AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Biogeosciences

RESEARCH ARTICLE
10.1002/2015JG003175

Key Points:
- Modeled historical plant functional types are evaluated with a novel data set
- Inaccurate PFT representation tied to climate-vegetation relationships
- Incorrect PFTs can bias albedo, net primary productivity, and transpiration

Supporting Information:
- Supporting Information S1

Correspondence to:
J. H. Matthes,
Jaclyn.H.Matthes@dartmouth.edu

Citation:
Matthes, J. H., S. Goring, J. W. Williams, and M. C. Dietze (2016), Benchmarking historical CMIP5 plant functional types across the Upper Midwest and Northeastern United States, *J. Geophys. Res. Biogeosci.*, *121*, doi:10.1002/2015JG003175.

Received 5 AUG 2015
Accepted 2 FEB 2016
Accepted article online 5 FEB 2016

Benchmarking historical CMIP5 plant functional types across the Upper Midwest and Northeastern United States

Jaclyn Hatala Matthes[1], Simon Goring[2], John W. Williams[2,3], and Michael C. Dietze[4]

[1]Department of Geography, Dartmouth College, Hanover, New Hampshire, USA, [2]Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin, USA, [3]Center for Climatic Research, University of Wisconsin-Madison, Madison, Wisconsin, USA, [4]Department of Earth and Environment, Boston University, Boston, Massachusetts, USA

Abstract Centennial-scale climate-ecosystem feedbacks are a major source of predictive uncertainty for land-atmosphere fluxes of energy, carbon, and water. Accurate representations of plant functional type (PFT) distributions through time and space are required for modeling centennial-scale feedbacks within Earth system models (ESMs). We tested the ability of ESMs from the Coupled Model Intercomparison Project Phase 5 (CMIP5) to capture historical PFT distributions at the time of Euro-American settlement in the Northeastern United States against a new subcontinental-scale data set of historical tree abundances derived from forest composition surveys. To identify and diagnose errors in ESM-simulated PFT distributions and quantify impacts on modeled albedo, net primary productivity, and transpiration, we analyzed actual and modeled PFT distributions with respect to historical mean annual climate and modeled elasticity among PFTs, climate, and vegetation-atmosphere fluxes. Historical PFT distributions were poorly matched between ESMs and the settlement-era data, often due to inaccurate PFT-climate relationships within ESMs, particularly for evergreen trees. Some models exhibited large local, but regionally compensating, errors in simulated albedo, net primary productivity, and transpiration due to inaccurate PFT distributions, while others had systematic regional biases in vegetation-atmosphere fluxes. Internal variable elasticity varied among ESMs, and these differences closely corresponded to model skill in predicting PFT distributions. New historical benchmarks like the settlement-era vegetation data provide opportunities to confront ESMs, parse sources of error, and improve simulations of historical and future vegetation-atmosphere feedbacks.

1. Introduction

Substantial uncertainty in projections of future climate arises from processes with long system memories at multiple temporal and spatial scales (e.g., between ecosystems and climate), particularly at multidecadal to centennial timescales [*Friedlingstein et al.*, 2006, 2013; *Ahlström et al.*, 2012; *Piao et al.*, 2013]. Given that these are the timescales most relevant to long-term infrastructural adaptation in response to global climate change, there is a large societal demand to improve predictive skill [*Meehl et al.*, 2009]. A central challenge to improving Earth system models (ESMs) at decadal to centennial timescales is the paucity of data sets informing land-atmosphere feedbacks. Assessing the ability of ESMs to capture historical ecosystem states—e.g., the land surface state at 1850 as a preindustrial baseline in Coupled Model Intercomparison Project Phase 5 (CMIP5) modeling [*Taylor et al.*, 2011]—and the effects of those states on modeled ecosystem-atmosphere feedbacks is an essential step toward improving the predictive power of long-term 21st century ecosystem-climate feedback trajectories. Reconstructions of historical forest composition provide a critical benchmark both for model initialization and characterization of transient modeled states, particularly at regional spatial scales where the largest uncertainties lie for future climate and ecosystem projections [*Sitch et al.*, 2008].

Initial land surface states within the CMIP5 experiment are typically based on models of potential vegetation [*Ramankutty and Foley*, 1999] due to the scarcity of preindustrial land cover data. Surface states are then modified by land use changes inferred from population and economic data [*Pongratz et al.*, 2009; *Hurtt et al.*, 2011]. However, modeled potential vegetation maps are constructed with implicit assumptions about species-climate relationships. Factors such as land use, dispersal limitation, natural disturbance, decadal- to centennial-scale climate variability, and other historical contingencies may prevent species from filling their potential niche [*Rhemtulla et al.*, 2009; *Thompson et al.*, 2013]. Hence, use of modeled potential vegetation for ESM initialization likely creates systematic and poorly understood biases in 21st century simulations.

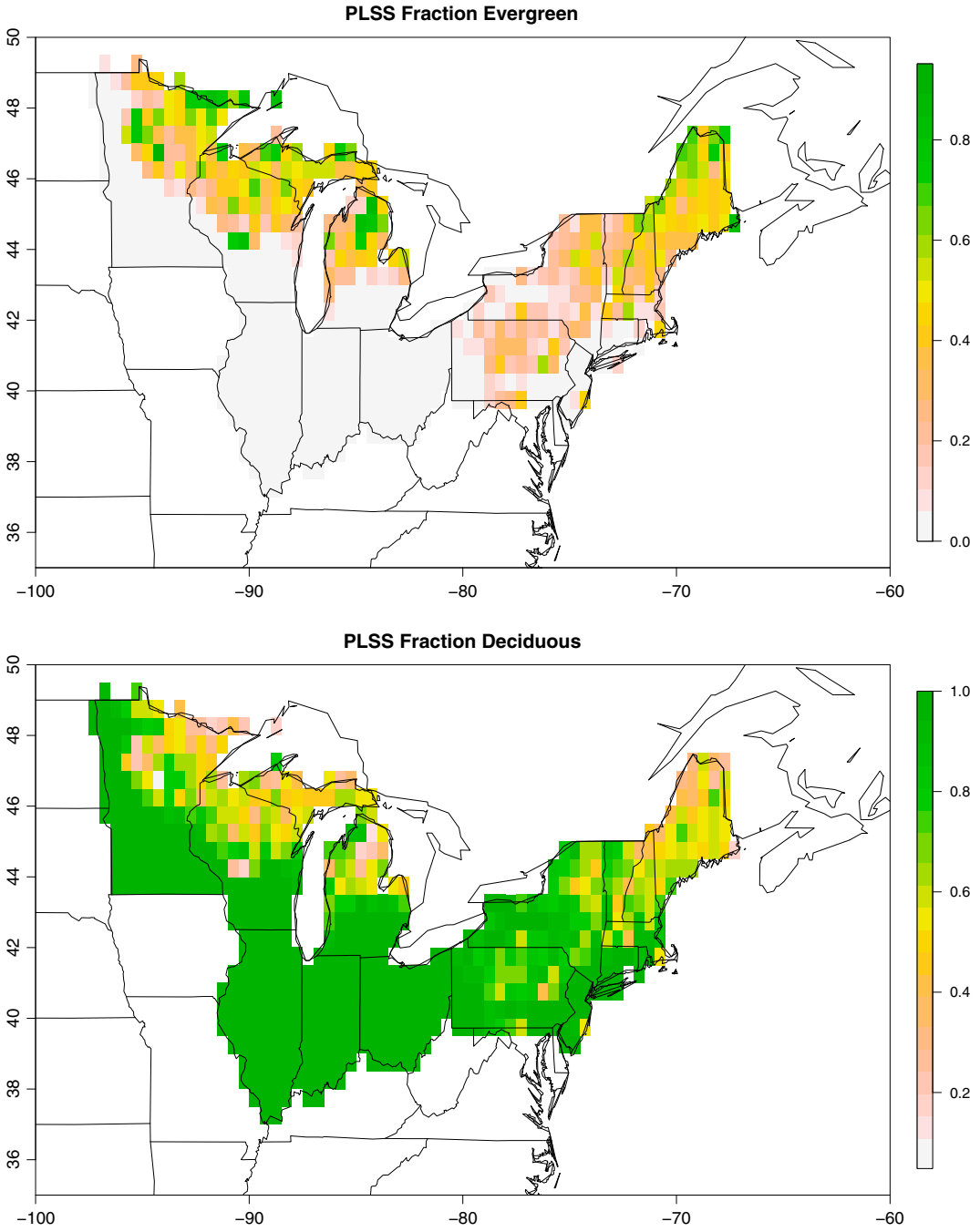

Figure 1. Reconstructed distribution of evergreen and deciduous PFTs at the time of Euro-American settlement. The Euro-American settlement vegetation (ESV) data set represents the state of vegetation at the time of Euro-American settlement, from circa 1700 on the eastern seaboard, to as late as 1907 in the Upper Midwest. This data set was digitized from Public Land Survey and proprietor township records and converted to a spatially continuous 0.5° grid through a Bayesian hierarchical statistical model. Fractional values represent the proportion of individual trees relative to total number of surveyed individuals.

