Institution: University of Surrey

Unit of Assessment: UOA 10 Mathematical Sciences

Title of case study:

Exploiting nonlinearity in operational data assimilation for weather prediction

1. Summary of the impact (indicative maximum 100 words)

Data assimilation is playing an ever increasing role in weather forecasting. Implementing *four-dimensional variational data assimilation* (4DVAR) is part of the long term strategy of the UK Met Office.

In this case study, an idealised 4DVAR scheme, developed by a team from the Universities of Surrey and Reading working with the UK Met Office, based on the integration of Hamiltonian dynamics and nonlinearity into data assimilation, has now been taken up by the Met Office. It is being used to evaluate options for improving operational 4DVAR. The simplicity of the scheme developed by this team has facilitated careful analyses of some generic problems with the operational model. The outcome includes direct impact on the environment and indirect impact on the economy, both through improvements in weather forecasting.

2. Underpinning research (indicative maximum 500 words)

Data assimilation (DA) is a technique for combining mathematical models of physical systems with measurements of those systems, in order to establish either the state of the system, or the parameters in the models. Such techniques have been used extensively in weather and climate prediction. 4DVAR calculates a forecast that best fits the available observations of weather, to within the observational error over a period of time. Since there will inevitably be insufficient data to calculate the present state of the atmosphere with certainty, data assimilation research often focuses on ways of using auxiliary information in the forecasting algorithms.

Surrey's Ian Roulstone (Professor of Mathematics) has been working in the area of DA, in the context of weather prediction, for over ten years. The motivation for this particular project was twofold: how to utilise conservation laws to mollify the problem of sparse data coverage in situations where nonlinearity becomes important, and how to rectify the inability of the operational scheme to represent rapidly growing modes. The approach taken, motivated by a Met Office strategy for evaluating new research directions by testing ideas on systems that are simpler than the full operational forecasting model, was to study a simple nonlinear system with the key attributes of nonlinearity and conservation laws as well as unstable modes, namely the 2- and 3-body problem [4].

It had been known for some time that conservation laws provide a rational basis for incorporating new observational data into forecast models, but the methodology was hitherto somewhat ad hoc, and often difficult to implement. In an earlier study of Wlasak, Nichols & Roulstone [2], potential vorticity (PV), an important conserved quantity, was exploited in a DA scheme for a simplified shallow water model, and improvements were found when it was used to project observational data onto the important modes of atmospheric motion. Hence, a more systematic study was in order.

The aim of the underpinning research [1,3,4] was to establish whether Hamiltonian properties of nonlinear dynamical systems could be exploited more generally in the formulation of 4DVAR. The results presented in [1,3] demonstrated conclusively that invariants of dynamical systems could be systematically incorporated into 4DVAR schemes. In particular, the inability of the operational scheme to represent rapidly growing modes and the problem of forecasting poorly observed modes have been studied in the simplified model, and new ways to improve the operational models have been formulated. The model developed in [1,3] supports rapidly growing

perturbations, so it is suitable for investigating why the so-called analysis error does not project strongly onto the rapidly growing modes of the forecast.

The underpinning research continued with the EPSRC-CASE funded 2009 PhD thesis of Alison Rudd on "*The effect of nonlinearity on the variational assimilation of satellite observations using a simple column model*", where nonlinearity, in the context of DA, was studied in a columnar model which was still simplified but closer to the meteorological context. The results were published in [5], showing that nonlinearity also dramatically affected the "tangent linear" assumption. The Leader of the Met Office unit, tasked with exploitation of satellite data in numerical weather prediction, wrote "*Rudd's work showed that we needed to sharpen our approach to understanding the basis for assimilating satellite radiances from cloudy regimes*." In current work with EngD student David Fairbairn (joint between Surrey and the Met Office) the aim is to apply the theory to models closer to the operational system [6].

The Surrey team is composed of Ian Roulstone (Professor), Sylvain Delahaies (NERC-funded Postdoc, 2008-present), Andrew Lorenc (Visiting Professor), Alison Rudd (EPSRC-CASE supported PhD student, completed in 2009), David Fairbairn (EPSRC supported EngD, started in 2010). The Reading team consists of Nancy Nichols (Professor), Amos Lawless (Lecturer) and Laura Watkinson (EPSRC-CASE supported PhD, completed in 2006).

3. References to the research (indicative maximum of six references)

1. L.R. Watkinson, A.S. Lawless, N.K. Nichols & I. Roulstone (2005) *Variational data assimilation for Hamiltonian problems*, International Journal Numerical Methods in Fluids, **47** 1361-1367. DOI: 10.1002/fld.844
2. M. Wlasak, N.K. Nichols & I. Roulstone (2006) *Use of potential vorticity for incremental data assimilation,* Quarterly J Royal Meteorological Society, **132**, 2867-2886. DOI: 10.1256/qj.06.02
3. L.R. Watkinson, A.S. Lawless, N.K. Nichols & I. Roulstone (2007) *Weak constraints in four-dimensional variational data assimilation,.* Meteorologische Zeitschrift, **16**, 767-776. DOI: 10.1127/0941-2948/2007/0249
4. I. Roulstone (2006) *Data assimilation and the 2- and 3-body problems*, Oberwolfach Reports **39**, 2356-2359.
5. A.C. Rudd, I. Roulstone, & J.R. Eyre (2012) *A simple column model to explore anticipated problems in variational assimilation of satellite observations*, J Env. Mod. & Software **27-28**, 23-29. DOI: 10.1016/j.envsoft.2011.10.001
6. D. Fairbairn, S.R. Pring, A.C. Lorenc, & I. Roulstone (2013) *A comparison of 4D-Var with ensemble data assimilation methods*, Q.J. Roy. Met. Soc. (in press, published online in May 2013). DOI: 10.1002/qj.2135

The research on DA has been supported by EPSRC (EP/C0006208/1, which looked at stochastic perturbations in DA), NERC through the National Centre for Earth Observation (£550k, 2008-13): Ian Roulstone is national co-theme leader of the DA Theme in NCEO. http://www.nceo.ac.uk

Roulstone presented a talk and report [4] on Hamiltonian methods in Data Assimilation at the Oberwolfach programme on the "*Mathematical Theory and Modelling in Atmosphere-Ocean Science*" in August 2006. Although that meeting was focussed on potential improvements to forecasting techniques, Roulstone's talk precipitated a discussion about a future Oberwolfach programme on the mathematics of data assimilation. That programme was approved and the meeting was held in December 2012.

