

# ASSIMILATING REFLECTANCE DATA INTO A ECOSYSTEM MODEL TO IMPROVE ESTIMATES OF TERRESTRIAL CARBON FLUX

T. Quaife<sup>a,\*</sup>, P. Lewis<sup>a</sup>, M. Disney<sup>a</sup>, M. De Kauwe<sup>a</sup>, M. Williams<sup>b</sup> and B. Law<sup>c</sup>.

<sup>a</sup> NERC Centre for Terrestrial Carbon Dynamics and Dept. of Geography, University College London, Pearson Building, Gower Street, London, WC1E 6BT UK - tquaife@geog.ucl.ac.uk

<sup>b</sup> NERC Centre for Terrestrial Carbon Dynamics and School of GeoSciences, Institute of Atmospheric and Environmental Sciences, Crew Building, Kings Buildings, University of Edinburgh, Edinburgh, EH9 3JN, UK.

<sup>c</sup> College of Forestry, Oregon State University, Corvallis, OR 97331, U.S.A.

**KEY WORDS:** Ensemble Kalman filter, MODIS, GORT, data assimilation, DALEC, GPP, NEP.

## ABSTRACT:

Ecosystem models are valuable tools for understanding the growth of vegetation, its response to climatic change and its role in the cycling of greenhouse gasses. Data Assimilation (DA) of synoptic coverage Earth Observation (EO) data into ecosystem models provides a statistically optimal mechanism for constraining the model state vector trajectory both spatially and temporally. EO “products” such as leaf area index (LAI) are attractive candidates for assimilation, but it is difficult to assign accurate uncertainty estimates to such products (a critical requirement of DA) and, more importantly, they are derived on the basis of assumptions that may be contradictory to those in the ecosystem model. An attractive alternative, therefore, is to assimilate reflectance data; the uncertainty in which is more easily understood. The assumptions made in generating the reflectance data are independent of assumptions in the ecosystem model and may consequently be treated as additional sources of uncertainty. To achieve this it is necessary to build a canopy reflectance model into the assimilation scheme. This paper describes the coupling of a canopy reflectance model to a simple ecosystem model. Reflectance data are assimilated over a boreal forest and improvements in predicted carbon fluxes are shown with comparison to field data. Previous work has highlighted problems of lost samples due to snow cover, resulting in poorly constrained flux estimates during winter months. This issue is addressed by incorporating a snow reflectance model. Results utilising the EnKF as a parameter estimator are also discussed.

## 1. INTRODUCTION

### 1.1 Background

Understanding and quantifying of the role of terrestrial vegetation in the carbon cycle is critical for climate change studies because of the feedbacks that exist with atmospheric CO<sub>2</sub> (Schimel et al., 2001). This in turn has major relevance for national and international policy (IPCC, 2001). Ecosystem models are an attractive tool for studying the terrestrial carbon cycle because they are based on process understanding. However, large uncertainties exist between models of terrestrial carbon dynamics (Churkina et al., 2005) despite advances in process based modelling (Law et al., 2001a; Rastetter, 2003) and improved networks of C flux measurements at the field scale (Valentini et al., 2000).

An issue when scaling such models up to regional or global levels is that they are not well constrained away from the field sites for which they are parameterised. An attractive option to tackle this problem is the use of Earth Observation (EO) data, which provides spatially and temporally synoptic data, to adjust model trajectories within a data assimilation scheme. Data assimilation methods have been used successfully for integrating EO data in numerical weather prediction models for some years but are only just gaining widespread attention in the terrestrial vegetation EO community. These techniques give a statistically optimal analysis of a model’s state vector and/or parameters against observations and allow for explicit representation of model and data errors.

