

Statistical post-processing of ensemble forecasts at the Belgian met service

Jonathan Demaeyer, Stéphane Vannitsem (both Royal Meteorological Institute of Belgium & EUMETNET), Bert Van Schaeybroeck (Royal Meteorological Institute of Belgium)

The new system for post-processing ECMWF ensemble forecasts at the stations of the Royal Meteorological Institute (RMI) of Belgium was described previously in a short newsletter article (Vannitsem & Demaeyer, 2020). This system has now been operational since the summer of 2020 and we provide a description of its functionality and a preliminary analysis of its added value.

Principles

Statistical post-processing of ensemble weather forecasts has become an essential step in the forecasting chain as it enables the correction of biases and uncertainty estimates of ensembles (Gneiting, 2014). In June 2020, the Royal Meteorological Institute of Belgium, an ECMWF Member State, launched an operational statistical post-processing suite based on ECMWF medium-range ensemble forecasts in 11 reference synoptic stations, with the goal of improving its forecasting chain.

More precisely, the purpose of this new system is to provide forecasts closer to the typical values observed in these stations. Indeed, while ECMWF forecasts have in general good skill in the centre of the country, the temperature forecasts for the seaside region to the north and for the hilly forest region in the south are commonly known by forecasters to display notable biases. In addition, in the southern region, the wind gust variable is also known to be problematic, hampering the assessment of storm intensities and the release of accurate warnings.

The algorithm selected to perform the correction of the weather forecasts for the minimum and maximum temperature and for wind gusts is a linear member-by-member (MBM) Model Output Statistics (MOS) system, post-processing each member of the ECMWF ensemble (Van Schaeybroeck & Vannitsem, 2015). This method consists in correcting the mean and variability of the ensemble members in line with the observed climatology. At the same time, it calibrates the ensemble spread such as to match, on average, the mean square error of the ensemble mean. The MBM method calibrates the ensemble forecasts based on the station observations by minimising the continuous ranked probability score

(CRPS). Slightly different configurations of the linear MBM approach are used, depending on the variable considered in order to optimise their reliability.

The forecast suite constitutes a proof-of-concept of research-to-operation implementation resulting from a fruitful collaboration between the different services of the institute.

Operational implementation

To generalise statistical post-processing for a large set of applications, a new Python library has been designed. For a smooth integration in the production chain, this library is then placed inside Docker platforms that are passed to the institute's team in charge of the operational duties.

The RMI post-processing application corrects four variables: temperature (T), minimum temperature (Tmin), maximum temperature (Tmax) and wind gusts. It does so for the 11 'canonical' Belgian synoptic observation stations mentioned above. The MBM MOS post-processing method uses ECMWF re-forecasts over the past 20 years and relates them to the corresponding past synoptic station observations. The ECMWF re-forecast products (Hagedorn, 2008) are issued twice a week (Monday and Thursday at 00 UTC). They consist of ensemble forecasts for the last 20 years at the same calendar date, with the newest version of the model available and with currently 10 ensemble members. A single predictor from the model is used. This means that, in essence, the relationship is obtained as a linear regression over a two-dimensional scatterplot.

The current post-processing applications at RMI are configured based on a window of five weeks of re-forecasts and the corresponding station observations to gather an optimal set of forecasts to evaluate the linear regression parameters. The five weeks cover the full set of available re-forecasts (<https://confluence.ecmwf.int/display/FUG/5.3+Model+Climates>). Twice a week, upon availability, the new re-forecasts are downloaded. Afterwards, the regression parameters are computed, ensuring that the statistical post-processing follows the seasonal variations. Once the relation between the forecasts and the observations is obtained, it is used to correct the ECMWF ensemble forecast

