Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Operational NWP System

Replacement of COSMO-D2 / COSMO-D2 EPS with ICON-D2 / ICON-D2-EPS

On **February 10, 2021**, the convection-permitting ensemble prediction system **COSMO-D2 (-EPS)** will be replaced with **ICON-D2(-EPS)**, which is based on the ICOsahedral Nonhydrostatic (ICON) Modeling System used for global numerical weather prediction (NWP) at DWD since January 2015. The transition will be accomplished between the 06-UTC and 09-UTC forecast runs. Besides the technical unification of DWD's NWP model chain, the introduction of ICON-D2 constitutes a major step forward in many aspects of forecast quality, as described in this document.

With this transition, the operational NWP-System of DWD with the global ICON, the two-way nested ICON-EU and the convection-permitting ICON-D2 is fully built on the uniform ICON modeling framework. ICON provides a consistent and efficient framework for atmospheric modeling and prediction from global to high-resolution scales.

ICON-D2(-EPS) employs the *Localized Ensemble Transform Kalman Filter* LETKF for the *Kilometer-Scale Ensemble Data Assimilation* KENDA, part of DACE[1]. The step to ICON-D2 also includes a new SYNOP data assimilation and important consistency steps for KENDA, where the observation operators are now available to ICON by a new ICON-DACE interface, such that the observation operator calls become uniform for the ICON-D2 assimilation cycle and the *Model Equivalent Calculator* MEC, on which verification is based for the COSMO consortium.

Technical changes relevant for users: The output of ICON-D2 and differences to COSMO-D2

Almost all fields that have been calculated by COSMO-D2 are also available from ICON-D2 (a few fields even have been added). However, in the ICON model the numerical computations take place on a horizontally unstructured, triangular grid. Therefore, it is natural to do the model output primarily on this ‚**native**‘ grid. Of course, the use of an unstructured grid is a bit less convenient for many customers. Therefore, many of the output fields additionally are interpolated to a structured, **rotated lat-lon** grid that is as similar as possible to the current COSMO-D2 grid.

To practically use the output on the **native** triangular grid, one needs the geographical positions of the output grid points. These are given by the 2-dimensional (2D) fields CLON and CLAT for the geographical longitude and latitude, respectively, for the center point of

1 Data Assimilation Coding Environment

each triangular grid cell. For 3D fields on the model layers, one additionally needs their height information that is given by the 3-dimensional (3D) height field HHL. For the grid center one has to average the two adjacent HHL values. Some variables (in particular the velocity) are staggered compared to the center point variables. For those, there exist the 2D fields ELON and ELAT.

For the use of the regular grid output, one should pay attention to the following facts. First, as in COSMO-D2, the output takes place on a **rotated lat-lon** grid, to more efficiently use the output file size (a geographical lat-lon grid would have unnecessarily small grid cells at higher latitudes). The related geographical coordinates are stored in the 2D fields RLON and RLAT. The regular output grid itself is exactly the same as for COSMO-D2, i.e. there are 651x716 horizontal output grid points at the same geographical positions. In the vertical, 3D model layer fields are defined on 65 layers as in COSMO-D2, too. This means one can read these fields without a change in the field dimensions. However, one should be aware that:
1.) although the output takes place on a rotated lat-lon grid, the velocity components are no longer rotated, but are aligned in the usual (geographical) zonal and meridional direction (in contrast to COSMO-D2),
2.) height levels of 3D fields are not the same because it turned out that the forecast quality benefits from distributing the levels a bit more evenly across the troposphere than it was the case in COSMO-D2. In any case one should read again the HHL field (which exists also for regular output),
3.) less important, but worth mentioning, is the fact that now all variables are interpolated to the same center position (i.e. no staggering of velocity variables as in COSMO-D2).

For a more detailed information about the output fields and their usage we refer to the DWD Database description for ICON. The output fields all are listed in section 11; some properties of these fields can be found in sections 6, 7, and 8, too, and in section 5 for analysis fields. Information about the native grid is given in section 4 and a few hints about the rotated lat-lon output grid is given at beginning of section 11, and in section 8.2. A rough impression about the vertical level spacing give figure 3.9 and tables A.3 and A.4.

The document '*DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System*' by Reinert et al. can be found under https://www.dwd.de --> Fachnutzer --> Forschung und Lehre --> Numerische Wettervorhersage.

Changes in model numerics and physics parameterizations

The dynamical core and the related numerical schemes for mass-consistent tracer transport are identical to the global ICON / ICON-EU system. They are described in Zängl et al. (2015) and Prill et al. (2020). The physics parameterizations are shared with the global system as well with the following exceptions necessitated by the convection-permitting resolution of ICON-D2

- The parameterization for non-orographic gravity waves is turned off, and only the low-level blocking component of the sub-grid scale orography (SSO) scheme is used. Correspondingly, the raw orography data needed for calculating the external input fields for the SSO scheme have a much higher horizontal resolution than those used for the global system (30 m vs. 1 km). In comparison with COSMO-D2, the SSO scheme replaces a highly simplified parameterization representing unresolved orography via an enhanced roughness length. Extensive verification tests during the preparatory phase indicated that the SSO scheme allows a much more realistic treatment of unresolved orography than the previous roughness length approach.

- Only the shallow convection component of the sub-grid scale convection scheme is used. This is also the case for COSMO-D2, but the Tiedtke-Bechtold convection scheme used in ICON, which was taken over from the Integrated Forecasting System (IFS) of ECMWF, is a further developed version of the 'old' Tiedtke scheme used in the COSMO model. In contrast to the Tiedtke scheme used in COSMO-D2, the Tiedtke-Bechtold shallow convection scheme is able to generate small amounts of convective precipitation, particularly over water surfaces. Another important difference is that the ICON-D2 scheme has been tuned such as to avoid excessive moisture transport out of the boundary layer, reducing a long-standing bias issue of COSMO-D2 with a dry bias in the boundary layer and a moist bias in the middle troposphere.

- In addition, ICON-D2 uses the same six-category graupel microphysics scheme as COSMO-D2, whereas the global ICON uses a simpler variant without graupel.

