



permafrost
cci

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

**CCN3 OPTION 6
IMPROVED SOIL DESCRIPTION THROUGH A LANDCOVER
MAP DEDICATED FOR THE ARCTIC**

**D2 Design justification (DJ), Technical Specifications (TS)
and Design Definition (DDF)**

VERSION 1.0

27 APRIL 2022

PREPARED BY

b·geos



J **GAMMA REMOTE SENSING**

Document status sheet

Issue	Date	Details	Authors
1.0	27.04.2022	First full version	AB, BW, AE, JP, GH, TS

Author team

Annett Bartsch, BGEOS
 Barbara Widhalm, BGEOS
 Aleksandra Efimova, BGEOS
 Gustaf Hugelius, SU
 Juri Palmtag, SU
 Tazio Strozzi, GAMMA

ESA Technical Officer:
 Frank Martin Seifert

<p>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.</p>

TABLE OF CONTENTS

Executive summary.....4

1 Introduction5

 1.1 Purpose of the document 5

 1.2 Structure of the document 5

 1.3 Applicable documents 5

 1.4 Reference Documents 5

 1.5 Bibliography..... 6

 1.6 Acronyms 6

2 Design justification7

 2.1 Methods and data 7

 2.2 Implementation options..... 11

3 Technical specifications21

4 Design definition22

 4.1 System components..... 22

 4.2 System requirements 23

 4.3 Trade-off analyses of implementation options..... 24

5 Summary25

6 References26

 6.1 Bibliography..... 26

 6.2 Acronyms 28

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). As part of the Permafrost_cci baseline project, ground temperature and active layer thickness were considered the primary variables that require climate-standard continuity as defined by GCOS. Permafrost extent and zonation are secondary parameters, but of high interest to users. The ultimate objective of Permafrost_cci is to develop and deliver permafrost maps as ECV products primarily derived from satellite measurements. Algorithms have been identified which can provide these parameters ingesting a set of global satellite data products (Land Surface Temperature LST, Snow Water Equivalent SWE, and landcover) in a permafrost model scheme that computes the ground thermal regime. Annual averages of ground temperature and annual maxima of thaw depth (active layer thickness) were provided at 1km spatial resolution during three phases of Permafrost_cci. The data sets were created from the analysis of lower level data, resulting in gridded, gap-free products. EO data sets are employed to determine the upper boundary condition of the differential equation, while its coefficients (e.g. heat capacity and thermal conductivity) are selected according to landcover information. Subgrid information on landcover must be used to generate ensembles of input parameters for permafrost modelling.

CCN3 Option 6 addresses the need for landcover information of relevance for Permafrost monitoring and modelling. The specific aim of this CCI+ Permafrost subproject is to implement a circumpolar landcover description with sufficient thematic content. It utilizes prototypes of ESA DUE GlobPermafrost, i.e. traditional landcover classification, vegetation height maps and surface roughness maps.

This document describes the design engineering process. This includes the justification for the choice of methods with respect to user needs. Product specifications are documented and information on levels of system design, engineering results and final selected algorithms provided. The landcover prototype has been extended regarding thematic content (four additional classes enhance the representation of disturbances and snow&ice). The original Maximum Likelihood Approach (MLH, with a preceding K-Means step) has been compared to Gradient Boosting. Three different dataset have been used for evaluation: soil pedons with focus on organic layer thickness as well as sites with comprehensive soil descriptions (both circumpolar) and vegetation height measurements (Western Siberia). MLH has been eventually selected with respect to the user requirements. The product specification includes class descriptions including correspondence to Landcover_cci. The focus is on tundra representation, forest classes have not been assessed, although main categories (mixed, deciduous, needle leaved) are provided for characterization of the transition zone. The use of Landcover_cci for forest related Plant Functional Types is recommended.

1 Introduction

The design engineering document provides information on the implementation of a circumpolar landcover description with sufficient thematic content as required for Permafrost_cci. It builds on prototypes of ESA DUE GlobPermafrost, i.e. traditional landcover classification and vegetation height maps. It provides an overall summary of choice of algorithms and product specification.

1.1 Purpose of the document

This document provides the user design engineering for CCN 3 Option 6 (option led by b.geos). The DF details the justification for the choice of methods with respect to user needs. The TS document the design engineering with focus in product specifications. The DDF provides information on all levels of system design, engineering results and final selected algorithms.

1.2 Structure of the document

Section 2 details the design justification, section the product specification and the design definition is provided in section 4.

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. 2016. The Global Observing System for Climate: Implementation Needs. GCOS Implementation Plan, WMO.

1.4 Reference Documents

[RD-1] Bartsch, A., Matthes, H., Westermann, S., Heim, B., Pellet, C., Onacu, A., Kroisleitner, C., Strozzi, T. 2021. ESA CCI+ Permafrost User Requirements Document, v2.0

[RD-2] 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>

[RD-3] GlobPermafrost team. 2016. Requirements Baseline Document. ESA DUE GlobPermafrost project. ZAMG, Vienna

[RD-4] Bartsch, A., Westermann, Strozzi, T., Wiesmann, A., Kroisleitner, C. 2019. ESA CCI+ Permafrost Product Specifications Document, v1.0

[RD-5] van Everdingen, Robert, ed. 1998 revised May 2005. Multi-language glossary of permafrost and related ground-ice terms. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. (<http://nsidc.org/fgdc/glossary/>; accessed 23.09.2009)

[RD-6] Bartsch, A., Widhalm, B., Pointner, G., Ermokhina, Ks., Leibman, M. and B. Heim (2019): DUE Globpermafrost Product documentation: Land cover prototype III – landcover classes https://download.pangaea.de/reference/98451/attachments/ESA_GlobPermafrost_PD_LCP_LANDC_20190128_v1.0.pdf

[RD-7] Bartsch, A., G. Hugelius, Strozzi, T.(2021): ESA CCI+ Permafrost CCN3 Option 6: improved soil description through a landcover map dedicated for the Arctic. User Requirements Document, v1.0

[RD-8] Bartsch, A., G. Hugelius, Strozzi, T.(2021): ESA CCI+ Permafrost CCN3 Option 6: improved soil description through a landcover map dedicated for the Arctic. Product Specification Document, v1.0

1.5 Bibliography

A complete bibliographic list that support arguments or statements made within the current document is provided in Section 6.1.

1.6 Acronyms

A list of acronyms is provided in section 6.2.

2 Design justification

Central to the design justification is the preparation of a benchmarking dataset, a revision of algorithms and comparison with respect to user requirements. Algorithms have been modified and the comparison to the benchmarking dataset has been completed.

The new landcover map is primarily developed for applications considered within the ESA Permafrost_cci project. This includes permafrost modelling for production of the climate data records as well as use cases targeted on improvements of Earth System Models. The user requirements discussion [RD-7] has been also extended to projects and groups using similar models outside of Permafrost_cci.

2.1 Methods and data

The Landcover prototypes (traditional landcover classification, vegetation height maps and surface roughness maps) developed from Sentinel-1 and Sentinel-2 in GlobPermafrost provide an alternative to CCI Landcover. It has been demonstrated in Bartsch et al. (2016) that C-band frozen backscatter (ENVISAT ASAR GM and WS) can be used as proxy for estimation of soil organic carbon in tundra regions (URQ_05). For tundra environments, this also coincides with specific landsurface wetness gradients (URQ_01) what has been initially shown for ENVISAT ASAR GM (1km) by Widhalm et al. (2015). The GM dataset has been previously applied for a permafrost equilibrium model soil parameterization (Obu et al. 2019) as well as for a recent estimation of the global methane budget (wetlands as input for landsurface modelling, Saunois et al. 2020). For improved spatial resolution, an implementation has been tested with Sentinel-1 for selected sites in GlobPermafrost. A simplified normalization scheme has been developed (Widhalm et al. 2018) in order to allow efficient pre-processing of the 10 to 40 m gridded data over large regions. This also enables integration of Sentinel-1 into traditional landcover classification as well as shrub height retrieval (Bartsch et al. 2020) (URQ_06). These landcover maps represent vegetation physiognomy as well as wetness levels associated with certain vegetation communities. The associated soil characteristics represent landscape types with differing sensitivity of frozen ground to temperature extremes (Bartsch et al. 2019, impact of hot summers on subsidence and active layer thickness). This scheme has been also already used for characterization of freeze/thaw (Bergstedt et al. 2020) as its detectability also depends on soil properties (water content).

In a first step a product comparison and assessment need to be made. This requires the preparation of a benchmarking dataset (soil in situ data) which can adequately describe the spatial variation of soil parameters of relevance for permafrost modelling. Such records have been compiled in the baseline project and more data are compiled through recent projects such as HORIZON2020 Nunataryuk. Soil organic carbon has been previously derived on the basis of such data using very high spatial resolution satellite data for selected sites through for example FP7 PAGE21 by the SU team (e.g. Siewert et al. 2016, building on Hugelius 2011).

Three types of in situ records are available: (1) full pedon descriptions for key regions, (2) organic layer thickness descriptions available from multiple sites globally, (3) shrub height records for several

tundra sites (assembled in Bartsch et al. 2020). The key regions will be in the following also referred to as primary sites. Some of the sites for (1) and (2) overlap (see Table 1).

