Investigating the Spatial Variability in Soil Geochemical and Colour Properties Across Two Contrasting Land Uses

Maria Luna[a], Alexander J Koiter[b,*], Taras E Lychuk[c], Arnie Waddel[c], Alan Moulin[c]

[a] *Brandon University, Masters in Environmental and Life Sciences, 270 18th St, Brandon, R7A 6A9*

[b] *Brandon University, Department of Geography and Environment, 270 18th St, Brandon, R7A 6A9*

[c] *Agriculture and Agri-Food Canada, Brandon Research and Development Centre, 2701 Grand Valley Road, Brandon, R7A 5Y3*

Abstract

Quantification and accurate assessment of the spatial variability and distribution of soil physical and biogeochemical properties are vital components of agri-environmental research and modeling, including sediment source fingerprinting. Understanding the distribution of soil properties is crucial in the development of appropriate, reliable, and efficient sampling campaigns. This study was aimed to investigate the spatial variability in soil geochemical and colour (i.e., spectral reflectance) soil properties (<63um) across two contrasting land uses. The main objectives of this study are to: 1) quantify the spatial variability of geochemical and colour properties at a field-scale (~ 40 ha) across agricultural and forested sites; 2) evaluate the spatial variability and distribution of soil properties and its relation to seven terrain attributes (e.g., catchment area, elevation). A combination of univariate analysis and geostatistical methods were applied to characterize the soil geochemistry and colour properties. This information was used to both quantify and assess the variability in soil properties. The variability and spatial autocorrelation were generally both site and soil property specific. For a selection of soil properties exhibiting some spatial autocorrelation, random forest regression was used to indentify the relative importance of terrain attributes on observed patterns of soil geochemical and colour properties. Elevation was found to explain the greatest amount of the variation in soil properties followed by the SAGA wetness index and relative slope position. These

*Corresponding author

Email addresses: LUNAMIMA56@brandonu.ca (Maria Luna), koitera@brandonu.ca (Alexander J Koiter), taras.lychuk@AGR.GC.CA (Taras E Lychuk), arnie.waddell@AGR.GC.CA (Arnie Waddel), apmaafc7788@gmail.com (Alan Moulin)

types of information can be used to help create efficient soil sampling designs by providing information that can inform sampling locations and number of samples collected in order to meet research needs and objectives.

Keywords: Soil geochemistry, Soil colour, Spatial analysis

1. Introduction

Variation in soil biological, chemical, and physical properties occurs across the landscape and in response to both regional and local (i.e., field-scale) variations in parent material, relief or topography, organisms, climate, time and land use/management (Mulla and McBratney, 2002). Understanding the patterns and drivers for this variation is an important component of many agri-environmental studies. For example, the spatial variability in soil properties is a key component for plant nutrient management plans as investigating the spatial variation influences how suitable producers sample their fields for fertilization and compliance with management practices and environmental regulations (Kariuki et al., 2009).

The spatial variability in soil properties is often caused by changes in topography that affect the storage and transport of water within the soil profile (Mulla and McBratney, 2002). The soil variability at a field-scale might be described by methods of geostatistics and soil mapping. Geostatistics is a technology used to estimate the soil property values in non-sampled areas or areas with sparse samplings and provides a set of statistical tools for a description of spatial patterns, quantitative modeling of spatial continuity, spatial prediction, and uncertainty assessment (Goovaerts, 1999). Geostatistical techniques that incorporate spatial information into predictions can improve estimation and enhance map quality (Zhang et al., 2007). Most applications of this statistical technique were estimated in small-scale areas or on a field-scale with minimal work completed in large land areas or soil regions even though standard geostatistical techniques have been suitable for describing the spatial distribution of soil and to successfully analyze spatial variability of soil properties (Zhang et al., 2007). Geostatistics is focused on spatial correlation, spatial interpolation, simulation, and visualization of values of random variables. Spatial interpolation is commonly used to predict a value of a variable of interest at unmeasured locations with the available measurements at sampled sites (Robertson, 2008). Kriging, for instance, is a multistep method of spatial interpolation, which assumes that the direction or distance between sample points reflects a spatial autocorrelation that can be used to explain variation in the surface (Oliver and Webster, 1990).

