



## **Soil Moisture Active Passive (SMAP)**

# **Algorithm Theoretical Basis Document SMAP Level 4 Carbon Data Product (L4\_C)**

Revision A  
December 9, 2014

John S. Kimball, Lucas A. Jones, Joseph P. Glassy  
*Numerical Terradynamic Simulation Group (NTSG)*  
*The University of Montana*  
*Missoula, MT*

Rolf Reichle  
*NASA Goddard Space Flight Center*  
*Greenbelt, MD*



Jet Propulsion Laboratory  
California Institute of Technology

© 2014. All rights reserved.

(Page intentionally left blank.)

The SMAP Algorithm Theoretical Basis Documents (ATBDs) provide the physical and mathematical descriptions of algorithms used in the generation of SMAP science data products. The ATBDs include descriptions of variance and uncertainty estimates and considerations of calibration and validation, exception control and diagnostics. Internal and external data flows are also described.

The SMAP ATBDs were reviewed by a NASA Headquarters review panel in January 2012 with initial public release later in 2012. The current version is Revision A. The ATBDs may undergo additional version updates after SMAP launch.

## TABLE OF CONTENTS

COMMON ACRONYMS AND ABBREVIATIONS .....	6
REFERENCE DOCUMENTS .....	8
1. INTRODUCTION .....	10
2. OVERVIEW AND BACKGROUND .....	11
2.1 The Soil Moisture Active Passive (SMAP) Mission .....	11
2.1.1 Background and Science Objectives .....	11
2.1.2 Measurement Approach .....	12
2.2 L4_C Product/Algorithm Objectives .....	15
2.3 Historical Perspective .....	16
2.4 Data Product Characteristics .....	18
2.4.1 Instrument/Calibration Aspects (affecting product) .....	18
2.4.2 Product Scope and Format .....	18
3. RETRIEVAL ALGORITHM .....	22
3.1 Theoretical Description .....	22
3.1.1 Mathematical Description of the Algorithms .....	22
3.1.2 Algorithm Baseline .....	28
3.1.3 Algorithm Options .....	32
3.1.3.1 GPP Calculation Using Ancillary VI Inputs .....	32
3.1.3.2 Disturbance and Recovery Effects .....	33
3.2 Ancillary Data Requirements .....	34
3.2.1 Impacts from Algorithm Options .....	37
3.3 Variance and Uncertainty Estimates .....	38
3.3.1 Impacts from Algorithm Options .....	43
3.4 Numerical Computation Considerations .....	45
3.5 Programming/Procedural Considerations .....	45
3.5.1 Impacts from Algorithm Options .....	46
3.6 Ancillary Data Availability/Continuity .....	47
3.6.1 Impacts from Algorithm Options .....	49
3.7 Calibration and Validation .....	49

3.8 Quality Control and Diagnostics.....	54
3.9 Exception Handling.....	55
3.10 Interface Assumptions.....	56
3.11 Test Procedures.....	56
3.12 Algorithm Baseline Selection.....	58
4. CONSTRAINTS, LIMITATIONS AND ASSUMPTIONS.....	58
5. REFERENCES.....	63
6. APPENDIX.....	75

## COMMON ACRONYMS AND ABBREVIATIONS

AirMOSS	Airborne Microwave Observatory of Subcanopy and Subsurface
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
ATBD	Algorithm Theoretical Basis Document
AVHRR	Advanced Very High Resolution Radiometer
BA	Burned Area
BPLUT	Biome Properties Look-Up Table
C5	MODIS Collection 5 reprocessing
CARVE	Carbon in Arctic Reservoirs Vulnerability Experiment
CUE	Carbon Use Efficiency
DAAC	Distributed Active Archive Center
DEM	Digital Elevation Model
EASE-Grid	Equal-Area Scalable Earth Grid
EOS	Earth Observing System
ESA	European Space Agency
ESDSWG	Earth Science Data System Working Group
EVI	Enhanced Vegetation Index
FGDC	Federal Geographic Data Committee
FLUXNET	Global network of tower CO <sub>2</sub> eddy covariance measurement sites
FOV	Field Of View
FPAR	Fraction of canopy-absorbed Photosynthetically Active Radiation
Gb	Gigabyte
GEOS	Goddard Earth Observing System (model)
GOSAT	JAXA Greenhouse gases Observing Satellite
GPP	Gross Primary Production
GSFC	Goddard Space Flight Center
GMAO	Goddard Modeling and Assimilation Office
HDF	Hierarchical Data Format
IGBP	International Geosphere-Biosphere Programme
IPO	NPOESS Integrated Program Office
JAXA	Japan Aerospace Exploration Agency
JPL	Jet Propulsion Laboratory

JPSS	Joint Polar Satellite System
LC	Land Cover
LSM	Land Surface Model
MCMC	Markov Chain Monte Carlo
MERRA	Modern Era Retrospective-analysis for Research and Applications
MODIS	MODerate-resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NetCDF	Network Common Data Form
NEE	Net Ecosystem Exchange of carbon dioxide (CO <sub>2</sub> )
NOAA	National Oceanic and Atmospheric Administration
NPP	NPOESS Preparatory Project or Net Primary Production
NRC	National Research Council
NTSG	Numerical Terradynamic Simulation Group
OCO	NASA Orbiting Carbon Observatory
P	Precipitation
PDF	Probability Density Function
PET	Potential Evapotranspiration
QA	Quality Assessment
QC	Quality Control
RFI	Radio Frequency Interference
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
SDS	SMAP Science Data System
SM	Soil Moisture
SMOS	Soil Moisture Ocean Salinity (ESA space mission)
SOC	Soil Organic Carbon
SPDM	Science Processing and Data Management
Tb	Terabyte
TBD	To Be Determined
VI	Vegetation Index
VIIRS	Visible Infrared Imager Radiometer Suite
VPD	Vapor Pressure Deficit
VWC	Vegetation Water Content

## REFERENCE DOCUMENTS

### Requirements:

- SMAP Level 1 Mission Requirements and Success Criteria. (Appendix O to the Earth Systematic Missions Program Plan: Program-Level Requirements on the Soil Moisture Active Passive Project.). NASA Headquarters/Earth Science Division, Washington, DC.
- SMAP Level 2 Science Requirements. SMAP Project, JPL D-45955, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Level 3 Science Algorithms and Validation Requirements. SMAP Project, JPL D-45993, Jet Propulsion Laboratory, Pasadena, CA.

### Plans:

- SMAP Science Data Management and Archive Plan. SMAP Project, JPL D-45973, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Science Data Calibration and Validation Plan. SMAP Project, JPL D-52544, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Applications Plan. SMAP Project, JPL D-53082, Jet Propulsion Laboratory, Pasadena, CA.

### ATBDs:

- SMAP Algorithm Theoretical Basis Document: L1B and L1C Radar Products. SMAP Project, JPL D-53052, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L1B Radiometer Product. SMAP Project, GSFC-SMAP-006, NASA Goddard Space Flight Center, Greenbelt, MD.
- SMAP Algorithm Theoretical Basis Document: L1C Radiometer Product. SMAP Project, JPL D-53053, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L2 & L3 Radar Soil Moisture (Active) Products. SMAP Project, JPL D-66479, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L2 & L3 Radiometer Soil Moisture (Passive) Products. SMAP Project, JPL D-66480, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L2 & L3 Radar/Radiometer Soil Moisture (Active/Passive) Products. SMAP Project, JPL D-66481, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L3 Radar Freeze/Thaw (Active) Product. SMAP Project, JPL D-66482, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L4 Surface and Root-Zone Soil Moisture Product. SMAP Project, JPL D-66483, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L4 Carbon Product. SMAP Project, JPL D-66484, Jet Propulsion Laboratory, Pasadena, CA.

### Ancillary Data Reports:

- Ancillary Data Report: Crop Type. SMAP Project, JPL D-53054, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Digital Elevation Model. SMAP Project, JPL D-53056, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Land Cover Classification. SMAP Project, JPL D-53057, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Soil Attributes. SMAP Project, JPL D-53058, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Static Water Fraction. SMAP Project, JPL D-53059, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Urban Area. SMAP Project, JPL D-53060, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Vegetation Water Content. SMAP Project, JPL D-53061, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Permanent Ice. SMAP Project, JPL D-53062, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Precipitation. SMAP Project, JPL D-53063, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Snow. SMAP Project, GSFC-SMAP-007, NASA Goddard Space Flight Center, Greenbelt, MD.
- Ancillary Data Report: Surface Temperature. SMAP Project, JPL D-53064, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Vegetation and Roughness Parameters. SMAP Project, JPL D-53065, Jet Propulsion Laboratory, Pasadena, CA.

## 1. INTRODUCTION

The Soil Moisture Active Passive (SMAP) mission is a NASA Decadal Survey mission consisting of a satellite L-band radar and radiometer instrument suite, with a projected launch in 2015 (Jan) that will provide global measurements and monitoring of soil moisture and landscape freeze/thaw state dynamics (Entekhabi et al. 2010). The SMAP mission includes a set of planned operational data products consisting of lower order (Level 1) brightness temperature and radar backscatter measurements, higher order (Level 2/3) soil moisture and freeze/thaw retrievals, and model enhanced (Level 4) soil moisture and carbon products. This document describes the algorithm and validation approach for the SMAP Level 4 Carbon (L4\_C) product.

The net ecosystem exchange (NEE) of CO<sub>2</sub> with the atmosphere is a fundamental measure of the balance between carbon uptake by vegetation gross primary production (GPP) and carbon losses through autotrophic (R<sub>a</sub>) and heterotrophic (R<sub>h</sub>) respiration. The sum of R<sub>a</sub> and R<sub>h</sub> defines the total ecosystem respiration rate (R<sub>tot</sub>), which encompasses most of the annual terrestrial CO<sub>2</sub> efflux to the atmosphere and typically represents 70-80 percent of the total magnitude of carbon uptake by GPP (Baldocchi 2008). The NEE term provides a measure of the terrestrial biosphere capacity as a net source or sink for atmospheric CO<sub>2</sub> and its ability to offset or reinforce anthropogenic greenhouse gas emissions purported to be a major driver of global warming (IPCC 2007). NEE and component GPP and respiration rates are spatially heterogeneous and temporally dynamic, and strongly influenced by changing environmental conditions encapsulated by the SMAP measurements. The baseline L4\_C algorithms will utilize daily level 3 freeze/thaw inputs (L3\_SM\_A) and level 4 soil moisture (L4\_SM) inputs with other ancillary geophysical data, including satellite (MODIS, VIIRS) derived canopy fraction of photosynthetically active radiation (FPAR) and daily surface meteorology from observation constrained global model reanalysis to compute NEE over all global vegetated land areas. The L4\_SM and L3\_SM\_A inputs will define low soil moisture and frozen temperature constraints to GPP and respiration calculations, providing a direct link between SMAP product retrievals and net ecosystem CO<sub>2</sub> exchange, and underlying vegetation productivity, soil decomposition and respiration processes. Primary science objectives of the L4\_C product are to:

- *Determine NEE regional patterns and temporal (daily, seasonal, and annual) behavior to within the accuracy range of in situ tower measurement based estimates of these processes;*
- *Link NEE estimates with component carbon fluxes (GPP and R<sub>tot</sub>) and the primary environmental constraints to ecosystem productivity and respiration.*

Primary science objectives for SMAP as directed by the National Research Council's Decadal Survey (NRC 2007) and directly relevant to the L4\_C product include improving understanding of processes linking terrestrial water, energy and carbon cycles; quantifying the net carbon flux in boreal landscapes and reducing uncertainties regarding the purported missing carbon sink on land.

The SMAP L4\_C baseline product will have a 9 km spatial resolution consistent with the L4\_SM product, but will retain sub-grid scale heterogeneity information determined from finer scale (1-3km resolution) land cover, FPAR and freeze/thaw (L3\_SM\_A) inputs. The resulting carbon product will be similar to the scale of tower CO<sub>2</sub> eddy covariance flux measurements (Baldocchi et al. 2008, Chen et al. 2012). The baseline L4\_C domain will encompass all global vegetated land areas and will attain a mean RMSE accuracy for NEE within 30 g C m<sup>-2</sup> yr<sup>-1</sup> (1.6

g C m<sup>-2</sup> d<sup>-1</sup>) over northern (≥45°N) boreal and arctic biomes, which is within the estimated ±30-100 g C m<sup>-2</sup> yr<sup>-1</sup> accuracy of *in situ* tower measurements (Baldocchi 2008; Richardson 2005; Richardson 2008). The L4\_C baseline product will have a mean daily temporal sampling to characterize the dynamic NEE response to daily variations in surface meteorology and associated moisture and thermal controls to GPP and respiration, and for greater precision in the computation of cumulative (weekly, monthly and annual) carbon fluxes. Operational implementation of the baseline L4\_C algorithms using L4\_SM and reanalysis (GMAO) surface meteorology inputs enables spatially and temporally continuous daily mapping of NEE for all vegetated land areas independent of data gaps, vegetation biomass and other constraints on SMAP parameter retrievals. The product will determine NEE and component carbon fluxes over all global vegetated land areas consistent with the scale of carbon processes of the major regional plant functional types.

The current algorithm baseline was developed from previous L4\_C ATBD external reviews (e.g. Kimball et al. 2009b) and ongoing algorithm development and assessment activities during the SMAP pre-launch phase (Kimball et al. 2009a, 2011, Jackson et al. 2012, McGuire et al. 2012, Yi et al. 2013). Two options are also being considered for L4\_C operational implementation in place of the current baseline algorithm. These options potentially enhance product accuracy, reliability and science utility, and include: 1) alternative GPP calculations using satellite optical-IR based Vegetation Indices (VIs) instead of higher order FPAR inputs from MODIS, and 2) representation of non-steady state fire disturbance and recovery effects on NEE and component carbon fluxes. This ATBD includes a description of the L4\_C algorithm and both baseline and optional implementation schemes. The ATBD also includes discussion of the major theoretical assumptions and procedures for refining and testing the algorithm to achieve the product objectives and mission requirements.

## 2. OVERVIEW AND BACKGROUND

### 2.1 The Soil Moisture Active Passive (SMAP) Mission

#### 2.1.1 Background and Science Objectives

The National Research Council's (NRC) Decadal Survey, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, was released in 2007 after a two year study commissioned by NASA, NOAA, and USGS to provide them with prioritization recommendations for space-based Earth observation programs (NRC 2007). Factors including scientific value, societal benefit and technical maturity of mission concepts were considered as criteria. SMAP data products have high science value and provide data towards improving many natural hazards applications. Furthermore SMAP draws on the significant design and risk-reduction heritage of the Hydrosphere State (HYDROS) mission (Entekhabi et al. 2004). For these reasons, the NRC report placed SMAP in the first tier of missions in its survey. In 2008 NASA announced the formation of the SMAP project as a joint effort of NASA's Jet Propulsion Laboratory (JPL) and Goddard Space Flight Center (GSFC), with project management responsibilities at JPL. The target launch date is January 2015.

The SMAP science and applications objectives are to:

- Understand processes that link the terrestrial water, energy and carbon cycles;
- Estimate global water and energy fluxes at the land surface;

- Quantify net carbon flux in boreal landscapes;
- Enhance weather and climate forecast skill;
- Develop improved flood prediction and drought monitoring capability.

### 2.1.2 Measurement Approach

A summary of the SMAP instrument functional requirements derived from the mission science measurement needs is presented in **Table 1**. The goal is to combine the attributes of the radar and radiometer observations (in terms of their spatial resolution and sensitivity to soil moisture, surface roughness, and vegetation) to estimate soil moisture at a resolution of 10 km, and freeze/thaw state at a resolution of 1-3 km.

**Table 1. SMAP Mission Requirements.**

Scientific Measurement Requirements	Instrument Functional Requirements
<u>Soil Moisture:</u> $\sim \pm 0.04 \text{ m}^3 \text{ m}^{-3}$ volumetric accuracy (1-sigma) in the top 5 cm for vegetation water content $\leq 5 \text{ kg m}^{-2}$ ; Hydrometeorology at $\sim 10 \text{ km}$ resolution; Hydroclimatology at $\sim 40 \text{ km}$ resolution	<u>L-Band Radiometer (1.41 GHz):</u> Polarization: V, H, T <sub>3</sub> and T <sub>4</sub> Resolution: 40 km Radiometric Uncertainty*: 1.3 K <u>L-Band Radar (1.26 and 1.29 GHz):</u> Polarization: VV, HH, HV (or VH) Resolution: 10 km Relative accuracy*: 0.5 dB (VV and HH) Constant incidence angle** between 35° and 50°
<u>Freeze/Thaw State:</u> Capture freeze/thaw state transitions in integrated vegetation-soil continuum with two-day precision, at the spatial scale of land-scape variability ( $\sim 3 \text{ km}$ ).	<u>L-Band Radar (1.26 GHz and 1.29 GHz):</u> Polarization: HH Resolution: 3 km Relative accuracy*: 0.7 dB (1 dB per channel if 2 channels are used) Constant incidence angle** between 35° and 50°
Sample diurnal cycle at consistent time of day (6am/6pm Equator crossing); Global, $\sim 3$ day (or better) revisit; Boreal, $\sim 2$ day (or better) revisit	Swath Width: $\sim 1000 \text{ km}$  Minimize Faraday rotation (degradation factor at L-band)
Observation over minimum of three annual cycles	Baseline three-year mission life
* Includes precision and calibration stability ** Defined without regard to local topographic variation	

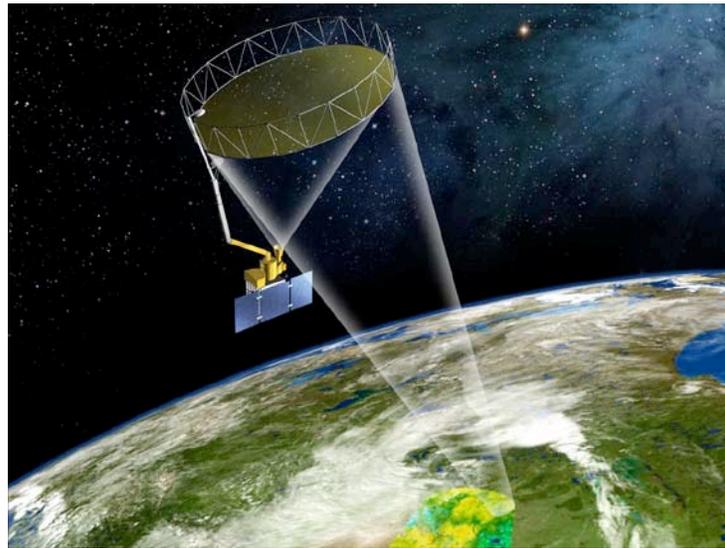
The SMAP observatory (see **Figure 1**) incorporates an L-band radar and an L-band radiometer that share a single feedhorn and parabolic mesh reflector. As shown in Figure 1 the reflector is offset from nadir and rotates about the nadir axis at 14.6 rpm (nominal), providing a conically scanning antenna beam with a surface incidence angle of approximately 40°. The provision of constant incidence angle across the swath simplifies the data processing and enables accurate repeat-pass estimation of soil moisture and freeze/thaw change. The reflector has a diameter of 6 m, providing a radiometer 3 dB antenna footprint of 40 km (root-ellipsoidal-area). The real-aperture radar footprint is 30 km, defined by the two-way antenna beamwidth. The real-aperture radar and radiometer data will be collected globally during both ascending and descending passes.

To obtain the desired high spatial resolution the radar employs range and Doppler discrimination. The radar data can be processed to yield resolution enhancement to 1-3 km spatial resolution over the 70% outer parts of the 1000 km swath. Data volume prohibits the

downlink of the entire radar data acquisition. Radar measurements that allow high-resolution processing will be collected during the morning overpass over all land regions and extending one swath width over the surrounding oceans. During the evening overpass data poleward of 45° N will be collected and processed as well to support robust detection of landscape freeze/thaw transitions.

The baseline orbit parameters are:

- Orbit Altitude: 685 km (2-3 days average revisit and 8-days exact repeat);
- Inclination: 98 degrees, sun-synchronous;
- Local Time of Ascending Node: 6 pm.



**Figure. 1.** *The SMAP observatory is a dedicated spacecraft with a rotating 6-m light-weight deployable mesh reflector. The radar and radiometer share a common feed.*

The SMAP radiometer measures the four Stokes parameters, V, H and  $T_3$ , and  $T_4$  at 1.41 GHz. The  $T_3$ -channel measurement can be used to correct for possible Faraday rotation caused by the ionosphere, although such Faraday rotation is minimized by the selection of the 6am/6pm sun-synchronous SMAP orbit.

At L-band anthropogenic Radio Frequency Interference (RFI), principally from ground-based surveillance radars, can contaminate both radar and radiometer measurements. Early measurements and results from the SMOS mission indicate that in some regions RFI is present and detectable. The SMAP radar and radiometer electronics and algorithms have been designed to include features to mitigate the effects of RFI. To combat this, the SMAP radar utilizes selective filters and an adjustable carrier frequency in order to tune to pre-determined RFI-free portions of the spectrum while on orbit. The SMAP radiometer will implement a combination of time and frequency diversity, kurtosis detection, and use of  $T_4$  thresholds to detect and where possible mitigate RFI.

Planned data products for the SMAP mission are listed in **Table 2**. Level 1B and 1C data products are calibrated and geolocated instrument measurements of surface radar backscatter cross-section and brightness temperatures derived from antenna temperatures. Level 2 products are geophysical retrievals of soil moisture on a fixed Earth grid based on Level 1 products and ancillary information; the Level 2 products are output on half-orbit basis. Level 3 products are daily composites of Level 2 surface soil moisture and freeze/thaw state data. Level 4 products are model-derived value-added data products that support key SMAP applications and more directly address the driving science questions.

**Table 2.** SMAP Data Products Table.

Product	Description	Gridding (Resolution)	Latency	
L1A_TB	Radiometer Data in Time-Order	-	12 hrs	Instrument Data
L1A_S0	Radar Data in Time-Order	-	12 hrs	
L1B_TB	Radiometer $T_B$ in Time-Order	(36x47 km)	12 hrs	
L1B_S0_LoRes	Low Resolution Radar $\sigma_0$ in Time-Order	(5x30 km)	12 hrs	
L1C_S0_HiRes	High Resolution Radar $\sigma_0$ in Half-Orbits	1 km (1-3 km)	12 hrs	
L1C_TB	Radiometer $T_B$ in Half-Orbits	36 km	12 hrs	
L2_SM_A	Soil Moisture (Radar)	3 km	24 hrs	Science Data (Half-Orbit)
L2_SM_P	Soil Moisture (Radiometer)	36 km	24 hrs	
L2_SM_AP	Soil Moisture (Radar + Radiometer)	9 km	24 hrs	
L3_FT_A	Freeze/Thaw State (Radar)	3 km	50 hrs	Science Data (Daily Composite)
L3_SM_A	Soil Moisture (Radar)	3 km	50 hrs	
L3_SM_P	Soil Moisture (Radiometer)	36 km	50 hrs	
L3_SM_AP	Soil Moisture (Radar + Radiometer)	9 km	50 hrs	
L4_SM	Soil Moisture (Surface and Root Zone)	9 km	7 days	Science Value-Added
L4_C	Carbon Net Ecosystem Exchange (NEE)	9 km	14 days	

## 2.2 L4\_C Product/Algorithm Objectives

The primary science objectives of the L4\_C product are to:

- Determine NEE regional patterns and temporal (daily, seasonal, and annual) behavior at the accuracy level of in situ tower measurement based estimates of these processes;
- Link NEE estimates with component carbon fluxes (GPP and  $R_{tot}$ ) and the primary environmental constraints to ecosystem productivity and respiration.

The NRC Decadal Survey for *Earth Science Applications from Space* (NRC 2007) recognized the importance of soil moisture and its freeze/thaw state in the global carbon cycle, and particularly for northern latitudes where biophysical processes are strongly limited by frozen temperatures for much of the year: “Soil moisture and its freeze/thaw state are key determinants of the global carbon cycle” and “Carbon uptake and release in boreal landscapes are a major source of uncertainty in assessing the carbon budget of the Earth system (the so-called missing carbon sink).” The Decadal Survey further establishes the importance of the SMAP mission by stating that “A soil moisture mission will directly support science to reduce that major uncertainty”, in reference to the purported “missing” carbon sink on land. The L4\_C algorithm

addresses carbon cycle science objectives as put forth in the Decadal Survey by enabling detailed mapping and monitoring of spatial patterns and temporal dynamics of land-atmosphere CO<sub>2</sub> exchange, and the underlying carbon fluxes and environmental drivers of these processes. The L4\_C product will also link SMAP land parameter measurements to global terrestrial CO<sub>2</sub> exchange, including boreal ecosystems, reducing uncertainties about the “missing sink” on land for atmospheric CO<sub>2</sub>.

Atmospheric transport model inversions of CO<sub>2</sub> concentrations indicate that the Northern Hemisphere terrestrial biosphere is responsible for much of the recent terrestrial sink strength for atmospheric carbon (Dargaville et al. 2002). Variability in land-atmosphere CO<sub>2</sub> exchange is strongly controlled by climatic fluctuations and disturbance, while uncertainty regarding the magnitude and stability of the sink are constrained by a lack of detailed knowledge on the response of underlying processes at regional scales (Denman et al. 2007, Houghton 2003). The SMAP mission provides the potential for much improved spatial resolution and L-band active/passive microwave sensitivity to land surface processes for monitoring soil moisture and thermal dynamics of global ecosystems, including boreal and arctic biomes. The baseline L4\_C algorithms will use daily inputs from the SMAP L3\_SM\_A and L4\_SM product streams to define frozen temperature and soil moisture constraints to vegetation productivity, ecosystem respiration and NEE. Landscape freeze/thaw state classification inputs from the SMAP L3\_SM\_A product will be used to define the proportional frozen area extent within each 9 km resolution grid cell and as an additional frozen temperature constraint to vegetation gross primary productivity (GPP) and ecosystem respiration calculations to determine NEE. The L4\_C algorithm will provide estimates of NEE (g C m<sup>-2</sup> day<sup>-1</sup>) and component carbon fluxes for global vegetated land areas at mean daily intervals; the product will have 9-km spatial resolution, but will define sub-grid scale mean and variability in carbon fluxes for dominant and sub-dominant vegetation classes within each grid cell as determined from finer scale (1-km resolution) ancillary land cover classification and FPAR inputs. The NEE product will attain a mean annual RMSE accuracy less than or equal to 30 g C m<sup>-2</sup> yr<sup>-1</sup> (1.6 g C m<sup>-2</sup> d<sup>-1</sup>), and commensurate with the estimated accuracy of *in situ* tower measurements (Baldocchi 2008, Richardson 2005, Richardson 2008). The baseline L4\_C product spatial resolution will be 9-km, consistent with the SMAP L4\_SM inputs, but the product will preserve sub-grid scale heterogeneity in carbon fluxes represented by finer (1-3km) resolution land cover and freeze/thaw (L3\_SM\_A) inputs, and algorithm processing. The resulting L4\_C product spatial resolution will be similar to the sampling footprint of CO<sub>2</sub> flux measurements from the global tower (FLUXNET) monitoring network (Running et al. 1999, Baldocchi et al. 2008). Secondary products of scientific value produced during L4\_C processing include surface (<10 cm depth) soil organic carbon (SOC) stocks (g C m<sup>-2</sup>), vegetation gross primary production (GPP), net primary production (NPP), ecosystem respiration (R<sub>tot</sub>), heterotrophic (R<sub>h</sub>) and autotrophic (R<sub>a</sub>) respiration components, and dimensionless (0-100 percent) frozen area, low temperature and moisture constraint indices for GPP and R<sub>tot</sub>.

The L4\_C product will enable quantification and mechanistic understanding of spatial and temporal variations in NEE over a global domain. NEE represents the primary measure of carbon (CO<sub>2</sub>) exchange between the land and atmosphere, and the L4\_C product will be directly relevant to a range of applications including regional mapping and monitoring of terrestrial carbon stocks and atmospheric transport model inversions of terrestrial source-sink activity for atmospheric CO<sub>2</sub>. The SMAP L4\_C product will also satisfy carbon cycle science objectives of the NRC Decadal Survey and advance our understanding of the way in which global ecosystems, including boreal-

Arctic biomes, respond to climate anomalies and their capacity to reinforce or mitigate global warming.

### 2.3 Historical Perspective

Current capabilities for regional assessment and monitoring of NEE are limited. Atmospheric transport model inversions of CO<sub>2</sub> concentrations from sparse measurement stations provide information on seasonal patterns and trends in atmospheric CO<sub>2</sub> but little information on underlying processes; these methods are also too coarse to resolve carbon source-sink activity at scales finer than broad latitudinal and continental domains (Piao et al. 2007, Dargaville et al. 2002; Yi et al. 2014). Tower CO<sub>2</sub> flux measurement networks provide detailed information on stand level NEE and associated biophysical processes, but little information regarding spatial variability in these processes over heterogeneous landscapes (Running et al. 1999). Estimates of NEE and component carbon fluxes from satellite remote sensing provide a means for scaling between relatively intensive stand level measurement and modeling approaches, and top down assessments from atmospheric model inversions.

The Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the NASA EOS Terra and Aqua satellites has been providing global, operational mapping of GPP at approximate 8-day intervals since 2000 (Running et al. 2004), while a similar operational product is planned for the JPSS Visible/Infrared Imager/Radiometer Suite (VIIRS) scheduled for NPOESS (NGST 2011). The GPP term quantifies the photosynthetic uptake of atmospheric CO<sub>2</sub>, but represents an incomplete picture of NEE because of a lack of information on ecosystem respiration. Several studies have applied satellite remote sensing to characterize NEE over northern landscapes using empirical relationships between CO<sub>2</sub> flux measurements and spectral vegetation indices (Hope et al. 1995, McMichael et al. 1999) or physiological models driven by optical-infrared (IR) remote sensing and surface meteorological data to characterize both vegetation productivity and ecosystem respiration (Vourlitis et al. 2000, Potter et al. 2003, Veroustraete et al. 2002). Empirical approaches can provide relatively high estimation accuracy, but are limited by the quality, quantity and representativeness of observational data used for model development and training (e.g. Jung et al. 2011); the resulting simulations are generally suitable for the specific regions and conditions under which the models were developed and provide limited diagnostic insight into underlying biophysical processes. Physiological models attempt to account for the primary environmental controls on productivity and respiration, but are often constrained by the availability and resolution of driving meteorological datasets from sparse observational networks or coarse (0.5–2.5 degree) resolution gridded products from atmospheric model reanalyses.

Recent developments in satellite remote sensing offer the potential for direct measurement and improved resolution of environmental conditions for estimating land-atmosphere carbon exchange. Satellite microwave radiometers are sensitive to variations in surface emissivity and dielectric constant associated with changes in soil moisture and temperature (Jackson et al. 1999, Njoku et al. 2003). Lower frequency microwaves (e.g., < 18.7 GHz) are capable of penetrating clouds and low-biomass vegetation to provide information more representative of the underlying soil than higher frequency microwave and thermal infrared (IR) observations. These favorable properties have been exploited for mapping surface soil moisture and temperature across a wide range of environments and vegetation types, including boreal forest and tundra (Fily et al. 2003, Jones et al. 2007, 2010). The NASA Advanced Microwave Scanning Radiometer for EOS (AMSR-E) is co-located with MODIS on the NASA Aqua satellite and has provided operational and experimental land products, including surface soil moisture (Njoku et al. 2003), landscape

freeze/thaw status and soil temperature (Jones et al. 2007, Kim et al. 2012b); these data have been used as surrogate measures of soil moisture and thermal controls to respiration calculations, and together with operational (MOD17) GPP calculations from MODIS have been used for estimating NEE in boreal-Arctic biomes (Kimball et al. 2009a, Yi et al. 2013).

Satellite active and passive microwave remote sensing retrievals of landscape freeze/thaw status have been shown to correspond closely with seasonal frozen temperature constraints to water mobility, vegetation productivity and ecosystem respiration (Kimball et al. 2004, 2006, Smith et al. 2004, Kim et al. 2012). The influence of the freeze/thaw signal on vegetation productivity and NEE is generally greater at upper elevations and higher (i.e.  $>45^{\circ}\text{N}$ ) latitudes where the frozen season represents a larger proportion of the annual cycle (Nemani et al. 2003; Kim et al. 2012). The freeze/thaw signal from moderate resolution satellite radars and radiometers generally does not distinguish individual (e.g. soil, snow, vegetation) elements within the satellite FOV, though lower (L-band) frequency and relatively fine (3-km resolution) scale freeze/thaw retrievals from SMAP offer the potential for improved freeze/thaw classification accuracy and greater sensitivity to vegetation canopy and soil layers (Du et al. 2014).

Synergistic information from satellite optical-IR and microwave remote sensing data series were applied within a simple decomposition model for regional mapping and monitoring of NEE and component GPP and respiration fluxes over a network of boreal forest, tundra and northern grassland monitoring sites, and 3-year (2002-2004) time period (Kimball et al. 2009a). The modeling framework is similar to the L4\_C algorithms except that model soil moisture and temperature inputs were derived from AMSR-E daily brightness temperature time series. The resulting RMSE differences between model estimates and tower  $\text{CO}_2$  flux observations were 1.2, 0.7, and  $1.2 \text{ g C m}^{-2} \text{ d}^{-1}$  for GPP,  $R_{\text{tot}}$  and NEE, while mean residual differences were 43 % of the RMSE. The model accuracies were also similar to detailed site level ecosystem process model (BIOME-BGC) simulations, while the model derived SOC estimates compared favorably with both site and global soil inventory records; the model results and their associated agreement with the soil inventory data indicated that the dominant SOC source for  $R_{\text{h}}$  comes from surface soil layers with a mean residence time of a decade or less, while  $R_{\text{h}}$  accounts for more than half of annual ecosystem respiration. An error sensitivity analysis determined that meaningful carbon flux estimates could be derived under prevailing climatic conditions at the study locations given documented error levels in the remote sensing inputs. The relatively coarse ( $\sim 12\text{-}60 \text{ km}$ ) spatial scale of the AMSR-E footprint limits the ability of the sensor to resolve sub-grid scale land surface properties, and the AMSR-E sensor is largely insensitive to surface soil moisture conditions where the overlying vegetation biomass water content is above approximately  $1.5 \text{ kg m}^{-2}$ . However, satellite microwave remote sensing provides the ability to monitor land surface conditions day or night, independent of solar illumination or signal degradation from cloud cover, smoke and other atmospheric aerosol effects; global coverage and temporal fidelity of the measurements is largely determined by the orbital geometry of the satellite measurements, potential data loss from sensor malfunctions and signal degradation from vegetation, snow and atmosphere effects. The combination of satellite retrievals and other observational data within a land model data assimilation framework provides capabilities for contiguous global coverage and continuous daily monitoring, with improved soil moisture estimates when compared to the model or retrieval estimates alone (Liu et al. 2011, Reichle et al. 2007).

The L4\_C algorithm elements are relatively mature. The MOD17 light use efficiency (LUE) algorithm logic and resulting GPP products have been extensively evaluated over global and northern land areas, and their uncertainty is well established (e.g. Zhao et al. 2005, Heinsch et al. 2006, Turner et al. 2006, Zhang et al. 2008, Schaefer et al. 2012). The L4\_C theoretical assumptions and respiration rate calculations are based on the well known Century (Parton et al. 1987, Ise and Moorcroft 2006) and CASA (Potter et al. 1993) soil decomposition models, but are adapted for use with daily biophysical inputs derived from both satellite optical-IR and microwave remote sensing time series as primary model drivers. The L4\_C model framework was also successfully tested over a latitudinal gradient of northern climate, land cover and vegetation biomass conditions using MODIS GPP and AMSR-E soil moisture and temperature inputs and was found to produce results with documented uncertainty that were similar to tower based CO<sub>2</sub> flux measurements (Kimball et al. 2009a). A subsequent test of the L4\_C terrestrial carbon flux (TCF) model was conducted over northern ( $\geq 45^\circ\text{N}$ ) land areas using MODIS vegetation and GMAO (GEOS-5 MERRA) reanalysis daily surface meteorology inputs; these simulations showed similar algorithm performance relative to more detailed process model simulations of regional land-atmosphere CO<sub>2</sub> exchange (McGuire et al. 2012) and generally favorable results against regional tower (FLUXNET) observation based carbon fluxes (Yi et al. 2013). These results provide a foundation for gauging the relative improvement in regional carbon flux measures provided by SMAP soil moisture and thermal information over existing capabilities.

## 2.4 Data Product Characteristics

### 2.4.1 Instrument/Calibration Aspects (affecting product)

The L4\_C algorithms will utilize SMAP freeze/thaw (L3\_SM\_A) and soil moisture (L4\_SM) product fields as primary model inputs. The L4\_C algorithms and product sensitivity to SMAP instrument and calibration aspects largely flows through from the L3\_SM\_A and L4\_SM inputs. These effects will be partially mitigated through L4\_SM reliance on multiple satellite remote sensing and biophysical data sources in addition to SMAP product inputs (Reichle et al. 2011). Sensor calibration aspects influencing the SMAP L3\_SM\_A based freeze/thaw retrievals may be imparted to the L4\_C product through their use as an additional environmental constraint to GPP and associated NEE calculations. The relative impact of these properties on the L4\_C product is being evaluated through pre-launch algorithm sensitivity studies (e.g. Yi et al. 2013).

### 2.4.2 Product Scope and Format

The primary (validated) L4\_C output variable is the net ecosystem exchange (NEE) of CO<sub>2</sub> ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) between the land and atmosphere on a per grid cell basis. The L4\_C product will also contain supporting model outputs including component respiration ( $R_h$ ) and GPP fluxes ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) that enable the derivation of other significant carbon variables including NPP ( $\text{NEE} - R_h$ ),  $R_a$  ( $\text{GPP} - \text{NPP}$ ), and  $R_{\text{tot}}$  ( $R_h + R_a$ ); surface (<10 cm depth) SOC pools ( $\text{g C m}^{-2}$ ); environmental constraint indices influencing ecosystem productivity and respiration calculations, and data quality flags for the NEE retrieval. The L4\_C derived outputs will be produced at a daily time step. The domain of the L4\_C baseline product encompasses all global vegetated land areas (**Figure 2**) as determined by ancillary land cover classification, FPAR and L4\_SM inputs. The L4\_C outputs will be posted to a 9 km resolution Earth Grid and global EASE-grid (version 2) projection format (Brodzik et al. 2012), and consistent with the resolution of SMAP L4\_SM

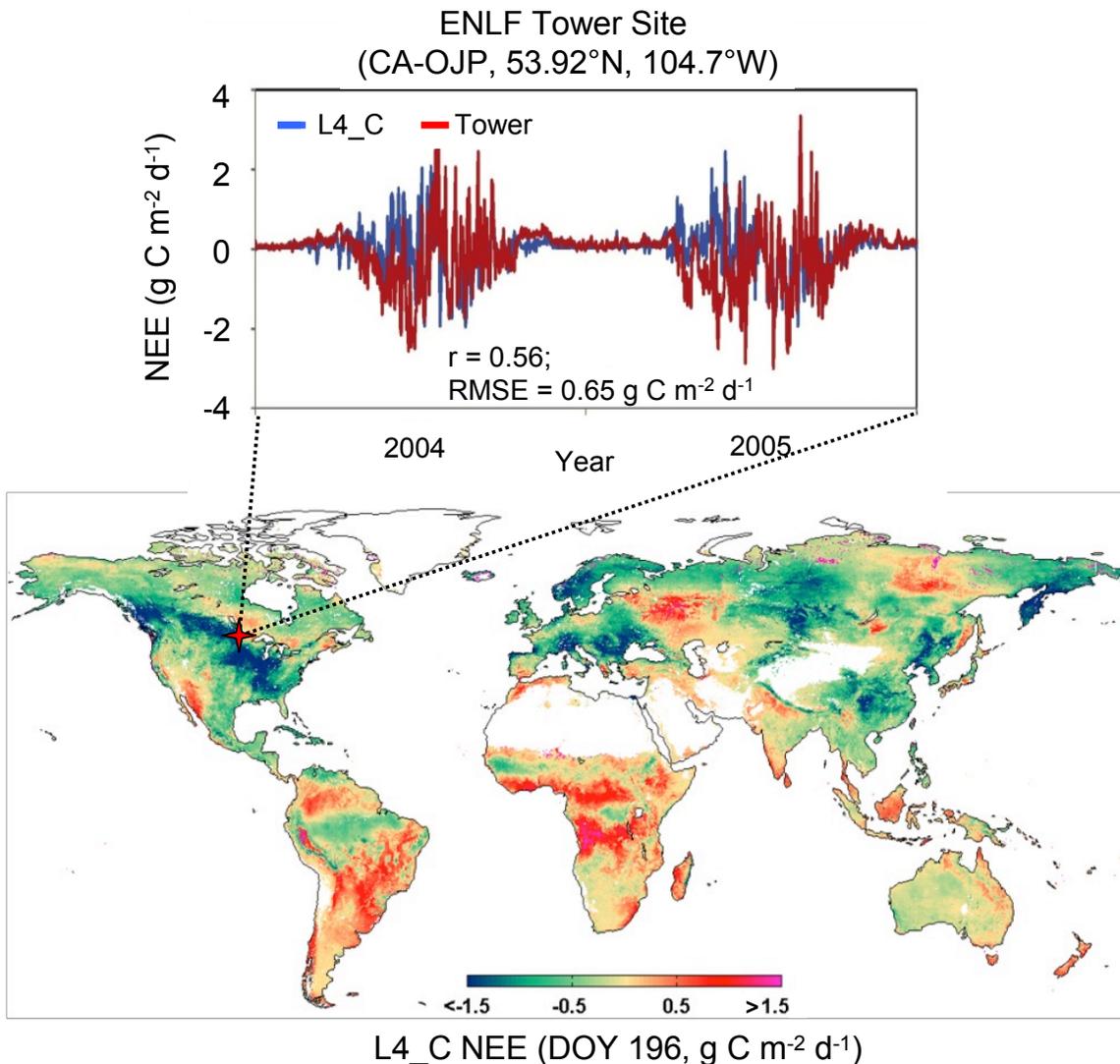
inputs. Operational L4\_C processing will be conducted at a finer (1-km) spatial resolution consistent with the ancillary satellite land cover classification and FPAR inputs. Sub-grid scale spatial means and variability (SD) of the 1-km resolution L4\_C calculations will be represented for the total area and individual land cover classes within each 9 km resolution grid cell. Additional metadata will be provided with the L4\_C product including the geographic (lat/lon) location of each grid cell and fractional (%) representation of individual land cover classes within each 9-km product grid cell, and defined from finer (1-km) scale ancillary land cover classification inputs used for the L4\_C calculations. These ancillary data are expected to remain static throughout the operational mission period though the land cover data may be updated under planned SMAP data reprocessing efforts to exploit potential improvements in available global land cover products that may benefit L4\_C product accuracy and utility.

Initiation of the L4\_C product stream is planned following initiation of the SMAP L3\_SM\_A and L4\_SM product streams following the in-orbit check-out (IOC) period and approximately 3 months after launch, with a planned beta release of the data product to the SMAP mission DAAC at NSIDC approximately 6 months after launch. The L4\_C product will have a target mean data latency of no more than 14-days. The L4\_C product latency is driven by latency and availability of ancillary FPAR (12-day latency) and L4\_SM (planned 7-day latency) inputs, and an additional 2-day data processing period. The targeted accuracy of the L4\_C NEE product is  $\leq 30 \text{ g C m}^{-2} \text{ yr}^{-1}$  or  $1.6 \text{ g C m}^{-2} \text{ d}^{-1}$  (RMSE), similar to the accuracy attained from tower based  $\text{CO}_2$  flux measurements (Baldocchi 2008, Richardson 2005, Richardson 2008).

The planned L4\_C daily product format is summarized in **Table 3**. Each daily product granule will include individual variable fields for each grid cell representing NEE, GPP, and  $R_h$  fluxes, surface SOC, and environmental constraint (EC) indices affecting the productivity and respiration calculations. Sub-grid spatial means and variability (SD) in carbon fluxes and SOC determined from the 1-km resolution processing within each 9-km grid cell will be represented as additional product fields within each granule, including regional means from up to 8 discrete global plant functional types (PFTs) represented within each 9-km grid cell; PFT representation within each 9-km grid cell will be defined from the 1-km resolution global land cover classification (Friedl et al. 2010) used to drive the L4\_C model simulations. The (8) PFT classes and initial BPLUT parameters associated with these classes are enumerated in the Appendix (Section 6). The EC product fields include four separate fields distinguishing: proportional frozen area within each 9-km grid cell and defined from the 3-km resolution freeze/thaw (L3\_SM\_A) inputs; the estimated mean bulk PAR conversion efficiency constraint ( $\epsilon_{\text{mult}}$ ) to the LUE model based vegetation productivity calculations; and estimated mean soil moisture ( $W_{\text{mult}}$ ) and temperature ( $T_{\text{mult}}$ ) constraints to  $R_h$  within each 9-km grid cell. The EC indices for  $\epsilon_{\text{mult}}$ ,  $W_{\text{mult}}$ , and  $T_{\text{mult}}$  are derived from the same dimensionless multipliers used to derive vegetation productivity and  $R_h$  (Section 3), but are rescaled in the product table to range from 0 (fully constrained) to 100 (no constraint) percent.

A separate global daily product granule will be derived that includes dimensionless data quality (QA) identifiers for the NEE calculation, and associated geolocation (latitude, longitude) information. The QA information will be derived from global model NEE performance (RMSE) diagnostics established from prior model sensitivity simulations, and daily quality control (QC) information obtained from lower order FPAR, freeze/thaw (L3\_SM\_A) and soil moisture (L4\_SM) inputs. The QA metric for NEE is derived in geophysical units ( $\text{g C m}^{-2}$ , RMSE) that are rescaled to a smaller set of 4 discrete general quality categories (e.g. 0=best, 1=good, 2=fair

and 3=poor) in the final product QA output fields. However, the detailed NEE RMSE QA geophysical units may be preserved as research data, separate from the L4\_C operational product, for supporting post-launch product performance and validation assessments. The L4\_C QA granules will have a similar format as the primary carbon model output granules described above. The QA fields represent model NEE quality metrics for each 9-km grid cell and up to 8 PFT classes represented within each cell. The QA granule fields also include the number of underlying NEE pixels within each 9-km grid cell used to estimate the aggregate QA value of the 9-km grid cell, and underlying pixel counts of each of the 8, 1-km resolution PFT classes represented within each 9-km product grid cell. Requirements for conducting the 1-km resolution model simulations include having vegetated (PFT 1-8) land area with valid FPAR. For the L4C output product raster dimensions referred to below (*Table 3*), the NROWS term equals 1624, and the NCOLS term equals 3856.



**Figure 2.** L4\_C global NEE product example for a selected day, and derived using MODIS (MOD15) FPAR and GMAO (MERRA) reanalysis daily surface meteorology inputs. The L4\_C product will encompass all global vegetated land areas. A comparison of L4\_C and tower eddy covariance measurement based daily NEE is also shown for a representative coniferous

evergreen needleleaf forest (ENLF) tower validation site. A negative (positive) carbon flux denotes net ecosystem uptake (loss) of atmospheric CO<sub>2</sub>. Non-vegetated land and other areas outside of the processing domain are shown in grey and white.

**Table 3.** Planned L4\_C daily product format, including daily granules for carbon model outputs and estimated NEE daily quality (QA) identifiers. Each granule is gridded to 9-km resolution, where each grid cell contains variable fields representing GPP, R<sub>h</sub> and NEE fluxes, surface SOC, environmental constraint (EC) indices and geolocation (latitude, longitude) of the grid cell center. A separate set of daily granules is produced containing dimensionless NEE QA fields for each 9-km grid cell and underlying PFT types, and counts of valid 1-km resolution processing (and PFT) pixels within each 9-km cell. Each granule will be stored as a separate file. Sub-grid spatial average (av) and SD (sd) in the underlying 1-km model calculations are defined as individual fields within each granule, including regional means from up to 8 global PFT classes within each cell. The EC variables include proportional frozen area (FA), estimated bulk environmental constraint to PAR conversion efficiency and vegetation productivity (Emult), and soil moisture (Wmult) and temperature (Tmult) constraints to R<sub>h</sub>.

Granule	Variable Class	Variable Fields	Dimensions	Data type	Units	Valid Range
Primary L4_C outputs	GPP_9km	GPP_cell_av GPP_cell_sd GPP_pft_1_av GPP_pft_2_av ..._8_av	NROW *NCOL	Flt32	g C m <sup>-2</sup> d <sup>-1</sup>	0-30
	Rh_9km	Rh_cell_av Rh_cell_sd Rh_pft_1_av Rh_pft_2_av ..._8_av	NROW *NCOL	Flt32	g C m <sup>-2</sup> d <sup>-1</sup>	0-20
	NEE_9km	NEE_cell_av NEE_cell_sd NEE_pft_1_av NEE_pft_2_av ..._8_av	NROW *NCOL	Flt32	g C m <sup>-2</sup> d <sup>-1</sup>	-30 - 20
	SOC_9km	SOC_cell_av SOC_cell_sd SOC_pft_1_av SOC_pft_2_av ..._8_av	NROW *NCOL	Flt32	g C m <sup>-2</sup>	0-25,000
	EC_9km	FA Emult_av Wmult_av Tmult_av	NROW *NCOL	Uint8	%	0-100
	GEO_9km	Latitude	NROW *NCOL	Flt32	DD	-90.00 – 90.00
	GEO_9km	Longitude	NROW *NCOL	Flt32	DD	-180.00 – 180.00

Granule	Variable Class	Fields	Dimensions	Data type	Units	Valid Range
NEE QA outputs	QA_9km	QA_cell QA_pft_1 QA_pft_2 ...pft_8	NROW *NCOL	UInt8	N/A	0-3
	CNT_9km	QA_cnt_cell QA_cnt_pft_1 QA_cnt_pft_2 ...pft_8	NROW *NCOL	UInt8	N/A	0-81
	GEO_9km	Latitude	NROW *NCOL	Flt32	DD	-90.00 – 90.00
	GEO_9km	Longitude	NROW *NCOL	Flt32	DD	-180.00 – 180.00

The L4\_C product will be provided in an HDF5 file format consistent with the other SMAP operational products. Each nominal data granule of the L4\_C product will represent 1 day of data. The product will be provided in a global cylindrical EASE-grid (version 2) projection containing 3856 columns and 1624 rows with an assumed compression factor of 0.60. The estimated L4\_C product data volumes are summarized in **Table 4** and will be approximately 105 MB per day for the carbon model granule (e.g. GPP, Rh, NEE, SOC), 12 MB per day for the time matched product QA granule, and 117 MB per day for both granules in uncompressed HDF5 file format. In production, all granules will be compressed. The estimated total annual product (uncompressed) data volume is 41.7 GB (~25 GB compressed) and represents less than 1% of an estimated mission annual data volume of ca 433 TB. The 1 km resolution L4\_C data processing yields larger estimated data processing volumes relative to the coarser 9-km product resolution but does not impact the final estimated data product storage volumes.

**Table 4.** Estimated L4\_C data volumes.

Compression factor	Domain	Resolution	Temporal fidelity	Size: per granule (MB d <sup>-1</sup> )	Size: all granules (MB d <sup>-1</sup> )	Annual volume (GB)
Uncompressed	Global	9 km	Daily	105 (12)	117	41.70

### 3. RETRIEVAL ALGORITHM

#### 3.1 Theoretical Description

##### 3.1.1 Mathematical Description of the Algorithms

NEE (g C m<sup>-2</sup> d<sup>-1</sup>) is computed on a daily basis as the residual difference between GPP and respiration from autotrophic (R<sub>a</sub>) and heterotrophic (R<sub>h</sub>) components:

$$NEE = (R_a + R_h) - GPP \quad (1)$$

where positive (+) and negative (-) NEE fluxes denote the respective terrestrial loss or uptake of CO<sub>2</sub>. The GPP term (g C m<sup>-2</sup> d<sup>-1</sup>) represents the mean vegetation gross primary production of a specified land cover class within a grid cell and is derived on a daily basis using a light use efficiency (LUE) model similar to the MODIS (MOD17) operational productivity algorithm (Running et al. 2004, Heinsch et al. 2003, Zhao et al. 2005):

$$\text{GPP} = \varepsilon * \text{APAR} \quad (2)$$

where  $\varepsilon$  is the conversion efficiency (g C MJ<sup>-1</sup>) of photosynthetically active radiation (PAR) to vegetation biomass, and APAR (MJ m<sup>-2</sup> d<sup>-1</sup>) is the amount of PAR absorbed by the canopy and available for photosynthesis. PAR (MJ m<sup>-2</sup> d<sup>-1</sup>) is estimated as a constant proportion (0.45) of incident shortwave solar radiation at the surface (R<sub>sw</sub>, MJ m<sup>-2</sup> d<sup>-1</sup>) and is used with the estimated fraction of incident PAR absorbed by the vegetation canopy (FPAR) to determine APAR:

$$\text{APAR} = \text{PAR} * \text{FPAR} \quad (3)$$

Alternatively, the GEOS-5 land model, which forms the basis of the planned L4\_SM model assimilation product routinely produces dynamic PAR estimates which could be used to estimate APAR in Eqn. (3). The PAR conversion efficiency ( $\varepsilon$ ) term is derived on a daily basis from an estimated maximum rate ( $\varepsilon_{\text{mx}}$ , g C MJ<sup>-1</sup>) prescribed for different land cover classes (Zhao et al. 2005), and is reduced for sub-optimal environmental conditions defined as the product ( $\varepsilon_{\text{mult}}$ ) of dimensionless rate scalars ranging from no effect (1) to complete rate reduction (0) for daily minimum air temperature (T<sub>mn\_scalar</sub>), atmosphere vapor pressure deficit (VPD<sub>scalar</sub>), landscape freeze/thaw status (FT<sub>scalar</sub>) and integrated (0-1 m depth) surface to root zone soil moisture (SM<sub>rz\_scalar</sub>) conditions:

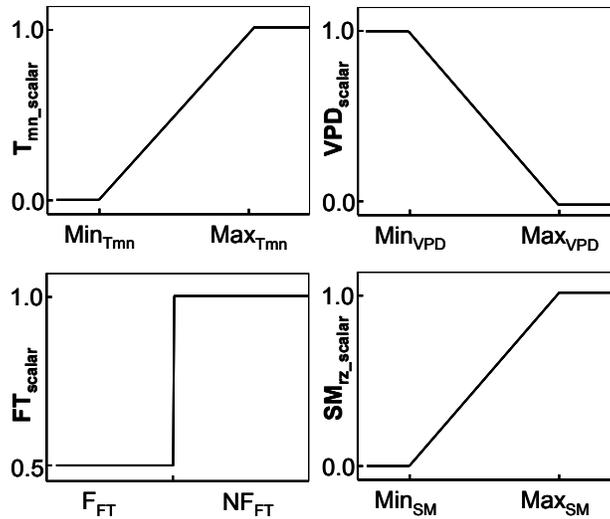
$$\varepsilon_{\text{mult}} = T_{\text{mn\_scalar}} * \text{VPD}_{\text{scalar}} * \text{FT}_{\text{scalar}} * \text{SM}_{\text{rz\_scalar}} \quad (4)$$

$$\varepsilon = \varepsilon_{\text{mx}} * \varepsilon_{\text{mult}} \quad (5)$$

The above attenuation scalars are defined as simple switch and linear ramp functions (**Figure 3**), and deviate from the original MOD17 algorithm logic (Zhao et al. 2005) by specifying additional environmental constraints for frozen landscape conditions and low root zone soil moisture (SM<sub>rz</sub>) levels. The attenuation functions vary according to prescribed minimum and maximum constraints determined for different global biome types (Field et al. 1995, Prince and Goward 1995, Turner et al. 2003, Zhao et al. 2005, Kim et al. 2012). The primary model environmental response characteristics in **Figure 3** are determined using control parameters defined for individual land cover classes within a general Biome Properties Look-Up Table (BPLUT). The BPLUT parameter definitions are summarized in **Table 5**. A detailed set of BPLUT parameters

has been developed for pre-launch L4\_C algorithm development and is summarized in the APPENDIX for individual biome types defined from a global land cover classification (Friedl et al. 2010). The BPLUT parameters were assembled from MODIS operational (MOD17) GPP product definitions (Zhao et al. 2005), detailed literature reviews and previous definition studies for different global biome types (e.g. White et al. 2000), and regional GPP and L4\_C model calibration and validation studies using in situ tower (FLUXNET) network based carbon fluxes (Heinsch et al. 2006, Kimball et al. 2009a, Yi et al. 2013).

The  $FT_{\text{scalar}}$  term represents the frozen temperature constraint to landscape water mobility and GPP as determined from regional comparisons between tower based GPP observations and daily FT retrievals from satellite microwave remote sensing (Kimball et al. 2004, Kim et al. 2012). The  $SM_{\text{rz\_scalar}}$  term provides a direct low soil moisture constraint to GPP in addition to the atmosphere VPD constraint. These additional terms provide for a direct link between SMAP FT and soil moisture products, and associated environmental constraints to GPP and terrestrial carbon flux calculations, whereas the original MOD17 algorithm relies solely on daily VPD and  $T_{\text{mn}}$  inputs to define the primary moisture and low temperature constraints to vegetation productivity (Running et al. 2004, Zhao et al. 2005).



