Spatial modelling of available aluminum in glacial till: informing a boreal forest soil carbon predictor

Scout M. Quinn[1,2], Benjamin Misiuk [1,2], Mackenzie E. Patrick[1,†], Susan E. Ziegler[1]

[1]Department of Earth Sciences, Memorial University of Newfoundland and Labrador, St John's, A1C 5S7, Canada

[2]Department of Geography, Memorial University of Newfoundland and Labrador, St John's, A1C 5S7, Canada

[†] Deceased

Correspondence to: smquinn@mun.ca

Abstract

Boreal forest soils store 30% of global forest soil carbon, making them a crucial component of the carbon cycle. However, climate change at high latitudes is increasing temperatures and producing unpredictable precipitation patterns. Soil organic carbon (SOC) formation and stabilization is tied to precipitation patterns, thus climate change will inevitably influence the stability and longevity of boreal forest SOC. The current size and distribution of this reservoir is poorly understood, creating uncertainty under current and future climate scenarios. Previous research demonstrates mineral soil properties may be used to model boreal forest SOC accurately. The surface slope, depth of carbon enriched horizon, and climate characteristics are important parameters for modelling SOC and can generally be obtained or estimated via remote sensing. However, information about the availability of reactive aluminum – non-crystalline forms capable of interacting with organic matter to form stable carbon rich organometal complexes in mineral soils – is not widely available but is controlled by the weatherable aluminum in soil parent material. To bridge this gap, the Newfoundland and Labrador till geochemistry dataset was used here to map and model weatherable aluminum availability in glacial till across the island of Newfoundland as a function of geology and climate. The Random Forest Algorithm was employed to develop two models: one relying solely on the strength of geological and climatic variables, and the other drawing additionally on the spatial context of sample points. The first model performed well (R^2=0.60), however, adding a spatial component increased the performance of the second model (R^2=0.71). Bedrock type and proximity of the samples to certain units were indicated to be the strongest controls on aluminum availability, while environmental factors were less influential. Additionally, model uncertainty was calibrated empirically and mapped spatially, providing reliable and actionable information about the confidence of predictions over the study area. This project demonstrates the value of predictive geospatial modelling for till geochemistry mapping and delivers key aluminum availability predictions useful for future efforts in deriving SOC reservoir estimates across Newfoundland.

1 Introduction

Soils are a crucial component of the global carbon cycle, with soil carbon (C) stocks on average three times larger than the atmospheric C reservoir (Scharlemann et al., 2014). Of this, boreal forest soils store 30% of global forest soil C alone (Scharlemann et al., 2014). This is primarily stored in podzols – an acidic mineral soil characterised by a nutrient- and organic-depleted eluvial (Ae) horizon overlying the enriched illuvial (B) horizon (Hagemann et al., 2010). Soil organic carbon (SOC) is sequestered when dissolved organic matter leached from the overlaying organic layer infiltrates into the mineral soil and interacts with reactive metals, precipitating as organometal complexes (OMCs) in an organic matter enriched soil horizon (Baldock & Skjemsted, 2000). Transport of SOC through the soil profile to the persistent deep mineral soil C pool depends on water infiltration, with SOC stabilisation influenced by climate regulated microbial composition and activity (Jackson et al., 2017; Wang et al., 2022). Soil C storage potential in boreal forests is linked to a variety of soil parameters such as texture, clay content, acidity, cation availability, and environmental factors such as climate, topography, and parent material (Oades, 1988; Slessarev et al., 2022; Patrick, 2023). However, in the young podzolic soils that dominate most boreal forests, reactive metal availability, dominated by reactive Al availability (Patrick et al., 2022), are one of the strongest predictors of SOC (Rasmussen et al., 2018).

Amplified climate change in high latitudes raises concerns about vulnerable boreal SOC stocks as the current and future size and distribution of this reservoir under various climate scenarios are poorly constrained, introducing uncertainty in models relying upon such estimates (Crowther et al., 2016; Friedlingstein et al., 2014; Wang et al., 2022). Consequently, unravelling climate-SOC feedbacks is crucial to predict the SOC stock evolution under future climate scenarios (Friedlingstein et al., 2014; Jackson et al., 2017). Additionally, evaluating this reservoir at global, regional, and local scales is key to mitigating climate change. Land use and resource extraction management can mitigate terrestrial CO_2 loss into atmospheric or aquatic reservoirs (Wang, 2020). Such practices include modifying agricultural techniques to include low tillage practices and fertilizer amendments, or reforestation and harvesting plans that encourage SOC retention (Hillier et al., 2012; Ameray et al., 2021).

Friedlingstein et al. (2014) analysed global SOC estimates from 11 participants in the 5th Climate Model Intercomparison Project and found estimates ranged from 510 to greater than 3000 Pg, with wide error margins disproportionately affecting high latitude regions. This is because SOC is modelled globally using earth system models (ESMs), which synthesize results of smaller models to provide insight into environmental context, climatic conditions, and soil properties (Pierson et al., 2022). Podzol-specific SOC storage mechanisms are often not incorporated into such models, and boreal regions are thus poorly modelled – likely a consequence of low sample density in high latitude areas (Pare et al., 2024). For example, productivity attributes often dictate SOC predictions (Zhang et al, 2020), however these are unreliable at high latitudes where organic input is scant and undergoes protracted decomposition due to low temperatures and elevated acidity (Pare et al., 2024; Zhou et al., 2024). Additionally, in some cases soil texture properties have been improperly positively related to SOC storage; the CENTURY climate model (Parton, 1993) associates higher clay fractions with increased SOC stocks, obscuring the importance of other mineralogical controls (Rasmussen et al., 2018; Zhang et al., 2020; Patrick, 2023; Pare et al., 2024). Research suggests

that a fine scale approach delivers higher accuracy and more realistic estimates over regional extents (Deluca & Boisvenue, 2012; Jackson et al., 2017; McNicol et al., 2019; Pare et al., 2024; Pierson et al., 2022). For example, constraining models to the biome scale may normalize the effects of climate and illuminate SOC mechanisms that may otherwise be concealed (Patrick, 2023).

A growing body of knowledge suggests that metal oxyhydroxides serve as a valuable SOC predictor in boreal and other humid forest soils. For example, Spinola et al. (2026) demonstrate the association between podzolic SOC distribution and Fe and Al oxyhydroxides. In soil, these forms of Fe and Al found in reactive poorly crystalline minerals (PCMs) stabilize carbon through the formation of organometal complexes (OMCs). The prevalence of these reactive metal forms in soil is a function of their content in weatherable mineral forms in the soil parent material (Slessarev et al., 2022), typically, the underlying glacial till in boreal regions. For example, higher Fe and Al in mineral soils were observed in boreal forests growing on highly weatherable ultramafic-rich till with high Al-rich feldspar content in contrast to similar forests overlaying more highly weathered, Al poor, sedimentary-based till (Patrick et al., 2022). Locally, the C saturation of organometal complexes (OMCs), the proportion of SOC in OMCs, and soil reactive Al content, (forms of Al in the soil that will directly participate in OMC formation) are the primary predictors for soil C in inaccessible, poorly sampled upland boreal forest soils on the Island of Newfoundland, further highlighting Al availability as a key predictor of SOC (Patrick, 2023). Therefore, Al availability – Al with the potential to contribute to reactive Alin soil – is a function of both Al content and mineral weatherability in soil parent material.The chemical index of alteration (CIA) which compares the content of highly stationary Al oxides to more mobile Ca, Na, and K oxides (Nesbitt & Young, 1984) is a commonly accessible approach to evaluating the degree of weathering a mineral has undergone. Combining soil parent material Al content with CIA, Al availability directly addresses a key geologic control on SOC storage (organometal complexation), as reactive metals prove to be a more reliable indicator of SOC storage capacity in podzols than clay and silt (Rasmussen et al., 2018; Patrick et al., 2022; Spinola et al., 2026).

