DOI: 10.22616/RRD.30.2024.047

MACHINE LEARNING BASED CLASSIFICATION OF PEAT LAYER THICKNESS IN LATVIA USING NATIONAL FOREST INVENTORY DATA

***Raitis Melniks, Janis Ivanovs, Andis Lazdins**
Latvian State Forest Research Institute 'Silava', Latvia
*Corresponding author's e-mail: raitis.melniks@silava.lv

Abstract
This study investigates the distribution and carbon content of organic soils in Latvia, leveraging machine learning techniques alongside remote sensing and National Forest Inventory (NFI) data to enhance the precision of organic soil mapping. Our approach integrates data from various sources, including airborne laser scanning (ALS) data, digital elevation models (DEM), depth-to-water (DTW) and wet area maps (WAM), and historical organic soil data. By classifying over 24,000 soil probing measurements across Latvia into distinct peat layer thickness categories, we develop a machine learning model that categorizes the thickness of the organic layer with notable accuracy. Our findings indicate that the model, particularly when employing the xgbTREE algorithm and over-sampling method, successfully identifies areas with peat layers thicker than 40 cm, demonstrating a significant improvement over traditional mapping methods. The study reveals an underestimation of organic soil coverage in Latvia by previous estimates, suggesting a broader distribution than recognized, with the model achieving an accuracy of 0.86 and a kappa value of 0.67. This research not only underscores the efficacy of integrating machine learning and remote sensing for soil mapping but also highlights the critical role of accurate data and models in determining organic soil distribution. The insights gained from this study are vital for policy-making and environmental planning, offering a more detailed understanding of Latvia's peatland resources and their conservation needs.
Key words: LiDAR, organic soils, topography.

Introduction

Peatlands provide essential services such as storing carbon, producing biomass, and regulating climate. However, they are being degraded by climate change and swift changes in land use, releasing their carbon (C) reserves (Joosten *et al.*, 2016; Minasny *et al.*, 2019). Understanding their size, condition, and carbon stocks is crucial for their conservation and to support the goals of the Paris Agreement. Despite covering just about 2.8% of the global land area, peatlands play a critical role in carbon storage, holding between 33% to 50% of the world's soil carbon reserves (Hilbert, Roulet, & Moore, 2000; Frolking *et al.*, 2011; Li *et al.*, 2018).

Organic soils are defined by their buildup of organic material based on their organic content, degree of decomposition, and water saturation levels. Within histosols, further distinctions are made, such as fibric, which consists of less decomposed peat; hemic, with partially decomposed organic material; and sapric, representing highly decomposed peat. These classifications help in understanding the characteristics and ecological functions of organic soils across different environments.

The landscape's topography and the bedrock beneath play a crucial role in creating moist conditions. The shape of the land affects water runoff, the connectivity of water networks, and the pooling of water, as noted by Jencso *et al*. (2009). The climate has a significant effect on the moisture content of the soil, with rainfall influencing the levels of groundwater, surface runoff, reduction reactions, and the buildup of organic materials. The temperature influences the activities of microorganisms and plants, which in turn affects the accumulation and breakdown of organic materials (Deluca & Boisvenue, 2012). It is also a factor in the rate of evaporation and the overall moisture regime of wetlands. In the boreal forests, the commonality of moist soil is attributed to elevated groundwater levels. These moist conditions, along with lower temperatures, retard the decomposition of organic materials, favoring the formation of peat (Luke *et al.*, 2007). Peat, with its superior capacity to retain water compared to mineral soils, enhances moisture retention in specific locales (Åström, Aaltonen, & Koivusaari, 2001). Wet soils are not merely repositories of moisture but also hubs for organic material. Observations indicate that areas prone to runoff and associated with wet soils exhibit similar concentrations of organic carbon compounds. This freshly derived organic carbon highlights the interplay between watercourses and soil. The transference of organic matter from moist soils mirrors in the flux of elements such as organic nitrogen, phosphorus, and sulfur, similar to the release of organic carbon (Ledesma *et al*., 2018).

