Detectability of Forest Peatlands in the Temperate Zone Using Random Forest and Open Remote Sensing Data – A Case Study of Northeast Brandenburg, Germany

Evelyn Wallor * , Hainner Aparicio , Nicole Voss , Jan-Peter Mund , Winfried Riek

Posted Date: 26 February 2026

doi: 10.20944/preprints202602.1553.v1

Keywords: organic soils in forest; Synthetic Aperture Radar (SAR); Advanced Land Observing Satellite (ALOS PALSAR); backscattering coefficient (σo); soil moisture (SM); Random Forest (RF)

 doi:10.20944/preprints202602.1553.v1

Article

Detectability of Forest Peatlands in the Temperate Zone Using Random Forest and Open Remote Sensing Data – A Case Study of Northeast Brandenburg, Germany

Evelyn Wallor [1,*], Hainner Aparicio [2], Nicole Voss [1], Jan-Peter Mund [1] and Winfried Riek [1]

[1] Department Forest, Wood & Environment; Eberswalde University of Sustainable Development; Alfred-Moeller-Str. 1; 16225 Eberswalde; Germany

[2] Skydda Skogen – Protect the Forest; Goetlind; Arlidsvaegen 31; 441 91 Alingås; Sweden

* Correspondence: evelyn.wallor@hnee.de

Highlights

What are the main findings?

- Soil moisture from L-band SAR shows high variable importance, and terrain indices are strongly linked to the formation of organic soils in forests.
- Resampling and random sampling effectively reduce data imbalance, while forest cover type can enhance the predictive accuracy of RF models.

What is the implication of the main finding?

- The method provides a transferable and operational framework for large-scale mapping of organic wet soils in temperate forests using satellite-based L-band SAR data.
- Improved detection of organic wet soils supports more efficient monitoring and conservation of peatland ecosystems under forest cover, which are critical for carbon storage and hydrological regulation.

Abstract

Environmental services of mires and peatlands and negative impacts of their alteration are comprehensively documented. A spatial detection of these organic soils is therefore essential. This paper examines the detectability of organic soils in forests using open geospatial and remote sensing data combined with mapped soil and water level information in two Random Forest (RF) approaches. Either surrounded in or covered by forest, organic soils of the study region exhibit elevated soil water content reaching saturation during the hydrological winter. Consequently, terrain indices from Digital Elevation Models (DEM) and soil moisture from L-band ALOS PALSAR signals are used as predictors in RF algorithms. CORINE Land Cover data help assess how different forest cover types (FCT) influence RF models. Substantial agreement is reached in the target classification when FCT is included and when using the higher spatial resolution DEM. The Boolean approach is less affected by different compositions of predictor variables, but is more sensitive to the level of imbalance in the reference data. This becomes evident when comparing the "event error" and "no event error". In all RF models, the variable importance of soil moisture pixel values retrieved from L-band ALOS PALSAR is the highest when FCT is not included.

Keywords: organic soils in forest; Synthetic Aperture Radar (SAR); Advanced Land Observing Satellite (ALOS PALSAR); backscattering coefficient (σ^o); soil moisture (SM); Random Forest (RF)

 doi:10.20944/preprints202602.1553.v1

1. Introduction

Organic soils, like mires, peatlands and gleyic soils, are ecosystems of vital importance due to their function on carbon sequestration, methane emission, water cycle regulation and biodiversity preservation [1–4]. Environmental services and negative impacts of amelioration of organic soils have been well documented [5–9]. Organic wet soils have acted as sinks for atmospheric carbon dioxide (CO2) for millennia and still today, they store more than one third of the global soil organic carbon [10]. We understand them as wetland ecosystems in which waterlogged conditions prevent plant material from fully decomposing. Consequently, the production of organic matter exceeds its decomposition, which results in a net accumulation of peat and carbon [1,2].

In Europe, many wetland ecosystems have become degraded during the last century due to drainage and amelioration, burning, pollution and climate change [11]. Locating and quantifying organic soils spatially at regional scale is an essential first step to monitor and manage them in an optimal manner. In Germany, a multitude of different data sources on organic soils is already available at smaller scales (> 1:50,000). Due to the Federal State structure of Germany, these data sources originate from a multitude of institutions and differ in input data, classification type, timeliness, and conformity with the Intergovernmental Panel on Climate Change (IPCC) definitions leading to fragmented, heterogeneous, and sometimes obsolete information [12,13]. The process of collecting contemporary in situ data on the location of organic soils in forests can become exceedingly arduous and resource-intensive in terms of time and labour. This is primarily due to the fact that these soils are characterised by a waterlogged environment and dense herbaceous vegetation, which significantly hinders accessibility for human investigation. The peat-forming mires, which are still relatively undisturbed, can be found in Germany, typically in wooded areas, due to their often-diminutive size (ranging from 1 to 20 hectares) and challenging accessibility. Nevertheless, the majority of forest peatlands are degraded and drained, thus ameliorated, primarily for agricultural and forestry purposes. Evidence indicates that the key ecological functions of forest mires are at risk of being depleted or destroyed [14]. These facts, as well as the availability of remotely sensed radar data products and machine learning methods, encouraged the authors to test open geospatial data products as an efficient and less resource-demanding alternative for detecting organic soils under forest showing different water regimes.

Freely available geospatial data that have been identified to potentially map characteristics of the geographical occurrence of organic soils under forest are: i) the Phased Array L-band Synthetic Aperture Radar (ALOS PALSAR) distributed for example by the Active Archive Centre of Alaska Satellite Facility [15], ii) the Shuttle Radar Topography Mission (SRTM) global Digital Elevation Model (DEM) to 1arc-second resolution [16,17] as well as the regional DEM at 25 m resolution [18], and iii) CORINE Land Cover Classes (CLC) based on multi-spectral satellite imagery and provided by Copernicus Land Monitoring Services [19].

The Random Forest (RF) model is known as an “ensemble” machine learning algorithm that conjugates two or more of what is called “weak learners” to generate a stronger one. It is capable of multivariate calibration, which is exactly the case in this study as it aims to detect soil properties based on remote sensing and ancillary data. RF is also known to be computationally efficient, easy to parameterize, accurate, stable and enables practical means to rank predictors based on their influence on the classification [20,21]. The RF classifier has been used successfully for wetland mapping in several studies [22–25].

The use of topographical indices for modelling the occurrence of organic soils has become established as part of the increasing provision of open geospatial data and the improvement of technical application possibilities [26,27]. Topographical indices such as elevation and slope play an important role, especially in connection with the modelling of hydrological properties that are essential for peat formation, e.g., groundwater dynamics, water balance [28]. The post-glacial morphological character of the study area and the presence of different hydrogenetic mire types (e.g., spring mires, kettle-hole mires, paludification mires etc.) favour the use of digital terrain models as explanatory variables in the machine learning model [14].

 doi:10.20944/preprints202602.1553.v1

Active remote sensing sensors including Synthetic Aperture Radar (SAR) have been proven capable of detecting and classifying wetland areas and their landscape dynamics in the tropical, temperate, and boreal zone [29–34]. The reflected electromagnetic radiation sensed as backscattering coefficient soundly responds to soil moisture through sparse vegetation [35]. The penetration of a microwave into the soil depends intrinsically on its wavelength [36–39]. Shorter wavelengths (e. g. C-band SAR with a wavelength of ~ 56 mm) respond more to soil roughness or superficial texture on bare soils and to vegetation density, orientation and texture in areas covered by low vegetation, for example in agricultural sites [40]. As highlighted by Wagner et al. [36,37], longer wavelengths (e. g. L-band SAR with a wavelength of ~ 236 mm) are less responsive to small obstacles and soil texture and allow measurements of the real dielectric constant (RDC) which is a reliable indicator of water content in soils, vegetation or in any other object or structure.

Spatial location of organic soils under vegetation has been estimated and mapped already from backscattering coefficient in several studies [41–44]. These studies have shown promising correlation between backscatter coefficient and water table depth in wetlands. Kim et al. [45] and Bechtold et al. [46] found encouraging results regarding the correlation between soil moisture, water table depth and L-band radar backscatter for a small number of observations in wetlands and some local peatlands in forests. C-band Sentinel-1 data has been applied for soil moisture estimation in organic soils [44] as well as in mineral soils [47]. However, the L-band radar signal is particularly suitable for observations characterized by dense vegetation cover, which can lead to strong attenuation of C-band radar signals [48].

Significant innovative methods on the application of L-band polarimetric data for peatland mapping and classification in tropical and in boreal forests have been contributed by several researchers [49–52], among others. However, these studies focused on peatlands in agricultural areas as well as in the boreal or tropical forests. To the best of the authors' knowledge, research focusing on the detection of organic soils with different water regimes in temperate forests at a high spatial resolution has not yet been carried out. The objective of this study is to propose an innovative approach for the detection of organic soils across diverse water regimes within the context of temperate forest regions. It encompasses an evaluation of the efficacy of L-band ALOS PALSAR data in conjunction with digital elevation models (DEM) and CORINE land cover data (CLC), as a potential tool for this purpose. For this study, we consider a study region in north-eastern Germany where organic soils are an abundant feature, as they are across Northeast Europe.

2. Materials and Methods

2.1. Study Region and Forest Land Cover

A study region spanning approximately 7,562 km², situated in the north-eastern area of the Federal State of Brandenburg (Figure 1), has been selected for the purpose of evaluating a novel, methodical approach. The landscape in the region has been formed by glacial and postglacial geomorphologic processes and is dominated by a typical lowland, mostly flat to smoothly undulated topography. Elevations range from 10 m to 160 m above sea level with short and steep slopes and aeolian fossil dunes located onto the postglacial surface. The lowlands were formed by meltwater channels and glacial valleys. The development of mires was influenced by the presence of groundwater, and the subsequent evolution of these wetlands varied according to specific hydrological conditions. According to DWD Climate Data Centre [53,54], the mean monthly precipitation ranked around 40 mm and mean monthly temperature amounted 5.3 °C within the six months timespan from November 2008 to April 2009 which is the period in which reference data was collected and SAR imagery was acquired. According to CORINE Land Cover (CLC) data from 2012 [19], the dominant land cover classes were agricultural land and forests. The latter covers 36% of the study region and composes of coniferous forests (24.5%), broadleaved forests (8%) and 3.5% of mixed forest (Figure 1). The land cover data utilized in this study was obtained from 2012 due to its proximity to the reference data collection and imagery acquisition date. The analysis in the this study

 doi:10.20944/preprints202602.1553.v1

is focussing on forest areas in the study region, consequently CLC 2012 forest land cover data was utilized to mask reference points located within respective forest areas.

