# Investigating the Spatial Variability in Soil Geochemical and Colour Properties Across Two Contrasting Land Uses in South-Central Manitoba

Maria Luna Miño[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 at investigating 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) assess 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 identify 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 Miño), 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)

findings 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, Terrain attributes

---

### *0.1. 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 the five soil forming factors: parent material, relief or topography, biota, climate, and time. Superimposed on this is the influence of changes in land use and current and historic management practices which can further modify soil properties. Quantifying and understanding the patterns and drivers for this variation is an important component of many agri-environmental studies. For example, to meet the desired level of precision for agronomic and environmental nutrient management plans the spatial variability in soil nutrients will influence the soil sampling design in terms of number and locations of soil samples [1, 2].

Sediment source fingerprinting is a watershed-scale technique that is used to identify and quantify the relative proportions of sediment derived from unique sources. This technique uses natural occurring biogeochemical properties as fingerprints (i.e., tracers) to discriminate between potential sources of sediment and are linked to downstream sediment using mixing models. From a sediment fingerprinting perspective, investigating the spatial variability of soil properties at a watershed-scale can be advantageous to identify, classify, and distinguish between potential sources of sediment [3]. However, investigating spatial variability at smaller scales is less common [e.g., 4, 5, 6, 7] and remains a research priority [8].

There are three main, interconnected, ways that spatial variability in fingerprint properties are an important aspect of sediment fingerprinting. First is to adequately quantify the fingerprint properties such that it is representative of that source. For some fingerprints the variability is not random but rather varies in a more systematic way. For example, the pattern of fallout radionuclides will reflect the patterns of soil erosion and deposition [9]. Designing and implementing source sampling plans need to take this into consideration as the sampling designed used has been shown to influence the characterization of wide range of commonly used fingerprints [7].

Secondly, the issue of spatial variability of fingerprint properties is further complicated by overlying spatial variability in the rates of erosion and sediment delivery. Incorporation of both types of variability into the mixing model will provide a more reliable estimate of the proportion of sediment derived from each source. Many mixing models have well defined inputs (sources) and outputs (sediment) that are characterized by their mean and standard deviation and the spatial distribution or pattern of fingerprints are not considered. This is not ideal as the values of samples that are collected closer, and more hydrologically connected, to the stream network may in fact present a better representation of that source despite potentially deviating from the mean value. This issue can be addressed by strategic sampling where the more likely to erode areas are targeted for sampling. However, a considerable amount of information and insight is lost through that approach. There has been some progress using information on erosion rates to calculate a erosion rate-weighted mean [9, 4] and using spatially interpolated maps of fingerprint values to provide a finer resolution of the fingerprint variability within each source [10].

Lastly, understanding the geomorphic, hydrologic, and biochemical processes that have led to the observed patterns in spatial variability helps in the selection of robust and reliable fingerprints and/or guide the sampling design for source characterization. In selecting fingerprints that provide good discrimination between sources many studies typically used a statistical-based approach [11]. However, there are concerns that this approach may result in the inclusion of false positives (i.e., type I error) or non-conservative fingerprints [12]. Consequently, there has been a call for the inclusion of a process-based (e.g., weathering, erosion) or geologic/lithologic-based explanation of the fingerprints selected to address these concerns [8]. Furthermore, there is also a lack of standardization in how sediment source areas are sampled (e.g., judgement, random, transect, grid, stratified) and it can be difficult to have an efficient sampling design without prior knowledge of why and how soil properties vary across the landscape [7]. Prior knowledge of the spatial variability of soil fingerprint properties would be beneficial; however, in practice this can be difficult, particularly with geochemical properties as routine lab analysis often return information on more than 50 elements. The spatial patterns of some soil properties are well studied because of their agronomic importance or ability to infer other important soil properties and processes and can include fallout radionuclides [e.g., $^{137}$Cs, $^{7}$Be; [13]], plant nutrients [e.g., N, P; [14]], soil colour [e.g., hue, value; [15]], major non-acid forming cations [e.g., Ca, Na; [16]]. In contrast, the processes that control the distribution of other soil properties, such as rare earth elements and trace metals, are less well studied or tend to be site-specific, making it difficult to draw generalizations.

Terrain attributes such as elevation, slope curvature, slope position, and soil wetness indices have been shown to be useful information in the understanding and modelling of a range of soil properties including soil moisture [17], texture [18], colour [19], organic matter [20], conductivity [21], and geochemistry [22]. Similar techniques may provide additional insight into the pedologic and geomorphic processes that drive the observed patterns of fingerprint properties within a given source. Since digital elevation models (DEMs) are more publicly available and can in some case be generated using drone imagery, while soil property data are often limited, terrain attributes derived from DEMs can be used to guide sampling design.

This study builds on the previous work of Luna Miño [7] where the impact of three different sampling designs on the characterization of source materials, within the framework of the sediment fingerprinting approach, was assessed. This study expands that study by using the data from grid sampling approach to assess the spatial autocorrelation, create soil property (i.e., fingerprint) maps, and identify important terrain attributes driving the observed patterns. The objectives of this study were (1) to investigate the spatial variability of a range of soil colour and geochemical properties in an agricultural and forested site; and (2) to assess the relative importance and correlation of terrain attributes with the spatial distribution of these soil properties. Together, these objectives address how terrain attributes may be used to understand spatial distributions of soil properties and help guide sampling design.

## *0.2. Methods*

### *0.2.1. Site description*

Two sites of contrasting land uses located in the Wilson Creek Watershed (WCW), near McCreary, Manitoba, Canada were selected to investigate the spatial variability in fingerprint properties. The headwaters of the WCW are located on top of the Manitoba Escarpment within the boundary of Riding Mountain National Park. There is a ~300m drop in elevation crosses the escarpment where the streams become deeply incised. At the base of the escarpment is a large alluvial fan situated in the lacustrine deposits of glacial lake Aggasiz where the main stem has a meandering form. However, beyond the national park boundary the stream flows straight through an engineered drain until it reaches the Turtle river (Figure 1). Both sites are both hydrologicaly connected to the mainstem of the Wilson Creek

The first site was a mixedwood forest including white and black spruce (Picea glauca, Picea mariana), balsam fir (Abies balsamea), larch (Larix laricina) and young stands of deciduous trees including trembling aspen (Populus tremuloides). The forested site is located within the boundaries of the national park where there is little disturbance beyond recreational hiking trails. The soils within the park are not well mapped but likely are part of the Grey Wooded soil association (Luvisol) consisting of fine sandy loam to clay loam soils developed on boulder till of mostly shale with some limestone, and granitic rocks [23]. The second site is under agricultural production and includes rotations of grain crops and forage. The site is mapped to the Edwards Soil Series (Cumulic Regosol) consisting of silty clay loam to silty clay soil developed on recent alluvial deposits [23].

The Köppen-Geiger climate classification of the WCW is cold, without dry season, and with warm summer (Dfb) [24]. The average annual precipitation is ~539 mm, with approximately 27% falling as snow with a mean annual temperature is 3.0°C [25]. The hydrology of the watershed is snowmelt dominated with ~ 80% of the cumulative runoff occurring during the spring season (May and June) [26].

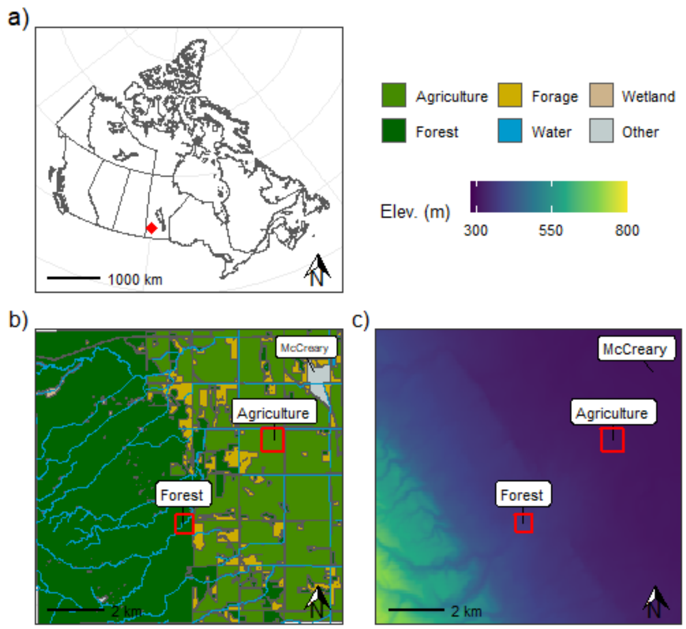


Figure 1: a) Map showing the location of the study sites within Canada. Location of the two study sites and nearby town of McCreary, and regional b) land use, and c) topography.