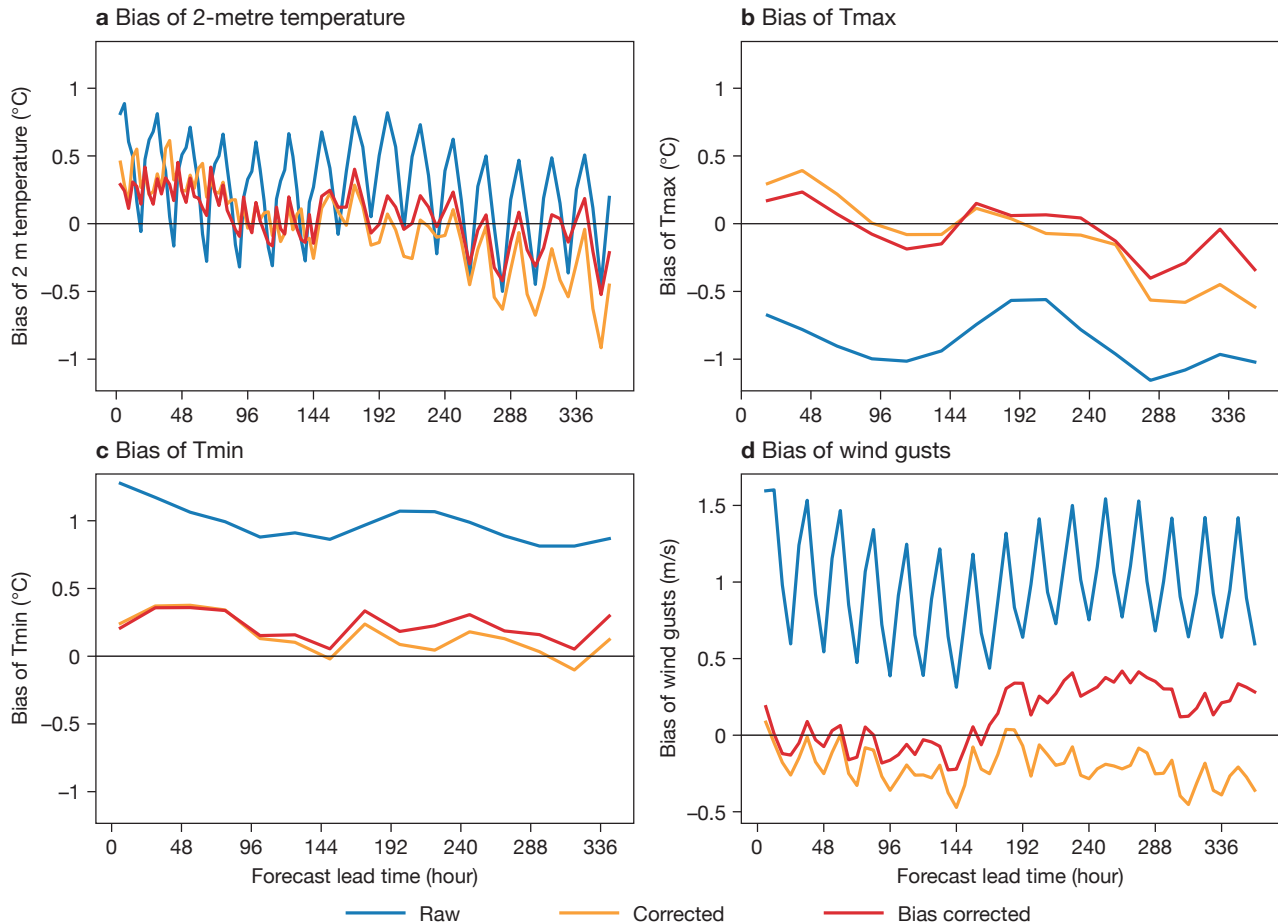


FIGURE 1 Averaged bias over all stations and over the JJAS months for (a) 2 m temperature, (b) Tmax, (c) Tmin and (d) wind gusts, as a function of the lead time. Two corrections are considered: one with a simple statistical correction of the bias of the ensemble ('Bias corrected'), and the full member-by-member correction minimising the CRPS ('Corrected').

issued daily at 00 UTC and transferred to the RMI through the dedicated dissemination channel.

