From land use to land cover: restoring the afforestation signal in a coupled integrated assessment – earth system model and the implications for CMIP5 RCP simulations

A. V. Di Vittorio¹, L. P. Chini², B. Bond-Lamberty³, J. Mao⁴, X. Shi⁴, J. Truesdale⁵, A. Craig⁵, K. Calvin³, A. Jones¹, W. D. Collins¹, J. Edmonds³, G. C. Hurtt², P. Thornton⁴, and A. Thomson³

¹Lawrence Berkeley National Laboratory, Berkeley, CA, USA
²University of Maryland, College Park, MD, USA
³Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA
⁴Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA
⁵Independent contractor with Lawrence Berkeley National Laboratory, Berkeley, CA, USA
Received: 4 April 2014 – Accepted: 2 May 2014 – Published: 19 May 2014
Correspondence to: A. V. Di Vittorio (avdivittorio@lbl.gov)
Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Climate projections depend on scenarios of fossil fuel emissions and land use change, and the IPCC AR5 parallel process assumes consistent climate scenarios across Integrated Assessment and Earth System Models (IAMs and ESMs). The CMIP5 project used a novel “land use harmonization” based on the Global Land use Model (GLM) to provide ESMs with consistent 1500–2100 land use trajectories generated by historical data and four IAM projections. A direct coupling of the Global Change Assessment Model (GCAM), GLM, and the Community ESM (CESM) has allowed us to characterize and partially address a major gap in the CMIP5 land coupling design: the lack of a corresponding land cover harmonization. The CMIP5 CESM global afforestation is only 22% of GCAM’s 2005 to 2100 RCP4.5 afforestation. Likewise, only 17% of GCAM’s 2040 RCP4.5 afforestation, and zero pasture loss, were transmitted to CESM within the directly coupled model. This is a problem because afforestation was relied upon to achieve RCP4.5 climate stabilization. GLM modifications within the directly coupled model did not increase CESM afforestation. Modifying the land use translator in addition to GLM, however, enabled CESM to simulate 66% of GCAM’s afforestation in 2040, and 94% of GCAM’s pasture loss as grassland and shrubland losses. This additional afforestation increases vegetation carbon gain by 19 PgC and decreases atmospheric CO$_2$ gain by 8 ppmv from 2005 to 2040, implying different RCP4.5 climate scenarios between CMIP5 GCAM and CESM. Although the IAMs and ESMs were not expected to have exactly the same climate forcing, due in part to different terrestrial carbon cycles and atmospheric radiation algorithms, the ESMs were expected to project climates representative of the RCP scenarios. Similar land cover inconsistencies exist in other CMIP5 model results, primarily because land cover information is not shared between models. High RCP4.5 afforestation might also contribute to inconsistencies as some ESMs might impose bioclimatic limits to potential forest area and have different rates of forest growth than projected by RCP4.5. Further work to harmonize land cover among models will be required to address this problem.
1 Introduction

Land use plays a major role in determining terrestrial–atmosphere mass and energy exchange (Adegoke et al., 2007; Raddatz, 2007), which in turn influences local to global climate (Brovkin et al., 2013; Jones et al., 2013a; Pitman et al., 2009). Despite much recent progress, we still have a limited understanding of how historical land use has affected, and continues to affect, climate (Brovkin et al., 2013; Jones et al., 2013a; Pitman et al., 2009) and carbon (Anav et al., 2013; Arora and Boer, 2010; Houghton, 2010; Houghton et al., 2012; Hurtt et al., 2006; Jain et al., 2013; Jain and Yang, 2005; Jones et al., 2013b; Smith and Rothwell, 2013), and high uncertainty as to how land use might evolve in the future (Hurtt et al., 2011; van Vuuren et al., 2011a; Wise et al., 2009). Part of the uncertainty in future land use trajectories is due to inherent unpredictability of human actions, and part to the high diversity of potential climate mitigation and adaptation scenarios. Several energy and land strategies have been proposed to mitigate climate change (Rose et al., 2012; Smith et al., 2013a), and while these strategies have similar overall goals, some strategies will likely compete for land and other resources if implemented simultaneously. For example, afforestation and bioenergy production both aim to reduce atmospheric CO$_2$ concentrations, but both activities require land area, and both strategies would impact crop production and markets through effects on crop area (Reilly et al., 2012).

Reflecting this limited understanding of land use effects on climate and carbon, Global Climate Models (GCMs), and also next generation Earth System Models (ESMs) that include fully coupled atmosphere–land-ocean carbon cycles, implement a wide range of land use/cover approaches with varying degrees of detail and limited inclusion of managed ecosystems and land use practices (Brovkin et al., 2013; Pitman et al., 2009). The Land Use and Climate, IDentification of robust impacts (LUCID) activity employed seven GCMs to determine whether land use change has significant regional climate impacts and farther-reaching teleconnections due to biophysical changes in land surface. The results for 1972–2002 revealed significant but incon-
sistent changes in temperature, precipitation, and latent heat in some areas where land use change had occurred. The authors concluded that the model disagreement was due mainly to differences in land use/cover change implementations and corresponding land type distributions, with contributions from methodological differences in crop phenology, albedo, and evapotranspiration (Pitman et al., 2009). The variables addressed by LUCID are also key variables for determining carbon uptake by vegetation, and thus it is not surprising that the Coupled Climate-Carbon Cycle Model Intercomparison Project (C^4MIP) activity generated ESM projections that range from the land being a carbon source to a large carbon sink by 2100 (Friedlingstein et al., 2006).

