JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2019MS001776

Special Section:
The Energy Exascale Earth System Model (E3SM)

Mathematical Reconstruction of Land Carbon Models From Their Numerical Output: Computing Soil Radiocarbon From ^{12}C Dynamics

Holger Metzler[1], Qing Zhu[2], William Riley[2], Alison Hoyt[1,2], Markus Müller[1], and Carlos A. Sierra[1]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany, [2]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Key Points:
- Numerical output of land carbon models can be used to reconstruct a compartmental dynamical system
- The reconstructed model can be used to compute isotope dynamics such as radiocarbon
- We demonstrate the approach using the land component of the E3SM model

Correspondence to:
C. A. Sierra,
csierra@bgc-jena.mpg.de

Citation:
Metzler, H., Zhu, Q., Riley, W. J., Hoyt, A. M., Müller, M., & Sierra, C. A. (2020). Mathematical reconstruction of land carbon models from their numerical output: Computing soil radiocarbon from ^{12}C dynamics. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001776. https://doi.org/10.1029/2019MS001776

Received 7 JUN 2019
Accepted 7 DEC 2019
Accepted article online 20 DEC 2019

©2019. The Authors.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Radiocarbon (^{14}C) is a powerful tracer of the global carbon cycle that is commonly used to assess carbon cycling rates in various Earth system reservoirs and as a benchmark to assess model performance. Therefore, it has been recommended that Earth System Models (ESMs) participating in the Coupled Model Intercomparison Project Phase 6 report predicted radiocarbon values for relevant carbon pools. However, a detailed representation of radiocarbon dynamics may be an impractical burden on model developers. Here, we present an alternative approach to compute radiocarbon values from the numerical output of an ESM that does not explicitly represent these dynamics. The approach requires computed ^{12}C stocks and fluxes among all carbon pools for a particular simulation of the model. From this output, a time-dependent linear compartmental system is computed with its respective state-transition matrix. Using transient atmospheric ^{14}C values as inputs, the state-transition matrix is then applied to compute radiocarbon values for each pool, the average value for the entire system, and component fluxes. We demonstrate the approach with ELMv1-ECA, the land component of an ESM model that explicitly represents ^{12}C, and ^{14}C in 7 soil pools and 10 vertical layers. Results from our proposed method are highly accurate (relative error <0.01%) compared with the ELMv1-ECA ^{12}C and ^{14}C predictions, demonstrating the potential to use this approach in CMIP6 and other model simulations that do not explicitly represent ^{14}C.

Plain Language Summary Models representing ecosystem carbon dynamics are generally complex and difficult to analyze. Comparing different models with different structures is even more challenging due to the variety of processes represented in the models. However, it is possible to use the numerical output of models to reconstruct the original structure using a common mathematical framework. In this manuscript, we demonstrate this approach and apply it to compute radiocarbon dynamics of a land carbon model. The proposed approach can reconstruct the carbon and radiocarbon dynamics of the original model very accurately and can be used to study system-level dynamics of complex contrasting models.

1. Introduction

Terrestrial ecosystems, and soils in particular, are important global carbon (C) reservoirs but with large uncertainties in the size of the stock, process rates, and their representation in Earth System Models (ESMs) (Luo et al., 2016; Sulman et al., 2018). The dynamics of radiocarbon in various components of terrestrial ecosystems can be used to interpret the rates at which CO_2 is removed from the atmosphere into plants through photosynthesis, moves from vegetation into soils through litter fall and mortality, and finally cycles back to the atmosphere via respiration. As a result, radiocarbon has been used in many empirical studies to evaluate rates of soil C cycling and, more recently, as a tool for model evaluation (Ahrens et al., 2015; Chen et al., 2019; Dwivedi et al., 2017) and intercomparison (He et al., 2016).

Measurements of radiocarbon offer great potential to constrain temporal dynamics of the land carbon cycle in ESMs as well as to assess their performance. However, to date, such comparisons have been limited to a small number of models that simulate C isotopes (e.g., Chen et al., 2019; Koven et al., 2013; Tifafi et al., 2018), as the majority of models do not explicitly represent isotope dynamics. The current coupled climate-carbon cycle model intercomparison project (C4MIP) of the Coupled Model Inter-comparison Project Phase 6

(CMIP6) requests model participants to explicitly model radiocarbon and include the corresponding output from the major land and ocean model compartments (Graven et al., 2017; Jones et al., 2016). However, this additional request can be time-consuming to implement and computationally intensive to run. Thus, the goal of this work is to develop a tool to make radiocarbon-based comparisons of C cycling practical and comparable across different ESMs even if they do not explicitly include these dynamics.

The availability of radiocarbon output from a range of models would allow an additional comparison and evaluation against global data sets, such as the International Soil Radiocarbon Database (ISRaD; Lawrence et al., 2019), and would provide a new way to differentiate patterns of C cycling across different model structures (Sierra et al., 2014). There are literally hundreds of models of soil carbon cycling described in the literature (Manzoni & Porporato, 2009), each relying on different hypotheses about internal system dynamics, with specific model structures and parameterizations. Generalizations of models have also been proposed based on basic assumptions of soil organic matter dynamics (Sierra & Müller, 2015). Computing radiocarbon output from these different classes of models, without the need to incorporate explicit isotopes dynamics in the model, would benefit data-model comparisons and improve the understanding of soil C dynamics in these models.

Linear response theory suggests that the impulse response function of a particular model can be used to compute isotope dynamics without explicitly simulating the isotope within the model but rather using the response function to convolve the isotope tracers (Thompson & Randerson, 1999). Therefore, it may not be necessary to explicitly represent isotope dynamics a priori in a model; rather, they can be obtained a posteriori through the convolution procedure. However, linear response theory relies on assumptions of steady-state and model linearity, which are not met for the type of transient simulations performed with ESMs.

Metzler et al. (2018) recently developed a method to obtain a mathematical object, the state-transition matrix (see details below), which generalizes response function theory for nonlinear models out of equilibrium. This approach removes previous assumptions that prevented the use of the linear response approach for the computation of isotope and tracer dynamics a posteriori from simulations. Therefore, the state-transition matrix approach offers an opportunity to compute isotope dynamics from models that do not explicitly represent them and could be applied to models participating in CMIP6, for example. However, since output from these models is provided in discrete time steps often much larger than the internal model time step, it is a challenge to translate the state-transition-matrix approach from the continuous to the discrete case.

We introduce in this manuscript an approach to mathematically reconstruct land carbon models from simulation output and use the reconstruction to compute radiocarbon values. To this end, we extend the method presented in Metzler et al. (2018) for the reconstruction of a differential equation model from an existing solution and the computation of the state-transition matrix in continuous time. We show how to translate the approach to the discrete-time case and apply it to reconstruct tracer and isotope dynamics from model output. We first give a mathematical overview of the approach and then present examples of its use. We test our approach by comparing our computed values to output from the ELMv1-ECA model (Chen et al., 2019; Riley et al., 2018; Tang & Riley, 2018; Zhu et al., 2019), the land component of the Exascale Earth System Model (E3SM) model, which explicitly represents radiocarbon and is therefore an appropriate test case to evaluate our proposed method. We evaluate the modeled radiocarbon values along the entire vertically resolved soil profile and across all represented soil and litter pools over time. We believe this new framework has potential for application to the CMIP6 simulations and to models with more diverse structures.

2. Methods

2.1. Carbon Cycle Models as Compartmental Systems

Carbon cycle models are subject to the law of mass conservation and therefore can be represented as compartmental systems (Anderson, 1983; Jacquez & Simon, 1993). A compartmental system with $d \geq 1$ compartments is of the general form of the initial value problem

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{C}(t) &= \mathrm{B}(\mathbf{C}(t), t)\ \mathbf{C}(t) + \mathbf{u}(\mathbf{C}(t), t), \quad t > t_0, \\ \mathbf{C}(t_0) &= \mathbf{C}^0. \end{aligned} \tag{1}$$

Here, $\mathbf{C}(t)$ is the vector of compartment contents at time t, B(C, t) is a matrix-valued function that governs the cycling of carbon through the system, $\mathbf{u}(\mathbf{C}, t)$ is the vector of external carbon inputs to the system, t_0

is some fixed initial time, and $\mathbf{C}^0$ is a given vector of initial system contents. Note that the internal cycling as well as the external inputs can each depend on time (e.g., due to changing environmental conditions) and on the current carbon content of the system itself. Consequently, system (1) is nonautonomous (time dependent) and nonlinear.

