

Sixth 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 1L36, 10am, 6th July 2016

1 Introduction

The land surface and oceans are responsible for removing around half of all human emitted carbon-dioxide from the atmosphere and therefore mediate the effect of anthropogenic induced climate change. Terrestrial ecosystem carbon uptake is the least understood process in the global carbon cycle [Ciais et al., 2014]. It is therefore vital that we improve understanding of the carbon uptake of terrestrial ecosystems and their response to climate change in order to better constrain predictions of future carbon budgets. Measurements of ecosystem 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]. These observations provide a valuable resource for insight into the current state of the carbon dynamics for existing terrestrial ecosystems.

In order to produce future predictions of forest carbon balance we must use mathematical models describing these systems. However, these model predictions are often unreliable and have high error. The mathematical technique of data assimilation allows us to combine current observations and model predictions to find optimised parameters and states for our system and produce forecasts with greater confidences. Many efforts have been made to combine available data with models of forest carbon balance using data assimilation techniques [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. In data assimilation it is important to characterise the errors in prior estimates and observations as accurately as possible. In ecosystem model data assimilation the treatment of prior and observational errors has largely been simple, assuming all errors are independent and uncorrelated. Making the assumption of independent uncorrelated errors, when correlations do exist, has been shown to have a negative impact on data assimilation results in numerical weather prediction [Smith et al., 2009, Weston et al., 2014]. The aims of this PhD are:

- Finding a better way to quantify prior and observation errors and their correlations.
- 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. Also assessing the effect of specifying error correlations on the information content in observations.

- Investigating the effect of disturbance on the Alice Holt research forest. The disturbance occurred in 2014 when one side of the forest was thinned and the other side left unmanaged. The flux tower measuring net ecosystem exchange of CO₂ is situated on the boundary between the two sides.

In the previous report we discussed the draft paper that had been produced on the work implementing the DALEC2 forest carbon balance model [Bloom and Williams, 2015] in a 4D-Var data assimilation routine for joint parameter and state estimation. We had moved on to work with the DALEC2 model from DALEC1 as it can be parameterised for both evergreen and deciduous forest sites (The CASE partners site is predominately deciduous) whereas the DALEC1 model from previous work was evergreen only. The presented draft paper explored the effect that different representations of prior and observation error statistics had on the assimilation. We found that specifying parameter-state error correlations in prior error statistics can improve data assimilation forecast results significantly. Including correlations in time between observation errors was also found to improve assimilation forecast results. We also discussed the extensive field work campaign that had been undertaken in the previous report, where three different methods were used to measure the leaf area index at the CASE partners research flux tower site Alice Holt.

Since completing the last report the paper draft was finalised and submitted to the Agricultural and Forest Meteorology journal [Pinnington et al., submitted 2016]. The reviews for the paper have been received with the paper requiring minor revisions. These revisions have been addressed and we have recently re-submitted the manuscript and response to reviewers. The fieldwork described in the last report has also now been completed. Another fieldwork campaign to measure woody biomass, using the method of point-centred quarters [Dahdouh-Guebas and Koedam, 2006], at Alice Holt has also been completed.

2 Paper on 4D-Var with DALEC2 and prior and observation error correlations

In the attached paper Four-Dimensional Variational data assimilation (4D-Var) is implemented with the DALEC2 functional ecology model 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. In our experiments we assimilate a single year (Jan 1999 - Dec 2000) of NEE observations and then run a 14 year (Jan 2000 - Dec 2013) forecast to judge the skill of our models after assimilation. We show that including these correlations in our assimilation scheme can improve forecasts of NEE significantly. The root mean square error in the 14 year forecast of daily NEE is reduced by 44%, decreasing from 4.22 g C m⁻² day⁻¹ to 2.38 g C m⁻² day⁻¹. We conducted 4 experiments to investigate the impact of the different matrices on the assimilation. In Table 3 we show which matrices were used in each experiment.

One of the main comments from the received reviews was that, from the included plots, it was hard to judge the skill of the model forecast for NEE after assimilation. It was suggested that we include plots of the model-data differences averaged over the multiple years of the forecast. These are shown in Figure 1. From these plots we can see that the largest errors in our posterior model forecast occur at the start of the season from April to June. This is a result of not capturing

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.

the phenology of the site correctly in our model forecast. We can see that these errors have been significantly reduced in Figure 1d where all correlations are included in the assimilation. Even when including all correlations we still find the largest errors in the period from April to June, as a result of mistiming the date of green-up for our forest site. This suggests a flaw with our phenology model in DALEC2. We have therefore begun work on including a new phenology model in DALEC2 that will more closely match the phenology of the site.

