SMAP L4 Global Daily 9 km EASE-Grid Carbon Net Ecosystem Exchange, Version 7

USER GUIDE

How to Cite These Data

As a condition of using these data, you must include a citation:

Kimball, J. S., K. A. Endsley, L. A. Jones, T. Kundig, and R. Reichle. 2022. *SMAP L4 Global Daily 9 km EASE-Grid Carbon Net Ecosystem Exchange, Version 7* [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/3K9F0S1Q5J2U. [Date Accessed].

FOR QUESTIONS ABOUT THESE DATA, CONTACT NSIDC@NSIDC.ORG

FOR CURRENT INFORMATION, VISIT https://nsidc.org/data/SPL4CMDL

National Snow and Ice Data Center

TABLE OF CONTENTS

1 DATA DESCRIPTION

1.1 Parameters

This SMAP data product contains daily estimates of global ecosystem productivity, including net ecosystem exchange (NEE), gross primary production (GPP), heterotrophic respiration (Rh), and soil organic carbon (SOC), along with quality control metrics. The NEE of CO_2 with the atmosphere is a fundamental measure of the balance between carbon uptake by vegetation gross primary production (GPP), and carbon losses through autotrophic respiration (Ra) and heterotrophic respiration (Rh). The sum of Ra and Rh defines the total ecosystem respiration rate (Rtot), which encompasses most of the annual terrestrial CO_2 efflux to the atmosphere. All parameters are expressed in units of g C m^{-2} day^{-1}. For each metric of global ecosystem productivity, the CO_2 flux state variable outputs are provided in SPL4CMDL files as eight vegetated land cover classes called Plant Functional Types (PFTs). For example, the CO_2 flux state variable outputs are provided in *NEE/nee_pft{1..8}_mean*, *GPP/gpp_pft{1..8}_mean*, and *RH/gpp_pft{1..8}_mean*. The soil carbon pool state variables are provided in *SOC/soc_pft{1..8}_mean*. Refer to Table 1 for descriptions of the eight PFTs. Refer to the Appendix of this document for details on all parameters.

Table 1. Plant Functional Type (PFT) Classifier Summary

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

Note on Table 1:

Within each 9 km grid cell the number of 1 km grid cells belonging to each vegetated land class is provided in QA/qa_count_pft{1..8}. Non-vegetated grid cells (i.e. cells where the algorithm is not applied), are determined by combining specified vegetation PFT classes and long-term VIIRS fPAR (VNP15A2) where available. Vegetated PFT grid cells lacking sufficient fPAR retrievals to produce the fPAR climatology and non-vegetated PFT grid cells with otherwise valid fPAR climatology are excluded from SPL4CMDL simulations and QA counts. QA counts are time-static and are therefore identical across files because the PFT classification does not change over the course of data generation within each SPL4CMDL version.

Users may use the QA count information to compute total non-vegetated 1 km grid cell coverage, compute percent coverage for each PFT, and account for non-vegetated regions when computing areal averages from SPL4CMDL state variables. For example, when computing the total GPP within a 9 km grid cell, a user would multiply the mean GPP (i.e. /GPP/gpp_mean in g C m^{-2} d^{-1}) by the vegetated PFT total QA count (i.e. /QA/qa_count).

1.2 File Information

1.2.1 Format

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

1.2.2 File Contents

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

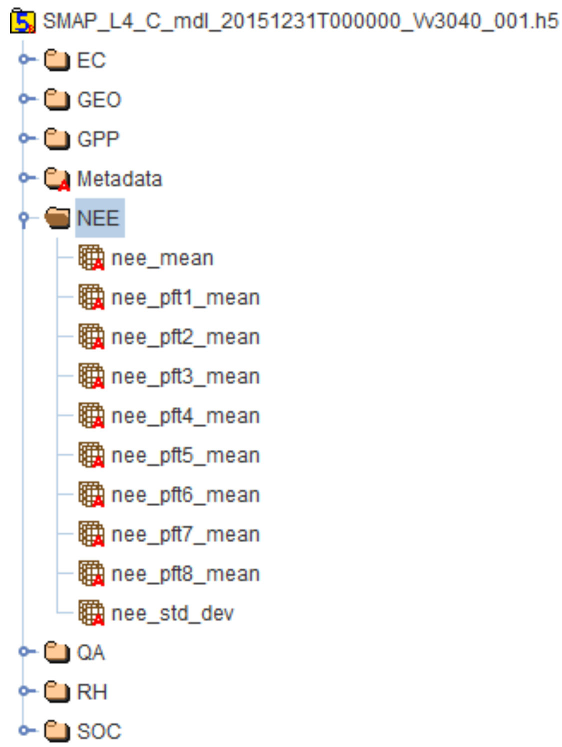

Figure 1. Subset of File Contents. For a complete list of file contents for the SMAP Level-4 carbon product, refer to the Appendix.

1.2.3 Data Fields

Each file contains the main data groups summarized in this section. For a complete list and description of all data fields within these groups, refer to the Appendix of this document.

All global data fields have dimensions of 1624 rows and 3856 columns (6,262,144 pixels per array).

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

NEE

Net Ecosystem CO_2 Exchange Data

QA

QA includes quality control flags, quality assessment, and valid grid cell counts.

RH

Heterotrophic Respiration Data

SOC

Soil Organic Carbon Data

1.2.4 Metadata Fields

For a description of all metadata fields for this product, refer to the Product Specification Document (Endsley et al., 2022b).

1.2.5 File Naming Convention

Files are named according to the following convention:

```
SMAP_L4_C_MDL_yyyymmddThhmmss_VLMmmm_NNN.[ext]
```

For example:

```
SMAP_L4_C_mdl_20151007T000000_Vv3040_001.h5
```

Table 2 describes the variables within a file name:

Table 2. File Naming Convention

<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 minute, 2-digit second</td></tr></table></td></tr>
<tr><td>VLMmmm</td><td>Composite Release ID, where:<table><tr><td>V</td><td>Version</td></tr><tr><td>L</td><td>Launch Indicator (V: Validated Data)</td></tr><tr><td>M</td><td>1-Digit Composite Release ID (CRID) Major Version Number (Note: the data set's major version does not necessarily coincide with the CRID major version)</td></tr></table></td></tr>
</table>

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td></td><td><table><tr><td>mmm</td><td>3-Digit CRID Minor Version Number</td></tr></table>Example: Vv7040 indicates a Validated data product with a version of 7.040.
Refer to the SMAP Version History page for version information.</td></tr>
<tr><td>NNN</td><td>Number of times the file was generated under the same version for a particular date/time interval (e.g., 002 denotes 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>

1.2.6 Data Volume Estimates

Daily SPL4CMDL granules that in Hierarchical Data Format version 5 (HDF5) are approximately 132 MB each. This leads to an approximate total of 48.2 GB for one (1) year of SPL4CMDL data.

1.3 Spatial Information

1.3.1 Coverage

Coverage spans from 180°W to 180°E, and from approximately 85.044°N and 85.044°S.