To advance the scientific understanding of the effects of land use change on climate, phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al., 2012) applied a novel “land use harmonization” approach to produce the required land use change information for GCMs and ESMs. The Global Land use Model (GLM) was used for this harmonization to generate the first set of continuous, spatially gridded land use change scenarios for the years 1500–2100 (Hurtt et al., 2011). GLM computes land use states and transitions annually at half-degree, fractional spatial resolution, including secondary land age, area, and biomass, and the spatial patterns of shifting cultivation and wood harvesting (Hurtt et al., 2006). Land use products from GLM have successfully been used as inputs to both regional and global dynamic land models (Baidya Roy et al., 2003; Hurtt et al., 2002; Shevliakova et al., 2009) and fully coupled ESMs (Jones et al., 2011; Shevliakova et al., 2013). The land use harmonization process ensures a continuous transition from the historical reconstructions to the future projections made by Integrated Assessment Models (IAMs).

The land use harmonization methodology was designed to satisfy the demands of a broad range of models and to provide a consistent set of land use inputs for GCMs and ESMs. The historical period of the land use harmonization (1500–2005) was based on version 3.1 of the Historical Database of the Environment (HYDE; Klein Goldewijk et al., 2011) and Food and Agriculture Organization (FAO) wood harvest data. For the future period (2005–2100), the harmonization process utilized land use data from the
four Representative Concentration Pathways (RCPs), each provided by a different IAM. The RCP scenarios were designed to each meet a different radiative forcing target (2.6, 4.5, 6.0, and 8.5 Wm$^{-2}$), and the implementations were unique to each model and thus spanned a range of approaches for all sectors, including land use (van Vuuren et al., 2011a). As a result, forest cover change varied widely from deforestation to afforestation across the scenarios. Once the land use data were passed through the land use harmonization, each GCM/ESM utilized a unique subset of the harmonized outputs, based on model capabilities, and applied it to a unique set of land use and land cover types (e.g. Lawrence et al., 2012). Although this process was largely successful in enabling the first spatially explicit land use driven climate change experiments, it introduced considerable uncertainty into the climate response for a given RCP in part because of model-specific translation requirements between harmonized outputs and simulated land use and cover (Brovkin et al., 2013). This implies that a more detailed specification of the relationship between land use and land cover may improve earth system simulations.

Recent analyses of CMIP5 results using prescribed CO$_2$ concentrations have also showed the land ranging from a carbon source to a sink in 2100 for a given scenario (Brovkin et al., 2013; Jones et al., 2013b). The LUCID activity was repeated for five CMIP5 ESMs and the results demonstrated that large inter-model spreads of key regional land surface variables (temperature, precipitation, albedo, latent heat, and available energy) were still due mainly to differences in land use/cover change implementations and corresponding land type distributions. Inter-model spreads of CO$_2$ emissions, however, were attributed mainly to differences in land carbon cycle process parameterizations. As a result, different land type distributions among the models gave significantly different regional changes in climate associated with land use change, but with insignificant effects on global mean temperature. Furthermore, the range of net cumulative land use change emissions from 2006 to 2100 for RCP8.5 was 34 to 205 PgC, with the high estimate likely due to the combination of relatively high levels of land carbon and the inclusion of all land use transitions rather than just net land use
change (Brovkin et al., 2013). Additionally, not all of the models used the GLM wood harvest data, further contributing to the spread of model results. For comparison, estimates of net cumulative carbon emissions during 1700–2000 (1850–2000) range from 138–250 PgC (110–210 PgC) (Table 3 in Smith and Rothwell, 2013). The differences in land implementations are also a main factor in projecting a large spread of land carbon uptake during the 21st century. The inter-model spreads for individual scenarios are greater than the inter-scenario spreads for individual models. The uncertainties generated by different land use implementations also contribute to the wide spread of compatible fossil fuel emissions allowable for a given RCP (Jones et al., 2013b). It is apparent that further work is needed to resolve inconsistencies among land use and land cover approaches to reduce climate uncertainty, especially for regional impact assessment.

Additional sources of climate uncertainty related to land use are the RCP radiative forcing targets, which include only emissions of GreenHouse Gases (GHGs) and some aerosols and reactive gases (van Vuuren et al., 2011a). These targets do not include radiative forcing from albedo change or other climate effects associated with land use change. In a recent modeling experiment, two different carbon tax policies with dramatically different land use scenarios met the same radiative forcing target (4.5 W m$^{-2}$) in the IAM used for RCP4.5 but had significantly different radiative forcing in an ESM due to albedo differences between the land use projections (Jones et al., 2013a). Furthermore, the shared socioeconomic pathways for mitigation, adaptation, and impact studies in the Intergovernmental Panel on Climate Change (IPCC) fifth Assessment Report (AR5) are likely to produce different land use distributions that meet the same RCP target. As a result of the wide range of land-related uncertainties in climate projections, an increased emphasis on land dynamics is a high priority for CMIP6 (Meehl et al., 2014).

