1: Department of Meteorology, University of Reading | 2: University of Oklahoma, US | 3: Met Office, Exeter

A comparison of forecast bust characteristics for different numerical weather prediction models over the European region.

Kaustubh Mittal[1] | Robert S Plant[1] | Suzanne L Gray[1] | Dave Parsons[2] | David L Flack[3]

k.mittal@pgr.reading.ac.uk

Overview

This study is a part of the project that will investigate the mechanisms behind forecast busts. These are poor forecasts, that can occur due to unresolved processes in the numerical weather prediction model. Despite of huge progress in the models (Lillo and Parsons 2017), the influence of mesoscale convective processes over the upper-tropospheric atmospheric flows can contribute to what is known as a forecast bust. Here we compare bust metrics for four leading operation centres: ECMWF, UKMO, NCEP and JMA.

Defining 'Forecast Busts'

One way of defining forecast busts is in terms of errors in the prediction of the geopotential height at 500-hPa, when the day-6 high resolution forecast of European Z500 has a root-mean-square error (RMSE) greater than 60 m and an anomaly correlation coefficient (ACC) less than 40% (Rodwell et al., 2013).

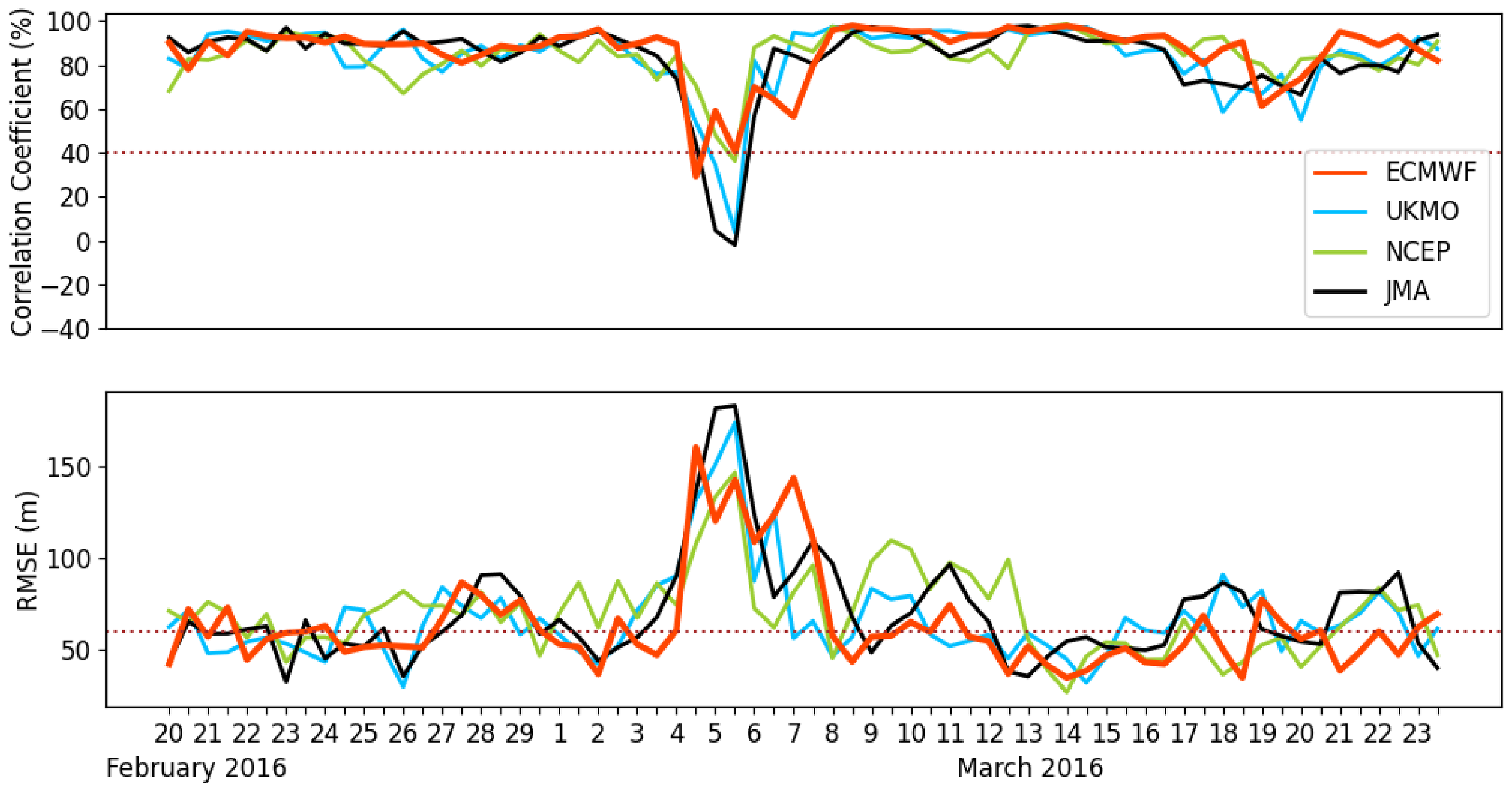

Fig. 1 Time series of (a) RMSE and (b) ACC for 500 hPa geopotential height for Feb-Mar 2016 including the bust case of 4th March 2016, 1200 UTC. The scores are calculated for 6-day forecasts over Europe (35°N–75°N, 12.5°W–42.5°E). ERA5 climatology from 1989-2022 is used to calculate the ACC.

An example bust case presented here (Fig 1 and Fig 2) is from March 2016. The mean values over the two-month period were 86% for ACC and 63m for RMSE, but for the HRES forecast from 31st August at 0000 UTC the scores were 38% for ACC and 150 m for RMSE (ECMWF).

