Wall-to-wall mapping of peat depth from Lidar terrain and airborne radiometrics in Norwegian landscapes

Julien Vollering[1], Naomi Gatis[2], Mette Kusk Gillespie[1], Karl-Kristian Muggerud[1], Sigurd Daniel Nerhus[1], Knut Rydgren[1], and Mikko Sparf[1]

[1]Department of Civil Engineering and Environmental Sciences, Western Norway University of Applied Sciences, Norway
[2]Department of Geography, University of Exeter, United Kingdom

Correspondence: Julien Vollering (julien.vollering@hvl.no)

Abstract. The abstract goes here. It can also be on *multiple lines*.

1 Introduction

Introduction text goes here. Read Gatis et al. (2019) and related work (Minasny et al., 2019).

2 Materials and methods

2.1 Sites

We assessed how well we could predict peat depth at two sites with conspicuously different physical geography: Skrimfjella in eastern Norway and Ørskogfjellet in western Norway (Fig. 1c). These sites were chosen because they were covered by radiometric data from airborne surveys, relatively little built-up area, and road access.

At Skrimfjella we delineated a study area of $34\,\mathrm{km}^2$ based on radiometric coverage and accessibility (Fig. 1b). The study area has a diverse bedrock, with 32 % alkali feltspat granite, 26 % mergelstein, 10 % granite, and eight other types with > 1 % coverage (NGU, 1:250 000 dataset). The landscape within our delineation is classified as *inland hills and mountains* (Simensen et al., 2021). It is almost without human infrastructure, dominated by forest, and borders on a large nature reserve. The study area has a mean elevation of $438\,\mathrm{m}$ above sea level (range 223–711, IQR 351–509), and its mean slope at $10\,\mathrm{m}$ resolution is 10.8° (IQR 4.6–15.1°). In Norway's AR5 national land capability dataset (Ahlstrøm et al., 2019), $1.5\,\mathrm{km}^2$ (4.5 %) of the study area is classified as mire — defined as areas with mire vegetation and at least $30\,\mathrm{cm}$ of peat depth.

At Ørskogfjellet we defined a study area of $124\,\mathrm{km}^2$ which basically followed the footprint of the radiometric survey (Fig. 1a). According to the Geological Survey of Norway, bedrock in the area is 84 % granitic gneiss, 11 % granite, and 5 % aluminium silicate gneiss (NGU, 1:250 000 dataset). This study area comprises a wide range of major landscape types: *coastal plains*, *coastal fjord*, *inland valleys*, as well as *inland hills and mountains* (Simensen et al., 2021). It is mostly forested, but also contains considerable farmland and open upland, and has several large lakes. Its mean elevation is $211\,\mathrm{m}$ above sea level

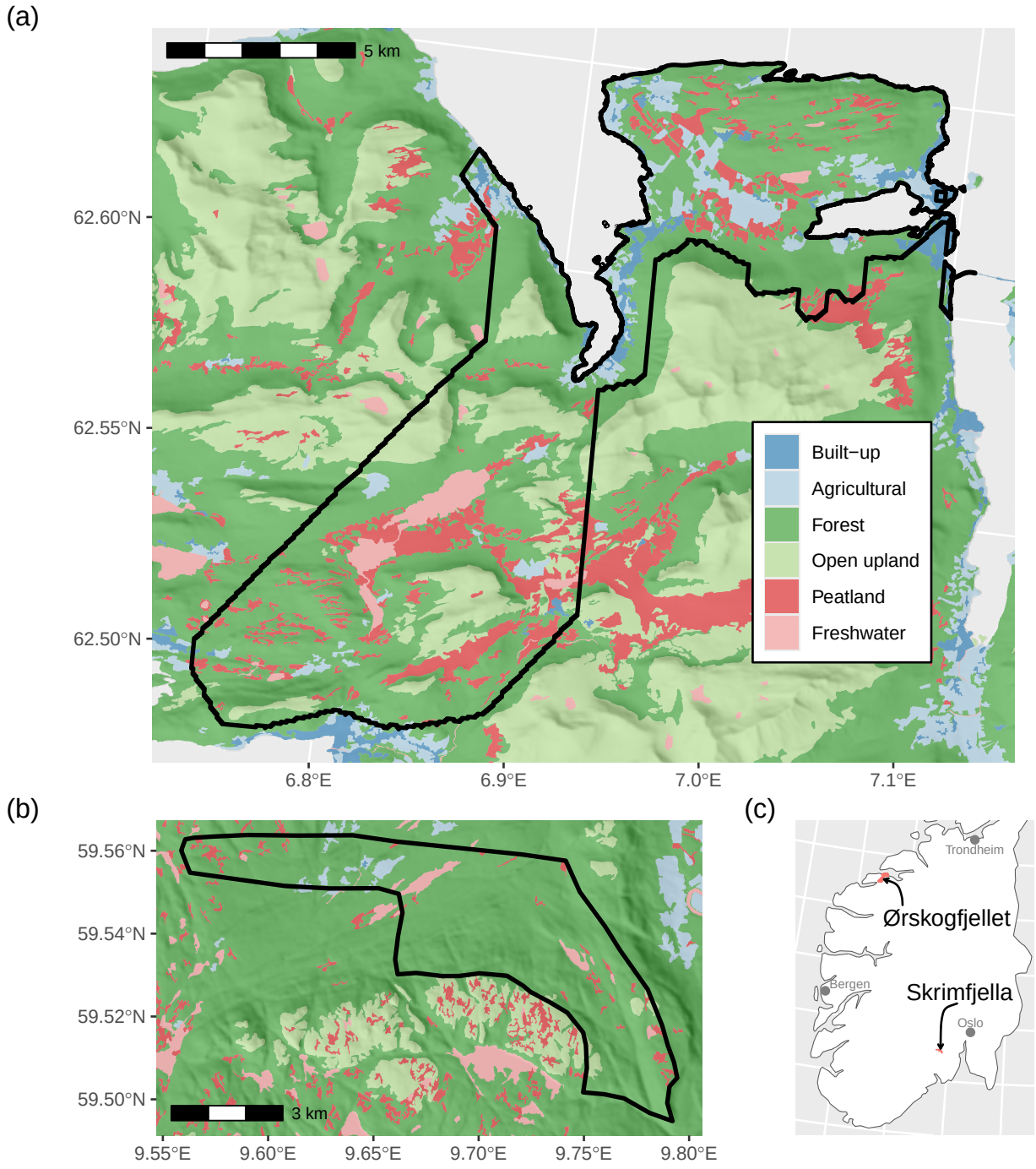

Figure 1. Study areas at Ørskogfjellet (a) and Skrimfjella (b) within southern Norway (c). Land cover shown here is from the AR50 national land resource database and has simplified geometry with respect to the AR5 database used in the study.

(range 0–807, IQR 73–310), and its mean slope at $10\,\mathrm{m}$ resolution is 13.0° (IQR 4.7–18.3°). The AR5 dataset counts $15.3\,\mathrm{km}^2$ (12.4 %) of the study area as mire.

