SMAP L4 Global Daily 9 km Carbon Net Ecosystem Exchange, Version 1

The Level-4 carbon product (SPL4CMDL) provides global gridded daily estimates of net ecosystem carbon (CO2) exchange derived using a satellite data terrestrial carbon flux model informed by the following: Soil Moisture Active Passive (SMAP) L-band microwave observations, land cover and vegetation inputs from the Moderate Resolution Imaging Spectroradiometer (MODIS) Visible Infrared Imaging Radiometer Suite (VIIRS), and the Goddard Earth Observing System Model, Version 5 (GEOS-5) land model assimilation system. Parameters are computed at 1 km spatial resolution using an Earth-fixed, global, cylindrical 9 km Equal-Area Scalable Earth Grid, Version 2.0 (EASE-Grid 2.0) projection.

Note: These data are Beta-release quality, meaning that they have not undergone full validation and may still contain significant errors.

Overview

Platforms	SMAP Observatory Aqua/Terra Suomi National Polar-orbiting Partnership (NPP) Satellite NASA Goddard Earth Observing System Model, Version 5 (GEOS-5)
Sensors	SMAP L-Band Radiometer Moderate Resolution Imaging Spectroradiometer (MODIS) Visible-Infrared Imaging-Radiometer Suite (VIIRS) NASA Goddard Earth Observing System Model, Version 5 (GEOS-5)
Spatial Coverage	Global, between 85.044°N and 85.044°S
Spatial Resolution	9 km
Temporal Coverage	13 April 2015 – present
Temporal Resolution	Daily
Parameters	Net Ecosystem CO2 Exchange (NEE) Gross Primary Productivity (GPP) Heterotrophic Respiration (Rh) Soil Organic Carbon (SOC) Environmental Constraints (EC)
Data Format	Hierarchical Data Format, Version 5 (HDF5)
Metadata Access	View Metadata Record
Version	V1. Refer to the SMAP Data Versions page for version information. **Maturity State:** Beta **Note:** These data are Beta-release quality, meaning that they have not undergone full validation and may still contain significant errors.
Error Sources	Errors in the land model Errors inherited from a variety of input data sources Errors in the specification of model and observation input error parameters
Get Data	FTP HTTPS Reverb | ECHO Worldview Subscription

Table of Contents

Citing These Data

As a condition of using these data, you must cite the use of this data set using the following citation. For more information, see our Use and Copyright Web page.

Kimball, J., L. A. Jones, J. M. Glassy, and R. Reichle. 2015. *SMAP L4 Global Daily 9 km Carbon Net Ecosystem Exchange.* Version 1. [Indicate subset used]. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:http://dx.doi.org/10.5067/22TFAUSNLO9R. [Date accessed].

1. Detailed Data Description

Format

Data are in HDF5 format. For software and more information, including an HDF5 tutorial, visit the HDF Group's HDF5 Web site.

File Structure

As shown in Figure 1, each HDF5 file is organized into the following main groups, which contain additional groups and/or data sets:

- Environmental Constraints (`EC`)
- Geolocation (`GEO`)
- Gross Primary Production (`GPP`)
- `Metadata`
- Net Ecosystem CO2 Exchange (`NEE`)
- Quality Assessment (`QA`)
- Heterotrophic Respiration (`RH`)
- Soil Organic Carbon (`SOC`)

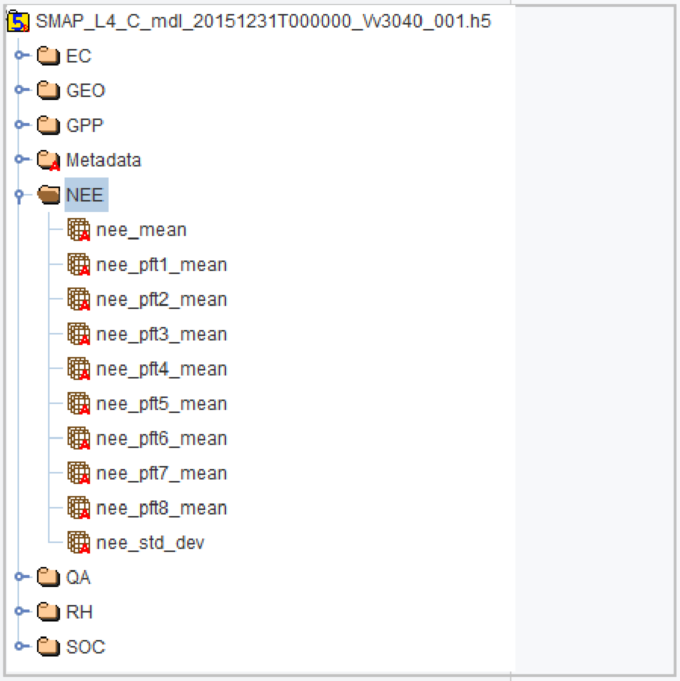

Figure 1. Sample of the HDF5 File Structure

Data Fields Overview

Each Level-4 carbon file contains the following:

EC

Environmental Constraints Data

GEO

Geolocation data, including latitude/longitude coordinate variables in decimal degree units that enable convenient geo-referenced viewing and analysis.

