

Conceptual Design of the Biogeochemistry Model Database BGC-MD

TEE Group
Max Planck Institute for Biogeochemistry, Jena, Germany

November 10, 2015

Abstract

This document serves as the basis for the conceptual development of the BGC-MD. It attempts at proposing a proptotype to implement computationally. It also presents examples of models that can be integrated within this framework.

Contents

1	Introduction	2
2	Mathematical abstraction	2
3	Example the IBIS model	3
4	Example: the G'DAY model	4
5	Implementation	7
5.1	Input files	7
5.2	Standard (MIRIAM?)	7
5.3	The model object	8
5.4	Dictionary/glossary	8
5.5	Tests	8
5.6	Reports	8
5.7	Automated code generation	8
5.8	Queries	8
5.9	Documentation	8
6	User types	8

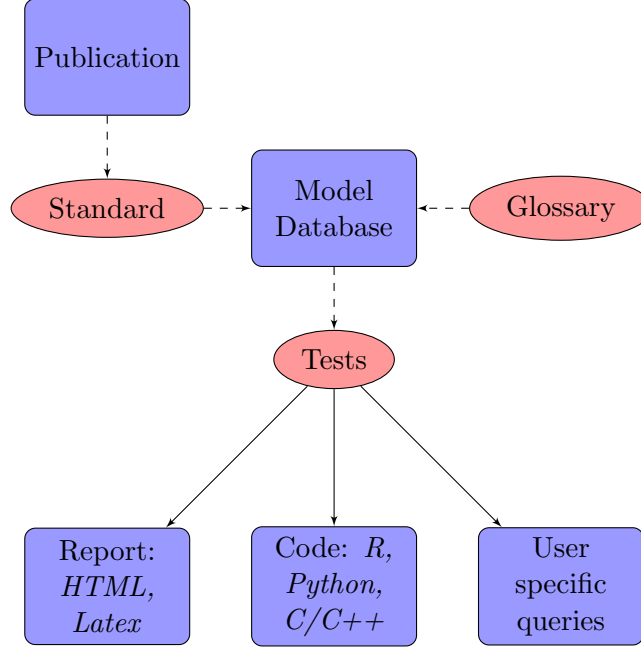


Figure 1: Conceptual design of the BGC-MD.

1 Introduction

2 Mathematical abstraction

The main objective of the conceptual framework is to propose a general abstraction to represent any ecosystem model. The simplest and most trivial abstraction is to represent ecosystem models as an n -dimensional dynamical system. However, this is too general to extract useful information from it. We need to find a representation of a dynamical system with some restrictions on the relations it represents.

A possible candidate is a model expressed as

$$f(\mathbf{x}, t) = u(t) \cdot \mathbf{b}(\mathbf{x}, t) + \mathbf{T}(\mathbf{x}, t) \cdot \mathbf{N}(\mathbf{x}, t) \cdot \mathbf{x}(t) \quad (1)$$

This equation combines the general linear model of Weng and Luo (2011) with the general soil carbon model we developed previously. The assumption here is that a scalar function determines photosynthetic inputs $u(t)$, which are allocated to different vegetation components according to the function valued vector $\mathbf{b}(\mathbf{x}, t)$. Notice that this vector allows for nonlinearities.

We can potentially split this framework between models for vegetation and models for soils. For vegetation, the model could be expressed as

$$f_v(\mathbf{x}_v, t) = u(t) \cdot \mathbf{b}(\mathbf{x}_v, t) + \mathbf{A} \cdot \mathbf{x}_v \quad (2)$$

and for the soil

$$f_s(\mathbf{x}_s, t) = \mathbf{I}(t) + \mathbf{T}(\mathbf{x}_s, t) \cdot \mathbf{N}(\mathbf{x}_s, t) \cdot \mathbf{x}_s(t) \quad (3)$$

Notice that the vector of states was split into a vector of states for vegetation \mathbf{x}_v and a vector of states for soil \mathbf{x}_s .

