

Manuscript prepared for Geosci. Model Dev.
with version 2014/09/16 7.15 Copernicus papers of the L^AT_EX class copernicus.cls.
Date: 11 June 2015

Addressing the Challenge of Nonnegativity in Coupling a Reactive Transport Code with a Global Land Surface Model for Mechanistic Biogeochemistry Representation

Guoping Tang¹, Fengming Yuan¹, Gautam Bisht^{1,2}, Glenn E. Hammond³, Peter C. Lichtner⁴, Nathaniel O. Collier¹, Jitendra Kumar¹, Richard T. Mills^{1,5}, Xiaofeng Xu^{1,6}, Ben Andre⁷, Forrest M. Hoffman¹, Scott L. Painter¹, and Peter E. Thornton¹

¹Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States

²Lawrence Berkeley National Laboratory, Berkeley, California, United States

³Sandia National Laboratories, Albuquerque, New Mexico, United States

⁴OFM Research, Redmond, Washington, United States

⁵Intel Incorporation, Portland, Oregon, United States

⁶University of Texas at El Paso, El Paso, Texas, United States

⁷National Center for Atmospheric Research, Boulder, Colorado, United States

Correspondence to: NAME (EMAIL)

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan(<http://energy.gov/downloads/doe-public-access-plan>).

Abstract. Reactive transport codes (e.g., PFLOTRAN) are increasingly used to improve the representation of biogeochemical processes in terrestrial ecosystem models (e.g., the Community Land Model, CLM). We have developed coupled CLM–PFLOTRAN modeling framework to utilize the reactive transport code PFLOTRAN for flexible and improved representation of biogeochemical processes in CLM. However, as CLM and PFLOTRAN use explicit and implicit time stepping, implementation of CLM biogeochemical reactions in PFLOTRAN can result in negative concentration, which is not physical and can cause numerical instability and errors. The objective of this work is to implement CLM subsurface biogeochemical reactions in CLM-PFLOTRAN with a focus on addressing the nonnegativity challenge to obtain accurate, efficient, and robust solutions. We implement the CLM-CN decomposition, nitrification, denitrification, and plant nitrogen uptake reactions and test the implementation at arctic, temperate, and tropical sites. We examine use of scaling back the update during each iteration (SU), log transformation (LT), and downregulating the reaction rate to account for reactant availability limitation to enforce nonnegativity. Both SU and LT guarantee nonnegativity but with implications. When a very small scaling factor occurs due to either consumption or numerical overshoot, and the iterations are deemed converged because of too small an update, SU can introduce excessive numerical error. LT involves multiplication of the Jacobian matrix by the concentration vector, which increases the condition number, decreases the time step size, and increases the computational cost. Neither SU nor SE prevents zero concentration. When the concentration is close to machine precision or 0, a small positive update stops all reactions for SU, and LT can fail due to a singular Jacobian matrix. The consumption rate has to be downregulated such that the solution to the mathematical representation is positive. A first-order rate downregulates consumption and is nonnegative, and adding a residual concentration makes it positive. For zero-order rate or when the reaction rate is not a function of a reactant, representing the availability limitation of each reactant with a Monod substrate limiting function provides a smooth transition between a zero-order rate when the reactant is abundant and first-order rate when the reactant becomes limiting. Marching through the transition may require small time step sizes when the half saturation is small. Our results from simple tests and CLM-PFLOTRAN simulations caution against use of SU and indicate that accurate, stable, and relatively efficient solutions can be achieved with LT and downregulation with Monod substrate limiting function and residual concentration.

1 Introduction

Methods developed for the reactive transport models (??) and geochemical codes (??) are increasingly used to improve the representation of biogeochemical processes in terrestrial ecosystem models (TEMs, e.g., the Community Land Model, CLM) for better climate prediction. An essential aspect of these TEMs is the ability to simulate competition for nutrients (e.g., mineral nitrogen, phosphate, etc.) among plants and microbes. For example, the CLM-CN decomposition cascade downregulates the demand based on the available nitrogen (N) (??). Specifically, marching from time step k to $k+1$ with a supply rate S^k and consumption rate D^k using the forward difference, $[N]^{k+1} = [N]^k + (S^k - D^k)\Delta t$. ($[]$ is used to denote concentration.) CLM replaces D^k with $\min(D^k, [N]^k/\Delta t)$ so that $[N]^{k+1} \geq S^k\Delta t \geq 0$. As a result, $[N]$ is nonnegative in CLM.

Geochemical codes generally use implicit time stepping such as the backward difference, $([N]^{k+1} - [N]^k)/\Delta t = S^{k+1} - D^{k+1}$. Solving this nonlinear equation with the Newton-Raphson method to iterate from p to $p+1$,

$$50 \quad [N]^{k+1,p+1} = [N]^{k+1,p} - \frac{\frac{[N]^{k+1,p} - [N]^k}{\Delta t} - S^{k+1,p} + D^{k+1,p}}{\frac{1}{\Delta t} - \frac{\partial S^{k+1,p}}{\partial [N]^{k+1,p}} + \frac{\partial D^{k+1,p}}{\partial [N]^{k+1,p}}} \quad (1)$$

Depending on $[N]^k$, Δt , $S^{k+1,p}$, $D^{k+1,p}$, and the derivatives, $[N]^{k+1,p+1}$ can be negative, which is not physical, and can cause numerical instability and errors (?).

Enforcing nonnegativity is a common challenge in science, engineering, and business, including, for example, image processing, optimization, and ecosystem and geochemical modeling (?????????).
55 The challenge increases in geochemical modeling because the concentration can be very low. For example, the threshold concentration for H₂ is 1.5 nM (10^{-9} M) for dechlorinators and 5~20 nM for methanogens (?). The EPA maximum contaminant level for drinking water for dioxin is 30 ppq (3.5×10^{-13} M). The redox potential Eh needs to be decreased to -0.35 V (corresponding to an O₂ concentration $< 10^{-22}$ M ?, ?) for methanogens to grow (?). With very low concentration, a small
60 consumption or numerical oscillation can lead to negative concentration.

Two methods are used to avoid negative concentration in geochemical codes. One is to use the logarithm concentration as the primary variable (??). Log transformation (LT) is well suited for solving the mass action equations because it converts these nonlinear equations into linear equations. However, LT converts linear advection and diffusion equations into nonlinear equations, which
65 may increase the computational cost (?). The other approach (SU) scales back the update in each of the Newton-Raphson iteration to enforce nonnegativity (??). Both methods are available in PFLOTRAN and some other geochemical codes (e.g., Geochemist's Workbench, ?, ?). However, to our knowledge, the implications of these approaches have not been thoroughly examined, particularly for application to CLM.

70 Competition for nutrients (for instance, labile carbon, phosphate, O₂, and H₂) is common in terrestrial ecosystems, and is increasingly incorporated in process-rich models. As the terrestrial

ecosystem models are often run under a variety of conditions around the globe for hundreds of years with a time step as small as half an hour, proper treatment of the nonnegativity issue is necessary for accurate, efficient, and robust simulation of biogeochemical processes using reactive transport codes for earth system models. The objective of this work is to implement CLM subsurface biogeochemical reactions in CLM-PFLOTRAN with the focus on addressing the nonnegativity challenge to obtain accurate, efficient, and robust solutions. We implement the CLM-CN decomposition (??); nitrification; denitrification (??); and plant nitrogen uptake reactions in CLM-PFLOTRAN and test the implementation at arctic, temperate, and tropical sites. In addition to SU and LT that are available in PFLOTRAN, we examine ways to downregulate consumption to account for the limitation of reactant availability on reaction rate. While this work focuses on addressing the nonnegativity challenge for biogeochemistry, detailed comparison of the results of CLM and CLM-PFLOTRAN biogeochemistry representation and comprehensive CLM-PFLOTRAN coupling in heat transfer (including freeze and thaw), hydrology, and biogeochemistry will be presented in future publications.

While we use CLM-PFLOTRAN to implement and test simple carbon and nitrogen reactions at a few sites, we hope that what we develop here will be relevant for a wide range of applications.

2 Biogeochemistry

2.1 CLM biogeochemistry

The terrestrial ecosystem models generally include biogeochemical reactions for carbon and nitrogen cycles, in particular, the organic matter decomposition, nitrification, dentrification and methane production and oxidation. The kinetics are usually described by a first-order rate modified by response functions for environmental variables (temperature, moisture, pH, etc.) (??). In this work, we use the CLM-CN decomposition (??), nitrification, denitrification (??), and plant nitrogen uptake reactions (Fig. 1) as an example.

2.1.1 CLM-CN decomposition

The CLM-CN decomposition cascade consists of three litter pools with variable CN ratios, four soil organic matter (SOM) pools with constant CN ratios, and seven reactions (Fig. 1A). The reaction can be described by

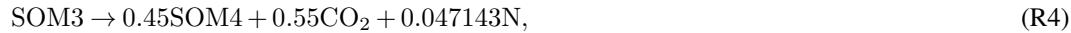
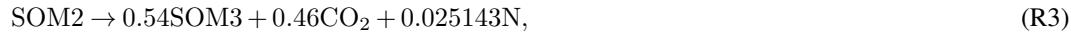


with CN_u and CN_d as the upstream and downstream pool (molecular formula), N as either NH_4^+ or NO_3^- , f as the respiration fraction, and $n = u - (1 - f)d$. The rate is

$$\frac{\partial[\text{CN}_u]}{\partial t} = -k_d f_T f_w [\text{CN}_u], \quad (2)$$

with k_d as the rate coefficient and f_T and f_w as the temperature and moisture response functions.

With a constant CN ratio, the decomposition reactions for the four SOM pools are

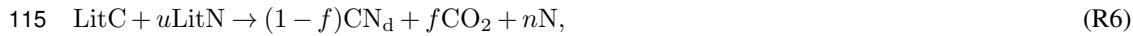


110 and

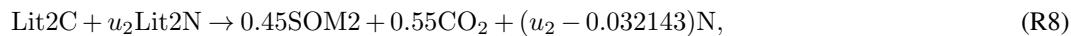
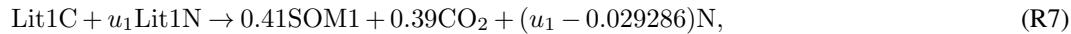


CLM4.5 has an option to separate N into NH_4^+ and NO_3^- . The N mineralization product is NH_4^+ .

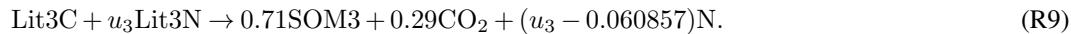
As the CN ratio is variable for the three litter pools, litter N pools need to be tracked such that reaction (R1) becomes



with $u = [\text{LitN}]/[\text{LitC}]$. The three litter decomposition reactions are



120 and



As the CN ratio of the litter pools is generally high, u_1 , u_2 , and u_3 are usually small, and n in these reactions (e.g., $n_1 = u_1 - 0.029286$ for Lit1) is normally negative. Namely, these reactions consume (immobilize) N, which can be NH_4^+ , NO_3^- , or both.

125 **2.1.2 Nitrification**

The nitrification reaction to produce NO_3^- is



with rate (Dickinson et al., 2002)

$$\frac{\partial[\text{NH}_4^+]}{\partial t} = -\frac{\partial[\text{NO}_3^-]}{\partial t} = -k_n f_T f_w [\text{NH}_4^+]. \quad (3)$$

130 The nitrification reaction to produce N₂O is



with one component related to decomposition as

$$\frac{\partial[\text{NH}_4^+]}{\partial t} = -2 \frac{\partial[\text{N}_2\text{O}]}{\partial t} = -f_{\text{nm}} f_T f_w f_{\text{pH}} \max(R_{\text{nm}}, 0) \quad (4)$$

with f_{nm} as a fraction (?) and R_{nm} as the net N mineralization rate,

$$135 \quad R_{\text{nm}} = \sum_i n_i R_i, \quad (5)$$

where R_i denotes the rate of reaction (R2, R3, R4, R5, R7, R8, R9). The second component is (?)

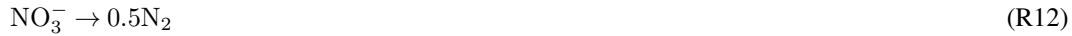
$$\frac{\partial[\text{NH}_4^+]}{\partial t} = -2 \frac{\partial[\text{N}_2\text{O}]}{\partial t} = -k_{\text{n2o}} f_T f_w f_{\text{pH}} (1 - e^{-0.0105[\text{NH}_4^+]}) . \quad (6)$$

Ignoring the high-order terms and moving the unit conversion factor into k_{n2o} , it can be simplified as a first-order rate as

$$140 \quad \frac{\partial[\text{NH}_4^+]}{\partial t} = -2 \frac{\partial[\text{N}_2\text{O}]}{\partial t} = -k_{\text{n2o}} f_T f_w f_{\text{pH}} [\text{NH}_4^+]. \quad (7)$$

2.1.3 Denitrification

The denitrification reaction is



with rate (?)

$$145 \quad \frac{\partial[\text{NO}_3^-]}{\partial t} = -2 \frac{\partial[\text{N}_2]}{\partial t} = -k_{\text{deni}} f_T f_w f_{\text{pH}} [\text{NO}_3^-]. \quad (8)$$

2.1.4 Plant nitrogen uptake

The plant nitrogen uptake reaction can be written as



and

$$150 \quad \text{NO}_3^- \rightarrow \text{PlantN}. \quad (\text{R14})$$

The rate is specified by CLM (plant nitrogen demand) and assumed to be constant in each half-hour time step.

2.1.5 Demand-based competition and distributing nitrogen demand between NH_4^+ and NO_3^-

Denote $R_{d,p}$, $R_{d,i}$, $R_{d,nitr}$, $R_{d,densi}$ as the potential plant, immobilization, nitrification, and denitrification demand (rate); $R_{a,tot} = R_{d,p} + R_{d,i} + R_{d,nitr}$ as the total NH_4^+ demand; and $R_{n,tot}$ as the total NO_3^- demand. CLM uses a demand-based competition approach to split the available sources in proportion to the demand rates to meet the demands (??). Specifically, for each time step, if $R_{a,tot}\Delta t \leq [\text{NH}_4^+]$, the uptakes are equal to potential demands, and $R_{n,tot} = 0$; otherwise, the uptakes for NH_4^+ are $[\text{NH}_4^+]R_{d,p}/R_{a,tot}\Delta t$, $[\text{NH}_4^+]R_{d,i}/R_{a,tot}\Delta t$, and $[\text{NH}_4^+]R_{d,nitr}/R_{a,tot}\Delta t$ for plants, immobilization, and nitrification, respectively; $R_{n,tot} = R_{a,tot} - [\text{NH}_4^+]/\Delta t + R_{d,densi}$. If $R_{n,tot}\Delta t < [\text{NO}_3^-]$, all of the remaining demand $R_{n,tot}$ is met with available NO_3^- . Otherwise, available NO_3^- is split to meet the remaining plant, immobilization, and denitrification demands in proportion to their rates.

2.2 CLM-PFLOTRAN biogeochemistry

In CLM-PFLOTRAN, CLM instructs PFLOTRAN to solve the partial differential equations for energy (including freezing and thawing), water flow, and reaction and transport in the surface and subsurface. This work focuses on the biogeochemistry (energy, hydrology and overall coupling will be described in future publications). Specifically, we focus on addressing the nonnegativity challenge for the geochemical reactions, with CLM solving the energy and water flow equations and handling the solute transport (mixing, advection, diffusion, and leaching).

In each time step, CLM provides production rates for Lit1C, Lit1N, Lit2C, Lit2N, Lit3C, Lit3N for litter fall; NH_4^+ and NO_3^- for nitrogen deposition and fixation; and plant N demand (rate); and specifies liquid water content, matrix potential and temperature for PFLOTRAN; PFLOTRAN solves the ordinary differential equations for the kinetic reactions and provides the final concentrations back to CLM.

PFLOTRAN does not track individual reaction rates such as total nitrogen mineralization rate. We add hypothetical species (diagnostic variables), e.g., PlantN, N2Od, and DeniN, to track NO_3^- uptake by plants as in reaction (R14); N_2O production from nitrification reaction (R11) due to net mineralization (Eq. 4); and denitrification (R12). At the end of each time step, CLM uses the change of these concentration to calculate the specific rates. To simplify the calculations, we reset these concentrations to 10^{-10} at the beginning of each time step instead of storing the values of the previous time step.