GPP

Gross Primary Production Data

Metadata

Includes all metadata that describe the full content of each file. For a description of all metadata fields for this product, refer to the Metadata Fields document.

NEE

Net Ecosystem CO2 Exchange Data

QA

Includes quality control flags, quality assessment, and valid grid cell counts.

RH

Heterotrophic Respiration Data

SOC

Soil Organic Carbon Data

Data Fields

All global data arrays have dimensions of 1624 rows and 3856 columns (6,262,144 pixels per layer). **Note:** The EASE-Grid 2.0 global 1 km reference grid is defined as 14616 lines by 34704 samples (507,233,664 pixels per layer).

For a complete list and description of all data fields, refer to the Data Fields document.

File Naming Convention

Files are named according to the following convention, which is described in Table 1:

`SMAP_L4_C_MDL_yyyymmddThhmmss_VLMmmm_NNN.[ext]`

For example:

`SMAP_L4_C_mdl_20151007T000000_Vb1010_001.h5`

Where:

Table 1. File Naming Conventions

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td>SMAP</td><td>Indicates SMAP mission data</td></tr>
<tr><td>L4_C_MDL</td><td>Indicates specific product (L4: Level-4; C: Carbon; MDL: Model)</td></tr>
<tr><td>yyyymmddThhmmss</td><td>Date/time in Universal Coordinated Time (UTC) of the first data element that appears in the product, where:
<table>
<tr><td>yyyymmdd</td><td>4-digit year, 2-digit month, 2-digit day</td></tr>
<tr><td>T</td><td>Time (delineates the date from the time, i.e. yyyymmddThhmmss)</td></tr>
<tr><td>hhmmss</td><td>2-digit hour, 2-digit month, 2-digit second</td></tr>
</table></td></tr>
<tr><td>VLMmmm</td><td>Science Version ID, where:
<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td>V</td><td>Version</td></tr>
<tr><td>L</td><td>Launch Indicator (b: Beta-quality data)</td></tr>
<tr><td>M</td><td>1-Digit Major Version Number</td></tr>
<tr><td>mmm</td><td>3-Digit Minor Version Number</td></tr>
</table>
Example: Vb1010 indicates a Beta-quality product with a version of 1.010.</td></tr>
<tr><td>NNN</td><td>Number of times the file was generated under the same version for a particular date/time interval (002: 2nd time)</td></tr>
<tr><td>.[ext]</td><td>File extensions include:
<table>
<tr><td>.h5</td><td>HDF5 data file</td></tr>
<tr><td>.xml</td><td>XML Metadata file</td></tr>
</table></td></tr>
</table>

File Size

Each global file is approximately 118 MB using HDF (gzip) compression.

Volume

The daily data volume is approximately 118 MB.

Spatial Coverage

Coverage spans from 180°W to 180°E, and from approximately 85.044°N and 85.044°S. The gap in coverage at both the North and South Pole, called a pole hole, has a radius of approximately 400 km.

Spatial Resolution

Level-4 carbon model inputs range from 500 m resolution MODIS-based global Plant Functional Type (PFT) classification (from MCD12Q1 Type 5) and 1 km Fraction of Photosynthetically Active Radiation (fPAR) data (from MOD15A2) to 9 km resolution SPL4SM and pre-processed global, daily averaged meteorology data from the ¼ degree GEOS-5 Forward Processing (FP) system. Level-4 carbon model processing is conducted at 1 km EASE-Grid 2.0 resolution consistent with MODIS PFT and fPAR inputs. Level-4 carbon model daily global outputs are gridded using a 9 km EASE-Grid 2.0 projection consistent with SPL4SM inputs, while preserving sub-grid scale PFT means from 1 km resolution processing and MODIS land cover inputs.

Projection and Grid Description

EASE-Grid 2.0

These data are provided on the global cylindrical EASE-Grid 2.0 (Brodzik et al. 2012). Each grid cell has a nominal area of approximately 9 x 9 km2 regardless of longitude and latitude.

EASE-Grid 2.0 has a flexible formulation. By adjusting a single scaling parameter, a family of multi-resolution grids that nest within one another can be generated. The nesting can be adjusted so that smaller grid cells can be tessellated to form larger grid cells. Figure 3 shows a schematic of the nesting to a resolution of 3 km (4872 rows x 11568 columns on global coverage), 9 km (1624 rows x 3856 columns on global coverage) and 36 km (406 rows x 964 columns). Note that the grid used for this product has been adjusted to accommodate a resolution of 1 km.

This feature of perfect nesting provides SMAP data products with a convenient common projection for both high-resolution radar observations and low-resolution radiometer observations, as well as for their derived geophysical products.

For more on EASE-Grid 2.0, refer to the EASE-Grid 2.0 Format Description.

