JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003625

Matrix Approach to Accelerate Spin-Up of CLM5

Cuijuan Liao[1], Xingjie Lu[2], Yuanyuan Huang[3], Feng Tao[1], David M. Lawrence[4], Charles D. Koven[5], Keith W. Oleson[4], William R. Wieder[4], Erik Kluzek[4], Xiaomeng Huang[1], and Yiqi Luo[6]

[1]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China, [2]School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China, [3]Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, [4]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [5]Earth and Environmental Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [6]School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

Abstract Numerical models have been developed to investigate and understand responses of biogeochemical cycle to global changes. Steady state, when a system is in dynamic equilibrium, is generally required to initialize these model simulations. However, the spin-up process that is used to achieve steady state pose a great burden to computational resources, limiting the efficiency of global modeling analysis on biogeochemical cycles. This study introduces a new Semi-Analytical Spin-Up (SASU) to tackle this grand challenge. We applied SASU to Community Land Model version 5 and examined its computational efficiency and accuracy. At the Brazil site, SASU is computationally 7 times more efficient than (or saved up to 86% computational cost in comparison with) the traditional native dynamics (ND) spin-up to reach the same steady state. Globally, SASU is computationally 8 times more efficient than the accelerated decomposition spin-up and 50 times more efficient than ND. In summary, SASU achieves the highest computational efficiency for spin-up on site and globally in comparison with other spin-up methods. It is generalizable to wide biogeochemical models and thus makes computationally costly studies (e.g., parameter perturbation ensemble analysis and data assimilation) possible for a better understanding of biogeochemical cycle under climate change.

Plain Language Summary Land carbon cycle models, a critical component of earth system models, have been developed to investigate and understand ecological responses to global changes. However, the model spin-up process to achieve steady state are usually time-consuming and pose great computational burden. The spin-up problem hinders our ability to study some key issues in land carbon cycle modeling, such as sensitivity analysis. This paper applies a new SASU framework to accelerate spin-up of Community Land Model matrix version 5. Compared with traditional native dynamic approach, SASU speeds up the spin-up by approximately 50 times, achieving the fastest spin-up speed than all the previously recorded methods. Our SASU method is applicable to biogeochemical models and, therefore, liberates the models for computationally costly studies, such as parameter perturbation ensemble analysis and data assimilation.

Key Points:

- A new semi-analytical spin-up (SASU) framework combines the default accelerated spin-up method and matrix analytical algorithm
- SASU accelerates CLIM5 spin-up by tens of times, becoming the fastest method to our knowledge
- SASU is applicable to most biogeochemical models and enables computationally costly study, for example, sensitivity analysis

Correspondence to:
X. Huang and Y. Luo,
hxm@tsinghua.edu.cn;
yl2735@cornell.edu

Citation:
Liao, C., Lu, X., Huang, Y., Tao, F., Lawrence, D. M., Koven, C. D., et al. (2023). Matrix approach to accelerate spin-up of CLM5. *Journal of Advances in Modeling Earth Systems*, *15*, e2023MS003625. https://doi.org/10.1029/2023MS003625

Received 1 FEB 2023
Accepted 28 JUL 2023

Author Contributions:
Conceptualization: Cuijuan Liao
Software: Cuijuan Liao
Validation: Cuijuan Liao
Writing – original draft: Cuijuan Liao
Writing – review & editing: Cuijuan Liao

1. Introduction

Terrestrial biogeochemical models usually contain several modules (e.g., photosynthesis, litter fall, carbon decomposition) by defining various state variables representing carbon and nitrogen pools to simulate the feedback of land ecosystem to global change over the past decades (Ahlström et al., 2012; Lehmann & Kleber, 2015; Wieder et al., 2018). Initialization of these state variables is essential for a transient model simulation, especially in the study of carbon-climate feedback. Observation of ecosystem carbon and nitrogen storage provides reliable information for model initialization (Fang et al., 2014; Murty & McMurtrie, 2000; Zhang et al., 2002). However, these observations are not commonly available, especially at global scale. Typically, a transient biogeochemical model simulations are initialized by a spin-up run. During spin-up, an arbitrary initial condition is given and the models are driven under recursive external meteorological forcing until an approximate equilibrium is reached for carbon and nitrogen. The computational cost to achieve the steady state is generally high due to the slow turnover rates of the soil carbon pools. Such computation cost would further surge with integrating more biogeochemical processes and adopting higher resolutions in the simulation of terrestrial models. This is known as the "spin-up problem" (Thornton & Rosenbloom, 2005). This spin-up problem become a computational bottleneck

that hampers important yet time-consuming analysis of biogeochemical models (e.g., parameter perturbation experiment (PPE) and model-data fusion) (Sun et al., 2023).

To resolve the "spin-up problem," a variety of methods have been proposed over the past two decades to improve the computational spin-up efficiency. The simplest method is native dynamics spin-up (ND). ND takes hundreds to thousands of years of model simulation, running under recursive external forcing (Bondeau et al., 2007; Randerson et al., 2009). The accelerated decomposition (AD) spin-up approach was developed by temporarily assuming higher decomposition rates (Thornton & Rosenbloom, 2005) and has been successfully applied to both the Biome-BGC and Community Land Model (CLM) to reduce 70% spin-up time (Fang et al., 2015; Koven et al., 2013; Shi et al., 2013; Thornton & Rosenbloom, 2005). It should be noted that both AD approaches require an additional period of ND simulation (the post-AD spin-up process) to further adjust the quasi-steady-state from the accelerated mode to a real steady-state, which is sometimes accompanied by a great computing burden. Some models are initiated with observations of plants, litter and soil in previous regional experiments (Fang et al., 2014; Murty & McMurtrie, 2000; Zhang et al., 2002). However, the model still take a long time to come to equilibrium so that this method may not practically improve efficiency (Shi et al., 2013; Wutzler & Reichstein, 2007).

In addition, a suite of approaches has been proposed based on the formulated carbon balance equation to solve the steady state in an analytical or semi-analytical manner. The approach works because most of the current generation of terrestrial carbon and nitrogen cycle models use the first-order kinetics to describe terrestrial carbon and nitrogen dynamics (Adair et al., 2008; Koven et al., 2013; Shi et al., 2018) while some of the nonlinear models also can be analytically solved to obtain steady state (Georgiou et al., 2017). Full analytical spin-up method such as gradient projection approach (Fang et al., 2015), numerically solves matrix-based expressions via the Jacobian matrix or Gauss-Jordan elimination algorithms (Lardy et al., 2011; Martin et al., 2007; Qu et al., 2018), which could save up to 85%–90% of the computational cost. However, the fully analytical method requires a series of complicated mathematical operations and currently it is only tested at a few sites (Fang et al., 2015; Lardy et al., 2011) and North America (Qu et al., 2018). Based on the matrix representation of biogeochemical models in the terrestrial ecosystems (Luo et al., 2003, 2017), Xia et al. (2012) first developed a semi-analytical spin-up (SASU) using the Australian Community Atmosphere Biosphere Land Exchange (CABLE) model. SASU offers great potential to effectively reduce the spin-up time of global models, which saves up to 92.4% and 86.6% of computational time in CABLE (Xia et al., 2012).

As more ecological processes are integrated, increasing complexity of current biogeochemical model structures further introduce substantial difficulties to achieve efficient spin-up. For examples, in recognition of the importance of vertical distribution and exchange of soil organic matter (SOM) (Baisden & Parfitt, 2007; Balesdent et al., 2018; Jobbagy & Jackson, 2000), vertically resolved SOM structure has been implemented in land models (Braakhekke et al., 2011; Jenkinson et al., 2008). In CLM Version 5 (CLM5), Koven et al. revised the vertically resolved structure (Koven et al., 2013), which can simulate dynamic changes of soil carbon up to 8.6 m deep (Lawrence et al., 2019). In addition, different from fixed carbon-nitrogen ratios in most of the current carbon-nitrogen coupled models (Clark et al., 2011; Oleson et al., 2010; Potter et al., 1993; Running & Hunt, 1993), CLM5 improved global carbon-nitrogen coupling simulation by defining flexible carbon-nitrogen ratios and adding a Fixation and Uptake of Nitrogen (FUN) module (Ghimire et al., 2015; Lawrence et al., 2019). All of the above advancements in representing biogeochemical processes lead to relatively longer carbon turnover times and more complex carbon-nitrogen interactions than its predecessors (Koven et al., 2015), which bring greater computational burden to the spin-up. Implementing SASU to such a complicated biogeochemical model like CLM will benefit future biogeochemical model development (Fang et al., 2015; Qu et al., 2018).

In this study, we propose a new SASU framework by combining the AD method and semi-analytical algorithm based on the CLM5 matrix version (Lu et al., 2020). We found that this new SASU framework further accelerates spin-up of global carbon–nitrogen coupled models over that introduced by Xia et al. (2012). Computational efficiency and steady-state consistency of ND, AD, and SASU approaches are evaluated using CLM5 at both one-site and the global scale. In addition, we applied SASU to a PPE using CLM5, which is designed to examine the parameter sensitivity and provide a pathway for systematic parameter calibration. Our results provide strong support for the applications of SASU to complicated biogeochemical models with vertically resolved structure and flexible carbon to nitrogen ratio representations.

2. Materials and Methods

2.1. Model Description

This study is based on CLM5, which is widely used and incorporate comprehensive biogeochemical processes. Similar to other land surface models, CLM5 represents the terrestrial carbon cycle via multiple carbon and nitrogen pools. The dynamics of pools can be characterized by the difference between carbon input and output (Figure 1). Terrestrial organic carbon is originally produced by photosynthesis of plants and distributed to six main pools (leaf, fine root, live stem, dead stem, live coarse root, dead coarse root). Vegetation carbon pools are regulated by multiple processes such as phenology, nutrient constraints, fire, etc. Subsequently, carbon is directly transferred from the vegetation or indirectly through the Coarse Woody Debris (CWD) pools to three different litter pools, including labile, cellulose and lignin. Downstream of the litter pools are three SOM pools with fast, slow, and passive decomposition. To simulate advection and diffusion of organic matter along a soil vertical profile, CLM5 divides soil into 20 layers that are up to 8.5 m along depth (Lawrence et al., 2019). The carbon turnover time varies in different pools, which depends on the transfer coefficients regulated by the availability of soil nutrients and environmental factors (such as temperature and moisture). Limited by depth and environmental factors, SOM in deep layers typically has little carbon input and low decomposition rates.