Parameterization schemes shared with the global ICON but differing from COSMO-D2 are the RRTM radiation scheme (Rapid Radiative Transfer Model) and the Köhler scheme for subgrid-scale cloud cover. They also contribute to the forecast quality improvements described below. In addition, the land-surface scheme TERRA and the turbulence parameterization, which are shared between the COSMO model and ICON, have been improved in many aspects in the preceding years, but most of these upgrades so far have been put into operations in the global ICON system only. Examples are a more realistic scheme for bare-soil evaporation, a skin-temperature scheme following the approach used in the IFS, an improved treatment of a partially snow covered surface, and a Richardson-number dependent formulation of the vertical diffusion coefficients used in the turbulence scheme under stable conditions. These improvements are now also used in ICON-D2.

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

DWD

References:

Zängl, G. D. Reinert, P. Ripodas and M. Baldauf, 2015: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc*. **141**, 563–579. https://doi.org/10.1002/qj.2378

Prill, F., D. Reinert, D. Rieger and G. Zängl, 2020: ICON Tutorial. Working with the ICON model. Available online under https://www.dwd.de/EN/ourservices/nwv_icon_tutorial/pdf_volume/icon_tutorial2020_en.pdf

Changes in data assimilation

The basic data assimilation system used to determine the initial states of ICON-D2 forecasts is the same as for COSMO-D2. This system called **KENDA** (Schraff et al., 2016) provides analyses at a 1-hourly interval based on a **LETKF** (Local Ensemble Transform Kalman Filter) scheme. For operational purposes, it currently assimilates **radiosonde** ascent and descent profiles, AMDAR and Mode-S **aircraft** data, **wind profiler** data, observations from **surface stations**, and volume data of **reflectivity** and the **radial wind** component from the German **radar** network. In addition, a **latent heat nudging** (LHN, Stephan et al., 2008) assimilates radar-derived precipitation rates from the European radars within the model domain during the 1-hourly forward integrations of the model between the analysis steps and during the first about 30 minutes of the operational forecasts. This is complemented by a sea surface temperature analysis once per day and a snow depth analysis every six hours.

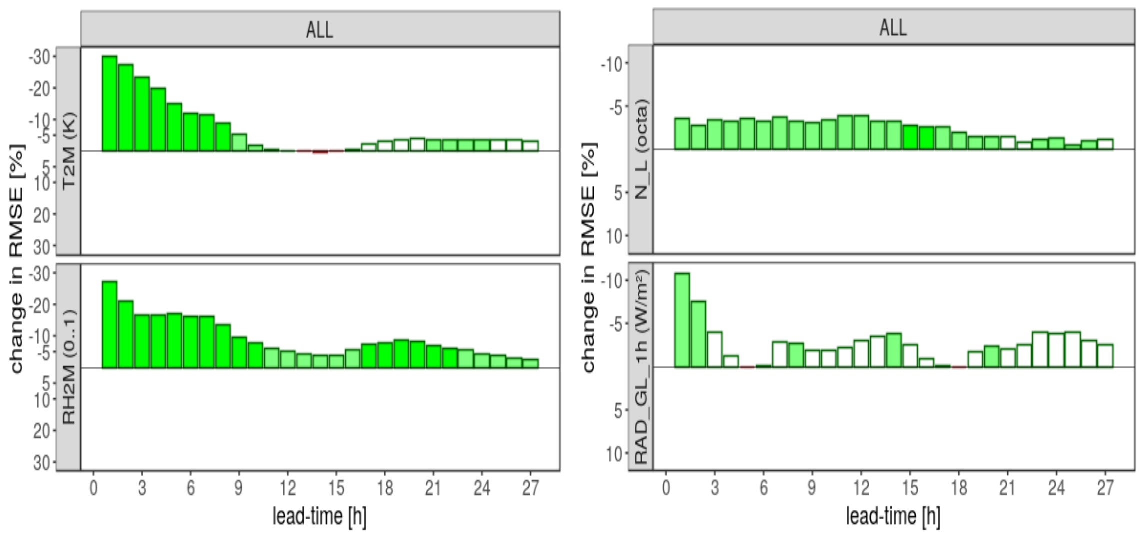

Fig 1: Change (in [%]) of root mean square error of deterministic ICON-D2 forecasts as a result of the additional assimilation of 2-m temperature and 2-m relative humidity observa-

tions from SYNOP surface stations. Scores are plotted as a function of forecast lead time verified against SYNOP data for 3 – 16 Nov. 2020. Green bars indicate improvement (i.e. reduction of errors), colour shading indicates statistical confidence. Upper left panel: 2-m temperature (T2M); lower left: 2-m relative humidity; upper right: low-level cloud cover (N_L); lower right: 1-hourly sum of global radiation (RAD_GL_1h). Note the different scales in left and right panels.

Even though this basic system remains the same, several improvements have been implemented for ICON-D2:

- While 2-m humidity observations are being assimilated in the global EnVar data assimilation system for ICON, neither 2-m temperature nor 2-m humidity data from surface synoptic stations have been used in the LETKF for COSMO-D2. For ICON-D2 however, it has become feasible to introduce the assimilation of both **2-m temperature** and **2-m relative humidity observations** in the LETKF. As illustrated in Fig. 1, this leads to an initially very large error reduction in 2-m temperature and humidity forecasts (up to 20 % of root mean square error in summer and 30 % in winter) which decreases rapidly during the first 9 hours, but small improvements prevail throughout the 27-hour forecast. A long-lived positive impact is also found on the prediction of wintertime low stratus (particularly in 12-UTC forecast runs). The impact on other parameters is small. For EPS forecasts, the forecast errors are reduced similarly, and the spread-skill ratio tends to be reduced as well for cloud cover, 2-m temperature and humidity. A station-dependent bias correction scheme for further enhancing the benefit from these observations is in preparation.
- In contrast to COSMO-D2, a less restrictive upper boundary relaxation zone in ICON-D2 allows for assimilation of the wind and temperature observations located between 300 and **200 hPa**. This leads to a small positive impact on forecast parameters in the upper troposphere, including high cloud cover.
- For COSMO-D2, the analysis increments (i.e. analysis correction) have been hydrostatically balanced and then added directly to the first guess (i.e. the model state prior to the analysis, i.e. a 1-hour forecast in KENDA) in order to obtain the initial conditions for the model forecast runs. For ICON-D2 instead, an **Incremental Analysis Update** (**IAU**) is applied without prior hydrostatic balancing. For this purpose, the model integration following a LETKF analysis step starts from the previous 55-minute forecast state, and the analysis increments are added to the model fields incrementally at each model time step during the initial 10-minute time window (i.e. from -5 to +5 min. relative to the analysis time). Subsequently, the model run continues as a free forecast (for 30 – 60 minutes with LHN). Additionally, a short IAU run from -5 to 0 min is carried out for purely diagnostic purposes to provide an ‘initialized’ analysis state.