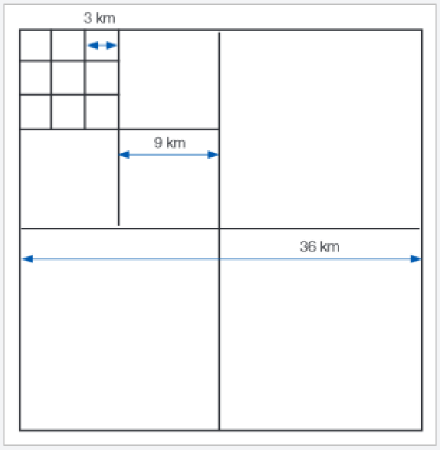

Figure 3. Perfect Nesting in EASE-Grid 2.0

Temporal Coverage

Data were collected from 13 April 2015 through present.

Temporal Resolution

Each Level-4 file is a daily composite.

Parameter Description

This SMAP data product contains daily estimates of global ecosystem productivity, including NEE, GPP, Rh, and SOC, along with quality control metrics. The net ecosystem exchange of CO2 with the atmosphere is a fundamental measure of the balance between carbon uptake by vegetation gross primary production (GPP) and carbon losses through autotrophic (Ra) and heterotrophic (Rh) respiration. The sum of Ra and Rh defines the total ecosystem respiration rate (Rtot), which encompasses most of the annual terrestrial CO2 efflux to the atmosphere. All parameters are expressed in units of g C m-2 d-1. The CO2 flux state variable outputs are provided in the SPL4CMDL granule as NEE/nee_pft{1..8}_mean, GPP/gpp_pft{1..8}_mean, RH/gpp_pft{1..8}_mean, whereas the soil carbon pool state variable output are provided in SOC/soc_pft{1..8}_mean.

As a spatio-temporal data product, the daily global outputs for this data set are organized on the basis of eight vegetated land-cover classes called Plant Function Types (PFTs), as defined within the Terra/Aqua MODIS land instrument team MOD12Q1/MYD12Q1/MCD12Q1 global land cover classification. The PFT two dimensional (2-D) variable is distributed by the Terra/Aqua MODIS team as the Type-5 classification Plant Function Type Scientific Data Set (SDS) in the MCD12Q1 (500 meter resolution) sinusoidal grid 2400 x 2400 tile granules. The original sinusoidal tiles are mosaicked and resampled to form a global 1 km EASEGRID version 2 PFT classification map. PFT classes are used as ancillary driving and spatial partitioning parameters in the SPL4CMDL algorithm, each possessing an entry in the SPL4CMDL Biome Properties Lookup Table (BPLUT); refer to Table 5 in the Algorithm Theoretical Basis Document (ATBD) for this product. The following table summarizes the Plant Function Type classifiers, and highlight the eight PFT classes used by the SPL4CMDL data algorithm to partition SPL4CMDL output variables.

Table 2. Plant Function Type (PFT) Classifier Summary

PFT Class Label	PFT Code	PFT Description	PFT Class used in SPL4CMDL
Water	0	For all ocean and perennial inland water bodies	No
Evergreen needleleaf	1	Evergreen needle-leaf trees (mostly conifers)	Yes
Evergreen broadleaf	2	Evergreen broadleaf trees	Yes
Deciduous needleleaf	3	Deciduous needle-leaf trees	Yes
Deciduous broadleaf	4	Deciduous broad-leaf trees	Yes
Shrub	5	Shrub (woody perennial)	Yes
Grass	6	Grasses (native Graminoids)	Yes
Cereal crop	7	Cereal cropland (domesticated agricultural crops such as wheat, oats, barley, rye)	Yes
Broadleaf crop	8	Broadleaf crop (domesticated agricultural)	Yes
Urban and Built-up	9	Urban and built-up (cities, towns, highways, etc)	No
Snow and ice	10	Snow and ice (may or may not be perennial)	No
Barren (rock) or sparsely vegetated	11	Barren, rock, or very sparsely vegetated land	No
Unclassified	254	Areas otherwise not classified as per above	No

Totals for each vegetated land class (i.e. count of vegetated 1 km nested grid cells in each 9 km grid cell) are provided in each SPL4CMDL granule (QA/qa_count_pft{1..8}). Non-vegetated grid cells are determined by the union of specified vegetation PFT classes in Table 2 and availability of long-term MOD16A2 fPAR for production of the fPAR climatology (refer to the Baseli ne Algorithm). Vegetated PFT grid cells lacking sufficient fPAR retrievals to produce the fPAR climatology and non-vegetated PFT grid cells with valid fPAR climatology are therefore excluded from SPL4CMDL simulations and granule QA counts. QA counts are time-static and therefore identical across granules because the PFT classification does not change within the course of data generation within each SPL4CMDL version. With this information users may calculate the count of non-vegetated 1 km grid cells and PFT percent coverage for each 9 km grid cell.

Refer to the Data Fields document for details on all parameters.

2. Data Access and Tools

Get Data

Data are available via FTP and HTTPS.

