

RESEARCH

Open Access



A traceability analysis system for model evaluation on land carbon dynamics: design and applications

Jian Zhou^{1,2}, Jianyang Xia^{1,2*} , Ning Wei^{1,2}, Yufu Liu^{3,4}, Chenyu Bian^{1,2}, Yuqi Bai^{3,4} and Yiqi Luo⁵

Abstract

Background: An increasing number of ecological processes have been incorporated into Earth system models. However, model evaluations usually lag behind the fast development of models, leading to a pervasive simulation uncertainty in key ecological processes, especially the terrestrial carbon (C) cycle. Traceability analysis provides a theoretical basis for tracking and quantifying the structural uncertainty of simulated C storage in models. Thus, a new tool of model evaluation based on the traceability analysis is urgently needed to efficiently diagnose the sources of inter-model variations on the terrestrial C cycle in Earth system models.

Methods: A new cloud-based model evaluation platform, i.e., the online traceability analysis system for model evaluation (TraceME v1.0), was established. The TraceME was applied to analyze the uncertainties of seven models from the Coupled Model Intercomparison Project (CMIP6).

Results: The TraceME can effectively diagnose the key sources of different land C dynamics among CMIP6 models. For example, the analyses based on TraceME showed that the estimation of global land C storage varied about 2.4 folds across the seven CMIP6 models. Among all models, IPSL-CM6A-LR simulated the lowest land C storage, which mainly resulted from its shortest baseline C residence time. Over the historical period of 1850–2014, gross primary productivity and baseline C residence time were the major uncertainty contributors to the inter-model variation in ecosystem C storage in most land grid cells.

Conclusion: TraceME can facilitate model evaluation by identifying sources of model uncertainty and provides a new tool for the next generation of model evaluation.

Keywords: CMIP6, land carbon cycle, model evaluation, traceability analysis, uncertainty

Introduction

Earth system models are an essential tool for understating and predicting the interactions between ecological processes and environmental changes at the global scale (Eyring et al. 2016a; Bonan and Doney 2018). In the past three decades, the structural complexity of models has

been increasing rapidly, which is featured by the incorporation of more and more ecological processes (Xia et al. 2020). However, the comprehensive and systematic model evaluations usually lag behind the fast development of Earth system models, leading to a pervasive uncertainty in Earth system models on key ecological processes, especially terrestrial carbon (C) cycle (Friedlingstein et al. 2006; Bonan et al. 2019; Fisher and Koven 2020; Xia et al. 2020). For example, the large uncertainty on global land C sink has been existing in Earth system models since the 3rd assessment report of IPCC (Arora et al. 2020; Zarakas et al. 2020). One key challenge is that how model evaluation can increase its

* Correspondence: jyxia@des.ecnu.edu.cn

¹Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China

²Research Center for Global Change and Ecological Forecasting, East China Normal University, Shanghai 200241, China

Full list of author information is available at the end of the article

pace to systematically trace the model uncertainty back to the key sources. For the land C cycle in Earth system models, the varied model structure among models (Bonan and Doney 2018), parameterization of C-related processes (Cui et al. 2019; Luo and Schuur 2020), and external climate forcings (Ahlström et al. 2012; Hoffman et al. 2014) are three major uncertainty contributors. Thus, a traceability analysis tool for efficiently evaluating terrestrial C cycles in Earth system models is useful to accelerate the pace of model inter-comparisons and model-data comparisons as well as their feedbacks to model developments.

A few new analytical tools have been recently developed to facilitate the evaluation of Earth system models, such as the International Land Model Benchmarking (ILAMB) System (Hoffman et al. 2016; Collier et al. 2018), the Earth System Model Evaluation Tool (ESM-ValTool) (Eyring et al. 2016b; Eyring et al. 2020), and the Land surface Verification Toolkit (LVT) (Kumar et al. 2012). The evaluation methods of these tools mainly focus on measuring the biases of a specific predicted variable across models or between models and observations using statistical metrics. For example, the ILAMB system uses a set of statistical methods to construct a data-driven scoring system to benchmark global C cycle models (Collier et al. 2018). These new tools have greatly increased the efficiency of model evaluations for Earth system models (Eyring et al. 2019). Even so, it is still difficult to quantitatively trace the structural sources of uncertainty among models. For the terrestrial C cycle, a traceability analysis has been developed to diagnose the inter-model variations in the land C cycle based on its fundamental properties (Xia et al. 2013; Luo et al. 2017). This method provides a traceability framework that can decompose the land C dynamics into a few traceable components, such as net primary productivity (NPP), C residence time, and environmental factors (temperature and precipitation). The traceability analysis has been applied to some local-level model evaluations (Jiang et al. 2017; Rafique et al. 2017). However, it remains unclear whether the traceability analysis is applicable to Earth system models, which simulate the terrestrial C cycle at a global scale. Thus, developing the traceability analysis as an available tool for analyzing Earth system models, especially those who have participated in the Coupled Model Intercomparison Project (CMIP), can effectively facilitate the simulations of the global terrestrial C cycle and its feedbacks to climate change.

The model evaluation process usually consists of three steps: downloading the model output data and archiving them locally, pre-processing the data to be suitable for analyses, and utilizing a dedicated program to finish the evaluation. Both the data volumes of model outputs and

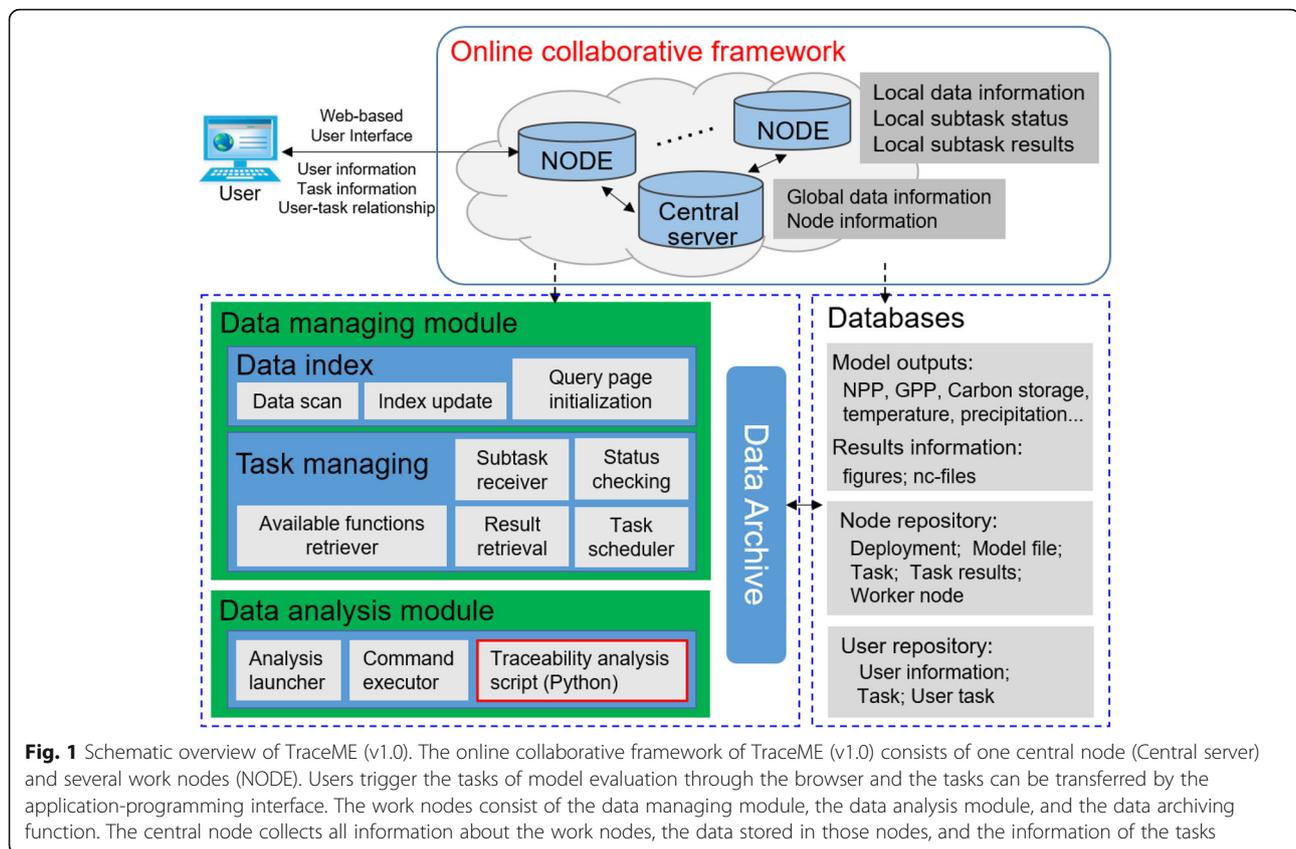
data products have been increased rapidly in the recent CMIPs (Overpeck et al. 2011; Stockhause and Lautenschlager 2017). It becomes more and more time-consuming for the routine evaluation of single or multiple models through downloading, managing, pre-processing, and analyzing the data comprehensively on their local equipment (Bai and Di 2012; Xu et al. 2019). Fortunately, cloud-based technology facilitates the processing of distributed big data and provides user-friendly web interfaces. Such web-based technology has been used in the field of ecological modeling and model evaluation (Abramowitz 2012; Huang et al. 2019). The advantage of the web-based cloud technology can help the researchers to focus on the scientific questions rather than processing the data.

Here, we introduce a new online traceability analysis system for model evaluation (TraceME v1.0), which can be applied to analyze the uncertainty of the terrestrial C cycle in the ongoing CMIP6. The specific aims of this study are (1) describing the design and workflow of TraceME, including the overview of TraceME, the introduction of the traceability analysis method, and the available data; (2) using TraceME to evaluate the performance of seven CMIP6 models in simulating terrestrial C cycle; and (3) discussing the potential applications and the implications of TraceME for the next generation of model evaluation.

