Environmental Research Letters

LETTER

The Arctic-Boreal vulnerability experiment model benchmarking system

Eric Stofferahn[1,2] , Joshua B Fisher[1] , Daniel J Hayes[3] , Christopher R Schwalm[4],
Deborah N Huntzinger[5], Wouter Hantson[3], Benjamin Poulter[6] and Zhen Zhang[7]

[1] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, United States of America
[2] Conservation Science Partners, 11050 Pioneer Trail Suite #202, Truckee, CA, 96161, United States of America
[3] School of Forest Resources, University of Maine, 233 Nutting Hall, Orono, ME, 04469-5755, United States of America
[4] Woods Hole Research Center, 149 Woods Hole Road, Falmouth, MA, 02540-1644, United States of America
[5] School of Earth & Sustainability, Northern Arizona University, PO Box 5694, Flagstaff, AZ, 86011-5697, United States of America
[6] NASA Goddard Space Flight Center, Biospheric Science Laboratory, Greenbelt, MD, 20771, United States of America
[7] Department of Geographical Sciences, University of Maryland, College Park, MD 20740, United States of America

E-mail: eric@csp-inc.org

Keywords: ABoVE, benchmarking, model, Arctic, Boreal, ecosystem, carbon

Supplementary material for this article is available online

Abstract

NASA's Arctic-Boreal Vulnerability Experiment (ABoVE) integrates field and airborne data into modeling and synthesis activities for understanding Arctic and Boreal ecosystem dynamics. The ABoVE Benchmarking System (ABS) is an operational software package to evaluate terrestrial biosphere models against key indicators of Arctic and Boreal ecosystem dynamics, i.e.: carbon biogeochemistry, vegetation, permafrost, hydrology, and disturbance. The ABS utilizes satellite remote sensing data, airborne data, and field data from ABoVE as well as collaborating research networks in the region, e.g.: the Permafrost Carbon Network, the International Soil Carbon Network, the Northern Circumpolar Soil Carbon Database, AmeriFlux sites, the Moderate Resolution Imaging Spectroradiometer, the Orbiting Carbon Observatory 2, and the Soil Moisture Active Passive mission. The ABS is designed to be interactive for researchers interested in having their models accurately represent observations of key Arctic indicators: a user submits model results to the system, the system evaluates the model results against a set of Arctic-Boreal benchmarks outlined in the ABoVE Concise Experiment Plan, and the user then receives a quantitative scoring of model strengths and deficiencies through a web interface. This interactivity allows model developers to iteratively improve their model for the Arctic-Boreal Region by evaluating results from successive model versions. We show here, for illustration, the improvement of the Lund–Potsdam–Jena-Wald Schnee und Landschaft (LPJwsl) version model through the ABoVE ABS as a new permafrost module is coupled to the existing model framework. The ABS will continue to incorporate new benchmarks that address indicators of Arctic-Boreal ecosystem dynamics as they become available.

1. Introduction

The Arctic-Boreal Region (ABR) is experiencing unprecedented terrestrial ecosystem change (Serreze *et al* 2000, Hinzman *et al* 2005, McGuire *et al* 2006, Chapman and Walsh 2007). During the past three decades, the Arctic surface temperature has warmed at a rate of 1 °C per decade, which is substantially higher than the midlatitudes and tropics (Christensen *et al* 2013).

On one hand, increased ABR temperatures are accelerating the permafrost-carbon feedback to climate, wherein thawing permafrost enables the decomposition and release to the atmosphere of previously inaccessible soil carbon as carbon dioxide (CO_2) and methane (CH_4) (Hayes *et al* 2014), resulting in increased warming and further permafrost thaw (Schuur *et al* 2015). Conversely, temperature increases in the ABR might facilitate plant growth via nutrient

release from thawing permafrost, thus acting as a negative feedback by increasing carbon CO_2 uptake from the atmosphere (Mack *et al* 2004, Natali *et al* 2012). Changes in disturbance and hydrology also impact these nonlinear feedbacks, necessitating the use of a process-based modeling approach to determine the total impact of warming on ecosystems of the ABR, particularly as it relates to terrestrial carbon source/sink dynamics (Oechel *et al* 1993, McGuire *et al* 2009, Schuur *et al* 2009, Hayes *et al* 2011, Koven *et al* 2011, McGuire *et al* 2018).