To guarantee mass conservation, the matrix $\mathrm{B}(\mathbf{C}, t)$ is required to be compartmental for all nonnegative $\mathbf{C}$ and $t > t_0$; that is, $\mathrm{B}(\mathbf{C}, t) = (\mathrm{B}_{ij}(\mathbf{C}, t))_{i,j=1,2,\dots,d}$ satisfies the properties

(i) $B_{ii}(\mathbf{C}, t) \le 0$ $B_{ii}(\boldsymbol{C}, t) \le 0$ for all i,
(ii) $B_{ij}(\mathbf{C}, t) \ge 0$ for all $i \ne j$,
(iii) $\sum_{i=1}^{d} B_{ij}(\mathbf{C}, t) \le 0$ for all j.

The diagonal values of B_{ii} can be interpreted as the cycling rate of compartment i, the off-diagonal entries B_{ij} represent the rate of carbon flux from compartment j to compartment i, and $z_j = -\sum_{i=1}^{d} B_{ij}$ is the rate of carbon leaving the system through pool j.

We now assume that we know a unique (numerical) solution $\mathbf{C} = \mathbf{C}(t)$ of system (1). Plugging this solution into the system, we obtain a new compartmental system:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\,\widetilde{\mathbf{C}}(t) &= \widetilde{\mathrm{B}}(t)\,\widetilde{\mathbf{C}}(t) + \widetilde{\mathbf{u}}(t), \quad t > t_0, \\ \widetilde{\mathbf{C}}(t_0) &= \mathbf{C}^0, \end{aligned} \tag{2}$$

where $\widetilde{\mathrm{B}}(t) = \mathrm{B}(\mathbf{C}(t), t)$ and $\widetilde{\mathbf{u}}(t) = \mathbf{u}(\mathbf{C}(t), t)$. This system is still time dependent, but linear. Under the assumption that it has a unique solution $\widetilde{\mathbf{C}} = \widetilde{\mathbf{C}}(t)$, the two solution trajectories $\widetilde{\mathbf{C}}$ and $\mathbf{C}$ are identical. Then the linear compartmental system (2) describes exactly the same solution trajectory as the nonlinear compartmental system (1). For the sake of simplicity of notation, we omit the tilde in equation (2) in the remainder of the manuscript.

The unique semianalytical solution of the linear compartmental system (2) is given by (Brockett, 2015)

$$\mathbf{C}(t) = \Phi(t, t_0)\,\mathbf{C}^0 + \int_{t_0}^{t} \Phi(t, \tau)\,\mathbf{u}(\tau)\,\mathrm{d}\tau.$$

We call the solution semianalytic because the system's state transition matrix Φ is only implicitly given as the unique solution of the matrix initial value problem:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\,\Phi(t, t_0) &= \mathrm{B}(t)\,\Phi(t, t_0), \quad t > t_0, \\ \Phi(t_0, t_0) &= \mathrm{Id}, \end{aligned}$$

where Id denotes the identity matrix in $\mathbb{R}^d$. The state transition matrix contains information about the entire dynamics of the original system along the solution trajectory. It can be used to reconstruct the original nonlinear model with the specific forcings specified for the particular simulation.

2.2. Representation of Depth-Resolved Systems

This framework can also incorporate spatially explicit systems that move carbon along vertical or horizontal gradients by, for example, advection and diffusion. For instance, consider a depth-resolved compartmental model with l depth layers and c compartment types per layer. Let $d := c \cdot l$. Values of d from $1, 2, \dots, l$ belong to Layer 1, Layer 2, …, and Layer l of compartment $c = 1$. The numbers $l + 1, l + 2, \dots, 2 \cdot l$ belong to Layer 1 through Layer l of compartment $c = 2$ and so forth. In general, the numbers $(c - 1) \cdot l + 1, (c - 1) \cdot l + 2, \dots, c \cdot l$ belong to Layer 1 through Layer l of compartment type c. We introduce $\mathrm{B} := (B_{ij})_{i,j=1,2,\dots,d}$ and $\mathbf{u} := (u_i)_{i=1,2,\dots,d}$. Then, the depth-resolved system fits in the framework of equation (1), and the nonlinear structure of the system accommodates interactions among system variables that are commonly used to represent movement along spatial gradients. For a system with 10 depth layers and 7 compartment types per layer, B is a 70×70 matrix, and $\mathbf{u}$ is a vector with 70 components. For example, $B_{12,11}$ is the flux downward from layer $l = 1$ to layer $l = 2$ of compartment $c = 1$, and $B_{35,55}$ is the exchange rate from compartment $c = 5$ to compartment $c = 3$ within depth layer $l = 5$ (recall that the first index represents the

target and the second index the source). Obviously, fluxes among different compartment types and vertical layers can be represented, for example, by $B_{23,52}$.

The negative diagonal entries $(B_{ii})_{i=1,2,\dots,70}$ are the total exit rates from each compartment, that is, the rates to other compartments in any layer plus the rate of material advecting or diffusing from the compartment.

2.3. The Radiocarbon Compartmental System

Once we know the representation (1) of a carbon cycle model, we can readily manipulate the equation to obtain a compartmental system for radiocarbon. We only need to substitute B by ^{14}B defined as

$$^{14}\mathrm{B}(\mathbf{C}(t), t) := \mathrm{B}(\mathbf{C}(t), t) - \lambda\ \mathrm{Id},$$

where λ is the radioactive decay constant of ^{14}C scaled to the relevant time unit. For example, if the system is observed in time units of years, then $\lambda = \ln(2)/5{,}568$ years. This manipulation leads to a state-transition matrix of the radiocarbon system given by

$$^{14}\Phi(t, t_0) = \Phi(t, t_0)\ e^{-\lambda\ (t-t_0)}, \quad t \geq t_0,$$

from which we can directly calculate the effect of radioactive decay.

2.4. Reconstruction of a Continuous-Time System From Discrete-Time Model Output

The approach presented above works only when the original model is completely described by a system of ODEs, which by definition implies that the system is continuous in time. However, we are interested in using the approach for ESMs for which we only have access to model output reported in discrete time steps and for which the underlying process descriptions may be a combination of ODEs, PDEs, noncontinuous representations, and so forth. Therefore, it is necessary to produce a Continuous-Time Approximation (CTA) of the model from its discrete model output. First, we will illustrate the approach using a one compartment system as an example and then for multiple interconnected compartments.

2.4.1. One compartment Case

Suppose we are given a one compartment time-dependent linear compartmental system

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\ C(t) &= \gamma(t)\ C(t) + u(t), \quad t > t_0, \\ C(t_0) &= C^0, \end{aligned} \tag{3}$$

where $\gamma(t)$ and $u(t)$ are unknown. Further, for discrete times $t_0 < t_1 < t_2 < \dots < t_n$, we are given

- the initial system state C^0,
- $u^k = \int_{t_k}^{t_{k+1}} u(t)\, dt, \quad k = 0, 1, \dots, n-1$, and
- $r^k = \int_{t_k}^{t_{k+1}} r(t)\, dt, \quad k = 0, 1, \dots, n-1.$

Here, u^k and r^k denote the accumulated external system input and the accumulated external system output in the time interval $I_k = [t_k, t_{k+1}]$, $k = 0, 1, \dots, n-1$, respectively.

We are interested in constructing a compartmental system, as simple as possible, that matches these given data as well as possible. To that end, we approximate the (unknown) system (3) on each interval I_k by a linear time-independent system

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\ \hat{C}(t) &= \hat{\gamma}^k\ \hat{C}(t) + \hat{\mathrm{u}}^k, \quad t \in I_k, \\ \hat{C}(t_k) &= \hat{C}^k. \end{aligned} \tag{4}$$

Suppose that at time step k the value $\hat{C}^k$ is already known from the previous time step, and for $k = 0$ we set $\hat{C}^0 = C^0$.

We write $\Delta_k = t_{k+1} - t_k$ and choose $\hat{\mathrm{u}}^k = u^k/\Delta_k$. Consequently, for the time interval I_k the cumulative external system input u^k of the unknown system (3) and the cumulative external system input $\Delta_k\ \hat{\mathrm{u}}^k$ of the approximating system (4) coincide. We are left with finding good approximations $\hat{\gamma}^k$ for $\gamma(t)$ on

$I_k, k = 0, 1, \ldots, n-1$. We can compute the cumulative external system output of the approximating system (4) on I_k, in dependence on the choice of $\hat{\gamma}^k$, by

$$\hat{r}^k(\hat{\gamma}^k) = -\hat{\gamma}^k \int_{t_k}^{t_{k+1}} \hat{C}(t)\, dt,$$

where, for $t \in I_k$,

$$\begin{aligned}\hat{C}(t) &= e^{-\hat{\gamma}^k\,(t-t_k)}\,\hat{C}^k + \int_{t_k}^{t} e^{-\hat{\gamma}^k\,(t-\tau)}\,\hat{u}^k\, d\tau \\ &= e^{-\hat{\gamma}^k\,(t-t_k)}\,\hat{C}^k + (\hat{\gamma}^k)^{-1}\,\left(e^{-\hat{\gamma}^k\,(t-t_k)} - 1\right)\,\hat{u}^k\end{aligned}$$

is the unique solution of (4) on I_k. Consequently, we obtain $\hat{\gamma}^k$ by solving the one-dimensional nonlinear optimization problem

$$\underset{\hat{\gamma}^k<0}{\text{minimize}}\ \left|\hat{r}^k(\hat{\gamma}^k) - r^k\right|.$$

We define $\hat{C}^{k+1} = \hat{C}(t_{k+1})$ and continue with the next time step $k+1$.