3 Example the IBIS model

The IBIS model was described in Foley et al. (1996). In this model, the annual amount of fixed carbon as GPP is given by

$$GPP = \int A_g(t) dt \quad (4)$$

where $A_g = \min(J_e, J_c, J_s)$; i.e. the minimum of three potential capacities to fix carbon. These photosynthetic capacities are functions that depend on the concentrations of CO_2 and O_2 in the atmosphere and other photosynthetic parameters. The integral in (4) is over an entire year, similarly as the calculation of net primary production

$$NPP = (1 - \eta) \int (A_g(t) - R_{leaf}(t) - R_{stem}(t) - R_{root}(t)) dt \quad (5)$$

where $\eta = 0.33$ is the fraction of carbon lost in the construction of plant material because of growth respiration. Respiration for each plant part is given by

$$R_{leaf} = R_l = \gamma \cdot V_m \quad (6)$$

$$R_{stem}(t) = R_s(t) = \beta_{stem} \cdot \lambda_{sapwood} \cdot C_{stem} \cdot f(T_{stem}(t)) \quad (7)$$

$$R_{root}(t) = R_r(t) = \beta_{root} \cdot C_{root} \cdot f(T_{soil}(t)) \quad (8)$$

where γ is the leaf respiration cost of Rubisco activity, V_m is the maximum capacity of Rubisco, β is a maintenance respiration coefficient, and $\lambda_{sapwood}$ is the sapwood fraction of the total stem biomass.

For the three compartments: leaf, stem, and roots, the system of ODEs is given by

$$\frac{dC_i}{dt} = a_i \cdot NPP - \frac{C_i}{\tau_i} \quad (9)$$

where a_i is the allocation coefficient for compartment i , and τ_i is the ‘residence time’ of carbon in each compartment. This version of the IBIS model has not a soil compartment.

We can express this model as a system of ODEs as

$$\frac{dC_l}{dt} = a_l(1 - \eta) \int (A_g(t) - R_l - R_s(t) - R_r(t))dt - \frac{C_l}{\tau_l} \quad (10)$$

$$\frac{dC_s}{dt} = a_s(1 - \eta) \int (A_g(t) - R_l - R_s(t) - R_r(t))dt - \frac{C_s}{\tau_s} \quad (11)$$

$$\frac{dC_r}{dt} = a_r(1 - \eta) \int (A_g(t) - R_l - R_s(t) - R_r(t))dt - \frac{C_r}{\tau_r} \quad (12)$$

which can be expressed in the form of equation (2) assuming the vectors

$$\mathbf{x}_v = \begin{pmatrix} C_l \\ C_s \\ C_r \end{pmatrix}, \mathbf{b} = \begin{pmatrix} a_l \\ a_s \\ a_r \end{pmatrix},$$

and a scalar valued function for t in years

$$u(t) = (1 - \eta) \int (A_g(t) - R_l - R_s(t) - R_r(t))dt,$$

and a matrix

$$\mathbf{A} = \begin{pmatrix} -1/\tau_l & 0 & 0 \\ 0 & -1/\tau_s & 0 \\ 0 & 0 & -1/\tau_r \end{pmatrix}.$$

The model is therefore expressed as

$$\begin{aligned} f_v(\mathbf{x}_v, t) &= \left((1 - \eta) \int (A_g(t) - R_l - R_s(t) - R_r(t))dt \right) \cdot \begin{pmatrix} a_l \\ a_s \\ a_r \end{pmatrix} \\ &+ \begin{pmatrix} -1/\tau_l & 0 & 0 \\ 0 & -1/\tau_s & 0 \\ 0 & 0 & -1/\tau_r \end{pmatrix} \cdot \begin{pmatrix} C_l \\ C_s \\ C_r \end{pmatrix}. \end{aligned} \quad (13)$$

4 Example: the G'DAY model

The G'DAY model is described in detail in Comins and McMurtrie (1993). Net carbon production (similar to NPP) in this model is given by

$$u(t) = G(t) = G_{max} \cdot I(C_f) \cdot E(\nu_f), \quad (14)$$

where G_{max} is a function of atmospheric CO₂ concentrations, incident photosynthetically active radiation (PAR), and a PAR utilization efficiency constant. Given that CO₂ concentrations change over time due to mostly anthropogenic activities, G_{max} is a function of time. $I(C_f)$ is the light interception factor, a function of the foliage biomass carbon C_f . The function

$E(\nu_f)$ is a rate modifying factor of PAR use efficiency as a function of the C:N ratio of the foliage ν_f .