An example of secondary impact is the application to modelling of the terrestrial carbon cycle. A recent talk on this was given by Delahaies at the EGU in April.

1. S. Delahaies, I. Roulstone & N. Nichols (2013) *A regularization of the carbon-cycle data-fusion problem*, Geophysical Research Abstracts **15**, EGU2013-4087-1.

The impact was facilitated by the visit of Gordon Inverarity (of the UK Met Office) to participate in a workshop on data assimilation held at the University of Surrey in October 2007.

2. G. Inverarity *Theoretical foundations of data assimilation using nonlinear forecast models,* Talk at the Surrey themed seminar series, 10th October 2007

4. Details of the impact (indicative maximum 750 words)

The principal impact of the case study has been in the take-up by the Met Office. It is using the simple model to identify key weaknesses in DA algorithms. Following the publication of the results [1,3], the Met Office DA group recognised the value of using the 3-body problem in an attempt to reconcile the theoretical limitations of 4DVAR with its practical success in situations far from those for which it is valid. The calculation of the background error covariance statistics of the variables used in data assimilation schemes is a crucial part of the algorithm. In an email communication, the Leader of the Met Office DA Unit wrote to Ian Roulstone "*The work has had a large influence on our 4DVAR strategy, as it probably explains why some of our recalculations of Cov were successful (and were implemented).*"

The Met Office recognised that the 4D VAR scheme for the 3-body problem, developed by Roulstone et al., would serve as a useful test bed in which "challenging but realistic scenarios" could be studied. The Met Office team set out to examine whether the standard method of calculating the background penalty was still optimal in the presence of model error. The relative simplicity of the 3-body problem (and the fact that in an idealised setting the notion of 'truth' can be defined precisely) enabled them to study this question in detail, and the outcome proved to be of major significance. Analysis of the standard 4DVAR technique applied to the 3-body problem revealed that the widely accepted notion of calculating the background error covariance from the mismatch between forecasts to estimates of the truth was flawed. Better forecasts could be obtained by excluding the mismatch between the forecast and the truth resulting from systematic model error.

The follow-up by the Met Office DA unit has had far-reaching implications: not only is the Met Office now pursuing several lines of research to improve the calculations of Cov, but they are also having to review how they conduct forecast verification. Recognising that the current calculation of Cov in the operational model was therefore flawed, implementing even a simple correction, mimicking the techniques applied to improve the simulation of the 3-body problem, led to marked improvements in forecasts. The Leader of the DA unit at the Met Office states;

"*As a result of subsequent projects carried out by the Met Office using this model, major investments have been made in improving the operational covariance model, which have had an impact on forecast accuracy, and new methods of generating forecast and analysis ensembles are being actively studied*."

The 3-body problem facilitated a thorough mathematical analysis of the implementation of the background error covariance term in 4DVAR. The two timescales of the 3-body problem, which play a key role in the definition, and representation, of short-range and long-range forecast errors, the inherent Hamiltonian properties, and the fact that the 'true state' of the system can be defined precisely in a toy model, enabled the Met Office DA unit to identify weaknesses in the standard formulation of 4DVAR in which the background is designed to approximate the 'truth'. The salient new idea, namely that the background term should be replaced by a 'regularization factor' (reported in the follow-up papers, Cullen (2010a, 2010b), cited below), the optimal choice of which minimises the short range forecast errors, was formulated on the basis of the study of the 3-body problem. This regularization procedure has also been employed by Delahaies et al [7] in a 4DVAR

approach to modelling the terrestrial carbon cycle.

5. Sources to corroborate the impact (indicative maximum of 10 references)

The impact is corroborated by emails from the Leader of the Met Office unit, tasked with exploitation of satellite data in numerical weather prediction, and emails plus a letter from the Leader of the Data Assimilation unit at the Met Office.

- Leader of the DA unit at the Met Office. Provided statement.

Following on from the results of the case study, results of the Met Office DA unit research give practical guidance as to ways of treating analysis error, and the choice of regularisation that should be adopted in operational 4D-VAR schemes. This follow up Met Office work is published in the open literature.

- Cullen, MJP. *A demonstration of 4D-Var using a time-distributed background term,* Q. J. R. Meteorol. Soc. **136** 1301-1315 (2010a).
- Cullen, MJP. *A demonstration of cycled 4D-Var in the presence of model error,* Q. J. R. Meteorol. Soc. **136** 1379-1395 (2010b).

and can be corroborated by;

- Leader of the unit tasked with exploitation of satellite data in numerical weather prediction at the Met Office. Contact details provided.

The implications of the research in the context of climate modelling can be found in the following technical report.

- D. Pearson (2011) *A Primer on data assimilation, parameter estimation and automatic differentiation with examples,* Hadley Centre Tech Report **85**, http://www.metoffice.gov.uk/media/pdf/k/o/HCTN_85.pdf