# Abstract

# Introduction

Life on earth is based on carbon; dry organic matter is 45 – 50% carbon. Tracking flows of carbon is therefore one of the main ways that scientists conceptualise the functioning of the biosphere. To grow, terrestrial plants need to acquire carbon from the atmosphere through the process of photosynthesis. Typically, about half of the carbon acquired through photosynthesis is respired, while the other half is used to grow new plant tissue (Waring et al. 1998). Studies of plant growth responses to environmental change therefore very commonly include measurements of photosynthetic responses. However, it is observed that in many circumstances, growth and photosynthesis do not correlate.

These circumstances include the onset of drought stress. It is well known that growth is more sensitive to water limitation than photosynthesis (e.g. Bradford & Hsiao 1982, Muller et al. 2011, Mitchell et al. 2014). Similarly, nutrient stress has been shown to impact on plant growth through reduction in allocation to growth rather than through a reduction in leaf photosynthetic rate (e.g. McDonald et al. 1986). Cross-species comparisons also frequently show that relative growth rate is not correlated with leaf-level photosynthetic rate (e.g. Curtis 1969, Poorter et al. 1990). At the forest stand scale, studies often show that temporal and spatial variations in gross primary production are not reflected in differences in wood production (e.g. Guillemot et al. 2015, Malhi et al. 2015).

This lack of correlation between photosynthesis and growth is a longstanding puzzle in plant ecophysiology that severely compromises our predictive understanding of the determinants of plant growth and survival. The reasons for the poor relationship between photosynthesis and growth are hotly debated. Körner (2003, 2013) has argued strongly that the reason for the lack of correlation is that plant growth is not carbon limited, but rather is limited by other environmental constraints. He argues that meristematic activity is limited by conditions such as cold temperatures and water stress and that under such conditions the supply of carbon from photosynthesis will exceed the plant’s ability to utilise the carbon in growth.

Additionally, plant survival of stressful events depends on adequate storage of reserves. Depletion of reserves during drought (‘carbon starvation’) is thought to be one of the causes of drought mortality (McDowell et al. 2008). Reserves are also needed for recovery following major disturbances such as fire or herbivory. There is an increasing awareness that if we are to understand plant vulnerability to stress and disturbance, we will need to quantify the rates at which plants form, and draw down from reserves (Sala et al. 2012). At present, vegetation models completely fail to capture these processes in a biologically meaningful way. We recently explored how leading ecosystem models represent storage of NSC (De Kauwe et al. 2014), and found that none included physiological hypotheses for the role of storage. Instead, storage was simply used as a buffer to allow net photosynthesis and growth to occur at different times of year. Different models implemented the buffer in quite different ways, resulting in radically different estimates of the storage pool size, ranging from 50 – 2000 g C m-2 (De Kauwe et al. 2014). The estimates of stored carbon are therefore not representative of the plant carbon reserves and could not be used to evaluate likely survival under stress.

One of the major stumbling blocks to advances in this area is a lack of quantitative information. Although it is often shown that growth and photosynthesis are not related, we do not generally know by how much they differ, on what timescale they differ, or even how much we might expect them to differ under alternative hypotheses. This lack of quantitative information makes it impossible to assess the importance of storage in determining plant growth and survival, and undermines our ability to predict likely future responses. The specific research goal of this paper is therefore to use quantitative approaches to calculate how much C *is* allocated to storage under belowground soil restricted conditions. Testing our theoretical predictions against the observations will provide important insights into the regulation of carbohydrate storage, and will significantly advance our ability to predict the impacts of environmental change on plant growth and vulnerability to stress.

We urgently need to quantify rates of carbon storage and utilisation. Many experiments do measure aspects of the carbon balance, including leaf photosynthesis, tissue respiration rates, NSC concentrations and plant biomass, but these measurements are very rarely connected to provide an overall estimate of the plant carbon balance. Here we use a mass balance approach, combining these measurements, to infer the rate and timing of transfer of NSC to and from storage. The goal of this research is to quantify from existing experimental data how much carbon is stored as reserves, and when. Storage and utilisation of reserves can be inferred from mass balance when there are frequent measurements of photosynthesis, respiration, NSC reserves, and growth. Mass balance approaches to infer carbon storage have been applied to data alone (Klein & Hoch 2015) or to process-based models parameterised with data (Schiestl-Aalto et al. 2015). Using data alone can be problematic because data are incomplete and subject to varying integration times, gaps and biases. There are also problems with applying process-based models to data because of the need to make model assumptions that can pre-determine the outcomes (Williams et al. 2005).