1.3.2 Resolution

Level-4 carbon model inputs include the following spatial resolutions:

- 500 m resolution MODIS-based global PFT classification (from MCD12Q1 Type 5)
- 500 m Fraction of Photosynthetically Active Radiation (fPAR) data (from VIIRS VNP15A2)
- 9 km resolution SMAP Level-4 soil moisture data (SPL4SMGP)
- ¼ degree pre-processed global, daily averaged meteorology data from the GEOS-5 Forward Processing (FP) system

Level-4 carbon model processing is conducted at 1 km EASE-Grid 2.0 resolution using spatially aggregated MODIS PFT and VIIRS fPAR inputs. Level-4 carbon model daily global outputs are gridded using a 9 km EASE-Grid 2.0 projection consistent with the SMAP L4 soil moisture data used as input.

Note that while this product is posted to a 9-km resolution grid, it also retains sub-grid scale heterogeneity information as determined from the 1 km resolution processing. The resulting product structure includes both the daily spatial mean carbon variables derived over each 9 km resolution grid cell, and the sub-grid spatial means of the carbon fluxes averaged over each PFT class depicted within the same grid cell as defined from the 1 km resolution MODIS (MCD12Q1)

PFT and VIIRS (VNP15A2) fPAR inputs. The multi-layer product structure allows for reconstructing finer scale (1 km resolution) maps of the estimated carbon variables that may be closer to the footprint of in situ measurements (Endsley et al. 2020).

For more details regarding inputs used in the carbon model, refer to the Data Acquisition and Processing section (Section 2.0) of this document.

1.3.3 Geolocation

These data are provided on the 9-km global cylindrical EASE-Grid 2.0 projection. The following tables provide information for geolocating this data set. For more on EASE-Grid 2.0, refer to the EASE Grids website.

Table 3. Geolocation details for the EASE-Grid 2.0 projections used in this product

	Global
Geographic coordinate system	WGS 84
Projected coordinate system	EASE-Grid 2.0 Global
Longitude of true origin	0
Standard Parallel	30° N
Scale factor at longitude of true origin	N/A
Datum	WGS 84
Ellipsoid / spheroid	WGS 84
Units	meter
False easting	0
False northing	0
EPSG code	6933
PROJ4 string	`+proj=cea +lon_0=0 +lat_ts=30 +x_0=0 +y_0=0 +ellps=WGS84 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs`
Reference	http://epsg.io/6933

Table 4. Grid details for the EASE-Grid 2.0 projections used in this product

	Global
Grid cell size (x, y pixel dimensions)	9,024.13 m (x) 9,024.13 m (y)
Number of columns	3856
Number of rows	1624

Geolocated lower left point in grid	85.044° S, 180.000° W
Nominal gridded resolution	9 km by 9 km
Grid rotation	N/A
ulxmap – x-axis map coordinate of the outer edge of the upper-left pixel	-17367530.45
ulymap – y-axis map coordinate of the outer edge of the upper-left pixel	7314540.83

1.4 Temporal Information

1.4.1 Coverage

Coverage spans from 31 March 2015 to 17 June 2025.

1.4.2 Satellite and Processing Events

Due to instrument maneuvers, satellite and data downlink anomalies, data quality screening, and other factors, small gaps in the SMAP time series will occur. Details of these events are maintained on two master lists:

SMAP On-Orbit Events List for Instrument Data Users
Master List of Bad and Missing Data

Significant gaps in coverage occurred between 19 June and 23 July 2019 and between 6 August and 20 September 2022 after the SMAP satellite went into Safe Mode. A brief description of the 2019 event and its impact on data quality is available in the SMAP Post-Recovery Notice. The SMAP data acquired after the 2022 event were determined to be of high quality and consistent with the data acquired prior to the event.

1.4.3 Latencies

Please see the following FAQ: What are the latencies for SMAP radiometer data sets?

1.4.4 Resolution

Each Level-4 file is a daily composite. Calculations for this product are conducted at a daily time step in order to provide the necessary precision for resolving dynamic vegetation phenology and carbon cycles (Kimball et al. 2009, Kim et al. 2012).

2 DATA ACQUISITION AND PROCESSING

This section has been adapted from the Algorithm Theoretical Basis Document (ATBD) for this product (Kimball et al. 2014).

2.1 Background

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 CO_2 concentrations from sparsely distributed measurement stations provide information on seasonal patterns and trends in atmospheric CO_2 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 CO_2 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 (Byrne et al. 2020, Endsley et al. 2022a).

To address these limitations, the primary objectives of the SPL4CMDL product are to:

- Determine NEE regional patterns and temporal behavior (daily, seasonal, and annual) with estimation errors not exceeding those of in situ tower measurements;
- Link NEE estimates with component carbon fluxes (GPP and Rtot) and the primary environmental constraints to ecosystem productivity and respiration.

The SPL4CMDL algorithm supports carbon cycle science objectives by enabling detailed mapping and monitoring of spatial patterns and temporal dynamics of land-atmosphere CO_2 exchange, and the underlying carbon fluxes, soil carbon storage, and environmental drivers of these processes. The SPL4CMDL product also links SMAP land parameter measurements to global terrestrial CO_2 exchange, including boreal ecosystems, reducing uncertainties about the 'missing sink' on land for atmospheric CO_2 (Liu et al. 2022).

Atmospheric transport model inversions of CO_2 concentrations indicate that the Northern Hemisphere terrestrial biosphere is responsible for much of the recent terrestrial sink strength for atmospheric carbon (Forkel et al. 2016, Ciais et al. 2019). Variability in land-atmosphere CO_2 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 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 exchange (as CO_2) 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 fluxes, climate and drought related impacts on vegetation productivity, and atmospheric transport model inversions of terrestrial source-sink activity for atmospheric CO_2.

For more background information, refer to Section 2.3 of the ATBD for this product (Kimball et al., 2014).

2.2 Instrumentation

For a detailed description of the SMAP instrument, visit the SMAP Instrument page at the Jet Propulsion Laboratory (JPL) SMAP website.

2.3 Acquisition

The following data sources are used as inputs to the calculations for this Level-4 Carbon product:

- SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 7 (SPL4SMGP)
- GMAO GEOS-5 Forward Processing (FP) Model Data: Daily surface meteorology from observation-corrected global atmospheric model analysis
- NASA EOS Terra MODIS (MCD12Q1) static land cover classification
- NOAA/NASA Suomi National Polar-Orbiting Partnership (Suomi NPP) VIIRS fPAR 8-day Data, Version 001 (VNP15A2H); if VNP15A2H data are unavailable, the following primary back-up fPAR source is used:
- SMAP L4 Carbon Model ancillary VIIRS fPAR 8-day Climatology

Ancillary data sources used as inputs to the calculation for this Level-4 carbon product are listed in Table 5. For in-depth information on ancillary data, refer to Section 3.2 of the ATBD (Kimball et al., 2014).

