Title
The Terrestrial Biosphere Model Farm

Permalink
https://escholarship.org/uc/item/3mc0d11t

Authors
Fisher, Joshua B
Sikka, Munish
Block, Gary L
et al.

Peer reviewed

Author Contributions:

Conceptualization: Joshua B. Fisher
Data curation: Munish Sikka, Yaxing Wei
Formal analysis: Joshua B. Fisher, Christopher R. Schwalm, Nicholas C. Parazoo
Funding acquisition: Joshua B. Fisher, Deborah N. Huntzinger
Investigation: Joshua B. Fisher, Nicholas C. Parazoo, Hannah R. Kolus, Malen Sok, Audrey Wang, Anna Gagne-Landmann, Shakirudeen Lawal, Alexandre Guillaume, Alyssa Poletti, Kevin M. Schaefer, Bassil Masri, Deborah N. Huntzinger
Methodology: Joshua B. Fisher, Gary L. Block, Anna Gagne-Landmann
Project Administration: Joshua B. Fisher

The Terrestrial Biosphere Model Farm

Joshua B. Fisher[1,2] ⓘ, Munish Sikka[1], Gary L. Block[1], Christopher R. Schwalm[3], Nicholas C. Parazoo[1] ⓘ, Hannah R. Kolus[1] ⓘ, Malen Sok[1], Audrey Wang[1], Anna Gagne-Landmann[1], Shakirudeen Lawal[1], Alexandre Guillaume[1], Alyssa Poletti[1] ⓘ, Kevin M. Schaefer[4] ⓘ, Bassil El Masri[5] ⓘ, Peter E. Levy[6] ⓘ, Yaxing Wei[7] ⓘ, Michael C. Dietze[8], and Deborah N. Huntzinger[9] ⓘ

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [2]Schmid College of Science and Technology, Chapman University, Orange, CA, USA, [3]Woodwell Climate Research Center, Falmouth, MA, USA, [4]National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [5]Department of Earth and Environmental Sciences, Murray State University, Murray, KY, USA, [6]Centre for Ecology and Hydrology, Penicuik, UK, [7]Environmental Sciences Division, Oak Ridge National Laboratory, Climate Change Science Institute, Oak Ridge, TN, USA, [8]Department of Earth and Environment, Boston University, Boston, MA, USA, [9]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ, USA

Abstract Model Intercomparison Projects (MIPs) are fundamental to our understanding of how the land surface responds to changes in climate. However, MIPs are challenging to conduct, requiring the organization of multiple, decentralized modeling teams throughout the world running common protocols. We explored centralizing these models on a single supercomputing system. We ran nine offline terrestrial biosphere models through the Terrestrial Biosphere Model Farm: CABLE, CENTURY, HyLand, ISAM, JULES, LPJ-GUESS, ORCHIDEE, SiB-3, and SiB-CASA. All models were wrapped in a software framework driven with common forcing data, spin-up, and run protocols specified by the Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) for years 1901–2100. We ran more than a dozen model experiments. We identify three major benefits and three major challenges. The benefits include: (a) processing multiple models through a MIP is relatively straightforward, (b) MIP protocols are run consistently across models, which may reduce some model output variability, and (c) unique multimodel experiments can provide novel output for analysis. The challenges are: (a) technological demand is large, particularly for data and output storage and transfer; (b) model versions lag those from the core model development teams; and (c) there is still a need for intellectual input from the core model development teams for insight into model results. A merger with the open-source, cloud-based Predictive Ecosystem Analyzer (PEcAn) ecoinformatics system may be a path forward to overcoming these challenges.

Plain Language Summary Comparing models is fundamental to our understanding of how the land surface responds to changes in climate. However, these comparisons are challenging to conduct, requiring the organization of multiple, decentralized teams throughout the world. We explored centralizing these models on a single supercomputing system. The models were all run the same way. We ran more than a dozen model experiments. We identify three major benefits and three major challenges. The benefits include: (a) the centralized system takes a lot of burden off individual teams; (b) running models the same way helps to identify differences in how the world is represented in the models; and (c) the system allows us to run many model experiments relatively quickly. The challenges are: (a) lots of models require lots of data storage and transfer needs; (b) model versions lag those from the core model development teams; and (c) there is still a need for intellectual input from the core model development teams for insight into model results. Another system, called PEcAn, which has a lot of tools that can help overcome these challenges, can potentially be used in future work.

1. Introduction

Models are among our best integrators of knowledge and understanding of how the world works. Terrestrial biosphere models (TBMs) encapsulate a myriad of land surface and ecosystem processes, and how those respond to one another and external forcings (Fisher et al., 2014). But, knowledge is biased and socially constructed (Berger & Luckmann, 1966; Nietzsche, 1977); consequently, we too may expect our models to be locally biased to the construction of knowledge that created them. How photosynthesis is understood to someone in the US may be

Resources: Joshua B. Fisher, Kevin M. Schaefer, Bassil Masri, Peter E. Levy, Deborah N. Huntzinger
Software: Munish Sikka, Gary L. Block, Nicholas C. Parazoo, Hannah R. Kolus, Malen Sok, Audrey Wang, Anna Gagne-Landmann, Shakirudeen Lawal, Alexandre Guillaume, Alyssa Poletti, Bassil Masri
Supervision: Joshua B. Fisher
Visualization: Joshua B. Fisher
Writing – original draft: Joshua B. Fisher
Writing – review & editing: Joshua B. Fisher, Christopher R. Schwalm, Nicholas C. Parazoo, Hannah R. Kolus, Malen Sok, Audrey Wang, Anna Gagne-Landmann, Shakirudeen Lawal, Alexandre Guillaume, Kevin M. Schaefer, Bassil Masri, Peter E. Levy, Yaxing Wei, Michael C. Dietze, Deborah N. Huntzinger

different than that to someone in Brazil or Australia or Japan (Rogers et al., 2017). Still, global models must be constructed for global science, though their construction is predominantly local (Figure 1).

The great "elephant in the room," so to speak, is not what is behind us, but what is ahead. Projections of the future hold no truth in the present, yet are best read by how well we see in the past. Like the fable of the blind men and the elephant, we want to know what is in front of us—the future of the Earth—that which we cannot see. Each model and modeler, like the blind men, is biased to their local knowledge, each saying something different, yet similar. These model differences define the uncertainty of our future (IPCC, 2014). By working together, we may piece together a more accurate picture than in isolation.

Model Intercomparison Projects (MIPs) are that "working together." MIPs are designed to hear different perspectives to sample the population of possible ways the world works. In theory, if that sample is sufficiently robust, then one might converge toward a central tendency of likely truth. Ensemble model means typically perform better than most individual models (Schwalm et al., 2015). The distribution of multimodel variability could be used to construct "emergent constraints" for benchmarks to base future projections (Bretherton & Caldwell, 2020; Brient, 2020; Cox et al., 2018; Eyring et al., 2019; Hall et al., 2019; Nijsse et al., 2020; Tokarska et al., 2020). In practice, however, that sampling is often not robust and likely skewed toward models that are derivatives of one another, thereby compounding local biases (Caldwell et al., 2018; Huntzinger et al., 2013, 2017; Sanderson et al., 2015, 2021). Still, MIPs are a necessary starting point for a conversation; otherwise, it is more like a lot of voices shouting in dissonance.

Due to numerous user decisions, the same model code run 10 times can have 10 different outputs. If our interest is in understanding variability in biophysical processes (Bonan & Doney, 2018), then this is a problem, as there should, in theory, be no process representation variability for a single model (with notable exceptions, e.g., Lawrence et al., 2019; Niu et al., 2011). What could potentially cause different outputs from the same model and can we control for it? Starting with the starting point or spin-up—a model should move forward from the same initial conditions every time, but this can easily vary depending on how one wants to create the world (Huntzinger et al., 2020; Shi et al., 2013). The forcing data used could vary, for example, from one meteorological data set to another, even if the model stays the same (Badgley et al., 2015; Lawrence et al., 2019). How a model/modeler uses the same forcing data could vary, for example, mapping of model plant functional types (PFTs) onto a land cover/land use change (LCLUC) data set (e.g., "mixed forest," different crops). A model might have functions that can be turned on or off and may be reliant on availability of certain forcing data (e.g., nitrogen cycle). Finally, the output may all be identical, but the interpretation of the output can vary from person to person or be altered in transfer or analysis.

Now, compound this all with multiple different models, and we can see how comparing models for variability in process representation can quickly become muddled by other factors that can influence model output. MIPs attempt to control for many of these factors, but struggle to ensure that their protocols are rigorously met. For TBMs alone, there have been a lot of MIPs (global, regional, site; offline, coupled), and there will be more—e.g., PILPS
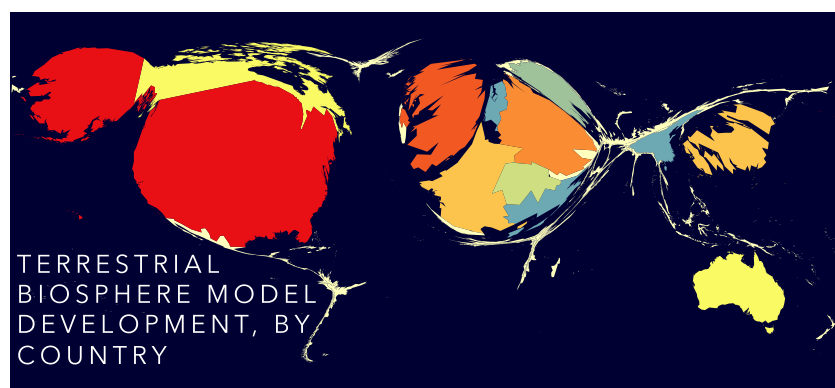
Figure 1. Most global terrestrial biosphere models (TBMs) are developed primarily in the US, Europe, Australia, and Japan. This cartogram inflates/deflates the size of the country based on the number of TBMs developed, $n = 58$. From Fisher et al. (2014).

