Algorithm Theoretical Basis Document (ATBD)

AirMOSS Level 4A/B Net Ecosystem Exchange of CO_2 (L4A/B-NEE) Data Products

Ke Zhang[1,2], Ashehad Ali[1], Alexander Antonarakis[1,3] and Paul Moorcroft[1*]

1. Department of Organismic & Evolutionary Biology, Harvard University, Cambridge, MA, USA

[2]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK, USA

[3]Department of Geography, University of Sussex, Brighton, UK

***Corresponding author:** Paul R. Moorcroft
Harvard University
Department of Organismic & Evolutionary Biology
Suite 43
26 Oxford Street
Cambridge, MA, 02138
USA
Email: paul_moorcroft@harvard.edu
Phone: +1-617-496-8703

AirMOSS Level 4A/B Net Ecosystem Exchange of CO_2 (L4A/B-NEE) Data Product Algorithm Theoretical Basis Document

EXECUTIVE SUMMARY

The object of Airborne Microwave Observatory of Subcanopy and Subsurface (AirMOSS) mission is to provide high-resolution observations of root-zone soil moisture over regions representative of the major North American Biomes, quantify the impact of variations in soil moisture on the estimation of regional carbon fluxes, and extrapolate the reduced-uncertainty on regional carbon fluxes to the continental scale of North America.

The main objectives of the AirMOSS Level 4 Net Ecosystem Exchange of CO_2 between atmosphere and ecosystems (NEE) products are:

(i) to quantify the impact of root-zone soil moisture (RZSM) on the estimation of regional carbon fluxes and provide estimates of NEE based on AirMOSS observations, and

(ii) to upscale the reduced uncertainty estimates of regional carbon fluxes to the continental scale of North America.

L4-NEE consists of two products: L4A-NEE and L4B-NEE. The L4A-NEE product consists of hourly estimates of NEE and its constituent carbon fluxes and corresponding estimates of the sensitivity these fluxes to changes in soil moisture at a spatial resolution of 30 arc-seconds (~1 km) for each of the 10 sites covered by AirMOSS flights, each site spanning ~2500 km^2. The hourly values span the measurement periods at each site and are also reported as monthly integrals. The L4B-NEE scales up the L4A-NEE product to provide a continental-scale estimate of NEE and its constituent carbon fluxes for the continental US and accompanying estimates of the sensitivity these fluxes to changes in soil moisture at hourly and monthly time scales at a spatial resolution of 50 km. The algorithm for the carbon flux and carbon flux sensitivity estimates is the integrated terrestrial biosphere model (ED2) that incorporates hydrology, land-surface biophysics, vegetation dynamics, and soil carbon and nitrogen biogeochemistry, and takes account of fine-scale ecosystem heterogeneity and its impact on large-scale ecosystem function.

RZSM and its spatial and temporal heterogeneity influences NEE. The airborne microwave remote sensing retrieved RZSM is used to estimate the sensitivity of carbon fluxes to soil moisture and to diagnose and improve estimation and prediction of NEE by constraining the model's predictions of soil moisture and its impact on above- and below-ground fluxes. Pre-flight calibration of parameterization of ED2 was conducted with observations from the FLUXNET carbon fluxes observations optimized using a Bayesian Markov-Chain Monte Carlo (MCMC) framework.

TABLE OF CONTENTS

1. Algorithms Overview & Objectives

North American ecosystems are critical components of the global carbon cycle, exchanging large amounts of carbon dioxide and other gases with the atmosphere. Net ecosystem exchange of CO_2 between atmosphere and ecosystems (NEE) quantifies these carbon fluxes, but current continental-scale estimates contain high levels of uncertainty. Spatial and temporal heterogeneity in Root-zone soil moisture (RZSM) exerts strong controls on NEE, and, as a result, uncertainty in soil moisture is often major source of variation in estimates of NEE and its component carbon fluxes. The objective of the AirMOSS L4A/B-NEE algorithm is to (1) quantify the impact of RZSM on the estimation of regional carbon fluxes and provide estimates of NEE based on AirMOSS observations, and (2) provide a new NEE product for the continental United States by upscaling the reduced uncertainty estimates of regional carbon fluxes to the continental scale.

L4-NEE products are divided into two categories: L4A-NEE and L4B-NEE. L4A-NEE combines the L4-RZSM products and their spatially explicit heterogeneity with the terrestrial biosphere model (ED2). The L4A-NEE is generated at hourly time steps at the spatial resolution of 30 arc-seconds (~1 km) for each of the sites covered by AirMOSS flights (~1°×0.25°; i.e. ~2500 km^2)) and also reported as monthly composite. The L4B-NEE scales up the L4A-NEE products to the North American continental scale via two methods: spatial extrapolation and continental-scale modeling. It provides continental-scale estimate of NEE for North America and accompanying assessments of the carbon flux sensitivities to changes in soil moisture. The L4B-NEE products are provided at hourly and monthly time scales with a spatial resolution of 50 km.

2. L4A-NEE Algorithm

The L4A-NEE product is comprised of hourly 30″ estimates of Net Ecosystem Exchange (NEE), its constituent fluxes (Gross Primary Productivity (GPP) and Ecosystem Respiration (R_{eco})) and the moisture sensitivities of these fluxes (β_{soil}, β_{GPP}, β_{Reco}) over the ~ 2500 km^2 radar flight lines. This section describes the ED2 model as well as the specification of the ecosystem's initial conditions and meteorological forcing needed for the model simulations.

2.1 The Ecosystem Demography model

The Ecosystem Demography Biosphere Model (ED2) is an integrated terrestrial biosphere model incorporating hydrology, land-surface biophysics, vegetation dynamics, and soil carbon and nitrogen biogeochemistry (Medvigy *et al*. 2009). Like its predecessor, ED (Hurtt *et al*. 1998, Moorcroft *et al*. 2001), ED2 tracks the changing abundance of plants of different sizes and plant functional types arising from plant growth, mortality, recruitment, and the impact of disturbances using the following set of size- and age-structured (SAS) partial differential equations (PDEs):

$$\underbrace{\frac{\partial}{\partial t}n^{(i)}(z,a,t)}_{\text{chang in plant density}} = \underbrace{-\frac{\partial}{\partial z}\left[g^{(i)}(z,\bar{r},t)n^{(i)}(z,a,t)\right]}_{\text{plant growth}} - \underbrace{\frac{\partial}{\partial a}n^{(i)}(z,a,t)}_{\text{aging of plant community}} - \underbrace{\mu^{(i)}(z,\bar{r},t)n^{(i)}(z,a,t)}_{\text{mortality}} \quad (1)$$

$$\underbrace{\frac{\partial}{\partial t}p(a,t)}_{\text{change in age structure}} = \underbrace{-\frac{\partial}{\partial a}p(a,t)}_{\text{aging}} - \underbrace{\lambda(a,t)p(a,t)}_{\text{disturbance}} \quad (2)$$

where

$$\frac{\partial}{\partial z} = \left\{\frac{\partial}{\partial z_s}, \frac{\partial}{\partial z_a}\right\} \quad , \quad \int_0^\infty p(a,t)da = 1$$

Equation (1) relates the change in expected plant density ($n^{(i)}$) of plant functional type *i* as a function of the growth, mortality, and age of a plant community, where *z* is the size of individuals, *a* is the time since last disturbance, *r* is a vector describing the resource environment (light, water, nitrogen) experienced by an individual of a certain size, and *t* is time. The functions $g^{(i)}(z,\bar{r},t)$ and $\mu^{(i)}(z,\bar{r},t)$ represent the growth and mortality factors at any time *t*. The growth can be further described as an array of structural (z_s) and active tissue (z_a) growth compartments. Equation (2) describes the changes in the distribution of landscape ages since the last disturbance event, where $\lambda(a,t)$ is the rate of disturbance. Equation (1) has two boundary conditions. The first describes the recruitment of new seedlings, which corresponds to a flux of individuals into the system at (z_0, *a*) under the assumption of random dispersal of seeds between gaps within a grid cell. The second describes the state of the ecosystem following a disturbance event, relating to the survivorship of individuals following the disturbance event of the plant of type *i* and size *z*. Equation (2) has a boundary condition describing the fraction of newly disturbed areas within a grid cell (see Moorcroft *et al.* 2001 for further details).

The size and age structure approximation is completed by initial conditions corresponding to the initial size distribution for Equation (1), and the initial age distribution for Equation (2) of the plant types:

$$\underbrace{n^{(i)}(z,a,t_0) = n_0{}^{(i)}(z,a)}_{\text{initial plant community}} \qquad \underbrace{p(a,t_0) = p_0(a)}_{\text{initial age disturbance}} \quad (3)$$

With this system of equations ED2 is able to realistically represent the dynamics of spatially heterogeneous plant communities incorporating the effects of natural disturbance processes such as fire, and anthropogenic disturbances such as forest harvesting or land clearing (e.g. Hurtt *et al*. 2004; Albani *et al*. 2006; Medvigy, *et al*. 2009).

The simulation region is subdivided into grid cells that experience the same meteorological forcing specified either from a meteorological forcing dataset, or from the boundary conditions of an atmospheric model (Figure 2a). Each grid cell is subdivided into a number of horizontal tiles/patches representing areas of forest that share a similar vegetation canopy structure and disturbance history. Long-term vegetation dynamics are calculated by integrating short-term

carbon dynamics (Figure 2b) of individual plants, that, in turn, drive the dynamics mortality, growth and recruitment (Figure 2c). The different plant functional types *i* differ in terms of their physiological, allometry, and other plant trait that that results in different rates of growth and mortality and sensitivity to environmental conditions.

The description of the above ground ecosystem state embodied in PDEs (1) and (2) enables the ED2 model to make realistic projections of both the fast-timescale exchanges of carbon, water and energy between the land and the atmosphere, and long-term vegetation dynamics incorporating the effects of ecosystem heterogeneity, including disturbance history and recovery.