In addition to having the same structure as the carbon cycle described above, the nitrogen cycle in CLM5 also regulates the availability of mineral nitrogen in the soil through immobilization and denitrification, which will constrain the photosynthesis of plants to form a relatively complete carbon-nitrogen coupling cycle. In particular, compared to CLM4, CLM5 has been updated in nutrient dynamics to model the nitrogen limitation in plant growth. In CLM4, denitrification was handled through a time constant applied to a bulk soil mineral N pool, and static carbon-nitrogen ratio is used to indicate that the availability of soil mineral nitrogen limits potential photosynthesis; in CLM4.5 nitrification and denitrification were modified based on CENTURY model from distinct and vertically resolved NO_3 and NH_4 pools (Koven et al., 2013; Lawrence et al., 2019). In CLM5, the FUN module proposes that plants will consume a certain amount of carbon when they uptake nitrogen based on the reality of symbiotic nitrogen fixation (Brzostek et al., 2014). The description of biogeochemical processes in CLM5 has been well documented in Lawrence et al. (2019) and more details can be found on this site (https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/).

2.2. The Matrix Representation of the Terrestrial Carbon and Nitrogen Cycle

In the carbon-nitrogen coupled cycle model, multiple carbon/nitrogen balance equations are used to calculate the dynamics of each carbon or nitrogen pool, which can be mathematically expressed by a matrix equation (Luo et al., 2003, 2017, 2022). The vegetation simulation in CLM5 contains six tissue pools (leaf, fine root, live stem, dead stem, live coarse root, dead coarse root) and each tissue pool is accompanied by a storage and a transfer pool to facilitate the description of vegetation biomass growth and other phenological processes. An additional nitrogen pool is used to store reusable nitrogen (Figure 1). These 18 Carbon (C) pools and 19 Nitrogen (N) pools are represented by 18×1 vector $C_{\mathrm{veg}}(t)$ and 19×1 vector $N_{\mathrm{veg}}(t)$. Following two matrix equations are constructed to track the state variables of carbon and nitrogen in CLM5 matrix version (Lu et al., 2020):

$$\frac{dC_{\mathrm{veg}}}{dt} = B_C(t)U_{C\mathrm{in}}(t) + \left(A_{C\mathrm{ph}}(t)K_{C\mathrm{ph}} + A_{C\mathrm{gm}}(t)K_{C\mathrm{gm}} + A_{Cfi}(t)K_{C\mathrm{fi}}\right)C_{\mathrm{veg}}(t) \quad (1)$$

$$\frac{dN_{\mathrm{veg}}}{dt} = B_N(t)U_{N\mathrm{in}}(t) + \left(A_{N\mathrm{ph}}(t)K_{N\mathrm{ph}} + A_{N\mathrm{gm}}(t)K_{N\mathrm{gm}} + A_{N\mathrm{fi}}(t)K_{N\mathrm{fi}}\right)N_{\mathrm{veg}}(t) \quad (2)$$

t represents each simulated timestep, which is half an hour in this study. A state variable with t will change dynamically with time. $B_C(t)$ (18×1 vector) and $B_N(t)$ (19×1 vector) indicate input allocation fractions of Total vegetation carbon (TOTVEGC) or nitrogen. $U_{C\mathrm{in}}(t)$ and $U_{N\mathrm{in}}(t)$ are total amount of carbon/nitrogen inputs to the system. In general, carbon input refers to the net primary productivity (NPP) and the sources of nitrogen input are usually nitrogen absorption and fixation. A are 18×18 matrices for vegetation carbon pools and 19×19 matrices for N pools, representing carbon/nitrogen transfer coefficients along prescribed pathway between pools. All A matrices have -1 diagonally, and the element a_{ij} (in row i and column j) in the matrix represents transfer coefficient from the pool j to i. $A_{C\mathrm{ph}}$ and $A_{N\mathrm{ph}}$ are used to quantify carbon and nitrogen fluxes among plant pools controlled by phenology processes. $A_{C\mathrm{gm}}(t)$ and $A_{N\mathrm{gm}}(t)$ describe the gap-mortality transfer fractions and fire-induced process is indicated by $A_{C\mathrm{fi}}(t)$ and $A_{N\mathrm{fi}}(t)$.

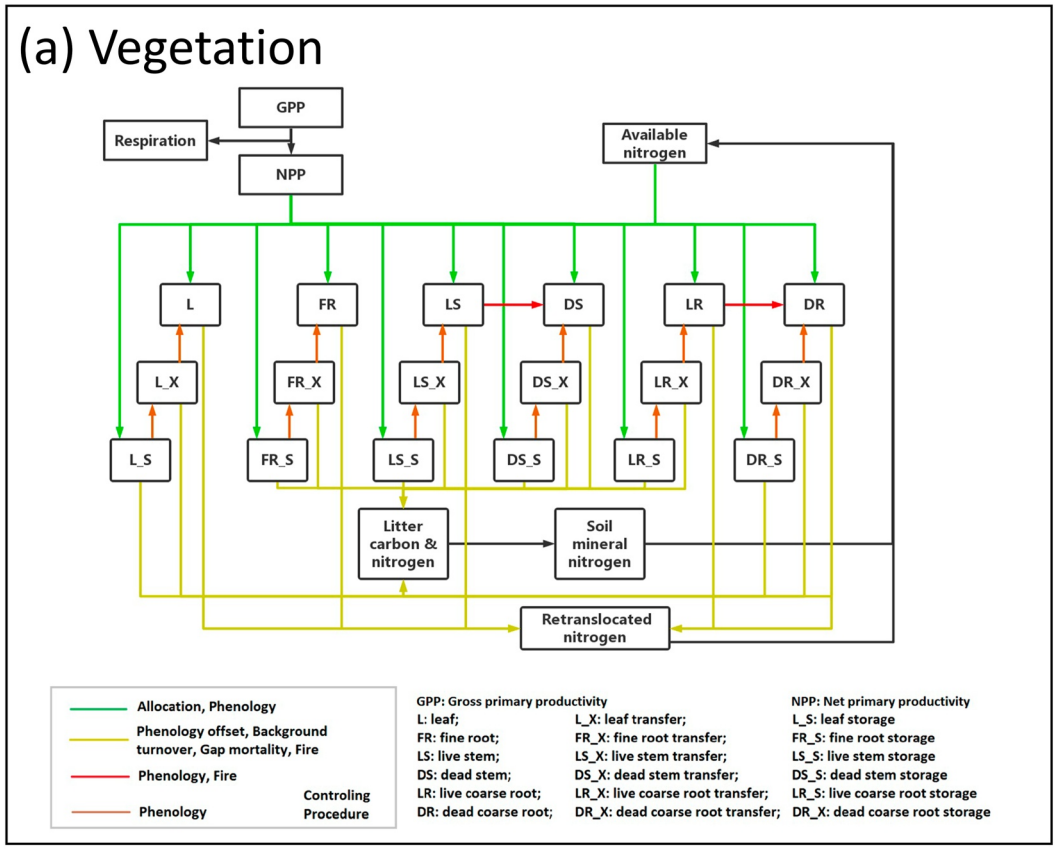

Figure 1.

K are diagonal matrices (18×18 matrices for carbon pools and 19×19 matrices for nitrogen pools) with each diagonal element representing the turnover rate of each plant carbon and nitrogen pools. Phenology matrices $K_{C\mathrm{ph}}(t)$ and $K_{N\mathrm{ph}}(t)$ represent plant tissue turnover rates according to phenology principles. Harvest rates and natural mortality are all included in gap mortality matrices $K_{C\mathrm{gm}}(t)$ and $K_{N\mathrm{gm}}(t)$. Fire matrices $K_{C\mathrm{fi}}(t)$ and $K_{N\mathrm{fi}}(t)$ refer to vegetation carbon/nitrogen losses from fire.

The soil carbon and nitrogen biogeochemical processes in CLM5 follow the CLM4.5bgc matrix structure developed by Huang et al. (2018), which includes CWD pool, three litter pools (metabolic litter, cellulose litter and lignin litter) and three SOM pools (fast SOM, slow SOM and passive SOM). Each pool can be subdivided into 20 layers along soil depth so that the carbon/nitrogen pools are represented by a 140×1 vector, that is, $(C_1(t), C_2(t), C_3(t), \ldots, C_{140}(t))^T$. The matrix equation for litter and SOM can be organized as following:

$$\frac{dC_{\mathrm{soil}}}{dt} = I_{C\mathrm{soil}} + \left(A_{Ch}\xi(t)K_h + V(t) + K_f(t)\right) C_{\mathrm{soil}}(t) \tag{3}$$

$$\frac{dN_{\mathrm{soil}}}{dt} = I_{N\mathrm{soil}} + \left(A_{Nh}\xi(t)K_h + V(t) + K_f(t)\right) N_{\mathrm{soil}}(t) \tag{4}$$

where $I_{C\mathrm{soil}}$ and $I_{N\mathrm{soil}}$ are 140×1 vector, representing vegetation carbon and nitrogen inputs to litter or CWD, respectively. As with vegetation, A_{Ch} and A_{Nh} a matrix of size (140×140), which is used to represent the transfer coefficient of carbon/nitrogen between pools. K_h are diagonal matrices (140×140), representing the potential decomposition coefficient of litter and soil pools. K_f refers to litter and SOM carbon/nitrogen loss induced by fire. $\xi(t)$ is a diagonal matrix (140×140) reflecting the environmental factors. Each element is a number from 0 to 1 to indicate the degree of carbon/nitrogen cycle affecting by environment such as soil temperature and moisture. $V(t)$ is a 140×140 matrix, used to capture carbon and nitrogen dynamics in the vertical soil profile through mixing mechanisms (i.e., diffusion and advection).