The use of semivariograms illustrate the spatial autocorrelation of the measured sample points at each field. To better understand semivariograms, three main characteristics are commonly used to describe these models. The range is the distance where the model first flattens. For instance, sample locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not. The sill is the value that the semivariogram model attains at the range (i.e., the value on the y-axis). The nugget refers to when at zero separation distance, the semivariogram value is zero. However, at a small separation distance, the semivariogram often exhibits a nugget effect, which is a value greater than zero (Oliver and Webster, 1990; Robertson, 2008). Spatial dependency or autocorrelation is often evaluated in terms of ratio of the nugget to sill expressed in percentage or proportion. For instance, a low ratio (less than 25%) means that a large part of the variance is introduced spatially, indicating a strong spatial dependency of the variable. A high ratio (more than 75%) generally indicates weak spatial dependency; otherwise, the spatial dependency is considered moderate (25% - 75%) (Cambardella et al., 1994).

From a sediment fingerprinting perspective, investigating the spatial variability of soil properties at a field-scale can be advantageous to identify, classify, and distinguish between potential sources of sediment (Boudreault et al., 2019). However, only some studies have explored fingerprint variability at individual sampling locations (e.g., Du and Walling, 2017; Pulley et al., 2017) as well as over larger scales (e.g., Wilkinson et al., 2015). Therefore, investigating spatial variability is not common in fingerprinting studies. The field-scale variability in fingerprint properties at different scales has been employed in sediment fingerprinting studies. For example, in a study conducted by Collins et al. (2019), determined that the main finding derives from reported observations on the variability of fingerprint properties (13C, 15N, and total carbon and nitrogen) that were used at a field-scale. Observations of variability from the study are related to sites that present different climates, slopes, soils, and management characteristics. Because in this study the soil was relatively aerobic during the sampling period, it is determined that diffusion of gaseous products (e.g., N and O) thorough the soil profile might be the cause of having lack of variation in geochemistry between soil layers. Similarly, in order to estimate the variability of soils fingerprint properties, topsoil characteristics, such as texture, and organic matter (OM) content are important data to take into account (Miranda et al., 2006).

The degree of spatial variability in fingerprint properties can be described by the coefficient of variation (CV) and R2, which is associated with the semivariogram. A study conducted by Du and Walling (2017) used radionuclides and geochemistry as the main fingerprint properties in samples of surface soils that were collected. Results present a range CV between 10-30%. Due to the fact that 137Cs fallout to the study field occurred over a period of 25 years, it was reasonable to assume that the spatial variation of the 137Cs activity of the surface soil was reflected by systematic variation related with the effect of soil redistribution during the sampling period. In the case of the other fingerprint properties, the variability could be principally associated with soil redistribution and topography. This study suggests that erosional history and topography are important aspects of designing an appropriate sampling approach.

A different study by Zhang et al. (2007) determined spatial variability of nutrient properties in black soil of Northeast China. The spatial variability of organic matter content, total N, and total P was principally associated with climate and terrain and perhaps the most possible reason was cultivation and the difference in potassium fertilizer application. The difference in annual average temperature in black soil study area was reported to be more than 7°C, which indicated that temperature could be a significant factor that might influence soil nutrient mineralization and accumulation. Therefore, at larger spatial scales, land management history and climate variables are likely important aspects to consider when investigating variability and characterizing sources of sediment.

Methods of geostatistics have been developed to estimate spatial variability of fingerprint and soil physical properties (Goovaerts, 1999). The major application of geostatistics to soil physical properties include the variability of soil colour, texture, structure, density, and porosity. Another application of geostatistics includes the estimation and mapping of soil attributes in unsampled areas. Moreover, a significant contribution of geostatistics is the assessment of the uncertainty about unsampled values, which normally takes the form of a map of the probability of exceeding critical values, such as regulatory thresholds in soil pollution or soil quality.

Geostatistical analyses allow to quantify the spatial distribution and variability based on spatial field-scale of the site(s) and spatial pattern of modeling semivariograms. These models have been broadly applied to assess spatial correlation in soil properties, including physical and geochemical properties. The main objectives in this study are to: 1) quantify the spatial variability in soil geochemical and colour fingerprint properties at a field-scale across two contrasting land uses (agriculture and forest) using geostatistics and terrain analysis; and 2) identify patterns and sources of variability in fingerprint properties over a large spatial scale. The range of spatial scales/sampling designs presented are intended to capture variation in fingerprint properties as part of sediment source fingerprinting studies.