The following in situ measurements are considered for the benchmarking:

- (1) Pedons (provided through SU)
 - Organic layer depth
- (2) Sites with comprehensive records (provided through SU)
 - Bare soil fraction
 - Lichen fraction
 - Graminoid fraction
 - Shrub height and fraction
 - Volumetric Water content (%) (preliminary)
 - Organic volumetric content (%)
- (3) Shrub height (provided through Ks. Ermokhina, documented in Bartsch et al., 2020)
 - Maximum height, or
 - Mean height

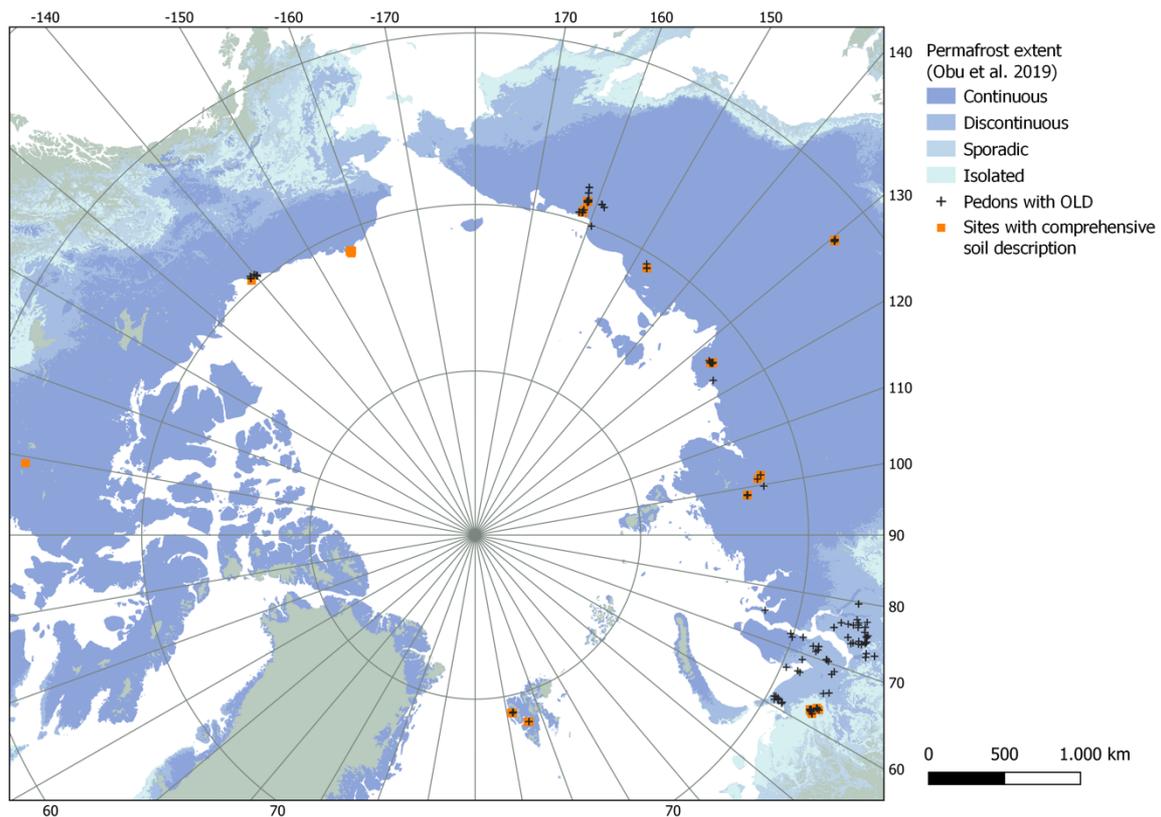


Figure 1: Sites with (1) comprehensive soil descriptions (key/primary sites) and (2) pedons with information on organic layer properties located within regions with for the benchmarking processed Semtinel-1/2 data.

Table 1: Overview of sites with soil descriptions (see also Figure 1)

Region	Country	S2 tile ID	Number of in-situ measurements	
			pedons with organic layer depth	sites with comprehensive soil and vegetation description
Barrow, Alaskan North Slope	USA	04WEE	141	0
Herschel Island, Yukon Territory	Canada	07WET	123	37
Tulemalu Lake, Nunavut Territory	Canada	14VMQ	18	0
Svalbard	Norway	33XVH, 33XWG	4	76
Yamal, Yamalo-Nenets Autonomous Okrug	Russia	41WMQ	153	0
Khatanga, Krasnoyarsk Region	Russia	47XNA, 47XMB	58	65
Lena River Delta	Russia	51XXA	54	56
Yakutsk, Sakha Republic	Russia	52VEQ	67	33
Tschokurdach, Sakha Republic	Russia	55WEU	39	27
Cherskiy, Sakha Republic	Russia	57WXT, 57WWS	51	57
sum			708	351

The availability of Sentinel-1 data is constrained as the polarization differs across the Arctic. The required data are unavailable for Greenland and the Canadian High Arctic (Figure 2). Only over Svalbard, both polarization combinations are existing. Separate cal/val needs to be considered before circumpolar implementation. A range of classes can be well characterized over Svalbard what will facilitate the trade-off option for these regions.

In order to address noise typical in SAR data, temporal averaging should be implemented. A minimum of three scenes is chosen. Acquisitions need to represent frozen conditions, but at maximum -10°C to minimize the effect of temperature on backscatter at C-band (Bergstedt et al. 2018). This requires the use of reanalyses data (ERA5) for scene selection.

Sentinel-2 availability is constrained by the frequent cloud cover in the Arctic. The growing season is short and the differentiation of the required landcover classes is based on peak vegetation season data. Due to various factors such as aerosols due to forest fires that can spread north, years with unusual precipitation/flooding conditions etc. several scenes should be combined. Although Sentinel-2 started data collection in 2016, the number of relevant cloud free scenes are often less than three (see Figure 3 for primary sites).

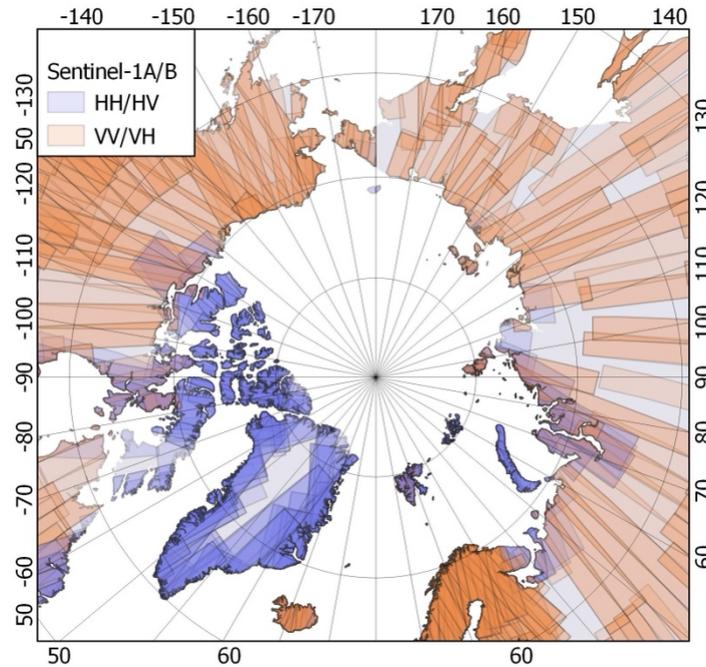


Figure 2: Example for Sentinel-1 coverage patterns across the Arctic land area. July 31 to August 20 2020 (figure source: Bartsch et al. (2021a); data source: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1>).

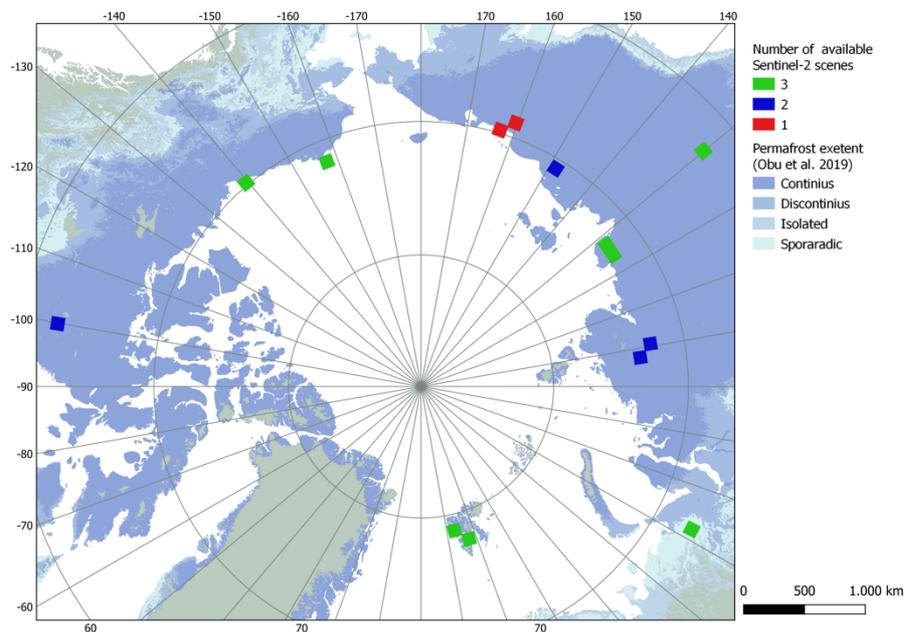


Figure 3: Number of available acquisitions since 2016 which match the requirements (cloud free, vegetation peak season, without flooding and increased aerosol content due to fires) for the primary benchmarking sites.