(Henderson-Sellers et al., 1993), VEMAP (VEMAP, 1995), CCMLP (Kicklighter et al., 1999), NPP MIP (Cramer et al., 1999), DGVM MIP (Cramer et al., 2001), C4MIP (Friedlingstein et al., 2006), AMMA (Redelsperger et al., 2006), GSWP (Dirmeyer et al., 2006), WETCHIMP (Melton et al., 2013), LBA-DMIP (de Gonçalves et al., 2013), NACP Interim Site and Regional Syntheses (Huntzinger et al., 2012; Schwalm et al., 2010), TRENDY (Sitch et al., 2013), MsTMIP (Huntzinger et al., 2013), ISI-MIP (Warszawski et al., 2014), FACE MDS (Medlyn et al., 2015), CMIP (Arora et al., 2020), and PalEON (Rollinson et al., 2021) (see Table 1 for acronyms and abbreviations). MIPs typically specify common forcing data, but some models need the data transformed (e.g., temporally) or require additional individualized data sets. Some MIPs specify spin-up criteria, but those are not necessarily applicable across all models (Huntzinger et al., 2013). Finally, a model might adhere strictly to the MIP protocol, but some element of tuning to common benchmark data occurs such that inter-model output converges, despite differences in processes (Dommenget & Rezny, 2018; Fer et al., 2021; Hourdin et al., 2017; Notz, 2015; Raäisäänen, 2007; Scheiter et al., 2013).

As a means to address some of these outstanding issues in MIPs, we sought to develop a prototype system whereby models are run with common operating software on the same hardware systems (Fer et al., 2021; Kumar et al., 2006). This approach is similar to multimodel systems, such as the Land Information System (LIS; Kumar et al., 2006), the Terrestrial Observation and Prediction System (TOPS; Nemani et al., 2002), and the Predictive Ecosystem Analyzer (PEcAn; Dietze et al., 2013). Such a system ensures that spin-up protocols are run consistently across models, user decisions on forcing data such as LCLUC-to-PFT mapping are systematic, there is no tuning, and that output is saved, transferred, and analyzed uniformly. Further, logistical coordination among multiple teams can be streamlined. We focused on global TBMs and used the relatively strict MsTMIP protocol

Table 1
Acronyms and Abbreviations

Acronym/Abbreviation	Definition
AMMA	African Monsoon Multidisciplinary Analysis
API	Application Programming Interface
C4MIP	Coupled Climate Carbon Cycle Model Intercomparison Project
CABLE	CSIRO Atmosphere Biosphere Land Exchange
CAM	Community Atmosphere Model
CASA	Carnegie-Ames-Stanford Approach
CCMLP	Carbon-Cycle Model Linkage Project
CESM	Community Earth System Model
CLASS-CTEM-N	Canadian LAnd Surface Scheme—Canadian Terrestrial Ecosystem Model—Nitrogen
CLM	Community Land Model
CMIP	Coupled Model Intercomparison Project
CMS	Carbon Monitoring System
DGVM	Dynamic Global Vegetation Model
DLEM	Dynamic Land Ecosystem Model
ESM	Earth System Model
FACE MDS	Free Air CO_2 Enrichment Model Data Synthesis
GFDL-CM3	Geophysical Fluid Dynamics Laboratory—Coupled Model
GSWP	Global Soil Wetness Project
GTEC	Global Terrestrial Ecosystem Carbon
HadGEM2-AO	Hadley Center Global Environment Model version 2—Atmosphere-Ocean
I/O	Input/Output
IPSL-CM5A-MR	Institut Pierre Simon Laplace—Coupled Model—Mid-Resolution
ISAM	Integrated Science Assessment Model
ISI-MIP	Inter-Sectoral Impact—Model Intercomparison Project
JPL	Jet Propulsion Laboratory

Table 1
Continued

Acronym/Abbreviation	Definition
JULES	Joint UK Land Environment Simulator
LBA-DMIP	Large-Scale Biosphere-Atmosphere—Data Model Intercomparison Project
LCLUC	Land Cover Land Use Change
LPJ-GUESS	Lund-Potsdam-Jena—General Ecosystem Simulator
LPJ-wsl	Lund-Potsdam-Jena—Wald Schnee und Landschaft
LSCE	Laboratoire des Sciences du Climat et de l'Environnement
LUH-SYNMAP	Land Use Harmonization—Synergetic Land Cover
MIP	Model Intercomparison Project
MPI-ESM-MR	Max Planck Institute—Earth System Model—Mixed Resolution
MsTMIP	Multi-scale Synthesis and Terrestrial Model Intercomparison Project
NACP	North American Carbon Program
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NPP	Net Primary Productivity
ORCHIDEE	ORganizing Carbon and Hydrology In Dynamic Ecosystems Environment
PalEON	Paleo-Ecological Observatory Network
PEcAn	Predictive Ecosystem Analyzer
PFT	Plant Functional Type
PILPS	Project for Intercomparison of Land Surface Parameterization Schemes
RCP	Representative Concentration Pathways
SG3	Simulation Global 3
SiB	Simple Biosphere
TBM	Terrestrial Biosphere Model
TEM	Terrestrial Ecosystem Model
VEGAS	VEgetation Global Atmosphere & Soil
VEMAP	Vegetation-Ecosystem Modeling & Analysis Project
VIC	Variable Infiltration Capacity
VISIT	Vegetation Integrative SImulator for Trace gases
WETCHIMP	Wetland CH4 Intercomparison of Models Project

for spin-up and factorial model experiments from 1901 to 2010 (Phase I) and 2011 to 2100 (Phase II; Huntzinger et al., 2013). Overcoming two major challenges was required for this to be successful: (a) development of a software system that could interface consistently across disparate models as well as supercomputing infrastructure and (b) compilation and setup of multiple models. Here, we describe this system, called the Terrestrial Biosphere Model Farm, named with tribute to a similar initial effort by others (Denning et al., 2009). The Model Farm conceptually builds on efforts such as LIS with hydrological strengths, but here with more of an emphasis on terrestrial carbon cycling. We show an example output for illustration, but the focus of this paper is descriptive rather than output analytical. Finally, we provide experiential lessons-learned and recommended paths forward.

2. Methods

2.1. Data and Models

We primarily used the MsTMIP data and simulation protocols (Huntzinger et al., 2013; Wei et al., 2014). Phase I consisted of simulations from years 1801 to 2010 and Phase II from 2011 to 2100, all globally at $0.5° \times 0.5°$ resolution. Phase I forcing data included climate, LCLUC, atmospheric CO_2, and nitrogen deposition. These

drivers were turned on/off (i.e., variable or constant) semi-factorially such that the influence of each driver can be isolated in analysis (Huntzinger et al., 2013). Spin-up to initial conditions was dictated by steady-state criteria (Huntzinger et al., 2013). Climate variables were given at a 6-hourly temporal resolution derived from a CRU-NCEP merged product (Harris et al., 2014; Kalnay et al., 1996). LCLUC was derived from a merged LUH-SYNMAP product, resulting in 47 classes annually (Hurtt et al., 2011; Jung et al., 2006). Monthly atmospheric CO_2 was constructed for consistency with GLOBALVIEW among other databases (Wei et al., 2014). Finally, annual nitrogen deposition was constructed based on Dentener et al. (2006).

MsTMIP Phase II simulations were forced with climate and CO_2 projections from five CMIP5 Earth System Models (ESMs) and two Representative Concentration Pathways (RCPs). The ESMs included CESM1-CAM5 (Meehl et al., 2013), GFDL-CM3 (Griffies et al., 2011), HadGEM2-AO (Bellouin et al., 2011), IPSL-CM5A-MR (Dufresne et al., 2013), and MPI-ESM-MR (Giorgetta et al., 2013). The ESMs span the full range of model climate sensitivity between low- and high-temperature response. The RCPs included business-as-usual (RCP8.5) and medium mitigation scenarios (RCP4.5) (Van Vuuren et al., 2011). LCLUC was held constant, and nitrogen deposition was not used in our Phase II simulations, though they are available as time-varying to 2100.