The above matrix equations have been successfully constructed and verified in previous studies (Huang et al., 2018; Liao et al., 2022a; Lu et al., 2020; Luo et al., 2022). More details about CLM5 matrix version can be found in Lu et al. (2020).

2.3. Spin-Up Methods

This study evaluated three spin-up methods, which are ND spin-up, AD spin-up, and semi-analytic spin-up (SASU). These three methods adopt the same spin-up criterion (the mean change in total ecosystem carbon over last two loops $\Delta C_{\mathrm{TOTECOSYSC}} \leq 1.0$ g C m^{-2} yr^{-1} for the Brazil site and each gird cell in global verification), which is one of criteria in Thornton and Rosenbloom (2005) and Shi et al. (2013). Below is a description of the three methods.

2.3.1. Native Dynamics (ND) Spin-Up

The ND spin-up is simple and universal. It keeps default model running from arbitrary initial conditions until steady-state is reached under cycled meteorological forcing (Thornton & Rosenbloom, 2005). Recursive forcing provides similar environmental conditions over time. With no additional disturbance, the state variables (carbon and nitrogen pools) in terrestrial model will gradually approach the quasi-equilibrium state after long-time simulation. It relies entirely on internal dynamics and results in high computation cost, but it is a reliable reference for other spin-up methods.

2.3.2. Accelerated Decomposition (AD) Spin-Up

Accelerated decomposition spin-up was first described by Thornton & Rosenbloom (2005) and was implemented into CLM4. The approach was slightly modified by Koven et al. (2013) to use different acceleration factors for

Figure 1. Diagram of the carbon and nitrogen processes of Community Land Model version 5 model. (a) The vegetation module tracks 18 carbon pools, 18 corresponding nitrogen pools, and an additional retranslocated nitrogen pool, which are controlled by phenology, gap-mortality and fire processes. L, leaf pool; L_X, leaf transfer pool; L_S, leaf storage pool; Fr, fine root; Fr_X, fine root transfer pool; Fr_S, fine root storage pool; Ls, live stem pool; Ls_X, live stem transfer pool; Ls_S, live stem storage pool; Ds, dead stem pool; Ds_X, dead stem transfer pool, Ds_S, dead stem storage pool; Lcr, live coarse root pool; Lcr_X, live coarse root transfer pool; Lcr_S, live coarse root storage pool; Dcr, dead coarse pool; Dcr_X, dead coarse root transfer pool; Dcr_S, dead coarse root storage pool. (b) The soil module tracks 7 carbon and 7 corresponding nitrogen pool categories that into 20 soil profiles, resulting in 140 pools in the matrix representations. SOM, soil organic matter; CWD, coarse woody debris. Figures are originally from Lu et al. (2020) and Huang et al. (2018).

each soil pool, leading to a 70% reduction in computational cost from ND. The spin-up efficiency is mainly limited by the pools with long turnover times. Thus, AD defines a series of fixed scaling scalars for litter and SOM pools to accelerate carbon decomposition while vegetation pools maintain their default turnover timescales. In this solution, the vegetation states achieve quasi-equilibrium quickly, but the litter and SOM carbon and nitrogen predicted state are inversely proportional to the decomposition scaling factors (Thornton & Rosenbloom, 2005). The scaling scalars can vary considerably for different pools and regions. For example, for passive soil organic carbon pool, the acceleration factor is higher to make it similar to the decomposition rates of other fast pools. Geographically, the decomposition rates at high latitude are much lower than in the tropics, where temperature and moisture factors strongly limit the carbon decomposition. In CLM5, these areas were assigned higher scalars to ensure spin-up efficiency, by adding a latitudinal dependence to the acceleration. Introducing scaling factors, the model will converge to a steady state in a short time due to the fact that fast SOM decomposition provides sufficient mineralized nitrogen for rapid plant growth (Shi et al., 2013). Consequently, a quasi-steady state can be obtained by multiplying by the corresponding scaling scalars. Thus, the model continues to run at the ND model after the end of the AD simulation until a steady state is reached.

To evaluate computational efficiency in this study, we used the same AD scaling scalars as Lawrence et al. (2019) and followed 2 spin-up steps suggested by Andre et al. (2020) to do AD spin-up: (a) 200 years of AD phase in which both carbon decomposition and vertical transport rates are accelerated by a set of scaling factors to ensure sufficient mineral nitrogen availability. At the end of AD spin-up, an exit-AD phase is executed automatically in the model to obtain quasi-steady-state values by multiplying the soil carbon and nitrogen states from the AD spin-up step by the scaling factors; (b) a long-time (at least 400 years) post-AD spin-up in normal mode (ND) to reach final equilibrium.

2.3.3. Semi-Analytical Spin-Up (SASU)

Xia et al. first proposed SASU for CABLE based on matrix equation (Xia et al., 2012). Fundamentally, litter and soil organic carbon decomposition in terrestrial models limits the spin-up efficiency due to the long carbon turnover time. The default first-order dynamical scheme can efficiently simulate the decay of litter and SOM, which can be described by the following matrix equation:

$$\frac{dC_{\text{soil}}}{dt} = B(t)U(t) + A\xi(t)KC_{\text{soil}}(t) \tag{5}$$

where $C_{\text{soil}}(t)$ is a vector of pool sizes; $U(t)$ is the total carbon input from vegetation, usually referring to NPP; $B(t)$ is a vector of allocation fraction to each pool; A is a matrix of transfer coefficients (or microbial carbon use efficiency) to quantify carbon transition; $\xi(t)$ is a diagonal matrix of environmental scalars. K is a diagonal matrix, representing the potential decomposition coefficient of litter and soil pools. It has been well documented and validated in different models (Huang et al., 2018; Lu et al., 2020; Luo et al., 2003; Xia et al., 2012).

Mathematically, theoretical steady state can be calculated by making Equation 5 equal to zero. As there are several time-dependent variables ($U(t)$, $B(t)$, and $\xi(t)$), we treated them by mean values during the forcing loops ($\overline{U}$, $\overline{B}$, $\overline{\xi}$) so that we can solve the equation as:

$$C_{ss} = -\left(A\overline{\xi}K\right)^{-1}\overline{BU} \tag{6}$$

where C_{ss} is the theoretical steady state, assuming that the carbon input is equivalent to the output. The dynamic changes of environmental factors $\xi(t)$ in the same climate loop are very similar and will be relatively stable after averaging. In addition to environmental factors, dynamic variables (e.g., $B(t)$ and $U(t)$), which are associated with the plant input and allocation, are also subject to feedback regulation of mineralized nitrogen. When vegetation growth is deficient in available mineral nitrogen, reduced gross primary productivity will lead to dynamic changes of downstream pools. In other words, the organic carbon and nitrogen produced by vegetation vary nonlinearly and generally dominate in the simulations. Regularly, the vegetation becomes stable within several atmosphere loops as it grows rapidly. Indeed, $\overline{B}$ and $\overline{U}$ will be close to steady state, which makes the analytical solution more accurate. The actual equilibrium can be achieved quickly when the analytical solution is applied as initial value for next simulation loop.

In the CLM5.0 matrix version, by setting Equations 3 and 4 to 0, at the end of each climate loop, the analytical solution of soil carbon and nitrogen can be calculated based on matrix operation as:

$$C_{\text{soil}}^{ss} = -\left(A_{Ch}\overline{\xi}K_{Ch} + \overline{V} + \overline{K_f}\right)^{-1}\overline{I_{C\text{soil}}} \tag{7}$$

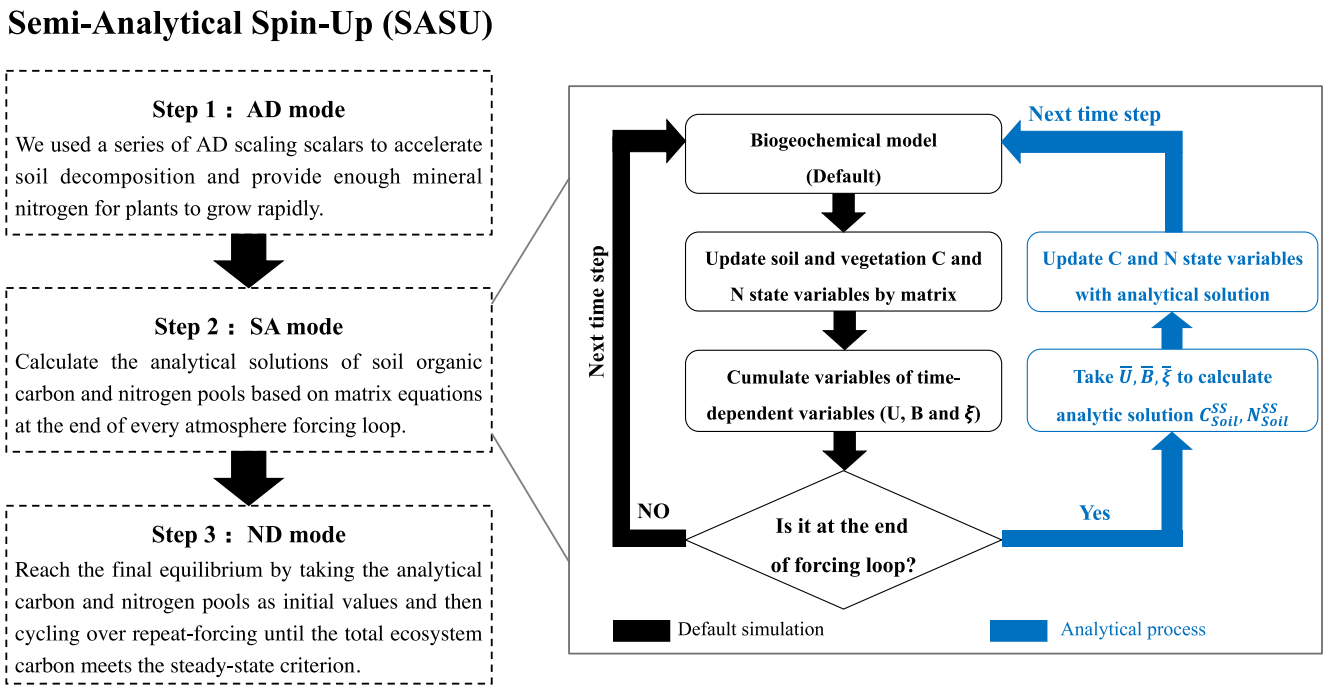

Figure 2. Procedures of the semi-analytical spin-up (SASU) method used in this study. There are three steps in SASU, including accelerated decomposition mode, semi-analytical (SA) mode, and native dynamics mode.

$$N_{\text{soil}}^{ss} = -\left(A_{Nh}\overline{\xi}K_{Nh} + \overline{V} + \overline{K_f}\right)^{-1}\overline{I_{N\text{soil}}} \tag{8}$$

In this study, we combined AD, semi-analytical and native dynamic modes to become a new SASU framework as the following three steps (Figure 2):

Step 1. Accelerated decomposition mode: We used the same AD scaling scalars from Andre et al. (2020) and Lawrence et al. (2019) to accelerate soil decomposition and provide enough mineral nitrogen for plants to grow rapidly. Thus, model will quickly obtain a near steady-state estimate of NPP (or all plant pools), which usually takes 100–200 years (looping over 20 years of atmospheric forcing).