Source: Research Site Locations

*0.2.2. Soil sampling and analysis*

This study uses samples and data collected as part of the grid sampling design outlined in Luna Miño [7]. Briefly, at each site 49 samples were collected using a soil auger on a 7x7 grid at a 100m spacing. Within the forested surface soil samples were collected below the LFH layer to a depth of 5cm, and the agricultural site was sampled to a depth of 15cm to account for the regular mixing of the soil due to tillage and other field operations.

Samples were dried, homogenized with a mortar and pestle, and sieved through a 63  m sieve to remove the sand fraction. The sand fraction was removed in an effort to reduce the differences in grain size and organic matter content between the two sites [27]. Samples were analyzed for a broad suite geochemical element using inductively coupled plasma mass spectrometry (ICP-MS) following a microwave-assisted digestion with aqua-regia (ALS Mineral Division, North Vancouver, BC, Canada). Spectral measurements were made with a spectroradiometer (ASD FieldSpecPro Malvern Panalytical Inc Westborough MA 01581, United States). Spectral reflectance measurements were taken in 1 nm increments over the 0.4-2.5  m wavelength range. Both samples and Spectralon standard (white reference) were illuminated with a white light source using a halogen-based lamp (12 VDC, 20 Watt). Following the method outlined in

Boudreault et al. [28], fifteen colour coefficients (R, G, B, x, y, Y, X, Z, L, a*, b*, u*, v*, c*, h*) were calculated for each sample [29]. Based on the results of Luna Miño [7], a composite fingerprint consisting of 10 geochemical elements (Ca, Co, Cs, Fe, Li, La, Nb, Ni, Rb, and Sr) and five colour coefficients (a*, b*, c*, h*, and x) were identifying as providing a strong discrimination between the agricultural and forested surface soils. These fifteen soil properties are the focus of the detailed spatial analysis detailed in this study.

*0.2.3. Geospatial and terrain analysis*

All geostatistics were performed with ArcGIS Pro [v 3.3.0 30]. Semivariograms were used to quantify spatial correlation for each of the 15 soil properties. The optimization tool, based on minimizing the mean square error, was used to parameterize the semivariogram model. Kriging was used to interpolate and generate maps of each soil property. The exploratory interpolation tool (Geostatistical Analyst extension) was used to select the kriging type with the highest ranked prediction accuracy.

A Digital Elevation Model (DEM) for the forested site was acquired from publicly available data [31]. A DEM for the agricultural site was generated by photogrammetry using UAV imagery, including the use of ground control and check points, with Agisoft Metashape Professional [v1.8.2 32]. Ordinary kriging was used to calculate a 1 m gridded digital elevation model for each site. Geographic information software [SAGA v2.1.4 33] was used to calculate six additional terrain attributes and included plan and profile curvatures, saga wetness index, catchment area, relative slope position, and vertical channel network distance (Table S2).

*0.2.4. Data analysis*

All subsequent statistical analysis was undertaken using R statistical Software v4.4.0 [34] through RStudio Integrated Development Environment v2024.04.2 [35]. Plots and maps were created using the R package `ggplot2` v 3.5.2 [36]. Skewness was categorized as values between -0.5 and 0.5 considered approximately symmetric, -1.0 to -0.5 or 0.5 to 1 as moderately skewed, and $<$ -1.0 or $>$ 1.0 as highly skewed. Coefficient of variation (CV) thresholds were categorized as low ($<$15%), moderate (15–35%), high (35–75%), and very high ($>$75%). Interpolated soil property and terrain attribute data were resampled to a 10 m resolution prior to analysis [`terra` v1.8.29 37]. Random Forest Regression [`randomForest` v4.7.1.2 38] was used to assess the relative importance of the terrain attributes on the spatial distribution of soil properties. The dataset was randomly split into training, validation, and testing datasets. Multicollinearity among the terrain attributed was assessed using the Variance Inflation Factor with a threshold of eight and correlated terrain attributes were removed [`usdm` v2.1.7 39]. The number of variables randomly sampled as candidates at each split within the random forest model was tuned using the training and validation data sets [`caret` v7.0.1 40]. The number of trees to grow was set to 500 and model performance was assessed using the Mean Square Error (MSE) and percent variance explained for both the training (Out of Bag Error) and the validation data sets. To test the model, actual and predicted values were plotted and the $R^2$ and MSE were calculated using the testing data set. Because analyzing interpolated data can cause issues, the random forest model was used to predict the original 49 non-interpolated data points at each site as an additional check.

## *0.3. Results*

### *0.3.1. Univariate summary*

Overall, the agricultural site had soil colour and geochemical properties that exhibited lower variability and more symmetrical data distributions as compared to the forested site (Table 1). All 15 colour properties at both sites displayed approximately symmetrical distributions. At the agricultural site, all colour properties were characterized by low coefficients of variation (CV), while the forested site showed slightly greater variability, with 10 colour properties having low CVs and five having moderate CVs.

Similarly, geochemical data at the agricultural site showed lower variability and greater symmetry. Most elements were approximately symmetrical, with only nine exhibiting moderate skewness and five highly skewed (Table 1). Variability was also limited, with the majority of elements having low CVs; 12 had moderate CVs and five had high CVs. In contrast, the forested site showed greater skewness and variability: seven elements exhibited moderate skewness, 14 were highly skewed, 28 had moderate CVs, six had high CVs, and two had very high CVs.

Table 1: Summary univariate statistics of selected geochemical and colour soil properties for each site (n = 49). Geochemical concentrations are reported in ppm, excecpt Ca and Fe(%).

| 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 |
| 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 |

Table 1: Summary univariate statistics of selected geochemical and colour soil properties for each site (n = 49). Geochemical concentrations are reported in ppm, excecpt Ca and Fe(%).

| Property | Mean | SD | Max | Min | Skewness | CV |
|---|---|---|---|---|---|---|
| | | | 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

The agricultural site has a relatively flat topography with an elevation change of approximately 3m, with the field draining toward a ditch in the northeast corner. The forested site has a relatively more complex topography, with a channel flowing from the southwest toward the northeast and an overall elevation difference of 18 m across the site. The mean plan and profile curvature measurements for both sites are near zero indicating an area of sediment transit and not accumulation or erosion (Table 2). The agricultural site had a higher SAGA Wetness Index but the forested site had a larger range in values and exhibited a higher degree of variability. The forested site exhibited a smaller mean Relative Slope Position value (streams and depressional areas) and a smaller Vertical Distance to Channel Network, and for both terrain attributes a greater variability as compared to the agricultural reflecting the presence of the stream crossing the forested site.

Table 2: Summary statistics for the interpolated values (10m resolution) for selected geochemical and colour soil properties and terrain attributes for each site. Geochemical concentrations are reported in ppm, excecpt Ca and Fe(%).

| 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. |

Table 2: Summary statistics for the interpolated values (10m resolution) for selected geochemical and colour soil properties and terrain attributes for each site. Geochemical concentrations are reported in ppm, excecpt Ca and Fe(%).

| Property | Mean | SD | Max | Min | Skewness | C |
|---|---|---|---|---|---|---|
| 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. |
| 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 | 0.473 | 0.00149 | 0.477 | 0.470 | −0.0168 | 0.3 |
| Plan Curvature | $1.65 \times 10^{-6}$ | $1.36 \times 10^{-4}$ | $6.57 \times 10^{-4}$ | $-5.07 \times 10^{-4}$ | $3.54 \times 10^{-1}$ | 8.24 |
| Profile Curvature | $-7.64 \times 10^{-6}$ | $1.53 \times 10^{-4}$ | $5.83 \times 10^{-4}$ | $-6.47 \times 10^{-4}$ | $9.51 \times 10^{-2}$ | −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 | $5.98 \times 10^{-2}$ | $4.10 \times 10^{-2}$ | $2.92 \times 10^{-1}$ | $4.25 \times 10^{-3}$ | 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. |
| Plan Curvature | $3.97 \times 10^{-4}$ | $3.27 \times 10^{-3}$ | $2.89 \times 10^{-2}$ | $-2.62 \times 10^{-2}$ | $7.91 \times 10^{-1}$ | 8.22 |
| Profile Curvature | $-1.83 \times 10^{-4}$ | $9.47 \times 10^{-3}$ | $6.37 \times 10^{-2}$ | $-7.37 \times 10^{-2}$ | $-5.31 \times 10^{-1}$ | −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 | $4.15 \times 10^{-1}$ | $4.43 \times 10^{-1}$ | 3.66 | $2.02 \times 10^{-2}$ | 2.96 | 1.07 |
| Elevation | 369 | 3.34 | 377 | 359 | −0.184 | 0.9 |