Materials and methods

Design of the TraceME

TraceME (v1.0) is an online framework for automatically analyzing and evaluating the performance of models using the traceability analysis. It is built on a collaborative analysis framework for distributed gridded environmental data (Collaborative Analysis Framework for Environmental data, CAFE; more details are described in Xu et al. (2019)) with different core functions and focuses. The basic cyberinfrastructure of TraceME consists of one central server (node) and more than one work node (Fig. 1). Work nodes can be set up in different data centers and can archive the data stored in these data centers. The central node is used to archive the descriptive information of each node and the information about the data stored on it and get the task request of users and send it to corresponding work nodes. In each node, it contains the data analysis module and the data management module. The data analysis module includes an analysis launcher, a command executor, and the traceability analytic script, to realize the traceability analysis and output the corresponding analysis results. The data managing module includes the data index submodule and task managing submodule. The data index submodule manages the descriptive information about data (data file name, storage path, and data attributes) stored



on each worker node. The task managing module is used for task submission, task dispatching, and task status/results query services on each node.

The web-based technology provides a straightforward way for users to interact with the system through a web browser and the model evaluation process of TraceME runs in the background. Users only need to filter data of interest from the entire system, and the selected data is then packaged into a task and delivered to the assigned work node for data processing, which includes data pre-processing, traceability analysis, and evaluation, and finally, the evaluation results are output and visualized for the users. The scientific workflow is essential for TraceME to realize online automated model evaluation.

Traceability analysis

The core functionality of TraceME is based on the framework of traceability analysis that is developed by Xia et al. (2013). This framework is extended to the transient dynamic by decomposing the C storage dynamics into a three-dimensional parameter space (Luo et al. 2017). The latter can be further partitioned into traceable components to track the sources of model uncertainty. In the framework of traceability analysis,

terrestrial C storage is at dynamic disequilibrium, which is collectively influenced by internal C-related processes, environmental forces, and their interactions (Luo and Weng 2011). Under given environmental conditions, the C storage of an ecosystem can reach the steady-state, which can be defined as C storage capacity (X_C). In the land C cycle model, we can obtain the X_C by spinning up the model to the steady-state (Xia et al. 2012). Because the external forces, such as climate, are never at steady state, so the X_C always deviates from the realistic C storage in natural ecosystems. Such deviation or difference between the transient C storage and X_C was defined as C storage potential (X_P) (Luo et al. 2017). The positive X_P means the potential of an ecosystem to store additional C while the negative X_P means the potential to lose C (Luo et al. 2017). Hence, the transient C storage of an ecosystem can be determined by X_C and X_P . Then, X_C is jointly determined by ecosystem C input (e.g., net primary production, NPP) and ecosystem C residence time (τ_E). As the net ecosystem C input, NPP is decomposed into gross primary production (GPP) and C use efficiency (CUE). CUE describes the capacity of an ecosystem to effectively absorb C from the atmosphere, which is defined as the ratio of NPP to GPP (DeLucia et al. 2007; Xia et al. 2017). The τ_E can be further traced

to the baseline C residence time (τ'_E) and the environmental scalar (ξ). τ'_E represents the ecosystem C residence time under optimal environmental conditions, which is usually determined by the preset soil properties and vegetation characteristics in the model (Xia et al. 2013). The ξ is influenced by several factors, such as climate, oxygen concentration, and land cover. The climate is the most common limiting factor in the land C cycle model. In this study, we focus on the effect of climate forcing (i.e., temperature and precipitation) on the ecosystem C residence time. The detail of the traceability analysis method is described in Xia et al. (2013), Luo et al. (2017), and Zhou et al. (2018).

Under the framework of traceability analysis, land C storage is ultimately attributed to its traceable components, which are related to the natural properties expressed by the model (Fig. 2). For example, GPP is the photosynthetic property of vegetation; baseline C residence time is related to the soil attributes (Fig. 2). To quantify the contributions of these traceable components to the uncertainty of models, we use a hierarchical partitioning method (Chevan and Sutherland 1991) to decompose the uncertainty of simulated C storage dynamics. This method can be used to calculate the independent effect of each explanatory variable ($x_1, x_2, x_3 \dots x_k$) on a single dependent variable (y). The independent effect of x_1 (I_{x1}) means the contribution of x_1 to the variable y , which is calculated by comparing the fit of all models (2^k possible models) including x_1 to that lacking x_1 by the hierarchical partitioning (Chevan and Sutherland 1991; Murray and Conner 2009). In

our system, we calculate the variance contribution of the variables using the “hier.part” package in R. Based on the relationships built by traceability, we first calculate the relative contributions of X_C and X_P to X . Then, the contributions of NPP and τ_E to X_C are calculated in their logarithmic form: $\ln(X_C)$, $\ln(\text{NPP})$ and $\ln(\tau_E)$. Third, the variation contributions of the components of NPP and C residence time are calculated in the same way. Finally, the contributions of these traceable components (GPP, CUE, baseline C residence time, temperature, and precipitation) can be calculated.

CMIP6 and modeling outputs

TraceME (v1.0) can be compatible with any model output that follows the NetCDF Climate and Forecast (CF) Metadata Convention (<http://cfconventions.org/>), for example, all data from CMIP5 and CMIP6. TraceME (v1.0) is a systematic framework for uncertainty analysis on the terrestrial C cycle for CMIPs. It requires a multivariable dataset to analyze and trace the sources of uncertainty in simulating ecosystem C storage. The time-series data of total ecosystem C storage are needed, which generally consist of living biomass C, litter C, and soil C pools. The time-series data of NPP, GPP, and forcing data (temperature and precipitation) are also used for further model inter-comparisons. In this study, the TraceME (v1.0) used CMIP6 model outputs as examples to describe the workflow of this platform. All data is from seven CMIP6 models (the release data before July 2019) and collected from ESGF (<http://esgf.llnl.gov/>) as shown in Table 1.

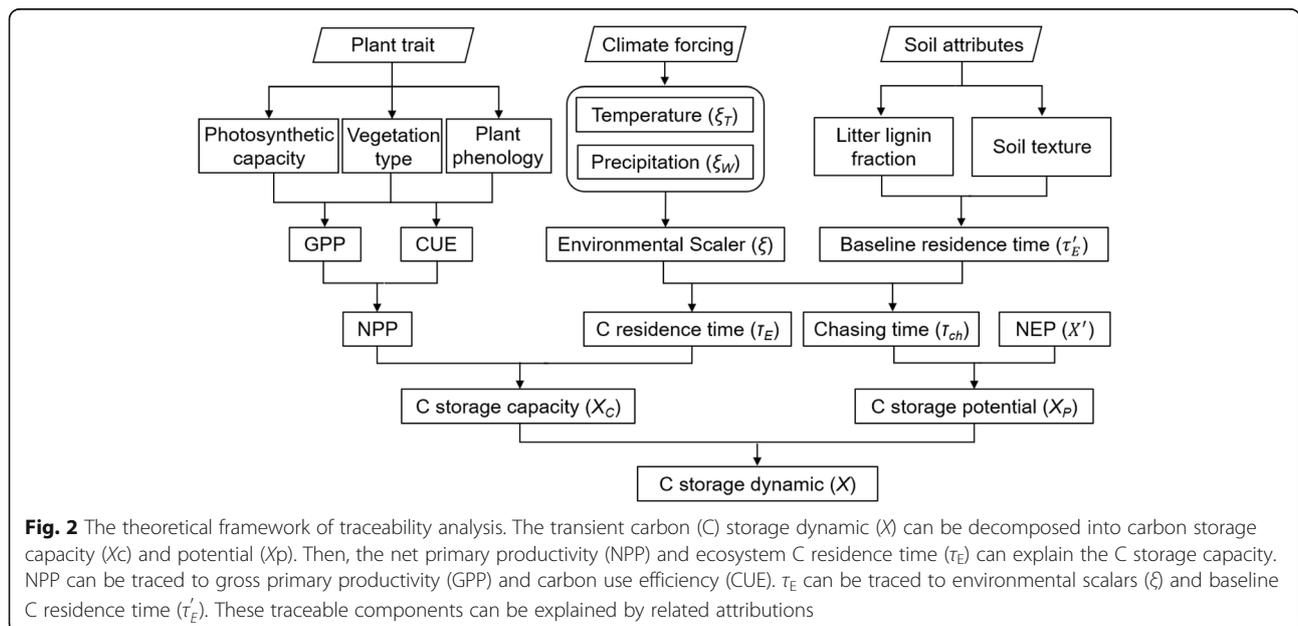


Fig. 2 The theoretical framework of traceability analysis. The transient carbon (C) storage dynamic (X) can be decomposed into carbon storage capacity (X_C) and potential (X_P). Then, the net primary productivity (NPP) and ecosystem C residence time (τ_E) can explain the C storage capacity. NPP can be traced to gross primary productivity (GPP) and carbon use efficiency (CUE). τ_E can be traced to environmental scalars (ξ) and baseline C residence time (τ'_E). These traceable components can be explained by related attributions

Table 1 The list of seven Earth system models used in this study from CMIP6

Institute	Earth system model	Land model
Beijing Climate Center, China Meteorological Administration, China	BCC-ESM1	BCC-AVIM2
Canadian Centre for Climate Modeling and Analysis, Victoria BC, Canada	CanESM5	CLASS-CTEM
National Center for Atmospheric Research, USA	CESM2	CLM5.0
IPSL Climate Modeling Centre, France	IPSL-CM6A-LR	ORCHIDEE
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, Japan	MIROC-ES2L	VISIT-e
National Center for Meteorological Research, Météo-France and CNRS laboratory, Toulouse, France	CNRM-ESM2-1	ISBA
EC-Earth Consortium, Europe	EC-Earth3-Veg	LPJ-GUESS