Previous work illustrates that slope, depth of enriched horizon, and climate can be modelled to extrapolate the saturation of OMCs and proportion of SOC within humid boreal forest soil OMCs (Patrick, 2023). However, soil Al% is not so easily estimated and is currently obtained through intensive field work and laboratory analyses. Assessing the controls influencing Al OMC precipitation is key to predicting SOC in these soils (Patrick et al., 2022; Patrick, 2023). Geological maps are often used as a proxy for parent material; however, this ignores the suitability of different lithologies to acquiesce more mineralogical material than others to soil forming processes and explains little about the persistence of effects of parent material (Lawley & Smith, 2008). Unsurprisingly, surficial geology has been found to be a consistently significant control of SOC at local to regional scales (Adhikari et al., 2020). On the Island of Newfoundland, and in many boreal regions, parent material is dominantly glacial till, not bedrock, rendering it especially relevant to incorporate this dimension into SOC modelling. First mapping the till composition and using this to support a future sampling design to capture a range of Al% in soils across a variety of slopes and climatic conditions would provide a means to explore this parent material- Al OMC connection (Patrick, 2023).

The aim of this paper is to create a high-resolution geospatial model predicting the distribution of Al-rich weatherable minerals in glacial till in Newfoundland. This is a critical step toward attaining accurate regional estimates of boreal forest soil SOC stocks across the Island of Newfoundland. We model Al% spatially as a function of covariates believed to influence Al geochemistry. Because till geochemistry is controlled by geological parent material, and because Al is a relatively immobile element, we hypothesise that bedrock geology will be the strongest predictor for Al%. Two models are trialled in this study: a highly specific spatial model using both environmental covariates and geographic location information, and a more general aspatial model relying solely on environmental covariates. We hypothesise that the spatial model may be useful in areas that have been extensively sampled and the aspatial version may have a higher capacity to extrapolate predictions to areas with fewer samples. Predictive maps of Al and calibrated spatial estimates of their uncertainty are produced as objectives of this project. Calibration of predictive uncertainty is crucial for delivering accurate estimates of model confidence that can be used directly in future SOC modelling. These objectives will contribute to the establishment of a continuous spatial SOC stock estimate for the Island of Newfoundland.

2 Methods

2.1 Till dataset

Publicly available till geochemistry data was sourced from the Geological Survey of Newfoundland and Labrador (GSNL; GSNL, 2024). This dataset was developed from till surveys conducted by the GSNL between 1985 and 2023 and spans much of the island (see Figure 1). The inconsistent sample distribution can be attributed to sampling primarily being conducted according to National Topographic System map sheets. Samples were collected 40-60 cm deep on average and were large enough to obtain ~0.50 kg of till matrix, with larger clasts discarded (GSNL, 2024). Samples were dried and sieved using either a <180 µm or <63 µm sieve. Samples underwent gravimetric analysis and the remaining solid was digested in 15 ml of concentrated HF, 5 ml of concentrated HCl, and 5 ml of 50 volume percent $HClO_4$. Further addition of concentrated 10% HCl was added if dissolution was not achieved. Geochemical composition of the solution was determined by the GSNL using inductively coupled plasma emission spectroscopy (ICP-OES). Results are given in wt% for elements of interest which were Na, Ca, K, and Al.

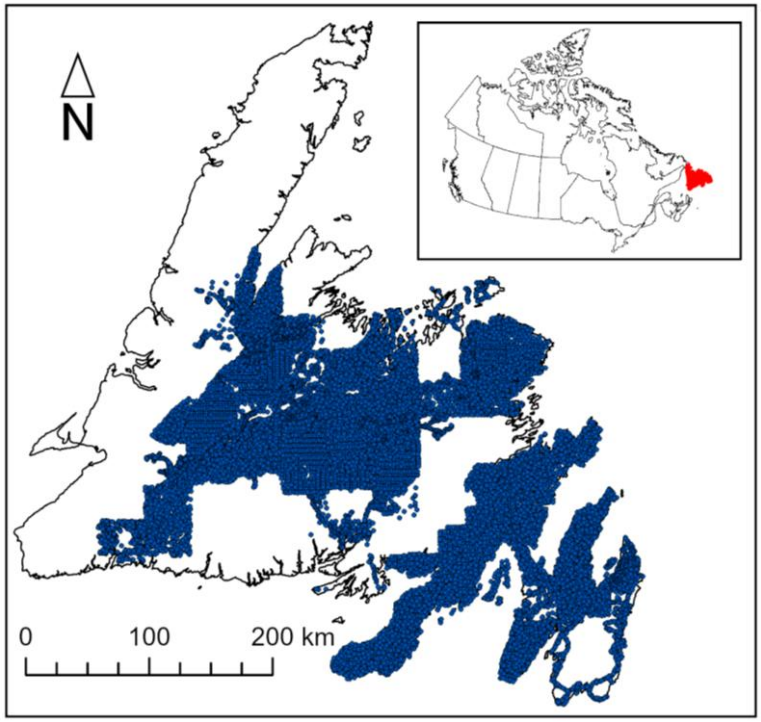
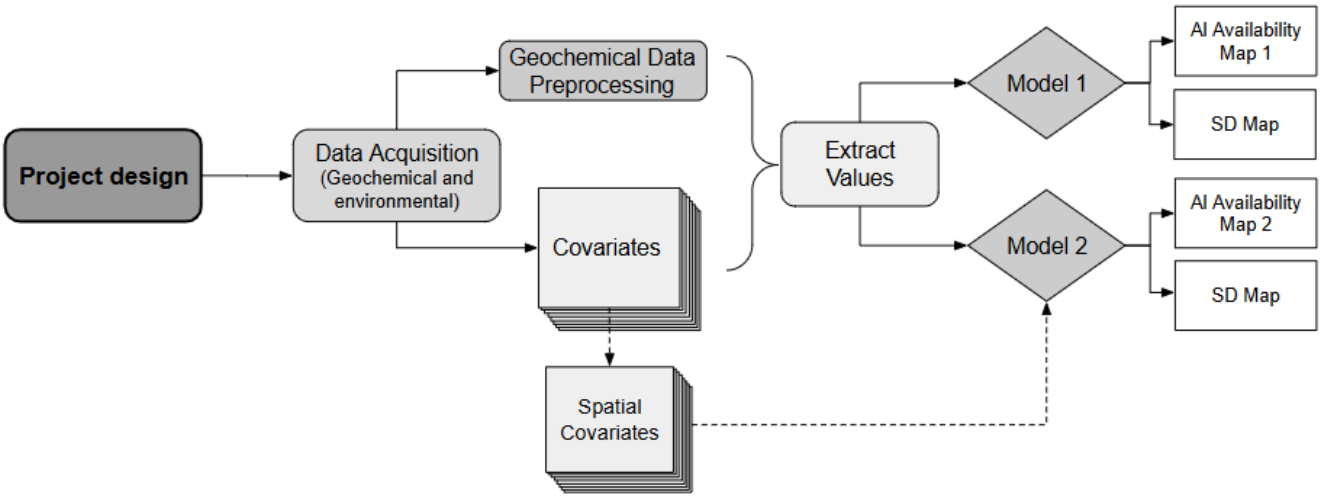


Figure 1 Distribution of till geochemistry samples from the Geological Survey of Newfoundland and Labrador across Newfoundland

2.2 Covariates

Covariates were selected to represent geological setting (surficial geology, regional geology), landscape characteristics (slope, ruggedness, elevation), and environmental context (climate region, distance from ocean). Additionally, a second model was developed that included layers representing the distance to each bedrock unit to provide spatial context. This ensures the model will consider the relationship of till geochemistry to the immediate bedrock unit, and within the greater geological setting (Figure 2).



