Geoderma 235–236 (2014) 59–73

Contents lists available at ScienceDirect

Geoderma

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

Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the *k*-nearest neighbor method

Nicolas Mansuy *, Evelyne Thiffault, David Paré, Pierre Bernier, Luc Guindon, Philippe Villemaire, Vincent Poirier, André Beaudoin

Natural Resources Canada, Canadian Forest Service, P.O. Box 10380, Stn. Ste-Foy, Québec, QC G1V 4C7, Canada

ARTICLE INFO

Article history:
Received 3 March 2014
Received in revised form 5 June 2014
Accepted 30 June 2014
Available online 7 July 2014

Keywords:
Climate
Raster
SCORPAN
Soil texture
Topography

ABSTRACT

Large-scale mapping of soil properties is increasingly important for environmental resource management. While forested areas play critical environmental roles at local and global scales, forest soil maps are typically at low resolution. The objective of this study was to generate continuous national maps of selected soil variables (C, N and soil texture) for the Canadian managed forest landbase at 250 m resolution. We produced these maps using the *k*NN method with a training dataset of 538 ground-plots from the National Forest Inventory (NFI) across Canada, and 18 environmental predictor variables. The best predictor variables were selected (7 topographic and 5 climatic variables) using the Least Absolute Shrinkage and Selection Operator method. On average, for all soil variables, topographic predictors explained 37% of the total variance versus 64% for the climatic predictors. The relative root mean square error (RMSE%) calculated with the leave-one-out cross-validation method gave values ranging between 22% and 99%, depending on the soil variables tested. RMSE values < 40% can be considered a good imputation in light of the low density of points used in this study. The study demonstrates strong capabilities for mapping forest soil properties at 250 m resolution, compared with the current Soil Landscape of Canada System, which is largely oriented towards the agricultural landbase. The methodology used here can potentially contribute to the national and international need for spatially explicit soil information in resource management science.

1. Introduction

Canada's forests span over 390 million ha of land, representing 10% of the world's forest cover and 30% of the world's boreal forest; mapping Canada's forest soil is therefore a huge challenge. Spatially explicit knowledge of Canada's soil properties is necessary not only for meeting national forest management needs but also for supporting participation in international environmental programs and studies.

In Canada, provinces and territories have a fiduciary responsibility for the stewardship of the natural resources within their administrative borders. Provincial forest inventory mapping programs follow standards and definitions specific to their respective jurisdictions, and soil information provided by these programs is difficult to harmonize across the country. Since the early 1900s, soil surveys have also been conducted for Canada's southern areas by a variety of federal and provincial agencies at scales ranging from 1:20,000 to 1:250,000 (Geng et al., 2010). Soil surveys traditionally focused on potential agricultural land for promoting the development of the most populated areas. Since the early 1980s, national soil maps and accompanying databases have been provided by the Soil Landscapes of Canada (SLC) as a product of the Canadian Soil Information Service (CanSIS), a component of Agriculture and Agri-Food Canada (Schut et al., 2011). The SLC database consists of 12,728 multi-component map units, with multiple taxonomic soil classes generalized from detailed soil surveys (Geng et al., 2010).

Although the latest public release of SLC version 3.2 (SLC, 2010) can be used for a variety of broad-scale spatial modeling applications, its main focus remains agricultural land with a relatively coarse resolution (1:1,000,000) using a vector mapping system (polygons) (Schut et al., 2011); no uncertainty or error assessment is provided. The greatest coverage of the Canadian commercial forest landbase can be found in earlier versions of SLC and has not been updated. The later SLC version provides coverage for areas with a mix of forest and agricultural lands (e.g. Atlantic Provinces). However, soil polygons in the forest landbase are very large and imprecise, and are therefore of limited use for forest policy and management decision-making.

There is a worldwide need for qualitative and spatial soil information for environmental monitoring and resource management (Hartemink et al., 2008; Lagacherie and McBratney, 2006; Sanchez et al. 2009). Soil is largely regarded as the foundation underlying a forest's capacity to provide environmental services; understanding, quantifying and monitoring soil patterns in relation to ecosystem health are therefore essential (Grunwald, 2009). Moreover, recent studies have shown how forest soil information can be used as predictive indicators of site

* Corresponding author.
E-mail address: Nicolas.Mansuy@RNCan-NRCan.gc.ca (N. Mansuy).

http://dx.doi.org/10.1016/j.geoderma.2014.06.032

suitability/sensitivity to intensive forest management practices such as biomass harvesting (Hazlett et al., 2014; Thiffault et al., 2014). Such functional linkage to forest nutrition and management increases the need for reliable and precise maps of soil properties in commercial forests. However, sparse or missing data is a common obstacle in soil mapping, and may lead to misrepresentations of soil characteristics and ultimately to inadequate or inappropriate resource management actions. In response to this situation, scientists around the world are increasingly turning to digital soil mapping and modeling approaches (DSMM) instead of conventional soil surveys (Grunwald, 2009; McBratney et al., 2003). DSMM is a broad class of soil map production methods that capitalizes on the availability of ancillary environmental predictive variables to generate maps of soil variables using a limited number of soil data. These methods are very explicit relative to conventional soil mapping. The predictive variables and their respective contributions are identified, error assessments are produced and results are repeatable.