A mass balance approach can clearly show the size and timing of the temporal disconnect between photosynthetic uptake and growth (Klein & Hoch 2015). However, putting together a mass balance is not a simple back-calculation from the component measurements because there are significant uncertainties in each measurement. We use data assimilation technique (Williams et al. 2005) to infer the mass balances from measurements. Data assimilation (DA) is an approach that combines models and data; rather than specifying model parameters in advance, it attempts to find parameter ensembles most consistent with observations (Lorenc, 1995). The model contains the underlying physical principles – in our case the principle of mass balance. The available data are then incorporated into the model parameters by a fitting process, which takes account of data uncertainty. Algorithms commonly used for this process include MCMC (Markov Chain Monte Carlo) and Kalman filter approaches (Fox et al. 2009). The end result is an analysis with clear confidence intervals on states, fluxes and parameters, and their error covariances.

This approach is ideal because it will allow us to place uncertainty bounds on fluxes based on measurement uncertainty. Importantly, having uncertainty bounds allows us to estimate the likelihood of alternative hypotheses for storage dynamics. The data assimilation framework that we develop is applied to a pot experiment that has previously been carried out at the Hawkesbury Institute for the Environment, mining the existing data to generate substantial new information about the dynamics of carbon storage in plants.

# Methods

This study requires a range of quantitative techniques, including data assimilation and data synthesis. Here we use data assimilation technique, applied to existing experimental data, to infer the amount and timing of carbon fluxes in and out of storage to develop a new predictive understanding of the relationship between photosynthesis, growth and storage. We synthesise experimental data on temporal patterns of growth vs. storage over various belowground sink limitations.

Prof Mat Williams’ group is at the forefront of the development and application of DA methods in forest ecosystem ecology. They have developed and refined a DA framework (Williams et al. 2005, Fox et al. 2009, Bloom & Williams 2014) that combines experimental data with a simple carbon mass-balance model. Previously this framework has principally been applied to developing data-constrained estimates of total forest or landscape carbon uptake, but in doing so it also provides estimates of individual components of the carbon balance, such as fluxes into and out of a labile carbon pool. The framework is thus an ideal tool for quantifying carbohydrate storage and utilisation rates from experimental data.

We customise this existing framework to optimise its ability to infer these fluxes from experiments which have photosynthesis, respiration, NSC and growth measurements available. The specific modifications required are as follows.

(i) Allow the model to represent individual plants, in addition to forest canopies. Currently, the model estimates total photosynthetic carbon gain using a simple canopy-scale model (ACM, Aggregated Canopy Model) that assumes a horizontally homogeneous canopy. We will replace this model with a light use efficiency approach that is applicable to single unshaded trees, that is based on our group’s work in simulating light interception and utilisation in individual plant canopies (Duursma & Mäkelä 2007; Duursma et al. 2012).

(ii) Allow the assimilation of tissue-level measurements, such as measurements of leaf photosynthesis and dark respiration, which provide important constraints on the carbon budget. Currently the DA framework assumes that photosynthetic rates are fixed, and that there is a fixed fraction of total photosynthesis used in respiration. It is straightforward to elaborate the structure of the light-use efficiency model to incorporate dynamic photosynthetic and respiration rates.

(iii) Allow time-varying rate constants for the fluxes. Currently the fluxes among pools are estimated as a constant fraction of their donor pools, but we will need to estimate how fluxes vary among time-steps. This can be done using an ensemble Kalman Filter method (Williams et al. 2005).

With these modifications in place, the DA framework will be an invaluable tool for quantifying carbon fluxes in experimental systems, enabling us to extract significant new information from existing datasets.A significant advantage of the DA framework is that it quantifies uncertainty around carbon fluxes in addition to giving best estimates (Williams et al. 2005). We will thus be able to state the level of confidence that we have in the calculated carbon storage fluxes, and to estimate the likelihood of alternative hypotheses explaining these fluxes. It is also straightforward to adjust the relatively simple model to evaluate alternative model structures related to labile (non-structural) carbon.

