can be considered very good and reliable.

Question 4 (Appropriateness of temporal and spatial extent)

- Yes. There were appropriate warnings with a lead time of at least 1 hour for all of the selected severe weather reports, in some cases even with a lead time of up to 2 hours.

Question 5 (Any other comments)

- These conclusions are drawn from the 18 June squall line case over Eastern Germany.

Feedback about 21 June:

Question 1 (Accuracy and usability of cell motion vectors):

- The predicted storm cell motion vectors for the June 21st 2012 case were rather accurate. Since this was a strongly forced synoptic setting, this would be expected. In airmass thunderstorm regimes where the steering flow is much weaker, one would expect the algorithm to be less successful, however.

Question 2 (Accuracy of representation of storm characteristics):

- The storm intensity levels shown in the warning areas gave a somewhat realistic picture of the intensity of the convective line crossing Germany on the afternoon of June 21st 2012. See question 3 for additional details.

Question 3 (Strengths and weaknesses of capturing storm attributes):

- Given the lack of high resolution radar data, it was difficult to estimate how well NowCastMIX was able to capture specific storm attributes. However, the transient nature of some of the radar echoes made the NowCastMIX algorithm jump from one intensity category to the other over short time periods which made it sometimes difficult to access the true intensity of the convective line. On the other hand, the overall intensity of the line was reasonably well represented by the algorithm.

Question 4 (Appropriateness of temporal and spatial extent)

- Given the fact that the algorithm bases the intensity of the convection on current radar, wind and lightning attributes it tends to extrapolate the warnings based on that intensity which may or may not be an accurate assessment of the overall nearby synoptic and mesoscale conditions (for example, the algorithm does not take into account the amount of instability located downwind of the convection which may intensify or weaken this convection as it advects).

Question 5 (Any other comments)

- One of the limitations of this automated method appears to be that it is essentially a diagnostic tool and not a prognostic tool which makes it poor at anticipating future changes in convective intensity. This may be a problem for forecasters as they would end up being late in issuing appropriate warnings. Also, it seems that forecasters would need to know the specific thresholds used in the algorithm in order to judge how the algorithm is performing or at least decide whether what the algorithm is showing is a reasonable assessment of the current convective activity and what made the algorithm decide to weaken or strengthen specific segments of the convection it is evaluating.

Feedback about 2 and 3 July:

Question 1 (Accuracy and usability of cell motion vectors):

- Yes, we supppose the cell motion vectors coincidence with the length extent of the boxes.

Question 2 (Accuracy of representation of storm characteristics):

- On that specific day (2nd to 3rd July 2012), Nowcastmix was a good guidance regarding the expected strength of thunderstorm activity. During the purple stage of the Nowcastmix, there wer many heavy rainfall reports in Saxony (E-Germany) and large hail (even up to tennisball size) over far NW Czech Republic.

Question 3 (Strengths and weaknesses of capturing storm attributes):

- Partly, you can not discriminate between severity in for instance hail size. It is important for other users as well to have a good overview what goes into this product. Mos, DMO, observations like lightening, radar, satellite,etc.

Question 4 (Appropriateness of temporal and spatial extent)

- Yes.

Question 5 (Any other comments)

- For the time being it gives me valuable information.

Feedback entered on 4 July:

- I add some extra comments above the ones we added yesterday (July 3 th Rob Groenland). The NowcastMix seems to me usefull for the following reasons: 1: I (as an operational meteorologist) will treat it as a FIRST ESTIMATE 2: it gives me more time to look and check real time data and fullfill the monitoring. Therefore there is time to adjust the edges of the box.

  Another comment I would like to make is that it is very important to train forecasters with this product in such a way they gain knowledge what goes into this product. They have to know what can cause shifts in colours in the boxes and in edges.

  Important again to repeat is that the forecaster will use this product always in combination with other information!

ESSL Testbed 2012 – Feedback

1. How did you like the scheduling of the daily activities (the daily programme) of the Testbed? Could you suggest any improvements?

2. What meteorological information did you miss at the ESSL Testbed?

3. Was the number of forecast-supporting tools sufficient?

- Too few
- Alright
- Too many

4. Can you suggest any tools, products or forecasting techniques to be tested at the next ESSL Testbed (e.g. from your institute)?

5. Will your future work benefit from your Testbed participation? If so, how much and in what way?

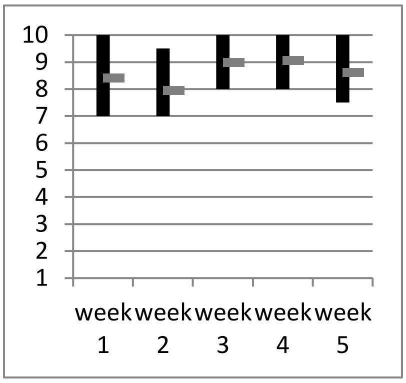
6 Would you recommend others to participate in the ESSL testbed? Why or why not?

7. How would you grade the ESSL Testbed as a whole on a scale from 1 (terrible) to 10 (excellent)?

8 Do you have any other wishes / suggestions for the future?

Thank you for your feedback!

week 1	9	7.5	8.5	7	10	8	8	9.5
week 2	8	7	8	8	8	7	8	7
week 3	9	8	8	10	10	9		
week 4	8	10	10	9	9	8	9	8.5
week 5	8	10	8.5	8	9.5	8	9	9



week 5

					grade		
7	9	8.5	9	week 1	7	10	8.4
9	9.5			week 2	7	9.5	8.0
				week 3	8	10	9.0
10	9			week 4	8	10	9.1
7.5				week 5	7.5	10	8.6
							8.6