In the special case of a time-independent unknown system (3), that is, $\gamma(t) = \gamma < 0$ and $u(t) = u \geq 0$, the approximation by system (4) is perfect, independent of the number n of given data points.

2.4.2. Multiple Compartment Case

If the compartmental system consists of more than one compartment, internal cycling among the different compartments may occur. In particular, all modern soil carbon models are multicompartmental (Manzoni & Porporato, 2009; Sierra & Müller, 2015), and they often represent vertical structure with explicit representation of diffusive or advective fluxes that connect adjacent soil layers in the discretized model. An approximating system is then required to also match the internal fluxes. Suppose we are given a d-compartment time-dependent linear system

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{C}(t) &= \mathrm{B}(t)\,\mathbf{C}(t) + \mathbf{u}(t), \quad t > t_0, \\ \mathbf{C}(t_0) &= \mathbf{C}^0,\end{aligned} \tag{5}$$

of which $\mathrm{B}(t) = (B_{ij}(t))_{i,j=1,2,\ldots,d}$ and $\mathbf{u}(t) = (u_i(t))_{i=1,2,\ldots,d}$ are unknown. But for discrete times $t_0 < t_1 < t_2 < \ldots < t_n$, we are given

- the initial system content vector $\mathbf{C}^0 = (C_1^0, C_2^0, \ldots, C_d^0)^T$,
- $\mathbf{u}^k = \int_{t_k}^{t_{k+1}} \mathbf{u}(t)\, dt, \quad k = 0, 1, \ldots, n-1,$
- $\mathbf{r}^k = \int_{t_k}^{t_{k+1}} \mathbf{r}(t)\, dt, \quad k = 0, 1, \ldots, n-1,$ and
- $F_{ij}^k = \int_{t_k}^{t_{k+1}} B_{ij}(t)\, C_j(t)\, dt, \quad i \neq j.$

Here, $\mathbf{u}^k$ and $\mathbf{r}^k$ denote the accumulated external system input vector and the accumulated external system output vector in the time interval $I_k = [t_k, t_{k+1}]$, $k = 0, 1, \ldots, n-1$, respectively. Furthermore, F_{ij}^k denotes the accumulated flux from compartment j to compartment i during I_k.

We are interested in constructing a compartmental system, as simple as possible, that matches these given data as well as possible. To that end, we approximate the (unknown) system (5) on each interval I_k by a linear time-independent system

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\hat{\mathbf{C}}(t) &= \hat{\mathrm{B}}^k\,\hat{\mathbf{C}}(t) + \hat{\mathbf{u}}^k, \quad t \in I_k, \\ \hat{\mathbf{C}}(t_k) &= \hat{\mathbf{C}}^k.\end{aligned} \tag{6}$$

Suppose that at time step k the vector $\hat{\mathbf{C}}^k$ is already known from the previous time step; for $k = 0$ we set $\hat{\mathbf{C}}^0 = \mathbf{C}^0$.

Again, we write $\Delta_k = t_{k+1} - t_k$ and choose $\hat{\mathbf{u}}^k = \mathbf{u}^k/\Delta_k$. Consequently, for the time interval I_k the cumulative external system input vector $\mathbf{u}^k$ of the unknown system (5) and the cumulative external system input vector $\Delta_k\ \hat{\mathbf{u}}^k$ of the approximating system (6) coincide. We are left with finding good approximations $\widehat{\mathrm{B}}^k$ for $\mathrm{B}(t)$ on I_k, $k = 0, 1, \ldots, n-1$ under the constraint that $\widehat{\mathrm{B}}^k$ is a compartmental matrix. To that end, we try to match the internal fluxes F_{ij}^k and the external output flux vectors $\mathbf{r}^k$ as well as possible. For $i \neq j$, we can compute the internal flux from j to i of the approximating system (6), in dependence on the choice of $\widehat{\mathrm{B}}^k$, during the time interval I_k by

$$\hat{F}_{ij}^k(\hat{\mathrm{B}}^k) = \hat{B}_{ij}^k \int_{t_k}^{t_{k+1}} \hat{C}_j(t)\ dt,$$

where, for $t \in I_k$,

$$\begin{aligned}\hat{\mathbf{C}}(t) &= e^{(t-t_k)\ \hat{\mathrm{B}}^k}\ \hat{\mathbf{C}}^k + \int_{t_k}^{t} e^{(t-\tau)\ \hat{\mathrm{B}}^k}\ \hat{\mathbf{u}}^k\ d\tau \\ &= e^{(t-t_k)\ \hat{\mathrm{B}}^k}\ \hat{\mathbf{C}}^k + (\hat{\mathrm{B}}^k)^{-1}\ (e^{(t-t_k)\ \hat{\mathrm{B}}^k}) - \mathrm{Id})\ \hat{\mathbf{u}}^k\end{aligned}$$

is the unique solution of (6) on I_k and $e^{(t-t_k)\ \hat{\mathrm{B}}^k}$ denotes the matrix exponential.

Furthermore, we define by $\hat{z}_j^k = -\sum_{i=1}^d \hat{B}_{ij}^k \geq 0$ the rate of external outflow from compartment j. Then we can compute the cumulative external system output through compartment j of the approximating system (6) during I_k, in dependence on the choice of $\widehat{\mathrm{B}}^k$, by

$$\hat{r}_j^k(\hat{\mathrm{B}}^k) = \hat{z}_j^k \int_{t_k}^{t_{k+1}} \hat{C}_j^k(t)\ dt.$$

Consequently, we obtain $\widehat{\mathrm{B}}^k$ by solving the m-dimensional nonlinear optimization problem

$$\underset{\hat{\mathrm{B}}^k \text{compartmental}}{\text{minimize}} \left[\max_{i \neq j} |\hat{F}_{ij}^k(\hat{\mathrm{B}}^k) - F_{ij}^k| + \left\|\hat{\mathbf{r}}^k(\hat{\mathrm{B}}^k) - \mathbf{r}^k\right\|\right],$$

where the maximum difference between predicted and provided fluxes is minimal as well as the difference between predicted and provided release fluxes. The dimension m of the optimization problem is at maximum d^2. If some internal fluxes or external output fluxes are inexistent, then $m < d^2$. We define $\hat{\mathbf{C}}^{k+1} = \hat{\mathbf{C}}(t_{k+1})$ and continue with the next time step $k+1$.

2.5. Reconstruction of a Discrete-Time Compartmental System

Model output is usually provided at discrete time points $t_0 < t_1 < \ldots < t_n$, either because there is no feasible method to represent continuous-time model output or the underlying process model works on the basis of discrete time steps. In either case, if we reconstruct the unknown original model using a CTA (6), we interpolate the given data for points in time that lie between t_k and t_{k+1}. This interpolation can be done in many ways, some of which could lead to large errors.

To ameliorate this problem, we can restrict the model reconstruction only to the data provided by the model at those particular time steps. Consequently, we reconstruct the unknown model in a discrete time setting, on the time grid $t_0 < t_1 < \ldots < t_n$ that is predetermined by the given data and make no interpolations for other time points. This approach leads to a reconstructed model that is represented by a d-compartment discrete-time system of the form

$$\begin{aligned}\hat{\mathbf{C}}(k+1) &= \hat{\mathrm{B}}(k)\ \hat{\mathbf{C}}(k) + \hat{\mathbf{u}}(k), \quad k = 0, 1, \ldots, n-1, \\ \hat{\mathbf{C}}(0) &= \hat{\mathbf{C}}^0.\end{aligned} \tag{7}$$

The discrete-time compartmental matrices $\hat{\mathrm{B}}(k)$ satisfy for all $k = 0, 1, \ldots, n-1$ two conditions:

(i) $\hat{B}_{ij}(k) \geq 0$ for all $i, j = 1, 2, \ldots, d$ and

(ii) $\sum_{i=1}^{d} \hat{B}_{ij}(k) \leq 1$ for all $j = 1, 2, \ldots, d$.