Our approach to addressing inconsistencies between IAMs and ESMs is to integrate an IAM and an ESM into the first fully coupled model that directly simulates human–environment feedbacks. The resulting integrated ESM (iESM) includes vege-
tation productivity feedbacks from the Community ESM (CESM) to the Global Change Assessment Model (GCAM) to facilitate land use projection every five years (Bond-Lamberty et al., 2014). The iESM uses GLM as in the CMIP5 land use harmonization, along with the CESM Land Use Translator (LUT) that converts harmonization outputs to CESM land cover and wood harvest area. Our initial iESM simulations were successful in passing evolving vegetation productivities from CESM to GCAM. However, these simulations also demonstrated that the large RCP4.5 afforestation signal was not being passed through from GCAM to CESM.

Here we test the feasibility of restoring the lost afforestation signal by using the iESM as a test bed to explore alternative coupling strategies. We focus on modifications to the CESM LUT because initial modifications to GLM did not restore CESM afforestation. One advantage of focusing on a post-land use harmonization approach is that it could be applied to other ESMs independently without changing the land use harmonization product. Section 2 includes model description and experimental design, Sect. 3 presents results and demonstrates that this problem exists in CMIP5, and Sect. 4 discusses the limitations of our current approach and the implications for the CMIP5 archive with respect to land use and climate. We conclude with suggestions for improving IAM to ESM land coupling for future model inter-comparisons.

2 Methods

2.1 iESM Description

The iESM integrates GCAM, GLM, and CESM to evaluate the effects of human-environment feedbacks on the earth system (Fig. 1). We have completed the first coupling stage that allows GCAM to project land use distribution in five year increments based on the previous five years of CESM vegetation productivity. The second stage will incorporate atmospheric linkages between GCAM and CESM at this 5 year interval (e.g. temperature, emissions). Here we give a brief overview of how the three
main components interact. A more detailed description of iESM development will be presented in a forthcoming paper (Collins et al., 2014).

GCAM v3.0 (Calvin et al., 2011; henceforth referred to as GCAM) is a tightly coupled IAM of human and biogeophysical processes associated with climate change. GCAM’s human system components simulate global economic activity within energy, agriculture, and forest product markets with respect to 14 geopolitical regions. A previous version of GCAM projected land use distributions for each of the 14 geopolitical regions (Wise et al., 2009) and was used to generate the CMIP5 RCP4.5 scenario (Thomson et al., 2011). Currently, GCAM incorporates a range of improvements to the Agriculture and Land Use (AgLU) module, including the capacity to operate on 151 land units to generate a more detailed and accurate spatial distribution of land use. There are three land types that remain constant over time (urban, tundra, and rock/ice/desert) and 24 land types available for redistribution, including 12 food and feed crops, five bioenergy crops, and seven managed and unmanaged ecosystems (Kyle et al., 2011; Wise and Calvin, 2011). The land units are defined by delineating 18 global agro-ecological zones (Lee et al., 2005) by the 14 geopolitical regions. Land use distributions are projected within these land units for the final year of GCAM’s time step, which is five years in the iESM.

In a second and intermediate step, GLM uses GCAM’s cropland, pasture, and forest areas (and wood carbon harvest) to compute all annual, fractional land use states and transitions. As part of this process it disaggregates the land unit data to a half-degree global grid by computing spatial patterns and also ensures consistency with the historical land use reconstructions (Hurtt et al., 2011, 2006). GLM has been slightly modified from its CMIP5 implementation to better facilitate forest area change matching with GCAM (Sect. 2.3.2). This modification enables GLM to use forest area output from GCAM that was not incorporated into the CMIP5 land use harmonization. Nonetheless, iESM follows the CMIP5 implementation for CESM land use harmonization by ingesting these GLM outputs: cropland, pasture, primary, and secondary land area, as well as wood harvest areas on primary and secondary forested and non-forested land.
CESM (Bitz et al., 2011; Gent et al., 2011) has fully coupled atmosphere, ocean, land, and sea ice components. Within CESM, the Community Land Model v4.0 (Lawrence et al., 2011) receives the selected GLM outputs via a translator that converts these outputs to 16 CLM Plant Functional Types (PFTs; eight forest, three grass, three shrub, one bare soil, and one crop) (Lawrence et al., 2012). The version of iESM used in this study was based on CESM v1.0beta9, which is a pre-release version of the model used for the CMIP5 simulations.

### 2.2 Simulations

Our iESM simulations cover 2005 to 2040 with fully coupled CESM components and prescribed RCP4.5 emissions and carbon price path. These simulations operate in transient land use mode with a dynamic ocean (Smith et al., 2013b), Community Atmosphere Model v4 physics (Gent et al., 2011), carbon-nitrogen biogeochemistry (Thornton et al., 2007), and active land–atmosphere–ocean carbon dynamics, at approximately 1° resolution (0.9375° × 1.25°). The iESM initial conditions are the culmination of a CESM spinup run followed by a CESM historical transient land use run from 1850 to 2005. GCAM initial conditions are calibrated to 2005 energy, economic, and agricultural values as reported by countries and processed and archived by international organizations (e.g. IEA, FAO). The GCAM scenario for RCP4.5 was described fully by Thomson et al. (2011).