The TBMs used in the Farm included CABLE (Wang et al., 2010), CENTURY (Parton et al., 1988), HyLand (Levy et al., 2004), ISAM (El-Masri et al., 2013), JULES (Clark et al., 2011), LPJ-GUESS (Smith et al., 2001), ORCHIDEE (Krinner et al., 2005), SiB-3 (Baker et al., 2008), and SiB-CASA (Schaefer et al., 2008). We also used the output from the MsTMIP version 1 models for comparison (Huntzinger et al., 2013), which included: CLASS-CTEM-N (S. Huang et al., 2011), CLM4 (Bonan et al., 1995), CLM4VIC (Lei et al., 2014), DLEM (Tian et al., 2012), GTEC (Ricciuto et al., 2011), ISAM (Jain & Yang, 2005), LPJ-wsl (Sitch et al., 2003), OR-CHIDEE-LSCE (Krinner et al., 2005), SiB-CASA (Schaefer et al., 2008), TEM6 (Hayes et al., 2011), VEGAS2.1 (Zeng et al., 2005), and VISIT (Ito, 2010). We do not describe the models in detail here, instead referring the reader to the respective references. Different models were run for different combinations of simulations and model experiments with all completing spin-up, some completing some or all of the MsTMIP Phase I experiments, and some completing all of the MsTMIP Phase I and II experiments (see Section 3), depending on application (see, e.g., Discussion).
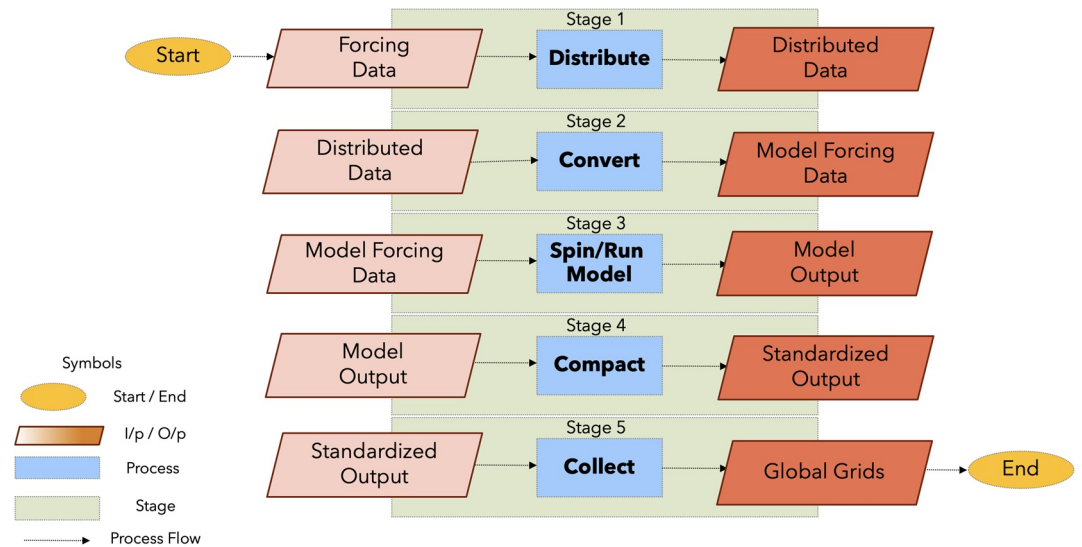
2.2. Model Farm

The Terrestrial Biosphere Model Farm was designed to keep individual TBM codes intact and treated as modular or compartmentalized, which can also facilitate the replacement of models by new versions. Each TBM is wrapped in the Farm software infrastructure, interfacing with the original model codes without modification, enabling different forcing data sets to be used, and unifying model output into defined formats.

The Terrestrial Biosphere Model Farm consists of five stages from distribution of forcing data to collection of model output (Figure 2): (a) Data Distribution, (b) Data Conversion, (c) Spin/Run Model, (d) Compact Output, and (e) Collect Output. Finally, a Diagnostics package enables automatic and relatively immediate visualization of model output to check for problems and/or success. The Model Farm was written in a combination of C++, Fortran, Matlab, and R for different functions. The Model Farm migrated over three different NASA JPL supercomputers (Zodiac, Halo, and Gattaca) during the course of development due to infrastructure upgrades (Table 2).

2.2.1. Stage 1: Data Distribution

Data Distribution restructures all input files into the supercomputing environment. This is the most time-consuming of all the stages, but is done only once per forcing data set, and used for all models. Here, the global forcing data are read, split into "patches" (i.e., groups) of cells, and distributed onto supercomputing nodes for parallel processing such that each patch has an approximately similar computational load. This split can be done because there is no lateral/horizontal cell-to-cell transport (e.g., water, carbon, and energy) for these models. Some patches may contain more grid cells than others due to empty ocean cells (i.e., land-only data). The global grid from MsTMIP was a 1-degree spatial resolution at 181×360 latitude and longitude or 65,160 cells (Wei et al., 2014). We distributed these data into 2,000 patches, each with ~32 cells (or "points") per patch, on 2,000 matching supercomputing processors/nodes based on memory and storage capacities of the nodes (with modifications as we migrated supercomputers). An index file assigns a patch and point number for each grid cell along with a time dimension, also forcing each model to operate with the same spatial resolution. This index file may be matched

Terrestrial Biosphere Model Farm: Conceptual Flowchart

Figure 2. Terrestrial Biosphere Model Farm software flowchart consists of five stages. Stage 1: forcing data (e.g., from MsTMIP) are partitioned and distributed among supercomputer nodes. Stage 2: forcing data are converted to individual model requirements (format, temporal resolution, and units). Stage 3: models are spun up and run. Stage 4: model output is converted again (format, temporal resolution, and units) for standardization across models. Stage 5: model output is gathered off supercomputing nodes and recombined into global (or otherwise) grids.

to index files from new data sets that need to align in space and time, for example, connecting historical data (1901–2010; Phase I) to future projections (2011–2100; Phase II), which may have disparate grid sizes and north/south orientations. MD5 checksums ensure that no data are corrupted in transfer from one computer to another.

The time required to complete Data Distribution varies depending on hardware infrastructure. Initially, the Data Distribution stage took 3–4 months to complete, but eventually came down to 3–4 days to complete per data set as we migrated across supercomputers. The bottleneck in the first supercomputer, Zodiac, was due to file transfer Input/Output (I/O) capacity and file count capacity, which were later expanded in the subsequent supercomputers, Halo and Gattaca. While I/O was expanded sufficiently in Halo, our project storage was reduced (from 30 TB in Zodiac to 10 TB in Halo), which required additional time to manage. MsTMIP forcing data required >500 GB (split roughly equally between Phase I and II). Finally, in Gattaca, the I/O and storage were sufficient for efficient Data Distribution.

2.2.2. Stage 2: Data Conversion

Data Conversion restructures the input forcing data to meet the unique requirements of individual models. Stage 2 is applied to each model after the model is successfully compiled and tested with provided test data or single point data on the supercomputer. Stage 2 ensures that there are no changes to the model code within the Farm.

Table 2
High-Performance Computing Specifications for the Terrestrial Biosphere Model Farm Project Generally Improved as the Project Migrated Across NASA JPL Supercomputers From Zodiac to Halo to Gattaca

System name	Zodiac	Halo	Gattaca
Project block quota	30 TB	10 TB	15 TB/high-performance scratch filesystem[a]
Project file # (inode)	1 million	10 million	9 million (scratch_lg[b])|4 million (scratch_sm[b])
Interconnect (I/O)	Dual 40 Gb/s[b]	40 Gb/s[b]	100 Gb/s[b]
Total cores	1,920	1,328/1,040	4,608
Number of cores/node	12	16	48
Memory/node	24 (2 GB/core)	64 (4 GB/core)	384 (8 GB/core)
Local scratch	None	1 TB HDD/node	365 GB SSD/node|1.9 TB HDD/node

[a]$n = 4$. [b]InfiniBand.

Specifically, forcing data may be converted to different: (a) formats such as NetCDF 3 or 4, ascii, *.mat, etc.; (b) temporal resolutions (e.g., hourly and daily); and (c) units, depending on the model. LCLUC is assigned to model PFTs. Configuration parameters are generated for each model to specify which data sets to access. This control module can be updated to swap out different forcing data sets for different domains and experiments. Spin-up data may be reorganized for length of runs. Time required for data conversion was approximately 6 hr per data set transformation. Additional storage was required for the converted data (e.g., >2 TB for SiB3). After Stage 2 is complete, a model may be ready to spin-up and run various global simulations.

2.2.3. Stage 3: Spin and Run Model

Spin and Run Model runs the model to meet defined steady-state initial conditions and/or executes a desired transient model experiment. Here, independent instances of the model for each of the 2000 patches are scheduled for parallel execution. In theory, Stage 3 should run smoothly; in practice, however, several issues must be managed for complete success. In spin-up, individual patches may fail to meet spin-up requirements within the run time allotted and must be rerun (Shi et al., 2013). Incorrect or incongruent LCLUC assignments may cause a model to crash a patch due to unspecified model parameters for the assignment. Forcing data sets may not align with each other (e.g., LCLUC assignment, but no climate). An automated failed-patch detector was written to rerun patches with longer run times to handle some spin-up issues. Other failed patches had to be identified visually, for example, noticeable as stripes or unexpectedly high/low values in a global map or as spikes in diagnostic zonal/latitudinal line plots. These patches were then gathered by the Farm operator for rerun and/or debugging. For efficiency, the Model Farm framework allowed for individual patches to be rerun without requiring the entire global data set to be rerun. Spin-up required 1–2 days for most patches to complete, depending on model and supercomputer. Transient run time also required 1–2 days, depending on model, supercomputer, and simulation (e.g., 100 years historical-only vs. 200 years historical plus future).

2.2.4. Stage 4: Compact Output

Compact Output converts output from individual models into common formats and units for analysis. Model output from Stage 3 is produced in individual model-dependent formats, units, variables, and time steps. Stage 4 unifies model output across models. Here, we convert all model outputs into NetCDF, common units (e.g., as defined by MsTMIP), standardized NaN/water/ice values, and monthly time steps. The monthly time step conversion reduces storage and transfer time demand, as most models operate and output at sub-monthly time steps (e.g., daily for HyLand and 6-hourly for CABLE), and which gives Stage 4 its name. For example, pre-Stage 4, HyLand used 1 TB per simulation; whereas, CABLE used nearly 4 TB per simulation. Monthly compaction requires approximately half a day (e.g., HyLand) to a day (e.g., CABLE) per simulation per model depending on the Stage 3 output. Converting output file formats (e.g., *.out files from CENTURY to NetCDF) requires approximately another day.

2.2.5. Stage 5: Collect Output

Collect Output converts all the individual patches into a common global grid. The final stage in the Model Farm requires collecting all the outputs, which are still partitioned into patches on individual supercomputer nodes and pieces the patches back together into a cohesive and standardized global grid (still at the monthly time step from Stage 4). Metadata are added at this stage, including, for example, model source code, model version, and contact information. Model output files are organized as one variable per file per model. Stage 5 collection requires less than half a day per simulation, depending on the number of variables saved.

