

Reproducible Carbon Cycle Models

Biogeochemical Model Database `bgc_md2`



Markus Müller, Holger Metzler, Verónica Ceballos Núñez, Kostiantyn Viatkin, Thomas Lotze, Jon Wells, Yu Zhou, Cuijuan Liao, Aneesh Chandel, Feng Tao, Yuanyuan Huang, Alison Bennett, Chenyu Bian, Lifeng Jiang, Song Wang, Chengcheng Gang, Carlos Sierra, Yiqi Luo

Why would **you** want a model data base?



Markus Müller, Holger Metzler, Verónica Ceballos Núñez,
Kostiantyn Viatkin, Thomas Lotze, Jon Wells, Yu Zhou,
Cuijuan Liao, Aneesh Chandel, Feng Tao, Yuanyuan Huang,
Alison Bennett, Chenyu Bian, Lifan Jiang, Song Wang,
Chengcheng Gang, Carlos Sierra, Yiqi Luo

- **Find** (rather than reinvent) the **right model** for a specific **task**?



Why would **you** want a model data base?

- **Find** (rather than reinvent) the **right model** for a specific **task**?
- **Implement** a **new** model but start from a similar one?



Why would **you** want a model data base?

- **Find** (rather than reinvent) the **right model** for a specific **task**?
- **Implement** a **new** model but start from a similar one?
- **Use common infrastructure** to compute diagnostics that are difficult to implement.

Why would **you** want a model data base?

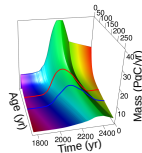
- **Find** (rather than reinvent) the **right model** for a specific **task**?
- **Implement** a **new** model but start from a similar one?
- **Use common infrastructure** to compute diagnostics that are difficult to implement. e.g. carbon flux diagrams, transit times, age distributions . . .

$$\frac{d}{dt} \begin{bmatrix} \text{leaf} \\ \text{wood} \end{bmatrix} = \begin{bmatrix} l_{\text{leaf}}(t) \\ l_{\text{wood}} \end{bmatrix} + \begin{bmatrix} -k_{\text{leaf} \rightarrow \text{wood}} - k_{\text{leaf},0}(t) & k_{\text{wood} \rightarrow \text{leaf}} \\ k_{\text{leaf} \rightarrow \text{wood}} & -k_{\text{wood} \rightarrow \text{leaf}} - k_{\text{wood},0} \end{bmatrix} \begin{bmatrix} \text{leaf} \\ \text{wood} \end{bmatrix}$$

■ tree ■ no data dependency ■ undetermined



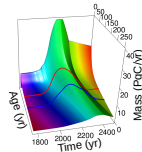
■ tree ■ no data dependency ■ undetermined



Why would **you** want a model data base?

- **Find** (rather than reinvent) the **right model** for a specific **task**?
- **Implement** a **new** model but start from a similar one?
- **Use common infrastructure** to compute diagnostics that are difficult to implement. e.g. carbon flux diagrams, transit times, age distributions . . .

$$\frac{d}{dt} \begin{bmatrix} \text{leaf} \\ \text{wood} \end{bmatrix} = \begin{bmatrix} I_{\text{leaf}}(t) \\ 0 \end{bmatrix} + \begin{bmatrix} -k_{\text{leaf} \rightarrow \text{wood}} - k_{\text{leaf} \rightarrow \text{lit}} & k_{\text{wood} \rightarrow \text{leaf}} \\ k_{\text{leaf} \rightarrow \text{wood}} & -k_{\text{wood} \rightarrow \text{leaf}} - k_{\text{wood} \rightarrow \text{cwd}} \end{bmatrix} \begin{bmatrix} \text{leaf} \\ \text{wood} \end{bmatrix}$$

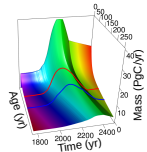


- Find and **reduce** sources of **uncertainty** in carbon predictions

Why would **you** want a model data base?