Results

Temporal dynamics of land carbon storage in CMIP6 models

TraceME (v1.0) provided an automatic traceability analysis for data of temporal interest, which can be used to evaluate the temporal dynamics of land C storage simulated by models. We used seven models that had submitted results in CMIP6 to analyze the uncertainty of these models in simulating historical land C storage from 1850 to 2014. From the results of TraceME, the temporal dynamics of global annual C storage simulated by different models were first calculated (Fig. 3a). The global annual C storage varied greatly among the seven models, ranging from 938.76 ± 11.36 to 2206.76 ± 50.14 Pg C (Fig. 3a). Decomposing the C storage into C storage capacity and potential, the C storage potential ranged considerably from about -21.66 ± 54.39 to 58.07 ± 57.62 Pg C (Fig. 3a). The C storage capacity of different models in response to external force was also quite different. For example, the lowest simulated C storage capacity was IPSL-CM6A-LR from 1850 to 2014, which was 944 ± 27.14 Pg C, and the other models were from about 1677.57 ± 57.21 to 2263.43 ± 106.61 Pg C (Fig. 3a). To further analyze the uncertainty of C storage capacity, the results of NPP and C residence time reflected the net C input capacity (38.48 ± 2.72 to 68.74 ± 5.88 Pg C year⁻¹) and the C turnover time of ecosystem (23.22 ± 1.75 to 56.23 ± 3.10 years) in the models (Figs. 3b, c and Fig. 4a). In detail, the lowest simulated NPP was CESM2 and the shortest C residence time was IPSL-CM6A-LR, while CanESM5 had the largest NPP and C residence time among all models (Figs. 3b, c and Fig. 4a).

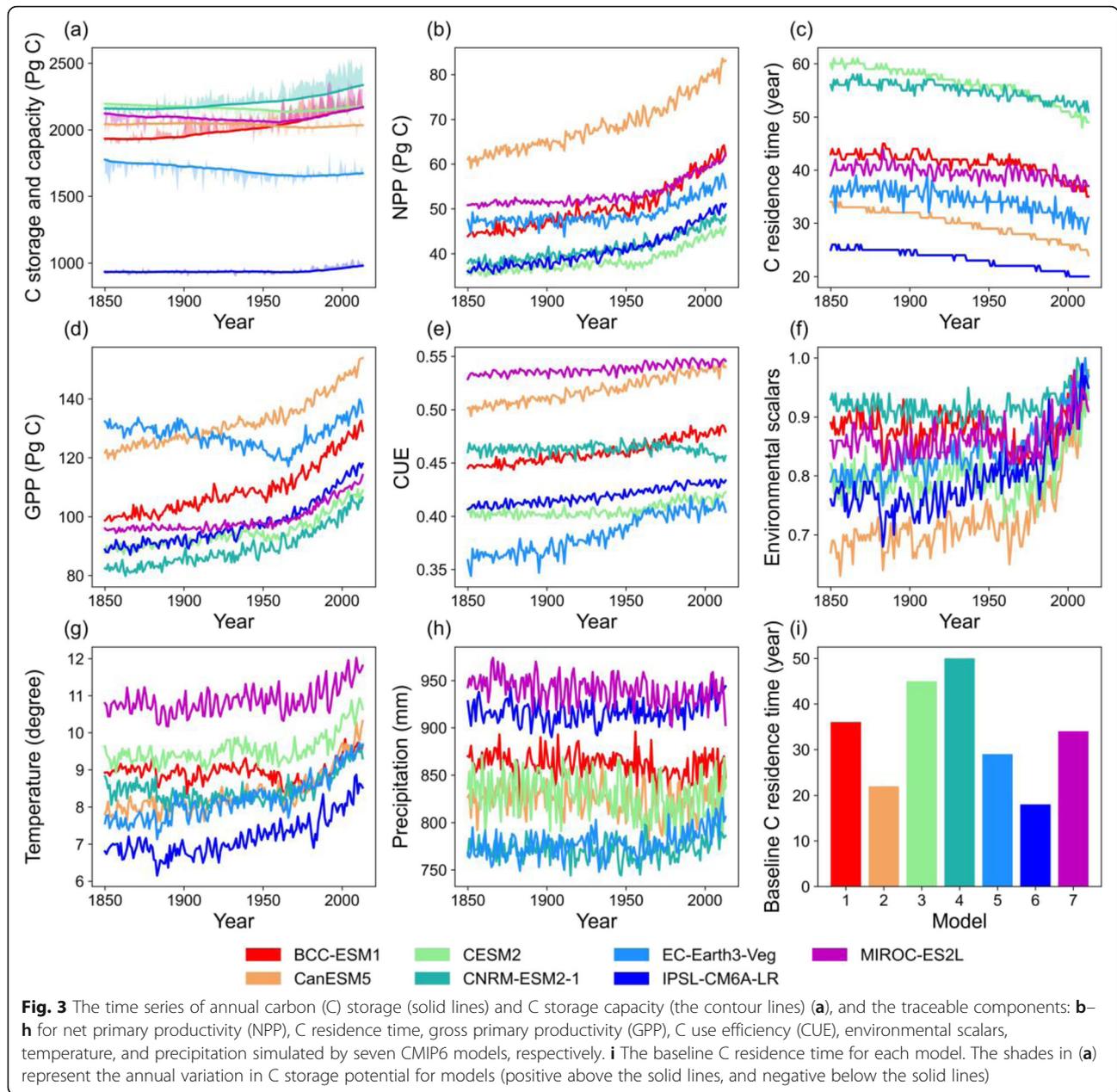
GPP and CUE were used to explain the uncertainty sources of NPP simulated by models (Figs. 3d, e and Fig. 4b). The differences of GPP and CUE in different models reflected the model's photosynthetic capacity and C transfer efficiency from the atmosphere to ecosystem biomass. Based on this process, TraceME could quantify the effects of models simulating photosynthesis and respiration on the uncertainty of NPP. For example, NPP simulated by CanESM5 and EC-Earth3-Veg had

larger uncertainty, which were 68.74 ± 5.88 and 48.96 ± 2.78 Pg C year⁻¹ respectively during 1850 to 2014, whereas their GPP was similar, which were 132.22 ± 8.18 and 127.72 ± 4.38 Pg C year⁻¹ respectively (Figs. 3b–e and Fig. 4b). Therefore, the uncertainty of NPP between the two models mainly came from CUE (0.52 ± 0.01 and 0.38 ± 0.02 , respectively), which was related to autotrophic respiration. Besides, to show the sources of C residence time, the uncertainties of baseline C residence time and environmental scalars were given in TraceME. For example, IPSL-CM6A-LR had the shortest C residence time (23.22 ± 1.75 years) than other models during 1850 to 2014, and compared with external forces, the main reason was that it had the shortest baseline C residence time (18 years) among all models (Figs. 3c, f–i, and Fig. 4c). Hence, the development of IPSL-CM6A-LR was suggested to pay more attention to the preset attributes of soil. Furthermore, the environmental scalar in TraceME here was the global annual scale. Its uncertainty reflected the variability of interannual variation of temperature and precipitation used in each model overall models rather than the direct difference of external forces among models (Figs. 3 f–h and Fig. 4c, d).

Overall, after analyzing the uncertainties of all traceable components, TraceME summarized the variance contributions of the components to the uncertainty of land C storage among models. This framework traced the uncertainty of land C storage to several sources. For example, the variation of land C storage among seven CMIP6 models was mainly from C residence time that contributed 74.8%, while NPP and the C storage potential contributed about 20.7% and 4.5%, respectively (Fig. 5). Comparing all traceable components, the variation in C storage simulated by these models was dominated by baseline C residence time (Fig. 5).

