



**soil moisture**  
cci

# System Prototype Description (SPD)

Version 1.0  
04 July 2013

Prepared by

TU Wien, VUA, GeoVille, ETH Zürich, AWST, FMI, UCC and NILU



System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

This document forms deliverable D3.1 and was compiled for the ESA Climate Change Initiative Phase 1 Soil Moisture Project (ESRIN Contract No: 4000106302 (4000104814/11/I-NB)). For more information on the CCI programme of the European Space Agency (ESA) see <http://www.esa-cci.org/>.

Number of pages: 29

Authors:		D. Chung R. Kidd, W. Dorigo	
Circulation (internal):		Project consortium and science partners	
External:		ESA	
Issue	Date	Details	Editor
0.1	10-02-12	Document created	R. Haas
0.2	25-06-13	Revision of structure and task allocation	R. Kidd
0.3	28-06-13	Draft delivery of document to ESA	R Kidd
0.4	01-07-13	Revision of section 5, 6, 7	R. Kidd
	03-07-13	Revision of section 7	D Chung
	04-07-13	Inclusion of data throughput and resource requirements	D Chung
0.5	04-07-13	Revision of narrative sections to final deliverable	R Kidd
1.0	04-07-13	Delivery to ESA	R Kidd

For any clarifications please contact Richard Kidd (richard.kidd@geo.tuwien.ac.at).

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## Project Partners

<b>Prime Contractor</b>	<b>TU Wien</b> , Vienna University of Technology (Austria)
<b>Scientific Lead</b>	
<b>Project Management</b>	<b>GeoVille</b> , GeoVille Information Systems GmbH (Austria)
<b>System Engineering Partners</b>	<b>AWST</b> , Angewandte Wissenschaft Software und Technologie GmbH (Austria)
<b>Earth Observation Partners</b>	<b>TU Wien</b> , Vienna University of Technology (Austria) <b>VUA</b> , Vrije Universiteit Amsterdam,(The Netherlands) <b>FMI</b> , Finnish Meteorological Institute, (Finland) <b>UCC</b> , University College Cork, (Ireland)
<b>Climate Research Partners</b>	<b>ETH</b> , Institute for Atmospheric and Climate Science, (Switzerland) <b>NILU</b> , Norsk institutt for luftforskning (Norway)

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## Table of Content

<b>TABLE OF CONTENT</b> .....	<b>III</b>
<b>LIST OF FIGURES</b> .....	<b>V</b>
<b>LIST OF TABLES</b> .....	<b>V</b>
<b>ACRONYMS AND ABBREVIATIONS</b> .....	<b>VI</b>
<b>1 EXECUTIVE SUMMARY</b> .....	<b>1</b>
<b>2 INTRODUCTION</b> .....	<b>2</b>
2.1 PURPOSE OF THE DOCUMENT .....	2
2.2 TARGETED AUDIENCE.....	2
<b>3 DOCUMENTS</b> .....	<b>2</b>
3.1 APPLICABLE DOCUMENTS .....	2
3.2 REFERENCE DOCUMENTS .....	3
3.3 BIBLIOGRAPHY .....	3
<b>4 PROTOTYPE OVERVIEW</b> .....	<b>4</b>
4.1 DATA PRODUCTS REQUIRED FOR THE PECVPS.....	6
<b>5 SYSTEM CONCEPT</b> .....	<b>7</b>
5.1 WARP FRAMEWORK .....	7
5.1.1 <i>Standard interfaces within the WARP Framework</i> .....	8
5.1.2 <i>WARP Framework object oriented approach</i> .....	9
5.1.3 <i>Versioning and version control</i> .....	9
<b>6 ARCHITECTURE</b> .....	<b>10</b>
6.1 HARDWARE AND NETWORK INFRASTRUCTURE.....	10
6.2 SOFTWARE ENVIRONMENT .....	10
6.3 EXTERNAL INTERFACES .....	10
6.4 INPUT DATA INGESTION INTERFACE.....	11
6.5 OUTPUT DATA DISSEMINATION INTERFACE .....	11
6.6 INTERNAL DATA STORAGE INTERFACE .....	11
6.7 OPERATOR COMMAND LINE INTERFACE .....	11
6.8 FOLDER STRUCTURE FOR DATA MANAGEMENT .....	11
6.9 PROCESSING ENVIRONMENT .....	12
<b>7 PROCESSING COMPONENTS</b> .....	<b>13</b>
<b>7.1 FIRST PROCESSING STEP 1: RESAMPLING</b> .....	<b>13</b>
<b>7.2 SECOND PROCESSING STEP: RESCALING</b> .....	<b>14</b>

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

7.3	<b>THIRD PROCESSING STEP: MERGING</b> .....	15
7.4	<b>FOURTH PROCESSING STEP: RESCALING</b> .....	15
7.5	<b>FIFTH STEP: MERGING</b> .....	16
<b>8</b>	<b>DATA THROUGHPUT AND RESOURCE REQUIREMENTS</b> .....	<b>17</b>
<b>9</b>	<b>DATA MANAGEMENT AND PRODUCTION CONTROLS</b> .....	<b>19</b>
9.1	DATA TRANSFER .....	19
9.2	BACK UP AND ARCHIVE .....	19
9.3	PRODUCT AND PRODUCTION CONTROLS .....	19
<b>10</b>	<b>BIBLIOGRAPHY</b> .....	<b>20</b>
	<b>ANNEX A PROCESSING LEVELS</b> .....	<b>21</b>

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## List of Figures

Figure 1 The modules and data flow of the Soil Moisture Prototype ECV Production System.	5
Figure 2: Composition of a process.....	7
Figure 3: Datasets distinguished by their formats. ....	9
Figure 4: Hierarchy of the object oriented design. ....	9
Figure 5 WFW Cell and grid based processing .....	12
Figure 6 Modules and data flow of the Soil Moisture Prototype ECV Production System showing their data volumes and data throughputs.....	17

## List of Tables

Table 1 Data product summary for PECVPS showing input data (blue), ancillary data (green) and output product (white) .....	6
Table 2 Processing Machines available to PECVPS .....	12
Table 3 Resource requirements for an example of each process step (based on the processing machine as specified in Table 2) .....	18
Table 4 Processing Level Codes for Remotely Sensed Datasets .....	21