Analogously to the CTA, $\hat{z}_j(k) = 1 - \sum_{i=1}^{d} \hat{B}_{ij}(k)$ is the rate at which mass at time step k leaves the system through compartment j. The discrete-time state-transition matrix is, for $0 \le k_1 < k_2 \le n$, given by

$$\hat{\Phi}(k_2, k_1) = \hat{\mathrm{B}}_{k_2-1} \cdot \hat{\mathrm{B}}_{k_2-2} \dots \hat{\mathrm{B}}_{k_1+1} \cdot \hat{\mathrm{B}}_{k_1} = \prod_{k=k_1}^{k_2-1} \hat{\mathrm{B}}_k,$$

and $\hat{\Phi}(k, k) = \mathrm{Id}$. The solution of the discrete-time initial value problem (7) is then given by

$$\hat{\mathbf{C}}(k) = \hat{\Phi}(k, 0)\ \hat{\mathbf{C}}^0 + \sum_{m=0}^{k-1} \hat{\Phi}(k, m)\ \hat{\mathbf{u}}(m), \quad k = 0, 1, \dots, n.$$

Suppose we are given a d-compartment time-dependent linear system with structure (5) and unknown B and $\mathbf{u}$. For discrete times $t_0 < t_1 < t_2 < \dots < t_n$ we are given

- the initial system content vector $\mathbf{C}^0 = (C_1^0, C_2^0, \dots, C_d^0)^T$,
- $\mathbf{u}^k = \int_{t_k}^{t_{k+1}} \mathbf{u}(t)\ dt, \quad k = 0, 1, \dots, n-1,$
- $\mathbf{r}^k = \int_{t_k}^{t_{k+1}} \mathbf{r}(t)\ dt, \quad k = 0, 1, \dots, n-1,$ and
- $F_{ij}^k = \int_{t_k}^{t_{k+1}} B_{ij}(t)\ C_j(t)\ dt, \quad i \ne j.$

Here, $\mathbf{u}^k$ and $\mathbf{r}^k$ denote the accumulated external system input vector and the accumulated external system output vector in the time interval $I_k = [t_k, t_{k+1}]$, $k = 0, 1, \dots, n-1$, respectively. Furthermore, F_{ij}^k denotes the accumulated flux from compartment j to compartment i during I_k.

We are interested in constructing a discrete-time compartmental system (7) that matches these given data as well as possible. To that end, we define $\hat{\mathbf{u}}(k) = \mathbf{u}^k$, $k = 0, 1, \dots, n-1$. Furthermore, for each time step $k = 0, 1, \dots, n-1$, we have to find $\hat{\mathrm{B}}(k)$ such that $\hat{B}_{ij}(k)\ \hat{C}_j(k) = F_{ij}^k$ and $\hat{z}_j(k)\ \hat{C}_j(k) = r_j^k$. Consequently, for $i \ne j$,

$$\hat{B}_{ij}(k) = \begin{cases} F_{ij}^k / \hat{C}_j(k), & \text{if } \hat{C}_j(k) \ne 0, \\ 0, & \text{if } \hat{C}_j(k) = 0, \end{cases}$$

and

$$\hat{B}_{jj}(k) = \begin{cases} 1 - \left[\sum_{i \ne j} \hat{B}_{ij}(k) + r_j^k / \hat{C}_j(k) \right], & \text{if } \hat{C}_j(k) \ne 0, \\ 1, & \text{if } \hat{C}_j(k) = 0. \end{cases}$$

We define $\hat{\mathbf{C}}(k+1) = \hat{\mathrm{B}}(k)\ \hat{\mathbf{C}}(k) + \hat{\mathbf{u}}(k)$ and continue with the step $k+1$.

For cases where the full model output temporal resolution is too coarse, which can lead to depletion of a pool and negative pool sizes during DTA reconstruction, the Continuous-Time Approximation (CTA) is required.

2.6. Application of the Framework to Output From the ELMv1-ECA Model

As the land component of E3SM, ELMv1-ECA (Riley et al., 2018; Zhu et al., 2019) simulates vertically resolved soil C, N, and P dynamics based on the Century soil model (Parton et al., 1993) with enhancements to include soil O_2 effects (Koven et al., 2013). Radiocarbon (^{14}C) is treated as a separate tracer in addition to ^{12}C. ^{14}C enters into a terrestrial ecosystem from the atmosphere (Graven et al., 2017), and it is stored in vegetation compartments through biomass growth. After vegetation turnover (e.g., natural phenology, mortality, and disturbances), ^{14}C flows into litter or coarse woody debris along with ^{12}C and is consequently decomposed or accumulated in soil organic matter. Besides microbial decomposition, ^{14}C also radioactively decays. ELMv1-ECA also added several new features that could affect C inputs to soil and SOC dynamics that potentially have great impacts on soil radiocarbon signals. Briefly, the changes relevant to the current study include (a) phosphorus limitation on soil organic carbon decomposition (Zhu et al., 2016); (b) application of the Equilibrium Chemistry Approximation of multiple-consumer-multiple-resource competition network to resolve microbial immobilization of soil available nutrients (Tang & Riley, 2013;

Zhu et al., 2017); and (c) dynamic plant stoichiometry that affects litter nutrient content and decomposability. ELMv1-ECA has been evaluated against multiple global-scale observations of ecosystem C, water, and energy stocks and fluxes using ILAMB (Collier et al., 2018; Zhu et al., 2019) and short-term nutrient cycling observations and global-scale partitioning of N losses (Riley et al., 2018; Zhu & Riley, 2015). Overall, model benchmarking shows improvement from the precursor CLM4.5 predictions, and in particular for this study, more accurate estimates of spatially distributed soil C stocks (Zhu et al., 2019).

In this study, we present simulation results from ESMv1-ECA at three sites with contrasting background climate dynamics: boreal (62.5° E, 67.3° N), temperate (72.2° W, 42.5° N), and tropical (60.0° W, 3.1° S) forest sites. The boreal forest site is located in Eurasia, dominated by needleleaf evergreen trees, and has a relatively dry and cold climate (Mean Annual Precipitation [MAP] 800 mm/year; Mean Annual Temperature [MAT] −6 ° C). The deciduous broadleaf temperate forest site is located at Harvard Forest with a moist and warm climate (MAP 1,200 mm/year; MAT 10 ° C). The tropical site is an evergreen broadleaf rainforest (Manaus, Brazil) with hot and wet climate (MAP 2,200 mm/year; MAT 27 ° C). These sites were selected due to their representative background climates and plant functional types representing high-, middle-, and low-latitude ecosystems.

Characteristics for the three sites were extracted from the standard global model datasets, which are used to set plant functional types, soil properties, and so forth. (Zhu et al., 2019). The simulation protocol for all three sites follows the standard approach described in Oleson et al. (2010). The model is first spun up (i.e., brought close to equilibrium) using an accelerated soil decomposition approach (Koven et al., 2013) for 1,000 years, followed by a 400-year regular spinup with regular soil decomposition. Soil phosphorus pools were initialized from an observationally derived data set (Yang et al., 2013) at the beginning of the regular spinup. The spinup simulations were forced with repeated meteorology (GSWP reanalysis; Dirmeyer et al., 2006; from years 1901 to 1920) and constant atmospheric CO_2 mole fraction (285 ppm). We evaluated the model's spun-up soil carbon stock to ensure changes at the end of the spinup are less than 0.01%/year. After spinup, the model was run in a transient simulation from 1850 to 2010 with GSWP reanalysis forcing (Dirmeyer et al., 2006), transient CO_2 concentrations, N deposition (Lamarque et al., 2005), and P deposition (Mahowald et al., 2008). Atmospheric $\Delta^{14}C$ is prescribed to be zero during spinup simulations and using observed values during transient simulations (Levin et al., 2010). The spinup procedure ultimately determines the distribution of radiocarbon in the vector of initial carbon contents. Other spinup procedures may lead to different initial radiocarbon distributions; however, this is an issue particular for each simulation setup without relevance for the reconstruction of the model in the transient simulation.

We performed analyses using different approaches to reconstruct ELMv1-ECA predictions from model output. To characterize the effect of temporal discretization, we applied the DTA approach using ELMv1-ECA output averaged over 1- and 10-day time steps. This analysis will inform required output frequency for other models to apply model reconstruction.

3. Results and Discussion

We first give two examples with simplified systems (i.e., one- and two-compartment systems) to demonstrate characteristics of the Continuous-Time Approximation (CTA) and then present comparisons using Discrete-Time Approximation (DTA) to the full ELMv1-ECA predictions.