Different spatial distributions of land carbon storage among CMIP6 models

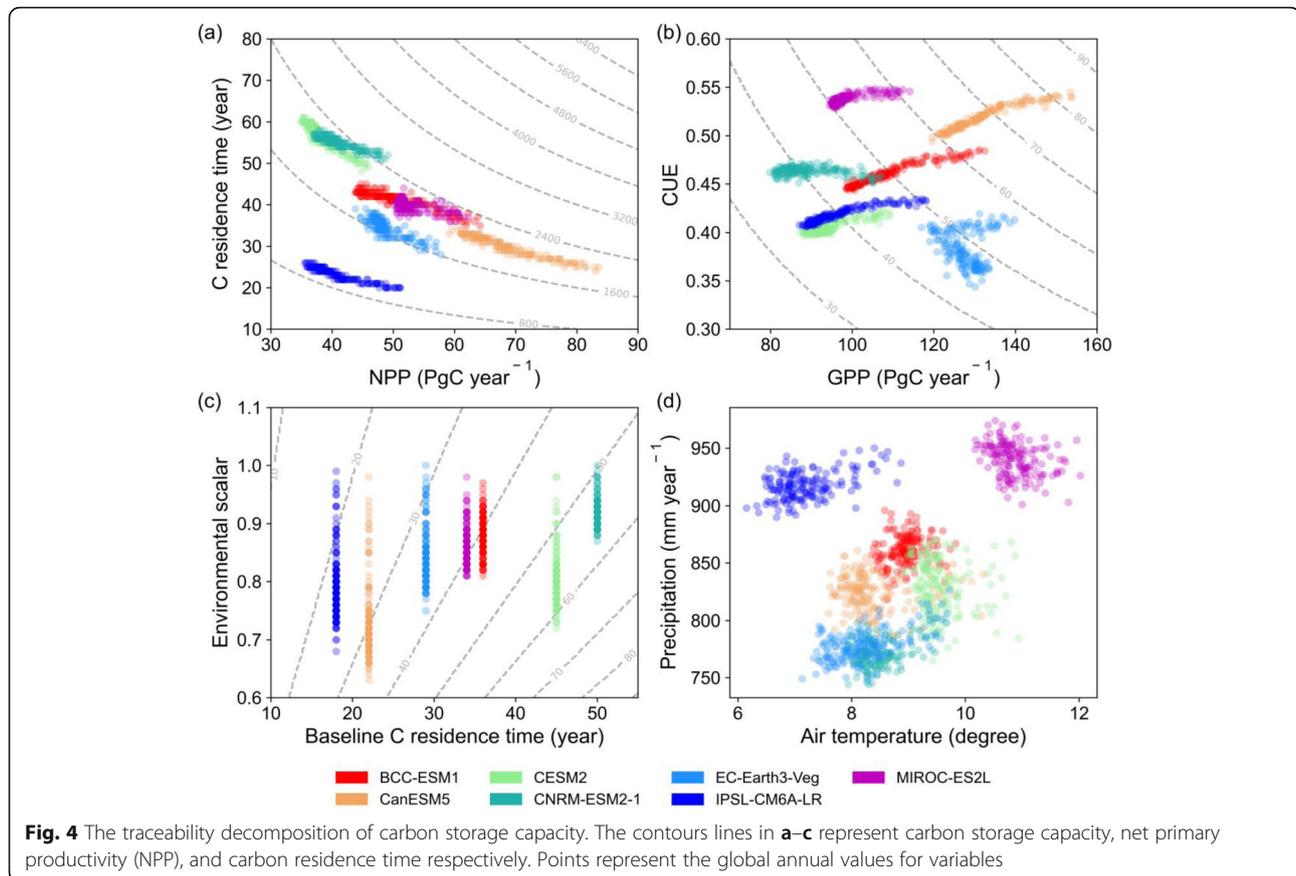
TraceME (v1.0) provided the ability to analyze the spatial uncertainty of models. It could trace the sources of the uncertainty of models in simulating C storage at each grid. From the results, the mean spatial pattern of



the seven models showed C storage in boreal regions was higher than in other regions (Fig. 6a). However, some models, such as IPSL-CM6A-LR, had no such spatial pattern (Fig. 7a), and the high variability of C storage simulated by these models also appeared in the boreal regions, such as Siberia and northern North America (Fig. 6b). To further research the sources of the uncertainty of models in simulating C storage, TraceME (v1.0) provided the spatial patterns of C storage capacity and C storage potential (Figs. 6 c–f and Fig. 7).

According to the traceability framework, the spatial distributions of NPP and C residence time were used to explain the uncertainty of land C storage capacity among

models (Fig. 7). From the results of seven CMIP6 models, the distribution of the variation in NPP among these models occurred in the lower latitude region, while the variation of C residence time was mainly distributed in the northern high latitude region (Fig. 8a and d). Following the workflow of TraceME (v1.0), the uncertainties of global distributions of NPP had a similar pattern to that of GPP (Fig. 8a–c). The distribution of the variation in baseline C residence time was mainly in the northern high latitude region and the Tibetan Plateau (Fig. 8e). To better guide model development, model evaluation needs to provide information on the spatial distribution of the dominant factor influencing the



simulation of land C storage. TraceME (v1.0) could analyze the variation contributions of all traceable components to land C storage at each grid and offered the spatial pattern of the dominant factor (Fig. 9a). For example, the baseline C residence time and GPP were the major contributors to the global distribution of the variation of simulated C storage by the seven models from CMIP6 (Fig. 9a). Compared to GPP, baseline C residence time dominated the uncertainties of simulated land C storage in northern high latitude, eastern Asian, and the northern part of South America (Fig. 9a).

Spatiotemporal changes in the dominant uncertainty sources of simulated carbon storage in CMIP6 models

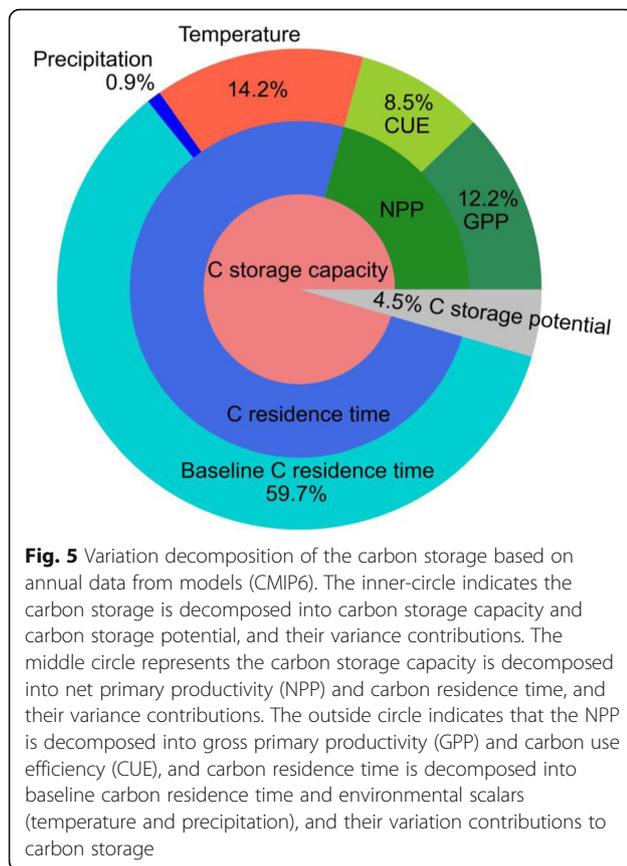
Assessing the performances of the model over different periods could provide a more comprehensive understanding of the model's ability to simulate land C storage. For example, the environmental scalars among the seven CMIP6 models had larger variability at the initial state (e.g., from 1850 to 1860) than those at the current state (e.g., 2004 to 2014) (Fig. 3f). It is necessary to examine whether and how the sources of model uncertainty change with time. For example, the dominant contributor to the inter-model variance of global land C storage was baseline C residence time from 1850–1860

to 2004–2014 (Fig. 9b, c). However, the contribution of C storage potential increased from 5.2% over 1850–1860 to 19.1% over 2004–2014 (Fig. 9b, c). In addition, GPP and C residence time were the major contributors to the inter-model variance of ecosystem C storage in most land grid cells (Fig. 9b, c). In the regions at northern high latitudes, GPP was the dominant contributor in more grid cells in the period of 1850–1860 than 2004–2014 (Fig. 9b, c).

Discussion

Evaluations on the uncertainty source of land C dynamics in CMIP6 models

The increase of model complexity and the rapid expansion of observational data volumes together promote the model evaluation into the next generation (Collier et al. 2018; Eyring et al. 2019; Xia et al. 2020). In our study, we introduce a new model evaluation platform, TraceME (v1.0), which uses traceability analysis and a collaborative cloud-based framework. As the core function of TraceME, the traceability analysis increases the traceability of the model evaluations (Luo et al. 2015). Rather than simply comparing the differences in simulated C storage among models, this method can trace and quantify the uncertainties to the traceable ecological



components (Figs. 3 and 7). For example, the annual C storage simulated by IPSL-CM6A-LR is much lower than other models, and TraceME can first track it to C storage capacity (Fig. 3a). Further analysis shows that the low estimates of ecosystem C storage capacity on the global scale in IPSL-CM6A-LR are mainly contributed from C residence time, especially the baseline C residence time (Figs. 3 and 4). Thus, TraceME not only shows the structural sources of the disagreement on global land C storage between models but also identifies the key uncertain component for a specific model for further development. Recent studies have highlighted the importance to develop model evaluations to explore and understand the sources of uncertainties in Earth system models (Lovenduski and Bonan 2017; Bonan and Doney 2018; Bonan et al. 2019). For example, the ILAMB package used the variable-to-variable relationships between metrics to benchmark Earth system models. Overall, TraceME gives model evaluation a new way to systematically trace the structural sources of the uncertainties in global C cycle models.

