Bayesian calibration of Yasso15 soil carbon model using global-scale litter decomposition and carbon stock measurements

Marko Järvenpää¹ Jari Liski² Anu Akujärvi² Mikko Kaasalainen¹

¹Department of Mathematics, Tampere University of Technology Tampere, Finland

²Finnish Environment Institute, Natural Environment Centre, Ecosystem Change Helsinki, Finland

10.6.2015 NBBC 2015 Reykjavik, Iceland
Outline

- Brief description of the Yasso15 model and data
- Bayesian calibration of the model
- Results and some analysis
Introduction

- Motivation of Yasso soil carbon model: need for a model that is
  - globally applicable
  - does not need detailed input data which is often hard to acquire
  - whose reliability can be assessed
The Yasso model

- Current Yasso model is based on the ideas in the papers

- Some applications of the model
  - Earth System Modelling (Max Planck Institute, Germany)
  - National greenhouse gas inventories
  - Climate change and land management effects of soil C
  - Several case studies
Motivation

- Use the information in a new global soil carbon dataset
- Study different temperature/precipitation dependencies for the compartments in the model
- Estimate parameters using this extended dataset and improved model
Basic assumptions of the Yasso model

- Soil carbon is split into 4+1 compartments: A, W, E, N + H
- Each pool decomposes at its own rate that depends on temperature and precipitation
- Carbon on pools A, W, E, N can decompose into CO₂ or to other pools or to H pool
- H pool is assumed to decompose slowly and it can only decompose into CO₂
- The decomposition rate of the woody litter (stumps, trunks, branches etc.) depends on their size
The Yasso model can be presented as linear differential equations with input $b(t)$ and initial condition $x(0) = x_0$,

$$x'(t) = \left( A(\theta, c) - \sum_{i=1}^{5} \delta_{S_i}(s) \omega_i P_a I_{5 \times 5} \right) x(t) + b(t),$$

$$A(\theta, c) = \begin{bmatrix} -1 & p_{WA} & p_{EA} & p_{NA} & 0 \\ p_{AW} & -1 & p_{EW} & p_{NW} & 0 \\ p_{AE} & p_{WE} & -1 & p_{NE} & 0 \\ p_{AN} & p_{WN} & p_{EN} & -1 & 0 \\ p_H & p_H & p_H & p_H & -1 \end{bmatrix} \cdot k(\theta, c),$$

$$k(\theta, c) = \text{diag}[k_A(\theta, c), k_W(\theta, c), k_E(\theta, c), k_N(\theta, c), k_H(\theta, c)],$$

$$k_i(\theta, c) = \frac{\alpha_i}{\sum_j} \exp(\beta_{i1} T_j + \beta_{i2} T_j^2)(1 - \exp(\gamma_i P_a)), \quad i \in \{A, W, E, N, H\}$$

where $x(t) = [x_A(t), x_W(t), x_E(t), x_N(t), x_H(t)]^T$ is the amount of litter,

$c = (T, P_a, d)$ denotes input data; $T = (T_1, ..., T_J)$ is temperature vector, $P_a$ annual precipitation, $d$ litter size,

$S_1, ..., S_5$ are woody size classes and $s$ is litter bag size,

$\alpha_i, p_{ij}, \omega_i, \beta_{i1}, \beta_{i2}, \gamma_i$ are the parameters; $\theta$ is the full parameter vector.
Decomposition rates are additionally scaled with

\[ h(d) = \min((1 + \phi_1 d + \phi_2 d^2)^r, 1), \]

where \( d \) is the size of litter and \( \theta_1, \theta_2 \) and \( r \) are additional parameters, to account slower decomposition of the woody litter.

In the earlier model it was assumed that \( \beta_{i1}, \beta_{i2} \) and \( \gamma_i \) parameters were the same for all compartments \( i \in \{A, W, E, N, H\} \).

