SPECIAL C4MIP COLLECTION

Scale-Dependent Performance of CMIP5 Earth System Models in Simulating Terrestrial Vegetation Carbon*

LIFEN JIANG,[+] YANER YAN,[#] OLEKSANDRA HARARUK,[+] NATHANIEL MIKLE,[+] JIANYANG XIA,[+,**] ZHENG SHI,[+] JERRY TJIPUTRA,[@] TONGWEN WU,[&] AND YIQI LUO[+]

[+] *Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma*
[#] *School of Life Science and Institute of Wetland Ecology, Nanjing University, Nanjing, China*
[@] *Uni Research Climate, and Bjerknes Centre for Climate Research, Bergen, Norway*
[&] *Beijing Climate Center, China Meteorological Administration, Beijing, China*

(Manuscript received 9 April 2014, in final form 18 March 2015)

ABSTRACT

Model intercomparisons and evaluations against observations are essential for better understanding of models' performance and for identifying the sources of uncertainty in their output. The terrestrial vegetation carbon simulated by 11 Earth system models (ESMs) involved in phase 5 of the Coupled Model Intercomparison Project (CMIP5) was evaluated in this study. The simulated vegetation carbon was compared at three distinct spatial scales (grid, biome, and global) among models and against the observations (an updated database from Olson et al.'s "Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: A Database"). Moreover, the underlying causes of the differences in the models' predictions were explored. Model–data fit at the grid scale was poor but greatly improved at the biome scale. Large intermodel variability was pronounced in the tropical and boreal regions, where total vegetation carbon stocks were high. While 8 out of 11 ESMs reproduced the global vegetation carbon to within 20% uncertainty of the observational estimate (560 ± 112 Pg C), the simulated global totals varied nearly threefold between the models. The goodness of fit of ESMs in simulating vegetation carbon depended strongly on the spatial scales. Sixty-three percent of the variability in contemporary global vegetation carbon stocks across ESMs could be explained by differences in vegetation carbon residence time across ESMs ($P < 0.01$). The analysis indicated that ESMs' performance of vegetation carbon predictions can be substantially improved through better representation of plant longevity (i.e., carbon residence time) and its respective spatial distributions.

1. Introduction

A substantial proportion of annual anthropogenic carbon dioxide emission is sequestered by terrestrial vegetation (Canadell et al. 2007; Houghton 2007; Le Quéré et al. 2014), but this carbon sink may decrease in the future as a result of the positive feedback between climate change and the carbon cycle (Cox et al. 2000; Dufresne et al. 2002; Friedlingstein et al. 2003, 2006; Le Quéré et al. 2009). Therefore, understanding the mechanisms regulating the changes in the terrestrial vegetation sink is crucial to improve future climate projections. In their study, Arora et al. (2013) highlighted the larger inconsistencies in the carbon–climate and carbon–concentration feedback parameters in the land carbon cycle component than in the ocean simulated by Earth system models (ESMs) involved in phase 5 of the Coupled Model Intercomparison Project (CMIP5). In addition, there was a significant model spread in the twenty-first-century compatible CO_2 emissions simulated by CMIP5 ESMs under four representative concentration pathways (RCPs; Jones et al. 2013). This spread was dominated by the variability in the projected land carbon changes, which was partly due to the diverse responses of land carbon cycle models to anthropogenic CO_2 increase

* Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/JCLI-D-14-00270.s1.

** Current affiliation: Tiantong National Forest Ecosystem Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China.

Corresponding author address: Lifen Jiang and Yiqi Luo, Department of Microbiology and Plant Biology, University of Oklahoma, 101 David L. Boren Blvd., Norman, OK 73019.
E-mail: lfjiang@ou.edu; yluo@ou.edu

DOI: 10.1175/JCLI-D-14-00270.1

and climate change and the different representations of land-use change. Finally, Friedlingstein et al. (2014) also showed that the uncertainty in the land carbon cycle projection was responsible for the large spread in the atmospheric CO_2 projections under the RCP8.5 scenario when driven by CO_2 emissions.

Currently, large uncertainties still exist in simulations of global land carbon storage (including vegetation and soil carbon), as shown in both historical simulations and future simulations of CMIP5 ESMs (Anav et al. 2013; Brovkin et al. 2013; Jones et al. 2013; Todd-Brown et al. 2013, 2014). In general, CMIP5 ESMs showed far less agreement in their projections of land carbon changes than in those of ocean carbon (Jones et al. 2013). The simulated intermodel variation in the land carbon uptake within a specific RCP scenario was even larger than the variation across the four RCPs (Jones et al. 2013). Despite the historical simulations of CMIP5 ESMs correctly reproducing the main climatic variables controlling the spatial and temporal characteristics of the carbon cycle, the global soil carbon and vegetation carbon showed very large variation across the models, even though the multimodel means were close to the reference data (Anav et al. 2013). Global soil carbon storage varied by nearly sixfold, and global soil carbon turnover times varied by almost fourfold across CMIP5 ESMs in their historical simulations (Todd-Brown et al. 2013). The changes in soil organic carbon over the twenty-first century projected by CMIP5 ESMs ranged from a loss of 70 Pg C to a gain of 250 Pg C (Todd-Brown et al. 2014). Analysis of permafrost thermal dynamics in the CMIP5 ESMs also revealed a wide range in current permafrost areas, active layer parameters, and model ability to simulate the coupling between soil and air temperatures (Koven et al. 2013). Additionally, projected loss of permafrost extent in response to climate change also varied greatly between the models.

Evaluating the models' performance and understanding the sources of uncertainties in the simulated contemporary state of the land carbon cycle are essential steps forward to improve the credibility of future climate projections. Model evaluations using observations can help us identify uncertainties in predictions and provide feedback for future model development priorities (Friedlingstein et al. 2006; Blyth et al. 2011; Luo et al. 2012). Hoffman et al. (2014) found a linear relationship between the contemporary and the future atmospheric CO_2 mole fractions in 17 emission-driven CMIP5 model simulations. They further constructed a contemporary CO_2-tuned model to estimate the atmospheric CO_2 trajectory for the twenty-first century based on this linear relationship and the long-term time series of atmospheric CO_2 from Mauna Loa. As a result, the uncertainties in future projections of atmospheric CO_2 for the emission-driven RCP8.5 scenario were considerably narrowed compared to the estimates from the full ESM ensemble. This study (Hoffman et al. 2014) calls for additional efforts to carefully evaluate and improve the simulations of contemporary terrestrial carbon stocks before we can have confidence in future projections (Jones et al. 2013).

Carbon in the vegetation pool is more dynamic compared with carbon in soils (IGBP Terrestrial Carbon Working Group 1998). The ability of ESMs to reproduce the observed terrestrial vegetation carbon dynamics is, therefore, a good indicator for model performance and is a critical prerequisite to improve the land carbon cycle component. Anav et al. (2013) evaluated historical simulations of the land carbon cycle in CMIP5 ESMs at the global scale and in the Northern Hemisphere, Southern Hemisphere, and tropics. Todd-Brown et al. (2013) benchmarked historical simulations of soil carbon stocks from CMIP5 ESMs at the gridcell, biome, and global scales. Both analyses have shown that scale does affect model performances. The study by Anav et al. (2013) was the first attempt to evaluate historical simulations of terrestrial vegetation carbon in CMIP5 ESMs along with many other variables. While it is likely that there were interactions between the variables, Anav et al. (2013) did not link other variables to vegetation carbon in their evaluations to explore the possible causes for uncertainties in the simulated vegetation carbon. Focusing on vegetation carbon, we evaluated the historical simulations of terrestrial vegetation carbon in 11 CMIP5 ESMs involved in the IPCC Fifth Assessment Report (IPCC 2013) against an observation-based estimate of the global land vegetation carbon pool developed by Gibbs (2006). We focused our analysis on evaluating and quantifying the sensitivity of the model–data fit to various spatial scales and identifying the causes of the intermodel variability.

2. Materials and methods

We evaluated vegetation carbon in terrestrial ecosystems simulated by 11 CMIP5 ESMs (Table 1) applying the observation-based estimate at three spatial scales: grid cells, biomes, and global. At the gridcell scale, we evaluated vegetation carbon density (kilograms of carbon per square meter); at the other two scales, we evaluated the total carbon stocks (petagrams of carbon) at each biome and globally. We also analyzed annual net primary productivity (NPP; petagrams of carbon per year) and residence times of vegetation carbon at the global scale. NPP, plant mortality, and losses from disturbances are the determinants of carbon stock sizes, and the combined

impacts of the loss factors can be quantified with the derived estimate of residence time (Xia et al. 2013).