2.2 Peat depth measurements

At both study sites, our measurements of peat depth were made for the purpose of training a Random Forest model of peat depth, and we designed our sampling with this in mind (Brus, 2019). Broadly, we aimed for a sample that was representative of the predictor space defined by the most important predictors of peat depth (Wadoux et al., 2019; Ma et al., 2020). A sample that preserves the properties of the multivariate distribution of predictor and outcome variables is most likely to maintain any complex, non-linear relationships that exist in the population while avoiding spurious ones (Brus, 2019). We chose for our sampling and modelling a spatial resolution of $10\,\mathrm{m}$. We considered this a reasonable compromise between digital terrain model (DTM) resolution ($1\,\mathrm{m}$) and small mires on the one hand, and airborne radiometric resolution ($50\,\mathrm{m}$) on the other.

2.2.1 Skrimfjella

We measured peat depth in selected locations ($10\,\mathrm{m}$ raster cells) at Skrimfjella. The locations were chosen only from areas delineated as mire in the AR5 national land capability dataset. Within this mire area, we stratified our sample across values of elevation, slope, and potassium ground concentration (from processed airborne gamma ray spectrometry, Baranwal et al., 2013). Specifically, we used the *eSample* function in the *iSDM* R package (v.1.0) to chose an environmentally systematic sample. This function defines the environmental space as a two-dimensional convex hull around the ordinated data, then creates a regular grid across that space, and lastly finds for each grid cell the datum that is nearest (Hattab et al., 2017). Elevation was extracted from the $10\,\mathrm{m}$ national DTM, slope calculated in degrees, and potassium ground concentration downscaled with bilinear resampling. We set a target sample size of 100, excluded the top and bottom percentile from the convex hull, and with these parameters *eSample* returned 105 raster cells.

In addition to the peat depth locations, we had another arm of our sampling design for measuring peatland occurrence, as binary variable. We wanted to measure peatland occurrence outside of mapped mire areas because the AR5 dataset is known to underestimate peatland coverage (especially in forests, Bryn et al., 2018), and because airborne radiometrics may help identify unmapped peatland (Gatis et al., 2019; O'Leary et al., 2022). The occurrence locations were sampled from the part of the study area that (1) was mapped as something other than mire in the AR5 database and (2) had a slope < 20°. We performed environmentally systematic sampling of this population with the same procedure as for the depth locations, and *eSample* returned 106 raster cells.

Field work at Skrimfjella was conducted in August 2020. We navigated to the centers of the raster cells in the depth and occurrence samples by handheld GPS, checking that positional error was below $3\,\mathrm{m}$. For each depth sample location, we measured peat depth three times (at the vertices of a triangle with $2\,\mathrm{m}$ sides) to get a more representative value for the $10\,\mathrm{m}$ raster cell, and to dampen the effect of outlying measurements (Parry et al., 2014). We used a metal probe pushed downward until resistance indicated the base of the peat column. Probe locations were adjusted up to $20\,\mathrm{cm}$ if the base of the peat column seemed to be blocked by an obvious artifact. For each occurrence sample location, we recorded the presence or absence of

peatland — primarily by digging and examining the top $20\,\mathrm{cm}$ of soil (where this was possible). We judged whether the soil was a peat soil based on its density, texture, and color. Occasionally, when the soil itself was difficult to judge, we made our determination also based on the presence or absence of mire vegetation. Although peat soil is strictly defined by organic content (which we did not analyse), we believe our protocol produced reasonable determinations of presence or absence that would generally satisfy most of the varying definitions of peatland (Minasny et al., 2023).

Besides the depth and occurrence measurements described above, we also measured peat depth in three subjectively-chosen, individual mires, using ground-penetrating radar (GPR). We used the Malå ProEx GPR system (Guideline Geo AB, Sweden) with its $500\,\mathrm{MHz}$ shielded antenna mounted in a plastic sledge, and its control unit connected to a GNSS receiver. At each of the three mires we recorded GPR traces along walking transects that covered the extent of the mire, mostly in traversing, zigzag patterns with between $5\,\mathrm{m}$ and $20\,\mathrm{m}$ spacing at their vertices. Along the GPR transects we also probed peat depth at marked trace locations, to be able to calibrate the GPR wave speed velocity. We processed the GPR data with Reflex2DQuick software (v.3.0, Sandmeier Scientific Software, Germany), applying a time-zero correction, a dewow filter, and a gain filter based on observed energy decay. Then we picked strong reflectors in the radargrams that we interpreted as the base of the peat column. We used picks at marked trace locations to calibrate wave speed velocity; we pooled calibration points across the three mires and fitted a linear regression of depth on one-way travel time with the intercept fixed at zero. In total we had 46 calibration points along $3.5\,\mathrm{km}$ of GPR transects. Finally, we used the calculated wave velocity ($0.0387\,\mathrm{m\,ns^{-1}}$, $R^2 = 0.874$) to convert the travel times of all picks to calibrated peat depths.

2.2.2 Ørskogfjellet

At Ørskogfjellet we also measured peat depth in a sample of $10\,\mathrm{m}$ raster cells, selected from the part of the study area classified in the AR5 dataset as mire. Before selecting locations, we determined a minimal sample size that would adequately capture the terrain and radiometric properties of the entire mire area. Specifically, we aimed to identify the size at which adding locations produced diminishing decreases in divergence between sample and population distributions — i.e. the elbow point in a curve of similarity between sample and population (Malone et al., 2019). This approach has been found to identify sample sizes that correspond with diminishing returns in predictive model performance on external evaluation data (Saurette et al., 2023). We defined a sequence of sample sizes (50–500) and for ten replicate samples at each size (drawn by conditioned latin hypercube sampling, Minasny and McBratney, 2006; Roudier, 2011), we calculated the mean Kullback-Leibler divergence between sample and population distributions (Malone et al., 2019; Saurette et al., 2023). The variables in the divergence calculation were terrain slope and four radiometrics: potassium, thorium, uranium, and total count. Next, we fitted a asymptotic regression of mean divergence on sample size, and identified the sample size at which the curve reached 95 % of the fitted asymptote. Through this procedure we found that we could adequately capture the population distribution with a sample of 160 locations.

To choose 160 locations, we performed feature space coverage sampling. This approach has been found to produce higher accuracy in Random Forest models than conditioned latin hypercube sampling (Wadoux et al., 2019; Ma et al., 2020). Feature space coverage sampling aims to disperse samples as uniformly as possible in multidimensional predictor space, and is

implemented by choosing locations that are closest to cluster centers in a k-means clustering of the standardized predictor space (Brus, 2019). Feature space coverage sampling works best when all dimensions are important predictors of the outcome (Wadoux et al., 2019), and we used the same five predictors that we used to choose sample size: terrain slope and four radiometrics. The radiometric variables were downscaled to $10\,\mathrm{m}$ resolution with cubic B-spline resampling in QGIS (v.3, QGIS Development Team, 2024). We adjusted the feature space coverage sampling to ensure that locations were accessible within time constraints, and assessed how this changed our sample from an ideal feature space coverage sample. Adjusting for accessibility is justified because the smaller sample size that would result if accessibility were ignored can degrade model accuracy as much or more as deviations from ideal sampling designs (Wadoux et al., 2019; Ma et al., 2020). To adjust, we first restricted the sampling population to mire areas that were within an arbitrary cost distance of publicly-accessible roads. Cost distance was calculated using GRASS's *r.walk* function, with friction costs defined by AR5 land classes (GRASS Development Team, 2022). After creating a feature space coverage sample with this restriction, we also inspected a map of the sample and substituted 16 inaccessible locations with accessible locations from the same or a nearby cluster. Our two accessibility adjustments increased the distance in standardized predictor space between sample locations and cluster centers by 78 % (with respect to the ideal sample), but distance in our sample was still only 46 % of the mean distance to cluster centers — i.e., accessibility did not force locations far from cluster centers relative to the size of the clusters.

