Response to the comments of the referees

Dear Dr. Arndat and referees,

We would like to thank the two referees for their careful and constructive comments for improving our manuscript. In the following, we address all comments and questions from the referees. The comments by the referees are given in blue color, our responses are in black. We have considered all comments for the preparation of a revised version of the manuscript.

Response to comments from Referee #2

General Comments
Fang et al. present an approach which allows the user to automatically and consistently set up large biogeochemical reaction models containing known and novel processes, which can be used in the Community Land Model (CLM). Technical correctness is shown by re-implementing an existing process model (CLM-CN).

With more details on biogeochemical processes becoming available, corresponding models become more complex and larger in size, so that their set-up and expansion becomes more and more difficult. Hence, the presented approach is welcome and surely of interest to the readers of Geoscientific Model Development.

We thank the reviewer for the positive comments.

The presentation and structure of the work, however, needs improvement, as detailed below.
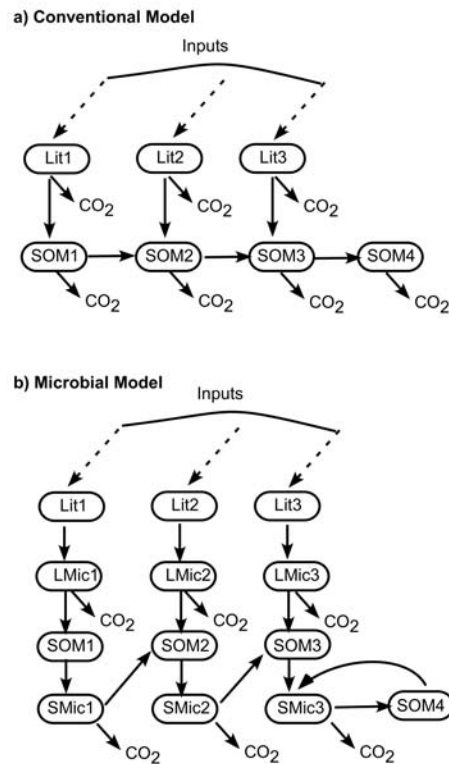
Major Issues
* The promise of a "generic biogeochemical module" is in contrast to the "reaction database consist[ing] of processes of nutrient flow" (P. 3212, l. 17-18). Also, the importance of microbial communities is discussed at some length (P. 3214, l. 13-15), but it is later not shown how their dynamics can be reflected in the proposed approach. Also, it is not obvious whether the module can be linked to other earth system models besides CLM. This would significantly improve it's value.

The generic module presented includes two parts: 1) a reaction database for the users to flexibly extract processes of interest in a particular application to form a reaction network, and 2) using the reaction network approach to derive the ODEs to be solved. The ODEs are solved using an explicit method in this model. The reaction database is not limited to nutrient flow. It is generic and expandable. The users can also create the database from scratch if the processes modeled differ too much from what we presented. Biogeochemical processes modeled in any earth system models can be added in a database. It is up to the user to decide which processes to include in order to model a system. CLM is the land model component of the Community Earth System Model (CESM). It can be linked to the Community Atmosphere Model (CAM) through a flux

coupler. This coupling, however, does not alter the way biogeochemical processes are included in CLM and therefore not relevant to the efforts discussed in this study.

The point we were trying to make by talking about microbial communities is that the level of complexity of a system will increase if microbial biomass and physiology were explicitly considered in the model. For example, in order to use the microbial enzyme model (Wieder et al., 2013), the processes modeled will be different from conventional model as shown in the conceptual diagram below:

a) Conventional Model

Inputs

Lit1   Lit2   Lit3

CO_2 CO_2 CO_2

SOM1 → SOM2 → SOM3 → SOM4

CO_2 CO_2 CO_2 CO_2

b) Microbial Model

Inputs

Lit1   Lit2   Lit3

LMic1   LMic2   LMic3

CO_2 CO_2 CO_2

SOM1   SOM2   SOM3

SMic1   SMic2   SMic3 → SOM4

CO_2 CO_2 CO_2

In the diagram, the boxes represent carbon/nitrogen pools for litter (Lit), microbial biomass (Mic) and soil (SOM). Compared to conventional models, the size of state variables increased by 6, total processes modeled also increased by 6. To simulate with this microbial model in the current code structure, one has to go into the code, manually remove the ODEs for the conventional model and add those for the new model. Using our approach, we only need to add the processes in our reaction database and provide the kinetics for these processes (such as first order kinetics, Michaelis–Menten kinetics) in the user subroutine, our solver will then automatically generate and solve the ODEs once the microbial model (processes tagged with "microbial_model") is selected from the database. The entry in the database for carbon flux from the litter pool 1 to the microbial pool 1 in the model can be written as:
*lit1_c(c) ↔ lmic1_c(c) !litr1_to_lmic microbial_model*

where *(c)* means the reaction is in the column of the model grid, *litr1_to_lmic* is the rate name that will be used in the user subroutine.