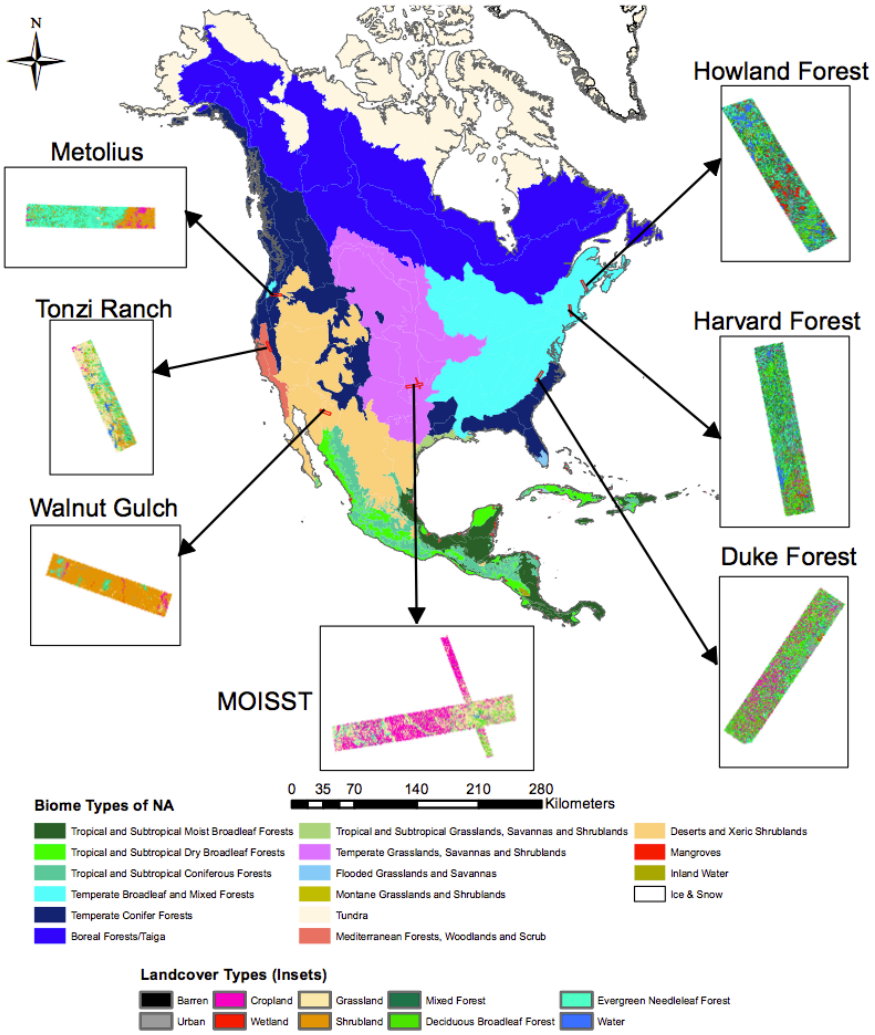


Figure 1. Map of the seven AirMOSS validation sites.

Further details on the ways in which how soil moisture influences the model's predictions of carbon fluxes can be found in Appendix A.

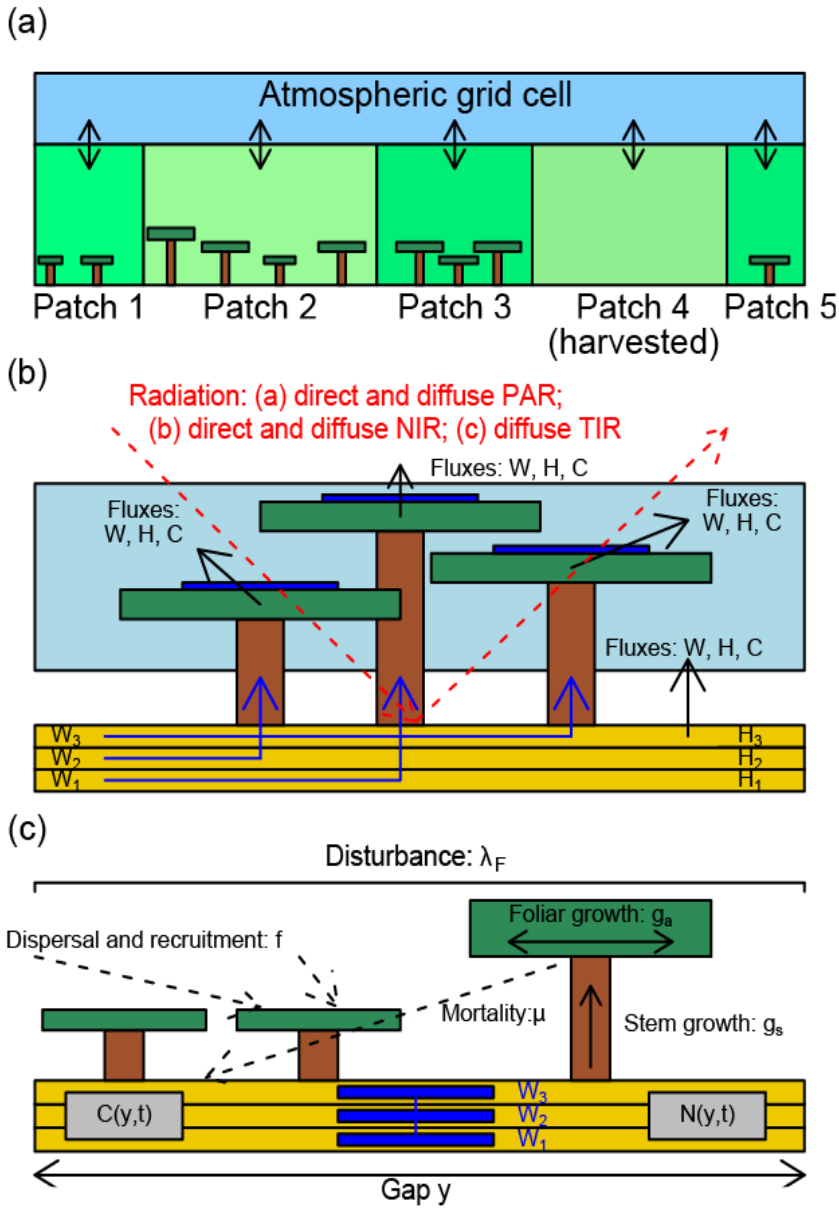


Figure 2. Schematic figure representing the ED2 model structure and process. (a) Each grid is subdivided into tiles with the relative area of each tile determined by the proportion of canopy-gap sized areas having a similar forest structure due to a similar disturbance history. (b) Within each tile, a multi-layer canopy model calculates the short term fluxes of water (W), internal energy (H), and carbon (C). (c) Illustration of the long term vegetation dynamics of the heterogeneous plant canopy resulting from the short term fluxes. The growth is represented in terms of stem and active tissue growth (g_s, g_a), the mortality as a rate μ, recruitment at rate of within and between gaps, and disturbance at rate λ_F (from Medvigy *et al*. 2009).

2.2 Simulation Protocol

A flowchart describing the inputs to the ED2 model is shown in Figure 3. The inputs include: meteorological forcing data, soil properties information, definition of initial ecosystem structure and composition, and prescribed phenology data regarding the timing of leaf flush and

leaf drop. The prognostic variables that will be produced as part of the L4A NEE algorithm are listed in Table 1.

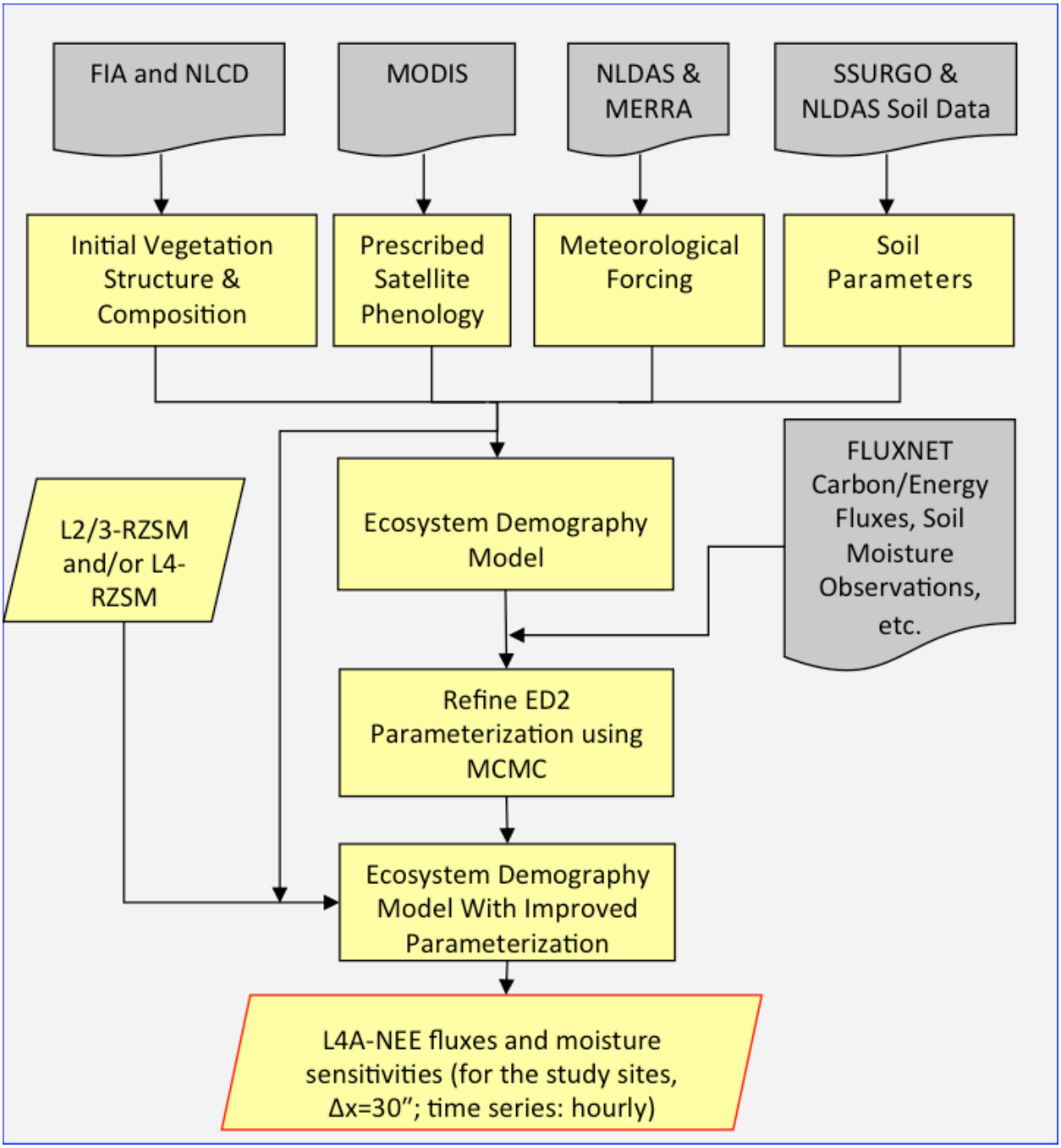


Figure 3. Flowchart describing the inputs, forcings, and initial conditions needed to produce the L4A-NEE products from the ED2 model and the processes to produce the L4A-NEE products.

Table 1. Ecosystem Demography V2 prognostic variables that form the L4 NEE data products.