**Figure 3.** Attenuation scalars used in the LUE model based GPP calculation for representing the PAR conversion efficiency ( $\epsilon$ ) reduction under suboptimal environmental conditions for daily minimum air temperature ( $T_{\text{mn}}$ ), vapor pressure deficit (VPD), root zone (0-1 m depth) soil moisture ( $SM_{\text{rz}}$ ), and landscape freeze/thaw (FT) classification defined frozen (F) and nonfrozen (NF) conditions. The dimensionless scalars are represented by linear ramp and switch functions that define the  $\epsilon$  reduction under prescribed maximum (Max) and minimum (Min) parameter ranges for different biome types. The scalars are dimensionless, ranging from 0 (fully constrained) to 1 (no constraint), and are used as equally weighted multipliers to define the daily bulk reduction ( $\epsilon_{\text{mult}}$ ) from a prescribed optimum ( $\epsilon_{\text{mx}}$ ) conversion efficiency. A detailed summary of BPLUT parameters that define the above response characteristics for different global biome types is presented in the Appendix (Section 6).

The autotrophic respiration ( $R_a$ ) term in Eqn. (1) represents the sum of vegetation growth and maintenance respiration, and is computed on a daily basis as a fixed proportion of GPP within individual land cover classes, based on observational evidence that variability in the ratio of  $R_a$  to GPP is conservative within individual plant functional types (Litton et al. 2007, Gifford 2003,

Waring et al. 1998). While this assumption provides a key simplification for a remote sensing based algorithm, the proportion of plant photosynthesis devoted to biophysical growth and maintenance may vary under changing environmental conditions and over the course of vegetation development (Amthor 2000, Makela and Valentine 2001, DeLucia et al. 2007).

Heterotrophic respiration is computed as the sum of variable decomposition and respiration rates from 3 distinct carbon pools as:

$$R_h = (K_{met} * C_{met} + [1-F_{str}] * K_{str} * C_{str} + K_{rec} * C_{rec}) \quad (6)$$

where  $C_{met}$ ,  $C_{str}$  and  $C_{rec}$  ( $g\ C\ m^{-2}$ ) represent metabolic, structural, and recalcitrant SOC pools, and  $K_{met}$ ,  $K_{str}$  and  $K_{rec}$  ( $d^{-1}$ ) are the corresponding decomposition rate parameters. The metabolic and structural SOC pools represent plant litter with relatively short (e.g.,  $\leq 5$  years) turnover periods, while the recalcitrant pool represents more physically and chemically protected SOC with longer turnover time.

The three-pool soil decomposition model approximates the complex variation of intrinsic SOC turnover rates, but has been found to produce results consistent with a wide range of observations from soil warming and incubation experiments (Knorr et al. 2005). Litter inputs to the  $C_{met}$  and  $C_{str}$  pools in Eqn. (6) are derived as proportions of estimated NPP, while input to  $C_{rec}$  is defined as a constant proportion ( $F_{str}$ ) of decomposed detritus from the  $C_{str}$  pool (Ise and Moorcroft 2006); outputs to the SOC pools represent daily sums of the respired components:

$$dC_{met}/dt = C_{fract} * NPP - R_{h,met} \quad (7)$$

$$dC_{str}/dt = (1 - C_{fract}) * NPP - R_{h,str} \quad (8)$$

$$dC_{rec}/dt = F_{str} * R_{h,str} - R_{h,rec} \quad (9)$$

where NPP is estimated as a fixed proportion of GPP ( $g\ C\ m^{-2}\ d^{-1}$ ) for individual land cover classes based on the assumption that vegetation carbon use efficiency is conserved (i.e., NPP/GPP is constant) within individual plant functional types (Waring et al. 1998, Gifford 2003, Litton et al. 2007). The  $C_{fract}$  term defines the rate in which litterfall from NPP is allocated to metabolic and structural SOC pools, and is specified as a fixed rate within individual PFT classes (Potter et al. 1993, Ise and Moorcroft 2006). Values for  $C_{fract}$ , and proportional allocations of GPP to  $R_a$  and NPP are defined in the BPLUT for individual PFT types (**Table 5**). This approach is based on the assumption that the litter input to the SOC pool is proportional to NPP under long-term, steady state conditions (Parton et al. 1987, Ise and Moorcroft 2006).

A dynamic litterfall scheme is employed for daily allocation of annual NPP to metabolic and structural SOC pools ( $C_{met}$  and  $C_{str}$ ). For evergreen forests (including needle-leaf and broadleaf types), the litter fraction of annual NPP is evenly distributed over each annual cycle; a simple phenology model based on either surface meteorological controls (e.g. Jolly et al. 2005) or a VI climatology may also be used to distribute litterfall according to seasonal phenology changes for deciduous and non-forest biome types (Randerson et al. 1996; White et al. 2000).

The L4\_C algorithms employ dimensionless rate curves to account for soil temperature and moisture constraints to soil decomposition. The soil decomposition rate ( $K$ ) is derived as the product of dimensionless multipliers for soil temperature ( $T_{\text{mult}}$ ) and moisture ( $W_{\text{mult}}$ ) and a theoretical optimum or maximum rate constant ( $K_{\text{mx}}$ ;  $\text{d}^{-1}$ ) under prevailing climate conditions:

$$K_{\text{met}} = K_{\text{mx}} * T_{\text{mult}} * W_{\text{mult}} \quad (10)$$

where  $T_{\text{mult}}$  and  $W_{\text{mult}}$  vary between 0 (fully constrained) and 1 (no constraint). The value for  $K_{\text{mx}}$  is specified as a constant within individual biomes, while decomposition rate parameters for  $K_{\text{str}}$  and  $K_{\text{rec}}$  are estimated as 40 % and 1 % of  $K_{\text{met}}$ , respectively (Ise and Moorcroft 2006). The estimation of  $K$  assumes constant soil decomposer efficiency (microbial  $\text{CO}_2$  production to carbon assimilation ratio) inherent in the  $K_{\text{mx}}$  term, and that soil moisture and temperature are the dominant controls on near-term (daily, seasonal, and annual) decomposition rates. However, we assume that changes in litter quality (e.g., physical protection and/or chemical resistance to microbial decomposition) influence  $R_h$  and NEE indirectly through associated changes in satellite optical-IR remote sensing derived vegetation indices (e.g. NDVI, EVI) and associated FPAR inputs, especially over generally N-limited boreal and tundra ecosystems.

The soil decomposition rate response to temperature is defined using an Arrhenius type function (Lloyd and Taylor 1994):

$$T_{\text{mult}} = \exp [308.56 * ((46.02 + T_{\text{opt}})^{-1} - (T_s + 46.02)^{-1})] \quad (11)$$

where  $T_{\text{opt}}$  and  $T_s$  are the respective reference and input daily surface soil temperatures ( $^{\circ}\text{C}$ ) for  $T_s \leq T_{\text{opt}}$ . The  $T_{\text{opt}}$  term defines the optimum temperature for soil decomposition and is prescribed as a biome-specific constant (Mahecha et al. 2010) in the BPLUT. The above relationship defines a low temperature constraint to soil decomposition;  $T_{\text{mult}}$  is assumed to be unity and soil decomposition no longer temperature limited for soil temperatures above  $T_{\text{opt}}$ . Under these conditions soil moisture is expected to decline with warmer soil temperatures and  $W_{\text{mult}}$  becomes the primary constraint to  $K_{\text{met}}$ . This assumption is generally valid for most global biome types, including temperate, boreal and Arctic ecosystems (Kimball et al. 2009a, Yi et al. 2013), but may not hold for warm and humid climate zones including tropical biomes (Jones et al 2003). A variety of functional types have been used to describe temperature effects on soil respiration including exponential (Oberbauer et al. 1992, Mikan et al. 2002) and Poisson (Parton et al. 1987, Ise and Moorcroft 2006) functions, while the Arrhenius functional type is physically based and provides a relatively accurate and unbiased estimate of soil respiration across a wide range of biomes and environmental conditions (Lloyd and Taylor 1994, Fang and Moncrieff 2001, Knorr et al. 2005, Yvon-Durocher et al. 2012).

The soil decomposition rate response to soil moisture (SM) has often been described using a parabolic function (Davidson et al. 1998, Davidson et al. 2000). This approach is suitable for a wide range of soils and accounts for the inhibitory effect of dry and saturated soil moisture conditions on aerobic decomposition and respiration (Doran et al. 1990). Although a parabolic-type response has been observed from both site observations and laboratory incubation

experiments (Ino and Monsi 1969, Bunnell and Tait 1974, Oberbauer et al. 1991, 1996), it may not adequately represent ecosystems adapted to wet or inundated soils or conditions where the duration of soil saturation is not sufficient to deplete soil oxygen levels to the point of restricting decomposition and CO<sub>2</sub> respiration (Altor and Mitsch 2008). Many studies also show evidence of sustained respiration and CO<sub>2</sub> emissions occurring under soil saturation (Chimner 2004, Lafleur et al. 2005, Kutzbach et al. 2007, Mäkiranta et al. 2009), which may be facilitated by oxygen transport through plant root systems (Colmer and Greenway 2005, Elberling et al. 2011).

In accordance with previous studies, the SM constraints on soil decomposition for unsaturated ( $SM \leq SM_{opt}$ ) conditions can be defined as:

$$W_{mult} = [1 + a * \text{EXP}(b * SM_{opt})] / [1 + a * \text{EXP}(b * SM)] \quad (12)$$

where SM is expressed as a proportion (%) of soil saturation;  $SM_{opt}$  is the optimum soil moisture level for heterotrophic decomposition (Doran et al. 1990) and is prescribed for different land cover types (**Table 5**). The  $a$  and  $b$  terms are empirical fitting parameters (dimensionless) that define the decomposition rate response to soil moisture variability and are specified for different land cover types in the BPLUT. The  $W_{mult}$  term is assumed to be unity (no restriction) for  $SM > SM_{opt}$ , which accounts for ecosystem adaptations to wet soil conditions and a general lack of landscape level observational evidence for extended SM saturation and associated reductions in aerobic decomposition (Chimner 2004, Reddy and DeLaune 2008, Elberling et al. 2011). The above algorithm is also based on the assumption that mean surface soil properties are similar within individual land cover classes and grid cells defined at the scale of relatively coarse (~1-40 km) global satellite footprints and modeling grids (Kimball et al. 2009a, Jones et al. 2007). The soil moisture response curve represented in Eqn (12) may be further refined to facilitate global model BPLUT calibration, including use of a simplified linear ramp function ranging between PFT specified minimum and maximum soil moisture levels, calibrated by minimizing RMSE differences between model estimated and in situ tower observed CO<sub>2</sub> flux observations (e.g., see Section 3.7).

**Table 5.** General Biome Properties Look-up Table (BPLUT) of primary ecophysiological parameters used for the L4\_C model calculations.

**LUE based GPP calculation**

Parameter	Units	Description
$\epsilon_{\text{mx}}$	(g C MJ <sup>-1</sup> )	Maximum PAR conversion efficiency
Min <sub>T<sub>mn</sub></sub>	(°C)	The daily minimum air temperature (T <sub>mn</sub> ) at which $\epsilon = 0.0$ (for non-limiting VPD, FT & SM <sub>rz</sub> conditions)
Max <sub>T<sub>mn</sub></sub>	(°C)	The daily T <sub>mn</sub> at which $\epsilon = \epsilon_{\text{mx}}$ (for non-limiting VPD, FT & SM <sub>rz</sub> conditions)
Min <sub>VPD</sub>	(Pa)	The daylight average vapor pressure deficit (VPD) at which $\epsilon = \epsilon_{\text{mx}}$ (for non-limiting T <sub>mn</sub> , FT & SM <sub>rz</sub> conditions)
Max <sub>VPD</sub>	(Pa)	The daylight average VPD at which $\epsilon = 0.0$ (for non-limiting T <sub>mn</sub> , FT & SM <sub>rz</sub> conditions)
Min <sub>SM<sub>rz</sub></sub>	(% Sat.)	The daily mean root zone (0-1 m depth) SM (SM <sub>rz</sub> ) level at which $\epsilon = 0.0$ (for non-limiting VPD, FT & T <sub>mn</sub> conditions)
Max <sub>SM<sub>rz</sub></sub>	(% Sat.)	The daily mean surface or SM <sub>rz</sub> level at which $\epsilon = \epsilon_{\text{mx}}$ (for non-limiting VPD, FT & T <sub>mn</sub> conditions)
F <sub>FT</sub>	(DIM)	The FT classified daily frozen (F) landscape status where $\epsilon_{\text{mult}} = 0.5$ (for non-limiting VPD, T <sub>mn</sub> & SM <sub>rz</sub> conditions)
NF <sub>FT</sub>	(DIM)	The FT classified daily nonfrozen (NF) landscape status where $\epsilon = \epsilon_{\text{mx}}$ (for non-limiting VPD, T <sub>mn</sub> & SM <sub>rz</sub> conditions)

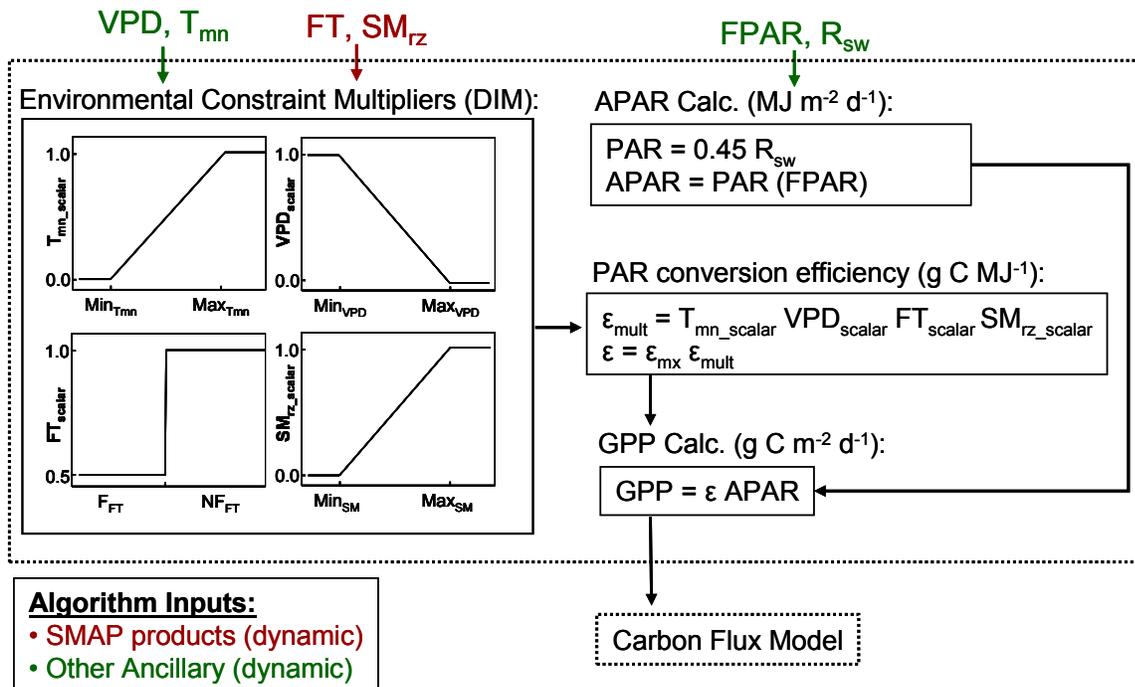
**Soil decomposition & respiration calculation**

Parameter	Units	Description
C <sub>fract</sub>	(DIM)	Proportion of NPP allocated to litterfall
CUE	(DIM)	Carbon use efficiency (NPP:GPP)
R <sub>a</sub> :GPP	(DIM)	Proportion of GPP allocated to autotrophic respiration (R <sub>a</sub> ); $\approx 1 - \text{CUE}$ .
K <sub>mx</sub>	(d <sup>-1</sup> )	Optimum soil decomposition rate
K <sub>str</sub> :K <sub>met</sub>	(%)	Ratio of structural to metabolic SOC decomposition
K <sub>rec</sub> :K <sub>met</sub>	(%)	Ratio of recalcitrant to metabolic SOC decomposition
F <sub>str</sub>	(DIM)	Proportion of structural SOC litter allocated to recalcitrant SOC pool
T <sub>opt</sub>	(°C)	Optimum soil temperature for SOC decomposition rate response (T <sub>mult</sub> ) calculation
Max <sub>sm</sub>	(% Sat.)	Maximum or optimum soil moisture level for the SOC decomposition rate response (W <sub>mult</sub> ) calculation; this is equivalent to SM <sub>opt</sub> .
Min <sub>sm</sub>	(% Sat.)	Theoretical minimum soil moisture level, where estimated W <sub>mult</sub> = 0 (i.e. full soil moisture limitation so SOC decomposition and R <sub>h</sub> )

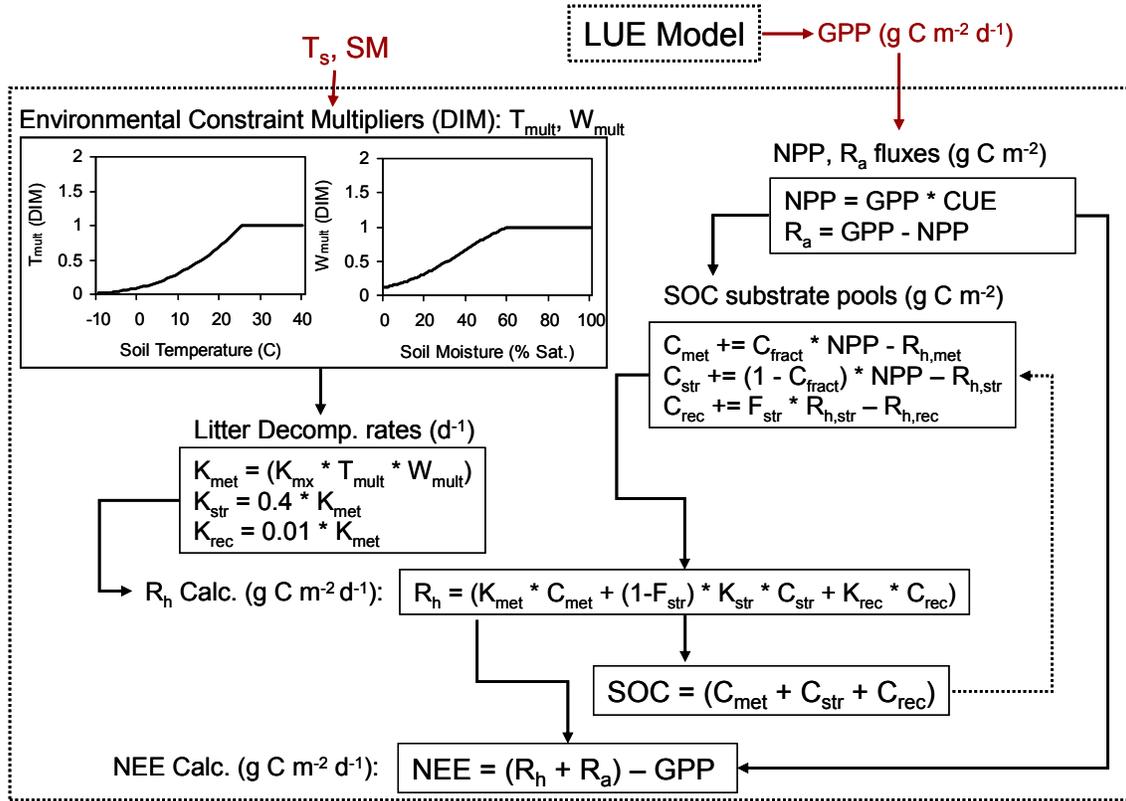
**3.1.2 Algorithm Baseline**

The L4\_C algorithm is composed of light use efficiency (LUE) and terrestrial carbon flux model components that are used to estimate GPP, autotrophic and heterotrophic respiration, residual NEE carbon fluxes, and underlying SOC pools on a daily basis. NEE represents the primary validated L4\_C output product, while the accompanying GPP, respiration and SOC

outputs will be included for enhanced validation and research activities. The baseline L4\_C algorithm is summarized in **Figure 4 (a, b)** for respective LUE and carbon flux model components. The algorithm approach has structural elements similar to the Century (Parton et al. 1987, Ise and Moorcroft 2006) and CASA (Potter et al. 1993) soil decomposition models and operational MOD17 GPP algorithm (Zhao et al. 2005, Zhao and Running 2010), but is adapted for use with daily biophysical inputs derived from both global satellite and model reanalysis data (Kimball et al. 2009, Yi et al. 2013). The current L4\_C algorithm baseline was developed from earlier versions and pre-launch algorithm development and testing, and incorporates recommendations from external L4\_C algorithm reviews (e.g. Kimball et al. 2009b).



**Figure 4a.** Baseline L4\_C LUE model structure for estimating GPP. Arrows denote the primary pathways of data flow, while boxes denote the major process calculations. Primary inputs include daily root zone soil moisture ( $SM_{rz}$ ) and landscape freeze/thaw (FT) status from SMAP L4\_SM and L3\_SM\_A products (in red), and other dynamic ancillary inputs (in green) including MODIS (MOD/MYD15) FPAR and reanalysis (GMAO) daily surface meteorology, including vapor pressure deficit (VPD), minimum air temperature ( $T_{mn}$ ) and incident solar shortwave radiation ( $R_{sw}$ ). Model calculations are performed at 1-km spatial resolution using dominant vegetation class and BPLUT response characteristics for each grid cell defined from a global land cover classification. The resulting GPP calculation is a primary input to the L4\_C terrestrial carbon flux model below (**Fig 4b**).



**Figure 4b.** L4\_C terrestrial carbon flux model for estimating NEE. Primary algorithm inputs (in red) include daily GPP from the LUE model (Fig 4a), and surface soil moisture (SM) and temperature (T<sub>s</sub>) from the SMAP L4\_SM product. NEE is the primary (validated) output, while GPP, respiration (R<sub>h</sub>+R<sub>a</sub>) and SOC are secondary (research) outputs.

Dynamic (daily) inputs to the L4\_C algorithms include satellite optical-IR remote sensing (MODIS) based FPAR, GMAO reanalysis surface meteorology (R<sub>sw</sub>, T<sub>mn</sub>, VPD) and associated L4\_SM based soil moisture (SM<sub>tz</sub>) that provide primary inputs to the LUE algorithm to determine GPP and NPP; the dynamic inputs also include L3\_SM\_A defined frozen temperature constraints to productivity and autotrophic respiration calculations. SMAP L4\_SM based surface (≤ 5 cm depth) soil moisture and soil temperature are used as primary drivers of the soil decomposition and R<sub>h</sub> calculations. Static inputs to the L4\_C algorithms include a global land cover (PFT) classification, which is used to define the major plant functional types and associated biome specific (BPLUT) response characteristics for each vegetated grid cell within the product domain. The land cover classification used for L4\_C processing will be consistent with the one used in the production of the FPAR inputs. All model inputs are available as satellite remote sensing derived products or from model reanalysis.

The L4\_C domain will encompass all global vegetated land areas. The global domain enables comprehensive determination of carbon fluxes, underlying SOC stocks and primary environmental drivers over all global biomes, and their cumulative impact on global terrestrial NEE source/sink activity. The global domain also increases the number and diversity of in situ tower observation sites for algorithm calibration and validation, enabling potentially improved algorithm accuracy and product utility (Running et al. 1999, Baldocchi 2008). While the L4\_C

product will be global in extent, product accuracy requirements and validation activities are primarily specified for northern ( $\geq 45^\circ\text{N}$ ) land areas consistent with NRC objectives for better understanding of terrestrial carbon source/sink activity in boreal regions (NRC 2007, Jackson et al. 2012).

The L4\_C calculations will be conducted at a daily time step that provides the necessary precision for resolving dynamic boreal vegetation phenology and carbon cycles (Kimball et al. 2009a, Kim et al. 2012). The L4\_C calculations will be conducted at a 1-km spatial scale consistent with the resolution of ancillary FPAR and land cover classification inputs. The simulations will also be conducted in a consistent global EASE-grid (version 2) projection format. Model simulations for each 1-km grid cell will be conducted using the corresponding (nearest-neighbor) 9-km resolution L4\_SM and GMAO (FP) inputs and 3-km resolution L3\_SM\_A inputs. The MODIS (MOD/MYD15) FPAR product is produced at 1-km spatial resolution and 8-day temporal fidelity from both NASA EOS Terra and Aqua sensor records, while a similar global product is planned for JPSS VIIRS (NGST 2011). The MODIS FPAR operational products use a tile based sinusoidal projection; preprocessing of these data prior to the L4\_C calculations will involve reprojecting from sinusoidal to 1-km resolution global cylindrical EASE-grid projection formats. The FPAR data will be screened and only best quality (QC) FPAR data used as L4\_C inputs. Missing or unfavorable QC 8-day FPAR data will be gap filled on a grid cell-wise basis using an ancillary FPAR mean 8-day climatology established from the long-term (10+ year) best QC MODIS FPAR record. The resulting 8-day FPAR data will be sub-sampled at the daily processing time step and combined with other daily biophysical inputs from GMAO and SMAP L4\_SM, and L3\_SM\_A inputs to estimate GPP and NPP, autotrophic and heterotrophic respiration, NEE and surface SOC pools. The model calculations and L4\_C outputs will also include daily environmental constraint (EC) indices influencing the GPP and NEE calculations, including the estimated bulk environmental reduction to PAR conversion efficiency ( $\epsilon_{\text{mult}}$ ), and the soil moisture and temperature constraints ( $W_{\text{mult}}$ ,  $T_{\text{mult}}$ ) to soil decomposition and  $R_h$  calculations. The resulting L4\_C variables will enable characterization of spatial patterns and daily temporal fidelity in NEE, underlying carbon fluxes and SOC pools, and the primary environmental drivers of these process calculations.

The resulting fine scale (1-km resolution) L4\_C outputs will be spatially aggregated to the coarser 9-km resolution grid of the final product by weighted linear averaging of outputs according to the fractional cover of individual vegetation classes represented within each 9-km grid cell and defined by the underlying 1-km resolution ancillary land cover map and FPAR climatology; the sub-grid scale means and SD variability from individual vegetation (PFT) classes will be preserved for each 9-km grid cell, while proportional vegetation cover information will be included in the product metadata, allowing the coarse resolution data to be decomposed into the relative contributions from individual PFT classes within each cell. An additional EC constraint will be defined as the proportional (%) frozen area (FA) within each 9-km grid cell and defined from finer (3-km) resolution SMAP L3\_SM\_A inputs. These outputs will facilitate improved algorithm and product accuracy over heterogeneous land cover areas, and product outputs that are more consistent with the mean sampling footprint of most tower CO<sub>2</sub> flux measurement sites (Baldocchi 2008, Chen et al. 2012).

### 3.1.3 Algorithm Options

Two primary algorithm options are being considered that affect the L4\_C design relative to the baseline algorithm. The options are summarized in **Table 6** and include: (1) use of lower order satellite optical-IR remote sensing based spectral vegetation index (VI) inputs to estimate FPAR for the LUE model based GPP calculations; (2) use of ancillary vegetation disturbance (fire) and recovery status inputs to perturb model steady state conditions and associated SOC and carbon flux calculations. Option (1) allows for estimation of FPAR inputs using more readily available VI data if an operational FPAR product stream becomes unavailable during the SMAP mission period; use of alternative VI inputs from other operational satellite sensors (e.g. VIIRS) potentially enhances reliability of algorithm inputs, but may degrade product latency. Option (2) is expected to enhance scientific merit and L4\_C product accuracy relative to the baseline, but at the expense of increased algorithm complexity; model accuracy and uncertainty may also be degraded where available ground observations (e.g. tower age chronosequence networks) needed for model disturbance recovery calibration are lacking.

**Table 6.** L4\_C algorithm options.

<b>Option</b>	<b>Description</b>
(1)	L4_C based FPAR and GPP calculation using lower order VI inputs
(2)	Representation of fire disturbance and recovery effects on SOC and carbon flux calculations

#### 3.1.3.1 GPP Calculation Using Ancillary VI Inputs

The baseline L4\_C algorithm relies on external FPAR inputs from the MODIS (MOD/MYD15) operational product stream available from the NASA EOS Terra and Aqua platforms. The MODIS FPAR product is well characterized (e.g. Yang et al. 2006, Masuoka et al. 2011) and has been providing regular global 8-day repeat FPAR records at 1-km spatial resolution from Terra and Aqua since 2000 and 2002, respectively. The MOD15 FPAR product is also a primary input to the MOD17 operational GPP product (Running et al. 2004). Both EOS MODIS sensors are currently in extended mission phases, and are projected to extend through 2017 (Hurt et al. 2011). Failure of a single MODIS sensor would be mitigated by adjusting L4\_C processing to ingest FPAR inputs from the remaining operational MODIS sensor, with no effect on L4\_C baseline processing or product latency. However, failure of both MODIS sensors would result in potential loss of a reliable FPAR input for SMAP baseline L4\_C production. There is currently no alternative operational FPAR product available from any other sensor, including the current set of environmental data records (EDRs) from NASA NPP VIIRS; a similar FPAR product is planned for JPSS VIIRS (NGST 2011), which has a projected launch in 2017, well beyond the initiation of SMAP operations.

