Hbr380: A Spatial Model of Soil Distribution for the Hubbard Brook Experimental Forest

Users Guide and Methods

Introduction and Usage

This model of the distribution of soils follows a similar approach as Gillin et al. (2015a) and updates and expands that model of the soils of Watershed 3 to the entire Hubbard Brook Experimental Forest. Ordination analysis was used to evaluate the relationships between soil horizon presence and thickness and topographic metrics derived from a LiDAR digital elevation model. Soil units clustered in ordination space, allowing identification of topographic metrics most predictive of the presence of each soil unit and an assessment of covariance among the metrics. Observations of soil profiles and water table fluctuations used in this analysis are available in separate data publications (Bailey 2024; Bailey et al. 2024).

Users should consider their specific objectives when choosing how to use model outputs. The most likely soil unit raster of the entire forest is useful for evaluating broad spatial relationships with other datasets collected at a similar scale. Soil units on the most likely soil unit raster (modelout2024-02-21.tif) are coded by number as follows: 1=Bh podzol, 2=Bhs podzol, 3=Bimodal podzol, 4=E podzol, 5=Histosol, 6=Inceptisol, 7= O podzol, 8=Typical podzol. It is important to remember that the most likely soil unit at a given location will vary a lot in its probability of occurrence.

For site specific questions, it is always best to have site specific soil observations rather than to rely on a model. If using the model to predict soil properties at the point, or at the scale of a small plot, it is best to look at the probabilities of all the soil units for a given site and consider all possible soil units that might occur at that location. Particularly at the scale of a small plot, unless the plot was specifically chosen for uniform soil conditions, most plots will have a mix of two or more of the soil units. Looking at the several most likely soil units predicted by the model can provide an indication of the range of soil variability to be expected. Raster files for each soil unit depict probability on a scale from 0=no chance of occurrence to 1=absolute certainty of occurrence.

Soil units

Eight soil units were defined, following a hydropedologic approach, based on relationships between soil genetic horizon presence and thickness and the frequency and depth of groundwater fluctuations (Bailey et al. 2014; Gannon et al. 2014). Most of these soils, typical of uplands of northern New England, would generally be classified following USDA soil taxonomy as Spodosols. However, due to the effects of steep slopes and common shallow to exposed bedrock, transient water tables in the shallow soil zone generate lateral water fluxes that promote lateral soil development, i.e. illuvial processes occurring downslope of eluviation rather than vertically within a single pedon (Bailey et al. 2014; Bourgault et al. 2015, 2017; Gannon et al. 2015; Bailey et al. 2019). At the pedon scale, using USDA soil taxonomy, soils influenced by shallow transient groundwater fluxes may classify as Entisols, Inceptisols, Spodosols, or Histosols (Villars et al. 2015). However, in the instance of lateral podzolization, we suggest that standard soil taxonomy applied at the pedon scale is misleading. Thus, we use functionally defined soil units labelled as types of podzols – to emphasize the central importance of podzolization as a soil forming factor and to emphasize that these soil units are not defined according to standard soil taxonomy used in the United States. The eight soil units, listed in general order from the top to the base of an idealized hillslope, also called a catena or toposequence, are described as follows:

O podzols are found along hilltops, ridgelines, and other convex slope positions with shallow bedrock. They consist of forest floor horizons (Oi, Oe, and Oa) directly on bedrock. If a mineral soil horizon is present, it is generally an E horizon and is thinner than the overlying Oa horizon. Soil moisture is highly variable. These soils may dry considerably during rain-free periods of the growing season. However, they tend to saturate with water during rainstorms, with groundwater receding immediately upon cessation of rain. During periods of saturation, groundwater has low pH and high concentrations of dissolved organic carbon (DOC), aluminum (Al), and iron (Fe) (Bailey et al. 2019). Illuviation of these solutes occurs downslope.