While model simulations can elucidate ABR ecosystem dynamics, the accuracy of those simulations is informed by observational data collected from the region (Fisher et al 2018a). Until recently, data collection from the region was relatively sparse, due largely to the extreme environment and sheer size of the area. (Sitch *et al* 2007, McGuire *et al* 2012, Melton *et al* 2013, Fisher *et al* 2014a, 2014b, Schuur *et al* 2015). To address this, NASA has embarked on a decadal-scale study: the Arctic-Boreal Vulnerability Experiment (ABoVE), to collect additional *in situ* and airborne measurements, and create new satellite data products, of the ABR (https://above.nasa.gov; Goetz *et al* 2011, Griffith *et al* 2012, Kasischke *et al* 2013).
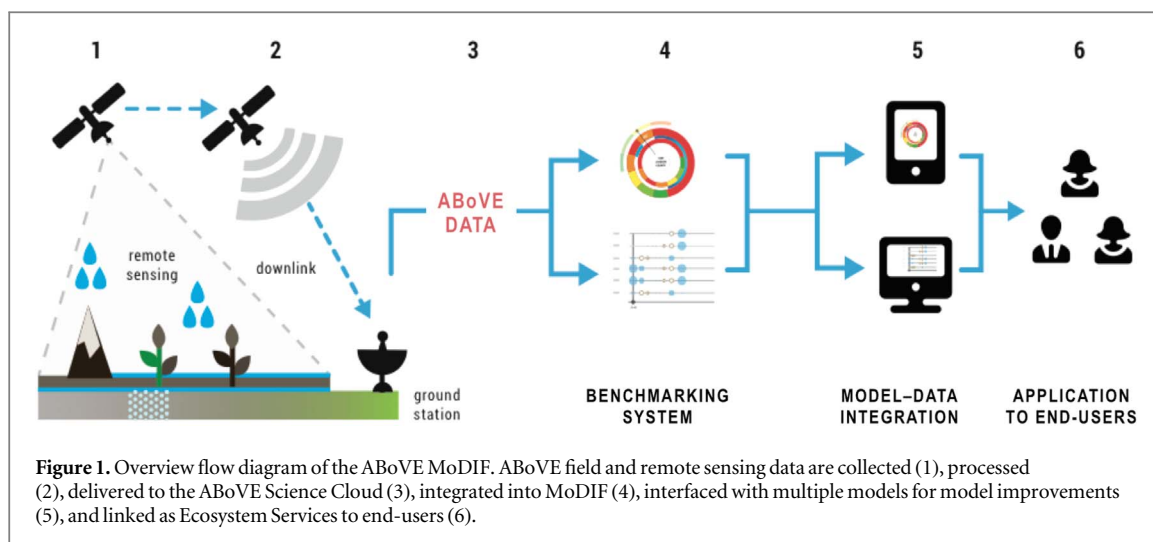
Several model benchmarking systems are in use by the modeling community to constrain model simulations with observed data, including the International Land Model Benchmarking (ILAMB) system (Luo *et al* 2012, Collier *et al* 2018, Hoffman *et al* 2017), the European Network for Earth System modeling (ENES) Benchmark Suite (https://redmine.dkrz.de/projects/enes-benchmark-suite/wiki), and the Permafrost Benchmarking System (PBS) (https://permamodel.github.io/pbs/). While ILAMB and ENES are comprehensive in their efforts to improve the performance of land models via model-data intercomparison, their global scope inhibits the development of benchmarks that specifically address key indicators of Arctic and Boreal ecosystem dynamics. The ABoVE Benchmarking System (ABS) utilizes permafrost benchmarks from the PBS (e.g. active layer depth) and adds additional benchmarks to address other ABR indicators (carbon dynamics, disturbance, ecosystem structure and function, and hydrology).

In addition to its mission of data collection, ABoVE is tasked with integrating data collection during the campaign with satellite data and model simulation efforts. Our Model-Data Integration Framework (MoDIF) is designed, in part, to ingest and organize ABoVE data into a format that is comparable with model output (e.g. NetCDF), thus facilitating the use of the system for benchmarking and model improvement (figure 1) (https://above.nasa.gov/cgi-bin/inv_pgp.pl?pgid=3394). At the core of the ABoVE MoDIF is the ABS, an ABR-focused model benchmarking system that is the main user interface between modelers and the ABoVE data (Stofferahn *et al* 2016). The ABS is operational software residing on the ABoVE Science

Cloud (ASC, https://above.nasa.gov/sciencecloud.html) and is designed to ingest different types of output from terrestrial biosphere models (TBMs) in order to evaluate their efficacy—and improvement through development—in simulating ABR ecosystem processes. An example of such model evaluation is shown in section 4, wherein the output from two versions of a sample TBM are evaluated through the ABS in order to make inference on the impact of changes between model versions (in this case, a change in treatment of permafrost) on overall model performance.

2. Observational data

The ABS incorporates data from *in situ*, airborne, and satellite measurements for comparison against model output. There has been one completed (2017) and one planned (2019) airborne campaign in the ABoVE region. The completed campaign has provided a variety of co-located data, summarized in supplementary table 1 is available online at stacks.iop.org/ERL/14/055002/mmedia: including Polarimetric Synthetic Aperture Radar (P&L-band), used for active layer thickness, AirSWOT (Ka-band Radar, used for detection of liquid water and topography), AVIRIS (hyperspectral data, for use in identifying vegetation), LVIS (LIDAR, useful for determining vegetation height), and CFIS (an instrument to measure solar-induced fluorescence (SIF)). In addition to data collected for the ABoVE campaign, there is also an enormous wealth of data and information existing or in development by non-ABoVE programs relevant to modelers. These include, for example: the Permafrost Carbon Network (Schuur and Abbott 2011, Schädel *et al* 2014), the International Soil Carbon Network (Jandl *et al* 2014, Nave *et al* 2016), the Northern Circumpolar Soil Carbon Database (Hugelius *et al* 2013), and individual AmeriFlux sites (e.g. Oechel *et al* 2014). This is in addition to measurements such as Gross Primary Production (GPP), Net Primary Production (NPP), and Evapotranspiration from the Moderate Resolution Imaging Spectroradiometer satellite system, SIF from the Orbiting Carbon Observatory 2, and soil moisture and freeze/thaw dynamics from the Soil Moisture Active Passive mission. The source data for each variable currently in the ABS is shown in supplementary table 2. It is expected that the list will never be 'complete': as new benchmarks become available, they will be added to the suite of existing benchmarks for a given variable, refining our estimate against which models will be evaluated. As any benchmark has an associated uncertainty (particularly those which are model-based or derived quantities), the addition of new benchmarks will reduce the overall uncertainty, providing a tighter constraint on model performance.

Figure 1. Overview flow diagram of the ABoVE MoDIF. ABoVE field and remote sensing data are collected (1), processed (2), delivered to the ABoVE Science Cloud (3), integrated into MoDIF (4), interfaced with multiple models for model improvements (5), and linked as Ecosystem Services to end-users (6).

