CCI+ PHASE 1 – NEW ECVS
PERMAFROST

CCN1 & CCN2
ROCK GLACIER KINEMATICS AS NEW ASSOCIATED PARAMETER OF ECV PERMAFROST

D3. System Development Document

VERSION 1.0

30 NOVEMBER 2020

PREPARED BY

Document Status Sheet

Issue	Date	Details	Authors
0.1	30.09.2019	Final version CNN1 - D3	T. Strozzi, A. Wiesmann, V. Poncos, A. Onaca, F. Sîrbu and A. Bartch
0.2	15.10.2020	Template created for CCN1&2 - D3	T. Strozzi
0.3	16.11.2020	Initial version	A. Wiesmann and R. Caduff
1.0	30.11.2020	Final version	A. Wiesmann, T. Strozzi and A. Bartsch

Author team

Andreas Wiesmann, Rafael Caduff and Tazio Strozzi, GAMMA

Alexandru Onaca and Flavius Sîrbu, WUT

Valentin Poncos, TERRASIGNA

Annett Bartsch, B.GEOS

ESA Technical Officer:

Frank Martin Seifert

TABLE OF CONTENTS

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, the System Development Document, includes the System Requirement Document (SRD), the System Specification Document (SSD), and the System Verification Report (SVR) of CCN1 and CCN2 on rock glacier kinematics as a new associated parameter of ECV Permafrost. It is intended as an appendix to the three deliverables D3.1 [RD-11], D3.2 [RD-12] and D.3.3 [RD-13] of the baseline Permafrost_cci project. Compared to the baseline project (with three iterations), CCN1 and CCN2 comprise only one algorithm and system development iteration over 18 months. In CCN1, according to the user requirements and algorithm development, the parameters to be retrieved in the Romanian Carpathians comprise a permafrost distribution model at regional scale using a random forest algorithm and trends in rock glaciers velocity from ALOS-2 PALSAR-2 and Sentinel-1 SAR interferometry. CCN2 consists of two options led by Swiss and Norwegian teams focusing on the investigation and definition of a new associated ECV Permafrost variable related to rock glacier kinematics through the two products “regional rock glacier inventory” (RGI) and “kinematical time series of selected rock glacier” (KTS).

The first section of this document outlines the requirements for the Permafrost_cci CCN1&2 processing system. This document provides details on the processing system, system scenarios, workflows and a detailed list of requirements. The requirements are grouped by type and identified by a unique 3 level identifier.

The second section of this document outlines the specifications for the Permafrost_cci CCN1&2 processing system. It describes purpose and intended use as well as the main requirements, functions and components. Further, it discusses the main operational scenarios and the necessary infrastructure and highlights the functional design from different perspectives (users, system operators and developers). Finally, it summarises the specifications concerning system life cycle design, implementation and maintenance costs and performance.

The third section of this document outlines the system verification procedures and results for the processing system. A benchmark test scenario is defined for the system verification. 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.

1 INTRODUCTION

1.1 Purpose of the document

This document provides the Task 3 deliverables of CCN1: **"**Rock glacier kinematics in the Carpathians (Romania)" and CCN2 "Rock glacier kinematics as new associated parameter of ECV permafrost" of the Permafrost_cci project, i.e. the System Requirement Document (SRD), the System Specification Document (SSD), and the System Verification Report (SVR). According to the user requirements and algorithm development, the parameters to be retrieved in the Romanian Carpathians for CCN1 comprise a permafrost distribution model at regional scale using a random forest algorithm and trends in rock glaciers velocity from ALOS-2 PALSAR-2 and Sentinel-1 SAR interferometry. CCN2 consists of two options led by Swiss and Norwegian teams focusing on the investigation and definition of a new associated ECV Permafrost variable related to rock glacier kinematics through the two products "regional rock glacier inventory" (RGI) and "kinematical time series of selected rock glacier" (KTS).

The purposes of Task 3, hereafter documented, include:

- the specification of the requirements of a processing system capable of generating the requested data products as specified in the first CCN1 and CCN2 deliverables (Task 1 [RD-4] and Task 2 [RD-10]) and the CCI project guidelines;
- the specification of an operational production system from a system engineering point of view;
- the verification that the system is properly working when executed in other hardware and/or software environments.

The Task 3 document of CCN1 and CCN2 are intended as an appendix to the three deliverables D3.1 [RD-11], D3.2 [RD-12] and D.3.3 [RD-13] of the baseline Permafrost_cci project.

1.2 Structure of the document

Section 2 of this document specifies the requirements of a processing system capable of generating the Permafrost_cci data products for CCN 1&2, i.e. regional rock glacier inventories, kinematical time series of selected rock glacier, and a permafrost distribution model at regional scale. The processing is first presented by giving the system overview. Then, different scenarios are discussed, and the workflow of the processing system is recalled. The detailed requirements for data processing functions of each step of the processing chain, including data volumes, the platform specification and the compliance to all processing needs defined by the Task 3 inputs, are finally listed.

Section 3 incorporates the system requirements and specifies the characteristics of an operational production system from a system engineering point of view. The system design is based on experience with prototype processors developed prior to this project.

Section 4 discusses the verification methodology background of the processing system, presents the module tests and reports the testing procedures.

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-SW-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

[AD-5] Bartsch, A.; Grosse, G.; Kääb, A.; Westermann, S.; Strozzi, T.; Wiesmann, A.; Duguay, C.; Seifert, F. M.; Obu, J.; Goler, R.: GlobPermafrost – How space-based earth observation supports understanding of permafrost. Proceedings of the ESA Living Planet Symposium, pp. 6

[AD-6] National Research Council. 2014. Opportunities to Use Remote Sensing in Understanding Permafrost and Related Ecological Characteristics: Report of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/18711.

