ASSIMILATING EARTH OBSERVATION DATA INTO LAND SURFACE MODELS

T. Quaife, P. Lewis and M. De Kauwe.

National Centre for Earth Observation & Department of Geography,
University College London, London, UK.

ABSTRACT

Data assimilation techniques such as the ensemble Kalman filter and the sequential Metropolis-Hastings algorithm provide a means of integrating satellite data with ecosystem models to optimally adjust their temporal trajectory. To some extent these methods can compensate for poor model parameterisations but a preferable scenario is to calibrate the model well in the first instance. This paper explores how a site specific model calibration can be adapted to a different site using only MODIS reflectance data. Results show that, using reflectance data only, estimates of the net carbon budget of a field site can be extended to a nearby site, but that this best facilitated by re–calibration rather than sequential data assimilation.

Index Terms— Data assimilation, NEP, Bayesian, GORT.

1. INTRODUCTION

Models of the terrestrial carbon (C) cycle are often poorly constrained by observation. Eddy covariance flux towers are one of the most significant data sources of Net Ecosystem Productivity (NEP), the difference between the amount of carbon sequestered by photosynthesis (GPP) and that released to the atmosphere via autotrophic and heterotrophic respiration. However, such instruments are costly to run in terms of human resource and can only provide very sparse spatial sampling of the terrestrial biosphere.

Consequently, spatially synoptic data from medium resolution satellite sensors such as MODIS provide an ideal data source to compliment flux tower measurements. Whilst these data are not directly related to net C flux they may be used to constrain ecosystem C models via Data Assimilation (DA) techniques. Such techniques seek to provide statistically optimal estimates of the ecosystem model state vectors based on all available observations and our best understanding of uncertainties in both model and data.

Derived 'high–level' products from satellite data, such as Gross Primary Productivity (GPP) or Leaf Area Index (LAI) are transformations of satellite data that are linearly related to an ecosystem model state vector and are thus straight forward to assimilate. It is often difficult to account for uncertainties in such products however and so the approach taken in this paper is to assimilate a 'low–level' MODIS product, surface reflectance. This is achieved by coupling an observation operator to the ecosystem model to make predictions of the satellite observed reflectance.

This paper examines the potential to spatialise the assimilation procedure from a well known, intensively studied field site to a nearby site where only information available is *i*) the calibration from the previous site and *ii*) MODIS surface reflectance data. Results are show in terms of ability to improve estimates of the NEP as measured by an eddy covariance flux tower.

2. ECOSYSTEM MODEL

The model used in this study is the Data Assimilation Linked ECosystem model (DALEC) [1]. DALEC models the flow of carbon through an ecosystem as a series of pools, that represent stocks of carbon (*e.g.,* foliar biomass). Carbon moves between these pools at rates defined by a set of internal parameters that require calibrating against data. The model was originally designed to use an ensemble Kalman filter (EnKF) [2] to assimilate field measurements to adjust the model state, provided an initial calibration. This approach was shown to improve estimates of the Net Ecosystem Production (NEP) and other stocks and fluxes of carbon when compared to the same model calibration running without data assimilation [1]. DALEC has been used with an EnKF to assimilate surface reflectance data from MODIS [3] which was also shown to improve estimates of the NEP. To achieve the assimilation of reflectance data an observation operator (in this case the canopy reflectance model, GORT [4]) was coupled to DALEC to enable it to predict surface reflectance.

Both of these previous studies use DALEC at a site where it had been calibrated against large quantities of field data. This site was a young age class ponderosa pine plantation and is referred to here as the 'young' site. This paper investigates a nearby, intermediate age class site of ponderosa pine, approximately 2km away.

3. DATA

Observations of net CO_2 flux (NEP) and meteorological data required to drive the DALEC model were acquired over 2002 for the intermediate site. Only observations of NEP that had not been gap filled are used in the results presented below.

MOD09 500m surface reflectance data, bands 1 and 2 (collection 4), were used as assimilates. Only observations passing all quality assurance tests were used in the assimilation procedure. As a further quality check, data points that exhibited a large residual when a Ross Thick–Li Sparse model was fit to the time series were discarded.

4. SEQUENTIAL MH ALGORITHM

In this paper a different technique from the EnKF is adopted for the state estimation problem: a variant of the Metropolis–Hastings algorithm designed for sequential data assimilation (seqMH) [5]. The seqMH algorithm is a particle filter and provides a fully Bayesian solution to the state estimation problem. The EnKF assumes all uncertainties to be Gaussian. This is unlikely to be the case when surface reflectance values are being predicted from an ensemble of LAI values. Consequently the ability to process none Gaussian distributions is desirable when using reflectance data.

