



LETTER

OPEN ACCESS

RECEIVED

10 September 2019

REVISED

23 December 2019

ACCEPTED FOR PUBLICATION

3 January 2020

PUBLISHED

7 February 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Evaluation of simulated soil carbon dynamics in Arctic-Boreal ecosystems

D N Huntzinger^{1,14}, K Schaefer², C Schwalm^{3,4}, J B Fisher⁵, D Hayes⁶, E Stofferahn⁵, J Carey⁷, A M Michalak⁸, Y Wei⁹, A K Jain¹⁰, H Kolus¹, J Mao⁹, B Poulter¹¹, X Shi⁹, J Tang¹² and H Tian¹³

¹ School of Earth and Sustainability, Northern Arizona University, PO Box 5694, Flagstaff AZ, United States of America

² National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, United States of America

³ Woods Hole Research Center, Falmouth MA, United States of America

⁴ Center for Ecosystems and Society, Northern Arizona University, Flagstaff, AZ, United States of America

⁵ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States of America

⁶ School of Forest Resources, University of Maine, Orono, ME, United States of America

⁷ Math and Science, Babson College, Babson Park, MA United States of America

⁸ Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, United States of America

⁹ Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

¹⁰ Department of Atmospheric Sciences, University of Illinois, Urbana, IL, United States of America

¹¹ National Aeronautics and Space Administration (NASA), Biospheric Sciences Lab, Greenbelt, MD, United States of America

¹² Marine Biology Laboratory, Woods Hole, MA, United States of America

¹³ School of Forestry and Wildlife Sciences, Auburn University, Auburn AL, United States of America

¹⁴ Author to whom any correspondence should be addressed.

E-mail: deborah.huntzinger@nau.edu

Keywords: soil carbon, Arctic-Boreal, terrestrial carbon cycle, soil respiration, functional benchmark

Supplementary material for this article is available [online](#)

Abstract

Given the magnitude of soil carbon stocks in northern ecosystems, and the vulnerability of these stocks to climate warming, land surface models must accurately represent soil carbon dynamics in these regions. We evaluate soil carbon stocks and turnover rates, and the relationship between soil carbon loss with soil temperature and moisture, from an ensemble of eleven global land surface models. We focus on the region of NASA's Arctic-Boreal vulnerability experiment (ABoVE) in North America to inform data collection and model development efforts. Models exhibit an order of magnitude difference in estimates of current total soil carbon stocks, generally under- or overestimating the size of current soil carbon stocks by greater than 50 PgC. We find that a model's soil carbon stock at steady-state in 1901 is the prime driver of its soil carbon stock a hundred years later—overwhelming the effect of environmental forcing factors like climate. The greatest divergence between modeled and observed soil carbon stocks is in regions dominated by peat and permafrost soils, suggesting that models are failing to capture the frozen soil carbon dynamics of permafrost regions. Using a set of functional benchmarks to test the simulated relationship of soil respiration to both soil temperature and moisture, we find that although models capture the observed shape of the soil moisture response of respiration, almost half of the models examined show temperature sensitivities, or Q10 values, that are half of observed. Significantly, models that perform better against observational constraints of respiration or carbon stock size do not necessarily perform well in terms of their functional response to key climatic factors like changing temperature. This suggests that models may be arriving at the right result, but for the wrong reason. The results of this work can help to bridge the gap between data and models by both pointing to the need to constrain initial carbon pool sizes, as well as highlighting the importance of incorporating functional benchmarks into ongoing, mechanistic modeling activities such as those included in ABoVE.

1. Introduction

The fastest rates of climate warming are occurring in the high northern latitudes (AMAP 2017, USGCRP 2017). Warming temperatures and rising atmospheric CO₂ could benefit plants by increasing plant productivity (Qian *et al* 2010, Natali *et al* 2012), accelerating nutrient cycling (Hobbie *et al* 2002, Mack *et al* 2004), and lengthening the growing season (Zeng *et al* 2011). However, it is likely that warming temperatures will also lead to more rapid rates of soil respiration (e.g. Hayes *et al* 2011, Koven *et al* 2017) and more extensive permafrost thaw (Schuur *et al* 2013, Hayes *et al* 2014, Hugelius *et al* 2014, Schuur *et al* 2015); both of which could feedback to further accelerate warming through the release of CO₂ and CH₄ to the atmosphere (Bond-Lamberty and Thomson 2010b, Schaefer *et al* 2011, Schuur *et al* 2015). Given the magnitude of soil carbon stocks at high latitudes (Hugelius *et al* 2014), and the potential vulnerability of these stocks to climate warming (Harden *et al* 2012, Schadel *et al* 2014, Crowther *et al* 2015, Phillips *et al* 2017), robust future climate projections require that global land surface models accurately represent soil carbon dynamics in high-latitude regions (Koven *et al* 2017), particularly under rapidly changing environmental conditions (Tang *et al* 2019).

Theoretically, soil carbon dynamics can be predicted given knowledge of the size of initial carbon pools stocks, carbon input rates, residence time of carbon in soil pools, and the sensitivity of stored carbon to environmental factors (Fisher *et al* 2014a, Luo *et al* 2015). However, results from previous evaluation studies show widely different estimates of both stocks and climate-carbon feedbacks across models (Todd-Brown *et al* 2013, Fisher *et al* 2014b, Tian *et al* 2015, McGuire *et al* 2016). This variability in model estimates has not, to date, been well constrained by conventional benchmarks (Luo *et al* 2016). A key challenge in model benchmarking is confronting models with observations that not only tell us whether models produce the right *endpoints*, such as magnitude of soil carbon pools or gross primary productivity (GPP), but also if they simulate the correct *pathway(s)* to those endpoints, such as the response of soil respiration to climate warming (Huntzinger *et al* 2017). Endpoints are critical for robust predictions of how much carbon is (or has the potential to be) stored within a given ecosystem. Pathways are crucial for predicting the vulnerability of stored carbon and ensuring the integrity of future projections of carbon fluxes under varied environmental conditions.

The use of observational data to evaluate model performance is an ongoing challenge due to the spatial and temporal mismatch between models and measurements, as well as the lack of concurrence between what is measured and what is modeled (Hayes and Turner 2012, McGuire *et al* 2012, Hoffman *et al* 2017,

Collier *et al* 2018). Another challenge is the lack of reported uncertainties on many observational data products, since the choice of the observational data product has as much or more influence on inferred model skill as the model itself (Schwalm *et al* 2015). Given these ongoing (and unresolved) challenges, the central questions remain as to (1) what can we learn about a model's ability to represent soil carbon stocks and losses in high-latitude regions using existing data products, and (2) whether current data products are sufficient to identify the largest sources of uncertainty in predicting soil carbon dynamics.

In this analysis, we focus on evaluating model-simulated soil carbon stocks and turnover, and the relationship between respiration and both soil temperature and moisture in the Arctic-Boreal region (ABR). Can models simulate reasonable soil storage and losses through heterotrophic respiration within the ABR? And, do modeled simulated soil carbon dynamics match the temperature and soil moisture responses obtained from observations? Combined, this information has the potential to: (1) provide a roadmap that modelers can use to reduce uncertainty in their predictions of terrestrial C cycle dynamics, not just within the high-latitude regions, but globally or in other vulnerable regions (Lenton *et al* 2008); and (2) identify the types of observationally-based data products that are needed in order to best support model evaluation and development moving forward.

2. Data and methods

2.1. Study domain

This work focuses on soil carbon dynamics within the ABR. Specifically, we focus on the region within NASA's Arctic-Boreal Vulnerability Experiment (ABoVE; figure S1 is available online at stacks.iop.org/ERL/15/025005/mmedia). ABoVE is a NASA campaign in Alaska and Western Canada that started in 2015 to study the response of Arctic and boreal ecosystems to environmental change. The ABoVE activity is divided into three phases, with the first two phases focused primarily on intensive airborne, satellite, and *in situ* data collection, and phase 3 focused more on analysis and synthesis. Modeling activities are included in all phases (Fisher *et al* 2018), ranging from initial benchmarking (Stofferahn *et al* 2019) with existing data in Phase 1 to integrated modeling (diagnosis and prediction) with ABoVE data in Phase 3. The work presented here was conducted during the first phase of ABoVE, and thus focuses on the initial benchmarking of process-based models using existing datasets.

3. Model ensemble

We use an ensemble of eleven global land surface models from the Multi-scale Synthesis and Terrestrial

Table 1. Model estimates of total soil carbon (TSC), soil carbon residence time (ResT), heterotrophic respiration (Rh), and net primary productivity (NPP) for the full ABoVE domain and for only those regions dominated (>50%) by soils other than peat and permafrost. Also shown are the multi-model ensemble (MME) minimum (min), maximum (max), and median (med) values and a corresponding estimate from the observational constraint.