A new regional reconstruction of historical plant functional type (PFT) distributions from the Euro-American settlement-era vegetation (ESV) data covers the Upper Midwest through the Northeastern United States [*Goring et al.*, 2015] (Figure 1) and provides the opportunity for detailed evaluation of PFT simulation accuracy in the CMIP5 preindustrial control experiment and the effects on simulated ecosystem-atmosphere feedbacks. ESV data are compiled from forest composition data collected by surveyors working for individual townships (in

Table 1. List of ESV Tree Genera Aggregated Into Evergreen and Deciduous PFTs[a]

Evergreen PFT	Deciduous PFT
Cedar	Ash
Fir	Basswood
Hemlock	Beech
Pine	Birch
Spruce	Black gum
	Elm
	Hickory
	Ironwood
	Maple
	Oak
	Other hardwood
	Poplar
	Tamarack
	Walnut

[a]"Other hardwood" included Buckeye, Cherry, Dogwood, Hackberry, Mulberry, and Sycamore, which were present in low abundance across the spatial domain.

New England) and the U.S. General Land Office [*Cogbill et al.*, 2003; *Thompson et al.*, 2013; *Goring et al.*, 2015]. The ESV data collected in this time transgressive survey reflect forest composition from the early 1700s to 1907, with the earliest records from the east, and the latest records in northwestern Minnesota. While Native American land use existed in this region prior to the time period captured by the aggregated ESV data, these impacts were highly patchy and restricted to within a few kilometers of settlements [*Munoz et al.*, 2014], well below the spatial resolution of the ESV data used here (8 km, aggregated to 0.5°). As such, the ESV data set describe the percent composition of tree taxa on the landscape prior to major land use change following settlement [*Goring et al.*, 2015].

The ability of CMIP5 models to capture presettlement PFT composition is critical to accurately simulate local feedbacks with climate, and in some cases for broader regional processes. Whether the land surface is dominated by evergreen or deciduous PFTs acts as a primary control on fluxes of energy, carbon, and water between the land and atmosphere, due to fundamental differences in the phenology of these two PFTs [*Bonan*, 2008]. For example, during winter the dark surface created by the perennial leaves of evergreen trees creates a much lower albedo compared to leafless deciduous trees and snow-covered ground, increasing the amount of incoming radiation that is absorbed at the surface, in turn increasing the local seasonal temperature [*Hall*, 2004; *Cheddadi and Bar-Hen*, 2009].

Our analysis focused on answering three questions: (1) How accurate are the CMIP5 ESMs at representing preindustrial PFTs within this region? (2) Are differences among the ESV data set and CMIP5 models due to differences in modeled climate or the climate-vegetation relationships encoded within the models? And (3) how does inaccurate representation of PFTs within this region affect key vegetation-atmosphere feedback variables, i.e., albedo, net primary productivity, and transpiration?

We evaluated biases in albedo, net primary productivity, and transpiration at a regional scale resulting from inaccurate PFT distributions in ESMs relative to the ESV. We further quantified the elasticity among variables related to PFT distributions, climate, and land-atmosphere flux within CMIP5 ESMs, which allowed us to identify within-model sources of error in PFT mapping by quantifying the extent to which changes in one variable initiated changes in other variables. Through this systematic analysis, we determined regional CMIP5 accuracy for preindustrial PFTs, diagnosed the reasons why some ESMs incorrectly simulated PFTs, and assessed the impact that this error had on land-atmosphere flux.

2. Methods

2.1. ESV and CMIP5 Forest Composition

The ESV data used in this analysis were developed from Public Land Survey (PLS) records in the Upper Midwest [*Liu et al.*, 2011; *Goring et al.*, 2015], and from Town Proprietor Surveys (TPSs) in New England, New York, and Pennsylvania [*Cogbill et al.*, 2003; *Thompson et al.*, 2013]. Original survey records consist of counts of representative trees within townships (the TPS), and a standardized data set representing the closest two trees at section and quarter-section points, aggregated to an 8×8 km grid (the PLS) [*Goring et al.*, 2015]. The data sets were combined using a multinomial probit model [*McCulloch and Rossi*, 1994] for the counts with latent Markov random field spatial models for each taxon with neighborhood structure defined by the four nearest neighbors of each cell [*Rue and Held*, 2005], also known as an intrinsic conditional autoregressive model [*Banerjee et al.*, 2004] to produce estimates on an 8×8 km grid across the entire Northwestern U.S. with uncertainty [*Paciorek et al.*, 2016].

Table 2. CMIP5 Model List[a]

Model Name	Native Model Resolution (Latitude × Longitude; deg)	Albedo (rsds/rsus)	Net Primary Productivity (npp)	Transpiration (tran)
ACCESS1-0	1.25 × 1.875	Yes	No	No
ACCESS1-3	1.25 × 1.875	Yes	No	No
HadGEM2-CC	1.25 × 1.875	Yes	Yes	No
HadGEM2-ES	1.25 × 1.875	Yes	Yes	No
IPSL-CM5A-LR	1.8947 × 3.75	Yes	Yes	Yes
IPSL-CM5A-MR	1.2676 × 2.5	Yes	Yes	Yes
IPSL-CM5B-LR	1.8947 × 3.75	Yes	Yes	Yes
MIROC-ESM-CHEM	2.7673 × 2.8125	Yes	Yes	Yes
MIROC-ESM	2.7673 × 2.8125	Yes	Yes	Yes
MPI-ESM-LR	1.8496 × 1.875	Yes	Yes	Yes
MPI-ESM-MR	1.8496 × 1.875	Yes	Yes	Yes
MPI-ESM-P	1.8496 × 1.875	Yes	Yes	Yes

[a]List of CMIP5 models used here, their native resolution in degrees, and whether they contributed output for albedo (where rsds is downwelling short-wave radiation and rsus is upwelling shortwave radiation), net primary productivity (npp), and transpiration (tran).

ESV taxa were aggregated to the fraction composition of evergreen and deciduous PFTs for comparison to CMIP5 model output by adding the fractional abundance of the mean posterior estimate for all taxa within each PFT at the 8 km grid cell resolution (Table 1). This ESV data set represents the percent forest composition and does not indicate the fraction of nonarboreal vegetation present. The ESV data were then reprojected and aggregated to a regular 0.5° latitude-longitude grid through bilinear interpolation using Earth System Modeling Framework software through the National Center for Atmospheric Research Command Language. We investigated the impact of regridding spatial resolution choice by conducting alternate analyses at 2.25°, 3.25°, and 4.25° resolutions and found that the choice of larger spatial resolutions did not impact the conclusions of this analysis (supporting information Figure S1).

Output for all CMIP5 models that contributed the variable PFT grid fraction (*landCoverFrac*) from the preindustrial control experiment (*piControl*) was obtained from the Earth System Grid (Table 2). piControl prescribed preindustrial CO_2 concentrations and did not include any prescribed vegetation state and is used to characterize carbon cycling just prior to 1850. As such, piControl output should closely approximate the ESV data set, which represents the state of the landscape just prior to major land use conversions during settlement.

To facilitate the comparison of PFT distributions among ESMs, which vary in their functional resolution, we aggregated all evergreen PFTs (e.g., boreal evergreen trees and temperate evergreen trees) and all deciduous PFTs within each ESM to single evergreen and deciduous PFT classes, excluding nonforest cover. Savanna and prairie comprise a small proportion of the 8 × 8 km ESV data set in the west along the prairie-forest margin of Wisconsin, Minnesota, and Illinois. At a 0.5° resolution this proportion falls even farther. CMIP5 models varied in their native spatial resolution (Table 2), so prior to our regional comparative analysis all models were resampled at a uniform 0.5° resolution using bilinear interpolation and clipped to the spatial extent of the ESV data set. CMIP5 models were left in their original resolution for within-model global PFT-climate analyses described below in section 2.3.