The system of equations for the foliage f , root r , and wood w pools is given by

$$\frac{dC_f}{dt} = \eta_f G(t) - \gamma_f C_f \quad (15)$$

$$\frac{dC_r}{dt} = \eta_r G(t) - \gamma_r C_r \quad (16)$$

$$\frac{dC_w}{dt} = \eta_w G(t) - \gamma_w C_w \quad (17)$$

where the γ_i are senescence rates for each pool. The vegetation component of the model is therefore given by

$$\begin{aligned} f_v(\mathbf{x}_v, t) = & G_{max}(t) \cdot I(C_f) \cdot E(\nu_f) \cdot \begin{pmatrix} \eta_f \\ \eta_r \\ \eta_w \end{pmatrix} \\ & + \begin{pmatrix} -\gamma_f & 0 & 0 \\ 0 & -\gamma_r & 0 \\ 0 & 0 & -\gamma_w \end{pmatrix} \cdot \begin{pmatrix} C_f \\ C_r \\ C_w \end{pmatrix}. \end{aligned} \quad (18)$$

The litter and soil compartments are divided in the four pools for the litter: surface structural litter u , surface metabolic litter m , soil structural litter v , and soil metabolic litter n . The soil pools are: active a , slow s , and passive p . The system of equations for these pools is given by

$$\frac{dC_u}{dt} = p_{uf}\gamma_f C_f + \gamma_w C_w - d_1 C_u \quad (19)$$

$$\frac{dC_m}{dt} = p_{mf}\gamma_f C_f - d_2 C_m \quad (20)$$

$$\frac{dC_v}{dt} = p_{vr}\gamma_r C_r - d_3 C_v \quad (21)$$

$$\frac{dC_n}{dt} = p_{nr}\gamma_r C_r - d_4 C_n \quad (22)$$

$$\frac{dC_a}{dt} = p_{au}d_1 C_u + p_{am}d_2 C_m + p_{av}d_3 C_v + p_{an}d_4 C_n + p_{as}d_6 C_s + p_{ap}d_7 C_p - d_5 C_a \quad (23)$$

$$\frac{dC_s}{dt} = p_{su}d_1 C_u + p_{sv}d_3 C_v + p_{sa}d_5 C_a - d_6 C_s \quad (24)$$

$$\frac{dC_p}{dt} = p_{pa}d_5 C_a + p_{ps}d_6 C_s - d_7 C_p \quad (25)$$

where the partitioning coefficients are either constants or functions of the lignin content and the C:N ratio of specific pools. The decomposition coefficients d are also modified by a temperature dependent function $f(T)$ and in some cases by an exponential function of the lignin content.

In matrix form, the soil model can be expressed as

$$\begin{aligned}
f(\mathbf{x}_s, t) &= \mathbf{I} + \mathbf{T} \cdot \mathbf{N} \cdot \mathbf{x}_s = \mathbf{I} + \mathbf{A} \cdot \mathbf{x}_s \\
&= \begin{pmatrix} p_{uf}\gamma_f C_f + \gamma_w C_w \\ p_{mf}\gamma_f C_f \\ p_{vr}\gamma_r C_r \\ p_{nr}\gamma_r C_r \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&+ \begin{pmatrix} -d_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -d_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -d_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -d_4 & 0 & 0 & 0 \\ p_{au}d_1 & p_{am}d_2 & p_{av}d_3 & p_{an}d_4 & -d_5 & p_{as}d_6 & p_{ap}d_7 \\ p_{su}d_1 & 0 & p_{sv}d_3 & 0 & p_{sa}d_5 & -d_6 & 0 \\ 0 & 0 & 0 & 0 & p_{pa}d_5 C & p_{ps}d_6 & -d_7 \end{pmatrix} \cdot \begin{pmatrix} C_u \\ C_m \\ C_v \\ C_n \\ C_a \\ C_s \\ C_p \end{pmatrix} \\
&\quad (26)
\end{aligned}$$