Model	Full ABoVE Domain				Non-permafrost and peatland dominated regions			
	TSC (PgC)	ResT (years)	Rh (PgC yr ⁻¹)	NPP (PgC yr ⁻¹)	TSC (PgC)	ResT (years)	Rh (PgC yr ⁻¹)	NPP (PgC yr ⁻¹)
CENTURY	23.68	49.34	0.55	0.55	6.03	29.81	0.33	0.33
CLM4.0	22.46	64.66	0.70	1.11	11.21	55.32	0.44	0.73
CLM4VIC	10.40	41.91	0.45	0.78	5.14	33.03	0.29	0.51
DLEM	39.85	65.98	0.75	0.79	22.78	73.95	0.46	0.49
HYLAND	25.95	2,501.67	0.41	0.43	11.59	90.58	0.30	0.31
ISAM	106.37	141.82	1.10	1.15	29.52	78.45	0.57	0.61
LPJ-wsl	162.68	116.39	2.13	2.68	60.16	88.60	1.03	1.31
ORCHIDEE	60.35	49.55	1.53	1.75	22.18	46.69	0.74	0.85
SIBCASA	32.21	29.57	1.50	1.93	11.91	26.46	0.73	0.93
TEM6	85.69	203.90	0.84	1.01	34.10	126.29	0.53	0.62
VEGAS2.0	72.06	91.28	1.23	1.31	29.11	76.49	0.63	0.66
MME min	10.40	29.57	0.41	0.43	5.14	26.46	0.29	0.31
MME max	162.68	2501.67	2.13	2.68	60.16	126.29	1.03	1.31
MME med	39.85	65.98	0.84	1.11	22.18	73.95	0.53	0.62
Constraint ^a	84.59	57.92	1.43	1.60	20.79	40.52	0.69	0.87

^a Observational constraints are as follows: Total soil carbon (TSC)—Northern Circumpolar Soil Carbon Database (NCSCD, Hugelius *et al* 2013a, 2013b), Heterotrophic respiration (Rh)—Hashimoto *et al* (2015), Soil carbon residence time (ResT)—TSC from NCSCD divided by Rh estimates from Hashimoto *et al* (2015). Net primary productivity (NPP)—MODIS NPP product (Zhao *et al* 2005).

Model Intercomparison Project (MsTMIP; (Huntzinger *et al* 2013). We also use MsTMIP's series of sensitivity simulations (table S1) to attribute changes in historical soil carbon storage and loss within the ABoVE domain (Loboda *et al* 2017) to key physical and biogeochemical drivers over the time period from 1901 through 2010. All models produce monthly output at half-degree spatial resolution from common forcing data for both spin-up and transient simulations, but differ in their representation and parameterization of soil C dynamics (Huntzinger *et al* 2014) (table S2). Therefore, each model can be viewed as a different realization of soil carbon uptake, loss, and storage within the ABoVE domain. MsTMIP models are run using a common protocol, where the environmental forcing data and sensitivity simulations are uniform across the ensemble (Huntzinger *et al* 2013, Wei *et al* 2014b). The MsTMIP sensitivity simulations are a set of semi-factorial runs where four time-varying drivers (climate, atmospheric CO₂, land-cover change, and nitrogen deposition) are sequentially turned-on (table S2) resulting in a set of five global simulations. Using these sensitivity simulations, we are able to quantify the relative contribution of each environmental driver to modeled changes in the soil carbon dynamics. There are eighteen models in the Version 2.0 release of Phase I MsTMIP (Huntzinger *et al* 2020). However, only eleven models met the following three criteria and are included in the analysis: (1) simulate the rate of soil carbon loss (heterotrophic respiration, Rh), soil carbon storage (total soil carbon (TSC)), and the rate of soil carbon inputs (net primary productivity, NPP); (2) submitted estimates for all sensitivity

simulations (Huntzinger *et al* 2013) (RG1-SG3 for C-only models and RG1-BG1 for C–N models; table 1); and (3) modeled output meets the criteria for carbon closure or mass balance across carbon fluxes, e.g. net ecosystem productivity (NEP) = photosynthesis – respiration, and total ecosystem respiration = heterotrophic respiration (Rh) + autotrophic respiration (Ra).

3.1. Analysis approach

We leverage the sensitivity simulations of the MsTMIP activity with a tiered benchmarking approach to evaluate simulated soil carbon stocks and residence time (i.e. turnover rate), and the relationship between simulated soil carbon loss, temperature, and soil moisture within the ABoVE domain (figure S1). We evaluate models in terms of: (1) large-scale state estimates (e.g. magnitude of simulated soil carbon stocks); (2) the sensitivity of modeled soil carbon stocks, inputs and losses to environmental forcing factors like climate and atmospheric CO₂; and (3) simulated functional relationships and emergent properties related to changing environmental conditions. The evaluation of large-scale state estimates provides an assessment of how well models simulate contemporary TSC stocks within the ABoVE domain. Simulation differencing (e.g. SG2 minus SG1; table S1) allows us to quantify and compare the influence of individual environmental forcing factors on model estimates of soil carbon stock size and inputs (through net primary production or NPP) and losses (through heterotrophic respiration or Rh) over the 110-yr simulation period in the ABoVE Domain. Finally, we

test each model's relationship of transient soil carbon loss as a function of both temperature and soil moisture across a range of temperature and soil moisture values. The benefit of 'functional benchmarks' is that they can provide more insight into the potential predictive power of a model (Hoffman *et al* 2017, Hall *et al* 2019) rather than a model's estimate of soil carbon stocks or Rh alone. The use of functional benchmarks in land carbon modeling is promising; and as long as the observations are concurrent (taken at the same location in space and time), enables the extrapolation of observations beyond sparse study sites (Fisher *et al* 2018). Functional relationships have been used to evaluate simulated above-ground productivity with changing evapotranspiration, and ultimately led to a 50% reduction in model spread in estimates of future productivity (Mystakidis *et al* 2016).

3.2. Model benchmarking

3.2.1. Gridded observationally-based state and flux products

In this study, the size of model-derived soil carbon stocks in the ABoVE Domain are evaluated against estimates from the northern circumpolar soil carbon database (NCSCD) (Hugelius *et al* 2013a, 2013b). The NCSCD is an observationally-constrained database of organic soil carbon storage in the northern circumpolar permafrost region, and contains estimates of contemporary (e.g. year ~2000) soil carbon stocks at multiple gridded spatial resolutions to a depth of 3 m (Hugelius *et al* 2013a). Models represent soil C dynamics to depths ranging from 0.3 to 3 m, with some models having variable soil carbon depth across gridcells (table S2). However, models did not report soil carbon output within specified depth ranges (e.g. 0–1 m, 1–2 m depth). Therefore, to maintain consistency with published datasets and other model-data comparisons (e.g. Todd-Brown *et al* 2013, Tian *et al* 2015, Koven *et al* 2017), we assume simulated soil carbon is contained within the top 1 meter and we use the 0.5° gridded NCSCD product of soil carbon from 0 to 1 m depth, clipped to the ABoVE Domain. Both the modeled and NCSCD estimates of soil carbon are aggregated spatially over the full domain using area-weighting. A significant portion of the ABoVE domain is covered by continuous and discontinuous permafrost. Although there are increasing efforts to improve process representation in models (Luo *et al* 2015), peatlands and permafrost carbon dynamics are not explicitly included in many global scale land carbon and Earth system models (Limpens *et al* 2008, Koven *et al* 2013a, Tian *et al* 2015). Therefore, we aggregate soil carbon stocks over: (1) the full ABoVE domain; (2) permafrost/peatland dominated soils; and (3) non-permafrost-peatland soils. We define permafrost and peatland regions as those regions covered by 50% or greater histels, histosols, gelisols, or orthel soil classes

based on the thematic classification in the NCSCD database (figure S2). We designate non-permafrost regions as those cells dominated (50% or greater) by other soils.

In addition to TSC, we also evaluate the magnitude of simulated component fluxes leading to the carbon input (i.e. NPP) and loss (i.e. Rh) from soil carbon pools against available, observationally-based, gridded data products. We use annual estimates from the MODIS NPP product (Zhao *et al* 2005, Zhao and Running 2010) and annual Rh estimates from a study by Hashimoto *et al* (2015). NPP is derived from Collection 5.5 of the MOD17 dataset (Zhao *et al* 2005) available from the Numerical Terradynamic Simulation Group (NTSG) (<http://ntsg.umd.edu>). The NPP was re-projected from 1 km resolution to a 0.5° grid to be consistent with modeled output. Rh estimates, taken from Hashimoto *et al* (2015), are derived from soil respiration (Rs) data (Bond-Lamberty and Thomson 2010a), an updated climate-driven model of Rs (Raich and Potter 1995, Raich *et al* 2002), and an empirical relationship based on meta-analysis (Bond-Lamberty *et al* 2004). We compare annual model estimates to these two data sets over the time period between 2000 and 2010, which is coincident with both observationally-constrained gridded products. Both the modeled and observationally-derived flux estimates are also aggregated spatially over the full ABoVE Domain using area-weighting.