The forecasts are being provided with a lead-time interval of 3 hours for the first 144 hours, and 6 hour-intervals afterwards. The re-forecasts, on the other hand, are provided 6-hourly over the whole time-range. Therefore, the statistical post-processing parameters are interpolated over the missing 3 hour-intervals in the first 144 hours. While experimental and based on the assumption that the post-processing parameters are smooth enough, this interpolation has proved to be skilful, as shown by the scores detailed in the following section.

Post-processing scores for the summer 2020

Statistics have been accumulated for the extended summer of 2020: June, July, August and September (JJAS). Some relevant scores are shown here to highlight the system's performance in providing improved forecasts. The results are obtained by averaging over all stations and all forecast days and months. Figure 1 shows the correction of the systematic

bias of the raw ensemble by the MBM adjustment (orange line). The red line is obtained using a simple correction of the systematic bias (no variability or spread correction) and therefore shows the smallest bias (except for the minimum temperature). The reason is that the MBM adjustment seeks to minimise the CRPS, at the expense of a less good correction of the bias.

The CRPS score, which measures the quality of the probabilistic information provided by the ensemble, is shown in Figures 2a,c and 3a,c. A smaller CRPS score corresponds to an improved ensemble forecast. Figures 2b,d and 3b,d show the reliability and resolution components of the CRPS (Hersbach, 2000).

The reliability measures the quality of the forecast probabilities with respect to the observed climatological frequencies. The lower the values of the reliability are, the better the calibration of the model's ensemble. The resolution, on the other hand, measures the information content of the ensemble forecast scheme. The higher it is, the more the ensemble forecasts improve upon the climatological forecasts (see Box A).

For all temperature variables (Figures 2 and 3), there is a substantial improvement due to the ensemble calibration

