Fifth Monitoring Committee Meeting

Understanding the Information Content in Diverse Observations of Forest Carbon Stocks and Fluxes for Data Assimilation and Ecological Modeling
NERC CASE partnership with Forest Research

E. Pinnington

Room ESSC 176, 10am Jan 7, 2016

1 Introduction

Forest ecosystems play an important role in sequestering human emitted carbon-dioxide from the atmosphere and therefore greatly reduce the effect of anthropogenic induced climate change. For that reason understanding their response to climate change is of great importance. Measurements of forest carbon balance are now routinely made in forests across the world using micrometeorological techniques, with many other relevant observations such as leaf area index and standing biomass also available [Baldocchi, 2008]. Many efforts have been made to combine this data with models of forest carbon balance using data assimilation techniques in order to improve modelled estimates [Fox et al., 2009, Niu et al., 2014, Quaife et al., 2008, Richardson et al., 2010, Zobitz et al., 2011, 2014]. Currently, however, the optimal set of observations for understanding the carbon balance of a forest is not known. The aims of this PhD are:

- Understanding which observations provide models of forest carbon balance with most information in a data assimilation framework, focusing on the CASE partners research site Alice Holt.
- Finding a better way to quantify background and observation errors and their correlations.
- Investigating the effect of disturbance on the Alice Holt research forest. The disturbance occured in 2014 when one side of the forest was thinned and the other side left unmanaged. The flux tower measuring net ecosystem exchange of CO_2 is situated on the boundary between the two sides.

In the previous report we had decided to move away from work using the Data Assimilation Linked Ecosystem Carbon model (DALEC) and to start using the new version of this model (DALEC2 [Bloom and Williams, 2015]). This was because DALEC2 can be parameterised for both deciduous and evergreen forests, whereas the version of DALEC used in previous work was an evergreen only model. The CASE partners research site (Alice Holt) is a mainly deciduous site. DALEC2 was implemented in a Four-Dimensional Variational data assimilation routine (4D-Var) for joint parameter and state estimation. It was shown that after assimilating 2 years of observations of the Net Ecosystem Exchange (NEE) of CO_2 from the Alice Holt flux site the forecast (2001-2006) root mean square error between modelled NEE and observed NEE fell from $4.37\text{gCm}^{-2}\text{day}^{-1}$ to $2.12\text{gCm}^{-2}\text{day}^{-1}$.

Since completing the previous report, a paper draft has been completed on the work implementing 4D-Var with the DALEC2 model. The final draft of this paper is attached to the monitoring committee report. In this paper the effect that background and observation error statistics and their correlations have on the assimilation have also been investigated. We found that specifying parameter-state error correlations in background error statistics can improve data assimilation forecast results significantly. Including correlations in time between observation error statistics was also found to improve assimilation forecast results. This work was expected to be completed in the new year (2016). However it has been brought forward ahead of completing the observation information content experiments and report.

In the last six months I have also been conducting a field work campaign taking leaf area index measurements at the CASE partners research flux site Alice Holt. These observations will contribute to both the information content experiments and the work investigating the effect of the disturbance on the Alice Holt research forest.

