Global Modeling and Assimilation Office

GMAO Office Note No. 11 (Version 2.1)

Soil Moisture Active Passive (SMAP) Mission
Level 4 Carbon (L4_C) Product Specification Document

Release Date: 11/15/2021

Global Modeling and Assimilation Office
Earth Sciences Division
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771

This page intentionally left blank.

Soil Moisture Active Passive (SMAP) Mission
Level 4 Carbon (L4_C) Product Specification Document

K. Arthur Endsley[1], Joe Glassy[1], John S. Kimball[1], Lucas A. Jones[1], Rolf H. Reichle[2], Joseph V. Ardizzone[2,3], Gi-Kong Kim[2], Robert A. Lucchesi[2,3], Edmond B. Smith[2,3], and Barry H. Weiss[4]

[1]NTSG, University of Montana, Missoula, MT, USA
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]Science Systems and Applications, Inc., Lanham, MD, USA
[4]Jet Propulsion Laboratory, Pasadena, CA, USA

Document maintained by Rolf Reichle (GMAO)

This document should be cited as:
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, 2021: Soil Moisture Active Passive (SMAP) Mission Level 4 Carbon (L4_C) Product Specification Document. GMAO Office Note No. 11 (Version 2.1), 69 pp, NASA Goddard Space Flight Center, Greenbelt, MD, USA. Available from http://gmao.gsfc.nasa.gov/pubs/office_notes.

Approved by:

____

Steven Pawson Date

Head, Global Modeling and Assimilation Office
Code 610.1, NASA GSFC

REVISION HISTORY

Revision	Date	Sections Changed	Reason for Change (ECR #)
1.0	26 February, 2013	All	J. Glassy
1.1	4 March, 2013	Product volume estimates added	J. Glassy
1.2	14 May, 28 May 2013	Added QA outputs partitioned by PFT, and updated narratives for QA and bit-flag scheme.	J. Glassy
1.3	17 July 2013	Refined definition of QA structure, added pft_freq{pft} and pft_mod_cnt{pft} definitions; GEO variables “latitude”,”longitude” changed to “cell_lat”, “cell_lon for Panoply interoperability	J.Glassy
1.4	23 August, 2013	Corrected several typos in QA field descriptions, and added gpp_method bit field to the QA bitflag variable.	J. Glassy
1.5	12 Sept 2013	Modified gpp_method bit to take on values {0,1,2}, indicating when FPAR climatology used vs FPAR 8-day or NDVI, and, added FT_method bitflag	J. Glassy
1.6	15 April, 2014	Deprecated output of 3km surface_flag variable, due to spatio-temporal incompatibility with 9-km variables	J. Glassy
		carbon_model_bitflag variable, **FT_method** bit flag (at bit 15) definition clarified as 0=used radar F/T, 1=used Tsurf (soil surf temp) driven F/T	J. Glassy
		TBD: pending L2/L3 changes to surface_flag will introduce new Tsurf F/T vs. radar F/T and	J. Glassy

		need for a F/T provenance bit indicating which was used.	
1.6.1	17 June 2014	Changed file name format to adopt L4 wide Science Version ID, and clarified valid_min, valid_max dataset properties	J. Glassy
1.7	23 July 2014	QA bitfield x-x GPP_Method changed to a single bit {0|1} and NDVI_Used {0|1} single bit metric was added , and adjusted metadata table to harmonize with L4_SM	J. Glassy
1.8	24 Oct, 2014	All QA spatial variables now merged into single L4C MDL output granule; added additional ISO metadata "source" sections needed for integration, and added prior missing DomainConsistency, CompletenessOmission	J. Glassy
1.8.1	22 June 2015	Updated EC min, max value documentation	J. Glassy
1.9	31 October 2015	Prefatory material: converted to GMAO Office Note for public beta-release of L4_C data. Revised definition of "Launch Indicator" in file name.	R. Reichle, J. Glassy
2.0	April 2016	Added references to new QA RMSE mean and (8) PFT wise RMSE float32 dataset/variables, and deleted QA qa_counts variables.	J. Glassy
2.1	15 November 2021	Updated SMAP Products Table (Table 1) and SMAP documents, added spatial coordinate datasets (**EASE2_global_projection, x, y**), and applied minor corrections throughout. Reflects Version 6 L4_C product release.	J. S. Kimball, K. A. Endsley, R. H. Reichle

TBD, TBR, TBS LOG

Section/Page	Description	Due Date
TBD		

Table of Contents

1 Contents

1 Contents 1
1 Introduction 4
1.1. Identification 5
1.2 Scope 5
1.3 The SMAP Experiment 5
1.4 SMAP Data products 7
Table 1: SMAP Data Products 7
1.5 Content Overview 9
Table 2: Plant Function Type (PFT) Classifier Summary 10
1.6 Related SMAP Project Documents 10
1.7 Applicable Documents 11
2 Data Product Organization 12
2.1 File Format 12
2.2 HDF File Notation 12
2.3 SMAP File Organization 13
2.3.1 Structure 13
2.3.2 Data 13
2.3.3 Element Types 13
Table 3: Element type definitions 14
2.3.4 File Level Metadata 15
2.3.5 Local Metadata 15
2.4 Data Definition Standards 16
Table 4: Definitions of Data element characteristics 16
2.4.1 Double Precision Time Variables 17
2.4.2 Array representation 17
3 Interface Characteristics 18
3.3 Coordinate Systems 18
3.2 The Global cylindrical EASE-Grid 2.0 19
3.3 Fill and Gap Values 21

Table 5: SMAP data product fill values. 22
4 Data Definition 23
4.1 Product Overview 23
4.1.1 SMAP L4_C Data Product 23
Table 6: L4_C Data Product Overview 24
4.1.2 Geolocation Information Overview 25
4.1.3 HDF5 Group QA Overview 25
4.1.4 L4_C Metadata Overview 26
4.2 Data Groups and Volume Estimates 26
Table 7: L4_C Data Volume Estimates 26
4.3 File Naming Conventions 27
4.4 Metadata Detail 29
4.4.1 L4_C Mandatory Metadata 30
4.4.2 L4_C Product Specific metadata 30
Table 8: Granule level metadata in the L4_C data product. 30
4.5 Data Structure 45
4.6 Element definitions 45
4.6.1 NEE Dataset Definition 46
Table 9: NEE Group Contents 46
4.6.2 GPP Dataset Definition 47
Table 10: GPP Group Contents 47
4.6.3 Heterotrophic respiration (R_h) Dataset definition 48
Table 11: Rh Group Contents 48
4.6.4 Soil Organic Carbon (SOC) Dataset definition 49
Table 12: SOC Group Contents 49
Table 13: EC -- Environmental Constraints Group Contents 50
4.6.5 Geolocation Information 50
Table 14: Overview of Geolocation Information 51
4.6.6 QA : the Quality Assurance Group 51
Table 15: Overview of Quality Group: QA 52
4.6.7 The carbon_model_bit-flag 53
Table 16: L4_C Quality Control Variable (carbon_model_bitflag) Definition 53

4.6.8 The NEE RMSE QA metrics: nee_rmse_mean, and nee_rmse_mean partitioned by PFT 54

Table 17: L4_C Quality Control Variable (nee_rmse_{pft:1..8}_mean Definition 55

4.7 L4_C Data Product Volume Estimates 55

Table 18: Daily L4_C Data Product Volume Estimates 56

5 Appendices 57

Appendix A: Acronyms and abbreviations 57

Appendix B: SMAP Data Product Specification Documents 65

Appendix C: L4C Dimensions 66

Table 18: Dimensions in the SMAP L4_C data product. 66

Appendix D: L4_C Shapes 67

Table 19: Shapes in the SMAP L4_C data product. 67

Appendix E: L4_C Units 68

Table 20: Units in the SMAP L4_C data product. 68

1 Introduction

The NASA Soil Moisture Active Passive (SMAP) mission's Level 4 Carbon (L4_C) data product provides daily estimates of global net ecosystem (CO_2) exchange (NEE), component carbon fluxes for vegetation gross primary production (GPP) and ecosystem respiration (R_h), and surface soil organic carbon (SOC) stocks. A satellite data-driven terrestrial carbon flux model is used for estimating the daily carbon fluxes. Key model inputs include: land cover and canopy intercepted fractional photosynthetically active radiation (fPAR) from the MODerate resolution Imaging Spectroradiometer (MODIS); surface- and root-zone soil moisture and temperature from the SMAP Level 4 Soil Moisture (L4_SM) product; and daily surface meteorology from the Goddard Earth Observing System (GEOS) Forward Processing weather analysis.

The L4_C product is computed using the daily Level 4 Carbon (L4_C) algorithm, which produces the above-mentioned global ecosystem productivity variables along with measures of the environmental constraint (EC) on daily fluxes and a set of QA/QC variables and bit-flags stratified by Plant Functional Type (PFT). All L4_C variables are modeled on a daily basis and internally registered to an EASE-Grid version 2 global 1-km extent (14616 lines by 34704 samples, for 507,233,664 pixels per layer), and are then posted to an EASE-Grid v2 global 9-km grid (3856 samples by 1624 lines, for 6,262,144 pixels per layer). The model derivation of each of these variables is described in detail in the SMAP L4_C Algorithm Theoretical Basis Document (ATBD).

The first Validated Release of the SMAP L4_C product, at Version 2, is distinguished from the prior Beta release (Version 1) of the product principally through the addition of the NEE Root Mean Square Error (RMSE) variable, which is described in detail in Section 4.6.8 of this document. (In this context, "Version n" refers to the data product version as archived at, and disseminated by, the National Snow and Ice Data Center, where n equals the major digit of the Science Version ID and also matches the ECS Version ID; section 4.3, Table 8.)

The SMAP L4_C baseline product has a 9-km spatial resolution consistent with the L4_SM product but retains sub-grid scale heterogeneity information determined from finer scale (1-km resolution) land cover and FPAR inputs. The resulting carbon product is similar to the sampling footprint of tower CO_2 eddy covariance flux measurements (Baldocchi et al. 2008, Chen et al. 2012). The baseline L4_C domain encompasses all global vegetated land areas and has a target mean unbiased RMSE (ubRMSE) accuracy for NEE within 30 g C m^{-2} yr^{-1} (1.6 g C m^{-2} d^{-1}), and emphasizing northern (≥45°N) boreal and arctic biomes, which is within the estimated ±30-100 g C m^{-2} yr^{-1} accuracy of *in situ* tower measurements (Baldocchi 2008; Richardson 2005; Richardson 2008). The L4_C baseline product has a mean daily temporal sampling to characterize the dynamic NEE response to daily variations in surface meteorology and associated moisture and thermal controls to GPP and respiration, and for greater precision in the computation of cumulative (weekly, monthly, and annual) carbon fluxes. Operational implementation of the baseline L4_C algorithms using L4_SM and GEOS Forward Processing surface meteorology inputs enables spatially and temporally continuous, daily mapping of NEE for all vegetated land areas independent of data gaps, vegetation biomass, and other constraints on SMAP parameter retrievals. The product determines NEE and component carbon fluxes over all global vegetated land areas, representing ecosystem carbon processes at the level of major plant functional types.

1.1. Identification

This GMAO Office Note No. 11 is the Product Specification Document (PSD) for the SMAP Level 4 Carbon (L4_C) dataset produced at the NASA Global Modeling and Assimilation Office (GMAO) for the SMAP mission project. Version 2.1 of this document replaces previous document versions (v1.9, v2.0) and refers to the latest (Version 6) release of the SMAP L4_C data product.

1.2 Scope

This Product Specification Document (PSD) describes the output file format of the L4_C data product. The intent of this document is to describe and clarify the L4_C data product structure as needed for external software interfaces. The SMAP Science Data Management and Archive Plan provides a more comprehensive explanation of these data within the complete context of the SMAP instrument, algorithms, and software.

1.3 The SMAP Experiment

The Soil Moisture Active Passive (SMAP) mission is designed to enhance the accuracy and the resolution of space-based measurements of terrestrial soil moisture and freeze-thaw state dynamics. SMAP data products have a noteworthy impact on multiple relevant and current Earth Science endeavors. These include:

- Understanding of the processes that link the terrestrial water, the energy and the carbon cycles;
- Estimations of global water and energy fluxes over the land surfaces;
- Quantification of the net carbon flux in boreal landscapes;
- Forecast skill of both weather and climate;
- Predictions and monitoring of natural disasters including floods, landslides and droughts; and
- Predictions of agricultural productivity.

To provide these data, the SMAP mission deploys a satellite observatory in a near-polar, sun-synchronous orbit. The observatory includes an L-band radiometer that operates at 1.40 GHz and an L-band radar that operates at 1.26 GHz. The instruments share a rotating reflector antenna with a 6-meter aperture that scans over a 1000-km swath.