Table 5. Primary Ancillary Inputs to the SPL4CMDL Algorithm

Parameter	Description	Units	Type	Spatial Resolution	Source
fPAR	fraction of photosynthetically active radiation	%	Dynamic (8-day)	1 km [IV]	VIIRS (VNP15A2H[I])
R_{sw}	incident solar shortwave radiation	$MJ\ m^{-2}\ d^{-1}$	Dynamic (daily)	9 km [II]	GEOS-5
T_{mn}	minimum air temperature	°C	Dynamic (daily)	9 km [II]	GEOS-5
VPD	vapor pressure deficit	Pa	Dynamic (daily)	9 km [II]	GEOS-5
SM	surface soil moisture	% Sat.	Dynamic (daily)	9 km	SPL4SMGP
SM_{rz}	root zone soil moisture	% Sat.	Dynamic (daily)	9 km	SPL4SMGP
T_s	surface temperature	°C	Dynamic (daily)	9 km	SPL4SMGP
FT	freeze/thaw status	Discrete class	Dynamic (daily)	9 km [II]	GEOS-5 [III]
Land Cover Class	land cover classification	Discrete class	Static	1 km [IV]	MODIS (MOD12Q1)
fPAR Climatology	long-term (10+ yrs) mean fPAR	%	Static (8-day)	1 km [IV]	VIIRS (VNP15A2H[I])

[I] VNP indicates data acquired by the VIIRS instrument on the NOAA/NASA Suomi NPP satellite.

[II] The native resolution of GEOS-5 FP fields is $\frac{1}{4}$ degree (latitude) by $\frac{3}{8}$ degree (longitude); SPL4CMDL processing internally resamples these fields to 9 km.

[III] Due to the loss of the SMAP radar instrument and operational freeze/thaw (FT) classification product, SPL4CMDL uses the GMAO GEOS-5-modeled T_{surf} parameter to define FT conditions in the carbon model.

[IV] Derived from finer scale (500 m resolution) MODIS and VIIRS data records spatially aggregated to 1 km resolution for carbon model processing. The best-quality fPAR retrieval within a 1-km window (4 pixels) is used based on the MODIS and VIIRS product QC flags.

2.4 Derivation Techniques and Algorithms

2.4.1 Baseline Algorithm

The baseline SPL4CMDL algorithm uses daily inputs from the SMAP Level-4 soil moisture stream to define soil moisture and frozen temperature constraints on vegetation productivity, ecosystem respiration, and NEE. The algorithm provides estimates of NEE (g C m^{-2} day^{-1}) and component carbon fluxes for global vegetated land areas at mean daily intervals; the product defines sub-grid scale mean and variability in carbon fluxes for dominant and sub-dominant land cover vegetation (PFT) classes within each grid cell as determined from finer scale ancillary land cover classification and fPAR inputs. The targeted accuracy for the SPL4CMDL product is a mean annual unbiased RMSE (ubRMSE) accuracy for NEE within 30 g C m^{-2} yr^{-1} or 1.6 g C m^{-2} day^{-1}, commensurate with the estimated accuracy of in situ tower measurements (Baldocchi 2008, Richardson and Hollinger 2005, Richardson et al. 2008). The baseline 1-km SPL4CMDL spatial resolution is similar to the sampling footprint of CO_2 flux measurements from the global tower network (Running et al. 1999, Baldocchi 2008). Secondary products of scientific value produced during SPL4CMDL processing include surface (~0-5 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) environmental constraint indices for GPP and Rh, and detailed data Quality Assessment (QA) metrics for NEE.

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. The baseline SPL4CMDL algorithm is summarized in Figure 2 and Figure 3 for respective LUE and carbon flux model components. The approach has structural elements similar to the Century model (Parton et al. 1987, Ise and Moorcroft 2006) and CASA model (Potter et al. 1993) for soil decomposition, and NASA EOS MODIS 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 analysis data (Jones et al. 2017).

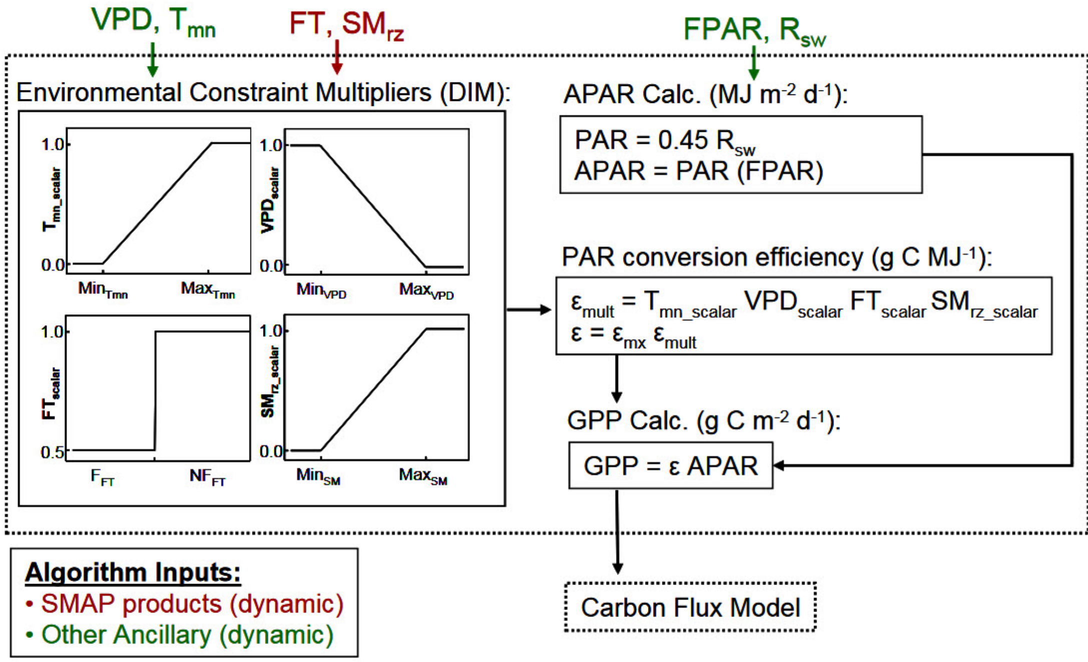

Figure 2. Baseline Light Use Efficiency (LUE) Carbon Model Structure for Estimating GPP

In Figure 2, arrows denote the primary pathways of data flow, while boxes denote the major process calculations. Following left to right in Environmental Constraint Multipliers of Figure 2, inputs include minimum air temperature (Tmn) and daily vapor pressure deficit (VPD) from the GEOS FP product (denoted in green) and daily landscape freeze/thaw (FT) status and root zone soil moisture (SMrz) from the SMAP Level-4 soil moisture product (denoted in red). Additional dynamic ancillary inputs (denoted in green) include VIIRS (VNP15A2H) fPAR and GEOS FP daily incident solar shortwave radiation (Rsw). Model calculations are performed at 1 km spatial resolution using dominant vegetation (PFT) class and Biome Properties Look-Up Table (BPLUT) response characteristics defined for each pixel using a global land cover classification (Jones et al. 2017); see Section 2.6 for detail on the BPLUT used for SPL4CMDL processing. The resulting GPP calculation is a primary input to the Level-4 Terrestrial Carbon Flux (TCF) model summarized below (Figure 3).