We compared the spatial pattern and magnitude of the CMIP5 model output to the ESV data set using Taylor plots [*Taylor*, 2001], which describe the correspondence between data and models by simultaneously summarizing the correlation, centered root-mean-square error, and standard deviation between two spatial variables. We included these statistics for only the fraction of deciduous trees in the domain, since the figure is nearly identical for evergreen trees due to the constraint that the evergreen and deciduous fractions sum to one.

2.2. Climatology for ESV and CMIP5 PFTs

We compared the simulated and observed relationships between PFT distributions and climate by examining ESV-based PFT distributions in the climate space of the CRUNCEP 0.5° climate data product, version 4 [*Mitchell and Jones*, 2005; *Wei et al*., 2014], and the distribution of the CMIP5 PFTs within the climate space of each ESM.

For CMIP5 climate variables, we downloaded output for monthly mean land surface temperature at 2 m above the Earth's surface and monthly mean precipitation flux for all models that also reported PFT grid fraction for the piControl experiment. We aggregated monthly mean model output to annual mean temperature (MAT; °C) and mean total annual precipitation (MAP; mm/yr) using the first 30 years of simulation to derive a 30 year climatology. We similarly aggregated monthly mean temperature and precipitation from the CRUNCEP data product into a 30 year mean annual climatology of MAT and MAP for years 1901–1930. Although the CRUNCEP product does not precisely align with the time period of the ESV data set, we used it because it is the temporally nearest observational gridded data product currently available. We regridded CMIP5 climate variables from their native resolution (~1.5–3° grid) to the uniform 0.5° CRUNCEP grid and clipped the output to the study region to facilitate spatial comparisons between models. We calculated the correspondence between CMIP5 modeled climatology and the CRUNCEP climatology through simple linear regression in R (supporting information Figure S2 and Table S1).

To examine similarity in PFT-climate relationships between the ESV data and the CMIP5 model output, we calculated the climate niche overlap between the ESV and CMIP5 PFT distributions as the Hellinger distance based on MAT and MAP [*Warren et al*., 2008]. The Hellinger distance is calculated as the squared difference between two probability distributions (i.e., the *f* divergence) [*van der Vaart*, 1998] and is used in biogeographical studies to assess niche overlaps among populations and species [*Warren et al*., 2008; *Evans et al*., 2009; *Heibl and Calenge*, 2013]. To determine whether climate niche overlap between PFTs in the ESV and CMIP5 piControl models differed significantly from random, we took 5000 random samples with replacement, from the ESV and CMIP5 PFT data in the MAT-MAP climate space. With each pseudo data set, we calculated the Hellinger distance metric for the niche overlap between the ESV data and the CMIP5 models, to estimate the mean distance (with a 95% interval) representing a random relationship generated from the data. We used these pseudodata to evaluate the difference between the actual ESV-CMIP5 PFT niche overlaps and the expected amount of niche overlap given a random spatial redistribution of the data.

2.3. Bias Calculation for CMIP5 PFT Distribution Errors

To quantify the bias in vegetation-atmosphere fluxes caused by inaccurate PFT distributions within the CMIP5 models, we evaluated modeled albedo (calculated as the ratio between upwelling (*rsus*) and downwelling (*rsds*) short-wave radiation; unitless), net primary productivity (NPP; kg C m^{-2} yr^{-1}), and transpiration (*tran*; mm yr^{-1}) from models that also contributed output for PFT fraction in the piControl experiment. We aggregated monthly mean albedo into an annual mean and monthly NPP and transpiration into 30 year averages of the annual sums averaged across the first 30 model years. Modeled albedo was available and calculated for all CMIP5 models analyzed here, but NPP output was only available for 10 of 12 models, and transpiration output was only available for 8 of 12 models (Table 2).

To estimate the impact of incorrect PFT distributions within individual ESMs, we evaluated the bias in the vegetation-atmosphere feedback variables albedo, NPP, and transpiration by analyzing the within-model relationships between climate and vegetation-atmosphere flux for each PFT. To assess these relationships, we identified ESM grid cells from the global simulations whose composition was more than 85% either evergreen or deciduous PFT cover, then calculated model-specific relationships among MAT, MAP, and albedo, NPP, or transpiration by fitting an Akima spline [*Akima*, 1970; *Gebhardt et al*., 2013]. Using each model's output for MAT and MAP, we then calculated the values of albedo, NPP, or transpiration that ought to have been modeled by using the ESV data to represent the actual PFT composition of the landscape. We calculated biases in albedo, NPP, and transpiration as the difference between the modeled ESM ecosystem-atmosphere flux variable for each grid cell and the "actual" calculated variable as a linear mixture of the actual PFTs from the ESV data set and their spline-fitted values of albedo, NPP, or transpiration. This method evaluated the direct within-model costs for inaccurately mapping evergreen and deciduous trees but did not evaluate secondary feedbacks between altered ecosystem-atmosphere feedbacks and MAT/MAP.

To examine the relationships among modeled PFT distributions, climate, and vegetation-atmosphere flux, we calculated the within-model elasticity ($e_{\text{var1-var2}}$) for variable pairs. A variable pair with high elasticity indicates that the response (var2) responds strongly to variability in the predictor (var1). We calculated $e_{\text{var1-var2}}$ for within-model variable pairs by relating var1 to var2 using a natural cubic spline ($g_{\text{var2}}(\text{var}_1)$), calculating sensitivity as the derivative of the spline at the median of var1 ($dg_{\text{var2}}/d(\text{var}_1)$) and standardizing the sensitivity by

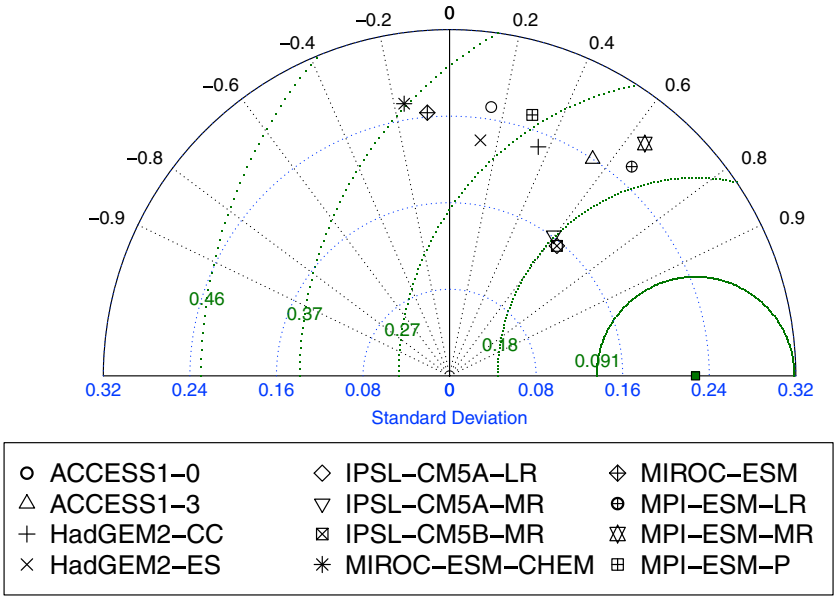

Figure 2. Taylor plot summarizing the correspondence between ESV settlement vegetation data set and simulated PFT distributions from the CMIP5 piControl experiment. Black symbols represent individual CMIP5 model simulations of deciduous PFT fraction; green square represents the ESV data set. The correlation between each CMIP5 simulation and ESV data is indicated by the radial branches (black numbers), with most models having moderately positive correlations (0.2 to 0.6) to the data and two with slightly negative (−0.2) correlations. Spatial heterogeneity in mapped PFTs is indicated by distance from the circle origin (blue numbers), with the ESV data set having a standard deviation of 0.22 and most simulations having similar or reduced spatial heterogeneity. In this visualization, smaller straight-line distances between model points and the green square (green numbers) indicate closer correspondence between modeled PFT fraction and the ESV data set.

mean of var1 and var2, $\frac{\frac{dg_{var2}}{d(var_1)}}{(\overline{var_2}/\overline{var_1})}$ [*LeBauer et al.*, 2013]. Because $e_{var1\text{-}var2}$ is normalized by the predictor and response means, the magnitude is comparable across variables of different units. We calculated within-model elasticity of (1) PFT fraction to climate variables ($e_{MAT/MAP\text{-}PFT}$), (2) vegetation-atmosphere flux to climate ($e_{MAT/MAP\text{-}albedo/NPP/tran}$), and (3) vegetation-atmosphere flux to PFT fraction ($e_{PFT\text{-}albedo/NPP/tran}$). Through this set of relationships, we assessed the relative importance of the mechanisms that link PFT fraction, climate, and vegetation-atmosphere flux within individual ESMs. The R code used for the entire analysis within this paper is available at http://github.com/jhmatthes/CMIP5_ESV_historicalPFTs/ and used several R packages for