The prototypes [RD-6] have been developed outside of regions with mountainous terrain or with snow and ice cover. The new scheme therefore requires new classes for shadows and snow/ice. The class

‘disturbed’ needs to be split up, especially to separate fire scars (with soils with a certain organic layer) from inundation areas. The training datasets need to be updated accordingly. An update is also necessary as the prototype was based on data which were not corrected for atmospheric and illumination effects. Also, super-resolution processing (enabling use of 20 m Sentinel-2 bands at 10 m) was originally not applied what resulted in a 20m product.

2.2 Implementation options

The GlobPermafrost prototype implementation is based on Maximum Likelihood (MLH) for the classification (building on K-means in a first step) and Sentinel-1 and Sentinel-2 in combination as input. Considering data access, no other alternative for the input data allows to achieve pan-arctic coverage (URQ_08) with the required spatial resolution (URQ_03). Machine learning may provide an alternative for classification. Gradient boosting machines (XGBoost) as used in Bartsch et al. (2020) in tundra environments has been therefore tested as alternative. In order to address URQ_04 (separation of artificial landcover, roads, settlements) available targeted classification schemes which consider such objects and the constraints by the spatial resolution can be used. Deep learning has been shown applicable for this purpose in Arctic environments (Bartsch et al. 2021). Results exclude in this case other landcover classes (than infrastructure). A combination of two landcover layers, a pixel based derived landcover map (MLH or machine learning approach) and a Deep Learning infrastructure mapping for masking, is therefore required.

Figure 4 to Figure 14 present the comparisons between the in situ data for the original scheme which aimed at 21 classes (old approach MLH versus new approach XGBoost) and the new scheme which aims at 25 classes (old approach MLH versus new approach XGBoost). Class descriptions are available in Table 2 and Table 3.

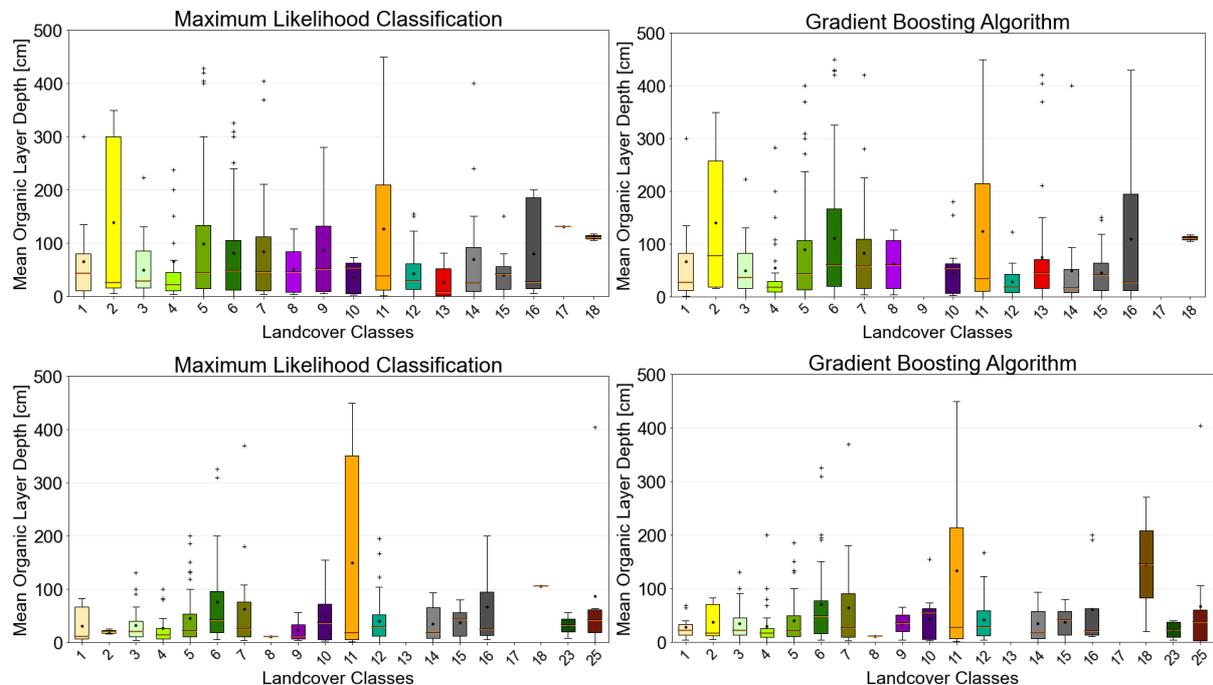


Figure 4: Mean organic layer depth (dataset #1) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

The MLH derived classes with highest organic layer depth are moist/wet/waterlogged with shrubs, grass/herb dominated meadows and vegetated floodplains (often lacustrine). This is well represented in both 25 class versions (Figure 4). XGBoost gives similar results except for a deviation for class 18 (barren). Organic volumetric content is highest for typical dwarf shrub tundra classes (dry to wet) (Figure 5). Volumetric water volume content measurements confirm the comparably dry conditions for ‘Dry cryptogamic-crust or sparse vegetation’ (Figure 14). The highest shrubs occur in the tundra – forest transition zone (forest classes, Figure 7). The classes with higher shrubs can be clearly separated. The number of data points in soil dataset #2 which contained vegetation description beyond shrub height are, however, very low and not all classes can be assessed (e.g. for class 7, 11 or 23). Therefore, two different data sources have been considered, one in situ (#3) and Sentinel-1 derived vegetation height following the approach by Bartsch et al. (2020). Both represent Yamal, spanning from high arctic to the tundra-taiga transition zone (Figure 7 and Figure 8). The moss fraction is highest for the wet to waterlogged shrub tundra class (Figure 11). The bare vegetation fraction measurements confirm the ‘sparse vegetation’ classes (Figure 13). No data are available in the datasets for the new class 13, which represents mostly recently burned areas.

The two tested schemes (MLH and Gradient Boosting) differ specifically for representation of the organic layer depth across the classes (Figure 4). The MLH25 version complies with most user requirements (Table 4) and is therefore chosen for the implementation.

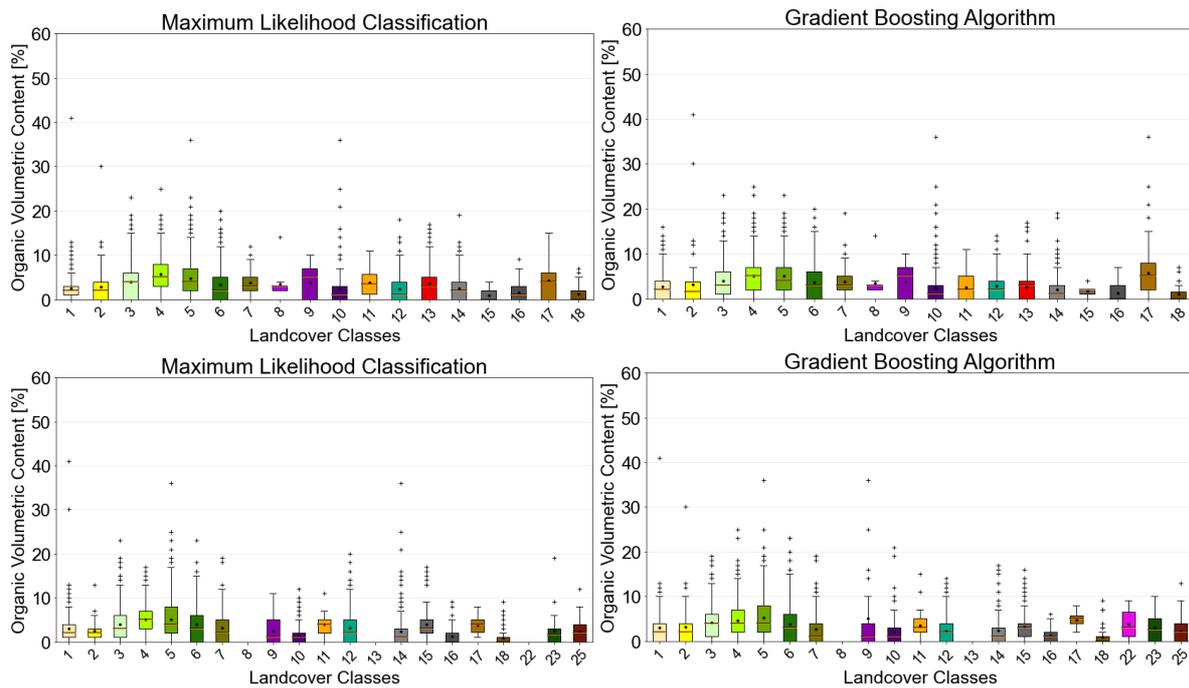


Figure 5: Organic volumetric content (dataset #2) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