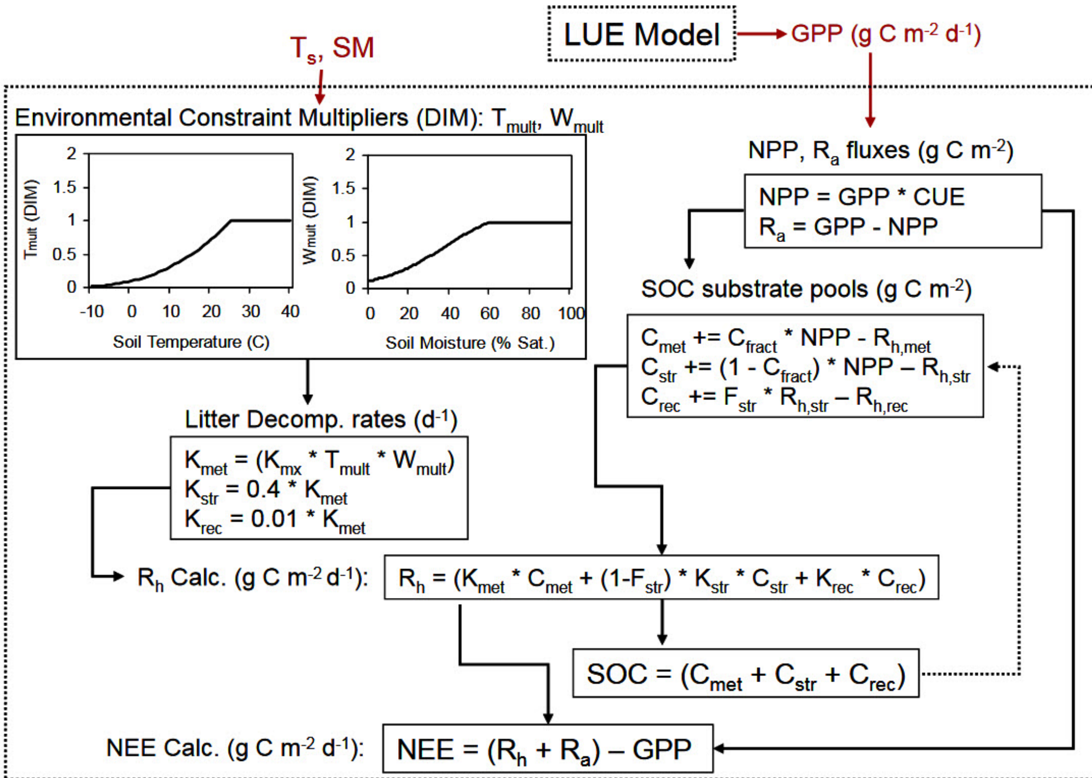

Figure 3. Terrestrial Carbon Flux（TCF）model structure used for estimating NEE

In Figure 3, primary Terrestrial Carbon Flux（TCF）model inputs (denoted in red) include daily GPP from the LUE model, and surface soil moisture (SM) and surface temperature (Ts) from the SMAP Level-4 Soil Moisture (SPL4SMGP) product. NEE is the primary (validated) output, while GPP, respiration (Rh + Ra), and SOC are secondary (research) outputs.

Dynamic daily inputs to the SPL4CMDL algorithms include satellite optical infrared (IR) remote sensing-based fPAR estimates from VIIRS, GEOS FP surface meteorology (Rsw, Tmn, VPD) and associated SPL4SMGP soil moisture (SMrz) which provide primary inputs to a LUE algorithm to determine GPP. 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 (0-1 m depth) soil moisture (% Sat.). The SPL4CMDL dynamic inputs also include GEOS-5 surface temperature (Ts, °C), defined frozen temperature (FT), constraints to GPP, 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 Look-Up Table (BPLUT) response characteristics for each vegetated grid cell within the product domain. The BPLUT parameters are defined for up to eight global vegetation (PFT) classes; the model parameters for each global PFT class were calibrated by optimizing carbon model NEE calculations against tower eddy covariance measurement-based daily NEE observations from global FLUXNET sites* representing the major PFT classes (Baldocchi 2008, Jones et al. 2017). 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 data assimilation (GEOS FP).

The SPL4CMDL parameters 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 land cover map. The mean values at the sub-grid (1-km) scale from individual PFT classes are preserved for each 9-km grid cell, while proportional vegetation cover information is included in the *qa_counts_pft* data field, 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 (Endsley et al. 2020), 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).

Footnote for this section:

* The more recent SPL4CMDL product versions (V4 and later) use an augmented FLUXNET global tower site record which includes more calibration sites (356 sites as of V5, 329 sites for V4, and 228 sites for V3 and earlier), expanded tower data records, and additional tower sites representative of more global land cover types.

2.4.2 Algorithm Options

The 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. 2014). Two major preprocessing options are used in the SPL4CMDL product, which are: 1) use of estimated clear-sky VIIRS fPAR climatology inputs for missing or lower quality VIIRS 8-day fPAR inputs, and 2) use of GMAO GEOS-5 surface temperature fields to estimate frozen temperature constraints to the GPP calculations instead of SMAP radar FT-defined constraints. The use of these preprocessing options is noted in the SPL4CMDL product bit flags as defined in Table 6 of this document and in the Appendix of this document.

For more information regarding the baseline algorithm and its options, refer to the ATBD for this product (Kimball et al. 2014).

2.5 Processing

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 SMAP Level-4 Soil Moisture (SPL4SMGP) production within the GMAO Level-4 SMAP Science Data Processing System (SDS).

To generate the SPL4CMDL product, the processing software:

1. Ingests SPL4SMGP daily files, VIIRS-derived 8-day fPAR files, and GEOS FP daily surface meteorology data.
2. The ingested data are then inspected for retrievability criteria according to input data quality, ancillary data availability, and land cover conditions.
3. 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 SPL4CMDL algorithm software. When retrievability criteria are met, the production software invokes the baseline retrieval algorithm to generate the daily carbon model outputs.

SPL4CMDL calculations are conducted at 1-km resolution, benefiting from finer scale (500 m) VIIRS fPAR and MODIS land cover (PFT) inputs. The simulations are also conducted in a consistent global EASE-Grid 2.0 projection format. Model simulations for each 1-km pixel are run using corresponding daily inputs from the closest (nearest-neighbor) 9-km resolution SPL4SMGP and GEOS FP grid cells. The VIIRS (VNP15A2H) fPAR product is produced at 500 m resolution and 8-day temporal fidelity from the NOAA NASA Suomi NPP VIIRS sensor records. The SPL4CMDL daily carbon flux and state variable calculations for each 1-km pixel are guided by PFT-specific environmental response parameters defined from the ancillary Biome Properties Look-Up Table (BPLUT) and global land cover (MOD12Q1) classification. The BPLUT is recalibrated for optimal performance with each SPL4CMDL product release and available as an ancillary product file (SMAP_L4_C_ANC_BPLUT).

VIIRS fPAR operational products are obtained in a tile-based sinusoidal projection. 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 VNP15A2H 8-day fPAR series to each daily time step, for any retrievals with good Quality Control (QC, e.g., relatively cloud-free with favorable surface conditions). 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 (2012-2021) VIIRS record. The resulting fPAR data are combined with daily biophysical inputs from GEOS FP and SPL4SMGP 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 (ε_{mult}), low soil moisture constraints and low temperature constraints (W_{mult}, T_{mult}) to both soil decomposition and Rh calculations, and freeze/thaw (FT) status within each 9-km grid cell. These environmental constraint indices are provided in SPL4CMDL files as the *EC/emult_mean, EC/wmult_mean, EC/tmult_mean* and *EC/frozen_area* respective data fields.

