REFLEX: REgional FLux Estimation eXperiment

A CarbonFusion project (www.carbonfusion.org)

Table of Contents:

Summary

REFLEX is a model-data fusion inter-comparison project aimed at assessing the capability of extrapolating observations of CO_2 fluxes from intensively studied sites (e.g. flux towers) to other locations (i.e. the surrounding region). This project will address many of the issues faced when predicting carbon fluxes at local-to-regional scales, and will compare and contrast a range of carbon-model parameter estimation/flux forecasting techniques. The outcome of this project will be an improved capability for prediction of carbon fluxes at regional to continental scales, and a quantification of the errors associated with extrapolation.

Aims

1. To compare the strengths and weaknesses of various model-data fusion techniques for estimating carbon model parameters and predicting carbon fluxes.
2. To quantify errors and biases introduced when extrapolating fluxes (and related measurements) made at flux tower sites in both space and time, using earth observation data and models constrained by model-data fusion methods.

Rationale

The flux tower networks around the world (FLUXNET) are important facilities enabling the measurement of limited-area carbon fluxes and the calibration of photosynthesis and respiration models at flux tower sites. These "point" estimates are combined with measurements and calibrated models from other network flux towers to scale fluxes of CO_2 upwards to regions. However, regional estimations based on these calibrated models contain errors due to:

- biases and noise associated with the representativeness of tower locations in the landscape;
- micrometeorological processes such as advection and preferential nocturnal drainage;
- inadequacy of model driver data;
- incompleteness of processing algorithms; and
- structural errors in the models.

This project will test the capability of carbon models to generate bias-free predictions of carbon fluxes using state-of-the-art model-data fusion techniques, while at the same time contributing to the development and refinement of those techniques for their routine application. **A particular focus will be on how earth observation (EO) data can be best used to assist the extrapolation process**, given that EO data provide the possibility for global and repeated measurements of land surface characteristics.

The value of this experiment is two-fold:

1. It will facilitate learning and technology transfer of model-data fusion techniques, by demonstrating an analysis of forest carbon dynamics using a simple carbon model. Comparison of various numerical techniques will assess both strength of analysis and computational efficiency.
2. It will quantify the capability of model-data fusion schemes (models + driver data + optimization) for within-site forecasting and between-site extrapolation skill.

Who can participate?

REFLEX is open to anyone who would like to participate. Those wishing to be involved can register their intention to participate, along with their chosen MDF techniques, to Andy Fox (a.m.fox@shef.ac.uk) by 15 August 2007.

Methodology

- The experiments are undertaken with data from European flux sites. These sites include two deciduous broad-leaf forests (DE1 and DE3) and two evergreen needle-leaf forests (EV1 and EV3). Sites have been selected which have relatively long, continuous records of fluxes and site meteorology which have been quality controlled. 'Gap filled' flux data itself are not required in this experiment and have been excluded. Two other "synthetic" sets of site data are provided, DE2 and EV2, for some experiments.
- Data provision for each site has been classified into one of two categories, high- and low-level (Tables 1 and 2). High-level data represent typical data available at a flux tower site (i.e. fluxes, albeit with gaps, and some ecological data). Low-level data are those that can be derived for any European location from remote sensing instruments (MODIS in this case), soil/forest databases and from interpolated weather station data, i.e. they lack fluxes. All data have been obtained via the Fluxnet site (http://www.fluxnet.ornl.gov/fluxnet/index.cfm), from

relevant, site specific literature and/or from site PIs. An attempt has been made to estimate the error in observations.

- We supply a simple reference carbon dynamics model for use by all participants. The use of a common reference model is required to assess variation in analyses resulting from differences in MDF techniques. The DALEC model (Williams et al., 2005) simulates C dynamics for evergreen stands, and we also provide DALEC-D, a version modified to incorporate phenology for deciduous trees. Model description, equations and FORTRAN code are described below.
- Participants may supply and use their own carbon model at their discretion, but their MDF technique must be applied to the common model to allow comparison.
- Participants apply the MDF technique of their choosing to a set of experiments.