[AD-7] IPA Action Group 'Specification of a Permafrost Reference Product in Succession of the IPA Map' (2016): Final report. https://ipa.arcticportal.org/images/stories/AG_reports/IPA_AG_SucessorMap_Final_2016.pdf

[AD-8] GlobPermafrost team (2017): Summary report from 3rd user Workshop. ESA DUE GlobPermafrost project. ZAMG, Vienna. https://www.globpermafrost.info/cms/documents/reports/ESA_DUE_GlobPermafrost_workshop_summary_ACOP_v1_public.pdf

1.4 Reference Documents

[RD-1] Bartsch, A., Matthes, H., Westermann, S., Heim, B., Pellet, C., Onaca, A., Kroisleitner, C., Strozzi, T. (2020): ESA CCI+ Permafrost User Requirements Document (URD), v2.0

[RD-2] Bartsch, A., Westermann, Strozzi, T., Wiesmann, A., Kroisleitner, C. (2020): ESA CCI+ Permafrost Product Specifications Document (PSD), v2.0

[RD-3] Bartsch, A., Westermann, S., Heim, B., Wieczorek, M., Pellet, C., Barboux, C., Kroisleitner, C., Strozzi, T. (2020): ESA CCI+ Permafrost Data Access Requirements Document (DARD), v2.0

[RD-4] Strozzi, T., Onaca, A., Poncos, V., Ardelean, F., Bartsch, A. (2019): ESA CCI+ Permafrost CCN1 D1. User Requirement, Product Specifications and Data Access Requirements Document, v1.0

[RD-5] Bartsch, A., Westermann, S., Strozzi, T. (2020): ESA CCI+ Permafrost Product Validation and Algorithm Selection Report (PVASR), v2.0

[RD-6] Westermann, S., Bartsch, A., Strozzi, T. (2020): ESA CCI+ Permafrost Algorithm Theoretical Basis Document (ATBD), v2.0

[RD-7] Westermann, S., Bartsch, A., Heim, B., A., Strozzi, T. (2020): ESA CCI+ Permafrost End-to-End ECV Uncertainty Budget (E3UB), v2.0

[RD-8] Westermann, S., Bartsch, A., Heim, B., A., Strozzi, T. (2020): ESA CCI+ Permafrost Algorithm Development Plan (ADP), v2.0

[RD-9] Heim, B., Wieczorek, M., Pellet, C., Delaloye, R., Barboux, C., Westermann, S., Bartsch, A., Strozzi, T. (2020): ESA CCI+ Permafrost Product Validation Plan (PVP), v2.0

[RD-10] A. Onaca, F. Ardelean, F. Sirbu,V. Poncos, T. Strozzi, A. Bartsch (2020): ESA CCI+ Permafrost CCN1 D2. Algorithm Development Document, v2.0

[RD-11] A. Wiesmann, A. Bartsch, S. Westermann, T. Strozzi (2020): ESA CCI+ Permafrost System Requirement Document (SRD), v2.0

[RD-12] A. Wiesmann, A. Bartsch, S. Westermann, T. Strozzi (2020): ESA CCI+ Permafrost System Specification Document (SSD), v2.0

[RD-13] A. Wiesmann, A. Bartsch, S. Westermann, T. Strozzi (2020): ESA CCI+ Permafrost System Verification Report (SVR), v2.0

[RD-14] C. Barboux, A. Bertone, R. Delaloye, A. Onaca, F. Ardelean, V. Poncos, A. Kääb, L. Rouyet, H. Christiansen, T. Strozzi, A. Bartsch (2019), ESA CCI+ Permafrost CCN1 & CCN2 User Requirement Document (URD), v1.0

[RD-15] C. Barboux, A. Bertone, R. Delaloye, A. Onaca, F. Ardelean, V. Poncos, A. Kääb, L. Rouyet, H. Christiansen, T. Strozzi, A. Bartsch (2019), ESA CCI+ Permafrost CCN1 & CCN2 Data Access Requirement Document (DARD), v1.0

[RD-16] C. Barboux, A. Bertone, R. Delaloye, A. Onaca, F. Ardelean, V. Poncos, A. Kääb, L. Rouyet, H. Christiansen, T. Strozzi, A. Bartsch (2020), ESA CCI+ Permafrost CCN1 & CCN2 Product Specification Document (PSD), v1.2

[RD-17] C. Barboux, A. Bertone, R. Delaloye, A. Onaca, F. Ardelean, V. Poncos, A. Kääb, L. Rouyet, H. Christiansen, T. Strozzi, A. Bartsch (2020), ESA CCI+ Permafrost CCN1 & CCN2 Algorithm Theoretical Basis Document (ATBD), v1.0

1.5 Bibliography

A complete bibliographic list that support arguments or statements made within the current document is provided in Section 5.1.

1.6 Acronyms

A list of acronyms is provided in Section 5.2.

1.7 Glossary

A comprehensive glossary of terms relevant for the parameters addressed in Permafrost_cci is available as part of the Product Specifications Documents of the baseline project [RD-2] and of CCN 1 [RD-4].

2 SYSTEM REQUIREMENT DOCUMENT (SRD)

2.1 Introduction

The aim of the SRD is to collect all requirements on the Permafrost_cci CCN1&2 Processing System (PS). The main sources of information are the SoW with its Annex [AD-1], the URD [RD-1], PSD [RD-2] and (DARD) [RD-3] of the baseline project as well as D1 of CCN1 [RD-4], and the CCI project guidelines. The scope of the SRD is to specify the requirements of a processing system capable of generating data products as specified in the applicable technical annex. The SRD shall include verifiable requirements on data processing function of each step of its processing chain, including data volumes, platform specification, and compliance to all processing needs defined by the Task 3 inputs. The processing algorithms and output products are being defined to some degree in parallel to this document, e.g. in the Task 2 deliverables. This affects some key areas such as the system performance and sizing.

2.2 Context of the algorithms

The Permafrost_cci baseline project is focused on the product ground temperature and relevant derivatives (active layer thickness and permafrost zones). CCN1&2 extent this to rock glacier kinematics.
The Processing System (PS) defined in the context of Permafrost_cci is a science-driven system that produces the required data products to satisfy the GCOS and evolved data requirements. It is under configuration control and maintenance (bug tracking, reprocessing, traceability) and is technically capable of being sustainable in the long term beyond funding from the CCI programme.

