



permafrost
cci

**CCI+ PHASE 1 – NEW ECVS
PERMAFROST**

D3.3 SYSTEM VERIFICATION REPORT (SVR)

VERSION 2.0

31 MAY 2020

PREPARED BY

b·geos



GAMMA REMOTE SENSING



UiO : University of Oslo



**UNI
FR**
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG



UN West University
of Timisoara

TERRASIGNA™

Document Status Sheet

Issue	Date	Details	Authors
0.1	25.04.2019	Initial version	A. Wiesmann, T. Strozzi (GAMMA), A. Bartsch (b.geos)
0.2	24.05.2019	Revision 1	S. Westermann (UIO)
1.0	31.05.2019	Ready for submission	A. Wiesmann (GAMMA)
2.0	31.05.2020	Updates for year 2	S. Westermann (UIO), A. Wiesmann (GAMMA) A. Bartsch (b.geos)

Author team

Andreas Wiesmann

Annett Bartsch, B.GEOS

Sebastian Westermann, UIO

Tazio Strozzi, GAMMA

ESA Technical Officer:

Frank Martin Seifert

EUROPEAN SPACE AGENCY CONTRACT REPORT

The work described in this report was done under ESA contract. Responsibility for the contents resides in the authors or organizations that prepared it.

TABLE OF CONTENTS

Executive summary.....	4
1. Introduction	5
1.1 Purpose of the document	5
1.2 Structure of the document	5
1.2 Document Status	6
1.3 Applicable documents	6
2. Verification Methodology	9
3. Qualification - Description of the test environment.....	10
4. Test Protocol	11
4.1 Objectives and Scope	11
4.2 Components Tested.....	11
4.3 Input Data and Reference Output Data	11
4.4 Test Procedure and requirements for successful testing	11
4.5 Implications for System Engineering.....	12
5. Conclusions	13

EXECUTIVE SUMMARY

Within the European Space Agency (ESA), the Climate Change Initiative (CCI) is a global monitoring program which aims to provide long-term satellite-based products to serve the climate modelling and climate user community. Permafrost has been selected as one of the Essential Climate Variables (ECVs) which are elaborated during Phase 1 of CCI+ (2018-2021).

This document outlines the system verification procedures and results for the Permafrost_cci Processing System.

1. INTRODUCTION

The European Space Agency (ESA) Climate Change Initiative aims to generate high quality Essential Climate Variables (ECVs) derived from long-term satellite data records to meet the needs of climate research and monitoring activities, including the detection of variability and trends, climate modelling, and aspects of hydrology and meteorology.

1.1 Purpose of the document

The system verification report (Deliverable 3.2, SVR) should confirm that the system as outlined in the SPD [AD 3] and described in more detail in the DPM [RD 5] and IODD [RD 4] is properly working when executed in other hardware and/or software environments. According to the SoW, the SVR “gives a complete report of all activities executed and the results achieved form a technical assessment of the end-to-end prototype system with all its subcomponents to verify that the prototype is compliant to the requirements outlined in the Product Specifications Document (PSD, [RD 6]) and technical specifications (IODD, DPM) and that it fulfils its intended purpose and replicates the results of the algorithms selected through the round-robin.”

Specifically, the SVR should include for each of the generated products:

- a description of the objectives and scope of the processor
- a list of all elements and components of the prototype that have been tested including a description of the platform, the network, and the interfaces with other systems
- a description of all test activities carried out and of the criteria on how the prototype was tested to ensure that the requirements are fulfilled and that the system performs as specified
- a summary description of all test cases, test procedures, and test data used
- a record of all test results
- a description of all acceptable and stated limitations in the prototype system and the steps taken to workaround anomalous, inappropriate, or undesired operating conditions.

Due to the required interaction with an analyst or operator for many of the products, the tests have been performed at the level of modules (e.g. glacier area, velocity, DEM differencing) by the Systems Engineers and EO team. It was ensured that the test operator was not involved in the development or implementation of the tested module.