Experiments

Experiment 1: These can be considered as “training” runs for the model. High-level data for two years at two sites (1 evergreen, EV1; 1 deciduous, DE1) are provided. The simple carbon model is used to analyse the carbon dynamics of both sites, using evergreen and deciduous versions of the model as appropriate. Participants provide estimates of parameters and times series of fluxes and pools over this two year period. In effect this is a gap-filling exercise, combined with model calibration. A useful extra output is an assessment of parameter uncertainty (probability density functions) and confidence intervals on time series of stocks and fluxes.

Experiment 2: As experiment 1, but the two data sets (evergreen, EV2; deciduous, DE2) provided have been synthetically generated by sampling the output of the DALEC and DALEC-D models, with added random error. This experiment will help to evaluate the performance of different MDF techniques as the truth is known and it removes to need to compare model output with observations which have sampling and instrumental errors.

Experiment 3: Within-site forecasting will be assessed by providing an additional year of meteorological driving data for the training run sites used in experiment 1 (EV1 and DE1). This experiment is an explicit extrapolation forward in time from the end of the model runs in experiment 1, using parameters from that analysis. Participants will provide predictions of forecast times series of fluxes and pools over this one year period, plus confidence intervals.

Experiment 4: As experiment 3, but using synthetically generated data to enable forecasting at the sites used in experiment 2 (EV2 and DE2).

Experiment 5: Between-site extrapolation skill will be measured by providing three years of low level data (Table 2) for two additional sites “twinned” with sites used in experiment 1. Twinned sites have similar vegetation type, age structure and disturbance history, so EV3 is paired with EV1 and DE3 with DE1. The main difference between this experiment and the within-in site forecasting (experiment 3) is that MODIS estimates of LAI will be available throughout the experimental period and the initial conditions will be poorly known. Participants will have to infer parameters and initial conditions using information from the designated twinned site in experiment 1 and limited information in the site description table (Appendix table A1).

Observation	Units	Interval	Source
Global Radiation	MJ m^{-2} d^{-1}	Daily	Fluxnet data portal
Min Temperature	°C	Daily	Fluxnet data portal
Max temperature	°C	Daily	Fluxnet data portal
CO_2 conc.	μmol mol^{-1}	Daily	Fluxnet data portal
NEE	g C m^{-2} d^{-1}	Daily	Fluxnet data portal
LAI	m^2 m^{-2}	When available	References/site PI
LAI (EO)	m^2 m^{-2}	When available	Modis ASCII subsets
Foliar N *	gN m^{-2} leaf area	Constant	References/site PI
Vegetation type*	n/a	Constant	References/site PI
Aboveground C mass*	Kg C m^{-2}	Initial condition	References/site PI
SOM C mass*	Kg C m^{-2}	Initial condition	References/site PI
Leaf mass per area*	g C m^{-2} leaf area	Constant	References/site PI

Table 1. "High-level" time series data available for use in experiments 1-4. Data with a "constant" interval are fixed values throughout model runs. *Ancillary data contained in Appendix table A5 for all experiments.

Observation	Units	Interval	Source
Global Radiation	MJ m^{-2} d^{-1}	Daily	Fluxnet data portal
Min Temperature	°C	Daily	Fluxnet data portal
Max temperature	°C	Daily	Fluxnet data portal
LAI	m^2 m^{-2}	When available	Modis ASCII subsets

Table 2. "Low-level" time series data available for use in experiments 5. See also Table A5.

MODIS and ground-based LAI data

The data indicate that there can be a (large) discrepancy between MODIS and ground-based assessments of LAI (see DE1_obs and EV1_obs). We suggest that participants undertake experiments 1 and 3 with 3 different sets of assimilation data:

1. Assimilating NEE, MODIS LAI and ground-based LAI (i.e. all data).
2. Assimilating NEE and ground-based LAI.
3. Assimilating NEE and MODIS LAI.

We can then assess the impact of the LAI discrepancies on C flux analyses.
For experiment 5 only MODIS LAI data are provided, so participants must make an assessment of MODIS errors in this assimilation (see also Appendix section on observational uncertainties).

Evaluation and outputs

Evaluation

For each experiment the participants will provide:

1. Times series of the C fluxes and pools associated with the DALEC model (i.e. the state variables), the daily NEE and total integrated NEE, and an assessment of the uncertainty in these estimates. **Please provide mean estimates and 90% confidence intervals on the state variable time series, the daily NEE and total integrated NEE.** Describe how confidence intervals were generated.