The Permafrost_cci CCN1&2 PS has to consider the following issues for each of the products:

- data archive;
- data production;
- data services.

Data archiving contains the need to retrieve and store the input data, auxiliary products (e.g. DEMs), intermediate products from the applied algorithms, and output data (final products). The data production deals with the generation of the final products, including meta-data and log files. The data services issue is related to data accessibility of the final products for the scientific community. For the Permafrost_cci system it must be considered that for all products mandated repositories exist so that the PS only has to consider the datasets that are required for data production. Indeed, selected common services are offered within the CCI projects for sharing among ECVs. Among them are a backup archive for the data, cloud services that can be used by other ECVs, and a CCI product viewer.

The CCN1 & CCN2 Permafrost_cci modules are considered along that of the CryoGrid module from the baseline project:

- Regional rock glacier inventories, incl. kinematics (RGI)
- Kinematical time series on selected rock glaciers (KTS)

- Mountain Permafrost in the Southern Romanian Carpathians (MPDM)

2.3 User Requirements

Specific user requirements are documented in the URD section of RD-14. All specific user requirements for CCN1 and CCN2 are listed in Table 1 to complement Table 2 for the baseline project [RD-11].

Table 1: Summary of user requirements. Background (BG) means that this is a continuous activity, production (P) means that the related requirement has to be considered during production. Parameters are Rock Glacier Inventories (RGI), Kinematical Time Series for selected rock glaciers (KTS), Ground Temperature (GT) and Active Layer Thickness (ALT).

ID	PARAMETER	REQUIREMENTS	TYPE
URQ_01	RGI	Regional and global coverage	BG
URQ_02	RGI	Current year overview	BG
URQ_03	RGI	Identification of rock glaciers by a point	P
URQ_04	RGI	Multi-unit differenciation	BG
URQ_05	RGI	Update every 10 years	-
URQ_06	RGI	Rock glacier activity according to IPA Action Group classification [RD-6]	P
URQ_07	RGI	Rock glacier destabilization (optional)	P
URQ_08	RGI	Kinematics attribute to be defined	BG/P
URQ_09	RGI	Moving areas identification and classification	P
URQ_10	RGI	Precision & accuracy (up to 30% of rock glaciers in an inventory could be undefined)	P
URQ_11	KTS	At east regional coverage	BG
URQ_12	KTS	Sufficient rock glaciers representative in a defined regional context ("representatives" has to be defined)	BG/P
URQ_13	KTS	Update every 5 years	-
URQ_14	KTS	Yearly or seasonally data with an annual time step	P
URQ_15	KTS	At least past 5 years have to be investigated	P
URQ_16	KTS	Exact value of velocity	P
URQ_17	KTS	Horizontal resolution of the velocity value has to be defined	BG
URQ_18	KTS	Precision and accuracy < 1 cm/y	BG
URQ_19	GT	Threshold: global with regional specific product, yearly, last decade, 1km, subgrid variability, RMSE<0.1k; Target: Southern Carpathians, monthly, 1979- present, 100m, subgrid variability, RMSE<0.1k.	BG
URQ_20	ALT	Threshold: global with regional specific product, yearly, last decade, 1km, RMSE<0.25m; Target: Southern Carpathians, monthly, 1979- present, 100m, RMSE<0.25 cm.	BG
URQ_21	RGI	Threshold & Target: Southern Carpathians, Last decade	BG
URQ_22	KTS	Threshold: Southern Carpathians, yearly, last decade,	BG

2.4 Processing System Scenarios

The sensor constellation and temporal coverage for CCN1 & CCN2 are in the corresponding DARD [RD-15]. For the permafrost distribution model at regional scale a cloud free satellite image that covers the entire area is required. For the calculation of trends in velocity of selected rock glaciers (KTS) ALOS-2 PALSAR-2 and Sentinel-1 data, both since 2014, are considered.

The product specifications for the permafrost distribution model at regional scale and for trends in velocity selected rock glaciers are described in detail in the PSD [RD-16]. TABLE 2 below gives a summary of the specifications.

Table 2: Permafrost_cci CCN1 & CCN2 product specifications (NA: Not Applicable).

	Regional Rock Glacier Inventories (RGI)	**Kinematic Time Series (KTS)**	**Mountain Permafrost Distribution Model (MPDM)**
Parameter	kinematical rock glacier inventories	velocity along the maximum slope direction	permafrost distribution model at regional scale
Description	rock glaciers and related moving area spatially identified and characterized	kinematic time-series of rock glaciers	probability
Spatial Coverage	European Alps, European Subarctic/Arctic and Extra-European sites	Selected rock glaciers in the European Alps, European Subarctic/Arctic and Extra-European sites	Southern Carpathians (areal extent: 14,000 sq km)
EO Data	Sentinel-1, TerraSAR-X, Cosmo-SkyMED, ALOS-2 PALSAR-2, ERS-1/2	Sentinel-1, ALOS-2 PALSAR-2	Landsat-8 or Sentinel-2
Spatial Resolution	0.01 km2	30 m	30 m
Spatial Aggregation	none	none	none
Spatial Filling	NA	NA	NA
Weather Station Data	NA	NA	NA
Period	1990-2020	2014-2020	2019
Frequency	NA	every 6 days	annually
Temporal Aggregation	NA	NA	NA
Update Frequency	NA	every 6 days	NA
Map Projection	UTM WGS84	UTM WGS84	UTM WGS84
Coding	8 bit	8 bit	8 bit
Format	NetCDF	NetCDF	NetCDF
Accuracy	1/10 of a	0.2 m/a	>0.7

	Regional Rock Glacier Inventories (RGI)	Kinematic Time Series (KTS)	Mountain Permafrost Distribution Model (MPDM)
Target	wavelength		
Uncertainty Metric	AUC (area under the curve)	coherence	AUC (area under the curve)
Metadata	Yes	Yes	Yes
Data Access	CCI data portal	CCI data portal	CCI data portal

It is an important requirement on the processing system that data can be reprocessed. Consequently, the process has to be reproducible and the system sufficiently powerful to allow reprocessing in time. Reprocessing can be due to improved input data quality, improved processing software or improved algorithms. An important feature is the possibility to only run parts of the system. That allows to keep intermediate results and allows to reduce the reprocessing resources and time.