## The study area

The pot experiment was located at the Hawkesbury Forest Experiment site (33°37'S 150°44'E) in Richmond, NSW, Australia. The site and experimental setup have been explained in detail by Campany et al. (Submitted). Briefly, at this site the topsoil is an alluvial formation of low-fertility sandy loam soils. It contains low organic matter and low capacity to hold moisture. A soil hard layer exists at ~1.0 m depth with a transition to heavy clay soils (Campany et al., Submitted). The site is in sub-humid temperate region with a distinct dry summer season. 20 weeks old Eucalyptus tereticornis seedlings in tube stock were chosen from a single local Cumberland plain cohort. Ten additional seedlings of similar age and growth were harvested to measure initial leaf area and dry mass of leaves, stems and roots.

Six container volumes were used ranging from 5 L to 35 L, with a 22.5 cm diameter, and lengths ranging from 15 to 100 cm. Containers were constructed of PVC pipe and were filled with local top soil (described above). Soil in each container was packed to achieve a target soil bulk density that matched local soil conditions of 1.7 g cm-3. A Imidacloprid (BAYER CropScience) insecticide tablet was planted 5 cm below the roots of each seedling. Containers were planted flush with the soil surface inside metal sleeves, designed to minimize excess air space between the container and outside soil while also allowing for container removal. This allowed for soil temperatures in containers to reflect conditions of naturally planted (free) seedlings. Each experimental block (n=7) contained a complete replicate set of six container volumes as well as one free seedling, with 1 m2 spacing. For each free seedling, used as the control, a 1 m2 subplot was excavated to the hard layer and replaced with the same soil used in each container. This ensured that any possible feedbacks from soil disturbance would be similar between seedlings in containers and free seedlings. A border of root exclusion material was buried 0.25 m deep and extended 0.25 m above the ground surface around each subplot to exclude local vegetation, which was further kept out by periodic weeding.

Plants were watered weekly or when needed to maintain soil moisture at field capacity (13-15 %). Drain systems were built into each pot to prevent pooling of water throughout the experiment. Pooling of water could lead to an anaerobic environment around the root that could hinder the uptake of water through reduced root conductance (Poorter et al. 2009), an undesired experimental artifact. A collection compartment in the bottom of containers, containing gravel covered by root exclusion mesh, was used to collect excess water for 20, 25, and 35 l containers. Plastic tubing (6 mm diameter) was inset into the gravel layer and extended through the top of the container. A lysimeter pump was then used to suction excess water, through the tubing, as needed. For small containers (5, 10, and 15 L) a simple bottom plug was used to drain excess water from the gravel compartment. Each container was inspected after every watering and rainfall event to determine if pooling had occurred.

**Carbon Balance Model (CBM)**

A simple Carbon Balance Model (CBM) of pools connected via fluxes is used to represent the C cycle similar to the previous studies (Bloom et al. 2016, Bloom and Williams 2015, Williams et al. 2005). There are six pools

for an individual free growing plant seedling (Eucalyptus Tereticornis) for daily time scale:

The inputs are on daily time scale:

1. Daily GPP (unit = gC),
2. Daily Respiration rates Rd (unit = gC g-1C).

The driver variables (data):

1. Cleaf (foliage biomass),
2. Cstem (stem biomass),
3. Croot (root biomass),
4. Sleaf (leaf turnover).

The 5 variables vary over time on temporal scale (just one parameter for whole 121 days / weekly / fortnightly / monthly etc.):

1. k (utilization coefficient),
2. Y (allocation fraction to biomass),
3. af (allocation fraction to foliage),
4. as (allocation fraction to stem),
5. sf (allocation fraction to leaf turnover).

GPP = Gross primary production (Photosynthesis)

Rd = Plant daily respiration (gC g-1C) = calculated based on leaf dark respiration and meteorological data (15 minutes temperature data): Rd \* (Cleaf + Cstem + Croot)

Not needed: Rd = Also calculated using modelled gross C gain (Cday\_gross) and net C gain (Cday\_net) from Court’s data: Rd = Ctotal\_gross(GPP) – Ctotal\_net(NPP)

Cleaf, Cstem, Croot = C pools of leaf, stem and root

Sleaf = leaf turnover

k = C fraction going out from storage pool (utilization coefficient)

af, as, ar = allocation fraction to foliage, stem and root (ar[i] + af[i] + as[i] = 1)

sf = leaf turnover fraction

Model Structure:

Sstem

Mleaf

Sroot

+

+

+

Clit

=

Mstem

=

Mroot

=

GPP

Plant respiration = Rd(Cl + Cs + Cr)

NPP

Storage pool, Cstorage

(Sleaf + Sstem + Sroot)

Allocation to Biomass

k.Cstorage

RGrowth

Y ~ 0.3

1-Y ~ 0.7

Cleaf

Cstem

Croot

af

as

ar

Sleaf

sf 

Model equations:

The notations are: M = for total mass (gC); S = for total storage (gC); C = for total C except the storage part (gC), so M = C + S.

