# Does the Met Office Consortium Need to Develop an Ensemble for Forecasting?

## Phil Andrews*1*

*1National Institute for Water and Atmospheric research (NIWA), New Zealand*

*p.andrews@niwa.co.nz*

**Introduction**

There are several reasons for running ensembles. One is to meet the needs of data assimilation which includes: providing background error covariance estimates; providing errors of the day estimates; and facilitating the use of data assimilation algorithms that circumvent the need for an explicit tangent linear model and its adjoint model (as required by 4d variational analysis). We will call these “data assimilation ensembles” (DAEs). Another is to provide improved forecasts by providing a range of possible outcomes with attached probabilities of occurrence. We will call these “forecast ensembles” (FEs). An important goal for an FE should be that it is able to make a good forecast even when the analysis has, by chance, an unusually large error. That is it should be more reliable than a deterministic forecast. At the very least it should be able to recognise when an unusually large analysis error may have occurred and that the forecast is unusually unreliable.

**Current Met Office Ensembles**

The Met Office currently runs an 18 member global ensemble, MOGREPS-G, and an 18 member high resolution regional ensemble over the UK: MOGREPS-UK. The MOGREPS-UK members are initialised by reconfiguring the MOGREPS-G perturbations and recentring them on the deterministic high resolution UK analysis.

Although also used for ensemble forecasting, MOGREPS-G itself is primarily designed as a DAE (Bowler et al 2008; Bowler et al 2017). One consequence of this is that the ensemble uses what we shall call a “type I” sampling of the analysis error. That is the distribution of the initial state perturbations of the ensemble matches the distribution of errors around the analysis.

This makes sense in the context of a DAE: most of the ensemble members are sampling the most likely part of the error space allowing one to estimate it with more accuracy.

It is less clear, however, that this makes sense in the context of a forecast ensemble. From an FE point of view we appear to be over sampling small perturbations and under sampling larger ones. This has the effect of increasing the mean distance between the closest ensemble member and the true analysis as the analysis error in any particular instance increases.

This would appear to be in conflict with the desired goal of an FE to be able to make better than deterministic forecasts in those cases when the deterministic forecast has unusually large errors. The raises the question: can an FE do better?

**What Might a Forecast Ensemble Look Like?**

In order to meet our goal for an FE we need two things: to change the way the ensemble members sample the analysis error; and a way to estimate which part of the ensemble space truth is most likely to lie in which is more accurate than the analysis used to construct the ensemble.

Instead of the “type I” sampling used by the DA ensemble the FE could use a “type II” sampling. That is the ensemble members should sample the analysis error uniformly for sigma (number of standard deviations) less than a cut off value; and don’t sample it all for sigma’s greater than the cut off value. The cut off value is of course required because we can only run a finite number of ensemble members. Its value determines the frequency of events where truth lies outside of the ensemble. The uniform sampling for sigmas smaller than the cut off ensures a constant maximum distance between the true analysis and the nearest ensemble member. Together, the constant sampling density and the cut off sigma determine the number of ensemble members required. Of course a practical method has to be found to generate ensemble members with this type II sampling of the analysis error.

At the time of making an ensemble forecast (Ta) the analysis is the best available estimate of the true analysis. Hence the best estimate of the relative probability of each ensemble member being truth is equal to the probability of their initial perturbation given the analysis error distribution function (the absolute probability will also depend upon the sampling density and the cut off sigma). However, at some later time Ta+δ (with Ta+δ < Ta+R where Ta+R is the maximum forecast range the ensemble is run out to) we can use observations made during the interval Ta+δ to refine our estimate of the probability of each ensemble member being truth (and, if we wish, to estimate the value of the analysis error at time Ta). This refinement could be done using various (or indeed multiple) methods including: using an analysis made at Ta+δ; a traditional regression algorithm using innovation vectors on the interval Ta+δ to predict the probability of a given ensemble member being truth at some later forecast range (up to Ta+R); or perhaps an appropriate machine learning algorithm to do the same which might benefit from having a representation of the nonlinear dynamics of the system. Regardless of the details of the refinement methodology the refinement process itself strongly suggests the formation of a lagged ensemble every member of which has an estimate of its probability of being truth. These, perhaps together with estimates of their accuracy (which might depend upon Ta+δ) can be used in generating forecast products from the lagged ensemble.

Another interesting possibility is that one could also use this information to modify the background for the analysis at the current time and hence reduce the analysis error.

**Including Other Sources of Forecast Error**

So far we have only discussed sampling the analysis error. Whilst this is the largest source of forecast error it is not the only one. Other sources of initial state error include all the boundary values used to run the model such as sea surface temperatures and all the values supplied via ancillary files from orography to vegetation types. We can potentially include all of these by uniformly sampling the forecast error at some chosen forecast range perhaps using a method akin to singular vectors. The exact details of such a method to generate type II sampling of the forecast error are still to be worked out.

Model error, caused by approximations and errors in the model formulation, is another source of forecast error. This is usually sampled by running different ensemble members with different perturbations to their model formulations. It is not clear at present if this can be incorporated into the type II sampling methodology. If not one may have to resort to running multiple ensembles sampling model error each centred on one of the members of the ensemble sampling initial state error.

**Conclusions**

In this talk I shown that the method of sampling error/generating ensemble member initial states required for data assimilation ensembles is sub optimal for forecast ensembles.

I have outlined an alternative method for generating forecast ensembles, though many important details remain to be worked out, not least devising a method of generating ensemble members with a type II sampling of the forecast error.

At NIWA we are going to begin by investigating the efficacy of the forecast refinement approach outlined here.

# References

Neill E. Bowler, Alberto Arribas, Kenneth R. Mylne, Kelvyn B. Robertson and Sarah E. Beare 2008: The MOGREPS short-range ensemble prediction system, Q. J. R. Meteorol. Soc. 134, 703 - 722.

N. E. Bowler, A. M. Clayton, M. Jardak, E. Lee, A. C. Lorenc, C. Piccolo, S. R. Pring,

M. A. Wlasak, D. M. Barker, G. W. Inverarity and R. Swinbank 2017: Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations, Q. J. R. Meteorol. Soc. 143, 1280 - 130.