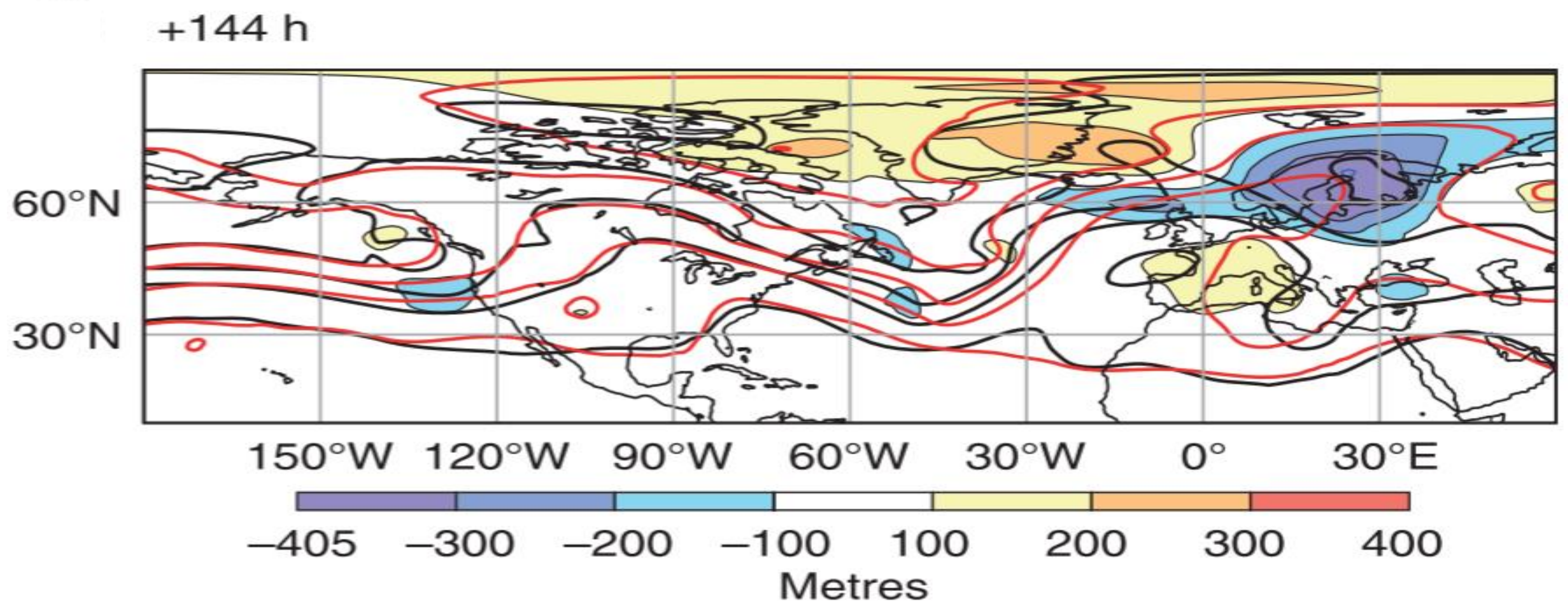

Fig. 2 Map of Z500 for the HRES forecast (black line), analysis (red line) and forecast error (shaded) on 4th March 2016.

ACC and RMSE

- Skill scores are generally used to quantify model performance in good or bad forecasts. For instance, fig 3. shows that majority of the ECMWF 6-day forecasts have high skill in both ACC and RMSE. There are 122 cases of bust forecasts.
- Since, RMSE provides limited discrimination compared to ACC, we will just use that as a criteria for defining a bust. We use the top and bottom 5th percentile as guides for very good and very poor model performance. Fig 4 shows that every model has improved in performance since 2009 in good as well as poor forecasts.

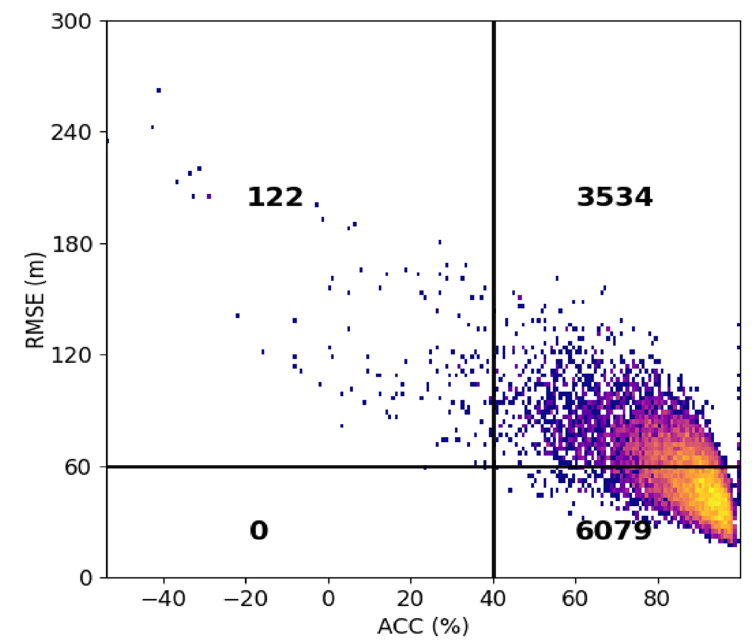

Fig. 3 ACC vs RMSE distribution of the ECMWF 6-day forecast of 500 hPa geopotential height from 2009 to 2023. The text refers to number of forecasts .in each quadrant.