Now we assume that the temperature/precipitation dependence is the same for A,W,E (so that e.g. \( \beta_{A1} = \beta_{W1} = \beta_{E1} =: \beta_1 \)) but can be different for N or H pools (so that e.g. \( \beta_1, \beta_{N1} \) and \( \beta_{H1} \) can have different values).
Available data 1/2

- **Decomposition measurements**
  - non-woody litter \((n \approx 11000)\), woody litter \((n \approx 2000)\)
  - mostly measured using litter bags (which causes systematic error that is corrected by modelling)
  - for some of this data individual AWEN masses are known, for some only total mass is known (initial AWEN masses are known for all data)
  - accuracy of data varies

- **Accumulation of soil carbon \((n = 26)\)**
  - measured on the southwest coast of Finland where the land rises from the sea and soil carbon accumulation was monitored

- **Steady state soil carbon stocks \((n = 60)\)**
  - measured in Finland
  - quite accurate
New data set: global steady state soil carbon measurements ($n \approx 4000$)
- data measured all over the world (blue circles in the image below)
- quite large measurement error
Likelihood

- Measurement error was assumed to follow the t-distribution (df \( \nu = 4 \)) to make inference robust to outliers:

\[
p(y \mid \theta) = \prod_{i=1}^{M} \prod_{j=1}^{N_i} \text{tpdf}_{\nu}(y_{ij} \mid m_i(\theta, c_{ij}, x_{0_{ij}}, b_{ij}), \sigma_{ij}^2)
\]

- Measurement error variances \( \sigma_{ij}^2 \) were estimated from the data beforehand and "plugged in" (thus ignoring uncertainty in them)

- Measurement errors were assumed to be additive and independent (although measurements at the same sites are likely correlated)

- Measurement errors in the input data \( c_{ij} \) etc. were assumed negligible (although they are also measured with error)

- Positivity constraints or correlations between AWEN fractions were not taken into account
## Prior information

<table>
<thead>
<tr>
<th>parameter</th>
<th>prior</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_A$</td>
<td>U(0,5)</td>
<td>decomposition rate of A</td>
</tr>
<tr>
<td>$\alpha_W$</td>
<td>U(0,20)</td>
<td>decomposition rate of W</td>
</tr>
<tr>
<td>$\alpha_E$</td>
<td>U(0,5)</td>
<td>decomposition rate of E</td>
</tr>
<tr>
<td>$\alpha_N$</td>
<td>U(0,2)</td>
<td>decomposition rate of N</td>
</tr>
<tr>
<td>$p_{ij}; i, j \in {A, W, E, N}, i \neq j$</td>
<td>see the next slide</td>
<td>relative mass flows</td>
</tr>
<tr>
<td>$\omega_i; i \in {1, ..., 5}$</td>
<td>U(-10,0)</td>
<td>leaching correction for size classes</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>U(0,0.2)</td>
<td>temperature dependence of A,W,E</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>U(-0.05,0)</td>
<td>temperature dependence of A,W,E</td>
</tr>
<tr>
<td>$\beta_{N1}$</td>
<td>U(0,0.2)</td>
<td>temperature dependence of N</td>
</tr>
<tr>
<td>$\beta_{N2}$</td>
<td>U(-0.05,0)</td>
<td>temperature dependence of N</td>
</tr>
<tr>
<td>$\beta_{H1}$</td>
<td>U(0,0.2)</td>
<td>temperature dependence of H</td>
</tr>
<tr>
<td>$\beta_{H2}$</td>
<td>U(-0.05,0)</td>
<td>temperature dependence of H</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>U(-20,0)</td>
<td>precipitation dependence of A,W,E</td>
</tr>
<tr>
<td>$\gamma_N$</td>
<td>U(-20,0)</td>
<td>precipitation dependence of N</td>
</tr>
<tr>
<td>$\gamma_H$</td>
<td>U(-20,0)</td>
<td>precipitation dependence of H</td>
</tr>
<tr>
<td>$\rho_H$</td>
<td>see the next slide</td>
<td>mass flow to H</td>
</tr>
<tr>
<td>$\alpha_H$</td>
<td>U(0,0.1)</td>
<td>decomposition rate of H</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>see the next slide</td>
<td>woody size dependence</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>see the next slide</td>
<td>woody size dependence</td>
</tr>
<tr>
<td>$r$</td>
<td>see the next slide</td>
<td>woody size dependence</td>
</tr>
</tbody>
</table>
prior + lhd = posterior

