MAPPING FOREST SOIL ORGANIC MATTER ON NEW JERSEY'S COASTAL PLAIN

Brian J. Clough, Edwin J. Green, and Richard G. Lathrop[1]

Abstract.—Managing forest soil organic matter (SOM) stocks is a vital strategy for reducing the impact of anthropogenic carbon dioxide emissions. However, the SOM pool is highly variable, and developing accurate estimates to guide management decisions has remained a difficult task. We present the results of a spatial model designed to map soil organic matter for all forested land in the Coastal Plain physiographic province of New Jersey. SOM stocks from 60 sampling locations, distributed across the region in a stratified random design based on vegetation type and drainage class, were used in a kriging model that incorporated several indices derived from Landsat Thematic Mapper data as predictor variables. This model reduced mean squared error at validation plots (n=26) by 10 to 23 percent when compared to kriging models that did not use a predictor variable. Our results suggest that this approach, combining SOM inventory and remote sensing data in a geostatistical framework, is a useful method for reducing uncertainty in forest SOM estimates.

INTRODUCTION

Forest inventory data have been an important tool in estimating forest carbon (C) stocks (Birdsey 1992). The widespread availability of spatially explicit, plot-level forest inventory data has allowed landscape ecologists to employ a variety of spatial modeling techniques to calculate forest C stocks at the landscape scale. Much progress has been made in estimating aboveground carbon, but belowground carbon storage is still poorly documented.

Combining geostatistical analysis with plot inventory data and remotely sensed covariates may represent an effective method for mapping soil organic matter distribution. In geostatistics, the spatial covariance among estimates at sampling locations is quantified, and statistical approaches such as variogram analysis are used to model the spatial pattern of the variable of interest (Isaaks and Srivastava 1989). The creation of a spatially explicit model of covariance provides a quantitatively rigorous framework for interpolating values across response surfaces. The results of such interpolation models are "rasterized," making them ideal for up-scaling plot-level data to broad scales, and they avoid the issues introduced by spatial aggregation and *a priori* assumptions in other approaches to scaling SOM.

The development of metrics for inferring forest soil organic matter stocks from Landsat imagery presents a fruitful line of research, and a reliable methodology would be of considerable interest to biogeochemists and climate change policy-makers. Formation of forest soil organic matter (SOM) is controlled by a variety of factors, including forest cover type, topography, and disturbance (Chapin et al. 2002). The widespread availability of landscape-scale remote sensing data, largely funded by governmental agencies such as the U.S. Geological Survey (USGS), provides researchers with tools to collect data on such factors at broad scales.

[1] Graduate Assistant (BJC) and Professors (EG and RGL), Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers University, 14 College Farm Rd., New Brunswick, NJ 08901. BJC is corresponding author: to contact, email at bclough84@gmail.com.

In this study, we report on a model to interpolate SOM across forests of the Coastal Plain region of New Jersey. The model incorporates soil inventory and normalized difference vegetation index (NDVI) data to map SOM across a grid covering the entire study region. Kriging with external drift (KED), also referred to as “kriging with a trend model” in the literature (Goovaerts 1999), was used as the framework for incorporating secondary information into the SOM interpolation. Our study has three objectives: (1) Establish a relationship between SOM and NDVI to validate the latter as a predictor variable for mapping SOM distribution, (2) Demonstrate that incorporating NDVI in an interpolation model reduces uncertainty of prediction estimates relative to a model that does not incorporate any secondary information, and (3) Apply our model to generate a map of SOM distribution for New Jersey’s Coastal Plain physiographic province.

STUDY AREA

This study was conducted on the Coastal Plain physiographic province of New Jersey. Three major upland forest communities dominate the region: (1) *Pinus rigida* Mill. forest, (2) *Quercus* L. spp. forest, and (3) mixed communities that span a gradient between the two pure types. On the inner coastal plain, these communities mix with other hardwood species such as *Fagus grandifolia* Ehrh. and *Carya* Nutt. species. Forested wetlands are common along river courses or in low areas. Most of these wetlands are hardwood swamps dominated by *Acer rubrum* L., *Liquidambar styraciflua* L., and *Nyssa sylvatica* Marsh. However, forested *Sphagnum* bogs with pure stands of *Chamaecyparis thyoides* (L.) B.S.P. are also present across the landscape.

