Dynamic modelling of the carbon cycle in forests: Data needs and uncertainty quantification

Marcel van Oijen (CEH-Edinburgh)
1. Process-based modelling of forests
2. Data assimilation: The Bayesian approach
3. What kind of data do we need to reduce model output uncertainty?
4. Discussion
1. Process-based modelling of forests
1.1 Ecosystem PBMs simulate biogeochemistry

Atmosphere

N → C → H₂O

Subsoil

N → C → H₂O

Tree

Soil

H₂O
1.2 I/O of PBM

Atmospheric drivers
- N-deposition
- CO₂
- Radiation
- Temperature
- Rain
- Humidity
- Wind speed

Parameters & initial constants
- vegetation

Parameters & initial constants
- soil

Management & land use

Input

Model

Output

Time series of plant and soil variables
1.3 Applications of forest PBMs

Estimating C-sequestration across the UK (1920-2000)

Assessing how much of soil C is in ‘fast’ resp. ‘slow’ pools

Explaining time series of greenhouse gas emissions

[Van Oijen & Thomson, 2010]  [Yeluripati et al., 2009]  [Van Oijen et al., 2011]
1.4 Forest models and uncertainty

[Levy et al, 2004]
1.4 Forest models and uncertainty

Photosynthesis

Allocation

C-pools

N-pools

Parameter value

Parameter value

bgc

century

hybrid

N_{dep\ UE} (kg C kg^{-1} N)

[Levy et al, 2004]
2. Data assimilation: The Bayesian approach
2.1 Bayesian Calibration

Bayes’ Theorem

\[
P(\theta|D) = \frac{P(\theta) P(D|\theta)}{P(D)}
\]

Prior pdf for the parameters

Likelihood of the data

Posterior pdf for the parameters

Scaling constant

( = \int P(\theta) P(D|\theta) d\theta )
2.2 Process-based forest modelling

Environmental scenarios → Initial values → Parameters → Model

Model → Trees → Soil → Subsoil (or run-off)

Atmosphere

Inputs:
- NPP
- Soil C
- Height

Outputs:
- C
- H₂O
- Nutr.
2.3 Process-based forest model BASFOR

40+ parameters → BASFOR → 12+ output variables
2.4 BASFOR: outputs

- Volume (standing)
- Carbon in trees (standing + thinned)
- Carbon in soil
- Carbon in soil
2.5 BASFOR: parameter uncertainty
2.6 BASFOR: prior output uncertainty

- **Volume (standing)**
- **Carbon in trees (standing + thinned)**
- **Carbon in soil**
2.7 Data Dodd Wood (R. Matthews, Forest Research)
2.8 Using data in Bayesian calibration of BASFOR

Prior parameter marginal probability distributions (beta)

Prior pdf

Bayesian Calibration using MCMC

Data

Dodd Wood

Posterior pdf
2.9 Bayesian calibration: posterior uncertainty

- Volume (standing)
- Carbon in trees (standing + thinned)
- Carbon in soil

Graphs showing changes over time in various parameters:
- VolTtot (m$^3$ ha$^{-1}$)
- Vol (m$^3$ ha$^{-1}$)
- CtreeTot (kg m$^{-2}$)
- Ctree (kg m$^{-2}$)
- Cstem (kg m$^{-2}$)
- Cbranch (kg m$^{-2}$)
- Cleaf (kg m$^{-2}$)
- Croot (kg m$^{-2}$)
- Nsoil (kg m$^{-2}$)
- Soil (kg m$^{-2}$)

Parameters include:
- VolTtot: Volume Total
- Vol: Volume
- CtreeTot: Carbon in Trees Total
- Ctree: Carbon in Trees
- Cstem: Carbon in Stems
- Cbranch: Carbon in Branches
- Cleaf: Carbon in Leaves
- Croot: Carbon in Roots
- Nsoil: Nitrogen in Soil

Data points are plotted against time (x10$^4$) with error bars indicating uncertainty.
2.8 Using data in Bayesian calibration of BASFOR

Prior pdf

Bayesian calibration

Posterior pdf
2.10 Continued calibration when new data become available

[Diagram showing Bayesian calibration process with prior and posterior distributions, data points, and map of Dodd Wood and Rheola.]
2.10 Continued calibration when new data become available
2.11 GHG Inventory for the UK: Forest C-sequestration

Figure 1: Process based modeling of C-sequestration in the UK, 1920-20x20 km. Left: mean temperature (source: UKCIP). Mid: average sequestration simulated using forest model BASFOR. Right: uncertainty outputs shown in mid panel.