Digital Elevation Models (DEMs) are instrumental in examining natural processes related to the landscape, and thematic maps provide further insights into their effects across different regions. Depth-to-Water (DTW) maps offer a model of groundwater proximity to surface water features like rivers and lakes (Lidberg, Nilsson, & Ågren, 2020), whereas maps of soil wetness reflect the influence of the underlying bedrock (Ivanovs & Lupikis, 2018). This study aims to identify the distribution of organic soils and its carbon content by comparing historical soil maps with predictions using remote sensing data.

Materials and Methods

The research area covers the entire territory of Latvia. According to historical soil maps, mire distribution data, forest growth condition data, peat extraction license data and other data sources, organic soils in Latvia cover an area of 6958 million square kilometers or 10.78% of land area 'Figure 1'. This estimation is based on the data on the organic soil coverage derived from the results of the Paliduculture in the Baltics project (Piirimäe *et al*., 2020).

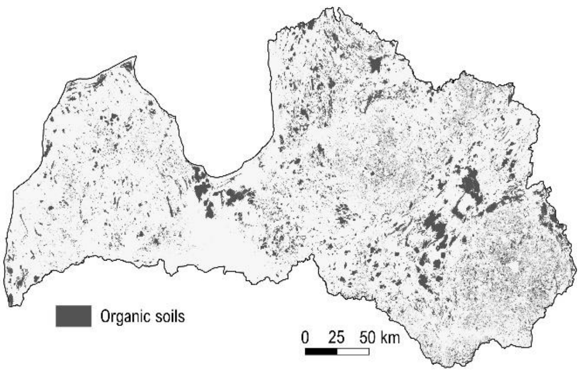

Figure 1. Distribution of organic soils in Latvia.

Data were collected from NFI sample plots located in forested areas, where the depth of the peat layer was assessed through soil probing. In each plot, this probing was conducted four times in each cardinal direction (North, East, South, and West), 12.5 meters away from the plot's center, resulting in a total of over 24,000 measurements. Subsequently, these measurements were categorized into three classes that represent four peat layer thickness ranges: plots with either no peat layer or one up to 5 cm thick, plots with peat layers up to 20 cm thick, plots where the peat layer's thickness is between 20 and 40 cm, and plots with peat layer thickness exceeding 40 cm.

The training dataset comprises a variety of variables sourced from ALS (Airborne Laser Scanning) data and additional cartographic materials, all at a 5 m horizontal resolution. The variables incorporated into the machine learning model include:

- DEM (Digital Elevation Model): A terrain model derived from ALS data, provided by the Latvian Geospatial Information Agency.
- Historical Organic Soil Data: A data layer generated by amalgamating historical soil maps, mire distribution data, data on forest growth conditions, peat extraction license information, and other data sources.
- DTW (Depth to Water): Maps indicating water depth with catchment areas of 10 and 30 hectares, created earlier in the research using ALS data.
- WAM (Wet Area Maps): Maps identifying wet areas, also prepared in prior research stages from ALS data.
- Normalized Height Maps: Models normalizing the Earth's surface relief.
- Slope: Models depicting the slope of the Earth's surface.
- Saga Wetness Index: A moisture index based on a modified calculation of the catchment area, as discussed by Böhner & Selige (2006).
- Soil Data: Data concerning soil texture.
- Proximity to Water: The distance to the nearest body of water, such as a river, lake, or sea.
- Continentality: The distance from the sea.
- X and Y Coordinates: Geographical positioning data.

The categorization of the organic layer's thickness was executed using the 'Caret' package in R. The NFI dataset, divided into three categories of peat thickness, was randomly split into 80% for training and 20% for testing. Various machine learning classification algorithms, including *xgbDART*, *xgbTREE*, among others, were evaluated. To mitigate the influence of an imbalanced dataset, different strategies were explored-over-sampling, under-sampling, and the SMOTE algorithm. All models underwent parameterization and optimization through a grid-search methodology coupled with 5-fold cross-validation to identify the optimal model. These fine-tuned models were then applied to the test data and assessed using Cohen's kappa index to measure agreement.

Results and Discussion

The machine learning model employing the *xgbTREE* algorithm, in conjunction with the over-sampling method, yielded the best classification outcomes and the highest kappa value. 'Figure 2' illustrates the influence of different remote sensing data and cartographic materials on the classification results. The most critical parameter was found to be continentality, followed by the depth-to-water index and DEM (Digital Elevation Model) values.