Option (1) would involve deriving FPAR within the L4\_C algorithm framework using lower order VI inputs such as the Normalized Difference Vegetation Index (NDVI) or Enhanced Vegetation Index (EVI), which are readily available from other operational satellite sensors, including NASA NPP VIIRS. The best quality top-of-canopy VI retrievals for each grid cell would be used for empirical estimation of daily FPAR using biome (BPLUT) specific coefficients developed between the respective best QC sensor VI records and the long-term (best

QC) MODIS FPAR record. The resulting FPAR estimates would be temporally interpolated or sub-sampled at a daily time step for input into the L4\_C LUE model. FPAR for missing or low quality VI retrievals would be obtained directly from the ancillary MODIS FPAR climatology.

Potential benefits of the VI based FPAR calculation include reduced algorithm reliance on operational FPAR products that may not be available during the SMAP mission period, relative to alternative VI inputs that are more readily available from multiple satellite sources. Potential disadvantages of this option include increased L4\_C algorithm complexity and computational processing costs associated with the additional step of deriving an intermediate FPAR product prior to the LUE model based GPP calculation. Current global VI records from MODIS are available at coarser 16-day temporal fidelity (relative to 8-day MODIS FPAR fidelity), which would decrease L4\_C product mean latency from 14-days (baseline) to approximately 26-days; a degraded product latency could reduce potential science application utility. Employing a simple empirical function to derive FPAR from lower order VI inputs may result in reduced GPP accuracy for some (e.g. higher biomass) areas due to non-linear effects or potential saturation of the VI signal; however, a global investigation of these effects using MODIS VI (MOD13A2) product inputs showed no significant reduction in GPP accuracy (Kimball et al. 2011, Yi et al. 2013).

### 3.1.3.2 Disturbance and Recovery Effects

The baseline L4\_C algorithms assume dynamic equilibrium between vegetation productivity (NPP) and surface SOC and respiration under prevailing climate conditions. This steady-state assumption has been successfully applied at global and regional scales and allows relatively straight-forward model initialization of surface SOC pools using recent (~10 yr) satellite FPAR records and model reanalysis daily surface meteorology (e.g. Ise and Moorcroft 2006, Potter et al. 1993). Initial tests of the L4\_C algorithms over northern ecosystems indicate that the steady-state assumption yields NEE results within L4\_C accuracy requirements (Kimball et al. 2009a). However, the global network of tower eddy covariance based CO<sub>2</sub> measurements indicate that most ecosystems are affected by disturbance and stand succession processes that create an imbalance between vegetation biomass and photosynthetic carbon uptake (GPP), and SOC storage, decomposition and carbon loss through R<sub>tot</sub>; these measurements also indicate that the degree of offset between GPP and R<sub>tot</sub> is similar across a broad global range of recently disturbed sites under varying succession stages (Baldocchi 2008). In boreal biomes the dominant disturbance mechanism is from fire, which may be increasing in frequency and severity with regional warming (Kasischke et al. 1995, McGuire et al. 2009).

The L4\_C algorithm option (2) would represent fire disturbance and recovery status using ancillary satellite derived disturbance maps to identify the timing and spatial extent of fire disturbance and associated vegetation recovery stage on a grid cell-by-cell basis within the domain. This information would be used with existing in situ tower eddy covariance CO<sub>2</sub> measurement records to define empirical corrections to the L4\_C dynamic steady-state assumption between NPP and surface SOC (Baldocchi 2008, Carvalhais et al. 2010). Potential sources of ancillary burned area information include the currently operational MODIS MCD45A1 burned area fire product, which is provided globally at a monthly time step and extends from 2000 onward (Giglio et al. 2006). Other global burned area products are available from ensemble satellite fire products (Giglio et al. 2010) or are being planned for VIIRS on NPP and JPSS. A disturbance recovery status map would be derived from existing global burned area and disturbance history records, and used as static ancillary information for empirical adjustment

of GPP and SOC pools in the L4\_C algorithm. Disturbance recovery status and associated impacts to GPP and SOC pools would be defined from previous studies of satellite VI variability over tower fire chronosequence sites (e.g. Goulden et al. 2011, Yi et al. 2013). The ancillary disturbance recovery map could remain static or be periodically updated to account for disturbance and recovery during the active mission period. Initial implementation and testing of the L4\_C disturbance recovery option over a northern ( $\geq 45^\circ\text{N}$ ) domain indicated accuracy improvement in model estimated carbon fluxes, over baseline steady-state model simulations, relative to sparse tower observations for recently disturbed boreal forest areas, though the relative impact of these changes on NEE was small compared to climate variability impacts on regional carbon fluxes (Yi et al. 2013). The potential benefits of option (2) include improved product accuracy and science utility, while drawbacks include increased data throughput, storage and computational costs, and increased algorithm complexity over baseline conditions; other limitations include a general lack of tower chronosequence site network observations spanning a representative global range of affected areas, which constrains disturbance model calibration and validation activities.

### 3.2 Ancillary Data Requirements

Anticipated ancillary inputs to the L4\_C algorithms are summarized in **Table 7**. The SMAP L4\_C baseline algorithms require dynamic daily FPAR inputs temporally sub-sampled from coarser (8-day) fidelity MODIS (MOD15 or MYD15) time series records; FPAR is used with global reanalysis (GMAO) daily surface meteorology ( $R_{sw}$ ,  $T_{mn}$ , VPD) and SMAP based FT and  $SM_{rz}$  inputs within the LUE model to estimate GPP. The FT inputs will be obtained from the SMAP L3\_SM\_A product, which includes a FT defined frozen flag derived from the SMAP AM orbit SAR FT classification and spanning the global domain, and with similar global EASE grid projection format. Alternatively, the SMAP L3\_FT\_A product provides a consistent FT classification, with enhanced temporal (AM and PM) FT fidelity, but is limited to a smaller northern ( $\geq 45^\circ\text{N}$ ) domain and polar EASE-grid projection format. The SMAP L4\_SM product is used to obtain daily  $SM_{rz}$  inputs for the L4\_C model GPP calculation; L4\_SM based daily surface (<10 cm depth) soil temperature and soil moisture inputs ( $T_s$ , SM) will be used within the soil decomposition and respiration model to compute ecosystem respiration and SOC, while NEE will be derived as a daily residual difference between GPP and  $R_{tot}$ .

Preprocessing of the MODIS FPAR inputs will involve spatial reprojection of individual MODIS tiles in the 1 km resolution sinusoidal projection format to the 1 km resolution global EASE-grid (version 2) projection format of the L4\_C product. Only best quality (QC) FPAR 8-day retrievals will be used for interpolation and estimation of the daily FPAR inputs; missing or low QC MODIS FPAR data for each 1 km resolution grid cell will be temporally gap filled using an ancillary global 8-day FPAR climatology estimated during the SMAP pre-launch phase from the long-term MODIS MOD/MYD15 FPAR record. The ancillary FPAR climatology will include the long-term mean and SD temporal variability for each 1 km resolution grid cell and 8-day time step established from the 10+ year MODIS best QC FPAR record. Model calculations using the FPAR climatology will be assigned a lower quality ranking for the product outputs. The MODIS FPAR climatology will also be used with the MODIS land cover classification to identify vegetated (PFT) pixels with sufficient minimum FPAR levels that define the global L4\_C potential modeling domain.

The SMAP L3\_SM\_A, L4\_SM and GMAO daily inputs for the L4\_C calculations will be provided in a consistent global EASE-grid projection format, but at coarser (3 km and 9 km

resolution) spatial scales. The 3-day repeat L3\_SM\_A based FT records will be subsampled to a daily time step using a nearest neighbor selection of the closest adjacent valid FT value; in cases of missing or low quality L3\_SM\_A freeze/thaw retrievals an alternative L3\_SM\_A frozen flag determined from reanalysis (GMAO) surface temperatures ( $T_{surf}$ ) will be used to define freeze/thaw inputs to the L4\_C algorithms; FT and  $T_{surf}$  defined frozen flags are both represented in the L3\_SM\_A global product. The alternative  $T_{surf}$  based freeze/thaw input may also be used during the initial post-launch period until reliable SMAP freeze/thaw seasonal reference state conditions are established and stability and accuracy of the associated L3\_SM\_A freeze/thaw flag is verified. The freeze/thaw quality control flags from the lower order L3\_SM\_A inputs will be represented in the L4\_C product quality (QC) bit flags and used to identify reliable 3-km resolution cells for determining proportional frozen area (FA) EC metrics and GPP within each 9-km resolution L4\_C grid cell. The finer (1 km) resolution L4\_C processing will involve spatial nearest neighbor selection of overlying coarser resolution grid cells of these model inputs.

**Table 7.** Anticipated primary ancillary inputs to the L4\_C algorithm.

<i>Parameter</i>	<i>Units</i>	<i>Type</i>	<i>Spatial Resolution</i>	<i>Potential Source</i>
FPAR	%	Dynamic (8-day)	1 km	MODIS (MOD15A2, MYD15A2), VIIRS (VVI3P)
$R_{sw}$	$MJ\ m^{-2}\ d^{-1}$	Dynamic (daily)	9 km	GMAO
$T_{mn}$	$^{\circ}C$	Dynamic (daily)	9 km	GMAO
VPD	Pa	Dynamic (daily)	9 km	GMAO
SM	% Sat.	Dynamic (daily)	9 km	SMAP L4_SM
$SM_{rz}$	% Sat.	Dynamic (daily)	9 km	SMAP L4_SM
$T_s$	$^{\circ}C$	Dynamic (daily)	9 km	SMAP L4_SM
FT	Discrete class	Dynamic (3-day)	3 km	SMAP L3_SM_A
Land Cover Class	Discrete class	Static	1 km	MODIS (MCD12Q1)
FPAR climatology	%	Static (8-day)	1 km	MODIS (MOD15A2, MYD15A2)
<b>Additional Inputs for Algorithm Options</b>				
VI (NDVI,	Dimensionless	Dynamic	1 km	MODIS

EVI)		(8-16 day)		(MOD/MYD13A2), VIIRS (VVI)
Recovery Status	Years	Static	1 km	MODIS (MCD45), GFED

Static inputs to the L4\_C baseline algorithms include a 1-km resolution global land cover classification consistent with the MODIS IGBP global land cover product (Friedl et al. 2010, 2011); the land cover classification will be aggregated into a reduced set of 8 global PFT classes that will be used with the BPLUT to define general physiological properties and environmental response characteristics of individual biomes. Care will be taken to ensure that the land cover classification used in the L4\_C calculations is consistent with the underlying land cover data used to derive the FPAR inputs. A static land cover mask will be applied to exclude open water bodies, permanent snow and ice, bare soil and other non-vegetated land areas from the model calculations. The MODIS FPAR climatology will also be used to mask PFT areas falling below a minimum FPAR level for successful GPP calculation. Quality assessment (QA) metrics included in the L4\_C product outputs will account for negative impacts on NEE RMSE from significant sub-grid scale terrain and land cover heterogeneity relative to the resolution of the overlying modeling window. These QA flags will be derived from model error sensitivity analyses (Section 3.3) and thresholds of elevation spatial variance and proportional coverage of discrete land cover classes within each 9 km resolution product grid cell and derived from the finer scale (1-km resolution) digital elevation (DEM) and land cover classifications. Additional negative impacts from lower quality (QC) flags of the model inputs will be carried through from ancillary FPAR, L4\_SM and L3\_SM\_A inputs and used for dynamic daily adjustment of the L4\_C EC and QA outputs.

Periodic updating of ancillary land cover information within the L4\_C framework is possible using available dynamic global land cover classification products such as the MODIS MCD12Q1 product (Friedl et al. 2011). These data provide a potential means for representing land cover changes during the SMAP mission period. Land cover and land use changes from direct and indirect human impacts exert a large influence on NEE over a global domain, especially over longer (decadal) time scales. The relative impact of dynamic and static land cover inputs on the L4\_C model calculations will be evaluated. These effects are expected to have less impact over a projected 3-year mission cycle and over sparsely populated northern land areas.

The L4\_C algorithms utilize general biophysical response characteristics defined from the BPLUT to estimate carbon fluxes and SOC stocks under variable vegetation and climate conditions. An initial BPLUT was developed for the L4\_C algorithms using general biophysical properties established from previous studies (e.g. Potter et al. 1993, White et al. 2000, Ise and Moorcroft 2006) and similar parameters defined for the MODIS MOD17 operational GPP algorithm (Heinsch et al. 2003, Zhao et al. 2005). The L4\_C BPLUT was further refined through regional and global comparisons and calibration of prototype L4\_C simulations driven by existing satellite (MODIS) and reanalysis (GMAO MERRA) inputs against global tower (FLUXNET) site network observations and soil inventory records (e.g. Yi et al. 2013). The resulting L4\_C BPLUT (APPENDIX) is being used for SMAP prelaunch simulations and global assessment of algorithm performance and uncertainty. The BPLUT may be recalibrated and updated during the SMAP operational period to refine algorithm performance and product

accuracy using SMAP operational inputs and information gained from post-launch calibration and validation activities (Jackson et al. 2012).

### 3.2.1 Impacts from Algorithm Options

The SMAP L4\_C algorithm options will require additional ancillary data requirements (e.g. **Table 7**). The optional use of lower order VI inputs to estimate GPP replaces the use of dynamic ancillary FPAR inputs from MODIS. A regional test of this option over a pan-Arctic and boreal domain using MODIS (MOD13A2) NDVI inputs showed favorable GPP and NEE results in relation to tower observation based carbon fluxes and algorithm performance within product accuracy requirements (Yi et al. 2013). The dynamic VI inputs could be obtained from VIIRS on NPP or JPSS platforms in the event of loss of an operational FPAR product stream. Lower order VI inputs would be used for estimating FPAR using land cover class (BPLUT) specific empirical relationships established between existing best quality NPP VIIRS or MODIS VI (MOD/MYD13) and FPAR (MOD/MYD15) records. The VI parameters for estimating FPAR include top-of-canopy NDVI or EVI. The NDVI is used for production of the MODIS FPAR product, while the EVI is also sensitive to FPAR and provides improved canopy sensitivity at higher vegetation biomass levels (Huete et al. 1997); however, the potential EVI advantage over NDVI would be offset by use of NDVI based MODIS FPAR products for developing empirical VI-FPAR relationships and the ancillary FPAR climatology inputs. FPAR values coinciding with missing or lower quality VI retrievals would be obtained directly from the ancillary FPAR 8-day climatology and assigned a lower QC ranking in the output product. The resulting FPAR series would be temporally interpolated to a daily time step on a grid cell-wise basis for estimation of GPP. A potential constraint of using dynamic VI time series from VIIRS or MODIS is a reduced 16-day VI product latency relative to current 8-day operational FPAR products. A reduced 16-day latency for the VI inputs could propagate to a longer (~26-day) mean data latency for the L4\_C product and negatively impact product science application utility. MODIS MOD13/15 VI/FPAR global products currently span more than 10 years of record and are relatively comprehensive, well calibrated and validated. Collection 5 reprocessed MODIS data are currently available, while Collection 6 data are forthcoming; these data would provide a standard from which other potentially available ancillary VI inputs (e.g. from VIIRS) would be calibrated against before operational use as ancillary inputs for L4\_C calculations.

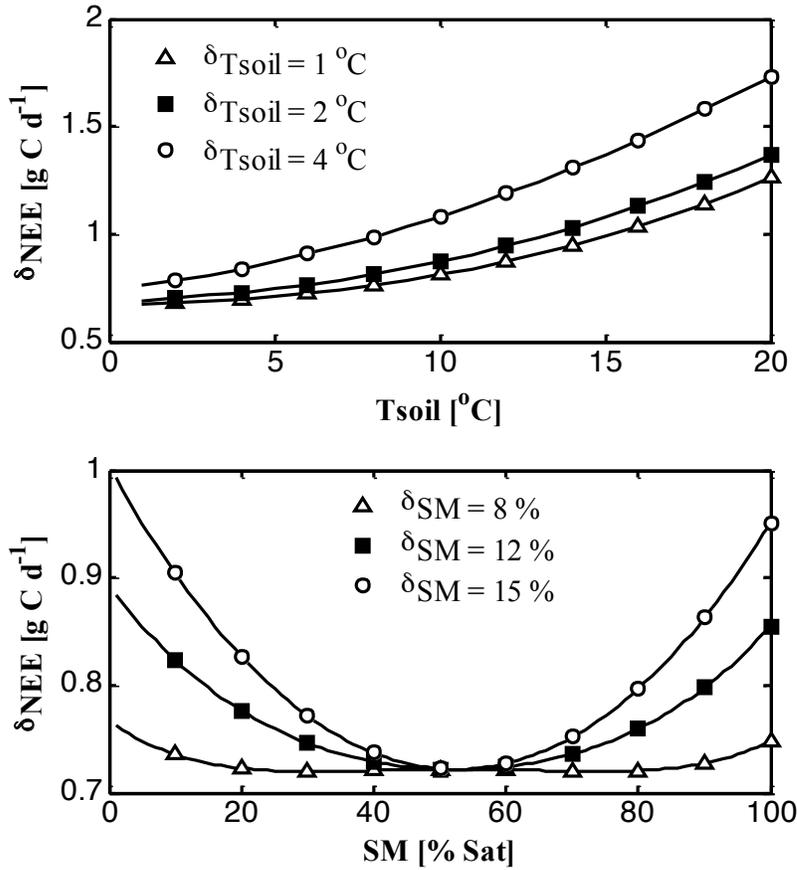
Additional ancillary inputs for the L4\_C algorithm disturbance option include geospatial data that identify disturbance recovery stage on a grid cell-wise basis. Recent (post 2000) disturbance recovery could be derived during the SMAP pre-launch period and implemented as a static input during SMAP mission operations. These static inputs could be updated periodically (though not required) under scheduled reprocessing activities and as additional data become available, and would define the approximate disturbance regime and recovery period (e.g. years) since the last major disturbance event as determined from the recent satellite record; currently available products that satisfy these requirements include the MODIS MCD45 operational disturbance product (Roy et al. 2008). An alternative Global Fire Emissions Database (GFED) is derived by merging multiple satellite fire products and provides a potential disturbance information source (Giglio et al. 2010), but would need to be resampled or downscaled from its native 0.5 degree format to a 1 km resolution for L4\_C processing. The ancillary recovery stage information would be used for prescribed empirical adjustment of model steady state derived SOC pools and NPP on a grid cell-wise basis based on the number of years since a major disturbance event and the affected land cover (PFT) class, and following previously developed methods (Yi et al. 2013).

Potential benefits of this option include improved L4\_C product accuracy and science utility, including boreal regions where fire disturbance has a major impact on terrestrial carbon sequestration and storage processes. Potential drawbacks of this option include difficulties in VI representation and model (BPLUT) parameterization of disturbance recovery effects on land-atmosphere carbon exchange. Tower CO<sub>2</sub> flux monitoring sites in recently disturbed areas or following regional recovery age chronosequences are critical for model development and parameterization of vegetation recovery effects (e.g. Yi et al. 2013), but are underrepresented globally (Baldocchi et al. 2008). Other drawbacks include increased processing and model complexity due to the additional ancillary inputs and the need to track recovery stage and associated impacts on SOC and NPP during mission operations.

### 3.3 Variance and Uncertainty Estimates

The L4\_C algorithm performance, including variance and uncertainty estimates of model outputs, will be determined during the mission pre-launch phase through model sensitivity studies using currently available model inputs similar to those that will be used during mission operations and evaluating the resulting model simulations over the observed range of northern ( $\geq 45^\circ\text{N}$ ) and global conditions. Model inputs being used for these studies include GMAO MERRA based soil temperature and soil moisture inputs (Yi et al. 2011, Reichle et al. 2011), satellite microwave remote sensing based freeze/thaw records (Colliander et al. 2012, Kim et al. 2012) and MODIS FPAR and GPP records (Kimball et al. 2011, Yi et al. 2013). These results indicate that the L4\_C accuracy requirements (i.e. NEE RMSE  $\leq 30 \text{ g C m}^{-2} \text{ yr}^{-1}$ ) can be met over more than 82% and 89% of global and northern vegetated land areas, respectively (e.g. **Figure 6**). These estimates will be refined following initiation of the SMAP operational data stream and associated L3\_SM\_A and L4\_SM production.

A previous L4\_C algorithm sensitivity simulation was conducted for a model evaluation study across a network of northern tower sites (Kimball et al. 2009a) and has since been extended to a global domain. The original study established an expected accuracy range in estimated carbon fluxes for varying  $T_s$  and SM uncertainty (RMSE) levels and over a characteristic range of northern vegetation and environmental conditions. The predicted NEE accuracy levels from this study are presented in **Figure 5**.



**Figure 5.** Estimated NEE RMSE uncertainty ( $\delta NEE$ ) under varying uncertainty levels for SM and  $T_s$  inputs ( $\delta T_s$ ,  $\delta SM$ ) and over a range of characteristic soil moisture and temperature conditions (Kimball et al. 2009a);  $\delta SM$  ranges from 8-12% ( $0.04$ - $0.075 \text{ m}^3 \text{ m}^{-3}$ ).

The sensitivity study assumed that errors are uncorrelated between  $T_s$  and SM inputs and uncorrelated through time. The sensitivity study also used a parabolic soil moisture response function that exaggerates model error under wetter soil conditions relative to the L4\_C algorithms. GPP inputs to the L4\_C algorithms were assumed to contribute a constant ( $1.2 \text{ g C m}^{-2} \text{ d}^{-1}$ ) representative error, derived as the mean RMSE difference between MODIS (MOD17) and tower GPP results over northern test sites (Heinsch et al. 2006). Other potential sources of model error including algorithm assumptions and land cover heterogeneity were not considered. For respective error levels in  $T_s$  and SM of  $2 \text{ }^\circ\text{C}$  and  $15 \%$  (of saturation), uncertainty in NEE ranges from  $0.72$ – $1.30 \text{ g C m}^{-2} \text{ d}^{-1}$  for  $T_s$  and SM conditions ranging from  $0$ - $20 \text{ }^\circ\text{C}$  and  $0$ - $100\%$ . These results translate into predicted uncertainties in annual NEE from  $13.75$ – $24.84 \text{ g C m}^{-2} \text{ yr}^{-1}$ .

The contributions of the different ancillary inputs to the total estimated NEE error vary over the characteristic range of  $T_s$  and SM conditions. The GPP inputs contribute most of the L4\_C estimation error for NEE when uncertainty in  $R_h$  is relatively small ( $< 0.64 \text{ g C m}^{-2} \text{ d}^{-1}$ ), which generally occurs when either  $T_s$  is low ( $< 10 \text{ }^\circ\text{C}$ ) or SM is near intermediate levels. The L4\_C sensitivity to SM uncertainty increases under drier SM levels and warmer  $T_s$  conditions. These results indicate that the L4\_C algorithm accuracy is sufficient to determine meaningful carbon

flux estimates over a broad range of  $T_s$  and SM conditions, including northern boreal forest, grassland and tundra biomes (Kimball et al. 2009a). The L4\_C algorithm performance was also confirmed over larger northern ( $\geq 45^\circ\text{N}$ ) and global domains (Kimball et al. 2011, Yi et al. 2013); these results indicate NEE accuracy within algorithm performance guidelines and relative to tower observation based carbon flux estimates. The L4\_C algorithm performance for NEE is also similar to other model approaches, including observation driven machine learning algorithms (Jung et al. 2011) and more detailed process model simulations (McGuire et al. 2012). These results provide an estimated range of model performance given expected uncertainty in MODIS derived GPP and satellite and reanalysis based  $T_s$  and SM inputs. Actual model error may be larger or smaller depending on correlations between model inputs, model or measurement bias, and potential error in model representation of biophysical processes.

An estimated error budget for the SMAP L4\_C product is summarized in **Table 8**. This table quantifies the expected primary error sources and individual and cumulative error contributions to the L4\_C based NEE estimates. This analysis extends the previous sensitivity study (**Figure 5**), but involves estimation of additional product uncertainty due to algorithm assumptions, land cover heterogeneity and expected accuracy levels for the GPP calculations and L4\_SM inputs. The NEE errors are expressed as daily and cumulative annual rates where the total expected uncertainty is a function of estimated GPP uncertainty, L4\_C algorithm structure and underlying assumptions, and expected uncertainty from the L4\_SM inputs. Uncertainties in L4\_C parameters and inputs were defined from the literature. A conservative estimate of model error from sub-grid scale land cover heterogeneity effects was determined by evaluating differences between model respiration calculations derived from 1-km and 25-km resolution land cover classification inputs. Each error source was individually propagated through the L4\_C algorithms to determine model sensitivity to each error source. Total algorithm uncertainty was then estimated by propagating all errors through the model together and assuming that errors were independent among sources and independent in time. Uncertainty is reported on an annual basis considering all known error sources. The total annual uncertainty ( $30 \text{ g C m}^{-2} \text{ y}^{-1}$ ) was estimated by summing daily uncertainties ( $1.6 \text{ g C m}^{-2} \text{ y}^{-1}$ ) and assuming independent errors. The results are in agreement with a previous L4\_C algorithm sensitivity study over northern tower sites (Kimball et al. 2009a) and similar Markov Chain Monte Carlo (MCMC) calibration studies using regional tower data (Richardson 2005, Richardson 2008, Yi et al. 2014); the resulting algorithm uncertainty is also comparable with reported daily tower  $\text{CO}_2$  measurement uncertainties (e.g. Baldocchi 2008).

The estimated soil temperature uncertainty in **Table 8** was determined from satellite (AMSR-E) microwave remote sensing and model reanalysis based studies of surface air and soil temperature retrievals relative to in situ measurements (Luo et al. 2003, Zhang et al. 2007, Dirmeyer 2006, Jones et al. 2007, Jones et al. 2010). Soil moisture uncertainty was defined from the SMAP L4\_SM ATBD and from satellite microwave remote sensing and land surface model derived soil moisture comparisons with in situ measurements (Luo et al. 2003, Berg 2005, Reichle et al. 2004, Ducharne et al. 2000, Dirmeyer 2006). The GPP uncertainty was determined through validation activities of the MODIS MOD17 GPP operational product (Heinsch et al. 2006, Zhao et al. 2006). Uncertainty associated with the autotrophic respiration fraction of GPP was determined from tower  $\text{CO}_2$  eddy flux measurement based estimates of carbon use efficiency for several temperate forest types (DeLucia et al. 2007). The L4\_C algorithms were initialized using parameter settings developed for northern biomes (Kimball et al. 2009a, Yi et al. 2013).