Step 2. Semi-analytical (SA) mode: Calculate the analytical solutions of soil organic carbon and nitrogen pools based on matrix equations at the end of every atmosphere forcing loop. Figure 2 shows an overview diagram of the SASU mode used in CLM5 matrix version.

Step 3. Native dynamic mode: Reach the final equilibrium by taking the analytical carbon and nitrogen pools as initial values and then cycling over repeat-forcing until the total ecosystem carbon meets the spin-up criterion.

We tested and adjusted the required simulation time in the three steps to maximize the spin-up efficiency. The following procedure was identified as the best-performing for site and global simulation: (a) 160 years of AD and the exit-AD mode as described in Section 2.3.2; (b) 200 years of semi-analytical mode, updating soil organic carbon and nitrogen pools with analytical solutions every forcing loop; (c) at least 40 years ND mode to reach final equilibrium.

2.4. Evaluation of These Three Spin-Up Methods

To evaluate the computational efficiency of different spin-up approaches, we ran AD and SASU at a Brazil site (55°W, 7°S). They are driven by recursive meteorological forcing from 1901 to 1920 of the Global Soil Wetness Project Phase 3 dataset (GSWP3) (Guo et al., 2006). In addition, we conducted the ND, AD and SASU for a 400-gridcell sparse grid (1.9° × 2.5°) (Hoffman et al., 2013) that is used for parameter perturbation studies with CLM. The 400 grid cells that are used were identified using a cluster analysis and represent a minimum number

of grid cells that can reasonably capture model behavior across different ecoregions around the globe (Figure 5). They are driven by repeated meteorological forcing from 2005 to 2014 to compare these approaches. All spin-up methods are initiated with cold-start, which means that plants will grow from baer land.

2.5. Parameter Perturbation Experiment (PPE) in CLM5

With the growing of complexity and comprehensiveness of land models, land carbon dynamics simulated by earth system models are highly variable and fit poorly with observations (Luo et al., 2015; Spafford & MacDougall, 2021; Tian et al., 2015). It is crucial to understand sources of uncertainties. There are more than 200 crucial parameters in CLM5 and the contribution of parameter uncertainty to total uncertainty expected to be large, but unquantified (Bradford et al., 2016). Parameter values in the current land models are mostly determined on an ad hoc basis and may be derived from the results of field experiments, other models, or informed from scientific studies (Luo et al., 2001). Systematic parameter calibration will enhance the accuracy of simulations, and increase suitability and accessibility of models for actionable science. Parameter perturbation experiment on CLM5 is proposed to examine the parameter importance and sensitivity on model results by ensemble analysis under parameter perturbation. Steady state under each parameter perturbation should be estimated to give us insight into how carbon cycle response to parameter change. PPE considered total 197 parameters in CLM5 across 14 categories and thousands of spin-up tasks are required for this ensemble analysis.

We applied SASU to PPE protocol based on a global 400 grid cell sparse grid driven by repeated meteorological forcing from 2005 to 2014. The test was to start from an equilibrium with default parameterization and we modified a model parameter (the changed parameter *stem_leaf* in this test was reduced by 50%, which is an allocation parameter that controls the amount of new stem carbon per new leaf carbon). We used the SASU method to achieve a new equilibrium with this modified parameter. We tested a range of simulation times in the three SASU method steps and identified the following procedures as optimal: (a) 20 years of AD phase and exit-AD spin-up as described in Section 2.3.2; (b) 120 years of semi-analytical mode, updating soil organic carbon and nitrogen pools with analytical solutions every forcing loop; (c) at least 20 years ND mode to reach final equilibrium. It is worth to noticed that the time allocation among these steps is not quite the same as description in Section 2.4. The time of step 1 (AD mode) is greatly reduced in PPE due to the relative mature vegetation but no growing from baer land.

3. Results

3.1. Spin-Up at the Brazil Site

We first compared the computational cost of each spin-up method at the Brazil site. For the Brazil site, the AD spin-up scaling scalars were set to 1, which means that it is fundamentally the same as ND. To reach the spin-up threshold ($\Delta C_{\mathrm{TOTECOSYSC}} < 1.0$ g C m^{-2} yr^{-1}), we ran SASU and took ND spin up as a control. We recorded the first year when the growth rate of total ecosystem carbon storage is below the spin-up criterion. Generally, it took SASU 420 years and ND 3,000 years to achieve the same steady state (Figure 3a). ND kept a slow growth rate for about 2,000 years to reach steady state. In SA mode, all the carbon pools quickly approached a quasi-steady-state after 360 years (Figure 3a). Moreover, in the following simulation of step 3, the state variables maintained a dynamic balance and the change rate approached zero (Figure 3b). Compared with ND, SASU significantly reduced the computational cost, saving 2,580 years (86.0% simulation time) to reach the same state (Figures 3a and 4). The steady-state estimations from ND and SASU are consistent, just with small bias. TOTVEGC reached the same equilibrium of 15.37 kg C m^{-2}. Total ecosystem carbon (TOTECOSYSC) is 25.25 kg C m^{-2} with ND and 25.28 kg C m^{-2} with SASU. Even the same spin-up criterion is adopted, SASU has reached the equilibrium with 0.56 g C m^{-2} yr^{-1} of change rate, while total ecosystem carbon in ND still increases with 1.00 g C m^{-2} yr^{-1}.

When a smaller criterion of the steady state ($\Delta C_{\mathrm{TOTECOSYSC}} < 0.05$ g C m^{-2} yr^{-1}) was set to evaluate computational efficiency of the two spin-up methods at the Brazil site, it took SASU a total of 480 years and ND 4,720 years to meet the smaller criterion (Figures 3a–3c). It took an additional 1720 years to reach the new criterion of $\Delta C_{\mathrm{TOTECOSYSC}} < 0.05$ g C m^{-2} yr^{-1} beyond the 3,000 years to reach the original criterion of ($\Delta C_{\mathrm{TOTECOSYSC}} < 1.0$ g C m^{-2} yr^{-1}) for ND (Figure 3c). It only took an additional 60 years to reach the new criterion of $\Delta C_{\mathrm{TOTECOSYSC}} < 0.05$ g C m^{-2} yr^{-1} beyond the 420 years to reach the original criterion of ($\Delta C_{\mathrm{TOTECOSYSC}} < 1.0$ g C m^{-2} yr^{-1}) for SASU (Figure 3c). Biases of the steady-state pool sizes are reduced to a minimal for both the methods. Total