3. Methods

3.1. Processing of observational data

Processing each benchmark dataset is done manually, and the full details of this processing are provided in the supplemental materials (S1). The data is transformed from its native file format to a netCDF file that has been re-projected to the ABoVE projection and grid (Loboda *et al* 2017). In addition, coarser spatial resolution files are formed, enabling the system to quickly access a dataset at the resolution that matches the incoming model file (e.g. an incoming model spatial resolution at 0.5° is matched to a pre-generated 0.5° benchmark product). This 'offline' matching greatly increases the speed at which the MoDIF ABS generates results. For *in situ* and fine scale airborne data, this coarsening can represent a significant scale mismatch, while models that run on a finer scale can take advantage of the finer resolution observational data. Future work may include utilizing the wealth of fine-scale data within a relatively coarse model cell to generate cell statistics/uncertainty for use in the scoring system.

3.2. ABoVE model benchmarking system

The ABS, written in Python, is an adaptation of the ILAMB system but focused specifically on indicators of ecosystem dynamics in the ABR. Its model and observational domain encompasses the ABR of North America (Loboda *et al* 2017). The ABS is also affiliated with the PBS through shared datasets, functional benchmarks, and statistical metrics associated with the permafrost objectives of ABoVE.

The process flow of the ABS, shown in figure 2, begins with a user submitting the model results of a specific version of their model outputs (either as a single NetCDF file or multiple NetCDF files) through the web upload interface, which utilizes the php language to move the model output to the ASC, wherein both the data and MoDIF software also reside (at present, only registered users to the ASC have direct access to

the ABS, necessitating an intermediary to upload model output or observational data; see section 5.3 for further discussion). Once the model output has been uploaded to the ASC, the ABS automatically begins the model evaluation. The process flow of the system code is shown in table 1 lists the, where each of the seven steps in the execution of the code has associated input elements (Column 'Input to Step') and produces elements (Column 'Output from Step') at the conclusion of the step. The full details of the process flow are outlined in the supplemental materials (S2), but a brief summary for each step is provided here. Step 1 is to load the shell of the scoring structure. The scoring structure is populated with the statistics generated by the benchmarks within the system, matched to the ABoVE Ecosystem Dynamics Objectives and Tier 2 Science questions/indicators as per the ABoVE Concise Experiment Plan (ACEP). The structure consists of nested custom Python classes that dynamically interact between data, model output, benchmarks, statistics, and summary scoring (supplemental figure (1)). Step 2 loads the model output that has been submitted by the user as a custom ModelOutput class, which includes a dictionary of variables of xarray DataArray instances. Step 3 determines which variables within the model output match to the observational benchmarks, and re-projects those model variables to the ABoVE projection and grid. Specifically, For every variable listed in the ModelOutput class, the system determines whether or not there is a matching observation within the system. If one or more matching benchmarks are found, then the model variable is re-projected over the temporal range of the benchmarks via the nearest neighbor algorithm using the pyresample Python package to the ABoVE projection and grid at a resolution that minimizes redundancy and gaps. Step 3 also generates the variable-observation pair entries in the statistics nested dictionary. Step 4 loads a version of the observational benchmark data from the system that matches the resolution of the model output into the Python
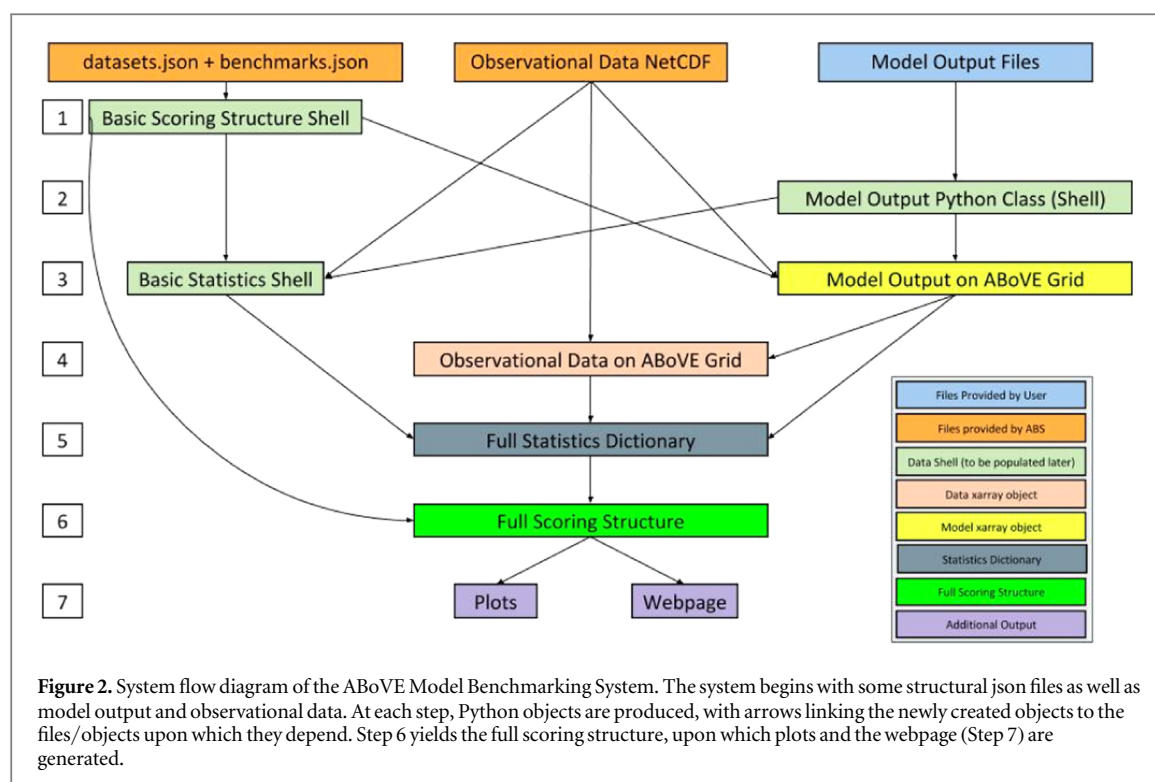
Figure 2. System flow diagram of the ABoVE Model Benchmarking System. The system begins with some structural json files as well as model output and observational data. At each step, Python objects are produced, with arrows linking the newly created objects to the files/objects upon which they depend. Step 6 yields the full scoring structure, upon which plots and the webpage (Step 7) are generated.