Field work at Ørskogfjellet was conducted in August 2023. We navigated to the centers of the raster cells in the sample using real time kinematic differential GNSS (Topcon Positioning Systems, USA), to ensure sub-meter positional accuracy. At each location we measured peat depth three times by manual probing, with probe locations spaced approximately $2.5\,\mathrm{m}$ apart. In areas with dense sampling locations, we also measured peat depth with GPR along snaking transects passing through the centers of the sampling cells (seven transects, $6.2\,\mathrm{km}$ total length). We used the same GPR system as at Skrimfjella, but with a $100\,\mathrm{MHz}$ Malå rough terrain antenna (Guideline Geo AB, Sweden) at some transects. To navigate the GPR transects, we placed flags at the cell centers of the sample locations, and used a handheld GNSS receiver to guide the GPR operator. At sampling locations crossed by a GPR transect, we arranged the manual probe positions along the transect (for better calibration of the GPR wave speed velocity), while other locations were probed in a triangular pattern around the cell center like at Skrimfjella.

We processed the GPR data with Reflexw software (v.8.5, Sandmeier Scientific Software, Germany), applying a dewow filter, time-zero correction, bandpass filter, gain filter, and a dynamic correction that accounts for the non-vertical wave path between offset transmitter and receiver antennae. The last correction is important for the rough terrain antenna, which has an antenna separation ($2.2\,\mathrm{m}$) — comparable to typical peat depths. As with the data from Skrimfjella, we picked the base of the peat column from strong reflectors in the radargram, and calibrated wave velocity with manual probe measurements in a linear regression. The points in the regression were created by joining to each probe measurement the travel time of the nearest pick, but only if these were within $2\,\mathrm{m}$ of each other. In total we had 78 calibration points along $7.8\,\mathrm{km}$ of interpretable GPR traces (transect length exceeded because of extra GPR data). Finally, we used the calculated wave velocity ($0.0427\,\mathrm{m\,ns^{-1}}$, $R^2 = 0.946$) to convert the travel times of all picks to calibrated peat depths.

We also used two sets of existing depth measurements from Ørskogfjellet. The first set was provided by the Norwegian Public Roads Administration, who commissioned GPR surveys of particular peatland areas in 2020 and 2021. The surveys

were conducted with a dual channel system (70 MHz and 300 MHz; ImpulseRadar AB, Sweden), connected to GNSS with CPOS correction. We used interpreted and calibrated traces from these surveys, and discarded some data where multiple depths were interpreted for the same locations. This summed to $7.4\,\mathrm{km}$ of interpreted traces. The second set of existing depth data we extracted from a paper map made by the Norwegian Soil and Mire Company in 1984. This map presents 44 borehole depths (in decimeters) across a $9\,\mathrm{ha}$ peatland area. We georeferenced the map and digitized the borehole locations and depths.

2.3 Peat depth predictors

We created the same suite of peat depth predictors for both sites (25 continuous and 1 categorical; Table 1). All continuous predictors were derived either from an airborne radiometric survey or from a DTM. From the radiometric surveys we simply used the four variables produced by the surveyors (Geological Survey of Norway): ground concentration of Potassium, Thorium, Uranium, as well as total count. From the DTMs we calculated several land surface parameters, ranging from simple terrain indices to more complex geomorphometric and hydrological variables (Maxwell and Shobe, 2022). The categorical predictor was peat depth class, from a national map dataset. Complete descriptions of predictors follow below.

2.3.1 Radiometric

The radiometric survey covering Skrimfjella was conducted in 2008–2011. The survey was flown at an average altitude of $75\,\mathrm{m}$ and average speed of $108\,\mathrm{km\,h^{-1}}$, with flight lines spaced $200\,\mathrm{m}$ apart. Spectrometer count rates were calibrated annually to known concentrations of Potassium, Thorium, and Uranium in mobile pads. The Geological Survey of Norway processed data from the spectrometer following standard procedures outlined by the International Atomic Energy Association, and the processing included: correction for aircraft and cosmic background radiation, correction for radon in the air, window stripping of the gamma ray spectrum, correction for flying height, conversion of count rates to ground concentrations, and finally gridding to $50\,\mathrm{m}$ resolution with micro-leveling. Further details about the survey and data processing are provided in Baranwal et al. (2013). We downscaled the processed data to $10\,\mathrm{m}$ resolution by cubic spline resampling, using the *terra* package in R.

A very similar radiometric survey covering Ørskogfjellet was conducted in December 2014 and January 2015. This survey was flown at an average altitude of $80\,\mathrm{m}$ and average speed of $88\,\mathrm{km\,h^{-1}}$, with flight lines also spaced $200\,\mathrm{m}$ apart. Spectrometer count rates were calibrated in 2013 to known concentrations of Potassium, Thorium, and Uranium in mobile pads. The spectrometer data were processed following the same procedure as for the survey at Skrimfjella, except that a convolution filter was added to smooth the gridded data. Further details about the survey and data processing are provided in Ofstad (2015). The $10\,\mathrm{m}$ resolution predictors from this survey were identical to the layers used in the sampling design (resampled from $50\,\mathrm{m}$ resolution with cubic B-splines).

2.3.2 Terrain

For terrain-derived variables, we obtained $1\,\mathrm{m}$ resolution DTMs from the Norwegian Mapping Authority. The DTM for Skrimfjella was produced from airborne laser scanning surveys in 2015 and 2022, with laser point density of $5\,\mathrm{pts\,m^{-2}}$.

Table 1. Candidate predictors of peat depth.

name	description
radK	Potassium ground concentration
radTh	Thorium ground concentration
radU	Uranium ground concentration
radTC	Total count of gamma radiation
elevation	Mean elevation
slope1m	Mean of 1 m slope
TPI1m	Mean of 1 m topographic position index
TRI1m	Mean of 1 m terrain ruggedness index
roughness1m	Mean of 1 m roughness
slope10m	10 m slope
TPI10m	10 m topographic position index
TRI10m	10 m terrain ruggedness index
roughness10m	10 m roughness
MRVBF	Multi-resolution valley bottom flatness
TWI5m	Mean of 5 m topographic wetness index
TWI10m	10 m topographic wetness index
TWI20m	Bilinear interpolation of 20 m topographic wetness index
TWI50m	Bilinear interpolation of 50 m topographic wetness index
DTW2500	Depth-to-water index, flow initiation area of 0.25 ha
DTW5000	Depth-to-water index, flow initiation area of 0.5 ha
DTW10000	Depth-to-water index, flow initiation area of 1 ha
DTW20000	Depth-to-water index, flow initiation area of 2 ha
DTW40000	Depth-to-water index, flow initiation area of 4 ha
DTW80000	Depth-to-water index, flow initiation area of 8 ha
DTW160000	Depth-to-water index, flow initiation area of 16 ha
DMK	DMK peat depth class, categorical with 3 levels

For Ørskogfjellet, the DTM was produced from a 2015 survey with $2\,\mathrm{pts\,m^{-2}}$. Where necessary, DTMs were resampled to the coordinate reference system of the radiometric data.