2. Methods

2.1. Site description

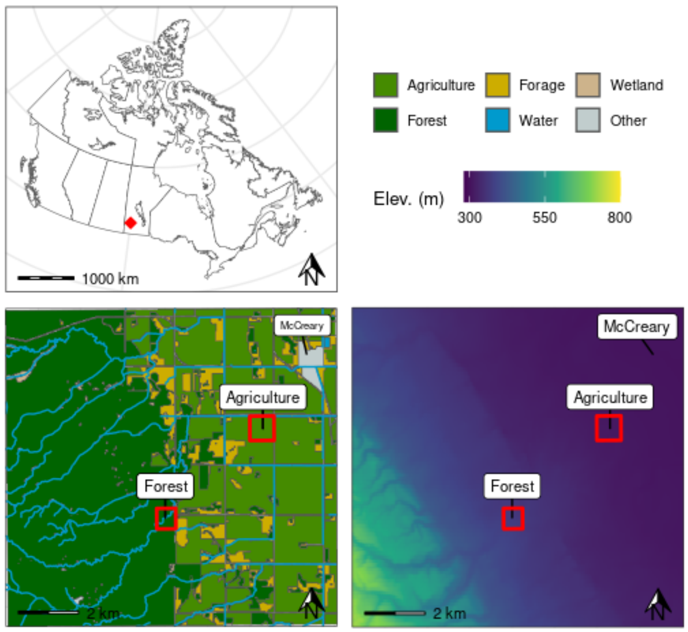


Figure 1: Map showing the location of the study sites within Canada, and the regional land use and topography.

Source: Research Site Locations

3. Results

3.1. Univariate summary

Table 1: Summary univariate statistics of selected geochemical and colour soil properties for each site (n = 49).

Property	Mean	SD	Max	Min	Skewness	CV
			Agriculture			
Ca	4.00	2.19	8.78	0.95	0.28	54.66
Co	8.76	0.83	10.60	7.50	0.52	9.48
Cs	0.75	0.15	1.07	0.47	0.18	19.93
Fe	1.92	0.09	2.11	1.71	−0.25	4.70
Li	15.62	1.42	19.80	12.80	0.62	9.11
La	18.23	1.22	20.20	15.50	−0.29	6.71

Table 1: Summary univariate statistics of selected geochemical and colour soil properties for each site (n = 49).

Property	Mean	SD	Max	Min	Skewness	CV
Nb	0.59	0.06	0.73	0.46	0.45	9.67
Ni	29.63	2.72	35.70	25.00	0.36	9.17
Rb	18.43	4.33	26.70	10.20	0.24	23.48
Sr	91.31	38.98	163.50	38.60	0.09	42.69
a*	3.38	0.32	4.15	2.59	−0.03	9.53
b*	8.84	0.97	10.59	6.69	−0.18	11.00
c*	9.47	1.02	11.32	7.17	−0.19	10.74
h*	1.20	0.01	1.23	1.18	0.19	1.12
x	0.47	0.00	0.48	0.47	0.06	0.46
			Forest			
Ca	1.89	1.53	5.46	0.47	1.07	81.12
Co	6.76	1.39	9.60	4.00	0.03	20.62
Cs	0.55	0.12	0.78	0.34	0.25	21.73
Fe	1.18	0.13	1.46	0.83	−0.58	11.24
Li	6.47	0.90	8.60	4.30	−0.02	13.89
La	15.00	2.60	21.80	10.30	0.33	17.31
Nb	0.37	0.06	0.56	0.17	−0.68	17.10
Ni	18.09	3.90	28.00	11.00	0.33	21.55
Rb	13.83	1.85	18.10	9.90	0.27	13.40
Sr	32.43	12.60	64.20	15.30	0.98	38.87
a*	5.73	0.41	6.56	4.41	−0.38	7.10
b*	12.47	2.01	15.91	8.02	0.22	16.11
c*	13.74	1.94	17.00	9.15	0.15	14.15
h*	1.13	0.05	1.23	1.06	0.34	4.13
x	0.49	0.00	0.49	0.48	−0.21	0.47

Source: Univariate summary

Table 2: Summary statistics for the interpoloated values (10m resolution) for slected geochemical and colour soil properties and terrain attributes for each site.