130 **Figure 2 Project workflow for aluminum availability (Al%) models 1 and 2 and their respective standard deviations (SD)**

2.3 Preprocessing

Till samples with Al, Na, Ca, and K values below the detection limit were removed, as were those with extremely low arbitrary values, which were identified as artefacts of the ICP OES analysis. Elemental composition was reported in terms of wt% oxide. Soil organic carbon is stored in Al organometal complexes which are formed from fresh, primary weathering products rather
135 than extensively weathered forms of Al in the underlying glacial till (Slessarev et al., 2022). To assess the degree of mineral weathering in each till sample, the chemical index of alteration (CIA; Nesbitt & Young, 1984) was calculated following Eq. (1):

$$\mathrm{CIA} = \frac{\mathrm{Al_2O_3}}{\mathrm{Al_2O_3 + CaO + K_2O + NaO}} \mathrm{X}\ 100. \quad (1)$$

While concentrations of stable Al may be higher in a more weathered sample due to loss of more mobile elements, it is less likely to contribute to the formation of soil Al oxyhydroxides, which are depleted with extensive weathering
140 (Slessarev et al., 2022). Therefore, the available Al (Al%), or the fresh weatherable Al available to support Al oxyhydroxides in the overlying soil, was then estimated using both the Al_2O_3 content and the CIA through Eq. (2):

$$\mathrm{Al\%} = \frac{100 - \mathrm{CIA}}{100}\ \mathrm{X}\ \mathrm{Al_2O_3}\ \mathrm{X}\ 100, \quad (2)$$

where Al% is expressed as a weight percent of oxides.

A set of environmental covariate rasters were developed to model Al% spatially (Table 1). Al% is intrinsically related to weathering, which is influenced strongly by water infiltration and climate. Climate covariates dictating water availability (precipitation, snowfall, temperature, and evapotranspiration) were selected to represent this information. Water infiltration is decisively affected by topographic factors resulting in the inclusion of slope, ruggedness, and elevation. Distance to the ocean was also included; following glacial retreat, marine inundation affected coastal areas resulting in reworking and weathering of surficial deposits near the coast. This form of weathering was thought to potentially impact Al% in these areas. Distance to bedrock layers were additionally developed using Euclidean distance from the nearest bedrock unit, which is potentially relevant since glacial till samples are not derived exclusively from the bedrock unit upon which they lie, but also from pulverised rock from nearby units, with distant units affecting the composition less than the nearby units.

GSNL data layers were all converted to rasters with a resolution of 20 m in ArcGIS Pro across the full extent of the Island of Newfoundland. Categorical variables were coded numerically for modelling (bedrock geology unit 1-46; surficial geology units 1-9; climate zones 1-7). The "Near" tool in ArcGIS Pro was used to obtain the distance from the coastline to the centre of each raster cell, represented as a point layer, which was rasterized to match the resolution and extent of the other layers. Mean monthly snowfall, evapotranspiration, and air temperature were acquired from the FLDAS Noah Land Surface Model at 0.1° resolution for the period 01/01/1982 to 30/04/2024 (McNally et al., 2018). The mean annual precipitation layer was derived from the WorldClim layer obtained from the ArcGIS Living Atlas at 3 km resolution (Fick & Hijmans, 2017). The data were masked to the study area, converted to vector format, and joined to one single 20 m grid-like polygon layer with an extent matching the raster layers. Values were interpolated in areas missing data using nearest neighbours within 10 km. This was done in preparation for use of the "Multivariate Clustering" tool in ArcGIS Pro, which was then used to identify seven climate zones using the mean annual precipitation and mean monthly snowfall, evapotranspiration, and air temperature values from the grid-like layer. The map of the climate zones was then converted to a raster. The NLDEM (GovNL, 2023) was used for elevation values. Although the NLDEM has a higher resolution, it comprises rounded integer values making the calculation of terrain attributes from this data source unsuitable. Therefore, slope and roughness were calculated from the Canada high resolution DEM (Government of Canada, 2024) in R using the MultiscaleDTM package (Ilich et al., 2023).

Table 1 Characteristics and sources of the geospatial data layers used for model development. N/A indicates that these are not applicable to a given layer.

Layer	Original Resolution	Transformation	Source
Till Geochemistry Dataset	N/A	N/A	GSNL, 2024

Regional Surficial Geology	1:250,000	Rasterize (20 m)[a]	GSNL, 2024
Regional Bedrock Geology	1:1,000,000	Rasterize (20 m)[a]	GSNL, 2024
Distance to Bedrock Unit	1:1,000,000	Rasterize (20 m)[a] Calculate Euclidean Distance[b]	Modified from GSNL, 2024
Annual Precipitation Annual Snowfall Annual Temperature Annual Evapotranspiration	3 km 0.10 degree 0.10 degree 0.10 degree	Multivariate Clustering, Rasterize (20 m)[a]	Fick & Hijmans, 2017 McNally et al., 2018
Distance From Ocean	20 m	Calculate Euclidean Distance[b]	N/A
Elevation	5 m	Resample (20 m)[b]	GovNL, 2023
Slope Ruggedness	16 m	Resample (20 m)[b]	Derived from Government of Canada, 2024

[a]resampling method: nearest neighbor, [b]resampling method: bilinear

Additional data layers were produced representing spatial and geological context of points. The Euclidean Distance tool in ArcGIS Pro was used to create layers indicating the nearest distance to each of the GSNL Regional Bedrock Geology units. This yielded an additional 37 raster layers that were used only in Model 2 (the spatial version).

2.4 Modelling

Till geochemistry data and all rasters were loaded into RStudio using the *terra* package (Hijmans, 2025). The rasters were stacked, and the values were extracted at all till sample locations and added to the till geochemistry data frame. Incomplete observations were removed, producing a dataset of 17,328 samples. The Regional Surficial Geology variable was reclassified to "drift poor" in instances of a till sample falling on bedrock

A Random Forest model was constructed to predict the Al% observed in the till samples using the environmental geospatial covariates. The Random Forest algorithm is effective at limiting overfitting, integrating both categorical and numerical data simultaneously, and parsing complex relationships between covariates and response variables, which are desirable characteristics for modelling SOC at a regional scale (Breiman, 2001; Heung et al., 2014). Random Forest is a

machine learning approach that uses an ensemble of decision trees to make predictions. Decision trees are grown from randomly selected "bootstrap" samples of the dataset, which each comprise approximately two thirds of the total sample size as a function of sampling with replacement. Decision trees split data observations (i.e., Al%) into increasingly homogenous groups – at each split assigning a variable (*m*) best explaining the division (Breiman, 2001). This *m* is selected from a randomly generated portion (*mtry*) of all total variables (*M*). Predictions are made based on an average of the ensemble, meaning the variables which best explain the Al content are used to make a prediction for each tree, and the average of these forms the final prediction. The samples that were not drawn for each decision tree (i.e., the "out-of-bag" samples) are used to estimate the model accuracy. This built-in validation avoids splitting the dataset into "train" and "validation" partitions during model training. Additionally, 10-fold cross validation was performed to test the model predictions. Finally, Random Forest contains functionality to evaluate relationships between predictor and response variables (i.e., between Al% and environmental variables).

Two models were developed to predict values of Al% across the island of Newfoundland. The Random Forest hyper-parameters for Model 1 were $M = 7$ (total variables), $mtry = 4$ (number of variables to draw at each split), and $ntree = 1000$ (number of trees to grow). Model 2 (the spatial model) used the same covariates as Model 1 but replaced the bedrock geology variable with the 37 individual distance to bedrock geology variables enabling the model to leverage spatial trends in the geochemistry data. Hyper-parameters for Model 2 were $M = 44$, $mtry = 13$, and $ntree = 2000$. The R^2 and RMSE were calculated to evaluate each model via out-of-bag cross-validation. The resulting predictions and the predicted standard deviations δ of each ensemble model were output as raster layers to visualize spatially the distribution of Al% and model uncertainty, respectively.