We used the *terra* R package to calculate from the DTMs: slope, topographic position index (difference from mean of eight neighbors), terrain ruggedness index (mean of absolute differences from eight neighbors), and roughness (range in the nine-cell neighborhood). These variables were derived at two scales to produce different predictors; we either calculated from $1\,\mathrm{m}$ DTM resolution and then aggregated, or aggregated to $10\,\mathrm{m}$ DTM resolution and then calculated the indices. This kind of multiscale feature engineering of land surface parameters has been found to improve machine learning predictions of soil properties (Miller et al., 2015; Dornik et al., 2022; Newman et al., 2023). We know that peat depth tends to vary at fine scales in Norway, which is why we chose $1\,\mathrm{m}$ and $10\,\mathrm{m}$ resolutions (Maxwell and Shobe, 2022). We also calculated the the Multi-Resolution Valley Bottom Flatness index, which indicates the degree of valley bottom flatness at a given location via a multiscale algorithm (Gallant and Dowling, 2003). We calculated this index in SAGA GIS (v.9.3.2, Conrad et al., 2015) with default parameters.

Next, we calculated the Topographic Wetness Index (Quinn et al., 1991). This index is notoriously scale-dependent and often matches real hydrological conditions best when calculated from moderate-to-coarse resolution DTMs (Ågren et al., 2014; Riihimäki et al., 2021), so we calculated it from $5\,\mathrm{m}$, $10\,\mathrm{m}$, $20\,\mathrm{m}$, and $50\,\mathrm{m}$ DTM resolution. The calculations were performed with Whitebox software (Lindsay, 2016), accessed through the *whitebox* R package (v2.4, Wu and Brown, 2022). We filled depressions in the DTM with the algorithm in Wang & Liu (2006), and used the deterministic infinity flow accumulation algorithm (Tarboton, 1997).

The last terrain-based predictor we included was the depth-to-water index (Murphy et al., 2007). This index approximates a location's vertical height above the surface water feature that it is likely to drain towards. It is calculated as the minimum cumulative slope (scaled by cell size) to a surface water feature (eq. 5 in Murphy et al., 2009). We calculated unitless slope from the $1\,\mathrm{m}$ DTM using the Whitebox software. Also using Whitebox, we defined surface water features from the DTM by filling depressions and then calculating flow accumulation to define catchment areas for each cell (Schönauer et al., 2021; Schönauer and Maack, 2021). This catchments area layer was then thresholded at seven different levels (*flow initiation areas* from $0.25\,\mathrm{ha}$ to $16\,\mathrm{ha}$) to estimate surface water features under moisture scenarios varying from wet to dry (Murphy et al., 2011; Ågren et al., 2014; Schönauer et al., 2021). In addition, all surface water features mapped in the AR5 dataset were also transferred to the raster layer. For each of the seven surface water layers, we derived the depth-to-water index using the *Distance Accumulation* tool in ArcGIS Pro (v.3.1, ESRI, USA), which has an efficient algorithm to find the cumulative distance over a cost surface to the least-cost source.

2.3.3 Peat depth class

We prepared one categorical predictor — peat depth class — from a historical national map dataset called *DMK* (Ahlstrøm et al., 2019). The DMK peat depth classes are: $< 1\,\mathrm{m}$ (*shallow*), $> 1\,\mathrm{m}$ (*deep*), and *unknown*. This dataset stems from field measurements made in 1964–2007 as part of a wider land cover mapping in Norway. The mapping was primarily for identifying agricultural and silvicultural resources, so it covers productive areas below the tree line and has greater coverage of peat depth class in peatlands judged to be arable or afforestable (Bjørdal, 2007; Ahlstrøm et al., 2019). Mappers generally assigned peat

depth classes to polygons of at least $0.5\,\mathrm{ha}$, although delineating polygons down to $0.2\,\mathrm{ha}$ was allowed if peat depth showed a "particularly marked difference" (Bjørdal, 2007). We rasterized the peat depth class attribute to our $10\,\mathrm{m}$ grid.

2.4 Predictive models of peat depth

2.4.1 Modelling approach

We used Random Forests (RF) to predict peat depth at both sites. RF is a tree-based ensemble machine learning algorithm that builds many decision trees on bootstrapped samples of the training data, randomly subsets predictors in the trees, and averages the predictions of the trees (Breiman, 2001). We chose RF because it can handle complex interactions between predictors, is robust to overfitting, and generally shows higher performance in DSM applications than other algorithms (Beguin et al., 2017; Nussbaum et al., 2018; Lamichhane et al., 2019). It is suited for use on relatively small training data sets and its predictions can be interrogated to learn about predictor importance (Khaledian and Miller, 2020). Evaluating variable importance in a maximally-predictive model aligns with the aim of this study.

RF by itself is not a spatial model, and it will only predict spatial structure in the outcome to the degree that the structure is captured by predictors. We considered using regression kriging — a hybrid between non-spatial and spatial techniques that would be achieved by adding to the RF predictions a geostatistically-interpolated surface of RF residuals (Hengl et al., 2004). The spatial component in regression kriging often improves map accuracy compared to a non-spatial model (Beguin et al., 2017; Lamichhane et al., 2019; Molla et al., 2023), but it can do so only if the spatial autocorrelation range in the non-spatial residuals is large compared to distances between samples and prediction locations (Hengl et al., 2004; Szabó et al., 2019; Takoutsing and Heuvelink, 2022). If the outcome varies at fine scales and the samples are clustered in small parts of the study area, a spatial component will hardly improve overall map accuracy. We used semivariograms to assess the spatial structure in the residuals of the RF predictions, and found that (non-spatial) RF rather regression kriging was justified at both sites.

We implemented models in the *tidymodels* framework in R (Kuhn and Wickham, 2020), with the *ranger* R package for RFs (v.0.16, Wright and Ziegler, 2017). RFs were fit with 1000 trees, minimum node size of 5, and the number of predictors randomly sampled at each split was the square root of the total number of predictors (*ranger* default). We did not tune these hyperparameters because RFs are relatively insensitive to tuning (Probst et al., 2019), and because it would require nested spatial cross-validation to prevent data leakage (Schratz et al., 2019).