2.6 Quality, Errors, and Limitations

2.6.1 Error Sources

Many sources of error contribute to uncertainty in the SPL4CMDL product. The key sources of error or uncertainty in the SPL4CMDL algorithm are:

1. Errors in the ancillary 8-day fPAR inputs
2. Errors in the SPL4SMGP soil moisture and temperature inputs
3. Errors in the GEOS FP 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 (Kimball et al. 2014).

2.6.2 Quality Assessment

For in-depth details regarding the quality of the SPL4CMDL product, including the latest product release and earlier versions, refer to Jones et al. (2017), Endsley et al. (2020, 2022a), and the most recent product Assessment Report.

2.6.3 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 Quality Flags section below, the Appendix of this document, or the Product Specification Document (Endsley et al. 2022b).

Each HDF5 file contains metadata with Quality Assessment (QA) metadata flags that are set by the GMAO prior to delivery to the 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 the NSIDC DAAC with the HDF5 file; it contains the same information as the HDF5 file-level metadata.

2.6.4 Quality Flags

Quality Assessment (QA) fields are also provided with metadata from VNP15A2H and SPL4SMGP inputs to the SPL4CMDL algorithms. The QA output incorporates expected model uncertainty propagating from input driver uncertainty including SPL4SMGP, GEOS FP, and VNP15A2H. QA input error information was assigned by comparing unbiased Root Mean Square Error (ubRMSE) differences relative to global historical flux tower benchmark data during SPL4CMDL pre-launch calibration. Input errors are propagated during SPL4CMDL 1-km model calculations using standard error propagation procedures that employ Jacobian and simplifying independence assumptions. Resulting 1-km NEE ubRMSE fields are quadratically averaged to 9 km output fields for each PFT class as defined from the 1-km resolution MOD12Q1 global land cover map and then posted as the NEE QA ubRMSE geophysical variable (g C m^{-2} d^{-1}). The resulting QA information has been evaluated and refined through post-launch SPL4CMDL Cal/Val activities using concurrent eddy covariance CO_2 flux measurements from global tower measurement networks (Baldocchi 2008), comparisons with other similar global carbon products, and algorithm sensitivity studies over the observed range of environmental variability (Jones et al. 2017). The above-described QA fields are provided in SPL4CMDL files as the *QA/nee_rmse_mean* and *QA/nee_rmse_pft{1..8}_mean* fields. Refer to the SPL4CMDL Product Specification Document (Endsley et al., 2022b) for additional details.

Quality control bit flags are provided in the SPL4CMDL files to identify retrieval conditions including use of alternative ancillary data sets and exceedance of expected output field value ranges. Alternative ancillary conditions indicated in the QC bit flags include the use of the alternative fPAR climatology in place of baseline VIIRS (VNP15A2H) 8-day fPAR inputs, and potential gaps in the GEOS FP input stream. 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+ years) daily data input records from pre-launch data sources similar to those used for post-launch SPL4CMDL production, including VIIRS fPAR, freeze-thaw status (Kim et al. 2017), and MERRA surface meteorology (Yi et al. 2011). These post-launch diagnostics are provided in SPL4CMDL files in the *QA/carbon_model_bitflag* data field for additional user evaluation. Definitions for bit-fields for flags described in this section are shown in Table 6; they are also documented in the granule-level metadata in the HDF5 Attribute *legend_carbon_model_bitflag*. Note that another HDF5 Attribute, *flag_meanings*, should be ignored because it is obsolete; it only inadvertently remains in the granule-level metadata of the L4_C Version 7 product and will be removed in future product versions.

Table 6. QC Bit Flag Fields, Names, Positions, and Description Metadata

Bit Flag Name	Bit Positions	Number of Bits	Value Range	Description
NEE bit	00	1	{0|1}	0 = NEE within valid range; 1 = out of valid range
GPP bit	01	1	{0|1}	0 = GPP within valid range; 1 = out of valid range
Rh bit	02	1	{0|1}	0 = Rh within valid range; 1 = out of valid range
SOC bit	03	1	{0|1}	0 = SOC within valid range; 1 = out of valid range
PFT dominant	04 – 07	4	{1..8}	Most frequently occurring (dominant) vegetated PFT class as defined from *qa_count*
QA score	08 – 11	4	{0,1, 2,3}	Relative *nee_mean* error as ranked by *nee_rmse_mean*: 0 = (RMSE<1 g C m^{-2} d^{-1}); 1 = (1<=RMSE<2 g C m^{-2} d^{-1}); 2= (2=<RMSE<3 g C m^{-2} d^{-1}); 3 = (RMSE> = 3 g C m^{-2} d^{-1})
GPP method	12	1	{0|1}	0 = derived GPP using 8-day fPAR input; 1 = derived GPP via fPAR climatology
fPAR source	13	1	{0|1}	0=used MODIS fPAR, 1=used VIIRS fPAR
FT method	14	1	{0|1}	0 = used SPL3SMA FT; 1 = used GEOS-5 surface temperature [SPL3SMA is no longer operational, so bitflag = 1]
IsFill*	15	1	{0|1}	0 = is NOT fill value (simulation performed for one or more 1 km grid cells within 9 km grid cell); 1 = is fill value (no 1 km simulation performed within 9 km grid cell). Fill values occur for non-land, non-vegetated, and/or grid cells otherwise lacking valid fPAR data record.
* When 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.				

Note: 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 SMAP mission science objectives for better understanding of terrestrial carbon source/sink activity in boreal regions (NRC 2007, Entekhabi et al. 2010, Jackson et al. 2011).

For more information, such as algorithm testing procedures, refer to the ATBD (Kimball et al., 2014). For more information regarding data flags, refer to the Appendix of this document.