2.5 Spatial model uncertainty

Following initial modelling and predictions the calibrated uncertainty of the models was derived and mapped spatially to convey confidence in Al predictions. Calibration refers to using the out-of-bag model error to scale δ to more realistic values. This is a necessary step to ensure that predicted δ reflects known model error, increasing confidence in these predictions. To accomplish calibration of the model uncertainty, the data were split into a test and training set comprising of 10% and 90% of the data respectively, with the training data used to generate calibration values and test data withheld to validate the optimized calibration values. Uncalibrated estimates of uncertainty were extracted from the out-of-bag standard deviation of the predictions of all trees of the Random Forest ensemble. To ascertain whether these estimates of uncertainty required calibration, graphical visual analysis of the training data was first performed (Palmer et al., 2022). The standard deviation (δ) of the out-of-bag predictions and the RMSE of residuals were calculated for the training data (90% of data) for both the spatial and aspatial models. The withheld test data (10% of data) was used to evaluate the calibration in both cases. . If the model prediction δ accurately represents the uncertainty of predictions, the ratio of the residuals to the δ (the r statistic) should be normally distributed with a standard deviation of 1 (Palmer et al., 2022). The r statistic was calculated and plotted as a histogram for

both models. While this emphasizes the relationship between δ and RMSE, it does not conclude whether larger RMSE values correlate with larger δ values (Palmer et al., 2022). To evaluate the linear relationship between model δ and cross-validated error, the predictions were divided into 15 equal frequency bins by δ, and the RMSE of residuals from each bin is calculated. The RMSE and δ are plotted against an identity function and checked for linearity. Preliminary results of both graphical analyses of the training data indicate that while the model uncalibrated standard deviation predictions δ_{UC} estimates are fair, they may be improved with calibration.

The Random Forest model prediction uncertainty was thus calibrated according to Palmer et al. (2022), which assumes that the error of predictions – or, residuals $R(x)$ from the out-of-bag data D_{OOB} – are linearly related to δ_{UC} for any given point, or $\delta_{\mathrm{Cal}} = a\,\delta_{\mathrm{UC}}(x) + b$. Calibration values were obtained by optimizing Eq. (3) using the Nelder-Mead algorithm in R:

$$a, b = \mathrm{argmin}_{a',b'} \sum_{x,y \in D_{OOB}} \ln(2\pi) + \ln((a'\delta_{UC}(x) + b')^2) + \frac{R(x)^2}{(a'\delta_{UC}(x) + b')^2}. \quad (3)$$

This minimizes the negative log likelihood that residuals $R(x)$ are normally distributed with $mean = 0$, and $SD = a\delta_{UC} + b$ to find optimized values of a and b (slope and intercept) that are used to calibrate δ_{UC} (uncalibrated standard deviation) via linear model. To validate these results the δ_{C} of the test data were then calculated by scaling δ_{UC} by a and b using Eq. (4):

$$\delta_{Cal} = a\delta_{UC} + b. \quad (4)$$

These were graphically and statistically validated using the same methods detailed above. Following this, the calibration factors were applied to the entire predicted δ data and mapped spatially.

3 Results

3.1 Mapped predictions and uncertainty

Both models predicted high values of Al% in several places throughout the island, notably the Mount Peyton (central northeast), Avalon (most easterly), and Deer Lake (south central west) areas (Figure 3 a, b). Highs also extend into the southwestern portion of the island, near Port Aux Basques. Low values were predicted in south central Newfoundland, over known weathered sedimentary packages. These patterns were predicted consistently by both models and are not surprising given the geological context. While both models predicted a high Al% near Gros Morne (central western coast), the spatial Model 2 predicted high values extending into the Great Northern Peninsula (northwest) and generally predicted higher values than the aspatial Model 1 in the southwestern area of the island.

The prediction uncertainty δ was highest closer to the ocean and with increasing distance from samples – especially on the Great Northern Peninsula (Figure 3 c, d) where samples are sparse or non-existent (Figure 1). While this pattern was observed in both outputs, it was more prominent in the spatial model. The aspatial model showed δ spread relatively evenly

throughout the island, with lower values in areas close to samples and areas near the coast or distant from samples showing higher values. Unsurprisingly, bedrock units with few samples also showed a high δ.

245

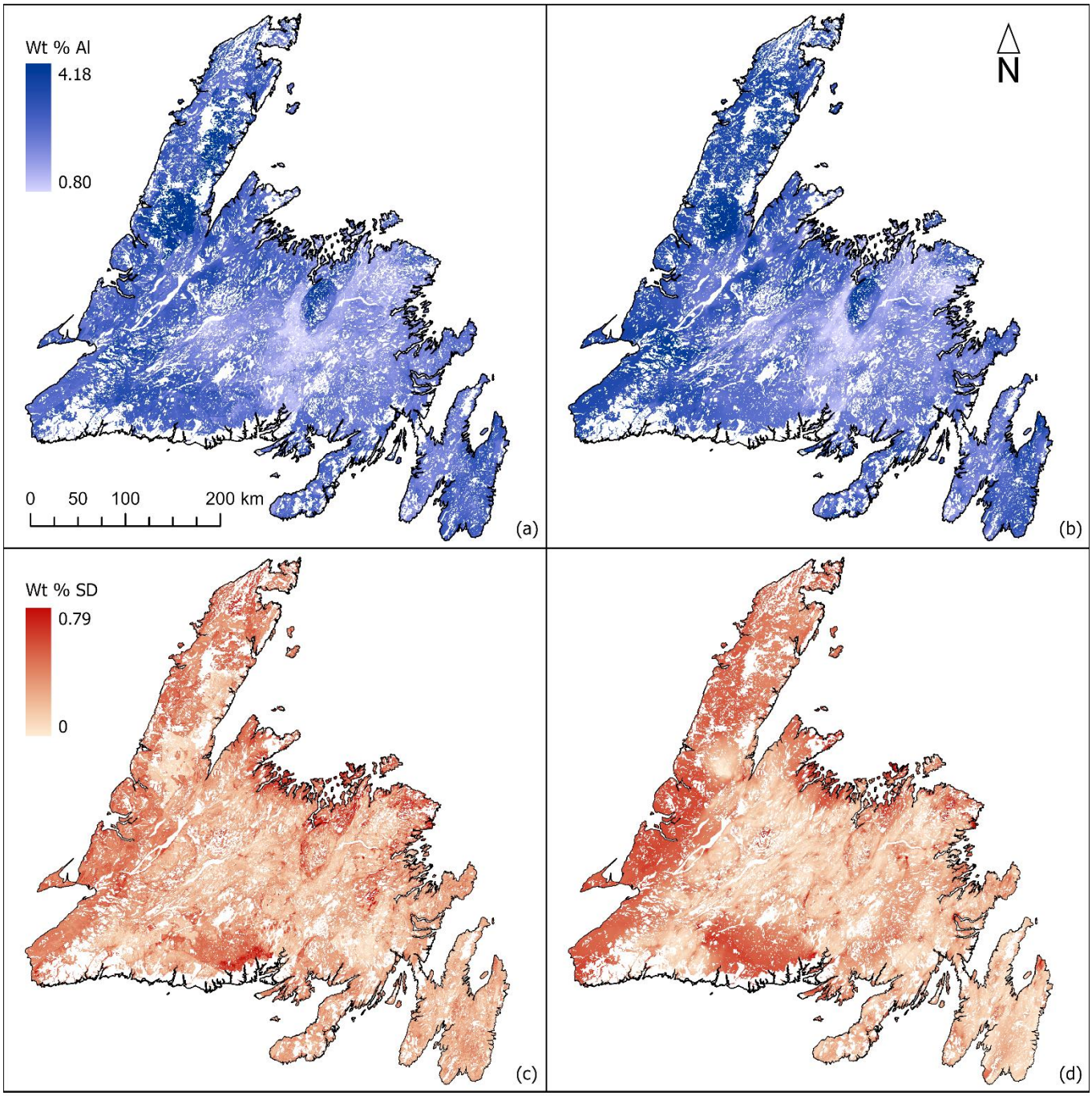


Figure 3 Aluminum availability (Al%) predictions and corresponding calibrated uncertainty predictions given as Al% (SD) across the island of Newfoundland with aspatial (a, c) and spatial (b, d) models.

