Very High Spatial Resolution Soil Moisture Observation of Heterogeneous Subarctic Catchment Using Nonlocal Averaging and Multitemporal SAR Data

Manninen, Terhikki

2022

Manninen , T , Jääskeläinen , E , Lohila , A , Korkiakoski , M , Räsänen , A , Virtanen , T , þÿ M u h i , F , M a r t t i l a , H , A l a - A h o , P , M a r k o v a a r a - K o i v i s t o , M , L i w a t , R & Hänninen , P 2022 , ' Very High Spatial Resolution Soil Moisture Observation of Heterogeneous Subarctic Catchment Using Nonlocal Averaging and Multitemporal SAR Data ' , IEEE Transactions on Geoscience and Remote Sensing , vol. 60 , 4405317 , pp. 1-17 . https://doi.org/10.1109/TGRS.2021.3109695

http://hdl.handle.net/10138/341935
https://doi.org/10.1109/TGRS.2021.3109695

acceptedVersion

Very high spatial resolution soil moisture observation of heterogenous subarctic catchment using non-local averaging and multi-temporal SAR data

Terhikki Manninen, Emmihenna Jääskeläinen, Annalea Lohila, Mika Korkiakoski, Aleksi Räsänen, Tarmo Virtanen, Filip Muhić, Hannu Marttila, Pertti Ala-Aho, Mira Markovaara-Koivisto, Pauliina Liwata-Kenttälä, Raimo Sutinen and Pekka Hänninen

***Abstract*—A soil moisture estimation method was developed for Sentinel-1 Synthetic Aperture Radar (SAR) Ground Range Detected High resolution (GRDH) data to analyze moisture conditions in a gently undulating and heterogeneous subarctic area containing forests, wetlands, and open orographic tundra. In order to preserve the original 10 m pixel spacing, PIMSAR (pixel-based multi-temporal nonlocal averaging) non-local mean filtering was applied. It was guided by multi-temporal statistics of SAR images in the area. Gradient Boosted Trees (GBT) machine learning method was used for the soil moisture algorithm development. Discrete and continuous *in situ* soil moisture values were used for training and validation of the algorithm. For surface soil moisture, the root mean square error (RMSE) of the method was 6.5% and 8.8% for morning and evening images, respectively. The corresponding maximum errors were 34.1% and 33.8%. The pixelwise sensitivity to the training set and method choice was estimated as the variance of the soil moisture values derived using the algorithms for the three best methods with respect to the criteria: smallest maximum error, smallest RMSE value and highest coefficient of determination (R^2) value. It was on the average 6.3% with a standard deviation of 5.7%. Our approach successfully produced instantaneous high-resolution soil moisture estimates on daily basis for subarctic landscape, and can further be applied to various hydrological, biogeochemical and management purposes.**

***Index Terms*—Synthetic aperture radar, non-local mean filtering, temporal classification, soil moisture, machine learning.**

I. Introduction

SOIL moisture retrieval has been one of the main topics of synthetic aperture radar (SAR) research for decades [1] [2][3]. The emphasis has long been on bare soil [4][5][6], and agricultural land in general [7], but recently also uncultivated land types have been studied [8][9][10]. However, SAR data is less applied and tested in forested areas, and in mosaic landscapes of subarctic regions with forest, wetlands, and open tundra. The small penetration depth of the SAR signal in C-band in high vegetation has been a limiting factor for soil moisture estimation in forests [11]. High correlation between backscatter and soil moisture at catchment scale but reduced correlation at more detailed scales has been observed in several studies [12][13][3]. Current operational global products covering all kinds of land cover types utilizing Sentinel-1 SAR (S1) are provided in 1 km spatial resolution [2][11][14] and a new method for soil moisture retrieval in 500 m resolution has been presented recently [15]. The aim is at reaching the spatial resolution 0.1 km operationally [2] [11].

The soil moisture retrieval applying S1 is currently actively studied in diverse resolutions and ecosystem types [16][17][18] [19][20][21][14][22][7][15][23][24][25][26][27][28][29][30]. The methods used have a wide variation. Change detection algorithms are based on the backscattering coefficient

This work was financially supported by the Academy of Finland in the project UPFORMET (grant number 308511). The manuscript was submitted March 17, 2021.

T. Manninen, is with Finnish Meteorological Institute, Helsinki, FI-00101 Finland (e-mail: Terhikki.Manninen@fmi.fi).

E. Jääskeläinen, is with Finnish Meteorological Institute, Helsinki, FI-00101, Finland (e-mail: Emmihenna.Jaaskelainen@fmi.fi).

A. Lohila Climate System Research, Finnish Meteorological Institute, Helsinki, FI-00101 Finland and Institute for Atmospheric and Earch System Research (INAR), University of Helsinki, Helsinki, Finland (e-mail: Annalea.Lohila@fmi.fi).

M. Korkiakoski, Institute for Atmospheric and Earch System Research (INAR), University of Helsinki, Helsinki, Finland (e-mail: Mika.Korkiakoski@helsinki.fi)

A. Räsänen Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland. Current affiliation: Natural Resources Institute Finland (Luke), Oulu, Finland (e-mail: aleksi.rasanen@luke.fi)

T. Virtanen Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland (e-mail: Tarmo.Virtanan@helsinki.fi)

F. Muhić, Water, Energy and Environmental Engineering Research Unit, University of Oulu, Oulu, Finland (e-mail: Filip.Muhic@oulu.fi).

H. Marttila, Water, Energy and Environmental Engineering Research Unit, University of Oulu, Oulu, Finland (e-mail: Hannu.Marttila@oulu.fi).

P. Ala-Aho, Water, Energy and Environmental Engineering Research Unit, University of Oulu, Oulu, Finland (e-mail: Pertti.Ala-Aho@oulu.fi).

M. Markovaara-Koivisto, Energy and Construction Solutions, Geological Survey of Finland, Espoo, Finland (e-mail: Mira.Markovaara-Koivisto@gtk.fi).

P. Liwata-Kenttälä , Environmental Solutions, Geological Survey of Finland, Rovaniemi, Finland (e-mail: Pauliina.Liwata-Kenttala@gtk.fi)

R. Sutinen, Energy and Construction Solutions, Geological Survey of Finland, Rovaniemi, Finland (e-mail: Raimo.Sutinen@gtk.fi)

P. Hänninen, Water Management Solutions, Geological Survey of Finland, Espoo, Finland (e-mail: Pekka.Hanninen@elisanet.fi)

dynamical variation between wet and dry conditions [9][14][15][17][19]. Inversion of bare soil and (when needed) vegetation backscattering models is another approach [6][20][21][26]. Polarimetry is also used [22]. However, the retrieval of soil parameters from radar measurements is an ill-posed problem because more than one combination of soil moisture, soil roughness and vegetation properties can lead to the same electromagnetic response [18]. Hence, the use of diverse machine learning methods is becoming more common in soil moisture retrieval [1][18][27][28][29][30].

Thermal infrared temperature has also shown to be useful for forested area soil moisture estimation using Artificial Neural Networks (ANN) [24]. Optical image-based normalized difference vegetation index, which is related to leaf area index (*LAI*) for low vegetation, has been shown to improve the soil moisture estimation accuracy [8][25], either applying the water cloud model or support vector regression (SVR). In the presence of high vegetation, optical imagery and laser scanner data have been used for analyzing the contribution of vegetation to the backscattered signal [9]. *LAI* retrieval of the canopy has been shown to be successful using the backscattering ratio of vertical and horizontal polarizations VV/HH [31]. Unfortunately, S1 does not provide this ratio. Instead, the backscattering coefficients of VV and VH polarizations are simultaneously available in S1 data.

The main goal of this study is to obtain spatially and temporally very high resolution (10 m pixel spacing, instantaneous daily) soil moisture maps to scale field-based soil methane (CH_4) flux measurements to a larger area. The area of interest is spatially heterogenous in terms of land cover and topography. There is a high need to better understand the role of very different surface types in landscape-scale methane budget and especially the role of the upland soils in CH_4 budget [32]. It is known that in wet enough conditions upland soils can act as CH_4 sources, but the interesting question is what kind of spatial and temporal variation in CH_4 exchange can be expected. Within a single catchment there can be surfaces or patches with high CH_4 emission and those with an order of magnitude lower CH_4 sink, controlled mainly by soil moisture [33]. Hence, normal spatial averaging or Lee filtering [8] to reduce speckle is not applicable. Previously, we have studied midsummer soil moisture patterns and related methane fluxes with a combination of SAR, optical and topographic data for this region, [34]. However, spatiotemporal changes in soil moisture were not assessed. In this paper, soil moisture maps including temporal variation are produced for the larger study area, and these maps will later be used in hydrological and dynamic methane models developed for the region.

Non-local despeckling can be used to develop high spatial and temporal resolution soil moisture maps. It has been a topic of active research since the first introduction of the principle of non-local averaging since 2009 [35]. The main challenge of the method is, how to choose the pixels to be averaged non-locally [36][37]. One possibility is to leverage information drawn from co-registered optical imagery [38]. This is not always possible, when the focus is on analyzing quick changes in the environment, as the area of interest may be cloudy, or the sun elevation is too low. Recently, multitemporal non-local despeckling methods have been developed [39][40], but the aim has been more in analyzing homogeneous targets, such as houses and cultivated fields. In this study, the non-local averaging is guided using SAR data only. The new PIMSAR method [41] combines slight spatial averaging (ground range multilook pixels) and temporal averaging with unsupervised classification to derive the guidance for non-local averaging.

In this study, the reduced simple ratio (RSR)[42][43] derived from Sentinel-2 (S2) is used as teaching material for machine learning methods to be developed for VV and VH backscattering for canopy effective *LAI* estimation. Further on, machine learning methods are used also for soil moisture retrieval using those S1 based effective *LAI* estimates and *in situ* soil moisture measurements. The advantage of deriving the soil moisture estimation method to be applicable also without optical data is that soil moisture estimation does not then suffer from cloudiness, cloud shadows when the solar zenith angle is large, or smoke plumes of forest fires.

The novelty of this study is that the soil moisture estimation method developed for S1 data without need of ancillary optical data is applicable also to forests in a gently undulating area including wetlands and that the resulting soil moisture map pixel spacing is as high as 10 m, the spatial resolution being of the order of 20 m. In the future it might be possible to produce instantaneous soil moisture maps for heterogeneous forested scenery every day in such high spatial resolution, which would open completely new possibilities for environmental and forestry applications for S1 data usage.

II. Data

A. Test area

The test area is Pallas in northern Finland, the corner co-ordinates being 67.9122° N, 24.0586° E and 68.0233° N, 24.2539° E [Fig. 1][34]. The terrain is partly hilly, the altitude varying in the range 260 m – 610 m above sea level. The diffuse tree line transition zone is at about 475–505 m a.s.l above which tundra starts [44]. The Norway spruce tree line has advanced there during the last 100 years [44]. The long-term (1981–2010) annual temperature and precipitation within the area are 1.0°C and 521 mm, respectively [45]. In spring, the soil moisture maximum in forest precedes the disappearance of snow by one month, but in tree line it is simultaneous with it [46][47]. National Land Survey of Finland provided a digital elevation model (DEM) and orthorectified aerial images with a spatial resolution of 2 m and 0.5 m, respectively. In addition, a cloud-free Sentinel-2 (S2) optical image was available on July 28, 2019. The nearest weather station was at Kenttärova (67.99°N, 24.24°E), where the intensity of precipitation and relative air humidity were available at 10 minutes interval.

B. SAR data

Sentinel-1 SAR Interferometric Wide (IW) Swath data available in ground range format (GRDH) in the time range May 15 – September 12, 2019, were used for this study, altogether 120 images, the only missing day being July 2. The data is delivered in 10 m pixel spacing, but the spatial resolution

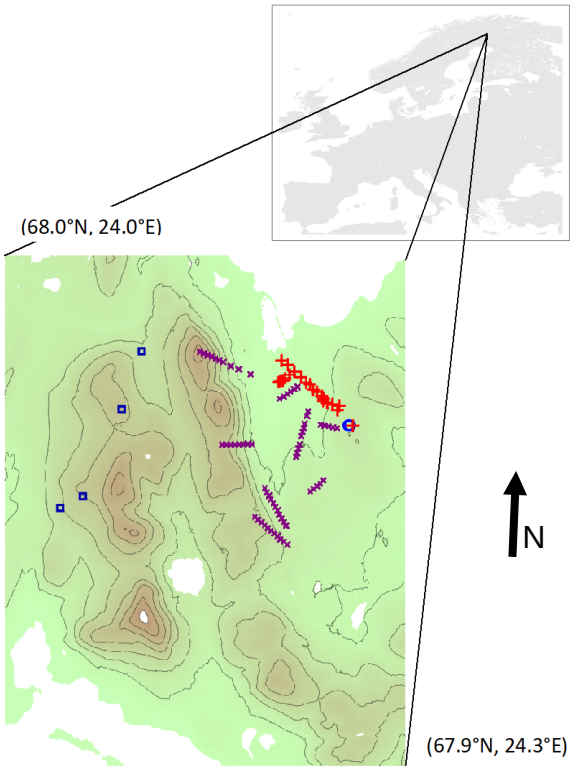

Fig. 1. Digital elevation model of the subarctic Pallas test area in northern Finland. The contour interval is 50 m. Open water is masked white. The *in situ* soil moisture measurement points are shown as well: red pluses for the time series points, purple crosses for the intensive campaign points, blue circle for continuous soil moisture measurements at 6 cm depth and dark blue squares for continuous electric measurements at 20 cm depth. The area is roughly 12 km x 9 km.

is 20 m x 22 m. The data was divided into 12 subsets according to polarization, vertical (VV) or cross (VH), swath (IW1, IW2 or IW3) and pass (ascending or descending, i.e. evening or morning, respectively). No discrimination between satellites S1-A and S1-B was made. The incidence angle ranges for the swaths in the test area are given in Table I. The preprocessing steps of all GRDH images were: 1) applying precise orbit files, 2) radiometric calibration, nearest neighbor terrain correction applying the DEM and 3) extracting the subset covering the test area.