The IAU is found to provide far better balanced initial conditions than the formerly used hydrostatic balancing, resulting in greatly reduced spin-up and spurious gravity wave noise during the first hour of the forecast.

- Due to the different model grid and dynamic response of the model to the temperature and humidity increments imposed by **latent heat nudging** (**LHN**), the LHN scheme had to be revisited and re-tuned for ICON-D2. One major change relates to the grid points where precipitation is observed by radar but not simulated by the model. While a search for a suitable grid point in the neighbourhood was performed in COSMO-D2 in order to derive appropriate heating rates, a pre-defined vertical heating profile is deployed in ICON-D2.

- For the global ICON and for ICON-EU, a variational soil moisture analysis (SMA) adjusts the soil moisture depending on daytime 2-m temperature forecast biases. In weather situations (particularly in spring and summer) with strong sensitivity to surface fluxes, this often improves these fluxes and reduces daytime 2-m temperature forecast errors. In contrast, the soil moisture in COSMO-D2 has been run unconstrained throughout the year(s) in the data assimilation cycle, and this could potentially lead e.g. to excessive drying below the plant wilting point in the summer. In order to improve on this and make the ICON-D2 system more robust against weather anomalies, the soil moisture index (SMI, i.e. a kind of relative humidity in the soil) at each grid is relaxed to the SMI of an appropriate grid point (possibly with the same soil type) of the ICON-EU. As the soil and surface layer parameterizations are very similar in ICON-EU, this **soil moisture nudging** allows ICON-D2 for taking the benefits of the SMA in an indirect but simple way.

  The SMA is well suited to correct for errors with long time scales, e.g. an over- or underestimated gradual drying of the soil in the summer months. However, it is not capable of reacting fast enough to certain temporary, weather dependent forecast errors, in particular large cold biases in sudden warm air outbreaks in winter and spring. This can be addressed by an automatic adjustment of model parameters relevant for bare soil and plant evaporation, namely the **stomata resistance**, the **unsealed fraction of the surface,** and, in the transitional seasons, additionally **the leaf area index**. Unlike for COSMO-DE, this scheme has been used for ICON since 2019 (see ICON Modification Note for 30.07.2019, available in German only), and is now also deployed for ICON-D2.

- In the **sea surface temperature** (SST) analysis, local ship and buoy data are used to adjust a first guess field on small scales. While the (interpolated) SST analysis from the global 13-km ICON model has been used as a first guess for COSMO-D2, the 6-km resolution OSTIA analysis (Donlon et al., 2012) is deployed for ICON-D2. Likewise, the sea ice information from BSH has been replaced by that from OSTIA.

References:

Donlon CJ, Martin M, Stark J, Roberts-Jones J, Fiedler E, Wimmer W, 2012, The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*. **116**: 140–158, doi: 10.1016/j.rse.2010.10.017.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periáñez, R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA), *Q. J. R. Meteorol. Soc.***142**, 1453-1472. https://doi.org/10.1002/qj.2748

Stephan, K., S. Klink, and C. Schraff, 2008: Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Q. J. R. Meteorol. Soc*. **134**, 1315-1326. https://doi.org/10.1002/qj.269

Changes specifically affecting the Ensemble Prediction systems (ICON-D2-EPS versus COSMO-D2-EPS)

The transition to ICON-D2-EPS does not imply major changes in the ensemble generation process compared to COSMO-D2-EPS. The forecast variability is represented by 20 EPS members based on initial conditions perturbations of KENDA (now with ICON-D2), perturbations of boundary conditions using forecasts of ICON-EU-EPS as before, and a randomized selection of default values and pre-defined perturbed values for certain parameters of the physics schemes.

However, there are changes in the set of perturbed parameters and some other technical changes and features for ICON-D2-EPS and its probabilistic products:

- To ensure qualitative analogy for the parameter perturbations of the two systems, most parameters are treated in the same way in ICON-D2-EPS as in COSMO-D2-EPS (possibly with different perturbed values optimized for ICON-D2). Certain parameters had to be replaced because some of the physics schemes differ between COSMO and ICON, and a few have been added in order to slightly increase the forecast variability (ensemble spread) for variables with particularly large underdispersion. Details of the perturbed parameters are listed in the *"DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System"* cited above.

- Twenty EPS members are produced every 3 hours (00 UTC, 03 UTC, …) up to 27 hours of lead time (45 hours for the 03 UTC run). The **output grid for EPS members is the native ICON-D2 icosahedral grid**. For a few variables, the output is on the former regular COSMO-D2 grid, but only as an interim solution for the visualization in NinJo.

- There is no significant change in the portfolio of probabilistic products going from COSMO-D2-EPS to ICON-D2-EPS. The GRIB2-shortname for snow-related products changes from SNOW_GSP to TOT_SNOW without changes to the products themselves. The output frequency of some products for periods of 6 and 12 hours has been increased.

- The **output grid for the probabilistic products of ICON-D2-EPS is the regular COSMO-D2 grid** (including the upscaling to 10 x 10 grid boxes for specific variables)**.** There is currently no output on the native grid.

- Technically, the generation of probabilistic products has been harmonized for ICON-EPS and ICON-D2-EPS, i.e. the identical software is used. The independent module for product generation of COSMO-D2-EPS is deprecated.

Weather parameters (WW) for ICON-D2

The approach for the generation of weather parameters (WW) for ICON-D2 has been adopted from COSMO-D2. A tuning of the criteria for the distinction between rain (W=6) and showers (W=8) adjusts the method to the characteristics of ICON-D2:

- The lower limit for precipitation is increased from 0.015 mm/h to 0.03 mm/h to reduce some slight noise in the field.