As the spacecraft flies from north to south on *descending* orbits, the SMAP instruments view Earth locations at approximately 06:00 local time. As the spacecraft flies from south to north, on *ascending* orbits, the SMAP instruments view Earth locations at approximately 18:00 local time. The spacecraft operates in a cycle of 117 repeatable orbits.

Each time that the spacecraft repeats the orbit cycle, the nadir path on the Earth's surface may not vary by more than 20 km. The flight plan enables scientists to collect data over any region of the Earth over seasonal and annual cycles and avoid diurnal variations. The combined flight pattern and viewing design enables the observatory to view almost all of the Earth's land mass once every three days.

The SMAP radiometer records microwave emissions from the top 5 cm in the soil with a spatial resolution of about 40 km. Scientific applications based on the radiometer's measurements in the same frequency range have established this approach as an accurate means to detect the presence or water in near-surface soil. The SMAP radar was designed to provide backscatter measurements at 3 km resolution. The combined instrumentation was designed enable SMAP to generate highly accurate global soil moisture estimates at 9-km resolution. However, the SMAP radar malfunctioned and ceased operating after within the first three months of operations, while the radiometer continues to conduct normal operations.

The SMAP radiometer data contribute to model enhanced products providing global estimates of surface to root zone (1m depth) soil moisture and carbon (CO_2) net ecosystem exchange between the Earth's surface and the atmosphere.

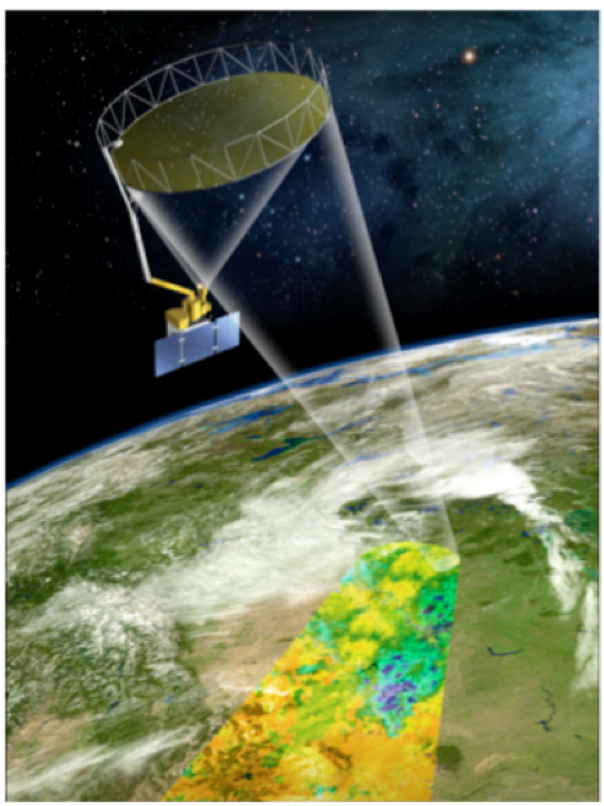

Figure 1: Artist's concept of the SMAP observatory.

1.4 SMAP Data products

During the life-of-mission (LOM) the SMAP project produces a series of data products, generally associated with their processing levels 1-4. Lower-level products embody data organization schemes closer to the observational platform (e.g., swath format, calibrated, geometrically corrected, and not gridded) while higher-level products have properties more conducive to a broader range of science users, who typically expect and work with data that possess a formal map projection, are gridded, and, more importantly, come with more user-oriented tools. The following table shows the Level 4 Carbon product in context with the other SMAP mission data products.

Table 1: SMAP Data Products

SMAP Data Products		
Mission Product Short Name	ECS Short Names	Description
L1A_Radar	SPL1AA	Parsed radar instrument telemetry
L1A_Radiometer	SPL1AP	Parsed radiometer instrument telemetry
L1B_S0_LoRes	SPL1BS0	Low resolution radar (x) in time order
L1C_S0_HiRes	SPL1CS0	High resolution radar (x) on swath grid
L1B_TB	SPL1BTB	Radiometer T (t) in time order
L1B_TB_E	SPL1BTBE	Backus-Gilbert interpolated, calibrated Radiometer T (t) in time order
L1C_TB	SPL1CTB	Radiometer T(t)
L1C_TB_E	SPL1CTBE	Backus-Gilbert interpolated, calibrated Radiometer T (t) on EASE2 grid
L2_SM_A	SPL2SMA	Radar soil moisture, includes freeze thaw state
L2_SM_P	SPL2SMP	Radiometer soil moisture

L2_SM_P_E	SPL2SMPE	Soil moisture derived from Backus-Gilbert interpolated Radiometer T
L2_SM_AP	SPL2SMAP	Active Passive soil moisture
L3_FT_A	SPL3FTA	Daily global composite freeze thaw state
L3_SM_A	SPL3SMA	Daily global composite radar soil moisture
L3_FT_P	SPL3FTP	Daily composite freeze/thaw state derived from Radiometer T
L3_FT_P_E	SPL3FTPE	Daily composite freeze/thaw state derived from Backus-Gilbert interpolated Radiometer T
L3_SM_P	SPL3SMP	Daily global composite radiometer soil moisture
L3_SM_P_E	SPL3SMPE	Daily global composite soil moisture derived from Backus-Gilbert interpolated Radiometer T
L3_SM_AP	SPL3SMAP	Daily global composite active passive soil moisture
L4_SM	SPL4SM	Surface and root zone soil moisture
L4_C	**SPL4C**	**Carbon net ecosystem exchange**

1.5 Content Overview

The SMAP L4_C data product contains daily estimates of global ecosystem carbon (CO_2) land-atmosphere fluxes (NEE, GPP, and R_h) and land surface soil organic carbon stocks (SOC) along with underlying environmental constraint (EC) metrics and data quality assurance and quality control (QA/QC) attributes. NEE is a fundamental measure of the balance between carbon uptake by vegetation GPP and carbon losses to the atmosphere through autotrophic (R_a) and heterotrophic (R_h) respiration. The sum of R_a and R_h defines the total ecosystem respiration rate (R_{tot}), which encompasses most of the annual terrestrial CO_2 efflux to the atmosphere. NEE, GPP, and R_h are expressed in units of g C m^{-2} d^{-1}, and SOC is expressed in units of g C m^{-2}. The L4_C data product appears on an Earth fixed, global, cylindrical Equal Area Scalable Earth grid, version 2.0 (EASE-Grid 2.0) at 9-km resolution.

As a spatio-temporal data product, the daily global L4_C outputs are organized on the basis of (8) land-cover Plant Functional Type (PFT) classes, as defined within the Terra/Aqua MODIS land instrument team MOD12Q1/MDY12Q1 dataset. The PFT 2D variable is stored as the Type-5, Plant Functional Type scientific data set (SDS) in both the MOD12Q1 (1-km resolution), and MDY12Q1 file (500meter resolution) forms. Global 1-km PFT classification maps in the original Sinusoidal 1-km map projection are thus distributed as part of the Terra/Aqua MODIS MOD12Q1 and related, global land-cover products. PFT classes directly used as driving parameters in the L4_C algorithm each possess an entry in the L4_C Biome Properties Lookup Table (BPLUT). The table below summarizes the Plant Functional Type classifiers and highlights the (8) PFT classes used by the L4_C data algorithm to partition L4_C output variables.

Table 2: Plant Function Type (PFT) Classifier Summary

PFT Class label	PFT code	PFT Description	PFT Class used in L4_C
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

1.6 Related SMAP Project Documents

SMAP L4 Global Daily 9 km EASE-Grid Carbon Net Ecosystem Exchange: User Guide. NASA National Snow and Ice Data Center Distributed Active Archive Center. 2021. Webpage. https://nsidc.org/data/SPL4CMDL/.

SMAP L4_C Algorithm Theoretical Basis Document (ATBD, v.1), J. Kimball et al, JPL D-66484, DDR No. SMAP-11-012

SMAP L4_SM Algorithm Theoretical Basis Document (Initial Release v.1), R. Reichle et al., JPL D-66483, October 1, 2012.

Kimball, J. S., K. A. Endsley, T. Kundig, J. Glassy, R. H. Reichle, and J. V. Ardizzone (2021), Validation Assessment for the Soil Moisture Active Passive (SMAP) Level 4 Carbon (L4_C) Data Product Version 5, NASA Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2018-104606, Vol. 56, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 19pp. Available at: https://gmao.gsfc.nasa.gov/pubs/.

SMAP Science Data Management and Archive Plan, JPL D-45973, August 29, 2011.

SMAP Pointing, Positioning, Phasing and Coordinate Systems, Volume 0: Definitions and Principal Coordinate Systems, JPL D-46018, Initial Release, May 18, 2010

SMAP Science Data Calibration and Validation Plan. SMAP Project, JPL D-52544, Jet Propulsion Laboratory, Pasadena, CA

SMAP Applications Plan. SMAP Project, JPL D-53082, Jet Propulsion Laboratory, Pasadena, CA

Kimball, John S., Lucas A. Jones,Joseph Glassy, E. Natasha Stavros, Nima Madani, Rolf H. Reichle, Thomas Jackson, and Andreas Colliander, 2015. Soil Moisture Active Passive (SMAP) Project Calibration and Validation for the L4_C Beta-Release Data Product. NASA/TM–2015-104606, Vol. 42, 37 pp. Available at: https://gmao.gsfc.nasa.gov/pubs/.

1.7 Applicable Documents

ISO 19115:2003(E) International Standard – Geographic Information – Metadata, May 1, 2003.

ISO 19115-2:2009 International Standard – Geographic Information – Part 2:Extensions for imagery and gridded data, December 12, 2009.

ISO 19139:2007 International Standard – Geographic Information – Metadata – XML schema implementation, May 14, 2009.

Introduction to HDF5, The HDF Group, http://www.hdfgroup.org/HDF5/doc/H5.intro.html.

HDF5: API Specification Reference Manual, The HDF Group http://www.hdfgroup.org/HDF5/doc/RM/RM_H5Front.html

HDF5 User's Guide Release 1.8.9, The HDF Group, http://hdfgroup.com/HDF5/doc/UG, May 2012.

EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets, Brodzik, M.J., et. al., National Snow and Ice Data Center, Cooperative Institute of Environmental Sciences, University of Colorado, ISPRS International Journal of Geo-Information, ISSN 2220-9964, DOI: 10.3390/igji1010032.

NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.6, December 5, 2011.

2 Data Product Organization

2.1 File Format

All SMAP standard products are in the Hierarchical Data Format version 5 (HDF5). HDF5 is a scientific, multi-platform binary file format, with access by software applications provided through an applications programming interface (API) library. A key distinguishing characteristic of the HDF5 file format is the extensive, multi-dimensional, self-referential data model it is built around; these properties allow for very flexible representation in both data and metadata, while providing good I/O performance, storage efficiency, and consistent access across many different compute platforms. The National Center for Supercomputing Applications (NCSA) at the University of Illinois developed HDF to help scientists share data regardless of the source. Modern high-performance compute platforms often differ in native word-size and byte ordering. These intrinsic properties govern how the machines internally represent integer and floating point numbers. Use of the HDF library gives users a consistent data model and tool to read HDF files on multiple platforms regardless of the differences in native architecture (endian byte-ordering) and word-size of the platform. Further, HDF files are equally accessible by routines written either in Fortran, C or C++.

The HDF Group is responsible for the development and maintenance of the HDF family of file formats. Users should reference The HDF Group website at http://www.hdfgroup.org to download HDF software and documentation.

2.2 HDF File Notation

HDF5 represents a significant departure from the conventions of previous versions of HDF, specifically, HDF v4.x. The changes that appear in HDF5 provide flexibility to overcome many of the limitations of previous releases, particularly regarding total file size and the ability to more naturally represent hierarchical groups of data and attributes. The key concepts of the HDF5 Abstract Data Model are Files, Groups, Datasets, Datatypes, Attributes and Property Lists. The most relevant of these for L4_C data product users are the group, datatype, dataset, and attributes entities. Briefly, HDF ***groups*** are named organizational levels, arranged in a hierarchy starting at the root or file level, within which data and attributes are encapsulated. Actual data is an HDF5 file are stored in ***Datasets***, a term that refers to the primary, multi-dimensional science data stored in the file. Datasets are defined as storing elements belonging to a supported HDF model ***datatype***. HDF datatypes ("uint16", "float32", etc) are expressively named to include a numeric variable's 'sign' property as signed or unsigned, its type as integer vs. floating point, along with a word-size identifier that helps convey the variables natural numeric representational limits. Attributes refer to a special class of named HDF variable that are most often used to represent metadata, or "data-about-data". The reader is referred to the HDF user documentation cited in Section 1.7 for formal descriptions of these foundation data model concepts.