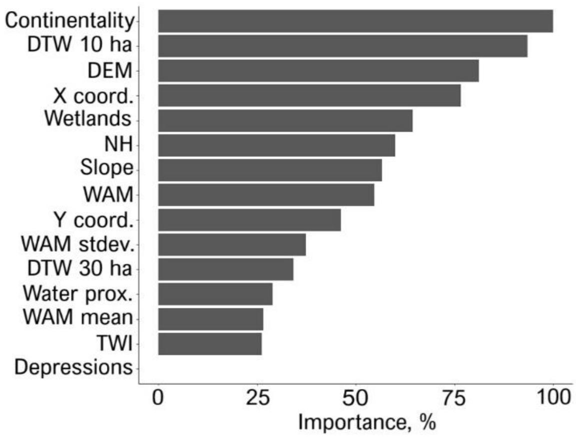

Figure 2. Feature importance in the model.

The accuracy of the overall machine learning classification algorithm reaches 0.86, while the kappa value is 0.67. Separately, by different classes, sensitivity reaches:

Soils without peat layer or up to 5 cm – 0.96;

Soils with a layer of peat from 5 to 20 cm – 0.45;

Soils with a layer of peat thickness 20-40 cm – 0.39;

Soils with a peat layer > 40 cm – 0.8.

'Figure 3' presents the classification outcomes for Latvia's landscape.

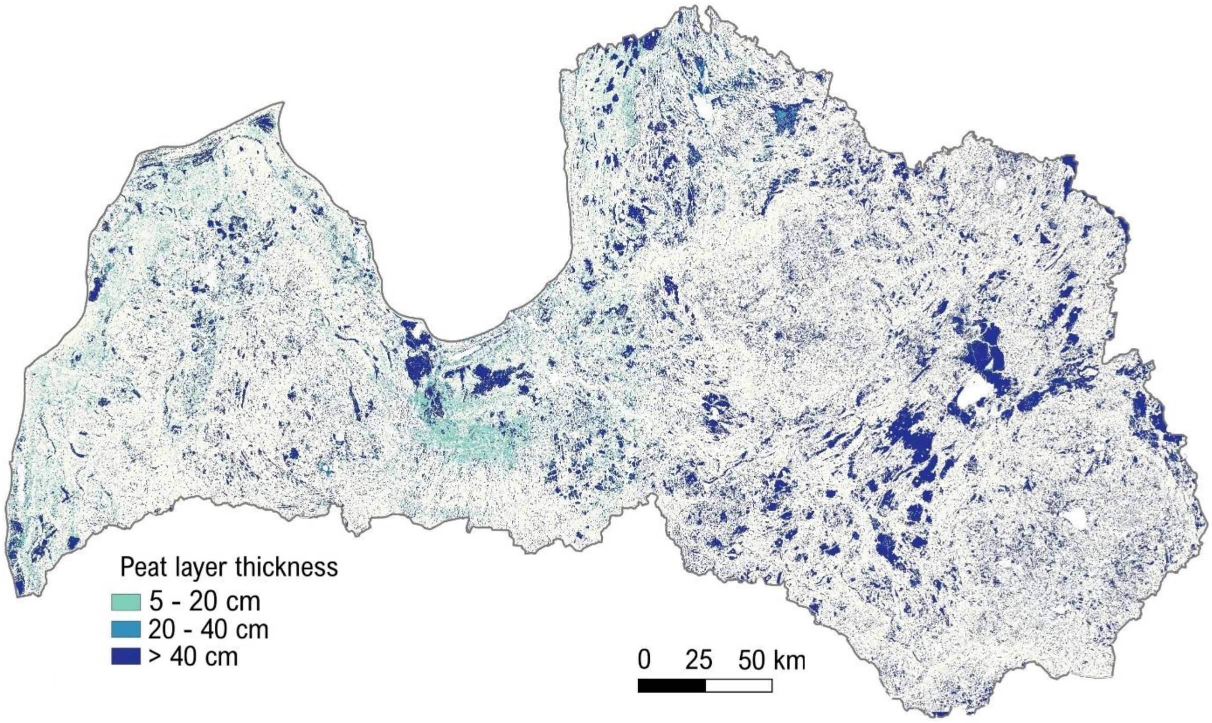

Figure 3. Map of modelled peat layer thickness in Latvia.

