Multi-model comparison

Lorenzo Menichetti

2023-03-08

Aims of the project

1. Produce aggregated national predictions on the C sequestration potential of Swedish agricultural soils with included model structural uncertainty

2. Understand the impact of various factors on prediction robustness To achieve these two goal, we will combine a selection of compartmental first-order SOC models, calibrate the whole ensemble on a set of Swedish long-term experiments, and then analyze the results of the ensemble in respect to various factors (for example climate, location, edaphic conditions).

We will the utilize the ensemble for predictions, aggregated on a regional scale, about the total C sequestration potential in Swedish agricultural soils. These predictions will include a detailed uncertainty evaluation including also the model structural uncertainty.

Specific tasks

1 Data consolidation

2 Models

2.1 Model description: similarities

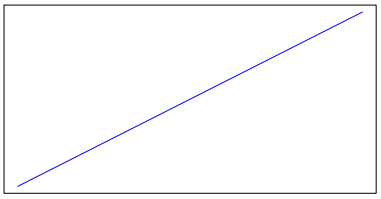
All models are possible to be written as:

$$\frac{d\Gamma}{dt} = I + \xi \cdot A \cdot \Gamma(t), \Gamma(t=0) = \Gamma_0 \tag{1}$$

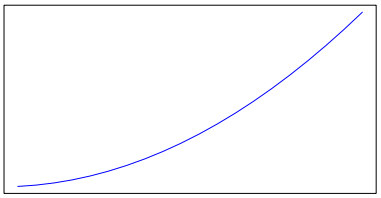
Where I is the inputs vector of lenght n, ξ is the climate or edaphic scaling (either one scalar or a vector of length n), A is the transfer matrix (square $n \times n$) and Γ_0 is the initialization vector of lenght n. Γ represents the total carbon C measured, distributed across all the pools of the model.
All models are assumed as non-autonomous and have forcing variables.
Different models will mainly differ in: * size (n) * transfer matrix A, which can have for example feedbacks, not all models are necessarily a cascade
And in the ways they estimate their forcing variables:
* in the way the estimate the inputs, usually variable over time * in the way they estimate the climatic scaling, also variable over time
All these differences will interact with each other. In order to test for the differences in model structures (n and A) without depending on the forcing variable estimation approaches, we will implement the latter into the Bayesian probabilistic model for calibration. The term ξ can be decomposed into separate temperature τ and moisture ψ effects.

2.1.2 The reduction functions τ and ξ

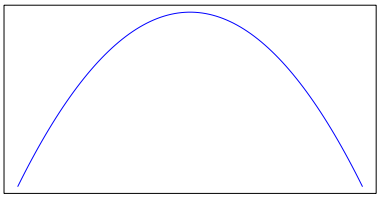
These functions are all classifiable under the following three groups:
Linear: representing the increase in activity with temperature or increase in activity with moisture in a more realistic way



Non-linear but monotonic: representing the increase in activity with temperature or increase in activity with moisture in a more realistic way



Non-linear and non-monotonic: representing the increase in activity with temperature and subsequent decrease due to processes like protein denaturation, or increase in activity with moisture with subsequent decrease due to lack of oxygen



2.1.3 The input estimation functions

Aboveground inputs are known. Belowground inputs are estimated based on total biomass yields. Ideally, belowground (roots) inputs tend to follow a linear relationship with aboveground (Bolinder et al. 2007) defined by one single coefficient, the shoot to root ration $S : R$ (so the relationship crosses the origin), $I = yields \cdot S : R$. One criticism to this approach is that plants can vary their allocation to belowground depending on their physiology, and it is possible that in more extreme treatments the relationship varies. We therefore considered as alternative method for input estimation a linear relationship including an intercept, $I = \alpha_1 + yields \cdot S : R$.

2.1.3 The exudates estimation functions

Root exudates are a relevant but also uncertain C input to the soil, and can be even more variable than root inputs. One possible way to estimate them is with a constant coefficient of total belowground inputs, as in (Bolinder et al. 2007): $E = I \cdot 0.65$. The original value 0.65 will be anyway considered uncertain.
But in order to include the uncertainty of the estimation approach itself, we consider in our models also the alternative of including an intercept: $E = \alpha_2 + I \cdot 0.65$

3 The Bayesian testing framework

3.1 Including model initialization in the calibration

Every model will need initial C values at the beginning of each simulation. Getting these values (usually called "initialization") will be done by including a solved (analytically or, in case of nonlinear models, numerically by running a spin-up) steady state version of the model into the Bayesian model. This means that the vector Γ_0 in Equation 1 is given by the vector of proportions (of length n) multiplied by the scalar representing the measurement of total C at time 0 $\Pi \cdot C_{t_0}$.

3.2 Including different amendments in the calibration

Each model includes one or more transfer terms between the first and the second pool. These terms, which are model parameters, are specific to different amendments.