We performed two fully integrated simulations to compare two iESM cases: (1) original CESM land use translator (OLDLUT) and (2) modified CESM land use translator (NEWLUT) (Table 1). In fact, OLDLUT was our initial fully integrated simulation with iESM and, as documented below, it revealed inconsistencies within iESM that needed to be addressed prior to scientific experimentation. OLDLUT also showed that the updated GLM did not increase CESM afforestation with respect to a previous simulation performed by manually passing data between the respective iESM models. The NEWLUT case was used to test our hypothesis that the lost afforestation signal could be recovered by modifying only the CLM component of iESM. These fully integrated runs
included vegetation productivity feedbacks into the land use projections at five-year intervals. Analysis of the effects of introducing these feedbacks on land use and climate will be presented in a forthcoming paper (Thornton et al., 2014).

2.3 Land use coupling

2.3.1 OLDLUT land use coupling within iESM

The OLDLUT iESM land use coupling followed the CMIP5 land use harmonization algorithm (Fig. 2), but with a slightly modified version of GLM (see Sect. 2.3.2). The coupling was designed to match GCAM and CLM changes in absolute cropland and pasture area. For CMIP5, GLM received only crop and pasture areas from GCAM, but for the iESM GLM also ingests forest area from GCAM to better facilitate forest area change matching (see Sect. 2.3.2). GLM also receives wood products demand from GCAM (in tons of carbon), which is spatially distributed to determine the extent of harvested area in each of five wood harvest types (primary forest harvest, primary non-forest harvest, secondary mature forest harvest, secondary immature forest harvest, and secondary non-forest harvest). The LUT (Fig. 3) uses only the cropland and pasture area outputs from GLM to update CLM PFTs in conjunction with maps of potential vegetation (Ramankutty and Foley, 1999). The LUT does not use the primary and secondary land area information for updating PFTs because CLM does not keep track of these land designations. The LUT does, however, use the primary and secondary land area to normalize the GLM total harvest area, which is applied only to forest area in CLM (Lawrence et al., 2012).

The vegetation feedbacks were implemented by passing annual climate scaling factors from CLM to GCAM based on Net Primary Productivity (NPP) and Heterotrophic Respiration (HR). These factors were used to scale GCAM vegetation and soil carbon densities, respectively. To calculate the scaling factors, the per-pixel, PFT-specific CLM 5 year annual average NPP and HR values for a given GCAM time step were divided by base-period average annual values (1990–2004) and linearly interpolated to individ-
ual years between the GAM time step endpoints. The inverse was used in the case of the HR factor to simulate reductions in soil carbon density with increasing HR. These CLM scaling factors were then filtered to exclude outliers based on a median absolute deviation method, and finally aggregated to GCAM’s land units and types and applied to the next GCAM time step (for details see Bond-Lamberty et al., 2014).

2.3.2 Modifying the GLM spatial distribution algorithm

For the iESM, GLM was modified to better facilitate forest area change matching with GCAM in an effort to increase the afforestation simulated by CESM. These modifications included operating on GCAM’s 151 land units (rather than the 14 regions used for CMIP5) in addition to using GCAM’s forest area output, which was not previously shared between the models. For CMIP5, GLM applied the cropland and pasture area changes to the 2005 half-degree maps of cropland and pasture while preserving the total cropland and pasture area changes within GCAM regions. Spatial allocation to the half-degree grids was done with a preference for expanding agricultural area onto non-forested land, and contracting agricultural area on potentially forested land, while also preserving the 2005 spatial patterns by allocating new cropland and pasture near to existing agricultural areas (Hurtt et al., 2011). The new GLM algorithm uses GCAM forest area from each land unit at each time step and attempts to preserve the forest area changes within each land unit in addition to preserving the cropland and pasture area changes. GLM has previously defined “forest” as natural vegetation that is growing on land where the potential biomass density is greater than 2 kgC m\(^{-2}\). Using this definition the potential forestland within GLM is fixed and, as a result, the GLM algorithm cannot grow forest outside of this forestland. In the new algorithm, GLM matches GCAM forest area changes by moving cropland and pasture around within each land unit to “expose” enough potential forestland for regrowth to meet the GCAM forest area changes. In addition, to meet GCAM’s land requirements for afforestation, GLM uses a different definition of “forest” (potential biomass density greater than 1 kgC m\(^{-2}\), rather than 2 kgC m\(^{-2}\)) than the definition used elsewhere in the GLM code (e.g. for comput-
ing the spatial pattern of wood harvesting). The new GLM algorithm operates in three main steps:

(a) Decreases in cropland and pasture occur first on the highest potential biomass land and increases in cropland and pasture occur first on the lowest potential biomass land.

(b) If the forest area change within a land unit is not met, a redistribution of cropland and pasture within that land unit occurs such that, when possible, existing cropland and pasture is moved from high biomass density land to low biomass density land.

(c) If the forest area change within a land unit is still not met, the algorithm attempts to allocate any “unmet” forest area change within another land unit (or across multiple land units) within the same region, using a similar method to (b) above.