We evaluated 21 historical simulations from 11 modeling centers. Here, we show the results of 11 simulations (we selected one simulation from each modeling center). This would not change our main findings and conclusions because models from the same centers usually have similar performance (e.g., Fig. S5 in the supplementary material). However, the global vegetation carbon budget in IPSL-CM5B-LR was considerably lower than those in the other two models from their group (IPSL-CM5A-LR and IPSL-CM5A-MR). This could be attributed to the fact that IPSL-CM5B-LR had an improved climatology and current climate variability in tropical regions (Dufresne et al. 2013). However, the equilibrium climate sensitivity in IPSL-CM5B-LR was drastically different from that in IPSL-CM5A models, and IPSL-CM5B-LR was in its early development stages [i.e., not completely validated and not yet perfectly tuned for CMIP5 runs (J.-L. Dufresne 2013, personal communication; G. Krinner 2014, personal communication)]. We selected one model from each modeling center with the closest global sums of vegetation carbon to the observational data, except for CESM1(BGC), IPSL-CM5A-MR, and NorESM1-ME (the latter two were recommended by the respective modeling groups). The results of the other 10 ESMs along with the total 21 ESMs are available in the supplementary material.

a. CMIP5 ESMs

CMIP5 ESMs and their major differences in simulating vegetation carbon are described in Table 1 and Table S1 in the supplementary material. We downloaded the output for monthly carbon mass in land vegetation (variable cVeg) and carbon mass flux out of the atmosphere resulting from net primary production on land (variable npp) for their historical simulations. Land area fraction (variable sftlf) in model output was also downloaded to compute the global vegetation carbon and total vegetation carbon for each biome. We selected the ensemble member r1i1p1 because most CMIP5 models had this ensemble member output, while the availability of other ensembles depended on specific models used. The letters r, i, and p in the label for the ensemble member refer to the initial condition, initialization method, and perturbed physics version, respectively, and the number 1 after each letter is the realization for the respective parameter (Taylor et al. 2010, 2012). In the cases where multiple ensemble members were available from a single model, we examined the differences between the ensemble members and found that the outputs were similar (not shown). To ensure comparability with the observations, we evaluated the simulated vegetation carbon in the year 2000, because many of the CMIP5 ESMs incorporated land-use change and had dynamic vegetation (Tables 1 and S1). Because of the high interannual variability in NPP compared to vegetation carbon storage, we evaluated the 10-yr means of NPP from 1996 to 2005 from CMIP5 output.

b. Observational dataset

We used an observational dataset from Gibbs (2006) as the benchmark for model evaluations. Olson et al. (1985) developed a database and corresponding map following more than 20 years of field investigations, consultations, and analyses of the published literature. It is an integrated dataset that provides a reference baseline for interpreting the role of terrestrial vegetation in the global cycling of CO_2 and other gases and a basis for improving estimates of vegetation and soil carbon, natural exchanges of CO_2, and net historic shifts of carbon between the biosphere and the atmosphere (Gibbs 2006). To reflect the changes in land cover over time, Gibbs (2006) updated the Olson et al. (1985) database to a more contemporary land-cover representation using the Global Land Cover 2000 database (European Commission, Joint Research Centre 2003) to estimate biomass carbon in living vegetation on a global scale. Data from Global Land Cover 2000 were developed using remotely sensed imagery acquired in 2000. Compared to the original dataset by Olson et al. (1985), which was derived from vegetation patterns of preagricultural vegetation, this updated dataset accounted for changes in land cover subject to human-induced land-use change, forest harvest, and natural disturbances.

The global vegetation carbon stock of the dataset developed by Gibbs (2006) was 560 Pg C. The IPCC (Watson et al. 2000) reported global vegetation carbon stocks to be 470 Pg C, taking into account land-use change, whereas Saugier et al. (2001) reported global vegetation carbon at 650 Pg C. The above two estimates of global vegetation carbon stocks were a deviation of approximately 20% from 560 Pg C, therefore we took 20% of 560 Pg C (112 Pg C) as the uncertainty range for the dataset developed by Gibbs (2006), as this dataset did not provide uncertainties associated with the errors in measurements or estimates.

c. Biome data

To make a comparison between performance of ESMs in simulating vegetation carbon and soil carbon, we used the biome dataset developed by Todd-Brown et al. (2013) to examine how well the models performed in simulating the vegetation carbon at the biome scale. In brief, the land-cover classification from the MODIS land cover type Climate Modeling Grid (CMG) yearly global 0.05° product (MCD12C1) (NASA LP DAAC 2008) was assigned to one of nine biomes: tundra, boreal

TABLE 1. Summary of CMIP5 ESMs and their land carbon cycle components. (Expansions of model name and component acronyms are available at http://www.ametsoc.org/PubsAcronymList.)

Model name of ESM	Version	Modeling group	Land carbon cycle components	No. of PFTs	No. of live carbon pools	Dynamic vegetation	Nitrogen cycle	Original resolution (lat × lon)	References
BCC_CSM1.1	20120918	Beijing Climate Center (BCC), China Meteorological Administration	BCC Atmosphere and Vegetation Interaction Model, version 1.0 (BCC_AVIM1.0)	15	3	No	No	2.81° × 2.81°	Ji et al. (2008); Wu et al. (2013)
BNU-ESM	20120504	College of Global Change and Earth System Science, Beijing Normal University (BNU)	Common Land Model (CoLM) and BNU DGVM [based on Lund–Potsdam–Jena model (LPJ)]	10	4	Yes	No	2.81° × 2.81°	Dai et al. (2003, 2004); Ji et al. (2014)
CanESM2	20120410	Canadian Centre for Climate Modeling and Analysis	Canadian Land Surface Scheme, version 2.7 (CLASS2.7) and CTEM, version 1 (CTEM1)	9	3	No	No	2.81° × 2.81°	Arora and Boer (2010)
CESM1 (Biogeochemistry) [CESM1(BGC)]	20121029	Community Earth System Model Contributors	Community Land Model, version 4 (CLM4)	15	4	No	Yes	0.94° × 1.25°	Thornton and Zimmermann (2007); Thornton et al. (2007, 2009); Lawrence et al. (2011)
GFDL-ESM2G	20121206	Geophysical Fluid Dynamics Laboratory	Land Model, version 3 (LM3)	5	5	Yes	No	1.99° × 2.48°	Shevliakova et al. (2009); Dunne et al. (2013)
HadGEM2-ES	20111007	Met Office Hadley Centre	TRIFFID	5	3	Yes	No	1.24° × 1.88°	Cox (2001); Collins et al. (2011); Jones et al. (2011); Martin et al. (2011)
INM-CM4.0	20110323	Institute of Numerical Mathematics	Land Surface Model (LSM), version 1.0	12	3	No	No	1.50° × 2.00°	Bonan (1996); Volodin (2007)
IPSL-CM5A-MR	20120430	L'Institut Pierre-Simon Laplace	ORCHIDEE	12	8	No	No	1.26° × 2.50°	Dufresne et al. (2013); Krinner et al. (2005)

TABLE 1. (*Continued*)

Model name of ESM	Version	Modeling group	Land carbon cycle components	No. of PFTs	No. of live carbon pools	Dynamic vegetation	Nitrogen cycle	Original resolution (lat × lon)	References
MIROC-ESM	20120710	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	Minimal Advanced Treatments of Surface Interaction and Runoff (MATSIRO) and Spatially Explicit Individual-Based DGVM (SEIB-DGVM)	13	4	Yes	No	2.81° × 2.81°	Sato et al. (2007); Watanabe et al. (2011)
MPI-ESM-MR	20120503	Max Planck Institute for Meteorology	JSBACH	12	3	Yes	No	1.88° × 1.88°	Raddatz et al. (2007); Brovkin et al. (2009); Reick et al. (2013)
NorESM1-ME	20120225	Norwegian Climate Centre	CLM4	15	4	No	Yes	1.88° × 2.50°	Tjiputra et al. (2013)

forest, tropical rain forest, temperate forest, desert and shrub land, grasslands and savannas, cropland and urban, snow and ice, or permanent wetland on a 1° × 1° global grid (Fig. S1 in the supplementary material). In this classification, the temperate rain forest of the North American Pacific Northwest was included in the boreal forest biome.

d. MODIS NPP dataset

To explain the uncertainties in vegetation carbon between CMIP5 ESMs and the dataset developed by Gibbs (2006), we used the MODIS product (MOD17A3) for annual NPP from the year 2000 as a reference NPP. The data were produced by the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) at 1-km spatial resolution (Heinsch et al. 2003). MOD17A3 did not include an uncertainty analysis, but uncertainties could be large because of possible errors related to inputs of the algorithm for MOD17A3, including land cover, fraction of photosynthetically active radiation/leaf area index (FPAR/LAI), and other meteorological data (Zhao et al. 2005).

e. Residence time

We calculated vegetation carbon residence time at the global scale as the ratio between the global vegetation carbon pool and the global NPP for both the Gibbs (2006) dataset (using MODIS NPP) and output of CMIP5 ESMs.

f. Resolution

The resolution of original model outputs from CMIP5 ESMs was quite diverse (Tables 1 and S1). All evaluation analysis was performed after we regridded vegetation carbon and NPP data to 1° × 1° cells assuming conservation of mass. The calculation of the yearly average of vegetation carbon and NPP of model outputs and sums of vegetation carbon and NPP for each biome and globe, as well as the regridding of all data were performed with the NCAR Command Language, version 6.1.2 (UCAR/NCAR/CISL/VETS 2013).