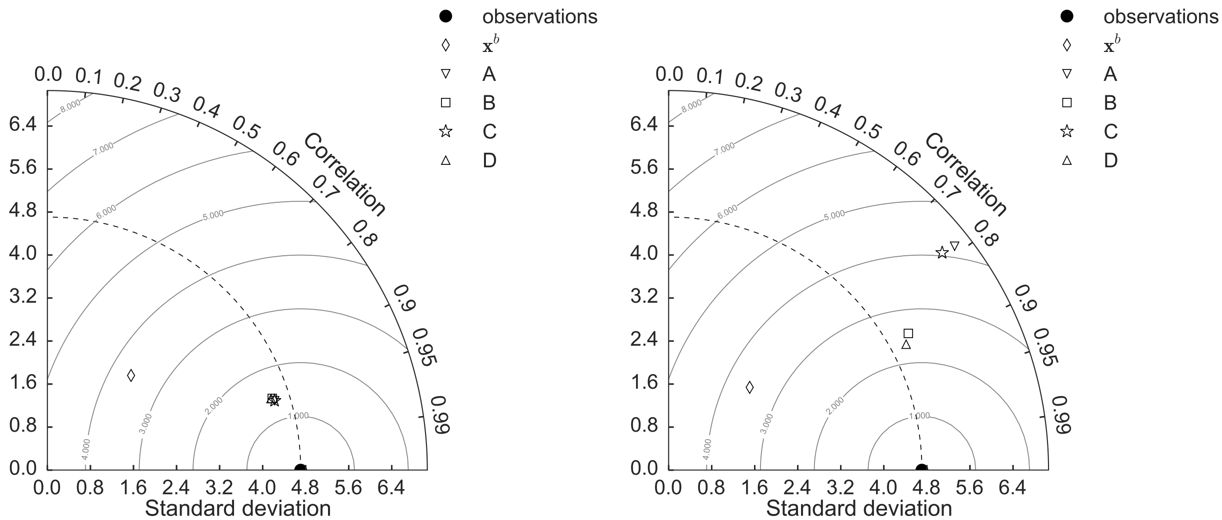
2 4D-Var

In the attached paper draft the DALEC2 functional ecology model is implemented for joint parameter and state estimation. The assimilation routine is then subjected to rigorous testing to check for correctness. We outline novel methods to create a background error covariance matrix (describing our knowledge of the error in prior model estimates before data assimilation) that includes correlations and a novel method to include time correlations between observation errors in the observation error covariance matrix. The background and observation error covariance matrices are largely treated as diagonal in carbon balance model data assimilation. We show that including these correlations can improve forecasts of NEE significantly. The root mean square error in the 14 year forecast of daily NEE being reduced by 44%, decreasing from 4.22 $\text{gCm}^{-2}\text{day}^{-1}$ to 2.38 $\text{gCm}^{-2}\text{day}^{-1}$. We conducted 4 experiments to investigate the impact of the different matrices on the assimilation. In table 1 we show which matrices were used in each experiment.

Experiment	$\mathbf{B}_{diag}$	$\hat{\mathbf{R}}_{diag}$	$\mathbf{B}_{corr}$	$\hat{\mathbf{R}}_{corr}$
A	×	×		
B		×	×	
C	×			×
D			×	×

Table 1: The combination of error covariance matrices used in each data assimilation experiment. $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ are the diagonal background and observation error covariance matrices, with no correlations and $\mathbf{B}_{corr}$ and $\hat{\mathbf{R}}_{corr}$ are the background and observation error covariance matrices including correlations.

In figure 1 we show Taylor diagrams displaying a statistical comparison of the four experiment and background analysis (Jan 1999 - Dec 1999) and forecast (Jan 2000 - Dec 2013) results with the observations of NEE. Here the radial distances from the origin to the points are proportional to the pattern standard deviations and the azimuthal positions give the correlation coefficient between the modelled and observed NEE [Taylor, 2001]. If a model predicted the observations perfectly it would have a correlation coefficient of 1 and a radial distance matching that of the observations (represented by the dotted line). Figure 1a shows that all the experiments give very similar results in the analysis window (Jan 1999 - Dec 1999) with all the experiment points closely grouped on top



(a) Analysis (Jan 1999 - Dec 1999)

(b) Forecast (Jan 2000 - Dec 2013)

Figure 1: Taylor diagrams displaying statistical comparison of the four experiment and background analysis (Jan 1999 - Dec 1999) and forecast (Jan 2000 - Dec 2013) results with observations of NEE ($gCm^{-2}day^{-1}$). The dotted line represents the standard deviation of the observations and the contours represent values of constant root mean square error between model and observations. Here $\mathbf{x}^b$ is the background augmented state vector.

of each other, whereas figure 1b shows the significant difference between the experiment results in the forecast (Jan 2000 - Dec 2013). We see that the model most closely matching the observations is from experiment D where both correlated matrices have been used in the assimilation algorithm.

