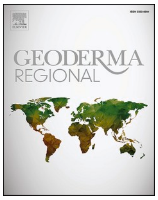
Geoderma Regional 29 (2022) e00495

Contents lists available at ScienceDirect

Geoderma Regional

journal homepage: www.elsevier.com/locate/geodrs



Incorporating spatial uncertainty maps into soil sampling improves digital soil mapping classification accuracy in Ontario, Canada

Christopher Blackford [a,*], Brandon Heung [b], Kara L. Webster [a]

[a] Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste Marie, ON, Canada
[b] Faculty of Agriculture, Dalhousie University, 15 Cox Road, Truro, NS, Canada

ARTICLE INFO

Keywords:
Digital soil mapping
Uncertainty
Adaptive sampling
Spodosols
Luvisols

ABSTRACT

Digital soil mapping combines soil plot data with environmental datasets to model variation in soil properties across a landscape. The quality of a digital soil map depends on both the quantity and distribution of soil plots within the study extent. Field campaigns to acquire soil data are time intensive and costly to undertake, requiring training and deployment of field crews and soil processing/analytical costs. Therefore, it is important to optimize site selection and sampling intensity to maximize digital soil map accuracy and minimize field costs. In many cases, soil sampling occurs across several years to gather sufficient soil data. Between successive field campaigns, preliminary digital soil maps and their corresponding uncertainty estimates can be generated. We hypothesize that preliminary uncertainty maps can be useful to guide sampling in subsequent field seasons by targeting areas of high uncertainty to significantly improve model accuracy. This hypothesis was tested by simulating a multi-year soil sampling campaign using an extensive soil moisture regime and soil texture dataset from the Hearst Forest in northeastern Ontario, Canada. We quantified how soil maps and soil models changed as new data points were added and how model/map improvement was influenced by performing additional sampling in areas of high uncertainty. We used multiple uncertainty metrics (Ignorance Uncertainty, Exaggeration Uncertainty and Confusion Index) and tested multiple levels of sampling intensity. The results showed modest but statistically significant improvements in model accuracy when subsequent sampling was targeted in high uncertainty areas (treatments) compared to sampling in random areas (controls) (38.7% control accuracy compared to 39.8%/40.4%/40.3% for moisture regime and 23.1% control accuracy compared to 24.3%/25%/24.9% for textural class). There were no significant differences in model performance between the three uncertainty metrics. The most common textural and moisture regime classes in the soil dataset rarely occurred in areas of high uncertainty suggesting that the environmental covariates used in the study tracked real soil variation. As subsequent sampling intensity increased, model performance increased as well (both in the control and treatment groups). There was also a significant treatment × sampling intensity interaction meaning that uncertainty guided sampling was increasingly beneficial as sampling effort increased. This paper demonstrates a proof of concept that generating preliminary uncertainty maps in digital soil mapping can be a useful tool for informing future field soil sample collections to improve model performance.

1. Introduction

Digital soil mapping (DSM) predicts soil variation across space by modelling the relationship between georeferenced soil observations and environmental covariates (McBratney et al., 2003; Scull et al., 2003; Minasny and McBratney, 2016). DSM accuracy depends not only on the size and quality of the input soil data (Grinand et al., 2008; Yang et al., 2016), but also on how well the soil sampling locations capture the variability of the soil and environmental properties (Biswas and Zhang, 2018). Therefore, to generate high quality soil maps, it is important to design soil sampling campaigns to optimize sampling size and sampling locations, given limited sampling time and financial costs. To this end, various approaches have been proposed in the literature to optimize soil sampling. These approaches aim to distribute sampling in structured ways throughout the study extent such that spatial and/or environmental variability is captured during sampling (Minasny and McBratney, 2006; Brus and Heuvelink, 2007; Vašát et al., 2010; Szatmári et al., 2019; Wadoux et al., 2019). Incorporating structured approaches to soil

* Corresponding author.
E-mail address: chris.blackford@ontario.ca (C. Blackford).

https://doi.org/10.1016/j.geodrs.2022.e00495
Received 7 December 2021; Received in revised form 2 March 2022; Accepted 6 March 2022
Available online 10 March 2022

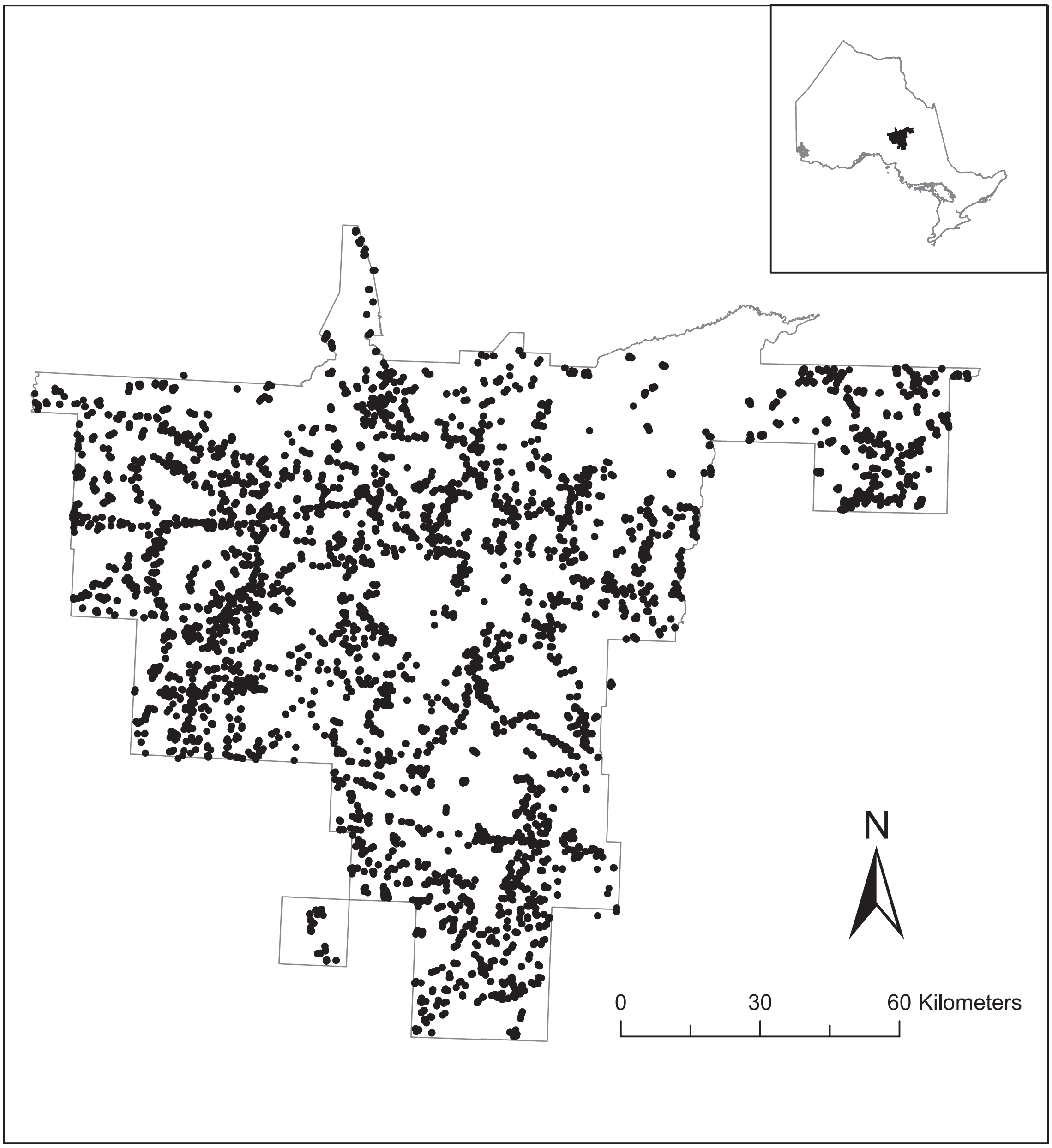


Fig. 1. Distribution of soil points within Hearst Forest. Inset: Hearst Forest with respect to Ontario, Canada.

sampling has been shown to improve digital soil map predictive accuracy. (for a review, see: Biswas and Zhang, 2018).

In tandem with structured sampling, incorporating previously collected soil data into DSM projects is another way to improve DSM performance. Legacy or existing soil data can supplement new soil sampling campaigns to generate a larger soil dataset (Bui and Moran, 2003; Odgers et al., 2014). Additionally, adaptive sampling designs, which use previous soil samples and soil spatial dependence models to guide additional sampling, can be employed to optimize sampling across space (Marchant and Lark, 2006; Musafer and Thompson, 2016). Finally, it is also possible to use uncertainty estimates of preliminary soil map predictions and then conduct subsequent sampling in areas of high uncertainty. For example, Huang et al. (2020) investigated the use of uncertainty maps for improving Gamma-ray potassium measurements on an agricultural field. In that study, Huang et al. (2020) acquired 10 initial soil samples; predicted the spatial distribution of the target variable using linear mixed model; generated local uncertainty estimates; and selected the next sampling locations based on uncertainty and travel time. Their results demonstrated that the adaptive sampling approach was more effective than applying a gridded or simple random sampling approach. Similarly, Stumpf et al. (2017) applied an approach for predicting particle size fractions and showed that uncertainty-guided sampling resulted in increased accuracy and decreased uncertainty. These uncertainty-guided approaches take advantage of a benefit of DSM over conventional soil mapping, in that DSM can provide uncertainty estimates of the soil model and map predictions (Malone et al., 2011; Minasny and McBratney, 2016).

In digital soil mapping, uncertainty can arise from errors in soil measurement, geolocation, digitization, data generalization, and interpolation (Arrouays et al., 2014b), as well as from modelling bias, parameterization, and predictive modelling errors associated with modelling the relationship between the predictor and response variables (Minasny and McBratney, 2002). When applied to DSM, a key outcome is the evaluation of spatially explicit uncertainty (i.e., “local error”), which provides an estimation of the predictive model uncertainty on a pixel-by-pixel basis. Uncertainty maps not only show where model biases may occur, but can also influence soil map application, since users may be more hesitant to rely on the soil map in areas of high uncertainty. For predicting continuous variables, local error may be represented as a prediction interval (e.g., the 90% prediction interval width; Arrouays et al., 2014a). In comparison, uncertainty in categorical variable predictions may be quantified through analysis of class probability layers. These layers may be generated using specific machine-learning models (e.g., support vector machines, Random Forest, and *k*-nearest neighbors) or by using a bootstrapping procedure (Heung et al., 2017). Some examples of uncertainty metrics for categorical variables include ignorance uncertainty (IU, Leung et al., 1993; Goodchild et al., 1994; Zhu, 1997; Heung et al., 2017), exaggeration uncertainty (EU, Zhu, 1997), and confusion index (CI, Burrough et al., 1997; Chaney et al., 2016). These metrics are calculated using class probability layers and evaluate the divergence of probability values across classes, whereby pixels with similar probability values for multiple classes will result in higher uncertainty. It has been suggested that these uncertainty analyses can be used to evaluate the quality of predictions, carry out sensitivity analysis on model variables, and identify where to allocate resources for the purposes of reducing uncertainty (Minasny and Bishop, 2008).

Despite the findings of Stumpf et al. (2017) and Huang et al. (2020), there has been little investigation into the use of uncertainty-guided sampling when categorical variables are being predicted. Furthermore, there are multiple metrics representing uncertainty for categorical variables as well (e.g., IU, EU, and CI) and it is unclear which, if any, of these metrics are best to inform future sampling locations. In this paper, we tested if using uncertainty evaluations of a digital soil map to guide subsequent sampling improves digital soil mapping performance. We used a large existing soil dataset in Hearst Forest, Ontario (Canada) to predict the categorical variables soil moisture regime and textural class. We simulated an adaptive soil sampling campaign by generating an initial DSM on a subset of the soil dataset; running an uncertainty analysis of the predictions; then updating the model/map with additional sample locations to simulate multiple field seasons. In Study 1, we tested: Does additional sampling in areas of high uncertainty (based on a preliminary digital soil map) significantly improve DSM performance compared to additional random sampling? This was tested using three different uncertainty metrics (IU, EU, and CI) to determine if certain uncertainty metrics perform better than others. We expected all digital soil models to improve with additional sampling but for that improvement to be more significant when these points come from areas of high uncertainty. Additionally, we qualitatively explored the relationship between uncertainty and soil class frequency in the original dataset. We expected soil classes that are underrepresented in the soil dataset will be found in areas of higher uncertainty.

