DATA ASSIMILATION: END-TO-END APPLICATION TO RESEARCH SATELLITE DATA

William Lahoz

DARC, University of Reading, RG6 6BB, UK, Email: swslahoz@rdg.ac.uk

ABSTRACT/RESUME

Research satellite data are a valuable but expensive resource to study and monitor the Earth System. Thus, it is important that the best possible use is made of this resource, end-to-end in the mission: from pre-launch mission planning to post-launch value-adding activities. This paper argues, in general and with specific examples, that space agencies should position data assimilation at the centre of their satellite activities, using it for the objective evaluation of the incremental value of future satellites, the objective evaluation of current satellites, and for adding value to the satellite data by combining their information with that from Earth System models.

1. INTRODUCTION

Data assimilation (DA) is a powerful technique that seeks to combine observations with output from a numerical model in an optimal manner. It is a tool based on techniques from estimation and control theories ([1-2]).

In the 1990’s, following years of development of meteorological data assimilation by the Numerical Weather Prediction (NWP) community, data assimilation methodology began to be applied to the assimilation of data from research satellite data, with a particular emphasis on stratospheric ozone [3-4]. Because of its comparatively later application, assimilation of research satellite data is less prevalent than that of operational satellite data, which is chiefly assimilated for NWP. Nevertheless, there has been an increase in the assimilation of research satellite data over the last 15 years, with the field evolving from initial efforts to test the methodology to later efforts focusing on products for monitoring ozone and other constituents. More recently, the production of ozone forecasts by a number of operational centres (e.g. the European Centre for Medium-range Weather Forecasts, ECMWF, [5]) assimilating ozone data from research satellites has become routine. A notable feature of the application of the data assimilation methodology to research satellite data has been the strong interaction between the NWP and research communities, for example, in the EU-funded ASSET project [6].

Research satellite data are thus a valuable resource for NWP, as well as for studying and monitoring the Earth System. Although the assimilation of research satellite data has primarily focussed on the atmosphere, they are also being used for other components of the Earth System.

Notwithstanding their value, research satellite data are an expensive resource (as are operational satellite data). Thus, it is important that the best possible use is made of this resource, in particular, end-to-end in a research satellite mission from pre-launch mission planning to post-launch value-adding activities.

This paper argues, in general and with specific examples, that space agencies should position DA at the centre of their satellite activities, using it for the objective evaluation of the incremental value of future satellites, the objective evaluation of current satellites, and for adding value to the satellite data by combining their information with that from Earth System models.

2. RESULTS

Arguably, to make the best use of satellite data and, in particular, of research satellite data, we need to: use the data end-to-end, from pre-launch to post-launch; perform exchange of knowledge between theory and applications; and confront models with observations. The outcomes of this approach are: an optimized observing system; information on observation error characteristics; stringently tested models; and value-added information accruing from melding observations and models. Using the examples below, structured around a question about the Earth System and potential solutions, we show that data assimilation is fundamental to achieving these outcomes.

Question 1: What is the incremental value of observations from a proposed instrument?

To answer this question we need to address the following: what new observations are of interest; what is the nature of the future Global Observing System (GOS) and what are the errors of the observing platforms; and that although in principle adding new
information should be beneficial, this may not always be the case (see, e.g., [4]). Thus, as well as testing the incremental value of new observations, we are also testing the current data assimilation system.

A solution is to set up an Observing System Simulation Experiment (OSSE; [7] – see also Fig. 1) to test if adding a new observation has a significant impact. Note that variants of the OSSE have been discussed recently in the literature (see, e.g., [8]).

In the example below we address the incremental value of horizontal wind observations from the proposed Canadian Space Agency (CSA) SWIFT instrument aboard the CHINOOK platform, due to be launched in 2010. Further details can be found in [9].

As described in [9], OSSE experiments show there is a strong suggestion that SWIFT observations have a positive impact on zonal wind analyses in the tropical stratosphere, and in the extra-tropical upper stratosphere when the flow regime is changing relatively fast and SWIFT observations are available (Fig. 2). There is also a strong suggestion that SWIFT observations have a positive impact on stratospheric ozone in regions where the vertical gradient of ozone is relatively high (tropopause and 10 hPa; not shown). As discussed in [9], most of the conclusions remain robust even after consideration of the caveats associated with OSSEs. This information was fed back to CSA, and helped in the decision to continue building the SWIFT instrument.