E podzols are associated with O podzols in shallow bedrock areas and tend to be below a break in slope, or in a small swale, particularly if it is surrounded by O podzols and exposed bedrock. These soils consist of forest floor horizons over an E horizon. The E horizon is thicker than the overlying Oa horizon and may be as much as 60 cm thick. If a B horizon (generally a Bhs horizon) is present between the E horizon and the bedrock surface, it is thinner than the E horizon. Groundwater saturation dynamics and chemistry are similar to O podzols.

Bhs podzols are immediately downslope of O and E podzols on landforms controlled by shallow and exposed bedrock. The Bhs horizon (Munsell color hue 7.5YR or redder, value and chroma both 3 or less) is the thickest mineral horizon and may be as much as a meter thick. E horizons tend to be absent or thin, typically less than 5 cm thick. These soils are considered to represent the zone of illuviation from upslope areas with exposed bedrock and O and E podzols. Groundwater saturation occurs during most rainstorm events, with the water table not reaching as high in the profile as in the O and E podzols, and with a slower recession, up to a few or several days. Groundwater pH is somewhat higher, and DOC, Al, and Fe concentrations are still high, although a bit lower than in the O and E podzols. Although illuviation is a dominant soil forming process in these soils, elevated DOC, Al, and Fe concentrations suggest that groundwater fluxes from Bhs podzols also support illuviation in soils further downslope.

Typical podzols generally consist of a thin forest floor and a thin or absent E horizon overlying a thicker spodic horizon. The spodic horizon generally consists of a thinner Bhs horizon overlying a thicker Bs horizon (value and/or chroma greater than 3). Typical podzols occur in areas with deeper glacial sediment deposits, generally in mid backslope positions and other areas with deeper soils on convex or planar slope shapes. Groundwater saturation is generally confined to the subsolum but may occasionally rise into the lower portion of the B horizon during larger rain events or the non-growing season. Groundwater chemistry has a higher pH and low concentrations of DOC, Al, and Fe, indicating that illuviation has occurred upslope and/or higher in the soil profile. This soil unit is most like a Spodosol as described in USDA soil taxonomy; most examples classify as Typic or Aquic Haplorthods.

Bimodal podzols have soil profiles that mostly resemble typical podzols except in the lower portion of the B horizon where redder colors, lower value and chroma, and higher carbon concentrations compared to the middle portion of the B horizon suggest a second zone of illuviation in the lower portion of the solum. These soils are found in areas of deeper glacial sediment deposits, typically with concave slopes, especially in footslope positions. Groundwater incursions in the lower B horizon are higher in the profile and more frequent than in the typical podzols, particularly during the dormant season. Groundwater chemistry is similar to that in typical podzols.

Bh podzols have dark profiles with relatively uniform colors where it can be difficult to distinguish between individual horizons. Soil colors tend to center around the 10YR hue, with value and chroma both 3 or less from the surface to the base of the B horizon. A horizons are common and tend to be thicker, or replace Oa horizons, suggesting more mixing of organic and mineral material, perhaps reflecting a more active or distinct decomposer community compared to soils upslope. The A horizon grades into a Bh horizon which is distinguished from a Bhs horizon by its tendency toward a less red hue (10YR). These soils are found in toeslope positions, on topographic benches (i.e., where the slope decreases or flattens along a hillslope profile), and in narrow corridors bordering headwater streams. Water table frequently saturates into the Bh horizon during rain events during the growing season, receding into the C horizon during dry periods, while remaining in the Bh horizon for more extended periods during the dormant season. The shallow groundwater fluctuations tend to be more muted than in upslope soil units. Groundwater chemistry is similar to bimodal and typical podzols.

Inceptisols are soils formed in deeper glacial sediment deposits that lack evidence of podzolization. E horizons are very thin, or more typically, absent. B horizons consist of cambic (Bw) or gleyed (Bg) materials. They occur in valley bottom areas and small depressions. Some are associated with groundwater seeps and may occur at a variety of slope positions and settings. The latter examples appear to be controlled by subsurface conditions conducive to groundwater discharge and are not well predicted by topographic metrics. Thus, a number of small groundwater seeps with Inceptisols are known around the valley that are not predicted by modelling presented here. These soils are permanently saturated into the B horizon, with the upper B and forest floor horizons being intermittently saturated during large storms and the dormant season. We have not sampled groundwater to determine solute chemistry in these soils.