2.4.2 Model performance

For both sites we compared the performance of models with five different configurations of three groups of predictors. Specifically, we trained models with:

1. only DMK peat depth class (2 binary indicator variables from 3 levels)
2. only terrain (21 continuous variables)
3. terrain and DMK peat depth class (23 variables)
4. terrain and radiometric predictors (25 continuous variables)

5. all predictors (26 variables)

These different configurations simulate different scenarios of data availability that are common in Norway or internationally. Comparing the different configurations allowed us to isolate the added value of each of the predictor groups. We did not explore configurations comprising radiometrics without terrain, because Lidar terrain surveys typically precede airborne radiometric surveys. The models with only DMK peat depth class were simple linear models rather than RFs, and served to provide a fair comparison between the accuracy of the RF models and the existing national map of peat depth, calibrated on the same data.

The performance of DSM must be evaluated with reference to a specific purpose (i.e., map vs. model validation, interpolation vs. extrapolation, Roberts et al., 2017; Milà et al., 2022), and here we aimed to evaluate maps of peat depth across the study areas. In the absence of additional field work to collect a design-based independent validation set, we used a spatial cross-validation scheme to evaluate model performance (Wadoux et al., 2021; Meyer and Pebesma, 2022). Specifically, we used k-Means Nearest Neighbor Distance Matching (kNNDM), which creates cross-validation folds that mimic the spatial prediction task that is defined as the goal (Linnenbrink et al., 2024). In particular, kNNDM looks for the spatial assignment of training data to folds that minimizes the difference between two distributions: nearest neighbor distances between training and test locations in the cross-validation, and nearest neighbor distances between training and prediction locations for the model. That way, the spatial separation between folds is similar to the separation between training and prediction locations — which increases the quality of the map accuracy estimate (Linnenbrink et al., 2024). For spatially clustered training data, this approach strikes a balance between the risk of optimistic metrics from random cross-validation and the risk of pessimistic metrics from other forms of spatial cross-validation (Wadoux et al., 2021). We implemented the kNNDM with the *CAST* R package (v.1.0.2, Meyer et al., 2024), setting prediction locations to all AR5 mire cells in the study area, and choosing a number of folds (5–20) that produced the best match between the two NND distributions. From the cross-validation we quantified *Root Mean Squared Error* (accuracy, original scale), R^2 (correlation, standardized scale), and *Lin's concordance correlation coefficient* (accuracy and correlation, standardized scale).

DSM products have much more value when their predictions are accompanied by uncertainty estimates, and all DSM should strive to assess uncertainty (Arrouays et al., 2020; Wadoux et al., 2020). Moreover, the quality of uncertainty estimates should be evaluated, just as the quality of predictions are (Heuvelink and Webster, 2022). Therefore, we produced prediction intervals with quantile regression forests (Meinshausen, 2006), and used the same spatial cross-validation to evaluate the prediction interval coverage probability (Shrestha and Solomatine, 2006). The quantile regression forests were trained with predictor configuration that showed the highest performance at each site (under the assumption that these models would be put into production) and we extracted 90 % prediction intervals.

While the primary aim of the DSM was to predict peat depth within peatlands (where peat > 30 cm), we also performed a minimal evaluation of our ability to estimate peat depth outside of mapped peatland. In other words, we tested if predictions of deep peat indicated unmapped peatland occurrence in the wider landscape. It is practically difficult to obtain suitable data for this problem, because locations with non-zero peat depth outside of mapped peatland are unknown and relatively rare. Therefore, we made do with the same data that we used for the peat depth predictions, which included a fraction of locations not mapped as peatland. We used the same spatial cross-validation folds, but this time we trained the models only on locations

mapped as peatland (in the training folds), and evaluated them only on locations not mapped as peatland (in the test fold). In this way, we simulated the situation where a model trained on known peatland is used to predict peat depth outside the known peatland. Since DMK peat depth class is always undefined outside known peatlands, we evaluated two predictor configurations: terrain predictors alone or terrain and radiometric predictors. From the cross-validation we quantified *Root Mean Squared Error* and *Mean Absolute Error*, and we tested the strength of the correlation between predicted and observed peat depth.

2.4.3 Model interpretation

We quantified global variable importance and examined partial dependence plots for the best-performing predictor configuration at each site. Global variable importance measures the influence of a given predictor on the output of the model, aggregated across all locations. Partial dependence plots depict the shape of the fitted relationship between a given predictor and the outcome. Both are useful for understanding the mechanisms behind the model's predictions and the roles of the predictors in the model.

For both sites, we interpreted a model trained on a non-collinear subset of variables from the best performing predictor configuration — because correlation between predictors degrades variable importance measures (Strobl et al., 2008; Biau and Scornet, 2016) and can produce misleading visualizations of predictor-outcome relationships (Biecek and Burzykowski, 2021; Dwivedi et al., 2023). Specifically, we eliminated variables from the best performing predictor configuration to obtain a set with no pairwise Pearson correlation above 0.7. Thus, highly-correlated sets of predictors are represented by a single variable for the purposes of model interpretation.

We calculated variable importance with the *vip* R package (v.0.4.1), by three different methods: *FIRM*, *permutation*, and *Shapley* (Greenwell and Boehmke, 2020). *FIRM* values measure the flatness of the partial dependence plot, *permutation* values measure the decrease in model performance when the predictor is permuted, and *Shapley* values are aggregated from local, game-theoretical measures of variable importance (Greenwell and Boehmke, 2020). *Permutation* values were obtained from ten iterations, with RMSE as the performance measure.

We calculated partial dependence with the *pdp* R package (v.0.8.1, Greenwell, 2017). For the top six most important variables, we plotted both partial dependence and individual conditional expectation (ICE), to show the average effect of the predictor on the outcome and the variation in the effect across observations, repectively (Goldstein et al., 2015). Non-parallel ICE lines indicate the presence of interactions between predictors.

3 Results

Include a 12cm width figure of Nikolaus Copernicus from Wikipedia with caption using R Markdown (Fig. 2).

3.1 Tables

You can add LaTeXtable in an R Markdown document to meet the template requirements (Table 2).

Or you can use markdown to create the table with booktabs = FALSE (https://github.com/rstudio/rticles/issues/558#issuecomment-19079

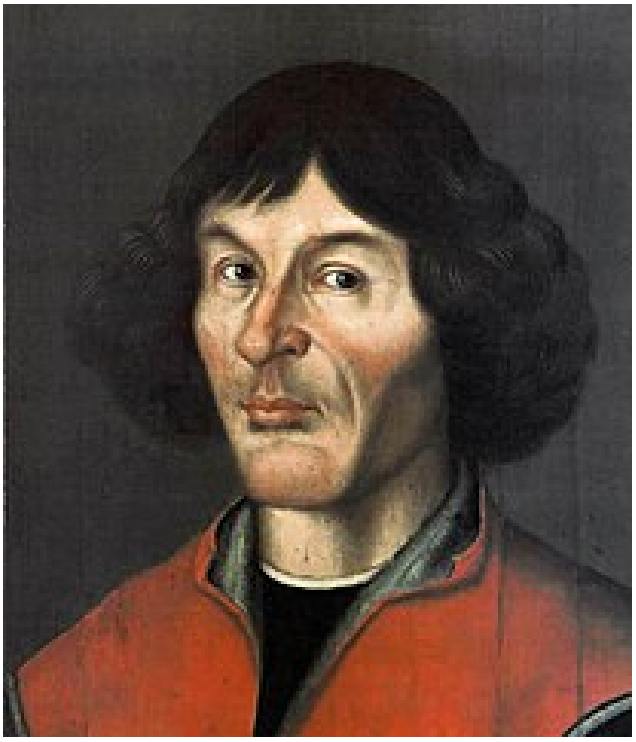

Figure 2. one column figure

Table 2. TEXT

a	b	c
1	2	3

Table Footnotes

See Table 3.