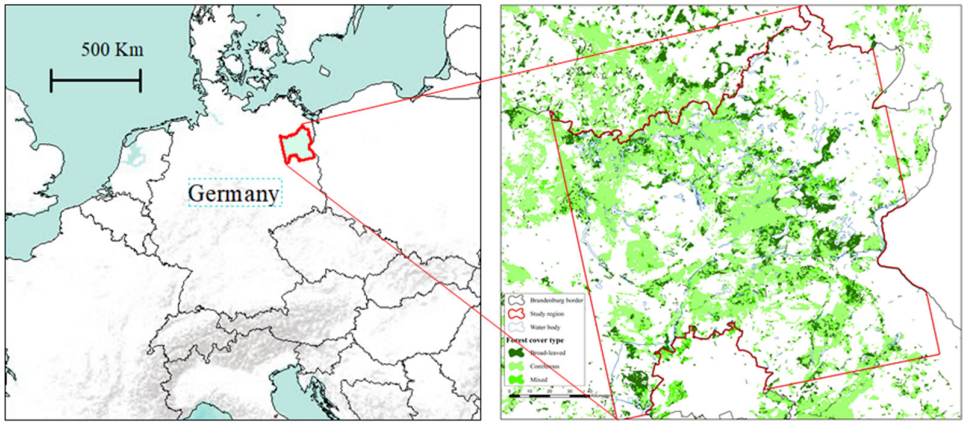


Figure 1. Study region in north-eastern Brandenburg, Germany.

2.2. Methodological Approach

This section describes the approach for the creation of the reference data set, the data pre-processing steps and the soil moisture (SM) retrieval from backscatter (σ^o). Also, the Random Forest (RF) model establishment and implementation as well as the applied accuracy assessment is introduced. The workflow diagram in Figure 2 illustrates the general approach that is presented in the next pages.

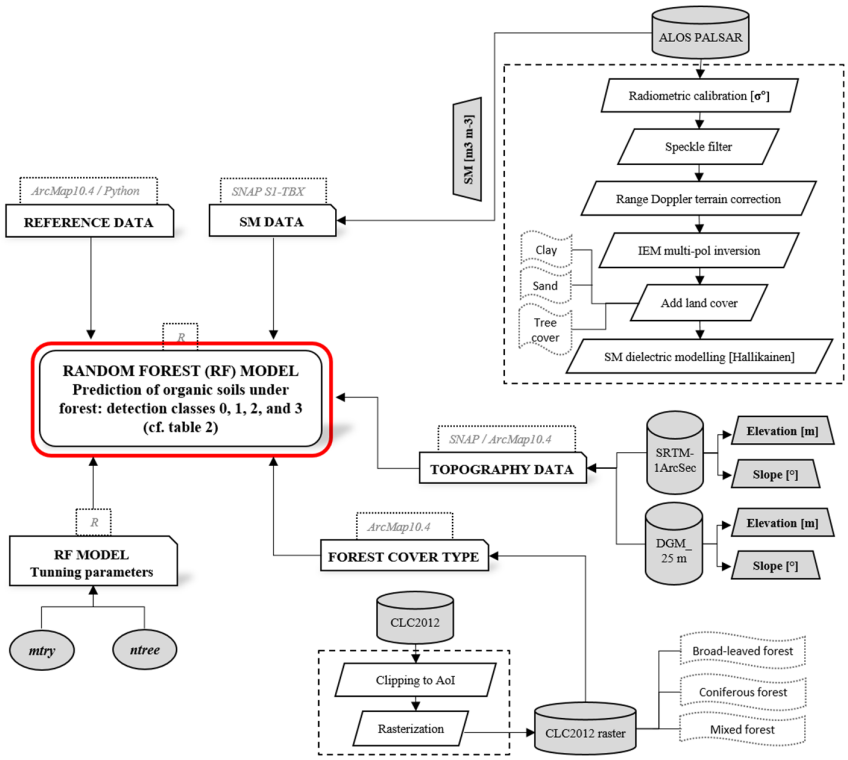


Figure 2. Workflow of the methodological approach; red-marked the RF models.

 doi:10.20944/preprints202602.1553.v1

2.2.1. Reference Data & Reference Data Preparation

We used the following ecological data sets as independent reference data in our study:

1. Pedo-ecological data of the Brandenburg-wide forest site assessment [55,56]
2. Pedological Data of the second National Forest Soil Inventory (NFSI-II) (2006 to 2009) [57,58]
3. Spatial interpolated groundwater level data of Brandenburg (reference year 2008) [59]

The Brandenburg-wide forest site assessment (1.) provides a classification of soil type including soil moisture level and is available on a scale of 1:10,000. The NFSI (2.) provides detailed soil profile information and measurement for 212 inventory points across Brandenburg. The Brandenburg groundwater level information (3.) is provided in a regular 100 x 100 m grid. Based on this groundwater grid and using the above data, Russ et al. [57,60,61] have developed a regionalization model that results in high-resolution soil information for the entire area of Brandenburg. In the present study, exclusive consideration is given to the groundwater grid point information (3.) that has been observed in the forest and has been identified as having the optimal prediction quality following Benning et al. [62]. Consequently, only analysed soil profile data and soil profile data derived from maps or digital soil mapping approaches have been used as reference data. Table 1 presents an overview of the combined soil classes and groundwater level classes to be detected in this study using the RF algorithm in combination with open geo-data. Thus, the reference data is split for training (70%) and testing (30%) the defined RF models (cf. section 2.2.4).

Table 1. Overview of combined soil classes and groundwater level detection classes (below soil surface) including the sample size of respective grid points (= data points) clipped by the study region.

Soil type	Soil class	Groundwater level [1] in [m]	Detection class [2]	N
Organic wet soils (Peat soils)	O1	>0.0	3	790
	O2	0.0-0.2	3	150
	O3	0.2-0.5	3	546
	O4	0.5-1.0	1	2,384
Mineral wet soils (Gleyic soils)	N0	<0.2	2	109
	N1	0.2-0.5	2	660
	N2	0.5-1.0	1	3,313
	N3	1.0-1.8	1	3,969
Mineral dry soils	T1	1.0-1.8	0	6,206
	T2g	1.8-3.0	0	10,908
	T2/3	>3.0	0	101,018

[1] Groundwater level for soil class O4 and N3 caused by groundwater lowering (amelioration). [2] Classes to be detected by ALOS-PALSAR.

In accordance to the objective of this study, the detection classes to be identified by the RF classifiers were reorganised on the basis of the following: i) soil class and groundwater level information, ii) the fact that depending on groundwater level, capillary rise influences soil moisture in respective (organic) soil layers, and iii) the capability of the SAR L-band frequency microwaves to penetrate soil depths up to approximately 80 mm below surface level. According to Table 1, detection class 3 refers to organic wet soils, detection class 2 to mineral wet soils, detection class 1 to intermediate wet soils (organic or mineral), and detection class 0 to sites far from groundwater, respectively. Figure 3 shows all grid points classified as organic wet soils O1 to O3 (= detection class 3) located in the federal state of Brandenburg.

Subsequent to the definition of the detection classes, the next stage of the process was to prepare and balance the reference data set. In order to eliminate any potential influence of lateral water fluxes on the recorded backscatter coefficient (σ^o), reference data points within 200 m to water bodies were subtracted. Subsequently, reference data points within a buffer of 100 m from artificial surfaces, urban areas, bare soil, and agricultural land, were subtracted to avoid edge effects altering the σo values

 doi:10.20944/preprints202602.1553.v1

recorded by the SAR sensor due to the influence of such land covers on the L-Band wavelength [63,64].

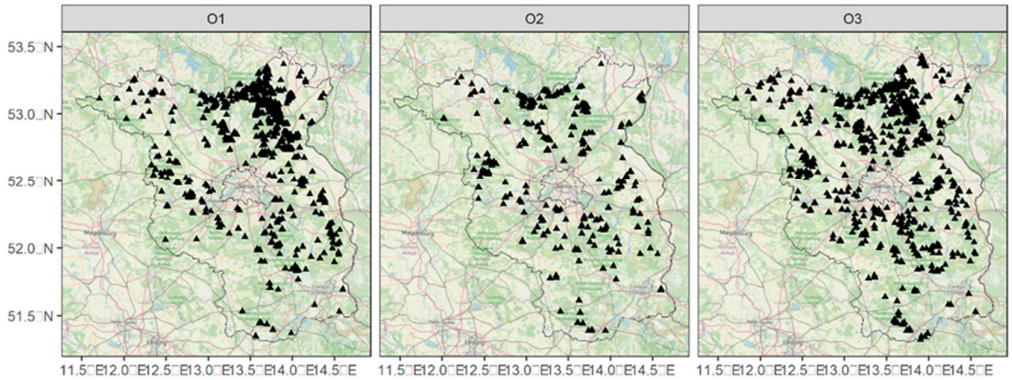


Figure 3. Occurrence of organic wet soils (O1 to O3) of detection class 3 in Brandenburg.

The RF algorithm is built on decision trees foundations, and decision trees are sensitive to class imbalance (cf. Table 1, column "N"). The RF classifier shows better results when it is trained with data sets where the number of data points per class is more balanced [65]. Therefore, a last data pre-processing step is recommended. An augmentation of the number of points from low occurrence classes (detection class 1, 2, and 3; cf. Table 1) was achieved through the process of resampling, whilst a specific percentage of the data points with a high occurrence class (detection class 0; cf. Table 1) were systematically selected for randomisation, in order to diminish the predominance of the latter. Table 2 presents the four created reference data sets including different levels of random sampling, resampling and edge effect masking. The four data sets, which differ with regard to their balancing tuning, form the basis of the construction of RF models.