3.2.2. Derived benchmarks

We evaluate the emergent or integrative behavior of soil carbon stocks and losses by computing an inferred soil carbon residence time for each model. We use the term 'inferred' here to acknowledge that each model has a different soil carbon pool structure (table S2). The inferred soil carbon residence time represents an approximate integrated turnover time for each model across its various soil carbon pools. At steady-state, carbon losses should equal inputs and the size of carbon pools should be constant with time. In order to calculate residence time, we assume quasi steady-state conditions in each decade of the simulations; a similar approach has been employed by others (Jeong *et al* 2018). The inferred quasi steady-state soil carbon residence time for each decade and each model is determined by the ratio of simulated decadal mean soil carbon stocks to decadal mean Rh for each gridcell. To reduce the impact of outliers, we use the median across all land cells to compute the inferred residence time for each model across the full ABoVE domain. Because this inferred residence time is computed by decade, we examine how simulated soil carbon turnover rates change over the 110 year simulation period. Through simulation differencing, we attribute changes in inferred residence time to key environmental forcing factors. To construct an observational constraint on inferred soil carbon residence time for the last decade of the simulations (i.e. 2000–2010), we use the gridded

soil carbon stocks reported between 0 and 1 m by the NCSCD, along with the gridded annual Rh from Hashimoto *et al* (2015) using the same approach as described for the models.

Spatial variability in soil or heterotrophic respiration is modulated by differences in vegetation cover, root distribution and depth, biological activity, temperature, and variations in soil characteristics, including soil moisture, texture, and geochemistry. Evaluating model representation of many of these factors is difficult due to the lack of concurrent measurements. In addition, models vary considerably in their treatment (inclusion/exclusion) of many of these effects. However, most models include climatic controls (e.g. temperature and moisture) on soil carbon decomposition. Therefore, we focus on observation-informed functional benchmarks of soil carbon loss with changing soil temperature and moisture to examine the modeled pathways to model endpoints. Models treat temperature and soil moisture effects separately using independent scaling factors; therefore it is appropriate to evaluate modeled functional responses of respiration with temperature and soil moisture separately. Field observations of soil respiration afford a dynamic view of respiration in response to changing environmental conditions including temperature and soil moisture. Soil respiration (Rs) is the product of both respiration by roots (part of Ra) and microbial decomposition of soil organic matter (Rh). However, Rs is not an output easily produced, nor commonly simulated, by models (Fisher *et al* 2014a, Phillips *et al* 2017). Rather, most models report the component fluxes Rh and Ra (which includes both above- and below-ground maintenance respiration). Here, we use direct measurements of Rs, soil temperature, and soil moisture (reported as volumetric water content) from both control and warming plots of experimental warming studies synthesized by Carey *et al* (2016a, 2016b) to provide observationally-based functional response curves of respiration with both temperature and soil moisture. We focus on observations from boreal forests and northern shrubland ecosystems, which include 810 individual data points from seven sites (latitude range 46.7°–63.9° N; Carey *et al* 2016a, 2016b). We chose the dataset synthesized by Carey *et al* (2016a, 2016b) for two key reasons: (1) it includes measurements in key ecosystem types found within the ABoVE domain; and (2) the measurements of temperature, moisture, and respiration are concurrent (i.e. taken at the same location in space and nearly simultaneously).

3.2.3. Respiration—temperature response

To create functional relationships of Rh with temperature for each model, we first construct Rh versus temperature curves for each grid cell within the ABoVE domain using monthly Rh output for each model from the climate only simulation (SG1; table S1) for the time period between 2000 and 2010. Most

models did not report soil temperature, and the number of soil layers and the thickness of each soil layer varies considerably across models (table S2). Therefore, we use monthly near-surface air temperature taken from the MsTMIP environmental driver data (Wei *et al* 2014a) and derived from the CRU-NCEP climate re-analysis data as a proxy for soil temperature. Since the Carey *et al* (2016b) dataset is not representative of the entire ABoVE domain (i.e. does not include regions underlain with continuous permafrost; or such tundra or taiga ecosystems) and only reports respiration values at temperatures above freezing, we restrict the comparison to modeled respiration values associated with temperatures greater than 0 °C and in non-permafrost or peatland dominated grid cells in boreal and northern shrubland ecosystem regions as defined by the MsTMIP environmental driver data; (Wei *et al* 2014a). To isolate the shape of the functional response curve and factor out the influence of modeled soil carbon stocks size on the magnitude of respiration, we normalize the respiration response of both modeled and observations by dividing by the magnitude of respiration at 0 °C. We fit an exponential model to respiration as a function of temperature over the temperature range of 0 °C–20 °C using equation (1), where R_T is respiration at a given temperature (T) and γ_1 and γ_2 are fitted parameters

$$R_T = \gamma_1 \times \exp^{\gamma_2 T}. \quad (1)$$

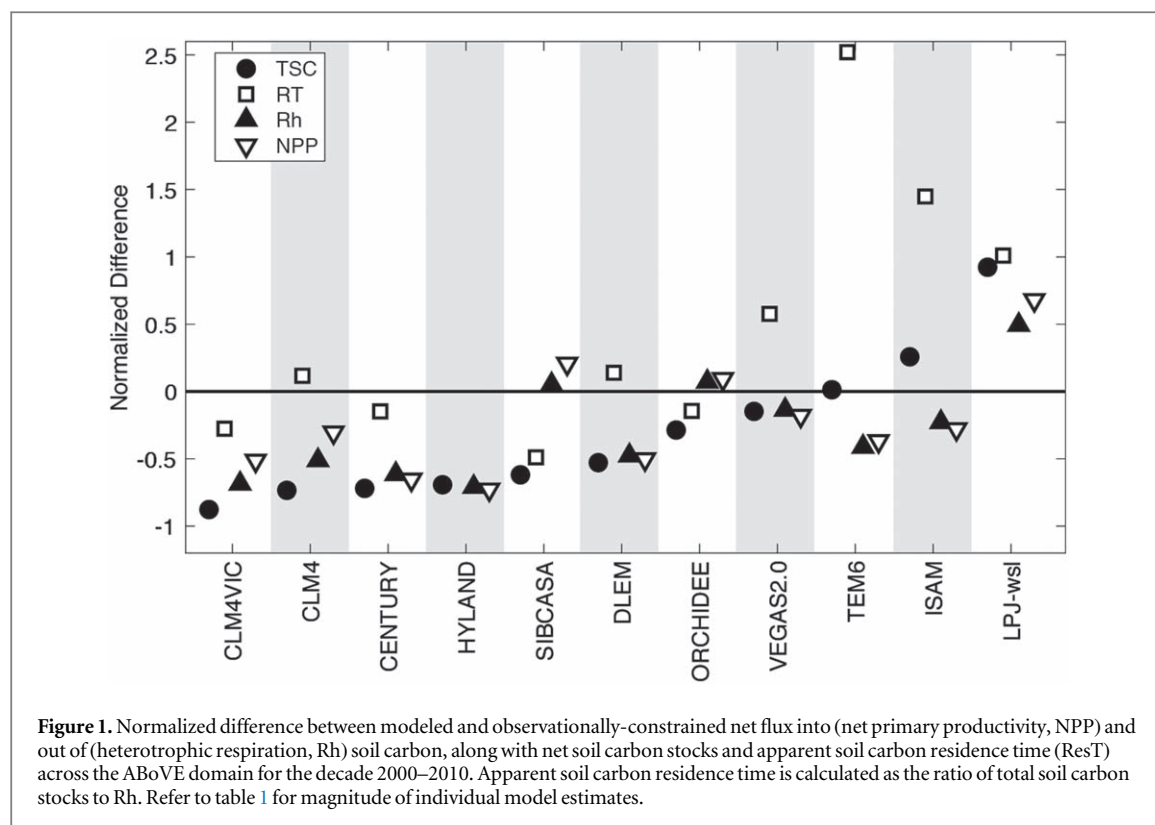
We extract an inferred temperature sensitivity of respiration, defined as the increase in soil respiration per 10 °C in temperature (or Q10), for both the modeled and observed curves for boreal forests and northern shrubland (separately and combined) using equation (2)

$$Q_{10} = \exp^{10 \times \gamma_2}. \quad (2)$$

We compute the median ‘inferred’ Q10 for each model and compare it to the observationally-constrained values derived from Carey *et al* (2016a, 2016b). The term inferred Q10 is used here to acknowledge that we are using air temperature rather than soil temperature in equation (1), and that the respiration versus temperature curves include non-temperature factors (such as soil moisture) that could influence respiration and by extension the inferred Q10 values for each model.

