Full Implementation of Matrix Approach to Biogeochemistry Module of CLM5

Xingjie Lu[1,2] , Zhenggang Du[2,3] , Yuanyuan Huang[4] , David Lawrence[5] , Erik Kluzek[5] , Nathan Collier[6] , Danica Lombardozzi[5] , Negin Sobhani[5], Edward A. G. Schuur[2] , and Yiqi Luo[2]

[1]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China, [2]Center for Ecosystem Science and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA, [3]Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, Center for Global Change and Ecological Forecasting, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China, [4]CSIRO Oceans and Atmosphere, Aspendale, Victoria, Australia, [5]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [6]Computer Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Abstract Earth system models (ESMs) have been rapidly developed in recent decades to advance our understanding of climate change-carbon cycle feedback. However, those models are massive in coding, require expensive computational resources, and have difficulty in diagnosing their performance. It is highly desirable to develop ESMs with modularity and effective diagnostics. Toward these goals, we implemented a matrix approach to the Community Land Model version 5 (CLM5) to represent carbon and nitrogen cycles. Specifically, we reorganized 18 balance equations each for carbon and nitrogen cycles among the 18 vegetation pools in the original CLM5 into two matrix equations. Similarly, 140 balance equations each for carbon and nitrogen cycles among the 140 soil pools were reorganized into two additional matrix equations. The vegetation carbon and nitrogen matrix equations are connected to soil matrix equations via litterfall. The matrix equations fully reproduce simulations of carbon and nitrogen dynamics by the original model. The computational cost for forwarding simulation of the CLM5 matrix model was 26% more expensive than the original model, largely due to calculation of additional diagnostic variables, but the spin-up computational cost was significantly saved. We showed a case study on modeled soil carbon storage under two forcing data sets to illustrate the diagnostic capability that the matrix approach uniquely offers to understand simulation results of global carbon and nitrogen dynamics. The successful implementation of the matrix approach to CLM5, one of the most complex land models, demonstrates that most, if not all, the biogeochemical models can be reorganized into the matrix form to gain high modularity, effective diagnostics, and accelerated spin-up.

Plain Language Summary Land models are widely used in climate change research. Due to the complex system, the model is not easily comprehended, nor the results are easily interpreted, even by a specialist. Enhancing the model tractability is imperative to make climate change prediction more effective, especially as models become more and more complex. In this study, we developed a matrix model by reorganizing six carbon and nitrogen modules of Community Land Model version 5 (CLM5) into four matrix equations to represent vegetation carbon, vegetation nitrogen, soil carbon, and soil nitrogen balance equations, respectively. The CLM5 matrix model gains high modularity, effective diagnostics, and accelerated spin-up. The success of applying the matrix approach to CLM5, one of the most complex land models, support the theoretical analysis that the matrix approach is applicable to almost all land biogeochemical models.

1. Introduction

Land biogeochemistry models are an essential component of Earth system models (ESMs) and have been extensively used to study climate change and its impacts on societally relevant commodities such as food, timber, energy, and crops (Bonan & Doney, 2018). Results of biogeochemistry models have been used to guide climate change mitigation (Eyring et al., 2016; Friedlingstein, Andrew, et al., 2014;

Jones et al., 2013; Masson-Delmotte et al., 2018; Meinshausen et al., 2009; Stocker et al., 2013), estimate global carbon budget (Friedlingstein et al., 2019; Houghton et al., 2012; Le Quéré et al., 2013; Schaphoff et al., 2013; Song & Woodcock, 2003), and assess climate change-carbon cycle feedback (Arora et al., 2013; Friedlingstein, 2015; Friedlingstein et al., 2006; Friedlingstein, Meinshausen, et al., 2014; Wenzel et al., 2014). On one hand, the computational requirement is much higher than before as land models have become more complex over time. On the other hand, as land models have become popular worldwide and versatile for exploring a variety of issues, their users are more diverse with different backgrounds, skill sets, and abilities. It is highly desirable to develop land models with simplicity in coding, modularization, effective diagnostics, and high computational efficiency so that scientists can use and further develop the models with ease in understanding, evaluation, and improvement.

An epitome of the land model development is the Community Land Model (CLM). The initial version of CLM was established by merging National Center for Atmosphere Research (NCAR) Land Surface Model (LSM) (Bonan, 1996), Institute of Atmosphere Physics (IAP), Chinese Academy of Sciences land model (IAP94) (Dai & Zeng, 1997), and the Biosphere-Atmosphere Transfer Scheme (BATS) (Dickinson et al., 1993). The early versions mainly focused on biophysical processes of the land surface. CLM has been continually developed from version 2 (CLM2) in 2002 to version 5 (CLM5) in 2018, including changes to model structure as well as the addition of process representation. In particular, CLM3.5 for the first time included the carbon cycle and CLM4 represented the coupled carbon-nitrogen cycle based on the model Biome-BGC (BioGeochemical Cycles) (Running & Hunt, 1993; Thornton et al., 2002) and the land use change. CLM4.5 incorporated a vertically resolved soil biogeochemistry scheme (Koven et al., 2013), fire model (Li et al., 2012, 2013), methane model (Riley et al., 2011), crop model (Drewniak et al., 2013; Sacks et al., 2009), and carbon isotope model (Koven et al., 2013). CLM5 further updated the biogeochemical modules, such as plant nutrient dynamics using the Fixation and Update of Nitrogen (FUN) model (Fisher et al., 2010), and photosynthetic capacity using Leaf Use of Nitrogen for Assimilation (LUNA) model (Xu et al., 2012). The development of land surface models has met the increasing need from the users and, meanwhile, makes models very complex. The lines of the source code were 32,651 in CLM2 and increased to over 150,000 in CLM5, of which nearly 50,000 lines are used for the biogeochemical module. The increasing complexity of CLM means that the development enterprise would benefit from greater modularization and methods that help diagnose model results. In addition, the long spin-up for soil biogeochemistry hampers uses of the model by a wider range of researchers.

Luo et al. (2017) proposed a matrix approach to land biogeochemistry modeling. The approach reorganizes carbon balance equations in the original models into one matrix equation without changing any modeled carbon cycle processes and mechanisms. The matrix approach was initially implemented in the Terrestrial ECOsystem model (TECO) for data assimilation studies (Luo et al., 2003; Xu et al., 2006). The matrix approach has recently been applied to global land models, such as the Community Atmosphere-Biosphere-Land Exchange model (CABLE) (Xia et al., 2012, 2013), LPJ-GUESS (Ahlstrom et al., 2015), ORCHIDEE (Huang, Zhu, et al., 2018), CLM3.5 (Hararuk et al., 2015; Hararuk & Luo, 2014; Rafique et al., 2016), CLM4 (Rafique et al., 2017), CLM4.5 (Huang, Lu, et al., 2018), and CLM5 in this study. The CABLE matrix model was used to accelerate its spin-up (Xia et al., 2012) and for traceability analysis to trace sources of uncertainty in model simulations of terrestrial carbon storage (Xia et al., 2013). The LPJ-GUESS matrix model was used to analyze the uncertainty of future climate-induced C uptake (Ahlstrom et al., 2015). Matrix versions of CLM4, CASA, and CABLE were compared to assess carbon dynamics using the traceability framework (Rafique et al., 2017). Overall, the matrix approach offers simplicity in coding, modularization, effective diagnostics, and high computational efficiency for spin-up. The latter makes it possible to conduct complex parameter sensitivity analysis (Huang, Lu, et al., 2018) and data assimilation (Hararuk & Luo, 2014; Shi et al., 2018; Tao et al., 2020).

This paper describes the full implementation of the matrix approach to CLM5. It is built upon the study by Huang, Lu, et al. (2018) who reorganized soil carbon balance equations in CLM4.5 into a stand-alone matrix model. This study reorganizes mass balance equations of carbon and nitrogen cycles in both vegetation and
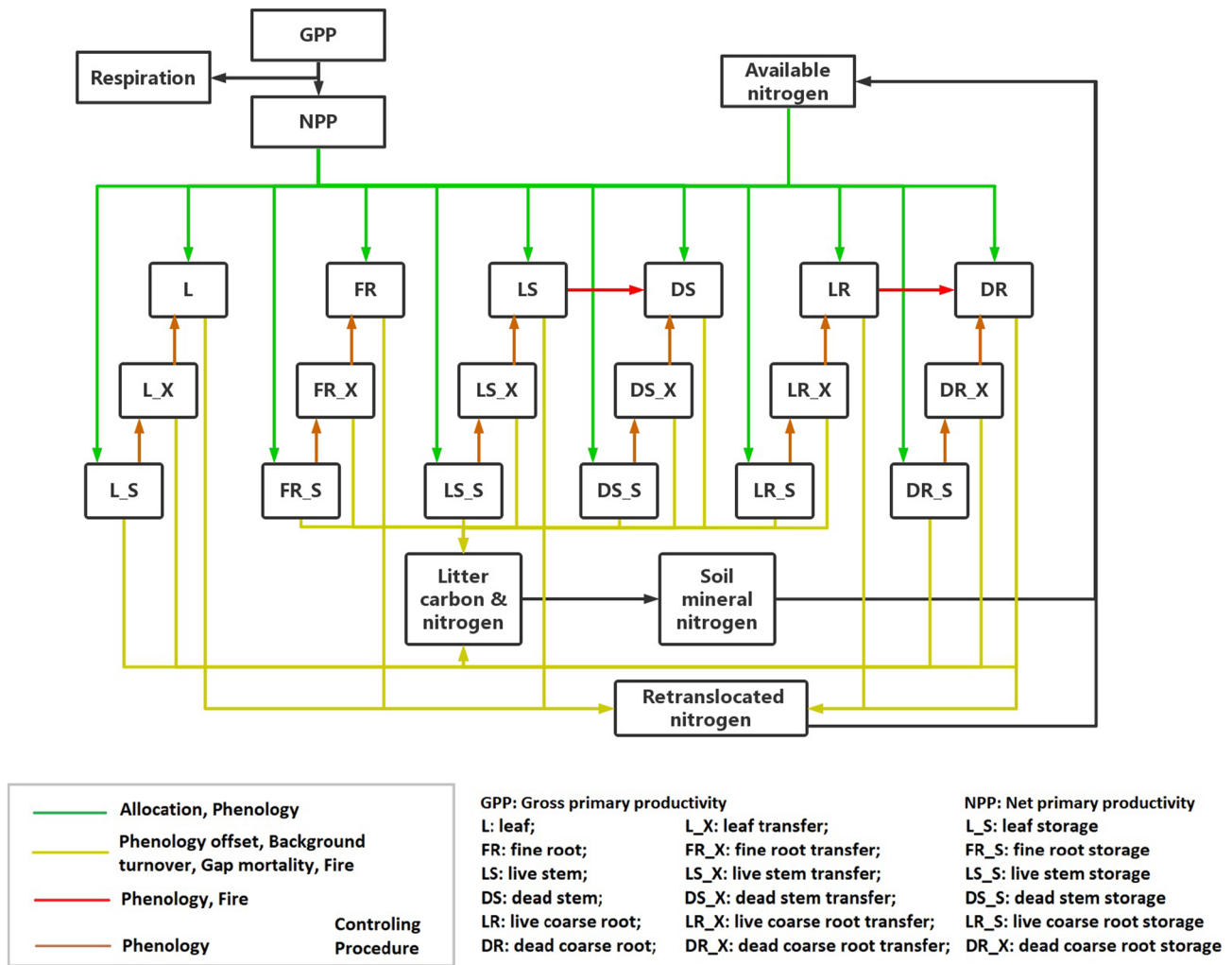
Figure 1. Carbon and nitrogen flow diagram of vegetation biogeochemical cycle. GPP stands for gross primary production. NPP stands for net primary production.

soil into four matrix equations and, more importantly, makes the matrix model executable with CLM5 and Community Earth System Model version 2 (CESM2). We first describe the mathematic details of the matrix model and numerical calculation using a sparse matrix technique for forward modeling. Then, we verify the CLM5 matrix model to exactly reproduce modeled global carbon and nitrogen dynamics in the original model. We use a technical test package developed by National Center for Atmospheric Research (NCAR) to ensure that the matrix carbon and nitrogen modules 100% meet the technical standards set by and can be compiled and executed with, CLM5 and the Community Earth System Model version 2 (CESM2). We also evaluate and compare the computational cost of CLM5 with and without the matrix model. In addition, we show the diagnostic capability that the matrix model uniquely offers to help understand and interpret simulation results of global carbon dynamics.