Table 3. Summary of Data-Model Comparisons[a]

Model	Evgrn Niche Corr	Decid Niche Ccorr	MAT Diff (°C)	MAP Diff (mm/yr)	Albedo Bias (%)	NPP Bias (%)	Transpir Bias (%)
ACCESS1.0	0.67	**0.72**	**1.30**	**197**	**11.9**	-	-
ACCESS1.3	0.49	**0.63**	**2.76**	**274**	−0.02	-	-
HadGEM2-CC	**0.76**	**0.83**	−0.35	111	**12.2**	−8.0	-
HadGEM2-ES	**0.76**	**0.85**	0.89	125	**25.2**	−1.3	-
IPSL-CM5A-LR	**0.71**	**0.71**	-1.35	168	5.3	**25.8**	**35.2**
IPSL-CM5A-MR	**0.75**	**0.83**	0.23	182	2.1	**14.5**	**27.8**
IPSL-CM5B-LR	0.62	**0.60**	**−2.33**	184	2.1	**20.6**	**17.0**
MIROC-ESM-CHEM	**0.19**	**0.63**	**3.50**	100	−0.6	**−12.4**	0.2
MIROC-ESM	**0.18**	**0.66**	**3.24**	103	3.7	**−8.4**	−0.05
MPI-ESM-LR	0.64	**0.61**	−1.07	**262**	0.9	2.0	1.9
MPI-ESM-MR	0.68	**0.58**	−0.68	**309**	2.8	4.3	1.0
MPI-ESM-P	0.76	**0.61**	−0.54	**251**	3.9	8.2	7.6

[a]Niche corr indicates the Hellinger index of niche overlap between observed and simulated evergreen (Evgrn) and deciduous (Decid) PFTs. MAT and MAP diff indicates the difference between CRUNCEP climatology and CMIP5 modeled climatology. The last three columns indicate the percent biases in albedo, net primary productivity (NPP), and transpiration (Transpir) for some of the CMIP5 models. Boldface indicates significantly different from random for the niche correlation test, and a mean difference greater than 2 standard deviations from zero for climate differences and flux biases. See supporting information Table S2 for the full distribution statistics and absolute magnitudes of modeled biases.

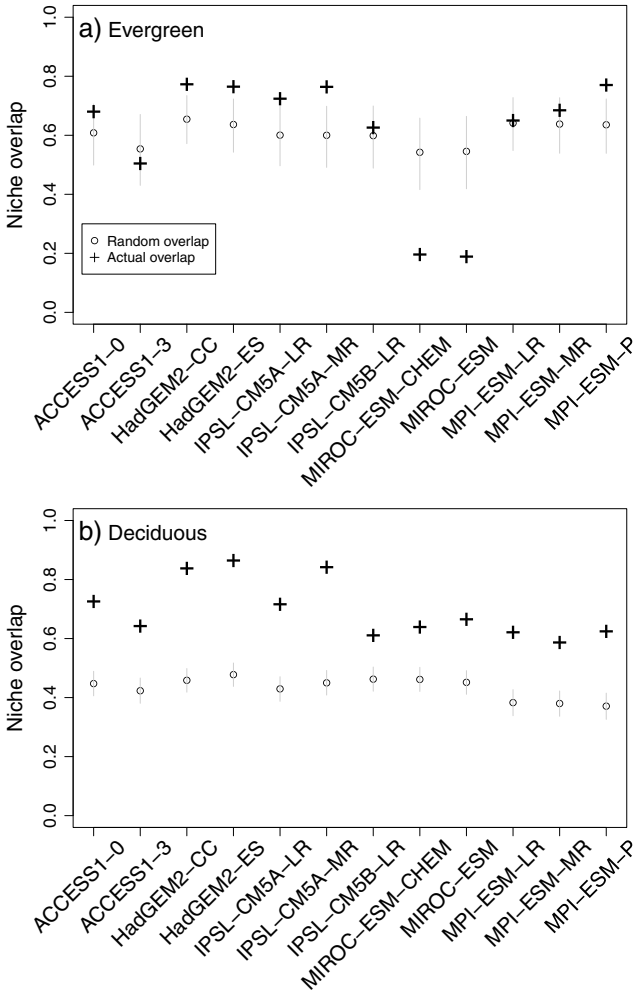

Figure 3. Hellinger distance niche overlap. The Hellinger distance measures the pairwise overlap between taxonomic distributions when plotted in climate space, with 1 indicating perfect overlap and 0 indicating no overlap. Here niche overlap between the ESV data set and CMIP5 models is assessed for (a) evergreen and (b) deciduous PFTs in a bivariate climate space represented by mean annual temperature and mean total annual precipitation, where grey dots and bars indicate the bootstrapped mean and 95% confidence interval for the randomized data set, and crosses indicate the actual niche overlap. For deciduous PFTs, niche overlap was consistently better than random, indicating skillful predictive ability by CMIP5 simulations. For the evergreen PFT, five models achieved better than random predictive skill, five were indistinguishable from random, and two models had significantly worse than random overlap.

plotting and data manipulation that are described within the code [*Furrer and Gerber*, 2014; *Soetaert*, 2014; *Becker et al.*, 2015; *Bivand et al.*, 2015; *Hijmans et al.*, 2015; *Nychka et al.*, 2015; *Pebesma et al.*, 2015; *Pierce*, 2015; *Plate and Heiberger*, 2015].

3. Results

3.1. ESV and CMIP5 PFT Spatial Comparison

Models varied widely in their ability to accurately capture the spatial pattern of PFTs in the ESV data set (Figure 2). Some of the discrepancy between ESV and CMIP5 PFTs can be attributed to the higher native resolution of the ESV data set and its finer-scale detail in capturing actual PFT distributions, which persists even after spatial aggregation to 0.5°. Data-model discrepancies persist and are similar in magnitude at coarser spatial resolutions (supporting information Figure S1). Data-model discrepancies cannot solely be attributed to issues of spatial resolution, and two models were anticorrelated with the spatial patterns in the ESV data (Figure 2). While simulations varied widely in their ability to capture underlying PFT distributions, the standard deviations of PFT abundance closely matched that of the ESV data, with a standard deviation in regional PFT distribution of approximately 0.24. However, the spatial correlation and centered root-mean-square difference varied more broadly across models (Figure 2). This indicates that, in general, the ESMs accurately represented the variability in PFT fraction across the region but often failed to accurately map the spatial locations of PFTs.

3.2. ESV and CMIP5 PFT-Climate Relationships

We examined the relationship between PFT abundance and MAT and MAP for both the ESV data set and the CMIP5 models to determine whether the data-model discrepancies were the result of (1) differences in modeled climates or (2) inaccuracies in model PFT-climate relationships. Diagnostic analyses indicated that these data-model discrepancies are due to inaccuracies in both the simulated climates and the model PFT-climate relationships. Many of the modeled CMIP5 climates varied significantly from the CRUNCEP-observed (1901–1930) climate; 6 of the 12 CMIP5 models had significant differences in regional MAT, and 5 of 12 models had significant