For all modules the objectives of the processor and the components tested, the performed test activities and the achieved results are documented in this report. The tests are completed by conclusions relevant for the system engineers towards system development and sustainability.

1.2 Structure of the document

This document is organised as follows:

- Section 1 gives an introduction on the deliverable, purpose, applicable and reference documents, its structure and release information.
- Section 2 discusses the verification methodology background.

- Sections 3 – 5 present the results of the module tests.

1.2 Document Status

This document is based on issue 1.0 of the Data Access Requirements Document (DARD), Issue 2.0 of the Product Specification Document (PSD), and Issue 1.0 of the User Requirements Document (URD); refinement of this document will be necessary of catchment of future issues of these documents.

1.3 Applicable documents

[AD-1] ESA 2017: Climate Change Initiative Extension (CCI+) Phase 1 – New Essential Climate Variables - Statement of Work. ESA-CCI-PRGM-EOPS-PF-17-0032

[AD-2] Requirements for monitoring of permafrost in polar regions - A community white paper in response to the WMO Polar Space Task Group (PSTG), Version 4, 2014-10-09. Austrian Polar Research Institute, Vienna, Austria, 20 pp

[AD-3] ECV 9 Permafrost: assessment report on available methodological standards and guides, 1 Nov 2009, GTOS-62

[AD-4] GCOS-200, the Global Observing System for Climate: Implementation Needs (2016 GCOS Implementation Plan, 2015.

1.4 Reference Documents

Table 1: Reference Documents.

REF	TITLE	VERSION
[R-1]	CCI+ SoW	Issue 1.4 r2 EOP-SEP/SOW/0031-1.4 r2
[R-2]	User Requirements Document (URD)	1.1
[R-3]	Product Specification Document (PSD)	2.0
[R-4]	Data Access Requirements Document (DARD)	1.0
[R-5]	Data Standards Guidelines (DSWG)	2.2
[R-6]	System Requirements Document (SRD)	2.0
[R-7]	System Specification Document (SSD)	2.0

1.6 Acronyms

ACOP	Asian Conference on Permafrost
ALT	Active Layer Thickness
Arctic CORDEX	Coordinated Regional Climate Downscaling Experiment
AWI	Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research
B.GEOS	b.geos GmbH
CALM	Circumpolar Active Layer Monitoring
CliC	Climate and Cryosphere project
CLM4	Land Community Model
CCI	Climate Change Initiative
CMIP-6	The Coupled Model Intercomparison Project
CMUG	Climate Modelling User Group
CRESCENDO	Coordinated Research in Earth Systems and Climate: Experiments, Knowledge, Dissemination and Outreach
CRG	Climate Research Group
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
ESA DUE	ESA Data User Element
GAMMA	Gamma Remote Sensing AG
GCOS	Global Climate Observing System
GCW	Global Cryosphere Watch
GT	Ground Temperature
GTN-P	Global Terrestrial Network for Permafrost
GTOS	Global Terrestrial Observing System
GUIO	Department of Geosciences University of Oslo
HIRHAM	High Resolution Limited Area Model
IASC	International Arctic Science Committee
ILAMB	International Land Model Benchmarking
IPA	International Permafrost Association
IPCC	Intergovernmental Panel on Climate Change
LS3MIP	Land Surface, Snow and Soil Moisture
MAGT	Mean Annual Ground Temperature
NetCDF	Network Common Data Format
NSIDC	National Snow and Ice Data Center
PCN	Permafrost Carbon Network
PE	Permafrost Extent
PERMOS	Swiss Permafrost Monitoring Network
PF	Permafrost
PSTG	Polar Space Task Group
RASM	Regional Arctic System Model
RD	Reference Document
RMSE	Root Mean Square Error

RS	Remote Sensing
SAR	Synthetic Aperture Radar
SCAR	Scientific Committee on Antarctic Research
SU	Department of Physical Geography Stockholm University
TSP	Thermal State of Permafrost
UNIFR	Department of Geosciences University of Fribourg
URD	Users Requirement Document
WCRP	World Climate Research Program
WMO	World Meteorological Organisation
WMO OSCAR	Observing Systems Capability Analysis and Review Tool
WUT	West University of Timisoara
ZAMG	Zentralanstalt für Meteorologie und Geodynamik

2. VERIFICATION METHODOLOGY

The system needs to be verified to ensure system integrity after software updates or installation on a new platform. It is not the purpose of the system verification process to validate the product scientifically but to ensure reproducibility of a defined process, processing system.

