[NPG-2017-45]
Response to Referee #2
November 22, 2017

We gratefully thank Referee #2 for the constructive comments. We have revised the paper accordingly. The point-by-point responses to the comments are detailed below.

The manuscript proposes an efficient method for choosing appropriate data selection criteria for new observing systems based on EFSOI. The usefulness of this approach is demonstrated with the assimilation of precipitation observations. The findings of the paper are interesting, it's very easy to read and well-written. It should be suitable for publications after addressing the following few remarks.

There is one issue that the authors should investigate a bit more. Usually, the number of beneficial observations should be slightly above 50%. When the number is much higher, I strongly suspect that the observations are correcting or compensating some model bias. Otherwise it's unlikely to achieve numbers up to 70%. Very low numbers likely indicate the opposite effect, i.e. that there is a model bias which prevents the effective use (which is also indicated by the plot of FSOI versus precipitating members). I think it would be very interesting to investigate this more and show more bias statistics (e.g. regional plots).

We agree with Referee #2 that the assimilation of precipitation observations may be correcting or compensating some model biases. Indeed, in the manuscript we show the 5-day forecast biases in Fig. 6j-l. Compared to their RMSE (Fig. 6a-c), the bias values are relatively small for 500-hPa u-wind and 700-hPa moisture, but the model does present considerable 500-hPa temperature bias. The precipitation assimilation generally reduces model biases for 500-hPa wind and temperature. Besides, we also show that, when measuring the forecast skill by the standard deviation of errors that does not take biases into account (Fig. 6g–i) (instead of by RMSE), although the improvement by precipitation assimilation becomes smaller, all the precipitation assimilation experiments are still better than CONTROL, and 1mR/24mR is still better than 24mR in all variables. We believe that this would be sufficient to show the superiority of 1mR/24mR and the usefulness of our EFSO methodology, despite of the existence of the model bias. The related discussion can be found in P.18, L.26–P.19, L.4. We revised the discussion regarding the model biases to better elaborate this aspect.

In addition, following Referee #2's suggestion, we show the regional bias plots (Fig. R1; c.f., Fig. 6j-l) and the regional (bias-free) standard deviation of errors (Fig. R2; c.f., Fig. 7) here. The results are more complicated in different regions, but we think the general conclusion remains as what we discussed above for the global verification: There are considerable model bias in the temperature field and the precipitation assimilation is correcting it, but in most variables and regions, the 1mR/24mR still outperforms the other experiments in the bias-free verification. We feel that these additional figures are not very essential for the paper and would be a distraction in the manuscript if included, so we would not add these figures in the manuscript.