3 VERSION HISTORY

Table 7. Version History

Version	Release Date	Description of Changes
V1	Oct 2015	First public data release
V2	Apr 2016	Changes to this version include: • Transitioned to Validated-Stage 2 • Using SPL4SMAU V2 Validated and SPL4SMGP V2 Validated data as input • Updated to process radiometer data from 2015-03-31 to 2015-04-12 • Some data fields renamed from *_av to *_mean • Updated to have continuous RMSE-based "quality" fields instead of the categorical quality flag in V1
V3	Jul 2017	Changes to this version include: • Uses dynamic 8-day fPAR inputs obtained from the latest (Collection 6) MODIS fPAR record at 500 m resolution. The preprocessor was updated to handle the finer resolution MODIS Collection 6 inputs, which are interpolated to 1 km resolution prior to model processing. The prior (Version 2) processor used MODIS Collection 5 fPAR inputs, which were derived at 1 km resolution. • Updated the ancillary MODIS fPAR 8-day climatology used for fPAR gap-filling as an L4C model preprocessing step to reflect new MODIS Collection 6 fPAR inputs. The fPAR climatology is derived from a longer 14-year (2000-2014) MODIS record relative to the original 12-year (2000-2012) Collection 5 fPAR record used in Version 2 processing. • For each grid cell, a sine-curve-based seasonal fPAR climatology curve is now used to identify and screen anomalous 8-day fPAR variations in the preprocessor. This change reduces impacts of anomalous fPAR temporal variations that may not be captured by the MODIS fPAR product quality control (QC) flags, particularly during seasonal transitions at northern latitudes. • Updated and recalibrated the ancillary Biome Properties Look-Up Table (BPLUT) and re-initialized the model initial global soil organic carbon (SOC) pools to reflect new MODIS Collection 6 fPAR inputs. The BPLUT calibration was conducted using global historical FLUXNET in situ tower eddy covariance CO2 flux measurement records for representative global land cover types using a similar step-wise calibration procedure employed for the Version 2 product. • A minor bug fix to the post-processor was made to ensure that all grid cell no-data fill values are identified with a consistent -9999 notation; the prior Version 2 product erroneously assigned some no-data values as -999900.
V4	Jun 2018	Changes to this version largely affect model inputs and ancillary files rather than changes to the internal model structure or code. Note that Version 4 is slightly better than Version 3 in RMSE terms, with improvement generally larger for drier sites. Specific changes include:

Version	Release Date	Description of Changes
		• The carbon model biome properties lookup table (BPLUT) has been calibrated using an augmented FLUXNET global tower site record which includes more calibration sites (335 sites compared to 228 sites for V3), expanded tower data records extending to at least 2015, and the addition of new tower sites representing more land cover types. • Revised Level-4 carbon global model calibration and SOC initialization using an extended (2000-2017) MODIS fPAR (V006) record and the latest SMAP Nature Run (NRv7.2) climate data records. • Implemented minor changes to spatial weighting of calibration tower sites within a model grid cell and reduced the outlier influence on model response curve fitting.
V5	Aug 2020	Changes to this version primarily represent changes to the model inputs and model Biome Properties Look-up Table (BPLUT) calibration rather than changes to the internal model structure or code. The Version 5 product accuracy and performance is largely consistent with Version 4. Specific changes include: • Revised Level-4 carbon global model calibration and soil organic carbon (SOC) initialization using the latest SMAP Nature Run (NRv8.3) soil moisture and soil temperature data, and MERRA-2 surface meteorology. • The carbon model BPLUT has been recalibrated using a FLUXNET2015 global tower site record that includes 329 sites representing all major global plant functional type (PFT) classes. • Implemented minor changes to spatial weighting of FLUXNET2015 tower sites in the Level-4 carbon model BPLUT calibration. Northern (≥ 45°N) tower sites now have greater impact in defining global PFT response characteristics; whereby, the adjusted weighting is designed to enhance product performance in boreal-arctic biomes. It also incorporated a more detailed set of PFT realistic range constraints for the BPLUT parameters prior to the model calibration and optimization. • Enhanced an attribute in the *legend_carbon_model_bitflag* data field to include descriptions of the bit values.
V6	Nov 2021	Changes to this version primarily reflect Biome Properties Look-up Table (BPLUT) recalibration and soil organic carbon (SOC) re-initialization in response to updates to the upstream input datasets. Specifically: • Revised SPL4CMDL global BPLUT calibration and SOC initialization using the latest SMAP Nature Run (NRv9.1) soil moisture and soil temperature data, and MERRA-2 surface meteorology. • The carbon model BPLUT, Vv6040, has been recalibrated using the most recent FLUXNET2015 global tower site synthesis record,

Version	Release Date	Description of Changes
		which includes 356 sites representing all major global plant functional type (PFT) classes. • Adjusted model SOC decay parameters based on comparison with other global assessments, which reduced SOC storage estimates in the tropics (20 S – 20 N) from the previous (V5) SPL4CMDL product release.
V7	Nov 2022	Changes to this version primarily involve replacing MODIS fPAR with VIIRS fPAR as the new baseline fPAR input for the carbon model processing; and Biome Properties Look-up Table (BPLUT) recalibration and soil organic carbon (SOC) re-initialization in response to updates to the upstream SPL4SMGP input datasets. Specifically: • Replaced NASA EOS MODIS (MOD15A2H) fPAR with NOAA/NASA Suomi NPP VIIRS (VNP15A2H) fPAR as the new baseline 8-day fPAR inputs to the L4C model; generated a new ancillary global clear-sky fPAR climatology from the long-term (2012-2021) VIIRS record. This update enables extended L4C operations beyond the EOS MODIS era through ongoing VIIRS operations. • Revised SPL4CMDL global BPLUT calibration and SOC initialization using soil moisture and soil temperature data from the latest SMAP L4SM Open Loop (OL7000) ensemble Catchment land model simulation. The revision results in a calibration that is more consistent with the ensemble SPL4SMGP inputs used in SPL4C production than the calibration with the single-member, unperturbed L4SM Nature Run (NR) data used in prior L4C product releases. • The carbon model BPLUT, Vv7040, was recalibrated using the FLUXNET2015 global tower site record, which includes 356 sites representing all major global PFT classes.

4 RELATED DATA SETS

SMAP Data at NSIDC | Overview

SMAP Radar Data at the ASF DAAC

5 RELATED WEBSITES

SMAP at NASA JPL

6 ACKNOWLEDGMENTS

Funding was provided by the NASA SMAP Mission. Computational resources were provided by the NASA High-End Computing program through the NASA Center for Climate Simulation. The authors are grateful for those who make the generation and dissemination of SMAP data products possible, including staff at GSFC, JPL, and NSIDC.

7 REFERENCES

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. https://doi.org/10.1071/BT07151.

Byrne, B., Liu, J., Lee, M., Baker, I., Bowman, K. W., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Iraci, L. T., Kiel, M., Kimball, J. S., Miller, C. E., Morino, I., Parazoo, N. C., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Velazco, V. A., Wennberg, P. O., & Wunch, D. 2020. Improved constraints on northern extratropical CO2 fluxes obtained by combining surface-based and space-based atmospheric CO2 measurements. *JGR Atmospheres* 125(15), e2019JD032029. https://doi.org/10.1029/2019JD032029.

Chen, B., N. C. Coops, D. Fu et al. 2012. Characterizing Spatial Representativeness of Flux Tower Eddy-Covariance Measurements across the Canadian Carbon Prgram Network Using Remote Sensing and Footprint Analysis. *Rem. Sens. Environ.* 124:742-755.

Ciais, P., Tan, J., Wang, X., Roedenbeck, C., Chevallier, F., Piao, S. L., Moriarty, R., Broquet, G., Le Quéré, C., Canadell, J. G., Peng, S., Poulter, B., Liu, Z., & Tans, P. 2019. Five decades of northern land carbon uptake revealed by the interhemispheric CO2 gradient. *Nature*, 568(7751):221–225. https://doi.org/10.1038/s41586-019-1078-6.

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_2 Measurements. *Climatic Change* 55:273–285. https://doi.org/10.1023/A:1020295321582.

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. https://archive.ipcc.ch/publications_and_data/ar4/wg1/en/ch7.html.

