Data assimilation in earth system model development.

Ian G Enting

MASCOS, The University of Melbourne, Australia.

Abstract: Techniques of automatic differentiation can facilitate data assimilation in many types of modeling. In the development of the land-surface component of ACESSS (Australian Community Earth System Simulator) a MASCOS project, supported by the ARC Network for Earth System Science is implementing and deploying automatic differentiation techniques in the land-surface model development.

Data assimilation

Development of a state-of-the-art earth system model has to address the close coupling between the hydrological cycle and the terrestrial carbon cycle. Considered separately, these system components share some important characteristics: (i) high spatial and temporal heterogeneity (ii) multiple characteristic time-scales (iii) disparate data available for model calibration, and (iv) primarily local interactions within the subsystem. While this last characteristic can be used to simplify calculations for calibration and data assimilation, the other three characteristics make land-surface data assimilation more challenging than "conventional" meteorological data assimilation. With the carbon and water cycles tightly coupled in the land surface, the difficulties of data assimilation become exacerbated.

At least two distinct types of role for data assimilation in land surface modelling can be identified. Firstly a range of operational needs require "real-time" assimilation of data into a land-surface model. Hydrological information is important for operational forecasting on both synoptic and seasonal time scales. For the carbon component, an important application in the medium to long term is the detection and quantification of carbon-climate feedbacks. The strength of such feedbacks over the 21st century is subject to great uncertainty. Estimates range from 20 to 200 ppm for the net amount of extra CO_2 expected to be added to the atmosphere this century from such feedbacks (Friedlingstein *et al*, 2006). Operational data assimilation, addressing the status of the terrestrial carbon system, can also be important in applications in Natural Resource Management (e.g. Steffan, 2005). Secondly, data assimilation techniques can be used as part of the model calibration process. This represents a recursive form of parameter estimation as an alternative to a "batch" approach of simultaneously adjusting all parameters to fit all data. Formally, a recursive application of Bayes rule shows that the two approaches would, if applied exactly, give equivalent results. However, in large problems, where "batch" approaches are computationally infeasible, even recursive approaches usually need to make simplifying approximations. Introducing such approximations destroys the formal equivalence. Consequently, data assimilation techniques need to be evaluated in terms of both accuracy and computational efficiency.

Advanced meteorological data assimilation takes one of two generic forms: "variational" and "ensemble" (Kalnay 2003). The choice between these involves a range of trade-offs. Ensemble techniques require large ensembles to span the relevant number of dimensions representing the uncertainty of the assimilated state. variational techniques rely on gradient calculations (adjoint models) that are difficult to produce.

Data assimilation for the terrestrial carbon cycle can be seen as part of a progression of inverse problems. The classic CO_2-inversion calculation (Enting, 2002) aims to deduce surface CO_2 fluxes as a function of space and time, given measurements of CO_2 concentrations. Early CO_2 inversions were divided between "batch" calculations in a synthesis mode, that interpolated source distributions and recursive "mass-balance" inversions that had to spatially interpolate concentration data. Over the last decade, variational techniques based on adjoints of transport models have led to a convergence of these approaches. Such calculations are being progressively modified to become "process inversions" that, rather than estimate carbon fluxes, estimate parameters in models that specify carbon fluxes.

Such calibration of processes has the potential to be extended to detection of processes. As noted above, a wide range of data streams are available for interpreting the carbon cycle (Canadell *et al*, 2000). These various diverse data streams will have correspondingly diverse statistical characteristics, as emphasised by Raupach *et al* (2005). What is needed in practice for data assimilation is a quantification of the various considerations raised by Raupach *et al.*

The MASCOS project on automatic differentiation is based on operator overloading (Greiwank, 2000). Initial applications have involved sensitivities of simple climate models. Forward sensitivities have been calculated to determine the incremental (marginal) utility of geosequestration as climate benefit over time (Enting *et al*, 2007). This involved automatic calculation of what is known as the tangent linear model. Backward sensitivities (or gradients) have been calculated as a way of attributing responsibility for climate change to emissions at particular times (Enting, 2005). Such calculations underlie the so-called "Brazilian Proposal" which, as part of the Climate Convention process, sought to have reduction targets for greenhouse gas emissions set on the basis of each nation's relative responsibility for global warming. The backward sensitivity corresponds to the adjoint of the tangent linear model. The particular results given by Enting (2005) were based on repeated "brute-force" application of the forward (tangent linear) calculation, derived by operator overloading in C++.