Source: Univariate summary

*0.3.2. Spatial analysis*

Soil colour and geochemical composition varied across both sites. In the agricultural field, all 15 soil color and geochemical properties exhibited spatial autocorrelation with most properties demonstrating a strong spatial dependency (Table 3). Some of the soil properties presented a pattern that roughly matches (e.g., Rb, Cs) or mirrors (e.g., Ca, Sr) the overall topography of the site with a gradation between the highest point in the south-west corner towards the lowest points in the north-east (Figure 2). Other properties appear to have more localized highs and low concentrations/values (e.g., *c*, h*). The geochemical concentrations of Ca and Rb had the largest range values and as result displayed a less patchy distribution across the site. The nugget (Co) was small for all soil properties (<1.5), and Sr had an exceptionally large sill value (900).

At the forested site, the geochemical concentrations of Fe and Rb, along with the color properties *a*, b*, c*,* and *x* showed no spatial autocorrelation and were excluded from further analysis and four and five properties exhibiting strong and moderate spatial dependency, respectively (Table 3). In comparison to the agricultural site, the soil properties at the forested site displayed a more moderate spatial dependency. The nugget (Co) was generally small for most soil properties (<2) with the exception of La and Ni. The range values were similar across the different soil properties and fell between 176 and 298 m. Overall, the influence of the channel and floodplain environment can be seen in the pattern of the nine soil properties (Figure 3).

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 | 0.0 | 0.0 | 91 | 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 |

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) | $r^2$ |
|---|---|---|---|---|---|---|
| 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 |

[1] Models are all isotropic.

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

Source: Semivariograms

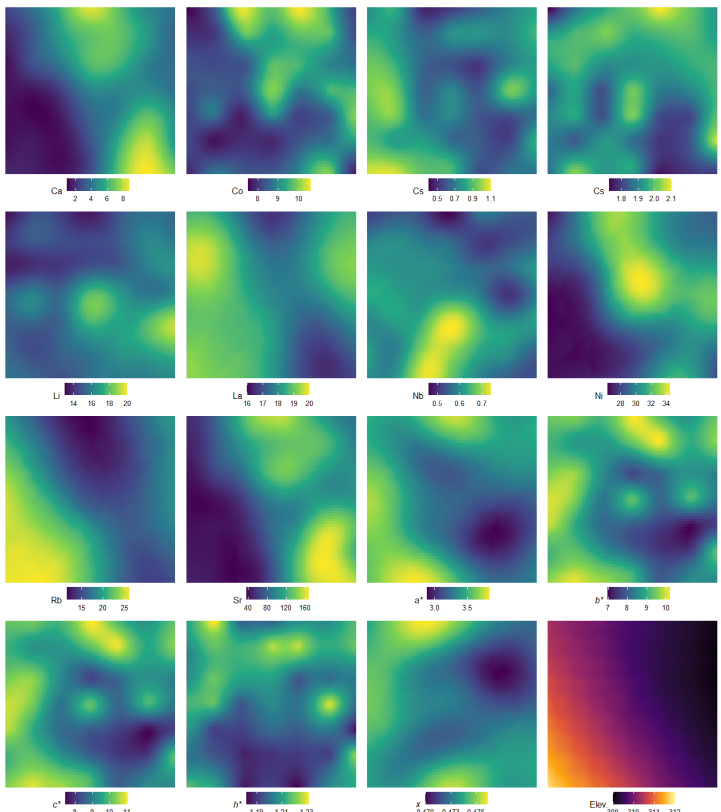

Figure 2: Kriged maps of select colour and geochemical properties and elevation across the agricultural site.

Source: Soil property mapping

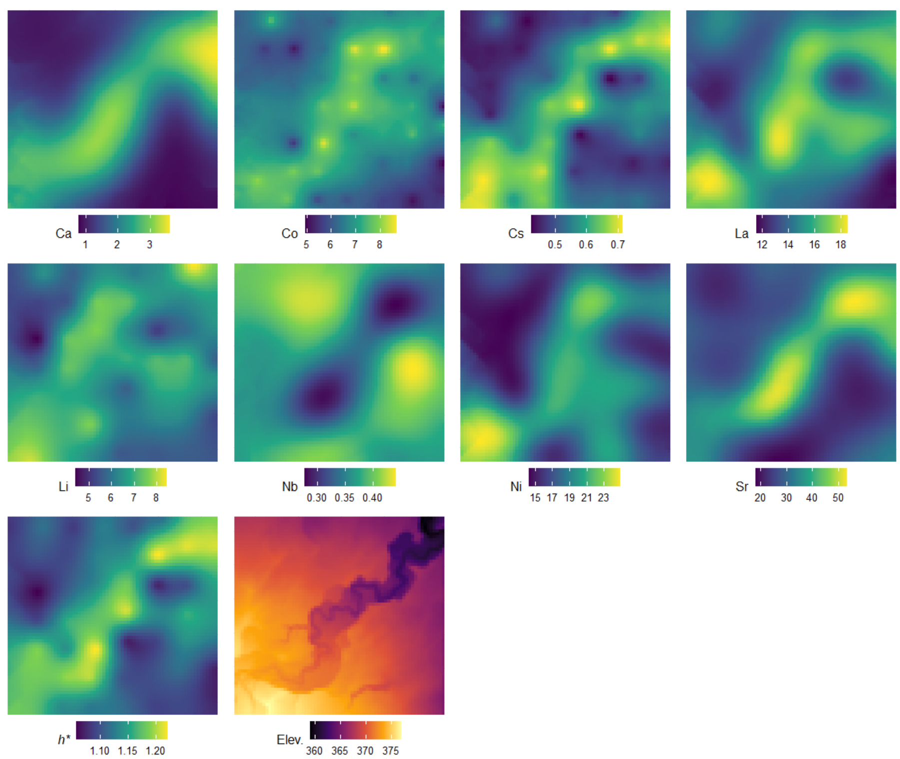


Figure 3: Kriged map of select colour and geochemical properties and elevation across the forested site.

Source: Soil property mapping

Across both sites, there was a significant ($p < 0.05$) correlation between the selected soil properties and the terrain attributes, with the exception of the plan and profile curvature attributes (**?@supptab-correlation2**). The elevation attribute generally had higher correlation coefficients; however, the direction and strength of the correlation did vary between both site and soil property. Overall, the random forest regression models exhibited relatively strong predictive performance, with the models better performing at the agricultural site compared to the forested site (Table 4). With the exception of the Ni concentration and *x* colour values at the agricultural site, elevation was consistently the terrain attribute that provided the greatest predictive power (Figure 4). SAGA Wetness and relative slope position were generally the second and third most informative terrain attributes. Plan curvature was consistently ranked least important predictive terrain attribute.