- To diagnose W=8 at a given grid point, the cloud cover of medium range clouds has to be below 85% in at least one grid box of the surrounding area within a distance of six grid boxes on the native grid. This value used to be 99.9% in COSMO-D2 to counteract an overestimation of mid-level cloud cover, which is no longer present in ICON-D2.

- Currently, the output is available on both, the native ICON-D2 grid and the regular COSMO-D2 grid.

Impact on forecast quality

Verification of deterministic forecasts

In the following, verification scores against SYNOP observations, radiosonde data and radar data are shown to demonstrate the improvement in forecast quality achieved with ICON-D2. The results are taken from the parallel suite running since late November 2019. Note that the assimilation of 2-m temperature and humidity data is not yet included in these results because the development of this feature has been completed only recently. This implies that the actual forecast quality improvement of 2-m temperature and humidity will be even larger than shown in the following figures, particularly during the first six forecast hours.

Selected continuous scores against SYNOP observations are depicted in Figs. 2-4 for March, July and November 2020, respectively. Generally, it can be noted that the systematic errors of ICON-D2 are much smaller than for COSMO-D2 and show less variability between different months. The forecast quality of ICON-D2 is thus more robust than it was for COSMO-D2, and the month-to-month variability in the amount of improvement is dominated by the quality fluctuations of COSMO-D2. March 2020 (Fig. 2) represents a month in which COSMO-D2 had particularly forecast errors for 2-m temperature and humidity. During daytime, the error reduction achieved by ICON-D2 exceeds a factor of 1.5 in this month. Typically, the error reduction achieved for these quantities is about 20%, as shown exemplarily for July and November in Figs. 3 and 4. For 10-m wind speed and direction and global radiation, error reductions are typically about 10% with relatively small month-to-month variability.

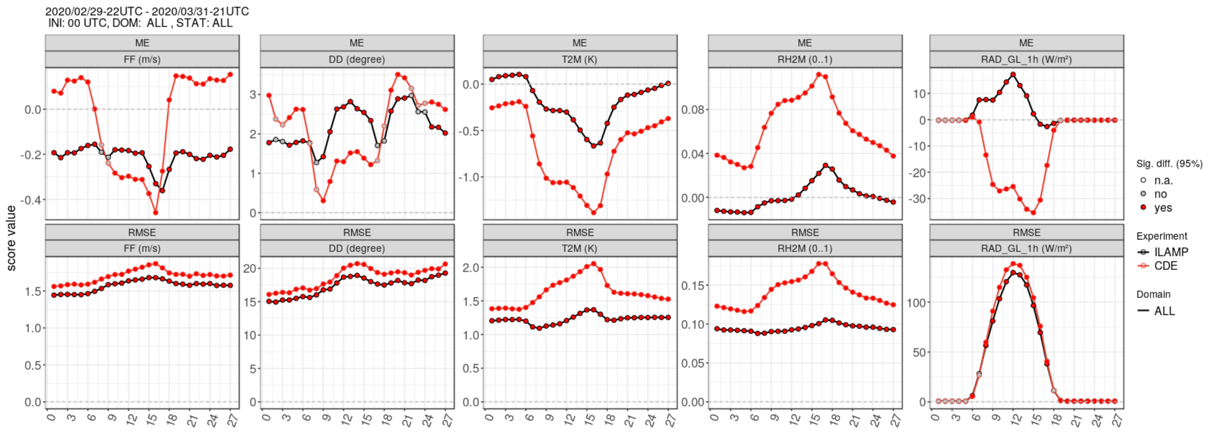

Fig. 2: Mean error (upper row) and root mean square error (lower row) of deterministic forecasts as a function of forecast lead time for the months March 2020. The verification is shown for 00-UTC forecast runs for 5 variables observed by SYNOP surface stations and ordered by column: 10-m wind speed (FF), 10-m wind direction (DD), 2-m temperature (T2M), 2-m relative humidity (RH2M), 1-hourly sum of global radiation (RAD_GL_1h). Black line: ICON-D2; red line: COSMO-D2. Statistical significance is marked by filled circles ("yes" in red, "no" in gray).

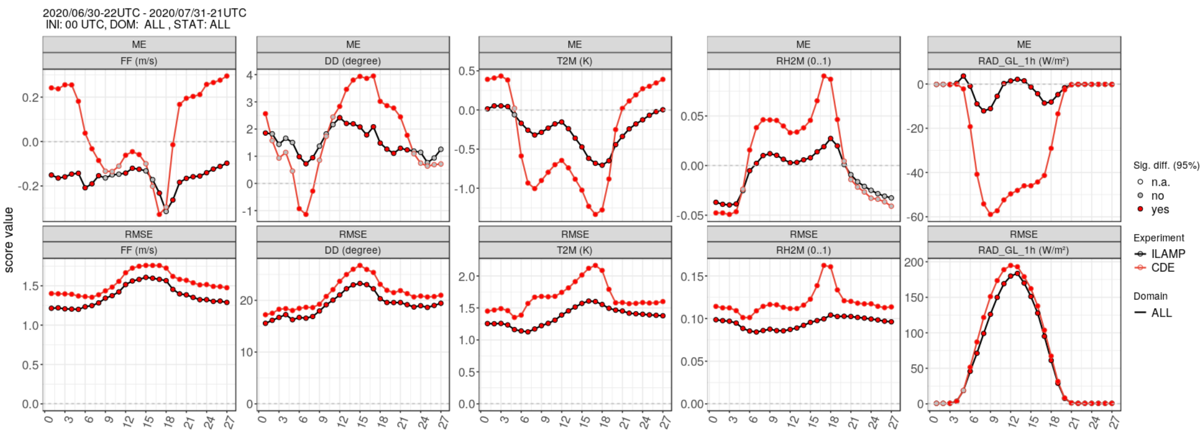

Fig. 3: Same as Fig. 2, but for July 2020.