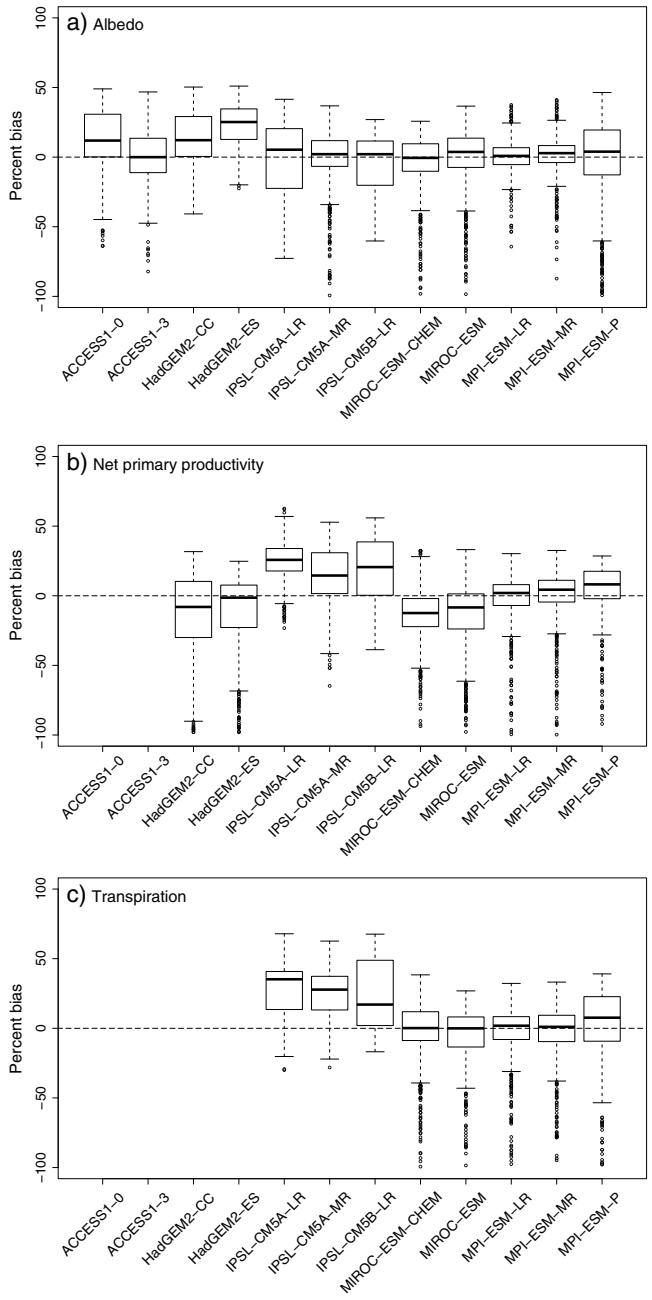

Figure 4. Biases in vegetation-atmosphere flux due to errors in PFT distributions. The percent grid cell bias in albedo, net primary productivity, and transpiration due to inaccurately simulated PFT distributions are plotted here for the spatial domain of the ESV data set. Box-and-whisker plots indicate 5%, 95%, 25%, and 75%, and the median of the distributions across all modeled grid cells and outliers that extend beyond the 95% range are indicated as dots. Most models had locally compensating errors in regional vegetation-atmosphere feedbacks, where the mean regional bias was not significantly different from zero. Models are absent from bias plots if output variables were not available in the Earth System Grid.

differences in regional MAP (supporting information Figure S2). Differences between CMIP5 models and CRUNCEP regional MAT ranged from −2.33°C to 3.5°C, and all CMIP5 models had higher regional MAP than the CRUNCEP data product by 100–300 mm yr^{-1} across the study region (Table 3).

Observed and modeled PFT-climate relationships also differed significantly. Across all CMIP5 models, correspondence between the ESV-CRUNCEP climate niches and the CMIP5 modeled climate niches were generally higher for the deciduous PFT than the evergreen PFT (Figure 3). This is due to the broader climate space and more diverse set of taxa covered by the deciduous PFT within this region during the study period, compared with the more constrained climate space and smaller set taxa within the evergreen PFT, thus constituting a more rigorous target to match. Deciduous PFT niche overlap between the ESV and all CMIP5 piControl models was significantly better than random (Figure 3). Niche overlap between the ESV and all CMIP5 piControl models for the evergreen PFT was better than random for five models, indistinguishable from random for five models, and worse than random for two models. Average niche overlap (i.e., the Hellinger distance) across all models with the ESV data set in CRUNCEP climate space was 0.60 for the evergreen PFT and 0.69 for the deciduous PFT.

3.3. Bias in Ecosystem-Atmosphere Feedback Variables

CMIP5 PFT distribution errors for evergreen and deciduous PFTs did not always produce significant directional biases in land-atmosphere feedback variables at the regional scale (Figure 4). Local biases, where individual cells showed large predictive errors, tended to be compensated elsewhere in the study region. This resulted in high regional variability but reduced regional-scale systematic bias. Mean regional albedo bias ranged from −0.6 to 25.2% across models, but 3 of the 12 models had a significant positive regional albedo bias (Table 3 and Figure 4a). The mean absolute albedo bias (−0.01–0.04) was similar

Table 4. Summary of the Elasticity ($e_{var1\text{-}var2}$) of Response Variables (var2) to Predictor Variables (var1) Within Individual CMIP5 Models for Relationships Between Mean Annual Temperature (MAT), Mean Annual Precipitation (MAP), Plant Functional Type Fraction (PFT), Albedo (alb), Net Primary Productivity (npp), and Transpiration (tran)

Model	MAT-PFT	MAP-PFT	MAT-alb	MAT-npp	MAT-tran	MAP-alb	MAP-npp	MAP-tran	PFT-alb	PFT-npp	PFT-tran
ACCESS1-0	1.39[b]	1.83[c]	0.68[a]			0.82[a]			0.43[a]		
ACCESS1-3	1.16[b]	1.36[b]	0.79[a]			0.93[a]			0.08[a]		
HadGEM2-CC	1.11[b]	1.77[c]	0.32[a]	1.09[b]		0.51[a]	1.65[c]		0.30[a]	1.18[b]	
HadGEM2-ES	1.49[b]	1.74[c]	0.38[a]	1.49[b]		0.27[a]	1.62[c]		0.27[a]	1.30[b]	
IPSL-CM5A-LR	0.28[a]	0.26[a]	0.16[a]	0.32[a]	0.35[a]	0.13[a]	0.42[a]	0.49[a]	0.63[a]	1.11[b]	1.43[b]
IPSL-CM5A-MR	0.42[a]	0.53[a]	0.27[a]	0.42[a]	0.45[a]	0.32[a]	0.51[a]	0.55[a]	0.43[a]	1.26[b]	0.73[a]
IPSL-CM5B-LR	0.26[a]	0.34[a]	0.19[a]	0.28[a]	0.27[a]	0.24[a]	0.36[a]	0.33[a]	0.79[a]	1.11[b]	1.13[b]
MIROC-ESM-CHEM	1.31[b]	0.91[a]	1.15[b]	1.31[b]	1.26[b]	0.81[a]	0.91[a]	0.88[a]	0.88[a]	1.00[b]	0.97[a]
MIROC-ESM	1.04[b]	0.83[a]	1.12[b]	1.11[b]	1.06[b]	0.90[a]	0.88[a]	0.86[a]	1.08[b]	1.07[b]	1.03[b]
MPI-ESM-LR	0.26[a]	0.63[a]	0.22[a]	0.51[a]	0.62[a]	0.69[a]	1.59[c]	1.92[c]	0.07[a]	0.93[a]	0.78[a]
MPI-ESM-MR	0.06[a]	0.41[a]	0.27[a]	0.53[a]	0.65[a]	1.07[b]	1.84[c]	2.30[d]	0.17[a]	1.47[b]	1.18[b]
MPI-ESM-P	1.35[b]	4.81[d]	0.43[a]	0.58[a]	0.82[a]	1.46[b]	1.38[b]	1.95[c]	0.09[a]	0.19[a]	0.24[a]

[a] $e_{var1\text{-}var2} < 1.0$.
[b] $e_{var1\text{-}var2} < 1.5$.
[c] $1.5 < e_{var1\text{-}var2} < 2.0$.
[d] $e_{var1\text{-}var2} > 2.0$.

to the range of global bias calculated by an analysis of Max-Planck-Institut (MPI)-ESM against a contemporary Moderate Resolution Imaging Spectroradiometer satellite albedo data set [*Brovkin et al.*, 2013].

Regional mean bias in NPP across the 10 CMIP5 models with reported NPP ranged from −12.4 to 25.8%, where 5 CMIP5 models significantly differed from 0 (Table 3 and Figure 4b). The three Institut Pierre-Simon Laplace (IPSL) models systematically overestimated NPP by 14.5 to 25.8% across the domain as a consequence of spatial incongruities between the evergreen and deciduous PFTs relative to ESV distributions, with lower predicted evergreen proportions resulting in higher deciduous PFT proportions, which have higher net NPP contributions within the models relative to the evergreen PFT.

The bias in regional NPP across models triggered a concomitant bias in transpiration from −0.05 to 35.2% across the CMIP5 models, where three of the eight CMIP5 models that contributed transpiration output had positive regional biases that were significantly different from zero (Table 2 and Figure 4c). The five models with significant transpiration biases also had significant positive biases in NPP as a result of mapping differences between the ESV and CMIP5 PFTs due to the mechanistic coupling between plant productivity and evapotranspiration.