Another common topic identified in the Systems Engineer Working Group is the versioning and improvement cycle. A favourite approach is suggested by the SST_CCI Science Leader C. Merchant, see SRD of the baseline project [RD-11].

There is no foreseen direct interaction between the PS and the users. Communication with the users is done through the project management. Data access for the users is enabled through the dedicated CCI product portals.

2.5 Processing System Production Workflow

The basic processing system production workflow of the baseline project is given in the corresponding SRD [RD-11]. It has the main parts preprocessing, retrieval, product generation and verification/validation. The basic processing system production workflow for the CCN1 & CCN2 processing systems is detailed in the Algorithm Technical Basis Document [RD-17].

2.5.1 Regional rock glacier inventories, including kinematics (RGI)

The regional rock glacier inventories are based on previously existing inventories (InSAR polygons from GlobPermafrost and/or existing morphological rock glacier inventories). The update/upgrade follows the procedure given in Figure 1 and follows the standards described in the CCN1 & CCN2 Product Validation and Assessment Report (PVASR). The two outputs are the moving areas related to rock glacier units and the rock glacier inventory.

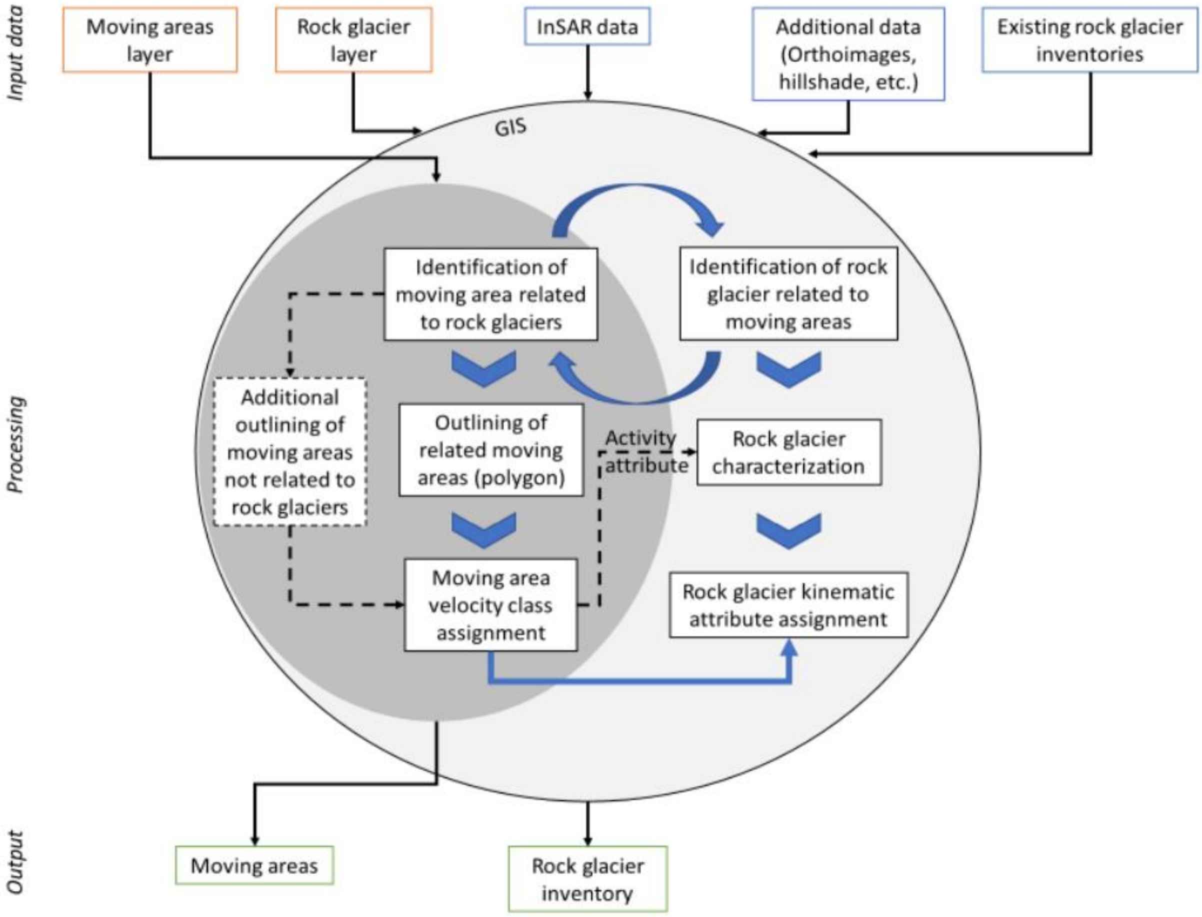

Figure 1: Procedure for the proposed standardization of rock glacier inventory using InSAR data. The analysis is performed in GIS software

2.5.2 Kinematical time series on selected rock glaciers (KTS)

The processing line for KTS consists of the measurement of the InSAR time series on selected rock glaciers, potentially complemented by time series from SAR offset tracking, feature tracking on repeat optical airphotos, and the development of standardized comparable products.

2.5.3 Mountain permafrost distribution model in Southern Carpathians (MPDM)

The Mountain Permafrost Distribution Model (MPDM) is a modelling product based on a RF classification algorithm that learns all the characteristics of the terrain and spectral data for the training area (for both the areas with permafrost and without permafrost) and searches for similar characteristics of the independent variables in the rest of the study area.