3.2.4. Respiration—soil moisture response

We use the soil respiration and soil moisture measurements reported by Carey *et al* (2016a, 2016b) to evaluate modeled functional response of respiration with changing soil moisture conditions. However, evaluating functional response curves of respiration with soil moisture presents several challenges. First, soil moisture is a prognostic variable rather than prescribed input in models. As such, it depends on model-specific treatments of evaporation, runoff, and soil parameters such as porosity and soil layer depth



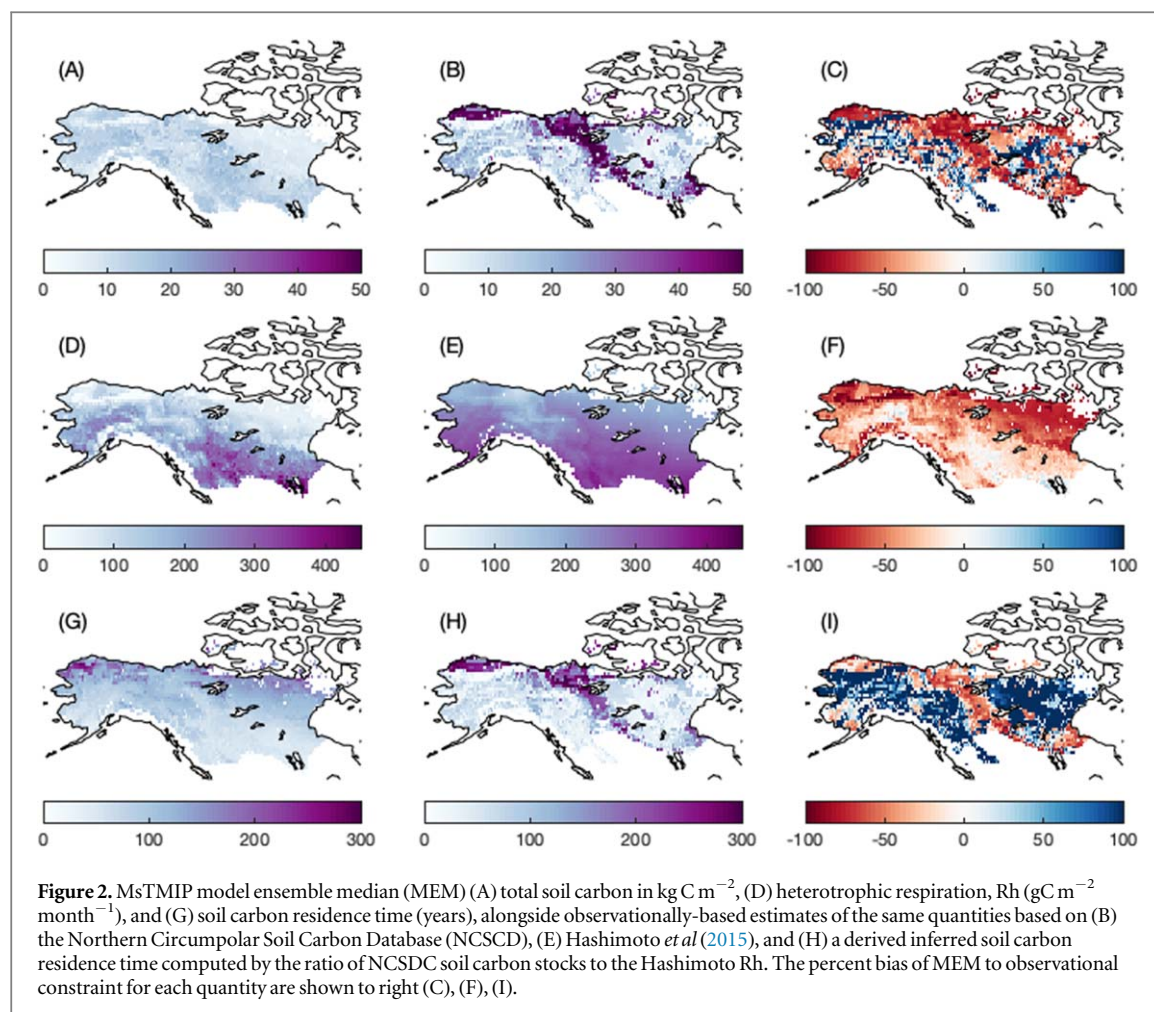
(Koster *et al* 2009). Second, only five (CENTURY, CLM, CLM4VIC, LPJ-wsl, and TEM6) of the eleven models reported soil moisture. Therefore, the analysis is limited to only a subset of the MsTMIP ensemble. Third, models typically report soil moisture in units of mass per volume of soil or kg m^{-3} . To compare with the observational constraint, we convert the soil moisture reported by models into volumetric water content (VWC)—a dimensionless quantity—by dividing reported soil moisture by the density of water (1000 kg m^{-3}) and the thickness of each model's top soil layer (in meters). For the model TEM6, which has variable soil layer thickness by grid cell, we used the TEM6's reported active layer thickness as an estimate of the thickness of the upper most soil layer in the model.

The influence of soil moisture on Rs is more complex than with temperature (Tang and Baldocchi 2005). Respiration tends to increase with increasing soil moisture until some critical soil moisture value is reached. Once the soil moisture exceeds this optimal value, respiration tends to decrease within further increases in soil moisture (Tang and Baldocchi 2005). We are interested in comparing this optimal volumetric water content across models and between models and observations. To do so, we extracted grid-cell monthly Rh and derived VWC for the last decade of simulations (2000–2010) for the climate-only simulation (SG1). We then bin respiration by VWC, calculate the median Rh per bin, and normalized the respiration values by the maximum median Rh across all bins. The same process is followed to construct a

respiration with VWC curve based on observational data for Carey *et al* (2016a, 2016b) using measurements from control-only plots, as well as both control and warming plots. We varied the bin width to assess the impact of bin width choice on the analysis. To be consistent with the observational data and the respiration-temperature analysis described above, we restricted the analysis to gridcells in boreal and northern shrubland ecosystems and in non-permafrost or peatland dominated regions.

4. Results and discussion

Models exhibit an order of magnitude difference in estimates of current TSC stocks within the ABoVE domain (table 1 and figures S4, S5). The observationally-based estimate of soil carbon from the NCSCD falls within the spread of model results (table 1). However, model performance against the NCSCD benchmark is relatively poor, with over half of the models underpredicted or overpredicting by more than 50 PgC the amount of soil carbon stored within the top 1 m of soil across the full ABoVE domain. Disagreement between the models and the NCSCD is greatest at the North Slope, the Mackenzie basin, and the Hudson Bay peatlands (figures 2(a)–(c), S5), where carbon burial rates by cryoturbation or peat development are the highest (Tarnocai *et al* 2009). For most models in the ensemble, performance against the NCSCD improves significantly if permafrost and peatland dominated regions are removed (figures 1 and S4 and S6). This is not surprising, since most models do