Endsley, K. A., Kimball, J. S., Reichle, R. H., & Watts, J. D. 2020. Satellite monitoring of global surface soil organic carbon dynamics using the SMAP Level 4 Carbon Product. *Journal of Geophysical Research: Biogeosciences*, 125(12). https://doi.org/10.1029/2020JG006100.

Endsley, K. A., Kimball, J. S., & Reichle, R. H. 2022a. Soil respiration phenology improves modeled phase of terrestrial net carbon exchange in northern hemisphere. *Journal of Advances in Modeling Earth Systems*, 14(2), e2021MS002804. https://doi.org/10.1029/2021MS002804.

Endsley, K. A., J. Glassy, J. S. Kimball, L. A. Jones, R. H., Reichle, J. V. Ardizzone, G.-K. Kim, R. A. Lucchesi, E. B. Smith, and B. H. Weiss, 2022b: Soil Moisture Active Passive (SMAP) Mission Level 4 Carbon (L4_C) Product Specification Document. GMAO Office Note No. 11 (Version 2.2), 71 pp, NASA Goddard Space Flight Center, Greenbelt, MD, USA (See PDF)

Entekhabi, D., E.G. Njoku, P.E. O'Neill, K.H. Kellogg, W.T. Crow, W.N. Edels ... J. Van Zyl. 2010. The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE* 98 (5), 704-716. https://doi.org/10.1109/JPROC.2010.2043918.

Forkel, M., Carvalhais, N., Rödenbeck, C., Keeling, R., Heimann, M., Thonicke, K., Zaehle, S., & Reichstein, M. 2016. Enhanced seasonal CO2 exchange caused by amplified plant productivity in northern ecosystems. *Science*, 351(6274):696–699. https://doi.org/10.1126/science.aac4971.

Houghton, R. A. 2003. Why are Estimates of the Terrestrial Carbon Balance so Different? *Global Change Biol.* 9:500-9. https://doi.org/10.1046/j.1365-2486.2003.00620.x.

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. https://doi.org/10.1007/s10533-006-9019-5.

Jackson, T. J., J. S. Kimball, R. Reichle, W. Crow, A. Colliander, and E. Njoku. 2011. SMAP Science Calibration and Validation Plan. SMAP Science Document, No. 014, Version 1.2 (Preliminary Release), JPL D-52544:1-93.

Jones, L.A., J.S. Kimball, R.H. Reichle, N. Madani, J. Glassy, J.V. Ardizzone ... Russell L. Scott. 2017. The SMAP Level 4 Carbon Product for Monitoring Ecosystem Land–Atmosphere CO_2 Exchange. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6517-6532. https://doi.org/10.1109/TGRS.2017.2729343.

Kim, Y., J. Kimball, K. Zhang, K. McDonald. 2012. Satellite Detection of Increasing Northern Hemisphere Non-Frozen Seasons from 1979 to 2008: Implications for Regional Vegetation Growth. *Remote Sens. Environ.*, 121:472-487. https://doi.org/10.1016/j.rse.2012.02.014.

Kim, Y., J. S. Kimball, J. Glassy, J. Du. 2017. An extended global Earth system data record on daily landscape freeze–thaw status determined from satellite passive microwave remote sensing. *Earth System Science Data*, 9(1): 133–147. https://doi.org/10.5194/essd-9-133-2017.

Kimball, J. S., Jones L. A., Zhang K., Heinsch F. A., McDonald K. C., and Oechel W. C. 2009. A Satellite Approach to Estimate Land-Atmosphere CO_2 Exchange for Boreal and Arctic Biomes using MODIS and AMSR-E. *IEEE Geosci. Remote Sens.* 47:569-87. https://doi.org/10.1109/TGRS.2008.2003248.

Kimball, J. S., L. A. Jones, J. P. Glassy, and R. Reichle. 2014. SMAP Algorithm Theoretical Basis Document: L4 Carbon Product, Release A. SMAP Project, JPL D-66484, Jet Propulsion Laboratory, Pasadena CA. 76 pp. (see PDF).

Liu, Z., Kimball, J. S., Ballantyne, A. P., Parazoo, N C., Wang, W. J., Bastos, A., Madani, N. Natali, S. M., Watts, J. D., Rogers, B. M., Ciais, P., Yu, K., Virkkala, A.-M., Chevallier, F., Peters, W., Patra, P. K., & Chandra, N. 2022. Respiratory loss during late-growing season determines the net carbon dioxide sink in northern permafrost regions. *Nature Communications* 13, 5626. https://doi.org/10.1038/s41467-022-33293-x.

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. 1-35. http://www.nap.edu/catalog/11820.html.

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. https://doi.org/10.2136/sssaj1987.03615995005100050015x.

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, https://doi.org/doi:10.1038/nature06444.

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. https://doi.org/10.1029/93GB02725.

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. https://doi.org/10.1016/j.agrformet.2005.05.008.

Richardson, A.D., M.D. Mahecha, E. Falge, J. Kattge, A.M. Moffat, D. Papale ... D.Y. Hollinger. 2008. Statistical Properties of Random CO_2 Flux Measurement Uncertainty Inferred from Model Residuals. *Ag. For. Meteor.* 148:38-50. https://doi.org/10.1016/j.agrformet.2007.09.001.

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. https://doi.org/10.1016/S0034-4257(99)00061-9.

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. https://doi.org/10.1175/2011JCLI4034.1.

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. https://doi.org/10.1016/j.rse.2004.12.011.

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.

8 DOCUMENT INFORMATION

8.1 Publication Date

December 2022

8.2 Date Last Updated

June 2025

APPENDIX – DATA FIELDS

This appendix provides a description of all data fields within the *SMAP L4 Global Daily 9 km Carbon Net Ecosystem Exchange* product. The data are grouped in the following main HDF5 groups:

- EC — Environmental Constraints
- GEO — Geolocation Information
- GPP — Gross Primary Production
- NEE — Net Ecosystem Exchange
- QA — Quality Assurance
- RH — Heterotrophic Respiration
- SOC — Soil Organic Carbon
- Metadata

For a description of metadata fields for this product, refer to the Product Specification Document (Endsley et al., 2022b).

SMAP Level-4 carbon files include HDF5 groups that contain six primary variables: EC, GEO, GPP, NEE, RH, and SOC. With the exception of EC and GEO, each group includes data fields for eight Plant Functional Type (PFT) classes (refer to Table A - 8). Note that the QA variables also includes PFT classes. All output variables have dimensions 1624 x 3856 and are summarized in the following tables.