This paper follows directly on from the work of Williams et al (2005) and Quaife et al. (2007). Williams et al (2005) present a simple ecosystem model (outlined below) and assimilate field observations using data assimilation techniques. The paper shows that the model is able to reproduce observed carbon fluxes well despite its simplicity by utilising large numbers of observations. Quaife et al. (2007) take this model and assimilate MODIS surface reflectance data into it as a demonstration of the potential use of EO data in such schemes. The authors argue for the use of reflectance over “high” level EO products (such as leaf area index, LAI) on the basis that it is easier to characterise errors in reflectance data and that high level products are typically generated using assumptions that contradict those in the ecosystem model itself. This latter point may be overcome by using surface reflectance products, as assumptions in its generation will be independent of those in the ecosystem model. Furthermore the required observation operator may be built to be entirely consistent with the ecosystem model itself.

Two issues in this approach highlighted by Quaife et al. (2007) are the lack of MODIS samples in winter months owing to snow/cloud contamination (resulting in a poorly constrained model) and the need for spatialised model parameters to allow for scaling to and beyond the regional scales. These issues are addressed in this paper.

---

\* Corresponding author.

## 2. METHOD

### 2.1 Field data

The field data for this study were acquired at the Metolius forest site (44° 26' N, 121° 34' W) in Oregon that has been extensively characterised (Law, 2001a) and is part of the AmeriFlux network. It is dominated by *Pinus Ponderosa* and is regenerating since being clearfelled in 1978. The understory is mixed bitterbush (*Purshia tridentata*) and manzanita (*Arctostaphylos patula*).

Extensive ecological measurement campaigns have been carried out at this site and the collected data have been used to calibrate the DALEC model (Williams et al. 2005). These data include CO<sub>2</sub> flux measurements from an eddy covariance system (Anthoni et al., 2002), leaf area index (LAI; Law, 2001b), specific leaf area, soil respiration (Irvine et al., 2002) and above- and belowground biomass. Meteorological data (incident PAR, temperature and vapour pressure deficit) were also collected. These are used here to drive the ACM and DALEC models.

### 2.2 MODIS data

Level 2, collection 4 TERRA-MODIS 500m surface reflectance data (MOD09) were used as the observations in this study. Only bands 1 (red) and 2 (near infra red), centred at 646 nm and 848 nm were considered, although this work could easily be extended to include any number of optical wavebands. Uncertainties in the reflectance data were set as 0.004 and 0.015 for bands 1 and 2 respectively (Roy et al., 2005).

In total there were 316 acquisitions, over the 3 years (200-2002) of this study that contained no snow or cloud. Within this set there were no observations during winter months however. Extracting pixels flagged in the MODIS QA data as being snow contaminated provided an additional 18 observations spread roughly evenly over the winter months of the time period.

### 2.3 DALEC

Carbon pools and fluxes for were modelled over the three year period using the Data Assimilation Linked Ecosystem model, DALEC (Williams et al., 2005).

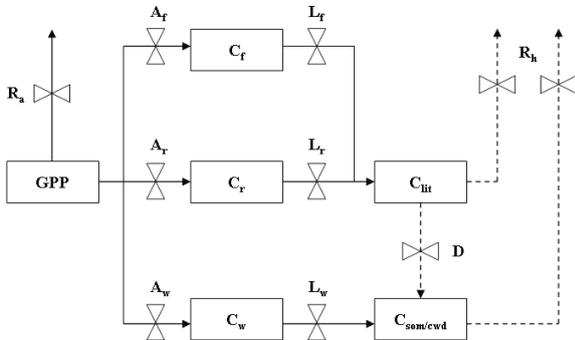


Figure 1. Schematic representation of the DALEC model.

Fig 1. Shows a diagrammatic representation of the DALEC model. The carbon pools are  $C_f$ ,  $C_r$ ,  $C_w$ ,  $C_{lit}$ , and  $C_{som/cwd}$  which represent foliar, root, woody, litter, and soil organic matter and

woody debris carbon, respectively. In addition there is a pseudo pool that represents the gross primary productivity (GPP). This is the uptake of carbon from photosynthesis. Fluxes are denoted  $R_a$  and  $R_h$  (autotrophic and heterotrophic respiration), where heterotrophic respiration is divided into  $D$  (decomposition from litter to soil organic matter), and  $L_f$ ,  $L_r$  and  $L_w$  (the rate of loss of foliar, fine root and woody carbon respectively). Dotted lines indicate a temperature dependant transform.