2.3.3 Modifying the CLM land use translation algorithm

To test our hypothesis that the lost afforestation signal could be recovered solely by the ESM component, we focused on modifying the LUT to capture GCAM afforestation via changes in agricultural land. This approach is more expedient than redesigning the coupling code and LUT to receive forest area changes directly from GLM because such redesign would logically require implementation of a single, consistent land surface and carbon cycle among all iESM components. Specifically, we modified the LUT to add tree PFTs when cropland and pasture are removed (Fig. 4). Furthermore, the LUT preferentially removes tree PFTs when cropland and pasture are added. Forest area information is still not shared between GLM and the LUT (other than forest harvest). We also fixed a few bugs in the LUT code, including one associated with normalizing area between GLM and CLM.
2.3.4 Pre- and post-harmonization CMIP5 RCP4.5 land use distributions for CESM

The OLDLUT iESM land use coupling was also used in CMIP5, albeit with 14 regions rather than 151 land units and without the GLM modifications and vegetation productivity feedbacks described above, and so we explored the extent to which the afforestation signal was lost in the CMIP5 simulations. We compared the RCP4.5 pre-harmonized forest and pasture area outputs with the harmonized GLM values and also with the corresponding PFT inputs for the CESM1.0-BGC simulations submitted to the CMIP5 archive. CESM1.0-BGC served as the base code for iESM and thus contains the same versions of the model components.

3 Results

3.1 CMIP5 RCP4.5 land area inconsistencies

The CMIP5 RCP4.5 afforestation signal was dramatically decreased in the CESM RCP4.5 simulations, and total herbaceous area increased while RCP4.5 pasture decreased (Fig. 5). CLM simulated 23 % of the 4.82 million km$^2$ of afforestation between 2005 and 2020, and 22 % of the 10.98 million km$^2$ increase by 2100. GLM captured 64 % and 56 % of the afforestation in 2020 and 2100, respectively. RCP4.5 and GLM pasture decreased by 4.69 million km$^2$ from 2005 to 2100 while CLM herbaceous PFTs increased by 1.11 million km$^2$ over the same period. The changes in global cropland area were faithfully transmitted (CLM decreases were only 7 % less than RCP4.5 decreases), but absolute CLM cropland area was approximately 1.5 million km$^2$ less than RCP4.5 cropland area throughout the simulation (data not shown). Changes in GLM pasture and cropland areas were essentially identical to RCP4.5 changes, and GLM absolute area values were slightly higher and lower, respectively, than RCP4.5 pasture and cropland areas (cropland data not shown).
3.2 Restored afforestation in iESM

The OLDLUT simulation revealed that only changes in crop area were being faithfully transmitted from GCAM to CLM (Fig. 6). In contrast, CLM simulated only 17% of GCAM’s 5.40 million km$^2$ of afforestation between 2015 and 2020, and only 17% of the 7.73 million km$^2$ of afforestation between 2015 and 2040. Changes in GLM forest area, on the other hand, reflected changes in GCAM forest area quite well (Fig. 6), but at the cost of dramatically overestimating absolute forest area within GLM due to a low biomass threshold for defining forest (Fig. 7). Within GLM, the new algorithm captured 93% of afforestation between 2015 and 2020 and 84% between 2015 and 2040, as compared to the original GLM algorithm that captured only 14% and 20% over the respective periods in a previous simulation performed by manually passing data between the respective iESM models (data not shown). Changes in GCAM pasture were not reflected by changes in CLM herbaceous (grass and shrub) PFTs, but were faithfully output by GLM (Fig. 6).

The NEWLUT simulation shows improved forest and cropland area changes in CLM with a corresponding change in CLM herbaceous vegetation. The main improvement is that CLM simulates 64% of GCAM’s 2015–2020 afforestation and 66% of the 7.71 million km$^2$ of afforestation from 2015–2040 (Fig. 6). This afforestation in NEWLUT reduces total herbaceous area in CLM by 94% of the 4.36 km$^2$ of GCAM pasture loss by 2040. Figure 8 shows the spatial tradeoff between forest and herbaceous PFTs that achieves this level of afforestation, and Fig. 9 demonstrates a sustained increase in average annual land carbon uptake after 2020 due to additional afforestation. NEWLUT also improves CLM’s simulation of absolute cropland area (Fig. 7) through normalization of GLM and CLM area. This area normalization is apparent in the cropland and pasture area changes from 2005 to 2006 (Figs. 6 and 7). GLM NEWLUT outputs follow the GCAM NEWLUT outputs with relationships between GLM and GCAM similar to those for OLDLUT (data not shown).
4 Discussion

The iESM and RCP4.5 CESM land area discrepancies (Figs. 5–7) result from a gap in the original CMIP5 land coupling design that allows inconsistent forest area and land cover type definitions across models (Fig. 2), along with different underlying carbon cycles. The land use harmonization was, however, ambitious and largely successful in developing consistent land use definitions and data without requiring extensive redevelopment of land use/cover components of all participant models (Hurtt et al., 2011). As this study attests, such redevelopment is challenging and model-specific, but may be required for ESMs to adequately simulate climate and energy policies generated by IAMs. Thus, while this is a specific case, the lost iESM afforestation signal is instructive of the shortcomings of the CMIP5 design and the restoration of this signal offers insights into improving land use coupling for model inter-comparisons.