Cstorage[i] = Cstorage[i-1] + GPP[i-1] - Rd[i-1]\*(Mleaf[i-1] + Mroot[i-1] + Mstem[i-1]) - k[i-1]\*Cstorage[i-1]

Sleaf[i] = Cstorage[i] \* 0.75 # 75% of storage goes to leaf (Duan's experiment)

Sstem[i] = Cstorage[i] \* 0.16 # 16% of storage goes to stem (Duan's experiment)

Sroot[i] = Cstorage[i] \* 0.09 # 9% of storage goes to root (Duan's experiment)

Cleaf[i] = Cleaf[i-1] + k[i-1]\*Cstorage[i-1]\*af[i-1]\*(1-Y) - sf[i-1]\*Cleaf[i-1]

Cstem[i] = Cstem[i-1] + k[i-1]\*Cstorage[i-1]\*as[i-1]\*(1-Y)

Croot[i] = Croot[i-1] + k[i-1]\*Cstorage[i-1]\*(1-af[i-1]-as[i-1])\*(1-Y)

No turnover for either stem and root as these are free growing small seedlings (20 weeks old)

Mleaf[i] = Cleaf[i] + Sleaf[i]

Mstem[i] = Cstem[i] + Sstem[i]

Mroot[i] = Croot[i] + Sroot[i]

# Results

## Synthetic experiments

**Processing the raw data sets**

Data source: Court's data files from the Soil Manipulation test

<https://github.com/kashifmahmud/CBM/Data_files>

Step 1: Find different soil volume pots from the data set-

Data: Volume designations by plot and pot ("plot\_summary.csv")

Step 2: Calculate daily leaf count-

Data: Weekly measurement of Leaf counts data ("leaf\_count\_1.csv"),

Steps:

* Count total leaf numbers for various soil manipulation tests (Weekly data)
* Interpolate daily leaf counts from the weekly measurements: Perform Cubic Spline interpolation to get daily leaf counts
* Plot interpolated daily leaf counts and save as “Leaf\_count\_daily.png”

Step 3: Calculate leaf area (LA) from harvest data-

Data: “harvest aboveground mass.csv”, daily leaf count from step 2

* Leaf area (t) = Leaf area (T) / Leaf count (T) \* Leaf count (t);

t = time, T = time of harvest

Step 4: Comparison between leaf area directly from harvest data (step 3) and considering the SLA changes-

Data: “harvest aboveground mass.csv”,

* sla\_harvest = leaf\_area\_harvest / leaf\_mass\_harvest / (100^2); unit – cm2 to m2
* Calculate leaf area considering the SLA changes
  + Modify SLA values considering changes of SLA due to new leaves only
* Plot the differences in SLA results (Measured SLA and Modified SLA)
* Find out the uncertainty (standard deviations) of weekly leaf mass from all 7 treatments

\*\*\* NO NEED TO CONSIDER SLA CHANGES TO GET THE LEAF AREA AS PER THE RESULTS AND DISCUSSION WITH BELINDA. THE SLA CHANGES ONLY REPRESENT THE UPPER CANOPY LEAVES (NEW ONES), NOT THE WHOLE CANOPY.

Step 5: Calculate leaf mass, Cleaf (leaf carbon pool) directly from harvest data-

Data: “harvest aboveground mass.csv”, daily leaf count from step 2

* Leaf mass (t) = Leaf mass (T) \* Leaf count (t) / Leaf count (T);

t = time, T = time of harvest

* Plot interpolated daily leaf mass using harvest data

Step 6: Analyse stem height and diameter data to estimate stem carbon pool, Cstem

Data: Weekly height diameter data ("height\_diameter.csv"; Diameter is in mm; Height is in cm), initial seedling data ("seedling\_initial.csv"), harvested seedling data for all different treatments ("seedling\_mass.csv")