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## Acronyms and abbreviations

AMI-WS	Active Microwave Instrument - Windscat (ERS-1 & 2)
AMSR-E	Advanced Microwave Scanning Radiometer-Earth Observing System
ASCAT	Advanced Scatterometer (Metop)
ATBD	Algorithm Theoretical Basis Document
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
CPU	Central Processing Unit
CRC	Cyclic Redundancy Code
CVS	Concurrent Versioning System (computing)
DFS	Distributed File System
DPM	Data Processing Model
DSWG	Data Standards Working Group (ESA - CCI)
ECV	Essential Climate Variable
FTP	File Transfer Protocol
GLDAS-Noah	Global Land Data Assimilation System- Noah model
GPI	Grid Point Index
I/O	Input Output
IDL	Interactive Data Language (Exelis VIS)
IODD	Input Output Data Description
L2	Retrieved environmental variables at the same resolution and location as the level 1 (EO) source.
L3S	Level 3 Super Collated observations from multiple instruments into a space time grid
LPRM	Land Parameter Retrieval Model
NASA	National Aeronautics and Space Administration
OS	Operation System
PECVPS	Prototype Essential Climate Variable Processing System
PSD	Product Specification Document

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

PVASR	Product Validation and Algorithm Selection Report
SM	Soil Moisture
SMMR	Scanning Multichannel Microwave Radiometer
SOW	Statement of Work
SPD	System Prototype Description
SSM/I	Special Sensor Microwave Imager
TMI	TRMM Microwave Imager
UUID	Universal Unique Identifier



System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## **1 Executive Summary**

This System Prototype Document (SPD) is provided as part of the suite of deliverables that document the implementation of the prototype essential climate variable processing system (PECVPS) for soil moisture (SM). When taken in conjunction with the detailed processing model (DPM) and the input output data description (IODD) it will enable system or software engineers to implement, replicate and further develop the PECVSP when provided with the PECVSP source code as developed by Phase 1 of the CCI SM project.

The SPD presents an overview of the main components of the prototype, the system concept underlying its implementation, and also provides details of the hardware and computer network architecture and configurations as required by the PECVPS. A detailed narrative of the main processing components is provided to assist the system or software engineers implementing the processing system in Phase 2 of the CCI SM project to obtain a clear and precise understanding of the PECVPS source code.

By providing an explicit record of data volumes and data throughputs of the prototype the SPD provides an accurate baseline for Phase 2 implement of the processing system.

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## **2 Introduction**

The System Prototype Description (SPD) provides an overview of the prototype soil moisture ECV production system (PECVPS) (hereafter also denoted ‘the prototype’) detailing its components, their functions, and interfaces. It clearly identifies its internal and external interfaces, data flow between processor components, data throughput, and resource requirements in terms of CPU load for processing demand, processing volumes, and timing of resources.

The SPD is based upon the Input Output Data Description document (IODD) provided as deliverable D2.8, and the Detailed Processing Model (DPM) provided as deliverable D2.7. These two documents in turn have been derived from the Algorithm Theoretical Baseline Document (ATBD). The ATBD identifies and provides the scientific background to the algorithms that are used within the PECVPS. The IODD defines the data items and flow within the PECVPS, defining the inputs and outputs to each of the processing steps, and essentially defines the processor data interfaces. The DPM takes the algorithms as selected in the Product Validation and Algorithm Selection Report (PVASR) and described in the ATBD and provides pseudo code of their implementation.

### **2.1 Purpose of the document**

The System Prototype Description (SPD) has been written to provide a complete high level description of the design of the prototype ECV production system for soil moisture. When the SPD is coupled with the prototype software it will enable system design engineers to understand the complete end to end processing system from a systems analysis perspective.

### **2.2 Targeted audience**

This document targets mainly

- System Design Engineers interested in understanding, or further implementing, the Prototype Essential Climate Variable Production System (PECVPS) for Soil Moisture.
- Scientist of other Essential Climate Variable (ECV) groups interested in the documentation of ECV prototype systems.

## **3 Documents**

### **3.1 Applicable documents**

The documents outlined here detail the scope and focus for the work that is reported in this document.

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

[AD-1] Phase 1 of the ESA Climate Change Initiative Soil- Moisture- cci. ESRIN Contract No: 4000104814/11/I-NB.

[AD-2] ESA Climate Change Initiative Phase 1, Statement of Work for Soil Moisture and Ice Sheets, European Space Agency, EOEP-STRI-EOPS-SW-11-0001.

[AD-3] Technical Proposal (Part 3) in response to ESA Climate Change Initiative Phase 1 ESRIN/AO/1-6782/11/I-NB, Vienna University of Technology.

### 3.2 Reference documents

This section provides information about references documents upon which this document is either based, or is required to be referenced by the reader in order to obtain the full information intended by the authors.

[RD-1] Climate Modelling User Group, Requirement Baseline Document, August 2011, CMUG\_D1.2\_URD\_v1.5.pdf (URL: <http://www.esa-cmug-cci.org/>)

[RD-2] Input/Output Data Definition Document, (IODD v1), Version 1.0, 21 December 2012, <http://www.esa-soilmoisture-cci.org/node/119>

[RD-3] Detailed Processing Model (DPM), Version 1.0, Climate Change Initiative Phase 1 Soil Moisture Project, 28<sup>th</sup> September 2012, <http://www.esa-soilmoisture-cci.org/node/119>

[RD-4] ESA, Guidelines for Data Producers - Climate Change Initiative Phase 1, CCI-PRGM-EOPS-TN-11-0003, Issue 3, Revision 1, Victoria Bennett, 20/07/2012

### 3.3 Bibliography

A complete bibliographic list, detailing scientific text or publications that support arguments or statements made within the current document is provided in section 10.