Parameter	*Quantity*	*Units*
NEE	Net Ecosystem Exchange	KgC m^{-2} y^{-1}
GPP	Gross Primary Productivity	KgC m^{-2} y^{-1}
R_{eco}	Ecosystem Respiration	KgC m^{-2} y^{-1}
β_{θ}	Moisture sensitivity of NEE	KgC m^{-2} y^{-1} m^3H_2O m^{-3}
β_{θ}^{GPP}	Moisture sensitivity of GPP	KgC m^{-2} y^{-1} m^3H_2O m^{-3}
β_{θ}^{Reco}	Moisture sensitivity of Reco	KgC m^{-2} y^{-1} m^3H_2O m^{-3}
ET	Evapotranspiration	gH_2O m^{-2} y^{-1}
SM	Volumetric soil water content	%

2.3 Ancillary Data Requirements

In addition to the L4-RZSM product, the L4A-NEE implementation of the ED2 model requires the following ancillary data for each pixel within the simulation domain.

2.3.1 Soil Hydraulic Properties

The soil physical and hydraulic properties are obtained from the Soil Survey Geographic (SSURGO) soil data set in conjunction with the application of the ROSETTA, a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. The soil physical and hydraulic properties needed by the Ecosystem Demography model include fraction of sand, clay and silt, soil class, soil stratification (soil layers and depths), saturated soil water content (θ_s), residual soil water content (θ_r), and two parameters (α and n) for the van Genuchten equation for each soil layer. The lower soil boundary condition is also derived from the SSURGO data set. The van Genuchten equation is used to describe the water retention curve within the ED2 model. The SSURGO products are produced by the United States Department of Agriculture and cover only the United States. Similarly, the North American Land Data Assimilation System (NLDAS) Land Surface Parameters (LSP) soil texture data sets also cover only the United States. For the parts of North America outside the United States, including the AirMOSS test sites in Canada, Mexico, and Costa Rica, soil characteristics will be prescribed from the IGBP Data and Information System (IGBP-DIS) soil dataset.

2.3.2 Meteorological Forcing

Meteorological input forcing includes time series of over-canopy air temperature, downward shortwave and longwave radiation, precipitation, specific humidity, wind velocities, and surface air pressure. These data are obtained from the North American Land Data Assimilation Version 2 (NLDAS-2) forcing dataset for these sites and regions inside the Continental United States. For other regions, the meteorological forcing data are prescribed from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) product, which is a NASA reanalysis for the satellite era using a major new version of the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5). Since both NLDAS-2 (1/8°×1/8°) and MERRA

(2/3°×1/2°) have coarser spatial resolutions than 30 arc-second, spatial downscaled meteorological forcing will be directly obtained from the downscaled meteorological forcing of the L4-RZSM products.

2.3.3 Current Ecosystem State

Forest Structure and Composition

The initial plant community describing the size of individuals (z) of initial density ($n_0^{(i)}$), of plant functional type (i) described in Equation (3) is determined using forest inventory data from the US forest service's Forest Inventory and Analysis (FIA) program that has an extensive network of plots across the US in which individual stems are identified to species, their diameters measured, and plot characteristics such as soil carbon and plot ages since last disturbance are recorded. The FIA database is divided up into plots, where the spatial density between each plot can be between 10-20 km^2 depending on the land-use, and the state. The plots design usually contains four subplots, where three are located 120 ft (36.58 m) away from the central subplot at azimuths of 360°, 120°, and 240° (Figure 4). Information on all trees greater than 5 inches (12.7 cm) diameter at breast height are recorded at the subplot level with an area of 168.11 m^2, and information on all trees less than 5 inches are recorded at the microplot level with an area of 13.5 m^2 (see Figure 4). This information is necessary to determine the initial density of individuals (n_0 in Eq 3). Figure 5 shows the locations of the FIA plots within the Harvard Forest Grid, as well as the whole of New England.

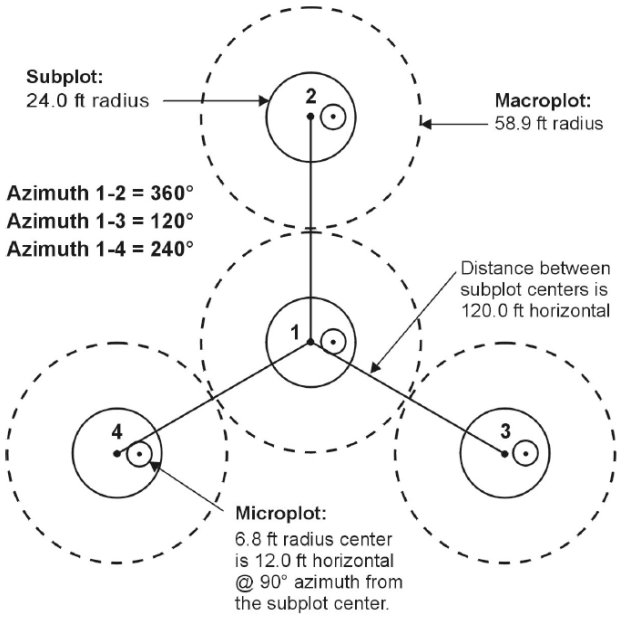


Figure 4. The FIA mapped plot design. Subplot 1 is the center of the cluster with subplots 2, 3, and 4 located 120 feet away at azimuths of 360°, 120°, and 240°, respectively (Woudenberg *et al.* 2010).

Forest ecosystem structure and composition is prescribed by assigning the trees reported in the FIA to the corresponding tree size class and plant functional types in ED2. For the initial forest structure, information extracted from the FIA includes the diameter at breast height (cm) and the plot sizes (see previous paragraph). For composition individual tree species are lumped into plant functional types. For example at Harvard Forest, white and red pine are classified as

early-successional conifers, hemlocks as late-successional conifers, birches as early-successional hardwoods, oak and red maple as mid-successional hardwoods, and beeches and sugar maple as late-successional hardwoods (Albani *et al*. 2006; Medvigy *et al*. 2009).

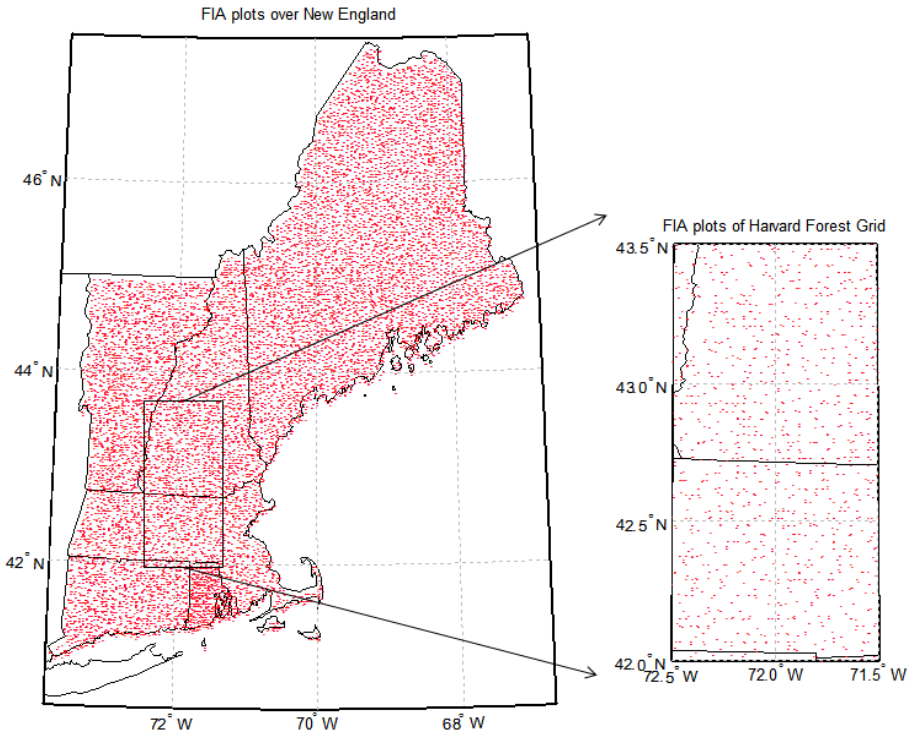


Figure 5. Spatial distribution of FIA plots across the New England region. Inset: FIA plots overlaid on the Harvard Forest L4A region.

An FIA based approach is used to create the initial forest condition needed to initialize the ED2 model to create sensitivity estimates of carbon fluxes to soil moisture. This is first determined for the radar swath discretized with a spatial resolution of 30 arc seconds, according to the L4A-NEE AirMOSS Science Product Grid Specifications. FIA plots are spatially extrapolated based on edaphic conditions such as soil texture (SSURGO), elevation (NED), topographic moisture index, and slope. All FIA plots located within the radar swath are first matched with the edaphic conditions, and then extrapolated into pixels with similar edaphic criteria. The forest structure at each pixel is then adjusted with the fraction of forest cover obtained using NLCD.

Land Cover Type

The FIA data contains structure and composition information for forested land cover, but the actual cover may be a mixture of forest and grass, grass, agriculture, or non-vegetation surfaces. For sites where grasses and herbaceous species are also present in the understory, e.g. at Tonzi Ranch, a grass layer is added where the land cover is defined as open or closed savanna or cropland determined from the NLCD (National Land Cover Database) and MODIS (Moderate Resolution Imaging Spectroradiometer)-IGBP (International Geosphere-Biosphere Programme) classification. Here, if there is *N* number of plots or patches defined from the FIA, then an extra patch is added with a representative grass layer, defined from the literature. For example Tonzi

Ranch has an understory grass layer with an LAI of 1, and heights of 15-50cm. Where the land cover class is defined just as grassland or pasture, then the forest overstory is removed, and a grass layer is considered only. Where the land cover class is non-vegetated, such as water, bare rock, urban, or fallow land, then all vegetated layers are removed, Pixels with no vegetation will be initialized similar to near-bare-ground simulations (i.e. first year since last disturbance Equation (3)) with seedlings of all plant functional types, and prescribed soil carbon information.

2.3.4 Soil carbon

The initial values of the soil organic carbon were extracted from FIA plot values of carbon in the litter pool and carbon in the soil organic material, with values converted from tons per acre to kg m^{-2}. Fast Soil Carbon is defined using the amount of carbon in the litter pool defined in the FIA data, and Structural Soil Carbon is defined from the soil organic material carbon pool defined in the FIA. Where data was unavailable for some sites, fast and structural carbon was set to the average of the swath, e.g. 1.8 kg m^{-2} and 7.75 kg m^{-2} respectively for Harvard Forest. The initial value of slow soil carbon was not available from the FIA, and was set from equilibrium runs at Harvard Forest. Other sites will also use simulated equilibrium values of slow soil carbon.