3. Results

The Model Farm includes two visual diagnostic packages: the Single Variable package and the Experiment package. The Single Variable package is automatically run after collecting the model output from spin-up or a transient run. The user repeats the Experiment package as needed. These packages allow the user to quickly evaluate the model output and identify any problems. The Single Variable package produces a single PDF page per variable per model for all output variables (Figure 3). Each page includes six panels, including three maps and three line plots: (a) map of average variable value over simulation time span, (b) map of average variable value over simulation time span from comparison models, (c) difference map between (a) and (b), (d) annual time series with comparison models, (e) average monthly values with comparison models, and (f) average zonal/latitudinal
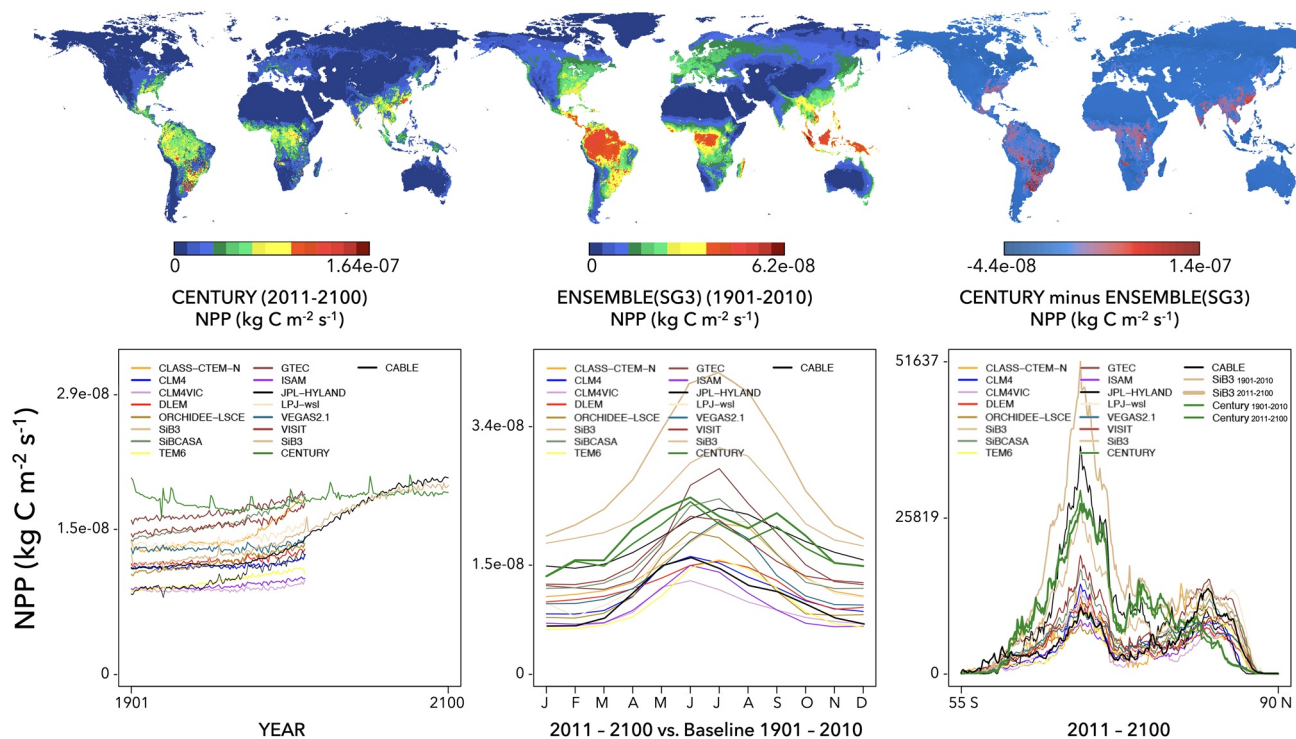
Figure 3. An automated diagnostic visualization PDF is generated from the Terrestrial Biosphere Model Farm. (top-left) Map of average variable value over simulation time span; (top-middle) comparison map of average variable value; (top-right) difference map between 1 and 2; (bottom-left) annual time series along with comparison models; (bottom-middle) average monthly values with comparison models; and (bottom-right) average zonal/latitudinal values with comparison models. Here, we show Net Primary Productivity (NPP) with CENTURY as the highlighted model of interest for the MsTMIP Phase II simulation (2011–2100) with CESM1-CAM5 climate and RCP4.5 CO_2 scenario. Comparison models are from MsTMIP Phase I SG3 simulation. Four additional models from the Farm are shown for 1901–2010 (CABLE, CENTURY, HyLand, and SiB-3) with the latter three also shown for the Phase II simulation.

values with comparison models. A specific model of interest is given a thicker line than the other models for visualization (Purdy et al., 2019). Axis and map ranges are automatically set to minimum and maximum extent of data visualized. The user must develop their own customized packages to answer specific scientific questions and to alter and improve esthetic visualization of the model output. The Experiment package produces inter-simulation comparisons. The Diagnostics package requires a few hours per variable to produce the plots for the global MsTMIP-structured output.

An example of the Single Variable package visualization is shown in Figure 3. Here, we show Net Primary Productivity (NPP) for CENTURY as the highlighted model for the MsTMIP Phase II simulation (2011–2100) with CESM1-CAM5 climate and RCP4.5 CO_2 scenario. The top-left map displays the 2011–2100 annual NPP. The top-middle map shows the mean NPP from the MsTMIP Phase I models, SG3 simulation (CO_2, climate, and LCLUC all "on"); we note that the 1901–2010 time period for SG3 is not an equivalent time comparison to the 2011–2100 runs, but it provides a first-order check. The top-right map shows the difference between the first two maps.

The bottom-left panel of Figure 3 shows the annual time series from 1901 to 2100. Here, we can see that the MsTMIP Phase I SG3 simulations are given only from 1901 to 2010. Additional Model Farm output is shown for 1901–2010 from CABLE, CENTURY, HyLand, and SiB-3. This figure also shows Model Farm Phase II simulations for 2011–2100 for CENTURY, HyLand, and SiB-3. The bottom-middle panel shows the mean seasonal cycle for all models; two lines are given each to CENTURY, HyLand, and SiB-3, one line for 1901–2010 and the other for 2011–2100. Finally, the bottom-right panel shows the zonal/latitudinal averages again with two lines each for the three Phase II models. These six figures combined are useful diagnostics to determine if there any obvious problems with the model runs. For example, a unit conversion error (e.g., Stage 2) can lead to model output significantly outside the spread of the other models.
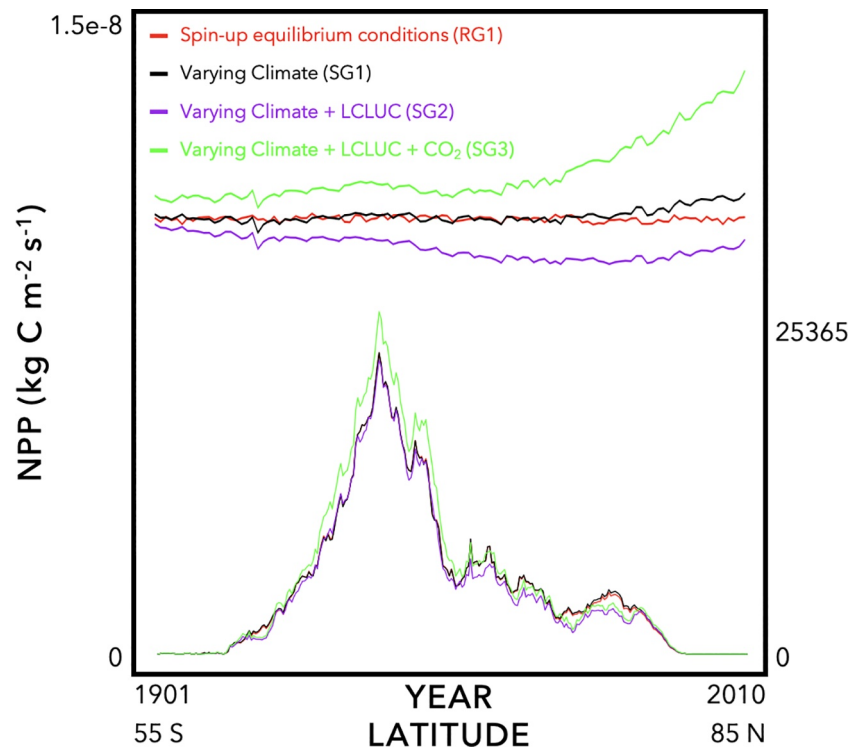
Figure 4. Inter-simulation visual diagnostics for MsTMIP Phase I experiments show how a single model changes output among different experiments. Here, we show output (e.g., Net Primary Productivity, NPP, for HyLand) from four experiments: (1) spin-up equilibrium conditions (RG1); (2) varying climate (SG1); (3) varying climate and land use/land cover change (LCLUC) (SG2); and (4) varying climate, LCLUC, and CO_2 (SG3). The top set of lines (primary y-axis) is the annual average from 1901 to 2010. The bottom set of lines (secondary y-axis) is the latitudinal/zonal sum from 55°S to 85°N.

An additional set of diagnostic visualizations are produced from the Experiment package to compare among different model experiments for a single given model (Figures 4 and 5). This is to ensure that there was indeed a difference among experiments, which may not happen if there was a problem in setting up the experiment. Moreover, this is also to ensure that the difference between model experiments coincides with theoretical expectations, for example, increasing NPP with increasing CO_2 (Schimel et al., 2015; Walker et al., 2021). Figures 4 and 5 are separated because the magnitude difference between Phase I experiments (Figure 4) and the output from Phase II (Figure 5) is so large that the Phase I differences would be too small to see on the raw Phase II scale. For the inter-simulation runs, we find that line plots are more useful for diagnosis than are maps, as differences and trends are more readily noticeable to the eye than in maps.