g. Statistical analysis

The goodness of fit of vegetation carbon of ESMs simulated at the grid and biome scale was quantified by the coefficient of determination of linear regressions (R^2) and root-mean-square error (RMSE). Slopes and intercepts of linear regressions were also shown as additional indicators for bias. At the biome scale, fewer data points (eight biomes) may have some limitation on the linear regressions in our analysis. The calculation of RMSE followed Janssen and Heuberger's (1995) equation:

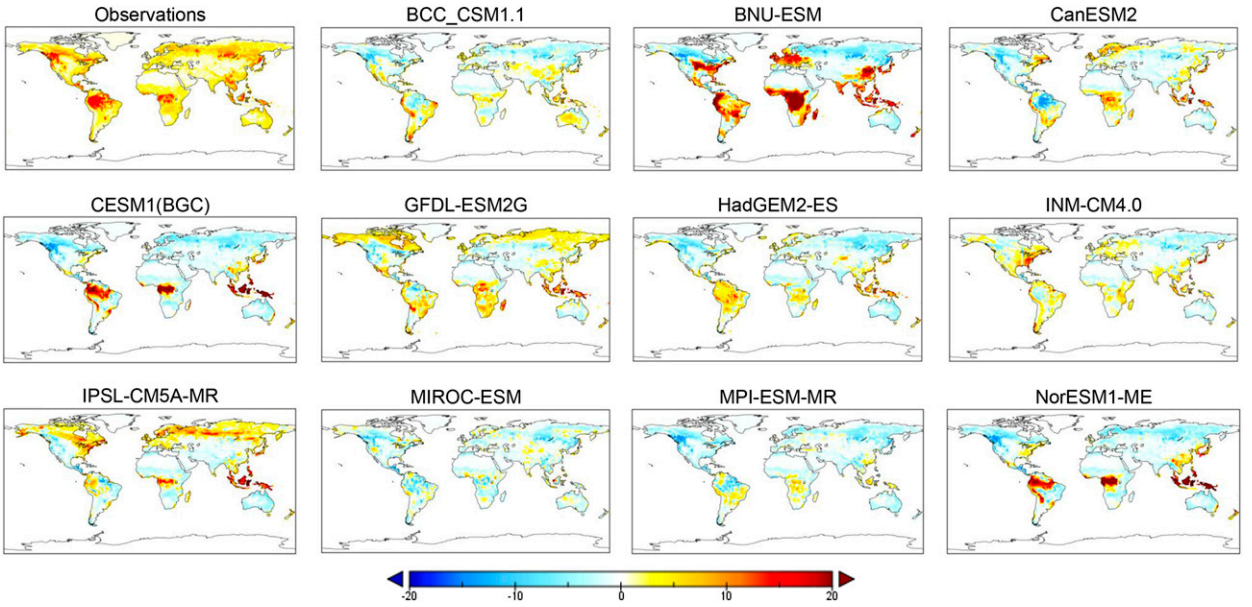

FIG. 1. Maps of vegetation carbon density of the observations and differences of vegetation carbon between historical simulations of CMIP5 ESMs and the observations (kg C m^{-2}).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}},$$

where P and O are modeled and observed data values, respectively, and N is the number of data points. The significance level of the linear relationships was analyzed with MATLAB. At the global scale, goodness of fit was assessed by the absolute differences in the global sums between ESMs and the observations.

3. Results

a. Goodness of fit of ESMs at grid scale

The spatial distribution of global vegetation carbon density simulated by ESMs showed consistent patterns to those in the observations (Fig. S2 in the supplementary material), with maximum density generally simulated in the tropical rain forest and Northern Hemisphere boreal forest biomes. Analogously, the biggest biases in biomass carbon between ESMs and the observations were located in these regions because of large vegetation carbon stocks there (Figs. 1 and S3). Overall, simulations of vegetation carbon density by ESMs were poor at the grid scale, with large differences in absolute values between ESMs and the observations at grid cells and with low R^2 (Table 2; Figs. 1, 2 and Figs. S3–S5 in the supplementary material). RMSE ranged from 3.01 to 7.18, which generally agreed with R^2 (i.e., higher R^2, lower RMSE). The maximum values of vegetation carbon densities at grid cells varied substantially across ESMs, but all ESMs overestimated the maximum vegetation carbon densities (Figs. 1, S3, and S4). The models MIROC-ESM and BCC_CSM1.1 (both had maximum values around 20 kg C m^{-2}) were the closest to the observations (16 kg C m^{-2}). The greatest overestimations were produced by CESM1(BGC) (47 kg C m^{-2}) and NorESM1-ME (49 kg C m^{-2}).

At grid scale, MPI-ESM-MR, MIROC-ESM, and INM-CM4.0 were the best-performing models with lower RMSE and higher R^2, which explained 38.0%, 36.9%, and 33.6% of the observed spatial variation, respectively (Table 2; Fig. 2); but slopes of MIROC-ESM and MPI-ESM-MR were much lower than 1, a sign of systematic bias. Agreement across ESMs was also poor, as indicated by the low R^2 of paired comparisons between ESMs (Figs. 2 and S5), with 48 out of 55 R^2 values being lower than 0.5. A very high R^2 was found between CESM1(BGC) and NorESM1-ME, which shared the same land carbon cycle model.

b. Variability of vegetation carbon at the biome scale

Most carbon-rich areas in the observations and ESMs were located in tropical and boreal regions (Fig. S2). It should be noted that the boreal biome in our classification included the temperate rain forest of the North American Pacific Northwest, which inflated the carbon content of this otherwise moderate biome. At the biome scale, vegetation carbon exhibited large variability between

TABLE 2. Coefficient of determination R^2 of linear regression and RMSE between vegetation carbon of the observations and historical simulations of 11 CMIP5 ESMs at gridcell (including slopes and intercepts) and biome scales and the absolute differences in vegetation carbon at global scale.

Models	Grid cell		Biome		Global
	R^2 (slope, intercept)	RMSE	R^2	RMSE	Absolute differences
BCC_CSM1.1	0.2665 (0.46, 1.22)	3.14	0.7725	24.96	75.35
BNU-ESM	0.1474 (0.85, 2.18)	7.14	0.6380	94.22	374.21
CanESM2	0.1867 (0.51, 1.60)	4.50	0.9396	12.51	18.71
CESM1(BGC)	0.2810 (0.89, 0.10)	4.96	0.5338	67.98	25.83
GFDL-ESM2G	0.1448 (0.48, 3.50)	5.16	0.8669	24.39	105.00
HadGEM2-ES	0.2952 (0.62, 0.72)	3.58	0.7608	31.50	99.39
INM-CM4.0	0.3363 (0.64, 1.55)	3.74	0.9234	15.59	52.68
IPSL-CM5A-MR	0.2927 (0.87, 1.94)	5.28	0.6712	37.61	87.83
MIROC-ESM	0.3685 (0.42, 0.57)	3.01	0.9825	31.83	203.89
MPI-ESM-MR	0.3796 (0.44, 0.32)	3.04	0.8494	32.07	212.12
NorESM1-ME	0.1545 (0.79, 1.11)	7.18	0.5283	66.68	3.53

ESMs and the observations and across ESMs (Fig. 3 and Fig. S6 in the supplementary material). Coefficients of variation across ESMs (calculated as standard deviation divided by mean) ranged from 33.0% in temperate forests to 115.7% in the tundra regions. The multimodel median of vegetation carbon was close to the reference data in temperate forests and permanent wetlands but higher in tropical rain forests and lower in the remaining biomes, especially in boreal forests.