In Study 2, we quantified how increasing sampling effort improved digital soil model performance. This was done by adding an increasing number of additional sampling points in high uncertainty areas versus adding additional sampling points in random areas. We expected increasing sampling within high uncertainty areas will improve model performance compared to random sampling, although there may be a point at which we observe diminishing returns. It is unclear what, if any, interaction there is between uncertainty-guided sampling and sampling intensity (i.e., are the effects of sampling in areas of high uncertainty more noticeable when added few or many additional soil samples?)

Table 1
Environmental covariates used in DSM for Hearst Forest.

Covariate	Measuring	Data source
Aspect	Local Relief	DEM
Downslope curvature	Local Relief	DEM
General curvature	Local Relief	DEM
Local downslope curvature	Local Relief	DEM
Local curvature	Local Relief	DEM
Local upslope curvature	Local Relief	DEM
Maximum curvature	Local Relief	DEM
Minimum curvature	Local Relief	DEM
Multiresolution ridgetop flatness (Gallant and Dowling, 2003)	Local Relief	DEM
Multiresolution valley bottom flatness (Gallant and Dowling, 2003)	Local Relief	DEM
Multi-scale topographic position index	Local Relief	DEM
Mid-slope position	Local Relief	DEM
Normalized height	Local Relief	DEM
Planar curvature	Local Relief	DEM
Profile curvature	Local Relief	DEM
Real surface area	Local Relief	DEM
Slop	Local Relief	DEM
Slope height	Local Relief	DEM
Tangential curvature	Local Relief	DEM
Topographic negative openness	Local Relief	DEM
Total curvature	Local Relief	DEM
Topographic positive openness	Local Relief	DEM
Terrain ruggedness index (Riley et al., 1999)	Local Relief	DEM
Terrain surface concavity	Local Relief	DEM
Terrain surface convexity	Local Relief	DEM
Upslope curvature	Local Relief	DEM
Upslope height	Local Relief	DEM
Valley depth	Landscape relief	DEM
Catchment area	Hydrology	DEM
Catchment slope	Hydrology	DEM
Modified catchment area	Hydrology	DEM
Topographic wetness index	Hydrology	DEM
Distance to stream	Landscape relief	Hydrology shapefile
Distance to water body	Landscape relief	Hydrology shapefile
Overstory height	Organisms	FRI shapefile
Understory height	Organisms	FRI shapefile
Overstory leading species	Organisms	FRI shapefile
Understory leading species	Organisms	FRI shapefile
Bedrock geology	Parent material	Geology shapefile
Quaternary geology	Parent material	Geology shapefile
Distance from centre of study extent	Spatial position	NA
Distance from Northeast extent point	Spatial position	NA
Distance from Northwest extent point	Spatial position	NA
Distance from Southeast extent point	Spatial position	NA
Distance from Southwest extent point	Spatial position	NA
Distance along x axis	Spatial position	NA
Distance along y axis	Spatial position	NA

2. Methods

2.1. Study area

The Hearst Forest, located in northeastern Ontario, Canada (49°41’’16" N, 83°40’21" W) is a large, managed forest (approximately 15,218 km^2) (Fig. 1). This forest was chosen as our study area because of the availability of a high-resolution light detection and ranging (LiDAR) derived digital elevation model (DEM), and the availability of a large soil pedon dataset. A full description of the study area and soil data can be found in Blackford et al. (2021). The forest sits on the Precambrian Shield, covered by Quaternary age sediments (Blackburn et al., 1985; Mackasey et al., 1974; Thurston et al., 1991). Clay plains occupy the northern and central areas of the forest (commonly known as the Clay Belt), which were deposited during inundation by the proglacial Lake Barlow-Ojibway about 9000 years ago (Dyke, 2004). In other areas of the forest, loamy and sandy soils can be found (Hearst Forest Management, 2019). Many soils are poorly drained and organic soil is common throughout (Hearst Forest Management, 2019). Esker complexes from previous glaciation can be found in the centre of the forest. Overall, the

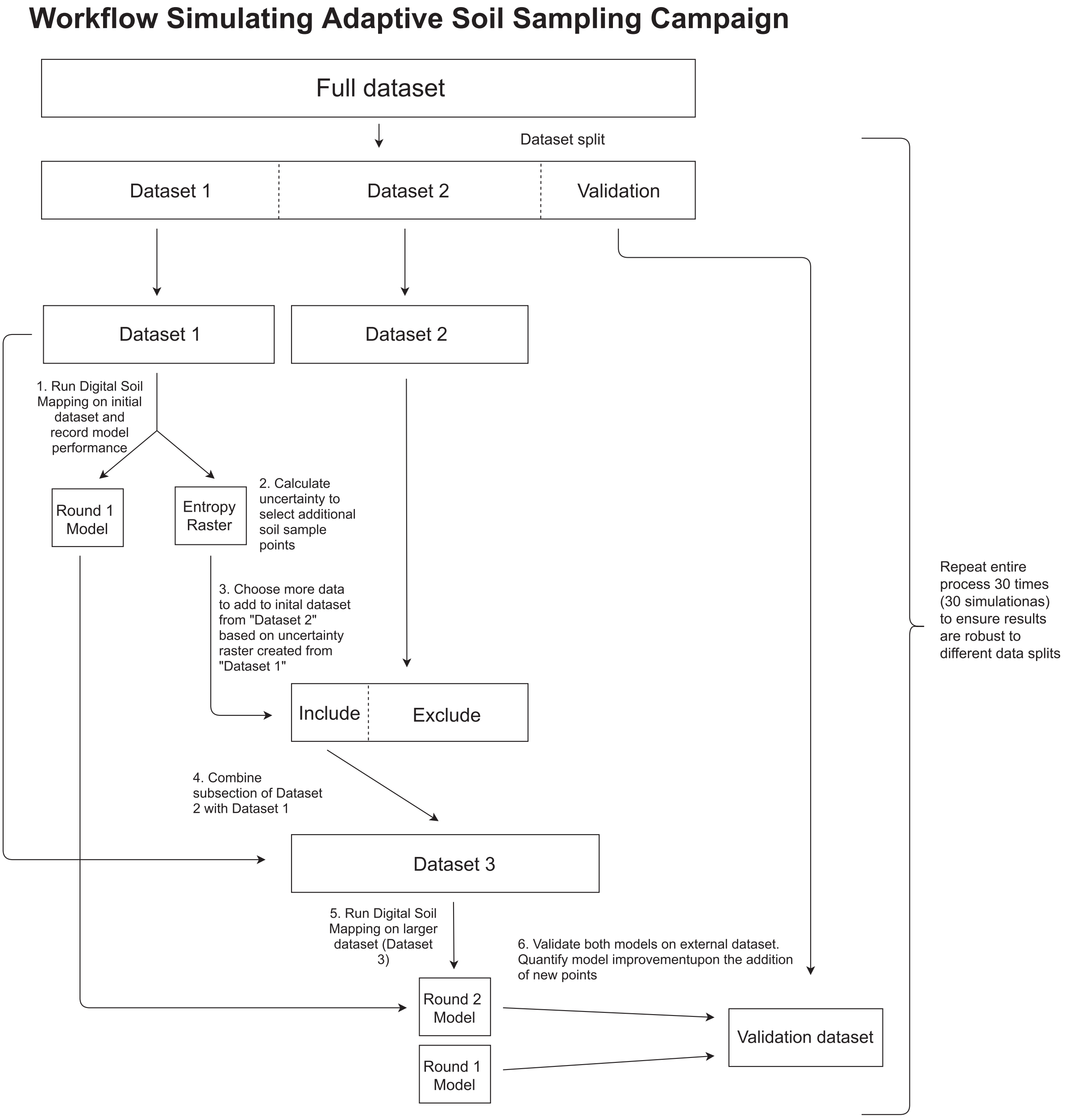

Workflow Simulating Adaptive Soil Sampling Campaign
Full dataset
Dataset split
Dataset 1
Dataset 2
Validation
Dataset 1
Dataset 2
1. Run Digital Soil Mapping on initial dataset and record model performance
Round 1 Model
Entropy Raster
2. Calculate uncertainty to select additional soil sample points
3. Choose more data to add to inital dataset from "Dataset 2" based on uncertainty raster created from "Dataset 1"
Include
Exclude
4. Combine subsection of Dataset 2 with Dataset 1
Dataset 3
5. Run Digital Soil Mapping on larger dataset (Dataset 3)
6. Validate both models on external dataset. Quantify model improvementupon the addition of new points
Round 2 Model
Round 1 Model
Validation dataset
Repeat entire process 30 times (30 simulationas) to ensure results are robust to different data splits

Fig. 2. Workflow for testing DSM upon addition of more soil datapoints. This workflow is repeated for multiple splits of Dataset 1/Dataset 2/ Validation dataset to ensure results are consistent no matter how dataset is split.

Hearst Forest is of moderate relief and representative of the 3E Boreal Shield ecoregion within Ontario (Crins et al., 2009).

2.2. Soil data

A combination of previously gathered federal, provincial, and targeted soil pedon data was used. These data came from Forest Resource Inventory surveys, National Forest Inventory plots, Provincial Growth and Yield plots, and Forest Ecological Classification plots. Altogether, these datasets contain 7893 spatial soil records within the Hearst (Fig. 1). From this dataset, we chose to model soil moisture regime and soil textural class (Johnson et al., 2015) as these were the most abundant soil variables measured with 7734 records of soil moisture regime class and 7213 records of textural class. Soil moisture regime is a relative ranking system that describes the soil's moisture supply during the growing seasons and is based on soil texture, pore patterns, soil depth, landscape position, and drainage. Textural classes are defined as the "effective texture" found at a site (i.e., the dominant soil texture of the pedon). These textural classes were either classified as organic, or if a mineral soil was present, by the relative proportions of sand, silt, and clay.

2.3. Environmental data

Table 1 shows the list of environmental predictors layers used to link to soil variation across space. These environmental predictors represented relief, hydrology, biota, geology, and spatial distance within the study area. A 10 × 10 m resolution LiDAR DEM was used to derive topographic metrics. The DEM was "smoothed" (locally averaged) by passing a 81 × 81 cell moving window filter across the extent which averaged the elevation vales within the window. This smoothing process was performed to reduce the effects of spatially uncorrelated noise from LiDAR derived DEMs and to remove their anomalous pits and peaks (Li et al., 2011). The derived topographic metrics characterized local-scale morphometry (e.g., slope, aspect, curvature), landscape-scale morphometry (e.g., topographic position index), and hydrology (e.g., topographic wetness index, valley depth, catchment area). These metrics were calculated using the SAGA program (Conrad et al., 2015), run through R (R core team, 2019). A distance to river and lakes layer was generated to correspond to landscape-relief patterns. Two geology layers were used to represent bedrock and sedimentation characteristics. Forestry layers of overstory height, understory height and overstory leading species (i.e., most common species) were used to represent biotic factors. Finally, Euclidean distance fields were used to incorporate spatial position into predictions (i.e., to account for spatial autocorrelation). Overall, 47 covariates were used (Table 1).

2.4. Machine learning model parametrization

All soil models/maps in this study were generated using the same machine learning modelling design. We used the Random Forest model to link the soil dataset with the environmental covariates as this model has been shown in previous work to perform well in our study area (Blackford et al., 2021). We partitioned the soil-environmental dataset for training and validation by performing a 10-fold cross validation, repeated 10 times. Values of the *mtry* hyperparameter, the main tuning parameter of Random Forest, were tested at values ranging from 3 to 15. (A common recommendation is to set *mtry* as the square root of the number of predictors; in our case 7.) 751 trees were generated for each Random Forest model. All simulations and data analysis were performed in R (R Core Team, 2019), using the *doParallel* (Microsoft Corporation and Weston, 2019), *caret* (Kuhn et al., 2019), *raster* (Hijmans, 2019), *rgdal* (Bivand et al., 2019), *rgeos* (Bivand and Rundel, 2019), and *tidyverse* (Wickham et al., 2019) packages.

2.5. Uncertainty analysis of digital soil map

For a digital soil map generated with the Random Forest model, uncertainty can be quantified for each pixel through the distribution of tree votes (within the Random Forest) between soil classes (Chaney et al., 2016; Stumpf et al., 2017). Three metrics of uncertainty were examined.