Numerous analyses have been published on 1901–2010 model and simulation experiment intercomparisons from MsTMIP and other MIPs (e.g., Cui et al., 2019; El Masri et al., 2019; Fang et al., 2017; Fisher et al., 2016; He et al., 2020; K. Huang et al., 2018; Huntzinger et al., 2017, 2020; Ito et al., 2016; Kolus et al., 2019; Liu et al., 2019; Mao et al., 2015; Schwalm et al., 2015, 2017, 2019, 2020; Shao et al., 2016; Thomas et al., 2016; S. Zhou et al., 2017, 2018). However, the 2011–2100 Phase II output is novel with multiple land models having been run with a suite of climate and RCP projections (see, also, Warszawski et al., 2014). In the annual time series example in Figure 5, we can see simultaneously the spread in NPP caused by differences (or uncertainties) in climate, which turn out to be relatively small compared to the large differences due to CO_2 scenario (RCP 4.5 vs. RCP 8.5). These results continue to suggest the overwhelming dominance of the CO_2 fertilization effect uncertainty on land carbon uptake uncertainty both historically and into the future (Arora et al., 2020; Huntzinger et al., 2017; Schimel et al., 2015; Sitch et al., 2015; Walker et al., 2021).
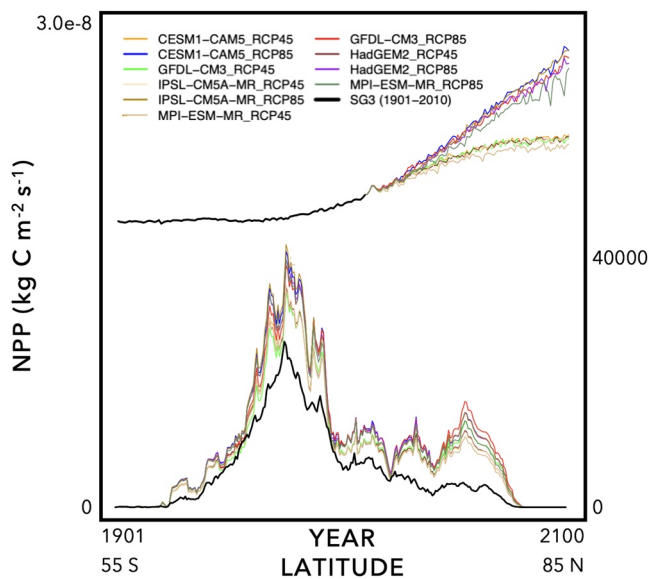
Figure 5. Inter-simulation visual diagnostics for MsTMIP Phase II experiments show how a single model changes output among five different climate projections and two Representative Concentration Pathways (RCPs). The climate projections are from five Earth System Models: (1) CESM1-CAM5, (2) GFDL-CM3, (3) HadGEM2, (4) IPSL-CM5A-MR, and (5) MPI-ESM-MR. The two RCPs are: (1) business-as-usual (RCP8.5) and (2) medium mitigation (RCP4.5). Output is for Net Primary Productivity (NPP) for the HyLand model. The top set of lines (primary *y*-axis) is the annual average from 1901 to 2100 (1901–2010 is from the retrospective SG3 scenario, black line). The bottom set of lines (secondary *y*-axis) is the latitudinal/zonal sum from 55°S to 85°N.

4. Discussion

The Terrestrial Biosphere Model Farm aims to reduce the time and skill required to execute a model experiment. An undergraduate student with no modeling experience and moderate programming skills can integrate a new TBM, spin-up the TBM, and execute at least one experiment with a 10-week summer internship. Indeed, many of the coauthors on this paper were those undergraduate interns. The core technical team would already have the TBM available and compiled with supported libraries installed and debugged before those internships began. Interns would spend the first few weeks learning about TBMs in general, studying the model code, and mapping the model I/O structure. They would then spend at least a couple of weeks mapping the LCLUC forcing data onto the TBM PFT structure. Another week or two would be spent connecting the model to the Farm and running tests. The next couple of weeks would be spent on spin-up and debugging. The final couple of weeks could allow the first MsTMIP simulation, assuming all else was successful to that point (which was not always). Some interns continued working on their model into their school year or returned for a following summer for more model runs and experiments. These examples illustrate the relative ease and challenges that the Farm afforded for running TBMs in contrast to, for example, traditional model development centers.

Generally, the Terrestrial Biosphere Model Farm provided several complements to traditional MIPs. Once the models were properly setup, it was relatively fast to run multiple experiments across models. Moreover, a team organizing a MIP might be interested in many different types of experiments—more than they would be comfortable specifying not wanting to bother so many people and risking incomplete runs. With a centralized Farm, once the forcing data and experiments are setup, multiple models can run with relative ease. One example of different types of experiments that the Farm supported was on LCLUC data set development for NASA's Carbon Monitoring System (CMS) (Kennedy et al., 2018; Neeti & Kennedy, 2016; Y. Zhou et al., 2021). Here, the team focused on new and different types of LCLUC data products and their impacts on TBMs. We were able to ingest these different LCLUC data sets and provide the corresponding variability in TBM output. Ultimately, future MIPs should expand to include not only multiple models, but also multiple forcing data sets (Dietze et al., 2014). Additionally, the Farm can be used as a "sandbox" with which to test and debug any problems with the new experiments. Once those are worked out, then the MIP can be extended to individual modeling teams so that each team does not have to waste time working out those bugs.

For each model, there were a number of user decisions that had to be made in setup, for example, how to map the LCLUC forcing data onto the model PFTs. For instance, a pixel may be classified as "mixed forest" in the forcing data, but a model may have multiple forest PFTs (Poulter et al., 2011). By erring consistently toward a single definition of mixed forest, for better or for worse, we could at least constrain differences in model output due to those types of user decisions. A similar process may be considered for classifying different crop types. Still, LCLUC mapping and debugging was a very time-consuming process. Likewise, while MIPs may specify spin-up requirements and forcing data, it is challenging to know if the modeling teams adhered completely to those specifications. In the Farm, all models were run uniformly across those specifications, even if models had different ways of going about spin-up, time steps, and use of data. However, while these aspects potentially reduce spread among model outputs, we also saw ways in which one could tune models to match benchmark data sets (e.g., scalars for stocks and fluxes). Being model and output agnostic, we did not tune. But, as noted earlier, models may converge toward each other in traditional MIPs simply because of tuning (Dommenget & Rezny, 2018; Fer et al., 2021; Hourdin et al., 2017; Notz, 2015; Räisänen, 2007; Scheiter et al., 2013).

Nonetheless, there were challenges in running the Farm and using its output. While "supercomputers" may be super at fast processing, they can be bottlenecked by file storage and transfer (Xie et al., 2012). Multiple models and multiple scenarios for hundreds of years for thousands of pixels result in a lot of data. Supercomputers are

not data archive centers, and so a conflict arises when these requirements are levied upon them at least in the relative short term. There needs to be a place to move intermediary model output before they are finalized for storage at a data archive, allowing time for debugging and output assessment. Moreover, supercomputers are a shared resource, and one cannot monopolize time on them for extended periods. Patches that require extra spin-up time to reach equilibrium run into that conflict, requiring disproportionate labor to complete a small number of pixels. Additionally, supercomputers are continually evolving, updating, and being replaced (Lim et al., 2015). Migration and re-setup of the Farm are very time-consuming for each new system.

Beyond the technical challenges, there are bounds to the scientific utility of the Model Farm output. Like super-computers, individual models are continuously evolving and updated by the respective modeling teams (e.g., Bonan, 1998; Bonan & Levis, 2006; Bonan et al., 2002; Lawrence et al., 2011, 2019). It requires significant labor to repeatedly replace Farm models with their most recent versions; inevitably, the models in the Farm may not be up to date. As such, examination of model Farm output may not necessarily be from the most recent model versions. This issue is common for traditional MIPs too, though modelers sometimes update their results later for the MIPs as their models change. Some MIPs manage to conduct regular updates (Friedlingstein et al., 2020). On the other hand, models in the Farm might be considered more stable than newer model versions that may still have bugs to be discovered. Further, when trying to publish results from most MIPs, reviewers typically hit those papers with, "Why did the models behave the way they did?" This is challenging for any individual modeler to answer about their own model, let alone a centralized system that runs other people's models as effective "black boxes."

So, how do we overcome these challenges, yet still advance the needs for robust MIPs? In Fer et al. (2021), we presented a roadmap to community cyberinfrastructure for ecological data-model integration. Five community tools are needed to accelerate the merging of models and data: (a) shared workflows, (b) scalable data ingest, (c) calibration, (d) benchmarking, and (e) iterative data assimilation. While the Farm addresses (a), (b), and some of (d), all five of these tools are implemented within a multimodel workflow management system called PEcAn (Dietze et al., 2013; LeBauer et al., 2013). PEcAn, like the Model Farm, is a tool to run multiple TBMs on a centralized system. However, unlike the Model Farm, PEcAn is an open-source utility that is available to anyone (pecanproject.org), can be installed on anything from a laptop to a supercomputer, or can be run in the cloud. As such, PEcAn operates as a distributed network of nodes rather than a single centralized server, which allows MIPs to either be fully centralized on one node—analogous to the Farm—or run by individual teams and synced via the PEcAn API, which may reduce the "other people's models" challenge mentioned earlier. Because PEcAn is containerized in Docker, it also greatly reduces the challenges of getting new models compiled and running on new serves as well migrating the system when systems are upgraded.