3.4. Elasticity Between CMIP5 PFT, Climate, and Ecosystem-Atmosphere Flux Variables

The modeled variable pairs with the largest elasticity varied across ESMs in this study. For the ACCESS, Hadley, and MIROC models, *e* was highest for the response of PFT fraction to climate, where $e_{MAP\text{-}PFT}$ was largest in ACCESS and Hadley models (1.36–1.83), and $e_{MAT\text{-}PFT}$ was largest in the MIROC models (1.04 and 1.31; Table 4). However, models that had lower elasticity between climate variables and PFT fraction more accurately simulated PFT fractions, with the exception of the ACCESS1-3 model (Figure 2). This could indicate that the ACCESS1-0, Hadley, and MIROC models had errors in PFT fraction primarily because these models simulated relationships that were particularly sensitive to regional climate variability.

Vegetation-atmosphere fluxes were not especially sensitive to MAT (median $e_{MAT\text{-}flux} = 0.52$) with the exception of models that also had high elasticity for climate-PFT relationships, where $e_{MAT\text{-}NPP}$ was 1.09 and 1.45 in the Hadley models and $e_{MAT\text{-}flux}$ for all vegetation-atmosphere flux variables in the MIROC models were 1.06–1.31 (Table 4). Within the majority of ESMs, vegetation-atmosphere fluxes were more sensitive to MAP (median $e_{MAP\text{-}flux} = 0.87$) with notably high $e_{MAP\text{-}flux}$ among the MPI models (0.69–2.30; Table 4). In general, across all models albedo had low elasticity to PFT fraction but was highly variable ($e_{PFT\text{-}albedo} = 0.07$–$1.08$); NPP and transpiration were more elastic in response to PFT fraction among ESMs ($e_{PFT\text{-}NPP} = 0.93$–$1.47$ and $e_{PFT\text{-}tran} = 0.73$–$1.43$), with the exception of the MPI-P model ($e_{PFT\text{-}NPP} = 0.19$ and $e_{PFT\text{-}tran} = 0.24$; Table 4). Within the IPSL models, the elasticity of NPP and transpiration to PFT fraction was the highest among the tested variable pairs (median $e_{PFT\text{-}NPP/tran} = 1.12$).

4. Discussion

4.1. Accuracy of Historical PFT Mapping in CMIP5 Models

CMIP5 models varied widely in their ability to capture historical regional Euro-American settlement patterns of evergreen and deciduous PFTs (Figure 2), due to differences in modeled climate, PFT-climate relationships, and in internal elasticities related to connections among climate, PFT distributions, and vegetation-atmosphere flux. The ESV data set for the Northeastern United States [*Paciorek et al*., 2016] provides a regional-scale benchmark at a critical time period, which acts as the baseline before the start of the industrial era in the CMIP experiments, binding data from the Upper Midwest [*Goring et al*., 2015] through the East Coast of the United States [*Cogbill et al*., 2003; *Thompson et al*., 2013] before European settlers initiated major land use changes. Local-scale variance in the ESV is partly the result of the higher spatial resolution and resampling process and also because historical contingency plays a role in vegetation structure on the landscape [*Jackson et al*., 2009]. Differences between PFT distributions within the ESV data set and CMIP5 models could support the need for incorporating more nuanced environmental drivers into potential vegetation calculations including subgrid heterogeneity in topography, soils, and hydrology, as well as historical information about ecological disturbances like forest pests and pathogens, fires, and storm damage [*Schulte et al*., 2005] and the fact that high-level PFTs are not always natural ecological groupings. Environmental constraints and ecological disturbances prevent ecosystems from attaining their modeled potential vegetation states or provide multiple "potential states." Thus, refining the representation of ecological disturbances with additional historical data sets capturing past disturbance processes [e.g., *Marlon et al*., 2013] has the potential to mechanistically improve the correspondence between actual and modeled historical vegetation distributions resulting in a more accurate representation of land-climate feedbacks [*Han et al*., 2012], while refined ecological groupings can further improve taxon-climate relationships and generate a broader ecological palette with which to map potential vegetation. Additional working groups within the PalEON project (http://paleonproject.org) are currently synthesizing data sets describing the role of historical ecological disturbances in shaping historical vegetation patterns in this region at spatial scales that are relevant for Earth system models.

4.2. Relationships Among PFT Fraction, Climate, and Vegetation-Atmosphere Flux

This analysis also highlighted significant room for improvement in the current representation of modeled PFT-climate relationships by CMIP5 models. Models varied in their ability to capture PFT-climate niche relationships across this region, particularly for evergreen trees that are more narrowly constrained in both climate space and taxonomic diversity within this region (Figure 3 and Table 1). Furthermore, ESMs varied widely in their elasticity between modeled PFT fraction and modeled climate (0.06–1.49 for $e_{MAT\text{-}PFT}$; 0.26–4.81 for $e_{MAP\text{-}PFT}$; Table 4), suggesting a lack of consensus among models regarding the sensitivity of PFT distributions to climate variability. The disparity among these relationships reflects a potential area for future model-data benchmarking to examine constraints between modeled climate and PFT fraction, particularly as more models develop the capacity to simulate the dynamic response of vegetation to climate.

Most efforts to parameterize PFT distributions in ESMs have focused on contemporary distributions of leaf traits in climate space from field data typically collected over the past 10–20 years that are based on broad-scale physiological constraints [*Harrison et al*., 2010]. These now provide a rich data set for understanding the current biogeography of plant traits (GLOPNET [*Wright et al*., 2004] and TRY [*Kattge et al*., 2011]). However, assimilation of these global PFT traits into ESMs is still at the early stages [*Bonan et al*., 2012], and improving the mechanistic relationships between PFTs and climate, through both trait data set analysis and model simulation, remains an area of developing research [*McMahon et al*., 2011; *Fisher et al*., 2015]. The analysis within this paper highlights the potential for building upon trait-based PFT-climate analyses collected at one point in time to incorporating historical data that can further refine our understanding of PFT-climate relationships through time, at the early stages of the Anthropocene. The ESV data set presented here is part of a larger historical data set. The Public Land Survey covers most of the contiguous United States; thus, efforts to use this data to improve models will rely on greater interaction between modelers and historical and paleoecologists who work directly with the data to understand not only the nature of the historical data but the implications of shifts between historical and modern forest composition and structure, including (but not limited to) the widespread loss of forest types in the Upper Midwest and the advent of novel forest types, previously unknown in the region [*Goring et al*., 2015].

Although historical data sets are commonly perceived to be more coarsely resolved than contemporary data sets, the ESV data set [*Goring et al.*, 2015; *Paciorek et al.*, 2016] offers a remarkable benchmark data set. It achieves an 8 km spatial resolution and genus-level taxonomic resolution, which is higher than the resolution of the models studied here. In addition to the ESV PFT fraction used in this analysis, PalEON researchers are working to produce statistical estimates of biomass and stem density for the same regional domain. Synthesis of other historical data sets similar to the ESV data (Sweden [*Östlund et al.*, 1997], Congo [*Brncic et al.*, 2007], and central Europe [*Bürgi and Gimmi*, 2007]) could also be used in future analyses to further test historical PFT mapping within ESMs. These high-quality data sets provide a rigorous test for ESMs to capture regional patterns in historical vegetation states, which can help to constrain centennial-scale ecosystem-climate feedbacks.

4.3. Biases in Vegetation-Atmosphere Flux From PFT Mapping Errors

Biases in land-atmosphere exchange can initiate important global changes through teleconnections within global circulation models [*Swann et al.*, 2014] or the creation of initialization biases that persist through 21st century simulations and improving the accuracy of historical PFT distributions within CMIP5 models has the potential to improve vegetation-atmosphere feedback differences created by discrepancies in mapped PFTs (Figure 4). Local differences in albedo, NPP, and transpiration within individual grid cells can play an important role for downscaling large-scale ESM output to regional climate scenarios [*Hawkins and Sutton*, 2009] and deducing impacts of historical climate on local ecological processes [*Harris et al.*, 2014]. For some CMIP5 models, these local-scale discrepancies scaled up to systematic regional biases in ecosystem-climate feedback variables (Figure 4). These within-model biases did not always correlate with models that had the highest error in mapping PFTs (Figure 2).