**Table 8.** Estimated total annual NEE error (RMSE) budget for the SMAP L4\_C product.

Type of Error	Error Source	Source Units	Range	Value	NEE Contribution (g m <sup>-2</sup> y <sup>-1</sup> )
Input Data	Temperature	°K	1.5-4	3.5	6.2
	Moisture	m <sup>3</sup> /m <sup>3</sup>	0.04-0.10	0.05	5.7
	GPP	g m <sup>-2</sup> d <sup>-1</sup>	1.0-2.0	1.5	14.5
Model Parameterization	Optimal Decomp. Rates/Response Curves	d <sup>-1</sup>	0.001-0.01	0.005	2.9
	Pool Representation/Steady State	g m <sup>-2</sup>	100-1000	1000	9.6
	Autotrophic Respiration fraction	dim	0.05-0.15	0.1	2.7
Heterogeneity	Land Cover Heterogeneity (Soil Respiration)	g m <sup>-2</sup> yr <sup>-1</sup>	0-60	16	22.6
Total NEE Error	Inputs (Soil Moist./Temp.) Only	g m <sup>-2</sup> yr <sup>-1</sup>			8.4
	Inputs (All) Only	g m <sup>-2</sup> yr <sup>-1</sup>			16.7
	Model Only	g m <sup>-2</sup> yr <sup>-1</sup>			10.4
	Inputs + Model	g m <sup>-2</sup> yr <sup>-1</sup>			19.7
	Inputs + Model + Het.	g m <sup>-2</sup> yr <sup>-1</sup>			30.0

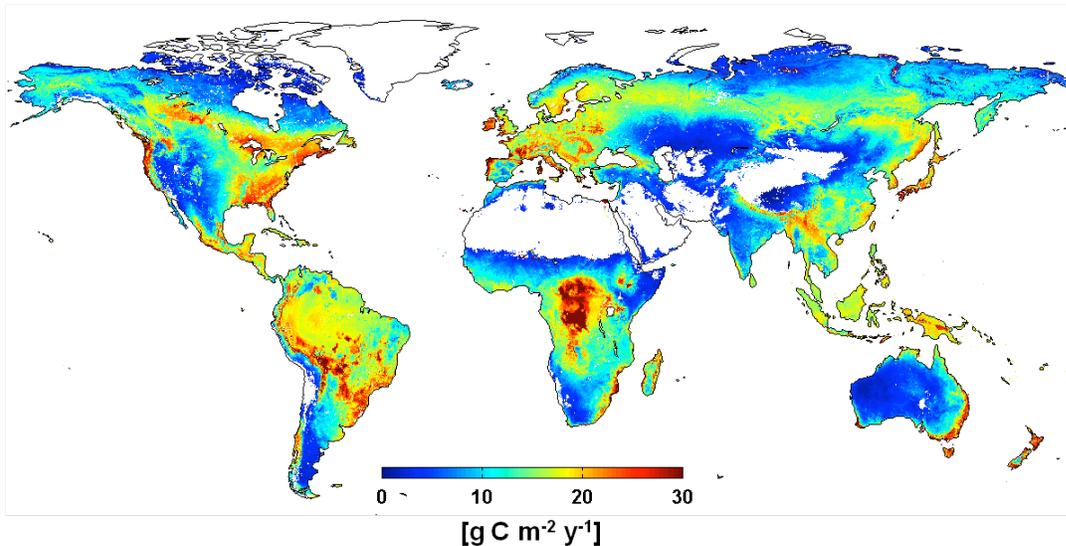
Sub-grid scale land cover heterogeneity contributions to L4\_C based NEE error were considered as the difference between a weighted average of individual land cover types and the dominant land cover type within each grid cell. The dominant land cover type and proportional (%) area representation of individual land cover classes within each coarse resolution grid cell was defined from the IGBP MODIS 1 km resolution land cover map over a northern ( $\geq 45^\circ\text{N}$ ) domain. Mean  $R_h$  and NPP values were assigned to each grid cell from steady-state simulations using daily AMSR-E (temperature and soil moisture) and MODIS derived GPP inputs for the 2003-2006 period. Simulations were solved for steady state conditions for each grid cell using the BPLUT parameters (APPENDIX) according to the local dominant land cover class, and then as a weighted average of simulations using BPLUT parameters from all land cover classes within the cell. The error contribution from land cover heterogeneity was then calculated as twice the 95th percentile difference between the land cover dominant and weighted simulations for each grid cell.

Model sensitivity was determined by taking the model derivative with respect to the parameter of interest and performing a linear transformation about the standard model input values. Input values were selected to represent the most sensitive portions of the soil moisture and temperature response curves and average conditions for representative northern biomes. The mean annual standard temperature and soil moisture for the error analysis was set as  $5^\circ\text{C}$  and 20% of saturation; GPP was assumed to be  $525 \text{ g C m}^{-2} \text{ yr}^{-1}$ . The resulting standard values for  $R_h$  and NPP were  $284 \text{ g C m}^{-2}$ , which are median values for the northern domain determined from the simulations. Each error source was considered independent among sources and stationary in time. This allows us to propagate errors using quadratic sums (i.e. by summing the error variances). The independent error sources from input data, model parameterization, and land cover heterogeneity categories are listed in the upper portion of **Table 8**, whereas combinations from each category are considered in the lower portion of the table. The proportional (%) contribution from each error source (or combination of sources) is computed by dividing the variance (or sum of squares) by the overall error ( $30 \text{ g C m}^{-2} \text{ yr}^{-1}$ ), which is the approximate sum of squares from all sources.

The resulting error budget indicates baseline L4\_C product performance within the target accuracy guidelines (i.e.  $\text{NEE RMSE} \leq 30 \text{ g C m}^{-2} \text{ yr}^{-1}$  and  $1.6 \text{ g C m}^{-2} \text{ d}^{-1}$ ). The error budget indicates that land cover heterogeneity contributes more than half (57%) of the total product

NEE uncertainty variance, while GPP, soil moisture and temperature inputs together contribute 31% of total error variance, and the remaining 12% of the expected total error variance attributable to model parameterization uncertainty. Errors contributed by model parameterization are the least certain component of the error analysis because it is difficult to precisely quantify global parameter variability and model structural inaccuracy. The error budget is defined relative to northern biomes, which are the primary focus of L4\_C science requirements and traceability. The relative (%) contributions of individual error components are likely to vary for other biomes and for variable weather and climate conditions. For example, the relative contribution of input soil moisture and temperature uncertainty is expected to be larger for warmer and drier grasslands relative to boreal forest biomes. The error budget in **Table 8** also represents a conservative measure of expected algorithm uncertainty because land cover heterogeneity is assessed relative to a coarser (25-km) product resolution rather than the 9-km baseline product resolution.

A spatial implementation of the L4\_C error (RMSE) budget over all global vegetated land areas was conducted using a forward model sensitivity analysis driven by MODIS FPAR and GMAO MERRA reanalysis daily surface meteorology inputs (**Figure 6**). These simulations include both random and systematic error components from model inputs and land cover heterogeneity effects at the 9-km spatial resolution of the global simulations and associated L4\_C product. For the simulations, soil moisture error was assumed to vary linearly with estimated GPP from 6% to 20% of soil saturation under low to high biomass productivity. GPP error was assumed to represent 30% (a conservative estimate) of daily GPP. The resulting global NEE error budget is similar to the previous error table (**Table 8**) and indicates that the L4\_C accuracy requirements are satisfied over more than 82 and 89 percent of respective global and northern ( $\geq 45^\circ\text{N}$ ) vegetated land areas. Systematic error from sub-grid scale land cover heterogeneity defined from the 1 km resolution MODIS land cover product and relative to the coarser 9 km resolution modeling grid is the largest source of estimated NEE uncertainty, though areas with large heterogeneity errors are confined to relatively few locations. Error contributions from the GPP calculations and soil moisture and temperature inputs are generally more spatially uniform, but vary with regional gradients in estimated vegetation biomass productivity. The estimated L4\_C NEE uncertainty increases in higher biomass productivity areas (e.g. forests) due to assumptions of increasing uncertainty in satellite microwave soil moisture retrievals and associated model assimilation based soil moisture inputs.



**Figure 6.** Estimated annual error (RMSE) budget for the L4\_C global NEE calculations; these results indicate that the L4\_C accuracy requirements (i.e. NEE RMSE  $\leq 30 \text{ g C m}^{-2} \text{ yr}^{-1}$ ) can be met over more than 82% and 89% of global and northern ( $\geq 45^\circ\text{N}$ ) vegetated land areas, respectively.

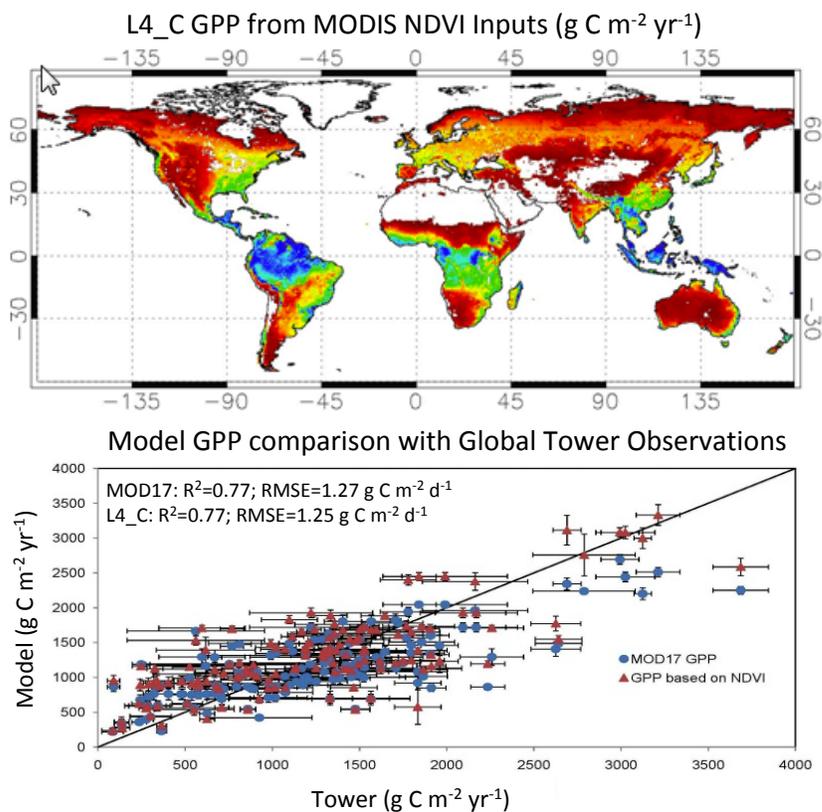
The estimated NEE uncertainty is lower than might be expected in some warmer tropical high biomass productivity areas (e.g. Amazon rainforest) because of reduced low temperature and moisture constraints to the L4\_C respiration calculations so that the bulk of model uncertainty is contributed by GPP in these areas. Model NEE uncertainty in the African Congo is relatively larger than Amazonia due to relatively drier climate conditions in central Africa defined from the MERRA surface meteorology inputs and associated uncertainty contributions from both respiration and GPP.

### 3.3.1 Impacts from Algorithm Options

The L4\_C algorithm options (**Table 6**) may impact product accuracy and relative error contributions from the ancillary inputs, model structure and parameterizations. Model parameterization uncertainty contributes approximately 12% of the total L4\_C based NEE estimation error variance; however, these estimates represent relative error contributions that will be refined using additional comparisons with observations and model sensitivity studies as discussed below (Section 3.7). The associated error contribution from model parameterization uncertainty would be reduced accordingly through the optional representation of land cover disturbance and recovery effects on L4\_C calculations. An investigation of disturbance recovery effects on prototype L4\_C calculations over a northern ( $\geq 45^\circ\text{N}$ ) domain indicated increased model GPP and NEE accuracy using an explicit model representation of disturbance recovery status over baseline, non-steady state simulations and relative to tower chronosequence based carbon fluxes (Yi et al. 2013); relative improvement in GPP accuracy was larger than the gain in NEE accuracy due to compensating effects of GPP and  $R_{\text{tot}}$ , while the baseline steady-state model performance was still within expected product accuracy guidelines.

Uncertainty associated with the GPP calculations may contribute up to 31% of the total NEE estimation error. The additional NEE uncertainty expected from using lower order VI inputs rather than MODIS FPAR to derive GPP is expected to be small. The potential effects of NDVI

based GPP calculations were evaluated through an initial global implementation of the L4\_C algorithms. These simulations were conducted over a 7-year record (2002–2008) at a daily time step and 1-km spatial resolution using MODIS GPP (Zhao and Running 2010) inputs and 0.5° resolution soil moisture and temperature inputs from the GMAO MERRA reanalysis (Kimball et al. 2011). The resulting uncalibrated global L4\_C simulations showed reasonable agreement (NEE RMSE = 1.78 g C m<sup>-2</sup> d<sup>-1</sup> and 34.0 g C m<sup>-2</sup> yr<sup>-1</sup>) with tower observations from 38 North America (Ameriflux) tower sites, though the NEE accuracy was lower than the expected baseline (**Table 8**). A second set of global simulations was conducted using ancillary MODIS (MOD13) NDVI inputs and L4\_C LUE model to estimate GPP. The resulting daily GPP calculations showed favorable agreement (R<sup>2</sup> = 0.77; RMSE = 1.25 g C m<sup>-2</sup> d<sup>-1</sup>) with coincident tower measurement based GPP from global FLUXNET sites (**Figure 7**). These calculations were similar or slightly better than the MODIS (C.5) MOD17 GPP product accuracy (R<sup>2</sup> = 0.77; RMSE = 1.27 g C m<sup>-2</sup> d<sup>-1</sup>) relative to the same tower site observations. These results indicate that the optional use of VI inputs to estimate GPP in the L4\_C model framework would have minimal negative impact on product accuracy.



**Figure 7.** Mean (2002-2008) annual GPP map (top) derived from the L4\_C optional GPP calculation using MODIS NDVI and MERRA meteorology inputs. The lower order NDVI inputs are used in the L4\_C option algorithm for empirical estimation of FPAR and LUE model based estimation of GPP. The L4\_C and MODIS (MOD17) results are compared with tower GPP estimates from global FLUXNET sites (lower), where error bars denote interannual variability in annual fluxes.

### 3.4. Numerical Computation Considerations

The SMAP L4\_C baseline product will be posted to an EASE-grid (version 2) global cylindrical equal area grid with 9-km grid cell resolution. Each nominal data granule will represent 1 day of data. The product will cover all global land areas but operational processing will only be conducted on vegetated grid cells defined from a static land cover map; open water and non-vegetated land areas including permanent ice and snow will be assigned unique mask value identifiers. The resulting baseline product data volume and HDF5 format will be approximately 117 MB per day (70 MB compressed data volume) and 41.7 GB per year (~25 GB compressed data volume). These data loads are within the resource capacity of most Desktop, Linux cluster and associated network computing environments.

While the final L4\_C product will be posted to a 9 km spatial resolution, operational processing will be conducted at a finer 1 km spatial resolution. Ancillary data inputs for L4\_C production will have variable spatial resolution and associated file sizes, including 1 km resolution for land cover classification and MODIS FPAR inputs; 3 km resolution for L3\_SM\_A inputs, and 9 km resolution for L4\_SM and GMAO surface meteorology inputs. The MODIS FPAR data are available in a tile based global sinusoidal projection with 8-day temporal fidelity from Aqua and Terra. A necessary preprocessing step will involve dynamic acquisition and reprojection of the MODIS FPAR tiles into the 1 km global EASE-grid format, and temporal subsampling or interpolation of these data to a daily time step for input into the L4\_C algorithms. It is anticipated that an on-line disk storage capacity of several terabytes will be needed for L4\_C production, including algorithm calibration, testing and validation activities. The L4\_C algorithms will require external satellite remote sensing inputs that may be processed at a different physical location. After production, the L4\_C product will be transmitted from the SMAP L4 Analysis Subsystem (L4\_SDS) at the NASA GSFC Global Modeling and Assimilation Office (GMAO) facility for storage to the SMAP mission DAAC at NSIDC. The network bandwidth for data transfer will therefore be an important consideration.

### 3.5 Programming/Procedural Considerations

The L4\_C algorithm science code will be written in ANSI C for use in a multi-processor Linux cluster environment. Source code version control will be conducted using Subversion, a mature multi-platform version control system. Within local prototyping environments, the GIT or Mercurial systems may be used. The final L4\_C product will be in HDF5 data format consistent with the other SMAP operational products. HDF5 provides rich support for multi-dimensional representations of all numeric data types, and support for modern metadata encapsulation within self-documenting 'container' technologies supported by extensive API libraries. The L4\_C science code will be transferred from the University of Montana's Numerical Terradynamic Simulation Group (NTSG) to the GMAO for translation and implementation as operational code in conjunction with L4\_SM production within the GMAO L4\_SDS. The L4\_C science code versioning will be conducted using the Subversion system, but can switch to Mercurial as needed, with test scripts implemented in Python (via numpy and H5py packages) and the R language. Programming and procedural considerations for the L4\_C product primarily involve the way in which the algorithms will be implemented and made operational after initial development and test cycles are complete. Adoption of specific metadata standards for the L4\_C and L4\_SM products will follow recommendations by the SMAP SDS and NASA Earth Science Data Systems Working Group (ESDSWG) on ISO-19115.x, 19139 and FGDC

variants as they evolve. Metadata design and implementation practices are also taking into account emerging best-practices in data preservation and data-curation (Duerr et al. 2011).

Baseline operations will ingest dynamic daily FPAR, L4\_SM, L3\_SM\_A and GMAO (FP) surface meteorology inputs external to the L4\_C algorithms. Pre-processing of these data prior to their input into the L4\_C algorithms will include reprojection, quality screening, gap filling and temporal subsampling or interpolation of the MODIS FPAR data to a continuous daily time step; sub-sampling of coarser scale L4\_SM, L3\_SM\_A and GMAO inputs to each 1 km resolution grid cell. Static ancillary inputs for the L4\_C baseline operations will include a global land cover map used to define the general biome (BPLUT) properties of each grid cell. An ancillary global 1 km resolution, 8-day FPAR climatology will be used for gap filling of lower quality MODIS 8-day FPAR inputs during the L4\_C preprocessing stage and prior to temporal subsampling or interpolation of the FPAR data for L4\_C daily processing. The ancillary land cover classification inputs will be consistent with the land cover data used to derive the FPAR inputs, and across the various SMAP operational products to the extent possible.

The L4\_C algorithms will require initialization of ancillary SOC stocks under average vegetation and climate conditions as a necessary pre-processing step. Model SOC initialization will be conducted during the mission pre-launch phase using available FPAR (MODIS), freeze/thaw and reanalysis (GMAO) daily inputs. The initialization process will involve cycling the global model simulations using a daily climatology established from a long-term (10+ year) record to achieve mean SOC steady-state conditions. The SOC pools will then be dynamically updated during L4\_C operations through the daily allocation of NPP using BPLUT defined litterfall rates and estimated  $R_h$ . Algorithm calibration and initialization activities will also include BPLUT calibration using ancillary tower eddy covariance  $CO_2$  measurements and global soil inventory data during the pre-launch science code development and testing phase. Re-initialization of the SOC pools and re-calibration of the BPLUT may occur during the mission post-launch phase using SMAP operational L4\_SM and L3\_SM\_A inputs.

### 3.5.1 Impacts from Algorithm Options

The L4\_C algorithm options (e.g. **Table 6**) would affect algorithm programming and procedures. The optional use of a static 1 km resolution global disturbance recovery map would record time (year of disturbance) of last prior disturbance, as defined from long-term (10+ year) satellite (e.g. MODIS MCD45) records and defined during the mission pre-launch period. The length (years since disturbance) of recovery would then be tracked for every 1-km grid cell during L4\_C operations and used for empirical adjustment of estimated litterfall and SOC during L4\_C operational production (e.g. Yi et al. 2013).

The optional use of lower order VI inputs (e.g. from MODIS or VIIRS) would require spatial reprojection of these data to a 1 km resolution global EASE-grid (version 2) format and temporal subsampling or interpolation to a continuous daily time step similar to the procedure for processing the baseline MODIS FPAR inputs. The satellite based VI records represent temporal composites (e.g. 8-16 day for MODIS or VIIRS) of daily observations; the dynamic best QC VI data would be used for estimating FPAR using the ancillary MODIS FPAR 8-day climatology and land cover class specific empirical coefficients defined in the BPLUT. The VI estimated FPAR record would then be subsampled or temporally interpolated to a daily time step between the current ( $t_i$ ) and prior ( $t_{i-1}$ ) time steps on a per grid cell basis. If there are insufficient best quality VI input data for FPAR estimation, the data would be gap filled using the corresponding

ancillary mean 8-day FPAR climatology values. The ancillary FPAR climatology will be established on a grid cell-wise basis from existing (e.g. MODIS MOD15 Collection 5 or higher) high quality FPAR records extending over a relatively long (e.g. from 2000) record. Potential VI inputs from other satellite records (e.g. VIIRS) would be calibrated to the respective MODIS FPAR climatology on a grid cell-wise basis to ensure relative consistency among potential alternative data sources. This activity would be conducted as a pre-process to the L4\_C algorithm implementation similar to preprocessing of the baseline FPAR inputs.

### 3.6 Ancillary Data Availability/Continuity

Anticipated primary ancillary inputs to the L4\_C algorithms are summarized in **Table 7**. All of the required ancillary data inputs for L4\_C processing are publicly available as existing operational products from NASA data archival (DAAC) facilities, the GMAO or will be produced during SMAP operations. Temporally dynamic (daily) inputs to the L4\_C algorithms include FPAR, which will be derived from the MODIS (MOD/MYD15) operational product streams (baseline) from Terra or Aqua. Other dynamic daily inputs include landscape freeze thaw status, surface and root zone soil moisture and surface soil temperature that will be provided from SMAP operational L3\_SM\_A and L4\_SM product streams. The L4\_C algorithms also require daily surface meteorology inputs, including incident solar shortwave radiation, vapor pressure deficit and minimum daily air temperature; these data will be provided by GMAO reanalysis data derived from the same land model (GEOS-5 or later) used to produce the SMAP L4\_SM product. Several similar surrogate products are being used for pre-launch L4\_C development and testing, including the GMAO GEOS-5 based MERRA reanalysis (Yi et al. 2011) and a satellite passive microwave remote sensing based global daily freeze/thaw classification record (Kim et al. 2012). The MODIS Collection 5 FPAR record is currently available and regularly updated through existing NASA public data archives; these data are also being used for pre-launch L4\_C algorithm testing, while similar (Collection 5 or 6) data will be used for post-launch operations.

Static ancillary inputs to the L4\_C algorithms include a global land cover classification that is used to define the product domain and spatial patterns of biome (BPLUT) properties and environmental response characteristics. There are a variety of suitable global land cover products available that have been derived from various operational satellite remote sensing datasets, including AVHRR (Hansen et al. 2000), MODIS (Friedl et al. 2010), SPOT (Bartholomé and Belward 2005) and Landsat (Tucker et al. 2004). The MODIS IGBP global land cover classification (Friedl et al. 2010, 2011) is available at 1-km spatial resolution and is used in the production of MOD15 LAI/FPAR and MOD17 GPP global products. This product has also been used for regional and global implementation and testing of the L4\_C algorithms during the SMAP pre-launch phase, and produces suitable product accuracy (Kimball et al. 2009a, McGuire et al. 2012, Yi et al. 2013). The MODIS land cover classification (Collection 5 or later) will be aggregated into a reduced set of 8 global PFT classes that will be used with the BPLUT to define the global vegetated domain for model calculations and grid cell-wise environmental response characteristics. An additional static ancillary input to the L4\_C algorithms includes a MODIS mean 8-day FPAR global climatology map series. The climatology is derived on a grid cell-wise basis using best quality MODIS (MOD/MYD15) FPAR retrievals from the long-term (10+ year) satellite record. These data will be derived during the SMAP pre-launch phase using existing MODIS land parameter records (Collection 5 or later).

The MODIS sensors on the EOS Terra and Aqua satellites are currently in good health and expected to remain operational through 2017, with Aqua operations potentially extending into 2022 (Ritchie et al. 2013). However, both MODIS sensors are in extended mission phases and could fail at any time, eliminating the FPAR (MOD15 and/or MYD15) operational data stream. A similar FPAR (VVI2P) product is planned for JPSS VIIRS (NGST 2011), which would provide continuity to the MODIS product series, but likely not until well after the initiation of SMAP operations. The NASA NPP VIIRS mission currently provides operational global VI (EVI and NDVI) products. If the MODIS FPAR operational product stream becomes unavailable during the SMAP mission period, FPAR could be derived from the lower order VI inputs using the L4\_C algorithm option (**Table 6**). Global VI products, including NDVI, will be operationally available from several potential sources during the SMAP mission period, including VIIRS (e.g. **Table 9**).

**Table 9.** Potential sources of satellite based operational products providing dynamic <sup>1</sup>vegetation inputs for L4\_C production during the SMAP mission development and operations time-frame.

Agency	Satellite	Sensor	Period	Current/Planned Operational Products
NASA	Terra	MODIS	2000- <sup>2</sup> 2017	LC, VI, FPAR, BA
NASA	Aqua	MODIS	2002- <sup>2</sup> 2017	LC, VI, FPAR, BA
NOAA	NOAA	AVHRR	1982- ongoing	VI
NASA	NPP	VIIRS	2011- <sup>3</sup> 2015	VI, BA
IPO/NOAA	JPSS-1	VIIRS	2016- <sup>3</sup> 2021	VI, BA, FPAR
ESA, NOAA	MetOp	AVHRR	2006- <sup>3</sup> 2020	VI

<sup>1</sup>Vegetation inputs include one or more of the following: Land cover classification (LC), VI, FPAR, active fire or burned area (BA) time series;

<sup>2</sup>Estimated duration of continuing operations based on sensor/satellite health and available fuel (Ritchie et al. 2013);

<sup>3</sup>Expected mission life.

Other ancillary data will be used for pre-launch L4\_C algorithm development and refinement and post-launch product calibration and validation activities, but are not required for L4\_C operational production. These data include tower CO<sub>2</sub> flux measurement based estimates of GPP, R<sub>tot</sub> and NEE, and other biophysical data from global FLUXNET sites (Baldocchi 2008). The FLUXNET data are available globally (<http://www.fluxnet.ornl.gov/fluxnet/index.cfm>), while many of these sites have multi-year data records with well defined accuracy. The La Thuile 2007 synthesis activity and associated global product was derived from FLUXNET observation records and provides a relatively consistent set of global, gap filled daily tower CO<sub>2</sub> flux records with well defined uncertainty (<http://www.fluxdata.org/DataInfo/default.aspx>). Static SOC inventory records are available from global site network measurements (e.g. Zinke et al. 1984) and existing global maps including IGBP-DIS (Global Soil Data Task 2000) and the Northern Circumpolar Soil organic Carbon database (Tarnocai et al. 2009). Global vegetation productivity records are available from existing MODIS (MOD/MYD17) GPP (Collection 5) operational data

records (Zhao and Running 2010). Other global carbon products available for L4\_C pre-launch development and post-launch validation activities include observation based empirical multi-tree ensemble (MTE) upscaled global monthly carbon products (Jung et al. 2011). Fire disturbance and recovery effects on the L4\_C calculations are also being evaluated using the MODIS (MCD45) operational burned area product (Roy et al. 2008) and a satellite ensemble based monthly Global Fire Emissions Database (GFED, Giglio et al. 2010).

### 3.6.1 Impacts from Algorithm Options

The L4\_C algorithm options (**Table 6**) have additional ancillary input requirements relative to the product baseline described above. The optional calculation of FPAR internal to the L4\_C algorithm involves using lower order VI inputs that are readily available as current or planned operational products under other missions (e.g. **Table 9**). The use of alternative VI inputs from different sensors would require recalibration and re-evaluation of the empirical FPAR models to mitigate potential impacts from cross-sensor bias and calibration differences. The VIs would then be used to compute FPAR based on BPLUT specific empirical relationships between the VI inputs and MODIS FPAR climatology. Initial testing of this optional approach showed minimal impact to GPP estimation accuracy or L4\_C based NEE calculations relative to the baseline simulations and tower network observations over northern land areas (Yi et al. 2013) and the larger global domain (**Figure 7**).

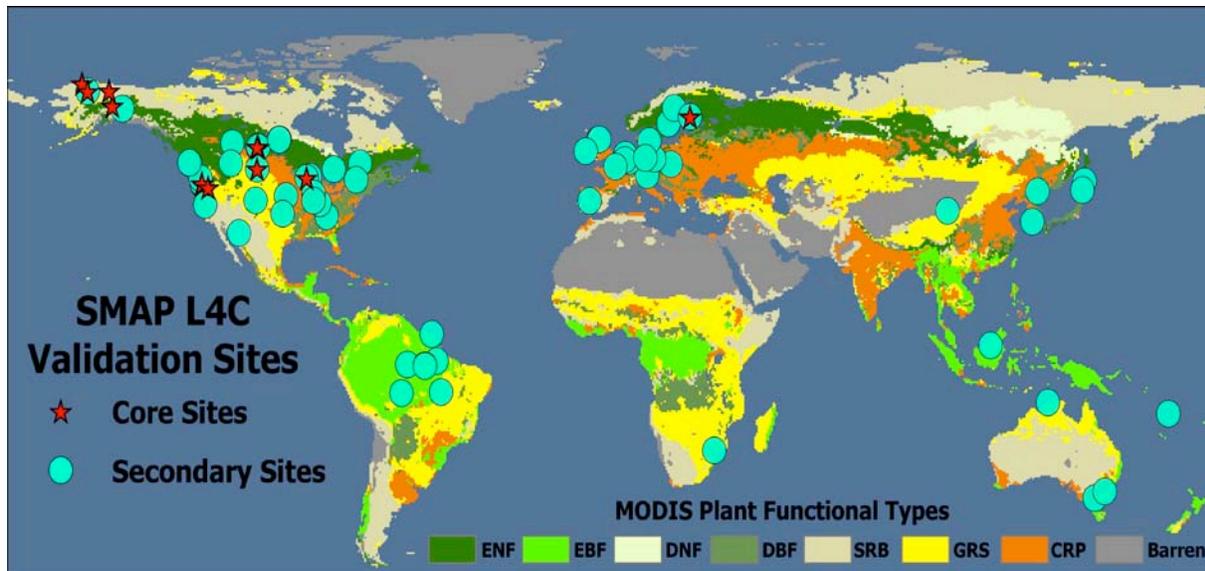
The L4\_C disturbance option would use a static disturbance recovery status map as an additional ancillary input to define the relative deviation of NPP and surface SOC calculations from dynamic steady-state conditions. The MODIS MCD45A1 burned area index product (Roy et al. 2008) is available globally at 500-m spatial resolution and provides suitable information to describe recent (from 2000) global burned area disturbance and recovery status for the L4\_C calculations. A similar MCD45A1 product may not be operationally available after the MODIS era, though an active fire EDR is currently underway or planned for VIIRS on NPP and JPSS, and could be processed to provide similar global disturbance recovery information.

