Data Assimilation Methodology Developments at ECMWF

S.J. English, P. Browne, M. Bonavita, M. Chrust, P. de Rosnay, A. Geer,
J. Goddard, M. Hamrud, E. Holm, S. Lang, P. Laloyaux, P. Lean,
O. Marsden, L. Isaksen, S. Massart, D. Salmond, D. Schepers

ECMWF, Shinfield Park, Reading, RG29AX, UK

ABSTRACT

In this paper we give a brief overview of current topics in data assimilation at ECMWF, of relevance to the assimilation of sounder radiances. ECMWF have developed a C++ code control layer to run the data assimilation in a single executable. This is known as the Object Orientated Prediction System (OOPS). This has previously been applied to data assimilation for simplified models. Over the past two years ECMWF's Integrated Forecast System (IFS) code has been refactored to allow IFS to be called and controlled from OOPS, with a view to using OOPS operationally. The initial goal is to replicate the current atmospheric 4D-Var capability with OOPS. Once this is achieved OOPS will provide a flexible framework that will enable introduction of new data assimilation developments, including more general forms of 4D-Var. OOPS is also being applied to other components of the IFS, such as the NEMOVAR ocean data assimilation, and will be applied to the land DA system in due course. During the process of this major technical development project, the opportunity has been taken to perform a modernisation of the IFS code base, and this allowed some errors and issues in the IFS code to be found and corrected. Alongside this major infrastructure project, a number of important science improvements have been made to the data assimilation methodology, and several others are being researched. Recent improvements include the application in 4D-Var of the the humidity background error covariance estimates from ECMWF's ensemble data assimilation (EDA). Areas of active development include overlapping 4D-Var 12 hour windows, asymmetric re-centred EDA and an augmented control vector implementation of the B operator. In addition, techniques for coupled ocean-atmosphere data assimilation are being developed and the impact for sounder radiances is described in more detail in the poster by Eresmaa, Lupu, Schepers et al.

1. Introduction

4D-Var has enabled satellite radiances to become critical observations for Numerical Weather Prediction leading to a huge reduction in analysis error since 1997, as shown below. ECMWF uses 4D-Var to generate both the high resolution analysis [1] and the ensemble covariances using an ensemble of 4D-Vars (EDA) [3]. The EDA 4D-Var approach continues regularly to give large skill improvements [3]. For example the extension of wavelet **B** to humidity, shown below. In this poster we show 4D-Var and EDA is still a developing concept with potential for further improvement. ECMWF are also making major code infrastructure changes to improve flexibility and scalability. Work also continues on weak constraint and longer windows, extending all-sky capability (Geer *et al*., ITSC-21), the inner and outer loop resolution and configuration and the stability and quality of the tangent-linear model.

The 24h RMSE error (Figure 1) gives an indication of the quality of the operational analysis. It is evident that improvements to the analysis since 4D-Var have been large and on-going, including the 12 hour window, the humidity control variable, wavelet Jb from an ensemble of 4D-Vars and flow dependent balance operators. The analysis quality also depends on improvements in the model and observations.