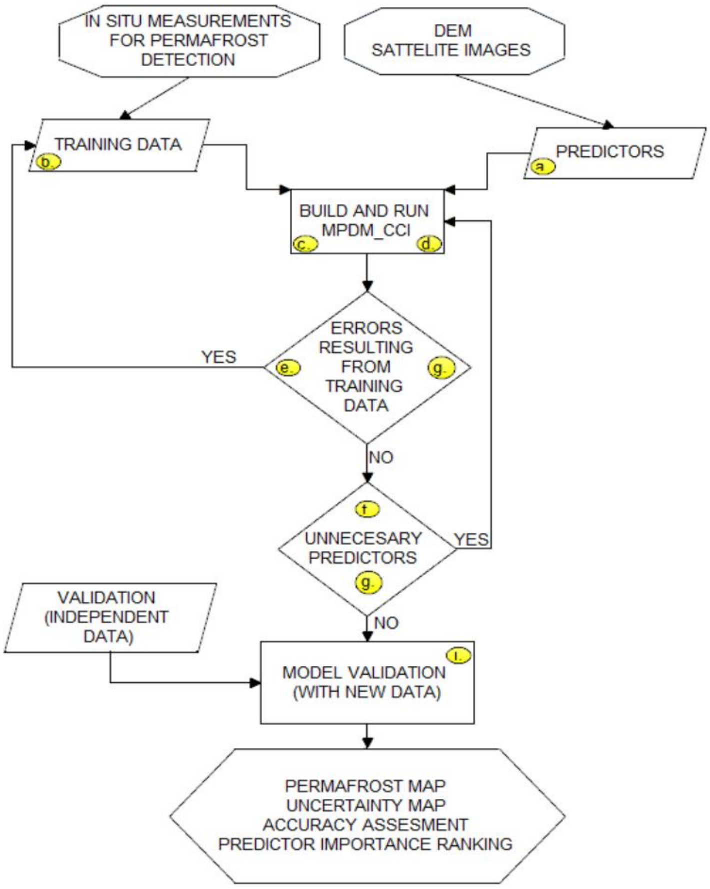

Figure 2: The workflow of the mountain permafrost distribution model in the Southern Carpathian (MDPM)

2.6 Detailed Processing System Requirements

The systems need to be designed to be sustainable. Hence, there is a need to plan for evolution from the prototype approach towards compliance with applicable software standards e.g. appropriate components of ECSS-E-ST-40C. This implies that the requirements for configuration control and maintenance (bug tracking, reprocessing, traceability), operability and transferability are priorities. They shall at the start of the project identify the correspondence between the documentation set within this project and those required by the applicable Software Standard.

Since the system is science-driven it must be capable of being regularly updated as scientific understanding improves and new algorithms are developed. The incorporation of new algorithms needs to consider trade-offs in cost, complexity and scientific impact on the quality and consistency of outputs, and the introduction of new algorithms must not jeopardise the output generation. The design should also be modular and flexible while at the same time capable of rapid reprocessing, thus the overall design needs to be developed with end-to-end throughput of the ECV production as a design priority.

The requirements of the Permafrost_cci baseline PS were collected in [RD-11]. They have a unique 3 level identifier of the format "PF-TYPE-ID". The "PF" stands for Permafrost_cci. The "TYPE" is defined similarly as in other CCI projects:

FUN: functional
PER: performance
SIZ: sizing
INT: interface
OPE: operational
RAM: availability, maintainability, security

The "ID" is a 4-digit number. Here, we include additional detailed processing system requirements related to the parameters to be computed for CCN1&2 on the rock glacier kinematics and for CCN1 regarding ground temperature modelling in the Carpathians (Romania).

High Level System Requirements

PF-FUN-0110	Develop and validate algorithms to compute a permafrost distribution model at regional scale in the Southern Carpathians
PF-FUN-0120	Produce, validate and deliver velocity trends of selected rock glaciers
FP-FUN-0130	Produce, validate and deliver a Rock Glacier Inventory

3 SYSTEM SPECIFICATION DOCUMENT (SSD)

3.1 Processing System Overview

This document incorporates the requirements described in the System Requirements Document (Section 2) and outlines the specifications for the CCN1 & CCN2 Processing System (PS) from a system engineering point of view. It describes its purpose and intended use as well as the main requirements, functions and components and discusses the main operational scenarios, the necessary infrastructure and highlights the functional design from different perspectives, the users, system operators and developers view. Finally, it summarises the specifications concerning system life cycle design, implementation and maintenance costs and performance.

The CCN1 and CCN2 PS are aligned with that of the baseline project described in [RD-12]. It generates products and supports the process of algorithm improvement, reprocessing and validation. It provides products and services to the permafrost community supporting their climate change impact assessment over a wide range of scales. The PS will be used by the Permafrost_cci consortium but can also be applied by others as the overall workflow is very generic. The key difference to data production in other science projects is their often-missing dissemination, i.e. the work ends with a publication and generated data products are not shared. The Permafrost_cci PS is specified to provide products in a transparent and documented way, with accompanying meta-data, documentation, uncertainty and validation reports (project outreach).

The Permafrost_cci PS can be understood as a value-adding layer between the data provider and the users. There are interfaces to the different user communities, which receive products and can provide feedback. Another interface is with the EO data providers. Depending on the module, EO data are obtained from the providers and ingested into the PS. Feedback is given to the providers about issues found with the data, processing improvements and requirements for the continuity of the service. Another interface is towards third-party sources to receive ancillary and validation datasets.

The SSD of the baseline project [RD-12] gives an overview of the processing system (PS) with its main modules, functions and components. It also summarises its designated use and the system requirements. In following, we remind the major characteristics of the Permafrost_cci PS and specify the peculiarities of the CCN1 and CCN2 modules.

3.2 Processing System Workflow and Operational Scenarios

As for the Permafrost_cci PS, the CCN1 and CCN2 development teams consist of scientists and operators that manage the production and continuous development. Actors in the operational scenarios are users with different roles depending on how they use the system:

- Permafrost_cci users (end user)
  → are interested in best existing permafrost products

→ are skilled in permafrost applications
→ provide feedback and proposals
→ request data format compatible with their communities

- Development team
  → mandated to run Permafrost_cci in dialogue with users
  → develops the PS further
  → issues product versions
- Validation experts
  → are part of the international community
  → support development team
  → have local expertise
  → feedback on the products
- Auditor/Project Manager
  → project supervision

The basic processing system production workflow of the Permafrost_cci PS can be found in [RD-12].