Table 3. My caption

	mpg	cyl	disp
Mazda RX4	21.0	6	160
Mazda RX4 Wag	21.0	6	160
Datsun 710	22.8	4	108

4 Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5 Conclusions

Nulla facilisi. Maecenas vel nunc nec purus tincidunt congue. Proin auctor, lectus eu pharetra malesuada, nisi nunc bibendum nunc, eget tincidunt nunc nisi id nunc. Sed euismod, nunc sit amet aliquam tincidunt, nunc nunc tincidunt nunc, nec tincidunt nunc nunc nec nunc. Donec auctor, nunc sit amet aliquam tincidunt, nunc nunc tincidunt nunc, nec tincidunt nunc nunc nec nunc.

. *Code and data availability.* Use this to add a statement when having data sets and software code available

Appendix A: For submission

"Appendices: all material required to understand the essential aspects of the paper such as experimental methods, data, and interpretation should preferably be included in the main text. Additional figures, tables, as well as technical and theoretical developments which are not critical to support the conclusion of the paper, but which provide extra detail and/or support useful for experts in the field and whose inclusion in the main text would disrupt the flow of descriptions or demonstrations may be presented as appendices. These should be labelled with capital letters: Appendix A, Appendix B etc. Equations, figures and tables should be numbered as (A1), Fig. B5 or Table C6, respectively. Please keep in mind that appendices are part of the manuscript whereas supplements (see below) are published along with the manuscript."

Appendix B: Figures and tables in appendices

Please also sort the appendix figures and appendix tables into the respective appendix sections. They will be correctly named automatically.

Appendix C: Copernicus from Rmarkdown

Please note: Per their guidelines, Copernicus does not support additional LaTeX packages or new LaTeX commands than those defined in their `.cls` file. This means that you cannot add any extra dependencies and a warning will be thrown if

so. **Important**: Always double-check with the official manuscript preparation guidelines at https://publications.copernicus.org/for_authors/manuscript_preparation.html, especially the sections "Technical instructions for LaTeX" and "Manuscript composition". Please contact Daniel Nüst, daniel.nuest@uni-muenster.de, with any problems.

. *Author contributions.* JV: Conceptualization, Investigation, Data curation, Formal analysis, Writing – original draft. NG: Conceptualization, Methodology, Writing - review & editing. MKG: Investigation, Writing - review & editing. KKM: Investigation, Data curation, Writing - review & editing. SDN: Investigation, Writing - review & editing. KR: Conceptualization, Investigation, Writing - review & editing. MS: Investigation, Data curation, Writing - review & editing.

. *Competing interests.* The authors declare that they have no conflict of interest.

. *Disclaimer.* The authors declare that the results, discussions, and interpretations presented in this study are solely their own. The views expressed herein do not necessarily reflect those of their respective institutions or funding agencies.

. *Acknowledgements.* We thank the Norwegian Public Roads Administration for sharing data from ground-penetrating radar surveys. We also thank Vikas Baranwal from the Geological Survey of Norway for helping us access the radiometric data from Skrim. This work contains data under the following licenses: (1) Creative Commons Attribution 4.0 International, © Kartverket, (2) *Norge digitalt* license, Norwegian Institute of Bioeconomy Research (NIBIO), © Geovekst, and (3) the Norwegian License for Public Data (NLOD), made available by the Geological Survey of Norway (NGU).

References

Ågren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J., and Arp, P. A.: Evaluating Digital Terrain Indices for Soil Wetness Mapping – a Swedish Case Study, Hydrology and Earth System Sciences, 18, 3623–3634, https://doi.org/10.5194/hess-18-3623-2014, 2014.

Ahlstrøm, A., Bjørkelo, K., and Fadnes, K. D.: AR5 Klassifikasjonssystem, Tech. rep., NIBIO, 2019.

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A. C., Morgan, C. L. S., Roudier, P., Poggio, L., and Mulder, V. L.: Impressions of Digital Soil Maps: The Good, the Not so Good, and Making Them Ever Better, Geoderma Regional, 20, e00 255, https://doi.org/10.1016/j.geodrs.2020.e00255, 2020.

Baranwal, V., Rodionov, A., Ofstad, F., Koziel, J., and Lynum, R.: Helicopter-Borne Magnetic, Electromagnetic and Radiometric Geophysical Surveys in the Kongsberg Region: Krøderen, Sokna, Hønefoss, Kongsberg and Numedalen., Tech. Rep. 2013.029, Geological Survey of Norway, 2013.

Beguin, J., Fuglstad, G.-A., Mansuy, N., and Paré, D.: Predicting Soil Properties in the Canadian Boreal Forest with Limited Data: Comparison of Spatial and Non-Spatial Statistical Approaches, Geoderma, 306, 195–205, https://doi.org/10.1016/j.geoderma.2017.06.016, 2017.

Biau, G. and Scornet, E.: A Random Forest Guided Tour, TEST, 25, 197–227, https://doi.org/10.1007/s11749-016-0481-7, 2016.

Biecek, P. and Burzykowski, T.: Explanatory Model Analysis, Chapman and Hall/CRC, New York, ISBN 978-0-367-13559-1, 2021.

Bjørdal, I.: Markslagsklassifikasjon i Økonomisk Kartverk. 2007-utgåva, Report, Norsk institutt for skog og landskap, ISBN 9788231100096, 2007.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brus, D. J.: Sampling for Digital Soil Mapping: A Tutorial Supported by R Scripts, Geoderma, 338, 464–480, https://doi.org/10.1016/j.geoderma.2018.07.036, 2019.

Bryn, A., Strand, G.-H., Angeloff, M., and Rekdal, Y.: Land Cover in Norway Based on an Area Frame Survey of Vegetation Types, Norwegian Journal of Geography, 72, 131–145, https://doi.org/10.1080/00291951.2018.1468356, 2018.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geoscientific Model Development, 8, 1991–2007, https://doi.org/10.5194/gmd-8-1991-2015, 2015.

Dornik, A., Cheţan, M. A., Drăguţ, L., Dicu, D. D., and Iliuţă, A.: Optimal Scaling of Predictors for Digital Mapping of Soil Properties, Geoderma, 405, 115 453, https://doi.org/10.1016/j.geoderma.2021.115453, 2022.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., and Ranjan, R.: Explainable AI (XAI): Core Ideas, Techniques, and Solutions, ACM Comput. Surv., 55, 194:1–194:33, https://doi.org/10.1145/3561048, 2023.

Gallant, J. C. and Dowling, T. I.: A Multiresolution Index of Valley Bottom Flatness for Mapping Depositional Areas, Water Resources Research, 39, https://doi.org/10.1029/2002WR001426, 2003.