2. Materials and Methods

2.1. Biogeochemical Modules of the Original CLM5

The CLM5 biogeochemistry module includes carbon cycle and nitrogen cycle for aboveground and belowground processes (Lawrence et al., 2019). The vegetation biogeochemical cycle in CLM5 contains 18 carbon pools, including six tissue pools: leaf, fine root, live stem, dead stem, live coarse root, and dead coarse root (Figure 1 and Table 1). Each tissue pool is accompanied by a storage pool and a transfer pool. The vegetation nitrogen cycle includes one additional pool, a retranslocation pool to store reusable

Table 1
Vegetation Pool Index and Name

Pool index	Pool name	Pool index	Pool name	Pool index	Pool name
1	Leaf	2	Leaf storage	3	Leaf transfer
4	Fine root	5	Fine root storage	6	Fine root transfer
7	Live stem	8	Live stem storage	9	Live stem transfer
10	Dead stem	11	Dead stem storage	12	Dead stem transfer
13	Fine coarse root	14	Fine coarse root storage	15	Fine coarse root transfer
16	Dead coarse root	17	Dead coarse root storage	18	Dead coarse root transfer
19	Grain	20[a]	Grain storage	21[a]	Grain transfer
22[b]/19[b]	Retranslocation				

[a]The crop module is turned on.   [b]The retranslocation pool is only activated for N matrix equation.

nitrogen from litter fall. A crop grain tissue pool, accompanied by a grain storage pool and a grain transfer pool, is added when the crop model is used. The belowground module of CLM5 has 20 soil layers by a default setting. Each layer contains seven pools for organic carbon and organic nitrogen, respectively, in metabolic litter, cellulose litter, lignin litter, coarse woody debris, fast soil organic matter, slow soil organic matter, and passive soil organic matter (Table 2). Thus, there are a total of 140 soil organic carbon pools (7 pools × 20 layers) and 140 soil organic nitrogen pools. Inorganic nitrogen pools, such as ammoniacal nitrogen and nitrate nitrogen pools, and related processes such as leaching, nitrification, denitrification, atmospheric nitrogen decomposition, and biological nitrogen fixation have all been preserved in original CLM5 biogeochemistry cycle modules but not reorganized into the matrix equation for soil nitrogen cycle.

Changes in vegetation carbon and nitrogen pool sizes are controlled by many processes. Photosynthesis, plant respiration, and plant nitrogen uptake, for example, control the net carbon and nitrogen input into vegetation carbon and nitrogen pools. Net carbon and nitrogen input are allocated to different storage pools and tissue pools of vegetation. Reusable nitrogen before litter falling is first stored in retranslocation pool, and then transferred to storage pools or display tissue pools for plant growth.

Vegetation carbon and nitrogen dynamics are also controlled by phenology for deciduous plant functional types (PFTs) while evergreen PFTs have constant leaf turnover rates. Plant phenology determines the leaf onset and offset dates according to air temperature and soil water conditions. During the onset period, plant carbon is transferred from storage pools through transfer pools to tissue pools within a half month, so that leaf area index grows rapidly. In the offset period, leaf and fine root carbon is lost to litterfall and leaf area index drops quickly. In addition, harvest, natural death, and fire alter the vegetation biogeochemical cycles. Harvest is included when land use change occurs according to a scheme of Land Use Harmonized version 2 (LUH2). LUH2 was used in Land Use Model Intercomparison Project (LUMIP; http://cmip.ucar.edu/lumip) as part of Coupled Model Intercomparison Project 6 (CMIP6). Part of plant carbon is removed from ecosystems and part goes to litter pools according to the harvest rate as a function of the PFT transition area. Conservation in mass, energy, carbon, and nitrogen has been considered to reconcile the change in area. Natural death occurs as a result of aging and is applicable to all plant functional types. Carbon moves from plant pools litter pools at defined mortality rates. The fire module triggers the occurrence of occasional fire events based on the amount of the fuel (i.e., litter) and the soil moisture. When fire occurs, carbon and

Table 2
Soil Component Index and Name

Pool index	Pool name	Pool index	Pool name	Pool index	Pool name
1–20	Metabolic litter	21–40	Cellulose litter	41–60	Lignin litter
61–80	Coarse woody debris	81–100	Fast soil organic matter	101–120	Slow soil organic matter
121–140	Passive soil organic matter				

nitrogen in plant pools are partly released to the atmosphere and partly transferred to the litter pools based on their tissue quality and burned area.

Following the vegetation biogeochemical cycles, soil carbon, and nitrogen cycles are controlled by various soil processes in CLM5. Carbon and nitrogen transfers in soil biogeochemical cycles are related to soil and litter decomposition, vertical mixing, mineralization, immobilization, atmospheric nitrogen deposition, biological nitrogen fixation, nitrification, denitrification, leaching, and fire.

In this study, we focus on reorganization of the organic carbon and nitrogen balance equations into a matrix form. As a consequence, code structure and the update of state variables within each time step are changed. All the other processes, such as decomposition, mineralization, immobilization, fire, nitrogen deposition, biological fixation, nitrification, and denitrification, remain intact as in the original code without any changes in this study.

2.2. Matrix Representation of the Vegetation Carbon and Nitrogen Cycles

In the new matrix representation, we reorganize CLM5 vegetation carbon and nitrogen balance equations into two matrix equations:

$$\frac{dC_{veg}}{dt} = BI_{Cin} + \left(A_{phc}(t)K_{phc} + A_{gmc}(t)K_{gmc} + A_{fic}(t)K_{fic}\right)C_{veg}(t) \tag{1}$$

$$\frac{dN_{veg}}{dt} = BI_{Nin} + \left(A_{phn}(t)K_{phn} + A_{gmn}(t)K_{gmn} + A_{fin}(t)K_{fin}\right)N_{veg}(t) \tag{2}$$

C_{veg} and N_{veg} are two time-dependent state variables, which are n entry vectors, each representing its respective vegetation pool size (g C m^{-2} and g N m^{-2}). I_{Cin} and I_{Nin} are scalars for carbon and nitrogen input, respectively. Carbon input is the net primary productivity, which is the difference between gross primary productivity and autotrophic respiration. Nitrogen input for vegetation nitrogen cycle includes both nitrogen fixation and uptake. B is also a n entry vector, representing allocation fraction of plant carbon or nitrogen input to individual pools. K is an $n \times n$ diagonal matrix. Its subscripts ph, gm, and fi indicate phenology, gap mortality (i.e., harvest from land use and natural mortality), and fire processes, respectively. The diagonal entries are the exit rates of all vegetation pools due to phenology (K_{ph}), gap mortality (K_{gm}), and fire (K_{fi}), respectively, as described by

$$K_{ph} = \begin{pmatrix} k_{p1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{pn} \end{pmatrix} \tag{3}$$

$$K_{gm} = \begin{pmatrix} k_{n1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{nn} \end{pmatrix} \tag{4}$$

$$K_{fi} = \begin{pmatrix} k_{f1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{fn} \end{pmatrix} \tag{5}$$

The exit rates in plant phenology matrix K indicate the leaf, root, live stem, and dead stem turnover due to phenology processes. The exit rates in gap mortality matrix K include harvest rates from land use plus natural mortality. The exit rates in fire matrix K represent the plant carbon loss rate due to the fire.

A is a transfer coefficient matrix, representing carbon and nitrogen transfer among pools as specified in (Equations 6–11) below for CLM5. Subscripts c and n are for carbon and nitrogen.

$$A_{phc} = \begin{pmatrix}
-1 & 0 & a_{1,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & a_{3,2} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & a_{4,6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a_{6,5} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & a_{7,9} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{9,8} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_{10,7} & 0 & 0 & -1 & 0 & a_{10,12} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{12,11} & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & a_{13,15} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{15,14} & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{16,13} & 0 & 0 & -1 & 0 & a_{16,18} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{18,17} & -1
\end{pmatrix} \quad (6)$$

$$A_{phn} = \begin{pmatrix}
-1 & 0 & a_{1,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{1,19} \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{2,19} \\
0 & a_{3,2} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{3,19} \\
0 & 0 & 0 & -1 & 0 & a_{4,6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{4,19} \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{5,19} \\
0 & 0 & 0 & 0 & a_{6,5} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{6,19} \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & a_{7,9} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{7,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{8,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{9,8} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{9,19} \\
0 & 0 & 0 & 0 & 0 & 0 & a_{10,7} & 0 & 0 & -1 & 0 & a_{10,12} & 0 & 0 & 0 & 0 & 0 & 0 & a_{10,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{11,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{12,11} & -1 & 0 & 0 & 0 & 0 & 0 & 0 & a_{12,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & a_{13,15} & 0 & 0 & 0 & a_{13,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & a_{14,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{15,14} & -1 & 0 & 0 & 0 & a_{15,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{16,13} & 0 & 0 & -1 & 0 & a_{16,18} & a_{16,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & a_{17,19} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{18,17} & -1 & a_{18,19} \\
a_{19,1} & 0 & 0 & a_{19,4} & 0 & 0 & a_{19,7} & 0 & 0 & a_{19,10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix} \quad (7)$$

$$A_{gmc} = \begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix} \quad (8)$$

$$
A_{gmn} = \begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix}
\tag{9}
$$

$$
A_{fic} = \begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_{10,7} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{16,13} & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix}
\tag{10}
$$

$$
A_{fin} = \begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_{10,7} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{16,13} & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix}
\tag{11}
$$

The off-diagonal entry, $a_{i,j}$, for matrix A represents a fraction of carbon or nitrogen leaving pool j that goes to pool i. The diagonal entry is set to -1 to represent that all the exiting carbon leaves pool j. The pool names referred by the subscript i or j can be found in Table 1.

Interactions between vegetation carbon and nitrogen cycles in the original CLM5 are fully preserved in CLM5 matrix model. The original CLM5 has two modules to regulate carbon and nitrogen interactions for vegetation. First, the photosynthetic capacity, an important variable driving the carbon cycle, interacts with nitrogen cycle in the LUNA module. LUNA optimizes nitrogen allocation to maximize the daily net photosynthetic carbon gain. Second, plant nitrogen uptake interacts with the carbon assimilation in the FUN module. FUN is based on the concept that nitrogen uptake requires the expenditure of energy in the form of exchange for carbon. Both the modules are fully preserved to drive the carbon and nitrogen interactions in the CLM5 vegetation matrix model.