2.3.5 Phenology

The ED2 model is able to either predict or prescribe the phenological cycle or the percentage of active leaf area throughout the year. Calculated phenology follows the method described in Botta *et al*. (2000). Here, leaf flush occurs when the springtime degree days averaging above 5°C exceeds the number of days averaging below 5°C. Leaf drop occurs when either the day length and/or soil temperature drop below a threshold. As shown by Medvigy *et al*. (2009), the magnitude of errors in the Botta *et al*. (2000) predictive phenology scheme significantly impacts the model's predictions of carbon fluxes. In the L4NEE product, the phenology is prescribed from the MOD12Q2 product (Zhang *et al*. 2003) using a generalized version of the methodology developed by Medvigy et al (2009). This is based on determining points of inflection from the NBAR EVI (Nadir-BRDF-Adjusted Reflectance Enhanced Vegetation Index). This product gives up to 8 inflection values. These are two dates for green-up, maturity, senescence, and dormancy of leaves. This means that the product can give a full phenological cycle, two half cycles, one and a half cycles, two full cycles, and one full cycle with 2 half cycles.[1]

$$F(JD;\mu,s) = \frac{1}{1+e^{-(JD-\mu)/s}} \quad \text{where } s = \frac{\sqrt{3}*(\text{Inflection2}-\text{Inflection1})}{6\pi} \tag{4}$$

[1]Future work will also need to address the issue of dividing the phenological cycle between grasses and trees, especially for savannas like Tonzi Ranch. For areas like Tonzi Ranch, the grasses have active leaf areas from the early winter to late spring, while the deciduous trees have active leaf areas from early spring to the midsummer. Furthermore, in savanna regions with a significant amount of grass, or in agricultural regions with a presence of trees, there needs to be a consideration of separating the phenology from the grasses and that of the forest canopy. This will be done by spatially extrapolating only grass pixels and only forest pixels into mixed grass / forest pixels. This means that in a savanna region, its combined grass/tree phenology will be separated into grass only (determined from an adjacent grass pixel), and forest only (determined from an adjacent forest pixel).

The mean of the above function is the middle of the greenup-maturity or senescence-dormancy dates, and the slope is derived from the distance of the greenup-maturity or senescence-dormancy dates.

$$\text{Flush Mean } (\mu) = 0.5 * (\text{Greenup} + \text{Maturity})$$

$$\text{Flush Slope } (s) = \frac{\sqrt{3} * (\text{Maturity} - \text{Greenup})}{6\pi}$$

$$\text{Leaf Drop Mean } (\mu) = 0.5 * (\text{Senescence} + \text{Dormancy})$$

$$\text{Leaf Drop Slope } (s) = \frac{\sqrt{3} * (\text{Dormancy} - \text{Senescence})}{6\pi} \quad (5)$$

where field-based phenology measurements are also available the above equations can be scaled. For example, the greenup slope observed at Harvard Forest was shallower than calculated by the above function.[1]

2.4 Model Fitting Procedure

L4A-NEE is a computationally modeled product. Prior to conducting the model simulations, the ED2 model formulation was constrained using in a Bayesian Markov-Chain Monte Carlo (MCMC) framework. The procedure to calibrate and constrain the ED2 model is summarized in Figure 6.

The Markov Chain Monte Carlo (MCMC) method was utilized to optimize the parameterization of ED2 to better predict NEE and soil moisture. The general goal is to determine the probability distribution of a vector of model parameters $\boldsymbol{p}$, given a set of measurements $\boldsymbol{f}$, in this case NEE and soil moisture. Whether a given vector $\boldsymbol{p}$ is consistent with $\boldsymbol{f}$ is determined by running the model **M**, such that:

$$\boldsymbol{f}_{\boldsymbol{M}}(\boldsymbol{p}) = \mathbf{M}(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{s}) \quad (6)$$

$\boldsymbol{f}_{\boldsymbol{M}}$ is the vector of model predictions (NEE and soil moisture), and $\boldsymbol{c}$ and $\boldsymbol{s}$ vectors of environmental conditions and model state variables, respectively. $\boldsymbol{f}$, $\boldsymbol{f}_{\boldsymbol{M}}$, $\boldsymbol{c}$ and $\boldsymbol{s}$ contain values across both time and types of data, while p is assumed constant over time. For a process-based prognostic model ED2, **M** is nonlinear and too complicated to be expressed as a set of mathematical functions. This can be efficiently solved by direct sampling of the PDF in parameter space using Monte Carlo techniques according to Mosegaard (1998). The MCMC generally consists of a stochastic technique that generates a random set of points $\boldsymbol{p}^1,\ldots,\boldsymbol{p}^N$ in parameter space with a distribution that approximates any given function f(p) for large values of *N*. For a Bayesian inversion, this function is chosen as the posterior PDF of model parameters, given by

$$\boldsymbol{f}(\boldsymbol{p}) = vL(\boldsymbol{p})\rho(\boldsymbol{p}) \quad (7)$$

with a normalization constant, *v* (Mosegaard & Sambridge, 2002). *L(p)* is the likelihood function, which describes the mismatch between model-derived values and measurements in relation to measurement error, and $\rho(p)$ is the prior probability distribution of parameters.

Following Medvigy et al. (2009), the model parameters that will be included in the optimization procedure will be determined following the evaluation of the initial model formulation at the different sites. However, the list of model parameters that were optimized were those significantly influencing the ecosystem's water budget and its impact on plant photosynthesis, and soil respiration (see Appendices A1 and A2).

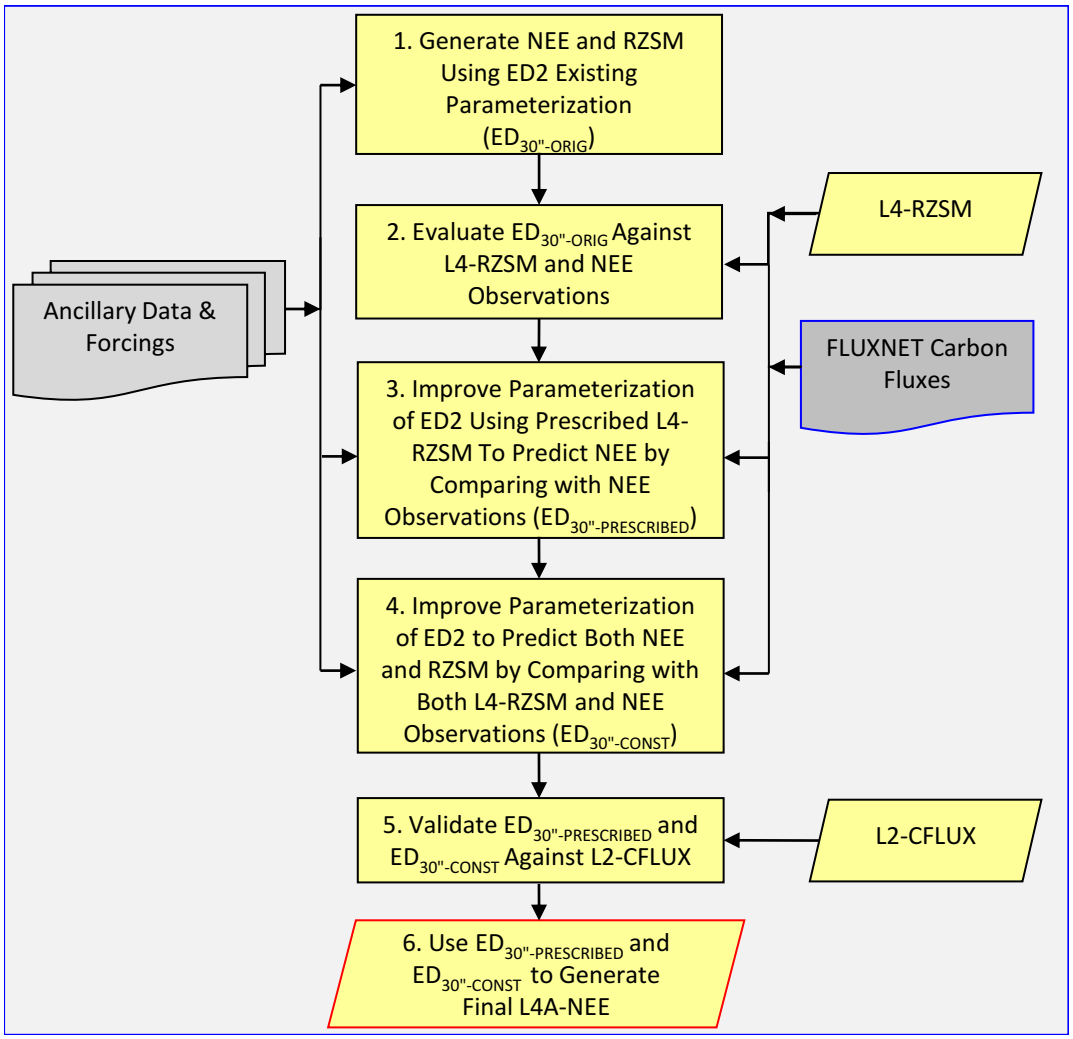


Figure 6. Flowchart that summarizes the processes to calibrate and constrain the ED2 model.

2.5 Regional Carbon Flux Estimates

An example of the carbon flux estimates produced for the seven validation sites calculated using the L4A-NEE algorithm is shown in Figure 7.