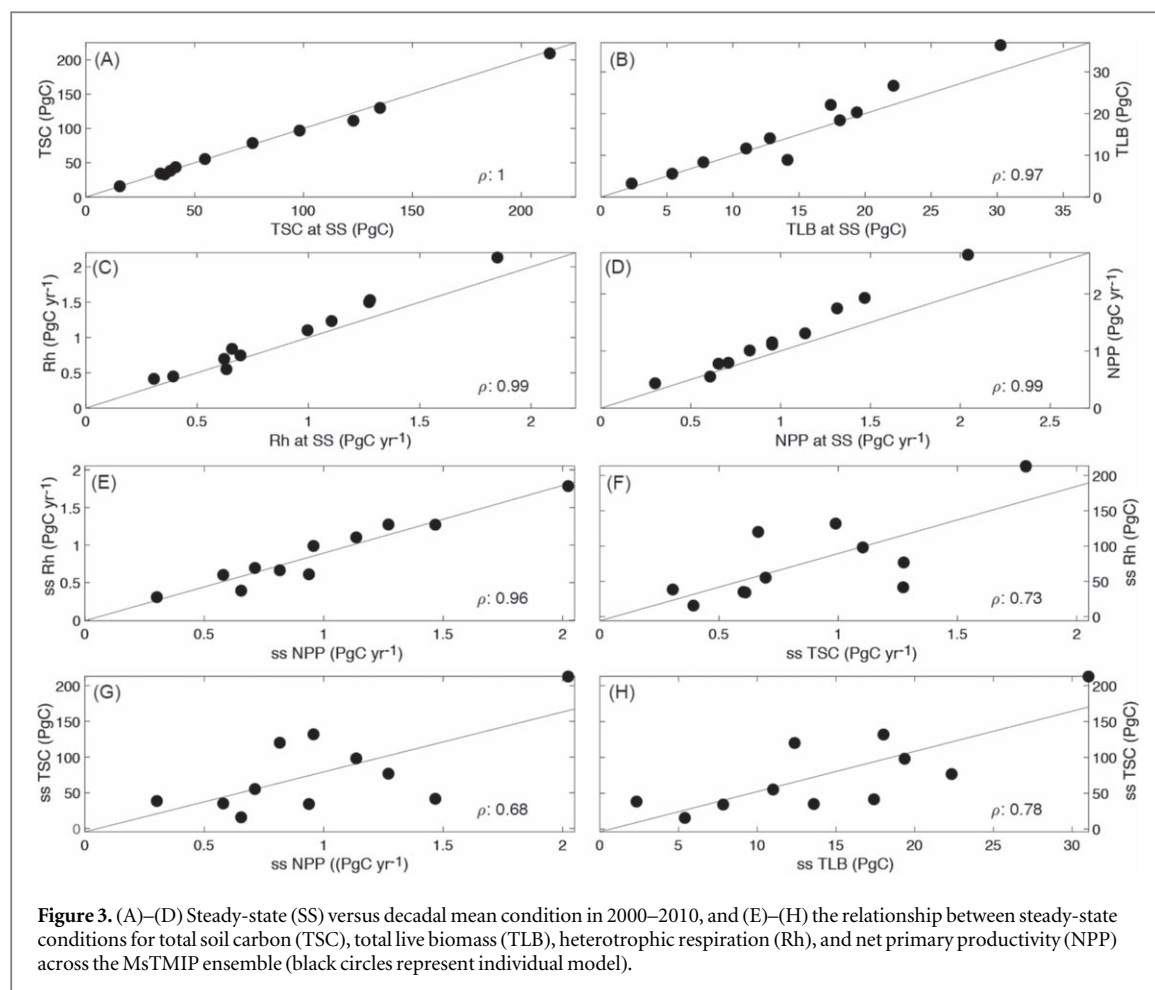


not explicitly model peatlands (Luo *et al* 2016). Also, the prescribed vegetation cover provided to modelers as part of the MsTMIP protocol did not explicitly include wetland and peatland land cover types (Wei *et al* 2014b, Tian *et al* 2015). In order to capture the frozen-soil carbon dynamics ubiquitous to permafrost regions, models need to include vertically resolved soil carbon pools. This result it also echoed by other studies (e.g. Koven *et al* 2013b, Luo *et al* 2016, McGuire *et al* 2018). Most models have multiple layers to simulate soil temperature and moisture. However, many have only a single, dimensionless soil carbon model representing all the carbon within the soil column (table S2). Even if the model partitions carbon into multiple pools, a dimensionless soil carbon model cannot capture the vertical mixing of frozen and thawed soil, or the vertical heterogeneity of organic matter within the soil column. To simulate the accumulation of carbon within permafrost, the model must also include burial by sediment and vertical mixing by cryoturbation (Koven *et al* 2013b, Burke *et al* 2017).

In addition to most models underestimating the size of TSC stocks across the full ABoVE domain, a majority of models (8 out of 11) also underestimate the magnitude of contemporary (i.e. 2000–2010) ecosystem fluxes (Rh and NPP) compared to

observationally-based constraints (figure 1 and table 1; figures S7, S8). Since most models assume first-order reaction kinetics for soil decomposition, the magnitude of a given model's initial TSC stocks controls the magnitude of Rh. We see this in comparison with observational constraints, where underestimation of Rh (and soil carbon residence time) is greatest at higher latitudes (North Slope, the Mackenzie basin, and the Hudson Bay peatlands) where models also tend to underestimate overall TSC stock size (figures 2(c), (f), (i)). Overall, model underestimation of Rh leads to inferred soil carbon residence times (ResT) that are longer than observations would suggest (figures 1 and 2(i)).

We find that the magnitude of carbon stocks at steady-state is the prime driver of carbon stock size at the end of the simulation period (figures 3(a), (b); table S3). This is true not only for carbon stocks, but also key ecosystems fluxes (figures 3(c), (d)). Almost all land carbon models, assume steady-state conditions prior to the start of transient simulations. While some models initialize carbon pools with observed biomass at the start of simulations (Schaefer *et al* 2008), without also optimizing model parameters, models will drift back towards their internal steady-state condition. The modeling community is starting to recognize the importance of steady-state conditions on model



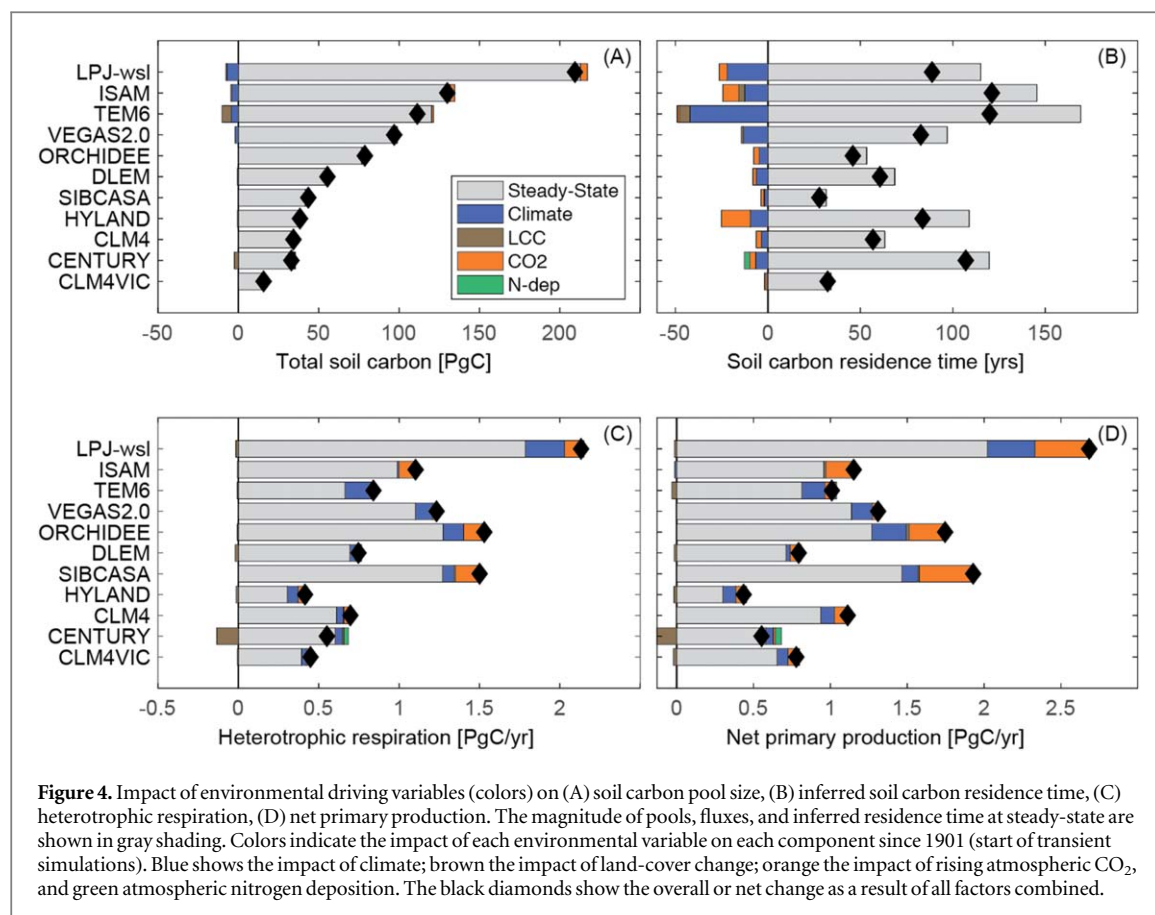
performance (Tian *et al* 2015, Luo *et al* 2017, Shi *et al* 2018), however it is clear that more work is needed.

The magnitude of steady-state TSC stocks is strongly driven by simulated above ground productivity or GPP, and both productivity and TSC drive changes the magnitude of respiration losses (figures 3(e)–(h)). GPP represents the primary carbon input into ecosystems and any bias in simulated GPP or NPP (defined as $GPP - R_a$) propagates through the model to produce bias in both simulated carbon stocks and respiration (Schaefer *et al* 2012). For example, models with lower NPP than MODIS also estimate lower soil carbon stocks compare to the NCSCD estimate (table 1; figure 1). Evaluating simulated GPP is beyond the scope of this analysis, but Schaefer *et al* (2012) found that improved representation of light use efficiency, drought stress, and low temperature inhibition improve simulated GPP.

The influence of steady-state conditions on pool and flux size overwhelms the effect of environmental forcing factors in terms of the spread or variability in model estimates of both stocks and fluxes over the 110-simulation period of the MsTMIP runs (figure 4). However, models do indicate an acceleration of soil carbon cycling within the ABoVE domain through a decrease in inferred soil carbon residence time and an increase in the magnitude (i.e. rate) of NPP and Rh

relative to 1901 conditions (figure 4). Across the ensemble, the primary driver for decreases in soil carbon residence time (and increase in flux magnitude) is climate, followed by rising atmospheric CO_2 ; however, the relative contribution of each varies considerably across models (figure 4).