The reactions and rates such as those described in Section 2.1 are implemented using the “reaction sandbox” concept in PFLOTRAN (?). For each reaction, we specify a rate and a derivative of the rate with respect to any components in the rate formula, given concentrations, temperature, moisture content, grid cell volume, and other environmental variables. PFLOTRAN accumulates these rates and derivatives into a residual vector and a Jacobian matrix, and the global equation is discretized in

time using the backward difference and solved using the Newton-Raphson method. Ignoring equilibrium reactions and transport for simplicity of discussion in this work, PFLOTRAN solves the
190 ordinary differential equation,

$$\frac{\partial \mathbf{c}}{\partial t} = \mathbf{R}(\mathbf{c}), \quad (9)$$

with \mathbf{c} as the concentration vector and \mathbf{R} as the kinetic reaction rate. Discretizing Eq. (9) in time using the backward difference,

$$\frac{\mathbf{c}^{k+1} - \mathbf{c}^k}{\Delta t} = \mathbf{R}(\mathbf{c}^{k+1}). \quad (10)$$

195 Solving the equation using the Newton-Raphson method,

$$\mathbf{f}(\mathbf{c}^{k+1,p}) = (\mathbf{c}^{k+1,p} - \mathbf{c}^k)/\Delta t - \mathbf{R}(\mathbf{c}^{k+1,p}), \quad (11)$$

$$\mathbf{J} = \frac{\partial \mathbf{f}(\mathbf{c}^{k+1,p})}{\partial \mathbf{c}^{k+1,p}}, \quad (12)$$

$$200 \quad \delta \mathbf{c}^{k+1,p} = \mathbf{J}^{-1} \mathbf{f}(\mathbf{c}^{k+1,p}), \quad (13)$$

and

$$\mathbf{c}^{k+1,p+1} = \mathbf{c}^{k+1,p} - \delta \mathbf{c}^{k+1,p}. \quad (14)$$

The iteration continues until either the residual $\mathbf{f}(\mathbf{c}^{k+1,p})$ or the update $\delta \mathbf{c}^{k+1,p}$ is less than a specified tolerance. Specifically,

$$205 \quad \|\mathbf{f}(\mathbf{c}^{k+1,p})\|_2 < \text{ATOL}, \quad (15)$$

$$\frac{\|\mathbf{f}(\mathbf{c}^{k+1,p})\|_2}{\|\mathbf{f}(\mathbf{c}^{k+1,0})\|_2} < \text{RTOL}, \quad (16)$$

$$\frac{\|\delta \mathbf{c}^{k+1,p}\|_2}{\|\mathbf{c}^{k+1,p}\|_2} < \text{STOL}, \quad (17)$$

210

$$\|\mathbf{f}(\mathbf{c}^{k+1,p})\|_\infty < \text{ITOL_RES}, \quad (18)$$

and

$$\|\delta \mathbf{c}^{k+1,p}\|_\infty < \text{ITOL_UPDATE}. \quad (19)$$

If none of these tolerances are met in MAXIT iterations or MAXF function evaluations, the iteration is considered to diverge, and PFLOTRAN decreases the time step size for MAX_CUT times. The default values in PFLOTRAN are ATOL = 10^{-50} , RTOL = 10^{-8} , STOL = 10^{-8} , ITOL_RES = 10^{-50} , ITOL_UPDATE = 10^{-50} , MAXIT = 50, MAXF = 10^4 , and MAX_CUT = 16.

Unlike the explicit time stepping in CLM for biogeochemistry where only the reaction rates need to be calculated, the implicit time stepping used in PFLOTRAN and other geochemical codes requires the derivatives. While PFLOTRAN provides the option to calculate the Jacobian numerically, geochemical codes generally prefer analytical Jacobian calculation (e.g., ?) because numerical calculation for accurate Jacobian approximation is a notoriously difficult task (?). Many reactions can be specified, and the rates and derivatives are accumulated in the residual and Jacobian, providing flexibility in specifying various reactions with a user-defined rate formula. In contrast, the number of pools and reactions were traditionally hard-coded in CLM. Therefore, this new approach facilitates implementation of increasingly mechanistic reactions and tests of various representations without code modifications.

3 Approaches

Like Eq. (1), one of the basic challenges for using geochemical code for CLM is that Eq. (14) is not guaranteed to be positive. Both SU (scaling back the update during each iteration) and LT (log transformation) are available in PFLOTRAN to enforce nonnegativity. However, to our knowledge, the limitations and implications of both methods have not been thoroughly examined.

3.1 Scaling back update in iterations

SU scales back the update (Eq. 14) with a scaling factor λ (??) such that

$$235 \quad \mathbf{c}^{k+1,p+1} = \mathbf{c}^{k+1,p} - \lambda \delta \mathbf{c}^{k+1,p} > 0, \quad (20)$$

where

$$\lambda = \alpha \min \left[1, \frac{\mathbf{c}^{k+1,p}(i)}{\delta \mathbf{c}^{k+1,p}(i)} \right] \quad (21)$$

for positive $\delta \mathbf{c}^{k+1,p}(i)$ and $i = 1$ to m , with m as the number of species times the number of numerical grid cells. With SU, the concentration of the species that is going negative decreases by $(1 - \alpha)$ times in each iteration instead. This can be shown by solving the zero-order uptake problem $dc/dt = -1$: $c^{k+1} = c^k - \Delta t$ when $\Delta t \geq c^k$; otherwise, $\lambda = \alpha c^k / \Delta t$, and $c^{k+1} = (1 - \alpha)c^k$. With a default $\alpha = 0.99$ in PFLOTRAN, each iteration reduces the concentration by 100 times. With multiple iterations, the concentration can approach machine precision or 0.

The choice of α and the value of λ are expected to influence the convergence process and may introduce numerical error under certain conditions. For example, if $\mathbf{c}^{k+1,p}(i)$ is 0 (or below the

machine precision), and $\delta\mathbf{c}^{k+1,p}(i) > 0$, then $\lambda = 0$, and the update is limited to 0 and the residual will not be decreased to satisfy Eqs. (15 or 16) to achieve convergence. If the iteration is deemed converged because the scaled update is 0, and Eq. (17) is met, the calculation will march through this time step without any change in the concentrations, numerically stopping all reactions. This can
250 continue for many time steps until $\delta\mathbf{c}^{k+1,p}(i)$ becomes negative.

Application of a very small scaling factor λ due to decrease of a small concentration for one species may have a similar consequence: the iteration may not decrease the residual $\mathbf{f}(\mathbf{c}^{k+1,p})$ or the iteration may not converge because Eqs. (15 or 16) are not met. If the iteration is considered converged because of too small an update $\lambda\delta\mathbf{c}^{k+1,p}$ (Eq. 17 is satisfied), the scaling factor λ (e.g.,
255 10^{-10}) may correctly limit the consumption reaction rates to account for availability limitation, but wrongly limit the production rate in the time step. This can be illustrated by adding to the zero-order uptake problem $dc/dt = -1$ an independent zero order production $de/dt = 1$: when $\Delta t \geq c^k$, $\lambda = \alpha c^k / \Delta t$, $c^{k+1} = (1 - \alpha)c^k$, $e^{k+1} = e^k + \alpha c^k$ rather than $e^{k+1} = e^k + \Delta t$. If $c^k = 0$, SU numerically stops the independent production reaction.

260 To avoid excessive numerical error, PFLOTRAN reports an error and exits when $\lambda < \lambda_{\min}$, with a default $\lambda_{\min} = 10^{-10}$. This translates the accuracy issue into a stability issue. It is necessary to investigate SU to resolve both the accuracy issue and the stability issue.

3.2 Log transformation

LT is widely used in geochemical codes to enforce positivity (???). Instead of solving Eq. (11) for
265 \mathbf{c}^{k+1} using Eqs. (12,13,14), LT solves for $(\ln \mathbf{c}^{k+1})$ from

$$\mathbf{J}_{\ln}(i,j) = \frac{\partial \mathbf{f}(i)}{\partial \ln(\mathbf{c}(j))} = \mathbf{c}(j) \frac{\partial \mathbf{f}(i)}{\partial \mathbf{c}(j)} = \mathbf{J}(i,j), \quad (22)$$

$$\delta \ln \mathbf{c}^{k+1,p} = \mathbf{J}_{\ln}^{-1} \mathbf{f}(\mathbf{c}^{k+1,p}), \quad (23)$$

and

$$270 \quad \mathbf{c}^{k+1,p+1} = \mathbf{c}^{k+1,p} \exp[-\delta \ln(\mathbf{c}^{k+1,p})]. \quad (24)$$

One downside is apparent from Eq. (22): suppose $\mathbf{c}^{k+1,p}(j) = 10^{-10}$; $\mathbf{J}(i,j)$ (column j of the Jacobian) decreases by 10^{10} times to $\mathbf{J}_{\ln}(i,j)$. As a result, the condition number of the Jacobian matrix may increase by orders of magnitudes, which may constrain the time step size to be substantially reduced, comparing to the case without LT. Secondly, the small eigenvalues may result in large
275 $\delta \ln \mathbf{c}^{k+1,p}$, which may cause overshooting to an unrealistically large concentration (positive update), or essentially 0 concentration (negative update), or even overflow in the exponential function in Eq. (24). To prevent these problems, PFLOTRAN limits the update to

$$\delta \ln(\mathbf{c}^{k+1,p}) = \text{sign}[\delta \ln(\mathbf{c}^{k+1,p})] \min[\text{abs}(\delta \ln \mathbf{c}^{k+1,p}), \delta_{\ln, \max}] \quad (25)$$

with a default $\delta_{\ln,\max} = 5$. Thirdly, LT does not prevent 0 concentration per se: a number of iterations with positive $\delta \ln \mathbf{c}^{k+1,p}(j)$ update is likely to bring $\mathbf{c}^{k+1,p+1}(j)$ to be below machine precision or 0. With zero concentration, \mathbf{J}_{\ln} becomes singular, and the numerical solution fails. These can be shown by solving $dc/dt = -1$ in the log transformed form $d \ln c/dt = -1/c$: $c^{k+1} = c^k e^{-\Delta t/c^{k+1}}$. LT converts the linear problem that does not require iteration to a nonlinear problem that needs to be solved iteratively. As c^{k+1} becomes small, Δt has to be decreased accordingly, or the concentration goes to below machine precision and a $c^{k+1} = 0$ causes division by 0 overflow.

3.3 Downregulation of reaction rate

Even though ensuring nonnegativity, the concentrations can become very small or essentially 0 using either SU or LT. With 0 or near 0 concentrations, the scaling factor for SU is small or 0, or the time step sizes for LT are too small if LT does not fail due to a singular Jacobian matrix. Both approaches require the solution to the mathematical representation itself be positive. To obtain such a mathematical representation, it is necessary to downregulate reaction rates to represent the limitation of the availability of each reactant on the reaction rate. In the geochemical modeling literature, this is mostly represented by a rate limiting function as a function of concentration in each reaction for each reactant. CLM downregulates demand (consumption) as a function of rate (demand-based competition). We examine downregulation of consumption as a function of concentration (DC) and rate (DR) for use with SU or LT to address the nonnegativity challenge.

3.3.1 Downregulation of consumption as a function of concentration

The first-order decay problem (Eqs. 2, 3, 6, 8) is nonnegative. This can be shown by solving the first-order problem $dc/dt = -c$ using the backward difference: $c^{k+1} = c^k/(1 + \Delta t)$; the concentration decreases but remains nonnegative. However, it can go to 0. To avoid 0 concentration, a residual term is often used to limit the decay to a residual concentration (?). For example, Eq. (2) becomes

$$\frac{\partial [\text{CN}_u]}{\partial t} = -k_d f_T f_w ([\text{CN}_u] - [\text{CN}_u]_r). \quad (26)$$

When the concentration goes below $[\text{CN}_u]_r$ (overshoots) in an iteration, Eq. (26) implies a hypothetical reverse reaction to bring it back to $[\text{CN}_u]_r$. An alternative is to introduce a cutoff, e.g., Eq. (2) becomes

$$\frac{\partial [\text{CN}_u]}{\partial t} = -k_d f_T f_w [\text{CN}_u] f([\text{CN}_u]), \quad (27)$$

with $f([\text{CN}_u]) = 1$ when $[\text{CN}_u] \geq [\text{CN}_u]_r$, and 0 otherwise. It is simple but does not prevent $[\text{CN}_u]$ from getting below $[\text{CN}_u]_r$ or 0 in theory. In addition, it introduces a discontinuity. A polynomial function can be used to smooth the cutoff

$$f([\text{CN}_u]) = 1 - \left[1 - \left(\frac{[\text{CN}_u] - [\text{CN}_u]_r}{[\text{CN}_u]_1 - [\text{CN}_u]_r} \right)^2 \right]^2. \quad (28)$$

This function varies from 0 at $[CN_u]_r$ to 1 at $[CN_u]_1$, with zero derivatives at both points. While not implying a nonphysical reverse reaction as Eq. (26),

$$\frac{df([CN_u])}{d[CN_u]} = 4 \frac{[CN_u] - [CN_u]_r}{([CN_u]_1 - [CN_u]_r)^2} \left[1 - \left(\frac{[CN_u] - [CN_u]_r}{[CN_u]_1 - [CN_u]_r} \right)^2 \right]. \quad (29)$$

Depending on $[CN_u]_1 - [CN_u]_r$, this cutoff does introduce a large derivative change during the transition: $df([CN_u])/d[CN_u]$ varies from 0 to 10^8 for $[CN_u]_r=10^{-10}$ and $[CN_u]_1=10^{-8}$ (Fig. 2). The maximum increases at about the same order of magnitude as the decrease of $[CN_u]_r$ and $[CN_u]_1$. Even though smoothed, the cutoff is still a sharp change. The smaller the cutoff concentrations, the sharper the transitions. Marching through a steep transition in time usually involves small time steps. In contrast, the residual concentration introduced in Eq. (26) does not adjust the derivative, which appears less nonlinear. Obviously, enforcing positivity even for a linear problem (here the first-order rate) introduces nonlinearity into the numerical problem. A smaller threshold may lead to higher nonlinearity. For example, even though a NO_3^- concentration of 10^{-10} M may not be detectable, and is not any different from 10^{-15} M, introducing the cutoff of Eq. (28) with a cutoff 10^{-15} produces five orders of magnitude more increase in the derivative than a cutoff of 10^{-10} . In the case when extremely low concentration matters (e.g., O_2), enforcing positivity with the implicit time stepping and Newton-Raphson method is numerically challenging.

For the litter decomposition reactions (R7, R8, R9) that immobilize nitrogen, the rate formulae (Eqs. 2, 26,) do not account for the limitation of the reaction rate by the availability of nitrogen. They are zero order with respect to nitrogen. Mechanistically, a nitrogen-limiting function needs to be added to those formulae to represent the limitation of decreasing nitrogen concentration on decomposition rate (downregulation), for example,

$$\frac{\partial [CN_u]}{\partial t} = -k_d f_T f_w ([CN_u] - [CN_u]_r) f([N]). \quad (30)$$

A widely used downregulation function is the Monod substrate limitation function (??):

$$f([N]) = \frac{[N]}{[N] + k_m}, \quad (31)$$

with half saturation k_m . In the case of $[N] = k_m$, $f([N]) = 0.5$. For $[N] \gg k_m$, Eq. (31) is zero order with respect to $[N]$ or has little impact on the decomposition reactions that immobilize N. For $[N] \ll k_m$, Eq. (31) approximates first order with respect to $[N]$ (Figure 3d). Nevertheless, since

$$\frac{df([N])}{d[N]} = \frac{k_N}{([N] + k_m)^2}, \quad (32)$$

the derivative increases to about k_m^{-1} as the concentration decreases to below k_m (Fig. 3). This is similar to the smoothed cutoff in Eq. (28) (Fig. 2): even though it is smoothed, it is a steep transition.

To represent the threshold concentration beyond which certain microorganisms can not use certain electron donors (?), a residual concentration was added to Eq. (31):

$$f([N]) = \frac{[N] - [N]_r}{[N] - [N]_r + k_m}. \quad (33)$$

Similar to Eq. (26), this implies a nonphysical backward reaction when $[N] < [N]_r$. If we use the 345 cutoff instead of a residual concentration, a combination of Eqs. (31) and (28) produces more non-linearity, as evidenced in Fig. (4), which is likely to result in small time step sizes to march through these highly nonlinear regions when the concentration gets there.

The plant nitrogen uptake reactions (R13, R14) are of zero order, and substrate limiting functions need to be added as well. For both plant uptake and immobilization, we also have to distribute the 350 demands between NH_4^+ and NO_3^- . If we simulate the NH_4^+ limitation on plant uptake or immobilization with

$$R_a = R_p \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m}, \quad (34)$$

the NO_3^- plant uptake or immobilization can be represented by

$$R_n = (R_p - R_a) \frac{[\text{NO}_3^-]}{[\text{NO}_3^-] + k_m} = R_p \frac{k_m}{[\text{NH}_4^+] + k_m} \frac{[\text{NO}_3^-]}{[\text{NO}_3^-] + k_m}. \quad (35)$$

355 This essentially assumes an inhibition of NH_4^+ on NO_3^- uptake or immobilization, and k_m becomes an inhibition coefficient. This simulates the observation that plant NO_3^- uptake rate remained low until NH_4^+ concentrations dropped below a threshold (?).

