Combining flux and ecosystem data with land surface models: the role of FLUXNET

Mathew Williams, Paul Stoy, Andrew Richardson, Enrico Tomelleri, Cathy Trudinger

From June 4-6, 2008, a meeting was convened in Edinburgh that combined scientists from the FLUXNET and carbon modelling communities. The objective of the meeting was to determine ways in which eddy covariance (EC) measurements might be used to improve the representation of ecosystem processes in land surface models (LSMs). Common examples of LSMs include Orchidee (Morales et al. 2005) and the Lund-Potsdam-Jena (LPJ) model (Sitch et al. 2003). The meeting focused particularly on the application of model-data fusion (MDF) techniques to bring together measurements and models in a rigorous, statistically-based analytical framework.

The participants recognized that currently model building is largely uncoordinated, and that there are no clear criteria for model improvement. Model outputs often lack assessments of error and bias, and this reduces their utility. However, large arrays of data are now available, particularly time series information such as EC data, which can provide insights into the model processes critically requiring improvement. MDF techniques provide the basis, theoretically, for linking models and observations optimally. Model-data fusion for carbon cycle science has been discussed before (Raupach et al. 2005). However, the application of MDF to ecological problems remains in its initial stages, with considerable knowledge gaps regarding practicalities.

The Edinburgh meeting addressed a series of key issues related to linking eddy covariance data to LSMs using MDF. The questions we addressed included:

1- What are the strengths and weaknesses of the various MDF approaches?

2- How should observation and model error be determined?

3- How do we assess observational and model bias?

4- What ancillary data (including Earth Observation) can and should be involved?

5- How can the LSM and FLUXNET communities best collaborate?

For accessing presentations from the meeting see www.carbonfusion.org/LSM.html.

A focused discussion that addresses these questions in detail is forthcoming as a result of the meeting (Williams et al. in prep). Below we discuss in brief some of the issues raised and some background.

The principles of model-data fusion

The mathematical foundations of DA can be derived from Bayes' Theorem (equation 1). The posterior probability distribution of the model *M* given observations *O*, written $p(M|O)$ is defined as the product of the prior model state $p(M)$ and the probability of *O* given previous model state $p(O|M)$ normalized by the probability of the observations, $p(O)$ (Lorenc 1995):

$$p(M|O) = \frac{p(O|M)p(M)}{p(O)} \qquad (1)$$

The important concept here is that distributions are required. The terms in (1) are probability distribution functions (pdfs), not simple measurements. Every observation and model state and rate term must have an associated error term, best expressed as a pdf. Quantifying uncertainty is just as important as making the measurement, as this allows multiple data streams to be incorporated into the model estimate with correct weightings (Williams et al. 2005, Figure 1).

Error estimation

Many of the fundamental papers in the EC literature address the issue of error estimation (Goulden et al. 1996), and this area of research continues to develop. Random error in instantaneous (half-hourly) flux observations has been argued to follow a Laplacian distribution (Hollinger and Richardson 2005). However, there is evidence that flux error is normally distributed but heteroscedastic, with variance increasing with flux magnitude (Lasslop et al. 2008). This current research focus on EC error is timely for MDF studies, given the requirement by MDF for accurate assessment of flux error. While there is progress in assessing random error, systematic error is less easily evaluated. We suggest that MDF provides an opportunity to assess systematic EC error, by linking other, independent data series to EC data through the model structure.

Multiple constraints

Net flux measurements do not provide enough information to constrain the simulation and

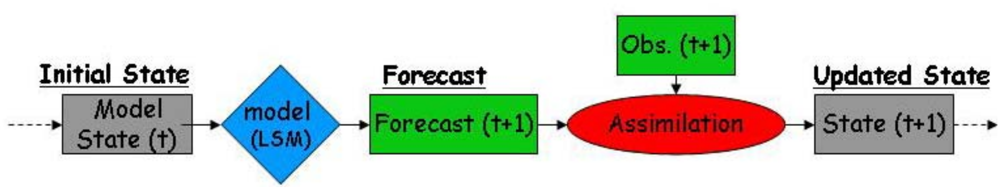

Figure 1: A schematic of sequential assimilation of data into a state-space land surface model (LSM), using an Ensemble Kalman filter (EnKF). A model ensemble (with associated error) is run forward from its initial state at time *t* to provide a forecast at time *t*+1 (likewise with associated error), and then combined with observations (again, with error) in the assimilation step (see equation 1). In the EnKF, the model error covariance matrix propagates the information from the assimilated observation(s) throughout the model state vector.

cont. on page 10

Combining flux and ecosystem data with land surface models

cont. from page 9

estimation of component gross fluxes – a net flux can be generated by many combinations of photosynthesis and ecosystem respiration. MDF approaches can provide useful quantification of this problem of *equifinality,* by generating pdfs of model parameters and model states. Narrow posterior pdfs can indicate that the EC data provided substantial information on the process associated with the particular parameter.