TABLE I
INCIDENCE ANGLE RANGE IN THE STUDIED AREA

Swath	Pass	Minimum incidence angle	Maximum incidence angle
IW1	Morning	33.8°	34.4°
	Evening	33.2°	33.8°
IW2	Morning	39.1°	39.6°
	Evening	38.6°	39.0°
IW3	Morning	43.8°	44.3°
	Evening	43.4°	43.8°

C. *Tree and field layer data*

Biomass and *LAI* of Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*) and Downy birch (*Betula pubescens*) and other deciduous species were available for 131 plots in the test area [34]. The *LAI* values were calculated using measured tree height and diameter values [48][49]. The total biomass and *LAI* values for coniferous and deciduous species and all trees were determined as well [Table II]. In addition, biomass and *LAI* of the ground and field layer vegetation were determined [Table III].

The coefficient of determination for the correlation between canopy biomass and *LAI* was 0.85, but it was dominated by the smallest *LAI* values. When only *LAI* values exceeding unity were taken into account, the R^2 values was only 0.74. The biomass and *LAI* values of that vegetation did not manifest any correlation.

TABLE II
FOREST PARAMETER STATISTICS OF 131 PLOTS[34]

Parameter	Species	Median	Mean	Maximum
Biomass [kg/m^2]	Spruce	0.86	2.9	24.0
	Birch	0	1.0	10.1
	Pine	0	1.0	14.7
	Other deciduous tree species	0	0.09	4.8
	Coniferous	1.9	3.9	24.0
	Deciduous	0.03	1.1	10.1
	Total tree	3.4	5.1	24.0
LAI	Spruce	0.88	1.8	11.7
	Birch	0	0.6	6.9
	Pine	0	0.4	4.5
	Other deciduous tree species	0	0.1	2.6
	Coniferous	1.38	2.1	11.7
	Deciduous	0.04	0.7	6.9
	Total tree	2.2	2.8	11.8

TABLE III
UNDERGROWTH VEGETATION STATISTICS OF 131 PLOTS [34]

Parameter	Species	Median	Mean	Maximum
Biomass [g/m^2]	Evergreen shrub	39.7	57.9	212
	Deciduous shrub	61.1	65.3	211
	Forb	1.0	11.3	151
	Graminoid	16.6	23.1	103
	Moss	490	438	610
	Ground and field total	633	595	781
LAI	Evergreen shrub	0.17	0.23	0.81
	Deciduous shrub	0.37	0.40	1.4
	Forb	0.03	0.18	2.4
	Graminoid	0.17	0.27	1.4
	Field layer total	1.0	1.2	3.4

D. *Soil moisture data*

Discrete volumetric soil moisture values were measured approximately every two weeks in several places at the surface and 10 cm, 20 cm, 30 cm, and 40 cm below the surface using WET-2 and PR2 Profile Probe sensors with an HH2 readout unit (Delta-T Devices Ltd). [34]. In addition, the water level depth was measured bi-weekly as well. The value of each point was based on three individual measurements within an area having a radius of 5 m, so that the three retrievals were typically within one pixel. Hence, the mean values of the triplets were used in the method retrieval. The discrete soil moisture time series was measured during May 28, 2019 – September 11, 2019, in 23 points [Table IV, Fig. 1, Fig. 2]. The correlation between the surface soil moisture and the soil moisture measured at various depths decreased with increasing depth, the coefficients of determination (R^2) being 0.61 (at 10 cm), 0.40 (at 20 cm), 0.32 (at 30 cm) and 0.26 (at 40 cm). The surface soil moisture was negatively correlated with the water level

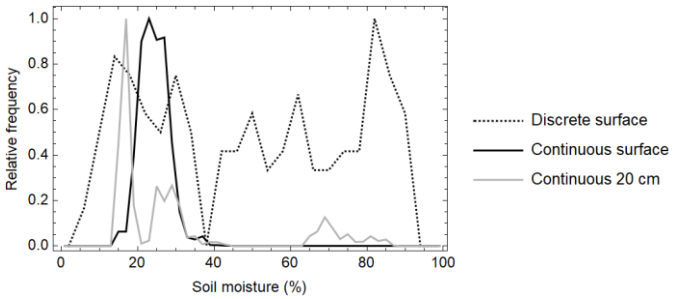

Fig. 2. Relative distributions of the soil moisture measured at the surface at discrete times and continuously and at 20 cm depth continuously.

depth, the R^2 value being 0.54. Besides the time series, surface soil moisture was measured in 64 additional points with WET-2 sensor, Delta-T Devices Ltd, Cambridge, UK, during the intensive campaign carried out on July 3, 2019 – July 13, 2019 [34]. Hence, the intensive campaign provides more information about the spatial variation of the surface soil moisture and the times series about the temporal variation.

Continuous soil moisture measurement data at 6 cm depth using soil scout sensors (Soil Scout Oy) were available in the test area with an interval of 15 – 19 min [Table V, Fig. 2]. These values were taken to directly represent surface soil moisture.

In addition, continuous soil moisture estimates at depths 20 cm, 40 cm and 60 cm based on electric measurement (EM) using the CS616 and CS650 Soil moisture probes by Campbell Scientific [50] were available in the test area with an interval of 30 minutes [Table VI, Fig. 2]. Values up to 100% were used in the machine learning training to cover the whole dynamic range, but only values smaller than 100% were included in the later analysis to remove actual liquid water cases from the soil moisture data. As the continuous measurements do not provide the soil moisture at the surface, a linear regression relationship ($sm_0 = 10.4 \cdot \exp(0.0254 \cdot sm_{20})$ was derived between measured soil moisture values at the surface sm_0 and at 20 cm depth sm_{20} for the 425 points available simultaneously in the discrete data.

TABLE IV
DISCRETE *IN SITU* SOIL MOISTURE [%] AND WATER LEVEL [CM] STATISTICS. THE NUMBER OF INDIVIDUAL MEASURED VALUES IS *n*.

Data set	Parameter	*n*	Minimum	Mean	Maximum
Time series	Surface	678	0.6	49.3	92.3
	10 cm depth	425	0.7	32.5	62.5
	20 cm depth	425	8	46.1	74.4
	30 cm depth	425	9	50.0	84.5
	40 cm depth	424	20.2	54.7	92.8
	Water level	354	0	18.5	71
Intensive campaign	Surface	195	1.3	26.7	91.6

TABLE V
CONTINUOUS SURFACE (6 CM DEPTH) SOIL MOISTURE [%] STATISTICS. THE NUMBER OF INDIVIDUAL MEASURED VALUES IS *n*.

Point	Latitude	Longitude	*n*	Min	Mean	Max
S3	67.98703°N	24.24193°E	7838	19	25.3	32
S13	67.98707°N	24.24238°E	7869	15	25.4	56
S17	67.98704°N	24.24193°E	7777	19	21.9	44
S19	67.98705°N	24.24238°E	7880	19	23.0	37

TABLE VI
CONTINUOUS SOIL MOISTURE [%] STATISTICS FROM ELECTRIC MEASUREMENTS. THE NUMBER OF INDIVIDUAL MEASURED VALUES IS *n*.

Point	Latitude	Longitude	*n*	Depth [cm]	Min	Mean	Max
GC4	67.970°N	24.095°E	5808	20	14.1	19,6	42.9
				40	16.3	20.8	36.7
				60	17.6	21.4	35.5
GC7	67.967°N	24.083°E	5808	20	13.1	20.4	55.0
				40	17.5	24.2	12.7
				60	13.4	17.6	31.2
GC8	67.988°N	24.115°E	5808	20	32.8	39.7	53.6
				40	9.7	12.7	22.1
				60	7.2	9.0	14.1
Well	68.001°N	24.124°E	3874	20	71.4	79.1	100
				40	78.7	84.8	100
				60	54.1	56.6	64.1

The relationship is rather vague ($R^2 = 0.34$ and the root mean square error is 8.3%), but it was used to convert the EM-based soil moisture data set from 20 cm depth to the surface, in order to get more extensive spatial coverage and wider land cover type range for the continuous surface soil moisture data set to be used for deriving the estimation algorithm.

E. Sensitivity of backscattering to soil moisture and LAI_{eff}

In principle the backscattering coefficient increases with increasing soil moisture and LAI_{eff}. However, in the hilly terrain of the study area the backscattered intensity is strongly related also to the local incidence and azimuth angles. Thus, the relationship between the soil moisture, LAI_{eff} and backscattering coefficient is not so simple [Fig. 3]. Although the high soil moisture values are common at the high end of a certain LAI_{eff} value, also opposite behavior appears. Hence, in an area like this a simple algorithm for soil moisture retrieval is not to be found.

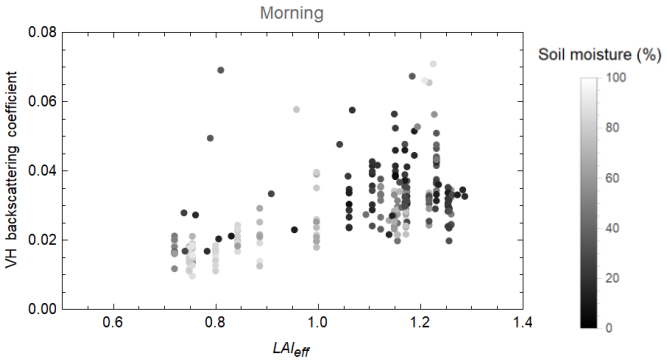

Fig. 3. The VH backscattering coefficient vs. LAI_{eff} for the morning images and the discrete soil moisture measurement points. The soil moisture is indicated by the darkness of the points.

III. METHODS

To preserve the high spatial resolution of the GRDH SAR images and cope with the speckle, a recently developed non-local averaging method PIMSAR [41] was applied to each image. For that purpose, the images were first divided into 30 classes based on multitemporal mean and standard deviation values. In each class, the backscattering neighbor locations were sought by sorting the multitemporal mean values of the

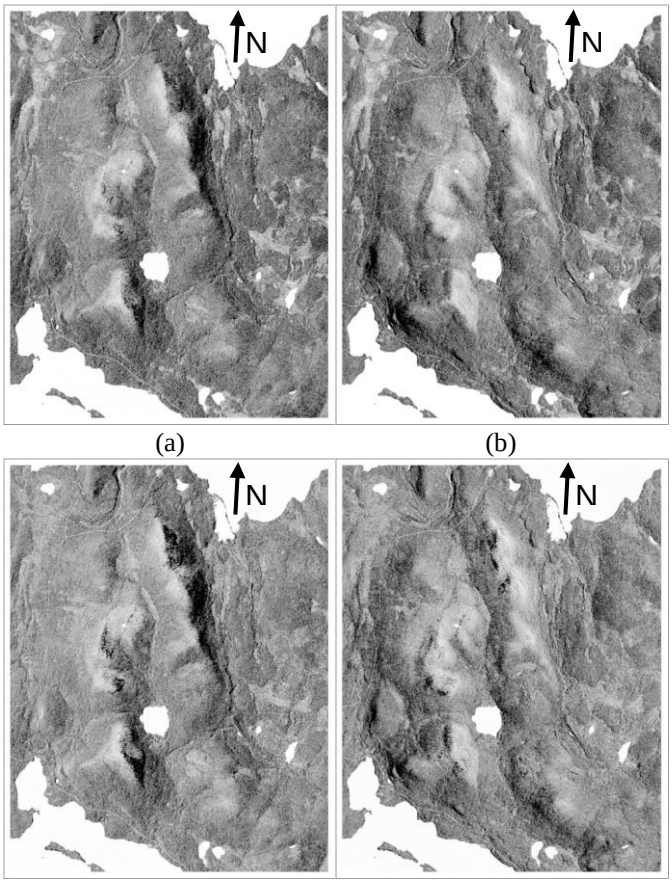

Fig. 4. The non-locally averaged VH backscattering coefficient of the S1 GRDH images of June 21 (a) and June 22 (b), 2019. The swath is IW3 and the passes are descending (morning) and ascending (evening), respectively. The magnitude increases from white to black to show forests dark and open areas bright. Open water is masked white. The corresponding images for VV polarization are shown in (c) and (d). The area is roughly 12 km x 9 km.

class. Indices of 49 pixels closest to each other in the multitemporal mean were stored for each pixel in a guidance matrix per pass and swath. The same guidance matrix was used for all images of the same pass and swath to derive the non-locally filtered images. PIMSAR preserves the image mean value with the high accuracy (better than 0.02% for each image). Details of the PIMSAR method are described in [41].