For comparison with CLM, we examine the uptake rate as a function of demands and available concentrations:

$$360 \quad f_{pi} = \frac{R_a + R_n}{R_p} = \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} + \left(1 - \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m}\right) \frac{[\text{NO}_3^-]}{[\text{NO}_3^-] + k_m}. \quad (36)$$

As an example, we consider uptake $R_p = 10^{-9} \text{ M s}^{-1}$ from a solution with various $[\text{NH}_4^+]$ and $[\text{NO}_3^-]$ for a 0.5 h time step. With CLM, $f_{pi} = 1$ when $[\text{NH}_4^+] + [\text{NO}_3^-] \geq R_p \Delta t$; otherwise, it decreases with decreasing $[\text{NH}_4^+] + [\text{NO}_3^-]$ (Fig. 5). The new representation (Eqs. 34, 35) is generally similar, with $f_{pi} = 1$ or 0 when $[\text{NH}_4^+]$ or $[\text{NO}_3^-] \gg$ or $\ll k_m$. For the intermediate concentrations, 365 f_{pi} in the new scheme is less than or equal to that in CLM because NH_4^+ “inhibits” NO_3^- uptake. The difference decreases with decreasing k_m , disappearing at $k_m = 10^{-10}$ (Fig. 6).

Various level of preferences of NH_4^+ over NO_3^- were observed (?????). The microbial uptake of inorganic and organic nitrogen species is similar (?????). CLM implies a strong preference for NH_4^+ over NO_3^- . For example, if NH_4^+ is abundant, NO_3^- will not be taken. The new scheme allows the 370 level of preference to be adjusted by varying k_m .

The nitrification reaction (R11) associated with decomposition to produce N_2O with the rate (Eq. 4) appears as zero order as well and needs to be further downregulated as

$$\frac{\partial [\text{NH}_4^+]}{\partial t} = -2 \frac{\partial [\text{N}_2\text{O}]}{\partial t} = -f_{nm} f_T f_w f_{pH} \max(R_{nm} \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m}, 0). \quad (37)$$

DC involves adding residual concentration and half saturations, which has the potential to provide
 375 mechanistic treatment of the processes. It also introduces a number of parameters that need to be determined. These parameters may not be well defined and can vary among different microbe and plant species under various conditions across the globe. An alternative is to downregulate consumption as a function of rate (DR) like CLM demand-based competition.

3.3.2 Downregulation of consumption as a function of rate

380 As described in Section 2.1.5, CLM splits the available nitrogen to meet the demands by microbes (immobilization) and plants in proportion to the potential rates (?). Suppose that the demand (consumption, sink, including immobilization, plant uptake, nitrification, etc.) rate is R_{dp} ; the supply (source, production, including deposition, mineralization, etc.) is rate R_s , and the concentration is [N] at the beginning of the time step, the demand is downregulated to

$$385 \quad R_d = \min \left[R_{dp}, R_s + \frac{[N]}{\Delta t} \right]. \quad (38)$$

CLM4.5 ignores the supply term R_s (?). Applying to $dc/dt = -1$, $c^{k+1} = c^k - \Delta t$ for $c^k \geq \Delta t$; for $c^k < \Delta t$, $c^{k+1} = c^k/2$: the concentration decreases by half in a time step. This is the similar to SU with $\alpha = 0.5$, which is suggested by ?. DR essentially downregulates the demands using the first-order rate when nitrogen is limiting. Eq. (38) is similar to Eq. (31), except that Eq. (31) switches
 390 from zero to the first-order rate smoothly, while Eq. (38) has a discontinuity.

Implementation of DR in a geochemical code like PFLOTRAN involves separating the supply and consumption rates for each species in each reaction, and checking and conducting downregulation when necessary after contributions from all of the reactions are accumulated. It involves not only the rate terms for the residual but also the derivative terms for the Jacobian. The complexity
 395 explodes when more species needs to be downregulated (e.g., NH_4^+ , NO_3^- , and organic N) and there are transformation processes among these species. This is shown in Appendix A, which describes the implementation of downregulation of NH_4^+ and NO_3^- with a nitrification reaction from NH_4^+ to NO_3^- . Basically, it becomes challenging to separate, track, and downregulate consumption and production rates for an indefinite number of species, and calculation of the Jacobian becomes
 400 convoluted.

4 Test problems, results, and discussions

We can use SU or LT with either DC or DR, or a combination of both, to enforce nonnegativity. SU has the potential to introduce numerical error due to false convergence when STOL is satisfied and the scaling factor λ is too small. Stopping the simulation when λ is smaller than λ_{\min} changes the
 405 accuracy issue into a stability or robustness issue. For both SU and LT, allowing for tiny time step

sizes and an excessive number of time step cuts may transfer a stability issue into an efficiency issue. We examine the causes of negative concentration; the advantages and disadvantages of SU, LT, DC, and DR; and the accuracy, stability, and efficiency issue using both simple test problems and coupled CLM-PFLOTRAN site simulations. For simple test problems, we assess the conditions under
410 which large updates occur. With coupled CLM-PFLOTRAN spin-up simulations for arctic, temperate, and tropical sites, we assess the conditions when the nonnegativity issue arises by relaxing the downregulation on nitrogen consumption (decreasing half saturation). Spreadsheet, python scripts, and PFLOTRAN input files for these simple test problems and for producing figures are provided as supplemental information.
415 Our implementation of CLM biogeochemistry introduces mainly two parameters: half saturation k_m and residual concentration. A wide range of k_m values were reported for NH_4^+ , NO_3^- , and organic nitrogen for microbes and plants. The median, mean, and standard deviations range from $10 \sim 100$, $50 \sim 500$, and $10 \sim 200 \mu\text{M}$, respectively (?). Reported residual (threshold) concentrations are limited and are considered to be 0 (e.g., ?), likely because of the detection limits of the analytical
420 methods. The detection limits are usually at the μM level, while up to the nM level was reported (?). In Ecosys, the k_m is 0.40 and 0.35 gN m^{-3} , and the residual concentration is 0.0125 and 0.03 gN m^{-3} (?) for NH_4^+ and NO_3^- for microbes. Depending on the moisture content, say 10, 100, and 1000 L m^{-3} for a dry, wet, and saturated case, these k_m values roughly translate to 10^{-3} , 10^{-4} , and 10^{-5} M . We start with $k_m = 10^{-6} \text{ M}$ or mol m^{-3} , and residual concentration 10^{-15} M or mol m^{-3}
425 for plants and microbes. To further investigate the nonnegativity challenge for the current study and for future application for other nutrients (e.g., H_2 and O_2) where the concentrations can be much lower, we examine k_m from 10^{-3} to 10^{-12} in our test problems. The k_m is expected to be different for different plants, microbes, and for NH_4^+ and NO_3^- , and different values can be assigned in the input file in our implementation. we do not differentiate them in this work as we focus on numerical
430 issues.

4.1 Plant nitrogen uptake, nitrification, and denitrification

It was observed that plants can decrease nitrogen concentration to below detection limits in hours (?). Plant nitrogen uptake is one of the major sinks for nitrogen in TEMs and contributes to decreasing nitrogen concentration numerically to 0 or negative. In CLM, the total plant nitrogen demand
435 is calculated based on photosynthesized carbon allocated for new growth and the C:N stoichiometry for new growth allocation, and the plant nitrogen demand from the soil is equal to the total nitrogen demand minus retranslocated nitrogen stored in the plants (?). The CLM calculated rate is provided as an input to PFLOTRAN, which is constant in a 0.5 h time step. Without downregulating plant nitrogen uptake rate, nitrogen concentration is likely to go negative when the net consumption rate overwhelms a low available concentration. As the Monod substrate limiting function is the most widely used downregulation function, we examine the numerical solutions, beginning with the
440

numerical issues during the iteration processes. Incrementally, we add first order reactions (e.g., nitrification, denitrification, and plant NO_3^- uptake) to look into these issues in increasingly complex systems.

445 4.1.1 Plant NH_4^+ uptake (Test 1)

We consider the plant NH_4^+ uptake reaction (R13) with the rate R_a downregulated by the Monod substrate limiting function:

$$\frac{d[\text{NH}_4^+]}{dt} = -R_a \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} = -R_{at}. \quad (39)$$

As an analytical solution is not available, discretizing Eq. (39) using the backward difference, it
450 becomes

$$[\text{NH}_4^+]^{k+1} + (k_m + R_a \Delta t - [\text{NH}_4^+]^k) [\text{NH}_4^+]^{k+1} - k_m [\text{NH}_4^+]^k = 0. \quad (40)$$

The two roots are

$$[\text{NH}_4^+]^{k+1} = 0.5 \left[[\text{NH}_4^+]^k - k_m - R_a \Delta t \pm \sqrt{([\text{NH}_4^+]^k - k_m - R_a \Delta t)^2 + 4k_m [\text{NH}_4^+]^k} \right]. \quad (41)$$

One root is positive, and the other is negative (not physical). With the positive root, $[\text{NH}_4^+]^{k+1} \rightarrow 0$ when $[\text{NH}_4^+]^k \rightarrow 0$, or $R_a \Delta t \rightarrow \infty$. As $k_m \rightarrow 0$, $[\text{NH}_4^+]^{k+1} \rightarrow 0$ when $[\text{NH}_4^+]^k \leq R_a \Delta t$. If we
455 replace $[\text{NH}_4^+]$ with $[\text{NH}_4^+] - [\text{NH}_4^+]_r$ in Eqs. (39) and (41), $[\text{NH}_4^+]^{k+1} \rightarrow [\text{NH}_4^+]_r$ rather than 0. We do not include the residual concentration in the simple test problems for simplicity in this subsection. Ignoring the negative root, the representation of plant NH_4^+ uptake by Eq. (39) ensures $[\text{NH}_4^+]^{k+1} \geq [\text{NH}_4^+]_r$ or 0.

460 Solving Eq. (39) using the Newton-Rapshon method from time step k to $k+1$, the update is

$$\delta[\text{NH}_4^+]^{k+1,p} = \frac{\frac{[\text{NH}_4^+]^{k+1,p} - [\text{NH}_4^+]_k}{\Delta t} + R_a \frac{[\text{NH}_4^+]^{k+1,p}}{[\text{NH}_4^+]^{k+1,p} + k_m}}{\frac{1}{\Delta t} + R_a \frac{k_m}{([\text{NH}_4^+]^{k+1,p} + k_m)^2}}. \quad (42)$$

For the first iteration, $[\text{NH}_4^+]^{k+1,0} = [\text{NH}_4^+]^k$,

$$\xi = \frac{\delta[\text{NH}_4^+]^{k+1,1}}{[\text{NH}_4^+]^{k+1,0}} = \frac{\frac{1}{[\text{NH}_4^+]^k + k_m}}{\frac{1}{R_a \Delta t} + \frac{k_m}{([\text{NH}_4^+]^k + k_m)^2}}. \quad (43)$$

The update is a function of uptake rate (R_a), time step size (Δt), concentration ($[\text{NH}_4^+]^k$), and
465 half saturation (k_m).

For $R_a\Delta t$ from 0 to ∞ , ξ increases from 0 to $1 + [\text{NH}_4^+]/k_m$. With $[\text{NH}_4^+]^k = 10^{-6}$ M, and $k_m = 10^{-6}$ M, the update is less than $[\text{NH}_4^+]^k$ when $R_a\Delta t \leq 5 \times 10^{-6}$ M (say using CLM $\Delta t = 1800$ s, $R_a \leq 2.8 \times 10^{-9}$ mol s $^{-1}$, Fig. 7); when $R_a\Delta t$ increases, the update increases, with an upper limit of 2, suggesting that large time step size and uptake rate alone usually do not lead to an excessively large update or a small scaling factor.

For $[\text{NH}_4^+]^k$ from 0 to ∞ , ξ starts from $R_a\Delta t/(k_m + R_a\Delta t)$, increases to peak at $\sqrt{R_a\Delta t/k_m}/2$ when $[\text{NH}_4^+] = \sqrt{k_m R_a \Delta t} - k_m$, and then decreases to 0. With $k_m = 10^{-6}$ M and $R_a\Delta t = 0.001$ M, ξ starts from ~ 1 , increases to peak at ~ 16 when $[\text{NH}_4^+]^k = 3 \times 10^{-5}$, and decreases to 0 as $[\text{NH}_4^+]^k \rightarrow \infty$.

For k_m from ∞ to 0, ξ increases from 0 to $R_a\Delta t/k_m$. With $R_a\Delta t = 10^{-3}$, and $[\text{NH}_4^+]^k = 10^{-6}$, the update increases several orders of magnitude as k_m decreases from 10^{-6} to 10^{-9} , with a limit of $\xi = 10^3$ (Fig. 7).

While the semi-analytical solution (Eq. 41) indicates that the problem itself is nonnegative, these calculations suggests that overshoot can occur, and large time step, uptake rate, low concentration and half saturation can contribute to large update that can exceed the available concentration by orders of magnitude.

Now we look into the iteration processes: the solution starts with an overshoot to 1.995×10^{-6} and converges to the positive semi-analytical solution 3.1127×10^{-5} (Eq. 41) in eight iterations when $R_a\Delta t = 0.001$ (Table 1). dR/dC changes from ~ 1 to 10^5 in the first iteration, although the change in the Jacobian is reduced due to a small $R_a\Delta t$. With $R_a\Delta t = 0.002$, the update δ is greater than $[\text{NH}_4^+]_k$; without scaling back the update, the solution converges to the negative solution -1.002×10^{-3} (Table 2, negative root from Eq. 41). Scaling back the update with $\alpha = 0.9999$, the concentration decreases by $1 - \alpha = 10^{-4}$ times in the first iteration rather than to negative root (Table 3). The solution converges to the positive root in seven iterations.

If LT is used with $R_a\Delta t = 0.002$, the solution oscillates and results in a δ of -1041.5 in the fifth iteration, which essentially makes the solution 0 (Table 4). Limiting $\delta \leq \delta_{\ln,\max}$, the solution converges to the semi-analytical solution (Eq. 41) in 7, 7, 6, 5, 6, and 8 iterations with $\delta_{\ln,\max} = 2, 3, 4, 5, 6$, and 7, respectively. For $\delta_{\ln,\max} \geq 8$, the iteration oscillates and converges slowly. If we split $R_a\Delta t = 0.002$ into two steps, the first step takes seven iterations to converge, while the second step takes nine iterations, with $\delta_{\ln,\max} = 5$. These simulations show that limiting the update to be less than $\delta_{\ln,\max}$ is helpful for LT to use large time step sizes for efficiency.

Finally, we consider a PFLOTRAN simulation: a 1 m \times 1 m \times 1 m grid cell, with 0.25 water content, initial 4 μM NH_4^+ , and a plant NH_4^+ demand 10^{-7} mol s $^{-1}$ (~ 3 mg d $^{-1}$, reported values for evergreen and deciduous forest range from $0.3 \sim 10$ g m $^{-2}$ year, ?, ?). It takes 10,000 s (2.78 h) to consume NH_4^+ . We use a time step of 0.5 h for a simulation duration of 10 h. Without downregulating the uptake rate, SU decreases $[\text{NH}_4^+]$ to 0 and then keeps it 0 as $\lambda = 0$. With LT, we obtain an accurate solution until it fails when the concentration goes to 0, and the Jacobian becomes singular.

Using SU with DC with $[\text{NH}_4^+]_r$ of 10^{-20} M (Eq. 33) and a half saturation k_m of 10^{-6} , 10^{-9} , or 10^{-12} , the calculated concentration $[\text{NH}_4^+]$ is kept to be greater than or equal to 10^{-20} (Figure. 8). With a k_m of 10^{-6} , the calculated $[\text{NH}_4^+]$ is above 10^{-20} , and the update does not exceed the concentration in any iteration during the 10 h simulation duration. With a k_m of 10^{-9} and 10^{-12} , the update exceeds the concentration from time 2.5 h to 3.0 h. In the case of $k_m = 10^{-12}$ M for the next time step, the two iterations decrease the concentration by 10,000 times, and the solution is deemed converged as the default STOL = 10^{-8} is met. This continues for another time step with one iteration (Fig. 8). In these two time steps, the solution does not converge to the exact solution. If we set the residual concentration to 10^{-15} , the false convergence results in a concentration below 10^{-15} (8.71×10^{-16}) in the first of the two time steps. If we set STOL = 10^{-50} to avoid the convergence due to Eq. (17), PFLOTTRAN cuts the time step sizes and quits due to too many cuts.

Using LT can avoid the false convergences in this example, but it requires more iterations, with 56/95, 36/110, and 32/87 for SU/LT for k_m values of 10^{-6} , 10^{-9} , and 10^{-12} M, respectively. These results demonstrate that SU ensures nonnegativity and allows large time step for efficiency at the risk of inaccuracy. LT is accurate but requires more iterations.