3.2 Uncertainty calibration

Calibrated uncertainty of both models conforms closely to the identify function and the standard normal distribution. Scatterplots or the r statistic clarified that for both models, uncalibrated points lie beneath the identity function, indicating that both models overestimated δ. They also indicated a positive slope between calibrated and uncalibrated values, demonstrating positive correlation between the size of residuals and the size of δ estimates. Deviations from this pattern exist – though primarily in poorly sampled bins (Figures 4 e and f). In both cases, calibration improved the fit of data to both the normal distribution and the identity function. After calibration, normality and accuracy of the r statistic was improved, with standard deviations calibrated from 0.46 and 0.41 to 0.97 and 0.94, and the RMSE of the δ compared to the out-of-bag errors decreased from 0.317 and 0.264 to 0.207 and 0.160 for the aspatial and spatial models respectively – a decrease of approximately 34% and 47%.

The calibration was most reliable in areas of moderate uncertainty. Figures 4 e and f show the highest uncertainty bin deviating significantly from the identity function, indicating that samples with a calibrated δ of greater than approximately 0.22 and 0.25 for the aspatial and spatial models respectively are underestimated. These points represent approximately 7% of samples from both models. Application of the calibration factors to the δ rasters translated to 19% and 40% of δ being underpredicted across Newfoundland aspatially and spatially, respectively (Figure 5). These underpredicted areas coincide entirely with areas of low confidence due to a paucity of samples. This is to be expected; calibrating uncertainty addresses systemic error within model predictions, not error associated with issues in sample distribution. Figures 4 a, b, e, and f clearly demonstrate that overall, the uncertainty calibration corrects the consistent underprediction of δ by both models.

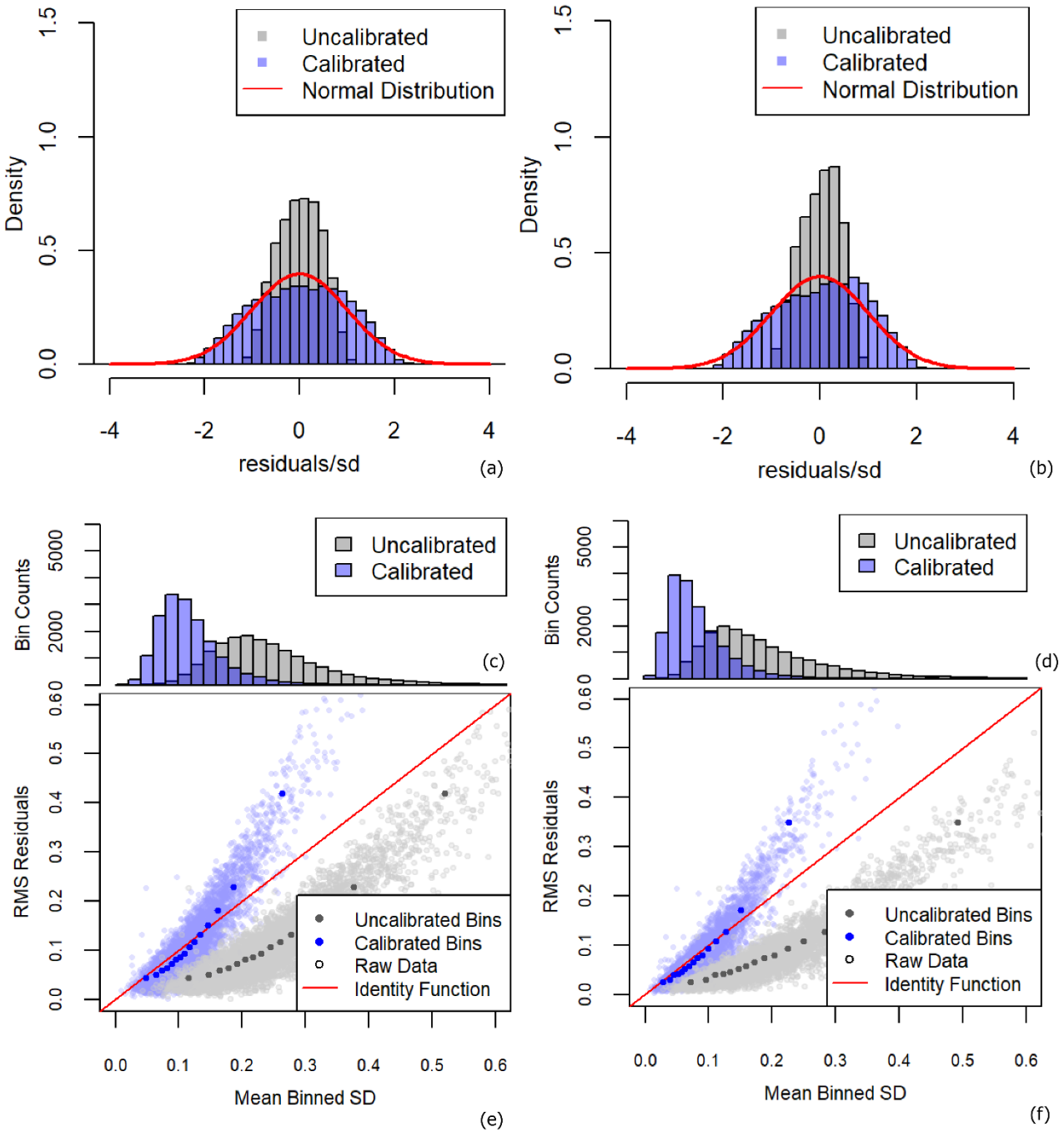


Figure 4 Distribution of the R statistic (a,b), distribution of sample δ (c,d), and residuals vs δ plots of both raw data and binned data (e, f) for aspatial (a, c, e) and spatial models (b, d, f).

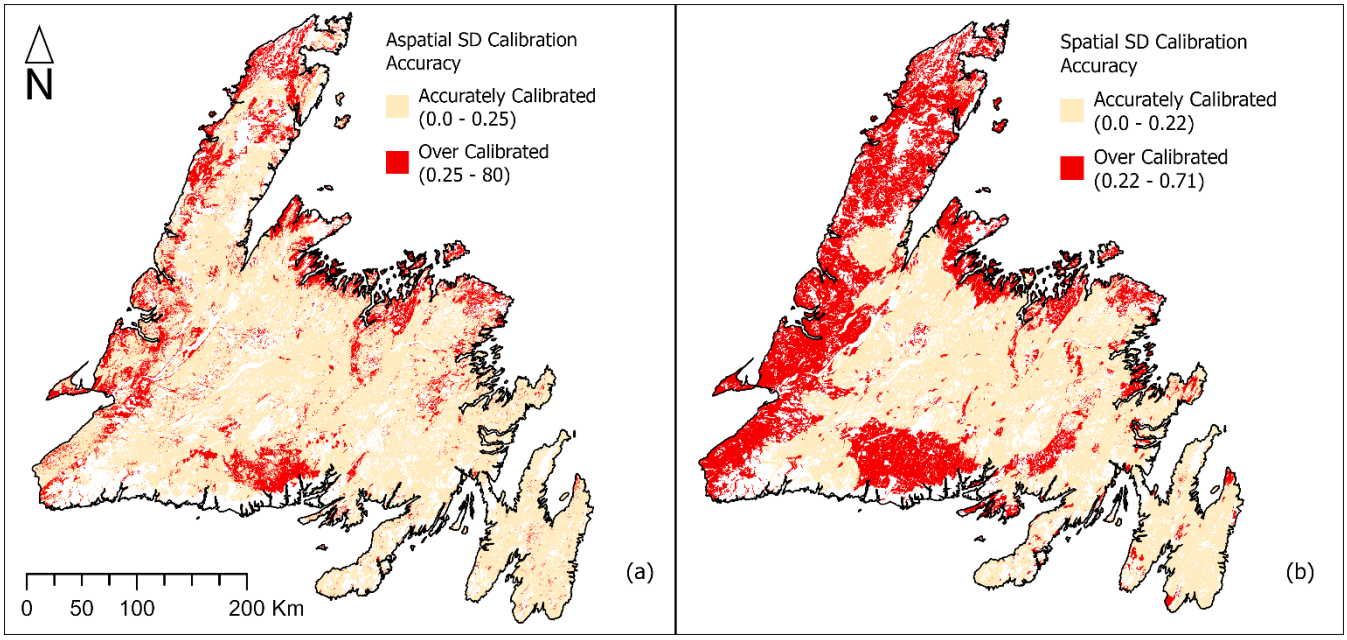


Figure 5 Map of δ calibration accuracy across Newfoundland for aspatial (a) and spatial (b) models.