Table 4: Model summary and performance statistics for the random forest regression using the training, validation, test and original (non-interpolated) data sets.

| Property | Training | | Validation | | Test | | Original | |
|---|---|---|---|---|---|---|---|---|
| | % Variance[1] | MSE[2] | % Variance[1] | MSE[2] | $R^2$ | MSE[2] | $R^2$ | MSE[2] |
| | | | | Agriculture | | | | |
| Ca | 91.6 | 0.37 | 91.8 | 0.36 | 0.91 | 0.38 | 0.95 | 0.23 |
| Co | 79.8 | 0.09 | 82.5 | 0.08 | 0.80 | 0.08 | 0.88 | 0.08 |
| Cs | 85.7 | 0.00 | 86.4 | 0.00 | 0.85 | 0.00 | 0.92 | 0.00 |
| Fe | 69.6 | 0.00 | 70.9 | 0.00 | 0.69 | 0.00 | 0.83 | 0.00 |
| Li | 59.3 | 0.54 | 59.8 | 0.53 | 0.64 | 0.51 | 0.88 | 0.24 |
| La | 93.0 | 0.05 | 93.1 | 0.04 | 0.93 | 0.05 | 0.96 | 0.03 |
| Nb | 57.3 | 0.00 | 59.1 | 0.00 | 0.55 | 0.00 | 0.71 | 0.00 |
| Ni | 93.1 | 0.34 | 93.7 | 0.33 | 0.93 | 0.34 | 0.95 | 0.25 |
| Rb | 95.3 | 0.73 | 96.1 | 0.64 | 0.95 | 0.79 | 0.98 | 0.39 |
| Sr | 93.5 | 97.22 | 93.6 | 93.97 | 0.93 | 105.59 | 0.97 | 44.77 |
| a* | 85.0 | 0.01 | 86.9 | 0.01 | 0.85 | 0.01 | 0.91 | 0.00 |
| b* | 72.5 | 0.14 | 75.3 | 0.12 | 0.72 | 0.15 | 0.89 | 0.09 |
| c* | 73.2 | 0.15 | 75.9 | 0.14 | 0.73 | 0.17 | 0.89 | 0.10 |
| h* | 58.3 | 0.00 | 58.6 | 0.00 | 0.56 | 0.00 | 0.73 | 0.00 |
| x | 73.3 | 0.00 | 73.6 | 0.00 | 0.69 | 0.00 | 0.82 | 0.00 |
| | | | | Forest | | | | |
| Co | 39.1 | 0.24 | 42.9 | 0.23 | 0.48 | 0.21 | 0.77 | 0.29 |
| Cs | 64.1 | 0.00 | 67.1 | 0.00 | 0.66 | 0.00 | 0.86 | 0.00 |
| Li | 41.3 | 0.28 | 42.0 | 0.28 | 0.46 | 0.28 | 0.66 | 0.26 |
| La | 43.3 | 1.40 | 47.5 | 1.32 | 0.48 | 1.23 | 0.78 | 0.60 |
| Nb | 55.0 | 0.00 | 55.9 | 0.00 | 0.58 | 0.00 | 0.84 | 0.00 |
| Sr | 59.4 | 29.43 | 59.1 | 29.66 | 0.59 | 29.25 | 0.82 | 13.88 |
| h* | 58.8 | 0.00 | 60.3 | 0.00 | 0.62 | 0.00 | 0.86 | 0.00 |

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

Source: Random Forest summary

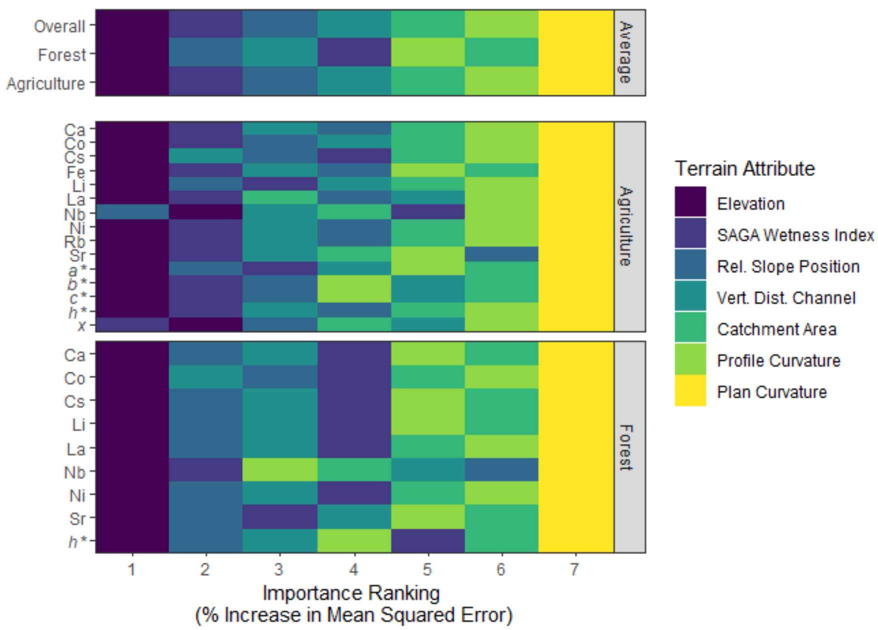


Figure 4: Heat map of the Random Forest regression results showing the ranking of the importance of terrain attributes (based on % increase in Mean Squared Error) in explaining the spatial variabilty of selected colour and geochemical properties within the agricultural and forested sites. Top panel shows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

Source: Random Forest summary

### *0.4. Discussion*

#### *0.4.1. Variability of soil properties*

Variability in soil geochemical properties have been studied at a range of scales including continental [41], regional [42], watershed [43], hillslope/catena, and farm field [44]. The objectives of these studies included addressing issues of pollution/contamination, providing benchmark/baseline information, investigating pedological and weathering properties and processes, and soil surveying and mapping [45]. Similarly, variability in soil colour, typically using the Munsell colour system, is a commonly reported diagnostic feature used in soil classification and ranges in spatial scale from reconnaissance to detailed soil surveys and maps. For sediment fingerprinting studies, these types of studies are often too site-specific or focus on a smaller subset of soil properties to effectively guide sample design to ensure the desired confidence is met characterizing sources of sediment.

Data distributions in soil science commonly exhibit a positively skewed distribution. This is likely due to several factors including that data of this nature are a semi-bounded distribution, with a lower bound of zero and no upper bound. Hot spots of soil processes, local variations in soil forming factors, and soil/land management practices can also lead to more extreme values [e.g., 46]. In many cases the cumulative effects of these processes, factors, and practices are multiplicative (i.e., interact) and not linearly additive, resulting in a skewed data distribution. Lastly, the distribution of data will also be a product of the scale of observation, number of samples, and sampling design.

Soil colour properties exhibited a near-normal distribution with a low CV which is consistent with claims that soil hue and value (Munsell colour system) have a low CV [47]. These data distribution properties are ideal for statistical and environmental modeling as it typically meets the model assumptions with out requiring transformations. For example, in sediment source fingerprinting, soil properties (i.e., fingerprints) are considered more reliable and robust for use in unmixing models when they show large differences between sources and low variability within each source. Additionally, most mixing models assume fingerprint data are normally distributed. [7] demonstrated that soil colour coefficients a*, b*, c*, h*, and x provided good discrimination between the agricultural and forested sites, and the low CV and skewness values reported in Table 1 makes these colour properties ideal fingerprints for sediment source apportionment.

The geochemical properties were more variable and skewed as compared to the soil colour properties. For many trace elements, concentrations are strongly correlated with the proportion of fine-grained material (<2 µm), due to its high specific surface area and enhanced chemical reactivity [48]. However, in this study the sand-size (>63 µm) material was removed prior to analysis to reduce the effects of grain-size on concentration. This likely resulted in lower variability and less extreme concentrations as compared to other studies that focus on bulk soil samples (<2 mm). In particular, the forested site exhibited a greater amount of variability which is likely due to the more complex topography and geomorphic setting. The floodplain within the forested site is likely accumulating shale-rich material derived from the Manitoba Escarpment which is enriched in trace metals [49]. This creates a zone of high concentrations relative to upland areas Figure 3. The forested site also had a higher and much more variable soil organic matter content ($\bar{x}$ = 8.5 %, CV = 51.9 %) as compared to the agricultural site ($\bar{x}$ = 11.6 %, CV = 16.1 %), which similarly to the grain size distribution, can influence the concentration of many major and trace elements [48]. These results provide evidence that both land use and landscape complexity both play a role in driving soil property variability.

*0.4.2. Spatial distribution*

The difference in the number of soil properties and the magnitude of the spatial auto correlation between the two sites can be used in designing an effect sampling campaign. The agricultural site, which has a simpler topography and a higher degree of spatial autocorrelation, the range values can be used to guide the distance between sampling points and a grid-style sampling regime may be an effective approach. In contrast, the forested site, which has a more complex geomorphic setting and a lower degree of spatial autocorrelation, a stratified sampling design may be the better approach. For example, at the forested site the stratas could include near-stream and hillslope environments. In situations where the soil properties of interest are not known or selected *a priori* (e.g., sediment fingerprinting) the differences in their spatial autocorrelation are difficult to accommodate in the sampling design. A sampling grid with irregular spacing, including spacing less than 100m, would have provided information on the spatial autocorrelation over shorter distances and reduced the uncertainty in the interpolation of soil properties [50].