4.1.2 Plant NH_4^+ uptake and nitrification (Test 2)

Adding a nitrification reaction (R10) with a first-order rate, Eq. (39) becomes

$$520 \quad \frac{d[\text{NH}_4^+]}{dt} = -R_a \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} - k_{nitr} [\text{NH}_4^+] = -R_{at} - R_{nitr}. \quad (44)$$

Discretizing the equation in time using the backward difference,

$$(1 + k_{nitr} \Delta t)[\text{NH}_4^+]^{k+1}^2 + (k_m + R_a \Delta T + k_{nitr} k_m \Delta t - [\text{NH}_4^+]^k) \text{NH}_4^+^{k+1} - k_m [\text{NH}_4^+]^k = 0. \quad (45)$$

Denote $-b = k_m + R_a \Delta T + k_{nitr} k_m \Delta t - [\text{NH}_4^+]^k$, the roots are

$$[\text{NH}_4^+]^{k+1} = \frac{\left[b \pm \sqrt{b^2 + 4(1 + k_{nitr} \Delta t)[\text{NH}_4^+]^k k_m} \right]}{2(1 + k_{nitr} \Delta t)}. \quad (46)$$

525 Similar to Eq. (39), one root is nonnegative. To solve the problem with the Newton-Raphson method, the equation is

$$\begin{bmatrix} \frac{1}{\Delta t} + J_{at} + J_{nitr} & 0 & 0 \\ -J_{at} & \frac{1}{\Delta t} & 0 \\ -J_{nitr} & 0 & \frac{1}{\Delta t} \end{bmatrix} \begin{pmatrix} \delta[\text{NH}_4^+]^{k+1,1} \\ \delta[\text{PlantA}]^{k+1,1} \\ \delta[\text{NO}_3^-]^{k+1,1} \end{pmatrix} = \begin{pmatrix} R_{at} + R_{nitr} \\ -R_{at} \\ -R_{nitr} \end{pmatrix}, \quad (47)$$

with $J_{at} = \frac{dR_{at}}{d[\text{NH}_4^+]} = R_a \frac{k_m}{([\text{NH}_4^+] + k_m)^2}$, and $J_{nitr} = \frac{dR_{nitr}}{d[\text{NH}_4^+]} = k_{nitr}$. The solutions are

$$\delta[\text{NH}_4^+]^{k+1,1} = \frac{R_{at} + R_{nitr}}{\frac{1}{\Delta t} + J_{at} + J_{nitr}}, \quad (48)$$

530 $\delta[\text{PlantA}]^{k+1,1} = J_{at}\delta[\text{NH}_4^+]^{k+1,1} - R_{at} = -\frac{\frac{1}{\Delta t} + J_{nitr}}{\frac{1}{\Delta t} + J_{at} + J_{nitr}} R_{at} + \frac{J_{at}}{\frac{1}{\Delta t} + J_{at} + J_{nitr}} R_{nitr}, \quad (49)$

and

$$\delta[\text{NO}_3^-]^{k+1,1} = J_{nitr}\delta[\text{NH}_4^+]^{k+1,1} - R_{nitr} = \frac{J_{nitr}}{\frac{1}{\Delta t} + J_{at} + J_{nitr}} R_{at} - \frac{\frac{1}{\Delta t} + J_{at}}{\frac{1}{\Delta t} + J_{at} + J_{nitr}} R_{nitr}. \quad (50)$$

It is not a surprise for $\delta[\text{NH}_4^+]^{k+1,1}$ to be positive as plant NH_4^+ uptake and nitrification consume NH_4^+ . It is a little surprising that a positive component, $\frac{J_{at}}{1/\Delta t + J_{at} + J_{nitr}} R_{nitr}$, is introduced to 535 $\delta[\text{PlantA}]^{k+1,1}$ even though there is no reaction that consumes PlantA. Depending on the relative magnitude of the rates (R_{at} and R_{nitr}) and the derivatives (J_{at} and J_{nitr}), as well as time step size Δt and $[\text{NH}_4^+]^k$, the update can be positive, which can result in too small a scaling factor that causes a nonnegativity issue if [PlantA] is very small. This demonstrates that products as well as reactants can be driven negative during the Newton-Raphson iterations (overshoot).

540 This can also be shown with another simple first-order reaction $A \rightarrow B$ with rate $k[A]$: taking out the equation for B, $-k\delta[A] + \delta[B]/\Delta t = -k[A]$, $\delta[B] = k\Delta t(\delta[A] - [A])$. Suppose $\delta[A] = 2 \times 10^{-6}$, $[A] = 10^{-6}$, and $k\Delta t = 10^{-3}$, then $\delta[B] = 10^{-9}$. If $[B]$ is very small, say 10^{-20} , this results in a $\lambda = 0.99 \times 10^{-11}$. This has an implication for DR: downregulating only some reactants may not be sufficient to guarantee positivity, while downregulating all species is complicated.

545 The update δNO_3^- can be positive, and cause negative consequence for the calculation as well. It can become worse when the same Monod substrate limiting function is used to further downregulate the nitrification rate like CLM demand based competition, namely, $R_{nitr} = k_{nitr}[\text{NH}_4^+] \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m}$, $J_{nitr} = k_{nitr} \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} + k_{nitr}[\text{NH}_4^+] \frac{k_m}{([\text{NH}_4^+] + k_m)^2} = k_{nitr} \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} \left(1 + \frac{k_m}{[\text{NH}_4^+] + k_m}\right)$. For $[\text{NH}_4^+] \ll k_m$, this increases J_{nitr} approximately by the order of k_m^{-1} , which can result in an increase of the 550 update.

4.1.3 Plant uptake, nitrification, and denitrification (Test 3)

Adding plant NO_3^- uptake reaction (R14) with rate $R_{nt} = R_p \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} \frac{[\text{NO}_3^-]}{[\text{NO}_3^-] + k_m}$, $J_{nt,n} = \frac{dR_{nt}}{d[\text{NO}_3^-]} = R_p \frac{[\text{NH}_4^+]}{[\text{NH}_4^+] + k_m} \frac{k_m}{([\text{NO}_3^-] + k_m)^2}$, and $J_{nt,a} = \frac{dR_{nt}}{d[\text{NH}_4^+]} = \frac{dR_n}{d[\text{NH}_4^+]} \frac{k_m}{([\text{NH}_4^+] + k_m)^2} \frac{[\text{NO}_3^-]}{[\text{NO}_3^-] + k_m}$, and denitrification reaction (R12) with rate $R_{deni} = k_{deni} [\text{NO}_3^-]$, and $J_{deni} = \frac{dR_{deni}}{d[\text{NO}_3^-]} = k_{deni}$, the equation becomes

$$\begin{bmatrix} \frac{1}{\Delta t} + J_{at} + J_{nitr} & 0 & 0 & 0 & 0 \\ -J_{at} & \frac{1}{\Delta t} & 0 & 0 & 0 \\ -J_{nitr} + J_{nt,a} & 0 & \frac{1}{\Delta t} + J_{nt} + J_{deni} & 0 & 0 \\ -J_{nt,a} & 0 & -J_{nt,n} & 1/\Delta t & 0 \\ 0 & 0 & -0.5J_{deni} & 0 & 1/\Delta t \end{bmatrix} \begin{pmatrix} \delta[\text{NH}_4^+]^{k+1,1} \\ \delta[\text{PlantA}]^{k+1,1} \\ \delta[\text{NO}_3^-]^{k+1,1} \\ \delta[\text{PlantN}]^{k+1,1} \\ \delta[\text{N}_2]^{k+1,1} \end{pmatrix} = \begin{pmatrix} R_{at} + R_{nitr} \\ -R_{at} \\ -R_{nitr} + R_{nt} + R_{deni} \\ -R_{nt} \\ -0.5R_{deni} \end{pmatrix}. \quad (51)$$

555

As a result,

$$\delta[\text{NO}_3^-]^{k+1,1} \left(\frac{1}{\Delta t} + J_{nt} + J_{deni} \right) = (J_{nitr} - J_{nt,a})(R_{at} + R_{nitr})\delta[\text{NH}_4^+]^{k+1,1} - R_{nitr} + R_{nt} + R_{deni}, \quad (52)$$

and

$$\delta[\text{PlantN}]^{k+1,1} / \Delta t = J_{nt,a}\delta[\text{NH}_4^+]^{k+1,1} + J_{nt,n}\delta[\text{NO}_3^-]^{k+1,1} - R_{nt}. \quad (53)$$

560 In addition to plant NH_4^+ uptake and nitrification, $\delta[\text{NO}_3^-]^{k+1,1}$ becomes a function of plant NO_3^- uptake and denitrification, and all of the rate coefficients and derivatives may contribute to a positive update that can result in too small a scaling factor. Even though the problem itself is nonnegative, coupling these reactions together in the Newton-Raphson iteration does introduce the potential to produce relatively significant positive update for NO_3^- and PlantN. It worsens when the preference-based uptake of NH_4^+ over NO_3^- (Eqs. 34 and 35) brings in the impact of any rates related to NH_4^+ (including deposition, uptake, immobilization, etc.), with potentially large derivative terms for the off-diagonal entries in the Jacobian matrix. It is the nonlinearity introduced in the downregulation and propagated through the reactions that causes the challenge. A smaller half saturation introduces a greater derivative change; therefore, it is more likely to cause the problem. These results also suggest
565 that the likelihood for NO_3^- to go negative is greater than NH_4^+ , and PlantN is greater than PlantA
570 as the latter are influenced by more rates and derivatives.

4.2 N immobilization, mineralization, and nitrification during decomposition (Test 4)

We further examine the implications of a small half saturation on the numerical solutions with a decomposition test problem. We consider a case of decomposing 0.2 M Lit1C + 0.005 M Lit1N to

575 produce SOM1 with an initial $4 \mu\text{M}$ NH_4^+ using the reactions (R7 and R2) in the CLM-CN reaction network (Fig. 1) at first. Then we add the nitrification reaction (R11) with rate (Eq. 37) to examine the implications of a complex rate formula. We use PFLOTRAN with a fully saturated grid cell of 1 m and porosity of 0.25.

580 Lit1 decomposes fast at the beginning and then slows down as NH_4^+ is depleted (Fig. 9). The Lit1 decomposition rate is controlled by the mineralization rate from SOM1 decomposition. As the immobilization rate decreases with decreasing Lit1, $[\text{NH}_4^+]$ rebounds. For k_m of 10^{-6} , 10^{-9} , and 10^{-12} M , Lit1 and SOM1 dynamics are similar, but the $[\text{NH}_4^+]$ values are decreased to $\sim 10^{-8}$, 10^{-11} , and 10^{-14} M , respectively. The number of iterations for SU/LT is 76/169, 68/194, and 54/207 for the three k_m values. Obviously, smaller k_m results in lower $[\text{NH}_4^+]$ and more iterations.

585 The implication of $k_m = 10^{-12}$ becomes obvious when the nitrification reaction (R11) with rate (Eq. 37) is included. In the time step when the mineralization rate surpasses the immobilization, $\delta[\text{N}_2\text{O}] > [\text{N}_2\text{O}]$, and N_2O is lowered to 10^{-23} M in three iterations as STOL is met. In the next time step, $\delta[\text{N}_2\text{O}] = \sim 10^{-13}$, resulting in a scaling factor λ of 10^{-10} that stops the simulation. Checking the calculations in the last iteration, the lower NH_4^+ concentration ($2 \times 10^{-13} \text{ M}$) introduces a derivative (6.94×10^{11}) into the column of the Jacobian with respect to NH_4^+ . Because N_2O production is a function of both immobilization and mineralization, a Monod substrate limiting function is added to downregulate the mineralization component (Eq. 37). In the row corresponding to N_2O , the off-diagonal terms are nonzero in the entries with respect to NH_4^+ , Lit1C, Lit1N, and SOM1. Even though these values are relatively small because the fraction of net mineralization rate 590 to N_2O is small (0.02), inversion of the Jacobian matrix results in substantial values in the row, and the 10^{-13} M update comes mainly from the entries with respect to Lit1C and Lit1N. This further illustrates that extremely small half saturation together with complex rate limiting formula can result in large derivatives at small concentration, which may lead to overshoot that results in a small update scaling factor.

600 4.3 CLM-PFLOTRAN simulations

We examine the nonnegative challenge in the coupled CLM-PFLOTRAN simulations for arctic (US-Brw), temperate (US-WBW), and tropical (BR-Cax) AmeriFlux sites. The CLM-PFLOTRAN simulations are run in the mode that PFLOTRAN only handles subsurface chemistry (decomposition, nitrification, denitrification, N plant uptake). For comparison, 1) depth and O_2 availability 605 impact on decomposition, 2) cryoturbation, and 3) SOM transport are ignored by setting 1) `comp_depth_efolding` to 10^6 m , `o_scalar` to 1, 2) `cryoturb_diffusion` and 3) `som_diffusion` to 0. For CLM-PFLOTRAN simulations, $k_m = 10^{-6}$ and residual concentration 10^{-15} are used for the base case. The nonnegativity challenge is assessed in particular by relaxing the downregulation (decreasing k_m). Nitrogen is the focus as it is the rate limiting nutrient in CLM4.5 biogeochemistry. Spin-up 610 simulations are used because they are generally more likely to face a nonnegativity challenge be-

cause the simulations start far away from equilibrium. In these site simulations, PFLOTRAN uses the same 10 layer grid for the 3.8 m one-dimensional column as CLM. The simulation duration is 500, 300, and 300 year for the arctic, temperate, and tropical sites as the system roughly reaches steady state within these simulation durations.

615 4.3.1 Site descriptions

The US-Brw site (71.35N,156.62W) is located near Barrow, Alaska. The mean annual temperature, precipitation, and snowfall are -12°C , 11 cm, and 69 cm, respectively (1971 \sim 2000) (?). The landscape is poorly drained polygonized tundra. The maximum thaw depth ranges from 30 to 40 cm, and the snow free-period is variable in length but generally begins in early June and lasts until early 620 September (?). The area is composed of several different representative wet-moist coastal sedge tundra types, including wet sedges, grasses, moss, and assorted lichens. The leaf area index (LAI) is ~ 1.1 .

The US-WBW site (35.96N, 84.29W) is located in the Walker Branch Watershed in Oak Ridge, Tennessee (?). The climate is typical of the humid southern Appalachian region. The mean annual 625 precipitation is ~ 139 cm, and the mean median temperature is 14.5°C . The soil is primarily Ultisols that developed in humid climates in the temperate zone on old or highly weathered material under forest. The temperate deciduous broadleaf forest was regenerated from agriculture land 50 years ago. LAI is ~ 6.2 (?).

The BR-Cax site (-1.72N, -51.46W) is located in the eastern Amazon tropical rainforest. The 630 mean annual rainfall is between 2000 and 2500 mm, with a pronounced dry season between June and November. The soil is a yellow oxisol (Brazilian classification latossolo amarelo) with a thick stony laterite layer at 3 \sim 4 m depth (?). The vegetation is evergreen broadleaf forest. The LAI is 4 ~ 6 (?).

4.3.2 Base case simulation results

635 For CLM-PFLOTRAN base case simulations, we use DC and LT because DC is the general geochemical modeling approach and is rigorous, and LT is less likely to cause numerical error than SU. The site climate data from 1998 to 2006, 2002 to 2010, and 2001 to 2006 are used to drive the spin-up simulation for the arctic, temperate, and tropical sites, respectively. This introduces a multi-year cycle in addition to the annual cycle (Figs. 10, 11, 12). Overall, CLM-PFLOTRAN roughly reproduces 640 CLM4.5 in predicting LAI and nitrogen distribution among vegetation, litter, SOM, NH_4^+ and NO_3^- pools for the arctic (Fig. 10), temperate (Fig. 11), and tropical site (Fig. 12). CLM calculation does accumulate more SOMN (nitrogen in the four SOM pools) than CLM-PFLOTRAN as the simulation is close to equilibrium, resulting in higher $[\text{NH}_4^+]$ and $[\text{NO}_3^-]$ in CLM than CLM-PFLOTRAN at later times. This will be further investigated in future work while the current work focuses on non-

645 negativity. Despite these differences, the calculated LAI, VEGN (vegetation N), and LITN (nitrogen in the three litter pools) are very close.

The arctic site shows a distinct summer growing season (Fig. 10): LAI and VEGN jump up at the beginning, then level off, and drop down at the end of the growing season when LITN jumps up due to litter fall. NH_4^+ and NO_3^- drop to very low level at the beginning of growing season 650 and accumulate at the other times. In addition to a longer growing season than the arctic site, the temperate site shows more litter fall by the end of the growing season as it is a temperate deciduous forest, which introduces immobilization demand that further lowers NH_4^+ and NO_3^- concentrations (Fig. 11e inset). The seasonality is much less apparent in the tropical site than in the arctic and temperate sites. LAI, VEGN, LITN, and SOMN accumulate with less seasonal variations to reach 655 an equilibrium. It takes much longer simulation duration to spin up the simulation for the arctic site than the temperate and tropical sites because the latter receive more solar radiation.

