Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2017-802-RC1, 2017

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "Extraction of wind and temperature information from hybrid 4D-Var assimilation of stratospheric ozone using NAVGEM" *by* Douglas R. Allen et al.

A.J. Geer (Referee)

alan.geer@ecmwf.int

Received and published: 1 November 2017

This is an interesting development in a series of studies by the authors investigating the possibility of wind extraction from tracers in the stratosphere. The latest study moves from using a shallow water model to a full operational-quality NWP system, albeit in an OSSE configuration with a restricted set of observations. The study is well presented and interesting, and the OSSE framework seems novel.

Major comments

1) The introduction needs to be improved in order to summarise the reasons why ozone

assimilation has not yet been successful in operational NWP systems. A short recap of the points made in Allen et al. (2015) would be useful here, and it is worth restating that one of the big problems seems to have been bias between model and observations, rather than particularly the deficiencies in the data assimilation framework. The introduction also needs to motivate the current study better, for example justifying why, since real MLS ozone observations are available, the framework of an OSSE been chosen. It would be useful to state here what benefit NAVGEM gets from MLS ozone assimilation, and to recap any studies that might have been done. A number of minor issues relate to these issues in the introduction:

- Page 2 Line 27/29: "Theoretical studies" may not be the best description of work like that of Peubey and McNally (2009) which tested a real operational NWP system

- Page 2 Line 32: "the operational benefit has not yet been obtained" needs to precisely relate to stratospheric tracers, rather than all tracers, as NWP centres routinely benefit from 4D-Var tracing of tropospheric water vapour from IR and MW radiances, as explained by Peubey and McNally (2009)

2) Based on shallow-water results, line 10 of the introduction says roughly that "including cross-correlations in B between ozone and other variables provides additional ozone-wind benefit". I would have expected to see experiments in this new study that would have explored whether this remains true in the full NAVGEM framework.

3) The OSSE design is described and tested in secs. 2.4 and 4.1. The approach of using real observations in the troposphere seems novel, so it deserves more (critical) investigation within this study. If there are any precedents to this design, they should be cited. There is one interesting parallel with the work of Harnisch et al. (2013) which also used a mixture of real and simulated observations, albeit in an EDA framework.

The statements in sec. 2.4 that the troposphere is "constant, ... essentially independent" and "this gives ... nearly identical tropospheric analyses for each experiment" are imprecise and omit a key idea for understanding this framework. That idea is later supplied in sec. 4.1: "additional observations ... change the numerics of the cost function ... resulting in slight changes to the troposphere". I have run some experiments (Geer, 2016) that explore exactly this issue: when assimilating an identical set of observations into an identical data assimilation system, even the slightest numerical perturbation will generate chaotic divergence between different runs of the data assimilation system. As with my experiments, the troposphere in these new OSSE experiments is not fully constrained by observations and hence will exhibit substantial chaotic variation from one run to the next, whenever the slightest numerical difference is introduced. This spread is somewhat smaller than the expected analysis error for reasons explained in Geer (2016) but is still appreciably large. The troposphere in each separate experiment can be seen as being drawn from a potential ensemble of tropospheric analysis states. The really key understanding is that the stratosphere in the truth experiment (TE) must also be just one realisation drawn from a potential ensemble of stratospheric states, this ensemble being constrained by the data being assimilated in the troposphere. Here also the parallel to Harnisch et al. (2013) is clear.

Thinking this way allows some of the more intriguing results of this study to be analysed better. For example the tropospheric errors shown in Fig. 9 will likely be a rough estimate of the spread of this "tropospheric ensemble" - this would help the authors to explain what they are already saying in Sec. 4.1 about the minimum possible errors in the stratosphere. However it is thus intriguing that by assimilating stratospheric radiance observations simulated from the TE as in Sec. 5.1, Fig. 15, this tropospheric "spread" gets larger, from around 0.2K to 0.5K. This suggests non-optimalities in the NAVGEM data assimilation system or problems in the OSSE framework that need further investigation. Assimilating observations in an optimal system should make errors go down, not up. For me a likely explanation is that the troposphere and the stratosphere are not in reality independent. In the stratospheric radiance assimilation experiment, the troposphere is not identical to that in the truth experiment, and the stratospheric state most compatible with that tropospheric realisation is different to the stratospheric state in the TE. Hence, by trying to make the experiment stratosphere fit that in the TE, it may

require the generation of incorrect increments in the troposphere, which then increase the "spread". Sub-optimalities in the NAVGEM system are also a possibility.