Mapping the soil properties that have a moderate to high spatial dependence can provide information on underlying soil forming processes and properties. At both sites, to some extent, the patterns appear to reflect the topography of the sites suggesting that geomorphic and hydrologic processes and properties are likely driving the observed patterns. Identifying patterns and understanding the underlying process and properties that drive these patterns are important considerations when designing as soil sampling campaign to successfully meet study objectives, including characterizing soil properties of a field site. In a related context, Koiter [51] discussed the issues surrounding the use of a statistical only approach to selecting fingerprints and that consideration of how fingerprints have developed improves the robustness of the sediment fingerprinting approach. However, local information on the spatial distribution of geochemical and colour properties at field scales ($< 1$ km$^2$) is often unavailable, and the processes driving these patterns are also not well documented or studied. When such information does exist, it typically focuses on agronomically important properties [e.g., 52] or is used for soil classification [e.g., 53]. These datasets usually include geochemical properties such as nitrogen (N), phosphorus (P), potassium (K), sulfur (S), calcium (Ca), magnesium (Mg), sodium (Na), iron (Fe), aluminum (Al), nitrate ($NO$), carbonate ($CO^{2}$), bicarbonate ($HCO$), chloride ($Cl$), and sulfate ($SO^{2}$). They may also include colour characteristics, such as Munsell hue, value, and chroma, as well as other soil properties like texture, organic matter content, and pH. The lack of information on the wide range of soil properties means the researchers are relying on other data, most often elevation, for informing sampling designs.

### *0.4.3. Terrain attributes and soil properties*

Both the correlation analysis and random forest regression identified elevation as the most influential terrain attribute, followed by the SAGA Wetness Index and relative slope position, in explaining most of the observed variation in soil geochemical and colour properties. This is consistent with the findings of Mashalaba [54] who also found that similar terrain attributes were important in predicting a range of other soil properties including texture, bulk density, and hydaulic conductivity. These attributes likely emerged as the most important factor in explaining the observed variability as they are strongly linked to a range of geomorphic and hydrologic process and conditions [[55]; [56]]. For example, in eroded landscapes in the Prairie region of Canada, Ca concentrations have been found to be higher in upper slope positions from erosion and subsquent exposure of high-carbonate subsoil [57]. In contrast, higher Ca concentrations have been noted in lower slope and depressional areas due to higher solubility of many Ca-minerals (e.g., $CaCO_3$) and the subsequent downslope transport in solution and reduced leaching losses in these accumulation zones. Landscape position can also have a strong influence on pedogenic process; for example, the translocation of Fe and clay down the soil profile is a diagnostic criteria used in classifying soils [58]. Soil colour also tends to change in a predictable manner in relation to local relief. Tillage and water erosion results in the net loss of darker organic-rich topsoil from upper slope positions resulting in the exposure of the lighter subsoil [57]. Moisture availability is also greater in the lower slope and depressional areas resulting in increased organic matter production resulting in darker organic-rich topsoil as compared to the upper slope positions. There is also evidence that suggests that soil texture varies with elevation and slope position, with coarser material on upper slopes and finer material accumulating in lower positions [18, 59]. Given the strong correlation of organic matter and texture with soil geochemistry [48] and colour [60], these properties may also help explain the observed spatial patterns .

The relative importance of terrain attributes in explaining soil property variability differs both among soil properties and between sites. The land use and the overall geomorphic complexity differences between the two study sites are likely interacting with terrain attributes and influencing the patterns of soil properties and modifying the nature of terrain attribute and soil property relationship. This suggests that these relationships observed in this study may not be broadly generalized. Similarly, information on how terrain attributes influence the spatial distribution of many trace elements and soil colour, beyond the Munsell system, at the field scale is limited in the scientific literature. Additional variables including climate and large-scale landscape features will also influence the observed patterns of soil properties. As a result, using terrain attributes to guide soil sampling or interpret spatial patterns of many soil properties remains challenging.

The impact of sampling design at the field scale on the characterization of soil properties can be substantial [7], which in turn can affect the interpretation of data, modeling results, and the conclusion drawn. High-quality LiDAR data or digital elevation models (DEMs) are increasingly openly available in many regions and can be used to create detailed terrain attribute maps. By incorporating terrain attributes into the sampling framework, researchers can ensure that key geomorphic and hydrologic gradients are adequately represented. Ultimately, integrating terrain analysis into sediment source fingerprinting is promising not only as a mechanism to improve the quality of source characterization but to also better link source material to downstream sediment.

### *0.5. Conclusions*

Understanding the spatial variability and distribution of soil geochemical and colour properties at a field-scale is important for agricultural and environmental research, monitoring, modeling, and management practices. This study conducted both univariate and spatial analyses of a suite of soil geochemical and colour properties at two sites with contrasting land uses. The agricultural site, characterized by gently sloping topography, exhibited lower coefficients of variation, approximately normal data distributions, and moderate to strong spatial autocorrelation across most measured properties. In contrast, the forested site featured more geomorphologically complex terrain, with greater variability in soil properties, data distributions that more frequently deviated from normality, and fewer properties exhibiting spatial autocorrelation. Despite these differences, random forest regression consistently identified elevation, the SAGA Wetness Index, and relative slope position as the three most important terrain attributes explaining the observed variability.

These findings underscore the role of topographic controls on many soil property distributions, regardless of land use. However, the strength and direction of the relationship between terrain attributes and soil property results were inconsistent between both site and soil property. While the study was limited to two sites, the approach demonstrates the value of integrating tools like random forest regression with spatial data to better understand soil-landscape relationships. Future research should expand to broader landscapes and incorporate additional biophysical variables to improve generalizability. Overall, this work highlights how terrain-driven spatial patterns can inform more targeted soil sampling, modeling, and land management strategies.

*Acknowledgments*

Special thanks and recognition for the field and technical support from A. Avila and the Riding Mountain National Park personnel.

*Statements and declarations*

*Funding*

This research was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant - From source to sink: Investigating the linkages between sources of sediment and downstream water quality in Canadian watersheds - awarded to AJK (RGPIN-2019-05273).

*Competing interests*

The authors have no competing interests to declare that are relevant to the content of this article.

*Data and code availability*

Data and source code for analysis and manuscript available on GitHub: https://github.com/alex-koiter/sampling-design-manuscript

*Author contributions*

**M Luna Miño** Methodology; Investigation; Data curation; Formal analysis; Writing - Original Draft; Writing - Review & Editing

**A Koiter**: Conceptualization; Funding acquisition; Methodology; Investigation; Data curation; Formal analysis; Visualization; Writing - Original Draft; Writing – review and editing; Software; Project administration

**T Lychuk**: Methodology; Formal analysis; Writing - Review & Editing

**A Waddel** Methodology; Formal analysis; Writing - Review & Editing

**A Moulin** Methodology; Formal analysis; Writing - Review & Editing

*References*


[1] J. L. Starr, J. J. Meisinger, T. B. Parkin, Influence of sample size on chemical and physical soil measurements, Soil Science Society of America Journal 59 (3) (1995) 713–719, _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj1995.03615995005900030012x. doi:10.2136/sssaj1995.03615995005900030012x. URL https://onlinelibrary.wiley.com/doi/abs/10.2136/sssaj1995.03615995005900030012x

[2] S. K. Kariuki, H. Zhang, J. L. Schroder, T. Hanks, M. Payton, T. Morris, Spatial variability and soil sampling in a grazed pasture, Communications in Soil Science and Plant Analysis 40 (9-10) (2009) 1674–1687. URL http://www.tandfonline.com/doi/abs/10.1080/00103620902832089

[3] S. Pulley, I. Foster, A. L. Collins, The impact of catchment source group classification on the accuracy of sediment fingerprinting outputs, Journal of Environmental Management 194 (2017) 16–26. URL http://www.sciencedirect.com/science/article/pii/S0301479716302195

[4] P. Du, D. E. Walling, Fingerprinting surficial sediment sources: Exploring some potential problems associated with the spatial variability of source material properties, Journal of Environmental Management 194 (1) (2017) 4–15. doi:https://doi.org/10.1016/j.jenvman.2016.05.066. URL http://www.sciencedirect.com/science/article/pii/S0301479716303218