Except for the tropical site where the higher k_m of 10^{-3} mol m $^{-3}$ results in lower immobilization, higher accumulation of LITN, and higher $[\text{NH}_4^+]$ and $[\text{NO}_3^-]$ during the spinup (Fig. 12), the range of k_m values (10^{-3} , 10^{-6} , and 10^{-9} mol m $^{-3}$) generally has limited impact on the overall calculations 660 except that the nitrogen concentrations drop lower with lower k_m values (e.g., inset in Figs. 10e,f, 11e), which is similar to Fig. (9). The lack of sensitivity is because these very low concentrations do not make up a mass of nitrogen that is significant enough to influence the carbon and nitrogen cycle. However, as a small k_m value means weak downregulation and steep transition between zero 665 order and first order, it has implications on the accuracy, stability, and efficiency of the numerical solutions.

4.3.3 Efficiency

We use OIC (ORNL Institutional Cluster phase5 esd13q) for the comparison of computing time for the three sites using DC or DR with SU or LT. The results suggest that SU takes about 30~80% more time than CLM, and LT costs about two to four times that of CLM (Table 5). For LT with 670 DC, the computing time increases with decreasing k_m . This is expected because a smaller k_m may require smaller time step size to march through steeper transition. The computing time appears not to be infeasibly high. The comparison between DC and DR is mixed. This is because DC has more complicated substrate limiting representation that requires smaller time step sizes to march through steep transitions, while DR has simpler substrate limiting representation, but a discontinuity that 675 may require more iterations. We expect that CLM-PFLOTRAN will require more computational cost as more complex biogeochemical processes are represented, transport is added, and thermal hydrology (including freeze and thaw) are incorporated. Future code optimization for performance on next-generation supercomputers is expected to mitigate the computational cost.

4.3.4 Accuracy

680 We use the tropical site to demonstrate the accuracy issue by SU (scaling update); i.e., the small scaling factor (λ) may not only limit N consumption (immobilization and uptake) to account for availability limitation but also “inhibit” accumulation from mineralization, deposition, and fixation. The latter is unintended consequence and can introduce numerical error when the solution is considered converged because STOL is met. With $k_m = 10^{-3}$ mol m⁻³, the LAI, VEGN, LITN, SOMN, 685 NH₄⁺, and NO₃⁻ are similar to that calculated using LT (Fig. 12 vs. 13). With decreasing k_m , the results diverge by SU but not by LT. With k_m of 10⁻⁶ mol m⁻³, much less nitrogen is predicted to accumulate with SU than with LT.

Checking NH₄⁺ and NO₃⁻ concentration in the first year, SU calculation is similar to LT when NH₄⁺ and NO₃⁻ are abundant due to nitrogen mineralization from SOM4 decomposition at early 690 times (Fig. 14). As nitrogen becomes limiting at late times due to increased plant nitrogen demand, NH₄⁺ and NO₃⁻ accumulate lower by SU than by LT. Further checking into the daily cycles (Fig. 14 inset) reveals that NH₄⁺ and NO₃⁻ accumulate during the night when photosynthesis and plant 695 nitrogen uptake pause. As the sun rises, photosynthesis and plant nitrogen uptake resume, decreasing nitrogen concentrations to very low level, limiting further uptake. As a result, nitrogen rebounds and then decreases again until sunset. From day 226 to 230, SU calculations are only slightly lower than LT calculations during the rebound. From day 250 to 254, the SU calculated rebounds are much lower than LT calculations. From day 295 to 300, accumulation during both night and day are decreased. Overall, NH₄⁺ and NO₃⁻ accumulation appear to be numerically “inhibited” when the system becomes increasingly nitrogen limiting and when SU is used.

700 Finally checking nitrogen dynamics between day 250 and 254, these “inhibited” intervals in Fig. (14) coincide with the concentration valleys in the diagnostic species N₂Od, which is used to track N₂O production (reaction R11) due to decomposition (Eq. 37, Fig. 15). At the beginning of each CLM 0.5 h time step, we reset [N₂Od] to 10⁻¹⁰ in PFLTRAN so that it can be used to calculate the reaction rate for CLM (instead of saving concentration in previous time step) As we do not have 705 any consumption reaction for N₂Od, the decreases are purely numerical. The contribution of the derivatives of the rate (Eq. 37) results in a large enough positive update to N₂Od, which results in a small scaling factor, and the solution is considered converged due to less than STOL update. This small scaling factor and false convergence result in the “inhibition” of NH₄⁺ and NO₃⁻ production, and manifest as much less nitrogen accumulation in Fig. (13). If we remove the contribution of the 710 derivatives for Eq. (37) for the entries with respect to N₂Od from the Jacobian, the “inhibition” disappears and SU results become close to LT (Fig. 15). If we decrease STOL from 10⁻⁸ to 10⁻⁵⁰, and λ_{min} from 10⁻¹⁰ to 10⁻⁵⁰, the false convergence occurs less often and the difference from LT decreases. This can also be mitigated by increasing the value from 10⁻¹⁰ to 10⁻⁶ that we reset these diagnostic species concentrations.

715 An alternative to mitigate the issue is to use DR rather than DC. The rate formulae for DR can
be very simple (no need for substrate limiting function). Because we use k_m to distribute nitrogen
demand between NH_4^+ and NO_3^- (Eqs. 34 and 35) for plant nitrogen uptake and immobilization, our
implementation ends up with the difference between DC and DR as that DR does not have NO_3^-
substrate limiting terms and uses Eq. (38) for downregulation of both NH_4^+ and NO_3^- . With DR, the
720 numerical error is much reduced (Fig. 13 vs. 16). However, it increases slightly with decreasing k_m
when comparing with LT (not shown). These results show that decreasing the complexity of the rate
limiting representation helps reducing the nonnegativity challenge for SU.

4.3.5 Stability and robustness

PFLOTTRAN may stop CLM-PFLOTTRAN run due to 1) too small a scaling factor for SU to avoid
725 excessive numerical error, 2) too small a time step size or too many time step cuts for either SU or
LT, and 3) singular Jacobian matrix for LT.

With SU and DC, the simulations run to conclusion for $k_m = 10^{-3}$, 10^{-6} , and $10^{-9} \text{ mol m}^{-3}$
but abort for $k_m = 10^{-12}$ for all of the three sites (Table 6). The reasons for the unfinished runs are
too small a scaling factor for NO_3^- (Table 6). As indicated in Test 3, it is not surprising to see posi-
730 tive updates to NO_3^- as they depend on source and sink terms not only for NO_3^- but also for NH_4^+
because of the nitrogen demand distribution scheme (Eq. 35) for plant uptake and immobilization.
With smaller k_m , the derivative of the substrate limiting function (Eq. 32) becomes bigger when the
concentration is lower, and the off-diagonal entries in the Jacobian matrix become more dominant,
increasing the probability of producing positive update. When a greater than $[\text{NO}_3^-]$ update is pro-
735 duced, $[\text{NO}_3^-]$ is expected to decrease by $1-\alpha$ times (0.01 by default) for each subsequent iterations,
making it lower than the residual concentration (10^{-15}). If the λ is not greater than 10^{-10} , and the
solution may be considered converged because STOL is satisfied, $[\text{NO}_3^-]$ can be very low, much
below residual concentration (10^{-26}) in Table 6. With such low concentrations, a positive update
can make the scaling factor to be less than 10^{-10} . The arctic site simulation aborts at a much later
740 time than the temperate and tropical site because the uptake and immobilization rate is much smaller
in the former site.

A lower STOL can be used to avoid or mitigate the false convergence. With $\text{STOL} = 10^{-12}$, the
simulations abort for $k_m = 10^{-3}$, 10^{-6} , and $10^{-9} \text{ mol m}^{-3}$ for the three sites due to too big an
update for the diagnostic variables (PlantN, DeniN, or N₂Od, Table 6). For each time step, these
745 concentrations start with 10^{-10} ; with an update as big as 10^{-6} , the concentration can be lowered in
several iterations using Eq. (20) to be below 10^{-10} times the update, resulting too small a scaling
factor. Removing the contribution of the derivatives to entries for these diagnostic species in the
Jacobian can mitigate this stability and accuracy issue but may slow down the convergence or cause
numerical errors. Increasing the reset value from 10^{-10} can mitigate the issue. However, as long as
750 there is a very low concentration for any species, there is a possibility for this issue to occur.

Instead of aborting the run when the scaling factor is less than the minimum allowable value λ_{min} , an alternative is to consider the iteration diverges and return to the time stepping subroutine to cut time step size to increase accuracy. This mitigates the stability and accuracy challenge at the expense of computing time. Our results show the default $\lambda_{min} = 10^{-10}$ can result in excessive error even
755 when STOL is extremely small. It is not clear how large a λ_{min} value can insure accuracy without introducing unnecessary computing time. For this reason, we caution against use of SU.

Similar to SU, the simulations using LT run to conclusion for $k_m = 10^{-3}, 10^{-6}$, and 10^{-9} mol m⁻³ but abort for $k_m = 10^{-12}$ for all of the three sites. The runs abort at 21.6, 2.46, and 1.1 year for the arctic, temperate, and tropical site because the number of time step cuts for one time step exceeds
760 MAX_CUT = 16 (default). Increasing MAX_CUT to 50, the simulations run to conclusion with longer computing time (Table 5). Looking into the reason for time step cuts, it usually involves iteration with NO₃⁻ concentration oscillation between the first order and zero order (Fig. 3b, for example, between 1.69×10^{-11} and 1.14×10^{-13} in the sixth layer in 0.5 year for BR-Cax site). While requiring an excessively large MAX_CUT, all of the simulations conclude with less than a 20% in-
765 crease of computing time for k_m of 10^{-12} over 10^{-9} . These results suggest that log transformation together with downregulation with reasonable half saturation and residual concentration can provide accurate, stable, and efficient solutions for CLM-PFLOTTRAN biogeochemistry simulations.

5 Summary and Conclusions

We implement CLM biogeochemistry, specifically, CLM-CN decomposition, nitrification, denitrification, and plant nitrogen uptake reactions in CLM-PFLOTTRAN. As CLM uses explicit time stepping (forward difference) and demand-based competition, the concentration is always nonnegative. PFLOTTRAN uses implicit time stepping (backward difference) and Newton-Raphson method; the concentration can become negative, which is not physical, may cause numerical instability, and introduce numerical error.
770

Both scaling back update in each iteration and log transformation can enforce nonnegativity. SU allows large time step sizes to achieve efficiency, but may introduce excessive numerical error if the scaling factor is small and the iterations are considered converged due to small updates. Log transformation involves multiplication of the Jacobian matrix by the concentration vector, decreasing the condition number by orders of magnitude as the concentration can be low. As a result, LT often
775 decreases the time step size and increases computational cost. Neither SU nor LT prevents too small or 0 concentrations. When the concentration becomes too small and essentially 0, SU may stop all reactions by a small positivity update caused by consumption, numerical overshoot, or truncation error and LT can fail due to too stiff or singular Jacobian matrix. Both SU and LT require that the limitation of reactant availability on reaction rates to be represented such that solution to the
780 mathematical representation does not introduce negative or 0 concentrations.
785

The first-order rate accounts for limitation of the reactant availability on the reaction rate, and the solution is nonnegative. Adding a residual concentration makes it positive. For the zero-order rate (or when the rate is not a function of a reactant), the Monod substrate limiting function provides a smooth transition from a zero order when the reactant is abundant to a first order when the
790 reactant becomes limiting. The downregulation relaxes with decreasing half saturation and residual concentration and disappears when both are 0.

Our CLM-PFLOTRAN spin-up simulation at an arctic, temperate, and tropical site indicates that accurate and stable solution can be achieved with log transformation with two to three times the computing time of CLM4.5 for a range of half saturation values from 10^{-3} to 10^{-9} and a residual
795 concentration of 10^{-15} for nitrogen. With half saturation of 10^{-12} , the number of maximum time cuts has to be increased from the default 16 to accommodate for small time step sizes to resolve small concentration changes in the transition of zero and first-order rate. The computing time increase from the half saturation of 10^{-12} to 10^{-9} is less than 20%. As physical half saturation ranges from 10^{-5} to 10^{-6} M for nitrogen, and the detection limits are often above 10^{-9} M, log trans-
800 formation together with downregulation by reasonable half saturation and residual concentration is expected to provide an accurate, stable, and efficient solution for CLM-PFLOTRAN implementation for current CLM biogeochemistry. As more substrate limiting processes, such as labile C, P, O₂, and H₂, are implemented with lower half saturation and residual concentration, and more complicated rate relationship, the maximum allowable time step cut may need to be increased, and the computing
805 time is expected to increase.

6 Code availability

PFLOTRAN is an open-source software. It is distributed under the terms of the GNU Lesser General Public License as published by the Free Software Foundation either version 2.1 of the License, or any later version. It is available at <https://bitbucket.org/pfotran>. Revision xxx is used for this work.
810 CLM-PFLOTRAN is Revision xxx is used for this work.

Acknowledgements. This research was funded by the U.S. Department of Energy, Office of Sciences, Biological and Environmental Research, Terrestrial Ecosystem Sciences and Subsurface Biogeochemical Research Program, and is a product of the Next-Generation Ecosystem Experiments in the Arctic (NGEE-Arctic) project. ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-
815 00OR22725.

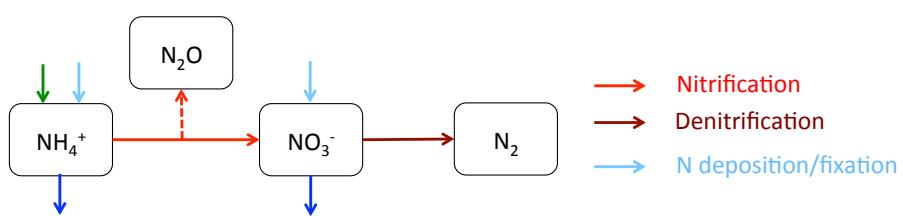
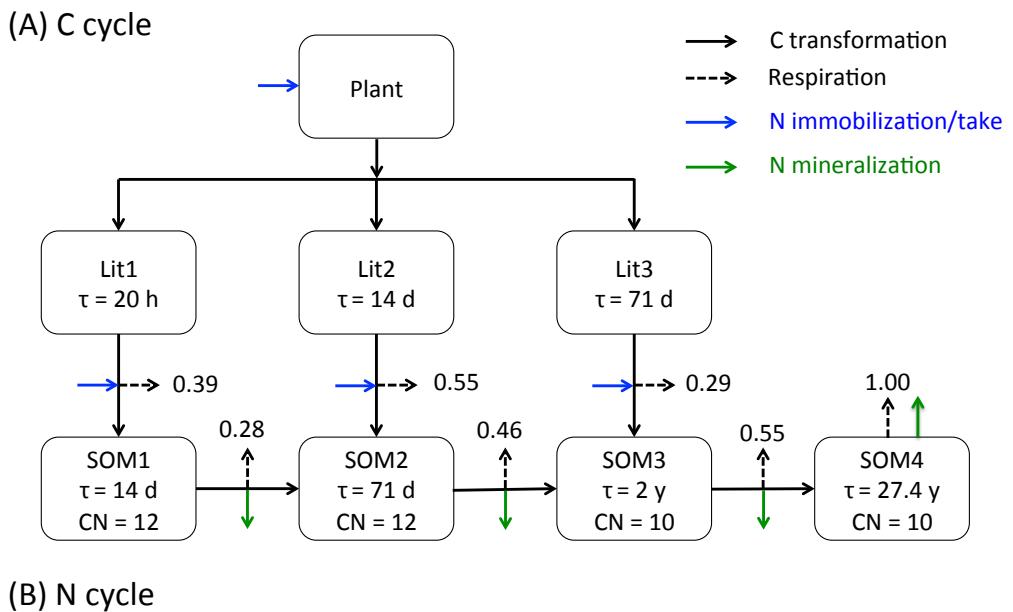


Figure 1. The reaction network for the carbon and nitrogen cycles implemented in this work. The carbon cycle (A) is modified from ? and ?