Examples of non-locally averaged morning and evening VH and VV images are shown in Fig. 4. The hilly topography causes obvious shading in different parts in the morning and evening images, so the soil moisture algorithm has to be derived for morning and evening images separately.

Examples of subareas of original and PIMSAR filtered images are shown in Fig. 5 and Fig. 6 for the highest and lowest mean image intensity cases for VH and VV polarizations. Notably, the highest intensity occurs on the same day (September 12), whereas the lowest intensity appears in the VH images on August 7 and in the VV images on May 15.

PIMSAR is able to produce a high spatial structure, which is completely blurred in the original images and their spatial averages. Even a narrow forest road (roughly 3 m wide without nourishings) is visible although the pixel spacing is 10 m.

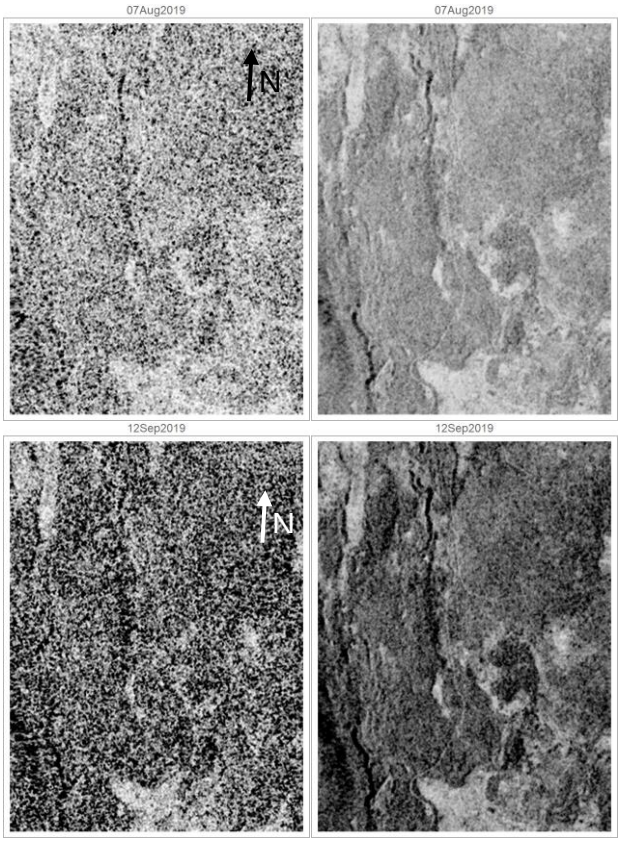

Fig. 5. Subarea of the S1 VH image for the lowest (top) and highest (bottom) image mean backscattering values. The original images are to the left and the non-locally filtered images to the right. The area covered is 4 km x 3 km.

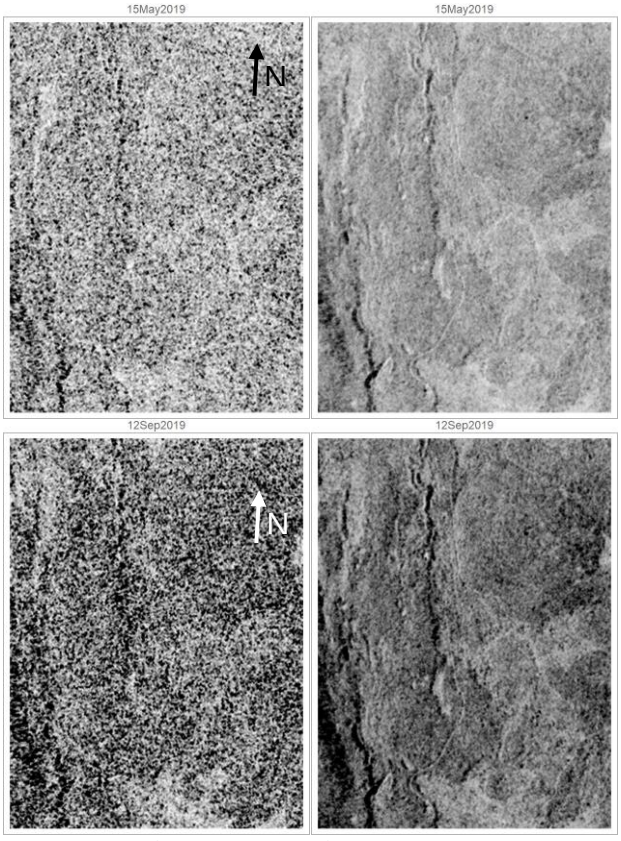

Fig. 6. Subarea of the S1 VV image for the lowest (top) and highest (bottom) image mean backscattering values. The original images are to the left and the non-locally filtered images to the right. The area covered is 4 km x 3 km.

The relationship between backscattering, leaf area index and soil moisture is briefly described in Section IIIII.EA using a water cloud with gaps approach. The effective *LAI* (LAI_{eff}) map based on the SAR images was derived using the random forest [51][52] machine learning method trained with the LAI_{eff} map obtained using the RSR index derived from the S2 image (Section IIIB). Then the relationship between soil moisture and backscattering was obtained using machine learning methods (Section II.EC) trained with *in situ* soil moisture measurements [Fig. 7].

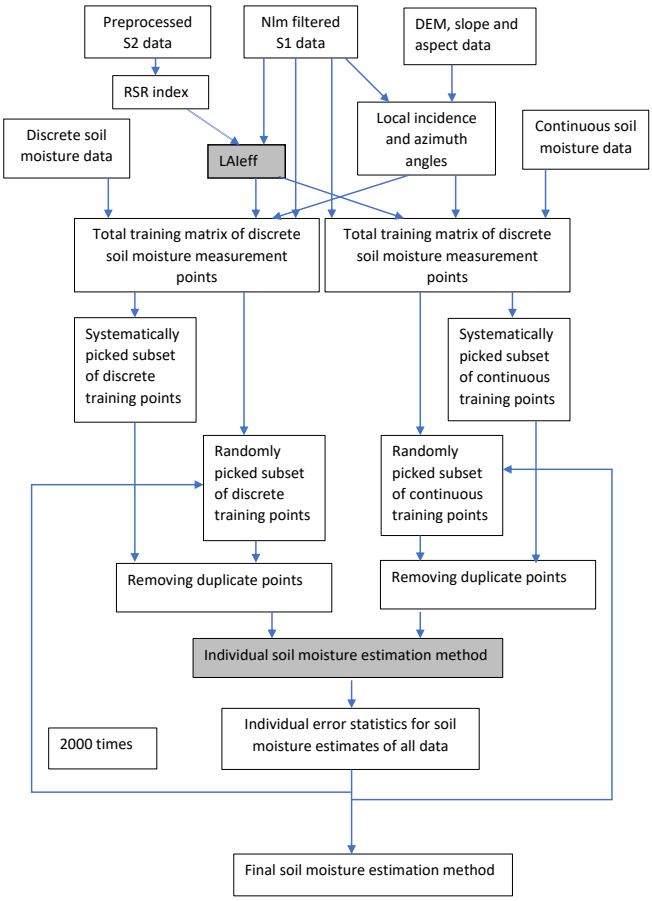

Fig. 7. The flowchart of the soil moisture method retrieval. The machine learning steps are indicated with gray shade. The random picking of the discrete and continuous training points is carried out 2000 times. The number of discretely and continuously measured points is roughly the same.

A. Ground contribution to backscattering

The total backscattering of the vegetated area σ^0 can be described with a water cloud with gaps [7][53][54][55]

$$\sigma^0 = (1-\eta)\sigma_{gr}^0 + \eta\,\tau_{can}^2\sigma_{gr}^0 + \eta\,\sigma_{can}^0(1-\tau_{can}^2) \quad (1)$$

where η denotes the canopy cover, $\sigma_{gr}{}^0$ and $\sigma_{can}{}^0$ are the backscattering contributions of the ground and canopy, respectively. The transmissivity of the canopy τ_{can} is related to the number of scatterers n per unit area and, each having the area A and thickness h, so that the volume of a single scatterer is $V = A\,h$. The amount of water of one scatterer is denoted by W_v (kg/m^3). The total amount of water per unit area in the canopy, i.e. the canopy water content m_v (kg/m^2), is then $n\,W_v\,A\,h$. Typically the canopy water content is taken to be related to the stem volume, especially when using longer wavelengths. In C band, however, also leaves contribute to scattering markedly, as their dimensions is comparable to the wavelength [58][59][60]. The branches of the dominating species of the area, the northern Norway spruce ('candle spruce'), are so small and hidden behind the shoots that their contribution in C band is not significant [61]. The question, whether the leaves (needles) or trunks dominate can be checked from the SAR data.

If the trunk-ground double bounce mechanism were the dominating scattering mechanism, the topography would reduce it so that the slopes facing the radar would produce weaker backscattering than flat areas, since the double bounce would not be directed to the radar due to the sloped terrain [Fig. 8]. However, the backscattering images do not show this kind of behavior. Hence, the trunk-ground double bounce backscattering is not the dominant scattering component in the area studied and the leaves (needles) appear to be the dominating scattering component from the canopy. Consequently, it is reasonable to characterize the canopy with LAI, not biomass.

The leaf area index *LAI* equals $n{\cdot}A$. Hence, the total amount of water in the canopy layer can be described by $m_v = W_v{\cdot}LAI{\cdot}h$. The canopy transmissivity is then exponentially decaying with the amount of water per unit area and can be given in the form

$$\tau_{can} = \exp(-Bm_v/\cos\theta) = \exp(-BW_vLAI\;h/\cos\theta) \quad (2)$$

where θ is the incidence angle of the microwave radiation and and B is a coefficient related to the canopy structure. For coniferous canopies, the true hemi-surface *LAI* is about 1.5 times the effective *LAI* [56][57]. One should also notice that both optical and microwave remote sensing *LAI* estimates may saturate at large *LAI* values, because the transmissivity is then negligible. As the leaves are taken to have identical thickness, the above equation can be simplified to the form

$$\tau_{can} = \exp\left(-B'W_vLAI_{eff}/\cos\theta\right) \quad (2)$$

Analogically, the backscattering coefficient of the canopy is

$$\sigma_{can}^0 = Am_v\cos\theta = A'W_vLAI_{eff}\cos\theta \quad (3)$$

where A' is a parameter related to the structure of the canopy. Because many individual SAR images taken in varying moisture conditions of the vegetation are used, the parameter W_v is explicitly shown both in the description of the backscattering and transmission of the vegetation.

Fig. 8. Schematic view of reduced trunk-ground double bounce backscattering of sloped terrain (right) compared to flat terrain (left).

Although LAI_{eff} can be estimated and θ is known for each pixel and η, A and B can be assumed to be constant during July 15 – August 31, there are still too many unknowns per pixel to derive the soil backscattering coefficient in a deterministic way. However, this analysis gives motivation for providing the machine learning method with LAI_{eff}. In addition, the low biomass [Table II][59] and sparse structure of the subarctic forests studied [62] support the view that soil moisture retrieval is possible using S1 data.

B. Effective LAI retrieval

For boreal forest, the reduced simple ratio (RSR) index suits for the canopy LAI_{eff} estimation using the S2 image [42]. We use the relationship $LAI_{eff} = 0.25 \cdot RSR + 0.27$ [43].

The SAR based LAI_{eff} estimation method was derived for the images June 26 – August 7, because then the LAI can be assumed to be about constant. As HH polarization was not available in the S1 data, a previously developed LAI retrieval method based on the VV/HH polarization ratio sensitivity to shoot orientation could not be applied [31]. On the other hand, the temporally averaged VH and VV polarization backscattering coefficients of single swaths showed only weak correlation with the S2 image based LAI_{eff} and saturated at about LAI_{eff} =1.3. To benefit from the availability of data from all three swaths, it was decided to derive a machine learning method for LAI_{eff} estimation using S1 data.

After removing open water areas and a cloud containing square from the S2 based LAI_{eff} image, there were 996897 pixels for machine learning training and validation. The LAI_{eff} distribution showed that a vast majority of the values were in the range of 0.99 – 1.35, which cover the percentiles from 0.2 to 0.95. In order to avoid their dominance in training the LAI_{eff} method, the trainset was chosen so that all values below 0.99 (20%) and above 1.35 (5%) were included in it. For every increment of 0.02 between the limit values 0.99 and 1.35, every 1600th value in order of magnitude was included. The total number of pixels in the trainset was then 484799. All of the smallest 20% and largest 5% LAI_{eff} values were taken in the training set to optimize the algorithm for the area in question. If a more general algorithm had been the objective, one would have left half of the smallest and largest values in the validation set.

The predictors used for the SAR based LAI_{eff} algorithm were the non-locally averaged backscattering coefficients of each swath and pass normalized with the respective average value of the land area in the same image, i.e. 12 individual values per one LAI_{eff} pixel. The normalization was adopted to avoid the effect of varying moisture conditions. No topography related parameters were used as predictors because the effect of topography is very complex for high vegetation.

The training was first tested using several methods: decision tree, gradient boosted trees, linear regression, nearest neighbors, neural network, random forest, and Gaussian process. The random forest approach [51][52] produced the best results with respect to standard deviation; therefore, it was chosen for the LAI_{eff} estimation. The derived algorithm was then applied to the average of individual non-locally averaged S1 images in the time range June 26 – August 7 to produce the LAI_{eff} map for the soil moisture retrieval.