Soils in the region are largely typic Hapludults and Quartzisappamments of marine or alluvial origin (Tedrow 1986). Podzolization occurs in some soils, owing to litter inputs rich in tannins and other recalcitrant organic compounds (Tedrow 1998). Soils range from very poorly to excessively drained and are primarily sandy in texture. However, clayey and mucky soils are frequent in wet areas. Mean annual temperature is 54 °C and average annual rainfall is 1055 mm.

METHODS

Eighty-six plots were established throughout all forests on New Jersey’s Coastal Plain, using a stratified random sampling design based on dominant forest type and drainage class. Soil was collected within each plot at three depths: 0-10 cm, 10-20 cm, and 20-30 cm. At each depth interval bulk density was sampled using the core method (Blake and Hartge 1986), and a second sample was taken for laboratory analysis. The analytical samples were air dried for at least 48 hours, while the bulk density samples were dried at 105 °C. Both were sieved to 2 mm, and the fine fraction material was ground into powder with a mortar and pestle and homogenized. Soil organic matter content was estimated using loss-on-ignition at 400 °C for 24 hours. Soil organic matter stock (t/ha) was then calculated for each plot using the calculated bulk density and gravimetric soil organic matter content data.

Cloud-free Landsat TM scenes (courtesy of USGS) were extracted for a single date during the study, July 14, 2011, and tiled into a mosaic of the study region. We generated raster files for two indices calculated from the Landsat data: NDVI and tasseled cap band 2 (TC2). Both are related to photosynthetic activity, and have been shown to correlate with net primary productivity (Asrar et al. 1984), so it was reasonable to expect a relationship with SOM. Values of NDVI were extracted for all 86 sampling plots, and the complete raster files were collated and retained to interpolate SOM across the study region.

The 86 plots were randomly divided into a prediction set (60 plots) and a validation set (26 plots). To test if incorporating remote sensing covariates as predictor variables reduces uncertainty, we compared two univariate models, universal kriging (UK) and ordinary kriging (OK) to three multivariate models:

NDVI only (NDVI), tasseled cap only (TC), and both NDVI and tasseled cap data (NDVI+TC). Multivariate interpolation of SOM was accomplished using a universal kriging procedure with a trend model that allows for dependency between the response variable and a chosen set of predictors. This approach, also referred to as ”kriging with external drift” (KED), derives predictions of the response variable by extending the covariance matrix of the kriging system to incorporate the predictor variables (Goovaerts 1999). Each model was used to interpolate SOM on 26 validation plots, and the sum of squared error (SSE) between predicted and observed values of SOM was computed for each model from 50 simulations.

To ensure that the composition of the modeling data set did not overly influence model results, a new set of validation plots was randomly selected at the beginning of each trial. Model performance was assessed by comparing the mean SSE of each model for all 50 trials. The results of the best kriging model were used to interpolate values of SOM density for all cells in a 90 m^2 grid covering the full extent of forested land on New Jersey's Coastal Plain. All geostatistical analysis was accomplished in the R statistical computing environment (R Development Core Team 2008), using the gstat package (Pebesma 2004).

Table 1.—Mean sum of squared error (SSE) of each model, and percent reduction by the best performing model (NDVI).