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## 4 Prototype Overview

The Soil Moisture ECV production system starts from level 1 calibrated backscatter values and level 1 calibrated brightness temperatures for scatterometers and radiometers, respectively. The ECV production system requires a modular system design where new or updated level 2 datasets can be easily ingested, quality controlled, and assimilated in the ECV production. In Phase I of the CCI SM project, only input level 2 (L2)<sup>1</sup> products based on the TU Wien method (for scatterometers) and the VUA-NASA LPRM algorithm (for radiometers) are considered. This results in the following structure of the ECV production system:

- A. Level 2 soil moisture retrieval for individual scatterometer datasets using the TU Wien method as first provided by (Wagner, Lemoine et al. 1999).
- B. Level 2 soil moisture retrieval for single radiometer datasets using the VUA-NASA method (Owe, de Jeu et al. 2008).
- C. Fusion of the active Level 2 datasets into a homogenized active surface soil moisture ECV.
- D. Fusion of the passive Level 2 data into a homogenized passive surface soil moisture ECV.
- E. Fusion of the merged active and passive datasets from steps 2 and 3 into homogenized active+passive surface soil moisture ECV.

The prototype only concerns itself with the process from the level 2 data input data to the generation of the level 3 ECV product (L3S) and therefore only addresses points C,D and E.

Level 2 derived soil moisture products from active and passive space borne systems form the main inputs to the merging algorithm of the ECV prototype production system and are supported by a limited set of ancillary data. The prototype consists of the following modules as shown in Figure 1:

- Processing Step 1: Resampling
- Processing Step 2: Rescaling
- Processing Step 3: Merging
- Processing Step 4 Rescaling
- Processing Step 5: Merging

---

<sup>1</sup> See Table 3 Processing Level Codes for Remotely Sensed Datasets

A detailed view of the 5 main processing steps in the prototype, namely resampling, rescaling, active merging and passive merging, rescaling to GLDAS climatology, and combing active and passive merged data are provided in the DPM [RD-3] and fuller details of data flows are provided in IOOD [RD-2].

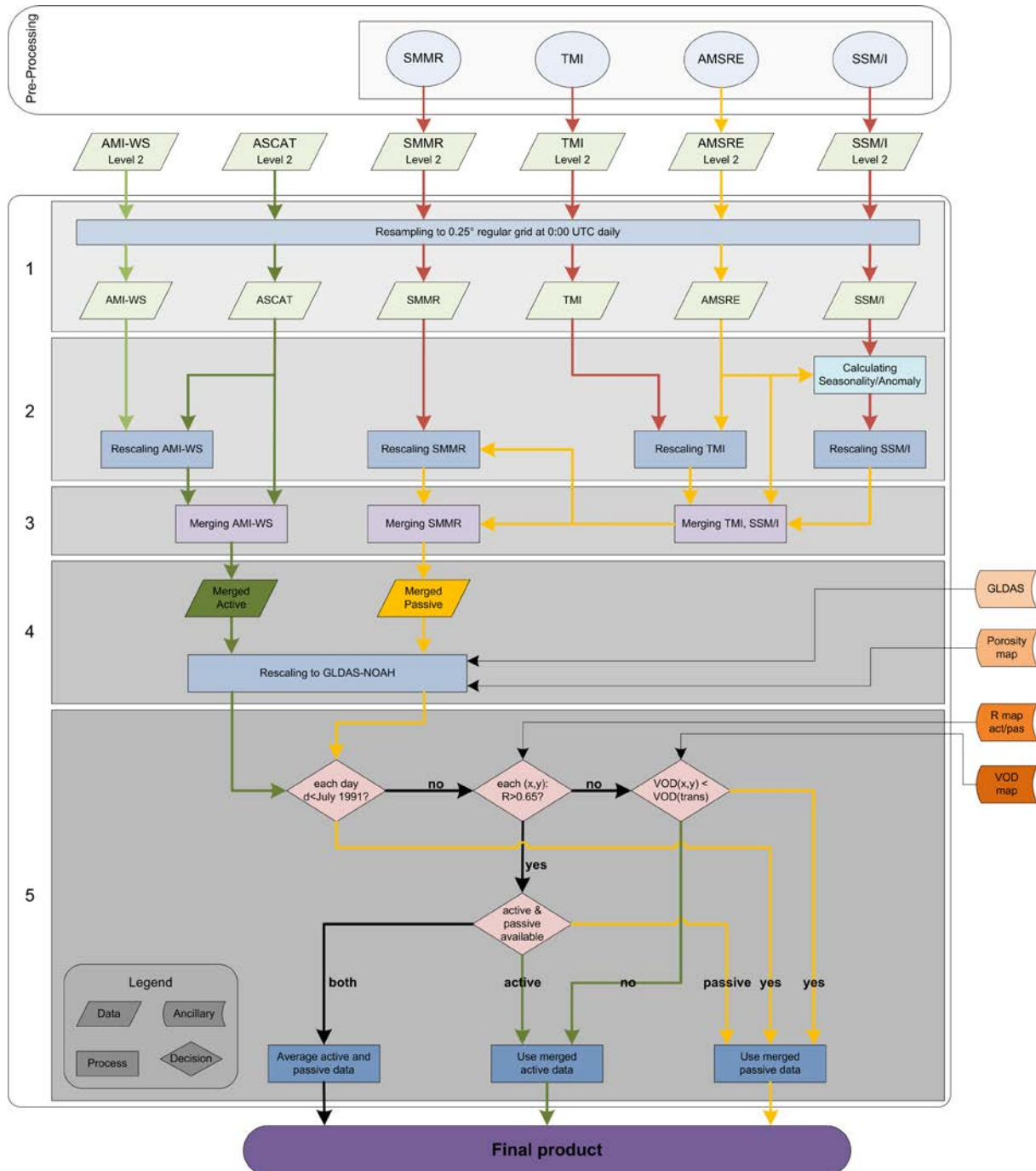


Figure 1 The modules and data flow of the Soil Moisture Prototype ECV Production System

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
------------------------------------	--	---------------------------------

#### 4.1 Data Products Required for the PECVPS

As seen in Figure 1 the prototype requires an number of input and ancillary datasets to drive the process and results in three output products. Information relating to these data products are summarised in Table 1

*Table 1 Data product summary for PECVPS showing input data (blue), ancillary data (green) and output product (white)*