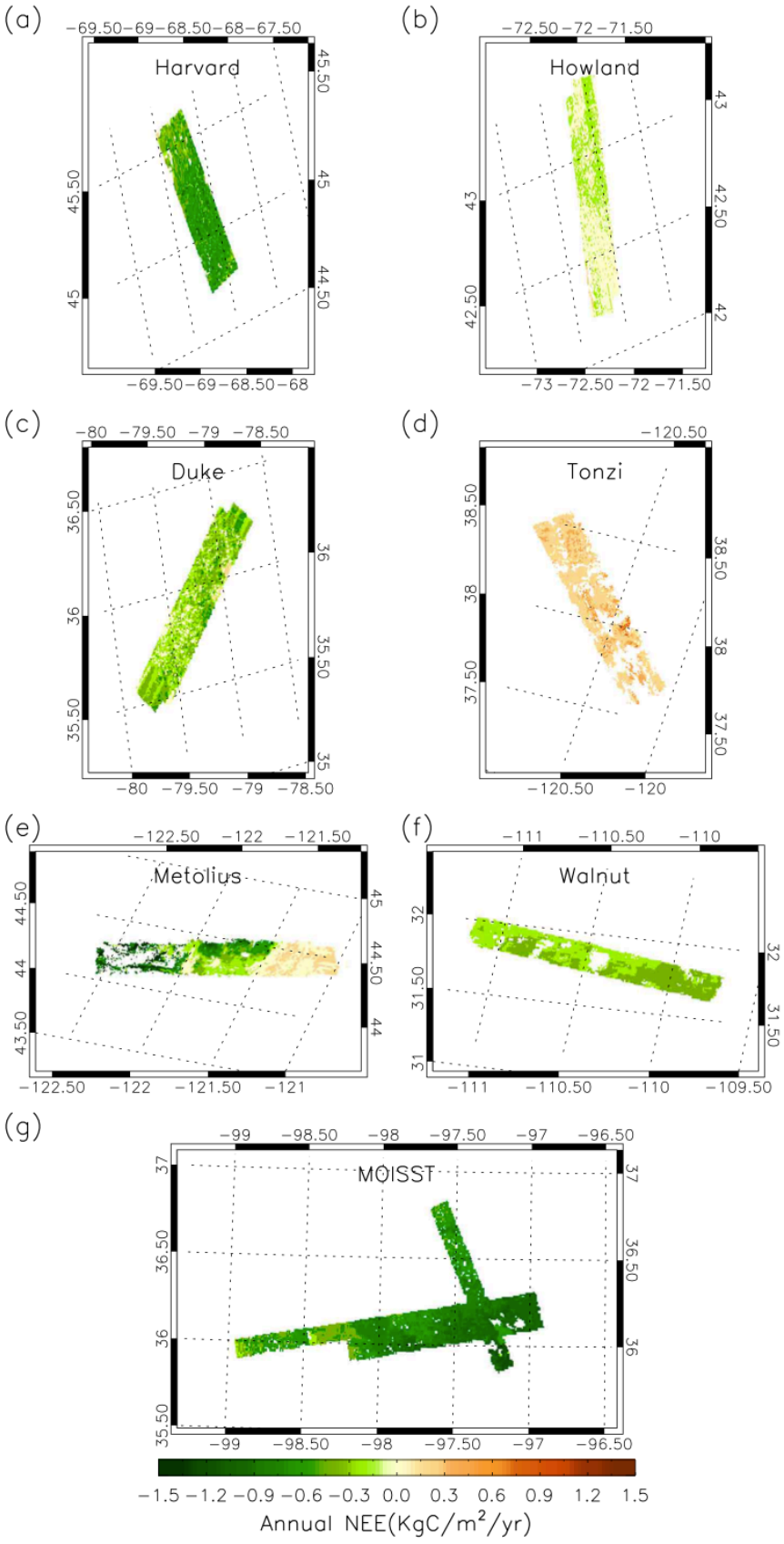


Figure 7. Spatial Patterns of Net Ecosystem Exchange (NEE) across the Seven AirMOSS evaluation sites calculated using the L4A-NEE algorithm.

2.5 Quantifying the Soil Moisture Sensitivities of L4A-NEE Carbon Fluxes

A key science objective of AirMOSS is a quantitative understanding of the impact of RZSM on ecosystem carbon fluxes. A key mathematical quantity that expresses this impact is the partial derivative of a given carbon flux with respect to soil moisture Adopting the notation of Friedlingstein et al. (2006), we denote this quantitative metric of a carbon flux's sensitivity to changes in soils moisture (θ) as β_θ. An estimate of β_θ for given carbon flux C (e.g. NEE, GPP or R_{eco}) can calculated from the model simulations by evaluating the magnitude the flux with and without the adjustment of the model's soil moisture field to the observations i.e.:

$$\beta_\theta = \left.\frac{\partial C_{\text{true}}}{\partial \theta}\right|_{\theta_{RZSM}} \approx \frac{C_{true}(\theta_{RZSM}) - C_{true}(\theta_{ED})}{\theta_{RZSM} - \theta_{ED}} \approx \frac{C_{ED}(\theta_{RZSM}) - C_{ED}(\theta_{ED})}{\theta_{RZSM} - \theta_{ED}}$$

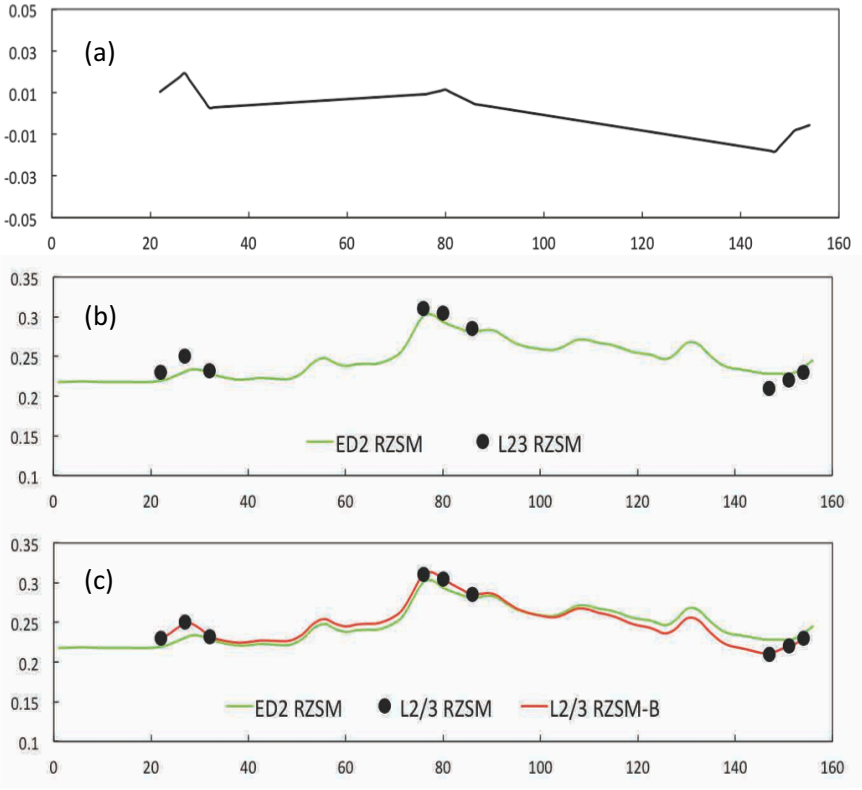
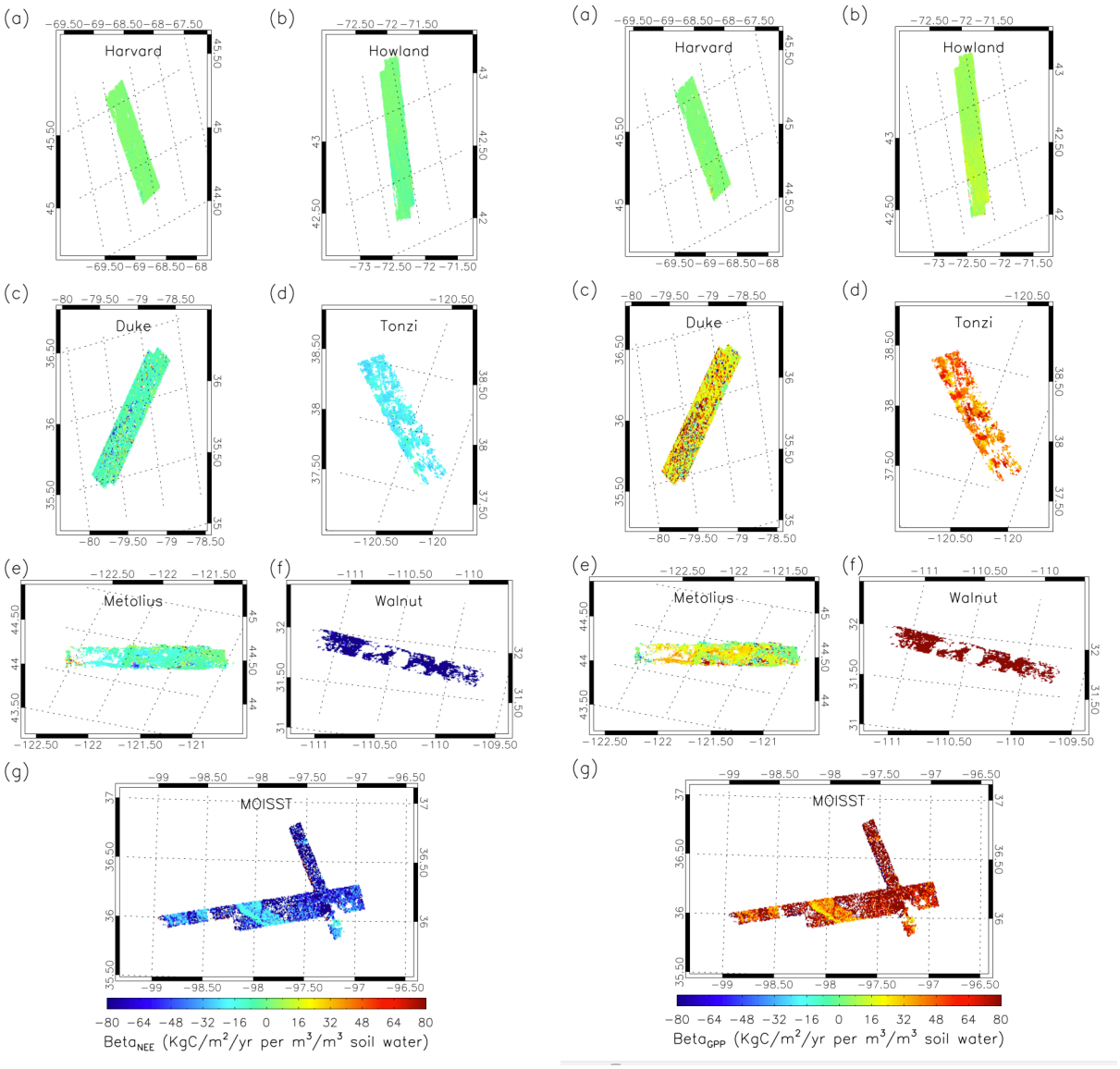


Figure 8. RZSM assimilation procedure. The bias in the ED2 model predictions of RZSM (black line in panel a) is calculated from the difference between the L2/3 RZSM measurements and the model's predictions (dots and green line in panel b respectively) with linear interpolation between L3/3 RZSM measurement dates. The bias correction is then applied to the model yielding a more accurate L2/3 RZSM-B soil moisture field (red line in panel (c)).

The above quantities were calculated by assimilating the L2/3 RZSM site-level snapshots of soil moisture in to the terrestrial biosphere model and using them to bias-correct the model's time-varying soil moisture field across the domain. Figure 8 illustrates the procedure for assimilation and bias-correcting the model's soil moisture using the L2/3 RZSM observations.

The sensitivities of *NEE*, *GPP* and R_{eco} were then estimated by computing the magnitude of the each carbon flux without assimilation and following assimilation (i.e. $C_{ED}(\theta_{RZSM})$ and $C_{ED}(\theta_{ED})$

respectively and dividing by the magnitude of the difference in soil moisture between the simulations (i.e. θ_{RZSM} - θ_{ED}). The annual average NEE and GPP carbon flux sensitivities for the seven AirMOSS validation sites (Figure 1) calculated using the L4A-NEE algorithm are shown in Figures 9 and 10.



Figures 9 and 10. Panels (a)-(g): Spatial patterns of the average soil-moisture sensitivities of Net Ecosystem Exchange (β_θ^{NEE}, Figure 9, left) and Gross Primary Productivity (β_θ^{GPP}, Figure 10, right) across the seven evaluation sites. Positive sensitivity values indicate that the carbon flux increases with increases in soil moisture.