A primary challenge for improving the CMIP5 land coupling is to increase the amount of specific land cover information being shared between the historical period, IAMs, and ESMs. For CMIP5, the land use harmonization was designed to harmonize land use data between models, not land cover data, and as such GLM did not ingest forest information from any of the IAMs. GLM does, however, keep track internally of whether the gridded primary or secondary land is forested or non-forested, and its mean biomass density (according to GLM’s definition of forest, which may differ from the IAM or ESM). To satisfy the demands of a broad range of models, the land use harmonization product includes only cropland, pasture, primary, and secondary land areas and transitions, and the age and biomass density of secondary land (and harvest areas, carbon amounts, and transitions, which we do not address here), while each ESM and IAM characterizes the land surface by its own suite of vegetation and management types (Brovkin et al., 2013; Masui et al., 2011; Riahi et al., 2011; Thomson et al., 2011; van Vuuren et al., 2011b). For example, GCAM has 19 crop types (the CMIP5 RCP4.5 version had 10) and seven managed and unmanaged land cover types while CESM has 16 PFTs, only one of which is a crop type. The LUT algorithm uses only the GLM crop
and pasture area information to adjust PFTs because CLM does not keep track of primary vs. secondary land. The resulting spatial pattern of non-crop PFTs is determined by the existing PFT distribution and CLM's internal representation of potential vegetation cover (Lawrence et al., 2012; Ramankutty and Foley, 1999). An additional source of error that we did not investigate here is the relationship between individual PFTs and land cover types that may comprise several PFTs (e.g. forest land may consist of 60% trees and 40% grass).

Thus, forest area changes in CESM (and iESM) are effectively residual changes that are only indirectly linked to GCAM forest area through changes in crop and pasture. The LUT calculates cropland area changes first and pasture second (Figs. 3 and 4). In CMIP5 CESM simulations, cropland area changes cause non-crop PFTs to be added or removed in proportion to their potential or existing grid-cell fractions, respectively. Pasture is more complicated because it is not tracked as such: pasture is not a single PFT and its changes are represented as changes in herbaceous and tree PFTs. Specifically, tree PFTs are removed when pasture is added, and non-crop PFTs are added in proportion to their potential vegetation grid-cell fractions when pasture is removed (Lawrence et al., 2012). This residual PFT determination, combined with independent and unique forest definitions across GCAM, GLM, and CLM, causes the bulk of RCP4.5 afforestation to not appear in the CMIP5 CESM land surface. As a direct consequence, the CMIP5 CESM RCP4.5 grass area (and shrub area to a lesser extent) increases while RCP4.5 pasture decreases dramatically (Fig. 5). CESM has this same limitation for all four RCP scenarios, and the other CMIP5 ESMs implement similar inconsistencies to varying degrees due to the lack of specific vegetation types in the land coupling between IAMs and ESMs. For example, Davies-Barnard et al. (2014) recently reported that the HadGEM2-ES RCP4.5 forest area increased 11% from 2005–2100, while the RCP4.5 forest area increased by 24%.

Even partial restoration of the lost afforestation signal in iESM has a significant impact on global carbon balance. The assumption that forest exclusively replaces abandoned cropland and pasture in GCAM's land use projection (Figs. 6–8) sets the upper
limit for CLM because there is no other information to constrain forest area. This is reasonable for the RCP4.5 case, yet even so NEWLUT simulates only two-thirds of the total afforestation. Nonetheless, the increased afforestation in NEWLUT results in an increase in net land carbon uptake over the OLDDLUT case due to a sustained increase in average annual land carbon uptake after 2020 (Fig. 9). As a result, the NEWLUT simulation increases vegetation carbon gain by 19 PgC and decreases atmospheric CO$_2$ gain by 8 ppmv from 2005 to 2040 in comparison to OLDDLUT (data not shown). Linear extrapolation of these carbon pools from 2020 to 2100 increases these changes to approximately 60 PgC and 25 ppmv, and extending CLM forest area to match GCAM total afforestation could potentially increase these changes to 100 PgC and 40 ppmv in 2100. For comparison, estimates of net cumulative land use change emissions during 1850–2000 range from 110–210 PgC (Table 3 in Smith and Rothwell, 2013). While these carbon cycle changes in the CESM component of iESM may have a significant effect on climate, it is important to note that the carbon cycle effects of afforestation in CESM are not identical to those in GCAM or GLM because these three models have different biogeochemistry and vegetation models.

Since the low OLDDLUT forest area increase was constrained by potential vegetation and the higher NEWLUT forest area increase is incomplete, it is likely that GCAM’s RCP4.5 afforestation signal is overly optimistic from a strictly bioclimatic standpoint, especially considering that it is not constrained by a map of potentially viable forest. However, it is important to note that GCAM’s projected afforestation assumes that it is cost effective to use agricultural inputs (e.g. water, fertilizer) to achieve the expected forest growth. CLM does not simulate these additions, and so its rate of forest carbon accumulation is constrained by environmental conditions. Figure 8 shows that most of the forest afforestation occurs on grassland and shrubland, and that these lands generally coincide with areas of limited potential forest. Additionally, we know that GLM’s spatial disaggregation of GCAM’s afforestation relies on a map of potential forest that is not necessarily in agreement with CLM’s map of potential forest. This disagreement among the three models is a major contributor of uncertainty to any communication of
forest area changes, especially when forest area information is not passed between models, as is the case for the CMIP5 land coupling. With respect to iESM, we plan to reconcile the land use inconsistencies by implementing a single carbon cycle with a consistent land characterization for all components. This iESM redevelopment could provide a template for improving the land coupling between IAMs and ESMs in the next CMIP. In fact, land cover information is already planned to be included in the CMIP6 land coupling, along with a more extensive land use model intercomparison project (Meehl et al., 2014).