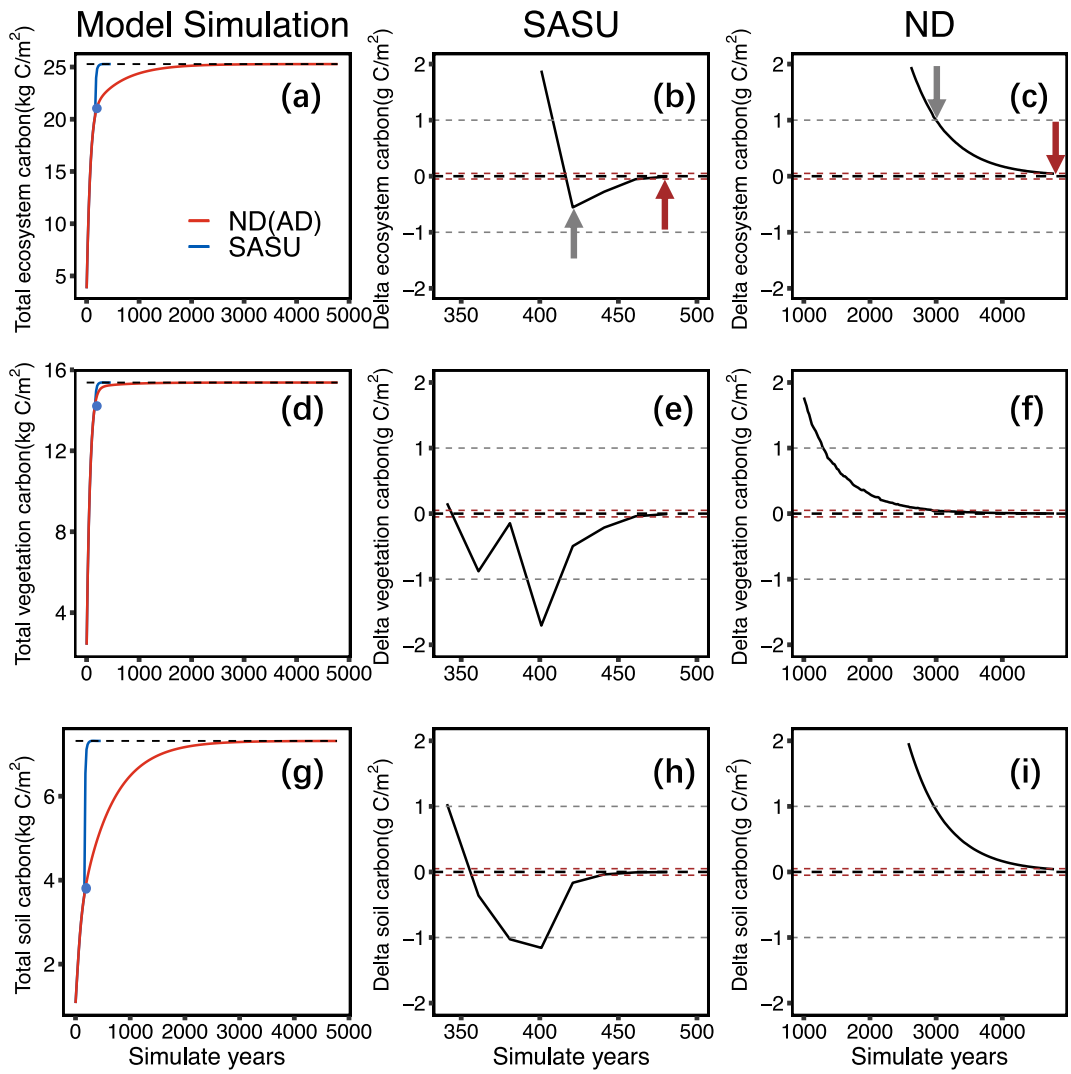

Figure 3. Carbon state trajectories (a, d, g) and the change of carbon between loops for semi-analytical spin-up (SASU) (b, e, h) and native dynamics (ND) (c, f, i) on the Brazil site. Results are organized as total ecosystem carbon (a–c), total vegetation carbon (d–f) and total soil carbon (g–i). These two methods ended up with the same steady state (horizontal black dashed line in a, d, g) on the Brazil site. Blue dots in a, d, g indicate where the SA mode starts. To reach the criterion of 1.0 g C m^{-2} yr^{-1} (horizontal gray dashed line in b, c, e, f, h, i), SASU took 420 years (b) and ND took 3,000 years (c), which were marked with gray arrows in (b, c). To meet a smaller criterion of 0.05 g C m^{-2} yr^{-1} (horizontal brown dashed line in b, c, e, f, h, i), SASU took 480 years (b) and ND took 4,720 years (c), which were marked with brown arrows in (b, c).

ecosystem carbon from both ND and SASU are 25.28 kg C m^{-2}. The difference of total ecosystem carbon is about 0.001 kg C m^{-2}, which is less than 0.005% of the pool size (Table 1). Each individual carbon pool (i.e., CWD, TOTVEGC, each litter and soil pool) was consistent as well (Table 1). The difference of CWD pools between the two methods was 6.51×10^{-5} kg C m^{-2}, while the difference of total soil carbon (TOTSOMC) was about 1.45×10^{-3} kg C m^{-2}. In general, the good agreement demonstrated that SASU can reach the same steady state as with the ND method but with much higher computational efficiency.

3.2. Spin-Up at Global Scale

We selected 400 sparse grid cells at global scale to compare the spin-up performance among ND, AD and SASU methods. All of grid cells were under $1.9° \times 2.5°$ resolution and the same for three spin-up approaches (Figure 5). For the global test, spin-up ends when $\Delta C_{TOTECOSYSC}$ is less than 1.0 g C m^{-2} yr^{-1} for more than 97% grid cells. The traditional ND and AD spin-up method spent 19,840 and 3,200 simulation years to reach equilibriums,

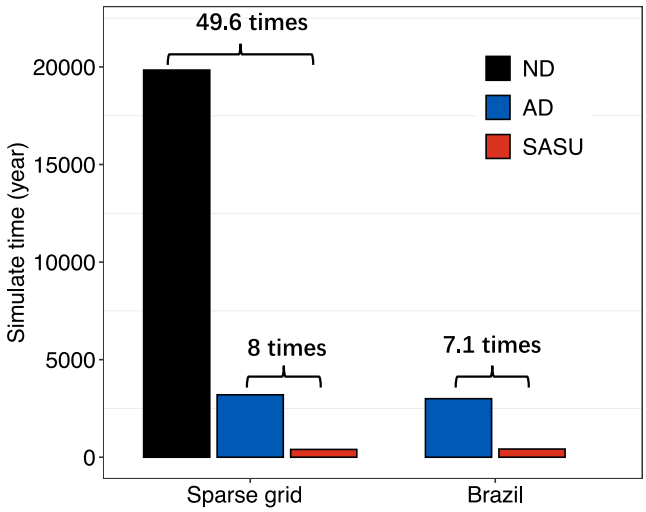

Figure 4. Model years needed for reaching equilibrium for different spin-up methods at 400 global grid cells ($\Delta C_{\text{TOTECOSYSC}} < 1.0$ g C m^{-2} yr^{-1} for each grid cell), the Brazil site ($\Delta C_{\text{TOTECOSYSC}} \leq 1.0$ g C m^{-2} yr^{-1}).

respectively (Figure 4). For SASU, it took a total of 400 years, including 200 years for AD mode, 160 years for SA mode and the last 40 years for ND. At the end of the spin-up, the number of grid cells that did not reach the spin-up criterion was 2.50% of the 400 grid cells for ND, 2.78% for AD, and 2.57% for SASU.

For the 400 global sparse grid cells, the steady states obtained by the three methods were similar (Figure 5). Total ecosystem carbon density was high in northern North America, northern Asia and the Tibetan Plateau, but low in the Sahara and Australia (Figures 5a, 5c, and 5e). Total soil organic carbon and total ecosystem carbon was mainly stored at high-latitude regions (~60°N), especially in Asia and northern North America (Figures 5b, 5d, and 5f).

3.3. Application of SASU in Parameter Perturbation Experiment (PPE)

When the parameter in plant processes was perturbed, gross primary productivity and total ecosystem carbon had changed compared with the default steady state. Total ecosystem carbon increased to 317.83 Pg C (Figure 6a) and NPP quickly approached relative equilibrium (from 1.97×10^{-7} to 2.03×10^{-7} g C m^{-2} s^{-1}) (Figure 6b). The whole spin-up process took 190 years by SASU from the default to a new steady state (Figure 6c). There were 2.61% of global areas that did not meet the spin-up criterion ($\Delta C_{\text{TOTECOSYSC}} < 1.0$ g C m^{-2} yr^{-1}) (Figures 6c and 6d).

4. Discussion

4.1. Improved Computational Efficiency for CLM5 With Vertical Structure

The development of vertical structure in SOM models posed great challenges to spin-up. Most terrestrial models regard the soil as a bulk without description of vertical distribution of SOM along the depth (Parton et al., 1987; Schimel et al., 1994). Recently, with increasing awareness of the importance of deep soil organic carbon, more and more models include an explicit representation of the vertical SOM distribution to improve predictions of carbon cycling, as well as facilitate the addition of new process descriptions (Braakhekke et al., 2011; Clark

Table 1
Mean Steady State Values (kg C/m^2) and the Corresponding Differences From Last Two Loops (g C/m^2) of State Variables With Semi-Analytical Spin-Up (420 years) and Accelerated Decomposition (3,000 years)

		SASU		ND/AD	
State variable	SASU-ND (kg C/m^2)	Pool size (kg C/m^2)	ΔC (g C/m^2)	Pool size (kg C/m^2)	ΔC (g C/m^2)
CWDC	-1.8×10^{-4}	2.34	-2.5×10^{-2}	2.34	8.7×10^{-3}
Metabolic Litter	6.2×10^{-7}	9.9×10^{-3}	2.0×10^{-4}	9.9×10^{-3}	3.4×10^{-5}
Cellulose Litter	1.5×10^{-5}	0.17	-2.0×10^{-3}	0.17	5.7×10^{-4}
Lignin Litter	5.5×10^{-6}	6.7×10^{-2}	-4.5×10^{-4}	6.6×10^{-2}	2.2×10^{-4}
Fast SOM	1.6×10^{-5}	0.10	7.2×10^{-5}	0.10	6.9×10^{-4}
Slow SOM	3.8×10^{-4}	2.75	-2.0×10^{-2}	2.75	1.7×10^{-2}
Passive SOM	2.5×10^{-2}	4.46	-1.8×10^{-2}	4.44	8.9×10^{-1}
TOTSOMC	2.7×10^{-2}	7.32	-3.8×10^{-2}	7.29	4.6×10^{-2}
TOTVEGC	9.7×10^{-4}	15.37	-2.1×10^{-1}	15.37	4.4×10^{-2}
TOTECOSYSC	2.7×10^{-2}	25.28	-2.8×10^{-1}	25.25	9.6×10^{-1}

Note. CWDC = Coarse Wood Debris Carbon; TOTSOMC = Total SOM Carbon; TOTVEGC = Total Vegetation Carbon; TOTECOSYSC = Total Ecosystem Carbon. Since scaling scalars at Brazil site are all 1, which means AD does the same way as ND for the Brazil site.