Data are also available through the services listed in Table 3.

Table 3. Data Access Services

Service	Description
Reverb | ECHO	NASA search and order tool for subsetting, reprojecting, and reformatting data.
Worldview	NASA visualization tool for browsing full-resolution imagery and downloading the underlying data.
Subscription	Subscribe to have new data automatically sent when the data become available.

Software and Tools

For tools that work with SMAP data, refer to the Tools Web page.

3. Data Acquisition and Processing

This section has been adapted from Kimball et al. (2014), the ATBD for this product.

Sensor or Instrument Description

For a detailed description of the SMAP instrument, visit the SMAP Instrument page at the JPL SMAP Web site.

Data Sources

The following data sources are used as input to calculating this Level-4 carbon product:

- SMAP Level-4 Soil Moisture Data (SPL4SM):
    - SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 1 (SPL4SMAU)

  - SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 1 (SPL4SMGP)
  - SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Land Model Constants, Version 1 (SPL4SMLM)
- GMAO GEOS-5 Forward Processing (FP) Model Data: Daily surface meteorology from observation-constrained global atmospheric model reanalysis
- MODIS Data: Canopy fPAR and land cover classification

In addition, ancillary data sources used as input to calculating this Level-4 carbon product are listed in Table 4.

Theory of Measurements

Current capabilities for regional assessment and monitoring of NEE are limited by mismatches between bottom-up and top-down information sources. Atmospheric transport model inversions of CO2 concentrations from sparse measurement stations provide information on seasonal patterns and trends in atmospheric CO2 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). Tower CO2 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.

Science Objectives

The primary science objectives of the SPL4CMDL product are to:

- Determine NEE regional patterns and temporal (daily, seasonal, and annual) behavior to within the accuracy range of in situ tower measurement estimates of these processes;
- Link NEE estimates with component carbon fluxes (GPP and Rtot) 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 SPL4CMDL 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 CO2 exchange, and the underlying carbon fluxes and environmental drivers of these processes. The SPL4CMDL product will also link SMAP land parameter measurements to global terrestrial CO2 exchange, including boreal ecosystems, reducing uncertainties about the "missing sink" on land for atmospheric CO2.

Atmospheric transport model inversions of CO2 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 CO2 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 microwave sensitivity to land surface processes for monitoring soil moisture and thermal dynamics of global ecosystems, including boreal and arctic biomes.

The baseline SPL4CMDL algorithm uses daily inputs from the SPL4SM products stream to define soil moisture and frozen temperature constraints to vegetation productivity, ecosystem respiration, and NEE. The SPL4CMDL algorithm will provide estimates of NEE (g C m-2 day-1) and component carbon fluxes for global vegetated land areas at mean daily intervals; the product has 9 km spatial resolution, but defines 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-2 yr-1 (1.6 g C m-2 d-1), commensurate with the estimated accuracy of in situ tower measurements (Baldocchi 2008, Richardson 2005, Richardson 2008). The baseline 1 km SPL4CMDL spatial resolution is consistent with the SMAP Level-4 Soil Moisture products (SPL4SMAU, SPL4SMGP, and SPL4SMLM) and is similar to the sampling footprint of CO2 flux measurements from the global tower network (Running et al. 1999, Baldocchi et al. 2008). Secondary products of scientific value produced during SPL4CMDL processing include surface (<10 cm depth) Soil Organic Carbon (SOC) stocks (g C m-2), vegetation Gross Primary Production (GPP), heterotrophic soil and litter respiration (Rh), dimensionless (0-100 percent) frozen area, low temperature and moisture constraint indices for GPP and Rh, and detailed data Quality Assessment (QA) metrics for NEE.

The SPL4CMDL product enables quantification and mechanistic understanding of spatial and temporal variations in NEE over a global domain. NEE represents the primary measure of carbon (CO2) exchange between the land and atmosphere, and the SPL4CMDL product is 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 CO2. The SPL4CMDL product also satisfies carbon cycle science objectives of the NRC Decadal Survey to 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.

For more background information, refer to Section 2.3: Historical Perspective in the ATBD for this product.

Derivation Techniques and Algorithms

This SMAP Level-4 carbon product is derived from a variety of Data Sources. Utilizing the retrieval algorithms discussed below, input sources are converted primarily to net ecosystem exchange (NEE) of CO2 (g C m-2 day-1) between the land and atmosphere on a per-grid-cell basis.

Beta Algorithm Preprocessing

The current Beta-release SPL4CMDL baseline product contains various processing options that are implemented in the algorithm preprocessing stage for handling of the daily model inputs. These processing options are distinct from other options that are more internal to the model algorithms (Kimball et al. 2012). Two major preprocessing options are used in the SPL4CMDL Beta-release product, including use of estimated clear-sky fPAR inputs for missing or lower quality MODIS fPAR inputs, and use of GEOS-5 surface temperature fields to estimate frozen temperature constraints to the GPP calculations instead of SMAP F/T-defined constraints. The use of these preprocessing options are noted in the SPL4CMDL product bit flags as defined on the Data Fields page.