C. Soil moisture retrieval

The boreal forest floor is mostly covered with field and ground layer vegetation including moss and shrub species. In addition, the wetland areas of the test site are also vegetated by shrubs, sedges, and mosses [34]. Hence, the ground backscattering coefficient is a complex combination of soil and low/high vegetation contributions. Therefore, the backscattering models for bare soil [4][5][20][22][24] were not used for the ground backscattering estimation in this study. Instead, an empirical relationship with the backscattering components and the soil moisture was sought on the basis of *in situ* soil moisture data and machine learning methods.

Due to the field and ground layer vegetation of the boreal forest, the backscattered signal does not trivially depend on the penetration depth in the soil layer [64], but may reveal soil moisture to which the surface vegetation reacts. Hence, it may be possible to relate the radar signal to soil moisture beneath the actual surface level.

A challenge in using the *in situ* soil moisture data set is that the continuous measurements provide a much larger data set than the spatially distributed discrete measurements. In addition, soil moisture has a quite marked temporal autocorrelation, so that the individual values of continuous measurements are not completely independent. Another complication is that part of the continuous measurements were converted from the values measured at 20 cm depth to the surface soil moisture. Naturally, their somewhat reduced quality will cause some inaccuracy in the surface soil moisture method retrieval.

The following values per pixel were paid attention to in picking the points for the machine learning training set at the time of satellite overpasses: 1) surface soil moisture, 2) day/time of soil moisture measurements, 3) altitude of terrain, 3) local slope angle of terrain θ_l, 4) local aspect angle of terrain φ_l, 5) class (see Section II.E), 6) local incidence angle θ, 7) azimuth difference of radar looking direction φ and terrain slope, 8) LAI_{eff} based on S1 data, 9) cos(θ) of S1 data, 10) time difference between the soil moisture measurement and SAR image acquisition, 11) σ_{VH}^0, 12) σ_{VV}^0, 13) $<\sigma_{VH}^0>_{land}$ and 14) $<\sigma_{VV}^0>_{land}$. Here $<\sigma_{VH}^0>_{land}$ and $<\sigma_{VV}^0>_{land}$ denote the average values of σ_{VH}^0 and σ_{VV}^0 of land pixels of each image.

In principle, one could use all available information (numbers 2 – 14 of the previous list) to predict the surface soil moisture (number 1). However, then the machine learning method might find solutions that are actually just typical soil moisture values related to altitude and land cover class or certain day, for example, and the SAR data would provide just noise on top of that. The real test for the suitability of SAR data in soil moisture estimation is to provide the machine learning method no parameters that don't depend on the SAR. Hence, the last nine parameters (numbers 6 – 14 in the list above) were used as the training information for the predictor to estimate the surface soil moisture (number 1).

The criteria 6 and 7 take into account the longer path the

radar signal has to go in the canopy before reaching the forest floor due to the terrain topography. The local incidence angle was provided for the predictor in the form of $\cos(\theta_l - \theta)\cdot\sin(\varphi_l - \varphi)$. The reason for the azimuthal weight is that the signal's penetration in the canopy in hilly terrain depends largely on that. The azimuth difference was provided in the form $\tan(\theta)\cdot\sin(\varphi_l - \varphi)$ to emphasize that the azimuth difference matters more for steep slopes. The radar incidence angle θ was given individually for each pixel of the image, but the radar azimuth angle φ was taken to be the same for all pixels of the pass in question. The criterion 8 (LAI_{eff}) provides the machine learning method information related to attenuation in the canopy. For bare surfaces taking into account only the local incidence angle of the terrain (criterion 6) would be sufficient, but for trees this is not the case. The trunks of trees are vertical also in sloped terrain, hence the crown scattering is dominated by the radar incidence angle (criterion 9), but also the local incidence angle (criterion 6) may have an effect, since the tree line to some extent follows the terrain topography. The time between the SAR image acquisition and the in situ soil moisture value (criterion 10) will help in taking into account the diurnal variation of the soil moisture.

The main criteria for soil moisture retrieval are numbers 11 and 12 providing the individual backscattering coefficients per pixel in the two polarizations available. The average backscattering values of the images (criteria 13 and 14) are provided in addition to the individual pixel backscattering coefficients, because possible winter hardening of coniferous species shows up in changed backscattering level and polarization ratio [62][64]. Then the attenuating effect of the canopy is markedly reduced, which affects the soil moisture retrieval. Providing the information about the general status in the area gives a possibility for the machine learning method to decide, whether the observed change in single pixel backscattering is due to soil moisture change only or possibly also to winter hardening.

Although points for the trainset were chosen so that they also cover the dynamic range of the timing of the soil moisture measurements and the topography related parameters (numbers 1 – 4 in the list), those parameters were not offered for the predictor in order to avoid producing information that is not observed by the SAR or related to its viewing configuration. Including the class information (number 5 in the list) to the trainset would have improved the results slightly, but it was skipped, because soil moisture measurements were not provided for all classes.

The training set was constructed so that it covered the dynamic variation range of all previously 14 listed parameters, although only 9 of them (all SAR dependent) were used as input for the machine learning algorithm. This was done to have good representativity for the training set. The discrete and continuously measured soil moisture values were picked separately and their fraction in the training set was roughly even, with slight preference of the discrete points. Roughly half of the discrete points were picked deterministically and the rest randomly. The continuously measured values were picked similarly in a half-random way. This was done as follows.

The data to be used for training (i.e. the above mentioned 9 parameter values per each soil moisture measurement at the time of the satellite overpass) was sorted in order of magnitude separately for each criterion, separately for the discrete and continuous measurements. For the discrete data, every 40th point was then picked in the order of magnitude of the criterion in question to be included in the training set in order to cover the dynamic variation of every parameter. In addition, for each criterion, a random point for every set of successive 41 points in the order of magnitude was also included in the training set. Possible redundant points (i.e. points picked on the basis of more than one criterion) were removed from the training set. The number of discrete measurement points in the training set was about 50% of available points. The continuous dataset was treated similarly, but in order to avoid its dominance, only every 9000th point was picked systematically and random points from all sets of 9001 successive sorted points. Possible redundant points were removed. The best morning method had 135 discrete and 126 continuously measured soil moisture values, the best evening method 131 and 122, respectively.

The time difference between the S1 image overpass and the soil moisture measurement was restricted to the range ± 1 day (i.e. [-24 h – 24 h]). Then, all soil moisture values measured at surface (including the converted ones) were used as training material. In addition, the time difference between the soil moisture measurement time and the S1 satellite overpass was used both as a criterion of picking points for the training set (besides the already mentioned criteria) and as an input parameter for the machine learning method developed.

First, all the machine learning methods already tested for the LAI_{eff} estimation were again tested for soil moisture retrieval for a few trainsets. The gradient boosted trees (GBT) method produced the best results with respect to standard deviation and was chosen as the method to be used for the final surface soil moisture estimation method retrieval [66][67][68][69][70]. This was carried out using 2000 different half-randomly picked training sets. The reason for GBT being the best method for the soil moisture estimation and the random forest method for the LAI_{eff} estimation is that GBT is able to provide better estimates for a finite variation range (0% — 100% for soil moisture) than random forest, which is otherwise a very good method (fast, stable and usually robust to outliers).

The R^2 values, RMSE and maximum errors of the predictor applied to all data were determined separately for all trainsets, the whole material and the validation set, which contained the points not included in the training set. The predictor producing the smallest maximum error for the whole material was chosen to be used. The reason for choosing the minmax norm instead of the highest R^2 or smallest RMSE value was to avoid overfitting to the spatially relatively sparsely representative material. Separate predictors were derived for the morning and evening images due to the hilly topography. The sensitivity of the method to the choice of training points and the chosen method was tested by producing soil moisture maps with three best point sets (with respect to minmax, RMSE and R^2) and calculating the difference of the soil moisture estimated per

pixel.

Machine learning methods may not apply outside the variation range of the parameter values used for training. In addition, major inaccuracy may appear in areas, which are geographically essentially different from those included in the trainset. Therefore, masks for morning and evening images were constructed for areas with possibly unreliable soil moisture estimates. The criteria checked were: 1) altitude (max), 2) terrain slope (max), 3) terrain aspect, 4) class, 5) local radar incidence angle, 6) local radar look angle and 7) LAI_{eff} (max). It should be noticed that the terrain aspect and local radar look angle matter only when the slope is relatively steep. The limit value of 0.2 was used for the slope to mask using the aspect and local look angle. Complications was caused by the LAI_{eff} = 0 values at high altitude above the tree line. No soil moisture measurements were available there. Yet the area is geographically very different from low altitude areas of low LAI_{eff} values. Hence, an additional criterion of LAI_{eff} = 0 values at an altitude higher than 450 m a.s.l. was used for masking areas of reduced reliability.

IV. Results

The RMSE of the S1 and S2 based LAI_{eff} estimates was 0.10 for the trainsets and 0.14 for all pixels. The relative 80% variation range of the S2 based LAI_{eff} estimate was in the whole land area 0.84 – 1.31 and the corresponding variation range of the S1 based LAI_{eff} estimate was 0.84 – 1.25. Hence, the microwave estimate had slightly smaller dynamic variation than the optical estimate and its mean value of 1.09 was also slightly smaller than the S2 based mean value 1.12 [Fig. 9]. However, the spatial variation of the LAI_{eff} estimates was quite similar [Fig. 10], the microwave estimate having slightly better pixel spacing (10 m) than the optical (20 m) due to the lower spatial resolution of the SWIR channel of S2.

The soil moisture predictors derived separately for morning and evening images are characterized statistically in Table VII and Fig. 11. The large difference between the maximum error and 90% percentile value suggests that the data contains a small number of points that deviate largely from the general behavior.

The relationship between the soil moisture estimation error and the parameters 1 – 14 listed in Section III.C revealed that there is a slight systematic dependence of the error on the Julian day of the soil moisture measurement. The error is more typically negative in May and early June. Since that period coincides with the start of the actual growth season of the canopy (and field and ground layer vegetation), the real LAI_{eff} is probably lower than the value derived using the summer images [Section III.B] and used for the whole period. Then the fraction of backscattering due to ground would be larger than assumed, which would lead to expected soil moisture being higher than the actual soil moisture, and thus cause negative estimation error. A first-order polynomial was fit to the error estimate dependence on the Julian day, but the R^2 values were very small, 0.03 for morning and 0.04 for evening. Hence, this systematic error trend is insignificant compared to the general estimation accuracy.

The estimated seasonal variation of soil moisture of the discrete measurement points was mostly in the range of the measurement variation of the three individual measured values with an accuracy of the RMSE variation of the estimation method. The morning images (Fig. 11, Table VIII) produced

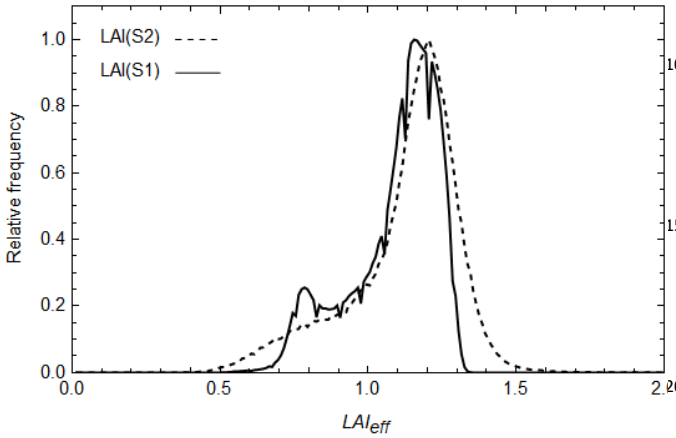

Fig. 9. The relative distribution of the S1 and S2 based *LAIeff* estimates.

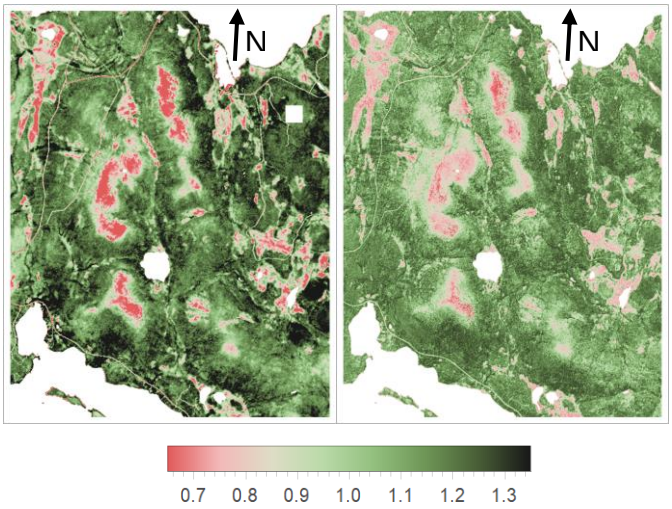

Fig. 10. *LAIeff* estimated from the RSR index of the S2 image (left) and using the random forest method applied to the VH and VV backscattering of the S1 images (right). Open water and the cloudy square in the S2 image are masked white.