- $p_{ij}, p_H$ parameters have joint uniform prior with support in the set

\[
\{ p \in [0, 1]^{12} \times [0, 0.1] : \sum_j p_{ij} \leq 1, i = A, W, E, N \}
\]

- Parameters related to the woody litter $\theta_1, \theta_2, r$ have joint uniform prior with support in the set

\[
\{ \theta_1 \in [-10, 0], \theta_2 \in [0, 10], r \in [0, 1] : 4\theta_2 - \theta_1^2 \geq 0 \}
\]

- Individual parameters and the blocks above were assumed apriori independent $\rightarrow p(\theta)$
- Samples from the posterior $p(\theta | y) \propto p(\theta)p(y | \theta)$ were generated using Markov Chain Monte Carlo (MCMC) methods
  - quasi-Newton type optimization algorithm to find initial point (MAP-estimate)
  - adaptive MCMC scheme (DRAM by Haario et al. 2006) to sample from the whole posterior
### Yasso15 parameters: posterior median and 95% CI

<table>
<thead>
<tr>
<th>param.</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₀ₐ</td>
<td>0.44</td>
<td>0.49</td>
<td>0.54</td>
</tr>
<tr>
<td>αₖ₇</td>
<td>4.5</td>
<td>4.9</td>
<td>5.4</td>
</tr>
<tr>
<td>αₖ₈</td>
<td>0.22</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td>αₖ₉</td>
<td>0.075</td>
<td>0.095</td>
<td>0.12</td>
</tr>
<tr>
<td>pₖ₇₈</td>
<td>0.4</td>
<td>0.44</td>
<td>0.48</td>
</tr>
<tr>
<td>pₖ₈₉</td>
<td>0.044</td>
<td>0.25</td>
<td>0.46</td>
</tr>
<tr>
<td>pₖ₉₊</td>
<td>0.84</td>
<td>0.92</td>
<td>0.98</td>
</tr>
<tr>
<td>pₖ₊₁₀</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>pₖ₁₀ₙ</td>
<td>0.0027</td>
<td>0.084</td>
<td>0.26</td>
</tr>
<tr>
<td>pₖ₉₉₉</td>
<td>0.00029</td>
<td>0.011</td>
<td>0.041</td>
</tr>
<tr>
<td>pₖ₆₉₅</td>
<td>1.6e-05</td>
<td>0.00061</td>
<td>0.0022</td>
</tr>
<tr>
<td>pₖ₆₅₂</td>
<td>1.2e-05</td>
<td>0.00048</td>
<td>0.0018</td>
</tr>
<tr>
<td>pₖ₆₅₃</td>
<td>0.0074</td>
<td>0.066</td>
<td>0.14</td>
</tr>
<tr>
<td>pₖ₅₉₅</td>
<td>2e-05</td>
<td>0.00077</td>
<td>0.0028</td>
</tr>
<tr>
<td>pₖ₉₅₇</td>
<td>0.086</td>
<td>0.1</td>
<td>0.12</td>
</tr>
<tr>
<td>pₖ₇₅₈</td>
<td>0.47</td>
<td>0.65</td>
<td>0.82</td>
</tr>
<tr>
<td>ω₁₅</td>
<td>-0.16</td>
<td>-0.15</td>
<td>-0.15</td>
</tr>
<tr>
<td>ω₂₅</td>
<td>-0.024</td>
<td>-0.02</td>
<td>-0.015</td>
</tr>
<tr>
<td>ω₃₅</td>
<td>-1.1</td>
<td>-0.92</td>
<td>-0.78</td>
</tr>
<tr>
<td>ω₄₅</td>
<td>-0.0014</td>
<td>-0.0004</td>
<td>-1e-05</td>
</tr>
<tr>
<td>ω₅₅</td>
<td>-0.00061</td>
<td>-0.00017</td>
<td>-4.4e-06</td>
</tr>
<tr>
<td>β₁₅</td>
<td>0.084</td>
<td>0.091</td>
<td>0.1</td>
</tr>
<tr>
<td>β₂₅</td>
<td>-0.00054</td>
<td>-0.00021</td>
<td>-1.2e-05</td>
</tr>
<tr>
<td>β₃₅</td>
<td>0.039</td>
<td>0.049</td>
<td>0.059</td>
</tr>
<tr>
<td>β₄₅</td>
<td>-0.00029</td>
<td>-7.9e-05</td>
<td>-1.9e-06</td>
</tr>
<tr>
<td>β₅₅</td>
<td>0.028</td>
<td>0.035</td>
<td>0.047</td>
</tr>
<tr>
<td>γ₅₇</td>
<td>-1.9</td>
<td>-1.8</td>
<td>-1.7</td>
</tr>
<tr>
<td>γ₆₇</td>
<td>-1.5</td>
<td>-1.2</td>
<td>-0.91</td>
</tr>
<tr>
<td>γ₇₇</td>
<td>-20</td>
<td>-13</td>
<td>-6.4</td>
</tr>
<tr>
<td>p₇₉₅</td>
<td>0.0038</td>
<td>0.0046</td>
<td>0.0054</td>
</tr>
<tr>
<td>α₉₅</td>
<td>0.0011</td>
<td>0.0013</td>
<td>0.0015</td>
</tr>
<tr>
<td>θ₁₅</td>
<td>-1.3</td>
<td>-0.44</td>
<td>-0.023</td>
</tr>
<tr>
<td>θ₂₅</td>
<td>0.71</td>
<td>1.3</td>
<td>2.2</td>
</tr>
<tr>
<td>r</td>
<td>0.23</td>
<td>0.26</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Some parameters are highly correlated