3.3 Variable importance

Cross-validation demonstrated that Al% in glacial till could be modelled accurately using environmental variables, and with lower uncertainty with the addition of spatial context. The aspatial and spatial models attained R2 values of 0.59 and 0.71, and RMSE values of 0.34 and 0.29 Al wt% respectively. These results suggest that the spatial model explained >10% more of the variance within sample data than the aspatial model, corresponding to a prediction uncertainty which was lower, on average, by ~0.12 Al wt%.

The most influential variables for the aspatial model were bedrock units, distance from the ocean, elevation, and unit size (Table 2). Climate zone, slope, ruggedness, and surficial deposit type had the least control on Al%. Distance to the ocean was not expected to exert this much influence, and could represent spatial autocorrelation between samples, which motivated the second spatial model (see Methods section). The importance % score of the bedrock variables suggests that while environmental variables can be used to model Al%, the till samples relationship to bedrock and to one another are the most important prediction factors. Partial dependence plots for the spatial model were reviewed during the modelling process but due to the large number of bedrock predictors the interpretation of the plots was convoluted. This is largely because individual

bedrock units are of great importance to some points (locally), but not to others (those at a distance from the unit). The value of these plots was diminished due to this, and as a result they were not included in the study.

Table 2 Aspatial and spatial model variables of importance given as a % of total variance explained.

Aspatial Model		**Spatial Model**	
Variable	**Importance (%)**	**Variable**	**Importance (%)**
Surficial Geology	4.08	Surficial Geology	4.02
Climate	4.35	Climate	0.22
Ruggedness	7.76	Ruggedness	1.36
Slope	7.77	Slope	1.18
Unit Size	13.68	Unit Size	1.94
Elevation	15.97	Elevation	1.83
Distance to Ocean	18.23	Distance to Ocean	2.16
Bedrock	28.14	Bedrock	86.96

4 Discussion

4.1 Utility of the approach

Results illustrate the value of geospatial environmental variables and spatial context for mapping and modelling available Al. The resulting map addresses an underlying need for incorporation of geological controls into future SOC modelling in boreal forests and will directly inform regional scale modelling efforts as recommended for Newfoundland (Patrick et al., 2022; Patrick, 2023). This approach has utility in areas where existing soil samples are sparse, but geological information is abundant. Till sampling is widespread due to its applications in mineral exploration, and Al, Ca, Na, and K contents are commonly measured. For example, the GSNL till geochemistry dataset yielded over 17,000 points while the National Soil Network yielded 231 samples in Newfoundland, only 65 of which recorded Al values. Further, when pedogenesis originates from glacial till, soil forms a direct and often indistinguishable boundary with the till. Till samples thus provide an excellent alternative to bedrock maps for soil parent material.

4.2 Novelty of this work

While previous mapping efforts have addressed the distribution of reactive Al in soils (Ren et al., 2024), this is the first attempt to map it in soil parent material. Support for modelling till geochemistry through existing datasets is likely more common than reactive Al in soil, as demonstrated in Newfoundland, with potential for providing a key factor supporting SOC modelling across more regions. Few studies investigate till geochemistry from a predictive geochemical standpoint, and those that do focus on economically viable trace elements such as rare earth elements (e.g. Casey et al., 2024). The results developed from this model can be used directly in future SOC models in Newfoundland and provides an important demonstration relevant to the development of SOC modelling in other boreal regions.

The uncertainty calibration is an uncommon yet important component to this project. Addressing uncertainty in SOC modelling has been identified as essential (Mishra et al., 2013); however, while many digital soil mapping studies produce spatial predictions (Pouladi et el., 2023) few provide spatial estimates of uncertainty, and even fewer provide calibrated uncertainty estimates. The calibration step may be critical for accurately reporting model uncertainty; here, calibration improved uncertainty estimates by over 30% in both cases. These calibrated estimates of uncertainty are critical for operationalizing these results to accurately constrain estimates of SOC stocks. An accurate spatial representation of uncertainty indicates regions of the study area that are predicted with lower confidence, which could be a) integrated into subsequent models or calculations of SOC stock for the island of Newfoundland, or b) used to guide future sampling to improve spatial estimates of Al%.

4.3 Limitations

Several constraints limit these models. While our results indicate bedrock geology is the largest control on Al% generally, there is a lack of understanding as to what variables, or interactions between variables, exactly control Al weathering within glacial till. For example, biotic factors, which have been shown to influence soil Al% at coarser scales (Ren et al., 2024), were not explored for the purpose of this study. Further, the model used geological context to leverage spatial autocorrelation within the samples, mimicking a geostatistical approach. While this improved model performance, it renders partial dependence plots for the spatial model difficult to interpret, and the individual variables difficult to rank. Finally, because the work was focused strongly on spatial predictive accuracy, multicollinearity was left largely unaddressed, making it possible that variable importance was split across multiple correlated predictor variables. Such limitations mean that this modelling approach is less useful for discerning strong predictors for available Al content in glacial till but rather is a useful approach for obtaining a spatial prediction for available Al in soil parent material. This represents a first order component for predictive modelling of boreal forest SOC. Next steps, however, should focus on assessing predictors of the relationship between available Al in parent material and reactive Al in the overlying boreal forest mineral soil including biotic topographic, and climatic factors and their interactions (Brantley et al., 2017; Maher, 2010, Patrick 2023). Finally, the CIA

is a one-dimensional value quantifying the proportion of mobile cations to immobile Al, and is the most common index relying on field and geochemical studies, verified by kinetic and thermodynamic modeling decades ago (Nesbitt & Young, 1982; Nesbitt & Young, 1984; Nesbitt et al., 1997). Another index using commonly available major element geochemistry, such as the mafic index of alteration (MIA; Babechuk et al., 2014), could also be applied. However, all such one-dimensional indices suffer from mineralogical controls independent of petrogenic effects (Nesbitt & Wilson, 1992; Babechuk et al., 2014). These are issues largely solved for the purposes of siliciclastic sedimentary geochemistry using three-dimensional approaches (Fedo & Babechuk, 2023). Furthermore, use of the chemical index of alteration (CIA) to inform on available Al in till is likely to vary with large variation in carbonate content. If incongruencies between in situ soil reactive Al and Al% are found, then use of complementary data such as total inorganic carbon could be evaluated as a means to improve the relationship between Al availability in the till and overlying mineral soil.

4.4 Utility of the spatial and aspatial models

The patterns of uncertainty across the spatial and aspatial models indicate that the spatial model may be more useful in previously sampled areas, while the aspatial model may deliver higher certainty in areas with scarce samples. While the overall accuracy of the aspatial model is lower, uncertainty is distributed more evenly across space than the spatial model. As seen in Figures 4 and 5 the calibration accuracy is slightly higher in the spatial model than the aspatial model. It may be possible to aggregate both models for regional scale mapping, while for finer scale mapping the model with higher local performance will be preferred.

4.5 Future work

Soil geochemistry is controlled largely by parent material – in this case till – and this in turn controls the ability of soils formed to store C. For example, reactive Al content of soil is a key factor controlling SOC content of boreal forest podzols in the study region (Patrick et al., 2022). The exact nature of the relationship between till and soil geochemistry is not fully understood, however, topography and climate likely play key interacting roles. To inform this relationship, soil and till should be sampled simultaneously at sites informed by this predictive model. Specifically, sampling where there is forecasted high or very low Al, as these extremes may best illustrate the relationships between soil Al% and till Al%. Sampling should be conducted on similar slopes to ensure that the relationship between till and soil is being captured without influence from terrain characteristics, where higher slope areas experience less vertical and more lateral flow resulting in a shallower horizon of SOC enrichment (Patrick, 2023). As depth of horizon is strongly correlated to other soil characteristics governing SOC storage and this may cause variations in SOC content not attributable to parent material alone (Patrick, 2023). Sampling on high and low slopes should then also proceed to assess controls of terrain on SOC storage.