To link both the vegetation and the soil model into a single ecosystem model, we can re-write the model as

$$\begin{aligned}
f(\mathbf{x}, t) &= u(t) \cdot \mathbf{b}(\mathbf{x}, t) + \mathbf{T}(\mathbf{x}, t) \cdot \mathbf{N}(\mathbf{x}, t) \cdot \mathbf{x}(t) \\
&= G_{max}(t) \cdot I(C_f) \cdot E(\nu_f) \cdot \begin{pmatrix} \eta_f \\ \eta_r \\ \eta_w \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&+ \begin{pmatrix} -\gamma_f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\gamma_w & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{uf}\gamma_f & 0 & \gamma_w & -d_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{mf}\gamma_f & 0 & 0 & 0 & -d_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{vr}\gamma_r & 0 & 0 & 0 & -d_3 & 0 & 0 & 0 & 0 \\ 0 & p_{nr}\gamma_r & 0 & 0 & 0 & 0 & -d_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{au}d_1 & p_{am}d_2 & p_{av}d_3 & p_{an}d_4 & -d_5 & p_{as}d_6 & p_{ap}d_7 \\ 0 & 0 & 0 & p_{su}d_1 & 0 & p_{sv}d_3 & 0 & p_{sa}d_5 & -d_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{pa}d_5C & p_{ps}d_6 & -d_7 \end{pmatrix} \\
&\cdot \begin{pmatrix} C_f \\ C_r \\ C_w \\ C_u \\ C_m \\ C_v \\ C_n \\ C_a \\ C_s \\ C_p \end{pmatrix}.
\end{aligned} \tag{27}$$

5 Implementation

This section should describe the technical aspects about the infrastructure that allows to import information about a paper, create a model object, manipulate the object, and print information.

5.1 Input files

Describe the Yaml files and their structure.

5.2 Standard (MIRIAM?)

Consider adopting a modified version of the MIRIAM standard (Novere et al., 2005).

5.3 The model object

Provide a description of the object created from the Yaml files.

5.4 Dictionary/glossary

Describe structure of dictionary.

5.5 Tests

Described automated tests.

5.6 Reports

Describe the automatic generation of reports.

5.7 Automated code generation

Describe procedure for translating code to different languages.

5.8 Queries

Describe possible strategy for doing calculations with the models.

5.9 Documentation

There are three levels at which these database is documented. First, standard Python documentation is created automatically from the source code. This documentation is available with the package and also at (url). A second source of documentation is this paper that describes the main motivation and scientific basis for the implementation of the database. A third source of documentation, are simple examples of user cases. These examples are available at (url).

6 User types

We distinguish between four types of users:

1. **Type I: The Browser.** This is the most general user level. This type of user is only interested in browsing through the dataset and exploring the automatically produced report. It uses any web browser to go around the database.

2. **Type II: The Author.** This type of user is interested in adding it's own model to the dataset. This user type must write an entry for the database in a Yaml file using Markdown syntax for special notation such as sup- subscripts, etc. This user must also write equations in the Yaml file using Sympy syntax.
3. **Type III: The Scripter.** This type of user would write scripts in python to make specific queries to the database.
4. **Type IV: The Hacker.** This type of user would be able to modify the source code of the database and modify its functionality.

Knowledge on programing, in particular python and sympy is increasingly required from user type I to IV. In fact, for user type I there is not requirement on programing to be able to browse and explore the automatically generated reports from the database. User type IV on the contrary, would need an in-depth knowledge of programing in python and would be able to modify important aspects of the functionality of the database.

References

- Comins, H. N. and McMurtrie, R. E. (1993). Long-term response of nutrient-limited forests to CO₂ enrichment; equilibrium behavior of plant-soil models. *Ecological Applications*, 3(4):666–681.
- Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A. (1996). An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles*, 10(4):603–628.
- Novere, N. L., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J. L., Spence, H. D., and Wanner, B. L. (2005). Minimum information requested in the annotation of biochemical models (miriam). *Nat Biotech*, 23(12):1509–1515.
- Weng, E. and Luo, Y. (2011). Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. *Ecological Applications*, 21(5):1490–1505.