Table 1. Table of system flow. There are seven steps in the system. Each step relies upon input data from outside the system, system code, or the previous step. The scoring structure (Step 6) is the output of interest, while Step 7 generates supplemental information (plots and the webpage).

Step num	Step name	Input to step	Output from step
1	Load structure	datasets.json and benchmarks.json	Basic scoring structure shell
2	Load model	Model output files in active model directory	Model output class shell
3	Determine matching datasets	Observational data NetCDF file (to get temporal coverage), shell structure	Model output in ABoVE grid, statistics shell
4	Load data from matching datasets	Observational data NetCDF files or source GeoTiffs	Observational data in ABoVE grid
5	Generate statistics	ABoVE gridded model output and observational data	Full statistics dictionary
6	Make full scoring structure	Scoring structure shell, statistics dictionary	Full scoring structure
7	Generate webpage and plots	Full scoring structure, Web generation code	Plots, webpage

environment. The observational data is already in the ABoVE projection and grid. Step 5 calculates the statistics for each model-observation pair and stores them in a statistics dictionary (calculations are further detailed in section 3.3 and S2). Step 6 populates the scoring structure shell with the results from the statistics nested dictionary, which is traversed from the bottom of the tree, starting with all of the observation instances (figure S1). Step 7 generates statistical maps and the webpage which displays the scoring information.

3.3. ABS Statistics

The statistical tests included in the ABS mirror those that are found in ILAMB. These were chosen because of the efficacy of these tests as demonstrated within ILAMB. The calculations are as follows: (1) Bias—the temporal mean of each benchmark pixel is subtracted

from the temporal mean of each model output pixel. The results are normalized over the domain, the bias score map is generated from equation (1), and the domain-wide spatial mean is calculated to produce the bias score. (2) Root Mean Square Error (RMSE)— the bias between the model and benchmark is calculated for each pixel in time and space as in the Bias calculation. The results are then squared, and the temporal mean of each squared bias pixel is taken. The square root of the result yields a spatial map of RMSE, which is then normalized, the RMSE score map is generated from equation (2), and the spatial mean of that map yields the RMSE score. (3) Spatial Distribution Score—calculated from the correlation between the temporal mean of the model results and the temporal mean of the benchmark. (4) Interannual Variability—two maps (benchmark, model) of inter-annual variability are produced from the temporal

standard deviation of each spatial point within the domain. The benchmark map is subtracted from the model output map, the result is normalized over the domain, the interannual variability score map is generated from equation (3), and the domain-wide spatial mean is calculated to produce the interannual variability score. (5) Seasonal Cycle—the model results and benchmarks are binned by month (i.e. all January results are grouped together). These groupings are then averaged, and a maximum function is applied to determine the month for which the value of the given variable is the largest. The maximum month for the benchmark is subtracted from the maximum month for the model output, a seasonal cycle score map is generated from equation (4), and the spatial mean of that domain-wide map is calculated to produce the Seasonal Cycle score (this statistic shows how well the timing of the cycle of a variable from the model matches with the benchmark). Each statistical score, as well as the Overall Score, can range from 0 (model output unaligned with observations) to 1 (model output aligned with observations). The Overall Score for each model-observation pair is calculated from a weighted average of the five previously described statistical test scores, with a weight of 2 for the RMSE score and a weight of 1 for all remaining statistical scores (this weighting structure is identical to the weights used in ILAMB).

$$\text{score}_{\text{bias}_{i,j}} = e^{-|(\text{bias}_{i,j}/(\text{obs}_{i,j} - \text{obs}_{\min}))|}, \quad (1)$$

$$\text{score}_{\text{cycle}_{i,j}} = e^{-(\text{rmse}_{i,j}/\text{obs}_{i,j})}, \quad (2)$$

$$\text{score}_{\text{iav}_{i,j}} = e^{-|(\sigma_t(\text{mod}_{i,j}) - \sigma_t(\text{obs}_{i,j})/\sigma_t(\text{obs}_{i,j}))|}, \quad (3)$$

$$\text{score}_{\text{cycle}_{i,j}} = 0.5*(1 + \cos(2\pi*|\text{shift}_{i,j}|)/12). \quad (4)$$

To calculate the Overall model score, we start with the Overall Score for each model-observation pair for a given variable. The system allows each Observation instance to have an associated weight to reflect the uncertainties associated with each observation (e.g. an *in situ* observation of active layer thickness might be weighted higher than an InSAR or Polarimetric SAR observation of that same variable). At present, all observations have a weight of 1, where future work will adjust those values as uncertainties are quantified. The score for a given variable is then the weighted average of all model-observation pairs for that variable. The Overall Score of each indicator in turn is a weighted average of the variables (with the weight for each variable being the sum of the weights used in calculating that variable's Overall Score), and the Overall model score in turn is the weighted average of each indicator score.