	Mean SSE	Reduction in mean SSE when compared to "best" model (percent)
Univariate models		
Universal kriging (UK)	1984734	0.17
Ordinary kriging (OK)	2118974	0.22
Multivariate (KED) models		
NDVI only (NDVI)	1657018	0
TC only (TC)	1883040	0.12
Both predictors (NDVI+TC)	1686912	0.02

RESULTS

All three multivariate models (NDVI, TC, NDVI+TC) reduced mean sum of squared error by at least 5 percent when compared to the two univariate models. The model which used NDVI as the only predictor had the lowest mean SSE while the ordinary kriging model had the highest SSE (Table 1). NDVI reduced mean SSE by 17 and 22 percent, respectively, when compared to UK and OK. NDVI also outperformed TC, reducing mean SSE by 12 percent, and provided slight improvement when compared to the full model (NDVI+TC). These results are probably explained by the higher correlation of SOM stock with the normalized difference vegetation index data than with the tasseled cap index (Fig. 1). In general, our results demonstrate that incorporating remote sensing covariates as predictor variables reduces uncertainty in regional SOM estimates by at least 17 percent.

The modeling data set that produced the lowest squared error was used for generating a map of SOM distribution for the entire study region. Our fitted variogram suggests a range of spatial covariance among data points of approximately 12,500 m (Fig. 2). However, note that considerable variation is unaccounted for by this model. The spatial pattern, owing to the fairly small sample size in the modeling data set (n=60), is not well defined. Additionally, the large nugget effect of the variogram indicates significant microscale spatial variability below the resolution of our sampling regime. Due to the weakly fitting variogram model, we observed only weak correlation between observed and predicted results at the 26 validation plots (Fig. 3). Although this model performs better than either univariate model ($r = 0.15$ for UK and $r = 0.05$ for OK using the same data set), there is still significant uncertainty owing to the small sample size in the modeling data set.

The model tends to predict higher SOM density on the interior portions of the Coastal Plain and declining SOM stocks to the southeast (Fig. 4). These predictions are consistent with our knowledge of the region, where the inland soils tend to be more nutrient rich and less excessively drained.