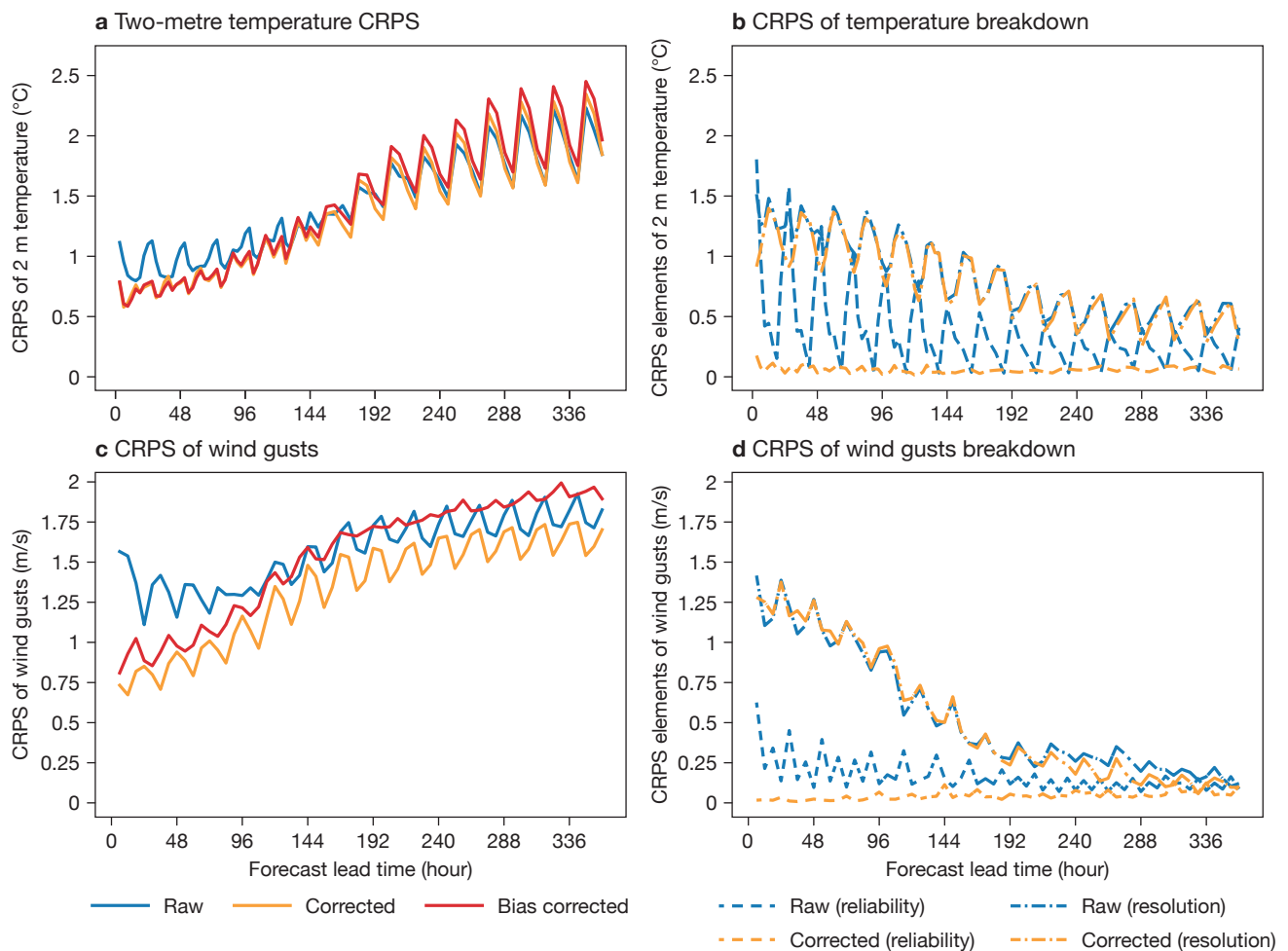


FIGURE 2 Averaged CRPS over all stations and over the JJAS months for (a) 2-metre temperature showing the full member-by-member correction minimising the CRPS ('Corrected') and a simple statistical correction of the bias of the ensemble ('Bias corrected'), with (b) the decomposition of the raw and MBM corrected ensemble CRPS for 2-metre temperature into the reliability and resolution components, (c) the corresponding wind gust variable corrections and (d) the corresponding breakdown of wind gust variable corrections, as a function of the lead time.

during the first half of the forecast period. Skill is gained at all lead times for the wind gust. While a simple bias correction (red line in Figure 2a) suffices to improve the CRPS of the 2-metre temperature to the same level of the MBM method, the variability correction is needed to further reduce the CRPS of the other variables. It indicates that for 2-metre temperature, the bias set aside, the ensemble is already well calibrated, in agreement with the results presented in Vannitsem & Hagedorn (2011).

The right-hand panels of Figures 2 and 3 show that, for each variable, the improvement of the CRPS score is due mainly to a decrease in the reliability contribution. As already indicated above, in the case of 2-metre temperature, the improvement of the reliability is mainly due to the correction of the statistical bias, while it also involves a contribution from the correction of the spread for the other variables. In contrast to the substantial ensemble reliability improvement, the resolution for the temperature and wind gust variables is not

a Breaking down the CRPS score

The CRPS score can be expressed by the reliability (Reli), the resolution (Resol) and the uncertainty (U) as follows:

$$\text{CRPS} = \text{Reli} - \text{Resol} + U$$

The forecast is the better the smaller the reliability value is – as close to zero as a possible – and the bigger the resolution. The ensemble system has positive resolution if it performs better than the climatological probabilistic forecast.