4. User case

For illustration, we show output from a state-of-the-art land surface model, the Lund–Potsdam–Jena (LPJ) Dynamic Global Vegetation Model (Sitch *et al* 2003)

Wald Schnee und Landschaft version (LPJ-wsl), used in two different modes. The LPJ-wsl modeling team has been developing a permafrost representation in their model, and they are using the ABS to assess the improvement for the ABoVE indicators with the new permafrost coupling. Output from both versions of the model (one with permafrost coupling, one without) was run through the ABS as described in Methods.

The new permafrost module in LPJ-wsl includes a soil thermal dynamics that simulates the freeze-thaw cycle, and a dynamic snow scheme to include some of the effects of snow aging on soil thermal insulation properties (Wania *et al* 2009). A multilayer Crank–Nicolson finite difference scheme was used to model soil thermal dynamics. The model has been applied in carbon cycle applications, and the simulated permafrost extent and timing of freeze/thaw cycle have been evaluated against a variety of observations, primarily at the global scale (Zhang *et al* 2016, Zhang *et al* 2017, Zhang *et al* 2018). For the representation of permafrost coupling with the carbon cycle, the modifications are mainly through (1) adding the influence of soil moisture on soil thermal conductivity; (2) using depth-weighted average temperature for the top soil layer (0–0.5 m) to replace surface air temperature in the calculation of GPP, NPP and ecosystem respiration, and wetland methane flux; and (3) introducing water-stress effects in the calculation of GPP due to the change in availability of water content in soil during the freeze phase.

Scores were computed for 4 carbon variables as supplied by the LPJ-wsl output: Wetland Methane Flux, GPP, NPP, and Net Ecosystem Exchange. Table 2 shows the full scoring results for both model versions. Both the Overall Score and scores of each of the four variables improved slightly when permafrost coupling was included in the model. Figure 3 shows a screenshot from the web output for GPP as illustration. The bias maps within the figure show that the output from the model with the permafrost module has a reduced bias in GPP, particularly in the discontinuous permafrost zone in Canada. Figure 4 shows the other mapped statistics (interannual variability, RMSE, and seasonal cycle), similarly showing slight performance improvement in the model version with the permafrost model.

5. Future work

5.1. Update the ABS with newly available ABoVE and complementary data

The ABS already hosts a large suite of datasets, primarily from existing satellite datasets (Supplementary table 1). There is an enormous wealth of complementary data and information existing or in development by programs outside of ABoVE that are relevant to modelers (Fisher *et al* 2018b). There are
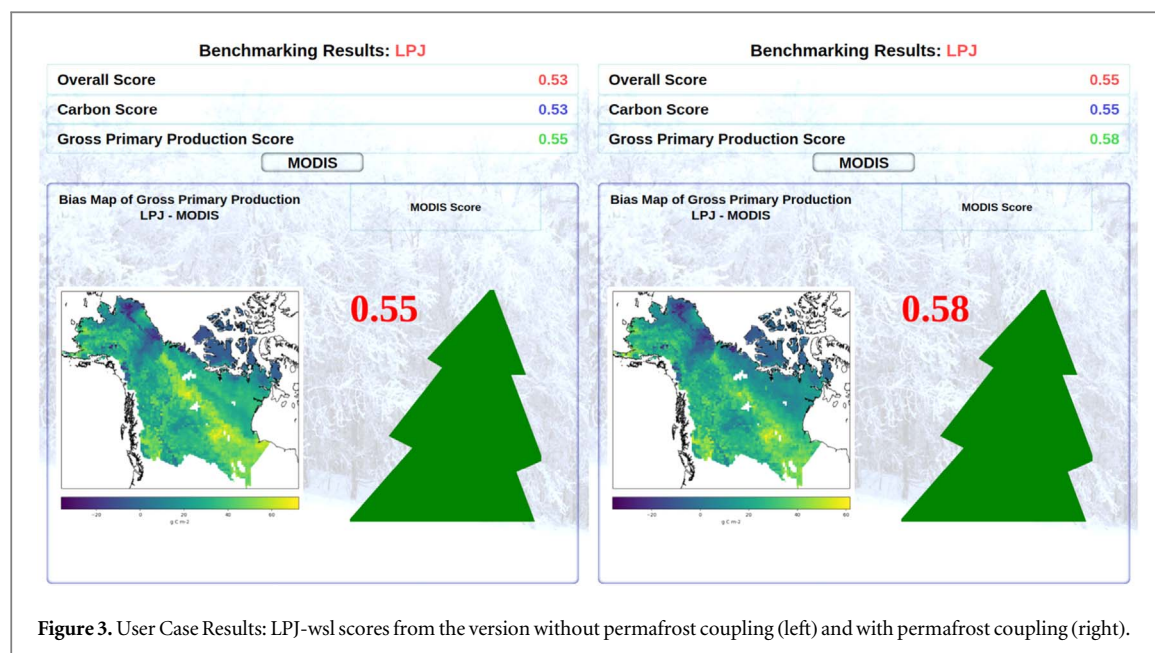
Figure 3. User Case Results: LPJ-wsl scores from the version without permafrost coupling (left) and with permafrost coupling (right).

Table 2. User Case Scoring Results—Scores for each variable in the test case: Gross Primary Production (GPP), Methane Flux, Net Primary Production (NPP) and Net Ecosystem Exchange (NEE) for each statistical scoring component: Bias, Root Mean Square Error (RMSE), Spatial Distribution Score, Interannual Variability (IAV), and the Seasonal Cycle… as well as the Overall Score for each variable, and the Overall Score for each model (the final 2 rows).