A brief outline of the method is as follows. An ensemble of model state vectors x^n (where n is the number of ensemble members) are propagated forward in time by a process model using stochastic forcing to represent uncertainty in the model (*e.g.,* process representation, parameterisation and so on) and driving data. Given observations of surface reflectance, ρ at time t, n candidate replacement particles are selected at random from the ensemble and propagated through DALEC. The acceptance probability, α, of each candidate particle $x^*_{t|t-1}$ compared to the previously accepted particle $x^{i-1}_{t|t}$ is given by:

$$\alpha = \min\left(1, \frac{p(\rho_t|h(x^*_{t|t-1}))}{p(\rho_t|h(x^{i-1}_{t|t}))}\right) \tag{1}$$

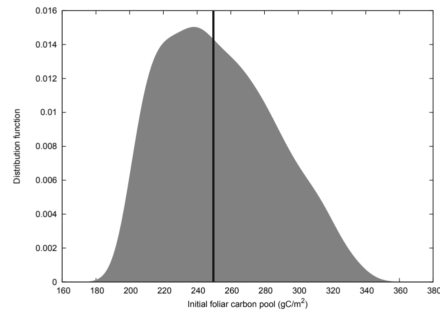
where h is the observation operator. Which of the two particles is accepted as the new i^{th} ensemble member is determined by comparing α with a draw z from the uniform random distribution, where $0 \leq z \leq 1$.

$$x^i_{t|t} = \begin{cases} x^*_{t|t-1} & \text{if } z \leq \alpha \\ x^{i-1}_{t|t} & \text{if } z > \alpha \end{cases} \tag{2}$$

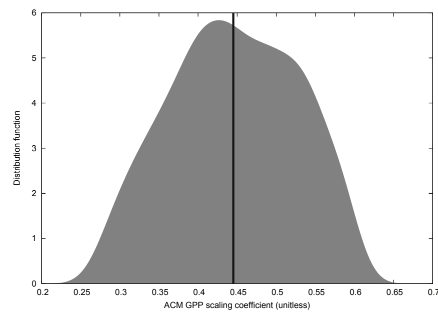
Note that the algorithm requires an initial particle to be selected for $x^1_{t|t}$, which in this case is set to the ensemble mean. Given a linear model with Gaussian errors this technique converges to the Kalman Filter as the number of particles used becomes large. For a complete description see [5].

5. PARAMETERISATION

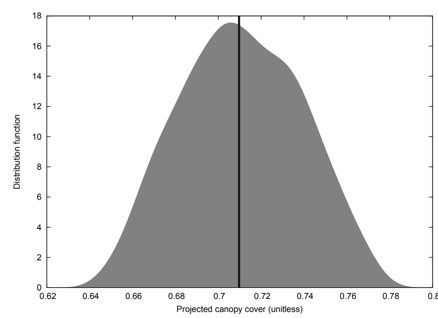
Previous studies with DALEC have run the model at the young site where it has been calibrated against large amounts



(a) Initial foliar carbon



(b) GPP scaling factor.



(c) Projected canopy cover

Fig. 1. Distribution of parameter values for the lowest 2% of RMSE in the brute force sampling.

of data [1], [3]. Clearly it is not feasible to collect large quantities of data at every point on the Earth's surface, and so it is desirable to know how effectively the model predictions of NEP and other C fluxes can be influenced by satellite data alone. Although it is possible to use a state–based assimilation to impose some degree of spatialisation [3] ideally the DALEC parameters would also be adjusted to remove any gross errors in the model. It is not possible, however, to optimise all of the DALEC model parameters against reflectance data alone. For example, reflectance is not sensitive to rates of leaf litter decomposition. Consequently a subset of parameters is chosen for optimisation and the remainder are taken from the results of the young site calibration. This is reasonable because the sites contain the same species and have very similar climatic conditions. The parameters chosen

	ACM5	PCC	C_{f0}
Young site	0.16	0.63	57.7
Re–calibration	0.45	0.71	249.7

Table 1. Re–calibrated, and original young site parameter values. ACM5 is a GPP scaling term, PCC is the GORT projected crown cover parameter, and C_{f0} is the initial foliar carbon pool (gC/m^2). Young site parameters are taken from previous publications [1], [3].

for optimisation were: the intial value of the foliar carbon pool (C_{f0}) a photosynthesis (or GPP) scaling factor (referred to as ACM5 in [1]) and the projected crown cover (used in GORT).