The goal of this study was therefore to produce standardized grid maps of forest soils at a scale relevant for strategic-level reporting and decision-making at regional to national scales featuring selected soil variables (C, N, and soil texture) for upland forests across the Canadian commercial forest landbase using the DSMM approach. The overall aim was to define and test a methodology for soil mapping at 250 m resolution within the Canadian managed forest, which could then be adapted to meet other needs. The specific objectives were to: 1) identify the best soil covariates (i.e., environmental predictive variables) for spatially predicting soil variables in the forest floor and the top mineral horizons (to a 15 cm depth), 2) run the *k*-nearest neighbor method (*k*NN) for interpolating point source forest soil data across landscapes, and 3) evaluate the output of predicted (imputed) soil maps by performing internal validation using a leave-one-out method, and by comparing them with independent soil databases for the area of interest.

2. Materials and methods

2.1. Mapping area

The study area consists of 290 million ha of managed forests within seven ecozones of Canada illustrated in Fig. 1: Boreal Shield (BS), Atlantic Maritime (AM), Pacific Maritime (PM), Montane Cordillera (MC), Boreal Plains (BP), Boreal Cordillera (BC), and Taiga Plains (TP). The ecozones are at the top level of the Ecological Framework of Canada which defines the ecological mosaic of the country on a subcontinental scale (Ecological Stratification Working Group, 1996). The analysis was limited to the forested Canadian ecozones as defined by the National Forest Inventory and that have available and exploitable georeferenced soil legacy data. Therefore, Northern Taiga, the Arctic, the Prairies, the Mixedwood Plains and the Hudson Plains were not mapped. Moreover, this analysis only focused on upland forests and therefore did not cover areas dominated by agricultural land (according to the SLC classification) and wetlands (Fig. 1).

2.2. Data and analysis

In this study, a *k*NN algorithm was used to populate each cell (pixel) of a raster map corresponding to the study area with soil information, using georeferenced point source forest soil data. The *k*NN method is a non-parametric, multivariate approach to imputing observations or combinations of observations from sampling units, i.e., the reference set, to estimation or mapping units, i.e. the target set using

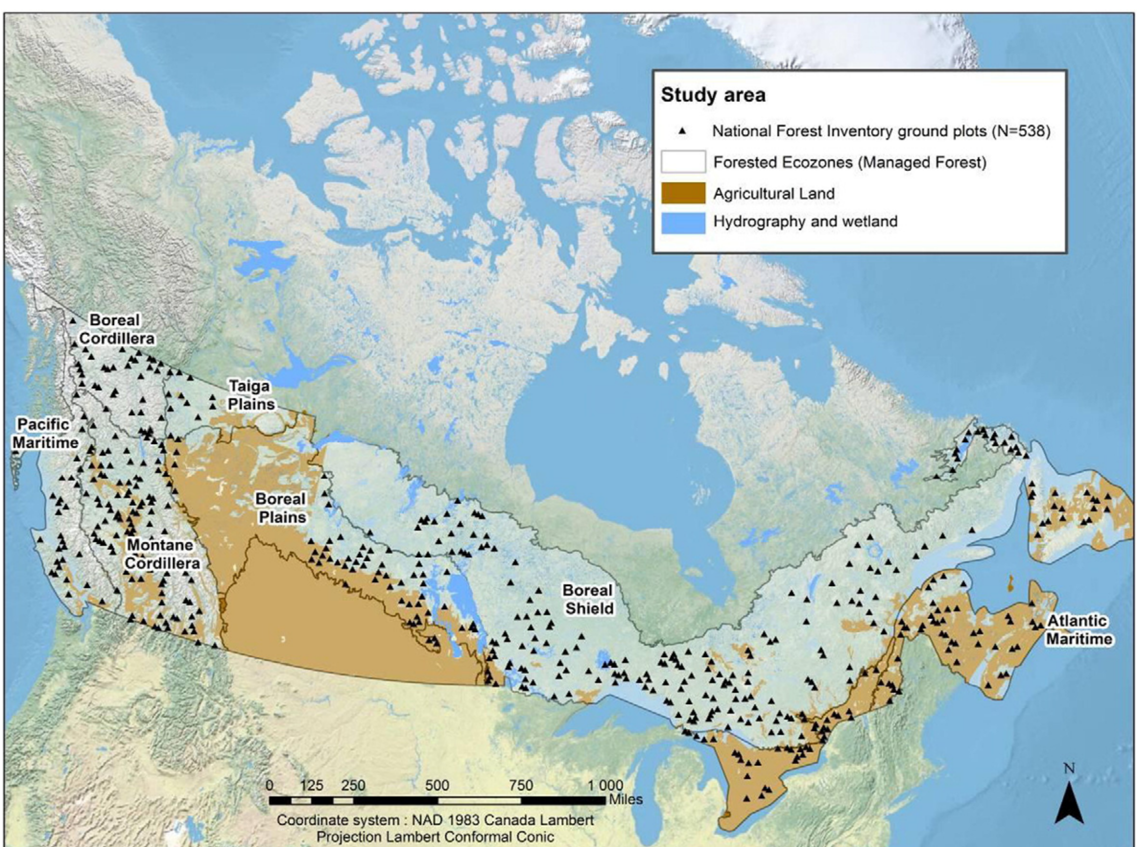

Fig. 1. National Forest Inventory (NFI) ground-plot data (538 points) across the Canadian forested ecozones. Areas dominated by agricultural land and wetlands were masked.