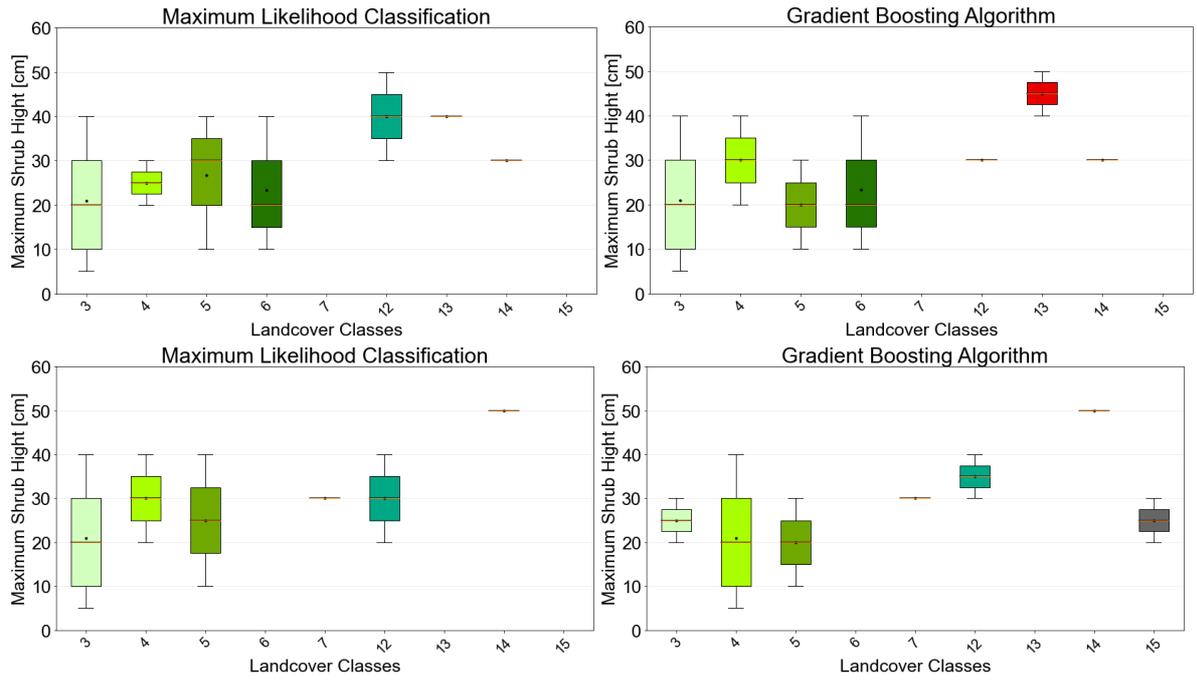


Figure 6: Shrub height (meta data of dataset #2) for classes of original (21, upper row) and new (25, lower row) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

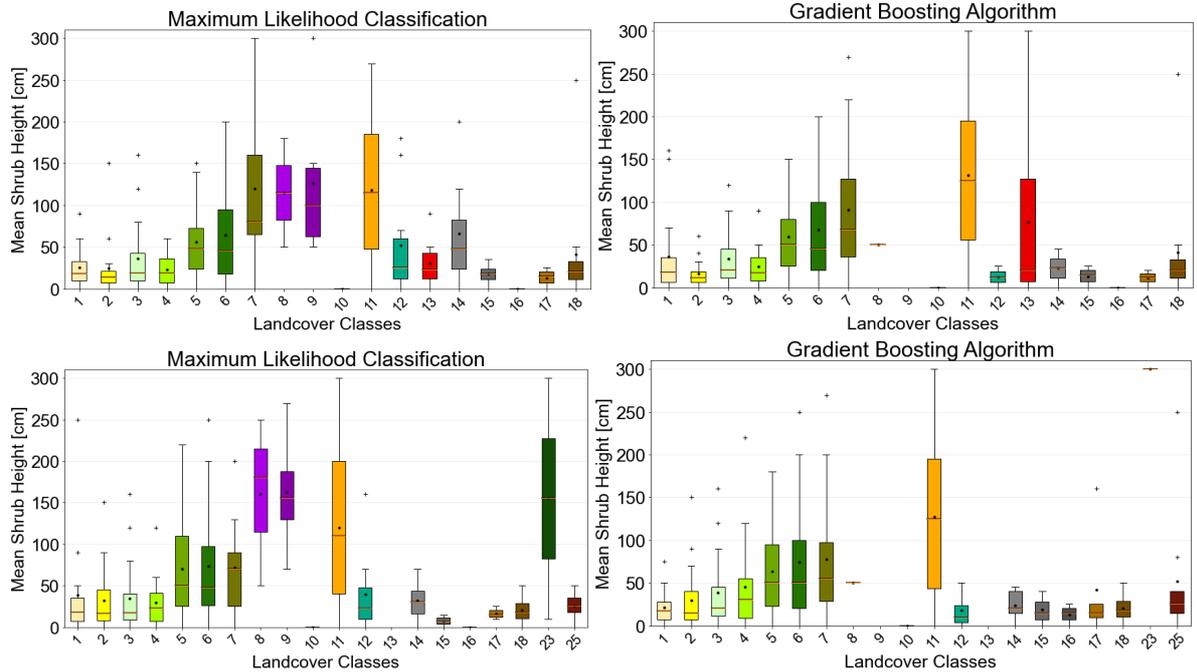


Figure 7: Shrub height (dataset #3) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

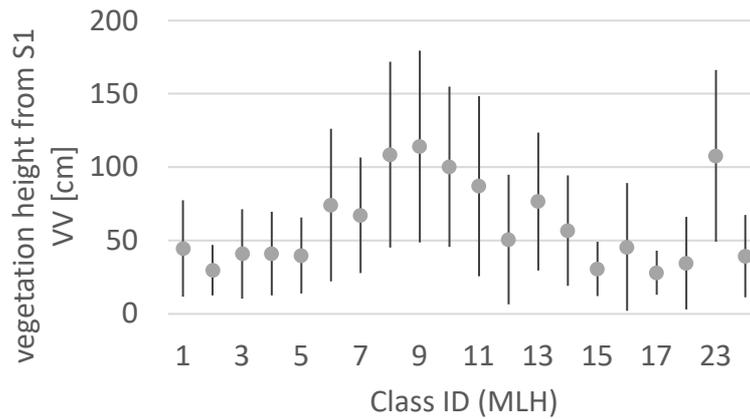


Figure 8: Vegetation height for areas with vegetation within MLH classes (new preliminary scheme). Vegetation height has been derived from Sentinel-1 VV based on the approach presented in Bartsch et al. (2020). For class descriptions see Table 3.

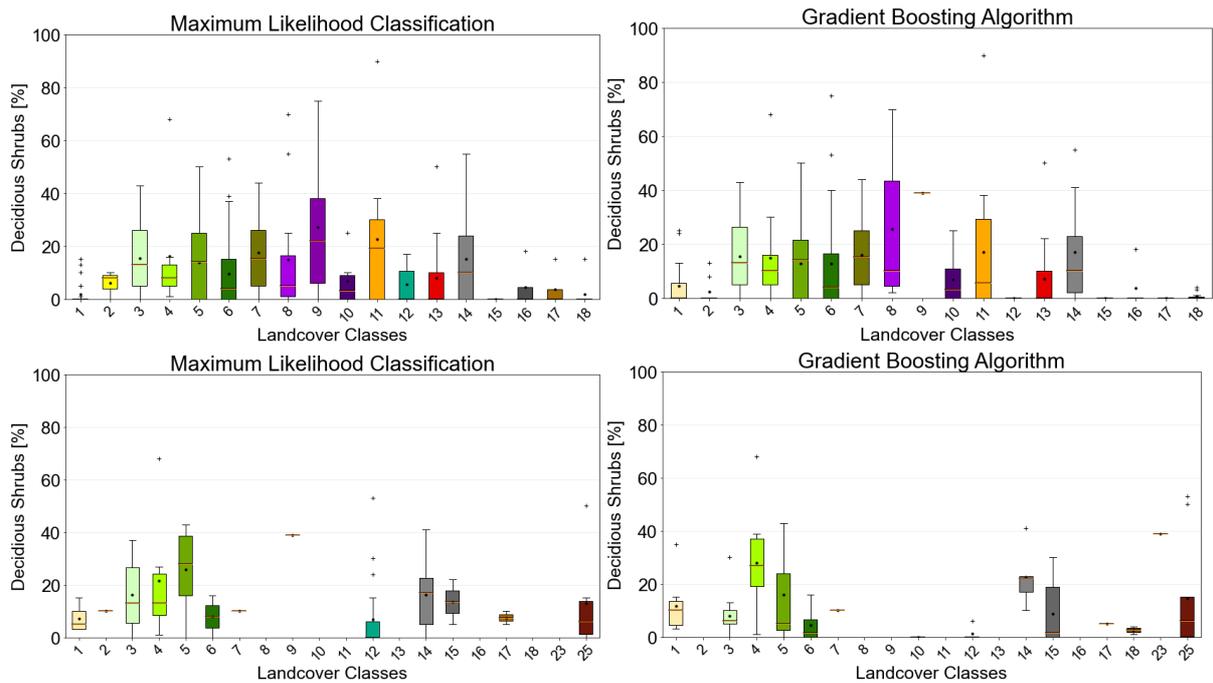


Figure 9: Deciduous shrub fraction (dataset #2) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