Histosols consist of organic materials overlying deeper glacial sediment deposits. We use a minimum thickness of a 40 cm O horizon to distinguish Histosols from Inceptisols. If a B horizon is present below the organic layers, it is generally a Bg horizon. Histosols occur in valley bottom settings but may be found in depressions all the way up to the Hubbard Brook catchment ridgeline. These soils are permanently saturated near to the soil surface. We have not studied the groundwater chemistry of these soils.

<u>Soil field observations and model training</u>

A soils dataset consisting of 609 soil profile observations was used to train the soil model (Bailey 2024). Full profile description and sampling was made of 396 soil profiles accumulated from various studies dating from 1995 to 2021. Individual studies had varying objectives, methods of locating sample pits, and distribution across the HBEF. All pedon observations are based on profiles with description and sampling of genetic horizons to bedrock, or to ~ 1 – 2 m where bedrock was not encountered.

To supplement the full soil profile observations, additional reconnaissance sampling was conducted at a lower level of detail during the summers of 2020-2022. Abbreviated profile descriptions, emphasizing color and genetic horizon determination from small pits 40 – 80 cm deep were written at a sufficient level of detail and depth of observation to classify the profile into one of the eight soil units represented in the model. In 2020 and 2021, reconnaissance observations were along transects in areas where the finest gradients of soil units occur. These included the transition from upland to wetland soils in valley bottoms, and transitions between shallow soils associated with bedrock outcrops and deeper soils downslope. In 2022, reconnaissance pedons were collected at points broadly distributed across the entire HBEF, generated to increase sampling in areas where model predictions were weakest, following the approach of Stumpf et al. (2017).

All sampling locations were geo-referenced by collecting rover files with various models of Trimble GPS units, generally with a hurricane external antenna, and post processing with CORS base station files. A minimum of 200 positions per pit were collected, resulting in precision of 2 m or less horizontal precision.

Development of Topographic Metrics

Topographic metrics served as predictor variables and were calculated from a coarsened 5 m DEM using mean cell aggregation of a 1 m LiDAR-derived, hydro-enforced DEM (Fraser et al. 2022). Slope (%) was calculated using the maximum slope algorithm (Travis, 1975). Topographic position indices (TPI; Guisan et al., 1999) were created with a moving circular-window centered on a target cell using 20, 100, 200, 2000 m radius windows, taking the difference between the elevation of each cell and mean elevation. Topographic wetness index (TWId; Beven and Kirkby, 1979) was computed as the natural logarithm of the ratio of the upslope accumulated area (UAA) using a multiple triangular flow direction algorithm (Seibert and McGlynn, 2007) and the gradient associated with 5 m downslope distance (Hjerdt et al., 2004). Multiresolution index of valley bottom flatness (MRVBF) was computed to identify valley bottoms as areas that are flat and low relative to their surroundings based on combining slope and elevation percentile at different scales into a single multi-resolution index (Gallant and Dowling, 2003). Euclidean distance from bedrock (EDb) and bedrock-weighted upslope accumulated area (UAAb) were calculated using predictive approaches from Fraser et al. (2020) and Gillin et al (2015a). UAAb is a normalized ratio between the Seibert and McGlynn (2007) UAA weighted by grid cells identified as bedrock outcrops and shallow soils (Fraser et al., 2020) and an unweighted UAA, where bedrock outcrops were assigned an arbitrarily large weighting value (10^5). Non-bedrock outcrops were assigned zero in the weighting function of the SAGA recursive catchment area tool. UAAb thus varies between 0 and 1, where a value of 1 indicates an entire upslope area is comprised of bedrock outcrops and associated shallow soil and zero indicates that upslope bedrock outcrops and shallow soils have little influence on that grid cell.