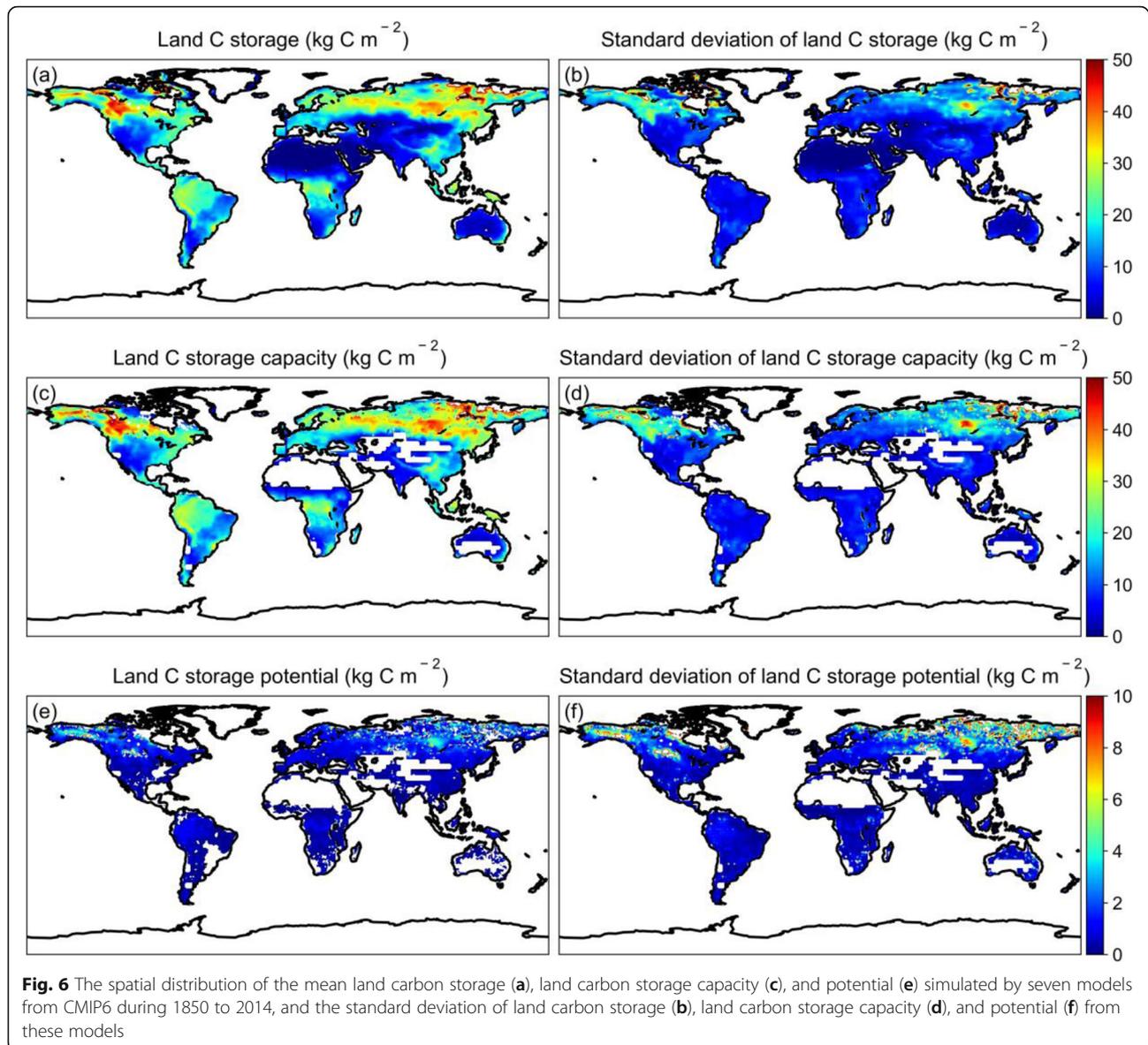
Potential applications of TraceME

An advantage from multi-model intercomparison projects (MIPs) is that model evaluation can provide a

multifaceted understanding of a given model by comparing its performance with its older versions or other models (Eyring et al. 2016b). Model evaluation needs to understand whether and how the fidelity of the models in simulating terrestrial C processes increases at different phases of MIP. For example, ESMValTool has been used to analyze whether the emergent constraints on equilibrium climate sensitivity in CMIP5 still hold for CMIP6 (Schlund et al. 2020). ILAMB has benchmarked and intercompared the terrestrial C cycle simulated by CMIP5 and CMIP6 models and presented the results in a detailed assessment report (<https://www.ilamb.org/CMIP5v6/historical/>). In our study, we analyzed the spatiotemporal changes in the uncertainty sources of simulated C storage in CMIP6 models at different periods using TraceME (Fig. 9). It also has the potential to research the terrestrial C cycle dynamic at the two phases of CMIP from a traceability perspective. Compared with other tools, it can diagnose whether the source of uncertainty simulated by CMIP6 models has shifted compared to CMIP5, and which processes cause the change. Furthermore, TraceME can provide detailed reports of traceability analysis on the performance of specific models in CMIP5 and CMIP6.

Global C cycle models have incorporated a broad set of terrestrial processes, such as human management and societal impacts (Fisher and Koven 2020). Model evaluation needs to comprehensively diagnose the effect of the new modules on the simulations of C cycle processes (Collier et al. 2018). TraceME has the advantage of traceability to measure which components of the C cycle can be affected by new processes represented in the model. For example, Du et al. (2018) has explored the effect of three different carbon-nitrogen coupling schemes on C storage capacity based on the framework of traceability analysis. Besides, some plant functional traits have been considered in models because of the robust relationship between traits (Wright et al. 2004; Fyllas et al. 2014; Sakschewski et al. 2015; Cui et al. 2020). A traceability framework has been used to analyze the uncertainty of simulated ecosystem productivity by linking different vegetation functional properties (Cui et al. 2019). Thus, TraceME can further update its traceable framework to evaluate the effect of some new processes on the performance of models.

Benchmarking analysis is an essential part of model evaluation. Some model evaluation systems (e.g., ILAMB and ESMValTool) have built large datasets of observation data as benchmarks to diagnose the performance of models (Eyring et al. 2020; Collier et al. 2018). The TraceME package can be applied together with those existing tools to offer additional diagnoses on model uncertainty. Recently, more and more observational products have been generated with the improvement of measurement means and algorithm technologies. For example, Wang et al. (2019) have

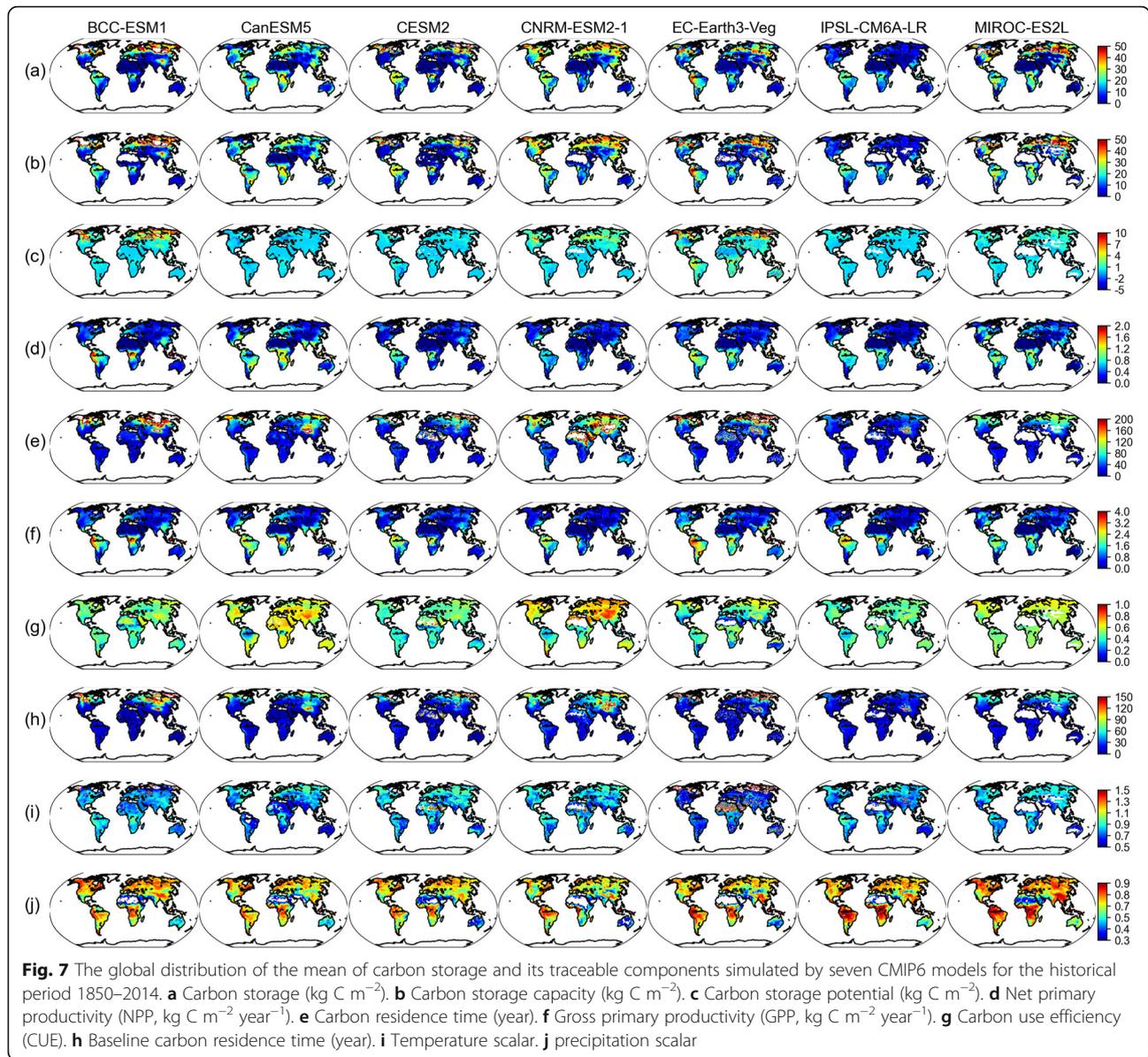


constructed a global soil C residence time database and used it to evaluate the simulated mean soil C transit times by Earth system models. Many new global datasets about other ecological processes based on both field measurements (Salunkhe et al. 2018; Li et al. 2019; Zheng et al. 2020; Zhu & Xia 2020; Ustin & Middleton 2021) and manipulative experiments (Song et al. 2019) are greatly valuable for model evaluation. These observational products make it possible for TraceME to develop datasets for evaluating those key processes which have not been incorporated in other tools.