The preprocessing options used in the Beta-release product include a grid cell-wise substitution of a MODIS fPAR 8-day climatology value where the data quality flag information from the operational MODIS fPAR inputs indicate missing or lower quality (QC) cloud contaminated data. The static MODIS global fPAR climatology is part of the ancillary data used for SPL4CMDL processing and was derived on a per grid-cell basis as the mean fPAR value for each 8-day time step over an annual cycle as determined from the best QC MODIS MOD15 fPAR long-term (2000-2012) record. The spatial extent of the global MODIS fPAR climatology also defines the global SPL4CMDL product domain. The use of the fPAR screening process and climatology generally improves model performance, especially in areas with persistent cloud cover, including the tropics and northern boreal/Arctic ecosystems. However, frequent substitution of current fPAR retrievals for alternative climatological values established from a long-term historical record may degrade model sensitivity to seasonal and annual climate variations, impacts for recent climate trends and extreme events, and recent land use and land cover changes. The GPP method quality bit flag information in the SPL4CMDL product provides a record of the spatial distribution and temporal frequency of these substitutions and facilitates future studies to evaluate these impacts.

The SPL4CMDL Beta-product includes the use of a land surface temperature defined F/T frozen flag to define frozen temperature constraints to the model GPP calculations (Kimball et al. 2012). The F/T frozen temperature flag is obtained from the lower order SPL3SMA inputs when available; when the SPL3SMA inputs are missing the F/T frozen flag information is obtained from similar daily surface temperature inputs from the GEOS-5 land model, where temperatures below a 0.0°C threshold are defined as frozen. Future SPL4CMDL product releases will be based on F/T frozen flags defined from SMAP passive radiometer retrievals at 36 km resolution. The major impact of using temperature defined frozen flags is that the F/T flags are derived from relatively coarse land model simulations that are not directly informed by SMAP observations. Future SPL4CMDL product releases will benefit from F/T frozen constraints defined from SMAP radiometer retrievals with enhanced L-band sensitivity to landscape F/T dynamics. These updates are expected to have the greatest benefit in northern ecosystems with frequent frozen conditions, complex terrain, and during seasonal F/T transitions with larger F/T spatial heterogeneity (Du et al. 2014). Employment of SMAP F/T inputs will also enhance SMAP carbon cycle science objectives encapsulated by the SPL4CMDL product, including improving understanding of the net carbon sink in boreal ecosystems (Entekhabi et al. 2010).

Baseline Algorithm

The SPL4CMDL algorithm consists of Light Use Efficiency (LUE) and terrestrial carbon flux model components used to estimate GPP, respiration, residual NEE carbon fluxes, and underlying SOC pools on a daily basis. NEE represents the primary validated SPL4CMDL output product, while the accompanying GPP, respiration and SOC outputs are included as supporting research products. The baseline SPL4CMDL algorithm is summarized in Figures 4a and 4b in the ATBD for respective LUE and carbon flux model components. The 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 the 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 SPL4CMDL algorithm baseline was developed from earlier versions and pre-launch development and testing, and incorporates recommendations from external SPL4CMDL algorithm reviews (for example, Kimball et al. 2009).

Dynamic daily inputs to the SPL4CMDL algorithms include satellite optical-IR remote sensing MODIS-based fPAR, GEOS-5 surface meteorology (Rsw, Tmn, VPD) and associated SPL4SM (SMrz) which provide primary inputs to a LUE algorithm to determine GPP, where Rsw is incoming shortwave solar radiation (MJ m-2 d-1); Tmn is minimum daily 2-m air temperature (°C), VPD is atmosphere vapor pressure deficit (Pa), and SMrz is the integrated surface to root zone (0-1m depth) soil moisture (% Sat.). The SPL4CMDL dynamic inputs also include GEOS-5 surface temperature (Ts, °C) or SMAP sensor freeze-thaw classification defined frozen temperature constraints to productivity and autotrophic respiration calculations. SMAP Level-4 surface soil moisture (≤ 5 cm depth) and soil temperature are used as primary drivers of the soil decomposition and Rh calculations. Static inputs to the SPL4CMDL algorithms include a global land cover classification, which is used to define the major plant functional types and associated biome-specific Biome Properties Lookup Table (BPLUT) response characteristics for each vegetated grid cell within the product domain. The land cover classification used for SPL4CMDL processing is 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 SPL4CMDL domain encompasses all global vegetated land areas with significant vegetation cover, as determined from the MODIS fPAR climatology. 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. Inclusion of 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). Although the SPL4CMDL product is global in extent, product accuracy requirements and validation activities were primarily specified for northern (≥45°N) land areas consistent with NRC objectives for better understanding of terrestrial carbon source/sink activity in boreal regions (NRC 2007, Jackson et al. 2011). The SPL4CMDL calculations are conducted at a daily time step which provides the necessary precision for resolving dynamic boreal vegetation phenology and carbon cycles (Kimball et al. 2009a, Kim et al. 2012). The SPL4CMDL calculations are conducted at 1 km spatial scale consistent with the resolution of ancillary fPAR and land cover inputs. The simulations have also been conducted in a consistent global EASE-GRID (version 2) projection format. Model simulations for each 1 km grid cell were conducted using the corresponding (nearest-neighbor) 9 km resolution SMAP Level-4 Soil Moisture and GEOS-5 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.