TABLE VII
STATISTICS FOR THE PREDICTORS OF S1 BASED SURFACE SOIL MOISTURE [%] ESTIMATION WITHIN ONE DAY OF S1 OVERPASS USING MINMAX NORM.

Images	Parameter	Trainset	Validation set	All points
Morning	*n*	260	34836	35096
	offset	4.49	1.33	1.45
	slope	0.90	1.03	1.02
	R^2	0.94	0.87	0.87
	RMSE	5.9	5.7	4.3
	90% error percentile	8.0	9.0	9.0
	Max error	34.1	30.5	34.1
Evening	*n*	253	34375	34627
	offset	5.1	1.44	1.50
	slope	0.86	0.98	0.98
	R^2	0.91	0.90	0.90
	RMSE	6.8	4.6	3.3
	90% error percentile	10.9	6.8	6.8
	Max error	33.6	33.9	33.9

slightly better results than the evening images (RMSE = 6.5, R^2 = 0.88 for morning and RMSE = 8.8, R^2 = 0.77 for evening images). Since soil moisture has a strong temporal autocorrelation, except for the beginning of precipitation events, the large continuously measured data sets dominate the statistics of the whole data set in Table VII and the RMSE values are too small to be representative in general. Hence, the statistics of Table VIII is considered to better represent the quality of the soil moisture maps.

Although the statistics of the whole data set containing also the continuous soil moisture measurements looks about as promising for the evening data set as for the morning data set [Table VII], the GBT method derived for the morning data set produced clearly better soil moisture estimates for the time series of the spatially distributed discrete *in situ* points (Table VIII, Fig. 12, Supplementary Fig 1). The soil moisture estimates derived using the morning images are mostly within the range

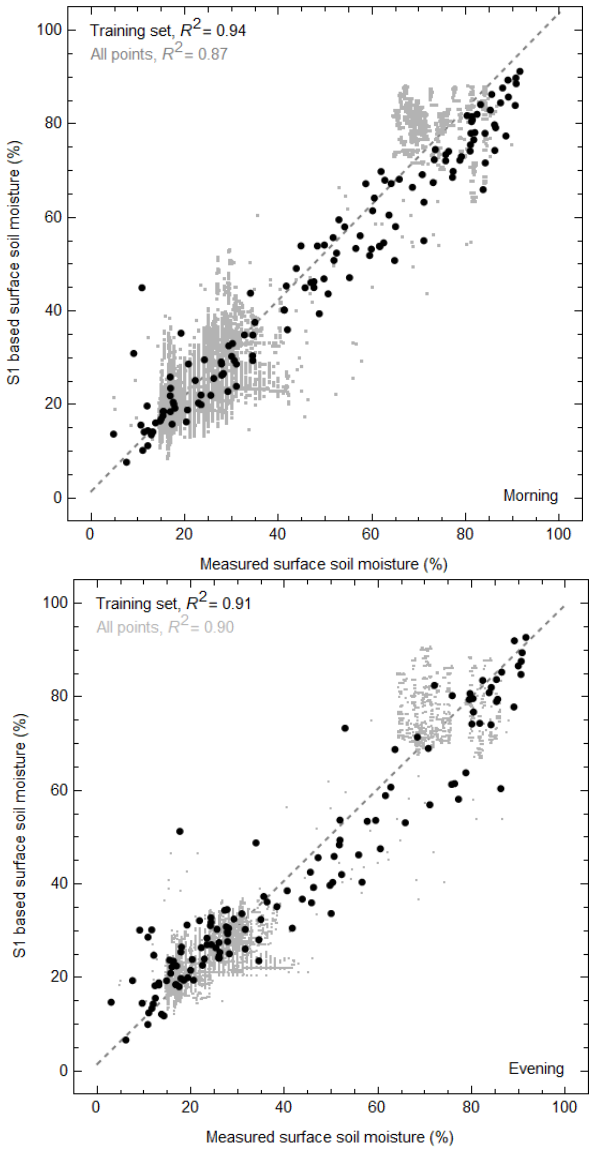

Fig. 11. The estimated surface soil moisture vs. the measured values for the trainset and all data using the algorithm derived for morning images (top) and evening images (bottom).

TABLE VIII
STATISTICS FOR THE PREDICTORS OF S1 BASED SURFACE SOIL MOISTURE [%] ESTIMATION WITHIN ONE DAY OF S1 OVERPASS FOR THE TIME SERIES OF THE DISCRETE POINTS.

Images	Parameter	Discrete points
Morning	*n*	202
	offset	6.3
	slope	0.84
	R^2	0.88
	RMSE	6.5
	90% error percentile	15.3
	Max error	30.1
Evening	*n*	165
	offset	8.2
	slope	0.80
	R^2	0.77
	RMSE	8.8
	90% error percentile	24.2
	Max error	32.9

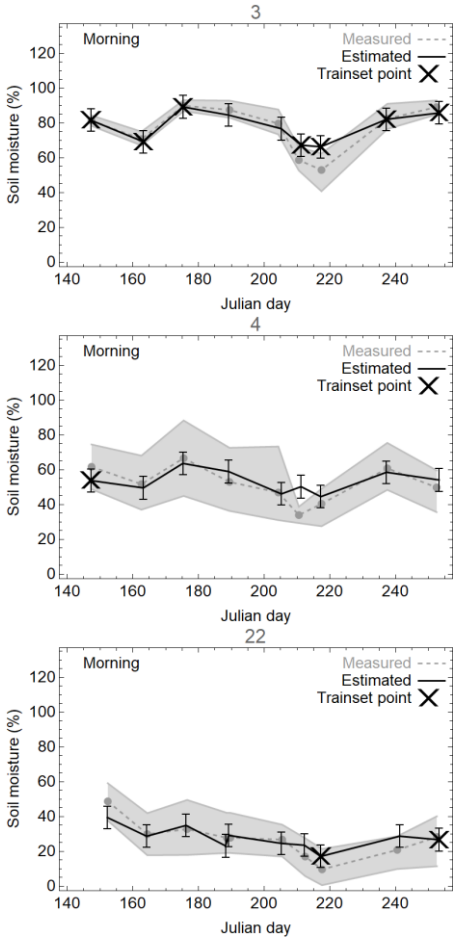

Fig. 12. Examples of the estimated surface soil moisture vs. the measured *in situ* values for the discrete points using the algorithm derived for morning images. The gray points show the individual triplet means and the shaded gray area the variation of the triplet values. The black points are the estimated values derived from the morning images of the same day, i.e. a few hours before the *in situ* values. The black error bars are the RMSE values for the GBT method derived for the morning images and applied to the discrete points only (Table VIII). The points marked with a cross were included in the training set.

of the *in situ* triplet variation measured during the same day, taking into account the RMSE value of the S1 based soil moisture estimates. For evening images (Table VIII), the

agreement of the derived soil moisture estimates with the *in situ* measurements of the same day had more scatter.

The potential of the method for daily soil moisture estimation on pixel basis was studied using the continuous surface soil moisture measurements. Two of the measured points were within one S1 pixel, the other two being situated in two neighbor pixels. The distance between those four points varied in the range 4 − 24 m. The estimated soil moisture per day is shown in for those three pixels in Fig. 13. Obviously the variation within the pixel is high and it is not known, for example, whether s13 or s19 represents the dominating characteristics of the pixel they are in. Although the diurnal variation of those points is rather small compared to the soil moisture estimation method uncertainty, the estimated values seem to catch the main features of the soil moisture seasonal variation. In addition, the difference in soil moisture values of neighbor pixels shows also in the estimated values. However, also the difference in swaths (successive points) appears occasionally in the estimated soil moisture curved. Partly this suggests that the machine learning method is not completely able to hide incidence angle variation related effects. Partly this may be a result of not having the same soil moisture variation in the training set per swath, as the precipitation events were not evenly distributed between swaths.

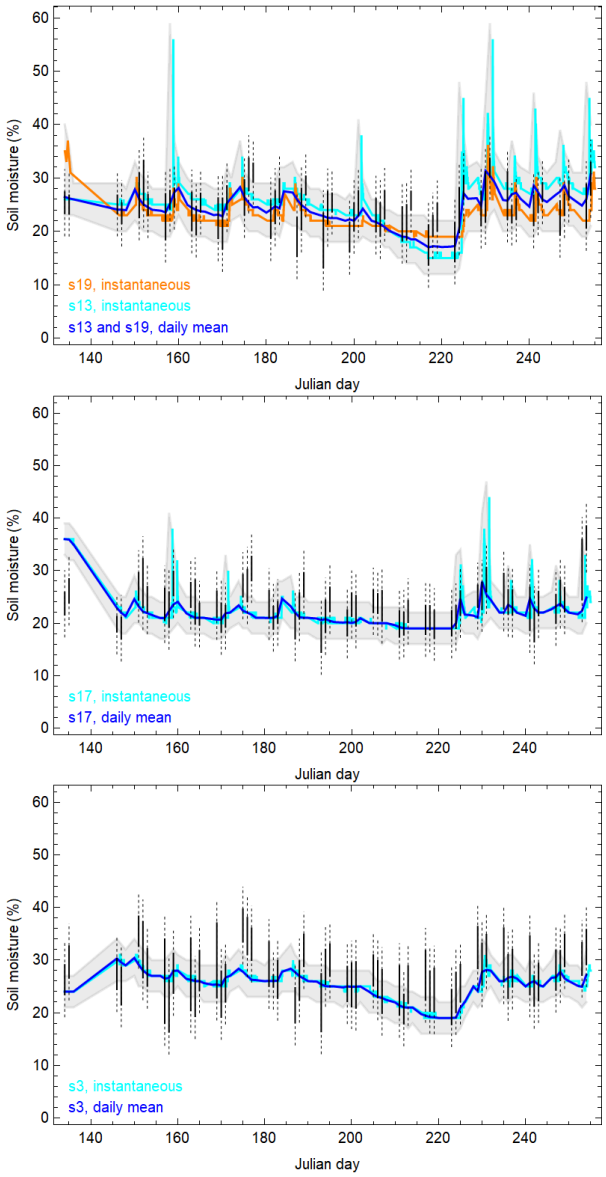

Fig. 13. Daily averaged measured soil moisture four three neighbor S1 pixels (blue) and the variation range (gray band). The individual instantaneous measured values are shown in cyan and orange. The variation range of corresponding estimated soil moisture values based on morning S1 images and the individual *in situ* measurement times are shown as black vertical lines. The variation range is extended by the method overall uncertainty 4.3% [Table VII] indicated as dotted lines.

Only three points of the discrete data were located in neighbor pixels during the intensive measurements. A comparison of nearby *in situ* and estimated soil moisture values supported the view that the soil moisture can in principle be estimated on pixel basis, but the number of points (Supplementary Fig. 2) is far too small to prove it.

Altogether 60 instantaneous midday soil moisture maps, one per day, were generated using morning images in order to perceive the temporal variation of the soil moisture of the whole area [Supplementary Fig 3]. Examples are shown in Fig. 14.

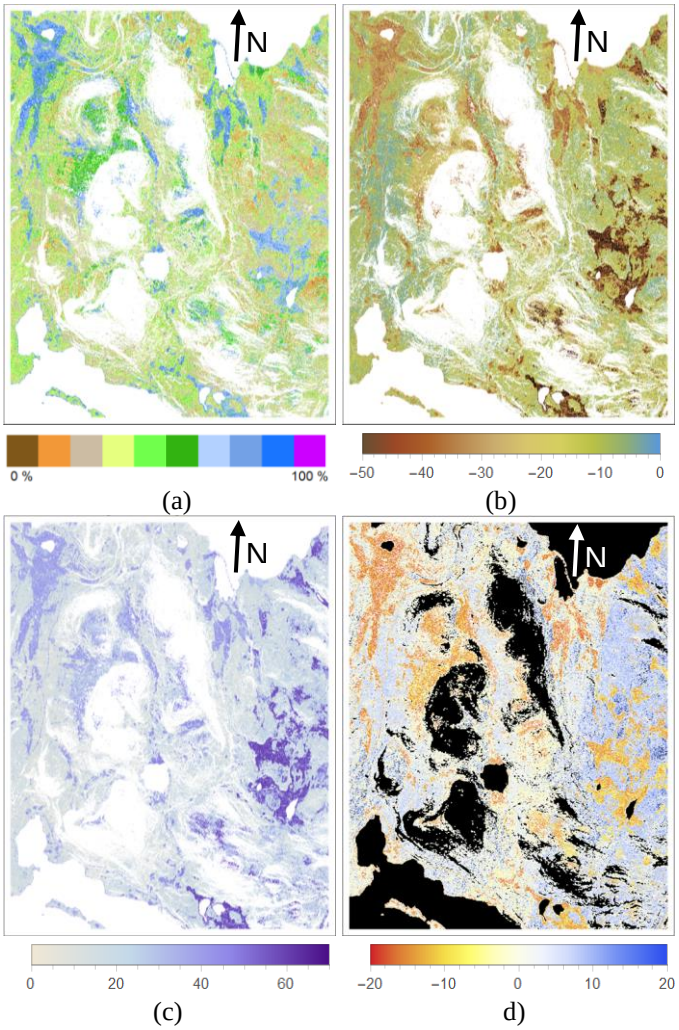

Fig. 14. Calculated soil moisture map for a) May 21, 2019 and difference maps for b) August 12 – May 21, 2019, c) August 13 – 12, 2019 and d) September 12 – May 21, 2019. All maps are calculated for midday (UTC 10) using morning images of S1. Open water and masked areas are white in a), b) and c) and black in d).