2. Parameter estimates (including probability density functions if possible). Provide details of initial weightings if used. **Please provide a sample of acceptable parameter sets,** from which

statistics of the distributions and correlations/covariances can be calculated. [These outputs not required in expts 3 and 4, which use parameters from expts 1 and 2].
3. Please tell us the starting values for all C pools, and whether these were guessed or fitted (and how).
4. Information on the MDF techniques employed (details of cost functions used), and typical computation time/resources required for each experiment. Please also explain procedures used for experiment 5.

Outputs should be emailed to a.m.fox@shef.ac.uk by 31 October 2007.

See provided FORTRAN code for suggested formatting of outputs.

In the case of experiment 1 the strengths and weaknesses of the different MDF techniques will have to be assessed in relation to each other. In the case of experiments 3 and 5 the different techniques can again be compared with each other, but also our best estimate of the actual system. This estimate will most likely be obtained through additional MDF using all the available high-level data (i.e. data not available to the participants). For experiments 2 and 4 the true system is known and the performance of the different MDF techniques (and carbon models if supplied) can be quantified.

Outputs
The outputs from this project will be at least two science journal manuscripts, one associated with assessing the relative strengths and weaknesses of alternative MDF techniques, and the other with demonstrating MDF capabilities for extrapolating flux data regionally, and describing the associated errors.

1. Describe set up and results of experiments 1 and 2. This will involve an assessment of strengths and weaknesses of alternative MDFs in comparison with each other and the 'truth' in the synthetic datasets.
2. Describe set up and results of experiments 3-5. This will concentrate on the temporal and spatial extrapolation. In particular, these will most likely emphases the issues associated with initial conditions, and errors in MODIS.

We will arrange a meeting in the UK of all participants to contribute to these publications, and we will provide funding through CarbonFusion to support attendance.

Data use
Data from the CarboEurope Integrated Project (IP) used in this project are those made freely available to the public and the scientific community in the belief that their wide dissemination will lead to greater understanding and new scientific insights. Data owners and site PIs have already been made aware of the use of these data in this project and will be informed of developments in line with the fair-use policy.

Reflex science team

Andrew Fox (University of Sheffield & CTCD) – Reflex coordinator
Mathew Williams (Univ. of Edinburgh & CTCD) – Science Leader & Steering Committee Chair
Mike Raupach (CSIRO and GCP) – Steering Committee
Markus Reichstein (MPI-Jena) – Steering Committee

REFLEX

Damian Barret (CSIRO) – Steering Committee.

Model Description

We have attempted to strike the correct balance between sufficient model complexity to capture the essential dynamics of the system while maintaining simplicity. We will use the Data Assimilation Linked Ecosystem (DALEC) model (Williams *et al.*, 2005), designed for evergreen forests, and a modified version (DALEC-D) for deciduous forests. For simplicity there is no hydrological modelling and the sites and site-years were selected for use in this experiment specifically because they experienced little or no drought stress to allow this gross simplification.

DALEC a simple box model of carbon pools connected via fluxes (Figure 1) running at a daily time-step. For the evergreen model there are five C pools representing foliage (C_f), woody stems and coarse roots (C_w), and fine roots (C_r) along with fresh leaf and fine root litter (C_{lit}) and soil organic matter and coarse woody debris ($C_{som/cwd}$). In the deciduous model there is an additional labile pool (C_{lab}). There is a further pseudo-pool representing the daily accumulation of photosynthate (*GPP*) that is entirely utilised each day. The following assumptions are made to determine the fluxes between the C pools:

1. All C fixed during a day is expended either in autotrophic respiration or else allocated to one of the three plant tissue pools, C_f, C_w or C_r.
2. Autotrophic respiration is a constant fraction of the C fixed during a day and so is not directly temperature sensitive.
3. Allocation fractions are donor-controlled functions which have constant rate parameters.
4. In the case of the deciduous model the timing of leaf out is controlled by a simple growing degree day model, and leaf-fall by a minimum temperature threshold. The maximum amount of C that can be allocated to leaves is also limited by a parameter (c_{fmax})
5. All C losses are via mineralization (i.e. no dissolved losses etc).