### 3.7 Calibration and Validation

The SMAP instrument and product calibration and validation activities are described in the SMAP Science Data Calibration and Validation plan (Jackson et al. 2012) while this section describes calibration and validation activities as they pertain to the L4\_C product. NEE represents the primary L4\_C product to be validated, while product validation activities and success criteria will emphasize northern ( $\geq 45^\circ\text{N}$ ) land areas consistent with SMAP carbon cycle science objectives for quantifying net ecosystem  $\text{CO}_2$  exchange in boreal landscapes.

The statistical methods and domains of validity envisaged for calibrating and testing the L4\_C algorithms and for demonstrating that their performance meets the SMAP science requirements will involve direct comparisons between model daily outputs and tower eddy covariance  $\text{CO}_2$  flux measurements from available FLUXNET tower sites, and representing dominant global biome types (**Figure 8**). Available tower sites will be selected for calibration and validation on the basis of being representative of the dominant vegetation class within a L4\_C grid cell, and having relatively long-term and complete data records with well defined measurement accuracy and uncertainty. The relative accuracy of L4\_C outputs will be established in relation to available tower network observations within regionally dominant vegetation classes and following established protocols (Running et al. 1999, Heinsch et al. 2006, Kimball et al. 2009a, Yi et al. 2013). Primary validation activities will involve direct

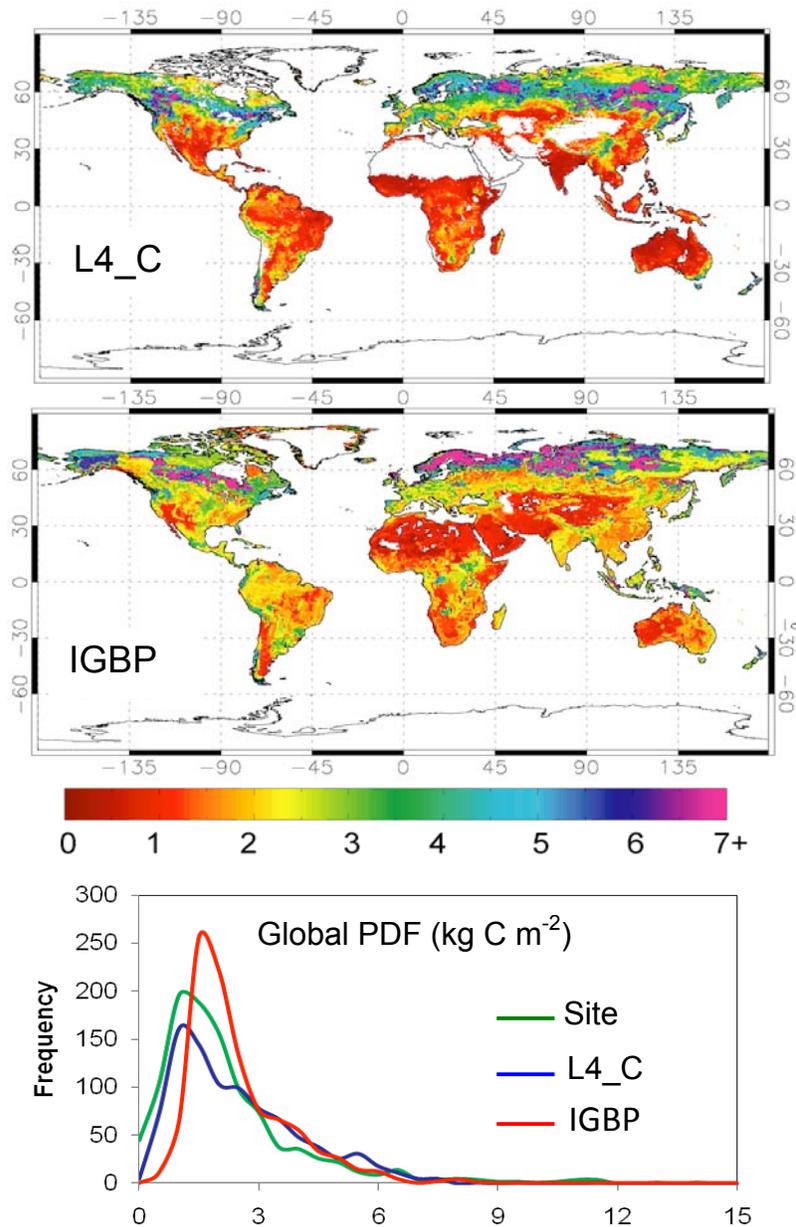
comparisons of SMAP L4\_C and corresponding tower NEE estimates on a daily basis for a subset of core tower sites. The core tower sites will represent a smaller subset of the global tower validation network (**Figure 8**) where formal data use agreements have been established between the SMAP mission and individual tower PIs to provide near real-time observational data during the post-launch SMAP validation period. The skill metrics for primary validation will include correlation, RMSE and bias between L4\_C and corresponding tower daily NEE estimates for northern land areas to document that model NEE meets documented accuracy requirements for product success (i.e. mean daily RMSE  $\leq 1.6 \text{ g C m}^{-2}$ ).



**Figure 8.** The FLUXNET global tower network consists of more than 500 sites covering the range of global biomes. More than 80 sites satisfy L4\_C validation criteria for having long-term (>1 yr) records, homogenous land cover conditions and representing regionally dominant biome types (above). A subset of ~14 core sites will provide near real-time tower data supporting post-launch validation activities. Additional validation activities will include comparisons against more numerous secondary tower site observations that may be consistent in space but not necessarily in time with the SMAP L4\_C retrievals. The above map also shows the MODIS MCD12Q1 (V5) global plant functional type (PFT) classification; similar ancillary land cover information will be used for the planned L4\_C calculations.

Additional (secondary) validation activities will involve product comparisons against a larger global network of tower carbon flux observations and synergistic global multi-year data records having generally well characterized uncertainty, but not necessarily temporally co-located with L4\_C operational outputs. These secondary activities will include global comparisons of L4\_C outputs, including NEE, SOC, GPP, and  $R_{\text{tot}}$ , against tower observation based estimates and soil carbon inventory records (**Figure 9**). The L4\_C product evaluation will also include secondary comparisons against spatially contiguous MODIS (MOD/MYD17) GPP operational data records and global MTE upscaled monthly carbon products. The primary skill metrics for these

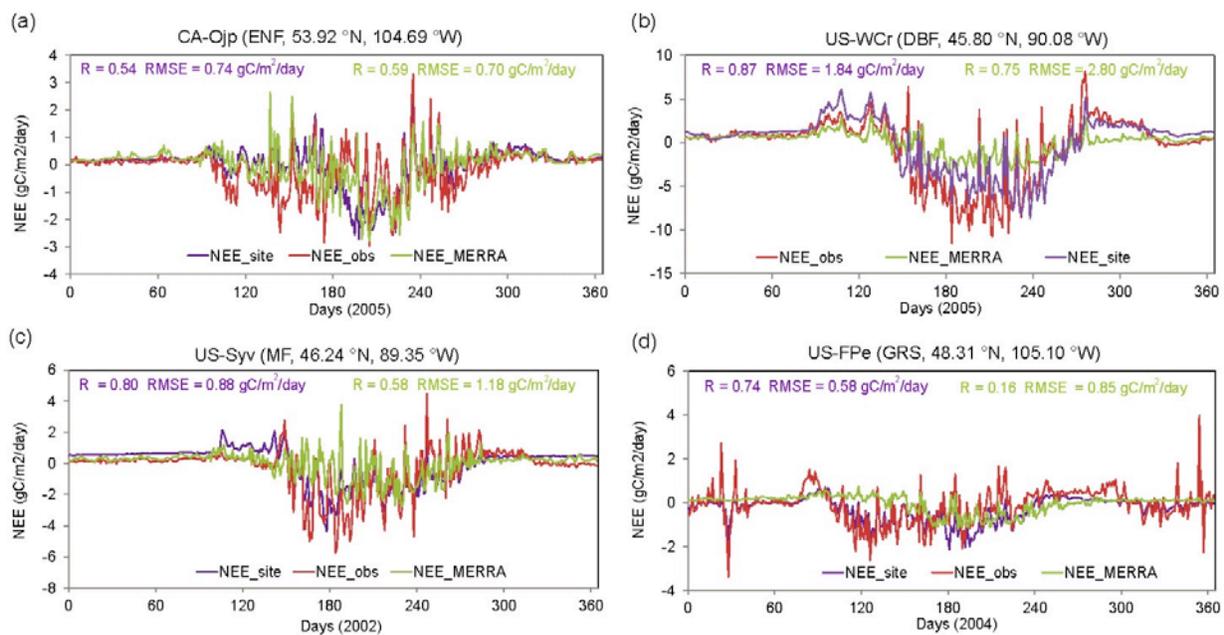
comparisons will include model sensitivity diagnostics, correlation, RMSE, bias and distribution matching, and consistency checks.



**Figure 9.** Global map of surface (<10 cm depth) soil organic carbon (SOC) stocks (kg C m<sup>-2</sup>) derived from the L4\_C algorithm using an 8 year (2002-08) record of MODIS FPAR and MERRA reanalysis daily meteorological inputs (top); non-vegetated areas (in white) were masked from L4\_C processing. An independent SOC map derived from IGBP soil inventory data is also shown (middle), while global Probability Density Function (PDF) distributions of the L4\_C results and two other global SOC inventory records are also shown (bottom; Global Soil

Data Task 2000 (IGBP), Zinke et al. 1986 (Site)). Similar comparisons will be used for SMAP L4\_C calibration and validation activities.

Calibration and validation of the L4\_C algorithms and products will involve model sensitivity studies in relation to observed variability in environmental conditions, uncertainties in the LUE based GPP calculations and L4\_SM inputs (e.g. **Figure 10**). Model sensitivity studies will be conducted by perturbing input parameters within their respective ranges of uncertainty independently and in combination, and documenting L4\_C algorithm responses. The L4\_C algorithms will be run using both tower and reanalysis based surface meteorology inputs and local and regional land cover conditions to quantify relative error contributions from remote sensing and meteorological reanalysis drivers, and model parameterizations and assumptions. Similar L4\_C algorithm sensitivity studies will be conducted at other FLUXNET sites during the SMAP pre-launch algorithm calibration and refinement period and post-launch validation period.

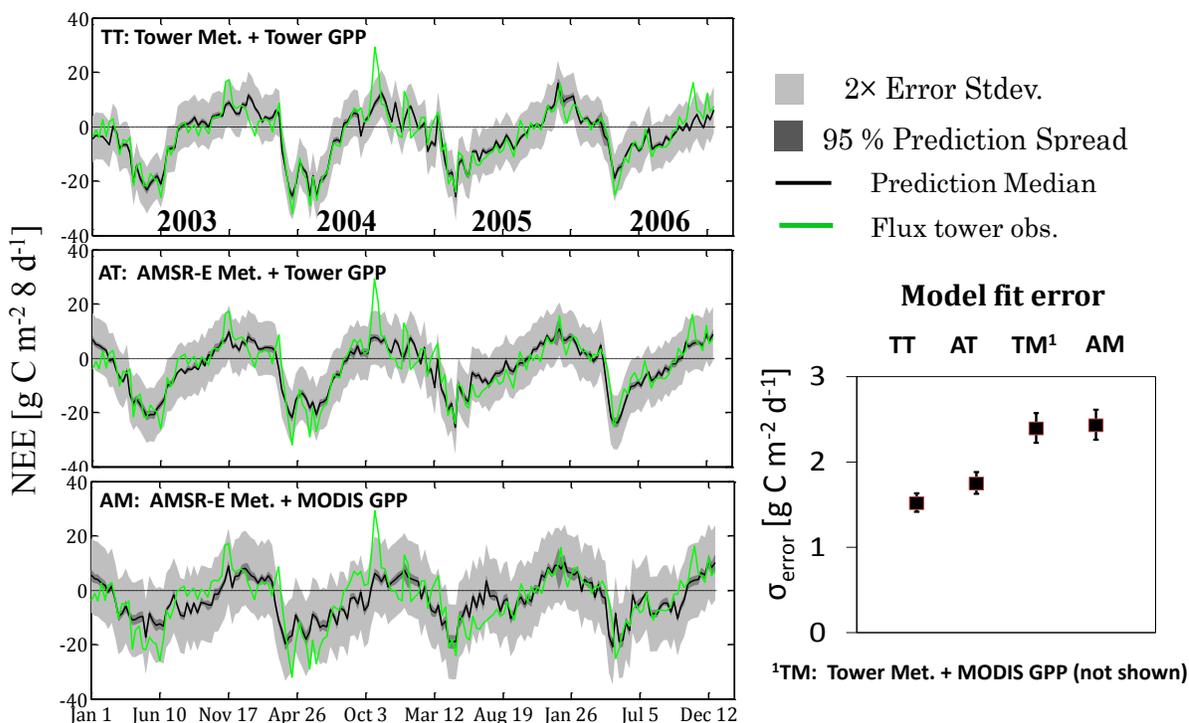


**Figure 10.** Comparison of tower derived and L4\_C model estimated daily NEE for four dominant northern biome types (Yi et al. 2013). The L4\_C simulations were derived using alternative model forcings, including MERRA surface meteorology and MODIS NDVI inputs (NEE\_MERRA), and tower observation based GPP and meteorology inputs (NEE\_site). The model results are evaluated against tower observation based daily NEE (NEE\_obs). Similar forward model simulations and comparisons will be conducted during the SMAP pre-launch and postlaunch periods to document L4\_C algorithm sensitivity and product accuracy.

Model-data fusion and data assimilation approaches have been used to quantify the relative value of remote sensing observations in land surface models (Renzullo et al. 2008, Crow 2007) and to estimate model parameters by inverting carbon flux measurements from regional tower networks (Sacks et al. 2007, Zobitz et al. 2008, Knorr and Kattge 2005). Daily eddy flux tower measurements are inherently noisy and influenced by local terrain and micrometeorology effects (Richardson and Hollinger 2005, Baldocchi 2008). Uncertainty also surrounds model assumptions and parameters, and remote sensing inputs. The data assimilation framework is useful

for parameter and state estimation in carbon cycle science, and differs from classical estimation techniques, by accounting for both observation and model estimation errors (Raupach et al. 2005).

The Bayesian synthesis framework is useful for parameter and state estimation in carbon cycle science, and differs from classical estimation techniques by accounting for uncertainty in observations as well as model estimates (Raupach 2005, Knorr 2005). A Bayesian synthesis approach will be used to optimize model fit by adjusting biome-specific (BPLUT) model parameters and to provide parameter probability distributions that reflect inherent sources of uncertainty (e.g. **Figure 11**). Model-data configurations will be evaluated for overall model fit and physical consistency of parameter estimates (Renzullo 2007). The optimization of selected L4\_C algorithm parameters will be conducted using daily time series carbon fluxes (GPP,  $R_{tot}$ , NEE) from tower CO<sub>2</sub> eddy covariance flux sites representing major biome types within the L4\_C product domain. These activities will be conducted during the SMAP pre-launch phase using available surface meteorology inputs from MERRA reanalysis, which is similar to the planned L4\_SM product, and remote sensing inputs from MODIS (Zhao and Running 2010), AMSR-E (Jones et al. 2010) and SMOS. Similar calibration and validation activities will be conducted during the post-launch phase using SMAP product retrievals.



**Figure 11.** Example Bayesian L4\_C model fit at the Tonzi Ranch, FLUXNET tower (D. Baldocchi PI) for 2002-2006. The Tonzi Ranch tower footprint represents grassland, but is classified as Woody Savanna (mixed oak forest and grassland) in the overlying 1-km resolution MODIS grid cell. Each panel shows model results for alternative temperature, moisture and GPP inputs, which indicate the respective RMSE error ( $\sigma_{error}$ ) attributed to model logic and ancillary inputs.

A Monte Carlo Markov Chain (MCMC) optimization approach will be applied to minimize an objective function weighted by the observation error and model error covariance matrices by

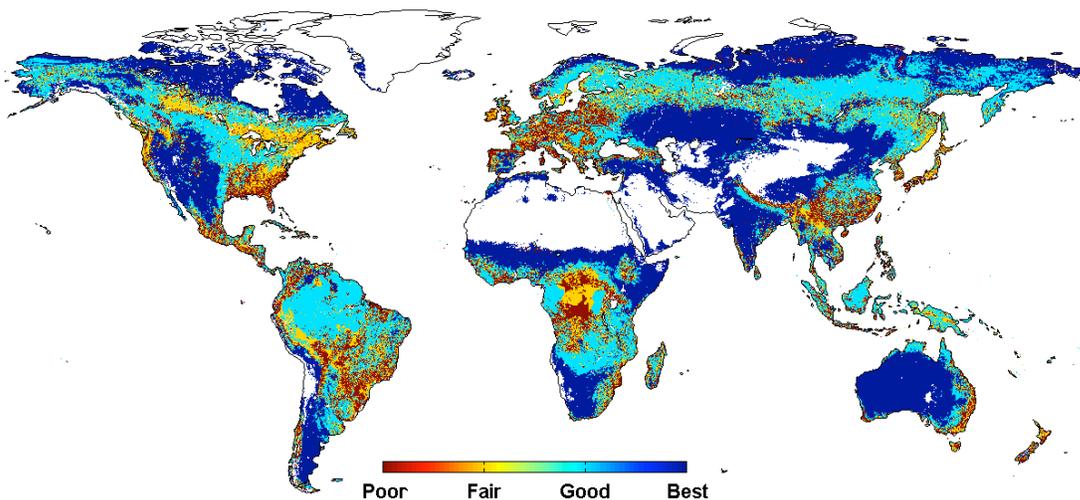
adjusting model parameters within expected uncertainty ranges for individual biome types as defined from the literature (e.g. Yi et al. 2013). Model parameters suitable for optimization include BPLUT defined proportional carbon allocations ( $C_{\text{fract}}$ ), carbon use efficiency (CUE) or proportional allocation of GPP to  $R_a$ , optimal soil decomposition rate constants ( $K_{\text{mx}}$ ), litterfall fraction ( $F_{\text{str}}$ ) and soil moisture response coefficients ( $W_{\text{mult}}$ ). Smaller values of the objective function are associated with more informative model-data configurations and resulting posterior distributions that allow for significance testing. The resulting simulated carbon pools and fluxes will be compared to the available tower observational data. Model-data configurations will be evaluated for overall model fit and physical consistency of parameter estimates (Renzullo 2007). Additional verification of the initial carbon pools will include comparisons of regional SOC patterns and magnitudes against site and regional soil inventories (Tarnocai et al. 2009; Global Soil Data Task 2000; Zinke et al. 1984). This approach will provide quantitative and uncertainty estimates of the relative value of L4\_C outputs in the overall match of model fit to noisy observations.

### 3.8 Quality Control and Diagnostics

Quality Control (QC) flags will be obtained as metadata from planned MODIS FPAR, SMAP L3\_SM\_A and L4\_SM inputs to the L4\_C algorithms. These quality flags will be carried through and combined with additional QC information acquired during L4\_C processing, and included with the L4\_C product. The L4\_C QC flags will represent a range of information, including validity of estimated carbon fluxes falling within predefined realistic min/max ranges; the proportion of successfully processed L4\_C 1-km pixels within each 9-km grid cell; the dominant PFT class within each 9-km resolution grid cell; whether GPP was derived using ancillary FPAR or alternative NDVI inputs; whether FPAR was derived using dynamic 8-day FPAR inputs or the ancillary FPAR 8-day climatology; whether the FT inputs were derived from ancillary L3\_SM\_A freeze/thaw or  $T_{\text{surf}}$  inputs. The global realistic min/max ranges for estimated carbon fluxes (NEE, GPP,  $R_h$ ) and SOC fields will be established during the mission pre-launch phase from long-term (e.g., 2000-2012) L4\_C model simulations driven by available ancillary inputs, including MODIS FPAR, FT-ESDR (Kim et al. 2012) and GMAO MERRA (Yi et al. 2011) records. The estimated model output ranges will be verified against in situ carbon flux observations from the global tower measurement network (Baldocchi 2008) and similar carbon variable estimates available from other global products (Zhang and Running 2009; Jung et al. 2011, Global Soil Data Task 2000). The resulting min/max ranges will be stored as an ancillary look-up table for assigning QC flags to the model outputs for each 1-km pixel and then aggregated to a single QC flag for each 9-km grid cell and daily time step; the QC spatial aggregation may involve a simple identification of one or more 1-km pixels identified as “invalid” within each 9-km grid cell. The resulting QC information will be evaluated and potentially refined through post-launch L4\_C cal/val activities using carbon flux measurements from in situ measurement networks and algorithm sensitivity studies conducted over the observed range of environmental variability. Post-launch diagnostics will also be performed on the operational L4\_C output fields, and grid cell values exceeding the specified performance thresholds will be identified and flagged for additional user evaluation and potential troubleshooting by the SMAP Science Data System (SDS) and Algorithm (ADT) teams.

The L4\_C product will also include more detailed quality assurance (QA) information on estimated carbon model performance for NEE as the primary validated product. The L4\_C QA performance metric will provide a spatially explicit and locally weighted estimate of NEE RMSE

uncertainty. The QA metric will be derived globally from forward model sensitivity simulations (Section 3.7) that account for uncertainty ranges in model assumptions and critical parameters, and expected uncertainty in ancillary data inputs (e.g., Yi et al. 2013). The model simulations will be derived during the mission pre-launch phase using available ancillary data extending over multi-year (e.g., 2000-2012) records, including MODIS FPAR and GMAO MERRA inputs. The resulting simulations will be validated against independent in situ NEE observations from the global tower eddy covariance CO<sub>2</sub> flux measurement network (Baldocchi 2008). The QA metric will include both temporally static and dynamic elements. Spatially variable but temporally static QA information will incorporate NEE RMSE uncertainty determined from model sensitivity simulations that account for uncertainties in model, inputs, assumptions and parameterizations, and expected lower accuracy in areas with higher vegetation biomass, and land cover and terrain heterogeneity; these conditions will also affect SMAP retrievals and associated L3\_SM\_A and L4\_SM inputs, while the L4\_C outputs may not adequately represent sub-grid scale variability under complex land cover and terrain conditions. This model performance QA information will be derived from pre-launch algorithm sensitivity runs using documented uncertainties in the L4\_C inputs and updated by the observed environmental conditions (e.g. **Figure 6**). An example NEE RMSE QA map is shown in **Figure 12** and was defined from spatially explicit estimation of the L4\_C NEE error budget over the global domain (Section 3.3). The detailed NEE RMSE (g C m<sup>-2</sup>) QA performance fields will be aggregated to a smaller set of discrete relative quality categories (e.g., 0=best, 1=good, 2=fair, 3=poor) and posted to a daily QA granule as part of the L4\_C operational product set. The detailed NEE RMSE QA carbon units may be provided as a research product to support post-launch L4\_C cal/val activities.



**Figure 12.** Example NEE RMSE product quality (QA) map derived from the L4\_C global error analysis (Section 3.3) and representing: Poor ( $RMSE > 30 \text{ g C m}^{-2} \text{ y}^{-1}$ ), Fair ( $20 \text{ g C m}^{-2} \text{ y}^{-1} < RMSE \leq 30 \text{ g C m}^{-2} \text{ y}^{-1}$ ), Good ( $10 \text{ g C m}^{-2} \text{ y}^{-1} < RMSE \leq 20 \text{ g C m}^{-2} \text{ y}^{-1}$ ), and Best ( $RMSE \leq 10 \text{ g C m}^{-2} \text{ y}^{-1}$ ) discrete QA categories.

### 3.9 Exception Handling

Exception handling for the L4\_C product will follow the approach of lower level SMAP data products to the extent possible. In the L4\_C codes, runtime messages emitted are classified as

follows (Fatal, Error, Warning, and Advisory); of these, the Fatal, Error, and Warning category messages may be considered exceptions. Initial work on defining a SMAP product wide exception handling protocol is treated in the JPL memo document SMAP-860-003-11, “Standard for Message File Generation for the SMAP/DESDynI SDS”. This document primarily addresses standards for composing Message Files, which include but are not strictly limited to exception handling. The exception handling approach at the granule algorithm level will be consistent and interoperable within the larger job scheduling and resource allocation system, whether based on the Science Processing and Data Management (SPDM) system from JPL, or a functional equivalent implemented at the GSFC GMAO.

### 3.10 Interface Assumptions

The baseline L4\_C product will use dynamic ancillary inputs of daily landscape freeze/thaw status and surface soil moisture and soil temperature from the SMAP L3\_SM\_A and L4\_SM products, daily surface meteorology inputs from the GMAO global forward processor (FP), and 8-day FPAR from MODIS MOD15 or MYD15 (Collection 5 or later) product streams. Static ancillary inputs to the L4\_C algorithms include a global land cover (PFT) classification and 8-day global FPAR climatology database established from the long-term MODIS best quality FPAR record. The dynamic inputs to the L4\_C algorithms will be obtained from the GMAO during the SMAP operational period, while static inputs will be obtained from the GMAO and SMAP SDS. Ideally, these products will be provided on the SMAP Earth-fixed nested global grids and also contain the appropriate quality control flags (e.g. Section 3.8).

The L4\_C options would require additional ancillary inputs, including dynamic land parameter inputs from lower order VI series as a substitute for the FPAR baseline, and a static global disturbance recovery status map. The VI and disturbance inputs would be provided by the GMAO, ideally from either MODIS (MOD13 or MYD13) or VIIRS (VVI) sensor data streams. All of these ancillary inputs would include associated QC information derived from these data sources, which would be carried through to the L4\_C product stream.

The dynamic ancillary inputs for the L4\_C algorithms must be available with latencies of at least two days less than the corresponding (14-day) latency of the L4\_C product to allow sufficient time for L4\_C processing. However, the potential use of lower order 16-day VI inputs in place of 8-day VI or FPAR inputs could double the resulting L4\_C product latency relative to the baseline. Once generated, the L4\_C product and associated metadata will be transferred to the SMAP mission DAAC at NSIDC for permanent archival.

### 3.11 Test Procedures

The L4\_C production may involve implementing the algorithms in conjunction with, but external to the GMAO assimilation system using dynamic L3\_SM\_A, L4\_SM, GMAO and MODIS inputs. An effort will be made to document and ensure consistency between ancillary inputs to the L4\_C algorithms and other SMAP products. The FPAR inputs will be provided as external inputs from ancillary data sources (baseline) or generated internal to the L4\_C algorithm framework (option) using lower order ancillary VI inputs from operational satellite records (e.g. MODIS, VIIRS) and GMAO daily surface meteorology. The use of L4\_SM and GMAO inputs and ancillary global FPAR climatology enables spatially and temporally continuous mapping of NEE and component carbon fluxes for all vegetated land areas in the domain, as these data are not constrained by SMAP sensor limitations on soil moisture and temperature retrievals. Potential gaps in the dynamic L3\_SM\_A freeze/thaw and MODIS FPAR inputs could result in a

lower product QC ranking, but would not prevent L4\_C production. Vegetated land areas will be defined from a global land cover classification and where the MODIS FPAR records indicate positive vegetation productivity on an annual basis. The use of existing model reanalysis data (e.g. MERRA) and currently available remote sensing (e.g. AMSR-E, SMOS and MODIS) products as ancillary inputs enables prototype L4\_C products to be generated during the SMAP pre-launch phase and well before initiation of the operational SMAP data stream. These activities provide for the development of mature calibration and validation protocols and a standard from which improved model calculations using SMAP derived inputs can be assessed.

Test procedures during L4\_C processing will include operational checks of model performance against land cover class specific QC range thresholds established during the pre-launch development phase and potentially refined during post-launch operations. Product outputs that exceed specified range thresholds will be flagged and routed to SDS monitors and ADT scientists for further analysis and diagnosis. Potential algorithm adjustments to rectify model performance in accordance with product accuracy guidelines may include re-initialization of SOC pools, BPLUT recalibration or adjustment of QA/QC criteria to clarify and potentially improve product accuracy. These changes would be tested on the offline L4 algorithm simulation system running parallel with L4 operations at the GMAO. Any actions taken to meet product performance and accuracy guidelines would require approval from the science (ADT) and SDS change control board.