The data reveal that areas with a peat layer thickness ranging from 5 to 20 cm account for 0.7 % (461 km²) of Latvia's land, while 1.7% (1120 km²) of the country is overlain by peat layers between 20 and 40 cm thick. Additionally, a substantial 16.8% (10858 km²) is characterized by peat layers exceeding 40 cm in thickness. These findings mark a deviation from the previously estimated coverage of organic soils in Latvia, which stood at 10.8%. This discrepancy is likely due to variations in the precision of organic soil mapping and the spatial resolution of the areas analysed. An evaluation of the existing organic soil distribution maps against the newly acquired NFI plot data reveals an accuracy rate of 0.76 and a kappa statistic of 0.39. This comparison underlines the significant enhancement in accuracy and insight into the distribution of organic soils across Latvia facilitated by the applied machine learning algorithm.

Peatland mapping techniques vary according to peatland accessibility and the resources at hand. In the EU, nations such as Finland and Sweden utilize their robust data infrastructures to generate accurate peat maps via country-wide gamma radiometric surveys, which allow for the distinction between shallow and deep peat layers (Lilja & Nevalainen, 2006, Väänänen *et al*., 2007).

Canada has successfully utilized remote sensing and DSM for extensive territory mapping. Recently, employing a variety of remote sensing methods has proven exceptionally effective for generating high-resolution outcomes in targeted, regional research efforts. For example, Hird *et al*. (2017) achieved the mapping of Alberta's peatlands by combining multispectral satellite data, digital elevation models (DEM), and synthetic aperture radar (SAR) imagery, with additional data from forest inventory plots. In a similar manner, Bourgeau-Chavez *et al*. (2017) utilized a technique for their research in regions rich in permafrost. Additionally, the application of airborne LiDAR has been acknowledged for its precision in mapping peatlands at a high level of detail, as demonstrated by research conducted by Millard & Richardson (2013) and Chasmer *et al*. (2016), even though its use is somewhat restricted. While Indonesia has tested digital soil mapping, it has not yet been implemented nationwide. To support effective spatial planning and policy formulation, it is necessary to have a peat map of at least 1:50,000 resolution or a spatial accuracy of 30 meters or better. Various research efforts have distinguished peatlands using satellite imagery (including visible and infrared wavelengths) (Wijedasa *et al*., 2012), as well as radar data (Novresiandi & Nagasawa, 2017). Some researchers have attempted to determine peat depth using only elevation data (Jaenicke *et al*., 2008). However, only a handful of studies have employed digital mapping methods to assess peat thickness. The peatlands in these regions are typically dispersed and difficult to access, suggesting that mapping strategies should combine remote sensing technology with direct field observations.

Accurately determining peat layer thickness at a granular scale presents challenges due to the complex and often unknown nature of the underlying mineral terrain (Kettridge *et al*., 2008). However, identifying regions where landforms suggest sustained high moisture levels enables the identification of zones with significant long-term peat accumulation, leading to denser peat layers (Comas, Slater, & Reeve, 2004). Our study included several LiDAR derived soil moisture indicators, like WAM and DTW maps, to train our machine learning model. These indicators are crucial in the feature importance hierarchy of the machine learning model's development (Lidberg, Nillson, & Ägren, 2020, Deluca & Boisvenue, 2012). It should be noted that the quality of the DEM has a significant impact on the accuracy of the used models, how well it is prepared for hydrological modelling. For example, from whether correct input data for existing roads, ditches, culverts, which correspond to DEM have been available, to make its corrections in these places.