- **Find** (rather than reinvent) the **right model** for a specific **task**?
- **Implement** a **new** model but start from a similar one?
- **Use common infrastructure** to compute diagnostics that are difficult to implement. e.g. carbon flux diagrams, transit times, age distributions . . .

$$\frac{d}{dt} \begin{bmatrix} \text{leaf} \\ \text{wood} \end{bmatrix} = \begin{bmatrix} l_{\text{leaf}}(t) \\ l_{\text{wood}} \end{bmatrix} + \begin{bmatrix} -k_{\text{leaf} \rightarrow \text{wood}} - k_{\text{leaf},0}(t) & k_{\text{wood} \rightarrow \text{leaf}} \\ k_{\text{leaf} \rightarrow \text{wood}} & -k_{\text{wood} \rightarrow \text{leaf}} - k_{\text{wood},0} \end{bmatrix} \begin{bmatrix} \text{leaf} \\ \text{wood} \end{bmatrix}$$



- Find and **reduce** sources of **uncertainty** in carbon predictions
- ... ???

Markus Müller, Holger Metzler, Verónica Ceballos Núñez,
Kostiantyn Viatkin, Thomas Lotze, Jon Wells, Yu Zhou,
Cuijuan Liao, Aneesh Chandel, Feng Tao, Yuanyuan Huang,
Alison Bennett, Chenyu Bian, Lifeng Jiang, Song Wang,
Chengcheng Gang, Carlos Sierra, Yiqi Luo



We need:

- **collections**



How can you build one?

We need:

- **collections**

- ▶ of **many models**: (from A to Z):

Arora2005GCB-1 ,CARDAMOM , . . . ,

Zelenev2000MicrobialEcology

How can you build one?

We need:

■ collections

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,. . . ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by

We need:

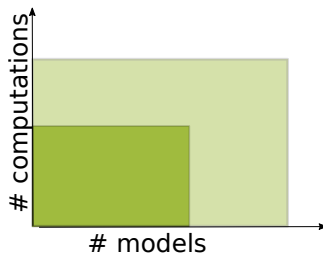
- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM , . . . ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by

How can you build one?

We need:

- **collections**

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,... ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by



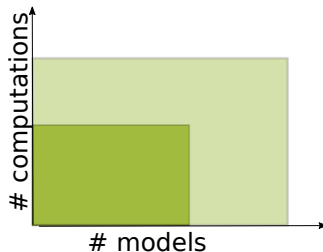
- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...

How can you build one?

We need:

- **collections**

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,... ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by



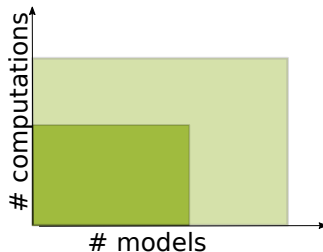
- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...
- better use the **dry** principle
Don't repeat yourself

How can you build one?

We need:

- **collections**

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,... ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by



- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...
- better use the **dry** principle
Don't repeat yourself

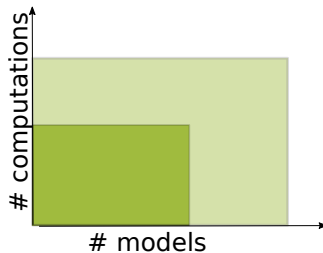
How can you build one?

We need:

- **collections**

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,... ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by

- Ways of **organizing** both collections



- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...
- better use the **dry** principle
Don't repeat yourself

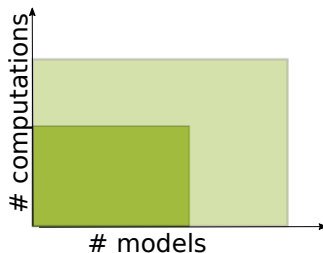
How can you build one?

We need:

■ collections

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,... ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by

- Ways of **organizing** both collections
 - ▶ building blocks for models



- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...
- better use the **dry** principle
Don't repeat yourself

How can you build one?

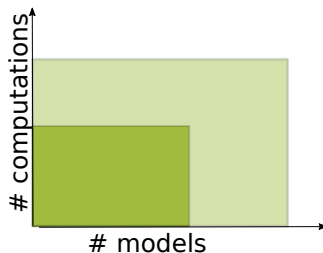
We need:

■ collections

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM ,... ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by

■ Ways of **organizing** both collections

- ▶ building blocks for models
- ▶ functions of building blocks



- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...
- better use the **dry** principle
Don't repeat yourself

How can you build one?