Figure 1. OSSE schematic. The “truth”: T, is created, usually from analyses or a model run. The truth is sampled to derive observations, with errors, for a baseline (B) GOS, and for a perturbation (P) GOS. The perturbation GOS is made of the baseline GOS plus the observation of interest, in this case SWIFT horizontal winds and ozone. After assimilation of the B and P versions of the GOS, the results are compared to T. If P-T is significantly smaller than B-T it is argued that the observation of interest has a significant impact on the GOS. In particular, we are interested if ABS(B-T) – ABS(P-T) is greater than zero (see Fig. 2).

Figure 2. Plot of the difference ABS(B-T) – ABS(P-T) for the zonal wind (m/s) averaged over a nominal January. The shaded area indicates where ABS(B-T) is greater than ABS(P-T) at the 95% confidence limit. Based on [9].

Question 2: What is the quality of Earth Observation data?

To answer this question we need to address the consistency of data in both space and time.

Solutions include: looking at the consistency of data, e.g., are observations of different stratospheric tracers consistent [10-11]; looking at the self-consistency of data using data assimilation diagnostics – this tests whether a priori assumptions on errors are correct [12-14]; comparison against independent data – this provides an estimate of biases [12-14]; and data monitoring – this looks at whether data characteristics change in time [12, 15].

We present below two examples: (i) a test of the self-consistency of Envisat MIPAS ozone data ([13], Fig. 3); and (ii) evaluation of Envisat MIPAS ozone data using independent data and data assimilation for interpolating between the MIPAS and independent data ([14], Fig. 4).

As can be seen from Fig. 3, the Observation minus Forecast (OmF) statistic is approximately zero in the stratosphere. This result suggests there is no bias between the MIPAS ozone data and ozone forecasts from the model, and is consistent with a priori assumptions that the MIPAS ozone data and the model representation of ozone are unbiased. Note, however, that this test does not identify whether the MIPAS data or model are themselves unbiased, nor quantify the bias. For this one needs to appeal to independent data (see Fig. 4).
Figure 3. Observation minus Forecast (OmF) statistics averaged for the period 14-28 September 2002. The observations are MIPAS ozone; the forecasts are short-range forecasts from the Met Office model used in the assimilation. Left-hand panel: standard deviation (dashed line) and bias (solid line) of the OmF differences – units: percent; right-hand panel: skewness (solid line) and kurtosis–3 (dashed line) of the OmF differences – units: dimensionless. Based on [13].

Figure 4 considers the period 18 August – 30 November 2003, binned over five latitude ranges: 90°S – 60°S; 60°S – 30°S; 30°S – 30°N; 30°N – 60°N; 60°N – 90°N. The statistic plotted is the difference between the MIPAS ozone observation and the BASCOE ozone analysis from the Belgian Institute of Space Aeronomy, BIRA-IASB, O1-A, minus the difference between an independent ozone measurement and the same ozone analysis, O2-A. This difference is O1-O2, and the ozone analysis serves as an interpolator between the MIPAS ozone data and the independent ozone data, in this case ozonesondes and HALOE.

Figure 4 shows that MIPAS ozone data generally have a positive bias of ~5-10% against ozonesonde data (100 hPa – 10 hPa) and against HALOE ozone data (50 hPa – 1 hPa).

Question 3: What is the quality of a model; what is the quality of an analysis?

To answer this question we need to address the following: a model has several components, e.g., transport and chemistry; analyses are a combination of observational and model information, including their errors.

A solution is to perform an intercomparison of analyses [14, 16]. An advantage of intercomparisons is that they are generally found to reveal shortcomings more quickly. Within the ASSET project [6], two intercomparisons were performed: one of ozone analyses [14]; the other of ozone chemistry parametrizations [16]. We focus on the first intercomparison (Figs. 5-6).

Figure 5. Colour key for Fig. 6. Based on [6].