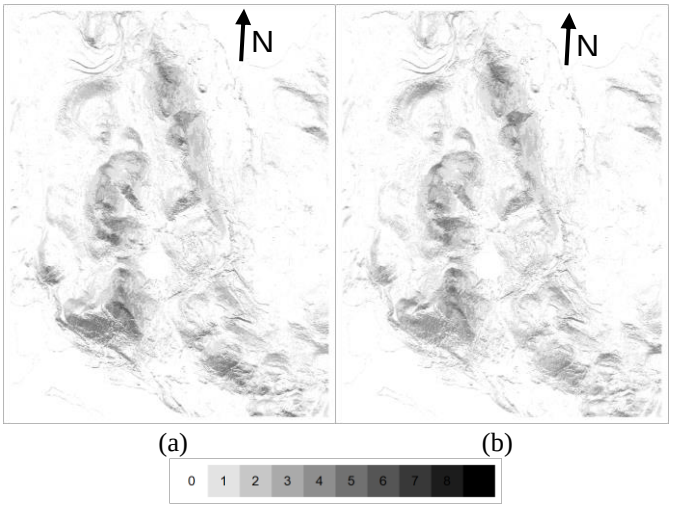

Fig. 15. Mask layers for the soil moisture retrieval using S1 morning data, a) morning images and b) evening images. The gray level indicates for how many criteria the pixel is masked.

May 21 is about the time of snowmelt at the tree line and thus represents an early summer case of high soil moisture. August 12 was preceded by a dry period of two weeks and the difference to the early summer situation is obvious [Fig. 14b].

One example of the effect of precipitation on soil moisture is shown as a difference between August 13 and August 12 [Fig. 14c]. Autumn soil moisture conditions are clear in the image of September 12 [Supplementary Fig. 3], the difference of which to the situation of May 21 is shown in Fig. 14d. The masks for morning and evening images are shown in Fig. 15.

The sensitivity of the result to the trainset and method was studied comparing the soil moisture maps produced using morning images and the chosen best minmax method and the methods based on lowest RMSE or highest R^2 value. The difference between the maximum and minimum value of the soil moisture estimates produced using the three methods was used to generate a map of pixelwise sensitivity to the chosen trainset and method [Fig. 16]. The corresponding relative sensitivity was obtained by dividing the difference of the largest and smallest value with the median value [Fig. 16]. The average difference was 6.3% with a standard deviation of 5.7%, both given in absolute soil moisture values. Those values are quite in line with the quality statistics of Table VIII.

V. Discussion

Our approach successfully produced very high resolution (~20 m with 10 m pixel spacing) soil moisture estimates on a daily scale to a subarctic landscape. Generally, estimates were matching with 6.5% - 8.8% RMSE accuracy (in absolute soil moisture percentages) and R^2 values within 0.8 – 0.9 in all study period and was especially good for morning images. The R^2 values obtained using the whole data sets (0.87 – 0.90) are probably so high and the RMSE values (3.3% – 4.3%) so small, because of the high temporal autocorrelation of the continuously measured data, as soil moisture changes abruptly only at the beginning of a precipitation event. Hence, we use here the higher RMSE and lower R^2 values obtained for the discrete time series points to characterize the accuracy of the method.

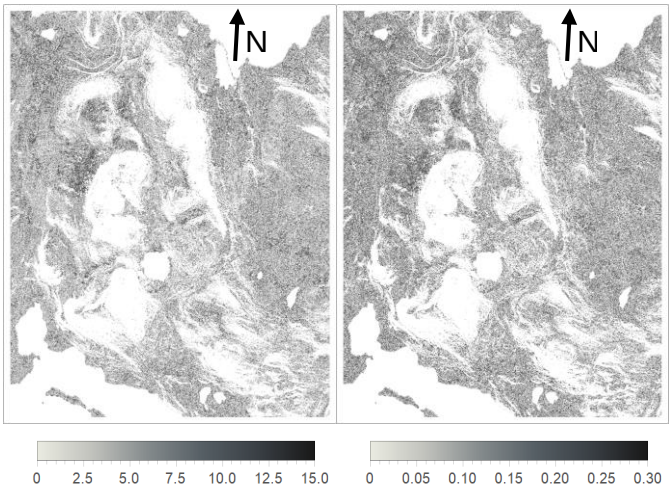

Fig. 16. The masked difference (in soil moisture%) of June 21, 2019 (a) and relative difference (b) between the largest and smallest soil moisture estimate values derived using the best morning algorithms based on 1) the minmax norm, 2) the smallest RMSE and 3) the largest R^2 values.

Results agreed with previous SAR studies done in other than agricultural areas, such as Mediterranean [24], Polish wetland [8] and subarctic conditions [10], in showing that C band SAR data can be used for soil moisture retrieval in diverse landscapes. Our approach, benefiting from the completely new PIMSAR nonlocal despeckling method [41], is unique and also produced good estimates in forest areas, which have been mostly lacking in previous S1 data using studies [24]. In boreal forests one complication is the field and ground layer vegetation, which may be quite thick, when the canopy is relatively small. However, the field and ground layer vegetation will reflect the surface soil moisture conditions. In addition, the crown shape of northern Norway spruce is very narrow, so that a larger fraction of the ground is visible between the trees than in southern Finland for the same *LAI* value. PIMSAR also enables retrieval of as accurate (~ 10%) soil moisture maps without need of optical data in higher spatial resolution (~20 m) compared to the previously mentioned three studies: 30 m [24], 500 m [8] and 100 m [10].

It was a challenge for the method development that the spatially distributed data was not measured at the times of the satellite overpasses and the continuous measurements were available practically only in five locations and in four of them the measurements were carried out only at 20 cm depth or deeper. In addition, spatially the data set was rather small for machine learning method development. Soil moisture values were available only for less than 0.1% of the land pixels. Yet the results are convincing and indicate that it is possible to develop a high-resolution soil moisture retrieval method that uses S1 images as input. Here we used one S2 image for LAI_{eff} training, but in an area, where notable changes in the canopy do not take place, one could probably use the same LAI_{eff} map for more than just one summer, so that only S1 data would be needed for the soil moisture map processing. This is a major advantage in areas that are often cloudy or at times of poor

illumination.

Including LAI_{eff} in the training set provided the machine learning algorithm information about the attenuation of the radar signal in the canopy and prevents the method from finding solutions that are not consistent with that. In the study area there was a vague increasing trend of soil moisture with decreasing LAI_{eff}. This might favor finding high soil moisture values in low LAI_{eff} area. However, the trend was so weak (R^2 = 0.28) that it should not have a marked effect on the results.

The reason for the morning images to have slightly smaller RMSE values than evening images in this study contrarily to the Polish wetland study [8] is partly due to the topography and location of the measurement points [Fig. 1][34]. Several of the discrete measurement points were in the distal slopes of the evening SAR images and prone to shading. In addition, this study profited by the continuous measurements, which provided an intelligent way to relate midday measurements to morning and evening S1 images.

Soil moisture estimates are very important for various applications including climate, hydrological, biogeochemical and management purposes [71][72]. As subarctic areas are facing very drastic changes due to climate change, including changes in precipitation and snowmelt timing [73], it is essential to have high-resolution estimates from soil moisture conditions in time and space [74]. Our methodology could be further applied to other subarctic and arctic landscapes to provide soil moisture estimates for i) ecohydrological and vegetation water usage studies [75], ii) to improve greenhouse gas estimates from forest lands, iii) to map influences of different land use actions such as logging at the landscape level, and iv) to assess the forest fire risk. Additionally, better estimates from soil moisture can further help local land managers to identify spatially important areas from soil moisture point of view.

Using machine learning methods requires extensive data sets to obtain generally applicable methods. When the data availability is limited in space or time, there is a considerable risk of overfitting. Hence, the results of this study rather demonstrate the capability of using S1 SAR images for seasonal soil moisture variation in one study area than provide a final method to be used in other areas. With larger amounts of soil moisture data, a generally applicable method could be developed, for example, using global soil moisture network databases.

The results obtained using machine learning methods depend largely on the training data provided. Here the soil moisture distributions of the discrete data sets were slightly bimodal [Fig. 2] and the continuous data sets were dominated by soil moisture values below 40%. The goal was to achieve a homogenous error structure, i.e. the low, average, and high soil moisture estimation accuracy should be about the same. For this purpose, the training data was chosen to cover the whole variation range of all parameters affecting the soil moisture about evenly. In addition, the choice of the best method was based on minmax norm, so that even the rare values are estimated at least at the accuracy of the minmax value. However, it depends on the application what kind of an error structure is desirable. If one would prefer to have the best accuracy for most cases at the expense of highly erroneous rare values, one would choose another norm (for example least squares) and pick the training set so that the most common values are most common also in that.

The effect of the slightly bimodal soil moisture distribution on the error statistics was checked by deriving the RMSE, 90% quantile and max error values also separately for soil moisture ranges 5 − 50% and 60 − 90 % (Tables VIIA and VIIB in the Supplement material). For the lower range the soil moisture RMSE was 5.2% and 3.9% for all points in the morning and evening, respectively. Hence, the estimation accuracy was about the same as for the whole material. For the high soil moisture variation range the corresponding RMSE values were 10.4% and 9.8%, respectively. Hence, it was slightly poorer than for the whole material. Very probably this is mainly due to the inaccuracy of the conversion of the values measured 20 cm depth to surface values [Fig. 2, Tables IV - VI]. The RMSE value for the conversion formula derived using the discrete measurements was namely only 8.3%. The fraction of the converted data was for morning cases: 96% for the high range, 35% for the low range and 40% for all points. For evening the corresponding values were 97%, 36% and 40%.

Balancing between spatially distributed discrete measurement points and a few continuous measurement points was a major issue in this study. In addition, the instruments used for continuous measurements were not identical with those used for the discrete data sets. This inevitably affected the results to some extent. Especially, transforming the soil moisture measured at 20 cm depth to surface values certainly caused extra uncertainty, which is most probably seen in the scatter of points in Fig. 11. However, the achieved soil moisture estimation accuracy was of the order of that of the *in situ* measurements.

In the autumn underestimation of the soil moisture may occur because of reduced canopy scattering due to fall of deciduous leaves and possibly also the decreasing impedance of Norway spruce needles, which may take place already in August, while the trees prepare for the winter [64]. However, precipitation will mask this effect.

Although topography constitutes a major complication for the use of SAR data, the machine learning method was able to adapt to hilly terrain surprisingly well. Only the completely shaded areas or slopes of serious foreshortening will not be suited for soil moisture estimation. The use of both morning and evening data is complicated by the hilly terrain, because the shaded or foreshortened pixels are in different locations. In addition, the discrete *in situ* measurements were taken in daytime after the morning SAR image retrieval, whereas the evening image overpasses succeeded the discrete *in situ* measurements. Comparing the average backscattered intensities of the images to precipitation and air humidity [Supplementary Figs 4 and 5] showed that the dynamic variation of the evening image intensity values is smaller than that of the morning images, especially in late summer. It may be that vegetation reacts differently to varying moisture conditions during late summer than at the beginning of summer.

Hence, also for this reason deriving a consistent soil moisture estimation method functioned better using morning images.

The difference of swaths was taken into account by the local incidence and azimuth angles. Within the dynamic range of the swath incidence angles corresponding to the *in situ* measurement sites, the soil moisture distributions of the whole season were very similar for all swaths. However, it may be that results would improve, if the LAI_{eff} values were derived separately for each swath, as the signal penetration in canopy depends on the incidence angle. When a larger *in situ* data set is available, specific methods adapted to the diverse swaths can be derived, which will improve the results even more.

VI. Conclusion

In this study a S1-based method for assessing soil moisture in a heterogeneous landscape was developed. It was demonstrated that with machine learning methods, PIMSAR non-local mean filtered S1 pixels of 10 m can be used for accurate instantaneous surface soil moisture estimation. The RMSE values were within 6.5% for the morning images and 8.8.% for the evening images, which is of the order of the accuracy of the *in situ* measurements. The soil moisture estimates followed consistently the seasonal variation of the *in situ* values. Precipitation events were also manifested in the soil moisture map time series. The machine learning method was quite robust, since the pixelwise sensitivity to the choice of the method was estimated to be on the average 6.3% with a standard deviation of 5.7% (in absolute soil moisture values).

Yet, the validity of the method could be checked only in the 92 different locations, for which soil moisture data was available: altogether 678 discrete values and 8 points of continuous data. Hence, although the filtered S1 data used has very high spatial resolution, ~20 m, with pixel size of 10 m, the spatially very limited *in situ* data set does not allow to conclude that the derived soil moisture maps would have the same soil moisture accuracy in the pixel resolution as the individual pixels used for developing and validating the soil moisture retrieval method.

To conclude, S1 has a high capability for observing seasonal soil moisture variation of subarctic forested and wetland areas. Furthermore, this methodology could be applied also to other subarctic and arctic regions to provide soil moisture estimates for the following purposes:

- ecohydrological and vegetation water usage studies
- to improve greenhouse gas estimates from forest lands
- to map influences of different land use actions such as logging at the landscape level
- to assess the forest fire risk.