GPP was calculated using the Aggregated Canopy Model (ACM) of Williams et al. (1997) calibrated for a Ponderosa pine ecosystem. The net ecosystem productivity (NEP) is given by the difference of the GPP and the  $R_a$  and  $R_h$  modelled by DALEC. This provides a quantification of the strength of the ecosystem as either a source (-ve NEP) or a sink (+ve NEP) of CO<sub>2</sub>.

### 2.4 Ensemble Kalman Filter (EnKF)

The EnKF is a variant of the Kalman Filter designed to work with non-linear models due to Evensen (1994, 2003). It uses an ensemble of model states to represent error statistics in the model, as opposed to the explicit covariance matrix representation used by the Kalman Filter. It has the form:

$$\mathbf{A}^a = \mathbf{A} + \mathbf{A}'\mathbf{A}'^T\mathbf{H}^T(\mathbf{H}\mathbf{A}'\mathbf{A}'^T\mathbf{H}^T + \mathbf{R}_e)^{-1}(\mathbf{D} - \mathbf{H}\mathbf{A}) \quad (1)$$

where,  $\mathbf{A}$  is the model ensemble;  $\mathbf{A}'$  is the ensemble perturbation;  $\mathbf{R}_e$  is the covariance observation error;  $\mathbf{D}$  is the observation ensemble and  $\mathbf{H}$  is the observation operator. The superscript  $\mathbf{a}$  denotes the analysed ensemble and the superscript  $\mathbf{T}$  denotes a matrix transpose.

In this form it is not possible to assimilate canopy reflectance into an ecosystem model as top of canopy BRF is not a linear transform of a typical ecosystem model state vector (as required by the observation operator matrix  $\mathbf{H}$ ). Evensen, (2003) suggests an augmented state vector approach to handle non linear observations:

$$\mathbf{A}^a = \mathbf{A} + \mathbf{A}'\hat{\mathbf{A}}'^T\hat{\mathbf{H}}^T(\hat{\mathbf{H}}\hat{\mathbf{A}}'\hat{\mathbf{A}}'^T\hat{\mathbf{H}}^T + \mathbf{R}_e)^{-1}(\mathbf{D} - \hat{\mathbf{H}}\hat{\mathbf{A}}) \quad (2)$$

Where  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{A}'}$  are the model ensemble and perturbation matrices augmented with predictions of BRF. In effect the BRFs become part of the model during analysis. The augmented ensemble is formed by:

$$\hat{\mathbf{A}} = \mathbf{h}(\mathbf{A}) \quad (3)$$

where  $\mathbf{h}$ , in this case, is a canopy reflectance model.

### 2.5 Canopy reflectance model

The hybrid Geometric Optic Radiative Transfer (GORT) model of Ni et al. (1999) was coupled with DALEC to provide estimates of the top-of-canopy reflectance for MODIS channels 1 and 2. The leaf area index (LAI) required to drive GORT is

provided by the DALEC model by assuming it to be a constant proportion of the foliar biomass. Leaf reflectance was determined by the PROSPECT model (Jacquemoud and Baret, 1990) and soil reflectance was modelled by the empirical spectral functions of Price (1990).

The “ancillary” parameters of the GORT model (i.e. those that are not provided by the ecosystem model) are taken from the Quaife et al. (2007) paper. These are derived by a look up table inversion of the GORT model for the field site in question.

## 2.6 Snow reflectance model

To enable the use of snow contaminated MODIS data in the EnKF scheme; a model of the spectral albedo of snow was used to modify the lower boundary condition of GORT. The refractive index (Kou et al., 1993) and Mie scattering coefficients (Wiscombe, 1980) of ice particles were calculated using routines available from the NASA GSFC FTP site.