2.3 SMAP File Organization

2.3.1 Structure

SMAP data products follow a common convention for all HDF5 Files with regard to both structure and naming. Use of this convention provides uniformity of data access and interpretation. The SMAP Project uses HDF5 Groups, which provide an additional level of data organization. All metadata that pertain to the complete data granule are members of the "/Metadata" Group and are typically stored using HDF Attributes as summarized above. All other data are organized within Groups that are designed specifically to handle the structure and content of each particular data product as specified later in this document.

2.3.2 Data

All data in HDF5 files are stored in individual Datasets. All of the Datasets in an SMAP product are assigned to an HDF5 Group, where the file's root level (denoted by the group called "/") also represents an implicit "base" group. A standard field name is associated with each Dataset. The field name is a unique string identifier. The field name corresponds to the name of the data element the Dataset stores. This document lists these names with the description of each data element that they identify.

Each Dataset is associated with an HDF5 Dataspace and an HDF5 Datatype. They provide a minimally sufficient set of parameters for reading the data using standard HDF5 tools.

2.3.3 Element Types

SMAP HDF5 employs the Data Attribute "Type" to classify every data field as a specific data type. The "Type" is an embellishment upon the standard HDF5 Datatypes that is designed specifically to configure SMAP data products.

Table 3 lists all of the "Type" strings that appear in the SMAP data products. The table maps each SMAP "Type" to a specific HDF5 Datatype in both the HDF5 file and in the data buffer. The table also specifies the common conceptual data type that corresponds to the "Type" in SMAP executable code.

Table 3: Element type definitions

Type	HDF5 Datatype (File)	HDF5 Datatype (Buffer)	Conceptual Type
Unsigned8	H5T_STD_U8LE	H5T_NATIVE_UCHAR	unsigned integer
Unsigned16	H5T_STD_U16LE	H5T_NATIVE_USHORT	unsigned integer
Unsigned24	H5T_STD_U16LE, with precision set to 24 bits, and size set to 3 bytes.	H5T_NATIVE_INT	unsigned integer
Unsigned32	H5T_STD_U32LE	H5T_NATIVE_UINT	unsigned integer
Unsigned64	H5T_STD_U64LE	H5T_NATIVE_ULLONG	unsigned integer
Signed8	H5T_STD_I8LE	H5T_NATIVE_SCHAR	signed integer
Signed16	H5T_STD_I16LE	H5T_NATIVE_SHORT	signed integer
Signed32	H5T_STD_I32LE	H5T_NATIVE_INT	signed integer
Signed64	H5T_STD_I64LE	H5T_NATIVE_LLONG	signed integer
Float32	H5T_IEEE_F32LE	H5T_NATIVE_FLOAT	floating point
Float64	H5T_IEEE_F64LE	H5T_NATIVE_DOUBLE	floating point
FixLenStr	H5T_C_S1	H5T_NATIVE_CHAR	character string
VarLenStr	H5T_C_S1, where the length is set to H5T_VARIABLE	H5T_NATIVE_CHAR	character string

SMAP HDF5 files employ two different types of string representation. “VarLenStr” are strings of variable length. “VarLenStr” provides greater flexibility to represent character strings. In an effort to make SMAP HDF5 more interoperable (friendly) for users who wish to use netCDF software, SMAP products restrict the use of “VarLenStr”. “FixLenStr” are strings with a prescribed fixed length. “FixLenStr” are useful for fixed length strings that are stored in large multi-dimension arrays. UTC time stamps are an excellent example of the type of data that store well in a "FixLenStr". The L4_C product uses fixed length strings unless otherwise noted.

2.3.4 File Level Metadata

Metadata describing the full content of each granule of the SMAP data product are stored within the explicitly named "/METADATA" Group. SMAP metadata are handled using exactly the same procedures as those that are used to handle SMAP data. The contents of each Attribute that stores metadata conform to one of the SMAP Types. Like data, each metadata element is also assigned a shape. Most metadata elements are stored as scalars. A few metadata elements are stored as arrays.

SMAP data products represent file level metadata in two forms. One form appears in one or more Attributes within the Metadata Group. Combined, those Attributes contain a complete representation of the product metadata. The content conforms to the ISO 19115-2 models in ISO 19139 compliant XML.

The second form of the metadata appears in a set of HDF5 Groups under the "/METADATA" Group. Each of these HDF5 Groups represents one of the major classes in the ISO 19115-2 model. These HDF5 Groups contain a set of HDF5 Attributes. Each HDF5 Attributes represents a specific ISO attribute of the associated ISO class. Although this hierarchical representation inherits its design from the ISO model, it does not completely conform to the model. In many cases, the names of the HDF5 Attributes match those used in the ISO model. In some situations, names were changed to provide greater clarity to SMAP users who are not familiar with the ISO model. Furthermore, to ease metadata searches, the structure of Groups within Groups was limited to four levels.

2.3.5 Local Metadata

SMAP standards incorporate additional metadata that describe each HDF5 Dataset within the HDF5 file. Each of these metadata elements appear in an HDF5 Attribute that is directly associated with the HDF5 Dataset. Where feasible, these HDF5 Attributes employ names that conform to the Climate and Forecast (CF) conventions.

2.4 Data Definition Standards

Sections 4.6 of this document specify the characteristics and definitions of every data element stored in this SMAP data product. Table 4 defines each of the specific characteristics that are listed in that section of this document. Some of these characteristics correspond with the SMAP HDF5 Attributes that are associated with each Dataset. Data element characteristics that correspond to SMAP HDF5 Attributes bear the same name. The remaining characteristics are descriptive data that help users better understand the data product content.

In some situations, a standard characteristic may not apply to a data element. In those cases, the field contains the character string ’n/a’. Hexadecimal representation sometimes indicates data content more clearly. Numbers represented in hexadecimal begin with the character string ’0x’.

Table 4: Definitions of Data element characteristics

Characteristic	Definition
Type	The data representation of the element within the storage medium. The storage class specification must conform to a valid SMAP type. The first column in Table 3 lists all of the valid values that correspond to this characteristic.
Shape	The name of the shape data element that specifies the rank and dimension of a particular data set. Appendix C lists all of the valid shapes that appear in this data product.
Valid_min	The expected minimum value for a data element. In most instances, data element values never fall below this limit. However, some data elements, particularly when they do not reflect normal geophysical conditions, may contain values that fall below this limit.
Valid_max	The expected maximum value for a data element. In most instances, data element values never exceed this limit. However, some data elements, particularly when they do not reflect normal geophysical conditions, may contain values that exceed this limit.
Valid Values	Some data elements may store a restricted set of values (e.g., categorical variables and others). In those instances, this listing specifies the finite list of values that the data element may store.
String Length	This characteristic specifies the length of the data string that represents a single instance of the data element. This characteristic appears exclusively for data elements of FixLenStr type.
Units	Units of measure. Typical values include “deg”, “degC”, “Kelvins”, “m/s”, “m”, “m**2”, “s” and “counts”. Appendix A and Appendix E include references to important data measurement unit symbols.

2.4.1 Double Precision Time Variables

SMAP double precision time variables contain measurements relative to the J2000 epoch. Thus, these variables represent a real number of Standard International (SI) compatible seconds since 11:58:55.816 on January 1, 2000 UTC.

2.4.2 Array representation

This document employs array notation to demonstrate and clarify the correspondence among data elements in different product data elements. The L4_C is implemented in C/C++ language (c99 standard where supported), so the standard (base0) C/C++ array notation is adopted in this document. All array indices are therefore zero based, and in multidimensional arrays the rightmost subscript index varies most rapidly. This scheme is typically referred to as a “row-major-order” (RMO) indexing scheme, where a linear sequencing through an orthogonal 2D raster extent has the column indices varying fastest and the row indices varying least quickly.

HDF5 is designed to read data seamlessly regardless of the computer language used to write an application. Thus, elements that are contiguous using the dimension notation in this document will appear in contiguous locations in arrays for reading applications in any language with an HDF5 interface.

3 Interface Characteristics

3.3 Coordinate Systems

The SMAP mission uses the Science Orbit Reference Frame (SRF) and the Earth Centered Rotating (ECR) coordinate systems to represent spacecraft attitude, position and relative motion.

The Science Orbit Reference Frame (SRF) is a right-handed coordinate system with its three axes mutually orthogonal. The SRF is defined such that the origin is at the spacecraft center of mass (CM). The +Z axis points toward Geodetic Nadir. Due to the oblateness of the Earth, the vector from the spacecraft to the geometric center of the Earth (Geocentric Nadir) is different from the vector from the spacecraft to the local WGS84 ellipsoid normal (Geodetic Nadir). The +X axis is coplanar with both the +Z axis and the spacecraft inertial velocity vector. The +X axis closely adheres to the direction of the spacecraft inertial velocity vector. The +Y axis completes the right-handed, orthogonal coordinate system. The +Y axis is normal to the orbit plane with positive sense in the direction opposite the orbit angular momentum vector.

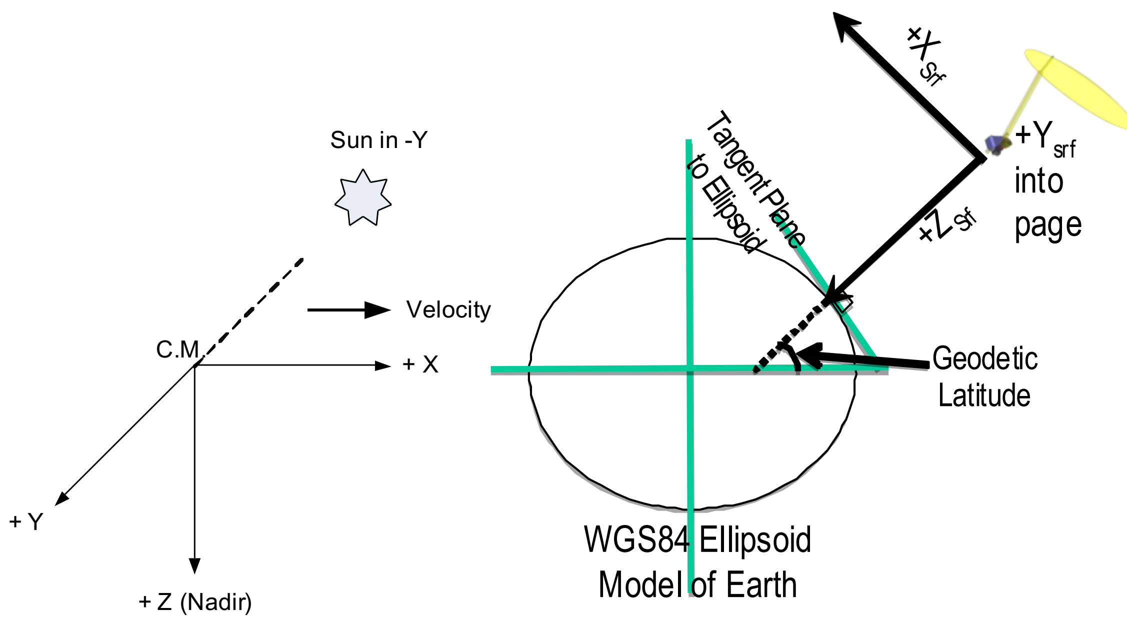

Figure 2: The Science Orbit Reference Frame coordinate system.

The Earth Centered Rotating (ECR) or Earth Centered Fixed coordinate system is a right-handed coordinate system with three mutually orthogonal axes. The origin of the system is the Earth's center of mass. The positive x-axis extends from the origin through the intersection of the Equator at 0° latitude and the Greenwich Meridian at 0° longitude. The positive z-axis extends directly North from the origin of the ECR system. Due to a slight wobbling of the Earth, the z-axis does not coincide exactly with the instantaneous rotation axis of the Earth. The y-axis completes the right- handed coordinate system as a vector from the origin to the intersection of the Equator and 90° East longitude.