Challenges and future developments of TraceME

Although TraceME (v1.0) provides a traceable and comprehensive system for evaluating global terrestrial C

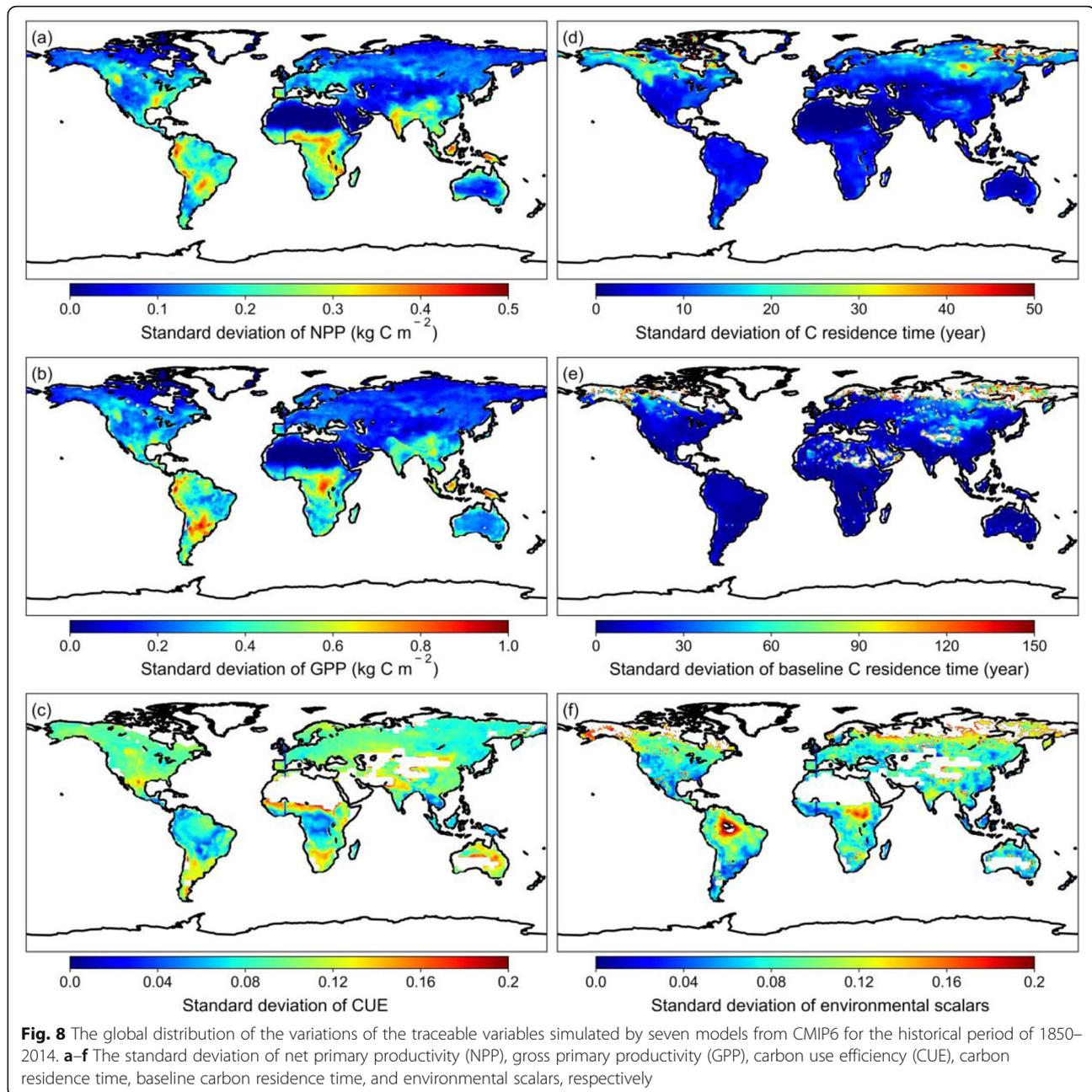
cycle models, some challenges remain in its future development. One challenge is the theoretical development of the traceability analysis in TraceME. The theoretical foundation of the traceability analysis is developed on the internal properties of the land C cycle, which can be described as a matrix equation (Xia et al. 2013; Luo et al. 2017). Some other terrestrial processes, such as nutrient cycles, hydrological processes, and energy fluxes, are difficult to be incorporated into the matrix equation (Wei et al. 2019). The second challenge is that it is difficult to obtain observational data for some traceable components in the framework of the traceability analysis, such as baseline C residence time. The third challenge is from the shortcomings of the cyberinfrastructure of the current TraceME. For



example, the efficiency of the evaluation process of TraceME significantly depends on the performance of the computer where the node of TraceME is located. Moreover, the installation of working nodes requires some specific environment settings in the operating system.

The development of TraceME is ongoing. Many efforts are being made to improve the framework of traceability analysis, to build up the observational datasets for benchmarking analysis, and to improve the infrastructure of the TraceME. In terms of developing the traceability analysis, some works can be considered. For example, recent studies have shown that GPP is jointly controlled by plant phenology and physiology, and it can be decomposed into the CO_2 uptake period (CUP) and the maximal GPP during the CUP that represents a property of plant canopy

physiology (Xia et al. 2015; Huang et al. 2018). Both of the phenological and physiological processes are influenced by environmental factors, such as temperature and water availability (Jaworski and Hilszczański 2013; Xie et al. 2015; Piao et al. 2019). Meanwhile, other environmental factors besides temperature and water, such as oxygen and nutrients availability, also affect C residence time (Tian et al. 1999; Wu et al. 2003; Melillo et al. 2011; Van Groenigen et al. 2014; Wieder et al. 2015). These traceable processes and factors still need to be added to the TraceME. On the other hand, the new advances in machine learning methods could be useful to produce datasets for some components in the framework of the traceability analysis. For example, Shi et al. (2020) has used the machine learning method to link the measurements of



radiocarbon with environmental factors to get the age distribution of global soil C. Finally, the infrastructure of TraceME is expected to evolve into a more open community for users and developers, so some aspects need to be further improved, such as version-control mechanism, intermediate analytical result, and encryption techniques (Xu et al. 2019). Developing an offline package is also one way to make TraceME more effective. Moreover, the databases in TraceME (v1.0) need to be updated in a timely and automated manner, especially since the amount of both observational and modeling data is increasing rapidly (Xia et al. 2020).

Conclusions

We developed an online tool for analyzing and evaluating the performance of CMIP6 models on the land C cycle using a traceability analysis (i.e., TraceME). TraceME can effectively diagnose the source of uncertainty of land C cycle models. As shown in this study, TraceME can accelerate the pace of model evaluation on land C cycle, and its evaluation results can be useful for specific models to further improve their representation of some ecological processes. Overall, new model evaluation tools like TraceME will provide new opportunities to understand the large uncertainty in the complex Earth system models.

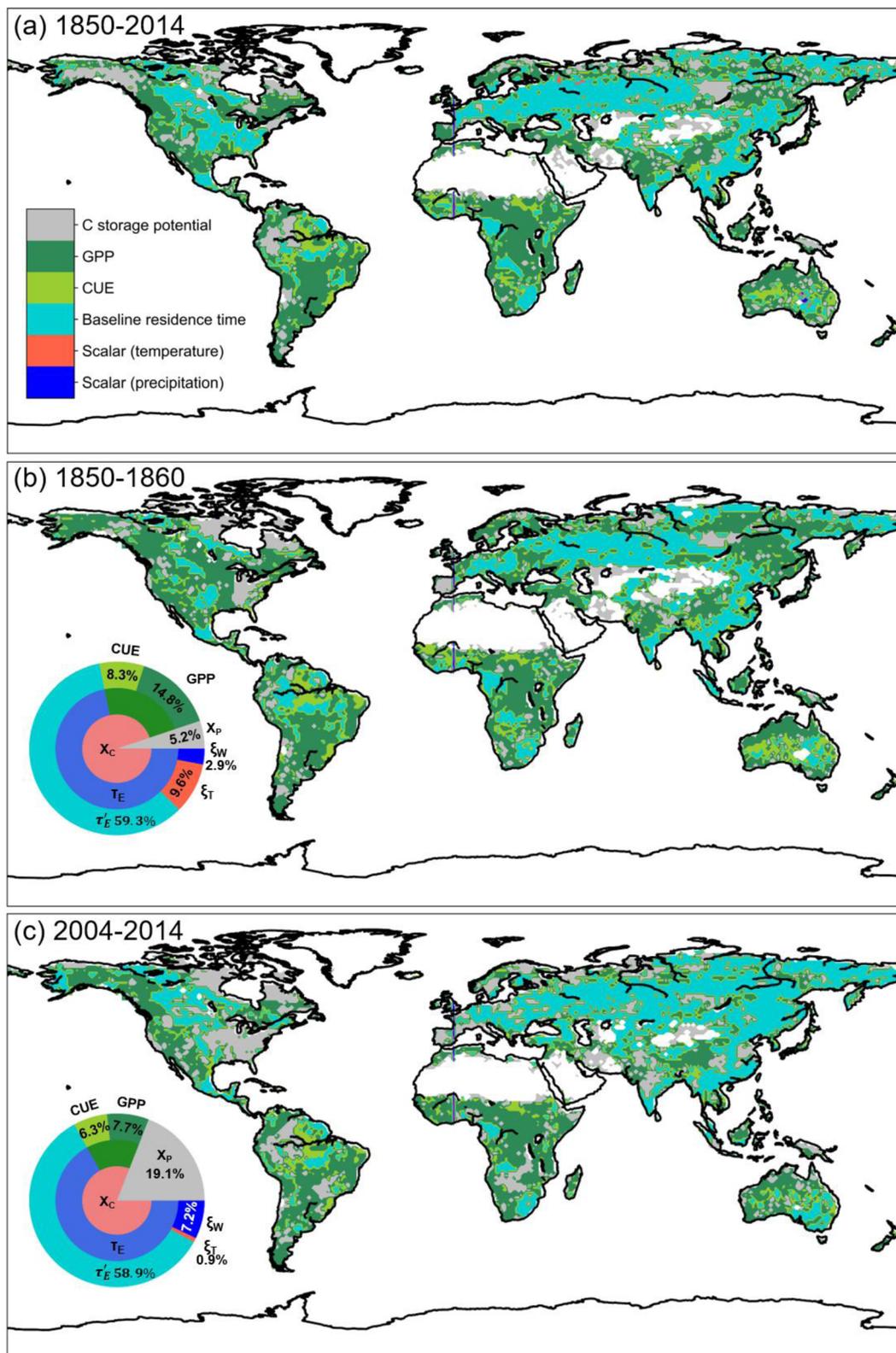


Fig. 9 The global distribution of the dominant variable for the variation in simulated land carbon storage by the models from CMIP6 at different periods: **a** 1850–2014, **b** 1850–1860, and **c** 2004–2014. The subplot of each panel is the variation decomposition of the carbon storage based on annual data

Abbreviations

TraceME: Traceability analysis system for Model Evaluation; C: Carbon; GPP: Gross primary production; CUE: Carbon use efficiency; CMIP: Coupled Model Intercomparison Project; NPP: Net primary productivity; ILAMB: International Land Model Benchmarking; ESMValTool: Earth System Model Evaluation Tool; LVT: Land surface Verification Toolkit; CAFE: Collaborative analysis framework for distributed gridded environmental data; NetCDF: Network Common Data Format; CF: Climate and Forecast

Acknowledgements

We acknowledge the World Climate Research Program (WCRP) that is responsible for CMIP, and we thank the modeling groups for providing their model output.