We have focused on understanding the effects of mismatched land type areas on global simulations, rather than on mismatched carbon cycles, because the spatial distribution of land types is effectively a boundary condition for simulating terrestrial processes. For global simulations this boundary condition is generally provided by historical data and IAMs, and, as we have shown, a mismatch in this boundary condition can determine whether or not an ESM is simulating a particular scenario (in this case, CESM and RCP4.5). Mismatched carbon cycles among IAMs and ESMs, on the other hand, along with differences in atmospheric radiation code, will preclude exact matches in radiative forcing for a given scenario, but should not cause significant deviations in climate scenarios among models. Furthermore, it is not desirable, nor feasible, for all IAMs and ESMs to have the same biogeochemistry and vegetation growth components. For example, a diversity of terrestrial models can help characterize uncertainty in global simulations. This uncertainty, however, is most useful if these models simulate the same spatial distribution of land cover and land use change.

5 Conclusion

We have identified the lack of specific land cover type information being shared among GCAM, GLM, and CESM in the iESM as the primary cause of CESM simulating very little afforestation and effectively no change in herbaceous vegetation in contrast to GCAM’s large RCP4.5 afforestation and corresponding pasture reduction. Initial efforts
to fix this problem through GLM modifications yielded no improvement, so we then fo-
cused on modifying the algorithm that translates GLM land use harmonization outputs
to CLM PFTs. While these land use translator modifications have been successful at
capturing two-thirds of GCAM’s RCP4.5 afforestation signal and corresponding reduc-
tions in herbaceous vegetation, they are not sufficient to completely overcome the limi-
tations imposed by not passing specific land cover types from GCAM through to CESM.
They have also not addressed the lack of constraints on GCAM forest area expansion,
nor mismatches between land cover and PFTs. Nonetheless, this partial restoration of
afforestation has a significant impact on iESM’s global carbon cycle through increased
vegetation carbon and decreased atmospheric CO₂ concentration.

The iESM framework follows the CMIP5 land coupling design, and as such we
have characterized a major gap in this design that precludes accurate translation of
projected IAM land types to ESMs by focusing on land use types such as cropland
and pasture (albeit successfully), rather than specific land cover types such as forest,
grassland, and shrubland. The relationship between land use and land cover type in-
formation is handled uniquely by individual ESMs, which means that the effects will
be model-specific, and is more relevant for some RCPs than others. The resulting
land type discrepancies are likely most pronounced for the large RCP4.5 afforesta-
tion signal, which was greatly reduced in the CMIP5 CESM and HadGEM2-ES (see
Davies-Barnard et al., 2014) simulations, but could also arise for other large land cover
changes such as the extensive deforestation of RCP8.5. As total land area is con-
servative, errors in the coverage of one land type are complemented by errors in the
coverage of other land types. In the RCP4.5 scenario, pasture decreases over the 21st
century, but the CMIP5 CESM runs have increasing grass and shrub areas over the
same period. It is very important that the land use and land cover changes (which de-
termine land use change emissions and the total capacity for vegetation carbon assim-
ilation) match between the IAMs and ESMs because the CMIP5 experimental design
is predicated on the IAMs and ESMs simulating similar, specific radiative forcings for
a given scenario, including CO₂ emissions from land use change (Moss et al., 2010).
Furthermore, future radiative climate targets are likely to include the biogeophysical forcings of land use change because it has been shown that the modeled climate system is sensitive to changes in these forcings due to the spatial distribution of land use and land cover type (Brovkin et al., 2013; Jones et al., 2013a; Pitman et al., 2009), making it imperative that IAM and ESM land use and land cover distributions match as closely as possible. Maintaining the diversity of global biogeochemical and vegetation models also calls for GCMs and ESMs to match historical and projected land cover and land use distributions as closely as possible, so as to isolate carbon cycle contributions to uncertainty from contributions due to differences in land types (as a boundary condition). Fortunately, our results indicate that it might be possible to adjust land cover in other CMIP5 models to better match RCP4.5 afforestation and the corresponding climate scenario, while still using the standard land use harmonization data.

We conclude that the land coupling between IAMs and ESMs for future model intercomparisons needs to ensure greater consistency in land cover and land use among the models. In short, the models need to agree on the actual land area and the annual spatial distribution of major (non-) vegetation land types and land uses. In other words, they need to simulate the same basic land surface. To achieve the required consistency, we suggest that the next CMIP land coupling design provides land cover and land use information, and a standard mapping between land cover and plant functional types. Fortunately, this is an emerging priority for the CMIP6 Land Use Model Intercomparison Project (LUMIP, http://www.wcrp-climate.org/index.php/modelling-wgcm-mip-catalogue/modelling-wgcm-mips/318-modelling-wgcm-catalogue-lumip, http://www.wcrp-climate.org/wgcm/WGCM17/LUMIP_proposal_v4.pdf). The following gridded data with fractional shares within grid cells are specifically recommended:

1. Annual land cover states with complete, contiguous spatial coverage within grid cells. Land cover needs to include at least the basic categories of cropland, grassland, shrubland, woodland, forest, and other (bare/sparse, ice, urban, water). This will allow consistency in major (non-) vegetation types for model intercomparison.
(with the “other” category having fixed area). The “other” categories could also be separated out for models that can use them, and in preparation for changing their areas also.