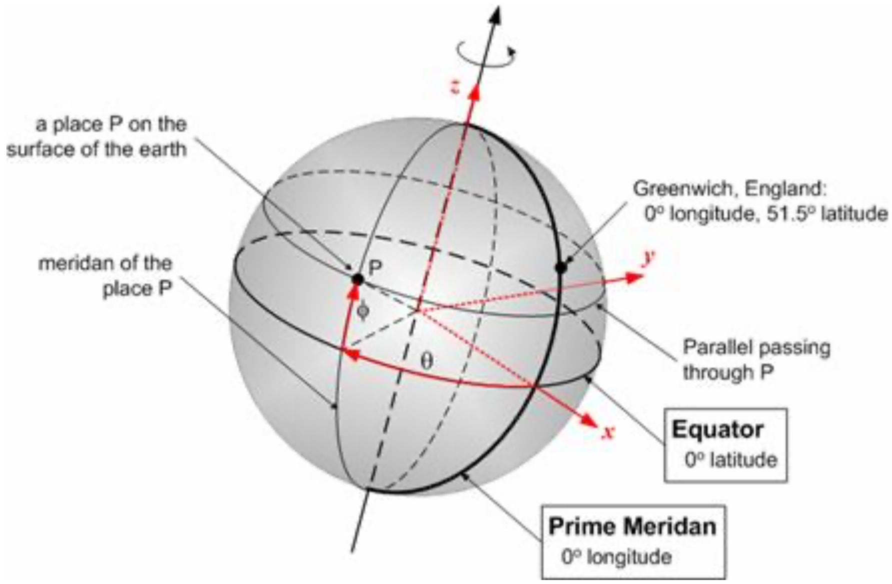

Figure 3: Earth Centered Rotating coordinate system.

The SMAP mission adopted the World Reference System WGS84 ellipsoid to define the horizontal Earth reference coordinates. The WGS84 geoid was adopted as the vertical Earth reference coordinates. Geodetic measure is used to define both the spacecraft location and the instrument target location relative to the Earth's surface.

3.2 The Global cylindrical EASE-Grid 2.0

The data in the SMAP L4_C data product are provided on the global cylindrical EASE-Grid 2.0 with a nominal grid spacing of 9 km. Each grid cell has an area of about 81 km^2, regardless of longitude and latitude. Using this projection, all global data arrays have dimensions of 1624 rows and 3856 columns (6,262,144 pixels per layer). Note that the EASE-Grid 2.0 global 1k reference grid is defined as 14616 lines by 34704 samples (507,233,664 pixels per layer). Figure 4 illustrates the region covered by the global cylindrical EASE-Grid 2.0 projection.

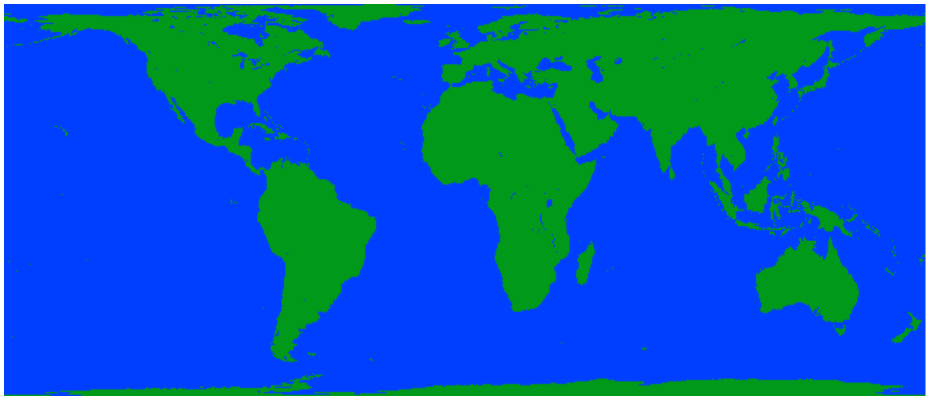

Figure 4: Global EASE-Grid 2.0.

The 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 made “perfect” so that smaller grid cells can be tessellated to form larger grid cells. Figure 5 shows a schematic of the nesting.

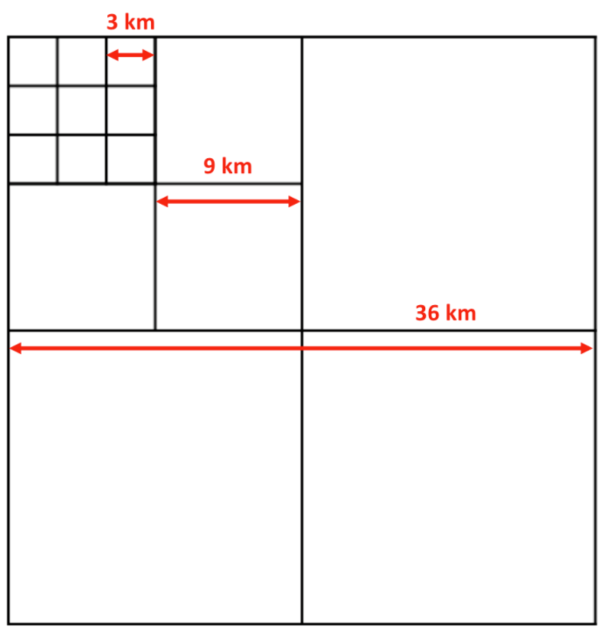

Figure 5: Perfect nesting in EASE-Grid 2.0. Smaller grid cells tessellate into larger grid cells.

The 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. Other SMAP data products use 3 km, 9 km and 36 km grid spacing. Figure 6 provides an example of sample Normalized Difference Vegetation Index (NDVI) data on the EASE-Grid 2.0 at these three resolutions.

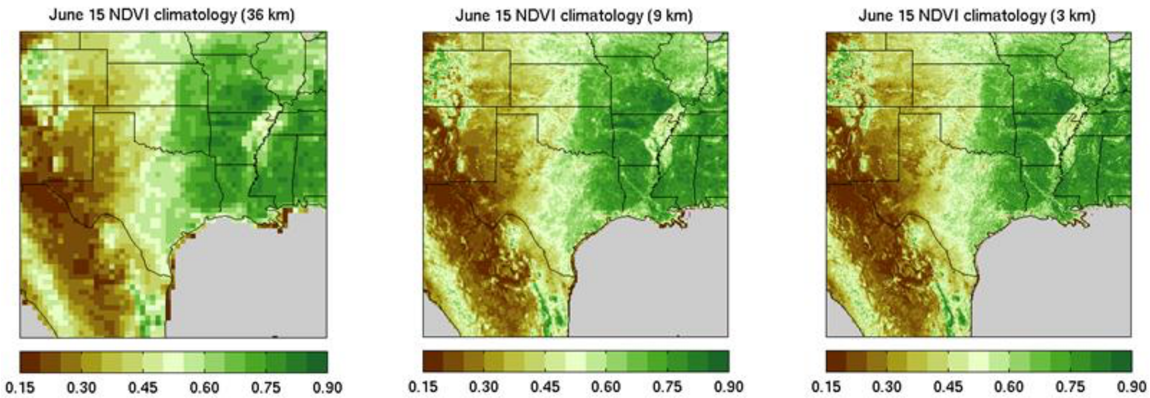

Figure 6: Sample NDVI data displayed on the SMAP EASE-Grid 2.0 with (left) 36 km, (middle) 9 km, and (right) 3 km nominal grid spacing.

3.3 Fill and 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 L4_C data product over non-vegetated regions including barren, urban, permanent ice/snow, ocean and permanent inland water surfaces. L4_C simulations are only performed when one or more vegetated 1-km grid cells with valid FPAR record are available within a 9-km grid cell, otherwise the 9-km grid cell receives fill values for all L4_C product fields. SMAP data products employ a standard set of data values to connote that an element is fill. The selected values that represent fill are dependent on the data type. Table 5 lists the values that represent fill in SMAP products based on data type:

No *valid* value in the L4_C data product is equal to the values that represent fill. If any exceptions should exist in the future, the L4_C content will provide a means for users to discern between elements that contain fill and elements that contain genuine data values. Operationally, we note that 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(fabs(a-b) <= 1E-14) method. This document will also contain a description of the method used to ascertain which elements are fill and which elements are genuine.

For discrete categorical variables such as the bit-flags (carbon_bitflag and carbon_qual_flag_{pft}, 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 temporal gaps are expected to occur in the L4_C output data stream. To accommodate gaps in the Freeze Thaw (F/T) input variable to the L4_C algorithm, alternative F/T estimates are derived from surface temperature (Tsurf, a modeled LDAS parameter) acquired from the GEOS

Forward Processing daily surface meteorology “gph” data product. In all cases, the final F/T state is represented on an EASE-Grid v2.0 3-km grid supplied to the L4_C algorithm.

Table 5: SMAP data product fill values.

Type	Value	Pattern
L4_C : Float32	-9999.0	Negative number outside of valid data range
Float32, Float64	+1e15	Large, negative number
Signed8, NormSigned8	-127	Datatype minimum + 1
Signed16, NormSigned16	-32767	Datatype minimum + 1
Signed24	-8388607	Datatype minimum + 1
Signed32	-2147483647	Datatype minimum + 1
Signed64	-9223372036854775807	Datatype minimum + 1
Unsigned8	254	Datatype maximum - 1
Unsigned16	65534	Datatype maximum - 1
Unsigned24	16777214	Datatype maximum - 1
Unsigned32	4294967294	Datatype maximum - 1
Unsigned64	18446744073709551614	Datatype maximum - 1
FixedLenString, VarLenString	NA	Not available

4 Data Definition

4.1 Product Overview

4.1.1 SMAP L4_C Data Product

The daily global 9-km resolution ecosystem productivity variables comprising the L4_C data product are the Net Ecosystem Exchange (NEE), gross primary productivity (GPP), heterotrophic respiration (R_h), surface (~0-5cm depth) soil organic carbon (SOC), along with (3) environmental constraint variables (emult, tmult, wmult), each of which are stored as IEEE single precision (4-byte) real values (e.g. “FLOAT32”|“H5T_IEEE_F32LE” HDFv4/HDFv5 data type. Each of these daily model variables, output at 9-km resolution on the EASE-Grid 2.0 grid projection, are represented by a “mean” and “standard deviation” dataset named after the variable. Further, the native mapped value of each variable is spatially partitioned by the Plant Function Type (PFT) associated with a given pixel, for PFT classes {1..8}, as summarized in Table 2. This organization is reflected in the HDF5 file using groups named after the ecosystem productivity variable itself.

To avoid repeating key properties (e.g., the datatype, units, and rank/dimensions) for the L4_C output variables in the **L4_C Group Contents** tables appearing later, the **L4_C Data Product Overview** table below summarizes these properties for each variable. The placeholder qualifier notation **“{pft}”** shown in Table 6 below is used to indicate that a given ecosystem productivity variable is partitioned the Plant Functional Type (PFT) classifier, for the relevant {1..8} terrestrial PFT’s. In addition to this partitioning by PFT (e.g., “NEE/nee_pft1_mean”), all-PFT mean datasets for NEE, GPP, R_h, and SOC (e.g., “NEE/nee_mean”) are included that are continuous for the gridded land domain.

Table 6: L4_C Data Product Overview

L4_C Data Product Overview				
Group	Dataset	Datatype	Units	Rank and dimensions
/NEE	nee_mean	Float32	g C m^{-2} d^{-1}	2: (3856 columns x 1624 rows)
/NEE	nee_{pft:1..8}_mean	Float32	g C m^{-2} d^{-1}	2: (3856 columns x 1624 rows)
/GPP	gpp_mean	Float32	g C m^{-2} d^{-1}	2: (3856 columns x 1624 rows)
/GPP	gpp_{pft:1..8}_mean	Float32	g C m^{-2} d^{-1}	2: (3856 columns x 1624 rows)
/Rh	rh_mean	Float32	g C m^{-2} d^{-1}	2: (3856 columns x 1624 rows)
/Rh	rh_{pft:1..8}_mean	Float32	g C m^{-2} d^{-1}	2: (3856 columns x 1624 rows)
/SOC	soc_mean	Float32	g C m^{-2}	2: (3856 columns x 1624 rows)
/SOC	soc_{pft:1..8}_mean	Float32	g C m^{-2}	2: (3856 columns x 1624 rows)
/EC	frozen_area	Float32	percent	2: (3856 columns x 1624 rows)
/EC	emult_mean	Float32	percent	2: (3856 columns x 1624 rows)
/EC	tmult_mean	Float32	percent	2: (3856 columns x 1624 rows)
/EC	wmult_mean	Float32	percent	2: (3856 columns x 1624 rows)

/QA	surface_flag	Uint16	dimensionless	2: (3856 columns x 1624 rows)
/QA	nee_rmse_mean	Float32	dimensionless	2: (3856 columns x1624 rows)
/QA	nee_rmse{pft:1..8}_mean	Float32	dimensionless	2: (3856 columns x 1624 rows)
/QA	carbon_model_bitflag	Uint16	dimensionless	2: (3856 columns x 1624 rows)
/GEO	cell_lat	Float32	decimal deg.	2: (3856 columns x 1624 rows)
/GEO	cell_lon	Float32	decimal deg.	2: (3856 columns x 1624 rows)

4.1.2 Geolocation Information Overview

The L4_C product shares certain properties with other SMAP data products, notably its inclusion of the ‘latitude’ and ‘longitude’ geolocation fields which play the same role as coordinate variables as defined within the Climate Forecasting (CF-1.x) conventions. Supporting these conventions facilitates additional software tool interoperability, allowing users of L4_C data to use common software tools such as Panoply and Hdfview to visualize and interactively analyze their data. The L4_C geolocation fields are described in greater detail in section 4.6.5.