Gatis, N., Luscombe, D., Carless, D., Parry, L., Fyfe, R., Harrod, T., Brazier, R., and Anderson, K.: Mapping Upland Peat Depth Using Airborne Radiometric and Lidar Survey Data, Geoderma, 335, 78–87, https://doi.org/10.1016/j.geoderma.2018.07.041, 2019.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E.: Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation, Journal of Computational and Graphical Statistics, 24, 44–65, https://doi.org/10.1080/10618600.2014.907095, 2015.

GRASS Development Team: Geographic Resources Analysis Support System (GRASS GIS) Software, Version 8.2, Open Source Geospatial Foundation, 2022.

Greenwell, B. M.: Pdp: An R Package for Constructing Partial Dependence Plots, The R Journal, 9, 421–436, 2017.

Greenwell, B. M. and Boehmke, B. C.: Variable Importance Plots—An Introduction to the Vip Package, The R Journal, 12, 343–366, 2020.

Hattab, T., Garzón-López, C. X., Ewald, M., Skowronek, S., Aerts, R., Horen, H., Brasseur, B., Gallet-Moron, E., Spicher, F., Decocq, G., Feilhauer, H., Honnay, O., Kempeneers, P., Schmidtlein, S., Somers, B., Kerchove, R. V. D., Rocchini, D., and Lenoir, J.: A Unified Framework to Model the Potential and Realized Distributions of Invasive Species within the Invaded Range, Diversity and Distributions, 23, 806–819, https://doi.org/10.1111/ddi.12566, 2017.

Hengl, T., Heuvelink, G. B. M., and Stein, A.: A Generic Framework for Spatial Prediction of Soil Variables Based on Regression-Kriging, Geoderma, 120, 75–93, https://doi.org/10.1016/j.geoderma.2003.08.018, 2004.

Heuvelink, G. B. M. and Webster, R.: Spatial Statistics and Soil Mapping: A Blossoming Partnership under Pressure, Spatial Statistics, 50, 100 639, https://doi.org/10.1016/j.spasta.2022.100639, 2022.

Khaledian, Y. and Miller, B. A.: Selecting Appropriate Machine Learning Methods for Digital Soil Mapping, Applied Mathematical Modelling, 81, 401–418, https://doi.org/10.1016/j.apm.2019.12.016, 2020.

Kuhn, M. and Wickham, H.: Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles., 2020.

Lamichhane, S., Kumar, L., and Wilson, B.: Digital Soil Mapping Algorithms and Covariates for Soil Organic Carbon Mapping and Their Implications: A Review, Geoderma, 352, 395–413, https://doi.org/10.1016/j.geoderma.2019.05.031, 2019.

Lindsay, J. B.: Whitebox GAT: A Case Study in Geomorphometric Analysis., Computers & Geosciences, 95, 75–84, 2016.

Linnenbrink, J., Milà, C., Ludwig, M., and Meyer, H.: kNNDM CV: *K*-Fold Nearest-Neighbour Distance Matching Cross-Validation for Map Accuracy Estimation, Geoscientific Model Development, 17, 5897–5912, https://doi.org/10.5194/gmd-17-5897-2024, 2024.

Ma, T., Brus, D. J., Zhu, A.-X., Zhang, L., and Scholten, T.: Comparison of Conditioned Latin Hypercube and Feature Space Coverage Sampling for Predicting Soil Classes Using Simulation from Soil Maps, Geoderma, 370, 114 366, https://doi.org/10.1016/j.geoderma.2020.114366, 2020.

Malone, B. P., Minansy, B., and Brungard, C.: Some Methods to Improve the Utility of Conditioned Latin Hypercube Sampling, PeerJ, 7, e6451, https://doi.org/10.7717/peerj.6451, 2019.

Maxwell, A. E. and Shobe, C. M.: Land-Surface Parameters for Spatial Predictive Mapping and Modeling, Earth-Science Reviews, 226, 103 944, https://doi.org/10.1016/j.earscirev.2022.103944, 2022.

Meinshausen, N.: Quantile Regression Forests, Journal of Machine Learning Research, 7, 983–999, 2006.

Meyer, H. and Pebesma, E.: Machine Learning-Based Global Maps of Ecological Variables and the Challenge of Assessing Them, Nature Communications, 13, 2208, https://doi.org/10.1038/s41467-022-29838-9, 2022.

Meyer, H., Ludwig, M., Milà, C., Linnenbrink, J., and Schumacher, F.: The CAST Package for Training and Assessment of Spatial Prediction Models in R, https://doi.org/10.48550/arXiv.2404.06978, 2024.

Milà, C., Mateu, J., Pebesma, E., and Meyer, H.: Nearest Neighbour Distance Matching Leave-One-Out Cross-Validation for Map Validation, Methods in Ecology and Evolution, 13, 1304–1316, https://doi.org/10.1111/2041-210X.13851, 2022.

Miller, B. A., Koszinski, S., Wehrhan, M., and Sommer, M.: Impact of Multi-Scale Predictor Selection for Modeling Soil Properties, Geoderma, 239–240, 97–106, https://doi.org/10.1016/j.geoderma.2014.09.018, 2015.

Minasny, B. and McBratney, A. B.: A Conditioned Latin Hypercube Method for Sampling in the Presence of Ancillary Information, Computers & Geosciences, 32, 1378–1388, https://doi.org/10.1016/j.cageo.2005.12.009, 2006.

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, D., Tuve, C., and Widyatmanti, W.: Digital Mapping of Peatlands – A Critical Review, Earth-Science Reviews, 196, 102 870, https://doi.org/10.1016/j.earscirev.2019.05.014, 2019.

Minasny, B., Adetsu, D. V., Aitkenhead, M., Artz, R. R. E., Baggaley, N., Barthelmes, A., Beucher, A., Caron, J., Conchedda, G., Connolly, J., Deragon, R., Evans, C., Fadnes, K., Fiantis, D., Gagkas, Z., Gilet, L., Gimona, A., Glatzel, S., Greve, M. H., Habib, W., Hergoualc'h, K., Hermansen, C., Kidd, D. B., Koganti, T., Kopansky, D., Large, D. J., Larmola, T., Lilly, A., Liu, H., Marcus, M., Middleton, M., Morrison, K., Petersen, R. J., Quaife, T., Rochefort, L., Rudiyanto, Toca, L., Tubiello, F. N., Weber, P. L., Weldon, S., Widyatmanti, W., Williamson, J., and Zak, D.: Mapping and Monitoring Peatland Conditions from Global to Field Scale, Biogeochemistry, https://doi.org/10.1007/s10533-023-01084-1, 2023.

Molla, A., Zhang, W., Zuo, S., Ren, Y., and Han, J.: A Machine Learning and Geostatistical Hybrid Method to Improve Spatial Prediction Accuracy of Soil Potentially Toxic Elements, Stochastic Environmental Research and Risk Assessment, 37, 681–696, https://doi.org/10.1007/s00477-022-02284-1, 2023.

Murphy, P. N. C., Ogilvie, J., Connor, K., and Arp, P. A.: Mapping Wetlands: A Comparison of Two Different Approaches for New Brunswick, Canada, Wetlands, 27, 846–854, https://doi.org/10.1672/0277-5212(2007)27[846:MWACOT]2.0.CO;2, 2007.