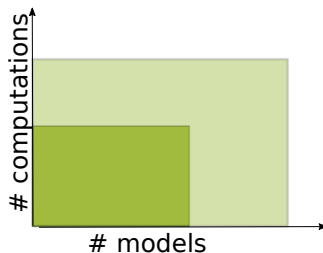
We need:

■ collections

- ▶ of **many models**: (from A to Z):
Arora2005GCB-1 ,CARDAMOM , . . . ,
Zelenev2000MicrobialEcology
- ▶ of **many** computable **properties** and diagnostics to **compare** them by

■ Ways of **organizing** both collections

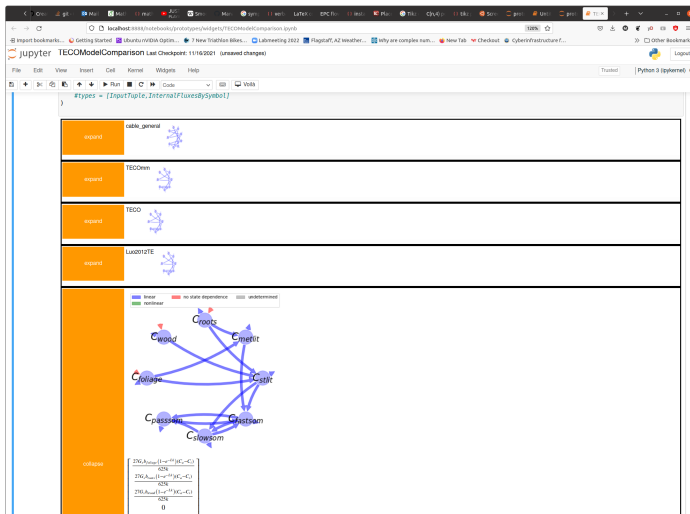
- ▶ building blocks for models
- ▶ functions of building blocks
- ▶ **computational graph** for **different / evolving** building blocks



- 'Copy and paste'?
 - bilinear increase of code
 - unmaintainable, untestable, errorprone, hard to change ...
- better use the **dry** principle
Don't repeat yourself



Example widget for query result





Analysis with symbolic tools (sympy) ...

The screenshot displays a Jupyter Notebook environment with the following content:

- Code Cell:**

```
import numpy, sympy
from inspectModel import *
from bc_model import *
from bc_model import *
from bc_model import *
```
- Output [3]:** A network diagram showing the relationships between carbon pools: C_{wood} , C_{leaf} , C_{root} , C_{DPM} , C_{RPM} , C_{HUM} , and C_{BIO} . Arrows indicate the flow of carbon between these pools.
- Code Cell:**

```
In [28]: mvs.get_CompartmentalMatrix()
```
- Output [28]:** A large matrix representing the compartmental matrix, with rows and columns corresponding to the carbon pools. The matrix is sparse, with many zero entries.
- Code Cell:**

```
In [1]: mvs.get_BibInfo()
## get_CompartmentalMatrix
## get_InfFluxesBySymbol
## get_InputTuple
## get_InternalFluxesBySymbol
## get_OutFluxesBySymbol
## get_SmoothReservoirModel
## get_StateVariableTuple
with nat.TimeCueModel
```