Property	Mean	SD	Max	Min	Skewness	C
			Agriculture			
Ca	4.12	2.10	8.76	0.918	0.0727	51
Co	8.75	0.664	10.6	7.52	0.431	7.
Cs	0.729	0.123	1.07	0.458	0.376	16
Fe	1.92	0.0644	2.10	1.73	−0.450	3.
Li	15.7	1.16	19.3	13.2	0.551	7.
La	18.2	0.817	19.8	16.5	−0.268	4.
Nb	0.593	0.0550	0.740	0.459	0.569	9.

Table 2: Summary statistics for the interpoloated values (10m resolution) for slected geochemical and colour soil properties and terrain attributes for each site.

Property	Mean	SD	Max	Min	Skewness	C
Ni	29.9	2.23	34.5	26.3	−0.0100	7.
Rb	18.0	3.94	26.1	11.5	0.498	21
Sr	93.4	38.6	167	36.3	0.00105	41
a*	3.34	0.211	3.83	2.88	0.0621	6.
b*	8.73	0.707	10.2	6.98	−0.162	8.
c*	9.34	0.762	11.0	7.41	−0.158	8.
h*	1.20	0.00977	1.23	1.18	−0.0603	0.8
x	23.1	1.31	26.6	18.6	−0.501	5.
Plan Curvature	1.65×10^{-6}	1.36×10^{-4}	6.57×10^{-4}	-5.07×10^{-4}	3.54×10^{-1}	8.24
Profile Curvature	-7.64×10^{-6}	1.53×10^{-4}	5.83×10^{-4}	-6.47×10^{-4}	9.51×10^{-2}	−2.00
SAGA Wetness Index	9.64	0.704	11.2	7.77	−0.122	7.
Catchment Area	475	1,010	10,100	4.35	4.76	2
Rel. Slope Position	0.718	0.288	1.20	0.0221	−0.946	40
Vert. Dist. Channel	5.98×10^{-2}	4.10×10^{-2}	2.92×10^{-1}	4.25×10^{-3}	1.21	6.85
Elevation	310	0.593	312	309	0.615	0.1
			Forest			
Ca	1.88	0.769	3.61	0.787	0.202	40
Co	6.80	0.632	8.66	4.93	−0.200	9.
Cs	0.551	0.0737	0.714	0.423	0.297	13
Li	6.43	0.694	8.46	4.39	−0.136	10
La	15.0	1.57	18.5	11.5	−0.0324	10
Nb	0.370	0.0356	0.440	0.278	−0.436	9.
Ni	18.2	2.49	24.9	14.3	0.314	13
Sr	31.6	8.50	53.1	18.1	0.716	26
h*	1.13	0.0371	1.22	1.06	0.257	3.
x	19.7	3.92	30.5	10.8	0.362	19
Plan Curvature	3.97×10^{-4}	3.27×10^{-3}	2.89×10^{-2}	-2.62×10^{-2}	7.91×10^{-1}	8.22
Profile Curvature	-1.83×10^{-4}	9.47×10^{-3}	6.37×10^{-2}	-7.37×10^{-2}	-5.31×10^{-1}	−5.18
SAGA Wetness Index	6.00	0.988	8.48	2.21	−0.430	16
Catchment Area	571	1,940	25,400	3.44	6.60	33
Rel. Slope Position	0.222	0.232	0.993	0.00617	1.56	10
Vert. Dist. Channel	4.15×10^{-1}	4.43×10^{-1}	3.66	2.02×10^{-2}	2.96	1.07
Elevation	369	3.34	377	359	−0.184	0.9

Source: Univariate summary

3.2. Spatial analysis

Table 3: Geostatistical parameters of the fitted semivariogram models of selec
geochemical properties within the agricultural and forested sites.