3 Field work

In the last monitoring report it was discussed that a field work campaign had been undertaken to measure Leaf Area Index (LAI) at the Alice Holt flux site, using three different methods. These measurements were taken along three transects capturing both thinned and unthinned sides of the forest. Since completing the last report I have completed this fieldwork campaign and am now beginning to process the collected data.

I have also conducted a biomass survey of the Alice Holt flux site. This survey used the Point-Centred Quarters Method (PCQM) [Dahdouh-Guebas and Koedam, 2006]. I have sampled along the three established transects. This will allow us to determine an estimate of the woody biomass for both sides of the forest. The observations that have been taken on the PhD are shown in Table 2.

These observations will be used in the third results chapter of the thesis. We plan to use all observations that I have made 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. From work already carried out at Forest Research it has been noted that there is no significant change in NEE between the thinned and unthinned sides of the forest, when considering only the flux towers partitioned record of NEE measurements [Wilkinson et al., 2016]. These experiments will hopefully allow us to answer questions we currently have about the effect of disturbance, for example:

- Do we see evidence that the understory and remaining dominant trees are compensating

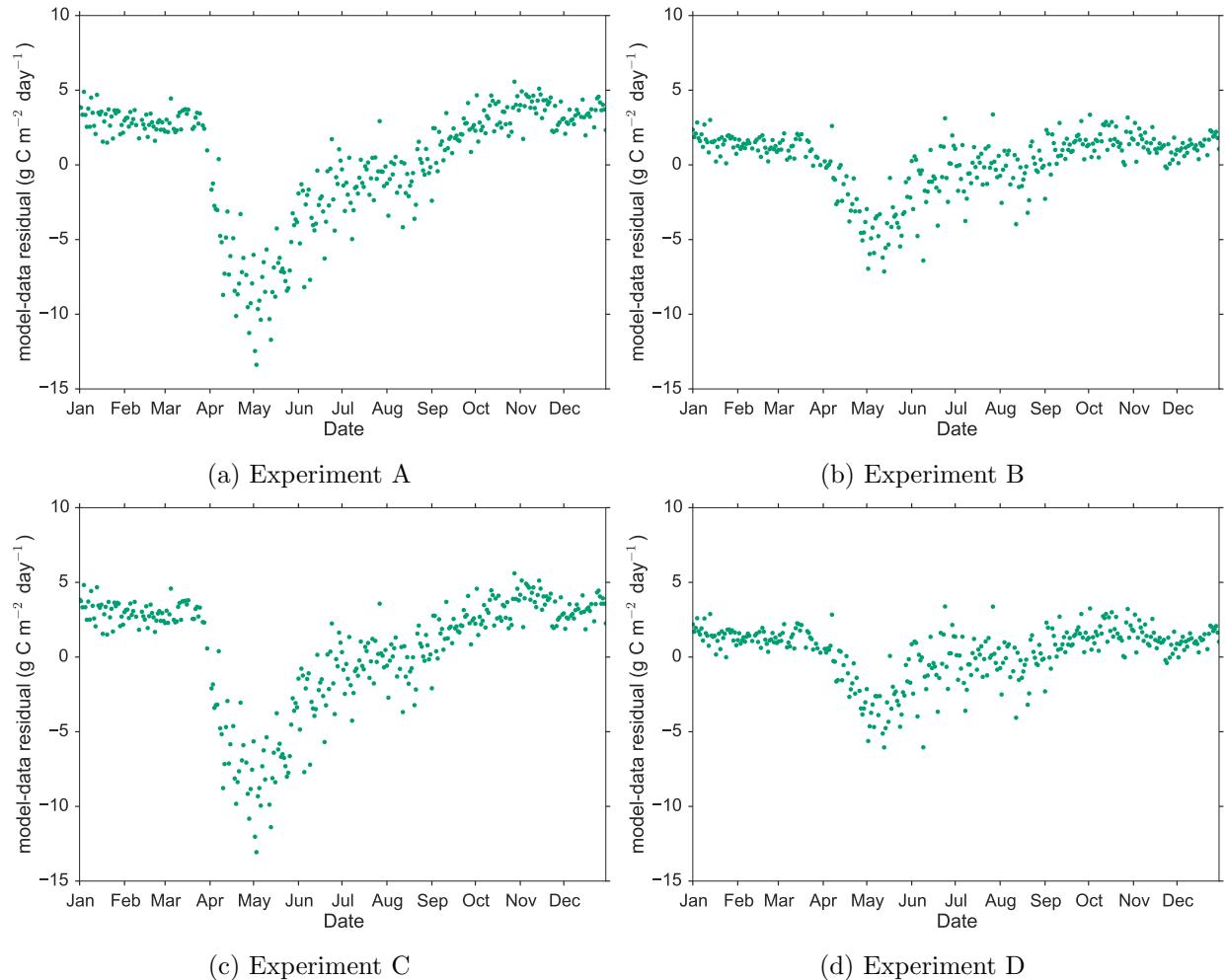


Figure 1: Net ecosystem exchange model-data differences for the four experiments. Here each point corresponds to the mean model-data difference for that day of the year over the 14 year model forecast (Jan 2000 - Dec 2013).

for the removed trees in the thinned section of the forest by demonstrating increased Gross Primary Productivity (GPP), as a result of reduced competition and higher light availability?