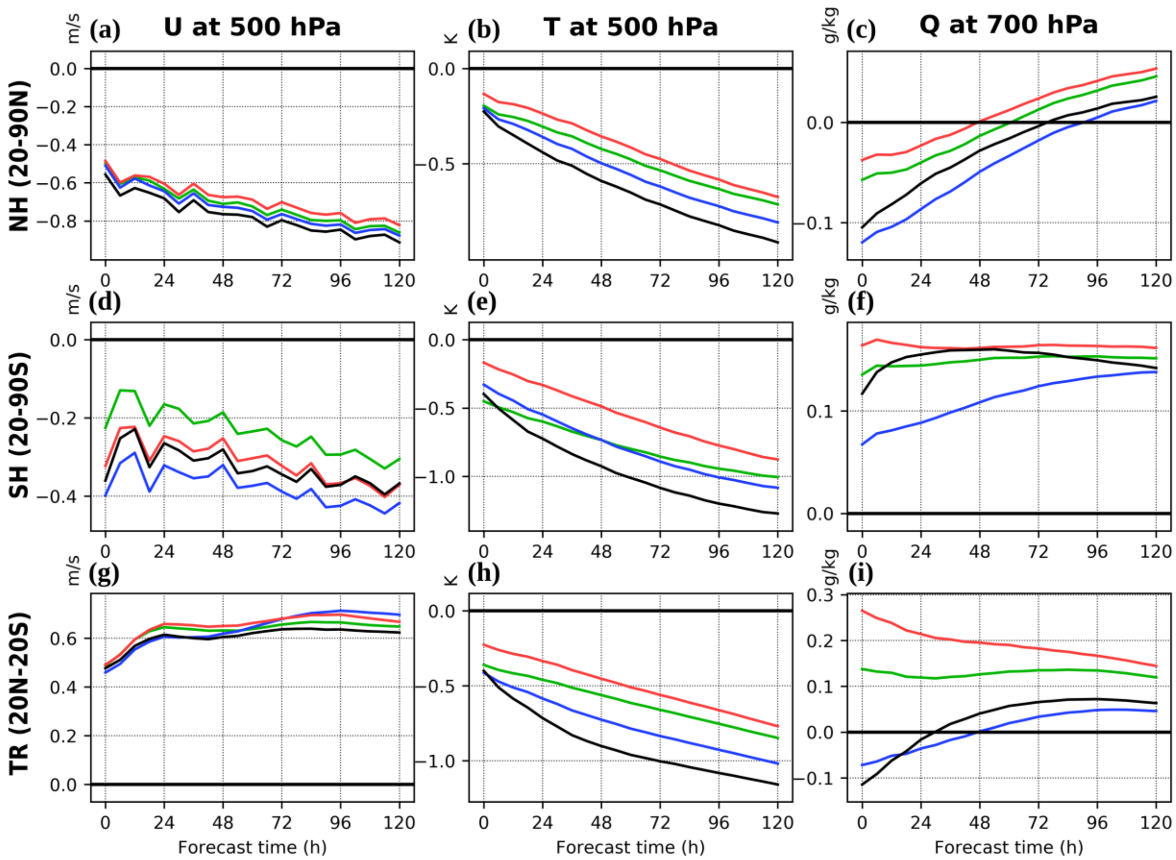

Figure R1: Similar to Fig. 6j-l, but for regional biases during the 5-day forecasts of (first column) 500-hPa u-wind, (second column) 500-hPa temperature, and (third column) 700-hPa specific humidity for the cycled OSEs, verified against the ERA interim reanalysis over one-year period. (a)–(c) Northern Hemisphere extratropics (NH; 20–90°N), (d)–(f) Southern Hemisphere extratropics (SH; 20–90°S), and (g)–(i) tropics (TR; 20°N–20°S).

