1) P3958, L22 "..the background error covariance does not include tracer-wind correlations.... This limitation can be overcome by using an ensemble Kalman filter...". Note that limitations of 4D-Var can also be addressed using hybrid methods and this is increasingly popular at operational centres (e.g. 4D-Var background errors can be generated using an ensemble method). More generally, though this should definitely not become a "4D-Var versus EnKF" paper, it would be nice if the authors could briefly compare back to their experiences when they performed a very similar study using 4D-Var (perhaps a short discussion or a few sentences in the conclusion?).

We are actually developing hybrid methods using the shallow water model as a follow-up to this paper. For the hybrid we will blend the static covariance from 4D-Var with the flow-dependent covariance from the EnKF. We will mention hybrid models as an additional option in the Introduction. We will also include some comments in the conclusion related to our 4D-Var experience. A direct comparison is not possible, since the observations in the 4D-Var paper differed in two major aspects: (1) only tracer data were assimilated, and (2) the observation grid was much different from that used in the EnKF.

2) P3961, L10 Why is the observation error covariance matrix referred to as "R_ens"? This implies, to me at least, there might be something special about how it is created. Would not "R" be sufficient?

R would be sufficient here since we use a diagonal matrix based on the specified standard deviations rather than a statistical covariance matrix calculated from the randomly perturbed observations. In principle these should be the same, but in practice, with limited ensemble size the statistical covariance is not exactly diagonal. We will change the notation to “R”.

3) P3962 L4, "To avoid filter divergence we apply an inflation factor to maintain reasonable spread in the ensemble". The text needs to say both what the inflation factor is applied to and how it is done.

The inflation factor is a scalar value multiplied by each state vector element in the ensemble perturbations in order to alter the globally-averaged SPREAD (calculated using one or more of the state variables) so that it matches the globally-averaged RMSE. For the EnKF-uv system we alter the SPREAD in the vector wind error (V_SPREAD in Eq. 6) so that it matches the vector wind RMSE (V_RMSE in Eq. 5). For the EnKF-$\psi\chi$ system we alter the SPREAD in the streamfunction so that it matches the streamfunction RMSE. Although the inflation factor is calculated using a subset of state variables, it is applied to all variables (including height and ozone for EnKF-uv and including height, ozone, and velocity potential for EnKF-$\psi\chi$). We will make this more explicit in the text.

4) Equations 8 and 9 seem slightly confusing to me. Why is only the latitude dependence made explicit? How was a latitude/longitude/time varying error reduced to a latitude/time varying RMSE? I think I can guess, but please add a bit more explanation. Note also that later, in equation 11 and associated discussions, you are giving a lot more detail about what are probably very similar calculations, but this time explicitly showing how the latitude-dependent RMSE has been computed.

The latitude dependence is made explicit since we show errors as a function of latitude in Figure 10. Eq. (11) does show how the RMS is calculated separately for each latitude. This should be explained as part of the discussion in Equations (8) and (9). Similarly, the area-weighting given in Eq. (12) is the explicit formulation of what is said in words at the bottom of p. 3969. We will edit the text to make this more explicit.

5) P3972, L16-23. Discussion around the chi-squared metric seems to start with a nonnormalised version (which is expected to be equal to the number of observations) and moves to a normalised version (expected to equal 1) without explaining properly how the second is obtained, and without changing terminology to distinguish the two. So it comes as a shock in the last sentence when a "chi-squared" of around 1 is considered good.

Thank you for pointing this out. It is indeed the normalized “chi-squared” that should equal 1. We will fix the text to make this clear.

6) p3976, l22: "It appears that combining height observations and ozone observations acts as a filter to dampen the GW that would otherwise be generated by the ozone observations alone". I am not sure the "filter" idea is necessary, and it would be good to explain what it means physically. Instead, is it possible that you simply require both height and ozone observations to properly constrain a shallow water model? Height observations just constrain height and the balanced part of the wind, but leave the unbalanced part of the wind unconstrained. That could allow GWs to develop. But the ozone observations could help constrain the unbalanced part of the wind field, stopping spurious GWs developing.

We will omit the “filter” concept from our discussion. Our results show that if we only assimilate ozone observations there is a lot of imbalance (Figure 9d), while height only assimilation produces much less imbalance (Figure 9i). Combined height and ozone assimilation produces an intermediate amount of imbalance. This suggests that adding height observations to the ozone system reduces the imbalance (excites fewer gravity waves) compared to ozone alone (Figure 9n vs. 9d). Similarly, adding ozone observations to the height system increases the imbalance (Figure 9n vs. 9i). We aren’t sure exactly how this process works, but we suspect that the cross-covariances of ozone/wind and ozone/height are not as balanced as the cross-covariance of height/wind. Your discussion of height/ozone constraining different parts of the wind is interesting. In our truth run there is almost no imbalance, so any imbalances is “unwanted”. But if the truth run had significant imbalance that we needed to analyze, it may be that ozone observations would be more useful for the unbalanced part of the wind, while height would be more useful for constraining the balanced part of the wind.