2.3. Matrix Representation of CLM5 Soil Carbon and Nitrogen Cycles

The matrix representation of CLM5 soil biogeochemical cycle follows the previous work in CLM4.5 (Huang, Lu, et al., 2018). The new implementation in this study extends the maximum soil layers from a fixed value of 10 in CLM4.5 to a default value of 20 with flexibility to change in CLM5. In addition, we convert nitrogen balance equations in the original CLM5 to a matrix representation. The soil organic carbon and nitrogen transfer among soil pools is formulated by following matrix equations:

$$\frac{dC_{soil}}{dt} = I_{Csoil} + \left(A_{hc}\xi(t)K_h + V(t) + K_f(t)\right)C_{soil}(t) \tag{12}$$

$$\frac{dN_{soil}}{dt} = I_{Nsoil} + \left(A_{hn}\xi(t)K_h + V(t) + K_f(t)\right)N_{soil}(t) \tag{13}$$

As with vegetation, C_{soil} and N_{soil} are two vectors of state variables representing soil organic carbon and nitrogen pool sizes in g C m^{-3} and g N m^{-3}, respectively. I_{Csoil} and I_{Nsoil} are vectors representing plant litterfall into different litter carbon and nitrogen pools, respectively. A_{hc} and A_{hn} represent the horizontal transfers of carbon and nitrogen, respectively, which means transfers among pools within one soil layer. V stands for the rate of vertical mixing between the same types of pools across soil layers, which is the same for carbon and nitrogen. K_f is the rate of fire-induced litter loss, which is the same for carbon and nitrogen.

Matrix A_h, including both A_{hc} and A_{hn}, is a block matrix constituted by several matrices as

$$A_{hc} \text{ or } A_{hn} = \begin{pmatrix} A_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 & 0 & 0 & 0 \\ A_{31} & 0 & A_{33} & 0 & 0 & 0 & 0 \\ A_{41} & 0 & 0 & A_{44} & 0 & 0 & 0 \\ 0 & A_{52} & A_{53} & 0 & A_{55} & A_{11} & A_{11} \\ 0 & 0 & 0 & A_{64} & A_{65} & A_{66} & 0 \\ 0 & 0 & 0 & 0 & A_{75} & A_{76} & A_{77} \end{pmatrix} \tag{14}$$

Each of the matrices A_{ij} within the block matrix A_{hc} or A_{hn} is 20 by 20, corresponding to 20 soil layers. These nonzero, off-diagonal matrices A_{ij}, $i \neq j$, indicate carbon and nitrogen transfer among pools within one soil layer as

$$
A_{ij} = \begin{cases}
\begin{bmatrix} -1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -1 \end{bmatrix} & i = j;\ \text{C and N cycles} \\[30pt]
\begin{bmatrix} (1-r_{ij})T_{ij} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1-r_{ij})T_{ij} \end{bmatrix} & i \neq j;\ \text{C cycle } (A_{hc}) \\[30pt]
\begin{bmatrix} (1-r_{ij})T_{ij}\dfrac{CN_{j,1}}{CN_{i,1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1-r_{ij})T_{ij}\dfrac{CN_{j,20}}{CN_{i,20}} \end{bmatrix} & i \neq j;\ \text{N cycle } (A_{hn})
\end{cases}
\tag{15}
$$

Each of the diagonal matrices A_{ii} is a negative identical matrix. The off-diagonal matrices A_{ij} have non-zero diagonal values, $(1-r_{ij})T_{ij}$ for carbon and $(1-r_{ij})T_{ij}\dfrac{CN_{j,k}}{CN_{i,k}}$ for nitrogen. A_{ij} represents transfer coefficients. The parameter r_{ij} is the respired fraction of carbon along the transfer pathway from pool j to i. T_{ij} represents a pathway fraction of carbon going to pool i from that leaving the jth pool due to decomposition. $CN_{j,k}$ represents carbon and nitrogen ratio of pool j in layer k.

The diagonal matrices K_h and K_f indicate the turnover rates, respectively, due to decomposition (horizontal transfer) and fire, at different layers:

$$
K_h = \begin{pmatrix} k_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_n \end{pmatrix}
\tag{16}
$$

$$
K_f = \begin{pmatrix} k_{f1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{fn} \end{pmatrix}
\tag{17}
$$

Environmental scalars ξ are time-dependent variables and are the product of temperature scalar ξ_T, water scalar ξ_W, oxygen scalar ξ_O, depth scalar ξ_D, and nitrogen scalar ξ_N:

$$
\xi(t) = \xi_T(t)\xi_W(t)\xi_O\xi_D\xi_N(t)
\tag{18}
$$

The vertical mixing coefficient matrix V is made up of six identical matrices v:

$$
V(t) = \begin{pmatrix}
v & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & v & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & v & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & v & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & v & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & v
\end{pmatrix}
\tag{19}
$$

Note that vertical mixing of coarse woody debris (CWD) is not allowed in CLM5; therefore, the corresponding vertical mixing matrix is 0 for CWD. The matrix v is a tridiagonal matrix, indicating the vertical mixing only transfers between adjacent layers:

$$
v = \mathrm{diag}(dz_1, dz_2, \ldots dz_{20})^{-1}
\begin{pmatrix}
g_1 & -g_1 & 0 & 0 & 0 & 0 \\
-h_2 & h_2 + g_2 & -g_2 & \cdots & 0 & 0 & 0 \\
0 & -h_3 & h_3 + g_3 & & 0 & 0 & 0 \\
& \vdots & & \ddots & & \vdots & \\
0 & 0 & 0 & & h_{18} + g_{18} & -g_{18} & 0 \\
0 & 0 & 0 & \cdots & -h_{19} & h_{19} + g_{19} & -g_{19} \\
0 & 0 & 0 & & 0 & -h_{10} & h_{20}
\end{pmatrix}
\tag{20}
$$

The subscripts represent the soil layer. g and h are vertical mixing rates related to upward and downward transfers.

As the same for the vegetation part, interactions between soil carbon and nitrogen cycles in the original CLM5 are fully preserved in the CLM5 soil matrix model. In the original CLM5, nitrogen limits soil organic carbon decomposition. The nitrogen limitation is represented by the ratio between available mineral nitrogen and the total soil nitrogen demand. The soil nitrogen demand includes soil immobilization during soil decomposition and plant uptake demand. The dynamics of mineral nitrogen processes that are involved in carbon and nitrogen interactions can be represented by one equation as done by Shi et al. (2016). The equation on mineral nitrogen dynamics can be coupled with the matrix equations on organic carbon and nitrogen processes to analytically or semianalytically explore their interactions.

2.4. Calculation of Storage Capacity and Storage Potential

One benefit of the matrix approach is expanded diagnostic capability. The transient dynamics of ecosystem carbon and nitrogen are interpretable from storage capacity and storage potential (Huang, Lu, et al., 2018; Jiang et al., 2017; Luo et al., 2017; Zhou et al., 2018). Theoretically, the storage capacity and potential can be calculated at any time point (Luo et al., 2017). For the purpose of diagnostics, we only calculate the annual average of vegetation storage capacity for both carbon and nitrogen from annual cumulative carbon and nitrogen inputs ($I_{Cveg,ann}$ and $I_{Nveg,ann}$) and yearly transfer coefficients ($\Pi_{ann,veg,C}$ and $\Pi_{ann,veg,N}$) as

$$
C_{\mathrm{cap,veg,ann}} = \left(\Pi_{ann,veg,C}\right)^{-1} I_{Cveg,ann}
\tag{21}
$$

$$
N_{\mathrm{cap,veg,ann}} = \left(\Pi_{ann,veg,N}\right)^{-1} I_{Nveg,ann}
\tag{22}
$$

$I_{Cveg,ann}$ and $I_{Nveg,ann}$ are vectors to allocate carbon and nitrogen uptake to different vegetation pools. The units are g C m^{-2} year^{-1} and g N m^{-2} year^{-1}, respectively. $\Pi_{ann,veg,C}$ for vegetation carbon and $\Pi_{ann,veg,N}$ for vegetation nitrogen are matrices of annual transfer coefficients representing annual transfer rate (in unit of year^{-1}) from one pool to another (nondiagonal entries) or exit rate leaving one pool (diagonal entries) as

$$
\Pi_{ann,veg,C} = \sum_{t=1}^{n=365 \times 48} \left(A_{phc}(t)K_{phc} + A_{gmc}(t)K_{gmc} + A_{fic}(t)K_{fic}\right) C_{veg,d}(t) C_{veg,d}^{-1}(t=1)
\tag{23}
$$

$$
\Pi_{ann,veg,N} = \sum_{t=1}^{n=365 \times 48} \left(A_{phn}(t)K_{phn} + A_{gmn}(t)K_{gmn} + A_{fin}(t)K_{fin}\right) N_{veg,d}(t) N_{veg,d}^{-1}(t=1)
\tag{24}
$$

where $C_{veg,d}(t)$ and $N_{veg,d}(t)$ are diagonal matrices of vegetation carbon and vegetation nitrogen, respectively. $C_{veg,d}^{-1}(t)$ and $N_{veg,d}^{-1}(t)$ are inverse diagonal matrices of vegetation carbon and vegetation nitrogen, respectively. Equations 23 and 24 calculate the annual transfer rate as annually accumulated transferred/exited flux divided by the initial pool size of this year.

Similarly, the annual average of the soil storage capacity is calculated from annual carbon and nitrogen inputs ($I_{Csoil,ann}$ and $I_{Nsoil,ann}$) and yearly transfer coefficients ($\Pi_{ann,soil,C}$ and $\Pi_{ann,soil,N}$):

$$
C_{\mathrm{cap,soil,ann}} = \left(\Pi_{ann,soil,C}\right)^{-1} I_{Csoil,ann}
\tag{25}
$$

$$
N_{\mathrm{cap,soil,ann}} = \left(\Pi_{ann,soil,N}\right)^{-1} I_{Nsoil,ann}
\tag{26}
$$

$I_{Csoil,ann}$ and $I_{Nsoil,ann}$ are vectors representing annual accumulated carbon and nitrogen litter fall to different soil layers and pools. The units are g C m^{-3} year^{-1} and g N m^{-3} year^{-1}. $\Pi_{ann,soil,C}$ for soil carbon, and $\Pi_{ann,soil,N}$ for soil nitrogen are calculated as

$$\Pi_{ann,soil,C} = \sum_{t=1}^{n=365 \times 48} \left(A_{hc}\xi(t)K_h + V(t) + K_f(t)\right)C_{soil,d}(t)C_{soil,d}^{-1}(t=1) \tag{27}$$

$$\Pi_{ann,soil,N} = \sum_{t=1}^{n=365 \times 48} \left(A_{hn}\xi(t)K_h + V(t) + K_f(t)\right)N_{soil,d}(t)N_{soil,d}^{-1}(t=1) \tag{28}$$

where $C_{soil,d}(t)$ and $N_{soil,d}(t)$ are diagonal matrices of soil carbon and soil nitrogen. $C_{soil,d}^{-1}(t)$ and $N_{soil,d}^{-1}(t)$ are inverse diagonal matrices of soil carbon and soil nitrogen.