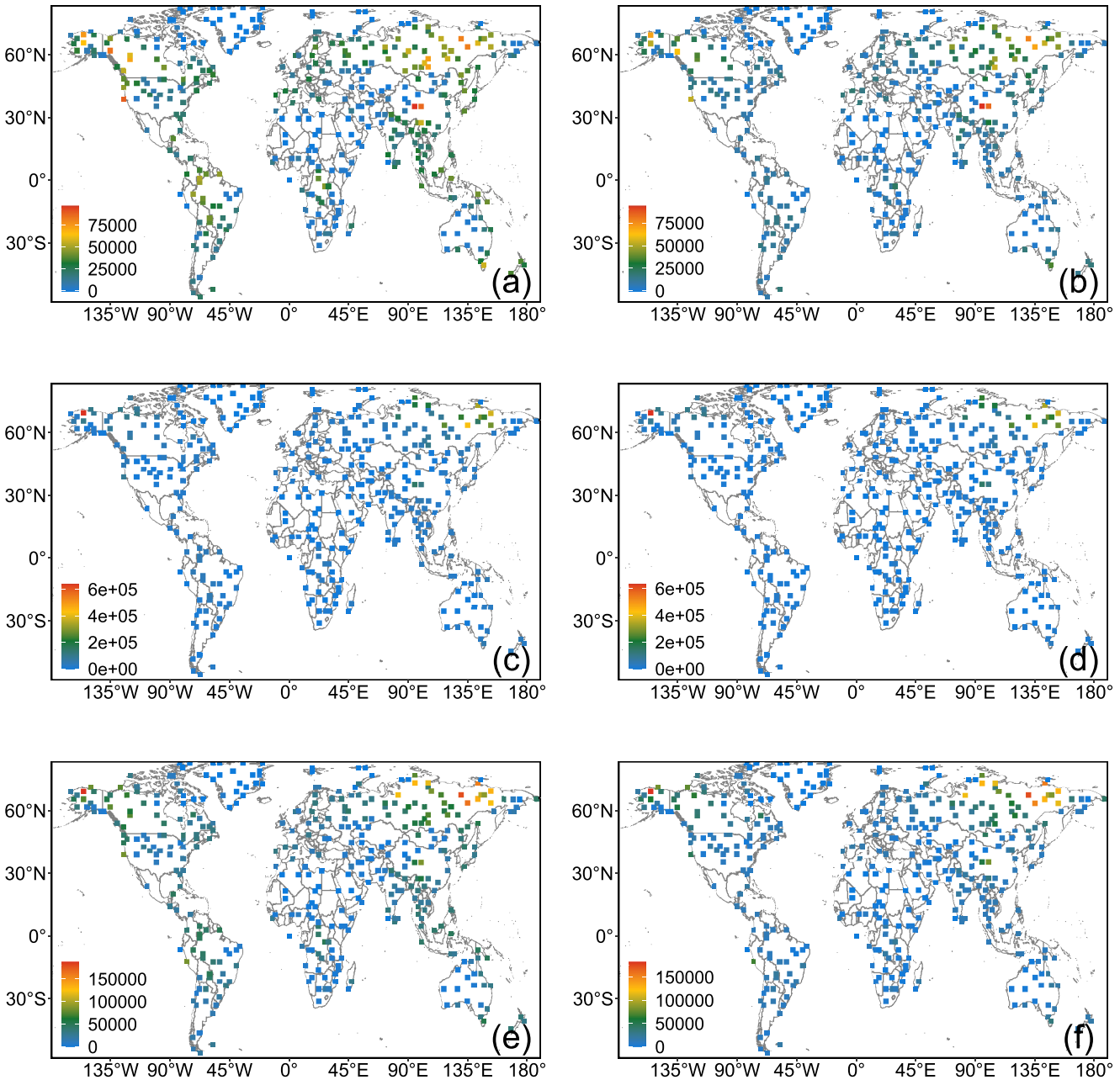

Figure 5. Global distributions of steady-state carbon density (g C m^{-2}) from native dynamics (a, b), accelerated decomposition (c, d) and semi-analytical spin-up (e, f) at 400 global grid cells. The figure is organized as the global carbon density of total ecosystem carbon (a, c, e), total soil carbon (b, d, f).

et al., 2011; Huang et al., 2018; Koven et al., 2013). However, the vertically resolved structure strictly limited the decomposition of deep soil by environmental factors, especially in high latitudes, resulting in little carbon input and output in deep soil and long carbon turnover time (Jenkinson et al., 2008; Jones et al., 2017), which means native dynamic simulation need an extremely long time to bring the SOM to the final steady state.

In CLM5, the extremely long turnover time in vertical structure brings a heavy computational burden on spin up and makes it difficult to accurately assess the steady state (Lawrence et al., 2019). ND is really a time-consuming method, requiring 19,840 years for the sparse grid, which took about 40 days of real time on a high-performance computing system. The widely used AD spin-up method was 6.2 times faster than ND (Figure 4), which took about a week of real time. However, due to the extremely long turnover time of deep SOM, the acceleration factors of deep soil layers are particularly high (greater than 200 in CLM5) to accelerate the decomposition and

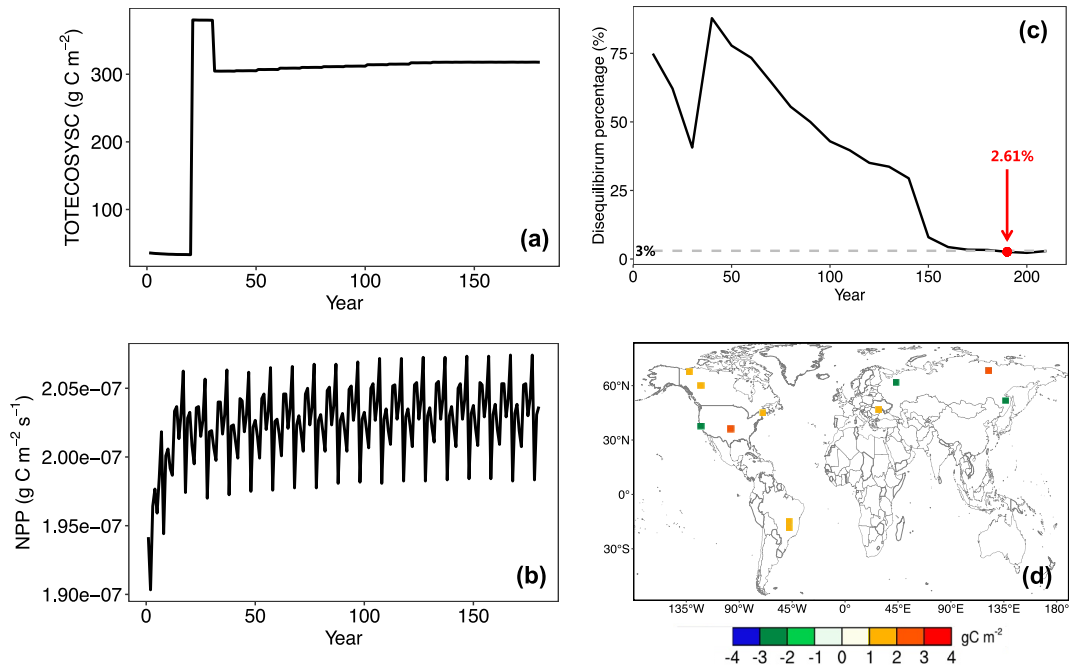

Figure 6. The application of semi-analytical spin-up in parameter perturbation experiment. Total ecosystem carbon (a) and net primary productivity state trajectories (b) in whole spin-up process. Changes in land area percentage of disequilibrium (c) and disequilibrium area distribution (d). Gray line in (c) is the threshold of 3% disequilibrium regions.

SOM cycling (Koven et al., 2013). Microbial activities are generally less active in deep soil, resulting in low SOM decomposition rates and small environmental factors (e.g., temperature, water and oxygen scalar), especially in high latitudes with permafrost (Koven et al., 2013, 2015; Schuur et al., 2015). In contrast, not limited by the long turnover time, SASU can estimate the theoretical steady-state once the annual soil carbon inputs and losses are obtained (Xia et al., 2012), which reduced the real running time to less than a day. The analytical principle of SASU is more conducive to solving the "spin up problem" brought by vertical structure in terms of computational efficiency and simulation accuracy in the context of model development.

The SASU method saved 86% of the computational time at the Brazil site, and 98.0% at 400 sparse grid cells worldwide than traditional ND spin-up. Compared to the efficient method explored by Thornton and Rosenbloom (AD method), SASU saved 87.5% of computational time at global scale. It is more efficient than methods currently reported in the literature. Our results are consistent with the results recorded in previous studies. AD can be successfully applied to land model and the calculation time is saved by 84% compared with the traditional ND method, which also showed the high efficiency of AD method as documented (Thornton & Rosenbloom, 2005). Xia et al. (2012) first applied SASU to CABLE and sped up spin-up by 20 times (Xia et al., 2012). To adapt to the strong carbon-nitrogen coupling cycle, compared with Xia et al. (2012), we introduced the AD step at the beginning of spin up, which can help the system stability in a short time. In this study, the computational efficiency is greatly improved, which is 50 times higher than 20 times in CABLE (Xia et al., 2012).

4.2. Applications of SASU to Various Biogeochemical Models

The developed SASU is directly applicable to most of terrestrial biogeochemical models that followed similar first-order decay principle with CLM5 in this study (Xia et al., 2012). The application of SASU in microbial models is under explored. With the increasing recognition of the role of microbial processes in soil carbon dynamics, dozens of microbial models have been developed in the past decades to consider microbial traits and nonlinear kinetics in simulating biogeochemical cycle (Allison et al., 2010; Wieder et al., 2015). For example, the rate of carbon assimilation by microbes and decomposition catalyzed by extracellular enzymes are not constant as assumed in linear model such as CLM5 but dependent on the substrate concentration (e.g., Michaelis-Menten Kinetics). Nevertheless, the nonlinear microbial models still can be represented in the matrix form (Sierra &

Müller, 2015) with either logistic or Michaelis-Menten equations in recent examples (Liao et al., 2022b; Tao et al., 2023). While various methods have been explored to obtain the steady state of microbial models, Georgiou et al. (2017) have developed a similar semi-analytical solution for accelerating model spin-up by solving differential equations and their method can be used to different microbial models (Tao et al., 2023). Thus, the SASU approach is likely appliable to different microbial models in the future.

The application of SASU to a dynamic global vegetation model (DGVM) is mathematically possible. However, it is still technically challenging. The development of DGVM in land models has expanded greatly in recent years, especially the vegetation demography model (VDM) (Fisher & Koven, 2020; Fisher et al., 2018; Koven et al., 2020). VDM introduces new representation of spatially heterogeneous canopy, which describes vegetation dynamics in two dimensions, individual plant size and the age of a forest gap since last disturbance. The steady state of vegetation dominancy hierarchy could be obtained by solving partial differential equations, which assume at long time scale that both plant size changes and gap age changes equal to zero at steady state. Although SASU is theoretically applicable on aboveground carbon cycle spin-up, the impact of weak nonlinearity or discretization in the VDM could still prevent the aboveground ecosystem from an immediate approaching to the steady state. Regardless, final convergence to the unique steady state should be still guaranteed based on the theory of compartmental model. Then, steady states of belowground soil carbon and nitrogen can be finally approached after plant size and age both reach the steady states.