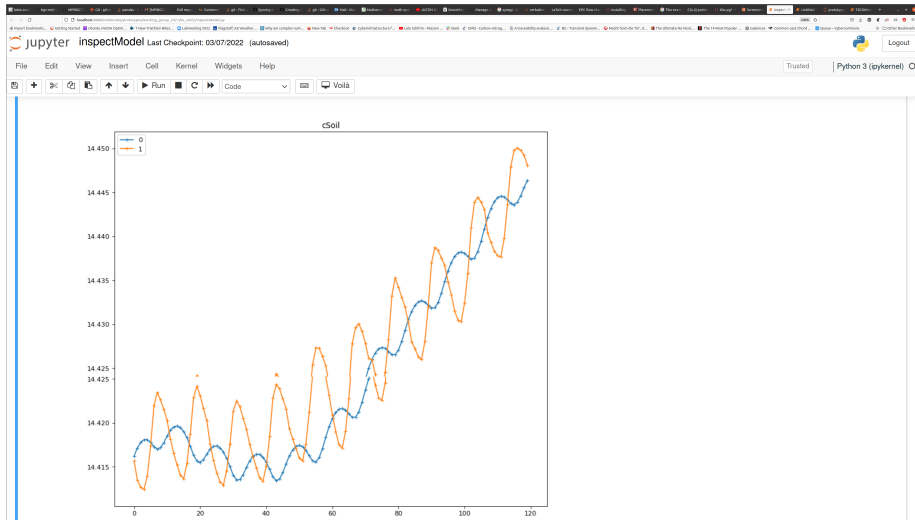
Reproducible Carbon Cycle Models

Biogeochemical Model Database `bgc_md2`



Markus Müller, Holger Metzler, Verónica Ceballos Núñez, Kostiantyn Viatkin, Thomas Lotze, Jon Wells, Yu Zhou, Cuijuan Liao, Aneesh Chandel, Feng Tao, Yuanyuan Huang, Alison Bennett, Chenyu Bian, Lifan Jiang, Song Wang, Chengcheng Gang, Carlos Sierra, Yiqi Luo

...or numerically



Reproducible Carbon Cycle Models

Biogeochemical Model Database bgc_md2

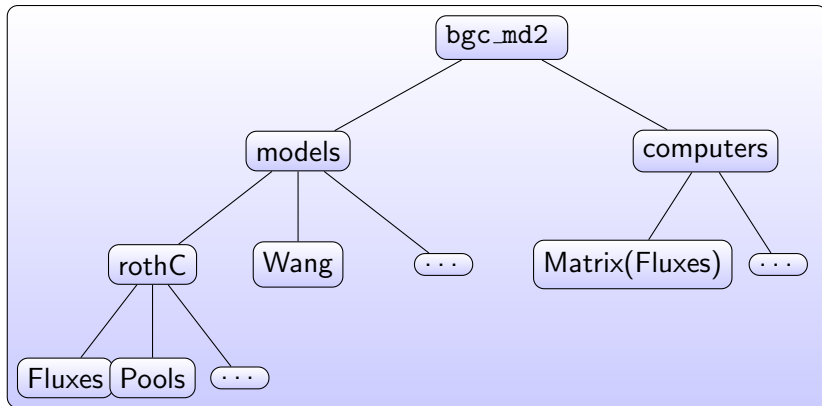


Markus Müller, Holger Metzler, Verónica Ceballos Núñez, Kostiantyn Viatkin, Thomas Lotze, Jon Wells, Yu Zhou, Cuijuan Liao, Aneesh Chandel, Feng Tao, Yuanyuan Huang, Alison Bennett, Chenyu Bian, Lifan Jiang, Song Wang, Chengcheng Gang, Carlos Sierra, Yiqi Luo