Based on Luo et al. (2017), the storage potential is the storage capacity minus the storage, representing the amount of additional C that an ecosystem can store to reach the storage capacity

$$C_{pot,veg,ann} = C_{cap,veg,ann} - C_{veg} \tag{29}$$

$$N_{pot,veg,ann} = N_{cap,veg,ann} - N_{veg} \tag{30}$$

$$C_{pot,soil,ann} = C_{cap,soil,ann} - C_{soil} \tag{31}$$

$$N_{pot,soil,ann} = N_{cap,soil,ann} - N_{soil} \tag{32}$$

2.5. Code Structure and Update of State Variables

Implementation of the matrix model helps modularize the code to more clearly describe the complex carbon and nitrogen transfer network in CLM5. The original code calculates the dynamics of carbon and nitrogen individually from different processes, such as phenology, gap mortality (i.e., plant death), harvest, fire, soil decomposition, and vertical mix in six subroutines. Updates of state variables are carried out stepwise in eight subroutines (Figure 2). With the matrix formulation, entries in matrix Equations 1, 2, 12, and 13 are taken from subroutines of the original CLM5. For example, all the entries for term $A_{phc}(t)$ of Equation 1 (as expressed in Equation 6) are calculated in the phenology subroutine and passed to the vegetation matrix module. Once all the entries from the different subroutines are calculated and passed to the vegetation and soil matrix modules, the matrix model updates the carbon and nitrogen state variables once per time step (Figure 2). The matrix model then calculates new diagnostic variables such as those described in section 2.4.

A namelist switch is implemented, which allows users to use either the matrix code or the original code. When the matrix code is on, state variables update once per time step in the matrix modules. When the matrix code is off, state variables update three times, respectively, after phenology and soil decomposition, gap mortality and harvest, and fire, stepwise per time step in the original CLM5 modules.

2.6. Sparse Matrix

Since the matrix equations include all the carbon and nitrogen pools of plants and soil layers, the plant transfer matrix is at least 18 by 18 and the soil transfer matrix is 140 by 140 for carbon, with similarly sized matrices for nitrogen. The solution of these large matrix equations is computationally expensive. However, the number of nonzero entries in A_{hc}/A_{hn}, K_h, V, and K_f of the soil matrix equations account for only 3.0%, 0.2%, 3.2%, and 0.2% of all the entries, and therefore these matrices can be considered sparse matrices. Since the time spent to access a matrix variable increases exponentially with the matrix size, the computational cost to operate the soil matrix in 140 by 140 is 300% more than that for the original CLM5. This is unacceptable for the code structure development. To improve the computation efficiency for CLM5 matrix modules, we use a sparse matrix solution method.

2.6.1. Sparse Matrix Format

We use a coordinate compact format to store the sparse matrix. The format records the rows, columns, and values of nonzero entries in three vectors with the same order. The sparse matrix formats use smaller memory for soil matrix equations (A_{hc}/A_{hn}, K_h, V, and K_f memory reduced by 91%, 99.4%, 90.4%, and 99.4%, respectively). The nonzero entries are recorded in a matrix from top to bottom and then from left to right
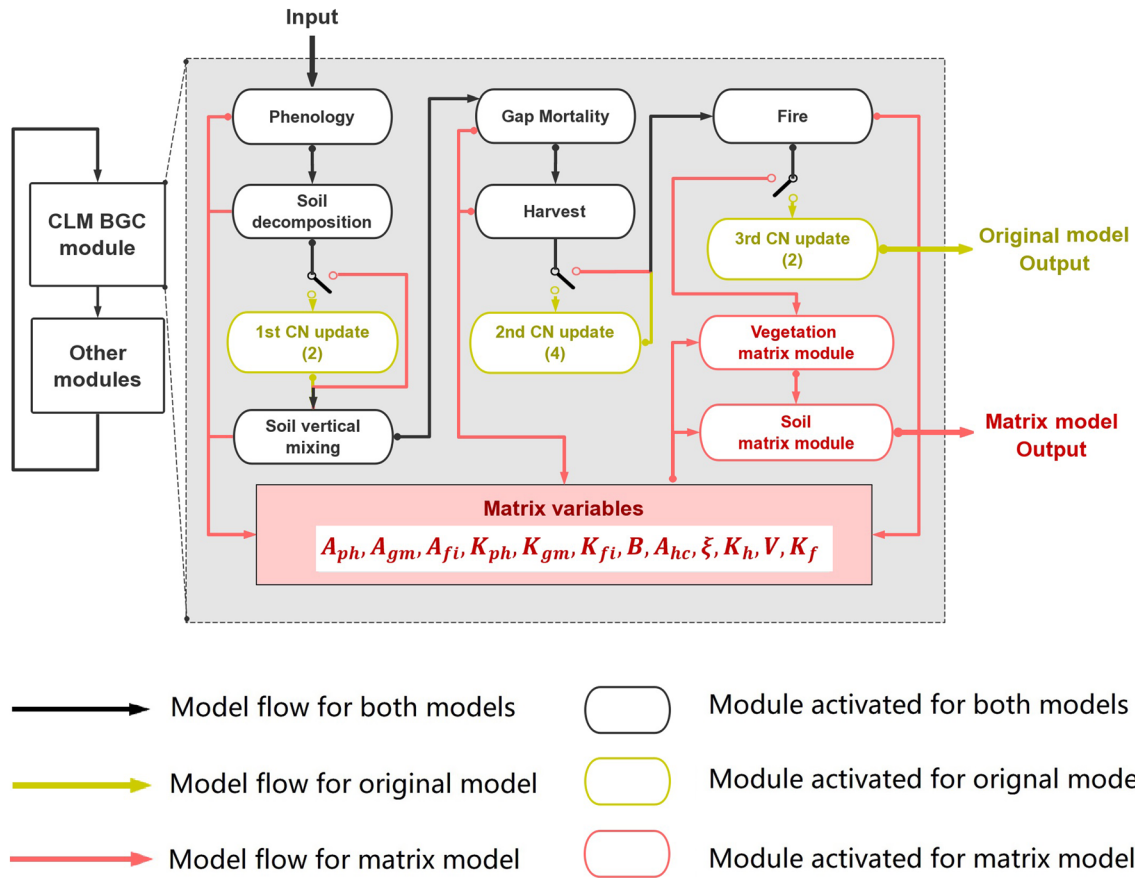
Figure 2. Schematic representation of CLM5.0 matrix model. The switch indicates whether to turn on or off the matrix model. The number in the yellow box indicates the numbers of updating subroutines.

columns. These coordinates of all the sparse matrices in CLM5 are shown in Tables S1–S4 in the supporting information.

In Equations 1, 2, 12, and 13, we save A_{phc}, A_{phn}, A_{gmc}, A_{gmn}, A_{fic}, A_{fin}, A_{hc}, A_{hn}, V, and K_f in the coordinate compact format. According to the format, the value of the ith nonzero entry in a sparse matrix A can be uniquely identified. Therefore, we notate them as $A(i)$. The row and column number of the ith nonzero entry can be also notated as $Row_A(i)$ and $Col_A(i)$. $A(i)$, $Row_A(i)$, and $Col_A(i)$ together uniquely define a sparse matrix. In those four equations, there are also seven diagonal matrices, K_{phc}, K_{phn}, K_{gmc}, K_{gmn}, K_{fic}, K_{fin}, and K_h. Those diagonal matrices K are converted to vectors using the diagonal vector format. The diagonal entry in row i can be notated as $K(i)$.

2.6.2. Matrix Operation in the Sparse Matrix Format

Generally, operation of matrix Equations 1, 2, 12, and 13 includes multiplication and addition. The matrix multiplication occurs between (1) coordinate compact format and diagonal vector format and (2) coordinate compact format and vector format.

Sparse matrix in coordinate compact format is multiplied by diagonal matrix in diagonal vector format, such as matrix A with matrix K in Equations 1, 2, 12, and 13. Since K is a diagonal matrix, the locations of nonzero entries in the production AK matrix should be the same with A. Thus, we have

$$AK(i) = A(i)K(Col_A(i)); \tag{33}$$

$$Col_{AK}(i) = Col_A(i); \tag{34}$$

$$Row_{AK}(i) = Row_A(i); \tag{35}$$

where $AK(i)$ represents the ith nonzero entry in the production AK, $Col_{AK}(i)$ is the column coordinate of the ith nonzero entry in the sparse matrix AK, and $Row_{AK}(i)$ is the row coordinate of the ith nonzero entry in the sparse matrix AK.

When matrix A multiply vector C in Equations 1, 2, 12, and 13, the production AC is also a vector. Thus, we have

$$AC(i) = \sum_{j=1}^{n, \, Row_A(j)=i} A(j) \, C \, (Col_A(j)) \tag{36}$$

where $AC(i)$ and $C(i)$ represent the ith entry in the vector AC and C. n is the size of the matrix A, which also equals the number of pools, for example, $n = 140$ for soil carbon cycle module in CLM5.

The matrix addition is carried out for (1) two coordinate compact formats when there is no fire and (2) three coordinate compact formats when fire is on. Sparse matrix addition operation is more complicated than sparse matrix multiplication, since the location of the nonzero entries in addition is not easily predicted. If the locations of nonzero entries in to-be-added matrices (e.g., $A_{phc}K_{phc}$, and $A_{gmc}K_{gmc}$) does not change for any specific addition (e.g., $A_{phc}K_{phc} + A_{gmc}K_{gmc}$), the addition should follow a constant mapping function. This is because the coordinates order in addition matrix should be exclusive in the coordinate matrix format. Therefore, we save the constant mapping function to avoid repeated calculation.

Two sparse matrices addition are used in coordinate compact format for the second term in Equations 1, 2, 12, and 13 when there is no fire. Both the to-be-added matrices (A and B) and addition matrix ($A\&B$) are sparse matrices in coordinate compact format. The nonzero entries in $A\&B$ should be no less than matrix A and B. The ith nonzero entry in $A\&B$ can be contributed from A, B or both A and B. The mapping function $j = I_{A \to AB}(i)$ records the mapping relation between jth nonzero entry in matrix A and the ith nonzero entry in matrix $A\&B$. Similarly, $k = I_{B \to AB}(i)$ records the mapping relation between kth nonzero entry in matrix B and the ith nonzero entry in matrix $A\&B$:

$$A\&B(i) = A(I_{A \to AB}(i)) + B(I_{B \to AB}(i)) \tag{37}$$

Three sparse matrices addition is done for the second term in Equations 1, 2, 12, and 13 when fire is on. All the denotations are similar with Equation 37. The addition is also based on the mapping function $I_{A \to ABC}$, $I_{B \to ABC}$, and $I_{C \to ABC}$ from added matrices (A, B, and C) to addition matrix (A&B&C):

$$A\&B\&C(i) = A(I_{A \to ABC}(i)) + B(I_{B \to ABC}(i)) + C(I_{C \to ABC}(i)) \tag{38}$$

2.7. Validation and Efficiency Test of the CLM5 Matrix Model

To validate the matrix model, we run global historical (1850–2014) simulations with rising atmospheric CO_2 and warming climate by two versions of CLM5 with the matrix or the original modules at a 4° × 5° resolution to compare their results. The CLM5 model simulations are conducted on Cheyenne system (2019). Simulations are driven by reanalysis meteorological forcing, the Global Soil Wetness Project Phase 3 data set (GSWP3) (Dirmeyer et al., 2006) ranged from 1901 to 2014. Meteorological forcing from 1901 to 1920 are recursively used in simulation period from 1850 to 1900. The atmospheric CO_2 concentration forcing uses the prescribed data from Climate Model Intercomparison Project Phase 6 (CMIP6) historical scenario (Eyring et al., 2016; Meinshausen et al., 2017). Plant hydraulics, land use, crop, and fire modules are turned on. For the nitrogen modules, the Fixation and Update of Nitrogen (FUN) model and Leaf Use of Nitrogen for Assimilation (LUNA) model are turned on.

To test the efficiency of the matrix model, we use the same spatial resolution and meteorological forcing as for the validation experiment. But we only run the model for one month for the efficiency test. Comparison of the efficiency are made by tracking the elapsed time in each subroutine.