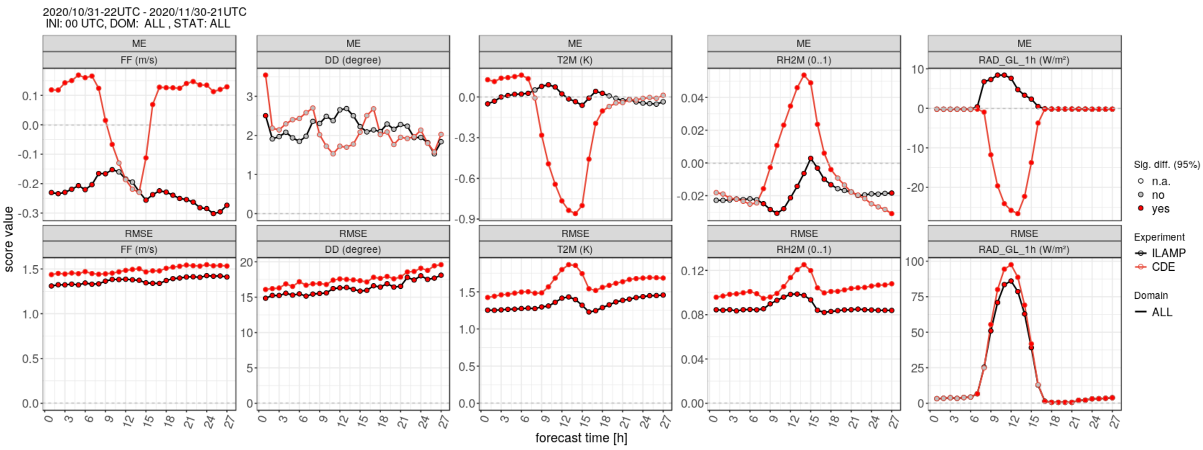

Fig. 4: Same as Fig. 2, but for November 2020.

Categorical scores against SYNOP observations are shown in Figs. 5 and 6 for July 2020, for wind gusts and cloud cover, respectively. For the wind gusts, it can be seen that the Equitable Threat Score (ETS) of ICON-D2 is better than for COSMO-D2 for all threshold values, with particularly large improvements for the weakest class (5 m/s), for which COSMO-D2 has a large positive Frequency Bias (FBI) during nighttime. For the strongest class (25 m/s), it needs to be mentioned that the statistical significance of the results is limited due to a relatively small number of events. For cloud cover (Fig. 6), comparatively small quality changes are found, but there is a pronounced shift from mid-level cloud cover to low-level cloud cover when comparing COSMO-D2 with ICON-D2. Besides changes in the cloud cover diagnostic, this is related to the fact that the shallow convection scheme tends to be too active in COSMO-D2, transporting too much humidity from the boundary layer into the middle troposphere on convectively active days. This tendency has been reduced in ICON-D2 by several tuning measures such as reducing the maximum convection depth allowed for shallow convection. For 2-octa low-level cloud cover (3rd column), a particularly large change in FBI can be seen, replacing an underprediction in COSMO-D2 by an overprediction in ICON-D2. To reduce the latter, a further tuning adjustment of the cloud cover diagnostic is in preparation and will be introduced in the first regular operational upgrade of ICON-D2, which is envisaged for spring 2021.

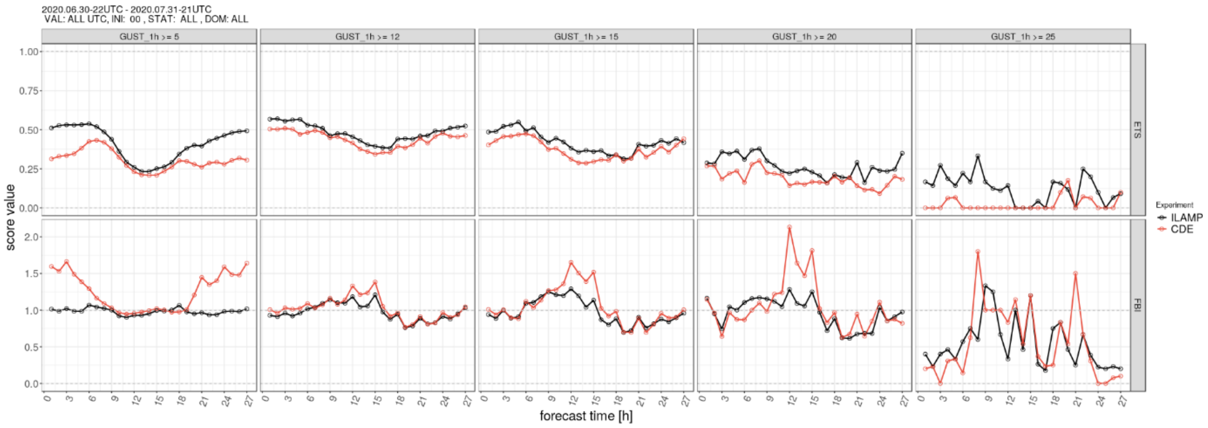

Fig 5: Equitable Threat Score (ETS, upper row) and Frequency Bias (FBI, lower row) of wind gusts for the 00-UTC deterministic forecast runs as a function of forecast lead time for July 2020. From left to right, the columns refer to gust thresholds of 5, 12, 15, 20, 25 m/s, respectively. Black line: ICON-D2; red line: COSMO-D2

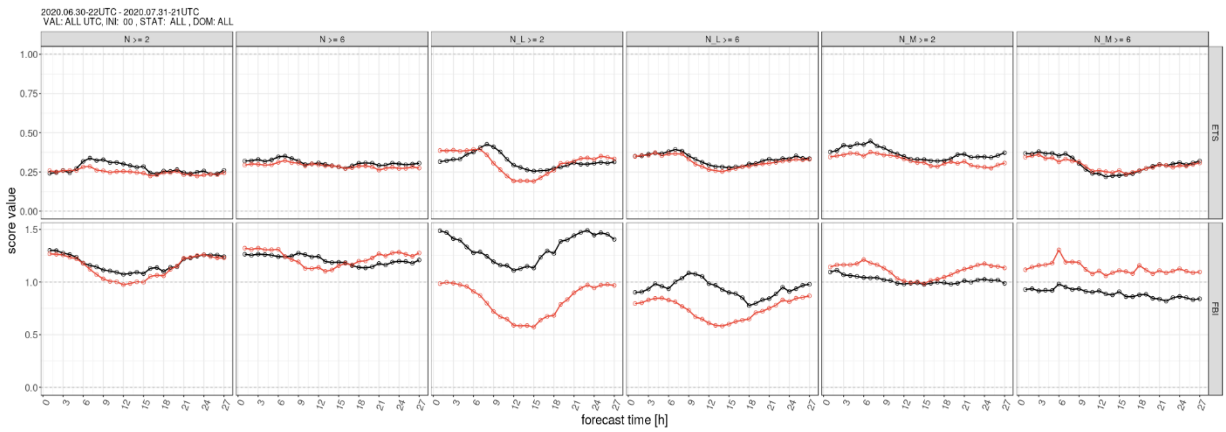

Fig 6: Same as Fig. 5, but for total cloud cover (two left columns), low-level cloud cover (two middle columns) and mid-level cloud cover (two right columns). For each cloud cover type, thresholds of 2 and 6 octas are shown in left and right panels, respectively. Black line: ICON-D2; red line: COSMO-D2