The aggregated canopy model (ACM) (Williams *et al.*, 1997) is used to calculate daily GPP in DALEC. ACM is a 'big leaf', daily time-step model that estimates GPP using a simple aggregated set of equations operating on cumulative or average values of leaf area index (LAI), foliar nitrogen, total daily irradiance, minimum and maximum daily temperature, day length, atmospheric CO_2 concentration, water potential gradient (ψ_d) and total soil-plant hydraulic resistance (R_{tot}). ACM contains 10 parameters which have been calibrated using a fine-scale model (the Soil-Plant-Atmosphere model (SPA) (Williams *et al.*, 1996)) across a wide range of driving variables producing a 'universal' parameter set which maintains the essential behaviour of the fine-scale model but at a much reduced complexity. REFLEX participants are NOT required to optimize all the ACM parameters. The sole ACM parameter included in the optimisation is the nitrogen use efficiency parameter (a1), which determines the maximum rate of carboxylation per g foliar N. For the purposes of this experiment the sites will be treated as being non-drought stressed. Those variables related to drought effects in ACM, specifically ψ_d and R_{tot}, are given a fixed value in the FORTRAN code provided.

The basic model equations used by ACM-DALEC are described below. FORTRAN code containing these equations is also provided.

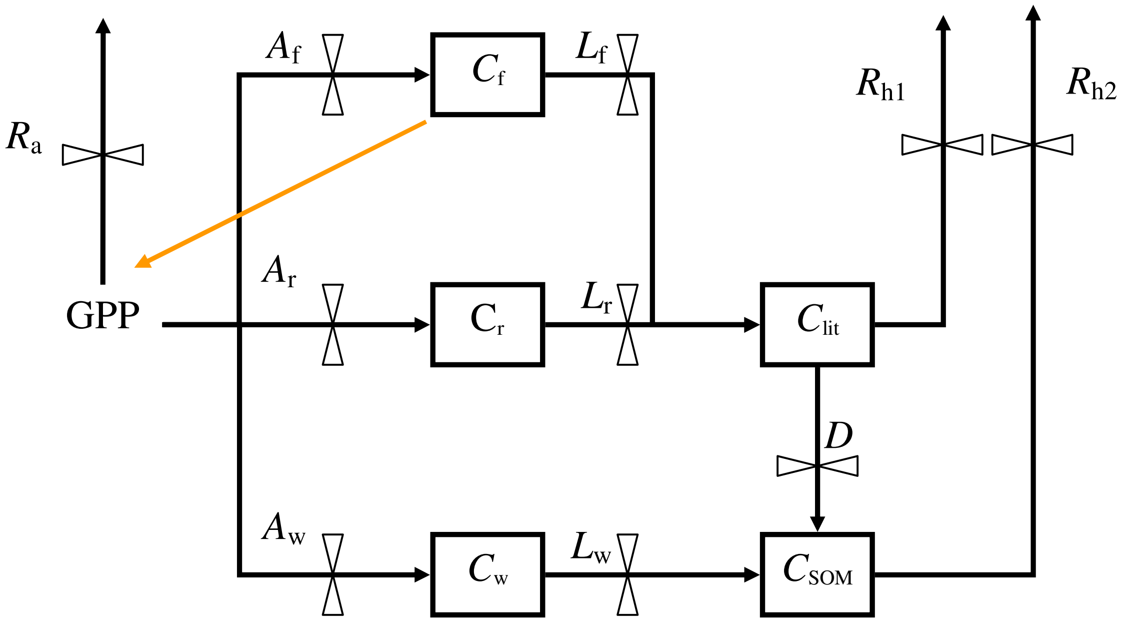

Figure 1a DALEC (evergreen).