3.3 Functional Design

In this section we discuss and present the major functional blocks of the Permafrost_cci PS and the peculiaries of the CCN1 modules.

3.3.1 Services

Most of the necessary services are provided through the CCI environment such as the data exchange storage, the document management system and user interaction. For the CCN1 PS development, a software repository and an issue tracker are hosted at WUT and TERRASIGNA, respectively. For the CCN2 PS development, a software repository and an issue tracker are hosted at GAMMA and NORUT.

Processor software repository
An important element of the modern software development process is source control (or version control). Cooperating developers commit their changes incrementally to a common source repository, which allows them to collaborate on code without resorting to crude file-sharing techniques (shared drives, email). Source control tools track all prior versions of all files, allowing developers to "time travel" backward and forward in their software to determine when and where bugs are introduced. These tools also identify conflicting simultaneous modifications made by two (poorly-communicating) team members, forcing them to work out the correct solution (rather than blindly overwriting one or the other original submission). The software repository contains the actual processing code and all prior versions. The write access to the processor repository is restricted to the development team. As all software changes are updated directly in the repository, the software changes are published almost immediately and are made available for review.

Issue Tracker

During software development for MPDM_CCI a Redmine issue tracker is used (http://www.redmine.org/). Redmine is a flexible project management web application written using Ruby on Rails framework. Redmine integrates the version control system into its user interface and manages the access control to the version control system resulting in a state-of-the-art FOSS software development environment.

The MPDM_CCI is developed by the WUT team and is planned to be used only once during the current CCI project. The model will be validated for both scientific accuracy and processing system integrity as stated at chapter 4 of the present document. Thus, an issue tracker for further bug tracking and system development is not needed.

3.3.2 Processors

The processors cover the necessary tools to produce the different Permafrost_cci products. The Permafrost_cci is organized in modules covering the production of the different products. In general, a distributed processing approach is followed. Consequently, the modules are portable.

Table 3: CCN1 and CCN2 Permafrost_cci modules

COMPONENT	PURPOSE	IMPLEMENTATION
Permafrost distribution model at regional scale	Generate pixel-based probability distribution of permafrost presence-absence in Southern Carpathians	New processor
Rock Glacier Kinematics Module	Compute rock glacier velocity along the maximum slope direction	New processor implemented around the Gamma software
Rock Glacier Inventory Module	Identify and characterize rock glaciers and related moving area spatially	New processor implemented around QGIS

3.3.3 Concept for continuous improvement

Continuous improvement is an important aspect in the CCI projects. In order to ensure a transparent process, Software Modularity, Software Version Concept, Version Control, and Version Numbering are important issues. This section defines the structures and functions that extend the production environment for continuous improvement. Focus is on flexibility, rapid testing and prototyping. The concepts described are processors, versioning, and a test environment. The concept of processors and versioning contribute to the modularity of the system.

The software of the PS and the processing algorithms code are under version control. The software repository contains the actual processing code and all prior versions. All software changes are documented in the repository. Version numbering of the processor is reflected in the repository by revisions and tags. Revisions are usually linked to commits and indicate the sequential order of documented changes. Tags are set to indicate software releases of frozen software states. Subversion is a good candidate for version control. Together with Redmine it is a complete FOSS version control and issue tracking system.

Data processors help to organize the data processing in modules. Due to the differing input and output datasets/formats, the modules are normally not shared among products (even if the functionality is the same). A processor is a software component that can be parametrised and that generates a (higher level) output product of a certain type from one or several input products of certain types. A PS module consists of the sequential call of processors. Each processor has its own version information. Parameters and environment variables are provided in dedicated parameter files within the code. Feedback is received by a return code, messages on stdout/stderr and in log files.

3.3.4 System Documentation

The documentation contains the PS documentation consisting of manuals, specifications and reports, as well as the product documentation consisting of product specifications, manual and validation reports. At this stage no advanced functionality such as collaborative editing etc. seems to be necessary so that the basic functionality of any FOSS CMS might be sufficient for this task.

The PS documentation includes requirement documents, design and interface control documents, test documents, manuals, and maintenance information. Permafrost_cci deliverables to name here are ATBD and the PUG. The SRD and SSD define requirements and design of the system.

3.4 Development, Life Cycle, Cost and Performance

This section discusses the system development in the future, potential development strategies, efforts and costs. The development is driven by several factors such as the availability of new technology, faster algorithms, new scientific findings and improved product algorithms, new available EO data, and user needs.

3.4.1 Re-use and Development

Development is needed to bring the existing prototypes of the PS modules to a higher operational level satisfying the requirements listed in the previous sections and to add the missing components such as those for user services, data handling, life cycle management, archiving etc.

Requirements addressed by this section are:

→ PF-OPE-4030 Development based on user requirements
→ PF-OPE-4020 Development decoupled from research
→ PF-OPE-4060 Freeze prototype
→ PF-FUN-5030 Verification of implementation

In the TABLE 4 we summarise the tools that were used, adapted, configured and integrated during development of the PS within the CCI and beyond.

Table 4: Permafrost_cci development tools.