Within the ESMs, NPP and transpiration had much more elastic relationships with climate and PFT fraction (median $e_{\mathrm{PFT/MAT/MAP\text{-}NPP/tran}} = 0.95$) than albedo did (median $e_{\mathrm{PFT/MAT/MAP\text{-}albedo}} = 0.43$; Table 4). This supports recent findings that indicate that not all model-data mismatch in the historical record propagates into future climate sensitivities at a global scale [*Schmidt et al.*, 2014]. However, for some ESMs, modeled NPP and transpiration were much more elastic in response to modeled PFT fraction, propagating into systematic regional biases in vegetation-atmosphere flux as a result of this within-model sensitivity. For example, while the IPSL models were among the most skillful at simulating PFTs (Figure 2), modeled NPP and transpiration had a relatively high elasticity with PFT fraction within this suite of models, whereas all other relationships among PFT distributions, climate, and vegetation-atmosphere flux were less elastic (Table 4). Thus, the biases in modeled NPP and transpiration for the IPSL suite responded with much more sensitivity to these small data-model differences in PFT distribution (Figure 4), due to the lack of elasticity among other drivers within this region (Table 4). This highlights the importance of considering the complex within-model connections and relative elasticity among modeled PFT fractions, climate, and vegetation-atmosphere flux.

Biases in NPP caused by inaccurate PFT distributions were of similar magnitude to several phenomena examined by *Anav et al.*, 2013: (1) differences in annual NPP between CMIP5 models, (2) a global atmospheric inversion for the Northern Hemisphere seasonal pattern in net biome productivity from 1986 to 2005, and (3) the seasonal pattern in gross primary productivity compared with the global FLUXNET database of eddy covariance measurements. In *Anav et al.*, 2013, differences between CMIP5 modeled carbon fluxes and the data sets were attributed to differences in the representation of nutrient limitations and ozone impacts on vegetation. However, discrepancies in mapping PFTs through space likely also contribute to these errors, because the proportion of deciduous versus evergreen trees on the landscape exerts a first-order control on the seasonal pattern of carbon flux [*Bonan*, 2008]. Since plant productivity and evapotranspiration fluxes are mechanistically coupled, we would expect that the transpiration fluxes would exhibit a similar range of biases in comparison with the plant productivity fluxes, but large differences in reference data sets for modern evapotranspiration have presented challenges for systematically assessing differences among CMIP5 models [*Schwalm et al.*, 2013].

5. Conclusions

These results indicate that model differences in internal relationships between climate, PFT distributions, and physiological variables are a primary cause of data-model discrepancies. Understanding and constraining the actual elasticity between PFT-climate relationships and emergent land-climate feedback variables is an area of critical need and promising future research both for modelers, field scientists, and paleohistorical scientists.

Historical data sets in particular offer a relatively underutilized opportunity for gaining mechanistic insights into the processes governing vegetation distributions and vegetation-climate feedbacks. Challenging ESMs with historical data sets that can benchmark land-climate feedbacks through time will play an important role in identifying and characterizing mechanistic improvements within ESMs. Better understanding of ecological interactions between PFTs and climate will be particularly important for improving the representation of decadal to centennial feedbacks between ecosystems and climate.

Acknowledgments
This work was supported by National Science Foundation grants DEB-1065848 and DEB-1065656. The authors are indebted to the work of the PalEON Settlement Vegetation team, and in particular C.V. Cogbill and D.J. Mladenoff, whose decades of research and experience with the TPS and PLS data made the compilation of the ESV data set possible. A. Thurman and C.J. Paciorek developed and ran the statistical model to produce domain-level estimates of ESV forest composition. The CMIP5 model output is available on the Earth System Grid (https://pcmdi9.llnl.gov/projects/esgf-llnl/); the PLS data set is available on GitHub (http://github.com/PalEON-Project/WitnessTrees), and code used to conduct the analyses within this manuscript is available on GitHub (http://github.com/jhmatthes/CMIP5_ESV_historicalPFTs/). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

References

Ahlström, A., G. Schurgers, A. Arneth, and B. Smith (2012), Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections, *Environ. Res. Lett.*, *7*(4), 44008, doi:10.1088/1748-9326/7/4/044008.

Akima, H. (1970), A new method of interpolation and smooth curve fitting based on local procedures, *J. Assoc. Comput. Mach.*, *17*(4), 589–602, doi:10.1145/321607.321609.

Anav, A., P. Friedlingstein, M. Kidston, L. Bopp, P. Ciais, P. Cox, C. Jones, M. Jung, R. Myneni, and Z. Zhu (2013), Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth system models, *J. Clim.*, *26*(18), 6801–6843, doi:10.1175/JCLI-D-12-00417.1.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman Hall, Boca Raton, Fla.

Becker, R. A., A. Wilks, R. Brownrigg, and T. P. Minka (2015), maps: Draw geographical maps.

Bivand, R., T. Keitt, B. Rowlingson, E. Pebesma, M. Sumner, R. J. Hijmans, and E. Rouault (2015), rgdal: Bindings for the geospatial data abstraction library.

Bonan, G. B. (2008), Forests and climate change: Forcings, feedbacks, and the climate benefits of forests, *Science*, *320*(5882), 1444–1449, doi:10.1126/science.1155121.

Bonan, G. B., K. W. Oleson, R. A. Fisher, G. Lasslop, and M. Reichstein (2012), Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4, *J. Geophys. Res.*, *117*, G02026, doi:10.1029/2011JG001913.

Brncic, T. M., K. J. Willis, D. J. Harris, and R. Washington (2007), Culture or climate? The relative influences of past processes on the composition of the lowland Congo rainforest, *Philos. Trans. R. Soc. B*, *362*(1478), 229–242, doi:10.1098/rstb.2006.1982.

Brovkin, V., L. Boysen, T. Raddatz, V. Gayler, A. Loew, and M. Claussen (2013), Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, *J. Adv. Model. Earth Syst.*, *5*(1), 48–57, doi:10.1029/2012MS000169.

Bürgi, M., and U. Gimmi (2007), Three objectives of historical ecology: The case of litter collecting in central European forests, *Landscape Ecol.*, *22*(1), 77–87, doi:10.1007/s10980-007-9128-0.

Cheddadi, R., and A. Bar-Hen (2009), Spatial gradient of temperature and potential vegetation feedback across Europe during the late Quaternary, *Clim. Dyn.*, *32*(2–3), 371–379, doi:10.1007/s00382-008-0405-7.

Cogbill, C. V., J. Burk, and G. Motzkin (2003), The forests of presettlement New England, USA: Spatial and compositional patterns based on town proprietor surveys, *J. Biogeogr.*, *29*, 1279–1304, doi:10.1046/j.1365-2699.2002.00757.x.

Evans, M. E. K., S. A. Smith, R. S. Flynn, and M. J. Donoghue (2009), Climate, niche evolution, and diversification of the "bird-cage" evening primroses (Oenothera, Sections Anogra and Kleinia), *Am. Nat.*, *173*, 225–240, doi:10.1086/595757.

Fisher, R. A., et al. (2015), Taking off the training wheels: The properties of a dynamic vegetation model without climate envelopes, CLM4.5 (ED), *Geosci. Model Dev.*, *8*(11), 3593–3619, doi:10.5194/gmd-8-3593-2015.

Friedlingstein, P., et al. (2006), Climate–carbon cycle feedback analysis: Results from the C4MIP Model Intercomparison, *J. Clim.*, *19*(14), 3337–3353, doi:10.1175/JCLI3800.1.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti (2013), Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, *J. Clim.*, *27*(2), 511–526, doi:10.1175/JCLI-D-12-00579.1.

Furrer, R., and F. Gerber (2014), spam: SPArse Matrix.

Gebhardt, A., T. Petzoldt, and M. Maechler (2013), akima: Interpolation of irregularly spaced data.

Goring, S., et al. (2015), Changes in forest composition, stem density, and biomass from the settlement era (1800s) to present in the Upper Midwestern United States, *bioRxiv*, doi:10.1101/026575.

Hall, A. (2004), The role of surface albedo feedback in climate, *J. Clim.*, *17*(7), 1550–1568, doi:10.1175/1520-0442(2004)017<1550:TROSAF>2.0.CO;2.

Han, Y. M., J. R. Marlon, J. J. Cao, Z. D. Jin, and Z. S. An (2012), Holocene linkages between char, soot, biomass burning and climate from Lake Daihai, China, *Global Biogeochem. Cycles*, *26*, GB4017, doi:10.1029/2011GB004197.