The geosequestration calculations used a Fortran-90 implementation. This scalar version has been extended to a vector version for use in land-surface modelling. The specific objective of the vector-oriented implementation is to achieve computational efficiency when the procedures are used in earth system modelling. In order to explore the strengths and weaknesses of deploying automatic differentiation in terrestrial carbon modelling, the MASCOS project is working through a spectrum of systematically increasing two distinct aspects of model complexity. The first aspect represents an aggregation dimension. The intended progression is to go from global-scale to biome type to grid-based to sub-grid tiling and then to distributions within tiles. The second aspect represents a process dimension starting from representing only carbon pools, adding hydrology, nutrients, competition, and possibly with isotopic diagnostics.

Within each class of models, differentiation is being deployed for a range of problems. The first is that of initialisation. Formally this means solving for state variables that lead to zero rate of change. One seasonal forcing is imposed one requires state variables that lead to zero year-to-year change. Solving such equations directly, rather than integrating to a steady state, has a long history in box models of the carbon cycle (Enting and Pearman, 1987). In earth system modelling, accelerated spin-up of ocean models is a standard procedure. Initial studies show that automatically computed derivatives can be used to accelerate the equilibration of nutrient-coupled carbon pool models. Automatic differentiation is also valuable for assessing parameter sensitivity in models. Data assimilation problems can be facilitated by careful analysis of the problem structure. The so-called "adjoint compilers" actually aim to produce code for generic gradient calculations. While these include adjoint models as special cases, there are problems for which the generic capability introduces unnecessary complexity and where simpler targetted formalisms can be more effective.

References

Canadell, J.P, H. A. Mooney, D. D. Baldocchi, J. A. Berry, J. R.Ehleringer, C. B. Field, S.T. Gower, D. Y. Hollinger, J. E, Hunt, R. B. Jackson, S. W. Running, G. R. Shaver,W. Steffan, S. E, Trumbore, R. Valentine and B.Y. Bond, 2000. Carbon metabolism of the terrestrial biosphere: a multi-technique approach for improved understanding. *Ecosystems,* **3**, 115–130.

Enting, I, G, 2002. *Inverse Problems in Atmospheric Constituent Transport*. CUP, Cambridge.

Enting, I.G., 2005. Automatic differentiation in the analysis of strategies for mitigation of global change. Presented at MODSIM05. Available online at MSSANZ website

Enting, I,G, and D. M. Etheridge and M. J. Fielding, 2007. A perturbation analysis of the climate benefit of geosequestration. Submitted to *International Journal of Greenhouse Gas Control*.

Enting, I. G, and G. Pearman, 1987. Description of a one-dimensional carbon cycle model calibrated using techniques of constrained inversion. *Tellus*, **39B**, 459–476.

P. Friedlingstein, P. Cox, R. Betts, L. Bopp, W. von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, and N. Zeng, 2006, Climate–Carbon Cycle Feedback Analysis: Results from the C^{4}MIP Model Intercomparison, *Journal of Climate,* **19**, 3337–3353.

Greiwank, A., 2000. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. SIAM, Phildelphia

Kalnay, E, 2003. Atmospheric Modelling, data Assimilation and Predictability. CUP, Cambridge.

Kowalczyk, E.A., Y.-P. Wang, J.L. McGregor and G. Abramowitz, 2006. CSIRO Atmosphere Biosphere Land Exchange model for use in climate models and as an offline model. CSIRO Marine and Atmospheric Research. Technical Report. Available on-line at CMAR website

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. de Fries, M. Heimann, D. S. Ojima, S. Quegan and C. C. Schmullius, 2005. Model-data synthesis in terrestrial carbon observations: methods, data requirements and data uncertainty specifications. *Global Change Biology*, **11**, 378–397.

Steffan, W. (ed.)., 2005. *Blueprint for Australian Carbon Cycle Research.* Australian Greenhouse Office, Canberra.

Acknowledgements.

The ARC Centre of Excellence for Mathematics and Statistics of Complex Systems)+(MASCOS) is supported by the Australian Research Council. Ian Enting’s fellowship at MASCOS is supported in part by CSIRO. The implementation of the Fortran modules for automatic differentiation was greatly assisted by Mark Fielding, supported by the ARC Network for Earth System Science.