Property	Kriging Type[1]	Nugget (Co)	Sill (Co + C)	C/(C + Co) (%)	Range (m)	r^2
Agriculture						
Ca	Universal	0.0	7.2	100	580	0.9
Co	Simple	0.0	0.7	100	208	0.4
Cs	Ordinary	0.0	0.0	100	210	0.5
Fe	Ordinary	0.0	0.0	100	185	0.2
Li	Universal	0.3	1.5	81	185	0.6
La	Simple	0.4	1.0	56	308	0.5
Nb	Universal	2.7	2.7	0	210	0.7
Ni	Ordinary	1.4	8.9	84	352	0.6
Rb	Ordinary	1.4	27.6	95	551	0.9
Sr	Ordinary	0.9	900.2	100	220	1.0
a*	Ordinary	0.4	1.0	59	288	0.3
b*	Simple	0.2	0.9	83	199	0.3
c*	Simple	0.1	0.9	87	199	0.3
h*	Simple	0.0	1.1	100	185	0.2
x	Simple	0.4	1.0	58	220	0.1
Forest						
Ca	Ordinary	1.6	2.7	41	269	0.2
Co	Ordinary	0.0	2.1	100	298	0.1
Cs	Ordinary	0.0	0.0	83	237	0.2
Li	Ordinary	0.0	0.8	100	222	0.3
La	Ordinary	3.1	7.4	59	176	0.1
Nb	Ordinary	0.0	0.0	51	224	0.2
Ni	Universal	6.7	15.8	57	187	0.2
Sr	Simple	0.4	1.0	65	229	0.4
h*	Universal	0.0	0.0	100	230	0.3
x	Ordinary	0.0	39.2	100	312	0.2

[1] Models are all isotropic.

[2] Strong spatial dependency (C/(C + Co) % >75); Moderate spatial dependency (C/(C + Co) % between 75