Ignorance uncertainty (IU; Leung et al., 1993; Zhu, 1997) is defined as:

$$IU(x) = -\frac{1}{ln(n)}\sum_{i=1}^{n} P_i(x)^{*}ln(P_i(x)) \tag{1}$$

where, for each pixel (x), n is the number of classes of the soil attribute being predicted and $P_i(x)$ is the proportion of votes that class i was given from the Random Forest model. When all the votes (i.e., trees) are assigned to a single class, IU = 0 (smallest uncertainty possible). When votes are assigned equally between classes, IU = 1 (highest uncertainty possible).

Exaggeration uncertainty (EU; Zhu, 1997) is defined as:

$$EU(x) = 1 - P_{max}(x) \tag{2}$$

where, for each pixel (x), P_{max} is the highest proportion of votes given to any class. In other words, P_{max} is the proportion of votes given to the predicted soil class the Random Forest model makes. When all the votes (i.e. trees) are given to a single class, EU = 0. The value of P_{max} cannot be lower than the reciprocal of the number of soil classes being predicted, thus the low limit of EU is bound by: 1–1/n, where n is the number of soil classes being predicted.

Confusion index (CI; Burrough et al., 1997; Chaney et al., 2016) is defined as:

$$CI(x) = 1 - (P_{max}(x) - P_{max-1}(x)) \tag{3}$$

where, for each pixel (x), P_{max} is the highest proportion of votes given to any class and $P_{max\ -\ 1}$ is the second highest proportion of votes given to any class. When all the votes are given to a single class, CI = 0. The lowest value CI could take would occur when the dominant class (P_{max}) has a very similar proportion of votes to the second-most dominant class ($P_{\text{max}\ -\ 1}$) and CI ~ 1. Setting the number of trees the Random Forest generates as an odd number, ensures there will never be an exactly even vote spilt (e.g., here we set number of trees = 751)—although in practice, this rarely occurs.

2.6. DSM workflow to simulate a sequential soil sampling campaign

Our workflow was designed to simulate a DSM project where multi-year soil sampling occurred. We built upon a previously developed procedure for DSM in forestry (Blackford et al., 2021) to generate multiple digital soil models/maps. We simulated a repeated soil sampling campaign in the Hearst Forest wherein the following steps were performed (Fig. 2):

1) The full soil pedon dataset (7893 observations) was randomly split into three datasets: "Dataset 1", "Dataset 2", and a "Validation dataset". The breakdown of number of observations in each dataset differed between study 1 and 2 (see next section.) Datasets 1 was used to simulate an initial soil sampling campaign. Dataset 2 was used to identify potential locations that could be sampled in a subsequent sampling campaign. The validation dataset was withheld to generate an unbiased estimate of model performance.
2) From Dataset 1, we generated an initial digital soil map (Model 1). Uncertainty was calculated using each uncertainty metric described above (IU, EU, CI) using class probability layers returned from the Random Forest model. Because the locations for subsequent soil sampling (i.e., Dataset 2) were acquired from a previously gathered dataset, we only calculated uncertainty in pixels where soils data was recorded, which helped reduce processing time.
3) Dataset 2 was subset to identify areas of high uncertainty to guide further "sampling" (i.e., simulated sampling). In Study 1, for the high uncertainty treatment, 500 points with the highest uncertainty values (IU, EU, or CI) were sampled from Dataset 2 to represent the additional soil sampling. For the random treatment, 500 points were randomly sampled from Dataset 2. In Study 2, a similar sampling approach was taken but we varied the number of points in the subset to determine if there was an interaction between dataset size, uncertainty treatment, and model performance.
4) Random Forest was run on Dataset 3 (Dataset 1 plus subset of Dataset 2) to generate an updated digital soil model (model 2).
5) Model 1 and Model 2 performance was validated using the Validation dataset. Model improvement was assessed by comparing the accuracy and kappa scores of the models.

Since the workflow involved withholding portions of the dataset for later inclusion and validation, model performance depended upon how the dataset was initially split (in Steps 1 and 2). To account for this, we replicated the workflow 30 times (i.e., 30 simulations) to account for performance variability in the initial and updated model (i.e., Models 1 and 2). The code used to run this workflow is freely available online (Blackford, 2022)

2.6.1. Study 1: Does additional sampling in areas of high uncertainty improve model performance?

Using the workflow described above, to test if sampling in areas of high uncertainty improves model performance over random sampling, the number of datapoints in each dataset were as follows: Dataset 1 contained 500 points for both moisture regime and textural class. Dataset 2 (initial size before uncertainty subset) contained 6000 points for both moisture regime and textural class. 500 points were subset from Dataset 2 in either areas of high uncertainty (uncertainty treatment) or randomly (control treatment). Dataset 1 was then combined with this

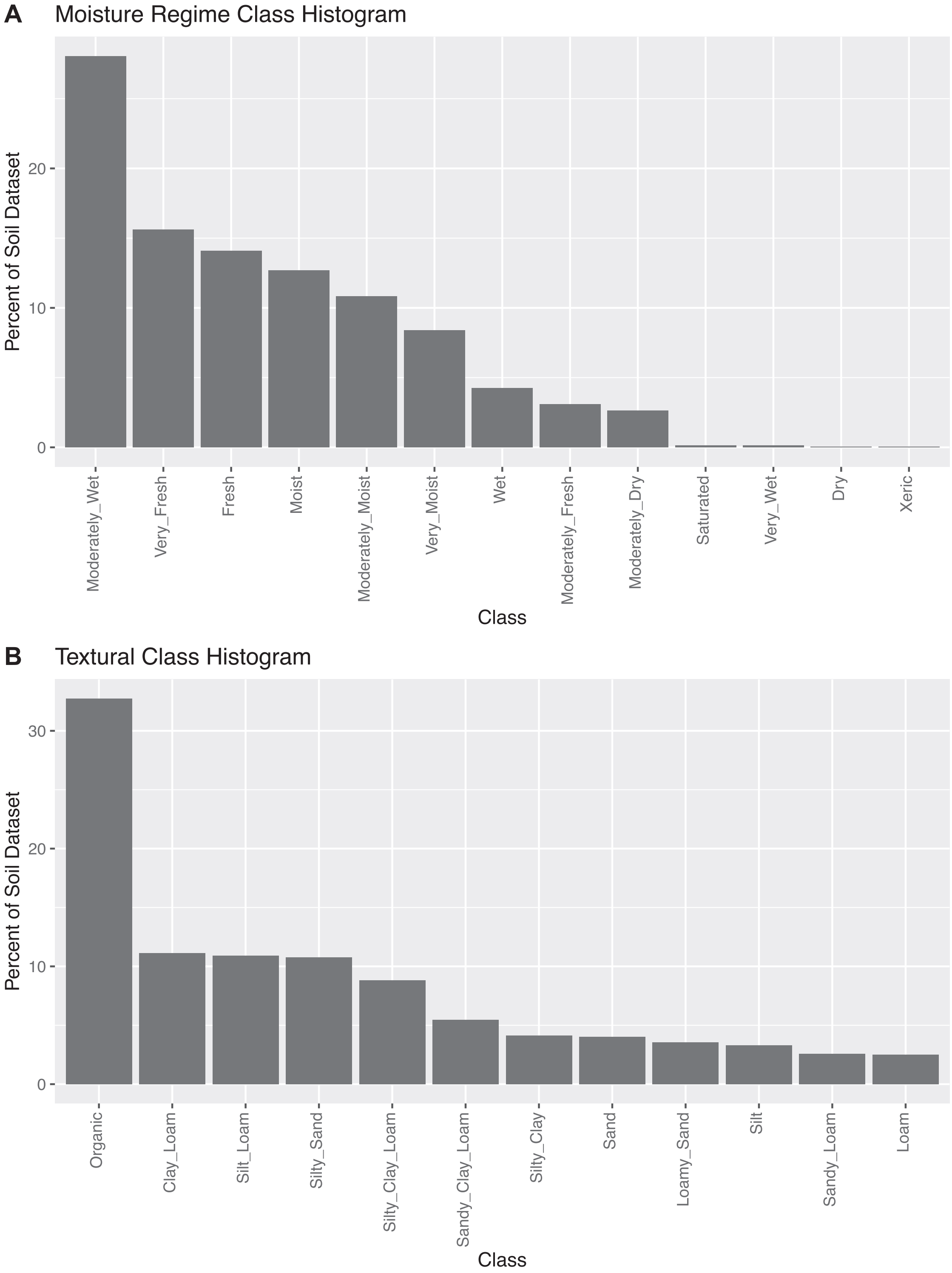


Fig. 3. (A–B). Breakdown of soil classes in (A) moisture regime and (B) textural class.

subset to create Dataset 3 of 1000 points (Dataset 1 + 500 points of Dataset 2). The validation dataset contained the remaining number of datapoints, which was 1234 for moisture regime and 713 for textural class. We recognize that 500 points is often more than would be captured in a single field season, however, these numbers were chosen to test a proof-of-concept that incorporating uncertainty can improve model predictions.

One-sided paired *t*-tests were performed to determine if accuracy and kappa scores were higher in Model 2 compared to Model 1. We performed Wilcoxon signed-rank tests to determine if the Model 2 treatment outperformed the Model 2 control. *p* values and effect sizes (Cohen's *d* and Pearson's *r*).

The distribution of soil classes from the high uncertainty points was also qualitatively compared to the distribution of soil classes across the entire soil dataset (Fig. 3) to determine if certain soil classes are identified as more uncertain to predict than others (i.e., are some soil classes

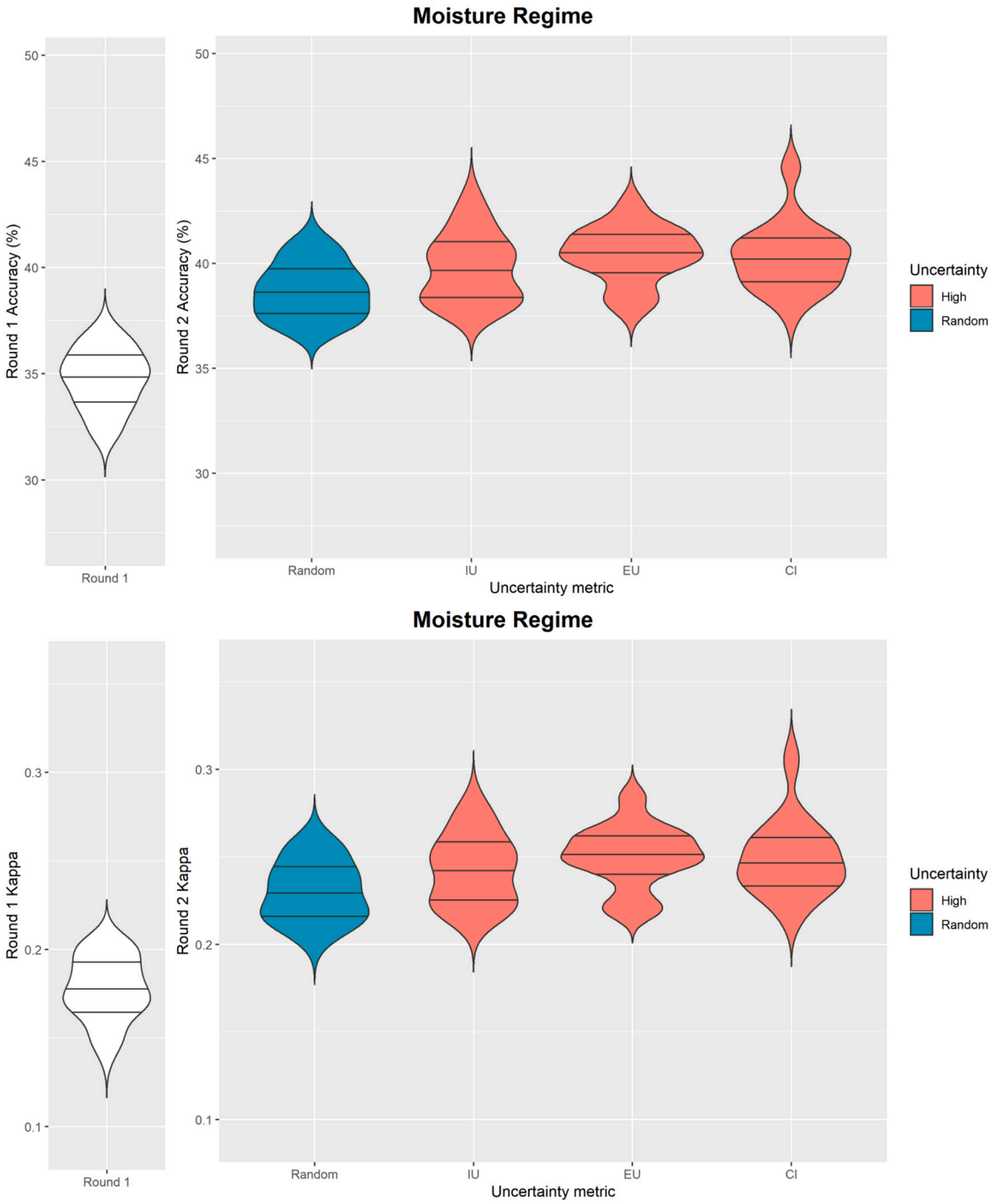


Fig. 4. (A–B). Accuracy and Kappa improvement for moisture regime upon addition of subsequent data selected randomly or from areas of high uncertainty as defined by Ignorance Uncertainty (IU), Exaggeration Uncertainty (EU), and Confidence Index (CI).

overrepresented in areas of high uncertainty).