[Van Oijen & Thomson 2010]
2.12 Bayesian calibration instead of model spin-up

System: Grassland, Oensingen (Switzerland)
Model: DAYCENT
Data: Soil respiration

Bayesian calibration as a tool for initialising the carbon pools of dynamic soil models
Jagadeesh B. Yeluripati, Marcel van Oijen, Martin Wattenbach, A. Neftel, A. Ammann, W.J. Parton, Pete Smith (2009)
2.13 Integrating Remote Sensing data

System: Corsican pine forest, U.K.
Data: Remote Sensing → {LAI, Height, Biomass}

RS-data: Hyper-spectral, LiDAR, SAR

Bayesian calibration

Model 3-PG

Integrating remote sensing datasets into ecological modelling: a Bayesian approach

G. Patenaude¹, R. Milne², M. Van Oijen², C.S.Rowland³ and R.A. Hill³ (2008)
3. What kind of data do we need to reduce model output uncertainty?
3.1 Prior predictive uncertainty & height-data

**Height**

**Biomass**

- **Prior pred. uncertainty**
- **Height data**
- **Skogaby**
3.2 Prior & posterior uncertainty: use of height data

[Graphs showing height, biomass, Cr, NPP, LAI, Ntree, NCI, Csoil, Nsoil, Nmin, Miny over time for both prior and posterior uncertainty, with red lines indicating posterior uncertainty (using height data) and black lines indicating prior prediction uncertainty.]

Skogaby
3.2 Prior & posterior uncertainty: use of height data

- **Height**
- **Biomass**

### Prior pred. uncertainty

- **Posterior uncertainty (using height data)**

- **Height data (hypoth.)**
3.2 Prior & posterior uncertainty: use of height data
3.3 How to use the data

- So data can be informative in two ways:
  1. Having high precision (small S.D.)
  2. Being numerous (large $n$)
- But … how realistic is the previous example?
  - Only stochastic measurement error considered
  - Measurements may be biased
  - Measurements may be taken in an ‘atypical’ forest (e.g. forest on polluted soil)
  - Measurements may not be at the ‘believable scale’ of the model

$P(\theta|D) = \frac{P(\theta) P(D|\theta)}{P(D)}$
A protocol for Model-Data Fusion in C-Extreme

Q1. What magnitude of random measurement error is possible for each individual point in your data set?
Q2. What magnitude of systematic measurement error is possible for each variable in your data set?
Q3. What magnitude of representativeness error is possible for each variable in your data set?
3.5 The most common likelihood function

\[ P(D|\varnothing) = \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[ -\frac{1}{2} \left( \frac{f_i(\varnothing) - D_i}{\sigma_i} \right)^2 \right] \]

MULTIPLICATION ⇔
information from different data-points considered uncorrelated

GAUSSIAN

Model-data mismatch for data-point i

Measurement uncertainty for data-point i
3.6 Decision tree for likelihood functions

- **Points**
  - Not time-averaged
    - 30 d
    - 15 d
    - 8 d
    - 4 d
  - Time-averaged
- **Patterns**
- **Within TS**
  - Statistical Moments
    - Standard deviation
    - Variance
    - Skewness
  - Spectral analysis
- **Across TS**
  - Conserved ratios
- **Gaussian**
  - CV same for all data
  - 10%
  - 20%
  - 30%
- **Non-Gaussian**
  - CV point-specific
  - SD same for all data

Note: the diagram focuses on time-series data, but similar options exist for spatially distributed data.
3.6 Decision tree for likelihood functions

- Data
  - Points
    - Not time-averaged
    - Time-averaged
  - Patterns
    - Within TS
    - Across TS
  - Statistical Moments
    - Variance
    - Skewness
  - Spectral analysis
  - Conserved ratios
  - Standard deviation

- Gaussian
  - CV same for all data
    - 10%
    - 20%
    - 30%
  - SD same for all data

- Stochastic error
  - 30 d
  - 15 d
  - 8 d
  - 4 d

- Systematic error
  - Non-Gaussian
    - CV point-specific

Note: the diagram focuses on time-series data, but similar options exist for spatially distributed data.
1. Process-based models for forest C-tracking are available
2. Bayesian Calibration: easy to implement and effective
3. Not a clear question: “What data does the model need”? Instead ask: “How much do we want to reduce model output uncertainty?”
4. Data quality = fitness for purpose = informative to models Must be clear how to write the likelihood function
5. Defining the likelihood function requires understanding of both data (errors and representativeness) and model (believable scale) Data that are given with +- stochastic error are NOT well-enough defined Precision of measurement equipment (stochastic error) is far less important than data being unbiased and representative Future projects: data-providers assess data-quality interactively with modellers