- Is heterotrophic respiration from the soil reduced for the thinned section of forest meaning that the NEE remains similar to that of the unthinned section of forest?

We hope that this work will form the basis for the next paper.

Observation	Method	Number of observations
LAI	Ceptometer [Fassnacht et al., 1994]	435
LAI	Hemispherical photographs [Jonckheere et al., 2004]	89
LAI	Litter traps [Kimmmins, 1973]	6 points measured throughout the season
Woody biomass	PCQM [Dahdouh-Guebas and Koedam, 2006]	114

Table 2: Observations taken during the PhD at the Alice Holt research forest.



Figure 2: Hemispherical photographs from the Alice Holt flux site showing the difference between the thinned and unthinned sides of the forest.

4 Observability and information content

A first draft of the second results chapter of the thesis on observability and information content for the DALEC2 ecosystem carbon model is close to completion. In this work we consider:

- The observability of the system for DALEC1 and DALEC2 with the available observations. This will allow us to see if we can find a unique analysis with the observational information alone.
- The analytic information content in carbon balance observations for evergreen ecosystems with DALEC1 and how this information content varies over a time window.
- The numerical information content in observations for DALEC2 for both evergreen and deciduous ecosystems over a time window.
- The effect of including error correlations on the information content in observations.

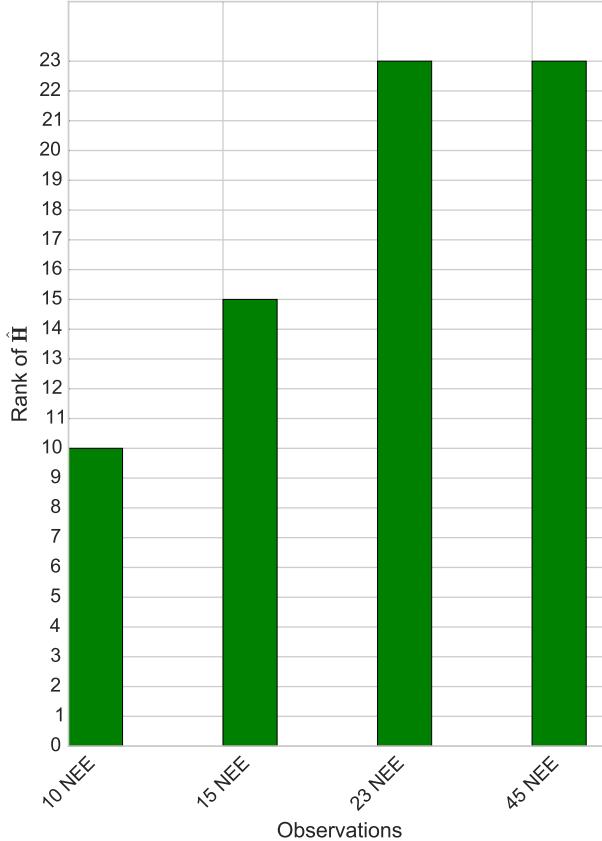


Figure 3: Rank of the observability matrix $\hat{\mathbf{H}}$ for increasing number of NEE observations over a years assimilation window. If the the rank of $\hat{\mathbf{H}}$ is equal to 23 (the size of the state vector) then our system is observable and we can find a locally unique minimum of the observational part of the cost function.