3.1. One compartment Example

In some cases, it is useful to represent soil carbon as one single compartment with an aggregated decomposition rate (Olson, 1963). This assumption may have important limitations to represent complex dynamics in soils (Manzoni & Porporato, 2009; Sierra & Müller, 2015), but it can be very useful to introduce complex concepts using a simplified system. We consider first a one compartment system with constant inputs and rates that represents the accumulation of carbon over time starting with a known initial amount of carbon

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\,C(t) = -0.04\,C(t) + 3, \quad t > 1909, \\ &C(1909) = 40. \end{aligned} \tag{8}$$

The CTA approach exactly matches the system's analytical solution despite the relatively small number of data points available for the reconstruction, $n = 3, 6$ (Figure 1).

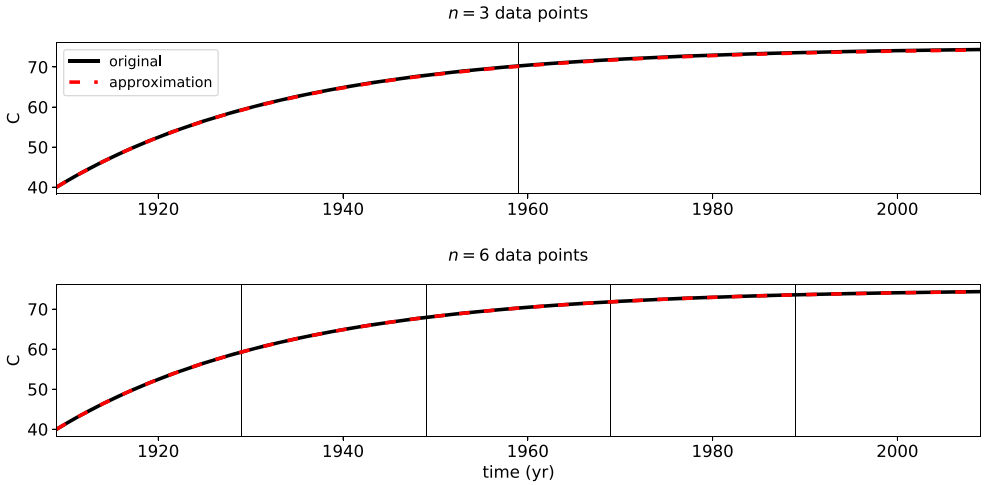

Figure 1. C stocks over time for a one-pool system with constant inputs and cycling rates. The solid black lines represents the carbon content of the original system (8); the dashed red lines represents the carbon content obtained with the CTA. Different panels represent a different number of given equidistant data points, represented by vertical lines.

However, the CTA can give approximations with large errors in the time-dependent case. Consider the one compartment system with time-dependent inputs beginning in year 1909:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\,C(t) = -0.04\,C(t) + \left[3 + \sin(t/20)\right], \quad t > 1919, \\ &C(1919) = 40. \end{aligned} \tag{9}$$

The CTA improves with increasing number n of data points used in the approximation (Figure 2). For small number of data points, the CTA can miss important dynamics occurring inside the time step. For instance, for the reconstruction using time steps of 30 years, the convexity of the curve is completely different between the original model and the reconstruction. Therefore, care must be taken in using the CTA when important changes in dynamics occur inside the time step.

3.2. Multicompartment Test

As a preliminary test before demonstrating the approach for the full ELMv1-ECA model reconstruction, we evaluated the approach for several configurations of multiple C pools and nonlinear time dependencies. We found that the accuracy of the CTA approach does not depend on the number of pools or the nonlinear time dependencies. The time step required for accurate reconstruction depends on the frequency of the forcing and nonlinear interactions among pools. For example, consider a two compartment system, which can represent the dynamics of a fast and a slow soil carbon pool (Andren & Katterer, 1997), initialized at year $t = 1909$ and propagated for 110 years:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix} = \begin{pmatrix} -\gamma_1(t) & 0.5\,\gamma_2 \\ \gamma_1(t) & -\gamma_2 \end{pmatrix}\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} + \begin{pmatrix} 0.5 \\ 0.5 + 0.5\,\sin(t/100) \end{pmatrix}, \\ &\begin{pmatrix} C_1(1909) \\ C_2(1909) \end{pmatrix} = \begin{pmatrix} 40 \\ 0 \end{pmatrix}, \end{aligned} \tag{10}$$

where $\gamma_1(t) = 0.1 + 0.05\,\sin(0.1\,t)$, $\gamma_2 = 0.05$.

In this case, data from the original model extracted at 5-year time steps leads to a very accurate CTA, whereas coarser temporal time steps lead to progressively worse approximations (Figure 3).

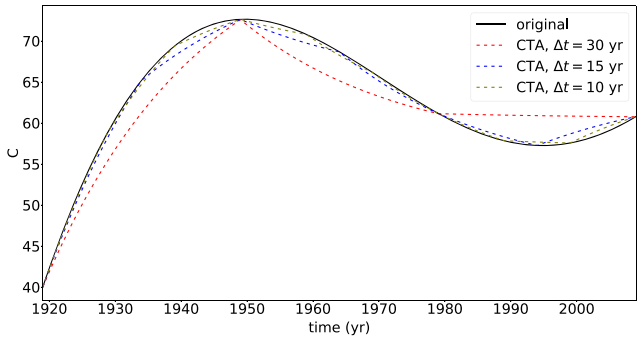

Figure 2. Transient C stocks of the one-compartment system with time-dependent inputs reconstructed by CTA for different data time steps Δt. Mean annual absolute bias is 3.92%, 0.96%, and 0.43% for $\Delta t = 30$, 20, and 10 years, respectively. The solid black line represents the carbon content of the original system (9); the dashed lines represent the carbon contents of the CTAs with different time steps.

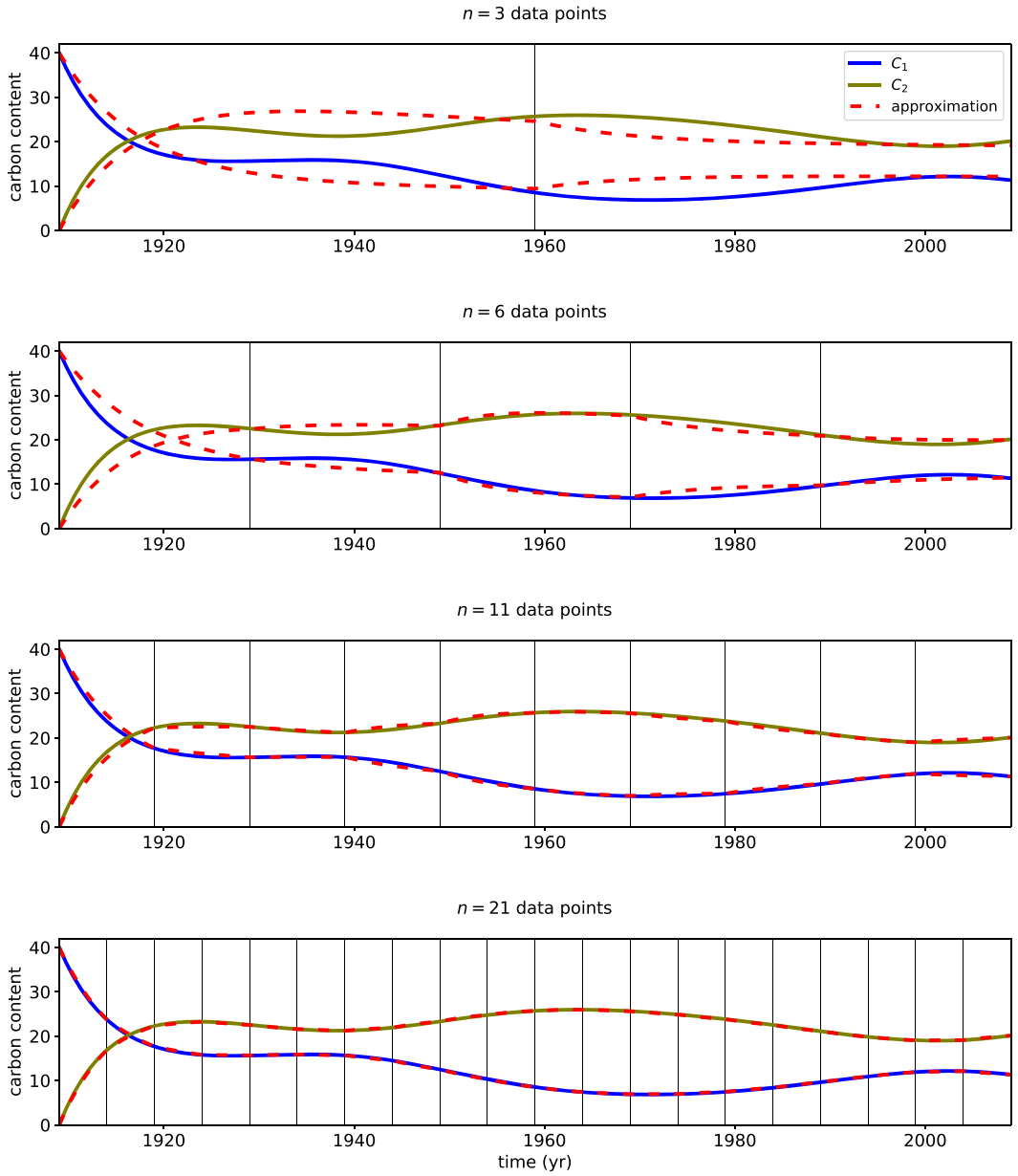

Figure 3. C stocks over time. The solid blue and olive curves represent the carbon contents of the two compartments of the original system (10); the dashed red lines represent the respective CTAs. Different panels represent different numbers of given equidistant data points, represented by vertical lines.