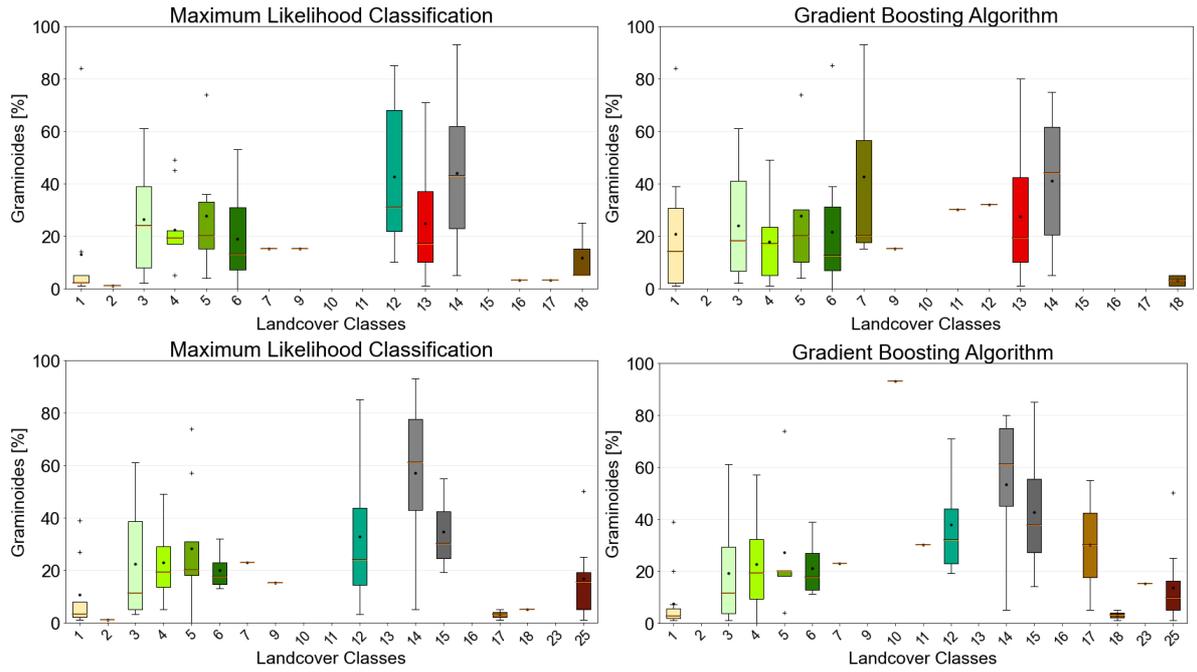


Figure 10: Graminoid fraction (dataset #2) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

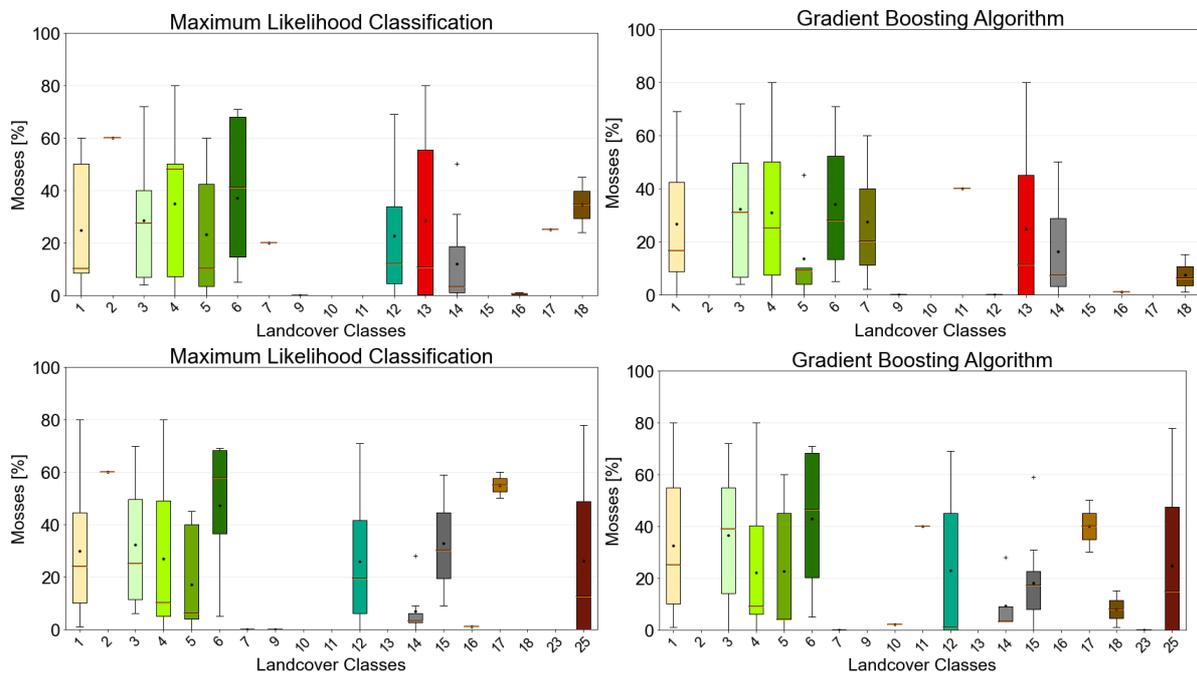


Figure 11: Moss fraction (dataset #2) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

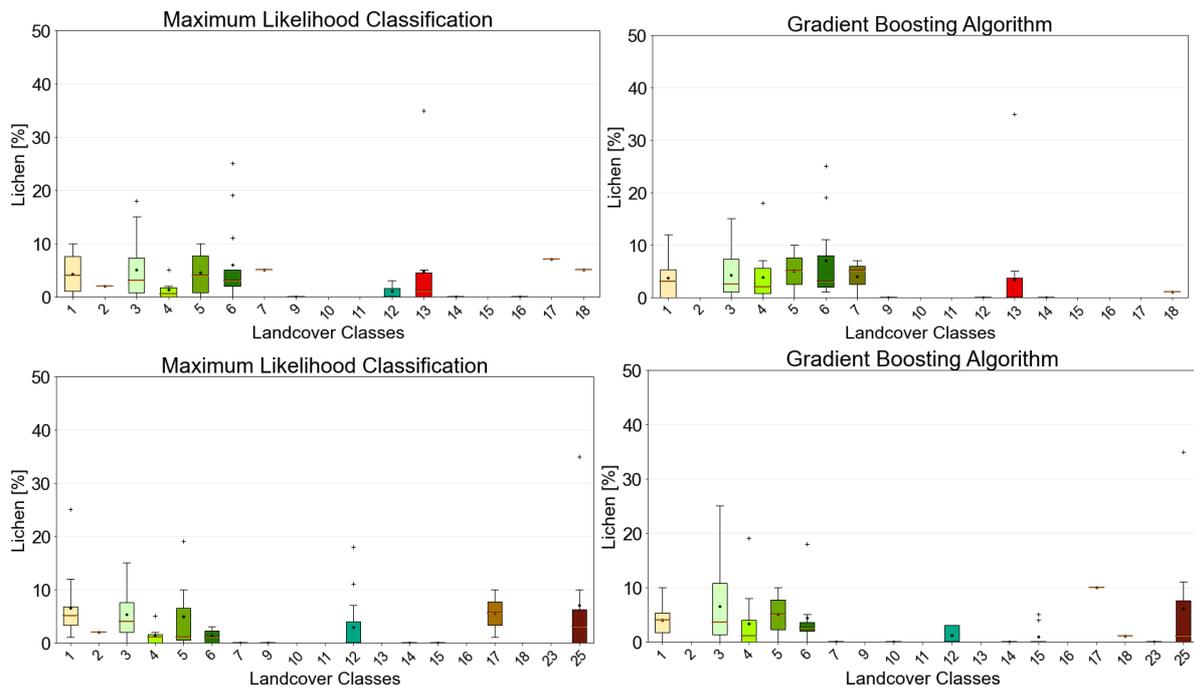


Figure 12: Lichen fraction (dataset #2) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