The uncertainty is the potential reliability for a forecast system based on the sample climatology.

substantially modified. The slight decrease in resolution for the minimum and maximum temperature can perhaps be partly circumvented if more predictors

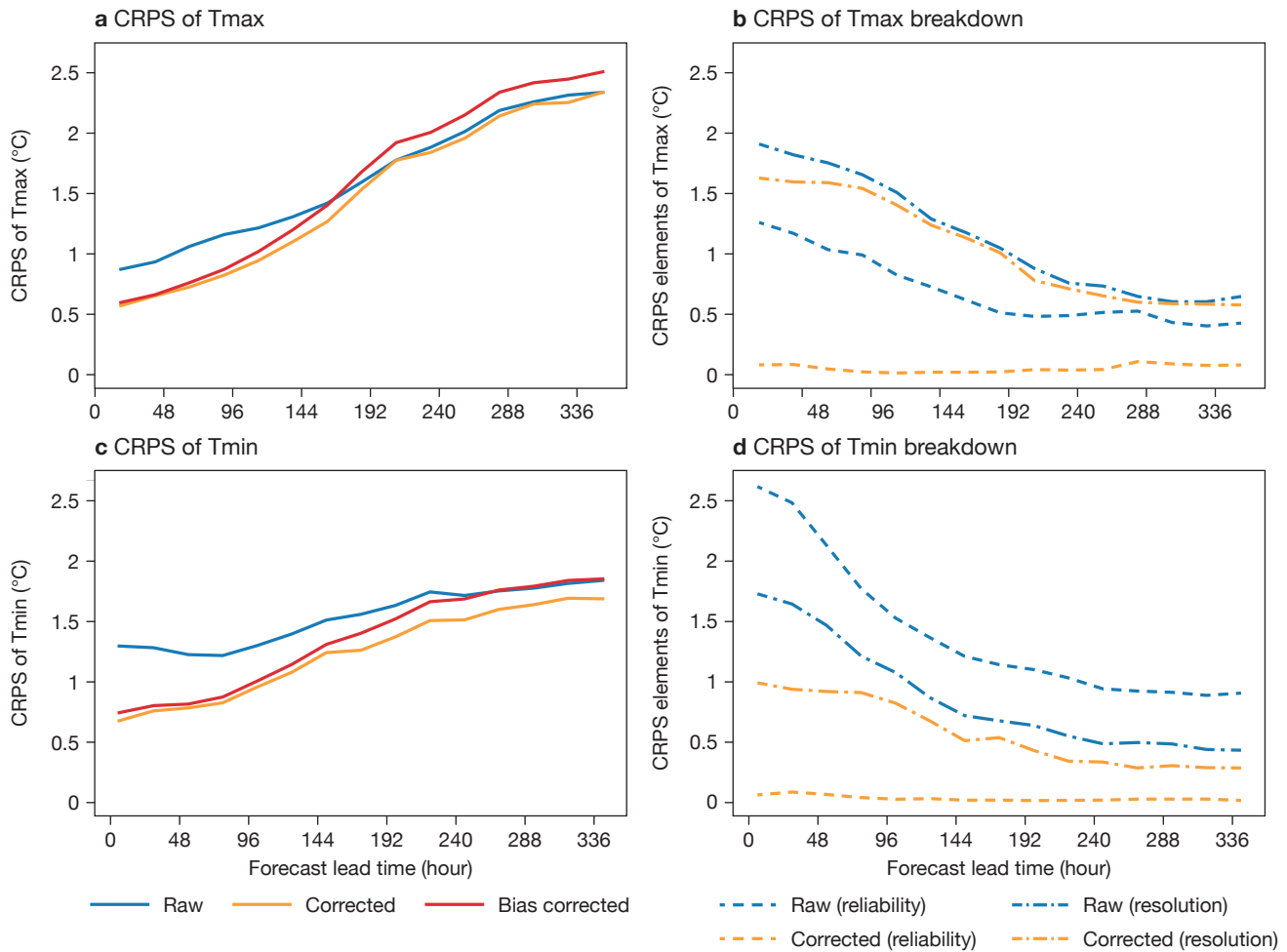


FIGURE 3 Same as Figure 2 but for the maximum and minimum temperatures, as a function of lead time.

are added to the correct variables. This aspect will be explored in the near future.