Finally, to benchmark the performance of round 1 and 2 models, we generated digital soil maps (30 simulations) using the entirety of Dataset 1 and Dataset 2. These "benchmarked" model were used to confirm that differences observed between round 1 and round 2 models were legitimate. If no differences were observed in performance between these benchmarked models and the round 2 models, it would indicate that model performance has hit a theoretical maximum performance. This could be caused by errors in the soil data or machine learning modelling process. In contrast, if these benchmarked models are significantly more accurate than the round 2 models, it means differences between the round 1 and 2 models are valid. In other words, lack of improvement of round 2 models over round 1 models would not be due to the machine learners have hit some theoretical maximum performance wherein performance cannot increase regardless of additional data.

2.6.2. *Study 2: How does sampling effort modulate the effect of incorporating uncertainty into sampling?*

Using the workflow described above, to determine how model improvement is affected by subsequent sampling effort, different quantities of data were sampled from Dataset 2. The number of datapoints in each dataset were as follows: Dataset 1 contained 300 points for both moisture regime and textural class. Dataset 2 (initial size before uncertainty filter) contained 6000 points for both moisture regime and

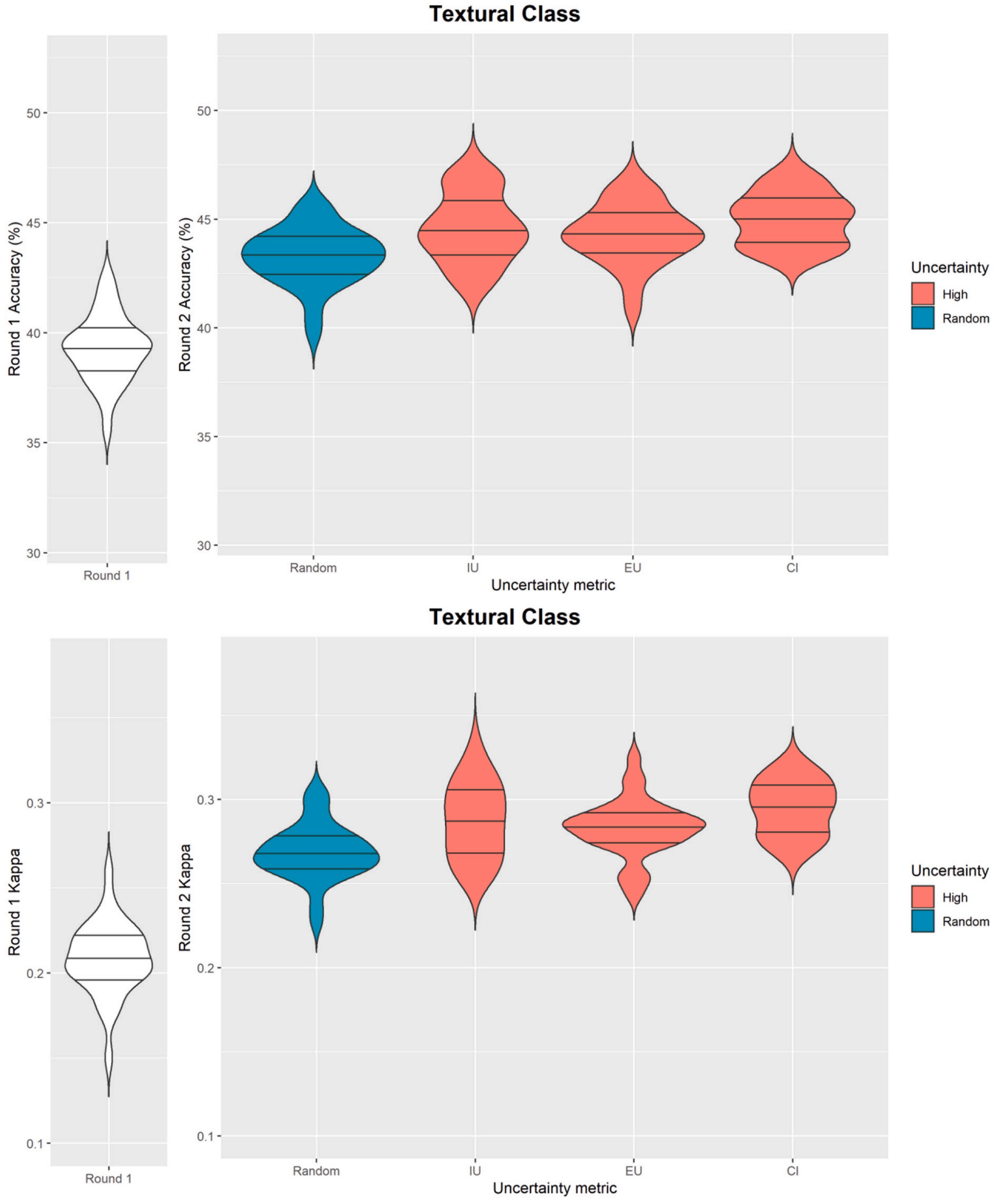


Fig. 5. (A–B). Accuracy and Kappa improvement for textural class upon addition of subsequent data selected randomly or from areas of high uncertainty as defined by Ignorance Uncertainty (IU), Exaggeration Uncertainty (EU), and Confidence Index (CI).

textural class. The top 300, 600, 900, or 1200 points of highest uncertainty were sampled (uncertainty treatment) or randomly sampled (control) from Dataset 2. The validation dataset contained the remaining number of datapoints, which was 1234 points for moisture regime and 713 points for textural class. Because Study 1 showed model improvement did not depend upon uncertainty metric, we performed Study 2 only using the Confidence index (CI) uncertainty metric. Multiple regression was performed to determine the effect of sample size, uncertainty treatment and the sample size × uncertainty interaction on model performance.

3. Results

3.1. Study 1: Does additional sampling in areas of high uncertainty improve model performance?

3.1.1. Comparing soil points of high uncertainty across metrics

The top 500 high uncertainty points showed some similarity across

Table 2

Summary statistics from Wilcoxon signed-rank tests comparing the round 2 control models to the round 2 treatment models.

Variable	Metric	Random (control) mean	IU mean	EU mean	CI mean	IU *p* value	IU effect size (*r*)	EU *p* value	EU effect size (*r*)	CI *p* value	CI effect size (*r*)
M	Accuracy	38.7	39.8	40.4	40.3	3.71E-04	0.65	5.36E-06	0.83	3.16E-05	0.76
M	Kappa	23.1	24.3	25	24.9	5.66E-04	0.63	2.36E-07	0.94	1.95E-05	0.78
T	Accuracy	0.43	0.45	0.44	0.45	3.03E-04	0.66	1.42E-03	0.58	2.03E-05	0.78
T	Kappa	0.27	0.29	0.28	0.29	3.53E-05	0.76	2.53E-04	0.67	9.96E-07	0.89

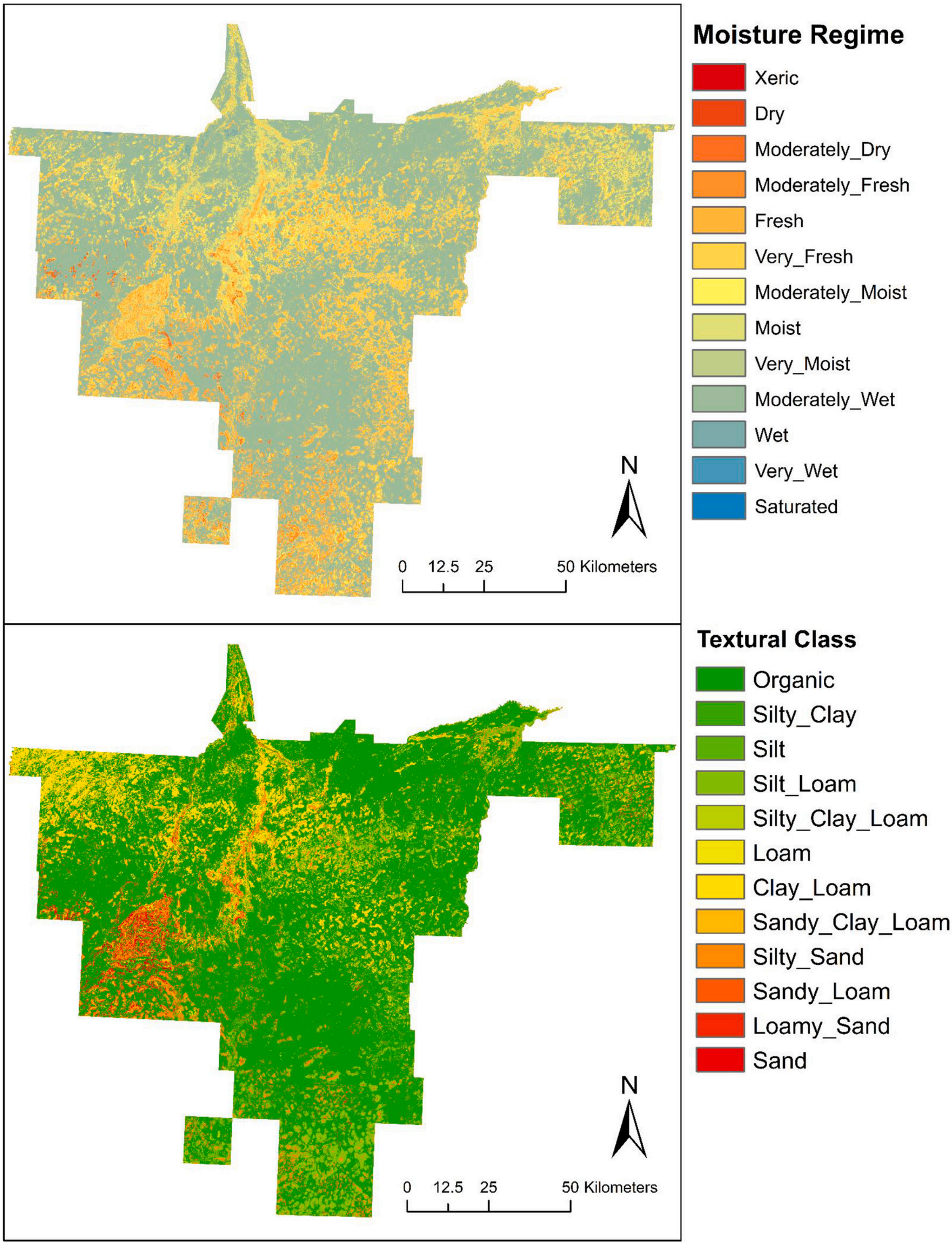


Fig. 6. Study 1 initial Digital Soil Map for Moisture Regime and Textural Class.