3. L4B-NEE Algorithm

The L4B-NEE product consists of hourly and monthly estimates of carbon fluxes over the continental United States at 50km resolution and the accompanying estimates of their soil moisture sensitivities (Table 1). We used two approaches to scale the L4A-NEE products to the continental United States (CONUS). The first method was to implement the ED2 simulation algorithm developed for L4A-NEE a coarser-scale spatial resolution of 50km (Figure 11). Details regarding the edaphic/surface conditions and auxiliary data for these simulations are given in section 3.1 below. The second was a statistical extrapolation of L4A-NEE products onto the 50-km resolution continental-scale grid. Details of the statistical extrapolation procedure can be found in Appendix B.

3.1 Ancillary Data Requirements for L4B-NEE Simulations

The physical properties of soil and soil depths were obtained from NLDAS at 0.5^{o} resolution. The meteorological forcing was obtained from NLDAS-2 at 1/8 degree resolution. The initial forest structure and composition is obtained from the FIA database, where for each 50 km pixel, the nearest 3 FIA plots are extracted to create the ED2 forest state initialization files. We prescribed the phenology in ED2 using the MOD12Q2 products at 0.5^{o} resolution. Model simulations on the CONUS grid were carried from July 2010 to December 2014. When the ED2 model was run on the CONUS grid at the 0.5^{o} resolution, the closest meteorological forcing from the NLDAS-2 to the center of the 50 km pixel was used. Since FIA information was not available for the desert and cropland biomes (i.e. the biomes associated with the Tonzi Ranch, Walnut Gulch, and MOISST evaluation sites, see Figure 1), in these biomes the ED2 model was spun up for five years reasonable steady state was reached. For the remaining biomes, the fraction of forest cover was obtained from the NLCD was used to specify the forested fraction within each 50 km pixel. The spin-up procedure for the non-forested fraction was a six-year spin-up, while for forested fraction the structure and composition of the ecosystem was specified from the FIA plot dataset. All FIA plots in each 50km pixel are first extracted and matched with elevation, sand and clay fractions (NLDAS). The FIA plots in each 50km pixel were then subset, removing plots outside of the mean plus/minus the standard deviations of elevation, sand and clay fractions and plots with either uncharacteristically high or low basal areas.

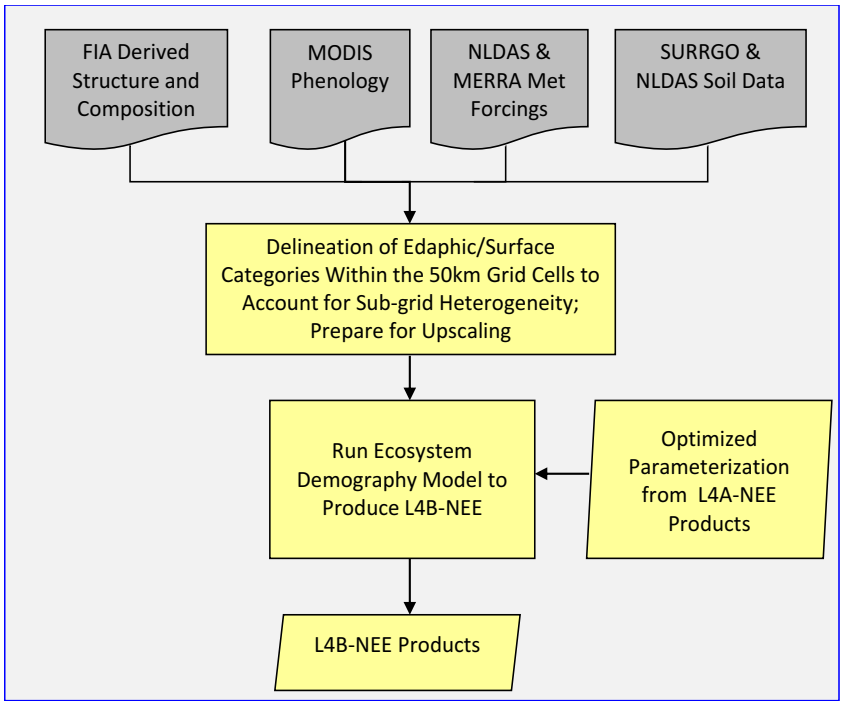


Figure 11. The L4B-NEE needs an aggregation of certain products to the 50km grid cell relative to certain edaphic/surface conditions. These edaphic / surface categories are classified by sub-grid heterogeneity of soil properties, ground slope, and topography.

3.2 Continental-Scale Carbon Fluxes

Examples of continental-scale carbon flux estimates calculated using the L4B-NEE algorithm is shown in Figure 12.

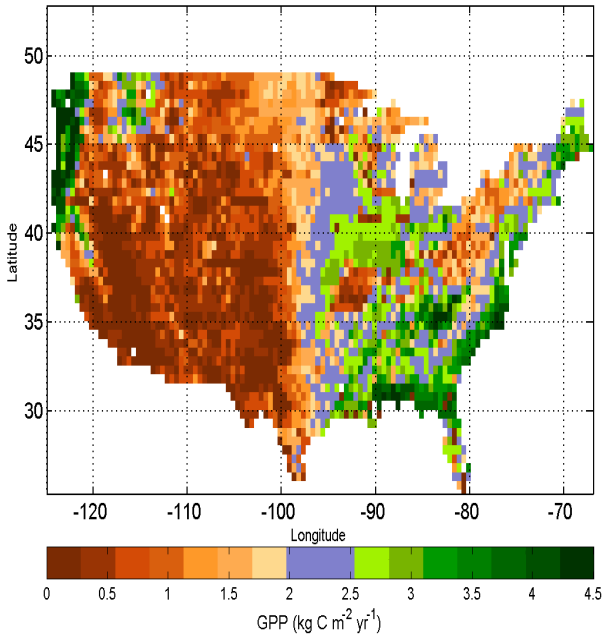


Figure 12. Patterns of annual Gross Primary Productivity (GPP) across the continental US calculated using the L4B-NEE algorithm.

3.3 Quantifying the Soil Moisture Sensitivities of L4B-NEE Carbon Fluxes

The soil moisture sensitivities of the continental-scale L4B-NEE carbon fluxes calculated from ED2 simulations were calculated using a similar procedure to the method applied to the regional scale L4A-NEE carbon fluxes as described in Section 2.5. The soil moisture sensitivity of the L4B-NEE carbon fluxes calculated by statistical extrapolation was calculated by spatial extrapolation of the L4A-NEE sensitivity estimates (e.g. Figures 9 and 10) using the same methodology used to extrapolate the fluxes as described in Appendix B. Figure 13 shows the estimated continental-scale sensitivity calculated by the statistical extrapolation method.

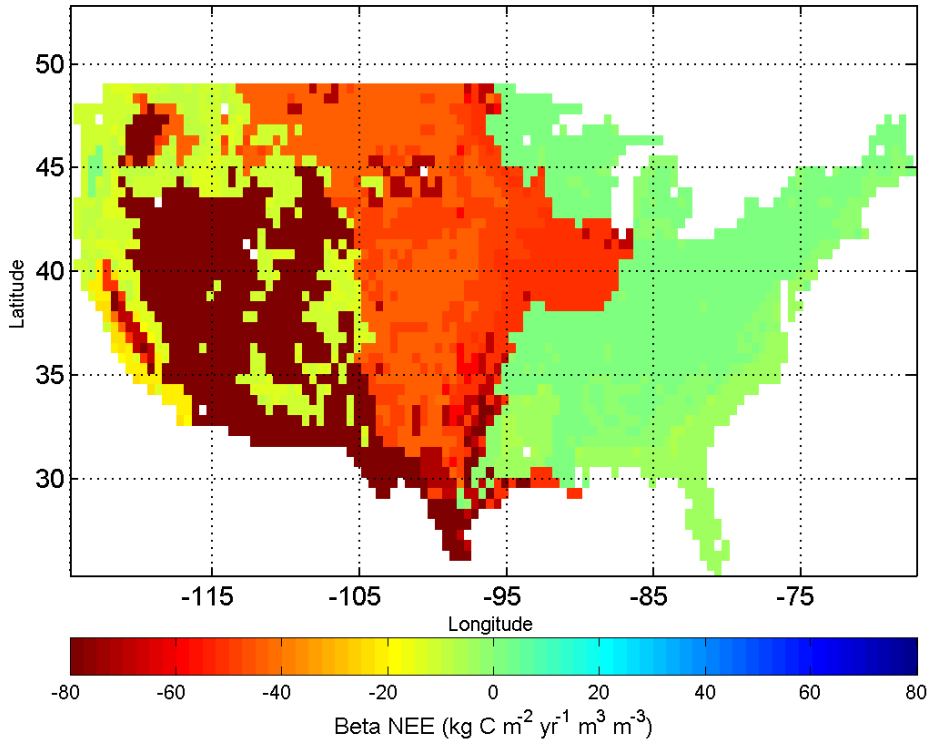


Figure 13. Patterns of NEE moisture flux sensitivity across the continental US based on statistical extrapolation of L4B-NEE. in soil moisture.

4. Model Uncertainty and Validation

4.1 Model Validation

The AirMOSS L4A-NEE products have been validated against:

1) The measured carbon fluxes and the measured latent heat fluxes from the eddy covariance flux towers at each site.
2) Soil moisture from the RZSM measurements.

A comparison of the AirMOSS L4A-NEE products against the L2-CFLUX estimates of NEE produced by Shepson et al. is also being conducted.

4.2 Quantifying Sources of Error

There are several qualitatively different sources of error in the L4A/B-NEE products. These include 1) uncertainty in initial ecosystem structure and composition, 2) error in phenology, 3) errors in the meteorological forcing, 4) errors in the L4 RZSM, and 5) model process error. The contributions of these different sources of error to the NEE estimates will be estimated in the following way:

1) The accuracy of the forest structure and composition defined from the FIA database has been evaluated through direct comparisons against ground measurements in the proximity of the flux towers in terms of tree size distributions of the different plant functional types. Also, where available, the structure will be assessed using remote sensing techniques such as airborne lidar and radar. Second, uncertainty will arise when determining forest structure and composition over a land cover type that is classified as a grassland or savanna. Standard errors determined from FIA plant functional type classifications will also be considered. These uncertainties will then be propagated to generate a distribution of canopy structure and composition estimates, which will then be used to initialize a corresponding series of ED2 model simulations.
2) The accuracy of the phenology data determined from MODIS has been assessed by comparing the derived cycles with observations where available.
3) The impact of error in the meteorological forcing has been estimated by comparing the NLDAS/MERRA based simulations with simulations conducted using observed meteorology.
4) The errors remaining after (1)-(3) is the process error – errors associated with the equations and parameters within the model. The approach to reducing the process error is described in the model re-calibration described section 4.3.

4.3 Using L4-RZSM to improve the ED2 Model Predictions of Carbon Fluxes

Evaluation of the ED2 soil moisture predictions against the L2/3 RZSM data product indicates a significant dry bias in the current model parameterization (Figure 14). The L4-RZSM product is

being used to re-constrain the parameters of the ED2 model formulation using the Bayesian Markov-Chain Monte Carlo (MCMC) framework described in Section 2.5.