Source: Semivariograms

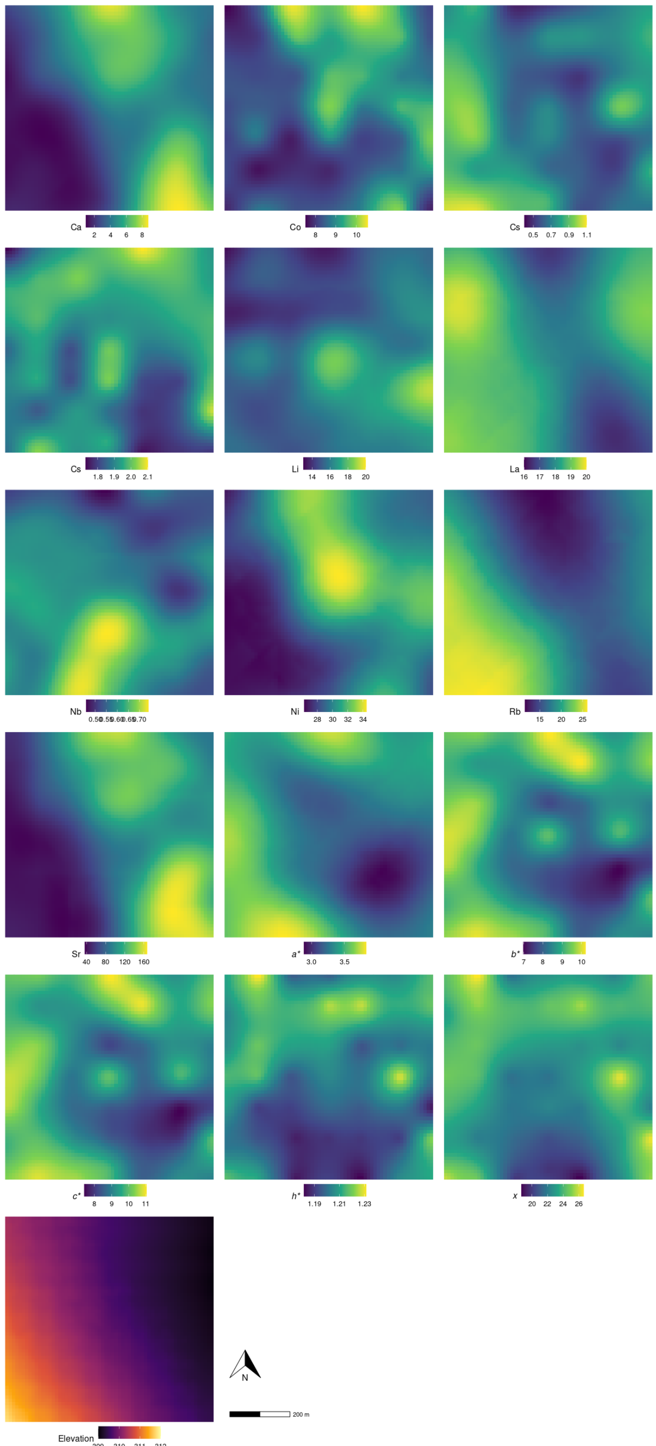


Figure 2: Kriged maps of select colour and geochemical properties and elevtion across the agricultural site.

Source: Soil property mapping

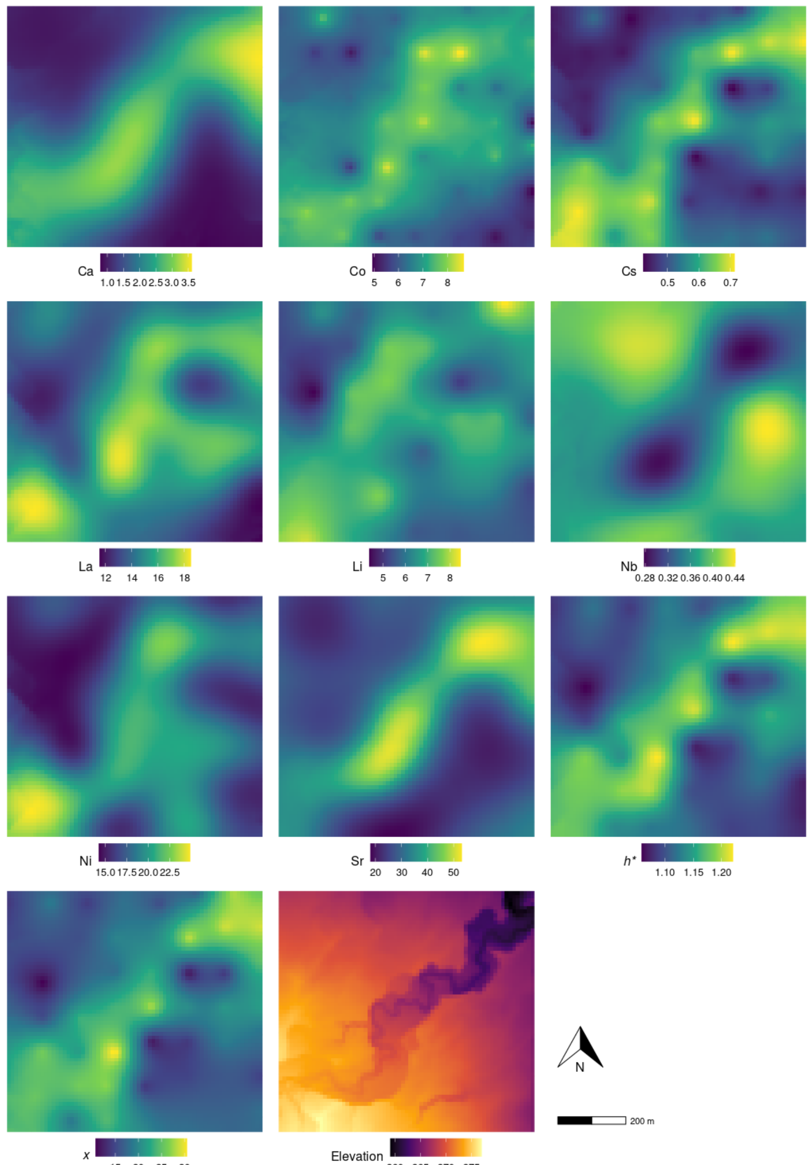


Figure 3: Kriged map of select colour and geochemical properties and elevation across the forested site.