2. Annual land use states including primary and secondary land, wood harvest, and pasture (cropland should coincide with the land cover state). These uses should be provided with respect to the land cover categories. Wood harvest and pasture should include both area and amount of biomass/carbon harvested or removed by grazing.

3. A standard present-day land area data set to be used by all models. Land area includes all land cover and land use categories as described above.

4. Annual land use and land cover transitions. Land use transitions need to be accompanied by corresponding land cover transitions with complete, contiguous spatial coverage within grid cells. Net transitions, which should be used for model intercomparison, correspond directly with changes in annual land use and cover states, and may include additional detail about sources of wood harvest and grazed biomass. Gross transitions should also be provided to characterize shifting cultivation and other gross land conversions occurring within a particular year. While gross transitions are very important and make a significant difference in the carbon cycle, until more models are able to make use of gross transitions they should not be included in model intercomparisons.

Acknowledgements. We are extremely grateful to Peter Lawrence (National Center for Atmospheric Research) for providing the original CLM land use translation code. This research was funded by the Director, Office of Science, Office of Biological and Environmental Research of the US Department of Energy under contract No. DE-AC02-05CH11231 as part of the Integrated Assessment Research and Earth System Modeling Programs. This project used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the US Department of Energy under contract No. DE-AC02-05CH11231. The CESM project is supported by the National Science Foundation and the Office
of Science (Biological and Environmental Research) of the US Department of Energy. We also gratefully acknowledge the support of the National Aeronautics and Space Administration.

References


Kyle, G. P., Luckow, P., Calvin, K., Emanuel, W., Nathan, M., and Zhou, Y.: Gcam 3.0 Agriculture and Land Use: Data Sources and Methods, Pacific Northwest National Laboratory, 58, 2011.


Table 1. Two integrated Earth System Model (iESM) simulations performed for this study.

<table>
<thead>
<tr>
<th></th>
<th>OLDLUT</th>
<th>NEWLUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Land Use Translator</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Vegetation productivity feedbacks</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Updated Global Land use Model</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
Fig. 1. Status of iESM implementation as of Spring 2014. The light blue arrows show information flow from GCAM to CESM. The light green arrows show information flow from CESM to GCAM. The dashed gray outline, including the crossed out arrows, represents the CMIP5 land coupling. The solid green outline, minus the arrow crossed out by green and including the 100 year emissions arrow, depicts the current iESM implementation. The dashed blue outline, minus both crossed out arrows, indicates ongoing development. The dashed red line, minus the crossed out arrows, includes the next stage of development. GCAM: Global Change Assessment Model. GLM: Global Land use Model. CESM: Community Earth System Model.
Fig. 2. General integrated Earth System Model (iESM) land use coupling algorithm. Forest area is not passed from the Global Change Assessment Model (GCAM) to the Global Land use Model (GLM) in the CMIP5 land use coupling, but it is passed in the iESM simulations used in this study. NPP: Net Primary Productivity. HR: Heterotrophic Respiration. PFT: Plant Functional Type.
Fig. 3. Old Land Use Translator (LUT) algorithm for dynamic Plant Functional Type (PFT) coverage. When cropland and pasture decrease, non-crop PFTs are added in proportion to potential vegetation fractions. When cropland and pasture increase, non-crop PFTs are removed in proportion to reference year fractions.
Fig. 4. New Land Use Translator (LUT) algorithm for dynamic Plant Functional Type (PFT) coverage. When cropland and pasture decrease, non-crop PFTs are added in proportion to potential vegetation fractions. When cropland and pasture increase, non-crop PFTs are removed in proportion to reference year fractions.
**Fig. 5.** Projected global forest, pasture, grass, and shrub areas for the CMIP5 4.5 W m\(^{-2}\) Representative Concentration Pathway (RCP4.5), in million km\(^2\). CLM: Community Land Model. GLM: Global Land use Model. PFT: Plant Functional Type.
Fig. 6. Integrated Earth System Model (iESM) land use and forest area changes with respect to 2015. The GLM-NEWLUT forest and pasture data are nearly identical to the GLM-OLDLUT data and are not shown for clarity. Similarly, the GLM-NEWLUT cropland data are nearly identical to the GCAM-NEWLUT data. CLM: Community Land Model. GCAM: Global Change Assessment Model. GLM: Global Land use Model.
Fig. 7. Integrated Earth System Model (iESM) land use and forest area. The GLM-NEWLUT forest and pasture data are nearly identical to the GLM-OLDLUT data and are not shown for clarity. Similarly, the GLM-NEWLUT cropland data track the GCAM-NEWLUT data, but with the same offset as for the GLM-OLDLUT data. CLM: Community Land Model. GCAM: Global Change Assessment Model. GLM: Global Land use Model.
Fig. 8. Spatial distributions of iESM increased forest Plant Functional Types (PFTs), decreased grass and shrub PFTs, and potential forest PFTs, as percentages of land area within each grid cell. (a) Difference in 2040 forest PFT area (NEWLUT – OLDLUT). (b) Difference in 2040 grass plus shrub PFT area (NEWLUT – OLDLUT). (c) Potential forest PFT area.
Fig. 9. Difference in Net Ecosystem Exchange (NEE) between iESM simulations (NEWLUT minus OLDLUT), as computed by CLM. These data show more land carbon uptake, associated with the additional trees, for the NEWLUT simulation during the afforestation period (2015 forward).