- $\beta_1$ (r=−0.94) (r=−0.16) (r=0.03)
- $\beta_2$ (r=0.13) (r=−0.03)
- $\beta_{N1}$ (r=−0.45)
- $\beta_{N2}$

- $\gamma$ (r=−0.35)
- $\gamma_N$ (r=−0.03)
- $\gamma_H$ (r=−0.06)

M. Järvenpää (TUT)
Analyzing the fit: decomposition measurements

![Residuals plot](image)
Analyzing the fit: decomposition measurements

M. Järvenpää (TUT)  Bayesian calibration of Yasso15 model  June, 2015  17 / 25
Analyzing the fit: decomposition measurements

![Graph showing predicted mass residual vs. predicted mass](image)

Predicted mass residual (model−measurement)

M. Järvenpää (TUT)
Analyzing the fit: decomposition measurements

Data origin residual (model−measurement)

CIDET ED1 ED2 LIDET HOB3 Benin woody1 woody2

−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Data origin

M. Järvenpää (TUT)
Analyzing the fit: Accumulation of soil carbon measurements

chronosequence data

model − measurement
datapoint

M. Järvenpää (TUT)
Bayesian calibration of Yasso15 model
June, 2015
Analyzing the fit: steady state soil carbon measurements (Finland+global)

Residuals for Zinke data

Olson class

Residuals for Zinke data

M. Järvenpää (TUT)
Analyzing the fit: steady state soil carbon measurements (Finland+global)
"Sensitivity analysis"

- The results (obviously) depend on the assumptions that have been made about the model and data.
- Thus we also calibrated the model under some different assumptions.
- It was found that most parameters were very stable while some could change remarkably.
- For final results "uninformative" priors were used.

"Overfitting"

- Making fixed parameters free cannot decrease the goodness-of-fit.
- Different "information criteria": BIC, AIC, WAIC, LOO-CV etc., BMC, Occam's Razor etc.
- We did not perform full & formal analysis (due to computational and other reasons) but note that estimating the additional parameters certainly improved the fit.
Summary

- Yasso is a global-scale process-based model for soil carbon decomposition with many applications
- Extended the previous Yasso model to account different temperature/precipitation dependencies of compartments
- Calibrated the model with extended dataset and using minimal prior constraints
- Differences in the temperature/precipitation dependencies were observed
- Challenges: noise and systematic errors, model misspecification, uncertainty of model structure, overfitting
Thank you for your attention!