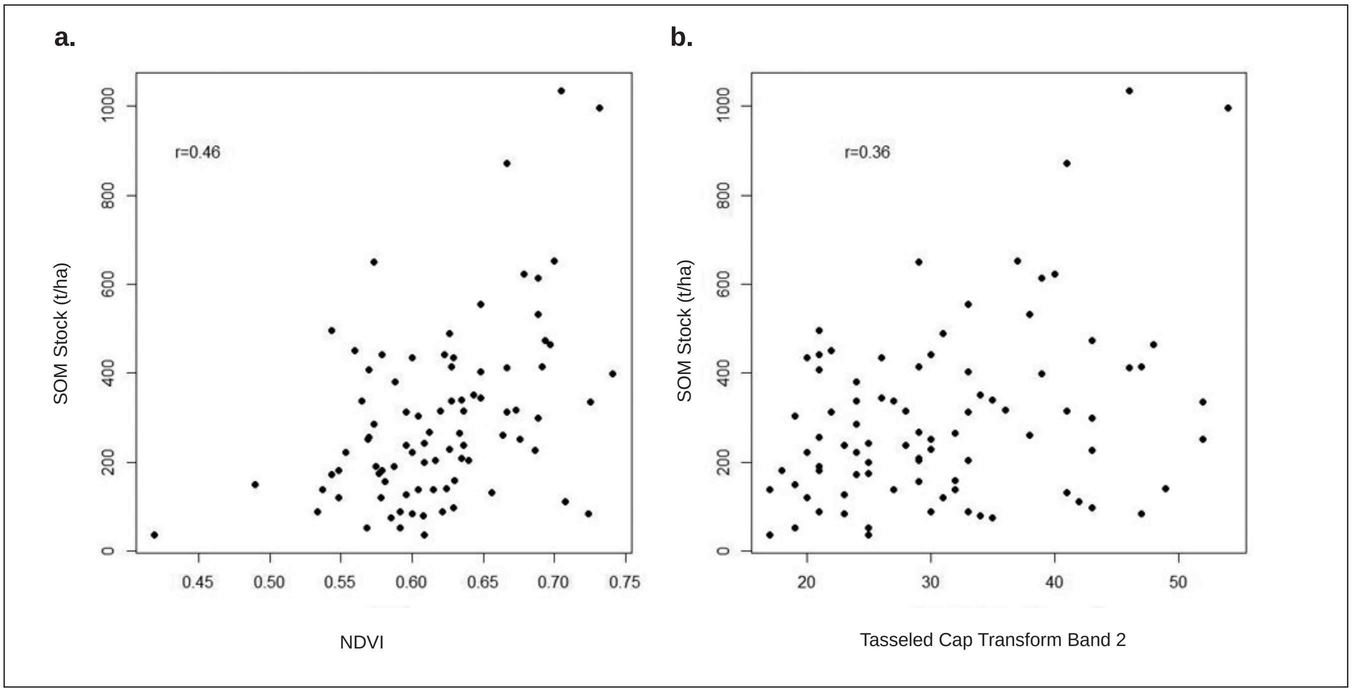

Figure 1.—Relationship between (a) soil organic matter stock (SOM) and normalized difference vegetation index (NDVI), and (b) SOM and tasseled cap band 2 (TC) at all 86 sampling locations.

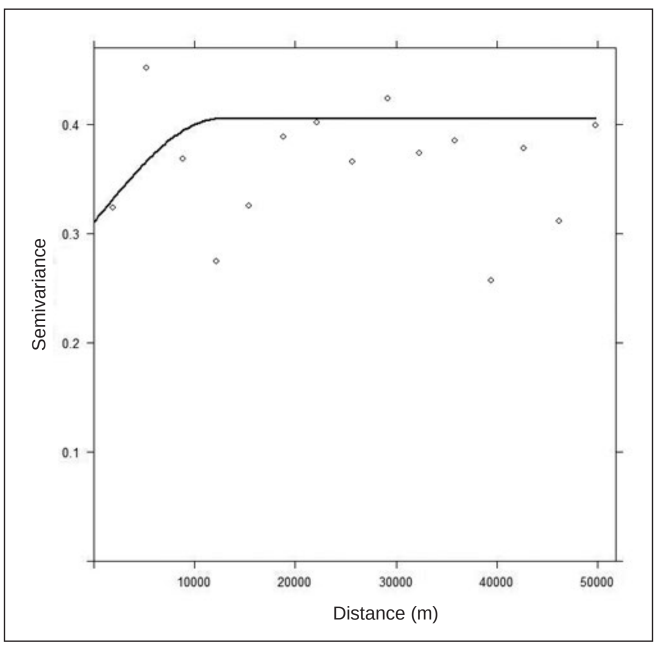

Figure 2.—Semivariogram for the 60 modeling plots. The open circles represent the binned empirical values of semivariance. The solid line is a spherical variogram model fitted with a restricted maximum likelihood optimization routine (partial sill = 0.09, range=12,500, nugget = 0.31).

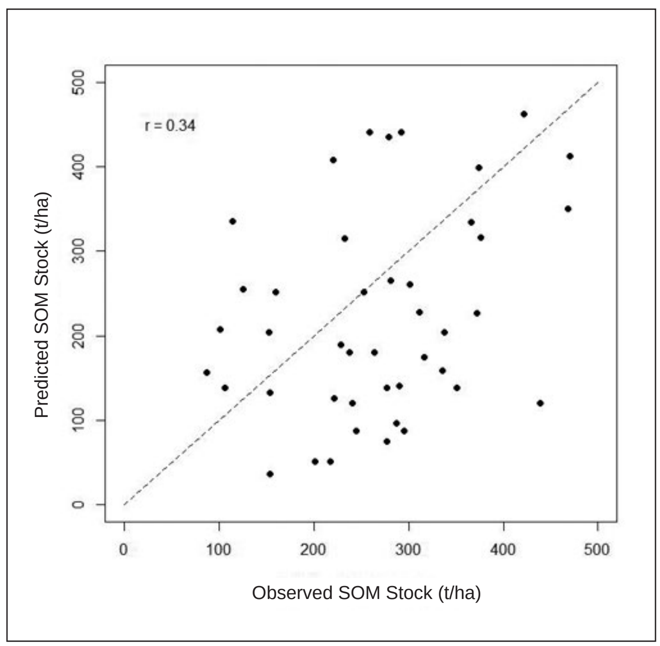

Figure 3.—Observed vs. predicted soil organic matter content at 26 independent validation plots. The dotted line represents a 1:1 relationship between the two variables.