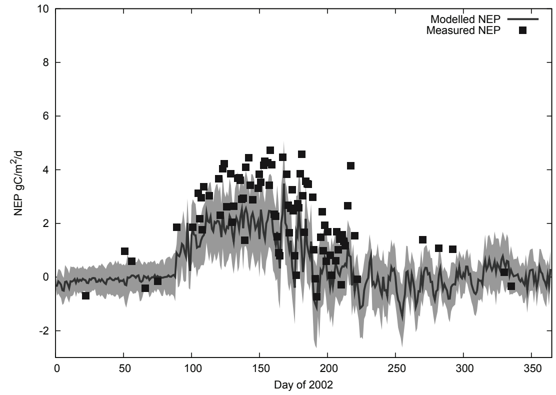
The optimisation was performed using a look–up table of 100000 combinations of the three selected parameters. The final selected set were the median parameter values from the lowest 2% of RMSEs compared to the MODIS surface reflectance data. The results and distributions of values are given in Fig. 1 and Table 1. The change in parameters from the initial site with young pine tree, to the re–calibrated values at intermediate site is reasonably intuitive: projected crown cover increases (*i.e.,* larger tree crowns) and the initial foliar carbon pool is nearly 5 times as large (*i.e.,* more leaves). The change in the GPP scaling factor suggests that the intermediate site is also more efficient at fixing carbon.

6. RESULTS

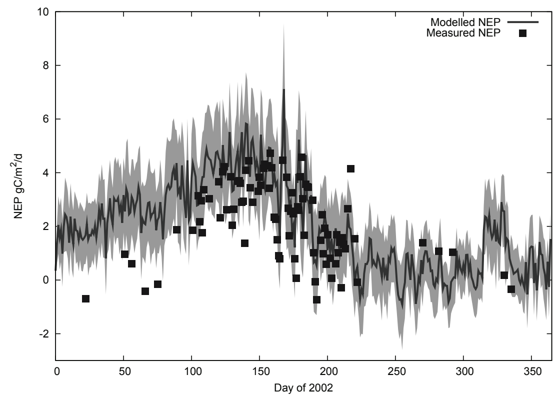
Temporal trajectories of modelled NEP are show in Fig. 2, using (a) the parameters from the young site only and (b) the parameters re–calibrated against MODIS reflectance data for the intermediate site. Both sets of results have been processed by the seqMH algorithm, assimilating reflectance data.

The re–calibrated model suggests a much stronger sink of carbon than the results that use the young site calibration. Values at the beginning of the year appear to be too high, although this could be attributable to the low number of flux tower observations during this period not properly representing the true signal. The majority of NEP observations fall within $\pm 1\sigma$ for the re–calibrated model, whereas around half fall outside using the young site parameters. This clearly indicates an improved representation of the carbon dynamics for this site.

Fig. 3 shows the NEP predictions made by DALEC running without seqMH data assimilation. For the re–calibrated model the results are not greatly different with or without assimilation. Errors are slightly inflated without performing data assimilation, but the mean trajectory is essentially the same ($r^2 = 0.97$). Clear difference exist between the assimilation and no–assimilation cases for the young site parameters however. The first available reflectance data are on day 88



(a) NEP, original model calibration.



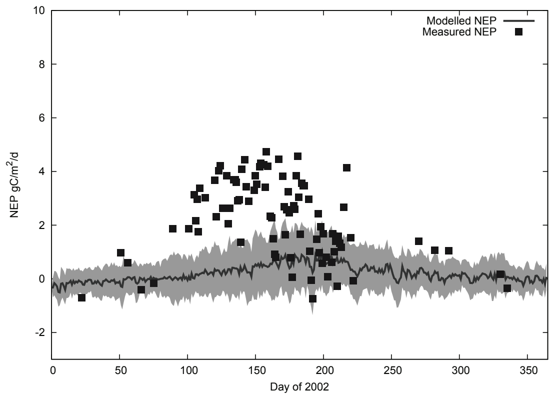
(b) NEP, re–calibrated model.

Fig. 2. NEP modelled by DALEC, and using the seqMH algorithm, for the intermediate age class flux tower site. Results in (a) use the existing model calibrations, whereas (b) uses a re–calibrated version of DALEC and the observation operator. The grey spread about the modelled NEP line shows the modelled uncertainty at $\pm 1\sigma$.