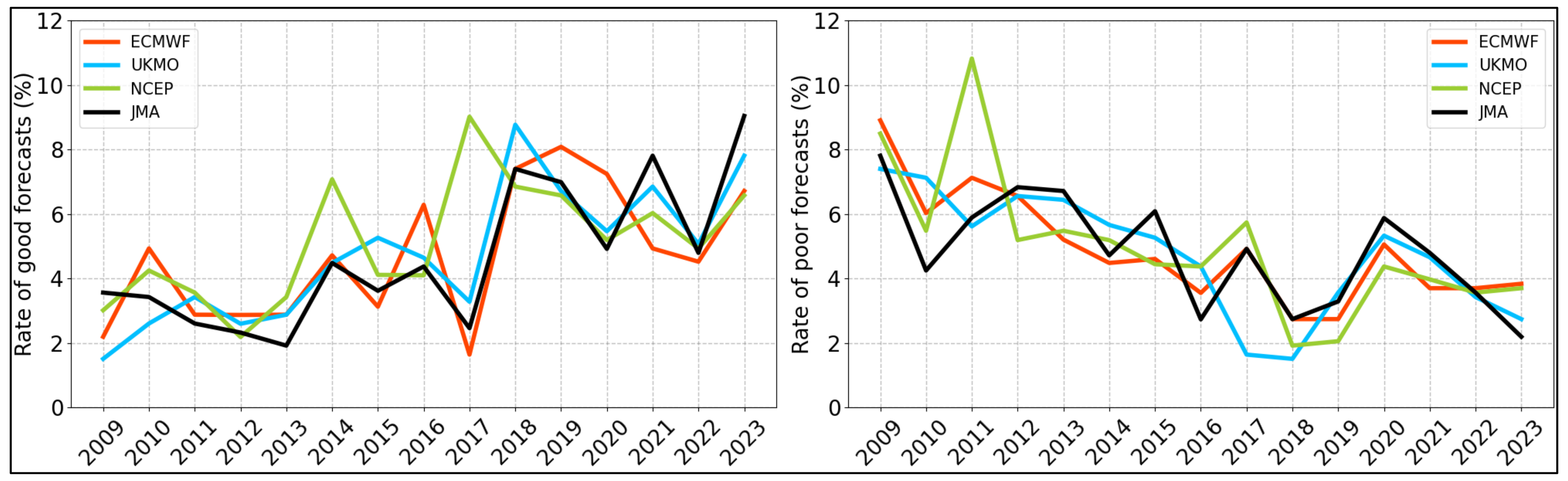

Fig. 4 Time series of (a) % of good forecasts and (b) % of poor forecasts for 500 hPa geopotential height for 2009-2023. It gives the frequency of forecasts which are above/below the top/bottom 5 percentile ACC.

- To compare models based on the frequency of forecasts becoming busts, we have tabulated the number of bust forecasts for each model (Table 1).
- To understand how often do bust forecasts coincide for every model, the table also states the common numbers for every model with respect to every other model.
- There are many cases where forecast busts don't coincide. This points out to the fact that there is variability between centres for same bust events.
- The forecast bust event of 4th March, 2016 (Fig 1) also shows that models behave differently before, during and after the bust event.
- Small bust cases for ECMWF might also reflects bias due to the usage of ERA5 climatology.

	ECMWF	UKMO	NCEP	JMA
ECMWF	122	35	50	52
UKMO	35	191	66	63
NCEP	50	66	286	85
JMA	52	63	85	271

Table. 1 Number of bust forecasts for each centre. The rows indicate number of busts for a centre given that the bust occurred in another centre (column wise).

Forecast Bust Metrics

To define forecast bust events, certain bust metrics are to be defined (Fig. 5). Different NWP models can be quantitatively compared based on these parameters.