NAME	USAGE	REMARKS
Redmine	Issue tracking	FOSS
Python3	Scripting, netcdf4 reader, pyresample, rasterio	FOSS
R	Scripting	FOSS
SAGA	GIS (accessed through R scripting)	FOSS
QGIS	GIS	FOSS
Gamma software (ISP, DIFF&GEO)	InSAR Scripting	Commercial

3.4.2 System Life Cycle Drivers and Considerations

The PS needs to be incrementally adapted to integrate new functional extensions, improved algorithms and input datasets. New EO data make adaptations necessary and most likely also have an impact on the hardware infrastructure. The life cycle plan cannot be static as it is not foreseeable. Currently, the following driving factors are identified:

- Availability of the existing processor module prototypes
- Functional extension of the system
- New workflows
- Improved algorithms
- New Sensors
- Hardware improvements
- Dependencies on 3rd parties (other ECVs, data providers, new users)

To answer the first two points, the system is initially based on the prototype. Incrementally, additional components and functions are added and interfaces to data providers and users are extended. The third and fourth point on workflow and algorithm development requires the addition of tools for validation and user feedback. The new sensors and the increased data volume and resolution are a qualitative change, too. The existing methods need to be adapted to make use of new sensors. For the longer perspective renewal of hardware and optional change of software layers must be taken into account. The PS design is prepared for this by the modularity of its functional components. The last item is not so relevant for Permafrost_cci at the moment as the dependence on other CCI projects is minor.

3.4.3 Sizing and Performance Analysis

In the SRD, no specific requirements are present concerning the processing time performance. At the moment it is mainly labour hours that drive the processing rather than CPU core hours. Full reprocessing of historical data requires a variable amount of work, depending on the product. The data storage budget for inputs and outputs for historical data and for the yearly increase of acquired data is in the low TB range initially.

The budgets for data storage and processing capabilities are estimated in TABLE 5. The budget for data storage mainly depends on the amount of input data to be managed. This comprises historical data and data acquired continuously.

Table 5: Permafrost_cci CCN1 & CCN2 annual data storage budget.

DATA	PRODUCT	TIME SPAN	HISTORICAL DATA	SIZE / Y
EO data (Landsat 8 / Sentinel-2) DEM and derived parameters	MPDM	NA	Thermal and geophysical data from in-situ measurements	1 GB
Sentinel-1, DEM	KTS	6 days	None	18 TB
Sentinel-1, ALOS-2 PALSAR-2, TerraSAR-X, Cosmo-SkyMED, DEM	RGI	NA	ERS-1/2	1 TB

3.4.4 Cost Estimation

The costs for the PS are composed of costs for storage, processing, network, development and integration, operations, dissemination and labour. The different modules have different needs.
The MPDM is developed as an open source script that runs on the open source scripting platform R and uses the GIS engine provided by the open source software SAGA. The costs for the development and testing of the product are covered by the current project. The computing environment is provided by the WUT through the Department of Geography. The storage of MPDM_CCI together with a sample data set is provided by the WUT through the Department of Geography and it is made available for download by the end users.
The KTS module is developed around the GAMMA software. The costs for the development and testing of the product are covered by the current project. The computing environment is provided by GAMMA. The storage of KTS_CCI is provided by GAMMA and it is made available for download by the end users.
RGI are mapped manually by external providers based on InSAR maps produced by the CCI team. The processing and mapping costs are covered by the current project. The storage of intermediate (InSAR maps) and final (RGI) products is provided by GAMMA. RGI will be made available for download through CCI.

4 SYSTEM VERIFICATION REPORT (SVR)

4.1 Introduction

The System Verification Report (SVR) should confirm that the system as outlined in the SSD and described in more detail in the CCN1&2 ATBD [RD-17] is properly working when executed in other hardware and/or software environments. 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.

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 work around anomalous, inappropriate, or undesired operating conditions.

For the CCN1&2 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. For testing of the PS it is ensured that the test operator is not involved in the development or implementation of the tested module.

4.2 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.

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.

4.3 Qualification and description of the test environment

4.3.1 KTS

The KTS processing system is implemented in Python and based on the GAMMA Software. It is developed on a linux workstation and implemented for testing on a different production linux workstation with 132 GB RAM, an Intel i9-9980XE processor with 16 cores and a 56 TB Raid6 storage.

4.3.2 MPDM

The processing chain for producing a permafrost distribution model at regional scale in the Southern Carpathians is implemented in the open source software R and run on a desktop computer, with 32GB DDR4, Intel® Core™ i7-7820HK processor. The required parameters from DEM and the landcover classes are derived using the open source software SAGA GIS running on a Windows 10 operating system. The processing chain for producing velocity trends of selected rock glaciers in the Southern Carpathians is implemented around the Gamma software on a Linux computer using scripts.

4.4 Test Protocol

The Permafrost_cci CCN1 and CCN2 processing chains consist of single sequences of modules from which all products are computed. For this reason, we can outline a verification procedure that is capable of testing the performance with respect to the products.

4.4.1 KTS

Objectives and Scope

To ensure the prototype processing is consistent on the production platform an end-to-end test is done based on a standard production script and a standard set of input data.

Components Tested

The process to create SLC subsets of S1 tracks and the calculation of line of sight (LOS) deformation maps based on new data is tested.

Input Data and Reference Output Data

Input data (md5sum, name)
9e4e66f394f6a105975215c0824b1387 S1_CH_T138D_20201113.rslc
a20470ad0fe69eab1a2d05781b5e4dc0 S1_CH_T138D_20201113.rslc.par

Benchmark result data (md5sum, name)
5a77aaa12d700f76d5b8e50f9445ff0f Distelhorn/Distelhorn_122_186.csv
2fabd71f756828dfa53e4c034a554611 Distelhorn/Distelhorn_144_147.csv
8670dc34500935465f685c1916584edc Distelhorn/Distelhorn_157_153.csv
b23c928575e4cf0216a0dd484235bf7d WassenRG/WassenRG_197_375.csv
2a74ab98497d9ea105085de4bb6d22d5 WassenRG/WassenRG_204_336.csv
1786122878791806718c62af29b66d5b WassenRG/WassenRG_216_394.csv
423a8958a9ef514b57272f69dec524d5 WassenRG/WassenRG_225_411.csv
abda9d86bb6ee60b333367b02e9bb9a7 WassenRG/WassenRG_240_379.csv

Test Procedure and requirements for successful testing

1. Check input data hashes
2. Run script S1_create_subsets_rockglaciers.py T138D 20201113
3. Check processing time within limits (< 1 min)
4. Check output data hashes versus benchmark data

All checks need to be successfully passed for a successful test.