MDF can further indicate how parameter confidence intervals are reduced according to the type, amount and quality of data involved in the fusion. For LSMs, ancillary information like leaf area index, biomass, or energy fluxes, provide useful extra constraints on model processes (Williams et al. 2005). The MDF process serves to check consistency among multiple independent data types, for instance between EC and biometric data. It is also possible to assimilate earth observation (EO) data, such as MODIS time series (Quaife et al. 2008). MDF with EO data opens up possibilities for extrapolating information from FLUXNET sites to their surroundings, with assessments of uncertainty.

Model-data fusion intercomparison

Two recent studies have focussed on comparing the capabilities of MDF techniques. The OptIC experiment used a simplified model and synthetic time series data, i.e. data generated from the model, with noise added. Multiple groups employing a range of MDF techniques attempted to estimate the parameters that were used to generate the synthetic data (Trudinger et al. 2007). The REFLEX experiment used a daily model of carbon fluxes and pools and provided both synthetic data and EC data from two FLUXNET sites. Participants used a range of MDF approaches, and attempted both to estimate parameters and determine fluxes of C and changes in stocks over multiple years (Fox et al. in review). REFLEX participants generated confidence intervals on all estimates, so it was possible to assess whether MDF approaches correctly assessed confidence by comparison with the synthetic data for which the "truth" was known. For more information on REFLEX see www.carbonfusion.org/Reflex.html.

The interaction between 'measurers' and modelers

The requirements for flux (and stock) data and error introduces a challenge for FLUXNET, which has to date excelled at reporting flux data, but less so at quantifying, reporting and organizing flux observation error, ancillary data, and particularly ancillary data error. These data requirements are heavy, and the perception may exist that delivery of data to modellers represents asymmetry in the modeller-measurer interaction (Figure 2). But measurers also benefit from LSMs, which provide a broader justification for observations and provide a mechanism to extrapolate observations to larger scales (Jung et al. 2007), to help answer the questions related to regional land-atmosphere interaction.

The future of MDF and FLUXNET

Modellers and measurers have a common goal of quantifying biosphere atmosphere fluxes (Figure 2). The meeting identified the need for a new model-data fusion exercise for LSMs using FLUXNET data. This exercise would build on previous model-data fusion projects, such as OptIC and REFLEX, using LSMs with a range of complexity, synthetic and EC data from several key biomes, including sites with long (>10 years) data time series, and with a range of auxiliary data. A major output of such an exercise would be forecasts of carbon dynamics *with confidence intervals* over multiple years beyond the period of MDF. The overarching goals of this exercise would be to identify systematic error in models and data, and to improve our skills in generating defensible and realistic model confidence intervals.