3 Field work

In order to address the third aim of the PhD from section 1 a field work campaign has been undertaken to measure Leaf Area Index (LAI) at the Alice Holt flux site. I have used three different methods to measure LAI:

- Using a ceptometer and an additional Photosynthetically Active Radiation (PAR) sensor. Here we measure below canopy PAR using the ceptometer while logging above canopy PAR using a data logger and PAR sensor positioned outside the canopy. We can then calculate LAI using the above and below canopy readings and a set of equations relying on some assumptions [Fassnacht et al., 1994]. The ceptometer represents the quickest method for estimating LAI.

- Using hemispherical photographs as shown in figure 3. Hemispherical photographs show a complete view of the sky in all directions, from these images we use software which can calculate the proportion of visible sky as a function of sky direction (gap fraction) this can then be used to calculate LAI [Jonckheere et al., 2004].

- Using litter traps. Here we place litter traps under the canopy which catch the litter in a bag these bags are changed every week during the litter fall period and the litter sorted into species. We then dry the litter in an oven at 70°C and weigh it. Towards the end of the season we scan a subsample for each species of 100 leaves to find an area we then dry and weigh each subsample, a relationship between dry weight and leaf area can then be built and used to infer the total LAI once the whole seasons litter has been collected. This method of LAI calculation is the most time consuming.

For this fieldwork we want to capture both the thinned and unthinned sides of the forest. For this reason I have taken measurements along three transects spanning both sides of the site. Figure 2 shows a map of the Alice Holt flux site with the three transects and 10m sampling points marked. For the ceptometer I took measurements every 10m giving us 435 readings in total. However the hemispherical photographs were taken every 50m as they are more time consuming. This gave a total of 89 images. I mapped out these points in the forest using a GPS unit and pink tree spray paint.

I deployed a total of 6 litter traps with 3 positioned in the western side and 3 in the east. However Forest Research already have 10 other traps in operation giving us some valuable extra data. The 6 litter traps is not enough to describe the LAI for the research site [Kimmins, 1973]. We will use these litter traps as a point of comparison and validation for the ceptometer and hemispherical photograph estimates of LAI made at the same locations. The litter traps have been sampled weekly throughout the autumn and winter which gives us an estimate to the rate of litter fall. This is another observation we can use for data assimilation and model validation.

I have finished taking observations with the ceptometer and hemispherical camera of summer peak LAI. However I am continuing the process of managing the litter traps weekly as this seasons leaf fall period has been longer than usual. I will be taking measurements of spring green up next year in order to constrain the phenology in DALEC2 more accurately. I am now processing the observations I have already taken to find an LAI product.

At the start of the new year (2016) I will be conducting a biomass survey of the Alice Holt flux site. This will use the method of point centred quarters [Dahdouh-Guebas and Koedam, 2006] to determine an estimate of the woody biomass for both sides of the forest.

We plan to use all observations that I have made and with the available observations of NEE for the thinned and unthinned sides of the forest (partitioned using a flux footprint model) for data assimilation. Two versions of the DALEC2 model will be parameterised (for the thinned and unthinned sides of the forest). This will allow us to see if there is a difference between the optimised parameters for the thinned/unthinned versions of DALEC2. We will test if by just assimilating NEE we can pick out a difference in LAI between the two sides.

We will investigate the effect of the disturbance further by parameterising DALEC2 for the thinned section of the forest both before and after the thinning occurred using the record of partitioned NEE observations. This will again allow us to see if the optimised parameters are consistent with the known changes to the forest ecosystem.