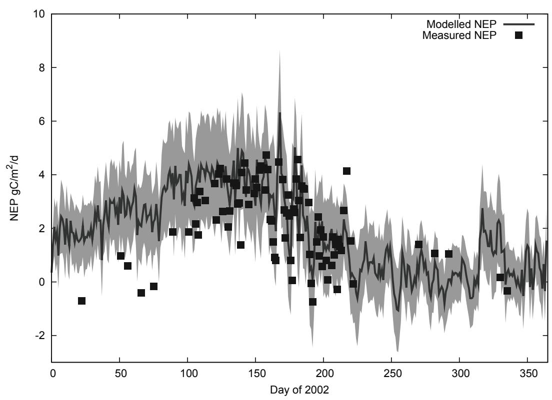
and the difference between Fig. 2(a) and 3(a) clearly shows the impact of seqMH algorithm. As data become available it starts to compensate for the poor model calibration. Without any data assimilation the young site calibration captures none of the variation of the observed NEP. The improvement of the NEP prediction in Fig. 3 between the different calibrations is marked.

7. DISCUSSION

The DALEC and GORT models can be re–calibrated using MODIS reflectance data to provide improved estimates of a net carbon flux (NEP) because the reflectance signal is controlled by LAI which, in the DALEC model, is directly proportional to the amount of foliar carbon. LAI, in turn, controls GPP (which is the C gain term of the NEP equation). Consequently the reflectance is sensitive to the GPP scaling factor



(a) NEP, original model calibration.



(b) NEP, re–calibrated model.

Fig. 3. As Fig. 2 but without applying the seqMH algorithm.

The C loss term in NEP is respiration from plants and soil. Given that parameters controlling these were not optimsed, and the NEP results are good, this suggests that the young site parameter values for the respiration components of the model are roughly appropriate for the intermediate site. Furthermore, the differences in the calibrated parameter values appears intuitive, given that they are changing from a young pine stand to an intermediate pine stand.

When the young site calibration is used the seqMH algorithm improves the DALEC predicted NEP. However, given the re–calibration against MODIS data for the intermediate site the sequential data assimilation has very little impact on the NEP trajectory. Uncertainty in the model prediction is reduced, but overall the impact is marginal. This is because gross errors in the model predictions have been eliminated by the optimisation. Clearly, when moving to sites differ more than the two considered here (*e.g.,* further away or different species) then are more of the model parameters may need re-calibrating, potentially requiring different types of data.

8. CONCLUSION

This paper demonstrates the feasibility of using an adapted MH algorithm for performing sequential data assimilation of surface reflectance data into an ecosystem model. It is shown to improve the performance of a poorly calibrated model, comparing against flux tower measurements of NEP. A strategy to spatialise the assimilation procedure by optimising a selected subset of process model and observation operator parameters is also described. Results show that calibration of this parameter subset against MODIS reflectance data is sufficient to improve model representation of carbon dynamics at a nearby site. The sites used contain the same species, albeit different age classes, and are close together. Further work is required to examine the domain over which this method is be valid. A key point raised is that state estimation techniques such as the EnKF and seqMH can compensate for poor model calibration, but whilst this maybe useful in some instances, a clearly preferable solution is to have good model calibration in the first instance.

9. ACKNOWLEDGEMENTS

We are grateful to Bev Law of Oregon State University for eddy covariance data used in this project and to Mat Williams of Edinburgh University for provision of the DALEC model. This work was funded by the UK National Centre for Earth Observation.

10. REFERENCES

[1] M. Williams, P.A. Schwarz, B.E. Law, J. Irvine, and M.R. Kurpius, “An improved analysis of forest carbon dynamics using data assimilation,” *Global Change Biology*, vol. 11, pp. 89–105, 2005.

[2] G Evensen, “The ensemble kalman filter: Theoretical formulation and practical implementation,” *Ocean Dynamics*, vol. 53, pp. 343–367, 2003.

[3] 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*, vol. 112, pp. 1347–1364, 2008.

[4] W. Ni, X. Li, C. E. Woodcock, M. R. Caetano, and A. H. Strahler, “An analytical hybrid gort model for bidirectional reflectance overdiscontinuous plant canopies,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, pp. 987–999, 1999.

[5] M. Dowd, “Bayesian statistical data assimilation for ecosystem models using markov chain monte carlo,” *Journal of Marine Systems*, vol. 68, pp. 439–456, 2007.