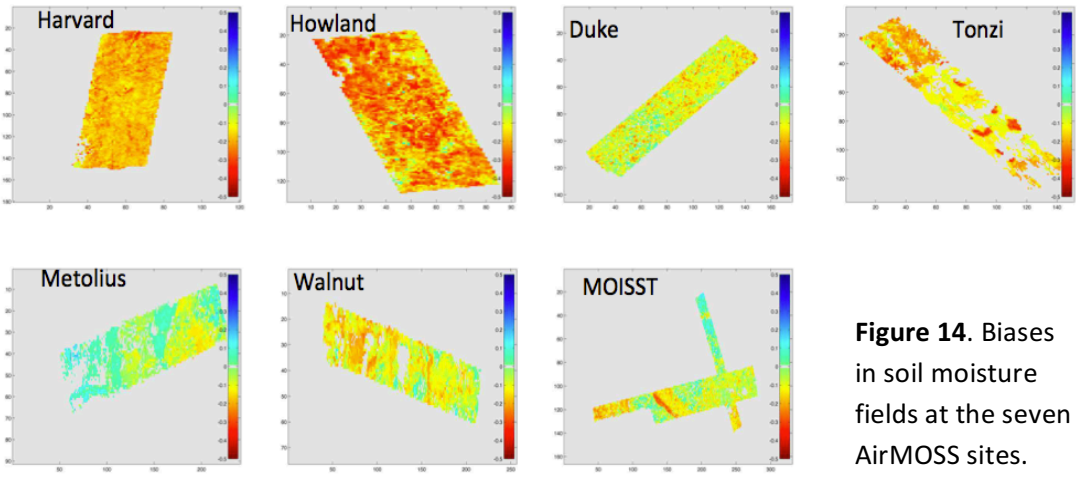


Figure 14. Biases in soil moisture fields at the seven AirMOSS sites.

5. List of Acronyms and Abbreviations

AirMOSS	Airborne Microwave Observatory of Subcanopy and Subsurface
ED2	Ecosystem Demography Model 2
DIS	Data and Information System
ET	Evapotranspiration
EVI	Enhanced Vegetation Index
FIA	Forest Inventory and Analysis
GEOS-5	Goddard Earth Observing System Data Assimilation System Version 5
GPP	Gross Primary Productivity
IGBP	International Geosphere-Biosphere Programme
LEAF	Land Ecosystem-Atmosphere Feedback model
MCMC	Markov-Chain Monte Carlo
MERRA	Modern-Era Retrospective analysis for Research and Applications
MESMA	Multiple End-member Spectral Mixture Analysis
MODIS	Moderate Resolution Imaging Spectroradiometer
NBAR	Nadir-BRDF-Adjusted Reflectance

NEE	Net Ecosystem Exchange of CO_2
NIR	Near Infrared Radiation
NLCD	National Land Cover Database
NLDAS	North American Land Data Assimilation System
NPP	Net Primary Productivity
PAR	Photosynthetic Active Radiation
PDE	Partial Differential Equation
Ra	Autotrophic Respiration
Reco	Ecosystem Respiration
Rh	Heterotrophic Respiration
RMSE	Root Mean Squared Error
RZSM	Root-Zone Soil Moisture
SAS	Size- and Age-Structured
SM	Soil Moisture
SSURGO	Soil Survey Geographic Database

References

Albani, M, D. Medvigy, G.C. Hurtt and P.R. Moorcroft (2006). The contributions of land-use change, CO_2 fertilization, and climate variability to the Eastern US carbon sink. *Global Change Biology,* **12**(12):2370--2390.

Antonarakis, A.S., S.S. Saatchi, , R.L. Chazdon and P.R. Moorcroft (2011). Using Lidar and Radar measurements to constrain predictions of forest ecosystem structure and function. *Ecological Applications*, **21**, 1120-1137.

Ball, J.T., I.E. Woodrow and J.A. Berry (1986), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis Research*, edited by I. Biggins, pp. 221–224, Martinus Nijhoff, Netherlands.

Beven, K.J., and M.J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, **24**, 43-69.

Botta, A., N. Viovy, P. Ciais, P. Friedlingstein, and P. Monfray (2000). A global prognostic scheme of leaf onset using satellite data. *Global Change Biology*, **6**, 709-725.

Farquhar, G.D. and T.D. Sharkey (1982), Stomatal conductance and photosynthesis, *Annual Review of Plant Physiology*, **33**, 317– 345.

Farquhar, G. D., S. von Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, *Planta*, **149**, 78–90.

Friedlingstein, P., P. Cox, R. Betts, L. Bopp, W. Von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H.D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A.J. Weaver, C. Yoshikawa, and N. Zeng (2006). Climate–Carbon Cycle Feedback Analysis: Results from the C^4MIP Model Intercomparison, *Jounral of Climate,* **19,** 3337-3353.

Hurtt, G.C., P.R. Moorcroft, S.W Pacala, and S.A. Levin (1998). Terrestrial models and global change: challenges for the future, *Global Change Biology*, **4**(5):581-590.

Hurtt, G.C., R. Dubayah, J. Drake, P.R. Moorcroft and M. Fearon (2004). Beyond potential vegetation: Combining Lidar remote sensing and a height-structured ecosystem model for improved carbon stock and flux estimates, *Ecological Applications*, **14**(3):873-883.

Hurtt, G. C. et al. (2011) Harmonization of Land-Use Scenarios for the Period 1500-2100: 600 Years of Global Gridded Annual Land-Use Transitions, Wood Harvest, and Resulting Secondary Lands, Climatic Change, DOI: 10.1007/s10584-011-0153-2.

Leuning, R. (1995), A critical appraisal of a combined stomatal-photosynthesis model for C3 plants, *Plant Cell Environ*., **18**, 339– 355.

Medvigy, D, SC Wofsy, J.W. Munger, D.Y. Hollinger and P.R. Moorcroft (2009). Mechanistic scaling of ecosystem function and dynamics in space and time: the Ecosystem Demography model version 2, *Journal of Geophysical Research – Biogeosciences*, **114**(G1).

Monteith, J.L. (1973), *Principles of Environmental Physics*, Edward Arnold, London.

Moorcroft, PR, G.C. Hurtt, and S.W. Pacala (2001). A method for scaling vegetation dynamics: the Ecosystem Demography model (ED), *Ecological Monographs*. **71**(4):557-586.

Mosegaard, K. (1998). Resolution Analysis of General Inverse Problems through Inverse Monte Carlo Sampling, *Inverse Problems*, **14**, 405-426.

Mosegaard, K. and M. Sambridge (2002). Monte Carlo analysis of inverse problems. *Inverse Problems*, **18**, R29-R54.

Parton, W., J. Stewart, and C. Cole. (1988). Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry*, **5**,109-131.

Shugart, H.H., S. Saatchi, and F.G. Hall (2010). Importance of structure and its measurement in quantifying function of forest ecosystems, *J. Geophys. Res*., **115**, G00E13.

von Caemmerer, S. and G.D. Farquhar (1981), Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves, *Planta*, **153**, 376– 387.

van Genuchten, M.Th. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, **44**(5), 892–898. doi:10.2136/sssaj1980.03615995004400050002x.

Walko, R.L., L.E. Band, J. Baron, T.G.F. Kittel, R. Lammers, T.J. Lee et al. (2000). Coupled atmosphere–biophysics–hydrology models for environmental modeling. Journal of Applied Meteorology, **39(6)**, 931–944.

Walko, R.L. and C.J. Tremback (2005). ATMET Technical Note 1, Modifications for the Transition from LEAF-2 to LEAF-3, ATMET, LLC, Boulder, Colorado 80308-2195, http://www.atmet.com/html/docs/rams/, 2005.

Woudenberg, S.W., B.L. Conkling, B.M. O'Connell, E.B. LaPoint, J.A. Turner and K.L. Waddell (2010). The Forest Inventory and Analysis Database: Database Description and Users Manual Version 4.0 for Phase 2. General Technical Report RMRS-GTR-245. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, 339 pp.

Zhang, X., M.A. Friedl, C.B. Schaaf, A.H. Strahler, J.C.F. Hodges, F. Gao et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, **84**, 471–475.

Appendix A: Impacts of Soil Moisture on Carbon Fluxes

The ecosystem exchange (NEE) of the ecosystem is defined as the heterotrophic respiration, r_h, minus the integral of the per plant net primary productivity (*NPP*) integrated over all plants within the grid cell:

$$Canopy\ NEE = \int_a \left(r_h(a) - \int_z \left(NPP^{(i)}(z,a) \right) * n^i(z,a)\, dz \right) p(a)\, da \tag{8}$$

where n^i and p are defined as plant density of plant functional type i and distribution of gap ages, respectively.

A1: Impacts of Soil Moisture on Plant Photosynthesis

$NPP^{(i)}(z,a)$ and $r_h^{(i)}(z,a)$ are calculated using the model of leaf-level carbon assimilation and water fluxes developed by Farquhar, Ball, Berry, and others (Farquhar et al., 1980; von Caemmerer and Farquhar, 1981; Farquhar and Sharkey, 1982; Ball et al., 1986). The processes described here are at the per plant level.

The availability of water in the soil directly affects the instantaneous rates of net photosynthesis (A_{net}) and evapotranspiration (Ψ_{net}). Photosynthesis and evapotranspiration are taken to be linear combinations of their rates under conditions of open (A_o Ψ_o) and closed (A_c Ψ_c) stomata, the weighting of being determined by a plant's water availability relative to its overall water demand:

$$A_{net} = f_{o,w} A_o + \left(1 - f_{o,w}\right) A_c \tag{9}$$

$$\Psi_{net} = f_{o,w} \Psi_o + \left(1 - f_{o,w}\right) \Psi_c \tag{10}$$

where the open stomata weighting is given as a function of the plants water demand and the water availability:

$$f_{o,w} = \frac{1}{1 + \frac{Demand}{Supply}} \quad , \quad Demand = \Psi_o\, SLA \cdot B_{leaf} \ , \ Supply = K_w W_{avail,tot} B_{root} \tag{11}$$

where SLA is the plant's specific leaf area, B_{leaf} is the plant's leaf biomass, $W_{avail,tot}$ is the total amount of water accessible to the plant, given its rooting depth B_{root} is the plant's root biomass, and K_w is the conductivity of water and is a constant. The leaf-level demand of photosynthesis is given by:

$$A_o = min(J_e, J_c) - \gamma V_m(T_v) \quad \text{for open stomata}$$

$$A_o = -\gamma V_m(T_v) \quad \text{for closed stomata} \tag{12}$$

where γV_m represents leaf respiration at the plant level, $V_m(T_v)$ is the maximum capacity of Rubisco to perform the carboxylase function at a given temperature, and γ is proportionality constant. In the fall, photosynthesis is ramped down according to the available active leaf area, or phenology. The equation currently used in ED2 to determine the active leaf area from MODIS is described in a subsequent section. J_e is defined as the leaf-level light limited photosynthesis term, J_c is the Rubisco-limited CO_2 demand. The light limited and Rubisco limited rate of photosynthesis is given by:

$$J_e = \alpha PAR_v \frac{C_{inter}-\Gamma}{C_{inter}+2\Gamma} \quad \text{and} \quad J_c = \frac{V_m(T_v)\,(\mathrm{C_{inter}}-\Gamma)}{C_{inter}+K_1(1+K_2)} \tag{13}$$

where α is the quantum efficiency, PAR_v is the PAR absorbed by the vegetation layer, C_{inter} is the intercellular CO_2 concentration, and Γ is the compensation point for gross photosynthesis directly related to the temperature, and $K_1$and $K_2$are the Michaelis-Menten coefficients for CO_2 and O_2 respectively. The intercellular boundary layer is directly related to the boundary mixing ratios for H_20 and CO_2 following Monteith (1973) and Leuning *et al*. (1995):

$$C_{inter} = C_S - \frac{A_o}{1.6 g_{sw}} \quad \text{and} \quad e_L = e_S + \frac{\Psi_o}{g_{sw}} \tag{14}$$

where g_{sw} is the stomatal conductance for water dependent on whether the stomata are open or closed, C_S and e_S are the CO_2 and H_2O concentrations within the leaf boundary layer both related to the boundary layer conductance of water from free and forced convection.