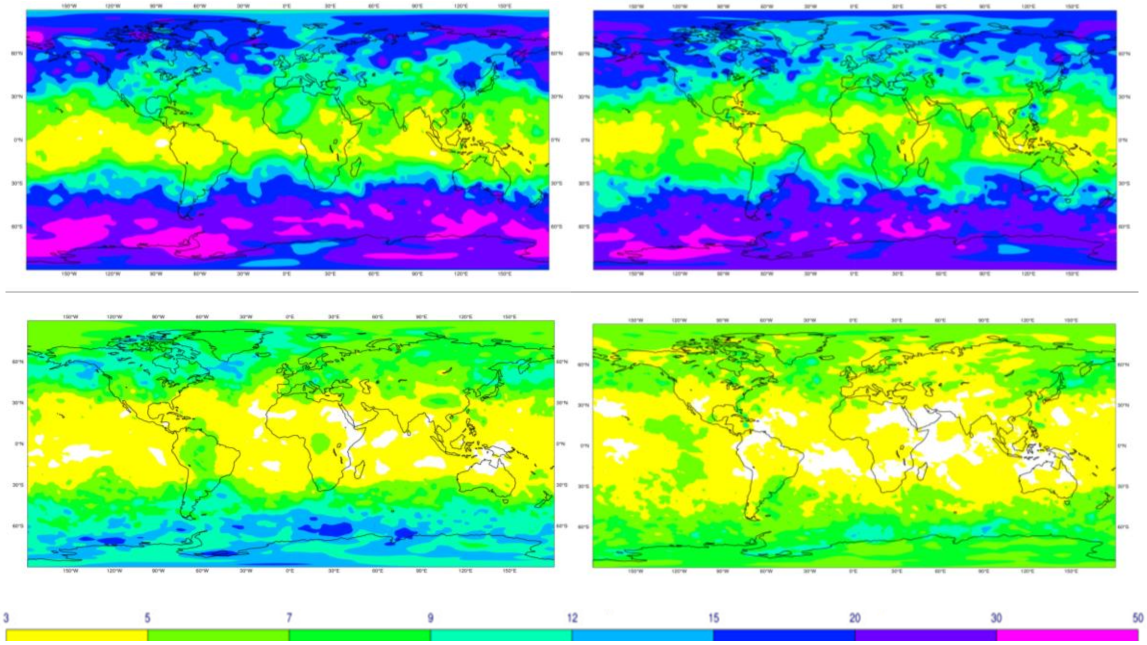

Figure 1: Root mean square of 24 hour operational forecast of 500 hPa geopotential height errors for top panels L to R: Oct 1997 (3D-Var) and Oct 1998 (first 4D-Var); bottom panels L to R: Oct 2007 (mature 4D-Var) and Oct 2016 (4D-Var + EDA).

Figure 1 illustrates how the on-going refinement of details of the 4D-Var implementation are where the main benefits arise.

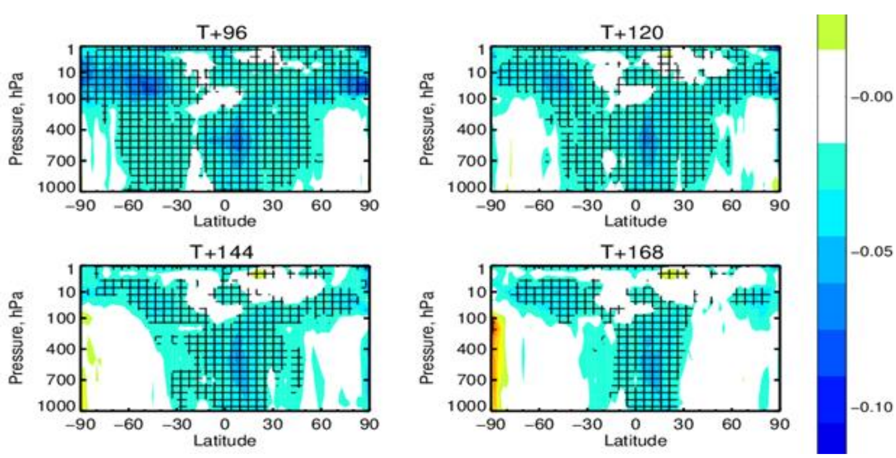

Figure 2: Normalised change in the wind vector RMSE for forecast range 96 to 168 hours over 1 June-30 September 2016 (blue = improvement, crosses indicate a statistically significant result)

The most recent example introduced later than the final panel of Figure 1 is the extension of the wavelet B ($\mathbf{B} = \mathbf{T}^{-1}\mathbf{\Sigma}_b^{½}\mathbf{C}\mathbf{\Sigma}_b^{½}\mathbf{T}^{-1T}$ [2]) formulation to humidity. This led to smaller analysis increments which usually implies a better analysis, as seen in Figure 2. A major improvement in wind scores (4D-Var tracer effect) was achieved, as shown here. It went operational in ECMWF Cycle 43r3 (July 2017). Improvements found widely up to 4-6 days ahead and to longer range in the tropics, and at higher altitudes.