Database records are python modules

```
1 from sympy import Symbol, Function
2 from CerebralBiosphereModels import CHNS
3 from bgc_md2.helper import module_computers
4 from bgc_md2.models.BiInfo import BiInfo
5 from bgc_md2.resolve.mvars import (
6     InFluxesBySymbol,
7     OutFluxesBySymbol,
8     InternalFluxesBySymbol,
9     TimeSymbol,
10     StateVariableTable,
11 )
12 import bgc_md2.resolve.computers as bgc_c
13
14 # Make a small dictionary for the variables we will use
15 sym_dict = {
16     'r_leaf_2_w': 'Internal flux rate from leaf to wood',
17     'r_w_2_l': 'Internal flux rate from wood to leaf',
18     'C_soil_fast': 'C',
19     'C_soil_slow': 'C',
20     'C_soil_passive': 'C',
21     'C_leaf': 'C',
22     'C_root': 'C',
23     'C_wood': 'C',
24     'C_leaf_litter': 'C',
25     'C_root_litter': 'C',
26     'C_wood_litter': 'C',
27     'r_C_leaf_2_C_leaf_litter': 'C',
28     'r_C_root_2_C_root_litter': 'C',
29     'r_C_wood_2_C_wood_litter': 'C',
30     'r_C_leaf_litter_rh': 'C',
31     'r_C_root_litter_rh': 'C',
32     'r_C_wood_litter_rh': 'C',
33     'r_C_soil_fast_rh': 'C',
34     'r_C_soil_slow_rh': 'C',
35     'r_C_soil_passive_rh': 'C',
36     'r_C_leaf_litter_2_C_soil_fast': 'C',
37     'r_C_leaf_litter_2_C_soil_slow': 'C',
38     'r_C_leaf_litter_2_C_soil_passive': 'C',
39     'r_C_wood_litter_2_C_soil_fast': 'C',
40     'r_C_wood_litter_2_C_soil_slow': 'C',
41     'r_C_wood_litter_2_C_soil_passive': 'C',
42     'r_C_root_litter_2_C_soil_fast': 'C',
43     'r_C_root_litter_2_C_soil_slow': 'C',
44     'r_C_root_litter_2_C_soil_passive': 'C',
45     'tau': 'Air temperature',
46     'tau_w': 'Air temperature',
47     'tau_l': 'Air temperature',
48     'tau_r': 'Air temperature',
49     'beta_leaf': 'C',
50     'beta_wood': 'C',
51 }
52
53 for k in sym_dict.keys():
54     codekv = sym_dict[k].format(k)
55     exec(code)
56
57 # some we will also use some symbols for functions (which appear with an argument)
58 func_dict = {
59     't': 'A scalar function of temperature and moisture and thereby ultimately of time',
60     'NP': 'C',
61 }
62 for k in func_dict.keys():
63     codekv = Function(k).format(k)
64     exec(code)
65
66 t=TimeSymbol('t')
67 beta_root = 1.0 - (beta_leaf+beta_wood)
68 mvs = CHNS()
69
70 t=
71 StateVariableTable(
72     C_leaf,
73     C_wood,
74     C_soil_fast,
75     C_soil_slow,
76     C_soil_passive,
77     C_leaf_litter,
78     C_root_litter,
79     C_wood_litter,
80     C_soil_fast_rh,
81     C_soil_slow_rh,
82     C_soil_passive_rh,
83 )
84 InFluxesBySymbol(
85     C_leaf: NP(t) * beta_leaf,
86     C_root: NP(t) * beta_root,
87     C_wood: NP(t) * beta_wood,
88 )
89 OutFluxesBySymbol(
90     C_leaf_litter: r_C_leaf_litter_rh * C_leaf_litter * t,
91     C_wood_litter: r_C_wood_litter_rh * C_wood_litter * t,
92     C_root_litter: r_C_root_litter_rh * C_root_litter * t,
93     C_soil_fast: r_C_soil_fast_rh * C_soil_fast * t,
94     C_soil_slow: r_C_soil_slow_rh * C_soil_slow * t,
95     C_soil_passive: r_C_soil_passive_rh * C_soil_passive * t,
96 )
97 InternalFluxesBySymbol(
98     (C_leaf, C_leaf_litter): r_C_leaf_2_C_leaf_litter * C_leaf,
99     (C_wood, C_wood_litter): r_C_wood_2_C_wood_litter * C_wood,
100     (C_root, C_root_litter): r_C_root_2_C_root_litter * C_root,
101     (C_leaf_litter, C_soil_fast): r_C_leaf_litter_2_C_soil_fast * C_leaf_litter * t,
102     (C_leaf_litter, C_soil_slow): r_C_leaf_litter_2_C_soil_slow * C_leaf_litter * t,
103     (C_leaf_litter, C_soil_passive): r_C_leaf_litter_2_C_soil_passive * C_leaf_litter * t,
104     (C_wood_litter, C_soil_fast): r_C_wood_litter_2_C_soil_fast * C_wood_litter * t,
105     (C_wood_litter, C_soil_slow): r_C_wood_litter_2_C_soil_slow * C_wood_litter * t,
106     (C_wood_litter, C_soil_passive): r_C_wood_litter_2_C_soil_passive * C_wood_litter * t,
107     (C_root_litter, C_soil_fast): r_C_root_litter_2_C_soil_fast * C_root_litter * t,
108     (C_root_litter, C_soil_slow): r_C_root_litter_2_C_soil_slow * C_root_litter * t,
109     (C_root_litter, C_soil_passive): r_C_root_litter_2_C_soil_passive * C_root_litter * t,
110     (C_soil_fast, C_soil_slow): r_C_soil_fast_2_C_soil_slow * C_soil_fast * t,
111     (C_soil_slow, C_soil_passive): r_C_soil_slow_2_C_soil_passive * C_soil_slow * t,
112 )
113
114 BiInfo(BiInfo(
115     name='bgc_md2',
116     longname='Biogeochemical Model Database',
117     version='1.0',
118     entryAuthor='Kostiantyn Viatkin',
119     entryAuthorId='1',
120     entryCreatedDate='2021-01-01',
121     doi='10.26434/chemrxiv-2021-01-01',
122 ))
```