Implications for System Engineering

Processing needs Python3 and GAMMA software installed

4.4.2 MPDM

Objectives and Scope

To ensure that every processing step outputs correct results.

To ensure that - whatever the quality of the final result of the measurements - the PS is functionally correct.

Components Tested

A. Permafrost distribution model is tested on scientific accuracy for:

A1. Correlation of input predictors (highly correlated predictors are removed)

A2. Evaluation of the uncertainty map for any positional bias and systematic errors

A3. Accuracy estimation of the modelling based on randomly extracted samples from the input data

B. Permafrost distribution model is tested on processing system integrity for:

B1. Compatibility of input data file formats with MPDM_CCI

B2. Compatibility and consistency of input data spatial resolution between the DEM and satellite images

B3. Compatibility of training data with the MPDM_CCI

B4. Size of the training data

B5. Script integrity

Input Data and Reference Output Data

Permafrost distribution model:

A1. Input: DEM and satellite images

Output: resampled DEM and derived parameters/satellite images specified in section 2.5

A2. Input: in situ data from geophysical and thermal measurements on selected rock glaciers

Output: Training data from in situ measurements

A3. Input: Model run with all the predictors

Output: uncertainty map, predictor importance ranking

A4. Input: Updated model

Output: Validated permafrost extension map, uncertainty map, accuracy assessment,

Permafrost distribution model processing system integrity test:

B1. Input: input data

Output: a system check message with “Input data compatible with MPDM_CCI”/ “Input data compatible with MPDM_CCI”

B1. Input: input data

Output: a system check message with “Input data with consistent spatial resolution”/ “Input data spatial resolution inconsistency”

B3. Input: training data (shapefile)

Output: the head of the attribute table for the shapefile used as training data (It has to be verified by the end user to respect the guide lines provided in the product description).

B4. Input: training data
Output: model accuracy
B5. Input: data sample (input data and training data)
Output: model results (permafrost map, uncertainty map, accuracy assessment)

Test Procedure and requirements for successful testing
Use MD5 hash to compare log file with reference .

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.
Because the MPDM_CCI is developed on an open source platform compatibility errors are to be expected when using different versions of the software. A description of the correct software version and a sample data for testing the product are provided by the WUT.

4.5 Testing Outcome

4.5.1 KTS

The system verification was done at GAMMA on the workstation gamma44 with access to the input data located on the prototype system on gamma43. The test was conducted and documented by Rafael Caduff.

1. Input data have consistent hashes with the test data PASSED
2. Start script S1_create_subsets_rockglaciers.py T138D 20201113 PASSED
3. Processing time is 22s that is within the limit of 1min PASSED
4. Check output data hashes versus benchmark data:
   5a77aaa12d700f76d5b8e50f9445ff0f Distelhorn/Distelhorn_122_186.csv
   2fabd71f756828dfa53e4c034a554611 Distelhorn/Distelhorn_144_147.csv
   8670dc34500935465f685c1916584edc Distelhorn/Distelhorn_157_153.csv
   b23c928575e4cf0216a0dd484235bf7d WassenRG/WassenRG_197_375.csv
   2a74ab98497d9ea105085de4bb6d22d5 WassenRG/WassenRG_204_336.csv
   17861228787918067 18c62af29b66d5b WassenRG/WassenRG_216_394.csv
   423a8958a9ef514b57272f69dec524d5 WassenRG/WassenRG_225_411.csv
   abda9d86bb6ee60b333367b02e9bb9a7 WassenRG/WassenRG_240_379.csv PASSED

- **Overall test status: PASSED**

4.5.2 MPDM

The verification and product generation within Permafrost_cci CCN1 for MPDM was accomplished with the existing infrastructure from WUT. Storage of DEM, satellite images and derived parameters along with the model was realized using the infrastructure from the Geography Department from WUT.

The verification and product generation within Permafrost_cci CCN&2 for the trends in velocity of selected rock glaciers from InSAR was accomplished with the infrastructure from Terrasigna Romania.

5 REFERENCES

5.1 Bibliography

N.A.

5.2 Acronyms

AD	Applicable Document
ATBD	Algorithm Theoretical Basis Document
AUC	Area Under the Receiver Operating Curve
B.GEOS	b.geos GmbH
BTS	Bottom Temperature of Snow Cover
CCI	Climate Change Initiative
CR	Cardinal Requirement (as defined in [AD-1])
CCN	Contract Change Notice
CRS	Coordinate Reference System
DARD	Data Access Requirement Document
DEM	Digital Elevation Model
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
FOSS	Free Open Source Software
GAMMA	Gamma Remote Sensing AG
GCOS	Global Climate Observing System
GFI	Ground Freezing Index
GPR	Ground Penetrating Radar
GST	Ground Surface Temperature
GTOS	Global Terrestrial Observing System
GUIO	Department of Geosciences University of Oslo
IPA	International Permafrost Association
IPCC	Intergovernmental Panel on Climate Change
L4	Level 4
MAGT	Mean Annual Ground Temperature
MAGST	Mean Annual Ground Surface Temperature
MPDM_CCI	Mountain Permafrost Distribution Model_ Climate Change Initiative
MRI	Mountains Research Initiative
MTD	Miniature Temperature Data Loggers
NDVI	Normalized Difference Vegetation Index
NDSI	Normalized-Difference Snow Index
NDWI	Normalized difference water Index
NIR	Near Infra Red

NMA	National Meteorological Administration
NSIDC	National Snow and Ice Data Center
PE	Permafrost Extent
PS	Processing System
PSD	Product Specifications Document
PUG	Product User Guide
QA4EO	Quality assurance framework for earth observation
RF	Random Forest
RD	Reference Document
RMSE	Root Mean Square Error
RS	Remote Sensing
SAR	Synthetic Aperture Radar
S4C	Science for the Carpathians
SVM	Support Vector Machine
T	Temperature
URD	Users Requirement Document
UTM	Universal Transverse Mercator
WGS	World Geodetic System
WUT	West University of Timisoara