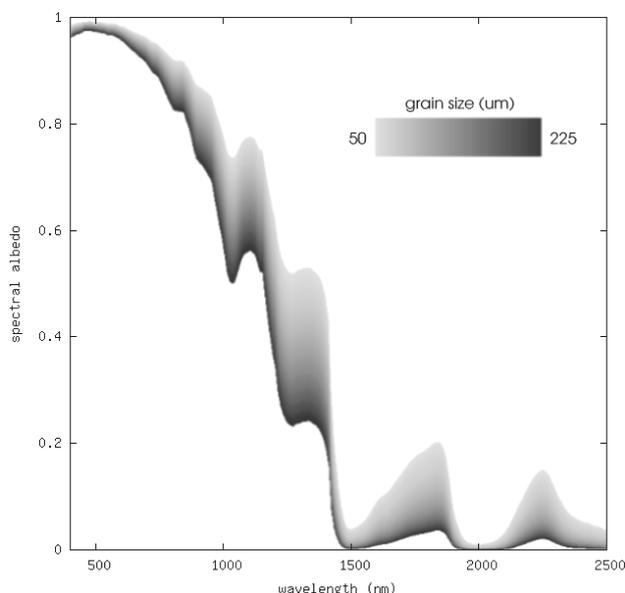


Figure 2. Snow spectral albedo as a function of wavelength and grain size (indicated by colour scale).

Integrated fluxes, taking into account multiple scattering were calculated using the DISORT code (Stamnes et al., 1988) available from the same site. This approach has been shown to provide reasonable predictions of the spectral albedo of snow by Green et al. (2002). For the purpose of extending the GORT model, the snow was assumed to be optically deep and only affect the lower boundary – no snow was taken to be resting in the tree crowns. Fig. 2 shows the spectral albedo for a range of wavelengths and grain sizes predicted by this model. Given the relative lack of sensitivity of albedo to grain size in the spectral regions of MODIS bands 1 and 2, an arbitrary grain size of 100µm was selected.

## 3. RESULTS

The DALEC model was run for the field site using meteorological drivers from the years 2000, 2001 and 2002. MODIS band 1 and 2 surface reflectance data were assimilated into the model using the EnKF. When the MODIS QA flags indicated snow in the retrieval the lower boundary of the GORT

model was covered with 50% snow, otherwise the GORT model was run without snow. Integrated values of GPP and NEP for the 3 years of the model run are given in Table 1 for the case where there is no assimilation, assimilation of reflectance *excluding* snow (previously reported in Quaife et al. 2007), assimilation of reflectance data *including* snow and the results of Williams et al. (2005) where large amounts of field data were assimilated.

Flux	Assimilated data	Total carbon uptake (g/m <sup>2</sup> )	Standard deviation
NEP	No data assimilation	240.2	212.2
	MODIS excluding snow	373.0	151.3
	MODIS including snow	404.8	129.6
	Williams et al. (2005)	406.0	27.8
GPP	No data assimilation	1646.4	834.5
	MODIS excluding snow	2620.3	96.8
	MODIS including snow	2525.6	42.7
	Williams et al. (2005)	2170.3	18.1

Table 1. Integrated C fluxes for 3 years.

### 3.1 Foliar Biomass

Fig. 3 shows the results of assimilating MODIS surface reflectance data into the DALEC model. Without the incorporation of snow contaminated pixels into the assimilation there is a clear period where the model is allowed to run without constraint and the ensemble spread increases monotonically. When the snow samples are added to the observations the model is adjusted accordingly, pulling the foliar biomass down during the winter months. The induced seasonality is quite large; in the order of 120 grams of carbon per m<sup>2</sup> over the course of a year. *Pinus Ponderosa* is evergreen and so seasonal cycles of this magnitude are unlikely. Some of the understory components are deciduous which will account, in part, for this effect but it is unlikely to explain all the variation.