The plant level net primary productivity is then calculated as:

$$NPP^{(i)}(z,a) = A_{NET}\, SLA \;\; B_{leaf} \;\; - \beta_{root} B_{root} \;\; - \; \beta_{storage}\, B_{storage}$$

where B_{root} and $B_{storage}$ are the size of the plant's fine-root and storage carbon pools and β_{root} and $\beta_{storage}$ are their respiration rates per unit mass.

A2: Impacts of Moisture on Soil Decomposition

The decomposition fluxes are based on the model of Parton *et al*. (1988). The process described here is at the per area level. For fast pool decomposition:

$$F_{f,decomp} = AK_2C_f \tag{15}$$

where K_2 is a rate constant defining the fast carbon pool (C_f) and A is the product of two functions $f(x_w)$ and $f(T)$ whose values vary between 0 and 1, that respectively account for the temperature and moisture dependence of heterotrophic respiration. The moisture dependency f is given as:

$$f(x_W) = \begin{cases} exp\left[\left(x_W - W_{opt}\right)w_1\right] & x_W < W_{opt} \\ exp\left[\left(W_{opt} - x_W\right)w_2\right] & x_W > W_{opt} \end{cases} \tag{16}$$

and the temperature dependence of A is given as:

$$f(T) = exp\left[\frac{logQ_{10}}{10}(T - 318.15)\right] \quad (17)$$

where x_W is the soil moisture, W_{opt} is the optimum soil moisture for decomposition, w_1 and w_2 are shape parameters for the moisture dependence of heterotrophic respiration, Q_{10} determines the response of decomposition to soil temperature, and T is the temperature in Kelvin. The structural pool decomposition flux is defined as:

$$F_{st,decomp} = AL_cK_1C_{s,struct}f_{std} \quad (18)$$

Here, K_1 is a rate constant setting the residence time of the structural pool $C_{s,struct}$, and L_c is the amount of lignin transferred out of the structural pool. Finally, the slow pool has a decomposition rate of:

$$F_{sl,decomp} = AK_3C_{s,slow} \quad (19)$$

The net decomposition rate or total heterotrophic respiration is given by:

$$r_h = F_{f,decomp} + F_{st,decomp} + F_{sl,decomp} \quad (20)$$

A3: Soil Water and Internal Energy Fluxes

The heat and moisture transfers among soil and snow layers in the ED2 are calculated by a biophysical scheme adopted from the Land Ecosystem-Atmosphere Feedback (LEAF-3) model (Walko and Tremback, 2005). Within each grid cell, the meteorology, soil physics, and topography are homogenous. However, the endogenous heterogeneity is generated within the subgrid scale due to the formations of different patches/gaps with different sizes, ages, and plant structure and composition, which are simulated by ED2 via the SAS approximation (Figure 1). The soil exists as pre-set number of vertically stacked layers having a constant and pre-specified thickness within each grid cell. The soil stratification, thicknesses and other physical properties are provided by the prescribed soil data described in section 2.2.4. Soil layers are labeled by an index *k*, where *k=1* and $k=N_g$ represent the deepest layer and the top layer, respectively. The vertical extent of each layer is given by D_k (i.e. soil thickness). Each soil layer is characterized by its volumetric water content (θ_k) and internal energy (Q_k). Due to the subgrid heterogeneity, internal energy and volumetric water content for each soil layer are prognosed on the patch level (Figure 1A). From Q_k the soil temperature (T_k) and the liquid water fraction $(f_{l,k})$ can be diagnosed. Measuring T_k in ^oC, the internal energy is

$$Q_k = \theta_k(1 - f_{l,k})C_iT_k + \theta_k f_{l,k}(C_lT_k + L_{il}) + C_sm_sT_k \quad (21)$$

where C_i is the specific heat of ice, C_l is the specific heat of liquid water, C_s is the specific heat of the dry soil layer, L_{il} is the latent heat of fusion and m_s is the mass density of the dry soil particles. The soil is entirely frozen when $Q_k < 0$, the soil is at $0\ ^oC$ and is partially frozen when $0 < Q_k < \theta_k L_{il}$, and the soil water is entirely liquid when $Q_k > \theta_k L_{il}$.

The θ_k of each layer is controlled by the flux of water between layers $(F_{w,gg})$, evaporation from the top soil layer $(F_{w,gc})$, percolation into the top soil layer $(F_{w,sg})$ and removal of water by plants through transpiration $(F_{w,gk,transp})$. Thus, specifying the density of water as ρ_w, the budget equation for θ_k is

$$\rho_w \frac{dW_{gk}}{dt} = \underbrace{-\frac{\partial F_{w,gg}}{\partial z}}_{diffusion} - \underbrace{\frac{1}{D_k} F_{w,g_k,transp}}_{transpiration} - \underbrace{\delta_{N_s,0} \frac{1}{D_k} F_{w,gc}}_{evaporation} - \underbrace{(1 - \delta_{N_s,0}) \frac{1}{D_k} F_{w,gc}}_{\text{percolation}} \quad (22)$$

where N_s is the number of temporary surface water layers, and $\delta_{i,j} = 0$ *if* $i \neq$ j and *1* otherwise.

The water fluxes between soil layers *($F_{w,gg}$)* are calculated according to the formulation of LEAF-3 (Walko and Tremback, 2005). They are given by

$$F_{w,gg} = -\rho_w K_\eta \frac{\partial(\psi + z)}{\partial z} \quad (23)$$

K_η is the hydraulic conductivity defined by

$$K_\eta = K_{\eta,0} \left(\frac{\theta}{\theta_s}\right)^{2b+3} \quad (24)$$

where θ and θ_s are actual and saturated soil water contents, respectively; $K_{\eta,0}$ is the hydraulic conductivity at saturation; b is a parameter; ψ is the water potential and can be calculated from the van-Genuchten equation (van Genuchten, 1980):

$$\theta = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha|\psi|)^n]^{1-1/n}} \quad (25)$$

where θ_r is residual soil water content; α and n are two parameters for the van Genuchten equation. All above parameters for soil physics and hydraulics are either obtained from the provided soil data (section 2.2.4) or derived from soil textural class provided by the soil data (section 2.2.4).

The characteristics of drainage from the bottom soil layer depend on the lower boundary condition for each grid cell provided by the prescribed soil data. There are only two boundary conditions: bedrock and non-bedrock. If the boundary is bedrock, there is no water flux through the soil lower boundary. If it is non-bedrock, water flux through the soil lower boundary is the gravitational flow.

Changes in layer energy Q_k are associated with all of these transfers as well as direct diffusion of heat between layers *($F_{h,gg}$),* short wave *(SW_k)* radiative fluxes, long wave *(LW_k)* radiative fluxes, and sensible heat exchange with the canopy air space *($F_{h,gc}$).* Thus the budget equation is

$$D_k \frac{dQ_{gk}}{dt} = \underbrace{-D_k \frac{\partial F_{h,gg}}{\partial t}}_{diffusion} - \underbrace{\delta_{N_s,0} F_{q,gc} + (1 - \delta_{N_s,0}) F_{q,gs}}_{sensible} - \underbrace{\delta_{N_s,0} \mathrm{LW_k}}_{\text{long wave}} - \underbrace{\mathrm{SW_k}}_{\text{short wave}} \quad (26)$$

Appendix B: Statistical Upscaling Methodology

At every AirMOSS site (~2500 km^2), the L4A-NEE product was available at a spatial resolution of 1 km. We extrapolated the L4A-NEE product from the site-level onto the CONUS grid at a spatial resolution of 50 km utilizing information about the soil texture (sand fraction (S_f; unitless), clay fraction (C_f; unitless) and silt fraction where S_f, C_f and silt fraction sum to one) and

elevation (E; m) of each pixel. The same type of information was available for each AirMOSS biome on the CONUS.

To segregate the L4A-NEE product, we developed a number of rules; that involved constructing classes of soil texture and elevation at the AirMOSS site and across the AirMOSS biomes across the continent (Figure 1). For example, at the Harvard site, we developed the following set of rules using the three variables; S_f, C_f and E:

$S_f \leq 0.5$ and $C_f \leq 0.5$ and $E \leq 300$	Class 1
$S_f \leq 0.5$ and $C_f \leq 0.5$ and $300 < E \leq 450$	Class 2
$S_f \leq 0.5$ and $C_f \leq 0.5$ and $E > 450$	Class 3
$S_f \leq 0.5$ and $C_f \geq 0.5$ and $E \leq 300$	Class 4
$S_f \leq 0.5$ and $C_f \geq 0.5$ and $300 < E \leq 450$	Class 5
$S_f \leq 0.5$ and $C_f \geq 0.5$ and $E > 450$	Class 6
$S_f > 0.5$ and $C_f < 0.5$ and $E \leq 300$	Class 7
$S_f > 0.5$ and $C_f < 0.5$ and $300 < E \leq 450$	Class 8
$S_f > 0.5$ and $C_f < 0.5$ and $E > 450$	Class 9

In the above classes, the numerical values were considered as critical values, whose value was determined by the distribution of that particular variable. Based on the above rule, we grouped L4A-NEE results into the above classes. Since we had soil texture and elevation information on the CONUS, we found similar classes on the CONUS. Then we dumped the median value of L4A-NEE from each of these classes onto the pixels on CONUS whose classes matched. Similar rules were developed and applied for the other AirMOSS sites.