Source: Soil property mapping

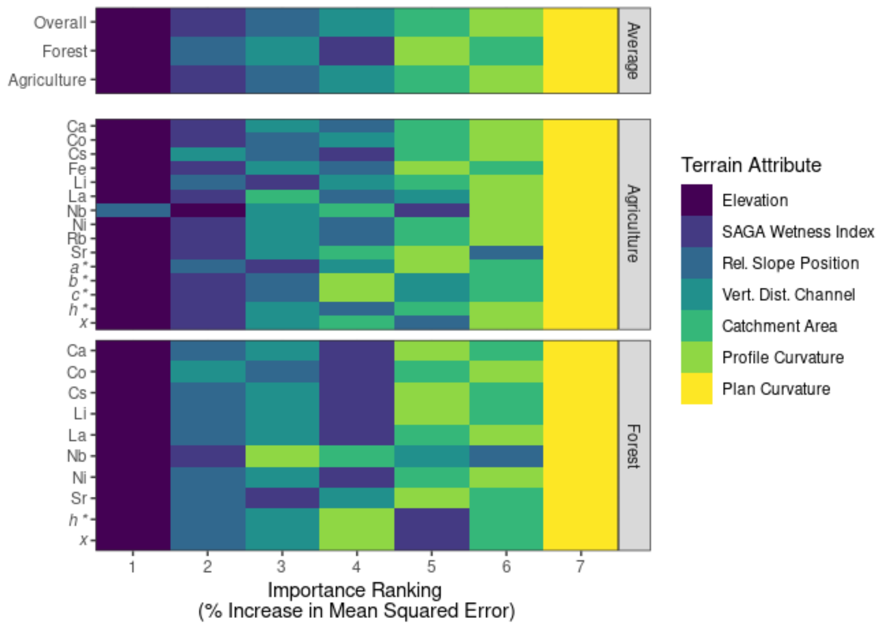


Figure 4: Heat map of the Random Forest regresssion results showing the ranking of the importance of terrain attributes (based on % increase in Mean Squared Error) in explaining the spatial variabilty of selected colour and geochemical properties within the agricultural and forested sites. Top panel shows an average ranking for each site and across both sites.

Source: Random Forest summary

Table 4: Model summary and performance statistics for the random forest regression using the training, validation, and test data sets.

Property	MSE Training[1]	% Var Training[2]	MSE Testing[1]	% Var Validation[2]	R^2 Testing
			Agriculture		
Ca	0.374	91.6	0.359	91.8	0.91
Co	0.089	79.8	0.080	82.5	0.80
Cs	0.002	85.7	0.002	86.4	0.85
Fe	0.001	69.6	0.001	70.9	0.69
Li	0.538	59.3	0.533	59.8	0.64
La	0.048	93.0	0.044	93.1	0.93
Nb	0.001	57.3	0.001	59.1	0.55
Ni	0.338	93.1	0.335	93.7	0.93
Rb	0.733	95.3	0.643	96.1	0.95
Sr	97.221	93.5	93.970	93.6	0.93
a*	0.007	85.0	0.006	86.9	0.85

Table 4: Model summary and performance statistics for the random forest regression using the training, validation, and test data sets.