Table 2. Balancing of reference data set.

Data set Id	Random sampling % high occurrence class (0)	Resampling low occurrence class (3;2)	Edge effect masking
DS_Tune_1	0.10	FALSE	FALSE
DS_Tune_2	0.15	TRUE	TRUE
DS_Tune_3	0.25	TRUE	FALSE
DS_Tune_4	0.20	TRUE	TRUE

The effect of the balancing of under-represented detection classes is shown exemplarily in Figure 4.

2.2.2. Processing of Open Geo-Data (SAR, DEM, CLC)

Several standardized geo-data products with almost global coverage are available from passive and active microwave observations, which are frequently utilized to detect soil moisture in the upper soil layers [66–68]. However, their spatial resolution is often too coarse to accurately detect and represent the high variability in medium and small sized wetlands [14]. Authors obtained nine images from L-band ALOS PALSAR polarimetric operational mode (PLR) via the Alaska Satellite Facility, distributed by the Active Archive Centre [15]. The products were acquired at quad polarization, covering a timespan from November 2008 to April 2009, collected in ascending pass direction. Images of periods with snow cover, dew, or precipitation during the last six hours prior to sensor data acquisition, were discarded. Such moist or wet conditions could lead to nonlinear microwave signal dynamics, alterations and disturbances, reducing their capability in monitoring

 doi:10.20944/preprints202602.1553.v1

topsoil moisture [69]. In order to identify such conditions, data and time of image detection and the hourly precipitation records from 16 climate stations in the study area were analysed [53,54].

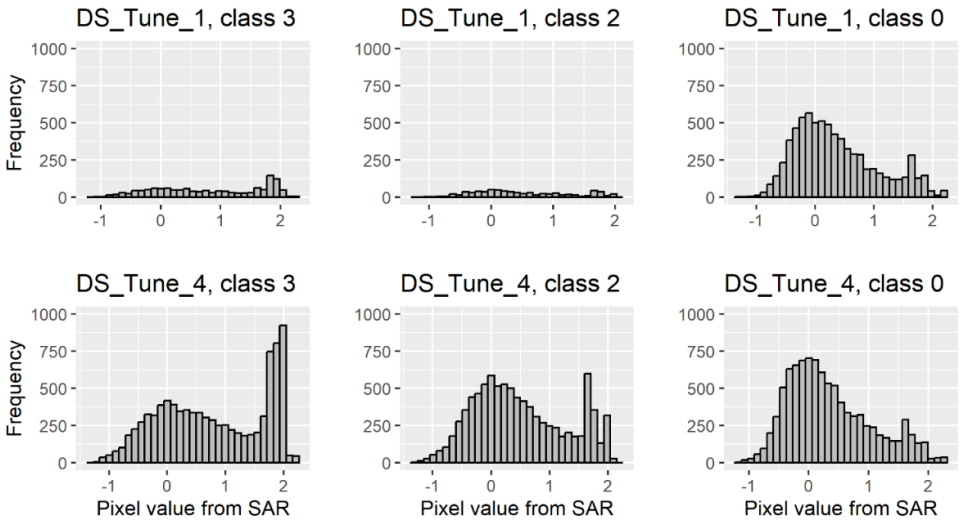


Figure 4. Absolute frequencies of data points per reference data set and detection class before (upper row) and after balancing (bottom row).

The Shuttle Radar Topography Mission (SRTM) global digital elevation model (DEM) with 1 arc-second (= 30 m pixel resolution) was integrated to pre-process the SAR images [16]. Furthermore, a regional topographic DEM of 25 m resolution (cf. section 2.2.4) was used as the source for topographic indices [18]. Such indices can significantly improve the general detection capability of SAR data [22,23]. This is particularly true in flat or geomorphologically less active landscapes, where the development and spatial location of organic soils is highly dependent on local topographic features that control water movement and sedimentation [70]. CLC data were used to examine the impact of forest cover type on the occurrence of forest organic soils. It is part of the Copernicus Land Monitoring Service and the 2012 CLC version used in this study is based on Landsat and SPOT data and has a spatial resolution of 100 m * 100 m. For the spatial assignment of the FCT, the data set had to be rasterised (cf. Figure 2). Hence, the spatially assigned forest cover type (deciduous, coniferous and mixed forest) was included as a predictor in the RF algorithm.

2.2.3. Soil Moisture Retrieval from SAR

The SAR bands of the selected images were calibrated to σo using the SNAP Sentinel 1 Toolbox from the European Space Agency [17]. Geometric correction using SRTM DEM, radiometric terrain correction, and 5*5 Sigma-Lee speckle filtering were applied to enhance contrast and lower random aspects. The relationship between SAR images σo and physical parameters used to characterize soil moisture have been studied by analysing various physical, semi-empirical and empirical radar backscattering models [71]. The most frequently used models are the Integral Equation Model (IEM) proposed by Fung et al. [72], the Advanced Integrated Equation Model (AIEM) by Wu et al. [73], and semi-empirical calibrated IEM for C-band and L-band SAR [74,75].

Algorithms have been developed to reduce the complexity of IEM application. They are based on fitting of IEM numerical simulations for a wide range of surface conditions [76,77]. These simplified IEM-based algorithms require fewer parameters and are much easier to use with remotely sensed data [78]. The result is a look-up table (LUT) of IEM simulations that serves to directly relate SAR σo to theoretical model predictions with known radar parameters. In this study, the IEM multi-polarization inversion has been used to obtain the real dielectric constant (RDC). The LUT used to feed the IEM was obtained from the soil moisture (SM) retrieval module in SNAP Sentinel 1 Toolbox, while all parameters were set to default [17]. Afterwards, a land cover model provided in SNAP was

 doi:10.20944/preprints202602.1553.v1

applied, including clay, sand, and MODIS-2010 tree cover percentage layer. To reach the desired output, the linear power was converted to retrieved SM following Hallikainen et al. [79]. A quality index was obtained in parallel along with the process of SM retrieval. This index supports the understanding of the quality and therefore accuracy of the calculation by assigning a quality value to each pixel of retrieved SM.

2.2.4. Random Forest Classification (Model Set-Up & Validation)

In this study, RF models were developed and applied to classify data points into the four described detection classes (cf. section 2.2.1, Table 1). Thereby, models differ according to the balancing procedure as well as the selected predictors. The summary of the predictors that were obtained from the different data sources are displayed in Table 3.

Table 3. Predictors obtained from the different data sources.

Source of predictor	Predicting variable
L-Band ALOS PALSAR	Soil moisture (m^3 m^{-3})
DEM SRTM 1arcSecHGT (30 m)	Elevation (m)
	Slope (°)
Regional DEM 25 × 25 m	Elevation (m)
	Slope (°)
Corine Land Cover 2012	Forest cover type (coniferous, broad leaved, mixed)

In this context, RF models were developed that on the one hand predict each of the four detection classes (target classification) and on the other hand use a Boolean approach. The latter then treat detection class 3 (organic wet soils) or classes 2 and 3 (organic and mineral wet soils) in the RF models as TRUE events (or "event"), whereas all other detection classes are treated as FALSE events (or "no event"). It is important to note that, in order to compare the influence of the two sources of topographic variables, it was necessary to set up independent RF models (Table 4).

Table 4. Random Forest model set-up.

Data set Id	DEM used	RF model Id
DS_Tune_2	DEM SRTM	*RFmdl_2*
	DEM 25	*RFmdl_3*
DS_Tune_4	DEM SRTM	*RFmdl_6*
	DEM 25	*RFmdl_7*

RF models were set up using the R software [80] in connection with the *randomForest* package [81]. The two main tuning parameters in the generated RF models were: a) the number of predictors randomly sampled as candidates at each split or tree node (*mtry*), and b) the number of classification trees (*ntree*) [82]. We tested these parameters iteratively by letting *mtry* one by one run through all the predictors. *ntree* was determined in an ascendent sequence every 25 trees up to a maximum of 200. A partition index of 0.7 train / 0.3 test was set and resampling of 25 replications bootstrapping across hyperparameters was implemented. No pruning of trees was set due to the small number of predictors. In addition, previous studies reported that bootstrap sampling that grows all trees to maximum size without pruning makes RF less susceptible to overfitting [83–85]. After the different RF models were set, trained, and tested, accuracy assessment matrices were calculated. The Kappa value as one output of the accuracy matrix was used to assess the quality of the RF models set up,

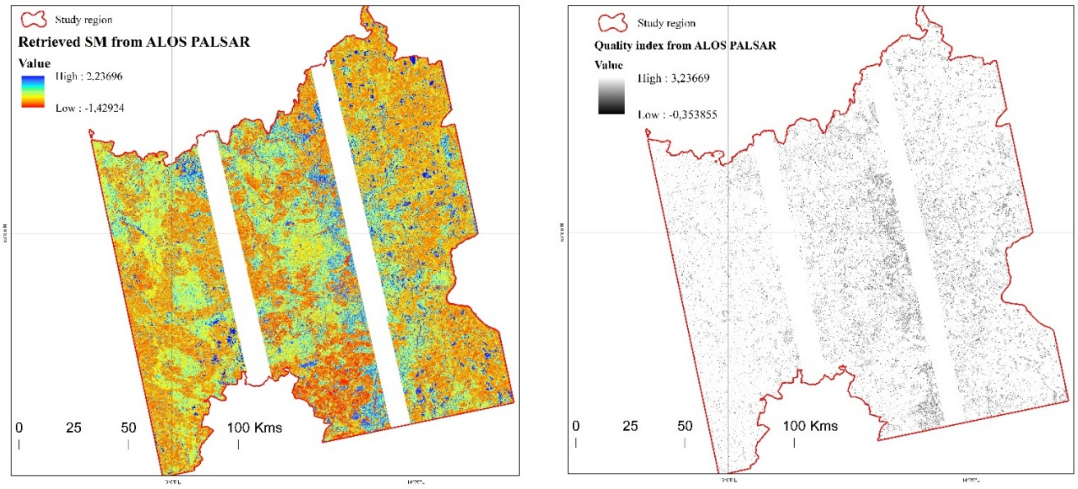
 doi:10.20944/preprints202602.1553.v1

and interpreted according to [86]. The importance of the predictors was evaluated based on the mean decrease in accuracy (MDA) criterion [23].