PEcAn provides a number of language-agnostic API tools to assist with model calibration, data fusion, analysis, and provenance tracking (Fer et al., 2018; LeBauer et al., 2013; Shiklomanov et al., 2020). As part of its data assimilation workflow, PEcAN can quantify, propagate, and analyze uncertainties associated with parameter, driver, and initial condition uncertainty distributions (Dietze et al., 2014). Further capabilities are in progress that are not too dissimilar from a community model including, for example, links to benchmarking and validation efforts (Abramowitz, 2012; Blyth et al., 2011; Collier et al., 2016; Kelley et al., 2013; Luo et al., 2012; Randerson et al., 2009; Schwalm et al., 2010). PEcAn's initial development has been focused on site- to continental-scale research; whereas, the Model Farm has focused on global runs, requiring the distribution and gathering of inputs and outputs. Still, like the Farm, PEcAn requires significant modeler time to develop couplers for inputs, runs, and updates. Nonetheless, PEcAn provides an excellent platform for unifying models for MIPs and ultimately saves researchers' time on developing the informatics and analysis tools built into PEcAn; we highly recommend that modelers begin to explore it. A blend of a centralized Farm team with the open platform of PEcAN could be a beneficial combination. The next frontier is a merger of the strengths of the Model Farm and those of PEcAn, which can create a whole that is larger than the sum of its parts.

5. Conclusion

Earth system modelers are like the blind men and the elephant, each trying to seek the truth in front of them—the future of the Earth; and, each saying something different, yet similar. By working together through MIPs, we may be able to piece that picture together (Figure 6). Yet, MIPs are challenging, and we need to ensure that each of those voices accurately depicts a defined perspective. We attempted to systematically unify and control for a
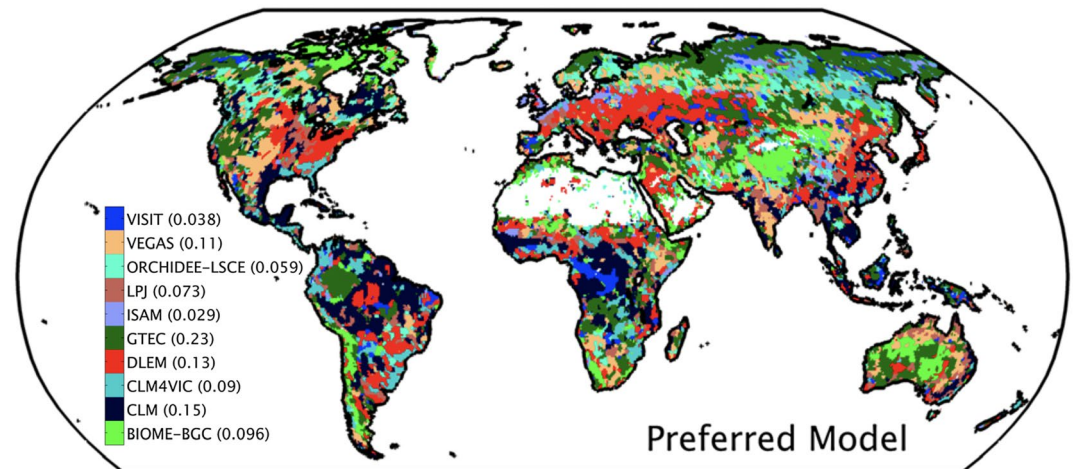
Figure 6. The terrestrial Earth as represented by the top-performing terrestrial biosphere model in MsTMIP based on a suite of benchmarks. From Schwalm et al. (2015).

number of "free parameters," or degrees of freedom, in MIPs through a centralized Model Farm. While solving these problems raised other challenges, we did bring something novel to the conversation of how MIPs are operated and interpreted. As we collectively continue to strive toward understanding the future, we must not only continue working together, but we must do so with improved consistency, transparency, and efficiency. And, we must do so quickly, as time is of the essence—the elephant may be gone by the time we figure out what was in front of us.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The authors primarily used the MsTMIP data and simulation protocols (Huntzinger et al., 2013; Wei et al., 2014).

References

Abramowitz, G. (2012). Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geoscientific Model Development*, *5*(3), 819–827. https://doi.org/10.5194/gmd-5-819-2012

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173–4222. https://doi.org/10.5194/bg-17-4173-2020

Badgley, G., Fisher, J. B., Jiménez, C., Tu, K. P., & Vinukollu, R. K. (2015). On uncertainty in global evapotranspiration estimates from choice of input forcing datasets. *Journal of Hydrometeorology*, *16*(4), 1449–1455. https://doi.org/10.1175/jhm-d-14-0040.1

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., & da Rocha, H. R. (2008). Seasonal drought stress in the Amazon: Reconciling models and observations. *Journal of Geophysical Research*, *113*(G1), G00B01. https://doi.org/10.1029/2007jg000644

Bellouin, N., Collins, W., Culverwell, I., Halloran, P., Hardiman, S., Hinton, T., et al. (2011). The HadGEM2 family of met office unified model climate configurations. *Geoscientific Model Development*, *4*(3), 723–757.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor.

Blyth, E., Clark, D., Ellis, R., Huntingford, C., Los, S., Pryor, M., et al. (2011). A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geoscientific Model Development*, *4*(2), 255–269. https://doi.org/10.5194/gmd-4-255-2011

Bonan, G. B. (1998). The land surface climatology of the NCAR Land Surface Model coupled to the NCAR Community Climate Model. *Journal of Climate*, *11*(6), 1307–1326. https://doi.org/10.1175/1520-0442(1998)011<1307:tlscot>2.0.co;2

Bonan, G. B., Chapin, F. S., III, & Thompson, S. L. (1995). Boreal forest and tundra ecosystems as components of the climate system. *Climatic Change*, *29*(2), 145–167. https://doi.org/10.1007/bf01094014

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

Bonan, G. B., & Levis, S. (2006). Evaluating aspects of the community land and atmosphere models (CLM3 and CAM3) using a dynamic global vegetation model. *Journal of Climate*, *19*(11), 2290–2301. https://doi.org/10.1175/jcli3741.1

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., et al. (2002). The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. *Journal of Climate*, *15*(22), 3123–3149. https://doi.org/10.1175/1520-0442(2002) 015<3123:tlscot>2.0.co;2

Bretherton, C. S., & Caldwell, P. M. (2020). Combining emergent constraints for climate sensitivity. *Journal of Climate*, *33*(17), 7413–7430. https://doi.org/10.1175/jcli-d-19-0911.1

Brient, F. (2020). Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects. *Advances in Atmospheric Sciences*, *37*(1), 1–15. https://doi.org/10.1007/s00376-019-9140-8

Caldwell, P. M., Zelinka, M. D., & Klein, S. A. (2018). Evaluating emergent constraints on equilibrium climate sensitivity. *Journal of Climate*, *31*(10), 3921–3942. https://doi.org/10.1175/jcli-d-17-0631.1

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

Collier, N., Hoffman, F. M., Mu, M., Randerson, J. T., & Riley, W. J. (2016). *International Land Model Benchmarking (ILAMB) Package v002. 00* (Report). BGCF-DATA (Biogeochemistry (BGC) Feedbacks).

Cox, P. M., Huntingford, C., & Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, *553*(7688), 319–322. https://doi.org/10.1038/nature25450

Cramer, W., Bondeau, A., Woodward, F. I., Prentice, I. C., Betts, R. A., Brovkin, V., et al. (2001). Global response of terrestrial ecosystem structure and function to CO_2 and climate change: Results six dynamic global vegetation models. *Global Change Biology*, *7*(4), 357–373. https://doi.org/10.1046/j.1365-2486.2001.00383.x

Cramer, W., Kicklighter, D., Bondeau, A., Iii, B. M., Churkina, G., Nemry, B., et al. (1999). Comparing global models of terrestrial net primary productivity (NPP): Overview and key results. *Global Change Biology*, *5*(S1), 1–15. https://doi.org/10.1046/j.1365-2486.1999.00009.x

Cui, E., Huang, K., Arain, M. A., Fisher, J. B., Huntzinger, D. N., Ito, A., et al. (2019). Vegetation functional properties determine uncertainty of simulated ecosystem productivity: A traceability analysis in the East Asian monsoon region. *Global Biogeochemical Cycles*, *33*(6), 668–689. https://doi.org/10.1029/2018gb005909

de Gonçalves, L. G. G., Borak, J. S., Costa, M. H., Saleska, S. R., Baker, I., Restrepo-Coupe, N., et al. (2013). Overview of the large-scale biosphere–atmosphere experiment in Amazonia Data Model Intercomparison Project (LBA-DMIP). *Agricultural and Forest Meteorology*, *182–183*, 111–127.

Denning, S., Stockli, R., & Lu, L. (2009). *A global, 1-km vegetation modeling system for NEWS* (Report, pp. 1–7). Colorado State University.