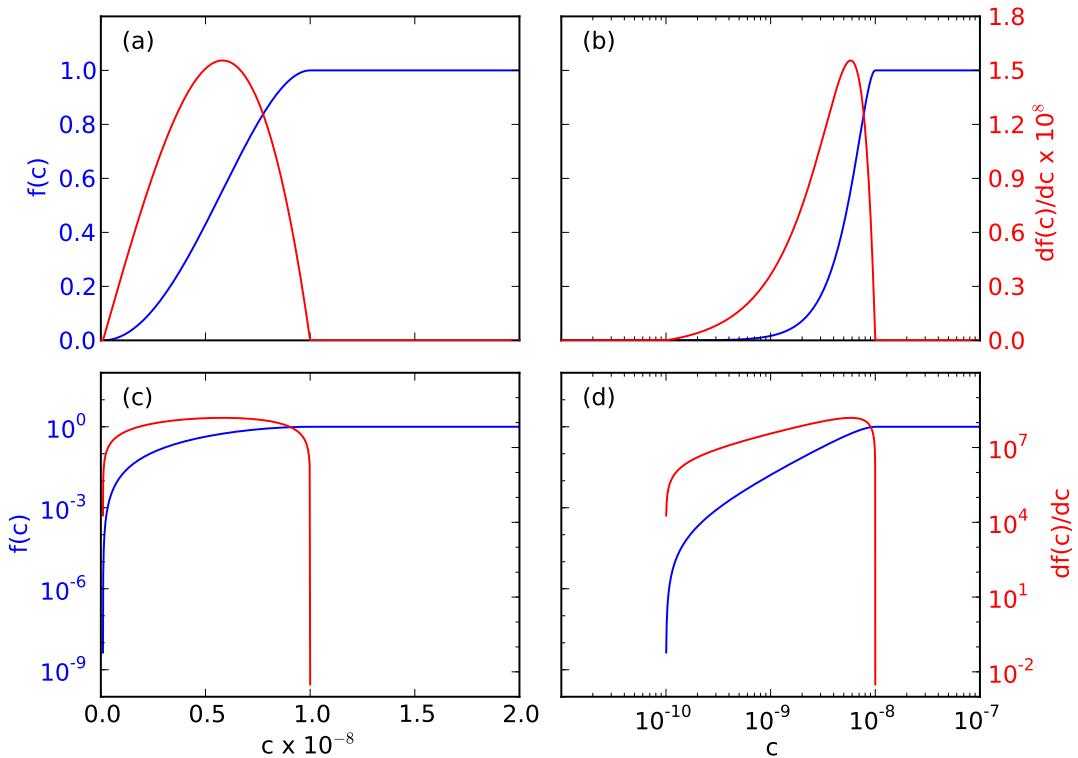


Figure 2. Smoothed cutoff function (Eq. 28) and derivative (Eq. 29). The y axis is in log in (c) and (d), while the x axis is in log in (b) and (d). Even though smoothed, the cutoff is a steep transition. The derivative varies by orders of magnitude in a small concentration range and requires small time step size to march through.

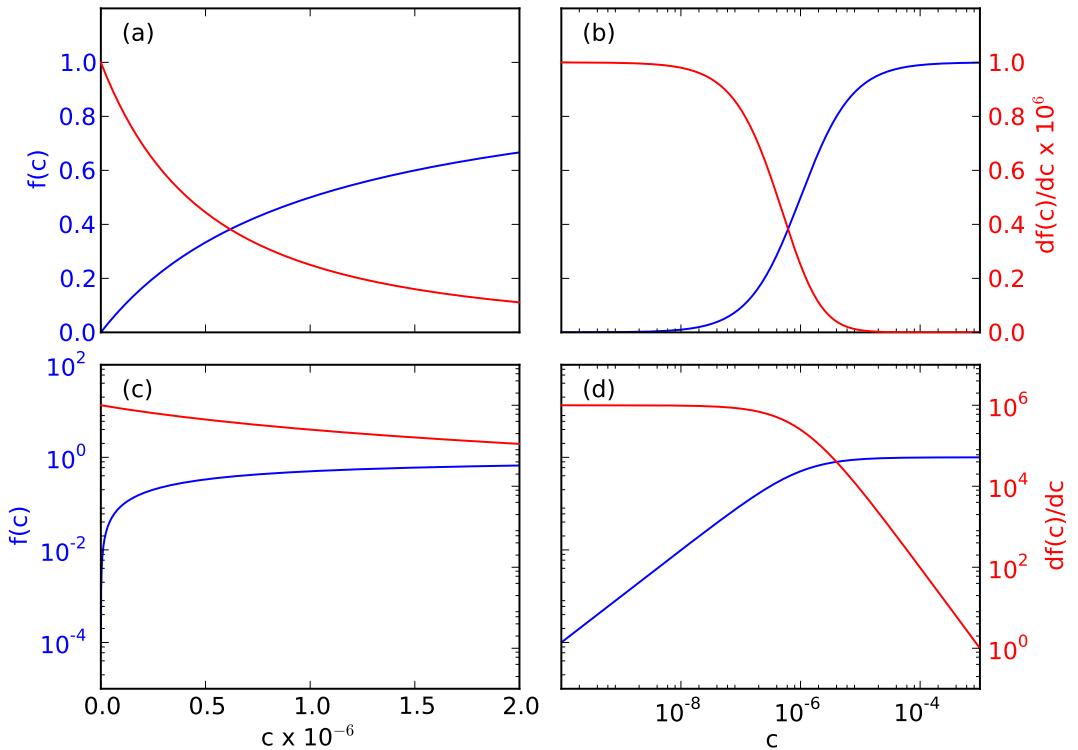


Figure 3. Monod substrate limiting function (Eq. 31) and derivative (Eq. 32). The y axis is in log in (c) and (d), while the x axis is in log in (b) and (d). As the function switches from zero-order to first-order, the derivative jumps six ($k_t \text{ext} m^{-1}$) orders of magnitudes, requiring small time step size to step through.

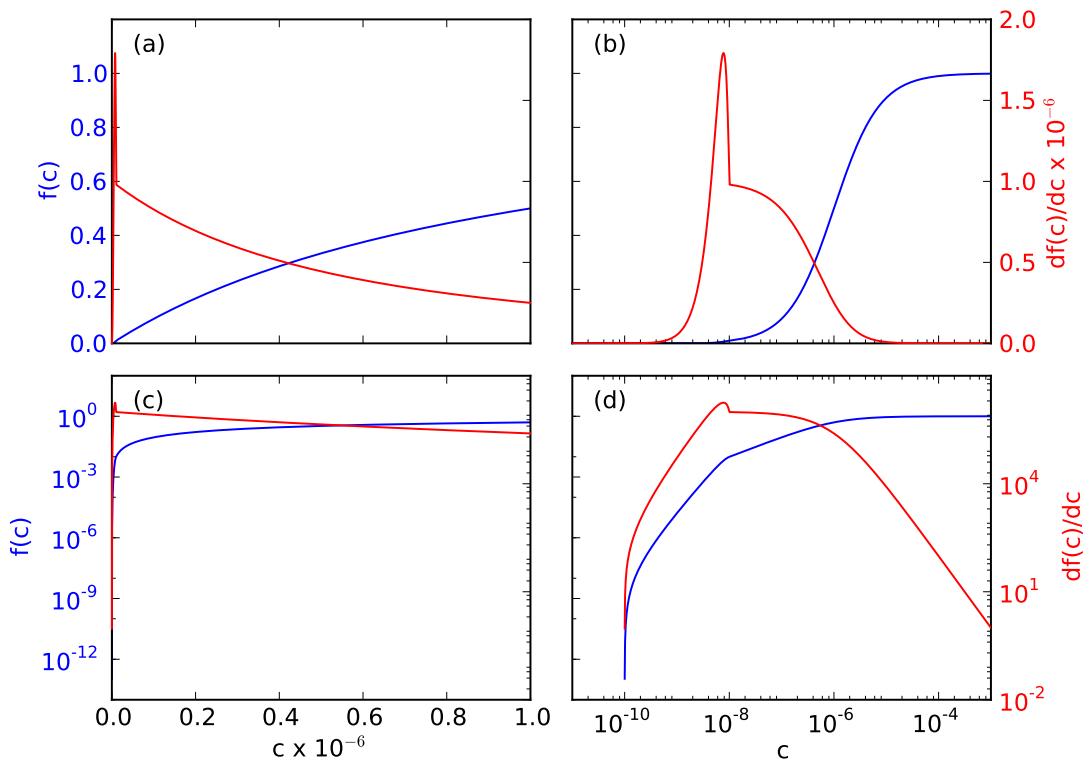


Figure 4. A combination of the Monod substrate limiting function (Eqs. 31 and 32, Fig. 3) and the smoothed cutoff function (Eqs. 28 and 29, Fig. 2) introduces steep transitions that require small time step sizes to march through.

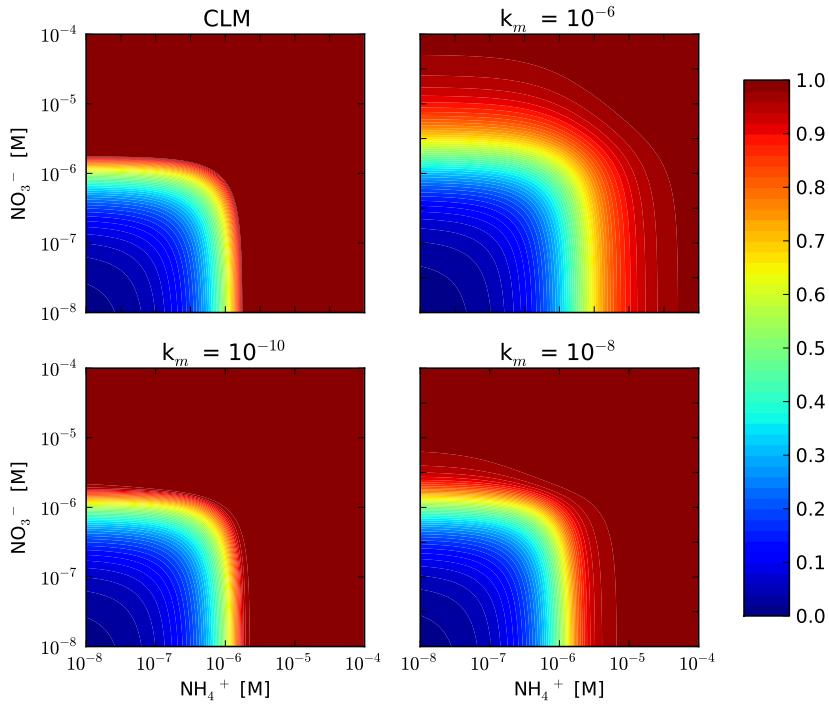


Figure 5. The ratio of uptake and demand (f_{pi} in Eq. 36) as a function of concentrations with CLM and representation by Eqs. (34 and 35) in a 0.5 h time step with an uptake rate of 10^{-9} M s^{-1} . f_{pi} for the new representation is less than or equal to that for CLM. The difference decreases with decreasing half saturation k_m .

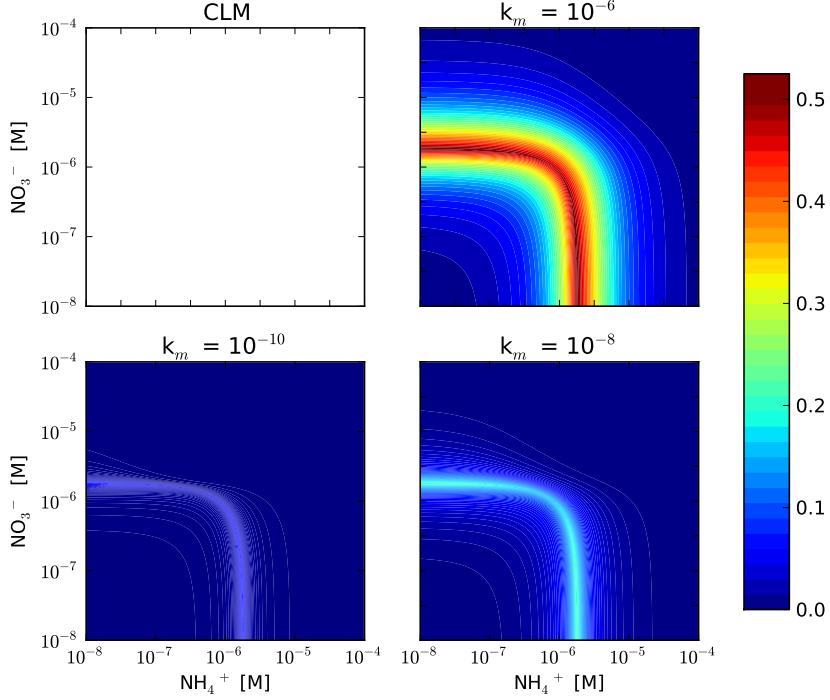


Figure 6. The difference plots for Fig. (5).

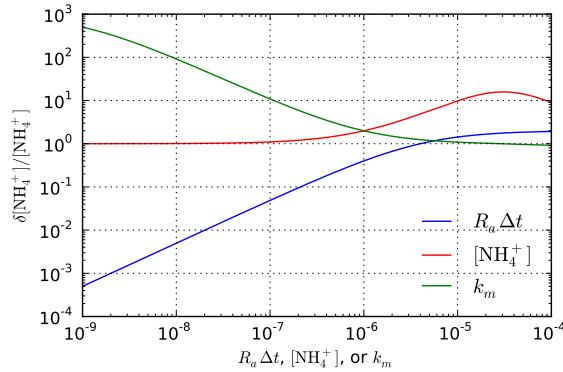


Figure 7. The calculated ratio of update over concentration for the first iteration as a function of time step size ($R_a \Delta t$), initial concentration (NH_4^+), and half saturation (k_m) for solving $d[\text{NH}_4^+]/dt = -R_a [\text{NH}_4^+]/(\text{NH}_4^+ + k_m)$ using the backward difference and Newton-Raphson method. Without scaling back the update (δ), the concentration in the first iteration step ($[\text{NH}_4^+]^{k+1,1}$) can go negative if $R_a \Delta t > 10^{-5}$, $[\text{NH}_4^+]^k > 10^{-6}$, or $k_m < 10^{-6}$. Unless specified in the figure, the base case parameters are $R_a \Delta t = 10^{-3}$, $\text{NH}_4^+ = 10^{-6}$, and $k_m = 10^{-6}$.

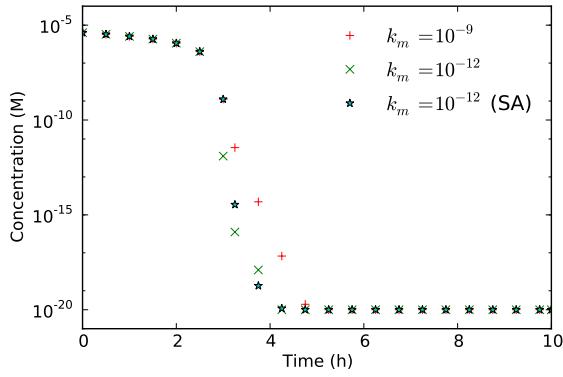


Figure 8. Influence of half saturation (k_m) on the simulated NH_4^+ concentration decrease due to plant uptake. Too small a half saturation may result in false convergence due to too small an update (scaling factor) during the iteration. The semi-analytical (SA) solution (Eq. 41) is used for comparison.

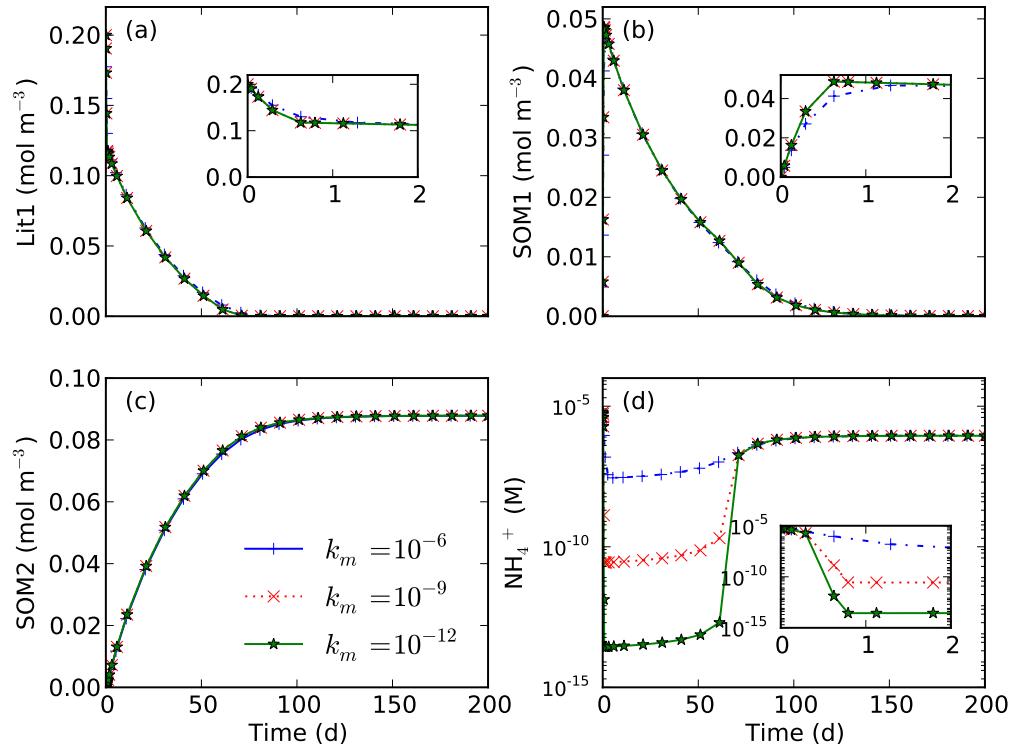


Figure 9. Influence of half saturation k_m on decomposition that involves both nitrogen immobilization and mineralization. Smaller half saturation can result in lower nitrogen concentration (d) but does not substantially impact the calculated concentrations other than NH_4^+ (a,b,c).