Finally, we show an example of a forecast at the Elsenborn station in Figure 4, a station located in the Ardennes region featuring large biases with respect to the other stations. The biases result from the complex orography, with an elevation at 570 m of the nearby region. Remarkably, one can see on this graph the correction of the bias, materialised by the shift of the ensemble mean, as well as the correction of the spread, depicted here by the 10–90% quantiles of the ensemble distribution shown with different levels of transparency. Notably, the correction of the bias for the wind gust induces a shift of the ensemble toward smaller values, and hence less extreme ones. We see also that while the observed minimum temperature is out of the raw ensemble distribution at the end of the forecast (Figure 4), the corrected ensemble distribution encompasses the observation, as expected.

The way forward

Following the first steps of the new post-processing programme of the RMI presented here, new developments are expected during the next couple of years:

- Currently only one predictor is used. Further improvements of the skill will be tested by introducing additional predictors.
- As mentioned above, for the moment only the midnight forecasts of ECMWF are post-processed, and not the forecasts issued at noon. Indeed, to correct the latter, one needs re-forecasts issued at noon as well, and these re-forecasts are not currently available at the Centre. One solution would be to post-process the noon forecasts with the parameters of the midnight forecasts. As a consequence, these parameters have to be shifted by 12 hours to match the diurnal cycle, which implies that the parameters are not optimal anymore. This could lead to a less optimal forecast correction and the impact of this shift must be carefully assessed. We nevertheless expect corrections that would justify the post-processing of these noon forecasts.
- Another development avenue is the implementation of a member-by-member post-processing application for gridded probabilistic forecasts. The core of the computation will again be done by the RMI post-processing library placed inside a Docker container. This post-processing application will either use the data