Topographic metrics were all generated in R (R Development Core Team, 2020). Previous predictive modeling efforts at Hubbard Brook Experimental Forest (Fraser et al., 2020; Gillin et al., 2015b) computed topographic metrics in ArcGIS© and SAGA-GIS (Conrad et al., 2015), but this study leveraged an open-source platform to support accessibility and reproducibility. Topographic metrics were created from a 1 or 5 m spatial resolution DEM (Fraser et al., 2019, 2020) using raster (Hijmans 2023), Whitebox and qgisprocess (Dunnington et al, 2023) packages. Whitebox is an R package front end for Whitebox GAT (Lindsay, 2016) which is an advanced geospatial data analysis platform that can be used to perform common geographical information systems (GIS). qgisprocess is an R package providing access to geoprocessing and terrain analysis functions of SAGA-GIS (Conrad et al., 2015) and QGIS (QGIS.org, 2024).

Ensemble machine learning models and performance

Ensemble machine learning from the Landmap package in R (https://github.com/Envirometrix/landmap) was used to perform spatial prediction with topographic metrics as covariates. Landmap leverages the SuperLearner method to stack all single learners using an extension of the mlr package. This package automates the spatial prediction mapping procedure by applying ensemble machine learning and estimates the performance of multiple machine learning models with cross-validation. The models were calculated using the train function within the mlr package which performs the training, pre-processing, tuning, and performance assessment. In addition, Landmap package provides automation of several steps, such as deriving principal components, overlaying the observations and covariates, spatial blocking, and oblique coordinates (Møller et al., 2019).

The three single machine learners we used included a fast implementation of random forests (ranger), support vector machine (svm), and multinomial logistic regression (multinom). These are commonly used for predicting soil types with relatively small datasets and hence were selected for predicting soil unit distribution (Hengl and MacMillan, 2019). We implemented spatial blocking to ensure spatially clustered points did not contribute to model over-fitting (Roberts et al., 2017) as well as oblique coordinates to improve prediction performance (Møller et al., 2019). The predictive performance, using a 0.12 probability threshold for each soil unit, was evaluated using a standard error matrix with the confusion matrix function in the caret package, the coefficient of agreement (κ) and the 95% significance level ($p < 0.05$) of a topographic metric as an independent variable.

References Cited

Bailey, S.W. 2024. Hubbard Brook Experimental Forest: Soil Profiles (Pedons), 1995-2022 ver 2. Environmental Data Initiative. https://doi.org/10.6073/pasta/e413be5a20ef9cf5344c7d7855a71c70 (Accessed 2024-02-29).

Bailey, S.W., J.P. Gannon, K.J. McGuire, M.B. Green, and A.M. Pennino. 2024. Hubbard Brook Experimental Forest: Watershed 3 well water level recordings, 2007 - ongoing ver 4. Environmental Data Initiative. https://doi.org/10.6073/pasta/210b60a3d2f5ee2bb2635ee2fb33b637 (Accessed 2024-02-29).

Bailey, S. W., McGuire, K. J., Ross, D. S., Green, M. B., & Fraser, O. L. (2019). Mineral weathering and podzolization control acid neutralization and streamwater chemistry gradients in upland glaciated catchments, northeastern United States. Frontiers in Earth Science, 7, 63.

Bailey, S. W., Brousseau, P. A., McGuire, K. J., & Ross, D. S. (2014). Influence of landscape position and transient water table on soil development and carbon distribution in a steep, headwater catchment. Geoderma, 226, 279-289.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrological sciences journal, 24(1), 43-69.

Bourgault, R. R., Ross, D. S., & Bailey, S. W. (2015). Chemical and morphological distinctions between vertical and lateral podzolization at Hubbard Brook. Soil Science Society of America Journal, 79(2), 428-439.

Bourgault, R. R., Ross, D. S., Bailey, S. W., Bullen, T. D., McGuire, K. J., & Gannon, J. P. (2017). Redistribution of soil metals and organic carbon via lateral flowpaths at the catchment scale in a glaciated upland setting. Geoderma, 307, 238-252.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., ... & Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. Geoscientific model development, 8(7), 1991-2007.