Data Product	Source	Data origin	Period	Product Format	Product Volume
METOP-A ASCAT 25 km Soil Moisture Time Series; Warp 5.5	TU Wien	Satellite	2007-01-01 - 2012-12-31	L2 SSM, binary	57 GB
ERS-1/2 AMI 50 km Soil Moisture Time Series; Warp 5.0	TU Wien	Satellite	1991-08-05 to 2007-05-31	L2 SSM, binary	27 GB
SMMR SSM VUA LPRM v03	VUA Amsterdam	Satellite	1978-10-26 to 1987-08-20	L2 SSM, binary	4 GB
SSM/I VUA LPRM v03	VUA Amsterdam	Satellite	1987-07-08 to 2008-06-30	L2 SSM, binary	2 GB
TMI VUA LPRM v05	NASA NSIDC	Satellite	1997-07-12 to 2013-06-12	L2 SSM, binary	23 GB
AMSR-E VUA LPRM v05	NASA NSIDC	Satellite	2002-06-19 to 2011-10-03	L2 SSM, binary	40 GB
GLDAS-NOAH Land Surface Model (GLDAS-1)	NASA GCMD	Model	2000 to present	SSM, binary	35 GB
Porosity Map	TU Wien	Model / In-situ	static	Soil properties, binary	12 MB
Vegetation Optical Depth Map	VUA Amsterdam	Satellite	1988-01-01 to 2011-10-03	VOD, binary	26 GB
Correlation Coefficient R Map	TU Wien	Model	1978 to 2011	R Map, binary	100 MB
CCI ECV SSM Active Merged Product	TU Wien	Satellite	1991 to 2012	L3s SSM, NetCDF	~50 GB
CCI ECV SSM Passive Merged Product	TU Wien	Satellite	1978 to 2011	L3s SSM, NetCDF	~75 GB
CCI ECV SSM Active & Passive Merged Product	TU Wien	Satellite	1978 to 2012	L3s SSM, NetCDF	~80 GB

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## 5 System Concept

The prototype consists of loosely coupled modules in a networked Windows environment, based upon the flexible integration of many components with standard data formats. The component modules are provided in a single language (IDL) within an in-house standard framework (WARP Framework). Interfaces between modules are largely file and data based.

### 5.1 WARP Framework

The WARP Framework (WFW) represents an object oriented processing environment, which has been originally developed to link the processing steps of TU Wien's soil Water Retrieval Package (WARP). The purpose of WARP is to perform retrieval of surface soil moisture information from scatterometer measurements using the TU Wien change detection algorithm. The first realisation of the change detection concept was based on ERS 1/2 datasets (Wagner, Lemoine et al. 1999)(Wagner et al. 1999a; Wagner et al. 1999b; Wagner et al. 1999c) and later the approach was successfully transferred to METOP-A Advanced Scatterometer (ASCAT) data (Bartalis, Wagner et al. 2007; Naeimi, Scipal et al. 2009). However, the WARP framework is neither dedicated to soil moisture retrieval nor scatterometer measurements, since the general and modular design of the framework gives the possibility to basically deploy and process any dataset.

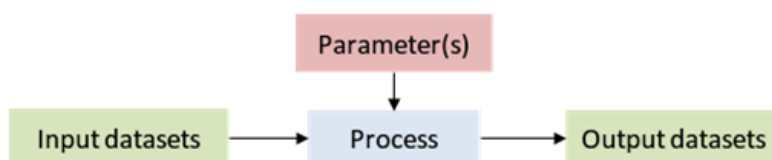


Figure 2: Composition of a process.

The core of the WARP framework basically reflects the natural composition of a process: input, output and parameters (see Figure 2). The definition of the input and output datasets, as well as the parameters has to be specified in a settings file. The importance of the settings file is twofold, on the one hand it represents the interface between the processing framework and the real physical data on the hard drive and on the other hand it ensures a distinct reproducibility of the complete processing setup. Furthermore, several processing steps can be defined in a settings file and the sequence of their execution will be passed as an argument to the WARP framework together with the settings file itself.

In addition, to the definition of processes and datasets a global object can also be created, which contains comprehensive list of parameters and their specified attributes. These global parameters are either relevant to the framework itself or required for the processing steps.

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

### **5.1.1 Standard interfaces within the WARP Framework**

The object oriented design of the WARP framework made it possible to develop standard interfaces between different modules. These are briefly highlighted in the following paragraphs

**Input and output:** The internal generic I/O format is used as a standard interface for reading and writing datasets. The general file structure that is supported starts with an 8-bit fixed field, which contains the structure name of the following generic header. The generic header contains all information to safely read or write a metadata block and data block of the file. The metadata and data block are both optional, however only one record can be hold in the metadata block whereas the data block is allowed to contain an indefinite number of records. Structure templates with distinct names for the metadata and data block are required to be defined beforehand, in order to allow a successful and unambiguous encoding and decoding of the file. However, if an input datasets is not provided in the standard generic I/O format or a dataset needs to be written in the different output format, it is possible to bypass the standard interface by overwriting it with an individual one.

**Settings file:** The settings file represents the standard interface between the WARP framework and the datasets stored on the hard drive. It contains the definition of the input and output datasets, the parameters and the processes. In addition, global ancillary parameters can be specified as well, which are either needed by the framework or by the processes.

**Logging:** The standardised logging mechanism is able to record different types of processing messages. High-level logs, such as start and end time of a process, are written into a global logfile, whereas detailed logs, such as warnings, errors, info and debug information are tracked for each processing step in an individual logfile. Each log contains a time stamp, routine name, user name and computer name.

**Dataset:** Even with the definition of standardised datasets, there must be a distinction concerning the formats of datasets (see Figure 3). Although datasets share common properties, their reading and writing operation are unique and require different formats. For example, datasets in time series format are always composed of two files: an index and a data file. Input and output datasets share similar properties and are treated in the same way, rather than being explicitly separated by their definition. This enables a dataset to be defined as an as output in the first processing step and as input in the second processing step.