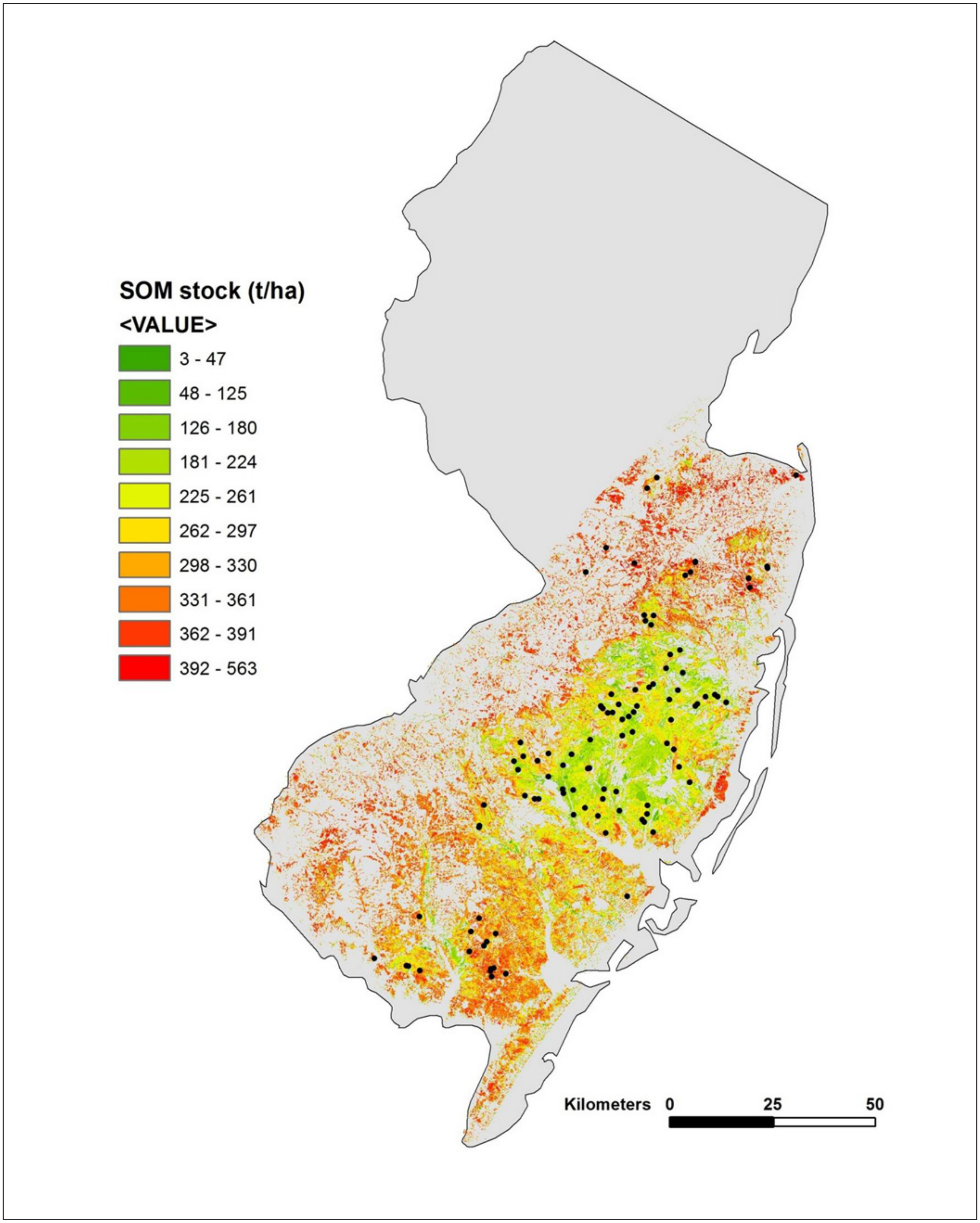

Figure 4.—Result of a KED interpolation for a 90-m^2 grid covering all forests in New Jersey's Coastal Plain. Our study region is defined by the outline. Note that the model predicts higher SOM stocks on the inner portion of the Coastal Plain, where soils tend to be less nutrient- poor.

DISCUSSION

Including soil sampling in large-scale forest inventories, such as the Forest Inventory and Analysis Program, will greatly increase the spatial coverage and availability of data on belowground carbon stocks. Incorporating NDVI reduced uncertainty in forest SOM estimates by 17 to 23 percent with a relatively small set of modeling plots (n=60) despite considerable error between measured and predicted values at the validation plots. Large coordinated efforts to sample forest SOM, where sampling density is increased and the range of variation in SOM can be better characterized by the data set, would provide improved performance. Our model framework will likely be an effective means for modeling SOM on a variety of forested landscapes, though appropriate predictor variables will have to be selected on a case-by-case basis. In general, predictor variables should be selected based on local knowledge of the environmental factors that are most likely to drive variation in SOM stocks.

Combining soil inventories with remotely sensed data and geostatistical analysis represents an effective framework for quantifying the SOM pool at broad spatial scales. Interpolation methods such as kriging, which provide a statistically rigorous framework for defining spatial patterns, provide an appealing alternative to traditional approaches for incorporating belowground C into landscape-scale forest carbon budgets. These methods are especially useful when predictor variables derived from rasterized remote sensing data sets are incorporated into the models. Our results confirm that such an approach constitutes an effective method for reducing uncertainty of soil organic matter stock estimates and would be a useful addition to forest inventory projects.