Dunnington, D., Vanderhaeghe, F., Caha, J., & Muenchow, J. (2023). R package qgisprocess: use QGIS processing algorithms. Version 0.1.0. https://github.com/r-spatial/qgisprocess/

Fraser, O.L., K.J. McGuire, and S.W. Bailey. 2022. Hubbard Brook Experimental Forest: 1 meter LiDAR-derived and Hydro-enforced Digital Elevation Models, 2012 ver 2. Environmental Data Initiative. https://doi.org/10.6073/pasta/dcab665da20b8e75a57506b19f90262a (Accessed 2024-03-11).

Fraser, O. L., Bailey, S. W., Ducey, M. J., & McGuire, K. J. (2020). Predictive modeling of bedrock outcrops and associated shallow soil in upland glaciated landscapes. Geoderma, 376, 114495.

Fraser, O., K. McGuire, and S. Bailey. 2019. Hubbard Brook Experimental Forest: 5 meter LiDAR-derived Topographic Metrics, 2018 ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/a24417c8e5a0dc97e35c114e3a1518f2 (Accessed 2024-03-15).

Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resources Research, 39(12).

Gannon, J. P., Bailey, S. W., & McGuire, K. J. (2014). Organizing groundwater regimes and response thresholds by soils: A framework for understanding runoff generation in a headwater catchment. Water Resources Research, 50(11), 8403-8419.

Gannon, J. P., Bailey, S. W., McGuire, K. J., & Shanley, J. B. (2015). Flushing of distal hillslopes as an alternative source of stream dissolved organic carbon in a headwater catchment. Water Resources Research, 51(10), 8114-8128.

Gillin, C. P., Bailey, S. W., McGuire, K. J., & Gannon, J. P. (2015a). Mapping of hydropedologic spatial patterns in a steep headwater catchment. Soil Science Society of America Journal, 79(2), 440-453.

Gillin, C. P., Bailey, S. W., McGuire, K. J., & Prisley, S. P. (2015b). Evaluation of LiDAR-derived DEMs through terrain analysis and field comparison. Photogrammetric Engineering & Remote Sensing, 81(5), 387-396.

Guisan, A., Weiss, S. B., & Weiss, A. D. (1999). GLM versus CCA spatial modeling of plant species distribution. Plant ecology, 143, 107-122.

Hengl, T., & MacMillan, R. A. (2019). Predictive soil mapping with R. OpenGeoHub Foundation: Wageningen, The Netherlands, 227-273.

Hijmans R (2023). _raster: Geographic Data Analysis and Modeling_. R package version 3.6-26, https://rspatial.org/raster.

Hjerdt, K. N., McDonnell, J. J., Seibert, J., & Rodhe, A. (2004). A new topographic index to quantify downslope controls on local drainage. Water resources research, 40(5).

Lindsay, J. B. (2016). Whitebox GAT: A case study in geomorphometric analysis. Computers & Geosciences, 95, 75-84.

Møller, A. B., Beucher, A. M., Pouladi, N., & Greve, M. H. (2019). Oblique geographic coordinates as covariates for digital soil mapping. SOIL Discussions, 2019, 1-20.

QGIS.org (2023). QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Roberts, D., Bahn, V., Ciuti, S., Boyce, M., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort,J., Schröder, B., Thuiller, W., Warton, D., Wintle, B., Hartig, F., & Dormann, C. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913-929.

Seibert, J., & McGlynn, B. L. (2007). A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. Water Resources Research, 43(4).

Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, S., Wadoux, A., Xiang, W. & Scholten, T. (2017). Uncertainty-guided sampling to improve digital soil maps. Catena, 153, 30-38.

Travis, M. R. (1975). VIEWIT: computation of seen areas, slope, and aspect for land-use planning (Vol. 11). Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station.

Villars, T. R., Bailey, S. W., & Ross, D. S. (2015). Four Soil Orders on a Vermont Mountaintop—One-Third of the World's Soil Orders in a 2500-Square-Meter Research Plot. Soil Horizons, 56(6), 1-5.