3.3. ELMv1-ECA Approximation

As long as the time-step is short enough, we can use either CTA or DTA to reconstruct the ELMv1-ECA model. CTA is generally more demanding in terms of computations because it needs to solve a system of ODEs for each time step. For practical reasons, we present here results using the DTA approach as described previously, compute radiocarbon values for all pools, and compare with the computations of radiocarbon performed in the original model. To evaluate errors related to the size of the time step, we performed simulations using 1- and 10-day ELMv1-ECA model outputs of ^{12}C inputs, fluxes between soil pools, and atmospheric $\Delta^{14}C$.

Our DTA accurately reproduced the ELMv1-ECA bulk soil ^{12}C and $\Delta^{14}C$ values across the three sites and at all soil depths (Figure 4). The bulk soil $\Delta^{14}C$ values shown in Figure 4 combined seven SOM component pools (coarse woody debris [CWD]; metabolic, cellulose, and lignin litter pools [Litters 1, 2, and 3,

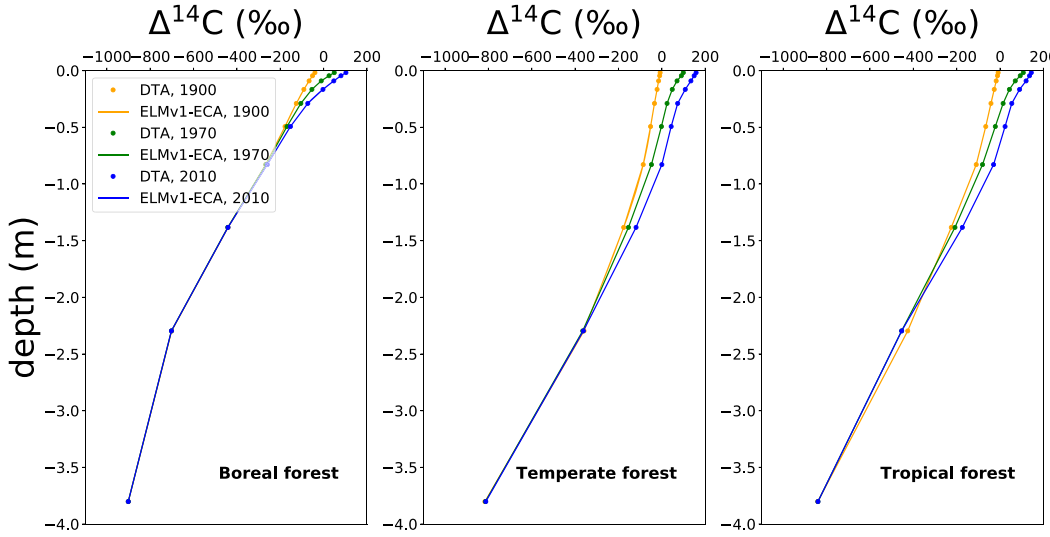

Figure 4. DTA using 10-day time step and ELMv1-ECA modeled $\Delta^{14}C$ values with depth in 1900, 1970, and 2010. The mean absolute DTA bias is less than 0.001% for $\Delta^{14}C$ in soils across all sites. Sites (left to right): boreal, temperate, and tropical. $\Delta^{14}C$ values decrease with depth and become more enriched post-1960s as they take up bomb-derived ^{14}C from the atmosphere.

respectively]; and fast, intermediate, and passive soil pools [Soils 1, 2, and 3, respectively]), which were separately reproduced with the DTA. The mean absolute biases were 0.0003%, 0.0014%, and 0.0003% for soil, litter, and coarse woody debris, respectively.

The DTA also accurately reproduced temporal variability of ELMv1-ECA modeled $\Delta^{14}C$ values for all these pools over the 110-year simulation from 1901 to 2010 (Figure 5). For all sites, the mean absolute error was <0.001% for soil carbon stocks and $\Delta^{14}C$ values at all depths (Table 1). These extremely low reconstruction errors are well within model uncertainty and far less than within-gridcell observed spatial heterogeneity of soil carbon stocks and radiocarbon values (Chen et al., 2019; Lawrence et al., 2019) therefore, tolerable within the context of soil C modeling.

During the 1960s, the radiocarbon concentration in the atmosphere increased sharply due to nuclear weapons testing (generally known as the bomb spike). The CWD, litter, and fast soil pools responded relatively rapidly to this change, as atmospheric $^{14}CO_2$ rapidly cycled through these pools. This bomb carbon ^{14}C signal was also evident in the top meter of soil (Figure 4). In contrast, the deeper and slower soil pools (i.e., intermediate and passive) had a more lagged and damped response to the atmospheric bomb-radiocarbon perturbation.

Although the relative errors were extremely low across all pools over time, the percent error was larger for the CWD and litter pools than for the soil pools, with the error in the litter pool reaching a maximum during the bomb-spike period at around 0.2% (Figure 6). The mean absolute percent errors for DTA ^{14}C values were, in most cases, at least 2 orders of magnitude larger than those for ^{12}C across sites and pools (Table 1). The DTA ^{14}C error increased for larger time steps, being an order of magnitude larger for CWD and litter pools in the 10-day versus 1-day time step cases. However, the mean absolute percent error for the 10-day time step case remained very low: $< 1 \times 10^{-4}$ % for ^{12}C and < 0.01 % for ^{14}C for all pools across different sites (Table 1), remaining well within model uncertainty bounds and soil heterogeneity.

3.4. Performance and Future Application

Using a 1-day time step, the DTA accurately matched ELMv1-ECA modeled $\Delta^{14}C$ values (<0.001% relative error for most sites and C pools). Although the error rate increases for larger time steps, acceptable low error (<0.01% relative error) is still achieved with a 10-day time step (Table 1). These results suggest that the error would likely remain within reasonable bounds for a 30-day time step, which is common output from most ESMs.

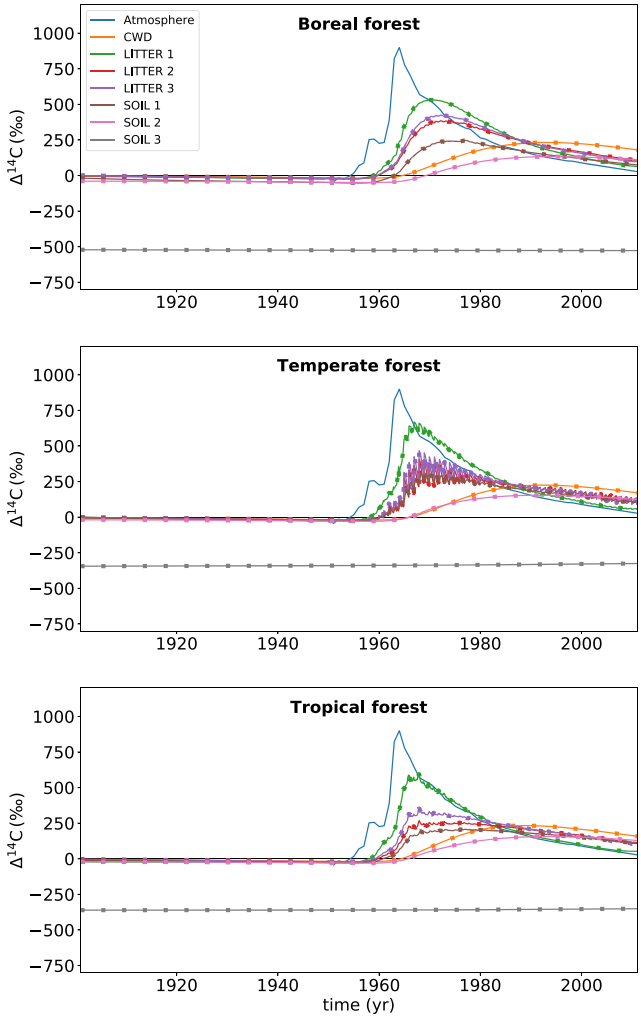

Figure 5. $\Delta^{14}C$ values in the different pools aggregated for all depths through time (Lines: ELMv1-ECA; dots: DTA using 10-day time step). Sites (top to bottom): boreal, temperate, and tropical forests. Litters 1, 2, and 3 denote metabolic, cellulose, and lignin litters. Soils 1, 2, and 3 are fast, median, and passive soil pools. Atmospheric $\Delta^{14}C$ values (blue) increase in response to nuclear bomb testing in the 1960s. Fast cycling pools respond more rapidly to the atmospheric $\Delta^{14}C$ changes. Slower cycling pools show a lagged and damped response. Differences in DTA and ELMv1-ECA modeled values are imperceptible in this figure.