Goodness of fit for ESM simulations at the biome scale was greatly improved compared to the grid scale, with more than double R^2 for most ESMs (Table 2; Table S2 and Fig. S7 in the supplementary material). Overall, goodness of fit indicated by R^2 at the biome scale was similar with those at the gridcell scale: MIROC-ESM, INM-CM4.0, and MPI-ESM-MR were still listed among the best-performing models. However, R^2 of CanESM2 was dramatically enhanced at the biome scale. CESM1 (BGC), HadGEM2-ES, and NorESM1-ME overestimated the vegetation carbon in tropical rain forests but underestimated carbon stocks in boreal forests (Fig. S7). MIROC-ESM and MPI-ESM-MR underestimated vegetation carbon for all biomes. While BNU-ESM greatly overestimated vegetation carbon for the majority of the biomes, it underestimated carbon stocks of boreal forests. GFDL-ESM2G performed well at other biomes, but it overestimated vegetation carbon in tropical rain forests, grassland and savanna, and tundra. CanESM2 and INM-CM4.0 simulated similar carbon stocks in relation to the observations in all biomes.

c. *Global sums of vegetation carbon*

Out of 11 ESMs, 8 reproduced global totals within ±20% of the observation-based estimate (560 ± 112 Pg C; Fig. 4 and Fig. S8 in the supplementary material). Global vegetation carbon totals simulated by ESMs varied between 340 (MPI-ESM-MR) and 930 Pg C (BNU-ESM) with a multimodel mean and standard deviation of 550 ± 160 Pg C. It should be noted that some ESMs showed nearly opposite performance at the global scale to those at the other spatial scales (Tables 2 and S2). MIROC-ESM and MPI-ESM-MR, which performed very well at the gridcell scale, had very large biases in the global sums (underestimated by 36.6% and 38.1%, respectively). In contrast, CESM1(BGC) and NorESM1-ME simulated very similar global sums as the observations (within ±5%), although their agreements with the observations at the other two scales were poor.

d. *Global NPP and residence time of vegetation carbon*

Global annual NPP simulated by ESMs was generally of the same order of magnitude as MODIS NPP (Fig. 5

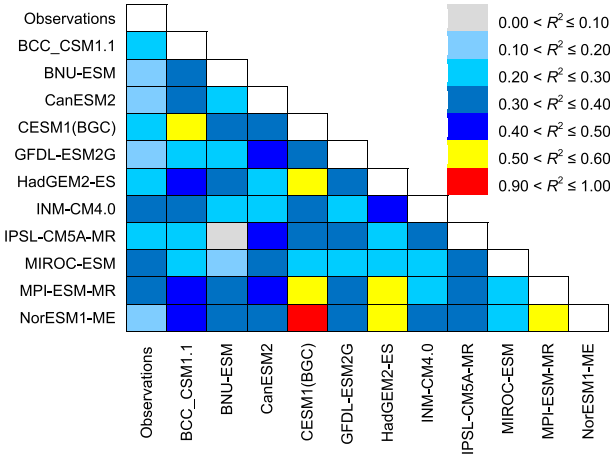

FIG. 2. A matrix of the coefficient of determination R^2 of linear regression of vegetation carbon density between each CMIP5 ESM and the observations and between CMIP5 ESMs at the grid cells.

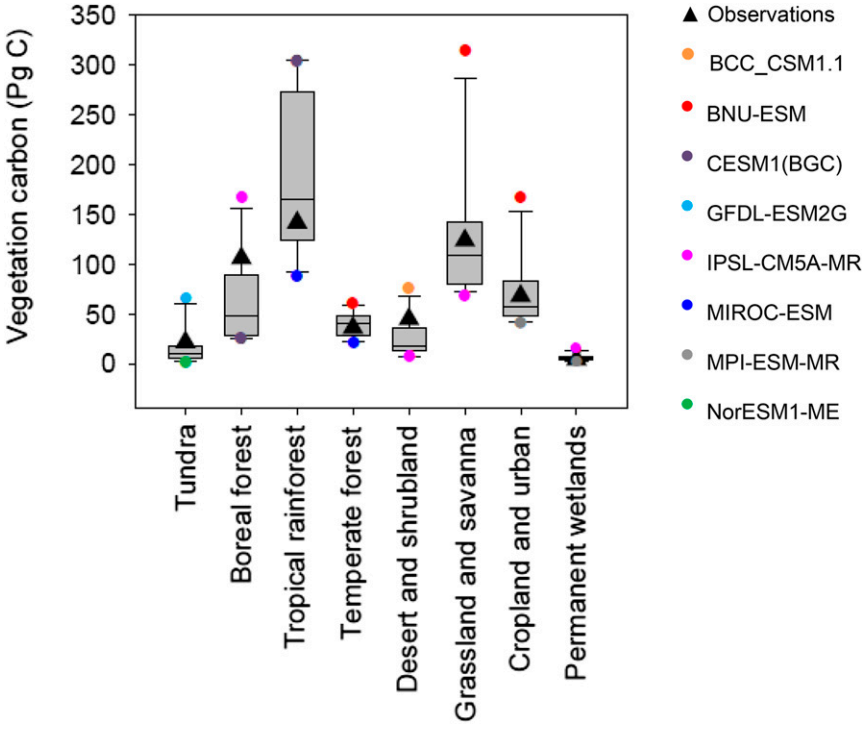

FIG. 3. Variability in vegetation carbon of different biomes of the observations and CMIP5 ESMs. Horizontal lines inside the boxes are medians of ESMs for each biome. The bottom and top edges of boxes stand for the first and third quartiles, respectively. The bars represent the extreme values (within 1.5 times the interquartile range from the upper or lower quartile). Color-filled circles represent those models for which estimates were greater or less than 1.5 times the interquartile range from the median. The black triangles represent the observational values for each biome.

and Fig. S9 in the supplementary material), but BCC_CSM1.1, GFDL-ESM2G, HadGEM2-ES, IPSL-CM5A-MR, and MPI-ESM-MR models simulated considerably higher NPP than the reference data. The residence time of vegetation carbon of ESMs ranged from 3.7 to 13.6 yr. Most models had similar or much shorter residence times than the observations, except for BNU-ESM. Variation of the global vegetation carbon storage between CMIP5 models exhibited a more pronounced relationship with the vegetation residence time than with NPP (Figs. 5 and S9). After excluding BNU-ESM, which might contribute too much to the high R^2, we observed a substantial reduction in the R^2, but it remained significant at the 0.1 level ($R^2 = 0.30$, $P < 0.10$).

4. Discussion

Hoffman et al. (2014) found that much of the intermodel variation in the projected CO_2 during the twenty-first century was tied to the biases that existed during the contemporary period. Because of this relationship, reliable historical simulations of land carbon storage by ESMs would increase our confidence in the projections of future land carbon storage and, hence, the future climate. Overall, vegetation carbon produced by CMIP5 ESM historical simulations had poor agreement with the observational data. Moreover, we observed little agreement in the simulations between ESMs. Below, we discuss the factors contributing to the inconsistencies in model predictions.

a. Divergent performances of ESMs at different scales

In our study, we found that the relative performance of ESMs in simulating vegetation carbon depends strongly on scale (Tables 2 and S2). Completely opposite conclusions about the performance of ESMs could be inferred when evaluating them at different scales. The performance of BCC_CSM1.1 and BNU-ESM were the most consistent across all scales. CESM1 (BGC), MIROC-ESM, MPI-ESM-MR, and NorESM1-ME exhibited opposite performance at the global scale compared to the other scales. MIROC-ESM and MPI-ESM-MR performed well at the grid scale and could well catch spatial distribution patterns of the observations at

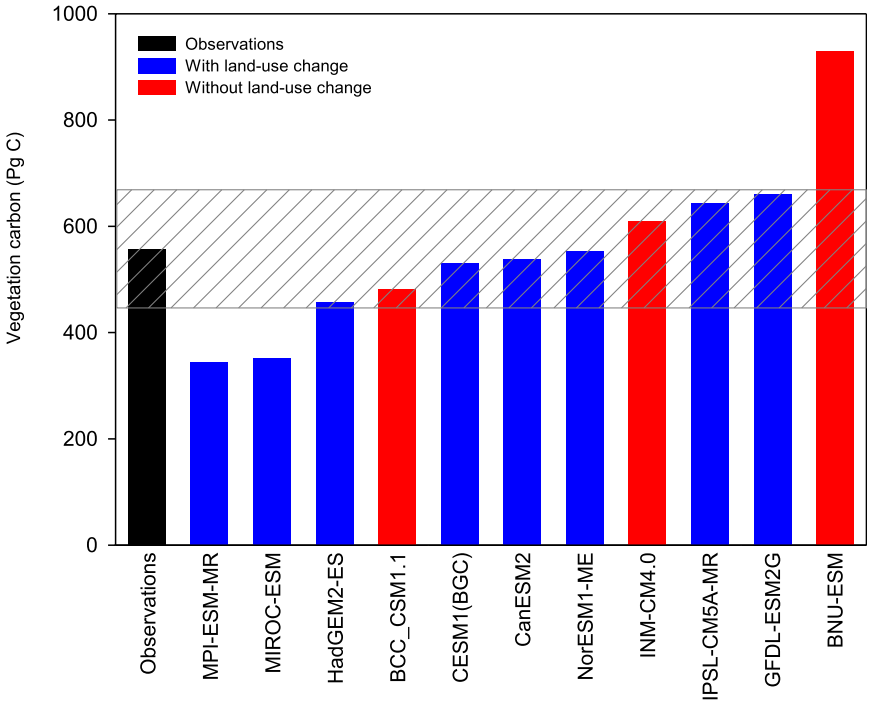

FIG. 4. Cumulative global vegetation carbon stocks of the observations and CMIP5 ESMs. The area with diagonal gray lines stands for ±20% of the observation value.

the biome scale (indicated by higher R^2), but they substantially underestimated the global sums. In contrast, the global vegetation carbon stocks simulated by CESM1(BGC) and NorESM1-ME were very close to the observed data, even though they had low R^2 and high RMSE at the other two scales. Therefore, it is critical to apply multiple metrics and criteria when evaluating the performance of multimodel ensembles. It should be noted that goodness of fit measured by R^2 at grid and biome scales emphasized the match of spatial distribution between ESMs and the observations. Thus, if the absolute differences are the major interest, the slopes and intercepts should be assessed as well. In contrast, goodness of fit at the global scale referred to the absolute difference in the global vegetation carbon between ESMs and the observations.