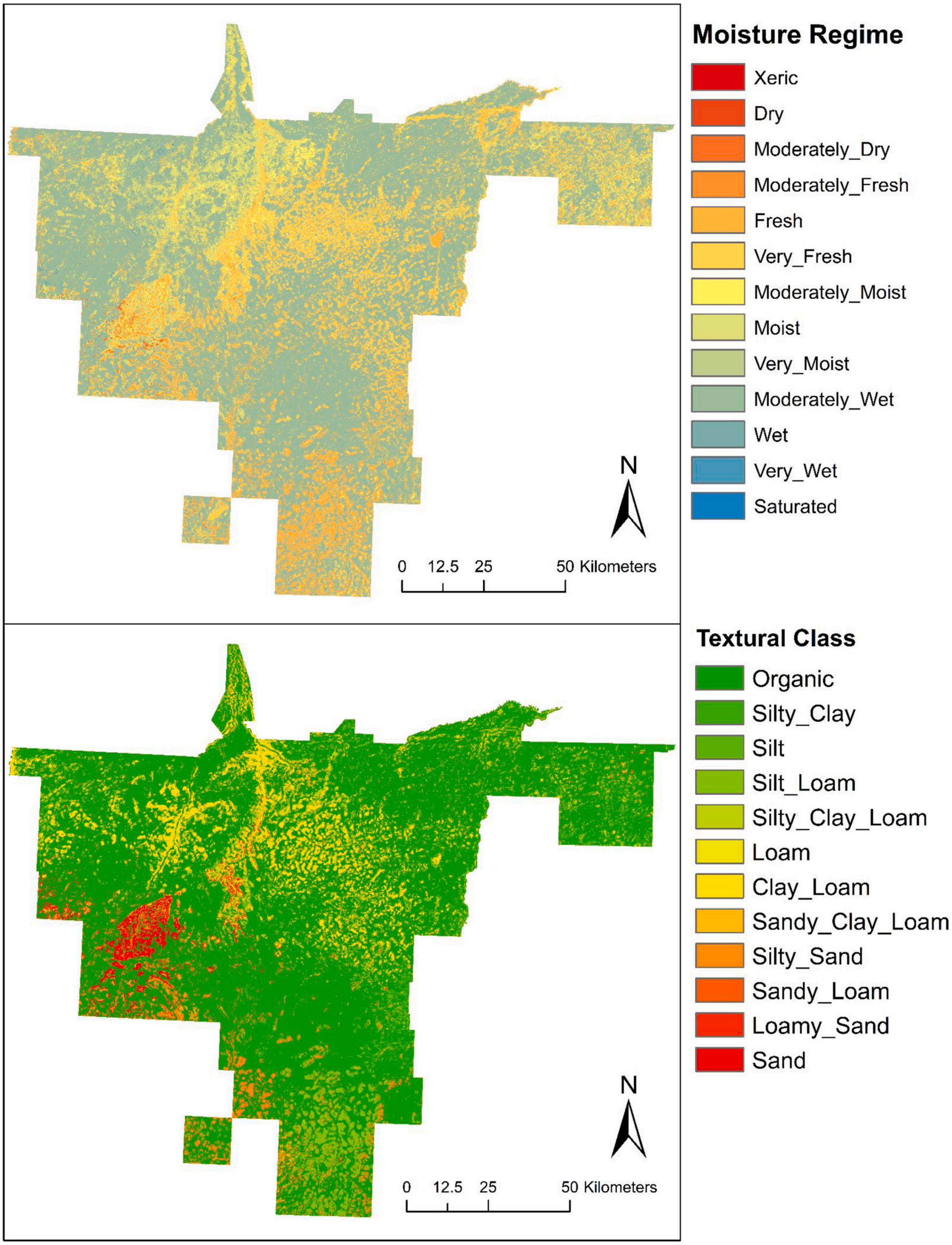


Fig. 7. Digital Soil Maps for inclusion of points from random spatial locations for Moisture Regime and Textural Class.

uncertainty metrics (Fig. S1). On average, for moisture regime, the IU and CI metrics shared 19.6% (98/500) of the same points, the EU and CI metrics shared 40.4% (202/500) of the same points, and the IU and EU metrics shared 49.8% (249/500) of the same points. Across all metrics (IU, EU, and CI), 17.2% (86/500) of the same points were identified as high uncertainty. On average, for textural class, the IU and CI metrics shared 24.8% (124/500) of the same points, the EU and CI metrics shared 44.8% (224/500) of the same points, and the IU and EU metrics shared 57.2% (286/500) of the same points. Across all metrics (IU, EU, and CI), 23% (115/500) of the same points were identified as high uncertainty.

3.1.2. Soil class values association with high uncertainty

In the control treatment, the relative frequency of each soil class was similar to the relative frequency of classes for the entire dataset (blue lines in Figs. S2–S4). For the high uncertainty treatments, there were much fewer points chosen in the dominant class for both moisture regime and textural class (i.e., “Moderately wet” and “Organic”,

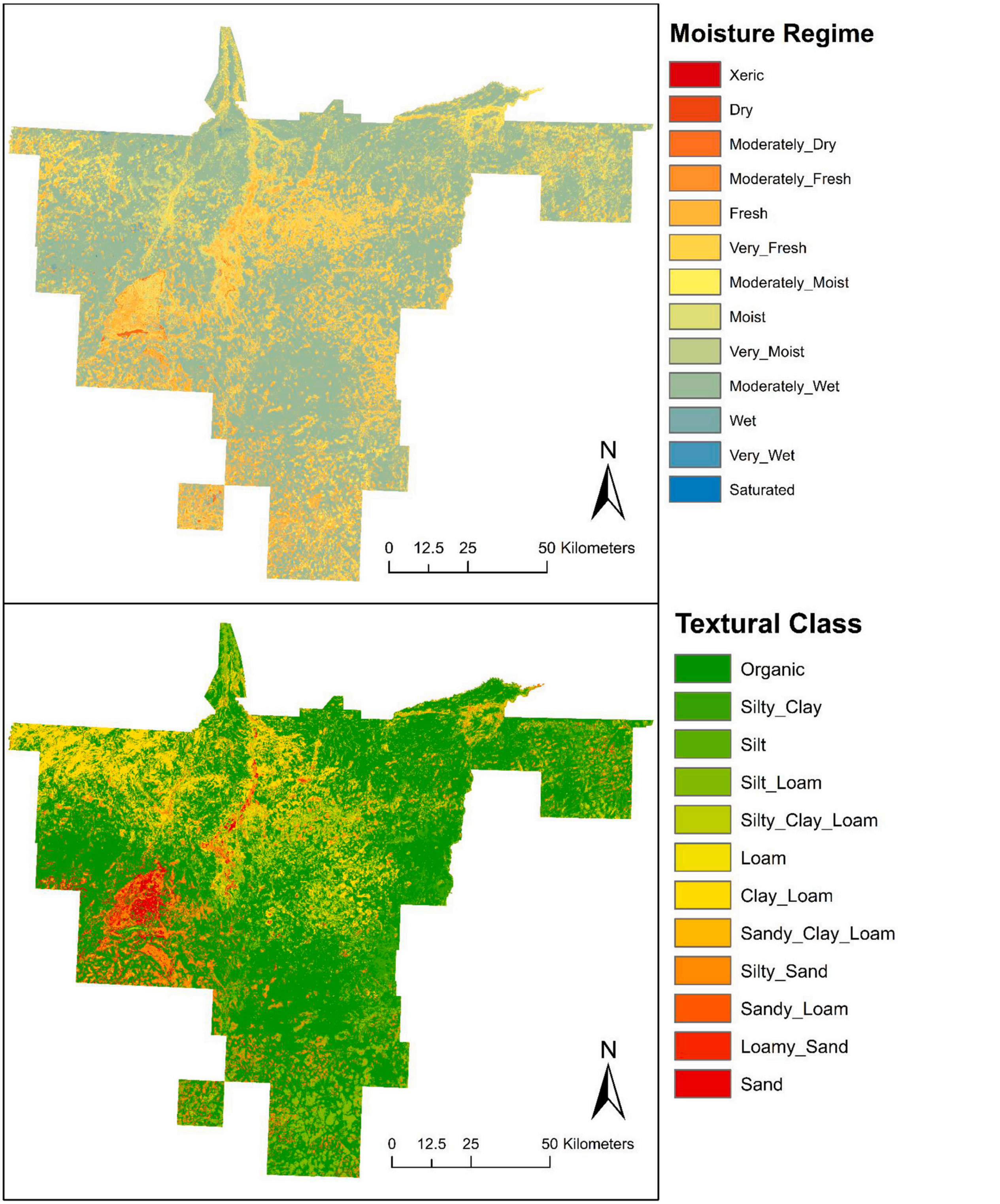


Fig. 8. Digital Soil Maps for inclusion of points from areas of high uncertainty as defined by Ignorance Uncertainty (IU) metric for Moisture Regime and Textural Class.

respectively).

In the high IU treatment, the dominant "Moderately wet" moisture regime class was 10.0% less common than the control and the dominant "Organic" textural class was 19.3% less common than the control (Fig. S2). In the high EU treatment, the "Moderately wet" class was 9.49% less common, and the "Organic" textural class was 18.6% less common (Fig. S3). In the high CI treatment, the "Moderately wet" class was 8.5% less common and the "Organic" class was 16.2% less common (Fig. S4).

3.1.3. Model improvement upon additional sampling in areas of high uncertainty

Additional sampling significantly improved model accuracy and kappa (Figs. 4–5). When comparing between the round 1 model and the

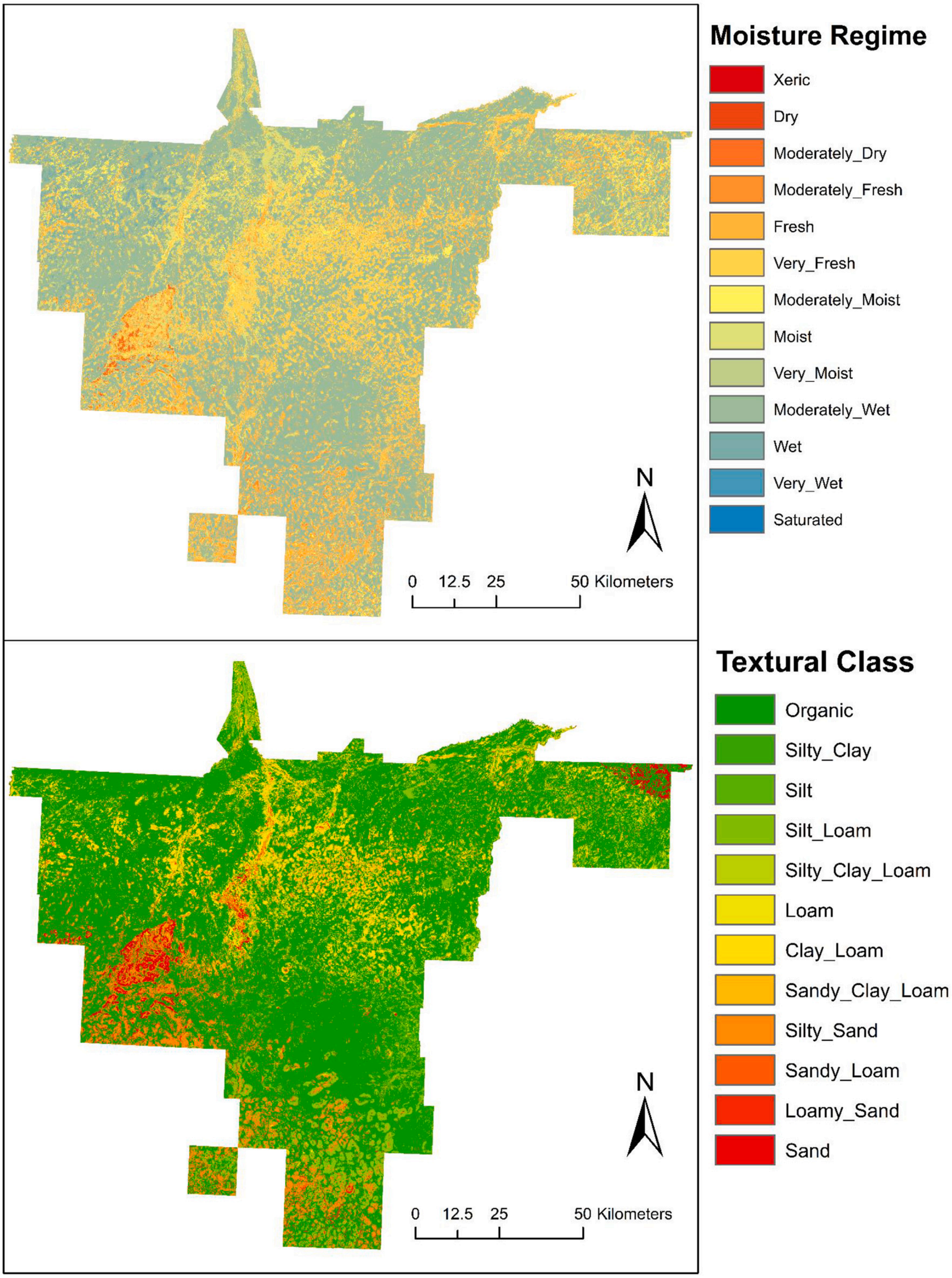


Fig. 9. Digital Soil Maps for inclusion of points from areas of high uncertainty as defined by Exaggeration Uncertainty (EU) metric for Moisture Regime and Textural Class.

round 2 control (random point addition) model, accuracy improved from 34.8% to 38.7% for moisture regime ($t_{(29)} = 19.3$, $p < 0.001$, *Cohen's* $d = 3.52$) and from 39.3% to 43.3% for texture class ($t_{(29)} = 15.8$, $p < 0.001$, *Cohen's* $d = 2.89$). The round 2 control also improved kappa from 0.18 to 0.23 for moisture regime ($t_{(29)} = 19.3$, $p < 0.001$, *Cohen's* $d = 3.52$) and from 0.21 to 0.27 for texture class ($t_{(29)} = 17.5$, $p < 0.001$, *Cohen's* $d = 3.19$).