The main contribution of this approach is to derive a system of ODEs automatically using matrix decomposition and reaction-based approach. The model reads in the stoichiometries of the reaction network and stores them in Matrix [A] of Eq (2). The creation of the stoichiometries is automated by running a perl script to parse the reaction network. The method presented in the manuscript is not only applicable to CLM, but also to other land surface models. It is a standalone module and can be implemented in other codes. Any terrestrial ecosystem modeler can make use of our approach in their models.

We have added the above discussion in the revised manuscript on page 5 and page 15.

* The general concept remains a little bit vague, and the description of the various components is scattered throughout the manuscript. What exactly compromises the "new biogeochemical module", what is the "generic algorithm"? The description of the PERL script is not clear, either. Please clearly state what it is used for, and describe the different functionalities separately. "dynamic variable stoichiometries" are mentioned here for the first time (P. 3221, l. 18), please describe in detail, what is meant by these, and how they fit into the previous description of the module. Also, the method for numerically solving the system is not clearly described. A conceptual figure could be helpful to make the operational structure of the approach clear, and also indicating the relationship between database, PERL script, and model, including the inputs and outputs of the different steps in setting up a reaction network model.

We thank the reviewer for this great suggestion. A conceptual diagram shown below is added in the revised manuscript and steps were summarized before detailed descriptions of each step to illustrate the procedure. We added the description of CLM model and how the ODEs for biogeochemistry are solved in the model in the introduction section on page 6 in the revised manuscript.
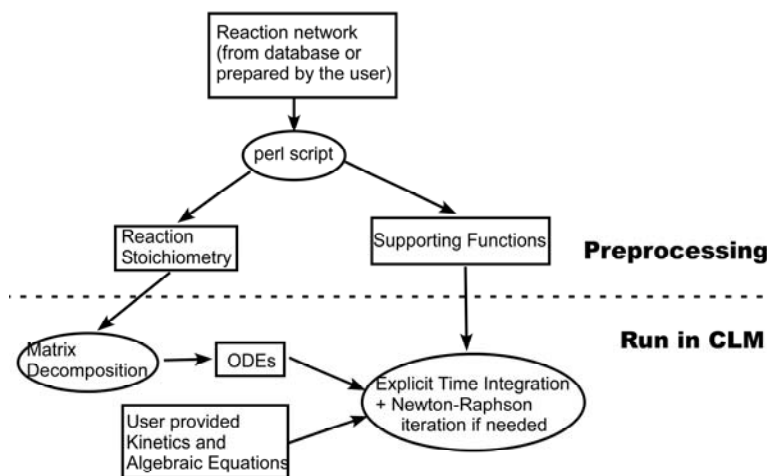
In the CLM model, the N:C ratios in the soil pools are fixed, while the N:C ratios for the pools in litters and coarse woody debris are adjusted depending on the decomposition

rates and nutrient availability. New ratios are calculated from the carbon and litter pools from the previous time step. The adjusted N:C ratios are what we called "dynamic variable stoichiometries". We clarified the terminology in page 17.

* Three current major challenges in earth system models are correctly identified in the abstract, however, how the approach helps on the second (computational cost) is not obvious. Are fast and slow reactions automatically separated? How are algebraic equations defined in the approach? The third challenge (different mathematical representations for different processes) sounds a little bit vague, and should be sharpened.

The approach presented in this manuscript will facilitate new model spin-up method to obtain the steady state of the system by setting the time derivatives of the state variables to zero, which gives a set of algebraic equations. For example, we can apply the Newton-Raphson method to the algebraic equations to solve the steady state solution. This is especially useful when the decomposition of carbon pools is represented by nonlinear kinetics.
As for the third challenge, different processes and kinetics can be easily evaluated by changing the reaction networks and user defined subroutines for kinetics.

We added the discussion in pages 17, 23-24.

* The stated challenges are not new, and a discussion and comparison to other approaches tackling these issues is missing (e.g., Aguilera et al., Geochemistry, Geophysics, Geosystems, 6 (7), and others).

The concepts described in Aguilera et al. (2005), Chilakapati et al. (2000) are similar to our approach. However, those models depend on the third-party commercial software MAPLE to do matrix operation given user's stoichiometric input in MAPLE language. We added these references and discussion in the revised manuscript.

* For readers not familiar with CLM, a short description of this model, including purpose, considered processes and numerical implementation would be very helpful.

Thank you for the suggestion. We have added a short description of the model in page 6 in the revised manuscript.

* While the stoichiometries in Equation (2) are automatically retrieved, how are the rates R defined? Can they be set by the user? It is mentioned that they might depend on moisture, temperature, etc., but details on how this can be implemented are missing. Are arbitrary mathematical expressions possible? See also next point.

The rates are defined and written by the user in the user subroutine for kinetics. It can be arbitrary mathematical expressions fitted from an experiment. We discussed this on page 15 in the revised manuscript.

* P. 3222, l. 1,2: How exactly are rate expressions incorporated in the module? This is an important point for modelling. Can they have arbitrary form?

The user has to provide the expression and write them in the user subroutine. This subroutine is called during the simulation. The expressions can be in any arbitrary form.