Harris, R. M. B., M. R. Grose, G. Lee, N. L. Bindoff, L. L. Porfirio, and P. Fox-Hughes (2014), Climate projections for ecologists, *Wiley Interdiscip. Rev. Clim. Change*, *5*(5), 621–637, doi:10.1002/wcc.291.

Harrison, S. P., I. C. Prentice, D. Barboni, K. E. Kohfeld, J. Ni, and J.-P. Sutra (2010), Ecophysiological and bioclimatic foundations for a global plant functional classification, *J. Veg. Sci.*, *21*(2), 300–317, doi:10.1111/j.1654-1103.2009.01144.x.

Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bull. Am. Meteorol. Soc.*, *90*(8), 1095–1107, doi:10.1175/2009BAMS2607.1.

Heibl, C., and C. Calenge (2013), phyloclim: Integrating phylogenetics and climatic niche modeling.

Hijmans, R. J., J. van Etten, J. Cheng, M. Mattiuzzi, M. Sumner, J. A. Greenberg, O. P. Lamigueiro, A. Bevan, E. B. Racine, and A. Shortridge (2015), raster: Geographic data analysis and modeling.

Hurtt, G. C., et al. (2011), Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, *Clim. Change*, *109*(1–2), 117–161, doi:10.1007/s10584-011-0153-2.

Jackson, S. T., J. L. Betancourt, R. K. Booth, and S. T. Gray (2009), Ecology and the ratchet of events: Climate variability, niche dimensions, and species distributions, *Proc. Natl. Acad. Sci. U.S.A.*, *106*(Supplement 2), 19,685–19,692, doi:10.1073/pnas.0901644106.

Kattge, J., et al. (2011), TRY—A global database of plant traits, *Global Change Biol.*, *17*(9), 2905–2935, doi:10.1111/j.1365-2486.2011.02451.x.

LeBauer, D., D. Wang, K. Richter, C. Davidson, and M. C. Dietze (2013), Facilitating feedbacks between field measurements and ecosystem models, *Ecol. Monogr.*, *83*(2), 133–154, doi:10.1890/12-0137.1.

Liu, F., D. J. Mladenoff, N. S. Keuler, and L. S. Moore (2011), Broadscale variability in tree data of the historical Public Land Survey and its consequences for ecological studies, *Ecol. Monogr.*, *81*(2), 259–275, doi:10.1890/10-0232.1.

Marlon, J. R., P. J. Bartlein, A.-L. Daniau, S. P. Harrison, S. Y. Maezumi, M. J. Power, W. Tinner, and B. Vanniére (2013), Global biomass burning: A synthesis and review of Holocene paleofire records and their controls, *Quat. Sci. Rev.*, *65*, 5–25, doi:10.1016/j.quascirev.2012.11.029.

McCulloch, R., and P. Rossi (1994), An exact likelihood analysis of the multinomial probit model, *J. Econ.*, *64*, 207–240, doi:10.1016/0304-4076(94)90064-7.

McMahon, S. M., S. P. Harrison, W. S. Armbruster, P. J. Bartlein, C. M. Beale, M. E. Edwards, J. Kattge, G. Midgley, X. Morin, and I. C. Prentice (2011), Improving assessment and modelling of climate change impacts on global terrestrial biodiversity, *Trends Ecol. Evol.*, *26*(5), 249–259, doi:10.1016/j.tree.2011.02.012.

Meehl, G. A., et al. (2009), Decadal prediction, *Bull. Am. Meteorol. Soc.*, *90*(10), 1467–1485, doi:10.1175/2009BAMS2778.1.

Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*(6), 693–712, doi:10.1002/joc.1181.

Munoz, S. E., D. J. Mladenoff, S. Schroeder, and J. W. Williams (2014), Defining the spatial patterns of historical land use associated with the indigenous societies of eastern North America, *J. Biogeogr.*, *41*(12), 2195–2210, doi:10.1111/jbi.12386.

Nychka, D., R. Furrer, and S. Sain (2015), fields: Tools for spatial data.

Östlund, L., O. Zackrisson, and A.-L. Axelsson (1997), The history and transformation of a Scandinavian boreal forest landscape since the 19th century, *Can. J. For. Res.*, *27*(8), 1198–1206, doi:10.1139/x97-070.

Paciorek, C., S. Goring, A. Thurman, C. Cogbill, J. Williams, D. Mladenoff, J. Peters, J. Zhu, and J. McLachlan (2016), Settlement-Era Gridded Tree Composition, Northeastern US: Level 2. Long Term Ecological Research Network, doi:10.6073/pasta/8544e091b64db26fdbbbafd0699fa4f9.

Pebesma, E., R. Bivand, B. Rowlingson, V. Gomez-Rubio, R. J. Hijmans, M. Sumner, D. MacQueen, J. Lemon, and J. O'Brien (2015), sp: Classes and methods for spatial data.

Piao, S., et al. (2013), Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends, *Global Change Biol.*, *19*(7), 2117–2132, doi:10.1111/gcb.12187.

Pierce, D. (2015), ncdf: Interface to Unidata netCDF data files.

Plate, T., and R. Heiberger (2015), abind: Combine multidimensional arrays.

Pongratz, J., T. Raddatz, C. H. Reick, M. Esch, and M. Claussen (2009), Radiative forcing from anthropogenic land cover change since A.D. 800, *Geophys. Res. Lett.*, *36*, L02709, doi:10.1029/2008GL036394.

Ramankutty, N., and J. A. Foley (1999), Estimating historical changes in global land cover: Croplands from 1700 to 1992, *Global Biogeochem. Cycles*, *13*(4), 997–1027, doi:10.1029/1999GB900046.

Rhemtulla, J. M., D. J. Mladenoff, and M. K. Clayton (2009), Historical forest baselines reveal potential for continued carbon sequestration, *Proc. Natl. Acad. Sci. U.S.A.*, *106*(15), 6082–6087, doi:10.1073/pnas.0810076106.

Rue, H., and L. Held (2005), *Gaussian Markov Random Fields: Theory and Applications*, Chapman Hall, Boca Raton, Fla.

Schmidt, G. A., et al. (2014), Using palaeo-climate comparisons to constrain future projections in CMIP5, *Clim. Past*, *10*(1), 221–250, doi:10.5194/cp-10-221-2014.

Schulte, L., D. Mladenoff, S. Burrows, T. Sickley, and E. Nordheim (2005), Spatial controls of Pre–Euro-American wind and fire disturbance in Northern Wisconsin (USA) forest landscapes, *Ecosystems*, *8*(1), 73–94, doi:10.1007/s10021-004-0052-8.

Schwalm, C. R., D. N. Huntinzger, A. M. Michalak, J. B. Fisher, J. S. Kimball, B. Mueller, K. Zhang, and Y. Zhang (2013), Sensitivity of inferred climate model skill to evaluation decisions: A case study using CMIP5 evapotranspiration, *Environ. Res. Lett.*, *8*(2), 24028, doi:10.1088/1748-9326/8/2/024028.

Sitch, S., et al. (2008), Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs), *Global Change Biol.*, *14*(9), 2015–2039, doi:10.1111/j.1365-2486.2008.01626.x.

Soetaert, K. (2014), plot3D: Plotting multi-dimensional data.

Swann, A. L. S., I. Y. Fung, Y. Liu, and J. C. H. Chiang (2014), Remote vegetation feedbacks and the mid-Holocene Green Sahara, *J. Clim.*, doi:10.1175/JCLI-D-13-00690.1.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192, doi:10.1029/2000JD900719.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2011), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thompson, J. R., D. N. Carpenter, C. V. Cogbill, and D. R. Foster (2013), Four centuries of change in Northeastern United States forests, *PLoS One*, *8*(9, e72540), doi:10.1371/journal.pone.0072540.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Univ. Press, Cambridge, U. K.

Warren, D. L., R. E. Glor, M. Turelli, and D. Funk (2008), Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution, *Evolution (N. Y.)*, *62*(11), 2868–2883, doi:10.1111/j.1558-5646.2008.00482.x.

Wei, Y., et al. (2014), NACP MsTMIP: Global and North American driver data for multi-model intercomparison. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge Natl. Lab. Distrib. Act. Arch. Center, Oak Ridge, Tenn. [Available at doi:10.3334/ORNLDAAC/1220, (Accessed 12 April 2014).]

Wright, I. J., et al. (2004), The worldwide leaf economics spectrum, *Nature*, *428*(6985), 821–827, doi:10.1038/nature02403.