4.1.3 HDF5 Group QA Overview

The SMAP L4_C QA group is contained within each daily L4_C granule, rather than written to a separate physical “.qa” file as for other SMAP products. The QA group includes product binary and integer bitfields indicating overall product quality (carbon_model_bitflag), estimated NEE root mean square error (RMSE with units g C m^{-2} d^{-1}; nee_rmse_mean), along with (8) L4_C specific nee_rmse_{pft}_mean variable(s) partitioned by Plant Function Class (PFT). The nee_rmse variables are float32 variables representing the root mean square error of the NEE term, and thus have the same units (g C m^{-2} d^{-1}) as the NEE and GPP variables.

4.1.4 L4_C Metadata Overview

The SMAP L4_C metadata are representative of the entire contents of the file. The metadata appear in two forms, that each provide a complete representation of the product metadata.

One form of the metadata appears as HDF5 Attributes that conform to the ISO 19115-2 model in ISO 19139 compliant XML. Metadata in ISO 19139 conformant XML enables users who are familiar with the ISO metadata standards to extract the metadata they need using software that operates with the ISO 19115-2 model and its formal representation.

The second form of the metadata appears in a set of HDF5 groups that contain a set of HDF5 Attributes. The arrangement and names of these groups and their Attribute components approximate major contents of the ISO model. This second form of HDF5 groups and Attributes enable users who are not familiar with the ISO standard to find the particular metadata elements they need to better comprehend product content and format.

4.2 Data Groups and Volume Estimates

L4_C daily 9-km data variables are stored in the following HDF root **groups** (/NEE, /GPP, /RH, /SOC, /EC, /QA, and /GEO). Note that the geo-location information for this 9-km product includes CF style coordinate variables (separate latitude and longitude bin-center values for each 9-km pixel) which reside in the /GEO group. All L4_C quality control information is stored in the /QA group of each L4_C “mdl” output granule, so there is no separate QA file as with some other SMAP products. Metadata unless otherwise noted, is stored in the /METADATA group. Table 7 summarizes the L4_C granule level data volumes.

Table 7: L4_C Data Volume Estimates

L4_C Data Volume Estimates		
Collection	Granule	Expected Data Volumes
L4_C “mdl”	Single granule combining L4C output variables and QA	133 Mb, **daily** volume
L4_C data product **annual** total (365 day)		48.54 Gb **annual** volume
L4_C data product (**LOM**, 3-yr estimated total **)		145.64 Gb, Life-of-Mission LOM volume

***Data product total is for the 3-year baseline mission, with a single “mdl” product granule, containing QA information, per day.*

4.3 File Naming Conventions

Distributable SMAP L4_C data product file names are 43 characters in length. The first 5 characters in the name of all mission distributable products are 'SMAP_'. These characters identify all products generated by the SMAP mission. The following 5 characters are always 'L4_C_'. These characters identify the L4_C data product. The subsequent 3 characters identify the file Collection ID (fixed as "MDL"). The following 27 characters uniquely identify the data stored in the file based on date, timestamp, version, and collection. The final 3 characters of each SMAP Product file name are '.h5'. These characters specify the format of the data in the file. More specifically, all SMAP L4_C data product file names must conform to the following convention:

> **SMAP_L4_C_**[CollectionID]_[Date/TimeStamp]_[ScienceVersionID]_[Product Counter].[extension]

The outline below describes the content of each field in the file naming convention:

Collection ID – L4_C product and QA granules carry a fixed collection ID of "**MDL**".

Date/Time Stamp – The date/time stamp of the data elements that appear in the product. Date/time stamps in SMAP file names are always recorded in Universal Coordinated Time (UTC). Date/time stamps conform to the following convention:

YYYYMMDD**T**hhmmss

where:

YYYY is the calendar year. The full calendar year must appear in the file name.

MM designates the month of the year. The month designator always occupies two digits. Months that can be represented with fewer than two digits must employ a leading zero.

DD designates the day of the month. The day designator always occupies two digits. Days of the month that can be represented with fewer than two digits must employ a leading zero.

T delineates the date from the time and is a required character in all time stamps in product names.

hh designates the hour of the day on a 24-hour clock in UTC. The hour designator always occupies two digits. Hours that can be represented with fewer than two digits must employ a leading zero.

mm designates the minute of the hour in UTC. The minute designator always occupies two digits. Minutes that can be represented with fewer than two digits must employ leading zeroes.

ss designates the truncated second of the minute in UTC. Fractional second specification is not necessary in file names. The second designator always occupies two digits. Seconds that can be represented with fewer than two digits must employ leading zeroes.

{S} - S{0|1} - data stream ID (0: forward processing (FP), 1: reprocessing (RP)
{M} - major release ID (as defined in the PSD)
{VVV} - L4 version number (e.g., 572 would represent triplet 5.7.2)
{PPP} - product counter (as defined in the PSD)
Extension: the standard .h5 file extension is used.

Science Version ID – The Science Version ID {v(L)(I)(III)} string reflects changes to L4_C algorithm updates that impact the science content of the product. The format of the Science Version ID is as follows:

"V"	The character "V" always precedes the version identifier
Launch indicator {L}	
	Distinguishes between pre-launch and post-launch operating conditions and indicates validation status. One digit launch indicators are assigned as follows: 0: pre-launch output (simulated data) a: output prior to public beta release ("alpha") b: beta-release output v: validated-release output
Major ID {M}	One digit that indicates the major version number. Major changes in algorithm or processing approach will generate an update to this identifier.
Minor ID {VVV}	Three digits that indicate the minor version number. Any change to any component that impacts the science content of the data product will lead to a change in this identifier.

Note that the data product Science Version ID (example: Vv2020) consists of the first six characters of the data product Composite Release ID. **The full Composite Release ID (CRID) includes four additional digits that are to be found in individual granule metadata within the DataIdentifcation / DatasetIdentification / CompositeReleaseID field (see Section 4.4).** The rationale for including only the Science Version ID, in the file name is to relieve user concerns with the more frequent changes in the additional four digits of the CRID. The Science Version ID is incremented whenever a change in the L4_C algorithm or its time-invariant ancillary inputs impacts the science content of the L4_C product, whereas the Composite Release ID uniquely specifies all changes in the L4_C production system, including changes in dynamic ancillary inputs and changes that have no impact on the science content.

Product Counter – The Product Counter tracks the number of times that a particular L4_C data file (or granule) was generated under the same Science Version ID and for the same Collection and time period (or instant). The system assigns a Product Counter of 001 to the first instance of a granule. Subsequent instances of the same granule (that is, representing the same Collection, time period, and Science Version ID) are assigned a Product Counter that represents the next consecutive integer. The Product Counter always occupies three digits. Product Counters that do not require three digits contain leading zeroes

Extension – The extension for all SMAP L4_C data products is “h5”. That extension indicates that the product contents are in HDF5 format.

Example File Names – Based on the above standard, the following name describes an operational data beta-release product from science version 1.003 of the L4_C data product for the period 00:00 UTC on June 9, 2015. The file represents the first time (001) an L4_C (“mdl”) product was generated for the date and time interval in question. An example of a fully formed SMAP L4_C file name follows:

SMAP_L4_C_mdl_20150609T000000_Vv2020_001.h5

Daily QA information for L4_C is stored in the /QA group within the primary “_mdl_” file documented above.

4.4 Metadata Detail

The L4_C data product contains a significant metadata component, designed to address the needs of users in the discovery, retrieval, and analysis phases of their work, as well as the needs of data managers charged with creating and maintaining data provenance information over time. Unless otherwise noted, all metadata elements conform to the NASA Earth Science Data Model.

The metadata emphasis in this document is on granule level (product specific) metadata, recognizing that some of the granule metadata directly contributes to collection level metadata as well. Structurally, two forms of metadata are included: HDF centric ‘attribute’ name-value pair style metadata, and a form of metadata (stored in nested HDF5 groups) used to help facilitate generation of ISO 19115 / 19139 compliant metadata.

ISO metadata injection for L4_C granules is performed at GSFC/GMAO, as part of a shared L4 metadata post-processing phase. Two phases of metadata injection are performed, where the first is to produce the metadata named in this document as HDF5 hierarchical group form (e.g., “group of groups”). In this treatment, within a given HDF5 group, the nested metadata elements themselves are stored as HDF5 attributes.

The second phase of metadata generation conforms with the ISO 19139 (XML) encoding of the ISO 19115 standard, with its associated ISO 19139 XML representation. This metadata post-processing performed by first extracting the HDF5 group metadata from each L4_C HDF5 granule (using h5dump or equivalent tools) to yield each metadata element’s content. The resulting metadata collection is then run through an XSLT post-process transformation using a Saxon parser, reformatting to be conformal to the ISO 19139 XML schema, for subsequent re-injection back into the granule as conformal XML.

4.4.1 L4_C Mandatory Metadata

In addition to the L4_C product specific metadata, a number of metadata elements (currently 43) are defined as mandatory metadata in order for the granule to conform to the Earth Science Data Model (ESDM).

4.4.2 L4_C Product Specific metadata

The following table defines the L4_C product specific metadata.

Table 8: Granule level metadata in the L4_C data product.

ISO Major Class	SMAP HDF5 Metadata SubGroup	Subgroup/Attribute in SMAP HDF5	Valid Values
DQ_DataQuality	DataQuality/L4_C	Scope	`nee, gpp, rh, soc, ec`
		CompletenessOmission/evaluationMethodType	directInternal
		CompletenessOmission/measureDescription	Percent of land grid (terrestrial domain) cells, excluding inland water and permanent ice, that lack geophysical data relative to the total number of land grid cells that fall within the global domain (excluding inland water and permanent ice). Metadata values are computed by pooling data from all elements listed in scope.
		CompletenessOmission/nameOfMeasure	Percent of Missing Data
		CompletenessOmission/value	*<A measure between 0 and 100>*
		CompletenessOmission/unitOfMeasure	Percent
		DomainConsistency/evaluationMethodType	directInternal
		DomainConsistency/measureDescription	Percent of L4_C modeled data that fall within a predefined acceptable range of measure.

		DomainConsistency/nameOfMeasure	Percent of L4_C modeled data that are within the acceptable range.
		DomainConsistency/value	*<A measure between 0 and 100>*
		DomainConsistency/unitOfMeasure	Percent
EX_Extent	Extent	description	Global land excluding inland water and permanent ice.
		westBoundLongitude	-180 degrees
		eastBoundLongitude	180 degrees
		southBoundLatitude	-85.04456 degrees
		northBoundLatitude	85.04456 degrees
		rangeBeginningDateTime	*<Time stamp that indicates the initial time element in the product.>*
		rangeEndingDateTime	*<Time stamp that indicates the final time of data in the product.>*
LI_Lineage/LE_Process Step	ProcessStep	processor	Soil Moisture Active Passive (SMAP) Mission Science Data System (SDS) Operations Facility
		stepDateTime	*<A date time stamp that specifies when the processing of the data product took place.>*

		processDescription	L4_C models FPAR 8-DAY composite, GEOS5-FP, and SMAP L4_SM and L3_SM_A data within a global carbon model to generate spatially complete and temporally continuous estimates of NEE (net ecosystem exchange), GPP, Rh, SOC and associated ecosystem variables at 9-km.
		documentation	*<Software description document title, version, and date stamp.>*
		identifier	L4_C_SPS
		runTimeParameters	*<Specify any run time parameters if they were used.>*
		SWVersionID	*<A software version identifier that runs from 001 to 999>*
		softwareDate	*<A date stamp that specifies when software used to generate this product was released.>*
		softwareTitle	Level 4 Surface and Root Zone Soil Moisture SPS
		timeVariableEpoch	J2000
		epochJulianDate	2451545.00
		epochUTCDate	2000-01-01T11:58:55.816Z
		ATBDTitle	Soil Moisture Active Passive (SMAP) Level 4 Carbon (L4_C) Algorithm Theoretical Basis Document (ATBD)
		ATBDDate	*<Time stamp that specifies the release date of the ATBD>*
		ATBDVersion	*<Latest ATBD version number.>*

		algorithmDescription	The SMAP L4_C global daily algorithm provides spatially and temporally complete net ecosystem exchange (NEE), gross primary productivity (GPP), heterotrophic respiration (Rh), Soil Organic Carbon (SOC), and related variables at 9-km for science and applications users.
		algorithmVersionID	*<An algorithm version identifier that runs from 001 to 999>*
		version	*<The SMAP Composite Release ID(s) of the input data product(s).>*
		identifier	SPL2SMDL
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/L4_SM	description	The surface meteorological data from L4_SM algorithm.
		fileName	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*
		version	*<Version ID(s) of the input data product(s).>*
		identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/MODIS FPAR 8-DAY	Description	MODIS FPAR 8-day (MOD15A2)
		filename	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*