For similar reasons, incorporating SASU into Community Earth System Model (CESM) may potentially increase the spin-up efficiency, but the acceleration rates will be negatively impacted by the nonlinearity. Compared to the offline version CLM5, the CESM add the feedback of the land surface change to the atmosphere. For example, an increase in leaf area index induced under climate change would also enhance latent heat fluxes (Forzieri et al., 2020), increase local precipitation (Muller & O'Gorman, 2011) and consequently alter the vegetation productivity patterns (Zhang et al., 2013). These feedbacks introduce nonlinearity into land carbon cycle, which therefore affect SASU spin up efficiency. Nevertheless, most of the new nonlinearity in CESM is external to land carbon cycle, so a quasi-steady state could still be derived from SASU.

4.3. Implications for Model Improvement

Acceleration of spin-up for biogeochemical models make some of the computationally costly studies possible, such as parameter sensitivity analysis (Huang et al., 2018), model inter-comparison (Liao et al., 2022a; Sulman et al., 2018) and data assimilation with complicated carbon cycle models (Hararuk et al., 2014; Shi et al., 2018; Tao et al., 2023). SASU accelerates spin-up, thus makes it computationally feasible to assimilate both flux- and pool-based big data to constrain full-dynamic model (e.g., earth system model) prediction through data assimilation and machine learning (Luo & Schuur, 2020; Reichstein et al., 2019; Tao et al., 2020; Xia et al., 2020). Constrained parameter values after data assimilation will improve SOM storage estimates and yielded better spatial and vertical distributions of SOM than the original model (Tao et al., 2020, 2023). In addition, SASU provides the possibility to implement parameter sensitivity analysis of complicated earth system models, such as PPE in CLM5. The total spin-up time of CLM5 was reduced to 190 years under parameter variations. Thus, computational resources can be reallocated to do more parameter perturbation experiments and ensemble analysis. SASU offers a new technical solution for most of terrestrial biogeochemical models that follow the first-order decay function in Equation 1 to increase the applicability of biogeochemical models toward an improved understanding of the land carbon cycle.

5. Conclusions

We applied a SASU framework to CLM5 to accelerate the spin-up of biogeochemical cycle to steady states. The SASU framework combined the AD mode, semi-analytical mode, and ND mode to improve the spin-up efficiency. SASU is 7.1 times faster than the AD spin-up to reach the same steady state at Brazil site. For the global simulation at 400 grid cells, SASU is 49.6 times faster than the ND method and 8.0 times faster than the AD method. Overall, the SASU method, to the best of our knowledge, is the most efficient spin-up method in comparison with all previously reported methods. Our study suggested that SASU is applicable to most of the biogeochemical models with the first-order kinetics and possibly with nonlinear microbial models and, thus, enabling computationally costly research, such as parameter sensitivity analysis and data assimilation with complex models.

Data Availability Statement

All simulations used in this work were performed using Version 5.0 of CLM https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/ (Lawrence et al., 2018). The code of the matrix model of CLM5 are available at this site https://github.com/chrislxj/ctsm/tree/cn-matrix_v3 (Lu, 2020). The model output data this study at https://zenodo.org/record/7593184#.Y9oTTOxBy3J (Liao, 2023) as an archival repository.

Acknowledgments
This work is supported by the National Key Research and Development Program of China (2022YFE0195900, 2021YFC3101600, 2020YFA0607900, 2020YFA0608000) and the National Natural Science Foundation of China (42125503, 42075137). Contributions from YL to this work were financially supported by US National Science Foundation (DEB 1655499, DEB 2017884), US Department of Energy (DE-SC0020227), and the subcontracts 4000158404 and 4000161830 from Oak Ridge National Laboratory (ORNL). The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. Computing and data storage resources, including the Cheyenne supercomputer (https://doi.org/10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. We thank all the scientists, software engineers, and administrators who contributed to the development of CESM2.

References

Adair, E. C., Parton, W. J., Del Grosso, S. J., Silver, W. L., Harmon, M. E., Hall, S. A., et al. (2008). Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates. *Global Change Biology*, *14*(11), 2636–2660. https://doi.org/10.1111/j.1365-2486.2008.01674.x

Ahlström, A., Schurgers, G., Arneth, A., & Smith, B. (2012). Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environmental Research Letters*, *7*(4), 044008. https://doi.org/10.1088/1748-9326/7/4/044008

Allison, S. D., Wallenstein, M. D., & Bradford, M. A. (2010). Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience*, *3*(5), 336–340. https://doi.org/10.1038/ngeo846

Andre, B., Kluzek, E., & Sacks, W. (2020). *CLM5.0 user's guide documentation*. National Center for Atmospheric Research (NCAR). from the documentation of the Community Land Model Retrieved from https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/users_guide/overview/introduction.html

Baisden, W. T., & Parfitt, R. L. (2007). Bomb ^{14}C enrichment indicates decadal C pool in deep soil? *Biogeochemistry*, *85*(1), 59–68. https://doi.org/10.1007/s10533-007-9101-7

Balesdent, J., Basile-Doelsch, I., Chadoeuf, J., Cornu, S., Derrien, D., Fekiacova, Z., & Hatte, C. (2018). Atmosphere-soil carbon transfer as a function of soil depth. *Nature*, *559*(7715), 599–602. https://doi.org/10.1038/s41586-018-0328-3

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., et al. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, *13*(3), 679–706. https://doi.org/10.1111/j.1365-2486.2006.01305.x

Braakhekke, M. C., Beer, C., Hoosbeek, M. R., Reichstein, M., Kruijt, B., Schrumpf, M., & Kabat, P. (2011). SOMPROF: A vertically explicit soil organic matter model. *Ecological Modelling*, *222*(10), 1712–1730. https://doi.org/10.1016/j.ecolmodel.2011.02.015

Bradford, M. A., Wieder, W. R., Bonan, G. B., Fierer, N., Raymond, P. A., & Crowther, T. W. (2016). Managing uncertainty in soil carbon feedbacks to climate change. *Nature Climate Change*, *6*(8), 751–758. https://doi.org/10.1038/nclimate3071

Brzostek, E. R., Fisher, J. B., & Phillips, R. P. (2014). Modeling the carbon cost of plant nitrogen acquisition: Mycorrhizal trade-offs and multipath resistance uptake improve predictions of retranslocation. *Journal of Geophysical Research: Biogeosciences*, *119*(8), 1684–1697. https://doi.org/10.1002/2014JG002660

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description—Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, *4*(3), 701–722. https://doi.org/10.5194/gmd-4-701-2011

Fang, Y., Liu, C., Huang, M., Li, H., & Leung, L. R. (2014). Steady state estimation of soil organic carbon using satellite-derived canopy leaf area index. *Journal of Advances in Modeling Earth Systems*, *6*(4), 1062–1064. https://doi.org/10.1002/2014ms000331

Fang, Y., Liu, C., & Leung, L. R. (2015). Accelerating the spin-up of the coupled carbon and nitrogen cycle model in CLM4. *Geoscientific Model Development*, *8*(3), 781–789. https://doi.org/10.5194/gmd-8-781-2015

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4). https://doi.org/10.1029/2018ms001453

Fisher, R. A., Koven, C. D., Anderegg, W. R. L., Christoffersen, B. O., Dietze, M. C., Farrior, C. E., et al. (2018). Vegetation demographics in earth system models: A review of progress and priorities. *Global Change Biology*, *24*(1), 35–54. https://doi.org/10.1111/gcb.13910

Forzieri, G., Miralles, D. G., Ciais, P., Alkama, R., Ryu, Y., Duveiller, G., et al. (2020). Increased control of vegetation on global terrestrial energy fluxes. *Nature Climate Change*, *10*(4), 356–362. https://doi.org/10.1038/s41558-020-0717-0

Georgiou, K., Abramoff, R. Z., Harte, J., Riley, W. J., & Torn, M. S. (2017). Microbial community-level regulation explains soil carbon responses to long-term litter manipulations. *Nature Communications*, *8*(1), 1223. https://doi.org/10.1038/s41467-017-01116-z

Ghimire, B., Riley, W. J., Koven, C. D., Mu, M., & Randerson, J. T. (2015). Representing leaf and root physiological traits in CLM improves global carbon and nitrogen cycling predictions. *Journal of Advances in Modeling Earth Systems*, *8*(2), 598–613. https://doi.org/10.1002/2015MS000538

Guo, Z., Zhao, M., Gao, X., Dirmeyer, P. A., Oki, T., & Hanasaki, N. (2006). Supplement to GSWP-2: Details of the forcing data. *Bulletin of the American Meteorological Society*, *87*(10), S10–S16. https://doi.org/10.1175/BAMS-87-10-Dirmeyer

Hararuk, O., Xia, J., & Luo, Y. (2014). Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 403–417. https://doi.org/10.1002/2013jg002535

Hoffman, F. M., Kumar, J., Mills, R. T., & Hargrove, W. W. (2013). Representativeness-based sampling network design for the State of Alaska. *Landscape Ecology*, *28*(8), 1567–1586. https://doi.org/10.1007/s10980-013-9902-0

Huang, Y., Lu, X., Shi, Z., Lawrence, D., Koven, C. D., Xia, J., et al. (2018). Matrix approach to land carbon cycle modeling: A case study with the community land model. *Global Change Biology*, *24*(3), 1394–1404. https://doi.org/10.1111/gcb.13948

Jenkinson, D. S., Poulton, P. R., & Bryant, C. (2008). The turnover of organic carbon in subsoils. Part 1. Natural and bomb radiocarbon in soil profiles from the Rothamsted long-term field experiments. *European Journal of Soil Science*, *59*(2), 391–399. https://doi.org/10.1111/j.1365-2389.2008.01025.x

Jobbagy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, *10*(2), 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

Jones, M. C., Harden, J., O'Donnell, J., Manies, K., Jorgenson, T., Treat, C., & Ewing, S. (2017). Rapid carbon loss and slow recovery following permafrost thaw in boreal peatlands. *Global Change Biology*, *23*(3), 1109–1127. https://doi.org/10.1111/gcb.13403

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., et al. (2020). Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) at Barro Colorado Island, Panama. *Biogeosciences*, *17*(11), 3017–3044. https://doi.org/10.5194/bg-17-3017-2020