* Linear regression model fitting log(stem\_mass) = b(1) + b(2)\*log(dia) + b(3)\*log(height)
* Fit the model with initial data (10) and harvested data for free seedlings (7)
* Estimate the stem mass from fitted linear regression equation
* Calculate the uncertainty (Standard deviations) of weekly stem mass from all 7 treatments
* Plot observation vs. modelled data to check the model accuracy

Step 7: Calculate daily GPP, NPP, R.plant (Plant respiration)-

Data: Daily leaf area (LA) calculated from harvested data (Step 3),

modelled daily net C gain in micromole CO2 m-2 d-1 (“cday\_120\_clean.csv”),

modelled daily gross C gain in micromole CO2 m-2 d-1 (“cday\_120\_clean\_gross.csv”),

self shading parameters (sigma) calculated from ‘YplantQMC’ package of Remko (“M\_leafarea\_model.csv"),

Total plant mass and leaf area at harvest (“harvest\_mass\_means.csv”),

\*\*\* cday\_120\_clean.csv and cday\_120\_clean\_gross.csv are calculated using coupled photosynthesis – stomatal conductance model with the ‘plantecophys’ package of Remko using mean photosynthetic parameters (Jmax, Vmax, R and g1)

* Generate total plant daily C gain (gross and net) using function:

modelledC\_func <- function(leafarea, shading, Cday)

* + Calculate daily self shading factor, M = b\*LA + intercept;

Where M is a linear function of leaf area; ‘b’ and ‘intercept’ were calculated from ‘YplantQMC’ for individual soil volume. This factor, M is accomplished by utilizing 61 previoulsy digitized Eucalyptus seedlings, covering 5 total species including E. Tereticornis.

* + Calculate gross daily C gain (GPP) with self shading:

GPP = Leaf\_area \* Cday\_gross \* M

* + Calculate net daily C gain (NPP) with self shading:

NPP = Leaf\_area \* Cday\_net \* M

* + Unit conversion for both GPP and NPP:

From micromole CO2 m-2 d-1to gDM m-2 d-1 by {\* (12/44) / 0.5};

**(12/44)** is for micromol CO2 to gC; **0.5** is for gC to gDM.

* Determine modelled daily plant respiration, R.plant (to compare with Rd.plant estimated from the leaf respiration, Rleaf measurements):

R.plant = GPP - NPP

* Calculate total seedling C gain over experiment (120 days) and compare to final harvest mass C
* Plot both the results (Modelled data and Measurements from harvest) for comparison

Step 8: Rd.plant prediction through time using rdarkq10 equation by volume-

Data: Leaf respiration measurements (“rdarkq10.csv") for the parameter rd12.3, which represents leaf dark respiration for various soil manipulation experiments at 12.30C, site weather data (“eucpve\_met.csv") for 15 mins temperature measurements

* Calculate Rd.plant through time using rdarkq10 equation:

Rd = rd12.3 \* q25\_drake^((temp-12.3)/10); Unit = micromol CO2 m-2 leaf d-1

where q25\_drake = 1.86 from John Drake’s experiment on Eucalyptus species, temp = temperatures at 15 mins interval from meteorological data.

* Rd\_daily = Rd \* sla\_harvest; Unit conversion to (micromol CO2 g-1 plant d-1)
* Unit conversion from micromole CO2 g-1 leaf d-1to gDM g-1 plant d-1 by {\* (12/44)}; **(12/44)** is for micromol CO2 to gC.
* Rd\_daily = Average of all (96) 15mins Rd\_daily data over one day

Step 9: Interpolate daily Cstem, Croot-

Data: Leaf area calculated from harvested data (step 3), stem mass, Cstem modelled based on stem diameter and height (step 6), root mass, Croot from initial and harvest data ("seedling\_initial.csv”, "seedling\_mass.csv")

* Interpolate (linear) daily stem mass from weekly modelled data
* Interpolate (linear) daily root mass from initial and harvest data
* Merge daily leaf mass with stem and root mass to get all Carbon pools together in one single data frame C.pool

Step 10: Estimate daily Plant respiration, rd.plant from Leaf respiration and temperature (15 mins data)

Data: Daily leaf mass, Cleaf calculated directly from harvest data (step 5),

* Sum up all daily Cleaf, Cstem, Croot (step 9) to get the total daily Carbon stock (Ctotal)
* Rd.plant = Rd\_daily \* (Cleaf + Cstem + Croot); Unit = gDM plant-1 d-1

Rd\_daily = Daily Plant respiration (gDM g-1 plant d-1)

* Comparison between whole plant respiration, Rd\_daily calculated from measured leaf respiration and modelled whole plant respiration, R.plant from GPP and NPP modelled data (step 7).