		Version	*<Version ID(s) of the input data product(s).>*
		Identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/L4_C_MET	Description	L4_C MET Pre-processor output (L4CMET) to daily L4_C model
		Filename	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*
		Version	*<Version ID(s) of the input data product(s).>*
		Identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/L4_C_MODIS	Description	L4_C MODIS Pre-processor output (L4CMOD) to daily L4_C model
		Filename	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*
		Version	*<Version ID(s) of the input data product(s).>*
		Identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*

	Source/L4_C_SOC	Description	L4_C Soil Organic Carbon (SOC) daily intermediate variables updated daily by L4_C model
		Filename	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*
		Version	*<Version ID(s) of the input data product(s).>*
		Identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/L4_C_NDVI	Description	L4_C VIIRs NDVI backup source for calculating FPAR
		Filename	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*
		Version	*<Version ID(s) of the input data product(s).>*
		Identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/L4_C FPAR Climatology	Description	L4_C FPAR 365-day climatology as backup FPAR source for daily L4_C model
		Filename	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*

		Version	*<Version ID(s) of the input data product(s).>*
		Identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
	Source/GEOS5_FP	description	The observation-corrected surface meteorological data from the NASA GEOS-5 system. Precipitation corrected using the gauge-based NOAA Climate Prediction Center Unified precipitation product.
		fileName	*<Complete file name(s) of the input data product.>*
		creationDate	*<Date stamp(s) of the input data product(s).>*
		version	*<Version ID(s) of the input data product(s).>*
		identifier	*<Short name(s) of the input data product(s).>*
		DOI	*<Digital object identifier(s) of the input data product(s), if available.>*
DS_Dataset/MD_DataIdentification	DatasetIdentification	creationDate	*<Date when the L4_C data product file was created>*
		VersionID	*<SMAP Science Version ID associated with this data product – See section (4.3) of the L4_C Product Specification Document.>*

		CompositeReleaseID	*<SMAP Composite Release ID (CRID) associated with this data product. The CRID is the 4 digit Science Version ID shown in the filename with an additional four-digit unique appendix. – See Section 4.3 of this document and Section 4.3 of the L4_SM Product Specification Document.>*
		ECSVersionID	*<Three-character string that identifies the type of ECS data stream and reflects the major science version of the product, for example: "001" = first operational processing stream; "002" = second operational processing stream (supersedes first stream after major algorithm update and includes reprocessing); "003" = third operational processing stream (supersedes second stream…); "199" = candidate operational stream (a.k.a. OASIS or parallel processing stream) >*
		UUID	*<Universally Unique identifier (36-character string)>*
		fileName	*<Name of the L4_C output data file.>*
		originatorOrganizationName	NTSG group at the University of Montana, Missoula, Montana, USA

		longName	SMAP Level 4 Carbon <*Collection ID*> Collection
		shortName	<*ECS Short Name*>
		SMAPShortName	L4_C <*Collection ID*> Collection
		abstract	The SMAP L4_C data product provides global, daily net ecosystem exchange, GPP and related variables 9 km resolution. The L4_CM data product consists of one Collection of data granules (or files): the model (MDL) data Collection (shortName=SPL4SMDL) .
		characterSet	utf8
		credit	The software that generates the L4_C data product and the data system that automates its production were designed and implemented at the NTSG group at the University of Montana, Missoula, Montana, USA.
		language	eng
		purpose	The SMAP L4_C data product provides spatially and temporally complete net ecosystem exchange, GPP, Rh, Soil Organic Carbon and related variables for science and applications users.
		status	"on-going" ("completed" after final reprocessing)
		topicCategoryCode	geoscientificInformation
		QACreationDate	<*The generation date of the QA product that accompanies the L4_C data granule.*>
		QAFileName	<*The name of QA product.*>

		QAAbstract	An HDF5 product that contains statistical information on data product results. These statistics enable data producers and users to assess the quality of the data in the data product granule.
DS_Series/MD_DataIdentification	SeriesIdentification	revisionDate	*<Date and time of the software release that was used to generate this data product.>*
		VersionID	*<SMAP Science Version ID associated with this data product – See section 4.3 of the L4_C Product Specification Document.>*
		CompositeReleaseID	*< SMAP Composite Release ID (CRID) associated with this data product. The CRID is the 4-digit Science Version ID shown in the filename with an additional four-digit unique appendix. – See Section 4.3 of this document and Section 4.3 of the L4_SM Product Specification Document.>*

		ECSVersionID	*<Three-character string that identifies the type of ECS data stream and reflects the major science version of the product, for example:* *"001" = first operational processing stream;* *"002" = second operational processing stream (supersedes first stream after major algorithm update and includes reprocessing);* *"003" = third operational processing stream (supersedes second stream…);* *"199" = candidate operational stream (a.k.a. OASIS or parallel processing stream)* >
		longName	SMAP Level 4 Carbon <*Collection ID*> Collection
		shortName	*< ECS Short Name>*
		identifier_product_DOI	*<digital object identifier>*
		resourceProviderOrganizationName	National Aeronautics and Space Administration

		abstract	The SMAP L4_C data product provides global, daily net ecosystem exchange, GPP and related variables 9 km resolution. The L4_CM data product consists of one Collection of data granules (or files): the model (MDL) data Collection (shortName=SPL4SMDL) .
		characterSet	utf8
		credit	The software that generates the L4_C data product and the data system that automates its production were designed and implemented at the NTSG group at the University of Montana, and at NASA/GSFC, Maryland, USA.
		language	Eng
		purpose	The SMAP L4_C data product provides spatially and temporally complete net ecosystem exchange, GPP, Rh, Soil Organic Carbon and related variables for science and applications users.
		status	“on-going” (“completed” after final reprocessing)
		topicCategoryCode	geoscientificInformation
		pointOfContact	NTSG group at the University of Montana, Missoula, Montana, and the National Snow and Ice Data Center, Boulder, Colorado, USA.
		PSDPublicationDate	*<Date of publication of the Product Specification Document>*
		PSDEdition	*<Edition identifier for the Product Specification Document>*

		PSDTtitle	Soil Moisture Active Passive Mission Level 4 Carbon (L4_C) Product Specification Document
		SMAPShortName	L4_C
		mission	Soil Moisture Active Passive (SMAP)
		maintenanceAndUpdateFrequency	asNeeded
		maintenanceDate	*<Specifies a date when the next update to this product might be anticipated.>*
		format	HDF5
		formatVersion	*<HDF5 version number>*
MD_GridSpatialRepresentation	GridSpatialRepresentation	Latitude/dimensionSize	1624
		Latitude/resolution	9 km on average
		Longitude/dimensionSize	3856
		Longitude/resolution	9 km on average
MD_AcquisitionInformation	AcquisitionInformation	platform/antennaRotationRate	*<Antenna rotation rate in rpm>*
		platformDocument/publicationDate	*<The date of publication of the document that describes the SMAP platform, if available to the general public.>*

		platformDocument/edition	*<The edition of publication of the document that describes the SMAP platform, if available to the general public.>*
		platformDocument/title	*<The title of the publication of the document that describes the SMAP platform, if available to the general public.>*
		platform/description	The SMAP observatory houses an L-band radiometer that operates at 1.4 GHz and an L-band radar that operates at 1.26-1.29 GHz. The instruments share a rotating reflector antenna with a 6-meter aperture that scans over a 1000 km swath. The bus is a 3-axis stabilized spacecraft that provides momentum compensation for the rotating antenna.
		platform/identifier	SMAP
		radarDocument/publicationDate	*<The date of publication of the document that describes the SMAP radar instrument, if available to the general public.>*
		radarDocument/edition	*<The edition of publication of the document that describes the SMAP radar instrument, if available to the general public.>*
		radarDocument/title	*<The title of the publication of the document that describes the SMAP radar instrument, if available to the general public.>*

		radar/description	The SMAP radar instrument employs an L-band conically scanned system and SAR processing techniques to achieve moderate resolution (1 km) backscatter measurements over a very wide 1000 km swath.
		radar/identifier	SMAP SAR
		radar/type	L-band Synthetic Aperture Radar
		radiometerDocument/publicationDate	<*The date of publication of the document that describes the SMAP radiometer instrument, if available to the general public.*>
		radiometerDocument/edition	<*The edition of publication of the document that describes the SMAP radiometer instrument, if available to the general public.*>
		radiometerDocument/title	<*The title of the publication of the document that describes the SMAP radiometer instrument, if available to the general public.*>
		radiometer/description	The SMAP L-band Radiometer records V-pol, H-pol, 3rd and 4th Stokes brightness temperatures at 40 km resolution at 4.3 Megatbits per second with accuracies of 1.3 Kelvin or better.
		radiometer/identifier	SMAP RAD
		radiometer/type	L-band Radiometer

4.5 Data Structure

The data structure of a L4_C granule (HDF5) file is described below, first in its own HDF Group table to show which ***group*** the indicated dataset is stored within in a HDF5 product granule file. Following these, tables describing a more detailed view of compound (bit-flag) variables and variables with more complex internal organization are included.

4.6 Element definitions

A single L4_C daily global 9-km resolution HDF5 product granule is produced each day. Internally, the primary L4_C output HDF5 granule contains (5) HDF5 groups, named after the primary variable indicated e.g. {NEE, GPP, RH, SOC, EC, GEO}, with a final group (GEO) storing geolocation information. Geolocation information in group (GEO) consists of {latitude, longitude} coordinate variables, in decimal degree units, that enable convenient geo-referenced viewing and analysis. All L4_C quality assurance information, in group /QA are stored within the primary L4_C daily HDF5 product file.

Within each of the primary L4_C granule's HDF5 groups, the indicated variable is partitioned by each of the (8) Plant Function Type classes, such that there are (8) scientific datasets per group. For example, in the NEE group, there are (8) "NEE" output variables, one per PFT class.

The output variables are summarized in Table 6. Note that the primary QA variable is also partitioned by PFT class. Both the QA (partitioned by PFT) and Geolocation information is stored in a companion daily 9-km HDF5 granule file, whose name follows the conventions as documented in Section 4.3 above.

4.6.1 NEE Dataset Definition

Table 9: NEE Group Contents

L4_C Contents of HDF Group: NEE		
Group	Dataset	Description
NEE	nee_mean	Mean Net Ecosystem Exchange value, not distinguished by PFT; (units = g C m^{-2} d^{-1} ; valid_min=-30.0; valid_max =20.0)
NEE	nee_std_dev	Standard deviation of Net Ecosystem Exchange, not distinguished by PFT.
NEE	nee_pft_1_mean	Mean value of NEE for grid cells classified in PFT class 1 (evergreen needleleaf)
NEE	nee_pft_2_mean	Mean value of NEE for pixels classified in PFT class 2 (evergreen broadleaf)
NEE	nee_pft_3_mean	Mean value of NEE for pixels classified in PFT class 3 (deciduous needleleaf)
NEE	nee_pft_4_mean	Mean value of NEE for pixels classified in PFT class 4 (deciduous broadleaf
NEE	nee_pft_5_mean	Mean value of NEE for pixels classified in PFT class 5 (shrub)
NEE	nee_pft_6_mean	Mean value of NEE for pixels classified in PFT class 6 (grass)
NEE	nee_pft_7_mean	Mean value of NEE for pixels classified in PFT class 7 (cereal crop)
NEE	nee_pft_8_mean	Mean value of NEE for pixels classified in PFT class 8 (broadleaf crop)

4.6.2 GPP Dataset Definition

The following table describes the datasets contained in the HDF group "GPP" in the L4_C product files:

Table 10: GPP Group Contents

L4_C Contents of HDF Group: GPP		
Group	Dataset	Description
GPP	GPP_mean	Mean Gross Primary Productivity per pixel, not distinguished by PFT (units = g C m^{-2} d^{-1} ; valid_min=0.0; valid_max =30.0)
GPP	GPP_std_dev	Standard deviation of Gross Primary Productivity per pixel, not distinguished by PFT
GPP	GPP_pft_1_mean	Mean value of GPP for pixels classified as PFT class 1
GPP	GPP_pft_2_mean	Mean value of GPP for pixels classified as PFT class 2
GPP	GPP_pft_3_mean	Mean value of GPP for pixels classified as PFT class 3
GPP	GPP_pft_4_mean	Mean value of GPP for pixels classified as PFT class 4
GPP	GPP_pft_5_mean	Mean value of GPP for pixels classified as PFT class 5
GPP	GPP_pft_6_mean	Mean value of GPP for pixels classified as PFT class 6
GPP	GPP_pft_7_mean	Mean value of GPP for pixels classified as PFT class 7
GPP	GPP_pft_8_mean	Mean value of GPP for pixels classified as PFT class 8

4.6.3 Heterotrophic respiration (R_h) Dataset definition

The following table describes the datasets contained in the HDF group "Rh" in the L4_C product files:

Table 11: Rh Group Contents

L4_C Contents of HDF Group: RH		
Group	Dataset	Description
RH	rh_mean	Mean heterotrophic respiration (Rh) value, not distinguished by PFT (units = g C m^{-2} d^{-1} ; valid_min=0.0 ; valid_max=20.0)
RH	rh_std_dev	Standard deviation of heterotrophic (Rh) respiration, not distinguished by PFT
RH	rh_pft_1_mean	Mean value of Rh for pixels classified as PFT class 1
RH	rh_pft_2_mean	Mean value of Rh for pixels classified as PFT class 2
RH	rh_pft_3_mean	Mean value of Rh for pixels classified as PFT class 3
RH	rh_pft_4_mean	Mean value of Rh for pixels classified as PFT class 4
RH	rh_pft_5_mean	Mean value of Rh for pixels classified as PFT class 5
RH	rh_pft_6_mean	Mean value of Rh for pixels classified as PFT class 6
RH	rh_pft_7_mean	Mean value of Rh for pixels classified as PFT class 7
RH	rh_pft_8_mean	Mean value of Rh for pixels classified as PFT class 8

4.6.4 Soil Organic Carbon (SOC) Dataset definition

The following table describes the datasets contained in the HDF group "SOC" (for soil organic carbon) in the L4_C product files:

Table 12: SOC Group Contents

L4_C Contents of HDF Group: Soil Organic Carbon (SOC)		
Group	Dataset	Description
SOC	soc_mean	Mean surface soil organic carbon (SOC) value, not distinguished by PFT (units = g C m^{-2} ; valid_min=0.0; valid_max=25,000)
SOC	soc_std_dev	Standard deviation of surface soil organic carbon (SOC), not distinguished by PFT (units = g C m^{-2}).
SOC	soc_pft_1_mean	Mean value of SOC for pixels classified as PFT class 1
SOC	soc_pft_2_mean	Mean value of SOC for pixels classified in PFT class 2
SOC	soc_pft_3_mean	Mean value of SOC for pixels classified in PFT class 3
SOC	soc_pft_4_mean	Mean value of SOC for pixels classified in PFT class 4
SOC	soc_pft_5_mean	Mean value of SOC for pixels classified in PFT class 5
SOC	soc_pft_6_mean	Mean value of SOC for pixels classified in PFT class 6
SOC	soc_pft_7_mean	Mean value of SOC for pixels classified in PFT class 7
SOC	soc_pft_8_mean	Mean value of SOC for pixels classified in PFT class 8

The table below describes the datasets contained in the environmental constraints "EC" HDF group in the L4_C product files. Note that these variables are not spatially partitioned by Plant Function Type (PFT) as the other science variables above were.

Table 13: EC -- Environmental Constraints Group Contents

L4_C Contents of HDF Group: EC		
Group	Dataset	Description
EC	frozen_area	Mean frozen_area, per 9x9 km pixel, not distinguished by PFT (units = percent constraint; min=0 fully constrained; max = 100.0 not constrained)
EC	emult_mean	Average estimated bulk environmental constraint to PAR conversion efficiency and vegetation productivity, not distinguished by PFT (units = percent constraint; min=0 fully constrained; max = 100.0 not constrained)
EC	tmult_mean	Average estimated bulk environmental constraint on heterotrophic respiration due to soil temperature, not distinguished by PFT (units = percent constraint; min=0 fully constrained; max = 100.0 not constrained)
EC	wmult_mean	Average estimated bulk environmental constraint on heterotrophic respiration due to soil moisture, not distinguished by PFT (units = percent constraint; min=0 fully constrained; max = 100.0 not constrained)

4.6.5 Geolocation Information

The following Table describes the HDF5 datasets providing geolocation information. Note that for interoperability with common science image data viewers such as Panoply (http://www.giss.nasa.gov/tools/panoply/) and HDFView (http://www.hdfgroup.org/hdf-java-html/hdfview/), the geolocation variables described here are intended to be functionally equivalent to CF convention "coordinate variables". In this way, they convey the geographic bin-center coordinate position of each L4_C science variable pixel to the viewer.

The *EASE2_global_projection* HDF5 Dataset in the root-level group contains attributes describing the parameters of the cylindrical 9 km Earth-fixed EASE-Grid 2.0 projection and defines the mapping from latitude/longitude to grid-native coordinates, following CF (Climate and Forecast) metadata conventions. The *x* and *y* HDF5 Datasets contain grid-native coordinate values from the cylindrical 9 km Earth-fixed EASE-Grid 2.0 projection. The *grid_mapping = EASE2_global_projection* attribute in all HDF5 Datasets signals that its grid-native coordinate system is defined in the *EASE2_global_projection* Dataset. The *standard_name = projection_x_coordinate* and *standard_name = projection_y_coordinate* attributes in the *x* and *y* HDF5 Datasets associate them as the grid-native coordinates of the projection, following CF metadata conventions. The dimension scales in all HDF5 Datasets associates the dimensions of that Dataset with the grid-native coordinates of the *x* and *y* Datasets. Dimension scales provide an additional (non-CF specific) way to associate the grid-native coordinate HDF5 Datasets with other Dataset dimensions in HDF5.

Table 14: Overview of Geolocation Information

L4_C Geolocation Information			
Group	Dataset Name	Datatype	Description
GEO	latitude	float32	Latitude, in decimal degrees, of the center of the Earth based grid cell.
GEO	longitude	float32	Longitude, in decimal degrees, of the center of the Earth based grid cell.
/	EASE2_global_projection	string	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. Type is string.
/	x	float64	The x coordinate values from the cylindrical 9 km Earth-fixed EASE-Grid 2.0 projection.
/	y	float64	The y coordinate values from the cylindrical 9 km Earth-fixed EASE-Grid 2.0 projection.

4.6.6 QA : the Quality Assurance Group

The overall quality of the L4_C daily 9-km product, by pixel, is described by several types of quality metrics:

- **carbon_qual_bitflag**.
- **nee_rmse_mean**
- **nee_rmse_pft(1..8)_mean**
- **qa_count_pft{1..8}**
- **qa_count**

Note that the **carbon_model_bitflag** provides ancillary quality related information in the form of a set of decision-tree bit-fields, to assist analysts in performing deeper inquiries describing overall L4_C product quality. The (4) types of L4_C quality metrics are further described in the below.

Table 15: Overview of Quality Group: QA

			L4_C Contents of HDF Group: QA
Group	Dataset Name	Datatype	Description
QA	carbon_model_bitflag	uint16	Bit fields indicating overall L4_C product quality for 9-km grid cell.
QA	nee_rmse_mean	float32	Daily NEE (nee_mean) estimated root mean square error, with a valid range of {0.0 <= nee_rmse_mean <= 20.0}.
QA	nee_rmse_{pft:1..8}_mean	float32	Daily NEE (nee_mean_{pft:1..8}) estimated root mean square error partitioned by Plant Function Type (PFT) {1..8}.
QA	qa_count_{pft:1..8}	uint8	Time-static count of 1-km grid cells simulated (from {0...81} possible) for each vegetated PFT present within 9-km cell, excluding cells which are non-land, non-vegetated, and/or otherwise lack a valid FPAR data record. Percent vegetated land coverage with 9-km grid cell calculated as qa_count/81 * 100.
QA	qa_count	uint8	Time-static count of 1-km grid cells simulated (from {0...81} possible) for all cells present within 9-km grid cell, excluding cells which are non-land, non-vegetated, and/or otherwise lack a valid FPAR data record. Percent coverage of each PFT within 9-km grid cell calculated as pft_qa_count{1..8}/81 * 100.

4.6.7 The carbon_model_bit-flag

The **carbon_model_bitflag** (uint16 bit-flag) variable is represented as a single, spatially explicit quality field output on the 9-km grid.

Table 16: L4_C Quality Control Variable (carbon_model_bitflag) Definition

L4_C Quality Control Variable (carbon_model_bitflag) Definition			
Bit field	Bit position (or range)	Possible Values	Description
NEE-bit	0	{0|1}	When ‘0’, all 1-km NEE values within 9-km grid cell lie within their valid range {min, max}
GPP-bit	1	{0|1}	When ‘0’, all 1-km GPP values within 9-km grid cell lie within their valid range {min, max}.
Rh-bit	2	{0|1}	When ‘0’, all 1-km Rh values within 9-km grid cell lie within their valid range {min, max}.
SOC-bit	3	{0|1}	When ‘0’, all 1-km SOC values within 9-km grid cell lie within their valid range {min, max}.
PFT_dominant	4-7 (4 bits)	Integer value {1…8} inclusive.	Time-static most frequently occurring (dominant) vegetated PFT class within each 9-km cell as defined from qa_count.
QA_score	8-11 (4bits)	Integer value {0…3} inclusive.	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	{0|1}	0=derived GPP using 8-day FPAR or NDVI input, 1=derived GPP via FPAR or NDVI climatology.
NDVI_method	13	{0|1}	0= derived GPP using FPAR; 1= derived GPP using NDVI
FT_method	14	{0|1}	0=SMAP F/T retrieval used, or: 1=SMAP F/T not available, so used Tsurf (surface temperature) to determine F/T state.
IsFill	15	{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. NOTE: if IsFill=1 the entire uint16 evaluates to 65534.

4.6.8 The NEE RMSE QA metrics: nee_rmse_mean, and nee_rmse_mean partitioned by PFT

The QA scheme used in the L4_C product defines an **NEE**-driven metric, as the float32 datatype dataset "**nee_rmse_mean**" in units of g C m^{-2} d^{-1}**.** This daily 9-km global **NEE** root mean square error (RMSE) term is produced using error propagation of L4_C daily input data fields with assumed constant error or constant proportional error depending on the field. The NEE RMSE estimates are also provided for each Plant Function Type (PFT; "**nee_rmse_pft{1..8}_mean**"), where the QA score refers to the mean-square quantity for each PFT. Even though the valid RMSE values range from 0.0 to 20.0 g C m^{-2} d^{-1}, typical RMSE values range from 0.0 to 4.0 g C m^{-2} d^{-1} (Table 17) and generally proportional to the daily total of GPP and Rh fluxes.

Table 17: L4_C Quality Control Variable (nee_rmse_{pft:1..8}_mean Definition

Dataset Variable	Value	Description
nee_rmse_pft1_mean	{0.0 to 20.0}	For PFT 1, in g C m^{-2} d^{-1}
nee_rmse_pft2_mean	{0.0 to 20.0}	For PFT 2, in g C m^{-2} d^{-1}
nee_rmse_pft3_mean	{0.0 to 20.0}	For PFT 3, in g C m^{-2} d^{-1}
nee_rmse_pft4_mean	{0.0 to 20.0}	For PFT 4, in g C m^{-2} d^{-1}
nee_rmse_pft5_mean	{0.0 to 20.0}	For PFT 5, in g C m^{-2} d^{-1}
nee_rmse_pft6_mean	{0.0 to 20.0}	For PFT 6, in g C m^{-2} d^{-1}
nee_rmse_pft7_mean	{0.0 to 20.0}	For PFT 7, in g C m^{-2} d^{-1}
nee_rmse_pft8_mean	{0.0 to 20.0}	For PFT 8, in g C m^{-2} d^{-1}

4.7 L4_C Data Product Volume Estimates

The table below summarizes early estimates of the L4_C pre-product data volume, with a granule data size of 133 Mb granule per day.