The GSNL continually updates their till sample geochemistry database and new samples should be incorporated as they become available. Till samples from western Newfoundland, particularly the Great Northern Peninsula, will be especially valuable. Additionally, independent glacial till sampling should be prioritized in high uncertainty areas to increase the robustness of these models.

Given the important role that parent material plays in regulating SOC (Slessarev et al. 2022; Patrick et al., 2022) increased attention must be given to improving parent material composition estimates for SOC modelling. This includes further investigation into key parameters controlling Al weathering and thus reactive Al in parent material. The results of the calibration in this study demonstrate the advantages of uncertainty quantification. Future SOC modelling should incorporate this dimension, as should any future mapping of parent material, as it is crucial to producing meaningful model predictions. In the case of modelling SOC, the more reliable the uncertainty the clearer the utility of the predictions for future model development or application for mitigation efforts.

5 Conclusion

Using both a spatial and aspatial approach, available Al% was successfully modelled and mapped in soil parent material across the Island of Newfoundland for the first time at a high resolution. Calibrated uncertainty was also estimated, providing spatial estimates of model confidence throughout the study area. There is broad utility in predicting parent material geochemistry because it is known to influence soil properties and modelling glacial till draws on a sample base much wider, in many cases, than available soil samples. While future research may expand the sample distribution, and explore relationship between glacial till and soil geochemistry, this work provides a critical component for a future mechanistic SOC model tailored to the soils of Newfoundland and Labrador. This will enable further investigation of the performance of previous modelling across various landscapes, topographies, and soil Al% delivering valuable insight into the spatial distribution of boreal forest SOC across the island. By enabling continuous predictions of SOC and better estimates of the terrestrial carbon reservoir in Newfoundland, this future work will provide a template for such research in other boreal regions.

Code Availability

Code developed for both modelling available Al and calibrating uncertainty are publicly available at https://doi.org/10.5281/zenodo.19883886.

.

Data Availability

All raster files, the till geochemistry shapefile, and key results are available at https://doi.org/10.5683/SP3/6AONRU.

Author Contributions

SMQ compiled the data, contributed to methodology, performed data analyses, validated the research outputs, visualized the data, interpreted findings, and prepared the original draft of the manuscript. MEP conceptualized and designed the project and initially supervised the research. BM contributed to coding and methodology, interpreted findings, supervised all aspects of the research, and reviewed and edited the manuscript. SEZ conceptualized the study, interpreted findings, supervised all aspects of the research, and reviewed and edited the manuscript.

Competing Interests

The authors declare they have no competing interests.

Acknowledgements

The authors gratefully acknowledge the insight, support, and advice of Heather Campbell of the Quaternary section of the GSNL as it pertains to details of till geochemistry and navigating the GSNL till geochemistry dataset, and that of Michael Babechuk for his insights into mineral chemical weathering indices and their application.

Financial Support

This research has been supported by the Canada Research Chairs and National Sciences and Engineering Research Council of Canada Discovery Grants (RGPIN-2018-05383 and RGPIN-2025-07219) programs.

References

Adhikari, K., Mishra, U., Owens, P. R., Libohova, Z., Wills, S. A., Riley, W. J., Hoffman, F. M., & Smith, D. R.: Importance and strength of environmental controllers of soil organic carbon changes with scale, Geoderma, 375, p.114472. https://doi.org/10.1016/j.geoderma.2020.114472, 2020.

Ameray, A., Bergeron, Y., Valeria, O. et al.: Forest Carbon Management: a Review of Silvicultural Practices and Management Strategies Across Boreal, Temperate and Tropical Forests, Curr Forestry Rep 7, 245–266, https://doi.org/10.1007/s40725-021-00151-w , 2021.

Baldock, J.A., Skjemstad, J.O.: Role of the soil matrix and minerals in protecting natural organic materials against biological attack, Org Geochem 31, 697–710, https://doi.org/10.1016/S0146-6380(00)00049-8, 2000.

Brantley, S. L., Eissenstat, D. M., Marshall, J. A., Godsey, S. E., Balogh-Brunstad, Z., Karwan, D. L., Papuga, S. A., Roering, J., Dawson, T. E., Evaristo, J., Chadwick, O., McDonnell, J. J., & Weathers, K. C.: Reviews and syntheses: on the roles trees play in building and plumbing the critical zone, Biogeosciences, 14, 5115–5142, https://doi.org/10.5194/bg-14-5115-2017, 2017.

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

Casey, P., Morris, G., & Sadeghi, M.: Derivation of Predictive Layers Using Regional Till Geochemistry Data for Mineral Potential Mapping of the REE Line of Bergslagen, Central Sweden, Minerals (Basel), 14, 753-. https://doi.org/10.3390/min14080753, 2024.

Crowther, T. W., Todd-Brown, K. E. O., Rowe, C. W., Wieder, W. R., Carey, J. C., Machmuller, M. B., Snoek, B. L., Fang, S., Zhou, G., Allison, S. D., Blair, J. M., Bridgham, S. D., Burton, A. J., Carrillo, Y., Reich, P. B., Clark, J. S., Classen, A. T., Dijkstra, F. A., Elberling, B., Bradford, M. A.: Quantifying global soil carbon losses in response to warming, Nature (London), 540, 104–108, https://doi.org/10.1038/nature20150, 2016.

Deluca, T. H., Boisvenue, C.: Boreal forest soil carbon: distribution, function and modelling, Forestry: An International Journal of Forest Research, 85, 161–184, https://doi.org/10.1093/forestry/cps003, 2012.

Fick, S.E. and Hijmans, R.J.: WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302-4315, 2017

Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., Knutti, R.: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, J Clim 27, 511–526, https://doi.org/10.1175/JCLI-D-12-00579.1, 2014.

Government of Canada, HRDEM, https://open.canada.ca/data/en/dataset/957782bf-847c-4644-a757-e383c0057995, 2024.

Government of Newfoundland and Labrador Fisheries, Forestry, and Agriculture, NLDEM, https://geohub-gnl.hub.arcgis.com/maps/7203aef91a024ece843c4be4b3727dda/about, 2023.

Heung, B., Bulmer, C. E., & Schmidt, M. G.: Predictive soil parent material mapping at a regional-scale; a Random Forest approach, Geoderma, 214–215, 141–154, https://doi.org/10.1016/j.geoderma.2013.09.016, 2014.

Hijmans, R.: terra: Spatial Data Analysis, R package version 1.8-2, https://CRAN.R-project.org/package=terra, 2025.

Hillier, J., Brentrup, F., Wattenbach, M., Walter, C., Garcia-Suarez, T., Mila-i-Canals, L., & Smith, P.: Which cropland greenhouse gas mitigation options give the greatest benefits in different world regions? Climate and soil-specific predictions from integrated empirical models, Global Change Biology, 18, 1880–1894, https://doi.org/10.1111/j.1365-2486.2012.02671.x, 2012.

Hagemann, U., Moroni, M. T., Shaw, C. H., Kurz, W. A., & Makeschin, F.: Comparing measured and modelled forest carbon stocks in high-boreal forests of harvest and natural-disturbance origin in Labrador, Canada, Ecological Modelling, 221, 825–839, https://doi.org/10.1016/j.ecolmodel.2009.11.024, 2010.

Ilich, A. R., Misiuk, B., Lecours, V., & Murawski, S. A.: MultiscaleDTM: An open-source R package for multiscale geomorphometric analysis, Transactions in GIS, 27, 1164-1204, https://doi.org/10.1111/tgis.13067, 2023.

Jackson, R.B., Lajtha, K., Crow, S.E., Hugelius, G., Kramer, M.G. and Pineiro, G.: The ecology of soil carbon: Pools, vulnerabilities, and biotic and biotic controls, ANNU REV ECOL EVOL S 48, 419-45, 2017.