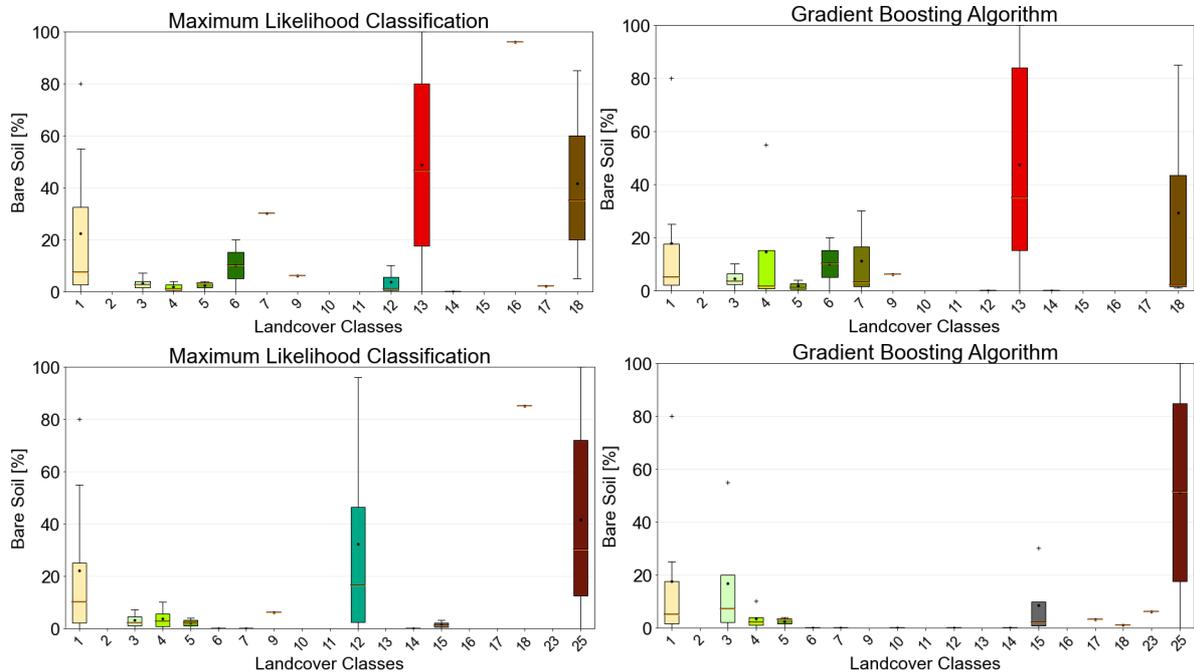


Figure 13: Bare soil fraction (dataset #2) for classes of original (21) and new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

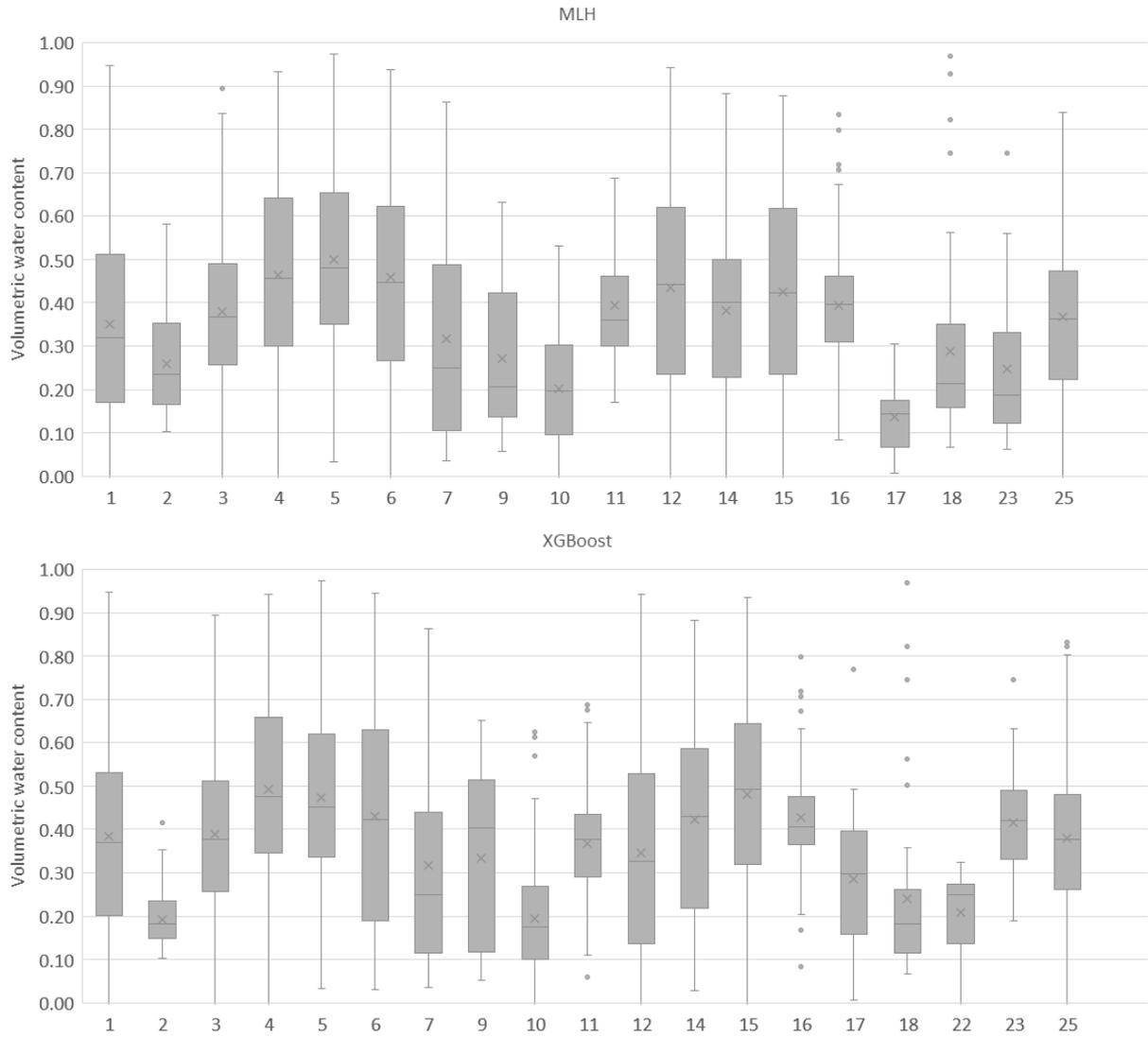


Figure 14: Preliminary comparison to Volumetric water content (dataset #2) for classes of the new (25) scheme, for two classification approaches. For class descriptions see Table 2 and Table 3.

Table 2: Data codes and class names of the GlobPermafrost prototype (Maximum Likelihood; source RD-6)

ID	Class name
1	Sparse vegetation (without shrubs), mostly sandy soil; flood plains, recent landslides, also within fire scars
2	Dry cryptogamic-crust or sparse vegetation
3	Graminoid, prostrate dwarf shrub, patterned ground, partially bare
4	Dry to moist prostrate to erect dwarf shrub tundra
5	Moist to wet graminoid prostrate to erect dwarf shrub tundra
6	Wet to waterlogged graminoid prostrate to low shrub tundra
7	Moist low dense shrubs
8	Tall shrubs, deciduous forest
9	Mixed forest
10	Coniferous (partially mixed) forest
11	Meadows, grass and herb-dominated
12	Wet ecotops, especially in floodplains
13	Disturbed, including forest fire scars, seasonally inundated areas and landslide scars
14	Floodplain, mostly fluvial
15	Floodplain, mostly lacustrine
16	Seasonally inundated
17	Barren, rare vegetation (petrophytes and psammophytes)
18	Barren, including artificial surfaces
19	Water (shallow or high sediment yield)
20	Water (medium depth or medium sediment yield)
21	Water (low sediment yield)

Table 3: Preliminary data codes and class names of the 25 class version, corresponding Landcover_cci classes, properties regarding organic layer depth (OLD) and shrub height

ID	Class name new	Landcover cci class name & Code	>40cm OLD	>40 cm shrubs
1	Sparse vegetation (without shrubs), mostly sandy soil; flood plains, recent landslides, also within fire scars	Lichen and mosses 140		
2	Dry cryptogamic-crust or sparse vegetation	Sparse vegetation 150		
3	Graminoid, prostrate dwarf shrub, patterned ground, partially bare	Grassland 130		
4	Dry to moist prostrate to erect dwarf shrub tundra	Grassland 130		
5	Moist to wet graminoid prostrate to erect dwarf shrub tundra	Grassland 130		x
6	Wet to waterlogged graminoid prostrate to low shrub tundra	Shrubland 120	x	x
7	Moist, dense low shrub tundra	Shrubland 120	x	x
8	Tall shrubs, deciduous forest	Tree cover, broadleaved, deciduous, closed to open 60		x
9	Mixed forest	Tree cover, mixed leave type 90		x
10	Coniferous (partially mixed) forest	Tree cover, needle leaved 70 & 80	no data	no data
11	Meadows, grass and herb-dominated, shrubs	Mosaic tree or shrub (<50%)/ herbacious cover 100	x	x
12	Wet ecotopes, shrubs, especially in floodplains	Shrub or herbacious cover, flooded 180		
13	Disturbed but vegetated, including specifically forest fire scars	Grassland 130	no data	no data
14	Graminoids, patches of erect dwarf shrubs, e.g. in inactive floodplains	Grassland 130		
15	Floodplain, mostly lacustrine, mosses, graminoids	Shrub or herbacious cover, flooded 180		
16	Seasonally inundated, sparse vegetation	Shrub or herbacious cover, flooded 180	x	
17	Barren, rare vegetation (petrophytes and psammophytes)	Sparse vegetation 150		
18	Barren, including artificial surfaces	Barren 190		
19	Water (shallow or high sediment yield)	Water bodies 210		
20	Water (medium depth or medium sediment yield)	Water bodies 210		
21	Water (low sediment yield)	Water bodies 210		
22	Snow, ice	Permanent snow and ice 220		
23	Moist low shrub tundra	Shrubland 120		x
24	Shadows	no data 0		
25	Partially barren due to disturbance, incl. wind-blown surfaces	Sparse vegetation 150		

Table 4: User requirements versus tested algorithms for landcover classification (Maximum Likelihood and Gradient Boosting approach, 21 versus 25 classes) and SACHI dataset (Bartsch et al. 2021). MLH21 corresponds to the GlobPermafrost prototype

ID	REQUIREMENTS	MLH21	XGB21	MLH25	XGB25	SACHI
URQ_01	Representation of dry, moist and wet	yes	yes	yes	yes	
URQ_02	Subcategories of prototype class ,disturbed'	no	no	yes	yes	
URQ_03	Spatial resolution 20 m	yes	yes	yes	yes	x
URQ_04	Separation of artificial landcover (roads, settlements)	partially	partially	partially	partially	x
URQ_05	Separation of peatlands, areas with > 40 cm organic layer	High class similarity	High class similarity	4 distinct classes	Barren confusion	
URQ_06	Separation of shrub tundra, higher than 40 m	yes	yes	yes	yes	
URQ_07	Compatible with Landcover cci	partially	partially	for none-forest	for none-forest	

3 Technical specifications

Product specifications including file naming are provided in the PSD [RD-07]. The technical specifications are summarized in Table 5.