3. Results

3.1. Soil Moisture (SM) Values

From L-band ALOS PALSAR imagery, RDC (Farad m-1) was calculated, and from it, SM was retrieved (cf. section 2.2.3). The latter ranged from 0.029 to 0.55 m3 m-3 before histogram stretching. Along with it, the calculation quality index was obtained. Authors like to note that during the visual analysis of the SAR images, a very slight displacement in the edges of the tiles was perceived. The resulting maps of retrieved SM and quality index can be observed in Figure 5a and 5b, respectively.



(**a**) (**b**)

Figure 5. (a) Soil moisture retrieved from SAR (ALOS PALSAR), and **(b)** Quality index results.

3.2. RF Classification & Validation (Target Classification)

The results of the accuracy assessment of the RF classification are presented for two scenarios: once when FCT was included as a predictor and once without. The error rates per detection class indicate the proportion of incorrectly classified instances per class and vary more between the different RF models when FCT was not included as a predictor (Figure 6). In both cases, the error rates are the highest for detection class 1 (intermediate wet soils), independently of the DEM and the percentage of random sampling of the high occurrence class used in the RF models. The low error rate for class 0 (mineral dry soils) increases significantly when the percentage of random sampling was increased from 15% to 20% (RFmdl_6, RFmdl_7). However, the error rate also decreases for this class when the FCT is included in the models as a predictor. The error rates for detection classes 2 and 3 are in the middle of the range, between 0.31 and 0.41, but are more balanced when FCT is included as a predictor variable in the models. If FCT is not included, RFmdl_7 performs best for detection class 3 (0.31). Along with the error rates per detection class, a clear pattern was found regarding the performance of the models as a response to the utilized DEM. Models in which the regional DEM (25 m pixel resolution) was used, showed superiority in terms of detection over SRTM DEM (30 m pixel resolution). This pattern can be observed in a consistent manner also in the attempts where the FCT was included as a predictor.

Table 5 shows a comparison of the confusion matrix of the four RF models, whereby the detection class-specific results for detection classes 3 (organic wet soils) and 2 (mineral wet soils) are highlighted as the target classes of the present study.

 doi:10.20944/preprints202602.1553.v1

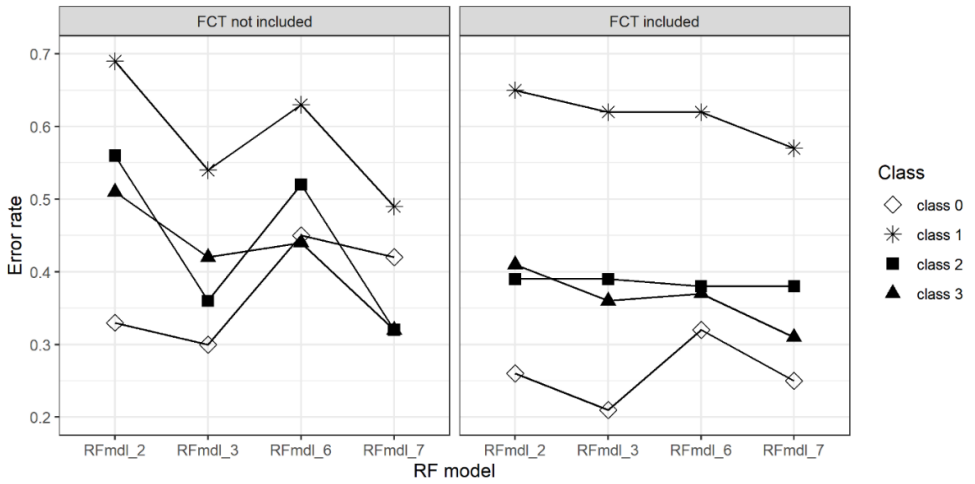


Figure 6. Error rate per detection class for different RF models; FCT not included (left) and FCT included as predictor (right).

Table 5. Accuracy assessment results detection class 3 and 2 (Kappa values in bold correspond to moderate to substantial agreement between 0.41 and 0.80).

Assessment indicator	Class	FCT not included				FCT included			
		RFmdl_2	RFmdl_3	RFmdl_6	RFmdl_7	RFmdl_2	RFmdl_3	RFmdl_6	RFmdl_7
Overall Accuracy	-	0.52	0.63	0.50	0.63	0.59	0.71	0.59	0.71
Kappa	-	0.34	0.49	0.34	0.50	0.44	0.61	0.45	0.61
Sensitivity (TPR)	3	0.50	0.62	0.57	0.70	0.59	0.74	0.65	0.77
Specificity (TNR)	3	0.88	0.89	0.86	0.87	0.90	0.92	0.89	0.90
Precision (PPV)	3	0.55	0.61	0.58	0.64	0.64	0.72	0.67	0.72
Prevalence	3	0.22	0.22	0.26	0.26	0.22	0.22	0.26	0.26
Balanced Accuracy	3	0.69	0.75	0.71	0.78	0.75	0.83	0.77	0.84
Sensitivity (TPR)	2	0.47	0.68	0.48	0.71	0.61	0.75	0.62	0.76
Specificity (TNR)	2	0.84	0.90	0.82	0.88	0.85	0.93	0.83	0.92
Precision (PPV)	2	0.46	0.65	0.48	0.64	0.53	0.74	0.55	0.76
Prevalence	2	0.22	0.22	0.26	0.26	0.22	0.22	0.26	0.26
Balanced Accuracy	2	0.66	0.79	0.65	0.78	0.73	0.84	0.72	0.84

The overall accuracy and Kappa values provide an insight into the overall performance of the RF models. Sensitivity (True Positive Result), specificity (True Negative Result), precision (Positive Predictive Value), prevalence and balanced accuracy are presented per detection class. The lowest overall accuracy and Kappa values are found for the RFmdl_2 and RFmdl_6 implementation in case FCT is not included as a predictor variable and indicate a fair agreement (Kappa ≥ 0.21 and ≤ 0.40). RFmdl_3 and RFmdl_7 show an increase in terms of overall accuracy, and Kappa values in both cases

 doi:10.20944/preprints202602.1553.v1

(FCT not included/included) exceed the limit for a moderate (Kappa ≥ 0.41 and ≤ 0.60) and even substantial agreement (Kappa ≥ 0.61 and ≤ 0.80). These results show high concordance with the error rates results per detection class (cf. Figure 6). It is evident that the disparity in the randomised sampling percentage of the high occurrence class appears to have a marginal influence on the comprehensive enhancement of the RF models.

This pattern is also observed when looking at the detection class specific accuracy indicators in the matrix. In the case that FCT is not considered as a predictor and using the SRTM DEM as an input variable (RFmdl_2, RFmdl_6), the class-specific indicators show a slightly better detectability of detection class 3 compared to detection class 2. This results in a higher value for the balanced accuracy, 0.69 and 0.71, opposite 0.66 and 0.65. The reverse is true if the regional DEM with a higher resolution is used as an input variable (RFmdl_3, RFmdl_7). This results in better detectability for detection class 2 with predominantly higher values for the sensitivity, specificity, precision and balanced accuracy. If the FCT is took into account as a predictor in the RF models, the pattern becomes more heterogeneous in terms of the detectability of the two target classes. Here, the detection class-specific accuracy values are also higher when the regional DEM is used (RFmdl_3, RFmdl_7), although detection class 2 appears to be slightly better predicted than detection class 3. Using the SRTM DEM (RFmdl_2, RFmdl_6), the results are even more diverse: sometimes RFmdl_2, sometimes RFmdl_6 shows a better performance, depending on which accuracy value is considered. Therefore, authors emphasise here the significance of using DEM data with a high spatial resolution.

The impact of the various balancing methods on the RF model performance is found to be negligible when considered against the reference data sets. Only in terms of sensitivity, the values increase in the case of detection class 3 and irrespective of the consideration of FCT as a predictor. As prevalence describes the proportion of the instances of a specific detection class, it is independent of the predictors involved. Therefore, prevalence is identical for the respective models of the same random sampling percentage.

3.3. RF Classification & Validation (Boolean Approach)

Specifying either detection class 3 or detection classes 2 and 3 as a positive, true event in the Boolean approach results in the following accuracy assessment values. Thereby, Table 6 shows the overall performance indicators when FCT is included, and Table 7 when FCT is not included as a predictive variable. Again, consideration of FCT improves the prediction accuracy of RF models, which is confirmed by lower Out-of-Bag (OOB) error rates in all model variants. It is observed, that the error rate for "no event" (detection classes 0, 1, and 2 = FALSE) is minimal, while it is between 43% and 60% for "event" (detection class 3 = TRUE). This discrepancy is levelled out if instances of detection classes 2 and 3 are considered together as a positive event. Concurrently, the OOB error rate exhibits a marked increase when FCT is omitted, as evident in Table 6. When FCT is included, the increase of OOB error rate is less (Table 7). A comparison of the different DEM data sources used in the RF models shows that the error rates are generally slightly reduced when using the regional DEM of 25 m pixel resolution.

Table 6. Error of event prediction for different RF models (FCT not included as predictor).

Model/ Error	event of success	RFmdl_2	RFmdl_3	RFmdl_6	RFmdl_7
OOB error	class 3	0.18	0.16	0.20	0.17
error no event	class 3	0.06	0.06	0.08	0.08
error event	class 3	0.60	0.51	0.54	0.43
OOB error	class 2 & 3	0.36	0.28	0.39	0.29
error no event	class 2 & 3	0.28	0.21	0.38	0.29
error event	class 2 & 3	0.47	0.37	0.39	0.30

 doi:10.20944/preprints202602.1553.v1

Table 7. Error of event prediction for different RF models (FCT included as predictor).