A detailed summary of the L4\_C calibration and validation approach is provided elsewhere (Jackson et al. 2012, Kimball et al. 2011, Yi et al. 2013) and summarized below. The statistical methods and domains of validity anticipated for testing the L4\_C algorithms and for demonstrating that their performance meets the SMAP science requirements will involve direct comparisons between model outputs and tower eddy covariance CO<sub>2</sub> flux measurements from northern (primary validation domain) and globally distributed monitoring sites from the FLUXNET network (Baldocchi 2008). Approximately 80 (from >400) FLUXNET sites meet L4\_C validation criteria for having spatially homogeneous land cover characteristics and multi-year data records with well characterized uncertainty. A subset of tower site investigators were solicited or have proposed to be SMAP early adopters (Jackson et al. 2012) from which formal agreements are being established with the mission for near real-time tower data access. The SMAP mission team will work with these collaborators to provide tower based observational data in a timely manner and concurrent with the SMAP product stream during the post-launch calibration and validation period to ensure robust L4\_C validation. A similar strategy was successfully implemented for testing MODIS MOD17 GPP products (Running et al. 1999, Baldocchi et al. 2001, Heinsch et al. 2006, Turner et al. 2006) and prototype L4\_C model based NEE and component carbon flux simulations (e.g. Yi et al. 2013). The L4\_C performance and error budgets will also be determined through model perturbation and sensitivity analyses spanning the range of observed environmental conditions and using model input accuracy information (e.g. Kimball et al. 2009a). Primary validation metrics for these activities will include correlation, RMSE and bias techniques.

Secondary validation activities will involve comparisons of L4\_C outputs against FLUXNET data records from up to 100 or more tower validation sites distributed globally and representing the major PFT and climate zones; these activities will involve consistency checks of L4\_C outputs against multi-year, gap filled tower records co-located in space but not necessarily in time (Baldocchi 2008). The L4\_C products will also be compared against other similar,

synergistic land products with relatively well characterized uncertainty, including spatially contiguous, observation based global carbon products derived from empirical upscaling techniques (Jung et al. 2011), global SOC inventory records (Tarnocai et al. 2009), and MODIS (MOD17) GPP records (Zhao and Running 2010). Validation metrics for these activities will include sensitivity diagnostics, correlation, RMSE, bias and distribution matching techniques.

### 3.12 Algorithm Baseline Selection

Criteria for the baseline algorithm selection include the availability and quality of ancillary satellite based VI and disturbance recovery inputs during the SMAP mission period, and the relative effectiveness of the algorithm options for enhancing product accuracy, science utility and reliability. Selection of the final L4\_C baseline algorithm version will also follow the results of algorithm tests involving comparisons against in situ observations and model sensitivity studies using pre-cursor satellite observations and reanalysis inputs. If an algorithm option provides a significant improvement in L4\_C accuracy, reliability or science utility and can be implemented without an unsupported cost burden to the mission, then the algorithm will be selected as the baseline, subject to approval from the SMAP ADT and SDS change control board.

## 4. CONSTRAINTS, LIMITATIONS AND ASSUMPTIONS

The L4 Carbon algorithms incorporate a number of simplifying assumptions consistent with a global satellite remote sensing product and may not sufficiently characterize all the major processes regulating CO<sub>2</sub> exchange. For example, soil decomposition studies indicate that the carbon assimilation efficiency of soil microbes and associated SOC decomposition rates vary with changes in soil nitrogen availability (Agren et al. 2001), and may not be adequately represented by a biome-specific optimum soil decomposition rate ( $K_{mx}$ ). Tower based studies over a northern temperate grassland site show large increases in vegetation photosynthetic light-use efficiencies and GPP during years with increased summer precipitation and soil moisture (Flanagan et al. 2002). At a mature boreal black spruce forest site, automated sampling and isotopic analysis of soil respiration indicate that  $R_h$  from deeper (> 20 cm depth) soil layers increases with soil warming, with a significant respiration contribution from older (centuries before present) SOC sources (Hirsch et al. 2003). These processes may not be well represented by regional GPP estimates and near surface freeze/thaw, soil temperature and moisture conditions from relatively coarse scale satellite remote sensing measurements and model assimilation data.

The L4\_C algorithm framework assumes that spatial and temporal variability in the relative magnitude and sign of land-atmosphere CO<sub>2</sub> exchange are largely driven by changes in photosynthetic leaf area and the environmental constraints to NPP, and surface soil wetness and temperature variations through direct environmental controls on  $R_h$ . The L4\_C productivity calculation has extensive heritage from the EOS MODIS MOD17 operational product, which provides for relatively accurate global mapping and monitoring of GPP and NPP (e.g. Heinsch et al. 2006, Zhao and Running 2010). Previous studies have also shown that surface soil temperature and moisture information can be retrieved with reasonable accuracy over heterogeneous landscapes from relatively coarse resolution satellite microwave remote sensing time series (Jones et al. 2007, Njoku et al. 2003), or obtained directly from observation constrained global reanalysis data (Yi et al. 2011). Other studies indicate that the freeze/thaw retrieval from satellite microwave remote sensing provides an effective measure of landscape

water mobility and associated frozen temperature constraints to vegetation productivity over the global cryosphere (Kimball et al. 2004, 2006, Kim et al. 2012). These data provide surrogate measures of primary environmental controls on vegetation productivity, soil decomposition and  $R_h$  across a broad range of northern and global biomes (Kimball et al. 2009a, 2011, Yi et al. 2013).

The L4\_C baseline algorithms additionally assume that surface SOC stocks are in relative equilibrium with these environmental conditions and NPP. This steady-state assumption produces a carbon neutral biosphere (long term cumulative net ecosystem-atmosphere  $CO_2$  exchange (NEE) = 0). Disturbance and recovery effects on L4\_C carbon flux calculations are partially accounted for through associated impacts to ancillary FPAR or VI (option) inputs. Forward model sensitivity simulations and regional evaluations of the baseline steady-state simulations indicate reasonable accuracy at daily, seasonal and annual time scales relative to tower observations and more detailed process model approaches for northern biomes (Kimball et al. 2009a, Yi et al. 2013, McGuire et al. 2012). However, an analysis of flux tower records across several biomes suggests that carbon source-sink strength at most locations is impacted by disturbance history, which adjusts above and below ground carbon stocks away from steady-state conditions (Baldocchi 2008). Alternative modeling approaches incorporate a relaxed steady-state assumption that first estimates steady-state conditions and then perturbs the system using an empirical ‘disturbance parameter’ based on the magnitude and sign of cumulative long term NEE fluxes observed by eddy flux tower measurements over representative biomes or inverted from atmospheric  $CO_2$  concentrations (Carvalhais *et al.* 2008, Rayner *et al.* 2005). The L4\_C disturbance recovery algorithm would use ancillary satellite remote sensing based disturbance or burned area products to track the timing and extent of burned area and general vegetation recovery stage on a grid cell-wise basis; carbon fluxes would be adjusted for disturbance and recovery effects using empirical parameters established from global tower  $CO_2$  eddy flux measurement networks (Baldocchi 2008, Carvalhais et al. 2010). Investigation of non-steady state fire disturbance and recovery effects on L4\_C calculations over a northern ( $\geq 45^\circ N$ ) domain showed relatively large impacts on NEE and component carbon fluxes following large fire events as determined from boreal tower chronosequence networks and satellite (GFED) based burned area products (Yi et al. 2012). However, these effects declined rapidly within the first 5-10 years following disturbance. These effects were also relatively small compared with regional temperature variability and drought impacts within the 11 year (2000-2010) simulation period. The L4\_C steady-state simulations were also found to produce similar accuracy relative to northern tower observation and model inversion based carbon fluxes, and more detailed process model simulations representing non-steady state conditions (McGuire et al. 2012). Nevertheless, fire disturbance during the SMAP mission period is expected to cause large deviations between estimated (baseline) carbon fluxes and actual conditions for the affected grid cells. The relative impact of disturbance is expected to be larger with increasing fire severity or duration, and for biomes with a larger component of woody vegetation cover.

Land cover and land use changes (LCLUC) from direct and indirect human impacts exert a large influence on NEE over a global domain, but are expected to have less impact over sparsely populated northern land areas. Satellite remote sensing based studies indicate that LCLUC from deforestation accounts for up to 1-3%  $yr^{-1}$  of forested land area in tropical regions (Lepers et al. 2005); these biome changes combined with urban and agricultural conversions have a substantial influence on global NEE patterns and recent trends (IPCC

2007). The current (C5) MODIS MOD15 algorithm and FPAR product series uses a static global land cover classification and does not explicitly represent disturbance and LCLUC impacts; disturbance and LCLUC impacts are only partially accounted for through associated changes to photosynthetic canopy cover represented by NDVI inputs to the LAI/FPAR (MOD15) algorithm. Land cover and land use changes occurring during the SMAP mission period would effectively alter biome (BPLUT) response characteristics of the affected region and would not be adequately represented using a static (baseline) ancillary land cover classification in the L4\_C algorithms, resulting in reduced NEE accuracy. The MODIS MOD12Q1 land cover classification and MOD44A vegetation cover conversion products are produced at annual and 16-day time periods (Friedl et al. 2010, Zhan et al. 2002). These data could be applied for periodic (e.g. annual) updating of ancillary land cover inputs and resulting BPLUT and SOC characteristics for more explicit representation of LCLUC impacts to the L4\_C NEE calculations; however, these MODIS products may be unavailable during the SMAP mission period, while similar operational products are not currently part of the VIIRS land product list on NPP or JPSS. The relative accuracy, uncertainty and consistency of available land cover change products are also uncertain over the global domain.

The potential productivity contribution and soil insulation effects of organic ground cover to NEE are not distinguished in the planned L4\_C algorithm apart from the FPAR and L4\_SM inputs and general land cover properties specified in the BPLUT. The Nitrogen (N) content of leaf litter and associated impacts to NPP,  $R_h$  and NEE are also not distinguished in the L4\_C algorithm apart from general land cover properties specified in the BPLUT. These model uncertainties are included in the L4\_C algorithm error budget analysis and indicate that model input and parameterization errors contribute up to 43% of total NEE uncertainty (RMSE), though model error due to lack of a distinct organic layer and litter N representation is a smaller (i.e.  $\leq 12\%$ ) component of the total relative error contribution estimated in **Table 8**. These results indicate that the baseline L4\_C algorithm structure produces an NEE accuracy that is within the uncertainty of tower  $\text{CO}_2$  flux measurements ( $\text{RMSE} \leq 30 \text{ g C m}^{-2} \text{ yr}^{-1}$ ), even without a distinct organic ground cover representation.

The inhibiting effects of low soil moisture on soil  $\text{CO}_2$  fluxes are included in the L4\_C algorithm, primarily through a non-linear soil heterotrophic response to surface soil moisture changes. High soil moisture conditions, especially in wetlands and boreal and arctic biomes, have been associated with reduced  $\text{CO}_2$  production by aerobic decomposition and respiration processes, and enhanced methane ( $\text{CH}_4$ ) production by anaerobic decomposition. Studies supporting  $R_h$  reduction under saturated soil conditions are largely based on controlled incubation experiments and extended inundation periods, while evidence is less consistent from studies involving natural, landscape level observations and heterogeneous surface conditions, including tower eddy covariance measurement footprints. Few wetland tower sites are available for robust model development and calibration. Pre-launch L4\_C algorithm simulations also indicated decreased NEE accuracy (relative to the baseline) by imposing an alternative (convex parabolic) soil moisture response curve and greater soil moisture constraints under saturated soil conditions relative to global FLUXNET tower records. The L4\_C algorithms therefore assume no soil moisture constraint to  $R_h$  under saturated soil conditions. The L4\_C calculations may therefore overestimate  $R_h$  and underestimate NEE carbon ( $\text{CO}_2$ ) sink strength for some areas under extended inundation conditions.

The SMAP L-band active/passive microwave measurements are sensitive to surface soil moisture and water inundation, while soil moisture and temperature influence both aerobic and anaerobic respiration processes; therefore, SMAP data and extension of L4\_C model framework are potentially useful for regional mapping and monitoring of both CO<sub>2</sub> and CH<sub>4</sub> fluxes (e.g., Watts et al. 2014). While CH<sub>4</sub> is a significant greenhouse gas and wetlands are a major component of northern and global ecosystems, CH<sub>4</sub> is beyond the scope of current L4\_C algorithm and product specifications.

The L4\_C baseline algorithm will employ GMAO defined daily minimum air temperature and L3\_SM\_A defined frozen temperature constraints to GPP with L4\_SM based surface (<10 cm depth) soil temperature inputs to define the R<sub>h</sub> response to soil temperature. The algorithm defines cold temperature constraints to NPP and heterotrophic respiration but assumes no soil temperature constraint on R<sub>h</sub> above a biome (BPLUT) specific optimum temperature threshold. Under warmer soil conditions, low soil moisture is assumed to be the primary constraint to R<sub>h</sub>. These assumptions may not hold in warm and relatively moist climate conditions (e.g. tropical rainforest), leading to potential overestimation of NPP and R<sub>h</sub>, and enhanced soil decomposition and underestimation of SOC. The net effect of these errors on NEE is less clear, but likely mitigated by compensating changes in GPP and respiration components.

The L4\_C soil decomposition algorithm is based on the assumption that the bulk of R<sub>h</sub> is derived from surface soil layers. This assumption generally holds for most ecosystems, including boreal-arctic biomes, because the bulk of annual litter decomposition is composed of relatively recent (i.e. <5 years old) leaf litter that is more labile than older soil litter layers. However, deeper soil layers can contribute up to 40% or more of total R<sub>h</sub>, especially later in the growing season as the seasonal warming of deeper layers progresses and lags behind shallower soil layers (Hirsch et al. 2003). The contribution of deeper SOC layers to R<sub>h</sub> may also increase over longer (decadal) time periods in boreal-Arctic regions due to the large reservoir of SOC stored in these colder soils and potential warming and destabilization permafrost and deeper SOC layers under global warming (Schuur et al. 2009). The L4\_C algorithm error budget analysis and model comparisons against regional and global tower observation based carbon fluxes (Kimball et al. 2009a, 2011, Yi et al. 2013, McGuire et al. 2012) indicate algorithm performance within the specified product accuracy requirements (RMSE ≤ 1.6 g C m<sup>-2</sup> d<sup>-1</sup> or 30 g C m<sup>-2</sup> yr<sup>-1</sup>). The baseline algorithm appears adequate to capture NEE seasonal and interannual variations over a 3-5 year SMAP mission life. However, the current algorithm would likely need to represent the R<sub>h</sub> contributions of deeper soil layers over longer time periods. The L4\_SM product will provide both surface and root zone soil moisture and temperature information down to 1m soil depth, so the potential exists for estimating R<sub>h</sub> and associated NEE contributions from deeper soil layers.

Sub-grid scale land cover heterogeneity is a potential source of L4\_C algorithm uncertainty, where landscape variability in land cover conditions and NEE may not be adequately represented by the 9-km grid cell resolution of the baseline L4\_C product. However, land cover heterogeneity effects will be reduced through the use of finer (1-km resolution) scale processing and representation of associated land cover cohorts within each 9-km product grid cell. The finer scale processing and representation of sub-grid scale heterogeneity defined from the ancillary (1 km resolution) global land cover classification inputs is also more consistent with the footprint of most tower (FLUXNET) observation sites, facilitating more robust algorithm calibration and product validation activities.

Additional algorithm uncertainty is contributed by coarse scale GMAO and L4\_SM daily surface meteorology inputs, which may not adequately represent local terrain variability and associated microclimate effects. This uncertainty will be estimated through pre-launch evaluations of sub-grid scale land cover and terrain heterogeneity defined from finer scale land products, and documented in the L4\_C product QA/QC metrics and associated metadata. These effects will also be partially mitigated using finer (1-3-km resolution) scale frozen temperature constraints to L4\_C based NPP calculations provided by the ancillary L3\_SM\_A inputs.

The L4\_C algorithms use a single set of land cover specific coefficients from a look-up table (BPLUT) with ancillary climate, land cover and FPAR inputs to estimate spatial and temporal variations in NEE and component carbon fluxes over a global domain. Calibration of the BPLUT PFT response characteristics using a limited set of global tower (FLUXNET) CO<sub>2</sub> flux monitoring sites likely does not represent the full range of variability in climate and vegetation conditions over the global domain, and may result in reduced accuracy and unspecified uncertainty in underrepresented areas. The BPLUT approach has extensive heritage and has been successfully used for similar satellite based global products, including the EOS MODIS operational (MOD17) GPP product (Zhao and Running 2010). However, use of singular coefficients to describe heterogeneous processes may lead to model prediction error where the underlying population response is skewed or multi-modal and not well represented by a single mean response characteristic. Ideally a Bayesian approach would be better suited to represent sub-grid scale population variability (and uncertainty) in model response characteristics. Here we refer to the use of a Bayesian approach to estimate the BPLUT parameters, and provide a distribution of model outcomes based on specified BPLUT input parameter distributions (e.g. *Figure 11*). However, these approaches are currently computationally prohibitive for an operational global product. The characteristic distributions of many of the BPLUT parameters are also uncertain based on the current literature and sparse tower observation networks.

The L4\_C product is intended to reduce uncertainty regarding the boreal carbon (CO<sub>2</sub>) sink on land (NRC 2007). However, NEE is an incomplete representation of CO<sub>2</sub> source-sink activity because it does not account for anthropogenic carbon emissions or terrestrial carbon losses due to fire, harvesting, and other disturbance (Baldocchi 2008). However, the SMAP L4\_C products will be appropriate for use with sparse station observations (e.g., FLUXNET, NOAA CMDL), regional fire, and fossil fuel emission estimates to initialize, constrain, and optimize atmospheric transport model inversions of atmospheric CO<sub>2</sub> for regional to global assessment and monitoring of terrestrial CO<sub>2</sub> source-sink activity. The NOAA CarbonTracker represents one such carbon data assimilation framework for tracking global CO<sub>2</sub> exchange for policy makers, industry, scientists, and the general public (Peters et al. 2007). Current carbon assimilation systems, including CarbonTracker, rely on ecosystem model based estimates of NEE derived from coarse (~1 degree resolution or greater) reanalysis based surface meteorology inputs with large uncertainty, particularly over northern latitudes (e.g., Zhao et al. 2005, Zhang et al. 2007, Yi et al. 2014). The L4\_C framework will provide NEE inputs with enhanced spatial resolution, temporal fidelity and accuracy over current methods, with additional information on underlying moisture and thermal controls to land-atmosphere CO<sub>2</sub> exchange. Atmospheric transport model inversions using the L4\_C outputs as prior conditions would enable quantification of terrestrial CO<sub>2</sub> source/sink activity (an NRC science objective) and additional validation of L4\_C outputs against other CO<sub>2</sub> observational records (e.g. OCO-2, CO<sub>2</sub> flask network) and baseline conditions from existing land models.



## 5. REFERENCES

- Ågren, G.I., E. Bosatta, and A.H. Magill, 2001. Combining theory and experiment to understand effects of inorganic nitrogen on litter decomposition. *Oecologia* 128, 94–98.
- Altor, A. E., and W. J. Mitsch. 2008. Methane and carbon dioxide dynamics in wetland mesocosms: effects of hydrology and soils. *Ecological Applications* 18, 1307-1320.
- Amthor, J.S., 2000. The McCree-de Wit-Penning de Vries-Thornley respiration paradigms: 30 years later. *Ann. Botany* 86, 1–20.
- Arevalo, C.B.M., J.S. Bhatti, S.X. Chang, R.S. Jassal, and D. Sidders, 2010. Soil respiration in four different land use systems in north central Alberta, Canada. *J. Geophys. Res.* 115, G01003.
- Baker, I., A.S. Denning, N. Hanan, L. Prihodko, M. Uliasz, P-L. Vidale, K. Davis, and P. Bakwin, 2003. Simulated and observed fluxes of sensible and latent heat and CO<sub>2</sub> at the WLEF-TV tower using SiB2.5. *Global Change Biology* 9, 1262-1277.
- Baldocchi, D. 2008. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Austr. J. Bot.* 56: 1–26.
- Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Holinger, S. Running, et al. 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82: 2415–2434.
- Bartholome, E., and A.S. Belward, 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* 26, 9, 1959-1977.
- Berg, A. A., J. S. Famiglietti, M. Rodell, R. H. Reichle, U. Jambor, S. L. Holl, P. R. Hauser. 2005. Development of a hydrometeorological forcing data set for global soil moisture estimation. *Internat. J. Climatol.* 25: 1697-14.
- Brodzik, M.J., B. Billingsley, T. Haran, B. Raup, and M.H. Savoie, 2012. EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets. *ISPRS Int. J. Geo-Inf*, 1, 32-45.
- Bunnell, F.L. and D.E.N. Tait, 1974. Mathematical simulation models of decomposition processes. In: Holding A.J., Heal O.W., MacLean S.F. Jr, Flanagan P.W. (eds) Soil organisms and decomposition in tundra. Tundra Biome Steering Committee, Stockholm, Sweden, pp. 207–225.
- Carvalhais, N., M. Reichstein, J. Seixas, et al., 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. *Global Biogeochem. Cycles* 22 (GB2007) doi:10.1029/2007GB003033.
- Carvalhais, N., M. Reichstein, P. Ciais, et al., 2010. Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints. *Global Change Biology* 16, 2813-2829.

- Chen, B., N.C. Coops, D. Fu et al., 2012. Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program network using remote sensing and footprint analysis. *Rem. Sens. Environ.* 124, 742-755.
- Chimner, R.A. 2004. Soil respiration rates of tropical peatlands in Micronesia and Hawaii. *Wetlands* 24, 51-56.
- Colmer, T.D., and H. Greenway. 2005. Oxygen transport, respiration, and anaerobic carbohydrate catabolism in roots in flooded soils. H. Lambers and M. Ribas-Carbo (eds.), *Plant Respiration*, 137-158. Springer, Netherlands.
- Crow, W. T. 2007. A novel method for quantifying value in spaceborne soil moisture retrievals. *J. Hydromet.* 8: 56–67.
- Dargaville, R. A.D. McGuire, and P. Rayner. 2002. Estimates of large-scale fluxes in high latitudes from terrestrial biosphere models and an inversion of atmospheric CO<sub>2</sub> measurements. *Climatic Change* 55: 273–285.
- Davidson, E.A., E. Belk, and R.D. Boone. 1998. Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. *Global Change Biology* 4, 217-227.
- Davidson, E.A., L.V Verchot, J.H. Cattânio, I.L. Ackerman, and J.E.M. Carvalho, 2000. Effects of soil water content on soil respiration in forests and cattle pastures of eastern Amazonia. *Biogeochemistry* 48(1), 53–69.
- Davis, K.J., P.S. Bakwin, C. Yi, B.W. Berger, C. Zhao, R.M. Teclaw, and J.G. Isebrands, 2003. The annual cycles of CO<sub>2</sub> and H<sub>2</sub>O exchange over a northern mixed forest as observed from a very tall tower. *Global Change Biology* 9, 1278-1293.
- DeLucia, E.H., J.E. Drake, R.B. Thomas, M. Gonzalez-Meler, 2007. Forest carbon use efficiency: is respiration a constant fraction of gross primary production? *Glob. Change Biol.* 13(6), 1157–1167.
- Denman, K.L., G. Brasseur, A. Chidthaisong, P. Ciais, P.M. Cox, R.E. Dickinson, D. Hauglustaine, C. Heinze, E. Holland, D. Jacob, U. Lohmann, S Ramachandran, P.L. da Silva Dias, S.C. Wofsy and X. Zhang. 2007: Couplings Between Changes in the Climate System and Biogeochemistry. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, N. Hanasaki. 2006. GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Am. Met. Soc.* 87: 1381-97.
- Doran, J.W., L.N. Mielke, and J. F. Power. 1990. Microbial activity as regulated by soil water-filled pore space. P. 94-99. Trans. 14 Int. Congr. of Soil Science, Kyoto, Japan, vol. 3.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, P. Kumar. 2000. A catchment-based approach for modeling land surface processes in a general circulation model 2. Parameter estimation and model demonstration. *J. Geophys. Res.* 105 (D20): 24,823-38.

- Duerr, R. et al. 2011. Utility of identification schemes for digital Earth science data: An assessment and recommendations. *Earth Science Informatics*, Springer-Verlag, DOI: 10.1007/s12145-011-0083-6.
- Du, J., J.S. Kimball, M. Azarderakhsh, R.S. Dunbar, M. Moghaddam, and K.C. McDonald, 2014. Classification of Alaska spring thaw characteristics using satellite L-band Radar remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, DOI: 10.1109/TGRS.2014.2325409.
- Elberling, B., L. Askaer, C.J. Jørgensen, H.P. Joensen, M. Kühl, R. N. Glud, and F. R. Lauritsen. 2011. Linking Soil O<sub>2</sub>, CO<sub>2</sub>, and CH<sub>4</sub> concentrations in a wetland soil: implications for CO<sub>2</sub> and CH<sub>4</sub> fluxes. *Environmental Science and Technology* 45, 3393-3399.
- Entekhabi, D., E. Njoku, P. Houser, M. Spencer, T. Doiron, J. Smith, R. Girard, S. Belair, W. Crow, T. Jackson, Y. Kerr, J. Kimball, R. Koster, K. McDonald, P. O'Neill, T. Pultz, S. Running, J. C. Shi, E. Wood, and J. van Zyl, 2004. The Hydrosphere State (HYDROS) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw. *IEEE Trans. on Geosci. and Remote Sensing*, 42(10), 2184-2195.
- Entekhabi, D., E. Njoku, P. O'Neill, K. Kellogg, W. Crow, W. Edelstein, J. Entin, S. Goodman, T. Jackson, J. Johnson, J. Kimball, J. Piepmeier, R. Koster, K. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. Spencer, S. Thurman, L. Tsang, J. Van Zyl, 2010. The Soil Moisture Active and Passive (SMAP) Mission. *Proceedings of the IEEE*, 98(5), 704-716.
- Fang, C., and J.B. Moncrieff, 2001. The dependence of soil CO<sub>2</sub> efflux on temperature. *Soil Biol. Biochem.*, 33(2), 155–165.
- Field, C.B., J.T. Randerson, and C.M. Malmstrom, 1995. Global net primary production: Combining ecology and remote sensing. *Remote Sensing of Environment* 51, 74-88.
- Fily, M. A. Royer, K. Goita, and C. Prigent, 2003. A simple retrieval method for land surface temperature and fraction of water surface determination from satellite microwave brightness temperatures in sub-Arctic areas. *Remote Sens. Environ.* 83(3): 328–338.
- Flanagan, L.B., L.A. Wever and P.J. Carlson, 2002. Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland. *Glob. Change Biol.* 8, 599–615.
- Friedl M.A., D.K. McIver, J.C.F Hodges *et al.*, 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.*, 83(1–2), 287–302.
- Friedl, M.A., D. Sulla-Menashe, B. Tan, et al., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114, 168-182,
- Friedl, M., X. Zhang and A. Strahler, 2011. Characterizing global land cover type and seasonal land cover dynamics at moderate spatial resolution with MODIS data. In: Land Remote Sensing and Global Environmental Change, Remote Sensing and Digital Image Processing, Vol. 11, Part 5, 709-724. DOI: 10.1007/978-1-4419-6749-7\_31.
- Ganguly, S., Samanta, A., Schull, M.A., Shabanov, N.V., Milesi, C., Nemani, R.R., Knyazikhin, Y., Myneni, R.B., 2008. Generating vegetation leaf area index Earth system data record from

- multiple sensors. Part 2: Implementation, analysis and validation. *Remote Sensing of Environment*, 112: 4318-4332.
- Gifford, R.M., 2003. Plant respiration in productivity models: conceptualization, representation and issues for global terrestrial carbon-cycle research. *Funct. Plant Biol.*, 30, 171–186.
- Giglio, L., T. Loboda, D.P. Roy, B. Quayle, and C.O. Justice, 2005. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sensing of Environment* 113, 2, 408-420.
- Giglio, L., J.T. Randerson, G.R. van der Werf, P.S. Kasibhatla, G.J. Collatz, D.C. Morton, and R.S. DeFries, 2010. Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosciences* 7, 1171-1186.
- Global Soil Data Task, 2000. Global Soil Data Products CD-ROM (IGBP-DIS). CD-ROM. International Geosphere-Biosphere Programme, Data and Information System, Potsdam, Germany. Available from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. [<http://www.daac.ornl.gov>].
- Goulden M.L., A.M.S. McMillan, G.C. Winston, A.V. Rocha, K.L. Manies, J.W. Harden, B.P. Bond-Lamberty, 2011: Patterns of NPP GPP, respiration, and NEP during boreal forest succession. *Global Change Biol.*, 17, 855-871.
- Hansen, M.C., R.S. DeFries, J.R.G. Townshend and R. Sohlberg, 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sensing* 21, 6 & 7, 1331-1364.
- Heinsch, F.A., M. Zhao, S.W. Running, J.S. Kimball, R.R. Nemani, et al. 2006. Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. *IEEE Transactions in Geoscience and Remote Sensing* 44(7): 1908–1925.
- Hicke, J.A., G.P. Asner, J.T. Randerson, C. Tucker, S. Los, R. Birdsey, J.C. Jenkins, and C. Field, 2002. Trends in North American net primary productivity derived from satellite observations, 1982-1998. *Global Biogeochemical Cycles* 16(2), 10.1029/2001GB001550.
- Hirsch, A.I., S.E. Trumbore, and M.L. Goulden, 2003. Direct measurement of the deep soil respiration accompanying seasonal thawing of a boreal forest soil. *J. Geophys. Res.* 108(D3), 8221, doi:10.1029/2001JD000921.
- Houghton, R. A. 2003. Why are estimates of the terrestrial carbon balance so different? *Global Change Biol.* 9: 500–9.
- House, J.I., I.C. Prentice, N. Ramankutty, R.A. Houghton and M. Heimann, 2003. Reconciling apparent inconsistencies in estimates of terrestrial CO<sub>2</sub> sources and sinks. *Tellus*, 55B, 345–363.
- Hope, A.S., J.B. Fleming, G. Vourlitis, D.A. Stow, W.C. Oechel, and T. Hack, 1995. Relating CO<sub>2</sub> fluxes to spectral vegetation indices in tundra landscapes: Importance of footprint definition. *Polar Rec.*, 31, 177, 245–250.
- Huete, A.R., H. Liu and W.J.D. van Leeuwen, 1997. The use of vegetation indices in forested regions: issues of linearity and saturation. *IEEE Trans. Geosci. Rem. Sens.* 4, 1966-1968.