### 3.2 Gross Primary Productivity

Without data in the winter months DALEC overestimates the GPP. The high, unconstrained, foliar biomass leads to an increase in light interception and thus greater uptake of atmospheric CO<sub>2</sub> by photosynthesis. This is rectified by the inclusion of the snow contaminated reflectance data, which pulls the foliar biomass down (Fig. 4). The integrated flux data (Table 1) only shows a slight improvement in the GPP however. This is because meteorological conditions (less incident PAR and lower temperatures) mean that GPP in the winter is low and so the correction at this time only has limited impact on the integrated quantity.

### 3.3 Net Ecosystem Productivity

The correction made to GPP is only small as a proportion of its integral. This amount of carbon as a proportion of NEP is much greater however, and so the correction to the results for NEP induced by inclusion of the winter reflectance values is significant. Fig. 5. shows that by assimilating the snow reflectance data the model predicts a stronger *source* of carbon during the winter months, which is in agreement with the observations taken at the field site. The values in Table 1. show a remarkable agreement with the values published by Williams et al (2005). Given that the GPP is overestimated this means that the total respiration from the system (autotrophic and heterotrophic) must also be overestimated to balance this out.

### 3.4 Parameter estimation

To use the techniques described above over wider areas it is important to be able to re-parameterise the DALEC model appropriately for the ecosystem in question. One option for determining the DALEC parameters is to use the EnKF itself. In this approach the parameters that are to be adjusted are placed in the model state vector and allowed to evolve with time under the influence of available observations. Fig. 7 show results of such an experiment assimilating MODIS reflectance data to adjust the rate of the carbon allocation to foliar biomass parameter.

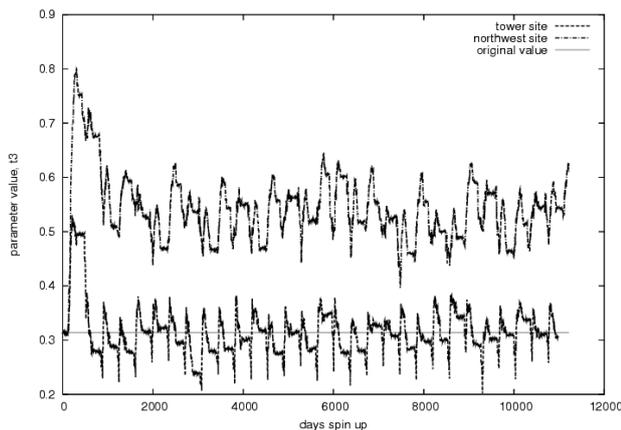


Figure 7. Parameter values derived using the EnKF and MODIS reflectance data. The “tower” site refers to the field site used elsewhere in this paper.

In these experiments a single year of meteorological data and MODIS observations were fed repeatedly into the model. When the reflectance data for the Metolius site was used the value of the parameter settles so that its mean is almost equal to that determined by Williams et al. (2005) using a quasi Newton minimisation method. When the reflectance data for a nearby site to the northwest is used the parameter takes on a higher mean value. This site has been more recently felled than the Metolius site and the younger trees are likely to be allocating more carbon to leaf material.

In these results the allocation parameter has been heavily damped by giving it a very small uncertainty and the same year of data was assimilated many times in repetition. Without damping, the parameter did not to converge toward a constant mean. This suggests that it is not possible to use such techniques at the same time as a normal assimilation run, but

instead they need to be performed off-line, much like a traditional calibration.

## 4. DISCUSSION AND CONCLUSION

This paper highlights some of the issues of using simple models in an EnKF scheme when there are long gaps between observation times. The DALEC model was designed to be used with a large number of observations, i.e. it gives reasonable, unbiased forecasts over some days or weeks, but not over several months. This is addressed here by building a modified version of an observation operator to permit assimilation of observations that had previously been disregarded (i.e. those contaminated by snow). Estimates of gross and net carbon fluxes were improved by the inclusion of these data.

Inclusion of winter observations induced a much higher variability in foliar biomass than expected. This is quite likely due to the assumption of a static 50% snow coverage for winter observations. A mechanism for determining this proportion is required to improve the observation operator for such cases. Ideally a snow hydrology model would be incorporated within the scheme and the EnKF used to adjust its parameters.