At the grid scale, it is hard to compare plant functional types (PFTs) and climate forcing between ESMs and observations, as well as between ESMs, because each modeling center adopts its own grid configuration and representation of PFTs, which can differ considerably from one another. Greater agreement could be reached at the biome scale because well-developed satellite-derived data products for land use and land cover become available, and climate forcing is better represented at a larger scale. However, relatively better performance of ESMs at the grid scale did not necessarily result in better performance at the biome or global scales. Similarly, good performance of an ESM at the biome scale might not guarantee a comparable global vegetation carbon stock to the observations. This could partly be attributed to the fact that we calculated the sums of grid carbon density over a specific region or the globe to get the respective total carbon storage for a specific biome or the globe. Aggregation at a coarser scale missed the variations at a finer scale. High R^2 and low RMSE implied that a model captured the observed spatial distribution well, although the slope and intercept of the regression might deviate away from 1 and 0, respectively, indicating a bias in the model.

There was a large divergence between model simulations and the observations and across the models at each scale, although 8 out of 11 ESMs could reproduce the global vegetation carbon within the 20% uncertainty range of the reference data (Fig. 4). We should direct more studies on the vegetation carbon in tropical and boreal forests (the latter included temperate rain forests), where higher vegetation carbon stocks were observed. Tropical regions may experience the earliest emergence of historically unprecedented climates (Mora et al. 2013) and, as a result, lose large amounts of carbon (Ahlström et al. 2012). Boreal regions have been predicted to undergo considerable changes in NPP over the next century, which, in turn, may change future vegetation carbon stocks (Todd-Brown et al. 2014).

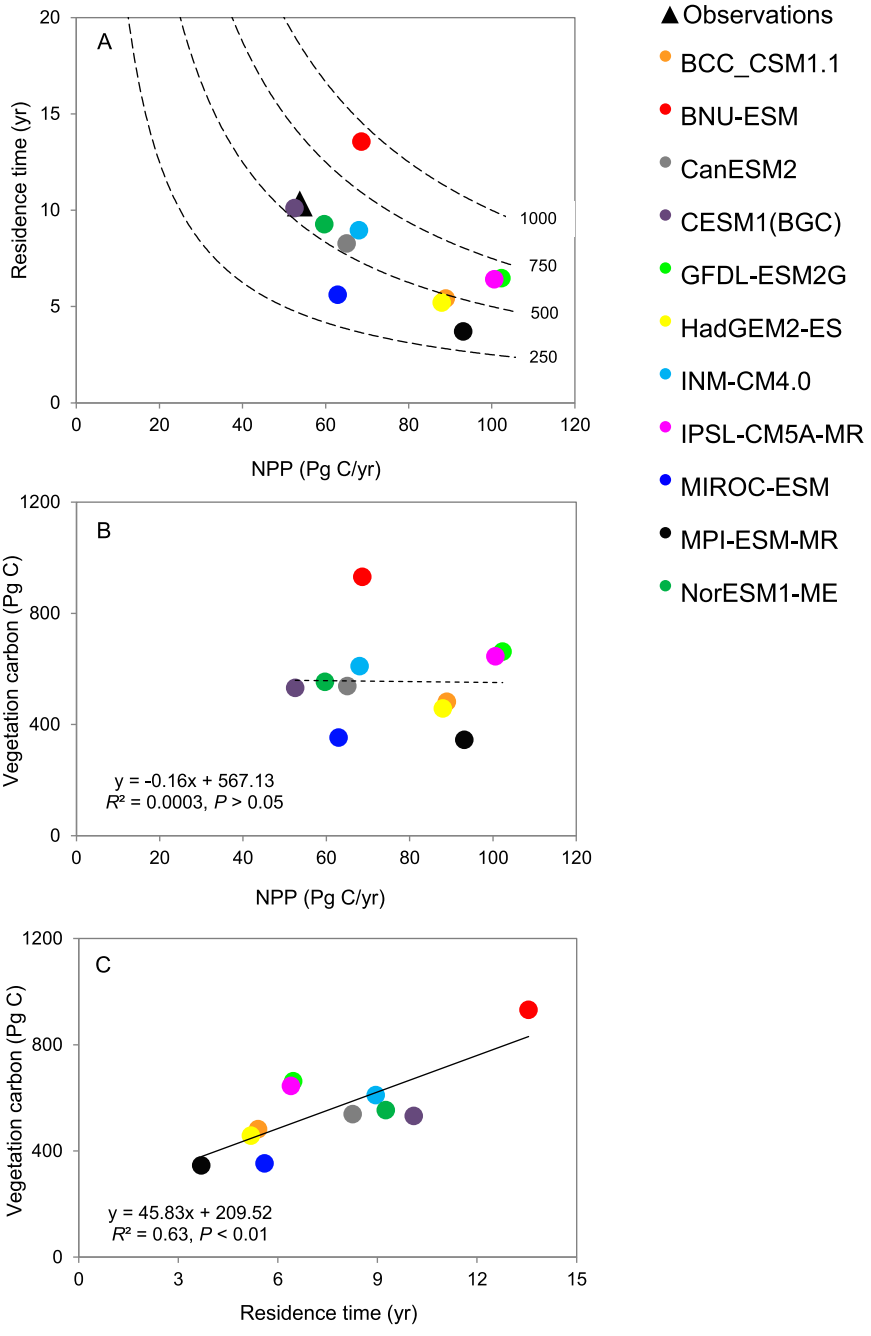

FIG. 5. (a) Relationship between global NPP and residence time of vegetation carbon from both the observations and CMIP5 ESMs; (b) relationship between vegetation carbon and NPP; and (c) relationship between vegetation carbon and residence time of ESMs. The dashed curves in (a) represent constant values of global vegetation carbon.

b. Comparison with simulations of soil organic carbon

Since the soil carbon pool and vegetation carbon pool together determine the bulk fraction of the land carbon storage, it is useful to compare the performance of CMIP5 ESMs in simulating both vegetation and soil carbon (Todd-Brown et al. 2013). The performance of ESMs in simulating historical vegetation carbon was quite similar to their performance in simulating soil organic carbon: that is, there was a poor ability to reproduce the spatial distribution

of the observed data at grid cells, but this was improved greatly at the biome scale. The R^2 of the regression for vegetation carbon simulations ranged from 0.14 to 0.38 at the grid scale and from 0.53 to 0.98 at the biome scale. The R^2 for soil carbon simulations was between 0.004 and 0.15 at the grid scale (calculated as the square of Pearson correlation coefficients in Todd-Brown et al. 2013) but was enhanced to 0.38–0.95 at the biome scale. The grid-scale simulations of vegetation carbon were better than simulations of soil carbon overall in terms of R^2. Interestingly, large variability of simulated vegetation carbon was found in tropical and boreal forests with higher vegetation carbon stocks, similar to the large bias in soil carbon simulated by ESMs at high northern latitudes, where greater soil carbon stocks existed (Todd-Brown et al. 2013) and were subject to complicated thawing-relevant physics and soil hydrology in permafrost (Koven et al. 2013).

At the global scale, most ESMs also performed better in simulating vegetation carbon than soil organic carbon, according to the percentage differences between modeled values and the observations (Figs. 4 and S8; Anav et al. 2013; Todd-Brown et al. 2013), with the exception of GFDL-ESM2G, HadGEM2, and IPSL-CM5. Better performance of ESMs in simulating vegetation carbon than in simulating soil carbon could be attributed to better understanding of aboveground than belowground processes, as well as better availability of aboveground data products to tune the models for accurate simulations of the respective key processes. While Wieder et al. (2013) improved global soil carbon projections with a new model that explicitly represented microbial mechanisms of soil carbon cycling, such improvement might have resulted from an unrealistic process representation, since the microbial models of soil decomposition exhibited some properties that have not been observed (Wang et al. 2014).