The verification processes should distinguish between testing after a system upgrade and the installation on a new platform. While after a software upgrade the results may differ if e.g. a classification algorithm was changed, no or minor differences are to be expected if the PS is installed on a new system. In any case deviations have to be understood and rectified.

In general a benchmark test scenario is defined for the system verification. Such a test scenario covers:

- Hardware requirements
- Software (availability) requirements
- Input Data
- Benchmark Data
- Scenario Process
- Other Resources

The Scenario Process description describes the processing steps and pass/fail tests to be conducted. The processing steps, if based on multiple executables, are best complemented by a script that conducts the different processing steps in an automated fashion to minimize operator errors.

All data (input, intermediate, output, benchmark) must be checked for integrity and consistency. Tests to be done are:

- Availability
- Integrity
- Format
- Content

The use of hash values derived from a hash function is the preferred method to confirm the content of a dataset with respect to a reference dataset. If the data are consistent the hash value is identical. The preferred hash function is based on the Message-Digest Algorithm 5 (MD5) a widely used and implemented cryptographic algorithm. It computes a 128-bit hash value of any dataset.

For some products where deviations on the byte level have to be expected (e.g. meta data holding processing dates), special tools may be necessary to only compare or hash the data part that is not affected by dynamic content. Fall-back strategies are value thresholds and visual inspection. The latter might be the only useful method after software upgrades affecting the product algorithm.

Other tests in the scenario address the processing environment. Tests need to cover the expected processing/production time, disk space as well as memory space usage. The results are OS and hardware dependent and will usually be checked against a threshold.

3. QUALIFICATION - DESCRIPTION OF THE TEST ENVIRONMENT

Processing of all permafrost ECV products in the Permafrost_cci (ground temperature, active layer thickness, permafrost fraction) is conducted with a single processing chain, with the permafrost simulation tool CryoGrid CCI in its core. For testing purposes, a parallel computing infrastructure with at least 16 CPU cores is required, which is the minimal requirement to run the regular processing system for smaller spatial domains. With this, it is possible to verify the two key aspects of the processing system: 1. the performance on global scale across representative climatic and environmental gradients; and 2. seamless mapping over larger areas.

In year 1 of Permafrost_cci, the processing has been performed on the Abel supercomputing cluster in Oslo, Norway. Following the discontinuation of Abel in December 2019, the year 2 processing was moved to the new SAGA HPC cluster situated in Trondheim, Norway. System verification according to the test protocol described in Sect. 4 was successfully performed, before the processing chain was adapted and supplemented for year 2 processing. For year 3, the processing chain is expected to remain on SAGA which has a life expectancy until at least the year 2023.

4. TEST PROTOCOL

The Permafrost_cci processing chain consists of a single sequence of modules from which all ECV products are computed. For this reason, we can outline a verification procedure that is capable of testing the performance with respect to all Permafrost_cci products.

4.1 Objectives and Scope

The three principal steps of the Permafrost_cci processing chain are

1. Preprocessing of satellite data and computation of eight-day averages of input data;
2. Simulation of ground temperature profiles with the CryoGrid CCI model;
3. Postprocessing of Permafrost ECV products from CryoGrid CCI output.

To limit the amount of intermediate output written to files, the second and third steps are performed together, which strongly decreases runtime and memory use. In the following, we refer to step 1 as “component 1”, while steps 2 and 3 are denoted “component 2”.