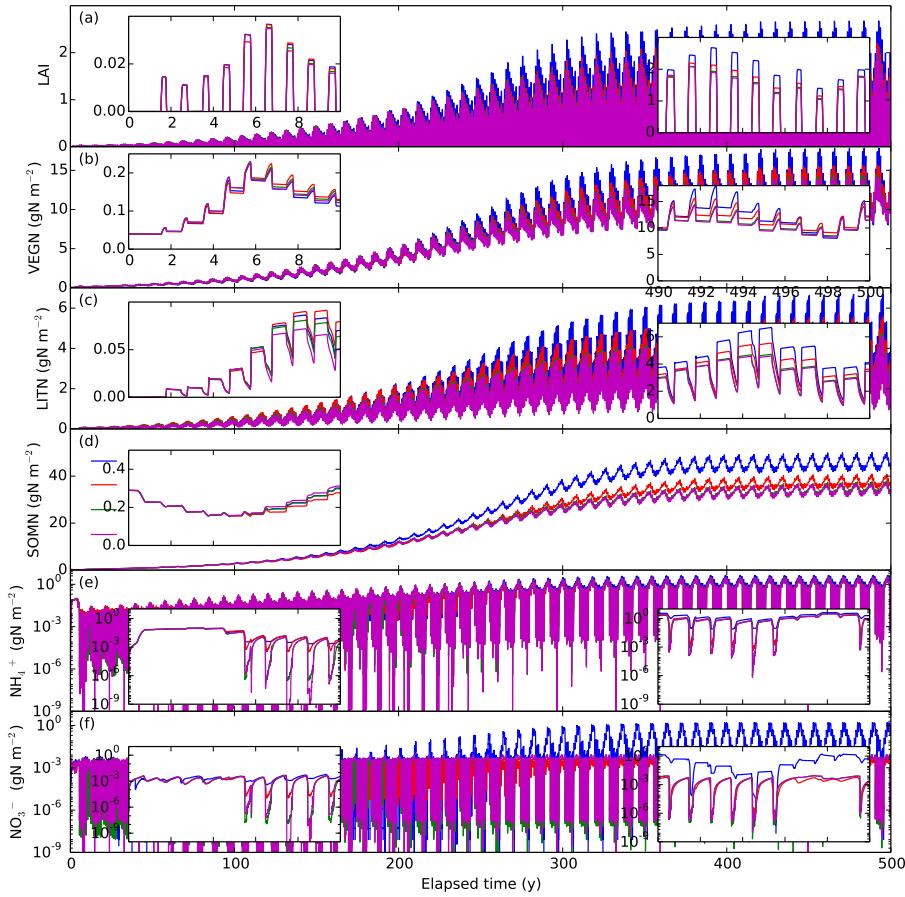


Figure 10. Calculated LAI and nitrogen distribution among vegetation, litter, SOM, NH_4^+ , and NO_3^- pools in spin-up simulations for the US-Brw site. Log transformation is used to enforce nonnegativity for CLM-PFLOTTRAN.

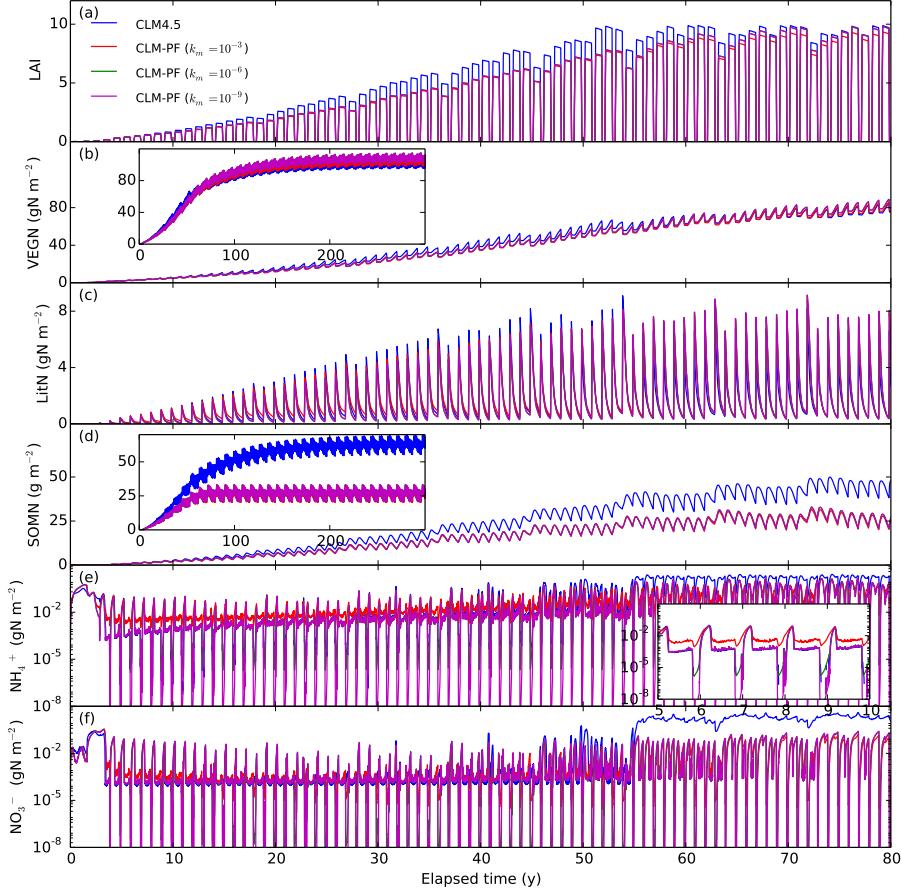


Figure 11. Calculated LAI and nitrogen distribution among vegetation, litter, SOM, NH_4^+ , and NO_3^- pools in spin-up simulations for US-WBW site. Log transformation is used to enforce nonnegativity for CLM-PFLTRAN calculations.

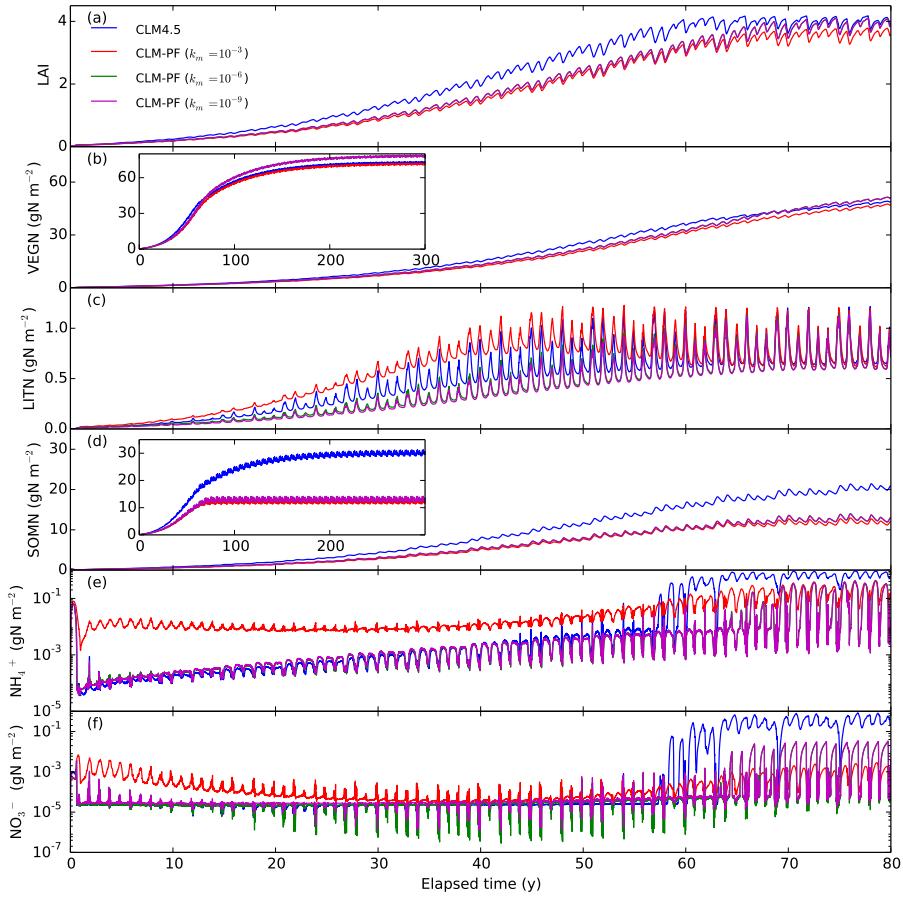


Figure 12. Calculated LAI and nitrogen distribution among vegetation, litter, SOM, NH_4^+ , and NO_3^- pools in spin-up simulations for BR-Cax site. Log transformation is used to enforce nonnegativity for CLM-PFLOTRAN calculations.

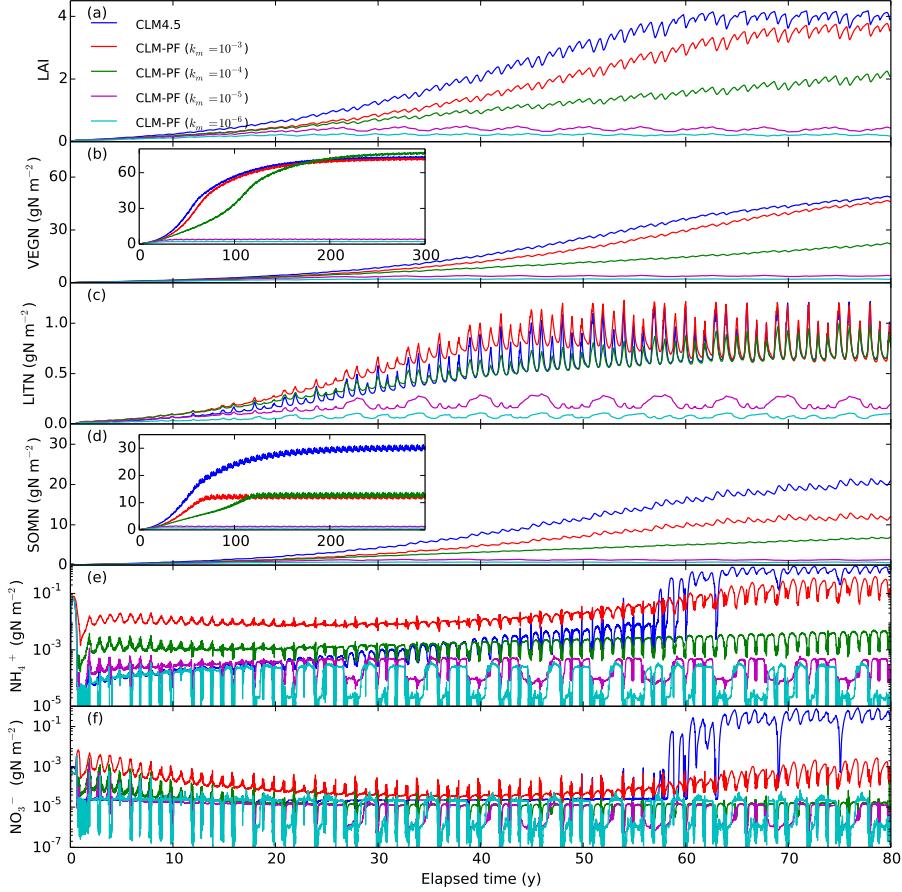


Figure 13. Calculated LAI and nitrogen distribution among vegetation, litter, SOM, NH_4^+ , and NO_3^- pools in spin-up simulations for BR-Cax site. Scaling back update in each iteration is used enforce nonnegativity for CLM-PFLOTTRAN.

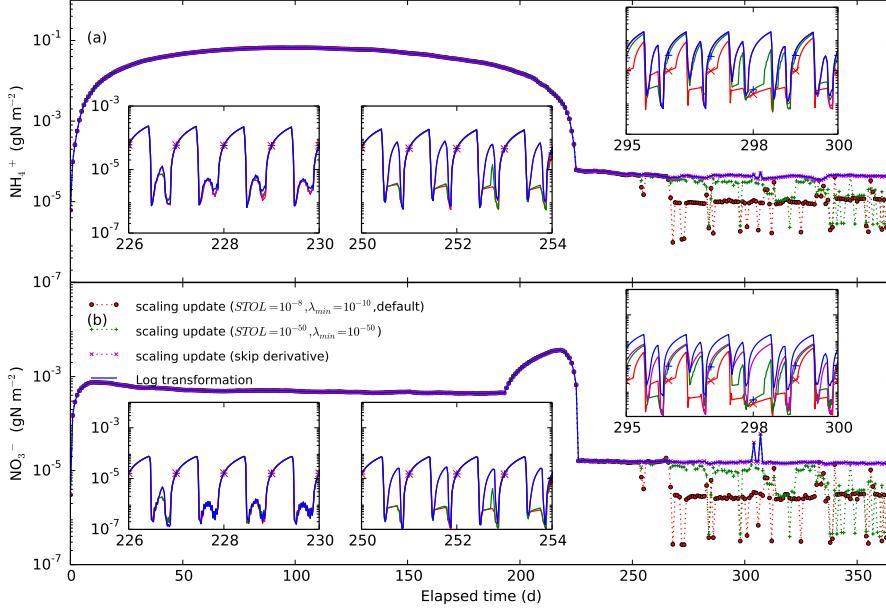


Figure 14. Calculated NH_4^+ and NO_3^- concentration in the first year using SU (scaling update) vs. LT (log transformation) in the spin-up simulation for BR-Cax site with $k_m = 10^{-6} \text{ mol m}^{-3}$. “skip derivative” refers to not including the derivatives for the reaction (R11) with rate (Eq. 37) in the entries for N_2Od in the Jacobian matrix.

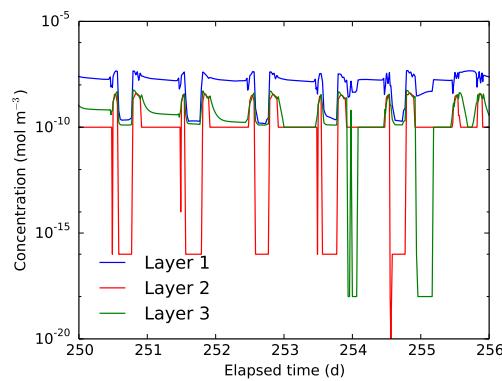


Figure 15. Diagnostic N_2O concentration (N_2Od from nitrification associated with net nitrogen mineralization Reaction R11 and rate Eq. 4) in the spin-up simulation for the BR-Cax site with $k_m = 10^{-6} \text{ mol m}^{-3}$. The concentration is reset to 10^{-10} at the beginning of each CLM half hour time step. Scaling back update in each iteration is used to enforce nonnegativity.

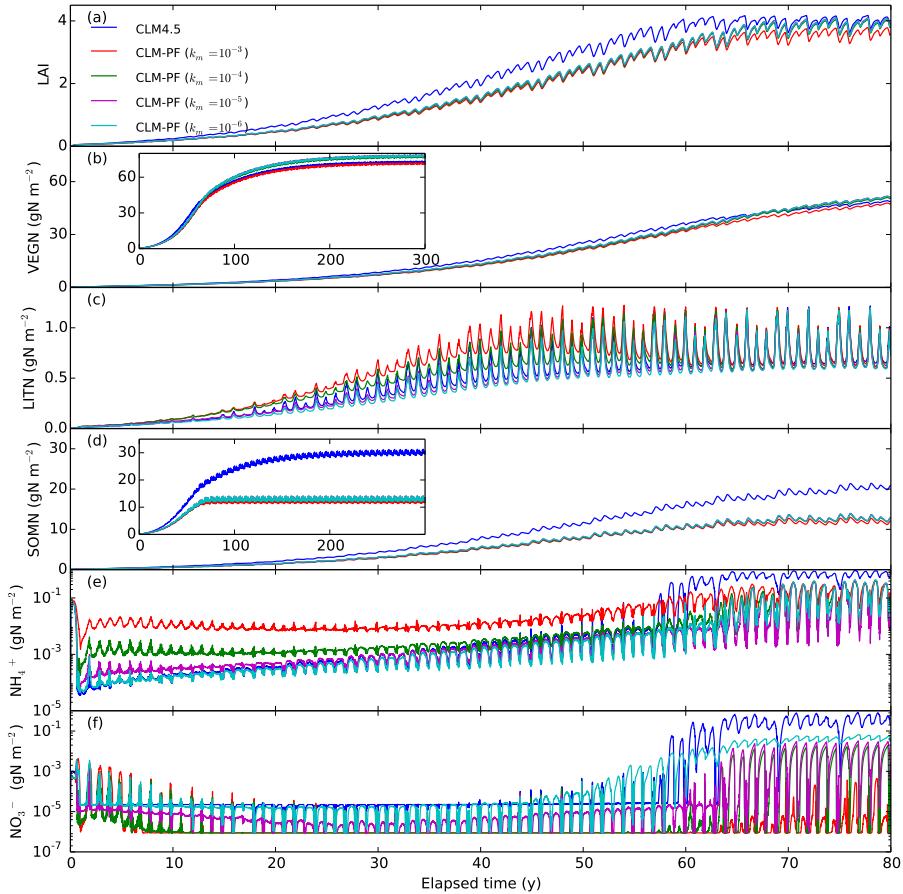


Figure 16. Calculated LAI and nitrogen distribution among vegetation, litter, SOM, NH_4^+ , and NO_3^- pools in spin-up simulations for BR-Cax site using the CLM demand based competition for consumption downregulation, and log transformation to enforce nonnegativity.