Figure 6. Ozonesonde measurements (triangles joined by black line) and analyses (see key Fig. 5), at 68 hPa over the South Pole, over the period 9 August – 30 November 2003. Shown are the ECMWF operational and MIPAS, DARC, KNMI TEMIS, BASCOE v3d24 and v3q33, MOCAGE-PALM Carolle v2.1 and Reprobus, Juckes and MIMOSA analyses. Units: ppmm. Based on [6].

From Fig. 6 it is seen that in the Southern Hemisphere ozone hole, only the analyses that correctly model heterogeneous ozone depletion are able to reproduce
the ozone destruction over the Pole. These analyses are those using Carriolle schemes versions 1.2 and 2.1 (see, e.g., [16]), which include a term to take account of heterogeneous chemistry (operational ECMWF system; operational ECMWF system with MIPAS; MOCAGE-PALM – see [6, 14]), or those using comprehensive chemical schemes (MOCAGE-PALM Reprobus; and BASCOE v3d24 and v3q33 – see [6, 14]).

There are two other points worth mentioning: First, ECMWF operational analyses only capture the full ozone depletion during October when they began assimilating MIPAS ozone data for the first time, the benefit coming from the relatively high vertical resolution of MIPAS, and the fact that before this only limited ozone data were assimilated [6, 14]; Second, the original BASCOE analyses (v3d24) were found to perform poorly in the ozone hole in the intercomparison – the scheme was improved and the new analyses (v3q33) performed better, demonstrating the value of intercomparisons for identifying shortcomings.

Most of the analyses (except KNMI TEMIS, who assimilate SCIAMACHY total column ozone data) at 68 hPa show too high ozone in November compared to ozonesondes (Fig. 6). This is thought to be due to the relatively broad resolution of MIPAS (and SCIAMACHY) ozone profile data compared to the ozonesondes, thus the analyses show an influence from the much higher ozone amounts at levels above, where the polar vortex has broken down. The conclusions drawn from the above intercomparison are that, in general, with current DA systems, in regions of good data quality and coverage, similarly good ozone analyses are obtained regardless of the DA method, or the model. This reflects the generally good quality of the MIPAS ozone observations. There were areas where some models performed better than others, and in general the improved performance could be explained by better modelling of transport and chemistry.

**Question 4: What is the “best” estimate of the system?**

To answer this question we must address the following: the system under study, in this example ozone in the atmosphere; that we combine observational and model information, including their errors; and we need to define what the “best” way of combining information is.

A solution is data assimilation, which provides an objective way of obtaining an “optimum” estimate of the system, the analysis. Note that, in general, analyses obtained using data assimilation are sub-optimal due to shortcomings in the methodology. However, arguably, using data assimilation is the best we can do.

In Fig. 7 we show an example of how we use data assimilation to provide the “best” estimate of the ozone field and, furthermore, how data assimilation adds value to both the observations and the model.

![Figure 7](image.png)

**Figure 7.** Schematic showing how the data assimilation methodology melds observational and model information. The MIPAS ozone observations at 10 hPa on 23 September 2002 (lower left hand panel) are combined with a short-range model forecast of the ozone field for that day. The lower right hand panel depicts a six day forecast of the ozone field for 23 September 2002. The combination of the observational and model information, and their errors, using data assimilation produces the ozone analyses for 12Z at 23 September 2002 shown in the top panel. Blue indicates relatively low ozone values; red indicates relatively high ozone values. Based on [13].

Figure 7 shows that the data assimilation method adds value to the observations by filling the data gaps in an intelligent way, and adds value to the model by constraining it with observations. In particular, the 6-day ozone forecast for 23 September 2002 shown in Fig. 7 does not capture the ozone hole split. This, however, is captured by the analyses, and confirmed by comparison with independent data [13].

**3. CONCLUSIONS**

The four questions posed in this paper, the solutions offered, and the examples illustrating these solutions, indicate that data assimilation adds value to both observational and model information. Furthermore, the data assimilation methodology allows the evaluation of both observational and model information.

Based on these results, this paper recommends that space agencies position data assimilation at the centre
of their activities, generally concerning research satellites. In particular, data assimilation should be used for:

- The objective evaluation of the incremental value of future satellites (using OSSEs);

- The objective evaluation of current satellite data (using Observing System Experiments, OSEs); and

- Adding value to satellite data by combining observational and model information.

4. REFERENCES