Acknowledgment

The authors are grateful to Annika Alsila, Stephanie Gerin, Valtteri Hyöky, Viivi Lindholm, Timo Penttilä and Petri Salovaara, for their field work.

References

[1] S. Paloscia P. Pampaloni,S. Pettinato,and E. Santi, “A Comparison of Algorithms for Retrieving SoilMoisture from ENVISAT/ASAR Images”, IEEE Transactions on Geosci. Remote Sensing, vol. 46, no. 10, 2008.

[2] J. Peng, C. Albergel, A. Balenzano, L. Brocca, O. Cartus, M.H. Cosh, W.T. Crow, K. Dabrowska-Zielinska, S. Dadson, M.W.J. Davidson, P. de Rosnay, W. Dorigo, A. Gruber, S. Hagemann, M. Hirschi, Y.H. Kerr, F. Lovergine, M.D. Mahecha, P. Marzahn, F. Mattia, J.P. Musial, S. Preuschmann, R.H. Reichle, G. Satalino, M. Silgram, P.M. van Bodegom, N.E.C. Verhoest, W. Wagner, J.P. Walker, U. Wegmüller, A. Loew, “A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements”, Remote Sensing of Environment, 252, 112162, 2021.

[3] W. Wagner, G. Blöschl, P. Pampaloni, J.-C. Calvet, B. Bizzarri, J.-P. Wigneron and Y. Kerr, ”Operational readiness of microwave remote sensing of soil moisture for hydrologic applications”, Nordic Hydrology, Vol. 38, No.1, pp 1-20, 2007. doi: 10.2166/nh.2007.029

[4] P.C. Dubois, J., Van Zyl, T. Engman, “Measuring soil moisture with imaging radars”, IEEE Trans. Geosci. Remote Sens. vol. 33, pp. 915–926, 1995.

[5] Y., Oh, “Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces”, IEEE Trans. Geosci. Remote Sens. vol. 42, pp. 596–601, 2004.

[6] N. Baghdadi, M. Choker, M. Zribi, M. El Hajj, S. Paloscia, N.E.C. Verhoest, H. Lievens, F. Baup and F. Mattia, “A New Empirical Model for Radar Scattering from Bare Soil Surfaces”, Remote Sensing, 8, 920, 14 p., 2016.

[7] H.F. Benninga, R. van der Velde, Z. Su, “Sentinel-1 soil moisture content and its uncertainty over sparsely vegetated fields”, Journal of Hydrology X, in press, 2020, doi: https://doi.org/10.1016/j.hydroa.2020.100066.

[8] K. Dabrowska-Zielinska, J. Musial, A. Malinska, M. Budzynska, R. Gurdak, W. Kiryla, M. Bartold and P. Grzybowski, “Soil Moisture in the Biebrza Wetlands Retrieved from Sentinel-1 Imagery”, Remote Sensing, 10, 1979, 24 p., 2018.

[9] M. Urban, C. Berger, T.E. Mudau, K. Heckel, J. Truckenbrodt, V. Onyango Odipo, I. P. J. Smit and C. Schmullius, “Surface Moisture and Vegetation Cover Analysis for Drought Monitoring in the Southern Kruger National Park Using Sentinel-1, Sentinel-2, and Landsat-8”, Remote Sens., 10, 1482, 2018, DOI:10.3390/rs10091482.

[10] S. Zwieback, A.A. Berg, “Fine-Scale SAR Soil Moisture Estimation in the Subarctic Tundra”, IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 7, pp. 4898-4912, July 2019, doi: 10.1109/TGRS.2019.2893908.

[11] M. W. Lang and E. S. Kasischke, "Using C-Band Synthetic Aperture Radar Data to Monitor Forested Wetland Hydrology in Maryland's Coastal Plain, USA," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 2, pp. 535-546, Feb. 2008, doi: 10.1109/TGRS.2007.909950.

[12] A.-L., Cognard, C. Loumagne, M. Normand, P. Olivier, C. Ottle´, D. Vidal-Madjar, S. Louahala, S. and A. Vidal, ”Evaluation of the ERS 1/synthetic aperture radar capacity to estimate surface soil moisture: two-yearresults over the Naizin watershed”,Wat. Res. Res., Vol. 31, No. 4, pp. 975–982, 1995.

[13] J. Alvarez-Mozos, J. Casali, M. Gonzalez-Audicana and N.E.C. Verhoest, ”Correlation between groundmeasured soil moisture and RADARSAT-1 derived backscatter coefficient over an agricultural catchment of Navarre (North of Spain)”,Biosyst. Engng., Vol., 92, No. 1, pp. 119–133, 2005.

[14] B. Bauer-Marschallinger, V. Freeman, S. Cao, C. Paulik, S. Schaufler, T. Stachl, S. Modanesi, C. Massari, L. Ciabatta, L. Brocca, W. Wagner, "Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 1, pp. 520-539, Jan. 2019, doi: 10.1109/TGRS.2018.2858004.

[15] M. Foucras, M. Zribi, C. Albergel, N. Baghdadi, J.-C. Calvet and T. Pellarin, “Estimating 500-m Resolution Soil Moisture Using Sentinel-1 and Optical Data Synergy”, Water 2020, 12, 866; doi:10.3390/w12030866.

[16] N.N. Das, D. Entekhabi, R.S. Dunbar, M.J. Chaubell, A. Colliander, S. Yueh, T. Jagdhuber, F. Chen, W. Crow, P.E. O'Neill, J.P. Walker, A. Berg, D.D. Bosch, T. Caldwell, M.H. Cosh, C.H. Collins, E. Lopez-Baeza, M. Thibeault,”The SMAP and Copernicus Sentinel 1A/B

microwave active-passive high resolution surface soil moisture product", Remote Sensing of Environment, vol. 233, 111380, 17 p., 2019.

[17] M. Hornáček, W. Wagner, D. Sabel, H.-L. Truong, P. Snoeij, T. Hahmann, E. Diedrich,and M. Doubková, "Potential for High Resolution Systematic Global Surface Soil Moisture Retrieval via Change Detection Using Sentinel-1", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 4, pp. 1303-1311, Aug. 2012, doi: 10.1109/JSTARS.2012.2190136.

[18] S. Paloscia, S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, A. Reppucci, "Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation", Remote Sensing of Environment, vol. 134, pp. 234-248, 2013.

[19] F. Mattia, A. Balenzano, G. Satalino, F. Lovergine, J. Peng, U. Wegmuller, O. Cartus, M.W.J. Davidson, S. Kim, J. Johnson, J. Walker, X. Wu, V.R.N. Pauwels, H. McNairn, T. Caldwell, M. Cosh and T. Jackson, "Sentinel-1 & Sentinel-2 for SOIL Moisture Retrieval at Field Scale," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 6143-6146, doi: 10.1109/IGARSS.2018.8518170.,

[20] S. Bousbih, M. Zribi, M. El Hajj, N. Baghdadi, Z. Lili-Chabaane, Q. Gao and P. Fanise, "Soil Moisture and Irrigation Mapping in A Semi-Arid Region, Based on the Synergetic Use of Sentinel-1 and Sentinel-2 Data", Remote Sens. 2018, 10, 1953; doi:10.3390/rs10121953.

[21] F. De Zan, G. Gomba. "Vegetation and soil moisture inversion from SAR closure phases: first experiments and results", Remote Sensing of Environment, vol. 217, pp. 562-572, 2018.

[22] A. Merzouki, H. McNairn, J. Powers and M. Friesen, "Synthetic Aperture Radar (SAR) Compact Polarimetry for Soil Moisture Retrieval", Remote Sensing, 11, 2227, 24 p., 2019; doi:10.3390/rs11192227.

[23] K. Morrison, and W. Wagner, "Explaining anomalies in SAR and scatterometer soil moisture retrievals from dry soils with sub-surface scattering", IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no.3, pp. 2190-2197, 2020. doi: https://doi.org/10.1109/TGRS.2019.2954771.

[24] D.D. Alexakis, F.-D.K. Mexis, A.-E.K. Vozinaki, I.N. Daliakopoulos and I.K. Tsanis, "Soil Moisture Content Estimation Based on Sentinel-1 and Auxiliary Earth Observation Products. A Hydrological Approach", Sensors, vol. 17, 1455, 16 p., 2017; doi:10.3390/s17061455.

[25] A.-K. Holtgrave, M. Förster, F. Greifeneder, C. Notarnicola and B. Kleinschmit, "Estimation of Soil Moisture in Vegetation-Covered Floodplains with Sentinel-1 SAR Data Using Support Vector regression", Journal of Photogrammetry, Remote Sensing and Geoinformation Science (2018) 86:85–101, https://doi.org/10.1007/s41064-018-0045-4.

[26] S. Guo, X. Bai, Y. Chen, S. Zhang, H. Hou, Q. Zhu and P. Du, "An Improved Approach for Soil Moisture Estimation in Gully Fields of the Loess Plateau Using Sentinel-1A Radar Images", Remote Sens., vol. 11, pp.349, 2019; doi:10.3390/rs11030349.

[27] A. Hachani, M. Ouessar, S. Paloscia, E. Santi & S. Pettinato (2019) Soil moisture retrieval from Sentinel-1 acquisitions in an arid environment in Tunisia: application of Artificial Neural Networks techniques, International Journal of Remote Sensing, 40:24, 9159-9180, DOI: 10.1080/01431161.2019.1629503

[28] F. Greifeneder, C. Notarnicola, W. Wagner, "A Machine Learning-Based Approach for Surface Soil Moisture Estimations with Google Earth Engine", Remote Sensing, vol. 11, no. 13, pp. 2099, 2021. https://doi.org/10.3390/rs13112099

[29] M.P. Mendes, M. Matias, R.C. Gomes, A.P. Falcão. "Delimitation of low topsoil moisture content areas in a vineyard using remote sensing imagery (Sentinel-1 and Sentinel-2) in a Mediterranean-climate region", Soil & Water Res., vol. 16, pp. 85−94, 2021.

[30] S. Schönbrodt-Stitt, N. Ahmadian, M. Kurtenbach, C. Conrad, N. Romano, H.R. Bogena, H. Vereecken and P. Nasta, "Statistical Exploration of SENTINEL-1 Data, Terrain Parameters, and in-situ Data for Estimating the Near-Surface Soil Moisture in a Mediterranean Agroecosystem", Frontiers in Water, Vol. 3, pp. 75, 2021; doi:10.3389/frwa.2021.655837.

[31] T. Manninen, P. Stenberg, M. Rautiainen, P. Voipio and H. Smolander, "Leaf area index estimation of boreal forest using ENVISAT ASAR," IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 11, pp. 2627-2635, Nov. 2005, doi: 10.1109/TGRS.2005.857325.

[32] A. Lohila, T. Aalto, M. Aurela, J. Hatakka, J.-P. Tuovinen, J. Kilkki, T. Penttilä, J. Vuorenmaa, P. Hänninen, R. Sutinen, Y. Viisanen and T. Laurila, "Large contribution of boreal upland forest soils to a catchment-scale CH4 balance in a wet year", Geophysical Research Letters, Vol. 43, pp. 2946–2953, 2016. doi: 10.1002/2016GL067718

[33] E. Vainio, O. Peltola, V. Kasurinen, A.-J. Kieloaho, E.-S. Tuittila and M. Pihlatie, "Topography-based modelling reveals high spatial variability and seasonal emission patches in forest floor methane flux", Biogeosciences Discuss. [preprint], https://doi.org/10.5194/bg-2020-263, in review, 2020.https://bg.copernicus.org/preprints/bg-2020-263/

[34] A. Räsänen, T. Manninen, M. Korkiakoski, A. Lohila, T. Virtanen, "Predicting catchment-scale methane fluxes with multi-source remote sensing", Landscape Ecol (2021). https://doi.org/10.1007/s10980-021-01194-x

[35] C.-A. Deledalle, L. Denis, F. Tupin, "Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights", IEEE Trans. on Image Process., vol. 18, no. 12, pp. 1-12, 2009.

[36] D. Gragnaniello, G. Poggi, G. Scarpa, and L. Verdoliva, "SAR Image Despeckling by Soft Classification", IEEE J. of Selected Topics in Appl. EO and Remote Sens., vol. 9, no. 6, pp. 2118-2130 2016.

[37] P.A.A. Penna and N.D.A. Mascarenhas, "(Non-) homomorphic approaches to denoise intensity SAR images with non-local means and stochastic distances", Computers and Geosciences, vol. 111, pp. 127–138, 2018.

[38] S. Vitale, D. Cozzolino, G. Scarpa, L. Verdoliva, G. and Poggi, "Guided Patchwise Nonlocal SAR Despeckling", IEEE Trans. Geosci. Remote Sensing, vol. 57, no. 9, pp. 6484-6498,2019.

[39] G. Chierchia, M. El Gheche, G. Scarpa, and L. Verdoliva,"Multitemporal SAR image despeckling based on block-matching and collaborative filtering,"IEEE Transactionson Geoscience and Remote Sensing, vol. 55, no. 10, pp.5467–5480, 2017.