The MODIS fPAR operational products are obtained in a tile based sinusoidal projection and preprocessing of these data prior to the SPL4CMDL ingestion involves reprojecting from sinusoidal to 1 km resolution global cylindrical EASE-grid projection formats, followed by trailing nearest-neighbor temporal interpolation of MOD16A2 best Quality Control (QC; cloud-free, favorable surface conditions) 8-day fPAR series to each daily time step. Missing or low QC 8-day fPAR data are gap filled on a grid cell-wise basis using an ancillary fPAR mean 8-day climatology constructed from the long-term (10+ year) MODIS record. The resulting fPAR data are combined with daily biophysical inputs from GEOS-5 and SPL4SM data to estimate NEE, component carbon fluxes (GPP and Rh) and surface SOC pools. SPL4CMDL computes daily Environmental Constraint (EC) indices which influence the GPP and NEE flux calculations, including the estimated bulk environmental reduction to PAR conversion efficiency (εmult), low soil moisture and temperature constraints (Wmult, Tmult) to soil decomposition and Rh calculations and proportional (%) frozen area (FA) within each 9 km grid cell. These constraint indices are provided in SPL4CMDL granules as the EC/emult_av, EC/wmult_av, EC/tmult_av and EC/frozen_area respective SDS fields.

The resulting SPL4CMDL variables enable characterization of spatial patterns and daily temporal fidelity in NEE, underlying carbon fluxes and SOC pools, and their primary environmental drivers. The resulting fine scale (1 km resolution) SPL4CMDL outputs are spatially aggregated to the coarser 9 km resolution final product grid by weighted linear averaging of outputs according to the fractional cover of individual PFT classes represented within each 9 km grid cell and defined by the underlying 1 km resolution MODIS PFT map; the sub-grid scale means from individual PFT classes are preserved for each 9 km grid cell, while proportional vegetation cover information is included in the product metadata, allowing the coarse resolution data to be decomposed into the relative contributions from individual PFT classes within each cell. These outputs are designed to 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_2 flux measurement sites (Baldocchi 2008, Chen et al. 2012).

Algorithm Options

Two algorithm options are considered that affect the SPL4CMDL design relative to the baseline algorithm. The options are summarized in Table 6 of the ATBD 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 initialization 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 (such as VIIRS) potentially enhances reliability of algorithm inputs, but may degrade product latency. Option (2) is expected to enhance scientific merit and SPL4CMDL product accuracy relative to the baseline, but at the expense of increased algorithm complexity.

Ancillary Data

Ancillary data required as input for the algorithms are summarized in Table 4. For in-depth information on ancillary data, refer to the ATBD, Section 3.2: Ancillary Data Requirements.

Table 4. Primary Ancillary Inputs to the SPL4CMDL Algorithm

Parameter	Units	Type	Spatial Resolution	Source
fPAR	%	Dynamic (8-day)	1 km	MODIS (MOD15A2, MYD15A2)
R_{sw}	MJ m - 2 d - 1	Dynamic (daily)	9 km	VIIRS (VVI3P)
T_{mn}	°C	Dynamic (daily)	9 km*	GEOS-5
VPD	Pa	Dynamic (daily)	9 km*	GEOS-5
SM	% Sat.	Dynamic (daily)	9 km	SPL4SM (Includes SPL4SMAU, SPL4SMGP, and SPL4SMLM)
SM_{rz}	% Sat.	Dynamic (daily)	9 km	SPL4SM (Includes SPL4SMAU, SPL4SMGP, and SPL4SMLM)
T_5	°C	Dynamic (daily)	9 km	SPL4SM (Includes SPL4SMAU, SPL4SMGP, and SPL4SMLM)
F/T	Discrete class	Dynamic (3-day)	3 km	SPL3SMA
Land Cover Class	Discrete class	Static	1 km	MODIS (MCD12Q1)
fPAR Climatology	%	Static (8-day)	1 km	MODIS (MOD15A2, MYD15A2)
Additional Inputs for Algorithm Options				
VI (NDVI, EVI)	Dimensionless	Dynamic (8-day)	1 km	MODIS (MOD/MYD13A2), VIIRS (VVI)
Recovery Status	Years	Static	1 km	MODIS (MOD/MYD13A2)

***** The native resolution of GEOS-5 FP fields is ¼ degree (latitude) by 3/8 degree (longitude); SPL4CMDL processing internally re-samples these fields to 9 km.

For more information regarding the algorithm, refer to the ATBD for this product.