Jenny, H., Gessel, S. P., & Bingham, F. T. (1949). Comparative study of decomposition rates of organic matter in temperate and tropical regions. Soil Science, 68, 419–432. https://doi.org/10.1097/00010694-194912000-00001

Lawley, R., Smith, B., de Lourdes Mendonca-Santos, M., Hartemink, A. E., McBratney, A. B., McBratney, A., Mendonça-Santos, M. de L., & Hartemink, A. E.: Digital soil mapping at a national scale; a knowledge and GIS based approach to improving parent material and property information, In Digital Soil Mapping with Limited Data, 173–182, https://doi.org/10.1007/978-1-4020-8592-5_14, 2008.

Maher, K.: The dependence of chemical weathering rates on fluid residence time, Earth and Planetary Science Letters, 294, 101–110, https://doi.org/10.1016/j.epsl.2010.03.010, 2010.

McNally, A., NASA/GSFC/HSL: FLDAS Noah Land Surface Model L4 Global Monthly 0.1 x 0.1 degree (MERRA-2 and CHIRPS), Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), 10.5067/5NHC22T9375G, 2018.

McNicol, G., Bulmer, C., D'Amore, D., Sanborn, P., Saunders, S., Giesbrecht, I., Arriola, S. G., Bidlack, A., Butman, D., & Buma, B.: Large, climate-sensitive soil carbon stocks mapped with pedology-informed machine learning in the North Pacific coastal temperate rainforest, Environmental Research Letters, 14, 14004-, https://doi.org/10.1088/1748-9326/aaed52, 2019.

Mishra, U., Jastrow, J. D., Matamala, R., Hugelius, G., Koven, C. D., Harden, J. W., Ping, C. L., Michaelson, G. J., Fan, Z., Miller, R. M., McGuire, A. D., Tarnocai, C., Kuhry, P., Riley, W. J., Schaefer, K., Schuur, E. A. G., Jorgenson, M. T., & Hinzman, L. D.: Empirical estimates to reduce modeling uncertainties of soil organic carbon in permafrost regions; a review of recent progress and remaining challenges, Environmental Research Letters, 8, 035020, https://doi.org/10.1088/1748-9326/8/3/035020, 2013.

Nesbitt, H. W., & Young, G. M.: Prediction of some weathering trends of plutonic and volcanic rocks based on thermodynamic and kinetic considerations, Geochimica et Cosmochimica Acta, 48, 1523–1534, https://doi.org/10.1016/0016-7037(84)90408-3, 1984.

Newfoundland and Labrador Geological Survey, Till Geochemistry, Newfoundland and Labrador GeoScience Atlas OnLine [dataset], https://geoatlas.gov.nl.ca/, 2024.

Oades, J. M.: The retention of organic matter in soils, Biogeochemistry, 5, 35–70, https://doi.org/10.1007/BF02180317, 1988.

Palmer, G., Du, S., Politowicz, A., Emory, J. P., Yang, X., Gautam, A., Gupta, G., Li, Z., Jacobs, R., & Morgan, D.: Calibration after bootstrap for accurate uncertainty quantification in regression models, Npj Computational Materials, 8, Article 115, https://doi.org/10.1038/s41524-022-00794-8, 2022.

Paré, D., Bognounou, F., Emilson, E. J. S., Laganière, J., Leach, J., Mansuy, N., Martineau, C., Norris, C., Venier, L., & Webster, K.: Connecting forest soil properties with ecosystem services: Toward a better use of digital soil maps—A review, Soil Science Society of America Journal, 88, 981–999, https://doi.org/10.1002/saj2.20705, 2024.

Patrick, M. E., Young, C. T., Zimmerman, A. R., & Ziegler, S. E.; Mineralogic controls are harbingers of hydrological controls on soil organic matter content in warmer boreal forests, Geoderma, 425, Article 116059, https://doi.org/10.1016/j.geoderma.2022.116059, 2022.

Patrick, M.E.: The roles of parent material, climate, and geomorphology in soil organic carbon response to short-term climate change in moist boreal forests, Doctor of Philosophy Dissertation, Department of Earth Sciences, Memorial University of Newfoundland, pp 265, https://doi.org/10.48336/SKF9-1132 2023, .

Pierson, D., Lohse, K. A., Wieder, W. R., Patton, N. R., Facer, J., de Graaff, M.-A., Georgiou, K., Seyfried, M. S., Flerchinger, G., & Will, R.: Optimizing process-based models to predict current and future soil organic carbon stocks at high-resolution, Scientific Reports, 12, 10824–10824, https://doi.org/10.1038/s41598-022-14224-8, 2022.

Pouladi, N., Gholizadeh, A., Khosravi, V., & Boruvka, L.: Digital mapping of soil organic carbon using remote sensing data; a systematic review, Catena (Giessen), 232, 107409-, https://doi.org/10.1016/j.catena.2023.107409, 2023.

Quinn, S. M., Misiuk, B., Ziegler, S., Patrick, M.: Replication Data for: Mapping and modelling a boreal forest soil organic carbon predictor in the glacial till of Newfoundland, Canada, Memorial University Dataverse (Borealis) [dataset], https://doi.org/10.5683/SP3/6AONRU, 2025.

Quinn, S. M., Misiuk, B., Ziegler, S., Patrick, M.: SMQuinn_Mapping_and_modelling_a_boreal_forest_soil_organic_carbon_predictor_in_glacial_tills_of_NL, Zenodo [code], https://doi.org/10.5281/zenodo.16109126, 2025.

Rasmussen, C., Heckman, K., Wieder, W. R., Keiluweit, M., Lawrence, C. R., Berhe, A. A., Blankinship, J. C., Crow, S. E., Druhan, J. L., Hicks Pries, C. E., Marin-Spiotta, E., Plante, A. F., Schädel, C., Schimel, J. P., Sierra, C. A., Thompson, A., & Wagai, R.: Beyond clay: Towards an improved set of variables for predicting soil organic matter content, Biogeochemistry, 137, 297–306. https://doi. org/10.1007/s10533-018-0424-3, 2018.

Ren, S., Wang, C., & Zhou, Z., Global Distributions of Reactive Iron and Aluminum Influence the Spatial Variation of Soil Organic Carbon, Global Change Biology, 30, e17576-n/a., https://doi.org/10.1111/gcb.17576, 2024.

Scharlemann, J. P., Tanner, E. V., Hiederer, R., Kapos, V.: Global soil carbon: understanding and managing the largest terrestrial carbon pool, Carbon Management, 5, 81-91, https://doi.org/10.4155/cmt.13.77, 2014.

Slessarev, E.W., Chadwick, O.A., Sokol, N.W., Nuccio, E.E., Pett-Ridge, J.: Rock weathering controls the potential for soil carbon storage at a continental scale, Biogeochemistry 157, 1-13, https://doi.org/10.1007/s10533-021-00859-8, 2022.

Wang, L.: Assessment of land use change and carbon emission: A Log Mean Divisa (LMDI) approach, Heliyon, 10, e25669 https://doi.org/10.1016/j.heliyon.2024.e25669, 2024.

Wang, X., Wackett, A.A., Toner, B.M., Yoo, K.: Consistent mineral-associated organic carbon chemistry with variable erosion rates in a mountainous landscape, Geoderma, 405, https://doi.org/10.1016/j.geoderma.2021.115448, 2022.

Zhang, H., Goll, D. S., Wang, Y., Ciais, P., Wieder, W. R., Abramoff, R., Huang, Y., Guenet, B., Prescher, A., Viscarra Rossel, R. A., Barré, P., Chenu, C., Zhou, G., & Tang, X.: Microbial dynamics and soil physicochemical properties explain large-scale variations in soil organic carbon, Global Change Biology, 26, https://doi.org/10.1111/gcb.14994, 2020.

Zhou, Z., Ren, C., Wang, C., Delgado-Baquerizo, M., Luo, Y., Luo, Z., Du, Z., Zhu, B., Yang, Y., Jiao, S., Zhao, F., Cai, A., Yang, G., & Wei, G.: Global turnover of soil mineral-associated and particulate organic carbon, Nature Communications, 15, https://doi.org/10.1038/s41467-024-49743-7, 2024.