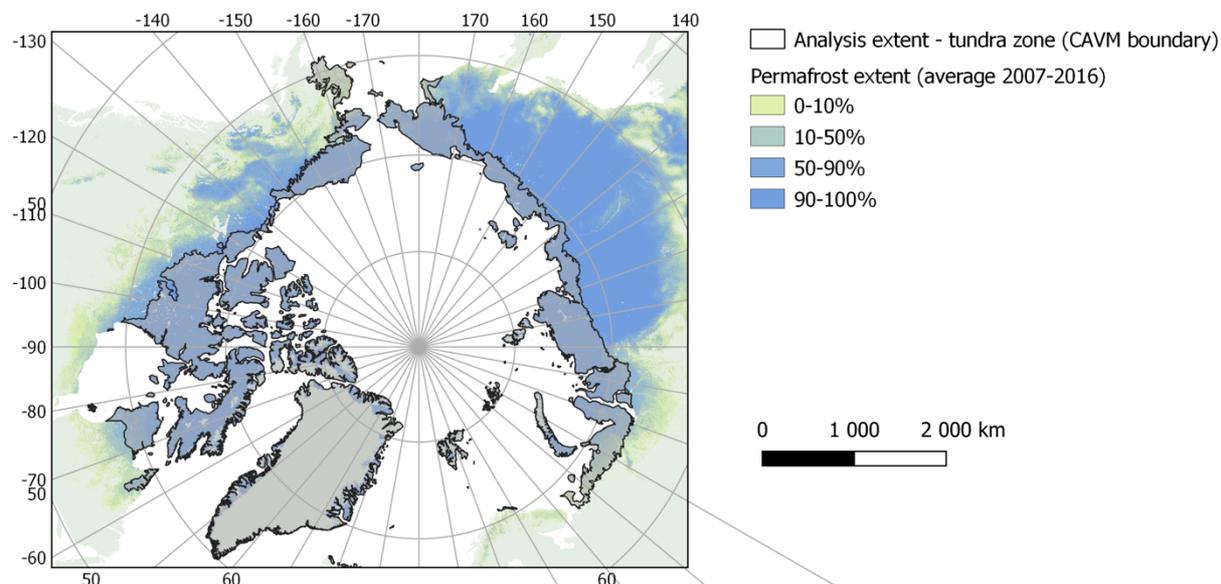


Figure 15: Product extent, confined through the tundra extent according to the CAVM (Raynolds et al. 2019)

Table 5: Preliminary specifications of the land cover product based on medium resolution satellite data

Subject	Specification
Variable	Land cover
Units	Discrete classes (see Table 3: Preliminary data codes and class names of the 25 class version, corresponding Landcover_cci classes, properties regarding organic layer depth (OLD) and shrub height)
Coverage	North of tree line where S1 C-VV data are available
Time period	From 2016
Temporal frequency	Static
Coordinate system	Polar stereographic
Spatial resolution (grid spacing)	20 m
Geometric accuracy	Subpixel
Thematic accuracy	Compatible with CCI land cover
Auxiliary information	On granule level: acquisition dates, number of scenes, quality flag
Data (file) format	Netcdf

4 Design definition

4.1 System components

Data selection:

Sentinel-2 data is available in granules of 100 x 100 km in UTM projection. To prevent errors due to undetected clouds the median of three acquisition dates are calculated for each granule. Because of often erroneous cloud-coverage metadata, the scenes have to be additionally manually inspected for cloud cover or haze. Scenes are selected within the timeframe of mid-July to mid-August. Due to frequent cloud cover, sporadic granules do not possess enough adequate acquisitions. In these cases, less than three acquisitions are used.

For Sentinel-1 data acquisitions under frozen soil conditions are used in order to exclude temporal backscatter variations due to changes in liquid water content. Since also strong negative temperatures show effects on backscatter values, they are omitted by selecting dates where temperature values were within the range of -1.5°C to -10°C (in some cases down to -15°C if not enough suitable dates were available). Each relative orbit is checked against reanalysis data and three scenes are subsequently selected.

Preprocessing:

For Sentinel-2, atmospheric effects and thereby connected differences between dates have to be mitigated. We therefore apply the Sen2Cor processor on Level 1C data. This process creates Bottom-of-Atmosphere Level 2A data. We further incorporate an optional terrain correction within this step using the Copernicus 90m resolution DEM (for Svalbard a DEM of the Norwegian Polar Institute is used, due to observed artefacts within the Sen2Cor process using the Copernicus DEM). The Sen2Cor process furthermore delivers masks for clouds (including thin cirrus clouds), cloud shadows and snow, which are later used for masking. Sentinel-2 offers data of 10 m spatial resolution for some bands. In order to enhance the spatial resolution of the coarser bands, super-resolution based on the tool Dsen2 is used, which uses a convolutional neural network. As described in Bartsch et al. (2020b), this approach by Lanaras et al. (2018) was adapted for Level-2 data and retrained for selected granules of the Arctic region. In a next step the median of three acquisition dates are calculated which further mitigates errors due to undetected clouds.

Processing steps of Sentinel-1 include border noise removal, based on the bidirectional all-samples method of Ali et al. (2018), calibration, thermal noise removal and orthorectification using the Copernicus 90m resolution DEM. Sentinel-1 data are normalized as described in Widhalm et al. (2018). They suggest a simplified method which is applicable to frozen tundra environments. A linear dependency is assumed with validity for an incidence angle range of approximately 20° to 45° . Normalization parameters have been published for HH, VV, VH and HV (Widhalm et al. (2018), Bartsch et al. (2020a), Bartsch et al. (2020b)). For this purpose a landcover map which includes a range of vegetation types reflecting vegetation physiognomy was used (Usa basin, Russia, Virtanen et al., 2004; see also Bartsch et al. (2016) and Widhalm et al. (2018), Bartsch (2020a)). Data from multiple orbits, representing a range of incidence angles for each location, were considered. A

function was fit to each landcover specific sample to describe the relationship of the local incidence angle with σ^0 . The relationship between the slope values k for all the different landcover classes and σ^0 at 30° (as frequently used in C-band SAR applications, e.g. Bartsch et al., 2008, 2009, 2012; Reschke et al., 2012; Trofaier et al., 2013) was derived in order to obtain the normalization function. After normalization, data has been reprojected and subset to match the Sentinel-2 granules and temporal averaging has been performed.

Classification:

The classification is built on the landcover prototype of Bartsch et al. (2019, <https://doi.org/10.1594/PANGAEA.897916>). For this prototype a k-means unsupervised classification was performed for a 100 km wide (1 Sentinel-2 granule) and 1400 km long transect of the Western Siberia region (along the 70 meridian, from 61 N to 74 N) incorporating the Sentinel-2 Level-1C data bands 3, 4, 8, 11 and 12 and the Sentinel-1 VV mean December backscatter data. The classes have been used as training areas for a supervised classification. The class name assignment is based on data from Western Siberia collected within the framework of ESA DUE GlobPermafrost in 2018 and the Austrian-Russian joint project COLD Yamal (Austrian Science Fund and Russian Foundation for Basic Research) in cooperation with RAS (Russian Academy of Science). Transferability has been assessed over the Lena Delta in cooperation with the Alfred Wegener Institute for Polar and Marine Research. As a result of the evaluation, further classes (and regions of interest) have been added to the training dataset in order to distinguish between forest classes and vegetated tundra floodplains of similar signatures.

In order to transfer this classification to the super-resolution Level 2A Bottom-of-Atmosphere dataset, new samples have been extracted for a subsequent maximum likelihood and XGBoost classification. Due to the vast number of samples, they have been selected in the following manner: a N-S strip of two adjacent granules has been used where samples have been further reduced by applying a negative buffer of 40 m to exclude pixels at the class margins. Of these samples only a maximum of 20000 random pixels was finally used. Due to occurring regions within the resulting landcover classification of coniferous forest (class 10) in areas north of the treeline these samples have been replaced by using only pixels of southern granules. Additionally, an additive class (class 23) has been defined comprised of samples of previously wrongly classified forest pixels of one granule above the treeline. For the 'disturbed' class 13 a possible subdivision was investigated by applying a k-means classification of these disturbed regions. The newly defined class 13 now comprises burned areas whilst the remaining 'disturbed' areas are now included in a new class 25. Two further classes have been added based on samples of one granule over Svalbard. The class 22 comprises pixels of snow and ice and the class 24 includes shadow areas.