Step 11: Find leaf storage (tnc) for corresponding dates (from Court's Gas Exchange measurement campaign)

Data: Gas measurement campaign fortnightly data (“leaf\_data.csv”)

* Sum up starch and sugar contents (fortnightly data):

tnc = starch\_mgperg + sugars\_mgperg

* Calculate average fortnightly tnc and standard deviation from 6 replicates
* Unit conversion from (mg g-1leaf) to (g plant-1) by \* leafmass / 1000

**MCMC Algorithm settings**

To test MCMC with a simple Carbon balance model:

Setting lower and upper bounds of the prior parameter pdf, and start point of the chain [lower value, starting value, upper value]:

k = [0, 0.45, 1]

Y = [0.2, 0.3, 0.4]

af = [0, 0.45, 0.7]

as = [0, 0.17,0.5]

sf = [0, 0.02, 0.04]

Generate synthetic data for GPP, Rd, Cstorage with Mean and SD:

GPP (with mean=15, sd=3)

Rd (with mean=4, sd=0.8)

Cstorage (with mean=7.5, sd=2), just to create a measurement sets of Cleaf, Cstem, Croot

Import all the processed data set with uncertainty (Standard errors) from HIE Pot Experiment (Daily GPP, daily Rd, weekly Cleaf, weekly Cstem, twice Croot, fortnightly Sleaf.

* Using random parameter sets within the lower and upper bounds calculate Cleaf, Cstem, Croot to form synthetic data sets and uncertainties associated with these data
* Perform MCMC algorithm with a prior probability distribution for the parameters and a likelihood function
* Defining the model to iteratively calculate Cstorage, Mleaf, Mstem, Mroot:

model <- function (GPP,Rd,j,Mleaf,Mstem,Mroot,Y,k,af,as,sf)

* Calculating model outputs for the starting point of the chain and then the log likelihood based on measurements of Mleaf, Mstem, Mroot, Sleaf
* Store the first parameter set with log likelihood
* Calculate the next candidate parameter vector, as a multivariate normal jump away from the current point
* Reflected back if the candidate is not within the range to generate another candidate value
* Calculate the prior probability density for the candidate parameter vector
* Calculate the outputs for the candidate parameter vector and log likelihood
* Calculate the logarithm of the Metropolis ratio
* Accept or reject the candidate vector
* Discard the first part of the MC (500 iterations) for Burn-in process from the total chain length of 10,500
* Get a representative sample of parameter sets (k1,…..,kn; ….) where n = length of chain­­­
* Calculate final output set from the predicted parameter set and then the cumulative sums over the length of time
* Find the acceptance rate of the chain
* Find the correlation coefficients between original measurements and predictions
* Plot few accepted parameter values over time to find whether the chain converged
* Plot original measurements vs predictions for Cstorage, Cleaf, Cstem, Croot

**Best model selection:**

Different numbers of parameter are modelled for each of the 5 variables (k, Y, af, as, sf) based on different temporal scales using the CBM model equations. Then log likelihood, AIC (Akaike information criterion), BIC (Bayesian information criterion) and time taken by the model run are measured for model comparison. The minimum values of log likelihood, AIC, BIC and model run time represent the best model to choose (precisely the right numbers of parameters to select on temporal scale).

Equations:

log likelihood = (Prediction[i] – Measurement[i])2 / Measured SD[i]2) - log(Measured SD [i])

AIC = -2 \* log likelihood + k \* npar

Where k = 2; npar = total number of parameters in fitted model

BIC = -2 \* log likelihood + k \* npar

Where k = log(n); n = Total number of observations; npar = total number of parameters in fitted model

**References**

Bloom, A., Exbrayat, J.-F., Van der Velde, I., Feng, L. and Williams, M. (2016) The decadal state of the terrestrial carbon cycle: global retrievals of terrestrial carbon allocation, pools and residence times. Proceedings of the National Academy of Sciences 113(5).

Bloom, A.A. and Williams, M. (2015) Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model–data fusion framework. Biogeosciences 12(5), 1299-1315.

Williams, M., Schwarz, P.A., Law, B.E., Irvine, J. and Kurpius, M.R. (2005) An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology 11(1), 89-105.