2.8. Traceability Analysis

To demonstrate the diagnostic capability, we conduct a traceability analysis of the soil carbon storage from two steady state simulations using different reanalysis data, (1) GSWP3, and (2) CRUNCEP version 7 (CRUNCEPv7) (Viovy, 2016) as the meteorological forcing. While it is feasible to conduct transient traceability analysis (Jiang et al., 2017; Zhou et al., 2018), we used this steady state example because a previous study

by Lawrence et al. (2019) has illustrated a huge difference in simulated soil carbon stock between the two forcing sets but could not fully identify the causes. The steady state of carbon cycle is achieved by spin-up of the CLM5 matrix model under the two sets of forcings. Twenty-year meteorological forcing (1901–1920) are recursively used for GSWP3 or CRUNCEPv7.

The traceability analysis (Luo et al., 2017; Xia et al., 2013) assumes that the soil carbon storage equals to soil carbon storage capacity at steady state. The soil carbon storage capacity is expressed by Equation 39 as

$$\overrightarrow{C_{cap,\,soil}} = \left(A_{hc}\xi(t)K_h + V(t) + K_f(t)\right)^{-1} \frac{\overrightarrow{I_{Csoil}}}{\left|\overrightarrow{I_{Csoil}}\right|} \left|\overrightarrow{I_{Csoil}}\right| \tag{39}$$

The soil carbon input $\overrightarrow{I_{Csoil}}$ is represented by a vector, including plant litterfall to 20 soil layers and 7 soil pools. We use $\left|\overrightarrow{I_{Csoil}}\right|$ to represent the total amounts of plant litterfall to soil. $\frac{\overrightarrow{I_{Csoil}}}{\left|\overrightarrow{I_{Csoil}}\right|}$ indicates the partition fraction of the total litterfall.

In the traceability framework, the soil carbons storage capacity can be decomposed into soil carbon input and soil carbon residence time. We use the total amounts of plant litterfall $\left|\overrightarrow{I_{Csoil}}\right|$ to indicate the soil carbon input. Soil carbon residence time is estimated by Equation 40 as

$$\tau = \left(A_{hc}\xi(t)K_h + V(t) + K_f(t)\right)^{-1} \frac{\overrightarrow{I_{Csoil}}}{\left|\overrightarrow{I_{Csoil}}\right|} \tag{40}$$

More complex than the traceability analysis for CABLE by Xia et al. (2013) is that CLM5 introduced the vertical transfer matrix $V(t)$ and fire matrix $K_f(t)$. The residence time as expressed in Equation 40 can be further decomposed into

$$\bar{\tau} = \left(I + (A_{hc}\xi(t)K_h)^{-1}\left(V(t) + K_f(t)\right)\right)^{-1} \xi^{-1}(t)(A_{hc}K_h)^{-1} \frac{\overrightarrow{I_{Csoil}}}{\left|\overrightarrow{I_{Csoil}}\right|} \tag{41}$$

Following the terminology by Xia et al. (2013), $\frac{(A_{hc}K_h)^{-1}\overrightarrow{I_{Csoil}}}{\left|\overrightarrow{I_{Csoil}}\right|}$ is the baseline residence time $\overrightarrow{\tau_{base}}$. $\xi(t)$ is the environmental scalar. I is an identity matrix. However, $I + (A_{hc}\xi(t)K_h)^{-1}(V(t) + K_f(t))$ is hard to be further decomposed and named as a process-related scalar, $\xi_{other}^{-1}(t)$. The closer $\xi_{other}^{-1}(t)$ is to 1, the more dominant the soil decomposition process is. Then, the matrix Equation 41 can be simplified to one dimensional equation:

$$|\tau| = \overline{\xi_{other}^{-1}(t)} \,\overline{\xi^{-1}(t)}|\tau_{base}| \tag{42}$$

$|\tau|$is the total soil carbon residence time, which is the products of three terms: averaged inverse of process-related scalars $\overline{\xi_{other}^{-1}(t)}$, averaged inverse of environmental scalars $\overline{\xi^{-1}(t)}$, and baseline residence time $|\tau_{base}|$. Thus, differences in the total ecosystem carbon storage between the two forcing data sets were attributed to seven traceable components, soil carbon input $\overrightarrow{I_{Csoil}}$, baseline residence time $\overrightarrow{\tau_{base}}$, temperature scalar $\frac{1}{\xi_T}$, water scalar $\frac{1}{\xi_W}$, oxygen scalar $\frac{1}{\xi_O}$, nitrogen scalar $\frac{1}{\xi_N}$, and process-related scalar $\frac{1}{\xi_{other}}$. Contribution from each component was estimated by $\frac{S_i}{\sum_i S_i} \times 100\%$. S_i represents the difference of the ith traceable component between two forcing sets. S_i is estimated from $S_i = \left|\ln\left(\frac{X_{i,\,GSWP}}{X_{i,\,CRU}}\right)\right|$, where $\frac{X_{i,\,GSWP}}{X_{i,\,CRU}}$ represents the ratio of the traceable components between two forcing sets. $X_{i,GSWP}$ and $X_{i,CRU}$ are traceable components from steady state simulations with GSWP and CRUNCEP-v7, respectively. In the global or regional analysis, each traceable component is represented by the spatial average $\overline{X}$. Details in the calculation can be found in Text S1.

3. Results

3.1. Global Validation of the Matrix Model to Carbon and Nitrogen Simulations

The temporal dynamics of carbon and nitrogen storage simulated by the matrix modules was compared with the dynamics by the original modules. The CLM5 matrix model and the original CLM5 use the same default initial value without spin-up for the transition simulation. Modeled carbon and nitrogen storage in total ecosystem, vegetation, and soil organic matter from the matrix model matched with those from the original CLM5 (Figure 3). The relative differences of soil carbon and nitrogen are less than 1% for most grid cells. Only 0.4% and 0.5% of the grid cells in vegetation carbon and nitrogen storage, respectively, diverged by more than 1% relative difference (Figures 3g and 3h). Only 2 of 1,466 global grid cells in ecosystem carbon and nitrogen storage diverged by more than 1%. Global annual means in all six carbon and nitrogen storages over 115 years perfectly lined on the 1:1 line (Figures 3a–3f). The good agreements demonstrate that the matrix model can reproduce both temporal dynamics and spatial patterns of carbon and nitrogen states from the original model.

It was worth repeating that the matrix modules update carbon and nitrogen state variables only once within each time step, whereas the original modules updated these state variables stepwise within each time step. As a consequence, simulated state variables may slightly differ between the two models. Nevertheless, the differences in modeled state variable due to different update methods of the two models were small enough not to generate notable differences as shown in global simulation of the carbon and nitrogen storage.

3.2. System Tests

To ensure no technique errors due to introduction of the matrix modules, we conducted system tests to verify four requirements: (1) the model should be able to run to completion, (2) results from direct run and restart run should be completely identical, (3) results should be independent on number of processors and threading, and (4) compilation with array bound check and floating pointer check should be successfully done. Those requirements had been verified under different configurations such as a single soil layer for biogeochemical state variables, crop model, fire model, flexible nitrogen (FUN) among others.

The policy of the CLM code development required any new modules to pass a suite of predefined tests before they can be merged into the trunk version of CLM. There were 148 tests in the test suites. For example, the test suites included basic smoke test (SMS) (i.e., a single run of the model) and exact restart test (ERS) to ensure that restart runs give bit-for-bit identical results to a continuous run. All 148 tests in the test suites passed with the new matrix modules. Details of the CLM system tests can be found on this site (https://github.com/ESCOMP/CTSM/wiki/System-Testing-Guide).

3.3. Computational Efficiency

The matrix version of CLM5 was slightly more computationally expensive due to the matrix operation and the calculation of additional diagnostic variables. We conducted two 1-month global 1.875° × 2.25° resolution simulations, with matrix on and off. Our test showed that the matrix version of CLM generates 26.1% additional computational cost using the sparse matrix algorithm. Without the sparse matrix algorithm, the computation cost of using the matrix modules was ~300% higher than that with the original modules. However, the matrix method has the advantage that it can accelerate spin-up by ~1 order of magnitude in comparison with the existing accelerated decomposition (AD) method. This spin-up approach has been successfully implemented into CABLE (Xia et al., 2012). The implementation into CLM5 will be described in another paper.

Approximately four-fifths of the additional computational cost incurred by the matrix modules was from matrix operation (i.e., 20.0% out of the 26.1% additional cost) (Figure 4). The matrix operation for the soil matrix module was more than 3 times that for the vegetation matrix module, due to the size of the soil matrix (140 × 140). The other 6.1% cost was from the matrix recording (i.e., 4.3%), and writing additional matrix diagnostic variables to the output files in the output module (1.2%). Since the test ran CLM5 for only one month, most of the writing cost was from writing the restart file. The relative contribution of rewriting cost may decrease as the model runs longer.
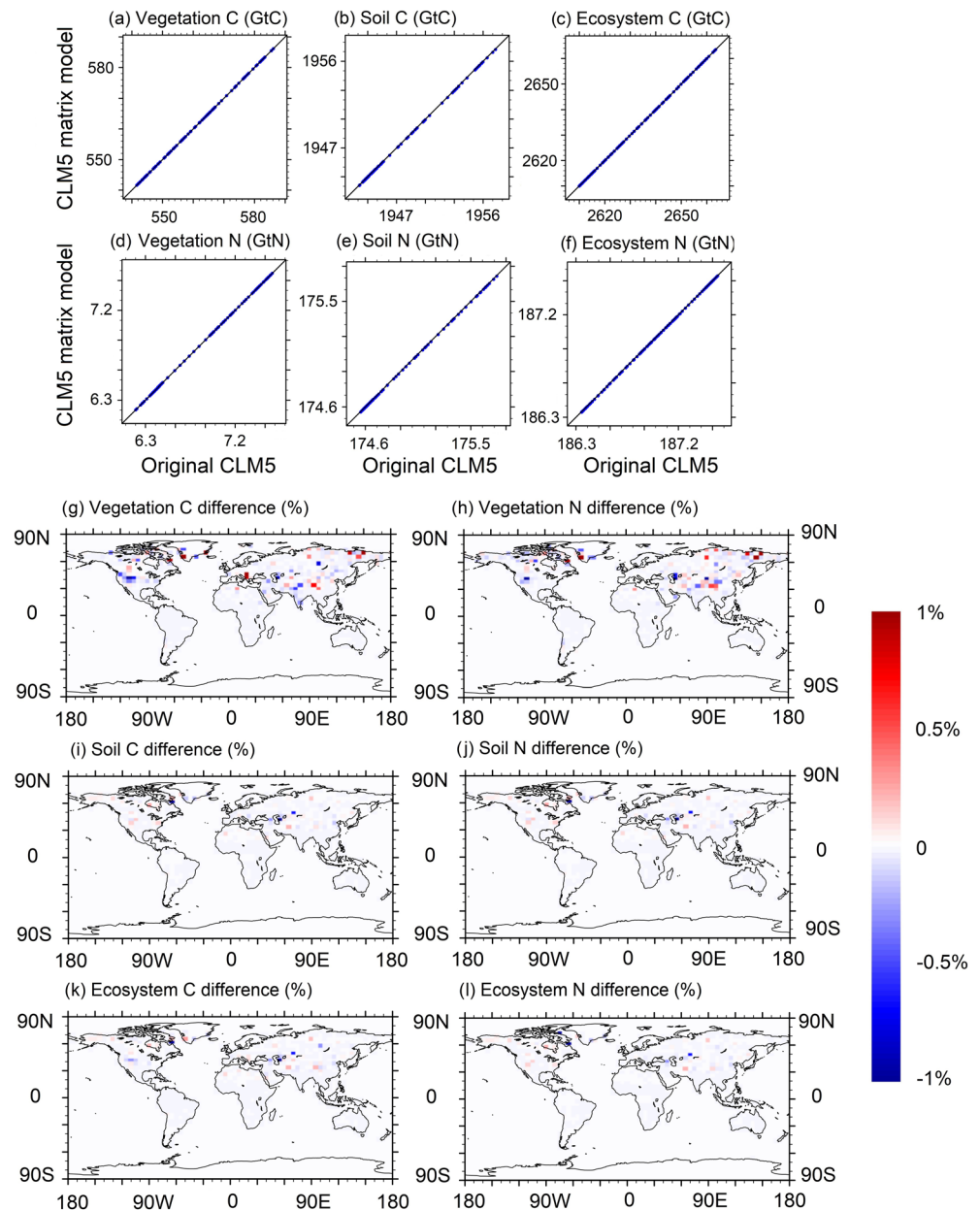
Figure 3. Comparison of simulated global carbon (C) and nitrogen (N) contents between the CLM5 matrix model and the original CLM5 from 1901 to 2014. (a) Vegetation carbon storage, (b) soil carbon storage, (c) ecosystem carbon storage, (d) vegetation nitrogen storage, (e) soil nitrogen storage, and (f) ecosystem nitrogen storage are summed up over all land grid cells each year for comparison. Relative differences averaged over the last four years (2011–2014) are calculated as $\frac{X_{matrix} - X_{original}}{X_{original}} * 100$. X_{matrix} represents (g) vegetation carbon storage, (h) vegetation nitrogen storage, (i) soil carbon storage, (j) soil nitrogen storage, (k) ecosystem carbon storage, and (l) ecosystem nitrogen storage from CLM5 matrix model. $X_{original}$ is the counterpart from the original CLM5.