[5] S. Pulley, A. L. Collins, B. Van der Waal, Variability in the mineral magnetic properties of soils and sediments within a single field in the cape fold mountains, south africa: Implications for sediment source tracing, CATENA 163 (2018) 172–183. doi:10.1016/j.catena.2017.12.019. URL https://www.sciencedirect.com/science/article/pii/S0341816217304216

[6] A. L. Collins, E. Burak, P. Harris, S. Pulley, L. Cardenas, Q. Tang, Field scale temporal and spatial variability of 13c, 15n, tc and tn soil properties: Implications for sediment source tracing, Geoderma 333 (2019) 108–122. doi:10.1016/j.geoderma.2018.07.019.
URL https://www.sciencedirect.com/science/article/pii/S0016706118303185

[7] M. A. Luna Miño, A. J. Koiter, D. A. Lobb, Effect of sampling design on characterizing surface soil fingerprinting properties, Journal of Soils and Sediments 24 (5) (2024) 2180–2198. doi:10.1007/s11368-024-03805-x.
URL https://doi.org/10.1007/s11368-024-03805-x

[8] A. L. Collins, M. Blackwell, P. Boeckx, C.-A. Chivers, M. Emelko, O. Evrard, I. Foster, A. Gellis, H. Gholami, S. Granger, P. Harris, A. J. Horowitz, J. P. Laceby, N. Martinez-Carreras, J. Minella, L. Mol, K. Nosrati, S. Pulley, U. Silins, Y. J. da Silva, M. Stone, T. Tiecher, H. R. Upadhayay, Y. Zhang, Sediment source fingerprinting: benchmarking recent outputs, remaining challenges and emerging themes, Journal of Soils and Sediments 20 (12) (2020) 4160–4193. doi:10.1007/s11368-020-02755-4.
URL https://doi.org/10.1007/s11368-020-02755-4

[9] S. N. Wilkinson, J. M. Olley, T. Furuichi, J. Burton, A. E. Kinsey-Henderson, Sediment source tracing with stratified sampling and weightings based on spatial gradients in soil erosion, Journal of Soils and Sediments 15 (10) (2015) 2038–2051. doi:10.1007/s11368-015-1134-2.
URL http://proxy.library.unbc.ca:2118/article/10.1007/s11368-015-1134-2

[10] A. Haddadchi, M. Hicks, J. M. Olley, S. Singh, M. Srinivasan, Grid-based sediment tracing approach to determine sediment sources, Land Degradation and Development 30 (17) (2019) 2088–2106, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ldr.3407. doi:10.1002/ldr.3407.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ldr.3407

[11] A. L. Collins, D. E. Walling, G. Leeks, Source type ascription for fluvial suspended sediment based on a quantitative composite fingerprinting technique, Catena 29 (1) (1997) 1–27. doi:https://doi.org/10.1016/S0341-8162(96)00064-1.
URL http://www.sciencedirect.com/science/article/B6VCG-3SVY7CV-1/2/17d8e824e1b6c26e15fcd31608eb6d05

[12] A. Koiter, P. Owens, E. Petticrew, D. Lobb, The behavioural characteristics of sediment properties and their implications for sediment fingerprinting as an approach for identifying sediment sources in river basins, Earth-Science Reviews 125 (2013) 24–42. doi:https://doi.org/10.1016/j.earscirev.2013.05.009.

URL http://www.sciencedirect.com/science/article/pii/S0012825213001074

[13] J. C. Ritchie, E. E. C. Clebsch, W. K. Rudolph, Distribution of fallout and natural gamma radionuclides in litter, humus and surface mineral soil layers under natural vegetation in the great smoky mountains, north carolina-tennessee, Health Physics 18 (5) (1970) 479.
URL https://journals.lww.com/health-physics/Abstract/1970/05000/Distribution_of_Fallout_and_Natural_Gamma.3.aspx

[14] D. Vasu, S. K. Singh, N. Sahu, P. Tiwary, P. Chandran, V. P. Duraisami, V. Ramamurthy, M. Lalitha, B. Kalaiselvi, Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management, Soil and Tillage Research 169 (2017) 25–34. doi:10.1016/j.still.2017.01.006.
URL https://www.sciencedirect.com/science/article/pii/S0167198717300065

[15] R. A. Viscarra Rossel, B. Minasny, P. Roudier, A. B. McBratney, Colour space models for soil science, Geoderma 133 (3) (2006) 320–337.

[16] Y. Sun, W. Guo, D. C. Weindorf, F. Sun, S. Deb, G. Cao, J. Neupane, Z. Lin, A. Raihan, Field-scale spatial variability of soil calcium in a semi-arid region: Implications for soil erosion and site-specific management, PEDOSPHERE 31 (5) (2021) 705–714, num Pages: 10 Place: Beijing Publisher: Science Press Web of Science ID: WOS:000671191300005. doi:10.1016/S1002-0160(21)60019-X.
URL https://www.webofscience.com/wos/woscc/summary/a5c79ace-f964-4a0b-90b2-d726580a394c-fae59ae2/relevance/1

[17] D. E. Beaudette, R. A. Dahlgren, A. T. O'Geen, Terrain-shape indices for modeling soil moisture dynamics, Soil Science Society of America Journal 77 (5) (2013) 1696–1710, _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2013.02.0048. doi:10.2136/sssaj2013.02.0048.
URL https://onlinelibrary.wiley.com/doi/abs/10.2136/sssaj2013.02.0048

[18] V. Kokulan, O. Akinremi, A. P. Moulin, D. Kumaragamage, Importance of terrain attributes in relation to the spatial distribution of soil properties at the micro scale: a case study, Canadian Journal of Soil Science 98 (2) (2018) 292–305, publisher: NRC Research Press. doi:10.1139/cjss-2017-0128.
URL https://cdnsciencepub.com/doi/10.1139/cjss-2017-0128

[19] D. J. Brown, M. K. Clayton, K. McSweeney, Potential terrain controls on soil color, texture contrast and grain-size deposition for the original catena landscape in uganda, Geoderma 122 (1) (2004) 51–72. doi:10.1016/j.geoderma.2003.12.004.
URL https://www.sciencedirect.com/science/article/pii/S0016706104000035

[20] S. Zhang, Y. Huang, C. Shen, H. Ye, Y. Du, Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information, Geoderma 171-172 (2012) 35–43. doi:10.1016/j.geoderma.2011.07.012.
URL https://www.sciencedirect.com/science/article/pii/S001670611100214X

[21] B. P. Umali, D. P. Oliver, S. Forrester, D. J. Chittleborough, J. L. Hutson, R. S. Kookana, B. Ostendorf, The effect of terrain and management on the spatial variability of soil properties in an apple orchard, CATENA 93 (2012) 38–48. doi:10.1016/j.catena.2012.01.010.
URL https://www.sciencedirect.com/science/article/pii/S0341816212000136

[22] F. R. D. Lima, P. Pereira, I. C. F. Vasques, E. C. Silva Junior, M. Mancini, J. R. Oliveira, M. T. A. Prianti, C. Windmöller, D. C. Weindorf, N. Curi, B. T. Ribeiro, J. Richardson, J. Marques, L. R. G. Guilherme, Predictive modeling of total hg background concentration in soils of the amazon rainforest biome with support of proximal sensors and auxiliary variables, Journal of South American Earth Sciences 129 (2023) 104510. doi:10.1016/j.jsames.2023.104510.
URL https://www.sciencedirect.com/science/article/pii/S0895981123003218

[23] W. Ehrlich, L. Pratt, F. Leclaire, Reconnaissance soil survey of west-lake map sheet area, Tech. rep. (1958).

[24] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, E. F. Wood, Present and future köppen-geiger climate classification maps at 1-km resolution, Scientific Data 5 (1) (2018) 180214, number: 1 Publisher: Nature Publishing Group. doi:10.1038/sdata.2018.214.
URL https://www.nature.com/articles/sdata2018214

[25] Environment, C. C. Canada, Canadian climate normals, meteorology, Weather (2024).
URL https://climate.weather.gc.ca/climate_normals/index_e.html

[26] G. MacKay, A quantitative study of geomorphology of the wilson creek watershed, manitoba, Msc thesis, University of Manitoba, Winnipeg, MB, m.Sc., Civil Engineering (1970).