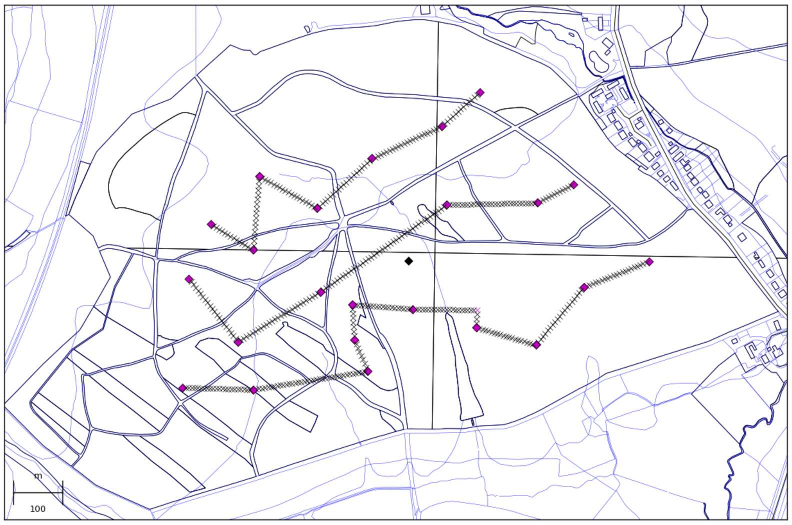
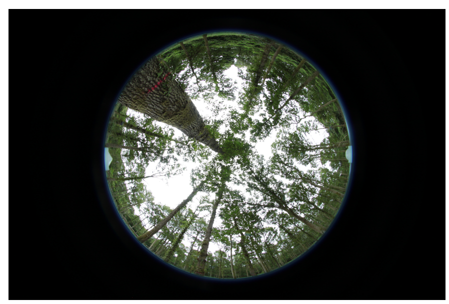


Figure 2: Map of the Alice Holt flux site. The crosses are 10m sampling points with the purple diamonds being Forest Research mensuration plots where measurements of woody biomass are made. The black diamond shows the position of the flux tower on the boundary between the thinned (West) and unthinned (East) sides of the forest.

(a) Unthinned



(b) Thinned

Figure 3: Hemispherical photographs from the Alice Holt flux site showing the difference between the thinned and unthinned sides of the forest.

4 Current work and future plans

I am currently in the process of getting the attached paper finished and we hope to have it submitted early in the new year (2016). I have begun repeating the information content experiments using DALEC2. these experiments had been previously conducted using the original DALEC model. In addition to the measures previously used for DALEC (SIC and DFS), I will be extending this work to include other measures. The influence matrix measures the sensitivity of the analysis in observation space to the observations [Cardinali et al., 2004]. The adjoint technique proposed by Langland and Baker [2004], approximates the sensitivity of a scalar forecast error norm to the observations . In Cardinali et al. [2004] the data assimilation problem is assumed to be Gaussian with a linear function mapping the state to observation space ($\mathbf{H}$), such that,

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b), \tag{1}$$

where $\mathbf{K}$ is the Kalman gain matrix, $\mathbf{K} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$. The influence matrix is then defined as,

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}} = \mathbf{K}^T\mathbf{H}^T. \tag{2}$$

In order to consider observations over a 4D-Var time window we rewrite equation 1 as,

$$\mathbf{x}_a = \mathbf{x}_b + \hat{\mathbf{K}}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_b), \tag{3}$$

using the defined matrices in the attached paper, with $\hat{\mathbf{K}} = (\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}} + \mathbf{B}^{-1})^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}$. Equation 2 then becomes,

$$\mathbf{S} = \frac{\partial \hat{\mathbf{H}}\mathbf{x}_a}{\partial \hat{\mathbf{y}}} = \hat{\mathbf{K}}^T\hat{\mathbf{H}}^T. \tag{4}$$

In equation 3 we have also assumed that we have a linear model, $\mathbf{M}_{i,0}$, evolving our state from time 0 to time i. We know that our model is in fact nonlinear but have tested the tangent linear hypothesis for DALECV2 in the attached paper and have seen it to be a good approximation.