All high uncertainty treatments improved model performance compared to the control (random point addition) (Figs. 4–5, Table 2). Wilcoxon signed-rank tests showed significant improvement when comparing treatments to the control with all p values significant at $p < 0.001$ and effect sizes all at $r > 0.50$ (Table 2).

The benchmarked models had significantly higher accuracy and kappa scores compared to the round 2 CI treatment. Average accuracy

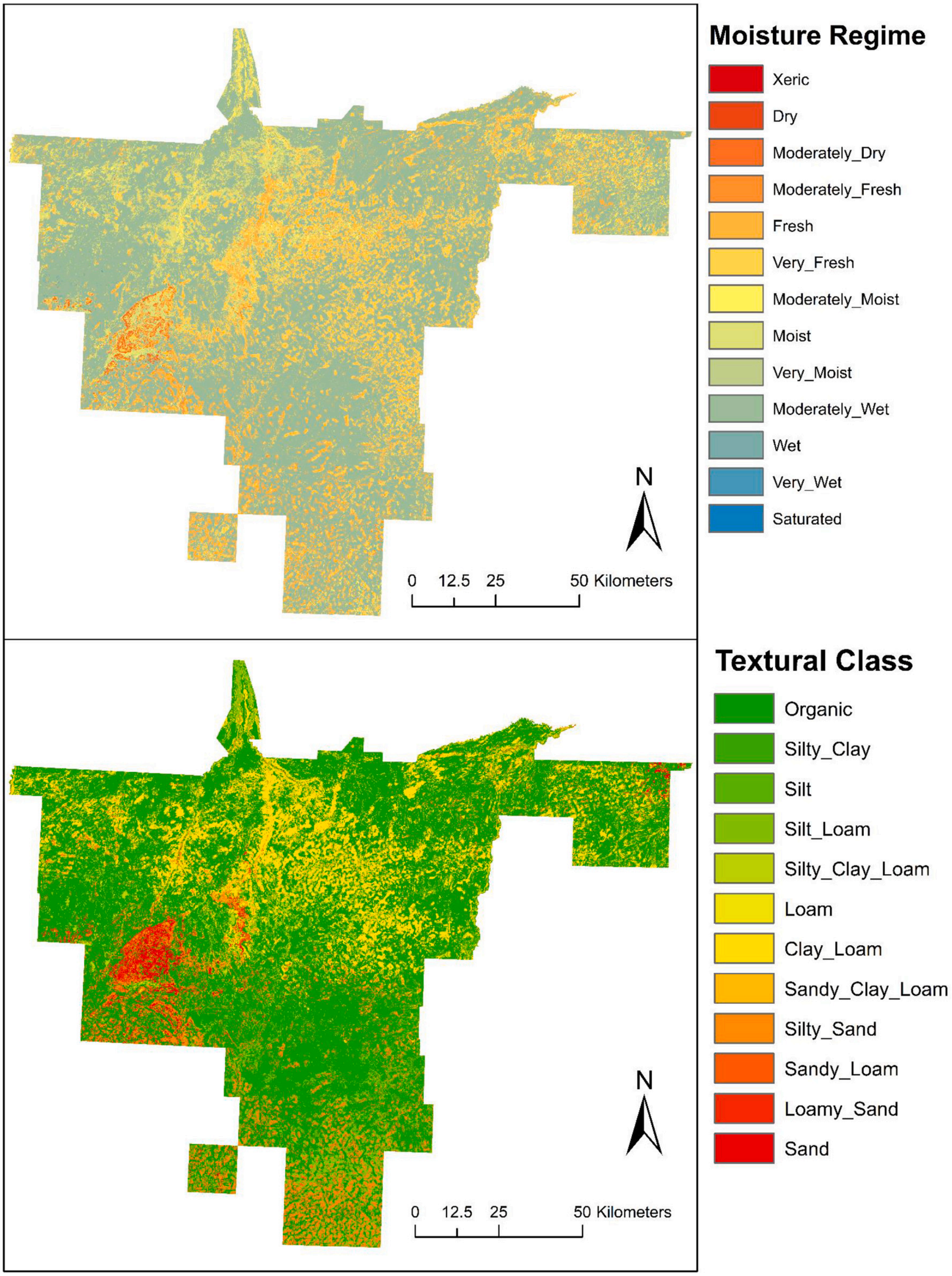


Fig. 10. Digital Soil Maps for inclusion of points from areas of high uncertainty as defined by Confidence Index (CI) metric for Moisture Regime and Textural Class.

for moisture regime was 59.8% ($p < 0.01$) and average kappa was 0.51 ($p < 0.01$). Average accuracy for textural class was 64.3% ($p < 0.01$) and average kappa was 0.56 ($p < 0.01$).

3.1.4. Soil maps

To simplify soil map analysis, we analyzed the soil maps produced using the models with the median accuracy across our 30 simulations. The soil maps generated from the round 1 models with median accuracy predicted a substantial amount of area included in the organic and moderately wet classes (Fig. 6). This aligns with our prior knowledge of soil in the Hearst Forest as well as the distribution of soil values in the dataset (Fig. 3). Drier, sandier soils were predicted in the central and central-west areas of the forest which correspond to areas of higher elevation. The moisture regime and textural class maps covary with each other since soil texture is a criterium in assigning moisture regime values (Johnson et al., 2015). For illustrative purposes, uncertainty maps were

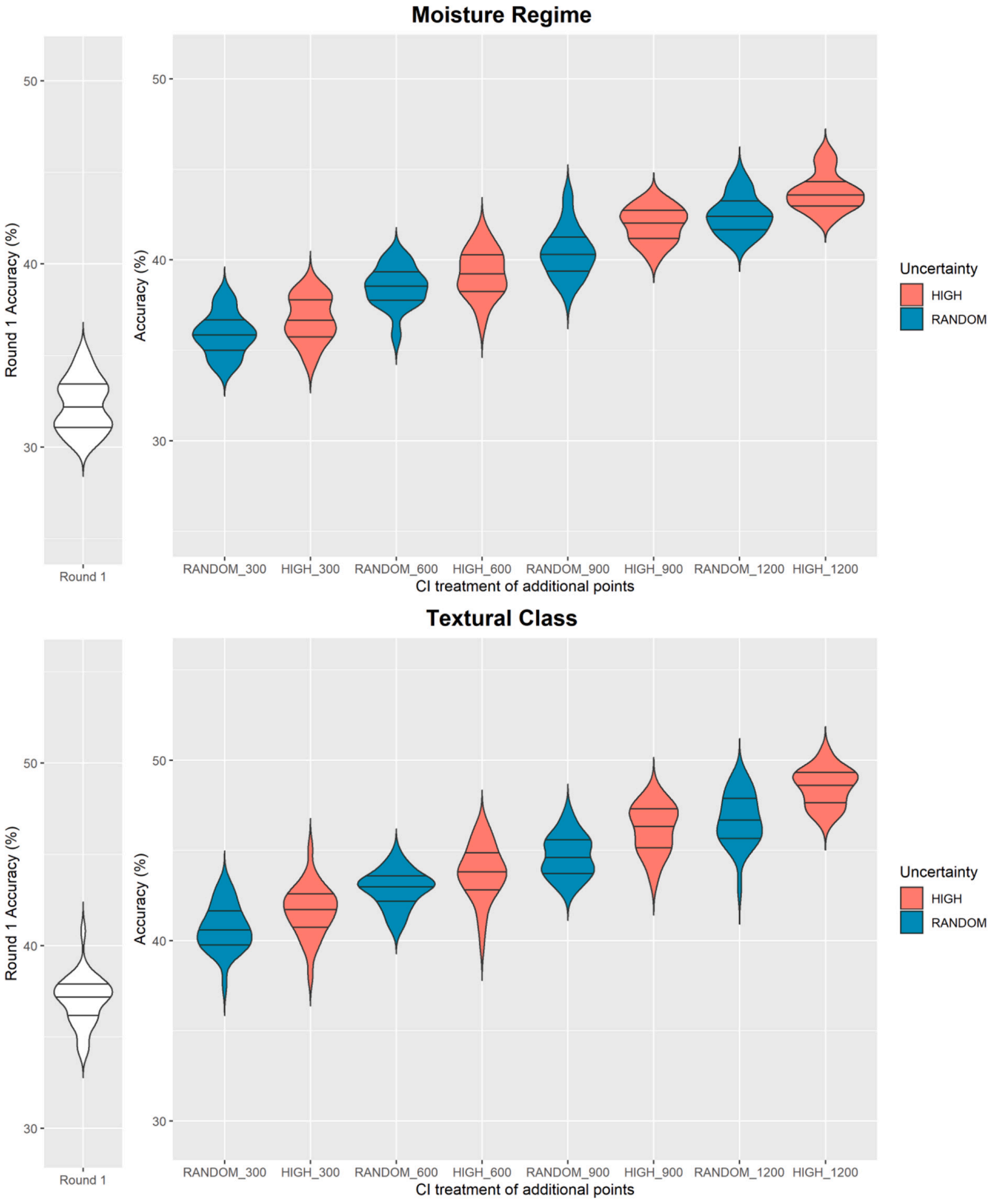


Fig. 11. Model accuracy of as a function of subsequent sample size and whether additional data came from areas of high uncertainty or were sampled randomly for (A) moisture regime and (B) textural class.

also generated from the round 1 model with the median accuracy (Figs. S5, S6). No analyses were performed on these uncertainty maps since we already quantify links between uncertainty and soil prediction in this section and in Sections 3.1.2, 3.1.3, but they may serve as useful visual aids for readers.

There was moderate similarity between the soil map generated from the round 1 model and the soil map generated from the round 2 model supplemented with data from random locations (Fig. 7). In the updated map, soil classes from minority classes were more common (Tables S1–S2). For example, sandy textural classes were more common in the central-west portion of the map (Figs. 6–7) and the "very moist" moisture regime class was overall more common in the round 2 map compared to the round 1 map (Tables S1–S2). The updated map predicted an increase in the majority class for textural class (Organic) but a decrease in the majority class (Moderately Wet) for moisture regime. The updated model predicted organic material occurring in 4.1% (618km^2) more of the study area than in the round 1 model (Table S1). The updated model predicted the Moderately Wet class occurring in

Table 3
Output of multiple linear regression for each soil variable and model performance statistic.

Variable	Model stat		Estimate	Std. Error	t stat	*p* value	R^2
Mositure Regime	Accuracy	(Intercept)	34.4582	0.26231	131.365	1.16E-222	0.83
		UncertaintyRANDOM	−0.56535	0.37096	−1.524	0.12884278	
		Data_pts_added	0.00794	0.00032	24.874	6.81E-68	
		UncertaintyRANDOM:Data_pts_added	−0.00069	0.00045	−1.5272	0.1280567	
Mositure Regime	Kappa	(Intercept)	0.17063	0.00342	49.8396	2.88E-127	0.84
		UncertaintyRANDOM	−0.00355	0.00484	−0.7333	0.46408059	
		Data_pts_added	0.00011	4.17E-06	25.9008	6.23E-71	
		UncertaintyRANDOM:Data_pts_added	−1.22E-05	5.89E-06	−2.0717	0.03937871	
Textural Class	Accuracy	(Intercept)	39.2849	0.28588	137.416	3.22E-227	0.79
		UncertaintyRANDOM	−0.50998	0.4043	−1.2614	0.20841428	
		Data_pts_added	0.00766	0.00035	22.0194	3.78E-59	
		UncertaintyRANDOM:Data_pts_added	−0.00103	0.00049	−2.0977	0.03699519	
Textural Class	Kappa	(Intercept)	0.19967	0.00424	47.1342	4.56E-122	0.80
		UncertaintyRANDOM	0.00246	0.00599	0.41068	0.68167858	
		Data_pts_added	0.00012	5.16E-06	23.6909	2.54E-64	
		UncertaintyRANDOM:Data_pts_added	−2.58E-05	7.29E-06	−3.5424	0.00047774	

1.8% (177km^2) less of the study area than in the round 1 model (Table S2).