While the relative magnitude of soil carbon losses through Rh across models is strongly tied to the rate of productivity and the size of soil carbon stocks of a given model, a model's sensitivity to environmental forcing factors (e.g. temperature) should control the rate of carbon turnover changes (i.e. acceleration) with changing environmental conditions. We evaluated model sensitivity of soil carbon losses with changing temperature by extracted an inferred Q10 for each model for the two major ecosystem types within the ABoVE domain (figure 5). Two groups of models emerge—those with an inferred Q10 that is too low compared to the observations and those with an inferred Q10 that is comparable to observations (figure 5). With the exception of CENTURY, using air temperature as opposed to soil temperature to derive the inferred Q10 does not have a significant impact on the results (figure S9). CENTURY reports soil temperatures that are several degrees warmer than air temperatures in the region and significantly warmer than soil temperatures reported by other models (figures S10,



S11). Given that CENTURY weights soil temperatures by day length (Parton 1984, Parton *et al* 1992), it is possible that high-latitude summer temperature is overestimated in the model and that this gives rise to the large differences seen here. The same analysis was performed by ecosystem type (boreal forest versus northern shrublands; figure S12). Most models apply a consistent or constant Q₁₀ regardless of ecosystem type, and this appears to be the case for most models within the MsTMIP ensemble (figure S12). However, the inferred Q₁₀ derived from the observational constraint varies between boreal forests and northern shrublands, suggesting that perhaps ecosystem specific Q₁₀s are more appropriate. The lack of observations prevented us from statistically evaluating respiration response below freezing. In frozen soils, microbial activity and associated respiration becomes limited to thin water films surrounding fine soil particles (Schaefer and Jafarov 2016). However, we do know that respiration decreases with temperature and effectively ceases below -8°C (Mikan *et al* 2002). However, we found that most models show small, but persistent respiration at temperatures below freezing (not shown). To improve simulated respiration in frozen soil, models need to account for thin water films and reduce respiration to near zero at about -8°C .

Interestingly, models that perform better against observational constraints of model endpoints (e.g. soil carbon stocks, component fluxes), do not necessarily perform better against observational constraints on

pathways to those endpoints (i.e. functional response of respiration versus temperature). In fact, the reverse tends to be true. Models that show weaker sensitivity of Rh to temperature (figure 5) tend to perform better against observational constraints in terms of their estimated soil carbon stock and flux magnitudes (figure 1). This suggests that models may be getting the right endpoints, but for the wrong reasons. It could also mean that the observational constraints themselves are insufficient for properly assessing model performance.

Consistent with observations, models tend to show an increase in respiration with soil moisture up to an optimal VWC. Past this optimal value, further increases in soil moisture lead to a decline in the amount of belowground carbon respired (figures S13, S14). The optimal VWC where maximum respiration occurred varies between models in the ensemble, but appears consistent with observations (figure 6) when using only control plots. When including both control and warming plots in the observational constraint, the uncertainty on the optimal VWC at maximum respiration narrows slightly (figure 6), leaving several models with either an optimal VWC lower (TEM6) or higher (LPJ-wsl, CLM4VIC) than observations suggest.

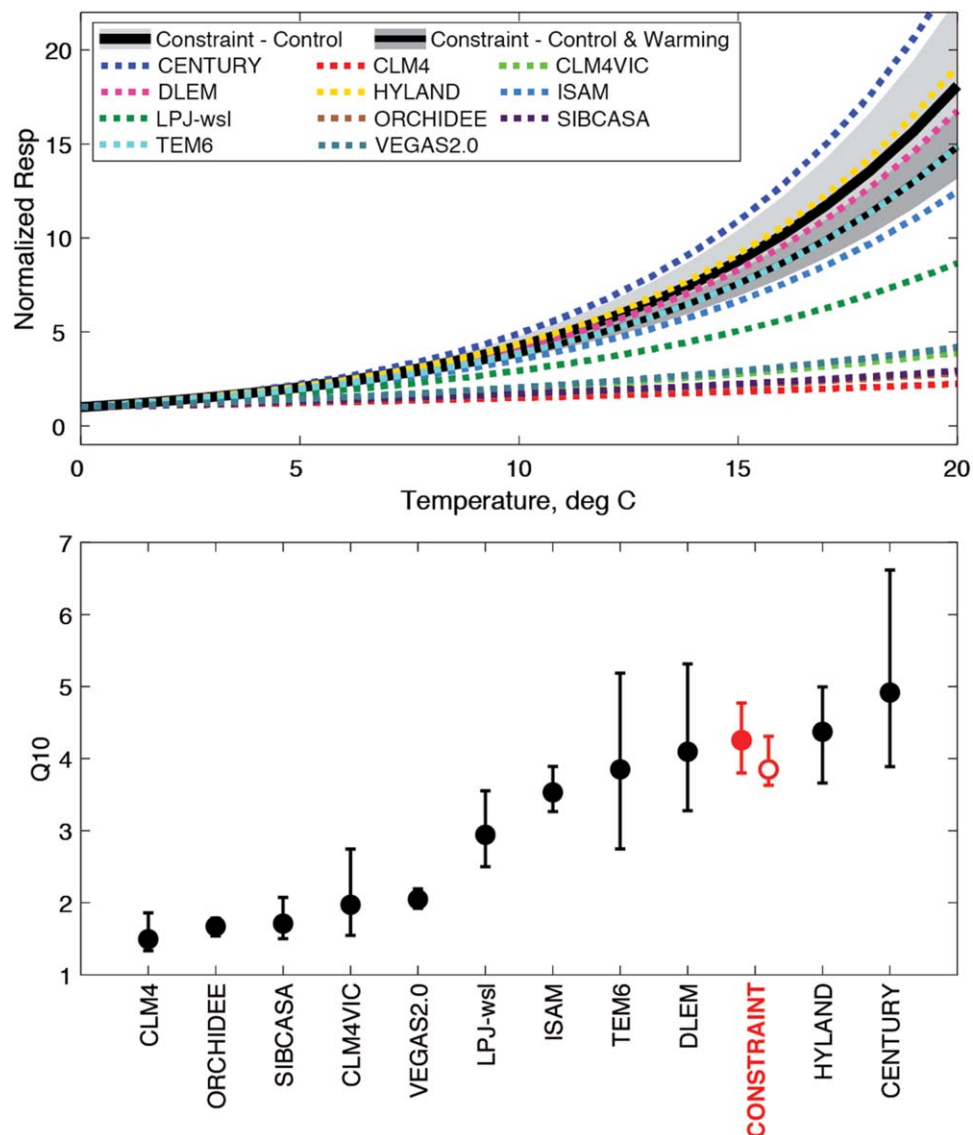


Figure 5. Functional response of simulated respiration with changing air temperature. (Top) Median normalized respiration with temperature extracted from gridcell response curves using monthly heterotrophic respiration (R_h) and air temperature for each model in the ensemble. Compared to similar curve obtained from an observational constraint taken from soil respiration measurements taken from control-only plots and both control and warming plots at warming experiments sites in Boreal and Northern Shrubland ecosystems (Carey *et al* 2016a, 2016b). (bottom) the inferred Q10 for each model based on heterotrophic respiration (R_h) as a function of temperature relationship shown in (A). Black circles represent the median inferred Q10 across all land cells within the ABoVE domain for Boreal and Northern Shrubland ecosystem in non-permafrost and peatland dominated regions. Error bars represent the spread (interquartile range) in inferred Q10 values across gridcells. Red circle and error bars show the inferred Q10 from control-only (filled circle) and both control and warming (open circle) sites at warming experiments reported by Carey *et al* (2016a, 2016b).

5. Conclusions

The results from this study suggest three potential pathways to improving simulated soil carbon dynamics, particularly in the ABR, that encompass both modeled *endpoints* and the *pathways* to those endpoints: (1) improving steady-state conditions in models; (2) using functional benchmarks to constrain model sensitivity to key environmental forcing factors; and (3) including vertically resolved soil biogeochemistry to better simulate soil carbon dynamics particularly in permafrost regions.

5.1. Improving steady-state initial conditions (endpoint)

To improve simulated steady-state soil carbon pool size, models need to improve simulated GPP. The initial, steady state pool size determines the magnitude of carbon stocks and respiration throughout the 110 year simulation, overwhelming all other factors. One can constrain initial carbon pools using observed carbon stocks (Schaefer *et al* 2008). However, without change a model's parameterization of GPP, pools and fluxes will rapidly drift back towards the model's own internal steady-state condition. Improving the representation of light use efficiency, drought stress, and

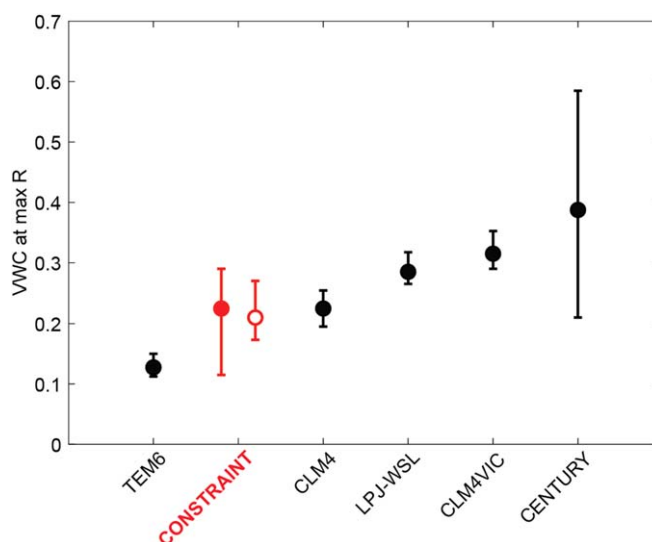


Figure 6. Volumetric water content (VWC) at maximum respiration extracted from gridcell response curves using monthly heterotrophic respiration (R_h) and volumetric water content for each model in the ensemble that reported soil moisture (see figure S12). Modeled optimal VWC are compared to similar value obtained from an observational constraint taken from soil respiration measurements taken from control-only (red fill circle) plots and both control and warming plots (red open circle) at warming experiments sites in Boreal and Northern Shrubland ecosystems (Carey *et al* 2016a, 2016b). Circles represent the median inferred VWC at max R across all land cells within the ABoVE domain for Boreal and Northern Shrubland ecosystem in non-permafrost and peatland dominated regions. Error bars represent the full spread (max and min) in inferred VWC at max R across various VWC bin choices (see figure S12).

low temperature inhibition will improve simulated GPP (Schaefer *et al* 2012), and will thus improve the magnitude of simulated carbon stocks and respiration.