2. Current research and development topics

a. EDA and ensemble B

Methods to reduce the cost of the EDA without reducing skill are being investigated. One of the most promising is to run the perturbed EDA members run with one outer loop whereas the unperturbed EDA control runs with three outer loops, and then to re-centre on the control member. This is sometimes referred to as asymmetric EDA, as the perturbed members have a different configuration to the control. We are investigating two re-centring options: 1) re-centring on the control analysis; 2) "soft" re-centring using $\mathbf{X}_{\mathbf{fg}}$ from the more accurate control member to "guide" minimization of perturbed members. The goal is to be able to increase the number of EDA members so that the EDA has the same number of members as the ensemble forecast (currently 50).

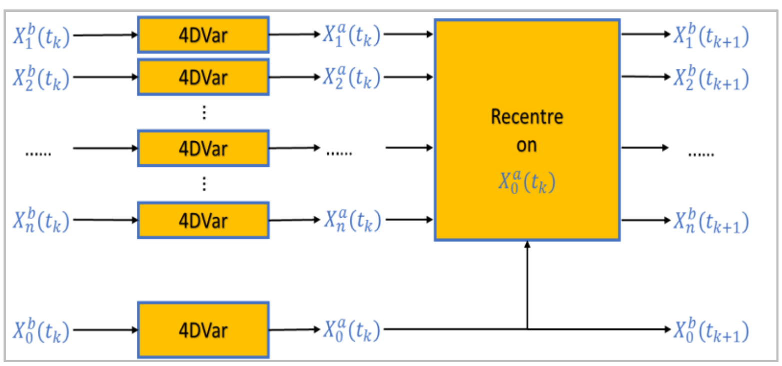

Figure 3: Asymmetric re-centred EDA concept

b. Hybrid DA Augmented Control Variable

Currently the **B** correlation matrix is 70% static, 30% flow-dependent at large scale and 7% static, 93% flow-dependent at small scale. A hybrid B is being tested where the flow-dependence comes from a physical space B from EDA members with localisation. The implementation based on augmented control vector following [4] with a localisation length scale from [5].

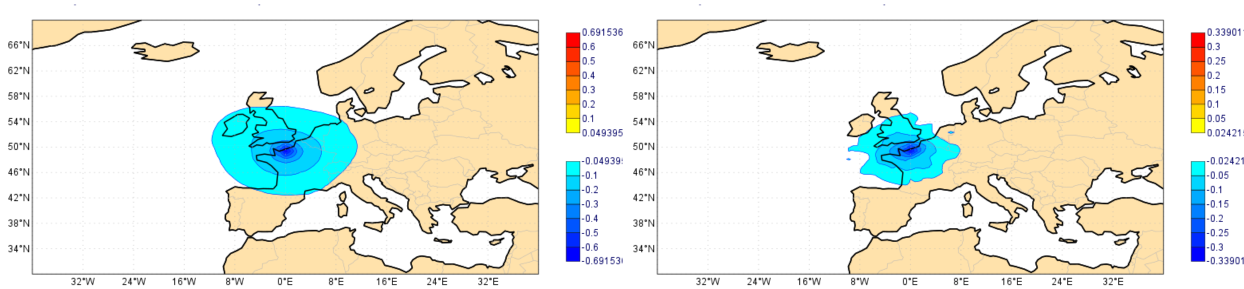

Figure 4: Temperature increment for a single observation at 0°E, 50°N for (left panel) pure static background error covariances and (right panel) EDA-based covariances with 350 Km localisation

c. Distributed use of observations in the EDA

EDA accounts for observation uncertainty by adding perturbations to the observations. An alternative, where possible, is to select different observations sets for each member. In the observation subset method the EDA members are more accurate as the observations are unperturbed. This method allows fuller use of high density observations by using spread to estimate uncertainty.