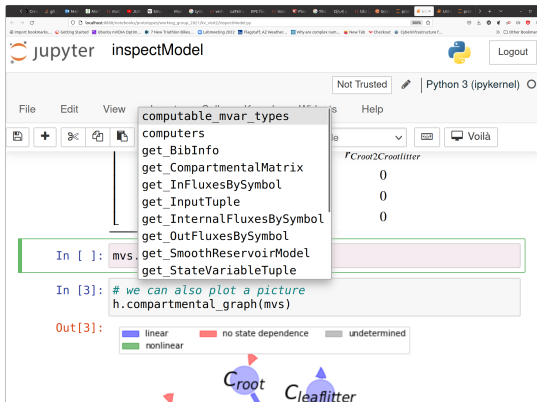
Internal Structure of `bgc_md2`



The `bgc_md` library provides I:

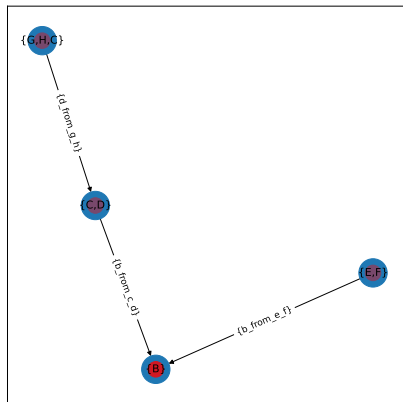
- ① Datatypes defining **building blocks** of models e.g.
`CompartmentalMatrix`, `InternalFluxesBySymbol`, ...
- ② Functions operating on those properties (forming the edges of the graph where the Datatypes are nodes)
- ③ A user interface based on graph algorithms to
 - ① compute the set of computable properties (e.g. the comparable criteria for a set of models, database queries)
 - ② actually compute the desired properties by recursively connecting several function applications.
 - ③ show what is missing to compute a desired property.

Userinterface using computability graphs



Suggested methods automatically created by ComputabilityGraphs

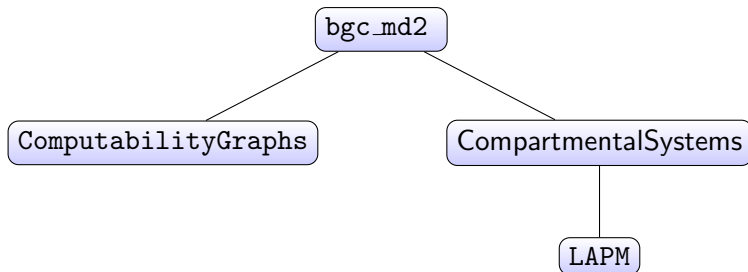
given a set of functions:
a(i), b(c,d), b(e,f),
c(b), d(b), d(g,h),
e(b), f(b) and the target variable **B** e.g. CompartmentalMatrix,
The algorithm computes all possible combinations and paths from which **B** can be computed.



The `bgc_md` library provides 11:

- 1 30+ vegetation, soil or ecosystem models for carbon and nitrogen cycling as reusable python modules using the building blocks in a flexible way.
- 2 An interface to *many algorithms* in CompartmentalSystems to compute diagnostic variables for *many models* in bgc_md2.

Relation to other Python Packages



Example computation via CompartmentalSystems

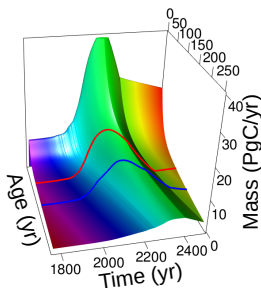


Figure: age distribution of a pool as function of time



Metzler, H., Müller, M., and Sierra, C. (2018).

Transit-time and age distributions for nonlinear time-dependent compartmental systems.

Proceedings of the National Academy of Sciences, 115:201705296.

Links

- The README of the package on github (with installation instructions): https://github.com/MPIBGC-TEE/bgc_md2
- To **explore** some rudimentary tutorials **without installation** use https://mybinder.org/v2/gh/MPIBGC-TEE/bgc_md2/binder
 - ▶ Click on the link!
 - ▶ After jupyter lab has started go to `/binder_notebooks/illustrativeExamples/`
 - ▶ right click on `createModel.py`
 - ▶ choose Open With
 - ▶ choose Jupyter Notebook

This will open an example notebook exploring some of the concepts. More applied examples are coming.