Conclusions

1. Our model using machine learning techniques, remote sensing and NFI data provides high accuracy of peat layers with different thicknesses spatial distribution, comparing to other available data sources.
2. Most important model input data variables was continentality, depth to water map and DEM. However, it should be noted that the accuracy of our model is affected by the accuracy of input data and models, such as depth to water and wet area map and other terrain indices.
3. The results reveal that best model performance is obtained for identifying peat which is thicker than 40 cm.
4. More research is needed which may improve the model performance regarding different peat thicknesses as well as considering usage of different satellite data products for training the model.
5. Our study provides more insight into organic soil distribution in Latvia, comparing to older data sources which can be taken into account by GHG inventory teams and policy makers.

Acknowledgements

This research was funded by the Norwegian Financial Mechanism for the 2014-2021, grant number 2021/6e-JP/SAD (ZM Nr.2021/20). The article processing charge was funded by the Latvian State Forest Research Institute 'Silava'. Contribution of Jānis Ivanovs is funded by the Forest Sector Competence Center funded project 'Carbon Forestry Planning Program Prototype and Activity Datasets' (No. 5.1.1.2.i.0/1/22/A /CFLA/007).

References

Åström, M., Aaltonen, E.-K., & Koivusaari, J. (2001). Effect of ditching operations on stream-water chemistry in a boreal forested catchment. *Science of The Total Environment*. 279(1–3), 117–129. DOI: 10.1016/S0048-9697(01)00757-4.

Böhner, J. & Selige, T. (2006). *Spatial prediction of soil attributes using terrain analysis and climate regionalisation*. In: Boehner, J., McCloy, K.R., Strobl, J. (Ed.). SAGA - Analysis and Modelling Applications, Goettingen: Goettinger Geographische Abhandlungen, 13-28.

Chasmer, L., Hopkinson, C., Montgomery, J., & Petrone, R. (2016). A Physically Based Terrain Morphology and Vegetation Structural Classification for Wetlands of the Boreal Plains. Alberta, Canada. *Canadian Journal of Remote Sensing.* 42(5), 521-540, DOI: 10.1080/07038992.2016.1196583.

Comas, X., Slater, L., & Reeve, A. (2004). Geophysical Evidence for Peat Basin Morphology and Stratigraphic Controls on Vegetation Observed in a Northern Peatland. *J. Hydrol*. 295, 173–184. DOI: 10.1016/j.jhydrol.2004.03.008.

Deluca, T.H., & Boisvenue, C. (2012). Boreal forest soil carbon: distribution, function and modelling. *Forestry*, 85(2), 161–184. DOI: 10.1093/forestry/cps003.

Frolking, S., Talbot, J., Jones, M.C., Treat, C.C., Kauffman, J.B., Tuittila, E.-S., & Roulet, N. (2011). Peatlands in the Earth's 21st century climate system. *Environ. Rev*. 19, 371–396. DOI: 10.1139/a11-014**.**

Lilja, H., & Nevalainen, R. (2006). Chapter 5 developing a Digital Soil Map for Finland. Lagacherie, A.B. McBratney, M. Voltz (Ed.). *Developments in Soil Science*. 31, 67-603. DOI: 10.1016/S0166-2481(06)31005-7.

Hilbert, D.W., Roulet, N., & Moore, T. (2000). Modelling and analysis of peatlands as dynamical systems. *Journal of Ecology* 88, 230-242. DOI: 10.1046/j.1365-2745.2000.00438.x

Hird, J.N., DeLancey, E.R., McDermid, G.J., & Kariyeva, J. (2017). Google Earth Engine, Open-Access Satellite Data, and Machine Learning in support of Large-Area Probabilistic Wetland Mapping. *Remote Sens*. 9(12), 1315. DOI: 10.3390/rs9121315.

Ivanovs, J., & Lupikis, A. (2018). Identification of wet areas in forest using remote sensing data. *Agronomy Research*. 16. DOI: 10.15159/AR.18.192.

Jaenicke, J., Rieley, J.O., Mott, C., Kimman, P., & Siegert, F. (2008). Determination of the amount of carbon stored in Indonesian peatlands. *Geoderma.* 147(3-4), 151–158. DOI: 10.1016/j.geoderma.2008.08.008**.**

Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Wondzell, S.M., Bencala, K.E., & Marshall, L.A. (2009). Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research*. 45(4). DOI: 10.1029/2008WR007225.