[27] J. P. Laceby, O. Evrard, H. G. Smith, W. H. Blake, J. M. Olley, J. P. G. Minella, P. N. Owens, The challenges and opportunities of addressing particle size effects in sediment source fingerprinting: A review, Earth-Science Reviews 169 (2017) 85–103. doi:https://doi.org/10.1016/j.earscirev.2017.04.009.
URL http://proxy.library.unbc.ca:2111/science/article/pii/S0012825216304548

[28] M. Boudreault, A. J. Koiter, D. A. Lobb, P. N. Owens, K. Liu, G. Benoy, S. Danielescu, S. Li, Using colour, shape and radionuclide sediment fingerprints to identify sources of sediment in an agricultural watershed in atlantic canada, Canadian Water Resources Journal 43 (3) (2018) 347–365. doi:https://doi.org/10.1080/07011784.2018.1451781.

[29] A. Koiter, Colour analysis r scriptsDOI: 10.5281/zenodo.5123327 (07 2021). doi:10.5281/zenodo.5123327.
URL https://github.com/alex-koiter/Colour-analysis

[30] Esri, ArcGIS Pro, Esri, 2024.
URL https://www.esri.ca/

[31] N. R. Canada, High resolution digital elevation model mosaic (hrdem mosaic) - canelevation series (2024).
URL https://open.canada.ca/data/en/dataset/0fe65119-e96e-4a57-8bfe-9d9245fba06b

[32] Agisoft, Agisoft metashape: Installer (2021).
URL https://www.agisoft.com/downloads/installer/

[33] O. Conrad, B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, J. Böhner, System for automated geoscientific analyses (saga) v. 2.1.4, Geoscientific Model Development 8 (7) (2015) 1991–2007, publisher: Copernicus GmbH. doi:10.5194/gmd-8-1991-2015.
URL https://gmd.copernicus.org/articles/8/1991/2015/

[34] R. C. Team, R: A language and environment for statistical computing (2024).
URL http://www.R-project.org

[35] RStudio, Rstudio: Integrated development environment for r (2024).
URL http://www.rstudio.org/

[36] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York NY U.S.A, 2016.

[37] R. J. Hijmans, terra: Spatial Data Analysis, r package version 1.7-78 (2024).
URL https://CRAN.R-project.org/package=terra

[38] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (3) (2002) 18–22.
URL https://CRAN.R-project.org/doc/Rnews/

[39] B. Naimi, N. a.s. Hamm, T. A. Groen, A. K. Skidmore, A. G. Toxopeus, Where is positional uncertainty a problem for species distribution modelling, Ecography 37 (2014) 191–203. doi:10.1111/j.1600-0587.2013.00205.x.

[40] Kuhn, Max, Building predictive models in r using the caret package, Journal of Statistical Software 28 (5) (2008) 1–26. doi:10.18637/jss.v028.i05.
URL https://www.jstatsoft.org/index.php/jss/article/view/v028i05

[41] L. J. Drew, E. C. Grunsky, D. M. Sutphin, L. G. Woodruff, Multivariate analysis of the geochemistry and mineralogy of soils along two continental-scale transects in north america, Science of The Total Environment 409 (1) (2010) 218–227. doi:10.1016/j.scitotenv.2010.08.004.
URL https://www.sciencedirect.com/science/article/pii/S0048969710008375

[42] M. Rattenbury, M. , B. , T. , K. Rogers, Geochemical baseline soil surveys for understanding element and isotope variation across new zealand, New Zealand Journal of Agricultural Research 61 (3) (2018) 347–357, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00288233.2018.1426616. doi:10.1080/00288233.2018.1426616.
URL https://doi.org/10.1080/00288233.2018.1426616

[43] N. Nanos, J. Rodríguez Martín, Multiscale analysis of heavy metal contents in soils: Spatial variability in the duero river basin (spain), Geoderma 189-190 (2012) 554–562. doi:10.1016/j.geoderma.2012.06.006.
URL https://www.sciencedirect.com/science/article/pii/S0016706112002418

[44] Y. Sun, W. Guo, D. C. Weindorf, F. Sun, S. Deb, G. Cao, J. Neupane, Z. Lin, A. Raihan, Field-scale spatial variability of soil calcium in a semi-arid region: Implications for soil erosion and site-specific management, PEDOSPHERE 31 (5) (2021) 705–714, num Pages: 10 Place: Beijing Publisher: Science Press Web of Science ID: WOS:000671191300005. doi:10.1016/S1002-0160(21)60019-X.
URL https://www.webofscience.com/wos/woscc/summary/a5c79ace-f964-4a0b-90b2-d726580a394c-fae59ae2/relevance/1

[45] M. A. Wilson, R. Burt, S. J. Indorante, A. B. Jenkins, J. V. Chiaretti, M. G. Ulmer, J. M. Scheyer, Geochemistry in the modern soil survey program, Environmental Monitoring and Assessment 139 (1-3) (2008) 151–171. doi:10.1007/s10661-007-9822-z.
URL http://link.springer.com/10.1007/s10661-007-9822-z

[46] P. Vidon, C. Allan, D. Burns, T. P. Duval, N. Gurwick, S. Inamdar, R. Lowrance, J. Okay, D. Scott, S. Sebestyen, Hot spots and hot moments in riparian zones: Potential for improved water quality management, Journal of the American Water Resources Association 46 (2) (2010) 278–298. doi:10.1111/j.1752-1688.2010.00420.x.
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2010.00420.x/pdf

[47] D. Pennock, T. Yates, J. Braidek, Soil sampling designs, 2nd Edition, CRC Press, Boca Raton, FL, USA, 2008, Ch. 1, pp. 1–14.

[48] A. J. Horowitz, A primer on sediment-trace element chemistry, 2nd Edition, Lewis Publishers, Chelsea, Michigan, USA, 1991.

[49] M. Nicolas, J. Bamburak, Geochemistry and mineralogy of cretaceous shale, southwestern manitoba (parts of nts 62f, g, j, k, n, 63c): phase 2 results, Tech. rep., Manitoba Innovation, Energy and Mines, Manitoba Geological Survey (2011).

[50] R. M. Lark, B. P. Marchant, How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters?, Geoderma 319 (2018) 89–99. doi:10.1016/j.geoderma.2017.12.022.
URL https://www.sciencedirect.com/science/article/pii/S0016706117309187

[51] A. Koiter, P. Owens, E. Petticrew, D. Lobb, The behavioural characteristics of sediment properties and their implications for sediment fingerprinting as an approach for identifying sediment sources in river basins, Earth-Science Reviews 125 (2013) 24–42. doi:https://doi.org/10.1016/j.earscirev.2013.05.009.
URL http://www.sciencedirect.com/science/article/pii/S0012825213001074

[52] M. Mzuku, R. Khosla, R. Reich, D. Inman, F. Smith, L. MacDonald, Spatial variability of measured soil properties across site-specific management zones, Soil Science Society of America Journal 69 (5) (2005) 1572–1579, wOS:000231817500026. doi:10.2136/sssaj2005.0062.

[53] S. C. W. Group, The Canadian System of Soil Classification, 3rd Edition, NRC Research Press, Ottawa, ON, Canada, 1998, google-Books-ID: 4aHVg18eDLYC.
URL https://sis.agr.gc.ca/cansis/taxa/cssc3/index.html

[54] L. Mashalaba, M. Galleguillos, O. Seguel, J. Poblete-Olivares, Predicting spatial variability of selected soil properties using digital soil mapping in a rainfed vineyard of central chile, Geoderma Regional 22 (2020) e00289. doi:10.1016/j.geodrs.2020.e00289.
URL https://www.sciencedirect.com/science/article/pii/S2352009420300389

[55] F. A. O. Mello, J. Demattê, R. Rizzo, D. C. d. Mello, R. R. Poppiel, N. Silvero, J. Safanelli, H. Bellinaso, B. R. Bonfatti, A. M. R. Gomez, G. P. B. Sousa, Complex hydrological knowledge to support digital soil mapping, Geoderma 409 (2022) 115638. doi:10.1016/j.geoderma.2021.115638.
URL https://www.sciencedirect.com/science/article/pii/S0016706121007187

[56] Z. Libohova, M. Mancini, H. E. Winzeler, Q. D. Read, N. Sun, D. Beaudette, C. Williams, J. Blackstock, S. Silva, N. Curi, K. Adhikari, A. Ashworth, J. O. Minai, P. R. Owens, Interpreting the spatial distribution of soil properties with a physically-based distributed hydrological model, Geoderma Regional 39 (2024) e00863. doi:10.1016/j.geodrs.2024.e00863.
URL https://www.sciencedirect.com/science/article/pii/S235200942400110X

[57] S. K. Papiernik, M. J. Lindstrom, J. A. Schumacher, A. Farenhorst, K. D. Stephens, T. E. Schumacher, D. A. Lobb, Variation in soil properties and crop yield across an eroded prairie landscape, Journal of Soil and Water Conservation 60 (6) (2005) 8.