Processing Steps

Written by the University of Montana's Numerical Terradynamic Simulation Group (NTSG), the SPL4CMDL science code was transferred from NTSG to the NASA Global Modelling and Assimilation Office (GMAO) for translation and implementation as operational code in conjunction with SPL4SM production within the GMAO Level-4 SMAP Science Data Processing System (SDS). To generate the SPL4CMDL product, the processing software ingests SPL4SMAU daily granules, MODIS-derived 8-day fPAR granules, and GEOS-5 daily surface meteorology data. The ingested data are then inspected for retrievability criteria according to input data quality, ancillary data availability, and land cover conditions. Two pre-processor codes, one for fPAR data and one for global meteorology data, are then executed each day to temporally aggregate and resample these respective inputs for use by the baseline algorithm software. When retrievability criteria are met, the production software invokes the baseline retrieval algorithm to generate the daily carbon model outputs.

Error Sources

Many sources of error contribute to the uncertainty in the SPL4C product. The key input uncertainties to the SPL4C algorithm are:

1. Errors in the ancillary 8 day fPAR inputs
2. Errors in the SPL4SM soil moisture and temperature inputs
3. Errors in the GEOS-5 daily surface meteorology inputs
4. Uncertainty in the internal model parameterization, initialization, and calibration parameters

For more information about error sources refer to the ATBD for this product.

Quality Assessment

These Version 1 Beta data employ preliminary algorithms that are still being validated and are thus subject to uncertainties. For in-depth details regarding the quality of these

Version 1 Beta data, refer to the Beta Assessment Report.

Quality Overview

SMAP products provide multiple means to assess quality. Each product contains bit flags, uncertainty measures, and file-level metadata that provide quality information. For information regarding the specific bit flags, uncertainty measures, and file-level metadata contained in this product, refer to the Data Fields document.

Each HDF5 file contains metadata with Quality Assessment (QA) metadata flags that are set by the GMAO prior to delivery to National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC). A separate metadata file with an `.xml` file extension is also delivered to NSIDC DAAC with the HDF5 file; it contains the same information as the HDF5 file-level metadata. If a product does not fail QA, it is ready to be used for processing, browse generation, active science QA, archive, and distribution. If a file/granule fails QA, the GMAO does not send the granule to NSIDC until it is reprocessed. A QA file is only produced when there are data that qualify for retrieval.

Quality Flags

Quality Assessment (QA) fields are also provided with metadata from MODIS fPAR and SPL4SM inputs to the SPL4CMDL algorithms. These QA fields incorporate expected model uncertainty propagating from input driver uncertainty including SPL4SM, GEOSS-5 FP, and MODIS fPAR. This QA information is assigned on the basis of look-up tables derived from pre-launch algorithm sensitivity runs using documented uncertainties in the SPL4CMDL inputs and were updated by comparing errors relative to global historical flux tower benchmark data during SPL4CMDL calibration. Input errors are propagated during SPL4CMDL 1-km model calculations using standard error propagation procedures employing the SPL4CMDL model Jacobian and simplifying independence assumptions. Resulting 1-km NEE error fields are quadratically averaged to produce 9-km output fields for each PFT. These floating-point continuous fields are then reclassified into three discreet integer quality classes: 3 (lower quality) to 0 (best quality), following similar protocols adopted from other NASA global land products (Kim et al. 2011). The resulting QA information has been evaluated and refined through post-launch SPL4CMDL cal/val activities using carbon flux measurements from in situ measurement networks, comparisons with other similar global carbon products, and algorithm sensitivity studies over the observed range of environmental variability. The above-described QA fields are provided in SPL4CMDL granules as QA/carbon_model_qual_pft{1..8}. Refer to Section 3.3 of the ATBD for additional details.

Quality control bit flags are provided in SPL4CMDL granules to identify retrieval conditions including use of alternative ancillary datasets and exceedance of expected output field value ranges. Alternative ancillary conditions indicated in the QC bit flags include the use of optional Normalized Difference Vegetation Index (NDVI) inputs in place of baseline MODIS fPAR inputs, potential gaps in the SPL3SMA input stream, and instances where the ancillary fPAR 8-day climatology is used in place of the dynamic best QC MODIS fPAR input stream to estimate GPP. Expected PFT class specific range thresholds for each state variable (NEE, GPP, Rh, and SOC) have been established from dynamic algorithm simulations using long-term (10+ year) daily data input records from pre-launch data sources similar to those used for post-launch SPL4CMDL production, including MODIS (MOD/MYD15) fPAR, freeze-thaw status (Kim et al. 2012), and MERRA surface meteorology (Yi et al. 2011). These post-launch diagnostics have been provided in SPL4CMDL granules as the QA/carbon_model_bitflags SDS field for additional user evaluation and potential troubleshooting by the SMAP Science Data System (SDS) and Algorithm (ADT) teams. Table 7 indicates the bit-field positions for the above-described flags. A copy of Table 7 is also provided within each SDS granule as metadata for quick reference.

