**Carbon Balance Model (CBM)**

A simple Carbon Balance Model (CBM) 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]

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