System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
------------------------------------	--	---------------------------------

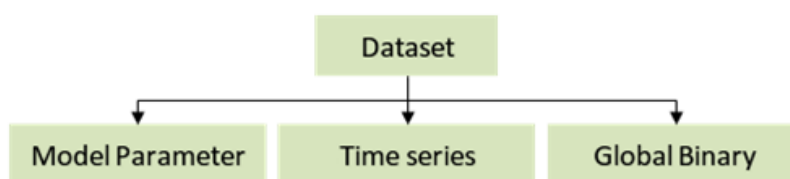


Figure 3: Datasets distinguished by their formats.

### 5.1.2 WARP Framework object oriented approach

Using an object oriented approach each individual process inherits the predefined methods and attributes of the standardised parent class. This is important in order to treat all processes in the same way within the WARP framework. The arguments of a process are the input datasets, output datasets and parameters (see Figure 4). The real processing takes place in a user-defined method, which has to be named “*calc*”. The interface of this method is standardised, however it is not compulsory to use all the parameters from the interface.

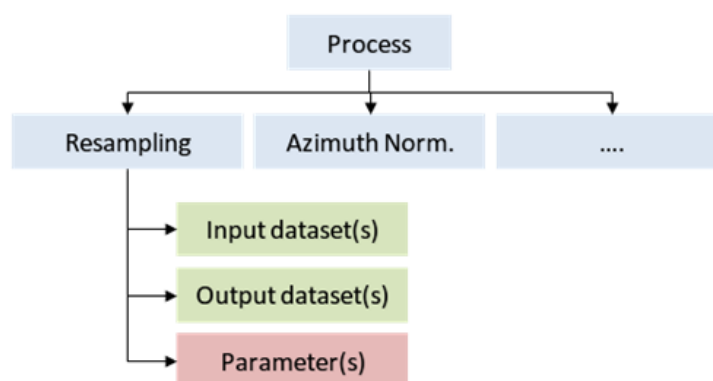


Figure 4: Hierarchy of the object oriented design.

### 5.1.3 Versioning and version control

To ensure that all code and code modifications can be easily traced across multiple development environments code versioning software, Concurrent Versions System (CVS), is rigorously employed in the development of all software and applications within the WARP Framework. This includes the PECVPS.

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## 6 Architecture

Although the prototype is implemented within a research environment, the host organisation, TU Wien, have more than 10 years experience in handling large EO datasets and establishing quasi operational near real time processing chains for the delivery of products. As such the prototype is implemented within robust networked environment hosting high end computers.

### 6.1 Hardware and Network Infrastructure

The hardware environment in which the PECVPS is deployed consists of two virtual networks, a DMZ and an internal Gigabit network, with two separate firewalls comprising numerous computers and associated networking installations, and backup archive capacities.

The internal Network runs on a Windows Active Domain and the hardware elements that are relevant for the PECVPS entail:

- 2 Domain Controllers
- 12 Terminal Servers, 10 of which are dedicated for fast data processing (processing machines). Seven processing machines are available to PECVPS.
- 600 TByte Fileserver
- Developer Workstations
- Backup Server

To ensure efficient access to the 600 TByte fileserver the files system of the fileserver is enabled via a distributed file system (DFS).

### 6.2 Software Environment

The development environment and processing environment run on different platforms and under different operating systems (OS). The development environment runs under Windows 7 environment with software being developed in an IDL development environment version 8.2.3. The processing environment runs under Windows Server 2003, or Windows Server 2008 (R2), x64, and within this environment processing is deployed using IDL run time virtual machines.

### 6.3 External interfaces

The PECVPS relies upon two types of external interface to enable the following functions:

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

- Data acquisition from Data Providers (human interfaces for manual transfers of L2 Data and Ancillary Data; software interfaces for automated transfers of L2 Data)
- Access to ECV Data Products by Data Users (FTP download).

Internal interfaces are handled within the WARP Framework as described in section 5.1.

#### **6.4 Input data ingestion interface**

There is no custom interface for data ingestion into the PECVPS, rather all L2 data, (after pre-processing, as shown in Figure 1), is made available to the PECVPS via manual transfer to the fileserver.

#### **6.5 Output data dissemination interface**

There is no custom interface for enable data dissemination of the products generated by the PECVPS rather all products (L3S) are manually transferred to a dedicated FTP site for authorised user access.

#### **6.6 Internal data storage interface**

The interface for internal data storage is handled within the WFW, with data storage location defaulting to the local processing environment. Data storage location can be overridden by correct parameterisation of the processor settings file.

#### **6.7 Operator command line interface**

The operator initiates processes within the PECVPS by creating and issuing command batch files as IDL virtual machines, with processes being configured by correct parameterisation of the process settings file.

The operator maintains manual control of the processes using standard OS administrative tools.

The operator is informed of process status by monitoring of log files. Optionally with WFW the operator can be automatically notified by e-mail of process completion or error status.

#### **6.8 Folder structure for data management**

The WFW supports a cell based grid processing system in which the area of interest is divided into indexed cells of defined spatial extent on a fixed earth grid. Each cell is normally 5 by 5 decimal degrees and would contain a number of clearly identified and described grid points. Each grid point is identified by a grid point index (GPI). This cell based structure

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
------------------------------------	--	---------------------------------

approach has been proven to have a better file I/O performance. The cell and grid point processing system is outlined in Figure 5, and the file structure for the cell based processing approach is defined and created within the WFW.