Model/ Error	event of success	RFmdl_2	RFmdl_3	RFmdl_6	RFmdl_7
OOB error	class 3	0.16	0.15	0.17	0.16
error no event	class 3	0.04	0.06	0.06	0.07
error event	class 3	0.56	0.44	0.50	0.41
OOB error	class 2 & 3	0.26	0.24	0.28	0.26
error no event	class 2 & 3	0.24	0.21	0.29	0.25
error event	class 2 & 3	0.29	0.28	0.27	0.27

Compared to the target classification (cf. section 3.2), the Boolean approach produces a higher overall accuracy and higher Kappa values (Table 8) in case detection class 3 is defined as TRUE. Similar to the target classification, a clear improvement of the RF models' performance can be stated when FCT is included as predictor. Furthermore, the positive impact of the higher spatial resolution of the DEM data source becomes visible. Hence, RFmdl_2 and RFmdl_6 produce Kappa values of 0.48 and 0.51 (FCT included), while RFmdl_3 and RFmdl_7 Kappa values are 0.62 and 0.67. The range of the Kappa values shifts downwards if detection classes 2 and 3 are jointly defined as a positive event, namely from 0.37 (RFmdl_2, FCT not included) to 0.67 (RFmdl_7, FCT included) to 0.28 (RFmdl_2, FCT not included) to 0.62 (RFmdl_7, FCT included).

Table 8. Accuracy assessment results for different RF models of the Boolean approach; event of success = detection class 3, event of success = detection class 2 & 3 (Kappa values in bold correspond to moderate to substantial agreement between 0.41 and 0.80).

Assessment	Event	FCT not included				FCT included			
		RFmdl_2	RFmdl_3	RFmdl_6	RFmdl_7	RFmdl_2	RFmdl_3	RFmdl_6	RFmdl_7
Overall Accuracy	3	0.81	0.83	0.80	0.84	0.84	0.87	0.83	0.88
Kappa	3	0.37	**0.47**	0.41	**0.54**	**0.48**	**0.62**	**0.51**	**0.67**
Sensitivity (TPR)	3	0.38	0.48	0.45	0.61	0.50	0.66	0.53	0.72
Specificity (TNR)	3	0.94	0.93	0.92	0.91	0.93	0.94	0.93	0.94
Precision (PPV)	3	0.64	0.68	0.65	0.70	0.69	0.75	0.72	0.79
Prevalence	3	0.23	0.23	0.25	0.25	0.23	0.23	0.25	0.25
Balanced Accuracy	3	0.66	0.71	0.68	0.76	0.72	0.80	0.73	0.83
Overall Accuracy	3 & 2	0.64	0.73	0.62	0.72	0.75	0.81	0.73	0.81
Kappa	3 & 2	0.28	**0.44**	0.24	**0.45**	**0.48**	**0.61**	**0.45**	**0.62**
Sensitivity (TPR)	3 & 2	0.54	0.68	0.61	0.75	0.71	0.78	0.74	0.83
Specificity (TNR)	3 & 2	0.73	0.76	0.62	0.70	0.78	0.77	0.74	0.83

 doi:10.20944/preprints202602.1553.v1

Precision (PPV)	3 & 2	0.61	0.70	0.63	0.72	0.72	0.79	0.72	0.81
Prevalence	3 & 2	0.45	0.45	0.51	0.51	0.45	0.45	0.51	0.51
Balanced Accuracy	3 & 2	0.63	0.72	0.62	0.72	0.74	0.80	0.73	0.81

The sensitivity (TPR) generally increases when detection classes 2 and 3 are jointly defined as positive event. In parallel, specificity (TNR) decreases. The average of both indicators, resulting in the balanced accuracy, remains higher in case detection class 3 is exclusively defined as TRUE. The same applies for the proportion of positive predictions that were actually correct (precision). According to the differences in data sets (DS_Tune_2, DS_Tune_4) used in the RF models as well as changing number of true instances, the prevalence varies. All in all, the confusion matrix in Table 8 confirms what has been stated previously: It is evident that the incorporation of FCT in conjunction with the regional high-resolution DEM has led to a substantial enhancement in the performance of RF models. The benefit of the marginal increase of the high occurrence class during the balancing procedure is only minor.

3.4. Importance of Predictors

Understanding the importance of predictors on the performance of the models is a fundamental part of the results (Figure 7). First, when FCT is not included as predictor, the pixel value retrieved from SAR backscatter was the most important predictor throughout all the models tested apart from RFmdl_6. The second most important predictor ranking was shared by elevation and slope. However, it is worth mentioning that depending on the DEM used, slope is more important compared to elevation and vice versa. For example, elevation is higher rated in case the SRTM DEM is used as a predictor (RFmdl_2, RFmdl_6). In case of RFmdl_6, elevation is even more important than the soil moisture retrieved from SAR backscatter. With increasing spatial resolution, the slope becomes more pronounced compared to the elevation. In total, the importance of predictors shifted significantly towards FCT when the RF models were trained and tested including it. The pixel value obtained from SAR backscatter was the second most important predictor for the RF models trained with the regional DEM.

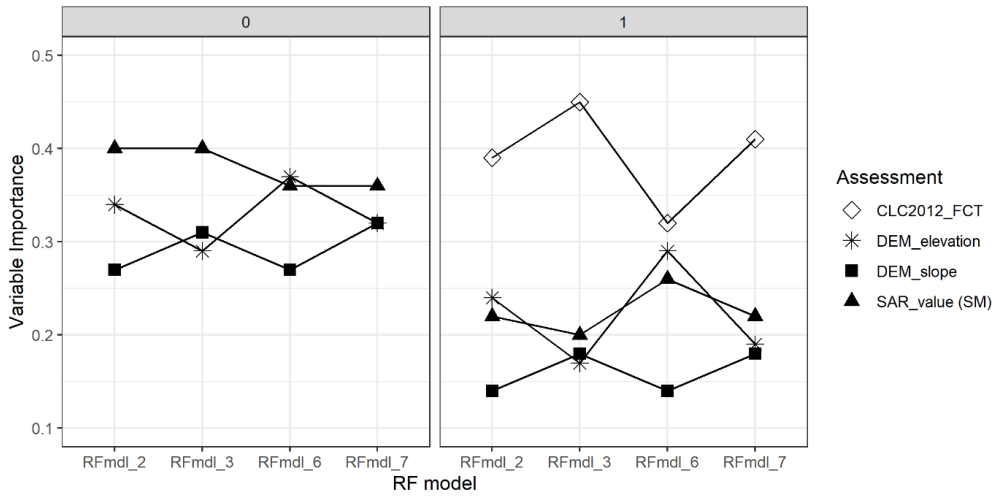


Figure 7. Predictor's importance results; 0 = FCT not included as predictor and 1 = FCT included as predictor.

 doi:10.20944/preprints202602.1553.v1

4. Discussion

4.1. L-Band SAR Detection

The data from 2009 L-Band SAR (ALOS PALSAR) mission were selected because of the quality and completeness of the soil survey for that specific time lapse (reference data availability). The selection of imagery had to be limited to a period as close as possible to the field data collection dates.

Results of originally retrieved and non-masked SM pixel values were congruent when compared with ground truth data ranges. Results are also in concordance with theoretical facts such as that open water is usually easily distinguishable in SAR imagery due to its specular reflectance properties [87]. The shuttle displacement in the edges of the tiles might be a consequence of the geometry of acquisition, the incidence angle of the radar signal on the swath, and/or as an additional effect of conducting the terrain correction using a DEM of medium resolution.

It was observed that areas with low surface roughness such as calm water, bare soil parcels and bare agricultural land showed the lowest values of retrieved soil moisture. In an intermediate range, some agricultural areas showed higher retrieved SM values, which is confirmed by several studies that have shown that farmland has specular scattering and double bounce scattering as dominant backscattering mechanisms [50].

The highest values of retrieved SM were obtained in areas nearby water bodies, waterlogged lands located where young/sparse/broad-leaved forest was the land cover. In areas close to water bodies, backscatter appeared to be sensitive to factors such as vegetation density and RDC of vegetation/soil, which tends to be higher around water bodies due to lateral currents and shallow water table depth. The quality index values range between -0.35 and 3.24 and revealed a higher calculation quality in areas with moderate to low vegetation density, young forest, broad-leaved forest, bare soils, agricultural and urban land. The higher quality index values correspond to areas where the backscatter matrices were less complex and therefore more accurately reconstructed. On the other hand, lower quality index values were observed in areas with coniferous/dense vegetation cover, which is also described in literature [63]. The foliage season for deciduous forests and younger stages of the forest where full canopy cover has not yet been achieved, are conditions that have positively influenced the quality of the retrieved SM values from SAR signals. Forest areas with dense, mature, or species that retain their foliage exhibited inadequate calculation precision.

With regard to target reachability, it was observed that an enhanced range to the target can be achieved when there are fewer obstacles along the trajectory of the SAR signal [37,40]. In summary, while the L-Band wavelength is expected to penetrate foliage, the presence of fewer leaves along the wave path is conductive to improved reachability of soil targets.

4.2. Reference Data Preparation & RF Model Performance

The study demonstrated that, with the aid of a comprehensive and balanced reference data set and taking into account additional prediction variables, SAR-generated soil moisture values are appropriate for the identification of organic wet soils, with substantial agreement using the RF approach. The original uneven distribution of the data points in the reference data set was corrected using the balancing approach [65,88] (cf. section 2.2.1). There were large differences, particularly with regard to the frequencies of reference points in detection class 0 and the, detection class 3, which had a negative impact on the performance of the RF models [61,62]. This study therefore presents and discusses only RF model results with balanced reference data sets that show small differences in performance. The reason for this is that the RF models, which were set up without resampling the low occurrence classes, have higher overall accuracy (0.71, 0.74; results not shown), but at the same time Kappa values are below 0.40, which indicates only fair agreement. It is evident that the observed model performance can be interpreted as a consequence of an imbalanced classification method applied to an imbalanced data set, wherein the classification model is extensively trained in a particular class (here: detection class 0). This finding is consistent with the conclusions presented in further studies [65,88]. This unbalanced classification was confirmed by examining the high values

 doi:10.20944/preprints202602.1553.v1

of specificity and low values of sensitivity of the corresponding RF models and appeared independent of the modelling approach (target classification vs. Boolean approach). In addition, a huge discrepancy occurred in the error rates of the target classification, where detection class 0 shows an error rate of 7%, and detection class 3 of 73%, respectively (results not shown). The same was the case for the Boolean approach: The "no event" error, when organic wet soils are defined as TRUE, amounts to only 3%, whereas the "event" error is 79% (results not shown). It has been demonstrated that a direct relationship exists between the balancing of the reference data set and the ability of the RF models to differentiate clearly between the detection classes and to make correct classifications.