The soil maps produced from the round 2 treatment models all agreed with our understanding of soil variation in the region. However, there were a few key differences when comparing between the maps generated from round 2 models with random point addition in comparison to the high uncertainty point addition treatments. For textural class, on average, Organic was predicted in 9.4% (1426 km^2) less of the study area and this was compensated largely through increased prediction of clay loam (+3.4%) and silty sand (+3.5%) (Table S1). For moisture regime, the round 2 models diverged, with the Moderately Wet class being predicted more often for the IU and CI metric (+0.7% and 2.7% respectively) but less often for the EU metric (−3.0%) (Table S2). In terms of the spatial distribution, increases in silty sand can be seen in the central-west and southern portion of the study extent (Figs. 8–10). Increases in clay loam for the IU and CI treatment can be seen throughout the map (Fig. 8, 10). The map produced from the IU treatment (Fig. 8) reclassified some moisture regime values in the northeast and central-south Hearst from moderately fresh to moderately moist. The map produced from the EU treatment (Fig. 9) reclassified some textural class values in the northeast from organic to sand. The map produced from the CI treatment predicted more sand in the northeast and central west (Fig. 10) and reclassified patches of soil in the central portion of the Hearst to drier moisture regime classes.

3.2. Study 2: How does sampling effort modulate the effect of incorporating uncertainty into sampling?

3.2.1. Effects of uncertainty and sampling effort on model improvement

As sampling effort increased, model performance improved (Fig. 11 and 12). Table 3 shows the output of the multiple linear regression for each soil variable and model performance stat (i.e. accuracy and kappa). Linear regression closely fit the data with all R^2 values ≥0.78 (Table 3). The number of soil points in the subsequent dataset (i.e., sampling effort) had a significant effect on model accuracy and kappa for both moisture regime and textural class ($p < 0.0001$; Fig. 11 and 12, Table 3). Across multiple sampling efforts, the high uncertainty treatments always failed to reach statistical significance at the $p < 0.05$ level. However, in some cases, the sample size × uncertainty interaction showed significance, indicating that as sampling effort increases, the benefits of adding data from areas of high uncertainty also increases. (See Fig. 11 and 12, Table 3)

4. Discussion

4.1. Study 1: Does additional sampling in areas of high uncertainty improve model performance?

4.1.1. Comparing high uncertainty locations across metrics

There was a fair bit of overlap between high uncertainty locations (where soil observations existed) across the uncertainty metrics. EU and CI had a large overlap of high uncertainty points, likely since EU and CI are calculated in a similar way, with CI defined as EI subtracted by the class with the second highest votes. IU and EU also shared a large amount of high uncertainty points. This is harder to explain but could happen if the predicted class had a "lead" in votes, but a low proportion of the overall vote and the rest of the votes were evenly split among the remaining class. In this scenario, IU and EU would represent this point as highly uncertain due to vote-splitting whereas CI would predict it as more certain since it only considers the vote difference between the predicted class and the "runner-up". Follow-up studies could consider how uncertainty metrics relate to each other both theoretically and in the field, since these association were quite noticeable in our study.

As expected, across all uncertainty metrics, locations where the dominant "moderately wet" and "organic" classes existed were rarely identified as high uncertainty. This pattern was more apparent for textural class than moisture regime likely because the textural class soil data was even more unbalanced than moisture regime (Figs. S2–S4). Importantly, these uncertainty scores were generated using the environmental covariates; they did not use the observed soil data in the validation dataset to calculate uncertainty. In other words, the model was not "deciding" these areas were low uncertainty because the soil classes present were common in the dataset, rather, the environmental feature space (along with the training soil observation) was used to determine these locations were low uncertainty. Classifying these locations as low uncertainty likely occurred because the environmental feature space for the dominant soil class was better represented in the soil-environmental dataset leading to less uncertainty when predicting in areas where moderately wet/organic soil exists. Further, this suggests that the environmental covariates used in this study map onto real soil variation since there was lower uncertainty for covariate values corresponding to the dominant class. This link between uncertainty and environmental representation could be investigated in future studies by comparing the environmental feature space captured in the initial soil dataset (used to build the round 1 model) with the environmental covariate values in the soil datapoints subsequently added. Dissimilarity indices (e.g., Bray and Curtis, 1957; Gower, 1971; Meyer and Pebesma, 2021) could be used to quantify, for each soil class, how well represented their corresponding environmental data is present in the initial soil dataset.

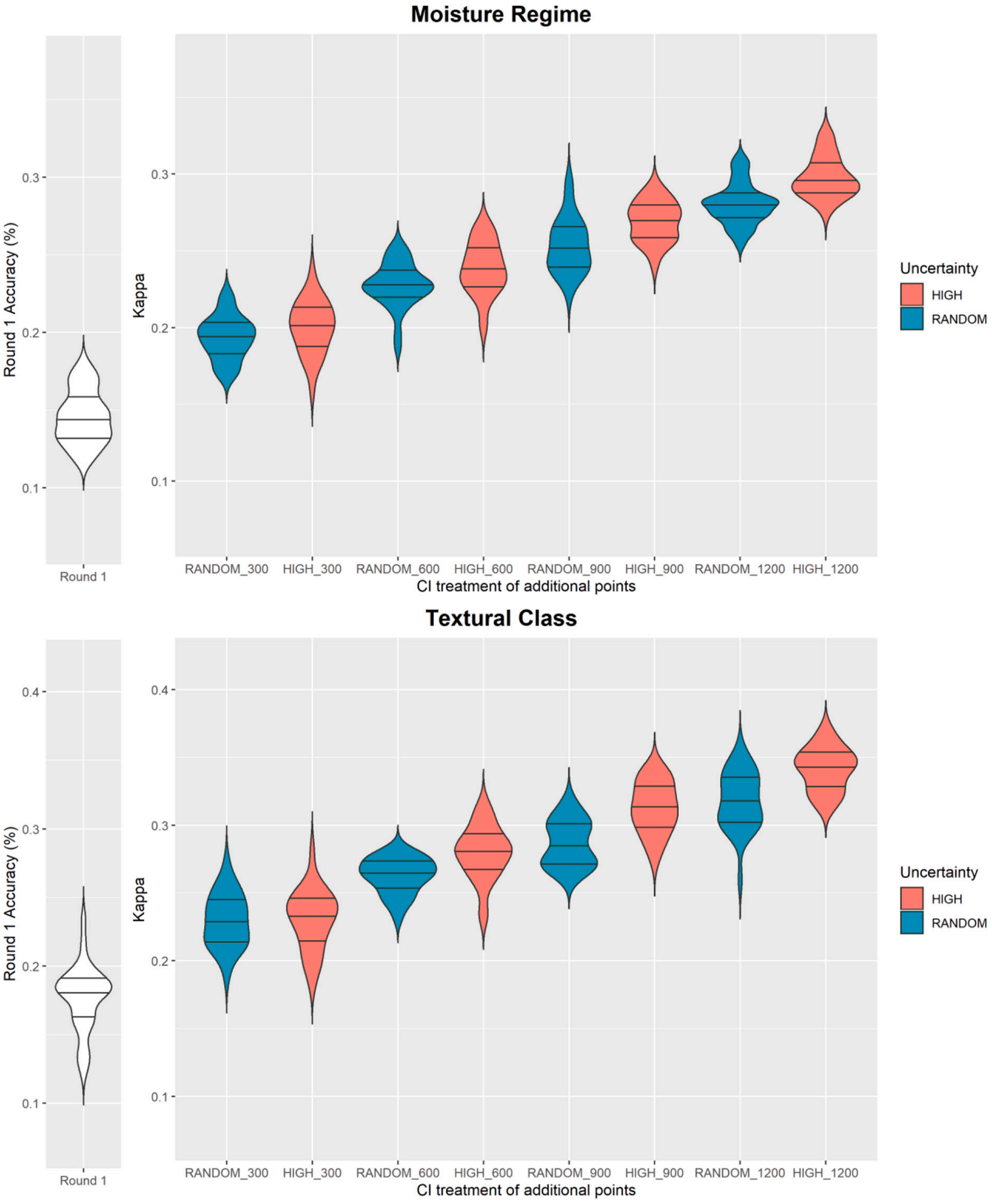


Fig. 12. Model kappa of as a function of subsequent sample size and whether additional data came from areas of high uncertainty or were sampled randomly for (A) moisture regime and (B) textural class.

Somewhat surprisingly, only the dominant class is underrepresented in the additional sample for the round 2 models (Figs. S2–S4). We expected more of a tapering curve for the high uncertainty treatments along Figs. S2–S4 but instead many of the other common classes were as well represented or even more represented in the high uncertainty treatment compared to the control. This likely occurred because the vote distributions for these classes were split between the dominant class and the correct class. This can be seen in the confusion matrices (Figs. S5–S6) where clay loam, silt loam, and silty sand were often misclassified as organic. This was also true for moisture regime where very fresh, fresh, and moist were often misclassified as moderately wet.

4.1.2. *Model/map improvement and choice of uncertainty metric*

All uncertainty metric treatments showed model improvement over the control of random point addition. We did not perform statistical tests comparing our treatments since the raw increases in accuracy and kappa values were minimal and no striking differences were apparent. The round 2 treatment maps for textural class showed less of the dominant class (organic) which was expected since this class was less common in the high uncertainty points. However, the round 2 treatments maps for moisture regime diverged in their dominant class (moderately wet) predictions with IU and CI predicting increases but EU predicting decreases. This may be because EU had the lowest proportion of high uncertainty points as moderately wet compared to IU and CI (Figs. 7–9).

4.2. *Study 2: How does sampling effort modulate the effect of incorporating uncertainty into sampling?*

It was unclear a priori what the shape of the trend would be between sampling effort and model performance. In our study, we observed a strong linear trend although this trend may not be consistent across soil datasets or sampling efforts. For example, the initial round 1 model was built using 500 soil observations so increases of 300/600/900/1200 are quite substantial compared to the initial dataset size. If the original model was built using more data or if additional sampling was smaller relative to the original sample size, it is unclear if the model improvement trend would still be linear. The effects of building round 1 models on 500 data points can also be seen in model accuracy/kappa scores which are lower than ideal. Nevertheless, in our study, the linear trend was strong, and we did not observe any diminishing returns as sampling effort increased.

The interaction between sampling effort and uncertainty treatment is important to note as it suggests the benefits of using high uncertainty maps to guide future sampling is more noticeable with intensive additional sampling. This fortuitous since uncertainty analyses take time to complete so it may make sense to perform them if there is opportunity to do considerable additional sampling.

4.2.1. *Incorporating uncertainty maps into additional sampling for DSM*

Although uncertainty-guided sampling showed statistical improvement over random sampling, from a practical perspective, increases of 1–2 percentage/kappa points are perhaps underwhelming. Note that the lack of performance improvement was not due to issues with the underlying soil dataset or modelling procedure the models being unable to improve with more data—the benchmarked models showed that adding the entirety of Dataset 1 + 2 led to much higher accuracy and kappa. Importantly, this paper compared the performance of multiple uncertainty metrics for uncertainty-guided sampling in the absence of other soil sampling restrictions such as spatial or environmental coverage. Given that we still observed model improvement using only uncertainty-guided sampling, it is likely model improvement will increase further if uncertainty-guided sampling is combined with other sampling considerations (Minasny and McBratney, 2006; Wadoux et al., 2019). For example, this uncertainty approach could be combined with conditioned Latin hypercube sampling (cLHS; Minasny and McBratney, 2006) – an approach that aims to capture environmental (and sometimes spatial) variability of the study area across the soil samples. Previous studies have combined cLHS with maps of terrain connectivity to optimize soil collection given that some areas are inaccessible (Clifford et al., 2014; Stumpf et al., 2016). Uncertainty maps could be incorporated into cLHS sampling by constraining the cLHS to only operate in areas above a certain uncertainty threshold value (Minasny and McBratney, 2006).

For our study area and soil data, much of the uncertainty patterns seemed to be driven by the dominant soil class. In our case, reducing uncertainty was achieved through sampling in areas containing less of the dominant class. This uncertainty guided approach improved overall model performance which is a common goal of DSM projects. However, in some cases, practitioners may be more interested in discriminating between specific soil classes as opposed to overall model performance which reflects the models ability to discriminate between all soil classes. In this case, the soil dataset could be modified to only include the soil classes of interest, or the uncertainty equation could be modified to reflect vote distribution between the specific classes of interest.