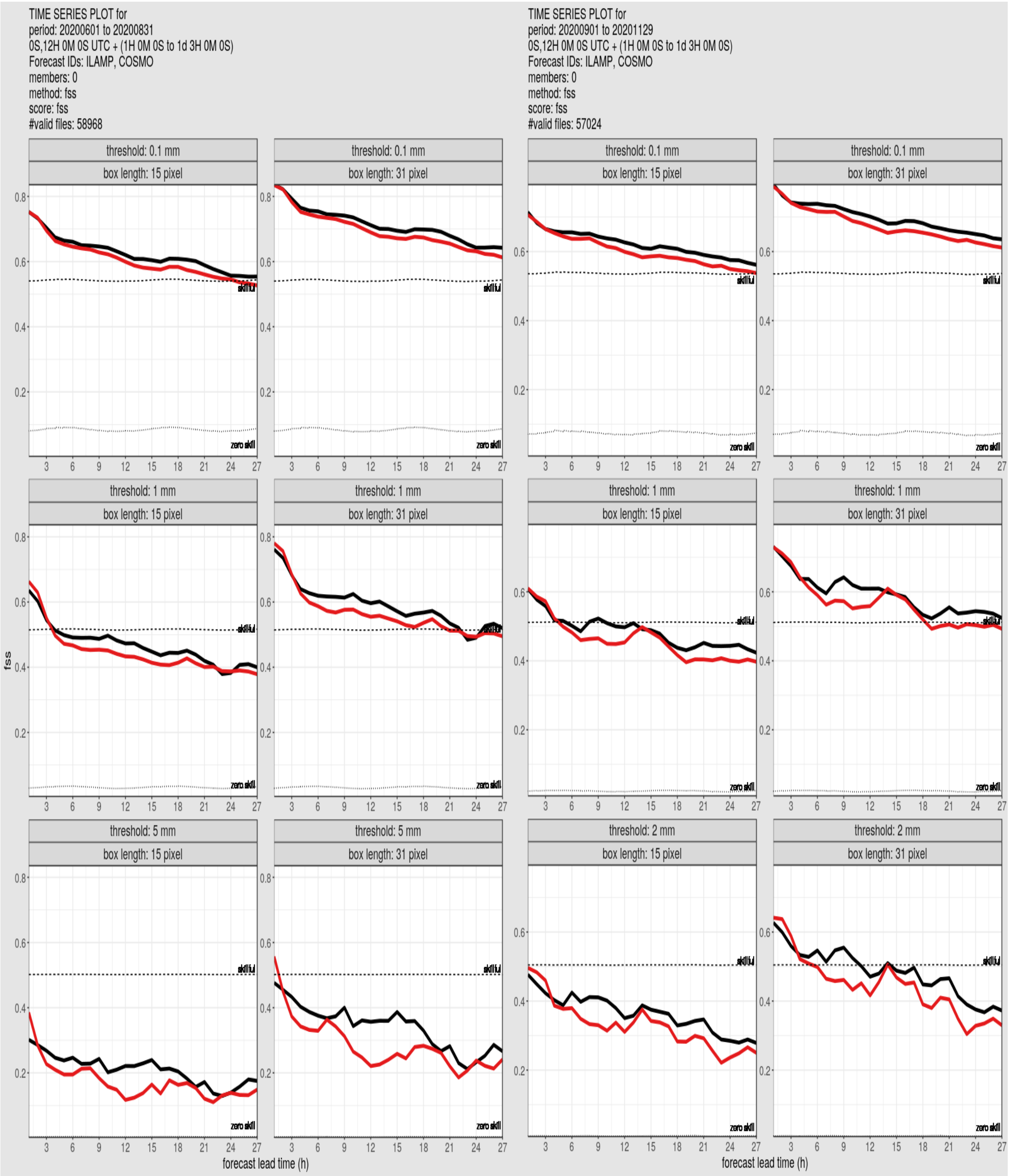

Fig. 7: Fraction Skill Score (FSS) of ICON-D2 (black lines) and COSMO-D2 (red lines) against radar-derived 1-hourly sums of precipitation for summer (June-August, two left columns) and autumn (September-November, two right columns). For each season, the left and right columns refer to box sizes of 15 and 31 model grid boxes, corresponding to about 30 km and 60 km, respectively. Results are shown for precipitation rates of 0.1 mm/h (upper

row), 1 mm/h (middle row) and 2 / 5 mm/h (lower row for autumn / summer); different intensity classes are chosen in the lower row because precipitation rates above 5 mm/h are too rare outside the convective season for statistically robust results.

To address the forecast quality of precipitation, which is one of the most important forecast quantities of COSMO-D2 and ICON-D2, Fig. 7 displays the Fraction Skill Score (FSS) of both models against the German radar network. To enhance the statistical significance for higher precipitation intensities, averages over seasons (summer and autumn) rather than single months have been selected. For all intensity classes and both averaging box sizes selected here (about 30 and 60 km), it can be seen that ICON-D2 performs better than COSMO-D2 except for the first 2-3 forecast hours. The slight degradation during the initial hours is related to the above-mentioned retuning of the latent heat nudging (LHN), for which the decision was made to push ICON-D2 not as strongly towards observed (radar-derived) precipitation than it was done for COSMO-D2. It was found that this tuning provides better forecast quality during most of the usable forecast range (lead times ≥ 2 hours; due to the cut-off times needed for data assimilation, forecasts are available about 1.5 hours after their initialization time).