Specific interest was given to RF models with low error rates or a low OOB error with respect to the classification of detection class 3, since these refer to the locations of organic wet soils. The preferable RF modelling approach in that case seems to be the target classification instead of the Boolean approach, resulting from the high discrepancy between the "no event" and "event" error described above. This high discrepancy is also reflected in the comparison of TPR and TNR of the respective RF models. The ratio only converges when detection class 2 and 3 are jointly defined as positive event. This can be attributed to the fact that the frequencies of the data points belonging to the "event" and "no event" categories show a tendency to approximate one another. In addition, the join of detection class 2 and 3 as TRUE also provides a distinct separation of different moisture conditions between the combined detection classes in the "event" and "no event" categories. The improving effect caused by the predictor FCT is most evident in these RF model variants. Ultimately, this means that the soil moisture content in the Boolean approach can be predicted with greater certainty than the high organic matter content caused by a higher soil moisture content. Conversely, it is possible that the high error rate in the pure classification of detection class 3 as positive event is generated by the fact that all other detection classes are categorised together as "no event". This means that the uniqueness of the SAR signal as predictor is reduced and that of the additional predictors is presumably reduced also.

Instead, the target classification classifies both detection classes – mineral wet soils and organic wet soils – with comparable quality, i.e., with similar error rates and a similar ratio of TPR and TNR. As demonstrated in Figure 4, there is a clear distinction in the SAR signal between mineral wet and organic wet. For detection class 3 in both the balanced and un-balanced data sets, the deflection in the range around two is observed to be more significant. However, there remains an overlap in the frequencies per SAR signal between the classes 0, 2 and 3 and it is hypothesized that the variations in groundwater levels attributed to the soil classes O1, O2 and O3 (cf. Table 1) may already be responsible for the observed disparities in SAR signals. The aggregation of these soil classes may therefore be considered suboptimal. Interestingly, the error rate for detection class 1 classification (intermediate wet soils) is the highest here, even higher compared to detection class 2 and 3. Regarding the groundwater information for this detection class in Table 1, it is clear that data points in this category have groundwater levels between 0.5 m and 1.8 m below surface. The lower groundwater levels also already apply to detection class 0. In addition, no further distinction is made between mineral and organic soils within detection class 1. As a result, these data points occupy an intermediate position, which also results in less clarity and greater scatter in the values of all predictors (soil moisture, elevation, slope, FCT) [61,62].

4.3. Importance of Predictors

Several studies found that topographic information on slope and elevation could be used to determine topographic locations where wetlands were likely to occur [24] [26–28]. In the present study, the performance of the RF models in terms of accuracy matrices depended also on such topography indices. The RF models in which the regional DEM with 25 m resolution was used, showed superiority in terms of detection over SRTM DEM. This could be attributed to the slightly higher spatial resolution of the regional DEM.

As a consequence of the training and testing of the RF models, incorporating FCT as a predictor, a marked shift in the relative importance of the predictors was observed. The significance of soil

 doi:10.20944/preprints202602.1553.v1

moisture pixel values from SAR signals diminished, while that of FCT increased. This result can be interpreted as a consequence of the random selection of predictors in the bootstrapping process when training occurs. The algorithm first selects the FCT variable given its better capabilities to induce an early effective separation to split the tree. In summary, the algorithm demonstrates a preference for initial execution of the most efficient predictor in order to execute the split. Consequently, as the remaining predictors are applied, the resulting effects will naturally decrease.

The apparent causality of FCT as a result of the ecological combination of soil, water table and terrain properties or vice versa also plays a role that could be further investigated. The employment of FCT alongside elevation and slope could result in confounding effects, thereby signifying the potential substitution of other predictors by FCT in this kind of classification, specifically within topographic conditions prevalent in the study region. The frequency of organic wet soils (detection class 3) regarding the FCT predictor was investigated additionally. Noticeable more data points of that specific detection class were located under broad-leaved forest cover. The study revealed that, despite the coniferous forest encompassing approximately three times more area than the broad-leaved forest in the region under investigation, organic wet soils occurred almost six times more frequently in the broad-leaved forest than in the coniferous forest and three times more frequently than in the mixed forest. This can be explained by the natural occurrence of *Alnus glutinosa*, which prefers wet or temporary water-logged soils in the study region.

5. Conclusions

SM pixel values retrieved from L-band SAR signals showed a good performance in the aim to detect the location of organic wet soils under forest cover in the study region. Retrieved SM values from ALOS-PALSAR were coherent with ground truth measurements of groundwater level or SM content. The largest improvements in the accuracy of the RF models were found in direct relation with the balance of the detection class distributions within the reference data set and the use of terrain indices from the 25 x 25 m DEM. This clearly indicates that in the region of study, terrain indices can characterize topographic features in the landscape with a strong linkage to peatland and/or wetland formation. With the aim of detecting organic wet soils, all RF models tested performed better than detection by random classification. The Kappa values in total range higher than the no-information rate and the presented results indicate a moderate to substantial agreement of detectability under the iteratively method developed [86].

The findings of the study demonstrated that the accuracy of the RF models was enhanced by the incorporation of the FCT as a predictor within the training procedure. When FCT was not included as a predictor, the SM pixel values obtained from SAR signals as well as terrain indices from the DEM became most important variables. However, when FCT was included as a predictor, the predictor's importance shifted to FCT as the most important.

The evaluation of the detection performance when using CORINE Land Cover data from 2018 would be a subject of interest. This is due to the fact that the CLC 2018 data offers a superior spatial resolution that is more compatible with SAR signals of a higher spatial resolution [19]. Furthermore, the Copernicus High Resolution Forest Type Layer, first provided in 2015, is of interest for further improvements of RF models' performance. In the same order of ideas, it would be of interest to try SAR polarimetric decomposition results as variables for training and testing the classifier. Such analysis would be an excellent further study to complement the scope of the present research. The recent launch of Argentinian SAOCOM-1 constellation in addition to the expected launch of new P- and L-band SAR missions such as ALOS-4 PALSAR-3, the P-band Biomass Mission from ESA, the NASA–ISRO Synthetic Aperture Radar (NISAR) mission strongly increases the importance and relevance of the present study.

The method explored in this study is relevant and shows potential in the still exploratory research path to detect the presence of organic wet soils in temperate forests based on L-Band SAR σ^o. Further assessments, as the ones previously mentioned, could improve this method and offer appropriate benefits. The detection of wet organic soils in temperate forests with global coverage

 doi:10.20944/preprints202602.1553.v1

could be of potential exploration and development, helping to reduce resource-intensive campaigns to know the exact spatial location of these valuable ecosystems.

Author Contributions: Specific contributions of each author are conceptualization, Jan-Peter Mund, Winfried Riek and Evelyn Wallor; methodology, Hainner Aparicio, Jan-Peter Mund and Evelyn Wallor; validation, Jan-Peter Mund, Winfried Riek and Nicole Voss; formal analysis, Hainner Aparicio and Evelyn Wallor; investigation, Hainner Aparicio, Nicole Voss and Evelyn Wallor; resources, Winfried Riek; writing—original draft preparation, Hainner Aparicio; writing—review and editing, Nicole Voss and Evelyn Wallor; visualization, Hainner Aparicio and Evelyn Wallor; supervision, Jan-Peter Mund. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Reference data used for this research is published by Benning et al. (2020) and can be accessed via DOI 10.3220/DATA20190625100522. The data publication additionally leads to a freely available database. Remote sensing and geo-data used as predictors is open source and referenced in the manuscript.

Acknowledgments: This research was conducted as part of the Forest Information Technology Master's programme. Authors appreciate the resources and academic environment provided by Eberswalde University for Sustainable Development, which made this work possible.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CLC	CORINE Land Cover
DEM	Digital Elevation Model
FCT	Forest cover type
RF	Random Forest
SAR	Synthetic Aperture Radar
SM	Soil moisture

References

1. IPCC. Summary of policymakers. In *Impacts, Adaption and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., Eds.; Publisher: Cambridge University Press, Cambridge, Great Britain, 2007, pp. 7–22.
2. UNFCC. United nations framework convention on climate change. In *Peatlands in Global Conventions: Status and Prospects*; Minayeva, T., Parish, F., Joosten, H., Silvius, M., Sirin, A., Eds.; IMCG Newsletter, **2008**, Volume 2, pp. 12-17.
3. Bridgham, S.D.; Cadillo-Quiroz, H.; Keller, J.K.; Zhuang, Q. Methane emissions from wetlands: Biogeochemical, microbial, and modeling perspectives from local to global scales. *Glob. Chang. Biol.* **2013**, *19*, 1325–1346.
4. Pohl, M.; Hoffmann, M.; Hagemann, U.; Giebels, M.; Albiac Borraz, E.; Sommer, M.; Augustin, J. Dynamic C and N stocks – key factors controlling the C gas exchange of maize in heterogenous peatland. *Biogeosciences* **2015**, *12*, 2737–2752.
5. Camporese, M.; Ferraris, S.; Putti, M.; Salandin, P.; Teatini, P. Hydrological modelling in swelling/shrinking peat soils. *Water Resour. Res.* **2006,** *42*, 1–15.
6. Kechavarzi, C.; Dawson, Q.; Leeds-Harrison, P.B. Physical properties of low-lying agricultural peat soils in England. *Geoderma* **2010**, *154*, 196–202.
7. Augustin, J.; Herrmann, A. Ökosystemleistungen aus der Sicht des Klimaschutzes. In *Moore in Brandenburg und Berlin*; Luthardt, V., Zeitz, J. Eds., Natur + Text, Rangsdorf, Germany, 2014, pp. 207–218.