We demonstrated here that the CTA and DTA approaches can be applied to numerical output from land carbon models to reconstruct a linear version of the model that very closely matches the output of the original model. The approximations can then be used to simulate isotope (e.g., ^{13}C or ^{14}C) dynamics consistent with the original model formulation of ^{12}C dynamics. Furthermore, the approximations can be used for both single-depth and vertically resolved soil radiocarbon models run at different spatial and temporal resolutions. The approach is not only restricted to soils but can also be used to reconstruct whole ecosystem carbon and isotope dynamics. We recommend using the DTA for reconstructing ESM-scale models, since the CTA makes implicit assumptions about what the original model does between time steps, and these assumptions may be unjustified in some cases depending on time-step size.

This DTA approach has widespread applicability, as it can be used across a range of models independent of their underlying structure. The approach does not require the model to implement additional isotope dynamics or tracers, but simply to output ^{12}C stocks and fluxes. For maximum accuracy, all model stocks and fluxes are required (e.g., all individual transfers between all pools represented in the model). Aggregating C stocks and fluxes results in greatly increased uncertainties. However, given the uncertainties inherent in

Table 1
Relative Error and Error Ratios of a 10- and a 1-Day Time Step Reconstruction of ELMv1-ECA

Relative error for 10-day time step (%):				Relative error for 1-day time step (%):				Ratio of 10- to 1-day errors			
^{12}C				^{12}C				^{12}C			
	CWD	Litter	Soil		CWD	Litter	Soil		CWD	Litter	Soil
Boreal	1.21e–04	2.48e–03	2.28e–06	Boreal	1.21e–04	2.48e–03	2.28e–06	Boreal	1.0	1.0	1.0
Temperate	9.61e–07	1.22e–05	1.34e–06	Temperate	9.61e–07	1.22e-05	1.34e-06	Temperate	1.0	1.0	1.0
Tropical	1.04e–06	3.78e–06	1.03e–06	Tropical	1.03e–06	3.76e–06	1.06e–06	Tropical	1.0	1.0	1.0
^{14}C				^{14}C				^{14}C			
Boreal	2.39e–03	6.40e–03	2.63e–04	Boreal	2.62e–04	2.34e–03	3.68e–04	Boreal	9.1	2.7	0.7
Temperate	2.48e–03	1.41e–02	6.46e–04	Temperate	2.55e–04	1.41e–03	2.56e–04	Temperate	9.7	10.0	2.5
Tropical	2.79e–03	4.30e–03	6.17e–04	Tropical	2.96e–04	4.77e–04	2.73e–04	Tropical	9.4	9.0	2.3
Ratio of ^{14}C error to ^{12}C error				Ratio of ^{14}C error to ^{12}C error							
Boreal	20	3	115	Boreal	2	1	161				
Temperate	2,578	1,155	482	Temperate	262	116	195				
Tropical	2,686	1,138	599	Tropical	288	127	258				

ESM land model soil carbon dynamics, reconstruction error may be lower than uncertainty in model structure and parameterization (Chen et al., 2019). Error arising from pool aggregation is of particular importance when models report carbon stocks only for aggregated pools (e.g., sum of all soil carbon pools in a particular soil layer). In such cases, the original and reconstructed model will differ in structure, leading to error in the reconstruction. This error would probably depend on the number and size of pools being aggregated, and on the temporal resolution of the output. Numerical errors and instabilities may also arise from empty or very small pools. For instance, we observed larger relative errors in the relatively small deep litter and CWD pools.

The potential of this approach to produce radiocarbon predictions across models that do not explicitly represent it will allow us valuable data-model comparisons and enable use of radiocarbon as a constraint for parameter estimation (He et al., 2016)

Although radiocarbon is commonly associated with C age, it cannot be directly interpreted as such in soils, due to mixing of C inputs of different ages in the open system (Trumbore et al., 2016). Instead, radiocarbon measurements and model output are primarily of value as a tracer, and for direct model-data comparison. In addition, by using a state-transition matrix approach as we did here, there is the added possibility to calculate age and transit time distributions for the system (Metzler et al., 2018). In the future, this approach could be applied to compute radiocarbon values, ages, and transit times across a suite of models and evaluate the time carbon spends in terrestrial ecosystems. This approach would provide meaningful comparisons of C ages and transit times between model structures and transient scenarios.

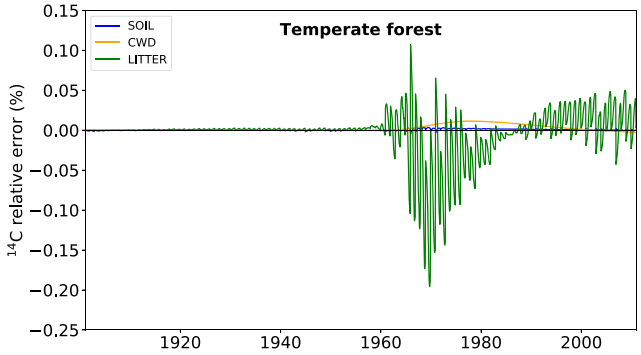

Figure 6. Relative error in radiocarbon values for CWD, aggregated litter, and soil pools over time for the temperate forest site. ^{14}C relative error is approximately 2 orders of magnitude larger than ^{12}C error but is always small in comparison to model uncertainty and spatial heterogeneity (Chen et al., 2019; Lawrence et al., 2019). The largest biases in the litter pool follow the bomb spike in atmospheric ^{14}C concentrations and vary seasonally.

In addition, the reconstruction we propose of numerical model output as a dynamical system can be used to compute other system-level diagnostics of the terrestrial carbon cycle. For instance, Luo et al. (2017) proposed a set of diagnostics to evaluate carbon storage capacity and potential from transient simulations and developed a framework to assess model performance based on the ability to trace particular components of ecosystems (Luo et al., 2015; Xia et al., 2013). From the numerical model reconstructions we proposed here, it is possible to perform these analyses without the need to analytically rewrite existing models in matrix form. This means that it is still possible to obtain time-dependent versions of these matrices and vectors in numeric form, without the need to explicitly write by hand all equations of the model. This can be an advantage for complex models with multiple interactions between the carbon cycle and the water, nutrient, and energy cycles.

4. Conclusions

Numerical output from carbon cycle models can be used to reconstruct a linear version of the model that exactly matches the solution trajectory of the original model. This reconstruction can be used to compute other quantities not included in the original model such as isotope dynamics and other system-level metrics. The approach consists of computing a state transition matrix that contains all information of the dynamics of the system. Since available model output is always reported in discrete time steps, we proposed here two methods to approximate the original system, approximating a continuous system between time steps or approximating a discrete dynamical system. As long as the time step of the numerical output is sufficiently small to capture complex dynamics, both approaches provide very good approximations to the solution trajectory of the original system.

Radiocarbon is a valuable tracer to evaluate, understand, and improve soil carbon models. However, most land models do not currently include explicit representation of soil radiocarbon dynamics, precluding comparison with observations. Here we demonstrated how continuous and discrete time approaches can reconstruct soil radiocarbon profiles using only land model ^{12}C stocks and fluxes and atmospheric ^{14}C values. We applied the approach here to the ELMv1-ECA model (Riley et al., 2018; Tang & Riley, 2018; Zhu et al., 2019), which explicitly represents soil organic matter ^{14}C dynamics with seven soil pools and ten vertical layers, and showed that our approach very accurately reproduces the full model simulations (bias < 0.01%). The approach can also be used for the new suite of CMIP6 simulations to evaluate whether the models are accurately representing soil C age and transit time, in addition to soil carbon stock (He et al., 2016; Lawrence et al., 2019). For applications to CMIP6 models, it is required that the models report all pools and all fluxes among pools as requested in the tier-2 output requirements of C4MIP (Jones et al., 2016).