3.4. Diagnostics of Global Biogeochemical Dynamics With Traceability Analysis

One major advantage of the matrix approach is to offer a more effective diagnostic capacity than the original biogeochemical models. To demonstrate the diagnostic capability of the matrix model, we compared soil carbon storage from two steady state simulations driven by GSWP3 and CRUNCEPv7 forcing data sets. Our results found a significantly higher soil carbon storage when the model is forced with CRUNCEPv7 data, especially in permafrost region at high latitudes (Figure 5). The high carbon storage in such regions can
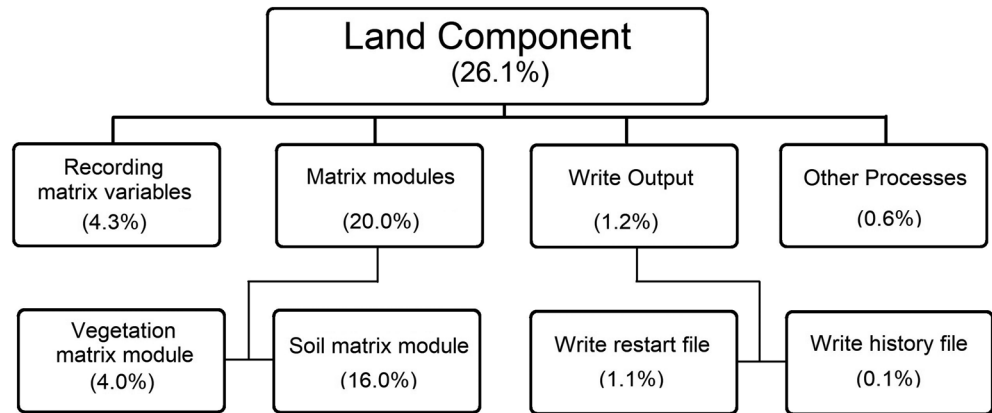
Figure 4. Additional computational cost (percentage in the parenthesis) by the CLM5 matrix model in comparison with the original CLM5. The computational cost test is conducted based on 1 month global CLM5 simulation ($1.875° \times 2.25°$) in 1901.

be over 250,000 g C m^{-2}. The rich carbon storage area simulated under CRUNCEPv7 is substantially larger than that under GSWP3. The differences in soil carbon storage can be over 50,000 g C m^{-2}.

To understand the causes of the large differences in soil carbon storage under the two sets of forcings, we conducted a traceability analysis. Our results showed the global soil carbon storage under CRUNCEPv7 (5,366 Gt C) more than doubled that under GSWP3 (2,462 Gt C). This difference in soil carbon storage was primarily attributed to a much longer soil carbon residence time with CRUNCEPv7 (74.4 years) than GSWP3 (38.0 years). The global soil carbon inputs played a minor role in different soil carbon storages between the two forcing data sets. The carbon input was 72.0 Gt C year^{-1} under CRUNCEPv7, which was just 11% higher than the 64.6 Gt C year^{-1} with GSWP3. The carbon residence time can be further decomposed into environmental scalars and baseline residence time. Overall, differences in carbon input, baseline
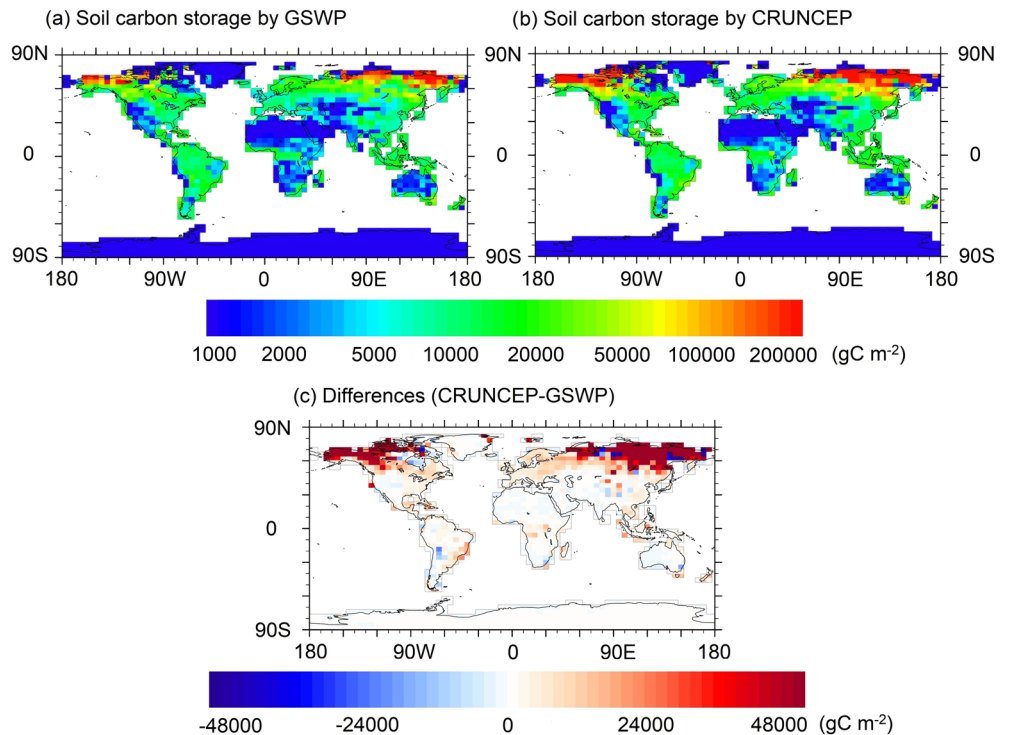
Figure 5. Global pattern of soil carbon storages at steady state are derived by CLM5 spin-up using 20-year recursive from (a) Global Soil Wetness Project Phase 3 (GSWP3) and (b) CRUNCEP v7. The differences in soil carbon storage is shown in (c). Positive indicates that soil carbon storage is greater when driven by CRUNCEP than by GSWP3.

Table 3

Traceability Analysis of Differences in Simulated Carbon Storage Between Two Forcing Data Sets, GSWP, and CRUNCEP-v7 to Seven Traceable Components (%)

Plant functional type	$\overline{I_{soil}}$	$\overline{\tau_{base}}$	$\overline{\dfrac{1}{\xi_T}}$	$\overline{\dfrac{1}{\xi_W}}$	$\overline{\dfrac{1}{\xi_O}}$	$\overline{\dfrac{1}{\xi_N}}$	$\overline{\dfrac{1}{\xi_{other}}}$
		$\overline{\tau}=\overline{\tau_{base}}\left(\overline{\dfrac{1}{\xi_T}}\right)\left(\overline{\dfrac{1}{\xi_W}}\right)\left(\overline{\dfrac{1}{\xi_O}}\right)\left(\overline{\dfrac{1}{\xi_N}}\right)\left(\overline{\dfrac{1}{\xi_{other}}}\right)$					
Global	10.0 (1.11)	0.31 (1.00)	23.4 (1.29)	57.2 (1.86)	0.59 (1.01)	0.23 (1.00)	8.3 (1.09)
NET temperate	48.5 (1.24)	1.20 (1.01)	4.7 (1.02)	43.9 (1.22)	0.38 (1.00)	0.81 (1.00)	0.5 (1.00)
NET boreal	12.5 (1.15)	0.77 (1.01)	13.4 (1.16)	68.4 (2.13)	0.36 (1.00)	3.09 (0.97)	1.4 (0.98)
NDT boreal	25.3 (1.45)	1.47 (1.02)	9.5 (1.15)	50.5 (2.10)	0.79 (0.99)	1.60 (1.02)	10.9 (1.17)
BET Tropical	33.6 (1.12)	1.42 (1.00)	0.5 (1.00)	54.7 (1.20)	1.70 (0.99)	0.06 (1.00)	8.0 (0.97)
BET temperate	23.9 (1.07)	0.36 (1.00)	4.9 (1.01)	60.2 (1.17)	0.04 (1.00)	4.49 (1.01)	6.1 (0.98)
BDT tropical	15.6 (1.07)	0.09 (1.00)	4.7 (0.98)	64.3 (1.33)	0.21 (1.00)	0.04 (1.00)	15.1 (0.94)
BDT temperate	50.0 (1.17)	0.71 (1.00)	9.5 (1.03)	33.3 (1.11)	0.08 (1.00)	3.95 (1.01)	2.4 (1.01)
BDT boreal	14.0 (1.11)	1.19 (1.01)	14.8 (1.12)	66.2 (1.64)	0.84 (0.99)	0.13 (1.00)	2.8 (0.98)
BES temperate	1.6 (1.00)	0.75 (1.00)	21.0 (0.94)	66.4 (1.21)	0.03 (1.00)	0.88 (1.00)	9.3 (0.97)
BDS temperate	36.2 (1.21)	1.36 (1.01)	8.2 (0.96)	48.9 (1.29)	0.01 (1.00)	0.04 (1.00)	5.3 (1.03)
BDS boreal	15.1 (1.19)	2.00 (1.02)	8.9 (1.11)	57.9 (1.96)	0.31 (1.00)	0.19 (1.00)	15.6 (1.20)
C_3 Arctic grass	5.7 (1.07)	0.94 (1.01)	10.6 (1.13)	65.9 (2.19)	0.09 (1.00)	0.17 (1.00)	16.5 (1.22)
C_3 grass	23.2 (1.13)	1.76 (1.01)	13.8 (1.07)	54.3 (1.32)	1.12 (1.01)	2.12 (1.01)	3.6 (1.02)
C_4 grass	24.3 (1.10)	1.55 (1.01)	2.6 (0.99)	62.3 (1.28)	0.26 (1.00)	0.81 (1.00)	8.2 (0.97)

Note. The number in parentheses is the ratio of one traceable component between the two forcing sets (CRUNCEP-v7 over GSWP). NET = needleleaf evergreen tree, NDT = needleleaf deciduous tree, BET = broadleaf evergreen tree, BDT = broadleaf deciduous tree, BES = broadleaf evergreen shrub, BDS = broadleaf deciduous shrub.

residence time, temperature scalar, water scalar, oxygen scalar, nitrogen scalar, and other scalars contributed 10.0%, 0.3%, 23.4%, 57.2%, 0.6%, 0.2%, and 8.3%, respectively, to the differences in soil carbon storage between the two forcing sets (Table 3). Among all the traceable components, the water scalar consistently made the most significant contribution to the different soil carbon storage under the two sets of forcings for most of the plant functional types.