Scores	Bias	RMSE	Spatial	IAV	Seasonal	Overall
GPP—No Permafrost	0.38	0.40	0.77	0.38	0.96	0.55
GPP—Permafrost	0.49	0.39	0.81	0.46	0.95	0.58
Methane flux—No Permafrost	0.40	0.58	0.72	0.39	0.89	0.59
Methane flux—Permafrost	0.51	0.60	0.69	0.48	0.85	0.62
NPP—No Permafrost	0.58	0.47	0.16	0.53	N/A	0.44
NPP—Permafrost	0.59	0.47	0.19	0.57	N/A	0.46
NEE—No Permafrost	0.57	0.38	0.42	0.67	0.71	0.52
NEE—Permafrost	0.58	0.38	0.41	0.63	0.86	0.54
Overall—No Permafrost	N/A	N/A	N/A	N/A	N/A	0.53
Overall—Permafrost	N/A	N/A	N/A	N/A	N/A	0.55

many more datasets being made available at present and in the near future (see: daac.ornl.gov/cgi-bin/dataset_lister.pl?p=34). Future work aims to integrate these new datasets into the ABS.
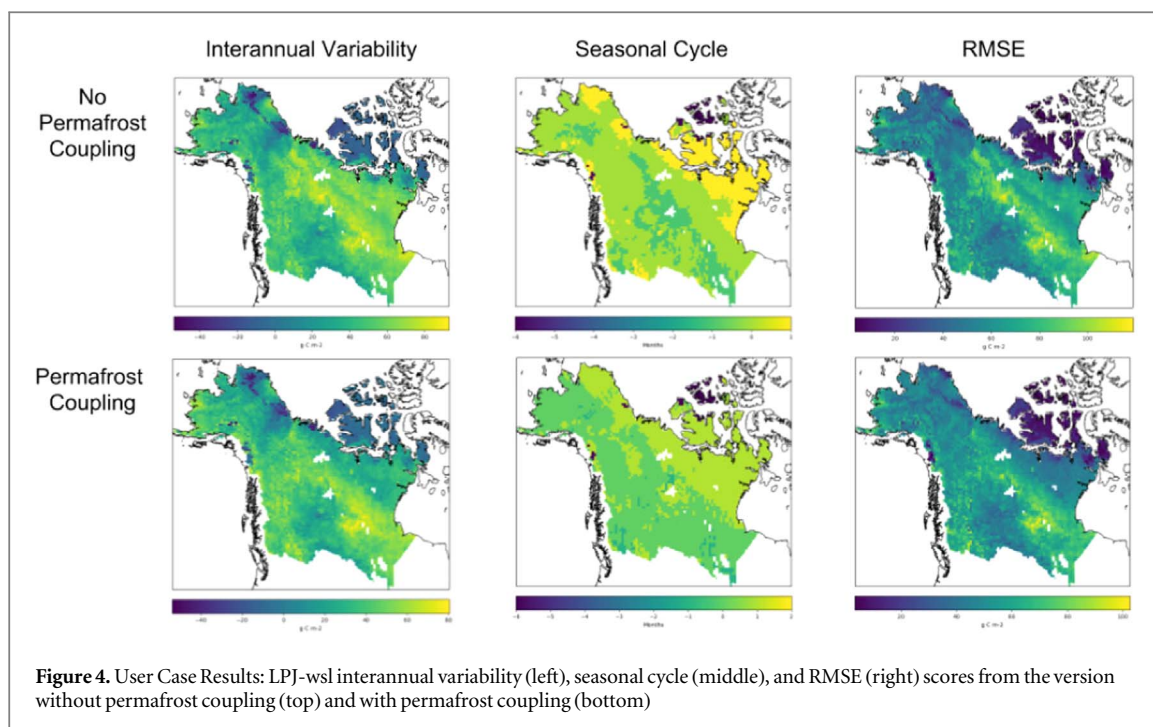
At present, the benchmark datasets included within the ABS are manually pre-processed so that they are in a suitable form for ingestion into the ABS software. It would increase future efficiency if there was a method to automate the pre-processing of a candidate dataset. This automation software would automatically recognize the spatio-temporal extent and resolution of the data, as well as the projection, and use that information to produce the ingestible form of the data via subsetting and transformations (as is currently done manually). If that automation is untenable, an alternative would be to provide a consistent method for those who are generating datasets to transform them into ingestible assets in the ABS. This would involve the dataset creator providing those same spatio-temporal properties and projection information to a piece of software that the creator can then run to transform the data into an ingestible asset.

5.2. Statistics/analysis refinement

While ILAMB forms a loose basis for the ABS system, the global nature of ILAMB necessitates a large volume of statistical metrics, whereas the ABS tailors these statistics specifically to the ABoVE Ecosystem Dynamics Objectives and Tier 2 Science questions/indicators, as detailed in the ACEP. Nonetheless, in so doing, the ABS identified key missing datasets, e.g. active layer thickness and surface water extent (Fisher *et al* 2018b), that limited holistic evaluation of the ABoVE indicators. Continued refinement of the statistical metrics will be done to satisfactorily evaluate the ABoVE objectives.

The basic statistics in this version of the ABS form the foundation on which more sophisticated metrics will performed. In addition to further refinement of appropriate scoring weights, functional benchmarks

Figure 4. User Case Results: LPJ-wsl interannual variability (left), seasonal cycle (middle), and RMSE (right) scores from the version without permafrost coupling (top) and with permafrost coupling (bottom)

will be used in model evaluation. These benchmarks are structured around similar response functions within models; e.g. temperature versus respiration. Additional functional benchmarks will be investigated, tested, and included in the proposed work to address key attributes of the ABoVE ecosystem indicators.