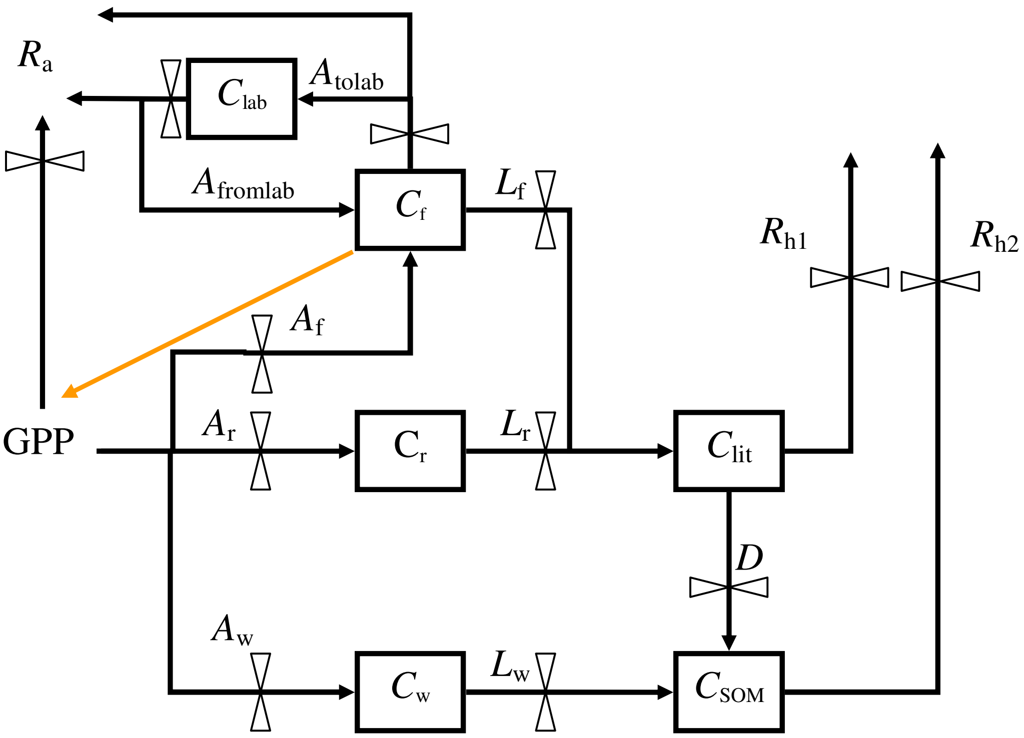

Figure 1b DALEC (deciduous). C dynamic model showing pools (boxes) and fluxes (arrows). Feed back between DALEC and ACM is indicated by orange arrow. Allocation fluxes are *A*, litterfall fluxes by *L*, and respiration by *R*, split between autotrophic (a) and heterotrophic (h). *D* is decomposition and GPP is gross primary productivity, a pseudo-pool simulated by ACM.

Model Equations

Temperature sensitive rate parameter:

$T = 0.5 \exp(p_{10}T_m)$ (Equivalent to a curve with $Q_{10} = 2$, $T = 1$ when $T_m = 10$ and $p_{10} = 0.0693$)
Where T_m = mean daily air temperature (°C)

Evergreen model equations

Require parameters p_1-p_{11}
Fluxes:

$G=f(C_f,p_{11})$ – f is the ACM model, described below

$R_a = p_2G$

$A_f = (G – R_a)\, p_3$

$A_r = (G – R_a – A_f)\, p_4$

$A_w = G – R_a – A_f – A_r$

$L_f = p_5C_f$

$L_w = p_6C_w$

$L_r = p_7C_r$

$R_{h1} = p_8C_{lit}T$

$R_{h2} = p_9C_{som}T$

$D = p_1C_{lit}T$

Pools:

$\Delta C_f = A_f – L_f$

$\Delta C_w = A_w – L_w$

$\Delta C_r = A_r – L_r$

$\Delta C_{lit} = L_f + L_r – R_{h1} – D$

$\Delta C_{som} = L_w + D – R_{h2}$

Deciduous model equations

Require parameters p_1-p_{17}
Fluxes:

$G=f(C_f,p_{11})$ – f is the ACM model, described below

$R_{alabt} = p_5\, (1 – p_{14})\, p_{16}\, C_f\, m_{tf}\, T$

$R_{alabf} = p_{15}\, p_{16}\, C_{lab}\, m_{tl}\, T$

$R_a = p_2G + R_{alabt} + R_{alabf}$

$NPP_1 = (1 - p2)\, G$

REFLEX

$A_\mathrm{f} = min(p_{17} - C_\mathrm{f},\ p_3\, NPP_1)\ m_\mathrm{tl} + A_\mathrm{fromlab}$