 doi:10.20944/preprints202602.1553.v1

8. Tiemeyer, B.; Albiac Borraz, E.; Augustin, J.; Bechtold, M.; Beetz, S.; Beyer, C.; Drösler, M.; Ebli, M.; Eickenscheidt, T.; Fiedler, S.; Förster, C.; Freibauer, A.; Giebels, M.; Glatzel, S.; Heinichen, J.; Hoffmann, M.; Höper, H.; Jurasinski, G.; Leiber-Sauheitl, K.; Peichl-Brak, M.; Roßkopf, N.; Sommer, M.; Zeitz, J. High emissions of greenhouse gases from grasslands on peat and other organic soils. *Glob. Change Biol.* **2016**, *22(12)*, 4134-4149.
9. Wallor, E.; Zeitz, J. How properties of differently cultivated fen soils affect grassland productivity – a broad investigation of environmental interactions in Northeast Germany. *Catena* **2016**, *147*, 288–299.
10. Yu, Z.; Loisel, J.; Brosseau, D.P.; Beilman, D.W.; Hunt, S.J. Global peatland dynamics since the Last Glacial Maximum. *Geophys. Res. Lett.* **2010**, *37*, L13402.
11. Swindles, G. T.; Morris, P. J.; Wheeler, J.; Smith, M. W.; Bacon, K. L.; Turner, T. E.; Headley, A.; Galloway, J. M. Resilience of peatland ecosystem services over millennial timescales: evidence from a degraded British bog. *J. Ecol.* **2016**, *104(3)*, 621–636.
12. Roßkopf, N.; Holger, F.; Jutta, Z. Organic soils in Germany, their distribution and carbon stocks. *Catena* **2015**, *133*, 157-170.
13. Fell, H.; Roßkopf, N.; Zeitz, J. Estimating vulnerability of agriculturally used peatlands in north-east Germany to carbon loss based on multi-temporal subsidence data analysis. *Catena* **2016**, *137*, 61–69.
14. Hasch, B.; Meier, R.; Luthardt, V., Zeitz, J. Restoration of forest mires in Brandenburg and first results to establish a Decision Support System for the management of forest mires. *Telma* **2007**, *37*, 165-183.
15. Alaska Satellite Facility (ASF) Data Search - VERTEX. Available online: https://asf.alaska.edu/data-sets/sar-data-sets/alos-palsar/ (accessed on 28 October 2020).
16. NASA JPL - Shuttle Radar Topography Mission (SRTM) Global 1 arc second. Distributed by the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC). Available online: https://lpdaac.usgs.gov (accessed on 28 October 2020).
17. European Space Agency (ESA) - Sentinel Application Platform (SNAP). Available online: https://step.esa.int/main/toolboxes/snap/ (accessed on 21 January 2021).
18. Landesvermessung und Geobasisinformation Brandenburg (LGB). *Digitales Geländemodell DGM25 – Brandenburg*. Publisher: LGB Potsdam, Germany, 2014.
19. Copernicus Land Monitoring Service (CLMS) - Corine Land Cover 2012 dataset version 2020_20u1. Available online: http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012/view (accessed on 28 October 2020).
20. Belgiu, M.; Dragut, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.*, **2016**, *114*, 24-31.
21. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156-168.
22. Whitcomb, J.; Moghaddam, M.; McDonald, K.; Kellndorfer, J.; Podest, E. Mapping vegetated wetlands of Alaska using L-band radar satellite imagery. *Can. J. Remote. Sens.* **2009**, *35*, 54–72.
23. Corcoran, M.J.; Knight, F.J.; Gallant, L.A. Influence of multi-source and multitemporal remotely sensed and ancillary data on the accuracy of Random Forest classification of wetlands in northern Minnesota. *Remote Sens.* **2013**, *5(7)*, 3212-3238.
24. Clewley , D.; Whitcomb, J.; Moghaddam, M.; McDonald, K.; Chapman, B.; Bunting, P. Evaluation of ALOS PALSAR data for high-resolution mapping of vegetated wetlands in Alaska. *Remote Sens.* **2015**, *7*, 7272–7297.
25. De Almeida Furtado, L.F.; Silva, T.S.F.; de Moraes Novo, E.M.L. Dual-season and full-polarimetric C band SAR assessment for vegetation mapping in the Amazon varzea wetlands. *Remote Sens. Environ.* **2016**, *174*, 212–222.
26. Rodhe, A.; Seibert, J. Wetland occurrence in relation to topography: a test of topographic indices as moisture indicators. *Agric. For. Metorol.* **1999**, *98-99*, 325-340.
27. Merot, Ph.; Squividant, H.; Aurousseau, P.; Hefting, M.; Burt, T.; Maitre, V.; Kruk, M.; Butturini, A.; Thenail, C.; Viaud, V. Testing a climato-topographic index for predicting wetlands distribution along an European climate gradient. *Ecol. Model.*, **2003**, *163(1–2)*, 51-71.

 doi:10.20944/preprints202602.1553.v1

28. Vaze, J.; Teng, J.; Spencer, G. Impact of DEM accuracy and resolution on topographic indices. *Environ. Model. Softw.* **2010**, *25(10)*, 1086-1098.
29. Rosenqvist, A.; Finlayson, C.; Lowry, J.; Taylor, D. The potential of long-wavelength satellite-borne radar to support implementation of the Ramsar Wetlands Convention. *Aquat. Conserv.: Mar. Freshw. Ecosyst.* **2007**, *17(3)*, 229-244.
30. Reschke, J.; Bartsch, A.; Schlaffer, S.; Schepaschenko, D. Capability of C-band SAR for operational wetland monitoring at high latitudes. *Remote Sens.* **2012**, *4*, 2923–2943.
31. Hess, L.L.; Melack, J.M.; Affonso, A.G.; Barbosa, C.; Gastil-Buhl, M.; de Moraes Novo, E.M.L. Wetlands of the lowland Amazon basin: extent, vegetative cover, and dual-season inundated area as mapped with JERS-1 synthetic aperture radar. *Wetlands* **2015**, *35*, 745–756.
32. White, L.; Brisco, B.; Dabboor, M.; Schmitt, A.; Pratt, A. A collection of SAR methodologies for monitoring wetlands. *Remote Sens.* **2015**, *7*, 7615–7645.
33. Bartsch, A.; Hofler, A.; Kroisleitner, C.; Trofaier, A.M. Land cover mapping in northern high latitude permafrost regions with satellite data: achievements and remaining challenges. *Remote Sens.* **2016**, *8*, 979.
34. White, L.; McGovern, M.; Hayne, S.; Touzi, R.; Pasher, J.; Duffe, J. Investigating the Potential Use of RADARSAT-2 and UAS imagery for Monitoring the Restoration of Peatlands. *Remote Sens.* **2020**, *12*, 2383.
35. Huang, S.; Ding, J.; Zou, J.; Liu, B.; Zhang, J.; Chen, W. Soil Moisture Retrival Based on Sentinel-1 Imagery under Sparse Vegetation Coverage. *Sensors* **2019**, *19(3)*, 589.
36. Wagner, W.; Blöschl, G.; Pampaloni, P.; Calvet, J.-C.; Bizzarri, B.; Wigneron, J.-P.; Kerr, Y. Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Nordic Hydrology* **2007**, *38(1)*, 1-20.
37. Wagner, W.; Hahn, S.; Kidd, R.; Melzer, T.; Bartalis, Z.; Hasenauer, S.; Figa-Saldaña, J.; de Rosnay, P.; Jann, A.; Schneider, S. The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. *Meteorol. Z.* **2013**, *22*, 5–33.
38. Horning, N.; Robinson, J.A.; Sterling, E.J. *Remote Sensing for Ecology and Conservation: A Handbook of Techniques*, 1st ed; Publisher: Oxford University Press. Oxford, UK, 2010.
39. Lillesand, T.M.; Kiefer, R.W.; Chapman, J. *Remote Sensing and Image Interpretation*, 7thed; Publisher: John Wiley & Sons Inc, New York, US, 2015.
40. Njoku, E.G., Entekhabi, D. Passive microwave remote sensing of soil moisture. *J. Hydrol.* **1996**, *184*, 101-129.
41. Kasischke, E.S.; Smith, K.B.; Bourgeau-Chavez, L.L.; Romanowicz, E.A.; Brunzell, S.; Richardson, C.J. Effects of seasonal hydrologic patterns in south Florida wetlands on radar backscatter measured from ERS2 SAR imagery. *Remote Sens. Envir.* **2003**, *88*, 423–441.
42. Kasischke, E.S.; Bourgeau-Chavez, L.L.; Robert, A.R.; Wyatt, K.H.; Waddington, J.M.; Turetsky, M.R. Effects of soil moisture and water depth on ERS SAR backscatter measurements from an Alaskan wetland complex. *Remote Sens. Environ.* **2009**, *113*, 1868–1873.
43. Lievens, H.; Verhoest, N.E.C. Spatial and temporal soil moisture estimation from RADARSAT-2 imagery over Flevoland, The Netherlands. *J. Hydrol.* **2012**, *456–457*, 44–56.
44. Dabrowska-Zielinska, K.; Musial, J.; Malinska, A.; Budzynska, M.; Gurdak, R.; Kiryla, W.; Bartold, M.; Grzybowski, P. Soil moisture in the Biebrza wetlands retrieved from Sentinel-1 imagery. *Remote Sens.* **2018**, *10*, 1979.
45. Kim, J.-W.; Lu, Z.; Gutenberg, L.; Zhu, Z. Characterizing hydrologic changes of the Great Dismal Swamp using SAR/InSAR. *Remote Sens. Environ.* **2017**, *198*, 187–202.
46. Bechtold, M.; Schlaffer, S.; Tiemeyer, B.; De Lannoy, G. Inferring water table depth dynamics from ENVISAT-ASAR C-band backscatter over a range of peatlands from deeply drained to natural conditions. *Remote Sens.* **2018**, *10*, 536.
47. Gao, Q.; Zribi, M.; Escorihuela, M.J.; Baghdadi, N. Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution. *Sensors* **2017**, *17*, 1966.
48. Ferro-Famil, L.; Potier, E. Synthetic Aperture Radar Imaging. In *Microwave Remote Sensing of Land Surfaces, Techniques and Methods*, ed.; Baghdadi, N., Zribi, M., Eds.; Publisher: ISTE Press London, UK, 2016, pp. 1-62.