5.3. User engagement

The full potential of the ABS is realized only when used by the modeling and data communities. Although modelers and observationalists have made use of the ABS at time of writing, it requires an active effort to engage and assist the modeling and data communities in utilizing the system. Moreover, only registered users to the ASC currently have direct access; at present, the process for inclusion of additional observations is by passing the data through an intermediary with access to the ASC. Future work aims to streamline this process to improve and expand ABS accessibility. Engagement with the community is facilitated in part by making the existence of the ABS more widely known, and in part by surveying the needs of the modeling and data communities and incorporating those findings into the ABS (Fisher *et al* 2018b). We also aim to continue connecting ABoVE Phase I data to ABoVE Phase II modeling efforts, thereby facilitating the advancement in modeling capabilities, and reduction in uncertainties, for ecosystem dynamics in the ABR. For example, one of the science questions modelers face in the ABR is: 'What processes are controlling changes in the distribution and properties of permafrost, and what are the impacts of these changes?' The ABS will ingest new permafrost data from the ABoVE campaign, thereby allowing model

evaluation against permafrost dynamics and an improved understanding of the relationships between permafrost change and other ABR ecosystem dynamics through functional benchmarks. The ABS will operate as a central hub for organization and reporting of model improvements. Finally, through user interaction and experience, we will continue to update the design of the system interface to improve usability. In sum, the ABS acts as part of the 'central nervous system', connecting multiple elements across data and models to one another, providing a centralized reporting system for model improvements (Fisher *et al* 2018a).

5.4. Expansion to larger domain

At present, the ABS provides model evaluation across the ABR of North America. Future applications of this system may be expanded to include global areas of Arctic and Boreal ecosystems with the requisite observational data (existing global or pan-Arctic model output would not require modification in an expanded ABS).

Acknowledgments

Author contributions

JBF formulated idea; ES designed software; ES implemented software; all authors contributed to the writing of the paper.

Conflict of interest

The authors declare no conflict of interest.

ORCID iDs

Eric Stofferahn ⓘ https://orcid.org/0000-0001-6960-4193
Joshua B Fisher ⓘ https://orcid.org/0000-0003-4734-9085
Daniel J Hayes ⓘ https://orcid.org/0000-0002-3011-7934
Benjamin Poulter ⓘ https://orcid.org/0000-0002-9493-8600

References

Chapman W L and Walsh J E 2007 Simulations of Arctic temperature and pressure by global coupled models *J. Clim.* **20** 609–32

Christensen J H *et al* 2013 Climate phenomena and their relevance for future regional climate change *Climate Change 2013: The Physical Science Basis. Contribution of working group I to the fifth Assessment Report of the Intergovernmental Panel on Climate Change.* ed T F Stocker *et al* (Cambridge and New York: Cambridge University Press) pp 1217–308 ch 14

Collier N *et al* 2018 The international land model benchmarking (ILAMB) system: design, theory, and implementation *J. Advances Modeling Earth Systems* **10** 2731–54

Fisher J B, Huntzinger D N, Schwalm C, Schaefer K, Hayes D, Stofferahn E and ABoVE Modeling Working Group 2018a *Bridge to the Future: Moving into ABoVE Phase II's Modeling Focus,* paper presented at NASA ABoVE Science Team Meeting 4 (Seattle, Washington)

Fisher J B, Huntzinger D N, Schwalm C R and Sitch S 2014a Modeling the terrestrial biosphere *Annu. Rev. Environ. Resour.* **39** 91–123

Fisher J B *et al* 2014b Carbon cycle uncertainty in the alaskan Arctic *Biogeosciences* **11** 4271–88

Fisher J B *et al* 2018b Missing pieces to modeling the Arctic-Boreal puzzle *Environ. Res. Lett.* **13** 020202

Goetz S, Kimball J, Mack M and Kasischke E 2011 Scoping completed for an experiment to assess vulnerability of Arctic and Boreal ecosystems *EOS Trans. Am. Geophys. Union* **92** 150–1

Griffith P C, Goetz S J, Kasischke E S, Mack M C and Wickland D E 2012 The Arctic-Boreal vulnerability experiment: a NASA terrestrial ecology field campaign AGU Fall Meeting (San Francisco, California, 3–7 December)

Hayes D J, Kicklighter D W, McGuire A D, Chen M, Zhuang Q, Yuan F, Melillo J M and Wullschleger S D 2014 The impacts of recent permafrost thaw on land-atmosphere greenhouse gas exchange *Environ. Res. Lett.* **9** 045005

Hayes D J, McGuire A D, Kicklighter D W, Gurney K R, Burnside T J and Melillo J M 2011 Is the northern high-latitude land-based CO_2 sink weakening? *Glob. Biogeochem. Cycles* **25** GB3018

Hinzman L D *et al* 2005 Evidence and implications of recent climate change in northern alaska and other Arctic regions *Clim. Change* **72** 251–98

Hoffman F M *et al* 2017 International Land Model Benchmarking (ILAMB) 2016 *Workshop Report* DOE/SC-0186 US Department of Energy, Office of Science, Germantown, Maryland, USA (https://doi.org/10.2172/1330803)

Hugelius G, Tarnocai C, Broll G, Canadell J G, Kuhry P and Swanson D K 2013 The northern circumpolar soil carbon database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions *Earth Syst. Sci. Data* **5** 3–13