Table 1. Iterations in solving $dc/dt = -R_a c/(c+k_m)$ using backward difference and Newton Raphson method for one time step $R_a \Delta t = 0.001$ with initial $c = 10^{-3}$ and $k_m = 10^{-6}$

Iteration	c	$c/(k_m+c)$	$s/(c+k_m)^2$	Residual	Jacobian	δ
1	0.001	9.99E-01	9.98E-01	9.99E-01	1.00E+03	0.000998005
2	1.99501E-06	6.66E-01	1.11E+05	-3.32E-01	1.12E+05	-2.95065E-06
3	4.94566E-06	8.32E-01	2.83E+04	-1.63E-01	2.93E+04	-5.57379E-06
4	1.05195E-05	9.13E-01	7.54E+03	-7.63E-02	8.54E+03	-8.93755E-06
5	1.9457E-05	9.51E-01	2.39E+03	-2.94E-02	3.39E+03	-8.68139E-06
6	2.81384E-05	9.66E-01	1.18E+03	-6.18E-03	2.18E+03	-2.838E-06
7	3.09764E-05	9.69E-01	9.78E+02	-2.97E-04	1.98E+03	-1.49981E-07
8	3.11264E-05	9.69E-01	9.69E+02	-6.85E-07	1.97E+03	-3.47801E-10
9	3.11267E-05	9.69E-01	9.69E+02	-3.65E-12	1.97E+03	-1.85288E-15
10	3.11267E-05	9.69E-01	9.69E+02	0	1.97E+03	0

Table 2. Iterations in solving $dc/dt = -R_a c/(c+k_m)$ using backward difference and Newton Raphson method for one time step $R_a \Delta t = 0.002$ with initial $c = 10^{-3}$ and $k_m = 10^{-6}$ (without scaling back δ)

Iteration	c	$c/(k_m+c)$	$s/(c+k_m)^2$	Residual	Jacobian	δ
1	0.001	9.99E-01	9.98E-01	9.99E-01	5.01E+02	0.001994022
2	-0.000994022	1.00E+00	1.01E+00	4.00E-03	5.01E+02	7.97596E-06
3	-0.001001998	1.00E+00	9.98E-01	6.44E-08	5.01E+02	1.28641E-10
4	-0.001001998	1.00E+00	9.98E-01	0.00E+00	5.01E+02	0

Table 3. Iterations in solving $dc/dt = -R_a c/(c+k_m)$ using backward difference and Newton Raphson method for one time step $R_a \Delta t = 0.002$ with initial $c = 10^{-3}$, $k_m = 10^{-6}$, and $\alpha = 0.9999$ for scaling back δ

Iteration	c	$c/(k_m+c)$	$s/(c+k_m)^2$	Residual	Jacobian	δ	λ
1	0.001	9.99E-01	9.98E-01	9.99E-01	5.01E+02	0.001994022	0.5
2	1E-07	9.09E-02	8.26E+05	-4.09E-01	8.27E+05	-4.9464E-07	1
3	5.9464E-07	3.73E-01	3.93E+05	-1.27E-01	3.94E+05	-3.22036E-07	1
4	9.16676E-07	4.78E-01	2.72E+05	-2.13E-02	2.73E+05	-7.80252E-08	1
5	9.94701E-07	4.99E-01	2.51E+05	-8.31E-04	2.52E+05	-3.29904E-09	1
6	9.98001E-07	4.99E-01	2.51E+05	-1.37E-06	2.51E+05	-5.45442E-12	1
7	9.98006E-07	5.00E-01	2.50E+05	-3.73E-12	2.51E+05	-1.48611E-17	1
8	9.98006E-07	5.00E-01	2.50E+05	0	2.51E+05	0	1

Table 4. Iterations in solving $dc/dt = -R_a c/(c+k_m)$ using backward difference and Newton Raphson method for one time step $R_a \Delta t = 0.002$ with initial $c = 10^{-3}$ and $k_m = 10^{-6}$ and with log transformation

Iteration	c	$c/(k_m+c)$	$s/(c+k_m)^2$	Residual	Jacobian	δ
1	0.001	9.99E-01	9.98E-01	9.99E-01	5.01E-01	1.994021918
2	0.000136147	9.93E-01	5.32E+01	5.61E-01	7.53E-02	7.446148992
3	7.94668E-08	7.36E-02	8.58E+05	-4.26E-01	6.82E-02	-6.247977459
4	4.10817E-05	9.76E-01	5.65E+02	4.97E-01	4.37E-02	11.35765187
5	4.79825E-10	4.80E-04	9.99E+05	-5.00E-01	4.80E-04	-1041.52556

Table 5. Wall time (hour) for spin-up simulation at the arctic, temperate, and tropical sites on OIC

Site	CLM	SU (10^{-3})	SU (10^{-6})	SU (10^{-9})	LT (10^{-3})	LT (10^{-6})	LT (10^{-9})	LT (10^{-12})
DC								
US-Brw	18.1	24.8	25.5	29.2	38.5	40.8	47.0	49.5
US-Pit	11.7	17.1	14.8	14.9	30.0	37.5	40.2	43.1
BR-Cax	10.9	16.2	18.6	18.0	40.2	40.5	45.7	52.0
DR								
US-Brw	18.1	21.5	21.5		35.8	35.9		
US-Pit	11.7	14.1	14.1		26.9	26.7		
BR-Cax	10.9	16.4	16.4		42.2	41.6		

SU = scaling update, LT = log transformation, 10^{-3} , 10^{-6} , and 10^{-9} are k_m , DC and DR = downregulating consumption as a function of concentration and rate, OIC = ORNL Institutional Cluster (Phase5), US-Brw simulation duration = 500 year and US-Pit and BR-Cax simulation duration = 300 year. For the last column LT (10^{-12}), MAX_CUT is increased from default 16 to 50.

Table 6. CLM-PFLOTRAN simulation termination before conclusion using scaling back update in iteration

Site	k_m	year	species	concentration	update	layer
STOL = 10^{-8}						
US-Brw	10^{-12}	346.66	NO_3^-	1.26×10^{-26}	8.20×10^{-15}	4
US-Pit	10^{-12}	3.56	NO_3^-	2.67×10^{-26}	2.49×10^{-14}	8
BR-Cax	10^{-12}	3.18	NO_3^-	2.54×10^{-26}	5.72×10^{-14}	8
STOL = 10^{-12}						
US-Brw	10^{-6}	41.78	DeniN	10^{-20}	2.31×10^{-10}	2
US-Brw	10^{-9}	5.56	PlantN	6.68×10^{-18}	8.12×10^{-8}	2
US-PIT	10^{-3}	7.81	N_2Od	10^{-20}	1.74×10^{-9}	1
US-PIT	10^{-6}	2.40	PlantN	10^{-18}	2.30×10^{-6}	1
BR-Cax	10^{-3}	1.59	N_2Od	10^{-20}	1.66×10^{-10}	1
BR-Cax	10^{-6}	0.64	N_2Od	10^{-20}	1.91×10^{-10}	2
BR-Cax	10^{-9}	0.59	PlantN	10^{-18}	4.92×10^{-8}	2

Appendix A: Downregulation of consumption as a function of rates

The contribution of n reactions to the rate component in the residual and Jacobian for species i are

$$\mathbf{R}(i) = \sum_{j=1}^n \mu_{ij} R_j, \quad (\text{A1})$$

and

$$820 \quad \mathbf{J}(i, k) = \frac{\partial \mathbf{R}(i)}{\partial \mathbf{C}(k)} = \sum_{j=1}^n \frac{\partial(\mu_{ij} R_j)}{\partial \mathbf{C}(k)}, \quad (\text{A2})$$

with \mathbf{C} as the concentration vector, \mathbf{R} as the rate vector, \mathbf{J} as the component of Jacobian matrix (m by m) that is associated with \mathbf{R} , μ_{ij} as stoichiometric coefficient of species i in reaction j ($\sum_{i=1}^m \mu_{ij} \mathbf{C}(i) = 0$), and R_j as the rate of reaction j , a function of the activities (concentration) of the reactants, products, and environmental variables (moisture, temperature, pH, redox, etc.).

825 A1 Downregulation of consumption for one species

Depending on whether a species a (e.g., NH_4^+) is a reactant ($\mu_{aj} > 0$) or a product ($\mu_{aj} < 0$), we divide the n reactions into a demand (subscript da for demand a) and a supply (subscript sa for supply a) group:

$$R_{da} = \sum_{j=1}^{n_{da}} \mu_{da,aj} R_{da,j}, \quad (\text{A3})$$

830 and

$$R_{sa} = \sum_{j=1}^{n_{sa}} \mu_{sa,aj} R_{sa,j}, \quad (\text{A4})$$

with n_{da} as the number of reactions that consume species a , $\mu_{da,aj}$ as the stoichiometric coefficient of species i in a consuming reaction j , n_{sa} as the number of reactions that produce species a , $\mu_{sa,ij}$ as the stoichiometric coefficient of species i in a producing reaction j , R_{da} as a consumption (demand) rate (mol s^{-1} , negative), and R_{sa} as a production (supply) rate (mol s^{-1} , positive).

835 Therefore,

$$\mathbf{R}(i) = \sum_{j=1}^{n_{sa}} \mu_{sa,aj} R_{sa,j} + \sum_{j=1}^{n_{da}} \mu_{da,aj} R_{da,j} = \mathbf{R}_{sa}(i) + \mathbf{R}_{da}(i), \quad (\text{A5})$$

and

$$\mathbf{J}(i, k) = \frac{\partial \mathbf{R}(i)}{\partial \mathbf{C}(k)} = \sum_{j=1}^n \frac{\partial(\mu_{ij} R_j)}{\partial \mathbf{C}(k)} = \sum_{j=1}^{n_{sa}} \frac{\partial(\mu_{sa,ij} R_{sa,j})}{\partial \mathbf{C}(k)} + \sum_{j=1}^{n_{da}} \frac{\partial(\mu_{da,ij} R_{da,j})}{\partial \mathbf{C}(k)} = \mathbf{J}_{sa} + \mathbf{J}_{da}. \quad (\text{A6})$$

840 We define a downregulation factor

$$d_a = \min \left(1, -\frac{R_{sa}\Delta t + [\mathbf{C}(a) - \epsilon]V}{R_{da}\Delta t} \right) = \min \left(1, -\frac{s_a}{D_a} \right), \quad (\text{A7})$$

with V as the bulk volume or volume of liquid water of the grid cell for species with concentration unit mol m^{-3} or M. After downregulation,

$$\mathbf{R}_a = \mathbf{R}_{sa} + d_a \mathbf{R}_{da}, \quad (\text{A8})$$

$$845 \quad \mathbf{J}_a(i, k) = \frac{\partial \mathbf{R}_a(i)}{\partial \mathbf{C}(k)} = \mathbf{J}_{sa}(i, k) + d_a \mathbf{J}_{da}(i, k) + \mathbf{R}_{da}(i) \frac{\partial d_a}{\partial \mathbf{C}(k)}, \quad (\text{A9})$$

$$\frac{\partial d_a}{\partial \mathbf{C}(k)} = - \left(\frac{\partial s_a}{\partial \mathbf{C}(k)} D_a - s_a \frac{\partial D_a}{\partial \mathbf{C}(k)} \right) D_a^{-2}, \quad (\text{A10})$$

$$\frac{\partial s_a}{\partial \mathbf{C}(a)} = \sum_{j=1}^{n_{sa}} \frac{\partial (\mu_{sa,aj} R_{sa,j})}{\partial \mathbf{C}(a)} \Delta t + V, \quad (\text{A11})$$

$$\frac{\partial s_a}{\partial \mathbf{C}(k)} = \sum_{j=1}^{n_{sa}} \frac{\partial (\mu_{sa,kj} R_{sa,j})}{\partial \mathbf{C}(k)} \Delta t, \quad (\text{A12})$$

$$\frac{\partial D_a}{\partial \mathbf{C}(k)} = \sum_{j=1}^{n_{da}} \frac{\partial (\mu_{da,kj} R_{da,j})}{\partial \mathbf{C}(k)} \Delta t, \quad (\text{A13})$$

850 Implementation in PFLOTRAN involves 1) adding variables \mathbf{R}_{sa} , \mathbf{R}_{da} , and \mathbf{J}_a , 2) accumulating the values in each reaction rate formula, and 3) conducting downregulation and adding the contribution to the global residual vector and Jacobian matrix.

A2 Downregulation of consumption for a second species

In addition to species a (e.g., NH_4^+), we want to downregulate another species l , for example, NO_3^- .

855 The treatment is the same except for the reactions that consume species a and produce species l . Suppose we have a nitrification reaction (R10) with rate R_{al} , and $R'_{al} = dR_{al}/d[\text{NH}_4^+]$. The rate and derivative are added in demand rate and derivative for NH_4^+ (\mathbf{R}_{sa} , \mathbf{R}_{da} , \mathbf{J}_{sa} , \mathbf{J}_{da}), not in the sink rate and derivative for NO_3^- (\mathbf{R}_{sl} , \mathbf{R}_{dl} , \mathbf{J}_{sl} , \mathbf{J}_{dl}). Define a downregulation factor

$$d_l = \min \left(1, -\frac{R_{sl}\Delta t + (\mathbf{C}(l) - \epsilon)V_w + \underline{R_{al}d_a\Delta t}}{R_{dl}\Delta t} \right) = \min(1, -\frac{s_l}{D_l}), \quad (\text{A14})$$

860 $\mathbf{R}_l = \mathbf{R}_{sl} + d_l \mathbf{R}_{dl},$ (A15)

$$\mathbf{J}_l(i, k) = \frac{\partial \mathbf{R}_l(i)}{\partial \mathbf{C}(k)} = \mathbf{J}_{sl}(i, k) + d_l \mathbf{J}_{dl}(i, k) + \mathbf{R}_{dl}(i) \frac{\partial d_l}{\partial \mathbf{C}(k)}, \quad (\text{A16})$$

$$\frac{\partial d_l}{\partial \mathbf{C}(k)} = - \left(\frac{\partial s_l}{\partial \mathbf{C}(k)} D_l - s_l \frac{\partial D_l}{\partial \mathbf{C}(k)} \right) D_l^{-2}, \quad (\text{A17})$$

$$\frac{\partial s_l}{\partial \mathbf{C}(l)} = \sum_{j=1}^{n_{sl}} \frac{\partial (\mu_{sl,lj} R_{sl,j})}{\partial \mathbf{C}(l)} \Delta t + V_w + \underline{\frac{\partial (R_{al} d_a)}{\partial \mathbf{C}(l)}}, \quad (\text{A18})$$

$$\frac{\partial s_l}{\partial \mathbf{C}(k)} = \sum_{j=1}^{n_{sl}} \frac{\partial (\mu_{sl,kj} R_{sl,j})}{\partial \mathbf{C}(k)} \Delta t + \underline{\frac{\partial (R_{al} d_a)}{\partial \mathbf{C}(K)}}, \quad (\text{A19})$$

865 and

$$\frac{\partial D_a}{\partial \mathbf{C}(k)} = \sum_{j=1}^{n_{da}} \frac{\partial (\mu_{da,kj} R_{da,j})}{\partial \mathbf{C}(k)} \Delta t. \quad (\text{A20})$$

In addition to variables \mathbf{R}_{sl} , \mathbf{R}_{dl} , and \mathbf{J}_l , downregulating the second species requires two additional variables, R_{al} , and R'_{al} . When accumulating the values in each reaction rate formula, the production rate from the reaction that consumes the first species has to be carefully treated as it is
870 downregulated for the first species. Conducting downregulation involves additional terms, as underlined in the equations. Compared with the downregulation for the first species, downregulating the second species that can be produced from the first species with a simple $A \rightarrow B$ reaction becomes much more complicated. If we add another reaction that consumes the second species to produce the first species, this approach has to be modified.

875 In general, many if not all species need to be downregulated. Extending this approach involves adding many variables (vectors and matrices) to track the consumption and production rates and their derivatives. The consumption and production relationship among many species can be complicated, making generalization of this approach challenging if not infeasible.