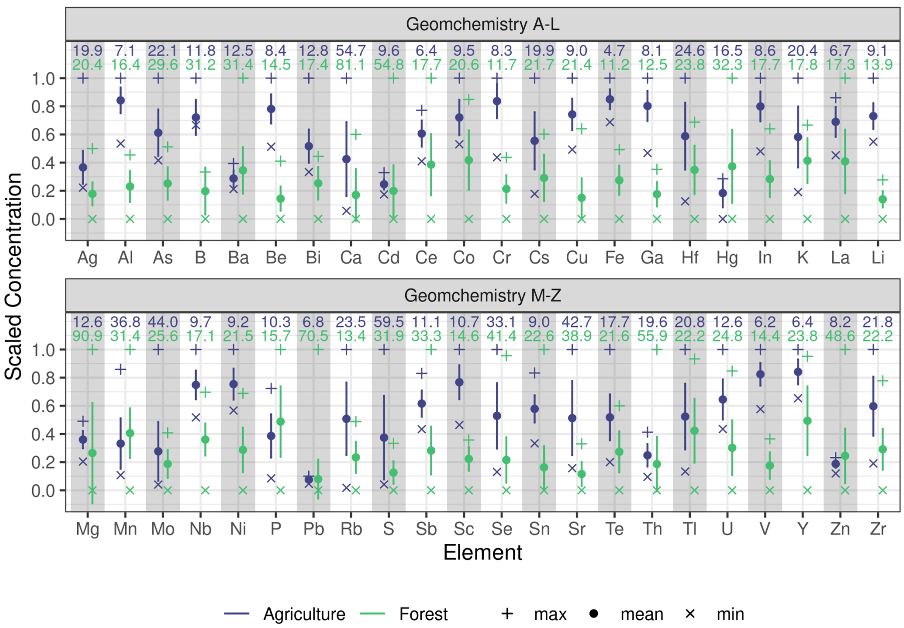
Property	MSE Training[1]	% Var Training[2]	MSE Testing[1]	% Var Validation[2]	R^2 Testing
b*	0.136	72.5	0.120	75.3	0.72
c*	0.155	73.2	0.136	75.9	0.73
h*	0.000	58.3	0.000	58.6	0.56
x	0.628	61.9	0.701	61.8	0.59
			Forest		
Ca	0.231	61.1	0.231	60.7	0.63
Co	0.244	39.1	0.234	42.9	0.48
Cs	0.002	64.1	0.002	67.1	0.66
Li	0.278	41.3	0.282	42.0	0.46
La	1.401	43.3	1.323	47.5	0.48
Nb	0.001	55.0	0.001	55.9	0.58
Ni	2.819	55.2	2.806	56.0	0.55
Sr	29.427	59.4	29.663	59.1	0.59
h*	0.001	58.8	0.001	60.3	0.62
x	5.646	62.6	5.810	63.4	0.64

[1] Mean square error
[2] Percent variance explained

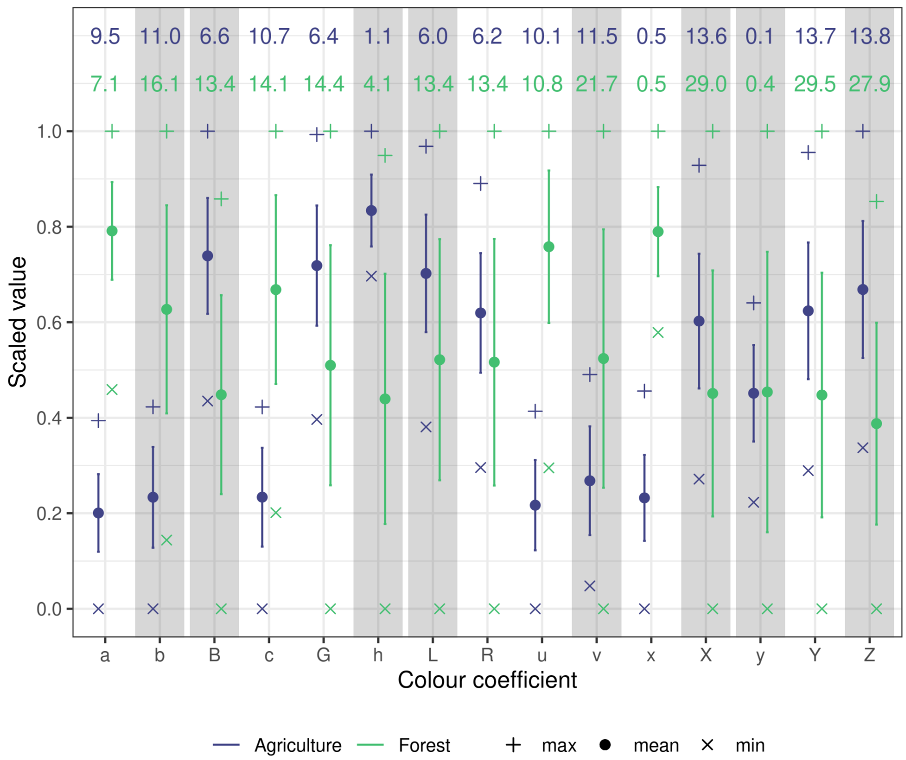
Source: Random Forest summary

References

Supplemental materials



Summary statistics of all measured geochemical soil properties at both sites. Error bars represent 1SD and the numeric values indicate the CV.



Summary statistics of all measured colour soil properties at both sites. Error bars represent 1SD and the numeric values indicate the CV.