Dentener, F., Drevet, J., Lamarque, J. F., Bey, I., Eickhout, B., Fiore, A. M., et al. (2006). Nitrogen and sulfur deposition on regional and global scales: A multimodel evaluation. *Global Biogeochemical Cycles*, *20*(4). https://doi.org/10.1029/2005GB002672

Dietze, M. C., Lebauer, D. S., & Kooper, R. (2013). On improving the communication between models and data. *Plant, Cell & Environment*, *36*(9), 1575–1585. https://doi.org/10.1111/pce.12043

Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., et al. (2014). A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 286–300. https://doi.org/10.1002/2013jg002392

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. https://doi.org/10.1175/bams-87-10-1381

Dommenget, D., & Rezny, M. (2018). A caveat note on tuning in the development of coupled climate models. *Journal of Advances in Modeling Earth Systems*, *10*(1), 78–97. https://doi.org/10.1002/2017ms000947

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dynamics*, *40*(9), 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

El-Masri, B., Barman, R., Meiyappan, P., Song, Y., Liang, M., & Jain, A. K. (2013). Carbon dynamics in the Amazonian Basin: Integration of eddy covariance and ecophysiological data with a land surface model. *Agricultural and Forest Meteorology*, *182*, 156–167. https://doi.org/10.1016/j.agrformet.2013.03.011

El Masri, B., Schwalm, C., Huntzinger, D. N., Mao, J., Shi, X., Peng, C., et al. (2019). Carbon and water use efficiencies: A comparative analysis of ten terrestrial ecosystem models under changing climate. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-50808-7

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Fang, Y., Michalak, A. M., Schwalm, C. R., Huntzinger, D. N., Berry, J. A., Ciais, P., et al. (2017). Global land carbon sink response to temperature and precipitation varies with ENSO phase. *Environmental Research Letters*, *12*(6), 064007. https://doi.org/10.1088/1748-9326/aa6e8e

Fer, I., Gardella, A. K., Shiklomanov, A. N., Campbell, E. E., Cowdery, E. M., De Kauwe, M. G., et al. (2021). Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, *27*(1), 13–26. https://doi.org/10.1111/gcb.15409

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., & Dietze, M. C. (2018). Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, *15*(19), 5801–5830. https://doi.org/10.5194/bg-15-5801-2018

Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., & Sitch, S. (2014). Modeling the terrestrial biosphere. *Annual Review of Environment and Resources*, *39*, 91–123. https://doi.org/10.1146/annurev-environ-012913-093456

Fisher, J. B., Sikka, M., Huntzinger, D. N., Schwalm, C., & Liu, J. (2016). Technical note: 3-hourly temporal downscaling of monthly global terrestrial biosphere model net ecosystem exchange. *Biogeosciences*, *13*(14), 4271–4277. https://doi.org/10.5194/bg-13-4271-2016

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *Journal of Climate*, *19*(14), 3337–3353. https://doi.org/10.1175/jcli3800.1

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global carbon budget 2020. *Earth System Science Data*, *12*(4), 3269–3340. https://doi.org/10.5194/essd-12-3269-2020

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., et al. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597. https://doi.org/10.1002/jame.20038

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., et al. (2011). The GFDL CM3 coupled climate model: Characteristics of the ocean and sea ice simulations. *Journal of Climate*, *24*(13), 3520–3544. https://doi.org/10.1175/2011jcli3964.1

Hall, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, *9*(4), 269–278. https://doi.org/10.1038/s41558-019-0436-6

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Hayes, D. J., McGuire, A. D., Kicklighter, D. W., Gurney, K. R., Burnside, T. J., & Melillo, J. M. (2011). Is the northern high-latitude land-based CO_2 sink weakening? *Global Biogeochemical Cycles*, *25*(3). https://doi.org/10.1029/2010gb003813

He, Y., Peng, S., Liu, Y., Li, X., Wang, K., Ciais, P., et al. (2020). Global vegetation biomass production efficiency constrained by models and observations. *Global Change Biology*, *26*(3), 1474–1484. https://doi.org/10.1111/gcb.14816

Henderson-Sellers, A., Yang, Z., & Dickinson, R. (1993). The project for intercomparison of land-surface parameterization schemes. *Bulletin of the American Meteorological Society*, *74*(7), 1335–1349. https://doi.org/10.1175/1520-0477(1993)074<1335:tpfiol>2.0.co;2

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. https://doi.org/10.1175/bams-d-15-00135.1

Huang, K., Xia, J., Wang, Y., Ahlström, A., Chen, J., Cook, R. B., et al. (2018). Enhanced peak growth of global vegetation and its key mechanisms. *Nature Ecology & Evolution*, *2*(12), 1897–1905. https://doi.org/10.1038/s41559-018-0714-0

Huang, S., Arain, M. A., Arora, V. K., Yuan, F., Brodeur, J., & Peichl, M. (2011). Analysis of nitrogen controls on carbon and water exchanges in a conifer forest using the CLASS-CTEMN+ model. *Ecological Modelling*, *222*(20–22), 3743–3760. https://doi.org/10.1016/j.ecolmodel.2011.09.008

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., et al. (2017). Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Scientific Reports*, *7*(4765), 1–8. https://doi.org/10.1038/s41598-017-03818-2

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., et al. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, *232*, 144–157. https://doi.org/10.1016/j.ecolmodel.2012.02.004

Huntzinger, D. N., Schaefer, K., Schwalm, C., Fisher, J., Hayes, D., Stofferahn, E., et al. (2020). Evaluation of simulated soil carbon dynamics in Arctic-Boreal ecosystems. *Environmental Research Letters*, *15*(2), 025005. https://doi.org/10.1088/1748-9326/ab6784

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., et al. (2013). The North American Carbon Program multi-scale synthesis and Terrestrial Model Intercomparison Project–Part 1: Overview and experimental design. *Geoscientific Model Development*, *6*, 2121–2133. https://doi.org/10.5194/gmd-6-2121-2013

Hurtt, G., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., et al. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, *109*(1), 117–161. https://doi.org/10.1007/s10584-011-0153-2

IPCC. (2014). Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. *Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 1132). Cambridge University Press.

Ito, A. (2010). Changing ecophysiological processes and carbon budget in East Asian ecosystems under near-future changes in climate: Implications for long-term monitoring from a process-based model. *Journal of Plant Research*, *123*(4), 577–588. https://doi.org/10.1007/s10265-009-0305-x

Ito, A., Inatomi, M., Huntzinger, D. N., Schwalm, C., Michalak, A. M., Cook, R., et al. (2016). Decadal trends in the seasonal-cycle amplitude of terrestrial CO_2 exchange resulting from the ensemble of terrestrial biosphere models. *Tellus B: Chemical and Physical Meteorology*, *68*(1), 28968. https://doi.org/10.3402/tellusb.v68.28968

Jain, A. K., & Yang, X. (2005). Modeling the effects of two different land cover change data sets on the carbon stocks of plants and soils in concert with CO_2 and climate change. *Global Biogeochemical Cycles*, *19*(2), GB2015. https://doi.org/10.1029/2004gb002349

Jung, M., Henkel, K., Herold, M., & Churkina, G. (2006). Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, *101*(4), 534–553. https://doi.org/10.1016/j.rse.2006.01.020

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., & Willis, K. O. (2013). A comprehensive benchmarking system for evaluating global vegetation models. *Biogeosciences*, *10*(5), 3313–3340. https://doi.org/10.5194/bg-10-3313-2013

Kennedy, R. E., Ohmann, J., Gregory, M., Roberts, H., Yang, Z., Bell, D. M., et al. (2018). An empirical, integrated forest biomass monitoring system. *Environmental Research Letters*, *13*(2), 025004. https://doi.org/10.1088/1748-9326/aa9d9e

Kicklighter, D. W., Bruno, M., Dönges, S., Esser, G., Heimann, M., Helfrich, J., et al. (1999). A first-order analysis of the potential role of CO_2 fertilization to affect the global carbon budget: A comparison of four terrestrial biosphere models. *Tellus B: Chemical and Physical Meteorology*, *51*(2), 343–366. https://doi.org/10.1034/j.1600-0889.1999.00017.x

Kolus, H. R., Huntzinger, D. N., Schwalm, C. R., Fisher, J. B., McKay, N., Fang, Y., et al. (2019). Land carbon models underestimate the severity and duration of drought's impact on plant productivity. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-39373-1

Krinner, G., Viovy, N., de Noblet-Ducoudrè, N., Ogèe, J., Polcher, J., Friedlingstein, P., et al. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, *19*(1), GB1015. https://doi.org/10.1029/2003gb002199

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., et al. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software*, *21*(10), 1402–1415. https://doi.org/10.1016/j.envsoft.2005.07.004

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018ms001583

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., et al. (2011). Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*(1). https://doi.org/10.1029/2011ms00045

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., & Dietze, M. C. (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs*, *83*(2), 133–154. https://doi.org/10.1890/12-0137.1

Lei, H., Huang, M., Leung, L. R., Yang, D., Shi, X., Mao, J., et al. (2014). Sensitivity of global terrestrial gross primary production to hydrologic states simulated by the Community Land Model using two runoff parameterizations. *Journal of Advances in Modeling Earth Systems*, *6*(3), 658–679. https://doi.org/10.1002/2013ms000252

Levy, P. E., Cannell, M. G. R., & Friend, A. D. (2004). Modelling the impact of future changes in climate, CO_2 concentration and land use on natural ecosystems and the terrestrial carbon sink. *Global Environmental Change*, *14*, 21–30. https://doi.org/10.1016/j.gloenvcha.2003.10.005

Lim, D.-J., Anderson, T. R., & Shott, T. (2015). Technological forecasting of supercomputer development: The March to Exascale computing. *Omega*, *51*, 128–135. https://doi.org/10.1016/j.omega.2014.09.009

Liu, Y., Piao, S., Gasser, T., Ciais, P., Yang, H., Wang, H., et al. (2019). Field-experiment constraints on the enhancement of the terrestrial carbon sink by CO_2 fertilization. *Nature Geoscience*, *12*(10), 809–814. https://doi.org/10.1038/s41561-019-0436-1

Luo, Y., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., et al. (2012). A framework for benchmarking land models. *Biogeosciences*, *9*(10), 3857–3874. https://doi.org/10.5194/bg-9-3857-2012

Mao, J., Fu, W., Shi, X., Ricciuto, D. M., Fisher, J. B., Dickinson, R. E., et al. (2015). Disentangling climatic and anthropogenic controls on global terrestrial evapotranspiration trends. *Environmental Research Letters*, *10*(9), 094008. https://doi.org/10.1088/1748-9326/10/9/094008

Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J., et al. (2015). Using ecosystem experiments to improve vegetation models. *Nature Climate Change*, *5*(6), 528–534. https://doi.org/10.1038/nclimate2621

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., et al. (2013). Climate change projections in CESM1 (CAM5) compared to CCSM4. *Journal of Climate*, *26*(17), 6287–6308. https://doi.org/10.1175/jcli-d-12-00572.1

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., et al. (2013). Present state of global wetland extent and wetland methane modelling: Conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences*, *10*(2), 753–788. https://doi.org/10.5194/bg-10-753-2013

Neeti, N., & Kennedy, R. (2016). Comparison of national level biomass maps for conterminous US: Understanding pattern and causes of differences. *Carbon Balance and Management*, *11*(1), 1–20. https://doi.org/10.1186/s13021-016-0060-y

Nemani, R., Votava, P., Roads, J., White, M., Running, S., & Coughlan, J. (2002). Terrestrial observation and prediction system: Integration of satellite and surface weather observations with ecosystem models. In *Paper presented at IEEE International Geoscience and Remote Sensing Symposium*. IEEE.