$NPP_2 = NPP_1 - min\ (p_{17} - C_\mathrm{f},\ p_3\, NPP1)\ m_\mathrm{tl}$

$A_\mathrm{r} = p_4\, NPP_2$

$A_\mathrm{w} = (1\text{-}p_4)\, NPP_2$

$A_\mathrm{tolab} = p_5\ (1 - p_{14})\ (1 - p_{16})\ C_\mathrm{f}\, m_\mathrm{tf}\, T$

$A_\mathrm{fromlab} = p_{15}\, (1 - p_{16})\ C_\mathrm{lab}\, m_\mathrm{tl}\, T$

$L_\mathrm{f} = p_5\, C_\mathrm{f}\ p_{14}\, m_\mathrm{tf}$

$L_\mathrm{w} = p_6 C_\mathrm{w}$

$L_\mathrm{r} = p_7 C_\mathrm{r}$

$R_\mathrm{h1} = p_8 C_\mathrm{lit}\, T$

$R_\mathrm{h2} = p_9 C_\mathrm{som}\, T$

$D = p_1 C_\mathrm{lit}\, T$

Pools:

$\Delta C_\mathrm{f} = A_\mathrm{f} - L_\mathrm{f} - A_\mathrm{tolab} - R_\mathrm{alabt}$

$\Delta C_\mathrm{w} = A_\mathrm{w} - L_\mathrm{w}$

$\Delta C_\mathrm{r} = A_\mathrm{r} - L_\mathrm{r}$

$\Delta C_\mathrm{lit} = L_\mathrm{f} + L_\mathrm{r} - R_\mathrm{h1} - D$

$\Delta C_\mathrm{som} = L_\mathrm{w} + D - R_\mathrm{h2}$

$\Delta \mathrm{C}_\mathrm{lab} = A_\mathrm{tolab} - A_\mathrm{fromlab} - R_\mathrm{alabf}$

Growing degree day factor:

Default $\quad GDD_\mathrm{doy} = 0$

If (DOY > 100)

$GDD_\mathrm{doy} = GDD_\mathrm{doy\text{-}1} + T_\mathrm{m}$

Phenology model

IF (yearday < 100) then gdd=0. max_fol=1.

If (GDD_doy < p12) !winter

mtf=1. !turnover of foliage on

mtl=0. !turnover of labile C off

ELSE

IF (max_fol = 1)THEN !spring

REFLEX

```
                        mtl =1.   !turnover of labile is ON
                        mtf =0.   !turnover of foliage is OFF
        ELSE                              !summer
                        mtl =0.   ! turnover of labile is OFF
                        mtf =0.   ! turnover of foliage is OFF
        ENDIF
ENDIF
IF (Cf > p17).or.(yearday > 200))THEN
        max_fol=0.        !switch to summer
        mtl =0.           !turnover of labile is OFF
ENDIF
If (yearday > 200).and.(mint < p13))         THEN    !autumn
                mtf =1.   !turnover of foliage is ON
ENDIF
```

ACM Model equations

(require one calibrated parameter, p_{11}, and 9 fixed parameters, $a_{2\text{-}10}$).

$$g_c = \frac{\left|\psi_d\right|^{a_{10}}}{0.5T_r + a_6 R_{tot}}$$

$$p = \frac{p_{11} NL}{g_c}\exp(T_{\max} a_8)$$

$$q = a_3 - a_4$$

$$C_i = \frac{1}{2}\left[C_a + q - p + \sqrt{(C_a + q + p)^2 - 4(C_a q - p a_3)}\right]$$

$$E_0 = \frac{a_7 L^2}{L^2 + a_9}$$

$$\delta = -23.4\cos^{-1}\left(\frac{360(D+10)}{365}\frac{\pi}{180}\right)$$

$$s = 24\cos^{-1}(-\tan(lat)\tan(\delta))/\pi$$

$$G = \frac{E_0 I g_c (C_a - C_i)}{E_0 I + g_c (C_a - C_i)}(a_2 s + a_5)$$

List of Symbols:

	Stocks (units = g C m^{-2})
C_f	Foliar C mass
C_w	Wood C mass
C_r	Fine root C mass
C_{lit}	Fresh litter C mass
$C_{som/wd}$	Soil organic matter and woody debris
C_{lab}	Labile C mass
	Fluxes (units = g C m^{-2} d^{-1})
A_f	Foliage allocation rate
A_w	Wood allocation rate
A_r	Fine root allocation rate
A_{tlab}	Allocation rate into labile
A_{flab}	Allocation rate out of labile
L_f	Foliage litter production rate
L_w	Wood litter production rate
L_r	Fine root litter production rate
D	Decomposition rate of litter
G	Gross primary productivity
R_a	Autotrophic respiration rate
R_{alabt}	Respiration cost for allocation to labile
R_{alabf}	Respiration cost for allocation from labile
R_{h1}	Mineralisation of litter
R_{h2}	Mineralisation of SOM
NPP_1	Net primary productivity
NPP_2	NPP after allocation to foliage
NEE	Net ecosystem exchange

Table A1. DALEC model stocks and fluxes.

List of symbols (cont.):

Symbol	Description
g_c	Canopy conductance (gC m^{-2} d^{-1})
ψ_d	Max soil-leaf water potential difference (MPa)
T_r	Daily temperature range (°C)
R_{tot}	Total plant-soil hydraulic resistance (MPa m^{2}s mmol^{-1})
N	Foliar nitrogen (g N m^{-2} leaf area)
L	Leaf area index (m^{2} m^{-2})
T_{max}	Maximum daily temperature (°C)
C_a	Atmospheric CO_2 concentration (µmol mol^{-1})
C_i	CO_2 concentration at site of carboxylation (µmol mol^{-1})
E_0	Canopy level quantum yield (g C MJ^{-1} m^{-2} day^{-1})
δ	Solar declination (radians)
D	Day of year
s	Day length (hrs)
lat	Site latitude (°)
I	Irradiance (MJ m^{-2} day^{-1})

Table A2. ACM model parameters and inputs.

Parameters to be optimised:

	Description	Range (low/high)
p1	Decomposition rate (per day)	1×10^{-6}/0.01
p2	Fraction of GPP respired	0.2/0.7
p3	Fraction of NPP allocated to foliage	0.01/0.5
p4	Fraction of NPP2 allocated to roots	0.01/0.5
p5	Turnover rate of foliage (per day)	1×10^{-4}/0.1
p6	Turnover rate of wood (per day)	1×10^{-6}/0.01
p7	Turnover rate of roots (per day)	1×10^{-4}/0.01
p8	Mineralisation rate of litter (per day)	1×10^{-5}/0.1
p9	Mineralisation rate of SOM/CWD (per day)	1×10^{-6}/0.01
p10	Parameter in exponential term of temperature dependent rate parameter	0.05/0.2
p11	Nitrogen use efficiency parameter (a1) in ACM	5/20
p12 *	*GDD* value causing leaf out	200/400
p13 *	Minimum daily temperature causing leaf fall	8/15
p14 *	Fraction of leaf loss transferred to litter	0.2/0.7
p15 *	Turnover rate of labile carbon (per day)	1×10^{-4}/0.1
p16 *	Fraction of labile transfers respired	0.01/0.5
p17 *	Maximum C_f value (gC m^{-2})	100/500

Table A3. Model parameters requiring calibration. NPP_2 is NPP remaining after allocation to foliage.

*** parameters p12-17 are used in DALEC-deciduous only.**

Parameters to be used as constants:

Parameter	Optimized Value
a2	0.0156
a3	4.22273
a4	208.868
a5	0.0453
a6	0.3783
a7	7.1929
a8	0.0111
a9	2.1001
a10	0.7897

Table A4. Model parameters not requiring calibration. These are all in the ACM GPP model.

Site Descriptions, Drivers and Observations

Site description

The following details are provided for all the sites, EV1,EV2, EV3 and DE1, DE2, DE3 (EV = evergreen vegetation type, DE = deciduous vegetation type). For initial stocks estimates data are provided only for $C_{som/cwd}$ and C_w. C_f can be estimated from LAI observations and LMA provided in Table A5. Users must make their own assessments of initial stocks on C_r, C_{lit} and C_{lab}. These are all likely to be in the range 20-200 g C m^{-2}.