[40] W. Zhao, C.-A. Deledalle, L. Denis, H. Maître, J.-M. Nicolas and F. Tupin, "Ratio-Based Multitemporal SAR Images Denoising: RABASAR", *IEEE Transactions on Geoscience and Remote Sensing*, Institute of Electrical and Electronics Engineers, 2019, 10.1109/TGRS.2018.2885683, hal-01791355v2.

[41] T. Manninen and E. Jääskeläinen, "Pixel based multitemporal Sentinel-1 SAR despeckling PIMSAR", accepted for publication in the IEEE Geoscience and Remote Sensing Letters, 2021.

[42] P. Stenberg, M. Rautiainen, T. Manninen, P. Voipio and H. Smolander, "Reduced simple ratio better than NDVI for estimating *LAI* in Finnish pine and spruce stands", Silva Fennica, vol. 38, no.1, pp. 3–14, 2004.

[43] J. Heiskanen, M. Rautiainen, L. Korhonen, M., Mõttus, and P. Stenberg, "Generating fine resolution leaf area index maps for boreal forests of Finland," in Proc. 2011 IEEE International Geoscience & Remote Sensing Symposium, Vancouver, Canada, 2011, pp. 2326-2329. https://doi.org/10.1109/IGARSS.2011.6049675.

[44] R. Sutinen, P. Närhi, M. Middleton, P. Hänninen, M. Timonen, M. and M.-L. Sutinen, "Advance of Norway spruce (Picea abies) onto mafic Lommoltunturi fell in Finnish Lapland during the last 200 years", Boreas, Vol. 41, pp. 367–378, 2012. 10.1111/j.1502-3885.2011.00238.x. ISSN 0300-9483.

[45] P. Pirinen, H. Simola, J. Aalto, J.-P. Kaukoranta, P. Karlsson, and R. Ruuhela, "Climatological statistics of Finland 1981–2010", Rep., 2012:1, Finnish Meteorological Institute, Helsinki, Finland, 2012.

[46] R. Sutinen, O. Äikää, M. Piekkari and P. Hänninen, "Snowmelt Infiltration Through Partially Frozen Soil in Finnish Lapland", Geophysica, Vol., 45, No. 1–2, pp. 27–39, 2009.

[47] R. Sutinen, M. Kuoppamaa, P. Hänninen, M. Middleton, P. Närhi, S. Vartiainen and M.-L. Sutinen, 'Tree species distribution on mafic and felsic fells in Finnish Lapland", Scandinavian Journal of Forest Research, pp. 1-10, First published on: 01 December 2010 (iFirst). DOI: 10.1080/02827581.2010.534109

[48] J. Repola, "Biomass equations for Scots pine and Norway spruce in Finland", Silva Fennica, Vol. 43, No. 4, pp. 625–647, 2009.

[49] T. Majasalmi, M. Rautiainen, P. Stenberg, P. Lukeš, "An assessment of ground reference methods for estimating *LAI* of boreal forests", Forest Ecology and Management, vol. 292, pp. 10–18, 2013.

[50] P. Liwata, P. Hänninen, J. Okkonen, and R. Sutinen. "Time-stability of soil water through boreal (60-68°N) gradient", Journal of Hydrology vol. 519, pp. 1584-1593, 2014.

[51] T.K. Ho, "Random Decision Forests", Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995, pp. 278–282, 1995. https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf

[52] L. Breiman, "Random Forests", Machine Learning, vol. 45, no. 1, pp. 5 – 32, 2001. doi:10.1023/A:1010933404324.

[53] E.P.W. Attema and F.T. Ulaby, "Vegetation modeled as water cloud", Radio Science, vol. 13, no. 2, 357–364, 1978.

[54] R. Bindlish and A.P. Barros, "Parameterization of vegetation backscatter in radar-based, soil moisture estimation", Remote Sensing of Environment, vol. 76, pp. 130-137, 2001.

[55] M. Santoro, C. Beer, O. Cartus, C. Schmullius, A. Shvidenko, I. McCallum, U. Wegmüller, A. Wiesmann, "Retrieval of growing stock volume in boreal forest using hyper-temporal series ofEnvisat ASAR ScanSAR backscatter measurements", Remote Sensing of Environment, vol. 115, pp. 490 –507, 2011.

[56] P. Stenberg, T. Nilson, H. Smolander, and P. Voipio, "Gap fraction based estimation of LAI in Scots pine stands subjected to experimental removal of branches and stems", Canadian Journal of Remote Sensing, vol. 29, pp. 363–370, 2003.

[57] P. Stenberg, M. Rautiainen, T. Manninen, and P. Voipio, "Boreal forest leaf area index from optical satellite images: model simulations and empirical analyses using data from central Finland", Boreal Environment Research, vol. 13, pp. 433–443, 2008.

[58] T. Le Toan, G. Picard, J.M. Martinez, P. Melon, M. Davidson," On the relationship between radar measurements and forest structure and biomass", Paper presented at 2001 Workshop on Retrieval of Bio- and Geophysical Parameters from SAR, 2001.

[59] E. Santi, S. Paloscia, S. Pettinato, G. Fontanelli, M. Mura, C. Zolli, F. Maselli, M. Chiesi, L. Bottai, and G Chirici, "The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas", Remote Sensing of Environment, vol. 200, pp. 63-73, 2017. https://doi.org/10.1016/j.rse.2017.07.038

[60] K. P. Papathanassiou, S.R. Cloude, M. Pardini, M.J. Quiñones, D. Hoekman, L. Ferro-Famil, D. Goodenough, H. Chen, S. Tebaldini, M. Neumann, L.M.H. Ulander and M.J. Soja, "Forest Applications". In I. Hajnsek, & Y-L. Desnos (Eds.), Remote Sensing and Digital Image Processing: Polarimetric Synthetic Aperture Radar, vol. 25, pp. 59-117, 2021. https://doi.org/10.1007/978-3-030-56504-6_2

[61] T. Manninen, A. Penttilä and K. Lumme, "C-band scattering simulation of a Scots pine shoot", Waves in Random and Complex Media, vol. 17, no. 1, pp. 85-98, 2007. DOI: 10.1080/17455030601011597

[62] Manninen, T. and Jääskeläinen, E., "The effect of boreal forest canopy on snow covered terrain broadband albedo", Geophysica, vol. 53, no.1, pp. 7–27, 2018.

[63] Manninen, T., Stenberg, P., Rautiainen, M., Voipio, P., Smolander, H. and Andersson, K., "LAI estimation of boreal forest with ENVISAT ASAR", Proc. of The ENVISAT & ERS Symposium, 6-10 September 2004, Salzburg, Austria, ESA SP-572 CD-ROM, 10 p, 2004.

[64] T. Repo, "Seasonal changes of frost hardiness in Piceaabies and Pinussylvestris in Finland", Canadian Journal of Forest Research, vol. 22, no. 12, pp. 1949-1957, 1992.

[65] J. Beale, B. Snapir, T. Waine, J. Evans and R. Corstanj, "The significance of soil properties to the estimation of soil moisture from C-band synthetic aperture radar", Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-294, 2019.

[66] L. Breiman, "Arcing The Edge", Technical Report 486. Statistics Department, University of California, Berkeley, June 1997. (https://statistics.berkeley.edu/sites/default/files/tech-reports/486.pdf)

[67] J.H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine", February 1999. (https://statweb.stanford.edu/~jhf/ftp/trebst.pdf)

[68] J. H. Friedman, "Stochastic Gradient Boosting", March 1999. (https://statweb.stanford.edu/~jhf/ftp/stobst.pdf)

[69] L. Mason, J. Baxter, P.L. Bartlett, and M. Frean, "Boosting Algorithms as Gradient Descent". In S.A. Solla and T.K. Leen and K. Müller (ed.), Advances in Neural Information Processing Systems 12. MIT Press. pp. 512–518, 1999. (https://proceedings.neurips.cc/paper/1999/file/96a93ba89a5b5c6c226e49b88973f46e-Paper.pdf)

[70] L. Mason, J. Baxter, P.L. Bartlett, and M. Frean, "Boosting Algorithms as Gradient Descent in Function Space", May 1999 (https://web.archive.org/web/20181222170928/https://www.maths.dur.ac.uk/~dma6kp/pdf/face_recognition/Boosting/Mason99AnyboostLong.pdf)

[71] S.I. Seneviratne, T. Corti, E.L. Davin, M. Hirschi, E.B. Jaeger, I. Lehner, B. Orlowsky, A.J. Teuling, "Investigating soil moisture–climate interactions in a changing climate: A review", Earth-Science Reviews, Volume 99, Issues 3–4, pp. 125-161, 2010. https://doi.org/10.1016/j.earscirev.2010.02.004.

[72] Kemppinen, J., Niittynen, P., Aalto, J., le Roux, P.C. and Luoto, M. (2019), Water as a resource, stress and disturbance shaping tundra vegetation. Oikos, 128: 811-822. https://doi.org/10.1111/oik.05764

[73] Prowse, T., Bring, A., Mård, J., Carmack, E., Holland, M., Instanes, A., Vihma, T., and Wrona, F. J. (2015), Arctic Freshwater Synthesis: Summary of key emerging issues, J. Geophys. Res. Biogeosci., 120, 1887– 1893, doi:10.1002/2015JG003128.

[74] Pertti Ala-aho and Anna Autio and Joy Bhattacharjee and Elina Isokangas and Katharina Kujala and Hannu Marttila and Meseret Menberu and Leo-Juhani Meriö and Heini Postila and Anssi Rauhala and Anna-Kaisa Ronkanen and Pekka M Rossi and Markus Saari and Ali Torabi Haghighi and Bjørn Kløve, "What conditions favor the influence of seasonally frozen ground on hydrological partitioning? A systematic review", accepted for publication in Environmental Research Letters, 2021. http://iopscience.iop.org/article/10.1088/1748-9326/abe82c

[75] Wang, C., Fu, B., Zhang, L. et al. Soil moisture–plant interactions: an ecohydrological review. J Soils Sediments 19, 1–9 (2019). https://doi.org/10.1007/s11368-018-2167-0

Terhikki Manninen received the D.Sc. (Tech) degree in physics from the Helsinki University of Technology, Espoo, Finland, in 1996.

Since 2003, she has been a Senior Research Scientist with the Finnish Meteorological Institute, Helsinki, Finland. She was also the Head of the Satellite and Radar Applications Group for about 15 years until 2020. Her current research interests include the surface description of the earth, such as broadband surface albedo, snow cover and soil moisture. She is also interested in developing mathematical methods for SAR image analysis.

Emmihenna Jääskeläinen received the M.Sc. degree in 2012 from Tampere University, Finland, and she is currently pursuing the Ph.D. degree in applied mathematics at the University of Helsinki, Finland. From 2011 onwards she has been Research Scientist with the Finnish Meteorological Institute, Helsinki, Finland. Her research interests include satellite remote sensing data analysis, mathematical methods, and machine learning applications.

Annalea Lohila received her PhD in environmental sciences in the University of Eastern Finland in 2008. She works nowadays as an associate professor at the Finnish Meteorological Institute and University of Helsinki, Institute for atmospheric and Earth system research. Her research interests are atmosphere-biosphere interactions and climate impacts of land-use, particularly those related to greenhouse gas exchanges.

Mika Korkiakoski received the PhD degree in meteorology from the University of Helsinki in 2021. He is currently working as a postdoctoral researcher at the University of Helsinki. His research focuses on carbon dioxide, methane, and nitrous oxide fluxes between the atmosphere and different boreal ecosystems, such as forests and peatlands.

Aleksi Räsänen received the PhD degree in environmental science and technology from the University of Jyväskylä, Finland in 2014. He is currently a senior scientist in the Natural Resources Institute Finland (Luke). In his research, he has concentrated on remote sensing of vegetation in boreal and Arctic landscapes, and on research related to climate change adaptation and disaster risk management.

Tarmo Virtanen received PhD degree in ecology from the University of Turku, Finland in 1999. Since 2003 he has worked as University lecturer in Environmental and Ecosystem studies in the University of Helsinki. In recent years his studies have been mostly about field and remote sensing based vegetation and land cover mapping in the Arctic and northern boreal environments.

Filip Muhić received M.Sc. degree in civil engineering from the University of Belgrade, Serbia in 2015. He is currently pursuing PhD degree in environmental engineering in Northern Hydrology group at the University of Oulu. His research focuses on soil isotope hydrology and ecohydrological processes in Northern regions.

Hannu Marttila received the PhD degree (tech) environmental engineering from the University of Oulu, Finland in 2011. He currently has Assistant Professor tenure position and leading the Northern Hydrology research group. His research focuses on hydrological processes and catchment management issues in northern landscapes with strong cross-disciplinary work across science fields.

Pertti Ala-Aho received his PhD degree in Environmental Engineering in 2014 from the University of Oulu, Finland. Currently Ala-Aho is a postdoctoral researcher funded by the Academy of Finland. His research focuses on hydrological processes in Arctic and other snow-influenced regions, with special emphasis on hydrological modeling and isotope hydrology.

Raimo Sutinen (emeritus) received his PhD in Quaternary geology at the University of Oulu in 1992. His research interests are in soil hydrology and bearing capacity, site requirements by tree species and the effects of forestry tillage on soil degradation. His current investigations are focused on glacially induced faulting.