It was interesting that MPI-ESM-LR and MIROC-ESM (average of MIROC-ESM and MIROC-ESM-CHEM) greatly overestimated the global soil organic carbon by 143.3% and 103.6%, respectively (Todd-Brown et al. 2013), but they underestimated the vegetation carbon by 39.9% for MPI-ESM-LR and 37.0% for the average of MIROC-ESM and MIROC-ESM-CHEM (Fig. S8). Another noticeable phenomenon was that CCSM4 and NorESM1 (average of NorESM1-M and NorESM1-ME), both having CLM4 for a land carbon cycle component, simulated extremely low global soil carbon (Todd-Brown et al. 2013). However, CESM1(BGC) (the analogs of CCSM4) and NorESM1 simulated reasonable global vegetation carbon. On the other hand, GFDL-ESM2G and HadGEM2 simulated global soil carbon that was close to the observations, but they got much higher (GFDL-ESM2G) or lower (HadGEM2) global vegetation carbon than the observations. When adding soil carbon (data from Todd-Brown et al. 2013) and vegetation carbon to get overall land carbon storage, MPI-ESM-LR and MIROC-ESM (average of MIROC-ESM and MIROC-ESM-CHEM) overestimated the total land carbon storage by 83.5% and 60.2%, respectively, compared with the observations. CESM1(BGC) [using soil carbon of CCSM4 to get total land carbon for CESM1(BGC)] and NorESM1 underestimated the total land carbon storage by 42.2% and 39.3%, respectively, indicating that bias in the land carbon storage was dominated by bias in soil carbon. We further explored the correlation between simulated vegetation carbon and soil carbon of 10 ESMs: BCC_CSM1.1, CanESM2, CESM1(BGC) (using soil carbon of CCSM4), GFDL-ESM2G, HadGEM2 (average of HadGEM2-CC and HadGEM2-ES), INM-CM4.0, IPSL-CM5 (average of IPSL-CM5A-LR, IPSL-CM5A-MR, and IPSL-CM5B-LR), MIROC-ESM (average of MIROC-ESM and MIROC-ESM-CHEM), MPI-ESM-LR, and NorESM1-M (average of NorESM1-M and NorESM1-M E). We found that the simulated soil carbon was negatively correlated with simulated vegetation carbon ($R^2 = 0.53$, $P < 0.05$). This might partly make the total land carbon stocks more comparable among the models, even though the total land carbon stocks were more dominated by soil carbon. Causes for the differences in soil or vegetation carbon across ESMs could vary. NPP and soil temperature can explain the differences in soil carbon between models (Todd-Brown et al. 2013), but, in this study, no significant correlation was found between vegetation carbon and NPP.

c. *Factors determining modeled vegetation carbon*

1) CARBON CYCLE STRUCTURE AND PARAMETERIZATION

Structures of the land carbon cycle in the ESMs adopted in this study were similar (most with 3 or 4 biomass pools; Tables 1 and S1). Thus, it appeared that the uncertainties in the simulated vegetation carbon were not affected by the structures of the land carbon cycle models. However, at all scales, two models [CESM1(BGC) and NorESM1-ME] that used the same land submodel (CLM4) always exhibited very similar behaviors in simulating vegetation carbon, as shown both in this study and in Anav et al. (2013). The same phenomenon was observed for their performance in simulating soil carbon (Todd-Brown et al. 2013). This may indicate that parameterization related to model structures for the land carbon cycle plays an important role in simulating land carbon storage, but this statement needs to be confirmed by a separate study, as CMIP5

protocol does not require models to provide parameter values. A useful method to test this would be to perform sensitivity experiments by changing various factors keeping the model structure and the rest of the model parameterization fixed, as Ahlström et al. (2013) have done to determine the roles of climate and atmospheric forcing.

2) NPP AND RESIDENCE TIME

Ecosystem carbon storage is codetermined by NPP and residence time (Luo et al. 2003; Luo and Weng 2011; Xia et al. 2013). For example, despite the increased carbon input to soils under rising atmospheric CO_2, reduced soil carbon residence time resulted in the insignificant net effect of increased atmospheric CO_2 on the equilibrium soil carbon storage (van Groenigen et al. 2014). In this study, however, the variability of modeled vegetation carbon across the ESMs was predominantly explained by the residence time of vegetation carbon and not by NPP (Figs. 5 and S9). Such a phenomenon was also found for uncertainty in the projected terrestrial vegetation (Friend et al. 2013). In this study, NPP of BNU-ESM was close to the observations, but it had much longer vegetation carbon residence times, resulting in a substantially higher vegetation carbon pool size than observed (Figs. 4, 5). The shorter vegetation carbon residence time of MIROC-ESM and MPI-ESM-MR made their total vegetation carbon pools very low, although MPI-ESM-MR had relatively high NPP. Disagreements still existed in simulations of NPP between CMIP5 ESMs (Fig. 5 and Todd-Brown et al. 2013) and were even worse in simulations of vegetation or ecosystem residence time (Fig. 5 and Yan et al. 2014); therefore, improvement of simulations of vegetation carbon residence time would be crucial for reducing the bias in the projections of vegetation carbon storage (Luo and Weng 2011; Friend et al. 2013; Todd-Brown et al. 2013; Yan et al. 2014). Improved representation of PFTs, allocation coefficients, and longevity of different biomass pools in models could improve simulations of vegetation carbon residence time [see also section 4c(3)].

3) PLANT FUNCTIONAL TYPES

As discussed above, variability of the simulated vegetation carbon storage by ESMs could be explained by the residence time, a trait related to PFTs (i.e., longevity of individual pools). Within a specific PFT, different biomass pools have different residence times, so allocation of NPP to different pools will also affect residence times of vegetation carbon substantially. Therefore, improved representation of regional PFTs and their respective prescribed allocation coefficients and longevity of different biomass pools are critical components for improving ESMs' performance. The number of PFTs represented in these ESMs varied from 5 to 15 (Table 1), and the combinations within grid cells differed considerably between ESMs (please refer to the references listed in Table 1 for each ESM). Therefore, improvements in the representation of PFTs toward better agreement between models might be a first step. The use of the same land-cover dataset between models is also important to elucidate other sources of models' uncertainty (e.g., model structures). Agreement is likely to be achieved with remote sensing land-cover products. Syntheses of published literature on allocation coefficients and the longevity of each biomass pool would be a useful way to improve their realistic representation in the ESMs. Data assimilation is a potential method to help optimize the parameterization associated with the allocation coefficients and longevity of biomass pools, given that we can generate enough observational datasets.

From the above discussion, we recommend that ESM developers use consistent and well-developed land-cover products together with compiled datasets on allocation coefficients and the longevity of different biomass pools to improve the accuracy of vegetation carbon projections at the regional scale first (e.g., the biomes). This is feasible because there are more available datasets at regional scales, especially for those critical biomes, such as the tropical rain forests and boreal forests. Reich et al. (2014) used temperature-dependent needle longevity and nitrogen concentration, as well as biomass allocation in a land surface model instead of constant values. They found the realistic parameterization of these variables improved predictions of carbon cycling processes, such as leaf area index and gross primary production in boreal forests, compared with observations from flux sites. Next, model developers can tune the ESMs to fit the global value. At a finer scale (i.e., at grid cells) it is currently not realistic to reach a very ideal goodness of fit because of the lack of global databases of biomass in which thousands of site-level measurements are compiled into grid values and can be used as benchmarks.

4) OTHER FACTORS

Because of the tremendous challenges in conducting investigations of vegetation carbon at the global scale, the benchmarks for evaluating performance of ESMs in simulating global vegetation carbon storage are highly limited, although biomass has been studied for quite a long time. As the only available contemporary dataset at the global scale, it is likely that the Gibbs (2006) dataset had uncertainties related to measurements or estimates, although the global vegetation carbon storage and the

spatial distribution were reasonable based on our knowledge of biomass carbon. However, this dataset was compiled using multiple approaches, and it was problematic to quantify the uncertainties. The lack of uncertainty analysis of the observations might impose some limitations but will not change our major conclusions. Moreover, our results from model intercomparisons can provide insights into the uncertainties in the simulated vegetation carbon and possible causes. While further efforts are needed to develop more robust global biomass databases with well-demonstrated uncertainties, the use of regional datasets is an alternative option to move forward, in particular those from important regions in terms of vegetation carbon storage, such as data products for the tropical regions (e.g., Saatchi et al. 2011).

For selected benchmarks, the metrics adopted to evaluate model performance could affect the outcome differently (Luo et al. 2012). Using the same reference data, we obtained slightly different performance in vegetation carbon simulations by CMIP5 ESMs from those found in Anav et al. (2013). Both studies have found that the goodness of fit changed considerably at different scales, but the goodness of fit at the global scale showed some difference between the two studies. Given the small difference in time of output of the ESMs selected, the global sums were very close between the two studies, and, therefore, the selected metrics are another factor that could bring a different understanding of model behaviors. They used the normalized mean bias between the models and the reference data to compute skill scores of ESMs at four scales: global, Southern Hemisphere, Northern Hemisphere, and tropics.

One possible contribution to the uncertainties across the models is climate forcing, such as temperature and precipitation (Piao et al. 2009; Xia et al. 2013). For example, simulations by dynamic global vegetation models often resulted in substantially varying results for carbon balance depending on the choice of forcing from the general circulation models (GCMs; Ahlström et al. 2012, 2013). GCMs explained the majority of uncertainty in the projected twenty-first-century terrestrial carbon balance (Ahlström et al. 2013). Climate forcing alters the carbon budget through influencing the simulated NPP and residence time, highlighting again the urgent demand to improve model representations of NPP and residence time.