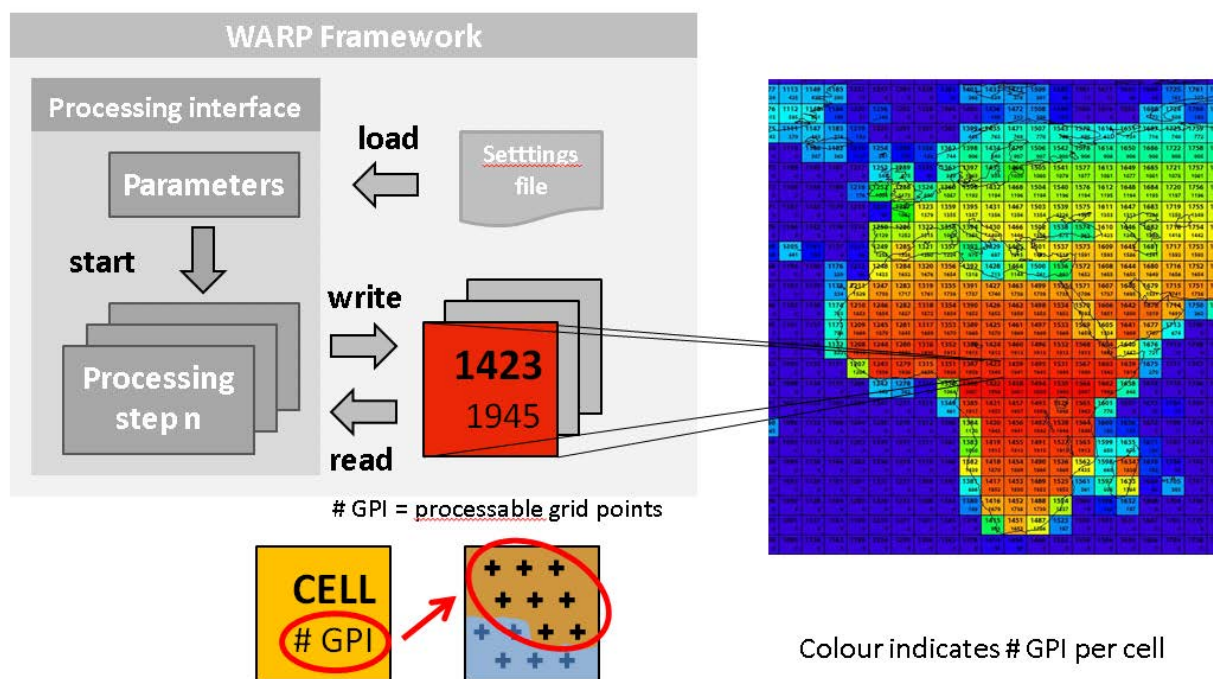


Figure 5 WFW Cell and grid based processing

## 6.9 Processing Environment

The specifications of the processing machines dedicated to the PECVPS processes are provided in Table 2. Typical processing times for specific processes are provided in section 8

Table 2 Processing Machines available to PECVPS

Machine	Quantity	Operating System	Processor	RAM (GB)	Disk (TB)
A	2	Windows Server 2008 R2, x64	Opteron 2 x 2.6 GHz Quad Core	32	4
B	1	Windows Server 2008 R2, x64	Opteron 2 x 2.6 GHz Quad Core	32	22
C	1	Windows Server 2008 R2, x64	Opteron 2 x 2.4 GHz, Octo Core, Multi-Threaded	32	1
D	1	Windows Server 2003, x64	Xeon 2 x 3.0 GHz Dual Core	20	1
E	1	Windows Server 2003, x64 Enterprise	Xeon 2 x 2.66 GHz Quad Core	16	2
F	1	Windows Server 2008 R2, x64	Opteron 2 x 2.2 GHz Dual Core	8	2

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## 7 Processing Components

The prototype consists of 5 distinct processing steps. Whilst each process is clearly defined in the DPM [RD-3] the following subsection provides a supporting narrative outlining the purpose for, providing a description of, and details the input and output products for each process.

### 7.1 First Processing Step 1: Resampling

#### **Purpose: Data preparation**

The homogenised and merged product provides surface soil moisture with a global coverage and a spatial resolution of 0.25° (decimal degree). The time period spans the entire period covered by the individual sensors, i.e. 1978 – present, while the temporal resolution is 1 day. All satellite soil moisture datasets will be collected and transformed into a harmonized format. Also, satellite datasets originating other than from TU Wien need to be converted into a common file format to make them readable by the resampling processor (pre-processing). This pre-processing step is shown in Figure 1 as being outside the environment of the prototype.

#### **Processing Description:**

Data are read from the various archives in their original formats and transformed into time series format based on the original grid (one time series per grid point) (pre-processing). After pre-pre-processing all input datasets are ready to be resampled temporally and spatially.

The temporal resolution of the merged product is one day. The reference time for the merged dataset set at 0:00 UTC. For each day, the observations within the reference time of 0:00 UTC  $\pm$ 12 hours are considered. If more than one observation falls within this period, the observation closest in time is selected (temporal resampling).

Nearest neighbour resampling is performed to convert the various grid systems into the common regular grid. For each grid point in the reference (regular grid) dataset, this resampling technique receives the value of the closest grid point in the input dataset (spatial resampling).

The regular grid points are saved in a structured cell based format. This structure has been proven to have a better file I/O performance, as the 0.25 degree regular grid is now grouped into 5°x5° cells. So instead of writing one file for each grid point, now the maximum of 400 grid points are saved in one file per cell.

**Input:** SMMR, SSM/I, TMI, AMSR-E, AMI-WS, ASCAT

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

**Output:** resampled to 0.25 degree regular grid: SMMR, SSM/I, TMI, AMSR-E AMI-WS, ASCAT

## 7.2 Second Processing Step: Rescaling

### **Purpose: Bringing datasets into a common reference by CDF matching**

Due to different observation frequencies, observation principles, and retrieval techniques, the contributing soil moisture datasets are available in different observation spaces. Therefore, before merging can take place at either level, the datasets need to be rescaled into a common climatology.

### **Processing Description:**

This rescaling procedure is mainly divided into two parts:

1. The rescaling of the resampled active microwave soil moisture observation to the climatology of the ASCAT dataset.
2. The rescaling of the resampled passive microwave soil moisture observation to the climatology of AMSR-E

Ad 1)

The AMI-WS dataset is rescaled by CDF matching using the ASCAT dataset as reference climatology.

Ad 2)

Before rescaling the passive SSM/I against the AMSR-E dataset, both datasets needs to be decomposed into their own seasonality and anomalies. Then the “anomalies of SSM/I” are rescaled against “anomalies of AMSR-E”. To obtain the reconstructed SSM/I dataset (rescaled), the “rescaled SSM/I anomalies” is added to the AMSR-E seasonality.

The TMI dataset is rescaled against the AMSR-E dataset based on their overlapping period.