Using these measures and an observing system simulation experiment that I have developed, I will produce another report investigating the best set of observations for understanding the carbon balance of a forest. We will explore the impact of using the correlated background and observational error covariance matrices defined in the attached paper on these information content measures. Using the twin experiments we will also investigate the effect of the correlated error covariance matrices on recovering the truth after assimilating a set of synthetic observations. I am also continuing with the fieldwork campaign as described in section 3. I have currently processed the ceptometer data to produce a Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) product as shown in figure 4. Work to produce an LAI product is still ongoing. From figure 4 we can see that the FAPAR produce demonstrates the distinct difference in canopy structure for the different sections of the forest site. Other future work will be based on the attached thesis outline and Gantt chart.

Figure 4: Map of the Alice Holt flux site displaying values of FAPAR for the 10m sampling points of the ceptometer in the described field work campaign in section 3.

5 Professional and Academic Development

5.1 Masters Courses

- MAMB10 (Data Assimilation) - 85%
- MAMNSO (Numerical Solutions to Ordinary Differential Equations) - 79%
- MTMG02 (Atmospheric Physics) - 66%
- MTMG49 (Boundary Layer) - 72%
- MTMD01 (Environmental Data Visualization) - 78%
- MTMD02 (Operational Data Assimilation) - 70%

5.2 Transferable Skills

During my PhD I have taken part in the following courses, workshops and activities:

- 28/01/2014 - Basic Statistics Refresher - RRDP
- 31/03/2014-01/04/2014 - Land Data Assimilation workshop at UCL - ESA
- 23/04/2014-25/03/2014 - Correlated Observation Errors in Data Assimilation Workshop - ESA
- 13/05/2014 - Social Media - Bloggs, Twitter and Your Online Presence - RRDP

- 29/05/2014 - How to Write a Paper - RRDP
- 25/06/2014-26/06/2014 - Software Carpentry Course - Git and Python
- 10/07/2014-11/07/2014 - Forest Research - Helped with field work LiDAR
- 21/07/2014-01/08/2014 - Fluxcourse Summer School - University of Colorado
- 29/09/2014-03/10/2014 - NERC course - Software Development for Environmental Scientists Level 1
- 08/10/2014-10/10/2014 - Environment YES - NERC "dragon's den" type competition at Syngenta, Jesops Hill
- 17/12/2014 - Presentation at Maths for Planet Earth Industry day
- 24/02/2015 - Reading Soil Centre Workshop - What can Land Surface Models do for you?
- 23/03/2015-27-03/2015 - NERC course - Software Development for Environmental Scientists Level 2
- 11/03/2015 - Quo Vadis presentation
- 08/09/2015-11/09/2015 - RSPSoc conference University of Southampton - Presented a poster
- 24/09/2015 - Department poster presentation - Recieved an honourable mention for poster on "Understanding the information content in observations of forest carbon balance"
- 02/11/2015-03/11/2015 - BES Ecosystems and Climate Change Mitigation Conference, Charles Darwin House, London - Presented a poster
- 02/12/2015 - RMetS SE centre meeting, Reading Town Hall - Invited to give a presentation after receiving honourable mention at the department poster presentation
- 20/01/2016 - Submitting your thesis electronically: what you need to know - RRDP
- 03/02/2016 - Open access and research data - RRDP
- 12/02/2016 - How to write a thesis - RRDP

5.3 Demonstrating

During my PhD I have helped demonstrate on the following courses:

- 15/09/2014-1909/2014 - NERC Data assimilation for environmental scientists training course
- 16/02/2015-20/02/2015 - NERC Software Development for Environmental Scientists Level 1
- 20/04/2015-23/04/2015 - MT26E Surface Energy Exchange Practicals

References

D. Baldocchi. Turner review no. 15.'breathing'of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, 56 (1):1–26, 2008.