5. Summary

This study evaluated the performance of 11 ESMs that were involved in CMIP5 in simulating terrestrial vegetation carbon at grid, biome, and global scales through model–model and model–data comparisons. Large disagreements were found between modeled vegetation carbon and the observations. The simulated maximum carbon density at grid cells varied by as much as a factor of 3 relative to the observed values. Performances of ESMs at the biome scale were better than at the grid scale. Even so, the absolute amounts of vegetation carbon in different biomes, particularly in tropical and boreal regions, varied greatly between ESMs. The global vegetation carbon stocks differed nearly threefold among ESMs.

The goodness of fit changed depending on the chosen spatial scales. For example, the performance of MIROC-ESM was better at the gridcell and biome scales than most other models in terms of R^2 and RMSE, but its global sum was very low compared to the observations. In contrast, global sums of CESM1(BGC) and NorESM1-ME were very close to the observed data, but they had low R^2 and high RMSE at finer scales. It should be noted that R^2 and RMSE at the grid and biome scales emphasized the goodness of fit of spatial distribution between ESMs and the observations, whereas goodness of fit at the global scale referred to the absolute differences between simulations and the observed data.

We found that the vegetation carbon residence time (i.e., plant longevity) explained the majority of the variability in vegetation carbon across ESMs. This means that the vegetation carbon storage depends mostly on how long the carbon will remain in the vegetation. These findings indicate that parameterization of residence time and its spatial distributions in ESMs may be a key factor in controlling the vegetation carbon simulations. Improvement of other drivers and processes that have effects on the residence times, including harvesting and natural disturbances, are also important for ESMs to get more accurate predictions of biomass carbon stocks.

Acknowledgments. The authors greatly appreciate Dr. Alessandro Anav for kindly sharing his vegetation carbon data with us, checking our data, and answering our questions on data processing; Dr. Katherine E. Todd-Brown for her constructive suggestions; and Dr. Gerhard Krinner for his great comments. We thank Drs. Ingo Bethke, Jean-Louis Dufresne, and Xueli Shi for answering our questions about their models. We thank Drs. Rashid Rafique, Xiaoming Xu, and Xia Xu for their help with figures. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the modeling groups (listed in Table 1) for producing and making their model outputs available. 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. We also thank Oak Ridge National Laboratory for open utilization of their datasets. The NCAR Command Language team gave us great support for using NCL. This research was financially supported by the U.S. Department of Energy, Terrestrial Ecosystem Sciences Grant DE SC0008270 and U.S. National Science Foundation (NSF) Grants DBI 0850290, EPS 0919466, DEB 0743778, DEB 0840964, and EF 1137293. JT acknowledges funding from the Center for Climate Dynamics and Research Council of Norway through Project EVA: ''Earth System Modelling of Climate Variations in the Anthropocene'' (Grant 229771).

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,** 044008, doi:10.1088/1748-9326/7/4/044008.

——, B. Smith, J. Lindström, M. Rummukainen, and C. B. Uvo, 2013: GCM characteristics explain the majority of uncertainty in projected 21st century terrestrial ecosystem carbon balance. *Biogeosciences,* **10,** 1517–1528, doi:10.5194/bg-10-1517-2013.

Anav, A., and Coauthors, 2013: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models. *J. Climate,* **26,** 6801–6803, doi:10.1175/JCLI-D-12-00417.1.

Arora, V., and G. J. Boer, 2010: Uncertainties in the 20th century carbon budget associated with land use change. *Global Change Biol.,* **16,** 3327–3348, doi:10.1111/j.1365-2486.2010.02202.x.

——, and Coauthors, 2013: Carbon–concentration and carbon–climate feedbacks in CMIP5 earth system models. *J. Climate,* **26,** 5289–5314, doi:10.1175/JCLI-D-12-00494.1.

Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, 2011: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geosci. Model Dev.,* **4,** 255–269, doi:10.5194/gmd-4-255-2011.

Bonan, G. B., 1996: A Land Surface Model (LSM version 1.0) for ecological, hydrological and atmospheric studies: Technical description and user's guide. NCAR Tech. Note NCAR/TN-417+STR, 156 pp, doi:10.5065/D6DF6P5X.

Brovkin, V., T. Raddatz, C. H. Reick, M. Claussen, and V. Gayler, 2009: Global biogeophysical interactions between forest and climate. *Geophys. Res. Lett.,* **36,** L07405, doi:10.1029/2009GL037543.

——, and Coauthors, 2013: Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century. *J. Climate,* **26,** 6859–6881, doi:10.1175/JCLI-D-12-00623.1.

Canadell, J. G., and Coauthors, 2007: Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl. Acad. Sci. USA,* **104,** 18 866–18 870, doi:10.1073/pnas.0702737104.

Collins, W. J., and Coauthors, 2011: Development and evaluation of an Earth-System model—HadGEM2. *Geosci. Model Dev.,* **4,** 1051–1075, doi:10.5194/gmd-4-1051-2011.

Cox, P. M., 2001: Description of the "TRIFFID" dynamic global vegetation model. Hadley Centre Tech. Note 24, 17 pp. [Available online at http://www.metoffice.gov.uk/media/pdf/9/h/HCTN_24.pdf.]

——, R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell, 2000: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature,* **408,** 184–187, doi:10.1038/35041539.

Dai, Y. J., and Coauthors, 2003: The Common Land Model. *Bull. Amer. Meteor. Soc.,* **84,** 1013–1023, doi:10.1175/BAMS-84-8-1013.

——, R. E. Dickinson, and Y. P. Wang, 2004: A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. *J. Climate,* **17,** 2281–2299, doi:10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2.

Dufresne, J.-L., L. Fairhead, H. Le Treut, M. Berthelot, L. Bopp, P. Ciais, P. Friedlingstein, and P. Monfray, 2002: On the magnitude of positive feedback between future climate change and the carbon cycle. *Geophys. Res. Lett.,* **29,** 43-1–43-4, doi:10.1029/2001GL013777.

——, and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dyn.,* **40,** 2123–2165, doi:10.1007/s00382-012-1636-1.

Dunne, J. P., and Coauthors, 2013: GFDL's ESM2 global coupled climate–carbon earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Climate,* **26,** 2247–2267, doi:10.1175/JCLI-D-12-00150.1.

European Commission, Joint Research Centre, 2003: Global Land Cover 2000 database. European Commission, Joint Research Centre, accessed 26 December 2012. [Available online at http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php.]

Friedlingstein, P., J.-L. Dufresne, P. M. Cox, and P. Rayner, 2003: How positive is the feedback between climate change and the carbon cycle? *Tellus,* **55B,** 692–700, doi:10.1034/j.1600-0889.2003.01461.x.

——, and Coauthors, 2006: Climate–carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. *J. Climate,* **19,** 3337–3353, doi:10.1175/JCLI3800.1.

——, M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti, 2014: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Climate,* **27,** 511–526, doi:10.1175/JCLI-D-12-00579.1.

Friend, A. D., and Coauthors, 2013: Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO_2. *Proc. Natl. Acad. Sci. USA,* 3280–3285, doi:10.1073/pnas.1222477110.

Gibbs, H. K., 2006: Olson's major world ecosystem complexes ranked by carbon in live vegetation: An updated database using the GLC2000 land cover product (NDP-017b). Carbon Dioxide Information Center, Oak Ridge National Laboratory, accessed 26 December 2012, doi:10.3334/CDIAC/lue.ndp017.

Heinsch, F. A., and Coauthors, 2003: User's guide GPP and NPP (MOD17A2/A3) products: NASA MODIS land algorithm. MODIS Land Team Tech. Rep., 57 pp. [Available online at http://www.ntsg.umt.edu/sites/ntsg.umt.edu/files/modis/MOD17UsersGuide.pdf.]

Hoffman, F. M., and Coauthors, 2014: Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models. *J. Geophys. Res. Biogeosci.,* **119,** 141–162, doi:10.1002/2013JG002381.

Houghton, R. A., 2007: Balancing the global carbon budget. *Annu. Rev. Earth Planet. Sci.,* **35,** 313–347, doi:10.1146/annurev.earth.35.031306.140057.

IGBP Terrestrial Carbon Working Group, 1998: The terrestrial carbon cycle: Implications for the Kyoto Protocol. *Science,* **280,** 1393–1394, doi:10.1126/science.280.5368.1393.

IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis,* T. F. Stocker et al., Eds., Cambridge University Press, 1–29. [Available online at http://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_SPM_FINAL.pdf.]

Janssen, P. H. M., and P. S. C. Heuberger, 1995: Calibration of process-oriented models. *Ecol. Modell.,* **83,** 55–66, doi:10.1016/0304-3800(95)00084-9.

Ji, D., and Coauthors, 2014: Description and basic evaluation of BNU-ESM version 1. *Geosci. Model Dev. Discuss.,* **7,** 1601–1647, doi:10.5194/gmdd-7-1601-2014.