The rescaled TMI, the reconstructed SSM/I, and the AMSR-E datasets are then merged. This merged passive dataset is used as reference for rescaling the SMMR dataset.

All rescaling procedures are performed using the piece-wise linear CDF matching technique. The [0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 100] percentiles are chosen for the CDF matching routine. Additionally the scaling parameter, slope and intercept, are saved separately.

**Input:** AMI-WS, ASCAT, SMMR, SSM/I, TMI, AMSR-E; (all resampled)

**Output:** rescaled: AMI-WS, SMMR, SSM/I, TMI

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

### 7.3 Third Processing Step: Merging

#### **Purpose: Blending active datasets and blending passive datasets**

This processing stage will produce the merged active and the merged passive soil moisture product.

#### **Processing Description:**

For the common time period between the active and passive datasets (2007-01-01 to 2007-05-31) when merging the rescaled ASCAT with the rescaled AMI-WS dataset, the following approach is applied: AMI-WS data is used to fill ASCAT data gaps if available.

The same approach is also applied when merging the passive datasets, where due to atmospheric effects SSM/I data is least preferred and ASMR-E data most preferred, if there is any overlapping time period when merging SSMR, SSM/I, TMI, and ASMR-E data.

**Input:** AMI-WS, ASCAT (rescaled active); SMMR, SSM/I, TMI, ASCAT (rescaled passive)

**Output:** merged active soil moisture product, merged passive soil moisture product

### 7.4 Fourth Processing Step: Rescaling

#### **Purpose: Bringing datasets into a common reference by CDF matching**

The merged passive and active microwave products obtained in the previous sections represent volumetric soil moisture ( $\text{m}^3 \text{m}^{-3}$ ) and degree of saturation (%), respectively. To combine these data, both the merged microwave products need to be adjusted to a common reference. The GLDAS-Noah dataset is employed as the reference dataset. Both merged passive and active microwave products are rescaled against GLDAS-Noah using the CDF matching technique.

#### **Processing Description:**

Apart from the GLDAS-Noah reference dataset, the same rescaling routines can be used to rescale the merged active and the merged passive microwave products. As before, the same set of percentiles [0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 100] is applied to the CDF matching procedure. The values from the porosity map are used as thresholds when after rescaling the soil moisture value becomes higher than the porosity or negative. Also, the scaling parameter, slope and intercept, are saved separately.

**Input:** Merged active microwave product, merged passive microwave product.

**Output:** Rescaled merged active and rescaled merged passive microwave product

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## 7.5 Fifth Step: Merging

### **Purpose: Generating the final merged microwave product**

In this final merging step the correlation coefficient R map, and the vegetation optical depth map are needed for the data selection between active and passive microwave data when merging these two products.

### **Processing Description:**

The R map shows the correlation between the merged active and the merged passive microwave product. For regions where R is higher than 0.65 (“transitional regions”) active, passive, or – when both are available – active and passive data (average) are used. Outside of the “transitional regions”, the merged passive microwave product is used over the regions with sparse vegetation density, and the merged active microwave product is used over the regions with moderate vegetation density. The annual average vegetation optical depth (VOD) during 2002 – 2008 derived from AMSR-E is used to represent vegetation density. The averaged VOD values over the “transitional regions” are, respectively, 0.38, 0.41, 0.44 and 0.43 for July 1991 – 1997, 1998–June 2002, July 2002 – 2006 and 2007 – 2012. Outside of the “transitional regions”, the merged passive microwave product is used over the regions with lower VOD values than the averaged VOD values of the “transitional regions”; otherwise the merged active microwave product is used.

For the time period before July 1991 SMMR and SSM/I data are used as shown in Figure 1.

**Input:** active microwave and passive microwave product

**Output:** final merged product



### 8 Data throughput and resource requirements

As Figure 1 already clearly shows the data flow within the prototype it is straight forward to annotate it to show data volumes and the throughput of these datasets at each processing stage. Data volumes and throughput are provided in Figure 6.