[58] H. B. Stonehouse, R. J. St. Arnaud, Distribution of iron, clay and extractable iron and aluminum in some saskatchewan soils, Canadian Journal of Soil Science 51 (2) (1971) 283–292, publisher: NRC Research Press. doi:10.4141/cjss71-036.
URL https://cdnsciencepub.com/doi/10.4141/cjss71-036

[59] M. S. Cox, P. D. Gerard, M. C. Wardlaw, M. J. Abshire, Variability of selected soil properties and their relationships with soybean yield, Soil Science Society of America Journal 67 (4) (2003) 1296–1302, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2003.1296. doi:10.2136/sssaj2003.1296.
URL https://onlinelibrary.wiley.com/doi/abs/10.2136/sssaj2003.1296

[60] R. A. Viscarra Rossel, S. R. Cattle, A. Ortega, Y. Fouad, In situ measurements of soil colour, mineral composition and clay content by vis–nir spectroscopy, Geoderma 150 (3) (2009) 253–266. doi:10.1016/j.geoderma.2009.01.025.
URL http://www.sciencedirect.com/science/article/pii/S0016706109000408

*Supplemental figures*

*Supplemental tables*

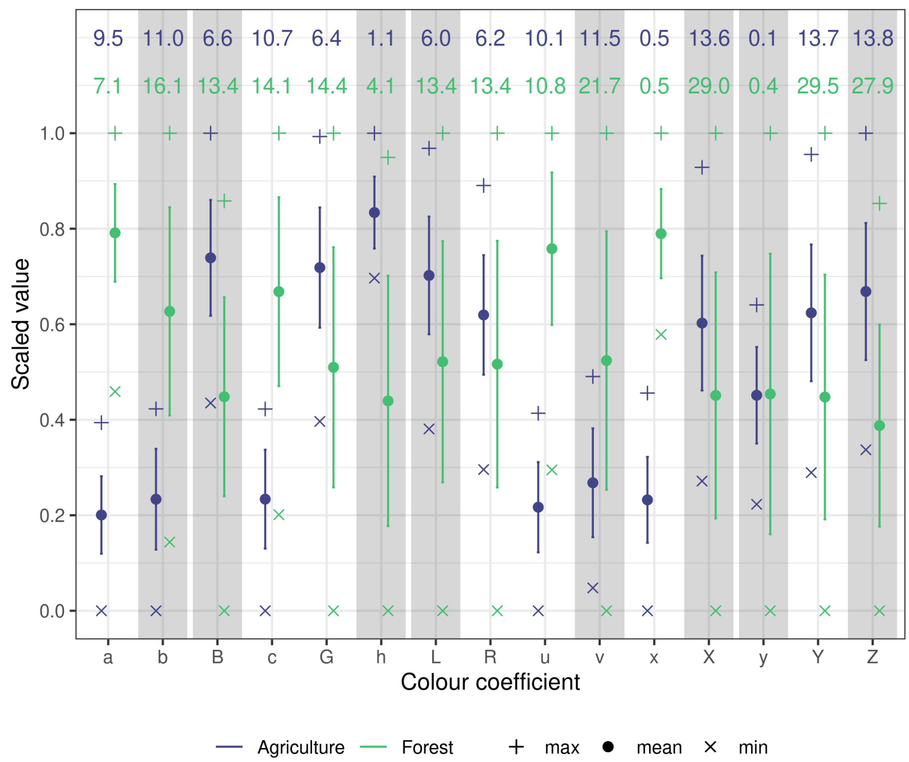


Figure S1: Summary statistics of all measured colour soil properties at both sites. Error bars represent 1SD and the numeric values indicate the CV.

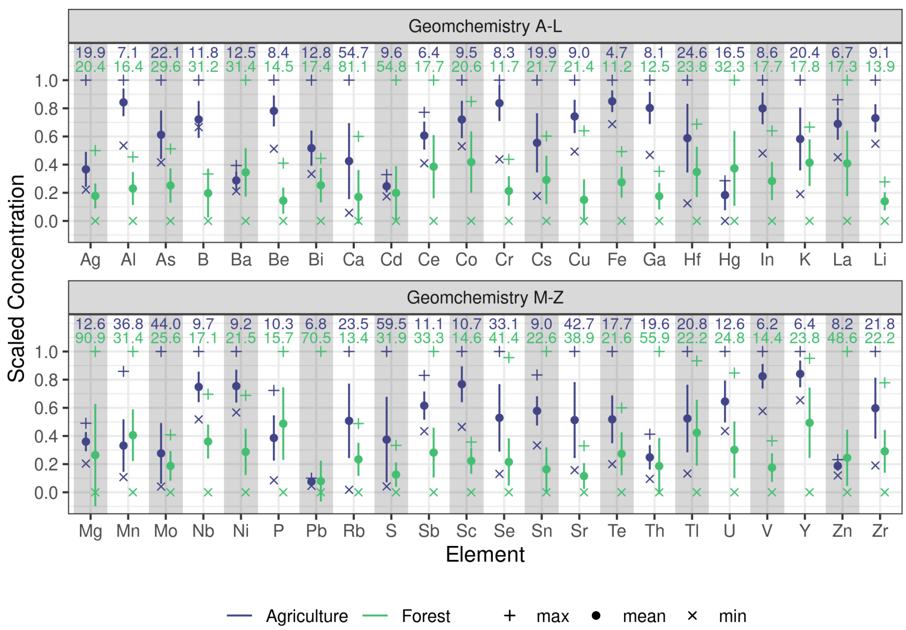


Figure S2: Summary statistics of all measured geochemical soil properties at both sites. Error bars represent 1SD and the numeric values indicate the CV.

| **Colour space model** | Parameter | Abbreviation |
|---|---|---|
| RGB | Red | R |
| RGB | Green | G |
| RGB | Blue | B |
| CIE xyY | Chromatic Coordinate x | x |
| CIE xyY | Chromatic Coordinate y | y |
| CIE xyY | Brightness | Y |
| CIE XYZ | Virtual component X | X |
| CIE XYZ | Virtual component Z | Z |
| CIE LAB | Metric lightness function | L |
| CIE LAB | Chromatic coordinate opponent red–green scales | $a*$ |
| CIE LAB | Chromatic coordinate opponent red–green scales | $b*$ |
| CIE LUV | Chromatic coordinate opponent blue–yellow scales | $u*$ |
| CIE LUV | Chromatic oordinate opponent red–green scales | $v*$ |
| CIE LCH | CIE hue | $c*$ |
| CIE LCH | CIE chroma | $h*$ |

Table S1: Description of spectral reflectance colour coefficients used as fingerprints. Reproduced from Boudreault et al. (2018)

| Terrain Attribute | Description |
|---|---|
| Elevation | Meters above sea level |
| Plan Curvature | Across slope curvature |
| Profile Curvature | Down slope curvature |
| SAGA Wetness Index | Similar to the ‘Topographic Wetness Index’ (TWI), but it is based on a modified catchment area calculation, which does not think of the flow as very thin film. As result it predicts for cells situated in valley floors with a small vertical distance to a channel a more realistic, higher potential soil moisture compared to the standard TWI calculation |
| Catchment Area | Area of upslope contributing area |
| Relative Slope Position | A value between 0 and 1 illustrating the position of a pixel within the landscape with values approaching 0 indicating streams to pits, and values approaching 1 indicating upper slope positions to peaks |
| Vertical Distance to Channel | The vertical distance to a channel network base level. The algorithm consists of two major steps:<br>1. Interpolation of a channel network base level elevation<br>2. Subtraction of this base level from the original elevations |

Table S2: Terrain attribute descriptions