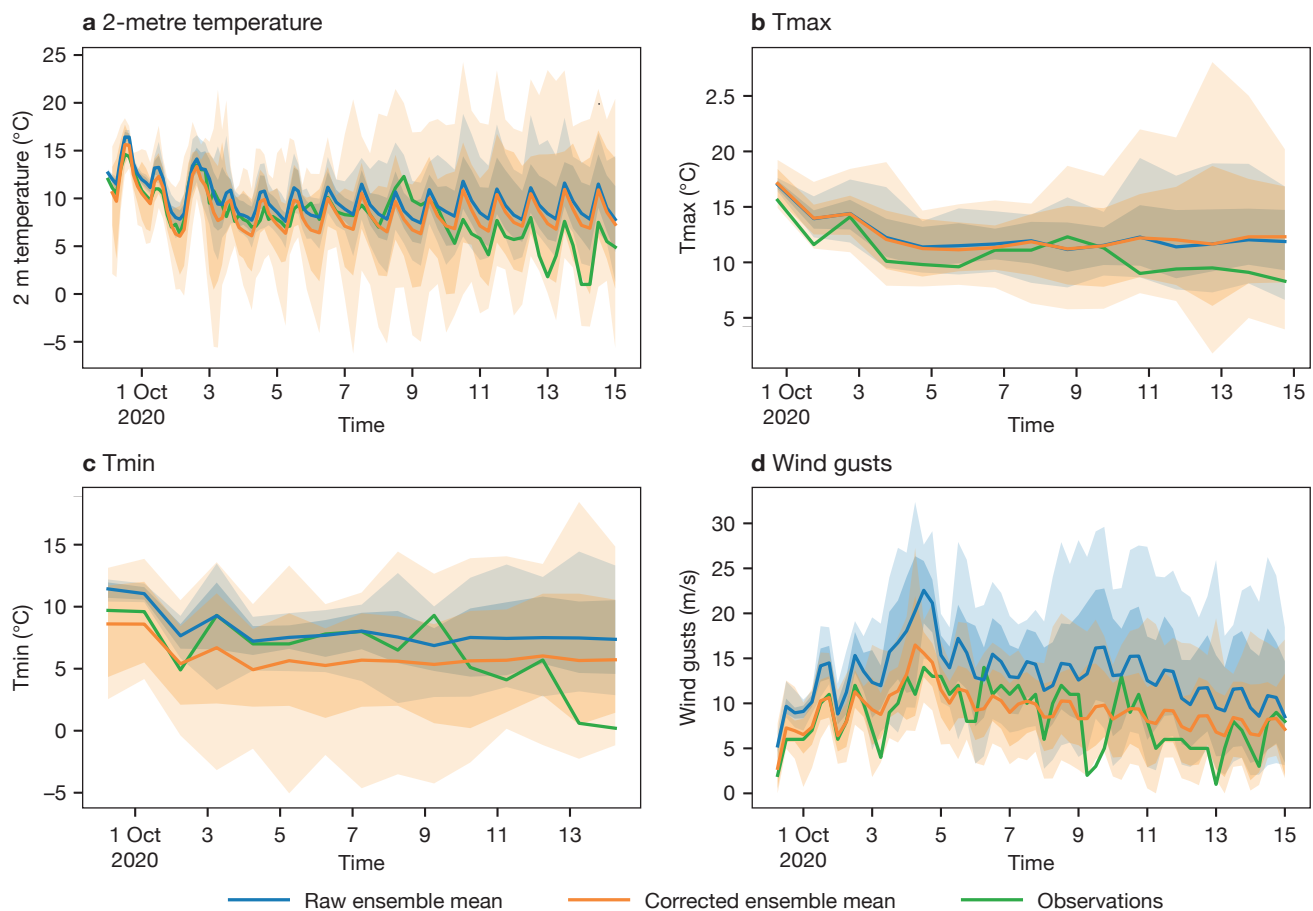


FIGURE 4 Example of a corrected forecast for each of the four variables: (a) 2-metre temperature, (b) maximum temperature (Tmax), (c) minimum temperature (Tmin) and (d) wind gusts, with the corresponding station observations for verification. The station being considered is Elsenborn and the raw ECMWF forecast was issued on 30 September 2020 at 00 UTC. The lower and upper lightly shaded areas represent respectively the 0 to 10% and the 90 to 100% quantiles. The darker shaded areas represent the 10 to 90% quantiles. Solid lines are the ensemble means.

of the RMI INCA system, which contains gridded combined observations, or of the ERA5-land reanalysis. The specific design of this application is still under discussion.

- Finally, longer developments involving more recent and sophisticated techniques are considered: spatial post-processing, machine learning, etc.

Discussions are also ongoing about the link of these activities with the EUMETNET post-processing benchmark, which is in preparation and for which an experimental proof-of-concept will be developed soon on the European Weather Cloud. This could foster the collaboration between national meteorological services on the development of a common platform for post-processing tools and best practices.

Further reading

Gneiting, T., 2014: Calibration of medium-range weather forecasts. *ECMWF Technical Memorandum No. 719*.

Hagedorn, R., 2008. Using the ECMWF reforecast dataset to calibrate EPS forecasts. *ECMWF Newsletter No. 117*, 8–13.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**, 559–570.

Vannitsem, S. & R. Hagedorn, 2011: Ensemble forecast post-processing over Belgium: comparison of deterministic-like and ensemble regression methods. *Meteorological Applications*, **18**, 94–104.

Vannitsem, S. & J. Demaeyer, 2020: Statistical post-processing of ECMWF forecasts at the Belgian met service, *ECMWF Newsletter No. 164*, 4–5.

Van Schaeybroeck, B. & S. Vannitsem, 2015: Ensemble post-processing using member-by-member approaches: theoretical aspects. *Quarterly Journal of the Royal Meteorological Society*, **141**, 807–818.

Wilks, D.S., 2011: *Statistical methods in the atmospheric sciences*, Volume 100, Academic Press.