Jandl R *et al* 2014 Current status, uncertainty and future needs in soil organic carbon monitoring *Sci. Total Environ.* **468–469** 376–83

Kasischke E S, Hayes D J, Griffith P C, Larson E K and Wickland D E 2013 NASA's Arctic-Boreal vulnerability experiment: a large-scale study of environmental change in Western North America and its implications for ecological systems and society 2013 Fall Meeting (AGU, San Francisco, California, 9–13 December)

Koven C D, Ringeval B, Friedlingstein P, Ciais P, Cadule P, Khvorostyanov D, Krinner G and Tarnocai C 2011 Permafrost carbon-climate feedbacks accelerate global warming *Proc. Natl Acad. Sci.* **108** 14769–74

Loboda T V, Hoy E E and Carroll M L 2017 *ABoVE: Study Domain and Standard Reference Grids* (Oak Ridge, Tennessee: ORNL DAAC) (https://doi.org/10.3334/ORNLDAAC/1367)

Luo Y Q *et al* 2012 A framework for benchmarking land models *Biogeosciences* **9** 3857–74

Mack M C, Schuur E A G, Bret-Harte M S, Shaver G R and Chapin F S 2004 Ecosystem carbon storage in Arctic tundra reduced by long-term nutrient fertilization *Nature* **431** 440–3

McGuire A D, Anderson L G, Christensen T R, Dallimore S, Guo L, Hayes D J, Heimann M, Lorenson T D, Macdonald R W and Roulet N 2009 Sensitivity of the carbon cycle in the Arctic to climate change *Ecol. Monogr.* **79** 523–55

McGuire A D, Chapin F S III, Walsh J E and Wirth C 2006 Integrated regional changes in Arctic climate feedbacks: implications for the global climate system *Annu. Rev. Environ. Resour.* **31** 61–91

McGuire A D *et al* 2012 An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions *Biogeosciences* **9** 3185–204

McGuire A D *et al* 2018 The dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change *Proc. of the National Academy of Sciences* (April 2018) **115** 3882–7

Melton J R *et al* 2013 Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP) *Biogeosciences* **10** 753–88

Natali S M, Schuur E A G and Rubin R L 2012 Increased plant productivity in Alaskan tundra as a result of experimental warming of soil and permafrost *J. Ecol.* **100** 488–98

Nave L, Johnson K, Van Ingen C, Agarwal D, Humphrey M and Beekwilder N 2016 International Soil Carbon Network (ISCN) Database v3-1 (https://doi.org/10.17040/ISCN/1305039)

Oechel W C, Hastings S J, Vourlitis G, Jenkins M, Riechers G and Grulke N 1993 Recent change of Arctic tundra ecosystems from a net carbon dioxide sink to a source *Nature* **361** 520–3

Oechel W C, Laskowski C A, Burba G, Gioli B and Kalhori A A M 2014 Annual patterns and budget of CO_2 flux in an Arctic tussock tundra ecosystem *J. Geophys. Res. Biogeosci.* **119** 2013JG002431

Schädel C, McGuire A D and Schuur E A G 2015 *Permafrost Carbon Network: 5-Year Synthesis Report*

Schädel C, Schuur E A, Bracho R, Elberling B, Knoblauch C, Lee H, Luo Y, Shaver G R and Turetsky M R 2014 Circumpolar assessment of permafrost C quality and its vulnerability over time using long-term incubation data *Glob. Change Biol.* **20** 641–52

Schuur E A, Vogel J G, Crummer K G, Lee H, Sickman J O and Osterkamp T E 2009 The effect of permafrost thaw on old carbon release and net carbon exchange from tundra *Nature* **459** 556–9

Schuur E A G *et al* 2015 Climate change and the permafrost carbon feedback *Nature* **520** 171–9

Serreze M C, Walsh J E, Chapin F S, Osterkamp T, Dyurgerov M, Romanovsky V, Oechel W C, Morison J, Zhang T and Barry R G 2000 Observational evidence of recent change in the northern high-latitude environment *Clim. Change* **46** 159–207

Sitch S, McGuire A D, Kimball J, Gedney N, Gamon J, Engstrom R, Wolf A, Zhuang Q, Clein J and McDonald K C 2007 Assessing the carbon balance of circumpolar Arctic tundra using remote sensing and process modeling *Ecol. Appl.* **17** 213–34

Sitch S *et al* 2003 Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model *Glob. Change Biol.* **9** 161–85

Stofferahn E, Fisher J B, Hayes D J, Huntzinger D N and Schwalm C R 2016 How well does your model capture the terrestrial ecosystem dynamics of the Arctic-Boreal Region? 2016 Fall Meeting *(AGU, San Francisco, California, 12–16 December)*

Wania R, Ross I and Prentice C I 2009 Integrating peatlands and permafrost into a dynamic global vegetation model: I. Evaluation and sensitivity of physical land surface processes *Glob. Biogeochem. Cycles* **23** GB3014

Zhang Z, Zimmermann E N, Calle L, Hurtt G C A and Poulter B 2018 Enhanced response of global wetland methane emissions to the 2015–2016 El Niño-Southern oscillation event *Environ. Res. Lett.* **13** 074009

Zhang Z, Zimmermann N E, Kaplan J O and Poulter B 2016 Modeling spatiotemporal dynamics of global wetlands: comprehensive evaluation of a new sub-grid TOPMODEL parameterization and uncertainties *Biogeosciences* **13** 1387–408

Zhang Z, Zimmermann N E, Stenke A, Li X, Hodson E L, Zhu G, Huang C and Poulter B 2017 Emerging role of wetland methane emissions in driving 21st century climate change *Proc. Natl Acad. Sci.* **114** 9647–52