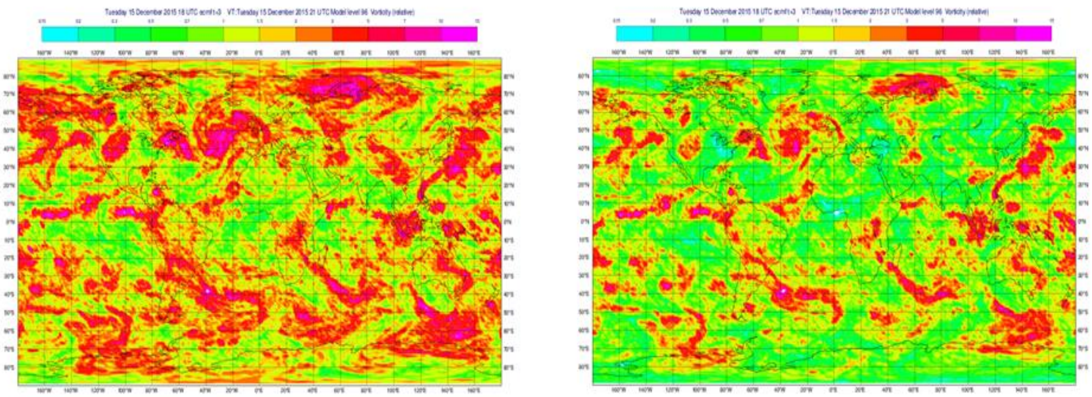

Figure 6: Vorticity spread at model level 96 (ca. 500 hPa, unit 10^{-5} s^{-1}) from four analyses with different observation selection and no added perturbations (right) compared to the equivalent spread from four EDA members using perturbed observations, as is done operationally (left). The structure in the spread is comparable demonstrating the feasibility of the idea.

d. Coupled DA

The Coupled ECMWF Re-Analysis (CERA) system was developed under the EC FP7 ERA-CLIM2 project for 20th century re-analysis [6]. Eresmaa et al. at ITSC-21 discuss CERA and the coupled re-analysis in the satellite era. CERA is being developed for an operational configuration.

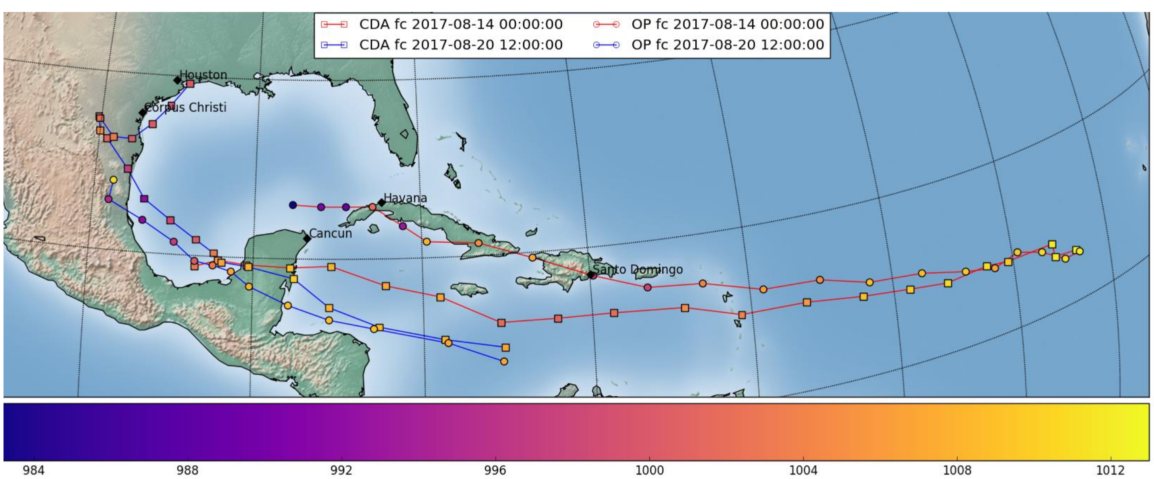

Figure 5: A technical demonstration of coupled DA on forecasts for hurricane Harvey. Two hurricane Harvey forecasts from initialised states with (CDA, squares) and without (OP, circles) ocean-atmosphere coupling.

e. Overlapping 12 hour 4D-Var

Currently the 12h DA windows are sequential. The overlapping configuration simplifies operations and allows for longer assimilation windows to be considered, e.g. 24 hours. In addition, more frequent updates will be possible, if affordable. The necessary separation of $\mathbf{X_b}$ and $\mathbf{X_{fg}}$ allows more flexibility, which enables the soft re-centring option to be tested in EDA.

