Atmos. Chem. Phys. Discuss., 6, S2208–S2212, 2006
www.atmos-chem-phys-discuss.net/6/S2208/2006/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "The ASSET intercomparison of ozone analyses: method and first results" *by* A. J. Geer et al.

A. J. Geer et al.

Received and published: 2 August 2006

Reviewer 2 is keen to draw out lessons learnt from the ASSET intercomparison of ozone analyses. This note attempts to do that and to explain our strategy. We will deal with the details of reviewer 2's comments in a separate interactive comment and we will include a short summary of the following notes in one or two paragraphs added to the conclusion of the paper. We thank the reviewer for provoking this discussion.

Why do we need to compare assimilation systems? Why not just compare the models on which they are based? Models do not always perform the same way in forecast runs as they do within an assimilation system. The assimilation of observational data changes temperatures, dynamics and budgets, as well as the parameter (e.g. ozone) we are interested in. If the system is working well, the analysis should be our best estimate of reality. We need to know if the models still provide the same answers under these conditions as they do in forecast mode. Often they do not. The spin-down in ERA-40 precipitation in the tropics is perhaps the most well-known example of this (e.g. Uppala et al 2006, QJ). In the stratosphere, the excessively fast Brewer-Dobson circulation (e.g. Schoeberl et al. 2003, JGR) is another.

We can consider an ideal intercomparison between ozone assimilation systems. There could be many reasons for the differences between them:

- the observations assimilated, as well as the observation operator, the specification of observation errors and the quality control (QC) applied.
- the model, i.e. its chemistry and transport properties
- the dynamical analysis (i.e. temperature and winds) whether imposed, in a CTM, or integral to the system, in a GCM.
- the assimilation technique (Kalman filter, 4D-Var etc.)
- the specification of background errors

With such a large range of possibilities, it can be hard to understand differences between systems. Much information can be gained from intercomparisons between forecast models alone, or from experiments that vary the model whilst keeping the rest of the assimilation system the same. One example from the ASSET intercomparison was the trial of different linear chemistry parametrizations in the DARC/Met Office assimilation system.

Another way to cut down the possibilities is to force every system to assimilate the same observations. However, quality control is always needed to remove bad data, so we would have to apply the same in each case. But might it be that the quality control algorithms are actually one of the things we want to test? More generally, we need

ACPD
6, S2208–S2212, 2006

Interactive Comment

to decide whether we are comparing systems, or products. If we want to compare products, such as the ozone analyses produced operationally by ECMWF or KNMI, we cannot apply such restrictions. The ASSET intercomparison was a hybrid, in that a subset of systems did assimilate exactly the same observations (though subject to varying QC approaches), but in other cases (e.g. ECMWF) we simply took operational products. However, within that subset of systems, the strategy of common observations was very important to understanding the differences between them.

Our ideal intercomparison would verify analyses against a fixed set of independent observations, but as suggested by the reviewer, in the case of satellite profiles like MIPAS it would do so in observation space. Each system would passively assimilate the independent observations, using the same observation operator. O-A differences would be calculated at the correct time from the full-resolution system, with no interpolation errors. To compare between different systems, O-A statistics could then be binned onto common levels.

As a more general principle, if interpolation is necessary it is better to interpolate from the finer grid to the coarser, and not the other way around. The very high vertical resolution of the sonde data means that it can be binned between model level boundaries. Hence, it is probably better to compare models to ozone sondes in model space.

The ASSET intercomparison failed to achieve the ideal intercomparison in a number of ways:

- all analyses were interpolated to a common grid before comparison to independent data
- differences with respect to independent observations were made on this common grid, not in observation space, and some observations were treated as point retrievals, without taking into account their averaging kernels.
- a common set of assimilated observations was used in only a subset of analyses

ACPD
6, S2208–S2212, 2006

Why was this? In practice, the difficulties were as much organisational as scientific. Even as part of a collaborative European project it needed to be as easy as possible for the participants and organisers. With hindsight, we could have insisted on more standardisation, though that would still have been difficult from an organisational perspective.

All analyses were collected in their original formats but then interpolated onto a common grid before further processing. As a result, the code used to generate statistics and figures was much simpler than it might have been. Since we were doing this for the first time, we did not know what statistics or comparisons would work best. The common grid made it easier to experiment. For example, rather than concentrating on independent data we might have looked at process-based statistics (e.g. measures of ozone hole area or stratosphere to troposphere transport). That would have been easier on the common grid. Also there needs to be a common grid to explore and visualise differences between analyses.

Intercomparisons often highlight gross errors. A number of these were discovered during the project. In many cases, the participants re-ran their assimilation experiments and supplied new analyses. If nothing else, to uncover these problems was a worthwhile outcome of the project. Again, to cope with a continual influx of new analysis datasets and re-runs, it was important to keep the project framework as simple as possible.

In the paper, we were able to estimate the impact of the gridding and interpolation on our statistics (Section 4.1). We did not separately consider the effect of vertical interpolation and that was an oversight. Another paragraph and a figure will be added on that subject, but away from the tropical tropopause it has little impact. Overall, we believe that though gridding and interpolation do affect our statistics, doing things more accurately would not have significantly altered our conclusions. A common grid will always be part of such an intercomparison, but we agree that validation against independent satellite data should in future be done in observation space.

ACPD
6, S2208–S2212, 2006

Finally, it is true that some of the conclusions highlight model problems that are already known. This may be because in practice ozone data assimilation is still decoupled from dynamical data assimilation. Operational NWP systems such as those of the ECMWF or Met Office do not allow prognostic ozone amounts to influence heating rates. Without the link between chemistry and dynamics, ozone analysis in NWP systems is very similar to that in CTM-based assimilation systems, where dynamical forcings are externally supplied. The problems affecting CTM model-only experiments (e.g. of tropospheric chemistry, or excessive transport in the Brewer-Dobson circulation) are well known, and often similar to those affecting model-only integrations (e.g. Austin et al., 2003, ACP). In this study we see that such problems can affect NWP and CTM data assimilation systems equally. However, once ozone and dynamics are fully coupled in NWP systems, intercomparisons between them will increasingly reveal separate issues.

In summary, there are areas where the intercomparison could have been made more scientifically accurate, and approximations were made for expediency, flexibility, and because this kind of activity has not been undertaken before. However, we show that the impacts were relatively small and that the conclusions of the study are not affected. The process of intercomparison has helped to improve many of the participating systems, and also showed many areas where further work is needed.

Interactive comment on Atmos. Chem. Phys. Discuss., 6, 4495, 2006.