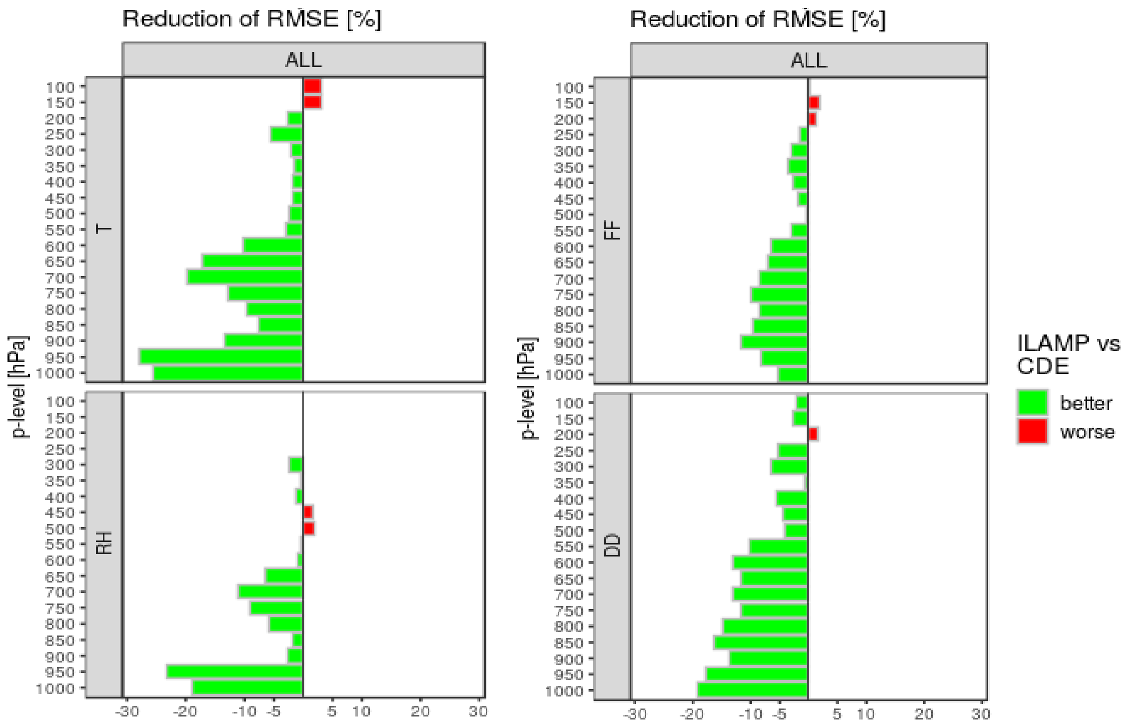

Fig. 8: Score card summarizing the relative change (in %) of forecast error against radiosonde measurements for temperature (upper left), relative humidity (lower left), wind speed (upper right) and wind direction (lower right) in July 2020. Green bars indicate an improvement of ICON-D2 with respect to COSMO-D2.

To complete the overview of the verification scores for the deterministic part of the system, Fig. 8 summarizes the relative change of forecast errors against radiosonde data for July 2020. For all depicted quantities (temperature, humidity and wind), substantial improvements are found throughout the lower and middle troposphere. At higher levels, the differences are smaller, but there is still a tendency for ICON-D2 being better than COSMO-D2.

Verification of EPS forecasts

In general, the scores of the ICON-D2-EPS improved compared to the COSMO-D2-EPS in accordance with the enhanced skill of the deterministic run of ICON-D2. This particularly affects the bias of the ensemble mean and its RMSE for many variables. The scores improve and the diurnal cycle of errors is reduced.

To point out some further characteristics of ICON-D2-EPS, Fig. 9 shows the Continuous ranked probability score (CRPS), the ensemble spread (standard deviation between the members), and the ratio between the spread and the standard deviation of the error of the EPS mean as a measure for the spread-skill-ration for selected variables. The forecasts of ICON-D2-EPS (black) and COSMO-D2-EPS (red) are verified against SYNOP observations over a six months period from June to November 2020.The CRPS (the lower, the better) shows a clear and significant improvement for all variables and almost all lead times.

The ensemble spread is smaller in ICON-D2-EPS with largest reduction in 10m wind gusts. Since ICON-D2-EPS as well as COSMO-D2-EPS tend to be underdispersive, such a spread reduction is not desirable at first glance. However, for most variables shown, the spread-skill-ratio as a measure of the amount of forecast variability captured by the EPS is increased thanks to the improved skill, i.e. the EPS is still, but less underdispersive in this respect. Since we choose the error standard deviation as a measure for skill, its improvement cannot be attributed to reduced overall biases of ICON-D2 and represents an additional improvement. However, the increased spread-skill ratio does not occur for wind gusts and there is a need for further improvement.

On the other hand, Fig. 10 and Fig. 11 show that there is an improvement in EPS forecast skill even though the spread is reduced. This is linked to a clearly decreased bias of 10m wind gusts in ICON-D2.