Nietzsche, F. (1977). *The portable Nietzsche*. Penguin.

Nijsse, F., Cox, P. M., & Williamson, M. S. (2020). An emergent constraint on Transient Climate Response from simulated historical warming in CMIP6 models. *Earth System Dynamics Discussions*, *2020*, 1–14.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12). https://doi.org/10.1029/2010jd015139

Notz, D. (2015). How well must climate models agree with observations? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2052), 20140164. https://doi.org/10.1098/rsta.2014.0164

Parton, W. J., Mosier, A. R., & Schimel, D. S. (1988). Dynamics of C, N, P, and S in grassland soils: A model. *Biogeochemistry*, *5*, 109–131. https://doi.org/10.1007/bf02180320

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., & Zimmermann, N. (2011). Plant functional type mapping for Earth system models. *Geoscientific Model Development*, *4*(4), 993–1010. https://doi.org/10.5194/gmd-4-993-2011

Purdy, A. J., Kawata, J., Fisher, J. B., Reynolds, M., Om, G., Ali, Z., et al. (2019). Designing drought indicators. *Bulletin of the American Meteorological Society*, *100*(11), 2327–2341. https://doi.org/10.1175/bams-d-18-0146.1

Rääisäänen, J. (2007). How reliable are climate models? *Tellus A: Dynamic Meteorology and Oceanography*, *59*(1), 2–29. https://doi.org/10.1111/j.1600-0870.2006.00211.x

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

Redelsperger, J.-L., Thorncroft, C. D., Diedhiou, A., Lebel, T., Parker, D. J., & Polcher, J. (2006). African Monsoon Multidisciplinary Analysis: An international research project and field campaign. *Bulletin of the American Meteorological Society*, *87*(12), 1739–1746. https://doi.org/10.1175/bams-87-12-1739

Ricciuto, D. M., King, A. W., Dragoni, D., & Post, W. M. (2011). Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: Effects of constraining variables and data record length. *Journal of Geophysical Research: Biogeosciences*(G1), 116. https://doi.org/10.1029/2010jg001400

Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., Von Caemmerer, S., Dietze, M. C., et al. (2017). A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytologist*, *213*(1), 22–42. https://doi.org/10.1111/nph.14283

Rollinson, C. R., Dawson, A., Raiho, A. M., Williams, J. W., Dietze, M. C., Hickler, T., et al. (2021). Forest responses to last-millennium hydroclimate variability are governed by spatial variations in ecosystem sensitivity. *Ecology Letters*, *24*(3), 498–508. https://doi.org/10.1111/ele.13667

Sanderson, B. M., Knutti, R., & Caldwell, P. (2015). A representative democracy to reduce interdependency in a multimodel ensemble. *Journal of Climate*, *28*(13), 5171–5194. https://doi.org/10.1175/jcli-d-14-00362.1

Sanderson, B. M., Pendergrass, A. G., Koven, C. D., Brient, F., Booth, B. B. B., Fisher, R. A., & Knutti, R. (2021). The potential for structural errors in emergent constraints. *Earth System Dynamics*, *12*(3), 899–918. https://esd.copernicus.org/articles/12/899/2021/

Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., et al. (2008). Combined simple biosphere/Carnegie-Ames-Stanford approach terrestrial carbon cycle model. *Journal of Geophysical Research*, *113*(G3), G03034. https://doi.org/10.1029/2007jg000603

Scheiter, S., Langan, L., & Higgins, S. I. (2013). Next-generation dynamic global vegetation models: Learning from community ecology. *New Phytologist*, *198*(3), 957–969. https://doi.org/10.1111/nph.12210

Schimel, D., Stephens, B. B., & Fisher, J. B. (2015). Effect of increasing CO_2 on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, *112*(2), 436–441. https://doi.org/10.1073/pnas.1407302112

Schwalm, C. R., Anderegg, W. R., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., et al. (2017). Global patterns of drought recovery. *Nature*, *548*(7666), 202–205. https://doi.org/10.1038/nature23021

Schwalm, C. R., Huntzinger, D. N., Fisher, J. B., Michalak, A. M., Bowman, K., Ciais, P., et al. (2015). Toward "optimal" integration of terrestrial biosphere models. *Geophysical Research Letters*, *42*(11), 4418–4428. https://doi.org/10.1002/2015gl064002

Schwalm, C. R., Huntzinger, D. N., Michalak, A. M., Schaefer, K., Fisher, J. B., Fang, Y., & Wei, Y. (2020). Modeling suggests fossil fuel emissions have been driving increased land carbon uptake since the turn of the 20th Century. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-66103-9

Schwalm, C. R., Schaefer, K., Fisher, J. B., Huntzinger, D., Elshorbany, Y., Fang, Y., et al. (2019). Divergence in land surface modeling: Linking spread to structure. *Environmental Research Communications*, *1*(11), 111004. https://doi.org/10.1088/2515-7620/ab4a8a

Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., et al. (2010). A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, *115*(G3). https://doi.org/10.1029/2009jg001229

Shao, J., Zhou, X., Luo, Y., Zhang, G., Yan, W., Li, J., et al. (2016). Uncertainty analysis of terrestrial net primary productivity and net biome productivity in China during 1901–2005. *Journal of Geophysical Research: Biogeosciences*, *121*(5), 1372–1393. https://doi.org/10.1002/2015jg003062

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

Shiklomanov, A. N., Cowdery, E. M., Bahn, M., Byun, C., Jansen, S., Kramer, K., et al. (2020). Does the leaf economic spectrum hold within plant functional types? A Bayesian multivariate trait meta-analysis. *Ecological Applications*, *30*(3), e02064. https://doi.org/10.1002/eap.2064

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., et al. (2013). Trends and drivers of regional sources and sinks of carbon dioxide over the past two decades. *Biogeosciences Discussions*, *10*(12), 20113–20177. https://doi.org/10.5194/bgd-10-20113-2013

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., et al. (2015). Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, *12*(3), 653–679. https://doi.org/10.5194/bg-12-653-2015

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., et al. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, *9*, 161–185. https://doi.org/10.1046/j.1365-2486.2003.00569.x

Smith, B., Prentice, I. C., & Sykes, M. T. (2001). Representation of vegetation dynamics in modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space. *Global Ecology and Biogeography*, *10*, 621–637. https://doi.org/10.1046/j.1466-822x.2001.t01-1-00256.x

Thomas, R. T., Prentice, I. C., Graven, H., Ciais, P., Fisher, J. B., Hayes, D. J., et al. (2016). Increased light-use efficiency in northern terrestrial ecosystems indicated by CO_2 and greening observations. *Geophysical Research Letters*, *43*(21), 11339–11349. https://doi.org/10.1002/2016gl070710

Tian, H., Chen, G., Zhang, C., Liu, M., Sun, G., Chappelka, A., et al. (2012). Century-scale responses of ecosystem carbon storage and flux to multiple environmental changes in the southern United States. *Ecosystems*, *15*(4), 674–694. https://doi.org/10.1007/s10021-012-9539-x

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), eaaz9549. https://doi.org/10.1126/sciadv.aaz9549

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., et al. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1), 5–31. https://doi.org/10.1007/s10584-011-0148-z

VEMAP, M. (1995). Vegetation/ecosystem modeling and analysis project: Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO_2 doubling. *Global Biogeochemical Cycles*, *9*, 407–437.

Walker, A. P., De Kauwe, M. G., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R. F., et al. (2021). Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO_2. *New Phytologist*, *229*(5), 2413–2445.

Wang, Y. P., Law, R. M., & Pak, B. (2010). A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. *Biogeosciences*, *7*, 2261–2282. https://doi.org/10.5194/bg-7-2261-2010

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. *Proceedings of the National Academy of Sciences*, *111*(9), 3228–3232. https://doi.org/10.1073/pnas.1312330110

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., et al. (2014). The North American Carbon Program multi-scale synthesis and Terrestrial Model Intercomparison Project–Part 2: Environmental driver data. *Geoscientific Model Development*, *7*(6), 2875–2893. https://doi.org/10.5194/gmd-7-2875-2014

Xie, B., Chase, J., Dillow, D., Drokin, O., Klasky, S., Oral, S., & Podhorszki, N. (2012). Characterizing output bottlenecks in a supercomputer. In *Paper presented at SC'12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. IEEE. https://doi.org/10.1109/sc.2012.28

Zeng, N., Qian, H., Roedenbeck, C., & Heimann, M. (2005). Impact of 1998–2002 midlatitude drought and warming on terrestrial ecosystem and the global carbon cycle. *Geophysical Research Letters*, *32*(22), L22709. https://doi.org/10.1029/2005gl024607

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

Zhou, S., Yu, B., Schwalm, C. R., Ciais, P., Zhang, Y., Fisher, J. B., et al. (2017). Response of water use efficiency to global environmental change based on output from terrestrial biosphere models. *Global Biogeochemical Cycles*, *31*(11), 1639–1655. https://doi.org/10.1002/2017gb005733

Zhou, Y., Williams, C. A., Hasler, N., Gu, H., & Kennedy, R. E. (2021). Beyond biomass to carbon fluxes: Application and evaluation of a comprehensive forest carbon monitoring system. *Environmental Research Letters*, *16*(5), 055026. https://doi.org/10.1088/1748-9326/abf06d