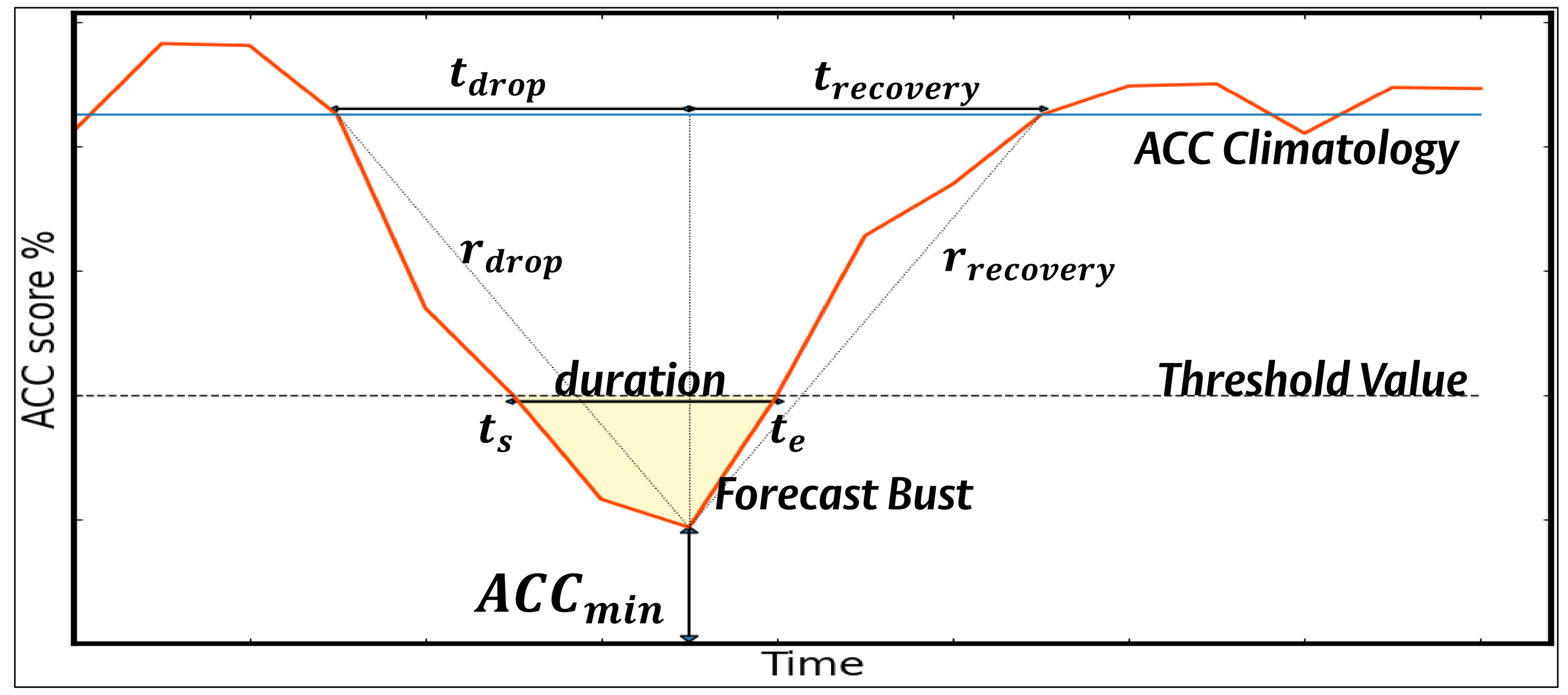

Fig. 5 Schematic of metrics used to define a forecast bust event. The figure shows the duration of a bust event which is defined by the time ACC remains below the threshold value, minimum ACC values during a bust event, start and end time of an event. Rate of drop and rate of recovery are calculated with reference to the ACC climatology, time taken by the ACC to drop to the minimum and the time taken by the ACC to rise back to the climatology.

Future Work

•How can we define a bust event in terms of its duration? What is the true end of a forecast bust event and how much time is required to differentiate between two consecutive events?

•What is the recovery time for a forecast, and does it depend on type of busts and model physics?

•Do different regimes/bust characteristics produce different types of busts?

Conclusions

Forecast busts can be defined based on just the ACC threshold as almost all times when ACC is below 40% the RMSE is always greater than 60m (Fig 3). Even though the rate of poor forecasts has decreased in the recent years, bust cases persists. There is a general disagreement in the models in total number of bust cases and the point of occurrence (Table. 1). This is an indicator that the models perform differently during a bust. Different models have different bust characteristics, which can be quantitatively summarised using the bust metrics.

References

- Lillo, S.P. and Parsons, D.B., 2017. Investigating the dynamics of error growth in ECMWF medium-range forecast busts. Quarterly Journal of the Royal Meteorological Society.
- Rodwell, M.J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez, A., Isaksen, L. and Källén, E., 2013. Characteristics of occasional poor medium-range weather forecasts for Europe. Bulletin of the American Meteorological Society.
- Magnusson, L., 2017. Diagnostic methods for understanding the origin of forecast errors. Quarterly Journal of the Royal Meteorological Society.

Acknowledgements

This work is a part of the project which is funded by the Advanced Frontier in Earth System Prediction (AFESP) research program and the Met Office through a CASE studentship.