Table 18: Daily L4_C Data Product Volume Estimates

Daily L4_C Data Product Volume Estimate			
HDF Group	Description	Bytes	Estimated compressed volume, in Mb
Metadata	Collective granule level Metadata, both ESDM and ISO forms	160000 (estimate)	0.125 Mb
GEO	Geolocation (latitude, longitude) and all QC, bit-flags	75145728	2 Mb
QA	(8) QA (nee_rmse_mean, plus (8) layers, 1 per PFT class.		25 Mb
QA	1 (uint16) surface_flags, 1 bitflags (uint16) layer		
QA	(8) QA pft_freq 9-km uint8 layers, 1 per PFT class and (1) pft_mod_cnt uint8 9-km layer		
GPP	Gross Primary productivity datasets for all (8) PFTs	250485760	106 Mb
Rh	Rh for (8) PFTs	250485760	
NEE	NEE for (8) PFTs	250485760	
SOC	SOC for (8) PFTs	250485760	
EC	(3) environmental constraint variables (emult, tmult, wmult)	25048576	
TOTALS	Granule, total Gb	= 0.12 Gb (133 Mb)	

5 Appendices

Appendix A: Acronyms and abbreviations

This is the standard Soil Moisture Active Passive (SMAP) Science Data System (SDS) list of acronyms and abbreviations. Not all of these acronyms and abbreviations appear in every SMAP SDS document.

ADT	Algorithm Development Team
AMSR	Advanced Microwave Scanning Radiometer
ANSI	American National Standards Institute
AOS	Acquisition of Signal
APF	Algorithm Parameter File
ARS	Agricultural Research Service
ASF	Alaska Satellite Facility
ATBD	Algorithm Theoretical Basis Document
ATLO	Assembly Test Launch and Operations
BFPQ	Block Floating Point Quantization
BIC	Beam Index Crossing
CARA	Criticality and Risk Assessment
CBE	Current Best Estimate
CCB	Configuration Control Board
CCSDS	Consultative Committee on Space Data Systems
CDR	Critical Design Review
CEOS	Committee on Earth Observing Systems
CF	Climate and Forecast
CM	Configuration Management
CM	Center of Mass
CONUS	Continental United States
COTS	Commercial Off the Shelf
CR	Change Request
DAAC	Distributed Active Archive Center
DB	Database

DBA	Database Administrator
dB	decibels
deg	degrees
deg/sec	degrees per second
deg C	degrees Celsius
DEM	Digital Elevation Model
DFM	Design File Memorandum
DIU	Digital Interface Unit
DN	Data Number
DOORS	Dynamic Object Oriented Requirements
DQC	Data Quality Control
DSK	Digital Skin Kernel
DVD	Digital Versatile Disc
EASE	Equal Area Scalable Earth
ECMWF	European Centre for Medium Range Weather Forecasts
ECHO	EOS Clearing House
ECI	Earth Centered Inertial Coordinate System
ECR	Earth Centered Rotating Coordinate System
ECR	Engineering Change Request
ECS	EOSDIS Core System
EDOS	EOS Data Operations System
EM	Engineering Model
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EPO	Education and Public Outreach
ESDIS	Earth Science Data and Information System Project
ESDT	Earth Science Data Type
ESH	EDOS Service Header
ESSP	Earth Science System Pathfinder
ET	Ephemeris TIme
EU	Engineering Units

FOV	Field of View
FRB	Functional Requirements Baseline
FS	Flight System
FSW	Flight Software
F/T	Freeze/Thaw
FTP	File Transfer Protocol
GByte	gigabyte
GDS	Ground Data System
GEOS-5	Goddard Earth Observing System, version 5
GHA	Greenwich Hour Angle
GHz	gigahertz
GLOSIM	Global Simulation
GMAO	Global Modeling and Assimilation Office
GMT	Greenwich Mean Time
GN	Ground Network
GPMC	Governing Program Management Council
GPP	Gross Primary Production
GPS	Global Positioning System
GSE	Ground Support Equipment
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
HK	Housekeeping (telemetry)
Hz	Hertz
HSD	Health and Status Data
ICE	Integrated Control Electronics
ICESat	Ice, Cloud and Land Elevation Satellite
IDL	Interactive Data Language
I&T	Integration and Test
ICD	Interface Control Document
IEEE	Institute of Electrical and Electronics Engineers
IFOV	Instantaneous Field of View

I/O	Input/Output
IOC	In-Orbit Checkout
IRU	Inertial Reference Unit
ISO	International Organization for Standardization
IV&V	Independent Verification and Validation
ITAR	International Traffic in Arms Regulations
I&T	Integration and Test
JPL	Jet Propulsion Laboratory
KHz	kilohertz
km	kilometers
LAN	Local Area Network
LBT	Loopback Trap
LDAS	Land Data Assimilation System
LEO	Low Earth Orbit
LEOP	Launch and Early Operations
LOE	Level Of Effort
LOM	Life Of Mission
LOS	Loss of Signal
LSK	Leap Seconds Kernel
LZPF	Level Zero Processing Facility
m	meters
MHz	megahertz
MIT	Massachusetts Institute of Technology
MMR	Monthly Management Review
MOA	Memorandum of Agreement
MOC	Mission Operations Center
MODIS	Moderate Resolution Imaging Spectroradiometer
MOS	Mission Operations System
m/s	meters per second
ms	milliseconds
MS	Mission System

NAIF	Navigation and Ancillary Information Facility
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Protection
NCP	North Celestial Pole
NCSA	National Center for Supercomputing Applications
NEDT	Noise Equivalent Diode Temperature
NEE	Net Ecosystem Exchange
NEN	Near Earth Network
netCDF	Network Common Data Form
NFS	Network File System/Server
NISN	NASA Integrated Services Network
NRT	Near Real Time
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
NVM	Non-Volatile Memory
NWP	Numerical Weather Product
n/a	not applicable
OCO	Orbiting Carbon Observatory
OEF	Orbit Events File
ORBNUM	Orbit Number File
OODT	Object Oriented Data Technology
ORR	Operational Readiness Review
ORT	Operational Readiness Test
OSSE	Observing System Simulation Experiment
OSTC	One Second Time Command
PALS	Passive and Active L-Band System
PALSAR	Phased Array L-Band Synthetic Aperture Radar
PcK	Planetary Constants Kernel
PDR	Preliminary Design Review
PPPCS	Pointing, Position, Phasing and Coordinate System
PR	Problem Report

PRF	Pulse Repetition Frequency
PRI	Pulse Repetition Interval
PROM	Programmable Read Only Memory
PSD	Product Specification Document
QA	Quality Assurance
rad	radians
RAM	Random Access Memory
RBA	Reflector Boom Assembly
RBD	Rate Buffered Data
RBE	Radiometer Back End
RDD	Release Description Document
RDE	Radiometer Digital Electronics
RF	Radio Frequency
RFA	Request For Action
RFE	Radiometer Front End
RFI	Radio Frequency Interference
RMS	root mean square
RSS	root sum sqaure
ROM	Read Only Memory
RPM	revolutions per minute
RVI	Radar Vegetation Index
SA	System Administrator
SAR	Synthetic Aperture Radar
S/C	Spacecraft
SCE	Spin Control Electronics
SCLK	Spacecraft Clock
SDP	Software Development Plan
SDS	Science Data System
SDT	Science Definition Team
SI	International System
SITP	System Integration and Test Plan

SMAP	Soil Moisture Active Passive
SMEX	Soil Moisture Experiment
SMOS	Soil Moisture and Ocean Salinity Mission
SMP	Software Management Plan
SNR	signal to noise ratio
SOC	Soil Organic Carbon
SOM	Software Operators Manual
SQA	Software Quality Assurance
SPDM	Science Process and Data Management
SPG	Standards Process Group
SPK	Spacecraft Kernel
SQA	Software Quality Assurance
SPS	Science Production Software
SRF	Science Orbit Reference Frame
SRR	System Requirements Review
SRTM	Shuttle Radar Topography Mission
SSM/I	Special Sensor Microwave/Imager
STP	Software Test Plan
sec	seconds
TAI	International Atomic Time
Tb	Brightness Temperature
TBC	To Be Confirmed
TBD	To Be Determined
TBR	To Be Resolved
TBS	To Be Specified
TCP/IP	Transmission Control Protocol/Internet Protocol
TEC	Total Electron Content
TM	Trademark
TOA	Time of Arrival
TPS	Third Party Software
UML	Unified Modeling Language

U-MT	University of Montana
USDA	United States Department of Agriculture
UTC	Coordinated Universal Time
V&V	Verification and Validation
VWC	Vegetation Water Content

Appendix B: SMAP Data Product Specification Documents

SMAP Level 1A Radar Product Specification Document. SMAP Project, JPL D-72543, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 1B Radar (L1C_S0_LoRes) Product Specification Document. SMAP Project, JPL D-72544, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 1A Radiometer Product Specification Document. SMAP Project, JPL D-72554, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 1B Radiometer (L1B_TB) Product Specification Document. SMAP Project, JPL D-72552, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 1C Radiometer (L1C_TB) Product Specification Document. SMAP Project, JPL D-72545, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 2 Active Soil Moisture (L2_SM_A) Product Specification Document. SMAP Project, JPL D-72546, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 2 Passive Soil Moisture (L2_SM_P) Product Specification Document. SMAP Project, JPL D-72547, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 2 Active/Passive Soil Moisture (L2_SM_AP) Product Specification Document. SMAP Project, JPL D-72548, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 3 Freeze-Thaw (L3_FT_A) Product Specification Document. SMAP Project, JPL D-72549, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 3 Active Soil Moisture (L3_SM_A) Product Specification Document. SMAP Project, JPL D-72550, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 3 Passive Soil Moisture (L3_SM_P) Product Specification Document. SMAP Project, JPL D-72551, Jet Propulsion Laboratory, Pasadena, CA.

SMAP Level 3 Active/Passive Soil Moisture (L3_SM_AP) Product Specification Document. SMAP Project, JPL D-72552, Jet Propulsion Laboratory, Pasadena, CA.

Reichle, R. H., R. A. Lucchesi, J. V. Ardizzone, G.-K. Kim, E. B. Smith, and B. H. Weiss (2018), Soil Moisture Active Passive (SMAP) Mission Level 4 Surface and Root Zone Soil Moisture (L4_SM) Product Specification Document, NASA GMAO Office Note, No. 10 (Version 1.5), National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 83pp.

Endsley, K. A., Glassy, J., 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 (2021), Soil Moisture Active Passive (SMAP) Mission Level 4 Carbon (L4_C) Product Specification Document, NASA GMAO Office Note, No. 11 (Version 2.1), National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 69pp.

Appendix C: L4C Dimensions

Table 18 lists all of the Dimensions that are used by data elements in the L4_C data product. The name of each Dimension matches the name given in the Dimension column below. The table also lists the anticipated nominal value and the maximum value for each dimension that appears in the L4_C data product.

Table 18: Dimensions in the SMAP L4_C data product.

Dimension	Nominal Size	Maximum Size
Latitude	1624	1624
Longitude	3856	3856

Appendix D: L4_C Shapes

Table 19: Shapes in the SMAP L4_C data product.

Shape	Rank	Nominal Product Dimensions	Maximum Product Dimensions
LatCell_LonCell_Array	2	(1624, 3856)	(1624, 3856)

Appendix E: L4_C Units

Table 20 lists the units that are used by the L4_C data product elements. The SMAP implementation of HDF5 stores unit information for each data element in local metadata. The first column in the Table 17 identifies units that apply to data in the L4_C data product. The second column lists the Common Symbol used to represent the unit. The third column lists the matching Label that appears in the local metadata in the L4_C data product.

Table 20: Units in the SMAP L4_C data product.

Unit	Common Symbol	L4_C Label	Typical Use
counts	Counts	counts	number of elements in a set
degrees	degrees	degrees	angular measure
dimensionless	n/a	dimensionless	dimensionless quantity
Gigahertz	GHz	GHz	frequency measure
megabytes	Mbytes	Mbytes	computer storage units
meters	m	m	distance measure
kilometers	km	km	distance measure
percent	%	percent	per hundred
seconds	s	sec	time measure
revolutions per minute	rpm	rpm	rotational measure
degrees Celsius	°C	Degrees Celsius	temperature measure
Kelvin	K	Kelvin	temperature measure
Pascal	Pa	Pa	pressure measure
kilograms per kilogram	$kg\ kg^{-1}$	kg kg-1	mass per mass measure
meters per second	$m\ s^{-1}$	m s-1	velocity measure
meters squared per meters squared	$m^2\ m^{-2}$	m2 m-2	area per area measure
meters cubed per meters cubed	$m^3\ m^{-3}$	m3 m-3	volume per volume measure
grams per meters squared	$g\ m^{-2}$	g m-2	mass storage per area measure

Unit	Common Symbol	L4_C Label	Typical Use
grams per meters squared per day	$g\ m^{-2}\ d^{-1}$	g m-2 d-1	mass flux measure
Watts per meters squared	$W\ m^{-2}$	W m-2	energy flux measure