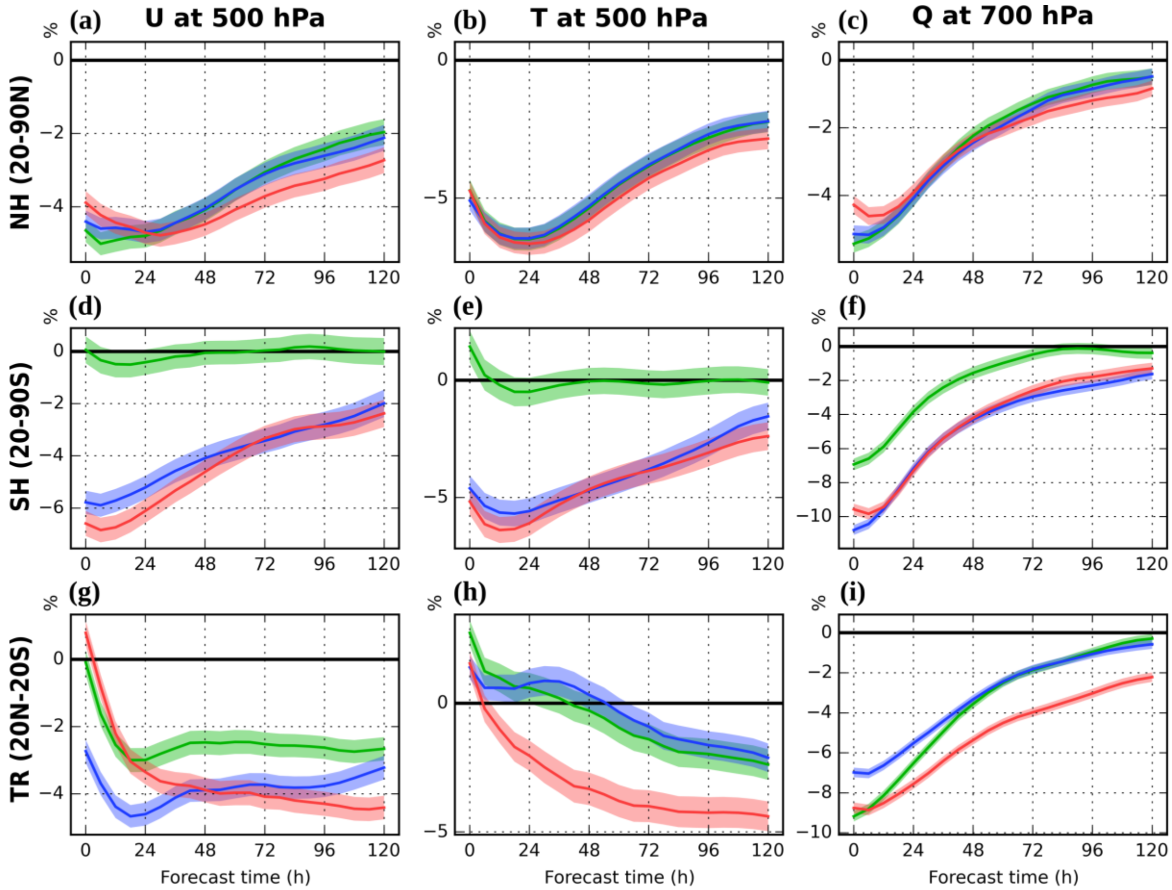

Figure R2: Similar to Fig. 7 and Fig. R1, but for standard deviations of errors (i.e., random errors) relative to CONTROL (%) in different verification regions.

Aside from the model bias, we think that the higher rate of beneficial observations can be understood by two other reasons. First, although the overall rate of beneficial observations is usually not too much higher than 50%, when taking some subsets of observations, due to the smaller sample sizes, it should be possible to obtain higher rates than the overall rate. For example, in our offline DA experiment, the positive impact rate for all precipitation observations are merely 53.5% and 51.8% in terms of the moist total energy norm and the dry total energy norm, respectively (added in P.10, L.22–24; c.f., Fig. 2c, d). The near 70% positive impact rate is only seen when we consider only the nonzero precipitation observations under the condition that less than half of the model background members are precipitating (Fig. 3c). Second, the positive impact rate of EFSO tends to be higher when the background is not very accurate, under which circumstance the observations are able to contribute a larger amount of information. In our experimental setting the CONTROL experiment assimilates only the rawinsonde data, and the assimilation of the precipitation effectively improves the forecast skill, so a larger positive rate compared to the experience in the modern operational system would be expectable. We included the above explanation in the manuscript (P.11, L.14–P.12 L.4).

Minor remarks: 1) The WMO DAOS group recently decided that the term FSOI should be used (where "I" stands for impact) instead of FSO. I recommend following this and using EFSOI. I think there is a document on the WMO website with more details.

We appreciate the comment of Referee #2, but we feel that for a procedure with a long name like “Ensemble Forecast Sensitivity to Observations”, adding “Impact” at the end makes it too long, without really clarifying the meaning of the procedure.

2) I don’t think QC is the appropriate term for data selection and it’s potentially misleading. Why not calling it "data selection criteria" or "observation preprocessing"?

Thank you very much for this suggestion. We have revised throughout the manuscript to use the term “data selection criteria” instead of “QC.” We also modified the explanation of the use of the terminology in the beginning of the manuscript (P.1, L.24–27).

3) Introduction: It would be good to make the discussion and literature review of strengths and weaknesses of EFSOI a bit broader and more critical. The method obviously has strengths, but also some weaknesses. E.g. there is a linearization involved, there are spurious correlations, potential bias (correction) issues and observations interact (adding a new type may decrease the impact of others).

We added a few sentences on the strengths and weaknesses of the EFSO method in the introduction section (P.2, L.16–21). Regarding the observation interaction issue, we added a sentence in P.5, L.17–18. The same paragraph has also broadly discussed some related issues.