Authors' contributions

JX and JZ designed this study. JZ build the system of TraceME (v1.0). NW provided the support of some algorithms in the system. YB and YF provided the code and technical support of CAFE. JZ wrote the first draft, and all other authors contributed to the revision and discussion of the results. The author(s) read and approved the final manuscript.

Funding

This work was financially supported by the National Key R&D Program of China (2017YFA0604600) and National Natural Science Foundation of China (31722009).

Availability of data and materials

The dataset(s) supporting the conclusions of this article is(are) available in the ESGF repository, <https://esgf-node.llnl.gov/search/cmip6/>.

Project name: TraceME (v1.0); Project home page: <http://traceme.org.cn/>;

Archived version: <https://github.com/ECNU-RCGCEF/TraceME>; Operating system: Linux; Programming language: PHP, JAVA, Python. Other requirements: Java 1.8 or higher, MySQL, Tomcat 4.0 or higher;

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China. ²Research Center for Global Change and Ecological Forecasting, East China Normal University, Shanghai 200241, China. ³Ministry of Education Key Laboratory for Earth System Modelling, Department of Earth System Science, Tsinghua University, Beijing 100084, China. ⁴Joint Center for Global Change Studies (JCGCS), Beijing 100084, China. ⁵Center for Ecosystem Science and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, USA.

Received: 20 October 2020 Accepted: 13 January 2021

Published online: 29 January 2021

References

- Abramowitz G (2012) Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geosci Model Dev* 5:819–827
- Ahlström A, Schurgers G, Arneeth A, Smith B (2012) Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environ Res Lett* 7:044008
- Arora VK, Katavouta A, Williams RG, Jones CD, Brovkin V, Friedlingstein P, Schwinger J, Bopp L, Boucher O, Cadule P (2020) Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences* 17(16):4173–4222
- Bai Y, Di L (2012) Review of geospatial data systems' support of global change studies. *Br J Environ Clim Change* 2(4):421–436
- Bonan GB, Doney SC (2018) Climate, ecosystems, and planetary futures: the challenge to predict life in Earth system models. *Science* 359(6375): eaam8328
- Bonan GB, Lombardozzi DL, Wieder WR, Oleson KW, Lawrence DM, Hoffman FM, Collier NJGBC (2019) Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850–2014). *Global Biogeochem Cy* 33(10):1310–1326
- Chevan A, Sutherland M (1991) Hierarchical partitioning. *Am Stat* 45:90–96
- Collier N, Hoffman FM, Lawrence DM, Keppel-Aleks G, Koven CD, Riley WJ, Mu M, Randerson JT (2018) The International Land Model Benchmarking (ILAMB) system: design, theory, and implementation. *J Adv Model Earth Sy* 10:2731–2754
- Cui E, Huang K, Arain MA, Fisher JB, Huntzinger DN, Ito A, Luo Y, Jain AK, Mao J, Michalak AM, Niu S, Parazoo NC, Peng C, Peng S, Poulter B, Ricciuto DM, Schaefer KM, Schwalm CR, Shi X, Tian H, Wang W, Wang J, Wei Y, Yan E, Yan L, Zeng N, Zhu Q, Xia J (2019) Vegetation functional properties determine uncertainty of simulated ecosystem productivity: a traceability analysis in the East Asian Monsoon Region. *Global Biogeochem Cy* 33:668–689
- Cui E, Weng E, Yan E, Xia J (2020) Robust leaf trait relationships across species under global environmental changes. *Nat Commun* 11:2999
- DeLucia EH, Drake JE, Thomas RB, Gonzalez-Meler M (2007) Forest carbon use efficiency: is respiration a constant fraction of gross primary production? *Glob Change Biol* 13:1157–1167
- Du Z, Weng E, Jiang L, Luo Y, Xia J, Zhou X (2018) Carbon–nitrogen coupling under three schemes of model representation: a traceability analysis. *Geosci Model Dev* 11:4399–4416
- Eyring V, Bock L, Lauer A, Righi M, Schlund M, Andela B, Arnone E, Bellprat O, Brötz B, Caron L-P, Carvalho N, Cionni I, Cortesi N, Crezee B, Davin EL, Davini P, Debeire K, de Mora L, Deser C, Docquier D, Earnshaw P, Ehbrecht C, Gier BK, Gonzalez-Reviriego N, Goodman P, Hagemann S, Hardiman S, Hassler B, Hunter A, Kadow C, Kindermann S, Koirala S, Koldunov N, Lejeune Q, Lembo V, Lovato T, Lucarini V, Massonnet F, Müller B, Pandde A, Pérez-Zanón N, Phillips A, Predoi V, Russell J, Sellar A, Serva F, Stacke T, Swaminathan R, Torralba V, Vegas-Regidor J, von Hardenberg J, Weigel K, Zimmermann K (2020) Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geosci Model Dev* 13:3383–3438
- Eyring V, Cox PM, Flato GM, Gleckler PJ, Abramowitz G, Caldwell P, Collins WD, Gier BK, Hall AD, Hoffman FM, Hurtt GC, Jahn A, Jones CD, Klein SA, Krasting JP, Kwiatkowski L, Lorenz R, Maloney E, Meehl GA, Pendergrass AG, Pincus R, Ruane AC, Russell JL, Sanderson BM, Santer BD, Sherwood SC, Simpson IR, Stouffer RJ, Williamson MS (2019) Taking climate model evaluation to the next level. *Nat Clim Change* 9:102–110
- Eyring V, Gleckler PJ, Heinze C, Stouffer RJ, Taylor KE, Balaji V, Guilyardi E, Joussaume S, Kindermann S, Lawrence BN, Meehl GA, Righi M, Williams DN (2016a) Towards improved and more routine Earth system model evaluation in CMIP. *Earth Syst Dynam* 7:813–830
- Eyring V, Righi M, Lauer A, Evaldsson M, Wenzel S, Jones C, Anav A, Andrews O, Cionni I, Davin EL, Deser C, Ehbrecht C, Friedlingstein P, Gleckler P, Gottschaldt K-D, Hagemann S, Juckes M, Kindermann S, Krasting J, Kunert D, Levine R, Loew A, Mäkelä J, Martin G, Mason E, Phillips AS, Read S, Rio C, Roehrig R, Senftleben D, Sterl A, van Ulft LH, Walton J, Wang S, Williams KD (2016b) ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci Model Dev* 9:1747–1802
- Fisher RA, Koven CD (2020) Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *J Adv Model Earth Sy* 12(4):e2018MS001453
- Friedlingstein P, Cox P, Betts R, Bopp L, von Bloh W, Brovkin V, Cadule P, Doney S, Eby M, Fung I (2006) Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J Clim* 19:3337–3353
- Fyllas NM, Gloor E, Mercado L, Sitch S, Quesada CA, Domingues T, Galbraith D, Torre-Lezama A, Vilanova E, Ramirez-Angulo H (2014) Analysing Amazonian forest productivity using a new individual and trait-based model (TFS v. 1). *Geosci Model Dev* 7:1251–1269
- Hoffman FM, Koven CD, Keppel-Aleks G, Lawrence DM, Riley WJ, Randerson JT, Ahlstrom A, Abramowitz G, Baldocchi DD, Best MJ (2016) 2016 International Land Model Benchmarking (ILAMB) Workshop Report
- Hoffman FM, Randerson JT, Arora VK, Bao Q, Cadule P, Ji D, Jones CD, Kawamiya M, Samar K, Lindsay K, Obata A, Shevliakova E, Six KD, Tjiputra JF, Volodin EM, Wu T (2014) Causes and implications of persistent atmospheric carbon dioxide biases in Earth system models. *J Geophys Res-Bioge* 119:141–162