ACKNOWLEDGMENTS

We would like to thank Dr. Charles Perry for his advice and guidance while developing this project. We would also like to acknowledge the U.S. Department of Agriculture and the U.S. Forest Service for their support.

LITERATURE CITED

Asrar, G.; Fuchs, M.; Kanemasu, E.T.; Hatfield, J.L. 1984. **Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat.** Agronomy Journal. 76: 300-306.

Birdsey, R.A. 1992. **Carbon storage and accumulation in United States forest ecosystems.** Gen. Tech. Rep. WO-59. Washington, DC: U.S. Department of Agriculture, Forest Service, Washington Office. 51 p.

Blake, G.R.; Hartge, K.H.; 1986. **Bulk density.** In: Klute, A., ed. Methods of soil analysis Part 1. Physical and mineralogical methods. Madison, WI: American Society of Agronomy, Inc.; Soil Science Society of America: 377-382.

Chapin, F.S.; Matson, P.A.; Mooney, H.A. 2002. **Principles of terrestrial ecosystem ecology.** New York, NY: Springer Science & Business Media Inc. 436 p.

Goovaerts, P. 1997. **Geostatistics for natural resources evaluation.** New York, NY: Oxford University Press. 496 p.

Isaaks, E.H.; Srivastava, M. 1989. **An introduction to applied geostatistics.** Oxford, UK: Oxford University Press. 561 p.

Pebesma, E.J. 2004. **Multivariable geostatistics in S: the gstat package.** Computers & Geosciences. 30: 683-691.

R Development Core Team. 2008. **R: a language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing. Available at www. R-project.org. [Date accessed unknown].

Tedrow, J.C.F. 1986. **The soils of New Jersey.** New Jersey Agricultural Experiment Station publication no. A-15134-1-82.

Tedrow, J.C.F. 1998. **Development of the Pine Barrens soils.** In: Forman, R., ed. Pine Barrens: ecosystem and landscape. New Brunswick, NJ: Rutgers University Press: 61-78.