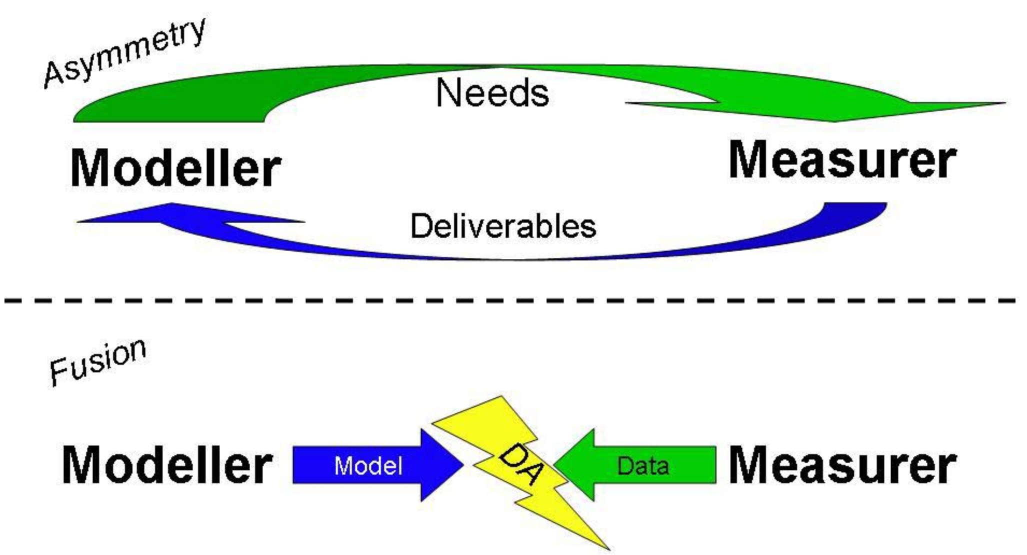

Figure 2: Asymmetry in the conventional modeller/measurer interaction can be replaced by the dual requirements for both models and measurements, with associated error terms, in a model-data fusion scheme.

cont. on page 11

Combining flux and ecosystem data with land surface models

cont. from page 10

Acknowledgements

We would like to thank FLUXNET, the CarbonFusion Programme funded by the National Environmental Resource Council (NERC) UK, and the workshop participants: Nuno Carvalhais (Lisbon), Martin Jung (MPI), Ray Leuning (CSIRO), Yiqi Luo (Oklahoma), David Pearson (UKMO), Philippe Peylin (LSCE), Shaun Quegan (Sheffield), Markus Reichstein (MPI), Hans Verbeeck (Antwerp/LSCE), and Ying-Ping Wang (CSIRO).

contact: Mathew Williams
Institute of Atmospheric and Environmental Sciences
University of Edinburgh
mat.williams@ed.ac.uk

Literature

Fox, A., M. Williams, A. D. Richardson, D. Cameron, J. H. Gove, T. Quaife, D. Ricciuto, M. Reichstein, E. Tomelleri, C. Trudinger, and M. T. V. Wijk. in review. Using model-data fusion to characterise confidence in parameterizations, analyses and forecasts of terrestrial C dynamics.

Goulden, M. L., J. W. Munger, S.-M. Fan, B. C. Daube, and S. C. Wofsy. 1996. Measurements of carbon storage by long-term eddy correlation: Methods and a critical evaluation of accuracy. Global Change Biology **2**:169-182.

Hollinger, D. Y., and A. D. Richardson. 2005. Uncertainty in eddy covariance measurements and its application to physiological models,. Tree Physiology **25**:873-885.

Jung, M., M. Vetter, M. Herold, G. Churkina, M. Reichstein, S. Zaehle, P. Ciais, N. Viovy, A. Bondeau, Y. Chen, K. Trusilova, F. Feser, and M. Heimann. 2007. Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models. Global Biogeochemical Cycles **21**:GB4021.

Lasslop, G., M. Reichstein, J. Kattge, and D. Papale. 2008. Influences of observation errors in eddy flux data on inverse model parameter estimation. Biogeosciences Discussions **5**:751-785.

Lorenc, A. C. 1995. Atmospheric Data Assimilation. Scientific Paper No 34, UK Meteorological Office, Bracknell.

Morales, P., M. Sykes, I. C. Prentice, P. Smith, B. Smith, H. Bugmann, B. Zierl, P. Friedlingstein, N. Viovy, S. Sabate, A. Sanchez, E. Pla, C. A. Gracia, S. Sitch, A. Arneth, and J. Ogee. 2005. Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. Global Change Biology **11**:2211-2233.

Quaife, T., P. Lewis, M. De Kauwe, M. Williams, B. E. Law, M. D. Disney, and P. Bowyer. 2008. Assimilating Canopy Reflectance data into an Ecosystem Model with an Ensemble Kalman Filter. Remote Sensing of the Environment **111**:1347-1364.

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

Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke, and S. Venevsky. 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology **9**:161-185.

Trudinger, C. M., M. R. Raupach, P. J. Rayner, J. Kattge, Q. Liu, B. Pak, M. Reichstein, L. Renzullo, A. D. Richardson, S. H. Roxburgh, J. Styles, Y. P. Wang, P. Briggs, D. Barrett, and S. Nikolova. 2007. OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. Journal of Geophysical Research-Biogeosciences **112**.

Williams, M., A. Richardson, M. Reichstein, P. Stoy, P. Peylin, H. Verbeeck, N. Carvalhais, M. Jung, D. Hollinger, J. Kattge, R. Leuning, Y. Luo, E. Tomelleri, C. Trudinger, and Y.-P. Wang. in prep. Improving land surface models with FLUXNET data.

Williams, M., P. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology **11**:89-105.