 doi:10.20944/preprints202602.1553.v1

49. Hidayat, H.; Hoekman, D.; Vissers, M.; Hoitink, A. Flood occurrence mapping of the middle Mahakam lowland area using satellite radar. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 1805–1816.
50. Kausarian, H.; Tetuko, J.; Sumantyo, S.; Kuze, H.; Aminuddin, J.; Waqar, M. Analysis of Polarimetric Decomposition, Backscattering Coefficient, and Sample Properties for Identification and Layer Thickness Estimation of Silica Sand Distribution Using L-Band Synthetic Aperture Radar. *Can. J. Remote Sens.* **2017**, *43:2*, 95-108.
51. Izumi, Y., Widodo, J.; Kausarian, H.; Demirci, S.; Takahashi, A.; Razi, P.; Nasucha, M.; Yang, H.; Tetuko, J. Potential of soil moisture retrieval for tropical peatlands in Indonesia using ALOS-2 L-band full-polarimetric SAR data. *Int. J. Remote Sens.* **2019**, *40:15*, 5938-5956.
52. Hoekman, D.; Van Rooij, W.; Quiñones, M.; Kooij, B.; Luiken, R.; Vissers M.; Vellekoop, S.; Carolita, I.; Budhiman, S.; Arief, S.; Roswintiarti, O. Tropical Peat View - NRT radar monitoring of peat swamp forests in Central Kalimantan and Riau. In Final report of the ESA EO Science for Society project: Tropical Peat View ESA AO/1-9101/17/I-NB (2020).
53. DWD Climate Data Center (CDC). Grids of monthly total precipitation over Germany, version v1.0. Available online: https://opendata.dwd.de/climate_environment/CDC/ (accessed on 30 October 2020).
54. DWD Climate Data Center (CDC) (2020). Grids of monthly averaged daily air temperature over Germany, version v1.0. Available online: https://opendata.dwd.de/climate_environment/CDC/ (accessed on 30 October 2020).
55. Kopp, D.; Jochheim, H. *Forstliche Boden- und Standortsdaten des Nordostdeutschen Tieflands als Datenbasis für die Landschaftsmodellierung*, 1st ed.; Publisher: Verlag Dr. Kessel, Remagen-Oberwinter, Germany, 2002, pp. 1-73.
56. Konopatzky, A. Das Wasser macht's - Substratfeuchte in der forstlichen Standortsgliederung - erste Skalierungs- und Bewertungsansätze in Brandenburg. In Wissenstransfer in die Praxis - Beiträge zum 7. Winterkolloquium, Eberswalder Forstliche Schriftenreihe 49, Eberswalde, Germany, 2012, pp. 73-82.
57. Russ, A., Riek, W., Wessolek, G. Regionalisierung bodenphysikalischer Kennwerte auf Grundlage forstlicher Standortskarten. In: *Ausgleichs- und Reaktorfunktionen von Waldböden im Stoff- und Wasserkreislauf*, 1st ed.; Albert-Ludwigs-Universität Freiburg, Ed.; Publisher: Freiburger Forstliche Forschung, Freiburg, Germany, 2013, Volume 96, pp. 113-127.
58. Riek, W.; Russ, A. (2019). Waldbodenbericht Brandenburg. Weitere Ergebnisse der landesweiten Bodenzustandserhebungen und Folgerungen für die nachhaltige Waldnutzung. In *Eberswalder Forstliche Schriftenreihe*, 1st ed.; Landesbetrieb Forst Brandenburg Eds.; Publisher: Landeskompetenzzentrum Forst Eberswalde, Eberswalde, Germany, 2019, Volume 68, pp. 1-238.
59. Hannappel, S.; Riek, W. Berechnung des Flurabstandes des oberflächennahen Grundwassers der Waldfläche Brandenburgs für hydrologisch typische Zeiträume. *Hydrol. Wasserwirtsch.* **2011**, *55(1)*, 4-15.
60. Russ, A., Riek, W., Wessolek, G. (2016). Entwicklung statistischer Vorhersagemodelle zur Identifizierung von Legendeneinheiten forstlicher Standortskarten am Beispiel Nordostdeutsches Tiefland. In Gefährdungen der Ökosystemdienstleistungen von Wäldern, Göttingen, Germany, 9.-10. Oktober 2014.
61. Russ, A.; Riek, W.; Wessolek, G. Three-Dimensional Mapping of Forest Soil Carbon Stocks Using SCORPAN Modelling and Relative Depth Gradients in the North-Eastern Lowlands of Germany. *Appl. Sci.* **2021**, *11*, 714.
62. Benning, R.; Ahrends, B.; Amberger, H.; Danigel, J.; Gauer, J.; Hafner, S.; Heinz, F.; Janot, M.; Mayer, S.; Mette, T.; Pieper, A.; Puhlmann, H.; Riek, W.; Steinicke, C.; Wallor, E.; Petzold, R. The Soil Profile Database for the National Forest Inventory Plots in Germany Derived from Site Survey Systems. *OpenAgrar* 2019, 1-12.
63. Lucas, R.; Armston, J.; Fairfax, R.; Fensham, R.; Accad, A.; Carreiras, J.; Kelley, J.; Bunting, P.; Clewley, D.; Bray, S. G. An evaluation of the ALOS PALSAR L-band backscatter—Above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure. IEEE J-STARS **2010**, *3(4)*, 576-593.
64. Williams, M.L.; Mitchell, A.L.; Milne, A.K.; Danaher, T.; Horn, G. Addressing critical influences on L-band radar backscatter for improved estimates of basal area and change. *Remote Sens. Environ.* **2022**, *272*, 112933.

 doi:10.20944/preprints202602.1553.v1

65. Dube, L.; Verster, T. Interpretability of the random forest model under class imbalance. *DSFE* **2024**, *4(3)*, 446-468.
66. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J. The soil moisture active passive (SMAP) mission. *Proc. IEEE* **2010**, *98*, 704–716.
67. Kerr, Y.H.; Waldteufel, P.; Richaume, P.; Wigneron, J.P.; Ferrazzoli, P.; Mahmoodi, A.; Al Bitar, A.; Cabot, F.; Gruhier, C.; Juglea, S.E. The SMOS soil moisture retrieval algorithm. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1384–1403.
68. Gruber, A.; Scanlon, T.; van der Schalie, R.; Wagner, W.; Dorigo, W. Evolution of the ESA CCI Soil Moisture Climate Data Records and their underlying merging methodology. *Earth Syst. Sci. Data* **2019**, *11*, 717–739.
69. Gevaert, A.; Parinussa, R.; Renzullo, L.; Van Dijk, A.; De Jeu, R. Spatio-temporal evaluation of resolution enhancement for passive microwave soil moisture and vegetation optical depth. *Int. J. Appl. Earth Obs. Geoinfom.* **2015**, *45*, 235–244.
70. Lang, M.; McCarty, G.; Oesterling, R.; Yeo, I.-Y. Topographic metrics for improved mapping of forested wetlands. *Wetlands* **2013**, *33*, 141–155.
71. Zribi, M.; Muddu, S.; Bousbih, S.; Al Bitar, A.; Kumar, S.; Baghdadi, N.; Bandyopadhyay, N. Analysis of L-Band SAR Data for Soil Moisture Estimations over Agricultural Areas in the Tropics. *Remote Sens.* **2019**, *11*, 1122.
72. Fung, A.K.; Li, Z.; Chen, K.S. Backscattering from a randomly rough dielectric surface. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 356–369.
73. Wu, T.D.; Chen; K.S., Shi, J.; Fung, A.K. A transition model for the reflection coefficient in surface scattering. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 2040–2050.
74. Baghdadi, N.; Saba, E.; Aubert, M.; Zribi, M.; Baup, F. Evaluation of radar backscattering models IEM, Oh, and Dubois for SAR data in X-band over bare soils. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1160–1164.
75. Baghdadi, N.; Zribi, M.; Paloscia, S.; Verhoest, N.; Lievens, H.; Baup, F.; Mattia, F. Semi-empirical calibration of the integral equation model for co-polarized L-band backscattering. *Remote Sens.* **2015**, *7*, 13626–13640.
76. Chen, K.; Yen, S.; Huang, W. A simple model for retrieving bare soil moisture from radar scattering coefficients. *Remote Sens. Environ.* **1995**, *54*, 121-126.
77. Shi, J.; Wang, J.; Hsu, A.Y.; O'Neill, P.E.; Engman, E.T. Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 1254-1266.
78. Moran, M.S.; Peters-Lidard, C.D.; Watts, J.M.; McElroy, S. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. *Canadian J. Remote Sensing* **2004**, *30*, 805-826.
79. Hallikainen, M.T.; Ulaby, F.T.; Dobson, M.C.; Elrayes, M.A.; Wu, L.K. Microwave Dielectric Behavior of Wet Soil - Part 1: Empirical-Models and Experimental-Observations. *IEEE Trans. Geosci. Remote Sens.* **1985**, *23*, 25–34.
80. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 15 January 2021).
81. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5-32.
82. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2(3)*, 18-22.
83. Heung, B.; Bulmer, C.E.; Schmidt, M.G. Predictive soil parent material mapping at a regional scale: a Random Forest approach. *Geoderma* **2014**, *214–215*, 141–154.
84. Wang, B.; Waters, C.; Orgill, S.; Gray, J.; Cowie, A.; Clark, A.; Liu, D.L. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Sci. Total Environ.* **2018**, *630*, 367–378.
85. Castaldi, F.; Hueni, A.; Chabrillat, S.; Ward, K.; Buttafuoco, G.; Bomans, B. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS J. Photogramm.* **2019**, *147*, 267–282.
86. Landis, J. R.; Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33(1)*, 159-174.

 doi:10.20944/preprints202602.1553.v1

87. Baghdadi, N.; Bernier, M.; Gauthier, R.; Neeson, I. Evaluation of C-band SAR data for wetlands mapping. *Int. J. Remote Sens.* **2001**, *22*, 71–88.
88. More, A.S.; Rana, D.P. Review of random forest classification techniques to resolve data imbalance. In *1st International Conference on Intelligent Systems and Information Management (ICISIM),* Aurangabad, India, 5-6 October 2017.