Ji, J., M. Huang, and K. Li, 2008: Prediction of carbon exchanges between China terrestrial ecosystem and atmosphere in 21st century. *Sci. China,* **51D,** 885–898, doi:10.1007/s11430-008-0039-y.

Jones, C. D., and Coauthors, 2011: The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.,* **4,** 543–570, doi:10.5194/gmd-4-543-2011.

——, and Coauthors, 2013: Twenty-first-century compatible CO_2 emissions and airborne fraction simulated by CMIP5 earth system models under four representative concentration pathways. *J. Climate,* **26,** 4398–4413, doi:10.1175/JCLI-D-12-00554.1.

Koven, C. D., W. J. Riley, and A. Stern, 2013: Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 earth system models. *J. Climate,* **26,** 1877–1900, doi:10.1175/JCLI-D-12-00228.1.

Krinner, G., and Coauthors, 2005: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem. Cycles,* **19,** GB1015, doi:10.1029/2003GB002199.

Lawrence, D. M., and Coauthors, 2011: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *J. Adv. Modell. Earth Syst.,* **3,** M03001, doi:10.1029/2011MS000045.

Le Quéré, C., and Coauthors, 2009: Trends in the sources and sinks of carbon dioxide. *Nat. Geosci.,* **2,** 831–836, doi:10.1038/ngeo689.

——, and Coauthors, 2014: Global carbon budget 2013. *Earth Syst. Sci. Data,* **6,** 235–263, doi:10.5194/essd-6-235-2014.

Luo, Y., and E. Weng, 2011: Dynamic disequilibrium of terrestrial carbon cycle under global change. *Trends Ecol. Evol.,* **26,** 96–104, doi:10.1016/j.tree.2010.11.003.

——, L. White, J. Canadell, E. DeLucia, D. Ellsworth, A. Finzi, J. Lichter, and W. Shclesinger, 2003: Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach. *Global Biogeochem. Cycles,* **17,** 1021, doi:10.1029/2002GB001923.

——, and Coauthors, 2012: A framework for benchmarking land models. *Biogeosciences,* **9,** 3857–3874, doi:10.5194/bg-9-3857-2012.

Martin, G. M., and Coauthors, 2011: The HadGEM2 family of Met Office Unified Model climate configurations. *Geosci. Model Dev.,* **4,** 723–757, doi:10.5194/gmd-4-723-2011.

Mora, C., and Coauthors, 2013: The projected timing of climate departure from recent variability. *Nature,* **502,** 183–187, doi:10.1038/nature12540.

NASA LP DAAC, 2008: Land Cover Type Yearly L3 Global 0.05Deg CMG (MCD12C1). USGS/Earth Resources Observation and Science (EROS) Center, accessed 26 December 2012. [Available online at https://lpdaac.usgs.gov/products/modis_products_table/mcd12c1.]

Olson, J. S., J. A. Watts, and L. J. Allison, 1985: Major world ecosystem complexes ranked by carbon in live vegetation: A database (NDP-017). Carbon Dioxide Information Center, Oak Ridge National Laboratory, accessed 26 December 2012, doi:10.3334/CDIAC/lue.ndp017.

Piao, S., P. Ciais, P. Friedlingstein, N. Noblet-Ducoudré, P. Cadule, N. Viovy, and T. Wang, 2009: Spatio-temporal patterns of terrestrial carbon cycle during the 20th century. *Global Biogeochem. Cycles,* **23,** GB4026, doi:10.1029/2008GB003339.

Raddatz, T. J., C. H. Reick, W. Knorr, J. Kattge, E. Roeckner, R. Schnur, K.-G. Schnitzler, P. Wetzel, and J. Jungclaus, 2007: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate Dyn.,* **29,** 565–574, doi:10.1007/s00382-007-0247-8.

Reich, P. B., R. Rich, X. Lu, Y.-P. Wang, and J. Oleksyn, 2014: Biogeographic variation in evergreen conifer needle longevity and impacts on boreal forest carbon cycle projections. *Proc. Natl. Acad. Sci. USA*, **111,** 13 703–13 708, doi:10.1073/pnas.1216054110.

Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler, 2013: Representation of natural and anthropogenic land cover change in MPI-ESM. *J. Adv. Modell. Earth Syst.,* **5,** 459–482, doi:10.1002/jame.20022.

Saatchi, S., and Coauthors, 2011: Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. USA,* **108,** 9899–9904, doi:10.1073/pnas.1019576108.

Sato, H., A. Itoh, and T. Kohyama, 2007: SEIB–DGVM: A new dynamic global vegetation model using a spatially explicit individual-based approach. *Ecol. Modell.,* **200,** 279–307, doi:10.1016/j.ecolmodel.2006.09.006.

Saugier, B., J. Roy, and H. A. Mooney, 2001: Estimations of global terrestrial productivity: Converging toward a single number? *Terrestrial Global Productivity,* J. Roy, B. Saugier, and H. A. Mooney, Eds., Academic Press, 543–557.

Shevliakova, E., and Coauthors, 2009: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. *Global Biogeochem. Cycles,* **23,** GB2022, doi:10.1029/2007GB003176.

Taylor, K. E., V. Balaji, S. Hankin, M. Juckes, and B. Lawrence, 2010: CMIP5 and AR5 Data Reference Syntax (DRS), version 0.25. WCRP Tech. Rep., 13 pp. [Available online at http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax_v0-25_clean.pdf.]

——, R. J. Stouffer, and G. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thornton, P. E., and N. E. Zimmermann, 2007: An improved canopy integration scheme for a land surface model with prognostic canopy structure. *J. Climate,* **20,** 3902–3923, doi:10.1175/JCLI4222.1.

——, J.-F. Lamarque, N. A. Rosenbloom, and N. M. Mahowald, 2007: Influence of carbon–nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability. *Global Biogeochem. Cycles,* **21,** GB4018, doi:10.1029/2006GB002868.

——, and Coauthors, 2009: Carbon–nitrogen interactions regulate climate–carbon cycle feedbacks: Results from an atmosphere–ocean general circulation model. *Biogeosciences,* **6,** 2099–2120, doi:10.5194/bg-6-2099-2009.

Tjiputra, J. F., C. Roelandt, M. Bentsen, D. M. Lawrence, T. Lorentzen, J. Schwinger, Ø. Seland, and C. Heinze, 2013: Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geosci. Model Dev.,* **6,** 301–325, doi:10.5194/gmd-6-301-2013.

Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison, 2013: Causes of variation in soil carbon predictions from

CMIP5 Earth system models and comparison with observations. *Biogeosciences,* **10,** 1717–1736, doi:10.5194/bg-10-1717-2013.

——, and Coauthors, 2014: Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences,* **11,** 2341–2356, doi:10.5194/bg-11-2341-2014.

UCAR/NCAR/CISL/VETS, 2013: The NCAR Command Language Version 6.1.2. NCAR, doi:10.5065/D6WD3XH5.

van Groenigen, K. J., X. Qi, C. Osenberg, Y. Luo, and B. Hungate, 2014: Faster decomposition under increased atmospheric CO_2 limits soil carbon storage. *Science,* **344,** 508–509, doi:10.1126/science.1249534.

Volodin, E. M., 2007: Atmosphere–ocean general circulation model with the carbon cycle. *Izv. Atmos. Oceanic Phys.,* **43,** 266–280, doi:10.1134/S0001433807030024.

Wang, Y., and Coauthors, 2014: Oscillatory behavior of two nonlinear microbial models of soil carbon decomposition. *Biogeosciences,* **11,** 1817–1831, doi:10.5194/bg-11-1817-2014.

Watanabe, S., and Coauthors, 2011: MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.,* **4,** 845–872, doi:10.5194/gmd-4-845-2011.

Watson, R. T., I. R. Noble, B. Bolin, N. H. Ravindranath, D. J. Verardo, and D. J. Dokken, Eds., 2000: *Land Use, Land-Use Change and Forestry*. Cambridge University Press, 30 pp. [Available online at http://www.ipcc.ch/ipccreports/sres/land_use/index.php?idp=0.]

Wieder, W. R., G. B. Bonan, and S. D. Allison, 2013: Global soil carbon projections are improved by modelling microbial processes. *Nat. Climate Change,* **3,** 909–912, doi:10.1038/nclimate1951.

Wu, T., and Coauthors, 2013: Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century. *J. Geophys. Res. Atmos.,* **118,** 4326–4347, doi:10.1002/jgrd.50320.

Xia, J., Y. Luo, Y. Wang, and O. Hararuk, 2013: Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Global Change Biol.,* **19,** 2104–2116, doi:10.1111/gcb.12172.

Yan, Y., Y. Luo, X. Zhou, and J. Chen, 2014: Sources of variation in simulated ecosystem carbon storage capacity from the 5th Climate Model Intercomparison Project (CMIP5). *Tellus,* **66B,** 22568, doi:10.3402/tellusb.v66.22568.

Zhao, M., F. Heinsch, R. Nemani, and S. Running, 2005: Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.,* **95,** 164–176, doi:10.1016/j.rse.2004.12.011.