Post processing:

The SACHI (Bartsch et al. 2021c) dataset will be used for masking of the classification results to indicate artificial landcover.

4.2 System requirements

Algorithm implementation requires:

- SNAP (Sentinel-1 preprocessing)
- Sen2Cor (atmospheric- and terrain correction of Sentinel-2 + mask generation)
- Dsen2 (super-resolution processing of Sentinel.-2 20 m bands)
- Python- selected modules:
 - o gdal
 - o snappy
 - o google.cloud
 - o tensorflow
 - o keras
 - o skimage
 - o rasterio
 - o scipy
 - o numpy
 - o pandas

For the topographic correction of Sentinel-2 data the Sen2Cor processor is designed to use the SRTM DEM. In order to utilize the Copernicus DEM, the Sen2Cor code need to be adapted to allow a topographic correction north of 60° N.

4.3 Trade-off analyses of implementation options

In cases, only one image can be used in case of unavailability of several cloud free (vegetation peak) Sentinel-2 scenes. This may result in higher uncertainties in the classification. Granule specific auxiliary data will be provided. It will comprise the number of scenes used as well as the date(s). This also needs to consider polarization issues in case of Greenland and the Canadian High Arctic. Uncertainties introduced through it will be quantified for Svalbard where VV as well as HH implementation is possible.

5 Summary

The Maximum Likelihood scheme with 25 target classes complies with most user requirements (Table 2) and is therefore chosen for the implementation. It needs to be combined with deep learning derived artificial landcover information in order to meet all user requirements eventually.

Data selection requires the consideration of phenology in case of Sentinel-2 and meteorologic conditions for Sentinel-1. Further constraints are imposed through polarization availability for Sentinel-1, cloud coverage, flooding patterns and forest fires which result in increased aerosol content in the air which cannot be fully accounted for with Sen2Cor in case of Sentinel-2. An adaption of Sen2Cor setup is also required regarding DEM usage as the target region extents north of 60°N.

6 References

6.1 Bibliography

- Ali, I.; Cao, S.; Naeimi, V.; Paulik, C.; Wagner, W.(2018): Methods to Remove the Border Noise From Sentinel-1 Synthetic Aperture Radar Data: Implications and Importance For Time-Series Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11, 777–786
- Bartsch, A., Widhalm, B., Kuhry, P., Hugelius, G., Palmtag, J., and Siewert, M. B. (2016): Can C-band synthetic aperture radar be used to estimate soil organic carbon storage in tundra?, *Biogeosciences*, 13, 5453–5470, <https://doi.org/10.5194/bg-13-5453-2016>, 2016.
- Bartsch, A.; Widhalm, B.; Leibman, M.; Ermokhina, K.; Kumpula, T.; Skarin, A.; Wilcox, E.J.; Jones, B.M.; Frost, G.V.; Höfler, A.; et al. (2020a): Feasibility of tundra vegetation height retrieval from Sentinel-1 and Sentinel-2 data. *Remote Sens. Environ.* , 237, 111515.
- Bartsch, A., Pointner, G., Ingeman-Nielsen, T., Lu, W. (2020b): Towards Circumpolar Mapping of Arctic Settlements and Infrastructure Based on Sentinel-1 and Sentinel-2. *Remote Sensing*, 12, 2368.
- Bartsch, A., G. Pointner, I. Nitze, A. Efimova, D. Jakober, S. Ley, E. Högström, G. Grosse, P. Schweitzer (2021a): Expanding infrastructure and growing anthropogenic impacts along Arctic coasts. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/ac3176>
- Bartsch, A.; Pointner, G.; Bergstedt, H.; Widhalm, B.; Wendleder, A.; Roth, A. Utility of Polarizations Available from Sentinel-1 for Tundra Mapping. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium*, Brussels, Belgium, 11–16 July 2021; IEEE: New York, NY, USA, 2021; p. 4.
- Bartsch, Annett, Pointner, Georg, & Nitze, Ingmar. (2021). Sentinel-1/2 derived Arctic Coastal Human Impact dataset (SACHI) (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4925911>
- Bergstedt, H.; Zwieback, S.; Bartsch, A.; Leibman, M.: Dependence of C-Band Backscatter on Ground Temperature, Air Temperature and Snow Depth in Arctic Permafrost Regions. *Remote Sens.* 2018, 10, 142. DOI: 10.3390/rs10010142
- Bergstedt H., A. Bartsch, A. Neureiter, A. Höfler, B. Widhalm, N. Pepin and J. Hjort (2020): Deriving a Frozen Area Fraction From Metop ASCAT Backscatter Based on Sentinel-1. *IEEE Transactions on Geoscience and Remote Sensing*, DOI:10.1109/TGRS.2020.2967364
- Hugelius, G.; Virtanen, T.; Kaverin, D.; Pastukhov, A.; Rivkin, F.; Marchenko, S.; Romanovsky, V.; Kuhry, P. High-resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic. *J. Geophys. Res.-Biogeosci.* 2011, 116, G03024

Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; Schindler, K. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS J. Photogramm. Remote Sens.* 2018, 146, 305–319

Obu, J., S. Westermann, A. Bartsch, N. Berdnikov, H.H. Christiansen, A. Dashtseren, R. Delaloye, B. Elberling, B. Etzelmüller, A. Kholodov, A. Khomutov, A. Kääb, M.O. Leibman, A.G. Lewkowicz, S.K. Panda, V. Romanovsky, R.G. Way, A. Westergaard-Nielsen, T. Wu, J. Yamkhin, D. Zou (2019). Northern Hemisphere permafrost map based on TTOP modelling for 2000-2016 at 1 km² scale. *Earth-Science Reviews*, Volume 193, Pages 299-316.

Reschke, J.; Bartsch, A.; Schlaffer, S.; Schepaschenko, D. Capability of C-band SAR for operational wetland monitoring at high latitudes. *Remote Sens.* 2012, 4, 2923–2943.

Raynolds, M.K., Walker, D.A., Balsler, A., Bay, C., Campbell, M., Cherosov, M.M., Daniëls, F.J., Eidesen, P.B., Ermokhina, K.A., Frost, G.V., Jedrzejek, B., Jorgenson, M.T., Kennedy, B.E., Kholod, S.S., Lavrinenko, I.A., Lavrinenko, O.V., Magnússon, B., Matveyeva, N.V., Metúsalemsson, S., Nilsen, L., Olthof, I., Pospelov, I.N., Pospelova, E.B., Pouliot, D., Razzhivin, V., Schaepman-Strub, G., Šibík, J., Telyatnikov, M.Y., Troeva, E., (2019): A raster version of the circumpolar arctic vegetation map (CAVM). *Remote Sens. Environ.* 232, 111297. <https://doi.org/10.1016/j.rse.2019.111297>.

Saunio, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton, J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J., Peng, C., Peng, S., Peters, G. P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf, G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: The Global Methane Budget 2000–2017, *Earth Syst. Sci. Data*, 12, 1561–1623, <https://doi.org/10.5194/essd-12-1561-2020>, 2020

Siewert, Matthias Benjamin; Hugelius, Gustaf; Heim, Birgit; Faucherre, Samuel (2016): Soil organic carbon storage and soil properties for 50 soil profiles in the Lena River Delta including land form description and map. Department of Physical Geography, University of Stockholm, PANGAEA, <https://doi.org/10.1594/PANGAEA.862961>, Supplement to: Siewert, MB et al. (2016): Landscape controls and vertical variability of soil organic carbon storage in permafrost-affected soils of the Lena River Delta. *CATENA*, 147, 725-741, <https://doi.org/10.1016/j.catena.2016.07.048>

Trofaier, A.M.; Bartsch, A.; Rees, W.G.; Leibman, M.O. Assessment of spring floods and surface water extent over the Yamalo-Nenets Autonomous District. *Environ. Res. Lett.* 2013, 8, 045026.

Virtanen, T.; Ek, M. The fragmented nature of tundra landscape. *Int. J. Appl. Earth Obs. Geoinf.* 2014, 27, 4–12

Widhalm, Barbara, Annett Bartsch, Birgit Heim (2015): A novel approach for the characterization of tundra wetland regions with C-band SAR satellite data. *International Journal of Remote Sensing* 11/2015; 36(22):5537-5556., DOI:10.1080/01431161.2015.1101505

Widhalm, B.; Bartsch, A.; Goler, R. Simplified Normalization of C-Band Synthetic Aperture Radar Data for Terrestrial Applications in High Latitude Environments. *Remote Sens.* 2018, 10, 551

6.2 Acronyms

CAVM	Circumarctic Vegetation Map
CCI	Climate Change Initiative
CCN	Contract Change Notice
CRS	Coordinate Reference System
DARD	Data Access Requirement Document
DDF	Design Definition File
DJ	Design Justification
DEM	Digital Elevation Model
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
GST	Ground Surface Temperature
GTOS	Global Terrestrial Observing System
IPA	International Permafrost Association
MAGT	Mean Annual Ground Temperature
MLH	Maximum Likelihood
MAGT	Mean Annual Ground Surface Temperature
NSIDC	National Snow and Ice Data Center
PSD	Product Specifications Document
RD	Reference Document
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
TS	Technical specifications
URD	Users Requirement Document
URQ	User Requirement
XGBoost	Gradient Boosting