- Ino, Y., and M. Monsi, 1969. An experimental approach to the calculation of CO<sub>2</sub> amount evolved from several soils. *Jpn. J. Bot.* 20, 153–188.
- Ise, T., and P.R. Moorcroft. 2006. The global-scale temperature and moisture dependencies of soil organic carbon decomposition: an analysis using a mechanistic decomposition model. *Biogeochemistry* 80: 217–231.
- Jackson, T.J., D.M. Le Vine, A.Y. Hsu, A. Oldak, P.J. Starks, C.T. Swift, J.D. Isham, and M. Haken. 1999. Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains Hydrology Experiment. *IEEE Trans Geosci. Rem. Sens.*, 37(5): 2136–2151.
- Jackson, T.J., J.S. Kimball, R. Reichle, W. Crow, A. Colliander, and E. Njoku, 2012. SMAP Science Calibration and Validation Plan. SMAP Science Document, No. 014, Version 1.2 (Preliminary Release), JPL D-52544, 93 pp.
- Jones, C.D., P. Cox, and C. Huntingford, 2003. Uncertainty in climate-carbon-cycle projections associated with the sensitivity of soil respiration to temperature. *Tellus* 55(2), 642-648.
- Jones, L.A., J.S. Kimball, K.C. McDonald, S.K. Chan, E.G., Njoku, and W.C. Oechel. 2007. Satellite microwave remote sensing of boreal and arctic soil temperatures from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing* 45(7): 2004–2018.
- Jones, L.A., C.R. Ferguson, J.S. Kimball, K. Zhang, S.K. Chan, K.C. McDonald, E.G. Njoku, and E.F. Wood, 2010. Satellite microwave remote sensing of daily land surface air temperature minima and maxima from AMSR-E. *IEEE Journal of Selected Topics in Earth Observations and Remote Sensing (JSTARS)* 3(1), 111-123.
- Jones, L.A., and J.S. Kimball, 2010. Daily Global Land Surface Parameters Derived from AMSR-E. Boulder Colorado USA: National Snow and Ice Data Center. Digital media (<http://nsidc.org/data/nsidc-0451.html>).
- Jung, M., M. Reichstein, H.A. Margolis, et al., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite and meteorological observations. *J. Geophys. Res.* 116, G00J07.
- Kasischke, E.S., N.L. Christensen, and B.J. Stocks, 1995. Fire, global warming and the carbon balance of boreal forests. *Ecological Applications* 5(2), 437-451.
- Kim, Y., J. S. Kimball, J. Glassy, and K. C. McDonald. 2010, Updated 2012. *MEaSUREs Global Record of Daily Landscape Freeze/Thaw Status, Version 2, [1979-2010]*. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center. Digital media (<http://dx.doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0477.002>).
- Kim, Y., J.S. Kimball, K.C. McDonald, and J. Glassy, 2011. Developing a global data record of daily landscape freeze/thaw status using satellite passive microwave remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3), 949-960, doi: 10.1109/TGRS.2010.2070515.
- Kim, Y., J.S. Kimball, K. Zhang, and K.C. McDonald, 2012. Satellite detection of increasing northern hemisphere non-frozen seasons from 1979 to 2008: Implications for regional vegetation growth. *Remote Sensing of Environment* 121, 472-487.

- Kimball, J.S., L.A. Jones, K. Zhang, F.A. Heinsch, K.C. McDonald, and W.C. Oechel. 2009. A satellite approach to estimate land-atmosphere CO<sub>2</sub> exchange for Boreal and Arctic biomes using MODIS and AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2), 569-587, DOI:10.1109/TGRS.2008.2003248.
- Kimball, J.S., K.C. McDonald, S.W. Running, and S. Frolking, 2004. Satellite radar remote sensing of seasonal growing seasons for boreal and subalpine evergreen forests. *Remote Sensing of Environment* 90, 243-258.
- Kimball, J.S., K.C. McDonald, and M. Zhao, 2006. Spring thaw and its effect on terrestrial vegetation productivity in the western Arctic observed from satellite microwave and optical remote sensing. *Earth Interactions* 10(21), 1-22.
- Kimball, J.S., R. Reichle, K.C. McDonald, and P. O'Neill, 2009. External review of the SMAP Level 4 carbon algorithm (L4\_C), Version 1.0, SMAP Science Document, No. 034, Version 1.0, 20 pp.
- Kimball, J.S., Y. Yi, L.A. Jones, R. Reichle, and K.C. McDonald, 2011. Monitoring of net ecosystem CO<sub>2</sub> exchange for the Soil Moisture Active Passive Mission. *Proceedings of the 34<sup>th</sup> International Symposium on Remote Sensing of Environment*. April 10-15, 2011, Sydney Australia, <http://www.isprs.org/proceedings/2011/ISRSE-34/211104015Final00370.pdf>.
- Knorr, W. and J. Kattge. 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biol.* 11: 1333–51.
- Knorr, W., I.C. Prentice, J.I. House, and E.A. Holland, 2005. Long-term sensitivity of soil carbon turnover to warming. *Nature* 433, 298–301.
- Knyazikhin, Y., J. Glassy, J.L. Privette, Y. Tian, A. Lotsch, Y. Zhang, Y. Wang, J.T. Morisette, P. Votava, R.B. Myneni, R.R. Nemani, S.W. Running, 1999. MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document, [http://modis.gsfc.nasa.gov/data/atbd/atbd\\_mod15.pdf](http://modis.gsfc.nasa.gov/data/atbd/atbd_mod15.pdf).
- Krishnan, P., T.A. Black, N.J. Grant, A.G. Barr, E.H. Hogg, R.S. Jassal, and K. Morgenstern, 2006. Impact of changing soil moisture distribution on net ecosystem productivity of a boreal aspen forest during and following drought. *Agricultural and Forest Meteorology* 139 (3-4), 208-223.
- Kurc, S.A., and E.E. Small, 2007. Soil moisture variations and ecosystem-scale fluxes of water and carbon in semiarid grassland and shrubland. *Water Resources Research* 43, W06416.
- Kutzbach, L., C. Wille, and E.-M. Pfeiffer. 2007. The exchange of carbon dioxide between wet arctic tundra and the atmosphere at the Lena River Delta, Northern Siberia. *Biogeosciences* 4, 869-890.
- Lafleur, P.M. T.R. Moore, N.T. Roulet, and S. Frolking. 2005. Ecosystem respiration in a cool temperate bog depends on peat temperature but not water table. *Ecosystems* 8, 619-629.
- Lepers, E. et al., 2005. A synthesis of information on rapid land-cover change for the period 1981-2000. *Bioscience* 55, 2, 115-124.

- Litton, C.M., J.W. Raich, and M.G. Ryan, 2007. Carbon allocation in forest ecosystems. *Glob. Change Biol.* 13, 2089–2109.
- Liu, Q., R. H. Reichle, R. Bindlish, M. H. Cosh, W. T. Crow, R. de Jeu, G.J.M. De Lannoy, G. J. Huffman, and T. J. Jackson, 2011. The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *Journal of Hydrometeorology* 12, 750-765, doi:10.1175/JHM-D-10-05000.
- Lloyd, J., and J.A. Taylor, 1994. On the temperature dependence of soil respiration. *Functional Ecol.* 8(3), 315–323.
- Luo, L., A. Robock, K.E. Mitchel, et al., 2003. Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains. *J. Geophys. Res.* 108 (D22): doi: 10.1029/2002JD003246.
- Mahecha, M.D., M. Reichstein, N. Carvalhais et al. 2010. Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science* 329, 838-840.
- Mäkelä, A., and H.T. Valentine, 2001. The ratio of NPP to GPP: evidence of change over the course of stand development. *Tree Physiol.* 21(14), 1015–1030.
- Mäkiranta, P., R. Laiho, H. Fritze, J. Hytönen, J. Laine, and K. Minkkinen. 2009. Indirect regulation of heterotrophic peat soil respiration by water level via microbial community structure and temperature sensitivity. *Soil Biology and Biochemistry* 41, 695-703.
- Masuoka, E., D. Roy, R. Wolfe, et al., 2011. MODIS land data products: Generation, quality assurance and validation. In: *Land Remote Sensing and Global Environmental Change, NASA's Earth Observing System and the Science of ASTER and MODIS*. B. Ramachandran, C.O. Justice and M.J. Abrams (Eds.), Volume 11, Part 5, 509-531. Springer.
- McDonald, K.C., J.S. Kimball, E. Njoku, R. Zimmermann, and M. Zhao, 2004. Variability in springtime thaw in the terrestrial high latitudes: Monitoring a major control on the biospheric assimilation of atmospheric CO<sub>2</sub> with spaceborne microwave remote sensing. *Earth Interactions* 8(20), 1-23.
- McGuire, A.D., L.G. Anderson, T.R. Christensen, S. Dallimore, L. Guo, D.J. Hayes, M. Heimann, T.D. Lorenson, R.W. MacDonald, and N. Roulet, 2009. Sensitivity of the carbon cycle in the Arctic to climate change. *Ecological Monographs* 79(4), 523-555.
- McGuire, A.D., T.R. Christensen, D. Hayes, A. Heroult, E. Euskirchen, J.S. Kimball, C. Koven, P. Lafleur, P.A. Miller, and W. Oechel, 2012. An assessment of the carbon balance of arctic tundra: Comparisons among observations, process models, and atmospheric inversions. *Biogeosciences Discuss* 9, 4543-4594.
- McMichael, C.E., A.S. Hope, D.A. Stow, J.B. Fleming, G. Vourlitis, and W. Oechel. 1999. Estimating CO<sub>2</sub> exchange at two sites in Arctic tundra ecosystems during the growing season using a spectral vegetation index. *International Journal of Remote Sensing*, 20(4): 683–698.
- Mikan, C.J., J.P. Schimel and A.P. Doyle, 2002. Temperature controls of microbial respiration in arctic tundra soils above and below freezing. *Soil Biol. Biochem.* 34(11), 1785–1795.
- Murphy, R.E., 2006. The NPOESS Preparatory Project. In: J.J. Qu, W. Gao, M. Kafatos, R.E. Murphy, and V.V. Salomonson (Eds.), *Earth Science Satellite Remote Sensing Vol. 1: Science and Instruments*, Springer Berlin Heidelberg, pp. 182-198.

- Murphy, R.E., P. Ardanuy, F.J. Deluccia, J.E. Clement, and C.F. Schueler, 2006. The Visible Infrared Imaging Radiometer Suite. In: J.J. Qu, W. Gao, M. Kafatos, R.E. Murphy and V.V. Salomonson (Eds.), *Earth Science Satellite Remote Sensing Vol. 1: Science and Instruments*, Springer Berlin Heidelberg, pp. 199-223.
- National Research Council (NRC). 2007. Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond (Executive Summary), <http://www.nap.edu/catalog/11820.html>. National Academy of Sciences, National Academies Press, Washington DC, 35pp.
- Nemani, R.R., C.D. Keeling, H. Hashimoto, W.M. Jolly, S.C. Piper, C.J. Tucker, R.B. Myneni, and S.W. Running, 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* 300 (5625), 1560-1563.
- Njoku, E.G., T.J., Jackson, V. Lakshmi, S.K.Chan, and S.V. Nghiem. 2003. Soil moisture retrieval from AMSR-E. *IEEE Trans Geosci. Rem. Sens.*, 41(2): 215–229.
- Northrop Grumman Space Technology (NGST), 2011. Joint Polar Satellite System (JPSS) VIIRS Vegetation Index (VVI) Algorithm Theoretical Basis Document (ref Y2400). NASA GSFC Document Number D43757, 82 pp.
- Oberbauer, F.F., C.T. Gillespie, W. Cheng, R. Gebauer, A. Sala Serra, J.D. Tenhunen, 1992. Environmental effects on CO<sub>2</sub> efflux from riparian tundra in the northern foothills of the Brooks Range, Alaska, USA. *Oecologia* 92: 568–577.
- Oberbauer, S.F., J.D. Tenhunen, J.F. Reynolds, 1991. Environmental effects on CO<sub>2</sub> efflux from water track and tussock tundra in Arctic Alaska, USA. *Arct. Alp. Res.* 23: 162–169.
- Oberbauer, S.F., W. Cheng, C.T. Gillespie, B. Ostendorf, A. Sala, G. Gebauer, R.A. Virginia, and J.D. Tenhunen, 1996. Landscape patterns of carbon dioxide exchange in tundra ecosystems. In: J.F. Reynolds, and J.D. Tenhunen (Eds.) *Ecological Studies*, Vol. 120, Springer-Verlag Berlin Heidelberg, pp. 223–256.
- Parton, W.J., D.S. Schimel, C.V. Cole, and D.S. Ojima, 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Sci Soc Am J.* 51: 1173–1179.
- Peters, W., A.R. Jacobson, C. Sweeney, A.E. Andrews, T.J. Conway, K. Masarie, J.B. Miller, L.M.P. Bruhwiler, G. Petron, A.I. Hirsch, D.E.J. Worthy, G.R. van der Werf, J.T. Randerson, P.O. Wennberg, M.C. Krol, and P.P. Tans. 2007. An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *PNAS* 104(48): 18925–18930.
- Piao, S., P. Ciais, P. Friedlingstein, P. Peylin, M. Reichstein, S. Luyssaert, H. Margolis, J. Fang, A. Barr, A. Chen, A. Grelle, D.Y. Hollinger, T. Laurila, A. Lindroth, A.D. Richardson, and T. Vesala, 2007. Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* 451, 49–52 doi:10.1038/nature06444.
- Potter, C., S. Klooster, R. Myneni, V. Genovese, P-N. Tan, and V. Kumar. 2003. Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982–1998. *Global and Planetary Change*, 39(304): 201–213.
- Potter, C.S., J.T. Randerson, C.B. Field, P.A. Matson, P.M. Vitousek, H.A. Mooney, and S.A. Klooster, 1993. Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles* 7(4), 811-841.

- Prince, S.D., S.N. Goward, 1995. Global primary production: A remote sensing approach. *Journal of Biogeography* 22(4-5), 815-835.
- Randerson, J.T., M.V. Thompson, C.M. Malmstrom, C.B. Field, and I.Y. Fung, 1996. Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO<sub>2</sub>. *Global Biogeochemical Cycles*. 10: 585-602.
- Raich, J. W., W. H. Schlesinger. 1992. The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus* 44B: 81-99.
- Raich, J. W., C. S. Potter, D. Bhabawati. 2002. Interannual variability in global soil respiration, 1980-94. *Global Change Biology* 8: 800-12.
- Raupach, M.R., P.J. Rayner, D.J. Barrett, et al., 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* 11: 378-97.
- Rayner, P.J., M. Scholze, W. Knorr, et al. 2005. Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles* 19, GB2026, doi:10.1029/2004GB002254.
- Reddy, R., and R. D. DeLaune. 2008. Adaptation of Plants to Soil Anaerobiosis. *Biogeochemistry of Wetlands*, 215-256. CRC Press, Boca Raton, FL.
- Reichle, R. H., R. D. Koster, J. Dong, A. A. Berg. 2004. Global soil moisture from satellite observations, land surface models, and ground data: implications for data assimilation. *J. Hydromet.* 5: 430-42.
- Reichle, R.H., R.D. Koster, G.J.M. De Lannoy, B.A. Forman, Q. Liu, S.P.P. Mahanama, and A. Toure, 2011. Assessment and enhancement of MERRA land surface hydrology estimates. *Journal of Climate* 24(15).
- Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research - Atmospheres*, 112, D09108, doi:10.1029/2006JD008033.
- Renzullo, L.J., D.J. Barrett, A.S. Marks, M.J. Hill, J.P. Guerschman, Q. Mu, S.W. Running. 2008. Multi-sensor model-data fusion for estimation of hydrologic and energy flux parameters. *Rem. Sens. Environ.* 112: 1306-19.
- Reynolds, C.A., T.J. Jackson, and W.J. Rawls, 1999. Estimated available water content from the FAO Soil Map of the World, Global Soil Profile Databases, Pedo-transfer functions. Data from the USDA Agricultural Research Service. Published by the NOAA National Geophysical Data Center, Boulder CO. <http://www.ngdc.noaa.gov/seg/fliers/se-2006.shtml>. <http://www.ngdc.noaa.gov/ecosys/cdroms/reynolds/reynolds/reynolds.htm>.
- Richardson, A. D, D. Y. Hollinger. 2005. Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty applied to three simple models. *Ag. For. Meteor.* 131: 191-208.

- Richardson, A.D, M.D. Mahecha, E. Falge, *et al.*, 2008. Statistical properties of random CO<sub>2</sub> flux measurement uncertainty inferred from model residuals. *Ag. For. Meteor.* 148: 38-50.
- Ritchie, E., A. Barros, R. Bell, A. Braun, R. Houghton, B.C. Johnson, G. Liu, J. Luo, J. Morrill, D. Posselt, S. Powell, W. Randel, T. Strub, and D. Vandemark, 2013. NASA Earth Science Senior Review 2013, 93 pp. <http://science.nasa.gov/media/medialibrary/2013/07/16/2013-NASA-ESSR-FINAL.pdf>.
- Roy, D.P., L. Boschetti, C.O. Justice and J. Ju, 2008. The collection 5 MODIS burned area product - Global evaluation by comparison with the MODIS active fire product. *Remote Sensing of Environment* 112, 9, 3690-3707.
- Running, S.W., D.D. Baldocchi, D.P. Turner, S.T. Gower, P.S. Bakwin, and K.A. Hibbard, 1999. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment* 70, 108-127.
- Running, S.W., R. Nemani, J.M. Glassy, P.E. Thornton, 1999. MODIS Daily Photosynthesis (PSN) and Annual Net Primary Production (NPP) Product (MOD17): Algorithm Theoretical Basis Document, Version 3.0. 59 pp. Available online: [http://modis.gsfc.nasa.gov/data/atbd/atbd\\_mod16.pdf](http://modis.gsfc.nasa.gov/data/atbd/atbd_mod16.pdf).
- Running, S.W., R.R. Nemani, F.-A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto. 2004. A continuous satellite-derived measure of global terrestrial primary production. *BioScience* 54(6): 547–560.
- Sacks, W. J., D. S. Schimel, R. K. Monson. 2007. Coupling between carbon cycling and climate in a high-elevation subalpine forest: a model-data fusion analysis. *Oecologia* 151: 54–68.
- Schaefer, K., C.R. Schwalm, C. Williams, et al. 2012. A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis. *J. Geophys. Res.* 117, G03010.
- Schuur, E.A.G., J.G. Vogel, K.G. Crummer, H. Lee, J.O. Sickman, and T.E. Osterkamp, 2009. The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* 459, 556-559.
- Sims, D.A., A.F. Rahman, V.D. Cordova, B.Z. El-Masri, D.D. Baldocchi, P.V. Bolstad, L.B. Flanagan, A.H. Goldstein, D.Y. Hollinger, L. MissOechel, H.P. Schmid, S.C. Wofsy, and L. Xu, 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment* 113, 4, 1633-1646.
- Smith, N.V., S.S. Saatchi, and J.T. Randerson, 2004. Trends in northern latitude soil freeze and thaw cycles from 1988 to 2002. *Journal of Geophysical Research* 109, D12101.
- Stöckli, R., T. Rutishauser, D. Dragoni, J. O’Keefe, P.E. Thornton, M. Jolly, L. Lu, and A.S. Denning, 2008. Remote sensing data assimilation for a prognostic phenology model. *Journal of Geophysical Research* 113, G04021.
- Sukhinin, A.I., N.H.F. French, E.S. Kasischke, J.H. Hewson, A.J. Soja, I.A. Csiszar, E.J. Hyer, T. Loboda, S.G. Conard, V.I. Romasko, E.A. Pavlichenko, S.I. Miskiv, and O.A. Slinkina, 2004. AVHRR-based mapping of fires in Russia: New products for fire management and carbon cycle studies. *Remote Sensing of Environment* 93, 4, 546-564.

- Tarnocai, C., J.G. Canadell, E.A.G. Schuur, P. Kuhry, G. Mazhitova, and S. Zimov, 2009. Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochemical Cycles* 23, GB2023, doi:10.1029/2008GB003327.
- Tarnocai, C., J. Kimble, and G. Broll, 2003. Determining carbon stocks in cryosols using the northern and Mid Latitudes soil database, in *Permafrost*, edited by Phillips, Springman, and Areson, pp. 1129–1134, Swets and Zeitlinger, Lisse.
- Tucker, C.J., D.M. Grant, and J.D. Dykstra, 2004. NASA's global orthorectified Landsat data set. *Photogrammetric Engineering & Remote Sensing*, 70(3), 313-322.
- Turner, D.P., W.D. Ritts, W.B. Cohen, S.T. Gower, S.W. Running, M. Zhao, M.H. Costa, A.A. Kirschbaum, J.M. Ham, S.R. Saleska, and D.E. Ahl., 2006. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment* 102: 282–292.
- Veroustraete, F., H. Sabbe, and H. Eerens, 2002. Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sensing of Environment* 83, 376-399.
- Vourlitis, G.L., W.C. Oechel, A. Hope, D. Stow, B. Boynton, J. Verfaillie Jr., R. Zulueta, and S.J. Hastings, 2000. Physiological models for scaling Plot measurements of CO<sub>2</sub> flux across and Arctic tundra landscape. *Ecological Applications*, 10(1): 60–72.
- Waring, R.H., J.J. Landsberg, and M. Williams, 1998. Net primary production of forests: a constant fraction of gross primary production? *Tree Physiol.* 18, 129–134.
- Watts, J.D., J.S. Kimball, F.-J. W. Parmentier, T. Sachs, J. Rinne, D. Zona, W. Oechel, T. Tagesson, M. Jackowicz-Korczyński, and M. Aurela, 2014. A satellite data driven biophysical modeling approach for estimating northern peatland and tundra CO<sub>2</sub> and CH<sub>4</sub> fluxes. *Biogeosciences*, 11, 1961-1980.
- Webb, R.W., C.E. Rosenzweig, and E.R. Levine 2000. Global soil texture and derived water-holding capacities. Data set. Available online [<http://daac.ornl.gov/SOILS/guides/Webb.html>] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tenn, USA doi:10.3334/ORNLDAAAC/548.
- White MA, P.E. Thornton, S.W. Running, and R.R. Nemani, 2000. Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls. *Earth Interactions*, 4, 1-85.
- Yang, W., D. Huang, B. Tan et al., 2006. Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the Terra MODIS sensor: 2000-2005.
- Yvon-Durocher, G., J.M. Caffrey, A. Cescatti, et al., 2012. Reconciling the temperature dependence of respiration across timescales and ecosystem types. *Nature* 487, 472-476.
- Yi, Y., J.S. Kimball, L.A. Jones, R.H. Reichle and K.C. McDonald, 2011. Evaluation of MERRA land surface estimates in preparation for the Soil Moisture Active Passive Mission. *Journal of Climate* 24(15), 3797-3816.
- Yi, Y., J.S. Kimball, L.A. Jones, R.H. Reichle, R. Nemani, and H.A. Margolis, 2013. Recent climate and fire disturbance impacts on boreal and arctic ecosystem productivity estimated using a satellite-based terrestrial carbon flux model. *J. Geophys. Res. Biogeosci.* 118, 1-17.

- Yi, Y., J.S. Kimball, and R.H. Reichle, 2014. Spring hydrology determines summer net carbon uptake in northern ecosystems. *Environmental Research Letters* 9, 064003.
- Zhan, X., R. Sohlberg, J.R.G. Townshend, C.M. DiMiceli, M.L. Carroll, J.C. Eastman, M.C. Hansen, R.S. Defries (2002), Detection of land cover changes using MODIS 250 meter data. *Rem. Sens. of Environ.*, 83, 336-350.
- Zhang, K., J.S. Kimball, M. Zhao, W.C. Oechel, J. Cassano, and S.W. Running. 2007. Sensitivity of pan-Arctic terrestrial net primary productivity simulations to daily surface meteorology from NCEP/NCAR and ERA-40 Reanalyses. *J. Geophys. Res. - Biogeosciences* 112 (G01011): 1–14.
- Zhang, K., J.S. Kimball, E.H. Hogg, M. Zhao, W.C. Oechel, J.J. Cassano and S.W. Running, 2008. Satellite-based model detection of recent climate driven changes in northern high latitude vegetation productivity. *J. Geophys. Res.* 113, G03033, DOI:10.1029/2007JG000621.
- Zhao, M., and S.W. Running, 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* 329, 5994, 940-943.
- Zhao, M., S. W. Running, R. R. Nemani. 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) to the accuracy of meteorological reanalyses. *J. Geophys. Res.* 111 (G01002), doi: 10.1029/2004JG000004.
- Zhao, M., F.A. Heinsch, R.R. Nemani, and S.W. Running, 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, 95(2), 164-175.
- Zinke, P.J., A.G. Stangenberger, W.M. Post, W.R. Emanuel, and J.S. Olson, 1986. Worldwide organic soil carbon and nitrogen data. ORNL/CDIC-18, NDP-018. Carbon Dioxide Information Analysis Center (CDIC) Numeric Data Collection. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA.
- Zobitz, J.M., D.J.P. Moore, W.J. Sacks, R.K. Monson, D.R. Bowling, D.S. Schimel. 2008. Integration of process-based soil respiration models with whole-ecosystem CO<sub>2</sub> measurements. *Ecosystems* 11: 250–69.

## 6. APPENDIX

*Initial L4\_C algorithm BPLUT used for global product development and testing; parameters are specified for individual plant functional types defined by a MODIS (C5) global land cover classification (Friedl et al. 2010), including evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF), grass (GRS), shrub (SRB), cereal crop (CCRP) and broadleaf crop (BCRP) types. Parameter definitions are provided in **Table 5** of the main text. The initial parameters were established from the literature and subsequently modified based on nested MCMC calibration minimizing model-tower NEE RMSE differences for >80 global tower (FLUXNET) calibration sites representing the global PFT classes.*

Parameter	Units	Plant Functional Type (PFT)							
		ENF	EBF	DNF	DBF	GRS	SHR	CCRP	BCRP
$\epsilon_{mx}$	(g C MJ <sup>-1</sup> )	1.84	1.87	0.75	1.22	1.37	1.72	1.81	2.41
$\text{Min}_{T_{mn}}$	(°C)	256	274	266	263	263	246	253	270
$\text{Max}_{T_{mn}}$	(°C)	300	305	281	286	288	302	300	300
$\text{Min}_{VPD}$	(Pa)	0	0	0	10	725	0	360	3430
$\text{Max}_{VPD}$	(Pa)	3800	7000	3600	6000	4300	4600	6000	6000
$\text{Min}_{SMRZ}$	(% Sat.)	0	0	0	0	0	0	0	0
$\text{Max}_{SMRZ}$	(% Sat.)	84	5	5	5	89	60	21	11
$\text{Min}_{SM}$	(% Sat.)	0	0	0	0	0	0	0	0
$\text{Max}_{SM}$	(% Sat.)	100	100	100	100	100	100	100	5
$F_{FT}$	(DIM)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$NF_{NF}$	(DIM)	1	1	1	1	1	1	1	1
$C_{fract}$	(DIM)	0.49	0.71	0.67	0.67	0.62	0.76	0.78	0.78
$CUE$	(DIM)	0.9	0.8	0.9	0.75	0.9	0.7	0.8	0.55
$R_a \cdot GPP$	(DIM)	0.1	0.2	0.1	0.25	0.1	0.3	0.2	0.45
$K_{mx}$	(d <sup>-1</sup> )	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301
$K_{str} : K_{met}$	(%)	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0
$K_{rec} : K_{met}$	(%)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$F_{str}$	(DIM)	0.3	0.3	0.7	0.3	0.35	0.55	0.5	0.8
$T_{opt}$	(°C)	20	20	20	20	20	20	20	20