In addition, the problem of selecting an appropriate snow grain size has been ignored in these experiments. The variability of the spectral albedo as a function of grain size is small at the wavelengths being considered and so this is justified. It will certainly be a minor effect in comparison to the assumptions of a constant proportion of ground covered by snow. To bring in observations from other regions of the optical spectrum some mechanism will be required to take grain size into account.

With comparison to the field data the modelled foliar biomass is high. The GPP, in turn, is overestimated because of this. The foliar biomass field measurements are acquired using an LAI2000 probe and scaled to biomass by taking the product of the LAI and specific leaf area. The plot size for these measurements is 100×100m and so the discrepancy may be attributable to the difference in scale with the 500×500m MODIS pixels. The reference GPP data are generated from these LAI data also and so may be subject to similar scale problems. Work is currently underway to address this issue using high resolution ASTER data over the same region.

Initial results of parameter estimation exercises show the potential of EO data to provide partial model calibration for wide areas: the method appears to retrieve reasonable values for the test cases. The parameters seem unable to settle to a single value however. This may reflect inadequacies in the simple model parameterisation but will almost certainly be partly due to noise in the reflectance data. In the results shown heavy damping has been introduced to allow the parameter to settle. An implication of this is parameter estimation may need to be carried out prior to the main run of the model-EnKF scheme.

An issue that is not explored in this paper is the determination of the ancillary parameters for the observation operator. The numerical procedures currently being used are probably too slow to be used for large amounts of data. The ideal situation would be that the ecosystem model itself described as many of them as possible, but failing that an approach similar to the parameter estimation may prove useful.

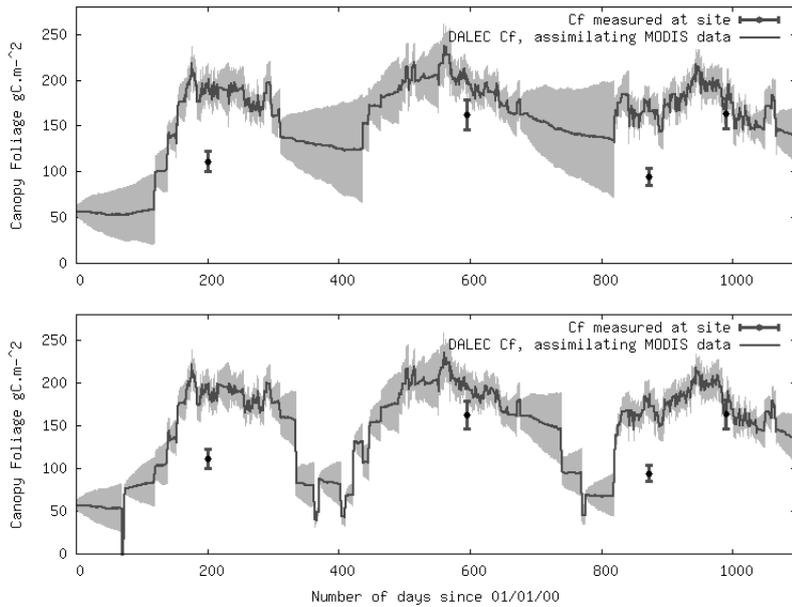


Figure 3. Foliar biomass predicted with DALEC assimilating observations of MODIS reflectance data. The top panel shows the case where snow contaminated pixels are exclude and the bottom panel show the results where snow contaminated data is used in the assimilation. The mean of the ensemble is shown with a black line and the spread about the line indicates one standard deviation of the ensemble. Field observations of canopy foliage (LAI scaled by specific leaf area) are plotted as single points with error bars.