Koven, C. D., Lawrence, D. M., & Riley, W. J. (2015). Permafrost carbon–climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(12), 3752–3757. https://doi.org/10.1073/pnas.1415123112

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., et al. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, *10*(11), 7109–7131. https://doi.org/10.5194/bg-10-7109-2013

Lardy, R., Bellocchi, G., & Soussana, J.-F. (2011). A new method to determine soil organic carbon equilibrium. *Environmental Modelling & Software*, *26*, 1759–1763. https://doi.org/10.1016/j.envsoft.2011.05.016

Lawrence, D. M., Fisher, R., Koven, C., Oleson, K., Swenson, S., Vertenstein, M., et al. (2018). Technical description of version 5.0 of the Community Land Model (CLM) [Software]. Retrieved from https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583

Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, *528*(7580), 60–68. https://doi.org/10.1038/nature16069

Liao, C. (2023). Matrix approach to accelerate spin-up of CLM5 [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7593184

Liao, C., Chen, Y., Wang, J., Liang, Y., Huang, Y., Lin, Z., et al. (2022a). Disentangling land model uncertainty via matrix-based ensemble model inter-comparison platform (MEMIP). *Ecological Processes*, *11*(1), 14. https://doi.org/10.1186/s13717-021-00356-8

Liao, C., Huang, W., Wells, J., Zhao, R., Allen, K., Hou, E., et al. (2022b). Microbe-iron interactions control lignin decomposition in soil. *Soil Biology and Biochemistry*, *173*, 108803. https://doi.org/10.1016/j.soilbio.2022.108803

Lu, X. (2020). Community terrestrial systems model (includes the community land model of CESM) matrix version [Software]. https://github.com/chrislxj/ctsm/tree/cn-matrix_v3

Lu, X., Du, Z., Huang, Y., Lawrence, D., Kluzek, E., Collier, N., et al. (2020). Full implementation of matrix approach to biogeochemistry module of CLM5. *Journal of Advances in Modeling Earth Systems*, *12*(11). https://doi.org/10.1029/2020ms002105

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., et al. (2022). Matrix approach to land carbon cycle modeling. *Journal of Advances in Modeling Earth Systems*, *14*(7). https://doi.org/10.1029/2022ms003008

Luo, Y., Keenan, T. F., & Smith, M. (2015). Predictability of the terrestrial carbon cycle. *Global Change Biology*, *21*(5), 1737–1751. https://doi.org/10.1111/gcb.12766

Luo, Y., & Schuur, E. A. G. (2020). Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. *Global Change Biology*, *26*(3), 1109–1117. https://doi.org/10.1111/gcb.14939

Luo, Y., Shi, Z., Lu, X., Xia, J., Liang, J., Jiang, J., et al. (2017). Transient dynamics of terrestrial carbon storage: Mathematical foundation and its applications. *Biogeosciences*, *14*(1), 145–161. https://doi.org/10.5194/bg-14-145-2017

Luo, Y., White, L. W., Canadell, J. G., DeLucia, E. H., Ellsworth, D. S., Finzi, A., et al. (2003). Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach. *Global Biogeochemical Cycles*, *17*(1). https://doi.org/10.1029/2002gb001923

Luo, Y., Wu, L., Andrews, J. A., White, L., Matamala, R., Schäfer, K. V. R., & Schlesinger, W. H. (2001). Elevated CO_2 differentiates ecosystem carbon processes: Deconvolution analysis of Duke Forest face data. *Ecological Monographs*, *71*(3), 357–376. https://doi.org/10.1890/0012-9615(2001)071[0357:Ecdecp]2.0.Co;2

Martin, M. P., Cordier, S., Balesdent, J., & Arrouays, D. (2007). Periodic solutions for soil carbon dynamics equilibriums with time-varying forcing variables. *Ecological Modelling*, *204*(3–4), 523–530. https://doi.org/10.1016/j.ecolmodel.2006.12.030

Muller, C. J., & O'Gorman, P. A. (2011). An energetic perspective on the regional response of precipitation to climate change. *Nature Climate Change*, *1*(5), 266–271. https://doi.org/10.1038/nclimate1169

Murty, D., & McMurtrie, R. E. (2000). The decline of forest productivity as stands age: A model-based method for analysing causes for the decline. *Ecological Modelling*, *134*(2–3), 185–205. https://doi.org/10.1016/s0304-3800(00)00345-8

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., et al. (2010). Technical description of version 4.0 of the community land model (CLM). NCAR Technical Notes, NCAR/TN-478+STR (pp. 1–266).

Parton, W. J., Schimel, D. S., Cole, C. V., & Ojima, D. S. (1987). Analysis of factors controlling soil organic matter levels in great plains Grasslands. *Soil Science Society of America Journal*, *51*(5), 1173–1179. https://doi.org/10.2136/sssaj1987.03615995005100050015x

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., & Klooster, S. A. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles*, *7*(4), 811–841. https://doi.org/10.1029/93GB02725

Qu, Y., Maksyutov, S., & Zhuang, Q. (2018). Technical Note: An efficient method for accelerating the spin-up process for process-based biogeochemistry models. *Biogeochemistry*, *15*(13), 3967–3973. https://doi.org/10.5194/bg-15-3967-2018

Randerson, J. T., Hoffman, F. M., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y. H., et al. (2009). Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Global Change Biology*, *15*(10), 2462–2484. https://doi.org/10.1111/j.1365-2486.2009.01912.x

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Running, S. W., & Hunt, E. R. (1993). Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. In *Scaling physiological processes* (pp. 141–158). https://doi.org/10.1016/b978-0-12-233440-5.50014-2

Schimel, D. S., Braswell, B. H., Holland, E. A., McKeown, R., Ojima, D. S., Painter, T. H., et al. (1994). Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Global Biogeochemical Cycles*, *8*(3), 279–293. https://doi.org/10.1029/94GB00993

Schuur, E. A., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., et al. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), 171–179. https://doi.org/10.1038/nature14338

Shi, M., Yang, Z.-L., Lawrence, D. M., Dickinson, R. E., & Subin, Z. M. (2013). Spin-up processes in the Community Land Model version 4 with explicit carbon and nitrogen components. *Ecological Modelling*, *263*, 308–325. https://doi.org/10.1016/j.ecolmodel.2013.04.008

Shi, Z., Crowell, S., Luo, Y., & Moore, B., III. (2018). Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nature Communications*, *9*(1), 2171. https://doi.org/10.1038/s41467-018-04526-9

Sierra, C. A., & Müller, M. (2015). A general mathematical framework for representing soil organic matter dynamics. *Ecological Monographs*, *85*(4), 505–524. https://doi.org/10.1890/15-0361.1

Spafford, L., & MacDougall, A. H. (2021). Validation of terrestrial biogeochemistry in CMIP6 earth system models: A review. *Geoscientific Model Development*, *14*(9), 5863–5889. https://doi.org/10.5194/gmd-14-5863-2021

Sulman, B. N., Moore, J. A. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., et al. (2018). Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry Letters*, *141*(2), 109–123. https://doi.org/10.1007/s10533-018-0509-z

Sun, Y., Goll, D. S., Huang, Y., Ciais, P., Wang, Y. P., Bastrikov, V., & Wang, Y. (2023). Machine learning for accelerating process-based computation of land biogeochemical cycles. *Global Change Biology*, *29*(11), 3221–3234. https://doi.org/10.1111/gcb.16623

Tao, F., Huang, Y., Hungate, B. A., Manzoni, S., Frey, S. D., Schmidt, M. W. I., et al. (2023). Microbial carbon use efficiency promotes global soil carbon storage. *Nature*, *618*(7967), 981–985. https://doi.org/10.1038/s41586-023-06042-3

Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., et al. (2020). Deep learning optimizes data-driven representation of soil organic carbon in earth system model over the conterminous United States. *Frontiers in Big Data*, *3*. https://doi.org/10.3389/fdata.2020.00017

Thornton, P. E., & Rosenbloom, N. A. (2005). Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecological Modelling*, *189*(1–2), 25–48. https://doi.org/10.1016/j.ecolmodel.2005.04.008

Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D. N., Schwalm, C. R., et al. (2015). Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochemical Cycles*, *29*(6), 775–792. https://doi.org/10.1002/2014GB005021

Wieder, W. R., Allison, S. D., Davidson, E. A., Georgiou, K., Hararuk, O., He, Y., et al. (2015). Explicitly representing soil microbial processes in Earth system models. *Global Biogeochemical Cycles*, *29*(10), 1782–1800. https://doi.org/10.1002/2015gb005188

Wieder, W. R., Hartman, M. D., Sulman, B. N., Wang, Y. P., Koven, C. D., & Bonan, G. B. (2018). Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Change Biology*, *24*(4), 1563–1579. https://doi.org/10.1111/gcb.13979

Wutzler, T., & Reichstein, M. (2007). Soils apart from equilibrium—Consequences for soil carbon balance modelling. *Biogeosciences*, *4*(1), 125–136. https://doi.org/10.5194/bg-4-125-2007

Xia, J., Wang, J., & Niu, S. (2020). Research challenges and opportunities for using big data in global change biology. *Global Change Biology*, *26*(11), 6040–6061. https://doi.org/10.1111/gcb.15317

Xia, J. Y., Luo, Y. Q., Wang, Y.-P., Weng, E. S., & Hararuk, O. (2012). A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state. *Geoscientific Model Development*, *5*(5), 1259–1271. https://doi.org/10.5194/gmd-5-1259-2012

Zhang, Y., Li, C., Zhou, X., & Moore, B., III. (2002). A simulation model linking crop growth and soil biogeochemistry for sustainable agriculture. *Ecological Modelling*, *151*(1), 75–108. https://doi.org/10.1016/S0304-3800(01)00527-0

Zhang, Y., Susan Moran, M., Nearing, M. A., Ponce Campos, G. E., Huete, A. R., Buda, A. R., et al. (2013). Extreme precipitation patterns and reductions of terrestrial ecosystem production across biomes. *Journal of Geophysical Research: Biogeosciences*, *118*(1), 148–157. https://doi.org/10.1029/2012jg002136