- Huang K, Xia J, Wang Y, Ahlstrom A, Chen J, Cook RB, Cui E, Fang Y, Fisher JB, Huntzinger DN, Li Z, Michalak AM, Qiao Y, Schaefer K, Schwalm C, Wang J, Wei Y, Xu X, Yan L, Bian C, Luo Y (2018) Enhanced peak growth of global vegetation and its key mechanisms. *Nat Ecol Evol* 2:1897–1905
- Huang Y, Stacy M, Jiang J, Sundi N, Ma S, Saruta V, Jung CG, Shi Z, Xia J, Hanson PJ, Ricciuto D, Luo Y (2019) Realized ecological forecast through an interactive Ecological Platform for Assimilating Data (EcoPAD, v1.0) into models. *Geosci Model Dev* 12:1119–1137
- Jaworski T, Hilszczański J (2013) The effect of temperature and humidity changes on insects development and their impact on forest ecosystems in the context of expected climate change. *For Res Pap* 74:345–355
- Jiang L, Shi Z, Xia J, Liang J, Lu X, Wang Y, Luo Y (2017) Transient traceability analysis of land carbon storage dynamics: procedures and its application to two forest ecosystems. *J Adv Model Earth Sy* 9:2822–2835
- Kumar SV, Peters-Lidard CD, Santanello J, Harrison K, Liu Y, Shaw M (2012) Land surface Verification Toolkit (LVT) – a generalized framework for land surface model evaluation. *Geosci Model Dev* 5:869–886
- Li S, Yuan W, Ciais P, Viovy N, Ito A, Jia B, Zhu D (2019) Benchmark estimates for aboveground litterfall data derived from ecosystem models. *Environ Res Lett* 14:084020
- Lovenduski NS, Bonan GB (2017) Reducing uncertainty in projections of terrestrial carbon uptake. *Environ Res Lett* 12(4):044020
- Luo Y, Keenan TF, Smith M (2015) Predictability of the terrestrial carbon cycle. *Glob Change Biol* 21(5):1737–1751
- Luo Y, Schuur EA (2020) Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. *Glob Change Biol* 26:1109–1117
- Luo Y, Shi Z, Lu X, Xia J, Liang J, Jiang J, Wang Y, Smith MJ, Jiang L, Ahlström A, Chen B, Hararuk O, Hastings A, Hoffman F, Medlyn B, Niu S, Rasmussen M, Todd-Brown K, Wang Y-P (2017) Transient dynamics of terrestrial carbon storage: mathematical foundation and its applications. *Biogeosciences* 14: 145–161
- Luo Y, Weng E (2011) Dynamic disequilibrium of the terrestrial carbon cycle under global change. *Trends Ecol Evol* 26:96–104
- Melillo JM, Butler S, Johnson J, Mohan J, Steudler P, Lux H, Burrows E, Bowles F, Smith R, Scott L (2011) Soil warming, carbon–nitrogen interactions, and forest carbon budgets. *P Natl Acad Sci USA* 108:9508–9512
- Murray K, Conner MM (2009) Methods to quantify variable importance: implications for the analysis of noisy ecological data. *Ecology* 90:348–355
- Overpeck JT, Meehl GA, Bony S, Easterling DR (2011) Climate data challenges in the 21st century. *Science* 331:700–702
- Piao S, Liu Q, Chen A, Janssens IA, Fu Y, Dai J, Liu L, Lian X, Shen M, Zhu X (2019) Plant phenology and global climate change: current progresses and challenges. *Glob Change Biol* 25:1922–1940
- Rafique R, Xia J, Hararuk O, Leng G, Asrar G, Luo Y (2017) Comparing the performance of three land models in global C cycle simulations: a detailed structural analysis. *Land Degrad Dev* 28:524–533
- Sakschewski B, von Bloh W, Boit A, Rammig A, Kattge J, Poorter L, Peñuelas J, Thonicke K (2015) Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. *Glob Change Biol* 21:2711–2725
- Salunkhe O, Khare PK, Kumari R, Khan ML (2018) A systematic review on the aboveground biomass and carbon stocks of Indian forest ecosystems. *Ecol Process* 7:17
- Schlund M, Lauer A, Gentine P, Sherwood SC, Eyring V (2020) Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6? *Earth Syst Dyn* 11(4):1233–1258
- Shi Z, Allison SD, He Y, Levine PA, Hoyt AM, Beem-Miller J, Zhu Q, Wieder WR, Trumbore S, Randerson JT (2020) The age distribution of global soil carbon inferred from radiocarbon measurements. *Nat Geosci* 13(8):555–559
- Song J, Wan S, Piao S, Knapp AK, Classen AT, Vicca S, Ciais P, Hovenden MJ, Leuzinger S, Beier C, Kardol P, Xia J, Liu Q, Ru J, Zhou Z, Luo Y, Guo D, Adam Langley J, Zscheischler J, Dukes JS, Tang J, Chen J, Hofmockel KS, Kueppers LM, Rustad L, Liu L, Smith MD, Templer PH, Quinn Thomas R, Norby RJ, Phillips RP, Niu S, Faticchi S, Wang Y, Shao P, Han H, Wang D, Lei L, Wang J, Li X, Zhang Q, Li X, Su F, Liu B, Yang F, Ma G, Li G, Liu Y, Liu Y, Yang Z, Zhang K, Miao Y, Hu M, Yan C, Zhang A, Zhong M, Hui Y, Li Y, Zheng M (2019) A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change. *Nat Ecol Evol* 3:1309–1320
- Stockhause M, Lautenschlager M (2017) CMIP6 data citation of evolving data. *Data Sci J* 16:30
- Tian H, Melillo J, Kicklighter D, McGuire A, Helfrich J (1999) The sensitivity of terrestrial carbon storage to historical climate variability and atmospheric CO₂ in the United States. *Tellus B* 51:414–452
- Ustin SL, Middleton EM (2021) Current and near-term advances in Earth observation for ecological applications. *Ecol Process* 10:1
- Van Groenigen KJ, Qi X, Osenberg CW, Luo Y, Hungate BA (2014) Faster decomposition under increased atmospheric CO₂ limits soil carbon storage. *Science* 344:508–509
- Wang J, Xia J, Zhou X, Huang K, Zhou J, Huang Y, Jiang L, Xu X, Liang J, Wang Y-P, Cheng X, Luo Y (2019) Evaluating the simulated mean soil carbon transit times by Earth system models using observations. *Biogeosciences* 16:917–926
- Wei N, Cui E, Huang K, Du Z, Xu X, Wang J, Yan L, Xia J (2019) Decadal stabilization of soil inorganic nitrogen as a benchmark for global land models. *J Adv Model Earth Syst* 11:1088–1099
- Wieder WR, Cleveland CC, Smith WK, Todd-Brown KJNG (2015) Future productivity and carbon storage limited by terrestrial nutrient availability. *Nat Geosci* 8:441–444
- Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JH, Diemer M (2004) The worldwide leaf economics spectrum. *Nature* 428:821–827
- Wu H, Guo Z, Peng C (2003) Land use induced changes of organic carbon storage in soils of China. *Glob Change Biol* 9:305–315
- Xia J, Luo Y, Wang YP, Hararuk O (2013) Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Glob Change Biol* 19: 2104–2116
- Xia J, McGuire AD, Lawrence D, Burke E, Chen G, Chen X, Delire C, Koven C, MacDougall A, Peng S, Rinke A, Saito K, Zhang W, Alkama R, Bohn TJ, Ciais P, Decharme B, Gouttevin I, Hajima T, Hayes DJ, Huang K, Ji D, Krinner G, Lettenmaier DP, Miller PA, Moore JC, Smith B, Sueyoshi T, Shi Z, Yan L, Liang J, Jiang L, Zhang Q, Luo Y (2017) Terrestrial ecosystem model performance in simulating productivity and its vulnerability to climate change in the northern permafrost region. *J Geophys Res-Bioge* 122:430–446
- Xia J, Niu S, Ciais P, Janssens IA, Chen J, Ammann C, Arain A, Blanken PD, Cescatti A, Bonal D, Buchmann N, Curtis PS, Chen S, Dong J, Flanagan LB, Frankenberg C, Georgiadis T, Gough CM, Hui D, Kiely G, Li J, Lund M, Magliulo V, Marcolla B, Merbold L, Montagnani L, Moors EJ, Olesen JE, Piao S, Raschi A, Rouspard O, Suyker AE, Urbaniak M, Vaccari FP, Varlagin A, Vesala T, Wilkinson M, Weng E, Wohlfahrt G, Yan L, Luo Y (2015) Joint control of terrestrial gross primary productivity by plant phenology and physiology. *P Natl Acad Sci USA* 112:2788–2793
- Xia J, Wang J, Niu S (2020) Research challenges and opportunities for using big data in global change biology. *Glob Change Biol* 26:6040–6061
- Xia JY, Luo YQ, Wang YP, Weng ES, Hararuk O (2012) A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state. *Geosci Model Dev* 5:1259–1271
- Xie Y, Wang X, Silander JA Jr (2015) Deciduous forest responses to temperature, precipitation, and drought imply complex climate change impacts. *P Natl Acad Sci USA* 112:13585–13590
- Xu H, Li S, Bai Y, Dong W, Huang W, Xu S, Lin Y, Wang B, Wu F, Xin X (2019) A collaborative analysis framework for distributed gridded environmental data. *Environ Model Softw* 111:324–339
- Zarakas CM, Swann AL, Laguë MM, Armour KC, Randerson JT (2020) Plant physiology increases the magnitude and spread of the transient climate response to CO₂ in CMIP6 Earth System models. *J Clim* 33(19):8561–8578
- Zheng Y, Shen R, Wang Y, Li X, Liu S, Liang S, Chen JM, Ju W, Zhang L, Yuan W (2020) Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017. *Earth Syst Sci Data* 12:2725–2746
- Zhou S, Liang J, Lu X, Li Q, Jiang L, Zhang Y, Schwalm CR, Fisher JB, Tjiputra J, Stith S, Ahlström A, Huntzinger DN, Huang Y, Wang G, Luo Y (2018) Sources of uncertainty in modeled land carbon storage within and across three mips: diagnosis with three new techniques. *J Clim* 31:2833–2851
- Zhu C, Xia J (2020) Nonlinear increase of vegetation carbon storage in aging forests and its implications for Earth system models. *J Adv Model Earth Syst* 12:e2020MS002304

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.