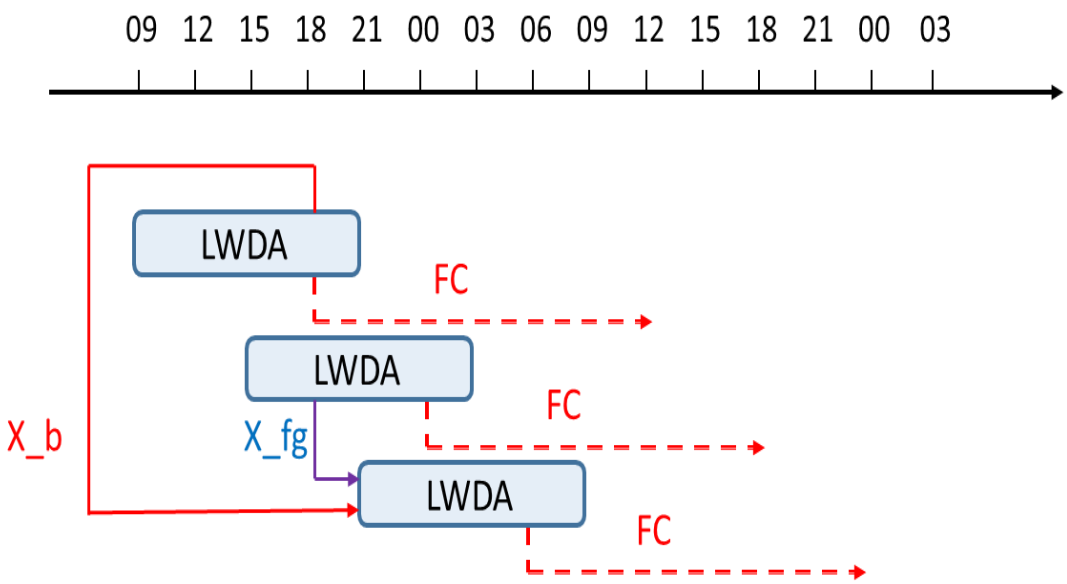

Figure 7: Overlapping 4D-Var configuration

f. Object Orientated Prediction System (OOPS)

The Object Orientated Prediction System (OOPS) was developed at ECMWF, principally by Yannick Trémolet and Mike Fisher. It has a C++ abstract layer that can run elements of the data assimilation system without needing to know what is being done inside each routine. ECMWF's Integrated Forecasting System (IFS) is now being developed to be called by OOPS. This has necessitated significant refactoring of the IFS Fortran code. ECMWF expect OOPS-IFS to be operational in 2019.

OOPS-IFS brings some significant benefits to the current IFS framework:

- Easier to develop and test alternative DA solvers;
- Enable approaches such as the saddle point solver to be tested with a full system;
- Potential for re-use of DA modules by different earth system components;
- Reduce interdependencies in the code;
- DA will be run as a single executable, reducing I/O costs;
- The OOPS code is open source under an Apache-2 license

OOPS is an international effort, involving ECMWF, Météo-France and the HIRLAM community. The Apache-2 license enables wider collaboration with other entities e.g. JCSDA, and the academic sector. Academic partners can benefit from the simplified models to test new DA approaches.

3. Conclusions

4D-Var has been a driving force behind analysis improvements in the last 20 years, in particular enabling effective use of radiances from infrared and microwave sounders. 4D-Var is continuing to improve through development and improvement of the core 4D-Var system as well as the ensemble hybrid **B** [3]. Significant science improvements, such as wavelet **B** for humidity, continue to regularly deliver large forecast accuracy improvements. Many new scientific improvements to the EDA 4D-Var configuration are currently underway. A major project, namely OOPS, is modernising the code to enable new science, increased flexibility and improved scalability of 4D-Var. A state of the art 4D-Var system with a hybrid ensemble component, and outer-loop coupling with ocean, sea-ice and land data assimilation, remains key to ECMWF's data assimilation strategy.

References

[1] Bouttier F, 2001: ECMWF Tech. Memo. 368

[2] Fisher, M., 2003, ECMWF seminar, September 2003. 45-63.

[3] Bonavita M., *et al.* 2017: ECMWF Tech. Memo. 800.

[4] Lorenc, A.C. 2003, doi:10.1256/qj.02.131.

[5] Ménétrier B. and T. Auligné, 2015: doi:10.1175/MWR-D-14-00404.1.

[6] Laloyaux P. *et al*. 2016, doi:10.1002/qj.2629