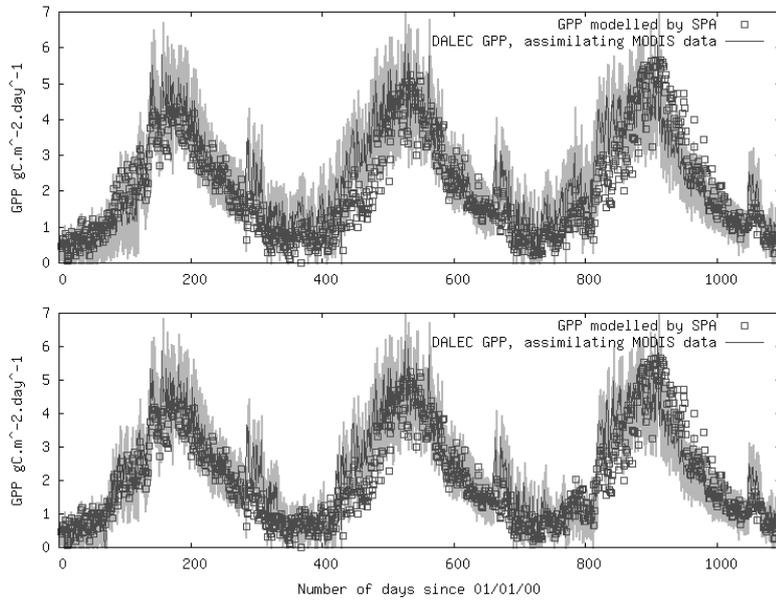


Figure 4. Gross primary productivity predicted with DALEC assimilating observations of MODIS reflectance data. The top panel shows the case where snow contaminated pixels are exclude and the bottom panel show the results where snow contaminated data is used in the assimilation. The mean of the ensemble is shown with a black line and the spread about the line indicates one standard deviation of the ensemble. Estimates of GPP modelled using the SPA model of Williams et al. (1996) are shown as small black squares.

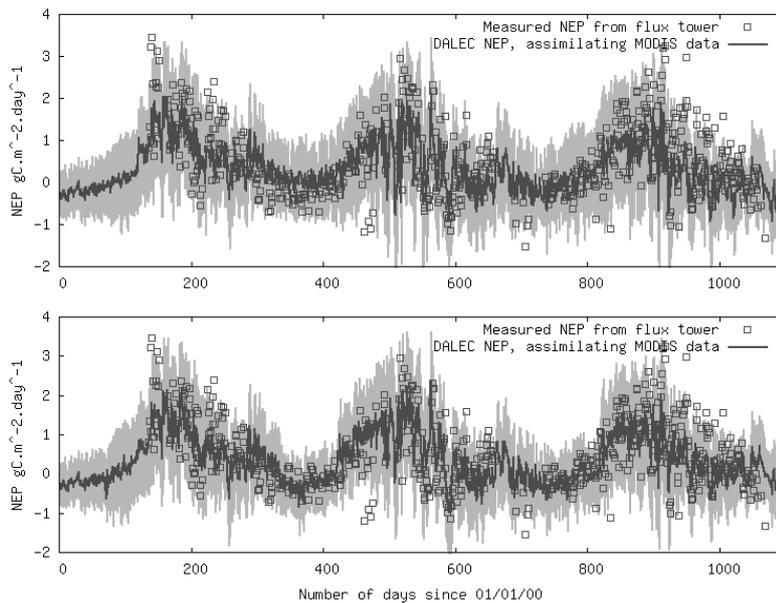


Figure 5. Net ecosystem productivity predicted with DALEC assimilating observations of MODIS reflectance data. The top panel shows the case where snow contaminated pixels are exclude and the bottom panel show the results where snow contaminated data is used in the assimilation. The mean of the ensemble is shown with a black line and the spread about the line indicates one standard deviation of the ensemble. Flux tower observations of NEP from the field site are shown as small black squares.