4. Discussion

4.1. Modularity in Coding

Without changing modeled processes, the matrix model reorganizes CLM5 biogeochemical modules according to elements in Equations 1 and 2 for vegetation carbon and nitrogen dynamics and Equations 12 and 13 for soil carbon and nitrogen dynamics. The original modules of CLM5 represent carbon and nitrogen cycles in six subroutines for updating state variables related to phenology, mortality, fire, and harvest, respectively. The reorganized CLM5 matrix model represent carbon and nitrogen cycles in two subroutines for vegetation and soil. Influences of phenology, mortality, and fire on carbon and nitrogen dynamics are expressed by the process rates and transfer matrices (i.e., phenology (K_{ph}, A_{ph}), gap mortality (K_{gm}, A_{gm}), fire (K_{fi}, A_{fi}), soil CN horizontal transfer (K_h, A_h), and vertical mixing (V). The gap mortality matrices (K_{gm}, A_{gm}) include influences of harvest. The organization of all the carbon and nitrogen processes in the matrix form makes multiple processes more traceable and enables better diagnostics analysis as described in section 4.2. Moreover, the modularity offered by the matrix model benefits teaching and learning about the model processes. One difficulty in learning carbon and nitrogen cycle models by students or postdoctoral scientists is the complicated carbon and nitrogen cycle networks. CLM5 uses hundreds of carbon and nitrogen balance equations to describe carbon and nitrogen cycles. Neither processes nor parameters are easily discerned from mass balance equations. A modularized representation in the matrix model summarizes all the mass balance equations into one matrix equation for either carbon or nitrogen cycle. In particular, the nonzero entries in those transfer coefficient matrices succinctly indicates connections within the networked pools.

4.2. Diagnostic Capability

One of the most useful advantages that the matrix approach offers is the diagnostic capability, such as traceability analysis and semianalytic attribution analysis. Xia et al. (2013) introduced the traceability framework and applied it to analyze sources of differences in modeled carbon cycle at steady state by CABLE. Modeled carbon storage in CABLE, CLM-CASA, and CLM4 were decomposed into the traceable components, net primary productivity (NPP) and residence time (Rafique et al., 2017). The comparison among the traceable components and corresponding observations provided fundamental explanations for the modeled carbon storage deviations at steady state. The traceability analysis was extended to understand transient dynamics of land carbon cycle in response to climate change according to the mathematic properties of the matrix equation (Luo et al., 2017). The transient traceability framework was applied to analyze differences in carbon dynamics between Harvard and Duke Forests (Jiang et al., 2017). The study found that the divergences in allocation coefficients and environmental scalars drove the residence time toward two opposite directions, although modeled carbon storage was only slightly different between Harvard and Duke Forests. At the global scale, uncertainties in modeled historical carbon storage among models in three Model Intercomparison Project (MIPs), which were CMIP5, MsTMIP, and TRENDY, were traced to NPP, residence time, and carbon storage potential (Zhou et al., 2018).

We implemented the traceability analysis in the CLM5 matrix model and illustrated its utility to understand sources of differences in model simulations with respect to large differences in soil carbon storage in CLM5 when driven by CRUNCEPv7 versus GSWP3, as reported in Lawrence et al. (2019). On the face of it, the large difference in soil carbon storage arising from two plausible historical climate reconstructions was puzzling. The global mean annual temperature and precipitation only differed by 0.4°C and 20 mm, respectively, between these data sets. Our traceability analysis enabled decomposition of the controls on soil carbon storage into seven components, including residence time and carbon inputs, with the residence time impact further decomposed into environmental scalars and baseline residence time.

Our traceability analysis indicates that the differences in modeled soil carbon storage between CRUNCEPv7 and GSWP3 were primarily due to the water scalars either based on global average or PFT average. Especially, the water scalar contributions from PFTs in permafrost regions are significantly higher (Table 3). The importance of water scalar in permafrost regions is consistent with experimental results that permafrost soil carbon dynamics were very sensitive to changes in soil thermal and hydrological conditions (Mauritz et al., 2017). CLM5 matrix model can further explain the high sensitivity mathematically. In matrix Equations 25 and 27, the water scalar in the carbon storage capacity is calculated as an inverse of the water scalar ($1/\xi_W$). Phase changes in soil water from frozen to active soil or the reverse can cause dramatic changes in soil water scalar. For example, water scalar is usually nearly 0 when soil water is at phase change. The inverse of a nearly 0 number can be extremely large and cause huge changes in simulated soil carbon even with a small change in precipitation or air temperature. Our results not only confirmed the conclusion on the large differences in permafrost regions by Lawrence et al. (2019) but also accurately identified the causes of the differences.

The diagnostic capability that the matrix model offers not only help identify sources in modeled carbon cycle dynamics caused by different sets of meteorological forcing but also help understand model structures. For example, Du et al. (2018) applied the traceability analysis to evaluate the large differences in carbon storage capacity among three existing schemes in carbon-nitrogen coupling. With assistance of the traceability analysis, the differences were mainly attributed to downregulation of photosynthesis, plant tissue C:N ratio, and plant nitrogen uptake.

Besides the traceability analysis, the matrix approach enables the semianalytic attribution to evaluate relative importance of different processes in modeled biogeochemical dynamics. For example, Huang, Lu, et al. (2018) used semianalytic attribution to evaluate CO_2 fertilization effects from different processes. The study evaluated relative importance of several processes, such as changes in carbon input, allocation, CO_2-induced nitrogen limitation, environmental scalars, and vertical mixing processes, for CO_2 effects on soil carbon storage. The semianalytic attribution identified carbon input as the most important contributor.

4.3. Matrix Representation of Terrestrial Carbon Cycle Models

Luo et al. (2017) have argued that almost all land carbon cycle models can be represented by the matrix form. So far, there are ~30 models that have been successfully represented in the matrix form, including some of

the commonly used models, such as CLM5, and nonlinear microbial models (Sierra & Muller, 2015). The matrix equation unifies carbon cycle models by varying its dimensions and details of expression of its terms related to carbon input, plant allocation, process rates, carbon transfer, and environmental modifiers. For examples, CABLE has only nine carbon pools including three vegetation pools, three litter pools, and three soil pools, whereas CLM5 has 18 vegetation carbon pools and 140 soil carbon pools. Thus, the matrix equation has nine dimensions for CABLE and 158 for CLM5 if one grid is occupied by one vegetation type and 194 if one grid is occupied by three vegetation types. Then, the number of the carbon transfers increased exponentially with the number of pools. CABLE includes three vegetation carbon transfers whereas CLM5 has a total of 56 carbon transfers in vegetations. Moreover, CLM5 has much more detailed process representation related to fire, vertical carbon transfers in soil, and nitrogen dynamics than CABLE. Nevertheless, all those detailed processes in CLM5 are all folded into the five terms of matrix equation as in Equations 1, 2, 12, and 13. Thus, CABLE and CLM5 share similar mathematical expression and properties.

The matrix equations can be semianalytically solved to estimate initial pool sizes at steady state. This is semianalytic spin-up (SASU). In comparison, a traditional method that runs a model to the steady state usually takes thousands of years with cycled meteorological forcing. An accelerated decomposition (i.e., AD) method requires hundreds of cycles (Koven et al., 2013; Thornton & Rosenbloom, 2005). Thus, SASU requires tends of cycles to reach steady states and, thus, can reduce the computational cost by 1 or 2 orders of magnitude for spinning up global land models (Xia et al., 2012). The elevated computational efficiency by SASU enables parameter sensitivity analysis (Huang, Zhu, et al., 2018) and pool-based data assimilation (Hararuk et al., 2015; Hararuk & Luo, 2014; Shi et al., 2018). Sensitivity of the carbon storage from 34 parameters have been evaluated (Huang, Zhu, et al., 2018). The active layer depth among 34 parameters has been identified to play important role in predicting high-latitude soil organic carbon. No parameter sensitivity analysis on modeling soil organic carbon can be conducted previously with complex biogeochemical models. The efficient matrix-based sensitivity analysis can help effectively evaluate such models to improve our understanding.

5. Conclusion

We fully implemented the matrix approach to CLM5 biogeochemistry modules by reorganizing hundreds of mass balance equations of carbon and nitrogen cycles into four matrix equations. Dynamics in vegetation carbon, vegetation nitrogen, soil carbon, and soil nitrogen are separately represented. The CLM5 matrix model fully reproduced the structure and simulations of the original biogeochemistry modules of CLM5. Moreover, the matrix model provided additional benefits to modeling, such as modularity, effective diagnostics, and high computational efficiency. The simplicity and modularity in code benefit teaching and learning. The matrix approach provides strong diagnostic capacity, such as traceability analysis, semianalytic attribution, and model structure comparison. Simulated carbon and nitrogen storage dynamics become more transparent and traceable than the original model.

The successful implementation of the matrix approach to CLM5 is a demonstration that most, if not all, the biogeochemical models can be represented in the matrix form. To date, about 30 models have been represented with the matrix form. Among them, CLM5 is among the most complicated models. Besides the benefit mentioned above, the standardized model structure will make biogeochemistry models more comparable. That is especially useful for model intercomparison projects.

Data Availability Statement

The code of the matrix model of CLM5, the model results, and the script for traceability analysis are available at this site (http://www2.nau.edu/luo-lab/download/Lu_2020_JAMES.php).

References

Ahlstrom, A., Xia, J. Y., Arneth, A., Luo, Y. Q., & Smith, B. (2015). Importance of vegetation dynamics for future terrestrial carbon cycling. *Environmental Research Letters*, *10*(5). https://doi.org/10.1088/1748-9326/10/5/054019

Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., et al. (2013). Carbon-concentration and carbon-climate feedbacks in CMIP5 earth system models. *Journal of Climate*, *26*(15), 5289–5314. <go to ISI>://WOS:000322327700001. https://doi.org/10.1175/JCLI-D-12-00494.1

Bonan, G. B. (1996). Land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and users guide (*Technical Note*). Boulder, CO: National Center for Atmospheric Research.

Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in earth system models. *Science*, *359*(6375), eaam8328. https://doi.org/10.1126/science.aam8328

Computational and Information Systems Laboratory (2019). *Cheyenne: HPE/SGI ICE XA System (Wyoming-NCAR Alliance)*. Boulder, CO: National Center for Atmospheric Research. https://doi.org/10.5065/D6RX99HX

Dai, Y., & Zeng, Q. (1997). A land surface model (IAP94) for climate studies part I: Formulation and validation in off-line experiments. *Advances in Atmospheric Sciences*, *14*(4), 433–460.