Joosten, H. (2016). Peatlands across the Globe. In: Bonn, A., Allott, T., Evans, M., Joosten, H., Stoneman, R. (Eds.). *Peatland Restoration and Ecosystem Services: Science, Policy and Practice*. Cambridge: Cambridge University Press. pp. 19–43.

Kettridge, N., Comas, X., Baird, A., Slater, L., Strack, M., Thompson, D., Jol, H., & Binley, A. (2008). Ecohydrologically Important Subsurface Structures in Peatlands Revealed by Ground-Penetrating Radar and Complex Con-ductivity Surveys. *J. Geophys. Res. Biogeosciences*. 113. DOI: 10.1029/2008JG000787.

Bourgeau-Chavez, L. L., Endres, S., Powell, R., Battaglia, M. J., Benscoter, B., Turetsky, M., Kasischke, E. S., & Banda E. (2017). Mapping boreal peatland ecosystem types from multitemporal radar and optical satellite imagery. *Canadian Journal of Forest Research*. 47(4): 545-559. DOI: 10.1139/cjfr-2016-019.

Ledesma, J. L. J., Futter, M. N., Blackburn, M., Lidman, F., Grabs, T., Sponseller, R. A., & Kö Hler, S. J. (2018). Towards an Improved Conceptualization of Riparian Zones in Boreal Forest Headwaters. *Ecosystems*, 21, 297–315. DOI: 10.1007/s10021-017-0149-5.

Li, C., Grayson, R., Holden, J., & Li, P. (2018). Erosion in peatlands: Recent research progress and future directions. *Earth-Sci. Rev*. 185, 870–886. DOI: https://doi.org/10.1016/j.earscirev.2018.08.005.

Lidberg, W., Nilsson, M., & Ågren, A. (2020). Using machine learning to generate high-resolution wet area maps for planning forest management: A study in a boreal forest landscape. *Ambio*. 49(2), 475–486. DOI: 10.1007/s13280-019-01196-9.

Luke, S. H., Luckai, N. J., Burke, J. M., & Prepas, E. E. (2007). Riparian areas in the Canadian boreal forest and linkages with water quality in streams. *Environmental Reviews*. 15(NA), 79–97. DOI: 10.1139/A07-001.

Millard, K. & Richardson, M. (2013) Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. *Canadian Journal of Remote Sensing*. 39(4), 290-307. DOI: 10.5589/m13-038.

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B., McBratney, A., Roudier, P., O'Rourke, S., Rudiyanto, Padarian, J., Poggio, L., ten Caten, A., Thompson, T., Tuve, C., Widyatmanti, W. (2019). Digital mapping of peatlands – A critical review, Earth-Science Reviews, 196, 102870. DOI: 10.1016/j.earscirev.2019.05.014.

Novresiandi, D. A. & Nagasawa, R. (2017). Polarimetric synthetic aperture radar application for tropical peatlands classification: a case study in Siak River Transect. Indonesia: Riau Province. *J. Appl. Remote. Sens*. 11 (1), 016040. DOI: 10.1117/1.JRS.11.016040.

Piirimäe, K., Salm, J.-O., Ivanovs, J., Stivrins, N., Greimas, E., Jarašius, L., Zableckis, N., & Haberl, A. (2020). Paludiculture in the Baltics-GIS Assessment report. Retrieved December 27, 2023, from https://media.voog.com/0000/0037/1265/files/GIS_EE_LV_LT_04.2020.pdf

Väänänen, T., Hyvönen, E., Jakonen, M., Kupila, J., Lerrsi, J., Leskinen, J., Liwata, P., Nevalainen, R., Putkinen, S., & Virkki, H. (2007). Maaperän yleiskartan tulkinta- ja kartoitusprosessi. *Tech. rep*. Retrieved January 21, 2024, from http://tupa. gtk.fi/raportti/arkisto/p22_4_2007_58.pdf.

Wijedasa, L. S., Sloan, S., Michelakis, D. G., & Clements, G. R. (2012). Overcoming Limitations with Landsat Imagery for Mapping of Peat Swamp Forests in Sundaland. *Remote Sensing*. 4(9), 2595-2618. DOI: 10.3390/rs4092595.