4) In section 4.1, line 10, the expectation of "zero analysis error relative to the TE" with "perfect stratospheric observations" is not so obvious, for a number of reasons:

a) The stratospheric analyses are, even in the presence of near-perfect observations, still just realisations from a hypothetical ensemble of possible stratospheric analyses - this is equivalent to what the authors already say in section 4.1

b) The 0.1 ppmv observation error is not zero, and hence does permit some additional spread within this hypothetical ensemble

c) there may be sub-optimalities of the data assimilation system, such as sampling error in the ensemble-derived part of the background error covariances

d) It is not clear that the stratospheric state is fully determined by the ozone field.

I agree with the authors that their experiments in section 4.1 define a minimum "limit on the level of errors we can reliably distinguish from the TE" but the reasons are more complex than currently stated. It would also be good to see the maximum possible level of error in this experiment, i.e. the "errors" (better, spread) between two realisations of the truth experiment, generated for example by starting another TE from perturbed initial conditions. These minimum and maximum errors roughly define the limits of sensitivity of this novel OSSE design.

5) An additional limitation to the sensitivity of this experimental design is the statistical significance of differences between experiments. In an ideal world all the relevant figures should have statistical significance bars added. For example, intriguing results like the decrease in analysis errors when ozone observation errors are increased (page 14, lines 11-13) could possibly be explained by a lack of statistical significance. However, the statistical significance would not be easy to estimate in this framework except by using an ensemble of perturbed experiments similar to what I was using in Geer

(2016), and analogous to the "spread" between experiments hinted at by Figure 9.

Considerations of sensitivity and statistical significance are important to the conclusion in sec. 5.2 that "in the presence of realistic radiance observations, it is likely that adding ozone assimilation from current ozone retrieval observations .. will have little impact". Again going back to Geer (2016), it is hard to detect the impact of small changes in the observing system in the forecast quality of an operational-quality NWP system. As shown there, adding a single new instrument in the troposphere will only become statistically significant in an experiment containing around 300 forecast samples. In the work under review, there is only a single sample, so it cannot hope to have the required sensitivity or statistical significance. This means the conclusion is unnecessarily pessimistic. If the OSSE were continued for several months, and statistics computed from that whole period, the benefit of ozone assimilation might be seen to be statistically significant. (Many other results presented in this study are more convincing and would likely be statistically significant, but this one is probably not.)

6) Page 16, line 11-14: For similar reasons to those explained in comment 5, the conclusion that current NWP systems can't benefit from ozone-wind interaction in a 4D-var system is probably incorrect. It is more a matter of quantifying the size of that benefit, which is something this current OSSE presentation does not have the sensitivity to explore.

Minor comments

Page 3, Lines 12-13, from "will perturb..." are difficult to follow and need rewriting for clarity.

Page 4, line 27: it is confusing to refer to X' first as an "ensemble state" and then as a "perturbation from the ensemble mean".

Page 5, line 8: "the standard suite" of observations would nowadays include satellite radiances, so this would be better described as a "baseline" or "conventional only" suite

of observations.

Page 6, line 26: Figure 3 caption is missing this piece of information: the level is 10.5hPa

Page 6, line 31-32: That the ozone initial conditions are biased with respect the ozone scheme in the experiments seems a major flaw and needs more explanation or investigation - probably it is no real problem, as implied by section 4.1, that the initial conditions don't matter to the results on 1 December.

Page 7, line 10: "By using a different stratospheric analyses" would be clearer if it was written "by replacing the initial conditions with a different stratospheric analysis"

Page 9, line 8: The high ensemble spread in the vortex is non-intuitive and deserves some explanation. How is this being generated? It should be fairly clear if, for example, the central location of the vortex varies quite a lot across the ensemble.

Page 10, line 10: "It is likely that radiances are the limiting factor..." - please explain this better - I don't understand it.

Citations

Allen, D. R., Hoppel, K. W., and Kuhl, D. D.: Wind extraction potential from ensemble Kalman filter assimilation of stratospheric ozone using a global shallow water model, Atmos. Chem. Phys., 15, 5835-5850, https://doi.org/10.5194/acp-15-5835-2015, 2015.

Geer, Alan J. "Significance of changes in medium-range forecast scores." Tellus A: Dynamic Meteorology and Oceanography 68, no. 1 (2016): 30229, http://dx.doi.org/10.3402/tellusa.v68.30229

Harnisch, F., S.B. Healy, P. Bauer, and S.J. English, 2013: Scaling of GNSS Radio Occultation Impact with Observation Number Using an Ensemble of Data Assimilations. Mon. Wea. Rev., 141, 4395–4413, https://doi.org/10.1175/MWR-D-13-00098.1

Peubey, C., and A. P. McNally. "Characterization of the impact of geostationary clearâĂŘsky radiances on wind analyses in a 4DâĂŘVar context." Quarterly Journal of the Royal Meteorological Society 135, no. 644 (2009): 1863-1876.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2017-802, 2017.