5.2. Improving model response to environmental forcing (pathway)

Uncertainty in model response to key drivers (e.g. climate warming) represents a major weakness in future carbon cycle and climate projections (Friedlingstein *et al* 2014, Huntzinger *et al* 2017). Several of the models in this study show temperature sensitivities, or Q10 values, that are half of observed. This suggests that modelers need to revisit their Q10 values and perhaps employ biome specific Q10 values, particularly in high-latitude. While models tend to reproduce the shape of the observed soil moisture response within uncertainty, we recognize the need for more data to improve this benchmark. Models likely also need to account for thin water films in frozen soils, but more data is needed to create an appropriate benchmark for models. Modelers and field-based scientists need to work more closely to bridge the gap between data and models and incorporate more functional benchmarks into ongoing activities such as those included in ABoVE (Fisher *et al* 2018).

5.3. Vertical resolved soil biogeochemistry (pathway)

The spread in simulated stocks and fluxes appear greatest in regions dominated by peat and permafrost, suggesting models do not capture frozen-soil carbon dynamics. A dimensionless carbon model cannot differentiate between frozen and thawed organic

matter distributed vertically in the soil column. Also, to simulate the accumulation on carbon in permafrost, a model needs to include sedimentation and burial as well as vertical mixing due to cryoturbation (Koven *et al* 2013b, Luo *et al* 2016, McGuire *et al* 2018). Vertically resolved soil carbon pools will likely also improve simulated respiration in permafrost regions.

Modelers should simultaneously compare model outputs against many benchmarks throughout development to evaluate model *endpoints* and *pathways*. Of course, we need more observations to improve model benchmarks, such as respiration response below freezing. Nevertheless, reducing overall model uncertainty often results in the ‘balloon’ effect: squeezing in one place causes it to pop out in another. Internal model components have become so complex and interconnected that fixing one process often changes another, seemingly unrelated process. Only with frequent comparison of model outputs against multiple benchmarks can we detect changes to these interconnected processes and reduce overall model uncertainty.

Acknowledgments

This work was supported by NASA’S Arctic Boreal Vulnerability Experiment (ABoVE; <https://above.nasa.gov>); NNN13D504T. Funding for the Multi-scale synthesis and Terrestrial Model Intercomparison Project (MsTMIP; <https://nacp.ornl.gov/MsTMIP.shtml>) activity was provided through NASA ROSES Grant #NNX10AG01A. Data management support


for preparing, documenting, and distributing model driver and output data was performed by the Modeling and Synthesis Thematic Data Center at Oak Ridge National Laboratory (MAST-DC; <https://nacp.ornl.gov>), with funding through NASA ROSES Grant #NNH10AN681. Finalized MsTMIP data products are archived at the ORNL DAAC (<https://daac.ornl.gov>). We also acknowledge the modeling groups that provided results to MsTMIP. The synthesis of site-level soil respiration, temperature, and moisture data reported in Carey *et al* 2016a, 2016b) was funded by the US Geological Survey (USGS) John Wesley Powell Center for Analysis and Synthesis Award G13AC00193. Additional support for that work was also provided by the USGS Land Carbon Program. JBF carried out the research at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. California Institute of Technology. Government sponsorship acknowledged.

Data availability

The data that support the findings of this study are openly available at DOI. Version 1 of MsTMIP Phase I model output is available at DOI: <https://doi.org/10.3334/ORNLDAAC/1225>. Version 2 of the MsTMIP Phase I output is currently in press and should be available following within 3 months of publication at <https://doi.org/10.3334/ORNLDAAC/159>. The synthesized soil respiration data from warming experiments is publicly available from the US Geological Survey (USGS): [10.5066/F7MK6B1X](https://doi.org/10.5066/F7MK6B1X).

ORCID iDs

J B Fisher  <https://orcid.org/0000-0003-4734-9085>

E Stofferahn  <https://orcid.org/0000-0001-6960-4193>

A K Jain  <https://orcid.org/0000-0002-4051-3228>

H Kolus  <https://orcid.org/0000-0001-9300-4585>

J Mao  <https://orcid.org/0000-0002-2050-7373>

References

- AMAP 2017 *Snow, Water, Ice and Permafrost in the Arctic* (Oslo: Arctic Monitoring and Assessment Programme)
- Bond-Lamberty B and Thomson A 2010a A global database of soil respiration data *Biogeosciences* **7** 1915–26
- Bond-Lamberty B and Thomson A 2010b Temperature-associated increases in the global soil respiration record *Nature* **464** 579–U132
- Bond-Lamberty B, Wang C and Gower S 2004 A global relationship between the heterotrophic and autotrophic components of soil respiration? *Glob. Change Biol.* **10** 1756–66
- Burke E, Chadburn S and Ekici A 2017 A vertical representation of soil carbon in the JULES land surface scheme (vn4.3_permafrost) with a focus on permafrost regions *Geosci. Model Dev.* **10** 959–75
- Carey J *et al* 2016a Temperature response of soil respiration largely unaltered with experimental warming *Proc. Natl Acad. Sci. USA* **113** 13797–802
- Carey J C *et al* 2016b Data compilation of soil respiration, moisture, and temperature measurements from global warming experiments from 1994–2014. US Geological Survey data release (<https://doi.org/10.5066/F7MK6B1X>)
- Collier N, Hoffman F M, Lawrence D M, Keppel-Aleks G, Koven C D, Riley W J, Mu M and Randerson J T 2018 The international land model benchmarking (ILAMB) system: design, theory, and implementation *J. Adv. Model. Earth Syst.* **10** 2731–54
- Crowther T W, Thomas S M, Maynard D S, Baldrian P, Covey K, Frey S D, van Diepen L T A and Bradford M A 2015 Biotic interactions mediate soil microbial feedbacks to climate change *Proc. Natl Acad. Sci. USA* **112** 7033–8
- Fisher J, Huntzinger D, Schwalm C, Sitch S, Gagli A and Liverman D 2014a Modeling the terrestrial biosphere *Annu. Rev. Environ. Resour.* **39** 91
- Fisher J *et al* 2014b Carbon cycle uncertainty in the Alaskan Arctic *Biogeosciences* **11** 4271–88
- Fisher J *et al* 2018 Missing pieces to modeling the Arctic-Boreal puzzle *Environ. Res. Lett.* **13** 020202
- Friedlingstein P, Meinshausen M, Arora V K, Jones C D, Anav A, Liddicoat S K and Knutti R 2014 Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks *J. Clim.* **27** 511–26
- Hall A, Cox P, Huntingford C and Klein S 2019 Progressing emergent constraints on future climate change *Nat. Clim. Change* **9** 269–78
- Harden J *et al* 2012 Field information links permafrost carbon to physical vulnerabilities of thawing *Geophys. Res. Lett.* **39** L15704
- Hashimoto S, Carvalhais N, Ito A, Migliavacca M, Nishina K and Reichstein M 2015 Global spatiotemporal distribution of soil respiration modeled using a global database *Biogeosciences* **12** 4121–32
- Hayes D, Kicklighter D, McGuire A, Chen M, Zhuang Q, Yuan F, Melillo J and Wullschlegel S 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 CO₂ sink weakening? *Glob. Biogeochem. Cycles* **25** GB3018
- Hayes D and Turner D 2012 The need for ‘apples-to-apples’ comparisons of carbon dioxide source and sink estimates *Eos, Trans. Am. Geophys. Union* **93** 404–5
- Hobbie S, Nadelhoffer K and Hogberg P 2002 A synthesis: the role of nutrients as constraints on carbon balances in boreal and arctic regions *Plant Soil* **242** 163–70
- Hoffman F M *et al* 2017 *International Land Model Benchmarking (ILAMB) 2016 Workshop Report* (Germantown, MD: US Department of Energy, Office of Science) (https://www.ilamb.org/meetings/washington2016/2016_ILAMB_Report_V10_web.pdf)
- Hugelius G *et al* 2013a A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region *Earth Syst. Sci. Data* **5** 393–402
- Hugelius G, Tarnocai C, Broll G, Canadell J G, Kuhry P and Swanson D K 2013b 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
- Hugelius G *et al* 2014 Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps *Biogeosciences* **11** 6573–93
- Huntzinger D N *et al* 2013 The North American carbon program multi-scale synthesis and terrestrial model intercomparison project: I. Overview and experimental design *Geosci. Model Dev.* **6** 2121–33
- Huntzinger D N, Schwalm C, Michalak A M, Schaefer K, Wei Y, Cook R B and Jacobson A R 2014 *NACP MsTMIP Summary of*