## REFERENCES

- Anthoni, P.M., Unsworth, M.H., Law, B.E., Irvine, J., Baldocchi, D. & Moore, D., 2002. Seasonal differences in carbon and water vapor exchange in young and old-growth ponderosa pine ecosystems. *Agricultural and Forest Meteorology*, 111, 203-222.
- Churkina, G., Schimel, D., Braswell, B.H., & Xiao, X., 2005. Spatial analysis of growing season length control over net ecosystem exchange. *Global Change Biology*, 11, 1777-1787
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143-10162.
- Evensen, G., 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, *Ocean Dynamics*, 53, 343-367.
- Green, R.O., Dozier, J., Roberts, D. and Painter, T., 2002. Spectral snow-reflectance models for grain-size and liquid-water fraction in melting snow for the solar-reflected spectrum. *Annals of Glaciology*, 34, 71-73.
- IPCC, 2001, *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A. (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York.
- Irvine, J. & Law, B.E. 2002. Contrasting soil respiration in young and old-growth ponderosa pine forests. *Global Change Biology*, 8, 1183-1194.
- Jacquemoud, S., and Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. *Remote Sensing of Environment*, 34, 75-91.
- Kou, L., Labrie, D. and Chylek, P., 1993. Refractive indices of water and ice in the 0.65- to 2.5- $\mu\text{m}$  spectral range. *Applied Optics*, 32 (19), 3531-3540.
- Law, B. E., Thornton, P., Irvine, J., Van Tuyl, S. & Anthoni, P., 2001a. Carbon storage and fluxes in ponderosa pine forests at different developmental stages. *Global Change Biology*, 7, 755-777.
- Law, B.E., Van Tuyl, S., Cescatti, A. & Baldocchi, D.D., 2001b. Estimation of leaf area index in open-canopy ponderosa pine forests at different successional stages and management regimes in Oregon. *Agricultural and Forest Meteorology*, 108, 1-14.
- Ni W.G., Li X.W., Woodcock C.E., Caetano M.R. and Strahler A.H., 1999. An analytical hybrid GORT model for bidirectional reflectance over discontinuous plant canopies. *IEEE Transactions on Geoscience and Remote Sensing*. 37(2), 987-999.
- Price, J., 1990. On the Information Content of Soil Reflectance Spectra, *Remote Sensing of Environment*. 33, 113-121.
- Quaife T., Lewis, P., De Kauwe, M., Williams, M., Law, B.E., Disney, M. and Bowyer, P., 2007, in press. Assimilating Canopy Reflectance data into an Ecosystem Model with an Ensemble Kalman Filter. *Remote Sensing of Environment*. In press.
- Rastetter, E.B., 2003. The collision of hypotheses: What can be learned from comparisons of ecosystem models? In Canham, C.D., Cole, J.J. & Lauenroth, W.K. (Eds.) *Models in Ecosystem Science*, (pp. 211-224). Princeton University Press: Princeton (NJ).
- Roy, D.P., Jin, Y., Lewis, P. & Justice, C.O., 2005. Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment*, 97, 137-162.
- Schimel, D.S. et al., 2001. Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature*, 414, 169-172.
- Stamnes, K., Tsay S-C., Wiscombe, W. and Jayaweera, K., 1988. Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics*. 27 (12), 2502-2509.
- Valentini, R. et al., 2000. Respiration as the main determinant of carbon balance in European forests, *Nature*, 404, 861-865.
- Williams, M., Rastetter, E.B., Fernandes, D.N., Goulden, M.L., Shaver, G.R., & Johnson, L.C., 1997. Predicting gross primary productivity in terrestrial ecosystems. *Ecological Applications*, 7, 882-894.
- Williams, M, Schwarz, P.A., Law, B.E., Irvine, J. & Kurpius, M., 2005. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11, 89-105.
- Williams, M., Rastetter, E.B., Fernandes, D.N., Goulden, M.L., Wofsy, S.C., Shaver, G.R., Melillo, J.M., Munger, J.W., Fan, S.-M., & Nadelhoffer, K.J., 1996. Modelling the soil-plant-atmosphere continuum in a Quercus-Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties. *Plant, Cell and Environment*, 19, 911-927.
- Wiscombe, W., 1980. Improved Mie scattering algorithms. *Applied Optics*, 19 (9), 1505-1509.