In this paper, we calculated uncertainty to guide independent sampling for moisture regime and textural class. Often, practitioners will measure multiple soil variables at a single sampling site. In this case, uncertainty maps for each variable could be averaged across soil properties to guide future sampling in areas of high uncertainty for multiple soil properties (Vašát et al., 2010; Szatmári et al., 2019).

5. Conclusions

Digital soil mapping success depends on an understanding of the landscape and an ability to leverage good methodology to generate a soil map. An understanding of the landscape is needed to know what soils data are available and how to appropriately select environmental covariates for modelling. Good methodological practices include site selection for additional soil sampling, optimal sampling effort, and appropriate machine learning parameterization. While uncertainty maps are often used to gain an understanding of the landscape by providing spatially explicit estimates of soil map accuracy, we show in this paper they can also be used as methodological tools to improve future DSM performance.

In this study, we were interested if uncertainty maps could be used to inform future soil sampling by performing additional sampling in high uncertainty as a strategy to increase model performance. By simulating a repeated soil sampling campaign, we demonstrate, as a proof of concept, that uncertainty analysis of a digital soil map can be used to guide future sampling since the updated soil maps showed a modest but statistically significant improvement in model accuracy and kappa scores. All uncertainty metrics tested in this study mapped onto real soil variation with highly uncertain areas unlikely to contain common soil classes in the soil pedon dataset. Furthermore, the benefits of using uncertainty maps to guide future sampling was consistent across multiple measures of uncertainty, was apparent for multiple degrees additional sampling effort and may be more significant as sampling effort increases. The approach outlined in this paper can incorporate legacy data into the initial DSM, incorporate initial soil sampling to guide future sampling in subsequent years (i.e., between field seasons), and can be performed multiple times for a single study area.

Declaration of Competing Interest

None.

Acknowledgments

We thank Hearst Forest Management Inc. for supplying the LiDAR DEM data and the Ontario Ministry of Natural Resources and Forestry (OMNRF) for use of the Growth and Yielf, Forest Resource Inventory, Ecological Land Classification, and Advanced Forest Resource Inventory Technologies soil data. Funding for this project was provided to KLW by Forestry Futures Trust and through Canadian Forest Service base funding through Sustainable Fibre Solutions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geodrs.2022.e00495.

References

Arrouays, D., Mcbratney, A., Minasny, B., Hempel, J., Heuvelink, G., Macmillan, R.A., Hartemink, A., Lagacherie, P., McKenzie, N., 2014a. The GlobalSoilMap project specifications. In: Glob. Basis Glob. Spat. Soil Inf. Syst. - Proc. 1st Glob. Conf. 9–12.

Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., McBratney, A.B., 2014b. GlobalSoilMap: Basis of the Global Spatial Soil Information System, 1st ed. CRC Press. https://doi.org/10.1201/b16500.

Biswas, A., Zhang, Y., 2018. Sampling designs for validating digital soil maps: a review. Pedosphere 28, 1–15. https://doi.org/10.1016/S1002-0160(18)60001-3.

Bivand, R., Rundel, C., 2019. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-2. https://CRAN.R-project.org/package=rgeos.

Bivand, R., Keitt, R., Rowlingson, B., 2019. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-8. https://CRAN.R-project.org/package=rgdal.

Blackburn, C.E., Bond, W.D., Breaks, F.W., Davies, D.W., Edwards, G.R., Poulsen, K.H., Trowell, N.F., Wood, J., 1985. Evolution of Archean volcanic-sedimentary sequences of the western Wabigoon subprovince and its margin: A review. In: Ayres, L.D., Thurson, P.C., Card, K.D., Weber, W. (Eds.), Evolution of Archean Supracrustal Sequences. Geological Association of Canada (Special Paper 28).

Blackford, C., Heung, B., Baldwin, K., Fleming, R.L., Hazlett, P.W., Morris, D.M., Uhlig, P., Webster, K., 2021. Digital Soil Mapping workflow for forest resource applications: a case study in the Hearst Forest. Ontario Can J For Res 51 (1), 59–77. https://doi.org/10.1139/cjfr-2020-0066.

Blackford, C., 2022. Spatial-Uncertainty-For-DSM-Sampling_code (version 1.0.0). [Accessed 15 March 2022] doi:10.5281/zenodo.6366470.

Bray, J.R., Curtis, J.T., 1957. An ordination of upland forest communities of southern Wisconsin. Ecol Monogr 27, 325–349. https://doi.org/10.2307/1942268.

Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138, 86–95. https://doi.org/10.1016/j.geoderma.2006.10.016.

Bui, E.N., Moran, C.J., 2003. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray–Darling basin of Australia. Geoderma 111, 21–44. https://doi.org/10.1016/S0016-7061(02)00238-0.

Burrough, P.A., van Ganns, P.F.M., Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77, 115–135. https://doi.org/10.1016/S0016-7061(97)00018-9.

Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C. W., Odgers, N.P., 2016. POLARIS: a 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025.

Clifford, D., Payne, J.E., Pringle, M.J., Searle, R., Butler, N., 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. Comput Geosci 64, 62–68. https://doi.org/10.1016/j.cageo.2014.03.005.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analysis (SAGA) v.2.1.4. Geosci Model Dev 8. https://doi.org/10.5194/gmd-8-1991-2015.

Corporation, Microsoft, Weston, S., 2019. doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.15. https://CRAN.R-project.org/package=doParallel.

Crins, W.J., Gray, P.A., Uhlig, P.W.C., Wester, M.C., 2009. The Ecosystems of Ontario, Part 1: Ecozones and Ecoregions. Marie, ON, Ontario Ministry of Natural Resources, Sault Ste.

Dyke, A.S., 2004. An outline of North American deglaciation with emphasis on central and northern Canada. In: Ehlers, J., Gibbard, P.L. (Eds.), Developments in Quaternary Sciences, Volume 2, Part B. Elsevier, pp. 373–424. https://doi.org/10.1016/S1571-0866(04)80209-4.

Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resource Res. 39 (12), 1347–1359. https://doi.org/10.1029/2002WR001426.

Goodchild, M.F., Chin-Chang, L., Leung, Y., 1994. Visualizing fuzzy maps. In: Hearnshaw, H., M., Unwin, D.J. (Eds.), Visualization in Geographical Information Systems. John Wiley & Sons, NY, pp. 158–167.

Gower, J.C., 1971. 1971 a general coefficient of similarity and some of its properties. Biometrics 27 (4), 857–871. https://doi.org/10.2307/2528823.

Grinand, C., Arrouats, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143, 180–190.

Hearst Forest Management, 2019. Hearst Forest Management Inc., Hearst Ontario. http://www.hearstforest.com/english/surficial.html (accessed 13 January 2020).

Heung, B., Hodúl, M., Schmidt, M.G., 2017. Comparing the use of legacy soil pits and soil survey polygons as training data for mapping soil classes. Geoderma 290, 51–68. https://doi.org/10.1016/j.geoderma.2016.12.001.

Hijmans, R., 2019. raster: Geographic Data Analysis and Modeling. R package version 3.0-7. https://CRAN.R-project.org/package=raster.

Huang, J., McBratney, A.B., Minasny, B., Malone, B., 2020. Evaluating an adaptive sampling algorithm to assist soil survey in New South Wales. Australia Geoderma Reg 21, e00284. https://doi.org/10.1016/j.geodrs.2020.e00284.

Johnson, J.A., Uhlig, P., Wester, M., 2015. Field Guide to the Substrates of Ontario. Marie, Ontario, Ontario Ministry of Natural Resources, Sault Ste.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Yuan, Candan, C., Hunt, T., 2019. caret: classification and regression training. R package version 6.0-84. https://CRAN.R-project.org/package=caret.

Leung, Y., Goodchild, M.F., Lin, C.C., 1993. Visualization of fuzzy scenes and probability fields. In: Newton, H.J. (Ed.), Computing Science and Statistics, Volume 24: Graphics and Visualization. (Proceedings of the 24th Symposium on the Interface). Interface Foundation of N America, Fairfax Station, VA, pp. 416–422.

Li, S., MacMillan, R.A., Lobb, D.A., McConkey, B.G., Moulin, A., Fraser, W.R., 2011. Lidar DEM error analyses and topographic depression identification in a hummocky landscape in the prairie region of Canada. Geomorphology 129, 263–275. https://doi.org/10.1016/j.geomorph.2011.02.020.

Mackasey, W.O., Blackburn, C.E., Trowell, N.F., 1974. A regional approach to the Wabigoon–Quetico belts and its bearing on exploration in Northwestern Ontario. Ontario Division of Mines, Ministry of Natural Resources, ON.

Malone, B.P., de Gruijter, J.J., McBratney, A.B., Minasny, B., Brus, D., 2011. Using additional criteria for measuring the quality of predictions and their uncertainties in a digital soil mapping framework. Soil Sci Soc Am J 75, 1032–1043. https://doi.org/10.2136/sssaj2010.0280.

Marchant, B.P., Lark, R.M., 2006. Adaptive sampling and reconnaissance surveys for geostatistical mapping of the soil. Eur J Soil Sci 57, 831–845. https://doi.org/10.1111/j.1365-2389.2005.00774.x.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol Evol 12, 1620–1633.

Minasny, B., Bishop, T.F.A., 2008. Analyzing uncertainty. Chapter 24. In: McKenzie, N., Grundy, M., Webster, R., Ringrose-Voase, A. (Eds.), Guidelines for Surveying Soil and Land Resources. CSIRO Publishing, Melbourne, pp. 383–393.

Minasny, B., McBratney, A.B., 2002. Uncertainty analysis for pedotransfer functions. Eur J Soil Sci 53, 417–429.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput Geosci 32, 1378–1388.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: a brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017.

Musafer, G.N., Thompson, M.H., 2016. Optimal adaptive sequential spatial sampling of soil using pair-copulas. Geoderma 271, 124–133. https://doi.org/10.1016/j.geoderma.2016.02.018.

Odgers, N.P., McBratney, A.B., Minasny, B., Sun, W., Clifford, D., 2014. DSMART: An algorithm to spatially disaggregate soil map units. In: Arrouays, D., McKenzie, N., Hempel, J., de Forges, A., McBratney, A.B. (Eds.), GlobalSoilMap: Basis of the Global Spatial Soil Information System. CRC Press, pp. 261–266.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain J. Sci. 5, 23–27.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog Phys Geogr 27, 171–197. https://doi.org/10.1191/0309133303pp366ra.

Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt-Stitt, S., Buzzo, G., Dumperth, C., Wadoux, A., Xiang, W., Scholten, T., 2016. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. J Plant Nutr Soil Sci 179, 499–509. https://doi.org/10.1002/jpln.201500313.

Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, C., Wadoux, A., Xiang, W., Scholten, T., 2017. Uncertainty-guided sampling to improve digital soil maps. Catena 153, 30–38. https://doi.org/10.1016/j.catena.2017.01.033.

Szatmári, G., László, P., Takács, K., Szabó, J., Bakacsi, Z., Koós, S., Pásztor, L., 2019. Optimization of second-phase sampling for multivariate soil mapping purposes: case study from a wine region, Hungary. Geoderma 352, 373–384. https://doi.org/10.1016/j.geoderma.2018.02.030.

Thurston, P.C., Osmani, I.A., Stone, D., 1991. Northwestern Superior Province: review and terrane analysis. In: Thurston, P.C., Williams, H.R., Sutcliffe, R.H., Stott, G.M. (Eds.), Geology of Ontario. Ontario Geological Survey, Special Volume 4, Part 1, pp. 81–144.

Vašát, R., Heuvelink, G.B.M., Borůvka, L., 2010. Sampling design optimization for multivariate sampling. Geoderma 155, 147–153. https://doi.org/10.1016/j.geoderma.2009.07.005.

Wadoux, A.M.J.-C., Brus, D.J., Heuvelink, G.B.M., 2019. Sampling design optimization for soil mapping with random forest. Geoderma 355, 113913. https://doi.org/10.1016/j.geoderma.2019.113913.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J Open Source Softw 4 (43), 1686. https://doi.org/10.21105/joss.01686.

Yang, L., Qi, F., Zhu, A.-X., Shi, J., An, Y., 2016. Evaluation of integrative hierarchical stepwise sampling for Digital Soil Mapping. Soil Sci Soc Am J 80, 637–651. https://doi.org/10.2136/sssaj2015.08.0285.

Zhu, A.X., 1997. Measuring uncertainty in class assignment for natural resource maps using fuzzy logic. Photogramm Eng Remote Sens 63, 1195–1202.