- Model Structure and Characteristics* (Oak Ridge, TN: ORNL DAAC) (<https://doi.org/10.3334/ORNLDAAC/1228>)
- Huntzinger D *et al* 2017 Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions *Sci. Rep.* **7** 4765
- Huntzinger D N *et al* 2020 NACP MsTMIP: Global 0.5-Degree Model Outputs in Standard Format, Version 2.0 (Oak Ridge, TN: ORNL DAAC) (<https://doi.org/10.3334/ORNLDAAC/1599>)
- Jeong S-J *et al* 2018 Accelerating rates of Arctic carbon cycling revealed by long-term atmospheric CO₂ measurements *Sci. Adv.* **4** ea01167
- Koster R D, Guo Z, Yang R, Dirmeyer P A, Mitchell K and Puma M J 2009 On the nature of soil moisture in land surface models *J. Clim.* **22** 4322–35
- Koven C, Riley W and Stern A 2013a Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 Earth system models *J. Clim.* **26** 1877–900
- Koven C, Riley W, Subin Z, Tang J, Torn M, Collins W, Bonan G, Lawrence D and Swenson S 2013b The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4 *Biogeosciences* **10** 7109–31
- Koven C, Hugelius G, Lawrence D and Wieder W 2017 Higher climatological temperature sensitivity of soil carbon in cold than warm climates *Nat. Clim. Change* **7** 817
- Lenton T M, Held H, Kriegler E, Hall J W, Lucht W, Rahmstorf S and Schellnhuber H J 2008 Tipping elements in the Earth's climate system *Proc. Natl Acad. Sci. USA* **105** 1786–93
- Limpens J, Berendse F, Blodau C, Canadell J, Freeman C, Holden J, Roulet N, Rydin H and Schaepman-Strub G 2008 Peatlands and the carbon cycle: from local processes to global implications—a synthesis *Biogeosciences* **5** 1475–91
- Loboda T V, Hoy E E and Carroll M L 2017 ABoVE: Study Domain and Standard Reference Grids, Version 2 (Oak Ridge, TN: ORNL DAAC) (<https://doi.org/10.3334/ORNLDAAC/1527>)
- Luo Y, Keenan T and Smith M 2015 Predictability of the terrestrial carbon cycle *Glob. Change Biol.* **21** 1737–51
- Luo Y *et al* 2016 Toward more realistic projections of soil carbon dynamics by Earth system models *Glob. Biogeochem. Cycles* **30** 40–56
- Luo Y *et al* 2017 Transient dynamics of terrestrial carbon storage: mathematical foundation and its applications *Biogeosciences* **14** 145–61
- Mack M, Schuur E, Bret-Harte M, Shaver G and Chapin F 2004 Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization *Nature* **431** 440–3
- McGuire A *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 *et al* 2016 Variability in the sensitivity among model simulations of permafrost and carbon dynamics in the permafrost region between 1960 and 2009 *Glob. Biogeochem. Cycles* **30** 1015–37
- McGuire A *et al* 2018 Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change *Proc. Natl Acad. Sci. USA* **115** 3882–7
- Mikan C J, Schimel J P and Doyle A P 2002 Temperature controls of microbial respiration in arctic tundra soils above and below freezing *Soil Biol. Biochem.* **34** 1785–95
- Mystakidis S, Davin E L, Gruber N and Seneviratne S I 2016 Constraining future terrestrial carbon cycle projections using observation-based water and carbon flux estimates *Glob. Change Biol.* **22** 2198–215
- Natali S, Schuur E and Rubin R 2012 Increased plant productivity in Alaskan tundra as a result of experimental warming of soil and permafrost *J. Ecol.* **100** 488–98
- Parton W 1984 Predicting soil temperatures in a shortgrass steppe *Soil Sci.* **138** 93–101
- Parton W J, McKeown R, Kirchner V and Ojima D 1992 *Users Guide for the CENTURY Model* (Fort Collins, CO: Colorado State University) (<https://www2.nrel.colostate.edu/projects/century/documentation2.htm>)
- Phillips C, Bond-Lamberty B, Desai A, Lavoie M, Risk D, Tang J, Todd-Brown K and Vargas R 2017 The value of soil respiration measurements for interpreting and modeling terrestrial carbon cycling *Plant Soil* **413** 1–25
- Qian H, Joseph R and Zeng N 2010 Enhanced terrestrial carbon uptake in the Northern high latitudes in the 21st century from the coupled carbon cycle climate model intercomparison project model projections *Glob. Change Biol.* **16** 641–56
- Raich J and Potter C 1995 Global patterns of carbon-dioxide emissions from soils *Glob. Biogeochem. Cycles* **9** 23–36
- Raich J, Potter C and Bhagawati D 2002 Interannual variability in global soil respiration, 1980–94 *Glob. Change Biol.* **8** 800–12
- Schadel C, Schuur E, Bracho R, Elberling B, Knoblauch C, Lee H, Luo Y, Shaver G and Turetsky M 2014 Circumpolar assessment of permafrost C quality and its vulnerability over time using long-term incubation data *Glob. Change Biol.* **20** 641–52
- Schaefer K, Collatz G, Tans P, Denning A, Baker I, Berry J, Prihodko L, Suits N and Philpott A 2008 Combined simple biosphere/carnegie-ames-stanford approach terrestrial carbon cycle model *J. Geophys. Res.-Biogeosci.* **113** G03034
- Schaefer K, Zhang T, Bruhwiler L and Barrett A P 2011 Amount and timing of permafrost carbon release in response to climate warming *Tellus B* **63** 168–80
- Schaefer K *et al* 2012 A model-data comparison of gross primary productivity: results from the North American Carbon Program site synthesis *J. Geophys. Res.-Biogeosci.* **117** G03010
- Schaefer K and Jafarov E 2016 A parameterization of respiration in frozen soils based on substrate availability *Biogeosciences* **13** 1991–2001
- Schuur E A G *et al* 2013 Expert assessment of vulnerability of permafrost carbon to climate change *Clim. Change* **119** 359–74
- Schuur E A G *et al* 2015 Climate change and the permafrost carbon feedback *Nature* **520** 171–9
- Schwalm C R *et al* 2015 Toward 'optimal' integration of terrestrial biosphere models *Geophys. Res. Lett.* **42** 4418–28
- Shi Z, Crowell S, Luo Y and Moore B 2018 Model structures amplify uncertainty in predicted soil carbon responses to climate change *Nat. Commun.* **9** 2171
- Stofferahn E, Fisher J B, Hayes D J, Schwalm C R, Huntzinger D N, Hantson W, Poulter B and Zhang Z 2019 The Arctic-Boreal vulnerability experiment model benchmarking system *Environ. Res. Lett.* **14** 055002
- Tang J and Baldocchi D 2005 Spatial-temporal variation in soil respiration in an oak-grass savanna ecosystem in California and its partitioning into autotrophic and heterotrophic components *Biogeochemistry* **73** 183–207
- Tang J, Bradford M A, Carey J, Crowther T W, Machmuller M B, Mohan J E and Todd-Brown K 2019 *Ecosystem Consequences of Soil Warming* ed J E Mohan (New York: Academic) pp 175–208
- Tarnocai C, Canadell J G, Schuur E A G, Kuhry P, Mazhitova G and Zimov S 2009 Soil organic carbon pools in the northern circumpolar permafrost region *Glob. Biogeochem. Cycles* **23**
- Tian H *et al* 2015 Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: current status and future directions *Glob. Biogeochemical Cycles* **29** 775–92
- Todd-Brown K, Randerson J, Post W, Hoffman F, Tarnocai C, Schuur E and Allison S 2013 Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations *Biogeosciences* **10** 1717–36
- USGCRP 2017 *Climate Science Special Report: Fourth National Climate Assessment* vol 1 ed D J Wuebbles *et al* (Washington, DC: US Global Change Research Program)
- Wei Y *et al* 2014a NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison (Oak Ridge, TN: ORNL DAAC) (<https://doi.org/10.3334/ORNLDAAC/1220>)

- Wei Y *et al* 2014b The North American Carbon program multi-scale synthesis and terrestrial model intercomparison project: II. Environmental driver data *Geosci. Model Dev.* **7** 2875–93
- Zeng H, Jia G and Epstein H 2011 Recent changes in phenology over the northern high latitudes detected from multi-satellite data *Environ. Res. Lett.* **6** 045508
- Zhao M, Heinsch F, Nemani R and Running S 2005 Improvements of the MODIS terrestrial gross and net primary production global data set *Remote Sens. Environ.* **95** 164–76
- Zhao M and Running S W 2010 Drought-induced reduction in global terrestrial net primary production from 2000 through 2009 *Science* **329** 940