Environmental Constraints (EC)

Table A - 1. Data Fields for EC Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
emult_mean	Float32	2	percent	0.0	100.0	-9999.0
frozen_area	Float32	2	percent	0.0	100.0	-9999.0
tmult_mean	Float32	2	percent	0.0	100.0	-9999.0
wmult_mean	Float32	2	percent	0.0	100.0	-9999.0

Geolocation Information (GEO)

Table A - 2. Data Fields for GEO Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
latitude	Float32	2	degrees	-89.999	89.999	-9999.0
longitude	Float32	2	degrees	-179.999	179.999	-9999.0

Gross Primary Production (GPP)

Table A - 3. Data Fields for GPP Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
gpp_mean	Float32	2	$gCm^{-2}day^{-1}$	0.0	30.0	-9999.0
gpp_pft[1-8]_mean	Float32	2	$gCm^{-2}day^{-1}$	0.0	30.0	-9999.0
gpp_std_dev	Float32	2	$gCm^{-2}day^{-1}$	0.0	30.0	-9999.0

Net Ecosystem Exchange (NEE)

Table A - 4. Data Fields for NEE Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
nee_mean	Float32	2	$gCm^{-2}day^{-1}$	-30.0	20.0	-9999.0
nee_pft[1-8]_mean	Float32	2	$gCm^{-2}day^{-1}$	-30.0	20.0	-9999.0
nee_std_dev	Float32	2	$gCm^{-2}day^{-1}$	-30.0	20.0	-9999.0

Quality Assurance (QA)

Table A - 5. Data Fields for QA Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
carbon_model_bitflag	Uint16	2	N/A	0.0	65534	65534
nee_rmse_mean	Float32	2	$gCm^{-2}day^{-1}$	0.0	20.0	-9999.0
nee_rmse_pft[1-8]_mean	Float32	2	$gCm^{-2}day^{-1}$	0.0	20.0	-9999.0
qa_count	Uint8	2	N/A	0	81	254
qa_count_pft[1-8]	Uint8	2	N/A	0	81	254

Heterotrophic Respiration (RH)

Table A - 6. Data Fields for RH Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
rh_mean	Float32	2	$gCm^{-2}day^{-1}$	0.0	20.0	-9999.0
rh_pft[1-8]_mean	Float32	2	$gCm^{-2}day^{-1}$	0.0	20.0	-9999.0
rh_std_dev	Float32	2	$gCm^{-2}day^{-1}$	0.0	20.0	-9999.0

Soil Organic Carbon (SOC)

Table A - 7. Data Fields for SOC Group

Data Field Name	Type	Bit	Unit	Valid Min	Valid Max	Fill/Gap Value
soc_mean	Float32	2	gCm^{-2}	0.0	25000.0	-9999.0
soc_pft[1-8]_mean	Float32	2	gCm^{-2}	0.0	25000.0	-9999.0
soc_std_dev	Float32	2	gCm^{-2}	0.0	25000.0	-9999.0

Data Field Definitions

EASE2_global_projection

Defines the parameters of the cylindrical 9-km Earth-fixed EASE-Grid 2.0 projection and the mapping from latitude/longitude to grid-native coordinates

x

The x coordinate values from the cylindrical 9-km Earth-fixed EASE-Grid 2.0 projection

y

The y coordinate values from the cylindrical 9-km Earth-fixed EASE-Grid 2.0 projection

emult_mean

Environmental-constraint multiplier. PFT class designations are listed in Table A - 8.

frozen_area

Frozen area percentage (environmental-constraint), not distinguished by PFT

tmult_mean

Environmental constraint on temperature.

wmult_mean

Environmental constraint on moisture.

latitude

Latitude coordinates in decimal degrees.

longitude

Longitude coordinates in decimal degrees.

Table A - 8. PFT Classes

Class	Designation
1	Evergreen Needleleaf Forest
2	Evergreen Broadleaf Forest
3	Deciduous Needleleaf Forest
4	Deciduous Broadleaf Forest
5	Shrub
6	Grass
7	Cereal Crop
8	Broadleaf Crop

gpp_mean

Global daily 9-km Gross Primary Productivity (GPP) mean.

gpp_pft_[1-8]_mean

Mean value of GPP for pixels classified as PFT class [1-8]. PFT class designations are listed in Table A - 8.

gpp_std_dev

Global daily 9-km Gross Primary Productivity (GPP) standard deviation.

nee_mean

Global daily 9-km Net Ecosystem Exchange (mean).

nee_pft_[1-8]_mean

Mean value of NEE for pixels classified in

PFT classes 1 through 8. The PFT class designations are listed in Table A - 8.

nee_std_dev
Global daily 9-km Net Ecosystem Exchange standard deviation.

carbon_model_bitflag
Carbon model quality bitflags. Refer to Table 6 in this User Guide for information on bit flags. A copy of Table 6 is also included as an attribute called *legend_carbon_model_bitflag* data field for quick reference.

nee_rmse_mean
NEE unbiased RMSE estimate.

nee_rmse_pft[1-8]_mean
NEE unbiased RMSE estimate for PFT class [1-8]. PFT class designations are listed in Table A - 8.

qa_count
Carbon model QA count 0 to 81 (overall).

qa_count_pft[1-8]
QA count for PFT class [1-8]. PFT class designations are listed in Table A - 8

rh_mean
Mean heterotrophic respiration (Rh) value, not distinguished by PFT.

rh_pft_[1-8]_mean
Mean value of Rh for pixels classified as PFT class [1-8]. PFT class designations are listed in Table A - 8

rh_std_dev
Global daily 9-km heterotrophic respiration standard deviation.

soc_mean
Global daily 9-km soil organic carbon mean.

soc_pft[1-8]_mean
Mean value of SOC for pixels classified as PFT class [1-8]. PFT class designations are listed in Table A - 8.

soc_std_dev
Global daily 9km soil organic carbon standard deviation.

Fill/Gap Values

SMAP data products employ fill and gap values to indicate when no valid data appear in a particular data element. Fill values ensure that data elements retain the correct shape. Gap values locate portions of a data stream that do not appear in the output data file. Fill values appear in the SPL4CMDL data product over ocean and water surfaces. SMAP data products employ a specific set of data values to connote that a field is fill. The selected values that represent fill are dependent on the data type.

No valid value in the SPL4CMDL data product is equal to the values that represent fill. If any exceptions should exist in the future, the SPL4CMDL content will provide a means for users to discern between fields that contain fill and fields that contain genuine data values. Operationally, software should not attempt to ever test two floating point values for equality, but instead test two given values using a relational inequality operator and a tolerance-based difference in absolute values (e.g. if abs(a-b) <= 1E-14) method.

For discrete categorical variables such as the bit-flags (*carbon_bitflag* and *carbon_qual_flag_pft[1-8]*), all values are defined, therefore no missing data values are expected with these variables. For the analytical variables, the fPAR climatology provides a fallback input source to help assure there are no spatio-temporal gaps in the modeled data record. Therefore, no gaps are expected to occur in the SPL4CMDL output data stream.

Acronyms and Abbreviations

Table A - 9. Acronyms and Abbreviations

Abbreviation	Definition
Char	8-bit character
Int8	8-bit (1-byte) signed integer
Int16	16-bit (2-byte) signed integer
Int32	32-bit (4-byte) signed integer
EC	Environmental Constraints
Float32	32-bit (4-byte) floating-point integer
Float64	64-bit (8-byte) floating-point integer
fPAR	Fraction of Photosynthetically Active Radiation
GEO	Geolocation Information
GPP	Gross Primary Production
N/A	Not Applicable
NEE	Net Ecosystem Exchange
PFT	Plant Functional Type
QA	Quality Assurance
RH	Heterotrophic Respiration
SOC	Soil Organic Carbon
Uint8	8-bit (1-byte) unsigned integer
Uint16	16-bit (2-byte) unsigned integer
UTC	Universal Coordinated Time