The verification procedure, in simple terms, makes use of sets of defined pixels/ regions of interest for which a reference output is provided, to which the test simulations can be compared. The goal of the verification is to ensure that the target performance of the processing system is reached, e.g. following a system crash or when porting the processing chain to a new HPC environment.

4.2 Components Tested

Due to the structure of the processing chain (see 4.1), verification is performed for two components, distinguishing the preprocessing (step 1) and the simulation/postprocessing components (steps 2 and 3). Both components will be advanced in year 3, so that also the verification procedures will be advanced alongside. However, it is in principle possible to advance/modify only one of the components, in which case the verification procedure of the other component can remain unchanged.

4.3 Input Data and Reference Output Data

As input data for the first component (preprocessing), the entire input of the processing chain as defined in [R-7], is required. For the second component (CryoGrid CCI simulation and postprocessing), the year 2 output of the preprocessing is employed as input.

As reference output data, the year 2 output of the two components is employed. Two spatial subsets are distinguished: first, the subset of globally distributed pixels corresponding to the validation sites (as defined in deliverable 2.1: Product validation) is employed. Second a coherent area is employed for verification of mapping, which should contain a gradient from non-permafrost to continuous permafrost, coastal areas, as well as mountain permafrost. We propose Scandinavia as verification region fulfilling all requirements, which can be processed in less than 24 hours on 16 CPU cores.

4.4 Test Procedure and requirements for successful testing

For component 1 (preprocessing) verification, all input data sets must be available globally. This is required so that the input data for the globally distributed subsets of the validation sites (Sect. 4.4) are available. For this purpose, a dedicated verification/validation script is called which applies preprocessing only for the subset of validation sites. The result of this script is sets of input data for component 2 (CryoGrid CCI simulation and postprocessing) which contains exclusively contains

pixels at the validation sites (Sect. 4.3). For testing of a mapping a coherent area (Sect. 4.4), only the input data sets for the region or interest are required and the regular processing chain is run, followed by comparison to the reference output data (Sect. 4.3).

For component 2 (CryoGrid CCI simulation and postprocessing) verification, it is not necessary to distinguish between the two different verification subsets of pixels. Based on available input data from component 1 of the regular processing, steps 2 and 3 of the regular processing chain (corresponding to the second component, see Sect. 4.1) are run, followed by comparison to the reference output data (Sect. 4.3).

For both components, we propose the following thresholds for successful testing, which will be adapted in year 3: temperatures (0.1K); lengths/depths (0.05m); permafrost fraction (one over number of realizations per pixel, i.e. 0.2 (unitless) with 5 model realizations in year 1. The verification is considered “passed” if all output files are available, they are in a readable format and the output data produced by the test are within the required accuracy of the reference data sets (Sect. 4.3).

4.5 Implications for System Engineering

As the verification is performed on the same system as the main processing, the same requirements as for processing must be met by the verification system. Therefore, system verification does not impose additional requirements on the processing system.

5. CONCLUSIONS

The processing within the Permafrost_cci project is done within the existing supercomputing infrastructure in Norway, which is managed nationally by the company UNINETT Sigma2 AS (<https://www.sigma2.no/>). Storage of input satellite data sets is currently realized on the Norwegian NIRD system (<https://documentation.sigma2.no/storage/nird.html>) which is also administered through UNINETT Sigma2. The verification and product generation of Year 1 was successfully accomplished on Intel E5-2670 (Sandy Bridge) processors at 2.6 GHz, yielding 16 physical compute cores per node. Each node has 64 GB of Samsung DDR3 memory operating at 1600 MHz, giving 4 GiB memory per physical core at about 58 GiB/s aggregated bandwidth using all physical cores. In year 2, the standard compute nodes of SAGA were employed, which feature 40 cores and 192 GiB memory each (about 4GB per core). During year 2 processing, up to 300 cores were employed simultaneously, depending on availability within the SLURM scheduling system, yielding a total processing time of about one month (ca. 150k CPU hours used in total).