Site	**Expt**	**Latitude (°N)**	**Soil organic matter C (g C m^{-2} ground area) , $C_{som/cwd}$**	**Above ground biomass (g C m^{-2} ground area) C_w**	**Leaf mass per area (g C m^{-2} leaf area) LMA**	**Foliar N (g N m^{-2} leaf area)**
EV1	1 + 3	52	11000	9200	110	4.0
EV2	2 + 4	50	9700	12400	110	3.8
EV3	5	51	12000	16400	115	3.7
DE1	1 +3	48	7100	8800	22	1.0
DE2	2 + 4	51	9900	8900	22	1.1
DE3	5	51	12200	17600	22	1.3

Table A5. Site details

Model drivers

Times series of daily drivers for three years for each site are provided in comma delimited .csv files (EV1_drivers.csv etc.). The column headers for these files are:

Runday	Min temp (°C)	Max temp (°C)	Radiation (MJ d^{-1})	Atmospheric CO_2 (µmol mol^{-1})	Day of year

Table A6. Drivier file column headers

Observations

Time series of daily observations for each site are provided in comma delimited .csv files (EV1_OBS.csv etc.). The column headers for these files are:

Runday	NEE (g C m^{-2})	MODIS LAI (m^2 m^{-2})	GROUND LAI (m^2 m^{-2})

Table A7. OBS file column headers

Daily NEE has been calculated by summing half-hourly observations. Daily values are only calculated if 44 or more of the possible 48 observations passed quality control. NEE is determined by summing fluxes for all available periods each day. Typically data coverage is 20-30% of days.

MODIS LAI values are the mean of all pixels in a 7km x 7km grid of 1km resolution pixels centred on the flux tower site (n=49) with the land-cover class corresponding to the flux tower forest. Only those pixels passing the highest QC measure were included in this calculation.

Uncertainty in observations

Assessing the uncertainty in NEE observations is always difficult. A number of approaches were tried on these data sets including binning observations made in similar conditions (DOY, temperature and radiation) and assessing their variance. This approach produced wide distributions of variances, but with a mode and mean value approximating to 0.5 g C m^{-2} d^{-1}. This is in keeping with the generally accepted estimates of 8-15% for half hourly measurements and it is appropriate to define error with a fixed value (rather than coefficient of variation) given that NEE varies around zero.

The uncertainty in the MODIS observations of LAI is unknown. However, the large variation seen between adjacent pixels of same vegetation type, and between observation times for the same pixel, and the large discrepancies with ground based estimates of LAI, suggest that error in this measurement is relatively large. We estimate it to be at least 15%.

Uncertainty in ground based observations in LAI has been estimated from the variation in replicated measurements and although differing methodologies makes direct comparisons impossible coefficients of variation were found between 5% and 20%. For this experiment we recommend an error value of at least 10%.

Data sets and experiments

The observation data provided have been “censored” for the appropriate experiments.

EV1_OBS and DE1_OBS contain two years of all available observations for sites EV1 and DE1. These data are to be used in Exp. 1, that is to say that in conjunction with driver files EV1_drivers and DE1_drivers they are to be used to provide the best estimate of the model parameters and C pools and fluxes during the first two years.

EV2_OBS and DE2_OBS contain two years of observations synthetically generated from model runs using driver files EV2_drivers and DE2_drivers It is the aim in Exp. 2 to again estimate the model parameters and C pools and fluxes over the two years – although in this case the actually values are known. The synthetic data have been ‘sampled’ with a frequency and error equivalent to the actual observations in Exp. 1

EV1_drivers, EV2_drivers, DE1_drivers and DE2_drivers also contain an additional, third year of model drivers. These are to be used to make forecasts of C pools and fluxes in Exp. 3 and 4. The quality of these forecasts, with an emphasis on the associated growth in uncertainty, will be assessed against observations held for these sites during these times periods, but not revealed to participants at this stage.

EV3_OBS and DE3_OBS contain three years of MODIS LAI observations at sites which can be considered to be similar to sites EV1 and DE1 respectively. EV3_drivers and DE3_drivers contain drivers for these two sites and can be considered to be ‘accurate’, local measurements. These data are to be used in Exp. 5 when knowledge gained from Exp. 1 and 2 is to be extrapolated in conjunction with the basic LAI measurements to estimate C pools and fluxes at the paired sites. The quality of these extrapolations will be assessed against observations held for these sites but not revealed to participants at this stage.

References:

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., 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.

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