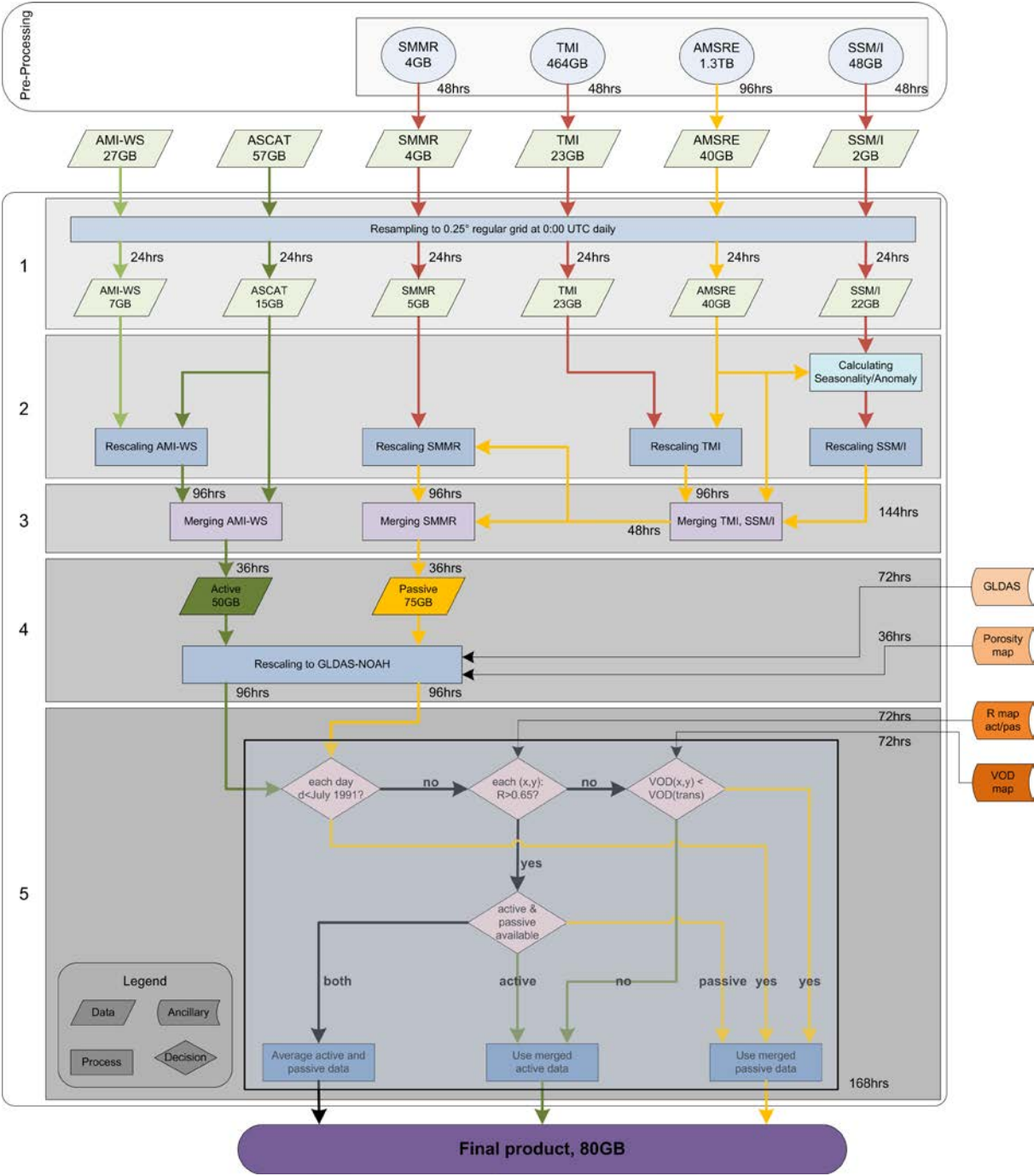


Figure 6 Modules and data flow of the Soil Moisture Prototype ECV Production System showing their data volumes and data throughputs

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

A simple calculation can be performed by 1548 hours (64 days) of processing to generate the global ECV product over the period 1978 to 2010.

Similarly baseline processing figure can be determined when considering various processing steps within the prototype. Table 3 provides example figures for process examples from each of the main the processing steps within the prototype.

*Table 3 Resource requirements for an example of each process step (based on the processing machine as specified in Table 2)*

Process	Machine	CPU Load	Volume	Time
Pre Processing: AMSRE	A, A, B	100%	1.3TB	96hrs
Resampling: AMSRE	A	30%	40GB	24hrs
Rescaling: TMI to AMSRE	A	40%	23GB	96hrs
Merging: AMI-WS / ASCAT	A	30%	50GB	36hrs
Rescaling: ACTIVE to GLDAS	A	40%	50GB	96hrs
Merging: Active / Passive	A	30%	80GB	168hrs

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## **9 Data management and production controls**

Data management is manually controlled by the prototype operator. The following sections detail some simple techniques that are routinely used.

### **9.1 Data Transfer**

When performing data transfer from systems external to the prototype the completeness of data transfer of input files (L2) is ensured by control of Cyclic Redundancy Code (CRC checksum code), when provided by data provider.

### **9.2 Back up and archive**

Critical PECVPS data and products are manually archived by copying and storing of these datasets in a separate location in DFS.

Code backup is facilitated by archive backup of the software repository

### **9.3 Product and Production controls**

Within the prototype there production control is ensured at each step of the process by the generation and inspection of log files for reported errors or anomalies. Further manual visual controls are undertaken at each relevant processing step for an inspection of product completeness and continuity of data.

A product version control is in place with each product being identified by a Universal Unique Identifier (UUID) for each product as specified by CCI DSWG [RD-4]

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
---------------------------------------	--	---------------------------------

## 10 Bibliography

Bartalis, Z., W. Wagner, et al. (2007). "Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT)." Geophys. Res. Lett 34: L20401.

Naeimi, V., K. Scipal, et al. (2009). "An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations." IEEE Transaction on Geoscience and Remote Sensing 47(7): 1999-2013.

Owe, M., R. de Jeu, et al. (2008). "Multisensor historical climatology of satellite-derived global land surface moisture." Journal of Geophysical Research-Earth Surface 113(F1): F01002.

Wagner, W., G. Lemoine, et al. (1999). "A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data." Remote Sensing of Environment 70: 191-207.

System Prototype Description (SPD)	ESA Climate Change Initiative Phase I Soil Moisture	Version 1.0 Date 4 July 2013
------------------------------------	--	---------------------------------

## ANNEX A *Processing Levels*<sup>2</sup>

Table 4 Processing Level Codes for Remotely Sensed Datasets

Level	<ProcessingLevel> Code	Description	Based on Source
Level 0	L0	Unprocessed instrument and payload data at full resolution. GHRSSST	
Level 1A	L1A	Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters, computed and appended, but not applied, to L0 data.	GHRSSST
Level 1B	L1B	Level 1A data that have been processed to sensor units.	GHRSSST
Level 1C	L1C	Level 1B data that have been further processed, eg by correcting radiances or by mapping onto A spatial grid, prior to deriving geophysical variables from the data.	SMOS data products definition
Level 2	L2	Retrieved environmental variables at the same resolution and location as the level 1 source.	CEOS handbook <sup>3</sup>
Level 2Pre-processed	L2P	Geophysical variables derived from Level 1 source data at the same resolution and location as the level 1 data, typically in a satellite projection with geographic information. These data form the fundamental basis for higher-level CCI products.	GHRSSST
Level 3	L3	Level 2 variables mapped on a defined grid with reduced requirements for ancillary data. Three types of L3 products are defined:	GHRSSST
	L3U	Uncollated (L3U): L2 data granules remapped to a space grid without combining any observations from overlapping orbits.	
	L3C	Collated (L3C): Observations combined from a single instrument into a space-time grid.	
	L3S	Super-collated (L3S): observations combined from multiple instruments into a space-time grid.	
Level 4	L4	Datasets created from the analysis of lower level data that result in gridded, gap-free products.	GHRSSST

<sup>2</sup> Extracted from [RD-4]

<sup>3</sup> [http://wiki.ieee-earth.org/@api/deki/files/7/=Handbook\\_0802.pdf](http://wiki.ieee-earth.org/@api/deki/files/7/=Handbook_0802.pdf)