A. A. Bloom and M. Williams. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a modeldata fusion framework. *Biogeosciences*, 12(5):1299–1315, 2015. ISSN 1726-4189. doi: 10.5194/bg-12-1299-2015. URL `http://www.biogeosciences.net/12/1299/2015/`.

C. Cardinali, S. Pezzulli, and E. Andersson. Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 130(603):2767–2786, 2004. ISSN 1477-870X. doi: 10.1256/qj.03.205. URL `http://dx.doi.org/10.1256/qj.03.205`.

F. Dahdouh-Guebas and N. Koedam. Empirical estimate of the reliability of the use of the point-centred quarter method (pcqm): Solutions to ambiguous field situations and description of the pcqm+ protocol. *Forest Ecology and management*, 228(1):1–18, 2006.

K. S. Fassnacht, S. T. Gower, J. M. Norman, and R. E. McMurtric. A comparison of optical and direct methods for estimating foliage surface area index in forests. *Agricultural and Forest Meteorology*, 71(1):183–207, 1994.

A. Fox, M. Williams, A. D. Richardson, D. Cameron, J. H. Gove, T. Quaife, D. Ricciuto, M. Reichstein, E. Tomelleri, C. M. Trudinger, et al. The reflex project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, 149(10):1597–1615, 2009.

I. Jonckheere, S. Fleck, K. Nackaerts, B. Muys, P. Coppin, M. Weiss, and F. Baret. Review of methods for in situ leaf area index determination Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121(1-2):19–35, 2004. ISSN 01681923. doi: 10.1016/j.agrformet.2003.08.027.

J. Kimmins. Some statistical aspects of sampling throughfall precipitation in nutrient cycling studies in british columbian coastal forests. *Ecology*, pages 1008–1019, 1973.

R. H. Langland and N. L. Baker. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 56(3):189–201, 2004. ISSN 02806495. doi: 10.1111/j.1600-0870.2004.00056.x.

S. Niu, Y. Luo, M. C. Dietze, T. F. Keenan, Z. Shi, J. Li, and F. S. C. Iii. The role of data assimilation in predictive ecology. *Ecosphere*, 5(5):art65, 2014. ISSN 2150-8925. doi: 10.1890/ES13-00273.1. URL `http://www.esajournals.org/doi/abs/10.1890/ES13-00273.1`.

T. Quaife, P. Lewis, M. De Kauwe, M. Williams, B. E. Law, M. Disney, and P. Bowyer. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, 112(4):1347–1364, 2008. ISSN 00344257. doi: 10.1016/j.rse.2007.05.020.

A. D. Richardson, M. Williams, D. Y. Hollinger, D. J. Moore, D. B. Dail, E. A. Davidson, N. A. Scott, R. S. Evans, H. Hughes, J. T. Lee, et al. Estimating parameters of a forest ecosystem c model with measurements of stocks and fluxes as joint constraints. *Oecologia*, 164(1):25–40, 2010.

K. E. Taylor. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106(D7):7183, apr 2001. ISSN 0148-0227. doi: 10.1029/2000JD900719. URL `http://doi.wiley.com/10.1029/2000JD900719`.

J. Zobitz, A. Desai, D. Moore, and M. Chadwick. A primer for data assimilation with ecological models using markov chain monte carlo (mcmc). *Oecologia*, 167(3):599–611, 2011.

J. M. Zobitz, D. J. P. Moore, T. Quaife, B. H. Braswell, A. Bergeson, J. a. Anthony, and R. K. Monson. Joint data assimilation of satellite reflectance and net ecosystem exchange data constrains ecosystem carbon fluxes at a high-elevation subalpine forest. *Agricultural and Forest Meteorology*, 195-196:73–88, 2014. ISSN 01681923. doi: 10.1016/j.agrformet.2014.04.011. URL `http://dx.doi.org/10.1016/j.agrformet.2014.04.011`.