We find that for both DALEC1 and DALEC2 our system is observable when assimilating observations of NEE only (providing we have enough observations of NEE) as shown in Figure 3. This means that we have enough information from observations of NEE alone to find a locally unique analysis state over a specific assimilation window with no background term [Zou et al., 1992]. In practice a background term is typically included in the cost function for 4D-Var data assimilation which regularises the problem and means that we always have a unique solution.

We show that both analytic and numerical implementations of the information content measures (Shannon information content, degrees of freedom for signal and the sensitivity of the analysis to observations) give the same results. This gives us confidence that for DALEC2 our numerical implementation of these measures is correct. For DALEC1 we find temporal variation in information content for an evergreen ecosystem with observations made at times with higher temperatures having higher information content, as discussed in previous reports. We also show that including correlations in time between observation errors reduces information content. This may explain some of the results found in the attached paper, where including correlations in time between observations of NEE improves the model forecast by reducing overfitting to the observations in the assimilation window.

For DALEC2 we find that for an evergreen ecosystem, information content in NEE observations is most reliant on temperature. However, for a deciduous site some of the NEE observations with the highest information content come during the period of green-up. This makes physical sense as the NEE for a deciduous site is controlled by phenology much more than for an evergreen site. We again find that including time correlations between observation errors reduces information content. This work is completed and in the process of being written up.

One of the main issues with assimilating only NEE is that the measurement represents the difference between GPP and Total ecosystem Respiration (RT), with $\text{NEE} = \text{RT} - \text{GPP}$. This means that assimilation of NEE alone can result in the over/under-prediction of RT and GPP. We hope at the end of this results chapter to consider the assimilation of day and night time NEE by using a modified observation operator in the 4D-Var DALEC2 scheme. Here night time NEE is equal to RT as no GPP occurs during the night. We hope this will help us to better partition NEE between GPP and RT. We plan to investigate this using a set of twin experiments that have already been developed. If successful this will also prove helpful for the final results chapter investigating the effect of disturbance on the Alice Holt research forest.

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 - Received 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
- 15/03/2016 - How will employers interview you? - RRDP
- 17/04/2016-22/04/2016 - EGU General Assembly, Vienna - Gave an oral presentation in the BG2.8 working group “Developments in terrestrial biogeochemical models using model-data integration”



Figure 4: Presenting at EGU

- 12/05/2016 - Three minute thesis, Graduate School - Competed in the heats for the 3 minute thesis competition and progressed to the final at the Doctoral Research Conference
- 23/06/2016 - Participated in the three minute thesis final at the doctoral research conference Reading

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

Experiment	Uncorrelated prior errors, \mathbf{B}_{diag}	Uncorrelated observation errors, $\hat{\mathbf{R}}_{diag}$	Correlated prior errors, \mathbf{B}_{corr}	Correlation observation errors, $\hat{\mathbf{R}}_{corr}$
A	×	×		
B		×	×	
C	×			×
D			×	×

Table 3: 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.

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/>.
- P. Ciais, C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, et al. Carbon and other biogeochemical cycles. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 465–570. Cambridge University Press, 2014.
- 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. McMurtrie. 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.
- 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>.
- E. Pinnington, E. Casella, S. Dance, A. Lawless, J. Morrison, N. Nichols, M. Wilkinson, and T. Quaife. Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using four dimensional variational data assimilation. *Agricultural and Forest Meteorology*, submitted 2016.
- 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.
- P. J. Smith, S. L. Dance, M. J. Baines, N. K. Nichols, and T. R. Scott. Variational data assimilation for parameter estimation: application to a simple morphodynamic model. *Ocean Dynamics*, 59(5):697–708, 2009.
- P. Weston, W. Bell, and J. Eyre. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429, 2014.
- M. Wilkinson, P. Crow, E. L. Eaton, and J. I. L. Morison. Effects of management thinning on co2 exchange by a plantation oak woodland in south-eastern england. *Biogeosciences*, 13(8):2367–2378, 2016. doi: 10.5194/bg-13-2367-2016. URL <http://www.biogeosciences.net/13/2367/2016/>.
- 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>.

X. Zou, I. Navon, and F. Le Dimet. Incomplete observations and control of gravity waves in variational data assimilation. *Tellus A*, 44(4):273–296, 1992.