3.3 The Bayesian model

We will build one single Bayesian model for each decomposition model, which will include all the possible combinations of forcing variables in subsequente hyerarchical levels. The graph of the probability model is represented below, where Θ represents the SOC decomposition model:

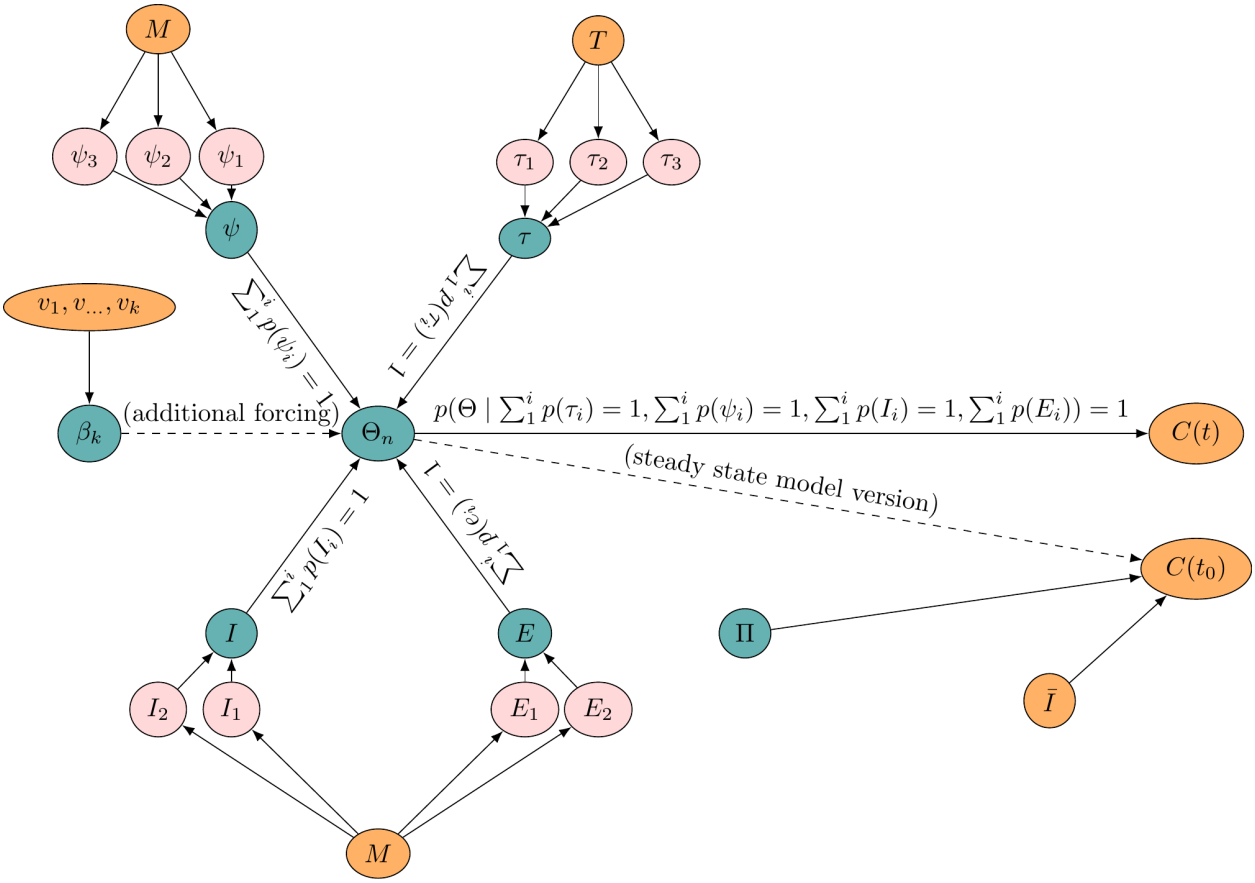


Figure 1: The core graph of the Bayesian model for each SOC decomposition model. Orange = data, pink = multiple choiches, teal = standard node

4 The models in detail

The first step is defining what a specific model is.
Here we will consider a SOC decomposition model as a compartmental model where compartments are in general connected with mass fluxes and defined by a decomposition kinetic associated with each pool. Fluxes can be in a cascade (mass has only one pathway through the pools) or in parallel (mass has more than one pathway through the pools).
Models can have feedbacks, with which we define only mass feedbacks (and not feedback mechanisms on the kinetics or transfers), so only fluxes going backwards.
The models which can present or not one or more inert pools, which are just a special case of pool with kinetic zero (and therefore present the peculiarity of simulate over time for SOC an asymptote translated above zero).
Models can also present interactions between the pools, which makes them nonlinear. We define as linear a model which presents as variable (besides forcing variables) only the pool masses. A nonlinear model presents other variables, usually the kinetics but in some cases something else like the input partitioning, which also vary as a function of the pools.
All these charachteristics can appear independently from each other.

4.1 Linear models

4.1.1 Cascade models

4.1.2 Parallel models

4.1.3 Feedback models

4.1.4 Models with inert pool

4.2 Nonlinear models

5 Comparing models

6 Extrapolating national predictions from the ensemble

References

Bolinder, M. A., H. H. Janzen, E. G. Gregorich, D. A. Angers, and A. J. VandenBygaart. 2007. "An Approach for Estimating Net Primary Productivity and Annual Carbon Inputs to Soil for Common Agricultural Crops in Canada." *Agriculture, Ecosystems & Environment* 118 (1-4): 29–42. https://doi.org/10.1016/j.agee.2006.05.013.