Table 7. QC Bit Flag Fields, Names, Positions, and Description Metadata

Bit Flag Name	Bit Positions {Start, End}	Number of Bits	Value Range	Description
NEE bit	00 – 00	1	{0|1}	0 = NEE within valid range; 1 = out of valid range
GPP bit	01 – 01	1	{0|1}	0 = GPP within valid range; 1 = out of valid range
Rh bit	02 – 02	1	{0|1}	0 = Rh within valid range; 1 = out of valid range
SOC bit	03 – 03	1	{0|1}	0 = SOC within valid range; 1 = out of valid range
PFT dominant	04 – 07	4	{1..8}	Dominant PFT class
QA score	08 – 11	4	{0,1,2,3}	0 = best quality; 3 = lowest quality
GPP method	12 – 12	1	{0|1}	0 = used MODIS FPAR; 1 = used FPAR climatology
NDVI method	13 – 13	1	{0|1}	0 = used MODIS FPAR or FPAR climatology; 1 = used VIIRS NDVI
F/T method	14 – 14	1	{0|1}	0 = used L3_SM_A F/T; 1 = used GEOSS-5 surface temperature
IsFill*	15 – 15	1	{0|1}	0 = is NOT a FillValue = is a FillValue

* If IsFill=1, then all other bit fields will have value 1 and the entire uint16 integer will evaluate to 65534. Users should therefore check the value of IsFill prior to referencing other bit fields.

For more information, such as algorithm testing procedures and cal/val activities, refer to the ATBD. For more information regarding data flags, refer to the Data Fields document.

4. References and Related Publications

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(1):32-45. http://dx.doi.org/10.3390/ijgi1010032.

Brodzik, M. J., B. Billingsley, T. Haran, B. Raup, and M. H. Savoie. 2014. Correction: Brodzik, M. J. et al. EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. *ISPRS Int. J. Geo-Inf 2012*. 1(1):32-45 *ISPRS Int. J. Geo-Inf.* 3(3):1154-1156. http://dx.doi.org/10.3390/ijgi3031154.

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 CO2 measurements. *Climatic Change* 55:273–285.

Entekhabi, D. et al. 2014. SMAP Handbook–Soil Moisture Active Passive: Mapping Soil Moisture and Freeze/Thaw from Space. Pasadena, CA USA: SMAP Project, JPL CL#14-

2285, Jet Propulsion Laboratory.

Glassy, J., J. S. Kimball, L. Jones, R. H., Reichle, R. A. Lucchesi, J. V. Ardizzone, G.-K. Kim, E. B. Smith, and B. H. Weiss. 2015. Soil Moisture Active Passive (SMAP) Mission Level 4 Carbon (L4_C) Product Specification Document. GMAO Office Note No. 11, Version 1.9. NASA Goddard Space Flight Center, Greenbelt, MD, USA. (https://nsidc.org/data/docs/daac/smap/sp_l4_cmdl/pdfs/Glassy790.pdf, 669 KB)

Kimball, J., L. A. Jones, J.M. Glassy, E.N. Stavros, N. Madani, R. Reichle, T. Jackson, A. Colliander. 2015. Soil Moisture Active Passive (SMAP) Project Calibration and Validation for the L4_C Beta-Release Data Product. Pasadena, CA USA: SMAP Project, Jet Propulsion Laboratory. (http://nsidc.org/data/docs/daac/smap/sp_l4_cmdl/pdfs/Kimball806.pdf, 13.4 MB)

Kimball, J. S., L. A. Jones, J. P. Glassy, and R. Reichle. 2014. SMAP Algorithm Theoretical Basis Document, Release A: L4 Carbon Product. SMAP Project, JPL D-66484, Jet Propulsion Laboratory, Pasadena CA. 76 pp. (https://nsidc.org/data/docs/daac/smap/sp_l4_cmdl/pdfs/271_L4_C_RevA_web.pdf, 2.7 MB)

National Research Council (NRC). 2007. Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond (Executive Summary). National Academy of Sciences, National Academies Press, Washington DC. 35pp. http://www.nap.edu/catalog/11820.html.

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.

5. Contacts and Acknowledgments

Investigators

John Kimball, Lucas Jones, Joe Glassy
Numerical Terradynamic Simulation Group (NTSG)
College of Forestry & Conservation
The University of Montana
Missoula, MT 59812-1049 USA

Rolf Reichle
NASA Goddard Space Flight Center
Global Modeling and Assimilation Office
Mail Code 610.1
8800 Greenbelt Rd.
Greenbelt, MD 20771 USA

Technical Contact

NSIDC User Services
National Snow and Ice Data Center
CIRES, 449 UCB
University of Colorado
Boulder, CO 80309-0449 USA
phone: +1 303.492.6199
fax: +1 303.492.2468
form: Contact NSIDC User Services
e-mail: nsidc@nsidc.org

6. Document Information

Document Creation Date

October 2015

Document Revision Date

N/A

Document URL

`http://nsidc.org/data/docs/daac/smap/sp_l4_cmdl/index.html`