Acknowledgments
Funding was provided by the Max Planck Society and the German Research Foundation through its Emmy Noether Program (SI 1953/2–1). A. M. H. received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (Grant Agreement 695101 (14Constraint)). W. J. R. and Q. Z. were funded by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract DE-AC02-05CH11231 as part of the Regional and Global Climate Modeling (RGCM), RUBISCO SFA, and the E3SM project. All code required to reproduce the results presented in this manuscript has been permanently archived (at https://doi.org/10.5281/zenodo.3520048).

References

Ahrens, B., Braakhekke, M. C., Guggenberger, G., Schrumpf, M., & Reichstein, M. (2015). Contribution of sorption, DOC transport and microbial interactions to the 14C age of a soil organic carbon profile: Insights from a calibrated process model. *Soil Biology and Biochemistry*, *88*, 390–402.

Anderson, D. H. (1983). *Compartmental modeling and tracer kinetics* (Vol. 50). Berlin, Heidelberg: Springer Science & Business Media.

Andren, O., & Katterer, T. (1997). Icbm: The introductory carbon balance model for exploration of soil carbon balances. *Ecological Applications*, *7*(4), 1226–1236.

Brockett, R. W. (2015). *Finite dimensional linear systems* (Vol. 74). Cambridge, Massachusetts: SIAM.

Chen, J., Zhu, Q., Riley, W. J., He, Y., Randerson, J. T., & Trumbore, S. (2019). Comparison with global soil radiocarbon observations indicates needed carbon cycle improvements in the E3SM land model. *Journal of Geophysical Research: Biogeosciences*, *124*, 1098–1114. https://doi.org/10.1029/2018JG004795

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The international land model benchmarking (ILAMB) system: design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*, 2731–2754. https://doi.org/10.1029/2018MS001354

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., & 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.

Dwivedi, D., Riley, W., Torn, M., Spycher, N., Maggi, F., & Tang, J. (2017). Mineral properties, microbes, transport, and plant-input profiles control vertical distribution and age of soil carbon stocks. *Soil Biology and Biochemistry*, *107*, 244–259.

Graven, H., Allison, C. E., Etheridge, D. M., Hammer, S., Keeling, R. F., Levin, I., et al. (2017). Compiled records of carbon isotopes in atmospheric CO_2 for historical simulations in CMIP6. *Geoscientific Model Development*, *10*(12), 4405–4417.

He, Y., Trumbore, S. E., Torn, M. S., Harden, J. W., Vaughn, LydiaJ. S., Allison, S. D., & Randerson, J. T. (2016). Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century. *Science*, *353*(6306), 1419–1424.

Jacquez, J. A., & Simon, C. P. (1993). Qualitative theory of compartmental systems. *SIAM Review*, *35*(1), 43–79.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., et al. (2016). C4MIP – the coupled climate–carbon cycle model intercomparison project: Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(8), 2853–2880.

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.

Lamarque, J.-F., Kiehl, J., Brasseur, G., Butler, T., Cameron-Smith, P., Collins, W., et al. (2005). Assessing future nitrogen deposition and carbon cycle feedback using a multimodel approach: Analysis of nitrogen deposition. *Journal of Geophysical Research*, *110*, D19303. https://doi.org/10.1029/2005JD005825

Lawrence, C. R., Beem-Miller, J., Hoyt, A. M., Monroe, G., Sierra, C. A., Stoner, S., et al. (2019). An open source database for the synthesis of soil radiocarbon data: Israd version 1.0. *Earth System Science Data Discussions*, *2019*, 1–37.

Levin, I., Naegler, T., Kromer, B., Diehl, M., Francey, R., Gomez-Pelaez, A., et al. (2010). Observations and modelling of the global distribution and long-term trend of atmospheric $^{14}CO_2$. *Tellus B: Chemical and Physical Meteorology*, *62*(1), 26–46.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., et al. (2016). Toward more realistic projections of soil carbon dynamics by earth system models. *Global Biogeochemical Cycles*, *30*, 40–56. https://doi.org/10.1002/2015GB005239

Luo, Y., Keenan, T. F., & Smith, M. (2015). Predictability of the terrestrial carbon cycle. *Global Change Biology*, *21*(5), 1737–1751.

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. http://doi.org/10.5194/bg-14-145-2017

Mahowald, N., Jickells, T. D., Baker, A. R., Artaxo, P., Benitez-Nelson, C. R., Bergametti, G., et al. (2008). Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. *Global Biogeochemical Cycles*, *22*, GB4026. https://doi.org/10.1029/2008GB003240

Manzoni, S., & Porporato, A. (2009). Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry*, *41*(7), 1355–1379.

Metzler, H., Müller, M., & Sierra, C. A. (2018). Transit-time and age distributions for nonlinear time-dependent compartmental systems. *Proceedings of the National Academy of Sciences*, *115*(6), 1150–1155.

Oleson, K., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., et al. (2010). Technical description of version 4.0 of the Community Land Model (CLM).

Olson, J. S. (1963). Energy storage and the balance of producers and decomposers in ecological systems. *Ecology*, *44*(2), 322–331. https://doi.org/10.2307/1932179

Parton, W. J., Scurlock, J. M. O., Ojima, D. S., Gilmanov, T. G., Scholes, R. J., Schimel, D. S., et al. (1993). Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochemical Cycles*, *7*(4), 785–809.

Riley, W. J., Zhu, Q., & Tang, J. (2018). Weaker land–climate feedbacks from nutrient uptake during photosynthesis-inactive periods. *Nature Climate Change*, *8*(11), 1002.

Sierra, C. A., & Müller, M. (2015). A general mathematical framework for representing soil organic matter dynamics. *Ecological Monographs*, *85*, 505–524.

Sierra, C. A., Müller, M., & Trumbore, S. E. (2014). Modeling radiocarbon dynamics in soils: SoilR version 1.1. *Geoscientific Model Development*, *7*(5), 1919–1931.

Sulman, B. N., Moore, JessicaA. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., et al. (2018). Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry*, *141*(2), 109–123.

Tang, J. Y., & Riley, W. J. (2013). A total quasi-steady-state formulation of substrate uptake kinetics in complex networks and an example application to microbial litter decomposition. *Biogeosciences*, *10*(12), 8329–8351.

Tang, J., & Riley, W. J. (2018). Predicted land carbon dynamics are strongly dependent on the numerical coupling of nitrogen mobilizing and immobilizing processes: a demonstration with the E3SM land model. *Earth Interactions*, *22*(11), 1–18.

Thompson, M. V., & Randerson, J. T. (1999). Impulse response functions of terrestrial carbon cycle models: method and application. *Global Change Biology*, *5*(4), 371–394.

Tifafi, M., Camino-Serrano, M., Hatté, C., Morras, H., Moretti, L., Barbaro, S., et al. (2018). The use of radiocarbon 14C to constrain carbon dynamics in the soil module of the land surface model ORCHIDEE (SVN r5165). *Geoscientific Model Development*, *11*(12), 4711–4726. https://doi.org/10.5194/gmd-11-4711-2018

Trumbore, S. E., Sierra, C. A., & Hicks Pries, C. E. (2016). Radiocarbon nomenclature, theory, models, and interpretation: Measuring age, determining cycling rates, and tracing source pools, *Radiocarbon and climate change: Mechanisms, applications and laboratory techniques* (pp. 45–82). Switzerland.

Xia, J., Luo, Y., Wang, Y.-P., & Hararuk, O. (2013). Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Global Change Biology*, *19*(7), 2104–2116. https://doi.org/10.1111/gcb.12172

Yang, X., Post, W. M., Thornton, P. E., & Jain, A. (2013). The distribution of soil phosphorus for global biogeochemical modeling. *Biogeosciences*, *10*, 2525–2537.

Zhu, Q., & Riley, W. J. (2015). Improved modelling of soil nitrogen losses. *Nature Climate Change*, *5*(8), 705.

Zhu, Q., Riley, W. J., & Tang, J. (2017). A new theory of plant–microbe nutrient competition resolves inconsistencies between observations and model predictions. *Ecological applications*, *27*(3), 875–886.

Zhu, Q., Riley, W. J., Tang, J., Collier, N., Hoffman, F. M., Yang, X., & Bisht, G. (2019). Representing nitrogen, phosphorus, and carbon interactions in the E3SM land model: Development and global benchmarking. *In revision at Journal of Advances in Modeling Earth Systems*, *11*, 2238–2258. https://doi.org/10.1029/2018MS001571

Zhu, Q., Riley, W. J., Tang, J., & Koven, C. D. (2016). Multiple soil nutrient competition between plants, microbes, and mineral surfaces: model development, parameterization, and example applications in several tropical forests. *Biogeosciences*, *13*, 341–363.