Murphy, P. N. C., Ogilvie, J., and Arp, P.: Topographic Modelling of Soil Moisture Conditions: A Comparison and Verification of Two Models, European Journal of Soil Science, 60, 94–109, https://doi.org/10.1111/j.1365-2389.2008.01094.x, 2009.

Murphy, P. N. C., Ogilvie, J., Meng, F.-R., White, B., Bhatti, J. S., and Arp, P. A.: Modelling and Mapping Topographic Variations in Forest Soils at High Resolution: A Case Study, Ecological Modelling, 222, 2314–2332, https://doi.org/10.1016/j.ecolmodel.2011.01.003, 2011.

Newman, D. R., Saurette, D. D., Cockburn, J. M. H., Dragut, L., and Lindsay, J. B.: Assessing Spatially Heterogeneous Scale Representation with Applied Digital Soil Mapping, Environmental Modelling & Software, 160, 105 612, https://doi.org/10.1016/j.envsoft.2022.105612, 2023.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M. E., and Papritz, A.: Evaluation of Digital Soil Mapping Approaches with Large Sets of Environmental Covariates, SOIL, 4, 1–22, https://doi.org/10.5194/soil-4-1-2018, 2018.

Ofstad, F.: Helicopter-Borne Magnetic and Radiometric Geophysical Survey in Romsdalsfjorden, Møre Og Romsdal, Tech. Rep. 2015.015, Geological Survey of Norway, 2015.

O'Leary, D., Brown, C., and Daly, E.: Digital Soil Mapping of Peatland Using Airborne Radiometric Data and Supervised Machine Learning – Implication for the Assessment of Carbon Stock, Geoderma, 428, 116 086, https://doi.org/10.1016/j.geoderma.2022.116086, 2022.

Parry, L., West, L., Holden, J., and Chapman, P.: Evaluating Approaches for Estimating Peat Depth, Journal of Geophysical Research: Biogeosciences, 119, 567–576, https://doi.org/10.1002/2013JG002411, 2014.

Probst, P., Wright, M. N., and Boulesteix, A.-L.: Hyperparameters and Tuning Strategies for Random Forest, WIREs Data Mining and Knowledge Discovery, 9, e1301, https://doi.org/10.1002/widm.1301, 2019.

QGIS Development Team: QGIS Geographic Information System, QGIS Association, 2024.

Quinn, P., Beven, K., Chevallier, P., and Planchon, O.: The Prediction of Hillslope Flow Paths for Distributed Hydrological Modelling Using Digital Terrain Models, Hydrological Processes, 5, 59–79, https://doi.org/10.1002/hyp.3360050106, 1991.

Riihimäki, H., Kemppinen, J., Kopecký, M., and Luoto, M.: Topographic Wetness Index as a Proxy for Soil Moisture: The Importance of Flow-Routing Algorithm and Grid Resolution, Water Resources Research, 57, e2021WR029 871, https://doi.org/10.1029/2021WR029871, 2021.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2017.

Roudier, P.: Clhs: A R Package for Conditioned Latin Hypercube Sampling., 2011.

Saurette, D. D., Heck, R. J., Gillespie, A. W., Berg, A. A., and Biswas, A.: Divergence Metrics for Determining Optimal Training Sample Size in Digital Soil Mapping, Geoderma, 436, 116 553, https://doi.org/10.1016/j.geoderma.2023.116553, 2023.

Schönauer, M. and Maack, J.: R-Code for Calculating Depth-to-Water (DTW) Maps Using GRASS GIS, https://doi.org/10.5281/zenodo.5718133, 2021.

Schönauer, M., Väätäinen, K., Prinz, R., Lindeman, H., Pszenny, D., Jansen, M., Maack, J., Talbot, B., Astrup, R., and Jaeger, D.: Spatio-Temporal Prediction of Soil Moisture and Soil Strength by Depth-to-Water Maps, International Journal of Applied Earth Observation and Geoinformation, 105, 102 614, https://doi.org/10.1016/j.jag.2021.102614, 2021.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A.: Hyperparameter Tuning and Performance Assessment of Statistical and Machine-Learning Algorithms Using Spatial Data, Ecological Modelling, 406, 109–120, https://doi.org/10.1016/j.ecolmodel.2019.06.002, 2019.

Shrestha, D. L. and Solomatine, D. P.: Machine Learning Approaches for Estimation of Prediction Interval for the Model Output, Neural Networks, 19, 225–235, https://doi.org/10.1016/j.neunet.2006.01.012, 2006.

Simensen, T., Erikstad, L., and Halvorsen, R.: Diversity and Distribution of Landscape Types in Norway, Norsk Geografisk Tidsskrift - Norwegian Journal of Geography, 75, 79–100, https://doi.org/10.1080/00291951.2021.1892177, 2021.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A.: Conditional Variable Importance for Random Forests, BMC Bioinformatics, 9, 307, https://doi.org/10.1186/1471-2105-9-307, 2008.

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., and Pásztor, L.: Mapping Soil Hydraulic Properties Using Random-Forest-Based Pedotransfer Functions and Geostatistics, Hydrology and Earth System Sciences, 23, 2615–2635, https://doi.org/10.5194/hess-23-2615-2019, 2019.

Takoutsing, B. and Heuvelink, G. B.: Comparing the Prediction Performance, Uncertainty Quantification and Extrapolation Potential of Regression Kriging and Random Forest While Accounting for Soil Measurement Errors, Geoderma, 428, 116 192, https://doi.org/10.1016/j.geoderma.2022.116192, 2022.

Tarboton, D. G.: A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models, Water Resources Research, 33, 309–319, https://doi.org/10.1029/96WR03137, 1997.

Wadoux, A. M. J.-C., Brus, D. J., and Heuvelink, G. B. M.: Sampling Design Optimization for Soil Mapping with Random Forest, Geoderma, 355, 113 913, https://doi.org/10.1016/j.geoderma.2019.113913, 2019.

Wadoux, A. M. J. C., Minasny, B., and McBratney, A. B.: Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions, Earth-Science Reviews, 210, 103 359, https://doi.org/10.1016/j.earscirev.2020.103359, 2020.

Wadoux, A. M. J. C., Heuvelink, G. B. M., de Bruin, S., and Brus, D. J.: Spatial Cross-Validation Is Not the Right Way to Evaluate Map Accuracy, Ecological Modelling, 457, 109 692, https://doi.org/10.1016/j.ecolmodel.2021.109692, 2021.

Wang, L. and Liu, H.: An Efficient Method for Identifying and Filling Surface Depressions in Digital Elevation Models for Hydrologic Analysis and Modelling, International Journal of Geographical Information Science, 20, 193–213, https://doi.org/10.1080/13658810500433453, 2006.

Wright, M. N. and Ziegler, A.: Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, Journal of Statistical Software, 77, 1–17, https://doi.org/10.18637/jss.v077.i01, 2017.

Wu, Q. and Brown, A.: 'whitebox': 'WhiteboxTools' R Frontend, 2022.