Dickinson, R., Henderson-Sellers, A., & Kennedy, P. (1993). Biosphere atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR community climate model. NCAR Technical Note NCAR/TN-387+STR. National Center for Atmospheric Research, Boulder, CO:

Dirmeyer, P. A., Gao, X. A., Zhao, M., Guo, Z. C., Oki, T. K., & Hanasaki, N. (2006). GSWP-2—Multimodel analysis and implications for our perception of the land surface. *Bulletin of the American Meteorological Society*, *87*(10), 1381–1398. https://doi.org/10.1175/BAMS-87-10-1381

Drewniak, B., Song, J., Prell, J., Kotamarthi, V. R., & Jacob, R. (2013). Modeling agriculture in the community land model. *Geoscientific Model Development*, *6*(2), 495–515. <Go to ISI>://WOS:000318438600015

Du, Z. G., Weng, E. S., Jiang, L. F., Luo, Y. Q., Xia, J. Y., & Zhou, X. H. (2018). Carbon-nitrogen coupling under three schemes of model representation: A traceability analysis. *Geoscientific Model Development*, *11*(11), 4399–4416. <Go to ISI>://WOS:000449154800001

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. <Go to ISI>://WOS:000376937800013

Fisher, J. B., Sitch, S., Malhi, Y., Fisher, R. A., Huntingford, C., & Tan, S. Y. (2010). Carbon cost of plant nitrogen acquisition: A mechanistic, globally applicable model of plant nitrogen uptake, retranslocation, and fixation. *Global Biogeochemical Cycles*, *24*, GB1014. https://doi.org/10.1029/2009GB003621

Friedlingstein, P. (2015). Carbon cycle feedbacks and future climate change. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, *373*(2054). <go to ISI>://WOS:000366270500005

Friedlingstein, P., Andrew, R. M., Rogelj, J., Peters, G. P., Canadell, J. G., Knutti, R., et al. (2014). Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience*, *7*(10), 709–715. https://doi.org/10.1038/ngeo2248

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate-carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. *Journal of Climate*, *19*(14), 3337–3353. https://doi.org/10.1175/JCLI3800.1

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., et al. (2019). Global carbon budget 2019. *Earth System Science Data*, *11*(4), 1783–1838. https://doi.org/10.5194/essd-11-1783-2019

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511–526. <go to ISI>://WOS:000329773100002

Hararuk, O., & Luo, Y. Q. (2014). Improvement of global litter turnover rate predictions using a Bayesian MCMC approach. *Ecosphere*, *5*(12). <go to ISI>://WOS:000347219000014

Hararuk, O., Smith, M. J., & Luo, Y. Q. (2015). Microbial models with data-driven parameters predict stronger soil carbon responses to climate change. *Global Change Biology*, *21*(6), 2439–2453. <go to ISI>://WOS:000353977500028

Houghton, R. A., House, J. I., Pongratz, J., van der Werf, G. R., DeFries, R. S., Hansen, M. C., et al. (2012). Carbon emissions from land use and land-cover change. *Biogeosciences*, *9*(12), 5125–5142. https://doi.org/10.5194/bg-9-5125-2012

Huang, Y. 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

Huang, Y. Y., Zhu, D., Ciais, P., Guenet, B., Huang, Y., Goll, D. S., et al. (2018). Matrix-based sensitivity assessment of soil organic carbon storage: A case study from the ORCHIDEE-MICT model. *Journal of Advances in Modeling Earth Systems*, *10*, 1790–1808. https://doi.org/10.1029/2017MS001237

Jiang, L. F., Shi, Z., Xia, J. Y., Liang, J. Y., Lu, X. J., Wang, Y., & Luo, Y. Q. (2017). Transient traceability analysis of land carbon storage dynamics: Procedures and its application to two Forest ecosystems. *Journal of Advances in Modeling Earth Systems*, *9*, 2822–2835. https://doi.org/10.1002/2017MS001004

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., et al. (2013). Twenty-first-century compatible CO_2 emissions and airborne fraction simulated by CMIP5 earth system models under four representative concentration pathways. *Journal of Climate*, *26*(13), 4398–4413. https://doi.org/10.1175/JCLI-D-12-00554.1

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

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., & Vertenstein, M. (2019). CLM5 documentation.

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*, 4245–4287. https://doi.org/10.1029/2018MS001583

Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., et al. (2013). The global carbon budget 1959–2011. *Earth System Science Data*, *5*(1), 165–185. <go to ISI>://WOS:000209415400013. https://doi.org/10.5194/essd-5-165-2013

Li, F., Levis, S., & Ward, D. S. (2013). Quantifying the role of fire in the earth system—Part 1: Improved global fire modeling in the community earth system model (CESM1). *Biogeosciences*, *10*(4), 2293–2314. <go to ISI>://WOS:000318434200008. https://doi.org/10.5194/bg-10-2293-2013

Li, F., Zeng, X. D., & Levis, S. (2012). A process-based fire parameterization of intermediate complexity in a dynamic global vegetation model. *Biogeosciences*, *9*(11), 4771–4772. <go to ISI>://WOS:000312667300041. https://doi.org/10.5194/bg-9-4771-2012

Luo, Y. Q., Shi, Z., Lu, X. J., Xia, J. Y., Liang, J. Y., Jiang, J., et al. (2017). Transient dynamics of terrestrial carbon storage: Mathematical foundation and its applications. *Biogeosciences*, *14*(1), 145–161. <go to ISI>://WOS:000393893500001. https://doi.org/10.5194/bg-14-145-2017

Luo, Y. Q., White, L. W., Canadell, J. G., DeLucia, E. H., Ellsworth, D. S., Finzi, A. C., et al. (2003). Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach. *Global Biogeochemical Cycles*, *17*(1), 1021. <go to ISI>://WOS:000182109500002. https://doi.org/10.1029/2002GB001923

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., et al. (2018). Global warming of 1.5°C. 1.

Mauritz, M., Bracho, R., Celis, G., Hutchings, J., Natali, S. M., Pegoraro, E., et al. (2017). Nonlinear CO_2 flux response to 7 years of experimentally induced permafrost thaw. *Global Change Biology*, *23*(9), 3646–3666. <go to ISI>://WOS:000406812100019. https://doi.org/10.1111/gcb.13661

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., et al. (2009). Greenhouse-gas emission targets for limiting global warming to 2 degrees C. *Nature*, *458*(7242), 1158–1162. <go to ISI>://WOS:000265754600045. https://doi.org/10.1038/nature08017

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*(5), 2057–2116. <Go to ISI>://WOS:000402419000002. https://doi.org/10.5194/gmd-10-2057-2017

Rafique, R., Xia, J. Y., Hararuk, O., Asrar, G. R., Leng, G. Y., Wang, Y. P., & Luo, Y. Q. (2016). Divergent predictions of carbon storage between two global land models: Attribution of the causes through traceability analysis. *Earth System Dynamics*, *7*(3) <go to ISI>://WOS:000382829500001, 649–658. https://doi.org/10.5194/esd-7-649-2016

Rafique, R., Xia, J. Y., Hararuk, O., Leng, G. Y., Asrar, G., & Luo, Y. Q. (2017). Comparing the performance of three land models in global C cycle simulations: A detailed structural analysis. *Land Degradation & Development*, *28*(2), 524–533. <go to ISI>://WOS:000394955500013. https://doi.org/10.1002/ldr.2506

Riley, W. J., Subin, Z. M., Lawrence, D. M., Swenson, S. C., Torn, M. S., Meng, L., et al. (2011). Barriers to predicting changes in global terrestrial methane fluxes: Analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences*, *8*(7), 1925–1953. <go to ISI>://WOS:000294153700013. https://doi.org/10.5194/bg-8-1925-2011

Running, S. W., & Hunt, E. R. (1993). Generalization of a forest ecosystem process model for other biomes, BIOME-BCG, and an application for global-scale models.

Sacks, W. J., Cook, B. I., Buenning, N., Levis, S., & Helkowski, J. H. (2009). Effects of global irrigation on the near-surface climate. *Climate Dynamics*, *33*(2–3), 159–175. <go to ISI>://WOS:000267103700001. https://doi.org/10.1007/s00382-008-0445-z

Schaphoff, S., Heyder, U., Ostberg, S., Gerten, D., Heinke, J., & Lucht, W. (2013). Contribution of permafrost soils to the global carbon budget. *Environmental Research Letters*, *8*(1) <go to ISI>://WOS:000316998300034, 014026. https://doi.org/10.1088/1748-9326/8/1/014026

Shi, Z., Crowell, S., Luo, Y. Q., & Moore, B. (2018). Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nature Communications*, *9*(1), 2171. <go to ISI>://WOS:000434015000009. https://doi.org/10.1038/s41467-018-04526-9

Shi, Z., Yang, Y. H., Zhou, X. H., Weng, E. S., Finzi, A. C., & Luo, Y. Q. (2016). Inverse analysis of coupled carbon-nitrogen cycles against multiple datasets at ambient and elevated CO_2. *Journal of Plant Ecology*, *9*(3), 285–295. https://doi.org/10.1093/jpe/rtv059

Sierra, C. A., & Muller, M. (2015). A general mathematical framework for representing soil organic matter dynamics. *Ecological Monographs*, *85*(4), 505–524. <go to ISI>://WOS:000364962200002. https://doi.org/10.1890/15-0361.1

Song, C. H., & Woodcock, C. E. (2003). A regional forest ecosystem carbon budget model: Impacts of forest age structure and landuse history. *Ecological Modelling*, *164*(1), 33–47. <go to ISI>://WOS:000182960700003. https://doi.org/10.1016/S0304-3800(03)00013-9

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., et al. (2013). Climate change 2013: The physical science basis.

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*, 17. https://doi.org/10.3389/fdata.2020.00017

Thornton, P. E., Law, B. E., Gholz, H. L., Clark, K. L., Falge, E., Ellsworth, D. S., et al. (2002). Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. *Agricultural and Forest Meteorology*, *113*(1–4), 185–222. <go to ISI>://WOS:000179188300011. https://doi.org/10.1016/S0168-1923(02)00108-9

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. <go to ISI>://WOS:000233374300002. https://doi.org/10.1016/j.ecolmodel.2005.04.008

Viovy, N. (2016). CRUNCEP data set.

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2014). Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *Journal of Geophysical Research: Biogeosciences*, *119*, 794–807. <Go to ISI>://WOS:000337607900006. https://doi.org/10.1002/2013JG002591

Xia, J. Y., Luo, Y. Q., Wang, Y. P., & Hararuk, O. (2013). Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Global Change Biology*, *19*(7), 2104–2116. <go to ISI>://WOS:000319963500011

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. <Go to ISI>://WOS:000310471600012. https://doi.org/10.5194/gmd-5-1259-2012

Xu, C. G., Fisher, R., Wullschleger, S. D., Wilson, C. J., Cai, M., & McDowell, N. G. (2012). Toward a mechanistic modeling of nitrogen limitation on vegetation dynamics. *PLoS ONE*, *7*(5), e37914. <go to ISI>://WOS:000305335800084. https://doi.org/10.1371/journal.pone.0037914

Xu, T., White, L., Hui, D., & Luo, Y. (2006). Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochemical Cycles*, *20*. <go to ISI>://WOS:000237446700001, GB2007. https://doi.org/10.1029/2005GB002468

Zhou, S., Liang, J., Lu, X., Li, Q., Jiang, L., Zhang, Y., et al. (2018). Sources of uncertainty in modeled land carbon storage within and across three MIPs: Diagnosis with three new techniques. *Journal of Climate*, *31*(7), 2833–2851. <go to ISI>://WOS:000429456500016. https://doi.org/10.1175/JCLI-D-17-0357.1