* P. 3216, l. 17 - P. 3217, l. 2 is introductorial / motivational material and should be moved to the introduction.

This paragraph is moved to the introduction as suggested by the reviewer.

* P. 3219, l. 4-14, l 24-27: is introductorial material and should be moved to the introduction.
We moved 1.4-14 to the introduction. We think 1 24-27 is relevant in the context.

* It seems that processes are derived from the source code of CLM-CN (P. 3217, l. 19-20). Note that this is an unusual step; for a well-documented model, this information should be unambigously retrievable from the documentation of the code.

The reviewer is right. Unfortunately, this information is not well documented.

Minor Issues / Language
P. 3212, l. 22: What is "CLM-CN"?
CLM-CN means the biogeochemistry module in CLM. We added the definition on page 6.

P. 2312, l. 4: "CLM" -> "CLMs"
Corrected.

P. 3212, l. 18: "litter" -> "litter,"
Corrected.

P. 3213, l. 9: "energy" -> "energy,"
Corrected.

P. 3213, l. 10: "e.g." -> "e.g.,"
Corrected.

P. 3213, l. 16: "contribution" -> "contributions"
Corrected.

P. 3213, l. 17: "in literature" -> "in the literature"
Corrected.

P. 3213, l. 24: "is not in" -> "is not included in"
Corrected.

P. 3214, l. 2: remove comma
Removed.

P. 3214, l. 5-6: remove "(" and ")"
Removed.

P. 3214, l. 15: "of the microbial" -> "of microbial"
Corrected.

P. 3215, l. 17: "in which," -> "in which"
Corrected.

P. 3216, Equation 2: remove "[I]"
The unit matrix cannot be removed because it is needed in the matrix decomposition. An example is shown in the revised manuscript on pages 9-11.

P. 3216, l. 11-15: avoid the same reference in consecutive sentences
Removed.

P. 3218, l. 1-2: Make clear that only an example is given.
We added a sentence to point out only a few ODEs are presented, the rest are given in the supplementary material.

P. 3222, l. 13: "refer our" -> "refer to our"
Corrected.

P. 3223, l. 5: "land model" -> "land models"
Corrected.

P. 3223, l. 8: "pools follow" -> "pools to follow"
The sentence was rewritten as "the allocation of P to live aboveground and belowground and the flow of P in plant litter and different organic soil pools are assumed to follow the allocation of N in CLM-CN".

P. 3223, l. 9: "N and P" -> "N, and P"
Corrected.

Unclear sentences or unusual wording:
P. 3213, l. 20: "raises a difficulty"
The sentence was reworded as "Not knowing the detailed processes simulated in a model makes it difficult to update or add new processes into the module".

P. 2315, l. 9-10: "P included model", "P effect"
The sentence was rewritten as "The model that includes the phosphorus dynamics can be used to evaluate the effect of phosphorus on productivity of terrestrial ecosystems." .

P. 3217, l. 22: "reaction-based decomposition cascade processes"
The sentence was rewritten as "we came up with the following decomposition cascade processes with balanced elements in Table 2 using reaction-based approach.".

P. 3217, l. 24: "inverse of C:N ratio" -> "N:C ratios"
Corrected.

P. 3220, l. 2: What is meant by "column averaged type by (p) and (c)"?
We rewrote the sentence as "For each process in the database, state variables are separated into plant functional type (p) or column averaged type (c)".

P. 3221, l. 11: "a tool written with the PERL script"
We rewrote it as "a script written using the PERL language".

P. 3221, l. 15: "that point the reaction names"
We changed it to "pointers that link the reaction names in the database to the defined rate expressions that are used in the code"

P. 3222, l. 22: "The database was used to pick the reactions"
The sentence was rewritten as "We selected all the reactions except for R_4, R_5, R_6 in Table 2, and all the reactions in Table 3 from the database."

P. 3223, l. 6-7: "assume the allocation of P to live aboveground and belowground"
The sentence was rewritten as "the allocation of P to live aboveground and belowground and the flow of P in plant litter and different organic soil pools are assumed to follow the allocation of N in CLM-CN".

P. 3223, l. 16: "inverse of the C:P ratio" -> "P:C ratios"
Corrected.

P. 3225, l. 5: "were inputted"
Rewritten as "written in a user defined subroutine of kinetics".

P. 3225, l. 20-21 "two years of simulation time" -> "two years of simulated time"
Corrected.

References:
Wieder, W.R., Bonan, G.B. and Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. doi:10.1038/NCLIMATE1951. Nature Climate Change.
Aguilera, D.R., Jourabchi, P., Spiteri, C. and Regnier, P., 2005. A knowledge-based reactive transport approach for the simulation of biogeochemical dynamics in Earth systems. Geochemistry Geophysics Geosystems, 6.

Chilakapati, A., Yabusaki, S., Szecsody, J. and MacEvoy, W., 2000. Groundwater flow, multicomponent transport and biogeochemistry: development and application of a coupled process model. Journal of Contaminant Hydrology, 43(3-4): 303-325.