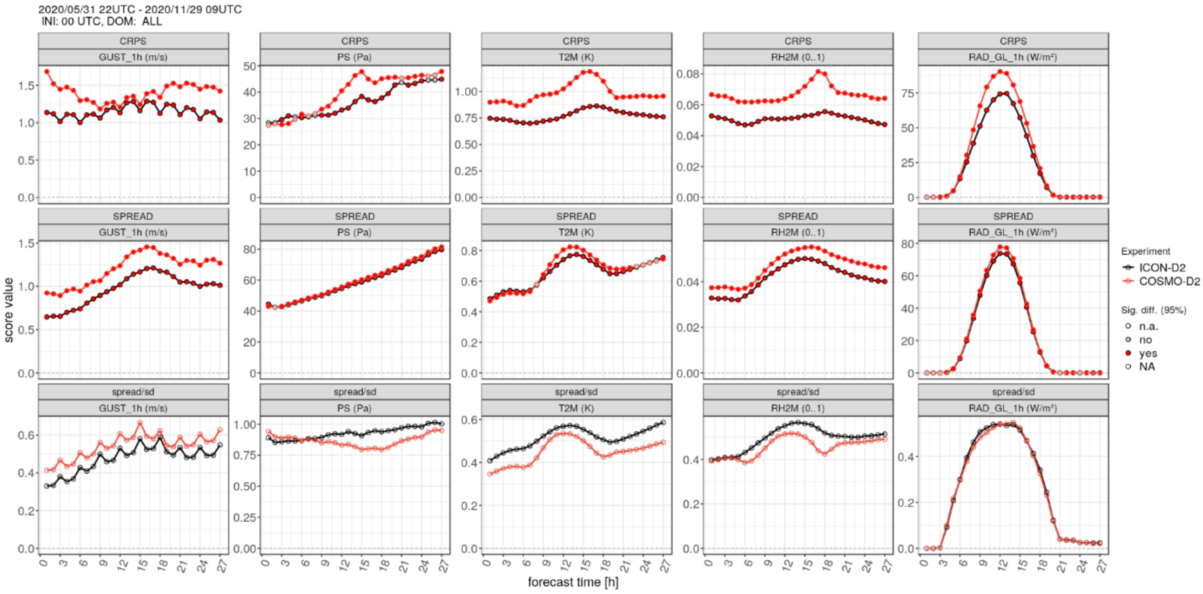

Fig 9: Continuous ranked probability score (CRPS, first row), ensemble spread (second row), and ratio of spread to error standard deviation (spread/sd, third row) as a function of lead time for five variables ordered by column: hourly 10m wind gusts (GUST_1h), surface pressure (PS), 2m temperature (T2M), 2m relative humidity (RH2M) and hourly global radiation (RAD_GL_1h). Lines for ICON-D2-EPS (without assimilation of 2-m temperature and humidity observations) in black, COSMO-D2-EPS in red. Statistical significance is marked by filled circles ("yes" in red, "no" in gray). The verification is for 00 UTC runs of the months June to November 2020. Observations for CRPS and spread-skill are from SYNOP stations.

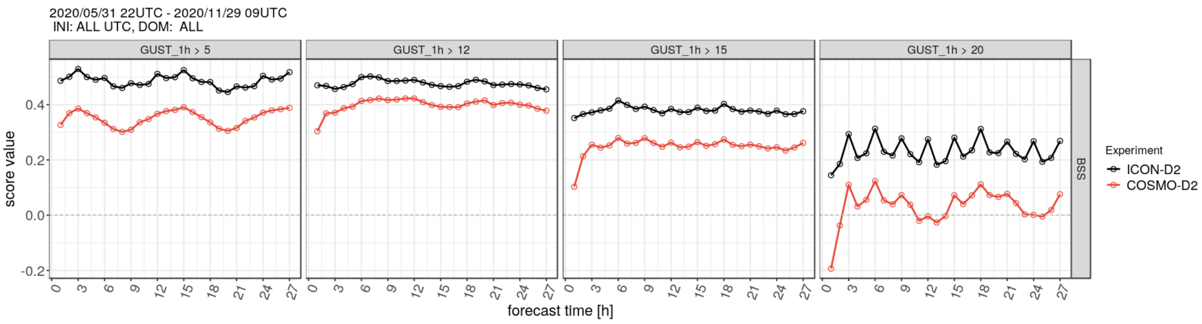

Figure 10: Brier skill score of 10m wind gusts as a function of lead time for four thresholds (5 m/s, 12 m/s, 15 m/s, 20 m7s) ordered by column. Lines for ICON-D2-EPS in black, COSMO-D2-EPS in red. The verification is averaged over 00 and 12 UTC runs of the months June to November 2020. Observations are from SYNOP stations.

The Brier Skill Score shows a clear improvement of around 10-15% by ICON-D2-EPS for the hourly 10m wind gusts above the thresholds 5, 12, 15, and 20m/s for all lead times. The most obvious improvement can be seen in the early forecast hours for the highest thresholds.

Fig. 11 shows the reliability diagrams for wind gusts and the same thresholds.

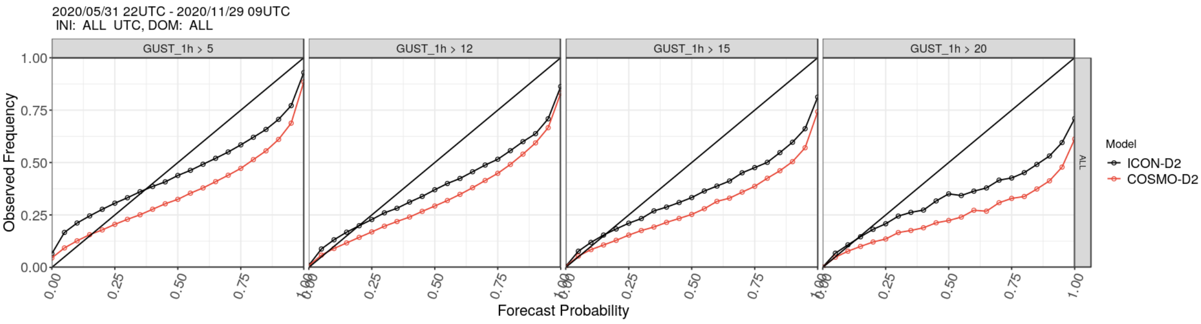

Figure 11: Reliability diagram of 10m wind gusts for four thresholds (5 m/s, 12 m/s, 15 m/s, 20 m7s) ordered by column. Lines for ICON-D2-EPS in black, COSMO-D2-EPS in red. The verification is averaged over all lead times of 00 and 12 UTC runs of the months June to November 2020. Observations are from SYNOP stations.

The comparison shows that the tendency to forecast too high event probabilities is reduced by ICON-D2-EPS for all thresholds, but only at the lowest threshold at the expense of too many low probabilities. Overall, the skill of wind gusts EPS forecasts is improved even though its EPS spread is reduced.

Contact

General contact address:
Günther Zängl, FE13, Tel.: (069) 8062 2728, guenther.zaengl@dwd.de

For questions regarding the data assimilation, please contact
Christoph Schraff, FE12, Tel: (069) 8062 2725, christoph.schraff@dwd.de

For questions regarding the ICON-D2-EPS in general, please contact
Christoph Gebhardt, FE15, Tel: (069) 8062 2743, christoph.gebhardt@dwd.de

For questions regarding the probabilistic products of ICON-D2-EPS, please contact
Chiara Marsigli, FE15, Tel: (069) 8062 2741, chiara.marsigli@dwd.de